UPPSALA UNIVERSITET
Matematiska institutionen
Raimundas Gaigalas

Statistik IP
2004-06-03

Lösningar till tenta 040603

1. Inför händelser E, A och D att respektive sträng går av. Då är E, A och D oberoende och $P(E) = 0.5$, $P(A) = 0.2$, $P(D) = 0.1$.

   (a) Sannolikheten att minst en av strängarna går av:

$$\begin{aligned} P(E \cup A \cup D) &= 1 - P(E^* \cap A^* \cap D^*) = 1 - P(E^*)P(A^*)P(D^*) \\ &= 1 - 0.5 \cdot 0.8 \cdot 0.9 = 0.64. \end{aligned}$$

   (b) Sannolikheten att exakt två av strängarna går av:

$$\begin{aligned} &P((E \cap A \cap D^*) \cup (E \cap A^* \cap D) \cup (E^* \cap A \cap D)) \\ &= P(E)P(A)P(D^*) + P(E)P(A^*)P(D) + P(E^*)P(A)P(D) \\ &= 0.5 \cdot 0.2 \cdot 0.9 + 0.5 \cdot 0.8 \cdot 0.1 + 0.5 \cdot 0.2 \cdot 0.1 = 0.14. \end{aligned}$$

   (c) $P(D|E \cup D) = \dfrac{P(D \cap (E \cup D))}{P(E \cup D)} = \dfrac{P(D)}{P(E \cup D)} = \dfrac{P(D)}{1 - P(E^* \cap D^*)}$
   $= \dfrac{0.1}{1 - 0.5 \cdot 0.9} = 0.1818.$

2. (a) Den totala massan för en sannolikhetsfunktion måste vara 1:

$$\begin{aligned} p(-1) + p(0) + p(1) + p(2) + p(3) &= \frac{1+c}{20} + \frac{0+c}{20} + \frac{1+c}{20} \\ &+ \frac{4+c}{20} + \frac{9+c}{20} = \frac{15+5c}{20} = 1 \quad \Leftrightarrow \quad c = 1. \end{aligned}$$

   Därmed $p(-1) = 0.1$, $p(0) = 0.05$, $p(1) = 0.1$, $p(2) = 0.25$, $p(3) = 0.5$.

   (b) $P(X = 0 \,|\, X \le 1) = \dfrac{P(X = 0, X \le 1)}{P(X \le 1)} = \dfrac{P(X = 0)}{P(X \le 1)}$
   $= \dfrac{0.05}{0.1 + 0.05 + 0.1} = 0.2.$

3. Låt X = vikten på en flygpassagerare, Y = vikten på hennes bagage, Z = vikten på 140 passagerare med bagage. Fördelningarna för X och Y är okända förutom att $E(X) = 75$, $D(X) = 18$, $E(Y) = 23$, $D(Y) = 7$. Men för $Z = \sum_{i=1}^{140} X_i + \sum_{i=1}^{140} Y_i$ gäller enligt centrala gränsvärdessatsen att $Z \overset{\cdot}{\sim} \mathrm{N}(140 \cdot 75 + 140 \cdot 23, 140 \cdot 18^2 + 140 \cdot 7^2) = \mathrm{N}(13720, 52220)$. Därmed

$$\begin{aligned} &P(13420 \le Z \le 13910) = P(Z \le 13910) - P(Z \le 13420) \\ &= P\Big(\frac{Z - 13720}{\sqrt{52220}} \le \frac{13910 - 13720}{\sqrt{52220}}\Big) - P\Big(\frac{Z - 13720}{\sqrt{52220}} \le \frac{13420 - 13720}{\sqrt{52220}}\Big) \\ &\approx \Phi\Big(\frac{13910 - 13720}{\sqrt{52220}}\Big) - \Phi\Big(\frac{13420 - 13720}{\sqrt{52220}}\Big) = \Phi(0.83) - \Phi(-1.31) \\ &= \Phi(0.83) - 1 + \Phi(1.31) = 0.7967 - 1 + 0.9049 = 0.7016. \end{aligned}$$

