1. Follow the steps of slide 8 of set 20:

   1) Find $n - rank(\boldsymbol{X}) = 3 - 2 = 1$ linearly independent vector $\boldsymbol{a}$ such that $\boldsymbol{a}'\boldsymbol{X} = \boldsymbol{0}'$. From the model we have $\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$, so one of the choices can be $\boldsymbol{a}' = (1, -1, 0)$.

   2) Find the MLE of σ^2 using $w \equiv \boldsymbol{a}'\boldsymbol{y} = y_1 - y_2$ as data.

$$w = \boldsymbol{a}'\boldsymbol{y} = \boldsymbol{a}'(\boldsymbol{X\beta} + \boldsymbol{\epsilon}) = \boldsymbol{a}'\boldsymbol{X\beta} + \boldsymbol{a}'\boldsymbol{\epsilon} = 0 + \boldsymbol{a}'\boldsymbol{\epsilon} = \boldsymbol{a}'\boldsymbol{\epsilon}$$

Thus $w \sim N(0, \boldsymbol{a}'\boldsymbol{\Sigma a})$ where

$$\begin{aligned}\boldsymbol{a}'\boldsymbol{\Sigma a} &= \begin{pmatrix}1, & -1, & 0\end{pmatrix} \begin{pmatrix} \sigma^2 & \sigma^2/2 & 0 \\ \sigma^2/2 & \sigma^2 & \sigma^2/2 \\ 0 & \sigma^2/2 & \sigma^2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix}\sigma^2/2 & -\sigma^2/2 & -\sigma^2/2\end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\ &= \sigma^2\end{aligned}$$

So $w \sim N(0, \sigma^2)$ and the log likelihood function is

$$l(\sigma^2|w) = -\frac{1}{2}\log(\sigma^2) - \frac{w^2}{2\sigma^2} - \frac{1}{2}\log(2\pi)$$

The score equation is

$$\frac{\partial l}{\partial \sigma^2} = \frac{1}{\sigma^2} - \frac{w^2}{\sigma^4} = 0 \implies \hat{\sigma}^2 = w^2$$

Therefore the REML estimator of σ^2 in this case is $w^2 = (y_1 - y_2)^2$.

2. (a) The likelihood function is

$$\begin{aligned}L(\lambda|\boldsymbol{y}) &= \prod_{i=1}^{n} \frac{\lambda^{y_i}\exp(-\lambda)}{y_i!} \\ &= \frac{\lambda^{\sum_{i=1}^n y_i}\exp(-n\lambda)}{\prod_{i=1}^n y_i!}\end{aligned}$$

   (b) The score equation is

$$\begin{aligned}&\frac{d\log L(\lambda|\boldsymbol{y})}{d\lambda} = \frac{d\left[\sum_{i=1}^n y_i \log\lambda - n\lambda - \log(\prod_{i=1}^n y_i!)\right]}{d\lambda} = 0 \\ &\implies \frac{\sum_{i=1}^n y_i}{\lambda} - n = 0\end{aligned}$$

   (c) The solution of the score equation is

$$\hat{\lambda} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}.$$

(d) Verify that the solution of the score equation is an MLE.

Because $\dfrac{d\log L(\lambda|\boldsymbol{y})}{d\lambda} = 0$ has $\bar{y}.$ as its only solution, and $\forall\lambda > 0$

$$\frac{d^2 \log L(\lambda|\boldsymbol{y})}{d\lambda^2} = -\frac{\sum_{i=1}^n y_i}{\lambda^2} < 0$$

So $\hat{\lambda} = \bar{y}.$ is the maximum point of the likelihood function, *i.e.* an MLE of λ.

(e) The Fisher information matrix is

$$\begin{aligned}\boldsymbol{I}(\lambda) &= -E\left[\frac{d^2 \log L(\lambda|\boldsymbol{y})}{d\lambda^2}\right] \\ &= E\left[\frac{\sum_{i=1}^n y_i}{\lambda^2}\right] \\ &= \frac{n}{\lambda} && \text{because } E(y_i) = \lambda\end{aligned}$$

(f) The inverse of the Fisher information matrix is

$$\boldsymbol{I}^{-1}(\lambda) = \frac{\lambda}{n}$$

(g) Verify that the inverse of the Fisher information gives the exact variance of the MLE.

