STOCKHOLMS UNIVERSITET
Statistiska institutionen
Regressionsanalys och undersökningsmetodik, vårterminen 2017

Tentamen i Regressions- och tidsserieanalys onsdagen den 26 april 2017

Datum	2017-04-26
Tid:	10.00-15.00
Ansvarig Lärare:	Jörgen Säve-Söderbergh
Antal frågor:	5
Maxpoäng:	50
Hjälpmedel:	1) Språklexikon
	2) Kalkylator utan lagrade formler eller lagrad text
Tentamensgenomgång	Måndag 15 maj kl. 15.00 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. Wotan vill undersöka vilket som är det mest effektiva sättet att använda tiden vid universitetet. Det finns många sidor av hans personlighet som måste utvecklas bortsett från att han, naturligtvis, behöver en examen.

Wotan, som är en mycket disciplinerad person, sätter upp ett experiment. Han tänker variera antalet dagar där hela dagen går åt till förberedelser inför tentamen. Under de nio kommande tentorna pluggar han en dag till den första, två dagar till den andra, etc.

Nedan återges resultatet där x står för antalet dagar tentamenspluggande och y för antalet poäng på tentan. Poängen på samtliga tentamina ligger mellan 0 och 100 där 50 är godkänd.

x	y
1	38
2	48
3	53
4	47
5	56
6	59
7	68
8	60
9	82

Wotan anpassar följande modell $Y = \beta_0 + \beta_1 X + \varepsilon$ till ovanstående siffror.

a) Beräkna minsta-kvadrat-skattningen av β_1. (1 p)

b) Beräkna minsta-kvadrat-skattningen av β_0. (1 p)

c) Beräkna ett 95%-igt konfidensintervall för β_1. (2 p)

d) Testa nollhypotesen $H_0 : \beta_1 = 0$ mot alternativet $H_1 : \beta_1 \neq 0$ på 5% signifikansnivå. (2 p)

e) Tolka punktskattningen av β_1. (1 p)

f) Använd det skattade sambandet för att svara på Wotans fråga: hur många dagar måste han plugga för att klara sig enligt experimentet? (3 p)

2. (Denna fråga är inte en fortsättning av fråga 1) Efter att ha anpassat en multipel linear regressionsmodell $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$ till ett datamaterial bestående av $n = 20$ observationer, så fann man följande ANOVA-tablå bland utskrifterna:

	SS	df	MS	F
Regression				
Residual	400			
Total	500			

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

b) Testa $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ mot H_1 : minst en av $\beta_i \neq 0$, med hjälp av ANOVA-tablån på 5% signifikansnivå. (5 p)

c) Beräkna determinationskoefficienten R^2. (2 p)

3. En amerikansk arbetsmarknadsforskare hade skattat en löneekvation

$$\hat{y}_i = 0.980 + 1.12385 x_{1i} + 0.9925 x_{2i},$$

där y står för en viss individs årliga löneinkomst i tusentals dollar, x_1 för det antal år utbildning som individen har genomgått och x_2 för individens yrkeserfarenhet mätt i antal år. Antag att en viss individ har nio års utbildning och sex års yrkeserfarenhet. Vilken årlig lön kan vi förvänta oss att individen skulle få om vi antar att modellen stämmer? (10 p)

4. Riksrevisionsverket har sammanställt inköpen av smarta mobiler för personer i chefsfunktion på ett antal mindre myndigheter.

t	2007	2008	2009	2010	2011	2012	2013
y_t	390	250	125	106	240	495	890

Anpassa en andragradskurva med hjälp av minsta-kvadrat-metoden och beräkna en prognos för 2014 om utvecklingen antas följa kurvan. (10 p)

5. Om man önskar anpassa en linje som går genom origo till ett datamaterial $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ är det modellen $Y = \beta X + \varepsilon$ som bör användas. Som vanligt antar vi att $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

Minsta-kvadrat-skattningen av β ges av

$$b = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Skattningen av feltermernas varians σ_ε^2 ges av

$$\hat{\sigma}_\varepsilon^2 = s_\varepsilon^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^2 - b\sum_{i=1}^{n} x_i y_i}{n-1}.$$

