

Kursus 02403

Introduktion til Statistik

Forelæsning 8: Inferens for varianser (kap 9)

Jan Kloppenborg Møller

DTU Informatik

Bygning 305 - rum 219

Danmarks Tekniske Universitet

2800 Lyngby – Danmark

e-mail: jkm@imm.dtu.dk

Oversigt

- 1 Motiverende eksempel
- 2 Estimator for varianser
 - χ^2 -fordelingen
- 3 Konfidensinterval for én varians
 - Eksempel 1
- 4 Hypotesetest af varianser
 - Test af én varians
 - Eksempel 1
 - Test af to varianser
 - F -fordelingen
 - Hypotesetestet - konkret
 - Eksempel 1 - fortsat
- 5 R (R note 6)

Oversigt

- 1 Motiverende eksempel
- 2 Estimator for varianser
 - χ^2 -fordelingen
- 3 Konfidensinterval for én varians
 - Eksempel 1
- 4 Hypotesetest af varianser
 - Test af én varians
 - Eksempel 1
 - Test af to varianser
 - F -fordelingen
 - Hypotesetestet - konkret
 - Eksempel 1 - fortsat
- 5 R (R note 6)

Motiverende eksempel

Ved produktion af tabletter blandes aktivt stof og pulver i en beholder hvorefter pulverblandingen trykkes til tabletter. Det er vigtigt, at blandingen er homogen, således at hver tablet får den samme styrke.

Vi betragter nu en blanding (dvs. færdigblandet aktivt stof og pulver) hvorfra der skal produceres et stort antal tabletter.

Man søger at producere blandingen (og dermed de færdige tabletter) så de har middelinhold 1 mg/g aktivt stof og så lille varians som muligt. Som stikprøvekontrol udtages prøver, og aktivt stof måles i prøverne. Det antages, at alle målinger følger en normalfordeling med enheden mg/g .

Oversigt

- 1 Motiverende eksempel
- 2 Estimator for varianser
 - χ^2 -fordelingen
- 3 Konfidensinterval for én varians
 - Eksempel 1
- 4 Hypotesetest af varianser
 - Test af én varians
 - Eksempel 1
 - Test af to varianser
 - F -fordelingen
 - Hypotesetestet - konkret
 - Eksempel 1 - fortsat
- 5 R (R note 6)

Estimator for varianser

Variansen for en stokastisk variabel X estimeres ved

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

hvor n er antal observationer

X_i er observation nr. i , og $i = 1..n$

$\bar{X}$ er estimat af middelværdien for X

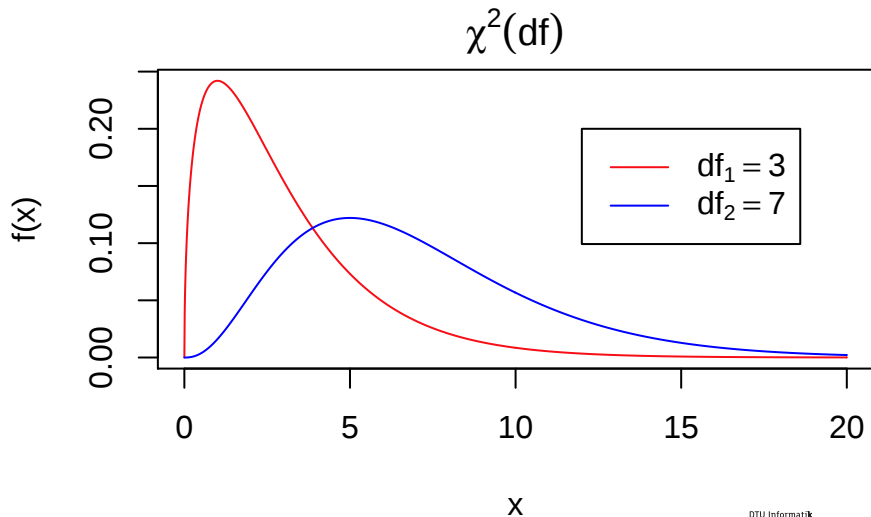
Stikprøvefordeling for variansen (Kap 6.4)

