STOCKHOLMS UNIVERSITET
Statistiska institutionen
Ellinor Fackle-Fornius

TENTAMEN I STATISTISK TEORI MED TILLÄMPNINGAR II
2016-11-24

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 2016-12-12 kl. 15 i sal B413.

OBS! Glöm inte att ange nödvändiga antaganden överallt.

Uppgift 1. (20 poäng)

Förklara kortfattat relationen mellan följande termer:
a) Statistisk signifikans och praktisk signifikans
b) Signifikansnivå och p-värde
c) Typ I-fel och typ II-fel
d) Apriorifördelning och aposteriorifördelning

Uppgift 2. (20 poäng)

48 belopp avrundas vart och ett till heltal. Avrundningsfelen $Y_1, Y_2, \ldots, Y_{48}$ antas vara likformigt fördelade i intervallet $(-0.5, 0.5)$ samt oberoende.

a) Beräkna väntevärdet och variansen för medelavrundningsfelet.

b) Vad är sannolikheten att medelavrundningsfelet hamnar i intervallet $(-0.1, 0.1)$?

c) Beräkna väntevärdet och variansen för det totala avrundningsfelet.

d) Vad är sannolikheten att det totala avrundningsfelet hamnar i intervallet $(-6, 6)$?

Uppgift 3. (20 poäng)

Vilopulsen uppmättes hos 11 slumpmässigt utvalda män och 11 slumpmässigt kvinnor. I följande tabell visas vilopulsen som antal slag/minut.

Män	Kvinnor
60	78
74	80
86	68
54	56
90	76
80	78
66	60
68	96
68	62
56	60
80	98

a) Använd ett lämpligt icke-parametriskt test för att testa om det är någon skillnad i läge mellan fördelningarna av vilopuls för män och för kvinnor.

b) Använd ett lämpligt parametriskt test för att testa om det är någon skillnad i läge mellan fördelningarna av vilopuls för män och för kvinnor.

Uppgift 4. (20 poäng)

Y är en stokastisk variabel med täthetsfunktion

$$f(y) = \frac{\lambda^3 y^2}{2} e^{-\lambda y}, \quad y > 0, \quad \lambda > 0.$$

Då är $E(Y) = \frac{3}{\lambda}$ och $V(Y) = \frac{3}{\lambda^2}$. Antag att ett slumpmässigt urval av n observationer görs.

a) Härled momentskattningen $\widehat{\lambda}_{MOM}$.

b) Härled maximumlikelihood-skattningen $\widehat{\lambda}_{ML}$.

c) Vad blir maximumlikelihood-skattningarna av $E(Y)$ och $V(Y)$?

Uppgift 5. (20 poäng)

Antag att en observation görs på den stokastiska variabeln Y, som har täthetsfunktion

$$f(y) = \begin{cases} \theta y^{\theta-1}, & 0 \leq y \leq 1, \quad \theta > 0 \\ 0, & \text{annars.} \end{cases}$$

a) Hur ser förkastelseområdet (RR) ut för det test som maximerar styrkan för att testa hypotesen $H_0 : \theta = 1$ mot $H_a : \theta = 2$? Bestäm kritisk gräns för signifikansnivån 0.1.

b) Visa att testet i a) är det likformigt starkaste ("uniformly most powerful") testet för att testa $H_0 : \theta = 1$ mot $H_a : \theta > 1$.

Statistiska institutionen

Stockholms universitet

Rättningsblad

Datum: 24/11-2016

Sal: Brunnsvikssalen

Tenta: Statistisk teori med tillämpningar II

Kurs: Statistisk teori med tillämpningar

ANONYMKOD: STM -0015

[X] 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. 8	20	15	20	20					

83 + 0 bonus

POÄNG	BETYG	Lärarens sign.
83	B	

SU, STATISTIK

Skrivsal: Brunnsvikssalen **Anonymkod:** STM-0016 **Blad nr:** 1

Uppg. 1

a) Statistisk signifikans och praktisk signifikans.

