Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 03-12-2018

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).
Återlämning av tentamen: hämtas på studentexpeditionen, plan 7 i B-huset fr.o.m. tisdagen den 18 december. Kolla på vår hemsida studentexpeditionens mottagningstider under terminstid.

Tentamen består av fyra 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: (50 poäng)
Inne i den tropiska djungeln i den lilla republiken Costa Rica pågår en studie om den exotiska paradis fågeln "splendid astrapia" (Astrapia splendidissima). I studien ingår 13 fåglar och man undersöker sambandet mellan Y = vinge längd (i cm.) och X = antal dagar efter kläckning.
Du har tillgång till följande information:

y	1,4	1,5	2,2	2,4	3,1	3,2	3,2	3,9	4,1	4,7	4,5	5,2	5,0
x	3	4	5	6	8	9	10	11	12	14	15	16	17

a). Anpassa regressionslinjen $\hat{y} = a + b\,x$ till materialet. (10 poäng)

b). Tolka verbalt de erhållna koefficienterna. (5 poäng)

c). Beräkna korrelationskoefficienten och tolka denna. (5 poäng)

d). Testa H_0: $\beta = 0$ mot H_1: $\beta \neq 0$ (dels med t-test (5 poäng), dels med F-test (5 poäng)). Använd signifikansnivå 5% ($\alpha = 0{,}05$). (totalt 10 poäng)

e). Testa H_0: $\beta = 0$ mot H_1: $\beta < 0$ (med t-test). Använd signifikansnivå 5% ($\alpha = 0{,}05$). (5 poäng)

f). Beräkna ett 95%-igt konfidensintervall för β. (5 poäng)

g). Beräkna ett 95%-igt konfidensintervall för det förväntat y-värdet för x = 14. (5 poäng)

h). Beräkna ett 95%-igt prediktionsintervall för det enskilda y-värdet för x = 14. (5 poäng)

Uppgift 2: (20 poäng)
Man vill skatta parametrarna β_0, β_1 och β_2 i regressionsmodellen $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$.
Slumpfelen (ε) antas vara oberoende och normalfördelade med väntevärde 0 och konstant varians σ^2.
Du har tillgång till följande information (datamaterial innehåller 13 observationer):

```
The regression equation is
y = 52,577 + 1,4683 x1 + 0,66225 x2


Predictor      Coef       SE Coef        T
Constant                               23,00
x1                                     12,10
x2                                     14,44

S =          R-Sq =         R-Sq(adj)  =

Analysis of Variance

Source            DF        SS         MS         F
Regression
Residual Error              57,9
Total                     2715,8
```

a). Beräkna värdet på s (5 poäng)
b). Beräkna en skattning av slumpfelsvariansen σ^2. (5 poäng)
c). Beräkna R^2 och förklara vad det erhållna värdet säger. (5 poäng)
d). Testa på 5% signifikansnivå om $\beta_1 < 0$ i modellen. (5 poäng)

Uppgift 3: (10 poäng)
Man har följande uppgifter om Bruttonationalprodukt (BNP i miljarder amerikanska dollar):

År	1970	1980	1990	2000	2010
BNP	4	10	20	35	60

a). Anpassa en exponentiell modell (exponentiell trendfunktion) till tidsserien. (5 poäng)
b). Tolka de skattade koefficienterna i ord och uttryckta i termer av de aktuella variablerna. (5 poäng)

Uppgift 4: (20 poäng)
Man studerat sambandet mellan Y = "död/lever" och X = blodtryck i mm. I ett material om 20 personer har man mätt blodtrycket samt följt dessa fem år framåt i tiden, varefter man har tagit reda på om personen är död eller lever. "död" är kodat 1 och "lever" 0.
Via en logistik regression modell har man fått följande skattningar:

$\hat{Y} = -15{,}59 + 0{,}105\ X$

a). Tolka verbalt koefficienten 0,105. (10 poäng)
b). Ange med hur mycket procent ökar oddset för "död" i genomsnitt om blodtrycket stiger 1 mm. (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

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

Konfidensintervall förväntat y-värde för ett nytt x-värde

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

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

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

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

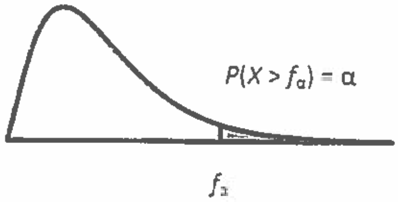
$$\text{odds}(D) = \frac{P(D)}{P(D \text{ inträffar inte})} = \frac{P(D)}{1 - P(D)}$$

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

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

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.