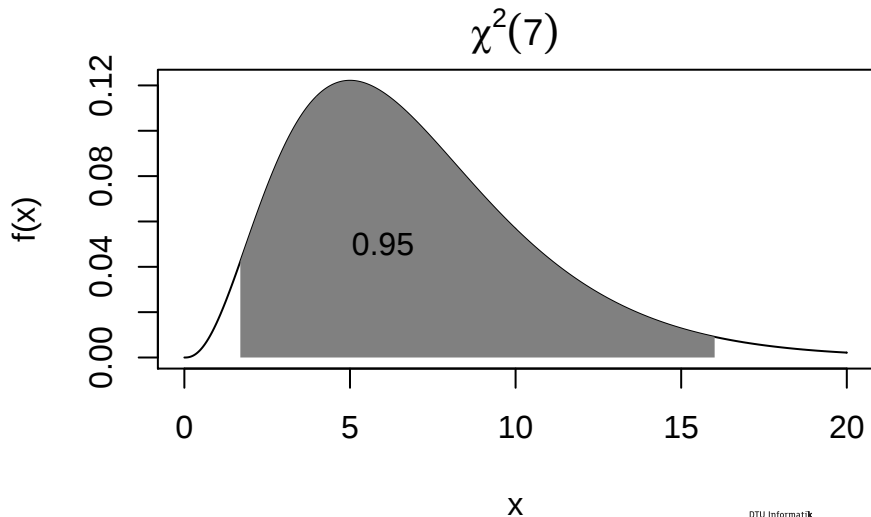
Lad S^2 være variansen af en stikprøve af størrelse n fra en normalfordeling med varians σ^2

Da er

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

en stokastisk variabel og følger en χ^2 -fordeling med parameter $v = n - 1$

χ^2 -fordelingen

χ^2 -fordelingen

Tabelopslag i χ^2 -fordelingen

- Tabelopslag i χ^2 -fordelingen gøres vha tabel 5
- Ved $\chi^2_{\alpha}(n-1)$ forstås den værdi, således at

$$P(\chi^2 \geq \chi^2_{\alpha}) = \alpha$$

Oversigt

- 1 Motiverende eksempel
- 2 Estimator for varianser
 - χ^2 -fordelingen
- 3 Konfidensinterval for én varians
 - Eksempel 1
- 4 Hypotesetest af varianser
 - Test af én varians
 - Eksempel 1
 - Test af to varianser
 - F -fordelingen
 - Hypotesetestet - konkret
 - Eksempel 1 - fortsat
- 5 R (R note 6)

Konfidensinterval for én varians (Kap 8.1)

Et $(1 - \alpha)\%$ konfidensinterval for en varians σ^2 fås ved

$$\frac{(n-1)S^2}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}$$

Fraktilerne for χ^2 har $v = n - 1$ frihedsgrader

Bemærk: Konfidensintervallet for én varians er ikke nødvendigvis symmetrisk

Eksempel 1

Der udtages en tilfældig stikprøve fra beholderen (prøverne tages fra forskellige lag i beholderen) på $n = 20$ tabletter og herfra estimeres middelværdi $\bar{X} = 1.01$ og varians $s^2 = 0.07^2$.

Opstil et 95% konfidensinterval for variansen σ^2

Oversigt

- 1 Motiverende eksempel
- 2 Estimator for varianser
 - χ^2 -fordelingen
- 3 Konfidensinterval for én varians
 - Eksempel 1
- 4 Hypotesetest af varianser
 - Test af én varians
 - Eksempel 1
 - Test af to varianser
 - F -fordelingen
 - Hypotesetestet - konkret
 - Eksempel 1 - fortsat
- 5 R (R note 6)

Hypotesetest af varianser

Hypotesetest af varianser foregår stort set som ved hypotesetest af middelværdier, blot ændres teststørrelse og stikprøvefordeling (fordeling for kritisk værdi under H_0).

1. Opstil hypoteser og vælg signifikansniveau α
2. Beregn teststørrelse
3. Beregn kritisk værdi (eller p-værdi)
4. Sammenlign teststørrelse og kritisk værdi og drag en konklusion

Test af én varians

nul hypotese testes mod en *alternativ hypotese* (her vist for et to-sidet alternativ)

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

Man vælger enten at acceptere H_0 eller at forkaste H_0 (hvor σ_0^2 er værdien der testes)

Test af én varians

Teststørrelse bliver

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Den kritiske værdi findes ved opslag i χ^2 fordelingen (tab. 5) med $v = n - 1$ frihedsgrader

Beregning af kritisk værdi

Ved hypoteseprøvning af én varians for data der antages normalfordelt fås

Alternativ hypotese	Afvis nul-hypotese hvis
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi_{1-\alpha}^2$
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi_{\alpha}^2$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi_{1-\alpha/2}^2$ eller $\chi^2 > \chi_{\alpha/2}^2$

Eksempel 1

Ved produktion af tabletter blandes aktivt stof og pulver i en beholder hvorefter pulverblandingen trykkes til tabletter. Det er vigtigt, at blandingen er homogen, således at hver tablet får den samme styrke.

