

Skriftlig prøve, den: 16. december 2002

Kursus nr : 02401

Kursus navn: Dataanalyse og Indledende Statistik

Tilladte hjælpemidler: Alle sædvanlige

Dette sæt er besvaret af:

(navn)_____
(underskrift)_____
(bord nr)

Der er i alt 18 spørgsmål fordelt på 18 opgaver, benævnt opgave 1,2,..., 18 i teksten. De enkelte spørgsmål er nummereret tilsvarende, og angivet som spørgsmål 1,2,...,18 i teksten. Bevarelserne af de 18 spørgsmål føres ind i nedenstående skema.

Opgave	1	2	3	4	5	6	7	8	9	10
Spørgsmål	1	2	3	4	5	6	7	8	9	10
Svar										

Opgave	11	12	13	14	15	16	17	18	
Spørgsmål	11	12	13	14	15	16	17	18	
Svar									

Svarmulighederne for hvert spørgsmål er nummereret fra 1 til 6. Indføres et forkert nummer i skemaet, kan dette rettes ved at "sværte" det forkerte nummer over og anføre det rigtige nedenunder. Er der tvivl om meningen med en rettelse, betragtes spørgsmålet som ubesvaret.

Kun forsiden skal afleveres. Afleveres blankt eller forlades eksamen i utide, skal forsiden alligevel afleveres. Kladder, mellemregninger og bemærkninger tillægges **ingen** betydning, kun tallene indført ovenfor registreres.

Der gives 5 point for et korrekt svar og -1 for et ukorrekt svar. Ubesvarede spørgsmål eller et 6-tal (svarende til "ved ikke") giver 0 point. Det antal point, der kræves for, at et sæt anses for tilfredstillende besvaret, afgøres endeligt ved censureringen af sættene.

Husk at forsyne opgaveteksten med navn, underskrift og bordnummer.

Der gøres opmærksom på at ideen med opgaverne er, at der er ét og kun ét rigtigt svar på de enkelte spørgsmål. Endvidere er det ikke givet, at alle de anførte alternative svarmuligheder er meningsfulde. Sættets sidste side er nr 25; blad lige om og se, at den er der.

Opgave 1

Som led i rutinekontrollen i produktionen af en plasticfilm udtog man 14 prøver tilfældigt fra en produktion og for hver af prøverne bestemte man trækstyrken (Y) i PSI.

Man fandt gennemsnittet, $\bar{y} = \sum_i y_i / 14 = 33.712$ PSI,
og standardafvigelsen $s = \sqrt{\sum (y_i - \bar{y})^2 / 13} = 0.798$ PSI.

Spørgsmål 1

Under antagelse af at fordelingen af trækstyrken kan beskrives ved en normalfordeling, er 99 % konfidensintervallet for middelværdien af trækstyrken i prøver fra produktionen:

- 1 $33.712 \pm 2.2816 \times 0.798 / \sqrt{14}$
- 2 $33.712 \pm 2.3264 \times 0.798 / \sqrt{14}$
- 3 $33.712 \pm 2.6810 \times 0.798 / \sqrt{14}$
- 4 $33.712 \pm 3.0123 \times 0.798 / \sqrt{14}$
- 5 $33.712 \pm 2.6810 \times 0.798 \times \sqrt{1 + \frac{1}{14}}$
- 6 Ved ikke

Ved besvarelsen kan nedenstående tabeluddrag eventuelt benyttes:

Friheds- grader	Kritiske værdier for t-fordelingen		
	$t_{k,0.98}$	$t_{k,0.99}$	$t_{k,0.995}$
12	2.3027	2.6810	3.0545
13	2.2816	2.6503	3.0123
14	2.2638	2.6245	2.9768

Fraktiler, z_P , i standardiseret normalfordeling			
P	0.98	0.99	0.995
z_P	2.0538	2.3264	2.5785

Opgave 2

En virksomhed, der fremstiller penicillin, ønsker at sammenligne 4 produktionsmetoder, A, B, C og D.

Ved produktionen benyttes et majsderivat som råvare. Det vides, at denne råvare udviser en vis variation fra den ene råvareblanding til den anden.