~~Kan ej lämnar plats för senare uppenbarelse.~~

Jag kan inte och vill ej gissa.

b) Signifikansnivå är ett sätt att beskriva hur stor säkerhet man väljer att ens slutsatser ska ha. Vanligtvis har man t.ex en nollhypotes som man förkastar om någon statistika hamnar utanför ett intervall. Har man då ett 5% signifikansnivå så betyder det att, givet parameterns värde under nollhypotesen, så finns det 5% sannolikhet att ett urval ger en statistika som hamnar utanför detta intervall. Alltså om och endast om nollhypotesen är sann.

p-värde är den signifikansnivån som krävs för att förkasta nollhypotesen. ✓

2

c) Vi har en nollhypotes och en alternativhypotes. Om nollhypotesen är sann men vi förkastar den, då har vi gjort ett typ I-fel. R

Om alternativhypotesen är sann, men vi inte förkastar nollhypotesen, då har vi gjort ett typ II-fel. R

5

d) Inom Bayesiansk teori.

Apriori-fördelning är sannolikhetsfördelningen man tror att en statistika har, baserat på tidigare teori. parameter!

Aposteriorifördelning är sannolikhetsfördelningen baserat på Apriori, data, och en likelihoodfunktion.

Från formelbladet: $g^*(\theta|y) = \frac{L(y|\theta)g(\theta)}{\int_\theta L(y|\theta)g(\theta)d\theta}$

där $g^*(\theta|y)$ är aposteriori- och $g(\theta)$ är apriorifördelningen.

1

8

SU, STATISTIK

Skrivsal: Brunnsviksalen Anonymkod: 8TM-0014 Blad nr: 2

Uppg. 2)

a) Taget direkt ur formelbladet:

$E(Y) = \frac{\theta_1 + \theta_2}{2} = \frac{-0.5 + 0.5}{2} = 0$ R

$V(Y) = (\theta_2 - \theta_1)^2 / 12 = (0.5 + 0.5)^2 / 12 = 1/12 \approx 0.0833$ R

Vi vill hitta motsvarande för $\bar{Y}$.

$E(\bar{Y}) = E(\sum_{i=1}^{n} Y_i / n) = E(\sum_{i=1}^{n} Y_i)/n = nE(Y)/n = 0$ R

$V(\bar{Y}) = V(\sum_{i=1}^{n} Y_i / n) = V(\sum_{i=1}^{n} Y_i)/n^2 = \sum_{i=1}^{n}(V(Y_i))/n^2 = V(Y)/n$

$= (1/12)/48 \approx 0.0017$ R 5

b) Då n är stort antar vi att $\bar{Y}$ är approximativt normalfördelat (p.g.a. CGS) och har då:

$\bar{Y} \sim N(\mu_{\bar{Y}} = 0, \sigma^2_{\bar{Y}} = 0.0017)$ R

Vi vill då hitta $p(-0.1 \leq \bar{Y} \leq 0.1)$.

Vi har smarta miniräknare som ger oss ungefär 0.98.

Har man inte det kan man göra såhär:

$p(-0.1 \leq \bar{Y} \leq 0.1) = p(\frac{-0.1}{\sqrt{0.0017}} \leq Z \leq \frac{0.1}{\sqrt{0.0017}}) \approx p(-2.43 \leq Z \leq 2.43)$

$= p(z \geq -2.43) - p(z \geq 2.43) = p(z \leq 2.43) - (1 - p(z \leq 2.43))$

$= \underbrace{2p(z \leq 2.43)}_{\text{Hitta i tabell}} - 1 \approx 0.98$ R 5

c) Vi vill nu hitta $E(\sum_{i=1}^{n} Y_i)$ och $V(\sum_{i=1}^{n} Y_i)$

$E(\sum_{i=1}^{n} Y_i) = \sum E(Y_i) = nE(Y) = 0$ R

$V(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} V(Y_i) = nV(Y) = 48/12 = 4$ R 5

d) Eftersom $\sum_{i=1}^{n}(Y_i)/n$ är approximativt normalfördelat måste även $\sum_{i=1}^{n}(Y_i)$ vara det.

