Antal blad / *Number of sheets* 10

TENTAMEN / EXAMINATION

Anvisningar: *Skriv din anonymitetskod på varje blad.*
Endast en uppgift får lösas på varje blad.
Var vänlig skriv tydligt!

Instructions: *Write your anonymous code on each sheet.*
Answer only one question on each sheet.
Please write clearly!

Vänligen texta anonymitetskoden i textboxen enligt exempel nedan!
Please write the Anonymous Code clearly in the textbox like example below!

Bokstäver/Letters:
A-B-C-D-E-F-G-H-I-J-K-L-M-N-O
P-Q-R-S-T-U-V-W-X-Y-Z-Å-Ä-Ö

Siffror/Numbers:
Ø-1-2-3-4-5-6-7-8-9

Exempel: A B C 1 7 Ø - Ø 1 7

STGAØ1
Kurskod + Kurs / *Course Code + Course:*

Statistik
Delkurs / *Part course:*

Anonymitetskod / *Anonymous code = Kurskod + kodnr / course code + code number*
S T G A Ø 1 - 1 1 7

Tentamensdatum / *Examination date:*
22/1-16

Behandlade uppgifter / *Solved problems*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	X	X	X	X	X									
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Ifylles av lärare / *To be completed by the examiner*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
7	10	9	7	4	10									
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Poäng / Marks gained: 47

Betyg / Grade: VG

Max poäng / Total marks gained:

För Gk poäng / Marks gainedto be passed:

A.W
Examin. lärare / Kursansvarig signatur / Signature of the examiner

Namnförtydligande / Clarification of the signature

3500000682

Häftområde

Skriv ej i detta område
Leave this area blank

Ange anonymitetskod / *Write your anonymity code* (Vid icke anonym tentamen ange kurskod + namn + personnummer) *(For non-anonymous exams write the course code + name + civic registration number)*	Löpande sidnr / *Consecutive no:*
STGA01-117	1

Uppgift nr / *Question no:* 1

Poäng / *Points awarded:* →

Lärarens anteckning *Examiner's remarks:*

1a, Vattenförbrukningen är (y) responsvariabeln, antal personer i ett hushåll är den förklarande variabeln (x)

b, Korrelationskoefficienten (r) beräknas:

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{(n\Sigma x^2-(\Sigma x)^2)\cdot(n\Sigma y^2-(\Sigma y)^2)}}$$

x	y	xy	x^2	y^2
2	59	118	4	3481
5	148	740	25	21904
1	52	52	1	2704
7	202	1414	49	40804
2	75	150	4	5625
9	140	1260	81	19600
3	104	312	9	10816
4	131	524	16	17161
Σ33	Σ911	Σ4570	Σ189	Σ122095

n = 8

$$r = \frac{8\cdot 4570 - 33\cdot 911}{\sqrt{(8\cdot 189 - 33^2)\cdot(8\cdot 122095 - 911^2)}} \approx$$

$$\approx \frac{36560 - 30063}{\sqrt{423 \cdot 146839}} \approx \frac{6497}{7881{,}17} \approx 0{,}824.$$

