

Tentamen i Statistisk analys av samhällsdata, 2008-06-04

Skrivtid: kl: 8-12

Hjälpmedel: Räknedosa.

Jourhavande lärare: Anders Nordgaard

Redovisa och motivera kort alla dina lösningar

Obs! Skriv namn och personnummer på varje papper du lämnar in.

1

I tabellen nedan visas totalförsäljningen (i löpande priser) för insektsmedel för hästar och för hundar. Även priser för medlet 'Stopp o bit' för hästar från gruppen insektsmedel för hästar samt 'fästinghalsband' för hundar från gruppen insektsmedel för hundar är givna. Data är från åren 2004 till 2006

År	Tot. förs. insektsmedel för hästar	Tot. förs. insektsmedel för hundar	Pris, Stopp o bit	Pris, fästinghalsband
2004	1150	630	257	210
2005	1210	755	267	215
2006	1530	835	285	275

- Använd varorna 'Stopp o bit' och fästinghalsband som representantvaror för sina grupper och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser med basår 2004. 3p
- Hur har företagets för grupperna insektsmedel för hästar och hundars priser förändrats mellan 2004 och 2006? 1p
- Om man vill se hur prisutvecklingen för varugrupperna insektsmedel för hästar och för hundar har utvecklats gentemot den allmänna prisutvecklingen, hur kan då en lämplig indexserie beräknas som beskriver detta och vad kallas en sådan indexserie? 1p

2

I en påhittad hundpopulation är dödsriskerna enligt tabellen:

Som vanligt är den övre gränsen i åldern inte inkluderad i intervallet.

Ålder (år)	Dödsrisk (%) ${}_n q_x$
0-1 (1 ej inkl)	15
1-3	16
3-6	7
6-9	11
9-12	44
12+	100

- Beräkna den återstående medellivslängden för de 6 åldersintervallen. Svara också på frågan: Hur gammal kan husse förvänta sig att hunden blir om den lyckats uppnå en ålder på 12 år?

Sätt förslagsvis $l_0 = 1000$ och ${}_n a_x = 0,5$. Formler hittar du sist i tentan.

4p

- Det kanske är olämpligt att sätta ${}_n a_x = 0,5$. Ge förslag på vettigare värden och förklara varför de är vettigare.

1p

3

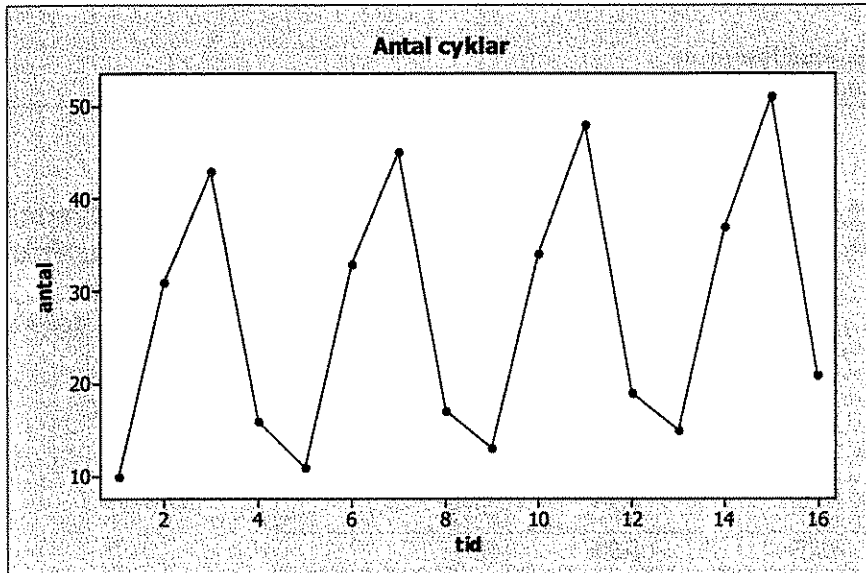
I figuren nedan visas försäljningen av antal cyklar i en viss affär. Det är kvartalsdata över fyra år. Två modeller är anpassade. Y= antal cyklar.

a) Prediktera antal sålda cyklar för första kvartalet år fem med båda modellerna.

