

**Preliminära lösningar till tentamensskrivning på kursen
Tillämpad statistik A5 (15hp) 2015-01-13 /BW, AG, RP**

Uppgift 1

A) Tolkning $b=746$: Månadshyran för rum med egen ingång uppskattas att vara i genomsnitt 746 kr större än för rum utan egen ingång.

B) $H_0: \beta = 0, H_1: \beta > 0$ Signifikansnivå: 5%

Testfunktion: $t = \frac{b-0}{s_b}$ som är t-fördelad med 40-2 f.g., $\varepsilon \text{NID}(0, \sigma_\varepsilon^2)$,

Förkasta nollhypotesen om $t_{\text{obs}} > \approx 1,69$

$t_{\text{obs}} = 4,22$. Skattningen 746 är signifikant > 0 . Följaktligen tror vi att β är större än noll och att därmed egen ingång höjer priset på rum i den bakomliggande populationen.

(Alternativt ser vi att parameterskattningen > 0 med ett p-värde $= 0,000/2 < \text{signifikansnivån}$)

C) Parametern α är den genomsnittliga hyran för rum utan egen ingång i den bakomliggande populationen. Ett 95% konfidensintervall för α ges av:

$$a \pm t^* s_a \quad \varepsilon \text{NID}(0, \sigma_\varepsilon^2).$$

$$3116,7 \pm 2,02 \cdot 111,7 \quad \text{eller } (2891 ; 3342)$$

Med 95% konfidens täcker detta intervall α .

(Alternativ lösning: $y_{\text{tak}} = 3117, s = 547,021, x_0 = 0, \bar{x} = 16/40, \sum (x_i - \bar{x})^2 = 16 \cdot 16^2 / 40 = 9,6$

$$\hat{\mu}_{y|x_0} \pm t_{n-2, \alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad \text{os v.)}$$

Uppgift 2 A)

Om variabeln standard i modell 3 inte bidrar till att förklara variationen i hyra måste parametrarna β_3, β_4 och β_5 alla vara noll.

$$H_0: \beta_3 = \beta_4 = \beta_5 = 0 \quad H_1: \text{Minst en av } \beta_j \neq 0 \quad j=3,4,5$$

Tre restriktioner $\rightarrow r=3$ Signifikansnivå: $\alpha = 0,05$

$$\text{Testfunktion: } F = \frac{(SSE_R - SSE_{UR}) / r}{SSE_{UR} / (n - k_{UR} - 1)} \sim F_{r, (n-k-1)}$$

Där k = antal förklarande variabler i Modell 3.

Förkastelseområde: H_0 förkastas då $F_{\text{obs}} > F_\alpha(r; n-k-1) = F_{0,05}(3; 34) = (2,84 - 2,92)$

$$\text{Resultat: } F_{\text{obs}} = \frac{(0,14908 - 0,06669) / 3}{0,06669 / 34} \approx 14,0 > 2,92$$

(14,0 kan även erhållas från variansanalystabellen i version 17, rad standard).

Signifikant resultat. Resultatet tyder på att åtminstone en av de aktuella parametrarna är skild från noll.

**Preliminära lösningar till tentamensskrivning på kursen
Tillämpad statistik A5 (15hp) 2015-01-13 /BW, AG, RP**

B) $e^{0,2284} = 1,2566$. $3000 * 0,2566 = 769,8$. Svar: Ca 770kr mer.

C)
$$R_{adj}^2 = 1 - \frac{s_e^2}{s_y^2} = 1 - \frac{SSE / (n - k - 1)}{SST / (n - 1)} = 1 - \frac{0,00403}{1,35235 / 39} = 0,884$$

Uppgift 3

A) Antal observationspar = 35-4=31. Obspar 1: $Y_5 = 386,5$; $Y_1 = 410,5$ Obspar 31: $Y_{35} = 378,5$; $Y_{31} = 382,9$

B) Prognos för y_{36} : $122,7 + 0,681 * 433,1 = 417,6$

Prognos för y_{37} : $122,7 + 0,681 * 451,6 = 430,2$

Prognos för y_{38} : $122,7 + 0,681 * 380,2 = 381,6$

C)	t	Utfall	Prognos	Prognosfel	Absolut prognosfel
	36	436,6	417,6	19,0	19,0
	37	453,6	430,2	23,4	23,4
	38	379,2	381,6	-2,4	2,4
			Medelvärde	13,3	14,9

I genomsnitt ligger prognoserna under perioden ca 13000 personer för lågt. Det absoluta prognosfelet är i genomsnitt ca 15000 personer. Svårt att uttala sig om prognosfelen är stora eller inte då ingen direkt jämförelse finns. (Det förtjänar att påpekas att utfallen utgörs av stickprovsuppskattningar, där felmarginalen oktober 2014 var ca 21000 personer. obs månadsdata.)

D) Vi noterar att DW-statistikan är mycket låg, varför det finns information i tidsserien som ej fångas upp i den använda modellen. Modellen bör gå att förbättra.

(DW-statistikan tolkas med försiktighet vid autoregressiva modeller)

**Preliminära lösningar till tentamensskrivning på kursen
Tillämpad statistik A5 (15hp) 2015-01-13 /BW, AG, RP**

Uppgift 4

Envägs variansanalys med träningsprogram som faktor.

$$\text{Modell: } x_{ij} = \mu + \alpha_i + e_{ij} \quad i = A, B, C, D \quad j = 1, 2, 3, 4, 5, 6, 7$$

$$\text{Förutsättningar: } e_{ij} = NID(0, \sigma_e^2)$$

$$\begin{aligned} \text{Hypoteser: } H_0: \mu_A = \mu_B = \mu_C = \mu_D & \quad \text{alt. } H_0: \alpha_A = \alpha_B = \alpha_C = \alpha_D = 0 \\ H_1: \text{Alla } \mu_i \text{ är ej lika} & \quad H_1: \text{Alla } \alpha_i \text{ är ej lika med 0} \end{aligned}$$

Signifikansnivå: $\alpha = 0,05$

$$\text{Testfunktion: } F_{obs} = \frac{MSA}{MSB}$$

Beslutsregel: H_0 förkastas om $F_{obs} > F_{3,24;0,05}$

$$F_{3,25;0,05} = 2,99 < F_{3,24;0,05} < 3,10 = F_{3,20;0,05}$$

$$\text{Beräkningar: } A = \sum_{i=1}^4 \sum_{j=1}^7 x_{ij}^2 = 6359915$$

$$B = \frac{(\sum_{i=1}^4 \sum_{j=1}^7 x_{ij})^2}{n} = \frac{13325^2}{28} \approx 6341272,32$$

$$C = \sum_{i=1}^4 \frac{(\sum_{j=1}^7 x_{ij})^2}{7} = \frac{3267^2 + 3164^2 + 3500^2 + 3394^2}{7} = \frac{44453421}{7} \approx 6350488,71$$

$$SST = A - B \approx 18642,68 \quad SSE = A - C \approx 9426,29 \quad SSA = C - B \approx 9216,39$$

ANOVA tabell

Variationsorsak	Frihetsgrader	Kvadratsummor	Medelkvadratsummor	F-test
Träningsprogram	3	9216,39	3072,1	7,82
Error	24	9426,29	392,8	
Totalt	27	18642,68		

Beslut: $F_{obs} = 7,82 > 3,10 > F_{3,24;0,05} \quad H_0$ förkastas

Slutsats: Testresultatet tyder, på 5% signifikansnivå, på att de fyra träningsprogrammen i genomsnitt skiljer sig åt i effektivitet.

Uppgift 5: Se lösning uppgift 701.