Assignment 1

Intermediate Regression and Data Science

Sophie Ennis and Elina Li

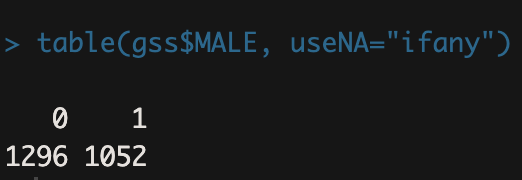
April 15, 2025

**Question 1**

gss$MALE <- ifelse(gss$sex==1, 1, 0)

describe(gss$MALE)

table(gss$MALE, useNA="ifany")

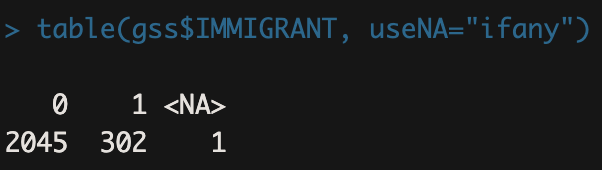


gss$IMMIGRANT[gss$born < 1 | gss$born > 2] <- NA

gss$IMMIGRANT <- ifelse(gss$born==2, 1, 0)

describe(gss$IMMIGRANT)

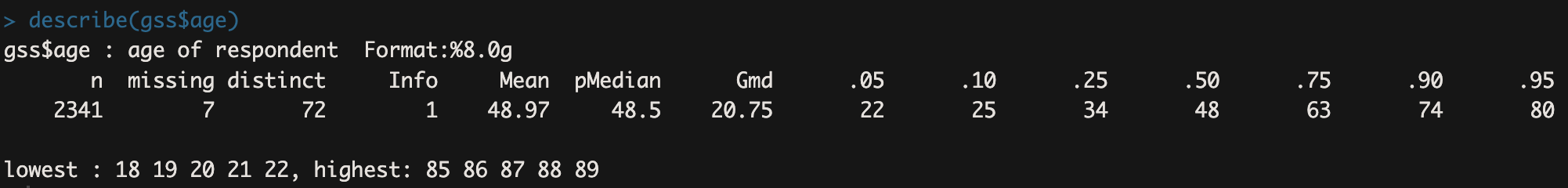
table(gss$IMMIGRANT, useNA="ifany")



gss$AGE <- gss$age

gss$AGE[gss$age==9] <- NA

describe(gss$AGE)



**Question 2**

col\_lst <- c("racecen1", "racecen2", "racecen3") *#creates a vector called col\_lst. The vector contains the names of three columns from your gss dataset: "racecen1", "racecen2", and "racecen3"*

gss[, colnames(gss) %in% col\_lst] *# just viewing the 3 columns*

gss <- gss %>%

mutate(

multiracial = !is.na(racecen2) | !is.na(racecen3),

RACE = case\_when(

multiracial ~ "Multiracial",

racecen1 == 1 ~ "White",

racecen1 == 2 ~ "Black",

racecen1 %in% 5:10 ~ "Asian",

racecen1 == 16 ~ "Hispanic",

racecen1 %in% c(3:4, 11:15) ~ "Other",

TRUE ~ NA\_character\_),

RACE = factor(RACE, levels = c("White", "Black", "Asian", "Hispanic", "Other", "Multiracial")))

table(gss$RACE, useNA = "ifany")



The primary advantage of consolidating racial categories in this way is that consolidating sparse categories improves both interpretability and precision. With sparse categories, it is difficult to make precise estimates and useful inferences because, with categorical variables like race, the smaller the number of observations in a racial category, the larger the standard errors will be.

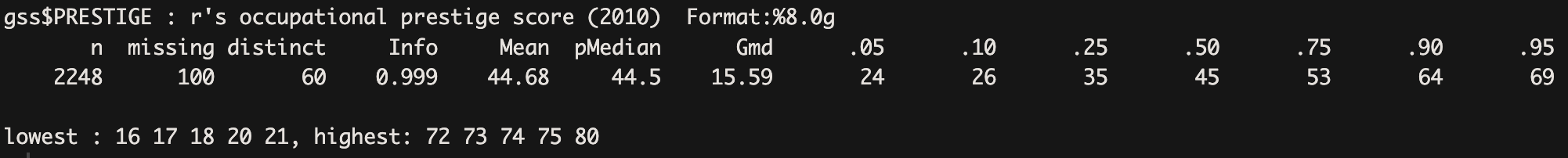
The primary disadvantage of consolidating racial categories in this way is that we lose nuance in the differences between the groups that have been consolidated. For example, within the Asian racial category, we lose out on potentially meaningful differences between Korean and Japanese individuals. Also, placing all “Native” or small groups into “Other” may marginalize those identities and mask inequalities or trends unique to them. Additionally, coding multiracial respondents into one category doesn’t reflect the complexity of racial identity or how society perceives them (e.g., someone Black and Asian may experience society differently than someone White and Asian).

**Question 3**

gss$PRESTIGE <- gss$prestg10

gss$PRESTIGE[gss$prestg10==0] <- NA

table(gss$PRESTIGE, useNA = "ifany")



**Question 4**

gss$CLASSORIGIN <- as.factor(case\_when(gss$incom16 == 1 ~ "FAR BELOW AVERAGE",

gss$incom16 == 2 ~ "BELOW AVERAGE",

gss$incom16 == 3 ~ "AVERAGE",

gss$incom16 == 4 ~ "ABOVE AVERAGE",

gss$incom16 == 5 ~ "FAR ABOVE AVERAGE",

gss$incom16 == 7 ~ "LIVED IN INSTITUTION"))

gss$CLASSORIGIN <- factor(gss$CLASSORIGIN, levels = c("AVERAGE", "FAR BELOW AVERAGE", "BELOW AVERAGE", "ABOVE AVERAGE", "FAR ABOVE AVERAGE", "LIVED IN INSTITUTION"))

gss$CLASSORIGIN[gss$incom16 == c(0, 8, 9)] <- NA

table(gss$CLASSORIGIN, useNA = "ifany")



We chose to code the “CLASSORIGIN” variable as a categorical variable because the numbers in the codebook are not meaningful, and are labeled with strings rather than numerics. For instance, the punch “1” corresponds with “Far Below Average” rather than a measurable income level. Coding this variable as a categorical variable allots more explanatory power, and thus a higher R² (0.03 when treated as categorical versus 0.02 when treated as continuous), in the model.

