STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2019-03-21

Skrivtid: 09.00-14.00
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)

Jeffs sport och fiske är en sport- och friluftsbutik belägen i en galleria. För att planera bemanningen i butiken vill VDn Jeff veta om det finns empiriskt evidens för att försäljningen under måndagar är större än försäljningen under lördagar. Man noterar därför försäljningen i tusentals kronor under ett slumpmässigt urval av 25 måndagar och 25 lördagar Försäljningen under måndagarna var då i medeltal 1078 tkr med urvalsstandardavvikelsen 633 och under lördagarna var försäljningen i medeltal 908.2 tkr med urvalsstandardavvikelsen 469.8. Sätt upp lämpliga hypoteser för att testa om försäljningen under måndagar är större än försäljningen under lördagar, ange nödvändiga antaganden för att testa hypoteserna och genomför testet på signifikansnivån 5 %.

Uppgift 2. (20 poäng)

Man kan anta att antal skott på mål som en fotbollsspelare gör per match är en Poissonfördelad stokastisk variabel med parameter λ. Under de senaste åtta matcherna har Lisa skjutit följande antal skott på mål

$$1 \quad 0 \quad 1 \quad 3 \quad 2 \quad 1 \quad 2 \quad 3$$

a) Bestäm maximumlikelihoodskattningen av λ.

b) Bestäm maximumlikelihoodskattningen av sannolikheten att Lisa inte skjuter något skott på mål under en match.

Uppgift 3. (20 poäng)

Ett slumpmässigt urval av 12 finansiella analytiker fick prediktera de procentuella prisförändringarna under kommande året på två företags aktier. Resultaten visas i nedanstående tabell.

Analytiker	A	B	C	D	E	F	G	H	I	J	K	L
Företag 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
Företag 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 finansiella analytiker bedömer att priserna på företagens aktier kommer att utvecklas lika.

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

Uppgift 4. (20 poäng)

Tiden det tar för en kund att handla i Monas livsmedelsbutik observerades hos 64 slumpmässigt utvalda kunder. Det observerade urvalsmedelvärdet blev 23 minuter med urvalsvariansen 256.

a) Uppskatta tiden det tar att handla i medeltal för kunderna i butikens kundkrets med ett 90 %-igt konfidensintervall.

b) Hur många observationer skulle behövas för att intervallets längd i uppgift a) skulle vara högst 5 minuter? Utgå från att populationens varians är 256.

Uppgift 5. (20 poäng)

Mimmi och Rasmus är intresserade av energibesparande åtgärder i hemmet och gör därför en del kartläggningar av hushållets energianvändning. Bl.a. vill de uppskatta vikten av det dagliga matavfallet. Man kan anta att vikten av matavfallet är en stokastisk variabel med täthetsfunktionen

$$f(y;\beta) = \frac{1}{6\beta^4} y^3 e^{-y/\beta}, \quad 0 < y < \infty, \; 0 < \beta < \infty.$$

För denna fördelning gäller det att förväntat värde är $E(Y) = 4\beta$ och variansen är $V(Y) = 4\beta^2$, dvs. förväntad vikt av matavfallet en dag är 4β och variansen för vikten av matavfallet en dag är $4\beta^2$. Din uppgift är att hjälpa Mimmi och Rasmus med att konstruera en bra estimator av det dagliga matavfallet och då i första hand en bra estimator av β.

a) Härled maximum likelihoodestimatorn av β.

b) Avgör om den estimator du härledde i uppgift a) är en unbiased estimator av β.

c) Avgör om den estimator du härledde i uppgift a) är en konsistent estimator av β.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 21/3-2019

Sal: Brunnsvikssalen

Tenta: Statistisk teori med tillämpningar 2

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: 0036-KHL

☒ 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					8
Lär.ant. 20	12	20	20	20					

POÄNG	BETYG	Lärarens sign.
92 + 6 = 98	A	JW

SU, STATISTIK

Skrivsal: Brunnsviksalen **Anonymkod:** 0036-KHL **Blad nr:** 1

Uppg 1.

Jett vill undersöka om det finns empiriskt stöd för att försäljningen under måndagar är större än försäljningen under söndagar.