Ett tvåsidigt konfindensintervall för β med konfidensgraden $100(1-\alpha)$ ges som

$$b \pm t_{n-1,\alpha/2}\sqrt{\frac{s_\varepsilon^2}{\sum_{i=1}^{n} x_i^2}}$$

Följande data härrör från en undersökning av tio papegojor där x står för ålder(antal dagar) och y för vikt (gram).

x	y
1	1.9
2	2.4
3	7.7
4	15.8
5	19.7
6	49.4
7	50.8
8	73.5
9	67.9
10	109.4

Använd ovanstående teori för att anpassa en kvadratisk modell utan intercept, d v s

$$Y = \beta_1 X^2 + \varepsilon.$$

a) Beräkna minsta-kvadrat-skattningen av β_1. (2 p)

b) Beräkna residualvariansen s_ε^2. (2 p)

c) Beräkna medelfelet för b_1. (2 p)

d) Beräkna ett 95%-igt konfidensintervall för β_1. (2 p)

e) Testa nollhypotesen $H_0 : \beta_1 = 1$ mot $H_1 : \beta_1 \neq 1$ på 5% signifikansnivå. (2 p)

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

α = 0,05

v_2 \ v_1 =	16	17	18	19	20	25	30	35	40	50	60	70	80	100	∞
v_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

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

$$a_0 = \bar{y} - b \cdot \bar{t}$$
$$T_t = a_0 + b \cdot t$$

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

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

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

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

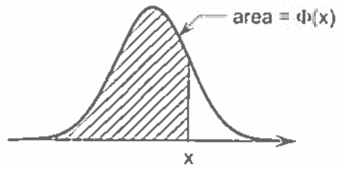
$$R_{Y|X_1,\dots,X_k} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}_i\right)^2}}$$

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

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 $X \in N(0, 1)$

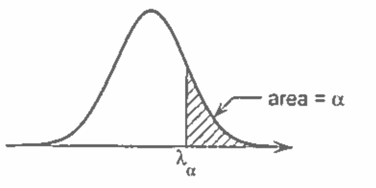
För negativa värden, utnyttja att $\Phi(x) = 1 - \Phi(-x)$

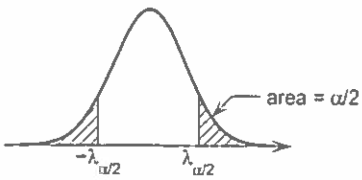
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.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 $X \in N(0, 1)$

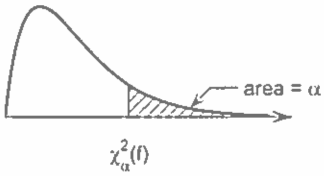
α	λ_α	α	λ_α
0.1	1.2816	0.001	3.0902
0.05	1.6449	0.0005	3.2905
0.025	1.9600	0.0001	3.7190
0.01	2.3263	0.00005	3.8906
0.005	2.5758	0.00001	4.2649





Tabell 4. χ^2-fördelningen

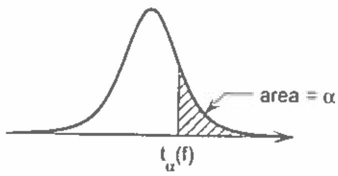
$P(X > \chi^2_\alpha(f)) = \alpha$ där $X \in \chi^2(f)$



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

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

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 26/4-2017

Sal: Värtasalen

Tenta: Regressions- och tidsserieanalys

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: RET-0026

☒ 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					13
Lär.ant. 10	10	10	10	0					

POÄNG	BETYG	Lärarens sign.
40	B	JSS

SU, STATISTIK

Skrivsal: Västra Anonymkod: RET-0026 Blad nr: 1

(1) $Y = B_0 + B_1 X + \varepsilon$ skattas av $\hat{Y} = b_0 + b_1 X$

där Y = antal poäng på tentan och
X = antal pluggdagar och n = 9

X	Y	X^2	XY	Y^2
1	38	1	38	1444
2	48	4	96	2304
3	53	9	159	2809
4	47	16	188	2209
5	56	25	280	3136
6	59	36	354	3481
7	68	49	476	4624
8	60	64	480	3600
9	82	81	738	6724
Σ 45	511	285	2809	30337

