



TENTAMEN / EXAMINATION



12307683

Fylls i av **student** / To be completed by the **student**

| | | | |
|---|----------------|--|---------|
| Skriv anonymiseringskoden på samtliga svarsblad / Write your anonymity code on each sheet | | Anonymiseringskod / Anonymity code | |
| S T G A 0 1 | | 0 0 4 7 - E M Y | |
| Provbenämning / Exam name | | | Oanmäld |
| Skriftlig tentamen | | | |
| Kurskod / Course code | Modul / Module | Tentamensdatum / Examination date | |
| S T G A 0 1 | 1 0 0 2 | 2 0 2 1 - 0 1 - 1 5 | |
| Jag har tagit del av regler som gäller vid tentamen / I have read the current rules for examinations | | Antal inlämnade blad med anonymiseringskod / Number of sheets with anonymity code | |
| <input type="checkbox"/> Ja / Yes | | | |

Fylls i av **skrivvakt** / To be completed by the **invigilator**

| | | |
|--|-----------------------------------|---|
| Kontroll av legitimation / Identification checked | <input type="checkbox"/> Ja / Yes | Härmed intygas att kontroller utförts / This is to certify that the checks have been carried out |
| Kontroll av inlämnade blad / Answer sheets checked | <input type="checkbox"/> Ja / Yes | |
| Inlämningstid / Time of submission | | Tydlig sign. / Signature |

Fylls i av **lärare** / To be completed by the **examiner**

| Bedömning av uppgifter / Questions attempted | | | | | | | | | | |
|--|----|----|-----|----|--|----|----|----|----|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | ~ |
| 10 | 10 | 7 | 6,5 | 9 | 9,5 | | | | | |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | ~ |
| | | | | | | | | | | |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | ~ |
| | | | | | | | | | | |
| Totalt antal poäng / Total points | | | | | Examin. lärare / Kursansvarig signatur / Signature of the examiner | | | | | |
| 52 | | | | | A.W | | | | | |
| Betyg / Grade | | | | | Namnförtydligande / Clarification of the signature | | | | | |
| VG | | | | | | | | | | |

12307683



Försättsbladet ska alltid lämnas in även om ingen uppgift behandlats /
Examination should always be submitted even if no questions are answered

STGA01-0047-EMY

1

Uppgift 1

a) Det är frekvensen som talar om för oss antalet observationer, så för Datamaterial 1:

$$5 + 4 + 3 + 5 + 3 = 20$$

Svar: Datamaterial 1 har 20 st observationer

För Datamaterial 2 är antalet observationer i detta fall 4 antalet x_2 vilket i detta fall är 5 st.

Svar: Datamaterial 2 har också 20 st observationer

b) För att kunna matcha datamaterialen med en utskrift skall vissa beräkningar utföras

$$\bar{x}_{\text{datamaterial 1}} = \frac{1 \cdot 5 + 2 \cdot 4 + 3 \cdot 3 + 4 \cdot 5 + 5 \cdot 3}{20} = 2,85$$

$$M_d_{\text{datamaterial 1}} = \underbrace{11111}_{m_1} \underbrace{2222}_{m_2} \underbrace{333}_{m_3} \underbrace{4444}_{m_4} \underbrace{555}_{m_5}$$

$$S_{\text{datamaterial 1}} = \frac{((1-2,85)^2 \cdot 5) + ((2-2,85)^2 \cdot 4) + ((3-2,85)^2 \cdot 3) + ((4-2,85)^2 \cdot 5) + ((5-2,85)^2 \cdot 3)}{n-1} = 1,461$$

För Datamaterial 1 gäller $\bar{x} = 2,85$ $M_d = 3$ och $S = 1,461$ därför passar den ihop med Utskrift E

1,5

1

0,5

1,5
0,5

STGA01-0047-EMY

2

1b)

