

# Assignment 07

## *Analysis of Covariance Using Multiple Regression: Part II*

In this assignment, you are going to focus on the question of whether there are differences in the average in-state tuition across the five states. Please submit your responses to each of the questions below in a printed document. Also, please adhere to the following guidelines for further formatting your assignment:

- All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption and labels.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Equation Editor, Markdown, or  $\text{\LaTeX}$ .

This assignment is worth 10 points. Each question is worth one point unless otherwise noted.

---

For this assignment, you will use the file *colleges-bordering-mn.csv*. This file contains institutional data for 104 colleges and universities in the five state area (MN, IA, WI, ND, and SD). These data were collected by the Department of Education for the [2013 College Score Card](#). The variables are:

- **name:** Name of college/university
- **tuition\_in\_state:** In-state tuition and fees
- **tuition\_out\_state:** Out-of-state tuition and fees
- **state:** State postal abbreviation (IA = Iowa, MN = Minnesota, ND = North Dakota, SD = South Dakota, WI = Wisconsin)
- **public:** Is this a public school? (1 = yes, 0 = no)
- **main:** Is this the main campus? (1 = yes, 0 = no)
- **admission:** Admission rate
- **act75:** 75th percentile of the ACT cumulative scores
- **avg\_fac\_salary:** Average faculty salary (per month)
- **completion:** Four-year completion rate for first-time, full-time students
- **pct\_pell:** Percentage of undergraduates who receive a Pell grant

---

## Preparation: Create Dummy Variables

Create five dummy variables, one for each state, for the analysis.

## Unadjusted Group Differences Model: ANOVA

Fit the regression model that uses the dummy predictors for state to predict variation in average in-state tuition. In this model, use Minnesota as the reference group.

1. Write the fitted regression equation.
2. Are there any statistically reliable differences in the average in-state tuition between Minnesota and its bordering states? Explain.
3. Report and interpret the  $R^2$  value for this model.
4. Which group difference(s) is/are not represented in this fitted model?

## Adjusted Group Differences Model: ANCOVA

Again, fit the regression model that uses the dummy predictors for state to predict variation in average in-state tuition, but this time control for differences in (1) sector (public vs. private), (2) ACT scores, and (3) percentage of students on Pell grants. Again, use Minnesota as the reference group.

5. Write the fitted regression equation.
6. Are there any statistically reliable differences in the average in-state tuition between Minnesota and its bordering states after controlling for these other predictors? Explain.
7. Report and interpret the  $R^2$  value for this model.

## Pairwise Differences

8. Create a table (suitable for publication) that presents each of the possible pairwise contrasts (hypothesis) of interest, the unadjusted  $p$ -values, and the Benjamini–Hochberg adjusted  $p$ -values for the controlled differences. (Note: To obtain all of these, you may need to fit additional models.)
9. Based on the Benjamini–Hochberg adjusted  $p$ -values, which states have a statistically reliable difference in their average in-state tuition after controlling for the other predictors in the model?
10. Create a heatmap of the information that you reported in the table in Question 8. You can see an example of a [heatmap for correlations here](#). We want to create a heatmap that shows the  $p$ -value for each mean comparison instead of the correlation between variables. So for example, the heatmap we want to create should be a 5x5 grid (the rows and columns would represent states) and the intersecting cells would include the  $p$ -value for the comparison (rather than the correlation coefficient). Color will be used to indicate the magnitude of the  $p$ -values. You may want to include different levels of color depending on the degree of significance ( $< .05$ ,  $< .01$ , etc. At the very least, make sure it is clear whether each comparison is statistically significant or not. Feel free to use any software tool you want to create this heatmap.