En Introduktion til Statistik

Bind 1B

Knut Conradsen

7. udgave

Lyngby 1999

IMM

Indhold

Kapitel 3

Hypoteseprøvning

Vi skal i dette kapitel beskæftige os med den anden hovedform for statistisk inferens ved siden af estimationsteorien, nemlig hypoteseprøvningen eller testteorien.

3.1 Generel problemstilling og metode

3.1.1 Indledning og definitioner

Vi introducerer de noget vanskeligt tilgængelige begreber ved hjælp af et gennemgående eksempel og giver så de præcise definitioner i afsnittets slutning.

Lad os antage, at vi har 2 giftblandinger A og B, der kun afviger på koncentrationerne, som er,

A: 400 mg giftstof/liter
B: 380 mg giftstof/liter.

Nu er mærkaterne forsvundet fra en beholder, der står der, hvor der sædvanligvis er anbragt beholdere med koncentrationen B. Da vi står over for at skulle anvende gift af en bestemt koncentration, vil vi - for at være "sikre" på ikke at lave fejl - undersøge, om det overhovedet kan antages, at beholderen er af koncentration A. Hvis ikke, må vi kunne gå ud fra, at den er af koncentration B.

Vi tager nu 36 prøver af giftstoffet fra beholderen og måler koncentrationen med en

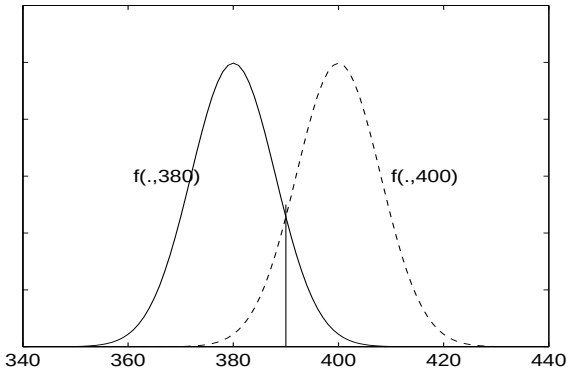
gennemprøvet teknik; vi ved således, at udfaldet af målingerne kan beskrives ved stokastiske variable $X_1, \cdots, X_{36}$, der er indbyrdes uafhængige og $\mathrm{N}(\mu, 48^2)$-fordelte, hvor μ er den sande koncentration og $\sigma = 48$ her antages at være kendt fra tidligere stikprøver.

Vores problemstilling omformuleres nu til

Hypotese:	H_0:	μ	$=$	400
Alternativ:	H_1:	μ	$=$	$380.$

Det må da være rimeligt at basere vor afgørelse på værdien af $\bar{X}$, gennemsnittet af målingerne.

Fordelingen af $\bar{X}$ vil i de 2 situationer være $\mathrm{N}(400, 48^2/36)$ henholdsvis $\mathrm{N}(380, 48^2/36)$, jvf. sætning 2.6, p. 226.



Lad os antage, at vi har målt $\bar{X} = \bar{x} = 396$.

En rimelig beslutningsregel vil være at acceptere H_0 for $\bar{x} \geq$ vist c og forkaste H_0 for $\bar{x} < c$. De tilsvarende områder $\{\bar{x} \geq c\}$ og $\{\bar{x} < c\}$ kaldes henholdsvis **acceptområdet** og **det kritiske område**.

Intuitivt indlysende forekommer det at vælge $c = 390$. Lad os prøve at aflæse nogle konsekvenser af dette. Hvilke fejl kan man begå i en sådan procedure? Der er 2 fundamentale:

Fejl af type I:	at forkaste en sand hypotese.
Fejl af type II:	at acceptere en falsk hypotese.

For at understrege forskellen mellem de 2 typer nævner vi, at i en retssag, hvor hypotese og alternativ er eller bør være

H_0: uskyldig $\quad H_1$: skyldig

er

Fejl af type I:	dømme uskyldig
Fejl af type II:	frikende skyldig

Skal man vurdere rimeligheden af et valgt c, må det ske ved vurdering af sandsynligheden for fejl af de 2 typer.

Vi sætter

$$\begin{aligned} \alpha &= \mathrm{P}\{\text{fejl af type I}\} \\ \beta &= \mathrm{P}\{\text{fejl af type II}\} \end{aligned}$$

og har da i eksemplet med giftstofferne

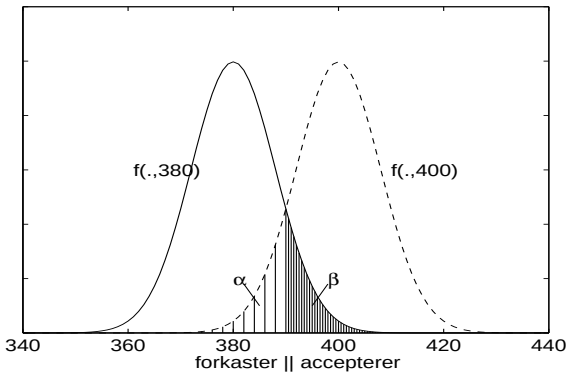
$$\begin{aligned} \alpha &= \mathrm{P}\{\text{forkaste } \mu = 400 | \mu = 400\} \\ &= \mathrm{P}\{\bar{X} < 390 | \mu = 400\} \\ &= \mathrm{P}\{\frac{\bar{X} - 400}{48/\sqrt{36}} < \frac{390 - 400}{48/\sqrt{36}} | \mu = 400\} \\ &= \mathrm{P}\{\mathrm{N}(0,1) < -10/8\} \\ &= 0.1056. \end{aligned}$$

Helt tilsvarende

$$\begin{aligned} \beta &= \mathrm{P}\{\text{acceptere } \mu = 400 | \mu = 380\} \\ &= \mathrm{P}\{\bar{X} > 390 | \mu = 380\} \\ &= \mathrm{P}\{\frac{\bar{X} - 380}{48/\sqrt{36}} < \frac{390 - 380}{48/\sqrt{36}} | \mu = 380\} \\ &= \mathrm{P}\{\mathrm{N}(0,1) > 10/8\} \\ &= 0.1056 \end{aligned}$$

At de to sandsynligheder er lige store, fremgår også umiddelbart af tegningen

Hvis man af en eller anden grund mener, at den fundne værdi for α er for stor, kan man ved at ændre på c gøre α mindre. Hvis man e.g. ønsker $\alpha = 5\%$, kan man finde det



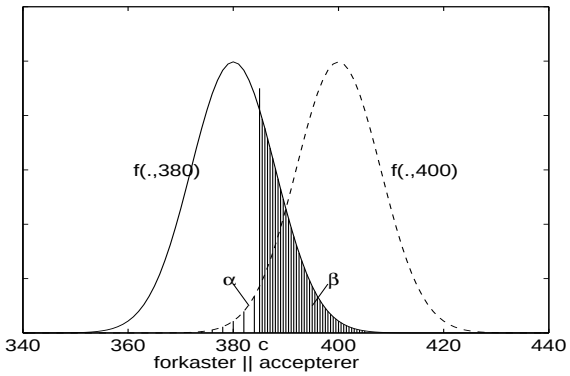
tilsvarende c som følger

$$\begin{aligned} 5\% \;=\; \alpha \;&=\; \mathrm{P}\{\text{forkaste } \mu = 400 | \mu = 400\} \\ &=\; \mathrm{P}\{\bar{X} < c | \mu = 400\} \\ &=\; \mathrm{P}\{\tfrac{\bar{X}-400}{48/\sqrt{36}} < \tfrac{c-400}{48/\sqrt{36}} | \mu = 400\} \\ &=\; \mathrm{P}\{\mathrm{N}(0,1) < \tfrac{c-400}{8}\} \end{aligned}$$

Ved hjælp af tabel fås da $\frac{c-400}{8} = -1.645$ eller $c = 400 - 13.36 = 386.84$. I dette tilfælde er

$$\begin{aligned} \beta \;&=\; \mathrm{P}\{\text{acceptere } \mu = 400 | \mu = 380\} \\ &=\; \mathrm{P}\{\frac{\bar{X} - 380}{48/\sqrt{36}} \geq \frac{c - 380}{8} | \mu = 380\} \\ &=\; \mathrm{P}\{\mathrm{N}(0,1) \geq \frac{386.84 - 380}{8}\} \\ &=\; 1 - \mathrm{P}\{\mathrm{N}(0,1) < 0.8553\} = 0.197 \end{aligned}$$

Grafisk kan disse udregninger tydeliggøres.



Generelt kan der naturligvis ikke siges noget om valg af $\mathbf{c}, \alpha, \beta$ etc. Dette afhænger af den aktuelle situation. Det er dog normalt - når man ikke har f.eks. økonomisk eller lignende kriterier for valg af α - da at vælge $\alpha = 1\%, 5\%$ eller 10%.

I kapitel 6 vil vi vise, hvorledes tilstedeværelsen af økonomiske kriterier for ens beslutningstagen kan føre til en bestemmelse af α.

Den næste dag står vi i en ny situation. En medhjælper har hældt vand i en af beholderne på A-lageret, og vi er interesseret i at erfare, om det kan antages at være den, vi står med. Den rimelige hypotese og det rimelige alternativ må nu være

$$H_0: \quad \mu = 400 \qquad H_1: \quad \mu < 400$$

En fornuftig beslutningsregel er fremdeles

Acceptområde: $\{\bar{x} \geq c\}$
Kritisk område: $\{\bar{x} < c\}$.

Vi søger nu igen for givet c sandsynlighederne for fejl af type I og af type II. Vi har

$$\begin{aligned} \alpha &= \mathrm{P}\{\text{fejl af type I}\} \\ &= \mathrm{P}\{\bar{X} < c | \mu = 400\} \\ &= \mathrm{P}\{\mathrm{N}(0,1) < (c - 400)/8\} \end{aligned}$$

og

$$\begin{aligned} \beta &= \mathrm{P}\{\text{fejl af type II}\} \\ &= \mathrm{P}\{\bar{X} \geq c | \mu < 400\}. \end{aligned}$$

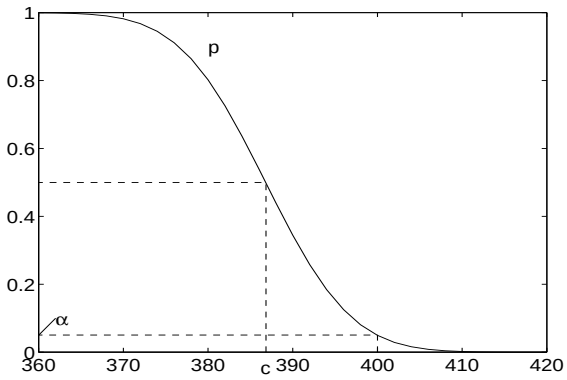
Denne kan ikke umiddelbart udregnes, idet $\bar{X}$'s fordeling afhænger af, hvilken værdi μ antager. Derfor defineres **styrken** $\mathrm{p}(\mu)$:

$$\mathrm{p}(\mu) = \mathrm{P}\{\text{forkaste hypotesen}|\mu\}$$

og i den aktuelle situation fås

$$\mathrm{p}(\mu) = \mathrm{P}\{\bar{X} < c|\mu\} = \Phi\left(\frac{c-\mu}{8}\right),$$

hvor Φ angiver fordelingsfunktionen for en $\mathrm{N}(0,1)$-fordeling, og vi ser, at $p(c) = \frac{1}{2}$, og $p(400) = \alpha$. I øvrigt er grafen

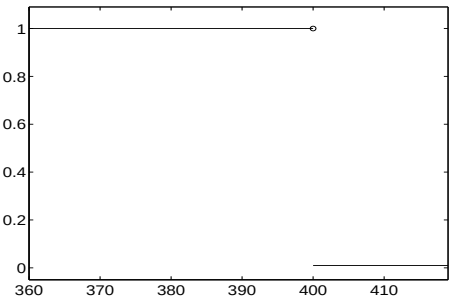


Man ser let, at $\lim_{\mu\to+\infty} p(\mu) = 0$ og $\lim_{\mu\to-\infty} p(\mu) = 1$.

Hvis vi fikserer $\alpha = 5\%$, kan vi finde det dertil hørende c:

$$\begin{array}{lll} 5\% = \Phi\left(\frac{c-400}{8}\right) & \Rightarrow & \frac{c-400}{8} = -1.645 \\ & \Rightarrow & c = 386.8. \end{array}$$

Det er åbenbart, at den ideelle styrkefunktion er



idet vi da ville have sandsynligheden 1 for at forkaste en falsk hypotese og sandsynligheden 0 for at forkaste en sand hypotese.

Hvorledes kan vi finde et test, hvis styrkefunktion nærmer sig denne? Intuitivt indlysende er det vel, at et test må få større "skelneevne", såfremt det baseres på flere observationer, hvorfor man jo kunne forsøge at gøre n større. Vi måler nu $X_1, \cdots, X_n$ og $\bar{X} \in \mathrm{N}(\mu, 48^2/n)$. Styrken bliver derfor

$$\mathrm{p}_n(\mu) = \mathrm{P}\{\bar{X} < c|\mu\} = \Phi\left(\frac{c-\mu}{48/\sqrt{n}}\right) = \Phi\left(\frac{c_n-\mu}{48/\sqrt{n}}\right),$$

hvor $c = c_n$ nu afhænger af n.

Hvis vi f.eks. ønsker $\alpha = 5\%$, fås

$$\tfrac{c-400}{48/\sqrt{n}} = -1.645 \quad \Rightarrow \quad c = 400 - 78.96/\sqrt{n}.$$

Styrken bliver altså

$$\mathrm{p}_n(\mu) \quad = \quad \Phi\left(\tfrac{400-78.96\sqrt{n}-\mu}{48/\sqrt{n}}\right) \quad = \quad \Phi\left(\tfrac{400-\mu}{48}\sqrt{n} - \tfrac{78.96}{48}\right)$$

Da $\Phi(\infty) = 1$ og $\Phi(-\infty) = 0$, fås

$$\mu > 400 \;\Rightarrow\; \lim_{n\to\infty} \mathrm{p}_n(\mu) = 0 \quad \wedge \quad \mu < 400 \;\Rightarrow\; \lim_{n\to\infty} \mathrm{p}_n(\mu) = 1,$$

d.v.s. at vi ved at tage tilstrækkelig mange observationer kan få en styrke, der ligger "vilkårligt tæt" ved den optimale.

Hvis vi f.eks. ønsker, at styrken i et punkt, f.eks. $\mu = 390$, skal være mindst 99%, får vi, idet vi stadig regner med $\alpha = 5\%$,

$$\mathrm{p}_n(400) = 5\% \quad \wedge \quad \mathrm{p}_n(390) = 99\%,$$

d.v.s.

$$\Phi\left(\tfrac{c-400}{48/\sqrt{n}}\right) = 5\% \quad \wedge \quad \Phi\left(\tfrac{c-390}{48/\sqrt{n}}\right) = 99\%$$

eller

$$\tfrac{c-400}{48/\sqrt{n}} = -1.645 \quad \wedge \quad \tfrac{c-390}{48/\sqrt{n}} = 2.326.$$

Disse ligninger omskrives til

$$c - 400 = -78.96\tfrac{1}{\sqrt{n}} \quad \wedge \quad c - 390 = 111.648\tfrac{1}{\sqrt{n}}.$$

Ved elimination af c fås

$$10 = (111.648 + 78.96)\frac{1}{\sqrt{n}} = 190.608\frac{1}{\sqrt{n}}$$

eller

$$\sqrt{n} = 19.0608,$$

d.v.s.

$$n = 363.3 \simeq \underline{\underline{364}}.$$

Indsættes dette resultat i ovenstående relation mellem n og c fås

$$c = 400 - \frac{78.96}{19.0608} \simeq \underline{\underline{395.86}}.$$

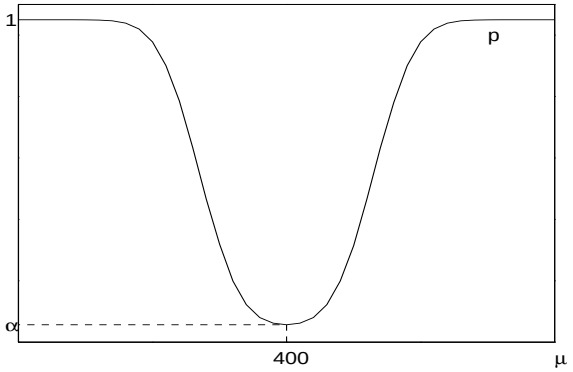
Hvis man havde haft et andet alternativ, f.eks.

$$H_0: \quad \mu = 400 \quad \Leftrightarrow \quad H_1: \quad \mu \neq 400,$$

var det kritiske område blevet af formen $\{\bar{X} < c_1\} \cup \{\bar{X} > c_2\}$.
I dette tilfælde var styrkefunktionen blevet

$$\begin{aligned} \mathrm{p}(\mu) &= \mathrm{P}\{\bar{X} < c_1|\mu\} + \mathrm{P}\{\bar{X} > c_2|\mu\} \\ &= \Phi\left(\frac{c_1 - \mu}{48/\sqrt{n}}\right) + 1 - \Phi\left(\frac{c_2 - \mu}{48/\sqrt{n}}\right). \end{aligned}$$

Vælges, som naturligt er, c_1 og c_2 symmetriske om 400, fås følgende graf



Her gælder det naturligvis også, at vi kan fastlægge værdien af p i forskellige punkter og bestemme det til disse værdier hørende n.

BEMÆRKNING 3.1 (BESTEMMELSE AF STIKPRØVESTØRRELSE (n)). Et almindeligt forekommende problem er at bestemme, hvor stor en stikprøve, der skal benyttes for at opnå en vis diskriminationsevne. Sæt f.eks., at der ønskes konstrueret et test med niveau α, og det samtidig kræves, at sandsynligheden for afvisning (styrken) skal mindst $1-\beta$ for et specificeret alternativ.

Et eksempel kunne være et test i Poisson-fordelingen. Lad $\{X_1, X_2, \ldots, X_n\}$, $X_i \in P(\lambda)$, være stikprøven. Lad hypotesen og alternativ være:

$$\begin{array}{lrcl} H_0: & \lambda & = & \lambda_0 \\ H_1: & \lambda & = & 5\lambda_0 = \lambda_1 \end{array}$$

Problemet kan da præciseres således:

Følgende relationer, som bestemmes ved hjælp af styrkefunktionen, skal være opfyldt:

$$\mathrm{P}\{\text{accept } H_0 | \lambda = \lambda_0\} \geq 1-\alpha$$

og

$$\mathrm{P}\{\text{afvis } H_0 | \lambda = \lambda_1\} \geq 1-\beta.$$

Som vi senere skal se, er det kritiske område $\sum_{i=1}^{n} X_i > c$. Bestem n.

Denne metode er illustreret i Eksemplerne 3.11-3.14 for normalfordelte observationer.

▼

Vi vil nu til sidst give de stringente matematiske definitioner på alle de begreber, vi har indført i de foregående eksempler (plus et par nye).

DEFINITION 3.1. En **statistisk hypotese** er et udsagn om fordelingen af en eller flere stokastiske variable. Hvis den statistiske hypotese fuldstændig fastlægger fordelingen, kaldes den en **simpel statistisk hypotese**, hvis ikke, kaldes den **sammensat**. ▲

Hvis $X_i \in \mathrm{N}(\mu, 1)$, $i = 1, \cdots, n$, er "$\mu = 400$" en simpel hypotese, og "$\mu < 400$" en sammensat hypotese.

DEFINITION 3.2. Et **test** er en beslutningsregel baseret på realiserede udfald af eksperimentet. Beslutningsreglen kan antage værdierne "accepterer hypotesen" og "forkast hypotesen". ▲

DEFINITION 3.3. Hvis der er givet en mængde C således, at vi

$$\begin{array}{lll} \text{forkaster}, & \text{hvis} & (x_1, \cdots, x_n) \in \mathrm{C} \\ \text{accepterer}, & \text{hvis} & (x_1, \cdots, x_n) \in \complement\mathrm{C} \end{array}$$

da kaldes C **det kritiske område** for testet og $\complement\mathrm{C}$ **acceptområdet**. ▲

Som vi har set i eksemplet, fastlægges det kritiske område ofte ved en relation som $\bar{x} \leq c$ etc. Da kaldes $\bar{X}$ vor teststørrelse. Vi formulerer det stringent i

DEFINITION 3.4. Hvis det kritiske område for et test er fastlagt ved en relation

$$\mathrm{t}(x_1, \cdots, x_n) \in C_1$$

hvor $\mathrm{t}(X_1, \cdots, X_n)$ er en stikprøvefunktion, kaldes $\mathrm{t}(X_1, \cdots, X_n)$ **teststørrelsen**. ▲

Lad nu fordelingen af $X_1, \cdots, X_n$ afhænge af den ukendte parameter θ. Parameterområdet er Ω (=mængden af mulige parameterværdier), og Ω_0 er en delmængde af Ω. I det følgende betragtes hypotesen

$$H_0: \quad \theta \in \Omega_0 \tag{3.1}$$

mod alternativet

$$H_1: \quad \theta \in \Omega \setminus \Omega_0 .$$

Vi har da

DEFINITION 3.5. Hvis vi som kritisk område for et test af hypotesen (3.1) anvender mængden C, da er **styrken** i θ for testet lig

$$\mathrm{p}(\theta) = \mathrm{P}\{(X_1, \cdots, X_n) \in \mathrm{C} | \text{ sand parameter} = \theta\} .$$

Afbildningen p kaldes **styrkefunktionen**. ▲

I det generelle tilfælde er hypotesen H_0 sammensat, således at vi ikke kan tale om sandsynligheden α for at begå en type I fejl, d.v.s. tale om sandsynligheden for at forkaste en sand hypotese. I stedet definerer vi begrebet niveau for et test.

DEFINITION 3.6. Ved **signifikansniveauet** (eller blot **niveauet**) α for et test forstås den maximale sandsynlighed for at forkaste hypotesen, når den er sand. ▲

Denne definition bliver langt mere overskuelig ved indførelse af styrkefunktionen p. Da har vi nemlig, at signifikansniveauet α er givet ved

$$\alpha = \sup_{\theta \in \Omega_0} \mathrm{p}(\theta), \tag{3.2}$$

d.v.s. at niveauet blot er supremum af styrkefunktionen, når H_0 er sand.

Vi skal ikke komme meget ind på optimalitetsegenskaber ved tests, men blot nævne to definitioner. De bygger begge på, at der er en bijektiv sammenhæng mellem et test og det dertil hørende kritiske område.

Vi betragter uafhængige stokastiske variable $X_1, \cdots, X_n$ med frekvensfunktion $\mathrm{f}(x;\theta)$. Vi har da

DEFINITION 3.7. Lad C være en delmængde af R^n. Da er C et **bedste kritisk område** af **størrelsen** α for et test af den simple hypotese
$H_0: \theta = \theta_0$ mod det simple alternativ $H_1: \theta = \theta_1$, såfremt

$$i) \qquad \mathrm{P}\{(X_1, \cdots, X_n) \in \mathrm{C} | H_0\} = \alpha,$$

og såfremt det for enhver delmængde A af R^n med $\mathrm{P}\{(X_1,\cdots,X_n)\in A|H_0\}=\alpha$, gælder

$$ii) \qquad \mathrm{P}\{(X_1,\cdots,X_n)\in \mathrm{C}|H_1) > \mathrm{P}\{(X_1,\cdots,X_n)\}\in A|H_1\}.$$

Vi siger endvidere, at det ved C definerede test er et **stærkeste test** af hypotesen H_0 med alternativet H_1. ▲

BEMÆRKNING 3.2. Definitionen går blot ud på, at man blandt alle tests af niveau α vælger det, der har den største styrke i θ_1, d.v.s. størst sandsynlighed for at forkaste hypotesen, når H_1 er sand, d.v.s. når H_0 er falsk. ▼

Hvis alternativet H_1 ikke længere er simpelt, men sammensat, har vi

DEFINITION 3.8. Det kritiske område C er et **uniformt bedste** kritisk område af størrelsen α for et test af den simple hypotese H_0 mod det sammensatte alternativ H_1, hvis C er et bedste kritisk område for test af H_0 mod enhver simpel hypotese i H_1. Vi siger, at det ved C definerede test er et **uniformt stærkeste test** (engelsk: UMP-test= uniformly most powerful test) med signifikansniveau α for test af H_0 mod H_1. ▲

Som nævnt skal vi ikke komme meget ind på teorien for uniformt stærkeste tests, men vi bemærker, at der ikke altid eksisterer et sådant for test af en simpel hypotese mod et sammensat alternativ. I næste afsnit skal vi se et enkelt eksempel på konstruktion af et uniformt stærkeste test.

Til sidst i afsnittet bemærker vi, at der er en sammenhæng mellem testteorien og intervalestimationsteorien. Denne sammenhæng kan f.eks. udtrykkes som i

SÆTNING 3.1. Lad $[\underline{t},\overline{t}]$ være et $(1-\alpha)$-konfidensinterval for parameteren θ, d.v.s.

$$\mathrm{P}\Big\{\theta\in\big[\underline{t}(X_1,\cdots,X_n),\overline{t}(X_1,\cdots,X_n)\big]\Big\}=1-\alpha.$$

Da er mængden

$$\mathrm{C}=\Big\{(x_1,\cdots,x_n)|\theta_0\notin\big[\underline{t}(x_1,\cdots,x_n),\overline{t}(x_1,\cdots,x_n)\big]\Big\}$$

kritisk område for et test af $H_0:\ \theta=\theta_0$ mod $H_1:\ \theta\neq\theta_0$ med signifikansniveau α.

BEMÆRKNING 3.3. Hvis man har de realiserede udfald $x_1, \cdots, x_n$, kan man altså teste H_0 ved at udregne $1-\alpha$ konfidensintervallet for θ og dernæst undersøge, om θ_0 er indeholdt i det. Hvis ja, accepteres hypotesen, hvis nej, forkastes den. Anderledes udtrykt: Konfidensintervallet består netop af de parameterværdier, der med de foreliggende data vil blive accepteret ved et test på niveau α. ▼

Bevis. Vi har

$$\begin{aligned} & \mathrm{P}\{\text{forkaste hypotesen}|\theta=\theta_0\} \\ = \; & \mathrm{P}\Big\{\theta_0 \notin \left[\underline{t}(X_1,\cdots,X_n), \overline{t}(X_1,\cdots,X_n)\right] |\theta=\theta_0\Big\} \\ = \; & 1-(1-\alpha) \\ = \; & \alpha, \end{aligned}$$

d.v.s. at testet ifølge definition 3.6 har niveauet α. ■

3.1.2 Testprincipper

I dette afsnit vil vi beskrive nogle retningslinier for konstruktion af teststørrelser og kritiske områder.

Vi bemærker først, at sætning 3.1 om sammenhængen mellem konfidensintervaller og kritiske områder i forbindelse med eksempel 2.23 side 262 direkte giver anledning til en mængde tests.

Vi giver dernæst en sætning, der handler om konstruktion af stærkeste tests i tilfældet med en simpel hypotese og et simpelt alternativ.

SÆTNING 3.2 (NEYMAN-PEARSON'S FUNDAMENTALE LEMMA). Lad der være givet stokastiske variable med simultan frekvensfunktion $\mathrm{f}(x_1, \cdots, x_n; \theta)$. Vi betragter den simple hypotese $H_0: \ \theta = \theta_0$ mod det simple alternativ $H_1: \ \theta = \theta_1$. Lad der være givet en mængde $\mathrm{C} \subseteq \mathrm{R}^n$, for hvilken nedenstående 3 betingelser er opfyldte.

$$
\begin{array}{lrcll}
i) & \dfrac{L(\theta_0)}{L(\theta_1)} & = & \dfrac{f(x_1, \cdots, x_n; \theta_0)}{f(x_1, \cdots, x_n; \theta_1)} & \leq \ k \ \forall (x_1, \cdots, x_n) \in \mathrm{C} \\
ii) & \dfrac{L(\theta_0)}{L(\theta_1)} & = & \dfrac{f(x_1, \cdots, x_n; \theta_0)}{f(x_1, \cdots, x_n; \theta_1)} & \geq \ k \ \forall (x_1, \cdots, x_n) \in \complement \mathrm{C} \\
iii) & \alpha & = & \mathrm{P}\{(X_1, \cdots, X_n) \in \mathrm{C} | \mathrm{H}_0\} &
\end{array}
$$

Da er C et bedste kritisk område af størrelse α for test af H_0 mod H_1.

Bevis. Forbigås. Se f.eks. [20][p. 201] ■

BEMÆRKNING 3.4. Sætningen siger ganske enkelt, at det bedste test har formen: Hvis $L(\theta_0)/L(\theta_1) \leq k$, så forkast H_0, ellers acceptéres H_0. Testet kaldes det bedste, fordi det har størst styrke, hvis $\theta = \theta_1$. ▼

Vi giver nu et eksempel på anvendelse af Neyman-Pearson's lemma.

EKSEMPEL 3.1. Vi har uafhængige stokastiske variable $X_1, \cdots, X_{20}$ med $X_i \in \mathrm{P}(\lambda)$. Vi ønsker at finde et uniformt stærkeste test med signifikansniveau $\alpha = 5\%$ af hypotesen

$$H_0: \quad \lambda = \frac{1}{10}$$

mod alternativet

$$H_1: \quad \lambda > \frac{1}{10}.$$

Vi betragter først et $\lambda' > \frac{1}{10}$ og anvender Neyman-Pearson's lemma til at konstruere et stærkeste test af $H_0: \ \lambda = \frac{1}{10}$ mod $H_1': \ \lambda = \lambda'$. Vi har, idet med $n = 20$,

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \mathrm{e}^{-\lambda} \\ &= \frac{1}{\prod x_i!} \lambda^{\Sigma x_i} \mathrm{e}^{-n\lambda}. \end{aligned}$$

Følgelig er

$$\begin{aligned} \frac{\mathrm{L}(\frac{1}{10})}{\mathrm{L}(\lambda')} \leq k \quad &\Leftrightarrow \quad \frac{\mathrm{e}^{-n\frac{1}{10}} \left(\frac{1}{10}\right)^{\Sigma x_i}}{\mathrm{e}^{-n\lambda'} (\lambda')^{\Sigma x_i}} \leq k \\ &\Leftrightarrow \quad \mathrm{e}^{20\lambda' - 2} (10\lambda')^{-\Sigma x_i} \leq k \\ &\Leftrightarrow \quad (-\Sigma x_i) \log_e (10\lambda') \leq \log_e k - 20\lambda' + 2 \\ &\Leftrightarrow \quad \Sigma x_i \geq -\frac{\log_e k - 20\lambda' + 2}{\log_e (10\lambda')} = c \end{aligned}$$

Ifølge Neyman-Pearson's lemma er mængden

$$\mathrm{C} = \{(x_1, \cdots, x_n) | \Sigma x_i \geq c\}$$

et bedste kritisk område for et test af $H_0: \ \lambda = \frac{1}{10}$ mod $H_1': \ \lambda = \lambda'$.

Det vil sige, at C angiver de stikprøveresultater for hvilke summen er større end et vist tal c. Dette tal kaldes ofte den **kritiske værdi**.

Vi skal nu bestemme den kritiske værdi c, så niveauet bliver 5%. Vi har altså,

$$\begin{aligned} 5\% &= \mathrm{P}\{\text{forkaste } H_0 | H_0\} \\ &= \mathrm{P}\left\{\sum_{i=1}^{20} X_i \geq c | \lambda = \frac{1}{10}\right\} \\ &= \mathrm{P}\{P(2) \geq c\}. \end{aligned}$$

Af tabel fås, at $c = 5$, idet $\mathrm{P}\{P(2) \leq 4\} = 0.95$. Vi har altså fået det kritiske område

$$\mathrm{C} = \left\{ (x_1, \cdots, x_{20}) | \sum_{i=1}^{20} x_i \geq 5 \right\}.$$

Vi ser nu, at havde vi gennemført de samme betragtninger med et $\lambda'' > \frac{1}{10}$, var vi kommet frem til den samme mængde C. Altså er C et bedste kritisk område for et test af H_0 mod et vilkårligt alternativ $\lambda = \lambda_1 > \frac{1}{10}$, og følgelig kaldes C et **uniformt bedste kritiske område** for test af H_0 mod H_1. ⧫

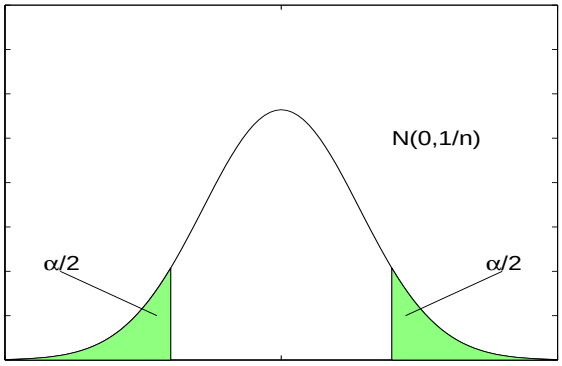
Som nævnt i det foregående afsnit vil der ikke altid eksistere et uniformt stærkeste test for en given hypotese og et givet alternativ. Vi må derfor opstille andre regler for konstruktion af et test.

En sædvanlig fremgangsmåde er at finde en stikprøvefunktion $T = \mathrm{t}(X_1, \cdots, X_n)$, hvis fordeling afhænger af den ukendte parameter og er kendt under H_0, da kan vi anvende T som teststørrelse og forkaste hypotesen, hvis vi observerer $T = t$ (for langt) ude i fordelingens haler.

EKSEMPEL 3.2. Lad $X_1, \cdots, X_n$ være $\mathrm{N}(\mu, 1)$-fordelte, og lad os teste

$$H_0: \quad \mu = 0 \quad \text{mod} \quad H_1: \quad \mu \neq 0.$$

Under H_0 vil $\bar{X}$ være $\mathrm{N}(0, \frac{1}{n})$-fordelt, d.v.s.



Vi forkaster for store og for små værdier af $\bar{X}$. ⧫

Reglen er imidlertid ikke formuleret generelt nok. Hvis vi havde hypotesen $H_0: \mu \geq 0$

mod alternativet $H_1\colon\ \mu < 0$, ville $\bar{X}$'s fordeling ikke være kendt under H_0. Alligevel er det klart, at det kritiske område for et rimeligt test må være

$$\{\bar{x} < c\}.$$

Det er imidlertid vanskeligt at give en fremgangsmåde, der dækker alle i praksis forekommende tilfælde.

Det gør imidlertid **kvotienttestet**. Fra afsnittet om maximum likelihood metoden erindrer vi, at $\mathrm{L}(\theta)$, hvor L er likelihoodfunktionen, er et udtryk for "rimeligheden" af parameteren θ, idet

$$\mathrm{L}(\theta) = \mathrm{f}(x_1, \cdots, x_n; \theta),$$

d.v.s. lig frekvensfunktionen taget i de observerede værdier, givet parameteren er θ. Vi fandt da et estimat for θ ved at vælge den værdi $\hat{\theta}$ der gjorde $\mathrm{L}(\theta)$ størst mulig. Den samme grundtanke ligger bag kvotienttestet. Man afgør, om man vil acceptere $\theta \in \Omega_0$ ved at vurdere forholdet mellem $\mathrm{L}(\theta)$, når $\theta \in \Omega_0$ (H_0 sand), og når $\theta \in \Omega \setminus \Omega_0$ (H_0 falsk).
Vi formulerer dette i

DEFINITION 3.9. Lad os antage, at vi ønsker at teste $H_0\colon\ \theta \in \Omega_0$ mod $H_1\colon\ \theta \in \Omega \setminus \Omega_0$, hvor $\Omega_0 \subseteq \Omega$, på basis af observationerne $X_1, \cdots, X_n$ med simultan frekvensfunktion $\mathrm{f}(x_1, \cdots, x_n; \theta)$. Vi definerer kvotienten q ved

$$\mathrm{q}(x_1, \cdots, x_n) = \frac{\sup\limits_{\theta \in \Omega_0} \mathrm{L}(\theta)}{\sup\limits_{\theta \in \Omega} \mathrm{L}(\theta)} = \frac{\sup\limits_{\theta \in \Omega_0} \mathrm{f}(x_1, \cdots, x_n; \theta)}{\sup\limits_{\theta \in \Omega} \mathrm{f}(x_1, \cdots, x_n; \theta)}$$

Kvotienttestprincippet (likelihood ratio test principle) går da ud på, at vi som kritisk område for testet H_0 mod H_1 anvender

$$\mathrm{C} = \{(x_1, \cdots, x_n) | \mathrm{q}(x_1, \cdots, x_n) < q_0\},$$

hvor $q_0 \leq 1$ fastlægges ved signifikansniveauet α

$$\alpha = \sup_{\theta \in \Omega_0} \mathrm{P}\{\mathrm{q}(X_1, \cdots, X_n) < q_0 | \theta\}.$$

Stikprøvefunktionen $Q = \mathrm{q}(X_1, \cdots, X_n)$ kaldes **kvotienttesstørrelsen**. ▲

Vi vil nu anskueliggøre denne meget vigtige metode ved at give 3 eksempler af stigende kompleksitet.

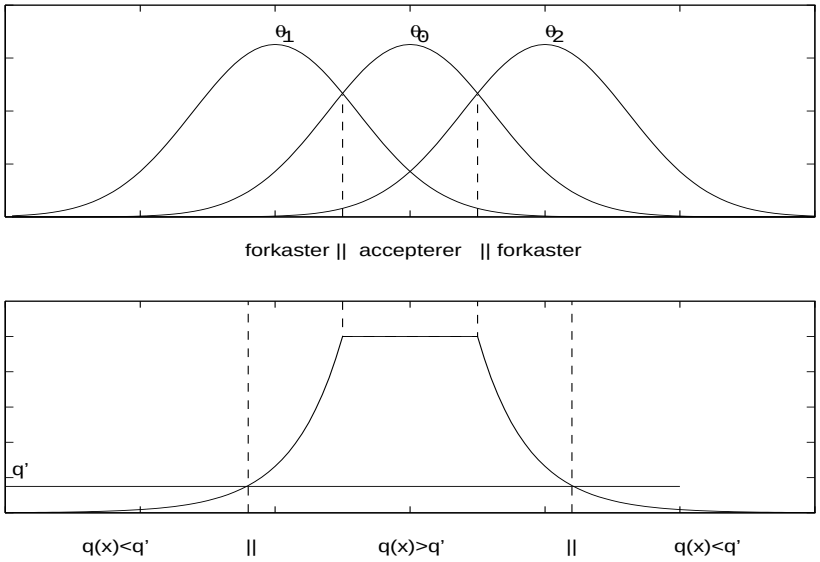
EKSEMPEL 3.3. Vi betragter en stokastisk variabel $X \in \mathrm{N}(\theta, 1), \theta \in \Omega = \{\theta_0, \theta_1, \theta_2\}$, og vi ønsker at teste

$$H_0: \quad \theta \in \Omega_0 = \{\theta_0\}$$

mod

$$H_1: \quad \theta \in \Omega \setminus \Omega_0 = \{\theta_1, \theta_2\}.$$

Ud fra en observation af X, vil vi afgøre, hvilket θ, vi vil antage. Parameterrummet Ω består altså kun af 3 punkter. På figuren nedenfor har vi skitseret, hvorledes det kritiske område for et kvotienttest af ovenstående hypoteser dannes ud fra et givet $q_0 = q'$. Den midterste fordeling er tætheden, hvis $\theta = \theta_0$, medens de to øvrige svarer til θ_1 og θ_2.



Vi ser, at det kritiske område er af formen $\{x < a\} \cup \{x > b\}$, hvilket jo er ganske rimeligt. ♦

I det næste eksempel betragter vi en lidt mere kompliceret problemstilling.

EKSEMPEL 3.4. Lad $X_1, \cdots, X_n$ være uafhængige $\mathrm{N}(\mu, 1)$-fordelte stokastiske variable. Vi ønsker at teste $H_0: \ \mu = 0$ mod $H_1: \ \mu \neq 0$. H_0 består altså alene af punktet

$\mu = 0$ (kaldes en simpel hypotese), medens H_1 består af alle andre værdier (kaldes et sammensat alternativ).

Likelihood-funktionen er

$$\mathrm{L}(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x_i-\mu)^2} = \sqrt{2\pi}^{-n} \exp\bigl(-\frac{1}{2}\Sigma(x_i-\mu)^2\bigr).$$

Vi ved at maximum likelihood estimatet er givet ved

$$\hat{\mu} = \bar{x}, \quad \text{d.v.s.} \quad \sup_{\mu} \mathrm{L}(\mu) = \sqrt{2\pi}^{-n} \exp\bigl(-\frac{1}{2}\Sigma(x_i-\bar{x})^2\bigr)$$

og følgelig er

$$\begin{aligned} \mathrm{q}(x_1, \cdots, x_n) &= \frac{\mathrm{L}(0)}{\mathrm{L}(\hat{\mu})} = \frac{\sqrt{2\pi}^{-n} \exp(-\frac{1}{2}\Sigma x_i^2)}{\sqrt{2\pi}^{-n} \exp\bigl(-\frac{1}{2}\Sigma(x_i-\bar{x})^2\bigr)} \\ &= \exp(-\frac{1}{2}\Sigma x_i^2 + \frac{1}{2}\Sigma x_i^2 - \frac{1}{2}n\bar{x}^2) \\ &= \exp(-\frac{1}{2}n\bar{x}^2). \end{aligned}$$

Nu er

$$\begin{aligned} \mathrm{q}(x_1, \cdots, x_n) < q_0 \quad &\Leftrightarrow \quad \log_e \mathrm{q}(x_1, \cdots, x_n) < \log_e q_0 \\ &\Leftrightarrow \quad -\frac{n}{2}\bar{x}^2 < k_1 \\ &\Leftrightarrow \quad \{\bar{x} < -c\} \vee \{\bar{x} > c\}, \end{aligned}$$

hvor k_1 og c er konstanter, der principielt kan udtrykkes ved q_0. Dette er imidlertid ikke så interessant, idet c kan fastlægges direkte ved niveauet α:

$$\mathrm{P}\{\bar{X} < -c \vee \bar{X} > c | \mu = 0\} = \alpha,$$

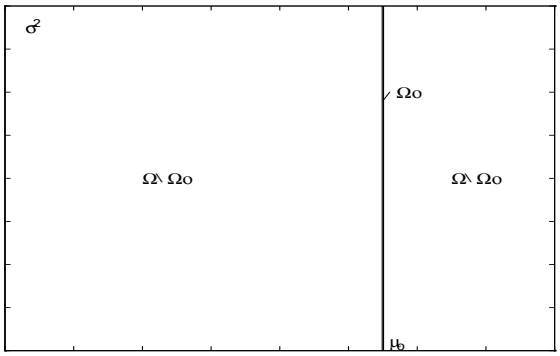
d.v.s. vi ender med samme test som i eksempel 3.2. ♦

Endelig betragter vi et eksempel, hvor såvel H_0 som H_1 er sammensatte. Eksemplet er langt og ret teknisk, så **udregningerne** kan eventuelt forbigås ved en første gennemlæsning.

EKSEMPEL 3.5. Lad $X_1, \cdots, X_n$ være indbyrdes uafhængige $\mathrm{N}(\mu, \sigma^2)$-fordelte, hvor σ^2 nu er ukendt. Lad hypotese og alternativ være

$$H_0: \quad \mu = \mu_0\,,\ \sigma^2 > 0 \quad \text{mod} \quad H_1: \quad \mu \neq \mu_0\,,\ \sigma^2 > 0.$$

Grafisk kan situationen afbildes således,



hvor den fede streg angiver parameterrummet under H_0 og den resterende del af parameterrummet (d.v.s. $\mu \neq \mu_0$, $\sigma^2 > 0$) svarer til H_1.

Vi har likelihood-funktionen

$$\mathrm{L}(\mu, \sigma^2) = \sqrt{2\pi\sigma^2}^{-n} \exp\big(-\frac{1}{2\sigma^2}\Sigma(x_i - \mu)^2\big),$$

og likelihood kvotienten bliver

$$\mathrm{q}(x_1, \cdots, x_n) = \frac{\sup\limits_{\sigma^2} \mathrm{L}(\mu_0\,, \sigma^2)}{\sup\limits_{\mu, \sigma^2} \mathrm{L}(\mu, \sigma^2)}$$

Vi erindrer fra tidligere, at

$$\mathrm{L}(\mu_0, \sigma^2) \text{ antager maximum i } \sigma^2 = \frac{1}{n}\Sigma(x_i - \mu_0)^2 = s'^2$$

og

$$\mathrm{L}(\mu, \sigma^2) \text{ antager maximum i } (\mu, \sigma^2) = \big(\bar{x}, \frac{1}{n}\Sigma(x_i - \bar{x})^2\big) = (\bar{x}, s_*^2)$$

De respektive maximumsværdier er

$$\begin{aligned}\sup_{\sigma^2} \mathrm{L}(\mu_0, \sigma^2) &= \sqrt{2\pi}^{-n} \exp(-\frac{n}{2}) \\ \sup_{\mu,\sigma^2} \mathrm{L}(\mu, \sigma^2) &= \sqrt{2\pi}^{-n} s_*^{-n} \exp(-\frac{n}{2}).\end{aligned}$$

Følgelig er $Q = \mathrm{q}(X_1, \cdots, X_n)$ lig

$$\frac{S_*^n}{S'^n} = \left(\frac{\Sigma(X_i - \bar{X})^2}{\Sigma(X_i - \mu_0)^2}\right)^{n/2}.$$

Nu er

$$\begin{aligned}\frac{\Sigma(X_i - \mu_0)^2}{\Sigma(X_i - \bar{X})^2} &= \frac{\Sigma(X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2}{\Sigma(X_i - \bar{X})^2} \\ &= 1 + \frac{1}{n-1} \frac{n(\bar{X} - \mu_0)^2}{\Sigma(X_i - \bar{X})^2/(n-1)} \\ &= 1 + \frac{1}{n-1} T^2,\end{aligned}$$

hvor betegnelsen T^2 åbenbart skyldes, at T er $\mathrm{t}(n-1)$-fordelt under H_0, d.v.s. såfremt $\mu = \mu_0$. Det kritiske område er bestemt ved

$$\mathrm{q}(x_1, \cdots, x_n) < \mu_0,$$

d.v.s. vi forkaster, hvis

$$\left(1 + \frac{1}{n-1} T^2\right)^{-n/2} \leq q_0.$$

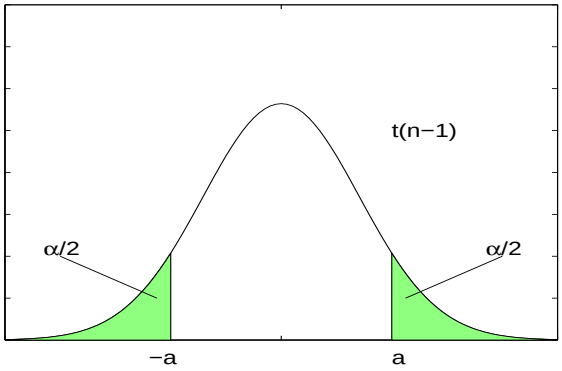
Denne ulighed løses med hensyn til T for $q_0 \geq 0$. Vi får derved et kritiske område af formen

$$\{T < -a\} \cup \{T > a\}.$$

Man fastlægger nu den kritiske værdi a ved hjælp af signifikansniveauet α, idet

$$\alpha = \sup_{H_0} \mathrm{P}\{\text{forkaste } H_0\} = \mathrm{P}\{T < -a \vee T > a | \mu = \mu_0\}.$$

Idet T er $\mathrm{t}(n-1)$-fordelt, når $\mu = \mu_0$,



sættes $a = \mathrm{t}(n-1)_{1-\alpha/2}$, d.v.s. lig $1-\alpha/2$ fraktilen i t-fordelingen med $n-1$ frihedsgrader.

Vi har altså fået følgende test:

Beregn teststørrelsen

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sqrt{\Sigma(X_i - \bar{X})^2/(n-1)}}$$

og forkast, hvis

$$T \leq \mathrm{t}(n-1)_{\alpha/2} = -\mathrm{t}(n-1)_{1-\alpha/2} \quad \text{eller} \quad T \geq \mathrm{t}(n-1)_{1-\alpha/2}.$$

♦

Som det fremgår af definitionen (og de to foregående eksempler), finder man kvotienttesstørrelsen ved at indsætte maximum likelihood estimatorer i stedet for parametrene i likelihood-funktionen. Det kan derfor ikke forbavse, at der eksisterer tilsvarende pæne sætninger om kvotientstørrelsens asymptotiske egenskaber, som der gør for maximum likelihood estimatorer.

Vi betragter parameterrummet

$$\Omega \subseteq \mathrm{R}^k$$

og

$$\Omega_{r0} = \{\theta \in \Omega | \theta_{r+1} = \theta_{(r+1)0}, \ldots, \theta_k = \theta_{k0}\},$$

d.v.s. Ω_{r0} er den delmængde af Ω, hvor de sidste $k - r$ koordinater er konstante.

Vi ønsker altså generelt at teste, om de sidste $k - r$ parametre kan have bestemte, forud fastsatte værdier, $\theta_{(r+1)0}, \cdots, \theta_{k0}$.

Lad nu $X_1, \cdots, X_n$ være stokastisk uafhængige og identisk fordelte med frekvensfunktionen $\mathrm{f}(x;\theta)$, $\theta \in \Omega$. Vi har da

SÆTNING 3.3. Lad $\mathrm{f}(x;\theta)$ være regulær med hensyn til alle θ_i, $i = 1, \cdots, k$. Kvotienttesstørrelsen for testet $H_0 : \ \theta \in \Omega_{r0}$ mod $H_1 : \ \theta \in \Omega \backslash \Omega_{r0}$ sættes lig $\mathrm{q}(X_1, \cdots, X_n)$. Da gælder, at

$$-2 \log_e \mathrm{q}(X_1, \cdots, X_n) \text{ asymtotisk } \in \chi^2(k - r),$$

hvis H_0 er sand, d.v.s. hvis $\theta \in \Omega_{r0}$.

Bevis. Forbigås. Beviset bygger på, at man kan vise, at det dominerende led i Taylor udviklingen er (-2) gange logaritmen til likelihood kvotienten er kvadratisk. Ved anvendelse af sætning 2.9, side 237, fås da, at størrelsen asymptotisk er fordelt som kvadratet på en normalt fordelt størrelse, d.v.s. som en χ^2-fordelt størrelse. For nøjere detaljer henvises til [55, p. 419]. ■

BEMÆRKNING 3.5. Sætningens store betydning er åbenbart, at man (ved hjælp af χ^2 fordelingen) kan finde et kritisk område, der asymptotisk har den rigtige størrelse, uden at kende Q's eksakte fordeling, som tit kan være svær at bestemme. Endvidere er grænsefordelingen uafhængig af begyndelsesfordelingen, hvilket naturligvis også er en fordel. ▼

BEMÆRKNING 3.6. Det må endvidere indskydes, at der gælder en lignende sætning i visse ikke-regulære tilfælde (e.g. for den rektangulære fordeling over intervallet $[0, \theta]$). Det er imidlertid bemærkelsesværdigt, at antallet af frihedsgrader bliver $2(k - r)$, og at fordelingen er eksakt og ikke asymptotisk. For nærmere detaljer henvises til [34, p. 237]. ▼

Vi skal nu give et eksempel på anvendelse af sætningen.

EKSEMPEL 3.6. Ved kast med en mønt har man observeret følgende antal kast, før man opnåede krone:

$$2, 2, 3, 0, 0, 1, 0, 0, 1, 0.$$

Kan det antages, at mønten er "fair", d.v.s. giver samme sandsynlighed for plat og krone.

Såfremt der er tale om "uafhængige" kast med en mønt, der har sandsynligheden p for at vise krone, kan ovenstående resultater opfattes som realisationer af stokastiske variable $X_1, \cdots, X_{10}$, hvor $X_i \in \mathrm{NB}(1, p)$ (jfr. afsnit 1.2.3).

Vi vil undersøge antagelsen ved at teste

$$H_0: \quad p = \frac{1}{2} \qquad \text{mod} \qquad H_1: \quad p \neq \frac{1}{2}.$$

Vi anvender kvotienttestet. Likelihood-funktionen er med $n = 10$

$$\begin{aligned} \mathrm{L}(p) &= \prod_{i=1}^{n} p(1-p)^{x_i} \\ &= p^n (1-p)^{\sum x_i} \end{aligned}$$

Ved at løse $L'(p) = 0$ m.h.t. p findes umiddelbart

$$\hat{p} = \frac{n}{n + \Sigma x_i},$$

d.v.s.

$$\sup \mathrm{L}(p) = \left(\frac{n}{n+\Sigma x_i}\right)^n \left(1 - \frac{n}{n+\Sigma x_i}\right)^{\Sigma x_i}.$$

Da

$$\sup_{H_0} \mathrm{L}(p) = \mathrm{L}(\frac{1}{2}) = 2^{-n} \cdot 2^{-\Sigma x_i},$$

og da

$$\sum x_i = 2 + 2 + 3 + 0 + 0 + 1 + 0 + 0 + 1 + 0 = 9$$

fås, at den observerede værdi af kvotienttesstørrelsen er

$$\mathrm{q}(x_1, \cdots, x_{10}) = \frac{\mathrm{L}(\frac{1}{2})}{\sup \mathrm{L}(p)} = \frac{2^{-10} \cdot 2^{-9}}{\left(\frac{10}{19}\right)^{10} \cdot \left(\frac{9}{19}\right)^{9}},$$

d.v.s.

$$\begin{aligned} -2\log_e q &= -2(-19\log_e 2 - 10\log_e 10 - 9\log_e 9 + 19\log_e 19) \\ &= 0.05274. \end{aligned}$$

Vi ser, at betingelserne for at anvende sætning 3.3 er til stede med $k = 1$ og $r = 0$, d.v.s. $-2\log_e Q$ er asymptotisk $\in \chi^2(1)$. Det kritiske område er derfor bestemt ved

$$\begin{aligned} & \mathrm{q}(x_1, \cdots, x_{10}) < q_0 \\ \Leftrightarrow \quad & -2\log_e \mathrm{q}(x_1, \cdots, x_{10}) > -2\log_e q_0, \end{aligned}$$

hvor $-2\log_e q_0$ fås som en fraktil i $\chi^2(1)$-fordelingen, i.e.

$$\mathrm{P}\{\chi^2(1) > -2\log_e q_0\} \simeq \alpha.$$

For $\alpha = 5\%$ fås $-2\log_e q_0 = \chi^2(1)_{0.95} = 3.84$, og da den observerede værdi af $-2\log_e q$ åbenbart er mindre end denne værdi, accepteres hypotesen altså på et 5% niveau. Af tabel ses, at hypotesen vil blive accepteret på alle niveauer, der er mindre end 80% (da $\chi^2(1)_{0.20} = 0.064$). Det må derfor være rimeligt at antage, at den pågældende mønt falder på begge sider med lige stor sandsynlighed. ♦

3.2 Specielle tests

Vi vil nu give en gennemgang af de almindeligt forekommende tests for parametre i nogle af de fordelinger, vi oftest beskæftiger os med. Vi indleder med

3.2.1 Tests i en binomialfordeling

Vi indleder med at betragte tests i binomialfordelingen. Lad $X_1, \cdots, X_k$ være uafhængige $\mathrm{B}(n_i, p)$-fordelte stokastiske variable. Vi betragter nu for $p_0 \in (0,1)$ hypoteserne

$$p = p_0, \quad p \geq p_0, \quad p \leq p_0$$

og alternativerne

$$p \neq p_0, \quad p < p_0, \quad p > p_0.$$

Nu vides, at summen er sufficient for parametrene i en binomialfordeling, således at vi som teststørrelse kan anvende

$$Z = \sum_{i=1}^{k} X_i.$$

Sætter vi

$$n = \sum_{i=1}^{k} n_i,$$

fås

$$Z \in \mathrm{B}(n, p).$$

Vi har da følgende skema over nogle hypoteser og alternativer angående p. I skemaet er endvidere anført de kritiske områder for tests på niveau α. Da binomialfordelingen er diskret, kan man ikke altid opnå, at niveauet eksakt er α. Endelig er styrken for de pågældende tests anført i et vilkårligt punkt p.

$Z \in \mathrm{B}(n,p)$			
Hypotese	Alternativ	Kritisk område	Styrke
$p=p_0$	$p\neq p_0$	$z<\mathrm{B}(n,p_0)_{\alpha/2}$ $\vee z>\mathrm{B}(n,p_0)_{1-\alpha/2}$	$\mathrm{P}\{\mathrm{B}(n,p)<\mathrm{B}(n,p_0)_{\alpha/2}\}$ $+\mathrm{P}\{\mathrm{B}(n,p)>\mathrm{B}(n,p_0)_{1-\alpha/2}\}$
$p\leq p_0$ $p=p_0$	$p>p_0$	$z>\mathrm{B}(n,p_0)_{1-\alpha}$	$\mathrm{P}\{\mathrm{B}(n,p)>\mathrm{B}(n,p_0)_{1-\alpha}\}$
$p\geq p_0$ $p=p_0$	$p<p_0$	$z<\mathrm{B}(n,p_0)_{\alpha}$	$\mathrm{P}\{\mathrm{B}(n,p)<\mathrm{B}(n,p_0)_{\alpha}\}$

Foruden ovenstående skema henleder vi opmærksomheden på, at konfidensintervallerne p. 265 også giver mulighed for at konstruere et test af e.g. hypotesen $H_0: p = p_0$ mod alternativet $H_1: p \neq p_0$ (ved hjælp af F-fordelingen).

EKSEMPEL 3.7. Ved en undersøgelse af virkningen af stress lærte en psykolog 18 studerende 2 forskellige måder at binde en knude på. Halvdelen af studenterne lærte metode A først og halvdelen metode B. Senere (henimod midnat) anmodedes de studerende efter en 4-timers eksamen om at binde knuden påny. Man registrerede, hvorvidt de brugte den første, de havde lært, eller den anden. Man ville på forhånd vente, at der var flest, der ville bruge den først lærte metode på grund af den psykisk stressede situation, man havde bragt studenterne i. Man fik følgende resultat:

Valg af metode	
Lært som nr. 1	Lært som nr. 2
16	2

Vi har her, at de 16 kan opfattes som et realiseret udfald af et $X \in \mathrm{B}(18,p)$, hvor p er sandsynligheden for at vælge den først tillærte metode. For at undersøge, om vor forestilling er korrekt, vil vi se, om den modsatte hypotese overhovedet kan antages at være rigtig, i.e. vi vil teste

$$H_0: \quad p \leq \frac{1}{2} \qquad \text{mod} \qquad H_1: \quad p > \frac{1}{2}.$$

Vælger vi et signifikansniveau $\alpha = 5\%$, fås $\mathrm{P}\{\mathrm{B}(18,\frac{1}{2}) \leq 12\} = 0.9519$, hvorfor det kritiske område bliver

$$\mathrm{C} = \{x | x > 12\}.$$

Da den observerede værdi af teststørrelsen ligger i det kritiske område, kan vi ikke på et 5% niveau antage hypotesen $p \leq \frac{1}{2}$. Vi må derfor gå ud fra, at stresspåvirkningen har en "regressions"-effekt, i.e. at forsøgspersonerne har en tendens til at vende tilbage til den først tillærte metode, når de handler under stress. ♦

3.2.2 Sammenligning af to binomialfordelinger

Lad der nu være givet 2 binomialt fordelte stikprøver

$$\begin{array}{ll} X_1, \cdots, X_{k_1} & X_i \in \mathrm{B}(n_i, p_1) \\ Y_1, \cdots, Y_{k_2} & Y_i \in \mathrm{B}(m_i, p_2). \end{array}$$

Vi betragter igen de sufficiente størrelser

$$\begin{aligned} X &= \sum_{i=1}^{k_1} X_i \in \mathrm{B}(n, p_1) \\ Y &= \sum_{i=1}^{k_2} Y_i \in \mathrm{B}(m, p_2) \end{aligned}$$

hvor vi har sat $n = \Sigma n_i$ og $m = \Sigma m_i$.

Vi er interesseret i at undersøge hypoteserne

$$p_1 = p_2, \quad p_1 \leq p_2, \quad p_1 \geq p_2$$

og alternativerne

$$p_1 \neq p_2, \quad p_1 > p_2, \quad p_1 < p_2.$$

Vi finder nu indledningsvis fordelingen af Y for givet $X + Y$, når $p_1 = p_2 = p$. Vi kan skrive

$$\begin{aligned} \mathrm{P}\{Y = y | X + Y = t\} &= \frac{\mathrm{P}\{Y = y \wedge X + Y = t\}}{\mathrm{P}\{X + Y = t\}} \\ &= \frac{\mathrm{P}\{Y = y \wedge X = t - y\}}{\mathrm{P}\{X + Y = t\}} \\ &= \frac{\mathrm{P}\{Y = y\}\mathrm{P}\{X = t - y\}}{\mathrm{P}\{X + Y = t\}}. \end{aligned}$$

Da $X + Y \in \mathrm{B}(n + m, p)$, får vi, at denne størrelse er

$$\begin{aligned} &= \frac{\binom{m}{y} p^y (1-p)^{m-y} nt - yp^{t-y}(1-p)^{n-t+y}}{\binom{m+n}{t} p^t (1-p)^{m+n-t}} \\ &= \frac{\binom{m}{y}\binom{n}{t-y}}{\binom{m+n}{t}} \end{aligned}$$

d.v.s at Y for givet $X+Y=t$ er hypergeometrisk fordelt $\mathrm{H}(t,m,m+n)$, hvis $p_1=p_2$.

Med ovenstående resultater er det nu muligt for os at angive teststørrelser og kritiske grænser for ovenstående hypoteser. Vi samler resultaterne i følgende skema.

$X\in\mathrm{B}(n,p_1)\wedge Y\in\mathrm{B}(m,p_2),\quad X+Y=t$		
Hypotese	Alternativ	Kritisk område
$p_1=p_2$	$p_1\neq p_2$	$y<\mathrm{H}(t,m,m+n)_{\alpha/2}\vee$ $y>\mathrm{H}(t,m,m+n)_{1-\alpha/2}$
$p_1\leq p_2$ $p_1=p_2$	$p_1>p_2$	$y<\mathrm{H}(t,m,m+n)_{\alpha}$
$p_1\geq p_2$ $p_1=p_2$	$p_1<p_2$	$y>\mathrm{H}(t,m,m+n)_{1-\alpha}$

Der gælder samme bemærkninger angående niveauet som anført til skemaet p 327. Styrken for ovenstående tests kan selvfølgelig findes, men fordelingen af Y for givet $X+Y$ er ret kompliceret, hvis $p_1\neq p_2$. Derfor forbigås styrken.

EKSEMPEL 3.8. Ved kvalitetskontrol af nogle emner fremstillet på hver sin produktionsgren fik man følgende resultater

	Defekte	I orden	I alt
I	2 stk.	18 stk.	20 stk.
II	2 stk.	13 stk.	15 stk.

Vi kan opfatte antallet af defekte fra de 2 produktionsgrene som realiserede udfald af binomialt fordelte stokastiske variable X og Y, hvor

$$\begin{aligned} X &\in \mathrm{B}(20,p_1)\\ Y &\in \mathrm{B}(15,p_2)\end{aligned}$$

Vi er interesserede i at undersøge, om de 2 produktionsgrene har samme defektprocent. Dette kan gøres ved at teste

$$H_0:\quad p_1=p_2 \qquad \text{mod} \qquad H_1: p_1\neq p_2\,.$$

Fordelingen af Y for givet $X+Y=4$ er $\mathrm{H}(4,15,35)$. Frekvensfunktionen for denne fordeling er

$$\mathrm{f}(y)=\frac{\binom{15}{y}\binom{20}{4-y}}{\binom{35}{4}}.$$

I nedenstående tabel anfører vi frekvensfunktionen $\mathrm{f}(y)$ samt fordelingsfunktionen $\mathrm{F}(y)$ for en $\mathrm{H}(4, 15, 35)$-fordeling. Til sammenligning har vi endvidere anført frekvensfunktionen g og fordelingsfunktionen G for en $\mathrm{B}(4, \frac{3}{7})$-fordeling, idet denne for store værdier af n og m skulle være en god approksimation til den hypergeometriske fordeling (sætning 1.8, p. 103).

y	$\mathrm{f}(y)$	$\mathrm{F}(y)$	$\mathrm{g}(y)$	$\mathrm{G}(y)$
0	0.093	0.093	0.107	0.107
1	0.327	0.419	0.320	0.427
2	0.381	0.800	0.360	0.787
3	0.174	0.974	0.180	0.967
4	0.026	1.000	0.034	1.000

Frekvens- og fordelingsfunktion for
$\mathrm{H}(4, 15, 35)$- og $\mathrm{B}(4, \frac{3}{7})$-fordelingerne

Ønsker vi at foretage et test på f.eks. 10% niveauet, ser, vi, at det nærmeste, vi kan komme, er det kritiske område

$$\{y < c_1 = 1\} \cup \{y > c_2 = 3\} \quad \sim \quad \{y = 0\} \cup \{y = 4\}.$$

Dette svarer til niveauet $0.093 + 0.026 = 0.119 = 11.9\%$. Beregnes niveauet ved hjælp af binomialapproksimationen fås $0.107 + 0.034 = 0.141 = 14.1\%$, hvilket afviger noget fra det eksakte niveau. ♦

EKSEMPEL 3.9. på anvendelse af approksimativ bestemmelse af kritisk område. I nedenstående tabel er der anført resultatet af 2 forskellige markedsanalyseinstitutters undersøgelser over folks stillingtagen til placeringen af en ny motorvej.

