

UNIVERSITETET I OSLO

ØKONOMISK INSTITUTT

Eksamen i: ECON3150/4150 – Elementær økonometri
Exam: ECON3150/4150 – Introductory econometrics

Eksamensdag: Tirsdag 30. mai 2006
Date of exam: Tuesday, May 30, 2006

Sensur kunngjøres: Torsdag 15. juni
Grades will be given: Thursday, June 15

Tid for eksamen: kl. 09:00 – 12:00
Time for exam: 09:00 a.m. – 12:00 noon

Oppgavesettet er på 8 sider (inkl. utskrifter)
The problem set covers 8 pages (incl. outputs) English version on page 4

Tillatte hjelpemidler:

- Alle trykte og skrevne hjelpemidler, samt lommekalkulator er tillatt

Resources allowed:

- *All printed and written resources, as well as calculator are permitted*

Eksamen blir vurdert etter ECTS-skalaen. A-F, der A er beste karakter og E er dårligste ståkarakter. F er ikke bestått.

The grades given: A-F, with A as the best and E as the weakest passing grade. F is fail.

Ledelsen for en fornøyelsespark ønsket å få bedre oversikt over faktorer som påvirker tilstrømningen av besøkende til parken i løpet av en sesong, og har gjennomført en utvalgsundersøkelse av familier i regionen der parken ligger. I undersøkelsen rapporterte familiene blant annet kjøreavstand til parken (X_{j1}) (målt i km), antall medlemmer i familien (X_{j2}) samt antall ganger (Y_j) familien hadde besøkt parken i løpet av siste sesong.

For å analysere de innsamlede data engasjerte ledelsen en økonometriker. I stedet for å hoppe direkte på en regresjonsanalyse, skisserer økonometrikeren følgende opplegg for analysen: Familiers tilfredsstillelse/nytte av besøk i parken oppsummeres av en nyttefunksjon $U(Y)$. Selv om utvalgsundersøkelsen gir oss nyttig informasjon om variable som påvirker antall besøk, kan fremdeles relevante faktorer være ignorert. Det er derfor naturlig å oppfatte antall besøk $\{Y\}$ som en stokastisk variabel. Spesielt skal vi anta at $\{Y\}$ er Poisson fordelt, eller:

$$(1) \quad P\{Y = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2, 3, \dots$$

Spørsmål 1

Anta at nytten for en vilkårlig familie j er gitt ved:

$$(2) \quad U(Y_j) = Y_j$$

Vis at familiens forventede nytte av besøk i parken er gitt ved:

$$(3) \quad E[U(Y_j)] = E(Y_j) = \lambda_j$$

Vi skal i denne oppgaven anta at λ_j er en lineær funksjon av familiens kjørelengde til parken (X_{j1}) og antall medlemmer i familien (X_{j2}), dvs.

$$(4) \quad \lambda_j = \beta_0 + \beta_1 X_{j1} + \beta_2 X_{j2}$$

Som bakgrunn for regresjonsanalysen postulerer vi at $Y_j = E[U(Y_j)] + \varepsilon_j$, og analyserer først tilfellet der nyttefunksjonen er lineær. Vi skal derfor først studere regresjonslikningen:

$$(5) \quad Y_j = \beta_0 + \beta_1 X_{j1} + \beta_2 X_{j2} + \varepsilon_j$$

der ε_j betegner stokastiske restledd. Restleddene ε_j har forventning null, og for enkelhets skyld antar vi at ε_j har konstant varians.

Spørsmål 2

Utskrift 1 viser resultatet av regresjonen (5). Kommenter og diskuter disse resultatene.

Spørsmål 3

- (a) Estimer forventet nytte (λ_j) for en familie med følgende verdier på forklaringsvariablene: $X_{j1} = 30$ og $X_{j2} = 4$.
- (b) Still opp et konfidensintervall for (λ_j) for denne familien når du får oppgitt at det estimerte standardavviket til estimeringsfeilen $F = (\lambda_j - \hat{\lambda}_j)$ er lik 0.172. Velg konfidenskoeffisient for intervallet lik 0.95.

Økonomer vil ofte hevde at den lineære nyttefunksjonen gitt ved (2) er for restriktiv. For å imøtekomme denne innvendingen, foreslår økonometrikeren nå å ta utgangspunkt i en kvadratisk nyttefunksjon og spesifiserer:

$$(6) \quad U(Y_j) = Y_j + cY_j^2$$

Spørsmål 4

Vis at forventet nytte nå vil være gitt ved:

$$(7) \quad E[U(Y_j)] = \lambda_j + c(\lambda_j^2 + \lambda_j) = (1+c)\lambda_j + c\lambda_j^2$$

Med bruk av uttrykket for λ_j gitt ved (4) og den forventede nytten (7), spesifiserte økonometrikeren regresjonen:

$$(8) \quad Y_j = b_0 + b_1 X_{j1} + b_2 X_{j2} + \gamma_1 X_{j1}^2 + \gamma_2 X_{j2}^2 + \gamma_3 X_{j1} X_{j2} + u_j,$$

der vi igjen antar at restleddene u_j har forventning null og konstant varians.

