STOCKHOLMS UNIVERSITET
Statistiska institutionen
Hans Nyquist

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2023-04-21
LÖSNINGAR

Uppgift 1.(20 poäng)

En kommun genomförde en enkätundersökning för att bland annat uppskatta andelen kommuninvånare som vill att kommunen anlägger fler cykelbanor. Av totalt 472 respondenter var 127 stycken positiva till att anlägga fler cykelbanor.

a) Beräkna ett 99%-igt konfidensintervall för andelen kommuninvånare som är för att anlägga fler cykelbanor.

b) Beräkna hur stort urval som skulle krävas för att längden på ett 99%-igt konfidensintervall inte ska överstiga 0.08. Antag vidberäkningen att den sanna andelen är okänd.

c) Hur påverkas erforderlig urvalsstorlek i b)-uppgiften om den statistiska felmarginalen minskas?

d) Hur påverkas erforderlig urvalsstorlek i b)-uppgiften om konfidensgraden minskas?

Lösning:
a) En uppskattning av proportionen är $\widehat{p} = \frac{127}{472} = 0.2691$. Variansen uppskattas därför till $\frac{0.2691(1-0.2691)}{472} = 0.00041671$. Ett 99%-igt kfi får därför som $0.2691 \pm 2.58 \cdot \sqrt{0.00041671}$, dvs. intervallet blir $(0.2164;\ 0.3218)$.

b) Längden på ett 99%-igt kfi är $2 \cdot 2.58 \cdot \sigma_p$, där $\sigma_p^2 = \frac{p(1-p)}{n}$ är variansen hos den uppskattade proportionen. Eftersom vi inte vet p använder vi det värde som ger den största variansen, vilket är $p = 0.5$. Vi får alltså

$$2 \cdot 2.58 \cdot \sqrt{\frac{0.5\,(1-0.5)}{n}} = 0.08$$

$$\frac{0.5}{\sqrt{n}} = \frac{0.08}{2 \cdot 2.58}$$

$$n = \left(\frac{2 \cdot 2.58 \cdot 0.5}{0.08}\right)^2 = 1040$$

c) Det krävs fler observationer om den statistiska felmarginalen minskas.

d) Det krävs färre observationer om konfidensgraden minskas.

Uppgift 2 (20 poäng)

I finalen av Melodifestivalen 2023 deltog 12 låtar. En internationell jury röstade på låtarna och deras poäng las ihop med poäng från tv-tittarnas röster för att kora den vinnande låten. I tabellen nedan visas resultaten. Testa med signifikansnivån 5% om poängen från den internationella juryn och tv-tittarna är positivt korrelerade.

Bidrag	Jurypoäng	Tittarpoäng
Where you are	23	58
Rythm of my Show	15	5
One day	35	16
Air	71	67
On my way	22	25
Never give up	10	37
Six feet under	53	59
Where did you go	37	39
Släpp alla sorger	8	35
Tattoo	92	85
Mer av dig	42	36
Royals	56	1

Lösning:

Observationerna på poängen har en ordinalskala. Vi använder Spearmans rangkorrelation. Vi vill pröva

$$H_0 : \rho_S = 0$$

mot

$$H_A : \rho_S > 0$$

på signifikansnivån $\alpha = 0.05$. Eftersom antalet observationer är $n = 12$ får vi förkastar vi H_0 om $RR = \{r_S \geq 0.497\}$.
Vi börjar med att rangordna poängen för juryn respektive tv-tittarna och beräknar differensen, d, av rangerna. Det spelar ingen roll om man ger rang 1 till den med högsta poängen eller till den med den lägsta poängen.

Bidrag	Rang för juryn	Rang för tittarna	d	d^2
Where you are	8	4	4	16
Rythm of my Show	10	11	-1	1
One day	7	10	-3	9
Air	2	2	0	0
On my way	9	9	0	0
Never give up	11	6	5	25
Six feet under	4	3	1	1
Where did you go	6	5	1	1
Släpp alla sorger	12	8	4	16
Tattoo	1	1	0	0
Mer av dig	5	7	-2	4
Royals	3	12	-9	81
Summa				154

Spearmans rangkorrelation ges nu av

$$r_S = 1 - \frac{6\sum d_i^2}{n\,(n^2-1)} = 1 - \frac{6 \cdot 154}{12\,(12^2-1)} = 1 - \frac{852}{1716} = 1 - 0.53846 = 0.46154$$

Vi ser att $r_S < 0.497$, dvs H_0 förkastas på signifikansnivån 5% och vi har inget empiriskt evidens för att poängen från den internationella juryn respektive från tv-tittarna samvarierar.

