STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)

HT2019

Tentamen i Regressions- och tidsserieanalys, (4,5 hp)

Kurs: Regressionsanalys och undersökningsmetodik

2020-01-13

Skrivtid:	kl. 9.00 - 14.00 (5 timmar)
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Vidhäftade hjälpmedel:	Formelsamling och Statistiska tabeller (endast de tabeller som krävs)

- Tentamen består av 5 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.

- Svar med fullständiga redovisningar ska lämnas.
  - Använd endast skrivpapper som tillhandahålls i skrivsalen.
  - För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
  - Kontrollera alltid dina beräkningar och lösningar! Slarvfel kan också ge poängavdrag!
  - Använd minst fem värdesiffror i dina beräkningar (1,2345 och 1234,5 är exempel på tal med fem värdesiffror). I förekommande fall är det inte möjligt pga. avrundning i t.ex. SAS-utskrifter men utgå då ifrån det som är givet. Du kan dock avrunda ditt slutliga svar.

- Tentamen kan maximalt ge 100 poäng och för godkänt resultat krävs minst 50.

- Betygsgränser:

  A: 90 – 100 p
  B: 80 – 89 p
  C: 70 – 79 p
  D: 60 – 69 p
  E: 50 – 59 p
  Fx: 40 – 49 p
  F: 0 – 40 p

  OBS! Fx och F är underkända betyg som kräver omexamination. Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.

- Lösningsförslag läggs ut på Athena kort efter tentamen.

LYCKA TILL!

Uppgift 1. (30p)

Ett företag analyserar ett antal utvecklingsprojekt som genomfördes för ett par år sedan. Man antar att Y = den årliga vinsten av ett genomfört projekt kan förklaras av X = projektbudgetens storlek, båda i mkr, enligt en linjär modell:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

Dina underlag förlorades när din hårddisk kraschade men du hittar en utskrift med följande sammanställning:

Projekt i	1	2	3	4	5	6	7	8	9	10	Summa
x_i	3	5	5	7	9	9	11	13	13	15	90
y_i	10	14	12	18	23	18	20	22	21	19	177
x_i^2	9	25	25	49	81	81	121	169	169	225	954
y_i^2	100	196	144	324	529	324	400	484	441	361	3303
$x_i y_i$	30	70	60	126	207	162	220	286	273	285	1719
e_i	-2,45	-0,20	-2,20	2,05	5,30	0,30	0,55	0,80	-0,20	-3,95	0,00

a) Skatta parametrarna i modellen med minsta-kvadrat-metoden. (6p)

b) Beräkna residualvariansen. (4p)

c) Beräkna förklaringsgraden och förklara kortfattat vad detta är så att din chef förstår. (6p)

d) Beräkna ett 95% prediktionsintervall för den vinsten när $X = 10$. Förklara kortfattat vad detta intervall säger oss så att din chef förstår. (8p)

e) Åskådliggör X och Y i ett spridningsdiagram tillsammans med den skattade regressionslinjen och kommentera kortfattat. Är modellen lämplig för dessa data eller vill du justera den på något sätt? (6p)

Uppgift 2. (20p)

För var och en av följande deluppgifter ska du svara kortfattat. Hela uppgiften bör kunna redovisas på en eller max två A4-sidor. Du får gärna komplettera med bilder och skisser.

a) Förklara begreppen homoskedasticitet respektive heteroskedasticitet. Vilket av dessa föredras i en linjär regressionsmodell med intercept? (5p)

b) Förklara begreppet extrapolering. Vad ska man tänka på när man extrapolerar? (5p)

c) Vad är skillnaden mellan ett konfidensintervall för $\mu_{Y|X=x}$ och ett prediktionsintervall för $Y|X = x$? (5p)

d) Vad är skillnaden mellan felterm och residual? Är inte det samma sak? (5p)

Uppgift 3. (30p)

Man ville veta om studenters resultat på en delkurs B i ett visst ämne kunde prediceras av deras resultat på delkurs A (som kommer före B) och deras gymnasiebetyg. Från ett stickprov skattade man den linjära regressionsmodellen:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

där $Y =$ poäng på B-kursens slutprov, $X_1 =$ poäng på A-kursens slutprov och X_2 är betyg från gymnasiet (på en skala 0-20). Nedan finner du en SAS-utskrift med resultat. Notera att en hel del saknas och du måste räkna ut en del av dessa för hand.

a) Ange de fem grundantaganden som gäller för modellen ovan och ge exempel för tre av dessa på hur man kan undersöka om de är uppfyllda eller inte. (8p)

b) Genomför ett formellt test på 5% signifikansnivå för att pröva om modellen som helhet är signifikant. Ange hypoteserna som du testar, testvariabel och dess fördelning, beslutsregel med kritiska gräns samt beräkningar och slutsats med förklaring. (8p)

c) Beräkna två konfidensintervall för lutningskoefficienten β_2, ett med 95% och ett med 99% konfidensgrad. Vilken slutsats drar du från dessa två intervall? Vad är det för hypoteser som du implicit testar för genom att analysera intervallen? (8p)

d) Variansinflationsfaktorn, som finns i utskriften, indikerar om det finns ett starkt linjärt samband mellan förklaringsvariablerna. Ange om du anser att faktorn är för hög eller om den är ok (2p). Beräkna sedan utifrån det angivna värdet, korrelationen mellan X_1 och X_2. (4p)

Number of Observations Read	30
Number of Observations Used	30

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model		7486.785			
Error		3056.681			
Total		10543.467			

Parameter Estimates						
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Variance Inflation**
Intercept	1	0.28251	14.5251			0
Prov A	1	0.59358	0.15442			2.00603
Gymn Betyg	1	2.81326	1.20348			2.00603

Uppgift 4. (10p)

Logistisk regression har använts för att modellera sannolikheten att kunder uppskattar en viss mjukvaruprodukt beroende på hur mycket man har använt den:

$$\text{LogOdds}(Y = 1|x) = \beta_0 + \beta_1 x$$

där

$$Y = \begin{cases} 1, & \text{uppskattar produkten} \\ 0, & \text{uppskattar ej produkten} \end{cases} \qquad x = \text{användning i timmar}$$

Från ett stickprov om $n = 20$ observationer skattades modellen med SAS och man fick bland annat resultaten nedan; se även den anpassade kurvan i diagrammet på nästa sida.

