1. (a) Enligt informationen i uppgiften har vi $\mathsf{P}(A) = 0.35$, $\mathsf{P}(B) = 0.05$ och $\mathsf{P}(A|B) = 0.1$. Detta ger att $\mathsf{P}(A \cap B) = \mathsf{P}(A|B)\mathsf{P}(B) = 0.1 \cdot 0.05 = 0.005$.

   (b) Vi söker $\mathsf{P}(A^c \cap B^c) = 1 - \mathsf{P}(A \cup B) = 1 - (\mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)) = 1 - (0.35 + 0.05 - 0.005) = 1 - 0.3950 = 0.6050$.

   (c) Enligt Bayes sats: $\mathsf{P}(B|A) = \mathsf{P}(A|B)\mathsf{P}(B)/\mathsf{P}(A) = 0.005/0.35 = 0.0143$.

2. (a) μ anger det förväntade antalet hårddiskar som kommer gå sönder under ett år.

   (b) Låt X beteckna antalet hårddiskar som går sönder under ett år. Vi söker $\mathsf{P}(X \le 3) = \mathsf{P}(X = 0) + \mathsf{P}(X = 1) + \mathsf{P}(X = 2) + \mathsf{P}(X = 3) = e^{-\mu}(1 + \mu + \mu^2/2 + \mu^3/6)$. Med $\mu = 2.5$ får vi $\mathsf{P}(X \le 3) = 0.7576$.

   (c) Efter nio månader kvarstår 3 månader av året, vilket är 0.25 år. Låt Y beteckna antalet hårddiskar som går sönder under 0.25 år, vi har då $Y \sim$Po(0.25μ). Vi söker $\mathsf{P}(Y \le 1) = \mathsf{P}(Y = 0) + \mathsf{P}(Y = 1) = e^{-0.25\mu}(1 + 0.25\mu)$. Med $\mu = 2.5$ får vi $\mathsf{P}(Y \le 1) = 0.8698$.

3. (a) Låt $X \sim \mathsf{N}(75, 10^2)$ beteckna vikten. Vikterna vi söker är talen a och b så att $\mathsf{P}(a \le X \le b) = 0.95$. Eftersom talen ska vara symmetriska kring 75kg gäller att b ska uppfylla

$$0.025 = \mathsf{P}(X > b) = \mathsf{P}((X - 75)/10 > (b - 75)/10) = \mathsf{P}(Z > (b - 75)/10)$$

   där $Z \sim \mathsf{N}(0, 1)$. Detta ger oss att $(b-75)/10 = \lambda_{0.025}$, där $\lambda_{0.025} = 1.96$ är 0.025-kvantilen i den standardiserade normalfördelningen. Detta ger $b = 10 \cdot 1.96 + 75 = 94.6$. På grund av symmetri får vi $a = -10 \cdot 1.96 + 75 = 55.4$.

   (b) Låt $Y = \sum_{i=1}^{13} X_i$, där $X_i \sim \mathsf{N}(75, 10^2)$ är oberoende. Vi söker $\mathsf{P}(Y > 1000)$. Vi har $\mathsf{E}(Y) = 13 \cdot 75$ och $\mathsf{V}(Y) = 13 \cdot 10^2$ och alltså är $Y \sim \mathsf{N}(975, 13 \cdot 10^2)$, vilket ger

$$\mathsf{P}(Y > 1000) = \mathsf{P}\left(\frac{Y - 975}{\sqrt{13} \cdot 10} > \frac{1000 - 975}{\sqrt{13} \cdot 10}\right) = \mathsf{P}\left(Z > \frac{25}{\sqrt{13} \cdot 10}\right) = \mathsf{P}(Z > 0.6934).$$

   Enligt formelsamlingen är $\mathsf{P}(Z < 0.69) = 0.7549$, och alltså är den sökta sannolikheten $\mathsf{P}(Z > 0.6934) \approx 1 - 0.7549 = 0.2451$. Det exakta svaret är 0.2440 (uträknat med `1-normcdf(25/(sqrt(13)*10))` i Matlab).

4. (a) Låt X vara antalet personer som får biverkningar. Vi har att $X \sim$ Bin$(200, p)$ och vill testa $H_0 : p = 0.1$ mot $H_1 : p < 0.1$. Med direktmetoden får vi att $\mathsf{P}(X \le 14|H_0) = \mathsf{P}(X \le 14|X \sim$Bin$(200, 0.1)) = 0.0929$. Eftersom denna sannolikhet är större än 0.01 kan vi inte förkasta H_0. Denna metod är krävande utan tillgång till Matlab eller liknande verktyg som har en implementering av fördelningsfunktionen för binomialfördelningen, vi kan därför alternativt använda normalapproximation för att göra testet. Detta är tillåtet eftersom $np(1-p) = 18$ under H_0. Vi har därför att X är approximativt $\mathsf{N}(np, np(1-p))$-fördelad. Alltså är

$$\begin{aligned}\mathsf{P}(X \le 14|H_0) &\approx \mathsf{P}(X \le 14|X \sim \mathsf{N}(20, 18)) = \Phi\left(\frac{14 - 20}{\sqrt{18}}\right) \\ &= \Phi(-1.4142) = 1 - \Phi(1.4142) \approx 1 - \Phi(1.41) = 1 - 0.9207 = 0.0793.\end{aligned}$$

   Eftersom sannolikheten är större än $\alpha = 0.1$ kan vi inte heller med denna metod förkasta H_0.

   (b) Ett hypotestests styrka anger sannolikheten att förkasta H_0 givet att H_1 är sann.

