

Tentamen Tillämpad statistik A5 (15hp) 2015-08-25

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: 8⁰⁰-13⁰⁰ 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

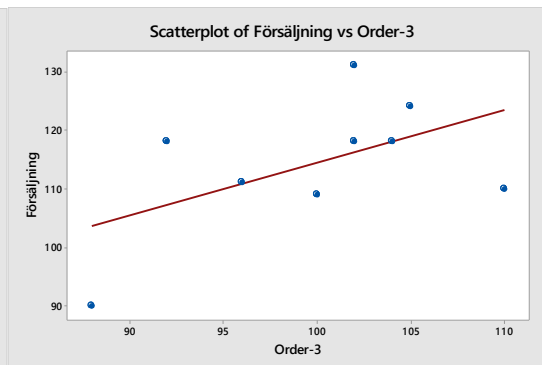
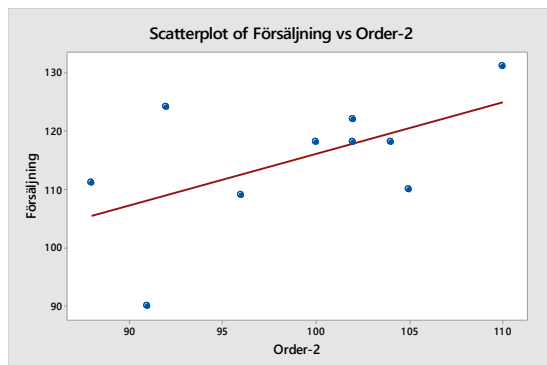
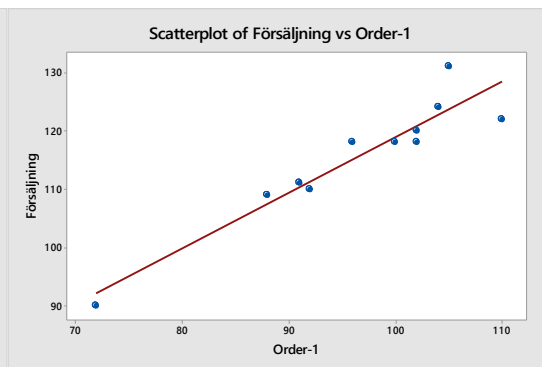
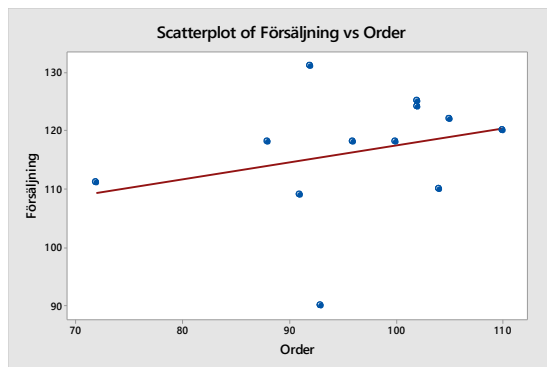
- A. Följ noga de anvisningar som finns på skrivningsförsättsbladet.
- B. Redovisa Dina lösningar i en form som gör det lätt att följa Din tankegång! (Det dunkelt uttryckta förutsätts av rättaren vara dunkelt tänkt). Motivera alla väsentliga steg i lösningen. Ange alla antaganden Du gör och alla förutsättningar 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(24 poäng)

Följande data är hämtade från ett grossistföretags årsredogörelse. Tidsserierna utgörs av kvartalsdata och är angivna i miljoner kronor.

