

## TENTAMEN I GRUNDLÄGGANDE STATISTIK FÖR EKONOMER 2018-03-15

---

<b>Skrivtid:</b>	kl. 10.00 - 15.00
<b>Godkända hjälpmedel:</b>	Miniräknare utan lagrade formler och text
<b>Bifogade hjälpmedel:</b>	Häftet <i>Formelsamling och Tabeller över statistiska fördelningar</i> (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.

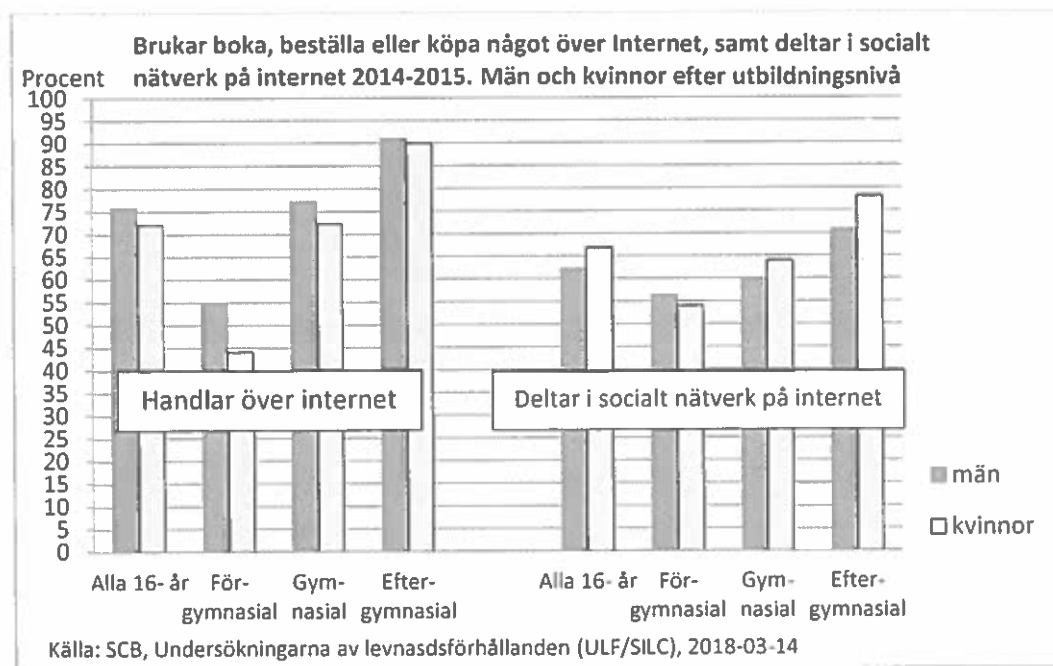
- Lösningförslag läggs ut på Mondo kort efter tentamen.

**LYCKA TILL!**

Endast svar ska anges för uppgifterna 1-5 med deluppgifter. Markera rätt svar på Svarsbilagan. OBS! Om du inte hittar rätt svar bland de givna svarsalternativen, skriv ditt svar och eventuell kommentar i marginalen på Svarsbilagan vid den aktuella frågan.

### Uppgift 1

Det blir allt vanligare att beställa varor, sköta bokningar med mera över internet och att vara med i något socialt nätverk. Följande diagram visar hur vi i Sverige använder internet för detta:



- a) Ange vilket av följande påståenden som inte är en korrekt tolkning av diagrammet. (5p)
- A. Ca 74 procent av befolkningen 16 år och äldre handlar över internet.
  - B. Att beställa varor på internet samvarierar starkt med utbildningsnivå.
  - C. Ca 54 procent av kvinnor med högst förgymnasial utbildning deltar i sociala nätverk.
  - D. En större andel män har en eftergymnasial utbildning jämfört med kvinnor.
  - E. Män handlar något mer på internet jämfört med kvinnor men kvinnor deltar i något större utsträckning i sociala nätverk.

