STOCKHOLMS UNIVERSITET
Statistiska institutionen
Ellinor Fackle-Fornius

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2016-04-13

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.
Genomgång av tentamen sker 2016-05-04 kl. 15-16 i sal B315.

OBS! Glöm inte att ange nödvändiga antaganden överallt.

Uppgift 1. (20 poäng)

Avgör om vart och ett av följande påståenden är sant eller falskt. Motivera ditt svar!

i) Sannolikheten att det 95 %-iga konfidensintervallet $(30, 50)$ innehåller parametern θ är 0.95.

ii) Sannolikheten att det 95 %-iga Bayesianska "credibility"-intervallet $(30, 50)$ innehåller parametern θ är 0.95.

iii) En lägre konfidensgrad ger ett bredare konfidensintervall (allt annat lika).

iv) p-värdet är sannolikheten att nollhypotesen är sann.

v) p-värdet är alltid större än signifikansnivån.

Uppgift 2. (20 poäng)

Under de 32 första börsdagarna 2016 gick aktien Eriksson A upp $(+)$ eller ner $(-)$ enligt följande sekvens

$- + - - + + + - - + + - + - - + + - + + - - + + - + - + + - ++$

a) Testa med ett Run-test om kursrörelserna för Eriksson A är slumpmässiga eller inte. Använd signifikansnivån 5 %.

b) Testa med ett tecken-test om sannolikheten att kursen för Eriksson A går upp är 0.5 eller inte. Använd signifikansnivån 5 %.

Uppgift 3. (20 poäng)

Det är känt att fett kan lagra inflammatoriska sjukdomar. I en medicinsk undersökning ville man därför undersöka om BMI är relaterad till en viss inflammatorisk sjukdom. BMI mättes hos 14 patienter och hos 14 kontroller (som inte har sjukdomen ifråga) där patient och kontroll har matchats med avseende på ålder, kön och rökvanor. De erhållna observationerna finns i följande tabell. BMI för patienter och BMI för kontroller antas vara normalfördelade.

Patient	Kontroll
25.41	26.93
26.15	29.04
35.15	26.08
30.60	23.62
27.00	21.11
23.20	27.97
27.70	28.09
29.00	24.31
35.98	29.07
26.30	35.91
31.00	27.47
31.31	26.10
28.03	28.22
28.41	27.61

a) Beräkna et 99 %-igt konfidensintervall för skillnaden mellan BMI för patienter och BMI för kontroller.

b) Testa på signifikansnivån 1 % om BMI för patienter skiljer sig åt från BMI för kontroller.

Uppgift 4. (20 poäng)

Rayleighfördelningen används ofta för att modellera vindhastighet. Antag att vindhastigheten i m/s (Y) är Rayleighfördelad med täthetsfunktion

$$f(y) = \frac{y}{\theta} e^{\frac{-y^2}{2\theta}}, \quad y \geq 0, \quad \theta > 0$$

Låt $Y_1, Y_2, \ldots, Y_n$ vara ett slumpmässigt urval av n vindhastigheter.

a) Härled maximumlikelihood-skattningen för θ.

b) Beräkna maximumlikelihood-skattningen för θ då ett slumpmässigt urval av $n = 40$ vindhastigeter har observerats. Observationerna sammanfattas enligt följande:

$$\sum_{i=1}^{40} y_i = 52.11, \quad \sum_{i=1}^{40} y_i^2 = 88.19, \quad \sum_{i=1}^{40} \ln(y_i) = 3.38$$

Uppgift 5. (20 poäng)

Antag att Y är en stokastisk variabel som är Rayleighfördelad (som i uppgift 4) och att man vill testa

$$\begin{aligned} H_0 &: \ \theta = 1 \\ H_a &: \ \theta \neq 1 \end{aligned}$$

med ett likelihoodkvottest.

a) Bestäm likelihoodfunktionen för $\theta = 1$.

b) Bestäm likelihoodfunktionen för $\theta = \widehat{\theta}_{ML}$.

c) Bestäm teststatistikan för likelihoodkvottestet.

d) Ange kritiskt område (RR) för signifikansnivån 0.05 (använd asymptotisk fördelning) och genomför testet för de data som angivits i uppgift 4 b).

