STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)
Ulf Högnäs

VT2018

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2018-04-19

Skrivtid:	kl. 15.00 - 20.00
Godkända hjälpmedel:	Miniräknare utan lagrade formler och text
Bifogade hjälpmedel:	Häftet *Formelsamling och Tabeller över statistiska fördelningar* (återlämnas efter skrivningen)

- Tentamen består av 7 uppgifter, i förekommande fall uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.

- **Uppgift 1 – 5**: Svar lämnas på särskild **SVARSBILAGA**,
  - Flervalsfrågor där ett av fem alternativ är korrekt svar.
  - Har fler än ett svarsalternativ markerats för en deluppgift ges noll poäng.
  - Uträkningar lämnas ej in för dessa, om uträkningar ändå lämnas in kommer de inte att beaktas vid bedömningen.

- **Uppgift 6 – 7**: Svar med **FULLSTÄNDIGA REDOVISNINGAR** ska lämnas in.
  - 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!

- Tentamen kan maximalt ge 60 + 40 = 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 Mondo kort efter tentamen.

LYCKA TILL!

Uppgift 1

Följande tabell visar årslönerna i tkr (1000 kr) för tolv anställda på ett mindre företag. Lönerna har ordnats i storleksordning från lägsta till högsta lön:

nr	1	2	3	4	5	6	7	8	9	10	11	12
Lön (tkr)	423	432	433	516	555	586	600	640	680	1798	2115	2408

a) Beräkna 10:e och 90:e percentilen enligt metoden som beskrivs i kurslitteraturen för årslönerna på företaget. (6p)

A. $P_{10} = 425{,}7 \quad P_{90} = 2320{,}1$

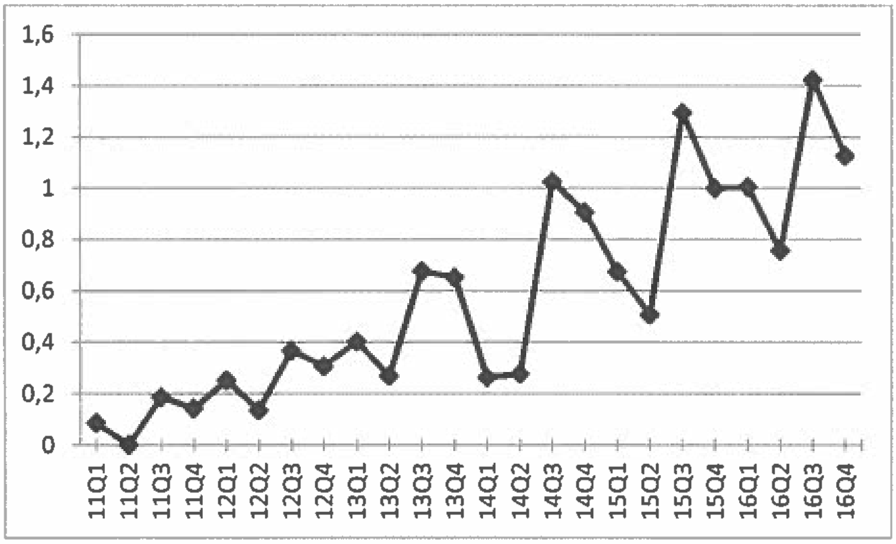
B. $P_{10} = 427{,}5 \quad P_{90} = 2261{,}5$

C. $P_{10} = 423{,}9 \quad P_{90} = 2378{,}7$

D. $P_{10} = 432{,}1 \quad P_{90} = 2083{,}3$

E. $P_{10} = 432{,}0 \quad P_{90} = 2115{,}0$

Följande diagram visar utvecklingen för ett tjänsteföretag från och med etableringsåret 2011 fram till och med året 2016. Diagrammet visar den kvartalsvisa försäljningen i mkr (miljoner kronor) av företagets tjänster.



