



# Försättsblad till skriftlig tentamen vid Linköpings Universitet

<b>Datum för tentamen</b>	2013-01-16
<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>	8-13
<b>Kurskod</b>	732G21
<b>Provkod</b>	TEN1
<b>Kursnamn/benämning</b> <b>Provnamn/benämning</b>	Sambandsmodeller Tentamen
<b>Institution</b>	IDA
<b>Antal uppgifter som ingår i tentamen</b>	3
<b>Jour/Kursansvarig</b> Ange vem som besöker salen	Lotta Hallberg
<b>Telefon under skrivtiden</b>	
<b>Besöker salen ca kl.</b>	10
<b>Kursadministratör/kontaktperson</b> (namn + tfnr + mailaddress)	carita.lilja@liu.se tel 1463
<b>Tillåtna hjälpmedel</b>	Räknedosa. Läroboken: <i>Applied linear statistical models</i> av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning
<b>Övrigt</b>	G=12, VG=16
<b>Vilken typ av papper ska användas, rutigt eller linjerat</b>	Rutigt
<b>Antal exemplar i påsen</b>	42

## TENTAMEN I SAMBANDSMODELLER, 2013-01-16

**Skrivtid:** kl: 8-13

**Hjälpmedel:** Räknedosa. Läroboken: *Applied linear statistical models* av Kutner, Nachtsheim m fl som inte får innehålla anteckningar men får ha markeringar och flärpar. Flärpar får ha en liten anteckning.

**Jourhavande lärare:** Lotta Hallberg.

Redovisa och motivera kort alla dina lösningar

---

Tolka (om möjligt) alla dina resultat!

1

Följande data ska analyseras.

Y = Försäljning av kosmetika (tusentals dollar)

x1= kostnad för direktreklam i salong

x2= kostnad för reklam i lokaltidning.

x3= kostnad för reklam i nationell tidning

Nedan visas några utskrifter från minitab med bland andra regressionsanalyser med Y som responsvariabel och x1, x2 och/eller x3 som förklarande variabler.

**Correlations: Y; x1; x2; x3**

	Y	x1	x2
x1	0,842		
x2	0,842	0,974	
x3	0,474	0,376	0,410

**Best Subsets Regression: Y versus x1; x2; x3**

Response is Y

Vars	R-Sq	R-Sq(adj)	Mallows		x x x		
			Cp	S	1	2	3
1	71,0	70,3	4,9	1,8882		X	
1	70,9	70,2	5,1	1,8923	X		
2	73,7	72,5	2,7	1,8177	X		X
2	73,0	71,7	3,9	1,8443		X	X
3	74,2	72,2	4,0	1,8254	X	X	X

---

### Modell 1

The regression equation is

$$Y = 1,02 + 0,966 x1 + 0,629 x2 + 0,676 x3$$

Predictor	Coef	SE Coef	T	P
Constant	1,023	1,203	0,85	0,400
x1	0,9657	0,7092	1,36	0,181
x2	0,6292	0,7783	0,81	0,424
x3	0,6760	0,3557	1,90	0,065

S = 1,82542    R-Sq = 74,2%    R-Sq(adj) = 72,2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	382,66	127,55	38,28	0,000
Residual Error	40	133,29	3,33		
Total	43	515,95			

Variabel	VIF
x1	20,07
x2	20,72
x3	1,22

## Modell 2

The regression equation is  
 $Y = 1,02 + 1,52 x_1 + 0,736 x_3$

Predictor	Coef	SE Coef	T	P
Constant	1,017	1,198	0,85	0,401
x1	1,5221	0,1701	8,95	0,000
x3	0,7362	0,3464	2,13	0,040

S = 1,81769    R-Sq = 73,7%    R-Sq(adj) = 72,5%

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	2	380,48	190,24	57,58	0,000
Residual Error	41	135,46	3,30		
Total	43	515,95			

### Unusual Observations

Obs	x1	Y	Fit	SE Fit	Residual	St Resid
30	3,50	2,960	8,553	0,337	-5,593	-3,13R

R denotes an observation with a large standardized residual.

För observation nummer 30 har minitab beräknat

$$h_{30,30} = 0,0344$$

$$\text{Studentized deleted residual}_{30} = -3,5462$$

$$DFFITS_{30} = -0,6697$$

$$\text{Cook's Distance} = 0,1166$$

- Vilka av de tre x-variablerna ska vara med i en regressionsmodell om man använder Mallows Cp som mått? Förklara. 1p
- Korrelationen mellan Y och x2 är 0,842 men x2 är inte signifikant i modell 1. Förklara vilken orsaken kan vara. 1p
- Tolka förklaringsgraden i modell 2. 1p
- Vilken typ av reklam är mest effektiv enligt modell 2? Förklara. 1p
- Beräkna ett 95% prediktionsintervall för Y då x1=4 och x3=4 med modell 2.

$$(X'X)^{-1} = \begin{pmatrix} 0,434 & -0,008 & -0,106 \\ -0,008 & 0,009 & -0,007 \\ -0,106 & -0,007 & 0,036 \end{pmatrix}$$

3p

- I utskriften för modell 2 så ges en varning för stor residual på observation nummer 30. Använd de fyra måtten som är givna i utskriften och avgör vad som är problemet med observationen. 2p

## 2

Ett företag utvärderade tre nya produktionslinjer. Dessa var låg, moderat och hög automatisering av teknologin. Linjeoperatörerna hade olika lång erfarenhet så tre grupper med operatörer med ungefär lika lång erfarenhet slumpades ut till de tre automatiseringslinjerna. Responsvariabeln  $Y$  är mätt i enheter per timme. Anta att  $Y$  är normalfördelad.

Operatörs erfarenhet	Teknologi		
	Hög Automatisering	Moderat Automatisering	Låg Automatisering
< 1 år	10	8	5
1-2 år	18	12	10
>2 år	19	13	12

Detta är ett randomiserat blockförsök.

- Plotta data på ett lämpligt sätt och avgör om det verkar finnas interaktion mellan faktorerna. 1p
- Beräkna ANOVA tabellen.  $SSTO=1589$ . 2p
- Testa om det är skillnad mellan automatiseringsnivåerna. Använd ett F-test. 1p
- Anta  $Y_{32} = 13$  saknas. Skatta  $\mu_{32}$  1p

## 3

Stickprov om 100 personer som väntar på rättegång för grova skattebrott valdes dels från glesbygd, förort och tätort i två amerikanska stater. Följande tabell sammanställer resultatet för tiden  $Y$  i månader mellan arrest och rättegång.  $Y$  kan antas vara normalfördelad.

Stat	Domstol		
	Glesbygd	Förort	Tätort
1	$\bar{Y} = 3,4; s = 1,3$	$\bar{Y} = 5,8; s = 1,2$	$\bar{Y} = 6,8; s = 1,5$
2	$\bar{Y} = 2,4; s = 1,5$	$\bar{Y} = 3,5; s = 1,7$	$\bar{Y} = 4,7; s = 1,7$

- Analysera data genom att använda ANOVA tabellen. 2p

Källa	DF	SS
Stat		486,00
Domstol		826,33
Interaction		49,00
Error	594	1327,59

- Använd Tukeys metod för att jämföra alla cellmedelvärden. Använd 95% familjekonfidensgrad. Det räcker att du visar hur detta görs för  $\mu_{13} - \mu_{23}$ . 2p
- Anta nu att de två staterna valdes slumpmässigt bland alla stater. Skatta  $\sigma_{\alpha}^2$  med ett 95% konfidensintervall. ( $\alpha$  slumpmässiga effekter för stater) 2p