- b) Anta att du har ett stickprov av storlek  $n = 100$  där man för en viss variabel observerar värdet noll (0) 30 gånger och värdet ett (1) 70 gånger. Beräkna och ange medelvärdet  $\bar{x}$  och varians  $s^2$  för variabeln. (5p)

- A.  $\bar{x} = 0,70$        $s^2 = 0,4606$
- B.  $\bar{x} = 0,70$        $s^2 = 0,2121$
- C.  $\bar{x} = 0,70$        $s^2 = 0,2100$
- D.  $\bar{x} = 70$           $s^2 = 49$
- E.  $\bar{x} = 70$           $s^2 = 21$

## Uppgift 2

Tullmyndigheten i ett visst land har tagit fram ett system för att profilera ankommande flygresenärer vid landets internationella flygplatser med avsikten att kunna upptäcka resenärer som för med sig mer alkoholhaltiga drycker än vad som är tillåtet.

Erfarenheten säger att 20 % av alla resenärer har med sig för mycket alkohol ( $A$ ). Efter en testperiod med den nya profileringsprocessen kommer man fram till att av alla resenärer som har med sig för mycket ( $A$ ) så är det 80 % som signaleras av systemet och får sitt bagage kontrollerat ( $B$ ). Vidare finner man att 20 % av de som inte har med sig för mycket ( $\bar{A}$ ) signaleras av systemet och får sitt bagage kontrollerat.

a) Hur stor är sannolikheten att en slumpvist vald resenär signaleras av systemet? (5p)

- A. 0,32
- B. 0,40
- C. 0,16
- D. 0,80
- E. 0,50

b) Givet att en resenär signaleras av systemet, vad är sannolikheten att denne har med sig för mycket alkohol? (5p)

- A. 0,32
- B. 0,40
- C. 0,20
- D. 0,80
- E. 0,50

Anta att vi har två (ospecificerade) händelser  $A$  och  $B$  sådana att  $P(A) > 0$  och  $P(B) > 0$ .

c) Ange vilket av följande påståenden som är ett sant påstående. (5p)

- A.  $P(A|B) = P(A)$  endast om  $A$  och  $B$  är beroende
- B.  $P(A \cap B) \geq P(A)$
- C. Om  $A$  och  $B$  är disjunkta så är de oberoende
- D. Om  $P(A) < P(B)$  så måste  $A \subset B$
- E. Alla ovanstående påståenden A-D är falska

TIPS! Rita Venndiagram!

### Uppgift 3

En inköpschef vid en stor livsmedelskoncern har sett till att man regelbundet kontrollerar leveranser av en storsäljande produkt. Leverantören har garanterat att den faktiska vikten för en förpackning är ungefär normalfördelad med genomsnittlig vikt  $\mu = 375$  gram så som det står på förpackningen men att det varierar mellan förpackningar; variansen sägs vara  $\sigma^2 = 9,3$  gram<sup>2</sup>. ← *821 24*

Man kontrollerar varje leverans genom att ta ett stickprov bestående  $n = 25$  enheter av produkten och mäta den genomsnittliga vikten i stickprovet. Inköpschefen underkänner varje leverans där den genomsnittliga vikten understiger 374 gram.

a) Vad är sannolikheten att inköpschefen underkänner en leverans? (5p)

- A. 0,295
- B. 0,050
- C. 0,008
- D. 0,950
- E. 0,038

Anta att inköpschefen överväger att beställa 20 leveranser av produkten. Man har bestämt att om 3 eller fler leveranser av dessa 20 underkänns så kommer man att byta leverantör.

Utgå ifrån att varje leverans är oberoende av varandra och att sannolikheten att en given leverans kommer att underkännas är konstant  $P = 0,10$ . OBS! Denna sannolikhet är inte svaret på a) uppgiften!

b) Beräkna och ange sannolikheten att man kommer att byta leverantör. (5p)

- A. 0,133
- B. 0,043
- C. 0,323
- D. 0,677
- E. 0,075

OBS! Svartalternativen i a) och b) har avrundats till 3 decimaler.