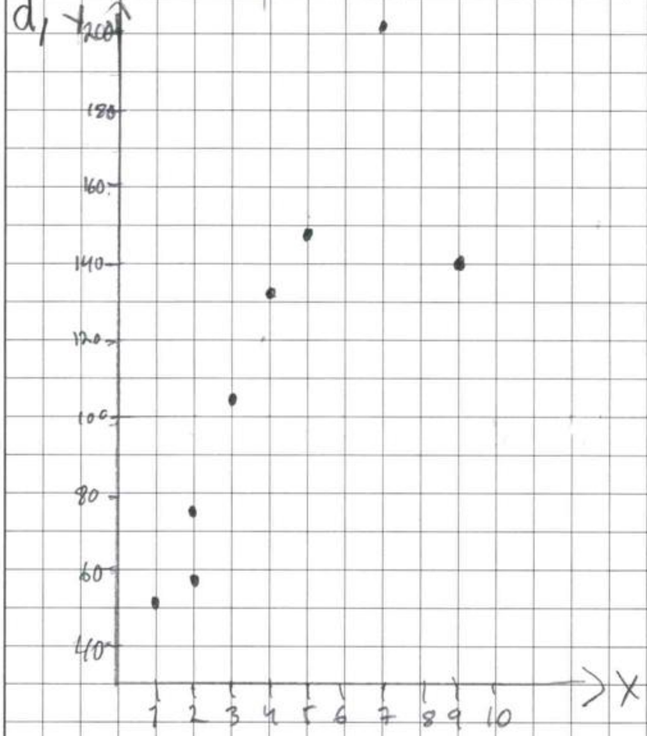
r = 0,824. Då r > 0,7 är det ett starkt linjärt samband mellan variablerna och det är värt att försöka anpassa en linje mellan observationerna.

c, $\hat{y} = 50{,}518 + 15{,}359x$

Att a är 50.518 innebär att det är årsförbrukningen av vatten när x, antal personer, är 0. Det är alltså alltid minst 50,518 m³ vatten som förbrukas per år i ett hushåll, även då ingen bor där.

Att b är 15,359 innebär att det är så mycket förbrukningen ökar för en till person i hushållet. Det är alltså hur mycket årsförbrukningen (y) påverkas av antalet personer i hushållet (x).

d,



Inte residual-plot

—

Universitetstryckeriet

ST6AØ1-117

Löpande sidnr / Consecutive no: 2

Uppgift nr / Question no: 1

Poäng / Points awarded: 7

d, forts. I diagrammet på föregående sida har jag ritat in alla observationer.
Då det handlar om 2 kvantitativa variabler valde jag ett spridningsdiagram.

För att få fram värdet på residualerna sätter jag in x-värdena i $\hat{y} = 50{,}518 + 15{,}359x$ för att få fram $\hat{y}$-värdena

√ $y - \hat{y}$ --

x	y	$\hat{y}$	$\hat{y} - y$
2	59	81,236	22,236
5	148	127,313	-20,687
1	52	65,877	13,877
7	202	158,031	-43,969
2	75	81,236	6,236
9	140	189,749	48,75
3	104	96,595	-7,405
4	131	111,954	-19,046

I höger stapeln ser man alltså residualernas värden

e, Genom att ta fram residualernas värde ser jag hur mycket de avviker från ekvationens linje.
Vissa residualer ligger rätt nära linjen, medan vissa ligger en bra bit bort från linjen.
Genom detta kan jag se att vissa hushåll använder mer resp. mindre vatten per år än vad ekvationens linje anger.

√ --

Behandla endast en uppgift per sida / Answer only one question per page

Häftområde

Skriv ej i detta område / Leave this area blank

KARLSTADS UNIVERSITET

Ange anonymitetskod / *Write your anonymity code*
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGA01- 117

Löpande sidnr / *Consecutive no:* 3

Uppgift nr / *Question no:* 2

Poäng / *Points awarded:* 10

Lärarens anteckning / *Examiner's remarks:*

2, H = betecknar händelsen att studenten har minst en förälder med högskoleutbildning.
K = betecknar händelsen att studenten känner krav från minst 1 förälder.

* slumpvis utvald

a, $P(H) = 0.48$: Sannolikheten att en* student har minst en förälder med högskoleutbildning är 48%. 0.5

$P(K^c \text{ och } H) = 0.40$: Sannolikheten att en slumpvis utvald student inte känner krav från minst 1 förälder samt att studenten har minst 1 förälder med högskoleutbildning är 40% 0.5

$P(K|H^c) = 0{,}10$: Sannolikheten att en student som inte har minst en förälder med högskoleutbildning, känner krav från minst en förälder är 10%. 0.5

För att svara på resterande frågor ritar jag en korstabell med en bestämd urvalsstorlek, jag väljer 1000.

