

Institutt for samfunnsøkonomi

## **Eksamensoppgave i SØK2900 – Empirisk metode Empirical methods in Economics**

**Faglig kontakt under eksamen: Per Tovmo**

**Tlf.: 73 55 02 59**

**Eksamensdato:** 9. desember 2013

**Eksamenstid:** 4 timer

**Sensurdato:** 9. januar 2014

**Tillatte hjelpemidler:** Flg formelsamling: Knut Sydsæter, Arne Strøm og Peter Berck (2006): Matematisk formelsamling for økonomer, 4utg. Gyldendal akademiske. Knut Sydsæter, Arne Strøm, og Peter Berck (2005): Economists' mathematical manual, Berlin.  
Enkel kalkulator Citizen SR-270x, HP 30S eller SR-270X College

**Annen informasjon:** Eksamensoppgaven består av 5 oppgaver med delspørsmål som alle skal besvares.

**Målform/språk:** Norsk og engelsk

**Antall sider:** 6 inkl forside

**Antall sider vedlegg:** 3

*Oppgave 1 (15 %)*

La X og Y være realavkastningen i prosent for to ulike verdipapirer. Den felles sannsynlighetsfordelingen for X og Y er gitt ved tabellen under:

	X	2	3	4	5	6
Y						
2		0.01	0.03	0.06	0.05	0.01
3		0.03	0.04	0.09	0.06	0.01
4		0.04	0.06	0.10	0.07	0.04
5		0.05	0.01	0.15	0.08	0.01

- Finne de marginale sannsynlighetsfordelingene for X og Y.
- Beregne forventning og varians for X og Y.
- Forklar hva som menes med kovarians og korrelasjon. Beregn kovariansen og korrelasjonen mellom X og Y.
- En investor har plassert halvparten av finansformuen i hvert av de to verdipapirene. Finn forventet avkastning på investorens portefølje. Hva blir variansen til avkastningen?

*Oppgave 2 (15 %)*

En arbeidsmarkedsundersøkelse blant unge voksne i Norge omfatter 150 personer i alderen 18-25 år. Arbeidsledighetsraten i utvalget er 10 prosent.

- Konstruer 95 og 99 prosenters konfidensintervall for arbeidsledighetsraten på bakgrunn av denne undersøkelsen. Presiser forutsetningene du benytter.
- Hva er tolkningen av et konfidensintervall?
- En tilsvarende arbeidsmarkedsundersøkelse i Sverige omfatter 250 personer og arbeidsledighetsraten i det svenske utvalget er 15 prosent. Gir de to undersøkelsene grunnlag for å si at arbeidsledigheten blant unge voksne er høyere i Sverige enn i Norge? Formuler hypotesene og presiser forutsetningene bak testen.

*Oppgave 3 (30 %)*

Følgende data viser etterspurt kvantum (D) og enhetsprisen (P) for vare Q på 10 ulike tidspunkt.

D	99	91	70	79	60	55	70	101	81	67
P	22	24	23	26	27	24	25	23	22	26

$$\bar{D} = 77,3 \quad \bar{P} = 24,2 \quad \sum_{i=1}^{10} (D - \bar{D})^2 = 2246,1 \quad \sum_{i=1}^{10} (P - \bar{P})^2 = 27,6$$

$$\sum_{i=1}^{10} (D - \bar{D})(P - \bar{P}) = -143,6$$

- Anta at  $D_i = \alpha + \beta P_i + \varepsilon_i$  hvor  $\varepsilon_i$  er et restledd og  $\alpha$  og  $\beta$  er parametere. Bruk minste kvadraters metode til å estimere  $\alpha$  og  $\beta$ . Gi en tolkning av estimatene. Beregn modellens forklaringskraft.

- b) Formuler påstanden «Etterspørselen etter Q er uavhengig av prisen» som en hypotesetest og test om påstanden kan forkastes. (Hint: For å estimere restleddsvariansen kan du utnytte at  $R^2=1-SSR/SST$ , der SSR er summen av kvadrerte residualer og  $SST=\sum_{i=1}^{10}(D_i - \bar{D})^2$ ).
- c) Konstruer 90 og 95 prosenters konfidensintervall for  $\beta$ . Kommenter resultatene, spesielt i forhold til testen i oppgave b).
- d) En forsker foreslår å estimere en logaritmisk transformasjon av modellen i a). Hvilken tolkning har da parameteren for logaritmen til prisen?
- e) Under ser du en utskrift fra STATA som viser resultatene fra estimering av den logaritmiske transformasjonen. Bruk utskriften til å teste en hypotese om at priselastisiteten for Q er lik -1.

```
. regress lnD lnP
```