Vänd

Vi har då $\sum_{i=1}^{n}(Y_i) \sim N(\mu_{\Sigma Y}=0, \sigma_{\Sigma Y}=2)$ och vill hitta $p(-6 \leq \sum Y_i \leq 6)$.

Använder miniräknaren och ser ≈ 0.997.

Annars: $\underbrace{2p(Z \leq 3)}_{\text{tabell}} - 1 \approx 0.997$
(som i b))

R

5

20

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: STM-0015 Blad nr: 3

Uppg. 3

$n_1 = 11$ $n_2 = 11$, alla observationer antas vara oberoende.
förtydliga

a) Jag gör ett Mann Whitney U test.

H₀: Ingen skillnad i läge mellan fördelningarna i vilopuls för män & kvinnor

H₁: Skillnad i läge mellan fördelningarna

$\alpha = 0.05$

	"wins+ties"		"wins+ties"
54	0	56	1.5
56	0.5	60	2.5
60	2	60	2.5
66	4	62	3
68	4.5	68	5
68	4.5	76	7
74	5	78	7
80	8.5	78	7
80	8.5	80	8
86	9	96	11
90	9	98	11
	$\Sigma = 55.5$		$\Sigma = 65.5$

$U = 55.5$ R

Titta i tabell ⇒ U finns ej för $n = 11$.

Vi antar då att U är approximativt normalfördelat (CGS) och har
n är stort nog så att

$$|z_{obs}| = \left|\frac{U - (n_1 n_2/2)}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}\right| = \left|\frac{55.5 - 60.5}{\sqrt{121(23)/12}}\right| \approx |-0.33| = 0.33$$ R

Med 5% signifikansnivå är $z_{krit} = 1.96$. R

$|z_{obs}| < z_{krit}$, och vi kan då inte förkasta nullhypotesen.
slutsats? 8

b) Jag väljer att göra ett t-test. Jag antar att medelvärdet för män respektive kvinnor är normalfördelat. (pga CGS igen) $+ \sigma_1^2 = \sigma_2^2$

Med $\alpha = 0.05$ har vi då $t^{(19)}_{krit} = 2.09$ (Taget ur tabell) (Borde egentligen stå här)

Y = Vilopuls (1 för män, 2 för kvinnor)

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$ R

$$T_{obs} = \frac{\bar{Y}_1 - \bar{Y}_2 - D_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\bar{Y}_1 - \bar{Y}_2 - D_0}{\sqrt{((n_1 - 1)S_1^2 + (n_2 - 1)S_2^2)/(n_1 + n_2 - 2)}\sqrt{\frac{2}{11}}} \approx \frac{71.09 - 73.82}{\sqrt{((10 \cdot 143.5 + 10 \cdot 202.0)/20)} \cdot 0.43}$$

≈ -0.48 R $|t_{obs}| < t_{krit}$ så vi förkastar ej H_0.
slutsats? 7

15

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: STM-0015 Blad nr: 4

Uppg. 4

$f(y) = \frac{\lambda^3 y^2 e^{-\lambda y}}{2}, \quad y>0, \quad \lambda>0$

$\Rightarrow Y \sim Gamma(\alpha = 3, \beta = 1/\lambda)$ R

$E(Y) = \alpha\beta = 3/\lambda$

$V(Y) = \alpha\beta^2 = 3/\lambda^2$

Om jag kommer ihåg rätt är enda skillnaden med en gammal tentafråga att vi skattar λ istället för β

