TENTAMEN: Matematisk statistik för K (8 mars, 2004)

Kortfattade lösningar:

1) a) $P(A \cap B) = P(A)P(B) = 0.08$
$P(A|B) = P(A) = 0.4$
$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.52$

b) $P(A|B) = \frac{P(A \cap B)}{P(B)}$ och man får att $P(A \cap B) = P(A|B)P(B) = 0.12$ och då är $P(A \cup B) = 0.48$
$P(B|A) = \frac{P(A \cap B)}{P(A)} = 0.30$

2) a) Väntevärdesriktighet och att variansen minskar då stickprovsstorleken ökar.

b) Ja. $\mathbf{E}[\bar{X}] = \mu$ så att det är en väntevärdesriktig skattning för μ och $Var(\bar{X}) = \sigma^2/n$, vilket blir mindre då n blir större. (Här $\mathbf{E}[X] = \mu$ och $Var(X) = \sigma^2$.)

3) a) Testa hypotesen $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 \neq \mu_2$ på signifikansnivån 0.1. Man får att $S_p^2 = 153.64$ och $T = (\bar{X}_1 - \bar{X}_2)/S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sim T_{64}$ och får ett värde 0.164. Nu är $t = 0.164 < 1.669 = t_{.05}$ (och $0.164 > -1.669$) så att H_0 kan inte förkastas på signifikansnivån 0.1. Dvs. att poängtalen för pojkar och flickor är inte olika. (Du kan också räkna p-värdet om du vill.)

b) $\bar{x}_1 - \bar{x}_2 \pm t_{0.05} s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.5 \pm 5.1$. Därför att $0 \in [-4.6, 5.6]$ kan man säga att det finns ingen skillnad mellan poängtal för pojkar och flickor enligt 90% konfidensintervallet. Man kan dra samma slutsats än i a).

c) Man måste anta att de två populationerna har samma varians. Man behöver inte anta normalfördelning därför att stickprovsstorlekerna är tillräckligt stora.

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.

5) a) Man får att $\hat{\mu}_{Y|x_i} = 6.25 - 0.00231x_i$

b) Figuren på vänster är en "scatterplot" där man har plottat (x_i, y_i)-värdena. Figuren på höger är en residualplot, dvs. $(x_i, y_i - (6.25 - 0.00231x_i))$. Det verkar som värdena av söthetsindexet minskar då pektinmängden ökar (se figuren till vänster) men punkterna är väldigt splittrade. Från residualplotten ser man att väntevärdet av $E_i \approx 0$ men variansen verkar inte vara konstant. Man har också ett hål i x-värdena. En linjär regressionsmodell verkar inte vara så bra.

6) a) $P(X > 55) = P((X - 50)/3.2 > (55 - 50)/3.2) = P(Z > 1.56) = 1 - P(Z \leq 1.56) = 1 - 0.9406 = 0.0594$, där $Z \sim N(0, 1)$.

b) Låt Y vara antalet vattenprov (bland 5) som har högre alkalinivån än 55mg/l. Då är $Y \sim Bin(5, 0.0594)$ och $P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - 0.0594)^5 = 0.2638$.

7) Man bör beräkna sannolikheten att man inte förkastar H_0 då $\mu = 9$, dvs. att man beräknar

$$
\begin{aligned}
&P(-z_{0.025} < (\bar{x} - 10)/(\sigma/\sqrt{n}) < z_{0.025}|\mu = 9) \\
&= P(-z_{0.025} + \sqrt{n}/\sigma < (\bar{x} - 9)/(\sigma/\sqrt{n}) < z_{0.025} + \sqrt{n}/\sigma) \\
&= P(-1.96 + \sqrt{20}/2 < Z < 1.96 + \sqrt{20}/2) \\
&= P(0.27 < Z < 4.20) = \Phi(4.20) - \Phi(0.27) = 1 - 0.6064 = 0.3956.
\end{aligned}
$$

($Z \sim N(0, 1)$ och $\Phi(4.20)$ måste vara större än 0.9998 så att vi kan säga att $\Phi(4.20) = 1$.)