Response Profile		
Ordered Value	Y	Total Frequency
1	0	11
2	1	9

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.3166	1.8405	5.5004	0.0190
X	1	0.4067	0.1705	5.6920	0.0170

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
X	1.502		

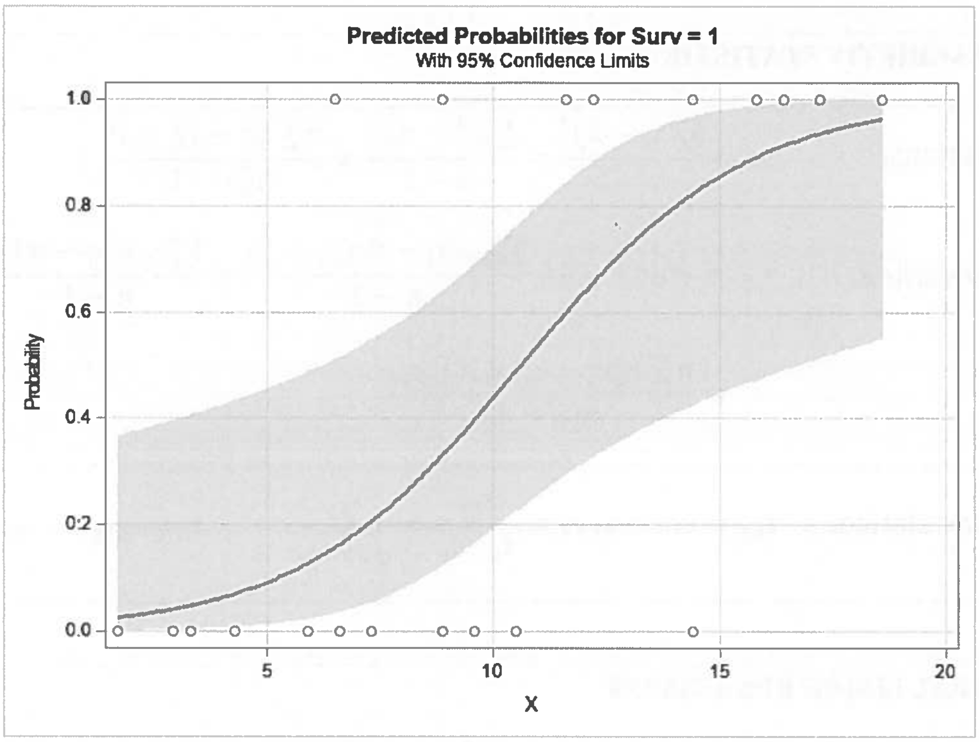
a) Välj <u>ett</u> av följande två alternativ att lösa. OBS! Du kan maximalt få 5p även om du gör båda.

Alt 1. Använd diagrammet på nästa sida och ange ett ungefärligt värde på x när sannolikheten $P(Y = 1|x) = 0.50$. Vad kallas detta värde? Observera att du inte behöver genomföra några beräkningar. (2p)

Alt 2. Beräkna det exakta värdet på x som ger $P(Y = 1|x) = 0.50$. TIPS: $Odds(Y = 1|x) = P(Y = 1|x)/P(Y = 0|x)$. Utnyttja formelsamlingen och använd punktskattningarna i utskriften ovan. (5p)

b) Beräkna ett 95% konfidensintervall för oddskvoten $\text{OR}(X)$. Intervallgränserna har ”ramlat bort” i utskriften ovan men du har standardfelet för skattningen av β_1. Utifrån detta intervall, skulle du påstå att antalet timmar har en signifikant effekt på sannolikheten $P(Y = 1|x)$? Om du inte lyckas beräkna intervallet kan du ändå ange vad du tittar efter med intervallet. (5p)

(forts. uppgift 4)



Uppgift 5. (10p)

I tabellen nedan visas y_t = produktionen (i mkr) för ett stort företag per kvartal under tre år. Du ser också att $\hat{T}_t$ = trenden redan har skattats med centrerade glidande medelvärden.

a) Trendrensa serien. Du kan använda antingen en additiv eller multiplikativ modell och du behöver <u>inte</u> motivera ditt val av modell, ange dock vilken du har använt och hur trendrensningen går till. (4p)

b) Beräkna de fyra justerade säsongsindexen, ett index för varje kvartal. (6p)

År	Kvartal	t	y_t	$\hat{T}_t$
2016	Kv1	1	156	*
	Kv2	2	76	*
	Kv3	3	116	168
	Kv4	4	304	180
2017	Kv1	5	196	196
	Kv2	6	132	208
	Kv3	7	188	220
	Kv4	8	328	236
2018	Kv1	9	268	252
	Kv2	10	188	268
	Kv3	11	260	*
	Kv4	12	384	*

FORMELSAMLING

DESKRIPTIV STATISTIK

Varians: $s_x^2 = \dfrac{\sum(x_i - \bar{x})^2}{n-1} = \dfrac{\sum x_i^2 - n\bar{x}^2}{n-1} = \dfrac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)}$

Kovarians: $s_{xy} = cov(x,y) = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \dfrac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{n-1}$

$= \dfrac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n(n-1)}$

Korrelation: $r_{xy} = corr(x,y) = \dfrac{s_{xy}}{s_x \cdot s_y} = \dfrac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}}$ Inferens: $t_{n-2} = \dfrac{r_{xy}\sqrt{n-2}}{\sqrt{1 - r_{xy}^2}}$

ENKEL LINJÄR REGRESSION

Modell: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ Betingat medelvärde för $Y|X = x$: $\mu_{Y|X=x} = \beta_0 + \beta_1 x$

Parameterskattningar och dessas varianser:

$b_1 = \dfrac{s_{xy}}{s_x^2} = r_{xy} \cdot \dfrac{s_y}{s_x}$ $s_{b_1}^2 = \dfrac{s_e^2}{(n-1)s_x^2} = \dfrac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$b_0 = \bar{y} - b_1\bar{x}$ $s_{b_0}^2 = s_e^2\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{(n-1)s_x^2}\right)$

Prediktion och skattat betingat medelvärde:	$\hat{y}_i = \hat{\mu}_{Y\|X=x_i} = b_0 + b_1 x_i$
Prediktionsintervall för prediktionen $\hat{y}_i$ givet $X = x$:	$\hat{y}_i \pm t_{n-2,\alpha/2} \cdot \sqrt{s_e^2\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}\right)}$
Konfidensintervall för betingade medelvärdet $\mu_{Y\|X=x}$ givet $X = x$:	$\hat{\mu}_{Y\|X=x} \pm t_{n-2,\alpha/2} \cdot \sqrt{s_e^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}\right)}$

ICKE-LINJÄR REGRESSION, exempel

Andragradspolynom:	$\hat{y}_i = a + b_1 x_i + b_2 x_i^2$
Exponentiell:	$\widehat{\ln(y_i)} = a + bx_i$ $\hat{y}_i = \exp(a + bx_i) = (e^a)(e^b)^{x_i} = (a') \cdot (b')^{x_i}$ $\widehat{\log_{10}(y_i)} = a + bx_i$ $\hat{y}_i = (10^a)(10^b)^{x_i} = (a') \cdot (b')^{x_i}$

ENKEL OCH MULTIPEL LINJÄR REGRESSION (sätt $k = 1$ om enkel regression)

Residualvarians: $$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} = \frac{SSE}{n-k-1} = MSE$$

Kvadratsummor: $$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (n-1)s_y^2 = SSR + SSE$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = (n-k-1)s_e^2$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = [\text{om \underline{enkel} regression}] = b^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Förklaringsgrad: $$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad R_{\text{adj}}^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

Inferens för β_j: KI: $b_j \pm t_{n-k-1,\alpha/2} \cdot s_{b_j}$ Test: $t_{n-k-1} = \dfrac{b_j - \beta_j^*}{s_{b_j}}$

Test för hela modellen: $$F_{k;n-k-1} = \frac{SSR/K}{SSE/(n-K-1)} = \frac{MSR}{MSE}$$

Beräkningsformler för KORRELATION och REGRESSIONSKOEFFICIENT

$$b_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})/(n-1)}{\sum(x_i - \bar{x})^2/(n-1)} = \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x^2} \cdot \frac{s_x s_y}{s_x s_y} = \frac{s_{xy}}{s_x s_y} \cdot \frac{s_y}{s_x} = r_{xy} \cdot \frac{s_y}{s_x}$$

$$r_{xy} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n\sum y_i^2 - (\sum y_i)^2}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{\sum x_i^2 - n\bar{x}^2} \cdot \sqrt{\sum y_i^2 - n\bar{y}^2}}$$

$$= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \cdot \sqrt{\sum(y_i - \bar{y})^2}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})/(n-1)}{\sqrt{\sum(x_i - \bar{x})^2/(n-1)} \cdot \sqrt{\sum(y_i - \bar{y})^2/(n-1)}}$$

$$= \frac{s_{xy}}{\sqrt{s_x^2} \cdot \sqrt{s_y^2}} = \frac{s_{xy}}{s_x s_y} = \frac{s_{xy}}{s_x s_y} \cdot \frac{s_x}{s_x} = \frac{s_{xy}}{s_x^2} \cdot \frac{s_x}{s_y} = b_1 \cdot \frac{s_x}{s_y}$$

TIDSSERIEANALYS

- Komponenter

Additiv modell: $Y_t = T_t + S_t + C_t + E_t$ Multiplikativ modell: $Y_t = T_t \cdot S_t \cdot C_t \cdot E_t$

där T = trend, S = säsong, C = cyklisk/konjunktur samt E = slumpkomponent

- Skattning av trendkomponenten:

– med glidande medelvärden utan säsongvariation, exempel:

3-punkter centrerat: $\hat{T}_t = \frac{1}{3} \cdot y_{t-1} + \frac{1}{3} \cdot y_t + \frac{1}{3} \cdot y_{t+1}$

5-punkter centrerat: $\hat{T}_t = \frac{1}{5} \cdot y_{t-2} + \frac{1}{5} \cdot y_{t-1} + \frac{1}{5} \cdot y_t + \frac{1}{5} \cdot y_{t+1} + \frac{1}{5} \cdot y_{t+2}$

– med centrerade glidande medelvärden med säsongvariation, exempel:

halvårsdata: $\hat{T}_t = \frac{1}{4} \cdot y_{t-1} + \frac{1}{2} \cdot y_t + \frac{1}{4} \cdot y_{t+1}$

kvartalsdata: $\hat{T}_t = \frac{1}{8} \cdot y_{t-2} + \frac{1}{4} \cdot y_{t-1} + \frac{1}{4} \cdot y_t + \frac{1}{4} y_{t+1} + \frac{1}{8} \cdot y_{t+2}$

månadsdata: $\hat{T}_t = \frac{1}{24} \cdot y_{t-6} + \frac{1}{12} \cdot y_{t-5} + \cdots + \frac{1}{12} \cdot y_{t+5} + \frac{1}{24} \cdot y_{t+6}$

– med regressionsanalys, linjär trend och exponentiell trend:

Modell: $Y_t = \beta_0 + \beta_1 t + \varepsilon_1$ Skattad modell: $\hat{y}_t = b_0 + b_1 t = \hat{T}_t$

$\ln Y_t = \beta_0 + \beta_1 t + \varepsilon_1$ $\hat{y}_t = \exp(b_0 + b_1 t) = \hat{T}_t$

- Justering av säsongsindex $\bar{S}_j$ med p säsonger (halvår, kvartal el. månader osv.):

Additiv modell: $S_j^+ = \bar{S}_j - \left(\frac{\sum \bar{S}_i}{p}\right)$ Multiplikativ modell: $S_j^+ = \frac{\bar{S}_j}{(\sum \bar{S}_i / p)}$