	H	H^c	
K	80	52	132
K^c	400	468	868
	480	520	1000

Jag satte in alla siffror i tabellen med hjälp av informationen i uppgiften.

2,5

b, $P(K) = 132/1000 \approx 0{,}132$ 13,2% 1

c, $P(K|H) = 80/480 \approx 0{,}167$ 16,7% 1

d, $P(H|K) = 80/132 \approx 0{,}606$ 60,6% 1

e, $P(K \text{ och } H) = 80/1000 \approx 0{,}08$ 8% 1

f, $P(K \text{ eller } H) = (80+52+400)/1000 \approx 0{,}532$ 53,2% 1

g, För att se om de påverkar varandra, ska jag ta fram om de är oberoende.
$P(K \text{ och } H) = P(K)P(H)$, om de är oberoende.

$0{,}08 = 0{,}132 \cdot 0{,}48 \Rightarrow 0{,}08 \neq 0{,}06336$.
Variablerna är inte oberoende av varandra, föräldrarnas utbildning påverkar därför hur mycket krav studenterna känner.

1

Universitetstryckeriet

Behandla endast en uppgift per sida / *Answer only one question per page*

Häftområde

Ange anonymitetskod / *Write your anonymity code*
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGA01 - 117

Löpande sidnr
Consecutive no: 4

Uppgift nr / *Question no:* 3

Poäng / *Points awarded:* 9

Lärarens anteckning
Examiner's remarks:

3a, Det handlar här om 15 oberoende upprepningar av ett försök där sannolikheten att lyckas är konstant (inte exakt, men då villkoret $n/N<0,1$ $(15/200=0,075)$ är uppfyllt får man tänka så). har vi $X \sim Bin(15, 0,1)$.
X = antalet brädor som är defekta i ett urval av 15.

Då det är en binomialfördelning använder jag formen: $P(x) = (n\,nCr\,x) \cdot p^x \cdot (1-p)^{n-x}$

Sannolikheten för att inte mer än 1 bräda är defekt $P(x>1)$
$P(x>1) = P(0) + P(1)$

$P(0) = (15\,nCr\,0) \cdot 0,1^0 \cdot 0,9^{15} \approx 0,2059$

$P(1) = (15\,nCr\,1) \cdot 0,1^1 \cdot 0,9^{14} \approx 0,3432$

$P(x>1) = 0,2059 + 0,3432 = 0,5491$ 3

Svar: Sannolikheten att max 1 bräda är defekt är ca. 54,91%

b, Man kan här säga att det handlar om 800 oberoende upprepningar av ett försök där sannolikheten att lyckas är konstant (inte exakt, men villkoret $n/N<0,1$ är uppfyllt, så man får tänka så).
X = antalet defekta brädor i ett urval av 800 st
Vi har då $X \sim Bin(800, 0,1)$
Det finns en tumregel som säger att om parametrarna n och p i en binomialfördelning uppfyller villkoren $np > 15$ och $n(1-p) > 15$ så får man approximera en normalfördelning med parametrarna $\mu = np$ och $\sigma = \sqrt{np(1-p)}$. 0,5 0,5
$np = 80$, $n(1-p) = 720 \rightarrow$ De är uppfyllda
$\mu = 80$, $\sigma \approx 8,485$

$X \sim N(80, 8,485)$ 0,5
$P(X \le 90) = P\left(Z \le \frac{90-80}{8,485}\right) \approx P(Z \le 1,18) \approx 0,8810$

Svar: Sannolikheten att högst 90 brädor är defekta är ca. 88,10%. 0,5

Häftområde

Skriv ej i detta område / Leave this area blank

Ange anonymitetskod / *Write your anonymity code*
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGA01-117

