Kurskod: TAMS65
Provkod: TEN1/TEN2

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningsförslag till tentamen onsdagen den 21 augusti 2019 kl 8–12

1. Vi har följande tabell

		Betyg från traine-program 3	4	5	n_i
Betyg på arbetet	Bra	$N_{11} = 23$	$N_{12} = 60$	$N_{13} = 29$	112
		$n_i\hat{p}_j = 20.47$	55.80	35.73	
	Mycket bra	$N_{21} = 28$	$N_{22} = 79$	$N_{23} = 60$	167
		30.53	83.20	53.27	
	N_j	51	139	89	$n = 279$
	$\hat{p}_j$	$51/279 \approx 0.183$	$139/279 \approx 0.498$	$89/279 \approx 0.319$	

Testa hypotesen H_0 : samma fördelning mellan stickproven $\Leftrightarrow$ Inget samband mellan betygen, mot H_1 : olika fördelning $\Leftrightarrow$ Samband mellan betygen.

Använd teststorheten

$$T = \sum_{i=1}^{3}\sum_{j=1}^{2}\frac{(N_{ij} - n_i\hat{p}_j)^2}{n_i\hat{p}_j}$$
$$= \frac{(23-20.47)^2}{20.47} + \frac{(60-55.80)^2}{55.80} + \cdots + \frac{(60-53.27)^2}{53.27} \approx 3.166.$$

Förkasta hypotesen att det skulle vara lika om $T > c$ där c fås ur en $\chi^2((3-1)(2-1))$-tabell, dvs. $c = \chi^2_{0.95}(2) = 5.99$. Vi har alltså $\underline{\underline{T = 3.166 < 5.99 = c \text{ och vi kan inte förkasta likhet av fördelning (vi kan inte påvisa samband)}}}$.

2. (a) Antag att x_{ij} är en observation från $X_{ij} \sim N(\mu_i, \sigma_i)$, $i = 1, 2, 3$ och $j = 1, ..., 11$. Testa hypotesen $H_0 : \sigma_1 = \sigma_2$ mot $H_1 : \sigma_1 \neq \sigma_2$ på nivån $\alpha = 5\%$.
Teststorhet: $v = \frac{s_2^2}{s_1^2} = 1.31$. H_0 förkastas om $v < a(< 1)$ eller om $v > b(> 1)$. Den s.v. $V \sim F(10, 10)$ om H_0 är sann. Tabellen ger $b = F_{0.975}(10, 10) = 3.72$. Eftersom $v > 1$ kan inte $v < a$ och eftersom dessutom $1 < v < b$ kan vi inte förkasta H_0 på nivån 5%. Det verkar rimligt att anta $\underline{\underline{\sigma_1 = \sigma_2 = \sigma_3}}$. (De andra testen på samma sätt)

   (b) Antag nu att $\sigma_1 = \sigma_2 = \sigma_3$, dvs x_{ij} är en observation från $X_{ij} \sim N(\mu_i, \sigma)$, $i = 1, 2, 3$ och $j = 1, ..., 11$.
Bilda intervallen $I_{\mu_i - \mu_k}$. Skatta parametrarna $\widehat{\mu}_i - \widehat{\mu}_k = \bar{x}_i - \bar{x}_k$.

Den s.v. $\bar{X}_i - \bar{X}_k \sim N\left(\mu_i - \mu_k, \sigma\sqrt{\frac{1}{11}+\frac{1}{11}}\right)$ med den sammanvägda σ^2-skattningen $s^2 = \dfrac{10s_1^2 + 10s_2^2 + 10s_3^2}{30} = 0.203,\ s = 0.45,\ df = 30.$

