Tutorial Exercises – Week 10

Question 1 (Textbook 17.2):
Let \( \text{grad} \) be a dummy variable for whether a student-athlete at a large university graduates in five years. Let \( \text{hsGPA} \) and \( \text{SAT} \) be high school grade point average and SAT score. Let \( \text{study} \) be the number of hours spent per week in organized study hall. Suppose that, using data on 420 student-athletes, the following logit model is obtained:

\[
\hat{P}(\text{grad}=1|\text{hsGPA}, \text{SAT}, \text{study}) = \Lambda(-1.17 + .24 \text{hsGPA} + .00058 \text{SAT} + .073 \text{study}),
\]

where \( \Lambda(z) = \exp(z)/(1+\exp(z)) \) is the logit function. Holding \( \text{hsGPA} \) fixed at 3.0 and \( \text{SAT} \) fixed at 1,200; compute the estimated difference in the graduation probability for someone who spent 10 hours per week in study hall and someone who spent five hours per week.

Question 2 (Textbook C17.1)

(i) The variable \( \text{favwin} \) is a binary variable if the team favored by the Las Vegas point spread wins. A linear probability model to estimate the probability that the favored team wins is

\[
P(\text{favwin}=1|\text{spread}) = \beta_0 + \beta_1 \text{spread}
\]

Explain why, if the spread incorporates all relevant information, we expect \( \beta_0 = 0.5 \).

(ii) Estimated model from part (i) by OLS is given as

\[
\hat{\text{favwin}} = .577 + .0194 \text{spread}
\]

\[
(0.028) (0.0023)
\]

\[
R^2 = .111.
\]

Test \( H_0: \beta_0 = 0.5 \) against a two-sided alternative. Use both the usual and heteroskedasticity-robust standard errors.

(iii) Is \( \text{spread} \) statistically significant? What is the estimated probability that the favored team wins when \( \text{spread} = 10 \)?

(iv) A probit model for \( P(\text{favwin}=1|\text{spread}) \) is estimated and its results are given in the table below. Interpret and test the null hypothesis that the intercept is zero. \([\text{Hint: Remember that } \Phi(0) = .5]\)
(v) Use the probit model to estimate the probability that the favored team wins when $spread = 10$. Compare this with the LPM estimate from part (iii).

(vi) When we add the variables $favhome$, $fav25$, and $und25$ to the probit model the Log Likelihood value for the new model becomes $-262.64$. Test joint significance of these variables using the likelihood ratio test. (How many $df$ are in the chi-square distribution?) Interpret this result, focusing on the question of whether the spread incorporates all observable information prior to a game.

**Question 3 (Textbook C17.3)**

(i) Out of 616 workers, 172 have zero benefits. For what percentage of the workers in the sample is $pension$ equal to zero? Given the range of $pension$ for workers with nonzero pension benefits is $7.28$ to $2880.27$, explain why is a Tobit model appropriate for modeling $pension$?

(ii) Estimated Tobit model explaining $pension$ in terms of $exper$, $age$ $tenure$, $educ$, $depends$, $married$, $white$, and $male$ is presented in the table given below. Do whites and males have statistically significant higher expected pension benefits?
(iii) Use the results from part (ii) to estimate the difference in expected pension benefits for a white male and a nonwhite female, both of whom are 35 years old, are single with no dependents, have 16 years of education, and 10 years of experience.