När det kommer till datamaterial 2 kan det vara lämpligt att beräkna regressionslinjen, och för att göra det är det behövt att sammanställa all data

| X | y | xy | X ² | y ² |
|---|---|----|----------------|----------------|
| 1 | 5 | 5 | 1 | 25 |
| 2 | 4 | 8 | 4 | 16 |
| 3 | 3 | 9 | 9 | 9 |
| 4 | 5 | 20 | 16 | 25 |
| 5 | 3 | 15 | 25 | 9 |

$$\sum x = 15 \quad \sum y = 20 \quad \sum xy = 57 \quad \sum x^2 = 55 \quad \sum y^2 = 84$$

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] [n \sum y^2 - (\sum y)^2]}}$$

$$\frac{5 \cdot 57 - (15 \cdot 20)}{\sqrt{(5 \cdot 55 - (15)^2) \cdot (5 \cdot 84 - (20)^2)}} = -0,474 \quad r = -0,474$$

$$b = \frac{5 \cdot 57 - (15 \cdot 20)}{5 \cdot 55 - (15)^2} = -0,3 \quad b = -0,3$$

Eftersom $\bar{x} = \frac{15}{5} = 3$ och $\bar{y} = \frac{20}{5} = 4$ så blir:

$$a = 4 - (-0,3 \cdot 3) = 4,9 \quad a = 4,9$$

Eftersom $r = -0,474$, $b = -0,3$, $a = 4,9$ så passar utskrift B bra ihop

0,5
0,5

1

1

0,5

+

STGA01-0047-EMY

3

Uppgift 1c)

För Datamaterial 1 har alla värden verifierats i A-uppgiften, För datamaterial 2 har alla värden förutom R square beräknats

$$R \text{ square} = r^2 = -0,474^2 \approx 0,225$$

0,5

Även R square matchar mellan datamaterial 2 och utskrift B. Allt är då verifierat

/ 10

STGA01-0047-EMY

4

Uppgift 2)

Jag har valt att göra en korstabell.
 Jag antar ett antal på 1000 personer och delar in i
 olika kategorier enligt de angivna procenttalen

För by 1:

| | B | B ^c | |
|----------------|-----|----------------|------|
| A | 280 | 420 | 700 |
| A ^c | 120 | 180 | 300 |
| | 400 | 600 | 1000 |

Enligt info skall $P(A|B) = P(A|B^c)$

$$P(A) = 0,7$$

b1 $P(B) = 0,4$

b2 $P(A) = 0,7$

c) $P(A \text{ och } B) = \frac{280}{1000} = 28\%$

Svar: Andelen barn som tycker om apelsiner i byn
 är 28%

a) För att händelserna A och B i by 1
 skall vara oberoende så skall:

$$P(A) \cdot P(B) = P(A \text{ och } B)$$

$$P(0,7) \cdot P(0,4) = P(0,28)$$

I by 1 är händelserna A och B oberoende

1,5
(se tidigare)

→

STGA01-0047-EMY

5

Uppgift 2)

Även här väljer jag att göra en korstabell och fyller i den information som angivits i antal

| | B | B ^c | |
|----------------|-----|----------------|------|
| A | 240 | 410 | 650 |
| A ^c | 60 | 290 | 350 |
| | 300 | 700 | 1000 |

d1) $P(B) = 0,3$

d2) $P(A|B) = 0,8$

d3) $P(A) = 0,65$

2a) För att övergå till frågan om händelserna A och B är oberoende i by 2, så gäller även här att $P(A) \cdot P(B) = P(A \text{ och } B)$
I det här fallet är $P(0,65) \cdot P(0,3) \neq P(0,24)$

I by 2 är alltså inte A och B oberoende, vilket det däremot är i By 1

e) $P(A \text{ och } B) = 0,24$. Genom att utgå från tabellen ovan gör det att se att andelen invånare i byn som är barn och tycker om apelsiner är 24%.

f) $P(B|A) = 0,37$. Genom att utgå från tabellen kan man utläsa att av de invånare som tycker om apelsiner är 37% barn

