Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 23/11/17

Sal: Värtasalen

Tenta: Statistik för ekonomer

Kurs: Grundläggande statistik för ekonomer

ANONYMKOD: GSFE-FYP-UOD

☒ 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			4
Lär.ant. 10	10	15	15	10	14	19			

POÄNG	BETYG	Lärarens sign.
93	A	ME

HT2017

STOCKHOLMS UNIVERSITET
Statistiska institutionen

SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer
2017-11-23

Skrivsal: Värtasalen

Anonymkod: GSFE-PYP-UOD *(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)	☐	☐	☐	☒	☐
	b)	☐	☐	☐	☐	☒
Uppgift 2	a)	☐	☐	☒	☐	☐
	b)	☒	☐	☐	☐	☐
Uppgift 3	a)	☒	☐	☐	☐	☐
	b)	☐	☐	☐	☐	☒
	c)	☐	☐	☐	☒	☐
Uppgift 4	a)	☐	☒	☐	☐	☐
	b)	☐	☒	☐	☐	☐
	c)	☐	☐	☒	☐	☐
Uppgift 5	a)	☐	☐	☐	☒	☐
	b)	☒	☐	☐	☐	☐

R

60/60

6. a) $n = 80$ (storleken på urvalet)

$\hat{p} = \frac{50}{80} = 0,625$ där $\hat{p}$ är punkt-skattningen för proportionen p.

Vi antar att det är ett slumpmässigt stickprov med oberoende "arbetstagare" vars fördelning går mot normalfördelningen. OK
(varför?) 2

Vi får ett konfidensintervall:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ 2

Vi kollar att $nP(1-P) > 5$

$\rightarrow 80 \cdot 0,625(1-0,625)$

$\Rightarrow 18,75 > 5$; ja, n är tillräckligt stort.

$\alpha = 0,05$

$\alpha/2 = 0,025$

$z_{\alpha/2} = 1,96$

Konfidensintervallet blir då:

$$0,625 \pm 1,96 \cdot \sqrt{\frac{0,625(1-0,625)}{80}}$$

$\downarrow$

$0,625 \pm 0,106$ (avrundat till 3 decimaler) 2

Svar: Det 95% konfidensintervallet är $0,625 \pm 0,106$ (eller $(0,519 ; 0,731)$). 2

/8

6. b) från del a) har vi att $\hat{p} = 0{,}625$ samt att konfidensintervallet beräknas:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

felmarginalen

1.96 ska vara med

Den del du räknat på är standardfelet (st.avv.) för $\hat{p}$

Vi sätter in värdet på $\hat{p}$ och löser för (n).

detta ger att

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0{,}05$$

$$\downarrow$$

$$\sqrt{\frac{0{,}625(1-0{,}625)}{n}} \leq 0{,}05$$

vi måste avrunda uppåt då vi inte kan ha ett urvalsstorlek på 93,75 personer

$$n \geq \frac{0{,}625 \cdot 0{,}375}{0{,}05^2}$$

$$n \geq 93.75 \rightsquigarrow n \geq 94$$

Svar: Urvalsstorlehen, n, måste vara minst 94 för att felmarginalen ska vara $\leq 0{,}05$. 2 3

6c) Ett p-värde på 0,0127 antyder hur säkra vi kan vara på vårt resultat. P värdet jämför sedan med olika signifikansnivåer t.ex i fråga a, har vi ett 95% konfidensintervall $\rightarrow \alpha/2 = 0{,}025$. Eftersom 0,0127 är mindre än 0,025 så är vårt resultat signifikant.

Vi får ett signifikant resultat när p värdet $< \alpha$.

3

Kopplingen med a)-uppgiften

här enkelsidigt test, i a) ett dubbelsidigt intervall $\Leftrightarrow$ dubbelsidigt test

7a, Den skattade modellen ges av $\hat{y}_i = b_0 + b_1 x_i$

där $b_1 = \frac{cov(x,y)}{s_x^2}$ och $b_0 = \bar{y} - b_1\bar{x}$.

$\bar{x} = \frac{15}{5} = 3$ $\quad \bar{y} = \frac{15}{5} = 3$ $\quad s_x^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{55 - (5 \cdot 3^2)}{5-1} = 2,5$

$cov(x,y) = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{n-1} = \frac{54 - (5 \cdot 3 \cdot 3)}{4} = 2,25$

värdena är tagna från tabellen i uppgift 7.

så $b_1 = \frac{2,25}{2,5} = 0,9$ R och $b_0 = 3 - (0,9 \cdot 3) = 0,3$ R

$\longrightarrow \hat{y}_i = 0,3 + 0,9 x_i$ R

Residualvariansen ges av $s_e^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}$

i	y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	0,7	1,2	−0,5	0,25
2	2,8	2,1	0,7	0,49
3	2,5	3	−0,5	0,25
4	4,8	3,9	0,9	0,81
5	4,2	4,8	−0,6	0,36
Σ	15	15	0	2,16

så $s_e^2 = \frac{2,16}{5-1-1} = \frac{2,16}{3} = 0,72$ R

Svar: modellens parametrar är $b_1 = 0,9$ och $b_0 = 0,3$, residualvariansen är 0,72.

