Homework
Applied Logistic Regression

WEEK 3

Exercise 2:

Use the ICU study (icu.dta)

a. Using the results from Week 1, problem 2 part (d) compute 95 percent confidence intervals for the slope and constant term. Write a sentence interpreting the confidence interval for the slope.

You can compute the compute the confidence intervals for the slope and the constant term through both Stata and through hand-calculations.

To do so through Stata, simply run the logistic model by typing "logit STA AGE" into the command box. The 95% confidence intervals for both covariates are listed in the table below:

```
. logit STA AGE

Iteration 0:  Log Likelihood =-100.08048
Iteration 1:  Log Likelihood =-96.288372
Iteration 2:  Log Likelihood =-96.153701
Iteration 3:  Log Likelihood = -96.15319

Logit Estimates                                 Number of obs =     200
                                                chi2(1)       =    7.85
                                                Prob > chi2   = 0.0051
Log Likelihood =  -96.15319                     Pseudo R2     = 0.0392

------------------------------------------------------------------------------
     STA |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .0275426   .0105645      2.607   0.009       .0068366    .0482487
   _cons |  -3.058513   .6961091     -4.394   0.000      -4.422862   -1.694165
```

Alternatively, you can calculate the confidence interval for the slope and constant term by the equations $\hat{\beta}_1 \pm z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_1)$ and $\hat{\beta}_0 \pm z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_0)$, respectively.

Endpoints of a 100(1-α)% confidence interval for slope coefficient

$$\hat{\beta}_1 \pm z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_1)$$

$$0.0275 \pm 1.96(0.0106)$$

$$(0.0068,\ 0.0482)$$

Endpoints of a 100(1-α)% confidence interval for constant:

$$\hat{\beta}_0 \pm z_{1-\alpha/2}\widehat{SE}(\hat{\beta}_0)$$

$$-3.0585 \pm 1.96(0.6961)$$
$$(-4.4229, \ -1.6941)$$

The 95% confidence interval for the slope suggests that the change in the log odds of dying in the ICU (STA=1) per one year increase in AGE is 0.0275 and the change could be as little as 0.0068 or as much as 0.0482 with 95% confidence.

b. Obtain the estimated covariance matrix for the model fit in Week 1, problem 2 part (d). Compute the logit and estimated logistic probability for a 60-year old subject. Compute a 95 percent confidence intervals for the logit and estimated logistic probability. Write a sentence or two interpreting the estimated probability and its confidence interval.

Again, to generate a variance/covariance matrix, simply run the regression and then type "vce" into the command window. You should see the output below:

```
. vce

        |      AGE     _cons
--------+------------------
     AGE|   .000112
   _cons| -.007104   .484568
```

From this table, we are able to see that the variance of AGE is 0.000112, the variance for the constant term is 0.484568, and the covariance between AGE and the constant term is -0.007104 (Note: the discordant pairs in a variance/covariance matrix represent covariance, whereas the matching pairs represent variance)

First, generate the estimate for the logit for a 60-year-old subject:

$$\begin{aligned}\hat{g}(60) &= -3.0585 + 0.0275(60) \\ &= -1.4060\end{aligned}$$

Next, calculate the variance for the logit for this observation using the values calculated from the variance/covariance matrix:

$$\begin{aligned}Var(\hat{g}(60)) &= 0.4846 + (60)^2 0.000112 + 2(60)(-0.007104) \\ &= 0.0353\end{aligned}$$

The standard error can then be determined by calculating the square root of the variance.

$$SE(\hat{g}(60)) = 0.1879$$

The estimated standard error of the logit for a 60-year old subject, can be used to calculate the endpoints of the confidence interval for the logit for a 60-year old subject:

$$\hat{g}(60) \pm z_{1-\alpha/2} SE[\hat{g}(60)]$$
$$-1.4060 \pm 1.96(0.1879)$$
$$(-1.7741, -1.0378)$$

The estimated logit and the endpoints of its confidence interval can be used to obtain the estimated logistic probability and its confidence interval:

$$
\begin{aligned}
\hat{\pi}(60) &= \frac{e^{\hat{g}(60)}}{1+e^{\hat{g}(60)}} \\
&= \frac{e^{-3.059+0.028(60)}}{1+e^{-3.059+0.028(60)}} \\
&= 0.1969
\end{aligned}
$$

The endpoints of the confidence interval for the estimated logistic probability can be obtained by following a similar process using the endpoints of the estimated logit. The lower limit is:

$$\frac{e^{-1.7741}}{1+e^{-1.7741}} = 0.1450$$

and the upper limit is:

$$\frac{e^{-1.0378}}{1+e^{-1.0378}} = 0.2616$$

The estimated logistic probability of dying in the ICU for a 60 year old subject, 0.1969, is an estimate of the proportion of 60 year old subjects in the population sampled that die in the ICU. The confidence interval suggests that this mean could be as low as 0.1450 or as high as 0.2616 with 95% confidence.