Vi betragter nu en blanding (dvs. færdigblandet aktivt stof og pulver) hvorfra der skal produceres et stort antal tabletter.

Man søger at producere blandingen (og dermed de færdige tabletter) så de har middelinhold 1 mg/g aktivt stof og så lille varians som muligt. Som stikprøvekontrol udtages prøver, og aktivt stof måles i prøverne. Det antages, at alle målinger følger en normalfordeling med enheden mg/g .

Eksempel 1 - fortsat

Man er indledningsvis interesseret i at teste om variansen af indholdet af aktivt stof i en tablet kan antages at være $\sigma^2 = 0.05^2$.

Formulér nul- og alternativ hypotese og specificer testets signifikansniveau

Eksempel 1 - fortsat

Der udtages en tilfældig stikprøve fra beholderen (prøverne tages fra forskellige lag i beholderen) på $n = 20$ tabletter og herfra estimeres middelværdi $\bar{X} = 1.01$ og varians $s^2 = 0.07^2$.

Udfør hypotestestet. Kan det påvises, at $\sigma^2 > 0.05^2$?

Eksempel 1 - fortsat

Opstil et 95% konfidensinterval for variansen σ^2

Udfør et hypotestestet for middelinholdet. Kan det påvises, at $\mu \neq 1$ mg/g?

Vurder ud fra de foregående analyser, hvad der er problemet i produktionen?

Sammenligning af 2 varianser

- Vi sammenligner varianser af 2 stikprøver
- Stikprøve 1: n_1 , $\bar{X}_1$ og s_1^2
- Stikprøve 2: n_2 , $\bar{X}_2$ og s_2^2

Det antages, at data for begge stikprøver er normalfordelt.

Stikprøvefordeling for sammenligning af varianser (Kap 6.4)

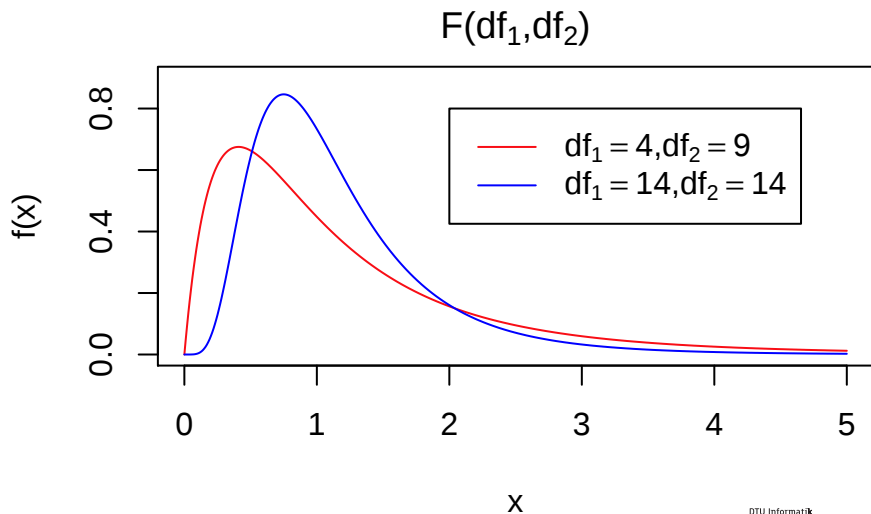
Lad S_1^2 og S_2^2 være varianser af stikprøver af størrelse henholdsvis n_1 og n_2 fra to normalfordelinger med samme varians

