Generalized Linear Regression Homework 2

Brian Cervantes Alvarez

October 15, 2024

Problem 1

Consider the Donner Party example. Answer the following questions based on the R output on Page 10 of Lecture 5 and the following covariance matrix of the estimated coefficient vector

$$\begin{pmatrix} \text{Covariances} & \text{Intercept} & \text{SexFemale} & \text{Age} \\ \text{Intercept} & 1.92338373 & 0.63080437 & -0.04670471 \\ \text{SexFemale} & 0.63080437 & 0.57073339 & -0.00823197 \\ \text{Age} & -0.04670471 & -0.00823197 & 0.001390134 \end{pmatrix}$$

The **logistic regression model** is

$$\text{logit}(p) = \beta_0 + \beta_1 \times \text{SexFemale} + \beta_2 \times \text{Age}$$

with estimated coefficients

- $\beta_0 = 1.63312$
- $\beta_1 = 1.59729$
- $\beta_2 = -0.07820$

Part A

For a 50-year-old female

- SexFemale = 1
- Age = 50

First, determine **Linear Predictor** η

$$
\begin{aligned}
\eta &= \beta_0 + \beta_1 \times \text{SexFemale} + \beta_2 \times \text{Age} \\
&= 1.63312 + 1.59729 \times 1 + (-0.07820) \times 50 \\
&= 1.63312 + 1.59729 - 3.910 \\
&= -0.67959
\end{aligned}
$$

Second, calculate the **Survival Probability** p

$$
p = \frac{e^{\eta}}{1+e^{\eta}} = \frac{e^{-0.67959}}{1+e^{-0.67959}} \approx \frac{0.5068}{1+0.5068} \approx 0.3365
$$

Then we can find the **Variance & Standard Error** of η

$$
\begin{aligned}
\text{Var}(\eta) &= \text{Var}(\beta_0) + \text{Var}(\beta_1) \times (\text{SexFemale})^2 + \text{Var}(\beta_2) \times (\text{Age})^2 \\
&\quad + 2 \times \text{Cov}(\beta_0, \beta_1) \times \text{SexFemale} + 2 \times \text{Cov}(\beta_0, \beta_2) \times \text{Age} + 2 \times \text{Cov}(\beta_1, \beta_2) \times \text{SexFemale} \times \text{Age} \\
&= 1.92338373 + 0.57073339 \times 1^2 + 0.001390134 \times 50^2 \\
&= 1.73739286 \\
\text{SE}(\eta) &= \sqrt{\text{Var}(\eta)} = \sqrt{1.73739286} \approx 1.318
\end{aligned}
$$

Penultimate, the **95% Confidence Interval** for η. Using $z_{0.975} = 1.96$, we can find the lower and upper bounds as follows

$$
\begin{aligned}
\text{Lower limit} &= \eta - 1.96 \times \text{SE}(\eta) = -0.67959 - 1.96 \times 1.318 \approx -0.67959 - 2.583 \approx -3.263 \\
\text{Upper limit} &= \eta + 1.96 \times \text{SE}(\eta) = -0.67959 + 1.96 \times 1.318 \approx -0.67959 + 2.583 \approx 1.903
\end{aligned}
$$

Ultmately, we can find the 95% Confidence Interval for p

$$
\begin{aligned}
p_{Lower} &= \frac{e^{-3.263}}{1+e^{-3.263}} \approx \frac{0.0383}{1+0.0383} \approx 0.0369 \\
p_{Upper} &= \frac{e^{1.903}}{1+e^{1.903}} \approx \frac{6.703}{1+6.703} \approx 0.870 \\
\text{95\% Confidence Interval} &= (0.0369, 0.870)
\end{aligned}
$$

Part B

Let's first find the **Log-Odds** $\Delta\eta$. I did the following,

$$\begin{aligned}
\Delta\eta &= (\beta_0 + \beta_1 \times 0 + \beta_2 \times 40) - (\beta_0 + \beta_1 \times 1 + \beta_2 \times 35) \\
&= \beta_2 \times (40 - 35) - \beta_1 \\
&= (-0.07820) \times 5 - 1.59729 \\
&= -0.391 - 1.59729 \\
&= -1.98829
\end{aligned}$$

Next, we can determine the **Odds Ratio** by,

$$\text{OR} = e^{\Delta\eta} = e^{-1.98829} \approx 0.1369$$

Then, we can compute the **Variance** & **Standard Error** of $\Delta\eta$

$$\begin{aligned}
\text{Var}(\Delta\eta) &= \text{Var}(-\beta_1 + 5\beta_2) \\
&= \text{Var}(\beta_1) + 25 \times \text{Var}(\beta_2) - 2 \times 5 \times \text{Cov}(\beta_1, \beta_2) \\
&= 0.57073339 + 25 \times 0.001390134 - 10 \times (-0.00823197) \\
&= 0.57073339 + 0.03475335 + 0.0823197 \\
&= 0.68780644 \\
\text{SE}(\Delta\eta) &= \sqrt{\text{Var}(\Delta\eta)} = \sqrt{0.68780644} \approx 0.829
\end{aligned}$$