/8

7b) Signifikansnivå: $\alpha = 0{,}05$, testar om b_1 är skild från noll (signifikant)

$H_0: b_1 = 0 \qquad H_1: b_1 \neq 0$ R

Nollhypotesen förkastas om $|t_{obs}| > t_{crit}$ R

$t_{crit} = t_{n-k-1,\,\alpha/2} = t_{3;\,0{,}025} = 3{,}182$ R

$t_{obs} = t_{n-k-1} = \frac{b_1 - \beta_1^*}{s_{b_1}} = \frac{b_1}{s_{b_1}}$

$t_{obs} = \frac{0{,}9}{\sqrt{0{,}072}} = 3{,}354$ → avrundat till 3 decimaler R

→ $s_{b_1}^2 = \frac{s_e^2}{(n-1)s_x^2} = \frac{0{,}72}{(5-1)\cdot 2{,}5} = 0{,}072$

$s_{b_1} = \sqrt{s_{b_1}^2} = \sqrt{0{,}072} =$

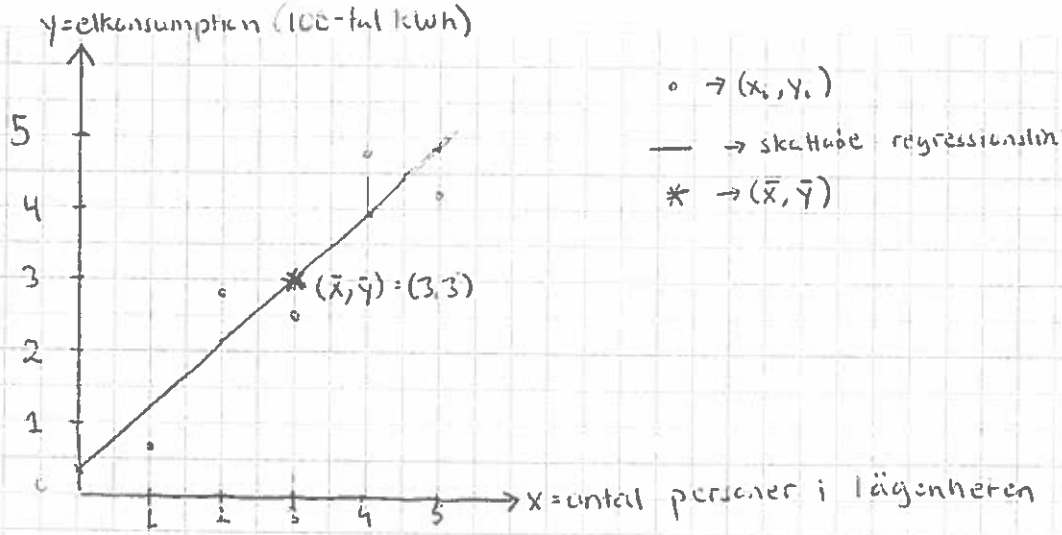
$|t_{obs}| = |3{,}354| > 3{,}182$

⇒ vi förkastar H_0 på signifikansnivån 5%.

Vi kan statistiskt säkerhetsställa att lutningen på vår regressionslinje, b_1, är signifikant skild från noll på en 5% signifikansnivå. D.v.s sambandet mellan X och Y är signifikant skilt från noll vid $\alpha = 5\%$.

vi

7c)



$\bar{x} = 3 \quad \bar{y} = 3 \rightarrow (3, 3)$ (värdena tagna från uppgift 7a.)

Punkten $(\bar{x}, \bar{y})$ ligger på regressionslinjen.
Punkten som består av medelvärdena $\bar{x}$ och $\bar{y}$ ligger alltid på linjen i en enkel linjär regression. Kan du förklara varför?

3

STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)
Per Gösta Andersson

HT2017

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2017-11-23

Skrivtid:	kl. 16.00 - 21.00
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

Nationella Viltolycksrådet publicerar statistik över viltolyckor i Sverige på sin webbsida, *www.viltolycka.se*, varifrån datamaterialet till denna uppgift har hämtats. Bland annat publiceras antal viltolyckor fördelat efter viltslag (dvs. vilken djurart som var inblandad), tiden för olyckan (månad och år) och var i landet olyckan ägde rum (läns- och kommunnivå).

I bilagan på sista sidan finns en tabell över antalet viltolyckor per kvartal (betecknat Kv1-Kv4) för tre djurarter (älg, vildsvin och dovhjort) för åren 2014-2016 för hela Sverige samt några olika diagram som åskådliggör olika egenskaper i datamaterialet.

a) Beräkna första och tredje kvartilen, Q_1 respektive Q_3, enligt metoden som beskrivs i kurslitteraturen, för antalet vildsvinsolyckor per kvartal över hela perioden 2014-2016. (5p)

A. $Q_1 = 624$ $Q_3 = 927$
B. $Q_1 = 532$ $Q_3 = 1580{,}25$
C. $Q_1 = 532$ $Q_3 = 1362{,}5$
D. $Q_1 = 578$ $Q_3 = 1580{,}25$
E. $Q_1 = 578$ $Q_3 = 1362{,}5$

b) Utgå ifrån egenskaperna för de ingående variablerna samt tabellen och diagrammen på sista sidan och ange vilket alternativ som inte är ett korrekt påstående. (5p)

A. Viltslag är en kategorisk variabel på nominalskala.
B. Antal olyckor över tid uppvisar en tydlig säsongsvariation för samtliga viltslag.
C. Antal vildsvinsolyckor per kvartal varierar mer under de tre åren jämfört med älg, men älg har i snitt ett större antal olyckor.
D. Antal olyckor med älg och antal med vildsvin är starkt positivt korrelerade.
E. Antal olyckor är en kontinuerlig numerisk variabel på kvotskala.