Stockholms universitet

Statistiska institutionen

Rättningsblad

Datum: 13/4-2016

Sal: Värtasalen

Tenta: Statistisk teori med tillämpningar 2

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: STM-0031

☒ 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. 20	19	20	20	16					

95 + 8 bonus

POÄNG	BETYG	Lärarens sign.
103	A	

SU, STATISTIK

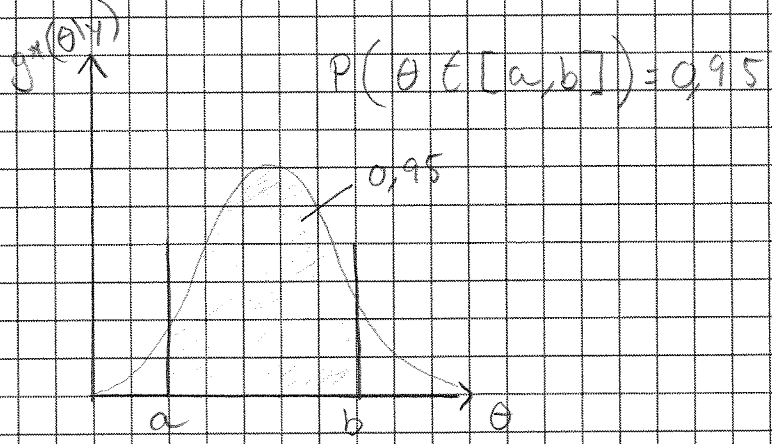
Skrivsal: VÅ   Anonymkod: STM-0031   Blad nr: 1

7)

i) Falskt. Inom klassisk inferens ses parametrarna (θ) som man vill skatta som fixa och datan som slumpmässig. Detta innebär att sannolikheten att θ inneslutes i intervallet antingen är 0 eller 1. Antingen finns θ i konfidensintervallet, eller så gör det inte det. Däremot kan man säga att med 95 % konfidens/säkerhet innehåller konfidensintervallet θ. Skapar man väldigt många konfidensintervall på liknande sätt, kommer $100(1-\alpha)$ %, i det här fallet 95 %, av dessa intervall att innehålla parametern θ, om $n \to \infty$. Bra

4

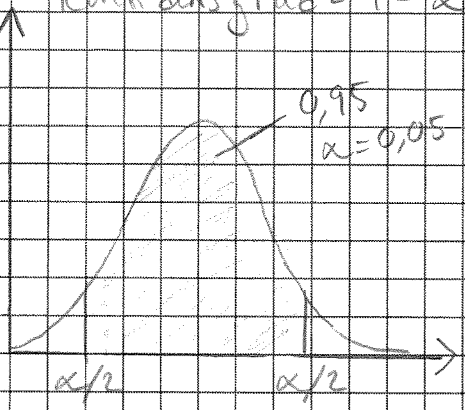
ii) Sant. Inom Bayesiansk teori ses parametrarna som stokastiska variabler, vilket gör att man kan tolka det 95 %-iga "credibility"-intervallet som att sannolikheten eller kredibiliteten att θ inneslutes i intervallet är 0,95.

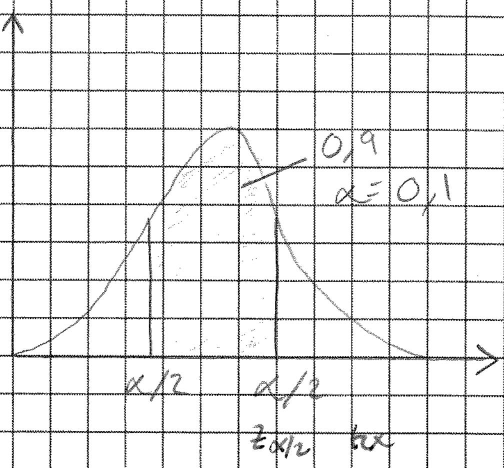


Bra!

