STOCKHOLMS UNIVERSITET
Statistiska institutionen
Ellinor Fackle-Fornius

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2016-03-17

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 2016-04-04 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) Enligt centrala gränsvärdessatsen närmar sig urvalets fördelning en standardiserad normalfördelning när urvalsstorleken ökar.

ii) $\widehat{\theta}_A$ och $\widehat{\theta}_B$ är två väntevärdesriktiga estimatorer för parametern θ. $\widehat{\theta}_A$ är mer effektiv än $\widehat{\theta}_B$ om $V(\widehat{\theta}_A)$ är mindre än $V(\widehat{\theta}_B)$.

iii) En biased (icke väntevärdesriktig) estimator kan aldrig vara konsistent.

iv) Inom klassisk inferens är parametern θ en stokastisk variabel.

v) Inom bayesiansk inferens är parametern θ en stokastisk variabel.

Uppgift 2. (20 poäng)

Kreditupplysningsföretaget F^3 håller på att utveckla en ny metod för att identifiera risker. Om metoden fungerar i mindre än 50 % av fallen läger man ner utvecklingen av metoden. Man genomför därför ett test av hypotesen

$$\begin{aligned} H_0 &: \quad p = 0.5 \\ H_a &: \quad p < 0.5 \end{aligned}$$

genom att tillämpa den på 12 fall som man vet är risker för att se hur många av dessa man kan identifiera. Nollhypotesen förkastas, d.v.s. utvecklingen läggs ner, om tre eller färre risker kan identifieras.

a) Bestäm testets signifikansnivå.

b) Bestäm testets styrka då $p = 0.3$.

c) Vad händer med signifikansnivån och styrkan (då $p = 0.3$) om man tillämpar metoden på fler än 12 fall men behåller samma förkastelseområde?

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.

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) Testa med ett tecken-test om BMI för patienter skiljer sig åt från BMI för kontroller.

b) Testa med ett Wilcoxon teckenrang-test om BMI för patienter skiljer sig åt från BMI för kontroller.

Uppgift 4. (20 poäng)

Antag att Y är en stokastisk variabel som är betafördelad med $\beta = 1$. Täthetsfunktionen är då

$$f(y) = \alpha y^{\alpha-1}, \quad 0 < y < 1, \quad \alpha > 0.$$

a) Visa att momentskattningen av α när ett slumpmässigt urval av n observationer görs är

$$\widehat{\alpha}_{MOM} = \frac{\overline{Y}}{1-\overline{Y}}.$$

b) Visa att maximumlikelihood-skattningen av α när ett slumpmässigt urval av n observationer görs är

$$\widehat{\alpha}_{ML} = -\frac{n}{\sum_{i=1}^{n} \ln(Y_i)}.$$

c) Vad blir maximumlikelihood-skattningen av $E(Y)$?

Uppgift 5. (20 poäng)

Antag att Y är en stokastisk variabel som är betafördelad med $\beta = 1$ (som i uppgift 4). Om $\alpha = 1$ så blir det en likformig fördelning på intervallet $(0, 1)$. Man kan testa för likformig fördelning genom att testa

$$H_0: \ \alpha = 1$$
$$H_a: \ \alpha \neq 1$$

med ett likelihoodkvottest. Antag att ett slumpmässigt urval av n observationer görs.

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

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

c) Bestäm teststatistikan för likelihoodkvottestet.

d) Ange kritiskt område (RR) för signifikansnivån 0.05 när n är stort.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 17/3-2016

Sal: Brunnsvikssalen

Tenta: Statistisk teori med tillämpningar 2

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: STM-0020

☑ 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. 13	17	15	20	14					

79 + 8 bonus

POÄNG	BETYG	Lärarens sign.
87	B	[signature]

SU, STATISTIK

Skrivsal: Brunnviksalen Anonymkod: STM-0026 Blad nr: 1

1

i) $\lim_{n\to\infty} Z \sim N(0,1)$

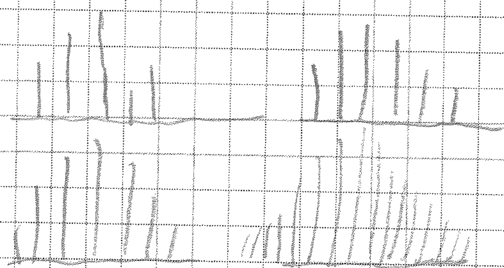
Sant. När antalet observationer ökar för sig fördelningen för urvalet en standardiserad n.f med väntevärde μ och varians σ^2

