Stockholms universitet

OBS! Läs noga igenom anvisningarna i tentamen, t.ex. hur du ska skriva svaren. Det är ditt ansvar som student att följa de anvisningar som ges.
NOTE! Read the examination instructions carefully, e.g. how to write the answers. It is your responsibility as a student to follow the given instructions.

Skriv din anonymiseringskod och dagens datum på allt material du lämnar in.
(*Enter your anonymization code and todays date on all submitted materials*)

Anonymiseringskod (*Anonymization code*)	3	1	1	-	0	0	0	3	-	D	M	A

Datum (*Date YYYY-MM-DD*)	2022-03-21	Plats nr. (*Seat No.*)	121

Kurs/Kurskod (*Course/Course code*)	ST211G
Kursmoment (*Course component*)	Tentamen i Statistisk teori med tillämpningar II

Fylls i av tentamensvärd (*To be filled in by invigilator*)

Direkt i skrivning: (kryss)		Svarsblankett: (kryss)		Lösa svarsblad: (antal)	11
Lämnat in blankt: (kryss)		Dator: (kryss)			

Inlämningstid: 18 : 50 Signatur tentamensvärd: Anelo

Fylls i av lärare/examinator (*To be filled in by teacher/examinator*)

Betyg:	A	Poäng:	93

Signatur rättande lärare/examinator: Jlw

Stockholms universitet

Datum: (Date YYYY-MM-DD)	2022-03-21	Kurs/Kurskod: (Course/Course code)	ST211G	Sidnr.: (Page no.)
Anonymiseringskod (Anonymization code)	311-0003-DHA			1

Uppg.nr.: (Task no.) 1

Lärarens kommentar: (Teacher's note)

1. $n = 236$ patienter

$Y = 189$ positiva utfall

$$\hat{p} = \frac{Y}{n} = \frac{189}{236} = 0{,}8008\ldots$$

a) 99%-igt konfidensintervall ges ($n > 30 \rightarrow$ CGS) av:

$\alpha = 0{,}01$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow \frac{189}{236} \pm 2{,}5758 \cdot \sqrt{\frac{189/236 \cdot \left(1 - \frac{189}{236}\right)}{236}}$$

$$V(\hat{p}) = V\left(\frac{Y}{n}\right) = \frac{V(Y)}{n^2} = \frac{n\hat{p}(1-\hat{p})}{n^2} = \frac{\hat{p}(1-\hat{p})}{n} \; ; \; \sigma_{\hat{p}} = \sqrt{V(\hat{p})}$$

$$\rightarrow (0{,}7339 \; ; \; 0{,}8678)$$

← 99%-igt konfidensintervall för andelen positiva hemtest bland patienter med Covid-19

5

b) För att längden av det 99%-iga konfidensintervallet inte ska överstiga 0,1 behöver: $2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0{,}1$

men du vet inte $\hat{p}$ (-3)

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{0{,}1}{2 \cdot z_{\alpha/2}} \; ; \; \frac{\hat{p}(1-\hat{p})}{n} = \left(\frac{0{,}1}{2 \cdot z_{\alpha/2}}\right)^2$$

$$n = \frac{\hat{p}(1-\hat{p})}{\left(\frac{0{,}1}{2 \cdot z_{\alpha/2}}\right)^2} = \frac{\frac{189}{236}\left(1 - \frac{189}{236}\right)}{\left(\frac{0{,}1}{2 \cdot 2{,}5758}\right)^2} = 423{,}27\ldots$$

$$\rightarrow \underline{n = 424}$$

2

Svar: Det behövs åtminstone 424 patienter i urvalet i så fall.

Poäng: (Points)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Stockholms universitet

Datum: (Date YYYY-MM-DD) 2022-03-21 | Kurs/Kurskod: (Course/Course code) ST211G | Sidnr.: (Page no.) 2

Anonymiseringskod (Anonymization code) 311-0003-DHA

Uppg.nr.: (Task no.) 1

Lärarens kommentar: (Teacher's note)

1.

c) Ju mindre den statistiska felmarginalen blir desto större blir urvalet då vi kommer dela $\hat{p}(1-\hat{p})$ med ett allt mindre tal. R

$$n = \frac{\hat{p}(1-\hat{p})}{\underbrace{\left(\frac{\text{"Statistiska felmarginalen"}}{2 \cdot z_{\alpha/2}}\right)^2}_{\text{"nämnaren"}}}$$

