NTNU
Norges teknisk-naturvitenskapelige universitet
Institutt for samfunnsøkonomi

EKSAMENSOPPGAVE I SØK1004

STATISTIKK FOR ØKONOMER

STATISTICS FOR ECONOMISTS

Faglig kontakt under eksamen: Hildegunn E. Stokke

Tlf.: 9 16 65

Eksamensdato: Fredag 8. juni 2012

Eksamenssted: Dragvoll

Eksamenstid: 4 timer

Studiepoeng: 7,5

Tillatte hjelpemidler: Alle trykte og håndskrevne hjelpemidler, samt godkjent kalkulator Citizen SR-270x el. HP 30S.

Sensur: 29. juni 2012

Antall sider bokmål: 2
Antall sider nynorsk: 2
Antall sider engelsk: 2

Merk! Det blir sendt automatisk varsel om sensur på e-post. Du kan se hva som er registrert ved å gå inn på Studentweb. Evt andre telefoner om sensur må rettes til instituttet. Eksamenskontoret vil ikke kunne svare på slike telefoner.

Eksamensoppgaven består av fire oppgaver, og alle skal besvares. Vekting er gitt i parentes.

Oppgave 1 (25%)

Da Anne var på besøk i Roma, fikk hun raskt problemer med språket. Anne snakker engelsk, men ikke italiensk, og kun 1 av 5 italienere behersker engelsk. Likevel, på tur i sentrum av byen er det noe lettere å treffe noen som snakker engelsk, da 1 av 4 personer som befinner seg i Roma sentrum er utenlandske turister. Og blant de utenlandske turistene er det hele 1 av 2 som behersker engelsk.

Merk at vi for enkelhets skyld antar at alle som befinner seg i Roma sentrum enten er italienere eller utenlandske turister.

For en vilkårlig person trukket fra Roma sentrum, definer hendelsene, E = 'personen snakker engelsk', I = 'personen er italiener' og T = 'personen er utenlandsk turist'.

a) Forklar ved Venn-diagram eller på annen måte at $E = (E \cap I) \cup (E \cap T)$. Er hendelsene $(E \cap I)$ og $(E \cap T)$ disjunkte? Forklar.
b) Finn følgende sannsynligheter: $P(E \mid I)$ og $P(E \cap I)$.
c) Hva er sannsynligheten for at en tilfeldig person som Anne treffer snakker engelsk?
d) Hva er sannsynligheten for at vedkommende er italiener, gitt at hun/han snakker engelsk?
e) Hva er sannsynligheten for at vedkommende er italiener, gitt at hun/han ikke snakker engelsk?
f) I en gate treffer Anne på 4 personer. La den stokastiske variabelen X være antall utenlandske turister blant disse fire.
   i) Drøft kort i hvilken grad vi kan anta at X er binomisk fordelt.
   ii) Anta nå at X er binomisk fordelt. Beregn sannsynligheten for at det er to eller flere utenlandske turister blant de fire som Anne møter i denne gaten.

Oppgave 2 (15%)

a) Anta at X er normalfordelt med middelverdi μ = 16 og varians $\sigma^2 = 25$. Finn:

i) $P(X > 20)$

ii) $P(20 < X < 25)$

iii) $P(X < 10)$

iv) $P(12 < X < 24)$

b) Anta at den stokastiske variabelen Z er standard normalfordelt. Finn verdien k slik at $P(-0.62 < Z < k) = 0.43$.

Oppgave 3 (20%)

Et tilfeldig utvalg bestående av 13 jusstudenter viser seg å ha en gjennomsnittlig studietid per uke i løpet av semesteret på 32 timer, med standardavvik lik 7 timer. For et tilfeldig utvalg bestående av 16 økonomistudenter er gjennomsnittlig studietid per uke i løpet av semesteret 38 timer med standardavvik lik 5.5 timer. Studietid antas normalfordelt i begge populasjoner.

a) Test om variansen i studietid er forskjellig i de to populasjonene. Benytt 5% signifikansnivå.
b) Test om gjennomsnittlig studietid er høyere for økonomistudenter enn for jusstudenter. Bruk både 5% og 1% signifikansnivå.

