Exp

Linköpings Universitet
IDA/Statistik
LH

HSTA21 Tentamen 5p
732G25 Tentamen 7,5hp
SDA1

Tentamen i Statistisk analys av samhällsdata, 2009-04-09

Skrivtid: kl: 8-12
Hjälpmedel: Räknedosa.
Jourhavande lärare: Lotta Hallberg

Redovisa och motivera kort alla dina lösningar

1

Tabellen nedan visar pris på färsk torskfilé samt på nötkött (högrev) åren 1996-2002. Även KPI med basår 1980 finns i tabellen.

År	Pris kr,öre på färsk torskfilé	Pris kr,öre på nötkött, högrev	KPI
1996	59,40	61,50	256,3
1997	67,10	59,10	258,0
1998	78,30	61,00	257,3
1999	83,90	57,80	258,5
2000	89,60	58,10	260,8
2001	97,40	63,40	267,1
2002	113,90	63,20	272,9

a) Beräkna indexserier för färsk torsk respektive högrev med basår 1996. Svara med en decimal. 1p
b) Räkna om KPI till basår 1996. Svara med en decimal. 1p
c) Hur många procent har färsk torsk ökat i pris samt hur många procent har högrev ökat i pris mellan 1996 och 2002 om hänsyn till inflationen tas? 2p
d) Beräkna ett sammansatt prisindex för torskfilé och högrev med vikterna 0.7 för högrev och 0.3 för torsk för åren 1996 till 2002 med basår 1996. 1p

2

Nedan ses data hämtade från SCBs hemsida. Data är för Sverige, 2001, hela riket.

moderns alder	tid	Levande födda
14 år	2001	3
15 år	2001	16
16 år	2001	60
17 år	2001	146
18 år	2001	312
19 år	2001	655
20 år	2001	1070
21 år	2001	1554
22 år	2001	2072
23 år	2001	2514
24 år	2001	3168
25 år	2001	3920
26 år	2001	5014
27 år	2001	5980
28 år	2001	6433
29 år	2001	7053
30 år	2001	7294

31 år	2001	7090
32 år	2001	6406
33 år	2001	6052
34 år	2001	5501
35 år	2001	4875
36 år	2001	4012
37 år	2001	3296
38 år	2001	2201
39 år	2001	1678
40 år	2001	1167
41 år	2001	820
42 år	2001	504
43 år	2001	317
44 år	2001	143
45 år	2001	79
46 år	2001	32
47 år	2001	15
48 år	2001	6
49 år	2001	8

moderns ålder_	totala antal kvinnor
14 år	54530
15 år	53064
16 år	51455
17 år	49387
18 år	48409
19 år	49007
20 år	49649
21 år	51266
22 år	50769
23 år	49454
24 år	50866
25 år	52449
26 år	55418
27 år	58554
28 år	58340
29 år	59554
30 år	60135
31 år	58859
32 år	58045
33 år	61301
34 år	64450
35 år	65301
36 år	65572
37 år	66064
38 år	61110
39 år	58702
40 år	57123
41 år	56698
42 år	57017
43 år	57014
44 år	58320
45 år	58053
46 år	57888
47 år	56698
48 år	58076
49 år	57929

Total medelfolkmängd år 2001 = 8897013
Totala antalet kvinnor år 2001 = 4495900
Totala antalet kvinnor mellan 14 och 49 år = 2036526
Totala antalet levande födda = 91466

a) Bestäm födelsetalet (Crude birth rate) CBR samt den generella fruktsamhetskvoten GBR och redogör för ev för- och nackdelar med de båda måtten. 2p
b) Bestäm den åldersspecifika fruktsamheten för åldersgruppen 30-39år 1p

c) Bestäm den summerade fruktsamheten (Total fertility rate) TFR och tolka resultatet. 2p

Du får i uppgift c använda avrundade 'asfr' i promille som är givna nedan om du kan beskriva hur de är beräknade.