Uppgift 2

På ett försäkringsbolag analyserar man för tre olika försäkringsprodukter (betecknade A_1, A_2, och A_3) och dessas risker, dvs. sannolikheterna att en skada som täcks av respektive typ ska inträffa. Riskerna och fördelningen mellan försäkringstyperna anges i följande tabell:

	Försäkringsprodukt		
	A_1	A_2	A_3
Andel	20 %	30 %	50 %
Risk	15 %	10 %	4 %

a) Vad är sannolikheten att en skada som medför utbetalning ska inträffa för en slumpmässigt vald försäkring? (5p)

A. 0,067
B. 0,100
C. 0,080
D. 0,967
E. 0,111

Låt X och Y vara två slumpvariabler med samma utfallsrum $S_X = S_Y = \{-1,0,1\}$ och med simultana sannolikheter $P(x,y)$ (dvs. bivariata sannolikhetsfördelning) enligt följande tabell:

$P(x,y)$	$y =$		
	-1	0	1
$x = -1$	0	0,25	0
0	0,25	0	0,25
1	0	0,25	0

b) Beräkna kovariansen σ_{XY} och ange vilket av följande alternativ som är sant. (5p)

A. $\sigma_{XY} = 0$ X och Y är beroende
B. $\sigma_{XY} = 0{,}5$ X och Y är beroende
C. $\sigma_{XY} = 0$ X och Y är oberoende
D. $\sigma_{XY} = 0{,}5$ X och Y är oberoende
E. $\sigma_{XY} = -0{,}5$ X och Y är beroende

Uppgift 3

Man har under ett par säsonger observerat tiderna för två 100-meters löpare. Löpare A's tider tenderar att vara normalfördelade med väntevärde 9,94 sekunder och standardavvikelse 0,05. Löpare B är i snitt något snabbare med tider som också är normalfördelade med väntevärde 9,90 sekunder men en större variation med en standardavvikelse på 0,08. Anta att deras löptider är oberoende av varandra.

a) Beräkna och ange sannolikheten för händelsen att både A och B springer fortare än 9,92 sekunder. (5p)

   A. 0,206
   B. 0,392
   C. 0,057
   D. 0,263
   E. Det finns inte tillräckligt med information ovan för att beräkna sannolikheten.

Vi fokuserar nu på löpare A. Under ett träningsläger springer A totalt tjugo simulerade tävlingslopp. Dessutom antas att loppen är oberoende av varandra.

b) Beräkna sannolikheten att genomsnittstiden för de tjugo loppen är högst 9,95 sekunder. (5p)

   A. 0,726
   B. 0,500
   C. 0,117
   D. 0,977
   E. 0,813

Utgå ifrån att sannolikheten för händelsen att A ska springa ett enskilt lopp under 9,952 sekunder är 0,6 (approximativt, egentligen är sannolikheten ca 0,595 men vi avrundar).

c) Beräkna sannolikheten att fler än hälften, dvs. fler än 10 av de 20 loppen går under 9,952 sekunder. (5p)

   A. 0,723
   B. 0,872
   C. 0,596
   D. 0,755
   E. 0,813

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

Uppgift 4

Ett företag lanserade ett nytt utbildningspaket för företagets säljare, givetvis med syftet att förbättra deras prestationer. En sammanställning av sex säljares försäljningssiffror (i tkr) under tre månader före och tre månader efter utbildningen ges i tabellen nedan:

Säljare (i)	1	2	3	4	5	6
Före utbildningen (y_i)	197	215	282	203	327	165
Efter utbildningen (x_i)	205	238	300	201	347	188

Man vill få stöd för att utbildningen har haft en positiv effekt på försäljningssiffrorna och du blir ombedd att testa detta på 1 % signifikansnivå ($\alpha = 0{,}01$), dvs. att testa om den genomsnittliga förändringen i försäljningssiffror har ökat efter utbildningen jämfört med innan.

a) Ange det kritiska värdet för detta test. (4p)

A. kritiskt värde = 4,032
B. kritiskt värde = 3,365
C. kritiskt värde = 3,169
D. kritiskt värde = 2,764
E. kritiskt värde = 2,3263

b) Genomför beräkningarna för testet och ange den korrekta slutsatsen. (6p)

A. observerat värde = 3,674 och H_0 förkastas inte
B. observerat värde = 3,674 och H_0 förkastas
C. observerat värde = 3,354 och H_0 förkastas inte
D. observerat värde = 0,418 och H_0 förkastas
E. observerat värde = 0,418 och H_0 förkastas inte

c) Ange vilket av följande påståenden som inte är korrekt, dvs. falskt. (5p)

A. Stickprovsmedelvärdet $\bar{X}$ är en väntevärdesriktig skattning av väntevärdet μ.
B. Konfidensgraden är den förväntade andelen av alla konfidensintervall som fångar in det sanna parametervärdet vid upprepade stickprovsdragningar.
C. För att få hälften så stor felmarginal måste man ha dubbelt så stort stickprov.
D. Styrkan hos ett test beräknas som sannolikheten att förkasta H_0 när H_0 är falsk.
E. Centrala gränsvärdessatsen används när den underliggande populationen inte är känd eller när det är säkert att den inte är normalfördelad.