Man tilrettelagde derfor et forsøg, hvor der benyttedes 5 forskellige råvareblandinger. For hver blanding blev der da foretaget en penicillinproduktion ved hver af de fire metoder, hvorefter man registrerede ydelsen ved denne produktion. Forsøgsresultaterne (ydelsen udtrykt som det relative indhold af penicillin i procent) er angivet i nedenstående tabel:

Råvare- blanding	Metode			
	A	B	C	D
1	89	88	97	94
2	84	77	92	79
3	81	87	87	85
4	87	92	89	84
5	79	81	80	88

Spørgsmål 2

Sammenligningen af ydelsen ved de fire metoder foretages bedst ved:

- 1 Regressionsanalyse
- 2 Ensided variansanalyse uden blokning (one-way model)
- 3 Ensided variansanalyse med blokning (RCBD)
- 4 χ^2 -test for uafhængighed i en tovejstabel for kategoriserede data
- 5 Ingen af ovenstående
- 6 Ved ikke

Opgave 3

En virksomhed, der producerer en bestemt type lejer, var interesseret i at undersøge betydningen af olieviskositet og belastning for levetiden af et leje.

Man foretog en række eksperimenter med forskellige kombinationer af viskositet og belastning, og for hver kombination registrerede man sliddet på lejet.

Resultaterne er vist i nedenstående tabel

Slid	Viskositet	Belastning
198	11.6	851
205	15.5	816
160	22.0	1058
98	43.0	1201
113	33.0	1357
125	40.0	1115

Der blev udført en lineær regressionsanalyse med henblik på at beskrive sliddet (`slid`) som funktion af viskositeten (`viskos`) af lejeolien og af belastningen (`belast`) af lejet. Udvalgte dele af output er vist nedenfor.

```
slid = viskos belast
Response Distribution: Normal
Link Function: Identity
```

```
Model Equation
slid = 312.1673 - 2.0157 viskos - 0.1002 belast
```

```
Summary of Fit
Mean of Response    149.8333    R-Square    0.9845
Root MSE            7.2306    Adj R-Sq    0.9742
```

```
Analysis of Variance
Source    DF    Sum of Squares    Mean Square    F Stat    Pr > F
Model     2     9969.9884         4984.9942      95.35     0.0019
Error     3     156.8449          52.2816
C Total   5     10126.8333
```

Spørgsmål 3

R-Square (R^2) værdien på 0.98 kan tages som udtryk for:

- 1 Variationen i lejeslid kan i udstrakt grad forklares af variationerne i viskositet og i belastning
- 2 Der er ingen multikollinearitet i data
- 3 Der er en høj grad af multikollinearitet i data
- 4 Fordelingen af lejeslid kan ikke beskrives ved en normalfordeling
- 5 Ingen af ovenstående
- 6 Ved ikke

Opgave 4

Som led i et biologisk studium med henblik på at vurdere relationen mellem kropsvægt og hjertevægt hos forskellige pattedyr indfangede man 19 murmeldyr og bestemte deres kropsvægt og hjertevægt i gram.

Nedenstående viser udskriften fra menuen Distribution for de variable kropsvægt body og hjertevægt heart

```
body
      Moments
N      19.0000  Sum Wgts      19.0000
Mean   3103.6842  Sum          58970.0000
Std Dev 1030.7196  Variance     1062382.89
Skewness  1.2068  Kurtosis      0.9321
USS      202147150  CSS          19122892.1
CV       33.2096  Std Mean     236.4633
```

```
heart
      Moments
N      19.0000  Sum Wgts      19.0000
Mean   11.9105  Sum          226.3000
Std Dev  1.5776  Variance      2.4888
Skewness  0.9943  Kurtosis      1.0560
USS      2740.1500  CSS          44.7979
CV       13.2453  Std Mean      0.3619
```

Endvidere bestemtes størrelsen

$$\sum_{i=1}^{19} (x_i - \bar{x})(y_i - \bar{y}) = -2120.2368$$

hvor x_i og \bar{x} angiver hhv. kropsvægten for det i 'te murmeldyr og den gennemsnitlige kropsvægt for de 19 dyr, og tilsvarende y_i og \bar{y} angiver hjertevægt og gennemsnitlig hjertevægt.

