STOCKHOLMS UNIVERSITET
Statistiska institutionen
Regressionsanalys och undersökningsmetodik, vårterminen 2017

Tentamen i Regressions- och tidsserieanalys fredagen den 2 juni 2017

Datum	2017-06-02
Tid:	9.00-14.00
Ansvarig Lärare:	Jörgen Säve-Söderbergh
Antal frågor:	5
Maxpoäng:	50
Hjälpmedel:	1) Språklexikon
	2) Kalkylator utan lagrade formler eller lagrad text
Tentamensgenomgång	Fredag 16 juni kl. 10.00 i B705.

Anvisningar

Redovisa dina lösningar i en form som gör det lätt att följa tankegången. Motivera alla väsentliga steg i lösningen. Ange alla antaganden och förutsättningar som du utnyttjar. Skriv endast på en sida av arket. Börja varje ny uppgift på nytt ark.

Lycka Till!

1. Antalet halsbandsflugsnappare som besökte Ottenby fågelstation under april har under åren 2008-2012 visat följande utveckling:

År	y_t
2008	5818
2009	16112
2010	73063
2011	51702
2012	139944

a) Approximera utvecklingen med en exponentialfunktion $\hat{y}_t = a \cdot b^t$. (6 p)

b) Vilken årlig förändringstakt anger den beräknade trenden? (2 p)

c) Hur många halsbandsflugsnappare kommer att besöka Ottenby nästa år i april, förutsatt att modellen är korrekt? (2 p)

2. Efter att ha anpassat en multipel lineär regressionsmodell $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$ till ett datamaterial bestående av $n = 100$ observationer, så fann man följande ANOVA-tablå bland utskrifterna:

	SS	df	MS	F
Regression				
Residual	200			
Total	1000			

a) Komplettera tabellen och fyll i de tomma fälten. (3 p)

b) Testa $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ mot H_1 : minst en av $\beta_i \neq 0$ med hjälp av ANOVA-tablån på 5% signifikansnivå. (7 p)

3. En börsanalytiker skattade följande samband för ett livsmedelsföretag

Utgifter för forskning = 7.83+0.026 Intäkter från försäljning+15.61 Rykten,

där utgifterna för forskning mäts i miljoner kronor, liksom intäkterna från försäljning. Rykten är en dummyvariabel som antar värdet 1 om många rykten florerar angående vad konkurrenterna gör och 0 vid stiltje på ryktessidan. Analytikern har en subtil metod för att skilja på *många* och *stiltje* som vi inte behöver bekymra oss om för närvarande.

a) Beräkna en prognos för utgifterna för forskning då intäkterna från försäljningen är tio miljoner kronor och ryktena kring vad konkurrenterna gör är många. (5 p)

b) Beräkna en prognos för utgifterna för forskning då intäkterna från försäljningen är tio miljoner kronor och det inte verkar hända något på konkurrenternas anläggningar rörande forskning. (5 p)

4. Personalutvecklingen på ett företag som utvecklar dataspel har de senaste sju åren haft följande utveckling

Årtal	2006	2007	2008	2009	2010	2011	2012
y_t	9	4	2	1	4	8	17

Anpassa en andragradskurva med hjälp av minsta-kvadrat-metoden och beräkna en prognos för antalet anställda 2013 om utvecklingen antas följa kurvan. (10 p)

5. a) Vad ska sannolikheten för en händelse A vara för att oddset för A ska vara lika med ett? (8 p)

b) Om oddset för en händelse A är tio, vad är sannolikheten för A? (2 p)

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	$n-1$		

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

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

$$P(Y_i = 1 \mid x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_i)}}$$
$$\log(\text{odds}) = \beta_0 + \beta_1 x_i.$$
$$\text{odds}(D) = \frac{P(D)}{P(D \text{ inträffar inte})} = \frac{P(D)}{1 - P(D)}$$
$$P(D) = \frac{\text{odds}(D)}{1 + \text{odds}(D)}$$

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

Formelsamling regressionsanalys

Enkel linjär regression

$$\hat{\beta}_1 = \frac{SSXY}{SSX} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{n\sum_{i=1}^{n}(x_i y_i) - \sum_{i=1}^{n}(x_i)\sum_{i=1}^{n}(y_i)}{n\sum_{i=1}^{n}(x_i^2) - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n}(x_i^2) - n(\bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

SSY=SSR+SSE

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$r = \frac{SSXY}{\sqrt{SSX \times SSY}} = \frac{S_X}{S_Y}\hat{\beta}_1$$

$$S_{Y|X}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$S_{\hat{\beta}_1} = \frac{S_{Y|X}}{S_X\sqrt{n-1}} \qquad S_{\hat{\beta}_0} = S_{Y|X}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)S_X^2}}$$

$$\hat{Y}_{X_0} \pm t_{n-2,1-\alpha/2} S_{Y|X}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)S_X^2}} \qquad \hat{Y}_{X_0} \pm t_{n-2,1-\alpha/2} S_{Y|X}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)S_X^2}}$$

$$b_1 \pm t_{n-2,1-\alpha/2} S_{b_1}$$

Multipel regression

$$F = \frac{MSR}{MSE} = \frac{(SSY - SSE)/k}{SSE/(n-k-1)}$$

$$b_j \pm t_{n-k-1,1-\alpha/2} S_{b_j}$$

$$R_{Y|X_1,\ldots,X_k} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}}_i)^2}}$$

$$R_{Y|X_1,\ldots,X_k}^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{SSY - SSE}{SSY}$$