Row	t	Försäljning	Order	Order-1	Order-2	Order-3
1	1	125	102	*	*	*
2	2	120	110	102	*	*
3	3	122	105	110	102	*
4	4	131	92	105	110	102
5	5	110	104	92	105	110
6	6	124	102	104	92	105
7	7	118	100	102	104	92
8	8	118	96	100	102	104
9	9	118	88	96	100	102
10	10	109	91	88	96	100
11	11	111	72	91	88	96
12	12	90	93	72	91	88

- A. Ange för minsta kvadratmetoden parameterskattningarna i en modell där försäljningen beror linjärt av orderingången ett halvår tidigare. Ange observationsparen och den skattade modellen. För resultat, se regressionsutskrifter på nästa sida.
- B. Beräkna ett 90% prognosintervall för försäljningen vid tidpunkt $t=13$ med hjälp av resultaten för modellen i uppgift A.
- C. Studera nedanstående figurer och avgör huruvida den i A-uppgiften valda modellen är den "bästa". Om ditt svar är nej upprepa uppgift A och B med en Du anser bättre modell av de skattade. Varför blir prognosintervallet kortare i C-uppgiften än i B-uppgiften?



Regression Analysis: Försäljning versus Order

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
10,5946	7,89%	0,00%	0,00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	88,1	30,6	2,87	0,017	
Order	0,293	0,317	0,93	0,376	1,00

Regression Equation

Försäljning = 88,1 + 0,293 Order

Regression Analysis: Försäljning versus Order-1

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3,71062	89,10%	87,89%	81,26%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	23,2	10,8	2,14	0,061	
Order-1	0,957	0,112	8,58	0,000	1,00

Regression Equation

Försäljning = 23,2 + 0,957 Order-1

Regression Analysis: Försäljning versus Order-2

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
9,79185	31,20%	22,60%	0,00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	27,5	46,1	0,60	0,568	
Order-2	0,885	0,465	1,90	0,093	1,00

Regression Equation

Försäljning = 27,5 + 0,885 Order-2

Regression Analysis: Försäljning versus Order-3

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
10,4104	28,57%	18,36%	0,00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	24,2	54,0	0,45	0,667	
Order-3	0,902	0,539	1,67	0,138	1,00

Regression Equation

Försäljning = 24,2 + 0,902 Order-3

Uppgift 2 (16 poäng)

Med användande av samma data som i uppgift 1 skattades modellen

$Försäljning_t = \alpha + \beta_1 \text{Order-1} + \beta_2 \text{Order-2} + \beta_3 \text{Order-3} + \varepsilon_t$ med följande resultat:

Regression Analysis: Försäljning versus Order-1; Order-2; Order-3

Method

Rows unused 3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	1000,67	333,557	27,19	0,002
Order-1	1	558,42	558,422	45,53	0,001
Order-2	1	0,55	0,548	0,04	0,841
Order-3	1	1,21	1,208	0,10	0,766
Error	5	61,33	12,266		
Total	8	1062,00			

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
3,50223	94,23%	90,76%	73,45%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5,1	20,9	0,24	0,817	
Order-1	1,041	0,154	6,75	0,001	1,66
Order-2	0,043	0,202	0,21	0,841	1,45
Order-3	0,067	0,215	0,31	0,766	1,41

Regression Equation

Försäljning = 5,1 + 1,041 Order-1 + 0,043 Order-2 + 0,067 Order-3.

- A. Har denna modell något förklaringsvärde vad avser försäljning? Genomför ett formellt test på 5% signifikansnivå.
- B. Ange två skäl till varför den enligt uppgift 1 bästa modellen kan vara att föredra framför modellen i denna uppgift. För poäng på deluppgiften krävs att ditt svar enbart innehåller två skäl och en mening per skäl.

Uppgift 3 (20 poäng)

Till en tidsserie bestående av 56 kvartalsvisa observationer på "hotellrumsbeläggning" för en viss hotellkedja har nedanstående modeller anpassats

$$\text{Modell 1: RoomAverQ}_t = \alpha + \beta_1 t + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \varepsilon_t$$

$$\text{Modell 2: lnRoomAverQ}_t = \alpha + \beta_1 t + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \varepsilon_t$$

där $t = 1, 2, 3, \dots, 56$ och D_{jt} ($j = 2, 3, 4$) är dummyvariabler som antar värdet 1 vid kvartal j och 0 vid övriga kvartal. "ln" är den naturliga logaritmen. Grafiska illustrationer presenteras på nästa sida och regressionsutskriften nedan.