4. (a) X = antalet trasiga plattor bland de 90 valda $\sim \mathrm{Bin}(90, 0.02)$. Eftersom $p \leq 0.1$ och $n \geq 10$, kan Poissonapproximation tillämpas: $X \overset{\cdot}{\sim} \mathrm{Po}(1.8)$. Detta ger

$$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.7306 = 0.2694.$$

(b) X = antalet trasiga plattor bland de 900 i partiet $\sim \mathrm{Bin}(900, 0.02)$. Vi har $npq = 900 \cdot 0.02 \cdot 0.98 = 17.64 > 10$, fördelningen kan approximeras med normal: $X \overset{\cdot}{\sim} \mathrm{N}(18, 17.64)$. Därmed

$$\begin{aligned} P(X > 20) = 1 - P(X \leq 20) = 1 - P(\frac{X - 18}{\sqrt{17.64}} \leq \frac{20 - 18}{\sqrt{17.64}}) \\ \approx \quad 1 - \Phi(\frac{20 - 18 + 0.5}{\sqrt{17.64}}) = 1 - \Phi(0.60) = 1 - 0.7257 = 0.2743. \end{aligned}$$

5. (a) Vi approximerar värden i varje grupp med mittvärdena:
$\bar{x} = \frac{1}{24}(0.55 \cdot 5 + 0.65 \cdot 8 + 0.75 \cdot 4 + 0.85 \cdot 5 + 0.95 \cdot 2) = \frac{17.1}{24} = 0.7125$,
$s^2 = \frac{1}{23}(0.55^2 \cdot 5 + 0.65^2 \cdot 8 + 0.75^2 \cdot 4 + 0.85^2 \cdot 5 + 0.95^2 \cdot 2 - \frac{1}{24}17.1^2) = 0.01636$, $s = 0.1279$.

(b) X = effektiviteten, $x_1, x_2, \ldots, x_{24}$ ett stickprov från $X \sim \mathrm{N}(\mu, \sigma^2)$, där μ och σ är okända. Punktskatting $\hat{\mu} = \bar{x}$, spridningsmått $D(\bar{X}) = \frac{\sigma}{\sqrt{n}}$. σ är okänd, skattas med s. Testvariabeln $\dfrac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ är $t(n-1)$-fördelad.

Ett 95 % konfidensintervall för μ blir

$$I_\mu = (\bar{x} \pm t_{0.025}(23)\frac{s}{\sqrt{24}}) = (0.7125 \pm 2.069 \cdot \frac{0.1279}{\sqrt{24}}) = (0.6585, 0.7665).$$

Eftersom 0.65 inte tillhör intervallet, är den genomsnittliga effektiviteten inte lika 65 % med 95 % säkerhet.

(c) Eftersom antalet observationer är större än 20, blir $\bar{X}$ enligt centrala gränsvärdessatsen approximativt normalfördelad. Testvariabeln $\dfrac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ blir i det här fallet approximativt $\mathrm{N}(0, 1)$-fördelad, fast en bättre approximation kan fås med $t(23)$-fördelningen. Därför är det möjligt att behålla samma konfidensintervall som ovan eller så kan $\lambda_{0.025}$-kvantilen användas i stället:

$$I_\mu = (\bar{x} \pm \lambda_{0.025}\frac{s}{\sqrt{24}}) = (0.7125 \pm 1.96 \cdot \frac{0.1279}{\sqrt{24}}) = (0.6613, 0.7637),$$

vilket leder till samma slutsats som det andra intervallet.