b) Vilken av följande beskrivningar passar bäst in på tidsserien ovan? (4p)

A. Additiv modell; exponentiell trend, ingen säsong, stark cyklicitet/konjunktur

B. Additiv modell; linjär trend, ingen säsongseffekt, mycket slump

C. Multiplikativ modell; ingen trend, stark säsongseffekt, mycket slump

D. Multiplikativ modell; linjär trend, stark säsongseffekt, lite slump

E. Multiplikativ modell; exponentiell trend, stark säsongseffekt, lite slump

Uppgift 2

Ett företag säljer produkter via internetbutik och för statistik över tiden mellan utskick av beställningar och mottagande av betalning för varorna och man observerar ser att många betalar sina fakturor senare än sista betalningsdag. Fördelningen över förseningen tid från sista betalningsdag och mottagande av pengarna ges i tabellen nedan:

X = antal veckor försenad	-1	0	1	2	3
Andel/sannolikhet	0,15	0,48	0,24	0,08	0,05

NOT: Att det finns ett negativt värde, dvs. -1, betyder bara att vissa betalar upp till en vecka tidigare än utsatt sista betalningsdag.

a) Beräkna väntevärde och standardavvikelse för antalet veckors försening. (5p)

A. $\mu_X = 0{,}4$ $\sigma = 1{,}581$

B. $\mu_X = 1{,}0$ $\sigma = 1{,}581$

C. $\mu_X = 0{,}4$ $\sigma = 1$

D. $\mu_X = 1{,}0$ $\sigma = 1$

E. $\mu_X = 0{,}4$ $\sigma = 1{,}414$

Anta att vi har två (ospecificerade) händelser A och B sådana att $P(A) = 0{,}8$ och $P(B) = 0{,}3$ samt $P(B|A) = 0{,}125$.

b) Ange vilket av följande påståenden som är ett <u>falskt</u> påstående. (5p)

A. $P(A \cap B) = 0{,}1$

B. $P(A \cup B) = 1$

C. $P(\bar{A} \cap \bar{B}) > 0$

D. A och B är beroende

E. A och B spänner upp hela utfallsrummet S, dvs. $A \cup B = S$

TIPS! Rita Venndiagram eller en tabell för utfallsrummet och händelserna!

Uppgift 3

Ett biluthyrningsföretag med ett stort antal hyrbilar behöver emellanåt serva sina bilar. Anta att servicetiden för en slumpmässigt vald bil är en slumpvariabel som med god approximation antas vara normalfördelad med väntevärde $\mu = 240$ och standardavvikelse $\sigma = 60$ minuter.

a) Beräkna sannolikheten att en slumpmässigt vald bil som ska servas kräver mindre än 160 minuter servicetid. (5p)

A. 0,206

B. 0,092

C. 0,057

D. 0,063

E. 0,144

Sannolikheten att en bil behöver servas under en given månad har bedömts vara $p = 0,10$. Bilarna antas vara oberoende av varandra avseende servicebehovet. En företagskund hyrde under en viss månad $n = 8$ bilar.

b) Beräkna sannolikheten att minst två av dessa åtta bilar behövde servas under månaden. (5p)

A. 0,200

B. 0,038

C. 0,117

D. 0,187

E. 0,813

Total har företaget $n = 300$ bilar för uthyrning. Anta även nu att $p = 0,10$.

c) Använd approximeringstekniken som beskrivs i kurslitteraturen och som demonstrerats i samband med undervisningen för att approximativt beräkna sannolikheten att högst 36 bilar av företagets samtliga 300 bilar behöver servas under en viss månad. OBS! Att använda en inbyggd funktion i miniräknaren för att få ett exakt svar räknas inte som en approximeringsteknik. (5p)

A. 0,903

B. 0,875

C. 0,596

D. 0,755

E. 0,685

OBS! Svarsalternativen i a) – c) har avrundats till 3 decimaler.