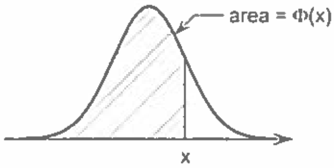
Korrelation

$$Z = \frac{\frac{1}{2}ln\left[\frac{(1+r)}{(1-r)}\right] - \frac{1}{2}ln\left[\frac{(1+\rho_0)}{(1-\rho_0)}\right]}{1/\sqrt{n-3}} \sim N(0,1)$$

Tabeller

Tabell 1. Standardiserad normalfördelning

$\Phi(x) = P(X \leq x)$ där $X \in N(0, 1)$
För negativa värden, utnyttja att $\Phi(x) = 1 - \Phi(-x)$

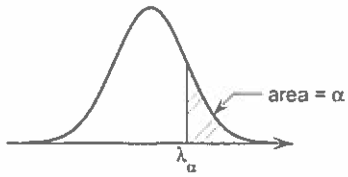


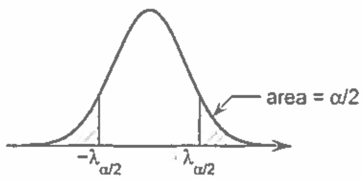
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865									
3.1	.99903									
3.2	.99931									
3.3	.99952									
3.4	.99966									
3.5	.99977									
3.6	.99984									
3.7	.99989									
3.8	.99993									
3.9	.99995									
4.0	.99997									

Tabell 2. Normalfördelningens kvantiler

$P(X > \lambda_\alpha) = \alpha$ där $X \in N(0, 1)$

α	λ_α	α	λ_α
0.1	1.2816	0.001	3.0902
0.05	1.6449	0.0005	3.2905
0.025	1.9600	0.0001	3.7190
0.01	2.3263	0.00005	3.8906
0.005	2.5758	0.00001	4.2649





Tabell 2. *F*-fördelningens kvantiler

$X \in F(\nu_1,\nu_2)$ där ν_1,ν_2 = antal frihetsgrader i täljaren respektive nämnaren. Vilket värde har f_α

om $P(X > f_\alpha) = \alpha$ där α är en given sannolikhet.

α = 0,05

	ν_1 = 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ν_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

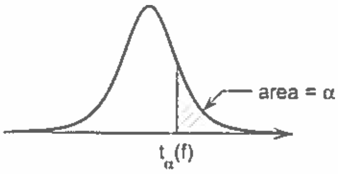
Tabell 2 forts. *F*-fördelningens kvantiler

