STK3100 Exercises, Week 6

Vinnie Ko, Jonas Moss

October 11, 2018

Exercise 5.10

i)
We have a logistic regression model with $\text{logit}(\pi) = \eta = \beta_0 + \beta_1 x$.

We let x_0 be given by $\dfrac{\exp(\beta_0 + \beta_1 x_0)}{1 + \exp(\beta_0 + \beta_1 x_0)} = \pi_0$ for a given probability π_0.

We will find a confidence interval for x_0 by inverting a α-level test for $H_0 : \pi = \pi_0$ or equivalently $H_0 : \eta = \text{logit}(\pi_0)$. A Wald test rejects H_0 when

$$\left| \frac{\widehat{\eta} - \text{logit}(\pi_0)}{\sqrt{\text{Var}(\widehat{\eta})}} \right| \geq z_{1-\frac{\alpha}{2}}.$$

As $\widehat{\text{Var}}(\widehat{\eta}) \xrightarrow{p} \text{Var}(\widehat{\eta})$, we have asymptotically

$$\left| \frac{\widehat{\eta} - \text{logit}(\pi_0)}{\sqrt{\widehat{\text{Var}}(\widehat{\eta})}} \right| \geq z_{1-\frac{\alpha}{2}}$$

or equivalently

$$\left| \frac{\widehat{\beta}_0 + \widehat{\beta}_1 x - \text{logit}(\pi_0)}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_0) + x^2 \widehat{\text{Var}}(\widehat{\beta}_1) + 2x\,\widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1)}} \right| \geq z_{1-\frac{\alpha}{2}}.$$

We obtain a $100(1-\alpha)\%$ confidence interval for x_0 by inverting this test, so the interval is given by all x that satisfy the inequality

$$\left| \frac{\widehat{\beta}_0 + \widehat{\beta}_1 x - \text{logit}(\pi_0)}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_0) + x^2 \widehat{\text{Var}}(\widehat{\beta}_1) + 2x\,\widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1)}} \right| < z_{1-\frac{\alpha}{2}}$$

ii)
The likelihood ratio test for $H_0 : \eta = \text{logit}(\pi_0)$ rejects if

$$-2\{L(\beta_0^*, \beta_1^*) - L(\widehat{\beta}_0, \widehat{\beta}_1)\} \geq \chi_1^2(\alpha),$$

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the maximum likelihood estimators under the logistic model, and β_0^* and β_1^* are the maximum likelihood estimators under the restriction $\beta_0 + \beta_1 x = \text{logit}(\pi_0)$ [so $L(\beta_0^*, \beta_1^*)$ depends on x].

A confidence interval is then given as the set of x that satisfy

$$-2\{L(\beta_0^*, \beta_1^*) - L(\widehat{\beta}_0, \widehat{\beta}_1)\} < \chi_1^2(\alpha),$$

Exercise 5.11

By the given conditions, the pmf of Y_i is given by $f(y_i) = \binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}$ for $i = 1, 2$, where $\text{logit}(\pi_i) = \beta_0 + \beta_1 x_i$ and $x_1 = 0$, $x_2 = 1$. Then, the log-likelihood is

$$
\begin{aligned}
L(\boldsymbol{\beta}) &= \log f(y_1) + \log f(y_2) \\
&= y_1 \log \pi_1 + (n_1 - y_1)\log(1-\pi_1) + y_2 \log \pi_2 + (n_2 - y_2)\log(1-\pi_2) + \log\binom{n_1}{y_1} + \binom{n_2}{y_2} \\
&= y_1 \log\left(\frac{e^{\beta_0}}{e^{\beta_0}+1}\right) + (n_1 - y_1)\log\left(\frac{1}{e^{\beta_0}+1}\right) + y_2 \log\left(\frac{e^{\beta_0+\beta_1}}{e^{\beta_0+\beta_1}+1}\right) + (n_2 - y_2)\log\left(\frac{1}{e^{\beta_0+\beta_1}+1}\right) \\
&\quad + \log\binom{n_1}{y_1} + \binom{n_2}{y_2} \\
&= y_1\left(\beta_0 - \log\left(e^{\beta_0}+1\right)\right) - (n_1 - y_1)\log\left(e^{\beta_0}+1\right) \\
&\quad + y_2\left(\beta_0 + \beta_1 - \log\left(e^{\beta_0+\beta_1}+1\right)\right) - (n_2 - y_2)\log\left(e^{\beta_0+\beta_1}+1\right) + \log\binom{n_1}{y_1} + \binom{n_2}{y_2} \\
&= y_1\beta_0 - n_1\log\left(e^{\beta_0}+1\right) + y_2(\beta_0+\beta_1) - n_2\log\left(e^{\beta_0+\beta_1}+1\right) + \log\binom{n_1}{y_1} + \binom{n_2}{y_2}
\end{aligned}
$$

.

