



ECONnect

NTNU

Faktor

- en eksamensavis utgitt av ECONnect



Eksamensbesvarelse:

SØK1004 – Statistikk for økonomer

Eksamen:
Antall sider:

Vår 2011
4



Om ECONnect:

ECONnect er en frivillig studentorganisasjon for studentene på samfunnsøkonomi- og finansøkonomistudiet ved NTNU. Vi arbeider for økt faglig kompetanse blant våre studenter samt tettere kontakt med næringslivet. Det gjør vi ved å arrangere fagdager, gjesteforelesninger, bedriftspresentasjoner m.m. I dag går det ca. 200 studenter på bachelornivå (1.-3. klasse) og ca. 70 studenter på masternivå (4.-5. klasse). Studentene på masternivå er fordelt på de to linjene samfunnsøkonomi (ca. 50 stk) og finansiell økonomi (ca. 20 stk). Mer om ECONnect og aktuelle arrangementer på www.econnect-ntnu.no.

ECONnect består av følgende personer ved utgivelsestidspunkt:

Ole Christian Grytten(Leder)	ole@econnect-ntnu.no
Daniel Johansson(Bedriftsansvarlig)	daniel@econnect-ntnu.no
Johan Berg Fossen(Fagdagsansvarlig)	johan@econnect-ntnu.no
Mariell Toven(Økonomi/Kandidattreffet)	mariell@econnect-ntnu.no
Georg Næsheim	georg@econnect-ntnu.no
Ellen Normann	ellen@econnect-ntnu.no
Ragnhild Grøv	ragnhild@econnect-ntnu.no
Martine Ødegård (Faktoransvarlig)	martine@econnect-ntnu.no
Inga Friis	inga@econnect-ntnu.no
Caroline Lesiewicz	caroline@econnect-ntnu.no

Post- og besøksadresse:

ECONnect, NTNU Dragvoll
 Institutt for samfunnsøkonomi
 Bygg 7, Nivå 5
 7491 Trondheim

Organisasjonsnummer:

NO 994 625 314

Hjemmeside:

www.econnect-ntnu.no

Merk: Eksamensbesvarelsene har i varierende grad feil og mangler, både oppsett og innhold. De vil også kun vise en av flere mulige fremgangsmåter. ECONnect står ikke ansvarlig for selve faginnholdet.



Norges teknisk-naturvitenskapelige universitet
Institutt for samfunnsøkonomi

EKSAMENSOPPGAVE I SØK1004

STATISTIKK FOR ØKONOMER

Faglig kontakt under eksamen: Torberg Falch

Tlf.: 9 67 57

Eksamensdato: Onsdag 8. juni 2011

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: 30. juni 2011

Antall sider bokmål: 3

Antall sider nynorsk: 3

Antall sider engelsk: 3

Bokmål

Oppgave 1 (teller 20 %)

En eksamen blir gitt som flervalgsprøve ("multiple choice test"). Eksamen består av tre spørsmål. For hvert spørsmål er det oppgitt fire svaralternativer hvorav kun ett er riktig. I oppgave a) og b) skal vi anta at spørsmålene omhandler helt ulike tema slik at selv om man vet svaret på ett spørsmål bidrar det ikke til å gjøre de andre spørsmålene lettere.

- a) En student som ikke vet svaret på noen av delspørsmålene bestemmer seg for å gjette helt vilkårlig. (i) Hva er sannsynligheten for at studenten gjetter riktig på spørsmål 1? (ii) Hva er sannsynligheten for at studenten får riktig svar på alle de tre spørsmålene? (iii) Hva er sannsynligheten for at en student gjetter riktig på ett av spørsmålene og galt på de to andre spørsmålene? Begrunn kort måten du regner ut svarene på.

For å bestå prøven må en student ha svart riktig på minst to av de tre spørsmålene.

- b) Hva er sannsynligheten for å bestå prøven for en student som gjetter helt vilkårlig på alle spørsmålene?