Uppgift 4

Ett stort företag var tvunget att genomföra stora kostnadsbesparingar och man ville pröva om externa organisationskonsulter kunde höja effektiviten i företagets olika enheter. Man anlitade fyra konsulter som fick jobba med varsin slumpmässigt vald enhet och jämförde dessa mot fyra enheter som fick jobba utan konsulthjälp. Ett halvår senare bedömdes effektivitetsvinsten inom varje enhet i tkr/medarbetare och man har nu följande resultat:

Med konsult (X):	$\bar{x} = 78{,}0$	$s_x = 14{,}4$	$n_x = 4$
Utan konsult (Y):	$\bar{y} = 63{,}5$	$s_y = 10{,}2$	$n_y = 4$

Man vill nu testa om förväntad effektivitetsvinst med konsult (μ_x) är större än förväntad vinst utan konsult (μ_Y).

a) Vilket par av nollhypotes och mothypotes ska användas? (4p)

A. $H_0: \mu_X - \mu_Y \neq 0$ mot $H_1: \mu_X - \mu_Y = 0$

B. $H_0: \bar{x} - \bar{y} = 0$ mot $H_1: \bar{x} - \bar{y} \neq 0$

C. $H_0: \mu_X - \mu_Y = 0$ mot $H_1: \mu_X - \mu_Y < 0$

D. $H_0: \bar{x} - \bar{y} = 0$ mot $H_1: \bar{x} - \bar{y} > 0$

E. $H_0: \mu_X - \mu_Y = 0$ mot $H_1: \mu_X - \mu_Y > 0$

För att förenkla det hela antas att enheternas vinst är $X \sim N(\mu_X, \sigma_X^2)$ respektive $Y \sim N(\mu_Y, \sigma_Y^2)$ dvs. normalfördelade slumpvariabler. Dessutom att varianserna är lika i båda populationerna. Vi antar även oberoende mellan observationerna/enheterna.

b) Låt Obs beteckna observerat värde på testvariabeln. Vilket kriterium ska användas för att förkasta H_0 på 5 % signifikansnivå? (5p)

A. $Obs > 1{,}860$

B. $|Obs| > 2{,}447$

C. $Obs > 1{,}960$

D. $Obs > 1{,}943$

E. $Obs > 1{,}645$

c) Vad blir det observerade värdet på testvariabeln? (6p)

A. 1,643

B. 5,847

C. 0,279

D. 1,618

E. 3,333

Uppgift 5.

Vid en kvalitetsgranskning av en statistisk undersökning av företag upptäcktes att flera företags svar behövde dubbelkollas, dvs. man måste kontakta dessa för att se om deras svar verkligen stämmer. Ibland svarar man t.ex. i kr istället för som angivet i mkr vilket medför att svaren kan bli orimliga och därför måste följas upp. Kontrollerna belastar förstås också budgeten för undersökningen som genomförs varje år.

Man insåg också att vissa företag svarade med hjälp av en datorstödd insamlingsmetod medan andra använde en pappersblankett. Företagen väljer själva vilken metod de vill använda. Följande frekvenstabell sammanställdes från den senaste undersökningen:

Antal företag	Dubbelkollades	
	Nej	Ja
Datorstöd	125	40
Pappersblankett	20	15

Du får i uppdrag att göra ett χ^2-test för att se om det finns ett beroende mellan insamlingsmetod och nödvändigheten att dubbelkolla. Du använder signifikansnivån 5 %.

a) Beräkna utifrån datamaterialet ovan det observerade värdet för testvariabeln. (6p)

A. $\chi^2_{\text{obs}} = 0{,}373$

B. $\chi^2_{\text{obs}} = 5{,}991$

C. $\chi^2_{\text{obs}} = 5{,}018$

D. $\chi^2_{\text{obs}} = 4{,}324$

E. $\chi^2_{\text{obs}} = 2{,}132$

b) Ange korrekt slutsats. (4p)

A. H_0 förkastas, de är beroende

B. H_0 förkastas, de är oberoende

C. H_0 förkastas inte, de är beroende

D. H_0 förkastas inte, de är oberoende

E. Varken H_0 eller H_1 förkastas, det finns inget stöd för någon av hypoteserna

Fullständig redovisning krävs för följande uppgifter.

