Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 27-11-2017

Skriftlig tentamen i **Regressionsanalys och tidsserieanalys** (4,5 hp), ingående som moment 1 i kursen **Regressionsanalys och undersökningsmetodik, 15 hp.**

Skrivtid: 5 timmar
Hjälpmedel: Miniräknare utan lagrade formler eller lagrad text. Vidhäftade formel- och tabellblad (obs! vidhäftas endast de tabellsidor som behövs för den här tentamen).
Tentamensgenomgång och återlämning: tisdagen den 12 december, kl. 16.00 i B315.
Därefter kan skrivningarna hämtas på studentexpeditionen, plan 7 i B-huset.

Tentamen består av fem uppgifter som kan ge totalt 100 poäng. För betyget A gäller 90-100 p., för betyget B gäller 80-89 p., för betyget C gäller 70-79 p., för betyget D gäller 60-69 p., för betyget E gäller 50-59 p., för betyget Fx gäller 40-49 p. och för betyget F gäller 0-39 p. För detaljerade betygskriterier se kursbeskrivningen på kurshemsidan.
För full poäng på en uppgift krävs fullständiga och väl motiverade lösningar.

Uppgift 1: (20 poäng)
Ett företag i Grönköping undersökte vid ett tillfälle sambandet mellan personers inkomst (INK) och deras årskonsumtion (KON) av en av företagets produkter. Följande information har erhållits:

```
The regression equation is
KON = 2,93 + 0,0588 INK

Predictor          Coef        SE Coef
Constant           2,93         1,755
INK                0,0588       0,0246

s =        R-sq=        R-sq (adj) =

Analysis of variance

SOURCE             DF             SS
Regression          1           8,500
Residual Error     17          11,500
Total              18
```

a). Undersök om INK är en signifikant förklarande variabel. Ställ upp hypoteser och gör sedan en hypotesprövning på signifikansnivå 5% ($\alpha = 0,05$). Vilken blir din slutsats? (10 poäng)

b). Beräkna R-sq och tolka den i ord. (5 poäng)

c). Beräkna residualvariansen. (5 poäng)

Uppgift 2: (20 poäng)
Vid en hushållsundersökning i Grönköping tog man bl.a. reda på kostnaden (utgifter för mat, kläder, nöjen etc.) för tonåringar. Speciellt intresserade man sig för hur kostnaden beror av ungdomarnas ålder. En av de regressionsmodeller som kom till användning var $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, där Y är årskostnaden i hundratal kronor, X_1 är tonåringens ålder i år, och X_2 är en dummyvariabel med värdet 1 för flickor och 0 för pojkar. Då man anpassade denna modell till ett material bestående av 20 tonåringar, 8 flickor och 12 pojkar, fick man bl.a. följande resultat.

```
The regression equation is
Y = 74,11 + 2,02 X1 - 23,74 X2

Predictor     Coef     SE Coef
Constant     74,11      7,58
X1            2,02      0,47
X2          -23,74      1,84

Analysis of Variance

Source             DF        SS
Regression          2     3213,7
Residual Error     17      272,5
Total              19     3486,2
```

a) Pröva på 5 % signifikansnivå om modellen som helhet är signifikant. Ange de hypoteser som du testar. (10 poäng)

b) Hur mycket större är årskostnaden i genomsnitt för en pojke jämfört med en flicka i samma ålder, enligt undersökningen? (10 poäng)

Uppgift 3: (20 poäng)
Följande tabell visar antalet bostadslägenheter (3 rum ock kök) i Grönköping, åren 2013-2017.

2013	2014	2015	2016	2017
10	16	40	100	160

a) Anpassa en exponentiell modell (exponentiell trendfunktion) till tidsserien. Tolka de skattade koefficienterna i ord och uttryckta i termer av de aktuella variablerna och sorterna. (15 poäng)

b) Enligt den anpassade modellen, gör en prognos för antalet bostadslägenheter (3 rok) i Grönköping, år 2018. (5 poäng)

Uppgift 4: (20 poäng)
Följande tabell visar utveckling på det genomsnittliga priset (i miljoner kr.) på bostadsrättslägenheter (3 rum ock kök) i Grönköping, åren 2013-2017. Bostadrättsbubblan har redan spruckit i Grönköping!

