Stockholms Universitet
Statistiska institutionen
Per Gösta Andersson

Statistikens grunder 2

SKRIFTLIG TENTAMEN
Måndagen den 25 november, 2019

Tillåtna hjälpmedel: miniräknare

Gräns för godkänt: 50 poäng av totalt 100.

För maximal poäng krävs på varje uppgift tydliga, utförliga och välmotiverade lösningar.

1. (18p) Sex bestämningar av μ gav följande resultat:

   0.4, 1.0, 0.8, 1.5, 1.3, 1.1.

   Antag att mätresultaten är oberoende och $N(\mu, \sigma^2)$.

   (a) Intervallskatta μ om standardavvikelsen är känd och lika med 0.4. Konfidensgrad: 95%.
   Hur många mätningar krävs för ett 99%-intervall av samma längd?

   (b) Beräkna ett konfidensintervall för μ under förutsättning att σ är okänd. Konfidensgrad: 95%.

2. (18p) Oberoende mätningar av mycket rent järn berett med två olika metoder, A och B, gav följande smältpunkter: (Celsius-grader)

   A: 1493, 1519, 1518, 1517, 1512, 1414, 1489, 1508, 1494

   B: 1509, 1494, 1512, 1483, 1507, 1491

   Formulera en statistisk modell baserad på normalfördelningar och samma varians.

   Kan vi påvisa skillnad i smältpunkt mellan beredningsmetoderna? Formulera och genomför lämpligt test på 10% signifikansnivå.

   Beräkna också ett 90% konfidensintervall för väntevärdesskillnaden mellan metoderna. Drar du samma slutsats från intervallet som från testet?

3. (16p) En speciell vitamin anses hämma uppkomsten av förkylningar. Betrakta ett försök under ett år med 200 personer som är slumpmässigt indelade i två lika stora grupper, där endast personer i ena gruppen fick vitamintabletter. Resultatet blev med angivna frekvenser av personer:

	Fler förkylningar	Ingen skillnad	Färre förkylningar
Kontrollgrupp	21	40	39
"Vitamingrupp"	20	29	51

Undersök med lämpligt test på approximativ signifikansnivå 5% huruvida vitaminintag har (positiv) inverkan på antalet förkylningar under ett år.

Drar du samma slutsats på 1%-nivå som på 5%-nivå?

4. (16p) Lystiden hos en viss typ av glödlampor följer en viss (okänd) fördelning. Tillverkaren påstår att medellystiden (väntevärdet för lystiden för en lampa) är 1000 timmar. Vi tvivlar dock på att den är så lång. Vi köper 50 lampor, tänder dem och väntar tills de går sönder. Vi noterar lystiden för var och en och får medelvärdet $\bar{x} = 953.2$ och standardavvikelsen $s = 910.5$.

Baserat på den beskrivna frågeställningen och det resultat vi fick, sätt upp lämplig noll- och mothypotes, samt utför ett test på approximativ signifikansnivå 5%.

5. (16p) Sannolikheten p att ett SJ-tåg är försenat vid ankomsten med mer än 5 minuter beror bla av vädret, antal resande och veckodag. Antag att man är intresserad av värdet på p en speciell dag. Man väljer därför slumpmässigt ur tidtabellen för alla SJ-tåg i Sverige n ankomster och undersöker hur mycket försenade just dessa ankomster är. Man vill pröva

$$H_0 : \ p = 0.3 \ \text{ mot } \ H_a : \ p > 0.3$$

på den approximativa signifikansnivån 5%.

(a) Antag att man har undersökt $n = 80$ ankomster och att vid 28 av dessa är tåget försenat med mer än 5 minuter. Kan H_0 förkastas?

(b) Vad är approximativt p-värdet för detta test?

6. (16p) Sant eller falskt? Motivering/kommentar nödvändig.

(a) Medelvärde och median är alltid båda lämpliga som lägesmått för kvalitativa data som är kodade till siffror.

(b) Medianen är medelvärdet av första och tredje kvartilen.

(c) En variationsvidd är alltid minst lika stor som en kvartilavvikelse.