	God linieføring	Dårlig linieføring	I alt
I	222	1778	2000
II	225	1275	1500

Nu anvender de 2 institutter forskellige interviewformer, og man er derfor interesseret i at erfare, om det er den samme "ting", de "måler" hos folk.

Vi opfatter de 222 henholdsvis 225 som realiserede udfald af 2 uafhængige binomialt stokastiske variable

$$X \in \mathrm{B}(2000, p_1)$$
$$Y \in \mathrm{B}(1500, p_2).$$

Med denne model kan vore spørgsmål fra før omformuleres til, om det kan antages, at p_1 og p_2 er ens. Vi undersøger dette ved at teste

$$H_1: \quad p_1 = p_2 \qquad \text{mod} \qquad H_1: p_1 \neq p_2.$$

Ifølge skemaet p. 329 er det kritiske område

$$y < c_1 \vee y > c_2,$$

hvor c_1 og c_2 passende kan findes ved at kræve

$$\begin{aligned} &\mathrm{P}\{\mathrm{H}(447, 1500, 3500) \leq c_1\} \simeq \frac{\alpha}{2} \qquad \text{og} \\ &\mathrm{P}\{\mathrm{H}(447, 1500, 3500) \leq c_2\} \simeq 1 - \frac{\alpha}{2}. \end{aligned}$$

Vi har nu, at

$$\begin{aligned} &\mathrm{P}\big\{\mathrm{H}(447, 1500, 3500) \leq x\big\} \simeq \\ &\quad \mathrm{P}\big\{\mathrm{B}(447, \frac{3}{7}) \leq x\big\} \simeq \\ &\quad \mathrm{P}\left\{\mathrm{N}(0,1) \leq \frac{x + \frac{1}{2} - 447 \cdot \frac{3}{7}}{\sqrt{447 \cdot \frac{3}{7} \cdot \frac{4}{7}}}\right\} = \Phi\left(\frac{x - 191.1}{10.5}\right). \end{aligned}$$

Vælger vi signifikansniveauet $\alpha = 5\%$, bestemmes c_1 og c_2 derfor approksimativt af

$$\Phi\left(\frac{c_1 - 191.1}{10.5}\right) = 2.5\% \ \wedge \ \Phi\left(\frac{c_2 - 191.1}{10.5}\right) = 97.5\%$$

Ved hjælp af normalfordelingstabel fås

$$\frac{c_1 - 191.1}{10.5} = -1.96 \ \wedge \ \frac{c_2 - 191.1}{10.5} = 1.96$$

Ved løsning af disse ligninger får vi approksimationerne

$$c_1 = 170.5 \wedge c_2 = 211.7.$$

Det kritiske område sættes derfor til

$$\mathrm{C} = \{y | y < 171 \vee y > 211\},$$

og da $225 \in \mathrm{C}$, vil vi forkaste hypotesen på et 5% niveau. Vi må derfor antage, at tallene fra de 2 analyseinstitutter er udtryk for forskellige ting. ♦

3.2.3 Tests i en Poissonfordeling

Vi går dernæst over til at betragte Poisson fordelingen. Lad $X_1, \cdots, X_n$ være uafhængige $\mathrm{P}(\lambda)$-fordelte stokastiske variable. Vi kan igen nøjes med at betragte den sufficiente stikprøvefunktion $Z = \Sigma X_i \in \mathrm{P}(n\lambda)$. Vi har nu ganske analogt til p. 327 følgende skema over test af forskellige hypoteser om λ.

$Z \in \mathrm{P}(n\lambda)$			
Hypotese	Alternativ	Kritisk område	Styrke i λ
$\lambda = \lambda_0$	$\lambda \neq \lambda_0$	$z < \mathrm{P}(n\lambda_0)_{\alpha/2}$ $\vee z > \mathrm{P}(n\lambda_0)_{1-\alpha/2}$	$\mathrm{P}\{\mathrm{P}(n\lambda) < \mathrm{P}(n\lambda_0)_{\alpha/2}\}$ $+ \mathrm{P}\{\mathrm{P}(n\lambda) > \mathrm{P}(n\lambda_0)_{1-\alpha/2}\}$
$\lambda \leq \lambda_0$ $\lambda = \lambda_0$	$\lambda > \lambda_0$	$z > \mathrm{P}(n, \lambda_0)_{1-\alpha}$	$\mathrm{P}\{\mathrm{P}(n\lambda) > \mathrm{P}(n\lambda_0)_{1-\alpha}\}$
$\lambda \geq \lambda_0$ $\lambda = \lambda_0$	$\lambda < \lambda_0$	$z < \mathrm{P}(n\lambda_0)_{\alpha}$	$\mathrm{P}\{\mathrm{P}(n\lambda) < \mathrm{P}(n\lambda_0)_{\alpha}\}$

3.2.4 Sammenligning af to Poissonfordelinger

Hvis der er givet 2 Poisson fordelte stikprøver

$$X_1, \cdots, X_n, \quad X_i \in \mathrm{P}(\lambda_1)$$

$$Y_1, \cdots, Y_m, \quad Y_i \in \mathrm{P}(\lambda_2),$$

reducerer vi igen sufficient, d.v.s. vi betragter

$$X = \sum_{i=1}^{n} X_i \in \mathrm{P}(n\lambda_1)$$

og

$$Y = \sum_{i=1}^{n} Y_i \in \mathrm{P}(m\lambda_2)$$

Vi går nu frem som ved binomialfordelingen, d.v.s vi finder den betingede fordeling af Y givet $X + Y$. Vi har som før

$$\begin{aligned}
\mathrm{P}\{Y = y | X + Y = t\} &= \frac{\mathrm{P}\{Y = y \wedge X = t - y\}}{\mathrm{P}\{X + Y = t\}} \\
&= \frac{\mathrm{P}\{Y = y\}\mathrm{P}\{X = t - y\}}{\mathrm{P}\{X + Y = t\}} \\
&= \frac{\mathrm{e}^{-m\lambda_2}\frac{(m\lambda_2)^y}{y!}\mathrm{e}^{-n\lambda_1}\frac{(n\lambda_1)^{t-y}}{(t-y)!}}{\mathrm{e}^{-n\lambda_1 - m\lambda_2}\frac{(n\lambda_1 + m\lambda_2)^t}{t!}} \\
&= \binom{t}{y}\left(\frac{m\lambda_2}{n\lambda_1 + m\lambda_2}\right)^y\left(\frac{n\lambda_1}{n\lambda_1 + m\lambda_2}\right)^{t-y}
\end{aligned}$$

d.v.s. den betingede fordeling af Y givet $X + Y = t$ er en binomialfordeling $\mathrm{B}\big(t, m\lambda_2/(n\lambda_1 + m\lambda_2)\big)$. Nu er et udsagn som $\lambda_1 = \lambda_2$ ensbetydende med, at parameteren i den udledte binomialfordeling er

$$\rho = \frac{m\lambda_2}{n\lambda_1 + m\lambda_2} = \frac{m}{n + m} = \rho_0$$

Vi ser altså, at en sammenligning af de 2 Poisson fordelinger kan foregå som et test i en binomial fordeling. Vi har nu følgende oversigt over tests på niveau α.

$X \in \mathrm{P}(n\lambda_1) \wedge Y \in \mathrm{P}(m\lambda_2), \quad X + Y = t$		
Hypotese	Alternativ	Kritisk område
$\lambda_1 = \lambda_2$	$\lambda_1 \neq \lambda_2$	$y < \mathrm{B}(t, \rho_0)_{\alpha/2} \vee$ $y > \mathrm{B}(t, \rho_0)_{1-\alpha/2}$
$\lambda_1 \leq \lambda_2$ $\lambda_1 = \lambda_2$	$\lambda_1 > \lambda_2$	$y < \mathrm{B}(t, \rho_0)_{\alpha}$
$\lambda_1 \geq \lambda_2$ $\lambda_1 = \lambda_2$	$\lambda_1 < \lambda_2$	$y > \mathrm{B}(t, \rho_0)_{1-\alpha}$

Udtrykket for den betingede styrke (dvs. styrken for givet $X + Y = t$) kan let findes v.h.a. binomialfordelingen. En metode er at indføre kvotienten

$$\gamma = \lambda_1/\lambda_2$$

hvoraf

$$\rho(\gamma) = \frac{m}{\gamma \cdot n + m}$$

Styrken for eksempelvis testet $\lambda_1 = \lambda_2$ mod $\lambda_1 \neq \lambda_2$ bliver da

$$\begin{aligned} p(\gamma) &= \mathrm{P}\{\mathrm{B}(t, \rho(\gamma)) < \mathrm{B}(t, \rho_0)_{\alpha/2}\} \\ &\quad + \mathrm{P}\{\mathrm{B}(t, \rho(\gamma)) > \mathrm{B}(t, \rho_0)_{1-\alpha/2}\} \end{aligned}$$

Beregning af den ubetingede styrke er noget mere kompliceret og medtages ikke her.

EKSEMPEL 3.10. I en virksomhed, hvor man har 2 drejebænke, har man en formodning om, at bænk I er hårdere belastet end bænk II. Inden for en uge konstaterede man 12 sammenbrud på bænk I og 7 på bænk II.

Vi antager, at antallet af sammenbrud følger en Poisson proces med intensitet λ_1 henholdsvis λ_2 sammenbrud pr. uge. Vi kan altså opfatte observationerne 12 og 7 som realiserede udfald af $X \in \mathrm{P}(\lambda_1)$ henholdsvis $Y \in \mathrm{P}(\lambda_2)$. For at undersøge vor formodning $\lambda_1 > \lambda_2$, vil vi se, om den modsatte hypotese overhovedet kan antages at være sand. Vi vil altså teste

$$H_0: \quad \lambda \leq \lambda_2 \qquad \text{mod} \qquad H_1: \quad \lambda_1 > \lambda_2 .$$

Det kritiske område er derfor $\{y < c\}$, hvor c bestemmes ved
$\mathrm{P}\big\{\mathrm{B}(12+7, \frac{1}{2}) \leq c\big\} \simeq \alpha$, da $m = n = 1$ medfører $\rho_0 = \frac{1}{2}$. Vi vælger $\alpha = 5\%$. Af tabel ser vi

$$\begin{aligned} \mathrm{P}\{\mathrm{B}(19, \frac{1}{2}) \leq 6\} &= 0.0835 \\ \mathrm{P}\{\mathrm{B}(19, \frac{1}{2}) \leq 5\} &= 0.0318. \end{aligned}$$

Det kritiske område er derfor $\{y < 6\}$, og vi ser, at hypotesen accepteres. Vi kan altså ikke på et 5% niveau afvise muligheden $\lambda_1 \leq \lambda_2$, d.v.s. vi kan **ikke** tillade os at konkludere, at $\lambda_1 > \lambda_2$. Det er klart, at hvis vi undersøger hypotesen

$$K_1: \quad \lambda_1 \geq \lambda_2 \qquad \text{mod} \qquad K_2: \quad \lambda_1 < \lambda_2$$

på et 5% niveau, får vi igen accept. Vi må derfor konkludere, at det ikke er muligt at konstantere forskelle i sammenbrudsintensiteten på de to drejebænke, som er signifikante på et 5% niveau. ♦

3.2.5 Tests i normalfordelingen

Vi skal i dette afsnit give tests vedrørende middelværdi og varians for normal fordelte variable. Vi har i de tidligere afsnit set, hvorledes man kan konstruere sådanne tests, så de anføres uden særlige kommentarer.

I tabellerne på side 336, 337, 338 og 339 betragtes uafhængige stokastiske variable $X_1,\cdots,X_n$ med $X_i \in \mathrm{N}(\mu,\sigma^2)$. Her er anvendt betegnelserne

$$\begin{aligned} \bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ S'^2 &= \frac{1}{n}\sum_{I=1}^{n}(X_i-\mu)^2 \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \end{aligned}$$

I tabellerne på side 340, 341, 342, 343, 344 og 345 betragtes uafhængige stokastiske variable $X_1,\cdots,X_n$ og $Y_1,\cdots,Y_m$ med $X_i \in \mathrm{N}(\mu_1,\sigma_1^2)$ og $Y_i \in \mathrm{N}(\mu_2,\sigma_2^2)$. Her er anvendt betegnelserne

$$\begin{aligned} \bar{X} &= \tfrac{1}{n}\sum_{i=1}^{n} X_i & \bar{Y} &= \tfrac{1}{m}\sum_{i+1}^{m} Y_i \\ S'^2_1 &= \tfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu_1)^2 & S'^2_2 &= \tfrac{1}{m}\sum_{i=1}^{m}(Y_i-\mu_2)^2 \\ S_1^2 &= \tfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 & S_2^2 &= \tfrac{1}{m-1}\sum_{i=1}^{m}(Y_i-\bar{Y})^2 \\ S^2 &= \frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2} \\ &= \tfrac{1}{m+n-2}\Big(\sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{m}(Y_i-\bar{Y})^2\Big) \\ f_s &= m+n-2 \\ \delta &= \frac{\mu_1-\mu_2}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} & \rho &= \frac{\mu_1-\mu_2}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \\ \frac{1}{r} &= \frac{c^2}{n-1}+\frac{(1-c)^2}{m-1} & c &= \frac{\frac{1}{n}S_1^2}{\frac{1}{n}S_1^2+\frac{1}{m}S_2^2} \end{aligned}$$

I tabellerne er der angivet udtryk for bestemmelse af stikprøvestørrelse. Metoden er illustreret i eksemplerne 3.11, 3.12, 3.13 og 3.14, der følger umiddelbart efter tabellerne.

$X_i \in N(\mu, \sigma^2) \quad i = 1, \ldots, n$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \mu = \mu_0$	$z < u_{\alpha/2}$	$\Phi\left(u_{\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$	$\mu = \mu_0 + \Delta\sigma$	$\Phi(u_{1-\alpha/2} - \delta\sqrt{n})$
$H_1 : \mu \neq \mu_0$	$\vee\, z > u_{1-\alpha/2}$	$+1 - \Phi\left(u_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$		$-\Phi(u_{\alpha/2} - \Delta\sqrt{n}) = \beta$
$H_0 : \mu = \mu_0$ el. $\mu \leq \mu_0$	$z > u_{1-\alpha}$	$1 - \Phi\left(u_{1-\alpha} - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$	$\mu = \mu_0 + \Delta\sigma$	$n = (u_{1-\alpha} - u_\beta)^2/\Delta^2$
$H_1 : \mu > \mu_0$				
$H_0 : \mu = \mu_0$ el. $\mu \geq \mu_0$	$z < u_\alpha$	$\Phi\left(u_\alpha - \frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$	$\mu = \mu_0 + \Delta\sigma$	$n = (u_{1-\alpha} - u_\beta)^2/\Delta^2$
$H_1 : \mu < \mu_0$				

Parameter	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
$\mu \quad (\sigma^2$ kendt)	$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$Z \in N\left(\frac{\mu - \mu_0}{\sigma/\sqrt{n}}\right)$

$X_i \in N(\mu, \sigma^2) \quad i = 1, \ldots, n$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$z < t(n-1)_{\alpha/2}$ $\vee\, z > t(n-1)_{1-\alpha/2}$	$P\left\{\left\lvert t\left(n-1, \sqrt{n}\sigma(\mu-\mu_0)\right)\right\rvert\right.$ $\left.> t(n-1)_{1-\alpha/2}\right\}$	$\mu = \mu_0 + \Delta\sigma$	Man prøver sig frem
$H_0 : \mu = \mu_0$ el. $\mu \leq \mu_0$ $H_1 : \mu > \mu_0$	$z > t(n-1)_{1-\alpha}$	$P\left\{t\left(n-1, \sqrt{n}\sigma(\mu-\mu_0)\right)\right.$ $\left.> t(n-1)_{1-\alpha}\right\}$	$\mu = \mu_0 + \Delta\sigma$	$t(n-1)_{1-\alpha} = t(n-1, \Delta\sqrt{n})_{\beta}$
$H_0 : \mu = \mu_0$ el. $\mu \geq \mu_0$ $H_1 : \mu < \mu_0$	$z < t(n-1)_{\alpha}$	$P\left\{t\left(n-1, \sqrt{n}\sigma(\mu-\mu_0)\right)\right.$ $\left.< t(n-1)_{\alpha}\right\}$	$\mu = \mu_0 + \Delta\sigma$	$t(n-1)_{\alpha} = t(n-1, \Delta\sqrt{n})_{1-\beta}$

Parameter	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
$\mu \quad (\sigma^2$ ukendt)	$Z = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$Z \in t\left(f_s, \dfrac{\sqrt{n}}{\sigma}(\mu - \mu_0)\right)$

$X_i \in N(\mu, \sigma^2) \quad i = 1, \ldots, n$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$z < \chi^2(n)_{\alpha/2}$ $\vee\, z > \chi^2(n)_{1-\alpha/2}$	$P\left\{\chi^2(n) < \frac{\sigma_0^2}{\sigma^2}\chi^2(n)_{\alpha/2}\right\} +$ $P\left\{\chi^2(n) > \frac{\sigma_0^2}{\sigma^2}\chi^2(n)_{1-\alpha/2}\right\}$	$\sigma^2 = \gamma\sigma_0^2$	Man prøver sig frem
$H_0 : \sigma^2 = \sigma_0^2$ el. $\sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$z > \chi^2(n)_{1-\alpha}$	$P\left\{\chi^2(n) > \frac{\sigma_0^2}{\sigma^2}\chi^2(n)_{1-\alpha}\right\}$	$\sigma^2 = \gamma\sigma_0^2$	$\frac{\chi^2(n)_{1-\alpha}}{\chi^2(n)_\beta} = \gamma$
$H_0 : \sigma^2 = \sigma_0^2$ el. $\sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$z < \chi^2(n)_\alpha$	$P\left\{\chi^2(n) < \frac{\sigma_0^2}{\sigma^2}\chi^2(n)_\alpha\right\}$	$\sigma^2 = \gamma\sigma_0^2$	$\frac{\chi^2(n)_\alpha}{\chi^2(n)_{1-\beta}} = \gamma$

Parameter		Stikprøvefunktion (Teststørrelse)	Fordeling af Z
σ^2	(μ kendt)	$Z = \frac{n}{\sigma_0^2}S'^2$	$Z \in \frac{\sigma^2}{\sigma_0^2}\chi^2(n)$

$X_i \in N(\mu, \sigma^2) \quad i = 1, \ldots, n$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$z < \chi^2(n-1)_{\alpha/2}$ $\vee\, z > \chi^2(n-1)_{1-\alpha/2}$	$P\left\{\chi^2(n-1) < \frac{\sigma_0^2}{\sigma^2}\chi^2(n-1)_{\alpha/2}\right\} +$ $P\left\{\chi^2(n-1) > \frac{\sigma_0^2}{\sigma^2}\chi^2(n-1)_{1-\alpha/2}\right\}$	$\sigma^2 = \gamma\sigma_0^2$	Man prøver sig frem
$H_0 : \sigma^2 = \sigma_0^2$ el. $\sigma^2 \leq \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$z > \chi^2(n-1)_{1-\alpha}$	$P\left\{\chi^2(n-1) > \frac{\sigma_0^2}{\sigma^2}\chi^2(n-1)_{1-\alpha}\right\}$	$\sigma^2 = \gamma\sigma_0^2$	$\frac{\chi^2(n-1)_{1-\alpha}}{\chi^2(n-1)_\beta} = \gamma$
$H_0 : \sigma^2 = \sigma_0^2$ el. $\sigma^2 \geq \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$z < \chi^2(n-1)_\alpha$	$P\left\{\chi^2(n-1) < \frac{\sigma_0^2}{\sigma^2}\chi^2(n-1)_\alpha\right\}$	$\sigma^2 = \gamma\sigma_0^2$	$\frac{\chi^2(n-1)_\alpha}{\chi^2(n-1)_{1-\beta}} = \gamma$

Parameter	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
σ^2 (μ ukendt)	$Z = \frac{n-1}{\sigma_0^2}S'^2$	$Z \in \frac{\sigma^2}{\sigma_0^2}\chi^2(n-1)$

$X_i \in N(\mu_1, \sigma_1^2) \quad i = 1, \ldots, n \quad \wedge \quad Y_i \in N(\mu_2, \sigma_2^2) \quad i = 1, \ldots, m$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \mu_1 = \mu_2$	$z < u_{\alpha/2}$	$\Phi\left(u_{\alpha/2} - \delta\right)$	$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \Delta$ for $m = n$	Man prøver sig frem
$H_1 : \mu_1 \neq \mu_2$	$\vee\, z > u_{1-\alpha/2}$	$+1 - \Phi\left(u_{1-\alpha/2} - \delta\right)$	$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_1\sigma_2}} = \Delta$ for $m = \frac{n\sigma_2}{\sigma_1}$	
$H_0 : \mu_1 = \mu_2$ el. $\mu_1 \leq \mu_2$	$z > u_{1-\alpha}$	$1 - \Phi\left(u_{1-\alpha} - \delta\right)$	$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \Delta$ for $m = n$	$n = (u_{1-\alpha} - u_\beta)^2/\Delta^2$
$H_1 : \mu_1 > \mu_2$			$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_1\sigma_2}} = \Delta$ for $m = \frac{n\sigma_2}{\sigma_1}$	
$H_0 : \mu_1 = \mu_2$ el. $\mu_1 \geq \mu_2$	$z < u_\alpha$	$\Phi\left(u_\alpha - \delta\right)$	$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \Delta$ for $m = n$	$n = (u_{1-\alpha} - u_\beta)^2/\Delta^2$
$H_1 : \mu_1 < \mu_2$			$\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_1\sigma_2}} = \Delta$ for $m = \frac{n\sigma_2}{\sigma_1}$	

Parametre	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
$\mu_1, \mu_2 \quad (\sigma_1^2, \sigma_2^2$ kendte)	$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$	$Z \in N\left(\delta, 1\right)$

$X_i \in N(\mu_1, \sigma_1^2) \quad i = 1, \ldots, n \quad \wedge \quad Y_i \in N(\mu_2, \sigma_2^2) \quad i = 1, \ldots, m$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$z < t(n+m-2)_{\alpha/2}$ $\vee\, z > t(n+m-2)_{1-\alpha/2}$	$P\left\{\left\|t\left(n+m-2,\rho\right)\right\|\right.$ $\left.> t(n+m-2)_{1-\alpha/2}\right\}$	$\frac{\mu_1-\mu_2}{\sigma} = \Delta$ for $m = n$	Man prøver sig frem
$H_0 : \mu_1 = \mu_2$ el. $\mu_1 \leq \mu_2$ $H_1 : \mu_1 > \mu_2$	$z > t(n+m-2)_{1-\alpha}$	$P\left\{t\left(n+m-2,\rho\right)\right.$ $\left.> t(n+m-2)_{1-\alpha}\right\}$	$\frac{\mu_1-\mu_2}{\sigma} = \Delta$ for $m = n$	$t(2n-2)_{1-\alpha}$ $= t(2n-2, \Delta\sqrt{2n})_{\beta}$
$H_0 : \mu_1 = \mu_2$ el. $\mu_1 \geq \mu_2$ $H_1 : \mu_1 < \mu_2$	$z < t(n+m-2)_{\alpha}$	$P\left\{t\left(n+m-2,\rho\right)\right.$ $\left.< t(n+m-2)_{\alpha}\right\}$	$\frac{\mu_1-\mu_2}{\sigma} = \Delta$ for $m = n$	$t(2n-2)_{1-\alpha}$ $= t(2n-2, \Delta\sqrt{2n})_{\beta}$

Parametre		Stikprøvefunktion (Teststørrelse)	Fordeling af Z
μ_1, μ_2	($\sigma_1^2 = \sigma_2^2 = \sigma^2$ ukendt)	$Z = \dfrac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}$	$Z \in t\left(f_s, \rho\right)$

$X_i \in N(\mu_1, \sigma_1^2) \quad i = 1, \ldots, n \quad \wedge \quad Y_i \in N(\mu_2, \sigma_2^2) \quad i = 1, \ldots, m$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$z < t(r)_{\alpha/2}$ $\vee\, z > t(r)_{1-\alpha/2}$	Afhænger af middelværdierne og spredningerne		Approximative beregninger nødvendige
$H_0 : \mu_1 = \mu_2$ el. $\mu_1 \leq \mu_2$ $H_1 : \mu_1 > \mu_2$	$z > t(r)_{1-\alpha}$	Afhænger af middelværdierne og spredningerne		Approximative beregninger nødvendige
$H_0 : \mu_1 = \mu_2$ el. $\mu_1 \geq \mu_2$ $H_1 : \mu_1 < \mu_2$	$z < t(r)_{\alpha}$	Afhænger af middelværdierne og spredningerne		Approximative beregninger nødvendige

Parametre		Stikprøvefunktion (Teststørrelse)	Fordeling af Z
μ_1, μ_2	($\sigma_1^2 \neq \sigma_2^2$ ukendte)	$Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$	$Z \sim t(r)$ hvis $\mu_1 = \mu_2$

$X_i \in N(\mu_1, \sigma_1^2) \quad i = 1, \ldots, n \quad \wedge \quad Y_i \in N(\mu_2, \sigma_2^2) \quad i = 1, \ldots, m$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$	$z < F(n,m)_{\alpha/2}$ $\vee\, z > F(n,m)_{1-\alpha/2}$	$P\left\{F(n,m) < \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n,m)_{\alpha/2}\right\}$ $+P\left\{F(n,m) > \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n,m)_{1-\alpha/2}\right\}$	$\gamma = \frac{\sigma_1^2}{\sigma_2^2}$	Man prøver sig frem
$H_0 : \sigma_1^2 = \sigma_2^2$ el. $\sigma_1^2 \leq \sigma_2^2$ $H_1 : \sigma_1^2 > \sigma_2^2$	$z > F(n,m)_{1-\alpha}$	$P\left\{F(n,m) > \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n,m)_{1-\alpha}\right\}$	$\gamma = \frac{\sigma_1^2}{\sigma_2^2}$	$\frac{F(n,m)_{1-\alpha}}{F(n,m)_{\beta}} = \gamma$
$H_0 : \sigma_1^2 = \sigma_2^2$ el. $\sigma_1^2 \geq \sigma_2^2$ $H_1 : \sigma_1^2 < \sigma_2^2$	$z < F(n,m)_{\alpha}$	$P\left\{F(n,m) < \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n,m)_{\alpha}\right\}$	$\gamma = \frac{\sigma_1^2}{\sigma_2^2}$	$\frac{F(n,m)_{\alpha}}{F(n,m)_{1-\beta}} = \gamma$

Parametre	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
σ_1^2, σ_2^2 (μ_1, μ_2 kendte)	$Z = \frac{S'^2_1}{S'^2_2}$	$Z \in \frac{\sigma_1^2}{\sigma_2^2} F(n,m)$

$X_i \in N(\mu_1, \sigma_1^2) \quad i = 1, \ldots, n \quad \wedge \quad Y_i \in N(\mu_2, \sigma_2^2) \quad i = 1, \ldots, m$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \sigma_1^2 = \sigma_2^2$	$z < F(n-1,m)_{\alpha/2}$	$P\Big\{F(n-1,m) < \big(\sigma_2^2/\sigma_1^2\big) F(n-1,m)_{\alpha/2}\Big\}$ +	$\gamma = \dfrac{\sigma_1^2}{\sigma_2^2}$	Man prøver sig frem
$H_1 : \sigma_1^2 \neq \sigma_2^2$	$\vee\, z > F(n-1,m)_{1-\alpha/2}$	$P\Big\{F(n-1,m) > \big(\sigma_2^2/\sigma_1^2\big) F(n-1,m)_{1-\alpha/2}\Big\}$		
$H_0 : \sigma_1^2 = \sigma_2^2$ el. $\sigma_1^2 \leq \sigma_2^2$ $H_1 : \sigma_1^2 > \sigma_2^2$	$z > F(n-1,m)_{1-\alpha}$	$P\Big\{F(n-1,m) > \big(\sigma_2^2/\sigma_1^2\big) F(n-1,m)_{1-\alpha}\Big\}$	$\gamma = \dfrac{\sigma_1^2}{\sigma_2^2}$	$\dfrac{F(n-1,m)_{1-\alpha}}{F(n-1,m)_{\beta}} = \gamma$
$H_0 : \sigma_1^2 = \sigma_2^2$ el. $\sigma_1^2 \geq \sigma_2^2$ $H_1 : \sigma_1^2 < \sigma_2^2$	$z < F(n-1,m)_{\alpha}$	$P\Big\{F(n-1,m) < \big(\sigma_2^2/\sigma_1^2\big) F(n-1,m)_{\alpha}\Big\}$	$\gamma = \dfrac{\sigma_1^2}{\sigma_2^2}$	$\dfrac{F(n-1,m)_{\alpha}}{F(n-1,m)_{1-\beta}} = \gamma$

Parametre		Stikprøvefunktion (Teststørrelse)	Fordeling af Z
σ_1^2, σ_2^2	(μ_1 ukendt, μ_2 kendt)	$Z = \dfrac{S_1^2}{S_2'^2}$	$Z \in \dfrac{\sigma_1^2}{\sigma_2^2} F(n-1,m)$

$X_i \in N(\mu_1, \sigma_1^2) \quad i = 1, \ldots, n \quad \wedge \quad Y_i \in N(\mu_2, \sigma_2^2) \quad i = 1, \ldots, m$				
Hypotese	Kritisk Område	Styrke	Alternativ	Bestemmelse af n
$H_0 : \sigma_1^2 = \sigma_2^2$	$z < F(n-1, m-1)_{\alpha/2}$	$P\Big\{F(n-1, m-1) < \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n-1, m-1)_{\alpha/2}\Big\}$ +	$\gamma = \frac{\sigma_1^2}{\sigma_2^2}$	Man prøver sig frem
$H_1 : \sigma_1^2 \neq \sigma_2^2$	$\vee\, z > F(n-1, m-1)_{1-\alpha/2}$	$P\Big\{F(n-1, m-1) > \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n-1, m-1)_{1-\alpha/2}\Big\}$		
$H_0 : \sigma_1^2 = \sigma_2^2$ el. $\sigma_1^2 \leq \sigma_2^2$	$z > F(n-1, m-1)_{1-\alpha}$	$P\Big\{F(n-1, m-1) > \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n-1, m-1)_{1-\alpha}\Big\}$	$\gamma = \frac{\sigma_1^2}{\sigma_2^2}$	$\frac{F(n-1, m-1)_{1-\alpha}}{F(n-1, m-1)_{\beta}} = \gamma$
$H_1 : \sigma_1^2 > \sigma_2^2$				
$H_0 : \sigma_1^2 = \sigma_2^2$ el. $\sigma_1^2 \geq \sigma_2^2$	$z < F(n-1, m-1)_{\alpha}$	$P\Big\{F(n-1, m-1) < \left(\frac{\sigma_2^2}{\sigma_1^2}\right) F(n-1, m-1)_{\alpha}\Big\}$	$\gamma = \frac{\sigma_1^2}{\sigma_2^2}$	$\frac{F(n-1, m-1)_{\alpha}}{F(n-1, m-1)_{1-\beta}} = \gamma$
$H_1 : \sigma_1^2 < \sigma_2^2$				

Parametre	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
σ_1^2, σ_2^2 (μ_1, μ_2 ukendte)	$Z = \frac{S_1^2}{S_2^2}$	$Z \in \frac{\sigma_1^2}{\sigma_2^2} F(n-1, m-1)$

EKSEMPEL 3.11. Vi betragter en virksomhed, som producerer emner med et givet karakteristikum (kvalitetsindeks), der er af afgørende betydning for emnets anvendelighed.

Lad $X_i \in N(\mu, \sigma^2),\ i = 1, \ldots, n$, betegne en stikprøve af emner fra produktionen, som antages at være indbyrdes uafhængige.

Vi ønsker at teste hypotesen

$$\mathrm{H}_0: \quad \mu = \mu_0$$

mod alternativet

$$\mathrm{H}_1: \quad \mu > \mu_0,$$

Såfremt H_0 er sand, ønskes en acceptsandsynlighed på mindst 1-α:

$$\mathrm{P}\{\text{ accept } | \mu = \mu_0\} \geq 1 - \alpha$$

Hvis μ er væsentligt større end μ_0, f.eks. $\mu_0 + \sigma$, ønskes en acceptsandsynlighed på højst β. Kravet bliver derfor

$$\mathrm{P}\{\text{ accept } | \mu > \mu_0 + \sigma\} \leq \beta$$

For $\alpha = \beta = 5\%$ fås af tabellen side 336

$$n = \frac{(u_{95\%} - u_{5\%})}{1^2} = \frac{(1.6449 - (-1.6449))^2}{1} = 3.2898$$

Altså skal stikprøven mindst indeholde 4 emner. ♦

EKSEMPEL 3.12. Vi betragter samme problemstilling som i Eksempel 3.11, men ønsker nu at undersøge produktionens variabilitet (fremfor niveauet μ) ved at teste hypotesen

$$\mathrm{H}_0: \quad \sigma^2 = \sigma_0^2$$

mod alternativet

$$\mathrm{H}_1: \quad \sigma^2 > \sigma_0^2$$

Som før ønskes en høj acceptsandsynlighed, hvis kvaliteten er tilfredsstillende:

$$\mathrm{P}\{\text{ accept }|\sigma^2 = \sigma_0^2\} \geq 1 - \alpha$$

Hvis σ^2 bliver væsentligt større end σ_0^2, f.eks. $\sigma^2 = 3\sigma_0^2$, ønskes naturligvis en lille acceptsandsynlighed β:

$$\mathrm{P}\{\text{ accept }|\sigma^2 = 3\sigma_0^2\} \leq \beta$$

Af tabellen side 339 fremgår det, at stikprøvestørrelsen n skal bestemmes af

$$\gamma = \frac{\chi^2(n-1)_{1-\alpha}}{\chi^2(n-1)_{\beta}} = 3,$$

som ikke kan løses analytisk med hensyn til n.

Vi vælger som eksempel $\alpha = 5\%$ og $\beta = 10\%$, som ofte benyttes i praksis. Ved at indsætte forskellige værdier af n fås følgende resultater

n	5	10	12	15	16
γ	8.92	4.07	3.52	3.04	2.92

Altså skal der udtages en stikprøve på n=16 emner for, at vi kan være mindst 95% sikre på, at emnerne opfylder kvalitetskravet ($\sim$ H_0 er sand). I dette tilfælde er der dog en sandsynlighed på (højst) 10% for at acceptere en stikprøve, som ikke er af en tilfredsstillende kvalitet (hvis $\sigma^2 \geq 3\sigma_0^2$ $\sim$ H_1 er sand). ♦

EKSEMPEL 3.13. Vi betragter en virksomhed, som producerer emner på to ensartede produktionslinjer. Lad $X_i \in N(\mu_1, \sigma_1^2),\ i = 1, \ldots, n,$ og $Y_j \in N(\mu_2, \sigma_2^2),\ j = 1, \ldots, m,$ betegne et kvalitetsindeks for emner fra de to produktionslinjer. På basis af tidligere stikprøver er det godtgjort, at $\sigma_1^2 = 2^2$ og $\sigma_2^2 = 1.5^2$.

Vi ønsker at teste hypotesen

$$\mathrm{H}_0: \quad \mu_1 = \mu_2$$

mod alternativet

$$\mathrm{H}_1: \quad \mu_1 > \mu_2$$

Det ønskes undersøgt hvor stor en stikprøve (forudsat vi vælger $n = m$), der skal udtages fra hver af de to produktionslinjer for at opfylde følgende krav:

$$\mathrm{P}\{\text{ accept }|\mu_1 = \mu_2\} \geq 1 - \alpha$$

i.e. en høj acceptsandsynlighed, såfremt kvaliteten af emner fra de to produktionslinjer er ensartet, og

$$\mathrm{P}\{\text{ accept }|\mu_1 - \mu_2 \geq 2\} \leq \beta,$$

i.e. en lille acceptsandsynlighed, hvis kvaliteterne afviger væsentligt fra hinanden.

Det antages at stikprøverne fra de to produktionslinjer skal være lige store ($n = m$). For $\mu_1 - \mu_2 = 2$, $\alpha = 5\%$ og $\beta = 5\%$ fås af tabellen side 340

$$\Delta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \frac{2}{\sqrt{2^2 + 1.5^2}} = 0.8$$

og

$$n = \frac{u_{95\%} - u_{5\%}}{\Delta^2} = \frac{(1.6449 - (-1.6449))^2}{0.8^2} = 16.91$$

Altså fås stikprøvestørrelserne $n = m = 17$. ♦

EKSEMPEL 3.14. Såfremt man i Eksempel 3.13 ønskede at tage hensyn til, at der er forskel på varianserne ($\sigma_1^2 \neq \sigma_2^2$) ved bestemmelsen af stikprøvestørrelserne (idet der nu gælder $\frac{m}{n} = \frac{\sigma_1}{\sigma_2}$) fås i stedet af tabellen side 340:

$$\Delta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_1\sigma_2}} = \frac{2}{\sqrt{2^2 + 2 \cdot 1.5}} = \frac{2}{\sqrt{7}},$$

hvorved man får

$$n = \frac{(u_{95\%} - u_{5\%})^2}{\Delta^2} = \frac{(1.6449 - (-1.6449))^2}{(2/\sqrt{7})^2} = 18.94 \approx 19$$

og

$$m = n\frac{\sigma_2}{\sigma_1} = 18.94\frac{1.5}{2} = 14.2 \approx 15$$

Altså fås den samme totale stikprøvestørrelse $n + m = 34$ i begge tilfælde.

Forhåndsviden om, at $\sigma_1^2 \neq \sigma_2^2$ giver dog anledning til en anden allokering af det totale antal emner i stikprøverne, således at der udtages flest emner fra produktionslinjen med størst variation. ♦

Vi giver nu et meget vigtigt eksempel, der skal illustrere anvendelsen af den teori, vi har formuleret.

EKSEMPEL 3.15 (PARRET/UPARRET T-TEST). For at sammenligne 2 sovemidler A (dextrohyoscyamin hydrobromid) og B (laevo-hyoscyamin hydrobromid) har man målt den forøgede søvnlængde, 10 personer har opnået ved brug af de 2 sovemidler. De opnåede resultater var - målt i timer -

Person	A	B
1	0.7	1.9
2	-1.6	0.8
3	-0.2	1.1
4	-1.2	0.1
5	-0.1	-0.1
6	3.4	4.4
7	3.7	5.5
8	0.8	1.6
9	0.0	4.6
10	2.0	3.4

For at kunne vurdere disse data må vi opstille en matematisk model. Vi antager, at observationerne er realiserede udfald af stokastisk uafhængige, normalt fordelte stokastiske variable $X_1, \cdots, X_{10}$ (sovemiddel A) og $Y_1, \cdots, Y_{10}$ (sovemiddel B). Disse ting må begrundes ved forsøgets art og udførelse (personerne får ikke at vide, hvilket middel de har fået, hvorledes det har virket på andre etc.).

Endvidere antager vi, at

$$X_i \in \mathrm{N}(\mu_i + \delta_1, \sigma_1^2) \quad i = 1, \cdots, 10$$

$$Y_i \in \mathrm{N}(\mu_i + \delta_2, \sigma_2^2) \quad i = 1, \cdots, 10.$$

Her angiver parameteren μ_i, om person nr. i generelt er påvirkelig af sovemidler og i hvor høj grad etc. Parameteren er altså en personparameter. Parametrene δ_1 og δ_2 angiver så effekten ved anvendelse af de specielle sovemidler A henholdsvis B. Disse

er altså personuafhængige. De størrelser, vi må være interesserede i at sammenligne, er altså δ_1 og δ_2. Da vi a priori ikke har nogen fornemmelse af, at den ene skulle være bedre end den anden, må det rimelige test være

$$H_0: \quad \delta_1 = \delta_2 \qquad \text{mod} \qquad H_1: \quad \delta_1 \neq \delta_2.$$

For at kunne udføre dette test betragter vi

$$D_i = X_i - Y_i \in \mathrm{N}(\delta_1 - \delta_2, \sigma_1^2 + \sigma_2^2).$$

Man skal her hæfte sig ved, at de dannede differenser D_i **ikke** er influerede af de enkelte personers underliggende gennemsnitlige søvnlængder, μ_i.

Sætter vi $\delta_1 - \delta_2 = \delta$ og $\sigma_1^2 + \sigma_2^2 = \sigma^2$, er altså $D_i \in \mathrm{N}(\delta, \sigma^2)$, og vi kan derfor sammenligne de to sovemidler ved at teste

$$H_0: \quad \delta = 0 \qquad \text{mod} \qquad H_1: \quad \delta \neq 0.$$

Af tabellen side 336 fremgår, at vi som teststørrelse kan anvende

$$Z = \frac{\bar{D}\sqrt{10}}{\sqrt{\Sigma(D_i - \bar{D})^2/9}},$$

som er $\in \mathrm{t}(9)$ under H_0. Hvis vi anvender et signifikansniveau på 5%, bliver den kritiske område

$$\begin{aligned} \mathrm{C} &= \{z | z < \mathrm{t}(9)_{2.5\%} \vee z > \mathrm{t}(9)_{97.5\%}\} \\ &= \{z | z < -2.262 \vee z > 2.262\}. \end{aligned}$$

De observerede værdier af $D_1, \cdots, D_{10}$ er

$$-1.2, -2.4, -1.3, -1.3, 0.0, -1.0, -1.8, -0.8, -4.6, -1.4,$$

og heraf fås

$$\begin{aligned} \bar{d} &= -1.58 \\ \Sigma d_i^2 &= 38.58 \\ \Sigma(d_i - \bar{d})^2 &= 13.62, \end{aligned}$$

d.v.s.

$$z = \frac{-1.58\sqrt{10}}{\sqrt{13.62/9}} = -4.06 \in \mathrm{C}.$$

Den observerede værdi af teststørrelsen ligger altså i det kritiske område, og vi forkaster derfor hypotesen på et 5% niveau. Vi må derfor antage, at de 2 sovemidler ikke har samme effekt.

For at illustrere vigtigheden af at vælge den korrekte matematiske model kunne vi nu alternativt forestille os, at de anførte målinger stammer fra forsøg med 20 forskellige patienter.

Denne ændring i forudsætningerne har den helt afgørende betydning, at vi nu **ikke** kan eliminere de underliggende gennemsnitlige søvnlængder ved at tage differenser som ovenfor. Derved bliver bedømmelsen af forskellen på de to sovemidler naturligvis mere usikker.

Den rimelige model kunne nu være

$$X_i \in \mathrm{N}(\theta_1, \sigma_1^2), \quad i = 1, \cdots, 10,$$

$$Y_i \in \mathrm{N}(\theta_2, \sigma_2^2), \quad i = 1, \cdots, 10,$$

hvor θ_1 og θ_2 angiver middelsovelængden for en tilfældigt udvalgt person ved at anvende henholdsvis A eller B som sovemiddel. Vi vil nu undersøge, om $\theta_1 = \theta_2$. Først bør vi imidlertid teste, om $\sigma_1^2 = \sigma_2^2$. Det rimelige alternativ er $\sigma_1^2 \neq \sigma_2^2$. Ifølge tabellen side 345 er det kritiske område ved et test på 5% niveauet givet ved

$$\frac{s_1^2}{s_2^2} < \mathrm{F}(9,9)_{2.5\%} \vee \frac{s_1^2}{s_2^2} > \mathrm{F}(9,9)_{97.5\%}$$

Af tabel fås $\mathrm{F}(9,9)_{97.5\%} = 4.03 = 1/\mathrm{F}(9,9)_{2.5\%}$. Beregningerne af teststørrelsen fremgår af følgende skema.

	A	B
Σz_i	7.5	23.3
Σz_i^2	34.43	90.37
$\Sigma(z_i - \bar{z})^2$	28.80	36.08

Den observerede værdi af teststørrelsen z er

$$z = \frac{s_1^2}{s_2^2} = \frac{28.80/9}{36.08/9} = 0.78,$$

der ligger i acceptområdet. Vi kan derfor nu antage, at $\sigma_1^2 = \sigma_2^2$, og vi kan danne et centralt, sammenvejet (pooled) skøn over det fælles σ^2:

$$s^2 = \frac{28.80 + 36.08}{18} = 3.60 = 1.9^2$$

Vi tester nu under antagelse af $\sigma_1^2 = \sigma_2^2$:

$$H_0: \quad \theta_1 = \theta_2 \qquad \text{mod} \qquad H_1: \quad \theta_1 \neq \theta_2 .$$

Af tabellen side 341 får vi, at teststørrelsen er

$$T = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Det kritiske område er - idet vi stadig anvender $\alpha = 5\%$ -

$$t < \mathrm{t}(18)_{2.5\%} = -2.101 \vee t > \mathrm{t}(18)_{97.5\%} = 2.101.$$

Den observerede værdi af T er

$$t = \frac{-1.58}{1.9\sqrt{\frac{1}{5}}} = -1.86.$$

som ligger i acceptområdet, d.v.s. vi må konkludere, at det med det foreliggende materiale ikke kan afvises, at de 2 sovemidler er ens under de sidst benyttede antagelser om forsøget. ♦

Dette resultat er meget lærerigt, idet vi jo under de først anførte forudsætninger om forsøgets udførelse fik, at der var en forskel på de 2 sovemidler. Det er altså af afgørende betydning for undersøgelsens udfald, at forudsætningerne nøje specificeres, at man holder variation inden for patienter og mellem patienter ude fra hinanden etc.

I det viste eksempel bliver usikkerheden meget mindre, hvis man kan undersøge $D_i = X_i - Y_i$ i stedet for som vist i sidste del af eksemplet. Testet, hvor man kan foretage differensdannelse, kaldes ofte et **parret** t-test eller et t-test for **parvise differenser**.

3.2.6 Test i Γ-fordelingen

Vi skal i dette afsnit anføre nogle tests i Γ-fordelingen, når formparameteren k er kendt. Vi bemærker, at dette inkluderer testning i eksponentialfordelingen.

I tabellerne side 355 og 356 har vi dels betragtet uafhængige stokastiske variable $X_1, \cdots, X_n$ med $X_i \in \mathrm{G}(k, \beta)$ og dels uafhængige variable $X_1, \cdots, X_n$ og $Y_1, \cdots, Y_m$ hvor $X_i \in \mathrm{G}(k_1, \beta_1)$ og $Y_i \in \mathrm{G}(k_2, \beta_2)$. Der er anført teststørrelser m.v. for test af β og for test af sammenhængen mellem β_1 og β_2. Signifikansniveauet er overalt lig α.

$X_i \in G(k,\beta) \quad i = 1,\ldots,n$		
Hypotese	Kritisk Område	Styrke
$H_0 : \beta = \beta_0$	$z < \frac{1}{2}\chi^2(2nk)_{\alpha/2}$	$P\left\{\chi^2(2nk) < \frac{\beta_0}{\beta}\chi^2(2nk)_{\alpha/2}\right\}$ +
$H_1 : \beta \neq \beta_0$	$\vee \quad z > \frac{1}{2}\chi^2(2nk)_{1-\alpha/2}$	$P\left\{\chi^2(2nk) > \frac{\beta_0}{\beta}\chi^2(2nk)_{1-\alpha/2}\right\}$
$H_0 : \beta = \beta_0$ el. $\beta \leq \beta_0$ $H_1 : \beta > \beta_0$	$z > \frac{1}{2}\chi^2(2nk)_{1-\alpha}$	$P\left\{\chi^2(2nk) > \frac{\beta_0}{\beta}\chi^2(2nk)_{1-\alpha}\right\}$
$H_0 : \beta = \beta_0$ el. $\beta \geq \beta_0$ $H_1 : \beta < \beta_0$	$z < \frac{1}{2}\chi^2(2nk)_{\alpha}$	$P\left\{\chi^2(2nk) < \frac{\beta_0}{\beta}\chi^2(2nk)_{\alpha}\right\}$

Parameter	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
β (k kendt)	$Z = \frac{1}{\beta_0}\sum_{i=1}^{n} X_i$	$Z \in G\left(nk, \frac{\beta}{\beta_0}\right)$

$X_i \in G(k_1,\beta_1) \quad i=1,\ldots,n \quad \wedge \quad Y_i \in G(k_2,\beta_2) \quad i=1,\ldots,m$		
Hypotese	Kritisk Område	Styrke
$H_0 : \beta_1 = \beta_2$ $H_1 : \beta_1 \neq \beta_2$	$z < F(2nk_1,2mk_2)_{\alpha/2}$ $\vee \quad z > F(2nk_1,2mk_2)_{1-\alpha/2}$	$P\left\{F(2nk_1,2mk_2) < \frac{\beta_2}{\beta_1}F(2nk_1,2mk_2)_{\alpha/2}\right\}$ + $P\left\{F(2nk_1,2mk_2) > \frac{\beta_2}{\beta_1}F(2nk_1,2mk_2)_{1-\alpha/2}\right\}$
$H_0 : \beta_1 = \beta_2$ el. $\beta_1 \leq \beta_2$ $H_1 : \beta_1 > \beta_2$	$z > F(2nk_1,2mk_2)_{1-\alpha}$	$P\left\{F(2nk_1,2mk_2) > \frac{\beta_2}{\beta_1}F(2nk_1,2mk_2)_{1-\alpha}\right\}$
$H_0 : \beta_1 = \beta_2$ el. $\beta_1 \geq \beta_2$ $H_1 : \beta_1 < \beta_2$	$z < F(2nk_1,2mk_2)_{\alpha}$	$P\left\{F(2nk_1,2mk_2) < \frac{\beta_2}{\beta_1}F(2nk_1,2mk_2)_{\alpha}\right\}$

Parametre	Stikprøvefunktion (Teststørrelse)	Fordeling af Z
β_1, β_2 (k_1, k_2 kendte)	$Z = \frac{\sum_{i=1}^{n} X_i/(nk_1)}{\sum_{i=1}^{m} Y_i/(mk_2)}$	$Z \in \frac{\beta_1}{\beta_2}F(2nk_1,2mk_2)$

EKSEMPEL 3.16. Vi betragter igen situationen i eksempel 2.12, p. 238, hvor der var givet 20 målinger af tidsafstanden mellem successive kundeankomster ved kasseapparaterne i et supermarked. Man har en formodning om, at tidsafstanden mellem kundeankomster her er mindre, end den er ved kasserne i en anden afdeling. For at undersøge dette har man ved disse andre kasser ligeledes målt tidsafstanden mellem kundeankomster. Af praktiske grunde har det dog kun været muligt at måle tidsafstanden mellem hver tredie kundes ankomst. Man fik følgende data, igen målt i $\frac{1}{100}$ min.

$$168, 351, 337, 274, 63, 155, 120, 406, 242, 152$$

Kan vi nu på basis af dette materiale slutte, at den ovenfor anførte formodning virker rimelig?

For at besvare dette spørgsmål må vi som sædvanlig først formulere en matematisk model. I eksempel 2.12, p. 238, arbejdede vi ud fra antagelsen, at kundeankomster fulgte en Poisson proces.

Dette medførte, at de anførte 20 observationer kunne opfattes som realiserede udfald af uafhængige $\mathrm{Ex}(\beta_1)$-fordelte stokastiske variable $X_1, \cdots, X_n$. Hvis forudsætningerne om Poisson processen også holder stik for de andre kasser, er de anførte 10 observationer realiserede udfald af uafhængige $\mathrm{G}(3, \beta_2)$-fordelte stokastiske variable, fordi summen af 3 uafhængige ensfordelte $\mathrm{Ex}(\beta_2)$ stokastiske variable følger en $\mathrm{G}(3, \beta_2)$-fordeling, jfr. sætning 1.19, side 122. Vi har altså de uafhængige variable

$$\begin{array}{ll} X_1, \cdots, X_{20}, & X_i \in \mathrm{Ex}(\beta_1) = \mathrm{G}(1, \beta_1) \\ Y_1, \cdots, Y_{10}, & Y_i \in \mathrm{G}(3, \beta_2) \end{array}$$

Middelafstanden mellem kundeankomster ved "kasse 1" er $\mathrm{E}(X_i) = \beta_1$ og ved "kasse 2" $\frac{1}{3}\mathrm{E}(Y_i) = \beta_2$. Vor antagelse kan nu formuleres som $\beta_1 < \beta_2$. Vi må nu undersøge, om det overhovedet kan antages, at $\beta_1 \geq \beta_2$, d.v.s. vi skal teste denne hypotese mod alternativet $\beta_1 < \beta_2$, i.e.

$$H_0: \quad \beta_1 \geq \beta_2 \qquad \text{mod} \qquad H_1: \quad \beta_1 < \beta_2 .$$

Af tabellen side 356 fås, at vi som teststørrelse kan anvende

$$Z = \frac{\Sigma X_i/(20 \cdot 1)}{\Sigma Y_i/(10 \cdot 3)} .$$

Det kritiske område er ved test på 5% niveau

$$\mathrm{C} = \{z | z < \mathrm{F}(40, 60)_{5\%}\} = \{z | z < 0.61\} .$$

Den observerede værdi af teststørrelsen er

$$z = \frac{1467/20}{2368/30} = 0.93 \notin \mathrm{C}.$$

Vi vil derfor acceptere hypotesen H_0, d.v.s. vi kan ikke afvise antagelsen $\beta_1 \geq \beta_2$. Afvigelsen mellem de målte gennemsnitsafstande er derfor ikke signifikant på et 5% niveau, og med det foreliggende datamateriale er der følgelig ingen basis for at hævde, at $\beta_1 < \beta_2$.

Hvis man fra supermarkedets ledelse ikke var tilfreds med denne konklusion, måtte man foretage yderligere indsamling af data. Som et krav til et test baseret på yderligere data kan man f.eks. stille, at sandsynligheden for at acceptere $\beta_1 \geq \beta_2$ skal være mindre end 2.5%, hvis $\beta_1 < \frac{1}{2}\beta_2$. Vi må her endvidere fastlægge, hvor mange observationer vi skal tage fra hver kasse. Da vi kun måler afstandene mellem hver tredie kunde ved "kasse nr. 2", vil det være rimeligt at tage 3 gange så mange observationer fra "kasse 1" som fra "kasse 2".

Vi har altså nu uafhængige stokastiske variable

$$\begin{array}{ll} X_1, \cdots, X_{3m}, & X_i \in \mathrm{G}(1, \beta_1) \\ Y_1, \cdots, X_m, & Y_i \in \mathrm{G}(3, \beta_2) \end{array}$$

og vi skal teste

$$H_0: \quad \beta_1 \geq \beta_2 \qquad \text{mod} \qquad H_1: \quad \beta_1 < \beta_2$$

på niveauet $\alpha = 5\%$. Kaldes styrkefunktionen for testet p_m, skal m vælges så stor, at

$$\beta_1 < \frac{1}{2}\beta_2 \quad \Rightarrow \quad p_m(\beta_1, \beta_2) > 97.5\%.$$

Ifølge tabellen side 356 er

$$p_m(\beta_1, \beta_2) = \mathrm{P}\{\mathrm{F}(6m, 6m) < \frac{\beta_2}{\beta_1}\mathrm{F}(6m, 6m)_\alpha\},$$

d.v.s. vi skal bestemme m, så

$$\mathrm{P}\{\mathrm{F}(6m, 6m) < 2\mathrm{F}(6m, 6m)_{5\%}\} \geq 97.5\%,$$

dvs., at den kritiske værdi $2\mathrm{F}(6m, 6m)_{5\%}$ skal være mindst 97.5% fraktil i $\mathrm{F}(6m, 6m)$-fordelingen. Vi kræver altså

$$\mathrm{F}(6m, 6m)_{97.5\%} \leq 2\mathrm{F}(6m, 6m)_{5\%}.$$

Af tabel fås

m	$\mathrm{F}(6m, 6m)_{97.5\%}$	$2\mathrm{F}(6m, 6m)_{5\%}$
10	1.67	1.31
20	1.43	1.48

Heraf ses, at de stillede krav er opfyldt for $m \geq 20$. Med henblik på en fornyet undersøgelse af kundeankomster ved kasseapparaterne vil man derfor anbefale, at der tages 60 målinger ved "kasse 1" og 20 målinger ved "kasse 2". Den statistiske analyse af disse observationer vil da foregå som før. ♦

3.2.7 Test i polynomialfordelingen

Vi vil i dette og de følgende to afsnit give en række eksempler på tests i polynomialfordelingen.

Lad $\underline{X} = (X_1, \cdots, X_k) \in \mathrm{Pol}(n, p_1, \cdots, p_k)$. Vi betragter da først hypotesen

$$H_0: \quad p_1 = p_{10}, \cdots, p_k = p_{k0}$$

mod alternativet

$$H_1: \quad \exists i(p_i \neq p_{i0}).$$

Det er vanskeligt at finde et eksakt test for denne situation, men vi har følgende asymptotiske resultat.

SÆTNING 3.4. Kvotienttestet på niveau α for test af H_0 mod H_1 er asymptotisk ækvivalent med testet givet ved det kritiske område

$$\mathrm{C} = \left\{ (x_1, \cdots, x_k) | \sum_{i=1}^{k} \frac{(x_i - np_{i0})^2}{np_{i0}} > \chi^2(k-1)_{1-\alpha} \right\}.$$

Bevis. Af sætning 3.3 p. 323 fås, idet vi erindrer, at polynomialfordelingen kun har frie $k-1$ parametre (p. 105), at

$$-2\log \mathrm{q}(X_1, \cdots, X_k) \text{ asymptotisk } \in \chi^2(k-1),$$

hvis H_0 er sand. Her betegner $\mathrm{q}(X_1, \cdots, X_k)$ kvotienttesstørrelsen for test af H_0 mod H_1. Det kan nu vises, at størrelsen

$$-2\log \mathrm{q}(X_1, \cdots, X_k) = -2\log Q$$

og størrelsen

$$\sum_{i=1}^{k} \frac{(X_i - np_{i0})^2}{np_{i0}}$$

har samme grænseværdi for $n \to \infty$, d.v.s

$$\sum_{i=1}^{k} \frac{(X_i - np_{i0})^2}{np_{i0}} \quad \text{asymptotisk} \in \quad \chi^2(k-1)$$

under H_0. Nu er det kritiske område for kvotienttestet

$$\begin{aligned} \mathrm{C}_1 &= \{(x_1, \cdots, x_k) | \mathrm{q}(x_1, \cdots, x_k) < q_0\} \\ &= \{(x_1, \cdots, x_k) | -2 \log \mathrm{q}(x_1, \cdots, x_k) > c_1\}, \end{aligned}$$

hvor $c_1 = -2 \log q_0$ fastlægges ved niveauet α. Heraf fås det i sætningen anførte resultat. ■

BEMÆRKNING 3.7. Ifølge sætning 1.9, p. 105, er $\mathrm{E}(X_i) = np_{i0}$ således, at leddet

$$(x_i - np_{i0})^2$$

angiver kvadratet på afvigelsen mellem det forventede antal i den i'te klasse og det observerede antal i klassen. Størrelsen

$$\sum_{i=1}^{k} \frac{(x_i - np_{i0})^2}{np_{i0}}$$

er da blot et relativt mål for afvigelsen mellem det observerede og det forventede antal i de forskellige klasser, og det er derfor såre rimeligt at forkaste hypotesen, hvis denne afvigelse er for stor. ▼

BEMÆRKNING 3.8. Approksimationerne i sætningen er specielt gode, hvis alle np_{i0} er lige store eller næsten lige store. Endvidere er de gode, hvis alle $np_{i0} \geq 5$ (hvis $k - 1 \geq 6$, må 1 eller 2 af np_{i0}'erne dog godt være små). ▼

EKSEMPEL 3.17. En fabrik af glødelamper oplyser, at brændetiden for en tilfældigt udvalgt lampe falder i intervallerne 0–700 timer, 700–800 timer og over 800 timer med sandsynligheder på henholdsvis 0.2, 0.7 og 0.1 (jfr. eksempel 3.9, p. 330). Ved målinger på 100 lamper fandt man følgende brændetider.

Brændetid	0–700	700–800	800–∞
Antal	14	75	11

Kan det nu fremdeles antages, at de fra fabrikken opgivne sandsynligheder stemmer?

Vi kan opfatte antallene (x_1, x_2, x_3) af lamper i de 3 klasser som et realiseret udfald af $\underline{X} = (X_1, X_2, X_3) \in \mathrm{Pol}(100, p_1, p_2, p_3)$. Vi undersøger, om der er overensstemmelse mellem de anførte data og fabrikkens opgivelser ved at teste

$$H_0: \quad p_1 = 0.2, \quad p_2 = 0.7, \quad p_3 = 0.1$$

mod alle alternativer. Det kritiske område er approksimativt ved test på niveau α

$$\mathrm{C} = \left\{ (x_1, x_2, x_3) | \sum \frac{(x_i - 100p_{i0})^2}{100p_{i0}} > \chi^2(3-1)_{1-\alpha} \right\}.$$

Den observerede værdi af teststørrelsen er

$$\frac{(x_1 - 20)^2}{20} + \frac{(x_2 - 70)^2}{70} + \frac{(x_3 - 10)^2}{10} = \frac{36}{20} + \frac{25}{70} + \frac{1}{10} = 2.26.$$

Da $\chi^2(2)_{70\%} = 2.41$, ser vi, at hypotesen vil blive accepteret på alle niveauer $\leq 30\%$. Der er derfor ingen grund til at tvivle på de fra fabrikken opgivne sandsynligheder. ♦

Ofte er man interesseret i at teste en hypotese

$$H_0: \qquad p_i = p_i(\theta), \qquad i = 1, \cdots, k$$

mod

$$H_1: \qquad \exists i \big(p_i \neq p_i(\theta)\big),$$

hvor $\underline{\theta} = (\theta_1, \cdots, \theta_r)$ er ukendte parametre, som bestemmer p_i. Det forekommer da intuitivt indlysende at indsætte estimatorer $\underline{\hat{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_r)$ i teststørrelsen således, at vi får

$$Z = \sum_{i=1}^{k} \frac{\big(X_i - np_i(\hat{\theta})\big)^2}{np_i(\hat{\theta})}$$

og så igen forkaste for store værdier af Z.

Det viser sig rent faktisk, at dette igen er et fornuftigt test, blot er grænsefordelingen af Z nu en anden, idet man ved hjælp af sætning 3.3 kan indse, at

$$Z \quad \text{asymptotisk} \quad \in \chi^2(k-1-r),$$

d.v.s frihedsgradsantallet er faldet med r svarende til, at vi har estimeret r parametre.

Vi formulerer resultatet af disse overvejelser mere præcist i

SÆTNING 3.5. Kvotienttestet på niveau α for test af

$$H_0: \quad p_i = p_i(\theta) \qquad \text{mod} \qquad H_1: \quad \exists i\big(p_i \neq p_i(\theta)\big),$$

hvor $\underline{\theta} = (\theta_1, \cdots, \theta_r)$ er en ukendt parameter, er asymptotisk ækvivalent med testet givet ved det kritiske område

$$\mathrm{C} = \left\{ (x_1, \cdots, x_k) | \sum_{i=1}^{k} \frac{\big(x_i - np_i(\hat{\theta})\big)^2}{np_i(\hat{\theta})} > \chi^2(k-1-r)_{1-\alpha} \right\},$$

hvor $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_r)$ er et maximum likelihood skøn over θ.

Bevis. Forbigås. Se f.eks. [35][p. 423]. ■

BEMÆRKNING 3.9. Sætningen er ikke helt præcis i udformningen. Der skal således bl.a. gælde, at ingen af parametrene θ_i kan beregnes ud fra de $r-1$ øvrige. Endvidere skal funktionerne $\mathrm{p}_i(\theta)$ have kontinuerte afledede med hensyn til alle θ_i. Endelig skal vi bemærke, at det er tilstrækkeligt, at estimatorerne $\hat{\theta}_1, \cdots, \hat{\theta}_r$ er asymptotisk effciente. Dette indebærer f.eks., at vi kan anvende estimatorer, der kun asymptotisk er lig maximum likelihood estimatorerne. ▼

Vi skal nu betragte nogle anvendelser af sætningen.

3.2.8 Test i kontingenstabel

Vi vil først omtale de såkaldte **kontingenstabeller**. I dette tilfælde er udfaldet af det tilfældige eksperiment karakteriseret ved 2 egenskaber, i.e. den ene egenskab er netop 1 af visse udtømmende og hinanden udelukkende hændelser $A_1, \cdots, A_r$ og den anden

tilsvarende med $B_1, \cdots, B_s$. Eksperimentet er gentaget n gange, og X_{ij} betegner hyppigheden af hændelsen $A_i \cap B_j$. Vi ordner vore data i et skema som

	B_1		B_s
A_1	X_{11}	$\cdots$	X_{1s}
$\vdots$	$\vdots$		$\vdots$
A_r	X_{r1}	$\cdots$	X_{rs}

(3.3)

Det er et skema som (3.3), der benævnes en **kontingenstabel** (engelsk: contingency=mulighed).

