1. (a) Låt B beteckna händelsen att en elev har druckit alkohol och låt A_7, A_8, A_9 beteckna händelserna att en elev går i årskurs 7, 8 respektive 9. Vi söker

$$\mathsf{P}(A_9 \cap B) = \mathsf{P}(B|A_9)\mathsf{P}(A_9) = 0.45 \cdot 0.35 = 0.1575.$$

(b) Vi har att händelserna A_7, A_8, A_9 är disjunkta och får

$$\begin{aligned}\mathsf{P}(B) &= \mathsf{P}(B \cap A_7) + \mathsf{P}(B \cap A_8) + \mathsf{P}(B \cap A_9)\\ &= \mathsf{P}(B|A_7)\mathsf{P}(A_7) + \mathsf{P}(B|A_8)\mathsf{P}(A_8) + \mathsf{P}(B|A_9)\mathsf{P}(A_9)\\ &= 0.15 \cdot 0.4 + 0.32 \cdot 0.25 + 0.45 \cdot 0.35 = 0.2975\end{aligned}$$

(c) Vi söker $\mathsf{P}(A_9|B)$ och använder Bayes sats:

$$\mathsf{P}(A_9|B) = \frac{\mathsf{P}(B|A_9)\mathsf{P}(A_9)}{\mathsf{P}(B)} = \frac{0.45 \cdot 0.35}{0.2975} = 0.5294.$$

2. (a) Låt A beteckna temperaturen och B luftfuktigheten. Vi får följande medelvärden och varianser

A	B	Skattningar
L	L	$\bar{y}_{11} = 243, s_{11}^2 = 242$
H	L	$\bar{y}_{21} = 133, s_{21}^2 = 8.0$
L	H	$\bar{y}_{12} = 183.5, s_{12}^2 = 4.5$
H	H	$\bar{y}_{22} = 93.5, s_{22}^2 = 40.5$

Skattningarna av effekterna ges nu av

$$\begin{aligned}\mu &= \frac{\bar{y}_{11} + \bar{y}_{21} + \bar{y}_{12} + \bar{y}_{22}}{4} = 163.5\\ A &= \frac{-\bar{y}_{11} + \bar{y}_{21} - \bar{y}_{12} + \bar{y}_{22}}{4} = -50\\ B &= \frac{-\bar{y}_{11} - \bar{y}_{21} + \bar{y}_{12} + \bar{y}_{22}}{4} = -24.75\\ AB &= \frac{\bar{y}_{11} - \bar{y}_{21} - \bar{y}_{12} + \bar{y}_{22}}{4} = 5\end{aligned}$$

(b) Den poolade skattningen av variansen ges av $s^2 = (s_{11}^2 + s_{21}^2 + s_{12}^2 + s_{22}^2)/4 = 73.75$.

(c) Antalet frihetsgrader är 2^2 och vi har t-kvantilen $t_{0.025}(4) = 2.7764$. Den skattade standardavvikelsen hos effektskattningarna ges av $s/\sqrt{8} = 3.0362$. Vi får konfidensintervallen

$$\begin{aligned}I_A &= [A \pm t_{0.025}(4)d(A)] = [-58.43, -41.57]\\ I_B &= [B \pm t_{0.025}(4)d(B)] = [-33.18, -16.32]\\ I_{AB} &= [AB \pm t_{0.025}(4)d(AB)] = [-3.43, 13.43].\end{aligned}$$

Eftersom I_A och I_B ej täcker noll är huvudeffekterna signifikanta på nivå 5%, men samspelseffekten är inte signifikant.

3. (a) Låt X beteckna mängden Ibuprofen i en tablett. Vi har då

$$\mathsf{P}(X < 200) = \mathsf{P}\left(\frac{X - 205}{10} < \frac{200 - 205}{10}\right) = \Phi(-0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085.$$

(b) Låt $Y = \sum_{i=1}^{10} X_i$, där $X_i \sim \mathsf{N}(205, 10^2)$ beteckna den totala mängden i en förpackning. Vi har då att $Y \sim \mathsf{N}(10 \cdot 205, 10 \cdot 10^2) = \mathsf{N}(2050, 1000)$ och

$$\mathsf{P}(Y > 2000) = 1 - \Phi\left(\frac{2000 - 2050}{10\sqrt{10}}\right) = \Phi(1.5811) = 0.9431.$$

