

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2019-02-04
Sal (2)	TER1(29) <u>TERF(1)</u>
Tid	8-12
Utb. kod	732G71
Modul	TENT
Utb. kodnamn/benämning Modulnamn/benämning	Statistik B Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Annika Tillander
Telefon under skrivtiden	013-28 12 14
Besöker salen ca klockan	ca kl 10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Annelie Almquist 013-282934 annelie.almquist@liu.se
Tillåtna hjälpmedel	Ett A4-blad med egna handskrivna anteckningar (båda sidor) samt räknedosa
Övrigt	
Antal exemplar i påsen	

Tentamen

Linköpings universitet, Institutionen för datavetenskap, STIMA

Kurskod och namn:	732G71 Statistik B
Datum och tid:	2019-02-04, 8-12
Jourhavande lärare:	Ann-Charlotte Hallberg (tentamen skapad av Bertil Wegmann)
Tillåtna hjälpmedel:	Ett A4-blad med egna handskrivna anteckningar (båda sidor) samt räknedosa.
Betygsgränser:	Tentamen omfattar totalt 20p. Godkänt från 12p, väl godkänt från 16p. Siffrorna i uppgifterna är delvis påhittade.

Redovisa och motivera tydligt alla dina lösningar!

Uppgift 1 (8p)

En branschorganisation i en amerikansk delstat vill bland annat undersöka hur medianpriset (i tusentals US dollar) på villor i en stadsdel påverkas av kriminalitet per capita i stadsdelen. I ett första steg av en större undersökning tog man därför ett slumpmässigt urval av 7 stycken stadsdelar och noterade medianpriset på villor och kriminalitet per capita i varje stadsdel. Detta gav följande tabell:

Stadsdel	Medianpris på villor	Kriminalitet per capita
1	24.0	0.006
2	22.8	0.092
3	11.8	2.780
4	23.3	0.046
5	24.8	0.043
6	19.9	3.837
7	14.3	5.581

Antag en enkel linjär regressionsmodell där medianpris på villor i en stadsdel förklaras av kriminalitet per capita. $SSE = 58.66$

- Skatta regressionslinjens skärning och lutning i den enkla linjära regressionsmodellen. Tolka skattningarna i ord. (2p)
- Beräkna ett 95 % dubbelsidigt konfidensintervall för lutningen. Tolka konfidensintervallet i ord. Använd det beräknade konfidensintervallet för att dra slutsats om kriminalitet per capita signifikant påverkar medianpriset på villor i en stadsdel och ange på vilken signifikansnivå man avgör det på. (3p)

- c) Beräkna ett 99 % dubbelsidigt konfidensintervall för det genomsnittliga medianpriset på villor i stadsdelar som har värdet 2.32 på kriminalitet per capita. Tolka konfidensintervallet i ord. $Distance\ value = 0.152$ (2p)
- d) Beräkna och tolka förklaringsgraden i ord. (1p)

Uppgift 2 (2p)

Ett tillverkningsföretag säljer en arbetsbänk för verkstäder. Arbetsbänken hade följande prisutveckling i kronor mellan 2016 och 2018:

År	Pris
2016	21000
2017	21500
2018	22900

Under samma tidsperiod var ett prisindex för varugruppen industri- och verkstadsprodukter enligt följande:

År	KPI_industri_verkstad
2016	245
2017	251
2018	254

Beräkna en indexserie för åren 2016, 2017 och 2018, med basår 2016, som visar hur priserna för företagets arbetsbänkar har utvecklats jämfört med den allmänna prisutvecklingen av industri- och verkstadsprodukter.

Uppgift 3 (6p)

I en större undersökning vill branschorganisationen i uppgift 1 undersöka hur medianpriset (i tusentals US dollar) på villor i en stadsdel påverkas av fler faktorer än bara kriminalitet per capita i stadsdelen. Därför tog man ett slumpmässigt urval av 113 stadsdelar i delstaten och noterade värden på följande variabler för respektive stadsdel:

y = medianpriset på villor

x_1 = kriminalitet per capita

x_2 = genomsnittligt antal rum per bostad

x_3 = fastighetskatt per 1000 US dollar

- a) Använd utskrifterna nedan för att på 1 % signifikansnivå välja bästa regressionsmodell (av de sju möjliga modellerna) för att förklara variationen i medianpris på villor med hjälp av procedurerna (var för sig)
- Bakåteliminering
 - Stegvis regression, där $\alpha_{entry} = \alpha_{stay}$

