Matematiska vetenskaper **Lösning till tentamen: 2016-12-20 kl** $14^{00} - 18^{00}$
Chalmers tekniska högskola **Matematisk statistik för K (TMA074)**

---

1. (a) Vi har $Z = X - Y \sim \mathsf{N}(1, 2^2 + 3^2)$ och söker

$$\mathsf{P}(Z > 0) = 1 - \mathsf{P}(Z < 0) = 1 - \Phi\left(\frac{-1}{\sqrt{2^2 + 3^2}}\right) = 1 - \Phi(-0.2774) = 0.6092$$

(b) Sannolikhetsfunktionen ges av $f(k) = F(k) - F(k-1)$ och blir alltså

k	1	2	3	4	5
f(k)	0.3	0.1	0.3	0.1	0.2

Väntevärdet beräknas som $\mathsf{E}(X) = \sum_{k=1}^{5} kf(k) = 2.8$. För att beräkna variansen beräknar vi först $\mathsf{E}(X^2) = \sum_{k=1}^{5} k^2 f(k) = 10$ och får $\mathsf{V}(X) = \mathsf{E}(X^2) - \mathsf{E}(X)^2 = 10 - 2.8^2 = 2.16$.

2. (a) Vi beräknar differenserna D_i som mätningarna efter minus mätningarna före:

Observation i	1	2	3	4	5	6	7
Differens	3	-10	-9	3	-20	-5	-14

En skattning av Δ ges av $\bar{D} = -7.4286$. Vi skattar variansen av D_i med stickprovsvariansen $s^2 = 72.2857$.

(b) Ett 95% konfidensintervall för Δ ges av

$$I_\Delta = \left(\bar{D} \pm t_{0.025}(6) \cdot \frac{s}{\sqrt{7}}\right) = (-7.4286 \pm 2.4469 \cdot 3.2135) = (-15.2917, 0.4346)$$

(c) Eftersom testet är ensidigt beräknar vi ett ensidigt konfidensintervall

$$I_\Delta = \left(-\infty, \bar{D} + t_{0.05}(6) \cdot \frac{s}{\sqrt{7}}\right) = (-\infty, 1.9432 \cdot 3.2135) = (-\infty, -1.1842)$$

Eftersom intervallet inte täcker noll så kan vi förkasta H_0.

3. (a) Log-likelihood funktionen ges av

$$\begin{aligned} l(p) = \log L(p) &= \log \prod_{i=1}^{10} \binom{k_i - 1}{r - 1} p^r (1-p)^{k_i - r} \\ &= \sum_{i=1}^{10} \log\left(\binom{k_i - 1}{r - 1} p^r (1-p)^{k_i - r}\right) \\ &= \sum_{i=1}^{10} \left(\log \binom{k_i - 1}{r - 1} + r\log(p) + (k_i - r)\log(1-p)\right) \end{aligned}$$

För att beräkna maximum likelihood-skattaren deriverar vi $l(p)$:

$$\begin{aligned} \frac{d}{dp} l(p) &= \sum_{i=1}^{10} \left(r\frac{1}{p} - (k_i - r)\frac{1}{1-p}\right) \\ &= \frac{10r}{p} - (\sum_{i=1}^{10} k_i - 10r)\frac{1}{1-p} \end{aligned}$$

Vi sätter derivatan till noll och löser ut p, detta ger skattningen

$$p = \frac{10r}{\sum_{i=1}^{10} k_i}$$

(b) Den negativa binomialfördelningen NB(k, p) beskriver det totala antalet försök som vi måste göra tills att k har lyckats.

4. (a) Det är ett försök med så kallad ensidig indelning, där Faktor A är färgen på sidan. Vi ansätter modellen

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

där $\varepsilon_{ij} \sim \mathsf{N}(0, \sigma^2)$, μ är det gemensamma väntevärdet för alla observationer, och α_i anger förändringen i vätevärdet som orsakas av färgen på hemsidan.

