1. (a) Vi har $X =$ "vikten hos en bakpotatis" $\in N(200,\ 32)$, vilket ger $\mathsf{P}(X > 246) = 1 - \mathsf{P}(X \le 246) =$
$= 1 - \mathsf{P}(\frac{X-200}{32} \le \frac{246-200}{32}) = 1 - \Phi(\frac{246-200}{32}) = 1 - \Phi(1.44) = 1 - 0.9251 = 0.0749.$

   (b) Vi har nu att $Y =$ "vikten hos 9 olika bakpotatisar" $= \sum_{i=1}^{9} X_i \in N\left(\mathsf{E}(\sum_{i=1}^{9} X_i),\ \sqrt{\mathsf{V}(\sum_{i=1}^{9} X_i)}\right) =$
$= N\left(\sum_{i=1}^{9}\mathsf{E}(X_i),\ \sqrt{\sum_{i=1}^{9}\mathsf{V}(X_i)}\right) = N\left(9\cdot 200,\ \sqrt{9\cdot 32^2}\right) = N\left(1800,\ 32\sqrt{9}\right) = N(1800,\ 96)$
så att $\mathsf{P}(Y > 2214) = 1 - \mathsf{P}(Y \le 2214) = 1 - \Phi(\frac{2214-1800}{96}) = 1 - \Phi(4.31) \approx 1 - 1.0 = 0.$

   (c) Vi har att $Y =$ "vikten av 9 bakpotatisar" $\in N(1800,\ 96)$ och vill hitta $a =$ "påsens hållfasthet" så att $\mathsf{P}(Y > a) \le 0.05$. Det innebär att $\mathsf{P}(Y > a) = 1 - \Phi(\frac{a-1800}{96}) \le 0.05 \Rightarrow$
$\frac{a-1800}{96} \ge \lambda_{0.05} \Rightarrow a \ge 1800 + 96\cdot \underbrace{\lambda_{0.05}}_{1.6449} = 1957.9$ g.

2. Vi anpassar modellen $y_i = \alpha + \beta x_i + \varepsilon_i$ där $\varepsilon_i \in N(0,\ \sigma)$ och oberoende. Vi vill testa om det finns en linjär trend, dvs $H_0\colon \beta = 0$ mot $H_1\colon \beta \neq 0$. (Vi har inte någon åsikt om åt vilket håll trenden skulle gå.)

   Vi har skattningarna $\beta^* = \frac{S_{xy}}{S_{xx}} = \frac{-13.475}{143} = -0.0942$ och $\sigma^* = s = \sqrt{\frac{Q_0}{n-2}} = \sqrt{\frac{0.08113}{12-2}} = 0.0901.$
   Vi vet också att $\beta^* \in N\left(\beta,\ \frac{\sigma}{\sqrt{S_{xx}}}\right) = N\left(\beta,\ \frac{\sigma}{\sqrt{143}}\right)$ med medelfelet $\mathsf{d}(\beta^*) = \frac{\sigma^*}{\sqrt{S_{xx}}} = \frac{s}{\sqrt{S_{xx}}} =$
   $= \frac{0.0901}{\sqrt{143}} = 0.0075.$

   Konfidensintervall: Eftersom $I_\beta = \left(\beta^* \pm t_{0.05/2}(n-2)\cdot \mathsf{d}(\beta^*)\right) = \left(-0.0942 \pm \underbrace{t_{0.025}(10)}_{2.23}\cdot\frac{0.0901}{\sqrt{143}}\right) =$
   $= (-0.11,\ -0.08)$ inte innehåller $\beta = 0$ kan H_0 förkastas på signifikansnivån 5 %.
   Det finns alltså en linjär trend.

   Teststorhet: Eftersom $t = \left|\frac{\beta^*-0}{\mathsf{d}(\beta^*)}\right| = \left|\frac{-0.0942-0}{0.0901/\sqrt{143}}\right| = |-12.5| = 12.5 > t_{0.05/2}(n-2) = t_{0.025}(10) =$
   $= 2.23$ kan H_0 förkastas på signifikansnivå 5 %. Det finns alltså en linjär trend.

