

EKSAMENSOPPGAVE

Institutt:	<u>IKBM</u>
Eksamen i:	<u>STAT 100</u> <u>Statistikk</u>
Tidspunkt for eksamen:	<u>12. mai 2016</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 11 sider.

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

Merk: Side 11: Skjema fylles ut og arket leveres inn.

Oppgave 1

Værmeldinger er usikre, og temperaturen som meldes dagen i forveien er ikke alltid den som faktisk observeres når dagen kommer. Vi skal i denne oppgaven se på differensen D mellom målt temperatur og den temperaturen værmeldingen predikerer dagen før. Nedenfor ser du et datasett med slike differenser registrert 10 tilfeldige dager i løpet av ett år. Du skal anta at differansene er uavhengige og kommer fra en normalfordelt populasjon med forventning μ_d og standardavvik σ_d .

-1 1 2 0 0 2 4 -3 1 -1.

Tabell 1. Differanser mellom målt og varslet temperatur.

mean	sd	n
0.5	1.96	10

Tabell 2. Opplysninger til bruk i oppgave 1.

- Forklar hva μ_d måler.
Gi et estimat for μ_d .
Finn estimatorens standardfeil.
- Sett opp nullhypotese og alternativ hypotese for å teste om det er grunnlag i dataene for å påstå at temperaturen i det lange løp er høyere enn det som værmeldingen varsler.
Utfør testen, konkluder på et 5 % signifikansnivå.
- Finn et 95% konfidensintervall for μ_d .
Gi en tolkning av intervallet.
- Data er analysert med en modell for parvis sammenligning. Forklar hvorfor en modell for ikke-parvis sammenligning hadde vært ubrukbar i dette tilfelle.

Oppgave 2

Mengden protein (målt i prosent av foret) er en svært viktig komponent ved foring av oppdrettsfisk. Som et ledd i å øke kunnskapen om dette ble 50 oppdrettsørret gitt forskjellig proteinmengde, samtidig som tilveksten (i gram) ble målt for en 8 måneders periode.

For å se effekten av proteinmengde på tilvekst ble en lineær regresjonsmodell tilpasset data.

Coefficients:

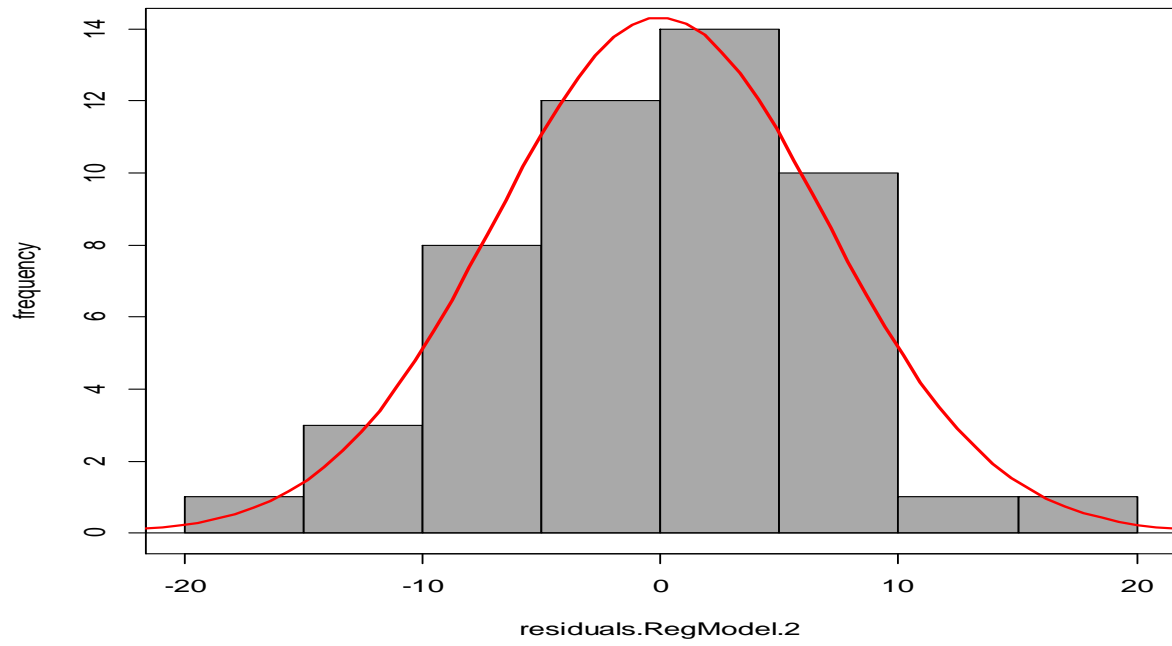
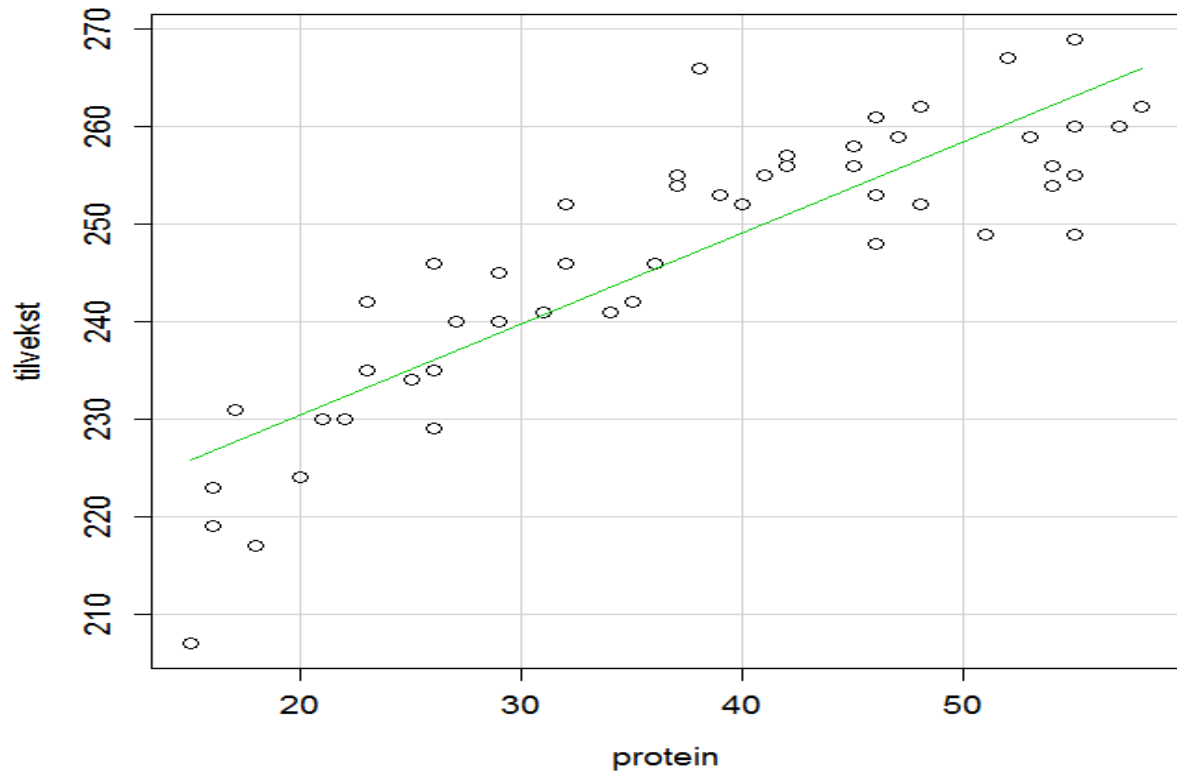
	Estimate	Std. Error
(Intercept)	211.8	3.079
protein	0.93	0.078

s: 7.1. Multiple R-squared: 0.75.

```
predict_CI_PI(RegModel.1, data.frame('protein'=40), level=0.99)
fit   lwr.CI   upr.CI   lwr.PI   upr.PI
      246.4   251.9   229.9   268.4
```