- Trend- och säsongsrensning:

Additiv modell: $y_t - \hat{T}_t$ resp. $y_t - S_t^+$ Multiplikativ modell: $y_t / \hat{T}_t$ resp. y_t / S_t^+

LOGISTISK REGRESSION och ODDS

Odds för en händelse A:	$\text{Odds}(A) = \frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1-P(A)} \Leftrightarrow P(A) = \frac{\text{Odds}(A)}{1+\text{Odds}(A)}$
Oddskvot för händelsen A mot B:	$\text{OR} = \frac{\text{Odds}(A)}{\text{Odds}(B)}$

- Logistisk regression:

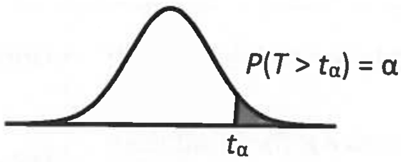
Enkel modell:	$P(Y=1\|x) = \frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)} = \frac{1}{1+\exp(-\beta_0-\beta_1 x)}$ $P(Y=0\|x) = \frac{1}{1+\exp(\beta_0+\beta_1 x)}$ $\text{Odds}(Y=1\|x) = \exp(\beta_0+\beta_1 x)$ $\text{LogOdds}(Y=1\|x) = \beta_0+\beta_1 x$
Multipel modell:	$\text{LogOdds}(Y=1\|x_1,\ldots x_k) = \beta_0+\beta_1 x_1+\cdots+\beta_k x_k$

Intercept β_0:	$P(Y=1\|x_1=\cdots=x_k=0) = \frac{\exp(\beta_0)}{1+\exp(\beta_0)}$
Oddskvot för $Y=1$ när $X_j = x_j+1$ mot $X_j = x_j$:	$\text{OR}(X_j) = \frac{\text{Odds}\left(Y=1\middle\|x_j+1, \text{allt annat lika}\right)}{\text{Odds}\left(Y=1\middle\|x_j, \text{allt annat lika}\right)} = \exp(\beta_j)$
KI för $\text{OR}(X_j)$:	$\left(\exp(b_j - z_{\alpha/2}\cdot s_{b_j})\,;\,\exp(b_j + z_{\alpha/2}\cdot s_{b_j})\right)$

TABELL 3. t-fördelningens kvantiler

$T \in t(\nu)$ där ν = antal frihetsgrader.

Vilket värde har t_α om $P(T > t_\alpha) = \alpha$ där α är en given sannolikhet. Utnyttja även $P(T \leq -t_\alpha) = P(T > t_\alpha)$.



ν	α = 0,1	0,05	0,025	0,010	0,005	0,0025	0,0010	0,0005
1	3,078	6,314	12,706	31,821	63,657	127,321	318,309	636,619
2	1,886	2,920	4,303	6,965	9,925	14,089	22,327	31,599
3	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659
30	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
35	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591
40	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
45	1,301	1,679	2,014	2,412	2,690	2,952	3,281	3,520
50	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496
55	1,297	1,673	2,004	2,396	2,668	2,925	3,245	3,476
60	1,296	1,671	2,000	2,390	2,660	2,915	3,232	3,460
65	1,295	1,669	1,997	2,385	2,654	2,906	3,220	3,447
70	1,294	1,667	1,994	2,381	2,648	2,899	3,211	3,435
75	1,293	1,665	1,992	2,377	2,643	2,892	3,202	3,425

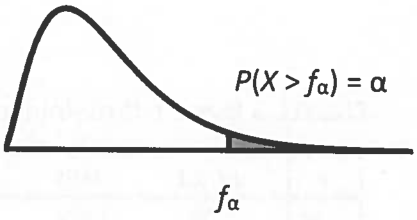
Forts. nästa sida

TABELL 3 forts. *t*-fördelningens kvantiler

ν	α = 0,1	0,05	0,025	0,010	0,005	0,0025	0,0010	0,0005
80	1,292	1,664	1,990	2,374	2,639	2,887	3,195	3,416
85	1,292	1,663	1,988	2,371	2,635	2,882	3,189	3,409
90	1,291	1,662	1,987	2,368	2,632	2,878	3,183	3,402
95	1,291	1,661	1,985	2,366	2,629	2,874	3,178	3,396
100	1,290	1,660	1,984	2,364	2,626	2,871	3,174	3,390
125	1,288	1,657	1,979	2,357	2,616	2,858	3,157	3,370
150	1,287	1,655	1,976	2,351	2,609	2,849	3,145	3,357
175	1,286	1,654	1,974	2,348	2,604	2,843	3,137	3,347
200	1,286	1,653	1,972	2,345	2,601	2,839	3,131	3,340
300	1,284	1,650	1,968	2,339	2,592	2,828	3,118	3,323
400	1,284	1,649	1,966	2,336	2,588	2,823	3,111	3,315
500	1,283	1,648	1,965	2,334	2,586	2,820	3,107	3,310
1000	1,282	1,646	1,962	2,330	2,581	2,813	3,098	3,300
2000	1,282	1,646	1,961	2,328	2,578	2,810	3,094	3,295
3000	1,282	1,645	1,961	2,328	2,577	2,809	3,093	3,294
4000	1,282	1,645	1,961	2,327	2,577	2,809	3,092	3,293
5000	1,282	1,645	1,960	2,327	2,577	2,808	3,092	3,292

TABELL 5. *F*-fördelningens kvantiler

$X \in F(\nu_1,\nu_2)$ där ν_1,ν_2 = antal frihetsgrader i täljaren respektive nämnaren. Vilket värde har f_α om $P(X > f_\alpha) = \alpha$ där α är en given sannolikhet.



