Assignment 1 - PH242C/STAT247C

John Semerdjian

September 14, 2015

Logistic Regression Review and Simulation

Run the associated dofile from STATA to help you complete the simulation and answer the questions below. (Note: I have converted the STATA dofile into R, so my randomly generated data may differ.)

1.) Based on the code used to simulate the data, describe the data-generating distribution, including the model of the regression, e.g.,

$$\text{logit}\{\text{E}(Y|X_1, X_2)\} = b_0 + b_1 X_1 + b_2 X_2, b_0 = -2.0, b_1 = 2.0, b_2 = -2.25$$

X_1 is drawn from a uniform distribution, $X_1 \sim \text{Uniform}(0, 5)$, while X_2 is drawn from the X_1 uniform distribution plus some error, $X_2 \sim 0.5X_1 + \text{Normal}(0, 2)$.

The model of regression is:

$$\begin{aligned}\text{logit}\{\text{E}(Y|X_1, X_2)\} &= -2.0 + 2.0X_1 - 2.25X_2 + \text{E}(e|X_1 = x_1, X_2 = x_2) \\ &= -2.0 + 2.0X_1 - 2.25(0.5X_1) + \text{E}(e|X_1 = x_1, X_2 = x_2) \\ &= -2.0 + 0.875X_1\end{aligned}$$

2.) Calculate the predicted value at $X_1 = 0, X_2 = 1$.

$$\text{E}(Y|X_1 = 0, X_2 = 1) = 0.014$$

3.) What is the true odds ratio when X_1 changes by 0.5, keeping X_2 fixed?

$$2.0(0.5X_1)$$
$$\exp^1 = 2.7183$$

4.) Repeat 1-3 for the estimated model but also give a 95% CI for the odds ratio. Do this for both sample sizes.

For $n = 100$:

$$\text{logit}\{\text{E}(Y|X_1 = x_1, X_2 = x_2)\} = -2.737 + 2.563x_1 + -2.752x_2$$
$$\text{E}(Y|X_1 = 0, X_2 = 1) = 0.0111$$

95% CI for Odds Ratio:

```
                      OR       2.5 %     97.5 %
(Intercept)   0.06478143 0.006669738  0.3442671
b1           12.97520624 4.233645068 68.2337257
b2            0.06381666 0.008491323  0.2111234
```

True odds ratio when X_1 changes by 0.5, keeping X_2 fixed:

```
                 Estimate lwr    upr
0.5 * X1 == 0 3.6020   1.8334 7.0766
```

For $n = 500$:

$$\text{logit}\{\text{E}(Y|X_1 = x_1, X_2 = x_2)\} = -1.762 + 1.759x_1 + -1.963x_2$$
$$\text{E}(Y|X_1 = 0, X_2 = 1) = 0.0615$$

95% CI for odds ratio:

```
                   OR      2.5 %    97.5 %
(Intercept) 0.1716313 0.09543207 0.2946313
b1          5.8043431 4.05088995 8.7589628
b2          0.1404379 0.09097498 0.2042333
```

True odds ratio when X_1 changes by 0.5, keeping X_2 fixed:

```
                 Estimate lwr    upr
0.5 * X1 == 0 2.4092   1.9879 2.9197
```

5.) Interpret to the best of your ability every number in the output on the row of results starting with X_2 (for sample size $n = 100$).

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.7367     0.9750  -2.807 0.005001 **
X1            2.5630     0.6891   3.719 0.000200 ***
X2           -2.7517     0.7888  -3.488 0.000486 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Estimate = $\hat{b}_2$ = -2.7517: For a one unit increase in X_2, there is a 2.7517 unit decrease in the log odds of Y holding X_1 constant.
- Std. Error = 0.7888: This is the estimated standard error of b_2, the coefficient of X_2 in the regression.
- z value = -3.488: The z value is a test statistic that is a part of the Wald test, comes from $H_0 : b_2 = 0$. Is calculated by $z = \dfrac{\hat{\beta} - 0}{\hat{\sigma}_{\hat{\beta}}} = \dfrac{-2.7517 - 0}{0.7888}$
- $Pr(> |z|) = 0.000486$: Assuming the null hypothesis is true, this is the probability of getting a z value this extreme or more extreme.
- 95% Conf. Int. = [-4.768711, -1.555313]: If the experiment is repeated infinitely many times and 95% confidence intervals are calculated each time, 95% of those intervals would contain the true parameter, $b_2 = -2.25$.

6.) Calculate and describes what happens to the estimated standard deviation (that is, the SE) of the estimate of b_1 when the sample size increases to $n = 500$ from $n = 100$.

The estimated standard error decreases from 0.6891 to 0.1961 when the sample size grows from 100 to 500.

$n = 100$:

```
        Estimate Std. Error z value Pr(>|z|)
b1        2.5630     0.6891   3.719 0.000200 ***
```

$n = 500$:

```
        Estimate Std. Error z value Pr(>|z|)
b1        1.7586     0.1961   8.967  < 2e-16 ***
```