4

iii) Konfidensgrad = $1-\alpha$





Om konfidensgraden minskar, innebär det att konfidensen att parametern finns i intervallet minskar, dvs. α ökar. Således är påståendet falskt.

och $z_{\alpha/2}$ minskar

4

iv) Falskt. P-värdet anger den lägsta signifikansnivå som vi kan förkasta H_0 på, givet det observerade värdet. Det anger alltså sannolikheten att få det observerade värdet, eller något ännu mer extremt.

4

v) Falskt. Om p-värdet är lägre än signifikansnivån, kan vi förkasta H_0. Är p-värdet däremot större än signifikansnivån, kan vi inte förkasta H_0. P-värdet är alltid detsamma, givet det observerade värdet, medan signifikansnivån kan variera, då vi själva bestämmer α. Om p-värdet alltid skulle vara större än signifikansnivån, skulle alla nollhypoteser förkastas.

4

20

SU, STATISTIK

Skrivsal: VÅ Anonymkod: STM-0031 Blad nr: 2

2) a) − + − − + + + − − + + − + − − + + − + + − − + + − + − + + − + +

H_0: Sekvensen / kursförelserna för Eriksson A är slumpmässig

H_a: Sekvensen är inte slumpmässig. R

R = 20 - antal Runs R

$n_1 = 14$

$n_2 = 18$

Då både n_1 och n_2 är större än 10, approximerar jag med normalfördelning, om H_0 är sann, enligt CGS.

Teststatistika:

$Z = \frac{R - E(R)}{\sqrt{V(R)}}$ R

$E(R) = \frac{2 n_1 n_2}{n_1 + n_2} + 1 = \left(\frac{2 \cdot 14 \cdot 18}{14 + 18}\right) + 1 = 16,75$ R

$V(R) = \frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} = \frac{2 \cdot 14 \cdot 18 (2 \cdot 14 \cdot 18 - 14 - 18)}{(14 + 18)^2 (14 + 18 - 1)}$

$= \frac{237888}{31744} = 7,49$ R

$Z_{obs} = \frac{R - E(R)}{\sqrt{V(R)}} = \frac{20 - 16,75}{\sqrt{7,49}} = 1,187$ R

$\alpha = 0,05$ och testet är dubbelsidigt:

$Z_{\alpha/2} = Z_{0,975} = 1,96 \Rightarrow RR = \{Z_{obs} > 1,96$ eller $Z_{obs} < -1,96\}$ R

Slutsats: Då $Z_{obs} = 1,19 < Z_{krit} = 1,96$ kan vi inte förkasta Ho på signifikansnivå $\alpha = 0,05$. Vi drar slutsatsen att kursrörelserna för Ericssonaktien är slumpmässiga. 10 p

b) Förutsättningar för teckentest:

Ej parvisa obs här

Parvisa observationer, dvs att det finns ett beroende mellan mätningar. Man kan ha mätt aktiens upp/nedgång på förmiddagen och sedan mätt upp/nedgången på kvällen igen, för att sedan jämföra om upp och/eller nedgången totalt sätt under en dag är upp(+) eller ner(-). Dessutom krävs oberoende mellan dessa parvisa mätningar i så fall.

$H_0: p = 0,5$

$H_a: p \neq 0,5$ R

$M = 18$ = antal (+)

$M \sim Bin(n=32; p=0,5)$ om H_0 är sann R

Eftersom $n_1 > 10$ och $n_2 > 10$ approximerar jag med N.F istället, om H_0 är sann.

Teststatistika:

$$Z = \frac{M - n/2}{(1/2)\sqrt{n}} = \frac{18 - 32/2}{1/2 \cdot \sqrt{32}} = 0,707$$ R

$\alpha = 0,05 \Rightarrow Z_{\alpha/2} = Z_{0,025} = 1,96$

$RR = \{|Z_{obs}| > 1,96\}$

Slutsats: Då det observerade Z-värdet är mindre än det kritiska, kan vi inte förkasta H_0. Därmed kan vi inte säga att sannolikheten för att kursen går upp är större eller mindre än 0,5. 9

19

SU, STATISTIK

Skrivsal: VÄ   Anonymkod: STM-0031   Blad nr: 3

3) Då patienter och kontroller har matchats med avseende på ålder, kön och rökvanor, antar jag beroende mellan patienter och kontroller. Bra

