Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 05-06-2018

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: fredagen den 15 juni, kl. 15.00 i B705.
Därefter kan skrivningarna hämtas på studentexpeditionen, plan 7 i B-huset.

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: (30 poäng)
En företagsledare i Grönköping undersöker sambandet mellan Y = konsumtion av lyxvaror (i tusentals kr.) och X = inkomst (i tiotusentals kr.). Data som gäller:

y	1	1	2	3	3	4	4	5	6	6
x	2	3	3	4	5	6	7	8	9	10

Du har tillgång till följande information:
The regression equation is
Y = -0,22 + 0,65 X

Predictor	Coef	SE Coef	T
Constant	-0,22		-0,70
X	0,65		12,79

Analysis of Variance

SOURCE	DF	SS
Regression		29,07
Residual Error		
Total		30,50

a). 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)
b). Testa H_0: $\beta = 0$ mot H_1: $\beta < 0$ (med t-test). Använd signifikansnivå 5% ($\alpha = 0{,}05$). (5 poäng)
c). Beräkna ett 95%-igt konfidensintervall för β. (5 poäng)
d). Beräkna ett 95%-igt konfidensintervall för det förväntat y-värdet för x = 8. (5 poäng)
e). Beräkna ett 95%-igt prediktionsintervall för det enskilda y-värdet för x = 8. (5 poäng)

Uppgift 2: (30 poäng)
En regressionsanalys har utförts med en beroende variabel (Y) och tre förklarande variabler (X_1, X_2 och X_3). Det modellantagande som ligger till grund för beräkningarna är att
$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$, där ε antas vara normalfördelat med väntevärde 0 och varians σ^2.
Du har tillgång till följande information:

```
The regression equation is

Y = 701 + 0,106 X1 + 18,5 X2 - 10,1 X3

Predictor      Coef     SE Coef        T
Constant      700,8                 6,44
X1           0,1058                 0,14
X2           18,523                 2,54
X3          -10,119                -4,39

s = ?

Analysis of Variance

Source            DF          SS
Regression
Residual Error             29850
Total             11      235241
```

a). Beräkna s. (5 poäng)
b). Beräkna residualvariansen. (5 poäng)
c). Beräkna R^2 och förklara vad det erhållna värdet säger. (5 poäng)
d). Testa på 5% signifikansnivå om de tre förklarande variablerna tillsammans kan förklara variationen i y. (5 poäng)
e). Testa på 5% signifikansnivå om $\beta_1 < 0$ i modellen. (5 poäng)
f). Beräkna ett 95% konfidensintervall för β_3. (5 poäng)

Uppgift 3: (20 poäng)
I Grönköping vill ett företag göra en prognos för det kommande året. Följande försäljningssiffror (100 000-tals kronor) finns tillgängliga.

År	2013	2014	2015	2016	2017
Försälj.	2	3	4	6	9

a). Anpassa en exponentiell modell (exponentiell trendfunktion) till tidsserien. Tolka de skattade koefficienterna i ord och uttryckta i termer av de aktuella variablerna. (15 poäng)
b). Enligt den anpassade modellen, gör en prognos för företagets försäljningssiffror år 2018. (5 poäng)

Uppgift 4: (20 poäng)
Följande tabell visar antal sålda smarta mobiler (i hundratals) i Grönköping under åren 2013-2017.

2013	2014	2015	2016	2017
2,8	2,9	3,0	2,8	2,6

a) Anpassa en andragradskurva till tidsserien med hjälp av minsta-kvadrat-metoden. (15 poäng)
b) Enligt den anpassade modellen, gör en prognos för antal sålda smarta mobiler i Grönköping, år 2018. (5 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	$n-1$		

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

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}: $e^{b_1 \pm z \times s_{b_1}}$

TABELL 3. *t*-fördelningens kvantiler

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

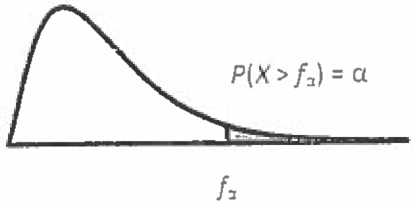
Vilket värde har t_α om $P(T > t_\alpha) = \alpha$ där α är en given sannolikhet. Utnyttja även $P(T \leq -t_\alpha) = P(T > t_\alpha)$.

