



EKSAMENSOPPGAVE

Institutt:	<u>IKBM</u>
Eksamen i:	<u>STAT 100</u> <u>Statistikk</u>
Tidspunkt for eksamen:	<u>10. desember 2015</u> <u>09.00-12.30. 3,5 timer</u>
Kursansvarlig:	<u>Trygve Almøy</u>

Tillatte hjelpemidler: C3. Alle typer kalkulatorer, alle andre hjelpemidler
Oppgaveteksten er på 10 sider.

Oppgave I og II teller 50 % av denne eksamen, og alle 8 delspørsmål teller likt.
Flervalgsspørsmål teller 50 % av denne eksamen.

Merk: Side 10: Skjema fylles ut og arket leveres inn.

Oppgave I

Avl på melkemengde går hovedsakelig gjennom okse. 5 okser (med navn Åge, Hynar, Lars, Bull og Gorm) ble valgt ut for å testes hvor gode de er med hensyn på melkemengde hos døtre. Alle disse oksene har mange døtre, vi plukker tilfeldig ut 8 av disse fra hver okse. For de 40 døtrene har vi observert årlig melkemengde (kilo melk) i ett år. Data ble analysert ved en enveis variansanalysemodell, der en gruppe er å oppfatte som en okse. Data finner du i Tabell 2 og resultat av analysen i Tabell 1.

	Df	Sum Sq	Mean Sq	F value	Pr (>F)
okse	4	1251009	312752	2.806	0.0404
Residuals	35	3901166	111462		

	mean	sd	n
Åge	6689	468	8
Hynar	6282	150	8
Lars	6325	286	8
Bull	6671	436	8
Gorm	6622	204	8

Tabell 1: Resultat fra kjøring i R-commander, oppgave I.

- Skriv ned modellen som blir brukt.
Gi en forklaring på alle størrelser som inngår i modellen.
Hva blir estimatene til alle parametere i modellen?
- Utfør en test på om okse har betydning for melkeproduksjon
Skriv ned nullhypotese og alternativ hypotese.
Trekk en konklusjon på bakgrunn av testen, bruk 5 % signifikansnivå.



Åge og Gorm er såkalt NRF okser (vanlig norsk rase), mens de 3 andre er såkalt Holstein Frieser okser.

- c) Finn en kontrast som definerer forskjellen i melkeproduksjon mellom disse rasene. Kan du med sikkerhet si at døtre av disse to NRF oksene gir mer melk enn døtre av disse tre Holstein Frieser oksene?
- d) Hva blir residualen for datter nr. 1 til Åge?
Hva betyr det at et residual er negativt eventuelt positivt?

Åge	Hynar	Lars	Bull	Gorm
6363	6337	6330	6250	6613
6697	6179	6571	6393	6317
6634	6478	6676	6992	6905
7033	6171	5970	7181	6609
6362	6465	6480	6147	6484
6038	6184	6022	6678	6777
6825	6073	6011	6434	6834
7562	6369	6546	7294	6442

Tabell 2: Data til oppgave I.

Oppgave II

Vi har samlet data om pris, størrelse og alder på leiligheter sentralt i Oslo. Pris er i tusen kroner, størrelse er i m^2 , og alder i år. Alle leiligheter er solgt nylig. Data finner du i Tabell 5.

Lag en modell der pris er respons og størrelse er forklaringsvariabel. Kall denne Modell 1. Resultatene av analysen finner du i Tabell 3.

- a) Skriv ned modellen.
Gi en forklaring på hva alle 3 parametre betyr, vær konkret.
Gi parameterestimaterne.
Gi en forklaring på hva R^2 måler, og finn verdien på denne størrelsen.

Anta at du har en leilighet på $20 m^2$ som du ønsker å selge.

- b) Hvilken pris vil du anslå at du oppnår for denne?
Et 95 % prediksjonsintervall for denne prisen er gitt ved (41; 2357).
Gi en tolkning av dette intervallet.
Kan du gi forskjellige årsaker til at intervallet blir såpass bredt?

I Figur 1 er det et histogram over residualene, og i Figur 2 er residualene plottet mot størrelsen på leilighetene.

- c) Bruk disse figurene til å kommentere antagelsene i modellen du har brukt.

En prøve også modellen $Y_i = \alpha + \beta x_i + \varepsilon_i$ der x_i er alder på leilighet nr. i , og ellers med vanlige antagelser. Se Tabell 4 for resultater.