Y_1 = "försäljningen under måndagar"

$\bar{Y}_1 = 1078$ tkr

$S = 633$ tkr

$n_1 = 25$

Y_2 = "försäljningen under lördagar"

$\bar{Y}_2 = 908,2$ tkr

$S_2 = 469,8$ tkr

$n_2 = 25$

[illegible]

Eftersom försäljningen på måndagar vs lördagar kan antas oberoende använder vi teststatistikan:

$$t_{obs} = \frac{\bar{Y}_1 - \bar{Y}_2 - D_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ där S_p är den poolade variansen.

$\alpha = 0,05$

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

1. Antaganden: ∘ Vi antar oberoende och slumpmässigt urval

∘ Vi antar normalfördelade pop.

∘ Vi antar lika varianser $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Hypoteser: $H_0: \mu_1 = \mu_2$

$H_A: \mu_1 > \mu_2$

Beslutsregel: förkasta H_0 om $t_{obs} > t_{0,05}(n_1+n_2-2)$

$= t_{0,05}(48) = 1,675$

Vi har endast $t_{0,05}(40) = 1,68$ och $t_{0,05}(60) = 1,67$ formelsamlingen så eftersom $t_{0,05}(48)$ ligger ungefär mitten (a vä came är sipass ll a) tar vi medelvärdet $\frac{1,67 + 1,68}{2}$ 1,675

$$S_p^2 = \frac{(25-1)633^2 + (25-1)469,8^2}{48} = 310700,52$$

$$S_p = \sqrt{310700,52} = 557,4052$$

$$t_{obs} = \frac{1078 - 909,2 - 0}{557,4052\sqrt{\frac{1}{25}+\frac{1}{25}}} = \frac{169}{157,6580} = 1,0770$$

Bra!

$t_{obs} = 1,0770 < 1,67 = t_{0,05}(48)$

→ Vi kan alltså ej förkasta H_0

Svar: Det finns inte något empiriskt stöd för någon skillnad i försäljning mellan lördagar och måndagar

20p

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0036-KHL **Blad nr:** 2

Uppg 2

$Y =$ "antal skott på mål under en match"

$Y \sim Po(\lambda)$ $\quad f(y) = \frac{\lambda^{y}}{y_i!} e^{-\lambda}$

$n = 8$

Match	1	2	3	4	5	6	7	8
$f(y) =$	1	0	1	3	2	1	2	3

$\hat{\lambda} = \bar{Y} = \frac{\sum y_i}{n} = \frac{13}{8} = \underline{1,625}$ enligt formelsamling 4

a) Vi söker $\hat{\lambda}_{ML}$

Först hittar vi likelyhoodfunktionen genom täthetsfunktionen.

$$L(\lambda) \overset{\text{vid oberoende}}{=} f(y_1) \cdot f(y_2) \cdot \ldots \cdot f(y_n) = \frac{\lambda^{y_1}}{y_1!} e^{-\lambda} \cdot \frac{\lambda^{y_2}}{y_2!} e^{-\lambda}$$

$$\cdot \frac{\lambda^{y_n}}{y_n!} e^{-\lambda} = \prod_{i=1}^{n} \frac{\lambda^{y_i}}{y_i!} e^{-\lambda} = \frac{\lambda^{\sum y_i}}{\prod_{i=1} y_i!} e^{-n\lambda}$$

Sedan logaritmerar vi likelyhoodfunktionen:

$$l(\lambda) = \sum y_i \cdot \ln(\lambda) + (\underbrace{\ln(1)}_{=0} - \sum \ln(y_i)) - n\lambda \cdot \ln(e)$$

$$= \sum y_i \cdot \ln(\lambda) - \sum \ln(y_i) - n\lambda$$

Sedan deriverar vi den logaritmerade funktionen

$$l'(\lambda) = \sum y_i \cdot \frac{1}{\lambda} - n = \frac{\sum y_i}{\lambda} - n$$

Vi sätter derivatan till noll för att få ut estimatorn

$$\frac{\sum y_i}{\hat{\lambda}_{ML}} - n = 0$$

$$\frac{\sum y_i}{\hat{\lambda}_{ML}} = n$$

$$\hat{\lambda}_{ML} = \frac{\sum y_i}{n} = \overline{Y} \qquad \text{V.S.V.}$$

8

Svar: $\hat{\lambda}_{ML} = \frac{\sum y_i}{n} = \underline{\overline{Y}}$

b) Y = "Lisa skjuter inte några skott på mål under en match" Men Y är ju $Po(\lambda)$!