Hjälpvariabeln $\dfrac{\bar{X}_i - \bar{X}_k - (\mu_i - \mu_k)}{S\sqrt{\frac{1}{11}+\frac{1}{11}}} \sim t(30)$ ger intervallen

$$I_{\mu_i - \mu_k} = \left(\bar{x}_i - \bar{x}_k \mp t_{0.99}(30)s\sqrt{\frac{2}{11}}\right) = (\bar{x}_i - \bar{x}_k \mp 0.47)\,,$$

där $t_{0.99}(30) = 2.46$, dvs.

$$\begin{aligned}
I_{\mu_1-\mu_2} &= (18.37 - 20.61 \mp 0.47) = \underline{\underline{(-2.71\ ;\ -1.77)}},\\
I_{\mu_1-\mu_3} &= (18.37 - 18.41 \mp 0.47) = (-0.51\ ;\ 0.43),\\
I_{\mu_2-\mu_3} &= (20.61 - 18.41 \mp 0.47) = \underline{\underline{(1.73\ ;\ 2.67)}}.
\end{aligned}$$

Vi kan inte påvisa skilland mellan rutt 1 och 3, men rutt 2 verkar ta längre tid än de båda.

3. (a) Vi har att

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \frac{1}{i\theta} e^{-x_i/(i\theta)} = \frac{1}{\theta^n}\left(\prod_{i=1}^{n}\frac{1}{i}\right) e^{-\frac{1}{\theta}\sum_{i=1}^{n}\frac{x_i}{i}}\\
l(\theta) &= \ln L(\theta) = -n\ln\theta + \ln\left(\prod_{i=1}^{n}\frac{1}{i}\right) - \frac{1}{\theta}\sum_{i=1}^{n}\frac{x_i}{i}\\
l'(\theta) &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}\frac{x_i}{i}
\end{aligned}$$

Lös $l'(\theta) = 0$ vilket ger $\underline{\underline{\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{i}}}$.

Vidare gäller att

$$l''(\theta)\big|_{\theta=\widehat{\theta}} = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n}\frac{x_i}{i}\Big|_{\theta=\widehat{\theta}} = -\frac{n}{\widehat{\theta}^2} < 0 \Rightarrow \text{ Max}$$

(Man kan också studera $Y_i = \dfrac{X_i}{i}$)

(b) (Endast TEN2) Det gäller att $E(\widehat{\Theta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{E(X_i)}{i} = \frac{1}{n}\sum_{i=1}^{n}\frac{i\theta}{i} = \theta$. Alltså är skattningen <u>väntevärdesriktig</u> (vvr).

(c) (Endast TEN2) Vi har också $var(\widehat{\Theta}) = \frac{1}{n^2}\sum_{i=1}^{n}\frac{var(X_i)}{i^2} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{i^2\theta^2}{i^2} = \underline{\underline{\frac{\theta^2}{n}}} \to 0$ när $n \to \infty$. Dessutom är skattningen vvr (b) vilket ger att skattningen är en <u>konsistent skattning av θ</u>.

4. (a) Testvariabel $Z = \dfrac{\bar{X} - 1}{2/\sqrt{16}}$ som är $N(0,1)$ under H_0. Förkasta H_0 om $\underline{\underline{|z| > 1.96}}$.

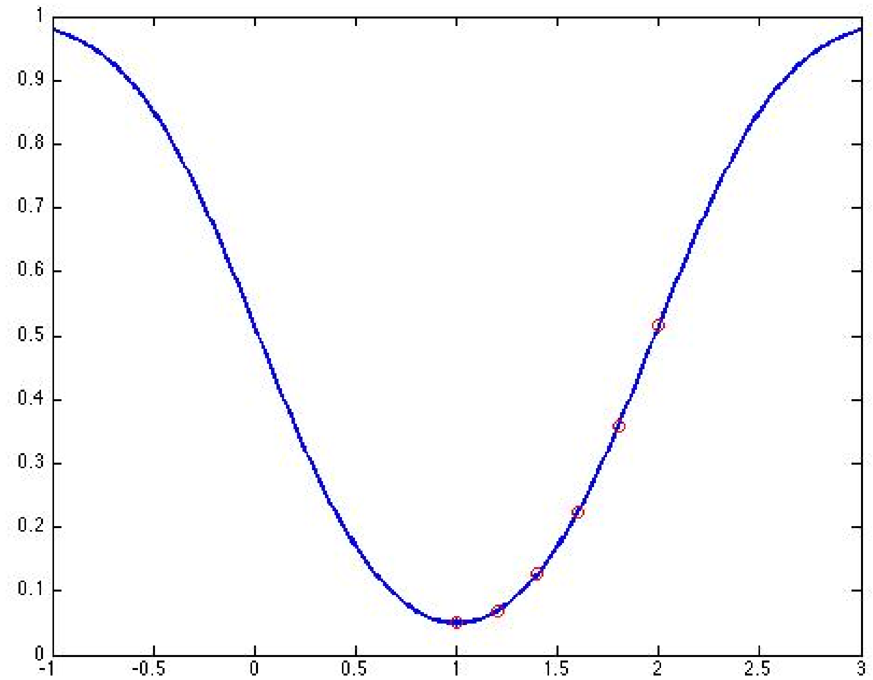
(b) Styrkefunktionen ges av

$$\begin{aligned} h(\mu) &= \text{P(förkasta } H_0 \text{ då } \mu \text{ är det sanna värdet)} \\ &= \text{P}(|Z| > 1.96 \text{ då } \mu \text{ är det sanna värdet}) = \\ &= 1 - \text{P}(-1.96 < Z < 1.96) = 1 - \text{P}(0.02 < \bar{X} < 1.98) \\ &= \underline{\underline{1 - \Phi\left(\frac{1.98 - \mu}{0.5}\right) + \Phi\left(\frac{0.02 - \mu}{0.5}\right)}}. \end{aligned}$$

Vidare ger

$$\begin{aligned} &\mu = 1 \Rightarrow h = \alpha = 0.05 \\ &\mu = 1.2 \Rightarrow h = 1 - \Phi(1.56) + \Phi(-2.36) = 0.0685 \\ &\mu = 1.4 \Rightarrow h = 1 - \Phi(1.16) + \Phi(-2.76) = 0.1259 \\ &\mu = 1.6 \Rightarrow h = 1 - \Phi(0.76) + \Phi(-3.16) = 0.2244 \\ &\mu = 1.8 \Rightarrow h = 1 - \Phi(0.36) + \Phi(-3.56) = 0.3596 \\ &\mu = 2.0 \Rightarrow h = 1 - \Phi(-0.04) + \Phi(-3.96) = 0.516 \end{aligned}$$

Symmetri ger sen skissen



5. (a) $x = 24$ observation från $X \sim Bin(200, p) \approx N(200p, \sqrt{200p(1-p)})$ om $200p(1-p) > 10$. Skatta p med $\hat{p} = \dfrac{x}{200} = 0.12$ som är en observation

från
$\widehat{P} = \dfrac{X}{200} \approx N\left(p, \sqrt{\dfrac{p(1-p)}{200}}\right)$ eftersom $var(\widehat{P}) = \dfrac{var(X)}{200^2} = \dfrac{200p(1-p)}{200^2} = \dfrac{p(1-p)}{200}$.
Approximationerna är ok eftersom $200\hat{p}(1-\hat{p}) = 21.1 > 10$. Hjälpvariabel

$$\frac{\widehat{P}-p}{\sqrt{\frac{p(1-p)}{200}}} \approx \frac{\widehat{P}-p}{\sqrt{\frac{\widehat{P}(1-\widehat{P})}{200}}} \approx N(0,1).$$

Stäng in och lös ut p ger intervallet

$$I_p = \left(\hat{p} \mp z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{200}}\right) = \underline{\underline{(0.075\ ;\ 0.165)}},$$

där $z_{0.975} = 1.96$.

