1. (a) Vi har att $\mathsf{E}(X-Y) = \mathsf{E}(X) - \mathsf{E}(Y) \Rightarrow 1 = 3 - \mathsf{E}(Y) \Rightarrow \mathsf{E}(Y) = 2$, samt

$$\mathsf{V}(X-Y) = \mathsf{V}(X) + \mathsf{V}(Y) - 2\mathsf{C}(X,Y) \Rightarrow 4 = 1 + \mathsf{V}(Y) - 2\cdot 0.3 \Rightarrow \mathsf{V}(Y) = 3.6.$$

Eftersom X och $X-Y$ är normalfördelade så är även Y normalfördelad och vi har alltså $Y \sim \mathsf{N}(2, 3.6)$. Den sökta sannolikheten ges av

$$\begin{aligned}\mathsf{P}(0 \le Y \le 1) &= \mathsf{P}(Y \le 1) - \mathsf{P}(Y \le 0) = \Phi\left(\frac{1-2}{\sqrt{3.6}}\right) - \Phi\left(\frac{-2}{\sqrt{3.6}}\right)\\ &= \Phi(-0.5270) - \Phi(-1.0541) = 1 - \Phi(0.5270) - 1 + \Phi(1.0541))\\ &= \Phi(1.0541)) - \Phi(0.5270) = 0.8541 - 0.7009\\ &= 0.1532.\end{aligned}$$

(b) Vi har $\mathsf{P}(B) = \mathsf{P}(A \cap B)/\mathsf{P}(A|B) = 0.4/0.7 = 4/7$. Bayes formel ger

$$\mathsf{P}(B|A) = \frac{\mathsf{P}(A|B)\mathsf{P}(B)}{\mathsf{P}(A)} = \frac{0.7\cdot 4/7}{0.5} = 0.8.$$

2. (a) Svaret ges av sannolikhetsfunktionens värde i $k = 0$, alltså $5^0 e^{-5}/0! = e^{-5} \approx 0.0067$.

(b) Om intensiteten av antalet olyckor under ett år är $\mu = 5$ så är antalet olyckor under sex månader $Y \sim \mathrm{Po}(0.5\mu)$. Vi söker alltså

$$\mathsf{P}(Y \le 3) = e^{-5}\left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!}\right) = 0.2650.$$

(c) För att testa $H_0 : \mu = 5$ mot $H_1 : \mu > 5$ med direktmetoden beräknar vi

$$\begin{aligned}\mathsf{P}(Y \ge 7|\mu = 5) &= 1 - \mathsf{P}(Y \le 6|\mu = 5)\\ &= 1 - e^{-2.5}\left(\frac{2.5^0}{0!} + \frac{2.5^1}{1!} + \frac{2.5^2}{2!} + \frac{2.5^3}{3!} + \frac{2.5^4}{4!} + \frac{2.5^5}{5!} + \frac{2.5^6}{6!}\right)\\ &= 1 - 0.9858 = 0.0142.\end{aligned}$$

Eftersom denna sannolikhet är lägre än α så förkastas H_0.

3. (a) Det totala medelvärdet av observationerna är

$$\bar{y} = \frac{1}{20+30+15+25}(20\cdot 0.82 + 30\cdot 2.05 - 15\cdot 0.41 + 25\cdot 3.09) = 1.6556.$$

Vi kan nu beräkna

$$\begin{aligned}SS_A &= \sum_{ij}(\bar{y}_i - \bar{y})^2\\ &= 20(0.82 - 1.6556)^2 + 30(2.05 - 1.6556)^2 + 15(-0.41 - 1.6556)^2 + 25(3.09 - 1.6556)^2\\ &= 134.0692.\end{aligned}$$

Vidare har vi att

$$\begin{aligned}SS_E &= \sum_{j=1}^{4}\sum_{i=1}^{n_j}(y_{ij} - \bar{y}_i)^2 = \sum_{j=1}^{4}(n_j - 1)s_i^2\\ &= 19\cdot 1.03^2 + 29\cdot 1.02^2 + 14\cdot 0.85^2 + 24\cdot 0.92^2\\ &= 80.7573.\end{aligned}$$

Slutligen är $SS_{Tot} = SS_E + SS_A = 214.8265$ och vi kan nu fylla i tabellen.