(c) Låt $Y = \sum_{i=1}^{10} X_i$, där $X_i \sim \mathsf{N}(205, \sigma^2)$ beteckna den totala mängden i en förpackning. Vi har då att $Y \sim \mathsf{N}(10 \cdot 205, 10 \cdot \sigma^2)$ och

$$\mathsf{P}(Y > 2000) = 1 - \Phi\left(\frac{2000 - 2050}{\sigma\sqrt{10}}\right) = \Phi\left(\frac{50}{\sigma\sqrt{10}}\right).$$

Vi vill att sannolikheten ska vara åtminstone 0.9 och ser i Tabell 1 i formelsamlingen att $50/(\sigma\sqrt{10})$ då måste vara åtminstone 1.29. Detta ger att

$$\frac{50}{\sigma\sqrt{10}} > 1.29 \Rightarrow \frac{50}{1.29\sqrt{10}} > \sigma.$$

Alltså måste $\sigma < 12.2569$.

4. (a) Skattningen av β_1 ges av $\beta_1^* = S_{xy}/S_{xx} = 20.3658/167.9901 = 0.1212$ och skattningen av β_0 ges av $\beta_0^* = \bar{y} - \beta_1^*\bar{x} = 5.4084 - 0.1212 \cdot 6.0981 = 4.6691$. Vi beräknar $Q_0 = S_{yy} - S_{xy}^2/S_{xx} = 14.1544$ och får $s^2 = Q_0/15 = 0.9436$.

(b) Vi använder $\alpha = 0.05$ och har då $t_{\alpha/2}(15) = 2.1314$. Vi vill göra en skattning med $x_0 = \log(44) = 3.7842$ och får prediktionsintervallet

$$\begin{aligned} I_{Y(x_0)} &= \left(\beta_0^* + \beta_1^* x_0 \pm t_{\alpha/2}(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}\right) \\ &= \left(5.1279 \pm 2.1314 \cdot 0.9714\sqrt{1 + \frac{1}{17} + \frac{(3.7842 - 6.0981)^2}{167.9901}}\right) \\ &= \left(5.1279 \pm 2.1314 \cdot 0.9714\sqrt{1 + \frac{1}{17} + 0.0319}\right) \\ &= [2.9655, 7.2902]. \end{aligned}$$

Vi ser att $\log(113.6) = 4.7327$ ligger i detta intervall.

(c) Vi vill testa $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ på nivå 5%. Ett konfidensintervall för β_1 ges av

$$I_{\beta_1} = \left[\beta_1^* \pm t_{\alpha/2}(15)\frac{s}{\sqrt{S_{xx}}}\right] = \left[0.1212 \pm 2.1314\frac{0.9714}{12.9611}\right] = [-0.0385, 0.2810].$$

Eftersom intervallet täcker noll kan vi inte förkasta H_0. Även om vi hade kunnat förkasta H_0 så betyder detta självfallet inte att antalet storkar påverkar födelsetalet. För att kunna svara på denna kausala fråga behöver vi mer information om datan. Det kan till exempel finnas “confounding variables”, som landets area, som påverkar båda variablerna.

5. (a) Låt X beteckna koktiden och M_X för medianen av fördelningen för X. Vi vill testa $H_0 : M_X = 200$ mot $H_1 : M_X > 200$. Om Y betecknar antal mätningar större än 200 så gäller att $Y \sim \mathrm{Bin}(10, 0.5)$ om H_0 är sann. Från resultaten har vi att 7 mätningar är större än 200 och vi får att p-värdet ges av

$$\begin{aligned} \mathsf{P}(Y \geq 7) &= \mathsf{P}(Y = 7) + \mathsf{P}(Y = 8) + \mathsf{P}(Y = 9) + \mathsf{P}(Y = 10) \\ &= \binom{10}{7}\frac{1}{2^{10}} + \binom{10}{8}\frac{1}{2^{10}} + \binom{10}{9}\frac{1}{2^{10}} + \binom{10}{10}\frac{1}{2^{10}} \\ &= \frac{1}{2^{10}}(120 + 45 + 10 + 1) = \frac{176}{1024} = 0.1719 \end{aligned}$$

Eftersom p-värdet är större än $\alpha = 0.05$ kan vi inte förkasta H_0.

