

## Tentamen *Tillämpad statistik A5 (15hp) 2015-01-13*

### UPPLYSNINGAR

**A. Tillåtna hjälpmedel:**

Miniräknare

Formelsamlingar: *A4/A8 Tabell- och formelsamling samt  
A5 Kompletterande Tabell- och formelsamling.*  
Inga anteckningar är tillåtna i formelsamlingarna.

- B. Skrivtid: 8<sup>00</sup>-13<sup>00</sup> Skrivningen omfattar 5 uppgifter om sammanlagt 100 poäng.
- C. För varje uppgift anges den maximala poäng som kan erhållas. Om en uppgift är uppdelad på deluppgifter anges den maximala poängen för varje deluppgift. Ibland kan inte deluppgifterna bedömas oberoende av varandra, vilket kan innebära att poäng inte utdelas på en senare uppgift om inte tidigare deluppgift lösts på ett i princip riktigt sätt. Dock gäller att utdelad poäng för varje deluppgift aldrig kan vara negativ.
- D. Om Du känner Dig osäker på någonting (skrivningens genomförande, någon formulering i en uppgift, om något hjälpmedel är otillåtet), fråga då jourhavande skrivningsvakt eller den skrivningsansvariga läraren (besök, alternativt telefon).
- E. Efter skrivningens slut får Du behålla sidorna med frågeställningarna. Preliminära lösningar anslås på Pingpong.

### UPPMANINGAR

- A. Följ noga de anvisningar som finns på skrivningsförsättsbladet.
- B. Redovisa Dina lösningar i en form som gör det lätt att följa Din tankegång! (Det dunkelt uttryckta förutsätts av rättaren vara dunkelt tänkt). Motivera alla väsentliga steg i lösningen. Ange alla antaganden Du gör och alla förutsättningar Du utnyttjar.
- C. Vid konfidensintervall måste Du dessutom ange vad intervallet avser att täcka samt teckna intervallet i symbolform innan de numeriska uppgifterna insätts. Verbal slutsats av det framräknade intervallet krävs för full poäng.
- D. Vid hypotestest måste Du utöver vad som sägs i punkt B ovan ange  $H_0$ ,  $H_1$ , signifikansnivå, testfunktion (inklusive antal frihetsgrader), förkastelseområde, resultat och verbal slutsats.
- E. Vid variansanalys måste Du utöver vad som sägs ovan ange modell.

## Uppgift 1 och 2

På sista sidan i skrivningen finner du uppgifter om fyrtio hyresrum i centrala Uppsala. Uppgifterna består av priser (hyra kr per månad), storlek (kvm), om rummen har "egen ingång" (1) eller inte (0) och standard. Dessa data utgör underlag för uppgift 1 och 2.

### Uppgift 1 (22 p)

I denna uppgift analyseras modell 1:  $hyra = \alpha + \beta Egen\ ingång + \varepsilon$   
Skattning av modellen ger följande resultat:

### Regression Analysis: Hyra versus Egen ingång

The regression equation is  
Hyra = 3117 + 746 Egen ingång

Predictor	Coef	SE Coef	T	P
Constant	3116,7	111,7	27,91	0,000
Egen ingång	745,8	176,6	4,22	0,000

S = 547,021    R-Sq = 32,0%    R-Sq(adj) = 30,2%

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	1	5340167	5340167	17,85	0,000
Residual Error	38	11370833	299232		
Total	39	16711000			

- Tolka parameterskattningen 746 (745,8).
- Genomför ett formellt test av huruvida egen ingång höjer det genomsnittliga priset på rum i den bakomliggande populationen.
- Beräkna ett 95% konfidensintervall för genomsnittlig hyra för rum utan egen ingång ( $\mu_{Hyra|Egen\ ingång=0}$ ).

### Uppgift 2 (22)

I denna uppgift skall du analysera följande två modeller

Modell 2:  $lnhyra = \alpha + \beta_1 lnstorlek + \beta_2 Egen\ ingång + \varepsilon$

Modell 3:  $lnhyra = \alpha + \beta_1 lnstorlek + \beta_2 Egen\ ingång + \beta_3 D_2 + \beta_4 D_3 + \beta_5 D_4 + \varepsilon$

Skattningar av modellerna finns på nästföljande två sidor dels med Minitab version 17, dels med version 16. Du väljer själv vilken version du vill använda.

