STAT 510 Homework 10 Solutions Spring 2018

1. (a) Follow the steps of slide 8 of set 20:

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

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

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

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

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

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

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

The score equation is

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

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

(b) 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.

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

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

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

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

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

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

2. (a) If we use the parametrization $\boldsymbol{\beta} = (\mu_1, \cdots, \mu_{100})'$ where $\mu_i = \mu + g_i$, $i = 1, ..., 100$, the model matrix is

$$\boldsymbol{X} = \begin{bmatrix} \underset{n_1\times 1}{\mathbf{1}} & & & & \\ & \underset{n_2\times 1}{\mathbf{1}} & & & \\ & & \ddots & & \\ & & & \underset{n_{99}\times 1}{\mathbf{1}} & \\ & & & & \underset{n_{100}\times 1}{\mathbf{1}} \end{bmatrix}$$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} n_1 & & & & \\ & n_2 & & & \\ & & \ddots & & \\ & & & n_{99} & \\ & & & & n_{100} \end{bmatrix} \text{ and } (\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} \frac{1}{n_1} & & & & \\ & \frac{1}{n_2} & & & \\ & & \ddots & & \\ & & & \frac{1}{n_{99}} & \\ & & & & \frac{1}{n_{100}} \end{bmatrix}$$

Thus, $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\bar{y}_{1\cdot}, ..., \bar{y}_{100\cdot})'$ and $\hat{\mu}_i = \widehat{\mu + g_i} = \bar{y}_{i\cdot}$ for $i = 1, ..., 100$. The R code below fits the cell means model to these data and provides estimates of $\mu_i = \mu + g_i$ using the parameterization $\boldsymbol{\beta} = (\mu_1, \cdots, \mu_{100})'$:

```
> dat=read.table("https://dnett.github.io/S510/hw10GenotypeYield.txt",
+                 header = T, col.names=c("genotype","yield"),
+                 colClasses = c("factor","numeric"))
> dat$genotype=factor(dat$genotype, levels = 1:100)
> ols.f=lm(yield˜0+genotype,data=dat)
> ols.f

Call:
lm(formula = yield ˜ 0 + genotype, data = dat)

Coefficients:
 genotype1    genotype2    genotype3    genotype4    genotype5    genotype6
     194.9        184.2        191.4        198.6        194.2        197.7

                                  ⋮

 genotype97   genotype98   genotype99  genotype100
      200.3        188.8        192.9        191.8
```

(b) Based on the output below, the REML estimates of σ_g^2 and σ_e^2 are $(2.6865)^2 = 7.2174$ and $(9.669)^2 = 93.4899$ respectively. The code and output are shown below:

```
> library(nlme)
> set.seed(1234)
> o=lme(yield~1,random=~1|genotype,data=dat)
> o
Random effects:
 Formula: ~1 | genotype
        (Intercept) Residual
StdDev:    2.686537 9.669021

Number of Observations: 304
Number of Groups: 100
```

(c) Note that $\boldsymbol{X} = \mathbf{1}, \boldsymbol{\beta} = \mu$,

$$\mathbf{Z} = \begin{bmatrix} \underset{n_1\times 1}{\mathbf{1}} & & & & \\ & \underset{n_2\times 1}{\mathbf{1}} & & & \\ & & \ddots & & \\ & & & \underset{n_{99}\times 1}{\mathbf{1}} & \\ & & & & \underset{n_{100}\times 1}{\mathbf{1}} \end{bmatrix}$$

