

Eksamen STAT100, **Høst 2011** (lettere revidert).

Tillatte hjelpemidler: C3: alle typer kalkulator, alle andre hjelpemidler

Ved alle hypotesetester skal både nullhypotese og alternativ hypotese skrives ned.

Dersom det ikke er oppgitt, må du forklare hvilket nivå du tester på.

Oppgave 1

Medisinen Nembuxil påstås å virke mot migrene hos kvinner som er plaget av dette. Anta at for denne populasjonen kvinner er antallet migreaneanfall i løpet av et år tilnærmet normalfordelt. For å teste ut medisinen ble 9 kvinner som alle hadde vært plaget av migrene trukket ut. Disse fikk registrert antall migreaneanfall i ett år. Dette ga følgende:

Kvinne	1	2	3	4	5	6	7	8	9
Antall	12	21	21	9	14	6	20	18	17

Tabell 1a. Antall migreaneanfall i året før kvinnene fikk medisin.

Året etter gikk de samme 9 på Nembuxil og fikk registrert antall migreaneanfall i dette året. Dette ga følgende:

Kvinne	1	2	3	4	5	6	7	8	9
Antall	5	17	18	7	14	5	11	17	18

Tabell 1b. Antall migreaneanfall i året der kvinnene fikk medisin.

Det er vedlagt to mulige analyser for dette forsøket (se Tabell 2).

a) Hvilken av de to analysene er den eneste akseptable å bruke? Svaret må begrunnes. Sett opp en modell for forsøket.

Estimer forventet effekt av Nembuxil med hensyn på reduksjon av migreaneanfall.

b) Har vi bevist at Nembuxil har positiv effekt? Svar ved hypotesetesting. Velg signifikansnivå selv.

Paired t-test		
sample estimates:		
mean of the differences	2.9	std.dev. of the differences
		3.3

Two Sample t-test	sample estimates:	
mean of antall før medisin	15.3	mean of antall etter medisin
		12.4
	pooled std.dev.	
	5.5	

Tabell 2. To mulige analyser av problemet i oppgave 1.

Det er tre forskjellige typer av Nembuxil, vil kaller disse for A, B og C. I et større forsøk ville en prøve om effekten er avhengig av type. Det ble valgt ut tilfeldig 100 kvinner, deretter ble de forskjellige typene fordelt etter loddtrekning mellom disse.

Dette førte til at 32 fikk A, 31 fikk B, og 37 fikk C. Det ble deretter rapportert hvor mange som hadde fått redusert antallet anfall for hver type (Tabell 3).

Medisintype	A	B	C
Antall som fikk redusert migreneanfall	24	26	27

Tabell 3. Antall som fikk redusert migreneanfall ved bruk av forskjellige typer Nembuxil.

c) Kan du påstå at de to variablene ”medisintype” og ”reduksjon i migreneanfall” er avhengige?

Oppgave 2 (Merk at det er mye å hente i tabellene 4, 5 og 6)

Denne oppgaven dreier seg om et datasett over ville bjørner, som er blitt bedøvet og deretter veid og målt på ulike måter. Et formål med studien er å kunne si noe om bjørners vekt basert på data om kroppsstørrelse og kjønn. Dette kan være nyttig ved framtidige observasjoner av bjørner som kan måles, men hvor en vekt ikke er tilgjengelig. I denne oppgaven skal vi benytte bjørnenes brystomkrets og deres nakkeomkrets som mulige forklaringsvariable (begge målt i cm). Vekt (målt i kg) er responsen. Data følger som vedlegg på slutten av oppgaven.

Vi bruker modellene:

Modell 1: $Y_i = \alpha + \beta x_i + \varepsilon_i$, der alle ε_i er uavhengige, og der $\varepsilon_i \sim N(0, \sigma)$. (Brystomkrets)

Modell 2: $Y_i = \alpha + \beta x_i + \varepsilon_i$, der alle ε_i er uavhengige, og der $\varepsilon_i \sim N(0, \sigma)$. (Nakkeomkrets)

Y_i er vekt på bjørn i , og x_i er henholdsvis brystomkrets og nakkeomkrets på bjørn i .

For $i = 1, 2, \dots, 24$.

a) For begge modeller:

Gi en tolkning av alle parametre (3 i hver modell).