(b) Vi har

$$\bar{y} = \frac{1}{55 + 47 + 33 + 68}(55 \cdot 33 + 47 \cdot 27 + 33 \cdot 45 + 68 \cdot 40) = 35.9064$$

och

$$\begin{aligned} SS_A &= \sum_{ij} (\bar{y}_i - \bar{y})^2 \\ &= 55(33 - 35.9064)^2 + 47(27 - 35.9064)^2 + 33(45 - 35.9064)^2 + 68(40 - 35.9064)^2 \\ &= 8061.2 \end{aligned}$$

Med $SS_E = SS_{Tot} - SS_A = 9938.8$ blir variansanalystabellen

Variation	Kvadratsumma	Frihetsgrader	Medelkvadratsumma	Teststorhet
Faktor A	8061.2	3	2687.1	53.8022
Residual	9938.8	199	49.9436	
Total	18000	202		

(c) Vi vill utföra testet

$$\begin{aligned} &H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \\ &H_1 : \text{åtminstone ett } \alpha_i \neq 0 \end{aligned}$$

Teststorheten är $F(3, 196)$-fördelad. Enligt formelsamlingen har vi

$$F_{0.05}(3, 199) \approx F_{0.05}(3, \infty) = 2.6507.$$

Det kritiska området för testet är $C_\alpha = \{T : T > 2.6507\}$. Eftersom $T_{obs} = 53.8022$ ligger i C_α kan vi förkasta nollhypotesen.

5. (a) Vi vill testa $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$. Eftersom konfidensintervallet för β_1 täcker noll så kan vi inte förkasta H_0 på nivå $\alpha = 0.05$. Vi kan alltså inte hävda att avstånd har en signifikant påverkan.

(b) Den totala variationen i datan, S_{yy}, kan delas upp i två delar:

$$S_{yy} = Q_0 + SSR,$$

där Q_0 är residualvariationen och SSR är variationen som förklaras av linjen. R^2 definieras som andelen av variationen som förklaras av linjen:

$$R^2 = \frac{SSR}{S_{yy}}.$$

(c) Modell 2 passar bäst. Residualerna för den modellen ser oberoende ut och verkar inte uppvisa något speciell struktur. Residualerna för Modell 1 har en klar struktur som funktion av avstånd. Normalfördelningsplotten ser bättre ut för Modell 2 och den har högre förklaringsgrad än Modell 1.

(d) Den förväntade logaritmerade effekten ges av $\beta_0 + 7\beta_1 = -15.74$ och ett konfidensintervall beräknas som

$$I = \left(\beta_0 + 7\beta_1 \pm t_{0.025}(13) s \sqrt{\frac{1}{n} + \frac{(7-4.89)^2}{154.66}} \right) = (-17.47, -14.09)$$

(e) (i) Sant, (ii) sant, (iii) falskt, (iv) sant.

6. (a) Den geometriska fördelningen beskriver fördelningen av antalet Bernoulliförsök tills det första lyckade. Den negativa binomialfördelningen $\mathrm{NB}(k, p)$ beskriver det totala antalet försök som vi måste göra tills att k har lyckats, och alltså är $\mathrm{Ge}(p)$-fördelningen samma som $\mathrm{NB}(1, p)$-fördelningen.

(b) Vi har

$$\begin{aligned} \mathsf{E}(Y_i) = \mathsf{P}(Y_i = 1) = \mathsf{P}(X_i < 3) &= \mathsf{P}(X_i = 1) + \mathsf{P}(X_i = 2) \\ &= p + p(1-p) = 0.5 + 0.5^2 = 0.75. \end{aligned}$$

Y är $\mathrm{Bin}(10, 0.75)$-fördelad, och vi har därför att $\mathsf{E}(Y) = 10 \cdot 0.75 = 7.5$.

(c) Vi har $\mathsf{E}(Y) = 10(p + p(1-p))$ och får momentskattningen av p genom att lösa ut p ur ekvationen $\mathsf{E}(Y) = 7$:

$$10(p + p(1-p)) = 7 \Rightarrow 2p - p^2 = 0.7 \Rightarrow p = 1 \pm \sqrt{1 - 0.7}.$$

En av de två lösningarna är större än 1 och alltså orimlig, vi har därför skattningen $p = 1 - \sqrt{0.3} = 0.4523$.

(d) Vi vill beräkna sannolikheten att få något som är åtminstone så extremt som $Y = 7$ givet att $p = 0.5$. Om $p = 0.5$ har vi från (a) att $Y \sim \mathrm{Bin}(10, 0.75)$ och sannolikheten är därför

$$\begin{aligned} \mathsf{P}(Y \geq 7 | H_0) &= \mathsf{P}(Y = 7) + \mathsf{P}(Y = 8) + \mathsf{P}(Y = 9) + \mathsf{P}(Y = 10) \\ &= \binom{10}{7} 0.75^7 0.25^3 + \binom{10}{8} 0.75^8 0.25^2 + \binom{10}{9} 0.75^9 0.25^1 + \binom{10}{10} 0.75^{10} 0.25^0 \\ &= 0.2503 + 0.2816 + 0.1877 + 0.0563 = 0.7759 \end{aligned}$$

Eftersom p-värdet är större än α kan vi inte förkasta H_0.