

Försättsblad till skriftlig tentamen vid Linköpings universitet



Datum för tentamen	2019-01-15
Sal (1)	<u>TER2(20)</u>
Tid	8-13
Utb. kod	732G46
Modul	TEN1
Utb. kodnamn/benämning Modulnamn/benämning	Regressions- och variansanalys Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	013-281657
Besöker salen ca klockan	10
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Annelie Almquist
Tillåtna 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.
Övrigt	
Antal exemplar i påsen	

TENTAMEN I REGRESSIONS- OCH VARIANSANALYS, 2019-01-15

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
Betygsgränser: För godkänt krävs minst 12 av 20 poäng och för väl godkänt krävs minst 16 av 20 poäng.

Redovisa och motivera kort alla dina lösningar
Tolka (om möjligt) alla dina resultat!

1

En medicinsk forskare vill undersöka vad som kan påverka pulsen hos studenter. Därför slumpade man ut 35 studenter som fick springa på ett band i 1 min och därefter mätte man deras puls. Det är denna puls som är responsvariabel. Övriga variabler som man mätte var Vilopuls, Vikt i kg, Längd i cm, Kön och Kondition. Kön kodades till 1 om de var Män och 0 annars. Denna dummy-variabel heter därför Män.

Kondition har tre nivåer. Mycket god, Måttlig och Lite kondition. Dummy-variabeln Mycket är 1 om Mycket god kondition och 0 annars. Variabeln Måttligt är 1 om Måttlig kondition och 0 annars.

Nedan har tre modeller anpassats. Uppgifterna kommer efter utskriftena.

Modell 1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	8881,5	1480,25	12,49	0,000
Error	28	3319,2	118,54		
Total	34	12200,7			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
10,8878	72,79%	66,96%	56,72%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	69,9	56,9	1,23	0,229	
Vilopuls	0,633	0,188	3,36	0,002	1,33
Längd	0,056	0,335	0,17	0,869	2,41
vikt	-0,324	0,243	-1,34	0,193	2,57
Män	-13,16	6,24	-2,11	0,044	2,48
Mycket	-8,40	7,81	-1,08	0,291	2,88
Måttligt	5,10	7,40	0,69	0,496	3,30

Modell 2

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	8633,2	2158,30	18,15	0,000
Error	30	3567,6	118,92		
Total	34	12200,7			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
10,9050	70,76%	66,86%	63,47%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	52,9	18,4	2,87	0,007	
Vilopuls	0,657	0,187	3,51	0,001	1,31
Män	-17,93	5,00	-3,59	0,001	1,59
Mycket	-6,37	7,58	-0,84	0,408	2,71
Måttligt	6,76	7,12	0,95	0,350	3,04

Modell 3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	7833	3916,5	28,69	0,000
Error	32	4368	136,5		
Total	34	12201			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
11,6829	64,20%	61,96%	58,09%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	65,6	16,2	4,06	0,000	
Vilopuls	0,583	0,195	2,99	0,005	1,24
Män	-23,40	4,73	-4,94	0,000	1,24

$$(X'X)^{-1} = \begin{pmatrix} 1,91469 & -0,0225411 & -0,331176 \\ -0,02254 & 0,0002786 & 0,002970 \\ -0,33118 & 0,0029696 & 0,164229 \end{pmatrix}$$

- a) Modell 1. Skatta med ett 95% konfidensintervall den förväntade pulsen för en man med vilopuls 60 och som är 180 cm lång, väger 90 kg och har mycket god kondition. Du får använda direkt att $s(\hat{Y}_h) = 4,74$. 2p
- b) Eftersom Längd och Vikt inte var signifikanta i modell 1 så tog medicinaren bort dessa två variabler. Är det korrekt att göra så eller skulle han ha gjort på ett annat sätt? 1p

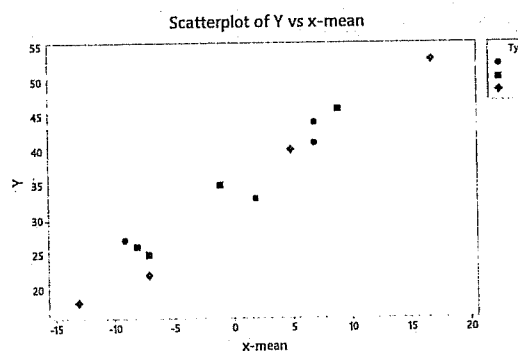
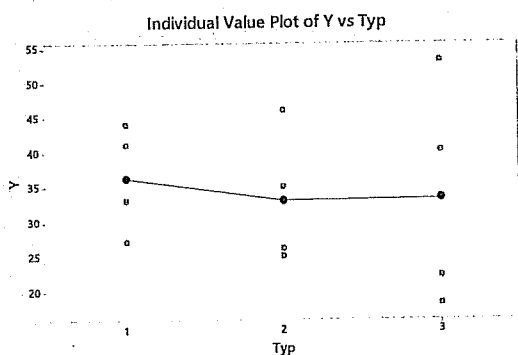
- c) Testa på 5% signifikansnivå om variabeln Kondition(Mycket , Måttligt) kan tas bort från Modell 2. 2p
- d) Tolka regressionskoefficienten för variabeln Mycket i Modell 2. 1p
- e) Modell 3. SE Coef för variabeln Män är 4,73 enligt utskrift. Visa hur denna kan räknas ut med hjälp av $(X'X)^{-1}$ och andra utskrifter. 1p

2

En leverantör av läskedrycker vill undersöka effektiviteten i leverans med tre olika typer av säckkärror.