R

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i Y_i - \frac{1}{n}(\sum X_i)(\sum Y_i)}{\sum X_i^2 - \frac{1}{n}(\sum X_i)^2} =$$

$$= \frac{2809 - \frac{1}{9}(45)(511)}{285 - \frac{1}{9}(45)^2} = \frac{127}{30} \approx 4{,}2333333 \quad R$$

(a) Svar: Minsta kvadratskattningen av B_1 är ca 4,23333 $\left(\frac{127}{30}\right)$ R

$$b_0 = \bar{Y} - b_1 \bar{X} = \frac{511}{9} - \frac{127}{30} \cdot \frac{45}{9} = \frac{641}{18} \approx 35{,}611111 \quad R$$

(b) Svar: Minsta kvadratskattningen av B_0 är ca 35,611111 $\left(\frac{641}{18}\right)$ R

SU, STATISTIK

Skrivsal: Värta Anonymkod: RET-0026 Blad nr: 2

1. Forts.

95% KI för β_1 ges av $b_1 \pm t_{\frac{0.05}{2}}(n-2) \cdot S_{b_1} =$

$= \frac{127}{30} \pm t_{0.025}(7) \cdot S_{b_1} \approx 4{,}2333 \pm 2{,}36 \cdot S_{b_1}$

$$S_{b_1} = \sqrt{\frac{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}{\sum(x_i - \bar{x})^2}} = \sqrt{\frac{\frac{SSE}{n-2}}{\sum(x_i-\bar{x})^2}} =$$

$$= \sqrt{\frac{\frac{\sum(y_i-\bar{y})^2 - b_1^2 \sum(x_i-\bar{x})^2}{n-2}}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}} = \sqrt{\frac{\frac{\sum(y_i-\bar{y})^2 - \left(\frac{127}{30}\right)^2 \cdot 285 - \frac{1}{9}(45)^2}{9-2}}{285 - \frac{1}{9}(45)^2}} =$$

$$= \sqrt{\frac{\frac{\sum y_i^2 - \frac{1}{n}(\sum y_i)^2 - \frac{16129}{15}}{7}}{60}} = \sqrt{\frac{\frac{30837 - \frac{1}{9}(511)^2 - \frac{16129}{15}}{7}}{60}} =$$

$$= \sqrt{\frac{\frac{10903/45}{7}}{60}} \approx 0{,}759513 \quad R$$

SU, STATISTIK

Skrivsal: Värta Anonymkod: RET-0026 Blad nr: 3

(1) forts

95% KI för B_1: $4{,}2333 \pm 2{,}36 \cdot 0{,}759513 =$

$= 4{,}2333 \pm 1{,}7925 \Rightarrow [2{,}4408 ; 6{,}0258]$ R

(c) Svar: Ett 95% KI för B_1 ges av $[2{,}4408 ; 6{,}0258]$ R

För att testa hypotesen $H_0: B_1 = 0$ mot $H_1: B_1 \neq 0$ på 5% sign nivå används

testvariabeln $T = \frac{b_1 - B_1^*}{S_{b_1}} \sim t_{\alpha/2}(n-2)$ när H_0 är sann

H_0 förkastas om $|T_{obs}| > T_{\alpha/2}(n-2) = T_{0{,}025}(7) = 2{,}36$ R

$T = \frac{\left(\frac{127}{30}\right) - 0}{0{,}759513} \approx 5{,}57 > 2{,}36 \Rightarrow H_0$ förkastas på 5% sign nivå

(d) Svar: Med ett T-värde på 5,57 (och $f_g = 7$) R kan H_0 förkastas på 5% sign-nivå R

(e) Att punktskattningen av B_1 ges av $\hat{B}_1 = b_1 = 4{,}23$ kan tolkas som att en mer pluggdag till tentan i genomsnitt kommer resultera i 4,23 fler poäng på tentan R