3. Vi vill beräkna $\mathsf{P}(Y \ge 58\,000)$ där $Y =$ "sammanlagt bidrag".

   Eftersom $Y = \sum_{i=1}^{1000} X_i$ där $X_i =$ "bidrag från medlem nr i" och X_i är oberoende av varandra men har samma fördelning, har vi att $\mathsf{E}(Y) = \mathsf{E}(\sum_{i=1}^{1000} X_i) = \sum_{i=1}^{1000}\mathsf{E}(X_i) = 1000\mathsf{E}(X_i)$, att
   $\mathsf{V}(Y) = \mathsf{V}(\sum_{i=1}^{1000} X_i) = \sum_{i=1}^{1000}\mathsf{V}(X_i) = 1000\mathsf{V}(X_i)$ och att $\mathsf{D}(Y) = \sqrt{\mathsf{V}(Y)} = \mathsf{D}(X_i)\sqrt{1000}.$

   Eftersom vi har många medlemmar ger centrala gränsvärdessatsen att $Y \underset{\sim}{\in} N\left(1000\mathsf{E}(X_i),\ \mathsf{D}(X_i)\sqrt{1000}\right).$

   För att beräkna $\mathsf{E}(X_i)$ och $\mathsf{D}(X_i)$ måste vi först ta fram fördelningen för X_i.

   Vi har fått veta att $\mathsf{P}(X_i = 50) = \mathsf{P}(X_i = 100)$ och att $\mathsf{P}(X_i = 0) = 0.20$. Eftersom sannolikheterna måste summera till ett får vi $p + p + 0.20 = 1 \Rightarrow p = \frac{1-0.20}{2} = 0.4$ och fördelningen för X_i ges alltså av

k (kr)	0	50	100
$p_X(k)$	0.2	0.4	0.4

Det ger nu att $\mathsf{E}(X_i) = \sum_k k \cdot p_X(k) = 0 \cdot 0.2 + 50 \cdot 0.4 + 100 \cdot 0.4 = 60$ kr.

Vi har också att $\mathsf{E}(X_i^2) = \sum_k k^2 \cdot p_X(k) = 0^2 \cdot 0.2 + 50^2 \cdot 0.4 + 100^2 \cdot 0.4 = 5000$ så att

$\mathsf{V}(X_i) = \mathsf{E}(X_i^2) - \mathsf{E}^2(X_i) = 5000 - 60^2 = 1400\,\text{kr}^2$ och $\mathsf{D}(X_i) = \sqrt{\mathsf{V}(X_i)} = \sqrt{1400} \approx 37.42$ kr.

Vi får alltså att $Y \underset{\sim}{\in} N\left(1000 \cdot 60,\ \sqrt{1400} \cdot \sqrt{1000}\right) = N(60\,000,\ 1183.2)$ och

$\mathsf{P}(Y > 58\,000) = 1 - \mathsf{P}(Y \le 58\,000) \approx 1 - \Phi(\frac{58\,000 - 60\,000}{1183.2}) = 1 - \Phi(-1.69) = \Phi(1.69) = 0.9545.$

Två vanliga lösningsansatser som dessvärre kräver kunskaper utanför kursens ram:

Alt.1: Multinomial Sätt X_0 = "antal som inte ger något bidrag" $\in Bin(1000,\ 0.2)$, X_{50} = "antal som ger 50 kr" $\in Bin(1000,\ 0.4)$, X_{100} = "antal som ger 100 kr" $\in Bin(1000,\ 0.4)$.

Eftersom $X_0 + X_{50} + X_{100} = 1000$ är X_0, X_{50} och X_{100} inte oberoende! I själva verket är (X_0, X_{50}, X_{100}) multinomialfördelad (se Blom kapitel 7.5). Det innebär att vi måste ta hänsyn till kovarianserna $\mathsf{C}(X_i, X_i) = -np_ip_j$ för $i \neq j$.

