STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2018-03-16

Skrivtid: 14.00-19.00
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 2 april.

OBS! Glöm inte att ange nödvändiga antaganden överallt

Uppgift 1. (20 poäng)

Förklara innebörden av följande begrepp:
a) Samplingfördelning
b) Konsistens
c) Signifikansnivå
d) Styrkan hos ett test
e) Likformigt starkaste test (UMP)

Uppgift 2. (20 poäng)

För att utvärdera en träningsgrupp genomförde man bl.a. ett Cooper-test (man mäter hur många meter man hinner springa under 12 minuter). Variansen av resultaten är ett mått på hur jämn träningsgruppen är. 18 personer genomförde Cooper-testet och en standardavvikelse på 42.8 meter uppmättes. Ange nödvändiga förutsättningar och beräkna ett 90 %-igt konfidensintervall för variansen. Ge en tolkning av det intervall du har beräknat.

Uppgift 3. (20 poäng)

En chipstillverkare lanserar lagom till finalen i Melodifestivalen en ny sorts chips, Mellochips. För att testa om de nya Mellochipsen upplevs som bättre än de tidigare Originalchipsen, fick åtta slumpmässigt utvalda studenter från någon population av studenter betygsätta chipsen på en skala från ett till tio. Se tabellen nedan.

Student	A	B	C	D	E	F	G	H
Originalchips	6	4	5	8	3	6	7	5
Mellochips	8	9	4	7	9	9	7	9

a) Sätt upp lämpliga hypoteser och gör nödvändiga antaganden för att testa om det går att påvisa att studenter i population av studenter föredrar Mellochipsen.

b) Testa hypoteserna i uppgift a) med hjälp av ett lämpligt icke-parametriskt test.

Uppgift 4. (20 poäng)

Antag att tiden det tar i minuter att besvara en viss typ av tentafråga är exponentialfördelad med väntevärde β. Med hjälp av n oberoende observationer, $y_1, y_2, \ldots, y_n$, vill man nu använda likelihoodkvottestet för att testa $H_0 : \beta = 25$ mot $H_a : \beta \neq 25$.

a) Bestäm likelihoodfunktionen för $\beta = 25$.

b) Härled maximum likelihood-skattningen av β.

c) Bestäm likelihoodfunktionen för $\beta = \widehat{\beta}_{ML}$.

d) Bestäm teststatistikan för likelihoodkvottestet, λ_{LR}.

e) 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.

f) I ett slumpmässigt urval av 40 frågor blev den totala tiden 1220 minuter. Genomför testet.

Uppgift 5. (20 poäng)

Mimmi och Rasmus är intresserade av energibesparande åtgärder i hemmet och gör därför en del kartläggningar av deras energianvändning. Bl.a. mäter de upp olika mängder vatten i en kastrull och mäter tiden det tar att värma vattnet från 20° C till 100° C. Olika mätfel och störningsfaktorer gör att den uppmätta tiden, Y_i, kan betraktas som en normalfördelad stokastisk variabel med väntevärdet βx_i och variansen σ^2, där x_i är mängden vatten i kastrullen och β är en okänd parameter. Variansen σ^2 kan antas vara känd. Härled maximum likelihoodestimatorn av β om Mimmi och Rasmus provar n stycken vattenmängder x_i och för varje mängd observerar tiden Y_i.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 16/3 2018

Sal: Brunnsvikssalen

Tenta: Statistisk teori med tillämpningar II

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: 0016-XFR

[X] 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					5
Lär.ant. 19	16	20	10	2					

POÄNG	BETYG	Lärarens sign.
67 + 6 = 73	D	[illegible]

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: 0016-XFR Blad nr: 1

1. a) samplingfördelning är en sannolikhets fördelning för en statistika. Statistika är en skattning av en okänt populationsparameter med hjälp av ett stickprov.

b) Konsistens är en av "properties" av estimatorer. En estimator är konsistent om (gäller om $\hat{\theta}$ är vvr)
$\lim_{n\to\infty} V(\hat{\theta})=0$ villkor betyder att variansen minskar och går mot noll när n ökar.
inte generellt (-1)

c) Signifikansnivå som är alfa (α) (eller testets signifikansnivå) är ett typ I-fel P(förkasta H0 | H0 är sann).
alfa används som en nivå/gräns för att avgöra om vi ska förkasta H0 eller inte. Ett nyss beräknat p-värde jämförs med signifikansnivån, om P-värdet är lägre än signifikansnivån så förkastas nollhypotesen

d) Styrkan hos ett test beräknas matematiskt $1-\beta$ där Beta är ett typ II-fel P(Ej förkasta H0 | Ha är sann)
Styrkan är power$(\hat{\theta}) = 1-\beta$

17p

e) likformigt starkaste test (UMP) är ett test som är det starkaste testet oavsett vilken alternativhypotes vi väljer (ex H0: $\lambda=1$ & Ha: $\lambda>1$)
kommer ihåg att vi skulle tänka på Ha som att den är sammansatt av många (oändligt) enkla hypoteser

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: 0016-XFR Blad nr: 2

(2) n = 18 standardavvikelse: 42,8 meter

90%-ki för variansen

$\alpha = 0,1$

vi antar ~~[illegible]~~ att urvalet är slumpmässigt (oberoende obs) och att vi observerar en normal fördelning stok. var.

