



**Høgskolen i Telemark**  
Fakultet for allmennvitenskapelige fag

**EKSAMEN**

**6019**  
**Statistikk II**

**2.12.2013**

Tid:	9-13 (4 timer)
Målform:	Bokmål
Sidetal:	6 (inkludert denne)
Hjelpemiddel:	Alle trykte og skrevne samt kalkulator.
Merknader:	Ingen
Vedlegg:	Tabell over binomisk fordeling til oppgave 1.

Sensuren finner du på StudentWeb.



## Oppgave 1

I et forsøk er det undersøkt om en ny krem A mot kviser blant tenåringer har bedre effekt enn krem K som mange bruker. I alt 20 tenåringer, som alle har vært plaget av kviser i ansiktet, deltok i forsøket. Hver forsøksperson brukte den ene kremen på venstre halvdel av ansiktet og den andre på høyre halvdel. For hver forsøksperson ble det trukket lodd om hvilken ansiktshalvdel som fikk A. Før behandlingen startet og etter 2 måneders behandling ble forsøkspersonene undersøkt av en hudlege som ikke visste hvilken ansikts-halvdel som hadde fått hvilken krem hos den enkelte forsøksperson.

Hudlegen konkluderte med at krem A hadde størst effekt for 16 av de 20 forsøkspersonene, og krem K hadde størst effekt for 4.

- a) Bruk tegntesten til å teste om krem A har større effekt enn krem K ved behandling av kviser i ansiktet hos tenåringer. Angi nullhypotese og alternativ hypotese, gjennomfør testingen og angi konklusjonen når signifikansnivået velges lik 1 %.

Du kan bruke tabellen over binomisk fordeling med parametere  $n = 20$  og  $p = 0.50$  i vedlegget til slutt i oppgavesettet.

## Oppgave 2

En idrettsforsker har undersøkt såkalt motorisk ferdighet for 8 år gamle gutter fra 3 skoler i henholdsvis Halden, Sarpsborg og Nøtterøy. Siktemålet med undersøkelsen er å finne ut om det er forskjeller i motorisk ferdighetsnivå for barn som har vokst opp i ulike sosiale og geografiske miljøer.

Guttene som deltok i undersøkelsen, gjennomgikk en grovmotorisk test bestående av 4 deloppgaver. På grunnlag av resultatene på de 4 oppgavene ble det regnet ut en motorisk kvotient (MQ).

Vi lar  $X_{ij}$  være målt motorisk kvotient for gutt nr  $j$  fra skole nr  $i$ , der  $i = 1$  er skolen i Halden,  $i = 2$  er skolen i Sarpsborg og  $i = 3$  er skolen på Nøtterøy. Antall deltakere i undersøkelsen var  $n_1 = 18$  (Halden),  $n_2 = 23$  (Sarpsborg) og  $n_3 = 22$  (Nøtterøy).

Resultatene er gitt til slutt i oppgaven.

Vi antar at  $X_{ij}$  er normalfordelt med forventning  $\mu_i$  og varians  $\sigma^2$  ( $i = 1, 2, 3$  og  $j = 1, 2, 3, \dots, n_i$ ). Alle observasjonene er uavhengige.

- a) Angi estimatorer for de ukjente parametrene i modellen ( $\mu_1, \mu_2, \mu_3$  og  $\sigma^2$ ). Finn et 95% konfidensintervall for forventet motorisk ferdighetsnivå for 8 år gamle gutter på skolen i Halden.
- b) Tyder forsøket på at det er forskjeller i motorisk ferdighetsnivå hos 8 år gamle gutter på de 3 skolene? Formuler dette spørsmålet som en hypotesetest. Gjennomfør testingen og angi konklusjonen når vi velger signifikansnivå 5 %.
- c) Hvilke tester fra pensum kan brukes for å analysere følgende problemstillinger knyttet til motorisk ferdighetsnivå for barn (du skal ikke gjennomføre noen analyse):
- 1) Problemstillingen er som i b), men vi antar ikke at målingene er normalfordelte.
  - 2) Resultater fra 2 skoler, Halden og Sarpsborg, skal sammenlignes. Hvilken annen test enn den som ble brukt i b), kan brukes om vi antar at målingene er normalfordelte, og hvilken test kan brukes om vi ikke antar at målingene er normalfordelte?



- 3) Problemstillingen er den samme som i b); Vi skal sammenligne gutter fra 3 skoler. I motsetning til i b) antar vi nå at motorisk ferdighet for hver gutt er klassifisert og registrert som enten **god**, **middels** eller **dårlig**. Hvordan kan en nå analysere problemstillingen i b)?

