Stockholms universitet
Statistiska institutionen
Raul Cano

SKRIVNINGSDATUM: 02-05-2019

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. måndagen den 20 maj. 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: (30 poäng)
Dalaköping har mycket dåligt luftkvalitet. Pridolin som är partiledaren för det oranga partiet vill införa en årlig miljöavgift. "I nästa riksdagsval måste man rösta rätt!" säger många oroliga pensionärer i området. Pridolin som är en fullfjädrad politiker vill lugna ner oroliga känslor. Han kommer att bli intervjuad i det kända TV-programmet "Almanacka". Innan han uttalar sig i massmedia för att ange det exakta beloppet på avgiften, undersöker han sambandet mellan X=inkomst (månadsinkomst i tusentalskronor efter skatt) och Y = antal kronor man är villig att betala i miljöavgift. Hjälp Pridolin att svara nedanstående frågor. Han har fått följande information:

y	10	20	25	40	60
x	10	12	15	20	30

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

b). Testa H_0: $\beta = 0$ mot H_1: $\beta \neq 0$ (med F-test). Använd signifikansnivå 5% ($\alpha = 0{,}05$). (5 poäng)
Obs! Testa med F-test, inte med t-test, om du testar med t-test blir det noll poäng.

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

d). Eftersom Pridolin vill fånga så många röster som möjligt från pensionärer i området då satsar han på personer som har en inkomst lika med 13. Skatta antal kronor man är villig att betala i miljöavgift om man har en inkomst lika med 13. Det är den här skattningen som Pridolin kommer att ange i det kända TV-programmet "Almanacka" som det exakta beloppet på den årliga miljöavgiften. (5 poäng)

e). Beräkna ett 95%-igt konfidensintervall runt skattning i d). ovan. (5 poäng)

Uppgift 2: (20 poäng).
En fastighetsmäklare undersöker prisutvecklingen på bostadsrätter i Centrala Stockholm under perioden 2014-2018. Variabel Y = prisutveckling är givet i tusentals kronor per kvadratmeter (tkr/kvm). Fastighetsmäklaren har följande information:

År	2014	2015	2016	2017	2018
Y	71	84	89	92	86

a). Anpassa en andragradskurva till tidsserien med hjälp av minsta-kvadrat-metoden. (10 poäng)

b). Enligt den anpassade modellen, gör en prognos för prisutvecklingen på bostadsrätter i Centrala Stockholm, år 2019. (10 poäng)

Uppgift 3: (30 poäng)
En varuhuskedja ville undersöka hur försäljningen av en viss vara beror av varans reklamkostnader i 3 olika massmedia. Följande data för en viss månad och för 12 olika varuhus har erhållits (alla enheter i 10 000 kr., FOR = försäljning, TV = television, BIO = biografer och TID = tidningar):

FOR	84	84	80	50	20	68	34	30	54	40	57	46
TV	13	13	8	9	9	13	12	10	8	10	5	5
BIO	5	7	6	5	3	5	7	3	5	5	6	7
TID	8	8	9	5	2	7	4	3	6	5	6	5

Du har tillgång till följande information:

```
The regression equation is
FOR = - 5,87 + 0,65 TV - 0,57 BIO + 9,99 TID

Predictor      Coef      SE Coef         T
Constant                    8,13
TV                          0,55
BIO                         1,30
TID                         0,86

S =            R-sq =            R-sq(adj) =

Analysis of Variance

Source              DF       SS          MS        F
Regression                            1630,97
Residual Error                          24,50
Total
```

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 FOR (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 4: (20 poäng)
Följande tabell visar elförbrukningen (i 1 000 kWh) i Grönköping, åren 2015-2018.

År	**Kvartal 1**	**Kvartal 2**	**Kvartal 3**	**Kvartal 4**
2015	9	2,5	3,5	6
2016	8,5	3	4	7,5
2017	10,5	5	5	9
2018	11	5,5	6,5	9,5

Nedan visas en SAS-utskrift från en regressionsanalys med värden på elförbrukning som beroende (undersöknings-) variabel och tid t (t = 1, 2, 3, …, 16), D1 (kodad 1 om kvartal 1 och 0 annars), D2 (kodad 1 om kvartal 2 och 0 annars) samt D3 (kodad 1 om kvartal 3 och 0 annars) som oberoende (förklarande) variabler.