1,5

1

2

2

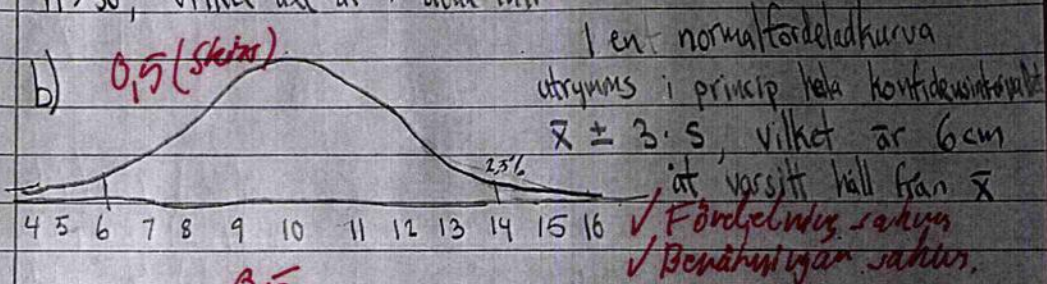
/ 10

STGA01-0047-EMY

6

Uppgift 3 (7)

a) För att man skall kunna använda sig om 95 fördelningen är ett antagande att det är normalfördelat. Enligt CBS kan man utgå från att samplingfördelningen för stickprovsmedelvärdet är lik en normalfördelning om $n > 30$, vilket det är i detta fall $n=10$



c) $\bar{x} \pm t \cdot se$ där $se = s/\sqrt{n}$ $df = n-1 = 9$
 $11,04 \pm 2,26 \times \left(\frac{1,93}{\sqrt{10}}\right) = 11,04 \pm 1,379322$
 Detta ger 0,5 konfidenstervall $[9,66; 12,42]$ 0,5

Detta betyder att med 95% finns väntevärdet för kottarnas längd mellan 9,66 och 12,42 cm. Även om n inte är högre än 30 i detta fall, går det ändå att utgå från resultatet eftersom kottars längd är normalfördelat

STGA01-0047-EMY

7

Uppgift 3d)

1. Antaganden.

- * Kvantitativ variabel
- * Urvalet har skett slumpmässigt **0,5**
- * Samplingfördelningen för stickprovsmedelvärdet är approximativt en normalfördelning (kottars längd är normalfördelat)

2. Hypoteser:

$$H_0: \mu = 10$$

$$H_a: \mu < 10$$

3. Teststatistika (vi har en situation där σ är okänd)

$$t = \frac{\bar{x} - \mu_0}{se} \quad \text{där } se = s/\sqrt{n} \quad \text{0,5}$$

$$t = \frac{8,69 - 10}{(1,82/\sqrt{10})} \approx -2,28 \quad \text{0,5}$$

4. P värde:

Då vi har ett ensidigt test, och $df=9$ i t-tabellen kan vi titta på den rad där det står "signifikansnivå vid ensidiga test och byta tecken på $df=9$ -raden till minustecken. Om vi fått $t=-2,26$ så hade P-värdet varit 0,025. Eftersom $-2,28$ är lägre än $-2,26$ så är P-värdet lägre än 0,025 **1**

5. Slutsats: Eftersom P-värdet $< \alpha$ (0,05) förkastas H_0 .
Data ger inte stöd för att påstå att kottars väntevärde är 10 cm på 5% signifikansnivå. **0,5**

Bön kottar "På 5% signifikansnivå så finns stöd för att μ är mindre än 10."

STGA01-0047-EMY

8

Uppgift 3 e)

Utifrån student 1 data, så skapades ett konfidenstervall mellan 9,66 - 12,42, vilket omfattar $\mu = 10$.

Därför leder resultatet från student 1:s data till ett sannolikhetsroget populationsmedelvärde **0,5**

För student 2 gäller följande konfidenstervall:

$$8,69 \pm 2,26 \times (1,82/\sqrt{n})$$

KI = (7,39 ; 9,99) vilket ej omfattar μ vilket är 10 cm. Därför stämmer den ej med verkligheten.