SU, STATISTIK

Skrivsal: Värtan Anonymkod: RET-0026 Blad nr: 4

1) forts.

f) $\hat{Y} = \frac{647}{18} + \frac{127}{30} X$ = Skattad modell

För att klara tentan måste $\hat{Y} = 50 \Rightarrow$

$$50 = \frac{647}{18} + \frac{127}{30} X$$

$$X = 50 - \frac{647}{18} \Big/ \frac{127}{30} \Rightarrow X = \frac{1295}{381} \approx 3,4$$

Svar: För att klara tentan (få minst 50 poäng) måste Wotan plugga minst 3,4 dagar (eller 4 dagar om hen bara tänkt plugga heldagar). Då antal dagar räknas från en skattad modell finns det dock inga garantier att Wotan klarar tentan bara för att hen pluggar 4 dagar. Detta kan man se i experimentet där 4 dagars plugg bara ledde till 47 poäng på tentan.

Men den tentan var kanske extra svår?

Utmärkt svar!!

SU, STATISTIK

Skrivsal: Värta Anonymkod: RET-0026 Blad nr: 5

(2) ANOVA-tabla

	SS	df	MS	F
Regression	100	5	20	0,7
Residual	400	14	200/7	
TOTAL	500	19		

(a) R

$\boxed{n=20}$

(b) $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: Minst en $\beta_i \neq 0$

För att testa denna H_0 används ett F-test där testvariabeln ges av $F = \frac{MSR}{MSE} \sim$

$F_\alpha(p, n-p-1)$ om H_0 är sann R

H_0 förkastas om $F_{obs} > F_\alpha(p, n-p-1) = F_{0.05}(5,14) =$

$= 2,96$ R

$F = \frac{MSR}{MSE} = \frac{20}{200/7} = 0,7 < 2,96 \Rightarrow H_0$ förkastas ej

på 5% sign-nivå

Svar: Då F-värdet är 0,7 kan H_0 inte förkastas på en 5% nivå när $p=5$; $n-p-1=14$. R

SU, STATISTIK

Skrivsal: Värta Anonymkod: BET-0026 Blad nr: 6

2) forts.

g) $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{400}{500} = 0{,}20$ R

$R^2_{adj} = 1 - \frac{MSE}{\frac{SST}{n-1}} = 1 - \frac{200/7}{500/19} = -0{,}09$ R

Svar: Determinationskoefficienten $R^2 = 0{,}2$ och det justerade R^2_{adj} som straffar för många förklarande variabler är $-0{,}09$. Att R^2 är så lågt och att H_0 inte kan förkastas är inte så konstigt då större delen av variationen i datan inte kan förklaras av modellen. Bra.

SU, STATISTIK

Skrivsal: Värta Anonymkod: RET-0026 Blad nr: 7

3. $\hat{Y}_i = 0{,}98 + 1{,}12385X_{1i} + 0{,}9925X_{2i}$

Y = årlig löneinkomst i tusentals dollar

X_1 = antal utbildningsår

X_2 = år av yrkeserfarenhet

$X_1 = 9$; $X_2 = 6 \Rightarrow$

$\Rightarrow \hat{Y}_{9;6} = 0{,}98 + 1{,}12385 \cdot 9 + 0{,}9925 \cdot 6 = 17{,}04965$

$\hat{Y}_{9;6} = 17{,}04965 \Rightarrow$ en skattad årlig lön

på $1000 \cdot 17{,}04965 = 17049{,}65$

Svar: Om modellen stämmer kan 9 års utbildning och 6 års yrkeserfarenhet antas i genomsnitt generera ett årlig löneinkomst på 17049,65 dollar R

SU, STATISTIK

Skrivsal: Värta **Anonymkod:** RET-0026 **Blad nr:** 8

(4.) $Y = \alpha + \beta_1 t + \beta_2 t^2 + \varepsilon$ kan i detta fall skattas av: $\hat{Y} = a + b_1 t + b_2 t^2$. $n = 7$