Spørgsmål 4

Pearson-korrelationen mellem kropsvægt og hjertevægt for disse data beregnes som:

1 $\frac{3103.6842 \cdot 11.9105}{18 \cdot 1030.7196 \cdot 1.5776}$

2 $\frac{3103.6842 \cdot 11.9105}{18 \cdot 236.4633 \cdot 0.3619}$

3 $\frac{-2120.2368}{18 \cdot 1030.7196 \cdot 1.5776}$

4 $\frac{-2120.2368}{1030.7196 \cdot 1.5776}$

5 $\frac{-2120.2368}{18 \cdot 3103.6842 \cdot 11.9105}$

6 Ved ikke

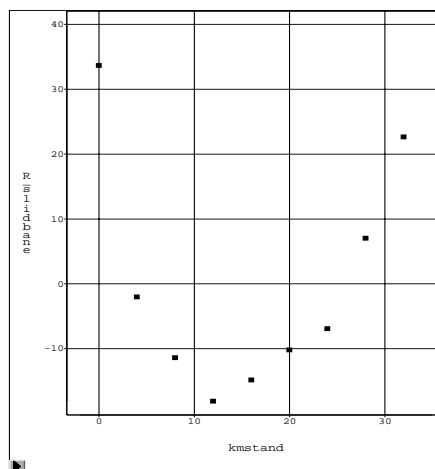
Opgave 5

I et studium til vurdering af dæksliddet for et bestemt dækfabrikat monterede man dæk af dette fabrikat på en bil. For hver 4000 miles målte man dybden af mønsteret i slidbanen (i tusindedel tommer) på seks udvalgte steder på hvert dæk og registrerede gennemsnittet af disse 24 målinger (slidbanedybde, y).

Der blev udført en lineær regressionsanalyse af slidbanedybde, y , mod kilometerstand, x . Nedenstående tabel viser de oprindelige målinger, de fittede værdier, \hat{y} og residualerne $e = y - \hat{y}$.

Obs nr i	Kilometerstand x_i (miles)	Slidbanedybde y_i	fittet værdi \hat{y}_i	residual e_i
1	0	394.33	360.64	33.69
2	4000	329.50	331.51	-2.01
3	8000	291.00	302.39	-11.39
4	12000	255.17	273.27	-18.10
5	16000	229.33	244.15	-14.82
6	20000	204.83	215.02	-10.19
7	24000	179.00	185.90	-6.90
8	28000	163.83	156.78	7.05
9	32000	150.33	127.66	22.67

Residualerne er afbildet mod kilometerstanden (i 1000 miles) i nedenstående figur:



Spørgsmål 5

Forløbet af residualerne er tegn på, at:

- 1 Variansen afhænger af kilometerstanden
- 2 Målingerne ikke er uafhængige
- 3 Sammenhængen mellem slid og kilometerstand er ikke lineær
- 4 Der er for få data til at kunne påvise en sammenhæng
- 5 Der er en lille forklaringsgrad, (R^2), i analysen
- 6 Ved ikke

Opgave 6

Som led i en undersøgelse af befolkningens rejsevaner udvalgte en tilfældig stikprøve af personer, der benyttede egen bil til transport mellem bolig og arbejdssted.

De udvalgte personer blev adspurgt om deres rejselængde (3 svarkategorier) og om størrelsen af deres bil (4 svarkategorier). Antallet af svar for hver af de 12 kombinationer af rejselængde og bilstørrelse er angivet i nedenstående tabel:

Bilens størrelse	Rejselængde		
	0-10 km	10-20 km	≥ 20 km
Mini	6	27	19
Compact	8	36	17
Standard	21	45	33
Caravan	14	18	6

Spørgsmål 6

Disse data analyseres bedst ved:

- 1 Regressionsanalyse
- 2 Ensided variansanalyse uden blokning (one-way model)
- 3 Ensided variansanalyse med blokning (RCBD)
- 4 χ^2 -test for uafhængighed i en tovejstabel for kategoriserede data
- 5 Ingen af ovenstående
- 6 Ved ikke

Opgave 7

En virksomhed producerer en bestemt type komponenter med en tilstræbt diameter på 1.50 cm. Virksomheden har erfaring for at målingerne af diametrene af de producerede komponenter kan beskrives ved en normalfordeling med middelværdi 1.50 cm og standardafvigelse 0.2 cm.