Använd separata pappersark för uppgift 6 resp. uppgift 7.

Uppgift 6

Spirometri är en lungfunktionsundersökning som utförs med spirometer som mäter hur lungorna fungerar. Forcerad vitalkapacitet (FVC) är den volym luft som en människa kan blåsa ut efter en maximal inandning. Från $n = 176$ patientjournaler beräknades FVC-medelvärdet för kvinnor i åldern 20-39 till $\bar{x} = 3,4$ liter och stickprovsvariansen till $s^2 = 0,64$.

a) Beräkna ett 95 % konfidensintervall för den genomsnittliga utandningsvolymen. Ange tydligt vilka antaganden som du gör, vilken formel du använder och tolka ditt resultat. (8p)

b) Förklara för någon som inte har läst statistik vad som menas med 95 % konfidens. Skriv kortfattat, max en ½ A4 behövs! (6p)

c) Bestäm stickprovsstorleken n om felmarginalen högst får vara $0,05$. (6p)

Uppgift 7

Ett finskt modeföretag hade precis introducerat sin senaste produkt, ett par strumpor. Försäljningsavdelningen anlitade en konsult som skattade en linjär modell som beskriver sambandet mellan antalet sålda enheter (Y) och pris i euro (X). Men tyvärr försvann konsulten och det enda som lämnades kvar av allt arbete var följande tabell som sammanställdes från $n = 8$ försäljningsställen och nu har du blivit ombedd att reda ut det hela:

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	$\hat{y}_i$
1	10	39	100	1521	390	42,0966
2	20	38	400	1444	760	39,4759
3	24	34	576	1156	816	38,4276
4	21	40	441	1600	840	39,2138
5	8	43	64	1849	344	42,6207
6	26	39	676	1521	1014	37,9034
7	15	44	225	1936	660	40,7862
8	20	43	400	1849	860	39,4759
Summa	144	320	2882	12876	5684	320

a) Formulera den linjära regressionsmodellen med pris (X) som förklaringsvariabel och antal sålda (Y) som beroende variabel samt skatta modellens parametrar med minstakvadrat-metoden. (6p)

b) Beräkna residualvariansen och förklaringsgrad för modellen och tolka/förklara resultatet för den senare. (6p)

c) Testa formellt om lutningskoefficienten är signifikant skild från noll på 5 % signifikansnivå. Ange dina hypoteser, testvariabel, beslutregel, beräkningar och din slutsats. (8p)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 19/04/18

Sal: Uggleviksalen

Tenta: Statistik för ekonomer

Kurs: Grundläggande statistik för ekonomer

ANONYMKOD: 0018-KBJ

☒ 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	X	X			6
Lär.ant. 10	10	15	15	10	18	20			

POÄNG	BETYG	Lärarens sign.
98	A	

STOCKHOLMS UNIVERSITET
Statistiska institutionen

VT2018 A-B

SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer 2018-04-19

Skrivsal: UG

Anonymkod: 0018-KBJ *(skriv tydligt!)*

Markera ditt svar med ett tydligt kryss (X) i rutorna nedan.
OBS! Endast ett kryss per uppgift. Har fler än ett svarsalternativ markerats ges noll poäng.

OBS! Om du efter att ha kontrollerat dina beräkningar ordentligt kommer fram till att svaret inte finns bland de angivna svarsalternativen, skriv ditt svar i mariginalen till höger.