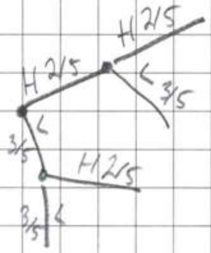
Löpande sidnr / *Consecutive no:* 5

Uppgift nr / *Question no:* 4

Poäng / *Points awarded:*

Lärarens anteckning / *Examiner's remarks:*

4a, 1,2,3 = lågt värde. 4,5 = högt värde
För att se sannolikhetsfördelningen för antalet gånger man får högt värde gör jag ett träd och en tabell:
X = antalet gånger man får högt värde i ett urval av 2. H = högt (2/5) L = lågt (3/5)



$HH = \frac{2}{5} \cdot \frac{2}{5} = 0{,}16$

$HL = \frac{2}{5} \cdot \frac{3}{5} = 0{,}24$

$LH = \frac{3}{5} \cdot \frac{2}{5} = 0{,}24$

$LL = \frac{3}{5} \cdot \frac{3}{5} = 0{,}36$

X	P(x)
0	0,36
1	0,48 (0,24+0,24)
2	0,16

I tabellen till vänster ser man sannolikhetsfördelningen för antal gånger man får högt värde.

/3

b,

Saknas

/0

Universitetstryckeriet

Behandla endast en uppgift per sida / *Answer only one question per page*

Häftområde

Skriv ej i detta område / Leave this area blank

KARLSTADS UNIVERSITET

Ange anonymitetskod / *Write your anonymity code*
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGAØ1-117

Löpande sidnr / *Consecutive no:* 6

Uppgift nr / *Question no:* 4

Poäng / *Points awarded:* 7

Lärarens anteckning / *Examiner's remarks:*

c, För att få fram sannolikhetsfördelningen för medelvärdet för de 2 omgångerna, gör jag en tabell.

Stickprovsstorlek => n = 2

Stickprovsutfall Omg. 1	Omg. 2	Stickprovs medelvärde	Sannolikhet för det aktuella utfallet
1	1	1	1/5 · 1/5 = 0,04
1	2	1,5	"
1	3	2	"
1	4	2,5	"
1	5	3	"
2	1	1,5	"
2	2	2	"
2	3	2,5	"
2	4	3	"
2	5	3,5	"
3	1	2	"
3	2	2,5	"
3	3	3	"
3	4	3,5	"
3	5	4	"
4	1	2,5	"
4	2	3	"
4	3	3,5	"
4	4	4	"
4	5	4,5	"
5	1	3	"
5	2	3,5	"
5	3	4	"
5	4	4,5	"
5	5	5	"

(lika stor sannolikhet för alla utfall)

Detta kan sammanfattas:

Stickprovsmedelvärde	total sannolikhet för det aktuella utfallet
1	0,04
1,5	0,08
2	0,12
2,5	0,16
3	0,2
3,5	0,16
4	0,12
4,5	0,08
5	0,04

I tabellen ovan har vi alltså sannolikhetsfördelningen för medelvärdet vid 2 spelomgångar.

/4

Universitetstryckeriet

Häftområde

Skriv ej i detta område / Leave this area blank

Ange anonymitetskod / *Write your anonymity code*
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGA01-117

Löpande sidnr / *Consecutive no:* 7

Uppgift nr / *Question no:* 5

Poäng / *Points awarded:* 4

Lärarens anteckning / *Examiner's remarks:*

5a, Det handlar här om 2 kvantitativa variabler där det ej är parvisa observationer. Jag ska därför använda mig av:

$(\bar{x}_1-\bar{x}_2) \pm t \cdot se$ ✓ $\quad se=\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ ✓ Situationen är "ganska"

$\bar{x}=\frac{\sum x}{n} \quad , \quad s=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$

$\bar{x}_{1\,(A)}=\frac{2{,}8+3{,}1+3{,}4+3{,}0+2{,}7+2{,}9+3{,}5+2{,}6}{8}=\frac{24}{8}=3.$