År	t	Y	Yt	t^2	t^4	Yt^2
2007	-3	390	-1170	9	81	3510
2008	-2	250	-500	4	16	1000
2009	-1	125	-125	1	1	125
2010	0	106	0	0	0	0
2011	1	240	240	1	1	240
2012	2	495	990	4	16	1980
2013	3	890	2670	9	81	8010
Σ		2496_R	2105_R	28_R	196_R	14865_R

där t definieras som $(År - \overline{År})$

För att kunna skatta parametrarna i $\hat{Y} = a + b_1 t + b_2 t^2$ används normalekvationer:

$$\Sigma y_i = a \cdot n + \cancel{b_1 \cdot \Sigma t_i} + b_2 \cdot \Sigma t_i^2 \quad (1)$$

$$\Sigma y_i t_i = \cancel{a \cdot \Sigma t_i} + b_1 \cdot \Sigma t_i^2 + \cancel{b_2 \cdot \Sigma t_i^3} \quad (2)$$

$$\Sigma y_i t_i^2 = a \cdot \Sigma t_i^2 + \cancel{b_1 \cdot \Sigma t_i^3} + b_2 \cdot \Sigma t_i^4 \quad (3)$$

Där Σt_i och Σt_i^3 kan strykas då dessa summer till 0. Detta ger:

(1) $\Rightarrow 2496 = a \cdot 7 + b_2 \cdot 28 \Rightarrow 28 b_2 = 2496 - 7a \Rightarrow$

$\Rightarrow b_2 = \dfrac{2496}{28} - \dfrac{7}{28} a \Rightarrow b_2 = \dfrac{624}{7} - 0{,}25a$

SU, STATISTIK

Skrivsal: Värta **Anonymkod:** RET-0026 **Blad nr:** 9

4. forts.

(2) $\Rightarrow 2105 = b_1 \cdot 28 \Rightarrow b_1 = \dfrac{2105}{28}$

(3) $\Rightarrow 14865 = a \cdot 28 + b_2 \cdot 196 \Rightarrow$

$14865 = 28a + 196\left(\dfrac{624}{7} - 0{,}25a\right) \Rightarrow$

$\Rightarrow 14865 = 28a + 17472 - 49a \Rightarrow$

$\Rightarrow 14865 = 17472 - 21a \Rightarrow 21a = 17472 - 14865 \Rightarrow$

$\Rightarrow 21a = 2607 \Rightarrow a = \dfrac{869}{7}$

(1) $\Rightarrow b_2 = \dfrac{624}{7} - 0{,}25a \Rightarrow b_2 = \dfrac{624}{7} - 0{,}25\left(\dfrac{869}{7}\right)$

$\Rightarrow b_2 = \dfrac{624}{7} - \dfrac{869}{28} \Rightarrow b_2 = \dfrac{1627}{28}$

Normalekvationerna ger alltså att:

$a = \dfrac{869}{7} \approx 124{,}1429$ kr

$b_1 = \dfrac{2105}{28} \approx 75{,}1786$ kr

$b_2 = \dfrac{1627}{28} \approx 58{,}1071$ kr

SU, STATISTIK

Skrivsal: Värta Anonymkod: RET-0026 Blad nr: 10

4. forts.

De skattade parametrarna ger följande skattade modell: $\hat{y} = \frac{869}{7} + \frac{2105}{28}t + \frac{1627}{28}t^2$

2014 motsvarar $t = 4$ R $\Rightarrow$

$\hat{y}_{t=4} = \frac{869}{7} + \frac{2105}{28} \cdot 4 + \frac{1627}{28} \cdot 4^2 =$

$= \frac{9482}{7} \approx 1354{,}5714$ R

Svar: Om den skattade modellen stämmer och utvecklingen följer kurvan kan $\hat{y}$ antas bli ca 1354,57 $\left(\frac{9482}{7}\right)$, det vill säga ca 1355 mobiler antas köpas in av personer i chefsfunktion på de givna myndigheterna år 2014.