$\alpha = 0,05$

	$\nu_1 =$ 16	17	18	19	20	25	30	35	40	50	60	70	80	100	∞
$\nu_2 =$ 1	246,5	246,9	247,3	247,7	248,0	249,3	250,1	250,7	251,1	251,8	252,2	252,5	252,7	253,0	254,3
2	19,43	19,44	19,44	19,44	19,45	19,46	19,46	19,47	19,47	19,48	19,48	19,48	19,48	19,49	19,50
3	8,69	8,68	8,67	8,67	8,66	8,63	8,62	8,60	8,59	8,58	8,57	8,57	8,56	8,55	8,53
4	5,84	5,83	5,82	5,81	5,80	5,77	5,75	5,73	5,72	5,70	5,69	5,68	5,67	5,66	5,63
5	4,60	4,59	4,58	4,57	4,56	4,52	4,50	4,48	4,46	4,44	4,43	4,42	4,41	4,41	4,37
6	3,92	3,91	3,90	3,88	3,87	3,83	3,81	3,79	3,77	3,75	3,74	3,73	3,72	3,71	3,67
7	3,49	3,48	3,47	3,46	3,44	3,40	3,38	3,36	3,34	3,32	3,30	3,29	3,29	3,27	3,23
8	3,20	3,19	3,17	3,16	3,15	3,11	3,08	3,06	3,04	3,02	3,01	2,99	2,99	2,97	2,93
9	2,99	2,97	2,96	2,95	2,94	2,89	2,86	2,84	2,83	2,80	2,79	2,78	2,77	2,76	2,71
10	2,83	2,81	2,80	2,79	2,77	2,73	2,70	2,68	2,66	2,64	2,62	2,61	2,60	2,59	2,54
11	2,70	2,69	2,67	2,66	2,65	2,60	2,57	2,55	2,53	2,51	2,49	2,48	2,47	2,46	2,40
12	2,60	2,58	2,57	2,56	2,54	2,50	2,47	2,44	2,43	2,40	2,38	2,37	2,36	2,35	2,30
13	2,51	2,50	2,48	2,47	2,46	2,41	2,38	2,36	2,34	2,31	2,30	2,28	2,27	2,26	2,21
14	2,44	2,43	2,41	2,40	2,39	2,34	2,31	2,28	2,27	2,24	2,22	2,21	2,20	2,19	2,13
15	2,38	2,37	2,35	2,34	2,33	2,28	2,25	2,22	2,20	2,18	2,16	2,15	2,14	2,12	2,07
16	2,33	2,32	2,30	2,29	2,28	2,23	2,19	2,17	2,15	2,12	2,11	2,09	2,08	2,07	2,01
17	2,29	2,27	2,26	2,24	2,23	2,18	2,15	2,12	2,10	2,08	2,06	2,05	2,03	2,02	1,96
18	2,25	2,23	2,22	2,20	2,19	2,14	2,11	2,08	2,06	2,04	2,02	2,00	1,99	1,98	1,92
19	2,21	2,20	2,18	2,17	2,16	2,11	2,07	2,05	2,03	2,00	1,98	1,97	1,96	1,94	1,88
20	2,18	2,17	2,15	2,14	2,12	2,07	2,04	2,01	1,99	1,97	1,95	1,93	1,92	1,91	1,84
25	2,07	2,05	2,04	2,02	2,01	1,96	1,92	1,89	1,87	1,84	1,82	1,81	1,80	1,78	1,71
30	1,99	1,98	1,96	1,95	1,93	1,88	1,84	1,81	1,79	1,76	1,74	1,72	1,71	1,70	1,62
35	1,94	1,92	1,91	1,89	1,88	1,82	1,79	1,76	1,74	1,70	1,68	1,66	1,65	1,63	1,56
40	1,90	1,89	1,87	1,85	1,84	1,78	1,74	1,72	1,69	1,66	1,64	1,62	1,61	1,59	1,51
45	1,87	1,86	1,84	1,82	1,81	1,75	1,71	1,68	1,66	1,63	1,60	1,59	1,57	1,55	1,47
50	1,85	1,83	1,81	1,80	1,78	1,73	1,69	1,66	1,63	1,60	1,58	1,56	1,54	1,52	1,44
60	1,82	1,80	1,78	1,76	1,75	1,69	1,65	1,62	1,59	1,56	1,53	1,52	1,50	1,48	1,39
70	1,79	1,77	1,75	1,74	1,72	1,66	1,62	1,59	1,57	1,53	1,50	1,49	1,47	1,45	1,35
80	1,77	1,75	1,73	1,72	1,70	1,64	1,60	1,57	1,54	1,51	1,48	1,46	1,45	1,43	1,32
100	1,75	1,73	1,71	1,69	1,68	1,62	1,57	1,54	1,52	1,48	1,45	1,43	1,41	1,39	1,28
∞	1,64	1,62	1,60	1,59	1,57	1,51	1,46	1,42	1,39	1,35	1,32	1,29	1,27	1,24	1,00

Tabell 3. *t*-**fördelningen**

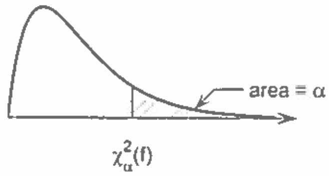
$P(X > t_\alpha(f)) = \alpha$ där $X \in t(f)$



f	α 0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.08	6.31	12.71	31.82	63.66	318.31	636.62
2	1.89	2.92	4.30	6.96	9.92	22.33	31.60
3	1.64	2.35	3.18	4.54	5.84	10.21	12.92
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61
5	1.48	2.02	2.57	3.36	4.03	5.89	6.87
6	1.44	1.94	2.45	3.14	3.71	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	4.02	4.44
12	1.36	1.78	2.18	2.68	3.05	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.85	4.22
14	1.35	1.76	2.14	2.62	2.98	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.69	4.01
17	1.33	1.74	2.11	2.57	2.90	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.50	3.79
23	1.32	1.71	2.07	2.50	2.81	3.48	3.77
24	1.32	1.71	2.06	2.49	2.80	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.45	3.73
26	1.31	1.71	2.06	2.48	2.78	3.43	3.71
27	1.31	1.70	2.05	2.47	2.77	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65
40	1.30	1.68	2.02	2.42	2.70	3.31	3.55
60	1.30	1.67	2.00	2.39	2.66	3.23	3.46
120	1.29	1.66	1.98	2.36	2.62	3.16	3.37
∞	1.28	1.64	1.96	2.33	2.58	3.09	3.29