α = 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

22

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 03/12/18

Sal: Laduvikssalen

Tenta: Regressions- och tidsserieanalys

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: 0019-LHJ

☒ 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						5
Lär.ant. 50p	20p	10p	10p						

POÄNG	BETYG	Lärarens sign.
90p	A	RC

SU, STATISTIK

Skrivsal: LA Anonymkod: 0019-LHJ Blad nr: 1

1) 50 p

1 a)

	y	x	x^2	$y \cdot x$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(x-\bar{x})(y-\bar{y})$
	1,4	3	9	4,2	-7	49	-2,0154	14,1078
	1,5	4	16	6	-6	36	-1,9154	11,4924
	2,2	5	25	11	-5	25	-1,2154	6,077
	2,4	6	36	14,4	-4	16	-1,0154	4,0616
	3,1	8	64	24,8	-2	4	-0,3154	0,6308
	3,2	9	81	28,8	-1	1	-0,2154	0,2154
	3,2	10	100	32	0	0	-0,2154	0
	3,9	11	121	42,9	1	1	0,4846	0,4846
	4,1	12	144	49,2	2	4	0,6846	1,3692
	4,7	14	196	65,8	4	16	1,2846	5,1384
	4,5	15	225	67,5	5	25	1,0846	5,423
	5,2	16	256	83,2	6	36	1,7846	10,7076
	5,0	17	289	85	7	49	1,5846	11,0922
Σ	44,4	130	1562	574,8	0	262	≈ 0	70,8

$\hat{y} = a + bx$ $n = 13$ $\bar{x} = \frac{130}{13} = 10$ $\bar{y} = \frac{44,4}{13} = 3,4154$

$b = b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x - \bar{x})^2} = \frac{70,8}{262} = 0,27022$

$a = b_0 = \bar{y} - b_1\bar{x} = 3,4159 - 0,27022 \cdot 10 = 0,7132$

Vår regressionslinje får då formeln

$\hat{y} = \underbrace{0,7132}_{b_0} + \underbrace{0,27022}_{b_1} x_1$

R

R

b) Vår b_0 säger oss egentligen inte så mycket eftersom det är vårat intercept, alltså när $x = 0$. En tolkning är dock att det är genomsnittlig längd vid kläckning.

Vårt b_1 säger oss att i genomsnitt växer arlens vingar med 0,27 cm per dag

R

c

$(y - \bar{y})^2$
4,0617
3,6687
1,4772
1,0310
0,0995
0,0464
0,0464
0,2348
0,4687
1,6502
1,1764
3,1848
2,511

$\Sigma(y - \bar{y})^2 = 19,611$

$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x-\bar{x})^2 \cdot \Sigma(y - \bar{y})^2}} = \frac{70,8}{\sqrt{262 \cdot 19,611}} = 0,9877$

R

Tolkning: Vår oberoende variabel x förklarar förändringen i vår beroende variabel y till 98,77%

R

d) (1) T-test

$H_0: \beta_1 = 0$
$H_A: \beta_1 \neq 0$

Testvariabel $T \ \frac{(b_1 - \beta_1)}{s_{b_1}}$ där $T \sim t(13-2 \text{ df}, \alpha=0{,}05)$ (β_1: 0 under H_0)