After that, we compute the 95% Confidence Interval for $\Delta\eta$

$$\begin{aligned}
\text{Lower limit} &= \Delta\eta - 1.96 \times \text{SE}(\Delta\eta) = -1.98829 - 1.96 \times 0.829 \approx -1.98829 - 1.62484 \approx -3.61313 \\
\text{Upper limit} &= \Delta\eta + 1.96 \times \text{SE}(\Delta\eta) = -1.98829 + 1.96 \times 0.829 \approx -1.98829 + 1.62484 \approx -0.36345
\end{aligned}$$

Finally, we simply determine 95% Confidence Interval for the Odds Ratio as follows,

$$\begin{aligned}
\text{OR}_{Lower} &= e^{-3.61313} \approx 0.027 \\
\text{OR}_{Upper} &= e^{-0.36345} \approx 0.695 \\
95\%\ \text{Confidence Interval} &= (0.027, 0.695)
\end{aligned}$$

Problem 2

The data below are the launch temperatures *(degrees Fahrenheit)* and an indicator of O-ring failures for 24 space shuttle launches prior to the space shuttle Challenger disaster of January 27, 1986.

Temperature	Failure	Temperature	Failure	Temperature	Failure
53	Yes	68	No	75	No
56	Yes	69	No	75	Yes
57	Yes	70	No	76	No
63	No	70	Yes	76	No
66	No	70	Yes	78	No
67	No	70	Yes	79	No
67	No	72	No	80	No
67	No	73	No	81	No

Part A

We are given data on launch temperatures and O-ring failures. Let's proceed with a logistic regression of `Failure` (binary: 1 for "Yes", 0 for "No") on `Temperature`.

1. Encode `Failure` as 1 for "Yes" and 0 for "No".
2. Fit the logistic regression model where the probability of failure is modeled as a function of `Temperature`.

   The estimated coefficients are:

   - Intercept ($\hat{\beta}_0$): 10.39366, SE: 5.52701
   - Temperature coefficient ($\hat{\beta}_1$): -0.16056, SE: 0.08014

Therefore,

- $\hat{\beta}_0 = 10.39366$, SE $= 5.52701$
- $\hat{\beta}_1 = -0.16056$, SE $= 0.08014$

```r
shuttleDs <- data.frame(
  Temperature = c(53, 56, 57, 63, 66, 67, 67, 67, 68, 69,
                  70, 70, 70, 70, 72, 73, 75, 75, 76, 76, 78, 79, 80, 81),
  Failure = c(1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1,
              1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0)
)
# Fit the logistic regression model
logisticModel <- glm(Failure ~ Temperature, family = binomial,
                     data = shuttleDs)
summary(logisticModel)
```

```

Call:
glm(formula = Failure ~ Temperature, family = binomial, data = shuttleDs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2976  -0.9187  -0.5123   0.6756   1.9102

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 10.39366    5.52701   1.881   0.0600 .
Temperature -0.16056    0.08014  -2.004   0.0451 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 30.553  on 23  degrees of freedom
Residual deviance: 24.986  on 22  degrees of freedom
AIC: 28.986

Number of Fisher Scoring iterations: 4
```

Part B

We will now test whether the coefficient of `Temperature` is 0 using Wald's test.

Let's define our hypotheses as follows,

- Null Hypothesis (H_0): $\beta_1 = 0$
- Alternative Hypothesis (H_a): $\beta_1 < 0$ (temperature decreases odds of failure)

Now, calculate Wald's test statistic

$$Z = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{-0.16056}{0.08014} \approx -2.004$$

```r
beta1 <- coef(summary(logisticModel))["Temperature", "Estimate"]
seBeta1 <- coef(summary(logisticModel))["Temperature", "Std. Error"]
Z <- beta1 / seBeta1
pVal <- pnorm(Z)
Z
```

```
[1] -2.003587
```

```r
pVal
```

```
[1] 0.02255718
```

Then, we can compute the one-sided p-value for $Z = -2.004$

$$p \approx 0.02255718$$

Thus, $p = 0.02255718 < 0.05$, we reject H_0. Hence, there is evidence that temperature negatively affects the odds of failure.

Part C

We will now compute a 95% confidence interval for the temperature coefficient.

$$\text{Lower limit} = \hat{\beta}_1 - 1.96 \times SE(\hat{\beta}_1) = -0.16056 - 1.96 \times 0.08014 \approx -0.3177$$

$$\text{Upper limit} = \hat{\beta}_1 + 1.96 \times SE(\hat{\beta}_1) = -0.16056 + 1.96 \times 0.08014 \approx -0.0034$$

95% Confidence Interval for β_1: (-0.3177, -0.0034)

Part D

Let's interpret the effect of temperature on the odds of O-ring failure.

Thus, let us look at the odds ratio, where for a one-degree increase in temperature leads to,

$$\mathrm{OR} = e^{\hat{\beta}_1} = e^{-0.16056} \approx 0.8517$$

Its 95% confidence interval is,

- Lower limit: $e^{-0.3177} \approx 0.7276$
- Upper limit: $e^{-0.0034} \approx 0.9966$

What does this mean? Well,

- Each one-degree increase in temperature reduces the odds of O-ring failure by approximately 14.83%.
- The true odds reduction is between 0.34% and 27.24%.

Part E

Finally, we compute the estimated probability of failure at 31°F, where

$$\eta = 10.39366 + (-0.16056) \times 31 \approx 5.4353$$
$$p = \frac{e^{5.4353}}{1 + e^{5.4353}} \approx 0.9956$$

Which means that the estimated probability of failure at 31°F is about 0.9956 (99.56%)