HUDM 5123 - Linear Models and Experimental Design
Lab 03 - Categorical Predictors and ANOVA

1 The Data

The **car** package in R comes with a data set called "Prestige" that includes 102 observations with six variables. Each observation is an occupation listed in the 1971 Census of Canada.

- **education**: Average education of occupational incumbents, years, in 1971.
- **income**: Average income of incumbents, dollars, in 1971.
- **women**: Percentage of incumbents who are women.
- **prestige**: Pineo-Porter prestige score for occupation, from a social survey conducted in the mid-1960s.
- **census**: Canadian Census occupational code.
- **type**: Type of occupation. A factor with levels: *bc*, blue collar; *prof*, professional, managerial, and technical; *wc*, white collar.

The primary research question is whether the categorical occupation type variable is a significant predictor of job prestige. The three levels of the occupation type variable are blue collar, white collar, and professional.

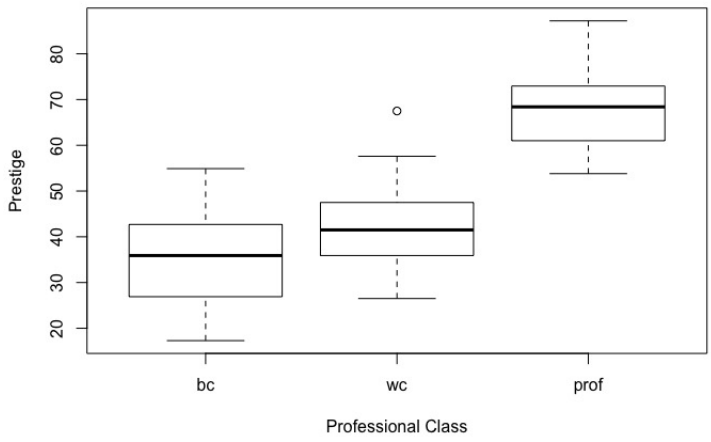

Figure 1: Boxplot of prestige ratings by profession type

Add two dummy-coded variables to the Prestige data frame, using bc as the reference category. Then use the dummies to run and save the following models:

$$\text{prestige}_{i_F} = \beta_0 + \beta_1 \text{WC}_i + \beta_2 \text{Prof}_i + \epsilon_{i_F}$$
$$\text{prestige}_{i_R} = \beta_0 + \epsilon_{i_R}$$

Task 1 *Run the full model and save it as lm_F1. Interpret the intercept and the slope coefficients in the context of this problem.*

Task 2 *Compare the full and reduced models using the* `anova()` *function and use the output to fill in an ANOVA table. Write out the null and alternative hypothesis for the incremental F test, and write your conclusion out (i.e., is job type a significant predictor of job prestige?) Use APA format for writing up F statistic, degrees of freedom, and p-value.*

Add two deviation-coded variables to the Prestige data frame, again using bc as the reference category. Then use the deviation-coded variables to run and save the following models:

$$\text{prestige}_{i_F} = \beta_0 + \beta_1 \text{WC}_i + \beta_2 \text{Prof}_i + \epsilon_{i_F}$$
$$\text{prestige}_{i_R} = \beta_0 + \epsilon_{i_R}$$

Task 3 *Run the full model and save it as lm_F2. Interpret the intercept and the slope coefficients in the context of this problem.*

Task 4 *Compare the full and reduced models using the* `anova()` *function and use the output to fill in an ANOVA table. Write out the null and alternative hypothesis for the incremental F test, and write your conclusion out (i.e., is job type a significant predictor of job prestige?) Use APA format for writing up F statistic, degrees of freedom, and p-value.*