1. Låt A vara händelsen att den testade individen har arthritis och låt D vara händelsen att testet detekterar arthritis. Vi vet då att
$$P(A) \approx 0.12 \quad P(D|A) \approx 0.87 \quad P(D|A^c) \approx 0.06$$
Bayes formel ger oss den sökta sannolikheten
$$P(A|D) = \frac{P(A)P(D|A)}{P(A)P(D|A)+P(A^c)P(D|A^c)} \approx \frac{0.1044}{0.1044+0.0528} \approx 0.66$$

2. (a) $\binom{7}{6} 0.95^6 \cdot 0.05^1 + \binom{7}{7} 0.95^7 \cdot 0.05^0 = 0.2573 + 0.6983 = 0.9556 \approx 0.956$

   (b) $0.9556^6 \approx 0.762$

   (c) Den sökta sannolikheten är $p = P(X \geq 36)$, där $X \sim \mathrm{Bin}(42, 0.95) \approx \mathrm{N}(39.9, \sqrt{1.995})$. Vi får
$$p \approx 1 - \Phi\left(\frac{35.5-39.9}{\sqrt{1.995}}\right) = 1 - \Phi(-3.12) = \Phi(3.12) \approx 0.999$$
med halvtalskorrektion och $\Phi(2.76) \approx 0.997$ utan. En exakt beräkning med Matlab ger resultatet $p = 0.9955$.

3. $P(T > s+t \,|\, T > s) = \frac{P(T>s+t,\, T>s)}{P(T>s)} = \frac{P(T>s+t)}{P(T>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t)$

4. (a) Nej, ty Cauchy-fördelningen saknar väntevärde och varians

   (b) Vi får
$$\begin{aligned}\mathrm{Var}[Z] &= \mathrm{Var}[2X-Y] = \mathrm{Cov}[2X-Y, 2X-Y] \\ &= 4\mathrm{Cov}[X,X] + \mathrm{Cov}[Y,Y] - 4\mathrm{Cov}[X,Y] \\ &= 4\mathrm{Var}[X] + \mathrm{Var}[Y] - 4\rho\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]} \\ &= 4 \cdot 1.5^2 + 2^2 - 4 \cdot 0.75 \cdot 1.5 \cdot 2 = 4\end{aligned}$$

5. Momentmetoden: $\mu = \lambda \Rightarrow \widehat{\lambda} = \bar{x}$. Trolighetsmetoden: medelst maximering av trolighetsfunktionen
$$L(\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_1!} = e^{-n\lambda}\lambda^{\sum_i x_i} / \left(\prod_i x_i!\right)$$
fås trolighetsskattningen $\widehat{\lambda} = \bar{x}$. Moment- och trolighetsskattningen överensstämmer alltså och är lika med $\widehat{\lambda} = \bar{x}$. Att denna skattning är väntevärdesriktig inses av att medelvärdet skattar väntevärdet väntevärdesriktigt. Det observerade utfallet av $\widehat{\lambda}$ är $\bar{x} = 103/10 = 10.3$. Poissonmodellen är ofta vettig när det gäller händelser som inträffar relativt sällan, som olyckor förhoppningsvis gör. Det antagande som jag kan tänka mig kanske är tveksamt är oberoendet. Det är ju så att om man under ett år får väldigt många olyckor, så kan det innebära en allmän uppryckning som leder till att antalet olyckor under de följande åren blir lägre än vad de skulle varit annars. Likaså kan ett antal i det närmaste olycksfria år leda till att folks olycksmedvetenhet minskar och att det allmänna slarvet ökar, vilket lätt leder till ett antal olyckor som inte skulle inträffat om vaksamheten hade varit högre. Men det kan ju vara så att olycksförekomsten inte diskuterats offentligt under åren som varit och då är nog oberoendeantagandet rimligt.

6. Man ska testa $H_0 : p \geq 0.15$ mot $H_1 : p < 0.15$ i modellen $f \sim \mathrm{Bin}(n,p)$. Praktiskt innebär detta att man ska slumpmässigt välja ut n oberoende enheter tillverkade med den nya metoden för kontroll och beräkna frekvensen felaktiga f. Att sannolikheten för ett typ-I-fel ska vara mindre än 0.05 betyder att testet ska vara på nivån $\alpha = 0.05$. Detta betyder i sin tur att $P(f \leq c) \leq 0.05$ under H_0, d.v.s då $p \geq 0.15$. Här är c det s.k kritiska värdet och testregeln är förkasta H_0 då $f \leq c$. Hur stort ska man då välja n? Antag att n är så stort att normalapproximation är tillåten. Då gäller att
$$0.05 = P(f \leq c) \approx \Phi\left(\frac{c-0.15n}{\sqrt{0.15 \cdot 0.85n}}\right) \Rightarrow \frac{c-0.15n}{\sqrt{0.15 \cdot 0.85n}} = -1.645$$
$$\Rightarrow c = 0.15n - 1.645\sqrt{0.15 \cdot 0.85n} = 0.15n - 0.5874\sqrt{n}$$
Då $p = 0.10$ vill man att sannolikheten att förkasta ska vara iallafall 0.90. Detta ger att