Uppgift 5

En universitetslärare vill se hur två olika undervisningsmetoder påverkar lärandet hos studenterna. För att utvärdera effekten av bytet bjuder hon in 70 frivilliga studenter och delar sedan slumpmässigt in dem i två grupper, de som får Metod A och de som får Metod B. Respektive grupp får delta på två lektioner var där motsvarande metod används och därefter genomförs ett skriftligt prov. Resultatet blev följande:

Observerat antal	Väl godkänt	Godkänt	Underkänt	Summa
Metod A	7	14	7	28
Metod B	21	14	7	42
Summa	28	28	14	70

Läraren vill nu testa om metod och resultat är oberoende eller inte. Under nollhypotesen beräknas de förväntade frekvenserna till:

Förväntat antal	Väl godkänt	Godkänt	Underkänt	Summa
Metod A	11,2	11,2	5,6	28
Metod B	16,8	16,8	8,4	42
Summa	28	28	14	70

a) Använd signifikansnivån $\alpha = 0{,}05$. Vilket av följande alternativ är en korrekt beslutsregel för detta test. (4p)

A. Om $\chi^2_{obs} > 3{,}841 \Rightarrow H_0$ förkastas

B. Om $\chi^2_{obs} \leq 90{,}531 \Rightarrow H_0$ förkastas inte

C. Om $\chi^2_{obs} > 14{,}067 \Rightarrow H_0$ förkastas

D. Om $\chi^2_{obs} > 5{,}991 \Rightarrow H_0$ förkastas

E. Inget av alternativen A-D är korrekt

b) Beräkna testvariabens observerade värde och ange korrekt alternativ. (6p)

A. $\chi^2_{obs} = 4{,}375$ metod och resultat är oberoende

B. $\chi^2_{obs} = 5{,}040$ metod och resultat är beroende

C. $\chi^2_{obs} = 4{,}375$ metod och resultat är beroende

D. $\chi^2_{obs} = 7{,}840$ metod och resultat är beroende

E. $\chi^2_{obs} = 7{,}840$ metod och resultat är oberoende

Fullständig redovisning krävs för följande uppgifter.

Använd separata pappersark för uppgift 6 resp. uppgift 7.

För uppgift 7c används Svarsbilagan

Uppgift 6

En myndighet ansvarig för arbetsmiljöfrågor genomför en enkät för att studera hur arbetstagare inom en viss bransch upplever arbetsmiljön. I ett urval om $n = 80$ arbetstagare som dragits från en större population angav 50 personer att de ansåg sig vara utsatta för buller.

a) Beräkna ett 95 % konfidensintervall för $P =$ andelen som är utsatta för buller på sin arbetsplats. Ange tydligt vilka antaganden som du utgår ifrån. (8p)

b) Hur stort urval skulle behövas om felmarginalen får vara högst 0,05? Utför exakta beräkningar och utgå från din punktskattning för andelen som du fick i a)-uppgiften. Om du inte har gjort uppgift a) kan du utgå från valfritt värde för andelen P men motivera i så fall ditt val. (6p)

c) Om man genomför ett hypotestest för att testa nollhypotesen $H_0: P = 0{,}5$ mot $H_1: P > 0{,}5$ skulle man med ovan givna data få ett p-värde lika med 0,0127. Vilken slutsats ska man dra av det resultatet och hur förhåller det sig till resultatet i a) ovan? Förklara kortfattat, max en ½ sida räcker. OBS! Det går att delvis besvara frågan utan att ha gjort a)-uppgiften. (6p)

Uppgift 7

En branschorganisation analyserar sambandet mellan antalet personer som bor i en lägenhet och elkonsumtionen för hushållsel. Man antar att sambandet kan modelleras enligt en linjär modell $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. Din kollega har sammanställt data och vissa delberäkningar för $n = 5$ observationspar i följande tabell där $X =$ antalet personer som bor i lägenheten och $Y =$ elkonsumtionen för hushållsel under en given månad (angivet i 100-tal kWh).

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	1	0,7	1	0,49	0,7
2	2	2,8	4	7,84	5,6
3	3	2,5	9	6,25	7,5
4	4	4,8	16	23,04	19,2
5	5	4,2	25	17,64	21,0
Summa	15	15	55	55,26	54,0

a) Skatta modellens parametrar och beräkna även residualvariansen. (8p)

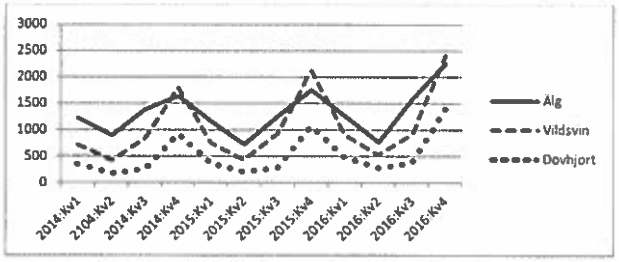
b) Är den erhållna lutningskoefficienten signifikant skild från noll? Testa på 5 % signifikansnivå. Tolka och förklara resultatet även i ord. (8p)