#### Uppgift 4

Karin äger tre andelar av investering A och en andel av investering B vid årets början. Enligt hennes modell kan värdet av investeringarna vid årets slut antas ha följande fördelningar:

$$\text{Varje andel av A: } X \sim N(105; 10^2)$$

$$\text{Varje andel av B: } Y \sim N(110; 20^2)$$

Karin bedömer att korrelationen mellan  $X$  och  $Y$  är  $\text{Corr}(X, Y) = 0,5$ .

- a) Under antagandet att Karins modell stämmer, beräkna och ange **variansen** för hela portföljen, 3 andelar av A och 1 andel av B. (5p)

- A. 500
- B. 700
- C. 1300
- D. 1600
- E. 1900

TIPS! Beräkna kovariansen först!

Med hjälp av informationen ovan kan man beräkna väntevärde och total standardavvikelse för hela portföljens värde. Beteckna väntevärdet för hela portföljen med  $\mu$  och den totala standardavvikelsen med  $\sigma$ .

- b) Vad är sannolikheten att portföljen är värd minst lika mycket som portföljens väntevärde minus två standardavvikelser vid årets slut, dvs. att det totala värdet på portföljen ska vara **större än  $\mu - 2\sigma$** ? (5p)

- A. 0,977
- B. 0,975
- C. 0,841
- D. 0,159
- E. 0,023

TIPS! Du behöver inte svaret från a) ovan för att svara på denna fråga!

OBS! Sannolikheterna i b) har avrundats till 3 decimaler.

### Uppgift 5

Ett marknadsföringsföretag genomför på uppdrag en enkätundersökning i en kommun kring ett planerat köpcentrum invid en genomfartsled. Av  $n_1 = 200$  tillfrågade svarade **100** att de var positiva till byggandet av det nya köpcentrumet. Utgå ifrån att svaren lämnades oberoende av varandra.

a) Vilket av följande alternativ är ett **99 %** konfidensintervall för  $P =$  andelen positiva? (5p)

- A. (0,442 ; 0,558)
- B. (0,431 ; 0,569)
- C. (0,418 ; 0,582)
- D. (0,409 ; 0,591)
- E. Inget av ovanstående

Efter den första undersökningen lanserade konsortiet för det planerade köpcentret en informationskampanj och ett halvår senare genomfördes en ny enkätundersökning, på samma sätt som den förra undersökningen men oberoende av den. Antalet positiva denna gång blev **140** av  $n_2 = 200$ . Du får i uppdrag att testa hypotesen att andelen positiva är oförändrad jämfört med förra undersökningen, mot att andelen positiva har ökat. Signifikansnivån är  $\alpha = 0,01$ .

b) Vilket av följande alternativ är en korrekt slutsats? (6p)

- A.  $z_{obs} = 4,082$   $H_0$  förkastas, ökningen är signifikant större än noll
- B.  $z_{obs} = 4,082$   $H_0$  förkastas inte, ökningen är inte signifikant större än noll
- C.  $z_{obs} = 5,000$   $H_0$  förkastas, ökningen är signifikant större än noll
- D.  $z_{obs} = 5,000$   $H_0$  förkastas, ökningen är inte signifikant större än noll
- E.  $z_{obs} = 6,172$   $H_0$  förkastas, ökningen är signifikant större än noll

Statistisk inferens avser metoder för att dra slutsatser från observationer, typiskt ett stickprov, och olika begrepp används i dessa sammanhang.

c) Vilket av följande alternativ är inte ett korrekt påstående? (4p)

- A. Konfidensgraden är den förväntade andelen intervall som innehåller det sanna populationsvärdet, om man drar många oberoende stickprov av samma storlek från samma population under identiska villkor.
- B. Om  $p$ -värdet är mindre än signifikansnivån förkastas nollhypotesen.
- C. Typ II fel inträffar när en sann nollhypotes förkastas.
- D. Ett "statistiskt säkerställt resultat" brukar betyda ett lågt  $p$ -värde för nollhypotesen.
- E. Signifikansnivån är lika med sannolikheten för Typ I fel.

