STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2019-04-25

Skrivtid: 15.00-20.00
Godkända hjälpmedel: Miniräknare, språklexikon.
Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.
Resultatet meddelas senast den 30 april.

OBS! Glöm inte att ange nödvändiga antaganden överallt

Uppgift 1. (20 poäng)

Ge ett kortfattat men uttömmande svar på följande frågor:
a) Vad är en momentestimator? (definiera begreppet)
b) Vad menas med konfidensgrad?
c) Vad säger Neyman-Pearsons lemma?
d) Vilka för- och nackdelar finns det med icke-parametriska metoder?

Uppgift 2. (20 poäng)

Antag att IQ hos en mycket stor grupp studenter är normalfördelad med det okända väntevärdet μ och den kända standardavvikelsen $\sigma = 15$. Man vill testa om μ överstiger 100 och testar därför hypotesen $H_0 : \mu = 100$ mot $H_a : \mu > 100$ på signifikansnivån 5 %.

a) Man väljer slumpmässigt ut $n = 4$ studenter ur gruppen. Ange kritiskt värde (alternativt RR, rejection region).

b) Bestäm testets styrka då $\mu = 115$.

c) Hur många observationer krävs för att testets styrka då $\mu = 115$ ska vara minst 0.95? Signifikansnivån ska fortfarande vara 5 %.

Uppgift 3. (20 poäng)

Två termometrar, A och B, jämförs genom att mäta utetemperaturen i grader C vid Södra husen, kl. 13.00 vid 12 olika dagar i april. Resultaten visas i nedanstående tabell.

Dag	A	B	C	D	E	F	G	H	I	J	K	L
Termometer A	6.8	9.8	2.1	6.2	7.1	6.5	9.3	1.0	-0.2	9.6	12.0	6.3
Termometer B	7.2	12.3	5.3	6.8	7.2	6.2	10.1	2.7	1.3	9.8	12.0	8.9

a) Använd teckentestet för att testa hypotesen om att det inte är någon systematisk skillnad mellan resultaten från de två termometrarna.

b) Testa samma hypotes som i deluppgift a) men med ett parametriskt test. Ange speciellt hur nödvändiga antaganden skiljer sig mellan testen i uppgift a) och i uppgift b).

Uppgift 4. (20 poäng)

Tiden det tar för en slumpmässigt vald kund att handla i Monas livsmedelsbutik antas vara en exponentialfördelad stokastisk variabel Y med väntevärde λ. Antag att tiden det tar att handla observeras hos $n = 64$ slumpmässigt utvalda kunder.

a) Bestäm maximum likelihoodestimatorn, $\widehat{\lambda}_{ML}$ av λ.

b) Är $\widehat{\lambda}_{ML}$ en väntevärdesriktig estimator av λ?

c) Bestäm variansen hos maximum likelihoodestimatorn $\widehat{\lambda}_{ML}$.

d) Är $\widehat{\lambda}_{ML}$ en konsistent estimator av λ?

e) Bestäm maximum likelihoodestimatorn av variansen för $\widehat{\lambda}_{ML}$.

Uppgift 5. (20 poäng)

Antag att antal gånger en slumpmässigt vald student behöver göra tentamen i STMT för att få ett godkänt betyg är en geometriskt fördelad stokastisk variabel, Y, med väntevärde $1/p$ (sannolikheten att klara tentamen vid ett enskilt tillfälle är p). Med hjälp av oberoende observationer på n slumpmässigt utvalda studenter, $y_1, y_2, \ldots, y_n$, vill man nu använda likelihood kvottestet för att testa hypotesen $H_0 : p = 0.8$ mot $H_a : p \neq 0.8$.

a) Bestäm teststatistikan för likelihoodkvottestet, λ_{LR}.
Ledning: Vi vet att maximum likelihood estimatorn av p är $\widehat{p}_{ML} = 1/\overline{y}$, där $\overline{y} = \sum y_i/n$.