α = 0,05

	ν_1 = 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ν_2 = 1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,7	245,4	245,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04	2,01	1,99	1,96
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97	1,94	1,92	1,89
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72	1,69	1,67

Forts. nästa sida

TABELL 5 forts. *F*-fördelningens kvantiler

***α* = 0,05**

	v_1 = 16	17	18	19	20	25	30	35	40	50	60	70	80	100	∞
v_2 = 1	246,5	246,9	247,3	247,7	248,0	249,3	250,1	250,7	251,1	251,8	252,2	252,5	252,7	253,0	254,3
2	19,43	19,44	19,44	19,44	19,45	19,46	19,46	19,47	19,47	19,48	19,48	19,48	19,48	19,49	19,50
3	8,69	8,68	8,67	8,67	8,66	8,63	8,62	8,60	8,59	8,58	8,57	8,57	8,56	8,55	8,53
4	5,84	5,83	5,82	5,81	5,80	5,77	5,75	5,73	5,72	5,70	5,69	5,68	5,67	5,66	5,63
5	4,60	4,59	4,58	4,57	4,56	4,52	4,50	4,48	4,46	4,44	4,43	4,42	4,41	4,41	4,37
6	3,92	3,91	3,90	3,88	3,87	3,83	3,81	3,79	3,77	3,75	3,74	3,73	3,72	3,71	3,67
7	3,49	3,48	3,47	3,46	3,44	3,40	3,38	3,36	3,34	3,32	3,30	3,29	3,29	3,27	3,23
8	3,20	3,19	3,17	3,16	3,15	3,11	3,08	3,06	3,04	3,02	3,01	2,99	2,99	2,97	2,93
9	2,99	2,97	2,96	2,95	2,94	2,89	2,86	2,84	2,83	2,80	2,79	2,78	2,77	2,76	2,71
10	2,83	2,81	2,80	2,79	2,77	2,73	2,70	2,68	2,66	2,64	2,62	2,61	2,60	2,59	2,54
11	2,70	2,69	2,67	2,66	2,65	2,60	2,57	2,55	2,53	2,51	2,49	2,48	2,47	2,46	2,40
12	2,60	2,58	2,57	2,56	2,54	2,50	2,47	2,44	2,43	2,40	2,38	2,37	2,36	2,35	2,30
13	2,51	2,50	2,48	2,47	2,46	2,41	2,38	2,36	2,34	2,31	2,30	2,28	2,27	2,26	2,21
14	2,44	2,43	2,41	2,40	2,39	2,34	2,31	2,28	2,27	2,24	2,22	2,21	2,20	2,19	2,13
15	2,38	2,37	2,35	2,34	2,33	2,28	2,25	2,22	2,20	2,18	2,16	2,15	2,14	2,12	2,07
16	2,33	2,32	2,30	2,29	2,28	2,23	2,19	2,17	2,15	2,12	2,11	2,09	2,08	2,07	2,01
17	2,29	2,27	2,26	2,24	2,23	2,18	2,15	2,12	2,10	2,08	2,06	2,05	2,03	2,02	1,96
18	2,25	2,23	2,22	2,20	2,19	2,14	2,11	2,08	2,06	2,04	2,02	2,00	1,99	1,98	1,92
19	2,21	2,20	2,18	2,17	2,16	2,11	2,07	2,05	2,03	2,00	1,98	1,97	1,96	1,94	1,88
20	2,18	2,17	2,15	2,14	2,12	2,07	2,04	2,01	1,99	1,97	1,95	1,93	1,92	1,91	1,84
25	2,07	2,05	2,04	2,02	2,01	1,96	1,92	1,89	1,87	1,84	1,82	1,81	1,80	1,78	1,71
30	1,99	1,98	1,96	1,95	1,93	1,88	1,84	1,81	1,79	1,76	1,74	1,72	1,71	1,70	1,62
35	1,94	1,92	1,91	1,89	1,88	1,82	1,79	1,76	1,74	1,70	1,68	1,66	1,65	1,63	1,56
40	1,90	1,89	1,87	1,85	1,84	1,78	1,74	1,72	1,69	1,66	1,64	1,62	1,61	1,59	1,51
45	1,87	1,86	1,84	1,82	1,81	1,75	1,71	1,68	1,66	1,63	1,60	1,59	1,57	1,55	1,47
50	1,85	1,83	1,81	1,80	1,78	1,73	1,69	1,66	1,63	1,60	1,58	1,56	1,54	1,52	1,44
60	1,82	1,80	1,78	1,76	1,75	1,69	1,65	1,62	1,59	1,56	1,53	1,52	1,50	1,48	1,39
70	1,79	1,77	1,75	1,74	1,72	1,66	1,62	1,59	1,57	1,53	1,50	1,49	1,47	1,45	1,35
80	1,77	1,75	1,73	1,72	1,70	1,64	1,60	1,57	1,54	1,51	1,48	1,46	1,45	1,43	1,32
100	1,75	1,73	1,71	1,69	1,68	1,62	1,57	1,54	1,52	1,48	1,45	1,43	1,41	1,39	1,28
∞	1,64	1,62	1,60	1,59	1,57	1,51	1,46	1,42	1,39	1,35	1,32	1,29	1,27	1,24	1,00

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 13/1-2020

Sal: Värtasalen

Tenta: Regressions- och tidsserieanalys

Kurs: Regressionsanalys och undersökningsmetodik

ANONYMKOD: 0058-CRO

☒ Jag godkänner att min tenta får läggas ut anonymt på hemsidan som studentsvar.

OBS! SKRIV ÄVEN PÅ BAKSIDAN AV SKRIVBLADEN

Markera besvarade uppgifter med kryss

1	2	3	4	5	6	7	8	9	Antal inl. blad
X	X	X	X	X					10
Lär.ant. 30	20	26	9	8					

POÄNG	BETYG	Lärarens sign.
93	A	ME

SU, STATISTIK

Skrivsal: ____________ **Anonymkod:** 0058-CRO **Blad nr:** 1

1.

a)