2013	2014	2015	2016	2017
4,8	4,9	5,0	4,8	4,6

a) Anpassa en andragradskurva till tidsserien med hjälp av minsta-kvadrat-metoden. (15 poäng)

b) Enligt den anpassade modellen, gör en prognos för det genomsnittliga priset på bostadsrättslägenheter (3 rok) i Grönköping, år 2018. (5 poäng)

Uppgift 5: (20 poäng) oddsen

I Trumps Amerika använder man följande variabler för att beräkna sannolikheten att en utländsk student kommer att bli antagen till forskarutbildning i statistik på de fina universiteten i landet:

Y = Resultat på antagning (blir antagen = 1, blir inte antagen = 0)

X_1 = GRE (betyg på provet "Graduate Record Examination ", skala 0-1000 poäng)

X_2 = TOEFL (betyg på provet "Test of English as a Foreign Language", skala 0-1000 poäng)

X_3 = Utbildning (1 = om universitets studier genomfördes i ett privat universitet, 0 = om universitets studier genomfördes i ett statligt universitet)

Via en logistik regression modell har man fått följande skattningar:

$\hat{Y} = 56{,}4 - 0{,}03X_1 - 0{,}05X_2 - 3X_3$

a). Beräkna sannolikheten att en utländsk student kommer att bli antagen till forskarutbildning i statistik, om studenten har genomfört sina studier i ett privat universitet, har fått 650 GRE-poäng och 650 TOEFL-poäng. (10 poäng)

a). Beräkna sannolikheten att en utländsk student kommer att bli antagen till forskarutbildning i statistik, om studenten har genomfört sina studier i ett statligt universitet, har fått 650 GRE-poäng och 650 TOEFL-poäng. (10 poäng)

Stockholms universitet
Statistiska institutionen
Regressionsanalys och undersökningsmetodik
Vårterminen 2017
Jörgen Säve-Söderbergh

Formelsamling – regressionsanalys

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right)^2$$

Enkel lineär regression

$$b_1 = \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

$$b_0 = \hat{\beta}_0 = \overline{y} - b_1 \overline{x}$$

$$\underbrace{\sum_{i=1}^{n} (y_i - \overline{y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^{n} e_i^2}_{\text{SSE}}.$$

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2 - b_1^2 \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

$$s_e^2 = \text{MSE} = \frac{\text{SSE}}{n-2} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}$$

Konfidensintervall för β_1 ges av

$$b_1 \pm t_{n-2,\frac{\alpha}{2}} s_{b_1},$$

där

$$s_{b_1} = \sqrt{\frac{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}{\sum_{i=1}^{n}(x_i - \overline{x})^2}}.$$

Prediktionsintervall

$$\underbrace{b_0 + b_1 x_{n+1}}_{\hat{y}_{n+1}} \pm t_{n-2,\frac{\alpha}{2}} \sqrt{s_e^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)}.$$

Konfidensintervall förväntat y-värde för ett nytt x-värde

$$\underbrace{b_0 + b_1 x_{n+1}}_{\hat{y}_{n+1}} \pm t_{n-2,\frac{\alpha}{2}} \sqrt{s_e^2 \left(\frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)}.$$

Multipel regression

n st observationer och p förklarande variabler.

Variationsorsak	SS	df	MS	F
Regression	SSR	p	$MSR = \frac{SSR}{p}$	MSR/MSE
Residual	SSE	$n-p-1$	$MSE = \frac{SSE}{(n-p-1)}$	
Totalt	SST	$n-1$		

$$R_{adj}^2 = 1 - \frac{SSE/(n-p-1)}{SST/(n-1)}$$

Normalekvationerna för fallet $\hat{y} = a + b_1 t + b_2 t^2$

$$\begin{aligned}
\sum_{i=1}^{n} y_i &= a \cdot n + b_1 \sum_{i=1}^{n} t_i + b_2 \sum_{i=1}^{n} t_i^2 \\
\sum_{i=1}^{n} y_i t_i &= a \sum_{i=1}^{n} t_i + b_1 \sum_{i=1}^{n} t_i^2 + b_2 \sum_{i=1}^{n} t_i^3 \\
\sum_{i=1}^{n} y_i t_i^2 &= a \sum_{i=1}^{n} t_i^2 + b_1 \sum_{i=1}^{n} t_i^3 + b_2 \sum_{i=1}^{n} t_i^4
\end{aligned}$$