Tabell 4. χ^2**-fördelningen**

$P(X > \chi^2_\alpha(f)) = \alpha$ där $X \in \chi^2(f)$



f	α 0.9995	0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.001	0.0005
1	0.00	0.00	0.00	0.00	0.00	0.00	3.84	5.02	6.63	7.88	10.83	12.12
2	0.00	0.00	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60	13.82	15.20
3	0.02	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27	17.73
4	0.06	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47	20.00
5	0.16	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52	22.11
6	0.30	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46	24.10
7	0.48	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32	26.02
8	0.71	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12	27.87
9	0.97	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88	29.67
10	1.26	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59	31.42
11	1.59	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26	33.14
12	1.93	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91	34.82
13	2.31	2.62	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82	34.53	36.48
14	2.70	3.04	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32	36.12	38.11
15	3.11	3.48	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80	37.70	39.72
16	3.54	3.94	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27	39.25	41.31
17	3.98	4.42	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72	40.79	42.88
18	4.44	4.90	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16	42.31	44.43
19	4.91	5.41	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58	43.82	45.97
20	5.40	5.92	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00	45.31	47.50
21	5.90	6.45	8.03	8.90	10.28	11.59	32.67	35.48	38.93	41.40	46.80	49.01
22	6.40	6.98	8.64	9.54	10.98	12.34	33.92	36.78	40.29	42.80	48.27	50.51
23	6.92	7.53	9.26	10.20	11.69	13.09	35.17	38.08	41.64	44.18	49.73	52.00
24	7.45	8.08	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56	51.18	53.48
25	7.99	8.65	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93	52.62	54.95
26	8.54	9.22	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29	54.05	56.41
27	9.09	9.80	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.64	55.48	57.86
28	9.66	10.39	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99	56.89	59.30
29	10.23	10.99	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34	58.30	60.73
30	10.80	11.59	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67	59.70	62.16
40	16.91	17.92	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77	73.40	76.09
50	23.46	24.67	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49	86.66	89.56
60	30.34	31.74	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95	99.61	102.69
70	37.47	39.04	43.28	45.44	48.76	51.74	90.53	95.02	100.43	104.21	112.32	115.58
80	44.79	46.52	51.17	53.54	57.15	60.39	101.88	106.63	112.33	116.32	124.84	128.26
90	52.28	54.16	59.20	61.75	65.65	69.13	113.15	118.14	124.12	128.30	137.21	140.78
100	59.90	61.92	67.33	70.06	74.22	77.93	124.34	129.56	135.81	140.17	149.45	153.17

$\sum \log(y) = 22.6951$

$\sum t_i^2 = 10$

$\sum t_i \log(y_i) = 3.2687$

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jörgen Säve-Söderbergh

Lösningsförslag till skriftlig tentamen i Regressionsanalys, Regressionsanalys och undersökningsmetodik, den 2 juni 2017

6

1. a) Vi sätter $t = \text{År} - 2010$ för att uppnå att $\sum_{i=1}^{n} t_i = 0$. Genom tabell eller på annat vis finner vi att $\sum_{i=1}^{n} \log(y) = 22.6951$, $\sum_{i=1}^{n} t_i^2 = 10$ samt $\sum_{i=1}^{n} t_i \log(y) = 3.2687$. Därmed blir

$$b' = \frac{\sum_{i=1}^{n} t_i \log(y)}{\sum_{i=1}^{n} t_i^2} = \frac{3.2687}{10} = 0.3269$$

samt

$$a' = \frac{\sum_{i=1}^{n} \log(y)}{n} = \frac{22.6951}{5} = 4.5390.$$

Alltså har vi att

$$a = 10^{a'} = 10^{4.5390} = 34596,$$

samt

$$b = 10^{b'} = 10^{0.3269} = 2.1226.$$

Därmed approximeras trenden med funktionen

$$\hat{y}_t = 34596 \cdot 2.1226^t$$

2 b) En ökning med 112% per år. fel (korrigerat)

2 c) År 2012 motsvarar att $t = 5$, så

$$\hat{y}_t = 34596 \cdot 2.1226^3 = 330849.4$$

3

2. a) Då $SST = SSR + SSE$ har vi att $SSR = SST - SSE = 1000 - 200 = 800$. Då $n = 100$ och antalet förklarnade variabler är $p = 4$, så blir frihetsgraderna $n - 1 = 99$ och $n - p - 1 = 100 - 4 - 1 = 95$. Därmed kan vi beräkna $MSR = SSR/p = 800/4 = 200$, samt $MSE = SSE/(n - p - 1) = 200/95 = 2.11$. Slutligen blir $F = MSR/MSE = 200/2.11 = 94.79$. Alltså blir tabellen:

	SS	df	MS	F
Regression	800	4	200	94.79
Residual	200	95	2.11	
Total	1000	99		

7 b) **Hypoteser:**

$$H_0 \;:\; \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_1 \;:\; \text{Ej samtliga lika med noll}$$

Signifikansnivå: 5%

Testvariabel: $F = MSR/MSE$ är F-fördelad med 4 fg i täljaren och 99 fg i nämnaren, då H_0 är sann.

Beslutsregel: H_0 förkastas om $F_{\text{obs}} > 2.46$

Resultat: $F_{\text{obs}} = 94.79 < 2.46$

Slutsats: H_0 förkastas på 5% signifikansnivå.