✓

exempel

en binomial fördelning kan approximeras med en standardiserad n.f där $np = \mu$ och $np(1-p) = \sigma^2$ för stora n (>30)

vilket ger $Z = \frac{\mu - \mu_0}{\sqrt{\frac{p(1-p)}{n}}}$



0

ii) $\hat{\theta}_A$ och $\hat{\theta}_B$ vvr

$V(\hat{\theta}_A) < V(\hat{\theta}_B)$

$$\text{eff}(\hat{\theta}_A, \hat{\theta}_B) = \frac{V(\hat{\theta}_B)}{V(\hat{\theta}_A)} > 1$$

Sant En lägre varians ger en högre effektivitet

En liten varians betyder att vår skattning är relativt träffsäker

4

iii) $B(\hat{\theta}) = E(\hat{\theta}) - \theta$

Konsistens def. $\lim_{n\to\infty} V(\hat{\theta}_n) \to 0$ om $\hat{\theta}$ vvr

$MSE = V(\hat{\theta}) + [B(\hat{\theta})]^2$

Falskt: Vid stora n kommer $V(\hat{\theta})$ att gå mot 0

Därimot kommer medelkvadratsfelet aldrig bli 0 pga bias. Jo, om biasen går mot noll

1

iv) Falskt. R θ finns eller finns inte i intervallet
som beror på α $[\hat{\theta}_L, \hat{\theta}_U]$ inom klassisk inferens och vi kan därför inte se den som
som stokastiskt. Det är ett 0-1-förhållande.
Vi säger att "med $100(1-\alpha)\%$ konfidens finns
θ i intervallet" vilket betyder att om vi
upprepar skattningen många gånger kommer θ befinna
sig i intervallet $100(1-\alpha)\%$ av gångerna. 4
OK

v) Sant. R Inom bayesiansk inferens kan vi betrakta
θ som en stokastisk variabel. Vi talar då
om kredibilitetsintervall i vilka vi med $100(1-\alpha)\%$
sannolikhet kan säga att θ finns i intervallet
$[\hat{\theta}_L, \hat{\theta}_U]$ eller inte. 4

13

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: STM-0020 Blad nr: 2

(2) ~~antal risker som kan identifieras~~

Y: "Ny metod för att identifiera risk"

$H_0: p = 0,5$

$H_A: p < 0,5$

$n = 12$

$Y \sim Bin(n=12, p)$ R

a) ← förkastelseområde

$P(Y \leq 3 \mid H_0 \text{ sann}) = [\text{läst ur tabell}] = 0,073$ R

Svar: testets signifikansnivå är 0,073 6

Hade vi satt en gräns på $\alpha = 0,05$ hade vi alltså ej ~~förkastat~~ H_0 eller sagt att den nya metoden var dålig. ej p-värde

b) Styrka $= 1 - \beta$

$\beta = P(\text{Behålla } H_0 \mid H_A \text{ sann}) = P(\text{Behålla } H_0 \mid p = 0,3)$ R

$= P(Y > 3 \mid p = 0,3) = 1 - P(Y \leq 3 \mid p = 0,3) = 1 - 0,49252 = 0,50748$

$1 - \beta = 1 - 0,50748 = 0,49252$ R

Svar: Styrkan är 0,49252 givet $H_A: p = 0,3$ 7

c) Använder vi samma förkastelseområde men tillämpar metoden på fler fall kommer signifikansnivån minska och styrkan öka. Detta eftersom vi blir mer säkra på vår skattning med större urval.

Ex $n = 16$

$P(Y \leq 3) = 0,01758$

$1 - \underbrace{\beta = 1 - P(Y \leq 3 \mid p = 0,3)}_{= \beta} = 1 - 0,29687 = 0,70313$ 4

(17)

3

X: "BMI hos patient med inflammatorisk sjukdom"

Y: "BMI hos kontrollgrupp"

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

a) Teckentest Förutsättn.?