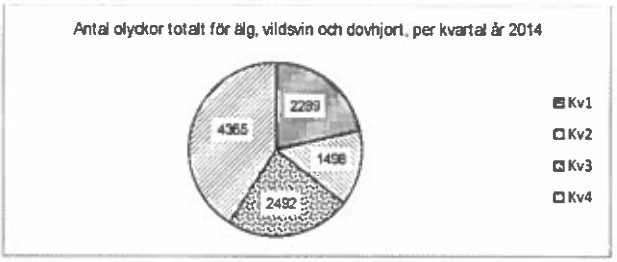
c) Åskådliggör X och Y i ett spridningsdiagram tillsammans med din skattade regressionslinje. Rita även in och markera tydligt punkten $(\bar{x}, \bar{y})$. Ligger den på eller utanför regressionslinjen? Är det alltid så att punkten ligger på eller utanför regressionslinjen eller kan det variera beroende på datamaterialet? (4p)

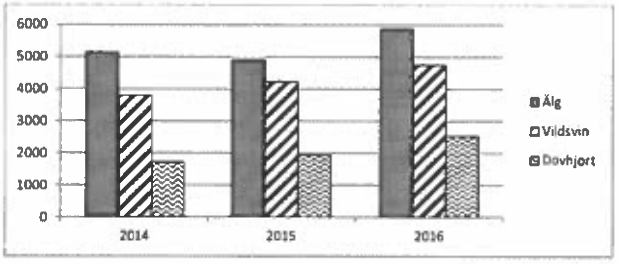
Bilaga till Uppgift 1.

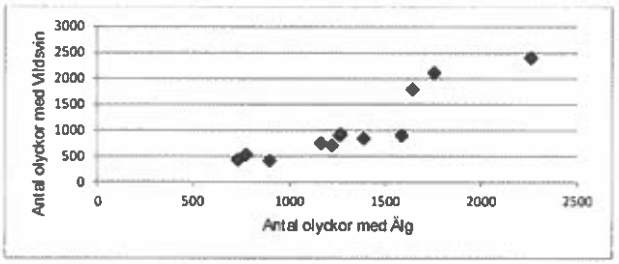
Uppgifterna har hämtats från Nationella Viltolycksrådet, www.viltolycka.se, (2017-11-12)

	År 2014				År 2015				År 2016			
	Kv1	Kv2	Kv3	Kv4	Kv1	Kv2	Kv3	Kv4	Kv1	Kv2	Kv3	Kv4
Älg	1221	894	1385	1641	1164	729	1265	1756	1258	772	1585	2259
Vildsvin	716	423	846	1798	755	432	927	2115	907	532	910	2408
Dovhjort	352	181	261	926	373	210	283	1074	483	288	375	1391









STOCKHOLMS UNIVERSITET
Statistiska institutionen
Michael Carlson (examinator)

HT2017

TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER

2017-11-23

LÖSNINGSFÖRSLAG

Slutversion, med reservation för tryck- och slarvfel / 2017-12-07 MC

Sammanfattning SVARSBILAGA Uppgifter 1-5
Utförliga beräkningar ges på efterföljande sidor

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

Uppgift 1

a) Rätt svar: **D**

Ordna de 12 kvartalssiffrorna för vildsvin i storleksordning, minsta till största:

1	2	3	4	5	6	7	8	9	10	11	12
423	432	532	716	755	846	907	910	927	1798	2115	2408

Första kvartilen: $0{,}25 \cdot (\square + 1) = 0{,}25 \cdot 13 = 3{,}25 \Rightarrow \square = 3$ och $\square = 0{,}25$

Använd 3:e och 4:e observationerna:

$\square = 3{:}\text{e} + 0{,}25 \cdot (4{:}\text{e} \square 3{:}\text{e}) = 532 + 0{,}25 \cdot (716 \square 532) = \square$

Tredje kvartilen: $0{,}75 \cdot (\square + 1) = 0{,}75 \cdot 13 = 9{,}75 \Rightarrow \square = 9$ och $\square = 0{,}75$

Använd 9:e och 10:e observationerna:

$\square = 9{:}\text{e} + 0{,}25 \cdot (10{:}\text{e} \square 9{:}\text{e}) = 927 + 0{,}75 \cdot (1798 \square 927) = \square$

b) Rätt svar: **E**

Antal olyckor är inte en kontinuerlig utan en diskret numerisk variabel på kvotskala.

Viltslag är en kategorisk variabel på nominalskala, kan ordnas på godtyckligt sätt. Övriga påståenden kan verifieras genom att studera tidseriediagrammet och spridningsdiagrammet.

Se kurslitteraturen och föreläsningsanteckningar.

Uppgift 2

a) Rätt svar: **C**

Beteckna händelsen att man drar en försäkring av typ med där 1, 2 eller 3. Sannolikheterna att man drar en av typ ges av deras respektive andelar dvs.

$P(\quad) = 0{,}2 \qquad P(\quad) = 0{,}3 \qquad P(\quad) = 0{,}5.$

Riskerna som ges är de betingade händelserna att en skada inträffar, här betecknat givet försäkringstyp, dvs.

