---

1. (a) Låt A beteckna händelsen att taxibilen är blå och låt B beteckna händelsen att vittnet pekar ut bilen som blå. Vi söker $\mathsf{P}(A|B)$, som via Bayes sats fås som

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B)} = \frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B|A)\mathsf{P}(A) + \mathsf{P}(B|A^c)\mathsf{P}(A^c)} = \frac{0.8 \cdot 0.25}{0.8 \cdot 0.25 + 0.2 \cdot 0.75} \approx 0.57$$

   (b) Vi söker det minsta talet $x = \mathsf{P}(B|A)$ som gör att $\mathsf{P}(A|B) \geq 0.9$, vilket vi får genom att lösa

$$\frac{9}{10} = \frac{x \cdot \frac{1}{4}}{x \cdot \frac{1}{4} + (1-x) \cdot \frac{3}{4}} \quad \Rightarrow \quad 27 - 18x = 10x \quad \Rightarrow \quad x = \frac{27}{28} \approx 0.9643$$

2. (a) Låt $Y = X_1 + X_2$. Eftersom summan av poissonfördelade variabler också är possionfördelad har vi att $Y \sim \text{Po}(1+2) = \text{Po}(3)$.

$$\begin{aligned}\mathsf{P}(Y > 2) &= 1 - \mathsf{P}(Y \leq 2) = 1 - (\mathsf{P}(Y=0) + \mathsf{P}(Y=1) + \mathsf{P}(Y=2)) \\ &= 1 - e^{-3}(1 + \frac{3}{1} + \frac{3^2}{2}) \approx 0.5768\end{aligned}$$

   (b) Låt $Z_n = \sum_{i=1}^{n} X_i$. Vi har $Z_n \sim \text{Po}(n)$ och får

$$\begin{aligned}\mathsf{P}(Z_n > 2) &= 1 - \mathsf{P}(Z_n \leq 2) = 1 - (\mathsf{P}(Z_n=0) + \mathsf{P}(Z_n=1) + \mathsf{P}(Z_n=2)) \\ &= 1 - e^{-n}(1 + n + \frac{n^2}{2})\end{aligned}$$

   Vi sätter in olika värden på n och får $\mathsf{P}(Z_1 > 2) = 0.0803, \mathsf{P}(Z_2 > 2) = 0.3233, \mathsf{P}(Z_3 > 2) = 0.5768, \mathsf{P}(Z_4 > 2) = 0.7619, \mathsf{P}(Z_5 > 2) = 0.8753, \mathsf{P}(Z_6 > 2) = 0.9380$. Alltså är $n = 6$ det lägsta antalet vi måste summera.

3. Låt X vara antal personer i en bil, och låt $p_k = p(k) = \mathsf{P}(X = k)$. Vi har då

$$\begin{aligned}\mathsf{E}(X) &= 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 + 5 \cdot p_5 = 1.72 \\ \mathsf{V}(X) &= (1-1.72)^2 \cdot p_1 + (2-1.72)^2 \cdot p_2 + (3-1.72)^2 \cdot p_3 + (4-1.72)^2 \cdot p_4 + (5-1.72)^2 \cdot p_5 \\ &= 0.8016\end{aligned}$$

   Om X_i är antal personer i bil i söker vi $\mathsf{P}(\sum_{i=1}^{570} X_i \geq 1000)$. Enligt centrala gränsvärdessatsen gäller att fördelningen för $\sum_{i=1}^{550} X_i$ är approximativt $\mathsf{N}(570 \cdot \mathsf{E}(X), 570 \cdot \mathsf{V}(X)) = \mathsf{N}(980.4, 456.91)$. Alltså har vi

$$\begin{aligned}\mathsf{P}\left(\sum_{i=1}^{570} X_i \geq 1000\right) &= 1 - \mathsf{P}\left(\sum_{i=1}^{570} X_i < 999\right) = 1 - \Phi\left(\frac{999 - 980.4}{\sqrt{456.91}}\right) \\ &= 1 - \Phi(0.8702) = 1 - 0.8079 = 0.1921,\end{aligned}$$