Motivera val av bästa modell väl för respektive procedur. (2p)

- b) Beräkna en prognos och ett 95 % prognosintervall med modell 7 för medianpriset på villor i en stadsdel där $x_1 = 5$, $x_2 = 7$ och $x_3 = 40$. Tolka prognosintervallet i ord. *Distance value = 0.0197* (2p)
- c) Utför residualanalys utifrån residualplotten till modell 7 för att undersöka om feltermerna är normalfördelade, har konstant varians och är oberoende av varandra. Motivera väl om varje antagande verkar vara uppfyllt eller ej. (2p)

MODELL 1:

Regression Analysis: y versus x1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1478	1477,82	21,08	0,000
x1	1	1478	1477,82	21,08	0,000
Error	111	7783	70,12		
Total	112	9261			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8,37363	15,96%	15,20%	9,13%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	23,876	0,839	28,47	0,000	
x1	-0,3064	0,0667	-4,59	0,000	1,00

MODELL 2:

Regression Analysis: y versus x2

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	4857,78	4857,78	122,46	0,000
x2	1	4857,78	4857,78	122,46	0,000
Error	111	4403,10	39,67		
Lack-of-Fit	105	4320,81	41,15	3,00	0,082
Pure Error	6	82,29	13,72		
Total	112	9260,88			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
6,29822	52,45%	52,03%	50,22%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-40,18	5,70	-7,05	0,000	
x2	9,972	0,901	11,07	0,000	1,00

MODELL 3:

Regression Analysis: y versus x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1906	1905,54	28,76	0,000
x3	1	1906	1905,54	28,76	0,000
Error	111	7355	66,26		
Lack-of-Fit	35	1880	53,72	0,75	0,830
Pure Error	76	5475	72,04		
Total	112	9261			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8,14029	20,58%	19,86%	17,62%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	33,10	2,11	15,69	0,000	
x3	-0,2586	0,0482	-5,36	0,000	1,00

MODELL 4:

Regression Analysis: y versus x1; x2 Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	5682,1	2841,07	87,33	0,000
x1	1	824,4	824,36	25,34	0,000
x2	1	4204,3	4204,33	129,23	0,000
Error	110	3578,7	32,53		
Total	112	9260,9			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5,70386	61,36%	60,65%	58,88%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-35,42	5,25	-6,75	0,000	
x1	-0,2312	0,0459	-5,03	0,000	1,02
x2	9,375	0,825	11,37	0,000	1,02

MODELL 5:

Regression Analysis: y versus x1; x3 Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	2210,3	1105,15	17,24	0,000
x1	1	304,8	304,75	4,75	0,031
x3	1	732,5	732,48	11,43	0,001
Error	110	7050,6	64,10		
Total	112	9260,9			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8,00601	23,87%	22,48%	19,04%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	31,07	2,27	13,66	0,000	
x1	-0,1660	0,0761	-2,18	0,031	1,42
x3	-0,1913	0,0566	-3,38	0,001	1,42

MODELL 6:

Regression Analysis: y versus x2; x3 Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	5596,40	2798,20	84,00	0,000
x2	1	3690,86	3690,86	110,79	0,000
x3	1	738,62	738,62	22,17	0,000
Error	110	3664,48	33,31		
Lack-of-Fit	108	3646,95	33,77	3,85	0,228
Pure Error	2	17,52	8,76		
Total	112	9260,88			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5,77178	60,43%	59,71%	56,76%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-27,13	5,91	-4,59	0,000	
x2	8,974	0,853	10,53	0,000	1,07
x3	-0,1662	0,0353	-4,71	0,000	1,07

MODELL 7:

Regression Analysis: y versus x1; x2; x3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	5879,9	1959,97	63,19	0,000
x1	1	283,5	283,52	9,14	0,003
x2	1	3669,6	3669,63	118,31	0,000
x3	1	197,8	197,77	6,38	0,013
Error	109	3381,0	31,02		
Total	112	9260,9			