a). Skatta trend samt säsongkomponenter i en additiv modell där trenden kan antas följa en linjär funktion. (10 poäng)

b). Gör prognoser för första och andra kvartalen 2019. (10 poäng)

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	16
Number of Observations Used	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	109.26250	27.31563	120.79	<.0001
Error	11	2.48750	0.22614		
Corrected Total	15	111.75000			

Root MSE	0.47554	R-Square	0.9777	
Dependent Mean		6.62500	Adj R-Sq	0.9696
Coeff Var	7.17793			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	5.43750	0.35665	15.25	<.0001
t	1	0.25625	0.02658	9.64	<.0001
D1	1	2.51875	0.34558	7.29	<.0001
D2	1	-3.48750	0.34043	-10.24	<.0001
D3	1	-2.99375	0.33731	-8.88	<.0001

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

Formelsamling – regressionsanalys

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

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

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

Enkel lineär regression

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

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

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

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

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

Konfidensintervall för β_1 ges av

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

där

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

Prediktionsintervall

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

Konfidensintervall förväntat y-värde för ett nytt x-värde

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

Multipel regression

n st observationer och p förklarande variabler.

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

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

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

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

Säsongsrensning med regression

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

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

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

Logistisk regression

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

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

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

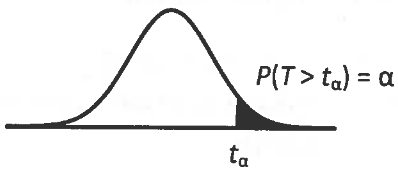
$$P(D) = \frac{\text{odds}\left(D\right)}{1 + \text{odds}\left(D\right)}$$

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

TABELL 3. t-fördelningens kvantiler

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

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

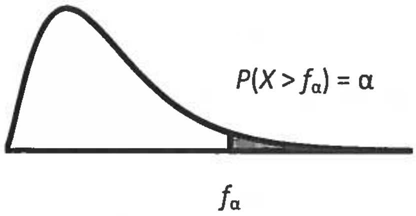


v	α = 0,1	0,05	0,025	0,010	0,005	0,0025	0,0010	0,0005
1	3,078	6,314	12,706	31,821	63,657	127,321	318,309	636,619
2	1,886	2,920	4,303	6,965	9,925	14,089	22,327	31,599
3	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	4,773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

Forts. nästa sida

TABELL 5. *F*-fördelningens kvantiler

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



$\alpha = 0{,}05$

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

Forts. nästa sida

1

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 2/5-2019

Sal: Uggleviksalen

Tenta: Regressionsanalys och undersökningsmetodik

Kurs: Regressions- och tidsserieanalys

ANONYMKOD: 0054-0EP

☒ 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						12
Lär.ant. ~~[illegible]~~ 24p	20p	30p	16p						

POÄNG	BETYG	Lärarens sign.
90p	A	RC

SU, STATISTIK

Skrivsal: Ugglevik **Anonymkod:** 0054-OEP **Blad nr:** 1

b) 24 p

1. Enkel linjär regression

X = Inkomst (1000tals kr/månad)

Y = Antal kr. villig att betala avgift

$\hat{y} = a + b \cdot x \Rightarrow \hat{y} = b_0 + b_1 \cdot x$

a) $b_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \dfrac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$

$b_0 = \bar{y} - b_1 \cdot \bar{x}$

Y	X	$X \cdot Y$	X^2	$\hat{Y}$	$(y_i - \hat{y}_i)^2$	$(\hat{y}_i - \bar{y})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2$
10	10	100	100	13.08	9.4864	321.1264	441	54.76
20	12	240	144	17.92	4.3264	171.0864	121	29.16
25	15	375	225	25.19	0.0361	33.7561	36	5.76
40	20	800	400	37.30	7.29	39.69	81	6.76
60	30	1800	900	61.51	2.2801	930.8601	841	158.76
155	87	3315	1769	155	23.419	1496.519	1520	255.2

$n = 5 \quad \bar{x} = 17.4 \quad \bar{y} = 31$

$b_1 = \dfrac{3315 - \frac{1}{5}(87 \cdot 155)}{1769 - \frac{1}{5}(87)^2} = \dfrac{618}{255.2} = 2.4216$