Om $A > B$ kommer

$$\underbrace{\frac{A}{2z_{\alpha/2}}}_{=A^*} > \underbrace{\frac{B}{2z_{\alpha/2}}}_{=B^*}$$

$$\Rightarrow \frac{\hat{p}(1-\hat{p})}{A^*} < \frac{\hat{p}(1-\hat{p})}{B^*}$$

Svar: Urvalet blir större

d) Minskad konfidensgrad kommer resultera i att $z_{\alpha/2}$ minskar i värde, vilket gör hela "nämnaren" större, vilket gör hela uttrycket ovan kommer minska i värde eftersom vi kommer dela $\hat{p}(1-\hat{p})$ med ett större tal

Konfidensgraden minskar $\rightarrow$ α ökar i värde

Om $\alpha_1 < \alpha_2 \rightarrow 2 \cdot z_{\frac{\alpha_1}{2}} > 2 \cdot z_{\frac{\alpha_2}{2}}$

Statistisk felmarginal = SF

$$\rightarrow \underbrace{\frac{SF}{2 \cdot z_{\frac{\alpha_1}{2}}}}_{=\alpha_1^*} < \underbrace{\frac{SF}{2 \cdot z_{\frac{\alpha_2}{2}}}}_{=\alpha_2^*}$$

$$\Rightarrow \frac{\hat{p}(1-\hat{p})}{\alpha_1^*} > \frac{\hat{p}(1-\hat{p})}{\alpha_2^*}$$

Svar: Urvalsstorleken blir mindre R

Poäng: (Points) 17

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Stockholms universitet

Datum: (Date YYYY-MM-DD) 2022-03-21 | Kurs/Kurskod: (Course/Course code) ST211G | Sidnr.: (Page no.) 3

Anonymiseringskod (Anonymization code) 311-0003-DHA

Uppg.nr.: (Task no.) 2

Lärarens kommentar: (Teacher's note)

(2) Vi har: $\hat{p}_1 = Y_1$; $\hat{p}_2 = \frac{Y_1 + Y_2}{2}$; $\hat{p}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$

Bernoullifördelade a) Besvaras på nästa sida

b) $E(\hat{p}_1) = E(Y_1) = p \rightarrow$ VVR estimator av p

$$E(\hat{p}_2) = E\left(\frac{Y_1+Y_2}{2}\right) = \frac{E(Y_1+Y_2)}{2} \overset{\text{oberoende}}{=} \frac{E(Y_1)+E(Y_2)}{2}$$

5

$\rightarrow = \frac{p+p}{2} = \frac{\cancel{2}p}{\cancel{2}} = p \rightarrow$ VVR estimator av p

c) Varians för ett Bernoulliförsök ges av: $V(Y) = p(1-p)$

$V(\hat{p}_1) = V(Y_1) = p(1-p)$

pga oberoende

$$V(\hat{p}_2) = V\left(\frac{Y_1+Y_2}{2}\right) = \frac{V(Y_1+Y_2)}{2^2} = \frac{V(Y_1)+V(Y_2)}{4}$$

$$= \frac{p(1-p)+p(1-p)}{4} = \frac{\cancel{2}(p(1-p))}{\cancel{4}\,2} = \frac{p(1-p)}{2}$$

5

$$\text{eff}(\hat{p}_1, \hat{p}_2) = \frac{V(\hat{p}_2)}{V(\hat{p}_1)} = \frac{\frac{\cancel{p(1-p)}}{2}}{\frac{\cancel{p(1-p)}}{1}} = \frac{1}{2}$$

R

Eftersom $\text{eff}(\hat{p}_1, \hat{p}_2) = \frac{1}{2} < 1 \Rightarrow \hat{p}_2$ är en bättre estimator

Poäng: (Points)

Tentamensservice/SU/typ_b

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Datum: (Date YYYY-MM-DD) 2022-03-21	Kurs/Kurskod: (Course/Course code) ST211G		Sidnr.: (Page no.) 4
Anonymiseringskod (Anonymization code) 311-0003-DHA			

Stockholms universitet

Uppg.nr.: (Task no.) 2

Lärarens kommentar: (Teacher's note)

(2) d) $E(\hat{p}_n) = E\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \frac{E\left(\sum_{i=1}^{n} Y_i\right)}{n} = \frac{\sum_{i=1}^{n} E(Y_i)}{n}$

$\rightarrow = \frac{np}{n} = p \rightarrow$ VVR

$V(\hat{p}_n) = V\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \frac{V\left(\sum_{i=1}^{n} Y_i\right)}{n^2} = \frac{\sum_{i=1}^{n} V(Y_i)}{n^2} = \frac{np(1-p)}{n^2}$ 6

