



Stockholms universitet

OBS! Läs noga igenom anvisningarna i tentamen, t.ex. hur du ska skriva svaren.
Det är ditt ansvar som student att följa de anvisningar som ges.

NOTE! Read the examination instructions carefully, e.g. how to write the answers.
It is your responsibility as a student to follow the given instructions.

Skriv din anonymiseringskod och dagens datum på allt material du lämnar in.
(Enter your anonymization code and today's date on all submitted materials)

Anonymiseringskod (Anonymization code)	3	1	1	-	0	0	0	7	-	X	X	F
Datum (Date YYYY-MM-DD)	2021-11-29						Plats nr. (Seat No.)	6				

Kurs/Kurskod (Course/Course code)	ST2236
Kursmoment (Course component)	31ET Regression analysis

Fylls i av tentamensvärd (To be filled in by invigilator)

Direkt i skrivning: (kryss)		Svarsblankett: (kryss)		Lösa svarsblad: (antal)	9
--------------------------------	--	---------------------------	--	----------------------------	---

Lämnat in blankt: (kryss)		Dator: (kryss)	
------------------------------	--	-------------------	--

Inlämningstid: 17:50

Signatur tentamensvärd: RM

Fylls i av lärare/examinator (To be filled in by teacher/examinator)

Betyg:	B	Poäng:	82
--------	---	--------	----

Signatur rättande lärare/examinator: On Lefter



Datum: (Date YYYY-MM-DD)	2021-11-29	Kurs/Kurskod: (Course/Course code)	S12236	Sidnr.: (Page no.)	1
Anonymiseringskod (Anonymization code)	311-0007-XXF				

Frågor 1

Uppg.nr.:
(Task no.)

1

Lärarens
kommentar:
(Teacher's
note)

1) svar: d) R

2) svar: d) R

3) svar: b) R

4) svar: d) - +5

3) svar: a) R

tot: 25
~~20~~/25

Poäng:
(Points)

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



a) Då modellen är en level-log och vi inte vill veta hur en procentuell förändring i x påverkar y , utan hur en enhetsförändring i x påverkar y . Tar jag derivatan med avseende på x_1 , se nedan:

$$\frac{d\beta_1}{dx_1} + \frac{d\beta_2}{dx_1} \Rightarrow -\frac{7,9}{x_1} - \frac{1,2x_2}{x_1}$$

Det syns att "Partiell effekten" inte är konstant och att en enhetsförändring i x_1 också påverkas av värdet på x_2 .

Men en enhetsförändring i x_1 leder till att y förändras med $-\frac{7,9}{x_1} - \frac{1,2x_2}{x_1}$ enheter. 5

b) Svar: Tolkningen är att om x_2 ändras med en enhet, givet ceteris paribus, kommer y ändras med 8,6 enheter då β_2 har ett parametervärde på 8,6.

Du behöver $x_1 = 1 \Leftrightarrow \log(x_1) = 0$ i interaktionen

c) Hitta dem värden som sökas.

① std.error $\log x_1 = t\text{-test} \Rightarrow \frac{-7,9}{se(x_1)} = -2,893 \Rightarrow se(x_1) = \frac{-7,9}{-2,893}$
 $se(x_1) \approx 2,7315$ R

② std.error $x_2 \Rightarrow T\text{-test} \Rightarrow \frac{8,6}{se(x_2)} = 6,26 \Rightarrow se(x_2) = \frac{8,6}{6,26} = 1,3738$ R
 $se(x_2) \approx 1,3738$
värdet är redan $\hat{\sigma}^2$

3. SSR $\Rightarrow \hat{\sigma}^2 = \frac{SSR}{n-k-1} \Rightarrow 0,25 = \frac{SSR}{200-4-1} \Rightarrow 0,25 = \frac{SSR}{195} \Rightarrow$

$0,25 \cdot 195 = SSR = 12,1875$ följaktligen

$SSR = 12,1875$

fort. nästa sida \Rightarrow

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2021-11-29	Kurs/Kurskod: (Course/Course code)	ST223G	Sidnr.: (Page no.)	3
Anonymiseringskod (Anonymization code)	3 1 1 - 0 0 0 7 - X X F				

Uppg.nr.:
(Task no.)
20, e)

Lärarens kommentar:
(Teacher's note)

R^2 . För att hitta R^2 måste vi först hitta SST och det görs genom att ställa upp formeln för "adjusted R^2 ".

$$R^2_{adj} = 1 - \frac{SSR}{SST} \cdot \frac{n-1}{n-k-1} \Rightarrow 0,765 = 1 - \frac{12,1875}{SST} \cdot \frac{200-1}{200-9-1}$$

blir fel med denna

$$0,765 = 1 - \frac{12,1875}{SST} \cdot 1,020512821 \Rightarrow -0,255 \cdot SST = -12,1875 \Rightarrow$$

$$SST = 51,861702$$

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{12,1875}{51,861702} = 0,765$$