Man ønsker nu at angive et interval, $1.50 \pm \Delta$, således at dette interval netop omfatter 95% af målingerne.

Spørgsmål 7

Værdien af Δ , der opfylder dette krav, bestemmes som:

- 1 1.64/0.2
- 2 $0.2 \cdot 1.64$
- 3 1.96/0.2
- 4 $0.2 \cdot 1.96$
- 5 $0.1 \cdot 1.96$
- 6 Ved ikke

Opgave 8

Ved måling, fx af kemiske koncentrationer, X , vil man ofte se, at fordelingen af data er unimodal og skæv til højre, sådan, at logaritmen til målingerne, $Y = \ln(X)$, kan beskrives ved en normalfordeling.

Spørgsmål 8

Såfremt dette er tilfældet, gælder for fordelingen af X :

- 1 median \leq middelværdi \leq mode
- 2 middelværdi \leq median \leq mode
- 3 mode \leq median \leq middelværdi
- 4 mode \leq middelværdi \leq median
- 5 middelværdi \leq mode \leq median
- 6 Ved ikke

Opgave 9

I en betonvirksomhed ønsker man at vurdere, hvorledes fugtoptagelsen i den færdigstøbte beton afhænger af betonens sammensætning.

Man undersøgte 5 forskellige recepter for fremstillingen (betonsammensætninger). For hver af disse blev udstøbt 6 betonprøver, som blev udsat for fugt i 48 timer, hvorefter man målte optagelsen af fugt i hver af de 30 prøver.

Eventuelle forskelle på fugtoptagelsen for de 5 recepter tænkes undersøgt ved brug af en ensidet variansanalysemodel uden blokning.

Spørgsmål 9

Modellens antagelse om normalfordeling kan for eksempel kontrolleres ved at tegne et samlet Q-Q plot for:

- 1 De 30 værdier af fugtoptagelsen
- 2 De 30 værdier af modellens fittede værdier af fugtoptagelsen
- 3 De 30 værdier af de standardiserede residualer fra den fittede model for fugtoptagelsen
- 4 De 30 værdier af logaritmen til fugtoptagelsen
- 5 Normalfordelingsantagelsen kan ikke kontrolleres ved Q-Q plot
- 6 Ved ikke

Opgave 10

Opgaven vedrører et forsøg, der havde til formål at bestemme styrken af en svejsning som funktion af elektrodeafstanden og svejsetiden. Ved forsøget benyttede man 3 indstillinger af elektrodeafstanden og 5 forskellige svejsetider. Der blev udført to uafhængige svejsninger for hver af de 15 kombinationer af elektrodeafstand og svejsetid.

Analyse af data i overensstemmelse med den generelle effekt model gav nedenstående udskrift:

```
styrke = afstand tid afstand*tid
Response Distribution: Normal
Link Function: Identity
```

Nominal Variable Information

Level	afstand	tid
1	1	1
2	2	2
3	3	3
4		4
5		5

Summary of Fit

Mean of Response	15.6000	R-Square	0.8849
Root MSE	3.3066	Adj R-Sq	0.7775

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
Model	14	1261.2000	90.0857	8.24	0.0001
Error	15	164.0000	10.9333	.	.
C Total	29	1425.2000	.	.	.