Vi sætter sandsynligheden for at få et udfald i klassen $A_i \cap B_j$ lig p_{ij}, i.e.

$$p_{ij} = \mathrm{P}(A_i \cap B_j).$$

Af beviset for sætning 3.4 fås nu, at

$$\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(X_{ij} - np_{ij})^2}{np_{ij}} \quad \text{asymptotisk} \quad \in \chi^2(rs-1). \tag{3.4}$$

Man vil nu hyppigt være interesseret i at undersøge, om der er uafhængighed mellem de to inddelingskriterier, i.e. om

$$\mathrm{P}(A_i \cap B_j) = \mathrm{P}(A_i)\mathrm{P}(B_j) \quad \forall i, j.$$

Her er den marginale sandsynlighed for hændelsen A_i:

$$\mathrm{P}(A_i) = \sum_{j=1}^{s} \mathrm{P}(A_i \cap B_j) = \sum_{j=1}^{s} p_{ij}.$$

Denne størrelse betegnes kort $p_{i\cdot}$, hvor punktummet angiver, at der er summeret over det pågældende index. Tilsvarende defineres $p_{\cdot j}$. Vi kan nu formulere vort problem som et testproblem med hypotese

$$H_0: \quad p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \quad \forall i, j,$$

hvor

$$p_{i\cdot} = \sum_{j=1}^{s} p_{ij}, \quad p_{\cdot j} = \sum_{i=1}^{r} p_{ij},$$

mod alle alternativer.

Vi estimerer parametrene $p_{i\cdot}$ og $p_{\cdot j}$ ved de respektive relative hyppigheder

$$\hat{p}_{i\cdot} = \frac{X_{i\cdot}}{n}, \quad X_{i\cdot} = \sum_{j=1}^{s} X_{ij}$$

$$\hat{p}_{\cdot j} = \frac{X_{\cdot j}}{n}, \quad X_{\cdot j} = \sum_{i=1}^{r} X_{ij}.$$

Vi indsætter nu $p_{ij} = \hat{p}_{i\cdot} \cdot \hat{p}_{\cdot j}$ i (3.4) og får derved størrelsen

$$Z = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\big(X_{ij} - n(X_{i\cdot}/n)(X_{\cdot j}/n)\big)^2}{n(X_{i\cdot}/n)(X_{\cdot j}/n)}. \tag{3.5}$$

Da $\Sigma p_{\cdot j} = \Sigma p_{i\cdot} = 1$, har vi estimeret $s-1+r-1$ parametre. Ifølge (bemærkningerne til) sætning 3.5 er derfor

$$Z \quad \text{asymptotisk} \quad \in \chi^2\big(rs - 1 - (r+s-2)\big) = \chi^2\big((r-1)(s-1)\big)$$

under H_0. Af den samme sætning følger nu direkte

SÆTNING 3.6. I kontigenstabellen (3.3) er kvotienttestet på niveau α for hypotesen om uafhængighed mellem inddelingskriterierne asymptotisk ækvivalent med testet givet ved det kritiske område

$$\mathrm{C} = \Big\{(x_{11}, \cdots, x_{rs}) \Big| \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\big(x_{ij} - n(x_{i\cdot}/n)(x_{\cdot j}/n)\big)^2}{n(x_{i\cdot}/n)(x_{\cdot j}/n)} > \chi^2\big((r-1)(s-1)\big)_{1-\alpha}\Big\}.$$

Vi skal nu give et eksempel på anvendelsen af sætningen.

EKSEMPEL 3.18. På et laboratorium karakteriserer man visse metallegeringer ved deres hårdhed (lille, normal, stor) og ved deres elasticitet (lille, normal, stor). For at få oplyst, om der er sammenhæng mellem disse 2 egenskaber, har man foretaget en analyse af 100 prøver. Man fik følgende data.

Antal		Hårdhed			Total
		Lille	Normal	Stor	
Elasticitet:	Lille	10	8	3	21
	Normal	12	21	7	40
	Stor	7	22	10	39
Total		29	51	20	100

Vi undersøger, om der er nogen sammenhæng ved at teste, om der er uafhængighed mellem inddelingskriterierne.

Vi bestemmer først estimater over de forventede antal i hver klasse. De bliver

Estimeret forventet antal	Lille	Normal	Stor	Total
Lille	6.09	10.71	4.20	21.00
Normal	11.60	20.40	8.00	40.00
Stor	11.31	19.89	7.80	39.00
Total	29.00	51.00	20.00	100.00

Her beregnes f.eks. værdien i klassen (E stor, H lille) som

$$\frac{x_{3\cdot} \cdot x_{\cdot 1}}{n} = \frac{39 \cdot 29}{100} = 11.31.$$

Ved hjælp af dette skema findes den observerede værdi af teststørrelsen (3.5) let. Den er

$$z = \frac{(10-6.09)^2}{6.09} + \cdots + \frac{(10-7.80)^2}{7.80} = 6.182.$$

Det kritiske område er ved et test på niveau α approksimativt lig med

$$\begin{aligned} \mathrm{C} &= \left\{(x_{11},\cdots,x_{33}) | z > \chi^2\big((3-1)(3-1)\big)_{1-\alpha}\right\} \\ &= \left\{(x_{11},\cdots,x_{33}) | z > \chi^2(4)_{1-\alpha}\right\}. \end{aligned}$$

Da $6.182 < \chi^2(4)_{0.90}$, vil vi derfor acceptere hypotesen på alle niveauer mindre end 10%. Med det foreliggende materiale er der derfor ikke påvist sammenhæng mellem de 2 inddelingskriterier. ♦

3.2.9 Homogenitetstestet

Vi vil til slut give et **test for homogenitet af k (observerede grupperede) fordelinger**. Vi har givet et skema af formen

Fordeling nr.	Gruppe nr.			Total
	1	$\cdots$	m	
1	X_{11}	$\cdots$	X_{1m}	n_1
$\vdots$	$\vdots$		$\vdots$	$\vdots$
k	X_{k1}	$\cdots$	X_{km}	n_k
Total	$X_{\cdot 1}$	$\cdots$	$X_{\cdot m}$	n

Her angiver X_{ij} antallet af observationer fra den i'te fordeling i den j'te gruppe. Totalen n_i er lig antallet af observationer fra den i'te fordeling, d.v.s.

$$n_i = X_{i\cdot} = \sum_j X_{ij}.$$

Endvidere har vi igen anvendt betegnelsen $X_{\cdot j} = \sum_i X_{ij}$.

Vi ønsker nu at undersøge hypotesen H_0, at de k observerede fordelinger er tilfældige stikprøver fra den samme population, i.e. at sandsynligheden for at falde i den j'te gruppe ($j = 1, \cdots, m$) er ens for alle fordelinger (homogenitet). Denne sandsynlighed kan derfor sættes lig θ_j under H_0.

Da $(X_{i1}, \cdots, X_{im}) \in \mathrm{Pol}(n_i, \theta_1, \cdots, \theta_m)$, hvis ovenstående hypotese er korrekt, er

$$\sum_{j=1}^{m} \frac{(X_{ij} - n_i\theta_j)^2}{n_i\theta_j} \quad \text{asymptotisk} \quad \in \chi^2(m-1).$$

Af χ^2-fordelingens reproduktivitetssætning (p. 106) følger nu, at

$$\sum_{i=1}^{k}\sum_{j=1}^{m} \frac{(X_{ij} - n_i\theta_j)^2}{n_i\theta_j} \quad \text{asymptotisk} \quad \in \chi^2\big(k(m-1)\big) \tag{3.6}$$

under H_0. Parametrene θ_j estimeres naturligvis ved de tilsvarende relative hyppigheder, i.e.

$$\hat{\theta}_j = \frac{X_{\cdot j}}{n}, \quad j = 1, \cdots, m.$$

Indsættes disse estimatorer i (3.6), fås størrelsen

$$Z = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(X_{ij} - n_i X_{\cdot j}/n)^2}{n_i X_{\cdot j}/n} \tag{3.7}$$

Da vi har estimeret $m-1$ uafhængige parametre (idet $\sum \theta_j = 1$), får vi som før, at

$$Z \quad \text{asymptotisk} \quad \in \chi^2\big(k(m-1) - m + 1\big) = \chi^2\big((k-1)(m-1)\big).$$

Efter disse betragtninger er det nu klart, at vi har

SÆTNING 3.7. Kvotienttestet på niveau α for homogenitetshypotesen H_0 mod alle alternativer er asymptotisk ækvivalent med testet ved det kritiske område

$$\mathrm{C} = \Big\{(x_1 1, \cdots, x_{km}) | \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(x_{ij} - n_i x_{\cdot j}/n)^2}{n_i x_{\cdot j}/n} > \chi^2\big((k-1)(m-1)\big)_{1-\alpha}\Big\}.$$

BEMÆRKNING 3.10. Vi ser, at teststørrelsen beregnes på samme måde som teststørrelsen for uafhængighedstestet i kontigenstabellen. Vi ser endvidere, at de asymptotiske fordelinger ligeledes stemmer overens, således at det, hvad angår de **numeriske** betragtninger, er underordnet, om man opfatter situationen på den ene eller den anden måde. Det er dog vigtigt, at man gør sig klart, at **modellerne** er forskellige. I kontigenstabellen er totalen n givet, og ved homogenitetstestet er det rækkemarginalsummerne $n_1, \cdots, n_k$, der er givne. ▼

Vi illustrerer sætningen med

EKSEMPEL 3.19. På en levnedsmiddelfabrik ønsker man at undersøge 3 forskellige konserveringsmidlers indflydelse på den færdige vares udseende.

Man har derfor bedømt en række prøver behandlet med et af de 3 konserveringsmidler efter udseendet. Man anvendte følgende klasser: Meget fint udseende, Fint udseende, Acceptabelt udseende og Dårligt udseende. Man fik herved data, som fremgår af nedenstående skema.

Behandling	Udseende				Total
	Meget fint	Fint	Acceptabelt	Dårligt	
1	45	27	20	12	104
2	25	10	9	10	54
3	56	47	30	18	151

Vi vil nu teste en hypotese om, at de 3 fordelinger har samme sandsynligheder for at falde i de respektive klasser, mod alle alternativer.

Vi finder først alle marginalsummer.

	1	2	3	4	Total
1	45	27	20	12	104
2	25	10	9	10	54
3	56	47	30	18	151
Total	126	84	59	40	309

Herefter kan vi let beregne estimater for de forventede antal i de enkelte grupper fra de 3 fordelinger.

$n_i\hat{\theta}_j$	1	2	3	4	Total
1	42.41	28.27	19.86	13.46	104.00
2	22.02	14.68	10.31	6.99	54.00
3	61.57	41.05	28.83	19.55	151.00
Total	126.00	84.00	59.00	40.00	309.00

Ved hjælp af dette skema findes den observerede værdi af teststørrelsen til

$$z = \frac{(45-42.41)^2}{42.41} + \cdots + \frac{(18-19.55)^2}{19.55} = 5.269.$$

Det kritiske område ved test på niveau α er approksimativt

$$\begin{aligned} \mathrm{C} &= \{(x_{11},\cdots,x_{34}) | z > \chi^2\big((3-1)(4-1)\big)_{1-\alpha}\} \\ &= \{(x_{11},\cdots,x_{34}) | z > \chi^2(6)_{1-\alpha}\}. \end{aligned}$$

Da $5.269 < \chi^2(6)_{50\%} = 5.35$, vil vi acceptere hypotesen på alle niveauer mindre end 50%. Vi vil derfor ikke på det foreliggende grundlag afvise hypotesen, at de 3 behandlingsmetoder har samme effekt. Man siger også, at hypotesen H_0 ikke er statistisk signifikant. ♦

3.3 Fordelingsfrie tests

I mange situationer vil det ikke være muligt at specificere en bestemt underliggende fordeling, når man har et testproblem. Da kan man ofte med fordel anvende et såkaldt fordelingsfrit eller ikke-parametrisk test. Vi indleder med

3.3.1 Fortegnstestet og Wilcoxon-testet

Vi vil dog først starte med

DEFINITION 3.10. Vi siger, at et test af en statistisk hypotese er **fordelingsfri**, når det bygger på en stikprøvefunktion, der har en fordeling, som under nulhypotesen er uafhængig af fordelingen af de oprindelige observationer. ▲

Læg mærke til, at det oftest forudsættes, at observationerne både er uafhængige og at de har samme (men dog ikke specificerede) fordeling.

Ved en del undersøgelser er man mere interesseret i, hvor stor medianen eller andre fraktiler er, end hvor stor middelværdien er. Lad os eksempelvis betragte studietidsfordelingen ved DTU for bygningsingeniører, der tog afsluttende eksamen i 1965. Man kunne nu være interesseret i et udsagn af følgende art: 90% af de studerende har afsluttet deres studium inden for et tidsrum på $5\frac{1}{2}$ år (= den normerede studietid + 1 år dengang). Observationerne var som følger:

Tid:	$4\frac{1}{2}$år	$5\frac{1}{2}$år	$6\frac{1}{2}$år	$7\frac{1}{2}$år	$8\frac{1}{2}$år	$9\frac{1}{2}$år
Antal:	55	31	19	4	2	2

Hvis vi sætter Z lig med antallet af observationer $\leq 5\frac{1}{2}$ år, er Z binomialtfordelt, dvs. $Z \in \mathrm{B}(113, 0.90)$ under H_0. Under det alternative, $H_1:$ det er færre end 90% af de studerende, der fuldfører i løbet af $5\frac{1}{2}$ år, vil Z være $\in \mathrm{B}(113, p)$, $p < 0.90$. Man vil derfor forkaste for små værdier af Z. Den kritiske værdi kan fastlægges ved kravet

$$\mathrm{P}\{\mathrm{B}(113, 0.90) \leq c\} \leq 5\%.$$

Benyttes en normalfordelingsapproksimation

$$\mathrm{P}\{\mathrm{B}(113, 0.90) \leq c\} \simeq$$
$$\mathrm{P}\{\mathrm{N}(113 \cdot 0.90, 113 \cdot 0.90 \cdot 0.10) \leq c + \frac{1}{2}\} \lesssim 5\%$$

findes

$$c \overset{<}{\sim} 95.95 \simeq 96,$$

d.v.s. vi må klart forkaste hypotesen, da den observerede værdi af Z er 86.

Et test af denne type kaldes et fortegnstest. Dette er fællesbetegnelsen for tests af denne og lignende art. De bygger på følgende princip. Lad $X_1, \cdots, X_n$ være identisk fordelte med $\mathrm{P}\{X \in A\} = p$. Da vil antallet Z af observationer, der falder i A, være $\in \mathrm{B}(n, p)$. Udsagn vedrørende p kan derfor testes som tidligere gennemgået for binomialsandsynligheder (afsnit 3.2.1).

Lad os betragte et andet eksempel:

EKSEMPEL 3.20. En fabrikant af fiskesnører er interesseret i at kunne reklamere med, at trækstyrken for snørerne er 80 kp. Fra et større parti udtoges 5 stk., som blev undersøgt. Man fik følgende resultater for trækstyrken (i kp):

$$81.7, 81.0, 79.9, 81.9, 79.2.$$

Vi antager nu, at disse målinger kan opfattes som realiserede udfald af uafhængige, identisk fordelte stokastiske variable $X_1, \cdots, X_5$, der er symmetrisk fordelte omkring deres middelværdi μ. Dette medfører, at medianen også er lig μ (median=50% fraktil). Vi ønsker at undersøge, om man kan antage H_0, at $\mu = 80$ kp. Under H_0 vil antallet Z af observationer under 80 kp være $\in \mathrm{B}(5, \frac{1}{2})$. Vi forkaster for Z stor (svarer til median < 80) og for Z lille (svarer til median > 80), d.v.s. det kritiske område er

$$\{z < c_1\} \cup \{z_2 > c_2\},$$

hvor $\mathrm{P}\{\mathrm{B}(5, \frac{1}{2}) < c_1\} = \mathrm{P}\{\mathrm{B}(5, \frac{1}{2}) > c_2\} = \frac{\alpha}{2}$ ved test på niveau α. Af binomialfordelingstabel fås for $\alpha = 5\%$ $c_1 \simeq 1$ og $c_2 \simeq 4$. Vi har observeret værdien 2, der falder i acceptområdet, hvorfor hypotesen om, at trækstyrken var 80 kp, ikke kan afvises.

Nu kan man mene, at vi i ovenstående udnytter lovlig lidt af den information, vi besidder. En bedre udnyttelse fås med det såkaldte
Wilcoxon-test. Vi betragter udfaldene $X_i - \mu$, d.v.s. $X_i - 80$, og får

$$1.7, 1.0, -0.1, 1.9, -0.8.$$

Tallene $|x_i - \mu|$ nummereres efter voksende størrelse. Herved defineres tallets **rang** R_i som det nummer, det har i opstillingen. I eksemplet har vi

$$|x_3 - \mu| < |x_5 - \mu| < |x_2 - \mu| < |x_1 - \mu| < |x_4 - \mu|,$$

d.v.s.

$$R_1 = 4, R_2 = 3, R_3 = 1, R_4 = 5, R_5 = 2.$$

Som teststørrelse kan vi nu anvende

$$W = \sum_{X_i > \mu} R_i = \sum_{X_i > 80} R_i,$$

idet fordelingen af W er kendt under H_0, **såfremt** fordelingsfunktionen F for X_i'erne er **kontinuert**.

Vi skal ikke komme ind på beviset for dette resultat, men blot henføre til f.eks. i [40, p. 46]. Nulhypotesefordelingen findes tabelleret (e.g. i [41, p. 325]).

For $n = 5$ fås følgende tabel over de kumulerede sandsynligheder:

w	0	1	2	3	4	5	6	7
$\mathrm{P}\{W \leq w\}$	0.031	0.062	0.094	0.156	0.219	0.312	0.406	0.500
w	8	9	10	11	12	13	14	15
$\mathrm{P}\{W \leq w\}$	0.594	0.688	0.781	0.849	0.906	0.938	0.969	1.000

Det er klart, at vi forkaster for store og små værdier af W, idet små værdier betyder, at vi har få observationer, der er større end μ, og de, der er større, ligger samtidig relativt nær ved μ. Analogt på niveau $\alpha = 5\%$ er

$$\mathrm{C} \simeq \{w < 1\} \vee \{w > 14\}.$$

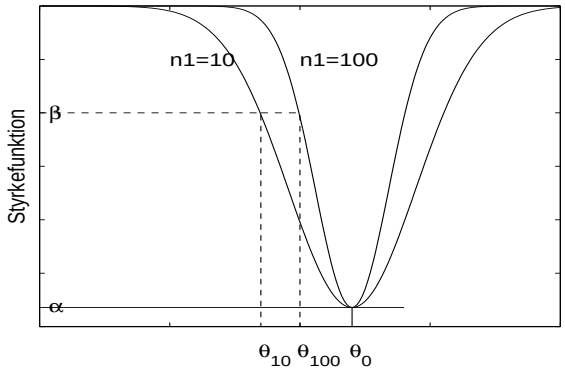
(Dette svarer til et niveau på $\alpha = 3.1\% + (100 - 96.9)\% = 6.2\%$, hvilket er det nærmeste, vi kan komme til 5%). Den observerede værdi af W er i det konkrete tilfælde

$$w = r_1 + r_2 + r_4 = 4 + 3 + 5 = 12.$$

Vi kan altså med det foreliggende materiale ikke afvise fabrikantens påstand. ⧫

Der melder sig nu det naturlige spørgsmål: Hvilken af de to metoder er at foretrække? For at løse dette spørgsmål har man defineret den såkaldte **Pitman-efficiens** af det ene test i forhold til det andet. Formelt: Lad testene ϕ_1 og ϕ_2 have samme niveau α, og

lad ϕ_1 have styrken β i θ_{n_1} med n_1 observationer. Vi fastsætter nu n_2, så test ϕ_2 også har styrke β i θ_{n_1}. Holdes α og β fast, og lades $n_1 \to \infty$, må θ_{n_1} nærme sig θ_0, nulhypoteseværdien.



Hvis nu grænseværdien

$$\lim_{n_1 \to \infty} \frac{n_1}{n_2} = \mathrm{e}_{\phi_2, \phi_1}$$

eksisterer og er uafhængig af α og β, kaldes den **den asymptotiske, relative efficiens** eller **Pitman-efficiensen** af ϕ_2 i forhold til ϕ_1.

Test med stor efficiens behøver altså færre observationer for at skelne et alternativ end det andet test, skønt de har samme niveau. Pitman-efficiensen er som sagt en grænseefficiens, men eksempler viser tit, at brøken n_1/n_2 er tæt ved grænseværdien selv for små n.

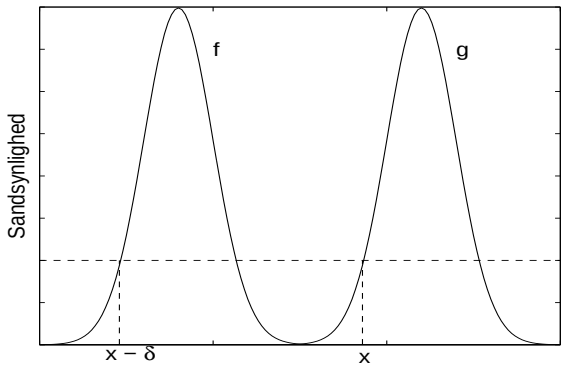
Vi nævner, at Pitman-efficiensen af fortegnstestet i forhold til Wilcoxon-test er ca. 0.7. I forhold til t- eller u-test er den $2/\pi$=0.63. Wilcoxon-tests efficiens i forhold til u- eller t-test er ca. 0.95, når den underliggende fordeling er normal. Hvis fordelingen ikke er normal, kan Wilcoxon-testet i nogle tilfælde være u- eller t-testet overlegent.

Vi forlader nu disse såkaldte **enstikprøveproblemer** og går over til **tostikprøveproblemerne**. Ligesom i det foregående afsnit vil vi af pladshensyn kun kunne give en summarisk indføring i metoderne. Af hensyn til tilegnelsen vil vi gøre dette ved hjælp af eksempler.

For at sammenligne 2 diæter har man fodret 2×5 rotter i 3 uger med hver af diæterne. To rotter, der blev fodret med den ene diæt, døde af eksperimentet uvedkommende grunde. Der foreligger altså observationer af vægttilvæksten: X_1, X_2, X_3 med samme kontinuerte fordelingsfunktion F og $Y_1, \cdots, Y_5$ med den kontinuerte fordeling G. Man har grunde til at tro, at Y-diæten er bedst, altså at Y'erne gennemgående er større end

X'erne. Man vil derfor teste

$$H_0: \quad \mathrm{F}(x) = \mathrm{G}(x) \qquad \text{mod} \qquad H_1: \quad \mathrm{G}(x) = \mathrm{F}(x-\delta), \quad \delta > 0.$$



Vi forudsætter altså (ret restriktivt), at de to fordelinger er identiske **alene** bortset fra en forskydning.

De observerede værdier blev

$$\begin{array}{ll} X: & 5, 3, 10 \\ Y: & 8, 16, 29, 25, 22. \end{array}$$

Idet vi sætter $\mathrm{R}(Z_j)$ lig rangen af observationen Z_j, d.v.s. $\mathrm{R}(Z_j)$ er Z_j's nummer i den ordnede stikprøve, kan det vises, at et test af H_0 mod H_1 kan baseres på teststørrelsen

$$W_x = \sum_{j=1}^{n} \mathrm{R}(X_j),$$

idet denne under H_0 har en fordeling, der er uafhængig af X'ernes og Y'ernes fordeling. (se f.eks. [40, p. 3]. Teststørrelsen W_x kaldes **Wilcoxon's tostikprøveteststørrelse**. Det kritiske område for test af H_0 mod H_1 er af formen.

$$\mathrm{C} = \{(x_1, \cdots, y_m) | w_x \leq c_1\}$$

hvor c_1 fastlægges (på sædvanlig vis) ved tabel over W's fordeling (nulhypotesefordelingen af W_x er rigt tabelleret).

Små værdier af W_x betyder, at X'erne har haft meget lave range, d.v.s. at X'erne må formodes at være de mindste, d.v.s. at H_1 må gælde. Derfor forkaster vi for W_x lille.

Hvis man har et tosidigt alternativ, skal man selvfølgelig også vælge et tosidigt kritisk område.

I vort konkrete tilfælde får vi

X_i	Y_i	$\mathrm{R}(X_i) = r_i$
3		1
5		2
	8	
10		4
	16	
	22	
	25	
	29	
Total		7

Altså er den observerede værdi af W_x lig 7. Af tabel fås, at det kritiske område ved test på niveau 5% er (idet $m = 3$ og $n = 5$ ($m \leq n$) benyttes i tabelopslaget)

$$\mathrm{C} = \{w_x \leq 7\},$$

således at vi vil forkaste hypotesen på dette niveau.

I tilfælde, hvor tabelmaterialet ikke er tilstrækkelig omfattende, kan man anvende

SÆTNING 3.8. Lad $X_1, \cdots, X_n$ og $Y_1, \cdots, Y_m$ være uafhængige stokastiske variable med kontinuerte fordelingsfunktioner F, henholdsvis G. Da har Wilcoxon-teststørrelsen

$$W = \sum_{i=1}^{n} \mathrm{R}(X_i),$$

hvor $\mathrm{R}(X_i)$ er rangen af X_i i den ordnede stikprøve, middelværdi og varians

$$\begin{aligned} \mathrm{E}(W) &= \frac{1}{2} n(m+n+1) \\ \mathrm{V}(W) &= \frac{1}{12} mn(m+n+1), \end{aligned}$$

såfremt $H_0 : \ F = G$ er sand. Endvidere er

$$W \quad \text{approksimativt} \quad \in \mathrm{N}\big(\mathrm{E}(W), \mathrm{V}(W)\big)$$

for $n, m \to \infty$.

Bevis. Forbigås, jfr. [40, p. 31]. ■

EKSEMPEL 3.21. Vi vil nu gennemgå eksempel 3.15, p. 349, ved hjælp af fordelingsfrie metoder.

Vi ser først på den korrekte model, hvor man ser på differenserne $D_1, \cdots, D_{10}$. Vi vil undersøge, om disse D'er kan tænkes at have en fordeling med median 0. Hvis dette er tilfældet, vil vi ikke kunne postulere nogen forskel mellem de 2 sovemidler.

Vi løser først problemet ved hjælp af et **fortegnstest**. Under hypotesen er antallet Z af D'er mindre end eller lig 0 en $\mathrm{B}(10, \frac{1}{2})$-fordelt stokastisk variabel. Den observerede værdi af Z er 1. Sandsynligheden for denne hændelse forenet med hændelsen "0 observationer mindre end 0" er (under H_0)

$$\binom{10}{0} 2^{-10} + \binom{10}{1} 2^{-10} = 11 \cdot 2^{-10} \simeq 0.01.$$

Vi vil derfor forkaste hypotesen på alle niveauer større end 0.02, d.v.s. at hypotesen om, at medianen er lig 0, er statistisk signifikant på et niveau på ca. 1%.

I en situation som ovenstående, hvor man har den ekstreme situation, at alle observationer på nær 1 er strengt større end 0, kan man udmærket anvende et fortegnstest - på trods af dets lavere efficiens - og konkludere, at de 2 sovemidler ikke har samme effekt.

Hvis man ville anvende det mere efficiente **Wilcoxon enstikprøvetest**, må vi forudsætte, at fordelingen af D'erne er kontinuert - hvilket må siges at være en yderst rimelig antagelse. Vi ordner observationernes numeriske værdier og får følgende skema:

i	D_i	$\mathrm{R}(\lvert D_i \rvert)$
1	1.2	4
2	2.4	9
3	1.3	$5\frac{1}{2}$
4	1.3	$5\frac{1}{2}$
5	0.0	1
6	1.0	3
7	1.8	8
8	0.8	2
9	4.6	10
10	1.4	7

Heraf fås, at den observerede værdi af Wilcoxon enstikprøveteststørrelsen er

$$w_1 = \sum_{d_i \leq 0} \mathrm{R}(|d_i|) = 1.$$

Ved hjælp af tabel (se f.eks. [41]) kan man finde

$$\mathrm{P}\{W_1 \leq w_1\} = \mathrm{P}\{W_1 \leq 1\} \simeq 0.001,$$

således at vi vil forkaste hypotesen ved test på alle niveauer større end 0.002. Den sidste teststørrelse er ikke uventet endnu mere signifikant end teststørrelsen fundet ved fortegnstestet.

Vi bemærker iøvrigt, at $X_3 = X_4$. Vi taler da om en såkaldt **tie**. Hvis fordelingen virkelig er kontinuert, er sandsynligheden for at få en tie lig 0. Hvis man alligevel får ties, er det almindeligt da at give en observation den rangværdi, der svarer til gennemsnittet af de rangværdier (midtrange), der skal tilordnes observationerne. I det aktuelle tilfælde er gennemsnittet lig $(6 + 5)/2 = 5.5$. Dette princip anvendes også i andre situationer, hvor man skal bruge rangværdier, og hvor man har observeret ties.

Endelig betragter vi den (hypotetiske) situation, at observationerne stammer fra målinger på 20 forskellige patienter. Da kan vi - stadig under forudsætning af, at fordelingen er kontinuerte - anvende Wilcoxon tostikprøvetest. Vi anfører et skema analogt til det p. 375 givne.

A	B	$\mathrm{R}(A_i)$
-1.6		1
-1.2		2
-0.2		3
-0.1	-0.1	$4\frac{1}{2}$
0.0		6
	0.1	
0.7		8
0.8	0.8	$9\frac{1}{2}$
	1.1	
	1.6	
	1.9	
2.0		14
3.4	3.4	$15\frac{1}{2}$
3.7		17
	4.4	
	4.6	
	5.5	
Total		$80\frac{1}{2}$

Af tabel fås (idet $n = m = 10$), at det kritiske område ved test på niveau 5% er

$$\{w_2 \leq 78\} \cup \{w_2 \geq 132\},$$

således at vi kan acceptere hypotesen om, at sovemidlerne virker ens. Sammenfattende kan vi altså sige, at de fordelingsfrie metoder overalt i dette eksempel fører til samme konklusion som metoderne, der bygger på normalfordelingsantagelsen.

Hvis vi anvender sætning 3.8, fås

$$W_2 \quad \text{approksimativt} \quad \in \mathrm{N}\Big(\frac{1}{2}10 \cdot 21, \frac{1}{12}100 \cdot 21\Big) = \mathrm{N}(105, 175),$$

således at

$$\begin{aligned} \mathrm{P}\{W_2 \leq 80.5\} &\simeq \mathrm{P}\{\mathrm{N}(105, 175) \leq 80.5\} \\ &= \mathrm{P}\{\mathrm{N}(0,1) \leq \frac{80.5 - 105}{\sqrt{175}} \\ &\simeq 3.2\% \end{aligned}$$

Da denne størrelse er større end 2.5%, vil vi igen acceptere på et 5% niveau. ♦

Efter denne gennemgang af fortegns- og Wilcoxon-testene betragter vi

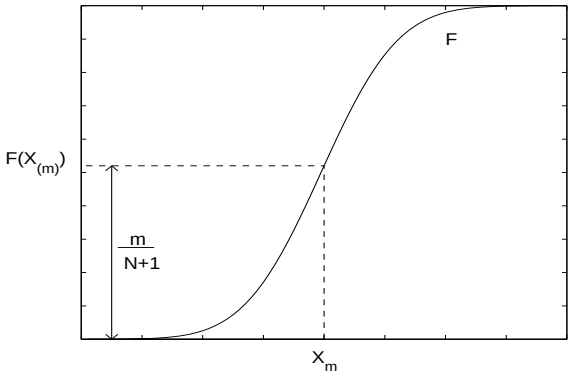
3.3.2 Invers normalvægttest (van der Waerden-test)

Vi vil nu gennemgå endnu et heuristisk test, hvor vi erstatter X'erne med fraktiler i en normalfordeling. Metoden er udviklet af van der Waerden.

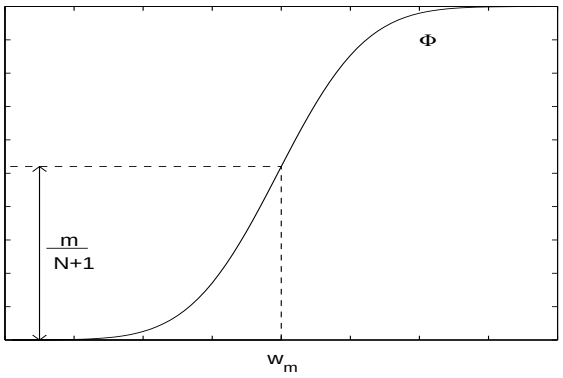
Vi benytter os af følgende resultat om ordnede observationer. Lad $X_1, \ldots, X_N$ være stokastisk uafhængige med fælles kontinuert fordelingsfunktion F. De ordnede observationer benævnes som sædvanlig $X_{(1)}, \ldots, X_{(N)}$. Det kan da vises, at middelværdien af den stokastiske variable $\mathrm{F}(X_{(m)})$ er

$$\mathrm{E}\big(\mathrm{F}(X_{(m)})\big) = \frac{m}{N+1},$$

hvilket grafisk kan anskueliggøres som følger:



Ideen i van der Waerden-testet er så at erstatte $X_{(m)}$ med $\frac{m}{N+1}$-fraktilen w_m i $\mathrm{N}(0,1)$-fordelingen:



hvor altså

$$w_m = \Phi^{-1}\Big(\frac{m}{N+1}\Big)$$

Vi går nu over til selve testproblemet. Lad der som før være givet indbyrdes uafhængige observationer $X_1, \ldots, X_n$ med kontinuert sumfordeling F og $Y_1, \ldots, Y_m$ med kontinuert sumfordeling G. Vi ønsker som før at finde et test for $\mathrm{F}(x) = \mathrm{G}(x)$, der er følsomt over for forskellig placering. Vi sætter $N = n + m$ og ordner som før alle observationer i én stikprøve. X_i's rang i den blandede stikprøve kaldes r_i.

Van der Waerden-teststørrelsen defineres nu ved

$$Z_w = \sum_{i=1}^{n} \Phi^{-1}\Big(\frac{r_i}{N+1}\Big).$$

Såfremt $F = G$ vil

Z_w approksimativt $\mathrm{N}(0, \sigma_w^2)$-fordelt

under grænseovergangen $N \to \infty$, hvor så

$$\sigma_w^2 = \frac{nm}{N(N+1)} \sum_{i=1}^{N} \Big(\Phi^{-1}\big(\frac{i}{N+1}\big)\Big)^2,$$

jfr. [6, p. 159].

Et test på niveau α for $\mathrm{F}(x) = \mathrm{G}(x)$ mod alternativet $\mathrm{G}(x) = \mathrm{F}(x - \delta)$, $\delta \neq 0$, vil derfor tilnærmet være bestemt ved det kritiske område

$$\Big\{\frac{z_w}{\sigma_w} \leq u_{\alpha/2}\Big\} \cup \Big\{\frac{z_w}{\sigma_w} > u_{1-\alpha/2}\Big\}$$

Modifikationerne for ensidede alternativer er åbenbare.

EKSEMPEL 3.22. Vi betragter igen situationen fra eksemplet p. 373.

Vi vil foretage en lignende analyse, blot under anvendelse af van der Waerden's test i stedet for Wilcoxon-testet.

Herefter ordnes materialet, og beregningerne opstilles i følgende skema:

X	Y	r_i	$\Phi^{-1}(\frac{i}{9})$	$\left(\Phi^{-1}(\frac{i}{9})\right)^2$
3		1	-1.22	1.4884
5		2	-0.77	0.5929
	8			0.1849
10		4	-0.14	0.0196
	16			0.0196
	22			0.1849
	25			0.5929
	29			1.4884
			-2.13	4.5716

Af skemaet ses, at den asymptotiske varians σ_w^2 skønnes ved

$$\hat{\sigma}_w^2 = \frac{3 \cdot 5}{8 \cdot 7} \cdot 4.5716 = 1.2245 = 1.11^2.$$

Den observerede værdi

$$\frac{z_w}{\hat{\sigma}_w} = -\frac{2.13}{1.11} = -1.92.$$

Da

$$\mathrm{P}\{Z_w/\hat{\sigma}_w \leq -1.92\} \simeq \Phi(-1.92) = 0.0274,$$

vil hypotesen også ved anvendelse af dette test blive forkastet. ♦

3.3.3 Rangtest for skalaparametre (Siegel-Tukey)

Lad der være givet en stokastisk variabel Z med en kontinuert sumfordeling H, der har median i 0, d.v.s.

$$\mathrm{H}(0) = \mathrm{P}\{Z \leq 0\} = \frac{1}{2}.$$

Lad nu Z_1 og Z_2 være stokastisk uafhængige med fordelingsfunktion H. Vi sætter

$$\begin{aligned} X &= Z_1 + \mu \\ Y &= \theta Z_2 + \mu \qquad \theta > 0. \end{aligned}$$

Da vil frekvensfunktionerne for X og Y, som vi kalder $\mathrm{F}(x)$ og $\mathrm{G}(y)$ være givet ved:

$$\begin{aligned} \mathrm{F}(x) &= \mathrm{P}\{X \leq x\} = \mathrm{P}\{Z_1 + \mu \leq x\} = \mathrm{H}(x-\mu) \\ \mathrm{G}(y) &= \mathrm{P}\{Y \leq y\} = \mathrm{P}\{\theta Z_2 + \mu \leq y\} = \mathrm{H}(\tfrac{y-\mu}{\theta}), \end{aligned}$$

jfr. p. 42. Vi ser nu, at $\mathrm{F}(\mu) = \mathrm{G}(\mu) = \mathrm{H}(0)$, hvorfor X og Y begge har medianen μ. Vi siger som bekendt, at X og Y's fordelinger kun afviger fra hinanden med **skalaparameter** θ. Som eksempel kan vi anføre, at hvis X er $\mathrm{N}(\mu, 1)$-fordelt og Y $\mathrm{N}(\mu, \sigma^2)$-fordelt, da afviger deres fordelinger alene med skalaparameteren σ.

Hvis Z_1 og Z_2 har en varians σ^2, fås

$$\begin{aligned} \mathrm{V}(X) &= \mathrm{V}(Z_1 + \mu) = \sigma^2 \\ \mathrm{V}(Y) &= \mathrm{V}(\theta Z_2 + \mu) = \theta^2\sigma^2 = \theta^2\,\mathrm{V}(X). \end{aligned}$$

Vi ser altså, at i dette tilfælde er $\mathrm{V}(X) = \mathrm{V}(Y)$, såfremt $\theta = 1$.

Vi går nu over til selve testproblemet. Vi observerer indbyrdes uafhængige stokastiske variable $X_1, \ldots, X_n$ med fælles kontinuert fordeling F og $Y_1, \ldots, Y_m$ med fælles kontinuert fordeling G. Vi forudsætter, at F og G har samme median, og at deres fordelinger kun afviger med skalaparameter θ. Vi vil nu teste

$$H_0: \quad \theta = 1, \text{ d.v.s. } F = G \qquad \text{mod} \qquad H_1: \quad \theta \neq 1.$$

Vi danner nu igen én blandet stikprøve af samtlige $m + n$ observationer, ordner disse i stigende rækkefølge og tilforordner range på følgende måde:

Observation nr.	1	2	3	4	$\cdots$	$N-3$	$N-2$	$N-1$	N
Rang	1	4	5	8	$\cdots$	7	6	3	2

dvs. $X_{(1)}$ får rang 1, $X_{(N)}$ rang 2, $X_{(2)}$ rang 3 etc.

Idet vi sætter X_i's rang lig r_i, vil

$$W = \sum_{i=1}^{n} r_i$$

under nulhypotesen følge Wilcoxon-teststørrelsens fordeling (tostikprøvestørrelsens fordeling), d.v.s. det kritiske område bliver

$$\{w \leq w_{\alpha/2}\} \cup \{w > w_{1-\alpha/2}\}$$

hvis alternativet er tosidet. Her angiver w_α α-fraktilen i W's fordeling.

Modifikationerne for ensidede alternativer er åbenbare (cf. nedenstående eksempel).

EKSEMPEL 3.23. Givet observationerne X_1, X_2, X_3 med kontinuert fordelingsfunktion F og observationer $Y_1, \ldots, Y_7$ med kontinuert fordelingsfunktion G. Vi antager, at F og G har samme median og kun adskiller sig ved en skalaparameter som ovenfor. Vi ønsker at teste $H_0: F = G$, d.v.s. $\theta = 1$ mod $H_1: \theta < 1$. Man fik følgende udfald:

$$\begin{array}{lrrrrrrr} X: & -16.5, & 7.1, & -28.1 & & & & \\ Y: & 1.5, & -4.2, & 9.8, & 5.0, & -11.0, & 2.1, & 3.6. \end{array}$$

Vi ordner observationerne som ovenfor beskrevet:

$$\begin{array}{cccccccccc} X_3 & X_1 & & & & & & & X_2 & \\ -28.1 & -16.7 & -11.0 & -4.2 & 1.5 & 2.1 & 3.6 & 5.0 & 7.1 & 9.8 \\ 1 & 4 & 5 & 8 & 9 & 10 & 7 & 6 & 3 & 2 \end{array}$$

Den observerede værdi af Wilcoxon-teststørrelsen er

$$1 + 4 + 3 = 8$$

Da vi har alternativet $\theta < 1$, d.v.s. variansen på Y mindst, vil det kritiske område være

$$\mathrm{C} = \{w \leq w_\alpha\}.$$

Sandsynligheden for at finde mere ekstreme udfald findes i tabellen for Wilcoxons tostikprøvetest. Det kritiske område (med $m = 3$ og $n = 7$) er

$$\mathrm{C} = \{w \leq 8\}$$

ved test på 5% niveau. Vi fandt $w = 8$ og må altså afvise, at skalaparameteren $\theta = 1$. Dvs, vi må antage, at $\mathrm{V}(Y) < \mathrm{V}(X)$. ♦

NB! Det må meget nøje indskærpes, at dette test for variationen kun har mening, såfremt de opstillede forudsætninger holder. F.eks. kan en accept af hypotesen meget vel skyldes, at X'erne og Y'erne har forskellige varianser, men også forskellige medianer. Man bør endvidere gøre sig klart, at der er tale om et test for skalaparametre og ikke for spredninger, idet forkastelse af en hypotese meget vel kan forekommen skønt fordelingerne har samme varians, men hvor forudsætningerne om samme "form" af fordelingerne ikke er til stede.

Kapitel 4

Modelkontrol

I dette kapitel skal vi gennemgå nogle metoder til at undersøge, om observerede data kan tænkes at følge en given model. I de 2 tidligere kapitler har vi altid arbejdet med forudsætningen, at der var givet uafhængige stokastiske variable $X_1, \cdots, X_n$, der fulgte en fordeling $\mathrm{F}(x;\theta)$, som var kendt på nær en ukendt parameter θ. På basis af sådanne forudsætninger har vi så foretaget inferens om θ. Det er klart, at disse analyser mangler en undersøgelse af

1. om observationerne kan tænkes at være realiserede udfald af uafhængige stokastiske variable

2. om den underliggende fordeling kan tænkes at være $\mathrm{F}(x;\theta)$.

Det er disse to problemer, vi vil beskæftige os med i dette kapitel.

4.1 Test for tilfældighed

Vi vil først omtale en metode, der er uafhængig af den underliggende fordeling, nemlig et såkaldt run test.

4.1.1 Run test

Vi betragter først det simple eksperiment at kaste en mønt 20 gange. Udfaldet kan f.eks. være

P P K K P P K K K P K P P P K P K K K P (4.1)

Vi har her 10 P og 10 K, og rækkefølgen virker tilfældig. Havde vi i stedet fået

P K P K P K P K P K P K P K P K P K P K (4.2)

ville de fleste vel nok fatte mistanke til det udførte eksperiment, skønt der stadig er 10 P og 10 K. Hvis vi tilsvarende havde fået

P P P P P P P P P K K K K K K K K K K K (4.3)

ville vi igen mene, at der måtte være noget i vejen.

Hvad er det nu, der adskiller udfaldene i (4.2) og (4.3) fra udfaldene i (4.1). Det er evident ikke den relative hyppighed af P'er, idet den er den samme i de 3 tilfælde. En måde at skelne på er at tælle antallet af runs, hvor en **run** er en række bestående af samme symbol. I (4.1) er antallet af runs lig 11, idet vi har

$$\underbrace{\text{P P}}_{1}\ \underbrace{\text{K K}}_{2}\ \underbrace{\text{P P}}_{3}\ \underbrace{\text{K K K}}_{4}\ \underbrace{\text{P}}_{5}\ \underbrace{\text{K}}_{6}\ \underbrace{\text{P P P}}_{7}\ \underbrace{\text{K}}_{8}\ \underbrace{\text{P}}_{9}\ \underbrace{\text{K K K}}_{10}\ \underbrace{\text{P}}_{11}$$

Det ses umiddelbart, at antallet af runs i (4.2) er 20, og at antallet i (4.3) er 2. Vi ser altså, at den "tilfældige" stikprøve (4.1) har et middelstort runs, hvorimod de "systematiske" stikprøver (4.2) og (4.3) har enten et meget stort eller også et meget lille antal runs.

Disse betragtninger foreslår, at vi som et test for tilfældighed i en stikprøve som (4.1) kan anvende antallet af runs R og forkaste for store og små værdier af R.

Før vi går videre, præciserer vi nu definitionen af de omtalte begreber.

DEFINITION 4.1. Lad der være givet en følge af 2 symboler. Ved en **run** forstår vi en succession af identiske symboler indeholdt mellem forskellige symboler (eventuelt yderpunkter i følgen). Det **totale antal runs** R er antallet af sådanne successioner. ▲

Hvis der er n_1 symboler af den ene slags og n_2 symboler af den anden slags, er det muligt at bestemme fordelingen af R, hvis symbolerne optræder i tilfældig rækkefølge. Vi har nemlig

SÆTNING 4.1. Hvis symbolerne optræder i tilfældig rækkefølge, er fordelingen af det totale antal runs R givet ved frekvensfunktionen

$$f(r) = \begin{cases} 2\dfrac{\binom{n_1-1}{\frac{1}{2}r-1}\binom{n_2-1}{\frac{1}{2}r-1}}{\binom{n_1+n_2}{n_1}} & r \text{ lige} \\ \dfrac{\binom{n_1-1}{\frac{1}{2}r-\frac{1}{2}}\binom{n_2-1}{\frac{1}{2}r-\frac{3}{2}} + \binom{n_1-1}{\frac{1}{2}r-\frac{3}{2}}\binom{n_2-1}{\frac{1}{2}r-\frac{1}{2}}}{\binom{n_1+n_2}{n_1}} & r \text{ ulige} \end{cases} \tag{4.4}$$

Endvidere er middelværdien

$$\mathrm{E}(R) = 2\frac{n_1 n_2}{n_1 + n_2} + 1$$

og variansen

$$\mathrm{V}(R) = 2\frac{n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}.$$

Bevis. Beviset er rent kombinatorisk og forbigås. Se f.eks. [55, p. 148]. ■

BEMÆRKNING 4.1. For store værdier af n_1 og n_2 kan vi ifølge den centrale grænseværdisætning anvende en approksimation med den normale fordeling, idet vi har, at R asymptotisk $\in \mathrm{N}(\mathrm{E}(R), \mathrm{V}(R))$. ▼

Vi kan nu teste, om rækkefølgen af symbolerne er tilfældig, idet vi går frem som anført på p.386. Vi formulerer det som en

SÆTNING 4.2. Lad der være givet en følge af 2 symboler med n_1 af den ene slags og n_2 af den anden slags. Et test for hypotesen om, at symbolerne optræder i tilfældig rækkefølge mod alle alternativer, er da givet ved det kritiske område

$$\{r|r \leq c_1\} \cup \{r|r \geq c_2\},$$

hvor teststørrelsen R er det totale antal runs, og hvor c_1 og c_2 er givet ved

$$\mathrm{F}(c_1) \simeq \alpha/2 \wedge \mathrm{F}(c_2 - 1) \simeq 1 - \alpha/2,$$

hvor F er den til (4.4) svarende fordelingsfunktion.

Bevis. Resultatet følger umiddelbart af overvejelserne på de foregående sider. ■

BEMÆRKNING 4.2. Ved tabelopslag af kritiske værdier i et runtest benyttes følgende relationer

$$\begin{aligned} m &= \min\{n_1, n_2\} \\ n &= \max\{n_1, n_2\}, \end{aligned}$$

idet m og n skal bruges som indgange i tabellen. ▼

EKSEMPEL 4.1. Kan en fordeling af symbolerne som i (4.1) anses for at være tilfældig? Af tabel fås med $\alpha = 5\%$, at c_1 og c_2 er lig

$$\begin{aligned} c_1 &= 6 \\ c_2 &= 16, \end{aligned}$$

idet $n_1 = n_2 = 10$. (NB! Disse må betragtes som givne, hvis niveauet skal være 5%). Da vi havde observeret $r = 11$, vil vi altså acceptere hypotesen om tilfældighed. ♦

Vi går nu over til at betragte problemet, om en række stokastiske variable $X_1, \cdots, X_n$ kan antages at være stokastisk uafhængige. Lad de realiserede udfald af $X_1, \cdots, X_n$ være

$$x_1, \cdots, x_n.$$

Hvis vi nu erstatter et x_i med a, hvis x_i er mindre end medianen i fordelingen, og med et b, hvis den er større, fås en følge som f.eks.

$$a, a, b, a, \cdots, b.$$

Hvis stikprøven består af uafhængige stokastiske variable, vil fordelingen af a'er og b'er være tilfældig. Hvis X_i'erne ikke er uafhængige, vil a'erne og b'erne fordele sig på en systematisk måde. Vi kan derfor teste for uafhængighed ved at danne denne række af a'er og b'er og så undersøge det totale antal runs. Hvis medianen ikke er kendt, kan man anvende den empiriske median. Testet må så opfattes som et betinget test givet den observerede værdi af den empiriske median.

Vi illustrerer denne teknik ved et par eksempler.

EKSEMPEL 4.2. Vi vil undersøge, om de i eksempel 2.12, p. 238, anførte målinger over kundeankomster ved en kasse i et supermarked kan antages at være realisationer af uafhængige stokastiske variable.

Da $x_{(10)} = 46$ og $x_{(11)} = 52$, vil vi inddele materialet i observationer, der er større end $(46 + 52)/2 = 49$, og i observationer, der er mindre end 49. Vi kan stille data op i et skema som anført nedenfor.

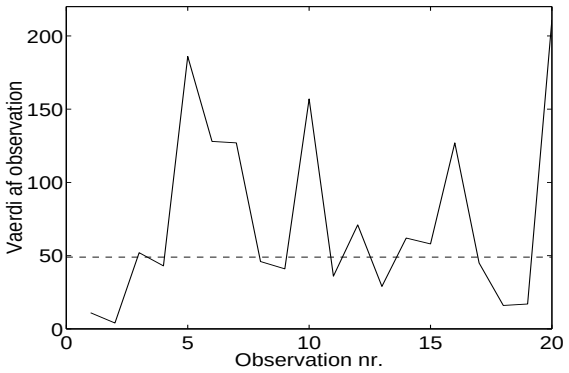
Observation	$a:\ <\ 49$ $b:\ >\ 49$		Observation	$a:\ <\ 49$ $b:\ >\ 49$	
11	a	1	36	a	7
4	a		71	b	8
52	b	2	29	a	9
43	a	3	62	b	
186	b		58	b	10
128	b	4	127	b	
127	b		45	a	
46	a	5	16	a	11
41	a		17	a	
157	b	6	211	b	12

Af tabel fås, at det kritiske område ved et et test på 5% niveau er

$$C = \{r|r \leq 6\} \cup \{r|r \geq 16\}.$$

Da det totale antal runs $= 12 \notin C$, vil vi acceptere hypotesen om tilfældighed. Vi kan derfor ikke afvise den antagelse, at de foreliggende målinger kan opfattes som realiserede udfald af uafhængige stokastiske variable.

Man får et bedre overblik over situationen, hvis man afbilder stikprøverne i et diagram som i nedenstående figur.



Linien parallelt med abscisseaksen angiver medianen. ♦

4.1.2 Gennemsnittet af kvadrerede successive differenser

I dette afsnit vil vi give et test for tilfældighed i det tilfælde, hvor $X_1, \cdots, X_n$ er normalt fordelte.

Vi betragter gennemsnittet af kvadrerede successive differenser

$$\frac{1}{n-1}\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2.$$

Hvis $X_i \in \mathrm{N}(\mu,\sigma^2)$, $i=1,\cdots,n$, og hvis X_i'erne er uafhængige, da vil

$$D_i = X_{i+1}-X_i \in \mathrm{N}(0,2\sigma^2), \qquad i=1,\cdots,n-1,$$

i.e.

$$\mathrm{E}(D_i^2) = \mathrm{V}(D_i) = 2\sigma^2.$$

Heraf fås umiddelbart, at

$$\mathrm{E}\left(\frac{1}{2(n-1)}\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2\right) = \sigma^2.$$

Sætter vi

$$D^2 = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2,$$

er D^2 derfor en central estimator for σ^2 under den antagelse, at observationerne er uafhængige. Det kan vises, at D^2 har større varians (er et mere usikkert skøn) end det sædvanlige centrale skøn over σ^2, nemlig

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

Til gengæld er D^2 mindre følsom over for f.eks. en svag stigning i middelværdien (under forudsætning af konstant varians). Her vil S^2 vokse forholdsmæssigt mere. Betragter vi til gengæld en situation, hvor vi har mange fluktuationer, vil D^2 vokse forholdsmæssigt mere end S^2.

Disse overvejelser giver anledning til at betragte størrelsen

$$R = \frac{D^2}{S^2}.$$

Vi kan nemlig nu anvende R som teststørrelse for et test af hypotesen om, at der er tale om en tilfældig stikprøve. Vi forkaster for store og små værdier af R. Store værdier tyder på hurtige fluktuationer og små værdier på, at der enten er en såkaldt trend (tendens) i datamaterialet eller langsomme cykliske bevægelser. Hvis man kan specificere alternativet til blot at være en af de 2 nævnte, anvender man selvfølgelig et ensidet test.

I nedenstående tabel har vi anført nogle fraktiler for fordelingen af R under H_0, i.e. hvis $X_1, \cdots, X_n$ er uafhængige og identisk normalt fordelte.

n	$r_{0.01}$	$r_{0.05}$	$r_{0.95}$	$r_{0.99}$
4	0.31	0.39	1.61	1.69
5	0.27	0.41	1.59	1.73
6	0.28	0.45	1.56	1.72
7	0.31	0.47	1.53	1.69
8	0.33	0.49	1.51	1.67
9	0.35	0.51	1.49	1.65
10	0.38	0.53	1.47	1.63
11	0.40	0.55	1.45	1.61
12	0.41	0.57	1.43	1.59
15	0.46	0.60	1.40	1.54
20	0.52	0.65	1.35	1.48
25	0.56	0.68	1.32	1.43

Hvis $n > 25$, kan vi anvende følgende

SÆTNING 4.3. For store n har vi approksimativt, at

$$R \in \mathrm{N}\left(1, \frac{n-2}{(n-1)(n+1)}\right).$$

hvis X_i'erne er uafhængige og $\in \mathrm{N}(\mu, \sigma^2)$.

Bevis. Forbigås. ■

Vi illustrerer metoden med et

EKSEMPEL 4.3. I nedenstående tabel er der anført middeludbyttet pr. uge (målt i % af den maksimale kapacitet) for et engelsk koksværk over en periode på et halvt år.

Uge nr.	Udbytte	Uge nr.	Udbytte
1	81.02	14	80.82
2	80.08	15	81.26
3	80.05	16	80.75
4	79.70	17	80.74
5	79.13	18	81.59
6	77.09	19	80.14
7	80.09	20	80.75
8	70.40	21	81.01
9	80.56	22	79.09
10	80.97	23	78.73
11	80.17	24	78.45
12	81.35	25	79.56
13	79.64	26	79.80

Af denne tabel fås

$$\begin{aligned} \sum x_i &= 2081.94 \\ \sum x_i^2 &= 166736.9454 \\ \sum (x_i - \bar{x})^2 &= 26.40016 \\ \sum (x_{i+1} - x_i)^2 &= 31.7348. \end{aligned}$$

Dette giver umiddelbart, at den observerede værdi af R er

$$r = \frac{31.7348}{26.40016} \cdot \frac{1}{2} = 0.60.$$

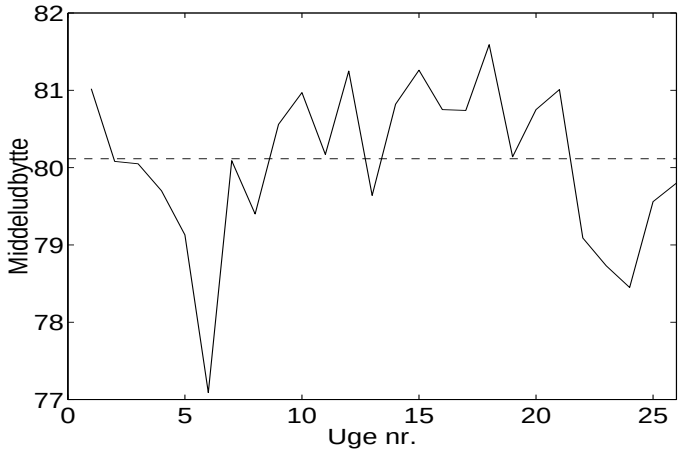
Det kritiske område for et tosidet test af hypotesen om tilfældighed mod alle alternativer er med niveau $\alpha = 10\%$ er

$$C = \{(x_1, \cdots, x_{26}) | r < r_{0.05} \vee r > r_{0.95}\},$$

hvor $r_{0.05} \simeq 0.69$ og $r_{0.95} \simeq 1.32$, hvor vi har ekstrapoleret i tabellen p. 391. Det ses, at $r \in C$, hvorfor vi forkaster hypotesen på et 10% niveau, d.v.s. vi må antage,

at stikprøven ikke består af uafhængige gentagelser af identisk fordelte variable. Da teststørrelsen er lille, tyder dette på, at der er cykliske svingninger i materialet.

Dette synes også at fremgå af nedenstående graf, hvor vi har indtegnet middeludbyttet som funktion af ugenummeret



Vi bemærker, at også et run test ville give denne konklusion. ♦

4.2 Kontrol af fordelingslov

4.2.1 Grafiske metoder

Vi skal nu se på, hvorledes man mest hensigtsmæssigt afbilder empiriske fordelinger grafisk.

Vi betragter først et eksempel med en diskret fordelt variabel.

EKSEMPEL 4.4. En batch bestående af 500 stålplader blev undersøgt for overfladefejl. Observationerne var

Antal fejl	0	1	2	3	4	5	6	7
Antal plader	90	154	133	77	29	13	3	1

Spørgsmålet er nu, om disse fejl optræder tilfældigt?

Ifølge sætning 1.20, p. 122, er dette spørgsmål ensbetydende med at spørge, om antallet af fejl pr. plade følger en Poisson fordeling.

Hvis vi nu indledningsvis går ud fra, at dette er tilfældet, har vi altså, at frekvensfunktionen er

$$\mathrm{f}(x) = \frac{1}{x!}\lambda^x \mathrm{e}^{-\lambda} \qquad x = 0, 1, 2, \cdots .$$

Da vi ikke kender λ, vil vi estimere den. Ifølge skemaet p. 248 er maximum likelihood skønnet for λ lig

$$\begin{aligned} \hat{\lambda} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{1}{500}(90 \times 0 + 154 \times 1 + \cdots + 1 \times 7) = 1.71 \end{aligned}$$

Vi vil nu sammenligne den empiriske fordeling med en $\mathrm{P}(1.71)$-fordeling. For at sikre en rimelig nøjagtighed vil vi nu beregne punktsandsynlighederne i $\mathrm{P}(1.71)$-fordelingen. Af tabel fås

$$\mathrm{f}(0) = \mathrm{e}^{-1.71} = 0.181.$$

Vi beregner nu $\mathrm{f}(1), \mathrm{f}(2), \cdots$ rekursivt, idet vi anvender formlen

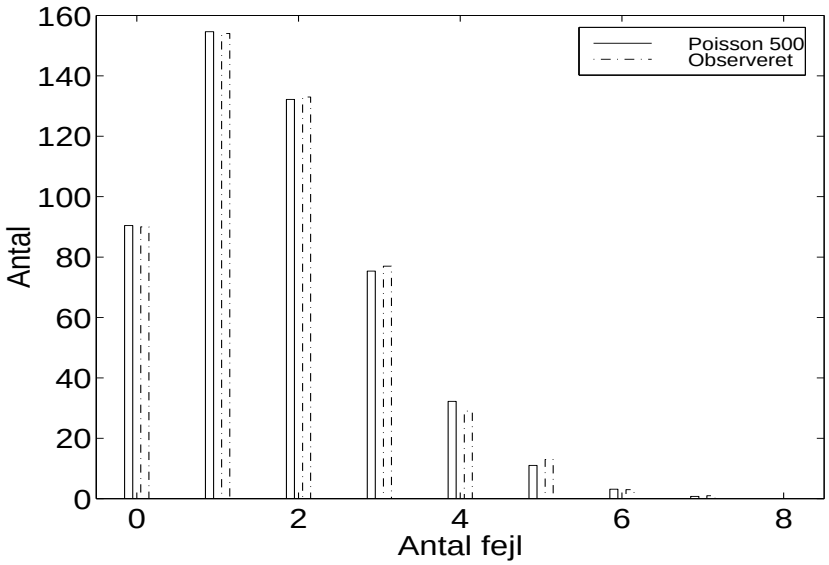
$$\begin{aligned} \mathrm{f}(x) &= \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!} \\ &= \mathrm{e}^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!}\frac{\lambda}{x} \\ &= \mathrm{f}(x-1)\frac{\lambda}{x}. \end{aligned}$$

Ved hjælp af denne rekursionsformel får vi nu let følgende skema

x	$\mathrm{f}(x)$	$500 \cdot \mathrm{f}(x)$	Observeret	$500\mathrm{f}(x)-$obs.
0	0.181	90.5	90	0.5
1	0.310	155.0	154	1.0
2	0.265	132.5	133	-0.5
3	0.151	75.5	77	-1.5
4	0.064	32.0	29	3.0
5	0.022	11.0	13	-2.0
6	0.006	3.0	3	0.0
7	0.001	0.5	1	-0.5
I alt	1.000	500.0	500	0.0

Vi ser, at der er en meget fin overensstemmelse mellem det observerede antal plader med et givet antal fejl og så det antal, man skulle forvente, hvis de fulgte en Poisson fordeling.

Denne overensstemmelse fremgår også af nedenstående figur, hvor vi har sammenlignet de 2 frekvensfunktioner. (Bemærk i øvrigt, at overgangen fra at betragte frekvensfunktionerne til at betragte de forventede antal blot svarer til en skalaændring på ordinataksen).



Ved vurderingen af figuren (eller det dermed ensbetydende skema p. 395) bør man lægge mærke til, at afvigelserne mellem de 2 frekvensfunktioner ikke er systematiske. Hvis man har et tilfælde, der er så åbenbart som ovenstående, vil næppe nogen være i tvivl om, at man bør godtage antagelsen om, at antallet af fejl følger en Poisson fordeling. I andre tilfælde vil det være rart at kunne give en mere præcis regel knyttet til en numerisk størrelse. Dette vender vi tilbage til i næste afsnit. ♦

Hvis man skal sammenligne kontinuerte fordelinger, er det ofte lettere at se på fordelingsfunktionerne i stedet for, som i ovenstående eksempel, frekvensfunktionerne.

Vi betragter en fordelingsfunktion

$$y = \mathrm{H}(x),$$

og vi vil undersøge, om den kan antages at være af samme type som F, der antages at være kontinuert og strengt monoton. Vi vil altså undersøge, om der for et (α, β) med $\beta > 0$ gælder

$$\forall x \left(\mathrm{H}(x) = F\left(\frac{x-\alpha}{\beta}\right) \right).$$

Idet vi med F^{-1} betegner den inverse funktion til F, har vi, at ovenstående er

$$\begin{aligned} &\Leftrightarrow \quad \forall x \Big(F^{-1}\big(\mathrm{H}(x)\big) = \frac{x-\alpha}{\beta} \Big) \\ &\Leftrightarrow \quad \forall x \Big(\big(x, F^{-1}(\mathrm{H}(x))\big) \in \ell \Big), \end{aligned}$$

hvor ℓ er linien med ligningen

$$y = \frac{x-\alpha}{\beta} = \frac{1}{\beta}x - \frac{\alpha}{\beta}.$$

Vi kan derfor undersøge, om H kan antages at være af samme type som F, ved at afbilde punkterne

$$\big(x_i, F^{-1}(\mathrm{H}(x_i))\big), \qquad i = 1, \cdots, n \tag{4.5}$$

i et koordinatsystem. Her er $x_1, \cdots, x_n$ en punktmængde, der ligger passende spredt i definitionsområdet for F og H. Hvis punkterne (4.5) ligger på eller tilnærmelsesvis på en ret linie, vil F og H være af samme type, henholdsvis tilnærmelsesvis af samme type.