- d)** Lag et 95 % konfidensintervall for β .
 Forklar hvordan du ved hjelp av dette intervallet kan påvise signifikant sammenheng mellom alder og pris på 5 % nivå.
 Hvis $\beta = 0$, hva er da $P(|\hat{\beta}| > 8,5)$?
 Hva vil du estimere gjennomsnittsprisen på nye leiligheter til?

Residuals:

Min	1Q	Median	3Q	Max
-923.7	-340.0	-120.0	168.9	1820.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	360	233	1.55	0.129
size	42	3.86	10.88	8.47e-15

s: 553 on 50 degrees of freedom

Multiple R-squared: 0.70

Tabell 3: Resultater fra Modell 1.

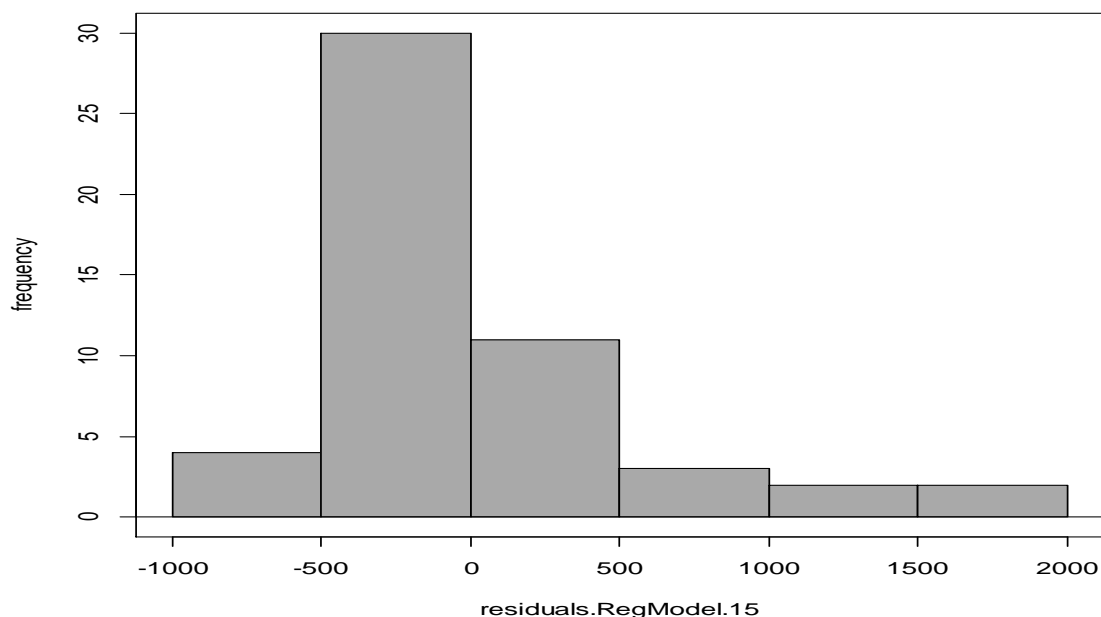
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3399			
age	-8.5	3.03	-2.81	0.007