Setting "Average" as the reference group creates a neutral comparison point that allows us to test the core question of whether individuals from below- or above-average origins achieve different levels of occupational prestige. It lets us evaluate if upward mobility is possible, and how much class origin still matters in adulthood.

**Question 5**

gss$YEARSED <- as.factor(case\_when(gss$educ == 0 ~ "NO FORMAL SCHOOLING",

gss$educ %in% c(1:5) ~ "PRIMARY SCHOOL",

gss$educ %in% c(6:8) ~ "MIDDLE SCHOOL",

gss$educ %in% c(9:12) ~ "HIGH SCHOOL",

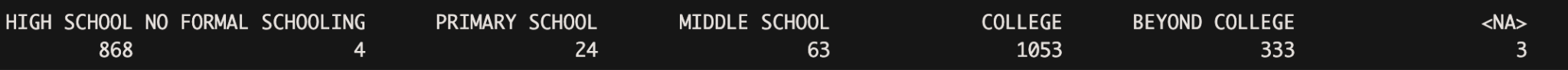
gss$educ %in% c(13:16) ~ "COLLEGE",

gss$educ %in% c(17:20) ~ "BEYOND COLLEGE"))

gss$YEARSED <- factor(gss$YEARSED, levels = c("HIGH SCHOOL", "NO FORMAL SCHOOLING", "PRIMARY SCHOOL", "MIDDLE SCHOOL", “COLLEGE”, "BEYOND COLLEGE"))

gss$YEARSED[gss$educ == c(98, 99)] <- NA

table(gss$YEARSED, useNA = "ifany")



We chose to code the variable “YEARSED” as a categorical variable because there are meaningful checkpoints throughout the educational experience as the years of education increase. The differences between primary and middle school are stark, while the differences between 1st and 2nd grade may not be as profound.

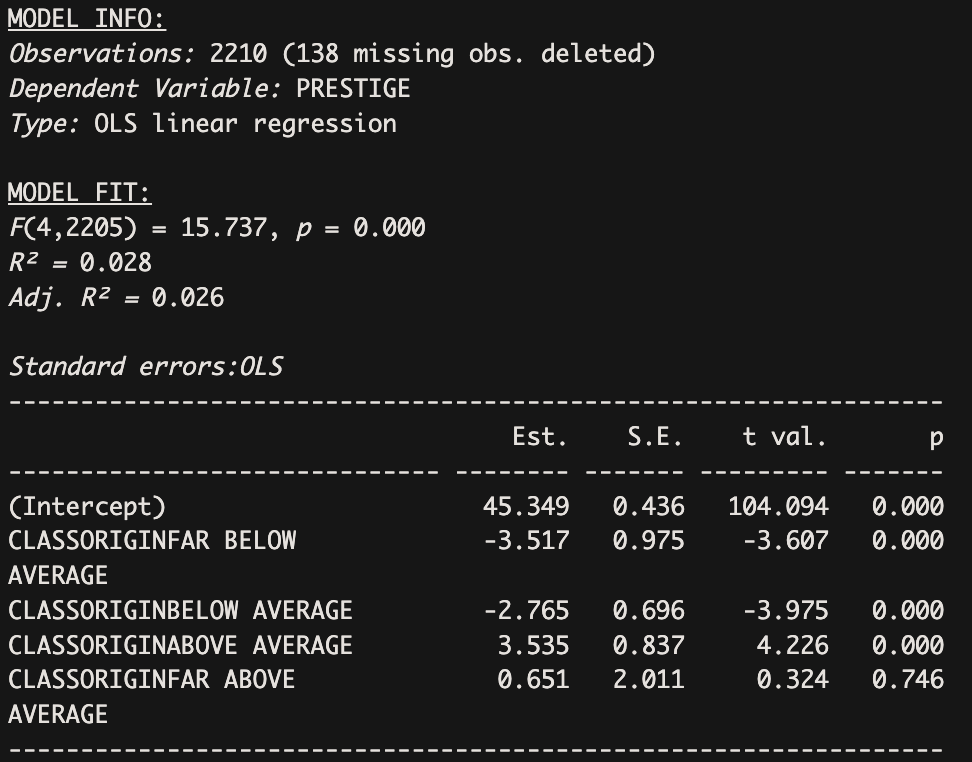
Since we are examining how formal schooling mediates social mobility, we set “High School” as the reference group. High school serves as a compelling reference group in analyzing educational effects on social mobility because it represents a widely attained and socially recognized baseline of formal education in the United States. As the culmination of compulsory schooling, high school completion marks the transition between basic education and post-secondary pathways, making it a meaningful dividing line between those who pursue higher education and those who do not. Using high school as the reference group allows researchers to assess both the advantages of obtaining education beyond high school (e.g., college or graduate school) and the disadvantages of falling short (e.g., only completing middle or primary school).

We also found out later that if we set “YEARSED” as a continuous variable, the R² is slightly higher than treating it as a categorical variable (0.23 versus 0.21) However, if our goal is to evaluate how educational thresholds relate to prestige (e.g., whether college or grad school offers a meaningful “jump”), the categorical model is more appropriate, even if its R² is slightly lower.

**Question 6**

model1 <- lm(PRESTIGE ~ CLASSORIGIN, data = gss)

jtools::summ(model1, digits = 3)



This model examines the association between respondents’ class origin and their current occupational prestige score, using “CLASSORIGIN” as a categorical independent variable with "Average" as the reference group. The F-Statistic of 15.737, p < 0.001 indicates that the overall model is statistically significant (that is, independent variable class origin categories significantly predicts the dependent variable occupational prestige). R² = 0.028 means that only 2.8% of the variance in occupational prestige is explained by class origin alone. This is a very low amount of explained variance, suggesting that other factors are likely also important in shaping occupational prestige but are not included in this model. Since the adjusted R² (0.026) is almost the same as R², it means there’s not much of a penalty for overfitting–the small set of predictors does contribute meaningfully, even if only modestly.