Vidare antar jag oberoende mellan paren, dvs. mellan patienter, respektive mellan kontroller sinsemellan.

n	Patient	Kontroll	D_i = Patient − kontroll
1	25,41	26,93	−1,52
2	26,15	29,04	−2,89
3	35,15	26,08	9,07
4	30,60	23,62	6,98
5	27,00	21,11	5,89
6	23,20	27,97	−4,77
7	27,70	28,09	−0,39
8	29,00	24,31	4,69
9	35,98	29,07	6,91
10	26,30	35,91	−9,61
11	31,00	27,47	3,53
12	31,31	26,10	5,21
13	28,03	28,22	−0,19
14	28,41	27,61	0,8

BMI för patienter, respektive kontroller, antas vara normalfördelade, men då $n < 30$ kan jag inte approximera med $Z(\mu, \sigma^2)$, så jag använder t-fördelningen istället.

$\bar{D} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ Den sanna variansen är okänd, så jag skattar den med urvalsvariansen.

$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{23,71}{14} = 1,6936$

$s^2 = \frac{\sum_{i=1}^{n} D_i^2 - n(\bar{D})^2}{n-1} = \frac{401,6 - 14 \cdot 1,6936^2}{14-1} = 27,8$

$s = \sqrt{27,8} = 5,27$

$t_{\alpha/2}(n-1) = t_{0,005}(13) = 3,01$

VGV →

Ett 99%-igt KI för skillnaden mellan BMI för patienter och kontroller ges av:

$$1,69 \pm 3,01 \cdot \frac{5,27}{\sqrt{14}} = 1,69 \pm 4,24$$

$$[-2,55 \ ; \ 5,93]$$ R

10

b) $H_0: \mu_D = 0$

$H_a: \mu_D \neq 0$ R

Jag använder samma antaganden som i a) och eftersom $n < 30$, använder jag ett t-test.

Antaganden för t-test är att fördelningarna för BMI för patienter, respektive kontroller, är normalfördelade. (Varianserna är okända men antas vara lika.) σ_D^2 okänd

$$T = \frac{\bar{D} - \mu_0}{s/\sqrt{n}} = \frac{1,69 - 0}{5,27/\sqrt{14}} = 1,1998 \approx 1,2$$

Kritiskt värde:

$\alpha = 0,01$, dubbelsidigt test $\Rightarrow t_{krit} = t_{\alpha/2}(n-1) =$

$t_{0,005}(14-1) = 3,01$

Förkasta H_0 om $t_{obs} > 3,01$ eller $t_{obs} < -3,01$ R

Slutsats: Då det observerade t-värdet (1,2) är mindre än det kritiska (3,01), kan vi ej förkasta H_0 på signifikansnivån 1%. Vi kan inte dra slutsatsen att BMI för patienter skiljer sig från BMI för kontroller.

10

20

SU, STATISTIK

Skrivsal: VÄ Anonymkod: STM-0031 Blad nr: 4

4) $f(y) = \frac{y}{\theta} e^{\frac{-y^2}{2\theta}}, \quad y \geq 0, \ \theta > 0$

a) $L(y_i|\theta) = \prod_{i=1}^{n} \frac{y_i}{\theta} e^{-\frac{y_i^2}{2\theta}} = \frac{1}{\theta^n} \prod_{i=1}^{n} y_i \, e^{-\frac{\sum y_i^2}{2\theta}}$ R

pga. oberoende
då s.u.