$Y \sim Bin(n = 1, P)$ $\qquad f(y) = p^y (1-p)^{n-y}$

Vi söker $\hat{P}_{ML}$ $\qquad n = 8$

Först hittar vi likelyhoodfunktionen genom täthetsfunktionen.

$$L(p) \overset{\text{om oberoende}}{=} f(y_1) \cdot f(y_2) \cdot \ldots \cdot f(y_n) = p^{y_1}(1-p)^{n-y_1} \cdot$$

$$p^{y_2}(1-p)^{n-y_2} \cdot \ldots \cdot p^{y_n}(1-p)^{n-y_n}$$

$$= \prod_{i=1}^{n} p^{y_i}(1-p)^{n-y_i} = p^{\sum y_i}(1-p)^{n-\sum y_i}$$

SU, STATISTIK

Skrivsal: Brunnsviksalen **Anonymkod:** 0036-KHL **Blad nr:** 3

forts. Uppg. 2

Sedan logaritmerar vi likelihoodfunktionen.

$$l(p) = \sum y_i \ln(p) + (n - \sum y_i) \ln(1-p)$$

Vi deriverar den logaritmerade likelihoodfunktionen

$$l'(p) = \sum y_i \cdot \frac{1}{p} + (n - \sum y_i) \cdot \frac{1}{1-p} \cdot (-1)$$

$$= \frac{\sum y_i}{p} - \frac{(n - \sum y_i)}{1-p}$$

Vi sätter derivatan till noll för att få vår estimatorn.

$$\frac{\sum y_i}{\hat{p}_{ML}} - \frac{n - \sum y_i}{1 - \hat{p}_{ML}} = 0$$

$$\frac{\sum y_i}{\hat{p}_{ML}} = \frac{n - \sum y_i}{1 - \hat{p}_{ML}}$$

$$\sum y_i (1 - \hat{p}_{ML}) = \hat{p}_{ML}(n - \sum y_i)$$

$$\sum y_i - \hat{p}_{ML} \sum y_i = \hat{p}_{ML} n - \hat{p}_{MLE} \sum y_i$$

$$\sum y_i = \hat{p}_{ML} n$$

$$\hat{p}_{ML} = \frac{\sum y_i}{n}$$

V.S.B. $\hat{p} = \frac{\sum y_i}{n} = \frac{1}{8} = 0.125$

12p

Svar: Enligt maximum likelihoodestimatorn för sannolikheten att Lisa inte skjuter ett skott på mål under en match är $\hat{p}_{ML} = \frac{\sum y_i}{n}$ vilket ger oss

$\hat{p} = \frac{1}{8} = 0.125$ ✓

SU, STATISTIK

Skrivsal: Brunnsvikssolen **Anonymkod:** 0036-KHL **Blad nr:** 4

Uppg. 3

Teckentest (icke parametriskt)

Analytiker	Företag C	Företag B	d_i	d_i^2	Tecken
A	6,8	7,2	-0,4	0,16	−
B	9,8	12,3	-2,5	6,25	−
C	2,1	5,3	-3,2	10,24	−
D	6,2	6,8	-0,6	0,36	−
E	7,1	7,2	-0,1	0,01	−
F	6,5	6,2	0,3	0,09	+
G	9,3	10,1	-0,8	0,64	−
H	1,0	2,7	-1,7	2,89	−
I	-0,2	1,3	-1,5	2,25	−
J	9,6	9,8	-0,2	0,04	−
K	12,0	12,0	0 (tie)	0	
L	6,3	8,9	-2,6	6,76	−
	Σ 76,5	Σ 89,8	Σ −13,3	29,69	

Antaganden: Observationerna är slumpmässigt utvalda och paren är ömsesidigt oberoende

Teckentestet används för att jämföra beroende par där de bivariata talen är ömsesidigt oberoende.

Hypotes: $H_0: p = \frac{1}{2}$ vs $H_A: p \neq \frac{1}{2}$

Testvariabel: M = antal positiva differenser = 7

$M \sim Bin(n = 11, p = \frac{1}{2})$ om H_0 är sann.