Resultatene fra regresjon (8) er gitt i Utskrift 2.

Spørsmål 5

Flere av estimatene på regresjonskoeffisientene i (8) viser seg å være upresise. Drøft mulige grunner til dette.

For å unngå problemene man møtte ved estimering av (8), ble variabelen (X_{j2}^2) ekskludert slik at man til slutt estimerte regresjonen:

$$(9) \quad Y_j = b_0 + b_1 X_{j1} + b_2 X_{j2} + \gamma_1 X_{j1}^2 + \gamma_3 X_{j1} X_{j2} + u_j$$

Resultatet fra denne beregningen er vist i Utskrift 3.

Spørsmål 6

- (a) Kommenter resultatene fra denne regresjonen.
- (b) Synes du regresjonsresultatene støtter påstanden om at lineær nytte er for restriktiv?
- (c) Hvordan vil du gå frem for å velge mellom regresjonene (5) og (9)?

Spørsmål 7

- (a) Når vi sammenligner forholdet mellom estimatene $\begin{bmatrix} \hat{\beta}_2 \\ \hat{\beta}_1 \end{bmatrix}$ i regresjon (5) (Utskrift 1)

og forholdet mellom de tilsvarende estimatene i regresjon (9) $\begin{bmatrix} \hat{b}_2 \\ \hat{b}_1 \end{bmatrix}$ (Utskrift 3),

finner vi at disse er ganske like, henholdsvis -7.01 og -7.17 . Forsøk å gi en forklaring på dette.

- (b) Uten at du gir et bestemt estimat, forsøk å skissere et opplegg for å estimere c .

ENGLISH VERSION

The management of a pleasure park wanted to improve their knowledge of factors that have impact on the number of people visiting the park during a season. For that reason they obtained a random sample of families in the region. In this sample the families among other things, reported their driving distance to the park (X_{j1}) (measured in km), the number of family members (X_{j2}), and the number of visits (Y_j) to the park in the last season. In order to analyse this data set the management hired an econometrician. Instead of jumping directly to a regression analysis, the econometrician approached the statistical modelling along the following lines: A family's satisfaction/utility of visits to the park can be summarized by a utility function $U(Y)$. Although the survey conducted have given useful information about variables influencing the number of visits, relevant factors can still have been overlooked. It is therefore reasonable to view the number of visits $\{Y\}$ as a random variable. Formally, we assume that $\{Y\}$ has a Poisson distribution, i.e.

$$(1) \quad P\{Y = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2, 3, \dots$$

Question 1

Assume that the utility of an arbitrary family j is given by:

$$(2) \quad U(Y_j) = Y_j$$

Show that the family's expected utility of visits to the park is given by:

$$(3) \quad E[U(Y_j)] = E(Y_j) = \lambda_j$$

We shall in this problem set assume that λ_j is a linear function of the family's driving distance to the park (X_{j1}) and the number of members in the family (X_{j2}), or

$$(4) \quad \lambda_j = \beta_0 + \beta_1 X_{j1} + \beta_2 X_{j2}$$

As a background for the regression analysis we postulate that $Y_j = E[U(Y_j)] + \varepsilon_j$, and analyse first the case with linear utility function. Thus, we shall first study the regression equation:

$$(5) \quad Y_j = \beta_0 + \beta_1 X_{j1} + \beta_2 X_{j2} + \varepsilon_j$$

where ε_j denote the random disturbances. The disturbances ε_j have means zero and for simplicity we assume that the variances of ε_j are constant.

Question 2

Output 1 shows the results of regression (5). Give your comments to and discuss these results.

Question 3

- (a) Estimate the expected utility (λ_j) for a family with the following values of the exogenous variables: $X_{j1} = 30$ and $X_{j2} = 4$.
- (b) Construct a confidence interval for (λ_j) pertaining to this family when you are informed that the standard deviation of the estimation error $F = (\lambda_j - \hat{\lambda}_j)$ is equal to 0.172. Choose a coefficient of confidence equal to 0.95.