Så den faktor som vi vill studera är Typ=typ av säckkärra. Det finns tre olika typer av kärror. Responsvariabeln Y= leveranstid i minuter. Denna beror troligen på x= mängden som ska levereras. Varje typ av säckkärra användes 4 gånger. x-mean är centrerade x-värden, centrerade runt sitt medelvärde.

Typ 1		Typ 2		Typ 3	
y	x	y	x	y	x
27	24	25	26	40	38
44	40	35	32	22	26
33	35	46	42	53	50
41	40	26	25	18	20



Modell 1

One-way ANOVA: Y versus Typ

Factor Information

Factor	Levels	Values
Typ	3	1; 2; 3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Typ	2	26,17	13,08	0,09	0,912
Error	9	1259,50	139,94		
Total	11	1285,67			

Means

Typ	N	Mean	StDev	95% CI
1	4	36,25	7,72	(22,87; 49,63)
2	4	33,00	9,76	(19,62; 46,38)
3	4	33,25	16,28	(19,87; 46,63)

Pooled StDev = 11,8298

Modell 2

General Linear Model: Y versus x-mean; Typ

Method

Factor coding (-1; 0; +1)

Factor Information

Factor	Type	Levels	Values
Typ	Fixed	3	1; 2; 3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
x-mean	1	1217,55	1217,55	232,20	0,000
Typ	2	11,65	5,82	1,11	0,375
Error	8	41,95	5,24		
Lack-of-Fit	7	37,45	5,35	1,19	0,610
Pure Error	1	4,50	4,50		
Total	11	1285,67			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	34,167	0,661	51,69	0,000	
x-mean	1,1733	0,0770	15,24	0,000	1,03
Typ					
1	0,226	0,943	0,24	0,817	1,36
2	1,082	0,946	1,14	0,286	1,37

Regression Equation

Typ	Equation
1	$Y = 34,39 + 1,1733 \text{ x-mean}$
2	$Y = 35,25 + 1,1733 \text{ x-mean}$
3	$Y = 32,86 + 1,1733 \text{ x-mean}$

Ta hjälp av utskrifterna ovan för att svara på följande frågor.

- Sätt upp modell 1. Ange vilka egenskaper som krävs. Testa sen om det är skillnad i leveranstid mellan de tre typerna av kärra. 1,5p
- Sätt upp modell 2. Ange vilka egenskaper som krävs. Testa sen om det är skillnad i leveranstid mellan de tre typerna av kärra. 1,5p
- Jämför medelvärdena för y för varje säckkärretyp i de båda modellerna på något lämpligt sätt för att utvärdera vilken effekt x har i modell 2. I modell 2 sätter du då x-mean=0. 1p
- Är det viktigt att kontrollera för x=mängden av dryck som ska levereras? Testa detta på 5% signifikansnivå. 1p

3

Vid en industri vill man undersöka förmågan är att fokusera på jobbet som utförs vid olika avstånd mätt från ögat. Fyra olika avstånd ska analyseras. Responsvariabeln Y = fokuseringstiden i minuter kan variera för olika personer i personalen så därför valde man ut fem personer slumpmässigt. Experimentet blir då ett randomiserat block-försök med slumpmässiga blockeffekter utan interaktion. Tabell 25,8 kan vara till din hjälp.

Resultat:

Avstånd cm	Person nr				
	1	2	3	4	5
12	10	6	6	6	6
18	7	6	6	2	6
24	5	2	3	2	5
30	6	1	4	2	3

Analysis of Variance for Y

Source	SS	MS
block	36,70	9,175
Avstånd	44,20	14,733
Error	15,30	1,275
Total	96,20	

- Rita en interaktionsplot. Är det rimligt att anta att ingen interaktion finns? 1p
- Oavsett vad du kom fram till i a-uppgiften så ska du nu anta att modellen är giltig. Testa om Avstånd har någon effekt på fokuseringstiden. Sätt upp hypoteser. Signifikansnivå 5%.
2p
- Anta att obs y_{11} saknas. Skatta värdet för denna tomma cell genom att skatta $\mu_{11} = E[Y_{11}]$. Använd lämplig formel från kapitel 20. 1p
- Testa om det var nödvändigt att använda block. Sätt upp hypoteser. Signifikansnivå 5%.
1p

4

Låt X vara antalet inkommande samtal till en telefonväxel under en minut. X kan antas vara Poissonfördelad med parameter λ . Man observerade antalet samtal under 8 slumpmässigt valda minuter. De observerade värdena blev 2, 3, 4, 0, 1, 2, 5, 2 samtal.

Det gäller att $E[X] = \lambda$. Härled maximumlikelihood-skattningen av det förväntade antalet samtal under en minut.

Sannolikhetsfunktionen för en Poissonfördelad slumpvariabel är $f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, \dots$

3p