The intercept is 45.349, representing the predicted prestige score for someone from an “average” class origin. This is the baseline group to which all others are compared.

Those who come from a "Far Below Average” class origin are predicted to have a 3.517-point lower prestige score on average compared to those from an average class origin. This result is highly statistically significant (p < 0.001), suggesting a meaningful disadvantage for individuals from the lowest class backgrounds.

Similarly, those from "Below Average" origins show a 2.765-point decrease in prestige score compared to the average group. This result is also highly statistically significant (p < 0.001), reinforcing the idea that lower class origins are associated with reduced occupational status in adulthood.

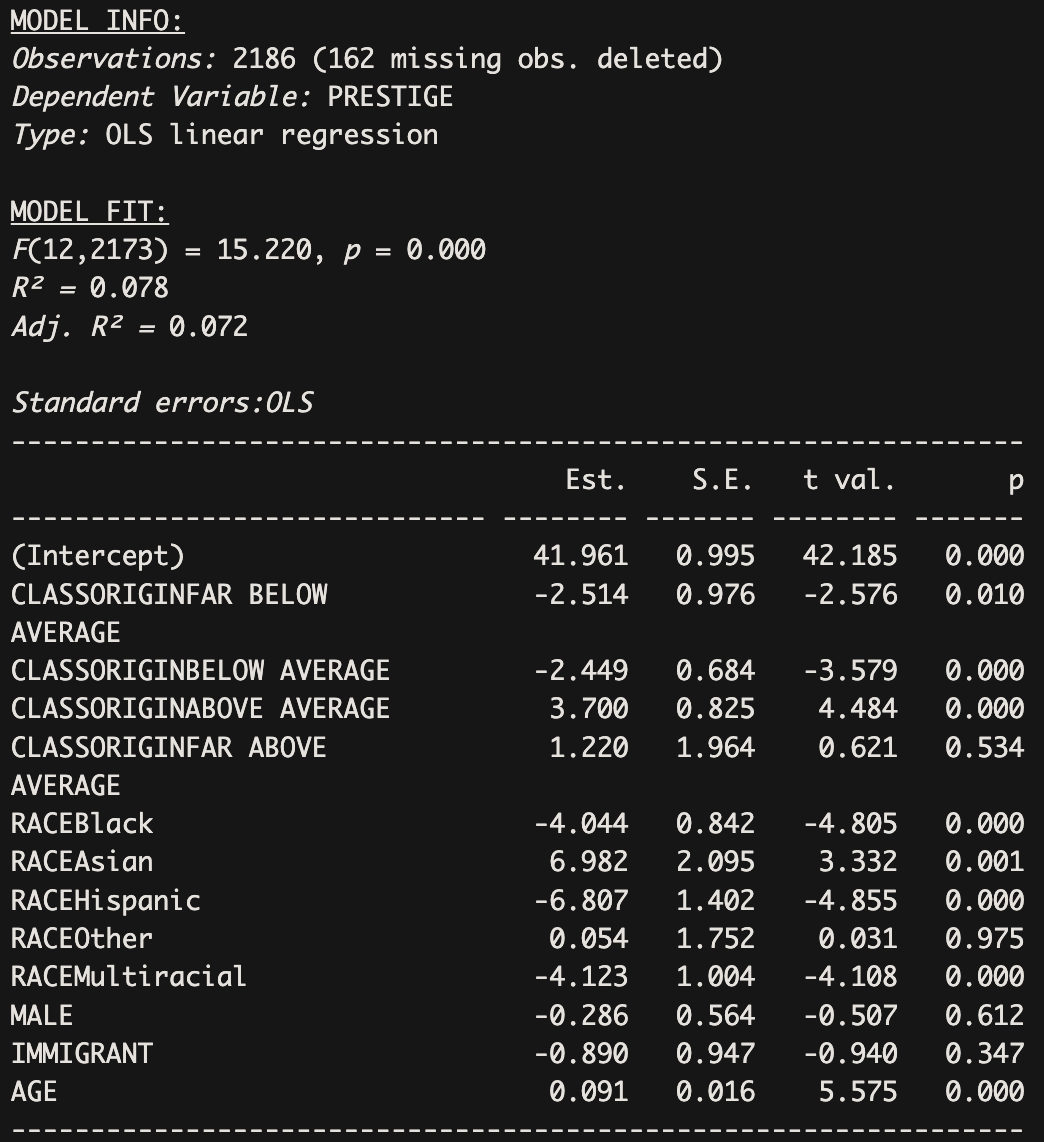
On the other hand, individuals from an "Above Average" class origin experience a 3.535-point increase in prestige scores compared to those from average backgrounds. This result is highly statistically significant (p < 0.001), indicating a meaningful advantage for those from more privileged backgrounds.

Interestingly, the coefficient for "Far Above Average" is 0.651, but this result is not statistically significant (p = 0.746). This suggests that beyond a certain point, additional class advantage does not significantly increase occupational prestige compared to the average group, at least in this model.

**Question 7**

model2 <- lm(PRESTIGE ~ CLASSORIGIN + RACE + MALE + IMMIGRANT + AGE, data = gss)

jtools::summ(model2, digits = 3)



This regression output models how various demographic characteristics, including class origin, race, gender, immigrant status, and age, are associated with economic success in adulthood, measured by prestige scores. The intercept of 41.961 represents the predicted prestige score for the reference group: non-immigrant White females whose class origin is classified as “Average.” All other coefficients represent deviations from this baseline group. The F-Statistic of 15.22, p < 0.001 indicates that at least one independent variable in the model significantly contributes to explaining variation in the dependent variable (PRESTIGE). In other words, the model as a whole fits the data better than an intercept-only model, and the probability that this fit occurred by chance is extremely low. R² = 0.078 means that only 7.8% of the variance in occupational prestige is explained by the included variables: class origin, race, gender, immigrant status, and age. This is a relatively low level of explanatory power, suggesting that important variables influencing occupational prestige are missing from this model. The adjusted R² (0.072) penalizes the model for adding independent variables that don’t improve the model fit substantially. The fact that it's close to the raw R² suggests that while the independent variables contribute to explaining prestige to a small degree, they don't overfit, but their combined explanatory power is limited.