Året etter har flervalgsprøven samme form, men nå er det slik at dersom studenten kjenner svaret på spørsmål 1 er det kun to alternativer som kan være riktig for spørsmål 2. Spørsmål 3 er igjen uavhengig av de andre spørsmålene.

- c) En student kjenner svaret på spørsmål 1 men ikke på spørsmål 2 og 3. (i) Hva er sannsynligheten for at studenten svarer rett på spørsmål 2? (ii) Hva er sannsynligheten for at studenten svarer rett på alle spørsmålene? (iii) Hva er sannsynligheten for at studenten består prøven? Begrunn kort måten du regner ut svarene på.

Oppgave 2 (teller 15 %)

Anta at variablene A og B er uavhengige og normalfordelte der $A \sim N(5, 3)$ og $B \sim N(4, 3)$

- a) Sett $C = A - B$. Forklar hvorfor $C \sim N(1, \sqrt{18})$
b) Finn sannsynlighetene $Pr(A < B)$, $Pr(A = B)$ og $Pr(A > B)$.

Oppgave 3 (teller 20 %)

Vi er interessert i omfang og betydning av lekser i matematikk i barneskolen. En lærer vi kjenner er av den oppfatning at ca. 50 % av elevene får matematikklektur til minst halvpartene av matematikktimene, men at det ikke gir noen bedring i matematikkferdighetene til elevene.

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.

SØK1004 – Statistikk for økonomer

TIMSS er en internasjonal undersøkelse av elevers ferdigheter. Undersøkelsen er nøye med å bruke representative utvalg. I en undersøkelse av realfagsferdighetene til norske 4. klassinger i 2007 ble lærerne spurt om omfanget på lekser. Nedenfor er resultater for matematikk

	Antall elever	Gjennomsnittresultat på matematikktesten	Standardavvik til testen
Lekser til minst halvparten av timene	2654	479,3	71,8
Lekser til mindre enn halvparten av timene	679	468,3	71,7

Testresultatet er normalfordelt.

- Hvor stor andel av elevene får lekse til minst halvparten av timene?
- Sett opp en hypotesetest og test om det er 50 % sannsynlighet for at en tilfeldig elev får lekser til minst halvparten av timene. Bruk 5 % signifikansnivå.
- Test om matematikkferdighetene er forskjellig i gruppen som får mye lekser enn i gruppen som får lite lekser. Bruk 5 % signifikansnivå.
- Basert på resultatet i c), bør vi anbefale læreren vi kjenner å gi lekser til minst halvparten av timene?

Oppgave 4 (teller 30 %)

- Det er en populasjonsmodell $Y_i = \alpha + \beta X_i + \varepsilon_i$. Gi en tolkning av parameteren β i denne modellen når Y benevner konsum og X benevner inntekt.
- Vi ønsker empirisk kunnskap om størrelsen på parameteren β . Beskriv kort hvordan man kan gå fram for å tilegne seg slik kunnskap.

Vi får tilgang til et tilfeldig utvalg på 10 arbeidstakere. Dataene er månedstall i 1000 kroner.

Individ	Inntekt (X)	Konsum (Y)
1	27,0	15,4
2	25,9	22,1
3	26,8	23,9
4	35,7	40,2
5	37,6	25,0
6	15,1	14,8
7	30,2	26,2
8	29,5	20,1
9	25,0	27,7
10	40,1	30,7

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.

$$\bar{X} = 29,29, \quad \bar{Y} = 24,61, \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 470,6, \quad \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 500,6,$$

$$\sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = 326,8$$

- Bruk minste kvadraters metode til å estimere verdien på α og β .
- Bruk modellen til å predikere konsumet til en arbeidstaker med månedsinntekt på 30 tusen kroner.
- “Konsumet er uavhengig av inntekt. Det ser man av at for eksempel individ 9 konsumerer mer enn inntekten”. Formuler denne påstanden som en hypotesetest og test om påstanden kan forkastes. Bruk 5 % signifikansnivå.
- “Det er kun inntekten som bestemmer konsumet. Tjener man en krone mer så konsumerer man også en krone mer”. Formuler denne påstanden som en hypotesetest og test om påstanden kan forkastes.
- Det er foreslått å øke stipendet til studenter med 1 000 kroner i måneden. Kan du bruke modellen til å predikere hvilken effekt dette vil ha på studenters konsum?
- Hva menes med modellens forklaringskraft (også kalt R^2)? Beregn denne.