Economists will often say that the linear utility (2) is too simple. In order to meet this criticism, the econometrician proposes also to try a quadratic utility and specify:

$$(6) \quad U(Y_j) = Y_j + cY_j^2$$

Question 4

Show that the expected utility in this case will be:

$$(7) \quad E[U(Y_j)] = \lambda_j + c(\lambda_j^2 + \lambda_j) = (1+c)\lambda_j + c\lambda_j^2$$

Using the expression for λ_j given by (4) and the value of the expected utility (7), the econometrician specified the regression:

$$(8) \quad Y_j = b_0 + b_1X_{j1} + b_2X_{j2} + \gamma_1X_{j1}^2 + \gamma_2X_{j2}^2 + \gamma_3X_{j1}X_{j2} + u_j,$$

and again we assume that the disturbances u_j have means zero and constant variances. The results of regression (8) are shown in Output 2.

Question 5

Some estimates of the regression coefficients appearing in (8) prove to be imprecise. Discuss possible reasons for this outcome.

In order to avoid the problems we encountered in estimating the regression equation (8), the variable (X_{j2}^2) has been excluded, so that finally we estimated the regression:

$$(9) \quad Y_j = b_0 + b_1X_{j1} + b_2X_{j2} + \gamma_1X_{j1}^2 + \gamma_3X_{j1}X_{j2} + u_j$$

The results of this regression is shown in Output 3.

Question 6

- (a) Give your comments to the results of this regression.
- (b) Do you think that these results support the assertion that a linear utility is too simple?
- (c) Explain how you will proceed if you want to choose between the regressions (5) and (9)?

Question 7

- (a) When we compare the ratio between the estimates $\left[\frac{\hat{\beta}_2}{\hat{\beta}_1} \right]$ in regression (5)

(Output 1), and the ratio between the corresponding estimates in regression (9)

$\left[\frac{\hat{b}_2}{\hat{b}_1} \right]$ it turns out that these ratios are quite similar, -7.01 and -7.17 , respectively.

Try to give an explanation of this finding.

- (b) Without trying to give an estimate of c , give a sketch of a procedure that can be used to estimate c .

Utskrift/Output 1

	Coefficient	Std.Error	t-value	t-prob
Constant	4.19330	0.3350	12.5	0.000
X_{j1}	-0.0449795	0.003420	-13.2	0.000
X_{j2}	0.315368	0.07472	4.22	0.000
sigma	1.65043	RSS	672.812106	
R^2	0.433003	F(2,247) =	94.31 [0.000]**	
no. of observations	250	no. of parameters	3	
mean(Y)	2.248	var(Y)	4.7465	

Covariance matrix of estimated parameters:

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
$\hat{\beta}_0$	0.11223	-0.00074483	-0.017019
$\hat{\beta}_1$	-0.00074483	1.1697e-005	-5.1266e-006
$\hat{\beta}_2$	-0.017019	-5.1266e-006	0.0055831

Utskrift/Output 2

	Coefficient	Std.Error	t-value	t-prob
Constant	5.29887	0.9152	5.79	0.000
X_{j1}	-0.0882092	0.01819	-4.85	0.000
X_{j2}	0.140071	0.4202	0.333	0.739
X_{j1}^2	0.000451080	0.0001198	3.77	0.000
X_{j2}^2	0.0782559	0.06171	1.27	0.206
$X_{j1}X_{j2}$	-0.00485863	0.002445	-1.99	0.048
sigma	1.59512	RSS	620.837966	
R^2	0.476803	F(5,244) =	44.47 [0.000]**	
no. of observations	250	no. of parameters	6	
mean(Y)	2.248	var(Y)	4.7465	

Correlation matrix:

	X_{j1}	X_{j2}	X_{j1}^2	X_{j2}^2	$X_{j1}X_{j2}$
X_{j1}	1.0000	0.020061	0.97782	0.037963	0.69953
X_{j2}	0.020061	1.0000	0.040931	0.98171	0.66819
X_{j1}^2	0.97782	0.040931	1.0000	0.056216	0.69439
X_{j2}^2	0.037963	0.98171	0.056216	1.0000	0.66710
$X_{j1}X_{j2}$	0.69953	0.66819	0.69439	0.66710	1.0000

Utskrift/Output 3

	Coefficient	Std.Error	t-value	t-prob
Constant	4.66108	0.7655	6.09	0.000
X_{j1}	-0.0865747	0.01817	-4.77	0.000
X_{j2}	0.621315	0.1806	3.44	0.001
X_{j1}^2	0.000443821	0.0001198	3.70	0.000
$X_{j1}X_{j2}$	-0.00494753	0.002447	-2.02	0.044
sigma	1.5971	RSS	624.929433	
R^2	0.473355	F(4,245) =	55.05 [0.000]**	
no. of observations	250	no. of parameters	5	
mean(Y)	2.248	var(Y)	4.7465	