Oppgave 4 (40%)

De følgende dataene gir antall timer studietid per uke (Y) og antall timer per uke i lønnet arbeid (X) for et tilfeldig utvalg av 10 studenter:

Y	40	35	34	36	40	32	44	46	48	45
X	7.5	15	7.5	4	0	10	0	4	2	0

$$\overline{X} = 5 \qquad \overline{Y} = 40 \qquad \sum_{i=1}^{10}(X_i - \overline{X})^2 = 223.5 \qquad \sum_{i=1}^{10}(Y_i - \overline{Y})^2 = 282 \qquad \sum_{i=1}^{10}(X_i - \overline{X})(Y_i - \overline{Y}) = -176$$

a) Anta at $Y_i = \alpha + \beta X_i + \varepsilon_i$, hvor ε er restledd, og bruk minste kvadraters metode (ordinary least squares, OLS) til å estimere koeffisientene α og β. Gi en tolkning av estimatene.
b) ”Antall timer studietid er uavhengig av antall timer i lønnet arbeid”. Formuler denne påstanden som en hypotesetest og test om påstanden kan forkastes. Bruk 5% signifikansnivå.
c) ”Dersom en bruker en time mer per uke i lønnet arbeid, reduseres tid brukt på studier med en time”. Formuler denne påstanden som en hypotesetest og test om påstanden kan forkastes. Bruk 10% signifikansnivå.
d) Beregn både 95% og 99% konfidensintervall for β. Kommenter resultatene.
e) Beregn hvor mange studenter du minst må ha med i utvalget for å få et 99% konfidensintervall for β på formen $b \pm 0.78$ eller mindre, der *b* er estimatet beregnet i oppgave a). Anta uendret standardavvik for *b*.
f) Hva menes med modellens forklaringskraft (også kalt R^2)? Beregn denne.

Merk! Det blir sendt automatisk varsel om sensur på e-post. Du kan se hva som er registrert ved å gå inn på Studentweb. Evt andre telefoner om sensur må rettes til instituttet. Eksamenskontoret vil ikke kunne svare på slike telefoner.

Eksamensoppgåva inneheld fire oppgåver, og alle skal svarast på. Vekting er gitt i parentes.

Oppgåve 1 (25%)

Då Anne var på besøk i Roma, fekk ho raskt problem med språket. Anne snakkar engelsk, men ikkje italiensk, og berre 1 av 5 italienarar beherskar engelsk. Likevel, på tur i sentrum av byen er det noko lettare å treffe nokon som snakkar engelsk, då 1 av 4 personar som oppheld seg i Roma sentrum er utanlandske turistar. Og blant dei utanlandske turistane er det heile 1 av 2 som beherskar engelsk.

Merk at vi for enkelheits skuld antar at alle som oppheld seg i Roma sentrum anten er italienarar eller utanlandske turistar.

For ein vilkårleg person trekt frå Roma sentrum, definer hendingane, E = 'personen snakkar engelsk', I = 'personen er italienar' og T = 'personen er utanlandsk turist'.

a) Forklar ved Venn-diagram eller på annan måte at $E = (E \cap I) \cup (E \cap T)$. Er hendingane $(E \cap I)$ og $(E \cap T)$ disjunkte? Forklar.
b) Finn fylgjande sannsyn: $P(E \mid I)$ og $P(E \cap I)$.
c) Kva er sannsynet for at ein tilfeldig person som Anne treffer snakkar engelsk?
d) Kva er sannsynet for at vedkommande er italienar, gitt at ho/han snakkar engelsk?
e) Kva er sannsynet for at vedkommande er italienar, gitt at ho/han ikkje snakkar engelsk?
f) I ei gate treff Anne på 4 personar. La den stokastiske variabelen X vere talet på utanlandske turistar blant disse fire.
   i) Drøft kort i kva for grad vi kan anta at X er binomisk fordelt.
   ii) Anta no at X er binomisk fordelt. Berekn sannsynet for at det er to eller fleire utanlandske turistar blant dei fire som Anne møter i denne gata.

Oppgåve 2 (15%)

a) Anta at X er normalfordelt med middelverdi μ = 16 og varians $\sigma^2 = 25$. Finn:

i) $P(X > 20)$
ii) $P(20 < X < 25)$
iii) $P(X < 10)$
iv) $P(12 < X < 24)$

b) Anta at den stokastiske variabelen Z er standard normalfordelt. Finn verdien k slik at $P(-0.62 < Z < k) = 0.43$.