Fullständig redovisning krävs för följande uppgifter.

Använd separata pappersark för uppgift 6 resp. uppgift 7.

### Uppgift 6

I en nyligen genomförd marknadsundersökning testades fem olika yoghurtsmaker för att avgöra om konsumenterna föredrar någon av sorterna. Varje person blev ombedd att ange sin favorit. Resultaten var som följer: av  $n = 80$  tillfrågade angav 12 smak A som sin favorit; B fick 16 röster, C fick 26 och D fick 9 och E fick resten av rösterna.

- Testa formellt på 5 % nivån om samtliga smaker är lika populära, dvs. att sannolikheterna att en slumpmässigt vald person väljer en viss smak är lika stor för de olika smakerna. Ange vilka antaganden som gäller för testet och kontrollera för eventuella tumregler. LEDNING: Det finns fem smaker. Hur ska du fördela sannolikheterna? (12p)
- Anta att man ska upprepa undersökningen i ett annat land och att man vill jämföra resultaten mellan länderna. Förklara hur man skulle testa detta formellt utan att genomföra några beräkningar (vi saknar ju data). Vilka hypoteser ställer man mot varandra? Vilken testvariabel används och hur är den fördelad? Kritiskt värde? Vad kallas testet? (8p)

### Uppgift 7

I en undersökning från mitten av 1970-talet drogs ett slumpmässigt urval av  $n = 10$  mindre geografiska områden i USA. Syftet var att studera sambandet mellan områdenas förväntade livslängd ( $Y = LifeExp$ ), genomsnittlig årsinkomst per capita ( $X_1 = Income$ ) och antalet mördade per 100 000 invånare och år ( $X_2 = MurderRate$ ). Tre linjära regressionsmodeller studerades och på nästa sida hittar du Excel-utskrifter för alla tre. Notera att  $Y$  är beroende variabel i alla tre:

$$\text{Modell 1: } Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \quad (Income)$$

$$\text{Modell 2: } Y_i = \beta_0 + \beta_2 X_{2i} + \varepsilon_i \quad (MurderRate)$$

$$\text{Modell 3: } Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \quad (Income \text{ och } MurderRate)$$

- Beräkna ett 95 % konfidensintervall för lutningskoefficienten i **Modell 1** och tolka resultatet. Kan man säga att lutningen är signifikant skild från noll? Förklara kortfattat! (6p)
- Beräkna ett 95 % prediktionsintervall för det betingade värdet på  $Y$  givet att  $x_{2,n+1} = 18$  i **Modell 2**. Använd  $\bar{x}_2 = 7,4$  och  $s_{x_2}^2 = 14,55$ . Kommentera eventuella problem med intervallet utifrån det faktum att det observerade maxvärdet på  $x_2$  var 13,9. (8p)
- Genomför ett hypotestest för att testa hela **Modell 3**. Ställ upp lämplig noll- och mothypotes och avgör med ledning av datorutskriften om du skulle förkasta nollhypotesen eller inte. Det räcker med hypoteserna och din slutsats och vad du baserar din slutsats på. Det ska räcka med 4-6 rader för att redovisa. LEDNING: Titta i formelsamlingen! (6p)

<b>Modell 1</b>	Multipel-R	0,554
	R-kvadrat	█
	Justerad R-kvadrat	█
	Standardfel	1,16026
	Observationer	10

ANOVA

	<i>fg</i>	<i>KvS (SS)</i>	<i>MKv (MS)</i>	<i>F</i>	<i>p-värde F</i>
Regression (R)	1	4,7616	█	█	█
Residual (E)	8	10,7697	█		
Totalt (T)	9	15,5313			