- Genomför ett formellt test av huruvida variabeln standard i Modell 3 bidrar till att förklara variationen i hyra.
- Hur mycket högre hyra i kr skulle ett rum utan egen ingång och med given standard som kostar 3000:- i månaden ha enligt den skattade modellen 3 om det hade egen ingång?
- Beräkna  $R^2_{adj}$  för modell 2.

## MINITAB 17

### Regression Analysis: ln(Hyra) versus ln(Storlek); Egen ingång

#### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1,20328	0,601639	149,32	0,000
ln(Storlek)	1	0,73944	0,739440	183,53	0,000
Egen ingång	1	0,54224	0,542235	134,58	0,000
Error	37	0,14908	0,004029		
Lack-of-Fit	14	0,04890	0,003493	0,80	0,659
Pure Error	23	0,10018	0,004356		
Total	39	1,35235			

#### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0634751	88,98%	xxxxx	87,13%

#### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	6,8993	0,0845	81,62	0,000	
ln(Storlek)	0,4483	0,0331	13,55	0,000	1,00
Egen ingång	0,2382	0,0205	11,60	0,000	1,00

#### Regression Equation

$\ln(\text{Hyra}) = 6,8993 + 0,4483 \ln(\text{Storlek}) + 0,2382 \text{ Egen ingång}$

### Regression Analysis: ln(Hyra) versus ln(Storlek); Egen ingång; Standard

Method

Categorical predictor coding (1; 0)

#### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	1,28566	0,257133	131,09	0,000
ln(Storlek)	1	0,60875	0,608752	310,36	0,000
Egen ingång	1	0,44967	0,449666	229,25	0,000
Standard	3	0,08239	0,027463	14,00	0,000
Error	34	0,06669	0,001961		
Lack-of-Fit	19	0,04238	0,002230	1,38	0,268
Pure Error	15	0,02431	0,001621		
Total	39	1,35235			

#### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0442881	95,07%	94,34%	93,04%

#### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	6,9769	0,0609	114,59	0,000	
ln(Storlek)	0,4192	0,0238	17,62	0,000	1,07
Egen ingång	0,2284	0,0151	15,14	0,000	1,11
Standard					
Låg standard	-0,0252	0,0190	-1,33	0,194	1,38
Mycket hög	0,0831	0,0202	4,11	0,000	1,33
Tveksam	-0,0555	0,0197	-2,82	0,008	1,27

#### Regression Equation

Standard

Ganska hög  $\ln(\text{Hyra}) = 6,9769 + 0,4192 \ln(\text{Storlek}) + 0,2284 \text{ Egen ingång}$

Låg standard  $\ln(\text{Hyra}) = 6,9517 + 0,4192 \ln(\text{Storlek}) + 0,2284 \text{ Egen ingång}$

Mycket hög  $\ln(\text{Hyra}) = 7,0600 + 0,4192 \ln(\text{Storlek}) + 0,2284 \text{ Egen ingång}$

Tveksam  $\ln(\text{Hyra}) = 6,9214 + 0,4192 \ln(\text{Storlek}) + 0,2284 \text{ Egen ingång}$

## MINITAB 16

### Regression Analysis: ln(Hyra) versus ln(Storlek); Egen ingång

The regression equation is

$$\ln(\text{Hyra}) = 6,90 + 0,448 \ln(\text{Storlek}) + 0,238 \text{ Egen ingång}$$

Predictor	Coef	SE Coef	T	P
Constant	6,89930	0,08452	81,62	0,000
ln(Storlek)	0,44825	0,03309	13,55	0,000
Egen ingång	0,23818	0,02053	11,60	0,000

S = 0,0634751    R-Sq = 89,0%    R-Sq(adj) = xxxxx

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	1,20328	0,60164	149,32	0,000
Residual Error	37	0,14908	0,00403		
Total	39	1,35235			

