Lösningar till tentamen i Matematisk statistik och diskret matematik (MVE055/MSG810/MSG820).
Den 7 april 2010.

1. Lösning:

   a) $0.15 = 0.5 * 0.3 = P(A)P(B) \neq P(A \cap B) = 0.2$. Alltså är A och B inte oberoende.
   b) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.3} = 0.6667$
   c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.3 - 0.2 = 0.6$

2. Lösning:

   a) $X=$ vikt för slumpvis vald individ ur populationen
   $P(X > 82) = P(\frac{X-75}{5.5} > \frac{82-75}{5.5}) = P(Z > 1.27) = 1 - P(Z \leq 1.27) = 0.1016 \approx 0.1$
   (där Z betecknar en standardnormalfördelad stokastisk variabel.)
   b) $Y=$ antal individer av tio som väger mer än 82kg.
   $Y \sim Bin(10, 0.1)$
   $P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - 0.7361 = 0.2639$
   c) $W=$ summan av de två individernas vikter. W är normalfördelad med väntevärde 75+75=150 och standardavvikelse $\sqrt{5.5^2 + 5.5^5} = 7.78$
   $P(W > 164) = P(\frac{W-150}{7.78} > \frac{164-150}{7.78}) = P(Z > 1.80) = 1 - P(Z \leq 1.80) = 0.0359$

3. Lösning:
   S = 'skyldig', D = 'dömd'

   Vi har $P(D) = P(D|S)P(S) + P(D|S^c)P(S^c) = 0.75 * 0.8 + 0.004 * 0.2 = 0.6008$.

   Bayes sats ger $P(S|D) = \frac{P(D|S)P(S)}{P(D)} = \frac{0.75*0.8}{0.6008} = 0.998668$.

   Alltså har vi $P(S^c|D) = 1 - 0.998668 = 0.001332$.

4. Lösning:

   a) Låt $X=$ längden på slumvis vald individ ur populationen och μ_X beteckna väntevärdet för X, dvs den sökta medellängden. Ett 95%-igt konfidensintervall för medellängden ges av $\bar{X} \pm t_{0.975} S/\sqrt{n}$, där t-fördelningen har $n-1$ frihetsgrader.
   Vi har $n = 5$, $\bar{x} = 173$, $s = 4.30$ och $t_{0.975} = 2.776$ (för 4 frihetsgrader).
   Alltså ges ett 95%-igt konfidensintervall för medellängden av $173 \pm 2.776 * 4.30/\sqrt{5}$
   dvs konfidensintervallet är $[167.66, 178.34]$
   b) X betecknar längden på slumvis vald individ ur populationen, μ_X betecknar väntevärdet för X och σ_X betecknar variansen.
   Eftersom σ_X är okänd måste denna skattas. Om X är normalfördelad gäller det att $\frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}}$ är standard normalfördelad. Men om σ_X är okänd och måste skattas med stickprovsvariansen S, då gäller att $\frac{\bar{X}-\mu_X}{s/\sqrt{n}}$ är t-fördelad.
   En t-fördelad stokastisk variabel har större varians än en standard normalfördelad stokastisk variabel (kommer från extra osäkerhet i skattningen av variansen) och därmed blir ett konfidensintervall baserat på en t-fördelning vidare. Om man hade baserat konfidensintervallet på en normalfördelning hade man alltså fått ett

snävare intervall (= bra). Dock hade detta endast varit ett approximativt 95%-igt konfidensintervall (= inte bra) för att man hade approximerat en t-fördelning med en standard normalfördelning.

5. Lösning:

   a) Tidsenhet: timmar. Intensitet: 30 telefonsamtal/timme.
   $X \sim Po(\text{antal tidsenheter} * \text{intensitet}) = Po(\frac{1}{6}30) = Po(5)$.
   $P(X \leq 1) = P(X = 0) + P(X = 1) = 0.0067 + 0.0337 = 0.0404$

   b) Låt T vara tiden från en tidpunkt t_0 till nästa inkommande samtal. Enligt sats är T exponentialfördelad med väntevärde 1/30h 2 minuter. $P(T < 5) = 0.9179$.

6. Lösning:
   Antalet sätt ges av koefficienten framför x^{45} i följande uttryck:

$$
\begin{aligned}
(1 + x + x^2 + x^3 + \ldots + x^{18})^4 &= (\tfrac{1-x^{19}}{1-x})^4 \\
&= \left(1 - x^{19}\right)^4 \left(\tfrac{1}{1-x}\right)^4 \\
&= \left(1 + 4(-x^{19}) + 6(-x^{19})^2 + 4(-x^{19})^3 + (-x^{19})^4\right) \textstyle\sum_{n=0}^{\infty} \begin{pmatrix} n+3 \\ 3 \end{pmatrix} \\
&= \left(1 - 4x^{19} + 6x^{38} - 4x^{57} + x^{76}\right) \textstyle\sum_{n=0}^{\infty} \begin{pmatrix} n+3 \\ 3 \end{pmatrix} x^n
\end{aligned}
$$

   Koefficienten framför x^{45} är alltså

$$
\begin{pmatrix} 45+3 \\ 3 \end{pmatrix} - 4 \begin{pmatrix} 26+3 \\ 3 \end{pmatrix} + 6 \begin{pmatrix} 7+3 \\ 3 \end{pmatrix} = 3400
$$

7. Lösning:

   a) Om vi använder en normalapproximation har vi att under H_0 är
   $Y = \frac{\hat{p}-0.5}{\sqrt{\hat{p}(\hat{p}-1)/40}} \overset{\text{appr}}{\sim} N(0,1)$. Med $\hat{p} = 0.45$ får vi $Y = -0.6356$. Låt Z beteckna en standard normalfördelad s.v. Vi har att $P(Z \leq -0.6356) = 0.2625$. Detta är inte mindre än signifikansnivån 0.05 så vi kan inte förkasta H_0.
   Om vi räknar exakt: W = antalet invånare som är positiva till projektet. Under H_0 gäller $W \sim Bin(40, 0.5)$. $P(W \leq 18) = 0.3179$. Detta är inte mindre än signifikansnivån 0.05 så vi kan inte förkasta H_0.

   b) Om 14 är den kritiska punkten för testet och det sanna värdet för andelen är 0.4 ges styrkan av $P(W \leq 14)$, där $W \sim Bin(40, 0.4)$, vilket är 0.3174.

8. Lösning:
   En enkel linjär regressionsmodell används för att beskriva ett linjärt samband mellan två variabler. Om X och Y är två variabler vars linjära samband man vill undersöka kan modellen skrivas $Y_i = \beta_0 + \beta_1 x_i + E_i$ där β_1 ger den förväntade förändringen för Y om X ökas med en enhet, β_0 ger det förväntade värdet för Y då $X = 0$ och E_i är avvikelsen för den i:te Y-observationen från dess förväntade värde.