6. Enligt påståendet borde säkerhetshål förekomma enligt proportionerna $6 : 3 : 2$. Alla proportioner ska testas samtidigt, därför är χ^2-test för anpassning lämpligast. Vi testar $H_0: \; p_1 = \frac{6}{11}, p_2 = \frac{3}{11}, p_3 = \frac{2}{11}$ mot H_1: andra sannolikheter förekommer. Testvariabeln

$$Q = \sum_{i=1}^{3} \frac{(x_i - np_i)^2}{np_i} = \sum_{i=1}^{3} \frac{x_i^2}{np_i} - n$$

är $\chi^2(3-1) = \chi^2(2)$-fördelad om H_0 är sann. Test: om $Q > \chi^2_{0.05}(2) = 5.991$, förkasta H_0. I vårt fall

$$Q = \frac{205^2}{207.8182} + \frac{95^2}{103.9091} + \frac{81^2}{69.2727} - 381 = 2.7874 < 5.991,$$

H_0 kan ej förkastas. Data tyder inte på att expertens påstående skulle vara felaktigt.

7. Ett linjärt samband beskrivs väl med linjär regression. Vi väljer X = kaloriintag, Y = kroppsvikt. $x_1, \ldots, x_7$ är givna, $y_1, \ldots, y_7$ är observationer av $Y_i = \alpha + \beta x_i + \epsilon_i$, där ϵ_i är $\mathrm{N}(0, \sigma^2)$-fördelade.

   (a) Graden av linjärt samband mäts med korrelationskoefficienten $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$. Vi har $S_{xx} = 0.9886$, $S_{xy} = 9.52286$, $S_{yy} = 94.5943$ och $r = 0.9848$. Ett starkt linjärt samband.

   (b) Den skattade regressionslinjen: $y = \hat{\alpha} + \hat{\beta}x$, $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ och $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. $\bar{x} = 2.0143$, $\bar{y} = 62.8714$, $\hat{\beta} = 9.6329$, $\hat{\alpha} = 43.4679$. Därmed $y = 43.4679 + 9.6329x$.

   (c) Insättning av $x = 2.9$ i den skattade linjens ekvation ger en skattning $y = 43.4679 + 9.6329 \cdot 2.9 = 71.4035$ kg för kroppsvikten. Detta är ett rimligt värde för de givna data, men skillnaden $71.4 - 57.5 = 13.9$ kg är ett farligt högt värde för en kroppsbyggare att gå upp på $7-8$ månader enligt experter...

8. X = antalet tvillingfödslar i Twin Peaks under 5 år $\sim \mathrm{Po}(5\lambda)$, λ okänd. $x = 9$ är en observation av X. Vi testar $H_0:\ \lambda = 0.8$ mot $H_1:\ \lambda > 0.8$. Direktmetoden:

$$\begin{aligned} P &= P(X \geq 9;\ \lambda = 0.8) = P(X \geq 9;\ 5\lambda = 4) = 1 - P(X \leq 8) \\ &= 1 - 0.9786 = 0.0214 < 0.05, \end{aligned}$$

H_0 förkastas. Med 95 % säkerhet har antalet tvillingfödslar ökat.

I den alternativa uppgiften är X = antalet tvillingfödslar under 24 år $\sim \mathrm{Po}(24\lambda)$, λ okänd, $x = 31$ en observation av X. Vi ska bilda ett konfidensintervall för 24λ först. Punktskattning $\widehat{24\lambda} = x = 31$. Spridningsmått $D(X) = \sqrt{24\lambda}$ beror på λ, skattas med $\sqrt{x}$. Eftersom $\widehat{24\lambda} = 31 > 15$, kan normalapproximation tillämpas. Testvariabeln $\dfrac{X - 24\lambda}{\sqrt{X}}$ är approximativt $\mathrm{N}(0,1)$-fördelad. Detta ger

$$I_{24\lambda} = (x \pm \lambda_{0.025}\sqrt{x}) = (31 \pm 1.96\sqrt{31}) = (20.0872, 41.9128).$$

Ett konfidensintervall för λ blir

$$I_\lambda = (\frac{20.0872}{24}, \frac{41.9128}{24}) = (0.8370, 1.7464).$$

0.8 tillhör inte intervallet, den genomsnittliga antalet tvillingfödslar per år är inte lika 0.8 med 95 % säkerhet.