   där $\Phi(0.8702)$ hämtas från tabell eller miniräknare. En något bättre approximation av sannolikheten fås om vi använder "half unit correction":

$$\begin{aligned}\mathsf{P}\left(\sum_{i=1}^{570} X_i \geq 1000\right) &= 1 - \mathsf{P}\left(\sum_{i=1}^{570} X_i < 999.5\right) = 1 - \Phi\left(\frac{999.5 - 980.4}{\sqrt{456.91}}\right) \\ &= 1 - \Phi(0.8935) = 1 - 0.8142 = 0.1858.\end{aligned}$$

4. (a) Vi vill testa $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ och gör det på nivå 95%. Vi beräknar

$$S_{xx} = \sum_i x_i^2 - \frac{1}{8}(\sum_i x_i)^2 = 424$$
$$S_{yy} = \sum_i y_i^2 - \frac{1}{8}(\sum_i y_i)^2 = 0.0096$$
$$S_{xy} = \sum_i x_i y_i - \frac{1}{8}(\sum_i x_i)(\sum_i y_i) = -2.012$$

Minsta-kvadratskattningen av β_1 ges av $\beta_1^* = S_{xy}/S_{xx} = -0.0047$ och ett 95%-konfidensintervall för β_1 ges av

$$I_{\beta_1} = [\beta_1^* \pm t_{0.025}(7)\frac{s}{\sqrt{S_{xx}}}]$$

Vi har $t_{0.025}(7) = 2.3646$, $s = \sqrt{\frac{1}{6}(S_{yy} - \frac{S_{xy}^2}{S_{xx}})} = 0.0026$ och får därför $I_{\beta_1} = [-0.0050, -0.0044]$. Intervallet täcker ej noll, och vi kan alltså förkasta H_0.

(b) Ett 95%-konfidensintervall för $\mu_Y(60)$ ska med 95% sannolikhet täcka det förväntade promillevärdet hos kvinnor som väger 60kg. Ett 95%-prediktionsintervall ska med 95% sannolikhet täcka det uppmätta promillevärdet hos en kvinna som väger 60kg. Prediktionsintervallet tar alltså hänsyn till den individuella variationen och är därför bredare.

(c) Minsta-kvadratskattningen av β_0 ges av $\beta_0^* = \bar{y} - \beta_1^*\bar{x} = 0.4435$. Prediktionsintervallet ges nu av

$$I_{Y(60)} = \left[\beta_0^* + 60\beta_1^* \pm t_{0.025}(6)s\sqrt{1 + \frac{1}{n} + \frac{(60 - \bar{x})^2}{S_{xx}}}\right] = [0.1520, 0.1655]$$

5. (a) En skattning av θ ges as $\theta^* = \bar{Y} - \bar{X}$. Eftersom vi känner varianserna för X och Y har vi $\mathsf{V}(\theta^*) = \frac{1}{n} + \frac{1}{n}$ samt $\theta^* \sim \mathsf{N}(\theta, \frac{2}{n})$. Som teststorhet används

$$T = \frac{\theta^*}{\sqrt{\mathsf{V}(\theta^*)}}$$

som är $\mathsf{N}(0,1)$-fördelad under H_0. Vi har $T_{obs} = (5.78 - 4.36)/\sqrt{2/5} = 2.2452$. Vi använder en signifikansnivå på 0.95 och ska då förkasta H_0 om $T_{obs} > \lambda_{0.05} = 1.6449$. Vi kan alltså förkasta H_0 på denna signifikansnivå.

(b) Ett fel av typ 1 görs då H_0 förkastas trots att H_0 är sann. Ett fel av typ 2 görs då H_0 ej förkastas trots att H_1 är sann. Testets styrka är sannolikheten för att inte göra ett fel av typ 2.