2p

b) Tolka för båda modellerna den skattade säsongeffekten för kvartal 3.

2p



Modell 1

Regression Analysis: y versus tid; kv_1; kv_2; kv_3

The regression equation is

$$y = 13,3 + 0,500 \text{ tid} - 4,50 \text{ kv}_1 + 16,5 \text{ kv}_2 + 29,0 \text{ kv}_3$$

Predictor	Coef	SE Coef	T	P
Constant	13,2500	0,5056	26,20	0,000
tid	0,50000	0,03769	13,27	0,000
kv_1	-4,5000	0,4900	-9,18	0,000
kv_2	16,5000	0,4827	34,19	0,000
kv_3	29,0000	0,4782	60,64	0,000

S = 0,674200 R-Sq = 99,8% R-Sq(adj) = 99,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	2990,00	747,50	1644,50	0,000
Residual Error	11	5,00	0,45		
Total	15	2995,00			

Durbin-Watson statistic = 2,2

Modell 2

Time Series Decomposition for y

Additive Model

Data Y
Length 16
NMissing 0

Fitted Trend Equation

$$Y_t = 23,531 + 0,496*t$$

Seasonal Indices

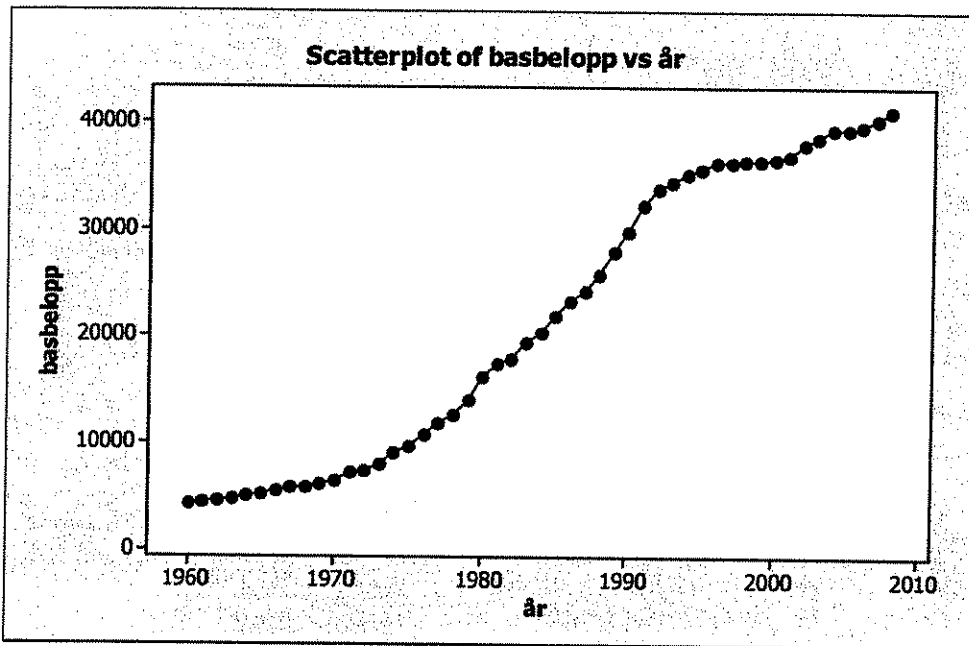
Period	Index
1	-14,8438
2	6,4063
3	18,4063
4	-9,9688

Accuracy Measures

MAPE	2,04255
MAD	0,46967
MSD	0,36983

4

I grafen nedan ser vi basbeloppet i kr i Sverige mellan 1960 och 2008.