Tabell 3: Resultater for oppgave 2

- Skriv ned modellen som er brukt i analysen.
Estimer alle parametere i modellen.
Gi en forklaring til fiskeoppdretteren hva parameterestimatene betyr.
Finn R^2 og gi en forklaring på hva denne måler.
- Skriv ned null hypotese og alternativ hypotese for å teste om det er en positiv sammenheng mellom mengdene protein i foret og tilvekst på fisken.
Utfør testen og konkluder.
- En fiskeoppdretter bestemmer seg for å bruke et fiskefor med 40 % protein.
Hvor mye vil du anslå at tilveksten blir for en ørret?
Angi et 99 % prediksjonsintervall for tilvekst av en enkelt fisk.
Gi en forklaring på hva dette intervallet betyr.
Hvilke faktorer påvirker bredden av et slikt 99 % prediksjonsintervall?
- Hva er et residual i denne analysen?
Finn residualen til den første ørreten i Tabell 4.
Det er også lagt ved 2 figurer.
Skisser grovt et plot med residualer på y-aksen og protein på x-aksen.
Hvor godt synes du modellen passer til data?



Ørret	tilvekst	protein
1	207	15
2	219	16
3	223	16
4	231	17
5	217	18
6	224	20
7	230	21
8	230	22
9	235	23
10	242	23
11	234	25
12	235	26
13	229	26
14	246	26
15	240	27
16	240	29
17	245	29
18	241	31
19	246	32
20	252	32
21	241	34
22	242	35
23	246	36
24	255	37
25	254	37
26	266	38
27	253	39
28	252	40
29	255	41
30	256	42
31	257	42
32	258	45
33	256	45
34	248	46
35	253	46
36	261	46
37	259	47
38	262	48
39	252	48
40	249	51
41	267	52
42	259	53
43	254	54
44	256	54
45	269	55
46	249	55
47	255	55
48	260	55
49	260	57
50	262	58

Tabell 4: data til oppgave 2

Flervalg

For Oppgavene **F1 - F4**

Dette er basert på artikkelen: *The antihypertensive effects of fish Oil*; av Knapp & FitzGerald publisert i *New England Journal of Medicine* 1989.

14 menn med høyt blodtrykk ble tilfeldig valgt ut, og deretter basert på loddtrekning ble 7 av disse satt på en diett basert på fiskeolje, mens de resterende 7 gikk på en diett som brukte helt vanlig matolje. Reduksjonen i diastolisk blodtrykk (målt i Torr, en målestokk for trykk) sammen med deler av en R-commander utskrift er gitt i Tabell 5.

La X_i være blodtrykksreduksjon for person nummer i på fiskeolje og Y_i blodtrykksreduksjon for person nummer i på matolje. Anta modellen $X_i \sim N(\mu_1, \sigma)$, og $Y_i \sim N(\mu_2, \sigma)$.

Vi antar at alle observasjoner er uavhengige:

Diett med fiskeolje:	8	12	10	14	2	0	0
Diett med vanlig olje:	-6	0	1	2	-3	-4	2

	mean	sd	n
Fiskeolje	6.57	5.85	7
Vanlig olje	-1.14	3.18	7

t = 3.0621, df = 12, p-value = 0.005

alternative hypothesis: true difference in means greater than 0

Tabell 5: Reduksjon i blodtrykk samt deler av R-commander utskrift

F1

Estimert forventet gevinst (i form av redusert blodtrykk) ved bruk av fiskeolje framfor vanlig olje er:

A) 5,43 B) -5,43 C) 6,57 D) 2,03 E) 0 F) 7,71

F2

Hva er det beste estimatet for σ ?