Oppgave 5 (teller 15 %)

De to variablene A og B har følgende simultanfordeling.

	B	0	2	3
A				
1		0,21	0,05	0,09
3		0,04	0,15	0,28
4		0,06	0	0,12

- Finn de marginale sannsynlighetsfordelingene til A og B.
- Finn kovariansen og korrelasjonskoeffisienten til A og B. Er variablene uavhengige?

Nynorsk

Oppgåve 1 (teller 20 %)

En eksamen blir gitt som flervalssprøve ("multiple choice test"). Eksamen består av tre spørsmål. For kvart spørsmål er det oppgitt fire svaralternativ men kor bare eitt er riktig. I oppgåve a) og b) skal vi anta at spørsmåla omhandlar heilt ulike tema. Sjølv om ein kjenner svaret på eitt spørsmål bidrar det ikkje til å gjere dei andre spørsmåla enklare.

- a) Ein student som ikkje veit svaret på noen av delspørsmåla bestemmer seg for å gjette heilt vilkårlig. (i) Kva er sannsynet for at studenten gjettar riktig på spørsmål 1? (ii) Kva er sannsynet for at studenten får riktig svar på alle dei tre spørsmåla? (iii) Kva er sannsynet for at ein student gjettar riktig på eit av spørsmåla og galt på dei to andre spørsmåla? Grunnge kort måten du regner ut svara på.

For å bestå prøven må ein student ha svart riktig på minst to av dei tre spørsmåla.

- b) Kva er sannsynet for å bestå prøven for ein student som gjettar helt vilkårlig på alle spørsmåla?

Året etter har flervalssprøven same form, men nå er det slik at dersom studenten kjenner svaret på spørsmål 1 er det bare to alternativ som kan være riktig for spørsmål 2. Spørsmål 3 er igjen uavhengig av de andre spørsmåla.

- c) Ein student kjenner svaret på spørsmål 1 men ikkje på spørsmål 2 og 3. (i) Kva er sannsynet for at studenten svarer rett på spørsmål 2? (ii) Kva er sannsynet for at studenten svarer rett på alle spørsmåla? (iii) Kva er sannsynet for at studenten består prøven? Grunnge kort måten du regner ut svara på.

Oppgåve 2 (teller 15 %)

Anta at variablane A og B er uavhengige og normalfordelte der $A \sim N(5,3)$ og $B \sim N(4,3)$

- a) Sett $C = A - B$. Forklar korfor $C \sim N(1, \sqrt{18})$
b) Finn sannsyna $Pr(A < B)$, $Pr(A = B)$ og $Pr(A > B)$.

Oppgåve 3 (teller 20 %)

Vi er interessert i omfang og konsekvensar av lekser i matematikk i barneskolen. Ein lærar vi kjenner er av den oppfatning at omlag 50 % av elevane får matematikklelse til minst halvdelen av matematikktimane, men at det ikkje gir noen betring i matematikkunnskapane til

SØK1004 – Statistikk for økonomer

elevane. TIMSS er ein internasjonal undersøking av elevers kunnskapar. Undersøkinga er nøye med å bruke representative utval. I en undersøking av realfagskunnskapane til norske 4. klassingar i 2007 ble lærarane spurt om omfanget på lekser. Nedanfor er resultatata for matematikk

	Antall elevar	Gjennomsnittresultat på matematikktesten	Standardavvik til testen
Lekser til minst halvparten av timane	2654	479,3	71,8
Lekser til mindre enn halvparten av timane	679	468,3	71,7

Testresultatet er normalfordelt.