Class origin plays a notable role. Compared to the reference group, individuals from “Far Below Average” and “Below Average” class origins have significantly lower predicted prestige scores than the reference group, by 2.514 and 2.449 points respectively, and both effects are statistically significant, holding other variables constant (respectively, p = 0.010, very significant; p < 0.001, highly significant). This indicates that coming from a disadvantaged class background is associated with lower occupational prestige in adulthood. In contrast, individuals from “Above Average” class origins have significantly higher prestige scores (+3.70) than the reference group, holding other variables constant, suggesting that upward class origin is linked to greater economic success (with p < 0.001). Interestingly, those from a “Far Above Average” background have a positive but statistically insignificant association (+1.22, p = 0.534), holding other variables constant, which may indicate that coming from a very privileged background does not always guarantee higher prestige. However, this coefficient is not statistically significant, so we can’t say this for sure.

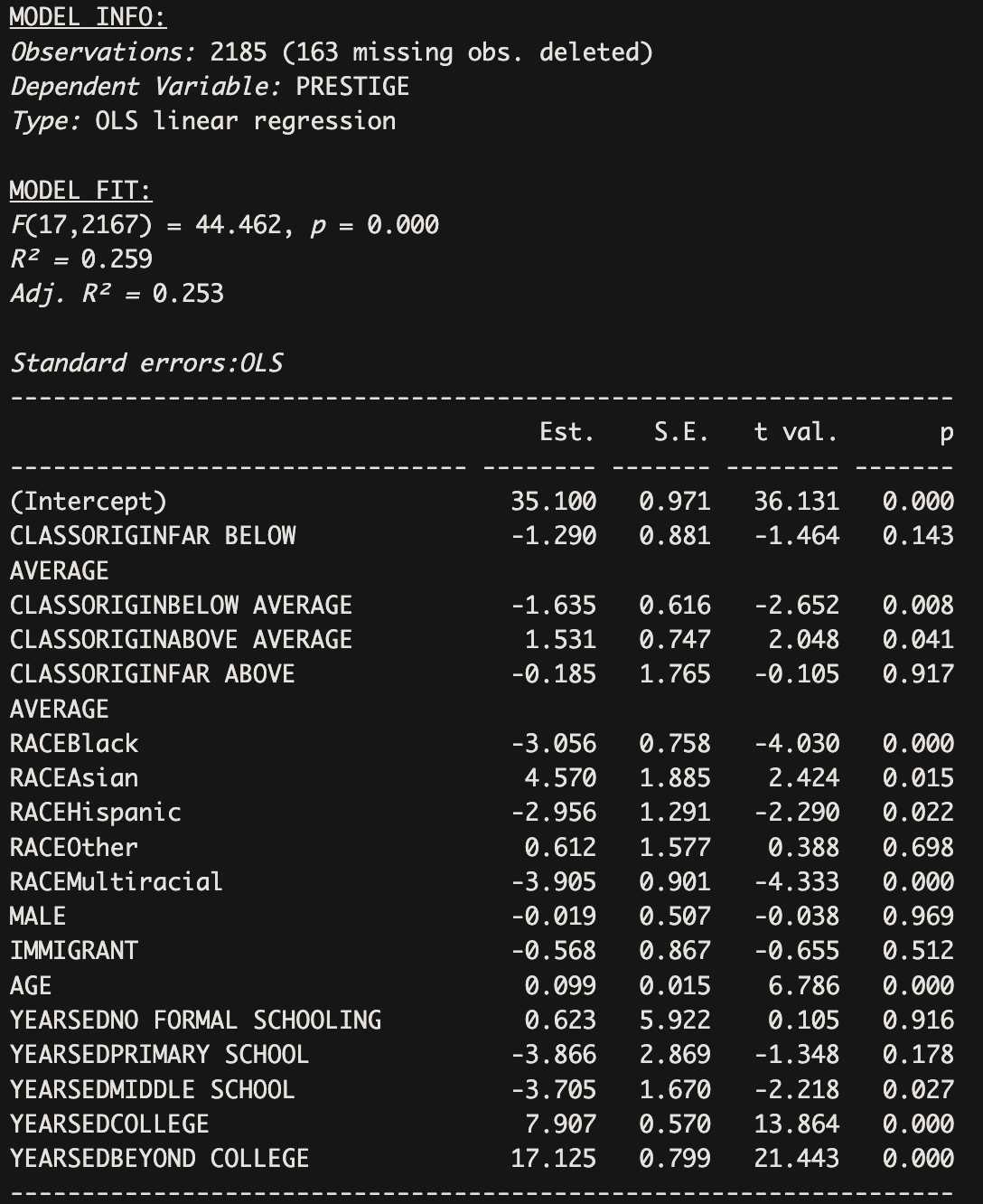
Race has a significant impact as well. Compared to White individuals in the reference group, Black, Hispanic, and Multiracial respondents have significantly lower prestige scores, with decreases of 4.044, 6.807, and 4.123 points respectively, holding other variables constant. These differences are statistically significant (p < 0.001 for each of them) and highlight racial disparities in occupational outcomes. Conversely, Asian respondents show a very significantly higher prestige score (+6.982, p < 0.01) than the reference group, holding other variables constant, suggesting greater economic success on average, while the "Other" racial group shows no significant difference.

The remaining demographic characteristics—gender, immigrant status, and age—show varying effects. Being male is not significantly associated with prestige score (-0.286, p = 0.612), nor is being an immigrant (-0.89, p = 0.347), suggesting these characteristics do not substantially shape occupational prestige in this model. However, age has a small but highly statistically significant positive association with prestige (+0.091 per year, p < 0.001), holding other variables constant, indicating that older individuals tend to have slightly higher prestige scores, perhaps reflecting accumulated experience or seniority.