$= \frac{p(1-p)}{n}$

För att $\hat{p}_n$ ska vara en konsistent behöver $\hat{p}_n$ vara VVR (har visat), dessutom behöver $\lim_{n\to\infty} V(\hat{p}_n) = 0$

Vi har: $\lim_{n\to\infty} \frac{p(1-p)}{n} = 0 \Rightarrow \hat{p}_n$ är en konsistent estimator

a) Samplingsfördelningarna till $\hat{p}_1$ och $\hat{p}_2$ är binomialfördelningen med $n=1$ respektive $n=2$ 2

Tentamensservice/SU/typ_b

Poäng: (Points) 18

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Stockholms universitet

Datum: (Date YYYY-MM-DD) 2022-03-21 | Kurs/Kurskod: (Course/Course code) ST211G | Sidnr.: (Page no.) 5

Anonymiseringskod (Anonymization code) 311-0003-DHA

Uppgg.nr.: (Task no.) 3

Lärarens kommentar: (Teacher's note)

(3) $k=2 \rightarrow f(y) = \frac{2y}{\lambda^2} e^{-(y/\lambda)^2} \quad 0<y<\infty$

$E(Y) = \lambda \frac{\sqrt{\pi}}{2} \; ; \; V(Y) = \lambda^2 \left(1 - \frac{\pi}{4}\right)$

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

$\mu_1' = E(Y) = \lambda \frac{\sqrt{\pi}}{2}$

$m_1' = \mu_1' \rightarrow \bar{y} = \lambda \frac{\sqrt{\pi}}{2}$

$\rightarrow \hat{\lambda}_{ME} = \frac{2\bar{y}}{\sqrt{\pi}}$ R 5

b) $E(\hat{\lambda}_{ME}) = E\left(\frac{2\bar{y}}{\sqrt{\pi}}\right) = E\left(\frac{2\sum y_i}{n\sqrt{\pi}}\right) = \frac{2E(\sum y_i)}{n\sqrt{\pi}}$

$= \frac{2n\lambda\frac{\sqrt{\pi}}{2}}{n\sqrt{\pi}} = \frac{\cancel{2n\lambda\sqrt{\pi}}}{\cancel{2n\sqrt{\pi}}} = \lambda \rightarrow$ VVR R 5

c) $L(y|\lambda) = \prod_{i=1}^{n} \frac{2y_i}{\lambda^2} e^{-\left(\frac{y}{\lambda}\right)^2} = 2^n \cdot \prod_{i=1}^{n} y_i \cdot \lambda^{-2n} \cdot e^{-\left(\frac{\sum_{i=1}^{n} y_i^2}{\lambda^2}\right)}$

$l(\lambda) = \ln L(y|\lambda) = n\ln 2 + \ln\left(\prod_{i=1}^{n} y_i\right) - 2n\ln\lambda - \frac{\sum_{i=1}^{n} y_i^2}{\lambda^2} \ln e$

$l'(\lambda) = 0 + 0 - \frac{2n}{\lambda} + 2\frac{\sum_{i=1}^{n} y_i^2}{\lambda^3} = \frac{-2n\lambda^2 + 2\sum_{i=1}^{n} y_i^2}{\lambda^3}$

$\rightarrow = \frac{2\left(-n\lambda^2 + \sum_{i=1}^{n} y_i^2\right)}{\lambda^3}$

Poäng: (Points) 20

Tentamensservice/SU/typ_b

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Datum: (Date YYYY-MM-DD)	2022-03-21	Kurs/Kurskod: (Course/Course code)	ST211G	Sidnr.: (Page no.)
Anonymiseringskod (Anonymization code)	311-0003-DHA			6

Stockholms universitet

Uppg.nr.: (Task no.) 3

Lärarens kommentar: (Teacher's note)

3. forts.

$l'(\hat{\lambda}) = 0 \Rightarrow 2\left(-n\hat{\lambda}^2 + \sum_{i=1}^{n} y_i^2\right) = 0$

$\rightarrow n\hat{\lambda}^2 = \sum_{i=1}^{n} y_i^2, \quad \hat{\lambda}^2 = \frac{\sum_{i=1}^{n} y_i^2}{n}$

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

10

Svar: Maximum likelihoodestimatorn av λ:

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

R

Poäng: (Points)

Tentamensservice/SU/typ_b

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Stockholms universitet

Datum: (Date YYYY-MM-DD)	2022-03-21	Kurs/Kurskod: (Course/Course code)	ST211G	Sidnr.: (Page no.)
Anonymiseringskod (Anonymization code)	311-0003-DHA			7

Uppg.nr.: (Task no.) 4

Lärarens kommentar: (Teacher's note)

4.