For $i = 1, \cdots, n$, $y_i \overset{iid}{\sim}$ Poisson(λ), and $Var(y_i) = \lambda$.

$$Var(\hat{\lambda}) = Var(\bar{y}.) = \frac{Var(y_i)}{n} = \frac{\lambda}{n} = \boldsymbol{I}^{-1}(\lambda)$$

(h) Provide an expression for an estimator of the variance of the MLE.

Plug the MLE of λ in the inverse Fisher information matrix, we have

$$\hat{Var}(\hat{\lambda}) = \hat{\boldsymbol{I}}^{-1}(\lambda) = \frac{\bar{y}.}{n}$$

3. By the results of problem 2,

In the full model, the MLEs of λ_1, λ_2 are $\hat{\lambda}_1 = \bar{y}_1.$ and $\hat{\lambda}_2 = \bar{y}_2.$, the maximized log likelihood is

$$l(\hat{\lambda}_1, \hat{\lambda}_2|\boldsymbol{y}) = \sum_{i=1}^{2}\left(\sum_{j=1}^{7} y_{ij}\log\hat{\lambda}_i - 7\hat{\lambda}_i - \log(\prod_{j=1}^{7} y_{ij}!)\right) = -37.10781$$

In the reduced model where $\lambda' \equiv \lambda_1 = \lambda_2$, the MLE of λ' is $\hat{\lambda}' = \bar{y}..$, the maximized log likelihood is

$$l(\hat{\lambda}'|\boldsymbol{y}) = \sum_{i=1}^{2}\sum_{j=1}^{7} y_{ij}\log\hat{\lambda}' - 14\hat{\lambda}' - \sum_{i=1}^{2}\sum_{j=1}^{7}\log(y_{ij}!) = -40.30237$$

(a) Compute AIC for the full model

$$AIC_{full} = -2l(\hat{\lambda}_1, \hat{\lambda}_2|\boldsymbol{y}) + 2k = 78.21563$$

(b) Compute BIC for the full model

$$BIC_{full} = -2l(\hat{\lambda}_1, \hat{\lambda}_2|\boldsymbol{y}) + k\log(n) = 79.49374$$

(c) Compute AIC for a simplified model in which $\lambda_1 = \lambda_2$.

$$AIC_{reduced} = -2l(\hat{\lambda}'|\boldsymbol{y}) + 2k = 82.60474$$

(d) Compute BIC for a simplified model in which $\lambda_1 = \lambda_2$.

$$BIC_{reduced} = -2l(\hat{\lambda}'|\boldsymbol{y}) + k\log(n) = 83.2438$$

(e) Which of the two models is preferred according to AIC?
The full model is preferred because it has the smaller AIC.

(f) Which of the two models is preferred according to BIC?
The full model is preferred because it has the smaller BIC.

(g) Compute the likelihood ratio test statistic $2\log\Lambda$ for testing $H_0 : \lambda_1 = \lambda_2$.

$$2\log\Lambda = 2\left(l(\hat{\lambda}'|\boldsymbol{y}) - l(\hat{\lambda}_1, \hat{\lambda}_2|\boldsymbol{y})\right) = 6.3891$$

(h) Find the p-value corresponding to the likelihood ratio statistic in part (g)
Under H_0, $2\log\Lambda \sim \chi_1$, $p-$value $= 0.01148$.

(i) Compute a Wald statistic for testing $H_0 : \lambda_1 = \lambda_2 \iff \lambda_1 - \lambda_2 = 0$.
By slide 12 of set 22 and the results of problem 2, we have

$$\begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{pmatrix} \dot{\sim} N\left(\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}, \hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\lambda}})\right)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$

$$\hat{\boldsymbol{I}}(\hat{\boldsymbol{\lambda}}) = \frac{-\partial^2 l(\boldsymbol{\lambda}|\boldsymbol{y})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\Big|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}} = \begin{bmatrix} \frac{7}{\hat{\lambda}_1} & 0 \\ 0 & \frac{7}{\hat{\lambda}_2} \end{bmatrix} = \begin{bmatrix} \frac{7}{\bar{y}_{1\cdot}} & 0 \\ 0 & \frac{7}{\bar{y}_{2\cdot}} \end{bmatrix}$$

Therefore $\bar{y}_{1\cdot} - \bar{y}_{2\cdot} \dot{\sim} N\left(\lambda_1 - \lambda_2, \frac{\bar{y}_{1\cdot}}{7} + \frac{\bar{y}_{2\cdot}}{7}\right)$.
The Wald statistic for testing $H_0 : \lambda_1 - \lambda_2 = 0$ is $\dfrac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{\sqrt{\frac{\bar{y}_{1\cdot}+\bar{y}_{2\cdot}}{7}}} = -2.5202$.

