1. $\Pr(disease) = 0.01$ $\quad\quad$ $\Pr(\overline{disease}) = 0.99$

$\Pr(positive \mid disease) = 0.99$ $\quad\quad$ $\Pr(\overline{positive} \mid disease) = 0.01$

$\Pr(positive \mid \overline{disease}) = 0.03$ $\quad\quad$ $\Pr(\overline{positive} \mid \overline{disease}) = 0.97$

$$\Pr(positive) = \Pr(positive \mid disease)\Pr(disease) + \Pr(positive \mid \overline{disease})\Pr(\overline{disease})$$
$$= (0.99)(0.01) + (0.03)(0.99) = 0.0396$$
$$\Pr(disease \mid positive) = \frac{\Pr(positive \mid disease)\Pr(disease)}{\Pr(positive)} = \frac{(0.99)(0.01)}{0.0396} = 0.25$$

2. Frequency table of the scores:

Class	Midpoint (m_i)	Frequency	Cumulative Frequency
0 – 300	150	15000	15000
300 – 500	400	50000	65000
500 – 600	550	35000	100000
600 – 700	650	45000	145000
700 – 850	775	45000	190000
850 – 1000	925	7500	197500

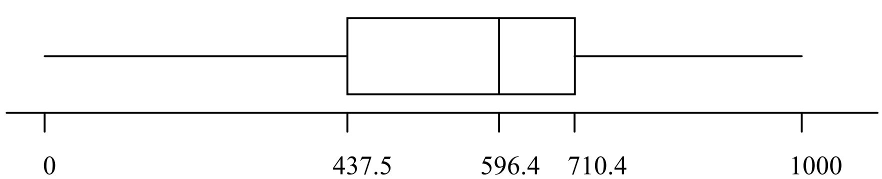
(a) $Min = 0$ $\quad\quad$ $Max = 1000$

$(197500)(0.25) = 49375$, $\quad\quad$ $Q_L = 300 + \dfrac{(200)(49375 - 15000)}{50000} = 437.5$

$(197500)(0.5) = 98750$, $\quad\quad$ $Median = 500 + \dfrac{(100)(98750 - 65000)}{35000} = 596.4$

$(197500)(0.75) = 148125$, $\quad\quad$ $Q_U = 700 + \dfrac{(150)(148125 - 145000)}{45000} = 710.4$

Boxplot :



(b) $$\mu = \frac{1}{197500}\sum_{i=1}^{k} m_i f_i = \frac{112562500}{197500} = 569.9$$

(c) $$\Pr(500 \le \overline{X} \le 600) = \Pr\left(\frac{500 - 569.9}{207/\sqrt{50}} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le \frac{600 - 569.9}{207/\sqrt{50}}\right)$$
$$\approx \Phi(1.03) - \Phi(-2.39) = 0.8485 - (1 - 0.9916) = 0.8401$$

(d) $(197500)(0.95) = 187625$

$$95^{\text{th}} \text{ percentile} = 700 + \frac{(150)(187625 - 145000)}{45000} = 842.1$$

3. (a) $\lambda = 1.5 \qquad t = 12$

$\lambda t = (1.5)(12) = 18$

(b) Let X be the number of failures in a year. Then $X \sim \wp(18)$.

$$\begin{aligned}\Pr\left(|X - E(X)| > \sqrt{Var(X)}\right) &= 1 - \Pr\left(|X - 18| \le \sqrt{18}\right)\\ &= 1 - \Pr\left(18 - \sqrt{18} \le X \le 18 + \sqrt{18}\right)\\ &= 1 - \Pr(13.757 \le X \le 22.243)\\ &= 1 - \Pr(14 \le X \le 22)\\ &= 1 - \sum_{i=14}^{22} \frac{e^{-18}(18)^i}{i!}\\ &= 1 - 0.7125 = 0.2875\end{aligned}$$

(c) $p = \Pr(\text{monthly loss less than } 1000)$

$$= \Pr(\text{number of monthly failures} \le 2) = \sum_{i=0}^{2} \frac{e^{-1.5}(1.5)^i}{i!} = 0.8088$$

Let Y be the number of months during the year with monthly loss less than 1000. Then $Y \sim b(12, 0.8088)$.

$$\Pr(Y \ge 8) = \sum_{i=8}^{12} \binom{12}{i} (0.8088)^i (0.1912)^{12-i} = 0.9386$$

(d) $T \sim \Gamma(25, 1.5)$

4. (a) A 95% confidence for p is

$$\hat{p} \pm Z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{13}{30} \pm (1.96)\sqrt{\frac{(13/30)(17/30)}{30}} = 0.4333 \pm 0.1773 = [25.6\%, 61.06\%]$$

(b) Although the sample proportion is more than 40%, the population proportion may not be more than 40% because the sampling error is 17.73% which is very large. Indeed the range of the confidence interval is so wide that the population proportion can be as low as 25.6% or as high as 61.6%. The newspaper had over-interpreted the experimental result in the statement.

(c) $$n = \frac{Z_{0.025}^2\, p(1-p)}{D^2} = \frac{(1.96)^2 (13/30)(17/30)}{(0.025)^2} = 1509.32 \approx 1510$$

(OR use the conservative sample size :

$$n = \frac{Z_{0.025}^2}{4D^2} = \frac{(1.96)^2}{4(0.025)^2} = 1536.64 \approx 1537$$)

5. (a) Test $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$ at $\alpha = 0.1$.