, $\boldsymbol{G} = \sigma_g^2\boldsymbol{I}, \boldsymbol{R} = \sigma_e^2\boldsymbol{I}$. The BLUP for $\boldsymbol{g} = (g_1, g_2, ..., g_{100})'$ is

$$\hat{\boldsymbol{g}} = \boldsymbol{G}\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}})$$

where $\boldsymbol{\Sigma} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + \boldsymbol{R}$. Then, the BLUP for $\mu + g_i$ is

$$\frac{n_i\sigma_g^2}{\sigma_e^2 + n_i\sigma_g^2}\left(\bar{y}_{i\cdot} - \hat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}\right) = \frac{n_i\sigma_g^2}{\sigma_e^2 + n_i\sigma_g^2}\bar{y}_{i\cdot} + \frac{\sigma_e^2}{\sigma_e^2 + n_i\sigma_g^2}\hat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}$$

where

$$\hat{\boldsymbol{\beta}}_{\Sigma} = (\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} = \frac{\sum_{i=1}^{100}\frac{n_i\bar{y}_{i\cdot}}{\sigma_e^2+n_i\sigma_g^2}}{\sum_{i=1}^{100}\frac{n_i}{\sigma_e^2+n_i\sigma_g^2}}.$$

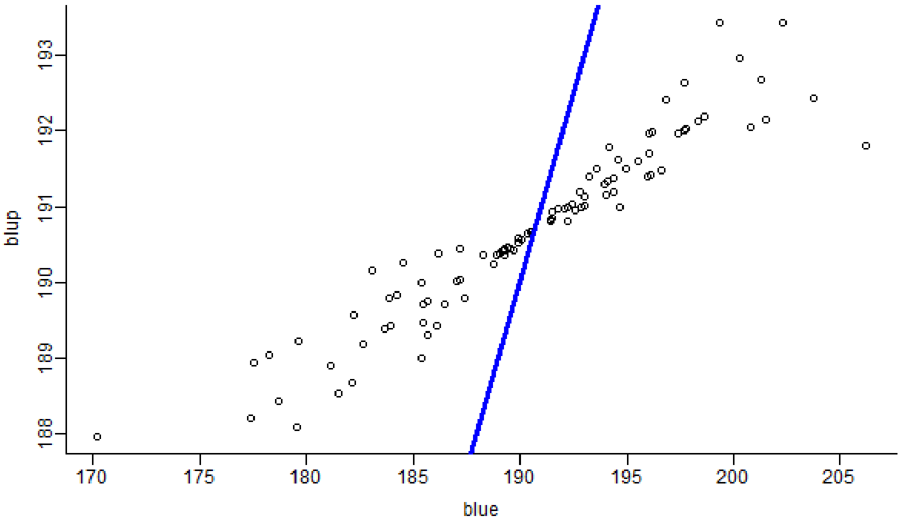
The following output is for the empirical BLUP:

```
> b=fixef(o)
> u=ranef(o)
> blup=as.matrix(b+u)
> blup
    (Intercept)
1      191.4947
2      189.8359
3      190.8365
            ⋮
```

```
97      192.9638
98      190.2445
99      190.9928
100     190.9641
```

(d) The plot of the eBLUPSs (vertical axis) vs. the BLUEs from part (a) (horizontal axis) is produced by the R code that follows:

```
> blue=as.vector(ols.f$coefficients)
> plot(blue,blup)
> abline(a=0,b=1,col=4,lwd=3)
```



In part (c), the BLUP of $\mu + g_i$ is a convex combination of the sample mean for the i^{th} genotype and the weighted average of sample means for all genotypes. In the middle part of the plot, the BLUE from part (a) and BLUP are similar because the BLUEs from part (a) are similar to $\hat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}$ in part (c). However, for large and small BLUEs from part (c), the BLUPs move toward $\hat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}$ in part (c) due to the effects of weights.

(e) According to the BLUEs from part (a), the top five highest yielding genotypes are as follows:

```
> blue.ord = order(blue,decreasing = T)
> top5 = blue.ord[1:5]
> print(data.frame(Top5=top5,Blue=blue[top5]))
  Top5    Blue
1   48 206.200
2   32 203.750
3    9 202.325
4   66 201.500
5   70 201.300
```

(f) According to the eBLUPs, the top five highest yielding genotypes are as follows:

```
> blup.ord = order(blup,decreasing = T)
> top5 = blup.ord[1:5]
> print(data.frame(Top5=top5,Blup=blup[top5]))
  Top5     Blup
1    9 193.4415
2   21 193.4319
3   97 192.9638
4   70 192.6920
5    6 192.6427
```