	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>
Konstant	63,67	3,619	█	█
Income	0,001573	0,0008362	█	█

<b>Modell 2</b>	Multipel-R	0,721
	R-kvadrat	█
	Justerad R-kvadrat	█
	Standardfel	0,96533
	Observationer	10

ANOVA

	<i>fg</i>	<i>KvS (SS)</i>	<i>MKv (MS)</i>	<i>F</i>	<i>p-värde F</i>
Regression (R)	1	8,0763	█	█	█
Residual (E)	8	7,4549	█		
Totalt (T)	9	15,5313			

	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>
Konstant	72,28	0,6949	█	█
Murder Rate	-0,2483	0,08435	█	█

<b>Modell 3</b>	Multipel-R	0,863
	R-kvadrat	█
	Justerad R-kvadrat	█
	Standardfel	0,75129
	Observationer	10

ANOVA

	<i>fg</i>	<i>KvS (SS)</i>	<i>MKv (MS)</i>	<i>F</i>	<i>p-värde F</i>
Regression (R)	2	11,5802	5,7901	10,258	0,0083
Residual (E)	7	3,9511	0,5644		
Totalt (T)	9	15,5313			

	<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>
Konstant	66,29	2,462	█	█
Income	0,001358	0,0005450	█	█
Murder Rate	-0,22966	0,06608	█	█





Stockholms  
universitet

Statistiska institutionen

## Rättningsblad

**Datum:** 15/03/18

**Sal:** Ugglevikssalen

**Tenta:** Statistik för ekonomer

**Kurs:** Grundläggande statistik för ekonomer

*Sic itur  
ad astra!*

**ANONYMKOD:**

0116-HMN

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			3
Lär.ant. 10	15	10	10	15	20	20			

*sc*

POÄNG	BETYG	Lärarens sign.
100	A	Me

**SVARSBILAGA till Tentamen i Grundläggande statistik för ekonomer  
2018-03-15**

Skrivsal: Ugglevikssalen

Anonymkod: 0116-HMN (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 marginalen till höger och kommentera.

		A	B	C	D	E	
Uppgift 1	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	}
	b)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Uppgift 2	a)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
	c)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
Uppgift 3	a)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	b)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Uppgift 4	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	
	b)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Uppgift 5	a)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
	b)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
	c)	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

60 av 60

6 Anpassningstest  $\mathcal{R}$

- a) Antaganden och förutsättningar:
- Oberoende och lika fördelade observationer (iid).
  - Slumpmässigt valda personer  $\mathcal{R}$
  - $n=80$
  - $K=5$

Hypoteser:  $H_0$ : Smakerna är lika populära.  $H_1$ : Smakerna är ~~inte~~ olika populära  $\mathcal{R}$

Testvariabel och fördelning:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \text{ som är } \chi^2\text{-fördelad med } (k-1) = (5-1) = 4 \text{ frihetsgrader. } \mathcal{?}$$

Där  $E_i = nP_i$   $\mathcal{R}$  ( $nP_i \geq 5$ )

Kritisk gräns och beslutsregel:

$$\chi^2_{krit} = \chi^2_{(k-1); \alpha} = \chi^2_{(5-1); \alpha} = \chi^2_{4; 0,05} = 9,488 \quad \alpha = 0,05$$

Förkasta  $H_0$  om  $\chi^2_{obs} > \chi^2_{krit} = 9,488$   $\mathcal{R}$

Beräkningar:

	$O_i$	$P_i$	$E_i$
A	12	0,2	16
B	16	0,2	16
C	26	0,2	16
D	9	0,2	16
E	17	0,2	16
Totalt	$n=80$	1,00	$n=80$