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	2,45	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

③

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 05/06/18

Sal: Brunnsvikssalen

Tenta: Regressions- och tidsserieanalys

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: 0021-DSJ

☒ 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						7
Lär.ant. 25p	20p	20p	20p						

1k

POÄNG	BETYG	Lärarens sign.
85p	B	RC

SU, STATISTIK

Skrivsal: Brunnsviksalen **Anonymkod:** 0021-DSJ **Blad nr:** 1

1\) 25p

1. $Y = -0{,}22 + 0{,}65X$

Predictor	Coef	SE Coef	T
Constant	-0,22 (B_0)	0,3142857143	-0,70
X	0,65 (B_1)	0,0508209539	12,79

Analysis of Variance

SOURCE	DF	SS	MS	F
Regression	1 (k)	29,07 (SSR)	29,07 (MSR)	162,6293706
Residual Error	8 (n-k-1)	1,43 (SSE)	0,17875 (MSE)	
Total	9 (n-1)	30,50 (SST)		

- $T = \frac{B_0}{s_{b_0}} \longrightarrow s_{b_0} = \frac{B_0}{T} = 0{,}3142857143$

  $s_{b_1} = \frac{B_1}{T} = 0{,}0508209539$

- $n = 10$, $k = 1$, $n-k-1 = 10-1-1 = 8$, $n-1 = 9$

- $SST = SSR + SSE \longrightarrow SSE = SST - SSR = 30{,}50 - 29{,}07 = 1{,}43$

- $MSR = \frac{SSR}{k} = \frac{29{,}07}{1} = 29{,}07$

  $MSE = \frac{SSE}{n-k-1} = \frac{1{,}43}{8} = 0{,}17875$

- $F = \frac{MSR}{MSE} = \frac{29{,}07}{0{,}17875} = 162{,}6293706$

a) $\boxed{\text{T-test}}$

Hypoteser: $H_0: \beta = 0$

$H_1: \beta \neq 0$

Signifikansnivå: $\alpha = 0{,}05 = 5\%$

Frihetsgrader: $n-k-1 = 10-1-1 = 8$

Testvariabel: $t_{(n-k-1)(\alpha/2)} = t_{(8)(0{,}025)} = 2{,}306$

Beslutsregel: om $t_{obs} > 2{,}306$ eller

$t_{obs} < -2{,}306$ så förkastas H_0

Resultat: $t_{obs} = 12{,}79 > 2{,}306$

Slutsats: H_0 förkastas på 5% signifikansnivå

R

$\boxed{\text{F-Test}}$

Hypoteser: $H_0 = \beta = 0$

$H_1: \beta \neq 0$

Signifikansnivå: $\alpha = 0{,}05 = 5\%$

Frihetsgrader: $k = 1$, $n-k-1 = 8$

Testvariabel: $F_{(k,\, n-k-1)\,(\alpha)} = F_{(1,8)(0{,}05)} = 5{,}32$

Beslutsregel: Om $F_{obs} > 5{,}32$ förkastas H_0

Resultat: $F_{obs} = 162{,}6293706 > 5{,}32$

R

Slutsats: H_0 förkastas på 5% signifikansnivå

R

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0021-DSJ **Blad nr:** 2

1. b) $\boxed{\text{T-test}}$

Hypoteser: $H_0: \beta = 0$

$H_1: \beta < 0$

Signifikansnivå: $\alpha = 0{,}05 = 5\%$

Frihetgrader: $n-k-1 = 8$

Testvariabel: $t_{(n-k-1)(\alpha)} = t_{(8)(0{,}05)} = 1{,}860$

Beslutsregel: H_0 förkastas om $t_{obs} > 1{,}860$

Resultat: $t_{obs} = 12{,}79 > 1{,}860$

Slutsats: H_0 förkastas på 5% signifikansnivå

F

SE FACIT

c) 95% konfidensintervall för β ges av formeln:

$$b \pm t_{n-2,\frac{\alpha}{2}} \cdot S_{b_1}$$

$b = 0{,}65$

$t_{n-2,\frac{\alpha}{2}} = t_{(8)(0{,}025)} = 2{,}306$

$$S_b = \sqrt{\frac{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}{\sum(x_i - \bar{x})^2}} \longrightarrow S_b = \sqrt{\frac{S_e^2}{\sum(x_i - \bar{x})^2}}$$

