Linköpings Universitet
IDA/Statistik
LH

TENTAMEN I STATISTISK ANALYS AV SAMHÄLLSDATA, 2014-10-20

Skrivtid: kl: 8-12
Hjälpmedel: Räknedosa. Bowerman, B.J., O'Connell, R, Koehler, A.: *Forecasting, Time Series and Regression*, alla upplagor tillåtna, som inte får innehålla anteckningar men får ha markeringar och flärpar. Liten anteckning på flärp tillåten. Med tentan vidhäftad formelsamling.
Jourhavande lärare: Lotta Hallberg
Redovisa och motivera kort alla dina lösningar

Tolka (om möjligt) alla dina resultat!

1

I tabellen nedan visas totalförsäljningen (i löpande priser) för varugrupperna Husvagnar och Husbilar för ett visst företag, samt priser för husvagnstyp Kabbevagn från varugrupp Husvagnar och husbilstyp Kabbebil från varugrupp Husbilar under åren 2009 till 2011. (alla uppgifter är påhittade)

År	**Tot. förs. värde, mkr, husvagnar**	**Tot. förs. värde, mkr, husbilar**	**Pris, tkr, Kabbevagn**	**Pris, tkr, Kabbebil**
2009	*335*	*185*	*250*	*750*
2010	*350*	*190*	*290*	*800*
2011	*370*	*210*	*320*	*830*

a) Använd varorna Kabbevagn och Kabbebil som representantvaror för sina varugrupper och beräkna ett kedjeprisindex av Laspeyre-typ för företagets priser med basår 2009.
2p

b) Hur har företagets priser för husvagnar och husbilar förändrats mellan 2010 och 2011?
1p

Formler finner du längst bak i tentan.

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 ${}_n a_x = 0,5$. 3p

Formler:

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

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

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

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

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

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

3

Detta exempel är baserat på ett exempel i Vejde-Leander, Ordbok i statistik.
En sommarförsäljare av glass är intresserad av att göra bra prognoser för hur mycket glass som hon kommer att sälja olika dagar. Hon använder de första sex dagarna under försäljningsperioden som "testperiod". Hon noterar varje dag morgontemperaturen i grader och försåld mängd glass under dagen i hela liter. I tabellen återfinns data för fem dagar, dagarna 2 till 6. Försäljningen den första dagen var 42 liter. Morgontemperaturen den sjunde dagen var 22 grader.
Tabellen visar såld glassmängd och morgontemperatur för fem dagar i maj.

Dag, nummer	Temp, x	Glassmängd, y
2	14	48
3	12	35
4	18	60
5	16	56
6	25	71

Hon vill pröva två relativt enkla prognosmetoder, nämligen

<u>Prognosmetod 1</u>
Hon prognosticerar dagens försäljning med gårdagens.

<u>Prognosmetod 2</u>
Hon använder data som beskriver morgontemperaturen en viss dag och försåld mängd samma dag för att med hjälp av regressionsanalys skatta sambandet. Detta samband använder hon sedan för att utifrån morgontemperaturen varje dag skatta hur mycket hon kommer att sälja respektive dag.
Regressionssambandet framtaget med hjälp av Minitab ser ut som följer:

Regression Analysis: glass versus temp

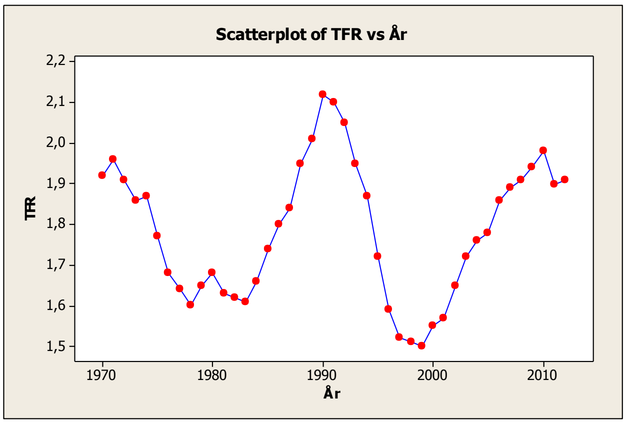
```
The regression equation is glass = 13,7 + 2,37 temp

Predictor         Coef         SE Coef       T        P
Constant          13,71        13,01         1,05     0,369
temp              2,3700       0,7401        3,20     0,049

S = 7,401         R-Sq = 77,4%
```

a) Beräkna de olika måtten för prognosfelet (MAD, MSE/MSD, MAPE) baserat på de fem dagarna 2 – 6 enligt <u>såväl prognosmetod 1 som prognosmetod 2</u>. 2p

c) Beräkna hur mycket glass försäljaren kan räkna med att sälja dag 7 med var och en av de två metoderna. Jämför också de båda metoderna. Vilken skulle Du välja för att prognosticera just dag 7 och varför? Fundera också på hur det skulle te sig på längre sikt. Glassförsäljaren vill ju göra prognos för varje dag för att kunna bedöma hur mycket glass hon skall ta med. och diskutera vilken som verkar vara rimligast. Diskutera! 2p

4

Nedan visas en tidsserie för TFR= Summerad fruktsamhet för hela riket Sverige åren 1970 till och med 2012.



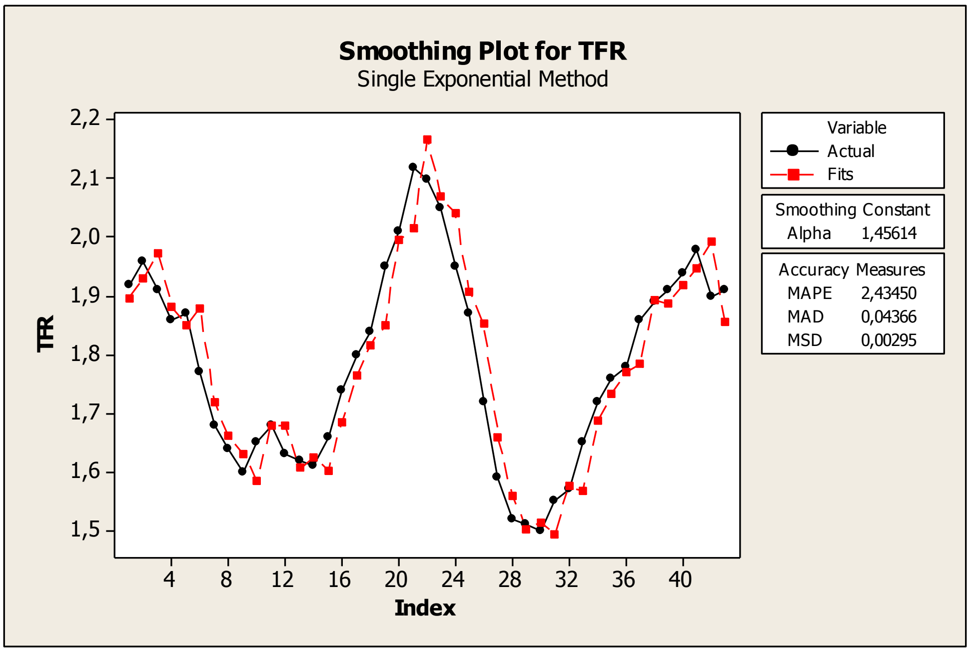
Nedan har tre modeller anpassats

a) Förklara hur TFR beräknas. 1p
b) Förklara varför modell 1 är sämre än modell 2. 1p
c) Beräkna prognos av TFR för 2013, 2014 samt 2015 med hjälp av modell 2. 1p
d) Hur väljer man startvärdet på utjämningsekvationen i modell 2 på ett lämpligt sätt? 1p
e) En Chi2-statistika i modell 3 har fått värde 23,6 (markerat med rött). Sätt upp vilka hypoteser man vill test med hjälp av denna statistika och utför testet. 2p

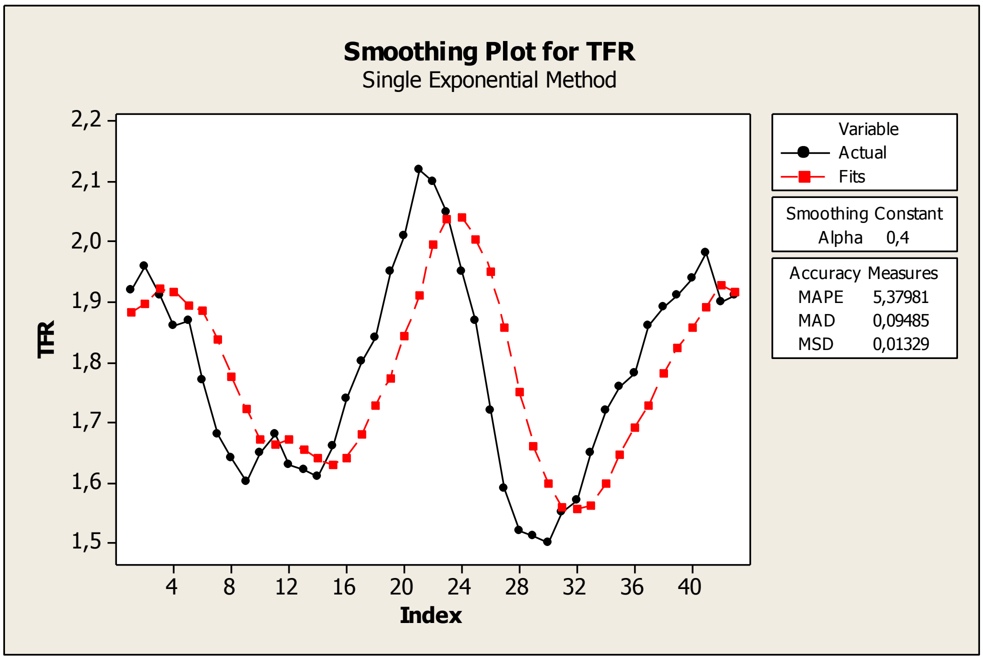
f) Utvärdera modellen med hjälp av utskriften nedan. Använd 5% signifikansnivå vid test. 2p

g) Beräkna prognos av TFR för 2013, 2014 samt 2015 med hjälp av modell 3. 2p

Modell 1



Modell 2



De sista värdena på utjämningsekvationen är:

1,85847
1,89108
1,92665
1,91599
1,91359

Modell 3
ARIMA Model: TFR

```
Final Estimates of Parameters

Type        Coef  SE Coef      T      P
MA   1  -0,4439   0,1515  -2,93  0,006
MA   2  -0,3577   0,1612  -2,22  0,032


Differencing: 1 regular difference
Number of observations:  Original series 43, after differencing 42
Residuals:    SS =  0,111903 (backforecasts excluded)
              MS =  0,002798  DF = 40


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            12     24     36   48
Chi-Square   23,6   35,5   48,4    *
DF             10     22     34    *
P-Value     0,009  0,034  0,052    *
```

De sista fyra värden i tidsserien:

TFR	Diff1	residualer	år
1,94	0,03	0,034511	2009
1,98	0,04	0,026484	2010
1,90	-0,08	-0,104098	2011
1,91	0,01	0,046732	2012