0,1	0,3	1,2	3	6,4	13,4	21,6
30,3	40,8	50,8	62,3	74,7	90,5	102,1
110,3	118,4	121,3	120,5	110,4	98,7	85,4
74,7	61,2	49,9	36	28,6	20,4	14,5
8,8	5,6	2,5	1,4	0,6	0,3	0,1
0,1						

3

För att studera försäljningen av en viss cykel så har följande modeller anpassats på kvartalsdata i 4 år.

a) Tolka den skattade säsongseffekten för kvartal 4 i de båda modellerna. 2p

b) Tolka hur stor ökningen är per år i försäljning av cyklar med hjälp av trendkomponenten i modell 1. 1p

b) Prediktera försäljningen för kvartal 1 och 2 nästkommande år både med modell 1 och modell 2. 1p

c) Vilken modell anser du vara den bästa? Motivera 1p

Modell 1

Regression Analysis: y versus tid; kv2; kv3; kv4

```
The regression equation is
y = 8,75 + 0,500 tid + 21,0 kv2 + 33,5 kv3 + 4,50 kv4

Predictor        Coef    SE Coef        T        P
Constant       8,7500     0,4281    20,44    0,000
tid           0,50000    0,03769    13,27    0,000
kv2           21,0000     0,4782    43,91    0,000
kv3           33,5000     0,4827    69,41    0,000
kv4            4,5000     0,4900     9,18    0,000

S = 0,6742      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 = 2,20
```

Modell 2

Time Series Decomposition

```
Data       y
Length     16,0000
NMissing   0

Trend Line Equation

Yt = 22,2 + 0,652941*t
```

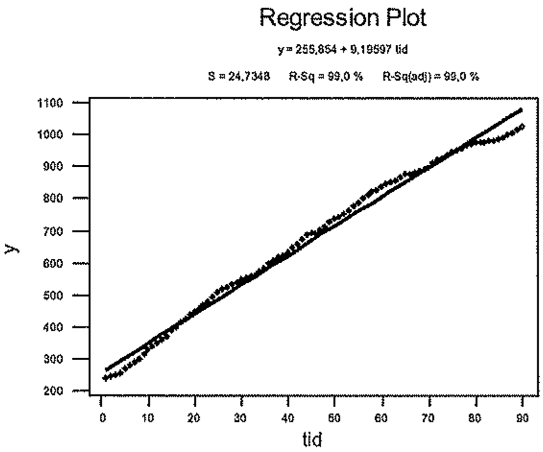
```
Seasonal Indices

Period     Index

   1       -14,8438
   2        6,40625
   3        18,4063
   4       -9,96875


Accuracy of Model

MAPE:      4,22616
MAD:       0,79210
MSD:       0,89107
```

4

I grafen nedan visas försäljning av en viss tandkräm under 90 veckor. En linjär trend är anpassad. För att undersöka om residualerna är oberoende så har Durbin-Watsons testfunktion beräknats.

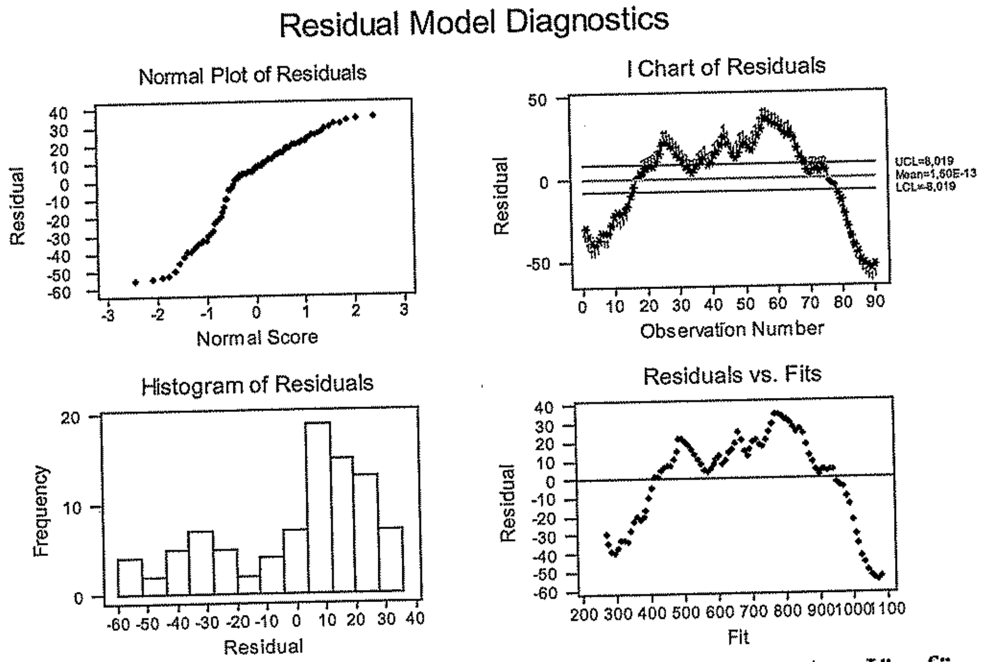
a) Vad kallas den typ av korrelation som Durbin-Watson testar? 0,5p

b) Ta hjälp av plotten med residualer mot anpassade värden och ställ upp den mothypotes som i detta fall är lämplig att använda i ett Durbin-Watson test. Kan vi förkasta nollhypotesen? 1p

