Tutorial Exercises – Week 5

Question 1. (Textbook 3.3)
The median starting salary for new law school graduates is determined by

\[
\log(salary) = \beta_0 + \beta_1 \text{LSAT} + \beta_2 \text{GPA} + \beta_3 \log(\text{libvol}) + \beta_4 \log(\text{cost}) + \beta_5 \text{rank} + u
\]

where \( \text{LSAT} \) is median LSAT score for the graduating class, \( \text{GPA} \) is the median college GPA for the class, \( \text{libvol} \) is the number of volumes in the law school library, \( \text{cost} \) is the annual cost of attending law school, and \( \text{rank} \) is a law school ranking (with \( \text{rank} = 1 \) being the best).

(i) Explain why we expect \( \beta_5 \leq 0 \).
(ii) What signs to you expect for the other slope parameters? Justify your answers.
(iii) Using the data in LAWSCH85.RAW, the estimated equation is

\[
\log(salary) = 8.34 + .0047 \cdot \text{LSAT} + .248 \cdot \text{GPA} + .095 \log(\text{libvol}) + .038 \log(\text{cost}) - .0033 \cdot \text{rank}
\]

\( n = 136, R^2 = .842 \).

What is the predicted ceteris paribus difference in salary for schools with a median GPA different by one point? (Report your answer as a percent.)
(iv) Interpret the coefficient on the variable \( \log(\text{libvol}) \).
(v) Would you say it is better to attend a higher ranked law school? How much is a difference in ranking of 20 worth in terms of predicted starting salary?

Question 2. (Textbook 4.7)
Consider the estimated equation from Example 4.3, which can be used to study the effects of skipping class on college GPA:

\[
\frac{\text{colGPA}}{.33} = 1.39 + .412 \cdot \text{hsGPA} + .015 \cdot \text{ACT} - .083 \cdot \text{skipped}
\]

\( n = 141, R^2 = .234 \).

(i) Using the standard normal approximation, find the 95% confidence interval for \( \beta_{\text{hsGPA}} \).
(ii) Can you reject the hypothesis \( H_0 : \beta_{\text{hsGPA}} = .4 \) against the two-sided alternative at the 5% level?
(iii) Can you reject the hypothesis \( H_0 : \beta_{hsGPA} = 1 \) against the two-sided alternative at the 5% level?

**Question 3. (Textbook 7.2)**

The following equations were estimated using the data in BWGHT.RAW:

\[
\begin{align*}
\log(bwght) &= 4.66 - 0.0044 \text{cigs} + 0.0093 \log(faminc) + 0.016 \text{parity} \\
&\quad + 0.027 \text{male} + 0.055 \text{white} \\
&\quad (0.22) \quad (0.009) \quad (0.0059) \quad (0.06) \\
&\quad (0.010) \quad (0.013) \\
&\quad n = 1,388, R^2 = .0472
\end{align*}
\]

and

\[
\begin{align*}
\log(bwght) &= 4.65 - 0.0052 \text{cigs} + 0.0110 \log(faminc) + 0.017 \text{parity} \\
&\quad + 0.034 \text{male} + 0.045 \text{white} - 0.0030 \text{motheduc} + 0.0032 \text{fatheduc} \\
&\quad (0.38) \quad (0.010) \quad (0.0085) \quad (0.06) \\
&\quad (0.011) \quad (0.015) \quad (0.0030) \quad (0.0026) \\
&\quad n = 1,191, R^2 = .0493
\end{align*}
\]

The variables are defined as in Example 4.9 (where \( bwght \) is birth weight, in pounds, \( cigs \) is average number of cigarettes the mother smoked per day during pregnancy, parity is the birth order of this child, \( faminc \) is annual family income, \( motheduc \) is years of schooling for the mother, and \( fatheduc \) is years of schooling for the father), but we have added a dummy variable for whether the child is male and a dummy variable indicating whether the child is classified as white.

(i) In the first equation, interpret the coefficient on the variable \( cigs \). In particular, what is the effect on birth weight from smoking 10 more cigarettes per day?

(ii) How much more is a white child predicted to weigh than a nonwhite child, holding the other factors in the first equation fixed? Is the difference statistically significant?

(iii) Comment on the estimated effect and statistical significance of \( motheduc \).

(iv) From the given information, why are you unable to compute the F statistic for joint significance of \( motheduc \) and \( fatheduc \)? What would you have to do to compute the F statistic?
Question 4. (Textbook 7.4)

An equation explaining chief executive officer salary is

\[
\log(\text{salary}) = 4.59 + .257 \log(\text{sales}) + .011 \text{ roe} + .158 \text{ finance} \\
(\ .30) (\ .032) (\ .004) (\ .089) \\
+ .181 \text{ consprod} - .283 \text{ utility} \\
(\ .085) (\ .099)
\]

\[n = 209, R^2 = .357.\]

The data used are in CEOSAL1.RAW, where finance, consprod, and utility are binary variables indicating the financial, consumer products, and utilities industries. The omitted industry is transportation.

(i) Compute the approximate percentage difference in estimated salary between the utility and transportation industries, holding sales and roe fixed. Is the difference statistically significant at the 1% level?

(ii) Use equation (7.10)* to obtain the exact percentage difference in estimated salary between the utility and transportation industries and compare this with the answer obtained in part (i).

(iii) What is the approximate percentage difference in estimated salary between the consumer products and finance industries? Write an equation that would allow you to test whether the difference is statistically significant.

*: Generally, if \( \hat{\beta}_i \) is the coefficient on a dummy variable, say \( x_i \), when \( \log(y) \) is the dependent variable, the exact percentage difference in the predicted \( y \) when \( x_i = 1 \) versus when \( x_i = 0 \) is

\[
100 \cdot [\exp(\hat{\beta}_i) - 1].
\]  \hspace{1cm} (7.10)

The estimate \( \hat{\beta}_i \) can be positive or negative, and it is important to preserve its sign in computing (7.10).