$P_i =$  samma för samtliga 5 grupper  $= \frac{1}{5} = 0,2$

$$\begin{aligned} \chi^2_{obs} &= \frac{(12-16)^2}{16} + \frac{(16-16)^2}{16} + \frac{(26-16)^2}{16} + \frac{(9-16)^2}{16} \\ &\quad + \frac{(17-16)^2}{16} = 1 + 0 + 6,25 + 3,0625 + 0,0625 = \\ &= 10,375 \quad \chi^2_{obs} = 10,375 \end{aligned}$$

Tumregel: Alla  $E_i \geq 5$ : Ok!  $\mathcal{R}$

Slutsats och tolkning:

$\chi^2_{obs} = 10,375 > \chi^2_{krit} = 9,488 \rightarrow H_0$  förkastas på 5% signifikansnivå. Det är statistiskt säkerställt på 5%-nivån att smakerna inte är lika populära.

☺

/12

(b) För att testa detta skulle ett homogenitetstest passabara, dvs. ett test där vi testar om fördelningen för de olika yoghurtsmakerna är lika i de två länderna.  $\mathbb{R}$

Vi skulle sätta upp följande hypoteser:

$H_0$ : Samma fördelning mellan föredragna yoghurt-smaker i båda länder.

$H_1$ : inte samma fördelning mellan föredragna yoghurt-smaker i länderna.  $\mathbb{BRA}$

Vår testvariabel skulle vara:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

som är  $\chi^2$ -fördelad med  $(R-1)(C-1)$  frihetsgrader, dvs  $R$ -rader i tabellen vi skulle få och  $C$ -kolumner i tabellen vi skulle få. Med 2 länder och 5 yoghurtsmaker får vi alltså  $(2-1) \cdot (5-1) = 4$  frihetsgrader.  $\mathbb{R}$

$$E_{ij} = \frac{R_i \cdot C_j}{n} \quad \mathbb{R}$$

• Tabellen skulle se ut såhär

		C=5					
		A	B	C	D	E	$\Sigma$
R=2	Land A						
	Land B						
	Totalt						

• Signifikansnivån skulle behöva bestämmas. En vanlig signifikansnivå är  $\alpha = 0,05$ .

• Kritiska gräns och beslutsregel:

Den kritiska gränsen beror på signifikansnivån och antalet frihetsgrader. Om de skulle vara  $\alpha = 0,05$  och f.g. = 4 får vi  $\chi^2_{krit} = 9,488$ .  $\mathbb{R}$

Vår beslutsregel är att förkasta  $H_0$  om  $\chi^2_{obs} > \chi^2_{krit}$ .

• Sedan räknar vi ut vårt observerade värde genom att först räkna ut alla förväntade värden genom:

$$E_{ij} = \frac{R_i \cdot C_j}{n} \quad \text{Därefter räknar vi ut } \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \text{ för}$$

Samtliga celler och summerar dessa. Detta blir vårt  $\chi^2_{obs}$ . Om det är större än  $\chi^2_{krit}$  förkastar vi  $H_0$  på den valda signifikansnivån. Det är då på denna nivå statistiskt säkert att fördelningen i de två länderna är olika. Om  $\chi^2_{obs} < \chi^2_{krit}$  kan vi inte förkasta  $H_0$  och vi får stå för acceptera  $H_0$ -hypotesen.  $\mathbb{R}$

# SU, STATISTIK

Skrivsal: Ugglevikssalen

Anonymkod: 0116-HMN

Blad nr: 2

7a) 95% KI för  $\beta_1$  ges av:  $b_1 \pm t_{n-k-1; \alpha/2} \times S_{b1}$   $\mathcal{R}$

$$b_1 = 0,001573$$

$$t_{n-k-1; \alpha/2} = t_{10-1-1; 0,05/2} = t_{8; 0,025} = 2,306 \mathcal{R}$$

$$S_{b1} = 0,0008362$$

$$\rightarrow 0,001573 \pm 2,306 \times 0,0008362$$

$$0,001573 \pm 0,0019282112$$

$$0,001573 \pm 0,001928 \text{ (avrundat)}$$

$$\rightarrow (-0,000355; 0,003501) \mathcal{R}$$

Tolkning: Det sanna värdet på  $\beta_1$  kommer med 95% konfidens att ligga i intervallet  $(-0,000355; 0,003501)$ . Om vi upprepar proceduren ett stort antal gånger kommer 95% av intervallen att täcka in det sanna värdet på  $\beta_1$ . **BRA!**

