1. Scientific and statistical abstraction

   (a) **(5)** All plants received the same amount of water at the same time each day.

   (b) **(4)** No; variety is an inherent trait of the plant that is not randomly assigned.

   (c) **(8)** Let X_i be a random variable associated with the survival indicator for plant i from variety A, where $i = 1, 2$. The sample space for X_i is $\{0, 1\}$. Let Y_i be a random variable associated with the survival indicator for plant i from variety B, where $i = 1, 2, 3$. The sample space for Y_i is $\{0, 1\}$.

2. **(6)** The data are binary

3. Inference for the variety B probability

   (a) **(5)** $2\log(p_B) + \log(1 - p_B)$

   (b) **(5)** $U(p_B) = 2/p_B - 1/(1 - p_B)$

   (c) **(5)** The MLE $\hat{p}_B$ satisfies $U(\hat{p}_B) = 0$. Solving for $\hat{p}_B$ gives $\hat{p}_B = 2/3$.

   (d) **(4)** A general form for the score equation is

   $$U(p_B) = \frac{n\bar{y}}{p_B} - \frac{n(1-\bar{y})}{1-p_B}.$$

   Differentiating $U(p_B)$ with respect to p_B and taking the negative sign gives as the observed information,

   $$I_3(p_B) = 3\bar{y}/p_B^2 + 3(1-\bar{y})/(1-p_B)^2,$$

   where $\bar{Y} = 3^{-1}(Y_1 + Y_2 + Y_3)$. Taking the expectation gives as the expected information

   $$\bar{I}_3(p_B) = E[I_3(p_B)] = 3\left(\frac{1}{p_B} + \frac{1}{1-p_B}\right) = \frac{3}{p_B(1-p_B)}.$$

   (e) **(8)** The estimated variance of $\hat{p}_B$ is

   $$\hat{V}\{\hat{p}_B\} = \frac{1}{3}(2/3)(1/3) = 0.074,$$

   and the standard error is $SE(\hat{p}_B) = 0.272$. A 95% Wald interval is $2/3 \pm 1.96(0.272)$. This is $[0.133, 1]$.

(f) **(6)** The estimate is

$$\hat{\theta} = \log(2/3/(1-2/3)) = 0.693.$$

We estimate the standard deviation based on the asymptotic normal distribution of the estimator of the proportion. The derivative of θ with respect to p_B is

$$\theta' = \frac{1}{p_B} + \frac{1}{1-p_B}.$$

An estimate of θ' is $\hat{\theta}' = 3/2 + 3 = 4.5$. Then, the standard error of $\hat{\theta} = 4.5(0.272) = 1.224$.

(g) **(7)** The canonical parametrization is

$$\exp[y\theta - B(\theta)],$$

where the canonical parameter is $\theta = \log(p_B/(1-p_B))$, and $B(\theta) = \log(1+exp(\theta))$. The sufficient statistic is Y.

(h) **(4)** Then, the mean of Y is $B'(\theta) = \exp(\theta)(1+\exp(\theta))^{-1}$.

(i) **(5)** Yes

(j) **(5)** Yes

(k) **(5)** No

4. **(12)** Under the full model, the MLE of p_A is $\hat{p}_A = 1$. The likelihood under the full model is

$$L(\hat{p}_A, \hat{p}_B) = 1^2(0^0)(2/3)^2(1/3)^1 = 0.148.$$

The MLE of $p = p_A = p_B$ under the reduced model is $\hat{p} = 4/5$. The likelihood under the reduced model is

$$L(\hat{p}) = (4/5)^4(1/5)^1 = 0.08192.$$

The log likelihood ratio statistic is

$$T_n = -2[\log(0.08192) - \log(0.148)] = 1.183.$$

A p-value is then $2P(Z < -\sqrt{1.183})$, where Z has a standard normal distribution. This is approximately $2P(Z < -1.1) = 0.2714$. We fail to reject the null hypothesis and conclude that the common probability appears to be appropriate.

5. **(6)** The study lacks replication.