Summative Quiz 2 (Simple Logistic Regression) Solutions:

1. Observational data were collected by researchers on 189 consecutive births at Baystate Medical Center in Springfield, Massachusetts (USA)[1]. The researchers were interested in using logistic regression to identify maternal and other factors associated with the low birth weight births (as per the World Health Organization's definition of low birth weight as < 2500 grams). The average birth weight for the 189 infants in the sample was 2,945 grams (s= 729, range -709 -4,990 grams) with 59 infants (30%) meeting the low birth weight criteria.

   The first regression relates the log (ln) odds of low birth weight (LBW) to mother's smoking status during pregnancy:

   $\ln(ODDs\ of\ LBW) = -1.09 + 0.70x_1$, where x_1= 1 if the mother smoked during pregnancy (and 0 if not). The estimated standard error of the slope of x_1 is 0.32.

   What is the estimated odds ratio for being low birth weight for children of mothers who smoked during pregnancy, compared to mothers who did not smoke during pregnancy?

   Answer: 2.0

   Reasoning: The slope for the indicator of whether the mother smoker during pregnancy, x_1, is the estimated difference in the ln(odds) of LBW between the two groups (smokers vs non smokers), i.e. the estimated ln(odds ratio). To get the corresponding odds ratio estimate, exponentiate this slope estimate: $e^{0.70} = 2.01 \approx 2.0$

2. (this item references the same logistic regression results as item #1)

   Compute and report a 95% CI for the odds ratio from the previous question (theodds ratio for being low birth weight for children of mothers who smoked during pregnancy, compared to mothers who did not smoke during pregnancy).

   Answer: (1.06, 3.82)

   Reasoning: The estimated 95% CI for the slope is $0.70 \pm 2*0.32 \rightarrow (0.06, 1.34)$. The estimated 95% CI for the odds ratio is $(e^{0.06}, e^{1.34}) \rightarrow (1.06, 3.82)$.

[1] Data from Hosmer D, and Lemeshow S. *Applied Logistic Regression Analysis*.

3. (this item references the same linear regression results as item #1)

   What is the estimated probability of being low-weight (proportion of low birth weight infants) for babies born to mothers who did not smoke during pregnancy?

   Answer: 0.25

   Reasoning: For this group, the estimated ln(odds of LBW) is -1.09. The estimated odds is $e^{-1.09}$ ≈0.34, and the estimated probability is $\hat{p} = \frac{\widehat{odds}}{1+\widehat{odds}} = \frac{0.34}{1.34} \approx 0.25$, or 25%.

4. (this item references the same logisitic regression results as item #1)

   What is the estimated probability of being low-weight (proportion of low birth weight infants) for babies born to mothers who smoked during pregnancy?

   Answer: 0.40

   Reasoning: For this group, the estimated ln(odds of LBW) is -1.09+0.70=-0.39. The estimated odds is $e^{-0.39}$ ≈0.68, and the estimated probability is $\hat{p} = \frac{\widehat{odds}}{1+\widehat{odds}} = \frac{0.68}{1.68} \approx 0.40$, or 40%

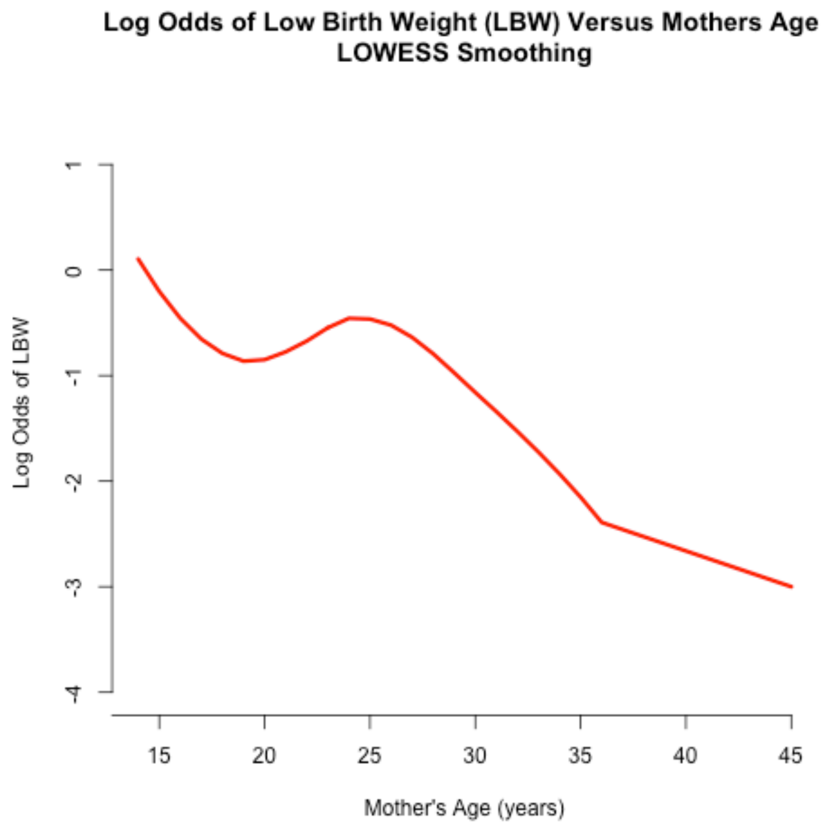
5. (this item references the same logistic regression results as item #1)