5. (a) Man ska kunna mäta den förklarande variabeln utan fel och responsvariabeln ska bero på den förklarande variabeln enligt den linjära modellen. Mätfelen e_i ska vara oberoende och normalfördelade med väntevärde noll. Enligt residualplottarna i figur 1 verkar dessa antaganden vara rimliga.

(b) Vi har $\beta_1^* = S_{xy}/S_{xx} = 0.257$ och $\beta_0^* = \bar{y} - \bar{x}\beta_1^* = 4.678$.

(c) Vi vill skatta $Y(x_0)$ för $x_0 = \log(8000) = 8.987$. Den förväntade logaritmerade lönen är $\mu_Y(x_0) = \beta_0^* + \beta_1^*\bar{x} = 6.99$. Ett prediktionsintervall ges av

$$I_{\mu_Y(x_0)} = \left(\mu_Y(x_0) \pm t_{\alpha/2}(175)s\sqrt{1 + \frac{1}{177} + \frac{(x_0-\bar{x})^2}{S_{xx}}}\right)$$

Vi beräknar först $s = \sqrt{(S_{yy} - S_{xy}^2/S_{xx})/175} = \sqrt{49.65906/175} = 0.532697$ och med $t_{0.025}(175) \approx 1.96$ får vi

$$I_{\mu_Y(x_0)} = \left(6.99 \pm 1.96 \cdot 0.532697\sqrt{1.0168}\right) = (6.99 \pm 1.05282) = (5.93718,\, 8.04282)$$

(d) Ett konfidensintervall anger var den förväntade lönen troligen ligger, eller med andra ord var regressionlinjen är. Ett prediktionsintervall anger var en ny observation för ett visst värde på x troligen kommer vara.

(e) För att få det förväntade marknadsvärdet löser vi ut x ur ekvationen $5 = \beta_0 + \beta_1 x = 4.678 + 0.257x$ och får $x = (5 - 4.678)/0.257 = 1.2529$. Detta värde är långt under de minsta observerade marknadsvärdena, så vi kan inte vara säkra på att den linjära modellen stämmer också i detta intervall.

6. (a) Vi har $\mathsf{E}(\bar{x}) = \mathsf{E}(\bar{y}) = p$, $\mathsf{V}(\bar{x}) = \sigma_1^2/m$ och $\mathsf{V}(\bar{y}) = \sigma^2/n$. Detta ger

$$\mathsf{E}(p^*) = \mathsf{E}(\alpha\bar{x} + \beta\bar{y}) = (\alpha+\beta)p$$
$$\mathsf{V}(p^*) = \mathsf{V}(\alpha\bar{x} + \beta\bar{y}) = \alpha^2\sigma_1^2/m + \beta^2\sigma_2^2/n.$$

(b) p^* är väntevärdesriktig om $\mathsf{E}(p^*) = p$, alltså måste vi ha $\alpha + \beta = 1$.

(c) Vi sätter $\beta = 1 - \alpha$, vilket garanterar att skattaren är väntevärdesriktig samt att den har varians $\alpha^2\sigma_1^2/m + (1-\alpha)^2\sigma_2^2/n$. Vi vill nu hitta värdet på α som minimerar variansen, vi deriverar därför variansen med avseende på α och sätter derivatan till noll:

$$\frac{\partial}{\partial\,\alpha}\mathsf{V}(p^*) = 0 \quad \Rightarrow \quad 2\alpha\sigma_1^2/n - 2(1-\alpha)\sigma_2^2/m = 0 \quad \Rightarrow$$
$$\alpha(\sigma_1^2/n + \sigma_2^2/m) = \sigma_2^2/m \quad \Rightarrow \quad \alpha = \frac{\sigma_2^2/m}{\sigma_1^2/n + \sigma_2^2/m}$$

(d) Med $\sigma_1 = \sigma_2 = 10$, $m = 5$ och $n = 10$ får vi att

$$\alpha = \frac{\sigma_2^2/m}{\sigma_1^2/n + \sigma_2^2/m} = \frac{1}{(\sigma_1/\sigma_2)^2(m/n) + 1} = \frac{1}{5/10 + 1} = \frac{2}{3}$$

(e) Vi har som tidigare $\mathsf{V}(\bar{x}) = \sigma_1^2/n$, men får nu $\mathsf{V}(\tilde{y}) = \sigma_2^2/n$ och

$$\mathsf{C}(\bar{x}, \tilde{y}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{C}(x_i, \tilde{y}_i) = \frac{\rho}{n}.$$

Detta ger oss att

$$\mathsf{V}(\tilde{p}^*) = \gamma^2\mathsf{V}(\bar{x}) + (1-\gamma)^2\mathsf{V}(\tilde{y}) + 2\gamma(1-\gamma)\mathsf{C}(\bar{x},\tilde{y}) = \gamma^2\frac{\sigma_1^2}{n} + (1-\gamma)^2\frac{\sigma_2^2}{n} + 2\gamma(1-\gamma)\frac{\rho}{n}$$

Vi deriverar med avseende på γ och sätter derivatan till noll:

$$\frac{\partial}{\partial \gamma} \mathsf{V}(\tilde{p}^*) = 0 \quad \Rightarrow \quad \gamma \frac{\sigma_1^2}{n} - (1-\gamma)\frac{\sigma_2^2}{n} + (1-2\gamma)\frac{\rho}{n} = 0 \quad \Rightarrow$$

$$\gamma(\sigma_1^2 + \sigma_2^2 - 2\rho) = \sigma_2^2 - \rho \quad \Rightarrow \quad \gamma = \frac{\sigma_2^2 - \rho}{\sigma_1^2 + \sigma_2^2 - 2\rho}$$