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$$

$$\hat{y}_i = b_0 + b_1 x_1$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{177}{10} = 17{,}7 \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{90}{10} = 9$$

$$b_1 = \frac{S_{xy}}{S_x^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{1719 - 10 \cdot 17{,}7 \cdot 9}{954 - 10 \cdot 9^2}$$

$$= \frac{126}{144} = 0{,}875 \quad R$$

$$b_0 = \bar{y} - b_1\bar{x} = 17{,}7 - 0{,}875 \cdot 9 = 9{,}825 \quad R$$

Svar: $b_1 = 0{,}875$ $\quad b_0 = \alpha = 9{,}825$

/6

b)

$$S_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1} = \frac{59{,}85}{10-1-1} = 7{,}48125 \quad R$$

i	1	2	3	4	5	6	7	8	9	10	Summa
e_i	-2,45	-0,20	-2,20	2,05	5,30	0,30	0,55	0,80	-0,20	-3,95	0
e_i^2	6,0025	0,04	4,84	4,2025	28,09	0,09	0,3025	0,64	0,04	15,6025	59,85

$$\sum_{i=1}^{n} e_i^2 = 59{,}85$$

Svar: $S_e^2 = 7{,}48125 \approx 7{,}4813$

/4

SU, STATISTIK

Skrivsal: ____________ Anonymkod: 0058-CRO Blad nr: 2

1. c). $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{59,85}{170,1} \approx 0,64815$

$SSE = \sum_{i=1}^{n} e_i^2 = 59,85$ $\left\{ \begin{array}{l} \text{Se uträkning} \\ \text{föregående deluppgift} \end{array} \right.$

$SST = (n-1) S_y^2 = (10-1) \cdot 18,9 = 170,1$

$S_y^2 = \frac{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}{n-1} = \frac{3303 - 10 \cdot 17,7^2}{10-1} = \frac{170,1}{9} = 18,9$

Förklaringsgraden är ett mått på hur mycket av variationen i Y som kan förklaras av dem oberoende förklaringsvariablerna X_k i modellen. Eftersom detta är en enkel regressionsmodell anger måttet hur mycket av variationen i Y som kan förklaras av X_1.

Cirka 65% av variationen i den årliga vinsten av ett genomfört projekt kan alltså förklaras av variationen i projektbudgetens storlek.

6

1 d) 95% PI

$$\hat{y}_i \pm t_{n-2;\frac{\alpha}{2}} \cdot \sqrt{S_e^2\left(1+\frac{1}{n}+\frac{(x-\bar{x})^2}{(n-1)S_x^2}\right)}$$ R

$$\hat{y}_i = 9,825 + 0,875x$$

$$\hat{y}_i|_{x=10} = 9,825 + 0,875(10) = 18,575$$ R

95% PI för $\hat{y}_i|x=10$

$$18,575 \pm t_{8;\,0,025} \cdot \sqrt{S_e^2\left(1+\frac{1}{n}+\frac{(x-\bar{x})^2}{(n-1)S_x^2}\right)}$$

$$18,575 \pm 2,306 \cdot \sqrt{7,48125\left(1+\frac{1}{10}+\frac{(10-9)^2}{(10-1)\,1,67}\right)}$$

$$18,575 \pm 2,306 \cdot \sqrt{7,48125 \cdot 1,106944444}$$

$$18,575 \pm 2,306 \cdot 2,877729682$$

$$18,575 \pm 6,636044648$$

$$[11,939;\ 25,211]$$ R

med ca ~~På~~ 95% sannolikhet ~~signifikansnivå~~ kan vi prediktera att dem framtida värdena på Y kommer ligga inom det givna intervallet då $x=10$

Då projektbudgetens str är 10 mkr kommer Y = den årliga vinsten av ett genomfört projekt i framtiden ligga inom det givna intervallet med ca 95% (säkerhet) sannolikhet

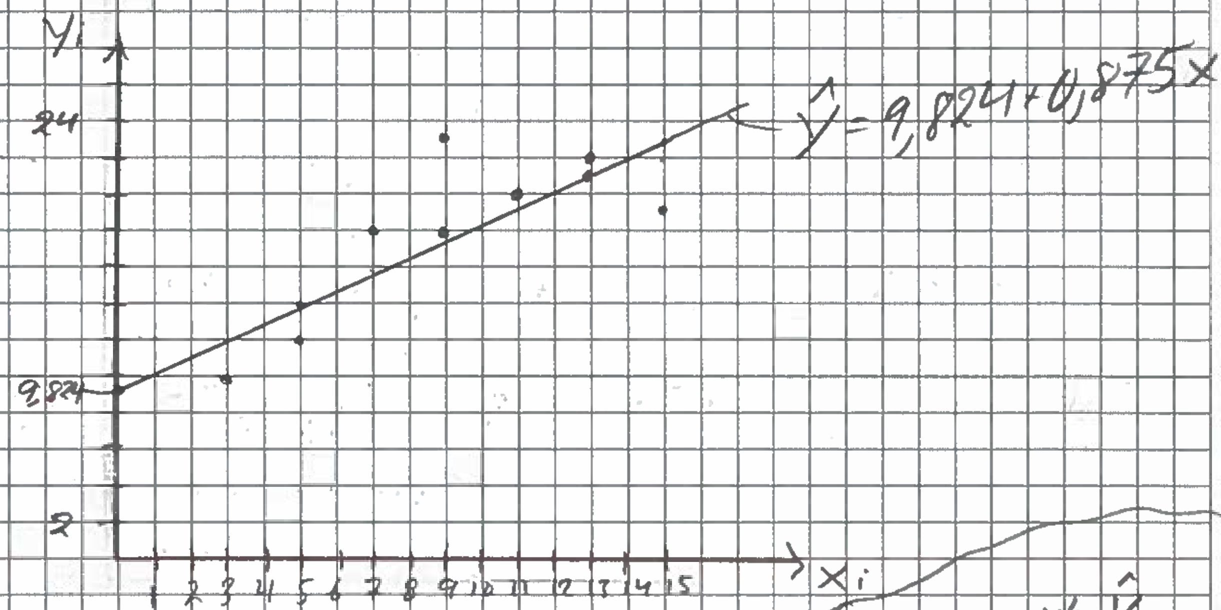
6

SU, STATISTIK

Skrivsal: ______ Anonymkod: 0058-CRO Blad nr: 3

1 e)

$\hat{y} = 9{,}824 + 0{,}875x$



$\hat{y}|x{=}15 = 9{,}824 + 0{,}875 \cdot 15 = 22{,}949$ } räknar ut $\hat{y}$ för ett givet x-värde för att kunna rita linjen

modellen ser ut att vara lämplig. Regressionslinjen går igenom svärmen av plotten och det är ungefär lika mycket punkter under som över linjen

Men var ligger de under resp över?