Estimer begge regresjonsparametre (α og β)

Vis at for Modell 1 er $\hat{\sigma} = 21,12$ og for Modell 2 er $\hat{\sigma} = 28,01$.

b) For begge modeller finn R^2 og gi en forklaring på hva denne måler.

c) Velg den modellen du synes passer best.

Lag et 95 % konfidensintervall for β i denne modellen.

Forklar hvordan du ved hjelp av dette konfidensintervallet kan teste hypotesen om at det er en sammenheng mellom vekt og forklaringsvariabel.

d) Vi skyter en ny bjørn og måler brystomkrets til 90 cm og nakkeomkrets til 54.3 cm.

Bruk begge modeller til å anslå vekt på denne bjørnen.

Bruk begge modeller til å lage et 99 % prediksjonsintervall for vekt på denne bjørnen.

Hvorfor blir prediksjonsintervallet bredere ved bruk av Modell 2 framfor Modell 1?

e) I Figur 1 har vi plottet residualer mot tilpassede y -verdier for Modell 1. På plottet er det også merket av 4 spesielle bjørner (med nummer), samt markert forskjellig kjønn.

Forklar hvordan vi kan bruke et residualplot til modellevaluering.

Kan du ved hjelp av Figur 1 se noen svakheter ved modellen?

Hva sier Figur 1 om bjørnene med nummer: 8, 14, 18 og 22?

Hvilke problemer ser du fra Figur 1 med hensyn på å ha hann- og hunnbjørner i samme modell?

Brystomkrets 90.0
 Nakkeomkrets 54.3

Tabell 4: Gjennomsnittsverdier for brystomkrets og nakkeomkrets.

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-133.52	16.68	-8.005
Brystomkrets	2.73	0.18	15.276

	Df	Sum Sq
Brystomkrets	1	104124
Residuals	22	9816

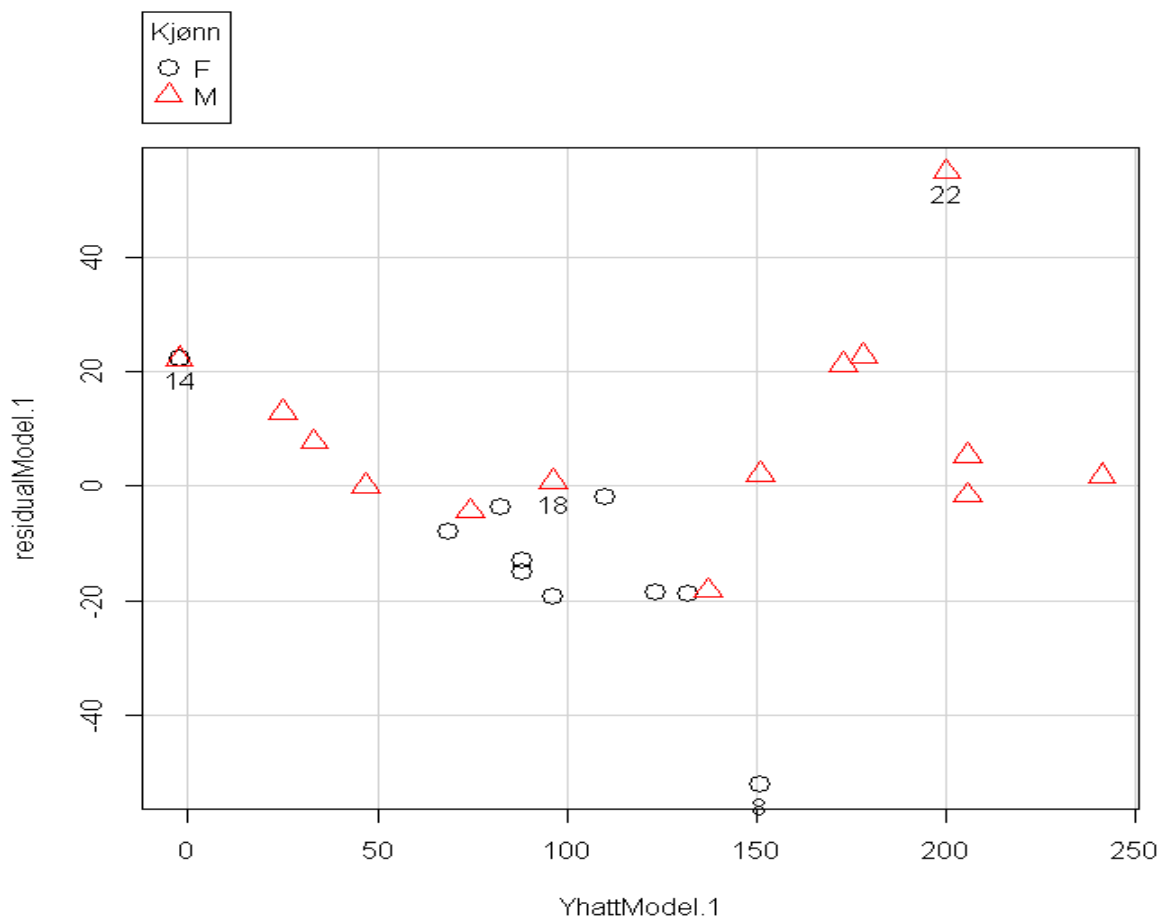
Tabell 5: Resultater for Modell 1.