Säsongsrensning med regression

$$a_0 = \overline{y} - b \cdot \overline{t}$$

$$T_t = a_0 + b \cdot t$$

$$\begin{aligned} S_1 &= a - a_0 + c_1 \\ S_2 &= a - a_0 + c_2 \\ S_3 &= a - a_0 + c_3 \\ S_4 &= a - a_0 \end{aligned}$$

Logistisk regression

$$P\left(Y_i = 1 \mid x_i\right) = \frac{\exp\left(\beta_0 + \beta_1 x_i\right)}{1 + \exp\left(\beta_0 + \beta_1 x_i\right)} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$

$$\log\left(\text{odds}\right) = \beta_0 + \beta_1 x_i.$$

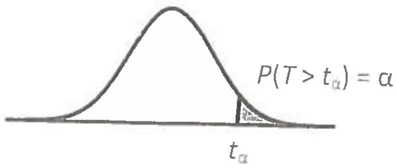
$$\text{odds}\left(D\right) = \frac{P(D)}{P(D \text{ inträffar inte})} = \frac{P(D)}{1 - P(D)}$$

$$P(D) = \frac{\text{odds}\left(D\right)}{1 + \text{odds}\left(D\right)}$$

Konfidensintervall för oddskvoten e^{β_1}: $e^{b_1 \pm z \times s_{b_1}}$

TABELL 3. *t*-fördelningens kvantiler

$T \in t(v)$ där v = antal frihetsgrader.



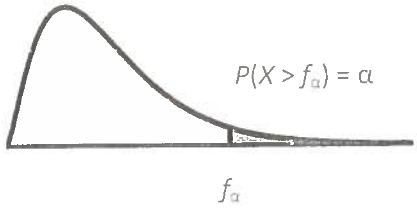
Vilket värde har t_α om $P(T > t_\alpha) = \alpha$ där α är en given sannolikhet. Utnyttja även $P(T \leq -t_\alpha) = P(T > t_\alpha)$.

v	α = 0,1	0,05	0,025	0,010	0,005	0,0025	0,0010	0,0005
1	3,078	6,314	12,706	31,821	63,657	127,321	318,309	636,619
2	1,886	2,920	4,303	6,965	9,925	14,089	22,327	31,599
3	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659
30	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
35	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591
40	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
45	1,301	1,679	2,014	2,412	2,690	2,952	3,281	3,520
50	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496
55	1,297	1,673	2,004	2,396	2,668	2,925	3,245	3,476
60	1,296	1,671	2,000	2,390	2,660	2,915	3,232	3,460
65	1,295	1,669	1,997	2,385	2,654	2,906	3,220	3,447
70	1,294	1,667	1,994	2,381	2,648	2,899	3,211	3,435
75	1,293	1,665	1,992	2,377	2,643	2,892	3,202	3,425

Forts. nästa sida

TABELL 5. *F*-fördelningens kvantiler

$X \in F(v_1,v_2)$ där v_1,v_2 = antal frihetsgrader i täljaren respektive nämnaren. Vilket värde har f_α om $P(X > f_\alpha) = \alpha$ där α är en given sannolikhet.



$\alpha = 0{,}05$

	v_1 = 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
v_2 = 1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,7	245,4	245,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04	2,01	1,99	1,96
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97	1,94	1,92	1,89
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72	1,69	1,67

Forts. nästa sida

(7)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 27/11/17

Sal: Ugglevikssalen

Tenta: Regressionsanalys och undersökningsmetodik

Kurs: Regressions- och tidsserieanalys

ANONYMKOD: REG-TGE-XCF

☒ Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					5 TF
Lär.ant. 20p	20p	20p	20p	20p					

POÄNG	BETYG	Lärarens sign.
100p	A	RC

Fråga 1

SU, STATISTIK

Skrivsal: Uggle Anonymkod: REG-TGE-XCF Blad nr: 1

(1) a) Hypoteser: $H_0 = \beta_1 = 0$ $\alpha = 0{,}05$

$H_1 = \beta_1 \neq 0$ (tvåsidigt)

$n = 19$

Testvariabel = $t = \dfrac{b - \beta_0}{S_b} \sim t(n-2)$ om Ho är sann

Beslutsregel = Förkasta Ho då

$|t_{obs}| > t_{17/0{,}025} = 2{,}110$ R

Observation: $\dfrac{0{,}0588 - 0}{0{,}0246} = 2{,}39$ R

Slutsats: $t_{obs} = 2{,}39 > 2{,}110 = t_{17/0{,}025}$

Vi förkastar Ho på sign. nivå 5% vilket betyder att variabeln INK är en signifikant förklarande variabel. R

b) $R^2 = \dfrac{SSR}{SST}$ SSR = 8,5 (från anova-tablå)

