**TENTAMEN:** Matematisk statistik för K (19 augusti, 2005)

**Kortfattade lösningar:**

1) a) $P(B|A) = P(A \cap B)/P(A) = 0.12/0.4 = 0.3$.

   b) Händelsen $B$ givet att $A$ har inträffat.

   c) Nej, $P(A \cap B) \neq 0$.

   d) $P(B) = 0.3$ och då är $P(A) \cdot P(B) = 0.12 = P(A \cap B)$, dvs. att $A$ och $B$ är oberoende.

2) Typ I fel: man förkastar $H_0$ även om den är sann. Typ II fel: man accepterar $H_0$ även om den inte är sann. Man kan garantera att både fel blir små genom att ha ett tillräckligt stort stickprov.

3) a) $X$ är kopparnivån. 90% konfidensintervall är $\bar{x} \pm t_{\alpha/2} \cdot s/\sqrt{n} = 0.408 \pm 0.145$, där $\bar{x} = 0.4083$, $s = 0.2495$ och $t^{(9)}_{0.05} = 1.833$.

   b) Medelkopparnivån är mellan 0.2637mg/l och 0.5529mg/l (med 90% sannolikhet).

   c) 90% av stickproven av storlek 10 skulle ge ett konfidensintervall som täcker det sanna medelkopparnivån (man är 90% konfident att medelnivån av koppar ligger på det beräknade intervallet). 10 gång i 100 skulle man få ett intervall som inte innehåller det sanna värdet.

   d) $X$ är normalfördelad.

4) 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 $\infty$-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 vilket verkar rimligt.

   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.)

5) Parvist test: $D = A - B$, $D$ antas vara normalfördelad. $H_0 : \mu_D = 0$ ($\mu_A = \mu_B$) och $H_1 : \mu_D \neq 0$. Observerade värdena av $D$ är 17, 9, -7, 14, -1, 12, 6 och 7 och $\bar{d} = 7.125$ och $s_d = 7.918$. Teststatistikan $T = \bar{D}/(S_d/\sqrt{n}) \sim T_{n-1} = T_7$. Nu är $t = 2.545$ och $t_{0.025}^{(7)} = 2.365$ och man kan förkasta $H_0$ på signifikansnivån 0.05 (därför att $2.545 > 2.365$). Det verkar som färg $B$ skulle vara mer resistant än färg $A$.

6) a) $P(X < 40) = P((X - 74)/16 < (40 - 74)/16) = P(Z < -2.125) = \Phi(-2.13) = 0.0166$, där $Z \sim N(0,1)$ och $\Phi$ är fördelningsfunktionen av $N(0,1)$- fördelningen.

   b) $Y$ är antalet feldiagnostiserade personer, $Y \sim Bin(4, 0.0166)$. Då är $P(Y \geq 1) = 1 - P(Y = 0) = 1 - 0.9834^4 = 0.065$.

7) a) Förening 1: inte så klart linjärt beroende. Förening 2: positivt linjärt beroende.

   b) $\hat{\beta}_1 = (n \sum x_i y_i - \sum x_i \sum y_i)/(n \sum x_i^2 - (\sum x_i)^2) = 41.713$ och $\hat{\beta}_0 = \hat{y} - \hat{\beta}_1 \bar{x} = 308.14$.

   c) Testa hypotesen $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$. Teststatistikan $\hat{\beta}_1/(S/\sqrt{\sum(x_i - \bar{x})^2})$ är $T_5$-fördelad och får värdet 3.019 ($s^2 = (\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2)/(n-2) = 142.64$). $t_{0.05} = 2.015$ och därför att $3.019 > 2.015$ förkastar man $H_0$. Dvs. att $\beta_1 \neq 0$ och att den maximala absorbanshastigheten verkar bero linjärt på Hammettkonstanten. Man har antagit att $Y$ är normalfördelad.