$l(\theta) = \ln L(\theta) = \ln 1 - n\ln(\theta) + \sum_{i=1}^{n} \ln y_i - \frac{\sum y_i^2}{2\theta}$

$= -n\ln(\theta) + \sum_{i=1}^{n} \ln y_i - \frac{\sum y_i^2}{2\theta}$ R

$\frac{d\,l(\theta)}{d\theta} = \frac{-n}{\theta} - \frac{\sum_{i=1}^{n} y_i^2}{2}(\theta^{-1-1}) \cdot (-1) = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} y_i^2}{2\theta^2}$ R

$\frac{d\,l(\theta)}{d\theta} = 0 \implies \frac{-n}{\theta} + \frac{\sum_{i=1}^{n} y_i^2}{2\theta^2} = 0$

$\frac{n}{\theta} = \frac{\sum_{i=1}^{n} y_i^2}{2\theta^2}$

$n \cdot 2\theta = \sum_{i=1}^{n} y_i^2$

$\hat{\theta}_{ML} = \frac{\sum_{i=1}^{n} y_i^2}{2n}$ R

17

b) Givet:

$n = 40$

$\sum_{i=1}^{40} y_i^2 = 88{,}19$

$\hat{\theta}_{ML} = \frac{88{,}19}{2 \cdot 40} = 1{,}1024$

ML-skattningen för θ är 1,1. R

3

20

SU, STATISTIK

Skrivsal: VÅ Anonymkod: STM-0031 Blad nr: 5

5)

$H_0: \theta = 1$

$H_a: \theta \neq 1$

$f(y) = \frac{y}{\theta} e^{-\frac{y^2}{2\theta}}$

a, $L(y_i \mid \theta = 1) \underset{\substack{\text{pga. ober} \\ \text{då s.a.}}}{=} \prod_{i=1}^{n} \left(\frac{y_i}{1} e^{-\frac{y_i^2}{2 \cdot 1}} \right) = \prod_{i=1}^{n} y_i \, e^{-\frac{\sum y_i^2}{2}}$

R

2

b, $\hat{\theta}_{ML} = \frac{\sum y_i^2}{2n}$ (enligt beräkningar i uppgift 4)

$L(y_i \mid \theta = \hat{\theta}_{ML}) \underset{\substack{\text{pga. ober} \\ \text{då s.a.}}}{=} \prod_{i=1}^{n} \left(\frac{y_i}{\hat{\theta}_{ML}} e^{-\frac{\sum y_i^2}{2\hat{\theta}_{ML}}} \right) = \left(\frac{1}{\hat{\theta}_{ML}} \right)^n \cdot \prod_{i=1}^{n} y_i \, e^{-\frac{\sum y_i^2}{2\hat{\theta}_{ML}}}$

$= \frac{1}{\left(\frac{\sum y_i^2}{2n} \right)^n} \cdot \prod_{i=1}^{n} y_i \, e^{-\sum y_i^2 / \left(2 \frac{\sum y_i^2}{2n} \right)}$

2

c, $\lambda_{LR} = \frac{\max_{\theta \in \Omega_0} L(\theta)}{\max_{\theta \in \Omega} L(\theta)} = \frac{L(\theta = 1)}{L(\theta = \hat{\theta}_{ML})} < k$

R

$\lambda_{LR} = \frac{\cancel{\prod y_i} \, e^{-\frac{\sum y_i^2}{2}}}{\left(\frac{1}{\hat{\theta}_{ML}} \right)^n \cancel{\prod y_i} \, e^{-\frac{\sum y_i^2}{2\hat{\theta}_{ML}}}} = \frac{e^{-\frac{\sum y_i^2}{2} - \left(-\frac{\sum y_i^2}{2\hat{\theta}_{ML}} \right)}}{\left(\frac{1}{\hat{\theta}_{ML}} \right)} =$

$e^{-\frac{\sum y_i^2}{2} + \frac{\sum y_i^2}{2\hat{\theta}_{ML}}} \cdot \hat{\theta}_{ML} < k$

R

$RR = \left\{ e^{-\frac{\sum y_i^2}{2} + \frac{\sum y_i^2}{2\hat{\theta}_{ML}}} \cdot \hat{\theta}_{ML} < k \right\}$

4

d) $\ln \lambda_{LR} < \ln k$

$2 \ln \lambda_{LR} < 2 \ln k$

$-2 \ln \lambda_{LR} > -2 \ln k$

$-2 \ln \lambda_{LR} \sim \chi^2(n-1) \quad \chi^2(1)$

$-2 \ln \lambda_{LR} = -2 \ln \left(e^{-\frac{\sum y_i^2}{2} + \frac{\sum y_i^2}{2\hat{\theta}_{ML}}} \cdot \hat{\theta}_{ML} \right)$