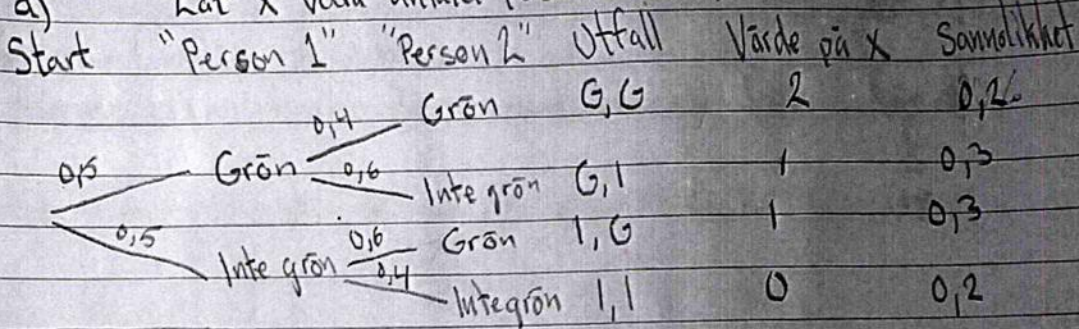
✓ Resultatet i 3d) skall användas.

STGA01-0047-EMY

9

Uppgift 4)

a) Låt x vara antalet personer i stickprovet som röstar grönt



| x | $P(x)$ |
|-----|--------|
| 0 | 0,2 |
| 1 | 0,6 |
| 2 | 0,2 |

$$\mu = \sum x P(x)$$

$$\mu = 0 \cdot 0,2 + 1 \cdot 0,6 + 2 \cdot 0,2 = 1$$

○ Väntevärdet "gröna" är 1 st vid $n=2$, vilket är en stickprovsandel på 0,5.

De möjliga populationsandelarna vid $n=2$ är 0, 0,5 eller 1 därför är sannolikheten att ha en max 7 procentenhet från populationsandelen 0,6

STGA01-0047-EMY

10

Uppgift 4b)

$X \sim \text{Bin}(n=10, p=0,45)$ $x =$ antalet gröna i stickprovet

Eftersom $n/N < 0,1$ kan man använda binomialfördelning.

Vi söker $0,38 \leq X \leq 0,52$, vilket är detsamma som $P(4) + P(5)$

$$P(4) = 10nCr4 \cdot 0,45^4 \cdot 0,55^6 = 0,238366$$

$$P(5) = 10nCr5 \cdot 0,45^5 \cdot 0,55^5 = 0,2340327$$

$$\text{Så } P(0,38 \leq X \leq 0,52) = 0,4723987$$

Svar: Det är till ungefär 47% säkerhet att stickprovsandelen gröna hamnar max 7 procentenheter från populationsandelen.

c) $n = 200$ $p = 0,42$ X anger antalet gröna

Eftersom $np > 15$ och $n(1-p) > 15$ så kan man använda sig av $X \sim N(np, \sqrt{np(1-p)})$ $\hat{p} \sim N(p, \sqrt{\frac{p(1-p)}{n}})$

$$X \sim N(84, 6,97997)$$

$$\text{Vi söker } 0,35 \leq X \leq 0,49, \text{ vilket är detsamma som}$$

$$P(70 \leq X \leq 98) = P\left(\frac{70-84}{6,97997} \leq Z \leq \frac{98-84}{6,97997}\right) = P(-1,0 \leq Z \leq 1,0)$$

$$= 0,8413 - 0,1587 = 0,6826$$

Svar Sannolikheten för att stickprovsandelen "gröna" hamnar max 7 procentenheter från populationsandelen är 68,26%.

