LÖSNINGAR TILL

Matematisk statistik
Matematikcentrum
Lunds tekniska högskola
Lunds universitet

Tentamen: 2015–08–18 kl 8^{00}–13^{00}
FMS 086 — Matematisk statistik för B, K, N och BME, 7.5 hp
MASB02 — Matematisk statistik för kemister, 7.5 hp

---

1. Låt X_i vara längden av kansten nr i och Y den totala längden av de 100 kantstenar stensättaren hinner på en dag, dvs

$$Y = \sum_{i=1}^{100} X_i,$$

$X_1, X_2, \ldots, X_n$ är **oberoende och likafördelade** och $n = 100$ är ganska **stort**. Alltså säger **centrala gränsvärdessatsen** att Y är ungefär normalfördelad, $Y \underset{\sim}{\in} N(E(Y), V(Y))$.

$$\begin{aligned} E(Y) &= E(\sum_{i=1}^{100} X_i) = \sum_{i=1}^{100} E(X_i) = 100 \cdot 0.5 = 50 \\ V(Y) &= V(\sum_{i=1}^{100} X_i) = \sum_{i=1}^{100} V(X_i) = 100 \cdot 0.06^2 = 0.36. \\ Y &\underset{\sim}{\in} N(50, 0.36) \end{aligned}$$

Den sökta sannolikheten blir

$$P(Y > 49) = 1 - P(Y \le 49) \approx 1 - \Phi(\frac{49 - 50}{\sqrt{0.36}}) = 1 - \Phi(-1.67) \approx 0.95$$

2. (a) Skattning av huvud- och samspelseffekter är

$$\begin{aligned} \widehat{A} &= \frac{-\mu_{11}^* + \mu_{21}^* - \mu_{12}^* + \mu_{22}^*}{2^2} = 3.10, \\ \widehat{B} &= \frac{-\mu_{11}^* - \mu_{21}^* + \mu_{12}^* + \mu_{22}^*}{2^2} = 3.90, \\ \widehat{AB} &= \frac{\mu_{11}^* - \mu_{21}^* - \mu_{12}^* + \mu_{22}^*}{2^2} = -2.95, \end{aligned}$$

(b) Vi ansätter en model där observationerna är

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad i = 1, 2;\ j = 1, 2;\ k = 1, \ldots, n$$

med **oberoende** och **normalfördelade** fel, $\varepsilon_{ijk} \in N(0, \sigma^2)$.
Den poolade variansskattningen är

$$s^2 = \frac{s_{11}^2 + s_{21}^2 + s_{12}^2 + s_{22}^2}{2^2} = 1.87$$

med $f = 2^2 \cdot (n - 1) = 4$ frihetsgrader, och medelfelet för respektive effekt är

$$d(\text{effekt}) = \frac{s}{\sqrt{2^2 n}} = \sqrt{\frac{1.87}{4 \cdot 2}} = 0.4835$$

Ett konfidensintervall för respektive effekt ges av

$$I_{\text{effekt}} = \widehat{\text{effekt}} \pm t_{0.025}(f) d(\text{effekt}) = \widehat{\text{effekt}} \pm 2.7764 \cdot 0.4835 = \widehat{\text{effekt}} \pm 1.3423,$$

och effekterna är signifikanta om $\left|\widehat{\text{effekt}}\right| > 1.3423$ (intervaller kommer då **inte** att innehålla 0).
Alla effekter är **signifikant**.

(c) Samspelseffekten indikerar hur två faktorer samverkar. En **negativ samspelseffekt** indikerar att den totala effekten av två höga/positiva faktorer är **lägre** än vad man kan vänta sig genom att bara addera de enskilda effekterna.

3. Låt $x_i, i = 1, \ldots, n_x = 5$ vara observationerna från fiskart 1 och $y_i, i = 1, \ldots, n_y = 6$ från fiskart två. Då har vi följande medelvärden och stickprovsstandardavvikelser

$$\bar{x} = 2.482, \quad \bar{y} = 2.0183, \quad s_x = 0.131, \quad s_y = 0.111$$

(a) Eftersom observationerna för både fiskart 1 och 2 antogs komma från normalfördelningar med gemensam varans σ^2, kan denna skattas med den sammanvägda variansskattningen

$$(\sigma^2)^* = s_p^2 = \frac{(5-1)s_x^2 + (6-1)s_y^2}{5-1+6-1} = 0.144$$

som har $f = 5 - 1 + 6 - 1 = 9$ frihetsgrader.