Person	1	2	3	4	5	6	7	8	9	10
Sommar	1,69	1,57	2,57	2,10	2,22	1,59	2,09	1,86	1,63	1,49
Vinter	1,65	1,74	1,74	2,02	2,36	1,09	2,09	1,63	1,56	1,36
d_i	0,04	−0,17	0,83	0,08	−0,14	0,5	0	0,23	0,07	0,13
	+	−	+	+	−	+	tie	+	+	+

Vi har parvisa observationer/relaterade urval

a) Väljer att göra ett teckentest.

Antar oberoende observationer mellan och inom personerna.

$H_0: p = 0{,}5$ $\quad$ M = antalet positiva differenser = 7

$H_A: p \neq 0{,}5$ $\quad$ $n = 9$ (pga tie hos person 7)

$\alpha = 0{,}05$

$RR = \{M = 0,1,2,7,8,9\}$; $Y \sim Bin(n,p)$ om H_0 sann

$\rightarrow Bin(9, 0{,}5)$

p-värde: $P(M = 0,1,2,7,8,9) = 2 \cdot P(M \leq 2) = 2 \cdot 0{,}08984$

$\rightarrow = 0{,}17968 > \alpha = 0{,}05$

10

Slutsats: Med hjälp av teckentestet kan vi inte förkasta H_0 på $\alpha = 0{,}05$, dvs vi kan inte förkasta att att energiförbrukningen är densamma på sommaren och vintern.

Poäng: (Points) 20

Tentamensservice/SU/typ_b

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Datum: (Date YYYY-MM-DD)	2022-03-21	Kurs/Kurskod: (Course/Course code)	ST211G	Sidnr.: (Page no.) 8
Anonymiseringskod (Anonymization code)	311-0003-DHA			

Uppg.nr.: (Task no.) 4

Lärarens kommentar: (Teacher's note)

4. b) Eftersom vi har parvisa observationer har vi som teststatistika:

$$T = \frac{\bar{d}}{S/\sqrt{n}} \sim t(n-1) \text{ om } H_0 \text{ sann}$$

$H_0: \mu_{sommar} = \mu_{vinter} \rightarrow \bar{d} = 0$

$H_A: \mu_{sommar} \neq \mu_{vinter} \rightarrow \bar{d} \neq 0$

$\alpha = 0,05 \quad n = 10$

Behöver anta att differenserna är normalfördelade, oberoende, med väntevärde μ och varians σ^2

$\sum d_i = 0,04 - 0,17 + 0,83 + 0,08 - 0,14 + 0,5 + 0 + 0,23 + 0,07 + 0,13 = 1,57$

$$\bar{d} = \frac{\sum d_i}{n} = \frac{1,57}{10} = 0,157$$

$\sum d_i^2 = 0,04^2 + (-0,17)^2 + 0,83^2 + 0,08^2 + (-0,14)^2 + 0,5^2 + 0,23^2 + 0,07^2 + 0,13^2 = 1,0701$

$$S^2 = \frac{\sum_{i=1}^{10} d_i^2 - n(\bar{d})^2}{n-1} = \frac{1,0701 - 10 \cdot 0,157^2}{10-1} \approx 0,091512\ldots$$

$$t_{obs} = \frac{\bar{d}}{S/\sqrt{n}} = \frac{0,157}{\sqrt{0,091512\ldots}/\sqrt{10}} \approx 1,641$$

10

Vi förkastar H_0 om $|t_{obs}| > t_{krit} = t_{\alpha/2}(9) = 2,26$

$\alpha = 0,05$ dubbelsidigt

Då $1,641 < 2,26$ kan vi inte förkasta H_0, dvs. vi kan inte förkasta att förbrukningen av energi under sommaren och vintern för detta urval är densamma.

Poäng: (Points)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Stockholms universitet

Datum: (Date YYYY-MM-DD)	2022-03-21	Kurs/Kurskod: (Course/Course code)	ST211G	Sidnr.: (Page no.)
Anonymiseringskod (Anonymization code)	311-0003-DHA			9

Uppg.nr.: (Task no.) 5

Lärarens kommentar: (Teacher's note)

5.