€ 6,5

STGA01-0047-EMY

Uppgift 5) a1)

$$\hat{p} \pm z \cdot se$$

$$\hat{p} = \frac{0,14 + 0,32}{2} = 0,23$$

Svar: 23% av gurkorna i stickprovet var skadade

$$a2) 0,23 - 0,14 = 0,09 \quad M = 0,09$$

$$\frac{0,09}{1,96} = \frac{1,96 \cdot se}{1,96} \quad se = 0,045918$$

$$0,045918 = \sqrt{\frac{0,17711}{n}} \quad 0,045918^2 = \frac{0,17711}{n}$$

$$\frac{0,00210846n}{0,00210846} = \frac{0,17711}{0,00210846} = 84$$

Tips: Formel

$$n = \frac{\hat{p}(1-\hat{p})z^2}{m^2}$$

finns

Svar: Stickprovet var på 84 2

$$a3) 0,23 \cdot 84 = 19$$

Svar: 19 gurkor var skadade i stickprovet

STGA01-0047-EMY

12

Uppgift 5b

Låt μ_1 beteckna medelvikten bland Tores gurkor
och μ_2 beteckna medelvikten bland Agnetas gurkor

1. Antaganden:

- * Kvantitativ variabel
- * Urvalen har skett slumpmässigt
- * n är nära 30 i bägge fallen så normalfördelning kan approximeras enligt CBS

2. Hypoteser

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 - \mu_2 < 0$$

3. Teststatistika (vi har en situation där σ är okänd)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se}, \text{ där } se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (df = 2)$$

$$\frac{-14}{\sqrt{44,46 + 36,96}} = \frac{-14}{9,0233} \quad t = -1,55$$

$$df = 26 + 26 - 2 = 50$$

P-värde: Vi har ett vänstersidigt test och i t-tabellen (df=50) kan vi använda oss av ensidigt test med höger svans, men byta ut alla tecken till minus

$t = -1,55$ ligger mellan 0,1 (-1,3) och 0,05 (-1,68)

Slutsats: Eftersom P-värdet $> \alpha (0,05)$ förkastas inte H_0 . Det går alltså inte att påstå att Tores gurkor är mindre.

9

STGA01-0047-EMY

13

Uppgift 6

1 Antaganden

- * 2 kategoriska variabler (Värmland eller Örebro)
- * Slumpmässigt urval
- *

2 Hypoteser

- H_0 : Linjeval och lan är oberoende
- H_1 : Det finns en association mellan variablerna

3 Teststatistika:

$$\chi^2 = \sum \frac{(\text{observerat} - \text{förväntat})^2}{\text{förväntat}}$$

| | Värmland | Örebro | |
|-----------|----------|--------|-----|
| Linje 1 | 45 | 42 | 87 |
| Linje 2 | 100 | 141 | 241 |
| Linje 3 | 105 | 117 | 222 |
| förväntat | 250 | 300 | 550 |

| | Värmland | Örebro |
|---------|----------|--------|
| Linje 1 | 39,5 | 47,5 |
| Linje 2 | 109,5 | 131,5 |
| Linje 3 | 100,9 | 121,1 |

$$\chi^2 = \frac{(45 - 39,5)^2}{39,5} + \frac{(100 - 109,5)^2}{109,5} + \frac{(105 - 100,9)^2}{100,9} + \frac{(42 - 47,5)^2}{47,5} + \frac{(141 - 131,5)^2}{131,5} + \frac{(117 - 121,1)^2}{121,1} = 0,7658 + 0,8242 + 0,1666 + 0,63684 + 0,6863 + 0,1388 \approx 3,21854$$

$$(3 - 1) \cdot (2 - 1) = 2 \quad \text{vid } \chi^2\text{-tabell}$$

forts →

0,5
0,5

1

0,5

1

1

0,5

STGA01-0047-EMY

19

Uppgift 6 forts

4. P-värde.

Ur tabellen får vi att $P(\chi^2 > 3,219) = 0,2$.

Så vårt sökta värde ligger i princip på den.

Så $P(\chi^2 > 3,21854)$ blir vid 0,2

5. Slutsats.

Eftersom P-värdet $> \alpha(0,05)$ förkastas inte H_0

Det gör alltså med en signifikansnivå på 5% att påstå att linjeval och län är oberoende

1,5

/ 2 9,5