A) 4,51 B) 4,71 C) 5.85 D) 22,18 E) 2,67 F) 7,6

F3

Testing av

H_0 : Valg av oljetype har ingen betydning for blodtrykksreduksjon, mot

H_1 : Fiskeolje har positiv betydning for blodtrykksreduksjon

kan formuleres som:

A) $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

B) $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$

C) $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 < \mu_2$

D) $H_0: \mu_1 \neq \mu_2$ $H_1: \mu_1 = \mu_2$

E) $\bar{X} = \bar{Y}$ $H_1: \bar{X} > \bar{Y}$

F) $\bar{X} > \bar{Y}$ $H_1: \bar{X} > \bar{Y}$

F4

Hva er rett?

- A) Hvis $\mu_1 = \mu_2$, så er $P(\bar{X} - \bar{Y} \geq 7,71) = 0,005$
- B) Hvis $\bar{X} - \bar{Y} \geq 7,71$, så er $P(\mu_1 = \mu_2) = 0,005$
- C) Hvis $\mu_1 > \mu_2$, så er $P(\bar{X} - \bar{Y}) = 0,005$
- D) $P(\mu_1 = \mu_2) = 0,005$
- E) $P(\mu_1 > \mu_2) = 0,995$
- F) Hvis $\bar{X} = \bar{Y}$, så er $P(\mu_1 = \mu_2) = 0,005$

For oppgavene F5 - F8

Forurensningsstoffene PCB (Polychlorinated biphenyls) representerer en stor trussel mot naturens økosystemer. Reproduksjonen hos pelikaner på USAs vestkyst er betraktelig redusert, da skallet på eggene ofte er så tynt at egget ikke makter å bære fram levende unger. Vi skal studere data som knytter tykkelsen på eggeskallet hos pelikaner til forekomsten av giftstoffene PCB. Vi har egg fra 54 pelikaner. Responsvariabel er tykkelsen på eggeskallet målt i mm, og forklaringsvariabel er konsentrasjonen av PCB i eggeplommen, målt i p.p.m. (parts pr million). Konsentrasjonen av PCB i eggeplommen er direkte knyttet til PCB-konsentrasjonen i blodet til pelikanen som la egget. For å analysere data brukte vi regresjonsmodellen:

$$\text{Skalltjukkelse}_i = \alpha + \beta \cdot \text{PCBkonsetrasjon}_i + \varepsilon_i,$$

$\varepsilon_i \sim N(0, \sigma)$ og alle ε_i -er er uavhengige.
 $i = 1, 2, \dots, 54$

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0,37	0,027	13,93	0,000
PCB	-0,0003	0,00012	-2,5	0,019

s: 0,059 on 52 degrees of freedom
R-squared: 0.102

Predicted Values for New Observations

Fit	SE Fit	95% CI	95% PI
0,31483	0,00807	(0,29863; 0,33103)	(0,19483; 0,43482)

New Obs PCB: 200

Tabell 6. For oppgavene F5 – F8

F5

Et estimat for gjennomsnittlig egg tjukkelse i mm når det ikke er PCB tilstede er:

- A) 5,43
- B) 0,37
- C) 6,57
- D) 2,03
- E) 0
- F) -5,43

F6

Hvilket utsagn er rett.

A: Ved en økning på 1 p.p.m. PCB anslår vi at gjennomsnittlig skalltjukkelse avtar med 0,0003 mm

B: Ved en økning på 1 p.p.m. PCB anslår vi at gjennomsnittlig skalltjukkelse øker med 0,0003 mm

C: Ved en økning på 1 mm skalltjukkelse anslår vi at gjennomsnittlig PCB innhold avtar med 0,0003 p.p.m.

D: Ved en økning på 1 mm skalltjukkelse anslår vi at gjennomsnittlig PCB innhold øker med 0,0003 p.p.m.