R

SU, STATISTIK

Skrivsal: Värta Anonymkod: RET-0026 Blad nr: 11

5

$Y = B_1 x^2 + \varepsilon$ vilken kan skattas av:

$\hat{y} = b_1 x^2$ $n = 10$

	x	y	x^2	xy	y^2
	1	1,9	1	1,9	3,61
	2	2,4	4	4,8	5,76
	3	7,7	9	23,1	59,29
	4	15,8	16	63,2	249,64
	5	19,7	25	98,5	388,09
	6	49,4	36	296,4	2440,36
	7	50,8	49	355,6	2580,64
	8	73,5	64	588	5402,25
	9	67,9	81	611,1	4610,41
	10	109,4	100	1094	11968,36
Σ	55	398,5	385	3136,6	27708,47

a) Minsta kvadratskattningen av $B_1 = \hat{B}_1 = b_1$

$$b_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{3136,6}{385} = \frac{15683}{1925} \approx 8,147012987$$

Svar: Minsta kvadratskattningen av B_1 ges av $\frac{15683}{1925}$ (ca 8,15) ✓

b) S_e^2 (MSE)

$$S_e^2 = \frac{\sum y_i^2 - b_1 \cdot \sum x_i y_i}{n-1} = \frac{27708,47 - \frac{15683}{1925} \cdot 3136,6}{10-1} \approx$$

$$\approx \frac{2154,489065}{9} \approx 239,3876739$$

Svar: Residualvariansen beräknas vara ca 239,3877 ✓

SU, STATISTIK

Skrivsal: Värta **Anonymkod:** RET-0026 **Blad nr:** 12

5. Forts. Modellen kan alltså skattas som:

$$\hat{y} = 8{,}1470x^2$$

6. Medelfelet för b_1 ges av

$$S_{b_1} = \sqrt{\frac{MSE}{\sum x_i^2}} = \sqrt{\frac{239{,}3876739}{385}} \approx$$

$$\approx \sqrt{0{,}6217861659} \approx 0{,}7885341882$$

OBS! I enkel linjär regression ges medelfelet av $\sqrt{\frac{MSE}{\sum(x_i-\bar{x})^2}}$ men då nämnaren i minstakvadratskattningen för b_1 i denna kvadratiska modell är $\sum x_i^2$ istället för $\sum(x_i-\bar{x})^2$ använder jag samma nämnare för att räkna ut S_{b_1}. Jag har även använt $n-1$ fg istället för $n-2$ fg vid uträknandet av MSE då detta verkar gälla för en kvadratisk modell utan intercept.

Svar: Medelfelet för b_1 beräknas vara ca 0,7885

SU, STATISTIK

Skrivsal: Värtan **Anonymkod:** RET-0026 **Blad nr:** 13

5 Forts.

d) Ett 95% KI för β_1 ges av:

$$b_1 \pm t_{\alpha/2}(n-1) \cdot S_{b_1} \Rightarrow 8,147013 \pm t_{0.025}(10-1) \cdot 0,788534$$

$$\Rightarrow 8,147013 \pm 2,26 \cdot 0,788534 \Rightarrow$$

$$\Rightarrow 8,147013 \pm 1,782087265 \Rightarrow [6,3649 ; 9,9297]$$

Svar: Ett 95% KI för β_1 ges av $[6,3649 ; 9,9297]$ ✓

e) $H_0: \beta_1 = 1$, $H_1: \beta_1 \neq 1$ $\alpha = 0.05$

För att testa H_0 används ett T-test

där $T = \dfrac{b_1 - \beta_1^*}{S_{b_1}} \sim t_{\frac{\alpha}{2}}(n-1)$ om H_0 är sann

där β_1^* står för värdet på β_1 under H_0.

Jag väljer att anta att $fg = (n-1)$ då detta brukar göras vid skattning av denna typ av modell

H_0 förkastas om $|T_{obs}| > T_{\frac{\alpha}{2}}(n-1) = T_{\frac{0.05}{2}}(10-1) = 2,26$

$$T = \frac{8,147013 - 1}{0,788534} \approx 9,0637 > 2,26 \Rightarrow H_0 \text{ förkastas}$$

Svar: Ett T-värde $= 9,0637$ gör att $H_0: \beta_1 = 1$ kan förkastas på 5% signifikansnivå med $fg = 9$ ✓