UPPSALA UNIVERSITET
Matematiska institutionen
Niclas Petersson

PROV I STATISTIK
Statistik IP1
2006-03-22

Lösningar till tenta 2006-03-22:

1. Decibelmätare.

   Definiera händelserna: A =*Mätare* A *ligger i lådan*, B =*Mätare* B *ligger i lådan* samt C =*Mätare* C *ligger i lådan*. Det är givet att händelserna är oberoende samt att

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{1}{3} \quad \text{och} \quad \mathbb{P}(C) = \frac{1}{4}.$$

   (a) Sannolikheten att alla mätare ligger i lådan:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{\mathbf{1}}{\mathbf{24}}.$$

   (b) Sannolikheten att det finns minst en decibelmätare i lådan:

$$\mathbb{P}(A \cup B \cup C) = 1 - \mathbb{P}(A^* \cap B^* \cap C^*) = 1 - \mathbb{P}(A^*) \cdot \mathbb{P}(B^*) \cdot \mathbb{P}(C^*) = 1 - \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} = \frac{\mathbf{3}}{\mathbf{4}}.$$

   (c) Sannolikheten att mätare C ligger i lådan givet att åtminstone en mätare finns i lådan:

$$\mathbb{P}(C \mid A \cup B \cup C) = \frac{\mathbb{P}(C \cap (A \cup B \cup C))}{\mathbb{P}(A \cup B \cup C)} = \frac{\mathbb{P}(C)}{\mathbb{P}(A \cup B \cup C)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{\mathbf{1}}{\mathbf{3}}.$$

2. Diskret stokastisk variabel X.

   (a) Bestäm konstanten c.

$$1 = \sum_{x=1}^{3} p(x) = \frac{1}{9} \sum_{x=1}^{3} (x + c) = \frac{1}{9}(1 + 2 + 3 + 3c) = \frac{2 + c}{3} \qquad \Rightarrow \qquad \mathbf{c = 1}.$$

   (b) Beräkna väntevärde och varians för X.

$$\mathbb{E}(X) = \sum_{x=1}^{3} x p(x) = \frac{1}{9} \sum_{x=1}^{3} x(x+1) = \frac{1}{9}(1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4) = \frac{\mathbf{20}}{\mathbf{9}}.$$

$$\mathbb{E}(X^2) = \sum_{x=1}^{3} x^2 p(x) = \frac{1}{9} \sum_{x=1}^{3} x^2(x+1) = \frac{1}{9}(1^2 \cdot 2 + 2^2 \cdot 3 + 3^2 \cdot 4) = \frac{50}{9}$$

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{\mathbf{50}}{\mathbf{81}}$$

   (c) Beräkna $\mathbb{P}(\{X = 2\} \mid \{X \neq 3\})$:

$$\begin{aligned}\mathbb{P}(\{X = 2\} \mid \{X \neq 3\}) &= \frac{\mathbb{P}(\{X = 2\} \cap \{X \neq 3\})}{\mathbb{P}(X \neq 3)} = \frac{\mathbb{P}(X = 2)}{1 - \mathbb{P}(X = 3)} = \\ &= \frac{p(2)}{1 - p(3)} = \frac{\frac{3}{9}}{1 - \frac{4}{9}} = \frac{\mathbf{3}}{\mathbf{5}}\end{aligned}$$

3. Registreringsplåtar.

   Det är rimligt att anta att alla nummer förekommer med samma frekvens.

   (a) $X \in Po(8)$.

   (b) $Y \tilde{\in} N(960, 960)$, ty $Y \in Po(8 \cdot 12 \cdot 10) = Po(960)$ och $Po(\lambda) \sim N(\lambda, \lambda)$ när $\lambda \geq 15$.

   (c) Beräkna $\mathbb{P}(Y \geq 999)$.

$$\begin{aligned} \mathbb{P}(Y \geq 999) &= 1 - \mathbb{P}(Y \leq 998) \approx 1 - \Phi(\frac{998 + 0.5 - 960}{\sqrt{960}}) = \\ &= 1 - \Phi(1.24) = 1 - 0.8925 \approx \mathbf{0.11} \end{aligned}$$

4. Maskin av typen *EPI Automatic*.

   Två stickprov med $n_x = 5$ respesktive $n_y = 6$ observationer.

   (a) Skatta μ_x med $\bar{x}$ och σ_x^2 med s_x^2.

