STOCKHOLMS UNIVERSITET
Statistiska institutionen
Jörgen Säve-Söderbergh

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2019-11-25

Skrivtid: 12.00-17.00
Godkända hjälpmedel: Miniräknare, språklexikon.
Tentamen består av fem uppgifter. För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade lösningar.

OBS! Glöm inte att ange nödvändiga antaganden där det behövs.

Uppgift 1. (20 poäng)
Förklara innebörden av följande begrepp:

a) Konsistens

b) Neyman-Pearsons lemma

c) Centrala gränsvärdessatsen

d) Samplingfördelning

e) Styrka

Uppgift 2. (20 poäng)
En forskare utvärderade tolv arbetstagares arbetsförmåga vid två tillfällen. Utvärderingen resulterade i poäng som anses befinna sig på den ordinala skalan. Det gjordes två utvärderingar vid två olika tillfällen över tid. De resulterade i följande värden

Arbetstagare	1	2	3	4	5	6	7	8	9	10	11	12
Första tillfället	45	23	78	17	73	66	38	83	92	35	27	76
Andra tillfället	49	24	68	23	78	66	19	89	97	32	28	70

a) Beräkna Spearmans rangkorrelationskoefficient r_S för arbetstagarnas arbetsförmåga vid de två tillfällena.

b) Vilken rangordning åstadkommer ett värde på ett för Spearmans rangkorrelationskoefficient r_S? Vad skulle ett sådant värde betyda i sammanhanget?

c) Visar data på en övertygande samstämmighet mellan de bägge bedömningarna av arbetstagarnas arbetsförmåga? Testa på signifikansnivån 5% att det inte finns någon sådan samstämmighet, d v s testa $H_0 : \rho_S = 0$. Använd ett ensidigt test.

Uppgift 3. (20 poäng)
Tolv tvillingpar genomgick ett psykologiskt test för att fastställa om den förstfödde av tvillingarna uppvisar ett mer aggressivt beteende än den senare födde. Resultaten var som följer, där ett högre värde betyder större aggressivitet.

Tvillingpar	Förstfödd	Född som andra barn
1	86	88
2	71	77
3	77	76
4	68	64
5	91	96
6	72	72
7	77	65
8	91	90
9	70	65
10	71	80
11	88	81
12	87	72

a) Sätt upp lämpliga hypoteser och gör nödvändiga antaganden för att testa om den förstfödde av tvillingarna uppvisar ett mer aggressivt beteende än den senare födde.

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

Uppgift 4. (20 poäng)
För att beskriva andelen av ett område som har en viss förorening över ett visst gränsvärde kan modellen

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

användas. Här är y andelen förorenad mark och β är en parameter. Denna modell är en betafördelning med $\alpha = 1$. (Betafördelningen med både α och β som parametrar används bl.a. av Naturvårdsverket för att beskriva andelen förorenad mark.) Antag nu att man har n oberoende observationer på Y.

a) Härled momentestimatorn av β.

b) Härled maximum likelihoodestimatorn av β.

Uppgift 5. (20 poäng) Antag att antalet tentamina en student på kursen «Statistisk teori med tillämpningar» behöver göra är geometriskt fördelad med parametern p och att man observerar antal nödvändiga tentamina som n slumpmässigt utvalda studenter behöver göra för att bli godkända på kursen.

a) Bestäm likehoodfunktionen. (2 poäng)

b) Bestäm maximumlikelihoodestimatorn $\hat{p}$ av p. (3 poäng)

c) Bestäm likelihoodfunktionens värde för $\hat{p}$. (2 poäng)

d) Bestäm likelihoodfunktionens värde för $p = 0.5$. (1 poäng)

e) Bestäm teststatistikan för likelihoodkvottestet för att testa $H_0 : p = 0.5$ mot $H_a : p \neq 0.5$. (6 poäng)

f) Ange kritiskt område för testet i uppgift (e) om ett stort antal observationer kan göras så att asymptotisk fördelning kan användas. Använd signifikansnivån 0.05. (6 poäng)

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 25/11-2019

Sal: Värtasalen

Tenta: Statistisk teori med tillämpningar

Kurs: Statistisk teori med tillämpningar 2

ANONYMKOD: 0025-FLL

0025

☒ 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					4
Lär.ant. 16	20	19	20	12					

HL

+0 bonusp.

POÄNG	BETYG	Lärarens sign.
87	B	JSS

SU, STATISTIK

Skrivsal: Vårtasalen Anonymkod: 0025-FLL Blad nr: 1

1 a) En estimator $\hat{\theta}$ av en parameter θ är konsistent om $\lim_{n\to\infty} P(|\hat{\theta}-\theta| \le \varepsilon) = 1$ för alla $\varepsilon > 0$.