E: Dersom skalltjukkelse holdes konstant vil PCB øke med 0,37 p.p.m.

F: Dersom skalltjukkelse øker med 1 mm, vil PCB holdes konstant.

F7

Hvordan vil du teste H_0 : Det er ingen sammenheng mellom PCB og skalltjukkelse, mot H_1 : Det er sammenheng mellom PCB og skalltjukkelse

A) $H_0: \beta = 0$, mot $H_1: \beta \neq 0$ B) $H_0: \beta \neq 0$, mot $H_1: \beta = 0$ C) $H_0: \alpha = 0$, mot $H_1: \alpha \neq 0$

D) $H_0: \alpha = \beta$, mot $H_1: \alpha \neq \beta$ E) $H_0: \beta = 0$, mot $H_1: \beta < 0$ F) $\beta = 0$, mot $H_1: \beta \neq 0,0003$

F8

Dersom PCB innholdet måles til 200 p.p.m., hva vil du anslå skalltjukkelse til å være (i mm)?

A) 0 B) 0,059 C) 0,37 D) 0,3694 E) 0,31 F) 0,0003

F9

Anta at $\text{var}(X_1) = \text{var}(X_2) = 1$. Anta at korrelasjonen mellom X_1 og X_2 er 0,9. Da er standardavviket til $(X_1 - X_2)$:

A) 2 B) 0,04 C) 3,8 D) 0,2 E) 1,95 F) 0,45

For oppgavene F10 og F11:

La Y_{ij} være observasjon nummer j av en tilfeldig variabel tilhørende gruppe i . Alle observasjoner antas å være uavhengige. Modellen vi bruker for å analysere disse data er:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, 5 \text{ og } j = 1, 2, \dots, 10. \text{ Her er } \varepsilon_{ij} \sim N(0, \sigma)$$

Noen resultater av analysen får du her.

	Df	Sum Sq	Mean Sq	F value
gruppe		4		
Residuals		9		

F10

Hva blir F verdien (F-value)?

A) 5 B) 0,44 C) 2,61 D) 2,04 E) 0 F) 2,25

F11Hva blir estimatet for σ ?

- A) 5 B) 0,45 C) 0,2 D) 1 E) 0 F) 3

For **F12 – F17**

Julestjernerdyrking har blitt big buisness, men for å få solgt plantene i en kort sesong rett før jul, er det et viktig at blomstene ser pene ut og at de er omtrent 25 cm høye. For å undersøke hvordan forskjellige gjødseltyper påvirker høyden på julerestjernene ble det utført et forsøk der høyde (målt i cm) ble registret etter 65 dager fra planting. Det ble benyttet 5 gjødseltyper (Kalt A, B, C, D, E)

Data:

A	B	C	D	E
22	27	22	23	23
25	24	24	22	29
27	25	23	19	26
28	25	23	29	24
29	30	26	20	22
28	25	23	22	24

Tabell 7: Data for F12 – F17.

Data ble analysert etter følgende modell

$Y_{ij} = \mu_i + \varepsilon_{ij}$, $i = 1, 2, \dots, 5$ og $j = 1, 2, \dots, 6$. Her er $\varepsilon_{ij} \sim N(0, \sigma)$ og med følgende resultat:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gjødsel	4	72.2	18.050	2.766	0.049
Residuals	25	163.2	6.527		

	mean	sd	data:n
A	26.50	2.58	6
B	26.00	2.19	6
C	23.17	1.60	6
D	22.50	3.50	6
E	24.67	2.50	6

Tabell 8: Analyse for F12 – F17.

F12Hvordan vil du estimere μ_1 ?