4

$$R^2 = 0,765$$

Standard error (log x_1) = 2,7915, standard error (x_2) = 1,3738,

SSR = 12,1875, R^2 = 0,765

e) Ett sett man kan testa det är genom ett RESET-test. Det testet så testar man om modellen saknar kvadratiske-termer (kan vara högre polynom) och interaktionstermer. Man testar alltså "misspecification" i modellen. Alltså att man har missat viktiga icke-linjära termer tex. Man inkluderar F^2 i testet för ett sedan test a p-r. exempel.

forts →

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2021-11-29	Kurs/Kurskod: (Course/Course code)	ST2236	Sidnr.: (Page no.)	
Anonymiseringskod (Anonymization code)	3 11 - 0 0 0 7 - X X F				4

exempel:

$$y = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 \log(x_1) x_2 + \delta \gamma^2 + \omega \gamma^3$$

H₀: $\delta = 0$ H_a: $\delta \neq 0$

Uppg.nr.:
(Task no.)
29

Lärarens kommentar:
(Teacher's note)

Det testet visar på att om man inte väljer för stora δ så blir termen inte tillräckligt felaktig term. Viktigt att tilläggs är att testet säger inte vilka icke-linjära parametrar som f. t. t. s och man kan ha med högre polynom. t. ex γ^3, γ^4 . (5)

d) Vi använder följande modell nedan för att hitta värdet på β_3 när x_3 ändras istället för x_2

$$\text{Plim}(\beta_3) = \beta_3 \left(\frac{\sigma_{x_3}^2}{\sigma_{x_3}^2 + \sigma_e^2} \right) = 2 \left(\frac{2}{2 + 0,33} \right) = 2,575107296$$

Svar: värdet på β_3 skulle då vara 2,575107296 alltså undervärderas på grund av attenuation bias

detta är rätt formel för enkel linjär regression kallas Waldstatistik för fler beroende variabler. Jag frågar efter åtgärd också.

(2)

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2021-11-29	Kurs/Kurskod: (Course/Course code)	GT223G	Sidnr.: (Page no.)	
Anonymiseringskod (Anonymization code)	311-0007-XXF				5

Fråga 3

Uppg.nr.:
(Task no.)
3, a), b)

Lärarens kommentar:
(Teacher's note)

a) Anledning till att "age" har ett negativt tecken och "age²" har ett positivt tecken är för att marginal effekten av att lägga ner pengar på mat avtar efter en viss ålder. Det vill säga att man lägger ner pengarna på mat tills man kommit till en viss ålder och då istället lägger man mindre pengar på mat. Det kan vara så att barnen i hushållet flyttar ut när föräldrarna i hushållet blir äldre och därmed köper dem mindre mat för att inga barn bor hemma längre.

Hitta när marginal effekten byter tecken:

$$\frac{dy}{dage} = 0,009 - 2 \cdot 0,0001 age = 0 \Rightarrow age = 45$$

när åldern > 45 år kommer marginal effekten bli negativ.

5

b) F-test för att undersöka om den nya variabelerna ska inkluderas

$H_0: \beta_4 = 0, \beta_5 = 0, \beta_6 = 0$ H_1 : Åtminstone en är skild från noll.
 $\alpha = 0,05$ Teststatistik = $\frac{R_{ur}^2 - R_r^2 / q}{(1 - R_{ur}^2) / (n - k - 1)}$

$$F\text{-test} = \frac{(R_{ur}^2 - R_r^2) / q}{(1 - R_{ur}^2) / (n - k - 1)} = \frac{(0,431 - 0,353) / 3}{(1 - 0,431) / (1000 - 6 - 1)} = 45,37$$

F = 45,37

Svar: H_0 kan förkastas på en 5% signifikansnivå då $45,37 > F_{0,05} = 2,6$. Alltså, de nya variabelerna ska inkluderas i modellen då H_0 kan förkastas.

q=3 2,6

4

Poäng:
(Points)

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2021-11-29	Kurs/Kurskod: (Course/Course code)	ST223C	Sidnr.: (Page no.)	6
Anonymiseringskod (Anonymization code)	311-0007-XXF				

1) Svar: En anledning till att koefficienterna inte är signifikanta (insignificant) kan bero på multikollinearitet i modellen (multicollinearity).
 Det vill säga att koefficienterna korrelerar med dem nya interaktionstermerna. Det är logiskt att dem korrelerar med varandra, då det är interaktioner med just dem variablerna i regression 2. Sammanfattningsvis kan det bero på "multicollinearity" att koefficienterna i regression två ej är signifikanta i regression 3.