The likelihood equations are

$$
\begin{aligned}
\frac{\partial L}{\partial \beta_0} &= y_1 - n_1\frac{e^{\beta_0}}{e^{\beta_0}+1} + y_2 - n_2\frac{e^{\beta_0+\beta_1}}{e^{\beta_0+\beta_1}+1} = 0 \\
\frac{\partial L}{\partial \beta_1} &= y_2 - n_2\frac{e^{\beta_0+\beta_1}}{e^{\beta_0+\beta_1}+1} = 0.
\end{aligned}
$$

We solve the equations and obtain

$$
\begin{aligned}
\widehat{\beta}_0 &= \text{logit}\left(\frac{y_1}{n_1}\right) = \log\left(\frac{y_1}{n_1 - y_1}\right) \\
\widehat{\beta}_1 &= \text{logit}\left(\frac{y_2}{n_2}\right) - \text{logit}\left(\frac{y_1}{n_1}\right) = \log\left(\frac{\frac{y_2}{n_2-y_2}}{\frac{y_1}{n_1-y_1}}\right).
\end{aligned}
$$

So, $\widehat{\beta}_1$ is the sample log odds ratio.

Exercise 5.16

a)

If we treat the data as N binomial observations by letting $y_i = \sum_{j=1}^{n_i} y_{ij}$, the pmf's become

$$
f(y_i) = \binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}
$$

Then, the log-likelihood is

$$
L(\boldsymbol{\pi}) = \sum_{i=1}^{N}\left[\log\binom{n_i}{y_i} + y_i \log \pi_i + (n_i - y_i)\log(1-\pi_i)\right].
$$

The kernel of the log-likelihood (by dropping the parts that don't depend on π_i) is

$$L(\boldsymbol{\pi}) = \sum_{i=1}^{N} \left[y_i \log \pi_i + (n_i - y_i) \log(1 - \pi_i) \right].$$

If we treat data as $n = \sum_{i=1}^{N} n_i$ Bernoulli observations, the pmf's become

$$f(y_{i,j}) = \pi_i^{y_{i,j}} (1 - \pi_i)^{1 - y_{i,j}}.$$

Then, the log-likelihood is

$$\begin{aligned} L(\boldsymbol{\pi}) &= \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left[y_{i,j} \log \pi_i + (1 - y_{i,j}) \log(1 - \pi_i) \right] \\ &= \sum_{i=1}^{N} \left[y_i \log \pi_i + (n_i - y_i) \log(1 - \pi_i) \right] \end{aligned}$$

and this is already the kernel.

b)

For a saturated model there are as many parameters as observations ($n = p$). When we treat the data as N binomial observations, there are N parameters $\pi_1, \cdots, \pi_N$. When we treat the data as $n = \sum_{i=1}^{N} n_i$ Bernoulli observations, there are n parameters $\{\pi_{i,j}\}$. So, the kernel of (log-)likelihood is different. Consequently, the deviance, which contains the log-likelihood of saturated model is also different.

c)

When we take difference (i.e. subtract) of deviance between 2 unsaturated models, the log-likelihood of the saturated model will be canceled out and the difference depends only on the log-likelihood of unsaturated models. In a), we showed that these log-likelihoods of unsaturated models are not affected by how we form the data entry.

Exercise 5.17

a)

```
> # Create data
> n.i = 4
> data.1 = data.frame(
+   x = c(rep(0, n.i), rep(1, n.i), rep(2, n.i)),
+   n = 1,
+   y = c(0,0,0,1,0,0,1,1,1,1,1,1)
+ )
>
> data.2 = data.frame(
+   x = c(0,1,2),
+   n = 4,
+   y = c(1,2,4)
```

```
+ )
>
> show(data.1)
   x n y
1  0 1 0
2  0 1 0
3  0 1 0
4  0 1 1
5  1 1 0
6  1 1 0
7  1 1 1
8  1 1 1
9  2 1 1
10 2 1 1
11 2 1 1
12 2 1 1
> show(data.2)
  x n y
1 0 4 1
2 1 4 2
3 2 4 4
>
> # Fit M.0 with 2 different data forms.
> M.0.data.1 = glm(y ~ 1, family = binomial(link = "logit"), data = data.1)
> M.0.data.2 = glm(cbind(y, n-y) ~ 1, family = binomial(link = "logit"), data = data
    .2)
> summary(M.0.data.1)

Call:
glm(formula = y ~ 1, family = binomial(link = "logit"), data = data.1)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.323  -1.323   1.038   1.038   1.038

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.3365     0.5855   0.575    0.566

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16.301  on 11  degrees of freedom
Residual deviance: 16.301  on 11  degrees of freedom
AIC: 18.301

Number of Fisher Scoring iterations: 4

> summary(M.0.data.2)

Call:
glm(formula = cbind(y, n - y) ~ 1, family = binomial(link = "logit"),
    data = data.2)

Deviance Residuals:
      1        2        3
-1.3536  -0.3357   2.0765
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.3365     0.5855   0.575    0.566

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6.2568  on 2  degrees of freedom
Residual deviance: 6.2568  on 2  degrees of freedom
AIC: 11.945

Number of Fisher Scoring iterations: 4

>
> # Fit M.1 with 2 different data forms.
> M.1.data.1 = glm(y ~ x, family = binomial(link = "logit"), data = data.1)
> M.1.data.2 = glm(cbind(y, n-y) ~ x, family = binomial(link = "logit"), data = data
    .2)
> summary(M.1.data.1)

Call:
glm(formula = y ~ x, family = binomial(link = "logit"), data = data.1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4216  -0.6339   0.3752   0.5193   1.8459

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.503      1.181  -1.272   0.2033
x              2.060      1.130   1.823   0.0682 .
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16.301  on 11  degrees of freedom
Residual deviance: 11.028  on 10  degrees of freedom
AIC: 15.028

Number of Fisher Scoring iterations: 4

> summary(M.1.data.2)

Call:
glm(formula = cbind(y, n - y) ~ x, family = binomial(link = "logit"),
    data = data.2)

Deviance Residuals:
      1        2        3
 0.3377  -0.5543   0.7504

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.503      1.181  -1.272   0.2034
x              2.060      1.130   1.823   0.0683 .
```

```
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6.2568  on 2  degrees of freedom
Residual deviance: 0.9844  on 1  degrees of freedom
AIC: 8.6722

Number of Fisher Scoring iterations: 4

>
> deviance.table = as.data.frame(matrix(NA, nrow = 2, ncol = 2))
> rownames(deviance.table) = c("M.0", "M.1")
> colnames(deviance.table) = c("data.1", "data.2")
> deviance.table[1,1] = M.0.data.1$deviance
> deviance.table[1,2] = M.0.data.2$deviance
> deviance.table[2,1] = M.1.data.1$deviance
> deviance.table[2,2] = M.1.data.2$deviance
> show(deviance.table)
      data.1    data.2
M.0 16.30064 6.2567798
M.1 11.02826 0.9843993
>
> logLik.table = as.data.frame(matrix(NA, nrow = 2, ncol = 2))
> rownames(logLik.table) = c("M.0", "M.1")
> colnames(logLik.table) = c("data.1", "data.2")
> logLik.table[1,1] = logLik(M.0.data.1)
> logLik.table[1,2] = logLik(M.0.data.2)
> logLik.table[2,1] = logLik(M.1.data.1)
> logLik.table[2,2] = logLik(M.1.data.2)
> show(logLik.table)
       data.1    data.2
M.0 -8.150319 -4.972265
M.1 -5.514129 -2.336075
>
> kernel.logLik.table = as.data.frame(matrix(NA, nrow = 2, ncol = 2))
> rownames(kernel.logLik.table) = c("M.0", "M.1")
> colnames(kernel.logLik.table) = c("data.1", "data.2")
> kernel.logLik.table[1,1] = logLik(M.0.data.1)
> kernel.logLik.table[1,2] = logLik(M.0.data.2) - (lchoose(4, 1) + lchoose(4, 2) +
    lchoose(4, 4))
> kernel.logLik.table[2,1] = logLik(M.1.data.1)
> kernel.logLik.table[2,2] = logLik(M.1.data.2) - (lchoose(4, 1) + lchoose(4, 2) +
    lchoose(4, 4))
> show(kernel.logLik.table)
       data.1    data.2
M.0 -8.150319 -8.150319
M.1 -5.514129 -5.514129
```

As we already saw in exercise 5.16 b), the log-likelihood of the saturated model is affected by the number of parameters. Thus, the deviance, which contains the log-likelihood of the saturated model is also affected by the data form.

b)

```
> deviance.table["M.0","data.1"] - deviance.table["M.1","data.1"]
[1] 5.27238
> deviance.table["M.0","data.2"] - deviance.table["M.1","data.2"]
[1] 5.27238
```

As we already saw in exercise 5.16 c), when we take difference (i.e. subtract) of deviance between 2 unsaturated models, the log-likelihood of the saturated model (and the constant part of the log-likelihood) will be canceled out and the difference depends only on the kernel of the log-likelihood of unsaturated models. Therefore, the difference of deviance between 2 unsaturated models is not affected by how we form the data entry.

Additional Exercise 15

i)

We know that $\text{logit}(\pi) = \beta_0 + \beta_1 x$. So, $x = \dfrac{\text{logit}(\pi) - \beta_0}{\beta_1}$ and $LD50 = \dfrac{\text{logit}\left(\frac{1}{2}\right) - \beta_0}{\beta_1} = -\dfrac{\beta_0}{\beta_1}$. The estimated version is $\widehat{LD50} = -\dfrac{\widehat{\beta}_0}{\widehat{\beta}_1} = 1.7716$

```
> # Read data.
> Beetle = read.table("http://www.stat.ufl.edu/~aa/glm/data/Beetles2.dat", header = T)
> head(Beetle)
logdose  n dead
1   1.691 59    6
2   1.724 60   13
3   1.755 62   18
4   1.784 56   28
5   1.811 63   52
6   1.837 59   53
>
> # Fit logistic regression
> Beetle.model.1 = glm(cbind(dead,n-dead) ~ logdose, family = binomial, data = Beetle)
> summary(Beetle.model.1)

Call:
glm(formula = cbind(dead, n - dead) ~ logdose, family = binomial,
data = Beetle)

Deviance Residuals:
Min       1Q   Median       3Q      Max
-1.5878  -0.4085   0.8442   1.2455   1.5860

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.740      5.182  -11.72   <2e-16 ***
logdose       34.286      2.913   11.77   <2e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.116  on 6  degrees of freedom
AIC: 41.314
```

```
Number of Fisher Scoring iterations: 4

>
> # Estimation of LD50
> LD50 = -as.numeric(Beetle.model.1$coef[1])/as.numeric(Beetle.model.1$coef[2])
> show(LD50)
[1] 1.771576
```

ii)
From exercise 5.10 of the book, we have that a 95% confidence interval for LD50 is given by all x that satisfy the inequality

$$\left| \frac{\widehat{\beta}_0 + \widehat{\beta}_1 x - \text{logit}(0.50)}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_0) + x^2\,\widehat{\text{Var}}(\widehat{\beta}_1) + 2x\,\widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1)}} \right| < 1.96$$

or equivalently

$$\left(\widehat{\beta}_0 + \widehat{\beta}_1 x\right)^2 < 1.96^2 \cdot \left(\widehat{\text{Var}}(\widehat{\beta}_0) + x^2\,\widehat{\text{Var}}(\widehat{\beta}_1) + 2x\,\widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1)\right)$$

$$\left(\widehat{\beta}_1^2 - 1.96^2 \cdot \widehat{\text{Var}}(\widehat{\beta}_1)\right) x^2 + 2\left(\widehat{\beta}_0\widehat{\beta}_1 - 1.96^2\widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1)\right) x + \widehat{\beta}_0^2 - 1.96^2 \cdot \widehat{\text{Var}}(\widehat{\beta}_0) < 0$$

Thus, the confidence interval is all x that satisfy the second degree inequality

$$ax^2 + bx + c < 0$$

where

$$\begin{aligned} a &= \widehat{\beta}_1^2 - 1.96^2 \cdot \widehat{\text{Var}}(\widehat{\beta}_1) \\ b &= 2\widehat{\beta}_0\widehat{\beta}_1 - 2 \cdot 1.96^2\widehat{\text{Cov}}(\widehat{\beta}_0, \widehat{\beta}_1) \\ c &= \widehat{\beta}_0^2 - 1.96^2 \cdot \widehat{\text{Var}}(\widehat{\beta}_0). \end{aligned}$$

By solving the inequality, we obtain the confidence interval

$$\left(\frac{-b - \sqrt{b^2 - 4ac}}{2a}, \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right)$$

```
> # 95% confidence interval of LD50
> alpha = 0.05
> z.value = qnorm(1 - alpha/2)
>
> beta.hat = as.numeric(Beetle.model.1$coeff)
> beta.hat.cov.mat = vcov(Beetle.model.1)
>
> a.val = beta.hat[2]^2 - (z.value^2)*beta.hat.cov.mat[2,2]
> b.val = 2*beta.hat[1]*beta.hat[2] - 2*(z.value^2)*beta.hat.cov.mat[1,2]
> c.val = beta.hat[1]^2 - (z.value^2)*beta.hat.cov.mat[1,1]
>
> LD50.CI95 = c(
+   (-b.val -sqrt(b.val^2 -4*a.val*c.val))/(2*a.val),
+   (-b.val +sqrt(b.val^2 -4*a.val*c.val))/(2*a.val)
+ )
> show(LD50.CI95)
[1] 1.763722 1.779054
```

Additional Exercise 16

a)

The probability mass function of the binomial with parameters n and π is $f(v;\pi,n)=\binom{n}{V}\pi^{v}(1-\pi)^{n-v}$, where $v\in\{0,1,\dots,n\}$.

The log likelihood is

$$l(\pi;v)=\log\binom{n}{V}+v\log\pi+(n-v)\log(1-\pi).$$

The score equals its derivative of the log likelihood with respect to π, namely

$$\begin{aligned}S(\pi;v) &= \frac{v}{\pi}-\frac{n-v}{1-\pi}.\\ &= \frac{v(1-\pi)-\pi(n-v)}{\pi(1-\pi)}\\ &= \frac{v-\pi n}{\pi(1-\pi)}\end{aligned}$$

Since the derivative of $\frac{v}{\pi}+\frac{n-v}{1-\pi}$ with respect to π is $-\frac{v}{\pi^2}-\frac{n-v}{(1-\pi)^2}$, the information matrix is

$$\begin{aligned}J(\pi) &= \frac{\pi n}{\pi^2}+\frac{n-\pi n}{(1-\pi)^2}\\ &= n\left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)\\ &= \frac{n}{\pi(1-\pi)}.\end{aligned}$$

b)

The Wald statistic is defined as

$$\begin{aligned}\frac{\widehat{\pi}-\pi}{\sqrt{J^{-1}(\widehat{\pi})}} &= \frac{\widehat{\pi}-\pi}{\sqrt{J^{-1}(\widehat{\pi})}}\\ &= \sqrt{n}\frac{\widehat{\pi}-\pi}{\sqrt{\widehat{\pi}(1-\widehat{\pi})}}.\end{aligned}$$

c)

The score statistic is

$$\begin{aligned}\frac{S(\pi)}{\sqrt{J(\pi)}} &= \frac{\frac{v-\pi n}{\pi(1-\pi)}}{\sqrt{\frac{n}{\pi(1-\pi)}}}\\ &= \sqrt{n}\frac{\widehat{\pi}-\pi}{\sqrt{\pi(1-\pi)}}\end{aligned}$$

d)

The likelihood statistic is

$$\begin{aligned}
-2\log\Delta &= -2 \quad (\log\binom{n}{V} + v\log\pi + (n-v)\log(1-\pi) - \\
& \qquad \log\binom{n}{V} + v\log\widehat{\pi} + (n-v)\log(1-\pi)), \\
&= \quad 2v\log\left(\frac{\widehat{\pi}}{\pi}\right) + 2(n-v)\log\left(\frac{1-\widehat{\pi}}{1-\pi}\right).
\end{aligned}$$

When n is large, $-2\log\Delta$ is approximately χ^2_1 by Wilks theorem.

e)

The solutions are as follows:

$$\begin{aligned}
\sqrt{n}\frac{\widehat{\pi}-\pi}{\widehat{\pi}(1-\widehat{\pi})} &= \sqrt{100}\frac{0.3-0.5}{\sqrt{0.3\cdot 0.7}} \\
&\approx -4.4 \\
\frac{S(\pi)}{\sqrt{J(\pi)}} &= \sqrt{100}\frac{0.3-0.5}{0.5} \\
&\approx -4 \\
-2\log\Delta &= 60\log\left(\frac{0.3}{0.5}\right) + 140\log\left(\frac{0.7}{0.5}\right) \\
&\approx 16.5
\end{aligned}$$

All tests agree to roughly the same degree that H_0 is false. Notice that $\sqrt{16.5} \approx 4.05$, and since the root of a χ^2_1 is the abosolute value of the normal, the likelihood ratio test agrees as well.

f)

The solutions are as follows:

$$\begin{aligned}
\sqrt{n}\frac{\widehat{\pi}-\pi}{\sqrt{\widehat{\pi}(1-\widehat{\pi})}} &= \sqrt{100}\frac{0.05-0.15}{\sqrt{0.05\cdot 0.15}} \\
&\approx -11.5 \\
\frac{S(\pi)}{\sqrt{J(\pi)}} &= \sqrt{100}\frac{0.05-0.15}{0.15} \\
&\approx -6.7 \\
-2\log\Delta &= 10\log\left(\frac{0.05}{0.15}\right) + 190\log\left(\frac{0.95}{0.75}\right) \\
&\approx 34
\end{aligned}$$

The likelihood ratio test is more conservative than the others.

g)

This is the case since $\sqrt{n}\frac{\widehat{\pi}-\pi}{\sqrt{\widehat{\pi}\left(1-\widehat{\pi}\right)}}$ is approximately normally distributed and ± 1.96 are the two-sided confidence limits for a standard normal variable.

h)

Recall that the score $\sqrt{n}\frac{\widehat{\pi}-\pi}{\sqrt{\pi\left(1-\pi\right)}}$ is asymptotically standard normal. This implies that $n\frac{\left(\widehat{\pi}-\pi\right)^2}{\pi\left(1-\pi\right)}$ is asymptotically χ_1^2. The $1-\alpha$ percentile of χ_1^2 is $-\Phi^{-1}\left(\alpha/2\right)=c_\alpha$. We find the confidence interval by inverting the hypothesis test that rejects when $n\frac{\left(\widehat{\pi}-\pi\right)^2}{\pi\left(1-\pi\right)}>c_\alpha$. This is done by solving the equation $n\frac{\left(\widehat{\pi}-\pi\right)^2}{\pi\left(1-\pi\right)}=c_\alpha^2$. This is equivalent to

$$n\left(\widehat{\pi}^2-2\widehat{\pi}\pi+\pi^2\right) \quad = \quad c_\alpha\pi-c_\alpha\pi^2.$$

Rearrange to obtain $\pi^2\left(n+c_\alpha^2\right)-\pi\left(2v+c_\alpha^2\right)+n\widehat{\pi}^2$. The solutions are

$$\frac{2v+c_\alpha^2\pm\sqrt{\left(2v+c_\alpha^2\right)^2-4\left(n+c_\alpha^2\right)n\widehat{\pi}^2}}{2\left(n+c_\alpha^2\right)}.$$

Simplify this to for instance

$$\frac{v+\frac{1}{2}c_\alpha^2}{n+c_\alpha^2}\pm\frac{\sqrt{\left(2v+c_\alpha^2\right)^2/4-\left(n+c_\alpha^2\right)\widehat{\pi}^2}}{n+c_\alpha^2}.$$

i)

```
> v = 30
> n = 100
> alpha = 0.05
> CI(n, v, alpha)

$wald
[1] 0.2101832 0.3898168

$score
[1] 0.2189489 0.3958485

$lrt
[1] 0.2160263 0.3940910
```

i)

```
> v = 5
> n = 100
> alpha = 0.05
> CI(n, v, alpha)

$wald
[1] 0.007283575 0.092716425
```

```
$score
[1] 0.02154368 0.11175047

$lrt
[1] 0.01823454 0.10438783
```