



# Försättsblad till skriftlig tentamen vid Linköpings Universitet

<b>Datum för tentamen</b>	2011-04-09
<b>Sal (1)</b> Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER1
<b>Tid</b>	8-12
<b>Kurskod</b>	732G71
<b>Provkod</b>	TENT
<b>Kursnamn/benämning</b> <b>Provnamn/benämning</b>	Statistik B Tentamen
<b>Institution</b>	IDA
<b>Antal uppgifter som ingår i tentamen</b>	5
<b>Jour/Kursansvarig</b> Ange vem som besöker salen	Kalle Wahlin
<b>Telefon under skrivtiden</b>	0709-719096
<b>Besöker salen ca kl.</b>	9.30
<b>Kursadministratör/kontaktperson</b> (namn + tfnr + mailaddress)	carita.lilja@liu.se tel 1463
<b>Tillåtna hjälpmedel</b>	Valfri räknedosa.
<b>Övrigt</b>	
<b>Vilken typ av papper ska användas, rutigt eller linjerat</b>	Rutigt.
<b>Antal exemplar i påsen</b>	25



# Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

---

- Kurskod och namn: 732G71 Statistik B
- Datum och tid: 2011-04-09, 8-12
- Jourhavande lärare: Kalle Wahlin
- Tillåtna hjälpmedel: Valfri räknedosa.
- Betygsgränser: Tentamen omfattar totalt 15p. Godkänt från 9p, väl godkänt från 12p.
- Formelsamling och tabeller följer efter uppgifterna. Svarsformulär till uppgifterna 2-5 finns på slutet.
- Siffrorna i uppgifterna är påhittade.

Till uppgift 1 ska fullständig lösning inlämnas. Till uppgifterna 2-5 lämnas endast svar på svarsblankett.

---

## Uppgift 1 (8p)

Vi studerar en fonds utveckling mellan år 2000 och 2009, mätt som en fondandels värde i kronor.

År	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Fondvärde (kr)	105	121	132	119	147	126	151	164	188	162

- Åskådliggör data i ett spridningsdiagram. (0.5p)
- Beräkna korrelationskoefficienten mellan fondvärde och tid. Välj som  $y$  en fondandels värde och som förklaringsvariabel  $t$  ett tidsindex satt till 1 för år 2000. (1p)
- Beräkna  $b_0$  och  $b_1$  i en enkel linjär regressionsmodell. (1p)
- Tolka  $b_0$  och  $b_1$  med ord. (0.5p)
- Är  $\beta_1$  signifikant skild från 0? Använd 5% signifikansnivå. (1.5p)
- Beräkna modellens förklaringsgrad. Är modellen välanpassad? (0.5p)
- Beräkna ett 95% konfidensintervall för hur fonden i genomsnitt utvecklas under ett år. (1.5p)
- Beräkna ett 95% prognosintervall för det förväntade värdet på en fondandel år 2010. (1.5p)

## Uppgift 2 (2.5p)

Ett företag studerar sambandet mellan total försäljning (Sales, i miljontals dollar) och hur mycket (i tusentals dollar) man har spenderat på reklam i tv, radio, tryckta medier (print) samt övrig marknadsföring (other) under ett antal månader. Nedan följer ett antal utskrifter från Minitab.

**Stepwise Regression: sales versus tv; radio; print; other**

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is sales on 4 predictors, with N = 18

Step	1	2	3
Constant	-57.58	-57.06	-58.68
tv	1.516	1.391	1.349
T-Value	12.59	13.88	15.48
P-Value	0.000	0.000	0.000
print		0.376	0.303
T-Value		3.41	3.09
P-Value		0.004	0.008
radio			0.196
T-Value			2.56
P-Value			0.023
S	4.07	3.16	2.70
R-Sq	90.83	94.84	96.48
R-Sq(adj)	90.26	94.15	95.73
Mallows Cp	23.0	8.8	4.2