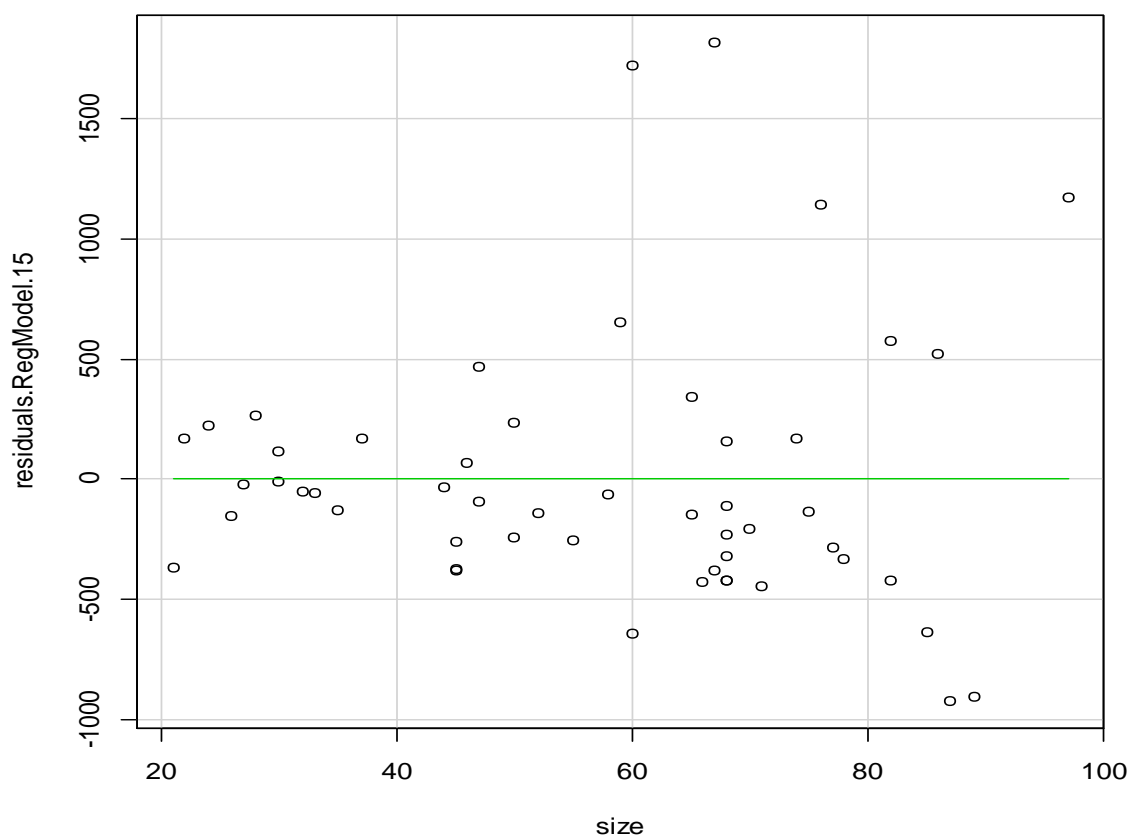
s: 944 on 50 degrees of freedom

Multiple R-squared: 0.14,

Tabell 4: Resultater fra Modell 2.



Figur1: Residualer fra Modell 1.



Figur 2: Residualer fra Modell 1 plottet mot størrelse.

size	age	price
71	129	2890
68	111	2790
67	107	2790
68	108	2790
60	121	2235
68	112	3100
65	121	2940
78	14	3300
76	7	4690
52	14	2400
66	75	2700
59	5	3490
30	75	1735
82	112	3380
55	78	2410
60	73	4600
67	4	4990
65	5	3430
68	80	2895
22	131	1450
87	115	3085
89	122	3190
45	6	1990



86	8	4490
77	75	3305
97	7	5600
82	71	4375
50	116	2215
68	81	2980
44	108	2175
24	61	1590
26	74	1300
35	118	1700
58	26	2730
30	121	1610
47	151	2240
45	44	1870
50	96	2690
46	75	2360
70	82	3090
33	81	1690
45	118	1875
85	71	3290
28	91	1800
74	93	3635
32	119	1650
21	73	875
75	118	3370
68	15	3370
37	84	2080
27	6	1470
47	4	2800

Tabell 5. Data for oppgave II.



Flervalgsoppgaver

Dersom du får andre svar enn de som er oppgitt, kan det skyldes avrundning. Velg det alternativet som er nærmest.

For oppgavene 1- 4 anta følgende:

Et firma leverer sikkerhetsprogram for industrien. De påstår at dette har en positiv effekt. For å kontrollere denne påstanden ble programmet gjennomført i 4 større bedrifter, der en registrerte ukentlige arbeidsulykker før og etter at programmet ble gjennomført. Dette ga:

Bedrift	Ulykker før	Ulykker etter
A	10	7
B	12	11
C	32	26
D	4	2

La X_i være antall ulykker i bedrift nr. i før innføring av programmet, mens Y_i er antall ulykker i samme bedrift etter innføring av programmet. La $D_i = X_i - Y_i$. Vi antar at D_i – ene er uavhengige og normalfordelte med forventning μ_d og standardavvik σ_d .

Oppgave 1

Hvordan vil du estimere μ_d ?

A: 14,5 B: 11,5 C: 1,87 D: 1,26 E: 3 F: -3

Oppgave 2

Hvordan vil du estimere σ_d ?

A: 2,16 B: 1,87 C: 4,67 D: 11,3 E: 3 F: 1,08

Oppgave 3

Et 90 % konfidensintervall for μ_d er gitt ved:

A: (-0,44; 6,44) B: (1,92; 4,08) C: $\mu_d \pm \sigma_d$
D: (-12,5; 18,5) E: (2,08; 8,08) F: (0,46; 5,54)

Oppgave 4

Testen av $H_0: \mu_d = 0$ mot $H_1: \mu_d > 0$, ga en p-verdi på 0,07. Hvordan tolker du dette resultatet:

A: Det er 7 % sannsynlig at H_0 er sann.

B: Det er 7 % sannsynlig at H_1 er sann.