$b_0 = 31 - 2.4216 \cdot 17.4 = -11.136$

Svar: Den anpassade regressionslinjen är

$\hat{y} = -11.136 + 2.4216\,x$

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0054-OEP Blad nr: 2

b) F-test med $\alpha = 0.05$

<u>Hypoteser</u>: $H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

<u>Signifikans</u>: $\alpha = 0.05$

<u>Testvariabel</u>: $F = \frac{SSR}{S_e^2}$ där F är $F \sim (1;3)$ om H_0 är sann

<u>Beslutsregel</u>: H_0 förkastas om $F_{obs} > F_{krit}$

$F_{krit} = F_{0.05}^{(1;3)} = 10.13$

<u>Resultat</u>: $SSR = 1496.519$

$S_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} = \frac{23.419}{3} = 7.806$

$F_{obs} = \frac{1496.519}{7.806} = 191.7$

<u>Slutsats</u>: $F_{obs} = 191.7 > F_{krit} = 10.13$

$\Rightarrow H_0$ förkastas på 5% signifikansnivå.

Svar: Det finns underlag för att tro att det föreligger ett linjärt samband mellan X och Y.

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0054-OEP Blad nr: 3

c) T-test med $\alpha = 0.05$

Hypoteser: $H_0: \beta_1 = 0$
$H_1: \beta_1 < 0$ } ensidigt test

Signifikans: $\alpha = 0.05$

Testvariabel: $T = \frac{\hat{\beta}_1 - \beta_1^{(0)}}{s_{\hat{\beta}_1}}$ där $T \sim t(3)$ om H_0 är sann

Beslutsregel: Förkasta H_0 om $T_{obs} > T_{krit}$ STORT FEL (SE FACIT)

$T_{krit} = T_{0.95}^{(3)} = 2.353$

Resultat: $s_{\hat{\beta}_1} = \sqrt{\frac{s_e^2}{\sum(x_i - \bar{x})^2}} = \sqrt{\frac{7.806}{255.2}} = 0.1749$ STORT FEL (SE FACIT)

$T_{obs} = \frac{2.4216}{0.1749} = 13.8456$

Slutsats: $T_{obs} = 13.8456 > T_{krit} = 2.353$

$\Rightarrow H_0$ förkastas på 5% signifikansnivå

Svar: Det finns skäl att tro att det föreligger ett positivt linjärt samband mellan X och Y.

−4p

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0051-OEP Blad nr: 4

d) Skatta $\hat{y}$ för $x = 13$

$$\hat{y} = -11.136 + 2.4216\,x$$

$$\hat{y} = -11.136 + 2.4216 \cdot 13 = 42.6168$$

−1p (SE FACIT)

Svar: En person som har 13 000 kr/månad som inkomst kommer i genomsnitt vara villig att betala ca 42.62 kr i miljöavgift.

e) 95 % KI för nytt värde på x

$$\hat{y}_{n+1} \pm t_{\alpha/2}^{(n-2)} \cdot \sqrt{S_e^2\left(\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)}$$

R

$n - 2 = 3$

$\alpha/2 = 0.025$ $\Longrightarrow t_{0.975}^{(3)} = 3.182$

$x = 13$

$$42.6168 \pm t_{0.975}^{(3)} \cdot \sqrt{7.806\left(\frac{1}{5} + \frac{(13 - 17.4)^2}{255.2}\right)}$$

R

$$42.6168 \pm 3.182 \cdot \sqrt{7.806\,(0.275862)}$$

$$42.6168 \pm 3.182 \cdot 1.4674$$

$$42.6168 \pm 4.669$$

Svar: Ett 95 % konfidensintervall för skattningen i d) ges av [37.9478 ; 47.2858]

−1p (SE FACIT)

SU, STATISTIK

Skrivsal: Ugglevik **Anonymkod:** 0054-OEP **Blad nr:** 5

b) 2) 20p

2. Tidsserie med andragradskurva Y = prisutveckling (tkr/kvm)

a) $\hat{y} = a + b_1 t + b_2 t^2$ $t = \text{År} - 2016$

År	y_i	t_i	t_i^2	t_i^3	t_i^4	$y_i \cdot t$	$y_i \cdot t_i^2$
2014	71	-2	4	-8	16	-142	284
2015	84	-1	1	-1	1	-84	84
2016	89	0	0	0	0	0	0
2017	92	1	1	1	1	92	92
2018	86	2	4	8	16	172	344
Σ	422	0	10	0	34	38	804

om $\Sigma t_i = 0$, $n = 5$

$$\boxed{\begin{aligned} &(1)\ \Sigma y_i = a \cdot 5 + b_2 \cdot \Sigma t_i^2 \\ &(2)\ \Sigma y_i t_i = b_1 \cdot \Sigma t_i^2 \\ &(3)\ \Sigma y_i t_i^2 = a \cdot \Sigma t_i^2 + b_2 \Sigma t_i^4 \end{aligned}}$$