Uppg.nr.:
(Task no.)
3, 2, 1

Lärarens kommentar:
(Teacher's note)

5

Någon vis-s hypoteserna som ska testas.
 H₀: β₇=0, β₈=0, β₉=0 H_a: Åtminstone en är skild från 0.
 För att kunna göra hypotestestet måste vi först veta R² för regression 1 se.

hitta R₂

Vi vet att F-statistic for overall significance of a Regression ser ut på följande, och vi vet F-värdet.

$$\frac{R^2/k}{(1-R^2)/(n-k-1)} = F\text{-värde} \Rightarrow 83,969 = \frac{R^2/9}{(1-R^2)/990} \Rightarrow$$

$$\frac{R^2}{9} = 83,969 \Rightarrow \frac{R^2 \cdot 990}{(1-R^2) \cdot 9} = 83,969 \Rightarrow \frac{R^2 \cdot 990}{(1-R^2)} = 755,721$$

$$990R^2 = 755,721 - 755,721R^2 \Rightarrow 1745,211R^2 = 755,721 \Rightarrow R^2 = 0,4327777777777777$$

När vi vet R₂ för Reg 3, kan F-testet utföras.

$$\frac{(R^2_{3r} - R^2_r)/q}{(1-R^2_{3r})/(n-k-1)} \Rightarrow \frac{(0,93277 - 0,43277)/3}{(1-0,93277)/(990)} = F = 0,988$$

F-värde = 0,988

det kritiskt värde, annars rätt.

4

Svar: då F-värdet = 0,988, kan H₀ inte förkastas på en 5% signifikansnivå, då 0,988 < 1,83. Det vill säga interaktionstermerna ska exkluderas.

Poäng:
(Points)

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



Datum: (Date YYYY-MM-DD)	2021-11-29	Kurs/Kurskod: (Course/Course code)	SI 2236	Sidnr.: (Page no.)	
Anonymiseringskod (Anonymization code)	311-0		007-XXF		7

e) Jag sätter in värdena i regressionsmodellen och extraherar
 dummy-variablernas som blir noll.

$$\begin{aligned}
 \hat{y} &= 2305 - 0,168 \log(162754,8) + 0,006 \cdot 50 - 0,00004 \cdot 50^2 \\
 &+ 0,060 \cdot 1 + 0,016 \log(162754,8) \cdot 1 = 1,486078321
 \end{aligned}$$

↑ Svar: "mean prediction" är 1,486078321

3

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



a) $y = 4 + 26x_1 + 3x_2 + u$, $x_2 = 3 + 0,5x_1 \Rightarrow y = 4 + 26x_1 + 3(3 + 0,5x_1) = 4 + 26x_1 + 9 + 1,5x_1$
 $y = 13 + 27,5x_1 + u$, $\beta_1 = 27,5$

Uppg.nr.: (Task no.)

4, 5, 6

1. regressionen kommer β_1 effektivt att överestimeras på grund av omitted variable bias. Variabeln x_2 kommer att vara endogen och korrelera med feltermen u . Så anledningen till att β_1 inte har det sanna värdet är för att x_2 är endogen och korrelerar därmed med feltermen. värdet β_1 kommer ha är 27,5.

Lärarens kommentar: (Teacher's note)

6

b) Metoden kallar vi för proxy-variabel, det vill säga att \tilde{x}_2 används som en proxy-variabel för x_2 .

För att proxy-variabeln ska vara "consistent" och leda till att β_1 blir en bra consistent estimator behöver några antaganden vara uppfyllda.

- Feltermen u måste vara icke-korrelerad med x_1 och \tilde{x}_2 .

- När proxy-variabeln (\tilde{x}_2) estimeras $\tilde{x}_2 = \delta_0 + \delta_1 x_1 + v_2$, så måste feltermen v_2 vara icke-korrelerad med x_1 , för att vara en bra proxy. Enligt sättet att

se det är genom att $E(\tilde{x}_2 | x_1) = E(x_2 | x_1) = \delta_0 + \delta_1 x_1$.

6

Om dessa antaganden är uppfyllda och \tilde{x}_2 är en bra prox för x_2 tar vi bort endogeniteten i x_1 försvinn och vara en exogen variabel, och därmed kunna estimeras rätt.

c) Två antaganden måste uppfyllas.

① "instrument exogeneity". Det innebär att instrument variabeln ska vara icke-korrelerad med feltermen, alltså $Cov(\tilde{x}_2, u) = 0$.

② Andra antagandet är att "instrument relevance", instrumentet ska vara relevant. Detta innebär att $Cov(\tilde{x}_2, x_1) \neq 0$.

Vi kan säga något om andra antagandet, att variabeln ska vara relevant, vi kan göra ett t-test för att se om instrument-variabeln är relevant.

T-test = $\frac{0,29}{0,06} = 4,833$. Värdet är stort och därmed signifikant.

en 5-procentig nivå. Instrumentet kan vara relevant. ok
 Nullhypotesen kan förkastas. Detta fortsättning \rightarrow

Poäng: (Points)

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)



c) (fortsättning) Däremot kan inte första antagandet testas. Det vill säga om instrumentet är exogent från feltermen, $Cov(z_i, \epsilon_i) = 0$. Anledningen till att det inte går att utföra testet är för att det måste finnas fler instrument än endogena variabler, om det ska vara möjligt att utföras. Detta fall finns det inte och därmed kan testet inte utföras.

7

d) F0

Uppg.nr.:
(Task no.)

Lärarens
kommentar:
(Teacher's
note)

Poäng:
(Points)