**Stepwise Regression: sales versus tv; radio; print; other**

Forward selection. Alpha-to-Enter: 0.25

Response is sales on 4 predictors, with N = 18

Step	1	2	3
Constant	-57.58	-57.06	-58.68
tv	1.516	1.391	1.349
T-Value	12.59	13.88	15.48
P-Value	0.000	0.000	0.000
print		0.376	0.303
T-Value		3.41	3.09
P-Value		0.004	0.008
radio			0.196
T-Value			2.56
P-Value			0.023
S	4.07	3.16	2.70
R-Sq	90.83	94.84	96.48
R-Sq(adj)	90.26	94.15	95.73
Mallows Cp	23.0	8.8	4.2

**Stepwise Regression: sales versus tv; radio; print; other**

Backward elimination. Alpha-to-Remove: 0.1

Response is sales on 4 predictors, with N = 18

Step	1	2
Constant	-55.50	-53.37
tv	1.17	1.06
T-Value	6.39	7.38
P-Value	0.000	0.000
radio	0.192	0.202
T-Value	2.53	2.69
P-Value	0.025	0.018
print	0.16	
T-Value	0.99	
P-Value	0.339	
other	0.98	1.70
T-Value	1.09	3.14
P-Value	0.295	0.007
S	2.68	2.68
R-Sq	96.78	96.53
R-Sq(adj)	95.79	95.79
Mallows Cp	5.0	4.0

**Best Subsets Regression: sales versus tv; radio; print; other**

Response is sales

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	r	p	o	a	r	t	d	i	h	t	i	n	e	v	o	t	r					
1	90.8	90.3	23.0	4.0720	X																					
1	81.7	80.5	59.9	5.7571																						X
2	94.8	94.2	8.8	3.1557	X																					X
2	94.7	94.0	9.2	3.1856	X																					X
3	96.5	95.8	4.0	2.6773	X	X																				X
3	96.5	95.7	4.2	2.6967	X	X	X																			X
4	96.8	95.8	5.0	2.6785	X	X	X	X																		X

### Regression Analysis: sales versus tv; radio; print; other

The regression equation is

$$\text{sales} = -55.5 + 1.17 \text{ tv} + 0.192 \text{ radio} + 0.161 \text{ print} + 0.984 \text{ other}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-55.504	5.489	-10.11	0.000	
tv	1.1727	0.1835	6.39	0.000	5.371
radio	0.19204	0.07603	2.53	0.025	1.195
print	0.1614	0.1625	0.99	0.339	3.491
other	0.9843	0.9019	1.09	0.295	9.750

S = 2.67850 R-Sq = 96.8% R-Sq(adj) = 95.8%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	2800.82	700.20	97.60	0.000
Residual Error	13	93.27	7.17		
Total	17	2894.08			

Source	DF	Seq SS
tv	1	2628.79
radio	1	94.15
print	1	69.33
other	1	8.55

#### Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	19.236	2.054	(14.799; 23.673)	(11.944; 26.528)

#### Values of Predictors for New Observations

New Obs	tv	radio	print	other
1	50.0	20.0	15.0	10.0

a) Vilket av följande påståenden stämmer bäst? (0.5p)

- I. Den bästa modellen enligt bakåtelimeringsprincipen är den med samtliga förklaringsvariabler.
- II. I en stegvis regression kan en variabel som inkluderats i modellen aldrig tas bort igen.
- III. Analyserna har baserats på 17 månaders data.
- IV. Framåtvalsprincipen och stegvis regression leder alltid till samma resultat men stegvis regression är en beräkningsmässigt snabbare metod.
- V. VIF-värdena visar på att det föreligger kraftig multikollinearitet mellan variablerna *print* och *other*.
- VI. Alpha-to-remove avser vald signifikansnivå.

b) Vilket av följande påståenden stämmer minst väl? (0.5p)