(g) Note that the BLUE of $\mu + g_i$ from part (a) is simply the sample mean for the i^{th} genotype while the BLUP of $\mu + g_i$ is a convex combination of the sample mean for the i^{th} genotype and the weighted average of every sample means for all genotypes. In the BLUPs, the weights for the sample mean (the BLUE from part (a)) and the weighted average of sample means (the BLUE from part (c)) depend on n_i, σ_e^2 and σ_g^2. Thus, even if a BLUE from part (a) is large, the corresponding BLUP might be smaller due to the effects of weights. So, the top-yielding genotype according to the BLUEs from part (a) is not so highly rated according to the BLUPs. In particular, note that $n_{48} = 1$, so that $\bar{y}_{48} = 206.2$ might not be a very reliable estimate of $\mu + g_{48}$. The small sample size for genotype 48 results in a large weight on $\hat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}$ when computing the eBLUP for genotype 48.

3. In the full model, the MLEs of θ_1, θ_2 are $\hat{\theta}_1 = \bar{y}_{1\cdot}$ and $\hat{\theta}_2 = \bar{y}_{2\cdot}$, the maximized log likelihood is

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

In the reduced model where $\theta^* \equiv \theta_1 = \theta_2$, the MLE of θ^* is $\hat{\theta}^* = \bar{y}_{\cdot\cdot}$, the maximized log likelihood is

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

(a) Compute AIC for the full model

$$AIC_{full} = -2l(\hat{\theta}_1, \hat{\theta}_2|\boldsymbol{y}) + 2k = 2(37.10781) + 2(2) = 78.21563$$

(b) Compute BIC for the full model

$$BIC_{full} = -2l(\hat{\theta}_1, \hat{\theta}_2|\boldsymbol{y}) + k\log(n) = 2(37.10781) + 2\log(14) = 79.49374$$

(c) Compute AIC for a simplified model in which $\theta_1 = \theta_2$.

$$AIC_{reduced} = -2l(\hat{\theta}^*|\boldsymbol{y}) + 2k = 2(40.30237) + 2(1) = 82.60474$$

(d) Compute BIC for a simplified model in which $\theta_1 = \theta_2$.

$$BIC_{reduced} = -2l(\hat{\theta}^*|\boldsymbol{y}) + k\log(n) = 2(40.30237) + (1)\log(14) = 83.2438$$

(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 $H_0: \theta_1 = \theta_2$.

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

(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 $H_0: \theta_1 = \theta_2 \iff \theta_1 - \theta_2 = 0$.
By slide 10 of set 22, we have

$$\begin{pmatrix} \bar{y}_{1\cdot} \\ \bar{y}_{2\cdot} \end{pmatrix} \dot{\sim} N\left(\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\theta}})\right)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)'$

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

Therefore $\bar{y}_{1\cdot} - \bar{y}_{2\cdot} \dot{\sim} N\left(\theta_1 - \theta_2, \frac{\bar{y}_{1\cdot}}{7} + \frac{\bar{y}_{2\cdot}}{7}\right)$.
The Wald statistic for testing $H_0: \theta_1 - \theta_2 = 0$ 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

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

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

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

(c) The solution of the score equation is

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

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

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

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

(e) The Fisher information matrix is

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

(f) The inverse of the Fisher information matrix is

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

(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(θ), and $Var(y_i) = \theta(1-\theta)$.

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

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

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

(i) By slide 17 of set 22, an approximate 95% confidence interval for the proportion of died plants of this type is

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

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

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

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

Here, we have $y_{1,1}, \ldots, y_{1,350} \overset{\text{iid}}{\sim} \text{Ber}(\theta_1)$ independent of $y_{2,1}, \ldots, y_{2,350} \overset{\text{iid}}{\sim} \text{Ber}(\theta_2)$, so that

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

An approximate 95% confidence interval for $\theta_1 - \theta_2$ is then

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

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.