(2) $38 = b_1 \cdot 10$

$b_1 = \frac{38}{10}$

$\boxed{b_1 = 3.8}$

R

(1) $(422 = a \cdot 5 + b_2 \cdot 10) \cdot -2$

(3) $804 = a \cdot 10 + b_2 \cdot 34$

(1) $-844 = \cancel{-a \cdot 10} - b_2 \cdot 20$

(3) $804 = \cancel{a \cdot 10} + b_2 \cdot 34$

$804 - 844 = b_2 (34 - 20)$

$-40 = b_2 \cdot 14$

$b_2 = \frac{-40}{14}$

$\boxed{b_2 = -2.857}$

R R

(3) $804 = a \cdot 10 + (-2.857) \cdot 34$

$804 = a \cdot 10 - 97.138$

$10a = 901.138$

$\boxed{a = 90.1138}$

R

Svar: Den anpassade andragradskurvan $\hat{Y} = 90.1138 + 3.8t - 2.857t^2$

R

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0051-OEP Blad nr: 6

6) År 2019 motsvarar $t=3$

$\hat{Y}_{2019} = \hat{Y}_{t=3} = 90.1138 + 3.8 \cdot 3 - 2.857 \cdot 9 = 75.8$

Svar: År 2019 kommer priset på bostadsrätter i centrala Stockholm i genomsnitt att vara 75800 kr/kvm.

SU, STATISTIK

Skrivsal: Ugglevik **Anonymkod:** 0054-OEP **Blad nr:** 7

3) 30p

3: Multipel regressionsanalys

$p = 3 \quad n = 12$

FOR = Försäljning
TV = Television
BIO = Biografer
TID = Tidningar
} i 1000-tals kronor

ANOVA-tabellen

SOURCE	DF	SS	MS	F
REGRESSION	3	4892.91	1630.97	66.57
RESIDUAL ERROR	8	196	24.50	
TOTAL	11	5088.91		

R R R

a) $s = \sqrt{s_e^2} = \sqrt{MSE} = \sqrt{24.50} \approx 4.95$ (alt. $\sqrt{MSR} = \sqrt{1630.97}$) R

b) Residualvariansen MSE (utläst ur ANOVA-tabellen) $= 24.50$

c) $R^2 = \dfrac{SSR}{SST}$

$SSR = MSR \cdot p = 1630.97 \cdot 3 = 4892.91$

$SST = SSR + SSE = 4892.91 + 196 = 5088.91$

$SSE = MSE \cdot (n-p-1) = 24.50 \cdot 8 = 196$

$R^2 = \dfrac{4892.91}{5088.91} \approx 0.9615$ R

Svar: Värdet på R^2 kan tolkas som att cirka 96.15 % av variationen hos den beroende variabeln FOR kan förklaras av de förklarande variablerna i modellen. R

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0054-OEP Blad nr: 8

d) Test på $\alpha = 0.05$

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

H_1: Alla ej $= 0$, dvs minst en $\neq 0$

Signifikans: $\alpha = 0.05$

Testvariabel: $F = \frac{MSR}{MSE}$ där F är $F \sim (p; n-p-1)$ om H_0 är sann

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

$F_{krit} = F^{(3;8)}_{0.05} = 4.07$ R

Resultat: $F_{obs} = \frac{MSR}{MSE} = \frac{1630.97}{24.50} = 66.57$ R

Slutsats: $F_{obs} = 66.57 > F_{krit} = 4.07$ R

$\Longrightarrow H_0$ förkastas på 5% signifikansnivå

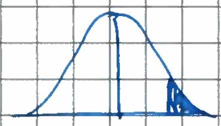
Svar: Det finns skäl att anta att de tre förklarande variablerna tillsammans förklarar variationen i FOR.