DEFINITION 4.2. En afbildning af punkterne (4.5) i et 2-dimensionalt koordinatsystem kaldes et **fraktildiagram**. ▲

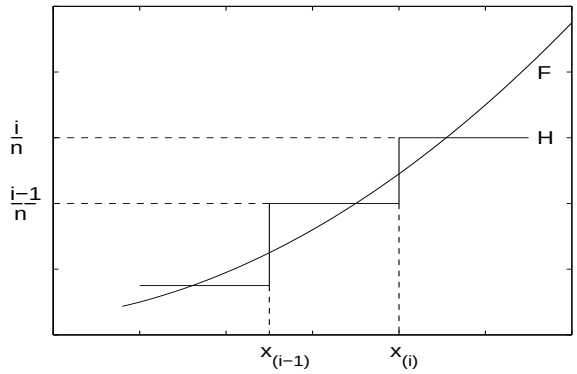
Der findes ark, hvor transformationen F^{-1} er fortrykt på ordinataksen. Mest almindelige er ark til sammenligning med en normal fordeling og en logaritmisk normal fordeling. Det er endvidere ikke svært at se, at almindeligt logaritmepapir kan bruges til at undersøge, om en fordeling er en $\mathrm{Ex}(\beta)$-fordeling.

Med henblik på at undersøge, om et realiseret udfald af den empiriske fordelingsfunktion kan antages at følge en kendt fordeling, betragter vi et øjeblik igen definition 4.2. Vi erindrer, af $\mathrm{F}_n(x)$ er en stokastisk variabel. Vi kalder det realiserede udfald af $\mathrm{F}_n(x)$ for $\mathrm{H}(x)$. Vi har da

$$\mathrm{H}(x) = \begin{cases} 0 & x < x_{(1)} \\ \frac{i}{n} & x_{(i)} \le x < x_{(i+1)} \quad i = 1, 2, \cdots, n-1, \\ 1 & x_{(n)} \le x \end{cases}$$

hvor $x_{(1)}, \cdots, x_{(n)}$ er de ordnede udfald.

Hvis vi vil sammenligne H med fordelingsfunktionen F, kan vi som ovenfor angivet gøre dette ved hjælp af et fraktildiagram.



Som det fremgår af tegningen, vil vi begå en systematisk skævhed, hvis vi afsætter punkterne

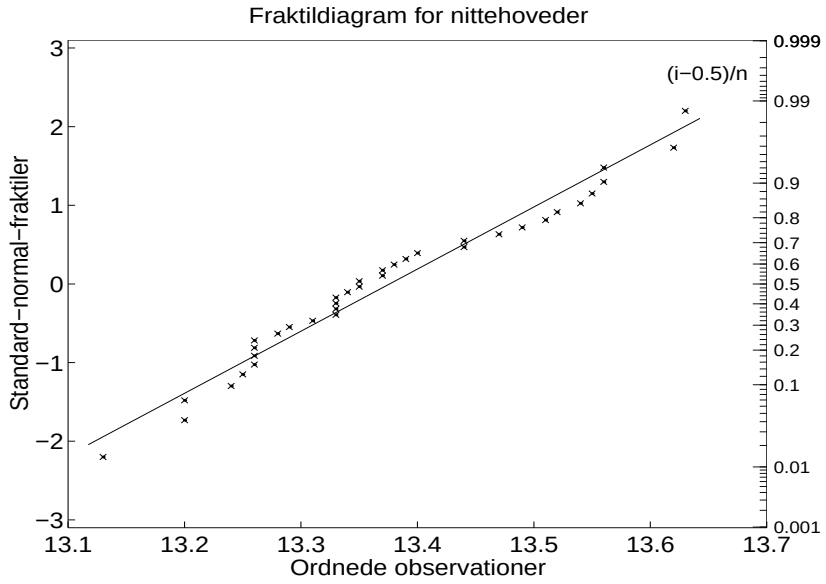
$$\Big(x_{(i)}, F^{-1}\big(\mathrm{H}(x_{(i)})\big)\Big) = \Big(x_{(i)}, F^{-1}(\frac{i}{n})\Big) \quad i = 1, \cdots, n.$$

Det vil være mere rimeligt at sammenligne $\mathrm{F}(x_{(i)})$ med midtpunktet mellem de 2 vandrette stykker, i.e. med $(i - \frac{1}{2})/n$. Vi skal altså i stedet indtegne punkterne

$$\Big(x_{(i)}, F^{-1}\Big(\frac{i - \frac{1}{2}}{n}\Big)\Big) \qquad i = 1, \cdots, n.$$

EKSEMPEL 4.5. Vi betragter de i eksempel 1.1 p. 88 og eksempel 2.13, p. 241, anførte målinger af diameteren af 36 nittehoveder. Vi postulerede i eksempel 2.13, at fordelingen var normal. Denne hypotese vil vi nu undersøge. I figuren herunder er indtegnet et fraktil diagram for observationerne, hvor den rette linje svarer til en $\mathrm{N}(\hat{\mu}, \hat{\sigma}^2)$-fordeling. Centrale skøn over μ og σ^2 er udregnet til

$$\begin{aligned} \hat{\mu} &= 13.376 \\ \hat{\sigma}^2 &= 0.125^2 \end{aligned}$$



Af figuren fremgår det at de stjerne-markerede punkter, der beskriver den empiriske fordeling, er fordelt rimeligt omkring den rette linje. Dvs. vi kan godtage antagelsen om at observationerne følger en normalfordeling.

♦

EKSEMPEL 4.6. Vi betragter igen de i eksempel 2.12, p. 238 anførte målinger over kundeankomsttider. Vi ønsker at undersøge, om antagelsen om, at de anførte data kan beskrives ved en eksponentialfordeling, er rimelig. Vi vil foretage denne sammenligning ved hjælp af et fraktildiagram.

Vi har altså, at

$$\mathrm{F}(x) = 1 - \mathrm{e}^{-x},$$

og det ses, at

$$F^{-1}(p) = -\log_e(1-p).$$

Heraf følger, at vi skal afsætte punkterne

$$\left(x_{(i)}, -\log_e(1 - \frac{i - \frac{1}{2}}{n})\right) \qquad i = 1, \cdots, n$$

i et retvinklet koordinatsystem, hvor disse punkter i bekræftende fald skulle gruppere sig om en ret linie med hældningen $\frac{1}{\beta}$. Heraf får vi, at punkterne

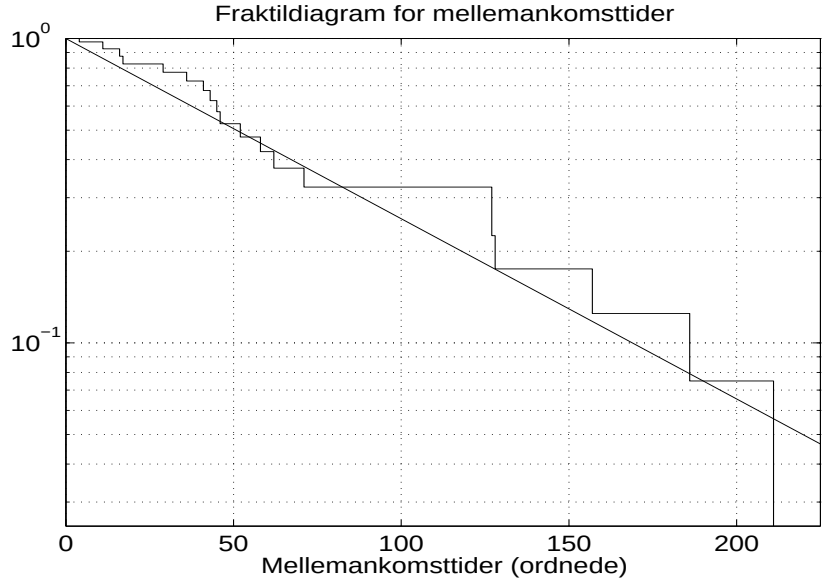
$$(x_{(i)}, 1 - \frac{i - \frac{1}{2}}{n}) \qquad i = 1, \cdots, n$$

indtegnet på almindeligt logaritmepapir igen vil fordele sig om en ret linie, hvis $\mathrm{H}(x)$ minder om en eksponentialfordeling.

Vi ordner observationerne og anfører de for tegning af fraktildiagrammet nødvendige størrelser i nedenstående tabel.

i	$x_{(i)}$	$1 - \frac{i-\frac{1}{2}}{n}$	i	$x_{(i)}$	$1 - \frac{i-\frac{1}{2}}{n}$
1	4	0.975	11	52	0.475
2	11	0.925	12	58	0.425
3	16	0.875	13	62	0.375
4	17	0.825	14	71	0.325
5	29	0.775	15	127	0.275
6	36	0.725	16	127	0.225
7	41	0.675	17	128	0.175
8	43	0.625	18	157	0.125
9	45	0.575	19	186	0.075
10	46	0.525	20	211	0.025

Herefter er det intet problem at tegne fraktildiagrammet, der bliver som vist i figuren herunder.



På grafen har vi endvidere indtegnet linien svarende til

$$\mathrm{F}(x) = 1 - \exp(-\frac{x}{73.35}),$$

idet vi jo erindrer, at maximum likelihood skønnet for β var 73.35. Denne linie fås f.eks. ved at forbinde punkterne

$$\big(0, 1 - F(0)\big) \wedge \big(146.70, 1 - (1 - \mathrm{e}^{-2})\big),$$

i.e. punkterne

$$(0, 1) \wedge (146.70, 0.135).$$

Det ses, at den trappekurve, den empiriske fordeling beskriver, fordeler sig rimeligt om linien, således at vi kan godtage antagelsen, at observationerne følger en eksponentialfordeling. ♦

Vi repeterer nu begrebet en gruppering og definitionen på histogrammet for at få en samlet fremstilling her (jvf. p. 88).

Hvis der foreligger mange observationer fra en kontinuert fordeling, bliver ovenstående fremgangsmåde med at se på data enkeltvis for besværlig. Vi foretager da en såkaldt **gruppering** af materialet, d.v.s. vi inddeler det interval, der indeholder alle observationerne, i et antal (lige store) delintervaller. Vi tænker os nu, at alle observationer, der ligger i et delinterval, er jævnt fordelt over delintervallet. En sådan inddeling giver anledning til

DEFINITION 4.3. Den tæthedsfunktion, som er konstant i ethvert delinterval, og hvis integral over ethvert interval er proportional med antallet af observationer i intervallet, kaldes det til inddelingen svarende **histogram**. ▲

DEFINITION 4.4. Den til et histogram svarende fordelingsfunktion kaldes en **sumpolygon**. ▲

Når man i en konkret situation skal vælge en inddeling, kan man som håndregel anvende følgende formel til bestemmelse af det nødvendige antal intervaller

$$k = 1 + 1.44 \cdot \log_e n.$$

Det må dog indskærpes, at der tale om en empirisk begrundet regel, som er udmærket i mange situationer, men som ingenlunde garanterer en fornuftig inddeling.

Ved en gruppering er det endvidere hensigtsmæssigt at angive klassegrænserne med én decimal mere end observationerne, således at der ikke opstår tvivl om, hvilken klasse en observation skal anbringes i.

Hvis man vil sammenligne sumpolygonen H for et grupperet materiale med en fordelingsfunktion F, afbilder man altså punkterne

$$\Big(t_i, F^{-1}\big(\mathrm{H}(t_i)\big)\Big) \qquad i = 1, \cdots, n,$$

hvor t_i er øvre intervalgrænse i klasse nr. i, i et fraktildiagram.

Vi giver et eksempel til belysning af den ovenfor omtalte teknik.

EKSEMPEL 4.7. Vi betragter de i eksempel 1.1 p. 88 og eksempel 2.13, p. 241, anførte målinger af diameteren af 36 nittehoveder. Vi postulerede i eksempel 2.13, at fordelingen var normal. Denne hypotese vil vi nu undersøge. Vi har, at

$$\begin{aligned} x_{(1)} &= 13.13 \\ x_{(36)} &= 13.63. \end{aligned}$$

Vi vælger derfor - som det også er gjort i eksempel 1.1 - at inddele intervallet $(13.105, 13.705)$, der omslutter alle observationer i 6 lige store dele. Vi bemærker, at intervalendepunkterne har én decimal mere end målingerne på nittehovedet. Tallet 6 er valgt i overensstemmelse med den ovenfor anførte håndregel ($1+1.44\log_e(36) = 6.1$).

Vi bestemmer nu antallet af observationer i hver klasse og får de p. 88–90 anførte skemaer og figurer. Vi kunne nu undersøge normalitetsantagelsen ved at indtegne den teoretiske tæthedsfunktion henholdsvis fordelingsfunktion i figurerne svarende til histogram og sumpolygon. Dette udelader vi dog og går direkte over til at betragte fraktildiagrammet.

På et såkaldt sandsynlighedspapir kan afsættes punkterne

$$(13.205, 0.08), \cdots, (13.605, 0.94).$$

Vi indtegner endvidere linien, der svarer til en $\mathrm{N}(\hat{\mu}, \hat{\sigma}^2)$-fordeling, hvor

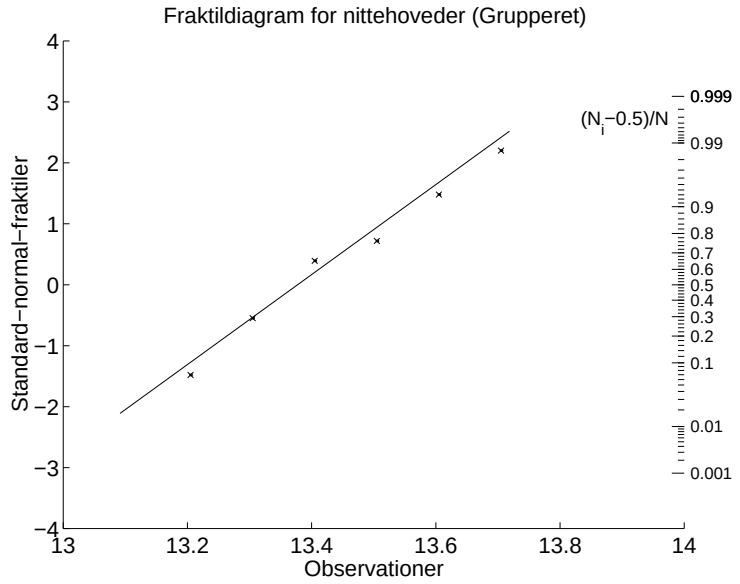
$$\begin{aligned} \hat{\mu} &= 13.376 \\ \hat{\sigma}^2 &= 0.125^2 \end{aligned}$$

er maximum likelihood skøn over μ og σ^2. Da

$$\begin{aligned}\Phi\left(\frac{\mu-\mu}{\sigma}\right) &= 50\% \\ \Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right) &= 84.1\% \\ \Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right) &= 15.9\%,\end{aligned}$$

ses det, at linien svarende til $\mathrm{N}(\mu,\sigma^2)$ går gennem punkterne

$$(\mu, 50\%) \wedge (\mu+\sigma, 84.1\%) \wedge (\mu-\sigma, 15.9\%).$$



Vi bemærker, at punkterne grupperer sig tilfældigt om linien, hvorfor vi ikke vil afvise hypotesen om normalitet. ♦

Vi vil ikke fordybe os mere i metoder til grafisk testning og til empirisk formulering af en fordelingslov, men blot henvise til [25], hvor der især, hvad angår normalfordelingen, findes en indgående beskrivelse af metoderne. I [23] findes der afbildet sandsynlighedspapir til brug for undersøgelser af ekstremværdifordelinger.

4.2.2 Tests for fordelingstype

De i foregående afsnit anførte grafiske metoder til undersøgelser af empiriske fordelinger er - som tidligere nævnt - tests, hvor man afgør, om et udfald er kritisk, ved at vurdere en graf. Disse tests kan suppleres med numeriske do. ved hjælp af sætning 3.5, p. 363, om test i en polymial fordeling. Metoden er den simple, at man tilnærmer de konkrete fordelinger med polynomialfordelinger og så udfører tests i disse.

Vi illustrerer teknikken med nogle eksempler. Vi betragter først situationen i eksempel 4.4, p. 394.

EKSEMPEL 4.8. Vi ser på hændelserne

$$\{X=0\}, \{X=1\}, \{X=2\}, \{X=3\}, \{X=4\}, \{X \geq 5\}$$

Under antagelsen, at de variable er Poisson fordelte, har vi estimeret sandsynlighederne, og de forventede antal i hver klasse fremgår af nedenstående skema.

x	Forventet antal	Observeret antal	$\frac{(\text{Obs.}-\text{forv.})^2}{\text{forv.}}$
0	90.5	90	0.003
1	155.0	154	0.006
2	132.5	133	0.002
3	75.5	77	0.030
4	32.0	29	0.281
≥ 5	14.5	17	0.431
Total	500.0	500	0.753

Vi har slået hændelsen $\{X = i\}$, $i \geq 5$, sammen i én for at få et forventet antal i klassen, der er større end eller lig med 5 (jfr. bemærkning 3.6, p. 323). Det er nu åbenbart, at antallene af observationer i de 6 klasser er en polynomialt fordelt variabel $(Z_1, \cdots, Z_6)$. Under hypotesen, at de oprindelige variable er Poisson fordelte, har vi estimeret de i skemaet anførte forventede værdier. Et test på niveau α er derfor (ifølge sætning 3.5, p. 363) approksimativt givet ved det kritiske område

$$C = \left\{ (x_1, \cdots, x_n) | \sum_{i=1}^{6} \frac{\left(z_i - np_i(\hat{\lambda})\right)^2}{np_i(\hat{\lambda})} > \chi^2(6-1-1)_{1-\alpha} \right\}.$$

Her betegner (for $i = 1, \cdots, 5$)

$$p_i(\hat{\lambda}) = \mathrm{e}^{-\hat{\lambda}} \frac{\hat{\lambda}^{i-1}}{(i-1)!}$$

og

$$p_6(\hat{\lambda}) = \sum_{i=5}^{\infty} \mathrm{e}^{-\hat{\lambda}} \frac{\hat{\lambda}^i}{i!},$$

hvor

$$\hat{\lambda} = \sum x_i/n.$$

Antallet af frihedsgrader bliver som anført

$$6 - 1 - 1 = 4$$

idet antallet af led minus 1 er lig $6 - 1$, og vi har så estimeret en parameter, nemlig $\hat{\lambda}$, således at vi skal trække yderligere 1 fra.

Den observerede værdi af teststørrelsen er 0.753. Ved test på 5% testniveau er den kritiske værdi $\chi^2(4)_{0.95} = 9.488$, medens den fundne værdi $0.753 \leq \chi^2(4)_{0.10} = 1.064$, dvs. meget langt fra signifikant. Vi kan altså acceptere hypotesen om Poisson fordelingen på alle niveauer $\leq 90\%$, d.v.s. Poisson fordelingen giver en udmærket beskrivelse af de anførte data over antallet af fejl pr. stålplade. ♦

Vi fortsætter nu undersøgelserne i eksempel 4.7.

EKSEMPEL 4.9. Vi kan uden videre danne et test analogt til ovenstående for den hypotese, at de i eksempel 4.7 anførte data er realiserede udfald af normalt fordelte variable. Vi vil imidlertid ændre løsningen en smule for at opnå et bedre test. Vi hæfter os igen ved bemærkning 3.6, p. 323, hvor det er anført, at approksimationerne med χ^2-fordelingen er specielt gode, hvis alle p_{i0} er lige store. Vi vil derfor danne en ny klasseinddeling, der sikrer, at dette krav er opfyldt.

Vi ønsker fortsat at have 6 klasser. Vi kan derfor som delepunkter anvende $\left(\frac{100}{6} \cdot i\right)$ %-fraktilerne $(i = 1, \cdots, 5)$ i den normale fordeling, vi vil sammenligne vore data med.

Vi bemærker nu, at hvis vi sætter p-fraktilen i en fordeling F lig x_p, da er p-fraktilen i den til F svarende fordeling med positionsparameter α og skalaparameter β lig

$$z_p = \beta x_p + \alpha,$$

idet

$$p = \mathrm{P}\{Z \leq z_p\} = \mathrm{P}\{\beta X + \alpha \leq z_p\} = \mathrm{P}\left\{X \leq \frac{z_p - \alpha}{\beta}\right\}$$

medfører, at

$$\frac{z_p - \alpha}{\beta} = x_p \quad \Leftrightarrow \quad z_p = \beta x_p + \alpha.$$

Ved hjælp af denne relation får vi nu let fraktilerne i $\mathrm{N}(13.376, 0.125^2)$-fordelingen. Vi samler resulaterne i følgende skema

$p = (\frac{100}{6} \cdot i)\%$	u_p	$\mathrm{N}(13.376, 0.125^2)_p$
$16\frac{2}{3}\%$	-0.97	13.255
$33\frac{1}{3}\%$	-0.43	13.322
50 %	0.	13.376
$66\frac{2}{3}\%$	0.43	13.430
$83\frac{1}{3}\%$	0.97	13.497

Svarende til disse punkter danner vi klasseinddelingen

Klasse	Antal	Forventet antal	$\frac{(\text{Obs.}-\text{forv.})^2}{\text{forv.}}$
$(-\infty, 13.255)$	||||| = 5	6	$\frac{1}{6}$
(13.255,13.322)	||||| || = 7	6	$\frac{1}{6}$
(13.322,13.376)	||||| |||| = 9	6	$\frac{9}{6}$
(13.376,13.430)	||| = 3	6	$\frac{9}{6}$
(13.430,13.497)	|||| = 4	6	$\frac{4}{6}$
$(13.497, \infty)$	||||| ||| = 8	6	$\frac{4}{6}$
I alt	36	36	$4\frac{2}{3}$

Teststørrelsen skal sammenlignes med $1 - \alpha$-fraktilen i en χ^2-fordeling med

$$6 - 1 - 2 = 3$$

frihedsgrader, idet vi har estimeret 2 parametre ved beregningen af

$$\sum_{i=1}^{6} \frac{\left(X_i - np_i(\hat{\mu}, \hat{\sigma}^2)\right)^2}{np_i(\hat{\mu}, \hat{\sigma}^2)}$$

Ved test på f.eks. 5% niveau er den kritiske værdi $\chi^2(3) = 7.814$. Vi forkaster for store værdier af teststørrelsen, og da faktisk $4\frac{2}{3} < \chi^2(3)_{0.90} = 6.25$, accepteres hypotesen på alle niveauer mindre end 10%. Vi må altså konkludere, at det foreliggende materiale kan beskrives ved en normal fordeling. ⧫

Til sidst i dette afsnit vil vi gennemgå det såkaldte **Kolmogorov-Smirnov test**. Vi betragter en række uafhængige, identisk fordelte stokastiske variable $X_1, \cdots, X_n$ med **kontinuert** fordelingsfunktion F. Vi ønsker at teste hypotesen

$$H_0: \quad F = F_0 \qquad \text{mod} \qquad H_1: \quad F \neq F_0,$$

hvor F_0 er en kendt, kontinuert fordelingsfunktion. Der gælder følgende sætning

SÆTNING 4.4 (KOLMOGOROV-SMIRNOV). Under H_0 er fordelingen af

$$D_n = \sup_x |\mathrm{F}_n(x) - \mathrm{F}_0(x)|,$$

hvor F_n betegner den empiriske fordelingsfunktion for $X_1, \cdots, X_n$, kendt og uafhængig af F_0. Et test på niveau α er derfor givet ved det kritiske område

$$C = \{(x_1, \cdots, x_n) | d_n > c\},$$

hvor c fastlægges ved

$$\mathrm{P}\{D_n > c | F = F_0\} = \alpha.$$

Bevis. Forbigås. Se f.eks. [35][p. 452]. ■

I nedenstående tabel har vi anført nogle kritiske grænser for teststørrelsen D_n.

n	$\mathrm{d}(n)_{0.80}$	$\mathrm{d}(n)_{0.90}$	$\mathrm{d}(n)_{0.95}$	$\mathrm{d}(n)_{0.99}$
1	0.90	0.95	0.98	1.00
2	0.68	0.78	0.84	0.93
3	0.57	0.64	0.71	0.83
4	0.49	0.56	0.62	0.73
5	0.45	0.51	0.57	0.67
6	0.41	0.47	0.52	0.62
7	0.38	0.44	0.49	0.58
8	0.36	0.41	0.46	0.54
9	0.34	0.39	0.43	0.51
10	0.32	0.37	0.41	0.49
11	0.31	0.35	0.39	0.47
12	0.30	0.34	0.38	0.45
13	0.28	0.33	0.36	0.43
14	0.27	0.31	0.35	0.42
15	0.27	0.30	0.34	0.40
16	0.26	0.30	0.33	0.39
17	0.25	0.29	0.32	0.38
18	0.24	0.28	0.31	0.37
19	0.24	0.27	0.30	0.36
20	0.23	0.26	0.29	0.36
25	0.21	0.24	0.27	0.32
30	0.19	0.22	0.24	0.29
35	0.18	0.21	0.23	0.27
40	0.17	0.19	0.21	0.25
45	0.16	0.18	0.20	0.24
50	0.15	0.17	0.19	0.23
>50	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Tabel over $1-\alpha$-fraktilen i fordelingen af D_n

Man kan naturligvis også bruge χ^2-testet til at teste H_0 mod H_1. Fordelen ved at bruge Kolmogorov-Smirnov testet er bl.a., at testet er eksakt og har større styrke end χ^2-testet. Det må her indskærpes, at vi **ikke** kan bruge Kolmogorov-Smirnov testet, når vi har estimeret en parameter i fordelingen. I sådanne tilfælde er vi henvist til at anvende χ^2-testet.

Vi illustrerer nu metoden ved hjælp af

EKSEMPEL 4.10. Vi betragter de i eksempel 2.20, p. 256, anførte data over sammenbrudsspændinger for papirkondensatorer. Vi ønsker at undersøge, om det kan antages, at $X_i \in \mathrm{Min}_1(1500, 130)$, $i = 1, \cdots, 25$. Vi bestemmer udfaldet af den empiriske fordelingsfunktion

Værdi	Antal	Kum antal	Kum. ant i %
900	1	1	4
1080	1	2	8
1200	2	4	16
1380	4	8	32
1440	7	15	60
1500	4	19	76
1560	2	21	84
1620	3	24	96
1680	1	25	100

Af tabellen fås, at den kritiske afvigelse er $d(25)_{0.95} = 0.27$, hvis vi anvender niveauet $\alpha = 5\%$. Vi accepterer en hypotese $F = F_0$, hvis F_0's afvigelse fra det realiserede udfald af den empiriske fordelingsfunktion er mindre end 0.27. Hvis vi derfor i ethvert punkt afsætter det realiserede udfald af $F_n(x) \pm 0.27$, accepterer vi hypotesen, hvis F_0 overalt ligger mellem de derved fremkomne bånd. Det er altså med andre ord et **95% konfidensområde** for fordelingsfunktionen, vi får dannet på denne måde.

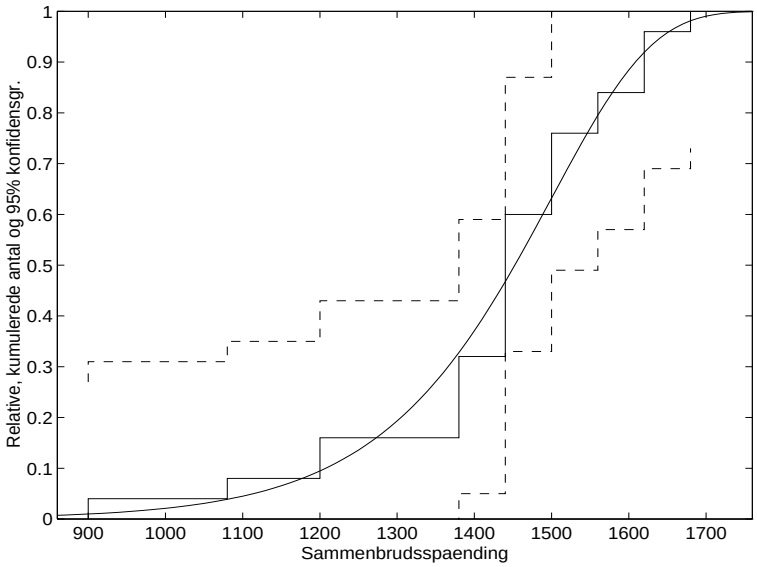
Vi har, at $X_i \in \mathrm{Min}_1(1500, 130)$ er ensbetydende med, at

$$\mathrm{F}_0(x) = 1 - \Lambda_1\left(-\frac{x - 1500}{130}\right)$$

ifølge definitionen p. 157. Vi anfører beregningen af nogle støttepunkter for F_0. Den beregnede F_0 er ligeledes indlagt i figuren.

x	$\frac{x-1500}{130}$	$1 - \Lambda_1\left(-\frac{x-1500}{130}\right)$
850	-5	0.01
980	-4	0.02
1110	-3	0.05
1240	-2	0.13
1370	-1	0.31
1500	0	0.63
1630	1	0.93
1760	2	1.00

Det ses, at F_0 intetsteds skærer grænserne i konfidensbåndet, således at vi kan acceptere hypotesen $H_0: \ X_i \in \mathrm{Min}_1(1500, 130)$ på et 5% niveau. ♦



4.2.3 Beregning af empiriske momenter

Vi slutter kapitlet med at give nogle generelle retningslinier for beregning af realisationer af empirisk middelværdi og varians, i.e. beregning af

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

og

$$s_*^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \text{eller} \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Det vil ofte være hensigtsmæssigt at indføre en **beregningsenhed** β og et **beregningsnulpunkt** α, d.v.s. vi sætter

$$t_i = \frac{x_i - \alpha}{\beta} \quad \Leftrightarrow \quad x_i = \beta t_i + \alpha. \tag{4.6}$$

Vi har da

$$\bar{x} = \frac{1}{n}\sum(\beta t_i + \alpha) = \beta\frac{1}{n}\sum t_i + \alpha = \beta\bar{t} + \alpha \tag{4.7}$$

og

$$\begin{aligned}\sum(x_i-\bar{x})^2 &= \sum(\beta t_i+\alpha-\beta\bar{t}-\alpha)^2\\ &= \beta^2\sum(t_i-\bar{t})^2. \end{aligned} \tag{4.8}$$

Her beregnes $\sum(t_i-\bar{t})^2$ som

$$\sum(t_i-\bar{t})^2=\sum t_i^2-\frac{1}{n}(\sum t_i)^2, \tag{4.9}$$

jfr. p. 235.

Ved hjælp af udtrykkene (4.6)–(4.9) kan man opnå væsentlige reduktioner i regnearbejdet ved bestemmelse af skøn over middelværdi og varians. Lad os eksempelvis betragte følgende data (eksempel 2.23, p. 262)

$$\begin{aligned} x_1 &= 58.59\\ &\quad 59.64\\ &\quad 58.45\\ &\cdot\quad 58.64\\ &\quad 58.00\\ &\cdot\quad 57.03\\ &\quad 57.33\\ &\cdot\quad 57.80\\ &\quad 58.04\\ x_{10} &= 58.41 \end{aligned}$$

Vi ønsker at beregne $\sum(x_i-\bar{x})^2$. Vi vælger $\alpha=58$ og $\beta=\frac{1}{100}$. Vi får da $t_i=(x_i-58)\cdot 100$, d.v.s.

i	t_i	t_i^2
1	59	3481
2	164	26896
3	45	2025
4	64	4096
5	0	0
6	−97	9409
7	−67	4489
8	−20	400
9	4	16
10	41	1681
I alt	193	52493

Heraf fås

$$\begin{aligned}\sum(t_i-\bar{t})^2 &= \sum t_i^2 - \frac{1}{10}(\sum t_i)^2 \\ &= 52493 - \frac{1}{10}37249 = 48768,\end{aligned}$$

d.v.s.

$$\sum(x_i-\bar{x})^2 = \left(\frac{1}{100}\right)^2 \cdot 48768 = 4.8768.$$

Hvis vi har grupperede data, opfatter vi blot materialet som om, alle elementer i en klasse var anbragt i klassemidtpunktet, og derefter anvender vi ovenstående formler. Her må vi dog være opmærksomme på, at en klasse med midtpunkt t_i og antal elementer a_i bidrager med

$$a_i t_i \qquad \text{og} \qquad a_i t_i^2$$

til den totale sum henholdsvis kvadratsum.

Kapitel 5

Varians- og regressionsanalyser

De metoder, vi skal gennemgå i dette kapitel, er nogle af de mest anvendte statistiske metoder i praksis. Variansanalyser består primært i at klassificere og krydsklassificere data og teste, om middelværdierne i specielle grupper afviger signifikant fra hinanden. Variansanalyserne anvendes især ved lidt mere indviklede (klassiske) **forsøgsplaner**. Emnet er meget stort, og vi vil kun give en introduktion til emnet og i øvrigt henvise til den i referencerne anførte litteratur. Regressionsanalysen beskæftiger sig med analyser af sammenhænge, som f.eks. at bestemme indflydelsen af forskellige produktionsfaktorer på det totale udbytte af en produktionsproces etc. Også disse metoder vil kun blive omtalt i deres mest simple form, og vi må igen henvise specielt interesserede til litteraturlisten.

5.1 Variansanalyser

5.1.1 Ensidet variansanalyse

Vi introducerer problemstillingen ved hjælp af

EKSEMPEL 5.1. På en virksomhed, hvor man fremstiller gødningsstoffer, har man målt det procentuelle indhold af kali K_2O i nogle stikprøver, man har taget fra 4 forskellige produktionsgrene. Man målte følgende data

Produktionsgren			
A	B	C	D
11.9	15.9	13.9	15.6
9.0	17.2	14.6	18.3
10.2	21.0	16.1	14.6
8.5	17.1	14.9	16.9
13.6	19.0	13.7	14.7
11.3	23.2	12.4	–

Som det fremgår, mislykkedes en af analyserne. Årsagen til, at man har taget ovenstående målinger, kunne være, at der har været fremført klager over for stærkt varierende indhold af kali i gødning stammende fra forskellige grene. Man ønsker derfor at få undersøgt variationen i kaliindholdet i ovenstående 4 stikprøver. ♦

For at kunne foretage en statistisk analyse af en situation som ovenstående må vi først formulere en matematisk model. Vi har givet k stikprøver (i eksemplet 4) med $n_1, \cdots, n_k$ (i eksemplet 6, 6, 6, 5) realiserede udfald af stokastiske variable

$$\begin{aligned} 1: &\quad X_{11}, \cdots, X_{1n_1} \\ 2: &\quad X_{21}, \cdots, X_{2n_2} \\ &\vdots \\ k: &\quad X_{k1}, \cdots, X_{kn_k}. \end{aligned}$$

Vi antager nu, at X'erne

1. er indbyrdes uafhængige
2. er normalt fordelte
3. har samme varians σ^2.

Vi forudsætter endvidere, at der er **homogenitet inden for grupper** , d.v.s. at variable i samme gruppe har samme middelværdi, i.e.

4. $\mathrm{E}(X_{ij}) = \mu_i$.

Det, man nu er interesseret i, er at teste, om de k gruppemiddelværdier er ens, i.e. teste

$$H_0: \quad \mu_1 = \cdots = \mu_k \qquad \text{mod} \qquad H_1: \quad \exists i, j(\mu_i \neq \mu_j).$$

Denne hypotese kaldes (meget naturligt) hypotesen om **fuldstændig homogenitet**. Det er sædvane at opspalte hver middelværdi μ_i i 2 størrelser μ og α_i, Vi har altså

$$\mathrm{E}(X_{ij}) = \mu_i = \mu + \alpha_i, \qquad j = 1, \cdots, n_i, \quad i = 1, \cdots, k,$$

hvor

$$n\mu = \sum_{i=1}^{k} n_i \mu_i \quad \text{og} \quad \sum_{i=1}^{k} n_i \alpha_i = 0,$$

Det ses umiddelbart, at disse relationer entydigt bestemmer μ og α_i. Her angiver μ altså et universelt niveau, og α_i angiver den i'te gruppes specielle effekt. Vi kan nu omformulere vor hypotese til

$$H_0 \quad : \qquad \alpha_1 = \cdots = \alpha_k = 0$$

mod

$$H_1 \quad : \qquad \exists i(\alpha_i \neq 0).$$

Det er en model som den her beskrevne, der benævnes en **ensidet variansanalyse-model**. Vi har nu

SÆTNING 5.1. Lad situationen være som beskrevet ovenfor. Da er kvotienttestet på niveau α for test af H_0 mod H_1 bestemt ved det kritiske område

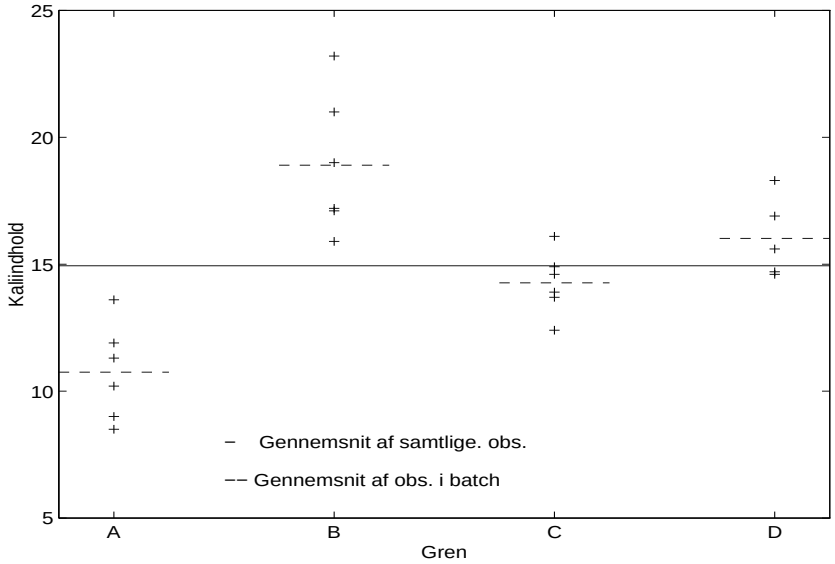
$$C = \left\{ (x_{11}, \cdots, x_{kn_k}) \middle| \frac{\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2/(k-1)}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2/(N-k)} > \mathrm{F}(k-1, N-k)_{1-\alpha} \right\},$$

hvor

$$\begin{aligned} N &= \sum_{i=1}^{k} n_i \\ \bar{X}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \\ \bar{X} &= \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij} = \frac{1}{N} \sum_{i=1}^{k} n_i \bar{X}_i. \end{aligned}$$

Bevis. Forbigås. Se f.eks. [48][p. 55]. ■

BEMÆRKNING 5.1. Det er muligt at give en heuristisk begrundelse for sætningen. Lad os betragte de i eksempel 5.1 anførte data. Vi afbilder dem i et koordinatsystem som anført nedenfor.



Vi ser nu, at tælleren i teststørrelsen, i.e.

$$\frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2$$

er summen (vægtet) af kvadraterne på de enkelte gruppemiddelværdiers afvigelse fra den totale middelværdi, d.v.s. den er et udtryk for **variationen mellem de enkelte grupper**. Hvis H_0 er sand, er den et centralt skøn over σ^2.

Nu er

$$\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2 = \frac{1}{\sum(n_i-1)}\big((n_1-1)S_1^2 + \cdots + (n_k-1)S_k^2\big),$$

hvor

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2$$

d.v.s. nævneren er et vejet gennemsnit af variansskønnene beregnet for hver gruppe. Ifølge reproduktionssætningen for χ^2-fordelingen (p. 195) er nævneren derfor et centralt skøn over σ^2 (jfr. p. 261). Dette skøn er beregnet på grundlag af **variationen inden for grupper**. Det er nu intuitivt klart, at et rimeligt test vil være at forkaste hypotesen H_0, at middelværdierne i grupperne er ens, hvis gruppegennemsnittene afviger "for meget" fra hinanden. Nu er det ikke klart, hvad "for meget" er; vi må finde en størrelse at måle afvigelsen relativt til, og det er da rimeligt at vælge et skøn over variansen beregnet på grundlag af variansskøn fra de enkelte grupper. Det virker derfor rimeligt, at det kritiske område har den anførte form. ▼

Det er mindre indlysende, at teststørrelsen er $\mathrm{F}(k-1, N-k)$-fordelt. Dette følger af

SÆTNING 5.2. Der gælder

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_i)^2 + \sum_{i=1}^{k} n_i(\bar{X}_i-\bar{X})^2 .$$

Under H_0 gælder endvidere, at de 2 kvadratsummer på højresiden er stokastisk uafhængige og $\sigma^2\chi^2$-fordelte med $N-k$ henholdsvis $k-1$ frihedsgrader.

Bevis. Spaltningen af kvadratsummen på venstresiden fås ved direkte regning:

$$\begin{aligned}
\sum_i\sum_j (X_{ij}-\bar{X})^2 &= \sum_i\sum_j (X_{ij}-\bar{X}_i+\bar{X}_i-\bar{X})^2 \\
&= \sum_i\sum_j (X_{ij}-\bar{X}_i)^2 + \sum_i\sum_j (\bar{X}_i-\bar{X})^2 \\
&\quad + 2\sum_i\sum_j (X_{ij}-\bar{X}_i)(\bar{X}_i-\bar{X}) \\
&= \sum_i\sum_j (X_{ij}-\bar{X}_i)^2 + \sum_i n_i(\bar{X}_i-\bar{X})^2 ,
\end{aligned}$$

idet

$$\begin{aligned}
\sum_i\sum_j (X_{ij}-\bar{X}_i)(\bar{X}_i-\bar{X}) &= \sum_i (X_i-\bar{X})\sum_j (X_{ij}-\bar{X}_i) \\
&= \sum_i (\bar{X}_i-\bar{X})\cdot 0 = 0
\end{aligned}$$

ifølge definitionen på $\bar{X}_i$.

Resultaterne vedrørende uafhængigheden og fordelingen af kvadratsummerne følger af Sætning 1.77, side 195 (se f.eks. [34] eller [48, p. 419 eller p. 127]). ■

BEMÆRKNING 5.2. Vi ser, at vi har fået spaltet observationernes variation om det "totale" middel op i observationernes variation om gruppemiddeltallene plus gruppemiddeltallenes variation om det totale middel. Vort test er baseret på disse komponenters indbyrdes størrelsesforhold, d.v.s. det er en analyse af variationerne omkring de forskellige middelværdier. Heraf kommer navnet **variansanalyse**. ▼

En umiddelbar følge af sætningen og definition 1.45, p. 201, er følgende

KOROLLAR 5.1. Under hypotesen om fuldstændig homogenitet er

$$Z = \frac{\sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2/(N-k)} \in \mathrm{F}(k-1, N-k).$$

Under den numeriske behandling kan det være hensigtsmæssigt at anvende følgende regneskema:

i	Observationer	n_i	S_i	SK_i	S_i^2/n_i	SAK_{1i}	f_i
1 $\vdots$ i $\vdots$ k	$X_{i1}, \cdots, X_{in_i}$	n_i	$\sum_{j=1}^{n_i} X_{ij}$	$\sum_{j=1}^{n_i} X_{ij}^2$		$\mathrm{SK}_i - \frac{S_i^2}{n_i}$	$n_i - 1$
Total		$N = \sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} S_i$	$\sum_{i=1}^{k} \mathrm{SK}_i$	$\sum_{i=1}^{k} \frac{S_i^2}{n_i}$	$\sum_{i=1}^{k} \mathrm{SAK}_{1i}$	$F = \sum_{i=1}^{k} f_i$

Vi beregner nu let kvadratafvigelsessummerne:

$$\begin{aligned}
\mathrm{SAK}_0 &= \sum_i \sum_j (X_{ij} - \bar{X})^2 = \sum_i \sum_j X_{ij}^2 - N\bar{X}^2 \\
&= \sum_i \mathrm{SK}_i - (\sum_i S_i)^2/N \\
\mathrm{SAK}_1 &= \sum_i \sum_j (X_{ij} - \bar{X}_i)^2 = \sum_i \mathrm{SAK}_{1i} \\
\mathrm{SAK}_2 &= \sum_i n_i(\bar{X}_i - \bar{X})^2 = \sum_i n_i \bar{X}_i^2 - N\bar{X}^2 \\
&= \sum_i \frac{S_i^2}{n_i} - (\sum_i S_i)^2/N.
\end{aligned}$$

Resultaterne samles sædvanligvis i et såkaldt **variansanalyseskema**:

Variation	SAK	f	S^2	Teststørrelse
Mellem grupper	SAK_2	$k-1$	$\mathrm{SAK}_2/(k-1)$	$Z = \frac{\mathrm{SAK}_2/(k-1)}{\mathrm{SAK}_1/(N-k)}$
Inden for grupper	SAK_1	$N-k$	$\mathrm{SAK}_1/(N-k)$	
Total	SAK_0	$N-1$		

Ved test på niveau α er det kritiske område

$$C = \{(x_{11}, \cdots, x_{kn_k}) | z > \mathrm{F}(k-1, N-k)_{1-\alpha}\}\,.$$

Hvis H_0 accepteres, kan vi som skøn over variansen σ^2 anvende

$$\frac{1}{N-1}\mathrm{SAK}_0 = \frac{1}{N-1}\sum_i \sum_j (X_{ij} - \bar{X})^2.$$

Hvis H_0 ikke accepteres, anvendes

$$\frac{1}{N-k}\mathrm{SAK}_1 = \frac{1}{N-k}\sum_i \sum_j (X_{ij} - \bar{X}_i)^2$$

som skøn over σ^2.

BEMÆRKNING 5.3. Formålet med en ensidet variansanalyse er som nævnt at forklare den totale variation SAK_0 ved at spalte den i et bidrag, der skyldes variationen mellem grupper, (SAK_2), og et bidrag, der skyldes variationen indenfor grupper, (SAK_1). Såfremt hypotesen H_0 accepteres, er der ikke nogen signifikant forskel mellem grupperne, og den totale variation kan ikke forklares ved andet end tilfældige målefejl. Omvendt, hvis H_0 afvises, så er der en signifikant forskel mellem grupperne, som forklares af forskellen mellem grupperne udtrykt gennem SAK_2. Tilbage er blot de tilfældige målefejl, som bidrager med SAK_1 til den totale variation SAK_0. Dette svarer til, at estimatet på σ^2, som er variansen på den tilfældige målefejl, afhænger af testets udfald. ▼

Vi vil nu bestemme maximum likelihood skøn over μ og α_i, $i = 1, \cdots, k$. Ifølge sætning 2.11 p. 253, kan disse findes ved hjælp af mindste kvadraters metode. Vi skal vælge μ og α_i således, at

$$\mathrm{f}(\mu, \alpha_1, \cdots, \alpha_k) = \sum_i \sum_j (x_{ij} - \mu - \alpha_i)^2$$

minimaliseres. Vi har

$$\begin{aligned} \frac{\partial f}{\partial \mu} &= -2 \sum_i \sum_j (x_{ij} - \mu - \alpha_i) \\ &= -2\Big[\sum_i \sum_j x_{ij} - N\mu - \sum_i n_i \cdot \alpha_i\Big] \\ &= -2\Big[\sum_i \sum_j x_{ij} - N\mu\Big], \end{aligned}$$

d.v.s.

$$\frac{\partial f}{\partial \mu} = 0 \quad \Leftrightarrow \quad \mu = \bar{x}.$$

Endvidere er

$$\begin{aligned} \frac{\partial f}{\partial \alpha_i} &= -2 \sum_j (x_{ij} - \mu - \alpha_i) \\ &= -2 n_i (\bar{x}_i - \mu - \alpha_i). \end{aligned}$$

Heraf fås

$$\frac{\partial f}{\partial \alpha_i} = 0 \quad \Leftrightarrow \quad \alpha_i = \bar{x}_i - \mu.$$

Ved en sammenfatning af disse relationer fås

$$\begin{aligned} \hat{\mu} &= \bar{X} = \frac{1}{N}\sum_i\sum_j X_{ij} \\ \hat{\alpha}_i &= \bar{X}_i - \bar{X} = \frac{1}{n_i}\sum_j X_{ij} - \frac{1}{N}\sum_\nu\sum_j X_{\nu j}. \end{aligned}$$

Eksempel 5.1 fortsættes i

EKSEMPEL 5.2. Vi er nu i stand til at gennemføre et test for, om kaliindholdet fra de 4 grene kan antages at være ens. Vi forudsætter, at vi har indbyrdes uafhængige stokastiske variable, der er normalt fordelte med samme varians. Endvidere forudsætter vi, at der er homogenitet inden for grupper. Vi skal ikke i dette eksempel komme ind på undersøgelsen af rimeligheden af disse forudsætninger, men blot konstatere, at en sådan undersøgelse ikke må mangle ved løsningen af et praktisk problem. Idet vi anvender de ovenfor anførte betegnelser, vil vi nu teste

$$H_0: \quad \alpha_1 = \cdots = \alpha_k = 0 \qquad \text{mod} \qquad H_1: \quad \exists i(\alpha_i \neq 0).$$

Vi har en ensidet variansanalysemodel. Beregningsskemaet bliver, idet vi udelader observationssøjlen:

i	n_i	S_i	SK_i	S_i^2/n_i	SAK_{1i}	f_i
1	6	64.5	711.55	693.375	18.175	5
2	6	113.4	2181.30	2143.260	38.040	5
3	6	85.6	1229.04	1221.227	7.813	5
4	5	80.1	1293.11	1283.202	9.908	4
Total	23	343.6	5415.00	5341.064	73.936	19

Heraf fås:

$$\begin{aligned} \mathrm{SAK}_0 &= 5415.00 - 5133.085 = 281.915 \\ \mathrm{SAK}_1 &= 73.936 \\ \mathrm{SAK}_2 &= 5341.064 - 5133.085 = 207.979. \end{aligned}$$

Variansanalyseskemaet bliver

Variation	SAK	f	s^2	Test
Mellem grene	207.979	3	69.326	17.817
Inden for grene	73.936	19	3.891	
Total	281.915	22		

Nu er $\mathrm{F}(3,19)_{99.95\%} = 9.42$, d.v.s. vi vil forkaste hypotesen, i hvert fald på alle niveauer større end 0.0005. Vi vil derfor konkludere, at der er et varierende indhold af kali i gødning fra de forskellige grene.

Vi har derfor, at kaliindholdet i en gødning fra batch nr. i beskrives ved

$$X_{ij} \in \mathrm{N}(\mu + \alpha_i, \sigma^2),$$

hvor

$$\hat{\sigma}^2 = \frac{\mathrm{sak}_1}{19} = 3.891 = 1.97^2$$

og

$$\begin{array}{lcrcr} \hat{\mu} & = & \bar{x} & = & 14.94 \\ \hat{\alpha}_1 & = & \bar{x}_1 - \bar{x} & = & -4.19 \\ \hat{\alpha}_2 & = & \bar{x}_2 - \bar{x} & = & 3.96 \\ \hat{\alpha}_3 & = & \bar{x}_3 - \bar{x} & = & -0.67 \\ \hat{\alpha}_4 & = & \bar{x}_4 - \bar{x} & = & 1.08 \end{array}$$

Vi kan nu formulere vor konklusion således, at kaliindholdet i gødningerne fordeler sig om en middelværdi, der er estimeret til 14.94%, **plus** et bidrag, der afhænger af produktionsenhed. Variansen af udfaldene er estimeret til $3.891 = 1.97^2$, dvs. en spredning på 1.97. ♦

5.1.2 Tosidet variansanalyse

Den foregående variansanalyse var relativt simpel, fordi der kun var en klassifikationsvariabel, nemlig batch nr. Man kunne forestille sig, at den valgte analysemetode måske ikke var den mest velegnede til denne form for analyse, således at man ville analysere de forskellige batches på flere forskellige metoder. Dette ville føre til en **tosidet variansanalysemodel**. Vi vil starte gennemgangen af denne med et lille eksempel.

På en forsøgsstation skal man vurdere, hvilken af 5 gødningstyper man skal foretrække ved gødskning af kartofler. Til rådighed for forsøget var 4 marker. Nu må man antage, at disse marker ikke er ens, hvad angår jordbundsforhold, lysforhold m.v. Derfor besluttedes det at inddele hver mark i 5 lige store stykker og knytte hver af de 5 gødningstyper tilfældigt til en af disse "delmarker". Der blev avlet kartofler på alle stykkerne, og disse blev gødsket i overensstemmelse med de foregående valg. Ved voksesæsonens slutning tog man kartoflerne op og målte udbyttet på hvert af de 20 stykker.

Herved opnåedes følgende data (målt i kg):

	Gødning				
Mark	A	B	C	D	E
1	310	353	366	299	367
2	284	293	335	264	314
3	307	306	339	311	377
4	267	308	312	266	342

Problemet er nu at teste, om de 5 gødninger er lige effektive med hensyn til middeludbytte af kartofler. Vi prøver at formulere en matematisk model, der dækker ovenstående situation.

Vi antager, at vore observationer er realiserede udfald af stokastiske variable

$$\begin{matrix} X_{11} & , \cdots , & X_{1m} \\ \vdots & & \vdots \\ X_{k1} & , \cdots , & X_{km} \end{matrix}$$

Disse antages at

1. være stokastisk uafhængige
2. være normalt fordelte
3. have samme varians σ^2.

Vi forudsætter endvidere, at der er en **additiv** struktur i middelværdierne, i.e. at

4. $\mathrm{E}(X_{ij}) = \mu_{ij} = \mu + \alpha_i + \beta_j$.

I 4. fastlægges α_i og β_j ved relationerne

$$\begin{aligned} \sum_{i=1}^{k} \alpha_i &= 0 \\ \sum_{j=1}^{m} \beta_j &= 0. \end{aligned}$$

Her angiver μ altså et niveau, α_i effekten af den i'te række og β_j effekten af den j'te søjle. Additivitetsforudsætningen 4. siger med andre ord, at middelværdien af den (i,j)'te variabel er lig niveauet plus den i'te rækkeeffekt plus den j'te søjleeffekt.

I eksemplet svarer 4. til at sige, at middeludbyttet på mark nr. i med anvendelse af gødning nr. j er lig et niveau plus et bidrag, der skyldes marken, plus et bidrag, der skyldes gødningen.

Vi er nu interesserede i at teste

$$H_{01}: \quad \alpha_1 = \cdots = \alpha_k = 0 \qquad \text{mod} \qquad H_{11}: \quad \exists i(\alpha_i \neq 0)$$

og

$$H_{02}: \quad \beta_1 = \cdots = \beta_m = 0 \qquad \text{mod} \qquad H_{12}: \quad \exists j(\beta_j \neq 0)$$

Før vi går videre med testproceduren, vender vi tilbage til additivitetsforudsætningen, som vi søger at belyse i

EKSEMPEL 5.3. For at få et indblik i, hvad 4. og hypoteserne H_{01} og H_{02} indebærer, betragter vi følgende situation. Lad α_i, $i = 1,2,3,4$, og β_j, $j = 1,2,3$, være givet ved skemaerne

i	1	2	3	4
α_i	4	-2	1	-3

j	1	2	3
β_j	1	2	-3

Idet vi sætter $\mu = 5$, definerer vi dernæst

$$\begin{aligned} \mu_{ij} &= \mu + \alpha_i + \beta_j \quad i = 1,2,3,4, \quad j = 1,2,3 \\ \mu_{i\cdot} &= \mu + \alpha_i \qquad\qquad i = 1,2,3,4 \\ \mu_{\cdot j} &= \mu + \beta_j \qquad\qquad j = 1,2,3 \end{aligned}$$

samt

$$\xi_{ij} \quad = \quad \mu + \alpha_i + \beta_j + \alpha_i \cdot \beta_j \quad i = 1,2,3,4, \quad j = 1,2,3.$$

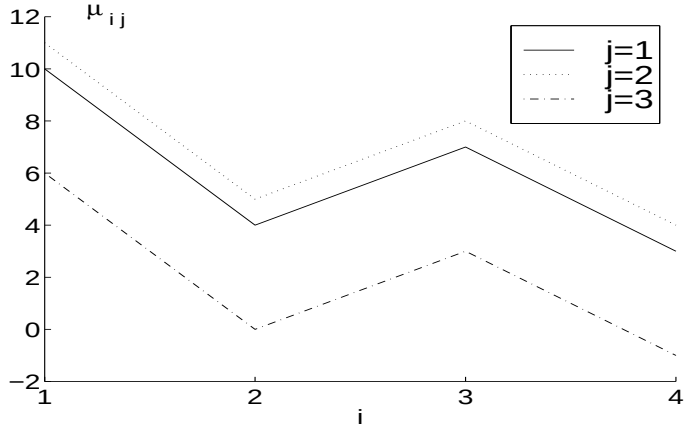
Vi danner nu let tabeller over μ_{ij}, $\mu_{i\cdot}$ og $\mu_{\cdot j}$:

	μ_{ij}	j 1	2	3
i	1	10	11	6
	2	4	5	0
	3	7	8	3
	4	3	4	−1

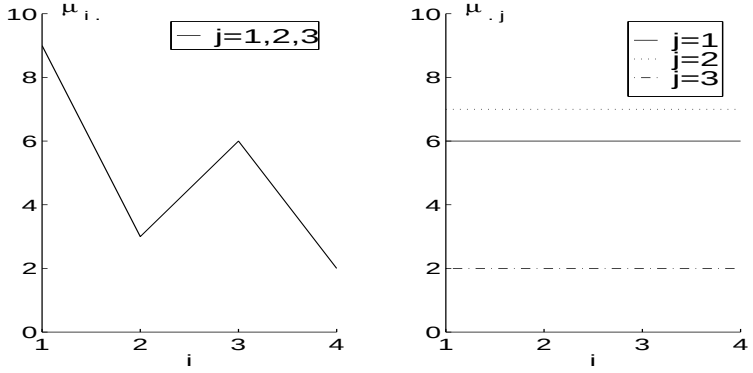
i	1	2	3	4
$\mu_{i\cdot}$	9	3	6	2

j	1	2	3
$\mu_{\cdot j}$	6	7	2

Hvis vi afbilder μ_{ij} som funktion af i, får vi følgende graf Til sammenligning anføres

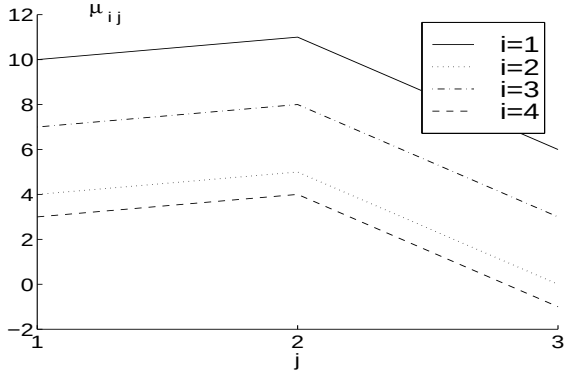


graferne for $\mu_{i\cdot}$ og $\mu_{\cdot j}$ som funktion af i

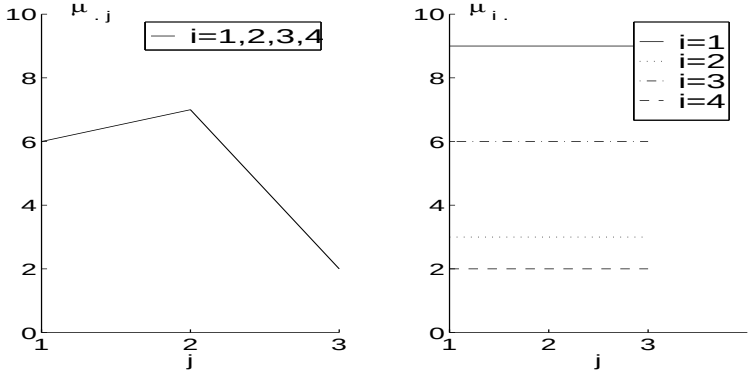


Vi bemærker, at linierne i alle grafer er parallelle.

Den analoge figur med μ_{ij} opfattet som funktion af j er



og de tilsvarende for $\mu_{i\cdot}$ og $\mu_{\cdot j}$:

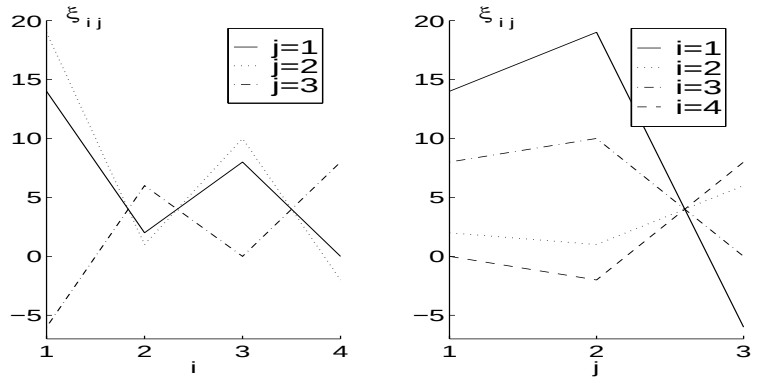


For at finde ud af, hvad der er særegent for den additive struktur, anfører vi også graferne for ξ_{ij}, der jo også indeholder et produktled.

Tabellen over ξ_{ij} bliver

		j		
	ξ_{ij}	1	2	3
i	1	14	19	−6
	2	2	1	6
	3	8	10	0
	4	0	−2	8

Graferne for ξ_{ij} opfattet som funktion af i henholdsvis j bliver



Vi bemærker nu, at forskellen mellem grafen for f.eks. μ_{1j} og μ_{2j} er konstant ($= \alpha_1 - \alpha_2 = 6$), og at kurverne derfor løber "parallelt" uden at skære hinanden. Derimod ses det, at graferne for e.g. ξ_{1j} og ξ_{2j} skærer hinanden, d.v.s. forskellen mellem ξ_{1j} og ξ_{2j} er **ikke** uafhængig af j. Dette udtrykker vi også ved at sige, at der er en **vekselvirkningseffekt**. ♦

Vi vender nu tilbage til spørgsmålet om testning af H_{01} og H_{02}. Som i den ensidede variansanalyse vil vi basere testene på sammenligninger af skøn over σ^2 - ét baseret på variationen mellem rækker, ét på variationen mellem søjler og ét, der udtrykker vekselvirkningsvariationen. Vi vil søge heuristisk at finde frem til teststørrelserne. Vi nævner først

SÆTNING 5.3. Maximum likelihood estimatorerne for parametrene er

$$\begin{aligned} \hat{\mu} &= \bar{X} \\ \hat{\alpha}_i &= \bar{X}_{i\cdot} - \bar{X} \qquad i = 1, \cdots, k \\ \hat{\beta}_j &= \bar{X}_{\cdot j} - \bar{X} \qquad j = 1, \cdots, m, \end{aligned}$$

hvor

$$\begin{aligned} \bar{X} &= \frac{1}{km} \sum_{i=1}^{k} \sum_{j=1}^{m} X_{ij} \\ \bar{X}_{i\cdot} &= \frac{1}{m} \sum_{j=1}^{m} X_{ij} \\ \bar{X}_{\cdot j} &= \frac{1}{k} \sum_{i=1}^{k} X_{ij}. \end{aligned}$$

Bevis. Forbigås. Forløber ganske analogt til de p. 420 anførte beregninger. ■

Vi anfører nu en til sætning 5.2 analog

SÆTNING 5.4. Der gælder

$$\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij}-\bar{X})^2 = m\sum_{i=1}^{k}(\bar{X}_{i\cdot}-\bar{X})^2 + k\sum_{j=1}^{m}(\bar{X}_{\cdot j}-\bar{X})^2 + \sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})^2.$$

Under H_{01} og H_{02} gælder endvidere, at de 3 kvadratsummer på højresiden er stokastisk uafhængige og $\sigma^2\chi^2$-fordelte med $k-1$, $m-1$ henholdsvis $(k-1)(m-1)$ frihedsgrader.

Bevis. Spaltningen af kvadratsummen på venstresiden følger ved direkte regning:

$$\begin{aligned}\sum_i\sum_j(X_{ij}-\bar{X})^2 &= \sum_i\sum_j\big(X_{ij}-(\bar{X}_{i\cdot}-\bar{X})-(\bar{X}_{\cdot j}-\bar{X})-\bar{X}\\ &\quad+(\bar{X}_{i\cdot}-\bar{X})+(\bar{X}_{\cdot j}-\bar{X})\big)^2\\ &= \sum_i\sum_j(X_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})^2+k\sum_j(\bar{X}_{\cdot j}-\bar{X})^2\\ &\quad+m\sum_i(\bar{X}_{i\cdot}-\bar{X})^2+2\sum_j(\bar{X}_{\cdot j}-\bar{X})\sum_i(X_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})\\ &\quad+2\sum_i(\bar{X}_{i\cdot}-\bar{X})\sum_j(X_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})\\ &\quad+2\sum_i(\bar{X}_{i\cdot}-\bar{X})\sum_j(\bar{X}_{\cdot j}-\bar{X}).\end{aligned}$$

Som ved beviset for sætning 5.2 ses, at de dobbelte produkter bliver 0, og heraf fremgår spaltningen.

