UPPSALA UNIVERSITET
Matematiska institutionen
Måns Thulin

TENTAMEN I MATEMATISK STATISTIK
Statistik för ingenjörer 1MS008
2012-03-13

PRELIMINÄRA LÖSNINGSFÖRSLAG

1. (a) Vi söker $\mathsf{P}(B|A) = \mathsf{P}(A \cap B)/\mathsf{P}(A)$. Vi har att $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$, så $\mathsf{P}(A \cap B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cup B) = 0.9 + 0.15 - 0.95 = 0.1$. Så $\mathsf{P}(B|A) = 0.1/0.9 = 1/9$.

   (b) $\mathsf{P}(X \leq 2) = \mathsf{P}(X = 0) + \mathsf{P}(X = 1) + \mathsf{P}(X = 2) = 0.3 + 0.4 + 0.1 = 0.8$.

   (c) Datamaterialets median är 4.95, vilket svarar mot det tjocka strecket inuti lådan. Därmed måste lådagram A höra till datamaterialet (de övriga värdena i lådagrammet stämmer också med givna data!).

   (d) $\mathsf{P}(X \leq 5) = \mathsf{P}((X-4)/\sqrt{3} \leq 1/\sqrt{3}) \approx \Phi(0.58) = 0.7190$.

   (e) $\mathsf{P}(X \geq 2) = 1 - \mathsf{P}(X \leq 1) = \mathsf{P}(X = 0) + \mathsf{P}(X = 1) = 1 - \left(\frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2}\right) = 1 - 3e^{-2} \approx 0.594$.

   (f) Centrala gränsvärdessatsen ger att Y *approximativt* tillhör en normalfördelning med väntevärdet $\mu = 1 \cdot 1500 = 0$ och variansen $\sigma^2 = 4 \cdot 1500 = 6000$.

   (g) $t_{0.05}(9) = 1.83$, så ett 90 % konfidensintervall ges av $I_\mu = 10.1 \pm 1.83 \cdot \sqrt{0.9}/\sqrt{10} \approx (9.551, 10.649)$.

   (h) $\hat{k} = S_{xy}/S_{xx} \approx 0.496$.

2. Vi antar att mätresultaten för de båda legeringarna är normalfördelade med väntevärde μ_1 respektive μ_2 och samma varians σ^2, samt att alla mätningar är oberoende av varandra. μ_i är den faktiska värmekonduktiviteten för legering i (avvikelser från μ_i antas vara mätfel).

   Den sammanvägda variansskattningen blir

$$s_p^2 = \frac{29s_1^2 + 29s_2^2}{30 + 30 - 2} = 6.525.$$

   Ett 95 % konfidensintervall för $\mu_1 - \mu_2$ ges av

$$I_{\mu_1 - \mu_2} = \bar{x}_1 - \bar{x}_2 \pm t_{0.025}(58) \cdot s_p \cdot \sqrt{1/30 + 1/30} \approx (0.98, 3.62),$$

   där $t_{0.025}(58) \approx t_{0.025}(60) = 2.00$. Eftersom hela intervallet ligger ovanför 0 så är det statistiskt säkerställt att μ_2 är lägre än μ_1.

3. (a) Givet är att $E(X_i) = 2$, $V(X_i) = 4$ och att olika signaler X_i är oberoende. För $Y = X_1 + X_2 + \cdots + X_{200}$ fås då genom räknereglerna för väntevärde och varians att

$$E(Y) = E(X_1) + \cdots + E(X_{200}) = 400 \qquad \text{och på grund av oberoendet}$$
$$E(Y) = V(X_1) + \cdots + V(X_{200}) = 800.$$

   (b) Centrala gränsvärdessatsen ger att $Y \approx N(400, 800)$. Alltså är

$$\mathsf{P}(Y > 490) = 1 - \mathsf{P}(Y \leq 490) = 1 - \mathsf{P}\left(\frac{X - 400}{\sqrt{800}} \leq \frac{490 - 400}{\sqrt{800}}\right) \approx 1 - \Phi(3.18).$$

   $\Phi(3.18) = 0.9993$, så $\mathsf{P}(Y > 490) \approx 1 - 0.9993 = 0.0007$.

(c) Nu ger centrala gränsvärdessatsen istället att $Y \approx N(190 \cdot 3, 190 \cdot 4) = N(570, 760)$. Vi får

$$\mathsf{P}(Y < 490) \approx \Phi\Big(\frac{570 - 490}{\sqrt{760}}\Big) \approx \Phi(-2.90) = 1 - \Phi(2.90) = 1 - 0.9981 = 0.0019.$$

4. (a) $\hat{k} = S_{xy}/S_{xx} \approx 1.009$ och $\hat{m} = \bar{y} - \hat{k}\bar{x} \approx -3.383$.

   (b) $R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.969$. Eftersom förklaringsgraden ligger nära 1 så drar vi slutsatsen att den anpassade linjen ger en bra beskrivning av hur entrepenadkostnaden beror på arean.

   (c) Standardavvikelsen σ skattas med

$$s = \sqrt{\frac{1}{3}\Big(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\Big)} = 6.278\ldots$$

   Ett 95 % prediktionsintervall då $x_0 = 90$ ges av $\hat{k} \cdot 90 + \hat{m} \pm t_{0.025}(3) \cdot s \cdot \sqrt{1 + \frac{1}{5} + \frac{(90-\bar{x})^2}{S_{xx}}} \approx (63.65, 111.15)$.

5. (a) Vi får skattningen $\hat{p}_1 = 149/1500 \approx 0.099$.

   (b) Medelfelet är $d(\hat{p}_1) = \sqrt{\hat{p}_1(1 - \hat{p}_1/1500)} \approx 0077$.

   (c) Eftersom $\hat{p}_1(1 - \hat{p}_1)1500 \approx 134 > 10$ så är tumregeln för när det vanliga konfidensintervallet kan användas uppfylld, vilket innebär att ett approximativt 95 % konfidensintervall för andelen ges av

$$\hat{p}_1 \pm \lambda_{0.025}\sqrt{\hat{p}_1(1 - \hat{p}_1)/1500} \approx (0.084, 0.114).$$

6. Andelen p_2 bland personer som inte bor nära vindkraftverk skattas som $\hat{p}_2 = 180/2000 = 0.09$. Vi får då $\hat{p}_2(1 - \hat{p}_2)2000 \approx 164 > 10$, så tumregeln för när konfidensintervallet för skillnad i andelar får användas är uppfylld. Vid 95 % konfidensgrad blir intervallet

$$\hat{p}_1 - \hat{p}_2 \pm \lambda_{0.025}\sqrt{\hat{p}_1(1 - \hat{p}_1)/1500 + \hat{p}_2(1 - \hat{p}_2)/2000} \approx (-0.010, 0.029).$$

   Eftersom 0 ingår i intervallet så kan vi inte utesluta att det inte finns någon skillnad mellan andelarna. Vi har därför inte några belägg för att säga att andelen personer som ofta lider av kraftig huvudvärk är större nära vindkraftverk.