Dä intervallet innehåller 0 kan vi inte säga att  $\beta_1$  är signifikant skilda från noll. Vi skulle alltså inte kunna förkasta nollhypotesen  $H_0: \beta_1 = 0$  vid ett dubbelsidigt hypotestest med signifikansnivån 5%. **BRA!**

/6

8b) 95% prediktionsintervall ges av:

$$b_0 + b_1 X_{n+1} \pm t_{n-2; \alpha/2} \times \sqrt{s_e^2 \left( 1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{(n-1)S_x^2} \right)} \mathcal{R}$$

$$\text{Modell 2: } Y_i = \beta_0 + \beta_2 X_{2i} + \varepsilon_i$$

$$b_0 = 72,28$$

$$b_2 = -0,2483$$

$$t_{n-2; \alpha/2} = t_{10-2; 0,05/2} = t_{8; 0,025} = 2,306$$

$$n = 10$$

$$X_{n+1} = 18$$

$$\bar{X} = 7,4$$

$$S_e^2 = 14,55$$

$$s_e^2 = \frac{SSE}{n-k-1} = \frac{7,4549}{10-1-1} = \frac{7,4549}{8} = 0,9318625 \mathcal{R}$$

$$72,28 + (-0,2483) \times 18 \pm 2,306 \times \sqrt{0,9318625 \cdot \left( 1 + \frac{1}{10} + \frac{(18-7,4)^2}{(10-1) \cdot 14,55} \right)}$$

$$67,8106 \pm 3,11491151427$$

$$(64,6956884857; 70,9255115143) \approx (64,696; 70,926) \mathcal{R} \rightarrow$$

fons 7b).

Ett 95% prediktionsintervall för det betingade värdet på  $Y$  givet att  $X_{2,n+1} = 18$  är  $(64,696; 70,926)$ .

Då det observerade maxvärdet var 13,9 och detta intervall är för  $X_{2,n+1} = 18$  är det alltså en extrapolering, dvs. att vi drar slutsatser utifrån modellen för värden som ligger utanför de observerade värdena. **BRA!**

Detta ger för det första breda intervall (termen  $(X_{n+1} - \bar{X})^2$  i formeln kommer ju att bli stor), vilket gör att vi inte får särskilt mycket information från det, men det största problemet är att vi inte kan veta att det linjära sambandet som regressionsmodellen beskriver mellan  $X$  och  $Y$  även gäller utanför de observerade värdena! Framför allt i fall som detta, då vi undersöker samhällsvetenskapliga ämnen. Man ska alltså vara försiktig för extrapolering.

18 ☺

1c) Hypotestest för hela modellen = F-test.

$$H_0: \beta_1 = \beta_2 = 0$$

$H_1$ : minst en av  $\beta_1$  och  $\beta_2 \neq 0$ . **R**

\* F-värde för modell 3: 10,258

\* p-värde för F-test: 0,0083. **R**

• Slutats: p-värdet är den lägsta signifikansnivån vid vilken vi kan förkasta  $H_0$  med de observerade värdena. Vi kan alltså för detta test förkasta  $H_0$  på som lägst en signifikansnivå på 0,83%, vilket är en låg nivå.

Så länge signifikansnivån är  $> 0,0083$ , vilket är högst troligt, kommer vi alltså att förkasta  $H_0$  i detta test. Det innebär att det är statistiskt säkerställt att minst en av  $\beta_1$  och  $\beta_2$  är signifikant skilda från noll och modellen i helhet är alltså inte värdelös.

16

20