

TENTAMEN I STATISTIK B, 2015-04-11

Skrivtid: kl: 8-12
Tillåtna hjälpmedel: Ett A4-blad med egna handskrivna anteckningar (båda sidor) samt räknedosa
Jourhavande lärare: Lotta Hallberg (endast telefonjour)
Betygsgränser: Tentamen omfattar totalt 20p. 12 poäng och uppåt ger betyget G, 16 poäng och uppåt ger betyget VG.

Redovisa och motivera kort alla dina lösningar samt tolka dina svar

1

Nu efter Påsk så vill ett företag som säljer kvastar och katter till häxor analysera försäljningen för de senaste tre åren. I tabellen nedan visas totalförsäljningen (i löpande priser) av utrustning till häxor (Påskkärringar). Det finns två utrustningskategorier: Kvastar och Katter. Även priser för kvasten "Speedkvast" från kategorin kvastar och priser för kattrasen "Svarten" från kategorin katter är givna. Data är från åren 2013 till 2015

År	Totalförsäljning av kvastar, tkr	Totalförsäljning av katter, tkr	Pris, i kr Speedkvast	Pris, i kr Svarten
2013	1150	630	1050	520
2014	1210	755	1060	540
2015	1530	835	1090	600

Använd varorna 'Speedkvast' och 'Svarten' som representantvaror för sina kategorier och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser med basår 2013. 3p

2

Datamängden som ska analyseras består av följande variabler:

Fat = mängden kroppsfett

Triceps = hudveckstjocklek (vet inte vad som menas med hudveck här)

Thigh = omkrets på lår

Midarm = omkrets på överarm

Dessa variabler har mätts på 20 friska kvinnor i åldern 20-34 år.

Variablen *Fat* är så komplicerad och dyr att mäta så man vill hitta en modell för att kunna prediktera den istället.

Så här ser data ut: Därefter kommer ett par utskrifter från Minitab

Fat	Triceps	Thigh	Midarm
11,9	19,5	43,1	29,1
22,8	24,7	49,8	28,2
18,7	30,7	51,9	37,0
20,1	29,8	54,3	31,1
12,9	19,1	42,2	30,9
21,7	25,6	53,9	23,7
27,1	31,4	58,5	27,6

25,4	27,9	52,1	30,6
21,3	22,1	49,9	23,2
19,3	25,5	53,5	24,8
25,4	31,1	56,6	30,0
27,2	30,4	56,7	28,3
11,7	18,7	46,5	23,0
17,8	19,7	44,2	28,6
12,8	14,6	42,7	21,3
23,9	29,5	54,4	30,1
22,6	27,7	55,3	25,7
25,4	30,2	58,6	24,6
14,8	22,7	48,2	27,1
21,1	25,2	51,0	27,5

Correlations:

	Fat	Triceps	Thigh
Triceps	0,843		
Thigh	0,878	0,924	
Midarm	0,142	0,458	0,085

Cell Contents: Pearson correlation

Regression Analysis: Fat versus Triceps; Thigh; Midarm

The regression equation is

$$\text{Fat} = 117 + 4,33 \text{ Triceps} - 2,86 \text{ Thigh} - 2,19 \text{ Midarm}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	117,08	99,78	1,17	0,258	
Triceps	4,334	3,016	1,44	0,170	708,843
Thigh	-2,857	2,582	-1,11	0,285	564,343
Midarm	-2,186	1,595	-1,37	0,190	104,606

S = 2,47998 R-Sq = 80,1% R-Sq(adj) = 76,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	396,98	132,33	21,52	0,000
Residual Error	16	98,40	6,15		
Total	19	495,39			

- Beskriv hur nollhypotesen ser ut som hör till F-testet ovan. Förklara därefter varför vi kan ha signifikans i detta test men inte signifikans i något av t-testen. 2p
- Beräkna den multipla kvadrerade korrelationen R^2 för *Thigh* och de övriga två förklarande variablerna. 1p
- Prediktera *Fat* för en kvinna med *Triceps*=25, *Thigh*=35 och *Midarm*= 30. Är det rimligt att använda dessa värden för att göra denna prognos? Motivera 2p

Man har till slut bestämt att följande enkla modell är bäst:

Regression Analysis: Fat versus Thigh

The regression equation is
 $Fat = -23,6 + 0,857 Thigh$

Predictor	Coef	SE Coef
Constant	-23,634	5,657
Thigh	0,8565	0,1100

$s = 2,51024$

Följande summor har också beräknats:

$$\sum Fat = 403,9 \quad \sum Thigh = 1023,4 \quad \sum (Fat * Thigh) = 21113,5 \quad \sum Thigh^2 = 52888$$

- Visa att skattningarna med minsta kvadrat (least squares) metoden på regressionskoefficienterna β_0 och β_1 i modellen ovan blir som i utskriften. Ta hjälp av summorna ovan. 2p
- Pröva på 1% signifikansnivå om β_1 är skild från 0. Tolka resultatet. 1p
- Prediktera Fat på en kvinna med låromfång = $Thigh$ på 45. Beräkna även ett prediktionsintervall med konfidensgrad 90%. 2p
- Beräkna R^2 i modellen och tolka den. Vad kallar man R^2 vanligen på svenska? 1p
- Man kom fram till modellen ovan genom framåtvalsmetoden. Förklara hur man gått tillväga. 1p

3

Betrakta följande tidsserie Y med kvartalsdata från tre år.

Y	tid	kvartal	kvartal_1	kvartal_2	kvartal_3	kvartal_4
30	1	1	1	0	0	0
26	2	2	0	1	0	0
24	3	3	0	0	1	0
50	4	4	0	0	0	1
36	5	1	1	0	0	0
28	6	2	0	1	0	0
32	7	3	0	0	1	0
52	8	4	0	0	0	1
38	9	1	1	0	0	0
28	10	2	0	1	0	0
30	11	3	0	0	1	0
54	12	4	0	0	0	1

- Plotta tidsserien på ett lämpligt sätt. 1p

Nedan har fem tidsseriemodeller anpassats.

- Två av modellerna nedan är direkt olämpliga att använda. Redogör vilka dessa modeller är och orsaken till olämpligheten. 1p
- Beräkna prognoser för två tidsenheter fram i tiden med modell fyra och med modell fem. 2p
- Vilken ytterligare typ av modell skulle kunna användas för att modellera tidsserien? 1p

Modell 1

The regression equation is
 $Y = 19,8 + 0,625 \text{ tid} + 4,71 \text{ kvartal}$

Predictor	Coef	SE Coef	T	P
Constant	19,833	7,274	2,73	0,023
tid	0,6250	0,8133	0,77	0,462
kvartal	4,708	2,511	1,87	0,094

S = 9,20145 R-Sq = 38,5% R-Sq(adj) = 24,8%

Modell 2

The regression equation is
 $Y = 22,3 + 5,33 \text{ kvartal}$

Predictor	Coef	SE Coef	T	P
Constant	22,333	6,372	3,51	0,006
kvartal	5,333	2,327	2,29	0,045

S = 9,01110 R-Sq = 34,4% R-Sq(adj) = 27,9%

Modell 3

The regression equation is
 $Y = 47,0 + 0,625 \text{ tid} - 15,5 \text{ kvartal}_1 - 23,4 \text{ kvartal}_2 - 22,7 \text{ kvartal}_3$

Predictor	Coef	SE Coef	T	P
Constant	47,000	1,890	24,87	0,000
tid	0,6250	0,1830	3,42	0,011
kvartal_1	-15,458	1,777	-8,70	0,000
kvartal_2	-23,417	1,729	-13,54	0,000
kvartal_3	-22,708	1,700	-13,36	0,000

S = 2,07020 R-Sq = 97,6% R-Sq(adj) = 96,2%

Modell 4

The regression equation is
 $Y = 52,0 - 17,3 \text{ kvartal}_1 - 24,7 \text{ kvartal}_2 - 23,3 \text{ kvartal}_3$

Predictor	Coef	SE Coef	T	P
Constant	52,000	1,826	28,48	0,000
kvartal_1	-17,333	2,582	-6,71	0,000
kvartal_2	-24,667	2,582	-9,55	0,000
kvartal_3	-23,333	2,582	-9,04	0,000

S = 3,16228 R-Sq = 93,5% R-Sq(adj) = 91,1%

Modell 5

Time Series Decomposition for Y

Additive Model

Data Y
 Length 12
 NMissing 0

Fitted Trend Equation

$Y_t = 31,48 + 0,644 * t$

Seasonal Indices

Period	Index
1	0,750
2	-8,875
3	-7,125
4	15,250

Accuracy Measures

MAPE	4,59011
MAD	1,35577
MSD	2,85714