3. a) Låt Intäkter från försäljning = 10 och Rykten = 1. Då blir

$$\text{Utgifter för forskning} = 7.83 + 0.026 \cdot 10 + 15.61 = \underline{\underline{23.7}}$$

Analytikerns prognos blir alltså att 23 700 000 kronor kommer att spenderas på forskning.

b) Låt Intäkter från försäljning = 10 och Rykten = 0. Då blir

$$\text{Utgifter för forskning} = 7.83 + 0.026 \cdot 10 = \underline{\underline{8.09}}$$

Analytikerns prognos blir alltså i detta fall att 8 090 000 kronor kommer att spenderas på forskning.

4. Vi har sju observationer vilket betyder att vi kan definiera nya värden för tiden som summerar till noll. Då antar normalekvationerna följande utseende:

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

Följande tabell ger oss de summor vi behöver:

År	t	y	t^2	t^4	$y \cdot t$	$y \cdot t^2$
2006	-3	9	9	81	-27	81
2007	-2	4	4	16	-8	16
2008	-1	2	1	1	-2	2
2009	0	1	0	0	0	0
2010	1	4	1	1	4	4
2011	2	8	4	16	16	32
2012	3	17	9	81	51	153
	0	45	28	196	34	288

Det är alltså ekvationssystemet

$$\begin{aligned} 7a + 28b_2 &= 45 & (1)\\ 28b_1 &= 34 & (2)\\ 28a + 196b_2 &= 288 & (3) \end{aligned}$$

som vi ska lösa. Som vanligt är ekvation 2 lättast att lösa:

$$28b_1 = 34 \Leftrightarrow \boxed{b_1 = \frac{34}{28} = 1.21429}$$

Då återstår

$$\begin{aligned} 7a + 28b_2 &= 45 & (4)\\ 28a + 196b_2 &= 288 & (5) \end{aligned}$$

Multiplicera ekvation 4 med −4 och addera den till ekvation 5 i syfte att eliminera a. Det ger

b_1

$$84b_2 = 108 \Leftrightarrow b_2 = \boxed{\frac{108}{84} = 1.28571}$$

Till sist använder vi ekvation 4 för att finna a. Substituera b_2, vilket ger

$$7a + 28 \cdot \frac{108}{84} = 45 \Leftrightarrow a = \frac{1}{7}\left(45 - 28 \cdot \frac{108}{84}\right) = \boxed{1.28571 = \frac{9}{7}}\ a$$

Alltså har vi den skattade modellen som

$$\hat{y} = 1.28571 + 1.21429t + 1.28571t^2.$$

eller som

$$\hat{y} = 1.28571 + 1.21429\,(\text{årtal} - 2009) + 1.28571 \times (\text{årtal} - 2009)^2$$

År 2013 är ekvivalent med att $t = 4$, vilket vi substituerar in i modellen och erhåller

$$\hat{y} = 1.28571 + 1.21429 \cdot 4 + 1.28571 \cdot 4^2 = 26.71423$$

Vår prognos för 2013 blir alltså 27 anställda.

5. **a)** Oddset för en händelse A ges ju av $O = p/(1-p)$ (skrivet med förenklad notation). Låter vi $O = 1$ har vi ekvationen

$$1 = \frac{p}{1-p} \Leftrightarrow 1 - p = p \Leftrightarrow 2p = 1 \Leftrightarrow p = \frac{1}{2}$$

b) Om $O = 10$ blir $p = O/(O+1) = 10/11 = 0.91$.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 2/6-2017

Sal: Ugglevikssalen

Tenta: Regressions- och tidsserieanalys

Kurs: Regressionsanalys och undersökningsanalys

ANONYMKOD: RTO-0021

☒ 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
Lär.ant. 10	10	10	10	10					

1 ps

Mycket fina svar! Bra!

POÄNG	BETYG	Lärarens sign.
50	A	JSS

SU, STATISTIK

Skrivsal: UG Anonymkod: RT0-0021 Blad nr: 1

Fråga 1

a) Approximering av utveckling med $\hat{y}_t = a \cdot b^t$

- Vi skriver om $\hat{y}_t = a \cdot b^t$ med hjälp av naturliga logaritmer för att kunna använda metoden för vanlig enkel linjär ekvation.

$$\ln y_t = \ln (a \cdot b^t)$$

$$\Leftrightarrow \ln y_t = \ln a + \ln b^t$$

$$\Leftrightarrow \underbrace{\ln y_t}_{\hat{y}_t'} = \underbrace{\ln a}_{\beta_0'} + \underbrace{t}_{X'} \cdot \underbrace{\ln b}_{\beta_1'}$$

Alltså kan vi skriva: $\hat{y}_t' = \beta_0' + \beta_1' X$

t	år	y_t	$\ln y_t$	$t \cdot \ln y_t$	t^2
1	2008	5818	8,668712	8,668712	1
2	2009	16 112	9,687320	19,374640	4
3	2010	73 063	11,199077	33,597231	9
4	2011	51 702	10,853252	43,413008	16
5	2012	139 944	11,848998	59,244990	25
Σ 15		286 639	52,257359	164,298573	55

