Assignment 3

Introduction to Applied Statistics and Data Science

Sophie Ennis & Hannah Yoon

2025-03-05

The traditional convention in the social sciences is to use a critical value of 0.05 (α = 0.05) (Jean, Lecture Day 6, Slide 15). Thus, we selected an alpha value of 0.05 to evaluate the p-values of the hypothesis tests for statistical significance. Any question that examines the statistical significance of p-values (in terms of the hypothesis tests) in this report uses this alpha value.

# QUESTION 1

### We’ll begin with the final model from the prior assignment – with occupational prestige being predicted by father’s occupational prestige, years of education, race, and gender. We’ll call this our base model. Run this model and note the results (include the output in your report), but you do not need to interpret them. In the analyses below, run each model as a variation on the base model. That is, as you work through the assignment, return to the base model from Question 1 and make only the relevant changes specified. Do not carry over changes from the previous question.

gss$Female <- ifelse(gss$Sex=="Female", 1, 0)   
  
base\_model <- summ(lm(OccPres ~ DadPres + Female + Race + YearsEd, data=gss), digits = 3)   
  
base\_model

## MODEL INFO:  
## Observations: 1776 (572 missing obs. deleted)  
## Dependent Variable: OccPres  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(5,1770) = 105.855, p = 0.000  
## R² = 0.230  
## Adj. R² = 0.228   
##   
## Standard errors:OLS  
## ---------------------------------------------------  
## Est. S.E. t val. p  
## ----------------- -------- ------- -------- -------  
## (Intercept) 14.435 1.563 9.236 0.000  
## DadPres 0.072 0.023 3.151 0.002  
## Female -0.297 0.572 -0.518 0.604  
## RaceBlack -2.417 0.887 -2.726 0.006  
## RaceOther 0.608 0.916 0.664 0.507  
## YearsEd 2.043 0.098 20.800 0.000  
## ---------------------------------------------------

The intercept is 14.435. The coefficient for the father's occupational prestige variable is 0.072. The coefficient for the female variable is –0.297. The coefficient for the Black race variable is    –2.417. The coefficient for the Other race variable is 0.608. The coefficient for the years of education variable is 2.043.

# QUESTION 2

### Modify the base model to center years of education at 12 years, and the other variables around their means (Hint: You’ll need to create dummy variables and center them manually; you cannot simply enter the cleaned factor variables into the model this time). Do not center the outcome variable.

gss$DadPresC <- gss$DadPres - mean(gss$DadPres, na.rm = T)   
  
gss$Female <- ifelse(gss$Sex=="Female", 1, 0)   
  
gss$FemaleC <- gss$Female - mean(gss$Female, na.rm=T)   
  
gss$Black <- ifelse(gss$Race=="Black", 1, 0)   
  
gss$BlackC <- gss$Black - mean(gss$Black, na.rm=T)   
  
gss$Other <- ifelse(gss$Race=="Other", 1, 0)   
  
gss$OtherC <- gss$Other - mean(gss$Other, na.rm=T)   
  
gss$AgeC <- gss$Age - mean(gss$Age, na.rm = T)   
  
gss$YearsEd12 <- gss$YearsEd - 12   
  
model2 <- summ(lm(OccPres ~ DadPresC + FemaleC + BlackC + OtherC + YearsEd12, data=gss), digits = 3)   
  
model2

## MODEL INFO:  
## Observations: 1776 (572 missing obs. deleted)  
## Dependent Variable: OccPres  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(5,1770) = 105.855, p = 0.000  
## R² = 0.230  
## Adj. R² = 0.228   
##   
## Standard errors:OLS  
## ----------------------------------------------------  
## Est. S.E. t val. p  
## ----------------- -------- ------- --------- -------  
## (Intercept) 41.664 0.344 121.192 0.000  
## DadPresC 0.072 0.023 3.151 0.002  
## FemaleC -0.297 0.572 -0.518 0.604  
## BlackC -2.417 0.887 -2.726 0.006  
## OtherC 0.608 0.916 0.664 0.507  
## YearsEd12 2.043 0.098 20.800 0.000  
## ----------------------------------------------------

## Part A

### Interpret the intercept. Explain the meaning of its test of statistical significance, including the null hypothesis.

The intercept is 41.664, which represents the expected occupational prestige of the reference group, which is high school educated white men with average father's occupational prestige. The p-value of the intercept is significant, so we reject the null hypothesis that the intercept is 0. This indicates the intercept is meaningful and has a measurable value different from 0 that this not attributable to chance alone when the years of education predictor variable is centered at 12 and the other predictor variables are centered around their means.