Source	SS	df	MS			
Model	.121021864	1	.121021864	Number of obs =	10	
Residual	.257602055	8	.032200257	F( 1, 8) =	3.76	
Total	.378623919	9	.042069324	Prob > F =	0.0885	
				R-squared =	0.3196	
				Adj R-squared =	0.2346	
				Root MSE =	.17944	

  

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnD						
lnP	-1.608723	.8298104	-1.94	0.089	-3.522269	.3048238
_cons	9.451022	2.642733	3.58	0.007	3.356869	15.54518

#### Oppgave 4 (25 %)

For et utvalg norske lønnstakere observeres følgende variable: Kjønn, timelønn, og utdanningslengde. Vi har et utvalg med 75 kvinner og 75 menn. Det hevdes ofte at kvinner blir underbetalt i arbeidslivet. Du skal undersøke denne påstanden ved bruk av regresjonsanalyse på dette datamaterialet. Følgende variable benyttes i analysen:

$W_i$  – timelønn for lønnstaker  $i$

$E_i$  – antall år utdanning for lønnstaker  $i$

$$D_{Fi} = \begin{cases} 1 & \text{hvis lønnstaker } i \text{ er kvinne} \\ 0 & \text{hvis lønnstaker } i \text{ er mann} \end{cases}$$

$$D_{Mi} = \begin{cases} 1 & \text{hvis lønnstaker } i \text{ er mann} \\ 0 & \text{hvis lønnstaker } i \text{ er kvinne} \end{cases}$$

Analysen baseres på følgende regresjonslikninger:

$$(1) W_i = \alpha_F D_{Fi} + \alpha_M D_{Mi} + \gamma_1 E_i + \varepsilon_i \quad i=1, 2, \dots, 150$$

$$(2) W_i = \beta_0 + \beta_1 D_{Fi} + \gamma_1 E_i + \varepsilon_i \quad i=1, 2, \dots, 150$$

- a) Gjør rede for sammenhengen mellom parametrene  $\alpha_F$  og  $\alpha_M$  i modell (1) og parametrene  $\beta_0$  og  $\beta_1$  i modell (2).

Modell (2) er estimert med minste kvadraters metode med følgende resultat:

$$W_i = 124,549 - 42,1839D_{Fi} + 2,651E_i$$

$$(19,80) \quad (8,163) \quad (1,535)$$

$$R^2=0,17, \text{ SSR}=365632,519$$

Tall i parentes under de estimerte parametrene er estimerte standardavvik,  $R^2$  er multipl determinasjonskoeffisient og SSR er summen av kvadrerte residualer.

- Presiser og test en hypotese om at kvinner blir underbetalt i norsk arbeidsliv.
- Ta utgangspunkt i resultatene over og beregn årslønnen for en mann en mann med 10 års utdanning som arbeider 1800 timer i året.
- Test om både  $D_{Fi}$  og  $E_i$  kan utelates fra modellen.

#### Oppgave 5 (15 %)

I perioden 1991 til 2005 måtte man betale en avgift for å kjøre med bil inn til Trondheim sentrum. Bomringen ble avviklet ved årsskiftet 2005/2006. Beskriv hvordan du ville gått frem for å teste empirisk hvilken effekt dette hadde på biltrafikken i Trondheim og presiser en hypotese.

Engelsk

#### Exercise 1 (15 %)

X and Y are the real returns in per cent on two different assets. The joint probability distribution for X and Y is presented in the table below:

X \ Y	2	3	4	5	6
2	0.01	0.03	0.06	0.05	0.01
3	0.03	0.04	0.09	0.06	0.01
4	0.04	0.06	0.10	0.07	0.04
5	0.05	0.01	0.15	0.08	0.01

- Find the marginal probability distributions for X and Y.
- Calculate expected value and variance for X and Y.
- Explain what is meant by covariance and correlation. Calculate the covariance and the correlation coefficient between X and Y.
- An investor has placed 50 percent of his financial wealth in each of the assets. Find expected return on his portfolio. Find the variance of the return.

#### Exercise 2 (15 %)

A labor market survey was conducted on 150 persons in the age group 18-25 years. The sample rate of unemployment is 10 per cent.

- Construct the 95 per cent and the 99 per cent confidence interval for the rate of unemployment based on this survey. State the assumptions you make.
- How do you interpret a confidence interval?

- c) A similar survey has been conducted on 250 persons in Sweden. The sample rate of unemployment was 15 per cent. Based on these two surveys, can you conclude that the rate of unemployment is higher in Sweden than in Norway? Formulate the hypotheses and state the assumptions behind the test.