$P(\quad \mid \quad) = 0{,}15 \qquad P(\quad \mid \quad) = 0{,}10 \qquad P(\quad \mid \quad) = 0{,}04$

Eftersom $\{\quad, \quad, \quad\}$ är en partition (de är parvis disjunkta och fyller upp hela utfallsrummet) beräknas enligt satsen om total sannolikhet:

$$\begin{aligned} P(\quad) &= P(\quad \cap \quad) + P(\quad \cap \quad) + P(\quad \cap \quad) \\ &= P(\quad \mid \quad) \cdot P(\quad) + P(\quad \mid \quad) \cdot P(\quad) + P(\quad \mid \quad) \cdot P(\quad) \\ &= 0{,}15 \cdot 0{,}2 + 0{,}10 \cdot 0{,}3 + 0{,}04 \cdot 0{,}5 = \end{aligned}$$

b) Rätt svar: **A**

Beräkna dubbelsumman:

$$\begin{aligned} \quad \quad E(\quad \cdot \quad) &= (\quad) \cdot (\quad) \cdot 0 + (\quad) \cdot 0 \cdot 0{,}25 + (\quad) \cdot 1 \cdot 0 \\ &\quad + 0 \cdot (\quad) \cdot 0{,}25 + 0 \cdot 0 \cdot 0 + 0 \cdot 1 \cdot 0{,}25 \\ &\quad + 1 \cdot (\quad) \cdot 0 + 1 \cdot 0 \cdot 0{,}25 + 1 \cdot 1 \cdot 0 = 0 \end{aligned}$$

Utöka tabellen med marginalfördelningarna för och och notera att de har samma sannolikhetsfördelning över samma utfallsrum och att de därför måste ha samma väntevärde:

$P(\quad, \quad)$		0	1	
	0	0,25	0	0,25
0	0,25	0	0,25	0,50
1	0	0,25	0	0,25
	0,25	0,50	0,25	1,00

$$\mu_{\quad} = \mu_{\quad} = E(\quad) = (\quad) \cdot 0{,}25 + 0 \cdot 0{,}50 + 1 \cdot 0{,}25 = 0$$

Kovariansen blir

$$\quad \quad E(\quad \cdot \quad) - \mu_{\quad} \cdot \mu_{\quad} = 0 \quad) \cdot 0 =$$

Att de är **beroende** inses bland annat genom att man ser att för varje kombination och gäller att

$$P(\quad, \quad) \neq P(\quad) \cdot P(\quad)$$

T.ex. har man att $P(0,0) = 0 \neq 0{,}25 = 0{,}5 \cdot 0{,}5 = P(0) \cdot P(0)$ och detta går igen för varje kombination av och Räcker att man hittar ett fall men här gäller det för alla fall.

Uppgift 3

a) Rätt svar: **A**

Beteckna A's tider med och B's tider med Det är given att och och att de är oberoende av varandra.

Sökt: [oberoende] = [standardisera] =

= [Tabell 1] = (1 0,65542) · 0,59871 = 0,206303 ≈

b) Rätt svar: **E**

Det är given att . Då följer att genomsnittet (stickprovsmedelvärdet) är

Sökt:

≈ [Tabell 1] = 0,81327 ≈

c) Rätt svar: **D**

Beteckna med antalet lopp av = 20 som går under 9,952 (lyckat utfall). Sannolikheten att gå under 9,952 i varje enskilt lopp är = 0,6. Tiderna för varje lopp är oberoende av varandra. Då är .

Sökt sannolikhet är men vi kan inte använda Tabell 7 utan att först definiera = 20 = antal lopp med tid över 9,952 (misslyckat utfall). Men är då per definition där 1 = 0,4 och den sökta sannolikheten kan skrivas

För att inse detta lista utfallen för och motsvarande utfall för och ringa in de gynnsamma utfallen:

	0	1	…	10	11	12	…	19	20
					↕				
20 −	20	19	…	10	9	8	…	1	0

Sökt: [enligt Tabell 7] = 0,75534 ≈

Uppgift 4

Notera att det är samma säljare som man undersöker, alltså är det ett *parvist beroende*.

Låt $D_i = Y_i - X_i$. Man får anta D_i är *iid* $N(\mu_D, \sigma_D^2)$. Beräkna de observerade differenserna d_i och sedan medelvärde och varians (standardavvikelse) för dessa:

Säljare (i)	1	2	3	4	5	6	Summa
Före utbildningen (x_i)	197	215	282	203	327	165	-
Efter utbildningen (y_i)	205	238	300	201	347	188	-
Differens ($d_i = y_i - x_i$)	8	23	18	-2	20	23	90
Differens i kvadrat (d_i^2)	64	529	324	4	400	529	1850

$$\bar{d} = 90/6 = 15 \qquad s^2 = \frac{1850 - 6 \cdot 15^2}{6 - 1} = 100 \qquad s = 10$$