Merk! Det blir sendt automatisk varsel om sensur på e-post. Du kan se hva som er registrert ved å gå inn på Studentweb. Evt andre telefoner om sensur må rettes til instituttet. Eksamenskontoret vil ikke kunne svare på slike telefoner.

Oppgåve 3 (20%)

Eit tilfeldig utval beståande av 13 jusstudentar viser seg å ha ein gjennomsnittleg studietid per veke i løpet av semesteret på 32 timer, med standardavvik lik 7 timer. For et tilfeldig utval beståande av 16 økonomistudentar er gjennomsnittlig studietid per veke i løpet av semesteret 38 timer med standardavvik lik 5.5 timer. Vi antek at studietid er normalfordelt i begge populasjonane.

a) Test om variansen i studietid er forskjelleg i dei to populasjonane. Bruk 5% signifikansnivå.
b) Test om gjennomsnittleg studietid er høgare for økonomistudentar enn for jusstudentar. Bruk både 5% og 1% signifikansnivå.

Oppgåve 4 (40%)

Dei fylgjande data gir talet på timer studietid per veke (Y) og talet på timer per veke i løna arbeid (X) for eit tilfeldig utval av 10 studentar:

Y	40	35	34	36	40	32	44	46	48	45
X	7.5	15	7.5	4	0	10	0	4	2	0

$$\overline{X} = 5 \qquad \overline{Y} = 40 \qquad \sum_{i=1}^{10}(X_i - \overline{X})^2 = 223.5 \qquad \sum_{i=1}^{10}(Y_i - \overline{Y})^2 = 282 \qquad \sum_{i=1}^{10}(X_i - \overline{X})(Y_i - \overline{Y}) = -176$$

a) Anta at $Y_i = \alpha + \beta X_i + \varepsilon_i$, der ε er restledd, og bruk minste kvadrats metode (ordinary least squares, OLS) til å estimere koeffisientane α og β. Gi ei tolking av estimata.
b) ”Talet på timer studietid er uavhengig av talet på timer i løna arbeid”. Formuler denne påstanden som ein hypotesetest og test om påstanden kan forkastast. Bruk 5% signifikansnivå.
c) ”Dersom ein bruker ein time meir per veke i løna arbeid, reduserast tid brukt på studiar med ein time”. Formuler denne påstanden som ein hypotesetest og test om påstanden kan forkastast. Bruk 10% signifikansnivå.
d) Berekn både 95% og 99% konfidensintervall for β. Kommenter resultata.
e) Berekn kor mange studentar du minst må ha med i utvalet for å få eit 99% konfidensintervall for β på forma $b \pm 0.78$ eller mindre, der *b* er estimatet berekna i oppgåve a). Anta uendra standardavvik for *b*.
f) Kva meines med modellen si forklaringskraft (også kalla R^2)? Berekn denne.

Merk! Det blir sendt automatisk varsel om sensur på e-post. Du kan se hva som er registrert ved å gå inn på Studentweb. Evt andre telefoner om sensur må rettes til instituttet. Eksamenskontoret vil ikke kunne svare på slike telefoner.

The exam consists of four questions, and all of them should be answered. Weights are given in parenthesis.

Question 1 (25%)

When Anne was visiting Rome, she quickly got problems with the language. Anne speaks English, but not Italian, and only 1 of 5 Italians knows English. Still, in the center of the city it is somewhat easier to meet someone who speaks English, since 1 of 4 persons in the center of Rome are foreign tourists. And among the foreign tourists as much as 1 of 2 knows English.

Note that for simplicity we assume that everyone who is in the center of Rome is either Italian or foreign tourists.

For a random person drawn from the center of Rome, define the events, E = 'the person speaks English', I = 'the person is Italian' and T = 'the person is a foreign tourist'.

a) Explain by Venn-diagram or otherwise that $E = (E \cap I) \cup (E \cap T)$. Are the events $(E \cap I)$ and $(E \cap T)$ mutually exclusive? Explain.
b) Find the following probabilities: $P(E \mid I)$ and $P(E \cap I)$.
c) What is the probability that a random person that Anne meets speaks English?
d) What is the probability that he/she is Italian, given that he/she speaks English?
e) What is the probability that he/she is Italian, given that he/she does not speak English?
f) In a street Anne meets 4 persons. Let the random variable X be the number of foreign tourists among these four.
   i) Discuss shortly to what extent we can assume that X is binomially distributed.
   ii) Assume now that X is binomially distributed. Calculate the probability that there are two or more foreign tourists among the four that Anne meets in this street.