Variation	Kvadratsumma	Frihetsgrader	Medelkvadratsumma	Teststorhet
Faktor A	134.0692	3	44.6897	47.5910
Residual	80.7573	86	0.9390	
Total	214.8265	89		

(b) Om vi låter μ_A, μ_B, μ_C och μ_D vara väntevärdena för de fyra instrumenten så testar vi $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ mot $H_1 : \mu_i \neq \mu_j$ för något $i \neq j$ och $i, j \in \{A, B, C, D\}$. Teststorheten är $F(3, 86)$-fördelad. Enligt formelsamlingen har vi

$$F_{0.05}(3, 86) \approx F_{0.05}(3, 80) = 2.71.$$

Det kritiska området för testet är $C_\alpha = \{T : T > 2.6507\}$. Eftersom $T_{obs} = 47.5910$ ligger i C_α kan vi förkasta nollhypotesen.

4. (a) Låt X vara en slumpvariabel med sannolikhetsfunktion $p(k)$. Vi har då

$$\mathsf{E}(X) = \sum_{k=3}^{6} kp(k) = 3 \cdot 0.2 + 4 \cdot 0.4 + 5 \cdot 0.3 + 6 \cdot 0.1 = 4.3.$$

$$\mathsf{E}(X^2) = \sum_{k=3}^{6} k^2 p(k) = 3^2 \cdot 0.2 + 4^2 \cdot 0.4 + 5^2 \cdot 0.3 + 6^2 \cdot 0.1 = 19.3.$$

Alltså är $\mathsf{V}(X) = \mathsf{E}(X^2) - \mathsf{E}(X)^2 = 19.3 - 4.3^2 = 0.81$.

(b) Enligt centrala gränsvärdessatsen har vi att antalet personer per månad, $\sum_{i=1}^{150} X_i$, är approximativt $\mathsf{N}(150 \cdot 4.3, 150 \cdot 0.81)$. Om vi låter p vara priset per person så är alltså intäckterna under en månad, $I = p\sum_{i=1}^{150} X_i$, fördelade enligt $\mathsf{N}(p150 \cdot 4.3, 121.5 \cdot p^2)$. Vi söker värdet på p så att

$$\mathsf{P}(I > 150000) = 0.95 \Rightarrow 1 - \Phi\left(\frac{150000 - 150 \cdot 4.3 \cdot p}{p\sqrt{121.5}}\right) = 0.95$$

Enligt formelsamlingen har vi $\lambda_{0.05} = 1.6449$ och vi får alltså

$$\frac{150000 - 150 \cdot 4.3 \cdot p}{p\sqrt{121.5}} = -1.6449 \Rightarrow p = \frac{150000}{150 \cdot 4.3 - 1.6449\sqrt{121.5}} \approx 239.$$

De måste alltså ta åtminstone 239kr per person.

5. (a) Eftersom fördelningen är symmetrisk kring $x = a$ kan vi direkt inse att $\mathsf{E}(X) = a$. Om vi istället vill visa detta genom att integrera så delar vi upp integralen i två delar

$$\mathsf{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{a} \frac{1}{2b} x \exp((x-a)/b)dx + \int_{a}^{\infty} \frac{1}{2b} x \exp(-(x-a)/b)dx$$