Resten af sætningen følger af Sætning 1.77, side 195, og forbigås. ■

BEMÆRKNING 5.4. Vi skal nu kort kommentere, hvorledes man kan komme frem til de anførte frihedsgradsantal. Ifølge sætning 2.6, p. 226 har de 2 første kvadratsummer

$k-1$ henholdsvis $m-1$ frihedsgrader. Vi betragter dernæst de stokastiske variable

$$X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X} \qquad i = 1, \cdots, k \quad j = 1, \cdots, m.$$

Vi vil bestemme antallet af (uafhængige) lineære bånd imellem disse (jfr. p. 227). Det er ifølge definitionen af $\bar{X}$, $\bar{X}_{i\cdot}$ og $\bar{X}_{\cdot j}$ klart, at

$$\sum_{i=1}^{k}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}) = 0 \qquad j = 1, 2, \cdots, m \tag{5.1}$$

og

$$\sum_{j=1}^{m}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}) = 0 \qquad i = 1, \cdots, k \tag{5.2}$$

Da f.eks. (5.1) medfører, at også

$$\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}) = 0,$$

ses det, at der ved (5.2) kun defineres yderligere $k-1$ lineære bånd, således at antallet af (uafhængige) lineære bånd defineret ved (5.1) og (5.2) er $k+m-1$. I bemærkningen p. 226 anførte vi, at antallet af frihedsgrader for kvadratsummen var lig antallet af led minus antallet af lineære bånd. Anvender vi denne regel, fås

$$km - (k + m - 1) = (k - 1)(m - 1),$$

hvilket netop er det i sætningen anførte. ▼

Da nu f.eks.

$$k\sum_{j}(\bar{X}_{\cdot j} - \bar{X})^2 = k\sum_{j}\hat{\beta}_j^2, \tag{5.3}$$

vil vi forvente, at kvadratsummen (5.3) er "stor", hvis nogle β_j'er er forskellige fra 0.

Her skal vendingen "stor" ses relativt til kvadratsummen

$$\sum\sum(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2.$$

Som ved corollaret p. 418 ses, at vi skal sammenligne med fraktiler i en F-fordeling. Analogt kan vi opstille et test for H_{01}. Vi kan samle disse overvejelser i følgende **variansanalyseskema**.

Variation	SAK	f	Test
Mellem søjler	$\mathrm{SAK}_4 = k\sum_{j=1}^{m}(\bar{X}_{\cdot j} - \bar{X})^2$	$m-1$	$\frac{\mathrm{SAK}_4/f_4}{\mathrm{SAK}_2/f_2}$
Mellem rækker	$\mathrm{SAK}_3 = m\sum_{i=1}^{k}(\bar{X}_{i\cdot} - \bar{X})^2$	$k-1$	$\frac{\mathrm{SAK}_3/f_3}{\mathrm{SAK}_2/f_2}$
Veksel-virkning	SAK_2 (se nedenfor)	$(m-1)(k-1)$	
I alt	$\mathrm{SAK}_0 = \sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X})^2$	$km-1$	

hvor SAK_2 er givet ved

$$\mathrm{SAK}_2 = \sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2$$

Nummeringen af SAK'erne får sin forklaring i det følgende.

SÆTNING 5.5. De kritiske områder ved tests på niveau α er

H_{01} mod H_{11} :

$$C_1 = \left\{(x_{11},\cdots,x_{km}) \middle| \frac{\mathrm{sak}_3/f_3}{\mathrm{sak}_2/f_2} > \mathrm{F}\big(k-1,(k-1)(m-1)\big)_{1-\alpha}\right\}$$

H_{02} mod H_{12} :

$$C_2 = \left\{(x_{11},\cdots,x_{km}) \middle| \frac{\mathrm{sak}_4/f_4}{\mathrm{sak}_2/f_2} > \mathrm{F}\big(m-1,(k-1)(m-1)\big)_{1-\alpha}\right\}.$$

Bevis. Umiddelbar følge af de foregående overvejelser. ■

Hvis ingen af hypoteserne accepteres, anvendes

$$\frac{\mathrm{SAK}_2}{(m-1)(k-1)} = \frac{1}{(m-1)(k-1)} \sum_i \sum_j (X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2$$

som skøn over σ^2. Accepteres f.eks. H_{01}, estimeres σ^2 ved

$$\frac{1}{f_2 + f_3}(\mathrm{SAK}_2 + \mathrm{SAK}_3),$$

som er en central estimator for σ^2. Accepteres begge hypoteser, anvendes estimatoren

$$\frac{1}{km-1}\mathrm{SAK}_0.$$

Af hensyn til en hensigtsmæssig numerisk behandling anføres følgende relationer

$$\begin{aligned}
\mathrm{SAK}_4 &= k\sum_j (\bar{X}_{\cdot j} - \bar{X})^2 = \frac{1}{k}\sum_j (\sum_i X_{ij})^2 - \frac{1}{km}(\sum_{i,j} X_{ij})^2 \\
\mathrm{SAK}_3 &= m\sum_i (\bar{X}_{i\cdot} - \bar{X})^2 = \frac{1}{m}\sum_i (\sum_j X_{ij})^2 - \frac{1}{km}(\sum_{i,j} X_{ij})^2 \\
\mathrm{SAK}_0 &= \sum_{i,j} (X_{ij} - \bar{X})^2 = \sum_{i,j} X_{ij}^2 - \frac{1}{km}(\sum_{i,j} X_{ij})^2 \\
\mathrm{SAK}_2 &= \mathrm{SAK}_0 - \mathrm{SAK}_3 - \mathrm{SAK}_4.
\end{aligned}$$

Ovenstående identiteter følger ved direkte regning.

Vi er nu i stand til at behandle det i indledningen til afsnittet omtalte eksempel.

EKSEMPEL 5.4. Vi antager, at de p. 423 anførte data kan beskrives ved en tosidet variansanalysemodel som anført p. 423. At undersøge, om de 5 gødninger er lige effektive med hensyn til middeludbyttet af kartofler, svarer til at undersøge, om søjleeffekterne er lig 0, i.e. om vi kan antage

$$H_{02}: \quad \beta_1 = \cdots = \beta_5 = 0.$$

Med henblik på at udføre det i sætning 5.5 anførte test beregner vi følgende summer og kvadratsummer

$$\begin{aligned}
\sum_i \sum_j x_{ij} &= 6320, & \sum_i \sum_j x_{ij}^2 &= 2018650 \\
\sum_j (\sum_i x_{ij})^2 &= 8039328, & \sum_i (\sum_j x_{ij})^2 &= 10017750,
\end{aligned}$$

og ved hjælp af ovenstående beregningsformler fås

$$\begin{aligned}
\mathrm{sak}_4 &= \tfrac{1}{4}8039328 - \tfrac{1}{20}6320^2 &&= 12712\\
\mathrm{sak}_3 &= \tfrac{1}{5}10017750 - \tfrac{1}{20}6320^2 &&= 6430\\
\mathrm{sak}_0 &= 2018650 - \tfrac{1}{20}6320^2 &&= 21530\\
\mathrm{sak}_2 &= 21530 - 12712 - 6430 &&= 2388.
\end{aligned}$$

Vi samler resultaterne i **variansanalyseskemaet**:

Variation	SAK	f	Test
Mellem gødninger	12712	4	16.0
Mellem marker	6430	3	10.8
Vekselvirkning	2388	12	
Total	21530	19	

Da $\mathrm{F}(4,12)_{0.999} = 9.63 < 16.0$, vil det ikke være rimeligt at antage, at de 5 gødninger er ens. Vi anfører estimaterne for gødningseffekterne $\beta_1, \cdots, \beta_5$:

i	1	2	3	4	5
$\hat{\beta}_i$	-24	-1	22	-31	34

Som skøn over σ^2 anvendes

$$\frac{1}{f_2}\mathrm{sak}_2 = \frac{1}{12}2388 = 199 \simeq 14.1^2.$$

Ved at se på værdierne af $\hat{\beta}_1, \cdots, \hat{\beta}_5$ er det klart, at gødningerne C og E er at foretrække fremfor de øvrige. Hvis forskellen ikke havde så udpræget som her, kunne man sammenligne disse værdier ved hjælp af nogle t-tests.

Vi bemærker til sidst, at vi heller ikke kan antage, at markerne er ens, idet

$$\mathrm{F}(3,12)_{0.95} = 3.49 < 10.8.$$

♦

Den hidtidige analyse har været baseret på antagelsen om en additiv struktur i middelværdien, nemlig

$$\mathrm{E}(X_{ij}) = \mu_{ij} = \mu + \alpha_i + \beta_j \,. \tag{5.4}$$

Hvis vi ikke mener, at vi direkte kan tillade os at postulere denne spaltning, men ønsker at teste den, må vi have flere observationer i hver celle for at kunne danne et variansskøn baseret på variationen inden for celler.

Vi betragter altså stokastiske variable

$$X_{ij\nu} \qquad i = 1, \cdots, k, \quad j = 1, \cdots, m, \quad \nu = 1, \cdots, n,$$

eller skrevet op i et skema:

	$j=1$		$j=m$
$i=1$	X_{111} $\vdots$ X_{11n}	$\cdots$	X_{1m1} $\vdots$ X_{1mn}
$\vdots$	$\vdots$		$\vdots$
$i=k$	X_{k11} $\vdots$ X_{k1n}	$\cdots$	X_{km1} $\vdots$ X_{kmn}

Vi antager, at disse stokastiske variable

1. er stokastisk uafhængige
2. er normalt fordelte
3. har samme varians σ^2
4. har middelværdier $\mathrm{E}(X_{ij\nu}) = \mu_{ij}$.

Her udtrykker betingelsen 4., at der er homogenitet inden for celler.

Vi spalter nu middelværdien μ_{ij}, idet vi sætter

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \tau_{ij} \tag{5.5}$$

hvor α_i, β_j og τ_{ij} fastlægges ved

$$\begin{aligned}
\sum_{i=1}^{k} \alpha_i &= 0 \\
\sum_{j=1}^{m} \beta_j &= 0 \\
\sum_{i=1}^{k} \tau_{ij} &= 0 \qquad \forall j \in \{1, \cdots, m\} \\
\sum_{j=1}^{m} \tau_{ij} &= 0 \qquad \forall i \in \{1, \cdots, k\}.
\end{aligned}$$

Estimatorerne for modellens parametre anføres i

SÆTNING 5.6. Maximum likelihood estimatorerne for parametrene er

$$\begin{array}{lcll}
\hat{\mu} & = & \bar{X} & \\
\hat{\alpha}_i & = & \bar{X}_{i\cdot} - \bar{X} & i = 1, \ldots, k \\
\hat{\beta}_j & = & \bar{X}_{\cdot j} - \bar{X} & j = 1, \ldots, m \\
\hat{\tau}_{ij} & = & \bar{X}_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X} & i = 1, \ldots, k,\ j = 1, \ldots, m
\end{array}$$

hvor

$$\begin{aligned}
\bar{X} &= \frac{1}{kmn} \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{\nu=1}^{n} X_{ij\nu} \\
\bar{X}_{i\cdot} &= \frac{1}{mn} \sum_{j=1}^{m} \sum_{\nu=1}^{n} X_{ij\nu} \\
\bar{X}_{\cdot j} &= \frac{1}{kn} \sum_{i=1}^{k} \sum_{\nu=1}^{n} X_{ij\nu} \\
\bar{X}_{ij} &= \frac{1}{n} \sum_{\nu=1}^{n} X_{ij\nu}
\end{aligned}$$

Bevis. Forbigås. Forløber ganske analogt til de p. 420 anførte beregninger. ■

Man er nu interesseret i at teste, om spaltningen (5.5) har formen (5.4), d.v.s. at teste

$$H_0: \quad \forall i,j(\tau_{ij}=0) \qquad \text{mod} \qquad H_1: \quad \exists i,j(\tau_{ij}\neq 0).$$

Som tidligere nævnt, vil vi basere testet på et skøn over σ^2, der ikke afhænger af hypotesen H_0 om forsvindende vekselvirkning. Et sådant skøn er

$$\frac{1}{km(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{\nu=1}^{n}(X_{ij\nu}-\bar{X}_{ij})^2,$$

hvor

$$\bar{X}_{ij}=\frac{1}{n}\sum_{\nu=1}^{n}X_{ij\nu}.$$

Dette skøn er centralt og $\sigma^2\chi^2(f)/f$-fordelt med $km(n-1)$ frihedsgrader. Dette kan vises som det analoge resultat p. 416. Vi kan derfor nu teste hypotesen $\tau_{ij}=0$ ved at sammenligne denne variation inden for grupper med vekselvirkningsvariationen. Inden vi præciserer resultatet, nævner vi, at sætning 5.4 nu antager formen

SÆTNING 5.7. Spaltningen af den totale variation er

$$\begin{aligned}\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{\nu=1}^{n}(X_{ij\nu}-\bar{X})^2 &= \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{\nu=1}^{n}(X_{ij\nu}-\bar{X}_{ij})^2\\ &+ n\sum_{i=1}^{k}\sum_{j=1}^{m}(\bar{X}_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})^2\\ &+ mn\sum_{i=1}^{k}(\bar{X}_{i\cdot}-\bar{X})^2\\ &+ kn\sum_{j=1}^{m}(\bar{X}_{\cdot j}-\bar{X})^2,\end{aligned}$$

d.v.s. den er spaltet i variationen inden for grupper plus vekselvirkningsvariationen plus variationen mellem rækker plus variationen mellem søjler.

Der gælder ydermere, at de 4 kvadratafvigelsessummer på højresiden er stokastisk uafhængige og $\sigma^2\chi^2$-fordelte med $km(n-1)$, $(k-1)(m-1)$, $k-1$ henholdsvis $m-1$ frihedsgrader under H_0, H_{01} og H_{02}, jvnf. Sætning 1.77, side 1.77.

Bevis. Forbigås. Forløber analogt til beviset for sætning 5.2. ■

Svarende til denne sætning har vi nu **variansanalyseskemaet**

Vari-ation	SAK	f	Test
Mellem søjler	$\mathrm{SAK}_4 = kn\sum_j(\bar{X}_{\cdot j} - \bar{X})^2$	$m-1$	$\frac{\mathrm{SAK}_4/f_4}{\mathrm{SAK}_{02}/f_{02}}$
Mellem rækker	$\mathrm{SAK}_3 = mn\sum_i(\bar{X}_{i\cdot} - \bar{X})^2$	$k-1$	$\frac{\mathrm{SAK}_3/f_3}{\mathrm{SAK}_{02}/f_{02}}$
Veksel-virkning	SAK_2 (se nedenfor)	$(k-1)$ $\cdot(m-1)$	$\frac{\mathrm{SAK}_2/f_2}{\mathrm{SAK}_1/f_1}$
Inden f. celler	$\mathrm{SAK}_1 = \sum_i\sum_j\sum_\nu(X_{ij\nu} - \bar{X}_{ij})^2$	$km(n-1)$	
Total	$\mathrm{SAK}_0 = \sum_i\sum_j\sum_\nu(X_{ij\nu} - X)^2$	$kmn-1$	

Her er

$$\begin{aligned}\mathrm{SAK}_2 &= n\sum_i\sum_j(\bar{X}_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2\\ \mathrm{SAK}_{02} &= \mathrm{SAK}_1 + \mathrm{SAK}_2\\ f_{02} &= f_1 + f_2 = km(n-1) + (k-1)(m-1).\end{aligned}$$

Under H_0, hypotesen om forsvindende vekselvirkning, er SAK_{02}/f_{02} et centralt skøn over σ^2. Hvis vi har fået accepteret hypotesen H_0, vil det derfor være rimeligt at sammenligne de øvrige variationer med denne "bedre" estimator (flere frihedsgrader, nemlig $f_1 + f_2$ og derfor mindre varians). Hvis vi ikke får accepteret hypotesen om forsvindende vekselvirkning, vil det sådan set ikke være rimeligt at fortsætte testningen, idet man da ikke får en simplere beskrivelse af situationen. (Der er $k \cdot m$ parametre μ_{ij} og $k \cdot m$ parametre τ_{ij}).

Vi samler nu de kritiske områder i

SÆTNING 5.8. Ved test på niveau α er det kritiske område for hypotesen H_0 om forsvindende vekselvirkning lig

$$C_0 = \Big\{(x_{111},\cdots,x_{kmn}) \mid \tfrac{\mathrm{sak}_2/f_2}{\mathrm{sak}_1/f_1} > \mathrm{F}\big((k-1)(m-1), km(n-1)\big)_{1-\alpha}\Big\}.$$

Hvis den hypotese accepteres, er de kritiske områder ved test på niveau α af hypotesen H_{01} om forsvindende rækkevirkning ($\alpha_1 = \cdots = \alpha_k = 0$), og hypotesen H_{02} om forsvindende søjlevirkning ($\beta_1 = \cdots = \beta_m = 0$) lig

$$C_1 = \Big\{(x_{111}, \cdots, x_{kmn}) | \tfrac{\mathrm{sak}_3/f_3}{\mathrm{sak}_{02}/f_{02}} > \mathrm{F}\big(k-1, km(n-1)+(k-1)(m-1)\big)_{1-\alpha}\Big\}$$

henholdsvis

$$C_2 = \Big\{(x_{111}, \cdots, x_{kmn}) | \tfrac{\mathrm{sak}_4/f_4}{\mathrm{sak}_{02}/f_{02}} > \mathrm{F}\big(m-1, km(n-1)+(k-1)(m-1)\big)_{1-\alpha}\Big\}.$$

Bevis. En umiddelbar følge af de foregående overvejelser. ■

Der gælder nogle analoge betragtninger til de p. 431 og p. 427 anførte angående **estimatorer for variansen og række- og søjleeffekterne**. (Angående de sidste skal man blot erstatte X_{ij} med $\bar{X}_{ij}$ i formlerne p. 427).

Vi anfører nogle nyttige beregningsskemaer.

i \ j	$\cdots$	j	$\cdots$	I alt
$\vdots$				
i		$S_{ij} = \sum_{\nu=1}^{n} X_{ij\nu}$ $\mathrm{SK}_{ij} = \sum_{\nu=1}^{n} X_{ij\nu}^2$		$S_{i\cdot} = \sum_{j=1}^{m} S_{ij}$ $\mathrm{SK}_{i\cdot} = \sum_{j=1}^{m} \mathrm{SK}_{ij}$
$\vdots$				
I alt		$S_{\cdot j} = \sum_{i=1}^{k} S_{ij}$ $\mathrm{SK}_{\cdot j} = \sum_{i=1}^{k} \mathrm{SK}_{ij}$		$S_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{m} S_{ij}$ $\mathrm{SK}_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{m} \mathrm{SK}_{ij}$

Endvidere beregnes

$$\sum_i \sum_j S_{ij}^2, \quad \sum_i S_{i\cdot}^2, \quad \sum_j S_{\cdot j}^2 \quad \text{og} \quad S_{\cdot\cdot}^2.$$

Da haves

$$
\begin{aligned}
\mathrm{SAK}_0 &= \sum_i \sum_j \sum_\nu X_{ij\nu}^2 - \frac{1}{kmn}(\sum_i \sum_j \sum_\nu X_{ij\nu})^2 = \mathrm{SK}_{\cdot\cdot} - \frac{1}{N} S_{\cdot\cdot}^2 \\
\mathrm{SAK}_1 &= \sum_i \sum_j \sum_\nu X_{ij\nu}^2 - \frac{1}{n} \sum_i \sum_j (\sum_\nu X_{ij\nu})^2 \\
&= \mathrm{SK}_{\cdot\cdot} - \frac{1}{n} \sum_i \sum_j S_{ij}^2 \\
\mathrm{SAK}_3 &= \frac{1}{mn} \sum_i (\sum_j \sum_\nu X_{ij\nu})^2 - \frac{1}{kmn}(\sum_i \sum_j \sum_\nu X_{ij\nu})^2 \\
&= \frac{1}{mn} \sum_i S_{i\cdot}^2 - \frac{1}{N} S_{\cdot\cdot}^2 \\
\mathrm{SAK}_4 &= \frac{1}{kn} \sum_j (\sum_i \sum_\nu X_{ij\nu})^2 - \frac{1}{kmn}(\sum_i \sum_j \sum_\nu X_{ij\nu})^2 \\
&= \frac{1}{kn} \sum_j S_{\cdot j}^2 - \frac{1}{N} S_{\cdot\cdot}^2 \\
\mathrm{SAK}_2 &= \mathrm{SAK}_0 - \mathrm{SAK}_1 - \mathrm{SAK}_3 - \mathrm{SAK}_4 .
\end{aligned}
$$

Vi giver nu et

EKSEMPEL 5.5. Ved en undersøgelse af forskellige blandemaskiners virkning på cements styrke har man foretaget følgende forsøg. Man har blandet en prøve Portland cement grundigt og opdelt den i mindre delprøver. I 3 blandemaskiner har man derefter fremstillet cementmørtel af disse prøver. Af denne mørtel har man i nogle forme støbt terninger. Efter 24 timers hærdning tog man terningerne ud af formene. Efter yderligere 7 dages lagring prøvede man terningernes styrke ved hjælp af 3 knusemaskiner. De opnåede resultater fremgår af følgende skema, hvor resulateterne er opgivet i pund pr. kvadrattomme.

Blandemaskine nr.	Knusemaskine nr.	Observationer			
1	1	5280	5520	4760	5800
1	2	4340	4400	5020	6200
1	3	4160	5180	5320	4600
2	1	4420	5280	5580	4900
2	2	5340	4880	4960	6200
2	3	4180	4800	4600	4480
3	1	5360	6160	5680	5500
3	2	5720	4760	5620	5560
3	3	4460	4930	4680	5600

Vi antager nu, at disse data er realiserede udfald af indbyrdes uafhængige, normalt fordelte stokastiske variable $X_{ij\nu}, i = 1,2,3, j = 1,2,3, \nu = 1,2,3,4$ med $\mathrm{E}(X_{ij\nu}) = \mu_{ij}$ og $\mathrm{V}(X_{ij\nu}) = \sigma^2$. Vi antager endvidere, at middelværdien af brudstyrken er skrevet op som en sum (som i (5.5))

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \tau_{ij} \qquad i = 1,2,3, \quad j = 1,2,3.$$

Her er μ altså det universelle niveau (der angiver det, man kan kalde "brudstyrken af beton"), α_i er den specielle effekt, der skyldes den i'te blandemaskine, og β_j den, der skyldes den j'te knusemaskine. Endelig angiver τ_{ij} vekselvirkningen mellem den i'te blandemaskine og den j'te knusemaskine.

Da formålet med forsøget har været at få belyst blandemaskinernes indvirkning på brudstyrken, vil vi først teste, om der er nogen vekselvirkning mellem blandemaskiner og knusemaskiner, i.e. vi vil teste

$$H_0: \quad \forall i,j(\tau_{ij} = 0) \qquad \text{mod} \qquad H_1: \quad \exists i,j(\tau_{ij} \neq 0).$$

Hvis denne hypotese accepteres, vil vi teste, om det kan antages, at der ikke er nogen blandemaskineeffekt, i.e. teste

$$H_{01}: \quad \forall i(\alpha_i = 0) \qquad \text{mod} \qquad H_{11}: \quad \exists i(\alpha_i \neq 0).$$

For at lette beregningerne indfører vi et beregningsnulpunkt $\alpha = 5000$ pund/kvadrattomme og en beregningsenhed $\beta = 10$ pund/kvadrattomme (jfr. p. 410). Vi får da følgende datamatrix

		Knusemaskine nr.		
		1	2	3
Blandemaskine nr.	1	28	-66	-84
		52	-60	18
		-24	2	32
	2	80	120	-40
		-58	34	-82
		28	-12	-20
	3	58	-4	-40
		-10	120	-52
		36	72	-54
		116	-24	-7
		68	62	-32
		50	56	60

Vi danner nu beregningsskemaet over summer og kvadratsummer.

		Knusemaskine nr.			Total
		1	2	3	
Blandema-skine nr.	1	136 10464	−4 22360	−74 10004	58 42828
	2	18 7612	138 15716	−194 11428	−38 34756
	3	270 21876	166 12740	−33 7589	403 42205
Total		424 39952	300 50816	−301 29021	423 119789

Ved hjælp af skemaet dannes

$$\begin{aligned}
\tfrac{1}{n}\sum_i\sum_j s_{ij}^2 &= \tfrac{1}{4}\left(136^2+\cdots+(-33)^2\right) &= 45634\\
\tfrac{1}{N}s_{..}^2 &= \tfrac{1}{36}178929 &= 4970.25\\
\sum_i s_{i.}^2 &= 167217\\
\sum_j s_{.j}^2 &= 360377.
\end{aligned}$$

Vi finder nu kvadratsummerne

$$\begin{aligned}
\mathrm{sak}_0 &= \mathrm{sk}_{..} - \tfrac{1}{N}s_{..}^2 &&= 119789 - 4970.25 &&= 114818.75\\
\mathrm{sak}_1 &= \mathrm{sk}_{..} - \tfrac{1}{n}\sum_i\sum_j s_{ij}^2 &&= 119789 - 45634 &&= 74155\\
\mathrm{sak}_3 &= \tfrac{1}{mn}\sum_i s_{i.}^2 - \tfrac{1}{N}s_{..}^2 &&= \tfrac{1}{12}167217 - 4970.25 &&= 8964.50\\
\mathrm{sak}_4 &= \tfrac{1}{kn}\sum_j s_{.j}^2 - \tfrac{1}{N}s_{..}^2 &&= \tfrac{1}{12}360377 - 4970.25 &&= 25061.17\\
\mathrm{sak}_2 &= \mathrm{sak}_0 - (\mathrm{sak}_1 + \mathrm{sak}_3 + \mathrm{sak}_4) &&&&= 6638.08.
\end{aligned}$$

Endvidere er

$$\frac{\mathrm{sak}_1+\mathrm{sak}_2}{km(n-1)+(k-1)(m-1)} = \frac{1}{31}80793.08 = 2606.2.$$

Vi samler nu disse størrelser i **variansanalyseskemaet**

Variation	SAK	f	SAK$/f$	Test
Mellem knuse-maskiner	25061.17	2	12530.6	4.81
Mellem blande-maskiner	8964.50	2	4482.3	1.72
Vekselvirkning	6638.08	4	1659.5	0.60
Inden for celler	74155.00	27	2746.5	
I alt	114818.75	35		

Ved test på 5% niveau bliver de kritiske værdier $\mathrm{F}(4,27)_{0.95} = 2.73$ og $\mathrm{F}(2,31)_{0.95} = 3.31$.

Med $0.60 < \mathrm{F}(4,27)_{0.50}$ og $1.72 < \mathrm{F}(2,31)_{0.90}$ vil vi altså acceptere hypotesen om forsvindende vekselvirkning på alle niveauer mindre end 50% og hypotesen om forsvindende blandemaskineeffekt på alle niveauer mindre end 10%.

Vi vil derfor konkludere, at man med det foreliggende materiale må konkludere, at blandemaskinerne ikke har signifikant indflydelse på brudstyrken af den færdige beton.

Vi bemærker i øvrigt, at

$$\mathrm{F}(2,31)_{0.975} < 4.81 < \mathrm{F}(2,31)_{0.99},$$

således at vi får afvist hypotesen om, at knusemaskinerne virker ens på f.eks. et 5% signifikansniveau. ♦

5.1.3 Romersk kvadrat

Somme tider er man i et forsøg nødt til at tage hensyn til, at visse faktorer ikke kan varieres frit, således at man ikke direkte kan eliminere de systematiske fejl, man frygter, er tilstede ved forsøget.

Lad os eksempelvis antage, at vi har et apparat til at efterprøve kvaliteten af tekstil-varer. Apparatet kan i en kørsel efterprøve 4 forskellige stykker tekstil. Nu kunne der i forsøgsresultaterne være en variation, der skyldes placeringen af tekstilstykket i maskinen, og en, der skyldes selve kørslen, d.v.s. vi kan ikke regne med at få samme resulater ved flere efterprøvninger af samme tekstiler.

Vi ønsker at få efterprøvet 4 forskellige vareprøver, A, B, C og D. Man kan nu sikre

sig, at man ved at foretage $4 \times 4 = 16$ forsøg i det mindste kan få efterprøvet hver af de 4 vareprøver A, B, C og D netop 1 gang i hver kørsel og netop 1 gang i hver position af maskinen. Dette kan f.eks. ske ved, at man efterprøver tekstilerne som i nedenstående skema

	Position			
Kørsel	1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

DEFINITION 5.1. Et sådant arrangement af n bogstaver i et kvadrat med siden n, således at hvert bogstav står netop 1 gang i hver række og i hver søjle, kaldet et **romersk kvadrat** (engelsk: latin square). ▲

Der findes udførlige tabeller over romerske kvadrater af forskellige størrelser. Vi kan f.eks. henvise til Fisher og Yates (1948).

Hvis vi udfører n^2 eksperimenter som anført i det romerske kvadrat, får vi observationer X_{ijk}, som vi vil beskrive ved **indbyrdes uafhængige** $\mathrm{N}(\mu_{ijk}, \sigma^2)$**–fordelte stokastiske variable**, hvor μ_{ijk} er middelværdien af observationen i (i, j, k)'te celle. Vi opdeler middelværdierne således, at

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k,$$

hvor μ er niveauet, α_i rækkeeffekten (d.v.s. kørselseffekten), β_j søjleeffekten (d.v.s. positionseffekten) og γ_k "prøve"effekten. For at sikre entydigheden af disse effekter må vi forudsætte, at $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$.

Den totale variation omkring det totale middeltal spaltes nu i variationer, der skyldes ovennævnte effekter, plus residualvariationen (restvariationen). Der gælder

$$\begin{aligned}\sum_i \sum_j (X_{ij} - \bar{X})^2 = & n \sum_j (\bar{X}_{\cdot j} - \bar{X})^2 + n \sum_i (\bar{X}_{i\cdot} - \bar{X})^2 \\ & + n \sum_k (\bar{X}_k - \bar{X})^2 \\ & + \sum_i \sum_j (X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} - \bar{X}_k + 2\bar{X})^2,\end{aligned}$$

hvor $\bar{X}_k$ er middeltallet af målingerne af prøve nr. k. Spaltningen kan fås ved en viderespaltning af vekselvirkningskvadratafvigelsessummen i den tosidede variansanalyse med én observation pr. celle.

Frihedsgraderne for komponenterne er $n-1$, $n-1$, $n-1$ og $(n-1)(n-2)$. Det kan ved hjælp af Sætning 1.77, side 195, vises, at de er indbyrdes uafhængige og $\sigma^2\chi^2$-

fordelte med de respektive antal frihedsgrader, såfremt de tilsvarende effekter er lig 0. Dette kan udnyttes til at konstruere tests for hypoteser om forsvindende effekter. F.eks. kan en hypotese om, at $\gamma_k = 0$, $k = 1, \cdots, n$ testes ved at anvende det kritiske område

$$\left\{ \frac{n\sum(\bar{X}_k - \bar{X})^2/(n-1)}{\sum_{i,j}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} - \bar{X}_k + 2\bar{X})^2/\left[(n-1)(n-2)\right]} > \mathrm{F}\big(n-1,(n-1)(n-2)\big)_{1-\alpha} \right\}$$

Vi samler resultaterne i et variansanalyseskema

Variation	SAK	f	Test
Mellem søjler	$\mathrm{SAK}_4 = n\sum_j(\bar{X}_{\cdot j} - \bar{X})^2$	$n-1$	S_4^2/S_1^2
Mellem rækker	$\mathrm{SAK}_3 = n\sum_i(\bar{X}_{i\cdot} - \bar{X})^2$	$n-1$	S_3^2/S_1^2
Mellem prøver	$\mathrm{SAK}_2 = n\sum_k(\bar{X}_k - \bar{X})^2$	$n-1$	S_2^2/S_1^2
Residual	$\mathrm{SAK}_1 = \sum_i\sum_j(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} - \bar{X}_k + 2\bar{X})^2$	$(n-1)\cdot(n-2)$	
Total	$\mathrm{SAK}_0 = \sum_i\sum_j(X_{ij} - \bar{X})^2$	n^2-1	

Vi anfører beregningsformler:

$$\begin{aligned}
\mathrm{SAK}_4 &= n\textstyle\sum_j(\bar{X}_{\cdot j} - \bar{X})^2 = \frac{1}{n}\sum_j(\sum_i X_{ij})^2 - \frac{1}{n^2}(\sum_i\sum_j X_{ij})^2 \\
\mathrm{SAK}_3 &= n\textstyle\sum_i(\bar{X}_{i\cdot} - \bar{X})^2 = \frac{1}{n}\sum_i(\sum_j X_{ij})^2 - \frac{1}{n^2}(\sum_i\sum_j X_{ij})^2 \\
\mathrm{SAK}_2 &= n\textstyle\sum_k(\bar{X}_k - \bar{X})^2 = \frac{1}{n}\sum_k \bar{X}_k^2 - \frac{1}{n^2}(\sum_i\sum_j X_{ij})^2
\end{aligned}$$

$$\begin{aligned}
\mathrm{SAK}_0 &= \sum_i\sum_j(X_{ij} - \bar{X})^2 = \sum_i\sum_j X_{ij}^2 - \frac{1}{n^2}(\sum_i\sum_j X_{ij})^2 \\
\mathrm{SAK}_1 &= \mathrm{SAK}_0 - \mathrm{SAK}_2 - \mathrm{SAK}_3 - \mathrm{SAK}_4.
\end{aligned}$$

Vi anfører nu resultaterne for tekstilprøvemaskinen

	Position i maskinen			
Kørsel	4	2	1	3
2	A(251)	B(241)	D(227)	C(229)
3	D(234)	C(273)	A(274)	B(226)
1	C(235)	D(236)	B(218)	A(268)
4	B(195)	A(270)	C(230)	D(225)

Vi trækker 240 fra samtlige observationer og får

	Position i maskinen				
Kørsel	4	2	1	3	Total
2	(A)+11	(B)+ 1	(D)−13	(C)−11	−12
3	(D)− 6	(C)+33	(A)+34	(B)−14	+47
1	(C)− 5	(D)− 4	(B)−22	(A)+28	− 3
4	(B)−45	(A)+30	(C)−10	(D)−15	−40
Total	−45	+60	−11	−12	− 8

Total: Prøve A: +103
Prøve B: − 80
Prøve C: + 7
Prøve D: − 38.

Vi får eksempelvis

$$\mathrm{SAK}_2 = \frac{1}{4}\left(103^2 + (-80)^2 + 7^2 + (-38)^2\right) - \frac{1}{16}(-8)^2 = 4621.5.$$

De øvrige størrelser beregnes analogt. Vi samler resultaterne

Variation	SAK	f	S^2	Test
Mellem positioner	1468.5	3	489.5	8.0
Mellem kørsler	986.5	3	328.8	5.4
Mellem prøver	4621.5	3	1540.5	25.0
Residual	367.5	6	61.25	
Total	7444.0	15		

Da $\mathrm{F}(3,6)_{95\%} = 4.76$, ses, at vi ved et test på $\alpha = 5\%$ signifikansniveau må forkaste alle hypoteser om forsvindende effekter. Det er vel især værd at bemærke, at der er positionseffekt og en kørselseffekt. Dette ville man måske ikke forvente i første omgang, således at man blot ville have anvendt en ensidet variansanalysemodel til sammenligning af prøverne. Dette ville have givet en stor variation inden for prøverne, hvilket ville have sænket følsomheden af sammenligningen mellem prøverne væsentligt.

5.1.4 Faktorforsøg

Lad os antage, at vi ønsker at bestemme optimale betingelser (maksimal effektivitet) for et pletteringsbad. De faktorer, vi kan ændre, er badets temperatur og sulfidkoncentrationen. Vi bestemmer os for at betragte koncentrationerne 4, 8 og 12 g/liter og temperaturerne $80°$ F og $160°$ F. Vi forsøger 3 gange i alle kombinationer og får følgende tal for reflektiviteten af det behandlede metal.

Faktor A Koncentration	Faktor B Temperatur	Forsøg 1	Forsøg 2	Forsøg 3
4	80	3.5	3.0	2.7
4	160	2.2	2.3	2.4
8	80	7.1	6.9	7.5
8	160	5.2	4.6	6.8
12	80	10.8	10.6	11.0
12	160	7.6	7.1	7.3

Vi vil nu give en matematisk model, der kan danne grundlag for en analyse af disse data. Vi har givet observationer

$$X_{ij\nu}, \qquad i = 1, \cdots, k; \quad j = 1, \cdots, m; \quad \nu = 1, \cdots, n.$$

Her angiver i niveauet for faktor A (d.v.s. 1,2 eller 3), j niveauet for faktor B (d.v.s. 1 eller 2), og n er antallet af gentagelser (her 3).

Vi antager, at observationerne er indbyrdes uafhængige og normalfordelte med fælles varians σ^2. Endvidere antager vi, at middelværdien af $X_{ij\nu}$ kan skrives

$$\mu_{ij\nu} = \mu + \alpha_i + \beta_j + \tau_{ij} + \rho_\nu.$$

Her angiver μ niveauet, α_i effekten af faktor A i niveau i, β_j effekten af faktor B i niveau j; τ_{ij} er vekselvirkningseffekten, og ρ_ν er effekten af den ν'te gentagelse. For at sikre entydigheden af ovennævnte størrelser må vi forudsætte, at

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \tau_{ij} = \sum_j \tau_{ij} = \sum_\nu \rho_\nu = 0.$$

DEFINITION 5.2. En sådan model som ovenfor beskrevet kaldes et $k \times m$ **faktorforsøg gentaget n gange**. ▲

Forskellen mellem denne model og en tosidet variansanalysemodel er, at vi har tilføjet leddet ρ_ν. Grunden er, at det ikke a priori kan anses for givet, at alle forsøgsbetingelser

er ens fra gang til gang, således at vi må regne med en systematisk forskel mellem resultater fra forskellige forsøg.

Det er vigtigt at notere, at i det beskrevne forsøg er de 3×2 målinger i Forsøg 1 udført i en tilfældig rækkefølge, men iøvrigt stort set samtidig. Derefter er Forsøg 2's 3×2 målinger udført, etc. Herved adskiller forsøget sig fra et almindeligt tosidet variansanalyse-forsøg, hvor **alle** data måles i en tilfældig rækkefølge uden opdeling i "Forsøg".

Det, man er interesseret i, er at undersøge, om virkningen af en eller begge af faktorerne A og B kan antages at være lig 0. For at få en mulighed for at foretage en sådan analyse vil vi spalte den totale variation i bidrag, der stammer fra de beskrevne effekter.

Fra den tosidede variansanalyse haves

$$\begin{aligned}\sum_i\sum_j\sum_\nu (X_{ij\nu}-\bar{X})^2 &= \sum_i\sum_j\sum_\nu (X_{ij\nu}-\bar{X}_{ij})^2 \\ &+ n\sum_i\sum_j (\bar{X}_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})^2 \\ &+ mn\sum_i (\bar{X}_{i\cdot}-\bar{X})^2 + kn\sum_j (\bar{X}_{\cdot j}-\bar{X})^2.\end{aligned}$$

Nu har vi imidlertid også en "gentagelsesfaktor" ρ_ν, hvorfor vi ønsker at få spaltet leddet

$$\sum_i\sum_j\sum_\nu (X_{ij\nu}-\bar{X}_{ij})^2$$

yderligere. Vi har

$$\begin{aligned}&\sum_i\sum_j\sum_\nu (X_{ij\nu}-\bar{X}_{ij})^2 = \\ &\quad \sum_i\sum_j\sum_\nu \left(X_{ij\nu}-\bar{X}_{ij}-(\bar{X}_{\cdot\cdot\nu}-\bar{X})+(\bar{X}_{\cdot\cdot\nu}-\bar{X})\right)^2 \\ &= \sum_i\sum_j\sum_\nu (X_{ij\nu}-\bar{X}_{ij}-\bar{X}_{\cdot\cdot\nu}+\bar{X})^2 \\ &\quad + km\sum_\nu (\bar{X}_{\cdot\cdot\nu}-\bar{X})^2 \\ &\quad + 2\sum_\nu (\bar{X}_{\cdot\cdot\nu}-\bar{X})\sum_i\sum_j (X_{ij\nu}-\bar{X}_{ij}-\bar{X}_{\cdot\cdot\nu}+\bar{X}).\end{aligned}$$

Det ses, at det dobbelte produkt er 0, hvorfor vi har spaltningen

$$\begin{aligned}\sum_i\sum_j\sum_\nu (X_{ij\nu}-\bar{X})^2 = \\ \sum_i\sum_j\sum_\nu (X_{ij\nu}-\bar{X}_{ij}-\bar{X}_{\cdot\cdot\nu}+\bar{X})^2 + km\sum_\nu (\bar{X}_{\cdot\cdot\nu}-\bar{X})^2 \\ + mn\sum_i (\bar{X}_{i\cdot}-\bar{X})^2 + kn\sum_j (\bar{X}_{\cdot j}-\bar{X})^2 \\ + n\sum_i\sum_j (\bar{X}_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})^2,\end{aligned}$$

d.v.s. vi har fået opspaltet den totale variation i én komponent, der skyldes "fejl" (den såkaldte residualvariation), plus én, der skyldes de gentagne forsøg plus én, der skyldes A-effekten, plus én, der skyldes B-effekten, plus én, der skyldes vekselvirkningen mellem faktorerne A og B.

Antallet af frihedsgrader i ovenstående led er

$$kmn-1 = (km-1)(n-1)+(n-1)+(k-1)+(m-1)+(k-1)(m-1).$$

Det følger af spaltningssætningen, at kvadratafvigelsessummerne er indbyrdes uafhængige og $\sigma^2\chi^2(f_i)$-fordelte, hvor f_i'erne er de ovenfor anførte frihedsgrader, hvis de pågældende effekter er lig 0.

Heraf fås let variansanalyseskemaet

Variation		SAK	f	Test
Gentagelse af forsøg		$\mathrm{SAK}_G = km\sum_\nu(\bar{X}_{\cdot\cdot\nu}-\bar{X})^2$	$n-1$	S_G^2/S_R^2
Hoved-effekter	A	$\mathrm{SAK}_A = mn\sum_i(\bar{X}_{i\cdot}-\bar{X})^2$	$k-1$	S_A^2/S_R^2
	B	$\mathrm{SAK}_B = kn\sum_j(\bar{X}_{\cdot j}-\bar{X})^2$	$m-1$	S_B^2/S_R^2
Vekselvirkning	AB	$\mathrm{SAK}_{AB} = n\sum_i\sum_j(\bar{X}_{ij}-\bar{X}_{i\cdot}-\bar{X}_{\cdot j}+\bar{X})^2$	$(k-1)\times(m-1)$	S_{AB}^2/S_R^2
Residual		$\mathrm{SAK}_R = \sum_i\sum_j\sum_\nu(X_{ij\nu}-\bar{X}_{ij}-\bar{X}_{\cdot\cdot\nu}+\bar{X})^2$	$(km-1)\times(n-1)$	S_R^2
Total		$\mathrm{SAK}_0 = \sum_i\sum_j\sum_\nu(X_{ij\nu}-\bar{X})^2$	$kmn-1$	

De kritiske områder er som sædvanlig store værdier af teststørrelserne, der er F-fordelte med de anførte frihedsgrader, såfremt de tilsvarende effekter er lig 0. Eksempelvis er det kritiske område for test af

$$H_0: \quad \forall i,j(\tau_{ij}=0) \qquad \text{mod} \qquad H_1: \quad \exists i,j(\tau_{ij}\neq 0)$$

på niveau α givet ved

$$C=\Big\{(x_{111},\cdots,x_{kmn})|\tfrac{s_{AB}^2}{s_R^2} > \mathrm{F}\big((k-1)(m-1),(km-1)(n-1)\big)_{1-\alpha}\Big\},$$

hvor

$$S_{AB}^2 = \frac{1}{(k-1)(m-1)}\mathrm{SAK}_{AB}$$

og

$$S_R^2 = \frac{1}{(km-1)(n-1)}\mathrm{SAK}_R.$$

Vi anfører nogle beregningsformler:

$$\begin{aligned}
\mathrm{SAK}_0 &= \sum_i\sum_j\sum_\nu X_{ij\nu}^2 - \frac{1}{kmn}\Big(\sum_i\sum_j\sum_\nu X_{ij\nu}\Big)^2\\
\mathrm{SAK}_G &= \frac{1}{km}\sum_\nu\Big(\sum_i\sum_j X_{ij\nu}\Big)^2 - \frac{1}{kmn}\Big(\sum_i\sum_j\sum_\nu X_{ij\nu}\Big)^2\\
\mathrm{SAK}_A &= \frac{1}{mn}\sum_i\Big(\sum_j\sum_\nu X_{ij\nu}\Big)^2 - \frac{1}{kmn}\Big(\sum_i\sum_j\sum_\nu X_{ij\nu}\Big)^2\\
\mathrm{SAK}_B &= \frac{1}{kn}\sum_j\Big(\sum_i\sum_\nu X_{ij\nu}\Big)^2 - \frac{1}{kmn}\Big(\sum_i\sum_j\sum_\nu X_{ij\nu}\Big)^2\\
\mathrm{SAK}_{AB} &= \frac{1}{n}\sum_i\sum_j\Big(\sum_\nu X_{ij\nu}\Big)^2 - \frac{1}{mn}\sum_i\Big(\sum_j\sum_\nu X_{ij\nu}\Big)^2\\
&\quad - \frac{1}{kn}\sum_j\Big(\sum_i\sum_\nu X_{ij\nu}\Big)^2 + \frac{1}{kmn}\Big(\sum_i\sum_j\sum_\nu X_{ij\nu}\Big)^2\\
\mathrm{SAK}_R &= \sum_i\sum_j\sum_\nu X_{ij\nu}^2 - \frac{1}{km}\sum_\nu\Big(\sum_i\sum_j X_{ij\nu}\Big)^2\\
&\quad - \frac{1}{n}\sum_i\sum_j\Big(\sum_\nu X_{ij\nu}\Big)^2 + \frac{1}{kmn}\Big(\sum_i\sum_j\sum_\nu X_{ij\nu}\Big)^2.
\end{aligned}$$

I det aktuelle tilfælde fås:

$$\frac{1}{kmn}(\sum_{ij\nu} x_{ij\nu})^2 = \frac{108.6^2}{18} = 655.22$$

$$\begin{aligned}
\text{sak}_0 &= 3.5^2 + \cdots + 7.3^2 - 655.22 = 149.38 \\
\text{sak}_G &= \frac{1}{2\cdot 3}(36.4^2 + 34.5^2 + 37.7^2) - 655.22 = 0.86 \\
\text{sak}_A &= \frac{1}{2\cdot 3}(16.1^2 + 38.1^2 + 54.4^2) - 655.22 = 123.14 \\
\text{sak}_B &= \frac{1}{3\cdot 3}(63.1^2 + 45.5^2) - 655.22 = 17.21 \\
\text{sak}_{AB} &= \frac{1}{3}(9.2^2 + \cdots + 22.0^2) - \frac{1}{6}(16.1^2 + 38.1^2 + 54.4^2) \\
&\quad - \frac{1}{9}(63.1^2 + 45.5^2) + 655.22 = 5.70 \\
\text{sak}_R &= (3.5^2 + \cdots + 7.3^2) - \frac{1}{6}(36.4^2 + 34.5^2 + 37.7^2) \\
&\quad - \frac{1}{3}(9.2^2 + \cdots + 22.0^2) + 655.22 = 2.47.
\end{aligned}$$

Vi har derfor variansanalyseskemaet

Variation		SAK	f	S^2	Test
Gentagelser		0.86	2	0.43	1.72
Hovedeffekter	A	123.14	2	61.57	246.0
	B	17.21	1	17.21	68.8
Vekselvirkning	AB	5.70	2	2.85	11.4
Residual		2.47	10	0.25	
Total		149.38	17		

Hvis vi tester på et 1% signifikansniveau, ses, at den eneste effekt, som vi kan antage er 0, er gentagelseseffekten ρ_ν. Man kan altså konkludere, at forsøgsbetingelserne har været ensartede fra forsøg til forsøg.

Ønsker man nu at bestemme optimale produktionsbetingelser findes disse ved at opskrive estimaterne for $\mu + \alpha_i + \beta_j + \tau_{ij}$ og se, hvilken (i, j) kombination, der giver det bedste resultat. Det ses fra data, at det er kombinationen (A=12, B=80), som giver størst reflektivitet i gennemsnit.

Denne metode udbygges let til at gælde flere faktorer. Det er dog almindeligt at begrænse sig til f.eks. 2 eller 3 niveauer. Dette vil vi dog ikke komme ind på her. En

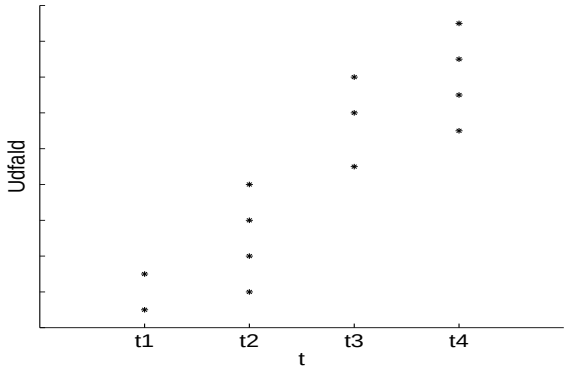
række eksempler på anvendelsen af faktorforsøg findes i [15].

5.2 Regressionsanalyser

5.2.1 Regressionsanalyse med 1 uafhængig variabel

Man er ofte i den situation, at man måler udfaldet af et eksperiment med et stokastisk udfald til forskellige tider $t_1, \cdots, t_k$; eller man kan måle egenskaber hos et produkt ved forskellige temperaturer $t_1, \cdots, t_k$ eller ved forskellige koncentrationer $t_1, \cdots, t_k$.

Man vil da få et diagram som f.eks.



hvor der altså foreligger 2 målinger for $t = t_1, \cdots, 4$ målinger til tid t_k. Idet vi regner med, at alle observationer til tid t_i har samme middelværdi μ_i, vil man ofte være interesseret i at undersøge en antagelse om, at

$$\mu_i = a + bt_i, \qquad i = 1, \cdots, k, \tag{5.6}$$

d.v.s. undersøge, om målingerne grupperer sig omkring en ret linie og i givet fald at estimere denne linie.

Estimationsproblemet kan løses ved hjælp af mindste kvadraters metode uden nogen anden antagelse om observationernes fordeling end (5.6):
$\mathrm{E}(X_{ij}) = a + bt_i$. For at finde mindste kvadraters estimatoren skal vi minimalisere

$$\mathrm{f}(a, b) = \sum_i \sum_j (X_{ij} - a - bt_i)^2,$$

hvor X_{ij} naturligvis er den j'te observation til $t = t_i$ $(j = 1, \cdots, n_i)$. Vi finder de

partielle afledede og sætter dem lig 0.

$$\frac{\partial \mathrm{f}(a,b)}{\partial a} = -2\sum_i\sum_j (X_{ij} - a - bt_i) = 0,$$

d.v.s.

$$N\bar{X} - Na - bN\bar{t} = 0$$

eller

$$a = \bar{X} - b\bar{t},$$

hvor

$$\begin{aligned} N &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} 1 = \sum_{i=1}^{k} n_i \\ \bar{X} &= \frac{1}{N}\sum_i\sum_j X_{ij} \\ \bar{t} &= \frac{1}{N}\sum_{i=1}^{k} n_i t_i. \end{aligned}$$

Tilsvarende

$$\frac{\partial \mathrm{f}(a,b)}{\partial b} = -2\sum_i\sum_j (t_i X_{ij} - t_i a - bt_i^2) = 0$$

d.v.s.

$$\sum_i\sum_j t_i X_{ij} - Na\bar{t} - b\sum_i n_i t_i^2 = 0,$$

eller, idet vi indsætter den for a fundne værdi

$$\sum_i\sum_j t_i X_{ij} - N\bar{X}\bar{t} + Nb\bar{t}^2 - b\sum_i n_i t_i^2 = 0$$

d.v.s.

$$b = \frac{\sum\sum_{i\ j} t_i X_{ij} - N\bar{X}\bar{t}}{\sum_i n_i t_i^2 - N\bar{t}^2} = \frac{\sum\sum_{i\ j}(X_{ij} - \bar{X})(t_i - \bar{t})}{\sum_i n_i (t_i - \bar{t})^2} = \frac{\mathrm{SAP}_{xt}}{\mathrm{SAK}_t}.$$

Det midterste lighedstegn fremgår af

$$\sum_i\sum_j (X_{ij} - \bar{X})(t_i - \bar{t}) = \sum_i\sum_j X_{ij} t_i - N\bar{X}\bar{t}.$$

Kaldes løsningen for b nu $\hat{b}$, fås altså

$$\hat{\mu} = \bar{X} - \hat{b}\bar{t} + \hat{b}t_i = \bar{X} + \hat{b}(t_i - \bar{t}).$$

Vi reparametriserer nu, og sætter $a = \alpha - \beta\bar{t}$ og $b = \beta$, hvorved modellen (5.6) bliver

$$\mu_i = \alpha + \beta(t_i - \bar{t}), \qquad i = 1, \cdots, k, \tag{5.7}$$

Dette giver os estimatorerne

$$\begin{aligned} \hat{\alpha} &= \bar{X} \\ \hat{\beta} &= \frac{\mathrm{SAP}_{xt}}{\mathrm{SAK}_t}. \end{aligned}$$

Vi samler disse overvejelser i

SÆTNING 5.9. Lad der være givet stokastiske variable

$$X_{ij}, \qquad i = 1, \cdots, k \quad j = 1, \cdots, n_i,$$

således at $\mathrm{E}(X_{ij}) = \mu_i$ tilfredsstiller (5.7) for alle (i, j). Da er mindste kvadraters estimatorerne for α og β lig

$$\begin{aligned} \hat{\alpha} &= \bar{X} = \frac{1}{N}\sum_i\sum_j X_{ij} \\ \hat{\beta} &= \frac{\mathrm{SAP}_{xt}}{\mathrm{SAK}_t} = \frac{\sum\sum(X_{ij} - \bar{X})(t_i - \bar{t})}{\sum n_i (t_i - \bar{t})^2}. \end{aligned}$$

Hvis de stokastiske variable er uafhængige og normalt fordelte med samme varians, er disse estimatorer ligeledes maximum likelihood estimatorer, og de er stokastisk uafhængige og normalt fordelte

$$\begin{aligned}\hat{\alpha} &\in \mathrm{N}(\alpha, \frac{\sigma^2}{N}) \\ \hat{\beta} &\in \mathrm{N}(\beta, \frac{\sigma^2}{\mathrm{SAK}_t}).\end{aligned}$$

Bevis. Sætningens første del er en umiddelbar følge af de tidligere overvejelser. Udsagnet om maximum likelihood estimatorer følger direkte af sætning 2.11, p. 253. Uafhængigheden følger af Sætning 1.77, side 195. ■

BEMÆRKNING 5.5. Vi ser, at disse skøn svarer til, at man vælger den linie $\mu = \alpha + \beta(t-\bar{t})$, for hvilken summen af kvadraterne på de lodrette afstande fra observationerne til linien er mindst mulig. ▼

BEMÆRKNING 5.6. Linien $\mu = \bar{x} + \hat{\beta}(t - \bar{t})$ går åbenbart gennem tyngdepunktet $(\bar{t}, \bar{x})$ for alle punkterne. Hældningen for linien gennem (t_i, x_{ij}) og $(\bar{t}, \bar{x})$ er

$$\frac{x_{ij} - \bar{x}}{t_i - \bar{t}}.$$

Det vejede gennemsnit af disse hældninger med vægtene $(t_i - \bar{t})^2$ er

$$\frac{1}{\sum_i \sum_j (t_i - \bar{t})^2} \sum_i \sum_j (x_{ij} - \bar{x})(t_i - \bar{t}) = \frac{\mathrm{sap}_{xt}}{\mathrm{sak}_t} = \hat{\beta}.$$

Den estimerede **regressionslinie** $\hat{\mu}(t) = \bar{x} + \hat{\beta}(t - \bar{t})$ går altså gennem observationernes tyngdepunkt og har en hældningskoefficient, som er et vejet gennemsnit af hældningskoefficienterne for de linier, der går gennem tyngdepunktet og de enkelte observationer. ▼

Hvis vi ønsker at teste, om der foreligger en sådan lineær sammenhæng, må vi gøre de i sætningen anførte forudsætninger om observationernes fordeling, d.v.s. vi forudsætter, at de stokastiske variable

$$X_{ij}, \qquad i = 1, \cdots, k, \quad j = 1, \cdots, n_i$$

1. er uafhængige
2. er normalt fordelte
3. har samme varians
4. har middelværdier $\mathrm{E}(X_{ij}) = \mu_i$.

Her angiver 4., at vi forudsætter, at der er homogenitet inden for grupper.

Under disse forudsætninger kan vi teste lineariteten ved en variansanalyseteknik, idet vi kan spalte den totale variation som følger.

$$\begin{aligned}\sum_i\sum_j (X_{ij}-\bar{X})^2 &= \sum_i\sum_j (X_{ij}-\bar{X}_i)^2 \\ &+ \sum_i n_i\big(\bar{X}_i-\bar{X}-\hat{\beta}(t_i-\bar{t})\big)^2 \\ &+ \sum_i n_i\big(\bar{X}+\hat{\beta}(t_i-\bar{t})-\bar{X}\big)^2 \\ &= \sum_i\sum_j (X_{ij}-\bar{X}_i)^2 + \sum_i n_i\big(\bar{X}_i-\bar{X}-\hat{\beta}(t_i-\bar{t})\big)^2 \\ &+ \hat{\beta}^2\sum_i n_i(t_i-\hat{t})^2 .\end{aligned}$$

Her angiver $\bar{X}_i$ naturligvis gennemsnittet i den i'te gruppe, i.e.

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} .$$

Vi har altså fået spaltet den totale variation i et bidrag, der skyldes observationernes afvigelse fra gruppemiddeltallet, plus et bidrag, der skyldes middeltallenes afvigelse fra regressionslinien, plus et bidrag, der skyldes regressionsliniens variation (d.v.s. hældningskoefficientens afvigelse fra 0). Frihedsgraderne for ovenstående kvadratafvigelsessummer ses at være

$$N-1 = \sum(n_i-1) \;+\; k-2 \;+\; 1 = N-k \;+\; k-2 \;+\; 1.$$

Heraf følger, at kvadratafvigelsessummerne er indbyrdes uafhængige og $\sigma^2\chi^2(f_i)$-fordelte, såfremt hypoteserne $\mu_i = \alpha + \beta(t_i-\bar{t})$ og $\beta = 0$ er opfyldte

(Sætning 1.77, side 195). Dette udnyttes på sædvanlig vis til at konstruere tests for disse hypoteser.

Resultaterne samles i et variansanalyseskema:

Variation	SAK	f	Test
Regressionslinien (d.v.s. β)	$\mathrm{SAK}_3 = \hat{\beta}^2 \sum_{i=1}^{k} n_i (t_i - \bar{t})^2$	1	$\frac{\mathrm{SAK}_3}{\mathrm{SAK}_{02}/f_{02}}$
Omkring regressionslinien	$\mathrm{SAK}_2 = \sum_{i=1}^{k} n_i \left(\bar{X}_i - \bar{X} - \hat{\beta}(t_i - \bar{t})\right)^2$	$k-2$	$\frac{\mathrm{SAK}_2/f_2}{\mathrm{SAK}_1/f_1}$
Inden for grupper	$\mathrm{SAK}_1 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$	$N-k$	
Total	$\mathrm{SAK}_0 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$	$N-1$	

Her er

$$\begin{aligned} \mathrm{SAK}_{02} &= \mathrm{SAK}_1 + \mathrm{SAK}_2 \\ f_{02} &= f_1 + f_2 = N - 2, \end{aligned}$$

jfr. p. 436

Ved test på niveau α er det kritiske område for hypotesen
$H_{01}: \quad \mu_i = \alpha + \beta(t_i - \bar{t})$

$$C_1 = \left\{ (x_{11}, \cdots, x_{kn_k}) \middle| \frac{\mathrm{sak}_2/f_2}{\mathrm{sak}_1/f_1} > \mathrm{F}(k-2, N-k)_{1-\alpha} \right\}.$$

Hvis denne hypotese accepteres, kan vi teste $H_{02}: \beta = 0$. Det kritiske område ved test på niveau α er for denne hypotese

$$C_2 = \left\{ (x_{11}, \cdots, x_{kn_k}) \middle| \frac{\mathrm{sak}_3}{\mathrm{sak}_{02}/f_{02}} > \mathrm{F}(1, N-2)_{1-\alpha} \right\}.$$

Hvis $n_1 = \cdots = n_k = 1$, bliver $\mathrm{SAK}_1 = 0$, og vi kan da ikke teste H_{01}. Hvis vi imidlertid **forudsætter**, at H_{01} er gyldig, kan vi teste H_{02} på sædvanlig vis.

Hvis ingen af hypoteserne accepteres, anvender vi SAK_1/f_1 som skøn over σ^2. Accepteres H_{01}, anvendes SAK_{02}/f_{02}, og accepteres endelig H_{02}, anvendes SAK_0/f_0.

Det er naturligvis også muligt at teste andre værdier for β end 0 og at teste værdier for α. Af spaltningssætningen (Sætning 1.77, side 195) følger nemlig endvidere, at skønnene over α og β er uafhængige af SAK_{02}. Da $\mathrm{SAK}_{02} \in \sigma^2\chi^2(N-2)$, kan vi på sædvanlig vis danne t-teststørrelser. Vi giver nogle værdifulde resultater i

SÆTNING 5.10. Under de i sætning 5.9 anførte forudsætninger med tilføjelsen

$$\mathrm{E}(X_{ij}) = \mu_i = \alpha + \beta(t_i - \bar{t})$$

gælder

$$\frac{\hat{\alpha} - \alpha}{S_{02}/\sqrt{N}} \in \mathrm{t}(N-2)$$

$$\frac{\hat{\beta} - \beta}{S_{02}\sqrt{\frac{1}{\mathrm{SAK}_t}}} \in \mathrm{t}(N-2)$$

$$\frac{\hat{\alpha} + \hat{\beta}(t-\bar{t}) - \alpha - \beta(t-\bar{t})}{S_{02}\sqrt{\frac{1}{N} + \frac{(t-\bar{t})^2}{\mathrm{SAK}_t}}} \in \mathrm{t}(N-2)$$

Her er

$$S_{02}^2 = \frac{1}{N-2}\mathrm{SAK}_{02} = \frac{1}{N-2}(\mathrm{SAK}_1 + \mathrm{SAK}_2)$$

Bevis. Forbigås. ■

BEMÆRKNING 5.7. Såfremt man ønsker at teste for en bestemt værdi af f.eks. hældningskoefficienten $\beta = \beta_0$ skal man altså benytte teststørrelsen

$$Z = \frac{\hat{\beta} - \beta_0}{S_{02}\sqrt{\frac{1}{\mathrm{SAK}_t}}},$$

som skal sammenlignes med en $\mathrm{t}(N-2)$-fordeling, jfr. sætning 5.10.

Ønskes f.eks. et tosidet test på niveau α vil det kritiske område være givet ved

$$C = \left\{ z | z < \mathrm{t}(N-2)_{\alpha/2} \quad \vee \quad z > \mathrm{t}(N-2)_{1-\alpha/2} \right\}$$

Såfremt man ønsker et ensidet test på niveau α skal teststørrelsen Z sammenlignes med α- eller $1-\alpha$-fraktilen i en $\mathrm{t}(N-2)$-fordeling – afhængig af om testet skal forkaste for små eller store værdier af Z.

Ønsker man at teste, om regressionslinien givet ved (α, β) går gennem et bestemt punkt, $y = \alpha + \beta(t - \bar{t})$, benyttes den sidste af de tre teststørrelser.