(b) Med denna och en skattning av kvicksilverhalten för fiskart 1, $\mu_1^* = \bar{x} = \frac{1}{5}\sum_1^5 x_i = 2.482$, fås ett 95% konfidensintervall till

$$I_{\mu_1} = \mu_1^* \pm t_{\alpha/2}(f)d(\mu_1^*) = \bar{x} \pm t_{0.025}(9)\frac{s_p}{\sqrt{n_x}} = 2.482 \pm 2.26 \cdot \sqrt{\frac{0.0144}{5}} = (2.36, 2.60)$$

(c) För den tredje fiskarten har vi $z = 3.13$ som är en observation av $Z \in N(\mu_3, \sigma^2)$, där σ^2 är densamma som tidigare. En skattning av medelmängden kvicksilver för denna fiskart blir $\mu_3^* = z = 3.13$ och ett konfidensintervall blir

$$I_{\mu_3} = \mu_3^* \pm t_{\alpha/2}(f)d(\mu_3^*) = z \pm t_{0.025}(9)\frac{s_p}{\sqrt{n_z}} = 3.13 \pm 2.26 \cdot \sqrt{\frac{0.0144}{1}} = (2.86, 3.40)$$

4. (a) Parametern μ är **medelantalet jordskalv per år**.

(b) $X \in Po(2.6)$. Sannolikheten att det blir högst två jordskalv under ett år blir

$$P(X \leq 2) = F_X(2) = [\text{Tabell 7}] = 0.5184$$

(c) Vi har fem observationer av $X \in Po(\mu)$. Summan av observationerna, 21, är en observation av $Y \in Po(5\mu)$. För att avgöra om detta är orimligt högt om $\mu = 2.6 \iff 5\mu = 13$ kan vi testa

$$H_0: \ 5\mu = 13$$
$$H_1: \ 5\mu > 13$$

Eftersom vi ligger något **under gränsen för normalapproximation** om H_0 är sann är det lämpligast att utföra testet med **direktmetoden**.

$$P = P(\text{Få det vi fått eller värre om } H_0 \text{ är sann}) = P(Y \geq 21 \text{ om } Y \in Po(13)) =$$
$$= 1 - P(Y \leq 20) = [\text{Tabell 7}] = 0.025$$

Vi kan förkasta H_0 på nivån $\alpha = 0.05$ eftersom P-värdet underskrider denna gräns. Det är alltså troligt att μ är större än 2.6.

5. (a) Skattningar av β_i ges av

$$\begin{bmatrix} \beta_0^* \\ \beta_1^* \\ \beta_2^* \end{bmatrix} = (X^TX)^{-1}X^TY \approx \begin{bmatrix} 0.1954 \\ 0.2385 \\ 0.1265 \end{bmatrix}$$

För variansen har vi att

$$s^2 = (\sigma_2^2)^* = \frac{Q_0'}{n-(p+1)} = \frac{90.77}{14-(2+1)} \approx 8.2518$$

med $f = 14 - 3 = 11$ frihetsgrader.

(b) Enligt modellen för linjärregression är skattningarna normalfördelade med varians $\sigma^2(X^TX)^{-1}_{ii}$. Tvåsidiga 95%-konfidensintervall för β_1:

$$\beta_1^* \pm t_{0.05/2}(n-3) \cdot \mathsf{d}(\beta_1^*) \approx 0.2385 \pm 2.201\sqrt{8.2518 \cdot 3.82 \cdot 10^{-5}} \approx \begin{bmatrix}0.1995 & 0.2776\end{bmatrix}$$

och β_2:

$$\beta_2^* \pm t_{0.05/2}(n-3) \cdot \mathsf{d}(\beta_2^*) \approx 0.1265 \pm 2.201\sqrt{8.2518 \cdot 4.46 \cdot 10^{-5}} \approx \begin{bmatrix}0.0843 & 0.1687\end{bmatrix}$$

Då både β_1 och β_2 är **signifikanta** kan vi **inte förenkla** modellen.