### Exercise 3 (30 %)

Below you find data for demand (D) and unit price (P) for the good Q on 10 different points of time.

D	99	91	70	79	60	55	70	101	81	67
P	22	24	23	26	27	24	25	23	22	26

$$\bar{D} = 77,3 \quad \bar{P} = 24,2 \quad \sum_{i=1}^{10}(D - \bar{D})^2 = 2246,1 \quad \sum_{i=1}^{10}(P - \bar{P})^2 = 27,6$$

$$\sum_{i=1}^{10}(D - \bar{D})(P - \bar{P}) = -143,6$$

- Assume  $D_i = \alpha + \beta P_i + \varepsilon_i$  where  $\varepsilon_i$  is the error term and  $\alpha$  and  $\beta$  are parameters. Estimate  $\alpha$  and  $\beta$  by using ordinary least squares (OLS). Calculate the coefficient of determination (goodness of fit) and interpret the estimated coefficients.
- Formulate the statement «The demand for Q is independent of the price» as a hypothesis and test whether the statement can be rejected. (Hint: To estimate the variance of the residuals you can utilize the following:  $R^2 = 1 - \text{SSR}/\text{SST}$ , where SSR is the sum of squared residuals and  $\text{SST} = \sum_{i=1}^{10}(D_i - \bar{D})^2$ ).
- Construct the 90 and 95 per cent confidence interval for  $\beta$ . Comment on the results, especially in relation to the test in question b).
- A researcher suggests to estimate a logarithmic transformation of the model in question a). How do you interpret the estimated coefficient for  $\log(\text{Price})$ ?
- Below there is a printout from STATA containing the estimation results for the logarithmic transformation of the model. Use the printout to test whether the price elasticity for Q equals minus 1.

```
. regress lnD lnP
```

Source	SS	df	MS	Number of obs = 10		
Model	.121021864	1	.121021864	F( 1, 8) =	3.76	
Residual	.257602055	8	.032200257	Prob > F =	0.0885	
Total	.378623919	9	.042069324	R-squared =	0.3196	
				Adj R-squared =	0.2346	
				Root MSE =	.17944	

  

lnD	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnP	-1.608723	.8298104	-1.94	0.089	-3.522269	.3048238
_cons	9.451022	2.642733	3.58	0.007	3.356869	15.54518

*Exercise 4 (25 %)*

For a sample of Norwegian workers consisting of 75 women and 75 men we have observed the following variables: Gender, wages per hour and years of education. It is often said that women are underpaid in the labor market. We wish to investigate this assertion by conducting regression analysis on our data set. We use the following variables in the analysis:

$W_i$  – wage per hour paid to worker  $i$

$E_i$  – years of education of worker  $i$

$$D_{Fi} = \begin{cases} 1 & \text{if worker } i \text{ is a woman} \\ 0 & \text{if worker } i \text{ is a man} \end{cases}$$

$$D_{Mi} = \begin{cases} 1 & \text{if worker } i \text{ is a man} \\ 0 & \text{if worker } i \text{ is a woman} \end{cases}$$

The analysis is based on the following models:

$$(1) W_i = \alpha_F D_{Fi} + \alpha_M D_{Mi} + \gamma_1 E_i + \varepsilon_i \quad i=1, 2, \dots, 150$$

$$(2) W_i = \beta_0 + \beta_1 D_{Fi} + \gamma_1 E_i + \varepsilon_i \quad i=1, 2, \dots, 150$$

- a) Clarify the relations between the parameters  $\alpha_F$  and  $\alpha_M$  in model (1) and the parameters  $\beta_0$  and  $\beta_1$  in model (2).

We get the following result when estimating model (2) with OLS:

$$W_i = 124,549 - 42,1839D_F + 2,651E_i$$

(19,80)      (8,163)      (1,535)

$$R^2=0,17, \text{ SSR}=365632,519$$

The figures in parenthesis below the estimated parameters are estimated standard errors,  $R^2$  is the coefficient of multiple determination and SSR is the sum of squared residual.

- b) Formulate a hypothesis and test whether women are underpaid in the labor market.  
 c) Use the results above and calculate the annual income for a man with 10 years of education and who works 1800 hours per year.  
 d) Test if both  $D_{Fi}$  and  $E_i$  can be left out of the model.

*Exercise 5 (15 %)*

From 1991 to 2005 you had to pay a fee to go by car into the city center of Trondheim. The fee was taken away in 2006. Describe how you would conduct an empirical analysis of the effect the removal of the fee had on the traffic in Trondheim. Formulate a hypothesis that can be tested empirically.