Hypoteserna är $H_0: \mu_D = 0$ dvs. ingen förändring och som man vill motbevisa och $H_1: \mu_D > 0$ dvs. att det har skett en genomsnittlig ökning. Testvariabeln

$$T = \frac{\bar{D} - 0}{S/\sqrt{n}} \sim t\text{-fördelad med } n - 1 = 5 \text{ frihetsgrader}$$

a) Rätt svar: **B**

Kritisk gräns för det enkelsidiga testet är $t_{krit} = t_{5;0,05} =$ [illegible] om $\alpha = 0,01$ används

b) Rätt svar: **B**

Enkelsidigt test till höger (> 0) medför att man förkastar H_0 om $t_{obs} > t_{krit}$. Insättning ger

$$t_{obs} = \frac{15 - 0}{10/\sqrt{6}} = 3,67 > 3,365 \Rightarrow H_0 \text{ förkastas}$$

c) Rätt svar: **C**

A. SANT. Om $E(\hat{\theta}) = \mu$ så är $\hat{\theta}$ en väntevärdesriktig skattning av väntevärdet μ (förutsatt att väntevärdet existerar vilket inte alltid är fallet).

B. SANT. Detta är definitionen på konfidensgrad.

C. **FALSKT**. Hälften så stor felmarginal kräver fyra gånger så stort stickprov:

$$\lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{4n}} = \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \cdot \frac{1}{\sqrt{4}} = \lambda_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \cdot \frac{1}{2}$$

Felmarginal för $4n$ — Felmarginal för n

Om man skattar variansen och använder t-fördelningen krävs faktiskt något mindre än fyra gånger så stort stickprov. Om det inte är riktigt små stickprov på $n = 2$ eller 3, om dubblar man till $2n = 4$ respektive 6 kommer felmarginalen mer än halveras.

D. SANT. Detta är definitionen av styrka.

E. SANT. Det är just då CGS kommer till användning.

Uppgift 5

Oberoende test, ett χ^2-test. Hypoteserna som ställs mot varandra är

H_0: variablerna är oberoende mot H_1: variablerna är beroende

Man har $r = 2$ rader och $k = 3$ kolumner vilket ger att testvariabeln är χ^2-fördelad med $(r-1)(k-1) = 2$ frihetgrader. Testvariabeln är

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{k} \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \quad \text{där}$$

där f_{ij} är de observerade frekvenserna och $e_{ij} = \frac{f_{i\cdot} f_{\cdot j}}{n}$ är de förväntade frekvenserna under H_0. Testet är enkelsidigt till höger och man förkastar H_0 om det observerade värdet är tillräckligt stort.

a) Rätt svar: **D**

Beslutsregel är förkasta H_0 om $\chi^2_{obs} > \chi^2_{krit} = \chi^2_{2;0,05} =$ [enligt Tabell 4] $= 5,991$

b) Rätt svar: **A**

För varje cell beräknas $(f_{ij} - e_{ij})^2 / e_{ij}$ vilket ger:

$(f_{ij} - e_{ij})^2 / e_{ij}$	Väl godkänt	Godkänt	Underkänt	Summa
A	1,575	0,7	0,35	2,625
B	1,05	0,4667	0,2333	1,750
				4,375

Slutsats: $\chi^2_{obs} = 4,375 < \chi^2_{krit}$ och H_0 kan inte förkastas på 5 % signifikansnivå, dvs. det kan inte antas att de är beroende. De kan alltså antas vara **oberoende**.

Uppgift 6

a) Vi skattar p = andelen som upplever buller på arbetsplatsen med stickprovsandelen $\hat{p} = \sum x_i / n$ där x_i = 1 om person i upplever buller och 0 annars. Observationerna x_i är alltså Bernoulli-fördelade.

Antaganden: Observationerna är **oberoende** av varandra och **lika fördelade** (iid) dvs. sannolikheten att man drar en person som upplever buller är konstant lika med p. Vidare antar vi att n = 80 är tillräckligt stort för att utnyttja **CGS** för att approximera fördelningen för $\hat{p}$ med en **normalfördelning**

$$\hat{p} \to N\left(p, \sqrt{\frac{p(1-p)}{n}}\right), \quad n \to \infty$$

Formel: Ett approximativt 95 % konfidensintervall för p ges av:

$$\hat{p} \pm \lambda_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

där $\lambda_{\alpha/2}$ = [enligt Tabell 1] = 1,96

Insättning ger $\hat{p}$ = 50/ 80 = 0,625 och

$$0{,}625 \pm 1{,}96 \cdot \sqrt{\frac{0{,}625 \cdot 0{,}375}{80}} = \text{[illegible]} \text{ eller } (\text{[illegible]})$$

Slutsats/tolkning: Med 95 % konfidens (ej sannolikhet) kan man säga att andelen i populationen som upplever buller ligger i intervallet 51,9 – 73,1 %.

b) Felmarginalen ska vara mindre än 0,05 dvs.

$$1{,}96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < 0{,}05 \quad \Leftrightarrow \quad n > \left(\frac{1{,}96}{0{,}05}\right)^2 \hat{p}(1-\hat{p})$$

Utgå ifrån $\hat{p}$ = 0,625 (resultatet i a) ovan) ger att n > 360,15. Alltså sätt n = [illegible]

En konservativ strategi är att sätt $\hat{p}$ = 0,5 vilket ger det maximala värdet på felmarginalen för givet n så fall får man n > 384,16. Alltså sätt n = [illegible]

c) Om p-värdet är lägre än signifikansnivån t ska H_0 förkastas, och här gäller att det är lägre än t.ex. t = 0,05 men högre än t = 0,01.

