

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK 1000 — Innføring i anvendt statistikk.

Eksamensdag: Fredag 3. desember 2004.

Tid for eksamen: 14.30 – 17.30.

Oppgavesettet er på 4 sider.

Vedlegg: Tabell over normalfordeling, tabell over  $t$ -fordeling.

Tillatte hjelpemidler: Lærebok: Moore & McCabe "Introduction to the practice of statistics", 3. el. 4. utgave. Ordliste for bruk i STK 1000, kalkulator.

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

### Oppgave 1.

Vi antar at liggelengde (tiden en pasient er innlagt) ved et større sykehus er normalfordelt med forventning  $\mu$  dager og standardavvik 3.68 dager (kjent).

- Anta først at  $\mu$  er kjent lik 4.52 dager. Hva er sannsynligheten for at en tilfeldig pasient blir liggende inne mer enn en uke (7.00 dager)?
- Forventet liggelengde  $\mu$  er egentlig ukjent. Du foretar nå en undersøkelse for å bestemme  $\mu$  ved å velge ut  $n$  tilfeldige pasienter som har vært innlagt i løpet av det siste året og registrere deres liggelengde. Hvor stor må  $n$  være for at et 90%-konfidensintervall for  $\mu$  ikke blir bredere enn 0.5 dager?

(Fortsettes side 2.)

## Oppgave 2.

Forskere ved NASA ønsket å undersøke om romflyvninger medfører endring i antall røde blodlegemer hos rotter. Spesielt så de på RBC (red blood cell) tellinger. Følgende data representerer RBC-tellinger 3 dager før avreise (RBC-L3) og rett etter retur (RBC-R0), samt differansen (DIFF) mellom disse, av 27 rotter sendt ut i rommet med en "Spacelab Sciences 1 flight":

	1	2	3	4	5	6	7	8	9
RBC-L3	7.53	7.79	7.84	6.86	7.93	7.48	7.94	7.95	8.44
RBC-R0	8.16	9.15	9.09	8.42	8.96	9.25	7.40	9.55	9.38
DIFF	-0.63	-1.36	-1.25	-1.56	-1.03	-1.77	0.54	-1.60	-0.94

	10	11	12	13	14	15	16	17	18
RBC-L3	6.70	6.95	6.73	7.21	6.95	7.79	8.24	7.69	7.23
RBC-R0	9.66	8.83	8.46	9.04	9.48	8.52	8.59	7.29	8.93
DIFF	-2.96	-1.88	-1.73	-1.83	-2.53	-0.73	-0.35	0.40	-1.70

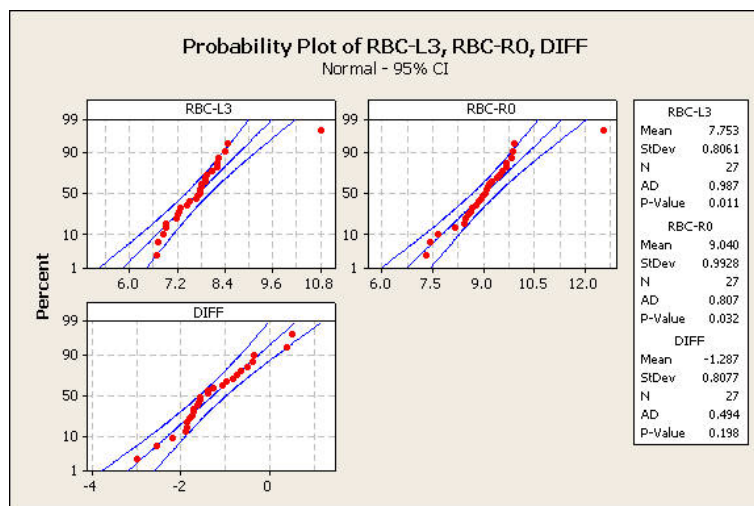
	19	20	21	22	23	24	25	26	27
RBC-L3	7.83	8.09	8.27	7.27	8.23	14.50	7.73	7.31	10.84
RBC-R0	8.64	9.65	8.76	9.11	9.81	9.86	9.90	7.64	12.54
DIFF	-0.81	-1.56	-0.49	-1.84	-1.58	4.64	-2.17	-0.33	-1.70

Oppsummerende statistikk for de tre variablene følger nedenfor

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
RBC-L3	27	7.975	0.294	1.526	6.700	7.230	7.790	8.090	14.500
RBC-R0	27	9.040	0.191	0.993	7.290	8.520	9.040	9.550	12.540
DIFF	27	-1.065	0.269	1.397	-2.960	-1.770	-1.560	-0.630	4.640

- Bør vi her bruke metoder for to-utvalgsdata eller er det naturlig å bruke parvis sammenlikning for å svare på det NASA ønsker å belyse? Begrunn svaret.
- Sett opp en nullhypotese og en alternativ hypotese for problemstillingen. Bør du bruke en  $z$ -test eller en  $t$ -test i dette tilfellet? Begrunn svaret.
- Beregn testobservator for testen. Beregn også P-verdien (eventuelt en øvre grense for P-verdien) og konkluder!
- Det viste seg i ettertid at observasjon nr. 24 av RBC-L3 skulle hatt verdien 8.50 istedet for 14.50. Hvordan vil en slik endring påvirke testobservatoren og P-verdien? Du behøver ikke her å utføre de faktiske utregninger, men kun indikere hvilken retningen endringene går. Vil konklusjonen av testen endres?
- Plottet nedenfor viser normalfordelingsplott (Probability plots) av RBC-L3, RBC-R0 og DIFF. Alle plottene er basert på at observasjon nr 24 av RBC-L3 er korrigert. Er antagelsene som testen baserer seg på tilstrekkelig tilstede her? Begrunn svaret.

