



EKSAMENSOPPGAVER

Institutt:	IKBM	
Eksamen i:	STAT100	STATISTIKK
Tid:	Torsdag 9.des 2010	09.00-12.30 (3.5 timer)
Emneansvarlig:	Solve Sæbø (Kontakt: Trygve Almøy, 64 96 58 20, under eksamen)	

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

Opgaveteksten er på: 4
antall sider inkl. vedlegg

Alle deloppgaver teller likt.

Alle steder der du utfører statistiske tester, skal du skrive opp både nullhypotesen og den alternative hypotesen.

Oppgave 1

Fører salting av veier om vinteren til flere drepte i trafikken? Motstandere av vei-salting om vinteren hevder at saltet fører til flere trafikkdrepte. Vei-salting ble innført i Norge på begynnelsen av 1970-tallet og for å undersøke om salting øker faren for dødsulykker, sammenliknet man antall drepte i vintermånedene (des-apr) i perioden 1965-1970 (kall denne perioden P1) med perioden 1971-1976 (P2). Data fra statistisk sentralbyrå :

Antall trafikkofer for de ulike periodene fordelt på måned

Mnd	P1	P2
des	259	295
jan	178	252
feb	146	196
mar	158	210
apr	134	198

Bruk følgende del-utskrifter fra R-commander i den grad du synes det nødvendig for å besvare spørsmålene nedenfor.

Paired t-test

data: P2 and P1

t = 8.5338, df = 4, p-value = *

alternative hypothesis: true difference in means is greater than 0

	mean	sd	n
P2-P1	55.2	14.46	5

Two Sample t-test

data: P2 and P1

t = 1.8842, df = 8, p-value = *

alternative hypothesis: true difference in means is greater than 0

	mean	sd	n
P1	175.0	49.69	5
P2	230.2	42.69	5

- Formulér en modell for dataene og tolk modell-parametrene. Finn estimater for parametrene i modellen.
- Test på 5% nivå om veisalting har ført til en økning i antall trafikkdrepte.

En kritiker til analysen gjort ovenfor mente at en eventuell forskjell i antall drepte i de to periodene kan skyldes en generell endring i antall ulykker over hele året og at man derfor også burde se på hva som skjer i sommermånedene (mai-okt). Tabellen nedenfor gir totalt antall ofre i trafikken i vintermånedene og sommermånedene i hver av de to periodene.

	P1	P2	Sum
vinter	875	1151	2026
sommer	1544	1407	2951
Sum	2419	2558	(n=4977)

- Test på 5% nivå om det er avhengighet mellom årstid (Vinter, Sommer) og periode (P1, P2), og konkluder om resultatet av testen støtter påstanden om at veisalting fører til økt antall drepte i trafikken.

Oppgave 2

Tiden det tar før blod koagulerer (størkner) kan avhenge av mange faktorer, og i en studie ble det undersøkt om ulike dietter kan ha en effekt på koaguleringstiden. I alt ble 4 dietter (A, B, C og D) prøvd ut på til sammen $N=24$ pasienter og koaguleringstiden ble registrert. Bruk R-utskriftene nedenfor i den grad du finner det nødvendig til å besvare spørsmålene. Noen tall er utelatt og erstattet med tallene (1)-(7).

	Mean	sd	n
A	61	1.8	4
B	66	2.8	6
C	68	1.7	6
D	61	2.6	8

Analysis of Variance Table

Response: coagulation

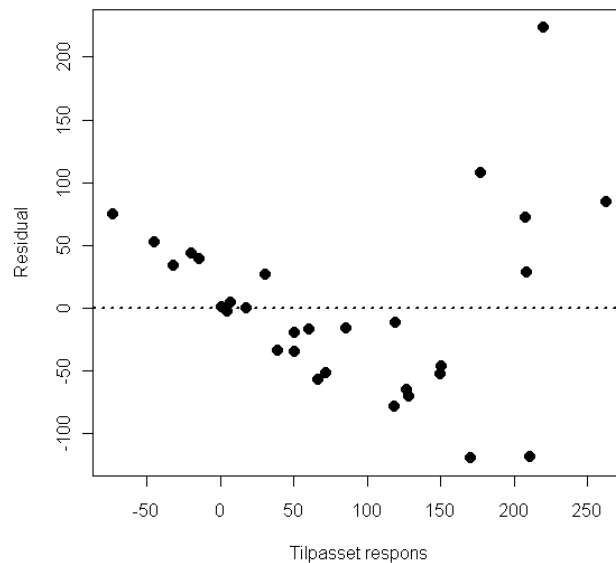
	Df	Sum Sq	Mean Sq	F value
Diet	(1)	228	(5)	(7)
Residuals	(2)	(4)	(6)	
Total	(3)	340		

- Sett opp en modell for forsøket og estimér alle modellparametrene. Finn alle tallene i utskriften, (1) – (7), som mangler.
- Formulér hypoteser for å teste om det er diett-effekt på koaguleringstiden. Bruk 5% nivå på testen.
- Diett D er en kontroll-diett som man ønsker å sammenlikne de andre diettene med. Definér en kontrast for å sammenlikne diett B med kontroldietten D og test på 1% nivå om det er forskjell på disse diettene.

Oppgave 3

En forskningsgruppe ville undersøke hvordan antall arter (Z) av skilpadder som finnes på de ulike Galapagosøyene avhenger av øyenes areal. Arealet (X) ble målt som den naturlige logaritmen til antall kvadratkilometer. Forskerne antok en enkel lineær regresjonsmodell for sammenhengen mellom Z og X og kjørte en regresjonsanalyse for å estimere modellparametrene.

- Forklar hva et residual er i en regresjonsmodell. Gi på grunnlag av residual-plottet nedenfor en forklaring på hvorfor forskernes modell ikke var god.



Forskerne prøvde også en alternativ responsvariabel (Y) der man i stedet målte antall arter på logaritmisk skala (naturlig logaritme), dvs $Y = \log(Z)$ (Merk: da er antall arter $Z = e^Y = 2.718^Y$)

- Formulér en modell for sammenhengen mellom Y og X og gi en tolkning av modellparametrene. Hvilke modellantagelser gjør du?
- Bruk del-utskriften (noen tall mangler) nedenfor til å finne estimater for modellparametrene og test om det er signifikant sammenheng mellom X og Y.

```
lm.default(formula = Y ~ X)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.9037      0.1571  18.484 <2e-16
X              0.3886      0.0416    *      *
```

Residual standard error: 0.7842 on 28 degrees of freedom
Multiple R-squared: 0.7571, Adjusted R-squared: 0.7484
F-statistic: * on 1 and 28 DF, p-value: *

- Gi en prediksjon av (anslag på) Y for en øy med areal lik gjennomsnittsarealet på 1.55 (dvs $X = \bar{X} = 1.55$).
Finn også et 95% prediksjonsintervall for Y når $X=1.55$ og gi en tolkning av intervallet.
Hva vil du predikere antall arter (Z) til for en øy av denne størrelsen?

Navn på emneansvarlig: Solve Sæbø
Navn på sensor: John Sølve Tyssedal