C: Selv om det ikke var effekt av programmet, er det 7 % sannsynlig å observere en så sterk positiv effekt eller sterkere i våre data.

D: Det er 7 % sannsynlig at vi gjør en feil.

E: $P(\bar{D} > 0) = 0,07$.

F: Selv om det var effekt av programmet, er det 7 % sannsynlig å observere en så svak positiv effekt i våre data.

For oppgavene 5 - 8 anta følgende:

Fra en gruppe har vi observert: X_1, X_2, \dots, X_n der alle $X_i \sim N(\mu_1, \sigma)$



Fra en annen gruppe har vi observert: Y_1, Y_2, \dots, Y_n der alle $Y_i \sim N(\mu_2, \sigma)$. Alle observasjoner er uavhengige. La

$$\text{La } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ og } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad s_1 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \text{ og } s_2 = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Oppgave 5.

Hvordan vil du på best mulig måte estimere $\mu_1 - \mu_2$?

A: $\bar{X} - \bar{Y}$ B: $= \frac{\bar{X} - \bar{Y}}{2}$ C: $s_1 - s_2$ D: $\sqrt{s_1^2 - s_2^2}$ E: $= \frac{\bar{X} - \bar{Y}}{s}$ F: 0

Oppgave 6.

Hvordan vil du på best mulig måte estimere σ ?

A: $\bar{X} - \bar{Y}$ B: $\frac{s_1 - s_2}{2}$ C: $\frac{s_1 + s_2}{2}$ D: $\sqrt{s_1^2 - s_2^2}$ E: $\sqrt{\frac{s_1^2 + s_2^2}{n}}$ F: $\sqrt{\frac{s_1^2 + s_2^2}{2}}$

Oppgave 7.

Hvis du skal teste en hypotese om at gruppene er like mot en påstand om at de er forskjellige, hva blir da nullhypotese (H_0) og alternativ hypotese (H_1)?

A: $H_0: \bar{X} = \bar{Y}$ mot $H_1: \bar{X} \neq \bar{Y}$ B: $H_0: \bar{X} - \bar{Y} = \mu_1 - \mu_2$ mot $H_1: \bar{X} - \bar{Y} \neq \mu_1 - \mu_2$
C: $H_0: \bar{X} = \mu_1$ mot $H_1: \bar{X} \neq \mu_1$ D: $H_0: \mu_1 + \mu_2 = \mu_1 - \mu_2$ mot $H_1: \mu_1 + \mu_2 \neq \mu_1 - \mu_2$
E: $H_0: \mu_1 + \mu_2 = 0$ mot $H_1: \mu_1 + \mu_2 \neq 0$ F: $H_0: \mu_1 - \mu_2 = 0$ mot $H_1: \mu_1 - \mu_2 \neq 0$

Oppgave 8.

Dersom testen i Oppgave 7 får en p-verdi på 0,03 hva er rett å gjøre?

- A) Forkaste H_1 på 5 % signifikansnivå. B) Forkaste H_0 på 5 % signifikansnivå.
C) Ikke forkaste H_0 på 5 % signifikansnivå. D) Påstå at sannsynligheten for H_1 er 0,03.
E) Påstå at sannsynligheten for H_0 er 0,03. F) Forkaste H_0 på 1 % signifikansnivå.

Oppgave 9.

Anta at X_1 og X_2 har samme forventning og samme standardavvik ($\sigma = 1$), men anta at de er avhengige og med korrelasjon på 0,7.

Da er standardavviket til gjennomsnittet av X_1 og X_2 :

A: 0,71 B: 0 C: 0,5 D: 0,92 E: 0,85 F: 1

For oppgavene 10 – 13 gjelder følgende.

I en større undersøkelse om arbeidsmiljøet på et stort universitet ble det plukket ut 100 tilfeldig personer. Disse måtte svare på hva slags stilling de hadde, og hvor fornøyd de var med arbeidsmiljøet.

Dette ga følgende tabell:

	Fornøyd	Misfornøyd
Professor	15	20
1. amanuensis	20	10
Tekniker	20	15

Det ble testet om det var avhengighet mellom yrkesgruppe og oppfatning av arbeidsmiljøet. Dette ga følgende utskrift fra R-commander:



Expected counts:

	Fornøyd	Misfornøyd
Professor	19.25	15.75
1. amanuensis	16.50	13.50
Tekniker	19.25	?

Chi-square components:

	Fornøyd	Misfornøyd
Professor	0.94	1.15
1. amanuensis	0.74	0.91
Tekniker	0.03	?