(Fortsettes side 3.)



### Oppgave 3.

Tabellen nedenfor viser årlig utslipp til luft av svoveldioksid (i 1000 tonn) i Norge for årene 1993–2002 (Kilde: Statistisk årbok 2004).

År	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Utslipp (i 1000 tonn)	35	35	33	33	30	30	28	27	25	22

Regression Analysis: SO2 versus year

The regression equation is

$$\text{SO}_2 = 2838 - 1.41 \text{ year}$$

Predictor	Coef	SE Coef	T	P
Constant	2838.4	198.2	14.32	0.000
year	-1.40606	0.09922	-14.17	0.000

$$S = 0.901178 \quad R\text{-Sq} = 96.2\% \quad R\text{-Sq}(\text{adj}) = 95.7\%$$

Predicted Values for New Observations

New	Obs	Fit	SE Fit	95% CI	95% PI
	1	22.067	0.616	(20.647, 23.486)	(19.550, 24.583)

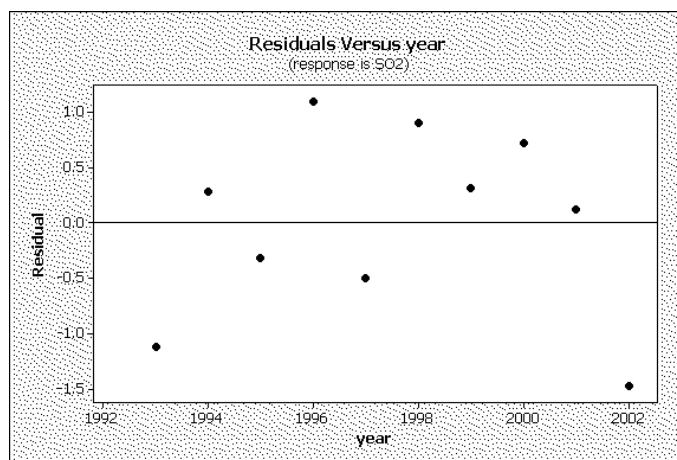
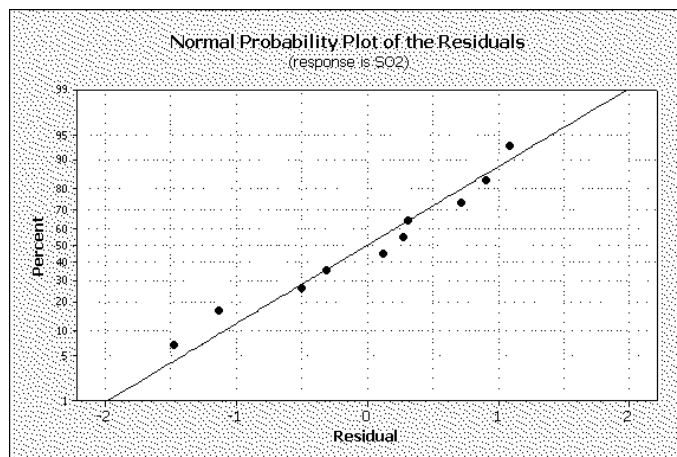
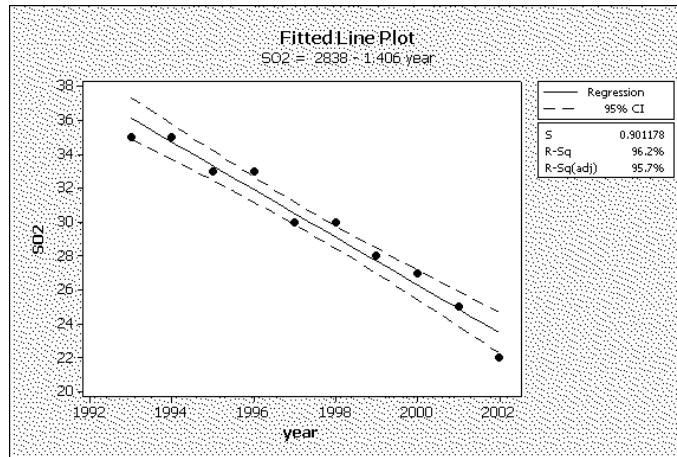
Values of Predictors for New Observations

New	Obs	year
	1	2003

- a) Sett opp en enkel lineær regresjonsmodell for disse dataene. Spesifiser antakelsene som ligger til grunn for modellen. Finn estimater for alle de tre parametrene i modellen fra Minitab-utskriften.

(Fortsettes side 4.)

- b) Diskuter kort hvor godt modellen passer til data på bakgrunn av vedlagte plott og utskrifter.
- c) I Minitabutskriften finner du 95%-konfidensintervallet (20.647,23.486) for forventet svoveldioksid-utslipp i 2003. Forklar hvordan dette er beregnet.



SLUTT