

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK1000 — Innføring i anvendt statistikk.

Eksamensdag: Torsdag 1. desember 2011.

Tid for eksamen: 09:00 – 13:00.

Oppgavesettet er på 4 sider.

Vedlegg: Ingen.

Tillatte hjelpemidler: Lærebok, ordliste for STK1000, godkjent kalkulator.

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Oppgave 1. Anta at 150 blodgivere ankommer Blodgiversentralen en gitt uke. Blodgiverne som møter opp kan antas å representere et tilfeldig utvalg fra befolkningen i Norge. Fra før vet vi at 8% av befolkningen i Norge har blodtype B.

a) La X være antall blodgivere med blodtype B blant de 150 som kommer til sentralen denne uken. Hvilken fordeling har X ? Finn forventning og standardavvik til X .

b) Bruk tilnærming til normalfordelingen og beregn sannsynligheten for at minst 15 av de 150 blodgiverne har blodtype B. Er normalfordelingen en god tilnærming i dette tilfellet? Begrunn svaret.

Oppgave 2. I en artikkel med tittelen 'Is Friday the 13th bad for your health?' i anerkjente British Medical Journal, har britiske forskere studert hvordan fredag den 13. påvirker folks oppførsel. Ett av spørsmålene var hvorvidt folk har en tendens til å holde seg mer hjemme på fredag den 13. Dataene i tabellen er antall biler som passerte et visst punkt på motorvei M25 to påfølgende fredager (den 6. og den 13.) for fem forskjellige tidsperioder:

Year	Month	6th	13th
1990	July	134012	132908
1991	September	133732	131843
1991	December	121139	118723
1992	March	124631	120249
1992	November	117584	117263

(Fortsettes på side 2.)

I artikkelen bruker forskerne en parret t-test for å besvare spørsmålet om folk har en tendens til å holde seg hjemme. Resultatet av analysen, utført i Minitab, finner du nedenfor.

a) Forklar hvorfor det er naturlig å bruke en parret analyse her. Forklar hvilke antakelser som ligger bak en slik parret t-test og spesifiser spesielt hvilken parameter t-testen dreier seg om.

b) Skriv ned nullhypotesen og alternativhypotesen. Begrunn valg av alternativhypotese. Estimer parameteren av interesse. Hva er standardfeilen (estimert standardavvik) til dette estimatet? Du kan bruke Minitab-utskriften (men noen tall er fjernet).

c) Vis hvordan t-observatoren og P-verdien er beregnet. Du kan hente tallene du trenger å sette inn, fra Minitab-utskriften. Spesifiser spesielt antall frihetsgrader i t-fordelingen du bruker. Kontroller at du får samme verdi for t-observator som i utskriften og at tilnærmet P-verdi fra tabell er konsistent med resultatet fra Minitab. Formuler en konklusjon på testen dersom signifikansnivået er satt til 0.05.

Paired T for 6th - 13th

	N	Mean	StDev	SE Mean
6th	5	126220	7417	
13th	5	124197	7549	
Difference	5	2022	1539	

T-Test of mean difference = 0 (vs > 0): T-Value = 2,94 P-Value = 0,021

Oppgave 3. Forurensningsstoffene PCB (Polychlorinated biphenyls) representerer en stor trussel mot naturens økosystemer. Reproduksjonen hos pelikaner på USAs vestkyst er betrakelig redusert, da skallet på eggene ofte er så tynt at egget ikke makter å bære frem 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 ppm. Konsentrasjonen av PCB i eggeplommen er direkte knyttet til PCB-konsentrasjonen i blodet til pelikanen som la egget.

a) Sett opp en passende regresjonsmodell, og finn estimater for alle parametrene i modellen fra Minitab-utskriften. Finn også et estimat for forventet eggeskall-tykkelse for egg med et PCB-innhold på 200 ppm. I Minitab-utskriften til slutt i oppgaven heter variablene hhv. THICKNESS og PCB.

(Fortsettes på side 3.)

b) Sett opp hypoteser og uttrykket for test-observator for å avgjøre om det er en signifikant lineær sammenheng mellom PCB-innhold og tykkelsen på eggeskallene. Finn P-verdi fra utskriften og konkluder med et signifikansnivå på 0.05.

c) Forklar hvordan r^2 (R-Sq i Minitab) skal fortolkes. For regresjonsanalysen av eggdataene ser vi at $r^2 = 0.102$. Kommenter kort denne (lave) verdien.

d) De 54 pelikanene i studien hekket på kysten av California. For en annen gruppe pelikaner ved kysten av Florida har vi PCB-målinger fra 48 individer, oppsummert med en gjennomsnittsverdi på 176,51 ppm og empirisk standardavvik på 59,89 ppm. Anta at PCB-verdiene nå representerer tilfeldige utvalg fra to populasjoner (California og Florida) med samme varians, og beregn et 99% konfidensintervall for forventet forskjell mellom PCB-innhold i eggeplommen hos pelikaner i California og i Florida.

e) Bruk sammenhengen mellom konfidensintervall og hypotesetesting til å finne ut om det er signifikant forskjell på forventet PCB-konsentrasjon i eggeplommen hos pelikaner i California og Florida på nivå 0.01.

Minitab-utskriften nedenfor gjelder pelikan-data fra California:

Descriptive Statistics: PCB

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median
PCB	54	203,04	8,71	64,02	46,00	165,50	204,00

Variable	Q3	Maximum
PCB	238,50	356,00

Regression Analysis: THICKNESS versus PCB

The regression equation is

THICKNESS = 0,377 - 0,000309 PCB

Predictor	Coef	SE Coef	T	P
Constant	0,37662	0,02704	13,93	0,000
PCB	-0,0003089	0,0001271	-2,43	0,019

S = 0,0592503 R-Sq = 10,2% R-Sq(adj) = 8,5%

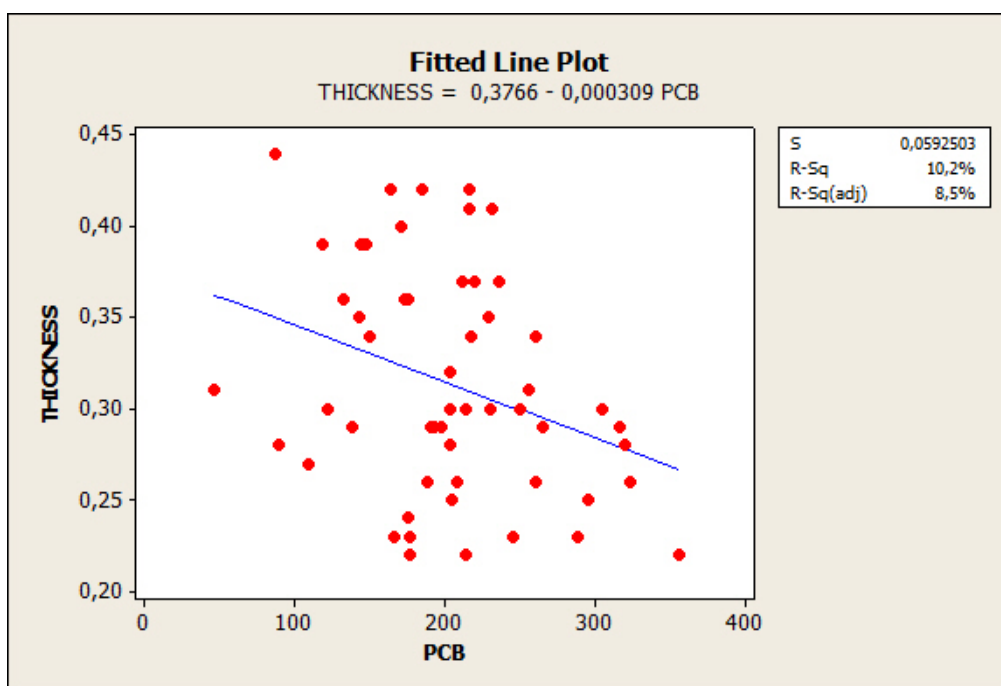
(Fortsettes på side 4.)

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	0,31483	0,00807	(0,29863; 0,33103)	(0,19483; 0,43482)

Values of Predictors for New Observations

New Obs	PCB
1	200



Takk for innsatsen!