

Homework #5

due Tuesday, November 26, at 2pm

Instructions: Complete all problems and turn in a set of answers either to me or under my door (office 335L) by the assigned due date.

Problem 1 You estimate the following regression model:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_1 * X_2 + \beta_4 * X_2^2 + \epsilon \quad (1)$$

You obtain the following estimates:

Coefficient	estimate
β_0	2
β_1	4
β_2	1
β_3	-1
β_4	-2

a. For a given observation, $X_2 = 5$. Based on your results, what is the marginal effect on Y of a one-unit increase in X_1 from $X_1 = 6$ to $X_1 = 7$?

b. What is the effect of a one-unit increase in X_2 from 5 to 6, given that $X_1 = 7$?

Problem 2 Alice is interested in how salaries for similar work differ across the public and private sectors. She obtains employment data on 30,000 workers randomly sampled from the entire country.

She uses the following regression model, where *SALARY* is annual salary in dollars, and *PUBLIC* equals 1 if an individual is employed in the public sector, and 0 otherwise:

$$SALARY = \beta_0 + \beta_1 * PUBLIC + \epsilon \quad (2)$$

She obtains the following results:

	R^2	.091
coefficient	estimate	p-value
Intercept	45,000	.00012
β_1	12,000	.03549

a. Interpret Alice's estimate of β_1 .

b. Alice suggests that her results prove that public sector wages are excessive, and that governments could lower the salaries they offer and attract the same workers. Do you agree? Explain.

Bob thinks that public sector employees may have different individual characteristics than private sector employees. Using Alice's data, Bob estimates the following regression:

$$SALARY = \beta_0 + \beta_1 * PUBLIC + \beta_2 * EDUCATION + \beta_3 * AGE + \beta_4 * AGE^2 + \epsilon \quad (3)$$

where *EDUCATION* is years of education and *AGE* is age of the employee. He obtains the following results:

	R^2 .63	
coefficient	estimate	p-value
β_0	25,000	.00005
β_1	4,000	.00312
β_2	1,200	.01579
β_3	300	.02697
β_4	-3	.00049

- c. Bob's estimate of β_1 is lower than Alice's. Explain in simple, intuitive terms why this might be the case.
- d. Based on a comparison of Bob's and Alice's results, do you think that public sector employees are more educated, or less educated than private sector employees, on average?
- e. Do Bob's results predict that a 40-year old public-sector employee or a 60-year old public sector employee will make more, all else equal?
- f. At approximately what age do Bob's results predict that SALARY will peak?

Problem 3 (use dataset 18-17) The manager of an amusement park would like to predict daily attendance in order to develop more accurate plans about how much food to order and how many ride operators to hire. After some consideration, he decided that the following three factors are critical:

- Yesterday's attendance
- Weekday or weekend
- Predicted weather

He then took a random sample of 40 days. For each day, he recorded the attendance, the previous day's attendance, day of the week, and weather forecast. The first independent variable is interval, but the other two are nominal. Accordingly, he created the following sets of indicator variables:

$$\begin{aligned}
 I_1 &= \begin{cases} 1 & \text{(if weekend)} \\ 0 & \text{(if not)} \end{cases} \\
 I_2 &= \begin{cases} 1 & \text{(if mostly sunny is predicted)} \\ 0 & \text{(if not)} \end{cases} \\
 I_3 &= \begin{cases} 1 & \text{(if rain is predicted)} \\ 0 & \text{(if not)} \end{cases}
 \end{aligned}
 \tag{4}$$

- a. Conduct a regression analysis in which you regress attendance on yesterday's attendance, weekend/weekday status, and weather. Which coefficients are significant?
- b. Can we conclude that weather is a factor in determining attendance?
- c. Do these results provide sufficient evidence that weekend attendance is, on average, larger than weekday attendance?

Problem 4 (use dataset GSS2008) The General Social Survey asks respondents to report the number of hours per average day of television viewing (TVHOURS). Conduct a regression analysis using the following independent variables:

- Education (EDUC)
- Age (AGE)
- Hours of work (HRS)
- Number of children (CHILDS)
- Number of family members earning money (EARNRS)
- Occupation prestige score (PRESTG80)

- a. For each coefficient, say whether or not it is significant.
- b. What is the coefficient of determination, and what does it tell you?
- c. A friend suggests that men watch more TV than women. Test this suggestion by including sex (SEX) as an independent variable, and interpreting its coefficient.
- d. Another friend suggests that men and women may be differentially affected by the number of children. Test this claim by including a term interacting sex and number of children, and interpret its results.