$H_0: p = 0,5$

$H_1: p \neq 0,5$ R

Testvariabel M = antal − = 6 $M \sim Bin(n=14, p)$ R

tvåsidigt test

$RR = \{M \leq 2,\ M \geq 11\}$ (12) ger $\alpha = 0,01294$ enligt tabell. Se sida 3 för slutsats!

SU, STATISTIK

Skrivsal: Brunnsviksalen Anonymkod: STM-0020 Blad nr: 3

3) c) forts

Eftersom M befinner sig inom intervallet för acceptansområdet $(2 < M < 11)$ kan vi ej förkasta H_0. Det verkar inte vara någon skillnad mellan patienternas och kontrollgruppens BMI på $\alpha = 0{,}01294$-nivån.

8

b) Wilcoxons teckenrangtest

Förutsättn.?

H_0: population X är lika med population Y i läge.

H_A: population X skiljer sig från population Y i läge. Population X ligger till höger om population Y. ?

Testvariabel $= \min(T^-, T^+) = T_{obs}$ R

$\alpha = 0{,}05$

$RO = \{T_{obs} \leq 21\}$

$T^+ = 71$

$T^- = 34$

$T^- = T_{obs}$ R

Vi kan ej förkasta H_0 på $\alpha = 0{,}05$ nivån. Det verkar inte vara någon skillnad mellan patienternas och kontrollgruppens BMI.

7

15

SU, STATISTIK

Skrivsal: Brunnvikssalen Anonymkod: STM-0020 Blad nr: 4

Obs alla $\sum = \sum_{i=1}^{n}$ och $\prod = \prod_{i=1}^{n}$

(4) $Y \sim Beta(\beta = 1, \alpha)$

$f(y) = \alpha y^{\alpha - 1} \quad 0 < y < 1 \quad \alpha > 1$

a) $\mu_1' = E(Y_i) = \frac{\alpha}{\alpha + 1}$ [enligt formelblad] R

$m_1' = \frac{\sum y_i}{n} = \bar{y}$ R

$\underline{m_1' = \mu_1'}$ R

$\bar{y} = \frac{\alpha}{\alpha + 1}$

$\frac{1}{\bar{y}} = \frac{\alpha + 1}{\alpha}$

$\frac{1}{\bar{y}} = \frac{\alpha}{\alpha} + \frac{1}{\alpha}$

$\frac{1}{\bar{y}} = 1 + \frac{1}{\alpha}$

$\frac{1}{\bar{y}} - 1 = \frac{1}{\alpha}$

$\frac{1}{\bar{y}} - \frac{\bar{y}}{\bar{y}} = \frac{1}{\alpha}$

$\frac{1 - \bar{y}}{\bar{y}} = \frac{1}{\alpha}$

$\frac{\bar{y}}{1 - \bar{y}} = \hat{\alpha}_{MOM}$

5

b) $f(y) = \alpha y^{\alpha - 1}$

- $L(\alpha) = f(y|\alpha) \overset{\text{pga ober.}}{=} \prod f(y_i) = \prod \left[\alpha y_i^{\alpha - 1} \right] = \alpha^n \prod y_i^{\alpha - 1}$ R
- $l(\alpha) = n \ln(\alpha) + (\alpha - 1) \sum \ln(y_i)$ R
- $l'(\alpha) = \frac{n}{\alpha} + \sum \ln(y_i)$ R
- $0 = \frac{n}{\alpha} + \sum \ln(y_i)$

$\frac{n}{\alpha} = -\sum \ln(y_i)$

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

$\hat{\alpha}_{ML} = \frac{-n}{\sum \ln(y_i)}$

10

c) $\widehat{E(Y)}_{ML} = \left(E(\hat{Y}_{ML})\right) = \dfrac{\dfrac{n}{\sum \ln(y_i)}}{\dfrac{n}{\sum \ln(y_i)} - 1}$

en maximumlikelihood-skattning av en funktion av θ blir funktionen substituerat med vår bästa skattning av θ vilket i det här fallet är vår bästa skattning av $\hat{\alpha}_{ML} = \dfrac{-n}{\sum \ln(y_i)}$