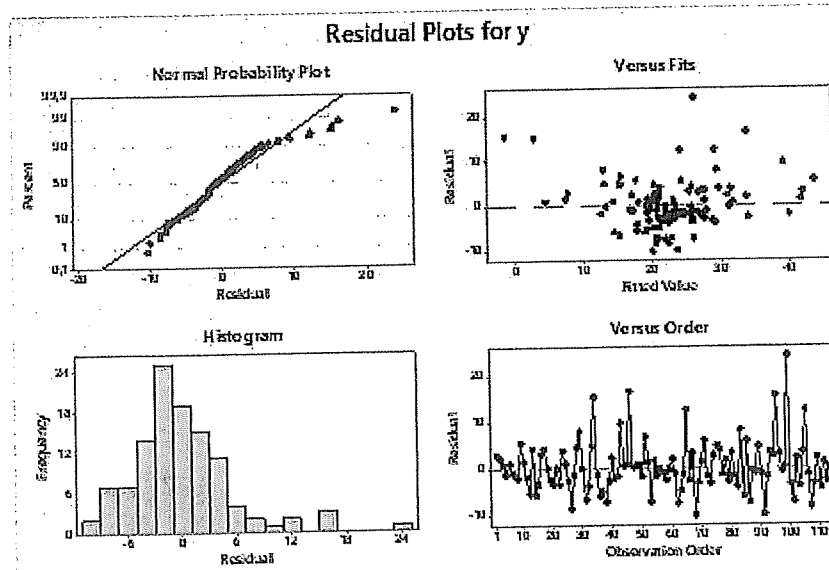
Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5,56938	63,49%	62,49%	59,71%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-28,91	5,74	-5,04	0,000	
x1	-0,1601	0,0530	-3,02	0,003	1,42
x2	8,949	0,823	10,88	0,000	1,07
x3	-0,1016	0,0402	-2,53	0,013	1,49

Residualplottar till modell 7



Uppgift 4 (4p)

a) Värden på ett börsindex ges nedan för varje kvartal under åren 2017 och 2018:

År	Kvartal	Börsindex
2017	1	1537
2017	2	1627
2017	3	1548
2017	4	1672
2018	1	1593
2018	2	1571
2018	3	1616
2018	4	1538

- a) Beräkna centrerade glidande medelvärden för åren 2017 och 2018 och ange för vilka kvartal de centrerade glidande medelvärdena gäller för. (2p)
- b) Nedan har klassisk komponentuppdelning anpassats i Minitab på börsindexets värden för varje kvartal mellan åren 2015 och 2018. Tolka säsongskomponenterna i ord. Beräkna en prognos för kvartal 1 år 2019 med hjälp av den klassiska komponentuppdelningen nedan. (2p)

Time Series Decomposition for Börsindex

Data Börsindex

Length 16

Fitted Trend Equation

$$Y_t = 1493,4 + 4,96 \times t$$

Seasonal Indices

Period	Index
1	1,5625
2	-12,1875
3	-10,0625
4	20,6875

Accuracy Measures

MAPE 5,37

MAD 80,69

MSD 9028,98

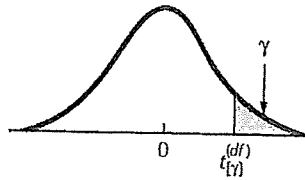


TABLE A2 A t-Table: Values of $t_{(gamma)}^{(df)}$

df	$t_{[.10]}^{(df)}$	$t_{[.05]}^{(df)}$	$t_{[.025]}^{(df)}$	$t_{[.01]}^{(df)}$	$t_{[.005]}^{(df)}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
inf.	1.282	1.645	1.960	2.326	2.576

Source: From "Table of Percentage Points of the t-Distribution," by Maxine Merrington, *Biometrika* 32 (1941), 300. Reproduced by permission of the *Biometrika* Trustees.

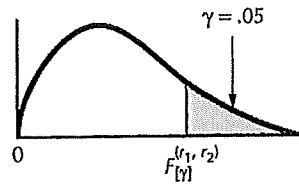


TABLE A3 An F-Table: Values of $F_{[.05]}$

Denominator Degrees of Freedom, r_2	Numerator Degrees of Freedom, r_1								
	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Source: From "Tables of Percentage Points of the Inverted Beta (F)-Distribution," by Maxine Merrington and Catherine M. Thompson, *Biometrika* 33 (1943), 73-88. Reproduced by permission of the *Biometrika* Trustees.