- I. Bakåteliminering genomförs genom att man utgår från en modell med samtliga förklaringsvariabler, och sedan utesluter den minst signifikanta variabeln. Modellen återanpassas och processen fortgår tills samtliga förklaringsvariabler är signifikanta.
- II. Mallows C beräknas enligt

$$C_p = \frac{SSE}{s_p^2} - (n - 2 \cdot (k + 1))$$

- III. Den modell som endast innehåller variabeln *other* har ett värde på Mallows C som indikerar stor bias.
- IV. För varje tusentals dollars ökning i marknadsföring i radio stiger försäljningen i genomsnitt med 192040 dollar.
- V. Ett 95% intervall för försäljningen en specifik månad när man lägger 50 tusen dollar på tv, 20 tusen dollar på radio, 15 tusen dollar på tryckta medier och 10 tusen dollar på övrig marknadsföring ges av (14.799; 23.673)
- VI. R-sq (adj), Mallows C och s kan ibland ge motstridiga resultat.

c) Bestäm ett 99% prognosintervall för den förväntade försäljningen en månad när företaget spenderar 50 tusen dollar på tv, 20 tusen dollar på radio, 15 tusen dollar på tryckta medier och 10 tusen dollar på övriga marknadsföring. (1.5p)

### Uppgift 3 (2p)

Följande data föreligger.

År	Vara A		Vara B	
	Pris/st	Antal sålda	Pris/st	Antal sålda
2007	105	2000	60	500
2008	118	1250	75	450
2009	135	850	78	380

Beräkna ett fastbasindex med länkar enligt Laspeyre och 2007 som basår.

### Uppgift 4 (2p)

På basis av data under åren 1980-2003 vill vi skatta följande modell för efterfrågan på havregryn:

$$Q = A \cdot P^{E_p} \cdot I^{E_I} \cdot \delta$$

där  $Q$  är efterfrågan representerad av konsumtionen av havregryn per capita i 1980 års priser (enhet: kronor),  $P$  är relativprisindex för havregryn,  $I$  är disponibel realinkomst per capita,  $E_p$  är priselasticiteten för havregryn,  $E_I$  är inkomstelasticiteten och  $\delta$  är en slumpvariabel med väntevärde 1.

En analys i Minitab har gett följande utskrift.

## Regression Analysis: logQ versus logP; logI

The regression equation is

$$\log Q = 1.16 - 0.162 \log P + 0.453 \log I$$

Predictor	Coef	SE Coef	T	P
Constant	1.1558	0.2943	3.93	0.001
logP	-0.16155	0.04596	-3.52	0.002
logI	0.45325	0.04862	9.32	0.000

S = 0.009674    R-Sq = 91.1%    R-Sq(adj) = 90.3%

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.020218	0.010109	108.01	0.000
Residual Error	21	0.001965	0.000094		
Total	23	0.022184			

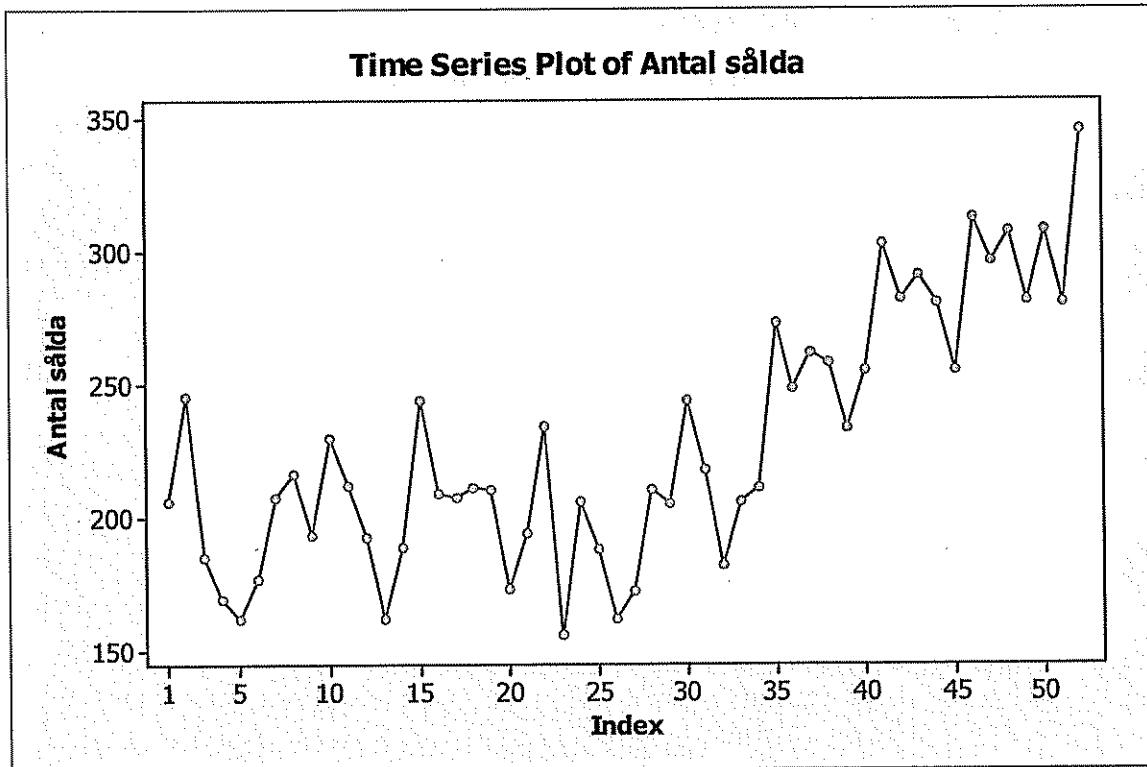
*log* i utskriften ovan står för 10-logaritmen.