- Kor stor del av elevane får lekse til minst halvparten av timane?
- Sett opp en hypotesetest og test om det er 50 % sannsyn for at ein tilfeldig elev får lekser til minst halvdel av timane. Bruk 5 % signifikansnivå.
- Test om matematikkunnskapane er ulik i gruppa som får mye lekser enn i gruppa som får lite lekser. Bruk 5 % signifikansnivå.
- Basert på resultatet i c), bør vi anbefale læraren vi kjenner å gi lekser til minst halvdel av timane?

Oppgåve 4 (teller 30 %)

- Det er en populasjonsmodell $Y_i = \alpha + \beta X_i + \varepsilon_i$. Gi ein tolking av parameteren β i denne modellen når Y representerer konsum og X representerer inntekt.?
- Vi ønskjer empirisk kunnskap om størrelsen på parameteren β . Beskriv kort korleis ein kan gå fram for å tileigna seg slik kunnskap.

Vi får tilgang til eit tilfeldig utval på 10 arbeidstakarar. Data er månadstall i 1 000 kroner.

Individ	Inntekt (X)	Konsum (Y)
1	27,0	15,4
2	25,9	22,1
3	26,8	23,9
4	35,7	40,2
5	37,6	25,0
6	15,1	14,8
7	30,2	26,2
8	29,5	20,1
9	25,0	27,7
10	40,1	30,7

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.

$$\bar{X} = 29,29, \quad \bar{Y} = 24,61, \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 470,6, \quad \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 500,6,$$

$$\sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = 326,8$$

- c) Bruk minste kvadrats metode til å estimere verdien på α og β .
- d) Bruk modellen til å predikere konsumet til ein arbeidstakar med månadsinntekt på 30 tusen kroner.
- e) “Konsumet er uavhengig av inntekt. Det ser ein for eksempel frå at individ 9 konsumerer meir enn inntekta”. Formuler denne påstanden som ein hypotesetest og test om ein må forkaste denne påstanden. Bruk 5 % signifikansnivå.
- f) “Det er bare inntekta som bestemmer konsumet. Aukar inntekta med ein krone så konsumerer ein også ein krone meir”. Formuler denne påstanden som ein hypotesetest og test om ein må forkaste påstanden.
- g) Det er foreslått å auke stipendet til studentar med 1 000 kroner i månaden. Kan du bruke modellen til å predikere effekten dette vil ha på studentars konsum?
- h) Kva er ein modells forklaringskraft (også kalla R^2)? Berekn denne.

Oppgåve 5 (teller 15 %)

Dei to variablane A og B har følgjande simultanfordeling.

A	B	0	2	3
1		0,21	0,05	0,09
3		0,04	0,15	0,28
4		0,06	0	0,12

- a) Finn dei marginale sannsynsfordelingane til A og B.
- b) Finn kovariansen og korrelasjonskoeffisienten til A og B. Er variablane uavhengig?

English

Question 1 (counts 20 %)

An exam is given as a multiple choice test. The exam consists of three questions. For each question there are four alternatives for which only one is correct. In part a) and b) we will assume that the questions deal with different issues such that if one knows the correct answer on one question that does not help in finding the correct answer on the other questions.

- a) A student who does not know the answer on any of the three questions decides to make guesses randomly. (i) What is the probability that the student will give a correct answer on Question 1? (ii) What is the probability that the student will give the correct answer on all three questions? (iii) What is the probability that the student gives the correct answer on one question and wrong answer on the other questions? Explain briefly how you have calculated your answers.

In order to pass the test the student must have correct answer on two of the three questions.

- b) What is the probability to pass the test for a student that guesses the answer randomly on all questions?

The next year the multiple choice test has the same format, but now it is the case that if the student knows the correct answer on Question 1 it is only two alternatives that can be correct on Question 2. Question 3 is again independent of the other questions.

- c) A student knows the correct answer on Question 1 but not on Questions 2 and 3. (i) What is the probability that the student give the correct answer on Question 2? (ii) What is the probability that the student gives the correct answer on all three questions? (iii) What is the probability that the student pass the test? Explain briefly how you have calculated your answers.