Med Y = "totalt bidrag" $= 0 \cdot X_0 + 50 \cdot X_{50} + 100 \cdot X_{100} = 50X_{50} + 100X_{100}$ får vi

$\mathsf{E}(Y) = 50\mathsf{E}(X_{50}) + 100\mathsf{E}(X_{100}) = 50 \cdot 1000 \cdot 0.4 + 100 \cdot 1000 \cdot 0.4 = 60\,000$ kr medan
$\mathsf{V}(Y) = 50^2\mathsf{V}(X_{50}) + 100^2\mathsf{V}(X_{100}) + 2 \cdot 50 \cdot 100 \cdot \mathsf{C}(X_{50}, X_{100}) =$
$= 50^2 \cdot 1000 \cdot 0.4(1 - 0.4) + 100^2 \cdot 1000 \cdot 0.4(1 - 0.4) - 2 \cdot 50 \cdot 100 \cdot 1000 \cdot 0.4 \cdot 0.4 = 1\,400\,000\,\text{kr}^2.$

Eftersom både X_{50} och X_{100} kan normalapproximeras $(1000 \cdot 0.4(1 - 0.4) > 10)$ får vi att
$Y \underset{\sim}{\in} N\left(60\,000,\ \sqrt{1\,400\,000}\right) = N(60\,000,\ 1183.2)$ och $\mathsf{P}(Y > 58\,000) = 1 - \mathsf{P}(Y \le 58\,000) \approx$
$\approx 1 - \Phi(\frac{58\,000 - 60\,000}{1183.2}) = 1 - \Phi(-1.69) = \Phi(1.69) = 0.9545.$

Alt.2: Betinga Sätt X_i = "individ i:s bidrag". Vi vet att $\mathsf{P}(X_i = 0) = 0.2$ och $\mathsf{P}(X_i \neq 0) = 1 - 0.2 = 0.8$. Dessutom vet vi att $\mathsf{P}(X_i = 50 \mid X_i \neq 0) = \mathsf{P}(X_i = 100 \mid X_i \neq 0) = 0.5$.

Det ger att $\mathsf{E}(X_i \mid X_i \neq 0) = 50 \cdot 0.5 + 100 \cdot 0.5 = 75$ kr och
$\mathsf{V}(X_i \mid X_i \neq 0) = \mathsf{E}(X_i^2 \mid X_i \neq 0) - \mathsf{E}^2(X_i \mid X_i \neq 0) = 50^2 \cdot 0.5 + 100^2 \cdot 0.5 - 75^2 = 625\,\text{kr}^2.$
Vi har naturligtvis också att $\mathsf{E}(X_i \mid X_i = 0) = 0$ kr och $\mathsf{V}(X_i \mid X_i = 0) = 0\,\text{kr}^2$.

Räknereglerna för betingade väntevärden och varianser (se Blom kapitel 5.7) ger att
$\mathsf{E}(X_i) = \mathsf{E}(\mathsf{E}(X_i \mid X_i)) = \mathsf{E}(X_i \mid X_i = 0) \cdot \mathsf{P}(X_i = 0) + \mathsf{E}(X_i \mid X_i \neq 0) \cdot \mathsf{P}(X_i \neq 0) = 0 \cdot 0.2 + 75 \cdot 0.8 = 60$ kr och $\mathsf{V}(X_i) = \mathsf{E}(\mathsf{V}(X_i \mid X_i)) + \mathsf{V}(\mathsf{E}(X_i \mid X_i))$. Eftersom

$\mathsf{E}(\mathsf{V}(X_i \mid X_i)) = \mathsf{V}(X_i \mid X_i = 0) \cdot \mathsf{P}(X_i = 0) + \mathsf{V}(X_i \mid X_i \neq 0) \cdot \mathsf{P}(X_i \neq 0) = 0 \cdot 0.2 + 625 \cdot 0.8 = 500\,\text{kr}^2$
medan $\mathsf{V}(\mathsf{E}(X_i \mid X_i)) = \mathsf{E}(\mathsf{E}^2(X_i \mid X_i)) - \mathsf{E}^2(\mathsf{E}(X_i \mid X_i)) =$
$= \mathsf{E}^2(X_i \mid X_i = 0) \cdot \mathsf{P}(X_i = 0) + \mathsf{E}^2(X_i \mid X_i \neq 0) \cdot \mathsf{P}(X_i \neq 0) - 60^2 = 0^2 \cdot 0.2 + 75^2 \cdot 0.8 - 60^2 = 900\,\text{kr}^2$
har vi att $\mathsf{V}(X_i) = 500 + 900 = 1400\,\text{kr}^2$.