- Avgör med ett lämpligt test på 5% nivå om varugruppen havregryn är oelastisk, dvs har en priselasticitet som är större än  $-1$ . Svara med teststorhetens värde samt om testet är signifikant eller ej. (1p)
- År 2003 har relativprisindex värdet 89.64, den disponibla realinkomsten per capita värdet 63057 och konsumtionen per capita i 1980 års priser är 1073 kronor. Antag att 2008 har relativprisindex ökat med 2% medan den disponibla realinkomsten per capita är oförändrad jämfört med 2003. Beräkna en prognos av den procentuella förändringen i efterfrågan 2008 jämfört med 2003. (1p)

## Uppgift 5 (0.5p)

Man har studerat den veckovisa försäljningen av 500-grams påsar med ris i en viss butik. En tidsserieplott av antalet sålda påsar per vecka ges enligt följande.





Tre analyser, som nedan följer delvis censurerade, har genomförts:

#### Single Exponential Smoothing for Antal sålda

Data Antal sålda  
Length 52

Smoothing Constant  
Alpha 0.366912

#### Accuracy Measures

MAPE 9.872  
MAD 21.875  
MSD 781.878

#### Forecasts

Period	Forecast	Lower	Upper
53	310.627	257.035	364.219
54	310.627	257.035	364.219

## Double Exponential Smoothing for Antal sålda

Data Antal sålda  
Length 52

### Smoothing Constants

Alpha (level) 0.489596  
Gamma (trend) 0.066912

### Accuracy Measures

MAPE 10.339  
MAD 22.370  
MSD 887.404

### Forecasts

Period	Forecast	Lower	Upper
53	324.219	269.414	379.025
54	328.441	266.602	390.280

## Regression Analysis: Antal sålda versus t

The regression equation is  
Antal sålda = 166 + 2.32 t

Predictor	Coef	SE Coef	T	P
Constant	166.396	8.760	19.00	0.000
t	2.3247	0.2876	8.08	

S = 31.1292 R-Sq = 56.6% R-Sq(adj) = 55.8%

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	63299			
Residual Error	50	48451			
Total	51	111750			

### Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	289.60	8.76	(272.01; 307.20)	(224.65; 354.56)

### Values of Predictors for New Observations

New Obs	t
1	53.0

Vilket av följande påståenden stämmer bäst?

- I. Prognoserna avser kvartal 1 och 2 året efter att studien avslutats.
- II. *F*-värdet för regressionsanalysen är 1.31.
- III. Genom att jämföra lämplig *accuracy measure* inses att dubbel exponentiell utjämning är den bästa modellen.
- IV. Regressionsanalysen visar ingen statistiskt säkerställd trend i data.
- V. Ingen av de föreslagna modellerna är lämplig, eftersom data har en exponentiell utveckling.
- VI. En prognos för antalet sålda rispåsar i tidpunkt 54 baserat på regressionsanalysen är 291.28.