Question 2 (counts 15 %)

Assume that two variables A and B are independently and normally distributed where $A \sim N(5, 3)$ and $B \sim N(4, 3)$

- a) Set $C = A - B$. Explain why $C \sim N(1, \sqrt{18})$
b) Find the probabilities $Pr(A < B)$, $Pr(A = B)$ and $Pr(A > B)$.

Question 3 (counts 20 %)

We are interested in the extent and the effect of assigning homework in mathematics in primary school. A teacher we know has the understanding that about 50 % of the pupils have homework to at least half of the lessons in school, but that the homework assigned does not improve the mathematical skills of the pupils. TIMSS is an international comparative test of pupils' skills. The test uses random sampling. In the test of the skills in science and mathematics for Norwegian 4th graders in 2007 the teachers were asked how often they assigned homework. In the table below the results for mathematics are summarized.

	Number of pupils	Mean test score in mathematics	Standard deviation of the test score
Homework assigned for at least half of the lessons	2654	479.3	71.8
Homework assigned for less than half of the lessons	679	468.3	71.7

The test score is normally distributed.

- What is the share of pupils who have homework for at least half of the lessons?
- Formulate a hypothesis test and test whether there is 50 % probability that a random pupil have homework to at least half of the lessons. Use a significance level of 5 %.
- Test whether the skills in mathematics are different in the two groups of students. Use a significance level of 5 %.
- Based on the results in c), would you recommend the teacher we know to assign homework for at least half of the lessons?

Question 4 (counts 30 %)

- There is a population regression equation $Y_i = \alpha + \beta X_i + \varepsilon_i$. Give an interpretation of the parameter β in this model when Y represents consumption and X represents income.
- We want empirical knowledge of the parameter β . Describe briefly an approach to gain such knowledge.

We get access to a random sample of 10 workers. The data include monthly consumption and income measured in 1000 Norwegian kroner.

Individual	Income (X)	Consumption (Y)
1	27.0	15.4
2	25.9	22.1
3	26.8	23.9
4	35.7	40.2
5	37.6	25.0
6	15.1	14.8
7	30.2	26.2
8	29.5	20.1
9	25.0	27.7
10	40.1	30.7

$$\bar{X} = 29,29, \quad \bar{Y} = 24,61, \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 470,6, \quad \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = 500,6,$$

$$\sum_{i=1}^{10} (X_i - \bar{X})(Y_i - \bar{Y}) = 326,8$$

- Use the method ordinary least-squares to estimate α and β .
- Use the model to predict the consumption to a worker with monthly income of 30 thousand Norwegian kroner.
- “Consumption is independent of income. That is easy to see from for example the behavior of individual 9 in the data who consumes more than the income”. Formulate this statement as a hypothesis test and test whether the statement can be rejected. Use a significance level of 5 %.
- “It is only income that determines consumption. If a worker earns another krone he/she will also consume one more krone”. Formulate this statement as a hypothesis test and test whether the statement can be rejected.
- Someone has proposed to increase the monthly scholarship to students by 1 000 kroner. Can you use the model to predict which effect this will have on students’ consumption?
- What is meant by the model’s coefficient of determination (also denoted R^2)? Calculate this coefficient.

Question 5 (counts 15 %)

The two variables A and B have the following joint probability distribution.

	B	0	2	3
A				
1		0.21	0.05	0.09
3		0.04	0.15	0.28
4		0.06	0	0.12

- Find the marginal probability distributions of A and B.
- Find the covariance and the correlation coefficient to A and B. Are the variables independent?

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.

Løsningsforslag eksamen SØK1004, vår 2011

Oppgave 1

a) (i) $\frac{1}{4} = 0,25$

(ii) $\frac{1}{4} * \frac{1}{4} * \frac{1}{4} = 1/64 = 0,0156$

(iii) $(\frac{1}{4} * \frac{3}{4} * \frac{3}{4}) * 3 = 27/64 = 0,422$

Begrunnelsen bør nevne at dette er uavhengige hendelser. Da er svarene gitt av produktet av sannsynlighetene for hver hendelse. Det er da også mulig å bruke formelen for binomisk fordeling.

b) Dette er summen av sannsynligheten for to rette + en feil og alle rette.

