

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK 1000 — Innføring i anvendt statistikk.

Eksamensdag: Mandag 4. desember 2006.

Tid for eksamen: 14.30 – 17.30.

Oppgavesettet er på 5 sider.

Vedlegg: Ingen.

Tillatte hjelpemidler: Lærebok: Moore & McCabe "Introduction to the practice of statistics". Ordliste for bruk i STK 1000, kalkulator.

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

### Oppgave 1.

Ved et fotgjengerfelt i en sterkt trafikkert gate står et trafikklys. Fotgjengere som ønsker å krysse gaten trykker på en knapp, og etter en tid vil signalet stoppe trafikken og fotgjengeren kan gå over gaten. For å få kunnskap om hvor lenge trafikken bør stoppes, foretas det en registrering av den tiden tilfeldig valgte fotgjengere bruker på å krysse gaten. Ti slike målinger ga følgende resultat, der tiden er angitt i sekunder:

5 5 8 8 9 12 12 13 17 25

- a) Finn median og gjennomsnitt for kryssingstidene. Disse representerer typisk tidsforbruk ved kryssing av veien. Hvorfor er det ikke tilrådelig å bruke f.eks. medianen som stoppetid for trafikken?

Trafikklyset er nå innstilt slik at fotgjengerne har 15 sekunder til å krysse gaten før trafikken starter på nytt. La  $X$  angi antall blant 10 tilfeldige

(Fortsettes side 2.)

fotgjengere som ikke klarer seg med 15 sekunder, og la  $p$  være sannsynligheten for at en tilfeldig fotgjenger bruker mer enn 15 sekunder på å krysse gaten.

- b) Forklar hvorfor  $X$  kan antas å være binomisk fordelt. Foreslå en estimator for  $p$ , og vis at den er forventningsrett (unbiased).  
Estimer  $p$  ved hjelp av datamaterialet ovenfor.
- c) Finn sannsynligheten for å få to eller flere fotgjengere (blant 10 tilfeldige) som ikke klarer seg med 15 sekunder, dersom sannsynligheten for at en tilfeldig fotgjenger bruker mer enn 15 sekunder egentlig er 0.08.

## Oppgave 2.

Det hevdes at man ved bloddoping (tilførsel av ekstra blod, gjerne eget) kan oppnå bedre prestasjoner i idrett. Et forsøk der man ønsker å studere effekten av bloddoping blir gjennomført på følgende måte:

Ti forsøkspersoner løper 500 meter to ganger. Den første gangen gjennomføres løpet uten doping. Andre gangen, noen dager senere, er forsøkspersonene bloddopet. Vi antar at forsøksbetingelsene er tilnærmet identiske for de to løpene (vind, temperatur etc.). Tidene målt i sekunder er registrert for hver løper og gitt i tabellen nedenfor:

Løper nr.	1	2	3	4	5	6	7	8	9	10
Tid uten doping	118	101	100	107	94	116	133	125	120	100
Tid med doping	115	90	106	100	92	100	124	115	125	92

La  $X_i$  og  $Y_i$  betegne hhv. tid uten og tid med doping for person nr.  $i$ . Anta at differansene  $D_i = X_i - Y_i$ ,  $i = 1, \dots, 10$ , er uavhengige og normalfordelte med forventning  $\delta$  og standardavvik  $\sigma_D$ .

- a) Foreslå en naturlig estimator for  $\delta$ .  
Hva er estimatorens forventning og standardavvik? Hvilken fordeling har estimatoren? Begrunn svaret.

Vi ønsker å undersøke om det ut fra de gitte observasjonene er grunnlag for å påstå at bloddoping forbedrer idrettsprestasjoner.

- b) Formuler nullhypotese og alternativ for denne problemstillingen, med utgangspunkt i  $\delta$ . Sett opp formelen for tilhørende testobservator, sett inn tallene og beregn observatorens numeriske verdi. Du kan gjerne plukke de størrelsene du trenger fra Minitab-utskriften (se nedenfor).
- c) Beregn tilnærmet  $P$ -verdi for testen ved hjelp av tabell D i læreboken. Skriv din konklusjon med ord.

(Fortsettes side 3.)

Data er lagt inn i Minitab. En kolonne kalt 'normal' inneholder de ti tidene uten doping, og en kolonne kalt 'doped' inneholder de ti tidene med bloddoping. Utskrift fra to analyser av datasettet er lagt ved til slutt i i oppgaven.

- d) Forklar kort hva som skiller de to analysene og hvorfor kun en av dem egner seg for det foreliggende datamaterialet. Finn nøyaktig  $P$ -verdi for testen i b).

1.

Paired T-Test and CI: normal, doped

Paired T for normal - doped

	N	Mean	StDev	SE Mean
normal	10	111.400	12.825	4.056
doped	10	105.900	13.195	4.173
Difference	10	5.50000	7.01189	2.21736

95% lower bound for mean difference: 1.43534

T-Test of mean difference = 0 (vs > 0): T-Value = 2.48 P-Value = 0.017

2.

Two-Sample T-Test and CI: normal, doped

Two-sample T for normal vs doped

	N	Mean	StDev	SE Mean
normal	10	111.4	12.8	4.1
doped	10	105.9	13.2	4.2

Difference = mu (normal) - mu (doped)

Estimate for difference: 5.50000

95% lower bound for difference: -4.62249

T-Test of difference = 0 (vs >): T-Value = 0.95 P-Value = 0.179 DF = 17

### Oppgave 3.

Sensommerkvelder i sydlige strøk fylles ofte av karakteristiske lyder: siris-senes sang, frembragt ved at de gnisser vingene mot hverandre meget raskt. Vi skal studere hvordan gnisse-frekvensen varierer med lufttemperaturen for en spesiell art (*Nemobius fasciatus*). Vi har 15 observasjoner av temperatur

(Fortsettes side 4.)