b) För stora urval kan en transformation av λ_{LR} approximeras med en känd fördelning. Ange den transformerade teststatistikan, dess approximativa samplingfördelning och kritiskt värde för signifikansnivån 0.01.

c) I ett slumpmässigt urval om 40 studenter blev det totala antalet genomförda tentamina lika med 52. Genomför testet.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 25/4-2019

Sal: Uggleviksssalen

Tenta: Statistisk teori med tillämpningar II

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: 0032-MMP

☒ 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					6
Lär.ant. 1	16	12	16	7					

POÄNG	BETYG	Lärarens sign.
52+0=52	E	

SU, STATISTIK

Skrivsal: Uggleviksssalen **Anonymkod:** 0032-MMP **Blad nr:** 1

Fråga 1

a) En momentestimator är ett verktyg för att estimera en variabels väntevärde. Hur? Förklara!

b) Under hypotesprövning brukar man säga att man kan anta ett visst värde eller intervall till en viss konfidensgrad. Om vi t.ex. gör ett 95%-igt konfidensintervall kan vi säga att det sanna värdet ligger i det intervallet till 95% konfidens. Detta är dock inte densamma som sannolikhetsgrad. ?

c) Neyman-Pearsons lemma används under test för likelihood för att hitta det kritiska värdet mellan två variabler.

$$\frac{L(\Omega_0)}{L(\Omega_a)} < K$$

1

Den säger att genom att använda formeln kan vi hitta det kritiska värdet av likelihood mellan två variabler. Vad betyder det?

d) När vi använder icke-parametriska metoder behöver vi, till skillnad från parametriska metoder, inte anta beroende mellan populationerna. Inte heller behövs det göra några antaganden om normalfördelning. ? Detta i sin tur gör dock i sin tur att våra resultat inte blir lika starka, då det inte lutar så mycket att luta sig tillbaka mot?

1p

SU, STATISTIK

Skrivsal: Ugglevikssalen **Anonymkod:** 0032-MMP **Blad nr:** 2

Fråga 2

a) $\mu = 100$ $\sigma = 15$ $n = 2$ $Z_\alpha = 1,645$ (ensidigt test)

Vi söker det värde vi minst måste observera för att kunna förkasta nollhypotesen på en 5%ig nivå.

$RR = Z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$ $Z \geq \frac{\bar{y} - 100}{15/2}$ $1,645 = \frac{\bar{y} - 100}{7,5}$

$1,645 \cdot 7,5 \geq \bar{y} - 100$ $112,3375 \geq \bar{y}$

Svar: Det kritiska värdet är 112,3375.

5

b) $P(Z \leq \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}) = P(Z \leq \frac{112,3375 - 115}{15/\sqrt{4}}) = P(Z \leq -0,355)$

$= 1 - (1 - \Phi(0,36)) = 1 - 1 - (0,6408) = 0,6406$

Svar: Testets styrka (β) då $\mu = 115$ är 0,6406

6

c) kritiskt värde förändras då n ändras!

Vi leter efter $-Z_\alpha = \frac{112,3375 - 115}{15/\sqrt{n}} = -1,645$ (−4)

5

$\frac{112,3375 - 115}{-1,645} = 15/\sqrt{n}$ $1,61854 = \frac{15}{\sqrt{n}}$ $n = \left(\frac{15}{1,61854}\right)^2$

$n = 85,89$.

Följdfel

Svar: Det krävs minst 86 observationer.

16p

SU, STATISTIK

Skrivsal: Ugglevikssalen **Anonymkod:** 0032-MMP **Blad nr:** 3

Fråga 3

a)

$p=0,5=H_0$: Det finns ingen systematisk skillnad,

$p\neq 0,5=H_1$: Det finns en systematisk skillnad

Signifikansnivå: 5%

T^+ = antalet positiva observationer

$T^+=1$

$n=11$ då obs K är samma för båda termometrar.

Vi använde binomialfördelningen för $n=11$ och $p=0,5$ för $Y=1$ och $Y=0$.