Sannsynligheten for alle rette har vi funnet i a) (ii) til 0,0156. Sannsynligheten for to rette kan det regnes på samme måte som i a) (iii): $(\frac{1}{4} * \frac{1}{4} * \frac{3}{4}) * 3 = 9/64 = 0,140$.

Svaret er $9/64 + 1/64 = 10/64 = 0,156$.

c) (i) $\frac{1}{2} = 0,5$

(ii) $1 * \frac{1}{2} * \frac{1}{4} = 1/8 = 0,125$

(iii) Vi vet at studenten svarer rett på oppgave 1. For å bestå prøven må da studenten unngå å svare feil på både oppgave 2 og oppgave 3. Dette er summen av en rett + en feil i tillegg til begge rette. Sannsynligheten for en rett og en feil er $\frac{1}{2} * \frac{3}{4} + \frac{1}{2} * \frac{1}{4} = 4/8$. Sannsynligheten for rett på begge har vi fra c) (ii). Sannsynligheten for å bestå er derfor $4/8 + 1/8 = 5/8 = 0,625$.

Begrunnelsen bør nevne at oppgave 1 og 2 ikke er uavhengig av hverandre. Da er formelen for rett på både oppgave 2 og 1 gitt ved

$$\Pr(\text{Oppg2 og Oppg1}) = \Pr(\text{Oppg2}|\text{Oppg1}) * \Pr(\text{Oppg1})$$

Oppgave 2

a) Det skal vises at $C \sim N(1, \sqrt{18})$. I følge læreboka er første leddet i parentesens

gjennomsnittet og det andre leddet variansen. I oppgaven er det standardavviket som er oppgitt som andre ledd i parentesens i uttrykket for C.

$$E(C) = E(A) - E(B) = 5 - 4 = 1.$$

$$\text{Var}(C) = \text{Var}(A) + \text{Var}(B) = 3^2 + 3^2 = 18 \text{ siden } A \text{ og } B \text{ er uavhengige.}$$

Ved å betrakte det andre leddet i parentesens som variansen får man

$$\text{Var}(C) = \text{Var}(A) + \text{Var}(B) = 3 + 3 = 6$$

- b) $\Pr(A > B) = \Pr(C > 0) =$
 $\Pr\left(Z > \frac{0 - E(C)}{\sqrt{\text{Var}(C)}} = \frac{0 - 1}{\sqrt{18}} = -0,236\right) = 0,5 + \Pr(0 < Z < 0,236) = 0,5 + 0,093 = 0,593$
 $\Pr(A = B) = 0$
 $\Pr(A < B) = 1 - \Pr(A > B) - \Pr(A = B) = 1 - 0,593 - 0 = 0,407$
 Det siste kan eventuelt regnes ut på samme måte som $\Pr(A > B)$.

Oppgave 3

Tallene i denne oppgaven er de faktiske resultatene fra TIMSS 2007.

- a) $2654 / (679 + 2654) = 0,796$
- b) A = andelen som får lekser i minst halvparten av timene.
 $H_0: A = 0,5$
 $H_A: A \neq 0,5$
 Siden dette er en binomisk fordeling er testobservatoren normalfordelt. Kritisk verdi på 5 % nivå er 1,96.
 Variansen til A under H_0 er gitt $A * (1 - A) = 0,5 * 0,5 = 0,25$. Under sensuren ble det også godtatt at utvalgsvarians ble brukt. Det er gitt ved $0,796 * (1 - 0,796) = 0,162$.
 $TS = \frac{0,796 - 0,5}{\sqrt{\frac{0,25}{3333}}} = 34,18$
 H_0 forkastes på 5 % nivå.
- c) B = Differansen i testresultat mellom gruppene
 $H_0: B = 0$
 $H_A: B \neq 0$
 Utvalgsvariansen til B er gitt ved $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 71,78$
 $TS = \frac{479,3 - 468,3}{71,78 * \sqrt{\frac{1}{679} + \frac{1}{2654}}} = 3,56$
 Testobservatoren er t-fordelt med 3331 frihetsgrader. Kritisk verdi på 5 % signifikansnivå er 1,96. H_0 forkastes.
- d) Ja, men resultatet er usikkert blant annet pga. at det kan være andre viktige faktorer som påvirker testresultatet. Det er ikke estimerte en effekt her, og hvis man skulle gjøre det kunne det være relevant å bruke multipl regressjon. Her er det mulig med flere argumenter.