Eftersom vi har en tie tar vi bort den och kvar har vi då $n = 12 - 1 = 11$

P-värdet fås ur binomialfördelningstabell för $M \leq 7$ och är 0,00586 ✓

?

VGV. (−3)

Vi kan alltså förkasta H_0 på signi.nivåer ända ned till p-värdet dvs 0,0058

Svar: Det finns empiriskt stöd för att bedömningen från finansiella analytiker av företag A:s och företag B:s aktierisk skiljer sig åt.

b) T-test (parametriskt)

Eftersom det är beroende par använder vi teststatistikan $t_{obs} = \frac{\bar{D}}{S_d/\sqrt{n}}$

Antaganden (vs Teckentest)

- Lika: Slumpmässigt utvalda par där d blir ~~riktat~~ talen är ömsesidigt ~~oberoende~~ beroende
- Olika: Teckentestet och t-testet skiljer sig åt i att t-testet kräver normalfördelad pop. (vilket ej icke-parametriska test som teckentestet gör). T-testet kräver även antagande om lika varians.

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: 0036-KHL Blad nr: 5

fort uppg 3

Testvariabel: $t_{obs} = \frac{\bar{D}}{s_d/\sqrt{n}}$ $\alpha = 0{,}05$

$\bar{D} = \frac{\sum d_i}{n} = \frac{-13{,}3}{12} = -1{,}1083$

$s_d^2 = \frac{\sum d_i^2 - n\bar{d}^2}{n-1} = \frac{29{,}69 - 12 \cdot (-1{,}1083)^2}{12-1} = \frac{14{,}9491}{11} = 1{,}3590$

Beslutsregel: förkasta H_0 om $|t_{obs}| > t_{0{,}025}(11)$
$= 2{,}20$

$t_{obs} = \frac{-1{,}1083}{\sqrt{\frac{1{,}3590}{12}}} = -3{,}2934$

$|t_{obs}| = 3{,}2934 > 2{,}20 = t_{0{,}025}(11) \Longrightarrow$

$\Longrightarrow$ vi kan förkasta H_0 på 5 % sig nivå

3

Svar: Det finns empiriskt stöd för att finanstidens analytikers bedömning av företag A och företag B:s aktier skiljer sig åt.

20p

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0036-KHL **Blad nr:** 6

Uppg 4

$n = 64$

Y = tiden det tar att handla i Monas "Livsmedelsbutik"

$\bar{Y} = 23$ $s^2 = 256$

a) Vi söker ett 90% k.i för tiden det tar att handla i medeltal.

Teststatistika: $\bar{Y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$

Eftersom $n<30$ ~~$n=64$~~ och den sanna pop variansen σ^2 ej är känd använder vi $t_{\alpha/2}$ (istället för $z_{\alpha/2}$)

$\alpha = 0{,}1 \Rightarrow t_{\alpha/2}(n-1) = t_{0{,}05}(n-1=63) = 1{,}67$
ok

$23 \pm 1{,}67 \sqrt{\frac{256}{64}}$

$23 \pm 3{,}34$

$\Downarrow$

$[19{,}66 \, , \, 26{,}34]$

Svar: Den sanna tiden det tar i medeltal att handla i Monas butik ligger med 90% konfidens inom intervallet $[19{,}66 \, ; \, 26{,}34]$

fort uppg 4

b) Hur många observationer skulle krävas för att intervallets längd (i a)) ska vara högst 5 min?

Enligt uppg är $\sigma^2 = 256$ vilket gör att vi kan använda ett K.I. för $z_{\alpha/2}$

Intervallets längd = 2B

$\underbrace{z_{\alpha/2} \frac{s}{\sqrt{n}}}_{= B}$

Vi söker n för: $2B = 5$

$B = \frac{5}{2} = 2{,}5$

$$z_{\alpha/2} \frac{s}{\sqrt{n}} = 2{,}5$$

$$\frac{s}{\sqrt{n}} = \frac{2{,}5}{z_{\alpha/2}}$$

$$\frac{s}{\left(\frac{2{,}5}{z_{0{,}05}}\right)} = \sqrt{n} \longrightarrow n = \frac{s^2}{\left(\frac{2{,}5}{1{,}6449}\right)^2} = 110{,}83$$