5

20

5)

$Y \sim Beta(\beta = 1, \alpha = 1)$

$H_0: \alpha = 1$

$H_1: \alpha \neq 1$

$f(y) = \alpha y^{\alpha - 1}$

a) $L(\alpha = 1) = f(y|\alpha = 1) \overset{\text{pga ober}}{=} \prod f(y_i) = \prod \left[\alpha y_i^{\alpha - 1}\right] = \alpha^n \prod y_i^{\alpha - 1}$

$= 1^n \prod y_i^0 = 1 \cdot 1 = 1$

2

Svar: $L(\alpha = 1) = 1$ R

b) $\hat{\alpha}_{ML} = \dfrac{-n}{\sum \ln(y_i)}$ från uppgif 4

$L(\alpha = \hat{\alpha}_{ML}) = \dfrac{-n}{\sum \ln(y_i)} \prod y_i^{\frac{-n}{\sum \ln(y_i)} - 1}$ R

2

5 forts

c)

$$\hat{\alpha}_{LR} = \frac{\max_{\theta \in \Omega_0} L(\theta) \quad MAX\, L(\alpha_0)}{MAX\, L(\alpha_A) \quad \max_{\theta \in \Omega} L(\theta)} < k$$

$$\frac{MAX\, L(\alpha_0)}{MAX\, L(\alpha_A)} = \frac{L(\alpha = 1)}{L(\alpha \neq 1)} = \frac{L(1)}{L(\hat{\alpha}_{ML})}$$

$$\hat{\alpha}_{LR} = \frac{1}{\frac{-n}{\sum \ln(y_i)} \prod y_i^{\frac{-n}{\sum \ln y_i} - 1}} < k$$

$$\hat{\alpha}_{LR} = \frac{-n}{\sum \ln(y_i)} \prod y_i^{\frac{-n}{\sum \ln(y_i)} - 1} > k$$

4

d) betecknar $\frac{-n}{\sum \ln(y_i)} = \hat{\alpha}_{ML}$ för husfridens skull ☺ (behövs tydligen inte mer)

$$-2\ln(\hat{\alpha}_{LR}) \overset{\text{approx.}}{\sim} \chi^2(1) = 7{,}88 \text{ för } \alpha = 0{,}05 \quad \checkmark \quad \alpha = 0.005$$

Beslutsregel: $-2\ln(\hat{\alpha}_{LR}) > -2\ln(k) = 7{,}88$

$-2\ln(\alpha_{LR}) < 7{,}88$ i vårt fall eftersom vi vände på uttrycket i slutet av c)

$$-2\ln\left[\frac{-n}{\sum \ln(y_i)} \prod y_i^{\frac{-n}{\sum \ln(y_i)} - 1}\right] = 2\ln(n) + 2\ln\sum \ln(y_i) - 2\left[\frac{n}{\sum \ln(y_i)} - 1\right]\sum \ln(y_i)$$

6

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 17/3-2016

Sal: Brunnsvikssalen

Tenta: Statistisk teori med tillämpningar 2

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: STM-0006

☒ 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					
Lär.ant. 15	17	15	15	16.					6 7

78 + 8 bonus

POÄNG	BETYG	Lärarens sign.
86	B	[illegible]

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: STM-0006 Blad nr: 1

1) i) Motivering: CGS kan tillämpas om $n \geq 30$. Formen för urvalets fördelning närmar sig formen av en standardiserad normal fördelning när n ökar. Om $n \geq 30$ säger vi att den är $Z \sim approx\ N(0,1)$

Svar: Sant ✓ 0

ii) Motivering: $eff(\hat{\theta}_A, \hat{\theta}_B) = \frac{V(\hat{\theta}_B)}{V(\hat{\theta}_A)}$

Om eff > 1 så är $\hat{\theta}_A$ en bättre estimator för θ

Om eff < 1 så är $\hat{\theta}_B$ en bättre estimator för θ

Om $V(\hat{\theta}_A)$ är mindre än $V(\hat{\theta}_B)$

Ex $V(\hat{\theta}_A) = 1$

$V(\hat{\theta}_B) = 5$

så blir $eff(\hat{\theta}_A, \hat{\theta}_B) = \frac{5}{1} = 5$

därför anses $\hat{\theta}_A$ vara en mer effektiv estimator än $\hat{\theta}_B$ för θ. Detta på grund av att eff > 1