$P\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]$ vilket ger oss ett 100(1-α)% igt ki

för $\sigma^2 \Rightarrow P \frac{(n-1)s^2}{\chi^2_{\alpha/2}} ; \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$

- $\chi^2_{\alpha/2}(n-1)$ $\chi^2_{0,05}(17) = 27,59$ [ur tabell]
- $\chi^2_{1-\alpha/2}(n-1)$ $\chi^2_{0,95}(17) = 8,67$ [ur tabell]

Det 90% iga ki för σ^2 ges av

(-2)

Svar: $\left(\frac{17 \cdot 42,8^2}{27,59}, \frac{17 \cdot 42,8^2}{8,67}\right) \approx \left(\underset{\hat{\theta}_L}{1128,716202}, \underset{\hat{\theta}_U}{3591,843137}\right)$

vi vet med 90% konfidens att den sanna variansen ligger mellan

Vad för def? (-2)

(1128,716202 ; 3591,843137)

16p

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: 0016-XFR Blad nr: 3

student	X originalchips	Y mellochips	Di	rang
3) A	6	8	2	3
B	4	9	5	6
C	5	4	-1	1,5
D	8	7	-1	1,5
E	3	9	6	7
F	6	9	3	4
G	7	7	tie	
H	5	9	4	5

X: originalchips fördras
Y: mellochips . . . –

Wilcoxons rangtest

Antaganden: parvisa observationer, beroende obs. mellan mars e försök
oberoende observationer mellan de olika studenterna

b)

Ho: X o Y har samma fördelning

Ha: Förd. för Y ligger till höger om förd. för X.

Test variabel: $T^- = 1,5 + 1,5 = 3$

$\alpha = 0,01$, $n = 7$ pga en "tie"

För $\alpha = 0,05$ blir $RR = \{T \leq 4\}$

Svar: H0 förkastas på 5% nivån ($T^- = 3$ som är i $RR = \{T \leq 4\}$)

Vi har stöd att med 0,05 signifikans säga att förd för Y ligger till höger om förd. för X.

2p

// a) är på andra sidan ... läste frågan fel... använder samma hypoteser ...
Dvs. icke parametriskt på den här sidan o T-test på baksidan
=)

a) T-test (CGS gäller inte, antar n.f ($Y_1, Y_2 \ldots Y_n$ random sample from a n.f med $E(Y_i) = m$)

parvisa Obs

$H_0: m = m_0$ $\quad \alpha = 0{,}05$ $\quad$ teststatistiken: $T = \dfrac{\bar{D} - 0}{S/\sqrt{n}} \sim t(n-1) = t(7)$

$H_a: m > m_0$ (upper-tail)

· $\bar{D} = \dfrac{18}{8} = 2{,}25$ $\quad S^2 = \dfrac{\sum d_i^2 - n\bar{D}^2}{n-1} = \dfrac{92 - 8 \cdot 2{,}25^2}{7} = 7{,}357142857$

· vi förkastar H_0 om $T_{obs} > T_{krit} = 1{,}89$ ← tabell 3, $fg = 7$ & $\alpha = 0{,}05$

· $T_{obs} = \dfrac{2{,}25}{\sqrt{7{,}357\ldots/8}} = 2{,}346242607$

Svar: vi förkastar H_0 på 5% sign.nivå pga. $T_{obs} > T_{krit}$ $(2{,}34\ldots > 1{,}89)$ samma anledning som med det icke parametriska testet.

vi har stöd med 0,05 sign nivå säga att fördelningen för Y ligger åt högre än förd. för X.