Eftersom vi har ett dubbelsidigt test tar vi $\alpha/2$

Beslutsregel: Förkasta H_0 om $|t_{OBS}| > 2{,}201$

R

$$s_{b_1} = \sqrt{\frac{\sum(y-\hat{y})^2 / n-2}{\sum(x-\bar{x})^2}}$$

x	y	$\hat{y}$	SSE $(y-\hat{y})^2$	SSR $(\hat{y}-\bar{y})^2$
3	1,4	1,5239	0,0153	3,5778
4	1,5	1,7941	0,0865	2,6286
5	2,2	2,0643	0,0184	1,8255
6	2,4	2,3345	0,00473	1,1683
8	3,1	2,875	0,0506	0,2920
9	3,2	3,1452	0,003	0,0730
10	3,2	3,4154	0,046	0
11	3,9	3,6856	0,046	0,073
12	4,1	3,9558	0,021	0,2920
14	4,7	4,4963	0,0415	1,1683
15	4,5	4,7665	0,071	1,8255
16	5,2	5,0367	0,0267	2,6287
17	5,0	5,3069	0,0942	3,5778
Σ			Σ 0,5245	Σ 19,1306

R

$$s_{b_1} = \sqrt{\frac{\frac{0{,}5245}{13-2}}{262}} \Rightarrow \sqrt{\frac{0{,}0403}{262}} = 0{,}012$$

$$t = \frac{0{,}27022}{0{,}012} = 22{,}5$$

R

Eftersom vårat observerade t är större än vårt kritisk t för $\alpha=0{,}05$ kan vi förkasta H_0

R

(2) forts nästa sida

SU, STATISTIK

Skrivsal: LA **Anonymkod:** 0019-LHJ **Blad nr:** 2

1d) forts.

(2) F-test

$H_0: \beta_1 = 0$
$H_A: \beta_1 \neq 0$

Testvariabel $F = \frac{MSR}{MSE}$ där $F \sim F_{1,11,\alpha=0{,}05}$

Beslutsregel: Förkasta H_0 om $F_{OBS} > 4{,}84$

$MSR = \frac{SSR}{p} = \frac{19{,}1306}{1}$ $MSE = \frac{SSE}{n-p-1} = \frac{0{,}5245}{13-1-1} = 0{,}04768$

$F = \frac{19{,}1306}{0{,}04768} = 401{,}21$

R

Eftersom vårt $F_{OBS} > 4{,}84$ kan vi förkasta H_0

R

---

e) $H_0: \beta_1 = 0$
$H_A: \beta_1 < 0$

0 under H_0

Testvariabel $T \left(\frac{b_1 - \beta_1}{S_{b_1}}\right)$, enkelsidigt test där $T \sim t_{11df, \alpha=0{,}05}$

Beslutsregel: Förkasta H_0 om $t_{OBS} < -1{,}796$

R

$t_{OBS} = \frac{0{,}27022}{0{,}012} = 22{,}5$

R

Eftersom vårt observerade $t > -1{,}796$ kan vi inte förkasta H_0

β_1 är med 95% signifikans positiv

R

SU, STATISTIK

Skrivsal: LA Anonymkod: 0019-LHJ Blad nr: 3

2) 20p

2 a) ANOVA n=13

Källa	df	SS	MS
Regression	2	2657,9	1328,95
Residual	10	57,9	5,79
Total	12	2715,8	

$s = \sqrt{MSE} = \sqrt{5,79} = 2,406$ R

b) $\sigma^2 = MSE = \frac{57,9}{13-2-1} = \frac{57,9}{10} = 5,79$ R

c) $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{57,9}{2715,8} = 0,9787$

Våra oberoende variabler (x_1, x_2) förklarar förändringen vår beroende variabel y till 97,87% R

d) $H_0: \beta_1 = 0$
$H_1: \beta_1 < 0$

0 under Ho

Testvariabel $T \; \frac{(b_1 - \beta_1)}{s_{b_1}}$ där $T \sim t_{(10 df \; \alpha = 0,05)}$

Enkelsidigt test

Beslutsregel: Förkasta Ho om $t_{obs} < -1,812$ R

$T = 12,10$ R

Eftersom vårat observerade värde på $t > -1,812$ kan vi inte förkasta Ho på $\alpha = 0,05$