4

b) Om man har två enkla hypoteser $H_0: \hat{\theta}_1 = \theta$ och $H_a: \hat{\theta}_2 = \theta$ så är likelihoodkvottestet $\frac{L(\hat{\theta}_1)}{L(\hat{\theta}_2)} < k$ det starkaste testet, där k bestäms av α.

4

c) Låt $Y_1, \ldots, Y_n$ vara stickprov från en population med väntevärde μ. Då kommer $Z = \frac{\bar{Y}_n - \mu}{\frac{S}{\sqrt{n}}} \sim$ approx $N(0,1)$ då n ökar.

4

d) Samplingfördelning är fördelningen som urvalet följer.

Nej

Om $X_1, \ldots, X_n$ är oberoende, där $X_i \sim N(\mu, \sigma^2)$ så dras stickprovet från $N(\mu, \sigma^2)$. Men t.ex. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$

punktskattare

Samplingfördelningen

e) Styrkan av ett hypotestest är sannolikheten att förkasta nollhypotesen under nollhypotesen, och betecknas som $\text{power}(\theta_a)$. Alternativt, $\text{power}(\theta_a) = 1 - \beta(\theta_a)$, där $\beta(\theta_a)$ är sannolikheten för ett typ-II fel.

4

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0025-FLL Blad nr: 2

2 a)

Arbetstagare	$R(x_i)$	$R(y_i)$	d_i
1	6	6	0
2	2	3	-1
3	10	8	2
4	1	2	-1
5	8	10	-2
6	7	7	0
7	5	1	4
8	11	11	0
9	12	12	0
10	4	5	-1
11	3	4	-1
12	9	9	0

Tabellen visar arbetstagarnas rang vid det första utvärderingen $(R(x_i))$, rang vid andra utvärderingen $(R(y_i))$ samt rangskillnaden mellan utvärderingarna (d_i).

$$r_s = 1 - \frac{6\sum_{i=1}^{12} d_i^2}{n(n^2-1)}$$ används för att beräkna rangkorrelationskoefficienten.

$$\Rightarrow \quad r_s = \frac{129}{143} \approx 0{,}9021$$

b) Om $r_s = 1$ innebär det att

$$1 = r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}.$$ Om likhet ska gälla

måste $\frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 0$, d.v.s $d_i = 0$ för alla $i = 1, 2 \ldots n$.

Detta händer endast om rangen är densamma, och i detta fall skulle det innebära att arbetstagarnas poängmässiga ordning är oförändrad mellan utvärderingarna.

V.G.V. →

c) $H_0: \rho_s = 0$

$H_a: \rho_s > 0$

$\alpha = 0.05 \Rightarrow r_k = 0,497.$ R

Vi förkastar H_0 om $r_s > r_k$.

Från uppgift a har vi att $r_s \approx 0,9021$, alltså förkastar vi H_0. Vi kan gå vidare och säga att vi kan förkasta H_0 på alla rimliga signifikansnivåer. R

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0025-FLL Blad nr: 3

3 a) H_0: tvillingarna visar ingen skillnad i aggresivitet $(\mu_1 - \mu_2 = 0)$

H_a: den förstfödde uppvisar mer aggresivt beteende $(\mu_1 > \mu_2)$

Vi antar oberoende mellan paren.

b) Vi väljer att utföra ett teckentest.

Om T = antalet förstfödda som visar högre aggresivitet, så är

$T \sim Bin(n=11, p)$ med $p = 0,5$ under H_0.

Vi har $T^* = 7$.

För att ta reda på om vi kan förkasta H_0 beräknar vi p-värdet, dvs;

Borde inte p-värde = $P(T^* \geq 7)$

$P(T > 6) = 1 - P(T \leq 6) = 1 - 0,72559 = 0,27441$.

-1p

Vi kan alltså inte förkasta H_0 på några rimliga nivåer.