Question 2 (15%)

a) Assume that X is normally distributed with mean μ = 16 and variance $\sigma^2 = 25$. Find:

   i) $P(X > 20)$
   ii) $P(20 < X < 25)$
   iii) $P(X < 10)$
   iv) $P(12 < X < 24)$

b) Assume that the random variable Z is standard normally distributed. Find the value k so that $P(-0.62 < Z < k) = 0.43$.

Merk! Det blir sendt automatisk varsel om sensur på e-post. Du kan se hva som er registrert ved å gå inn på Studentweb. Evt andre telefoner om sensur må rettes til instituttet. Eksamenskontoret vil ikke kunne svare på slike telefoner.

Question 3 (20%)

A random sample of 13 law students turns out to have an average study time per week during the semester of 32 hours, with a standard deviation of 7 hours. For a random sample of 16 economics students the average study time per week during the semester is 38 hours with a standard deviation of 5.5 hours. The study time is assumed normally distributed in both populations.

a) Test whether the variance in study time is different in the two populations. Use 5% significance level.
b) Test whether the mean study time is higher among economics students than among law students. Use both 5% and 1% significance level.

Question 4 (40%)

The following data give the number of hours study time per week (Y) and the number of hours per week in paid employment (X) for a random sample of 10 students:

Y	40	35	34	36	40	32	44	46	48	45
X	7.5	15	7.5	4	0	10	0	4	2	0

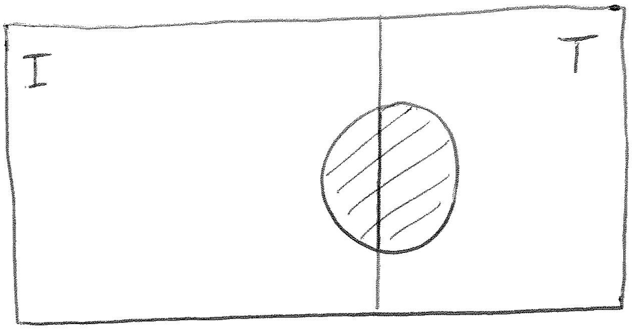
$$\bar{X} = 5 \quad \bar{Y} = 40 \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 223.5 \quad \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 282 \quad \sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = -176$$

a) Assume that $Y_i = \alpha + \beta X_i + \varepsilon_i$, where ε is a disturbance, and use ordinary least squares (OLS) to estimate the coefficients α and β. Give an interpretation of the estimates.
b) ”The number of hours study time is independent of the number of hours paid employment”. Formulate this statement as a hypothesis test and test whether the statement can be rejected. Use a significance level of 5%.
c) ”If you spend one hour more per week in paid employment, the time spent on studies is reduced by one hour”. Formulate this statement as a hypothesis test and test whether the statement can be rejected. Use a significance level of 10%.
d) Calculate both the 95% and 99% confidence interval for β. Comment on the results.
e) Calculate how many students you at least need in the sample to get a 99% confidence interval for β in the form $b \pm 0.78$ or less, where *b* is the estimate calculated in sub-question a). Assume the same standard deviation for *b* as before.
f) What is meant by the model’s coefficient of determination (also called R^2)? Calculate this coefficient.

Fasit eksamen SØK1004 våren 2012

Oppg. 1

a)



$= E$

$= E \cap I$

$= E \cap T$

$\Rightarrow E = (E \cap I) \cup (E \cap T)$

$E \cap I$ og $E \cap T$ er disjunkte hendelser.