		A	B	C	D	E
Uppgift 1	a)	X				
	b)				X	
Uppgift 2	a)			X		
	b)			X		
Uppgift 3	a)		X			
	b)				X	
	c)		X			
Uppgift 4	a)					X
	b)				X	
	c)	X				
Uppgift 5	a)			X		
	b)	X				

R

60/60

SU, STATISTIK

Skrivsal: ~~EG~~ UG Anonymkod: 0018-KBJ Blad nr: 1

6 a) Antagande: 176 oberoende observationer där varje kvinnas FVC antas (iid) dvs kommer från identiska oberoende fördelningar BRP

σ^2 antas okänt.

antar vidare att n är tillräckligt stort (för CGS) dvs $n \geq 30$ BBR

⇒ Från formelsamlingen med ~~en~~ med $\alpha = 0{,}05$

$\bar{x} \pm z_{\alpha/2} \cdot \frac{s_x}{\sqrt{n}}$; $z_{0,05/2} = z_{0,025} = [\text{TABELL } 2] = 1{,}96$

$\Rightarrow 3{,}4 \pm 1{,}96 \cdot \sqrt{\frac{0{,}64}{176}}$

Tolka!

$3{,}4 \pm 0{,}118$

⇒ ~~[3,2~~ $[3{,}282 \,;\, 3{,}518]$ R

6

b) Att det sanna värdet ligger i intervallet med en konfidens på 95%. Dvs vid många upprepade stickprov så kommer 95% av intervallen innehålla det sanna/riktiga värdet R

Mer allmänt har vi alltså 95% konfidens

~~Dvs risken för att att sant~~

Dvs risken för att en sann hypotes förkastas är 1 − konfidensen = α (signifikansen) ett sådant fel är av typ I.

~~Ett annan~~ Ett typ II fel är när en falsk hypotes accepteras

konfidens = 1 − α

6

SU, STATISTIK 0018-

Skrivsal: UG **Anonymkod:** KBJ **Blad nr:** 2

forts. 6

c) Om felmarginalen högst får vara 0,05 så får vi ekvationen

$\underbrace{z_{\alpha/2} \cdot \frac{s_x}{\sqrt{n}}}_{\text{felmarginalen}} \leq 0{,}05$ om antagande i a) gäller

$\Rightarrow \alpha = 0{,}05 \quad z_{0{,}025} = 1{,}96$ [från a) och Tabell 2]

$s_x = \sqrt{0{,}64}$ [från uppgift]

$\Rightarrow \sqrt{n} \geq \frac{1{,}96 \cdot \sqrt{0{,}64}}{0{,}05}$

$\Rightarrow n \geq \left(\frac{1{,}96}{0{,}05}\right)^2 0{,}64 \approx 983{,}4$

$\Rightarrow$ n som är heltal måste vara minst 984

$\underline{\underline{n \geq 984}}$ R

6

SU, STATISTIK

Skrivsal: UG Anonymkod: 0018-KBJ Blad nr: 3

7. a) Den linjära modellen ges av $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ R

där $\varepsilon_i \in N(0, \sigma_\varepsilon^2)$ (iid) med konstant varians

Detta ger oss den skattade modellen BRA

$\hat{y}_i = b_0 + b_1 x_i$ och med minsta kvadrat metoden

ges b_0 och b_1 via normalekvationerna

$\Rightarrow b_0 = \bar{y} - b_1 \bar{x}$ – $\left\{ \bar{x} = 144/8 = 18 \quad \bar{y} = 320/8 = 40 \right\}$

$b_1 = \dfrac{Cov(x,y)}{s_x^2}$

$Cov(x,y) = \dfrac{\sum_{i=1}^{8} x_i y_i - 8\bar{x}\bar{y}}{8-1} = \dfrac{5684 - 8 \cdot 40 \cdot 18}{7} = -\dfrac{76}{7}$

$s_x^2 = \dfrac{\sum_{i=1}^{8} x_i^2 - 8\bar{x}^2}{7} = \dfrac{2882 - 2592}{7} = \dfrac{290}{7}$