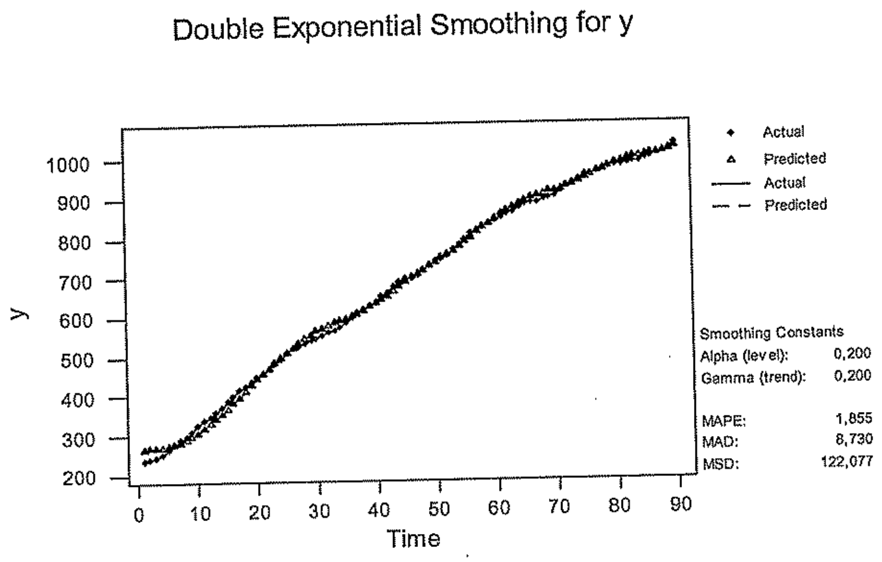


Regression Analysis: y versus tid

```
The regression equation is
y = 256 + 9,20 tid

Predictor        Coef     SE Coef          T        P
Constant      255,854       5,258      48,66    0,000
tid            9,1960      0,1004      91,63    0,000

S = 24,73       R-Sq = 99,0%      R-Sq(adj) = 99,0%

Analysis of Variance

Source           DF        SS         MS          F        P
Regression        1   5136739    5136739    8395,97    0,000
Residual Error   88     53839        612
Total            89   5190579


Durbin-Watson statistic = 0,02
```



c) Nedan har dubbel exponentiell utjämning anpassats. Jämför denna modell med den linjära trendmodellen ovan. Vilken modell föredrar du i syfte att göra en bra prognos? Vilka mått är lämpliga att jämföra i de båda modellerna? Motivera noggrant. 1,5p



d) Ge prognoser för tidpunkterna 91 och 92 för båda modellerna då du står vid tidpunkt 90.
Du får använda direkt att $S_{90} = 1021.67$ och $T_{90} = 4.8513$ 2p

Formelsamling
Index

Sammansatta fastbasindex:

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

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

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

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

Kedjeprisindex:

$I_t = \dfrac{L_{0,1}}{100} \cdot \dfrac{L_{1,2}}{100} \cdot \ldots \cdot \dfrac{L_{t-1,t}}{100} \cdot 100$

där

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

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

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

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

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

Implicitprisindex:

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

Relativprisindex:

$I_t^R = \dfrac{I_t^v}{I_t^0} \cdot 100$

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

Formelblad, tidsseriedelen av (formelnummer) och [sidnummer] enligt AJÅ.

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 .

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

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

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

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

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

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

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

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

2 Multipel linjär regression

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

3 Glidande medelvärden

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

4 Mått på prognosers noggrannhet

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

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

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

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

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

Symbolen i i de 5 formlerna ovan kommenteras under rubriken "Buggar i boken" i kursinfon 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

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

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

5.2 Linjär trend

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

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

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

5.3 Exponentiell trend

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

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

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

5.4 Linjär trend och multiplikativ säsong

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

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

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

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

5.5 Linjär trend och additiv säsong

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

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

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

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

6 Autokorrelation

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

7 Tabell över t-fördelningen

t-koefficienter vid dubbelsidiga intervall [283]

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