Vi använder partialintegration för att beräkna dessa två integraler:

$$\begin{aligned}\int_{-\infty}^{a} \frac{1}{2b} x \exp((x-a)/b)dx &= \left[\frac{x}{2}\exp((x-a)/b)\right]_{-\infty}^{a} - \int_{-\infty}^{a} \frac{1}{2}\exp((x-a)/b)dx \\ &= \frac{a}{2} - \left[\frac{b}{2}\exp((x-a)/b)\right]_{-\infty}^{a} = \frac{a}{2} - \frac{b}{2},\end{aligned}$$

$$\begin{aligned}\int_{a}^{\infty} \frac{1}{2b} x \exp(-(x-a)/b)dx &= \left[-\frac{x}{2}\exp(-(x-a)/b)\right]_{a}^{\infty} - \int_{a}^{\infty} -\frac{1}{2}\exp(-(x-a)/b)dx \\ &= \frac{a}{2} + \left[-\frac{b}{2}\exp(-(x-a)/b)\right]_{a}^{\infty} = \frac{a}{2} + \frac{b}{2}.\end{aligned}$$

Alltså är $\mathsf{E}(X) = a/2 - b/2 + a/2 + b/2 = a$. För att beräkna variansen använder vi att funktionen $(x-a)^2 f(x)$ precis som $f(x)$ är symmetrisk kring $x = a$, och får

$$\begin{aligned}\mathsf{V}(X) &= \int_{-\infty}^{\infty} (x-a)^2 f(x)dx = 2\int_a^{\infty} (x-a)^2 \frac{1}{2b}\exp(-(x-a)/b)dx = [\text{variabelbyte } y = a - x] \\ &= 2\int_0^{\infty} y^2 \frac{1}{2b}\exp(-y/b)dy = 2\left[-y^2\frac{1}{2}\exp(-y/b)\right]_0^{\infty} + 2\int_0^{\infty} y\exp(-y/b)dy \\ &= 2\left[-yb\exp(-y/b)\right]_0^{\infty} + 2\int_0^{\infty} b\exp(-y/b)dy = 2\left[-b^2\exp(-y/b)\right]_0^{\infty} = 2b^2.\end{aligned}$$

(b) X har parametrar $a = 0$ och $b = \sigma/\sqrt{2}$. Vi vill hitta värdet på c så att

$$\frac{\alpha}{2} = \int_{c\sigma}^{\infty} f(x)dx = \int_{c\sigma}^{\infty} \frac{1}{\sqrt{2}\sigma}\exp(-\sqrt{2}|x|/\sigma)dx = \left[\frac{1}{2}\exp(-\sqrt{2}|x|/\sigma)\right]_{c\sigma}^{\infty} = \frac{1}{2}\exp(-\sqrt{2}c).$$

Alltså har vi

$$\alpha = \exp(-\sqrt{2}c) \Rightarrow c = -\frac{\log(\alpha)}{\sqrt{2}}.$$

Med $\alpha = 0.05$ har vi $c = -\log(0.05)/\sqrt{2} \approx 2.12$.

(c) Vi beräknar först de två stickprovsmomenten

$$M_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = 2.0541, \quad M_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 8.4072.$$

Vi har $\mathsf{E}(X) = a$ och $\mathsf{E}(X^2) = \mathsf{V}(X) + \mathsf{E}(X)^2 = 2b^2 + a^2$. Alltså ska vi lösa ekvationssystemet

$$\begin{cases} M_1 &= a \\ M_2 &= 2b^2 + a^2. \end{cases}$$

Vi får alltså $a^* = M_1 = 2.0541$ och $2b^2 = M_2 - M_1^2 \Rightarrow b^* = \sqrt{(M_2 - M_1^2)/2} = 1.4470$.