$= -2 \left(-\frac{\sum y_i^2}{2} + \frac{\sum y_i^2}{2\hat{\theta}_{ML}} \right) \cdot \ln \hat{\theta}_{ML}$

$= -2 \left(-\frac{88,19}{2} + \frac{88,19}{2 \cdot 1,1} \right) \cdot \ln(1,1)$

$= (88,19 - 80,17) \cdot \ln(1,1) = 0,7644$

$\chi^2_{0,05}(n-1) = \chi^2_{0,05}(39) \approx 55$

$-2 \ln k = 55$

$\ln k = \frac{55}{-2} = 27,5$

8

16

Stockholms universitet

Statistiska institutionen

Rättningsblad

Datum: 13/4-2016

Sal: Värtasalen

Tenta: Statistisk teori med tillämpningar 2

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: STM-0029

☒ 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 st
Lär.ant. 16	20	17	17	8					

78 + 8 bonus

POÄNG	BETYG	Lärarens sign.
86	B	

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: STM-0029 Blad nr: 1

Uppgift 1.

i) Falskt! När konfidensintervallet är uträknat är gränserna inte längre s.v. och det går inta att tala om sannolikheter. Bra!
Istället kan en tala om konfidens att parametern finns inom intervallet.

4

ii) Sant. Då parametern θ har en sannolikhets-fördelning inom den Bayesianska inferensen kan en säga att det är en sannolikhet.

4

iii) Konfidensgraden är $1-\alpha$; en lägre konfidensgrad ger ett lägre α-värde och därför ett smalare konfidensintervall.
Falsk.

4

iv)

–

v) p-värdet är den lägsta signifikansnivån där H_0 kan förkastas, så det beror på vilken signifikansnivå som väljs.
Falsk.

4

16

SU STATISTIK

Skrivsal: Värtasalen   Anonymkod: STM-0029   Blad nr: 2

· Uppgift 2

a) 32 börsdagar $(n_1 + n_2)$

n_1 = antal upp (+) = 18
n_2 = antal ned (−) = 14
} > 10 så kan approximera med Z.

R = antal runs = 20 $\alpha = 0{,}05$

Hypoteser

H_0: kursrörelserna är slumpmässiga

H_A: kursrörelserna är inte slumpmässiga

Beslutsregel

Förkasta H_0 om $|Z^{obs}| > Z_{0,025}$

$|Z^{obs}| > 1{,}9600$

Teststatistika

$$Z = \frac{R - E(R)}{\sqrt{V(R)}} \quad \text{där } E(R) = \frac{2n_1n_2}{n_1+n_2} + 1, \quad V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}$$

Test

$$Z^{obs} = \frac{20 - \left(\frac{2\cdot18\cdot14}{32} + 1\right)}{\sqrt{\frac{2\cdot18\cdot14(64-18-14)}{32^2\cdot(18-14-1)}}} = \frac{20 - 14{,}75}{\sqrt{\frac{16128}{3072}}} = \frac{5{,}25}{\sqrt{5{,}25}} \approx 2{,}2913$$

$Z^{obs} \approx 2{,}2913$ större än 1,96 så H_0 förkastas och kursrörelserna är troligen inte slumpmässig.

ok, följdfel

10

Uppgift 2 forts.

b) Teckentest

Hypoteser

H_0: s/h att kursen går upp $p = 0,5$
H_A: —||— $p \neq 0,5$ } dubbelsidigt

R

M = antalet positiva tecken = 18 R

$M \sim Bin(n, p)$ Då urvalet är relativt stort
är M approximativt n.f. och

Teststatistikan

$$Z = \frac{M - n/2}{(1/2)\sqrt{n}}$$ R

Beslutsregel

förkasta H_0 om $|Z^{obs}| > z_{0,025}$
$|Z^{obs}| > 1,96$ R

Test

$$Z^{obs} = \frac{18 - 32/2}{(1/2)\sqrt{32}} = \frac{2}{2,8284...} \approx 0,7071$$ R

$Z^{obs} \approx 0,7071$ alltså mindre än det
kritiska värdet och H_0 kan inte
förkastas. s/h att kursen går upp
tycks vara 0,5 och slumpmässig...
Går emot testet i a) kanske för
att teckentestet är mindre pålitligt.