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	-118.23	21.57	-5.482
Nakkeomkrets	4.25	0.38	11.100

	Df	Sum Sq
Nakkeomkrets	1	96678
Residuals	22	

Tabell 6: Resultater for Modell 2.



Figur 1: Residualer mot tilpassede verdier (Yhatt) for Modell 1 beskrevet i oppgave 2. Nummer representerer enkeltbjørner og det er brukt forskjellig symbol for hvert kjønn, der F er hunnbjørn (binne) og M er hannbjørn

Oppgave 3

Denne oppgaven dreier seg om å se hvorvidt miljøgifter som PCB (polychlorinated biphenyls) påvirker reproduksjon hos fugler. Vi har valgt oss ut Pelikaner og studert hvordan PCB påvirker skalltjukkelsen på egg fra disse. Blir skallet for tynt går eggene i stykker og fuglene kan få store problemer med reproduksjon. Det ble valgt ut 3 områder

- Ett område i California, der PCB er til stede i sterk grad.
- Ett område i Florida, der det er moderat PCB forurensning.
- Ett område i Honduras, der PCB nesten ikke er til stede.

Fra hvert område samlet vi 20 egg og målte skalltjukkelsen i millimeter. Data ble analysert etter hva vi kaller Modell 3. Noen av resultatene er gitt i Tabell 7.

- a) Forklar detaljert hvilken modell som er brukt?
Estimer alle parametre i denne modellen.
Sett opp en test der du undersøker om mengden PCB (her blir dette representert ved sted) har betydning for skalltjukkelse.
Konkludere på testen.
P-verdien på testen er 0,034. Hvordan vil du forklare hva dette betyr til en fugleekspert som ikke kan noe statistikk?
- b) Test på 10 % signifikansnivå om det er forskjell i skalltjukkelse mellom området i California og det i Florida.

	Df	Sum Sq	Mean Sq	F value
Sted		0.015		
Residuals (error)		0.126		
	mean	standard deviation		
California	0.288	0.055		
Florida	0.317	0.043		
Honduras	0.325	0.042		

Tabell 7. Resultater for analyse i oppgave 3.

Vedlegg: Data til Oppgave 2.

Vekt er i kg, lengde, nakke (omkrets) og bryst (omkrets) er i cm, alder er i måneder, For kjønn så er F hunn- (Female) og M hannbjørn (Male).

Nr	Vekt	Lengde	Alder	Nakke	Bryst	Kjønn
1	38	114	19	38	58	M
2	41	121	19	51	61	M
3	47	135	19	41	66	M
4	201	171	55	71	114	M
5	243	183	81	79	137	M
6	204	183	115	80	124	M
7	108	157	104	56	89	F
8	99	178	100	53	104	F
9	153	187	56	67	104	M
10	211	174	51	69	124	M

11	113	163	57	51	97	F
12	79	147	53	46	79	F
13	194	185	68	74	112	M
14	20	94	8	33	48	M
15	75	160	44	27	81	F
16	105	170	32	55	94	F
17	61	132	20	44	74	F
18	97	150	32	55	84	M
19	119	163	45	61	99	M
20	20	91	9	30	48	F
21	70	150	21	48	76	M
22	255	183	177	76	122	M
23	73	146	57	48	81	F
24	77	155	81	51	84	F