$\bar{x}_{2\,(B)}=\frac{3{,}2+3{,}1+2{,}9+3{,}5+2{,}4+3{,}0+3{,}2+2{,}8}{8}=\frac{24{,}1}{8}=3{,}0125$

s_1 (A):

x	$x-\bar{x}$	$(x-\bar{x})^2$
2,8	−0,2	0,04
3,1	0,1	0,01
3,4	0,4	0,16
3,0	0	0
2,7	−0,3	0,09
2,9	−0,1	0,01
3,5	0,5	0,25
2,6	−0,4	0,16
		Σ0,72

s_2 (B):

x	$x-\bar{x}$	$(x-\bar{x})^2$
3,2	0,1875	0,0352
3,1	0,0875	0,00766
2,9	−0,1125	0,01266
3,5	0,4875	0,2377
2,4	−0,6125	0,3752
3,0	−0,0125	0,000156
3,2	0,1875	0,0352
2,8	−0,2125	0,0452
		Σ0,748976

$s_1=\sqrt{\frac{0{,}72}{8-1}}\approx 0{,}3207$ 0,5 $\quad s_2=\sqrt{\frac{0{,}748976}{8-1}}\approx 0{,}3271$ 0,5

För att få fram t-värdet behöver jag veta df. $Df=n_1+n_2-2$ om n_1, n_2 samt s_1, s_2 är tillräckligt nära varandra.
Koll om de är nära varandra:
$n_1=n_2$
(s_1) $0{,}3207\cdot 1{,}5\approx 0{,}481$. s_2 ligger under detta nummer, så s_1, s_2 är tillräckligt nära varandra.
$df=8+8-2=14. \Rightarrow t=2{,}14.$ (95% KI)

$se=\sqrt{\frac{0{,}3207^2}{8}+\frac{0{,}3271^2}{8}}\approx\sqrt{0{,}012856+0{,}013374}\approx 0{,}161957$

Övre gräns: $(3-3{,}0125)+2{,}14\cdot 0{,}161957\approx 0{,}3341$

0,5 0,5

Nedre gräns: $(3-3{,}0125)-2{,}14\cdot 0{,}161957\approx -0{,}3591$

Svar: Skillnaden finns med 95% säkerhet i intervallet $[-0{,}3591, 0{,}3341]$ 0,5

Skriv ej i detta område / Leave this area blank

Häftområde

Ange anonymitetskod / Write your anonymity code
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGA01-117

Löpande sidnr / Consecutive no: 8

Uppgift nr / Question no: 5

Poäng / Points awarded:

Lärarens anteckning / Examiner's remarks:

5b. Antaganden:
- Kvantitativ variabel i 2 populationer.
- Urvalet har skett slumpmässigt.
- fördelningen för proven på avloppsvattnet var approximativt normalfördelade. ✓ Parvis

Det måste de vara, då $n_1 < 20$, $n_2 < 20$, fördelningen måste alltså vara normalfördelad, annars hade man inte kunnat lösa denna uppgift.

c. Nej, jag skulle inte dra slutsatsen att det finns någon 0,5 skillnad mellan metoderna. Då 0 finns med i intervallet, kan det betyda att det inte är någon skillnad. Jag kan med 95% säkerhet inte säga att det finns någon skillnad mellan metoderna.

Universitetstryckeriet

Behandla endast en uppgift per sida / Answer only one question per page

Häftområde

Skriv ej i detta område / Leave this area blank

Ange anonymitetskod / *Write your anonymity code*
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGA01-117

Löpande sidnr / *Consecutive no:* 9

Uppgift nr / *Question no:* 6

Poäng / *Points awarded:* 10

Lärarens anteckning / *Examiner's remarks:*

6a,

1. Antaganden:
- Kategorisk variabel, (röster på soc/ej). 1
- Urvalet har skett slumpmässigt
- Tillräckligt stort urval för att normalapproximation ska gälla, villkoren $np > 15$ och $n(1-p) > 15$ är uppfyllda
(200) (600)