### Resultater

$$X_{1j} \text{ (Halden): } \begin{array}{cccccccccccccccc} 100 & 86 & 71 & 80 & 95 & 77 & 104 & 96 & 91 & 82 & 85 & 72 & 92 & 97 & 82 \\ & 85 & 71 & 78 & & & & & & & & & & & & \\ n_1 = 18 & \bar{X}_1 = 85.78 & \sum (X_{1j} - \bar{X}_1)^2 = 1743.11 & & & & & & & & & & & & & \end{array}$$

$$X_{2j} \text{ (Sarpsborg): } \begin{array}{cccccccccccccccc} 90 & 75 & 111 & 80 & 105 & 91 & 69 & 105 & 108 & 79 & 80 & 83 & 119 & 97 \\ & 105 & 111 & 79 & 76 & 60 & 113 & 80 & 102 & 95 & & & & & \\ n_2 = 23 & \bar{X}_2 = 91.87 & \sum (X_{2j} - \bar{X}_2)^2 = 5742.61 & & & & & & & & & & & & \end{array}$$

$$X_{3j} \text{ (Nøtterøy): } \begin{array}{cccccccccccccccc} 90 & 114 & 110 & 127 & 102 & 72 & 80 & 94 & 93 & 96 & 96 & 85 & 106 & 108 \\ & 108 & 83 & 89 & 89 & 91 & 99 & 111 & 118 & & & & & & \\ n_3 = 22 & \bar{X}_3 = 98.23 & \sum (X_{3j} - \bar{X}_3)^2 = 3807.86 & & & & & & & & & & & & \end{array}$$

$$\text{Totalt gjennomsnitt: } \bar{X} = 92.35$$

### Oppgave 3

En miljøvernorganisasjon har studert hvordan bensinforbruket for nye biler avhenger av motorstyrken. For 15 biler er det målt kjørelengde  $Y$  i kilometer (km) på 10 liter bensin. Motorstyrken  $X$  er målt i hestekrefter (hk).

Observasjonene går fram av Minitab-utskriften til slutt i oppgaven (neste side). Studér denne før du svarer på spørsmålene nedenfor.

Vi skal bruke en regresjonsmodell:

$$Y = \beta_0 + \beta_1 X + U$$

- a) Skriv opp den estimerte sammenhengen mellom kjørelengde og motorstyrke. Forklar kort hva den estimerte verdien for  $\beta_1$  uttrykker og angi et 95% konfidensintervall for  $\beta_1$ .

Forklar kort hva  $R^2$  uttrykker (R-Sq i Minitab).

Gjør kort greie for antakelsene i regresjonsmodellen. Til slutt i Minitab-utskriften er det to plott, først residualplott (residual mot observasjon nummer) og deretter normalplott for residualene. Tyder plottene kan på at disse antakelsene er rimelige? Begrunn svaret.

- b) For en valgt verdi av motorstyrken  $X$  (hk) kan en angi både et konfidensintervall og et prediksjonsintervall for kjørelengden (km) på 10 liter bensin. Forklar forskjellen på de to intervallene. Finn et 95 % prediksjonsintervall for en ny bil med 90 hestekrefter.



## Minitab-utskrift – oppgave 3

### Data Display

Row	X	Y
1	85	153
2	80	210
3	66	220
4	63	206
5	90	166
6	95	127
7	139	125
8	92	143
9	88	167
10	58	230
11	84	160
12	100	141
13	65	235
14	76	198
15	71	180