(d) Värdet $\hat{F}(x_1)$ för en empirisk fördelningsfunktion är alltid minst lika stor som $\hat{F}(x_2)$, om $x_1 < x_2$.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 25/11-2019

Sal: Värtasalen

Tenta: Statistikens grunder

Kurs: Statistikens grunder 2

ANONYMKOD: 0047-GOB

☑ 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				~~8~~ 4
Lärant. 15	11	16	16	16	16				

MW

POÄNG	BETYG	Lärarens sign.
90	A	PG

SU, STATISTIK

Skrivsal: Värta Anonymkod: 0044-GOB Blad nr: 1

1, a, $\bar{x} = \frac{0,4 + 1 + 0,8 + 1,5 + 1,3 + 1,1}{6} \approx 1,0167$

n.f ! $\Rightarrow$ z

$\sigma = 0,4 \quad n = 6 \quad \alpha = 0,05$

$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$1,0167 \pm 1,96 \cdot \frac{0,4}{\sqrt{6}}$

$1,0167 \pm 0,32$

K.I. $[0,6967 ; 1,3367]$

99% Samma längd: $2,58 \cdot \frac{0,4}{\sqrt{n}} = 0,32$

$\left(\frac{0,4 \cdot 2,58}{0,32}\right)^2 = 10,4$

Svar: $n = 11$

b Eftersom vi vet att normalfördelning föreligger kan vi fortsatt använd $\underline{Z}$ som testvariabel.

$\bar{X} = 1,0167$ $\sum x = 6,1$ $\sum x^2 = 6,95$

$$S^2 = \frac{6,95 - \frac{6,1^2}{6}}{5} \approx 0,1497$$

$$S = \sqrt{0,1497} \approx 0,3869$$

$$X = Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

$$1,0167 \pm 1,96 \cdot \frac{0,3869}{\sqrt{6}}$$

$$1,0167 \pm 0,3096$$

$$[0,7071 ; 1,3263]$$

/15

SU, STATISTIK

Skrivsal: Värta Anonymkod: 0044-GOB Blad nr: 3-4)

2. $\alpha = 0,1$ $d.f = 13$

$H_0: \mu_A = \mu_B$

$H_A: \mu_A \neq \mu_B$

$\bar{X}_A = 1496$ $\bar{X}_B = 1499,33$

$\sum x_A = 13464$ $\sum x_B = 8996$

$\sum x_A^2 = 20150784$ $\sum x_B^2 = 13488680$

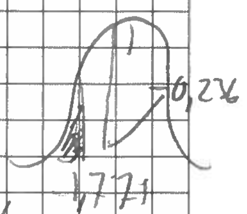
$$S_A^2 = \frac{20150784 - \frac{13464^2}{9}}{8} = 1080 \qquad S_B^2 = \frac{1348868 - \frac{8996^2}{6}}{5} = 135,4$$

$$S_p^2 = \frac{(9-1)\cdot 1080 + (6-1)\cdot 135,47}{9+6-2} = 716,72$$

$$T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{1496 - 1499,33}{\sqrt{716,72\cdot\left(\frac{1}{9} + \frac{1}{6}\right)}} = -0,236$$

Ho förkastas om Tobs > 1,771 eller Tobs < -1,771
(Tove (13))

Ho kan inte förkastas eftersom -0,236 ligger inom spannet.



Således kan vi inte påvisa skillnad på 10% sign nivå

K.1 Väntevärdesskillnad

$\bar{X}_A - \bar{X}_B = -3{,}33$

$-3{,}33 \pm 1{,}771 \frac{26{,}772}{\sqrt{15}} =$

$-3{,}33 \pm 12{,}242$

$[-15{,}572 ; 8{,}912]$

I 90% av intervallen beräknade genom samma metod ska det korrekta värdet = 0 ligga inom intervallet. om H_0 sann

Slutsatsen ändras inte eftersom ~~Tobs~~ 0 ligger inom intervallet.

///

SU, STATISTIK

Skrivsal: Värta Anonymkod: 0047-GOB Blad nr: ~~5~~ ~~6~~ 3

3. H_0: Ingen skillnad mellan grupperna

H_A: Skillnad mellan grupperna

$k = (3-1)(2-1) = 2$

n	Fler	Ingen	Färre	
Kontroll	21	40	39	100
Vitamin	20	29	51	100
	41	69	90	200

E(n)

Samtliga större än 5!

$$\frac{41 \cdot 100}{200} = 20{,}5 \quad \frac{69 \cdot 100}{200} = 34{,}5 \quad \frac{90 \cdot 100}{200} = 45$$

$$\frac{41 \cdot 100}{200} = 20{,}5 \quad \frac{69 \cdot 100}{200} = 34{,}5 \quad \frac{90 \cdot 100}{200} = 45$$

$$\chi^2 = \frac{(21-20{,}5)^2}{20{,}5} + \frac{(20-20{,}5)^2}{20{,}5} + \frac{(40-34{,}5)^2}{34{,}5}$$

$$+ \frac{(29-34{,}5)^2}{34{,}5} + \frac{(39-45)^2}{45} + \frac{(51-45)^2}{45}$$

$$= 3{,}378$$

H_0 förkastas om $\chi^2 > \chi^2_{0,05}(2) = 5{,}991$

H_0 förkastas inte på 5%-sig eftersom $3{,}378 < 5{,}991$

Samma slutsats dras på 1% eftersom $3{,}378 < 9{,}210$

Vi kan inte förkasta ett påstående att vitaminerna inte gör någon skillnad på varken 5% eller 1% sign. nivå.

1/1

4. $\mu = 1000$ $n = 50$ $\bar{x} = 953,2$ $s = 910,5$

$\alpha = 0,05$

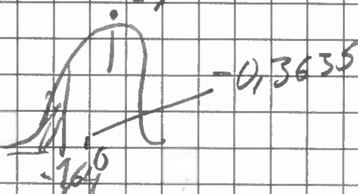
Ensidigt eftersom vi misstänker kortare lystid.

$H_0: \mu = 1000$

$H_1: \mu < 1000$

$n > 30$ därmed approx N.f enligt C.G.S $\Rightarrow Z$

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{953,2 - 1000}{910,5/\sqrt{50}} = -0,3635$$



Ho förkastas om $Z_{obs} < -1,64$

Ho kan inte förkastas eftersom $-0,3635 > -1,64$

Vi kan inte förkasta att medellystiden skulle vara 1000 timmar.

1/6

SU, STATISTIK

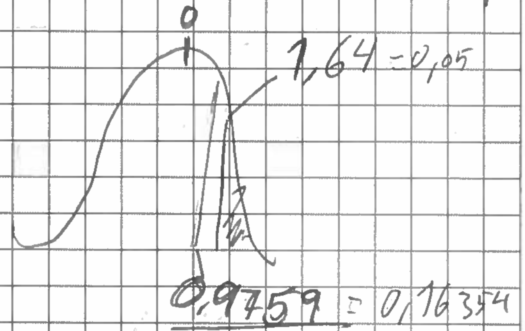
Skrivsal: Värta Anonymkod: 0047-GOB Blad nr: ~~7~~ ~~8~~ 4

5a $n = 80$ $p = 28/80 = 0{,}35$ $H_0: \pi_p = 0{,}3$

$H_1: \pi_p > 0{,}3$

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0{,}35 - 0{,}3}{\sqrt{\frac{0{,}3 \cdot 0{,}7}{80}}} = \underline{0{,}9759}$$

$\boxed{\begin{array}{l} n \cdot \pi_0 > 5 \\ n \cdot (1-\pi) > 5 \end{array}}$



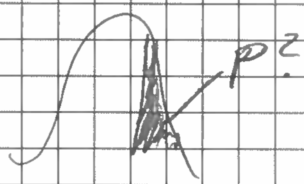
H_0 förkastas om $Z_{obs} > 1{,}64$

H_0 kan inte förkastas på 5% sign.
eftersom $0{,}9759 < 1{,}64$

b. $p(Z > 0{,}9759) = 1 - p(Z < 0{,}9759) \approx 1 - 0{,}83646$

$\approx 0{,}16354$

P-värde $\approx 0{,}16354$



/16

6.

a. Falskt. För kvalitativa data kodade som siffror redovisas fördelaktigtsvis frekvenser.

b. Falskt. medianen är det mittersta värdet i en storleksordnad talföljd.

c. Sant. Variationsvidden är minst en kvartilavvikelse eftersom $\sigma = \sqrt{\sigma^2}$

d. Falskt. $f(x_0) + f(x_1) = \hat{F}(x_1)$. $\hat{F}(x_1) < \hat{F}(x_2)$ förutsatt att $\hat{F}(x_2) > 0$.

/16