TENTAMEN: Matematisk statistik för K (21 augusti, 2008)

Kortfattade lösningar:

1) a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B|A)$ och vi får att $P(A) = (P(A \cup B) - P(B))/(1 - P(B|A)) = \frac{2}{7}$.

   b) $P(A \cap B) = P(A)P(B|A) = \frac{2}{7} \cdot 0.3 = \frac{3}{35}$.

   c) $P(A \cap B) = \frac{3}{35} \neq 0$, dvs. att A och B inte är disjunkta händelser.

   d) $P(A \cap B) = \frac{3}{35} \neq 0.174 = \frac{2}{7} \cdot 0.6 = P(A)P(B)$ och därför är A och B inte oberoende.

2) a) Man kan säga någonting om noggranheten av skattningen.

   b) Man kan inte säga att det finns en skillnad mellan pojkar och flickor därför att konfidensintervallet innehåller värdet 0.

3) I samtliga fall (faktorer) har man med 95% sannolikhet hittat ett intervall, som täcker det "normala" värdet. Dvs. att sannolikheten att intervallet inte täcker det "normala" värdet är 0.05. Låt X vara antalet intervall som inte täcker det "normala" värdet. Då är $X \sim Bin(50, 0.05)$ och
$P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 1) - P(X = 0) = 0.72$.

4) a) Testa hypotesen $H_0 : \mu = 100$ mot $H_1 : \mu > 100$. Då är $T = (\bar{X} - 100)/(s/\sqrt{n}) \sim T_{71}$ och får ett värde 1.697. Nu är $t = 1.697 > 1.294 = t_{0.10}$ och man förkastar H_0 på signifikansnivån 0.10. Cyanidnivån verkar vara högre än 100mg/kg.

   b) $\alpha = 0.05$: $t_{71,0.05} = 1.667$ och H_0 förkastas
   $\alpha = 0.01$: $t_{71,0.005} = 2.380$ och H_0 accepteras
   Desto större är α desto större är risken att man förkastar H_0 även om den är sann. Desto mindre risk man tar desto svårare det är att förkasta H_0.

   c) Behövs inga antaganden (stickprovet tillräckligt stort).

5) Testa hypotesen $H_0 : \mu_X = \mu_Y$ mot $H_1 : \mu_X \neq \mu_Y$. Teststatistikan är $T = (\bar{X} - \bar{Y})/S\sqrt{1/n_1 + 1/n_2}$ som är T_{1037}-fördelad och $S^2 = ((n_1-1)S_X^2 + (n_2-1)S_Y^2)/(n_1+n_2-2)$. Impulsiv sensation sökande: $\bar{x}-\bar{y} = -0.1$, $s = 4.4$, $\sqrt{1/n_1 + 1/n_2} = 0.0626$, och $t = -0.363$. $\alpha = 0.01$ och $t_{0.005} = 2.576$ (se ∞-raden i T-fördelningstabellen). Nu är $-2.576 < -0.363 < 2.576$ och därför kan man inte förkasta H_0 på signifikansnivån 0.01. Det finns ingen skillnad mellan kokainmissbrukare och studenter utan missbruk när det gäller impulsivitet. Man har antagit att varianserna är lika.

   Aggression-fientlighet: $\bar{x} - \bar{y} = 1.3$, $s = 4.01$ och $t = 5.179$. Nu är $5.179 > 2.576$ och man förkastar H_0 på signifikansnivån 0.01. Kokainmissbrukare verkar vara mer aggressiva än studenter utan missbruk. (Varianserna kan antas vara lika.)

6) Låt X vara densiteten. Då är $X \sim N(0.0046, 9.6 \cdot 10^{-8})$.

   a) $P(0.004 < X < 0.005) = P(\frac{0.004-0.0046}{0.0003} < \frac{X-0.00046}{0.0003} < \frac{0.005-0.0046}{0.0003}) = P(-2 < Z < 1.33) = \Phi(1.33) - \Phi(-2) = 0.8854$, där $Z \sim N(0,1)$ och Φ är fördelningsfunktionen av den standardiserade normalfördelningen.

   b) $P(X > x_0) = 0.05$ och man borde hitta x_0.
   $P(X \leq x_0) = P(\frac{X-0.0046}{0.0003} \leq \frac{x_0-0.0046}{0.0003}) = \Phi(\frac{x_0-0.0046}{0.0003}) = 0.95 = \Phi(1.645)$ och $x_0 = 0.0051$.

7) a) När det finns ett linjärt samband mellan två variabler och man vill förklara en av dem m.h.a. den andra. Man kan använda residualplottar för att se om modellen är lämplig.

   b) Man vet att
$$\frac{B_0 - \beta_0}{S\sqrt{\sum x_i^2/n\sum(x_i - \bar{x})^2}} \sim T_{n-2}$$
   om $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma)$ (se formelbladet för beteckningarna). Ett 95% konfidensintervall för β_0: Man vet att
   $P(-t_{0.025} \leq \frac{B_0-\beta_0}{S\sqrt{\sum x_i^2/n\sum(x_i-\bar{x})^2}} \leq t_{0.025}) = 0.95$. Då får man att
   $P\left(-t_{0.025}S\sqrt{\sum x_i^2/n\sum(x_i - \bar{x})^2} \leq B_0 - \beta_0 \leq t_{0.025}S\sqrt{\sum x_i^2/n\sum(x_i - \bar{x})^2}\right)$
   $= P\left(B_0 - t_{0.025}S\sqrt{\sum x_i^2/n\sum(x_i - \bar{x})^2} \leq \beta_0 \leq B_0 + t_{0.025}S\sqrt{\sum x_i^2/n\sum(x_i - \bar{x})^2}\right)$
   $= 0.95$ och konfedensintervallet blir $B_0 \pm t_{0.025}^{(n-2)}S\sqrt{\sum x_i^2/n\sum(x_i - \bar{x})^2}$, där B_0 (skattningen för β_0) och S hittar man i formelbladet.