2. Hypoteser:

$H_0: p = 0{,}25$ (p_0)

$H_a: p > 0{,}25$ 1

3. Teststatistika:

$\hat{p} = 225/800 \approx 0{,}28125$ $n = 800$

$Z = \frac{\hat{p} - p_0}{se_0}$ 0,5 $se_0 = \sqrt{\frac{p_0(1-p_0)}{n}}$

$se_0 = \sqrt{\frac{0{,}25 \cdot 0{,}75}{800}} \approx 0{,}015309$

$Z = \frac{0{,}28125 - 0{,}25}{0{,}015309} \approx 2{,}04$ 1

4. P-värde: 0,5 0,5

$P(Z > 2{,}04) \approx 1 - 0{,}97931 = 0{,}0207$. (enligt normalfördelnings-tabellen)

P-värdet är alltså: 0,0207

5. Slutsats:

0,5

Då P-värdet < α (0,05) förkastas H_0 på 5% signifikansnivå.

Lord: Med 5% signifikansnivå vågar jag påstå att andelen som röster på socialdemokraterna är högre än 25%.

0,5

Universitetstryckeriet

Häftområde

Skriv ej i detta område
Leave this area blank

KARLSTADS UNIVERSITET

Ange anonymitetskod / *Write your anonymity code*
(Vid icke anonym tentamen ange kurskod + namn + personnummer)
(For non-anonymous exams write the course code + name + civic registration number)

STGAØ1 - 117

Löpande sidnr
Consecutive no: 10

Uppgift nr / *Question no:* 6

Poäng / *Points awarded:*

Lärarens anteckning
Examiner's remarks:

6b, Att göra ett typ II-fel innebär att acceptera Ho, trots att den är falsk.
För att ta reda på sannolikheten att göra ett typ II-fel behöver jag räkna ut värdet på $\hat{p}$ där gränsen för att förkasta Ho går.
Gränsen för att förkasta Ho går då P-värdet $= \alpha$, dvs. det värde på test statistikan z då $P(Z > z) = 0.05$.
Detta inträffar då $z = 1,645$ (enligt N-fördelnings tab.).
För att hitta värdet på $\hat{p}$, behöver jag lösa:

$$\frac{\hat{p} - 0,25}{\sqrt{\frac{0,25 \cdot 0,75}{800}}} = 1,645 \implies \hat{p} \approx 0,27518$$

0,5

Man förkastar alltså Ho när $\hat{p} > 0,27518$ och accepterar Ho när $\hat{p} < 0,27518$. 0,5

Den verkliga normalfördelningen:

$$\hat{P} \sim N\left(0,29, \sqrt{\frac{0,29(1-0,29)}{800}}\right) \approx \hat{P} \sim N(0,29, 0,01604)$$ 0,5

Nu ska jag ta reda på sannolikheten att acceptera Ho, alltså: 0,5

$$P(\hat{P} < 0,27518) = P\left(Z < \frac{0,27518 - 0,29}{0,01604}\right) \approx P(Z < -0,92) \approx$$

$\approx 0,1788$ (enligt N-fördelningstab.). 0,5

Svar: Sannolikheten att göra ett typ II-fel är 17,88% 0,5

c, Sannolikheten att förkasta Ho om det sanna värdet är 29%, alltså att $\hat{p} > 0,27518$.

$$P(\hat{P} > 0,27518) = P\left(Z > \frac{0,27518 - 0,29}{0,01604}\right) \approx P(Z > -0,92) \approx$$

$\approx 1 - 0,1788 = 0,8212$ 1 (enligt N-fördelningstab.).
Svar: Sannolikheten att förkasta Ho, när Ho är sann är 82,12%.

0,5

d, Sannolikheten för att förkasta Ho när Ho är sann brukar kallas för att man gör ett "typ I-fel".

Universitetstryckeriet