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: 0016-XFR Blad nr: 4

$Y \sim Exp(\beta)$

(4) Y: tiden det tar för att svara en viss typ av lenta frågor

$f(y) = \frac{1}{\beta} e^{-y/\beta}$ $\beta > 0$, $0 < y < \infty$ $H_0: \beta = 25$, $\beta \neq 25$

a o b) $L(Y) = \prod_{i=1}^{n} \frac{1}{\beta^n} \cdot e^{-y/\beta} \Rightarrow \frac{1}{25} \cdot e^{-y/25}$ ✓ 0

Logaritmerar $\prod_{i=1}^{n} \frac{1}{\beta^n} \cdot e^{-y/\beta}$ pga ober

• $\ell(Y) = \ln \beta^{-n} + \ln e^{-\sum y_i/\beta} = -n\ln(\beta) + \left(-\frac{\sum y_i}{\beta}\right) \ln e = -n\ln\beta - \frac{\sum y_i}{\beta}$

Deriverar m.a.p β

$\frac{d\ell(Y)}{d\beta} = \frac{-n}{\beta} - \sum y_i \cdot (-\beta^{-2}) = \frac{-n}{\beta} + \frac{\sum y_i}{\beta^2}$

Sätt Derivatan $= 0$

$\frac{-n}{\beta} + \frac{\sum y_i}{\beta^2} = 0 \Rightarrow \frac{\sum y_i}{\beta^2} = \frac{n}{\beta} \Rightarrow \frac{\sum y_i \beta}{\beta^2} = n \Rightarrow \hat{\beta}_{ML} = \frac{\sum y_i}{n} = \bar{Y}$ 4

c) $\beta = \hat{\beta}_{ML} \left(= \frac{\sum y_i}{n}\right)$ 2

$L(\hat{\beta}_{ML}) = \prod_{i=1}^{n} \frac{1}{\hat{\beta}_{ML}} e^{-y/\hat{\beta}_{ML}} \Rightarrow \frac{1}{\sum y_i/n} \cdot e^{-y/\sum y_i/n}$

d) Teststatistikan för ett L-R test ges av $\lambda_{LR} = \frac{\max_{\theta \in \Omega_0} L(\theta)}{\max_{\theta \in \Omega} L(\theta)} = \frac{\frac{1}{25} \cdot e^{-y/25}}{\frac{1}{\sum y_i/n} \cdot e^{-y/\frac{\sum y_i}{n}}}$ ✓ 0

e) kritiskt område RR för sign.nivå 0,01 ges utav

- $RR = \{\lambda < k \mid H_0 \text{ är sann}\} \Rightarrow RR\{-2\ln(\lambda_{LR}) > \underbrace{-2\ln k}_{k^*} \sim \chi^2_{(1)}$ för sign.nivå 0,01

R $k^* = 6{,}63$ 4

$RR = \{-2\ln(\lambda_{LR}) > 6{,}63\}$

f) $\lambda_{LR} = \frac{\frac{1}{25} e^{-y/25}}{\frac{1}{\sum y_i/n} \cdot e^{-y/\sum y_i/n}} = \frac{\frac{1}{25} e^{-1220/25}}{\frac{1}{30{,}5} e^{-1220/30{,}5}} = \frac{2{,}56147E^{-23}}{1{,}34290303E^{-19}} = 1{,}838943514E^{-4}$? ✓ 0

$-2\ln(\lambda_{LR}) = -2\ln(1{,}838943514E^{-4}) = 17{,}20224828$ $(17{,}20224828 > 6{,}63)$

Svar: Vårt Beräknade λ_{LR} ligger i RR. Vi förkastar H_0 på signifikans nivå $\alpha = 0{,}01$.

~~Vi accepterar H_0 istället~~

10p

SU, STATISTIK

Skrivsal: Brunnsviksalen Anonymkod: 0016-XFR Blad nr: 5

(5) $Y_i \sim N(\beta x_i, \sigma^2)$ $f(y) = \frac{1}{\sigma\sqrt{2\pi}} Exp\left[-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2\right]$

$-\infty < y < +\infty$

x_i = mängden vatten i kastrullen

β = okänd parameter

Variansen antas vara känd. pga ober?

$L(\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} Exp\left[-\frac{(y_i - \beta \cancel{x_i})^2}{2\sigma^2}\right] = \frac{1}{(2\pi\sigma)^{n/2}} exp\left[-\frac{\sum(y_i - \beta \cancel{x_i})^2}{2\sigma^2}\right]$

loggar: $\ell(\beta) = -\frac{n}{2}\ln(2\pi\sigma^2) + \frac{\sum(y_i - \beta)^2}{2\sigma^2}$ 2

$\frac{d\ell(\beta)}{d\beta} = \frac{\sum y_i - n\beta}{\sigma^2}$

Sätt derivatan $= 0 \Rightarrow \hat{\beta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{Y}$ 2p

Ber om ursäkt att jag inte sätter in x_i på β innan stycket. Du brukar vilja ha det när vi ska sätta derivatan $= 0$..

gjorde samma sak på fråga 4...

Svar: $\hat{\beta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{Y}$ ✓