Fremgangsmåden er dybest set analog til testene i sætning 5.10. Følgelig er det kritiske område identisk for de nævnte tre tests. ▼

Af hensyn til beregningerne anføres

$$\begin{aligned}
S_t &= \sum_{i=1}^{k} n_i t_i & S_{\bar{X}} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} \\
\mathrm{SK}_t &= \sum_{i=1}^{k} n_i t_i^2 & \mathrm{SK}_{\bar{X}} &= \sum_{i=1}^{k} \tfrac{1}{n_i}\left(\sum_{j=1}^{n_i} X_{ij}\right)^2 \\
\mathrm{SAK}_t &= \mathrm{SK}_t - \tfrac{1}{N} S_t^2 & \mathrm{SAK}_{\bar{X}} &= \mathrm{SK}_{\bar{X}} - \tfrac{1}{N} S_{\bar{X}}^2 \\
& & \mathrm{SP} &= \sum_{i=1}^{k} t_i \left(\sum_{j=1}^{n_i} X_{ij}\right) \\
& & \mathrm{SAP} &= \mathrm{SP} - \tfrac{1}{N} S_{\bar{X}} S_t
\end{aligned}$$

Med disse betegnelser bliver

$$\hat{\alpha} = \frac{1}{N} S_{\bar{X}} \qquad \text{og} \qquad \hat{\beta} = \frac{\mathrm{SAP}}{\mathrm{SAK}_t}$$

$$\begin{aligned}
\mathrm{SAK}_1 &= \sum_i \sum_j (X_{ij} - \bar{X}_i)^2 = \sum_i \sum_j X_{ij}^2 - SK_{\bar{X}} \\
\mathrm{SAK}_2 &= \sum_i n_i \left(\bar{X}_i - \bar{X} - \hat{\beta}(t_i - \bar{t})\right)^2 = \mathrm{SAK}_{\bar{X}} - \frac{\mathrm{SAP}^2}{\mathrm{SAK}_t} \\
\mathrm{SAK}_3 &= \hat{\beta}^2 \sum_i n_i (t_i - \bar{t})^2 = \frac{\mathrm{SAP}^2}{\mathrm{SAK}_t} \\
\mathrm{SAK}_0 &= \sum_i \sum_j (X_{ij} - \bar{X})^2 = \sum_i \sum_j X_{ij}^2 - \frac{1}{N} S_{\bar{X}}^2
\end{aligned}$$

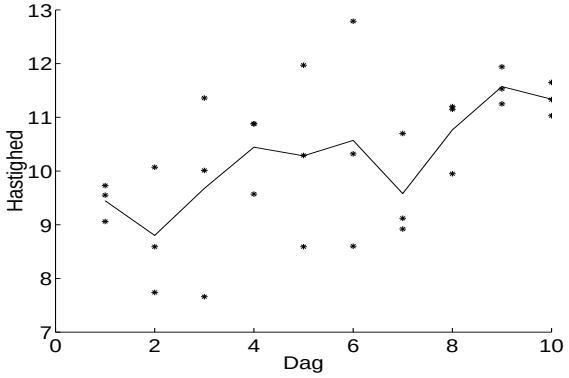
Man ser, at alle de størrelser, der alene vedrører X'er kan aflæses af det regneskema,

der er givet for den ensidede variansanalyse p. 418, hvorfor man starter med dette. Vi illustrerer nu regressionsanalyseteknikken ved hjælp af nogle eksempler

EKSEMPEL 5.6. Efter en periode med megen nedbør ønsker man at få belyst, om vandets hastighed i en bestemt elv varierer fra dag til dag. Man har derfor hver dag i 10 dage på samme sted foretaget tre uafhængige målinger af vandets hastighed. Resultatet af målingerne er samlet i tabellen nedenfor. Hastigheden er angivet i meter pr. sekund.

Dag	1	2	3	4	5	6	7	8	9	10
Hastig-	9.55	7.74	7.66	10.88	10.29	10.32	9.12	11.15	11.53	11.03
heder	9.06	10.07	10.01	9.57	11.97	8.60	10.70	9.95	11.25	11.33
	9.73	8.59	11.36	10.88	8.59	12.79	8.92	11.20	11.94	11.65

For at få et overblik over disse data, afbilder vi dem i et koordinatsystem med vandhastigheden som funktion af tiden



Den fuldt optrukne linie forbinder gennemsnitshastigheden de 10 dage. Der synes at være en lineær stigning i vandhastigheden. Det må a priori anses for ret sandsynligt, at vandhastigheden vil stige med tiden efter perioden med stærk nedbør (inden for et passende tidsinterval), og vi vil nu ved hjælp af en regressionsanalysemodel undersøge, om denne stigning kan beskrives ved en ret linie.

Vi forudsætter derfor, at de målte observationer kan opfattes som realiserede udfald af uafhængige, normalt fordelte stokastiske variable X_{ij}, $i = 1, \cdots, 10$, $j = 1, \cdots, 3$, der har samme varians σ^2, og middelværdier
$\mathrm{E}(X_{ij}) = \mu_i$.

Vi vil nu teste

$$H_{01}: \quad \mu_i = \alpha + \beta(t_i - \bar{t})$$

mod alle alternativer. Her er $t_i = i$, dvs. lig antallet af dage, der er gået fra forsøget

startede. Hvis denne hypotese accepteres, vil vi undersøge, om linien kan antages at være parallel med abscisseaksen, i.e. teste

$$H_{02}: \quad \beta = 0 \qquad \text{mod} \qquad H_{12}: \quad \beta \neq 0.$$

Vi anfører nu det regneskema, der svarer til den ensidede variansanalyse:

$i = t_i$	Observationer -10			n_i	S_i	SK_i	S_i^2/n_i	SAK_{1i}	f_i
1	-0.45	-0.94	-0.27	3	-1.66	1.1590	0.91853	0.24047	2
2	-2.26	0.07	-1.41	3	-3.60	7.1006	4.32000	2.78060	2
3	-2.34	0.01	1.36	3	-0.97	7.3253	0.31363	7.01167	2
4	0.88	-0.43	0.88	3	1.33	1.7337	0.58963	1.14407	2
5	0.29	1.97	-1.41	3	0.85	5.9531	0.24083	5.71227	2
6	0.32	-1.40	2.79	3	1.71	9.8465	0.97470	8.87180	2
7	-0.78	0.70	-1.08	3	-1.16	2.2648	0.44853	1.81627	2
8	1.15	-0.05	1.20	3	2.30	2.7650	1.76300	1.00200	2
9	1.53	1.25	1.94	3	4.72	7.6670	7.42610	0.24090	2
10	1.03	1.33	1.65	3	4.01	5.5523	5.36003	0.19227	2
	Total			30	7.53	51.3673	22.35498	29.01232	20

Dernæst beregnes de p. 458 anførte størrelser.

$$\begin{array}{rcl} s_t & = & \sum n_i t_i \;=\; 3\sum i \;=\; 165 \\ \mathrm{sk}_t & = & \sum n_i t_i^2 \;=\; 3\sum i^2 \;=\; 1155 \\ \mathrm{sak}_t & = & \mathrm{sk}_t - s_t^2/N \;=\; 247.5 \end{array}$$

$$\begin{array}{rcl} s_{\bar{x}} & = & \sum s_i \;=\; 7.53 \\ \mathrm{sk}_{\bar{x}} & = & \sum s_i^2/n_i \;=\; 22.35498 \\ \mathrm{sak}_{\bar{x}} & = & \mathrm{sk}_{\bar{x}} - s_{\bar{x}}^2/N \;=\; 20.46495 \end{array}$$

$$\begin{array}{rcl} \mathrm{sp} & = & \sum t_i s_i \;=\; 100.92 \\ \mathrm{sap} & = & \mathrm{sp} - s_{\bar{x}} s_t/N \;=\; 59.505 \\ \hat{\alpha} & = & s_{\bar{x}}/N \;=\; 0.251 \\ \hat{\beta} & = & \mathrm{sap}/\mathrm{sak}_t \;=\; 0.2404 \end{array}$$

Ved hjælp af disse størrelser er det ikke vanskeligt at beregne kvadratafvigelsessum-

merne i variansanalyseskemaet p. 456.

$$\begin{aligned}
\mathrm{sak}_1 &= \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 &&= \sum_i \mathrm{sak}_{1i} &&= 29.01232 \\
\mathrm{sak}_2 &= \sum_i n_i(\bar{x}_i - \bar{x} - \hat{\beta}(t_i - \bar{t}))^2 &&= \mathrm{sak}_{\bar{x}} - \tfrac{\mathrm{sap}^2}{\mathrm{sak}_t} &&= 6.15851 \\
\mathrm{sak}_3 &= \hat{\beta}^2 \sum_i n_i(t_i - \bar{t})^2 &&= \tfrac{\mathrm{sap}^2}{\mathrm{sak}_t} &&= 14.30644 \\
\mathrm{sak}_0 &= \sum_i \sum_j (x_{ij} - \bar{x})^2 &&= \sum_i \mathrm{sk}_i - (\sum_i s_i)^2/N &&= 49.47727 \\
s_{02}^2 &= (\mathrm{sak}_1 + \mathrm{sak}_2)/(N-2) &&= &&1.256
\end{aligned}$$

Vi samler disse resultater i variansanalyseskemaet.

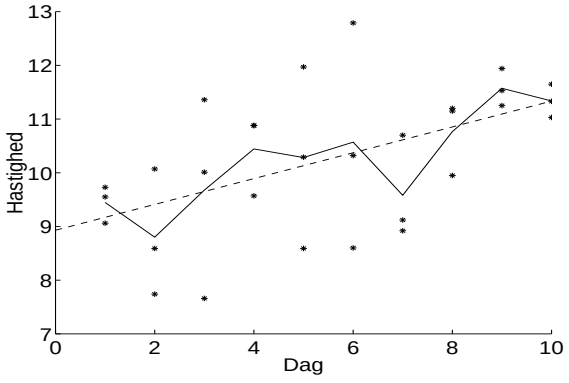
Variation	SAK	f	S^2	Test
Regressionslinien	14.30644	1	14.30644	11.4
Om regressionslinie	6.15851	8	0.7698	0.53
Inden for dage	29.01232	20	1.4506	
Total	49.47727	29		

De kritiske værdier er $\mathrm{F}(8,20)_{0.95} = 2.45$ og $\mathrm{F}(1,28)_{0.95} = 4.20$. Nu er $0.53 < \mathrm{F}(8,20)_{0.50}$ og $11.4 > \mathrm{F}(1,28)_{0.995}$, hvorfor vi vil acceptere hypotesen om linearitet på alle niveauer mindre end 50% og forkaste hypotesen om, at hældningen af regressionslinien er 0 på alle niveauer større end 0.5%.

Vi vil derfor konkludere, at middelvandhastigheden kan beskrives ved den affine funktion (rette linie)

$$\begin{aligned}
\hat{\mu}(t) &= 0.251 + 0.2404(t - 5.5) + 10 \\
&= 8.93 + 0.2404t
\end{aligned}$$

Vi indtegner denne linie i den tidligere figur og får



Til slut nævner vi, at vi selvfølgelig også kunne have forsøgt at løse spørgsmålet ved at anvende en ensidet variansanalysemodel, d.v.s. ved at undersøge, om middelvandstandene kan antages at være ens de 10 dage. Vi får da variansanalyseskemaet

Variation	SAK	f	S^2	Test
Mellem dage	20.46495	9	2.2739	1.57
Inden for dage	29.01232	20	1.4506	
Total	49.47727	29		

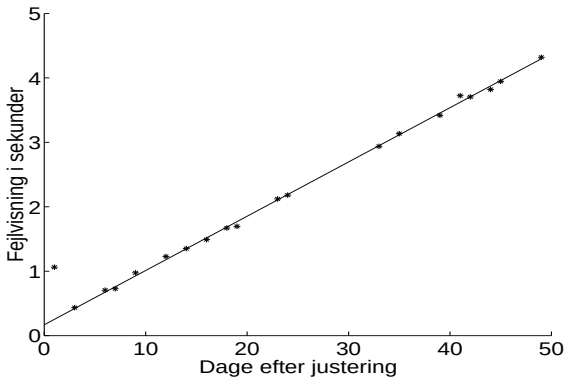
Den kritiske værdi er $\mathrm{F}(9,20)_{0.95} = 2.30$. Da faktisk $1.57 < \mathrm{F}(9,20)_{0.90}$, vil vi acceptere hypotesen om fuldstændig homogenitet på alle niveauer mindre end 10%. Nu svarer hypotesen om fuldstændig homogenitet jo åbenbart til hypotesen $\beta = 0$ i regressionsanalysen, og denne fik vi jo forkastet.

Forklaringen på denne tilsyneladende modstrid er følgende. Variationen mellem dage er lig summen af variationen omkring regressionslinien og regressionsliniens variation. Nu er variationen om regressionslinien meget lille i forhold til antallet af frihedsgrader og regressionsliniens variation meget stor, således at vi får en overbevisende accept af hypotesen om linearitet og en endnu mere overbevisende forkastelse af hypotesen om hældningen 0. Da denne sidste teststørrelse kun har 1 frihedsgrad, får den ikke så stor vægt ved gennemsnitsdannelsen, således at variansanalysehypotesen accepteres. Grunden til uoverensstemmelsen er altså, at vi i variansanalysen ikke skelner mellem de enkelte dages variation omkring regressionslinien og regressionsliniens egen variation (hældning). ♦

EKSEMPEL 5.7. I nedenstående tabel er der opført fejlvisningen for et kvartsur beregnet ud fra passageobservationer foretaget fra 3 til 49 dage efter, at uret var blevet justeret.

Antal dage efter justering	Fejlvisning i sek.	Antal dage efter justering	Fejlvisning i sek.
3	0.435	23	2.122
6	0.706	24	2.181
7	0.729	33	2.938
9	0.975	35	3.135
11	1.063	39	3.419
12	1.228	41	3.724
14	1.342	42	3.705
16	1.491	44	3.820
18	1.671	45	3.945
19	1.696	49	4.320

Vi vil nu vurdere den foretagne justering. Vi afbilder først ovenstående data i et koordinatsystem.



Ved justeringen kan man tænke sig 2 primære fejlkilder. Dels kan nulstillingen være forkert, og dels kan hastighedsjusteringen være forkert således, at fejlvisningen øges med et konstant bidrag hver dag. Hvis der ikke er andre fejlkilder, har vi derfor, at middelfejlvisningen den i'te dag er

$$\mu(t_i) = \alpha + \beta(t_i - \bar{t})\,. \tag{5.8}$$

Vi kan nu undersøge, om **nulpunktsindstillingen har været korrekt ved at undersøge, om linien (5.8) går gennem** $(0,0)$**. Hvis hastighedsjusteringen har været i orden, vil** $\beta = 0$ **i (5.8)**.

Vi er interesserede i at teste disse hypoteser ved hjælp af en lineær regressionsanalysemodel. Vi ser, at der kun foreligger et X_i til hvert tidspunkt. Vi kan derfor ikke teste hypotesen om linearitet, men må forudsætte den. Et blik på figuren p. 462 viser, at linearitetshypotesen ser rimelig ud. Vi forudsætter endvidere, at aflæsningsnøjagtigheden er konstant, så vi kan antage, at de stokastiske variable har samme varians.

Af figuren fremgår i øvrigt, at der ikke er meget håb om at få accepteret en hypotese $\beta = 0$. Vi anfører dog alligevel den fuldstændige analyse.

Vi har $N = k = 20$ observationer. De nødvendige summer, kvadratsummer og produktsummer fremgår af nedenstående skema.

	t	x
s	490	44.645
sk	16324	130.322683
s^2/k	12005	99.658801
sak	4319	30.663882
sp_{xt}	1457.543	
$s_x s_t/k$	1093.803	
sap	363.740	

Ved hjælp af disse størrelser finder vi estimaterne

$$\begin{array}{rclcrcl} \hat{\alpha} & = & \frac{1}{k}\Sigma x_i & = & s_x/k & = & 2.2323 \\ \hat{\beta} & = & \frac{\Sigma(x_i-\bar{x})(t_i-\bar{t})}{\Sigma(t_i-\bar{t})^2} & = & \frac{\mathrm{sap}}{\mathrm{sak}_t} & = & 0.08422 \end{array}$$

Endvidere finder vi kvadratsummerne

$$\begin{array}{rclcrcr} \mathrm{sak}_2 & = & \sum\limits_i \big(x_i - \bar{x} - \hat{\beta}(t_i-\bar{t})\big)^2 & = & \mathrm{sak}_x - \frac{\mathrm{sap}^2}{\mathrm{sak}_t} & = & 0.030220 \\ \mathrm{sak}_3 & = & \hat{\beta}^2 \sum\limits_i (t_i-\hat{t})^2 & = & \frac{\mathrm{sap}^2}{\mathrm{sak}_t} & = & 30.633662 \\ \mathrm{sak}_0 & = & \mathrm{sak}_x & & & = & 30.663882 \end{array}$$

Vi får herved variansanalyseskemaet

Variation	SAK	f	Test
Regressionslinien	30.633662	1	18246
Omkring regressionslinien	0.030220	18	
Total	30.663882	19	

Da $\mathrm{F}(1,18)_{0.999} = 15.38$, ser vi, at vi forkaster hypotesen om hældningen $\beta = 0$ endog særdeles kraftigt.

At undersøge, om linien $\mu(t) = \alpha + \beta(t-\bar{t})$ går gennem $(0,0)$ svarer til at undersøge, om $\alpha - \beta\bar{t} = 0$. Vi vil derfor nu teste

$$H_0' : \quad \alpha - \beta\bar{t} = 0 \qquad \text{mod} \qquad H_1' : \quad \alpha - \beta\bar{t} \neq 0.$$

Ifølge sætningen p. 457 er

$$T = \frac{\hat{\alpha} - \hat{\beta}\bar{t}}{\sqrt{\frac{1}{k-2}\mathrm{SAK}_2}\sqrt{\frac{1}{N} + \frac{\bar{t}^2}{\mathrm{SAK}_t}}} \in \mathrm{t}(k-2)$$

under H_0' (bemærk, at $k = N$). Dette giver mulighed for at danne et sædvanligt t-test. Vi får, at den observerede værdi af T er

$$t = \frac{2.2323 - 0.08422 \cdot 24.5}{\sqrt{\frac{1}{18}0.030220}\sqrt{\frac{1}{20} + \frac{600.25}{4319}}} = 9.48.$$

Det kritiske område ved test på niveau α af H_0' mod H_1' er

$$C' = \left\{(x_1, \cdots, x_{20}) | \mathrm{t} < \mathrm{t}(18)_{\alpha/2} \vee \mathrm{t} > \mathrm{t}(18)_{1-\alpha/2}\right\},$$

Ved test på 5% signifikansniveau er de kritiske værdier $\mathrm{t}(18)_{0.95}$ og $-\mathrm{t}(18)_{0.95}$, i.e. -1.73 og 1.73 henholdsvis. Da $9.48 > \mathrm{t}(18)_{0.9995} = 3.922$, vil vi forkaste H_0' på alle niveauer større end 0.1%.

Sammenfattende kan det derfor siges, at såvel justering af hastighed som af nulstilling har været unøjagtige. Estimatet for fejlvisningen t_i dage efter justeringen er (målt i sekunder)

$$\begin{aligned} \hat{\mu}(t_i) &= 2.2323 + 0.08422(t_i - 24.5) \\ &= 0.1689 + 0.08422 t_i. \end{aligned}$$

♦

5.2.2 Sammenligning af 2 regressionslinier

Vi antager, at vi har givet 2 empiriske regressionslinier, og vi vil undersøge, om de tilsvarende teoretiske linier

1. kan antages at være parallelle
2. kan antages at være ens.

Der er altså givet uafhængige, normalt fordelte stokastiske variable

$$X_{ij}, \qquad i = 1, \cdots, k, \quad j = 1, \cdots, n_i,$$

svarende til værdierne $t_1, \cdots, t_k$, og givet uafhængige, normalt fordelte stokastiske variable

$$Y_{ij}, \qquad i = 1, \cdots, h, \quad j = 1, \cdots, m_i,$$

svarende til $v_1, \cdots, v_h$. Vi forudsætter, at

1.
$$\begin{array}{rcll} \mathrm{E}(X_{ij}) & = & \alpha_x + \beta_x(t_i - \bar{t}) & \forall i,j \\ \mathrm{E}(Y_{ij}) & = & \alpha_y + \beta_y(v_i - \bar{v}) & \forall i,j \end{array} \tag{5.9}$$

2.
$$\begin{array}{rcll} \mathrm{V}(X_{ij}) & = & \sigma^2 & \forall i,j \\ \mathrm{V}(Y_{ij}) & = & \sigma^2 & \forall i,j. \end{array} \tag{5.10}$$

Idet vi sætter $N = \Sigma n_i$ og $M = \Sigma m_i$, har vi ifølge forrige afsnit estimatorerne

$$\begin{array}{rcll} \hat{\alpha}_x & = & \frac{1}{N}\sum\sum_{i\ j} X_{ij} & = \bar{X} \in \mathrm{N}(\alpha_x, \frac{\sigma^2}{N}) \\ \hat{\alpha}_y & = & \frac{1}{M}\sum\sum_{i\ j} Y_{ij} & = \bar{Y} \in \mathrm{N}(\alpha_y, \frac{\sigma^2}{M}) \\ \hat{\beta}_x & = & \frac{\sum\sum(X_{ij}-\bar{X})(t_i-\bar{t})}{\sum n_i(t_i-\bar{t})^2} & = \frac{\mathrm{SAP}_{xt}}{\mathrm{SAK}_t} \in \mathrm{N}(\beta_x, \frac{\sigma^2}{\mathrm{SAK}_t}) \\ \hat{\beta}_y & = & \frac{\sum\sum(Y_{ij}-\bar{Y})(v_i-\bar{v})}{\sum n_i(v_i-\bar{v})^2} & = \frac{\mathrm{SAP}_{yv}}{\mathrm{SAK}_v} \in \mathrm{N}(\beta_y, \frac{\sigma^2}{\mathrm{SAK}_v}) \end{array}$$

Vi har 2 skøn over σ^2, nemlig

$$\begin{array}{rcl} S_x^2 & = & \dfrac{1}{N-2}\mathrm{SAK}_{x02} \in \sigma^2\chi^2(N-2)/(N-2) \\ S_y^2 & = & \dfrac{1}{M-2}\mathrm{SAK}_{y02} \in \sigma^2\chi^2(M-2)/(M-2). \end{array}$$

Disse samles til et fælles skøn (eventuelt testes $\sigma_x^2 = \sigma_y^2$ først):

$$S^2 = \frac{(N-2)S_x^2 + (M-2)S_y^2}{N+M-4} \in \sigma^2 \chi^2 (N+M-4)/(N+M-4).$$

Ifølge bemærkningerne p. 463 er disse skøn endvidere stokastisk uafhængige, således at vi på sædvanlig vis kan danne t-teststørrelser.

Vi er først interesserede i at undersøge hypotesen

$$H_{01}: \quad \beta_x = \beta_y \qquad \text{mod} \qquad H_{11}: \quad \beta_x \neq \beta_y.$$

Ifølge de ovenstående betragtninger er

$$\hat{\beta}_x - \hat{\beta}_y \in \mathrm{N}\left(\beta_x - \beta_y, \sigma^2\left(\frac{1}{\mathrm{SAK}_t} + \frac{1}{\mathrm{SAK}_v}\right)\right),$$

således at

$$Z_1 = \frac{\hat{\beta}_x - \hat{\beta}_y}{S\sqrt{\frac{1}{\mathrm{SAK}_t} + \frac{1}{\mathrm{SAK}_v}}} \in \mathrm{t}(M+N-4)$$

under H_{01}. Vi kan derfor danne et test på niveau α ved at anvende det kritiske område

$$\begin{aligned} C_1 = \big\{ (x_{11}, \cdots, y_{hm_h}) | z_1 < \mathrm{t}(M+N-4)_{\alpha/2} \\ \vee \quad z_1 > \mathrm{t}(M+N-4)_{1-\alpha/2} \big\}. \end{aligned}$$

Hvis H_{01} accepteres antages, at $\hat{\beta}_x$ og $\hat{\beta}_y$ er estimatorer for den samme (ukendte) hældningskoefficient $\beta = \beta_x = \beta_y$.

Vi kan derfor danne et bedre skøn over β (i.e. et skøn med mindre varians), nemlig

$$\hat{\beta} = \frac{\mathrm{SAK}_t \hat{\beta}_x + \mathrm{SAK}_v \hat{\beta}_y}{\mathrm{SAK}_t + \mathrm{SAK}_v} = \frac{\mathrm{SAP}_{xt} + \mathrm{SAP}_{yv}}{\mathrm{SAK}_t + \mathrm{SAK}_v}.$$

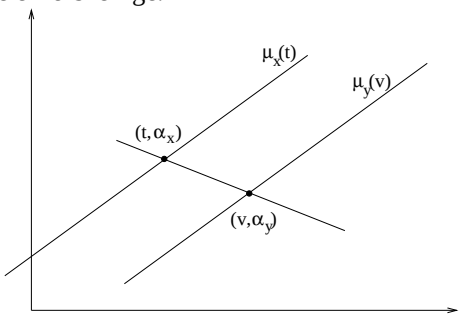
Vi bemærker, at $\hat{\beta}$ er et vejet gennemsnit af $\hat{\beta}_x$ og $\hat{\beta}_y$ med de reciprokke varianser som vægte. Det ses, at

$$\hat{\beta} \in \mathrm{N}\left(\beta, \frac{\sigma^2}{\mathrm{SAK}_t + \mathrm{SAK}_v}\right).$$

Hvis vi har fået accepteret H_{01}, vil vi dernæst interessere os for hypotesen

H_{02}: de to linier er identiske

mod alternativet, at de er forskellige.



Da linierne har samme hældning, er de ens, hvis deres konstantled er lig hinanden, i.e. såfremt

$$\alpha_x - \beta\bar{t} = \alpha_y - \beta\bar{v}$$

eller

$$\beta = \frac{\alpha_x - \alpha_y}{\bar{t} - \bar{v}}.$$

Dette udtrykker blot, at linierne er ens, hvis den linie ℓ, der forbinder tyngdepunkterne $(\bar{t}, \alpha_x)$ og $(\bar{v}, \alpha_y)$ har samme hældning β som de oprindelige linier, jfr. vedstående figur. Som skøn over $\beta = (\alpha_x - \alpha_y)/(\bar{t} - \bar{v})$ kan derfor anvendes

$$\tilde{\beta} = \frac{\bar{X} - \bar{Y}}{\bar{t} - \bar{v}},$$

og man finder, at

$$\tilde{\beta} \in \mathrm{N}\left(\beta, \frac{\sigma^2}{(\bar{t} - \bar{v})^2}\left(\frac{1}{N} + \frac{1}{M}\right)\right)$$

under H_{02}. Af Sætning 1.77, side 195, følger, at $\hat{\beta}$ og $\tilde{\beta}$ er stokastisk uafhængige og uafhængig af S^2. Derfor er

$$\tilde{\beta} - \hat{\beta} \in \mathrm{N}\left(0, \sigma^2\left(\frac{1}{(\bar{t} - \bar{v})^2}\left(\frac{1}{N} + \frac{1}{M}\right) + \frac{1}{\mathrm{SAK}_t + \mathrm{SAK}_v}\right)\right)$$

og

$$\begin{aligned} Z_2 &= (\tilde{\beta}-\hat{\beta})/\left(S\left[\frac{1}{(\bar{t}-\bar{v})^2}\left(\frac{1}{N}+\frac{1}{M}\right)+\frac{1}{\mathrm{SAK}_t+\mathrm{SAK}_v}\right]^{\frac{1}{2}}\right) \\ &\in \mathrm{t}(N+M-4) \end{aligned}$$

under H_{02}. Som kritisk område for et test på niveau α kan vi derfor anvende

$$\begin{aligned} C_2 = \{ &(x_{11},\cdots,y_{hm_h})|z_2 < \mathrm{t}(M+N-4)_{\alpha/2} \\ &\vee \quad z_2 > \mathrm{t}(M+N-4)_{1-\alpha/2}\} . \end{aligned}$$

Vi bemærker, at den anvendte teknik kan udvides til at omfatte sammenligning af m regressionslinier. I litteraturen omtales sådanne analyser ofte som **kovariansanalyser**.

Det skal sluttelig anføres, at hvis vi forkaster et test for $\sigma_x^2 = \sigma_y^2$, kan problemet ikke løses eksakt. Man kan finde en approksimativ løsning, men dette skal vi ikke komme ind på her. Vi henviser blot til [25, p. 573].

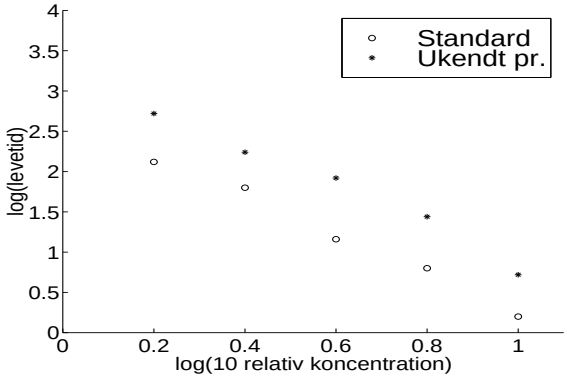
Vi anfører nu et illustrativt

EKSEMPEL 5.8. På et laboratorium ønsker man at sammenligne et giftstof med en kendt standard for at få et indtryk af det ukendte giftstofs styrke. Man anvender følgende fremgangsmåde. Man fortynder giftstofferne med vand, så man opnår forskellige (relative) koncentrationer. Derefter indsprøjter man giftstofferne i laboratoriemus og noterer, hvor længe musene lever efter indsprøjtningen.

Ved undersøgelsen fik man følgende sammenhæng mellem (10-tals-) logaritmen til overlevelsestiden i minutter og logaritmen til 10 gange den relative koncentration. (Erfaringen har vist, at det er hensigtsmæssigt at logaritmere måleresultaterne. Det sikrer, at forudsætningerne for en lineær regressionsanalyse (tilnærmelsvis) er til stede).

$\log_{10}$ (10·relativ koncentration)	$\log_{10}$ (overlevelsestid)	
	Standard	Ukendt prøve
0.20	2.12	2.72
0.40	1.80	2.24
0.60	1.16	1.92
0.80	0.80	1.44
1.00	0.20	0.72

Vi indtegner disse data i et koordinatsystem og får følgende figur



Det ser ud til, at vi kan antage linearitetshypotesen for begge analyser. Vi udfører derfor en lineær regressionsanalyse med henblik på en senere sammenligning af regressionslinierne. Vi forudsætter nu - og erfaringen fra lignende forsøg underbygger dette - at ovenstående observationer kan opfattes som realisationer af uafhængige normalt fordelte stokastiske variable, der tilfredsstiller linearitetsforudsætningerne (5.9).

Vi anfører beregningerne i følgende skema

	$t=v$	y	x
Værdier	0.20	2.12	2.72
	0.40	1.80	2.24
	0.60	1.16	1.92
	0.80	0.88	1.44
	1.00	0.20	0.72
s	3.00	6.16	9.04
sk	2.20	9.8944	18.6944
sp		2.744	4.464
sak	0.40	2.3053	2.3501
sap		-0.952	-0.960

Ved hjælp af disse værdier finder vi

$$\begin{aligned} \mathrm{sak}_{2x} &= 2.3501 - \tfrac{0.960^2}{0.40} = 0.0461 \\ \mathrm{sak}_{2y} &= 2.3053 - \tfrac{0.952^2}{0.40} = 0.0395, \end{aligned}$$

og vi får da let følgende estimatorer

$$\begin{array}{rcccl}
\hat{\alpha}_x & = & \frac{9.04}{5} & = & 1.808 \\
\hat{\beta}_x & = & -\frac{0.960}{0.40} & = & -2.400 \\
\hat{\sigma}_x^2 & = & \frac{0.0461}{5-2} & = & 0.0154 \\
\hat{\alpha}_y & = & \frac{6.16}{5} & = & 1.232 \\
\hat{\beta}_y & = & -\frac{0.952}{0.40} & = & -2.380 \\
\hat{\sigma}_y^2 & = & \frac{0.0395}{5-2} & = & 0.0132.
\end{array}$$

Vi undersøger nu, om det kan antages, at $\sigma_x^2 = \sigma_y^2$. Da

$$\mathrm{F}(3,3)_{0.50} < \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} = 1.16 < \mathrm{F}(3,3)_{0.70},$$

vil vi acceptere hypotesen $\sigma_x^2 = \sigma_y^2$ på alle niveauer mindre end 30%, hvorfor vi vil arbejde med denne antagelse i det følgende. Som et forbedret skøn over σ^2 har vi da

$$s^2 = \frac{1}{6}(3\hat{\sigma}_x^2 + 3\hat{\sigma}_y^2) = 0.014266 = 0.119^2.$$

Vi vil nu undersøge, om hældningerne kan antages at være ens. Den observerede værdi af teststørrelsen er

$$z_1 = \frac{\hat{\beta}_x - \hat{\beta}_y}{s\sqrt{\frac{1}{\mathrm{sak}_t} + \frac{1}{\mathrm{sak}_t}}} = \frac{-2.40 + 2.38}{0.119\sqrt{\frac{2}{0.40}}} \simeq -0.075.$$

Da

$$-0.265 = \mathrm{t}(6)_{0.40} < z_1 < \mathrm{t}(6)_{0.50} = 0,$$

vil vi acceptere antagelsen om, at linierne har samme hældning på alle niveauer mindre end 80%. Som skøn over den fælles hældning har vi

$$\hat{\beta} = \frac{-0.960 - 0.952}{0.40 + 0.40} = -2.39.$$

Vi skal nu undersøge, om linierne kan antages at være ens. Da $\bar{t} = \bar{v}$ (idet $t_i = v_i \quad \forall i$), kan vi **ikke** anvende den p. 468 beskrevne procedure. Vi har imidlertid, at

$$\alpha_x - \beta\bar{t} = \alpha_y - \beta\bar{t} \quad \Leftrightarrow \quad \alpha_x = \alpha_y\,.$$

Da

$$\begin{aligned}\hat{\alpha}_x &\in \mathrm{N}\left(\alpha_x, \frac{\sigma^2}{N}\right),\\ \hat{\alpha}_y &\in \mathrm{N}\left(\alpha_y, \frac{\sigma^2}{N}\right),\end{aligned}$$

og

$$S^2 \in \sigma^2\chi^2(2N-4)/(2N-4),$$

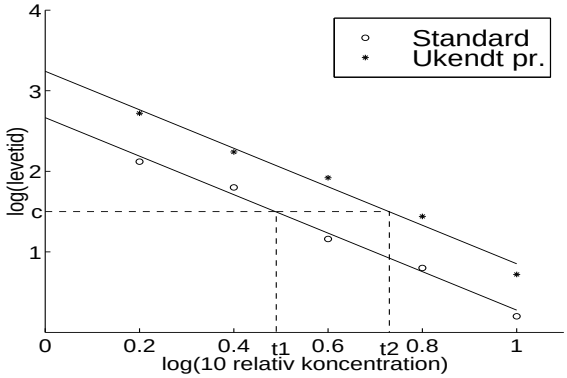
er stokastisk uafhængige (jfr. p. 457), er

$$\frac{\hat{\alpha}_x - \hat{\alpha}_y}{S\sqrt{\frac{1}{N}+\frac{1}{N}}} \in \mathrm{t}(6),$$

under antagelsen $\alpha_x = \alpha_y$. Da

$$\frac{\hat{\alpha}_x - \hat{\alpha}_y}{s\sqrt{\frac{1}{5}+\frac{1}{5}}} = \frac{1.808 - 1.232}{0.119\sqrt{\frac{2}{5}}} \simeq 10.8 > \mathrm{t}(6)_{0.9995},$$

vil vi forkaste hypotesen $\alpha_x = \alpha_y$. Vi må altså konkludere, at de 2 regressionslinier er parallelle, men forskellige. I nedenstående figur har vi indtegnet de parallelle regressionslinier.



Vi vil nu opsummere de resultater, vi har fundet. Idet vi sætter logaritmen til overlevelsestiden i minutter lig X for det ukendte giftstof og lig Y for den kendte standard

og logaritmen til 10 gange den relative koncentration lig t, har vi fundet

$$\begin{array}{rcccccl} \mathrm{E}(X_t) & \simeq & \hat{\mu}_x(t) & = & \hat{\alpha}_x - \hat{\beta}(t-\bar{t}) & = & 1.808 - 2.39(t-0.6) \\ \mathrm{E}(Y_t) & \simeq & \hat{\mu}_y(t) & = & \hat{\alpha}_y - \hat{\beta}(t-\bar{t}) & = & 1.232 - 2.39(t-0.6). \end{array}$$

Løser vi nu ligningen

$$c = \hat{\mu}_x(t_1) = \hat{\mu}_y(t_2),$$

fås

$$\hat{\alpha}_x - \hat{\beta}(t_1 - \bar{t}) = \hat{\alpha}_y - \hat{\beta}(t_2 - \bar{t})$$

eller

$$t_1 - t_2 = \frac{\hat{\alpha}_x - \hat{\alpha}_y}{\hat{\beta}} = \frac{0.576}{2.39} = 0.241.$$

Vi har altså fået, at giftene har samme styrke (d.v.s. samme overlevelsestid for musene), hvis forskellen mellem $\log_{10}$ (10 gange koncentrationerne) er (konstant lig) 0.241. Dette svarer til, at den ukendte gift skal anvendes i en koncentration, der er

$$10^{0.241} = 1.74$$

gange større end standarden for at have samme virkning som denne. Vi kan også sige, at den ukendte gift har den relative styrke $1/1.74 = 0.57$.

Til sidst bemærker vi, at parallelliteten af regressionslinierne er en forudsætning for, at man kan tale om den relative styrke. Hvis regressionslinierne ikke er parallelle, er de 2 giftstoffer ikke direkte sammenlignelige. ♦

Hermed slutter vi behandlingen af problemet med sammenligning af regressionslinier og går over til at betragte

5.2.3 Regressionsanalyse med 2 uafhængige variable

I regressionsanalysen benævnes størrelserne $t_1, \cdots, t_k$ de **uafhængige variable** (hvor forveksling med begrebet uafhængige stokastiske variable naturligvis er udelukket). I dette afsnit vil vi generalisere modellen fra afsnit 5.2.1 med 1 observation for hver t-værdi til følgende.

Der er givet stokastiske variable

$$X_i, \qquad i = 1, \cdots, k$$

svarende til de 2 "uafhængige" variable t_{i1} og t_{i2}, i.e.

$$t_{ij}, \qquad i = 1, \cdots, k, \quad j = 1, 2.$$

Vi antager, at X_i'erne

1. er stokastisk uafhængige

2. er normalt fordelte

3. har samme varians σ^2.

Endvidere forudsætter vi, at der er givet nogle størrelser α, β_1 og β_2 således, at middelværdien af de stokastiske variable X_i er

$$\mathrm{E}(X_i) = \mu_i = \alpha + \beta_1(t_{i1} - \bar{t}_1) + \beta_2(t_{i2} - \bar{t}_2), \tag{5.11}$$

hvor

$$\bar{t}_j = \frac{1}{k}\sum_{i=1}^{k} t_{ij} \qquad j = 1, 2.$$

Sætter vi

$$\begin{aligned}
\mathrm{SAK}_{t_j} &= \sum_{i=1}^{k}(t_{ij}-\bar{t}_j)^2 = \sum_{i=1}^{k} t_{ij}^2 - \frac{1}{k}(\sum_{i=1}^{k} t_{ij})^2 \qquad j=1,2 \\
\mathrm{SAP}_{t_1t_2} &= \sum_{i=1}^{k}(t_{i1}-\bar{t}_1)(t_{i2}-\bar{t}_2) \\
&= \sum_{i=1}^{k} t_{i1}t_{i2} - \frac{1}{k}(\sum_{i=1}^{k} t_{i1})(\sum_{i=1}^{k} t_{i2}) \\
\mathrm{SAP}_{t_jx} &= \sum_{i=1}^{k}(t_{ij}-\bar{t}_j)(X_i-\bar{X}) \\
&= \sum_{i=1}^{k} t_{ij}X_i - \frac{1}{k}(\sum_{i=1}^{k} t_{ij})(\sum_{i=1}^{k} X_i), \quad j=1,2,
\end{aligned}$$

har vi følgende analog til sætning 5.9.

SÆTNING 5.11. Maximum likelihood estimatorerne for parametrene i relationen (5.11) er givet ved

$$\hat{\alpha} = \bar{X} = \frac{1}{k}\sum_{i=1}^{k} X_i$$

samt ligningssystemet

$$\begin{array}{lcll}
\hat{\beta}_1\mathrm{SAK}_{t_1} & + & \hat{\beta}_2\mathrm{SAP}_{t_1t_2} & = \mathrm{SAP}_{t_1x} \\
\hat{\beta}_1\mathrm{SAP}_{t_1t_2} & + & \hat{\beta}_2\mathrm{SAK}_{t_2} & = \mathrm{SAP}_{t_2x}.
\end{array} \tag{5.12}$$

Bevis. Sætningen følger ved direkte regning. ■

BEMÆRKNING 5.8. Ligningssystemet (5.12) som kaldes **normalligningerne**, kan ikke altid løses. Der er netop 1 løsning, hvis determinanten

$$D_{t_1t_2} = \mathrm{SAK}_{t_1}\mathrm{SAK}_{t_2} - (\mathrm{SAP}_{t_1t_2})^2$$

er forskellig fra 0. ▼

Hvis man er interesseret i at undersøge, om $\beta_1 = \beta_2 = 0$, kan man ved en spaltning af den totale variation komme frem til følgende variansanalyseskema

Variation	SAK	f	Test
Regressions-planen	$\mathrm{SAK}_3 = \hat{\beta}_1 \mathrm{SAP}_{t_1 x} + \hat{\beta}_2 \mathrm{SAP}_{t_2 x}$	2	$\frac{\mathrm{SAK}_3/2}{\mathrm{SAK}_2/(k-3)}$
Omkring regressions-planen	$\mathrm{SAK}_2 = \sum_{i=1}^{k} \Big(X_i - \bar{X} - \hat{\beta}_1 (t_{i1} - \bar{t}_1) - \hat{\beta}_2 (t_{i2} - \bar{t}_2) \Big)^2$	$k-3$	
Total	$\mathrm{SAK}_0 = \sum_{i=1}^{k} (X_i - \bar{X})^2$	$k-1$	

Her har vi anvendt betegnelserne SAK_2 og SAK_3 for at bevare analogien med det endimensionale tilfælde (jfr. skemaet p. 456). Som skøn over σ^2 anvendes $\mathrm{SAK}_2/(k-3)$.

Vi skal ikke gå i detaljer med den multiple regressionsanalyse, men blot konstatere, at resultater og metoder fra den endimensionale analyse i vid udstrækning kan overføres.

Vi slår nogle af begreberne fast ved hjælp af

EKSEMPEL 5.9. I nedenstående tabel er udbyttet (i % af det maximalt opnåelige) af en proces anført som funktion af reaktionstemperaturen (i °F) og mængden af en katalysator (målt i % af hovedreaktanten)

Katalysator	Temperatur	Udbytte
0.05	200	80.9
0.05	250	67.3
0.05	300	60.9
0.05	350	60.6
0.10	200	85.9
0.10	250	72.2
0.10	300	61.1
0.10	350	56.3
0.15	200	89.3
0.15	250	76.4
0.15	300	64.1
0.15	350	60.8

Vi vil undersøge dette materiale ved hjælp af en regressionsanalyse. Vi benævner katalysatormængderne (0.05, 0.10, 0.15) v_{11}, v_{21}, v_{31} og temperaturerne (200, 250, 300, 350) v_{12}, v_{22}, v_{32}, v_{42}. Vi antager nu, at udbyttet kan opfattes som realisationer af uafhængige normalt fordelte stokastiske variable $X_1, \cdots, X_{12}$ med

$$\mathrm{E}(X_i) = \alpha + \beta_1 (v_{i1} - \bar{v}_1) + \beta_2 (v_{i2} - \bar{v}_2)$$

og

$$\mathrm{V}(X_i) = \sigma^2.$$

Vi vil estimere parametrene α, β_1, β_2 og σ^2. For at få en lettere numerisk behandling indfører vi for $i = 1, \cdots, 12$

$$\begin{aligned} Y_i &= X_i - 50 \\ t_{i1} &= v_{i1} \cdot 20 \\ t_{i2} &= v_{i2}/50. \end{aligned}$$

I regressionsmodellen med de nye uafhængige variable t_{i1} og t_{i2} kaldes hældningskoefficienterne $\beta_1' = \frac{\beta_1}{20}$ og $\beta_2' = 50\beta_2$.

Beregningerne af summer, kvadratsummer og produktafvigelsessummer fremgår da af følgende skema

	t_1	t_2	y
Værdier	1	4	30.9
	1	5	17.3
	1	6	10.9
	1	7	10.6
	2	4	35.9
	2	5	22.2
	2	6	11.1
	2	7	6.3
	3	4	39.3
	3	5	26.4
	3	6	14.1
	3	7	10.8
s	24	66	235.8
sk	56	378	5986.72
sak	8	15	1353.25
sp_{yt}	492.5	1164.4	
$s_y \cdot s_t$	5659.2	15562.8	
sap_{yt}	20.9	-132.5	
sp_{tt}		132	
$s_t \cdot s_t$		1584	
sap_{tt}		0	

Da $\mathrm{sap}_{tt} = 0$, bliver normalligningerne til bestemmelse af $\hat{\beta}'_1$ og $\hat{\beta}'_2$ særligt simple, og vi taler om en **ortogonal** regressionsanalyse. Ligningerne bliver

$$\begin{aligned} 8 \cdot \hat{\beta}'_1 &= 20.9 \\ 15 \cdot \hat{\beta}'_2 &= -132.5. \end{aligned}$$

Løsningerne bliver åbenbart

$$(\hat{\beta}'_1, \hat{\beta}'_2) = (2.613, -8.833),$$

hvorfor

$$\begin{aligned} \hat{\beta}_1 &= 20 \cdot \hat{\beta}'_1 = 52.26 \\ \hat{\beta}_2 &= \tfrac{1}{50} \cdot \hat{\beta}'_2 = -0.1767 \end{aligned}$$

Da $\bar{x} = 69.65$, får vi altså følgende udtryk for den empiriske regressionsplan

$$\begin{aligned} \hat{\mu}(v_1, v_2) &= 69.65 + 52.26(v_1 - 0.10) - 0.1767(v_2 - 275) \\ &= 113.01 + 52.26 v_1 - 0.1767 v_2. \end{aligned} \tag{5.13}$$

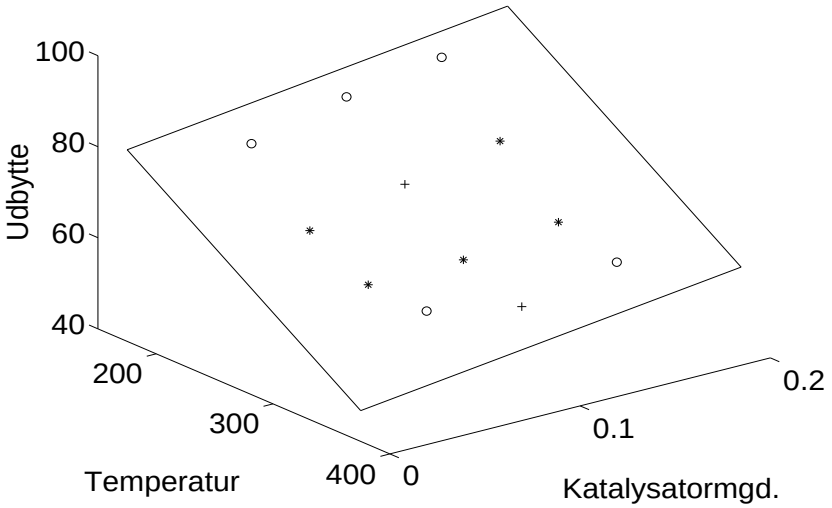
Denne plan er indtegnet i figur 5.1.

Vi vil nu undersøge, om planen kan antages at være parallel med (v_1, v_2)-planen, i.e. om $(\beta_1, \beta_2) = (0, 0)$. Ved hjælp af det p. 475 anførte skema får vi let følgende variansanalyseskema

Variation	SAK	f	s^2	Test
Regresionsplanen	1224.98	2	612.49	42.98
Omkring regressionsplanen	128.27	9	14.25	
Total	1353.25	11		

Den kritiske værdi for testet af $(\beta_1, \beta_2) = (0, 0)$ ved test på 5% niveau er $\mathrm{F}(2, 9)_{0.95} = 4.26$.

Da $42.98 >> \mathrm{F}(2, 9)_{0.9995} = 19.9$, vil vi faktisk forkaste hypotesen $(\beta_1, \beta_2) = (0, 0)$ på alle niveauer større end 0.0005. Vi kan altså ikke med rimelighed antage, at udbyttet er uafhængigt af katalysatormængden og reaktionstemperaturen. Til brug for forudsigelse af et udbytte ved en bestemt kombination af katalysatormængde v_1 og reaktionstemperatur v_2 kan vi anvende (5.8), hvis v_1 er af størrelsesordenen 5–15% og v_2 af størrelsesordenen 200–350 °F (hvor undersøgelsen er foretaget).



Figur 5.1: Plot af udbytte mod temperatur og katalysatormgd. Der er anvendt følgende betegnelser. $\circ$ angiver observation over fladen. $+$ angiver punkt på fladen. $*$ angiver observation under fladen.

Variansen på én måling estimeres til

$$\hat{\sigma}^2 = \frac{1}{9}128.27 = 14.25 = 3.77^2$$

Det er muligt at undersøge forudsætningen for regressionsanalysen lidt nærmere. Såfremt disse holder, vil for $i = 1, \cdots, k$

$$\frac{1}{\sigma}\big(X_i - \alpha - \beta_1(v_{i1} - \bar{v}_1) - \beta_2(v_{i2} - \bar{v}_2)\big) \in \mathrm{N}(0,1),$$

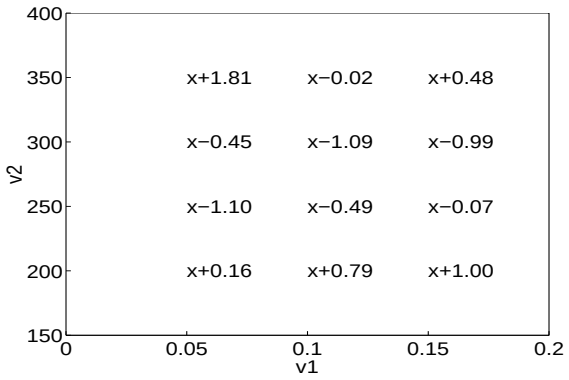
Hvis vi derfor beregner de tilsvarende estimerede størrelser

$$\hat{z}_i = \frac{1}{\hat{\sigma}}\big(X_i - \hat{\alpha} - \hat{\beta}_1(v_{i1} - \bar{v}_1) - \hat{\beta}_2(v_{i2} - \bar{v}_2)\big),$$

vil de være approksimativt $\mathrm{N}(0,1)$-fordelte, hvis linearitetsforudsætningen og normalitetsforudsætningen holder. Disse størrelser er anført i nedenstående skema. I skemaet er medtaget de nødvendige beregninger for udarbejdelse af et fraktildiagram.

x_i	$\hat{\mu}_i$	$x_i - \hat{\mu}_i$	$\hat{z}_i$	$(i-0.5)/12 = p_i$	$u(p_i)$
80.9	80.28	$+0.62$	$+0.16$	0.041	-1.74
67.3	71.45	-4.15	-1.10	0.125	-1.15
60.9	62.61	-1.71	-0.45	0.208	-0.81
60.6	53.78	$+6.82$	$+1.81$	0.292	-0.55
85.9	82.90	$+3.00$	$+0.79$	0.375	-0.32
72.2	74.06	-1.86	-0.49	0.458	-0.11
61.1	65.23	-4.13	-1.09	0.542	0.11
56.3	56.39	-0.09	-0.02	0.625	0.32
89.3	85.51	$+3.79$	$+1.00$	0.708	0.55
76.4	76.67	-0.27	-0.07	0.792	0.81
64.1	67.84	-3.74	-0.99	0.875	1.15
60.8	59.00	$+1.80$	$+0.48$	0.958	1.74

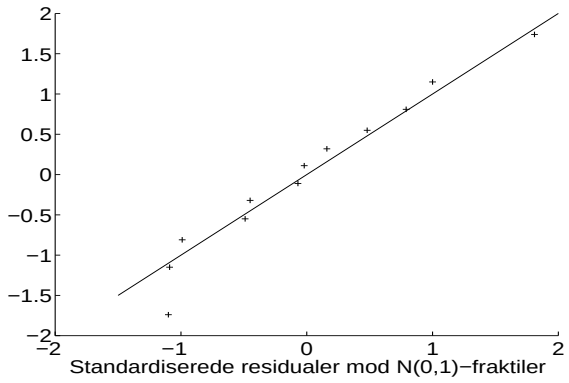
Vi tegner en v_1, v_2-plan, hvor vi ved hvert punkt (v_{1i}, v_{2i}) angiver den tilsvarende værdi af z_i. Vi får da følgende figur



Med det ringe antal observationer taget i betragtning er det vanskeligt at påpege nogen udpræget systematisk i fortegns- og størrelsesvariationen af disse "normerede residualer". Et fraktildiagram af $z_1, \cdots, z_n$ på normalfordelingspapir ser ud som vist på tegningen p. 481.

De standardiserede residualer, $Z_i, \ i = 1, 2, \cdots, 12$, er afbildet på x-aksen og de tilsvarende y-værdier er standardiserede normalfordelingsfraktiler u_{p_i}, hvor $p_i = (i - 0.5)/12$. På såkaldt 'normalpapir' får afbildningen også det viste udseende.

Dette viser heller ingen systematisk afvigelse fra den normale fordeling. Sammenfattende må vi derfor sige, at de for regressionsanalysen nødvendige forudsætninger synes at være til stede. ♦



Med dette eksempel har vi afsluttet gennemgangen af regressionsanalysen. Vi har koncentreret os om den endimensionale lineære regressionsanalyse og antydet en udvidelse af teorien til flere dimensioner. En anden mulig udvidelse er at betragte andre middelværdifunktioner end de lineære, f.eks. polynomier eller eksponentialfunktioner. Vi skal ikke komme ind på disse emner, men henvise til de i referencerne anførte værker om emnet.

5.3 Tests for varianshomogenitet

Vi afslutter kapitlet med et afsnit om nogle af de mest benyttede tests for homogenitet af flere (end to) varianser. Denne hypotese har været antaget opfyldt for alle de modeller, der har været omtalt i dette kapitel.

I forhold til andre antagelser, som ligger til grund for (især) varians– og regressionsanalyser er antagelsen om varianshomogenitet meget betydningsfuld. Derimod er det i praksis mindre væsentligt, f.eks. om data er eksakt normalfordelt; det er i reglen fuld tilstrækkeligt, at de tilfældige afvigelser mellem middelværdimodel og data fordeler sig symmetrisk eller næsten symmetrisk omkring 0, samt, at der ikke optræder 'vilde' afvigelser (såkaldte outliers) i data.

5.3.1 Bartlett's test

Vi betragter uafhængige normalt fordelte stokastiske variable

$$X_{ij} \qquad j = 1, \cdots, n_i, \quad i = 1, \cdots, k$$

med

$$\begin{array}{rcll} \mathrm{E}(X_{ij}) & = & \mu_i & \\ \mathrm{V}(X_{ij}) & = & \sigma_i^2 & \forall i, j. \end{array}$$

Vi ønsker at teste

$$H_0: \quad \sigma_1^2 = \cdots = \sigma_k^2 \qquad \text{mod} \qquad H_1: \quad \exists i, j(\sigma_i^2 \neq \sigma_j^2).$$

Vi har da følgende

SÆTNING 5.12 (BARTLETT). Kvotienttestet på niveau α for test af H_0 mod H_1 er for $n_1 \to \infty, \cdots, n_k \to \infty$ asymptotisk ækvivalent med testet givet ved det kritiske område

$$C = \left\{ (x_{11}, \cdots, x_{kn_k}) | f \log_e s^2 - \sum_{i=1}^{k} f_i \log_e s_{1i}^2 > \chi^2(k-1)_{1-\alpha} \right\}.$$

Her er

$$\begin{aligned} f_i &= n_i - 1 \\ f &= \sum_{i=1}^{k}(n_i - 1) = N - k \\ s_{1i}^2 &= \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2 \\ s^2 &= \frac{1}{f}\Sigma f_i s_{1i}^2. \end{aligned}$$

Bevis. Følger ved at opskrive kvotienttesstørrelsen for testet og dernæst anvende sætning 3.3, p. 323. ■

BEMÆRKNING 5.9. Testet er ret følsomt over for afvigelser i form af asymmetri fra den normale fordeling, således at en eventuel forkastelse kan skyldes en sådan afvigelse og ikke, at varianserne er forskellige. ▼

Vi giver et eksempel

EKSEMPEL 5.10. Vi betragter de i eksempel 5.1 og 5.2 anvendte data. I den ensidede variansanalysemodel vi dér anvendte, forudsatte vi, at varianserne af de tilfældige afvigelser mellem forventningsværdier og data var ens i alle 4 grupper. Da denne forudsætning er af afgørende betydning, vil vi vise, hvorledes dette undersøges.

Vi supplerer beregningsskemaet til variansanalysen med

i	SAK_{1i}	f_i	S_{1i}^2	$\log_e S_{1i}^2$
1	18.175	5	3.635	1.2906
2	38.040	5	7.608	2.0292
3	7.813	5	1.563	0.4466
4	9.908	4	2.477	0.9070
Total	73.936	19		

Vi finder nu let

$$\begin{aligned} \Sigma f_i \log_e s_{1i}^2 &= 22.4596 \\ s^2 = \frac{73.936}{19} &= 3.891 \\ 19 \log_e s^2 &= 25.8147 \\ b &= 25.8147 - 22.4596 \\ &\simeq 3.36. \end{aligned}$$

Da $\chi^2(4-1)_{0.70} = \chi^2(3)_{0.70} = 3.67 > 3.36$, vil vi derfor acceptere hypotesen om $\sigma_1^2 = \cdots = \sigma_4^2$ på alle niveauer mindre end 30%. Den i eksempel 5.2 gjorte antagelse må derfor siges at have været rimelig. ♦

5.3.2 Andre tests for varianshomogenitet

Vi anfører først en sætning, som følger af den centrale grænseværdisætning

SÆTNING 5.13. Med de i afsnit 5.3.1 anvendte betegnelser har vi for $i = 1, \cdots, k$, at

$$\sqrt{n_i - 1}[\log_e S_{1i}^2 - \log_e \sigma_i^2] \quad \text{asymptotisk} \quad \in \mathrm{N}(0, 2)$$

for $n_i \to \infty$.

Bevis. Forbigås ■

Dette kan anvendes til et test af H mod H_0; thi af sætningen får vi, at

$$Z_i = \log_e S_{1i}^2 \quad \text{approksimativt} \quad \in \mathrm{N}\left(\log_e \sigma_i^2, \frac{2}{n_i - 1}\right),$$

således at vi blot skal undersøge, om middelværdierne i k normalfordelinger med kendte varianser kan antages at være ens. Idet vi sætter $2/(n_i - 1) = a_i^2$, kan vi som kritisk område ved test på niveau α anvende

$$C_1 = \left\{ (x_{11}, \cdots, x_{kn_k}) \,|\, \sum (\frac{z_i}{a_i})^2 - \frac{1}{\sum \frac{1}{a_i^2}} (\sum \frac{z_i}{a_i^2})^2 > \chi^2(k-1)_{1-\alpha} \right\}.$$

Her er teststørrelsen fremkommet ved den sædvanlige opspaltning

$$Z = \sum_i \frac{1}{a_i^2}\left(Z_i - \frac{\sum Z_j/a_j^2}{\sum 1/a_j^2}\right)^2 = \sum_i \left(\frac{Z_i}{a_i}\right)^2 - \frac{1}{\sum 1/a_j^2}\left(\sum_i \frac{Z_i}{a_i^2}\right)^2$$

Vi skal ikke komme ind på begrundelsen for, at det kritiske områdes størrelse bestemmes ved fraktiler i χ^2-fordelingen.

Dette test, hvis originale idé skyldes D.G. Kendall og Bartlett, stammer i den her viste version fra [37, p. 272]. Det udmærker sig ved ikke at være så følsomt som Bartlett's test overfor mindre afvigelser fra forudsætningerne om, at observationerne er normalt fordelte. Vi kan her indskyde, at heller ikke variansanalyserne er følsomme over for moderate afvigelser fra normalitet (men meget følsomme over for inhomogene varianser).

EKSEMPEL 5.11. Vi vil nu anvende det nys konstruerede test på det samme datamateriale som i det foregående eksempel. Vi samler beregningerne i nedenstående skema.

i	n_i	a_i^2	a_i	$Z_i' = \log_e S_{1i}^2$	Z_i'/a_i^2	Z_i'/a_i
1	6	0.4	0.6325	1.2906	3.2265	2.0405
2	6	0.4	0.6325	2.0292	5.0730	3.2082
3	6	0.4	0.6325	0.4466	1.1165	0.7061
4	5	0.5	0.7071	0.9070	1.8140	1.2827

Heraf får vi nu

$$\begin{aligned} \sum 1/a_j^2 &= 9.5 \\ \sum (Z_i'/a_i)^2 &= 16.6001 \\ \sum Z_i'/a_i^2 &= 11.2300. \end{aligned}$$

Vi beregner nu let teststørrelsen

$$z = 16.6001 - \frac{1}{9.5}11.2300^2 = 3.33,$$

d.v.s. et resultat, der ligger meget tæt op ad det i eksempel 5.11 fundne. Da vi skal sammenligne med den samme fraktil i χ^2-fordelingen, vil vi derfor komme til samme konklusion, nemlig at varianserne kan antages ens. ♦

Til slut vil vi nævne et test, der er mere hensigtsmæssigt, hvis vi mistænker en enkelt af varianserne for at være større end de øvrige $k-1$. Det bygger på teststørrelsen

$$\frac{\max(S_{11}^2, \cdots, S_{1k}^2)}{\sum S_{1i}^2}. \tag{5.14}$$

Testet skyldes Cochran og kræver ydermere, at $n_1 = \cdots = n_k$. Tabeller over fordelingen af (5.14) findes bl.a. i Techniques of Statistical Analysis (1947).

5.4 Fordelingsfrie tests

I dette afsnit skal vi kort omtale nogle fordelingsfrie metoder, der kan anvendes i situationer svarende til simple varians- og regressionsanalysemodeller. Indledningsvis vil vi kort omtale forskellige former for måleskalaer. Det vil blive godtgjort, at nogle af disse nødvendigvis fører til, at man må anvende nogle ikke-parametriske metoder, nemlig de såkaldte rangtests.

5.4.1 Måleskalaer

Når man foretager almindelige, fysiske målinger som f.eks. at veje nogle emner, tillader man sig e.g. at beregne gennemsnitsvægte af emnerne, eller man udtaler sig om, hvad vægten af halvdelen af et emne vil være (nemlig det halve). Når man foretager disse tilsyneladende uskyldige operationer, underforstår man en meget vigtig ting, nemlig at de reelle tals struktur er **isomorf** med strukturen af den fysiske genstand, der måles på. At dette ikke trivielt er opfyldt, kan det følgende - måske lidt søgte, men ret simple - eksempel vise.

EKSEMPEL 5.12. Ved en sortering af nogle farvede emner knyttes for simpelheds skyld tallene 1, 2, 3 og 4 til de forekommende fire farver. Resultatet af optællingen kan eksempelvis være

1, 1, 2, 1, 4, 3, 3, 4, 2, 4.

En svag sjæl kunne muligvis da føle sig fristet til at beregne gennemsnittet af målingerne. Dette er lig $2\frac{1}{2}$. En sådan operation er selvfølgelig ganske uden mening. Det fundne gennemsnit kan på ingen måde karakterisere populationen af farvede emner. ♦

Det ovenfor anførte eksempel illustrerer en måling foretaget i en såkaldt **nomial-skala** eller en **klassifikationsskala**. En måling i en sådan skala består i, at måleobjektet klassificeres som hørende til netop en af k hinanden udelukkende klasser $A_1, \cdots, A_k$, $k \in \mathrm{N}$. Den eneste relation, der er involveret her, er **ækvivalens**, i.e. at alle medlemmer af en bestemt underklasse må være ækvivalente, hvad angår den egenskab, der klassificeres efter. Vi betegner ækvivalensrelationen med lighedstegnet $=$. Ved målingen er det selvsagt principielt ligegyldigt, om vi i eksempel 5.1 kalder rød for farve nr. 1 eller grøn for farve nr. 1. Det væsentlige er blot, at der til hver farve svarer netop et tal. Dette kan også udtrykkes, at de strukturbevarende transformationer netop er alle **bijektive afbildninger**.

Denne simple struktur influerer selvfølgelig også på de metoder, der kan tages i anvendelse ved en statistisk analyse. Hvis man vil beskrive et sandsynlighedsmål på

en nomial skala, kan dette ikke ske på samme måde som ved mål, defineret på den reelle akse. Som "beliggenhedsmål" kan vi anvende fordelingens **modus**, i.e. angive den klasse, der har den største sandsynlighed. Som **variationsmål** kan vi anvende **entropien**. Kaldes $\mathrm{P}(A_i) = p_i$, defineres entropien ved

$$\mathrm{H}(p_1, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log_a p_i,$$

hvor $\log_a$ betegner en logaritmefunktion med vilkårligt grundtal a. Det kan vises, at

$$\max_{p_1,\cdots,p_n} \mathrm{H}(p_1, \cdots, p_n) = \mathrm{H}(\frac{1}{n}, \cdots, \frac{1}{n})$$

og

$$\min_{p_1,\cdots,p_n} \mathrm{H}(p_1, \cdots, p_n) = \mathrm{H}(1, 0, \cdots, 0).$$

Af disse relationer udledes, at entropien virkelig er et anvendeligt variationsmål. Yderligere oplysninger om entropien kan e.g. findes i [32] eller [36].