$S_e^2 = MSE = 0{,}17875$

$$\sum(x_i - \bar{x})^2 = n \cdot \sum x_i^2 - \left(\sum x_i\right)^2$$

$n = 10$

$\sum x_i^2 = 393$

$(\sum x_i)^2 = 57^2 = 3249$

$$s_b = \sqrt{\frac{0{,}17875}{10 \cdot 393 - 3249}} = 0{,}0162012853$$

Konfidensintervallet blir: $0{,}65 \pm 2{,}306 \cdot 0{,}0162012833 = 0{,}65 \pm 0{,}0373601638$

$\{0{,}6126398362 \ ; 0{,}6873601638\}$ Vi kan med 95% säkerhet säga att värdet på parametern β kommer att ligga någonstans inom det ovan givna intervallet

c) $x = 8$ ger det förväntade y-värdet:

$$\hat{y}_{n+1} = -0{,}22 + 0{,}65 \cdot 8 = 4{,}98$$

Konfidensintervall för förväntat y-värde ges av:

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

$\boxed{\hat{y}_{n+1} = 4{,}98 \ ; \ t_{n-2,\frac{\alpha}{2}} = 2{,}306; \ s_e^2 = 0{,}17875 \ ; \ \sum (x_i - \bar{x})^2 = n \sum x_i^2 - (\sum x_i)^2 = 681}$ Detta vet vi redan

$\bar{x} = 5{,}7$

$(x_{n+1} - \bar{x})^2 = (8 - 5{,}7)^2 = 5{,}29$

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: 0021-DSJ Blad nr: 3

Fortsättning fråga 1 d)

Konfidensintervallet blir:

$$4{,}98 \pm 2{,}306\sqrt{0{,}17875\left(\frac{1}{10}+\frac{5{,}29}{68{,}1}\right)} = 4{,}98 \pm 0{,}320056916$$

$$\left[4{,}659943084 ; 5{,}300056916\right]$$

R

e) Prediktionsintervall för $\hat{y}_{n+1} = 4{,}98$ ges av:

$$\underbrace{b_0 + b_1 \cdot x_{n+1}}_{\hat{y}_{n+1}} \pm t_{n-2, \frac{\alpha}{2}} \sqrt{s_e^2\left(1+\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right)} \longrightarrow$$

$$4{,}98 \pm 2{,}306\sqrt{0{,}17875\left(1+\frac{1}{10}+\frac{5{,}29}{68{,}1}\right)} = 4{,}98 \pm 1{,}026140275$$

$$\left[3{,}953859725 ; 6{,}006140275\right]$$

R

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0021-DSJ **Blad nr:** 4

2. $Y = 701 + 0,106x_1 + 18,5x_2 - 10,1x_3$

2) 2 0p

Predictor	Coef	SE Coef	T
Constant	700,8 (β_0)	108,8198758 (s_{b_0})	6,44
x_1	0,1058 (β_1)	0,7557142857 (s_{b_1})	0,14
x_2	18,523 (β_2)	7,292519685 (s_{b_2})	2,54
x_3	-10,119 (β_3)	2,30501139 (s_{b_3})	-4,39

- $t = \frac{b}{s_b} \longrightarrow s_b = \frac{b}{t}$

$s_{b_0} = \frac{b_0}{t} = \frac{700,8}{6,44} = 108,8198758$

$s_{b_1} = \frac{b_1}{t} = \frac{0,1058}{0,14} = 0,7557142857$

$s_{b_2} = \frac{b_2}{t} = \frac{18,523}{2,54} = 7,292519685$

$s_{b_3} = \frac{b_3}{t} = \frac{-10,119}{-4,39} = 2,30501139$

Source	DF	SS	MS	F
Regression	3	205391 (SSR)	68463,66667 (MSR)	18,34872139
Residual Error	8	29850 (SSE)	3731,25 (MSE)	
Total	11	235241 (SST)		