V.G.V. →

$$b_1 = \frac{\sum_{i=1}^{n} t_i \ln y_i - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)\left(\sum_{i=1}^{n} \ln y_i\right)}{\sum_{i=1}^{n} t_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)^2}$$

$$b_1 = \frac{164{,}298573 - \frac{1}{5}(15)(52{,}257359)}{55 - \frac{1}{5}(15)^2}$$

$$b_1 = 0{,}7526496$$

- Alltså blir: $\ln b = 0{,}7526496$

$$\Rightarrow \quad b = e^{0{,}7526496} \approx \underline{\underline{2{,}1226}} \quad R$$

$$b_0 = \bar{y} - b_1 \bar{x} = \left(\frac{52{,}257359}{5}\right) - 0{,}7526496\left(\frac{15}{5}\right)$$

$$b_0 = 8{,}193523$$

- Alltså blir $\ln a = 8{,}193523$

$$\Rightarrow \quad a = e^{8{,}193523} \approx \underline{\underline{3617{,}44}}$$

$\boxed{\text{Modellen blir då: } \hat{y} = 3617{,}44 \cdot 2{,}1226^{t}}$ R

b) En exponentiell förändring av $2{,}1226^t$ för varje år.

c) $\hat{y}_6 = 3617{,}44 \cdot 2{,}1226^6 \approx 330\ 834$ fåglar nästa år. R

SU, STATISTIK

Skrivsal: UG Anonymkod: RT0-0021 Blad nr: 2

Fråga 2 $n = 100$ observationer

a) $p = 4$ förklarande variabler

	SS	df	MS	F
Regression	800	4	200	95
Residual	200	95	2,105	
Total:	1000	99		

R

Följande beräkningar görs:

(A) $SSR = SST - SSE = 1000 - 200 = 800$

(B) $p = 4$

(C) $n - p - 1 = 100 - 4 - 1 = 95$

(D) $MSR = \frac{SSR}{p} = \frac{800}{4} = 200$

(E) $MSE = \frac{SSE}{n-p-1} = \frac{200}{100-4-1} = \frac{200}{95} \approx 2,105$

(F) $F = \frac{MSR}{MSE} = \frac{200}{\left(\frac{200}{95}\right)} = 95$

b) Hypotestest

Hypoteser: $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

$H_1:$ minst en av $\beta_i \neq 0$

Signifikansnivå: 5 % signifikansnivå.

V.G.V. →

Fråga 2 (fortsättning)

Testvariabel: $F = \frac{MSR}{MSE}$

F-fördelning, med 4-frihetsgrader i täljaren samt 95-frihetsgrader i nämnaren.

$F_{4,95} = 2,4675$ R (approximativt, då $F_{4,80} = 2,49$ samt $F_{4,100} = 2,46$ enl. tabellsamling. $\frac{\left(\frac{2,46+2,49}{2}\right) + 2,46}{2}$)

Beslutsregel:

Förkasta H_0, då $F_{obs} > F_{krit} = 2,4675$ R

Resultat:

Från ANOVA-tablån kan vi avläsa $F = 95$. Alltså, F påvisar kraftig signifikans. R

Slutsats:

Vi förkastar H_0, då $F_{obs} > F_{krit}$ $(95 > 2,4675)$. Slutsatsen är att minst en av $\beta_i \neq 0$, och det innebär att minst en av de förklarande variablerna bidrar till modellen. R

SU, STATISTIK

Skrivsal: UG Anonymkod: RTO-0021 Blad nr: 3

Fråga 3 Givet följande modell:

Utg. för forskning = 7,83 + 0,026 · Int. försj. + 15,61 · Rykten

a) Prognos 1 för utgifterna för forskning

Försäljning = 10 miljoner kronor
Rykten = 1

Utg. för forskning = 7,83 + 0,026 · 10 + 15,61 · 1
= 23,7 miljoner kronor

Alltså 23 700 000 kronor i utgifter för forskning. R

b) Prognos 2 för utgifterna för forskning

Försäljning = 10 miljoner kronor
Rykten = 0

Utg. för forskning = 7,83 + 0,026 · 10 + 15,61 · 0

= 8,09 miljoner kronor

Alltså 8 090 000 kronor i utgifter för forskning, då inga rykten kring vad konkurrenterna gör, finns. R

SU, STATISTIK

Skrivsal: UG Anonymkod: RTO-0021 Blad nr: 4

Fråga 4

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

Normalekvationerna antar följande utseende:

$$(1) \quad \sum_{i=1}^{n} y_i = a \cdot n + b_2 \sum_{i=1}^{n} t_i^2$$

$$(2) \quad \sum_{i=1}^{n} y_i t_i = b_1 \sum_{i=1}^{n} t_i^2$$

$$(3) \quad \sum_{i=1}^{n} y_i t_i^2 = a \sum_{i=1}^{n} t_i^2 + b_2 \sum_{i=1}^{n} t_i^4$$

Vi ställer upp följande beräkningstabell:

År	t_i	y_i	$y_i t_i$	t_i^2	t_i^4	$y_i t_i^2$
2006	-3	9	-27	9	81	81
2007	-2	4	-8	4	16	16
2008	-1	2	-2	1	1	2
2009	0	1	0	0	0	0
2010	1	4	4	1	1	4
2011	2	8	16	4	16	32
2012	3	17	51	9	81	153
$\sum$	0	45	34	28	196	288

V.G.V. →

Fråga 4 (fortsättning)

- Vi sätter nu in värdena från beräkningstabellen i normalekvationerna.