$$0.90 = P(f \leq c) \approx \Phi\left(\frac{c-0.10n}{\sqrt{0.10 \cdot 0.90n}}\right) = \Phi\left(\frac{0.05n-0.5874\sqrt{n}}{0.3\sqrt{n}}\right)$$

Således gäller

$$1.282 = \frac{0.05n-0.5874\sqrt{n}}{0.3\sqrt{n}} \Rightarrow \sqrt{n} = 19.44 \Rightarrow n \approx 378$$

(Det går alltså bra att normalapproximera i detta problemet.) Väljes stickprovsstorleken $n = 380$ och regeln förkasta $H_0 : p \geq 0.15$ till förmån för $H_1 : p < 0.15$ då frekvensen felaktiga $f \leq 45$, så blir nivån $\alpha \approx 0.05$ och upptäcktssannolikheten $1 - \beta \approx 0.90$, vilket var vad som krävdes. Väljes ett större värde på n gäller att förkastelseregeln ska vara $f \leq 0.15n - 0.5874\sqrt{n}$. Då håller testet nivån $\alpha \approx 0.05$ och styrkan $1 - \beta$ är $\approx \Phi\left(\frac{0.05n-0.5874\sqrt{n}}{0.3\sqrt{n}}\right) \gtrapprox 0.90$.

7. Ur formelsamlingen (se Ett exponentialfördelat stickprov) fås att $2\lambda n\bar{x} \sim \chi^2(2n)$. Antalet frihetsgrader är alltså $2n = 50$. Ur tabell $\chi^2(50)$-fördelningen, ser vi att 0.025- och 0.975-kvantilerna är 32.3574 och 71.4204 (jag väljer konfidensen 0.95 eftersom inget är sagt om denna i problemet). Således gäller

$$P(32.3574 \leq 2\lambda n\bar{x} \leq 71.4204) = 0.95$$

vilket räknas om till konfidensintervallet $0.000\,06037 \leq 10^6\,\lambda \leq 0.000\,13326$, d,v,s $60.37 \leq 10^6\,\lambda \leq 133.26$. Alternativt kan man, eftersom normalapproximationen av medelvärdet av $n = 25$ oberoende exponentialvariabler är ganska bra (j.f.r projektuppgift 1), bilda

$$1/\lambda = \mu = 10719.0 \pm 2.064 \cdot 10366.9/\sqrt{25} = 10719.0 \pm 2073.38$$

(obs $2.064 = t_{0.025}(24)$) och medelst invertering av alla leden komma fram till intervallet $0.000\,07817 \leq \lambda \leq 0.000\,11567$ eller $78.17 \leq \lambda \leq 115.67$. Obs att konfidensen i detta intervall endast är approximativ. Att det blev kortare än det exakt beräknade, beror nog mest på att utfallet av s råkade bli mindre än utfallet av $\bar{x}$. Normalt ska vi ha $s \approx \bar{x}$ då data är exponentialfördelade.

8. Vi får

$$\begin{aligned}
(n-2)s^2 &= \textstyle\sum_i (y_i - b_0 - b_1 x_i)^2 \\
&= \textstyle\sum_i (y_i - \bar{y} + b_1\bar{x} - b_1 x_i)^2 = \sum_i \left[(y_i - \bar{y}) - b_1(x_i - \bar{x})\right]^2 \\
&= \textstyle\sum_i (y_i - \bar{y})^2 - 2b_1 \sum_i (x_i - \bar{x})(y_i - \bar{y}) + b_1^2 \sum_i (x_i - \bar{x})^2 \\
&= S_{yy} - 2b_1 S_{xy} + b_1^2 S_{xx} = S_{yy} - 2S_{xy}^2/S_{xx} + S_{xy}^2/S_{xx} \\
&= S_{yy} - S_{xy}^2/S_{xx}
\end{aligned}$$

vilket skulle visas.