Svar: Sant R 4

iii) Motivering: En estimator är konsistent när $\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$. Dvs sannolikheten att en estimator kommer godtyckligt nära det sanna värdet av parametern ökar och går mot 1 när n ökar. En unbiased estimator kan vara konsistent men det krävs då att n är väldigt stort för att estimatorn ska komma godtyckligt nära det sanna värdet på parametern. Svar: falskt R 3

biasen är mot 0

iv) Inom klassisk inferens är θ en parameter.

Det kan t.ex vara μ (genomsnittslängden t.ex)

Svar: Falskt (stok. variabler inom klassisk inferens skrivs ofta som "Y" "X" "W") R 4

v) Ja θ är en stokastisk variabel inom bayesiansk inferens. R 4

Från Hans Nyquists föreläsning definierade han.

θ_1 = lågtryck

θ_2 = högtryck

Svar: Sant

15

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: STM-0006 Blad nr: 2

2) Y: "antalet fall av 12 risker som man kan identifiera som risker"

$Y \sim bin(n=12, p=0,5)$ R

Förkasta H_0 om tre eller färre risker kan identifieras dvs. $Y \leq 3$

a) Bestäm testets signifikansnivå

$\alpha = P(\text{Typ-I fel}) = P(\text{Förkasta } H_0 \mid H_0 \text{ är sann}) = P(Y \leq 3 \mid p=0,5\}$

$= \{\text{ur binomialtabell } n=12 \; p=0,5 \; y \leq 3\} = 0,07300$ R

Svar: testets signifikansnivå är 0,073

7

b) Testets styrka [power(θ)] då $p=0,3$

Testets styrka kan beräknas genom $1-\beta$ R

$\beta = P(\text{Typ-II fel}) = P(\text{Ej förkasta } H_0 \mid H_A \text{ är sann}) =$ R

$= P(Y \geq 4 \mid p=0,3) = 1 - P(Y \leq 3 \mid p=0,3) = \{\text{ur binomialtabell}\}$

$= 1 - 0,49252 = 0,50748$

Svar: Testets styrka är: $1 - 0,50748 = 0,49252$ R

7

och testets styrka

c) Vad händer med α / och testets styrka då $p=0,3$ om n ökar.

Om man behåller samma förkastelseområde dvs

Förkasta H_0 om $Y \leq 3$.

$\underline{n=13}$

$\alpha = P(Y \leq 3 \mid p=0,3) = 0,42061$

$p=0.5$

$\underline{n=14}$

$\alpha = P(Y \leq 3 \mid p=0,3) = 0,35517$

0.5

$\underline{n=16}$

$\alpha = P(Y \leq 3 \mid p=0,3) = 0,29687$

0.5

$\vdots$

$\underline{n=19}$

$\alpha = P(Y \leq 3 \mid p=0,3) = 0,13317$

Om man jämför dessa värden mot $P(Y \leq 3 \mid p=0,3)$ då $n=12$ vilket blir $\alpha = 0,49252$ så kan vi se att när n ökar så kommer signifikansnivån α att minska för varje n.

Men styrkan däremot $(1-\beta)$ kommer att stiga för varje n. Vi ser att $1 - 0,49252 = 0,50748 = \beta$ och $1 - 0,42061 = 0,57939, \ldots \; 1 - 0,13317 = 0,86683$

Svar: signifikansnivån minskar och styrkan ökar

3

17

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** STM-0006 **Blad nr:** 3

3)

Patient (X)	Kontroll (Y)	Differens $(x_i - y_i)$	Tecken	Rang
25,41	26,93	-1,52	-	4
26,15	29,04	-2,89	-	5
35,15	26,08	9,07	+	13
30,60	23,62	6,98	+	12
27,00	21,11	5,89	+	10
23,20	27,97	-4,77	-	8
27,70	28,09	-0,39	-	~~3~~ 2
29,00	24,31	4,69	+	7
35,98	29,07	6,91	+	11
26,30	35,91	-9,61	-	14
31,00	27,47	3,53	+	6
31,31	26,10	5,21	+	9
28,03	28,22	-0,19	-	1
28,41	27,61	0,8	+	~~2~~ 3

a) Tecken test (parvisa observationer)