Source	DF	Sum of Squares	Mean Square	F Stat	Pr > F
afstand	2	278.6000	139.3000	12.74	0.0006
tid	4	385.5333	96.3833	8.82	0.0007
afstand*tid	8	597.0667	74.6333	6.83	0.0008

Spørgsmål 10

Man kan derfor konkludere:

- 1 Der er ikke samme varians af styrken ved de 15 kombinationer af elektrodeafstand og svejsetid
- 2 Der er en klar indikation af, at forskellen mellem styrken ved forskellige elektrodeafstande afhænger af svejsetiden
- 3 Der er ingen indikation af, at forskellen mellem styrken ved forskellige elektrodeafstande afhænger af svejsetiden
- 4 Man bør foretage en logaritmetransformation af de målte styrker
- 5 Man bør foretage en kvadratrodsttransformation af de målte styrker
- 6 Ved ikke

Opgave 11

I en undersøgelse af luftforureningen i et udviklingsland udtog man prøver af luften på fire forskellige tidspunkter i perioden 1975-1976, og på fem forskellige lokaliteter i landet. For hver prøve bestemte man partikelindholdet i mg/m^2 .

Resultatet er angivet i nedenstående skema.

Tidspunkt	Lokalitet					Gennemsnit
	1	2	3	4	5	
Okt 75	76	67	81	56	51	66.20
Jan 76	82	69	96	59	70	75.20
Maj 76	68	59	67	54	42	58.00
Sep 76	63	56	64	58	37	55.60
Gennemsnit	72.25	62.75	77.00	56.75	50.00	63.75

Spørgsmål 11

Under antagelse af en additiv effektmodel (dvs en model uden vekselvirkning) finder man den fittede værdi for lokalitet 3 i maj 76 som:

- 1 63.75
- 2 $(58.00 + 77.00)/2$
- 3 $58.00 + 77.00 - 63.75$
- 4 $67 - 63.75 + (58.00 + 77.00)/2$
- 5 $67 - 63.75 + \sqrt{58.00 \cdot 77.00}$
- 6 Ved ikke

Opgave 12

En virksomhed, der fremstiller farsbrød i en industriel ovn, var interesseret i at vurdere varmemefordelingen i ovnen.

Man valgte at vurdere varmemefordelingen ved at måle dryptabet fra farsbrød på 8 udvalgte positioner i ovnen. Der blev tilberedt 3 partier fars i ovnen, hvert parti bestående af 8 brød. Brødene fra hvert parti blev placeret tilfældigt i de 8 positioner.

Variation	SS
Position	40.396
Farsparti	1762.067
Error	9.290
Total	65.945

Spørgsmål 12

Den relevante F-teststørrelse for et test af hypotesen om, at der ikke er forskel på dryptabet i de forskellige positioner er:

- 1 $(1762.067/2)/(9.290/14)$
- 2 $(1762.067/7)/(9.290/14)$
- 3 $(40.396/7) / (1762.067/2)$
- 4 $(40.396/7)/(9.290/7)$
- 5 $(40.396/7)/(9.290/14)$
- 6 Ved ikke

Opgave 13

En virksomhed fremstiller plasticdunke ved en støbeprocess med 6 forskellige støbemaskiner.

Man er interesseret i at vurdere, hvorvidt middelvægten for de producerede dunke er den samme for alle 6 maskiner.

Man udtog derfor tilfældigt 8 dunke fra hver af de 6 maskiners produktion og bestemte vægten af hver dunk (ialt 48 dunke).

Spørgsmål 13

Sammenligningen foretages bedst ved:

- 1 Vurdering af konfidensintervaller for de seks individuelle middelvægte
- 2 Test i en ensidet variansanalysemodel uden blokning (one-way model)
- 3 Test i en tosidet variansanalysemodel (two-way model)
- 4 Tukey's test for additivitet
- 5 χ^2 -test for uafhængighed i en tovejstabel for kategoriserede data
- 6 Ved ikke

Opgave 14

En virksomhed, der fremstiller kuglelejer med en nominal diameter på 1 micron, benytter to produktionslinier til fremstillingen.

Man ønsker at undersøge, om der er forskel på de to produktionslinier.

Nedenstående tabel viser diameteren i micron for en tilfældigt udtaget stikprøve på 10 lejer fra hver af disse to produktionslinier.

