Kurskod: TAMS65
Provkod: TEN1/TEN2

MATEMATISK STATISTIK I FORTSÄTTNINGSKURS

Lösningsförslag till tentamen torsdagen den 10 juni 2019 kl 8–12

1. a) Låt y vara totala antalet benbrott, dvs. $y = 8 \cdot 0 + 14 \cdot 1 + \cdots + 4 \cdot 5 = 104$, och n antalet veckor, $n = 52$.

   Skatta μ med $\hat{\mu} = \dfrac{y}{n} = \dfrac{104}{52} = 2$.

   b) $H_0 : X \sim Po(2)$ mot H_1 : ej $Po(2)$.

   Under H_0 så har vi $p_k = \dfrac{2^k}{k!}e^{-2}$ och

i	p_i	np_i	N_i
0	$e^{-2} = 0.1353$	7.0356	8
1	$2e^{-2} = 0.2707$	14.0764	14
2	$2e^{-2} = 0.2707$	14.0764	13
3	$\frac{4}{3}e^{-2} = 0.1804$	9.3808	8
4	$\frac{2}{3}e^{-2} = 0.0902$	4.6904	5
5	$1 - 7e^{-2} = 0.0527$	2.7404	4

   där $p_5 = P(X \geq 5) = 1 - (p_0 + p_1 + p_2 + p_3 + p_4)$. Eftersom $np_5 = 2.7404 < 5$ så måste vi slå ihop ”fack” 4 och 5.

   Teststorhet: $T = \sum_{i=0}^{4} \dfrac{(N_i - np_i)^2}{np_i} = 0.75$.

   Förkasta H_0 om $T > c = \chi^2_{0.95}(5-1-1) = \chi^2_{0.95}(3) = 7.81$. Förkasta inte H_0. Vi kan ha anpassning till Poissonfördelning.

2. Låt $\widehat{M}_1$ och $\widehat{M}_2$ vara $\widehat{M}_1 = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ och $\widehat{M}_2 = \dfrac{X_1 + X_n}{2}$.

   (a) Då gäller att väntevärdena ges av

$$\mathrm{E}\left(\widehat{M}_1\right) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathrm{E}\left(X_i\right)}_{=\mu} = \frac{1}{n}n\mu = \mu \quad \text{vvr.}$$

$$\mathrm{E}\left(\widehat{M}_2\right) = \mathrm{E}\left(\frac{X_1 + X_n}{2}\right) = \frac{\mathrm{E}(X_1) + \mathrm{E}(X_n)}{2} = \frac{\mu + \mu}{2} = \mu \quad \text{vvr.}$$

(b) Varianserna ges av

$$\operatorname{var}\left(\widehat{M_1}\right) = \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\operatorname{var}(X_i)}_{=\sigma^2} = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$\operatorname{var}\left(\widehat{M_2}\right) = \operatorname{var}\left(\frac{X_1 + X_n}{2}\right) = \frac{\operatorname{var}(X_1) + \operatorname{var}(X_n)}{4} = \frac{\sigma^2 + \sigma^2}{4} = \frac{\sigma^2}{2}$$

Alltså, för $n > 2$ så har vi $\operatorname{var}\left(\widehat{M_1}\right) < \operatorname{var}\left(\widehat{M_2}\right)$ vilket ger att skattning 1 ($\widehat{M_1}$) är effektivast.

(c) Båda skattningar är vvr, se (a). Vidare gäller att

$$\operatorname{var}\left(\widehat{M_1}\right) \to 0 \quad \text{när } n \to \infty \qquad \text{och} \qquad \operatorname{var}\left(\widehat{M_2}\right) = \frac{\sigma^2}{2} \quad \text{för alla } n.$$

Alltså, $\widehat{M_1}$ är en konsistent skattning av μ men $\widehat{M_2}$ är det inte.

3. Vi har att

(a) $L(\theta) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \dfrac{\theta}{x_i^{\theta+1}} = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)}$

$l(\theta) = \ln L(\theta) = n \ln \theta - (\theta + 1) \sum_{i=1}^{n} \ln x_i$

$\dfrac{\partial l}{\partial \theta} = \dfrac{n}{\theta} - \sum_{i=1}^{n} \ln x_i$

$\dfrac{\partial l}{\partial \theta} = 0 \iff \theta = \dfrac{n}{\sum_{i=1}^{n} \ln x_i} = \dfrac{n}{\ln \prod_{i=1}^{n} x_i}$

$\dfrac{\partial^2 l}{\partial \theta^2} = -\dfrac{n}{\theta^2} < 0$, dvs. maxpunkt.

Maximum-likelihood-skattningen är alltså $\widehat{\theta}_{ML} = \dfrac{n}{\sum_{i=1}^{n} \ln x_i}$.

(b) (Endast TEN2) Beräkna väntevärdet enligt

$$\mathrm{E}(X) = \int_1^{\infty} x \frac{\theta}{x^{\theta+1}} dx = \theta \int_1^{\infty} x^{-\theta} dx = -\frac{\theta}{\theta - 1}\left[x^{-\theta+1}\right]_1^{\infty} = \frac{\theta}{\theta - 1} = \mu(\theta),$$

om $\theta > 1$ om $0 < \theta \leq 1$ så existerar inte väntevärdet och således inte heller momentskattningen.

För $\theta > 1$ gäller nu att $\mu(\widehat{\theta}_{MM}) = \bar{x}$ dvs. $\widehat{\theta}_{MM} = \dfrac{\bar{x}}{\bar{x} - 1}$.