(b) $|I_p| = 2z_{0.975}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} \leq z_{0.975}\dfrac{1}{\sqrt{n}}$ för alla $\hat{p}$ eftersom $\hat{p}(1-\hat{p}) \leq \dfrac{1}{4}$ för alla $\hat{p}$.
Eftersom vi vill att $|I_p| \leq 0.08$ gäller $z_{0.975}\dfrac{1}{\sqrt{n}} \leq 0.08$ vilket ger

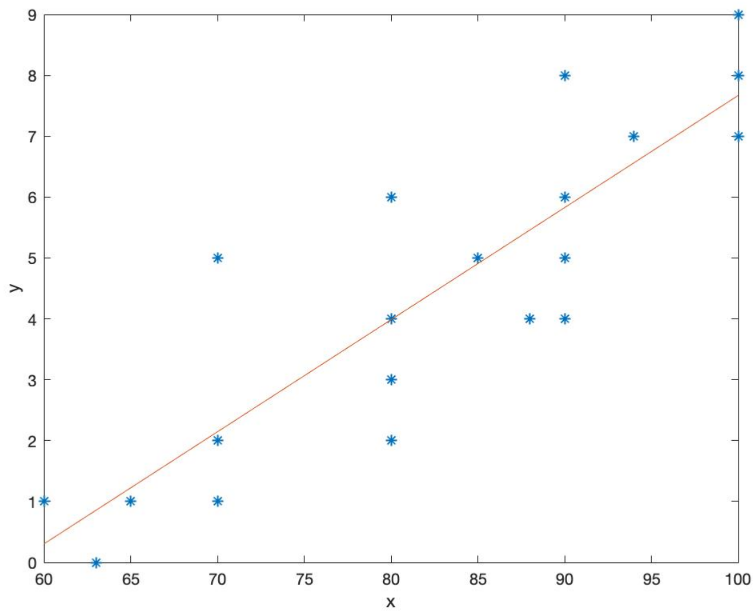
$$n > \left(\frac{1.96}{0.08}\right)^2 = \underline{\underline{600}}$$

.

6. (Endast TEN1)

(a) $\underline{\underline{df_{REGR} = 1}}$ och $\underline{\underline{df_{RES} = 18}}$

(b)



(c) Testa hypotesen $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ på nivån 5%. Förkasta H_0 eftersom $\left|\dfrac{\widehat{\beta}_1}{s\sqrt{h_{11}}}\right| = 7.43 > t_{0.975}(18) = 2.10$. Bullernivåerna verkar påverkar blodtrycksförhöjningen.

(d) Betrakta $Y_0 = \beta_0 + 70\beta_1 + \varepsilon_0 = \boldsymbol{u}'\boldsymbol{\beta} + \varepsilon_0$ där $\boldsymbol{u}' = \begin{pmatrix} 1 & 70 \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ och $\varepsilon_0 \sim N(0, \sigma)$. Skatta σ^2 med $s^2 = \dfrac{Q_{RES}}{18} = \dfrac{33.16}{18} = 1.842$, dvs $s = 1.357$. Skatta $\boldsymbol{u}'\boldsymbol{\beta}$ med $\boldsymbol{u}'\widehat{\boldsymbol{\beta}} = 2.144$ som är en observation från den s.v.

$$\boldsymbol{u}'\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{u}'\boldsymbol{\beta}, \sigma\sqrt{\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}}\right),$$

där $\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u} = 0.082$. Hjälpvariabel $\dfrac{\boldsymbol{u}'\widehat{\boldsymbol{\beta}} - \boldsymbol{u}'\boldsymbol{\beta}}{s\sqrt{\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}}} \sim t(18)$ som ger intervallet

$$I_{\beta_0+70\beta_1} = \left(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} \mp \underbrace{t_{0.975}(18)}_{=2.10} s\sqrt{\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}}\right) = (1.328\ ;\ 2.961).$$