En person kan ikke være både italiener og utenlandsk turist.

b) Opplysninger gitt i oppgaveteksten:

$$P(T) = 0{,}25$$

$$P(E|T) = 0{,}5$$

$$\rightarrow \quad \underline{P(E|I) = 0{,}2}$$

$$P(I) = 0{,}75$$

$$\Rightarrow P(E \cap I) = P(E|I) \cdot P(I)$$

$$= 0{,}2 \cdot 0{,}75$$

$$= \underline{0{,}15}$$

c) $P(E) = P(E \cap I) + P(E \cap T)$ siden de er disjunkte

$= 0{,}15 + P(E|T) \cdot P(T)$

$= 0{,}15 + 0{,}5 \cdot 0{,}25$

$= 0{,}15 + 0{,}125$

$= \underline{0{,}275}$

d) $P(I|E) = \dfrac{P(E \cap I)}{P(E)} = \dfrac{0{,}15}{0{,}275} = \underline{0{,}545}$

e) $P(I|\bar{E}) = \dfrac{P(\bar{E} \cap I)}{P(\bar{E})} = \dfrac{P(\bar{E}|I) \cdot P(I)}{0{,}725} = \dfrac{0{,}8 \cdot 0{,}75}{0{,}725}$

$= \underline{0{,}828}$

f) i) - To mulige utfall, suksess vs. fiasko
- Individene er uavhengige av hverandre
- Konstant sannsynlighet for suksess

ii) X = antall utenlandske turister

$$p(X) = \frac{n!}{(n-x)!\, x!} \pi^x (1-\pi)^{n-x}$$

Vi har: $n = 4$, $\pi = 0{,}25$

$$P(2) = \frac{4!}{2!\,2!} \, 0{,}25^2 \cdot 0{,}75^2 = 0{,}2109$$

$$P(3) = \frac{4!}{1!\,3!} \, 0{,}25^3 \cdot 0{,}75 = 0{,}0469$$

$$P(4) = \frac{4!}{0!\,4!} \, 0{,}25^4 \cdot 0{,}75^0 = 0{,}0039$$

$$\Rightarrow P(X \geq 2) = 0{,}2109 + 0{,}0469 + 0{,}0039 = \underline{0{,}2617}$$

<u>Oppg. 2</u>

a) $X \sim N(16, 25) \Rightarrow Z = \dfrac{X - 16}{5} \sim N(0,1)$

i) $P(X > 20) = P(Z > \frac{4}{5}) = 0{,}5 - P(0 < Z < \frac{4}{5})$

$$= 0{,}5 - 0{,}2881 = \underline{0{,}2119}$$

ii) $P(20 < X < 25) = \cancel{P(X > 20) - P(X > 25)}$

$\cancel{= 0{,}2119 - (0{,}5 - P(0 < Z < 25}$

$$= P\left(\frac{4}{5} < Z < \frac{9}{5}\right) = P(0{,}8 < Z < 1{,}8)$$

$= P(Z > 0,8) - P(Z > 1,8)$

$= 0,2119 - (0,5 - P(0 < Z < 1,8))$

$= 0,2119 - 0,5 + 0,4641$

$= \underline{0,176}$

iii) $P(X < 10) = P(Z < -1,2)$

$= P(Z > 1,2)$

$= 0,5 - P(0 < Z < 1,2)$

$= 0,5 - 0,3849$

$= \underline{0,1151}$

iv) $P(12 < X < 24) = P(-0,8 < Z < 1,6)$

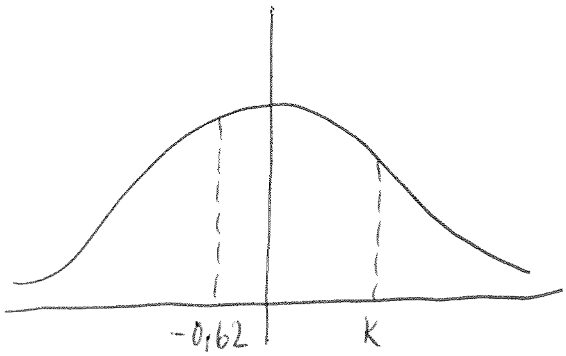
$= P(0 < Z < 0,8) + P(0 < Z < 1,6)$

$= 0,2881 + 0,4452$

$= \underline{0,7333}$

b) $Z \sim N(0,1)$

$P(-0{,}62 < Z < k) = 0{,}43$



$P(-0{,}62 < Z < 0) = P(0 < Z < 0{,}62) = 0{,}2324$

$\Rightarrow P(0 < Z < k) = 0{,}43 - 0{,}2324 = 0{,}1976$

Ser at: $P(0 < Z < 0{,}51) = 0{,}195$

$P(0 < Z < 0{,}52) = 0{,}1985$

$\Rightarrow \underline{k \approx 0{,}52}$.