10

(26)

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: STM-0029 Blad nr: 3

Uppgift 3

parvisa observationer med beroende inom paren men oberoende mellan par. R

X = patient Y = kontroll $n = 14$

X & Y antas normalfördelade.

a)

X_i	Y_i	$D_i = X_i - Y_i$
25,41	26,93	-1,52
26,15	29,04	-2,89
35,15	26,08	9,07
30,60	23,62	6,98
27,00	21,11	5,89
23,20	27,97	-4,77
27,70	28,09	-1,09 -0.39
29,00	24,31	4,69
35,98	29,07	6,91
26,30	35,91	-9,61
31,00	27,47	3,53
31,31	26,10	5,21
28,03	28,22	-0,19
28,41	27,61	0,8

$\Sigma D_i = 23,01$

Få observationer men från en normalfördelad population gör t-fördelning lämplig.

Konfidensintervall

$\bar{D} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$ där R

$$S^2 = \frac{\sum_{i=1}^{n}(D_i - \bar{D})^2}{n-1}$$

$$\bar{D} = \frac{\Sigma D_i}{n} = \frac{23,01}{14} = 1,64357\ldots$$

$\approx 1,6436$ ≈R

$$S^2 = \frac{\sum_{i=1}^{14}(D_i - \bar{D})^2}{13} = \frac{364,8397}{13}$$

$\approx 28,0646$ ≈R

$S \approx 5,2976$

$t_{0,005}(\cancel{14}) \Rightarrow 2,98$

$n-1 = 13$

KI för skillnaden i BMI mellan patienter och kontoller

$$1,6436 \pm 2,98 \cdot \frac{5,2976}{\sqrt{14}} = 1,6436 \pm 4,2192$$

$[-2,5756; 5,8628]$ 9

Uppgift 3 forts.

b) Hypoteser

$\mu_0 = 0$ $\quad H_0: D = 0$ Det finns ingen skillnad i BMI för patient och kontroll

$\mu_0 \neq 0$ $\quad H_A: D \neq 0$ Det finns en skillnad.

Teststatistika

$$T = \frac{\bar{D} - D_0}{S/\sqrt{n}}$$ R

Beslutsregel

förkasta H_0 om $|t^{obs}| > t_{0,005}(\overset{13}{\cancel{14}})$

$|t^{obs}| > 2,98$

$$t^{obs} = \frac{1,6436 - 0}{5,2976/\sqrt{14}} = 1,1609$$ ≈R

H_0 kan inte förkastas på 1%-signifikansnivå

Stämmer överens med att $\bar{D} = 1,6436$? $\mu_0 = 0$! inneslutst av konfidensintervallet.

Det är inte troligt att det finns en skillnad i BMI mellan patienter och kontroller.

8

17

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: STM-0029 Blad nr: 4

· Uppgift 4

Y: "vindhastighet i m/s" Rayleighfördelad med täthetsfunktionen

$$f(y) = \frac{y}{\theta} e^{\frac{-y^2}{2\theta}}, \quad y \geq 0, \theta > 0$$

$Y_1, Y_2, \dots, Y_n$ är s.u. av n vindhastigheter.

a) Maximumlikelihoodskattningen för θ

Likelihoodfunktionen för θ

$$L(\theta) = L(y_1, y_2, \dots, y_n | \theta) = \prod_{i=1}^{n} \frac{y_i}{\theta} e^{\frac{-y^2}{2\theta}} = \frac{y_1}{\theta} e^{-y_1^2/2\theta} \cdot \frac{y_2}{\theta} e^{-y_2^2/2\theta} \cdot \frac{y_n}{\theta} e^{-y_n^2/2\theta} \quad R$$

pga ober.

$$= \left(\frac{1}{\theta}\right)^n \cdot \prod_{i=1}^{n} y_i \cdot e^{-\sum_{i=1}^{n} y_i^2/2\theta}$$