## Part B

### Interpret the coefficients for DadPres, YearsEd, Female, and Black, along with their tests of statistical significance.

For the DadPres variable, the expected (average) occupational prestige of a high school educated white man increases by 0.072 units when one's father's occupational prestige increases by one unit. This increase has a p-value that is statistically significant, so we reject the null hypothesis that father's occupational prestige has no effect on the respondents' occupational prestige. This suggests that the effect of the DadPres variable may have a profound effect on occupational prestige which is unlikely to be due to chance.

For the YearsEd variable, the expected (average) occupational prestige of a high school educated white man with an average father’s occupational prestige increases by 2.043 units when one's years of education (in addition to their high school education) increases by one year. This increase has a p-value that is statistically significant, so we reject the null hypothesis that years of education has no effect on occupational prestige. This suggests that the effect of the YearsEd variable may have a profound effect on occupational prestige which is unlikely to be due to chance.

For the Female variable, the mean difference between the expected average occupational prestige of high school educated white women and high school educated white men both with average father’s occupational prestige decreases by –0.297, indicating that white high school educated females may have a lower expected occupational prestige. However, this coefficient has a p-value that is not statistically significant, so we fail to reject the null hypothesis that the mean difference between these groups is 0. This suggests that the effect of the Female variable may not have a profound effect on occupational prestige.

For the Race – Black variable, the mean difference between expected average occupational prestige of high school educated black men and high school educated white men both with average father’s occupational prestige is –2.417, indicating that black men may have a lower expected occupational prestige. This coefficient has a p-value that is statistically significant, so we reject the null hypothesis that the mean difference between these groups is 0. This suggests that the effect of the Race – Black variable may have a profound effect on occupational prestige which is unlikely to be due to chance.

## Part C

### Did your coefficients change from the base model? Why or why not?

The coefficients did not change from the base model. This is because centering only shifts the data along the x-axis, maintaining the relative relationships between the variables. The intercept, however, did change from the base model to become a more meaningful representation of occupational prestige.

# QUESTION 3

### Modify the base model by standardizing your outcome variable of occupational prestige, your predictor variable of father’s prestige, and years of education.

gss$OccPresSd <- as.numeric(scale(gss$OccPres))  
  
gss$DadPresSd <- as.numeric(scale(gss$DadPres))  
  
gss$YearsEdSd <- as.numeric(scale(gss$YearsEd))   
  
model3 <- summ(lm(OccPresSd ~ DadPresSd + Female + Race + YearsEdSd, data=gss), digits = 3)   
  
model3

## MODEL INFO:  
## Observations: 1776 (572 missing obs. deleted)  
## Dependent Variable: OccPresSd  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(5,1770) = 105.855, p = 0.000  
## R² = 0.230  
## Adj. R² = 0.228   
##   
## Standard errors:OLS  
## ---------------------------------------------------  
## Est. S.E. t val. p  
## ----------------- -------- ------- -------- -------  
## (Intercept) 0.074 0.033 2.252 0.024  
## DadPresSd 0.069 0.022 3.151 0.002  
## Female -0.022 0.042 -0.518 0.604  
## RaceBlack -0.177 0.065 -2.726 0.006  
## RaceOther 0.045 0.067 0.664 0.507  
## YearsEdSd 0.445 0.021 20.800 0.000  
## ---------------------------------------------------

## Part A

### Interpret the coefficients for DadPres, YearsEd, and Female.

For the standardized regression Model 3, the interpretations of the coefficients (the standardized effect sizes) are as follows:

1. The coefficient for the standardized DadPres variable, DadPresSd, is 0.069. This is the expected change in OccPres (in standard deviation – SD – units) for a one-SD change in DadPresSd. This constitutes a “small” standardized effect size in the sociology discipline (Jean, Lecture Day 13, Slide 20).
2. The coefficient for the standardized YearsEd variable, YearsEdSd, is 0.445. This is the expected change in OccPres (in SD units) for a one-SD change in YearsEdSd. This constitutes a “large” standardized effect size in the sociology discipline (Jean, Lecture Day 13, Slide 20).
3. The coefficient for the Female variable is –0.022. This is the mean difference in OccPres (in SD units) between male and female respondents. More specifically, white female respondents have an occupational prestige score (in SD units) that is –0.022 lower than white male respondents. This constitutes a “small” standardized effect size in the sociology discipline (Jean, Lecture Day 13, Slide 20).

## Part B

### Explain the advantages of standardizing these variables in this particular context.

The variables that we standardized were the outcome variable of occupational prestige, the predictor variable of father’s prestige, and years of education. The units of “measurement” for prestige and years of education are different, which makes the comparison between these two difficult. For instance, we cannot say that an additional year of education has the same effect as one point increase in father’s prestige.

However, by translating our variables into standard-deviation units, we are putting them on the same scale, which allows us to compare the variables against each other. The size of coefficients for the standardized variables can be productively compared to each other because we are comparing the changes in SD units relative to the variance of our variables – with large variances, small differences are less meaningful and with small variances, large differences are less meaningful. Additionally, putting it in SD units can describe a big/small effect without knowing about the underlying distribution of the variables. Essentially, standardizing these variables in this particular context is advantageous because we are able to assess the relative effects of variables that are on different scales.

## Part C

### What do your results suggest about the relative influence of social origins (including race, gender, and father’s occupation) versus educational attainment on occupational prestige?

Consider the p-values of each of the variables – DadPresSd, RaceBlack, and YearsEdSd are all under 0.05, which means that these variables may have a profound effect on occupational prestige, and it is unlikely to be due to chance. On the other hand, Female and RaceOther variables have p-values that are above 0.05, so these variables are not significant and may not have as profound of an effect on occupational prestige.

Even if DadPresSd, RaceBlack, and YearsEdSd are all significant, they do not have the same effect. Consider the value (magnitude) of the coefficients. The (absolute) value of the coefficients increase in the following order: DadPresSd, RaceBlack, and YearsEdSd. Specifically for the standardized variables, a one SD unit increase in YearsEdSd has a greater effect on occupational prestige than DadPresSd. In other words, educational attainment (the years of education variable) has a greater positive influence on (expected) occupational prestige than the relative influence of social origin – specifically compared to father’s occupational prestige.

This also applies to RaceBlack vs YearsEdSd because the latter has a larger coefficient (more than twice), suggesting that the effect of educational attainment is greater than being in the Black racial category. (One might interpret the two variables as such: being in the Black racial category may decrease one’s occupational prestige, but educational attainment can overcome this reduction in occupational prestige associated with race.)

Thus, one may conclude that educational attainment has a greater impact (“matters more”) in influencing one’s occupational prestige than one’s relative influence of social origins (e.g., race, gender, father’s occupation).

# QUESTION 4

### We can speculate that the influence of parental SES (operationalized as father’s occupational prestige) may differ between men and women. Investigate whether this is true by adding an interaction term to the base model. Center DadPres around its mean and YearsEd at 12.

model4 <- summ(lm(OccPres ~ DadPresC\*Female + DadPresC + Female + Race + YearsEd12, data=gss), digits = 3)   
  
model4

## MODEL INFO:  
## Observations: 1776 (572 missing obs. deleted)  
## Dependent Variable: OccPres  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(6,1769) = 88.507, p = 0.000  
## R² = 0.231  
## Adj. R² = 0.228   
##   
## Standard errors:OLS  
## -------------------------------------------------------  
## Est. S.E. t val. p  
## --------------------- -------- ------- -------- -------  
## (Intercept) 42.174 0.489 86.185 0.000  
## DadPresC 0.042 0.033 1.256 0.209  
## Female -0.299 0.572 -0.523 0.601  
## RaceBlack -2.446 0.887 -2.758 0.006  
## RaceOther 0.571 0.916 0.623 0.533  
## YearsEd12 2.042 0.098 20.794 0.000  
## DadPresC:Female 0.056 0.044 1.261 0.207  
## -------------------------------------------------------

## Part A

### Interpret the coefficient for your new interaction term, including the hypothesis test for statistical significance. Also explain in plain language what this means.

The expected difference in average occupational prestige for the main effect of father’s occupational prestige between white women and men with a high school education is 0.056. In other words, for each unit increase in the father’s occupational prestige from its mean, being female increases occupational prestige by 0.056 units. However, this coefficient has a p-value that is not statistically significant, so we fail to reject the null hypothesis that the there is no interaction effect between DadPresC and Female. This suggests that the interaction term of DadPresC and Female may not have a profound effect on occupational prestige.

## Part B

### Calculate the predicted occupational prestige for a white man with twelve years of education whose father had a job with prestige 10 points higher than the mean.

The equation we used to make this prediction was:

ŷi = β0 + β1 DadPresCi + β2 Femalei + β3 Blacki + β4 Otheri + β5 YearsEd12i  + β6 DadPresCi \* Femalei

Since we have centered the DadPres variable around its mean, the DadPresCi variable (for a white man with a father who had a job with prestige 10 points higher than the mean) is 10. Using this, the coefficients identified above, and that we have a white man with 12 years of education (i.e., Female = 0, Black = 0, Other = 0, YearsEd = 0), we plug in these values into the equation:

ŷi = 42.174 + (0.042 \* DadPresC) + (-0.299 \* Female) + (-2.446 \* RaceBlack) + (0.571 \* RaceOther) + (2.042 \* YearsEd12) + (0.056 \* DadPresC \* Female)

ŷi = 42.174 + (0.042\*10) + (-0.299\*0) + (-2.446\*0) + (0.571\*0) + (2.042\*0) + (0.056\*10\*0)

## [1] 42.594

ŷi = 42.594

Thus, we expect the estimated occupational prestige to be a score of 42.594 for a white man with twelve years of education whose father had a job with prestige 10 points higher than the mean.

## Part C

### Calculate the predicted occupational prestige for a white woman with twelve years of education whose father had a job with prestige 10 points higher than the mean.

The equation we used to make this prediction was:

ŷi = β0 + β1 DadPresCi + β2 Femalei + β3 Blacki + β4 Otheri + β5 YearsEd12i  + β6 DadPresCi \* Femalei

Since we have centered the DadPres variable around its mean, the DadPresCi variable (for a white woman with a father who had a job with prestige 10 points higher than the mean) is 10. Using this, the coefficients identified above, and that we have a white woman with 12 years of education (i.e., Female = 1, Black = 0, Other = 0, YearsEd = 0), we plug in these values into the equation:

ŷi = 42.174 + (0.042 \* DadPresC) + (-0.299 \* Female) + (-2.446 \* RaceBlack) + (0.571 \* RaceOther) + (2.042 \* YearsEd12) + (0.056 \* DadPresC \* Female)

ŷi = 42.174 + (0.042\*10) + (-0.299\*1) + (-2.446\*0) + (0.571\*0) + (2.042\*0) + (0.056\*10\*1)

## [1] 42.855

ŷi = 42.855

Thus, we expect the estimated occupational prestige to be a score of 42.855 for a white woman with twelve years of education whose father had a job with prestige 10 points higher than the mean.

# QUESTION 5

### We learned in the previous assignment that educational attainment is an important mediating factor in the relationship between father’s and children’s occupational prestige. Perhaps this mediating relationship varies across racial groups? Investigate this possibility by modifying the base model to include the interaction between years of education and race. Center DadPres around its mean and YearsEd at 12.

model5 <- summ(lm(OccPres ~ DadPresC + Female + Race + YearsEd12 + YearsEd12\*Race, data=gss), digits = 3)   
  
model5

## MODEL INFO:  
## Observations: 1776 (572 missing obs. deleted)  
## Dependent Variable: OccPres  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(7,1768) = 76.306, p = 0.000  
## R² = 0.232  
## Adj. R² = 0.229   
##   
## Standard errors:OLS  
## -----------------------------------------------------------  
## Est. S.E. t val. p  
## ------------------------- -------- ------- -------- -------  
## (Intercept) 41.923 0.508 82.478 0.000  
## DadPresC 0.072 0.023 3.126 0.002  
## Female -0.288 0.572 -0.504 0.614  
## RaceBlack -2.340 1.043 -2.244 0.025  
## RaceOther 1.184 0.960 1.233 0.218  
## YearsEd12 2.151 0.119 18.066 0.000  
## RaceBlack:YearsEd12 -0.021 0.315 -0.067 0.947  
## RaceOther:YearsEd12 -0.467 0.231 -2.027 0.043  
## -----------------------------------------------------------

## Part A

### Interpret the coefficients of the main effects of race and education, and the interaction terms, including the hypothesis tests for statistical significance.

The coefficient of the main effect of the Race – Black category is –2.340. This is the mean difference (in occupational prestige) between Black and White men whose fathers held a job with average prestige and 12 years of education (i.e., high school education). In other words, being in the Race – Black category lowers one’s occupational prestige by –2.340. This coefficient is significant because its p-value is less than 0.05, so we reject the null hypothesis (which is that the mean difference between Black and White men whose fathers held a job with average prestige and 12 years of education is 0). This suggests that the effect of Race – Black may have a profound effect on occupational prestige, and it is unlikely to be due to chance.

The coefficient of the main effect of Race – Other category is 1.184. This is the mean difference (in occupational prestige) between men in the “Other” and White race categories whose fathers held a job with average prestige and 12 years of education (i.e., high school education). In other words, being in the “Other” race category increases one’s occupational prestige by 1.184. However, this coefficient is not significant because its p-value is greater than 0.05, so we fail to reject the null hypothesis (which is that the mean difference between men in the “Other” and White race categories whose fathers held a job with average prestige and 12 years of education is 0). This suggests that the effect of being in the “Other” race category may not have a profound effect on occupational prestige.

The coefficient of the main effect of Years of Education 2.151. This is the expected change in occupational prestige for a one-year increase in years of education for White men with 12 years of education and a father who held a job with average prestige. In other words, for every additional year of education (in addition to 12 years), occupational prestige is expected to increase by 2.151. This coefficient is significant because its p-value is less than 0.05, so we reject the null hypothesis (which is that the Years of Education has no effect on the respondents’ occupational prestige). This suggests that the years of education may have a profound effect on occupational prestige, and it is unlikely to be due to chance.

The coefficient for the interaction term Race – Black and Years of Education is –0.021. This is the expected difference (in occupational prestige) in the main effect of years of education between Race categories Black and White (for men with 12 years of education & father who held a job with average prestige). In other words, for every additional year of education (in addition to 12 years), being in the Black racial category lowers occupational prestige by –0.021. However, this coefficient is not significant because its p-value is greater than 0.05, so we fail to reject the null hypothesis (which is that there is no interaction effect between Years of Education and Race – Black). This suggests that the interaction term of Race – Black and Years of Education may not have a profound effect on occupational prestige.

The coefficient for the interaction term Race – Other and Years of Education is –0.467. This is the expected difference (in occupational prestige) in the main effect of years of education between Race categories “Other” and White (for men with 12 years of education & father who held a job with average prestige). In other words, for every additional year of education (in addition to 12 years), being in the “Other” racial category lowers occupational prestige by –0.467. This coefficient is significant because its p-value is less than 0.5, so we reject the null hypothesis (which is that there is no interaction effect between Years of Education and Race – Other). This suggests that the interaction term of Race – Other and Years of Education may have a profound effect on occupational prestige (unlikely to be due to chance).

## Part B

### Calculate the expected change in occupational prestige for each additional year of education for white respondents, black respondents, and those of other races, holding other variables in the model constant.

The equation we used for all three race categories:

ŷi = β0 + β1 DadPresC i + β2 Female i + β3 Black i + β4 Other i + β5 YearsEd12 i  + β6 YearsEd12i \* Race i

The equation we used for all three race categories, with coefficients substituted:

ŷi = 41.923 + (0.072 \* DadPresC) + (–0.288 \* Female) + (–2.340 \* RaceBlack) + (1.184 \* RaceOther) + (2.151 \* YearsEd12) + (–0.021 \* RaceBlack \* YearsEd12) + (–0.467 \* RaceOther \* YearsEd12)

1. To find the expected change in occupational prestige for *each additional year of education* for *White* respondents, we substituted 1 for YearsEd12, 0 for RaceBlack & RaceOther, and kept the rest “constant” (i.e., leave variables as it is). Thus, for White respondents, we have:

ŷi = 41.923 + (0.072 \* DadPresC) + (-0.288\*Female) + (-2.340\*0) + (1.184\*0) + (2.151\*1) + (-0.021\*0\*1) + (-0.467\*0\*1)

Calculating the numerical portion of our equation in R:

(-2.340\*0) + (1.184\*0) + (2.151\*1) + (-0.021\*0\*1) + (-0.467\*0\*1)

## [1] 2.151

ŷi = 41.923 + (0.072 \* DadPresC) + (-0.288\*Female) + 2.151

From this, we find that for each additional year of education that a White respondent receives, the occupational prestige is expected to change (increase) by 2.151 points.

1. To find the expected change in occupational prestige for *each additional year of education* for *Black* respondents, we substituted 1 for YearsEd12, 1 for RaceBlack, and 0 for RaceOther, and kept the rest “constant” (i.e., leave variables as it is). Thus, for Black respondents, we have:

ŷi = 41.923 + (0.072 \* DadPresC) + (-0.288\*Female) + (-2.340\*1) + (1.184\*0) + (2.151\*1) + (-0.021\*1\*1) + (-0.467\*0\*1)

Calculating the numerical portion of our equation in R:

(-2.340\*1) + (1.184\*0) + (2.151\*1) + (-0.021\*1\*1) + (-0.467\*0\*1)

## [1] -0.21

ŷi = 41.923 + (0.072 \* DadPresC) + (-0.288\*Female) + (- 0.21)

From this, we find that for each additional year of education that a Black respondent receives, the occupational prestige is expected to change (decrease) by –0.21 points.

1. To find the expected change in occupational prestige for *each additional year of education* forrespondents in the *Other* race category, we substituted 1 for YearsEd12, 0 for RaceBlack, and 1 for RaceOther, and kept the rest “constant” (i.e., leave variables as it is). Thus, for those in the *Other* race category, we have:

ŷi = 41.923 + (0.072 \* DadPresC) + (-0.288\*Female) + (-2.340\*0) + (1.184\*1) + (2.151\*1) + (-0.021\*0\*1) + (-0.467\*1\*1)

Calculating the numerical portion of our equation in R:

(-2.340\*0) + (1.184\*1) + (2.151\*1) + (-0.021\*0\*1) + (-0.467\*1\*1)

## [1] 2.868

ŷi = 41.923 + (0.072 \* DadPresC) + (-0.288\*Female) + 2.868

From this, we find that for each additional year of education that a respondent in the Other race category receives, the occupational prestige is expected to change (increase) by 2.868 points.

## Part C

### Calculate the estimated occupational prestige for a woman in the “Other” racial category with 17 years of education whose father held a job with average prestige.

The equation we used to make this prediction was:

ŷi = β0 + β1 DadPresC i + β2 Female i + β3 Black i + β4 Other i + β5 YearsEd12 i  + β6 YearsEd12i \* Race i

Since we have centered the DadPres variable around its mean, the DadPresCi variable (for a woman with a father who held a job with average prestige) is zero. Using this, the coefficients identified above, and that we have a woman in the “Other” racial category with 17 years of education (i.e., Female = 1, Black = 0, Other = 1, YearsEd = 17 but with YearsEd12 = 17-12 = 5), we plug in these values into the equation:

ŷi = 41.923 + (0.072 \* DadPresC) + (–0.288 \* Female) + (–2.340 \* RaceBlack) + (1.184 \* RaceOther) + (2.151 \* YearsEd12) + (–0.021 \* RaceBlack \* YearsEd12) + (–0.467 \* RaceOther \* YearsEd12)

ŷi = 41.923 + (0.072\*0) + (-0.288\*1) + (-2.340\*0) + (1.184\*1) + (2.151\*5) + (-0.021\*0\*5) + (-0.467\*1\*5)

## [1] 51.239

ŷi = 51.239

Thus, we expect the estimated occupational prestige to be a score of 51.239 for a woman in the “Other” racial category with 17 years of education whose father held a job with average prestige.

## Part D

### Explain in plain language what this model indicates about how race and education affect occupational prestige.

For the RaceBlack and YearsEd12 interaction term, one main effect is positive (YearsEd12, with a coefficient of 2.151) and the other is negative (RaceBlack, with a coefficient of –2.340). The interaction term has a negative interaction effect (the coefficient of RaceBlack & YearsEd12 is   –0.021). This means that the effects of the positive main effect will be weakened and the effects of the impact of the negative main effect will be strengthened. Increasing both variables together results in a more negative outcome than increasing each one individually. In other words, since the interaction term has a negative sign, increasing one’s years of education will have less of an impact on occupational prestige, while being in the Black racial category will have more of an impact on occupational prestige. If one “increases” both variables together (i.e., if you have more years of education & are Black), it will reduce the occupational prestige more than increasing each of the variables individually (i.e., if you aren’t Black and have more education OR if you are Black and do not increase the years of education). However, the p-value of the interaction term is greater than 0.05, which suggests that the interaction term of RaceBlack & YearsEd may not have a profound effect on occupational prestige (so the effect/relationship described above may not occur).

For the RaceOther and YearsEd12 interaction term, both main effects are positive. The interaction term has a negative interaction effect (the coefficient of RaceOther & YearsEd12 is   –0.467). This means that as you increase one or both variables, the predicted changes to the outcome grow weaker and there are diminishing returns. In other words, if one has more years of education and/or is in the “Other” race category, the predicted changes to one’s occupational prestige grows weaker and occupational prestige will decrease – “diminish return” (because we have a negative interaction effect). However, it is important to note that Race – Other has a p-value that is greater than 0.05. This means that this coefficient may not be significant, and its effects (on occupational prestige) may be due to chance. The main effect YearsEd12 and the interaction term both have significant p-values, so these variables may have a profound effect on the outcome variable.

# QUESTION 6

### Create a visualization drawing from examples from class that depicts the mean occupational prestige of the three racial groups in the GSS sample. Title this visualization: “Raw differences in occupational prestige by race”

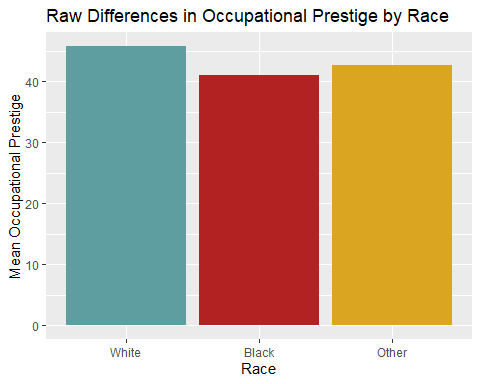
Calculate the mean of occupational prestige by race:

race\_occpres <- gss %>%   
 group\_by(Race) %>%   
 summarise(OccPresMean = mean(OccPres, na.rm = TRUE))   
race\_occpres

## # A tibble: 3 × 2  
## Race OccPresMean  
## <fct> <dbl>  
## 1 White 45.8  
## 2 Black 41.0  
## 3 Other 42.7

Create a visualization to compare occupational prestige by race.

ggplot(race\_occpres, aes(   
 x = Race,   
 y = OccPresMean,   
 fill = Race)) +   
 geom\_col(fill=c("cadetblue", "firebrick", "goldenrod")) +   
 labs(   
 title = "Raw Differences in Occupational Prestige by Race",   
 x = "Race",   
 y = "Mean Occupational Prestige")



# QUESTION 7

### Create a visualization drawing from examples from class that depicts the estimated mean occupational prestige of these three racial groups – controlling for gender, and typical values of education and father’s occupational prestige. Use the results from your model in Part 2 to calculate these estimates. [Hint: It is likely simplest to perform these calculations manually, and then feed these estimates into R] Title this visualization: “Adjusted differences in occupational prestige by race”

We first performed calculations manually using the model in Part 2. The model was:

model2 <- summ(lm(OccPres ~ DadPresC + FemaleC + BlackC + OtherC + YearsEd12, data=gss), digits = 3)

To control for gender, we used Female = 0. We used the typical value of education as 12 years, and typical father’s occupational prestige as the average. Model 2 centered the YearsEd variable at 12 years and centered DadPres around the mean, so both these will be zero when we calculate our model. RaceBlack and RaceOther was substituted (either 0 or 1) depending on the race we were calculating for. Below are our calculations for each race category.

white <- 41.664 + (0.072\*0) + (-0.297\*0) + (-2.417\*0) + (0.608\*0) + (2.043\*0)   
white

## [1] 41.664

black <- 41.664 + (0.072\*0) + (-0.297\*0) + (-2.417\*1) + (0.608\*0) + (2.043\*0)   
black

## [1] 39.247

other <- 41.664 + (0.072\*0) + (-0.297\*0) + (-2.417\*0) + (0.608\*1) + (2.043\*0)   
other

## [1] 42.272

Using these calculations, we create the following visualization.

adjusted\_race\_occpres <- data.frame( Race = c("White", "Black", "Other"),   
 OccPres = c(41.664, 39.247, 42.272))   
adjusted\_race\_occpres$Race <- factor(adjusted\_race\_occpres$Race, levels = c("White", "Black", "Other"))

ggplot(adjusted\_race\_occpres, aes(x = Race, y = OccPres, fill = Race)) +   
 geom\_bar(stat = "identity", fill=c("cadetblue", "firebrick", "goldenrod")) +   
 labs(title = "Adjusted Differences in Occupational Prestige by Race",   
 x = "Race",   
 y = "Adjusted Occupational Prestige")