$\Rightarrow b_1 = \dfrac{-76}{290} \approx -0{,}2621 = -0{,}2621$ R

$\Rightarrow b_0 = 40 + 0{,}2621 \cdot 18 = 44{,}7178$ R

b) Residualvariansen $s_e^2 = \sum_{i=1}^{8} e_i^2 / n-2$

där $e = y_i - \hat{y}_i$ e^2 $(y_i - \hat{y}_i)^2$

$\Rightarrow$ i	e	e	$\approx e^2$
1	~~-3,0966~~	-3,0966	≈ 9,5889
2	~~-1,4759~~	-1,4759	≈ 2,1783
3	-~~4,4276~~	-4,4276	≈ 19,6036
4		0,7862	≈ 0,6181
5		0,3793	≈ 0,1439
6		1,0966	≈ 1,2025
7		3,2138	≈ 10,3285
8		3,5241	≈ 12,4193
$\sum_{i=1}^{8}$			≈ 56,0832

$\Rightarrow s_e^2 = \dfrac{56{,}0832}{8-2}$

$s_e^2 \approx 9{,}3472$ R

SU, STATISTIK

Skrivsal: UG Anonymkod: 0018-KBJ Blad nr: 4

forts 7 b)

$s_e^2 = 9,3472$

Förklaringsgrad $R^2 = 1 - \frac{SSE}{SST}$

$SSE = \sum_{i=1}^{8} e_i^2 = 56,0832$

$SST = SSR + SSE$

$SSR = \sum_{i=1}^{8} (\hat{y}_i - \bar{y})^2 \approx 19,9176$

$\bar{y} = 40$ $\hat{y}_i$ är givet ur uppgiften

i		$(\hat{y}_i - \bar{y})^2$
1	$\approx$	4,3957
2	$\approx$	0,2747
3	$\approx$	2,4724
4	$\approx$	0,6181
5	$\approx$	6,8681
6	$\approx$	4,3957
7	$\approx$	0,6181
8	$\approx$	0,2747
Σ	$\approx$	19,9176

$SST = 56,0832 + 19,9176$
$= 76,0008$

$\Rightarrow R^2 = 1 - \frac{56,0832}{76,0008} \approx 0,2621$

Det betyder att $\approx 26,21\%$ av variationen i den beroende variabeln y kan förklaras av variationen i den oberoende variabeln x BRA

6

SU, STATISTIK

Skrivsal: UG Anonymkod: 0018-KBJ Blad nr: 5

3 forts. 7 ($\not{e}$)

c) $H_0: b_1 = 0$

$H_1: b_1 \neq 0$ (tvåsidigt test) } Hypotes

Testvariabel R

~~$t_{obs} = \frac{\ldots}{\ldots}$~~ där b_1^* är den skattade lutningen i a)

$$t_{obs} = \frac{b_1 - 0}{s_{b_1}}$$ R {-0 ty H_0: $b_1 = 0$}

~~Har~~ $s_{b_1}^2 = \frac{s_e^2}{(n-1)s_x^2} = \frac{9,3472}{7 \cdot 290/7} = \frac{9,3472}{290}$

$\Rightarrow s_{b_1} \approx 0,1795$ $\{\sqrt{s_{b_1}^2} = s_{b_1}\}$

$\Rightarrow t_{obs} = \frac{-0,2621}{0,1795} \approx -1,4602$

Beslutsregel

Eftersom tvåsidigt test förkasta H_0 om $|t_{obs}| > t_{8-2,\,0,05/2}$ dvs $|t_{obs}| > t_{6,\,0,025}$ på signifikansnivån $\alpha = 0,05$ (5 %)

~~t kritisk~~ $t_{6,\,0,025}$ = [tabell 3] = 2,447 R

$\Rightarrow |t_{obs}| \not> t_{kritisk}$ dvs H_0 förkastas ej

och lutningen kan inte sägas vara signifikant skild från 0.

BRA

8