CGS ger $\sum_{i=1}^{1000} X_i \underset{\sim}{\in} N\left(1000 \cdot 60,\ \sqrt{1000 \cdot 1400}\right) = N(60\,000,\ 1183.2)$ och $\mathsf{P}(Y > 58\,000) =$
$= 1 - \mathsf{P}(Y \le 58\,000) \approx 1 - \Phi(\frac{58\,000 - 60\,000}{1183.2}) = 1 - \Phi(-1.69) = \Phi(1.69) = 0.9545.$

4. (a) Sätt X_i = "hastighet hos fordon i på 50-sträckan" $\in N(\mu_x,\ \sigma_x)$. Vi har fått skattningarna

$$\mu_x^* = \bar{x} = \frac{1}{49}\sum_{i=1}^{49} x_i = 37.8\ \text{km/h och}\ \sigma_x^* = s_x = \sqrt{\frac{1}{49-1}\sum_{i=1}^{49}(x_i - \bar{x})^2} = 5.95\ \text{km/h}.$$

Eftersom $\mu_x^* = \bar{X} \in N\left(\mu_x,\ \frac{\sigma_x}{\sqrt{49}}\right)$ där medelfelet blir
$\mathsf{d}(\mu_x^*) = \frac{\sigma_x^*}{\sqrt{49}} = \frac{s_x}{\sqrt{49}} = \frac{5.95}{\sqrt{49}} \approx 0.85$ ges konfidensintervallet av

$$I_{\mu_x} = \left(\mu_x^* \pm t_{\alpha/2}(49-1) \cdot \mathsf{d}(\mu_x^*)\right) = \left(37.8 \pm \underbrace{t_{0.025}(48)}_{2.01} \cdot \frac{5.95}{\sqrt{49}}\right) = (36.1,\ 39.5)\ \text{km/h}.$$

Alternativ: Om vi dessutom sätter $Y_i =$ ”hastigheten hos fordon i på 30-sträckan” $\in N(\mu_y, \sigma_y)$ med skattningarna $\mu_y^* = \bar{y} = 32.0$ och $\sigma_y^* = s_y = 4.74$ och antar att $\sigma_x = \sigma_y = \sigma$ kan vi få en bättre skattning av standardavvikelsen:

$$\sigma^* = s = \sqrt{\frac{(49-1)s_x^2 + (36-1)s_y^2}{49-1+36-1}} = \sqrt{\frac{(49-1)\cdot 5.95^2 + (36-1)\cdot 4.74^2}{49-1+36-1}} \approx 5.47$$

som ger medelfelet $\mathsf{d}(\mu_x^*) = \dfrac{\sigma^*}{\sqrt{49}} = \dfrac{5.47}{\sqrt{49}} \approx 0.78$ så att intervallet istället blir

$$I_{\mu_x} = \left(\mu_x^* \pm t_{\alpha/2}(49-1+36-1)\cdot \mathsf{d}(\mu_x^*)\right) = \left(\bar{x} \pm \underbrace{t_{0.025}(83)}_{1.99}\cdot \frac{5.47}{\sqrt{49}}\right) = (36.2,\ 39.4)\text{ km/h}.$$

Detta intervall är aningen smalare och därmed lite bättre.

(b) Sätt $Y_i =$ ”hastighet hos fordon i på 30-sträckan” $\in N(\mu_y, \sigma_y)$. Vi har fått skattningarna

$$\mu_y^* = \bar{y} = \frac{1}{36}\sum_{i=1}^{36} y_i = 32.0\text{ km/h och } \sigma_y^* = s_y = \sqrt{\frac{1}{36-1}\sum_{i=1}^{36}(y_i - \bar{y})^2} = 4.74\text{ km/h}.$$