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \mathbf{23.8}, \qquad s_x^2 = \frac{1}{5-1}(\sum_{i=1}^{5} x_i^2 - \frac{1}{5}(\sum_{i=1}^{6} x_i)^2) = \mathbf{28.2}.$$

   Skatta μ_y med $\bar{y}$ och σ_y^2 med s_y^2.

$$\bar{y} = \frac{1}{6}\sum_{i=1}^{6} y_i = \mathbf{44.8}, \qquad s_y^2 = \frac{1}{6-1}(\sum_{i=1}^{6} y_i^2 - \frac{1}{6}(\sum_{i=1}^{6} y_i)^2) = \mathbf{206}.$$

   (b) 95% konfidensintervall för μ_x: ($\alpha = 0.05$, använd $t_{n_x-1,\frac{\alpha}{2}} = t_{4,0.025}$)

$$\begin{aligned} I_{\mu_x} &= (\bar{x} \pm t_{4,0.025}\frac{s_x}{\sqrt{n_x}}) = (23.8 \pm 2.776\sqrt{\frac{28.2}{5}}) \\ &= (23.8 \pm 6.6) = (\mathbf{17.2}, \mathbf{30.4}) \end{aligned}$$

   (c) Skatta variansen σ^2.

$$s^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + x_y - 2} = \frac{4 \cdot 28.2 + 5 \cdot 206}{9} = \mathbf{127}$$

   (d) 95% konfidensintervall för $\mu_x - \mu_y$: ($\alpha = 0.05$, använd $t_{n_x+n_y-2,\frac{\alpha}{2}} = t_{9,0.025}$)

$$\begin{aligned} I_{\mu_x-\mu_y} &= (\bar{x} - \bar{y} \pm t_{9,0.025}\sqrt{s^2(\frac{1}{n_x} + \frac{1}{n_y})}) = (-21.0 \pm 2.262\sqrt{127(\frac{1}{5} + \frac{1}{6})}) \\ &= (-21.03 \pm 15.43) = (\mathbf{-36.5}, \mathbf{-5.60}) \end{aligned}$$

   (e) Uppvärmningstiden för maskinen är med 95% säkerhet längre då den brukats i 3 års tid ty intervallet (-36.5, -5.61) innehåller inte värdet 0.

5. Genomsnittlig månadshyra för 2:a med kök.

   Beräkna:

$$\bar{x} = 2001, \qquad \bar{y} = 3743.43$$

och

$$S_{xx} = \sum_{i=1}^{7} x_i^2 - \frac{1}{7}(\sum_{i=1}^{7} x_i)^2 = 28, \qquad S_{yy} = \sum_{i=1}^{7} y_i^2 - \frac{1}{7}(\sum_{i=1}^{7} y_i)^2 = 171517.7$$

$$S_{xy} = \sum_{i=1}^{7} x_i y_i - \frac{1}{7}(\sum_{i=1}^{7} x_i)(\sum_{i=1}^{7} y_i) = 2106$$

   (a) Korrelationskoefficienten, r, tyder på ett mycket starkt linjärt samband ty

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot s_{yy}}} \approx \mathbf{0.96}.$$

   (b) Skatta β, α och σ^2 med:

$$b = \frac{S_{xy}}{S_{xx}} = \mathbf{75.2143}, \qquad a = \bar{y} - b\bar{x} = -\mathbf{146760}$$

$$s^2 = \frac{1}{n-2}(S_{yy} - \frac{S_{xy}^2}{S_{xx}}) = \mathbf{2623}$$

   Den skattade linjens ekvation blir således $y = -146760 + 75.2143x$.

   (c) Vilket år är genomsnittliga månadshyran ca 4500 kr?

$$4500 = -146760 + 75.2143 \cdot x \qquad \Rightarrow \qquad x = 2011.06 \approx \mathbf{2011}$$

   (d) 90% prediktionsintervall för genomsnittlig månadshyra år 2006: (Använd $t_{n-2,0.10/2} = t_{5,0.05}$-fördelningen)

$$y(2006) = -146760 + 75.2143 \cdot 2006 = 4119.5$$

$$\begin{aligned} I_y &= (y(2006) \pm t_{5,0.05} \cdot s\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}) = \\ &= (4119.5 \pm 2.015 \cdot \sqrt{2623}\sqrt{1 + \frac{1}{7} + \frac{(2006-2001)^2}{28}}) = \\ &= (4119.5 \pm 147.3) = (\mathbf{3970}, \mathbf{4270}) \end{aligned}$$