Oppgave 10

Hva blir testobservatoren (kalt Q i læreboka eller W på forelesning)

A: 10,48 B: 5,99 C: 3,80 D: 2 E: 3,77 F: 0,05

Oppgave 11

Testen fikk en P-verdi (P-value) på 0,15. Kan du dermed forkaste H_0 om at det ikke er sammenheng mellom yrke og oppfatning av arbeidsmiljø på signifikansnivå 10 %, begrunnelsen må også være korrekt?

- A) Ja, fordi Q er større enn 0.
- B) Ja, fordi p-verdien er større enn 0,10.
- C) Nei, fordi så stor p-verdi betyr at H_1 er sann.
- D) Nei, fordi p-verdien er større enn 0,10.
- E) Nei, fordi p-verdien sier ikke noe om sannsynligheten for type 1 feil.
- F) Ja, fordi p-verdien er mindre enn Q.

Oppgave 12

Undersøkelsen ble også brukt for å se om flertallet av universitetets ansatte var fornøyd med arbeidsmiljøet. La p være sannsynligheten for at en vilkårlig valgt ansatt ville si seg fornøyd. (Vi antar at universitetet består kun av disse tre yrkesgruppene)

Hva blir estimert verdi av p:

A: 0,45 B: 0,55 C: 0,5 D: 0 E: 50 F: 0,0003

Oppgave 13

La p være som definert i oppgave 12. Følgende ble testet: $H_0: p = 0,5$ mot $H_1: p > 0,5$.

Hva blir p-verdien til denne testen?

A: 0 B: 1 C: 0,5 D: 0,16 E: 0,02 F: 0,55

For oppgavene 14 – 16. Anta at vi har plukket ut 5 forskjellige distrikter i Tanzania, der vi i hvert distrikt har registrert årlig inntekt på 9 tilfeldig valgte småbrukerfamilier.

Vi ønsker primært å se om det er distriktvise forskjeller. Følgende deler av en tabell fra R-commander er tilgjengelig:

	Df	Sum Sq	Mean Sq	F value
Distrikt		340711		
Residuals		414100		



Oppgave 14

Hva blir F-verdien (F-value) her?

A: 8,23 B: 754811 C: 9202 D: 0,11 E: 9,26 F: 11,25

Oppgave 15.

Vi ønsker å teste H_0 om at forventet inntekt er den samme i alle distrikter, mot H_1 at minst to av distriktene har ulik forventet inntekt. La F være verdien du regnet ut i oppgave 14. Da skal du på 5 % signifikansnivå forkaste H_0 hvis:

A: $F > 2,07$ B: $F > 2,61$ C: $F > 9,26$
D: $F < 2,07$ E: $F < 2,61$ F: $F > 8,23$

Oppgave 16.

Stikkord som passer for modellen som er brukt i Oppgavene 14 og 15 er:

A: Enveis variansanalyse. B: Enkel lineær regresjon. C: Parvise data.
D: To-utvalg. E: Krysstabeller. F: Binomisk modell.

For oppgavene 17 - 20 Anta at modellen $Y_i = \alpha + \beta x_i + \varepsilon_i$ blir tilpasset av disse data:

<u>x:</u>	<u>1</u>	<u>2</u>	<u>4</u>
<u>Y:</u>	<u>2</u>	<u>4</u>	<u>5</u>

Noe som gir tilpasset linje: $\hat{Y} = 1,5 + 0,93x$, og en korrelasjon mellom x og Y som er 0,93.

Oppgave 17

Hva blir residualen til den første observasjonen ($Y = 2$ og $x = 1$)?

A: 2 B: -2,43 C: -0,43 D: 2,43 E: 0,93 F: 1,5

Oppgave 18.

Den predikerte verdien for Y når $x = 3$ er:

A: 3 B: 2,43 C: 0,93 D: 4,29 E: 4,5 F: (-7,9; 16,5)

Oppgave 19

Hva blir R^2 ?

A: 0,86. B: -0,86. C: 0,96. D: 0,14. E: 0,80. F: 0,64.

Oppgave 20

Hvilken tolkning er rett?

A: Når x øker med 1, så øker Y med 0,93.
B: Når x øker med 1, så øker forventet verdi av Y med 0,93.
C: Når x øker med 1, så er estimeres forventet økning i Y med 0,93
D: Når Y øker med 1, så øker x med 1,5.
E: Når x øker med 1, så er Y lik 0,93
F: Når Y øker med 1, så øker x med 0,93



**Riv ut arket og levere dette sammen med besvarelsen.
Bare ett kryss i hver rute.**

Oppgave	A	B	C	D	E	F
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						