Nr	Diameter	
	Linie 1	Linie 2
1	1.18	1.72
2	1.42	1.62
3	0.69	1.69
4	0.88	0.79
5	1.62	1.79
6	1.09	0.77
7	1.53	1.44
8	1.02	1.29
9	1.19	1.96
10	1.32	0.99

Spørgsmål 14

En sammenligning mellem middeldiameteren for de to produktionslinier foretages bedst ved:

- 1 Regressionsanalyse
- 2 t-test for uparrede data
- 3 t-test for parrede data
- 4 χ^2 -test for uafhængighed i en tovejstabel for kategoriserede data
- 5 Ingen af ovenstående
- 6 Ved ikke

Opgave 15

I en virksomhed foretages en blanding af to saltholdige væsker ved at tage 1 liter af hver væske og hælde sammen i én dunk (ialt 2 liter).

De oprindelige væsker opbevares i 1 liter dunke, og det vides, at saltkoncentrationen i dunkene med den første væske varierer fra dunk til dunk i overensstemmelse med en normalfordeling med middelværdi 10 mg/ml og standardafvigelse 3 mg/ml, og koncentrationen i dunkene med den anden væske tilsvarende varierer i overensstemmelse med en normalfordeling med middelværdi 15 mg/ml og standardafvigelse 4 mg/ml.

Spørgsmål 15

Da bliver standardafvigelsen i fordelingen for saltkoncentrationen (i mg/ml) i dunkene med blandingen:

- 1 7
- 2 $\sqrt{25/4}$
- 3 $\sqrt{25/2}$
- 4 $\sqrt{25}$
- 5 3.5
- 6 Ved ikke

Opgave 16

På en produktionsvirksomhed, hvor der arbejdes i tre skift (dag, aften og nat), er man interesseret i at vurdere, om der kan påvises forskelle i defektprocenten for varer produceret af de tre skift.

Der blev derfor udtaget en stikprøve af varer produceret af hvert af de tre skift, og antallet af defekte varer i hver stikprøve blev opgjort.

Resultatet er anført i nedenstående tabel:

Klassifikation	Skift		
	Dag	Aften	Nat
Defekte	45	55	70
Ikke-defekte	905	890	870

Man udførte et χ^2 -test i en tovejstabel for kategoriserede data, og fandt følgende værdi af teststørrelsen, $X^2 = 6.29$.

Spørgsmål 16

Teststørrelsen skal sammenlignes med fraktilerne i en χ^2 -fordeling med:

- 1 1 frihedsgrad
- 2 2 frihedsgrader
- 3 3 frihedsgrader
- 4 4 frihedsgrader
- 5 Ingen af ovenstående
- 6 Ved ikke

Opgave 17

Nedenstående tabel viser sandsynlighedsfordelingen for antallet af retransmissioner, der er nødvendige for at transmittere en 1024K datapakke i en satellittransmission.

x	0	1	2	3
$p(x)$	0.40	0.30	0.25	0.05

Spørgsmål 17

Det forventede antal retransmissioner, der er nødvendige (forventningsværdien i fordelingen), bestemmes som:

- 1 $(0 + 1 + 2 + 3)/4$
- 2 $(0 \cdot 40 + 1 \cdot 30 + 2 \cdot 25 + 3 \cdot 5)/4$
- 3 $0 \cdot 40 + 1 \cdot 30 + 2 \cdot 25 + 3 \cdot 5$
- 4 $0 \cdot 0.40 + 1 \cdot 0.30 + 2 \cdot 0.25 + 3 \cdot 0.05$
- 5 $(0^2 \cdot 0.40 + 1^2 \cdot 0.30 + 2^2 \cdot 0.25 + 3^2 \cdot 0.05)/3$
- 6 Ved ikke

Opgave 18

Et multiple-choice opgavesæt består af 15 spørgsmål, hver med fem svarmuligheder.

Antag, at en opgaveløser ved besvarelsen af hvert spørgsmål har sandsynligheden $1/5 = 0.2$ for at svare korrekt, og at besvarelsen af de 15 spørgsmål er indbyrdes uafhængige.

Spørgsmål 18

Sandsynligheden for at denne opgaveløser højst opnår 3 korrekte svar er da:

- 1 $(0.2)^3$
- 2 0.6482
- 3 $0.8358 - 0.6482$
- 4 $1 - 0.8358$
- 5 $15 \cdot (0.2)^3 \exp(-0.2)$
- 6 Ved ikke

Slut på opgavesættet.