(j) Find the p-value corresponding to the Wald statistic in part (i).
Under H_0, the Wald statistic $\dot{\sim} N(0,1)$, $p-$value $= 0.01173$.

R code used in problem 3:

```
> geno1 <- c(14,9,10,5,18,9,9)
> geno2 <- c(17,10,17,18,13,17,16)
> y = c(geno1,geno2)
> type <- as.factor(rep(c(1,2),each=7))
> dat <- data.frame(y,type)
> o = glm(y ~ type,family=poisson,data=dat) # the full model
> logLik(o)
'log Lik.' -37.10781 (df=2)
> AIC(o)
[1] 78.21563
> BIC(o)
[1] 79.49374
> oreduce = glm(y ~ 1,family=poisson,data=dat) # the reduced model
> logLik(oreduce)
'log Lik.' -40.30237 (df=1)
> AIC(oreduce)
[1] 82.60474
> BIC(oreduce)
[1] 83.2438
> anova(oreduce,o,test="Chisq") # Likelihood ratio test
Analysis of Deviance Table

Model 1: y ~ 1
Model 2: y ~ type
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        13     19.606
2        12     13.217  1   6.3891  0.01148 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> est=mean(geno1)-mean(geno2)    # the Wald test
> var=(mean(geno1)+mean(geno2))/7
> est/sqrt(var)
[1] -2.520248
> pnorm(est/sqrt(var))*2
[1] 0.01172723
```

4. (a) The likelihood function is

$$\begin{aligned} L(\pi|\boldsymbol{y}) &= \prod_{i=1}^{n} \pi^{y_i}(1-\pi)^{1-y_i} I(0 \leqslant \pi \leqslant 1) \\ &= \pi^{\sum_{i=1}^{n} y_i}(1-\pi)^{n-\sum_{i=1}^{n} y_i} I(0 \leqslant \pi \leqslant 1) \end{aligned}$$

(b) For $\pi \in [0,1]$ the score equation is

$$\frac{d \log L(\pi|\boldsymbol{y})}{d\pi} = \frac{d\left(\sum_{i=1}^n y_i \log \pi (n - \sum_{i=1}^n y_i) \log(1-\pi)\right)}{d\pi} = 0$$
$$\implies \frac{\sum_{i=1}^n y_i}{\pi} - \frac{n - \sum_{i=1}^n y_i}{1-\pi} = 0$$

(c) The solution of the score equation is

$$\hat{\pi} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}.$$

(d) Verify that the solution of the score equation is an MLE.
Because $\frac{d \log L(\pi|\boldsymbol{y})}{d\pi} = 0$ has $\bar{y}.$ as its only solution, and $\forall \pi \in [0,1]$

$$\frac{d^2 \log L(\pi|\boldsymbol{y})}{d\pi^2} = -\frac{\sum_{i=1}^n y_i}{\pi^2} - \frac{n - \sum_{i=1}^n y_i}{(1-\pi)^2} < 0$$

So $\hat{\pi} = \bar{y}.$ is the maximum point of the likelihood function, *i.e.* an MLE of π.

(e) The Fisher information matrix is

$$\begin{aligned}
\boldsymbol{I}(\pi) &= -E\left[\frac{d^2 \log L(\pi|\boldsymbol{y})}{d\pi^2}\right] \\
&= E\left[\frac{\sum_{i=1}^n y_i}{\pi^2} + \frac{n - \sum_{i=1}^n y_i}{(1-\pi)^2}\right] \\
&= \frac{n}{\pi} + \frac{n}{1-\pi} \qquad \text{because } E(y_i) = \pi \\
&= \frac{n}{\pi(1-\pi)}
\end{aligned}$$