- A) 25 B) 26,5 C) 18,05 D) 2,58 E) 24,3 F) 6,527

F13

Anta at $\varepsilon_{11} < 0$, det betyr i så fall:

- A) Den første blomsten med gjødsel A er kortere enn alle de andre blomstene.
- B) Den første blomsten med gjødsel A er kortere enn alle de andre med samme gjødsel.
- C) Det er 50 % sjanse for at blomsten med gjødsel A er kortere enn 25 cm.
- D) Den første blomsten med gjødsel A er kortere enn snittet av de 6 med samme gjødsel.
- E) Dersom vi velger å bruke gjødsel A på et stort antall blomster, vil den første blomsten med gjødsel A være kortere enn snittet av disse.
- F) Den første blomsten med gjødsel er større enn resten av blomstene.

F14

Hvordan vil du forklare σ ?

- A): Spredning målt som standardavvik mellom gjødseltyper
- B): Populasjons-Standardavvik i høyde for alle julestjerner som bruker samme gjødseltype
- C) Gjennomsnittshøyde i populasjonen.
- D) Populasjons-Standardavvik i høyde for alle julestjerner.
- E) Standardavvik i utvalget for julestjerner som bruker en spesiell gjødseltype.
- F) Standardavvik i høyde mellom gjødseltyper.

F15

Hva er e_{11} (residualet til første blomst med gjødseltype A)?

- A) 25 cm B) -3 cm C) 2,3 cm D) -2,3 cm E) 3 cm F) -4,5 cm

F16

Hvilket utsagn er rett?

- A): Sannsynligheten for at gruppene er like er 0,049
- B): Sannsynligheten for at gruppene er forskjellige er 0,049
- C) Sannsynligheten for at vi får det resultatet vi fikk er 0,049
- D) Bare 4,9 % av observasjonene er sannsynlige.
- E) Sannsynligheten for å observere så store forskjeller mellom gruppene, hvis gruppeforventningene er like er 0,049
- F) Sannsynligheten for å observere så store forskjeller mellom gruppene, hvis gruppeforventningene er forskjellige er 0,049.

F17

Gjødseltypene A og E er rike på nitrogen, mens de resterende er fattige på nitrogen. Kontrasten som fanger forventet effekt av sterk nitrogengjødsling estimeres til.

- A) 1,69 cm B) 0,5 cm C) 25,6 cm D) 1,93 cm E) 0 cm F) 24,6 cm

F18

En annet gartneri valgte å prøve ut 3 gjødseltyper, de oppsummerte forsøket slik:

	mean	sd	data:n
A	26	1	3
B	25	1	3
C	24	2	3

Tabell 9: Oppsummerte data for F18

Hva blir F-verdien i dette forsøket?

- A) 0,5 B) 0 C) 1,5 D) 6 E) 1,22 F) 0,67

For F19 og F20

1154 barn ble trukket ut tilfeldig ut. Deretter undersøkte vi om de var ulydige, og hvilken plass de hadde i søskenrekka.

Dette ga følgende tabell.

	Eldst	Midten	Yngst	Enebarn	Total
Ulydig	24	29	35	23	111
Lydig	450	312	211	70	1043
Total	474	341	246	93	1154

Fra En kjøring i R-commander sakser vi?

Expected counts

45.6	32.8	23.7	8.9
228.4	308.2	222.3	84.1

Chi-square components:

	1	2	3	4
1	10.23	0.44	5.43	22.08
2	1.09	0.05	0.58	2.35

Pearson (Q) 42.245

F19

Dersom du tester en nullhypotese om at rekkefølge i søskenrekka og ulydighet er uavhengig variable, får du en testobservator kalt Q i læreboka (og W på forelesning).

Hvor stor må denne være dersom du på 5 % signifikansnivå vil forkaste nullhypotesen.

- A) 3 B) 3,84 C) 7,81 D) 42,25 E) 2,35 F) 8

F20

P-verdien på testen vil være

- A) større enn 10 % B) mellom 1 % og 2 % C) mellom 2 % og 5 %
D) mellom 5 % og 10 % E) mindre enn 0,2 % F) Mellom 0,2 % og 1 %

**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						