Oppg. 3

a) $H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$

$$TS = \frac{s_1^2}{s_2^2} = \frac{7^2}{5{,}5^2} = 1{,}62 \quad \sim F(12,15)$$

$$F_{0{,}025} = 2{,}9633$$

$$F_{0{,}975} = \frac{1}{F_{0{,}025}} = \frac{1}{2{,}9633} = 0{,}3375$$

$\Rightarrow H_0$ forkastes ikke.

b) $H_0: \mu_1 = \mu_2$

$H_A: \mu_1 < \mu_2$

$$TS = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \sim t\text{-fordelt med 27 frihetsgrader}$$

Små utvalg $\Rightarrow$ t-fordeling.

Antar lik varians basert på resultatet i a).

Felles varianseestimator:

$$s^2 = \frac{12 \cdot 49 + 15 \cdot 30{,}25}{27} = 38{,}583$$

$$\Rightarrow s = 6{,}2115$$

$$\Rightarrow TS = \frac{32-38}{6{,}2115\sqrt{\frac{1}{13}+\frac{1}{16}}} = -2{,}587$$

Kritisk verdi:

1% sign.nivå: $-t_{0,01} = -2{,}473$

5% sign.nivå: $-t_{0,05} = -1{,}703$

H_0 forkastes til både 1% og 5% nivå.

Oppg. 4

a) $b = \frac{-176}{223,5} = -0,7875$

$a = 43,9375$

b) $H_0: \beta = 0$

$H_A: \beta \neq 0$

$TS = \frac{b}{s_b}$

der $s_b = \sqrt{\frac{s^2}{\sum(x-\bar{x})^2}}$ $\quad s^2 = \frac{1}{n-2}\left[\sum(y-\bar{y})^2 - b\sum(x-\bar{x})(y-\bar{y})\right]$

$\Rightarrow s^2 = \frac{282 + 0,7875 \cdot (-176)}{8} = 17,925$

$\Rightarrow s_b = \sqrt{\frac{17,925}{223,5}} = 0,2832$

$\Rightarrow TS = \frac{-0,7875}{0,2832} = -2,78$

$t_{0,025} = 2,306$ med 8 frihetsgrader $\Rightarrow H_0$ forkastes.

c) $H_0: \beta = -1$

$H_A: \beta \neq -1$

$$TS = \frac{b-\beta}{s_b} = \frac{-0{,}7875+1}{0{,}2832} = 0{,}75$$

$t_{0{,}05} = 1{,}86$

H_0 beholdes selv til 10 % sign. nivå.

d) 95 % : $-0{,}7875 \pm 2{,}306 \cdot 0{,}2832$

$= -0{,}7875 \pm 0{,}6531$

$\Rightarrow \langle -1{,}4406 \,,\, -0{,}1344 \rangle$

99 % : $-0{,}7875 \pm 3{,}355 \cdot 0{,}2832$

$= -0{,}7875 \pm 0{,}9501$

$\Rightarrow \langle -1{,}7376 \,,\, 0{,}1626 \rangle$

Kommentarer: – Forklare hvorfor 99 % intervallet er større
– Relatere til testen i oppg. b, H_0 hadde ikke blitt forkasta til 1 % sign. nivå siden 99 % konf. int. inneholder 0.

e) Antar at s_b er uendra.

99% konf. intervall: $b \pm t_{0,005} \cdot s_b$

$\Rightarrow t_{0,005} \cdot 0,2832 \leq 0,78$

$\Rightarrow t_{0,005} \leq 2,754$

$\Rightarrow$ Må ha 30 frihetsgrader, dvs 32 observasjoner.

f) $R^2 = \frac{SSE}{SST} = \frac{\sum(\hat{Y}-\bar{Y})^2}{\sum(Y-\bar{Y})^2}$ → må beregnes! → gitt i oppg. tekst

$\Rightarrow R^2 = \frac{138,6}{282} = \underline{0,49}$