SST = SSR + SSE = 8,5 + 11,5 = 20

$R^2 = \dfrac{8{,}5}{20} = 0{,}425$ R

Svar: $R^2 = 0{,}425$ vilket betyder att 42,5% av variationen i y (konsumtion) förklaras av vår förklarande variabel INK R

→

c) Residualvariansen = $MSE = \frac{SSE}{n-2}$

$SSE = 11{,}5$ $n-2 = 17$

$S_e^2 = \frac{11{,}5}{17} = 0{,}6764705882$

Svar: Residualvariansen (S_e^2) är ca 0,67647

R

Fråga 2

SU, STATISTIK

Skrivsal: Uggle Anonymkod: REG-TGE-XCF Blad nr: 2

2) 20p

(2) a) Hypoteser: $H_0 = \beta_1 = \beta_2 = 0$ $\alpha = 0{,}05$

$n = 20$ $H_1 =$ Minst en $\beta_i \neq 0$

Testvariabel: $F = \frac{MSR}{MSE} \sim F(k, n-k-1)$ om Ho är sann

Beslutsregel: Förkasta Ho då $F_{obs} > F_{(2,17)} = 3{,}59$

Observation: $MSR = \frac{SSR}{k} = \frac{3213{,}7}{2} = 1606{,}85$

$MSE = \frac{SSE}{n-k-1} = \frac{272{,}5}{17} = 16{,}02941176$

$\frac{MSR}{MSE} = \frac{1606{,}85}{16{,}02941176} = 100{,}24138532$ ← F-värdet.

Slutsats: $F_{obs} = 100{,}24 > 3{,}59 = F_{(2,17)}$

Vi förkastar Ho på signnivå 5% vilket betyder att modellen som helhet är signifikant.

R R R R R

b) $y = 74{,}11 + 2{,}02x_1 - 23{,}74x_2$

$y =$ Årskostnad (Hundratal kronor)

$x_1 =$ Ålder (År)

$x_2 = \begin{cases} 1 = \text{Flicka} \\ 0 = \text{Pojke} \end{cases}$

= Information från uppgift

Enligt modellen ser vi att $\beta_2 = -23{,}74$. Om personen är en flicka kommer $-23{,}74$ räknas med i modellen eftersom x_2 då är 1. Om det är en pojke kommer $-23{,}74$ inte räknas med eftersom x_2 då är 0.
Detta betyder att årskostnaden i genomsnitt är 2374 kronor högre för pojkar än för flickor (förutsatt samma ålder).

R

R

Fråga 3

SU, STATISTIK

Skrivsal: Uggle   Anonymkod: REG-TGE-XCF   Blad nr: 3

3) 20p

(3)

År:	t	y_t	$\lg(y)$	$t \cdot \lg(y)$	t^2
2013	-2	10	1	-2	4
2014	-1	16	1,204119983	-1,204119983	1
2015	0	40	1,602059991	0	0
2016	1	100	2	2	1
2017	2	160	2,204119983	4,408239965	4
$\Sigma =$	0	326	8,010299957	3,204119982	10

a) $b' = \dfrac{\Sigma\, t \cdot \lg(y)}{\Sigma\, t^2}$ $a' = \dfrac{\Sigma \lg(y)}{n}$

$b' = \dfrac{3,204119982}{10} = 0,3204119982 \approx 0,32$ R

$a' = \dfrac{8,010299957}{5} = 1,602059991 \approx 1,6$ R

$b = 10^{0,3204119982} = 2,09$ R | $a = 10^{1,602059991} = 40$ R

$\hat{y}_t = 40 \cdot 2,09^t$ R

a = 40 - Då t = 0 är antalet bostadslägenheter 40 stycken (precis som vi fått i tabellen). R

$b_1 = 2,09$ - För varje t, alltså för varje år ökar antalet bostadslägenheter i genomsnitt med ca 109%. R R

b) År 2018 är $t = 3$

$\hat{y}_3 = 40 \cdot 2{,}09^3 = 40 \cdot 9{,}129329 = 365{,}17316$

Svar: År 2018 förväntas det finnas ca 365 bostadslägenheter enligt den anpassade modellen

R

R

Fråga 4

-2·-2·-2·-2

SU, STATISTIK

Skrivsal: Uggle Anonymkod: REG-TGE-XCF Blad nr: 4

4) 20p

(4)

a)

År	t	y_t	t_i^2	t_i^4	$y_i \cdot t_i$	$y_i \cdot t_i^2$
2013	-2	4,8	4	16	-9,6	19,2
2014	-1	4,9	1	1	-4,9	4,9
2015	0	5	0	0	0	0
2016	1	4,8	1	1	4,8	4,8
2017	2	4,6	4	16	9,2	18,4
	0	24,1	10	34	-0,5	47,3

Normalekvationer:

$\Sigma y_i = a \cdot n + \cancel{b_1 \Sigma t_i} + b_2 \Sigma t_i^2$ $\quad 5a + 10b_2 = 24,1 \leftarrow (\cdot -2)$

$\Sigma y_i t_i = \cancel{a \Sigma t_i} + b_1 \Sigma t_i^2 + \cancel{b_2 \Sigma t_i^3}$ $\quad = \quad 10b_1 = -0,5$

$\Sigma y_i t_i^2 = a \Sigma t_i^2 + \cancel{b_1 \Sigma t_i^3} + b_2 \Sigma t_i^4$ $\quad 10a + 34b_2 = 47,3$

$$-10a - 20b_2 = -48,2$$
$$\underbrace{10a + 34b_2 = 47,3}$$
$$14b_2 = -0,9$$

$$\boxed{\begin{aligned} b_1 &= \frac{-0,5}{10} = -0,05 \\ b_2 &= \frac{-0,9}{14} = -0,064 \\ a &= \frac{24,742857\,14}{5} = 4,9485 \approx 4,95 \end{aligned}}$$

R R R

För att räkna ut a:

$5a + 0,6428571429 = 24,1$

$\uparrow$

$10 \cdot b_2$

$5a = 24,74285714$

R R

Modellen blir:

$$\hat{y}_t = 4,95 - 0,05t - 0,064t^2$$

b) År 2018 är $t=3$ R

$\hat{y}_3 = 4{,}95 - 0{,}05 \cdot 3 - 0{,}064 \cdot 3^2 =$ R

$4{,}95 - 0{,}15 - 0{,}576 = 4{,}224$

Svar: Det genomsnittliga priset på bostadslägenheter år 2018 kommer att vara ca 4,2 miljoner kronor

R

Fråga 5

SU, STATISTIK

Skrivsal: Uggle Anonymkod: REG-TGE-XCF Blad nr: 5

5) 20p

(5) $\hat{Y} = 56{,}4 - 0{,}03x_1 - 0{,}05x_2 - 3x_3$

Y = Resultat antagning, (1 = Antagen, 0 = Ej antagen)

X_1 = GRE (Betyg på GRE prov, 0–1000 p)

X_2 = TOEFL (Betyg på TOEFL prov, 0–1000 p)

X_3 = Utbildning, (1 = Privat universitet, 0 = Statligt universitet)

a) $P(Y=1 \mid X_1=650\ X_2=650,\ X_3=1) =$

$$\frac{e^{56{,}4-0{,}03\cdot 650-0{,}05\cdot 650-3\cdot 1}}{1+e^{56{,}4-0{,}03\cdot 650-0{,}05\cdot 650-3\cdot 1}} = \frac{e^{1{,}4}}{1+e^{1{,}4}} = \frac{4{,}055199967}{5{,}055199967} =$$

0,8021 R

R

Svar: Sannolikheten att en utländsk student blir antagen, om studenten gått på privat universitet, fått 650 poäng på GRE-prov samt 650 poäng på TOEFL-prov är ca 0,8.

R

b) $P(Y=1 \mid X_1=650,\ X_2=650,\ X_3=0) =$

$$\frac{e^{56{,}4-0{,}03\cdot 650-0{,}05\cdot 650-3\cdot 0}}{1+e^{56{,}4-0{,}03\cdot 650-0{,}05\cdot 650-3\cdot 0}} = \frac{e^{4{,}4}}{1+e^{4{,}4}} = \frac{81{,}45086866}{82{,}45086866} =$$

R

0,9878 R

Svar: Sannolikheten att en utländsk student blir antagen, om studenten gått på statligt universitet, fått 650 poäng på GRE-prov samt 650 poäng på TOEFL-prov är ca 0,99!

R