Antaganden: oberoende mellan paren, beroende mellan patient och kontroll. Fördelningen är okänd

Bra

$H_0: p = 0,5$

$H_1: p \neq 0,5$

R

$M \sim bin(n = 14, p = 0,5)$ R

$m = 8+$

$n = 14$

$n_1 = n_2 = 14$ ($n_1 > 10$ $n_2 > 10$ så vi kan $Z \sim$ approx $N(0,1)$)

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

ok

Förkasta H_0 om $|Z_{obs}| > 1,96$ på 5 % signifikansnivå

Förkasta H_0 om $|Z_{obs}| > 2,5758$ på 1 % signifikansnivå

$$Z_{obs} = \frac{8 - (14/2)}{(1/2)\sqrt{14}} \approx 0,53$$

Resultat: 0,53 < 1,96

0,53 < 2,5758

Slutsats: Det finns ej tillräckligt med stöd att förkasta H_0 på signifikansnivån $\alpha = 0,05$ eller $\alpha = 0,01$

Det finns statistiskt stöd för att påstå att BMI för patienter inte skiljer sig åt från BMI för kontroller.

~~vi kan ej påvisa att de skiljer sig men~~ inte bevisa att de inte skiljer sig 10

b) Antaganden: se a) Det är mellan parvisa observationer så samma antaganden, men här antas även att det är oberoende mellan differenserna. ?

Willcoxons tecken rangtest (parvisa observationer)

$T = \min(T^-, T^+)$

$T^+ = 13+12+10+7+11+6+9+2 = 70$ (rangsumman för de positiva differenserna)

~~Även här blir det Z ~ approx N(0,1) eftersom~~ $\begin{cases} n_1 > 10 \\ n_2 > 10 \end{cases}$ $n > 25$ krävs

H_0: BMI för patienter skiljer sig inte åt från BMI för kontroller

H_A: BMI för patienter skiljer sig åt från BMI för kontroller

Förkasta H_0 om $|Z_{obs}| > 1,96$ på signifikansnivån $\alpha = 0,05$

Förkasta H_0 om $|Z_{obs}| > 2,5758$ på signifikansnivån $\alpha = 0,01$

$$Z = \frac{T^+ - [n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}$$

Här är n=14 för litet för normalapprox.

$$Z_{obs} = \frac{70 - [14(14+1)/4]}{\sqrt{14(14+1)(2\cdot 14+1)/24}} = \frac{17,5}{\sqrt{280}} \approx 1,046$$

Resultat: $-1,046 < 1,96$

$1,046 < 2,5758$

Se slutsats nästa sida

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** STM-0006 **Blad nr:** 4

3b) fortsättning

Slutsats: Det finns ej tillräckligt med stöd att förkasta H_0 på signifikansnivå $\alpha = 0{,}05$ och $\alpha = 0{,}01$

Det finns statistiskt stöd för att påstå att BMI för patienter inte skiljer sig åt från BMI för kontroller.

5

15

SU, STATISTIK

Skrivsal: Brunnsviksalen Anonymkod: STM-0006 Blad nr: 5

4) $f(y) = \alpha y^{\alpha-1}, \quad 0 < y < 1 \quad \alpha > 0$

a) $\mu_1' = E(Y) = \int_0^1 y(\alpha y^{\alpha-1})\,dy = \left[\frac{\alpha y^{\alpha+1}}{\alpha+1}\right]_{y=0}^{y=1} = \frac{\alpha}{1-\alpha}$

$m_1' = \frac{\sum y_i}{n} = \bar{Y}$ R

$\mu_1' = m_1' \Rightarrow \frac{\alpha}{1-\alpha} = \bar{Y}$

$\alpha = \bar{Y}(1-\alpha)$

$\alpha = \bar{Y} - \bar{Y}\alpha \Rightarrow \alpha = \frac{\bar{Y}}{1-\bar{Y}}$