R $\begin{cases} (1) \quad 45 = 7a + 28b_2 & \boxed{-4} \\ (2) \quad 34 = 28b_1 \\ (3) \quad 288 = 28a + 196b_2 \end{cases}$

- Vi multiplicerar ekvation (1) med -4 och adderar till ekvation (3).

(1) $45 = 7a + 28b_2$

(2) $34 = 28b_1$

(3) $108 = 84b_2$

- Vi kan nu lösa ut b_1 och b_2 från (2) och (3).

(2) $b_1 = \frac{34}{28} \approx \underline{1,21429}$ R (3) $b_2 = \frac{108}{84} \approx \underline{1,28571}$ R

- Från ekvation (1): $a = \frac{45 - 28\left(\frac{108}{84}\right)}{7} \approx \underline{1,28571}$ R

- Alltså blir den skattade modellen:

$$\hat{y} = 1,28571 + 1,21429 \cdot t + 1,28571 \cdot t^2$$

- Antalet anställda 2013 $(t = 4)$ blir:

$$\hat{y}_4 = 1,28571 + 1,21429 \cdot 4 + 1,28571 \cdot 4^2 \approx \underline{\underline{27 \text{ st}}}$$

Alltså 27 st anställda 2013 om modellen antas stämma.

R

SU, STATISTIK

Skrivsal: UG Anonymkod: RTo-0021 Blad nr: 5

Fråga 5

Vi har att: $\text{odds}(A) = \dfrac{P(A)}{1-P(A)}$

a) Vi söker $P(A)$ då $\text{odds}(A) = 1$

Detta är möjligt då $P(A) = 0,5$.

$P(A) = 0,5 \Rightarrow \dfrac{0,5}{1-0,5} = \dfrac{0,5}{0,5} = 1$.

- Alltså är sannolikheten för en händelse A, $P(A) = 0,5$, då oddset för A är ett. R

b) Givet $\text{odds}(A) = 10$, då blir:

$10 = \dfrac{P(A)}{1-P(A)} \Leftrightarrow P(A) = 10(1-P(A))$

$\Leftrightarrow P(A) = 10 - 10P(A)$

$\Leftrightarrow 11P(A) = 10$

$\Leftrightarrow P(A) = \dfrac{10}{11} \approx 0,9091$

- Alltså är sannolikheten för A cirka $0,91$ eller $\left(\frac{10}{11}\right)$, om oddset för A är tio.

R

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 2/6-2017

Sal: Ugglevikssalen

Tenta: Regressions- och tidsserieanalys

Kurs: Regressionsanalys och undersökningsanalys

ANONYMKOD: RTU-0020

☒ 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
Lär.ant. 10	10	10	10	10					

JRS

Bra!

POÄNG	BETYG	Lärarens sign.
50	A	JSS

SU, STATISTIK

Skrivsal: Vagdevilessalen Anonymkod: RTO-0020 Blad nr: 1

1.

År	t	Y_t	$\log Y_t$	$t \log Y_t$	t^2
2008	-2	5818	3,764	-7,528	4
2009	-1	16112	4,207	-4,207	1
2010	0	73063	4,863	0	0
2011	1	51702	4,713	4,713	1
2012	2	139944	5,145	10,29	4
Σ	0	286639	22,692	3,268	10

a) $Y_t = a \cdot b^t$ (1)

* (1) kan skrivas om till en linjär funktion genom att uttrycken logaritmeras.

$\rightarrow \log Y_t = \log a + \log b \cdot t$ (2)

* Därefter kan vi förenkla ytterligare

$\rightarrow \log Y_t = a' + b' t$

$$* \; b_1 = \frac{\Sigma t \log Y - \frac{1}{n}\cdot \underbrace{(\Sigma t) \cdot \Sigma(\log Y_t)}_{0}}{\Sigma t^2 - \underbrace{\frac{1}{n}\cdot(\Sigma t)^2}_{0}} = \frac{\Sigma t \log Y}{\Sigma t^2}$$

$$= \frac{3,268}{10} = 0,3268 \text{ R}$$

$$b' = \log b = 0,3268 \longrightarrow b = 10^{0,3268} \approx 2,12 \text{ R}$$

vänd!

$a' = \overline{\log y} - b_1 \bar{t} = \frac{\Sigma \log y}{n} = \frac{22{,}692}{5} = 4{,}5384$ R

$a = 10^{4{,}5384} \approx 34546{,}2$ R

Svar: Den slutliga exponentialfunktionen approximeras till $y_t = 34546{,}2 \cdot 2{,}12^t$ R

b) Ett b-värde på 1 indikerar 0% förändringstakt.