   What is the estimated relative risk for being low birth weight for children of mothers who smoked during pregnancy, compared to mothers who did not smoke during pregnancy?

   Answer: 1.6

   Reasoning: $\widehat{RR} = \frac{0.4}{0.25} = 1.6$. Notice this is smaller in magnitude than the OR estimate.

6. The next analysis involves estimating the association between low birth weight and age of the mother (at the time of birth). The following "lowess" plot show the association between the log (ln) odds of infants being low birthright and the age of the mothers.



=

The following logistic regression relates the two measures:

$\ln(ODDS\ of\ LBW) = 0.38 + -0.05x_1$, where x_1= 1 age of the mother. The estimated standard error of the slope of x_1 is 0.03, and the range of mother's ages is 14-45 years.

What is the nature of the estimated relationship between the risk infants being low birth weight and the age of the mother?

Answer: The risk of being low birth weight decreases with an increase in mother's age.

Reasoning: increased maternal age is associated with reduced ln(odds), and hence reduced risk of LBW. (recall that ln(odds), odds, and risk (probability, proportion) all track together: i.e. an increase in one of the three results in an increase in the other two quantities, and a decrease in one of the three results in a decrease in the other two quantities).

7. (This item references the same logistic regression results as item #6)

   What is the estimated odds ratio, and 95% CI for the odds of delivering a low birth weight baby for two groups of mothers who differ by 1 year in age?

   Answer: 0.95 (0.89, 1.01)

   Reasoning: $e^{-0.05} \approx 0.95$.

   The 95% for the slope is $-0.05 \pm 2(0.03) \rightarrow (-0.11, 0.01)$.
   The 95% for the odds ratio is $(e^{-0.11}, e^{0.01}) \rightarrow (0.89, 1.01)$.

8. (This item references the same logistic regression results as item #6)

   Is the relationship between low birth weight and mothers' ages statistically significant ($\alpha=0.05$)?

   Answer: No.

   Reasoning: Notice that the 95% CI for ln(OR), i.e. the slope ,includes 0 , and hence the 95% CI for the odds ratio includes 1.

9. (This item references the same logistic regression results as item #6)

   What is the estimated odds ratio, and 95% CI for the odds of delivering a low birth weight baby for 40 years old mothers compared to 30 year old mothers?

   Answer: 0.60 (0.31, 1.10)

   Reasoning: The estimated odds ratio is $(0.95)^{10} \approx 0.60$.
   The estimated 95% CI is $(.89^{10}, 1.01^{10}) \rightarrow (0.31, 1.10)$

   (Notice: the above is the end result of doing the following computations on the regression scale and exponentiating the results:

   $$\text{For 40 year olds: } \ln(ODDS\ of\ LBW) = 0.38 + -0.05(40)$$
   $$\text{For 30 year olds: } \ln(ODDS\ of\ LBW) = 0.38 + -0.05(30)$$

   So the difference in ln(odds), ie: the ln(odds ratio) is -0.05(40 -30)= -0.05(10) = -0.5.
   So the estimated odds ratio is $e^{-0.5} \approx 0.60$. The 95% CI for the slope β_1 is (-0.11,0.01).
   The 95% CI for $10\beta_1$ is (10(-0.11), 10(-0.01)) = (-1.10, -0.1). Exponentiating these endpoints give $(e^{-1.10}, e^{-0.1})$ = (0.33,1.1)
   (the first endpoint is slightly different than the one above because of rounding).

10. (This item references the same logistic regression results as item #6)

What is the estimated probability of a 30 year old mother giving birth to a low birth weight baby?

Answer: 0.25 (24%)

Reasoning: For this group, the estimated ln(odds of LBW) is 0.38+30*-0.05=-1.12. The estimated odds is $e^{-1.12}$ ≈0.33, and the estimated probability is $\hat{p} = \frac{\widehat{odds}}{1+\widehat{odds}} = \frac{0.33}{1.33} \approx 0.25$, or 25%.