Test statistic: $F = \dfrac{S_1^2}{S_2^2}$

Reject H_0 if $F > F(7,7,0.05) = 3.79$ or $F < F(7,7,0.95) = \dfrac{1}{3.79} = 0.2639$.

From the data, $\bar{X} = 2577.75$, $S_1^2 = 8158.79$, $\bar{Y} = 2550.5$, $S_2^2 = 5485.43$.

$$F_{obs} = \frac{8158.79}{5485.43} = 1.4874 \Rightarrow 0.2639 < F_{obs} < 3.79$$

Hence do not reject H_0 at $\alpha = 0.1$. The data does not provide enough evidence that the population variances are unequal.

(b) Test $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$ at $\alpha = 0.05$.

Assume equal population variances.

Test statistic: $T = \dfrac{\bar{X} - \bar{Y}}{S_{pool}\sqrt{\frac{1}{8} + \frac{1}{8}}}$

Reject H_0 if $|T| > t_{14,0.025} = 2.145$.

From the data, $\bar{X} = 2577.75$, $S_1^2 = 8158.79$, $\bar{Y} = 2550.5$, $S_2^2 = 5485.43$.

$$S_{pool}^2 = \frac{(7)(8158.79) + (7)(5485.43)}{7+7} = 6822.11$$

$$|T_{obs}| = \left| \frac{2577.75 - 2550.5}{\sqrt{(6822.11)\left(\frac{1}{8} + \frac{1}{8}\right)}} \right| = 0.6598 < 2.145$$

Hence do not reject H_0 at $\alpha = 0.05$. The data does not provide enough evidence that in average the measurements from the two testers are difference.

(c) A 95% confidence interval for μ_1 is

$$\bar{X} \pm t_{7,0.025}\frac{S_1}{\sqrt{8}} = 2577.75 \pm (2.365)\sqrt{\frac{8158.79}{8}} = 2577.75 \pm 75.53 = [2502.22\,,2653.28]$$

A 95% confidence interval for μ_2 is

$$\bar{Y} \pm t_{7,0.025}\frac{S_2}{\sqrt{8}} = 2550.5 \pm (2.365)\sqrt{\frac{5485.43}{8}} = 2550.5 \pm 61.93 = [2488.57\,,2612.43]$$

(OR use the pooled sample variance,

$$\bar{X} \pm t_{14,0.025}\frac{S_{pool}}{\sqrt{8}} = 2577.75 \pm (2.145)\sqrt{\frac{6822.11}{8}} = 2577.75 \pm 62.64 = [2515.11\,,2640.39]$$

$$\bar{Y} \pm t_{14,0.025}\frac{S_{pool}}{\sqrt{8}} = 2550.5 \pm (2.145)\sqrt{\frac{6822.11}{8}} = 2550.5 \pm 62.64 = [2487.86\,,2613.14] \;)$$

(d) Let A be the event that the first interval contains μ_1, B be the event that the second interval contains μ_2 . Then the probability that both intervals will contain the corresponding parameters is

$$\Pr(A \cap B) = 1 - \Pr(\overline{A} \cup \overline{B}) \geq 1 - \Pr(\overline{A}) - \Pr(\overline{B}) = 1 - 0.05 - 0.05 = 0.9 \ .$$

Therefore we only know that the overall confidence is at least 90%. However, we don't know whether the overall confidence can be 95% or not.

(e) A 90% confidence interval for $\mu_1 - \mu_2$ is given by

$$(\overline{X} - \overline{Y}) \pm t_{14,0.05} S_{pool} \sqrt{\frac{1}{8} + \frac{1}{8}} = (2577.75 - 2550.5) \pm (1.761)\sqrt{(6822.11)\left(\frac{1}{8} + \frac{1}{8}\right)}$$
$$= 27.25 \pm 72.73 = [-45.48 , 99.98]$$

6. (a) $H_0 : p_V \geq p_N$ vs $H_0 : p_V < p_N$.

(b) Test statistic : $$Z = \frac{\hat{p}_V - \hat{p}_N}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_V} + \frac{1}{n_N}\right)}}$$

Reject H_0 if $Z < -Z_{0.05} = -1.645$.

From the data, $\hat{p}_V = \frac{9}{51} = 0.1765$, $\hat{p}_N = \frac{17}{45} = 0.3778$, $\hat{p} = \frac{9+17}{51+45} = 0.2708$

$$Z_{obs} = \frac{0.1765 - 0.3778}{\sqrt{(0.2708)(0.7292)\left(\frac{1}{51} + \frac{1}{45}\right)}} = -2.2149 < -1.645 \ .$$

Reject H_0 at $\alpha = 0.05$. The data provided strong evidence that the vaccination is effective in protecting the laboratory animals from that disease.

(c) A 95% confidence interval for $p_V - p_N$ is given by

$$(\hat{p}_V - \hat{p}_N) \pm Z_{0.025}\sqrt{\frac{\hat{p}_V(1-\hat{p}_V)}{n_V} + \frac{\hat{p}_N(1-\hat{p}_N)}{n_N}}$$
$$= (0.1765 - 0.3778) \pm (1.96)\sqrt{\frac{(0.1765)(0.8235)}{51} + \frac{(0.3778)(0.6222)}{45}}$$
$$= -0.2013 \pm 0.1761 = [-0.3774 , -0.0252]$$

Since all the possible values in this interval are less than zero, we have high confidence that p_V is less than p_N, i.e. the vaccination is effective.