- Tolka skattningarna av β_1 och β_2 i de båda modellerna.
- Gör prognoser för hotellrumsbeläggningen med den skattade modellen 1 för varje kvartal nästa år ($t=57, 58, 59$ och 60).
- Hur kommer dessa prognoser att slå med tanke på tidsseriens mönster.

Regression Analysis: RoomAverQ versus tid, D2t, D3t, D4t

The regression equation is

$$\text{RoomAverQ} = 1396 + 17.6 \text{ tid} + 299 D_{2t} + 670 D_{3t} + 108 D_{4t}$$

Predictor	Coef	SE Coef	T	P
Constant	1396.50	16.21	86.13	0.000
tid	17.5503	0.3855	45.52	0.000
D2t	299.38	17.59	17.02	0.000
D3t	670.26	17.60	38.08	0.000
D4t	108.35	17.62	6.15	0.000

S = 46.52 R-Sq = 98.7% R-Sq(adj) = 98.6%

Regression Analysis: lnRoomAverQ versus tid, D2t, D3t, D4t

The regression equation is

$$\text{lnRoomAverQ} = 7.30 + 0.00818 \text{ tid} + 0.150 D_{2t} + 0.303 D_{3t} + 0.0567 D_{4t}$$

Predictor	Coef	SE Coef	T	P
Constant	7.30441	0.00539	1356.05	0.000
tid	0.0081819	0.0001281	63.89	0.000
D2t	0.149671	0.005842	25.62	0.000
D3t	0.302949	0.005846	51.82	0.000
D4t	0.056696	0.005854	9.69	0.000

S = 0.01545 R-Sq = 99.3% R-Sq(adj) = 99.3%

Uppgift 4 (16 poäng)

Sociologer i USA undersöker ofta sambandet mellan socioekonomisk status och *college performance*. Socioekonomisk status delas i allmänhet in i tre grupper: *lower class*, *middle class* och *upper class*. *Grade point averages (GPAs)* för slumpmässigt utvalda stickprov om sju college freshmen från vardera av de tre socioekonomiska grupperna plockades fram i slutet av det akademiska läsåret och redovisas nedan.

Grade point average for three socioeconomic groups

<i>Lower class</i>	<i>Middle class</i>	<i>Upper class</i>
1,80	2,78	1,36
2,16	2,97	2,25
2,16	3,01	2,44
2,51	3,23	2,54
2,87	3,45	2,81
3,01	3,53	3,13
3,14	3,77	3,27

Undersök om de tre socioekonomiska grupperna har samma *GPAs*. Normalfördelning kan inte förutsättas. Använd 5% signifikansnivå.

Uppgift 5 (24 poäng)

En kommun ville få en uppfattning om antalet parkerade bilar i en stad vid en viss tidpunkt. För att undersöka detta genomfördes ett systematiskt urval genom att lägga ett rutnät bestående av 400 rutor över kommunkartan. Sedan valdes systematiskt, med slumpmässig startruta, var 10:e ruta.

Antalet parkerade bilar (x) i de valda rutorna var:

8 12 13 16 12 9 6 8 10 12
12 3 5 8 9 13 12 3 13 14
7 7 15 8 12 9 10 9 10 11
8 11 9 5 13 11 12 15 12 6

$$\Sigma x = 398, \Sigma x^2 = 4356$$

- Ge en punktskattning för antalet parkerade bilar i staden.
- Beräkna ett 90-procentigt konfidensintervall för antalet parkerade bilar i staden. Tolka intervallet! Var noga med förutsättningar!
- Grannkommunen önskar genomföra en motsvarande undersökning. Staden är av ungefär samma storlek och rutnätet består därför av samma antal rutor. Om vi antar att spridningen vad gäller antalet parkerade bilar är samma i båda städerna, hur många rutor ska kommunen då välja om felmarginalen för det totala antalet bilar får vara högst 200 bilar.

Illustration av ett rutnät