- $k = 3$, $n-1 = 11 \longrightarrow n = 12$, $n-k-1 = 12-3-1 = 8$
- $SST = SSR + SSE \longrightarrow SSR = SST - SSE = 235241 - 29850 = 205391$

$\longrightarrow$

• $MSR = \frac{SSR}{k} = \frac{205391}{3} = 68463,6667$ • $F = \frac{MSR}{MSE} = 18,34872139$

• $MSE = \frac{SSE}{n-k-1} = \frac{29850}{8} = 3731,25$

F

a) $S = \frac{Sb_0 + Sb_1 + Sb_2 + Sb_3}{n} = (108,8198758 + 0,7557742857 + 7,292519685 + 2,30501739)/n = 29,79328029$ (SE FACIT)

b) Residualvariansen $= \sigma^2 = \hat{\sigma}_e^2 = MSE = 3731,25$ (se beräkning ovan) R

c) $R^2 = \frac{SSR}{SST} = \frac{205391}{235241} = 0,8731088593$ R

R^2-~~värdet~~ anger hur stor del av variationen i Y som kan förklaras av regressionssambandet (de oberoende variablerna tillsammans). I denna modell kan ca 87,31% av variationen i Y förklaras av x_1, x_2 och x_3 R

d) $\boxed{\text{F-test}}$

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

H_1: $\beta_i \neq 0$

Signifikansnivå: $\alpha = 0,05 = 5\%$

Frihetsgrader: $k = 3$, $n-k-1 = 8$

Testvariabel: $F_{(k,n-k-1)(\alpha)} = F_{(3,8)(0,05)} =$

Beslutsregel: H_0 förkastas om $F_{obs} > 4,07$

Resultat: $F_{obs} = 18,34872139 > 4,07$

Slutsats: H_0 förkastas på 5% signifikansnivå

R

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0021 - DSJ **Blad nr:** 5

Fortsättning fråga 2,

e) $\boxed{T-test}$

Hypoteser: $H_0: \beta_1 \geq 0$

$H_1: \beta_1 < 0$

Signifikansnivå: $\alpha = 0{,}05 = 5\%$

Frihetsgrader: $n-k-1 = 12-3-1 = 8$

Testvariabel: $t_{(n-k-1)(\alpha)} = t_{(8)(0{,}05)} = 1{,}860$

Beslutsregel: Om $t_{obs} > 1{,}860$ förkastas H_0

FEL (SE FACIT)

Resultat: $t_{obs} = 0{,}14 < 1{,}860$

F (SE FACIT)

Slutsats: H_0 kan inte förkastas på 5% signifikansnivå

F

f) 95% konfidensintervall för β_3 fås av:

$\beta_3 \pm t_{n-2, \frac{\alpha}{2}} \cdot S_{b_3}$

$\beta_3 = -10{,}119$

$t_{n-k-1, \frac{\alpha}{2}} = t_{(8)(0{,}025)} = 2{,}306$

$S_{b_3} = 2{,}30501139$

$\rightarrow -10{,}119 \pm 2{,}306 \cdot 2{,}30501139 = -10{,}119 \pm 5{,}315356265$

$\{-15{,}43435627 ; -4{,}803643735\}$

R

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0021-DSJ **Blad nr:** 6

3. a) exponentiell funktion: $\hat{y} = a \cdot b^x$ logaritmerats:

$\log \hat{y} = \underbrace{\log a}_{a'} + x \cdot \underbrace{\log b}_{b'} \Rightarrow \log \hat{y} = a' + b'x \rightarrow \boxed{\log \hat{y} = a' + b't} \leftarrow$ eftersom tidsserie

Skattning av parametrarna b' och a' ges av:

$$b' = \frac{\sum t \cdot \log(y) - \frac{1}{n}\sum t \sum \log(y)}{\sum t^2 - \frac{1}{n}\left(\sum t\right)^2}$$

$$a' = \frac{\sum \log(y)}{n} - b' \frac{\sum t}{n}$$

Vi ställer upp en tabell för att få fram de värden vi behöver samt centrerar årtalen:

År	$t = År - 2015$	y	t^2	log(y)	$t \cdot \log(y)$
2013	-2	2	4	0,3010299957	-0,6020599913
2014	-1	3	1	0,4771212547	-0,4771212547
2015	0	4	0	0,6020599913	0
2016	1	6	1	0,7781512504	0,7781512504
2017	2	9	4	0,9542425094	1,908485019
	0	24	10	3,112605002	1,607455023