Ka-

30

SU, STATISTIK

Skrivsal: ____________ Anonymkod: 0058-CRO Blad nr: 4

2 a) homoskedastisitet är fallet då feltermerna har en konstant varians oberoende av värdena på X:n dvs

$\sigma^2_{\varepsilon i} = \sigma^2_{\varepsilon}$ R

Heteroskedasticitet är det omvända dvs då feltermerna inte har en konstant varians oberoenda av värdena på X:n dvs $\sigma^2_{\varepsilon i} \neq \sigma^2_{\varepsilon}$

I en linjär regressionsmodell föredras homoskedastisitet annars kan det vara tecken på det inte råder ett linjärt samband mellan X och Y eller att det råder linjära beroenden mellan prediktorvariablerna i en multipel linjär regressionsmodell.

/5

b) extrapolering sker när man använder en modell för att undersöka $\hat{Y}$ för sådanna värden på X som inte fanns att tillgå när man skattade koefficienterna i modellen. Det finns härmed en risk att modellen inte håller för dessa värden på X som inte användes vid skattningen av modellen $\hat{Y}$ BRA!

→

man ska alltså vara försiktig när man extrapolerar, vara uppmärksam på avvikande observationer så kallade outliers då detta kan vara en indikator på att modellen inte är användbar här. R

c) Ett konfidensintervall anger att med exempelvis 95% signifikansnivå kommer medelvärdet för stickprovet sammanfalla med det faktiska medelvärdet för populationen. Detta medelvärde sammanfaller i sin tur inom det givna intervallet.

Ett prediktionsintervall anger istället vart framtida y men som vi inte ännu observerat kommer att hamna med ex 95% säkerhet för 95% PI. prediktionsintervallet är längre än konfidensintervallet eftersom det även råder en naturlig variation hos observationerna kring medelvärdet. BRA!

SU, STATISTIK

Skrivsal: ______ Anonymkod: 0058-CRO Blad nr: 5

2 d)

feltermerna betäcknas ε_i och är en slumpterm som inte går att observera i verkligheten därför används istället residualerna e_i som proxyn för ε_i vid beräkning och analysering av att modelantaganden för en linjär regressionmodell är uppfyllda exempelvis att feltermerna är normalfördelade och har en konstant varians, är oberoende sinsemellan och oberoende av x'n. Vi undersöker i blandannat olika typer av spridningsdiagram

(Definiera gärna tydligare vad ε_i och e_i är)

/5

20

SU, STATISTIK

Skrivsal: ____________ Anonymkod: 0058-CRO Blad nr: 6

3.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

a) modelantaganden multipel linjär regressionsmodell

1. Det råder ett linjärt samband mellan den beroende y-variabeln och förklaringsvariablerna $x_1, +..+, x_K$ ~~(plus en felterm ε)~~ kan missförstås R

2. feltermerna är oberoende sinsemellan R

3. feltermerna är oberoende av värdena på x'n R

4. feltermerna är nfl med väntevärdet 0 $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ R

5. feltermerna har en konstant varians x'n är alltså oberoende av feltermerna $\sigma_{\varepsilon_i}^2 = \sigma_\varepsilon^2$. Feltermerna är med andra ord homoskedastiska R

6. prediktorvariablerna får korrelera med varandra men dem får inte vara ~~tra~~ linjärt bestämda av varandra så det existerar en uppsättning R FORA! Konstanter

→

Man kan undersöka att antagande nr. 6 är uppfyllt genom att beräkna VIF för modellen. VIF > 10 eller t.om större är 5 kan vara tecken på multikollinearitet R

Man kan undersöka att antagande nr. 5 är uppfyllt genom att plotta residualerna mot dem motsvarande X_i värdena. Ev mönster i spridningsdiagrammet R kan vara en indikator på heteroskedasticitet

man kan undersöka antagande 1 att det råder ett linjärt samband mellan Y och X'n genom att beräkna F-värdet och jämföra det med motsvarande F-krit för att se om modellen som helhet är signifikant skild från noll R

/8

SU, STATISTIK

Skrivsal: ________ Anonymkod: 0058-LRO Blad nr: 7

3 b) Hypoteser

$H_0: \beta_1 = \beta_2 = 0$ $\alpha = 0{,}05$

H_a: någon av β_1 eller $\beta_2 \neq 0$ R

Testvariabel

$F = \frac{MSR}{MSE} \sim F_{2;27}$ R

Frihetsgrader	
modell = k	= 2
Error	= n-k-1 = 27
Total	= n-1 = 29

Beslutsregel

Förkasta H_0 om $F_{obs} > F_{kritisk}$

$F_{kritisk} = F_{2;27;0,05} \approx 3{,}32 - 3{,}39$

(obs finns ej 27 frihetsgrader i tabellen men finns för 30 och 25 så någonstans mellan dessa två värden är $F_{kritisk}$) BRA!

$MSR = SSR / k = 7486{,}785/2 = 3743{,}3925 \approx 3743{,}4$

$MSE = SSE / n-k-1 = 3056{,}681/27 = 113{,}2104074 \approx 113{,}21$

$MST = SST / n-1 = 10543{,}467/29 = 363{,}5678276 \approx 363{,}57$

Beräkningar

$F_{obs} = \frac{MSR}{MSE} = \frac{3743{,}4}{113{,}21} \approx 33{,}066$ R

Slutsatser

$F_{obs} = 33{,}066 > F_{krit} = 3{,}32$ någon $- 3{,}39$

Vi kan alltså på 5% signifikansnivå förkasta nollhypotesen och modellen som helhet är signifikant skild från noll. X variablerna ämnar alltså förklara variationen i y i ett ensemble med varandra. R