(d) Skattningen är väntevärdesriktig eftersom

$$\mathsf{E}(a^*) = \mathsf{E}\left(\frac{1}{20}\sum_{i=1}^{20} X_i\right) = \frac{1}{20}\sum_{i=1}^{20}\mathsf{E}\left(X_i\right) = a.$$

Variansen av a^* är

$$\mathsf{E}(a^*) = \mathsf{V}\left(\frac{1}{20}\sum_{i=1}^{20} X_i\right) = \frac{1}{20^2}\mathsf{V}\left(\sum_{i=1}^{20} X_i\right) = \frac{20\cdot 2b^2}{20^2} = \frac{b^2}{10}.$$

Enligt centrala gränsvärdessatsen har vi att a^* är approximativt $\mathsf{N}(a, b^2/10)$-fördelad. Alltså får vi ett konfidensintervall som

$$I_a = \left[a^* \pm z_{0.025}\cdot\frac{b}{\sqrt{10}}\right] = \left[2.0541 \pm 1.96\cdot\frac{1.4470}{\sqrt{10}}\right] = [2.0541 \pm 0.8969] = [1.1573,\ 2.9510].$$

6. (a) Låt x_i beteckna mätningarna i Göteborg och y_i mätningarna i den nya staden. Vi beräknar först stickprovsmedelvärde och stickprovsvarians hos datan för respektive stad:

	Väntevärde	Varians
Göteborg	10.0714	2.9724
Ny stad	11.0571	21.3162

Vi antar att $x \sim \mathsf{N}(\mu_x, \sigma_x^2)$ och $y \sim \mathsf{N}(\mu_y, \sigma_y^2)$ och vill testa $H_0 : \sigma_x = \sigma_y$ mot $H_1 : \sigma_x \neq \sigma_y$ på nivå $\alpha = 0.05$. Vi beräknar teststorheten $T_{obs} = s_y^2/s_x^2 = 21.3162/2.9724 = 7.1714$, som under H_0 är $F(6,6)$-fördelad. Enligt formelsamlingen är $F_{\alpha/2}(6,6) = 5.8198$ och eftersom $T_{obs} > F_{\alpha/2}(6,6)$ kan vi förkasta H_0.

(b) Vi vill nu testa $H_0 : \mu_x = \mu_y$ mot $H_1 : \mu_x < \mu_y$. Eftersom vi kunde förkasta att varianserna är lika i de två städerna utför vi testet under antagandet att varianserna är olika. Vi beräknar ett ensidigt konfidensintervall för $\mu_y - \mu_x$ enligt

$$I_{\mu_y - \mu_x} = \left[\bar{y} - \bar{x} - t_\alpha(f)\sqrt{\frac{s_x^2}{7} + \frac{s_y^2}{7}}, \infty\right],$$

där

$$f = \frac{(s_x^2/7 + s_2^2/7)^2}{\frac{(s_x^2/7)^2}{6} + \frac{(s_y^2/7)^2}{6}} = 6\frac{(s_x^2 + s_y^2)^2}{s_x^4 + s_y^4} \approx 7.64.$$

Med $\alpha = 0.05$ har vi $t_\alpha(8) = 1.8595$ och får

$$I_{\mu_y - \mu_x} = [0.9857 - 1.8595 \cdot 1.8627, \infty] = [-2.4780, \infty].$$

Eftersom intervallet täcker noll kan vi inte förkasta H_0.

(c) Vi vill nu testa $H_0 : \mu_y \leq 10$ mot $H_1 : \mu_y > 10$. Ett undre begränsat konfidensintervall för μ_y ges av

$$\begin{aligned} I_{\mu_y} &= \left[\bar{y} - t_\alpha(6)s_y/\sqrt{7}, \infty\right] \\ &= \left[\bar{y} - 1.94324.6169/\sqrt{7}, \infty\right] \\ &= [7.6662, \infty]. \end{aligned}$$

Eftersom intervallet täcker 10 kan vi inte förkasta H_0.

(d) Om vi inte förkasta H_0 trots att H_1 är sann gör vi ett Typ 2 fel. Sannolikheten att inte göra ett typ 2 fel kallas för testets styrka, och ges alltså av $1 - \beta$ där $\beta = \mathsf{P}($förkasta ej $H_0 | H_1$ sann). Styrkefunktionen för testet i (c) visar testets styrka som funktion av det sanna värdet på μ_y. Denna funktion visas nedan: