Preliminära lösningar till tentamensskrivning på kursen Tillämpad statistik A5 (15hp) 2015-08-25 /BW, AG, RP

Uppgift 1A) De 10 observationsparen är:

Föräljning (Y_t)	Order-2 (X_{t-2})
122	102
131	110
110	105
124	92
118	104
118	102
118	100
109	96
111	88
90	91

Modell: $Y_t = \alpha + \beta X_{t-2} + \varepsilon$. Från regressionsutskriften erhålls den skattade modellen som

Försäljning = 27,5 + 0,885 Order-2.

B) Ett 90% prognosintervall för försäljningen vid tidpunkt 13 ges av.

$$\hat{y} \pm t_{n-2,\alpha/2} \cdot s_e \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$, ε_i är $NID(0,\sigma_\varepsilon^2)$, t=1,86

n = 10, X_0 = 72, $\bar{X} = \frac{102+\cdots+91}{10} = 99$. $\sum(X-\bar{X})^2 = (102-99)^2 + \cdots + (91-99)^2 = 444$

$27{,}5 + 0{,}885 * 72 \pm 1{,}86 * 9{,}79185\sqrt{1+\frac{1}{10}+\frac{(72-99)^2}{444}}$, Intervallet blir 91,22 ± 30,16.

Med 90% sannolikhet hamnar försäljningen vid tidpunkt tretton i detta intervall.

C) Av figurerna framgår att sambandet för de aktuella observationerna är starkare mellan försäljning vid tidpunkt t och order vid tidpunkt t-1. Av regressionsutskrifterna framgår även att denna relation har klart högst förklaringsgrad.

Skattad modell: Försäljning = 23,2 + 0,957 Order-1

Ett 90% prognosintervall för försäljningen vid tidpunkt 13 ges av.

$$\hat{y} \pm t_{n-2,\alpha/2} \cdot s_e \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$, ε_i är $NID(0,\sigma_\varepsilon^2)$, t=1,833, n = 11, X_0 = 93,

$\bar{X} = \frac{102+\cdots+72}{10} = 96{,}545$. $\sum(X-\bar{X})^2 = \sum X^2 - n\bar{X}^2 = 103638 - 11 * 96{,}545^2$=1107,7

$23{,}2 + 0{,}957 * 93 \pm 1{,}833 * 3{,}71062\sqrt{1+\frac{1}{11}+\frac{(93-96{,}545)^2}{1107{,}7}}$, Intervallet blir 112,2 ± 7,14.

Med 90% sannolikhet hamnar försäljningen vid tidpunkt tretton i detta intervall.

Intervallet blir kortare av främst två skäl: 1) residualspridningen är avsevärt mindre än tidigare.

2) Värdet 93 ligger avsevärt närmare x-variabelns medelvärde.

Preliminära lösningar till tentamensskrivning på kursen Tillämpad statistik A5 (15hp) 2015-08-25 /BW, AG, RP

Uppgift 2

A. $H_0: \beta_1 = \beta_2 = \beta_3 = 0, H_1$: åtminstone ett β skilt från noll. Signifikansnivå: 5%

Testfunktion: $\frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$, som är F-fördelad med k=3 resp n-k-1=5 f.g., ε NID(0,σ_ε^2),

Förkasta nollhypotesen om $F_{obs} > F_{0,05,3,5} = 5{,}41$. $F_{obs} = 27{,}19$ med p-värde 0,002 enligt utskrift.

Resultatet är signifikant. Resultatet tyder på att modellen förklarar åtminstone någon del av variationen i y.

B. Ex:

1) Skattningarna av β_2 och β_3 är långt ifrån signifikanta.

2) Modellen i uppgift 1 baseras på 11 observationspar med 9 f.g. medan modellen i uppgift 2 baseras på 9 observationspar med enbart 5 f.g..

Uppgift 3.

A. Skattningarna av β_1: <u>Modell 1:</u> Rumsbeläggningen har under perioden ökat med i genomsnitt 17.6 enheter per kvartal.
Modell 2: Rumsbeläggningen har under perioden ökat med i genomsnitt 0.82% per kvartal.

Skattningarna av β_2: Modell 1: Rumsbeläggningen kvartal 2 ligger säsongmässigt i genomsnitt 299 enheter över kvartal 1.
Modell 2: ($e^{.149671}$≈1.16) Rumsbeläggningen kvartal 2 ligger säsongmässigt i genomsnitt 16% över rumsbeläggningen för kvartal 1(sedan hänsyn tagits till trenden).

B. Modell 1:

RoomAverQtak(57) = `1396.50 + 17.5503*57 + 0        = 2396.9`

RoomAverQtak(58) = `1396.50 + 17.5503*58 + 299.38   = 2713.8`

RoomAverQtak(59) = `1396.50 + 17.5503*59 + 670.26   = 3102.2`

RoomAverQtak(60) = `1396.50 + 17.5503*60 + 108.35   = 2557.9`

C. Av graferna framgår att tidsserien ser ut att ha en linjär trend med multiplikativ säsong (relativt konstanta svängningar i absoluta tal kring den ngt avtagande positiva trenden för lnRoomAverQ). Detta innebär att trenden är korrekt specificerad i modell 1 och prognoserna bör därför hamna rätt i genomsnitt. Dock kommer säsongsvängningarna under prognosperioden att underskattas eftersom modell 1 förutsätter att säsongvariationen är additiv. Man bör därför förvänta sig ett särskilt stort prognosfel för kvartal 3.

Preliminära lösningar till tentamensskrivning på kursen
Tillämpad statistik A5 (15hp) 2015-08-25 /BW, AG, RP

Uppgift 4

Normalfördelning kan inte förutsättas →icke-parametriskt alternativ.

Kruskal Wallis test

Lower class: i=1   *Middle class*: i=2   *Upper class*: i=3

$H_0: \eta_1 = \eta_2 = \eta_3$ $i = 1, 2, 3$

H_1:Alla η_i är ej lika

Signifikansnivå: $\alpha = 0{,}05$

Teststatistika: $H = \frac{12}{n(n+1)} \sum_{i=1}^{3} \frac{R_i^2}{n_i} - 3(n+1)$

Förutsättningar: *Grade point average* är mätt på minst ordinalskalenivå.
Stickproven är oberoende av varandra utvalda med OSU.
Grade point average har samma fördelning i alla tre populationerna.
Minst fem observationer i alla stickprov.

H_0 förkastas om $H > \chi^2_{a-1;\alpha} = \chi^2_{2;0{,}05} = 5{,}991$

Grade point average for three socioeconomic groups

Lower class		*Middle class*		*Upper class*	
1,80	*2*	2,78	*9*	1,36	*1*
2,16	*3,5*	2,97	*12*	2,25	*5*
2,16	*3,5*	3,01	*13,5*	2,44	*6*
2,51	*7*	3,23	*17*	2,54	*8*
2,87	*11*	3,45	*19*	2,81	*10*
3,01	*13,5*	3,53	*20*	3,13	*15*
3,14	*16*	3,77	*21*	3,27	*18*
□	56,5		111,5		63

GPAs till vänster och *rangordningstal till höger*.

$$H = \frac{12}{21(22)}\left(\frac{56{,}5^2}{7} + \frac{111{,}5^2}{7} + \frac{63^2}{7}\right) - 3(22) = 6{,}70$$

$6{,}70 > 5{,}991$ H_0 förkastas

Testresultatet tyder, på 5% signifikansnivå, på att de tre socioekonomiska grupperna inte har samma *GPAs*.

Uppgift 5

Kommer senare