**Question 8**

model3 <- lm(PRESTIGE ~ CLASSORIGIN + RACE + MALE + IMMIGRANT + AGE + YEARSED, data = gss)

jtools::summ(model3, digits = 3)



Compared to Model 2, the model fit of Model 3 improves substantially, with R² increasing from 0.08 in the second model to 0.26 in the third. This suggests that education explains a considerable portion of the variance in occupational prestige and is a strong independent variable in itself. The F-statistic of 44.462 (p < 0.001) indicates that Model 3, as a whole, is highly statistically significant—meaning the set of independent variables included significantly explains variance in occupational prestige.

In terms of baseline interpretation, the intercept of 35.10 now represents the predicted prestige score for non-immigrant White females, from an “Average” class origin, who completed high school.

The coefficient for the class origin variable “Far Below Average” increases to -1.29, becoming less negative than in the previous model, but also becoming insignificant in this model (p = 0.143). The coefficient for the class origin variable “Below Average” increases to -1.635, becoming less negative than in the previous model, while remaining very significant (p < 0.01). The coefficient for the class origin variable “Above Average” decreases from 3.70 in the previous model to 1.53 in this model while remaining significant (p = 0.041 < 0.05). The “Far Above Average” class origin coefficient decreases, even flipping signs, but remains insignificant (p = 0.917). Taken together, these changes suggest that once education is accounted for, the direct impact of class origin on occupational prestige is notably reduced. The weakening of both negative and positive coefficients implies that educational attainment mediates much of the association between class background and economic outcomes. In other words, while individuals from more privileged class origins still tend to achieve higher levels of education, which in turn boosts their occupational prestige, the disadvantages of lower class origins are partially mitigated through the same pathway.

The coefficients for the racial category variables Black, Hispanic, and Multiracial all increase, becoming less negative. For instance, the coefficient for Black is now -3.056 (as compared to -4.044 in the previous model), the coefficient for Hispanic is now -2.956 (as compared to -6.807 in the previous model, showing the most drastic change among all of the variables), and the coefficient for Multiracial is now -3.905 (as compared to -4.123 in the previous model). These coefficients all remain significant (respectively, p < 0.001, highly significant; p < 0.05, very significant; p < 0.001, highly significant), holding other variables constant. On the other hand, the coefficient for the Asian racial category decreases, and is now 4.57 as compared to being 6.982 in the previous model, nonetheless remaining significant. The Other racial category does increase minimally, but this coefficient remains insignificant (p = 0.698). These shifts suggest that educational attainment may partially mediate the relationship between race and occupational prestige, particularly for Hispanic individuals, whose disadvantage is substantially reduced when education is taken into account. However, because the coefficients for Black, Hispanic, and Multiracial groups remain significantly negative even after controlling for education, it is clear that racial disparities in occupational prestige persist beyond differences in educational attainment. This points to the continued influence of structural inequalities and racial discrimination in shaping economic outcomes. Conversely, the reduced but still significant advantage for Asian individuals suggests that their higher prestige scores are also partially explained by educational attainment, but not entirely.

The Male and Immigrant variables remain insignificant, although their coefficients both become less negative (insignificantly so). The coefficient for the Age variable increases from 0.091 to 0.099, remaining significant (p < 0.001), indicating that for every year older individuals get, their occupational prestige score increases by 0.099 when holding all other variables in the model constant.

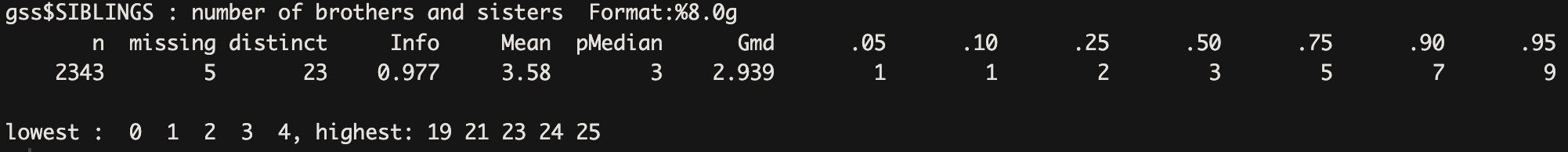
**Question 9**

*#additional variable 1: SIBS (How many brothers and sisters did you have? Please count those born alive, but no longer living, as well as those alive now. Also include stepbrothers and stepsisters, and children adopted by your parents.) We believe that family size can impact the distribution of economic and social resources (e.g., educational investment, inheritance), thereby determining their social class background. We set it as a continuous variable.*

gss$SIBLINGS <- gss$sibs

gss$SIBLINGS[gss$sibs == c(98, 99)] <- NA

describe(gss$SIBLINGS)



*#additional variable 2: RES16 (Which of the categories on this card comes closest to the type of place you were living in when you were 16 years old?) Growing up in an urban, suburban, or rural setting can strongly influence access to quality education, job networks, and cultural capital. These factors all contribute to shaping occupational outcomes in adulthood.*

gss$RES <- as.factor(case\_when(gss$res16 == 1 ~ "OPEN COUNTRY",

gss$res16 == 2 ~ "FARM",

gss$res16 == 3 ~ "SMALL CITY/TOWN",

gss$res16 == 4 ~ "MEDIUM CITY",

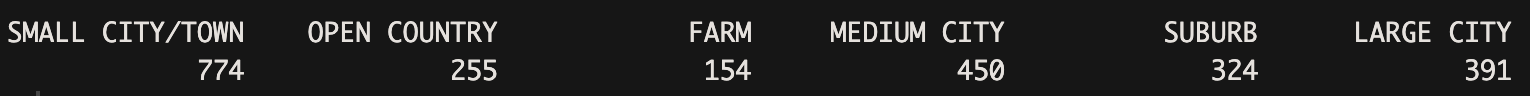
gss$res16 == 5 ~ "SUBURB",

gss$res16 == 6 ~ "LARGE CITY"))

gss$RES <- factor(gss$RES, levels = c("SMALL CITY/TOWN", "OPEN COUNTRY", "FARM", "MEDIUM CITY", "SUBURB", "LARGE CITY"))

gss$RES[gss$res16 == c(8, 9, 0)] <- NA

table(gss$RES, useNA = "ifany")



We set "Small City/Town" as our reference group since we are interested in comparing across the rural–urban gradient and understanding how growing up in different residential contexts affects adult prestige.