	A	B	Tecken
A	6,8	7,2	−
B	9,8	12,3	−
C	2,1	5,3	−
D	6,2	6,8	−
E	7,1	7,2	−
F	6,5	6,2	+
G	9,3	10,7	−
H	1,0	2,7	−
I	−0,2	1,3	−
J	9,6	9,8	−
K	12,0	12,0	/
L	6,3	8,9	−

$P(Y=1)=\binom{11}{1}0,5^1\cdot 0,5^{10}=0,00537$

$P(Y=0)=\binom{11}{0}0,5^0\cdot 0,5^{11}=0,00049$

$0,00537+0,00049=0,00586$

Då det är ett dubbelsidigt test måste vi gångra med 2. $2\cdot 0,00586=0,01172$

Vi har ett p-värde på 0,01172

Vi förkastar H_0 då $0,01172<0,05$ att det inte finns någon systematisk skillnad och antar samtidigt alternativhypotesen att det finns en systematisk skillnad mellan termometrarna.

R

8

b) De antaganden vi måste göra är:

Observationerna är oberoende av varandra

Populationerna är normalfördelade

Konstant varians

Vi använder teststatistikan

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\bar{Y}_1 = \frac{76,5}{12} = 6,375$$

$$\bar{Y}_2 = \frac{89,8}{12} = 7,483...$$

	Y_1	Y_2
	6,8	7,2
	9,8	12,3
	2,1	5,3
	6,2	6,8
	7,1	7,2
	6,5	6,2
	9,3	0,1
	1,0	2,7
	-0,2	1,3
	9,6	9,8
	12,0	12,0
	6,3	8,9
Σ:	76,5	89,8

För att hitta sp måste vi hitta s_1^2 och s_2^2.

$$s^2 = \frac{\Sigma y_i^2 - n(\bar{y})^2}{n-1}$$

$\Sigma Y_1^2 = 641,17$ $\Sigma Y_2^2 = 797,98$

$$s_1^2 = \frac{641,17 - 12 \cdot 6,375^2}{12-1} = 13,953$$

$$s_2^2 = \frac{797,98 - 12 \cdot 7,483^2}{12-1} = 11,452$$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{11 \cdot 13,953 + 11 \cdot 11,452}{22} = 12,7025$$

$$s_p = \sqrt{s_p^2} = \sqrt{12,7025} = 3,564$$

(Kom på senare att man kan använda ett enklare test enbart för differensen)

SU, STATISTIK

Skrivsal: Ugglevikssalen **Anonymkod:** 0032-MMP **Blad nr:** 4

Fråga 3 fortsättning

Nu har vi alla tal för att kunna genomföra testet

$$t = \frac{6,375 - 7,483}{3,564\sqrt{\frac{1}{12}+\frac{1}{12}}} = -0,76$$

Vi använder en 5%-ig signifikansnivå

H_0: Det finns ingen systematisk skillnad mellan termometrarna

H_1: Det finns en systematisk skillnad mellan termometrarna

Vi förkastar H_0 om $|t_{obs}| > t_{krit}$.

$t_{krit} = t_{\alpha/2}(n_1+n_2-2)$ $t_{0,025}(22) = 2,07$

$|-0,76| \not> 2,07$

Svar: Vi kan inte förkasta nollhypotesen att det inte finns någon skillnad. Vi kan alltså inte påvisa någon skillnad mellan termometrarna.

Detta är fri relaterade mål

12p

SU, STATISTIK

Skrivsal: Ugglevikssalen **Anonymkod:** 0032-MMl **Blad nr:** 5

Fråga 4

a) En exponentialfördelning med väntevärde λ ser ut på följande vis: $\frac{1}{\lambda} \cdot e^{-y/\lambda}$

$L(\lambda) = \frac{1}{\lambda^n} e^{-\Sigma Y_i/\lambda} = \lambda^{-n} e^{-\Sigma Y_i/\lambda}$

$\ell(\lambda) = -n \ln \lambda - \frac{\Sigma Y_i}{\lambda}$

$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{-n}{\lambda} + \frac{\Sigma Y_i}{\lambda^2}$

$\frac{-n}{\lambda} + \frac{\Sigma Y_i}{\lambda^2} = 0 \qquad \frac{n}{\lambda} = \frac{\Sigma Y_i}{\lambda^2} \qquad n = \frac{\Sigma Y_i}{\lambda} \quad \hat{\lambda}_{ML} = \frac{\Sigma Y_i}{n} = \bar{Y}$

Svar: $\hat{\lambda}_{ML} = \bar{Y}$ R

b) $E(\hat{\lambda}_{ML}) = E\left(\frac{\Sigma Y_i}{n}\right) = \frac{1}{n} E(\Sigma Y_i) = \frac{1}{n} \cdot n E(Y) = \frac{1}{n} \cdot n \lambda = \frac{n}{n} \cdot \lambda = \lambda$

Svar: λ_{ML} är väntevärdesriktig. R

c) $V(\lambda_{ML}) = V\left(\frac{\Sigma Y_i}{n}\right) = \frac{1}{n^2} V(\Sigma Y_i) = \frac{1}{n^2} \cdot n V(Y) = \frac{1}{n^2} \cdot n \cdot \lambda^2$

$= \frac{n\lambda}{n^2} = \frac{\lambda^2}{n}$ Svar: $V(\hat{\lambda}_{ML}) = \frac{\lambda^2}{n}$ R

d) λ_{ML} är en konsistent estimator av λ då R

$n \to \infty \Rightarrow \frac{\lambda^2}{n} \to 0$ Svar: Ja, λ_{ML} är en konsistent estimator.

Vänd $\rightarrow$

Fråga 4

e) $V(\lambda_{ML}) = \frac{\lambda^2}{n} \Longrightarrow \frac{1}{\frac{\lambda^2}{n}} \cdot e^{-\frac{\frac{y}{\lambda^2}}{\frac{\lambda^2}{n}}} = \frac{n}{\lambda^2} \cdot e^{-\frac{ny}{\lambda^2}}$

$L(\lambda) = \frac{n}{\lambda^{2n}} \cdot e^{-\frac{n\Sigma y_i}{\lambda^2}}$

?

$l L(\lambda) = \ln(n) - 2n \ln\lambda - \frac{n\Sigma y_i}{\lambda^2}$

$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{-2n}{\lambda} + \frac{n\Sigma y_i}{\lambda^3} = 0$

$\frac{2n}{\lambda} = \frac{n\Sigma y_i}{\lambda^3} \qquad 2n = \frac{n\Sigma y_i}{\lambda^2} \qquad 2 = \frac{\Sigma y_i}{\lambda^2}$ ✓

$\lambda^2 = \frac{\Sigma y_i}{2} \qquad \lambda = \left(\frac{\Sigma y_i}{2}\right)^2 = \frac{(\Sigma y_i)^2}{4}$

Svar: Maximum likelihoodestimatorn av variansen för $\hat{\lambda}_{ML}$ är $\frac{(\Sigma y_i)^2}{4}$

16p

SU, STATISTIK

Skrivsal: Ugglevikssalen **Anonymkod:** 0032-MMP **Blad nr:** 6

Fråga 5

a) $L(p) = p^n (1-p)^{\Sigma(y_i - 1)}$ $\left(p_{ML} = \frac{1}{\bar{y}}\right)$ $= \left(\frac{1}{\bar{y}}\right)^n \left(1 - \frac{1}{\bar{y}}\right)^{\Sigma y_i - 1}$

$L(0,8) = 0,8^n (1-0,8)^{\Sigma(y_i - 1)}$

$$\frac{L(0,8)}{L(p)} < k \implies \frac{0,8^n \cdot 0,2^{\Sigma(y_i - 1)}}{\left(\frac{1}{\bar{y}}\right)^n \cdot \left(1 - \frac{1}{\bar{y}}\right)^{\Sigma(y_i - 1)}} < k$$

om H_0

b)

c) $n = 40$ $p = \frac{40}{52} = 0,77$