(c) Vi vill beräkna värdet av regressionslinjen då

$$x^0 = \begin{bmatrix}1\\75\\20\end{bmatrix}$$

En skattning av medelvärdet ges av

$$\mu^*(x^0) = \beta_0^* + 75\beta_1^* + 20\beta_2^* \approx 20.62$$

Och ett 95%-intervall för den **förväntade** upptagningsförmågan blir

$$\begin{aligned}\mu^*(x^0) \pm t_{0.05/2}(n-3) \cdot \mathsf{d}(\mu^*(x^0)) &= \mu^*(x^0) \pm t_{0.05/2}(n-3) \cdot \sqrt{s^2 \cdot (x^0)^\top (X^TX)^{-1}x^0} \\ &\approx 20.61 \pm 2.201 \cdot \sqrt{8.2518 \cdot 0.0768} \\ &\approx \begin{bmatrix}18.87 & 22.37\end{bmatrix}\end{aligned}$$

6. (a) i) Den sökta sannolikheten ges av

$$P(X < 2) = P\left(\underbrace{\frac{X-3}{2}}_{N(0,1)} < \underbrace{\frac{2-3}{2}}_{-1/2}\right) = \Phi(-0.5) = 1 - \Phi(0.5) \approx 0.3085$$

ii) Z är **normalfördelade** eftersom det är en **linjärkombination** av normal variabler. Väntevärde och varians ges av

$$\begin{aligned}E(Z) &= E(2X - Y) = 2E(X) - E(Y) = 2 \cdot 3 - 1 = 5 \\ V(Z) &= V(2X - Y) = 2^2V(X) + (-1)^2V(Y) = 2^2 \cdot 2^2 + 3^2 = 25\end{aligned}$$

och $Z \in N(5, 25)$.

(b) A och B är oberoende händelser om sannolikheten att den ena skall inträffa inte beror på om den andra inträffat eller ej.

A och B är disjunkta eller oförenliga händelser om de inte kan inträffa samtidigt. (De är därmed inte oberoende).

(c) i) En fullständigt randomiserad försöksuppställning fås genom att **helt slumpmässigt** placera de åtta provrören på de åtta tillgängliga platserna.

ii) I ett randomiserat block försök väljs slumpmässigt **två provrör med bakterie 1 och två med bakterie 2** (tex A,B och F,H) dessa fyra rör slumpas närmst väggen (plats 1–4), resterande fyra rör slumpas på plats 5–8. Målsättningen är att **lika många rör** med respektive bakteriestam ska utsättas för **samma temperatur**.

(d) i) Då testet har lägst sannolikhet att förkasta $\mu = 4$ så testas $H_0 : \mu = 4$ mot $H_1 : \mu \neq 4$

ii) Signifikansnivån ges av sannolikheten att förkasta H_o då H_o är sann, dvs $\alpha = h(4) = 0.1$.

(e) Gauss approximations formler ger väntevärde för R:

$$\mathrm{E}(R) = \mathrm{E}\left(-\frac{\Delta C}{\Delta t}\right) \approx -\frac{\Delta C}{\mathrm{E}\,(\Delta t)} = -\frac{-1.5}{4} = 0.375$$

För variansen behöver vi först beräkna derivatan

$$R(\Delta t) = -\frac{\Delta C}{\Delta t} \qquad\qquad R'(\Delta t) = \frac{\Delta C}{(\Delta t)^2}$$

och varians är

$$\mathrm{V}(R) \approx R'\,(\mathrm{E}\,(\Delta t))^2\,\mathrm{V}(\Delta t) = \left(\frac{\Delta C}{(\mathrm{E}\,(\Delta t))^2}\right)^2 \mathrm{V}(\Delta t) \approx \left(\frac{-1.5}{4^2}\right)^2 \cdot 0.07^2 \approx 4.31 \cdot 10^{-5}$$