(b) Vi antar $X \sim \mathsf{N}(\mu_x, \sigma_x^2)$ och vill testa $H_0 : \mu_x = 200$ mot $H_1 : \mu_x > 200$. Vi beräknar $\bar{x} = 204.3$ och $s_x = 7.2426$ och får teststorheten

$$T_{obs} = \frac{204.3 - 200}{s/\sqrt{10}} = 1.8775.$$

Under H_0 ska denna vara $t(9)$-fördelad. I formelsamlingen ser vi att 5%-kvantilen för denna fördelning är 1.8331. Eftersom T_{obs} är större än detta värde kan vi förkasta H_0.

(c) Fördelen med teckentestet är att vi inte behöver göra något antagande om att X är normalfördelad, men nackdelen är att testet har lägre styrka i fallet då X faktiskt är normalfördelad.

(d) Låt $Y \sim \mathsf{N}(\mu_y, \sigma_y^2)$ beteckna koktiderna för den dyra vattenkokaren. Vi vill testa $H_0 : \mu_x = \mu_y$ mot $H_1 : \mu_x > \mu_y$. Vi beräknar $\bar{y} = 199.2$ och $s_y = 13.1892$. Eftersom vi antar att X och Y har samma varians beräknar vi en poolad skattning av variansen, $s_p^2 = (s_x^2 + s_y^2)/2 = 10.6398^2$. Vi har nu teststorheten

$$T_{obs} = \frac{\bar{x} - \bar{y}}{s_p\sqrt{1/10 + 1/10}} = 1.0718$$

som ska vara $t(18)$-fördelad under H_0. I formelsamlingen ser vi att 5%-kvantilen för denna fördelning ges av 1.7341. Eftersom T_{obs} inte är större än detta värde kan vi inte förkasta H_0.

6. (a) Vi får marginalfördelningen för X_1 genom att integrera täthetsfunktionen över x_2:

$$\begin{aligned} f(x_1) &= \int_1^\infty f(x_1, x_2)dx_2 = a(a+1)\int_1^\infty (x_1 + x_2 - 1)^{-(a+2)}dx_2 \\ &= a(a+1)\left[-\frac{1}{a+1}(x_1 + x_2 - 1)^{-(a+1)}\right]_{x_2=1}^\infty = ax_1^{-(a+1)}. \end{aligned}$$

(b) Sannolikheten ges av

$$\mathsf{P}(X_1 > 2) = \int_2^\infty f(x_1)dx_1 = \int_2^\infty ax_1^{-(a+1)}dx_1 = a\left[-\frac{1}{a}x_1^{-a}\right]_2^\infty = 2^{-a}.$$

(c) Givet att $a > 1$ har vi

$$\begin{aligned} \mathsf{E}(X_1) &= \int_1^\infty x_1 f(x_1)dx_1 = \int_1^\infty x_1 a x_1^{-(a+1)}dx_1 = \int_1^\infty a x_1^{-a}dx_1 \\ &= a\left[\frac{1}{1-a}x_1^{-a+1}\right]_1^\infty = \frac{a}{a-1}. \end{aligned}$$

(d) Den betingade fördelningen ges av

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \frac{a(a+1)(x_1 + x_2 - 1)^{-(a+2)}}{ax_1^{-(a+1)}} = (a+1)x_1^{a+1}(x_1 + x_2 - 1)^{-(a+2)}.$$

(e) Givet stickprovet har vi likelihooden $L(a) = \prod_{i=1}^n ax_{1,i}^{-(a+1)}$ och log-likelihooden

$$l(a) = \sum_{i=1}^n (\log(a) - (a+1)\log(x_{1,i})) = n\log(a) - (a+1)\sum_{i=1}^n \log(x_{1,i}).$$

Vi deriverar $l(a)$ med avseende på a och sätter derivatan till noll:

$$\frac{\partial}{\partial a}l(a) = \frac{n}{a} - \sum_{i=1}^n \log(x_{1,i}) = 0 \Rightarrow a = \frac{n}{\sum_{i=1}^n \log(x_{1,i})}.$$