(i °C) og frekvens (gniss per sekund), og vil utføre en regresjonsanalyse med temperatur som forklaringsvariabel og frekvens som respons.

Data er analysert i *Minitab*, der temperaturene er lagt inn i kolonnen 'temp' og frekvensene i kolonnen 'frekv'. Plott og utskrifter er lagt ved til sist i oppgaven.

- Sett opp en enkel lineær regresjonsmodell for disse dataene. Spesifiser antakelsene som ligger til grunn for modellen. Finn fra *Minitab*-utskriften estimer for alle de tre parametrene i modellen.
- Diskuter kort hvor godt modellen passer til data på bakgrunn av vedlagte materiale.

Uavhengig av svaret ditt i b), beholder vi den enkle lineære modellen.

- Finn et 95% konfidensintervall for forventet frekvens utenfor hotellvinduet ditt på Riva del Sole en kveld det er 31°C i luften. Bruk *Minitab*-utskriften!  
Hvor mange frihetsgrader har *t*-fordelingen som brukes i beregningene, og hvilken kritisk verdi  $t^*$  gir det?

Regression Analysis: frekv versus temp

The regression equation is  
 frekv = 6.47 + 0.381 temp

Predictor	Coef	SE Coef	T	P
Constant	6.472	1.877	3.45	0.004
temp	0.38147	0.06968	5.47	0.000

S = 0.971518    R-Sq = 69.7%    R-Sq(adj) = 67.4%

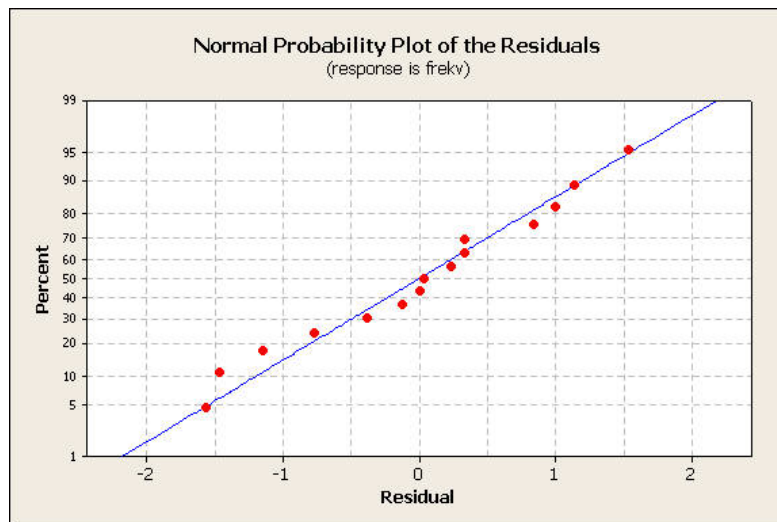
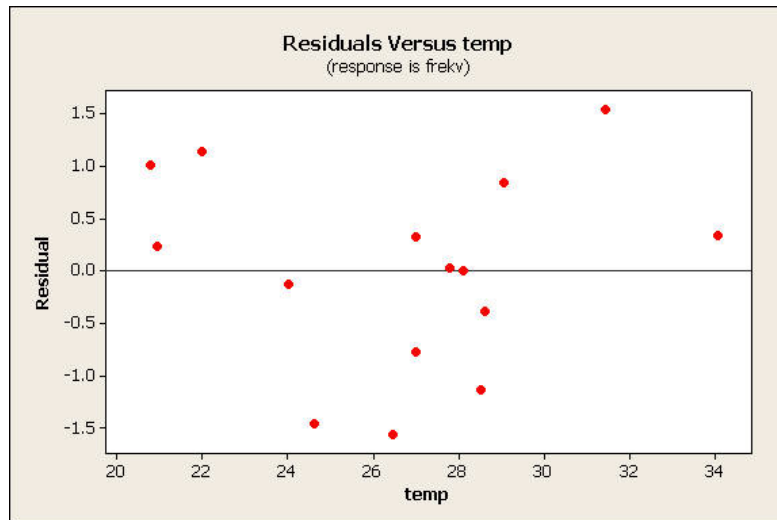
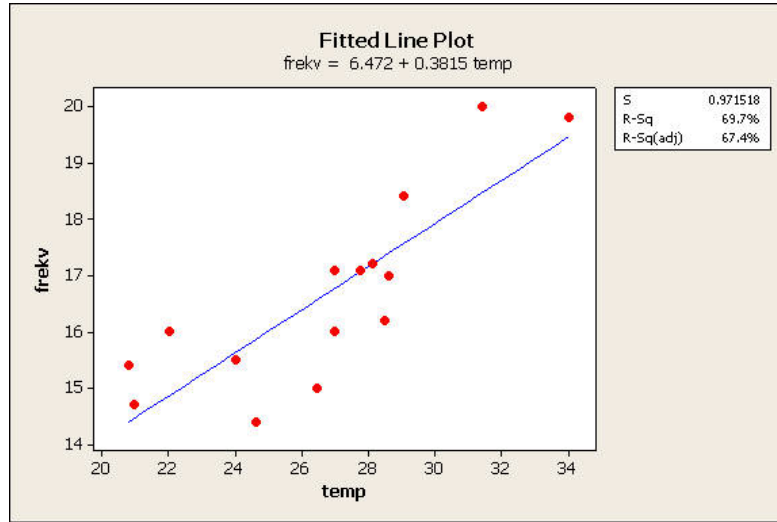
Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	18.298	0.391	(17.452, 19.143)	(16.035, 20.561)

Values of Predictors for New Observations

New Obs	temp
1	31.0

(Fortsettes side 5.)



SLUTT