/8

3c) 95% KI för β_2

$b_2 \pm t_{27;0,025} \cdot S_{b_2}$

$2,81326 \pm 2,052 \cdot 1,20348$

$2,81326 \pm 2,46954096$

$0,34371904 ; 5,28280096$

$[0,3437 ; 5,2828]$ R

Med 95% säkerhet kommer det sanna värdet på b_2 dvs β_2 hamna inom detta intervall

99% KI för β_2

$b_2 \pm t_{27;0,005} \cdot S_{b_2}$ 8

$2,81326 \pm 2,771 \cdot 1,20348$

$2,81326 \pm 3,33484308$

$-0,52158308 ; 6,14810308$

$[-0,52158 ; 6,1481]$ R

Med 99% säkerhet kommer det sanna värdet på b_2 dvs β_2 hamna inom detta intervall

implicit testar man för vilka värden på β_2 som koefficienten är signifikant skild från noll dvs vilka värden på β_2 koefficienten som krävs för att förklaringsvariabeln ämnana förklara variationen i Y dvs ger mervärde för modellen. Man ser efter om noll ligger i intervallet

$H_0 : \beta_2 = 0 \mid x_1$ i modellen

$H_1 : \beta_2 \neq 0 \mid$ — " —

7

SU, STATISTIK

Skrivsal: ______ Anonymkod: 0058-CRO Blad nr: 8

3 d) VIF värdet på 2,00603 anser jag är okej tumregeln är att VIF ska vara under 10 eller i vissa fall under 5. 2,00603 är alltså okej och tyder på att ingen multikollinearitet går att misstänka i modellen.

3

$r_{x_1x_2} = Sb_1 \cdot Sb_2 = 0,15442 \cdot 1,20348 \approx 0,18584$

26

SU, STATISTIK

Skrivsal: ______ Anonymkod: 0058CRO Blad nr: 9

4 a)

$$\text{Odds}(y=1|x) = \frac{p(y=1|x)}{p(y=0|x)} = \frac{0,5}{0,5} = 1$$

$$\text{Log Odds}(y=1|x) = -4,3166 + 0,4067x = \log 1$$

$$-4,3166 + 0,4067x = 0$$

Löser ut x

$$0,4067x = 4,3166$$

$$x = \frac{4,3166}{0,4067} = 10,6137\,2019 \approx 10,614$$

BRA!

/5

b) 95% KI för $OR(x_j)$

$$\left(\exp\left(b_j - z_{\frac{\alpha}{2}} \cdot s_{b_j}\right); \exp\left(b_j + z_{\frac{\alpha}{2}} \cdot s_{b_j}\right)\right)$$

$$\exp(0,4067 - 1,96 \cdot 0,1705); \exp(0,4067 + 1,96 \cdot 0,1705)$$

$$\exp(0,07252); \exp(0,74088)$$

$$e^{0,07252}, e^{0,74088}$$

$$1,07521431; 2,0977807 5$$

$$[1,0752; 2,0978]$$ R

punktskattningen av $OR(x) = 1,502$. Detta värde faller inom det givna intervallet och jag vill därför påstå att antalet timmar har en signifikant effekt på sannolikheten $P(Y=1|x)$

KI täcker ej värdet 1

~~5~~ 4

9

SU, STATISTIK

Skrivsal: ______ Anonymkod: 0058-CRO Blad nr: 10

5.

a) Trendrensning för Additiv modell

t	$= Y_t - \hat{T}_t$		
1	$Y_1 - \hat{T}_1 = 156 - * = \cancel{0}$	Kv 1	räknas ej
2	$Y_2 - \hat{T}_2 = 76 - * = \cancel{0}$	Kv 2	– " –
3	$Y_3 - \hat{T}_3 = 116 - 168 = -52$	Kv 3	
4	$Y_4 - \hat{T}_4 = 304 - 180 = 124$	Kv 4	
5	$Y_5 - \hat{T}_5 = 196 - 196 = 0$	Kv 1	
6	$Y_6 - \hat{T}_6 = 132 - 208 = -76$	Kv 2	
7	$Y_7 - \hat{T}_7 = 188 - 220 = -32$	Kv 3	R
8	$Y_8 - \hat{T}_8 = 328 - 236 = 92$	Kv 4	
9	$Y_9 - \hat{T}_9 = 268 - 252 = 16$	Kv 1	
10	$Y_{10} - \hat{T}_{10} = 188 - 268 = -80$	Kv 2	
11	$Y_{11} - \hat{T}_{11} = 260 - * = \cancel{0}$	Kv 3	– " –
12	$Y_{12} - \hat{T}_{12} = 384 - * = \cancel{0}$	Kv 4	– " –

4

b)

$\bar{S}_{Kv\,1} = (0 + 0 + 16) / \cancel{3}\,2 = 16/3 = 5{,}333333333$

$\bar{S}_{Kv\,2} = (0 - 76 - 80) / \cancel{3}\,2 = -156/3 = -52$

$\bar{S}_{Kv\,3} = (-52 - 32 + 0) / \cancel{3}\,2 = -84/3 = -28$

$\bar{S}_{Kv\,4} = (124 + 92 + 0) / 2 = 216/\cancel{3}\,2 = 72$

Fortsättning b → nästa sida

	Kv 1	Kv 2	Kv 3	Kv 4	Σ
$\bar{S}_j$	5,3333	−52	−28	72	−2,6667
S_j^+	5,999975	−51,333325	−27,333325	72,666675	0

Rätt idé, fel siffror in

$$\frac{\Sigma \bar{S}_i}{P} = \frac{-2,6667}{4} = -0,666675$$

$$S_j^+ = \bar{S}_j - \left(\frac{\Sigma \bar{S}_i}{P}\right)$$

$$S_1^+ = 5,3333 - (-0,666675) = 5,3333 + 0,666675 = 5,999975$$

$$S_2^+ = -52 - (-0,666675) = -52 + 0,666675 = -51,333325$$

$$S_3^+ = -28 - (-0,666675) = -28 + 0,666675 = -27,333325$$

$$S_4^+ = 72 - (-0,666675) = 72 + 0,666675 = 72,666675$$

$\bar{S}_j$ = säsongsindex

$\bar{S}_j^+$ = justerade säsongsindex