(f) The inverse of the Fisher information matrix is

$$\boldsymbol{I}^{-1}(\pi) = \frac{\pi(1-\pi)}{n}$$

(g) Verify that the inverse of the Fisher information gives the exact variance of the MLE.
For $i = 1, \cdots, n$, $y_i \overset{iid}{\sim}$ Bernoulli(π), and $Var(y_i) = \pi(1-\pi)$.

$$Var(\hat{\pi}) = Var(\bar{y}.) = \frac{Var(y_i)}{n} = \frac{\pi(1-\pi)}{n} = \boldsymbol{I}^{-1}(\pi)$$

(h) Provide an expression for an estimator of the variance of the MLE.
Plug the MLE of π in the inverse Fisher information matrix, we have

$$\hat{Var}(\hat{\pi}) = \hat{\boldsymbol{I}}^{-1}(\pi) = \frac{\bar{y}.(1-\bar{y}.)}{n}$$

(i) By slide 14 of set 22, an approximate 95% confidence interval for the proportion of died plants of this type is

$$\hat{\pi} \pm z_{0.975}\sqrt{\hat{Var}(\hat{\pi})} = 0.17 \pm 1.96\sqrt{\frac{0.17(1-0.17)}{100}} = (0.0964, 0.2436)$$

5. Let i denote the treatment group ($i = 1, 2$) and j denote the subject within the treatment group ($j = 1, \ldots, 350$). Assume $y_{ij} \overset{\text{ind}}{\sim} \text{Ber}(\pi_i)$. Recall that for $y_1, \ldots, y_n \overset{\text{iid}}{\sim} \text{Ber}(\pi)$,

$$\frac{\widehat{\pi}_{\text{mle}} - \pi}{\sqrt{\widehat{\text{Var}}(\widehat{\pi}_{\text{mle}})}} \xrightarrow{\text{d}} \mathcal{N}(0, 1) \text{ as } n \to \infty,$$

where $\widehat{\pi}_{\text{mle}} = \bar{y}_{\boldsymbol{\cdot}}$ and $\widehat{\text{Var}}(\widehat{\pi}_{\text{mle}}) = \frac{\bar{y}_{\boldsymbol{\cdot}}(1-\bar{y}_{\boldsymbol{\cdot}})}{n}$.

Here, we have $y_{1,1}, \ldots, y_{1,350} \overset{\text{iid}}{\sim} \text{Ber}(\pi_1)$ independent of $y_{2,1}, \ldots, y_{2,350} \overset{\text{iid}}{\sim} \text{Ber}(\pi_2)$, so that

$$\widehat{\pi}_1 = \bar{y}_{1\boldsymbol{\cdot}} = 172/350, \qquad \widehat{\text{Var}}(\widehat{\pi}_1) = \frac{\bar{y}_{1\boldsymbol{\cdot}}(1 - \bar{y}_{1\boldsymbol{\cdot}})}{n} = \frac{172/350(1 - 172/350)}{350},$$
$$\widehat{\pi}_2 = \bar{y}_{2\boldsymbol{\cdot}} = 137/350, \qquad \widehat{\text{Var}}(\widehat{\pi}_2) = \frac{\bar{y}_{2\boldsymbol{\cdot}}(1 - \bar{y}_{2\boldsymbol{\cdot}})}{n} = \frac{137/350(1 - 137/350)}{350}.$$

An approximate 95% confidence interval for $\pi_1 - \pi_2$ is then

$$\begin{aligned}
&\widehat{\pi}_1 - \widehat{\pi}_2 \pm z_{0.975}\sqrt{\widehat{\text{Var}}(\widehat{\pi}_1 - \widehat{\pi}_2)} \\
&\quad = \widehat{\pi}_1 - \widehat{\pi}_2 \pm z_{0.975}\sqrt{\widehat{\text{Var}}(\widehat{\pi}_1) + \widehat{\text{Var}}(\widehat{\pi}_2)} \quad \text{(by independence)} \\
&\quad = 172/350 - 137/350 \pm 1.96\sqrt{\frac{172/350(1 - 172/350)}{350} + \frac{137/350(1 - 137/350)}{350}} \\
&\quad = (0.027, 0.173).
\end{aligned}$$

Since this confidence interval is entirely above zero, there is evidence that treatment 1 is more effective than treatment 2 at enabling people to quit smoking for at least four weeks.