*#additional variable 3: PARBORN (Were both your parents born in this country?) This matters because having immigrant parents may limit early access to social, economic, and cultural resources, shaping one’s class background and later occupational prestige.*

gss$PARENTBORN <- as.factor(case\_when(gss$parborn == 0 ~ "BOTH BORN IN US",

gss$parborn %in% c(1,2) ~ "ONE BORN IN US",

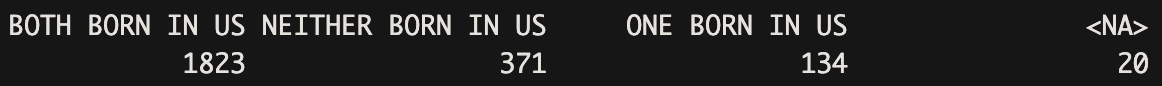
gss$parborn == 8 ~ "NEITHER BORN IN US"))

gss$PARENTBORN[gss$parborn == c(3:7)] <- NA

gss$PARENTBORN[gss$parborn == c(9, -1)] <- NA

gss$PARENTBORN <- factor(gss$PARENTBORN, levels = c("BOTH BORN IN US", "NEITHER BORN IN US", "ONE BORN IN US" ))

table(gss$PARENTBORN, useNA = "ifany")



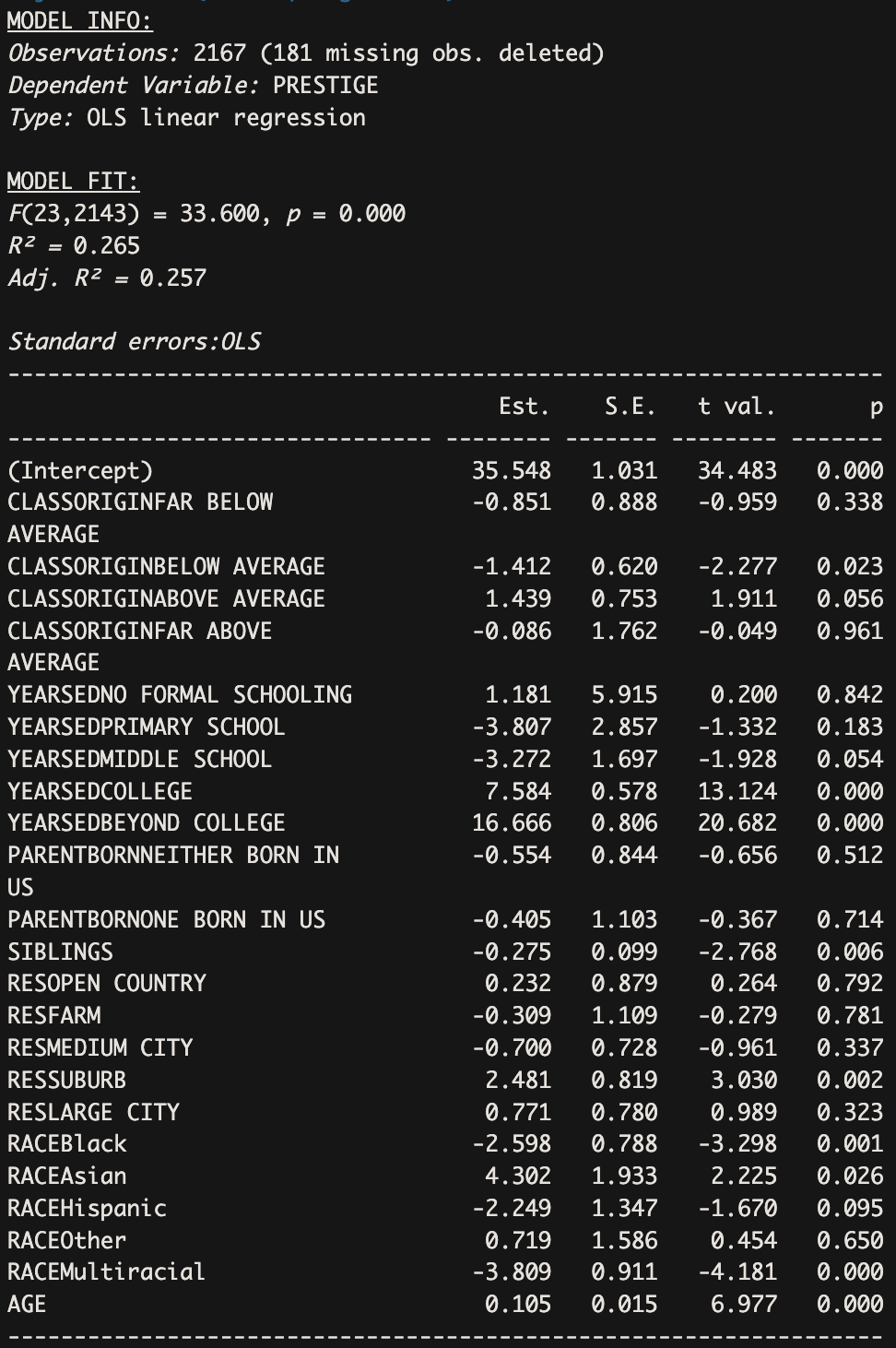
We included the variable PARBORN to account for intergenerational nativity as an important dimension of social background. Parental nativity status captures the structural and cultural resources—or constraints—that may shape an individual's opportunities beyond their own immigrant status. For example, those with U.S.-born parents may have greater inherited familiarity with U.S. educational and occupational institutions, which can complement or buffer the effects of social class background.

In this model, we set “Both Born in U.S.” as the reference group to mirror our earlier choice of setting “IMMIGRANT” as ‘1’ in the cleaned binary for the respondent's own nativity so that non-immigrants would be the reference group. This allows for meaningful comparisons: we can assess whether individuals with one or neither U.S.-born parents experience upward occupational mobility beyond what would be expected based on class origin alone, particularly when contrasted against those from fully U.S.-born family backgrounds. This approach helps us explore whether having U.S.-born parents offer additional social advantage in occupational outcomes—independent of one’s own immigration status and class origin.