Da er

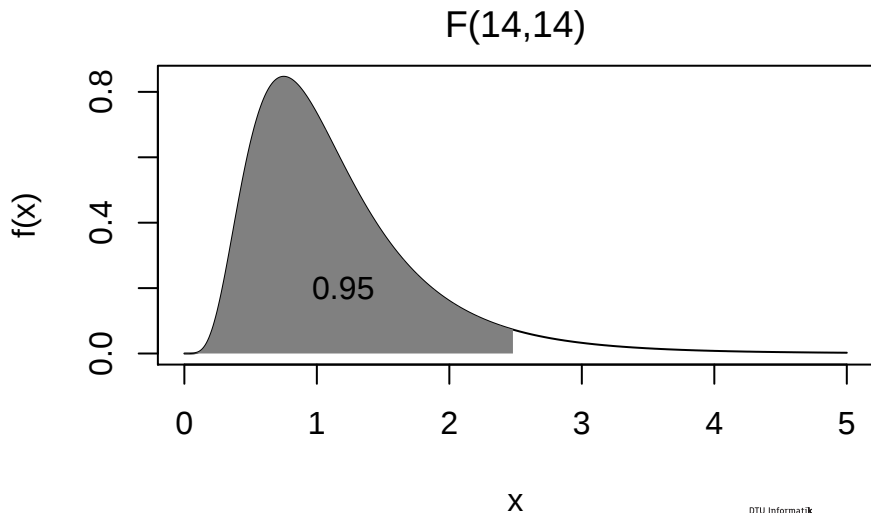
$$F = \frac{S_1^2}{S_2^2}$$

en stokastisk variabel og følger en F -fordeling med parameter $v_1 = n_1 - 1$ og $v_2 = n_2 - 1$

F -fordelingen



F -fordelingen



Tabelopslag i F -fordelingen

- Tabelopslag i F -fordelingen gøres vha tabel 6
- Ved $F_\alpha(n_1 - 1, n_2 - 1)$ forstås den værdi, således at

$$P(F \geq F_\alpha) = \alpha$$

1. Opstil Hypoteser (Kap 8)

nul hypotese testes mod en *alternativ hypotese* (her vist for et to-sidet alternativ)

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Man vælger enten at acceptere H_0 eller at forkaste H_0

Beregning af kritisk værdi

Ved hypoteseprøvning af to varianser for data der antages normalfordelt fås

Alternativ hypotese	Test størrelse	Afvis nul-hypotese hvis
$\sigma_1^2 < \sigma_2^2$	$F = \frac{S_2^2}{S_1^2}$	$F > F_\alpha(n_2 - 1, n_1 - 1)$
$\sigma_1^2 > \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$	$F > F_\alpha(n_1 - 1, n_2 - 1)$
$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_M^2}{S_m^2}$	$F > F_{\alpha/2}(n_M - 1, n_m - 1)$

(i sidste tilfælde gælder $S_M^2 > S_m^2$)

Eksempel 1 - fortsat

Det viser sig, at et problem i produktion af tabletter er, at der er tendens til lagdeling i blandingen, f.eks. når aktivt stof samler sig i bunden af beholderen. Der udtages nu en prøve fra hhv øvre og nedre lag i beholderen.

$$n_1 = 10, \bar{X}_1 = 0.98, \text{ og } s_1^2 = 0.03^2$$

$$n_2 = 12, \bar{X}_2 = 1.04, \text{ og } s_2^2 = 0.08^2$$

Vi er nu interesseret i at undersøge, om der er forskel i varianserne for de to stikprøver. *Formulér nul- og alternativ hypotese, og udfør testet med signifikansniveau $\alpha = 5\%$*

Eksempel 1 - fortsat

Vi er nu interesseret i at undersøge, om der er forskel i middelværdierne for de to stikprøver. *Formulér nul- og alternativ hypotese, og udfør testet med signifikansniveau $\alpha = 5\%$*

Vurder ud fra de foregående analyser, hvad der er problemet i produktionen?

Oversigt

- 1 Motiverende eksempel
- 2 Estimator for varianser
 - χ^2 -fordelingen
- 3 Konfidensinterval for én varians
 - Eksempel 1
- 4 Hypotesetest af varianser
 - Test af én varians
 - Eksempel 1
 - Test af to varianser
 - F -fordelingen
 - Hypotesetestet - konkret
 - Eksempel 1 - fortsat
- 5 R (R note 6)

R (R note 6)

R	Betegnelse
t	t-fordelingen
chisq	χ^2 -fordelingen
f	F-fordelingen

- d Tæthedsfunktion $f(x)$ (probability distribution).
- p Fordelingsfunktion $F(x)$ (cumulative distribution function).
- q Fraktil (quantile) i fordeling.

Eksempel: $P(F \leq 2), F \sim F(3, 12)$

```
> pf(2, 3, 12)
```

Oversigt

- 1 Motiverende eksempel
- 2 Estimator for varianser
 - χ^2 -fordelingen
- 3 Konfidensinterval for én varians
 - Eksempel 1
- 4 Hypotesetest af varianser
 - Test af én varians
 - Eksempel 1
 - Test af to varianser
 - F -fordelingen
 - Hypotesetestet - konkret
 - Eksempel 1 - fortsat
- 5 R (R note 6)