4. a) Teststorhet: $z = \dfrac{\bar{x}-\bar{y}}{\sqrt{\frac{1}{8}+\frac{1}{10}}} = 1.62.$

Förkasta H_0 om $z > c = z_{0.95} = 1.645$. Alltså, vi kan inte förkasta H_0, μ_1 och μ_2 kan vara lika.

b) Styrka blir

$$
\begin{aligned}
h &= P\left(\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{8}+\frac{1}{10}}} > 1.645 \text{ då } \bar{X} \sim N\left(2, \frac{1}{\sqrt{8}}\right), \bar{Y} \sim N\left(1, \frac{1}{\sqrt{10}}\right)\right) \\
&= P\Bigg(\underbrace{\frac{\bar{X}-\bar{Y}-(2-1)}{\sqrt{\frac{1}{8}+\frac{1}{10}}}}_{\sim N(0,1)} > \underbrace{1.645 - \frac{1}{\sqrt{\frac{1}{8}+\frac{1}{10}}}}_{=-0.463}\Bigg) \\
&= 1-\Phi(-0.463) = \Phi(0.463) \approx 68\%
\end{aligned}
$$

Ok styrka.

5. Vi har att $X = \{\text{antalet fel hos en enhet}\} \sim Po(\mu)$ och $Y = \{\text{antalet felaktiga enheter bland 1000 kontrollerade}\} \sim Bin(1000, p)$, där p är sannolikheten att en enhet har minst ett fel, dvs $p = 1 - P(X = 0) = 1 - e^{-\mu}$.

(a) $y = 88$ ger att $\hat{p} = \frac{88}{1000}$. Samtidigt gäller att $\hat{p} = 1 - e^{-\hat{\mu}}$, vilket ger att $\underline{\underline{\hat{\mu} = -\ln(1-\hat{p}) \approx= 0.092.}}$

(b) Bilda först ett konfidensintervall för p.

Det gäller att $Y \sim Bin(1000, p) \approx N\left(1000p, \sqrt{(}1000p(1-p)\right)$, eftersom $n\hat{p}(1-\hat{p}) = 80 > 10$. Vidare gäller att $\hat{P} = \dfrac{Y}{1000} \approx N\left(p, \sqrt{\dfrac{p(1-p)}{1000}}\right)$.

Vi har då hjälpvariabeln

$$
\frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{1000}}} \approx \frac{\hat{P}-p}{\sqrt{\frac{\hat{P}(1-\hat{P})}{1000}}} \approx N(0,1).
$$

Instängning ger intervallet

$$
I_p = \left(\hat{p} \mp z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{1000}}\right) = (0.0704\ ;\ 0.1056), T
$$

där $z_{0.975} = 1.96$. Eftersom funktionen $\hat{\mu} = -\ln(1-\hat{p})$ är strängt växande fås konfidensintervallet för μ av

$$
\underline{\underline{I_\mu = (-\ln(1-0.0704)\ ;\ -\ln(1-0.1056)) = (0.073\ ;\ 0.112)}}.
$$

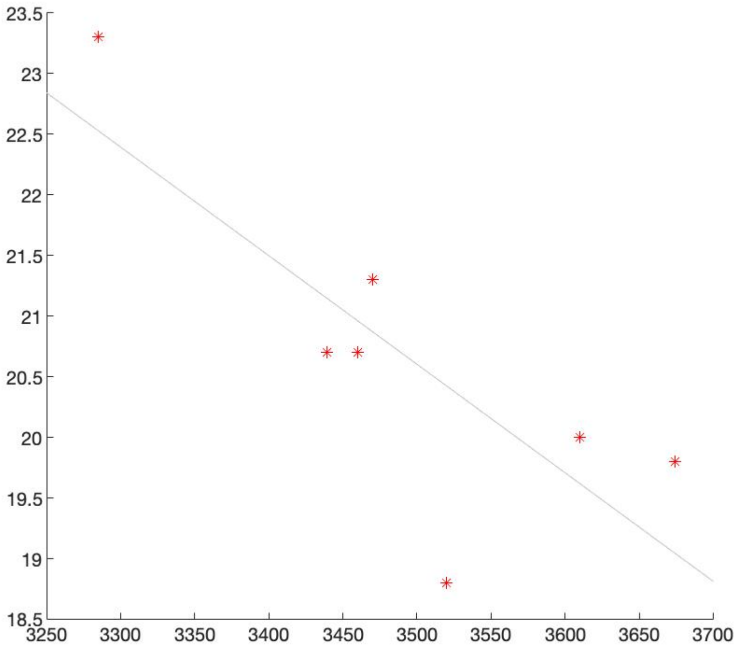
6. (Endast TEN1)

(a) Vi vill testa $H_{0i} : \beta_i = 0$ mot $H_{1i} : \beta_i \neq 0$ på nivån 5% för varje $i = 1, 2$. Teststorhet

$$T_i = \frac{\widehat{\beta}_i}{d(\widehat{\beta}_i)} = \begin{cases} -2.75, & i = 1, \\ 0.95, & i = 2. \end{cases}$$

Förkasta H_{0i} om $|T_i| > t_{0.975}(4) = 2.78$, dvs vi kan inte förkasta $H_{0i} : \beta_i = 0$ för något $i = 1, 2$. Men det finns tendens att förklaringsvariabeln vikt (x_1) påverkar mest.

(b) Observationer samt den skattade regressionslinjen ges av



(c) Fel i uppgiften, så rätt uttryck för prediktionsintervallet

$$I_{\boldsymbol{u}'\boldsymbol{\beta}} = \left(\boldsymbol{u}'\widehat{\boldsymbol{\beta}} \mp t_{0.975}(5) s \sqrt{\boldsymbol{u}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{u}} \right),$$

där $\boldsymbol{u}' = \begin{pmatrix} 1 & ? \end{pmatrix}'$, $t_{0.975}(5) = 2.57$ samt $s^2 = \dfrac{SS_{RES}}{5}$, ger 2p.