We chose not to include answers where respondents did not know either one or both of their parents’ nativity in the NA group because these ‘don’t know’ categories were sparse.

model4 <- lm(PRESTIGE ~ CLASSORIGIN + YEARSED + PARENTBORN + SIBLINGS + RES + RACE + AGE, data = gss)

jtools::summ(model4, digits = 3)



Adding parental nativity (PARENTBORN), number of siblings (SIBLINGS), and residential context (RES) into Model 4 does introduce some meaningful shifts in the coefficients, though the changes are subtler compared to the transition from Model 2 to Model 3 (i.e., before and after education was included). These additional background variables help refine the estimates but do not fundamentally alter the key patterns established in Model 3. The F-statistic of 33.60 (p < 0.001) indicates that Model 4, as a whole, is statistically significant—meaning the set of independent variables included significantly explains variance in occupational prestige. The R² of 0.265 tells us that 26.5% of the variance in the outcome variable, PRESTIGE, is explained by the model's independent variables. This is a moderate amount of explanatory power, suggesting the model does a reasonable job capturing key factors that affect occupational status.

The Adjusted R² of 0.257 adjusts for the number of independent variables in the model. It’s only slightly lower than the raw R², which means that the additional variables in the model are not simply overfitting the data—they provide meaningful explanatory power.

The intercept of 35.548 now represents the predicted prestige score for non-immigrant White females, from an “Average” class origin, who have parents born in the U.S., lived in small cities or towns at 16 years old, and completed high school.

For class origin, the coefficients continue to shrink slightly. The coefficient for “Far Below Average” becomes less negative (-0.851, p = 0.338), and is now even further from significance, holding other variables constant. “Below Average” remains significant (p < 0.05) at -1.412, though weaker than in Model 3, holding other variables constant. “Above Average” drops slightly to 1.439 and is now insignificant (p = 0.056), holding other variables constant. These trends reinforce the earlier finding that much of class origin’s impact on prestige is channeled through education. The additional controls in Model 4 further dilute any remaining direct effect, indicating that family structure and residential background do not add much explanatory power to class origin once education is considered.

For race, most coefficients remain consistent with Model 3. The coefficient for Black respondents remains very significantly negative (-2.60, p < 0.01), though slightly reduced in magnitude, holding other variables constant. Asian respondents continue to have significantly higher prestige (+4.30, p < 0.05) but drops in significance (from 0.015 to 0.026), holding other variables constant. Multiracial individuals still experience a highly significantly lower prestige score (-3.81, p < 0.001), holding other variables constant. However, the coefficient for Hispanic respondents becomes insignificant (from p = 0.022 to p = 0.095), holding other variables constant, suggesting that some of the disadvantages observed in Model 3 may be partially explained by residential location or family background. The "Other" race category remains statistically insignificant.

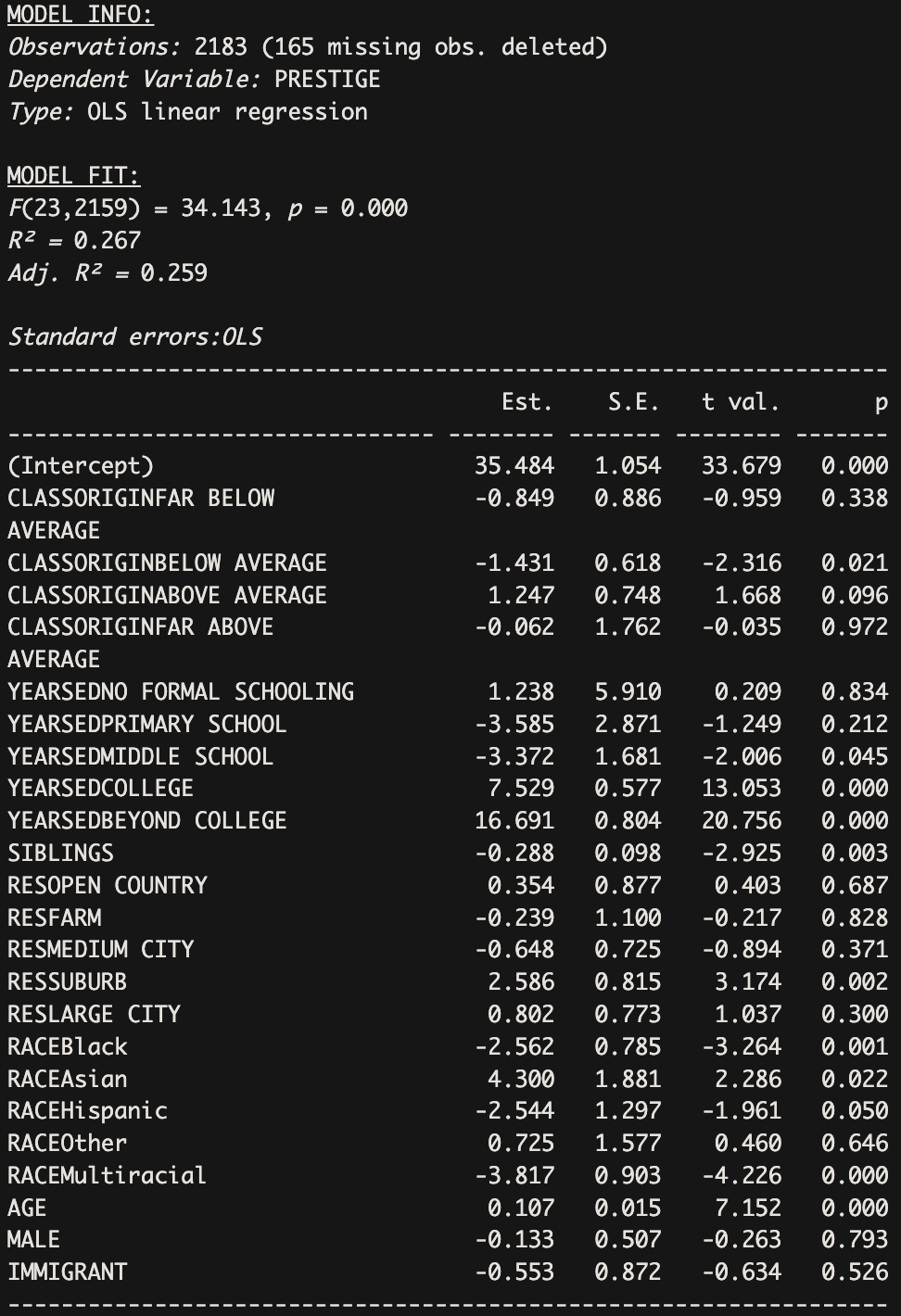
Education continues to have a strong effect on occupational prestige, with college and post-college education levels maintaining large, highly significant effects (7.584, p < 0.001; 16.666, p < 0.001). While college and post-college education experience lower prestige scores when accounting for these new variables, their effect remains strong and significant (as compared to the previous model: 7.907, p < 0.001; 17.125, p < 0.001), holding other variables constant. On the other hand, the education levels of “No Formal Schooling”, “Primary School”, and “Middle School” have insignificant effects on occupational prestige. While “Primary School” increases but continues to have a negative insignificant effect (-3.807, p = 0.183, as compared with -3.866, p = 0.178 in the previous model) and “No Formal Schooling” increases but continues to have a positive insignificant effect (1.181, p = 0.842, as compared with 0.623, p 0.916 in the previous model), “Middle School” increases, remaining negative, but becomes insignificant (-3.272, p = 0.054 as compared to -3.705, p < 0.05 in the previous model). Together, these findings underscore that education remains the most robust pathway to occupational prestige, and the addition of other variables does not drastically change this core insight.