Kopplingen till uppgift a) är inte självklart så att man kan titta efter om nollhypotesens värde p_0 ligger i intervallet eller inte. Detta eftersom konfidensintervallet är dubbelsidigt medan testet här i c) är *enkelsidigt*. Ett dubbelsidigt konfidensintervall är (mer eller mindre) ekvivalent med det *dubbelsidiga* testet $H_0: p = p_0$ mot $H_1: p \neq p_0$ (se t.ex. föreläsningsanteckningarna F13 s. 14-16). I detta fall mer eller mindre ekvivalent eftersom testvariabeln beräknas med det under H_0 antagna värdet p_0 = 0,5 men intervallet beräknas helt utifrån det skattade värdet på P dvs. $\hat{p}$

NOT: Med lite beräkningar kan man visa att beslutsregeln kan skrivas "förkasta H_0 om

$$\hat{p} > p_0 + \lambda_{\alpha} \sqrt{p_0(1-p_0)/n} = 0.5 + 1.6449 \cdot \sqrt{0.25/80} = 0.592"$$

Uppgift 7

a) Beräkna medelvärdena för och variansen för och kovariansen mellan och

$$= = \frac{\sum}{} = \frac{\sum}{} = \frac{15}{5} = 3 \qquad = \frac{\sum}{} = \frac{55 \cdot 3^2}{4} = 2{,}5$$

$$= \frac{\sum}{} = \frac{54{,}0 \cdot 3 \cdot 3}{4} = 2{,}25$$

Skattningen av lutningskoefficienten β_1: $= \frac{}{} = \frac{2{,}25}{2{,}5} =$

Skattningen av interceptet β_0: $= = 3 \; 0{,}9 \cdot 3 =$

Den skattade modellen: $= + = +$

Beräkna sedan $= +$ för samtliga observationer, sedan residualerna $=$

	1	2	3	4	5	Summa
	0,7	2,8	2,5	4,8	4,2	-
	1,2	2,1	3	3,9	4,8	-
	-0,5	0,7	-0,5	0,9	-0,6	0
	0,25	0,49	0,25	0,81	0,36	2,16

Residualvariansen: $= = \frac{}{} = [= 1] = \frac{\sum}{2} = \frac{2{,}16}{3} =$

b) Antaganden: efterfrågades inte men i korthet antar man att $\sim (0, \sigma^2)$, att de är iid och att de är oberoende av värdet på Vi konstaterar att antalet förklaringsvariabler är $= 1$ och stickprovsstorleken är $= 5$.

Hypoteser: $: \beta_1 = 0$ mot $: \beta_1 \neq 0$

Testvariabel: $_1 = = \frac{}{} \sim$ dvs. fördelad med $= 3$ f.g.

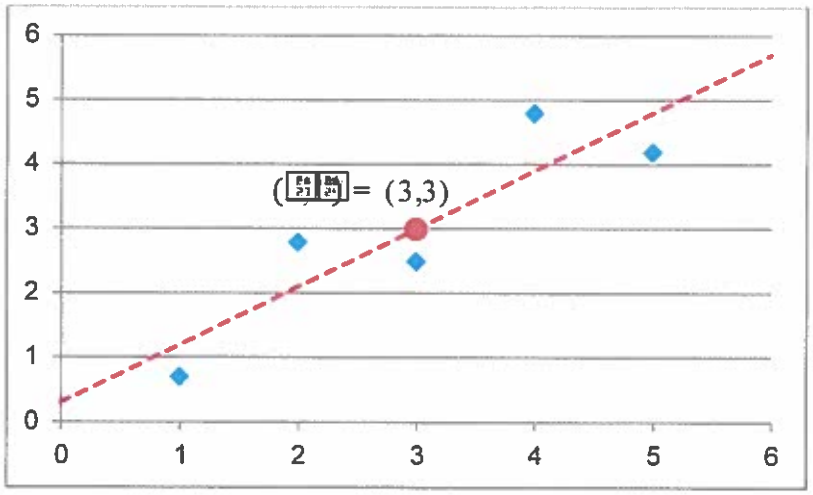
Beslutregel: Förkasta på nivån $\alpha = 0{,}05$ om $| | > _{0{,}025} =$ [Tabell 3] $=$

Beräkningar: $= \frac{}{( )} =$ [från a)] $= \frac{0{,}72}{4 \cdot 2{,}5} = 0{,}072$ $= \sqrt{0{,}072} = 0{,}26833$

$$| | = \frac{0{,}9}{0{,}26833} = > _{0{,}025}$$

Slutsats: **förkastas** på 5 % signifikansnivå, det observerade värdet $= 0{,}9$ är signifikant skiljt från noll. Antalet personer i lägenheten förklarar elkonsumtionen.

c) Spridningsdiagram med den skattade regressionslinjen:



Punkten (som i detta fall är lika med (3,3) ligger exakt på regressionslinjen.

I en linjär regressionsmodell kommer (alltid lika exakt på linjen oavsett vilket datamaterial man har. Detta följer av det faktum att linjen bestäms av normalekvationen = vilket är detsamma som + Parameterskattningarna är alltså konstruerade så att detta är uppfyllt.