Logaritmera

$$\ell(\theta) = \ln L(\theta) = n \cdot (\underbrace{\ln 1}_{0} + \ln \theta) + \sum_{i=1}^{n} \ln y_i - \underbrace{\frac{\sum_{i=1}^{n} y_i^2}{2\theta}}_{\frac{\sum_{i=1}^{n} y_i^2}{2} \cdot \theta^{-1}}$$

Derivera m.a.p. θ

$$\frac{d\,\ell(\theta)}{d\theta} = \frac{n}{\theta} - (-1) \cdot \sum_{i=1}^{n} \frac{y_i^2}{2} \cdot \theta^{-2} \quad R$$

$$= \frac{n}{\theta} + \frac{\sum_{i=1}^{n} y_i^2}{2\theta^2} = 0$$

Sätt derivatan till noll hitta maxpunkten

$$\frac{n}{\theta} = -\frac{\sum_{i=1}^{n} y_i^2}{2\theta^2}$$

$$\frac{2\theta^2}{\theta} = -\frac{\sum_{i=1}^{n} y_i^2}{n}$$

$$\hat{\theta}_{ML} = -\frac{\sum_{i=1}^{n} y_i^2}{2n}$$

+ , $\theta > 0$

14

Uppgift 4 forts.

b) Beräkna maximumlikelihood-skattningen för θ då s.u. av $n=40$ vindhastigheter observerats från (a)

$$\hat{\theta}_{ML} = -\frac{\sum_{i=1}^{n} y_i^2}{2n} = -\frac{88,19}{2 \cdot 40} = -1,102375 \approx \underline{-1,1024}$$

ok, följdfel

3

17

STATISTIK

rivsal: Värtasalen   Anonymkod: STM-0029   Blad nr: 5

Uppgift 5

Y är en s.v. med Rayleighfördelning

Testa $H_0: \theta = 1$

$H_A: \theta \neq 1$

med likelihoodkvottest.

a) från uppgift 4

$$L(1) = 1 \cdot \prod_{i=1}^{n} y_i \cdot e^{-\sum_{i=1}^{n} y_i^2/2}$$ R 2

b) $$L(\hat{\theta}_{ML}) = \left(\frac{1}{\hat{\theta}_{ML}}\right)^{n} \cdot \prod_{i=1}^{n} y_i \cdot e^{-\sum_{i=1}^{n} y_i^2/2\hat{\theta}_{ML}}$$ R 2

c) $$\lambda_{LR} = \frac{\max_{\theta \in \Omega_0} L(\cancel{x}\,\theta)}{\max_{\theta \in \Omega} L(\cancel{\hat{\theta}_{ML}}\,\theta)} = \frac{L(\Omega_0)}{L(\Omega)}$$ 1

d) kritiskt område

$$-2\ln(\lambda_{LR}) \sim \chi^2(\cancel{n}\,1)$$

$$-2\ln(\lambda_{LR}) > \chi^2_{\alpha/2}(f) \qquad \chi^2_{0,025}(40) \Rightarrow 59,34$$

✓ eller $-2\ln(\lambda_{LR}) < \chi^2_{1-\alpha/2}(f) \qquad \chi^2_{0,975}(40) \Rightarrow 24,43$

RR: $\{24,43 > -2\ln(\lambda_{LR}) > 59,34\}$ ← förkasta H_0 om ✓

$$\lambda_{LR} = \frac{\max_{\theta\in\Omega_0} L(1)}{\max_{\theta\in\Omega} L(-1,1024)} = \frac{\cancel{\prod_{i=1}^{40} y_i} \cdot e^{-88,19/2}}{\left(\frac{1}{-1,1024}\right)^{40} \cdot \cancel{\prod_{i=1}^{40} y_i} \cdot e^{-88,19/(2\cdot(-1,1024))}}$$

$$= \frac{e^{-44,095}}{0,02025 \cdot e^{39,999}} = \frac{e^{-44,095-39,999}}{0,02025} = \frac{e^{-84,094}}{0,02025}$$

3

$$-2\ln(\lambda_{LR}) = 160,3888 \qquad = 1,4859 \cdot 10^{-35}$$