De 10 sista åren är basbeloppet:

36400, 36600, 36900, 37900, 38600, 39300, 39400, 39700, 40300, 41000

- a) Använd dubbel exponentiell utjämning för linjär trend och visa hur de 3 sista åren utjämnas. Du kan använda direkt $S_{2005} = 39109,3$ och $T_{2005} = 286,153$. Gör även prognos för 2009, 2010 samt 2011 då du står vid tidpunkt 2008. Låt $\alpha = \beta = 0,3$ 4p
- b) Visa hur 3 punkters glidande medelvärde beräknas för åren 2005, 2006 samt 2007. 2p

Formler:

$${}_n d_x = {}_n q_x \cdot l_x$$

$${}_n L_x = \frac{n(l_x + l_{x+n})}{2}$$

$$L_{12+} \approx 3 \cdot l_{12}$$

$$T_x = T_{x+n} + {}_n L_x$$

$$T_{12+} = L_{12+}$$

$$e_x = \frac{T_x}{l_x}$$

Formelsamling Index

Sammanstatta fastbasindex:

$$I_t = i_{1,t} \cdot w_1 + i_{2,t} \cdot w_2 + \dots + i_{n,t} \cdot w_n$$

där n är antalet ingående varor/tjänster, $i_{1,t}, \dots, i_{n,t}$ är enkla prisindex för ingående varor, alla med basår t_0 och w_1, \dots, w_n väljs enligt ett viktsystem:

$$\text{Laspeyre: } w_i = \frac{p_{i,t_0} \cdot q_{i,t_0}}{\sum_j p_{j,t_0} \cdot q_{j,t_0}}$$

$$\text{Paasche: } w_i = \frac{p_{i,t_0} \cdot q_{i,t}}{\sum_j p_{j,t_0} \cdot q_{j,t}}$$

Kedjeprisindex:

$$I_t = \frac{L_{0,1}}{100} \cdot \frac{L_{1,2}}{100} \cdot \dots \cdot \frac{L_{t-1,t}}{100} \cdot 100$$

där

$$L_{t-1,t} = \sum_{i=1}^n \frac{p_{i,t}}{p_{i,t-1}} \cdot 100 \cdot w_{i,t-1,t}$$

är årslänken från år $t-1$ till t för n ingående varor/tjänster. $w_{i,t-1,t}$ väljs enligt ett viktsystem:

$$\text{Laspeyre: } w_{i,t-1,t}^L = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t-1}{\text{Totala försäljningsvärdet år } t-1}$$

$$\text{Paasche: } w_{i,t-1,t}^P = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t \text{ i priser för år } t-1}{\text{Totala försäljningsvärdet år } t \text{ i priser för år } t-1}$$

Med representantvaror byts "Försäljningsvärdet för vara i " mot "Försäljningsvärdet för varugrupp i " i vikterna.

Implicitprisindex:

$$I_t = \frac{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i löpande priser}}{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i basårets priser}} \cdot 100$$

Relativprisindex:

$$I_t^R = \frac{I_t^V}{I_t^0} \cdot 100$$

där I_t^V = Prisindex för aktuell vara/tjänst/grupp och I_t^0 = Prisindex för den större jämförelsegruppen, t ex KPI.

1 Enkel linjär regression

Den förklarande variabeln har under kursens gång skrivits t eller x beroende på om den förklarande variabeln har varit tid eller inte. Här har jag valt att skriva med x .

$$y = \alpha + \beta x + \epsilon \quad (1.2) \text{ efter förenkling [16]}$$

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (2.4) [32]$$

$$a = \frac{\sum y_i}{n} - b \frac{\sum x_i}{n} = \bar{y} - b\bar{x} \quad (2.5) [32]$$

$$\hat{y} = a + bx \quad (\text{efter omskrivning}) [33]$$

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} \quad (2.7) [36]$$

$$\sum (y_i - \hat{y}_i)^2 = \sum y_i^2 - a \sum y_i - b \sum x_i y_i = \sum (y_i - \bar{y})^2 - b^2 \sum (x_i - \bar{x})^2 \quad (\text{fotnot 6}) [36]$$

$$\text{Konfidensintervall för } \mu_{y|x_0} : a + bx_0 \pm t \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}} \quad (2.20) [66]$$

$$\text{Prediktionsintervall för } y|x_0 : a + bx_0 \pm t \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}} \quad (2.23) [67]$$