Om vi antar att $\sigma_x = \sigma_y = \sigma$ kan vi få en bättre skattning av standardavvikelsen:

$$\sigma^* = s = \sqrt{\frac{(49-1)s_x^2 + (36-1)s_y^2}{49-1+36-1}} = \sqrt{\frac{(49-1)\cdot 5.95^2 + (36-1)\cdot 4.74^2}{49-1+36-1}} \approx 5.47.$$

Eftersom $\mu_x^* = \bar{X} \in N\left(\mu_x, \dfrac{\sigma}{\sqrt{49}}\right)$ och $\mu_y^* = \bar{Y} \in N\left(\mu_y, \dfrac{\sigma}{\sqrt{36}}\right)$ har vi att

$$\mu_x^* - \mu_y^* = \bar{X} - \bar{Y} \in N\left(\mathsf{E}(\bar{X}-\bar{Y}), \sqrt{\mathsf{V}(\bar{X}-\bar{Y})}\right) = N\left(\mathsf{E}(\bar{X}) - \mathsf{E}(\bar{Y}), \sqrt{\mathsf{V}(\bar{X}) + (-1)^2\mathsf{V}(\bar{Y})}\right) =$$

$$= N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma^2}{49} + \frac{\sigma^2}{36}}\right) = N\left(\mu_x - \mu_y, \sigma\sqrt{\frac{1}{49} + \frac{1}{36}}\right) \text{ där medelfelet blir}$$

$\mathsf{d}(\mu_x^* - \mu_y^*) = \sigma^*\sqrt{\dfrac{1}{49} + \dfrac{1}{36}} = 5.47\sqrt{\dfrac{1}{49} + \dfrac{1}{36}} \approx 1.20$. Konfidensintervallet för differensen ges då av $I_{\mu_x - \mu_y} = \left(\mu_x^* - \mu_y^* \pm t_{\alpha/2}(49-1+36-1)\cdot \mathsf{d}(\mu_x^* - \mu_y^*)\right) =$

$$= \left(37.8 - 32.0 \pm \underbrace{t_{0.025}(83)}_{1.99}\cdot 5.47\sqrt{\frac{1}{49} + \frac{1}{36}}\right) = (3.4,\ 8.2)\text{ km/h}.$$

5. (a) Vi har en observation $x = 15$ av $X =$ ”antal defekta komponenter i en förpackning” $\in Bin(50, p)$ och skattar p med $p^* = \dfrac{x}{n} = \dfrac{15}{50} = 0.3$.

Eftersom $np(1-p) \approx 50\cdot 0.3\cdot 0.7 = 10.5 > 10$ kan vi normalapproximera X, dvs $X \underset{\sim}{\in} N(\mathsf{E}(X), \mathsf{D}(X)) = N\left(np, \sqrt{np(1-p)}\right) = N\left(50p, \sqrt{50p(1-p)}\right)$. Det ger i sin tur att

$$p^* = \frac{X}{n} = \frac{X}{50} \underset{\sim}{\in} N\left(\mathsf{E}(\frac{X}{n}), \sqrt{\mathsf{V}(\frac{X}{n})}\right) = N\left(\frac{\mathsf{E}(X)}{n}, \sqrt{\frac{\mathsf{V}(X)}{n^2}}\right) = N\left(\frac{np}{n}, \sqrt{\frac{np(1-p)}{n^2}}\right) =$$

$$= N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = N\left(p, \sqrt{\frac{p(1-p)}{50}}\right).$$

Vi har också medelfelet $\mathsf{d}(p^*) = \sqrt{\dfrac{p^*(1-p^*)}{50}} = \sqrt{\dfrac{0.3(1-0.3)}{50}} \approx 0.0648$ och konfidensintervallet

$$I_p = \left(p^* \pm \lambda_{\alpha/2}\cdot \mathsf{d}(p^*)\right) = (0.3 \pm \underbrace{\lambda_{0.025}}_{1.96}\cdot\sqrt{\frac{0.3\cdot 0.7}{50}}) = (0.17,\ 0.43).$$

(b) Sätt $X_i =$ ”antalet defekta komponenter i förpackning nr i” $\in Bin(50, p)$.

Kunden behöver köpa mer än tre förpackningar om $\displaystyle\sum_{i=1}^{3} X_i > 150 - 120 = 30$.

