

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2017-04-08
Sal (1)	TER3(41)
Tid	8-12
Kurskod	732G71
Provkod	TENT
Kursnamn/benämning Provnamn/benämning	Statistik B Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Bertil Wegmann
Telefon under skrivtiden	070-1128321
Besöker salen ca klockan	09 ³⁰
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, Statistik

Kurskod och namn:	732G71 Statistik B
Datum och tid:	2017-04-08, 8-12
Jourhavande lärare:	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 (9p)

I en studie vill man förklara uppvärmningskostnad för bostadshus under en månad (y i kronor) med hjälp av bostadsytan i kvadratmeter (x). Därför valde man slumpmässigt ut 8 olika bostadshus och resultatet av detta visas i tabellen nedan:

Bostadshus	y	x
1	1850	140
2	2450	210
3	1900	130
4	2150	150
5	1500	90
6	1700	100
7	2200	120
8	2000	180

Antag följande enkla linjära regressionsmodell: $y = \beta_0 + \beta_1 x + \varepsilon$

- Skatta skärningen och lutningen i regressionsmodellen ovan och tolka skattningarna i ord. Motivera om tolkningen av skärningen är relevant. (2p)
- Testa på 5 % signifikansnivå om bostadsytan i kvadratmeter signifikant bidrar till att förklara uppvärmningskostnaden för bostadshus. SSE = 228885. (2p)

- c) Beräkna en prognos och ett 99 % prognosintervall för uppvärmningskostnaden för ett bostadshus på 110 kvadratmeter genom att använda den skattade regressionsmodellen i uppgift a). (2p)
- d) Beräkna regressionsmodellens förklaringsgrad och tolka förklaringsgraden i ord. (2p)
- e) Ett av modellantagandena för regressionsmodellen är konstant varians. Förklara vad som menas med detta antagande och hur residualerna bör se ut för att detta antagande ska vara uppfyllt. (1p)

Uppgift 2 (6p)

Datamaterialet i uppgift 1 är en del av ett större datamaterial med 48 slumpmässigt valda bostadshus.

Följande är en lista över samtliga variabler i datamaterialet:

uppv (y)	Uppvärmningskostnad för bostadshuset
yta (x_1)	Bostadsyta i kvadratmeter
alder (x_2)	Bostadshusets åldersfaktor (beräknas på byggnadsår och renoveringsår). Högre värde ju äldre huset är eller ju mindre renoverat huset är.
olja (x_3)	En variabel som är lika med 1 om bostadshuset värms upp med oljepanna och 0 annars.

Man vill pröva att förklara uppvärmningskostnaden med variablerna alder och olja samt interaktionsvariabeln alder*olja som en produkt mellan alder och olja. Detta ger följande multipla linjära regressionsmodell

$$y = \beta_0 + \beta_1 x_2 + \beta_2 x_3 + \beta_3 x_2 x_3 + \varepsilon$$

Skattning av regressionsmodellen ovan gav följande resultat i Minitab:

Regression Analysis: uppv versus alder; olja; alder*olja

The regression equation is

$$\text{uppv} = 2.99 - 1.50 \text{ alder} - 0.251 \text{ olja} + 1.42 \text{ alder*olja}$$

Predictor	Coef	SE Coef
Constant	2.9925	0.2513
alder	-1.4967	0.4869
olja	-0.2508	0.5025
alder*olja	1.4163	0.9597

Analysis of Variance

Source	DF	SS
Regression	3	3.1181
Residual Error	44	3.8422
Total	47	6.9603

Predicted Values for New Observations

New

Obs	Fit	SE Fit	95% CI	95% PI
1	2.6774	0.2438	(2.1867; 3.1680)	(1.9123; 3.4425)

Values of Predictors for New Observations

New

Obs	alder	olja	alder*olja
1	0.800	1.00	0.800

- Redovisa den skattade regressionslinjen för uppvärmningskostnaden för bostadshus som värms upp med oljepanna. (1p)
- Testa på 5 % signifikansnivå om den multipla linjära regressionsmodellen ovan är signifikant eller inte. Dra fullständig slutsats av ditt test i ord. (2p) (om antalet frihetsgrader för kritiska värdet för testet inte finns i tabell, använd närliggande värde)
- Beräkna ett 90 % konfidensintervall för interaktionsparametern. Tolka konfidensintervallet i ord och avgör om interaktionsvariabeln bidrar signifikant till modellen. (2p)
- Vad är 95 % CI och 95 % PI och vad är det för skillnad mellan 95 % CI och 95 % PI i utskriften ovan? Motivera väl. (1p)

Uppgift 3 (2p)

I nedanstående tabell visas försäljning av en viss vara år 2000 – 2005 i löpande och fasta priser (miljoner kr).

ÅR	FASTA	LÖPANDE
2000	114	112
2001	115	115
2002	119	124
2003	121	131
2004	125	134
2005	127	137

- a) Vad menas med försäljning i löpande och fasta priser? Motivera. Ange också vilket år som är basår. (1p)
- b) Beräkna implicitprisindex med hjälp av försäljningsserierna ovan för åren 2000 – 2005. (1p)

Uppgift 4 (3p)

Klassisk komponentuppdelning har anpassats på ett prisindex för kopieringsmaskiner nedan. Prisindexet sträcker sig från kvartal 1 år 2005 till kvartal 4 år 2013.

Time Series Decomposition for Kopieringsmaskiner

Data Kopieringsmaskiner
Length 36
NMissing 0

Fitted Trend Equation

$$Y_t = 95,325 + 0,113 \times t$$

Seasonal Indices

Period	Index
1	0,252
2	2,268
3	1,153
4	0,327

Accuracy Measures

MAPE 0,343
MAD 0,059
MSD 0,003

- a) Vilken metod (additiv eller multiplikativ) har använts för den klassiska komponentuppdelningen? Motivera. (1p)
- b) Tolka samtliga säsongsindex i ord. (1p)
- c) Gör en prognos för kvartal 1 år 2014. (1p)

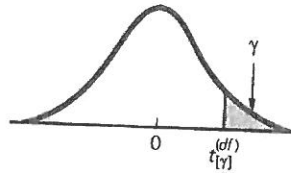


TABLE A2 A t-Table: Values of $t_{(y)}^{(df)}$

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.

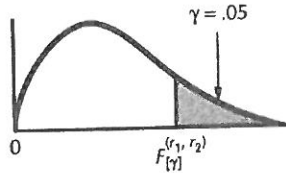


TABLE A3 An F-Table: Values of $F_{[.05]}$

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.