2 Multipel linjär regression

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon \quad (1.2) [16]$$

3 Glidande medelvärden

$$M_t = \sum w_s y_{t+s}, \sum w_s = 1, 0 \leq w_s \leq 1 \quad (5.10) [187]$$

4 Mått på prognosers noggrannhet

$$\text{MSE} = \frac{\sum_t (y_{t+1} - M_{it})^2}{n} \quad (6.4) [221]$$

$$\text{MAD} = \frac{\sum_t |y_{t+1} - M_{it}|}{n} \quad (6.5) [221]$$

$$\text{MAPE} = \frac{1}{n} \sum_t \frac{|y_{t+1} - M_{it}|}{y_{t+1}} \times 100, y_{t+1} > 0 \quad (6.6) [221]$$

$$\text{ME} = \frac{\sum_t (y_{t+1} - M_{it})}{n} \quad (\text{ej med i AJÅ})$$

$$\text{MPE} = \frac{1}{n} \sum_t \frac{y_{t+1} - M_{it}}{y_{t+1}} \times 100, y_{t+1} > 0 \quad (\text{ej med i AJÅ})$$

Symbolen i i de 5 formlerna ovan kommenteras under rubriken "Buggar i boken" i kursinfor på kursens webb. Observera hur AJÅ definierar ett prognosfel och håll reda på det när du ska tolka tecknet i de senare två måtten.

5 Metoder för exponentiell utjämning

$$(0 < \alpha < 1), (0 < \beta < 1), (0 < \gamma < 1)$$

5.1 Enkel exponentiell utjämning

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1} \quad (6.8) [225]$$

$$\hat{y}_{t+h} = S_t, h = 1, 2, 3, \dots \quad (6.9) [225]$$

5.2 Linjär trend

$$S_t = \alpha y_t + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (6.10a) [230]$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (6.10b) [230]$$

$$\hat{y}_{t+h} = S_t + h \cdot T_t \quad (6.10c) [231]$$

5.3 Exponentiell trend

$$S_t = \alpha y_t + (1 - \alpha)(S_{t-1} \cdot T_{t-1}) \quad (6.11a) [234]$$

$$T_t = \beta(S_t/S_{t-1}) + (1 - \beta)T_{t-1} \quad (6.11b) [234]$$

$$\hat{y}_{t+h} = S_t \cdot T_t^h \quad (6.11c) [234]$$

5.4 Linjär trend och multiplikativ säsong

$$S_t = \alpha \frac{y_t}{I_{t-s}} + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (6.12a) [235]$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (6.12b) [235]$$

$$I_t = \gamma \frac{y_t}{S_t} + (1 - \gamma)I_{t-s} \quad (6.12c) [235]$$

$$\hat{y}_{t+h} = (S_t + h \cdot T_t)I_{t-s+h} \quad (6.12d) [235]$$

5.5 Linjär trend och additiv säsong

$$S_t = \alpha(y_t - I_{t-s}) + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (6.13a) [238]$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (6.13b) [238]$$

$$I_t = \gamma(y_t - S_t) + (1 - \gamma)I_{t-s} \quad (6.13c) [238]$$

$$\hat{y}_{t+h} = (S_t + h \cdot T_t) + I_{t-s+h} \quad (6.13d) [238]$$

6 Autokorrelation

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (6.14) [241]$$

7 Tabell över t-fördelningen

t-koefficienter vid dubbelsidiga intervall

[283]

Fg	Konfidensnivå (%)						
	80.0	90.0	95.0	98.0	99.0	99.8	99.9
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