Efter et forsøg med målinger i en nomial skala vil man som regel betragte den stokastiske variable $(X_1, \cdots, X_n)$, der angiver antallet af udfald i de forskellige klasser. $(X_1, \cdots, X_n)$ vil være polynomialt fordelte, og estimation og testning foregår som omtalt under **estimation og testning i binomial- og polynomialfordelingen**.

Hvis det er muligt at **rangordne** de emner, vi måler, siger vi, at vi måler på **ordinal-skala**. En sådan foreligger f.eks., når man klassificerer soldater efter militær rang, når man klassificerer børn efter intelligens etc., etc. I en ordinal-skala foreligger der foruden den simple ækvivalensrelation = også en ordningsrelation, som vi kan skrive $<$. Hvis vi vil transformere målinger, der er foretaget i en ordinal-skala, må vi tage hensyn til denne, således at de strukturbevarende transformationer bliver de **monotone afbildninger**.

En sandsynlighedsfordeling, der er defineret på en ordinal-skala, kan selvfølgelig delvis beskrives ved de mål, der er omtalt under nomiale skalaer. Som beliggenhedsmål kan vi endvidere anvende **medianen** og som variationsmål **"forskelle" mellem fraktiler**, f.eks. angive, hvor mange klasser, der ligger mellem 75%- og 25%-fraktilen.

Efter forsøg med målinger i en ordinal-skala vil man som regel betragte de stokastiske variable, der angiver observationens **rang** i den totale stikprøve, i.e. observationens nummer, hvis disse er ordnet efter "størrelse". De testmetoder, der kommer på tale, er de såkaldte **rangtests**.

En mere struktureret måleskala er den såkaldte **interval-skala**. Den fremkommer, når man foruden en rangfølge blandt målingerne også har en **afstand** mellem to vilkårlige målinger. For at sikre skalauafhængighed måles afstandene oftest relativt til afstanden mellem to faste punkter, i.e.

$$\mathrm{d}_0(x,y) = \frac{\mathrm{d}(x,y)}{\mathrm{d}(x_0,y_0)}.$$

Der forudsættes ikke noget "naturgivent" nulpunkt. De strukturbevarende transformationer er de **affine afbildninger**. Et eksempel på målinger i en intervalskala er f.eks. temperaturmålinger. En affin afbildning mellem to ækvivalente måleskalaer er e.g.

$$\mathrm{F} = \frac{9}{5}\mathrm{C} + 32,$$

der giver sammenhængen mellem målinger i Fahrenheit- og Celcius-skalaen.

Sandsynlighedsmål på en intervalskala kan beskrives ved de sædvanlige beliggenheds- og spredningsmål, nemlig **middelværdi** og **varians**. Som stokastisk variabel anvendes sædvanligvis blot den afbildning, der angiver måletallet i en given skala (f.eks. afbildningen, der fører $500°\mathrm{C} \rightarrow 500$). Forsøg med målinger i intervalskala fører til de sædvanlige statistiske modeller som f.eks. **normalfordelingsmodeller** og andre **kontinuerte** sandsynlighedsfordelinger defineret på hele den reelle akse, R.

Hvis man på en intervalskala har et naturgivent nulpunkt, taler man om en **ratio-skala**. En sådan foreligger f.eks. ved vægtmålinger eller ved ventetidsmålinger. De strukturbevarende transformationer er de **lineære afbildninger**. Som eksempel kan angives overgangen fra målinger i engelske fod til meter, i.e.

$$F = 0.3048m$$

Sandsynlighedsmål på en ratioskala (NB. kun den positive del) kan foruden ved de sædvanlige beliggenhedsmål og spredningsmål også beskrives ved de såkaldte geometriske mål, nemlig **geometrisk middelværdi** og **geometrisk standardafvigelse**. Disse defineres præcist ved

$$\begin{aligned} \mu_g &= \exp\{\mathrm{E}(\log_e X)\} \\ \sigma_g &= \exp\{\sqrt{\mathrm{V}(\log_e X)}\}, \end{aligned}$$

naturligvis forudsat, at forventningsværdierne eksisterer. Forsøg med målinger i en ratioskala fører til de modeller, man kender på R_+, nemlig Γ**-fordelingsmodeller** og **logaritmiske normalfordelingsmodeller**.

5.4.2 Invarians og rangtests

Som det fremgår af det foranstående, har vi velegnede statistiske metoder til at behandle måleresulater i de forskellige skalaer bortset fra målinger i ordinale skalaer.

De strukturbevarende afbildninger er her de monotone afbildninger, d.v.s. der er ingen grund til at skelne mellem sæt af måleresultater, der kun afviger ved en monoton transformation. Lad os eksemplificere det for at gøre meningen helt klar. Ved målinger af en gruppes (sociale) status kunne man f.eks. have fået følgende målinger

1, 2, 7, 9

i en given ordinal skala. Nu er der intet afstandsbegreb i en sådan, så man kunne lige så godt have anvendt en skala med de kvadrerede værdier. Da ville man have fået værdierne

1, 4, 49, 81.

Det er nu indlysende, at de statistiske analysemetoder, vi ønsker at anvende, **skal** give samme resulatat, hvad enten vi bruger den ene eller den anden af de to skalaer.

Dette kan udtrykkes, at vores metode skal være **invariant under gruppen af monotone transformationer**.

Lad os f. eks. betragte **tostikprøvesituationen**, i.e. lad der være givet indbyrdes uafhængige observationer

$$X_1, \cdots, X_n$$

og

$$Y_1, \cdots, Y_m.$$

Fordelingsfunktionen for X'er er F og for Y'er G. Vi vil teste hypotesen

$$H_0: \quad \mathrm{F} = \mathrm{G} \qquad \text{mod} \qquad H_1: \quad \mathrm{F} \leq \mathrm{G}, \quad \mathrm{F} \neq \mathrm{G},$$

Det kan da vises, at et **invariant test** er en funktion af observationernes rangværdier. Sættes X_i's rang i den totale stikprøve lig r_i, kan man specielt betragte tests, hvor det kritiske område er givet ved

$$\mathrm{h}(r_1) + \cdots + \mathrm{h}(r_n) < c,$$

hvor h er en monoton afbildning $\mathrm{R} \to \mathrm{R}$ og c er en passende konstant (givet ved signifikansniveauet).

Vælges h lig identiteten, fås det kritiske område

$$\left\{(x_1, \cdots, y_m) | r_1 + \cdots + r_n < c\right\},$$

d.v.s. hypotesen forkastes, hvis summen af X'ernes rangværdier er lille, svarende til, at X'erne fortrinvis er de mindste i den totale stikprøve. Dette test er **Wilcoxons tostikprøvetest**, jvf. p. 383.

Defineres h ved

$$\mathrm{h}(r) = \Phi^{-1}\left(\frac{r}{n+m+1}\right),$$

fås det kritiske område

$$\left\{(x_1, \cdots, y_m) | \sum_{i=1}^{n} \Phi^{-1}\left(\frac{r}{n+m+1}\right) < c\right\}.$$

Dette test er det såkaldte **inverse normalvægtstest** eller **van der Waerdens test**, jvf. p. 377.

Disse tests, der bygger på observationernes rangværdier, kaldes **rangtests**, og de er de eneste tests, der er invariante under monotone afbildninger. Hvis man derfor foretager målinger på en ordinalskala eller på en skala, der principielt godt kan være kontinuert, men hvor man ikke har mulighed for at fastlægge enheder m.v., så den kan udbygges til en intervalskala, da bør man i testsituationer kun anvende rangtests.

I det følgende skal vi give en kort omtale af nogle rangtests, der naturligt supplerer de tidligere givne resultater angående fordelingsfrie metoder. Udvidelserne vil angå situationer, der svarer til simple varians- og regressionsanalysemodeller. Mere komplicerede analyser må søges i litteraturen. Her kan vi anføre [24], [53] og [44].

5.4.3 Kruskal-Wallis' test

Med dette test er vi i stand til at analysere en model, der svarer til en **ensidet variansanalyse model**. Teststørrelsen bygger på rangværdier, og det må derfor forudsættes, at data mindst er målt på en ordinal skala. Hvis dette ikke er tilfældet, kan man anvende χ^2-test for homogenitet af k polynomialfordelinger

Der foreligger observationer

$X_{11}, \cdots, X_{1n_1}$ med kontinuert fordelingsfunktion. F_1
$\cdots$
$X_{k1}, \cdots, X_{kn_k}$ med kontinuert fordelingsfunktion F_k.

Vi er nu interesserede i at teste

$$H_0: \quad F_1 = F_2 = \cdots = F_k$$

mod alternativet, at F'erne er forskellige. Vi ønsker dog, at testet især skal være følsomt over for afvigelser i placeringen af fordelingerne (i.e. deres medianer) og mindre følsomt over for moderate afvigelser i skalaparametre (eller varianser).

Teststørrelsen, vi vælger, bygger på, at samtlige observationer rangordnes, og dernærst erstattes enhver observation med dens nummer i den totale rangordning. Hvis fordelingerne er ens, må man formode, at summerne af rangene i de forskellige grupper er nogenlunde lige store. Hvis dette ikke er tilfældet, forkastes hypotesen.

Idet vi sætter $N = \Sigma n_i$ og $R_i =$ summen af rangværdierne for den i'te gruppe, bliver teststørrelsen

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1). \tag{5.15}$$

Det kritiske område ved test på niveau α er

$$C = \{x_{11}, \cdots, x_{kn_k} | h > h_{1-\alpha}\},$$

hvor h er "den observerede værdi" af H og $h_{1-\alpha}$ er $1-\alpha$-fraktilen i fordelingen af H under H_0.

For moderate værdier af k og n_i er H's nulhypotesefordeling tabelleret. Ellers må man benytte følgende

SÆTNING 5.14. Under H_0 er H givet i (5.15) approximativt $\chi^2(k-1)$-fordelt.

Bevis. Forbigås. ■

Hvis der er mange sammenfaldende observationer (de såkaldte ties), må man korrigere teststørrelsen H. I de fleste tilfælde vil det være tilstrækkeligt at tildele de pågældende

observationer gennemsnittet af de rangværdier, der kommer på tale. Hvis f.eks. de 3 mindste observationer er ens, tilordnes de hver rangværdien 2 $(= \frac{1}{3}(1+2+3))$. Hvis der er mange "ties", må vi dog justere H yderligere.

Lad der eksempelvis være n grupper med sammenfaldende observationer, og lad antallet af observationer i disse grupper være $t_1, \cdots, t_n$. Vi skal da dividere den størrelse H, der beregnes på ovenfor nævnte måde, med størrelsen

$$1 - \frac{\sum_{\nu}(t_\nu^3 - t_\nu)}{N^3 - N}$$

Da denne korrektionsfaktor er mindre end 1, vil den beregnede værdi af teststørrelsen blive forøget. Hvis vi derfor har en signifikant værdi uden at have korrigeret, vil dette blot blive endnu mere udtalt efter en korrektion.

Til belysning af forholdene giver vi nu et

EKSEMPEL 5.13. Vi betragter de i eksempel 5.1 p. 413 anførte værdier over kaliindholdet i nogle gødninger produceret i 4 forskellige omgange. Man fik følgende data

Batch			
A	B	C	D
11.9	15.9	13.9	15.6
9.0	17.2	14.6	18.3
10.2	21.0	16.1	14.6
8.5	17.1	14.9	16.9
13.6	19.0	13.7	14.7
11.3	23.2	12.4	

Vi vil analysere disse data ved hjælp af Kruskal-Wallis' test. Vi finder rangværdierne

	A	B	C	D
	5	15	9	14
	2	19	10.5	20
	3	22	16	10.5
	1	18	13	17
	7	21	8	12
	4	23	6	
Total	22	118	62.5	73.5

Som en kontrol bemærkes, at

$$22 + 118 + 62.5 + 73.5 = 276 = \frac{1}{2}23(23+1).$$

Den beregnede værdi af H bliver

$$\begin{aligned} h &= \frac{12}{N(N+1)}\sum_j \frac{r_j^2}{n_j} - 3(N+1) \\ &= \frac{12}{23\cdot 24}(22^2/6 + 118^2/6 + 62.5^2/6 + 73.5^2/5) - 3\cdot 24 \\ &= 17.84. \end{aligned}$$

Korrektionen for ties er - da der kun er observeret en gruppe med to elementer

$$\sum_\nu (t_\nu^3 - t_\nu) = 2^3 - 2 = 6$$

d.v.s.

$$\text{Korrektion} = 1 - \frac{6}{23^3 - 23} = 1 - \frac{1}{2024} = 0.9995.$$

Den korrigerede værdi af H er derfor

$$h_{\text{kor.}} = \frac{17.84}{0.9995} \simeq 17.85.$$

Da $\chi^2(3)_{.9995}$=17.7, forkastes hypotesen om, at de 4 fordelinger er ens – i hvert tilfælde på alle niveauer større end 0.0005. Det bemærkes, at dette resultat i øvrigt er præcis det samme, som vi fik under anvendelse af F-testet i eksempel 5.1 p. 413. ♦

BEMÆRKNING 5.10. Det er anført, at Kruskal-Wallis' teststørrelse især er følsom over for afvigelser i fordelingernes beliggenhed og mindre over for eventuelle afvigelser i varianser. Som illustration af dette kan vi anføre nogle konstruerede data, alle med median 0 og forskellige spredninger.

$$\begin{array}{lcccccccc} 1) & -20, & -8, & -4, & -2, & 2, & 4, & 8, & 20 \\ 2) & -9, & -5, & -3, & -1, & 1, & 3, & 5, & 9 \\ 3) & -3\frac{1}{2}, & -2\frac{1}{2}, & -1\frac{1}{2}, & -\frac{1}{2}, & \frac{1}{2}, & 1\frac{1}{2}, & 2\frac{1}{2}, & 3\frac{1}{2}. \end{array}$$

Disse rangordnes, og vi får

$$\begin{array}{lcccccccc} 1) & 1, & 3, & 5, & 9, & 16, & 20, & 22, & 24 \\ 2) & 2, & 4, & 7, & 11, & 14, & 18, & 21, & 23 \\ 3) & 6, & 8, & 10, & 12, & 13, & 15, & 17, & 19. \end{array}$$

For alle 3 grupper har vi, at rangsummene er lig 100, d.v.s. Kruskal-Wallis teststørrelsen bliver 0. Dette må dog ikke forlede en til at tro, at "varianser" ingen betydning har; men eksemplet understreger dog, at teststørrelsen ikke (nødvendigvis) er meget følsom over for inhomogeniteter i "varianser". ▼

BEMÆRKNING 5.11. En mere appellerende måde at komme til teststørrelsen på er følgende. Vi bemærker først, at

$$\sum_{i=1}^{k} R_i = \frac{N}{2}(N+1),$$

idet det er lig summen af alle rangværdier, i.e. lig $1 + 2 + \cdots + N$. Gennemsnittet af samtlige rangværdier er altså

$$\bar{R} = \frac{1}{N}\frac{N}{2}(N+1) = \frac{N+1}{2}.$$

Vi beregner nu kvadratafvigelsessummen

$$\sum_i n_i \left(\frac{R_i}{n_i} - \bar{R}\right)^2,$$

hvor R_i/n_i jo er gennemsnittet af rangværdierne i den i'te gruppe. Vi har

$$\begin{aligned}\sum_i n_i \left(\frac{R_i}{n_i} - \bar{R}\right)^2 &= \sum_i n_i \left(\frac{R_i}{n_i} - \frac{N+1}{2}\right)^2 \\ &= \sum_i n_i \frac{R_i^2}{n_i^2} - N\bar{R}^2 \\ &= \sum_i \frac{R_i^2}{n_i} - \frac{N}{4}(N+1)^2.\end{aligned}$$

Vi finder dernæst, at

$$\frac{12}{N(N+1)} \sum_i n_i \left(\frac{R_i}{n_i} - \bar{R}\right)^2 = \frac{12}{N(N+1)} \sum_i \frac{R_i^2}{n_i} - 3(N+1) = H.$$

Vi ser altså, at Kruskal-Wallis' test netop svarer til at lave en ensidet variansanalyse på rangværdierne i stedet for på selve observationsmaterialet. ▼

5.4.4 Friedmans test

Dette test er et rangtest, der svarer til F-testet for **forsvindende søjleeffekt** i **en ensidet variansanalyse med én observation pr. celle** (der findes generaliseringer til mere indviklede variansanalyser, men det skal vi ikke komme ind på her).

Situationen er den, at der er givet uafhængige observationer

$$\begin{array}{c} X_{11}, \cdots, X_{1m} \\ \vdots \qquad \vdots \\ X_{k1}, \cdots, X_{km}, \end{array}$$

der har kontinuerte fordelingfunktioner

$$\begin{array}{c} F_{11}, \cdots, F_{1m} \\ \vdots \qquad \vdots \\ F_{k1}, \cdots, F_{km}. \end{array}$$

Vi ønsker nu at teste, om fordelingerne i hver enkelt række er ens, i.e.

$$H_0: \begin{array}{c} F_{11} = \cdots = F_{1m} \\ \vdots \\ F_{k1} = \cdots = F_{km} \end{array}$$

mod alle alternativer.

Vi rangordner observationerne i hver række og finder summen af rangværdierne i hver søjle:

$$\begin{array}{ll} & \begin{array}{ccccc} R_{11}, & \ldots, & R_{1j}, & \ldots, & R_{1m} \\ \vdots & & \vdots & & \vdots \\ R_{k1}, & \ldots, & R_{kj}, & \ldots, & R_{km} \end{array} \\ \text{Total:} & \overline{R_1, \ldots, R_j, \ldots, R_m} \end{array}$$

Vi bemærker, at rækketotalen altså er lig $m(m+1)/2 = 1 + \cdots + m$.

Ideen i testet er så, at hypotesen forkastes, hvis R_j'erne er meget forskellige, thi da må der eksistere søjler, hvor vi e.g. har meget store rangværdier, hvorfor vi må antage, at de pågældende søjlers fordelinger afviger meget fra de øvriges.

For at bestemme et rimeligt mål for "ensheden" af rangene R_j betragter vi deres afvigelser fra gennemsnittet af rangværdierne. Dette sidste er

$$\bar{R} = \frac{1}{m}\sum_j R_j = \frac{1}{m}k\frac{m}{2}(m+1) = \frac{k(m+1)}{2},$$

således at

$$S^2 = \sum_{j=1}^{m}(R_j - \bar{R})^2 = \sum_{j=1}^{m} R_j^2 - m\bar{R}^2 = \sum_j R_j^2 - \frac{m(m+1)^2k^2}{4}.$$

Et fornuftigt test forkaster nu for S^2 stor. Det er dog mere almindeligt at betragte størrelsen

$$\begin{aligned} C &= \frac{12}{km(m+1)}S^2 = \frac{12}{km(m+1)}\sum_{j=1}^{m}\left(R_j - \frac{k(m+1)}{2}\right)^2 \\ &= \frac{12}{km(m+1)}\sum_{j=1}^{m} R_j^2 - 3k(m+1). \end{aligned} \tag{5.16}$$

Det er denne størrelse, der sædvanligvis benævnes **Friedmans teststørrelse**. Teststørrelsens fordeling under H_0 er tabelleret for $m = 3$ og $k = 2$–9 samt $m = 4$ og $k = 2$–4. Hvis man har et større materiale, kan man benytte

SÆTNING 5.15. Under H_0 er teststørrelsen (5.16) approksimativt $\chi^2(m-1)$-fordelt.

Bevis. (**Heuristisk**) Da størrelserne R_j er summer af uafhængige, stokastiske variable, vil de være asymptotisk normalt fordelte, hvorfor en kvadratsum af disse vil være approximativt χ^2-fordelt. Antallet af frihedsgrader er lig antallet af led minus 1, da vi har det lineære bånd

$$\sum_{j=1}^{m}\left(R - \frac{k(m+1)}{2}\right) = 0.$$

■

Vi illustrerer nu metoderne i det følgende

EKSEMPEL 5.14. Vi betragter de p. 423 anførte data over kartoffeludbyttet ved nogle markforsøg med forskellige gødningstyper. Resultaterne (målt i kg) blev

Mark	Gødning				
	A	B	C	D	E
1	310	353	366	299	367
2	284	293	335	264	314
3	307	306	339	311	377
4	267	308	312	266	342

Problemet er nu at undersøge, om gødningstyperne kan anses for ens, hvad angår udbytte. Vi opfatter derfor resultaterne som realiserede udfald af uafhængige, kontinuert fordelte, stokastiske variable med fordelingsfunktioner

$$F_{iA}, F_{iB}, F_{iC}, F_{iD}, F_{iE}, \qquad i = 1, \cdots, 4,$$

og vi undersøger antagelsen ved at teste

$$H_0: \quad F_{iA} = \cdots = F_{iE}, \qquad i = 1, \cdots, 4,$$

ved hjælp af Friedman's test. Vi rangordner observationerne i hver række og får

	A	B	C	D	E
1	2	3	4	1	5
2	2	3	5	1	4
3	2	1	4	3	5
4	2	3	4	1	5
$\sum$	8	10	17	6	19

Teststørrelsen (5.16) bliver nu

$$C = \frac{12}{4 \cdot 5 \cdot 6}(64 + 100 + 289 + 36 + 361) - 3 \cdot 4 \cdot 6 = 13$$

Da $\mathrm{P}\{\chi^2(4) < 13\} \simeq 98.8\%$, forkastes hypotesen ved test på niveau (approksimativt) $\alpha > 1.2\%$.

Til sammenligning kan anføres, at det analoge F-test i hvert fald forkaster for $\alpha > 0.05\%$.

Havde vi blot undersøgt typerne A, B og D, havde vi fået følgende range

	A	B	D
1	2	3	1
2	2	3	1
3	2	1	3
4	2	3	1
$\sum$	8	10	6

og teststørrelsen ville være blevet

$$\frac{12}{4 \cdot 3 \cdot 4}(64 + 100 + 36) - 3 \cdot 4 \cdot 4 = 2.$$

Ifølge tabel er $\mathrm{P}\{C \geq 2.0\} = 0.431$, d.v.s. hypotesen accepteres for $\alpha \leq 43.1\%$. ♦

5.4.5 Rangkorrelationskoefficienter

Vi skal i dette sidste afsnit angive to størrelser, der, lige som den almindelige korrelationskoefficient

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)\,\mathrm{V}(Y)}},$$

skal være et mål for den sammenhæng, der er mellem 2 stokastiske variable X og Y i den forstand, at de skal angive et mål for, i hvilken udstrækning store (små) værdier af den ene variable er forbundet med store (små) værdier af den anden variable, men de pågældende mål skal være invariant under ordensbevarende transformationer.

Vi betragter en 2-dimensional, stokastisk variabel (X,Y) med fordelingsfunktionen $\mathrm{F}(x,y)$. Lad (X_1,Y_1) og (X_2,Y_2) være uafhængige og identisk fordelte med samme fordeling som (X,Y). Vi har da

DEFINITION 5.3. Ved **Kendalls korrelationskoefficient** $\tau = \tau(X,Y)$ forstås størrelsen

$$\begin{aligned}\tau &= \mathrm{P}\{X_1 < X_2 \wedge Y_1 < Y_2\} + \mathrm{P}\{X_1 > X_2 \wedge Y_1 > Y_2\} \\ &\quad - \mathrm{P}\{X_1 < X_2 \wedge Y_1 > Y_2\} - \mathrm{P}\{X_1 > X_2 \wedge Y_1 < Y_2\}.\end{aligned}$$

Vi ser, at τ virkelig er en størrelse af den søgte art. Dels er $\tau(X,Y) = \tau\big(\mathrm{f}(X),\mathrm{g}(Y)\big)$, hvor f og g er monotont voksende, og dels ser vi, at $\tau = 0$, hvis X og Y er stokastisk uafhængige, at $\tau = 1$, hvis sandsynligheden, for at X og Y er i "modfase", er lig 0, og at $\tau = -1$, hvis sandsynligheden, for at X og Y er i "medfase", er lig 0. De sidste relationer gælder kun for kontinuert fordelte variable, og de fås ved at betragte identiteten

$$\begin{aligned}1 &= \mathrm{P}\{X_1 < X_2 \wedge Y_1 < Y_2\} + \mathrm{P}\{X_1 > X_2 \wedge Y_1 > Y_2\} \\ &\quad + \mathrm{P}\{X_1 < X_2 \wedge Y_1 > Y_2\} + \mathrm{P}\{X_1 > X_2 \wedge Y_1 < Y_2\}.\end{aligned}$$

Det må indskærpes, at relationen "$\tau = 0 \Rightarrow X$ og Yuafhængige" **ikke** er gyldig. ▲

Det kan bemærkes, at vi i en todimensional normalfordeling har

$$\tau = \frac{2}{\pi}\arcsin\rho.$$

Vi går nu over til problemet med estimation af τ. Lad der være givet uafhængige observationer

$$(X_1, Y_1), \cdots, (X_n, Y_n).$$

Vi betragter nu størrelsen

$$S = S_n = \sum_{i<j} \text{sign}(X_i - X_j)\text{sign}(Y_i - Y_j),$$

hvor

$$\text{sign}(x) = \begin{cases} 1, & \text{hvis } x > 0 \\ 0, & \text{hvis } x = 0 \\ -1, & \text{hvis } x < 0 \end{cases}.$$

Hvis vi ordner X'erne i stigende rækkefølge, bliver

$$S = -\sum_{i<j} \text{sign}(Y_i - Y_j).$$

Vi bemærker, at S er uafhængig af, om vi anvender selve observationerne eller blot deres rangværdier. Det kan nu vises, at

$$\tilde{\tau} = \tilde{\tau}_n = \frac{2S}{n(n-1)}$$

er et **centralt skøn** over τ. Estimatoren $\tilde{\tau}_n$ kaldes også **Kendall's** rangkorrelationskoefficient.

Hvis $\tau = 0$, er $\tilde{\tau}$'s (og S_n's) fordeling uafhængig af observationernes fordeling. For enkelte værdier af n er fordelingen af S tabelleret. Det skal præciseres, at S's fordeling er symmetrisk, når $\tau = 0$.

Disse resultater kan anvendes ved test af hypotesen

$$H_0: \quad \tau = 0 \qquad \text{mod} \qquad H_1: \quad \tau \neq 0.$$

For større værdier af n kan man anvende resultatet i

SÆTNING 5.16. Hvis $\tau = 0$, er

$$\tilde{\tau}_n \quad \text{asymptotisk} \quad \in \mathrm{N}\Big(0, \frac{2(2n+5)}{9 \cdot n(n-1)}\Big).$$

Bevis. Forbigås. ■

EKSEMPEL 5.15. Man har anmodet 2 smagsdommere om at ordne 4 forskellige ølbryg efter smag. Man fik følgende resultat.

Bryg	A	B	C	D
Dommer I	3	4	2	1
Dommer II	3	1	4	2

Vi omorganiserer data, således at dommer I's resultater er anført i stigende rækkefølge

Dommer I	1	2	3	4
Dommer II	2	4	3	1

Vi er interesserede i at erfare, om der er nogen sammenhæng mellem de to dommeres resultater, og vi vælger at belyse dette ved af Kendall's τ.

Den observerede værdi af S er

$$s = -\big((1-2) + (2-0) + (1-0)\big) = -2,$$

hvor e.g. tallet $(2-0)$ forklares ved, at der under tallet 2 for dommer I står et 4 ud for dommer II. Vi ser så, at der til højre for dette 4 er 2 observationer mindre end 4 og 0 observationer større end 4. Vi finder dernæst

$$\tilde{\tau} = \frac{2 \cdot (-2)}{4 \cdot 3} = -0.33.$$

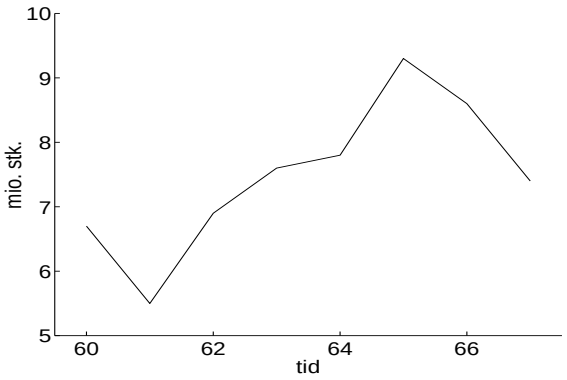
Af tabel fås, at $\mathrm{P}\{S \leq -2\} = \mathrm{P}\{S \geq 2\} = 0.375$, d.v.s. en hypotese, om at $\tau = 0$ ville blive accepteret på alle niveauer $\alpha < 75\%$. Med det foreliggende materiale er der altså ingen grund til at postulere, at der er udtalt sammenhæng mellem de to smagsdommeres vurderinger af øltypernes smag. ♦

BEMÆRKNING 5.12. Der eksisterer formler for korrektion for ties, men for moderate antal ties er det tilstrækkeligt at anvende midtrange (dvs. den sædvanlige gennemsnitlige range) og ikke korrigere herfor ved fastlæggelse af det kritiske område. ▼

Det i det foregående anførte test for hypotesen $\tau = 0$ kan også anvendes som et test for "**trend**" (tendens), hvis vi er i den situation, at X'erne ikke opfattes stokastiske, men som faste tal, og vi ønsker at undersøge, om Y'erne vokser eller aftager med X, jvf. regressionsanalysen. Beregningerne og fordelingen af S er uforandrede, men resultaterne må selvfølgelig tolkes på en ganske anden måde.

EKSEMPEL 5.16. I nedenstående graf og tabel er anført antallet af solgte personbiler i USA i årene 1960–67.

År	60	61	62	63	64	65	66	67
Antal (i mio.)	6.7	5.5	6.9	7.6	7.8	9.3	8.6	7.4



Vi vil undersøge, om der er et trend i disse data. Vi benytter et test, som ovenfor diskuteret. Vi finder

$$\begin{aligned} s &= ((1-6)+(0-6)+(0-5)+1-3)+(1-2)+(2-0)+(1-0)) \\ &= 16 \end{aligned}$$

Ifølge tabel er $\mathrm{P}\{S \geq 16\} = 3.1\%$, d.v.s. en hypotese om, at der ej er trend til stede, vil blive accepteret ved tests med niveau $\alpha < 6.2\%$ (nemlig ved tosidet alternativ: 3.1%+3.1%=6.2%). ♦

Et andet mål for rangkorrelation skyldes [52]. Vi betragter igen uafhængige observationer

$$(X_1, Y_1), \ldots, (X_n, Y_n),$$

og vi sætter

$$\begin{aligned} \mathrm{rg}(X_i) &= S_i \\ \mathrm{rg}(Y_i) &= T_i \end{aligned}$$

Vi har da

DEFINITION 5.4 (SPEARMAN'S RANGKORRELATIONSKOEFFICIENT). r_s for observationerne $(X_1, Y_1), \ldots, (X_n, Y_n)$ er givet ved

$$r_s = \frac{\sum (S_i - \bar{S})(T_i - \bar{T})}{\sqrt{\sum (S_i - \bar{S})^2 \sum (T_i - \bar{T})^2}}$$

d.v.s. som den empiriske korrelationskoefficient mellem observationernes range. ▲

Det følger, at r_s tilfredsstiller den almindelige korrelationsulighed

$$-1 \leq r_s \leq 1.$$

I modsætning til skønnet $\tilde{\tau}$ over Kendall's τ er det derimod vanskeligt umiddelbart at give en sandsynlighedsmæssig fortolkning af r_s, d.v.s. den er ganske vist et mål for sammenhængen mellem X'erne og Y'erne. men ikke estimat for nogen simpel parameter i den simultane fordeling af X og Y.

Der findes forskellige uligheder, som forbinder de to mål. En er **Daniels** ulighed

$$-1 \leq \frac{3(n+2)}{n-2}\tilde{\tau} - \frac{2(n+1)}{n-2} r_s \leq 1,$$

(se f.eks. [33]). Denne bliver for store værdier af n til

$$-1 \leq 3\tilde{\tau} - 2r_s \leq 1.$$

Der findes andre forbindelser mellem τ og r_s, men vi skal ikke komme videre ind herpå.

Indfører vi størrelsen

$$D = \sum_{i=1}^{n} (S_i - T_i)^2,$$

kan det vises, at

$$r_s = 1 - \frac{6D}{n^3 - n},$$

at fordelingen af D er symmetrisk om $\frac{n^3-n}{6}$, at $0 \leq D \leq \frac{1}{3}(n^3 - n)$, og at D kun antager lige værdier. Den er tabelleret på side 512 for $n \leq 11$. Fordelingen er givet under forudsætning af, at der ikke er nogen sammenhæng mellem X'er og Y'er. Den kan derfor bruges ved test af hypotesen

H_0 : ingen sammenhæng mellem X'er og Y'er

mod

H_1 : der er en sammenhæng

EKSEMPEL 5.17. Vi betragter data fra eksempel 5.15 og finder

$$d = (1-2)^2 + (2-4)^2 + (3-3)^2 + (4-1)^2 = 14$$

For $n = 4$ er symmetripunktet $(4^3 - 4)/6 = 10$, så vi finder

$$\mathrm{P}\{D \geq 14\} = \mathrm{P}\{D \leq 6\} = 0.3750,$$

og vi kan derfor acceptere en hypotese om forsvindende sammenhæng for $x \leq 75\%$, den samme værdi som i eksempel 5.15. Den observerede værdi af Spearman's r_s bliver

$$r_s = 1 - \frac{6 \cdot 14}{60} = -0.4.$$

♦

EKSEMPEL 5.18. Vender vi os mod data i eksempel 5.18 fås følgende rangværdier for data i tabellen

$$\begin{aligned} \mathrm{rg}(X_i) &: 1, 2, 3, 4, 5, 6, 7, 8, \\ \mathrm{rg}(Y_i) &: 2, 1, 3, 5, 6, 7, 8, 4. \end{aligned}$$

Dette giver anledning til følgende værdier af d og r_s:

$$\begin{aligned} d &= 1^2 + 1^2 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 4^2 = 22, \\ r_s &= 1 - \frac{6 \cdot 22}{504} = 0.738. \end{aligned}$$

Af tabellen findes

$$\mathrm{P}\{D \leq 22\} = 0.229,$$

d.v.s. en hypotese om tilstedeværelse af et trend vil ved et tosidet test blive accepteret på alle niveauer mindre end lig med 4.58%, igen i rimelig overensstemmelse med eksempel 5.16. ♦

Hvis antallet af observationer overstiger 11 kan man ikke benytte tabellen. Man kan da anvende

SÆTNING 5.17. Under antagelsen om manglende sammenhæng mellem X'er og Y'er er D approksimativt normalt fordelt med

$$\begin{aligned} \mathrm{E}(D) &= \frac{n^3 - n}{6}, \\ \mathrm{V}(D) &= \frac{n^2(n+1)^2(n-1)}{36}. \end{aligned}$$

Bevis. Forbigås. Se f.eks. [38]. ■

BEMÆRKNING 5.13. Ved anvendelse af sætningen bør man have in mente, at fordelingen er koncentreret på de **lige tal**, således at der anvendes en **kontinuitetskorrektion**. For $n = 8$ fås f.eks.

$$\begin{aligned} \mathrm{P}\{D \leq 22\} &= \mathrm{P}\{D \leq 23\} \\ &\simeq \mathrm{P}\big\{\mathrm{N}(84, 1008) \leq 23\big\} \\ &= \Phi\left(\frac{23 - 84}{31.749}\right) = \Phi(-1.921) = 0.0273 \end{aligned}$$

▼

BEMÆRKNING 5.14. Afslutningsvis kan anføres, at man i tilfældet med ties i analogi med de tilsvarende anførte tests kan anvende midtrange, dvs. den gennemsnitlige rang for de indgående ties. ▼

5.4.6 Tabeller

På de næstfølgende sider er anført en række tabeller til brug ved analyse af rangordnede data.

Tabellerne vedrører

Tabel over sandsynligheder for værdier, der er større end eller lig med de observerede værdier i Kruskal Wallis' ensidede rangvariansnalyse. [1]

Stikprøve-størrelse			H	P
n_1	n_2	n_3		
2	1	1	2.7000	.500
2	2	1	3.6000	.200
2	2	2	4.5714	.067
			3.7143	.200
3	1	1	3.2000	.300
3	2	1	4.2857	.100
			3.8571	.133
3	2	2	5.3572	.029
			4.7143	.048
			4.5000	.067
			4.4643	.105
3	3	1	5.1429	.043
			4.5714	.100
			4.0000	.129
3	3	2	6.2500	.011
			5.3611	.032
			5.1389	.061
			4.5556	.100
			4.2500	.121
3	3	3	7.2000	.004
			6.4889	.011
			5.6889	.029
			5.6000	.050
			5.0667	.086
			4.6222	.100
4	1	1	3.5714	.200
4	2	1	4.8214	.057
			4.5000	.076
			4.0179	.114
4	2	2	6.0000	.014
			5.3333	.033
			5.1250	.052
			4.4583	.100
			4.1667	.105
4	3	1	5.8333	.021
			5.2083	.050
			5.0000	.057
			4.0556	.093
			3.8889	.129
4	3	2	6.4444	.008
			6.3000	.011
			5.4444	.046
			5.4000	.051
			4.5111	.098
			4.4444	.102
4	3	3	6.7455	.010
			6.7091	.013
			5.7009	.046
			5.7273	.050
			4.7091	.092
			4.7000	.101
4	4	1	6.6667	.010
			6.1667	.022
			4.9667	.048
			4.8667	.054
			4.1667	.082
			4.0667	.102
4	4	2	7.0364	.006
			6.8727	.011
			5.4545	.046
			5.2364	.052
			4.5545	.098
			4.4455	.103
4	4	3	7.1439	.010
			7.1364	.011
			5.5985	.049
			5.5758	.051
			4.5455	.099
			4.4773	.102
4	4	4	7.6538	.008
			7.5385	.011
			5.6923	.049
			5.6538	.054
			4.6539	.097
			4.5001	.104
5	1	1	3.8571	.143
5	2	1	5.2500	.036
			5.0000	.048
			4.4500	.071
			4.2000	.095
			4.0500	.119

[1] Fra Kruskal, W.H., and Wallis, W.A. 1952. Use of ranks in one-criterion variance analysis. J. Amer. Statist. Ass., 47, 614–617.

Stikprøve-størrelse			H	P
n_1	n_2	n_3		
5	2	2	6.5333	.008
			6.1333	
			5.1600	.034
			5.0400	.056
			4.3733	.090
			4.2933	.122
5	3	1	6.4000	.012
			4.9600	.048
			4.8711	.052
			4.0178	.095
			3.8400	.123
5	3	2	6.9091	.009
			6.8218	.010
			5.2509	.049
			5.1055	.052
			4.6509	.091
			4.4945	.101
5	3	3	7.0788	.009
			6.9818	.011
			5.6485	.049
			5.5152	.051
			5.5333	.097
			4.4121	.109
5	4	1	6.9545	.008
			6.8400	.011
			4.9855	.044
			4.8600	.056
			3.9873	.098
			3.9600	.102
5	4	2	7.2045	.009
			7.1182	.010
			5.2727	.049
			5.2682	.050
			4.5409	.098
			4.5182	.101
5	4	3	7.4449	.010
			7.3949	.011
			5.6564	.049
			5.6308	.050
			4.5487	.099
			4.5231	.103
5	4	4	7.7604	.009
			7.7440	.011
			5.6571	.049
			5.6176	.050
			4.6187	.100
			4.5527	.102
5	5	1	7.3091	.009
			6.8364	.011
			5.1273	.046
			4.9091	.053
			4.1091	.086
			4.0364	.105
5	5	2	7.3385	.010
			7.2692	.010
			5.3385	.047
			5.2462	.051
			4.6231	.097
			4.5077	.100
5	5	3	7.5780	.010
			7.5429	.010
			5.7055	.046
			5.6264	.051
			4.5451	.100
			4.5363	.102
5	5	4	7.8229	.010
			7.7914	.010
			5.6657	.049
			5.6429	.050
			4.5229	.099
			4.5200	.101
5	5	5	8.0000	.009
			7.9800	.010
			5.7800	.049
			5.6600	.051
			4.5600	.100
			4.5000	.102

Tabel over sandsynligheder for værdier, der er større end eller lig med de observerede værdier i Friedman's tosidede rangvariansanalyse[2]

m=3

k=2		k=3		k=4		k=5	
c	p	c	p	c	p	c	p
0	1.000	.000	1.000	.0	1.000	.0	1.000
1.0	.833	.667	.944	0.5	.931	0.4	.954
3.0	.500	2.000	.528	1.5	.653	1.2	.691
4.0	.167	2.667	.361	2.0	.431	1.6	.522
		4.667	.194	3.5	.273	2.8	.367
		6.0	.028	4.5	.125	3.6	.182
				6.0	.069	4.8	.124
				6.5	.042	5.2	.093
				8.0	.0046	6.4	.039
						7.6	.024
						8.4	.0085
						10.0	.00077

k=6		k=7		k=8		k=9	
c	p	c	p	c	p	c	p
.00	1.000	.000	1.000	.00	1.000	.000	1.000
.33	.956	.286	.964	.25	.967	.222	.971
1.00	.740	.857	.768	.75	.794	.667	.814
1.33	.570	1.143	.620	1.00	.654	.889	.865
2.33	.430	2.000	.486	1.75	.531	1.556	.569
3.00	.252	2.571	.305	2.25	.355	2.000	.398
4.00	.184	3.429	.237	3.00	.285	2.667	.328
4.33	.142	3.714	.192	3.25	.236	2.889	.278
5.33	.072	4.571	.112	4.00	.149	3.556	.187
6.33	.052	5.429	.085	4.75	.120	4.222	.154
7.00	.029	6.000	.052	5.25	.079	4.667	.107
8.33	.012	7.143	.027	6.25	.047	5.556	.069
9.00	.0081	7.714	.021	6.75	.038	6.000	.057
9.33	.0055	8.000	.016	7.00	.030	6.222	.048
10.33	.0017	8.857	.0084	7.75	.018	6.889	.031
12.00	.00013	10.286	.0036	9.00	.0099	8.000	.019
		10.571	.0027	9.25	.0080	8.222	.016
		11.143	.0012	9.75	.0048	8.667	.010
		12.286	.00032	10.75	.0024	9.556	.0060
		14.000	.000021	12.00	.0011	10.667	.0035
				12.25	.00086	10.089	.0029
				13.00	.00026	11.556	.0013
				14.25	.000061	12.667	.00066
				16.00	.0000036	13.556	.00035
						14.000	.00020
						14.222	.000097
						14.889	.000054
						16.222	.000011
						18.000	.0000006

[2]Taget fra Friedman, M. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Amer. Statist. Ass., 32, 688–689.

m=4

k=2		k=3		k=4			
c	p	c	p	c	p	c	p
.0	1.000	.2	1.000	.0	1.000	5.7	.141
.6	.958	.6	.958	.3	.992	6.0	.105
1.2	.834	1.0	.910	.6	.928	6.3	.094
1.8	.792	1.8	.727	.9	.900	6.6	.077
2.4	.625	2.2	.608	1.2	.800	6.9	.068
3.0	.542	2.6	.524	1.5	.754	7.2	.054
3.6	.458	3.4	.446	1.8	.677	7.5	.052
4.2	.375	3.8	.342	2.1	.649	7.8	.036
4.8	.208	4.2	.300	2.4	.524	8.1	.033
5.4	.167	5.0	.207	2.7	.508	8.4	.019
6.0	.042	5.4	.175	3.0	.432	8.7	.014
		5.8	.148	3.3	.389	9.3	.012
		6.6	.075	3.6	.355	9.6	.0069
		7.0	.054	3.9	.324	9.9	.0062
		7.4	.033	4.5	.242	10.2	.0027
		8.2	.017	4.8	.200	10.8	.0016
		9.0	.0017	5.1	.190	11.1	.00094
				5.4	.158	12.0	.000072

Tabel over sandsynligheder for værdier, der er større end eller lig med observerede værdier af S i Kendall's rang korrelationskoefficient.[3]

S	n				S	n		
	4	5	8	9		6	7	10
0	.625	.592	.548	.540	1	.500	.500	.500
2	.375	.408	.452	.460	3	.360	.386	.431
4	.167	.242	.360	.381	5	.235	.281	.364
6	.042	.117	.274	.306	7	.136	.191	.300
8		.042	.199	.238	9	.068	.119	.242
10		.0083	.138	.179	11	.028	.068	.190
12			.089	.130	13	.0083	.035	.146
14			.054	.090	15	.0014	.015	.108
16			.031	.060	17		.0054	.078
18			.016	.038	19		.0014	.054
20			.0071	.022	21		.00020	.036
22			.0028	.012	23			.023
24			.00087	.0063	25			.014
26			.00019	.0029	27			.0083
28			.000025	.0012	29			.0046
30				.00043	31			.0023
32				.00012	33			.0011
34				.000025	35			.00047
36				.0000028	37			.00018
					39			.000058
					41			.000015
					43			.0000028
					45			.00000028

[3] Fra: Kendall, M.G., Rank correlation methods, Charles Griffin & Company, Ltd., London, 1948, Appendix Table 1, p. 141

Tabel over værdier , der er mindre end eller lig med d i fordelingen af D i Spearman's rangkorrelationskoefficient. [4]

[4] Fra [38]

d	P	d	P	d	P	d	P	d	P	d	P
$N = 2$		12	.0240	60	.2504	80	.1927	64	.0334	.	.	116	.0729
0	.5000	14	.0331	62	.2682	82	.2050	66	.0367	.	.	118	.0773
———		16	.0440	64	.2911	84	.2183	68	.0403	.	.	120	.0817
$N = 3$		18	.0548	66	.3095	86	.2315	70	.0441	12	.0000	122	.0865
0	.1667	20	.0694	68	.3323	88	.2467	72	.0481	14	.0000	124	.0913
2	.5000	22	.0833	70	.3517	90	.2603	74	.0524	16	.0001	126	.0964
———		24	.1000	72	.3760	92	.2759	76	.0569	18	.0001	128	.1015
$N = 4$		26	.1179	74	.3965	94	.2905	78	.0616	20	.0001	130	.1070
0	.0417	28	.1333	76	.4201	96	.3067	80	.0667	22	.0002	132	.1125
2	.1667	30	.1512	78	.4410	98	.3218	82	.0720	24	.0003	134	.1183
2	.2083	32	.1768	80	.4674	100	.3389	84	.0774	26	.0003	136	.1242
6	.3750	34	.1978	82	.4884	102	.3540	86	.0831	28	.0005	138	.1304
8	.4583	36	.2222	84	.5116	104	.3718	88	.0893	30	.0006	140	.1365
10	.5417	38	.2488	———		106	.3878	90	.0956	32	.0007	142	.1431
———		40	.2780	$N = 9$		108	.4050	92	.1022	34	.0009	144	.1496
$N = 5$		42	.2974	0	.0000	110	.4216	94	.1091	36	.0011	146	.1566
0	.0083	44	.3308	2	.0000	112	.4400	96	.1163	38	.0014	148	.1634
2	.0417	46	.3565	4	.0001	114	.4558	98	.1237	40	.0016	150	.1708
4	.0667	48	.3913	6	.0002	116	.4742	100	.1316	42	.0020	152	.1780
6	.1167	50	.4198	8	.0004	118	.4908	102	.1394	44	.0023	154	.1857
8	.1750	52	.4532	10	.0007	120	.5092	104	.1478	46	.0027	156	.1932
10	.2250	54	.4817	12	.0010	———		106	.1564	48	.0032	158	.2012
12	.2583	56	.5183	14	.0015	$N = 10$		108	.1652	50	.0037	160	.2091
14	.3417	———		16	.0023	0	.0000	110	.1744	52	.0043	162	.2174
16	.3917	$N = 8$		18	.0030	2	.0000	112	.1839	54	.0049	164	.2255
18	.4750	0	.0000	20	.0041	4	.0000	114	.1935	56	.0056	166	.2342
20	.5250	2	.0002	22	.0054	6	.0000	116	.2035	58	.0064	168	.2427
———		4	.0006	24	.0069	8	.0001	118	.2135	60	.0072	170	.2517
$N = 6$		6	.0011	26	.0086	10	.0001	120	.2241	62	.0081	172	.2604
0	.0014	8	.0023	28	.0107	12	.0002	122	.2349	64	.0091	174	.2697
2	.0083	10	.0036	30	.0127	14	.0003	124	.2459	66	.0102	176	.2787
4	.0167	12	.0054	32	.0156	16	.0004	126	.2567	68	.0113	178	.2883
6	.0292	14	.0077	34	.0184	18	.0006	128	.2683	70	.0126	180	.2975
8	.0514	16	.0109	36	.0216	20	.0008	130	.2801	72	.0139	182	.3073
10	.0681	18	.0140	38	.0252	22	.0011	132	.2918	74	.0153	184	.3168
12	.0875	20	.0184	40	.0294	24	.0014	134	.3037	76	.0168	186	.3269
14	.1208	22	.0229	42	.0333	26	.0019	136	.3161	78	.0184	188	.3366
16	.1486	24	.0288	44	.0380	28	.0024	138	.3284	80	.0201	190	.3469
18	.1778	26	.0347	46	.0429	30	.0029	140	.3410	82	.0220	192	.3568
20	.2097	28	.0415	48	.0484	32	.0036	142	.3536	84	.0239	194	.3673
22	.2486	30	.0481	50	.0540	34	.0044	144	.3665	86	.0260	196	.3773
24	.2819	32	.0575	52	.0603	36	.0053	146	.3795	88	.0281	198	.3879
26	.3292	34	.0661	54	.0664	38	.0063	148	.3925	90	.0304	200	.3982
28	.3569	36	.0756	56	.0738	40	.0075	150	.4056	92	.0328	202	.4090
30	.4014	38	.0855	58	.0809	42	.0087	152	.4191	94	.0354	204	.4192
32	.4597	40	.0983	60	.0888	44	.0101	154	.4326	96	.0380	206	.4302
34	.5000	42	.1081	62	.0969	46	.0117	156	.4458	98	.0409	208	.4406
———		44	.1215	64	.1063	48	.0134	158	.4592	100	.0438	210	.4516
$N = 7$		46	.1337	66	.1149	50	.0153	160	.4730	102	.0470	212	.4621
0	.0002	48	.1496	68	.1250	52	.0173	162	.4865	104	.0502	214	.4731
2	.0014	50	.1634	70	.1348	54	.0195	164	.5000	106	.0536	216	.4837
4	.0034	52	.1799	72	.1456	56	.0219	———		108	.0571	218	.4947
6	.0062	54	.1947	74	.1563	58	.0245	$N = 11$		110	.0609	220	.5053
8	.0119	56	.2139	76	.1681	60	.0272	0	.0000	112	.0647		
10	.0171	58	.2309	78	.1793	62	.0302	2	.0000	114	.0688		

Kapitel 6

Beslutningsteori

Et af de spørgsmål, der trænger sig mest på efter vor gennemgang af specielt testteorien, er, om der findes en rationel metode til bestemmelse af et rimeligt niveau α. Det er ikke altid tilfredsstillende blot a priori at fastlægge det til e.g. 5%, selv om det i praksis ofte er den værdi, man uden videre overvejelser vælger. En sådan metode kan man i nogle tilfælde konstruere ved hjælp af **decisions** eller **beslutningsteorien**, især hvis man er i besiddelse af en tabsfunktion, i.e. hvis man kan rangordne forskellige fejlbeslutninger og angive, hvor meget værre det er at begå én fejl end en vilkårlig anden. Det ligger uden for denne fremstillings rammer at give en dybere indføring i den statistiske beslutningsteori. På den anden side ville vores gennemgang af vigtige statistiske principper og værktøjer være ufuldstændig, hvis man ikke fik antydet løsningerne på sådanne problemer. Vi vil her omtale nogle vigtige begreber i den **statistiske** beslutningsteori, i.e. den gren, der berører de tilfælde, hvor der foreligger observationer, som er realiserede udfald af stokastiske variable. Den rene såkaldte spilteori vil derimod **ikke** blive omtalt. Her kan henvises til f.eks. [39].

6.1 Generelt om beslutningsteori

Som antydet indledningsvis vil vor gennemgang være af causerende karakter: sætninger og definitioner vil ikke altid blive anført med den matematiske stringens, man kunne ønske, idet en sådan fremstilling ville være unødigt tung og omfattende til vort formål.

6.1.1 Definitioner og metoder

Vi antager, at vi er i den situation, at vi skal træffe beslutninger, der afhænger af mængden af **tilstande**

$$\Omega = \{\theta | \theta \in \Omega\}, \tag{6.1}$$

de såkaldte **parametre**. Disse (naturens) tilstande er ukendte for os. Vi ved, hvilke muligheder der er, nemlig Ω, men vi ved ikke, hvad den relevante værdi af θ er. De mulige **beslutninger** vi kan antage, er **aktionerne**

$$\acute{A} = \{a | a \in \acute{A}\}. \tag{6.2}$$

Det er naturligvis ikke underordnet, hvilken aktion vi vælger, idet der er knyttet et **tab** til valget af en aktion a, hvis den sande parameterværdi er θ, nemlig $\mathrm{L}(\theta, a)$, hvor

$$\mathrm{L}: \quad \Omega \times \acute{A} \to \mathrm{R} \tag{6.3}$$

altså er tabsfunktionen. For at kunne træffe den "rigtige" beslutning foretager vi - med henblik på at skaffe os oplysning om θ - et **eksperiment**, hvis udfald er en stokastisk variabel

$$X = (X_1, \cdots, X_n). \tag{6.4}$$

Fordelingen af udfaldet afhænger af, hvilken parameter der er involveret. Hvis parameteren er θ, har X frekvensfunktionen

$$\mathrm{f}(x_1, \cdots, x_n | \theta) = \mathrm{f}(x | \theta). \tag{6.5}$$

For at bestemme den aktion, vi vil tage efter at have konstateret udfaldet x, definerer vi **beslutningsfunktionen** eller **strategien**

$$\mathrm{d}: \quad \mathrm{R}^n \to \acute{A}, \quad \text{d.v.s.} \quad \mathrm{d}(x) = \mathrm{d}(x_1, \cdots, x_n) \in \acute{A}. \tag{6.6}$$

Hvis man ønsker en konkret situation som et eksempel til "forklaring" af begreberne, kan man anvende følgende

EKSEMPEL 6.1. En fabrikant, der vil markedsføre et nyt produkt, er i tvivl om, hvilket navn han skal give produktet, hvilken emballage han skal anvende etc. For at skaffe

sig et overblik over, hvad hans potentielle kunder ville foretrække, bekoster han en markedsundersøgelse, og på basis af udfaldet af denne træffer han sit valg af navn, emballage etc. Her svarer hans potentielle kunders opfattelser til tilstandene Ω og et konkret valg af navn og emballage til en aktion a. Det er naturligvis ikke underordnet, hvad produktet hedder, og hvordan emballagen ser ud. Hvis kunderne foretrækker røde æsker, og varen sælges i gule, vil fabrikanten ikke tjene så meget, som hvis han havde valgt røde æsker. Der er altså knyttet et tab $\mathrm{L}(\theta, a)$ til enhver tilstand og enhver aktion. Markedsundersøgelsen svarer naturligvis til eksperimentet, og udfaldet af undersøgelsen afhænger naturligvis af den mening, kunderne har, i.e. θ. Endelig vælger fabrikanten jo navn og emballage på basis af undersøgelsens udfald, i.e. fabrikanten har en beslutningsfunktion d. ♦

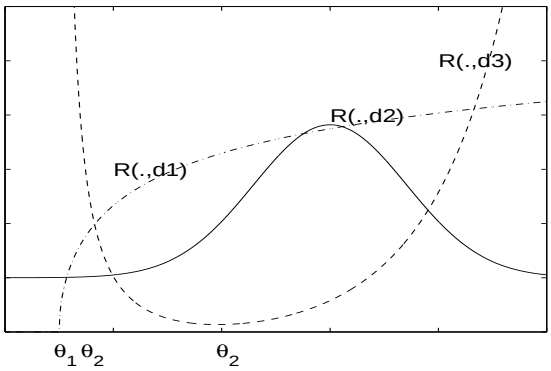
Vi vender nu tilbage til den almindelige situation. Formålet med en "rationel" beslutningstagen må være at vælge de aktioner, der i en eller anden passende forstand minimaliserer vort tab. Med henblik på dette defineres **risikofunktionen**.

$$\mathrm{R}(\theta, d) = \mathrm{E}_\theta\Big(\mathrm{L}\big(\theta, \mathrm{d}(X)\big)\Big) = \begin{cases} \sum\limits_{x} \mathrm{L}\big(\theta, \mathrm{d}(x)\big)\mathrm{f}(x|\theta) \\ \int\limits_{-\infty}^{\infty} \mathrm{L}\big(\theta, \mathrm{d}(x)\big)\mathrm{f}(x|\theta)\, dx \end{cases} \tag{6.7}$$

hvor vi selvfølgelig anvender summen, hvis $X = (X_1, \cdots, X_n)$ er diskret, og integralet, hvis $X = (X_1, \cdots, X_n)$ er kontinuert. Risikoen er altså det forventede tab, hvor vi tager forventningsværdi over de mulige udfald af vort eksperiment. Vi vil nu søge at bestemme en beslutningsfunktion, der bevirker, at vor risiko R bliver lille. Hvis vi kan finde et d_0, så vi for **alle** θ og alle d har

$$\mathrm{R}(\theta, d_0) \leq \mathrm{R}(\theta, d)\,,$$

er d_0 uniformt bedre end alle andre d og må derfor foretrækkes. Nu viser det sig imidlertid, at det er yderst sjældent, et sådant d_0 eksisterer. Langt hyppigere har vi en situation som antydet i nedenstående figur, hvor risikofunktionernes grafer skærer hinanden



Her er det vanskeligere at vælge en beslutningsfunktion. I θ_1 er $\mathrm{R}(\cdot, d_1)$ mindst, i θ_2 $\mathrm{R}(\cdot, d_2)$ etc. En mulig løsning på dette dilemma er at vælge den forsigtige strategi d_1, idet d_1 har den egenskab, at den maksimale risiko, vi løber, er mindst mulig. Vi taler om en **minimax-strategi**, og definitionen er altså

DEFINITION 6.1. Strategien d_1 siges at være en **minimax-strategi**, hvis

$$\sup_{\theta} \mathrm{R}(\theta, d_1) = \inf_{d} \sup_{\theta} \mathrm{R}(\theta, d), \tag{6.8}$$

eller, hvis ekstremumsværdierne antages

$$\max_{\theta} \mathrm{R}(\theta, d_1) = \min_{d} \max_{\theta} \mathrm{R}(\theta, d). \tag{6.9}$$

▲

I ovenstående figur vil d_2 være minimax-strategien.

I mange situationer vil parameteren θ variere stokastisk. Hvis man f.eks. har en automatisk tappemaskine, vil den faktiske fyldehøjde X i flaskerne formentlig følge en fordeling med indstillingshøjden θ som middelværdi. Vi sætter frekvensfunktionen for X lig $\mathrm{f}(x|\theta)$. Når man nu hver morgen indstiller maskinen til at give fyldehøjden θ, vil dette være forbundet med nogen usikkerhed (jfr. afsnittet om compound fordelinger). Vi kan derfor opfatte den faktiske indstilling θ som det realiserede udfald af en stokastisk variabel Θ, og den usikkerhed, der ligger i værdien af Θ, kan vi udtrykke ved en frekvensfunktion

$$\mathrm{g}(\theta), \quad \theta \in \Omega, \tag{6.10}$$

den såkaldte **a priori fordeling** for Θ.

Når man nu har målt fyldehøjden i en flaske, ville det muligvis falde én ind at spørge, hvor godt har man nu indstillet maskinen, i.e. hvad kan jeg - efter at have målt X - sige om Θ's fordeling?

Hvis vi forudsætter, at både X og Θ kun antager endeligt mange værdier, kan vi mere konkret sige, at vi søger den **betingede fordeling**

$$\mathrm{P}\{\Theta = \theta | X = x\} = \mathrm{k}(\theta|x)\,.$$

Her kan k let bestemmes, idet vi jo har

$$\begin{aligned}\mathrm{P}\{\Theta = \theta | X = x\} &= \frac{\mathrm{P}\big(\{\Theta = \theta\} \cap \{X = x\}\big)}{\mathrm{P}\{X = x\}} \\ &= \frac{\mathrm{P}\{X = x|\Theta = \theta\}\mathrm{P}\{\Theta = \theta\}}{\mathrm{P}\{X = x\}} \\ &= \frac{\mathrm{f}(x|\theta)\mathrm{g}(\theta)}{\mathrm{h}(x)}\,, \end{aligned} \tag{6.11}$$

hvor $\mathrm{h}(x)$ er givet ved

$$\mathrm{h}(x) = \mathrm{P}\{X = x\} = \sum_i \mathrm{f}(x|\theta_i) = \mathrm{g}(\theta_i)\,,$$

jfr. sætning 1.70 p. 189. Vi bemærker i øvrigt analogien til Bayes regel (p. 38).

Den således fremkomne betingede frekvensfunktion kaldes **a posteriori fordelingen** for Θ **givet** observationen $X = x$. Generelt vil vi blot definere a posteriori fordelingen ved udtrykket (6.11), d.v.s.

DEFINITION 6.2. Lad Θ være en stokastisk variabel med (a priori) frekvensfunktion $\mathrm{g}(\theta)$, og lad $X = (X_1, \cdots, X_n)$ være en stokastisk variabel, der for givet $\Theta = \theta$ har frekvensfunktionen $\mathrm{f}(x|\theta)$. Da defineres **a posteriori fordelingen** k for Θ givet $X = x$ ved

$$\mathrm{k}(\theta|x) = \frac{\mathrm{f}(x|\theta)\mathrm{g}(\theta)}{\mathrm{h}(x)}\,, \tag{6.12}$$

hvor

$$\mathrm{h}(x) = \begin{cases} \int_{-\infty}^{\infty} \mathrm{f}(x|\theta)\mathrm{g}(\theta)\, d\theta, & \text{hvis } \Theta \text{ type K} \\ \sum_{\theta} \mathrm{f}(x|\theta)\mathrm{g}(\theta), & \text{hvis } \Theta \text{ type D.} \end{cases}$$

Vi bemærker, at også Θ kan være flerdimensional. ▲

Af hensyn til senere referencer anfører vi nogle nyttige a priori - a posteriori fordelinger i nedenstående skema:

A priori fordeling for Θ	Fordeling af X for $\Theta = \theta$	A posteriori fordeling af Θ for $X = x$
$\mathrm{Be}(\alpha, \beta)$	$\mathrm{B}(n, \theta)$	$\mathrm{Be}(\alpha + x, \beta + n - x)$
$\mathrm{G}(k, \beta)$	$\mathrm{P}(\theta)$	$\mathrm{G}(k + x, \frac{\beta}{\beta+1})$
$\mathrm{N}(\mu, \sigma^2)$	$\mathrm{N}(\theta, \tau^2)$	$\mathrm{N}(\mu_1, \sigma_1^2)$

I skemaet er anvendt betegnelserne:

$$\begin{aligned} \mu_1 &= \frac{\mu/\sigma^2 + x/\tau^2}{1/\sigma^2 + 1/\tau^2} \\ \frac{1}{\sigma_1^2} &= \frac{1}{\sigma^2} + \frac{1}{\tau^2}. \end{aligned}$$

Bevis. Vi viser f.eks. resultatet for $\mathrm{G}(k, \beta)$- fordelingen. Vi har ifølge resultatet p. 189, at

$$\mathrm{h}(x) = \frac{\Gamma(x+k)}{x!\Gamma(k)} \left(\frac{1}{1+\beta}\right)^k \left(\frac{\beta}{\beta+1}\right)^x,$$

d.v.s. ifølge definitionen (6.12)

$$\begin{aligned} \mathrm{k}(\theta|x) &= \frac{1}{\Gamma(k)} \frac{1}{\beta^k} \theta^{k-1} \mathrm{e}^{-\frac{\theta}{\beta}} \frac{\theta^x}{x!} \mathrm{e}^{-\theta} \frac{x!\Gamma(k)}{\Gamma(x+k)} (1+\beta)^k \left(\frac{1+\beta}{\beta}\right)^x \\ &= \frac{1}{\Gamma(x+k)} \left(\frac{\beta+1}{\beta}\right)^{k+x} \theta^{k+x-1} \mathrm{e}^{-\theta(\beta+1)/\beta}, \end{aligned}$$

hvilket netop er tætheden for en $\mathrm{G}(x+k, \frac{\beta}{\beta+1})$-fordeling. De øvrige vises analogt. ■

BEMÆRKNING 6.1. Hvis man indfører vendingen **præcisionen** for den reciprokke varians, ser vi, at a posteriori præcisionen i det normale tilfælde er summen af a priori præcisionen og observationens præcision, og a posteriori middelværdien er det vejede gennemsnit af a priori middelværdien og observationen med præcisionerne som vægte. Vi ser i øvrigt, at a posteriori fordelingen i alle tre tilfælde er af samme art som a priori fordelingen. En a priori fordeling og en stikprøvefordeling, der har denne egenskab, kaldes **konjugerede**. ▼

Vi vender nu tilbage til vort oprindelige beslutningsproblem. Hvis vi er i besiddelse af en a priori viden om Θ, som kan udtrykkes ved en a priori fordeling g, vil det naturligvis ikke være rimeligt at se bort fra denne merinformation, når man skal træffe en beslutning. Vi vil naturligt foretrække strategier d, som er sådan beskafne, at risikoen $\mathrm{R}(\theta, d)$ i en eller anden forstand er lille i punkter θ, der har stor sandsynlighed for at indtræffe. Vi indfører derfor den såkaldte **Bayes-risiko**

$$\mathrm{r}(g,d) = \mathrm{E}\big(\mathrm{R}(\Theta,d)\big) = \begin{cases} \int_{\Omega} \mathrm{R}(\theta,d)\,\mathrm{g}(\theta)\,d\theta & \Theta \text{ af type K} \\ \sum_{\theta} \mathrm{R}(\theta,d)\,\mathrm{g}(\theta) & \Theta \text{ af type D} \end{cases} \tag{6.13}$$

d.v.s. Bayes-risikoen er middelrisikoen, hvor vi altså har taget forventningsværdi over de mulige parameterværdier.