Eftersom X_i är oberoende och binomialfördelade med samma p gäller att

$$\sum_{i=1}^{3} X_i \in Bin(3\cdot 50, p) = Bin(150, p).$$

Vi har skattningen $p^* = 0.3$ och, eftersom $150p(1-p) \approx 150 \cdot 0.3(1-0.3) = 31.5 > 10$,

kan vi normalapproximera så att $\sum_{i=1}^{3} X_i \underset{\approx}{\in} N\left(150p,\ \sqrt{150p(1-p)}\right) = N\left(150p,\ \sqrt{150p(1-p)}\right)$.

Med skattade värden får vi $N\left(150 \cdot 0.3),\ \sqrt{150 \cdot 0.3(1-0.3)}\right) = N\left(45,\ \sqrt{31.5}\right)$ och

$\mathsf{P}(\sum_{i=1}^{3} X_i > 30) = 1 - \mathsf{P}(\sum_{i=1}^{3} X_i \leq 30) \approx 1 - \Phi(\dfrac{30-45}{\sqrt{31.5}}) = 1 - \Phi(-2.67) = \Phi(2.67) = 0.9962$.

6. (a) Vi har $X =$ "livslängden hos en apparat med en komponent" $\in Exp(\dfrac{1}{\mu})$ och

$Y =$ "livslängden för en apparat med två komponenter" $\in Exp(\dfrac{2}{\mu})$, dvs

$f_X(x) = \dfrac{1}{\mu}\, e^{-x/\mu}$ för $x > 0$, och $f_Y(y) = \dfrac{2}{\mu}\, e^{-2y/\mu}$ för $y > 0$. Dessutom är X och Y oberoende.

ML-skattningen av μ fås då som det μ-värde som maximierar

$$L(\mu) = f_X(x) \cdot f_Y(y) = \frac{1}{\mu}\, e^{-x/\mu} \cdot \frac{2}{\mu}\, e^{-2y/\mu} = \frac{2}{\mu^2}\, e^{-(x+2y)/\mu}.$$

Logaritmering ger $\ln L(\mu) = \ln 2 - 2\ln\mu - \dfrac{x+2y}{\mu}$.

Derivering ger $\dfrac{d \ln L(\mu)}{d\mu} = -\dfrac{2}{\mu} + \dfrac{x+2y}{\mu^2} = 0 \Rightarrow \mu^* = \dfrac{x+2y}{2}$.

(b) Eftersom $\mathsf{E}(X) = \mu$, $\mathsf{V}(X) = \mu^2$, $\mathsf{E}(Y) = \mu/2$ och $\mathsf{V}(Y) = (\mu/2)^2$ får vi

$\mathsf{E}(\mu^*) = \mathsf{E}(\dfrac{X+2Y}{2}) = \dfrac{\mathsf{E}(X) + 2\mathsf{E}(Y)}{2} = \dfrac{\mu + 2 \cdot \frac{\mu}{2}}{2} = \mu$ och

$\mathsf{V}(\mu^*) = \mathsf{V}(\dfrac{X+2Y}{2}) = [\text{oberoende}] = \dfrac{\mathsf{V}(X) + 2^2\mathsf{V}(Y)}{2^2} = \dfrac{\mu^2 + 4 \cdot (\frac{\mu}{2})^2}{2^2} = \dfrac{\mu^2}{2}$.

(c) Eftersom $\mathsf{E}(Z) = \mu/3$ och $\mathsf{V}(Z) = (\mu/3)^2$ har vi att $\mathsf{E}(\mu_z^*) = \mathsf{E}(3Z) = 3\mathsf{E}(Z) = 3 \cdot \dfrac{\mu}{3} = \mu$ och

$\mathsf{V}(\mu_z^*) = \mathsf{V}(3Z) = 3^2\mathsf{V}(Z) = 3^2 \cdot (\dfrac{\mu}{3})^2 = \mu^2$.

Båda skattningarna är väntevärdesriktiga men ML-skattningen har hälften så stor varians och är alltså bättre. Det är inte ett bra förslag (men man kan gå hem tidigare och slipper vänta ut båda apparaterna).