$\underline{\underline{n = 111}}$ (pga runda upp)

Svar: För att intervallets längd ska vara högst 5 minuter måste n vara minst 111

20p

(Hade vi ej vetat $\frac{\quad}{2}$

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 0036-KHL **Blad nr:** 7

Uppg 5

Y = "vikten av metanfoket"

$$f(y_i;\beta) = \frac{1}{6\beta^4} y^3 e^{-y/\beta} \quad 0 < y < \infty \quad 0 < \beta < \infty$$

$$E(y) = 4\beta \qquad V(Y) = 4\beta^2$$

a) Härled maximum likelihood skattning för β dvs $\hat{\beta}_{ML}$

Vi tar fram likelihoodfunktionen genom täthetsfunktionen

$$L(\beta) = \prod_{i=1}^{n} \frac{1}{6\beta^4} y_i^3 e^{-y_i/\beta} = \left(\frac{1}{6\beta^4}\right)^n \cdot \prod_{i=1}^{n} y_i^3 \cdot e^{-\sum y_i/\beta}$$

Vi logaritmerar likelihoodfunktionen

$$l(\beta) = n \cdot \left(\underbrace{\ln(1)}_{=0} - (\ln(6) + 4\ln(\beta))\right) + 3\sum \ln(y_i) - \frac{\sum y_i}{\beta} \ln(e)$$

$$= n \cdot (-4\ln(\beta) - \ln(6)) + 3\sum \ln(y_i) - \frac{\sum y_i}{\beta}$$

$$= -4n\ln(\beta) - n\ln(6) + 3\sum \ln(y_i) - \frac{\sum y_i}{\beta}$$

Vi deriverar den logaritmerade funktionen

$$l'(\beta) = -4n \cdot \frac{1}{\beta} - \frac{\sum y_i}{\beta^2}(-1)$$

$$= -\frac{4n}{\beta} + \frac{\sum y_i}{\beta^2}$$

Vi sätter derivatan till noll för att få ut estimatorn.

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

$$\frac{4n}{\hat{\beta}_{ML}} = \frac{\sum y_i}{\hat{\beta}_{ML}^2}$$

$$\frac{4n\hat{\beta}_{ML}^2}{\hat{\beta}_{ML}} = \sum y_i$$

$$4n\hat{\beta}_{ML} = \sum y_i$$

$$\hat{\beta}_{ML} = \frac{\sum y_i}{4n}$$

Svar: $\hat{\beta}_{ML} = \frac{\sum y_i}{4n}$

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** 3036-KHL **Blad nr:** 8

forts. Uppg. 5.

b) En estimator är väntevärdesriktig om:

$E(\hat{\theta}) = \theta$

$E(\hat{\beta}_{ML}) = E\left(\frac{\sum y_i}{4n}\right) = \frac{1}{4n} E(\sum y_i) = \frac{1}{4n} \cdot n E(y)$

$= \frac{n}{4n} \cdot 4\beta = \frac{4\beta}{4} = \beta$ VVR

$E(\hat{\beta}_{ML}) = \beta$

Svar: Ja estimatorn $\hat{\beta}_{ML}$ är väntevärdesriktig

c) En estimator är konsistent om den är väntevärdesriktig och om sannolikheten att komma godtyckligt nära det sanna värdet ökar och går mot 1 då $n \to \infty$

Det kan även skrivas som

$$\boxed{\lim_{n \to \infty} V(\hat{\theta}_n) = 0}$$

$V(\hat{\beta}_{ML}) = V\left(\frac{\sum y_i}{4n}\right) = \left(\frac{1}{4n}\right)^2 \cdot V(\sum y_i) =$

$= \left(\frac{1}{4n}\right)^2 \cdot n \cdot V(y_i) = \left(\frac{1}{4n}\right)^2 \cdot n \cdot 4\beta^2$

$= \frac{1}{4^2 n^2} \cdot n \cdot 4\beta^2 = \frac{4n\beta^2}{4^2 n^2} = \frac{\beta^2}{4n}$

då $n \to \infty$ går $\frac{\beta^2}{4n} \to 0$ R

Estimatorn $\hat{\beta}_{ML}$ är alltså konsistent

Svar: Ja, estimatorn $\hat{\beta}_{ML}$ är en konsistent estimator av β_{ML}

20