Eftersom $\sum t = 0$ blir formlerna nu:

$$b' = \frac{\sum t \cdot \log(y)}{\sum t^2} = \frac{1,607455023}{10} = 0,1607455023$$

$$a' = \frac{\sum \log(y)}{n} = \frac{3,112605002}{5} = 0,6225210004$$

$b = 10^{b'} = 1{,}447923118$

$a = 10^{a'} = 4{,}192962714$

$\hat{y} = 4{,}192962714 \cdot 1{,}447923118^{t}$

Parametern a betyder att år 2015 ($t=0$) var försäljningen ca 4,193 (100 000-tals kronor) eller 419 300 kr.

Parametern b betyder att försäljningen i genomsnitt kommer att öka med ca 44,79% per år

b) År 2018 $\longrightarrow$ $t=3$

$\hat{y}_{2018} = 4{,}192962714 \cdot 1{,}447923118^{3} = 12{,}72792206$

Modellen förutsäger att försäljningen år 2018 kommer vara ca 1,27 miljoner kronor (1,272,792,206 kr)

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0021-DSJ **Blad nr:** 7

4. a) Andragradskurva fås av: $\hat{y} = a + b_1 t + b_2 t^2$ och skattas mha. normalekvationerna:

4) 2p

$$
\left.
\begin{aligned}
a \cdot n + b_1 \sum t + b_2 \sum t^2 &= \sum y \\
a \cdot \sum t + b_1 \sum t^2 + b_2 \sum t^3 &= \sum y \cdot t \\
a \cdot \sum t^2 + b_1 \sum t^3 + b_2 \sum t^4 &= \sum y \cdot t^2
\end{aligned}
\right\}
$$

Vi gör en tabell för att få fram de summor vi behöver samt centrerar årtalen:

År	$t = År - 2015$	t^2	t^3	t^4	y	$y \cdot t$	$y \cdot t^2$
2013	-2	4	-8	16	2,8	-5,6	11,2
2014	-1	1	-1	1	2,9	-2,9	2,9
2015	0	0	0	0	3,0	0	0
2016	1	1	1	1	2,8	2,8	2,8
2017	2	4	8	16	2,6	5,2	10,4
	0	10	0	34	14,1	-0,5	27,3

$n = 5$

Normalekvationerna blir efter insättning av summorna:

(1) $5a + 10b_2 = 14,1$

(2) $10b_1 = -0,5 \longrightarrow b_1 = -0,05$

(3) $10a + 34b_2 = 27,3$

R

$\longrightarrow$

Multiplicera ekvation (1) med 2

(1) $5a \cdot 2 + 10b_2 \cdot 2 = 14{,}1 \cdot 2 \rightarrow 10a + 20b_2 = 28{,}2 \longrightarrow 10a = 28{,}2 - 20b_2$

(3) $10a + 34b_2 = 27{,}3 \longrightarrow 10a = 27{,}3 - 34b_2$

sätt (1) = (3):

$28{,}2 - 20b_2 = 27{,}3 - 34b_2 \longrightarrow 28{,}2 - 27{,}3 = -34b_2 + 20b_2 \longrightarrow$

$0{,}9 = -14b_2 \longrightarrow b_2 = -0{,}064285714 3$ R

sätt in b_2 i ursprungliga (1)

$5a + 10 \cdot (-0{,}0642857143) = 14{,}1 \longrightarrow 5a - 0{,}642857143 = 14{,}1 \longrightarrow$

$5a = 14{,}742857 14 \longrightarrow a = 2{,}948571 429$ R

$\hat{y} = 2{,}948571429 + 0{,}05t - 0{,}0642857143 \cdot t^2$ R

b) År 2018 $\longrightarrow t = 3$

$\hat{y} = 2{,}948571429 - 0{,}05 \cdot 3 - 0{,}0642857143 \cdot 3^2 \Longrightarrow$ R

$\hat{y} = 2{,}948571429 - 0{,}15 - 0{,}5785714287 = 2{,}22$

År 2018 förväntas 222 smarta mobiler säljas i Grönköping