Source	DF	Seq SS
ln(Storlek)	1	0,66104
Egen ingång	1	0,54224

### Regression Analysis: ln(Hyra) versus ln(Storlek); Egen ingång; ...

The regression equation is

$$\ln(\text{Hyra}) = 6,95 + 0,419 \ln(\text{Storlek}) + 0,228 \text{ Egen ingång} - 0,0303 \text{ D2} \\ + 0,0252 \text{ D3} + 0,108 \text{ D4}$$

Predictor	Coef	SE Coef	T	P
Constant	6,95167	0,06372	109,09	0,000
ln(Storlek)	0,41919	0,02379	17,62	0,000
Egen ingång	0,22839	0,01508	15,14	0,000
D2	-0,03025	0,02195	-1,38	0,177
D3	0,02523	0,01903	1,33	0,194
D4	0,10833	0,02112	5,13	0,000

S = 0,0442881    R-Sq = 95,1%    R-Sq(adj) = 94,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	1,28566	0,25713	131,09	0,000
Residual Error	34	0,06669	0,00196		
Total	39	1,35235			

### Uppgift 3 (20)

På basis av data från och med 2005:2 till och med 2013:4 över antal arbetslösa (se tabell nedan) har Modellen  $\text{Arbetslösa}(t) = \alpha + \beta \text{Arbetslösa}(t-4) + \varepsilon_t$  skattats. Regressionsutskrift presenteras nedan. Källa AKU,SCB.

- A. Hur många observationspar baseras regressionen på? Skriv ut det första och det sista observationsparet som används vid regressionskattningen.
- B. Använd den skattade modellen till att göra prognoser för kvartal 1, 2 och 3 2014.
- C. Hur har prognoserna slagit? Kommentera kortfattat. Som grund för dina kommentarer skall du beräkna genomsnittligt prognosfel och genomsnittligt absolut prognosfel.

Utfall 2014: Kvartal 1: 436,6      Kvartal 2: 453,6      Kvartal 3: 379,2

- D. Utifrån given information i regressionsutskriften tror du att modellen går att förbättra?

#### Data: Antal arbetslösa (tusental) kvartal 2 2005 till och med kvartal 4 2013

t	År	Kvartal	Arbetslösa(t)
1	2005	2	410,5
2	2005	3	344,5
3	2005	4	333,9
4	2006	1	367,5
5	2006	2	386,5
6	2006	3	306,9
7	2006	4	284,4
8	2007	1	322,7
9	2007	2	334,5
-----			
29	2012	2	439,6
30	2012	3	383,4
31	2012	4	382,9
32	2013	1	433,1
33	2013	2	451,6
34	2013	3	380,2
35	2013	4	378,5

#### Regression Analysis: Arbetslösa(t) versus Arbetslösa(t-4)

Rows unused 4

##### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
45,7103	43,14%	41,18%	35,16%

##### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	122,7	53,8	2,28	0,030	
Arbetslösa(t-4)	0,681	0,145	4,69	0,000	1,00

##### Regression Equation

$\text{Arbetslösa}(t) = 122,7 + 0,681 \text{Arbetslösa}(t-4)$

##### Durbin-Watson Statistic

Durbin-Watson Statistic = 0,265656

#### Uppgift 4 (16)

En amerikansk fysiolog ville jämföra fyra träningsprogram avsedda för ”business executives”. Tjugoåtta personer fördelades slumpmässigt till de fyra träningsprogrammen A, B, C och D. Nedanstående tabell visar förändringen för ett mått på fysisk förmåga från före till efter deltagande i respektive program.

	A	B	C	D	Summa
	434	447	472	448	1 801
	473	465	499	502	1 939
	484	462	524	458	1 928
	457	441	500	489	1 887
	472	443	482	490	1 887
	455	452	495	481	1 883
	492	454	528	526	2 000
Summa	3 267	3 164	3 500	3 394	13 325

Summan av alla kvadrerade observationer är 6 359 915.

Tyder dessa data på att träningsprogrammen i genomsnitt skiljer sig åt i effektivitet? Använd 5% signifikansnivå. Summor som kan vara användbara redovisas i och i samband med tabellen.