Således bör 2,12 indikera en förändringstakt i procent på $(2{,}12 - 1) \cdot 100 = 112\%$.

Svar: 112%! R   Vilken tillväxt, va?!

c) Sätt in t = 3 i formeln!

$y_t = 34546{,}2 \cdot 2{,}12^3 \approx 329160{,}61 \approx 329161$ R

SU, STATISTIK

Skrivsal: Vegahuiksalen RTO-0020 Anonymkod: Blad nr: 2

2.

a)

	SS	df	MS	F
Regression	800	4	200	95
Residual	200	95	2,105	
Total	1000	99		

R

b) Hypotes: $H_0: B_1 = B_2 = B_3 = B_4 = 0$

$H_1: B_i \neq 0$

Signifikansnivå: 5%

Testvariabel: F_{obs}

Beslutsregel: Förkasta H_0 om $F_{obs} > F_{krit}$

* F_{krit} utläses ur tabell

$F_{krit} \approx 2,46 \longrightarrow F_{obs} = 95 > 2,46$ R

Slutsats: H_0 kan förkastas på 5% signifikansnivå.

R

SU, STATISTIK

Skrivsal: Uppfeniksalen Anonymkod: RTO-0020 Blad nr: 3

3. $Y = 7{,}83 + 0{,}026X_1 + 15{,}61X_2$

Y = Utgifter för forskning (miljoner)
X_1 = Intäkter från försäljning (miljoner)
X_2 = Rykten (Dummy som antar 0 eller 1)
$X_2 = 0 \rightarrow$ Inga rykten $X_2 = 1 \rightarrow$ Många rykten

a) * Beräkna Y givet följande:
$X_1 = 10$
$X_2 = 1$

$\rightarrow Y = 7{,}83 + 0{,}026 \cdot 10 + 15{,}61 = 23{,}7$

Svar: 23,7 miljoner blir utgifterna för forskning då intäkter är 10 miljoner och rykten florerar R

b) * Beräkna Y givet följande:
$X_1 = 10$
$X_2 = 0$

$Y = 7{,}83 + 0{,}026 \cdot 10 + 15{,}61 \cdot 0 = 8{,}09$

Svar: Vid 10 miljoner i intäkter och stiltje på ryktessidan hamnar utgifterna på 8,09 miljoner kronor R

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: RTO-0020 Blad nr: 4

4.

Årtal	t	y_t	t^2	t^4	$y_t t$	$y_t t^2$
2006	-3	9	9	81	-27	81
2007	-2	4	4	16	-8	16
2008	-1	2	1	1	-2	2
2009	0	1	0	0	0	0
2010	1	4	1	1	4	4
2011	2	8	4	16	16	32
2012	3	17	9	81	51	153
Σ	0	45	28	196	34	288

* För att hitta andragradskurvan kan vi använda oss av normalekvationerna (för fallet $\hat{y}_t = a + b_1 t + b_2 t^2$)
Vi får då följande 3 ekvationer:

(1) $7a + 28b_2 = 45$

(2) $28b_1 = 34$

(3) $28a + 196b_2 = 288$

* Därefter löser vi ut respektive variabel

- Multiplicera (1) med $-4 \rightarrow -28a - 112b_2 = -180$
- Addera (1) och (2) $\rightarrow 28a - 28a + 196b_2 - 112b_2 = 288 - 180$
  $\rightarrow 84b_2 = 108 \rightarrow b_2 \approx 1{,}286$

- $b_1 = \frac{34}{28} = 1{,}214$ R

- $a = \frac{45 - 28 \cdot 1{,}286}{7} \approx 1{,}285$ R

Svar på nästa sida!

* Slutliga ekvationen blir: $Y_t = 1,285 + 1,214t + 1,286t^2$

Prognos för år 2013:

Sätt in $t = 4$

$Y_t = 1,285 + 1,214 \cdot 4 + 1,286 \cdot 4^2 = 26,717$

Svar: Antal anställda 2013 kommer vara ungefär 27 st

SU, STATISTIK

Skrivsal: Ugglevikssalen Anonymkod: RTO-0020 Blad nr: 5

5.

a) $\text{Odds}(A) = \frac{P(A)}{1-P(A)}$

Sätt $\text{Odds}(A) = 1$ och lös ut $P(A)$

$$1 = \frac{P(A)}{1-P(A)}$$

$\rightarrow 1 - P(A) = P(A)$

$1 = 2P(A)$

$\rightarrow P(A) = \frac{1}{2} = 0,5$

← Elegant bevis!

Svar: För att oddset för A ska vara lika med 1 ska sannolikheten för A vara 50%. R

b) Sätt $\text{Odds}(A) = 10$ och lös ut $P(A)$

$$10 = \frac{P(A)}{1-P(A)}$$

$10 - 10P(A) = P(A)$

$\frac{10}{11} = P(A) \approx 0,909$

Svar: Om oddsen för A är 10 är sannolikheten att A inträffar ungefär 90,9%. R