Med 95% signifikans är β_1 positiv och skild från 0

R

SU, STATISTIK

Skrivsal: LA **Anonymkod:** 0019-LHJ **Blad nr:** 4

3) 10p

	t	t^2	y	log(y)	log(y)t
1970	-2	4	4	0,602	-1,204
1980	-1	1	10	1	-1
1990	0	0	20	1,301	0
2000	1	1	35	1,544	1,544
2010	2	4	60	1,778	3,556
Σ	0	10	129	6,225	2,896

R

$y = a \cdot b^t \Rightarrow \log(y) = \underbrace{\log(a)}_{a'} \cdot \underbrace{\log(b)}_{b'} t$ $n=5$

$b' = \frac{\Sigma \log(y) t}{\Sigma t^2} = \frac{2{,}896}{10} = 0{,}2896$

$a' = \frac{\Sigma \log(y)}{n} = \frac{6{,}225}{5} = 1{,}245$

R

$a = 10^{1{,}245} = 17{,}5792$

$b = 10^{0{,}2896} = 1{,}948$

~~Vi får då modellen~~

$y = 17{,}5792 \cdot 1{,}948^t$ R (i miljarder US dollar)

b) a ~~är det förväntade värde y antar då x = 0~~

b kan uttryckas som den procentuella förändringen per år

dvs förändringen = 1,948 − 1 = 0,948. I genomsnitt ökar alltså

BNP med i genomsnitt 94,8% per 10 år ~~ALLA~~

R

SU, STATISTIK

Skrivsal: LA Anonymkod: 0019 - LHJ Blad nr: 5

4) 10p

(4) Y = "död/lever" där död = 1, lever = 0
X = blodtryck i mm n = 20

$\hat{Y} = -15,59 + 0,105X$

a) En ökning av blodtrycket med 1 mm ökar vårt värde på y som är vår beroende variabel med 0,105

En ökning av blodtrycket med 10 mm skulle således öka värdet på y med 1,05

F (SE FACIT)

b) För en person med blodtryck = 200 får vi följande sannolikhet att personen dött inom 5 år

$$P = \frac{-15,59 + 0,105 \cdot 200}{1 + (-15,59 + 0,105 \cdot 200)} = 0,84399$$

Odds för att personen dött $= \frac{P(D)}{1 - P(D)} = \frac{0,84399}{1 - 0,84399} = 5,41$

För en person med blodtryck = 201

$$P = \frac{-15,59 + 0,105 \cdot 200}{1 + (-15,59 + 0,105 \cdot 200)} \quad 0,84650$$

$$Odds = \frac{0,84650}{1 - 0,84650} = 5,515$$

Procentuell förändring $= \frac{5,515}{5,41} - 1 = 0,0194$

Eftersom vi har en logistik regression är dock i genomsnitt förändringen 10,5% (0,105)

11%

R

OBS!

4g) 95% KI för förväntat y då $x=14$

$$0{,}7132+0{,}27022\cdot 14\pm 2{,}201\cdot\sqrt{0{,}04768\left(\frac{1}{13}+\frac{(14-10)^2}{262}\right)}$$

$$\Rightarrow 4{,}49628\pm 2{,}201\cdot\sqrt{0{,}04768\cdot 0{,}13799}$$

$$\Rightarrow 4{,}49628\pm 2{,}201\cdot 0{,}0811=[4{,}32;\,4{,}67]$$

R R

OBS

4h) 95% Prediktionsintervall för enskilt y då $x=14$

$$0{,}7132+0{,}27022\cdot 14\pm 2{,}201\cdot\sqrt{0{,}04768\cdot\left(1+\frac{1}{13}+\frac{(14-10)^2}{262}\right)}$$

$$\Rightarrow 4{,}49628\pm 2{,}201\cdot\sqrt{0{,}04768\cdot 1{,}13799}$$

$$\Rightarrow 4{,}49628\pm 2{,}201\cdot 0{,}2329=[3{,}98;\,5{,}01]$$

R R

OBS! f) 95% KI för β

$$0{,}27022\pm 2{,}201\cdot 0{,}012=[0{,}2438;\,0{,}2966]$$

R