Dvs modellen är signifikant i sin helhet. R

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0051-OEP Blad nr: 9

e) Test för β_1 på $\alpha = 0.05$



Hypoteser: $H_0 : \beta_1 = 0$, $H_1 : \beta_1 < 0$ ensidigt test

Signifikans: $\alpha = 0,05$

Testvariabel: $T = \dfrac{\hat{\beta}_1 - \beta_1^{(0)}}{S_{\hat{\beta}_1}}$ där T är $t \sim (n-p-1)$ om H_0 är sann

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

$$T_{krit} = T_{0.95}^{(8)} = 1.860$$

Resultat: $T_{obs} = \dfrac{0.65}{0.55} \approx 1.18182$

Från ANOVA:
$\hat{\beta}_1 = 0.65$ (TV Coef)
$S_{\hat{\beta}_1} = 0.55$ (SE Coef)

Slutsats: $T_{obs} = 1.18182 < T_{krit} = 1.860$

$\Rightarrow$ H_0 kan ej förkastas på 5% signifikansnivå

Svar: Kan utifrån datamaterialet ej dra slutsatser om ifall det föreligger ett positivt linjärt samband mellan den förklarande variabeln TV och den beroende variabeln FOR, med 95% konfidensgrad.

SU, STATISTIK

Skrivsal: Ugglevik **Anonymkod:** 0051-OEP **Blad nr:** 10

f) 95% KI för β_3 ges av:

$b_3 \pm t_{\alpha/2}^{(n-p-1)} \cdot s_{b_3}$

$9.99 \pm 2.306 \cdot 0.86$

9.99 ± 1.98316

Från ANOVA:

$b_3 = 9.99$ (TID Coef)

$s_{b_3} = 0.86$ (SE Coef)

$t_{0.975}^{(8)} = 2.306$

Svar: Ett 95% konfidensintervall för β_3 ges av $[8.00684 ; 11.97316]$

SU, STATISTIK

Skrivsal: Ugglevik Anonymkod: 0051-OEP Blad nr: 11

4) 16p

4. Säsongsrensning med regression

Y = Elförbrukning $t = 1, 2, 3, \ldots, 76$

$D1 \begin{cases} 1 = \text{kvartal } 1 \\ 0 = \text{Annars} \end{cases}$ $D2 \begin{cases} 1 = \text{kvartal } 2 \\ 0 = \text{Annars} \end{cases}$

$D3 \begin{cases} 1 = \text{kvartal } 3 \\ 0 = \text{Annars} \end{cases}$ $D1 = D2 = D3 = 0$ om kvartal 4

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

$Y = a_0 + b \cdot t + c_1 D_{1t} + c_2 D_{2t} + c_3 D_{3t}$

FRÅN ANOVA:

$b = 0.25625$ R

$c_1 = 2.51875$

$c_2 = -3.48750$

$c_3 = -2.99375$

$\bar{y} = 6.625$ R

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

$\bar{t} = \dfrac{1+2+3+\ldots+76}{76} = 9.0625$ 8,5

−1p

$a_0 = 6.625 - 0.25625 \cdot 9.0625$
$= 4.3027$

R

R

−1p

Den skattade trenden med säsongskomponenter

$Y_t = T_t = 4.3027 + 0.25625 \cdot t + 2.51875 \cdot D_{1t} - 3.4875 \cdot D_{2t} - 2.99375 \cdot D_{3t}$

$S_1 = \underbrace{a + c_1}_{a_1} - a_0 = 5.43750 + 2.51875 - 4.3027 = 3.65355$

$S_2 = \underbrace{a + c_2}_{a_2} - a_0 = 5.43750 - 3.48750 - 4.3027 = -2.3527$

$S_3 = \underbrace{a + c_3}_{a_3} - a_0 = 5.43750 - 2.99375 - 4.3027 = -1.85895$

$S_4 = a - a_0 = 543750 - 4.3027 = 1.1348$

R

−1p

SU, STATISTIK

Skrivsal: Ugglevik **Anonymkod:** 0051-OEP **Blad nr:** 12

b) Första och andra kvartalen 2019 $\Rightarrow$ $t = 17, 18$

$y = T_{t=17, S_1} = 4.3027 + 0.25625 \cdot 17 + 2.51875 = 11.1777 \approx \underline{11.2}$

$y = T_{t=18, S_2} = 4.3027 + 0.25625 \cdot 18 - 3.4875 = 5.4277 \approx \underline{5.4}$

R

$-1p$