Uppgift 3 (20 poäng)

Statistiska institutionen har skaffat en ny kaffemaskin där man får ange den styrka av kaffet man önskar, Styrkan anges med att tal mellan 0 (mycket svagt kaffe) och 1 (mycket starkt kaffe). En övningslärare undersöker hur stor styrka som lektorerna väljer och antar att den valda styrkan kan modelleras med en betafördelning med $\beta = 1$, dvs. om Y är styrkan som en slumpmässigt vald lektor väljer så har Y täthetsfunktionen

$$f(y) = \left[\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(1)}\right] y^{\alpha-1} = \alpha y^{\alpha-1},\ 0 \leq y \leq 1.$$

Antag att det finns n oberoende observationer på Y.

a) Bestäm momentestimatorn av α.

b) Bestäm maximum likelihoodestimatorn av α.

Lösning:

a) För en betafördelning gäller det att $E(Y) = \frac{\alpha}{\alpha+\beta}$ och med $\beta = 1$ blir det $E(Y) = \frac{\alpha}{\alpha+1}$. Momentestimatorn $\widetilde{\alpha}$ av α definieras därför av

$$\frac{\widetilde{\alpha}}{\widetilde{\alpha}+1} = \overline{y},$$

där $\overline{y} = n^{-1}\sum_{i=1}^{n} y_i$ är medelvärdet av observationerna. Detta gör att

$$\widetilde{\alpha} = \frac{\overline{y}}{1-\overline{y}}.$$

b) För betafördelningen med $\beta = 1$ gäller att

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

så att likelihoodfunktionen är

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

Logaritmering ger

$$\ln L(\alpha) = n \ln \alpha + (\alpha - 1)\sum_{i=1}^{n} \ln y_i.$$

Derivering ger

$$\frac{d \ln L(\alpha)}{d\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \ln y_i.$$

Sätt derivatan lika med noll och lös ut $\widehat{\alpha}$

$$\frac{n}{\widehat{\alpha}} + \sum_{i=1}^{n} \ln y_i = 0$$

$$\widehat{\alpha} = -\frac{n}{\sum_{i=1}^{n} \ln y_i}$$

Uppgift 4 (20 poäng)

Antal studenter som söker upp övningsläraren under en jourtimme på kursen STMT antas vara Poissonfördelad med parametern λ. Studierektorn undrar om det kommer för få studenter till jourtimmarna och vill därför undersöka om det förväntade antalet besök är större än 1.

a) Hjälp studierektorn (som är en mycket upptagen person) att konstruera ett likformigt starkaste test av hypotesen $H_0 : \lambda = 1$ mot $H_A : \lambda > 1$.

b) Under tio jourtimmar observerades följande antal besök:

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

Antag att observationerna på antalet besök kan anses vara stokastiskt oberoende och ange p-värdet för testet i uppgift a.

Lösning:

a) Vi har ett slumpmässigt urval från $Po(\lambda)$ och börjar med att konstruera ett starkaste test av den enkla nollhypotesen $H_0 : \lambda = 1$ mot den enkla alternativhypotesen $H_A : \lambda = \lambda_A$ för något fixt men godtyckligt värde $\lambda_A > 1$. Likelihoodfunktionen för $Po(\lambda)$ är

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{y_i}}{y_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!} e^{-n\lambda}$$

Neyman-Pearsons lemma ger att H_0 förkastas då

$$\frac{L(\lambda = 1)}{L(\lambda = \lambda_A)} < k$$

för något tal k, men detta är ekvivalent med

$$\frac{\frac{1}{\prod_{i=1}^{n} y_i!} e^{-n}}{\frac{\lambda_A^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!} e^{-n\lambda_A}} = \frac{e^{n(\lambda_A - 1)}}{\lambda_A^{\sum_{i=1}^{n} y_i}} < k$$

Logaritmering ger

$$n(\lambda_A - 1) - \ln \lambda_A \sum_{i=1}^{n} y_i < \ln k$$

Eftersom $\lambda_A > 1$ är $\ln \lambda_A > 0$. Detta ger

$$\sum_{i=1}^{n} y_i > \frac{n(\lambda_A - 1) - \ln k}{\ln \lambda_A} = k'$$

för något tal k. H_0 förkastas alltså för stora värden på $\sum_{i=1}^{n} y_i$. Eftersom $\lambda_A > 1$ valdes godtyckligt är detta det likformigt starkaste testet.

b) Vi förkastar H_0 om $S = \sum_{i=1}^{n} y_i$ är stor. Eftersom observationerna är oberoende och Poissonfördelade är $S \sim Po\,(n\lambda)$. Om H_0 är sann är alltså $S \sim Po\,(10)$. Vi har att observerat värde på S är 16. Detta ger $p = P\,(S \geq 16) = 1 - P\,(S \leq 15) = 1 - 0.951 = 0.049$.

Uppgift 5 (20 poäng)

a) Vad menas med effektivitet hos en estimator?

b) Hur används Bayes sats i Bayesiansk inferens?

c) Hur tolkas ett kredibilitetsintervall?

d) Två dataanalytiker estimerar en parameter i en modell. Båda har samma observationer, men de får olika kredibilitetsintervall. Förklara varför de kan få olika resultat förutsatt att båda inte gör några matematiska fel.

e) Vad är a priorifördelning?