DEFINITION 6.3. Ved en **Bayes strategi** d_1 forstår vi en strategi, der minimaliserer Bayes-risikoen, d.v.s. d_1 er en Bayes strategi, hvis og kun hvis

$$\mathrm{r}(g,d_1) = \inf_d \mathrm{r}(g,d).$$

▲

Hvis man leder efter en Bayes strategi, skal man altså minimalisere $\mathrm{r}(g,d)$. Dette er ofte besværligt, hvorfor vi omskriver $\mathrm{r}(g,d)$. Vi betragter f.eks. tilfældet, hvor X er

kontinuert, og Θ er kontinuert. Vi har da, idet vi udelader grænserne i integralerne, at

$$\begin{aligned}
\mathrm{r}(g,d) &= \int \mathrm{R}(\theta,d)\mathrm{g}(\theta)\,d\theta \\
&= \iint \mathrm{L}\big(\theta,\mathrm{d}(x)\big)\mathrm{f}(x|\theta)\,dx\,\mathrm{g}(\theta)\,d\theta \\
&= \iint \mathrm{L}\big(\theta,\mathrm{d}(x)\big)\mathrm{f}(x|\theta)\mathrm{g}(\theta)\,d\theta\,dx \\
&= \int \mathrm{h}(x)\int \mathrm{L}\big(\theta,\mathrm{d}(x)\big)\mathrm{k}(\theta|x)\,d\theta\,dx \\
&= \int h(x)\,\mathrm{E}_x\Big(\mathrm{L}\big(\Theta,\mathrm{d}(x)\big)\Big)\,dx.
\end{aligned}$$

Vi har her anvendt (6.12) og forudsat, at det er tilladeligt at ombytte integrationsordenen. Vi ser nu, at et d, der for ethvert x minimaliserer det inderste integral, naturligvis også minimaliserer dobbeltintegralet og dermed $\mathrm{r}(g,d)$. Vi har derfor følgende

SÆTNING 6.1. Under visse (milde) regularitetsforudsætninger bestemmes Bayes strategien punktvis ved at minimalisere

$$\mathrm{E}_x\Big(\mathrm{L}\big(\Theta,\mathrm{d}(x)\big)\Big) = \begin{cases} \int \mathrm{L}\big(\theta,\mathrm{d}(x)\big)\mathrm{k}(\theta|x)\,d\theta & \theta \text{ af type K} \\ \sum_\theta \mathrm{L}\big(\theta,\mathrm{d}(x)\big)\mathrm{k}(\theta|x) & \Theta \text{ af type D} \end{cases} \tag{6.14}$$

BEMÆRKNING 6.2. Fodtegnet x i E_x angiver, at der er tale om middelværdien i a posteriori fordelingen. Hvis vi udtrykker (6.14) verbalt, får vi, at Bayes strategien er den strategi, der for ethvert x minimaliserer det forventede tab. ▼

Vi vil nu illustrere de indførte begreber ved hjælp af et større eksempel. Da eksemplet er ret omfattende, formulerer vi det som et særligt afsnit.

6.1.2 Eksempel på analyse af et beslutningsproblem

Det eksempel, vi skal gennemgå, er en simplificeret version af et case fra [45].

Den texanske oil wildcatter[1] Mr. Jones står i den situation at skulle vælge mellem enten selv at bore efter olie på en ranch, han har anskaffet, eller sælge sine boringsrettigheder.

[1] Wildcatter: A person who drills for oil in territory not known to be oil bearing, p. 2092 Webster's New Twentieth Century Dictionary of the English Language, sec. ed., Cleveland and New York 1971.

Vi kalder disse mulige aktioner a_1 og a_2, i.e.

a_1 : at bore selv
a_2 : at sælge boringsrettighederne.

Nu afhænger gevinsten ved at bore naturligvis af den mængde olie, der er det pågældende sted; men for simpelheds skyld betragter vi kun to tilstande, nemlig

θ_1 : ingen olie til stede
θ_2 : olie til stede.

Man regner a priori med, at sandsynligheden for, at der er olie det pågældende sted, er 20%. Denne a priori sandsynlighed er resultatet af en mængde overvejelser vedrørende olieforekomster i nærheden, områdets geologiske karakter etc. Vi har altså a priori fordelingen

$$\begin{aligned} \mathrm{g}(\theta_1) &= 0.8 \\ \mathrm{g}(\theta_2) &= 0.2 \end{aligned}$$

Mr. Jones, som er en erfaren oliemand, regner med, at han taber ca. $1.500.000\$$, hvis han sælger olierettighederne, og der findes olie. Dette tab er et såkaldt **opportunity loss** (eller **regret**), idet der jo ikke er tale om direkte omkostninger, men betydelige indtægter, dels fra salget af borerettighederne, dels fra afgifter af den indvundne olie. Tabet er blot et udtryk for den sum, han kunne have tjent mere, hvis han havde valgt den i denne situation optimale beslutning, nemlig at bore selv. Tabet ved den optimale beslutning sættes derfor lig 0 her.

Hvis der ingen olie er, lider Mr. Jones nogle konkrete tab. Hvis han sælger sine rettigheder, taber han $1.000.000\$$ (bl.a. grundet den merpris, han har givet ud over ranchens landbrugsmæssige værdi af den formodede tilstedeværelse af olie), og hvis han borer, har han yderligere omkostninger på $2.000.000\$$.

Mr. Jones vælger derfor at træffe sine videre beslutninger på basis af tabsfunktionen L givet ved

$\mathrm{L}(\theta, \alpha)$	a_1	a_2
	bore selv	sælge
θ_1 (ingen olie)	300	100
θ_2 (olie)	0	150

Vi skal ikke komme nærmere ind på at begrunde rimeligheden af (eller det urimelige i) Mr. Jones' valg af tabsfunktion. Vi nøjes med at konstatere, at det generelt er meget

vanskeligt at opstille en sådan, idet der jo ofte kræves, at man skal kunne vurdere de økonomiske konsekvenser af ikke trufne beslutninger under forhold, om hvilke man ikke ved, om de opstår.

Før Mr. Jones tager sin beslutning, kan han få foretaget en række seismologiske undersøgelser til yderligere belysning af forholdene. De mulige udfald af en sådan forsøgsrække er

x_1 : der afsløres ingen struktur,

x_2 : der er en åben struktur,

x_3 : der er en lukket struktur.

Erfaringen viser, at man, hvis der ingen olie er, har, at sandsynligheden for at få x_1 henholdsvis x_2 og x_3 er 0.68 henholdsvis 0.28 og 0.04. Såfremt der er olie i undergrunden, er de tilsvarende sandsynligheder 0.30, 0.35 og 0.35.

Frekvensfunktionen $\mathrm{f}(x|\theta)$ for udfaldet af eksperimentet er altså givet ved

$\mathrm{f}(x\|\theta)$	x_1	x_2	x_3
θ_1	0.68	0.28	0.04
θ_2	0.30	0.35	0.35

Under forudsætning af, at Mr. Jones lader forsøgsrækken udføre, vil vi nu bestemme hans minimax og hans Bayes strategi.

Da der er to mulige aktioner og 3 mulige udfald af eksperimentet, er der $2^3 = 8$ mulige beslutningsfunktioner, nemlig

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
x_1	a_1	a_1	a_1	a_1	a_2	a_2	a_2	a_2
x_2	a_1	a_1	a_2	a_2	a_1	a_1	a_2	a_2
x_3	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2

Vi skal nu bestemme risiciene

$$\mathrm{R}(\theta_i, d_j) \quad = \quad \mathrm{E}_{\theta_i}\Big(\mathrm{L}\big(\theta_i, \mathrm{d}_j(X)\big)\Big) = \sum_{\nu=1}^{3} \mathrm{L}\big(\theta_i, \mathrm{d}_j(x_\nu)\big)\mathrm{f}(x_\nu|\theta_i).$$

Vi har f.eks.

$$\mathrm{R}(\theta_1, d_3) \quad = \quad 300 \cdot 0.68 + 100 \cdot 0.28 + 300 \cdot 0.04 = 244.$$

På tilsvarende måde beregnes de øvrige risici, og vi samler resultaterne i nedenstående skema.

$\mathrm{R}(\theta, d)$	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
θ_1	300	292	244	236	164	156	108	100
θ_2	0	52.5	52.5	105	45	97.5	97.5	150
Max	300	292	244	236	164	156	108	150

Heraf ses, at

$$\max_{\theta} \mathrm{R}(\theta, d_7) = \min_{d} \max_{\theta} \mathrm{R}(\theta, d),$$

d.v.s. minimaxstrategien er d_7. Mr. Jones skal altså kun bore selv, hvis der konstateres en lukket struktur ved de seismologiske undersøgelser. I de øvrige tilfælde skal han sælge.

Vi vil nu bestemme Bayes strategien. Vi finder først a posteriori fordelingen for Θ, i.e.

$$\mathrm{k}(\theta_i | x_j) = \frac{\mathrm{f}(x_j|\theta_i)\mathrm{g}(\theta_i)}{\mathrm{h}(x_j)},$$

hvor

$$\mathrm{h}(x_j) = \mathrm{f}(x_j|\theta_1)\mathrm{g}(\theta_1) + \mathrm{f}(x_j|\theta_2)\mathrm{g}(\theta_2).$$

Vi bestemmer først h. Vi har eksempelvis

$$\mathrm{h}(x_1) = 0.68 \cdot 0.8 + 0.30 \cdot 0.2 = 0.604.$$

Analogt findes

$$\begin{aligned} \mathrm{h}(x_2) &= 0.294 \\ \mathrm{h}(x_3) &= 0.102 \end{aligned}$$

Herefter har vi e.g.

$$\mathrm{k}(\theta_2|x_1) = \frac{0.30 \cdot 0.2}{0.604} = 0.10$$

Vi samler samtlige a posteriori sandsynligheder i nedenstående tabel

$\mathrm{k}(\theta\|x)$	x_1	x_2	x_3
θ_1	0.90	0.76	0.31
θ_2	0.10	0.24	0.69

Ifølge sætning 6.1 skal vi nu for ethvert x blot bestemme den aktion, der minimaliserer

$$\mathrm{E}_x\big(\mathrm{L}(\Theta, a)\big) = \mathrm{L}(\theta_1, a)\mathrm{k}(\theta_1|x) + \mathrm{L}(\theta_2|a)\mathrm{k}(\theta_2|x)\,.$$

For $x = x_1$ og $a = a_2$ finder vi

$$\mathrm{E}_{x_1}\big(\mathrm{L}(\Theta, a_2)\big) = 100 \cdot 0.90 + 150 \cdot 0.10 = 105.$$

Analogt bestemmes de øvrige, som vi samler i

$\mathrm{E}_x\big(\mathrm{L}(\Theta, a)\big)$	x_1	x_2	x_3
a_1	270	228	**93**
a_2	**105**	**112**	134.5

Vi konstruerer nu Bayes løsningen ved for ethvert x at vælge det a, der giver det mindste middeltab (markeret ved en firkant $\boxed{}$ i tabellen). Det ses, at Bayes løsningen er givet ved

$$\mathrm{d}(x_1) = a_2\,, \qquad \mathrm{d}(x_2) = a_2\,, \qquad \mathrm{d}(x_3) = a_1\,,$$

d.v.s. $d = d_7$, den samme som minimaxløsningen.

Vi ser nu i øvrigt let, at den minimale Bayes-risiko er

$$\mathrm{r}(g, d_7) = 105 \cdot 0.604 + 112 \cdot 0.294 + 93 \cdot 0.102 = 105.8.$$

Hvis Mr. Jones ikke lader foretage noget eksperiment, kan han beregne det forventede tab ved hjælp af a priori fordelingen for Θ i stedet for a posteriori fordelingen. Vi får da

$$\mathrm{E}_g\big(\mathrm{L}(\Theta, a_i)\big) = \mathrm{L}(\theta_1, a_i)\mathrm{g}(\theta_1) + \mathrm{L}(\theta_2, a_i)\mathrm{g}(\theta_2),$$

d.v.s.

$$\begin{aligned}\mathrm{E}_g\big(\mathrm{L}(\Theta, a_1)\big) &= 300 \cdot 0.8 + 0 \cdot 0.2 = 240\\ \mathrm{E}_g\big(\mathrm{L}(\Theta, a_2)\big) &= 100 \cdot 0.8 + 150 \cdot 0.2 = 110.\end{aligned}$$

Hvis Mr. Jones intet eksperiment foretager, da er det altså aktionen a_2, der giver det mindste forventede tab.

Forskellen mellem de forventede tab i de to situationer er

$$\mathrm{E}_g\big(\mathrm{L}(\Theta, a_2)\big) - \mathrm{E}_x\big(\mathrm{L}(\Theta, d_2)\big) = 110 - 105.8 = 4.2,$$

og man kan da sige, at de 4.2 er **øvre grænse for den pris, vi er villige til at betale for eksperimentet for at få den merinformation, der ligger i eksperimentets udfald**. I det aktuelle tilfælde må vi derfor råde Mr. Jones til at lade eksperimentet udføre, hvis dette er billigere end

$$4.2 \cdot 10.000\$ = 42.000\$.$$

En analyse af denne slags kaldes en **præ posteriori analyse**. Den udvikles i det generelle tilfælde naturligvis til at afgøre, hvilket af flere mulige eksperimenter man skal vælge.

Vi forlader nu eksemplet og de almindelige betragtninger om den statistiske beslutningsteori og vender os mod anvendelsen af beslutningsteorien i estimations- og testproblemer.

6.2 Beslutningsteoriens anvendelse i statistikken

Det er åbenbart, at man kan opfatte et estimations- eller et testproblem som et beslutningsproblem af en karakter som omtalt i det foregående afsnit. Skal man f.eks. estimere en parameter θ på basis af målinger $X_1, \cdots, X_n$, er estimatoren $t(X_1, \cdots, X_n)$, hvor t afbilder $\mathrm{R}^n \to \Omega = \{\theta | \theta \in \Omega\}$ jo en beslutningsfunktion som omtalt p. 516. Skal man teste $H_0 \colon \theta = \theta_0$ mod $H_1 \colon \theta = \theta_1$, skal man jo vælge enten θ_0 eller θ_1 på basis af udfaldet af stokastiske variable $X_1, \cdots, X_n$, d.v.s. at vi igen har en beslutningssituation. Vi skal nu vise, hvorledes tilstedeværelsen af en tabsfunktion muliggør konstruktion af andre estimatorer eller muliggør fastlæggelsen af et rimeligt niveau. Vi indleder med

6.2.1 Beslutningsteoriens anvendelse i estimationsteorien

I mange estimationsproblemer vil en tabsfunktion afhænge af afstanden mellem estimatet $\mathrm{t}(x) = \mathrm{t}(x_1, \cdots, x_n)$ og den sande parameter θ. De to simpleste former for tabsfunktioner vil da være

$$\mathrm{L}_1\big(\theta, \mathrm{t}(x)\big) = c|\theta - \mathrm{t}(x)|$$

og

$$\mathrm{L}_2\big(\theta, \mathrm{t}(x)\big) = c\big(\theta - \mathrm{t}(x)\big)^2,$$

hvor c er en positiv konstant og θ en endimensional parameter. Vi vil nu søge at bestemme minimax- og Bayes løsninger svarende til disse tabsfunktioner. Beslutningsfunktionerne er her estimatorerne $\mathrm{t}(X) = \mathrm{t}(X_1, \cdots, X_n)$, og en aktion er et valg af en bestemt værdi af θ. Vi har risikofunktionerne

$$\mathrm{R}_1(\theta, t) = \mathrm{E}_\theta\big(c|\theta - \mathrm{t}(X)|\big)$$

og

$$\mathrm{R}_2(\theta, t) = \mathrm{E}_\theta\Big(c\big(\theta - \mathrm{t}(X)\big)^2\Big),$$

og det er klart, at minimax estimatorerne bestemmes ved at minimalisere

$$\sup_\theta \mathrm{E}_\theta\big(|\theta - \mathrm{t}(X)|\big)$$

henholdsvis

$$\sup_{\theta} \mathrm{E}_\theta\Big(\big(\theta - \mathrm{t}(X)\big)^2\Big).$$

Uden yderligere forudsætninger om, hvilke estimatorer t vi betragter, og hvilke fordelinger der er involveret, er det ikke muligt at opstille eksplicitte udtryk for minimaxløsningerne.

Vi vender os derfor til et konkret tilfælde i følgende eksempel.

EKSEMPEL 6.2. Vi betragter uafhængige stokastiske variable $X_1, \cdots, X_n$ med $\mathrm{E}(X_i) = \theta$ og $\mathrm{V}(X_i) = \sigma^2$, $i = 1, \cdots, n$. Vi vil anvende den kvadratiske tabsfunktion

$$\mathrm{L}\big(\theta, \mathrm{t}(x_1, \cdots, x_n)\big) = \mathrm{L}\big(\theta, \mathrm{t}(x)\big) = \big(\theta - \mathrm{t}(x)\big)^2.$$

Vi vil kun betragte estimatorer, der er **lineære** i $X_1, \cdots, X_n$, d.v.s. af formen

$$\mathrm{t}(x) = \mathrm{t}(x_1, \cdots, x_n) = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n. \tag{6.15}$$

Vi vil nu bestemme minimaxestimatoren af denne form. Vi har, at

$$\mathrm{E}\big(\mathrm{t}(X)\big) = \mathrm{E}(\alpha_0 + \alpha_1 X_1 + \cdots + \alpha_n X_n) = \alpha_0 + \alpha_1\theta + \cdots + \alpha_n\theta$$

og

$$\mathrm{V}\big(\mathrm{t}(X)\big) = \mathrm{V}(\alpha_0 + \alpha_1 X_1 + \cdots + \alpha_n X_n) = \alpha_1^2\sigma^2 + \cdots + \alpha_n^2\sigma^2.$$

Sætter vi $\alpha_1 + \cdots + \alpha_n = \beta$ og $\alpha_1^2 + \cdots + \alpha_n^2 = \delta^2$, har vi altså

$$\begin{aligned} \mathrm{E}\big(\mathrm{t}(X)\big) &= \alpha_0 + \beta\theta \\ \mathrm{V}\big(\mathrm{t}(X)\big) &= \delta^2\sigma^2. \end{aligned}$$

Følgelig er

$$\begin{aligned} \mathrm{R}(\theta, t) &= \mathrm{E}_\theta\Big(\big(\theta - \mathrm{t}(X)\big)^2\Big) \\ &= \mathrm{E}_\theta\big(\theta^2 + \mathrm{t}(X)^2 - 2\theta\mathrm{t}(X)\big) \\ &= \theta^2 + \mathrm{E}_\theta\big(\mathrm{t}(X)^2\big) - 2\theta\,\mathrm{E}_\theta\big(\mathrm{t}(X)\big) \\ &= \theta^2(\beta - 1)^2 + \theta 2\alpha_0(\beta - 1) + \delta^2\sigma^2 + \alpha_0^2. \end{aligned}$$

Vi ser nu, at $\mathrm{R}(\theta, t)$ er ubegrænset for alle beslutningsfunktioner, for hvilke $\beta \neq 1$. En minimaxløsning t må derfor tilfredsstille $\beta = 1$, d.v.s.

$$\alpha_1 + \cdots + \alpha_n = 1.$$

For et sådant t haves

$$\mathrm{R}(\theta, t) = \delta^2\sigma^2 + \alpha_0^2 = (\alpha_1^2 + \cdots + \alpha_n^2)\sigma^2 + \alpha_0^2.$$

Det er klart, at α_0 må være 0, og at $\alpha_1, \cdots, \alpha_n$ bestemmes ved at minimalisere

$$\mathrm{g}(\alpha_1, \cdots, \alpha_n) = \alpha_1^2 + \cdots + \alpha_n^2$$

under bibetingelsen

$$\alpha_1 + \cdots + \alpha_n = 1.$$

At bestemme dette minimum er en simpel øvelse i anvendelse af teorien for Lagrange multiplikatorer. Vi anfører betragtningerne for fuldstændighedernes skyld. Vi får, at α_i, $i = 1, \cdots, n$, skal tilfredsstille

$$\frac{\partial g}{\partial \alpha_i} + \lambda = 0 \qquad i = 1, \cdots, n$$

$$\alpha_1 + \cdots + \alpha_n = 1,$$

hvor λ er Lagrange multiplikatoren. Dette giver

$$2\alpha_i + \lambda = 0 \qquad i = 1, \cdots, n$$

$$\alpha_1 + \cdots + \alpha_n = 1.$$

Ved addition af de første n ligninger fås

$$2(\alpha_1 + \cdots + \alpha_n) + n\lambda = 1$$

d.v.s.

$$\lambda = -\frac{2}{n}.$$

Dette giver løsningen

$$\alpha_1 = \cdots = \alpha_n = \frac{1}{n}.$$

Det verificeres nu let, at dette virkelig er et globalt minimumspunkt for g, således at vi har fået vist, at minimaxestimatoren er $\bar{X}$. ♦

Vi resumerer indholdet af ovenstående eksempel i

SÆTNING 6.2. Lad $X_1, \cdots, X_n$ være uafhængige stokastiske variable med $\mathrm{E}(X_i) = \theta$ og $\mathrm{V}(X_i) = \sigma^2$. Hvis vi anvender en kvadratisk tabsfunktion og kun betragter estimatorer, der er lineære i $X_1, \cdots, X_n$, er minimaxestimatoren for θ lig $\bar{X} = \frac{1}{n}\sum X_i$.

Vi vil nu betragte Bayes estimatorer. Det er her noget lettere at angive præcise udtryk for estimatorerne. Vi har altså nemlig følgende

SÆTNING 6.3. Vi betragter uafhængige identisk fordelte stokastiske variable $X_1, \cdots, X_n$, der for givet $\Theta = \theta$ har frekvensfunktionen $\mathrm{f}(x|\theta)$. Den stokastiske variable Θ har a priori fordeling med frekvensfunktion $\mathrm{g}(\theta)$. Hvis vi anvender den absolutte tabsfunktion (6.2.1), er Bayes estimatoren for θ lig enhver median i a posteriori fordelingen for Θ, og hvis vi anvender den kvadratiske tabsfunktion (6.2.1), er Bayes estimatoren lig middelværdien i a posteriori fordelingen.

BEMÆRKNING 6.3. Vi må her indskyde, at sætningen ikke er fuldstændigt formuleret. Vi må bl.a. forudsætte, at de p. 521 anførte ombytninger af integrationsordenen er tilladte. ▼

Bevis. Vi viser først resultatet vedrørende den kvadratiske tabsfunktion. Vi har, at Bayes strategien bestemmes ved at minimalisere

$$\mathrm{E}_x\Big(\big(\Theta - \mathrm{t}(x)\big)^2\Big)$$

hvor E_x altså angiver middelværdien i a posteriori fordelingen. Vi har imidlertid

$$\begin{aligned}
\mathrm{E}_x\Big(\big(\Theta-\mathrm{t}(x)\big)^2\Big) &= \mathrm{E}_x\Big(\big(\Theta-\mathrm{E}_x(\Theta)+\mathrm{E}_x(\Theta)-\mathrm{t}(x)\big)^2\Big)\\
&= \mathrm{E}_x\Big(\big(\Theta-\mathrm{E}_x(\Theta)\big)^2\Big)+\mathrm{E}_x\Big(\big(\mathrm{E}_x(\Theta)-\mathrm{t}(x)\big)^2\Big)\\
&\quad +2\,\mathrm{E}_x\Big(\big(\Theta-\mathrm{E}_x(\Theta)\big)\Big)\big(\mathrm{E}_x(\Theta)-\mathrm{t}(x)\big)\\
&= \mathrm{E}_x\Big(\big(\Theta-\mathrm{E}_x(\Theta)\big)^2\Big)+\mathrm{E}_x\Big(\big(\mathrm{E}_x(\Theta)-\mathrm{t}(x)\big)^2\Big)\\
&= \mathrm{V}_x(\Theta)+\big(\mathrm{E}_x(\Theta)-\mathrm{t}(x)\big)^2.
\end{aligned}$$

Da det første led er uafhængigt af t, ses det, at vi får minimum for

$$\mathrm{t}(x)=\mathrm{E}_x(\Theta)=\begin{cases}\int \theta \mathrm{k}(\theta|x)\,dx\\ \sum \theta \mathrm{k}(\theta|x)\end{cases},$$

hvilket skulle vises.

Ved lidt mere indviklede betragtninger følger tilsvarende, at

$$\mathrm{E}_x\big(|\theta-\mathrm{t}(x)|\big)$$

antager sin minimumsværdi i m_x bestemt ved

$$\mathrm{P}\{\Theta<m_x|X=x\}\le\frac{1}{2}\le\mathrm{P}\{\Theta\le m_x|X=x\},$$

d.v.s. m_x = medianen i a posteriori fordelingen (se f.eks. [8] p. IX 7). ■

Vi giver et eksempel til yderligere belysning af forholdene.

EKSEMPEL 6.3. En bygmester, der modtager alle sine sten fra samme teglværk, har gennem lang tid konstateret, at defektprocenten blandt de leverede partier er stærkt svingende. Den hyppigste defektprocent (modus) er ca. 1%, og middeldefektprocenten er ca. 2%. Han modtager nu et nyt parti sten og konstaterer, at der er én dårlig ud af 5, han tilfældigt udvælger. Han kunne derfor estimere defektprocenten θ i dette parti ved den relative hyppighed af defekte, i.e. ved $\frac{1}{5}=20\%$. Dette er selvsagt ikke særlig rimeligt, mandens a priori viden taget i betragtning. Vi vil søge at formulere en a priori fordeling for defektprocenten og dernæst bestemme en Bayes estimator, idet vi vil anvende en kvadratisk tabsfunktion.

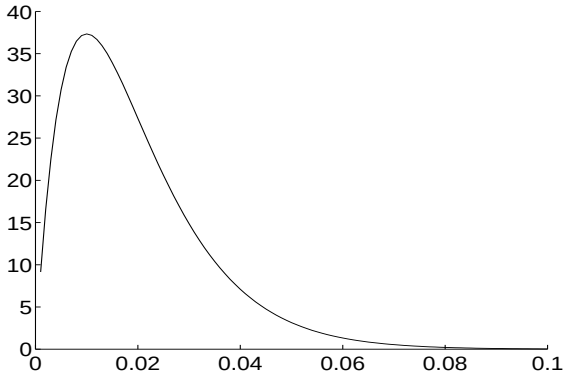
Det er ikke muligt uden videre at opstille en kanonisk model for Θ's fordeling. Da Θ er en sandsynlighed, søger vi derfor blot en fordeling på intervallet $[0,1]$, der virker rimelig, og som har den anførte middelværdi og modus. Det vil derfor være naturligt at tænke på en beta-fordeling. En $\mathrm{Be}(\alpha,\beta)$-fordeling har middelværdi $\alpha/(\alpha+\beta)$ og modus $(\alpha-1)/(\alpha+\beta-2)$ (ses ved differentiation af tætheden). Vi kan derfor opstille ligningerne

$$\begin{aligned} \frac{\alpha}{\alpha+\beta} &= 0.02 \\ \frac{\alpha-1}{\alpha+\beta-2} &= 0.01. \end{aligned}$$

En løsning til disse ligninger er (afrundet) $\alpha=2$ og $\beta=100$. En $\mathrm{Be}(2,100)$-fordeling har tætheden

$$\mathrm{f}(\theta)=\frac{1}{\mathrm{B}(2,100)}\theta(1-\theta)^{99}=10100\cdot\theta(1-\theta)^{99}.$$

Vi indtegner grafen for f i et koordinatsystem og får



Vi ser, at grafen vokser fra værdien 0 i punktet $\theta=0$ til ca. 37 i 0.01 og derefter aftager, til den (næsten) har nået abscissen i $\theta=0.1$. Den virker derfor yderst plausibel som a priori fordeling af defektprocenterne.

For givet $\Theta=\theta$ vil det være rimeligt at antage, at X-antallet af defekte i stikprøven på de 5 - vil følge en $\mathrm{B}(5,\theta)$-fordeling.

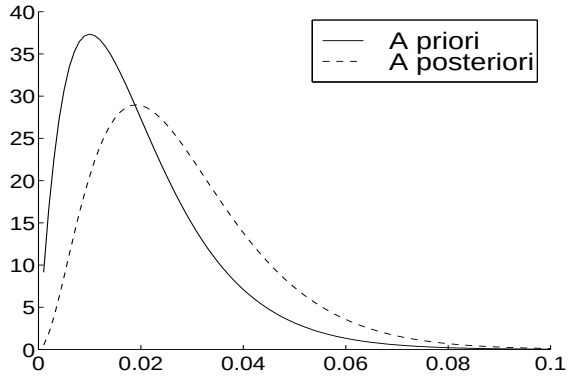
Ifølge skemaet 520 vil a posteriori fordelingen af Θ givet $X=x$ være en $\mathrm{Be}(2+x,100+5-x)$-fordeling, i.e., da vi observerede $x=1$, en $\mathrm{Be}(3,104)$-fordeling.

Middelværdien i denne fordeling er

$$\frac{3}{3+104} = 0.028,$$

og dette er ifølge sætning 6.3 Bayes estimatet for θ. Dette resultat virker langt mere plausibelt end de 20%. Størrelsesordenen af de 0.028 er den samme, som vi er vant til for de øvrige defektprocenter (1–2%), men selvfølgelig større end middelværdien 2%, da vi jo rent faktisk har en stor relativ hyppighed af defekte i vor stikprøve.

I nedenstående graf har vi indtegnet a posteriori tætheden for Θ.

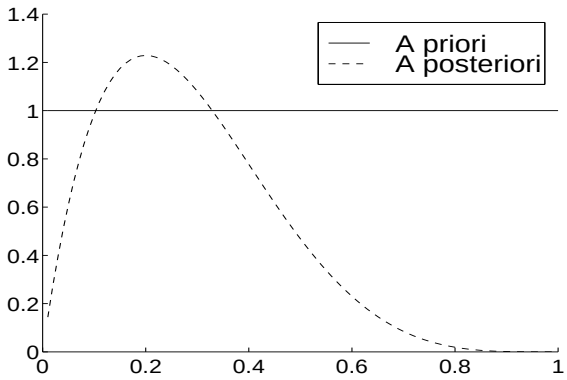


Vi ser, at tætheden er rykket til højre i forhold til a priori tætheden. Endvidere har vi fået en større spredning svarende til, at vi har fået en "ekstrem" observation.

Vi vil nu forsøge at få et overblik over betydningen af valget af a priori fordeling, idet vi vil undersøge, hvad der sker, hvis vi i stedet for en $\mathrm{Be}(2,100)$-fordeling havde valgt en $\mathrm{Be}(1,1)$-fordeling, i.e. en rektangulær fordeling. Vi får da en $\mathrm{Be}(1+1,1+5-1)$-fordeling som a posteriori fordeling, i.e. en $\mathrm{Be}(2,5)$-fordeling. En sådan har tætheden

$$\mathrm{k}(\theta|x) = \frac{1}{\mathrm{B}(2,5)}\theta^1(1-\theta)^4 = 15\theta(1-\theta)^4.$$

Graferne for de to fordelinger er indtegnet i nedenstående figur.



Vi bemærker, at der er en meget stor forskel mellem a posteriori fordelingen her og så a posteriori fordelingen i eksemplets første halvdel. Dette er et udtryk for, at overgangen fra a priori til a posteriori fordeling er "træg", d.v.s. a posteriori fordelingen ligger - naturligt nok - "nær" ved a priori fordelingen.

Med det noget særprægede valg af a priori fordeling bliver Bayes estimatet for θ nu lig $2/7 = 0.29$, hvilket jo igen ligger langt fra det først opnåede Bayes estimat. ♦

Med ovenstående eksempel skulle det være antydet, at valget af a priori fordeling (selvfølgelig) er meget vigtigt. Man kan ikke blot vælge en temmelig vilkårlig og så gå ud fra, at modificeringen til a posteriori fordeling "klarer skærene", således at man alligevel får fornuftige estimater.

Vi går nu over til at behandle beslutningsteoriens anvendelser i testteorien.

6.2.2 Beslutningsteoriens anvendelse i testteorien

Vi skal i det følgende kun beskæftige os med test af en simpel hypotese

$$H_0: \quad \theta = \theta_0$$

mod et simpelt alternativ

$$H_1: \quad \theta = \theta_1,$$

d.v.s. vi har parameterrummet

$$\Omega = \{\theta_0, \theta_1\}$$

og vi ønsker på basis af de uafhængige stokastiske variable $X_1, \cdots, X_n$ at teste H_0 mod H_1.

$X_1, \cdots, X_n$ har den simultane frekvensfunktion

$$\mathrm{f}(x|\theta) = \mathrm{f}(x_1, \cdots, x_n|\theta).$$

Der er de mulige aktioner

a_0: at vælge θ_0
a_1: at vælge θ_1

hvor a_0 altså svarer til det vanlige begreb at acceptere en hypotese etc. For at kunne behandle problemerne som beslutningsproblemer er det naturligvis nødvendigt at have en tabsfunktion L. Vi forudsætter, at der ikke er noget tab ved at vælge den rette parameter. Vi har derfor

$$\begin{aligned} \mathrm{L}(\theta_0, a_0) &= \mathrm{L}(\theta_1, a_1) = 0 \\ \mathrm{L}(\theta_0, a_1) &= L_0 \\ \mathrm{L}(\theta_1, a_0) &= L_1 \end{aligned}$$

Da der hersker en éntydig korrespondance mellem det kritiske område C for et test og selve testet, kan vi fastlægge beslutningsfunktionerne ved at specificere det C, de derved svarer til. Vi har altså for ethvert C beslutningsfunktionen

$$\mathrm{d}_C(x) = \begin{cases} a_0, & \text{hvis } x \in \complement C \\ a_1, & \text{hvis } x \in C \end{cases}$$

Vi kan nu vise følgende sætning om minimaxløsningen til ovenstående beslutningsproblem.

SÆTNING 6.4. Lad situationen være som ovenfor beskrevet. Da er minimaxtestet givet ved det kritiske område

$$C = \left\{ (x_1, \cdots, x_n) | \frac{\mathrm{f}(x_1, \cdots, x_n|\theta_0)}{\mathrm{f}(x_1, \cdots, x_n|\theta_1)} \leq c \right\},$$

hvor c fastlægges ved

$$L_1 \mathrm{P}\{X \in \complement C|\theta = \theta_1\} = L_0 \mathrm{P}\{X \in C|\theta = \theta_0\}.$$

Bevis. Vi gennemgår beviset i det tilfælde, hvor X er kontinuert fordelt.

Vi har risikoen $\mathrm{R}(\theta, d)$. Den er givet ved

$$\begin{aligned}\mathrm{R}(\theta_0, d) &= \mathrm{E}_{\theta_0}\Big(\mathrm{L}\big(\theta_0, \mathrm{d}(X)\big)\Big) \\ &= \int_{x\in C} \mathrm{L}\big(\theta_0, \mathrm{d}(x)\big)\mathrm{f}(x|\theta_0)\, dx \\ &\quad + \int_{x\in \complement C} \mathrm{L}\big(\theta_0, \mathrm{d}(x)\big)\mathrm{f}(x|\theta_0)\, dx \\ &= L_0 \mathrm{P}\{X \in C|\theta = \theta_0\} + 0 \cdot \mathrm{P}\{X \in \complement C|\theta = \theta_0\}.\end{aligned}$$

Analogt finder vi

$$\mathrm{R}(\theta_1, d) = L_1 \mathrm{P}\{X \in \complement C|\theta = \theta_1\}.$$

Vi forudsætter nu, at d er en minimaxstrategi. Hvis vi sætter

$$\mathrm{P}\{X \in C|\theta = \theta_0\} = \alpha,$$

og hvis vi for et andet A har

$$\mathrm{P}\{X \in A|\theta = \theta_0\} = \alpha,$$

er

$$\mathrm{P}\{X \in \complement A|\theta = \theta_1\} > \mathrm{P}\{X \in \complement C|\theta = \theta_1\},$$

thi ellers ville A svare til en minimaxløsning. Denne relation giver ved overgang til komplementærmængderne

$$\mathrm{P}\{X \in A|\theta = \theta_1\} < \mathrm{P}\{X \in C|\theta = \theta_1\},$$

og ifølge definition 3.7, p. 311, er C et bedste kritisk område. Formen for et sådant er givet ved Neyman-Pearson's lemma. Tilbage bliver alene at bestemme c; men det er trivielt, at det skal vælges, så

$$L_1 \mathrm{P}\{X \in \complement C|\theta = \theta_1\} = L_0 \mathrm{P}\{X \in C|\theta = \theta_0\},$$

idet funktionen på højresiden er voksende som funktion af C og venstresiden aftagende. Maximum af disse 2 må derfor minimaliseres, når de er lig hinanden. ■

Vi forudsætter nu endvidere, at Θ har en a priori fordeling $\mathrm{g}(\theta)$, $\theta = \theta_0, \theta_1$. Vi kan da på sædvanlig vis beregne a posteriori fordelingen $\mathrm{k}(\theta|x)$ og får

$$\mathrm{k}(\theta|x) = \frac{\mathrm{f}(x|\theta)\,\mathrm{g}(\theta)}{\mathrm{h}(x)}.$$

Vi kan nu bestemme Bayes strategien. Vi har nemlig

SÆTNING 6.5. Lad situationen være som ovenfor beskrevet. Da er Bayes strategien givet ved det kritiske område

$$C = \left\{ (x_1, \cdots, x_n) | \frac{\mathrm{f}(x_1, \cdots, x_n|\theta_0)}{\mathrm{f}(x_1, \cdots, x_n|\theta_1)} \leq \frac{\mathrm{g}(\theta_1)L_1}{\mathrm{g}(\theta_0)L_0} \right\}.$$

Bevis. I følge sætning 6.1 bestemmes Bayes strategien ved at minimalisere

$$\begin{aligned} \mathrm{E}_x\Big(\mathrm{L}\big(\Theta, \mathrm{d}_C(x)\big)\Big) &= \mathrm{L}\big(\theta_0, d_C(x)\big)\mathrm{k}(\theta_0|x) + \mathrm{L}\big(\theta_1, \mathrm{d}_C(x)\big)\mathrm{k}(\theta_1|x) \\ &= \begin{cases} L_1\mathrm{k}(\theta_1|x) & x \in \complement C \\ L_0\mathrm{k}(\theta_0|x) & x \in C. \end{cases} \end{aligned}$$

Hvis denne størrelse skal minimaliseres for ethvert x, må vi definere C ved

$$\begin{aligned} x \in C \quad &\Leftrightarrow \quad L_0\mathrm{k}(\theta_0|x) \leq L_1\mathrm{k}(\theta_1|x) \\ &\Leftrightarrow \quad \frac{L_0\mathrm{f}(x|\theta_0)\mathrm{g}(\theta_0)/\mathrm{h}(x)}{L_1\mathrm{f}(x|\theta_1)\mathrm{g}(\theta_1)/\mathrm{h}(x)} \leq 1 \\ &\Leftrightarrow \quad \frac{\mathrm{f}(x|\theta_0)}{\mathrm{f}(x|\theta_1)} \leq \frac{\mathrm{g}(\theta_1)L_1}{\mathrm{g}(\theta_0)L_0} \\ &\Leftrightarrow \quad \frac{\mathrm{f}(x_1, \cdots, x_n|\theta_0)}{\mathrm{f}(x_1, \cdots, x_n|\theta_1)} \leq \frac{\mathrm{g}(\theta_1)L_1}{\mathrm{g}(\theta_0)L_0}. \end{aligned}$$

■

BEMÆRKNING 6.4. Vi konstaterer, at såvel minimaxtestet som Bayes testet er bedste tests ifølge Neyman-Pearson's lemma. De adskiller sig alene ved konstanten c, d.v.s. ved niveauet. ▼

Vi vil nu diskutere og illustrere de forskellige aspekter ved disse to sætninger ved hjælp af et konkret eksempel.

EKSEMPEL 6.4. Vi betragter indbyrdes uafhængige identisk Poisson fordelte stokastiske variable $X_1, \cdots, X_n$, hvor altså $X_i \in \mathrm{P}(\theta)$. Vi ønsker at teste

$$H_0: \quad \theta = \theta_1 = 1 \qquad \text{mod} \qquad H_1: \quad \theta = \theta_1 = 2.$$

Vi har med den i det foregående anvendte symbolik tabsfunktionen L givet ved

$$\begin{aligned} \mathrm{L}(\theta_0, a_1) &= L_0 = 6L_1 \\ \mathrm{L}(\theta_1, a_0) &= L_1, \end{aligned}$$

hvor vi som et eksempel altså har valgt at sætte $L_0 = 6L_1$. Dvs., at tabet ved at begå en såkaldt type I fejl sættes til 6 gange tabet ved at begå en type II fejl.

Da observationerne er stokastisk uafhængige, er

$$\mathrm{f}(x_1, \cdots, x_n|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} \mathrm{e}^{-\theta} = \frac{1}{\prod x_i!} \theta^{\sum x_i} \mathrm{e}^{-n\theta}.$$

For at bestemme de forskellige tests løser vi nu følgende ulighed (jfr. eksempel 3.1 p. 314)

$$\begin{aligned} \frac{\mathrm{f}(x_1, \cdots, x_n|1)}{\mathrm{f}(x_1, \cdots, x_n|2)} \leq c \quad &\Leftrightarrow \quad \frac{\mathrm{e}^{-n}}{\mathrm{e}^{-2n}} \frac{1^{\sum x_i}}{2^{\sum x_i}} \leq c \\ &\Leftrightarrow \quad \sum x_i \geq \frac{n - \log c}{\log 2}, \end{aligned}$$

d.v.s. det kritiske område er

$$C = \left\{ (x_1, \cdots, x_n) | \sum x_i \geq \frac{n - \log c}{\log 2} \right\},$$

hvor c så fastlægges som i sætning 6.4 eller sætning 6.5. Niveauet for testet bliver

$$\begin{aligned}
&\mathrm{P}\left\{\sum X_i \geq \frac{n-\log c}{\log 2}|\theta=1\right\} \\
&\quad=\mathrm{P}\left\{\mathrm{P}(n)\geq\frac{n-\log c}{\log 2}\right\} \\
&\quad\simeq\mathrm{P}\left\{\mathrm{N}(n,n)\geq\frac{n-\log c}{\log 2}\right\} \\
&\quad=\mathrm{P}\left\{\mathrm{N}(0,1)\geq\frac{n-\log c-n\log 2}{\log 2\sqrt{n}}\right\} \\
&\quad=\mathrm{P}\left\{\mathrm{N}(0,1)\geq\frac{0.31n-\log c}{0.69\sqrt{n}}\right\}.
\end{aligned}$$

Denne størrelse vil for Bayes testet gå mod 0 for $n\to\infty$, da c her er uafhængig af n. Vi ser altså, at **niveauet for Bayes testet går mod** 0 **for stikprøvestørrelsen gående mod** ∞. Dette resultat gælder også for minimaxtestet, men det er lidt vanskeligere at vise dette. Disse forhold illustrerer en af de fejl, man kan begå ved at fiksere et tests niveau til e.g. 5%, **når man har involveret omkostningsfunktioner**. ♦

Vi vil med $n=10$ observationer bestemme minimaxløsningen til testet i det følgende

EKSEMPEL 6.5. Ifølge sætning 6.4 skal vi bestemme c, så

$$L_1\mathrm{P}\{X\in\complement C|\theta=\theta_1\}=L_0\mathrm{P}\{X\in C|\theta=\theta_0\},$$

d.v.s.

$$L_1\mathrm{P}\left\{\sum X_i<c_1|\theta=2\right\}=6L_1\mathrm{P}\left\{\sum X_i\geq c_1|\theta=1\right\},$$

hvor c_1 er størrelsen $(n-\log c)/\log 2$. Vi har altså, at

$$\mathrm{P}\{\mathrm{P}(20)\leq c_1-1\}=6\mathrm{P}\{\mathrm{P}(10)\geq c_1\}.$$

Vi prøver os frem. Vi får ved hjælp af tabel

c_1	$\mathrm{P}\{\mathrm{P}(20)\leq c_1-1\}$	$6\cdot\mathrm{P}\{\mathrm{P}(10)\geq c_1\}$
15	0.105	$6\cdot 0.083=0.498$
16	0.157	$6\cdot 0.049=0.294$
17	0.221	$6\cdot 0.027=0.162$

Da første søjle er voksende i c_1 og anden søjle aftagende, ser vi, at c_1 lig 17 giver minimaxløsningen. Det kritiske område ved en minimaxstrategi er altså

$$C = \left\{ (x_1, \cdots, x_{10}) | \sum_{i=1}^{10} x_i \geq 17 \right\}.$$

Det svarer til et test på niveau

$$\mathrm{P}\left\{ \sum X_i \geq 17 | \theta = 1 \right\} = \mathrm{P}\{\mathrm{P}(10) \geq 17\} = 2.7\%,$$

d.v.s. af en størrelsesorden, vi er vant til (1–5%). ♦

Vi bestemmer nu en Bayes løsning.

EKSEMPEL 6.6. Vi antager nu, at der er givet en a priori fordeling over de mulige værdier af θ, i.e.

$$\begin{aligned} \mathrm{P}\{\Theta = 1\} &= \mathrm{g}(1) = \frac{1}{3} \\ \mathrm{P}\{\Theta = 2\} &= \mathrm{g}(2) = \frac{2}{3}. \end{aligned}$$

Ved hjælp af sætning 6.5 får vi, at konstanten i det bedste test er

$$c = \frac{2/3 \cdot L_1}{1/3 \cdot L_0} = \frac{2/3 \cdot L_1}{1/3 \cdot 6L_1} = \frac{1}{3}.$$

Heraf fås

$$c_1 = \frac{n - \log c}{\log 2} = \frac{10 + \log 3}{\log 2} = 16.0.$$

Det kritiske område ved en Bayes strategi er derfor

$$C_1 = \left\{ (x_1, \cdots, x_{10}) | \sum_{i=1}^{10} x_i \geq 16 \right\}.$$

Dette svarer til et test på niveau

$$\mathrm{P}\left\{\sum X_i \geq 16|\theta = 1\right\} = \mathrm{P}\{\mathrm{P}(10) \geq 16\} = 5\%,$$

igen et niveau af sædvanlig størrelse. Styrken for testet i $\theta = 2$ er

$$\mathrm{P}\left\{\sum X_i \geq 16|\theta = 2\right\} = \mathrm{P}\{\mathrm{P}(20) \geq 16\} = 84.3\%.$$

♦

Endelig vil vi illustrere betydningen af at vælge e.g. $\theta_0 = 1$ som hypotese og $\theta_1 = 2$ som alternativ i

EKSEMPEL 6.7. Vi så i foregående eksempel, at Bayes testet af $H_0 \colon \theta = 1$ mod $H_1 \colon \theta = 2$ svarende til et test på niveau 5%. Vi vil nu bestemme niveauet for Bayes testet af den modsatte hypotese, i.e. $H_0' \colon \theta = 2$ mod $H_1' \colon \theta = 1$. Vi har da

$$\begin{aligned} \frac{\mathrm{f}(x_1, \cdots, x_n|2)}{\mathrm{f}(x_1, \cdots, x_n|1)} \leq c \quad &\Leftrightarrow \quad \frac{\mathrm{e}^{-2n}}{\mathrm{e}^{-n}} \frac{2^{\sum x_i}}{1^{\sum x_i}} \leq c \\ &\Leftrightarrow \quad \sum x_i \leq \frac{n + \log c}{\log 2}, \end{aligned}$$

hvor c bestemmes til

$$\frac{1/3 \cdot 6L_1}{2/3 \cdot L_1} = 3,$$

d.v.s.

$$\frac{n + \log c}{\log 2} = \frac{10 + \log 3}{\log 2} = 16.0.$$

Det kritiske område er derfor

$$C_2 = \left\{ (x_1, \cdots, x_{10}) | \sum_{i=1}^{10} x_i \leq 16.0 \right\}.$$

Niveauet er

$$\mathrm{P}\left\{\sum X_i \leq 16|\theta = 2\right\} = \mathrm{P}\{\mathrm{P}(20) \leq 16\} = 22.1\%,$$

d.v.s. noget højere, end vi er vant til. På den anden side er styrken i $\theta = 1$ lig

$$\mathrm{P}\left\{\sum X_i \leq 16|\theta = 1\right\} = \mathrm{P}\{\mathrm{P}(10) \leq 16\} = 97.3\%,$$

d.v.s. acceptabelt høj.

Hvis vi på forhånd havde fikseret niveauet α af testet H_0' mod H_1' til at være så nær som muligt ved 5%, havde vi fået det kritiske område

$$C_3 = \left\{(x_1, \cdots, x_{10})|\sum_{i=1}^{10} x_i \leq c_3\right\},$$

hvor c_3 fastlægges ved

$$\mathrm{P}\left\{\sum X_i \leq c_3|\theta = 2\right\} = \mathrm{P}\{\mathrm{P}(20) \leq c_3\} \simeq 5\%.$$

Af tabel fås, at $c_3 = 12$ giver $\alpha = 3.9\%$, hvilket er det nærmeste, vi kan komme. Styrken for testet er i $\theta = 1$

$$\mathrm{P}\left\{\sum X_i \leq 12|\theta = 1\right\} = \mathrm{P}\{\mathrm{P}(10) \leq 12\} = 79.2\%.$$

Vi vil nu sammenligne de i dette og det foregående eksempel anførte tests til afgørelse af, om $\theta = 1$ eller $\theta = 2$. Vi samler resultaterne i nedenstående skema.

	Test 1		Test 2		Test 3	
	$\theta = 1$	$\theta = 2$	$\theta = 1$	$\theta = 2$	$\theta = 1$	$\theta = 2$
$\mathrm{P}(a_1) = \mathrm{P}$(vælge $\theta = 1$)	95%	15.7%	97.3%	22.1%	79.2%	3.9%
$\mathrm{P}(a_2) = \mathrm{P}$(vælge $\theta = 2$)	5%	84.3%	2.7%	77.9%	20.8%	96.1%

For bedre at kunne vurdere disse tests anfører vi igen tabsfunktionen L og a priori sandsynlighederne for Θ.

$\mathrm{L}(\theta, a)$	$\theta = 1$	$\theta = 2$
a_1=vælge $\theta = 1$	0	L_1
a_2=vælge $\theta = 2$	6 L_1	0
A priori sandsynlighed	$\frac{1}{3}$	$\frac{2}{3}$

Vi ser, at Bayes testene er næsten ens, hvad angår de involverede sandsynligheder (at de ikke er helt ens, skyldes alene, at de stokastiske variable er diskrete, således at $\mathrm{P}\{X \leq c\} \neq 1 - \mathrm{P}\{X \geq c\}$). Vi bemærker endvidere, at de begge har den egenskab, at sandsynligheden for at begå en fejl er lavest der, hvor (middel-)tabet ved at begå en fejl er størst. Denne egenskab besidder test 3 åbenbart ikke. Disse forhold kommer også til udtryk, når man betragter Bayes-risikoen for de tre tests. Vi har nemlig generelt, at Bayes-risikoen ved test af $\theta = \theta_0$ mod $\theta = \theta_1$ er

$$\mathrm{r}(g,d) = L_0 \mathrm{g}(\theta_0)\mathrm{P}\{X \in C|\theta = \theta_0\} + L_1 \mathrm{g}(\theta_1)\mathrm{P}\{X \in \complement C|\theta = \theta_1\},$$

jfr. definitionen p. 521, idet

$$\begin{aligned} \mathrm{R}(\theta_0,d) &= \mathrm{E}_{\theta_0}\Big(\mathrm{L}\big(\theta_0,\mathrm{d}(X)\big)\Big) = L_0\mathrm{P}\{X \in C|\theta = \theta_0\} \\ \mathrm{R}(\theta_1,d) &= \mathrm{E}_{\theta_1}\Big(\mathrm{L}\big(\theta_1,\mathrm{d}(X)\big)\Big) = L_1\mathrm{P}\{X \in \complement C|\theta = \theta_1\}. \end{aligned}$$

Udtrykt ved fejl af type I og type II fås derfor

$$\begin{aligned} \mathrm{r}(g,d) &= (\text{Tab ved fejl af type I})\mathrm{P}(\text{Fejl af type I})\mathrm{g}(\theta_0) \\ &\quad + (\text{Tab ved fejl af type II})\mathrm{P}(\text{Fejl af type II})\mathrm{g}(\theta_1) \end{aligned}$$

For de tre betragtede tests fås

$$\begin{aligned} \mathrm{r}(g,d_1) &= 6L_1 \cdot 5\% \cdot \frac{1}{3} + L_1 \cdot 15.7\% \cdot \frac{2}{3} = 0.20L_1 \\ \mathrm{r}(g,d_2) &= L_1 \cdot 22.1\% \cdot \frac{2}{3} + 6L_1 \cdot 2.7\% \cdot \frac{1}{3} = 0.20L_1 \\ \mathrm{r}(g,d_3) &= L_1 \cdot 3.9\% \cdot \frac{2}{3} + 6L_1 \cdot 20.8\% \cdot \frac{1}{3} = 0.45L_1, \end{aligned}$$

d.v.s. vi ser igen, at test 3 med det fikserede niveau $\alpha \simeq 5\%$ (præcis $= 3.9\%$) er det klart dårligste hvad angår Bayes-risikoen.

Man kunne naturligvis opstille nogle ganske analoge betragtninger vedrørende minimaxtestet i eksempel 6.5. ♦

I mange testsituationer vil man uden at have en konkret tabsfunktion oftest være i den situation, at man ikke mener, de forskellige fejl, man kan begå, er lige graverende. Hvis man af en eller anden grund (f.eks. et eksisterende standardprogram, et tabelværk med

5%-fraktiler, ens foresattes traditionsbundne holdning etc., etc.) vælger at operere med niveauet $\alpha = 5\%$, er det åbenbart ikke underordnet, hvad man vælger som hypotese, og hvad man vælger som alternativ. Den morale, man kan udlede af ovenstående eksempler er, **at man da bør vælge som sin hypotese den (eller de) parameterværdi(er), som det er "dyrt" fejlagtigt at forkaste**, idet man da kun vil begå denne handling med en sandsynlighed, der højest er lig $\alpha = 5\%$. Hvis man tester den modsatte hypotese, vil man ikke umiddelbart have bånd på sandsynligheden for at begå de "slemme" fejl.

For at eksemplificere ovenstående kommer vi med følgende (ekstreme)

EKSEMPEL 6.8. En laborant på et analyselaboratorium har fået til opgave at undersøge, om et bestemt firma A markedsfører varer, der ikke er i overensstemmelse med varedeklarationen. Efter at have foretaget sine målinger har laboranten følgende muligheder for valg af hypotese og alternativ

1. H_0 : A's varer er "forfalskede"
   H_1 : A's varer er i orden,

eller

2. H_0' : A's varer er i orden
   H_1' : A's varer er "forfalskede".

Laboranten har ingen konkret tabsfunktion, men han har dog på fornemmelsen, at det vil være mere alvorligt at beskylde A for at bedrage, hvis han er hæderlig, end det omvendte, så ifølge ovenstående princip bør han vælge at teste

H_0' : A's varer er i orden

mod

H_1' : A's varer er ikke i orden

♦

Med dette eksempel afslutter vi vor gennemgang af beslutningsteorien og skal henvise læsere, der måtte være interesserede i en uddybning og yderligere eksemplifikation af teorien til de i Bibliografien anførte artikler og lærebøger.

Bibliografi

[1] Artin, E.: *Einführung in die Theorie der Gammafunktion*. Leipzig 1931.

[2] Barnett, Vic: *Comparative Statistical Inference*. John Wiley & Sons, New York & London 1973.

[3] Bar-Shalom, Yaakov: *On the Asymptotic Properties of the Maximum-Likelihood Estimate Obtained from dependent Observations*. J.R. Statist. Soc., B, vol. 33 1971, p. 72.

[4] Bennet, C. A., & Franklin, N. L.: *Statistical Analysis in Chemistry and the Chemical Industri*. John Wiley & Sons, New York 1954.

[5] Bourbaki, N.: *Element de Mathematique. Fonctions d'une variable réelle*. Hermann, Paris 1961.

[6] Bradley, J. V.: *Distribution-free Statistical Tests*. Prentice-Hall, Englewood Cliffs 1968.

[7] Bradley, Ralph A. and Cart, John H.: *The asymptotic properties of ML estimators when sampling from associated populations*. Biometrika, vol. 49 1962, p. 205.

[8] Brøns, H. K.: *Forelæsninger i matematisk statistik og sandsynlighedsregning 9–15, 2. udgave*. Institut for Matematisk Statistik, Københavns Universitet 1968.

[9] Brøns, H. et al.: *Statistik 1*. Institut for matematisk statistik. Københavns Universitet, 1974.

[10] Buch, K.R. m.fl.: *Sandsynlighedsregning i grundrids*. Gyldendal, København 1967.

[11] Chow, V.T.: *The Log-Probability Law and Its Engineering Applications*. Proc. ASCE, vol. 80, No. 536, 1954.

[12] Cox, D.R. and Hinkley, D.V.: *Theoretical Statistics*. Chapman and Hall, London 1974.

[13] David, H.A.: *Order Statistics*, John Wiley & Sons, New York 1970.

[14] Davies, O. L. (Ed.): *Statistical Methods in Research and Production*. Oliver and Boyd, London 1961.

[15] Davies, O. L. (Ed.): *Design and Analysis of Industrial Experiments*. Oliver and Boyd, London 1967.

[16] Dorffman, R.: *The detection of defective members of large populations*. Annals of Math. Stat. 14 (1943), pp. 436-440.

[17] Epstein, B.: *Statistical aspects of fracture problems*. Journal of Applied Physics, vol. 19, (1948), pp. 140-147.

[18] Epstein, B. & Brooks, H.: *The Theory of Extreme values and Its Implications in the Study of the Dielectric Strength of Paper Capacitors*. Journal of Applied Physics 19, (1948), pp. 544-550.

[19] Feller, W.: *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, New York 1965.

[20] Ferguson, T. S.: *Mathematical Statistics. A Decision Theoretic Approach*. Academic Press, New York 1967.

[21] Fisz, M.: *Probability Theory and Mathematical Statistics*. John Wiley & Sons, New York 1963.

[22] Gnedenko, B.: *Sur la distribution limits du terme maxime d'une série aléatoire*. Ann. of Math. 44 (1943), pp. 423-453.

[23] Gumbel, E.J.: *Statistics of Extremes*. Columbia University Press, New York 1958.

[24] Hájek, S. & Šidák, Z.: *Theory of Rank Tests*. Academia, Prague 1967.

[25] Hald, A.: *Statistical Theory with engineering Applications*. John Wiley & Sons, New York & London 1952.

[26] Hansen, N.H.: *Stokastiske modeller*. IMSOR, Lyngby 1975

[27] Jensen, Arne: 1948.

[28] Jensen, Arne: *A distribution model*. Munksgaard, København 1954.

[29] Jensen, Arne: *An Elucidation of Erlangs Statistical Works through the Theory of Stochastic Processes*. Acta Polytechnica Scandinavica, Danish Contribution No. 12, København 1960.

[30] Jensen, Arne: *Application of the Decision Theory to Law and Administration*. Bull. Inst. Int. de Stat., Paris 1961.

[31] Jensen, Arne: *2. rapport til Thyborønudvalget af 1957 vedr. stormflod og digebygning*, 1963.

[32] Jørsboe, Ole Groth: *Noter til sandsynlighedsregning*. DTH, 1972.

[33] Kendall, M. G.: *Rank Correlation Methods (Third Ed.)*. Charles Griffin & Co., Ltd., London 1962.

[34] Kendall, M. G., & Stuart, A.: *The advanced Theory of Statistics*, vol. 1, 2, 3, Charles Griffin & Co., London 1963, 1967, 1966.

[35] Kendall, M.G. and Stuart, A.: *The Advanced Theory of Statistics*, vol. 2. Charles Griffin and Co., London 1967.

[36] Khichin, A. I.: *Mathematical Foundations of Information Theory*. Dover, New York 1957.

[37] Lehmann, E. L.: *Testing Statistical Hypotheses*. John Wiley & Sons, New York 1959.

[38] Lehmann, E. L.: *Nonparametrics: Statistical Methods Based on ranks*. Holden Day, Inc., San Francisco 1975.

[39] Luce, R. D., & Raiffa, H.: *Games and Decisions*. John Wiley & Sons, New York 1957.

[40] Noether, G. E.: *Elements of Nonparametric Statistics*. John Wiley & Sons, New York 1967.

[41] Owen, D. B.: *Handbook of Statistical Tables*. Addison-Wesley, Reading Mass. 1962.

[42] Patil, G.P.: *On the evolution of the negative binomial distribution with examples*. Technometrics, vol. 2, No. 4 (1960), pp. 501-505.

[43] Plackett, R. L.: *Principles of Regression Analysis*. Clarendon Press, Oxford 1960.

[44] Puri, M. L. & Sen, P. K.: *Nonparametric Methods in Multivariate Analysis*. John Wiley & Sons, Inc., New York 1971.

[45] Raiffa, H., & Schlaifer, R.: *Applied Statistical Decision Theory*. Graduate School of Business Administration, Harvard 1961.

[46] Rao, C. Radhakrishna: *Linear Statistical Inference and Its Applications*. John Wiley & Sons, New York 1965.

[47] Renyi, A.: *Wahrscheinlichkeitsrechnung*. VEB Deutscher Verlag der Wissenschaften, Berlin 1966.

[48] Scheffé, H.: *The Analysis of variance*. John Wiley & Sons, New York 1959.

[49] Simonsen, K.Aa.: *Om Portland-Cementens Konstitution*. Proc. 11 fra Nordiska Kemistmötet, 20.- 25. august 1962 i Åbo, pp. 193-214.

[50] Sobel, M. & Groll, P.A.: *Group testing to eliminate efficiently all defectives in a binomial sample*. Bell Syst. Tech. Journal 38 (1959), pp. 1179-1252.

[51] Sobel, M.: *Group testing to classify all all defectives in a binomial sample*. A Contribution in Information and Decision Prosesses. (Ed. R.E. Machol). McGraw Hill, New York 1960, pp. 127-161.

[52] Spearman, C.: *The Proof and Measurement of Association Between Two Things*. Am j. Psych., vol. 15, 1904.

[53] Walsh, J. E.: *Handbook of Nonparametric Statistics*, Vol. i–iii. D. van Nostrand Company, Inc., Princeton 1962, 1965, 1968.

[54] Weibull, W.: *A Statistical distribution function of wide applicability*. J. Appl. Mech. 18 (1952), p. 293.

[55] Wilks, S.S.: *Mathematical Statistics*. John Wiley & Sons, New York 1962.

Stikordsregister

S

T

Det græske alfabet

	Bogstavnavn	*Udtale*	*Gengivelse*
A α	alfa	[a][a:]	a
B β	b$\bar{e}$ta	[b]	b
Γ γ	gamma	[g]	g
Δ δ	delta	[d]	d
E ϵ	epsilon	[e]	e
Z ζ	z$\bar{e}$ta	[ts,s]	z
H η	$\bar{e}$ta	[æ:]	$\bar{e}$
Θ ϑ θ	th$\bar{e}$ta	[θ,th,t]	th,t
I ι	i$\bar{o}$ta	[i][i:]	i
K κ	kappa	[k]	k
Λ λ	lambda	[l]	l
M μ	my	[m]	m
N ν	ny	[n]	n
Ξ ξ	ksi	[ks]	ks(x)
O o	omikron	[o]	o
Π π	pi	[p]	p
P ρ	ro	[r]	r
Σ σ ς	sigma	[s]	s
T τ	tau	[t]	t
Y υ	ypsilon	[y][y:]	y
Φ ϕ	fi	[f]	f(ph)
X χ	khi	[x,ç,kh,k]	ch(kh)
Ψ ψ	psi	[ps]	ps
Ω ω	$\bar{o}$mega	[å:]	$\bar{o}$