Oppgave 4

- a) Den marginale konsumtilbøyelighet
- b) Hypotesetesting basert på representativt utvalg. Utvalget kan være makroinformasjon over tid eller individer.

$$c) \beta = \frac{\sum xy}{\sum x^2} = \frac{326,8}{470,6} = 0,694, \quad \alpha = \bar{Y} - \beta \bar{X} = 24,61 - 0,694 * 29,29 = 4,28$$

$$d) \hat{Y} = a + bX = 4,28 + 0,694 * 30 = 25,1$$

$$e) H_0: \beta = 0$$

$$H_A: \beta \neq 0$$

Dette er en t-test med 8 frihetsgrader. Kritisk verdi på 5 % er 2,306

Estimert varians til b er gitt ved

$$s_b^2 = \frac{s^2}{\sum x^2} = \frac{\sum e^2 / (n-2)}{\sum x^2} = \frac{\sum y^2 - b^2 \sum x^2}{(n-2) \sum x^2} = \frac{500,6 - 0,694^2 * 470,6}{8 * 470,6} = 0,073$$

$$TS = \frac{0,694 - 0}{s_b} = \frac{0,694}{\sqrt{0,073}} = 2,57$$

Testobservatoren er større enn kritisk verdi. H_0 forkastes.

$$f) H_0: \beta = 1$$

$$H_A: \beta \neq 1$$

Dette er en t-test med 8 frihetsgrader. Absoluttverdien til kritisk verdi på 5 % er 2,306.

Estimert varians til b er den samme som i e).

$$TS = \frac{0,694 - 1}{s} = \frac{-0,306}{\sqrt{0,073}} = -1,13$$

Absoluttverdien til testobservatoren er 1,13. Det er mindre enn kritisk verdi. H_0 kan ikke forkastes.

- g) Nei. Undersøkelsen gjelder kun arbeidstakere, studenter er ikke representert. Evt., studenter har typisk lavere inntekt enn det vi observerer i utvalget.

- h) Andelen av variasjonen i Y som forklares av modellen. Det er flere formler for R^2 som kan benyttes, blant annet $R^2 = \frac{b^2 \sum x^2}{\sum y^2} = 0,453$.

Oppgave 5

A	B	0	2	3	Pr(A)
1		0,21	0,05	0,09	0,35
3		0,04	0,15	0,28	0,47
4		0,06	0	0,12	0,18
Pr(B)		0,31	0,20	0,49	

a) Er inkludert i tabellen over.

$$b) E(AB) = \sum \Pr(A, B) * A * B = 0,21 * 1 * 0 + 0,05 * 1 * 2 + 0,09 * 1 * 3 + \dots + 0,12 * 4 * 3 = 5,23$$

$$E(A) = \sum \Pr(A) * A = 0,35 * 1 + 0,47 * 3 + 0,18 * 4 = 2,48$$

$$E(B) = 1,87$$

$$Cov(A, B) = E(AB) - E(A)E(B) = 5,23 - 2,48 * 1,87 = 0,592$$

$$Var(A) = \sum (A - E(A))^2 p(A) =$$

$$(1 - 2,48)^2 * 0,35 + (3 - 2,48)^2 * 0,47 + (4 - 2,48)^2 * 0,18 = 1,31$$

$$Var(B) = 1,71$$

$$\rho = \frac{Cov(A, B)}{\sqrt{Var(A)}\sqrt{Var(B)}} = \frac{0,592}{\sqrt{1,31}\sqrt{1,71}} = 0,396$$

De er ikke uavhengige siden korrelasjonen er ulik 0.