### Descriptive Statistics: X

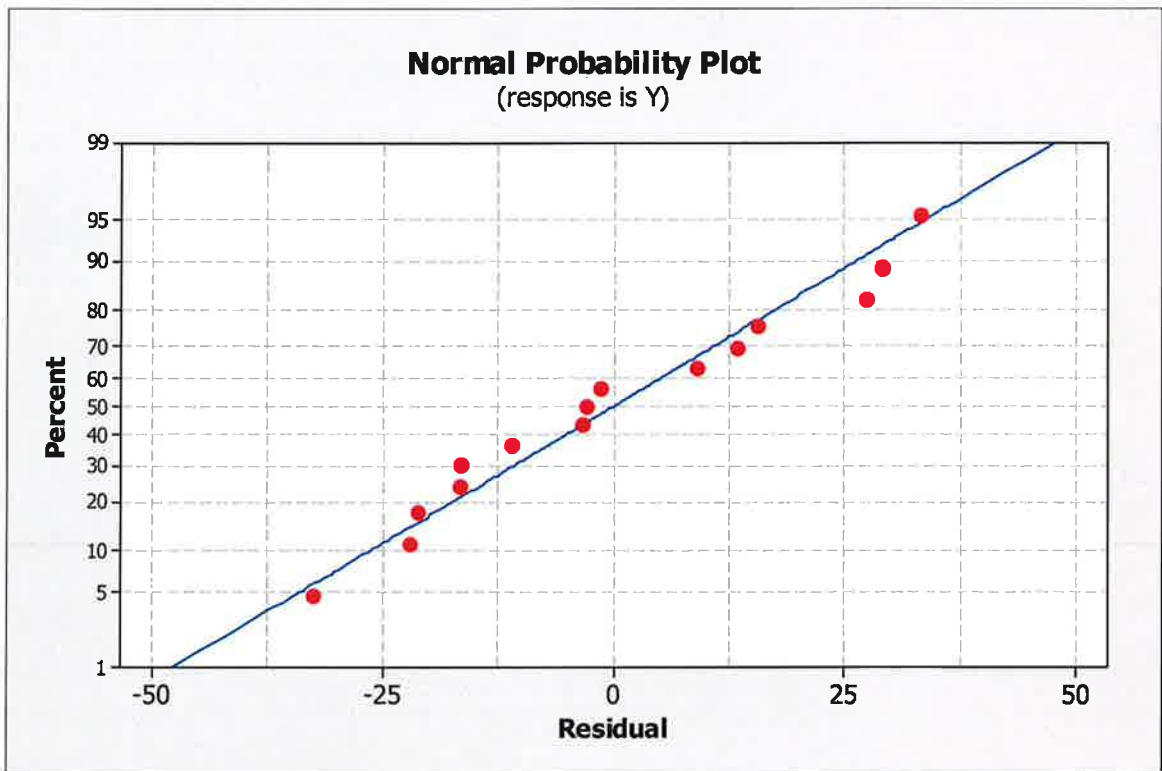
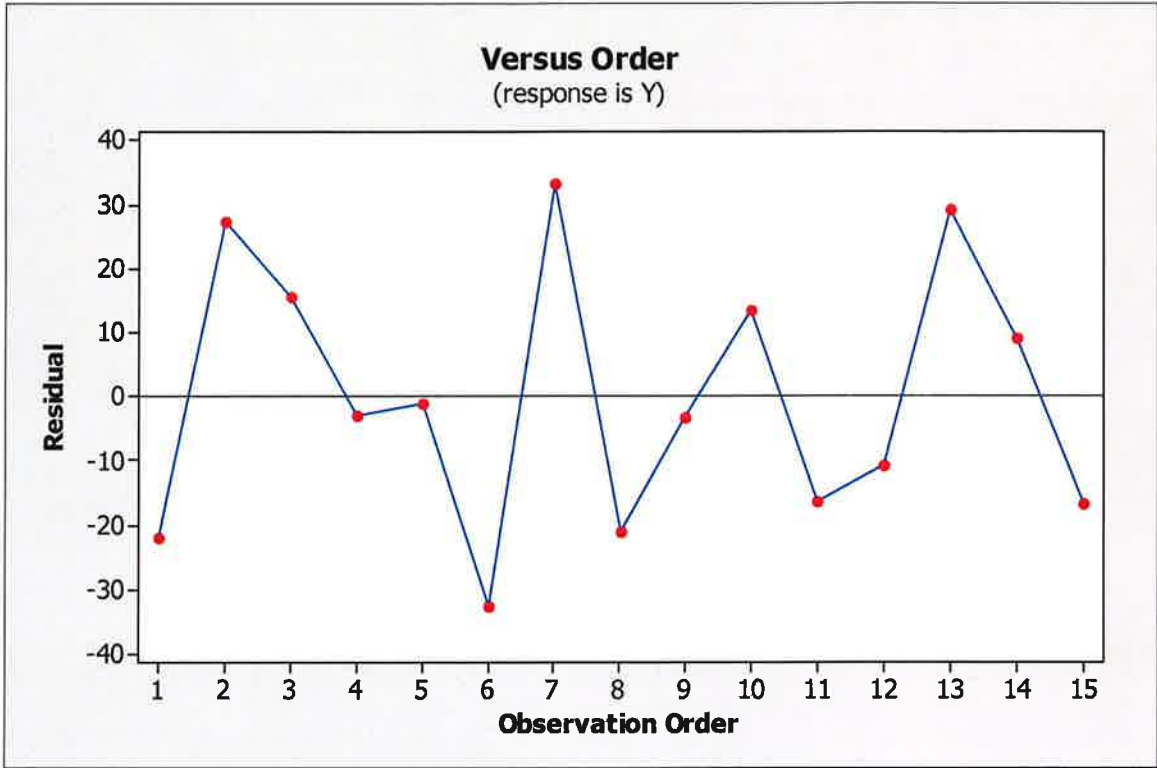
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
X	15	0	83,47	5,15	19,94	58,00	66,00	84,00	92,00	139,00

### Regression Analysis: Y versus X

The regression equation is  
 $Y = 306 - 1,54 X$

Predictor	Coef	SE Coef	T	P
Constant	306,17	24,45	12,52	0,000
X	-1,5428	0,2854	-5,41	0,000

S = 21,2927    R-Sq = 69,2%    R-Sq(adj) = 66,8%





## Vedlegg til oppgave 1

### Cumulative Distribution Function

Binomial with  $n = 20$  and  $p = 0,5$

x	P( X <= x )
0	0,00000
1	0,00002
2	0,00020
3	0,00129
4	0,00591
5	0,02069
6	0,05766
7	0,13159
8	0,25172
9	0,41190
10	0,58810
11	0,74828
12	0,86841
13	0,94234
14	0,97931
15	0,99409
16	0,99871
17	0,99980
18	0,99998
19	1,00000
20	1,00000