Svar $\hat{\alpha}_{MoM} = \frac{\bar{Y}}{1-\bar{Y}}$ 3

b) $L(\alpha) = \prod_{i=1}^{n} \alpha y^{\alpha-1} = \alpha^n \prod \sum y_i^{\alpha-1}$ ✓

pga oberoende

$\ell(\alpha) = \ln[L(\alpha)] = n\ln\alpha + \sum \ln y_i^{\alpha-1} = n\ln\alpha + (\alpha-1)\sum \ln y_i$

Deriverar m.a.p α och sätter derivatan lika med noll

$\ell'(\alpha) = \frac{n}{\alpha} + \sum \ln(y_i) = 0$

$\frac{n}{\alpha} = -\sum \ln(y_i)$

Svar: $\hat{\alpha}_{ML} = -\frac{n}{\sum \ln(y_i)}$

7

c) Vad blir maximum likelihood-skattningen av $E(Y)$?

$$E(Y) = \frac{\alpha}{1-\alpha}$$

Maximum likelihood skattningen av $E(Y)$ blir:

$$\frac{\hat{\alpha}_{ML}}{1-\hat{\alpha}_{ML}} = \frac{\left(-\frac{n}{\sum \ln y_i}\right)}{1-\left(-\frac{n}{\sum \ln y_i}\right)} =$$

$$= \frac{\left(-\frac{n}{\sum \ln y_i}\right)}{\left(1+\frac{n}{\sum \ln y_i}\right)}$$

ok, följdfel

$\overline{5}$

15

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: STM-0006 Blad nr: 6

5)

$H_0: \alpha = 1$

$H_a: \alpha \neq 1$

$\hat{\alpha}_{ML} = -\dfrac{n}{\sum \ln(y_i)}$

$f(y) = \alpha y^{\alpha - 1}, \quad 0 < y < 1, \quad \alpha > 0$

$L(\alpha) = \prod_{i=1}^{n} \alpha y^{\alpha-1} = \alpha^n \sum y_i^{\alpha-1}$

pga. ober. !

a) $L(\alpha = 1) = 1^n \cdot \sum y_i^{1-1} = 1 \cdot \sum y_i^{(1-1)} = \sum y_i^0 \neq 1$

$= n$ i så fall

1

b) $L(\alpha = \hat{\alpha}_{ML}) = \left(-\dfrac{n}{\sum \ln y_i}\right)^n \cdot \sum y_i^{\left(-\frac{n}{\sum \ln y_i}\right) - 1}$

ok, följdfel

2

c) Teststatistikan för likelihood kvottestet blir:

$\lambda_{LR} = \dfrac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \dfrac{L(\alpha = 1)}{L(\alpha = \hat{\alpha}_{ML})} = \dfrac{L(1)}{L(\hat{\alpha}_{ML})}$

$\lambda_{LR} = \boxed{\dfrac{1}{\left(-\dfrac{n}{\sum \ln y_i}\right)^n \cdot \sum y_i^{\left(-\frac{n}{\sum \ln y_i}\right) - 1}}}$

ok, följdfel

6

d) Kritiskt område (RR) för signivån 0,05 när n är stort:

$$RR = \left\{ -2\ln\lambda_{LR} > \chi^2(n-1) \right\}$$ ✓

där antalet frihetsgrader i χ^2 fördelningen bestäms av antalet parametrar i nollhypotesen, dvs 1 parameter.

$\chi^2_{0,05}(1) = 3,84$ R

$$RR = \left\{ -2\ln\left[\frac{1}{\left(-\frac{n}{\sum\ln y_i}\right)^n \sum y_i^{\left(-\frac{n}{\sum\ln y_i}\right)-1}}\right] > 3,84 \right\}$$

ok, följdfel

Vi förkastar H_0 dvs $\alpha=1$ på 5% signifikansnivå, om $-2\ln\lambda_{LR} > 3,84$. Om vi förkastar H_0 innebär det att $\alpha \neq 1$ och vi har ej en likformig fördelning på intervallet (0,1). Om vi inte förkastar H_0 innebär det att $\alpha = 1$ och vi har en likformig fördelning på intervallet (0,1).

8

16