

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i: STK1000 — Innføring i anvendt statistikk
Eksamensdag: Onsdag 3. desember 2014
Tid for eksamen: 09.00 – 13.00
Oppgavesettet er på 3 sider.
Vedlegg: Ingen
Tillatte hjelpemidler: Lærebok: Moore, McCabe & Craig:
Introduction to the practice of statistics,
ordliste for bruk i STK1000, godkjent kalkulator.

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

Oppgave 1

Tallene i tabell 1 er hentet fra en undersøkelse for en ny metode for fysioterapi etter kneoperasjon. Et mål for bevegeligheten ble registrert på 7 pasienter før og etter behandling og forskjellen beregnet.

Table 1: Bevegelighet før og etter fysioterapi.

Person	Før behandling	Etter behandling	Differanse
1	31	32	-1
2	53	59	-6
3	45	46	-1
4	57	64	-7
5	50	49	1
6	43	45	-2
7	32	40	-8

- a) Hva kalles denne typen forsøksplan? Diskuter hvordan randomisering kan benyttes i slike forsøksplaner, og om det er mulig i dette tilfellet.

Minitab utskriften nedenfor inneholder noen størrelser som brukes i analysen av slike forsøk.

Descriptive Statistics: Differanse

Variable	N	Mean	SE Mean	StDev
Diff	7	-3,43	1,32	3,51

- b) Forklar hvordan dataene i tabell 1 er brukt til å beregne standardfeilen (SE Mean) og standardavviket (StDev).

(Fortsettes på side 2.)

- c) Bruk tallene i utskriften til å beregne et 95% konfidensintervall for forventningen til differansen. Presiser forutsetningene som må være til stede for at konfidenskoeffisienten skal være nøyaktig 95%.
- d) Forklar hvordan konfidensintervallet fra punkt c) kan brukes til en hypotesetest. Angi nullhypotesen og den alternative hypotesen. Hva er nivået til testen? Bruk tabellene i boka til også å finne en øvre og nedre grense for P-verdien til denne testen.
- e) Anta at en ny pasient blir behandlet etter denne metoden for fysioterapi. Angi et intervall der det er 95% sannsynlighet for at differansen i bevegelighet mellom før og etter behandling er mindre enn den øvre grensen og større enn den nedre, med andre ord det som kalles et prediksjonsintervall.

Oppgave 2

Figuren nedenfor viser sammenhengen i en tiårsperiode mellom dødsrater for sykdommer relatert til åreforkalking, dth (antall døde per 100 000 i et år), og fettforbruk, rt (kg per person per år).

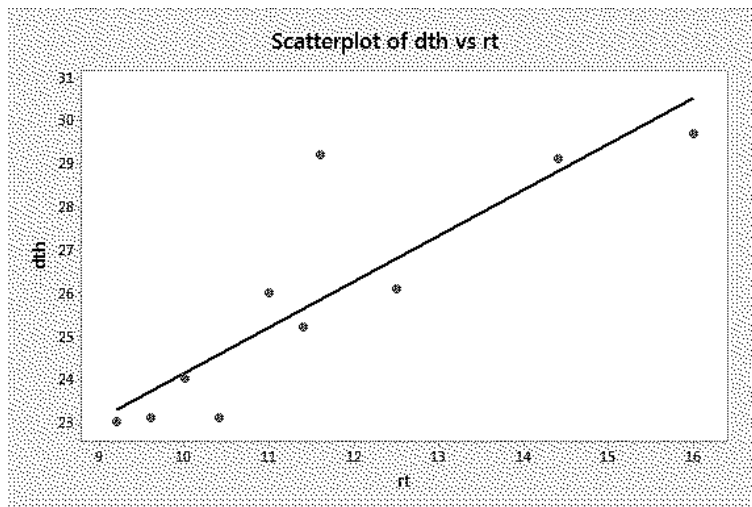


Figure 1: Plott av dødsrate mot fettforbruk.

For disse observasjonene kan man tilpasse en enkel lineær regresjon

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, 10$$

der dødsraten, dth , er respons og fettforbruk, rt , forklaringsvariabel.

Tilpasning av en slik modell med Minitab gir resultatet:

(Fortsettes på side 3.)

Coefficients

Term	Coef	SE Coef	T-Value
Constant	13,49	2,56	5,26
rt	1,065	0,218	4,89

Regression Equation

$$\text{dth} = 13,49 + 1,065 \text{ rt}$$

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1,41273	74,97%	71,84%	64,24%

- Hva er den predikerte dødsraten hvis fettforbruket er 15 kg per person per år?
- Finn et konfidensintervall med konfidenskoeffesient 95% for stigningsforholdet β_1 .
- Hva er P-verdien hvis nullhypotesen er $H_0 : \beta_1 = 0$ og alternativet $H_a : \beta_1 > 0$? Bruk tabellene i boka til å angi en så nøyaktig numerisk verdi som mulig.

Oppgave 3

- La X være en binomialfordelt tilfeldig variabel, $B(n,p)$. Forklar hvorfor forventningen μ_X er np og hvorfor variansen σ_X^2 er $np(1-p)$.
- La X være $B(5,0.3)$ og Y være $B(5,0.5)$ fordelte og anta at X og Y er uavhengige. La Z være summen, $Z=X+Y$. Finn $P(Z \leq 1)$.

SLUTT