Vi har: Tiden mellan att två radioaktiva partiklar träffar en viss yta: $f(y) = \frac{1}{\beta} e^{-y/\beta}$

$E(Y) = \beta$; Normala förhållanden: $t_{\text{två träffar}} = \beta_0$ ← i medeltal

Förhöjd strålnivå: $t_{\text{två träffar}} = \beta < \beta_0$

→ Vill testa: $H_0: \beta = \beta_0$

$H_A: \beta < \beta_0$

$$L(y|\beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-y/\beta} = \beta^{-n} e^{-\frac{\sum_{i=1}^{n} y_i}{\beta}}$$

Likformigt starkaste test får vi genom att använda Neyman-Pearsons lemma:

$$\rightarrow \frac{L(H_0)}{L(H_A)} < k \rightarrow \frac{L(\beta = \beta_0)}{L(\beta < \beta_0)} < k$$

väljer godtyckligt $\beta < \beta_0$

Alla summatecken framöver avser $i=1$ till n om inte annat anges

$$\Rightarrow \frac{\beta_0^{-n} e^{-\sum_{i=1}^{n} y_i / \beta_0}}{\beta^{-n} e^{-\sum_{i=1}^{n} y_i / \beta}} < k, \quad \left(\frac{\beta_0}{\beta}\right)^{-n} e^{-\frac{\sum y_i}{\beta_0} + \frac{\sum y_i}{\beta}}$$

$$\rightarrow \left(\frac{\beta}{\beta_0}\right)^{n} e^{\frac{\beta_0 \sum y_i - \beta \sum y_i}{\beta_0 \beta}} < k$$

Poäng: (Points)

Tentamensservice/SU/typ_b

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

$$\rightarrow \ln\left(\left(\frac{\beta}{\beta_0}\right)^n e^{\frac{\sum y_i(\beta_0-\beta)}{\beta_0\beta}}\right) < \ln k$$

$$\rightarrow n\ln\left(\frac{\beta}{\beta_0}\right) + \frac{\sum y_i(\beta_0-\beta)}{\beta_0\beta}\ln e < \ln k$$

$$\rightarrow \frac{\sum y_i(\beta_0-\beta)}{\beta_0\beta} < \ln k - n\ln\left(\frac{\beta}{\beta_0}\right) \qquad (\beta_0\beta > 0)$$

$$\rightarrow \sum y_i(\beta_0-\beta) < \beta_0\beta\left(\ln k - n\ln\left(\frac{\beta}{\beta_0}\right)\right)$$

$$\rightarrow \sum y_i < \underbrace{\frac{\beta_0\beta\left(\ln k - n\ln\left(\frac{\beta}{\beta_0}\right)\right)}{(\beta_0-\beta)}}_{k^*}$$

ty $(\beta_0-\beta)>0$

ty $\beta_0 > \beta$

konstant

ok (7p)

Vi har alltså att $\sum y_i < k^*$. Detta ger oss teststatistikan $W = \sum_{i=1}^{n} y_i$ och vi förkastar H_0 om $W < k^*$,

dvs. $RR = \{W < k^*\}$

Hur är det med 2-sidigt, $\beta < \beta_0$? (−3)

Vi standardiserar teststatistikan genom att subtrahera med väntevärdet och sedan dividera med standardavvikelsen.

Poäng: (Points) 17

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)

Datum: (Date YYYY-MM-DD)	2022-03-21	Kurs/Kurskod: (Course/Course code)	ST211G	Sidnr.: (Page no.)
Anonymiseringskod (Anonymization code)	311-0003-DHA			11

Uppg.nr.: (Task no.) 5

Lärarens kommentar: (Teacher's note)

forts.

Vi har: $Y \sim Exp$; $E(Y) = \beta$
$V(Y) = \beta^2$

$\rightarrow E(\sum y_i) = n\beta$
$V(\sum y_i) = n\beta^2$

$$\rightarrow Z = \frac{W - E(W)}{\sqrt{V(W)}} = \frac{\sum y_i - n\beta}{\sqrt{n\beta^2}} = \frac{\frac{1}{n}(\sum y_i - n\beta)}{\frac{1}{n}\sqrt{n\beta^2}}$$

$$\rightarrow \frac{\frac{\sum y_i}{n} - \frac{\cancel{n}\beta}{\cancel{n}}}{\frac{\sqrt{\cancel{n}}\sqrt{\beta^2}}{\sqrt{\cancel{n}}\sqrt{n}}} = \frac{\bar{y} - \beta_0}{\sqrt{\beta_0^2}/\sqrt{n}} \sim \text{approx } N(0,1)$$

om n är stort

7

Vi har fått vår standadisering på formen:

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$$

som vi vet är approx $N(0,1)$ om n är tillräckligt stort på grund av den centrala gränsvärdes satsen CSG

Poäng: (Points)

Uppg.nr.:
(*Task no.*)

Lärarens kommentar:
(*Teacher's note*)

Poäng:
(*Points*)