### Uppgift 5 (20)

I ett nybyggt höghusområde önskar man uppskatta daghemsbehovet. Man gör därför ett OSU (utan återläggning) av 100 lägenheter och frågar hur många daghemsplatser som invånarna i de uttagna lägenheterna är i behov av. Totala antalet lägenheter i området är 800. Resultatet blev:

Antal önskade daghemsplatser	Antal lägenheter
0	56
1	34
2	7
3	3

- A. Definiera population, element och urvalsenhet.
- B. Beräkna ett 95% konfidensintervall för det totala antalet önskade daghemsplatser i området.
- C. Beräkna ett 95% konfidensintervall för antalet lägenheter i området i vilka invånarna inte önskar daghems plats alls (d v s i vilka invånarna önskar 0 daghemsplatser).

## Data Display för uppgift ett och två

Row	Månads		ln		Egen	Standard	D2	D3	D4
	Hyra	Storlek	(Hyra)	(Storlek)	ingång				
1	4100	10	8,31874	2,30259	1	Mycket hög	0	0	1
2	3150	15	8,05516	2,70805	0	Tveksam	1	0	0
3	2450	8	7,80384	2,07944	0	Tveksam	1	0	0
4	3700	12	8,21609	2,48491	1	Låg standard	0	0	0
5	4400	18	8,38936	2,89037	1	Låg standard	0	0	0
6	4900	24	8,49699	3,17805	1	Ganska hög	0	1	0
7	3100	12	8,03916	2,48491	0	Ganska hög	0	1	0
8	2600	10	7,86327	2,30259	0	Ganska hög	0	1	0
9	2900	12	7,97247	2,48491	0	Ganska hög	0	1	0
10	3400	14	8,13153	2,63906	0	Ganska hög	0	1	0
11	3600	10	8,18869	2,30259	1	Låg standard	0	0	0
12	3400	10	8,13153	2,30259	1	Låg standard	0	0	0
13	4000	15	8,29405	2,70805	0	Mycket hög	0	0	1
14	4500	23	8,41183	3,13549	0	Mycket hög	0	0	1
15	2600	8	7,86327	2,07944	0	Ganska hög	0	1	0
16	2800	9	7,93737	2,19722	0	Ganska hög	0	1	0
17	3200	15	8,07091	2,70805	0	Låg standard	0	0	0
18	2950	13	7,98956	2,56495	0	Tveksam	1	0	0
19	3800	10	8,24276	2,30259	1	Mycket hög	0	0	1
20	3450	10	8,14613	2,30259	1	Tveksam	1	0	0
21	3700	10	8,21609	2,30259	1	Mycket hög	0	0	1
22	3050	15	8,02290	2,70805	0	Tveksam	1	0	0
23	2450	8	7,80384	2,07944	0	Tveksam	1	0	0
24	3000	12	8,00637	2,48491	0	Låg standard	0	0	0
25	3500	18	8,16052	2,89037	0	Låg standard	0	0	0
26	4400	24	8,38936	3,17805	0	Ganska hög	0	1	0
27	3700	12	8,21609	2,48491	1	Ganska hög	0	1	0
28	2600	10	7,86327	2,30259	0	Ganska hög	0	1	0
29	3100	12	8,03916	2,48491	0	Ganska hög	0	1	0
30	4200	14	8,34284	2,63906	1	Ganska hög	0	1	0
31	3300	10	8,10168	2,30259	1	Låg standard	0	0	0
32	3600	10	8,18869	2,30259	1	Låg standard	0	0	0
33	3500	15	8,16052	2,70805	0	Mycket hög	0	0	1
34	5100	23	8,53700	3,13549	1	Mycket hög	0	0	1
35	2600	8	7,86327	2,07944	0	Ganska hög	0	1	0
36	3400	9	8,13153	2,19722	1	Ganska hög	0	1	0
37	3200	15	8,07091	2,70805	0	Låg standard	0	0	0
38	2850	13	7,95507	2,56495	0	Tveksam	1	0	0
39	2900	10	7,97247	2,30259	0	Mycket hög	0	0	1
40	3450	10	8,14613	2,30259	1	Tveksam	1	0	0