R

V.G.V →

4 a) $\frac{1}{n}\sum_{i=1}^{n} y_i = \int_0^1 y \cdot \beta (1-y)^{\beta-1}\, dy$

$$= \left[-(1-y)^{\beta} \cdot y\right]_0^1 - \int_0^1 -(1-y)^{\beta}\, dy$$

$$= -0 - 0 - \left[\frac{(1-y)^{\beta+1}}{\beta+1}\right]_0^1 = \frac{1}{\beta+1}$$

$$\Rightarrow \beta = \frac{1}{\bar{y}} - 1$$

R

b) $L(\beta) = \prod_{i=1}^{n} \beta (1-y_i)^{\beta-1} = \beta^n \prod_{i=1}^{n} (1-y_i)^{\beta-1}$

$$\ln[L(\beta)] = n \times \ln(\beta) + (\beta-1) \ln\left(\prod_{i=1}^{n} (1-y_i)\right)$$

$$\frac{d \ln[L(\beta)]}{d\beta} = \frac{n}{\beta} + \ln\left(\prod_{i=1}^{n} (1-y_i)\right)$$

sätt lika med 0:

$$\frac{n}{\hat{\beta}} + \ln\left(\prod_{i=1}^{n} (1-y_i)\right) = 0$$

$$\Rightarrow \hat{\beta} = \frac{-n}{\ln\left(\prod_{i=1}^{n} (1-y_i)\right)}$$

R

SU, STATISTIK

Skrivsal: Värtasalen Anonymkod: 0025-FLL Blad nr: 4

5 a) $p(y) = p(1-p)^{y-1}$

$L(p) = \prod_{i=1}^{n} p(1-p)^{y_i-1} = p^n \times (1-p)^{\sum_{i=1}^{n} y_i - n}$ R

b) $\ln[L(p)] = n\ln(p) + \left(\sum_{i=1}^{n} y_i - n\right) \times \ln(1-p)$

$\frac{d\ln[L(p)]}{dp} = \frac{n}{p} - \frac{\sum_{i=1}^{n} y_i - n}{1-p}$

Sätt lika med 0:

$\frac{n}{\hat{p}} = \frac{\sum_{i=1}^{n} y_i - n}{1-\hat{p}}$

$\Rightarrow \hat{p} = \frac{n}{\sum_{i=1}^{n} y_i} = \frac{1}{\bar{y}}$ R

c) $L(\hat{p}) = \hat{p}^n \times (1-\hat{p})^{\sum_{i=1}^{n} y_i - n} = \left(\frac{1}{\bar{y}}\right)^n \times \left(1 - \frac{1}{\bar{y}}\right)^{\sum_{i=1}^{n} y_i - n}$ R

d) $L(p=0.5) = 0.5^n \times (1-0.5)^{\sum_{i=1}^{n} y_i - n}$

$= 0.5^n \times 0.5^{\sum_{i=1}^{n} y_i} \times 0.5^{-n} = 0.5^{\sum_{i=1}^{n} y_i}$ R

(e nästa sida)

f) $\frac{L(p=0.5)}{L(\hat{p})} = \frac{0.5^{\sum_{i=1}^{n} y_i}}{\hat{p}^n \times (1-\hat{p})^{\sum_{i=1}^{n} y_i - n}} < k$

$\ln\left(\frac{L(p=0.5)}{L(\hat{p})}\right) = \sum_{i=1}^{n} y_i \times \ln(0.5) - n \times \ln(\hat{p}) - \left(\sum_{i=1}^{n} y_i - n\right) \times \ln(1-\hat{p})$

$= \ln(0.5) \times \sum_{i=1}^{n} y_i + n \times (\ln(1-\hat{p}) - \ln(\hat{p})) - \ln(1-\hat{p}) \times \sum_{i=1}^{n} y_i$

$< \ln(k)$

Du har ett uttryck för $L(\hat{p})$ ovan som måste användas här

V.G.V

$$\left(\ln(0.5) - \ln(1-\hat{p})\right) \times \sum_{i=1}^{n} y_i + n \times \left(\ln(1-\hat{p}) - \ln(\hat{p})\right) < \ln k$$

$$\sum_{i=1}^{n} y_i < \frac{\ln(k) - n \times \left(\ln(1-\hat{p}) - \ln(\hat{p})\right)}{\ln(0.5) - \ln(1-\hat{p})}$$

$$\bar{y} < \frac{\frac{\ln(k)}{n} - \ln(1-\hat{p}) - \ln(\hat{p})}{\ln(0.5) - \ln(1-\hat{p})} = k^*$$

Du är inte klar! Du kan inte ha $\hat{p}$ med i höger ledet

Vi har att k^* är vårt kritiska värde som är beroende på n.

Förkasta H_0 om $\bar{y} > k^*$, Med $\alpha = 0{,}05$, är $k^* = 3{,}84$.

e) Vi har test statistikan $\ln\left(\frac{L(0.5)}{L(\hat{p})}\right)$

som är χ^2-fördelad med 1 frihetsgrad. Nej,

det (krävs) ~~en~~ multiplikation med -2

$8p + 4p = 12p$