(c) Vi söker det värde på n så att $\mathsf{P}(T > \lambda_{0.05}) = 0.9$. Om $\theta = 1$ har vi enligt beräkningen ovan att $\theta^* \sim \mathsf{N}(1, \frac{2}{n})$ och alltså att $T \sim \mathsf{N}(\sqrt{\frac{n}{2}}, 1)$. Med $Z \sim \mathsf{N}(0,1)$ har vi

$$\mathsf{P}(T > \lambda_{0.05}) = \mathsf{P}\left(Z + \sqrt{\frac{n}{2}} > \lambda_{0.05}\right) = 1 - \mathsf{P}\left(Z \leq \lambda_{0.05} - \sqrt{\frac{n}{2}}\right) = 1 - \Phi\left(\lambda_{0.05} - \sqrt{\frac{n}{2}}\right).$$

Värdet på n får nu genom att lösa ut n ur ekvationen $1 - \Phi\left(\lambda_{0.05} - \sqrt{\frac{n}{2}}\right) = 0.9$:

$$1 - \Phi\left(\lambda_{0.05} - \sqrt{\frac{n}{2}}\right) = 0.9 \quad \Rightarrow \quad \Phi\left(\lambda_{0.05} - \sqrt{\frac{n}{2}}\right) = 0.1 \quad \Rightarrow$$
$$\lambda_{0.05} - \sqrt{\frac{n}{2}} = \Phi^{-1}(0.1) = -\lambda_{0.1} \quad \Rightarrow \quad \sqrt{\frac{n}{2}} = \lambda_{0.05} + \lambda_{0.1} \quad \Rightarrow$$
$$n = 2(\lambda_{0.05} + \lambda_{0.1})^2 = 17.1288$$

Vi ska alltså välja åtminstone $n = 18$.

6. (a) Vi har

$$\mathsf{E}(\theta_1^*) = \mathsf{E}\left(\frac{1}{2}\left(X + \frac{Y}{3}\right)\right) = \frac{1}{2}\left(\mathsf{E}(X) + \frac{\mathsf{E}(Y)}{3}\right) = \frac{1}{2}\left(\theta + \frac{3\theta}{3}\right) = \theta$$
$$\mathsf{E}(\theta_2^*) = \mathsf{E}\left(\frac{X+Y}{4}\right) = \frac{\mathsf{E}(X) + \mathsf{E}(Y)}{4} = \frac{\theta + 3\theta}{4} = \theta$$

(b) Vi har

$$\mathsf{V}(\theta_1^*) = \mathsf{V}\left(\frac{1}{2}\left(X + \frac{Y}{3}\right)\right) = \frac{1}{4}\left(\mathsf{V}(X) + \frac{\mathsf{V}(Y)}{9}\right) = \frac{1}{4}\left(\theta + \frac{3\theta}{9}\right) = \frac{\theta}{3}$$
$$\mathsf{V}(\theta_2^*) = \mathsf{V}\left(\frac{X+Y}{4}\right) = \frac{\mathsf{V}(X) + \mathsf{V}(Y)}{16} = \frac{\theta + 3\theta}{16} = \frac{\theta}{4}$$

Den andra metoden är alltså mer effektiv.

(c) Vi använder den andra metoden och får att $\theta_2^* = 27/4$. Eftersom X och Y båda är Poissonfördelade har vi att $Z = X + Y \sim \text{Po}(4\theta)$. Vi normalapproximerar och får $Z \sim \mathsf{N}(4\theta, 4\theta)$ och alltså att $\theta_2^* = Z/4 \sim \mathsf{N}(\theta, \theta/4)$. Ett konfidensintervall med approximativ konfidensgrad 95% ges därför av

$$I_\theta = \left[\theta_2^* \pm \lambda_{0.025}\sqrt{\frac{\theta_2^*}{4}}\right] = [4.204, 9.296]$$