Alltså pinsamt om jag gör fel

:)

a) $\mu_1 = E(Y) = 3/\lambda \Rightarrow \lambda = 3/E(Y)$

$m_1 = \bar{Y} \Rightarrow \hat{\lambda}_{MM} = 3/\bar{Y}$ R

3

b) $L(y_1, y_2, \ldots, y_n | \lambda) = \prod_{i=1}^{n} f(y_i|\lambda)$ $\Rightarrow$ antar oberoende observationer

$= \prod_{i=1}^{n} (\lambda^3 y_i^2 e^{-\lambda y_i}/2)$ R

$\ln\left(\prod_{i=1}^{n}(\lambda^3 y_i^2 e^{-\lambda y_i}/2)\right) = \sum_{i=1}^{n}(\ln(\lambda^3) + \ln(y_i^2) - \lambda y_i - \ln(2))$

$= 3n\ln(\lambda) - n\ln(2) + \sum_{i=1}^{n}\ln(y_i^2) - \lambda\sum_{i=1}^{n} y_i$ R

Derivera med avseende på λ

$= 3n/\lambda - \sum_{i=1}^{n} y_i$ R

Sätt lika med 0

$0 = 3n/\lambda - \sum_{i=1}^{n} y_i \Rightarrow 3n/\lambda = \sum_{i=1}^{n} y_i \Rightarrow \lambda = 3n/\sum_{i=1}^{n} y_i = 3/\bar{Y}$

$\hat{\lambda}_{ML} = 3/\bar{Y}$ R

12

c) $E(Y)$ och $V(Y)$ är funktioner utav λ, därför räcker det att stoppa in $\hat{\lambda}_{ML}$ för att hitta respektive ML skattning.

$\widehat{E(Y)}_{ML} = 3/\hat{\lambda}_{ML} = 3/(3/\bar{Y}) = \bar{Y}$ R

$\widehat{V(Y)}_{ML} = 3/\hat{\lambda}_{ML}^2 = 3/(9/\bar{Y}^2) = \bar{Y}^2/3$ R

5

20 Bra!

SU, STATISTIK

Skrivsal: Brunnsvikssalen Anonymkod: STM-0015 Blad nr: 5

Uppg. 5 $f(y) = \begin{cases} \theta y^{\theta-1}, & 0 \le y \le 1, \quad \theta > 0 \\ 0, & \text{annars} \end{cases}$

Beta-fördelning med $\beta=1$ och $\alpha=\theta$. Standardform för $\theta \in \mathbb{N}$

a) Vi hitter först likelihoodratien (använder Neyman-Pearsons lemma (tror jag det heter))

Ja!

$\prod f(y_i|\theta=2) / \prod f(y_i|\theta=1)$

$= \prod 2y_i / \prod 1 = \prod 2y_i$

Det test som maximerar styrkan har då formen att H_0 förkastas om $\prod 2y_i$ är större än ett kritiskt värde.

Eftersom vi bara har en observation får vi då

$\prod 2y_i = 2y \implies RR: 2y \ge k$, där k är något kritiskt värde bestämt av signifikansnivån.

Under H_0 är Y uniformt fördelat och skriver vi om $RR: 2y \ge k$ till $RR: y \ge l$ har vi introducerat att $l = 0.9$.

(Lösning l)

Alltså: $RR: y \ge 0.9$ R

1/1

b) $\prod f(y_i|\theta=\alpha) / \prod f(y_i|\theta=1)$, $\alpha > 1$

$= \alpha y^{\alpha-1}$

$RR: \alpha y^{\alpha-1} \ge k \implies y^{\alpha-1} \ge k/\alpha \implies y \ge (k/\alpha)^{\frac{1}{\alpha-1}} = l$

$RR: y \ge l$

I ord: Vi hitter återigen det starkaste testet genom likelihoodratioon, och den blir $\alpha y^{\alpha-1}$. Sedan skriver vi om den på samma sätt som i a). Det slutgiltiga kritiska värdet l bestäms enkelt av 1 − signifikansnivån.

6

(2/1)