Among the new variables, the only coefficient showing a statistically significant effect is residing in a suburb (RESSUBURB), which is associated with a +2.481 point increase in prestige (p < 0.01, very statistically significant), holding other variables constant. This suggests that a suburban upbringing may provide some social or institutional advantages linked to occupational outcomes. In addition, having more siblings shows a small but very significant negative association with prestige (-0.275, p < 0.01; This means that adding one sibling is associated with a decrease of 0.27 points in occupational prestige on average), possibly reflecting resource dilution within larger families.

**Question 10**

final\_model <- lm(PRESTIGE ~ CLASSORIGIN + YEARSED + SIBLINGS + RES + RACE + AGE + MALE + IMMIGRANT, data = gss)

jtools::summ(final\_model, digits = 3)



This final model is very similar to a combination of the third and fourth models, essentially with the addition of the MALE and IMMIGRANT variables as well as the deletion of the PARENTBORN variables to the fourth model. We chose to include MALE and IMMIGRANT variables because individuals’ gender and immigration status are theoretically highly influential in their upbringing. We chose to delete PARENTBORN variables because they were statistically insignificant in the previous model and did not contribute to the overall model’s statistical significance. We also felt that the IMMIGRANT variable was a stronger predictor of the effect of non-citizenship on occupational prestige than PARENTBORN.

The final model shows an F-statistic of 34.143 (p < 0.001), indicating that the overall model is statistically significant. The R² of 0.267 and Adjusted R² of 0.259 suggest that approximately 26% of the variation in occupational prestige is explained by the independent variables included, accounting for the number of independent variables. When we only have CLASSORIGIN, RACE, MALE, IMMIGRANT, AGE, and YEARSED as the independent variables (i.e., model 3), the R² was 0.259 and Adj. R² was 0.253. So in order to improve the general fit, we chose to include the variables CLASSORIGIN, YEARSED, SIBLINGS, RES, RACE, AGE, MALE, IMMIGRANT in the final model.

Our reference group is non-immigrant high school educated White females from an average class origin, who lived in small cities or towns at 16 and have, on average, an occupational prestige score of 35.484.

Class origin largely does not have a significant effect on occupational prestige, except for individuals from a below average class origin, which, on average, significantly (p < 0.05) decreases occupational prestige by -1.431, holding other variables constant. This suggests that early class background still matters (but only for people from certain backgrounds), but its impact may be partially suppressed by education.

The number of years of education a respondent obtains has mixed effects on occupational prestige. As the years of education increase and respondents finish college and grad school, the effects of their education on their occupational prestige are positive (+7.529 for college and +16.691 for beyond college) and are highly statistically significant (p < 0.001), as compared to the reference group that includes high school educated individuals, holding other variables constant. Meanwhile, when respondents only finish primary school and middle school, the effects of their education on their occupational prestige are negative and insignificant, as compared to the reference group, holding other variables constant. This suggests that as one becomes more educated, and especially as they finish higher education, their occupational prestige significantly increases.

The number of siblings a respondent grew up with has a significantly (p < 0.01) negative effect on their occupational prestige (-0.288), so as the number of siblings increases by 1, the occupational score decreases by 0.288. This suggests that the number of siblings one grows up with matters for occupational prestige, and if one has many siblings (potentially indicating that their parents potentially give them less attention or allot them fewer resources), their occupational prestige may suffer.

The residential context where respondents as when they grew up largely had an insignificant effect on occupational prestige, except the suburbs, which had a very significant (p < 0.01) positive effect on occupational prestige (+2.586). This suggests that growing up in the suburbs may provide individuals with resources unavailable to those in rural areas and cities that enable them to achieve higher occupational prestige.

In terms of race, Black (-2.562, p < 0.01) and Multiracial (-3.817, p < 0.001) respondents have significantly lower occupational prestige scores than their white counterparts, holding all other variables constant. This indicates that racial disparities persist. Hispanic, similarly, becomes just significant (-2.544, p = 0.050) again, as it was in the second and third models. This may suggest that once broader structural and contextual variables are included, the specific disadvantage associated with being Hispanic again shows. In contrast, Asian respondents (+4.30, p = 0.022 < 0.05) show significantly higher prestige scores than the reference group, holding other variables constant. The “Other” category is insignificant (p = 0.646).

As individuals age, they experience a highly significant (p < 0.001) positive effect on occupational prestige (0.107), meaning that for every year older individuals get, their occupational prestige score increases by 0.107, holding other variables constant. This suggests that older individuals have higher occupational prestige, perhaps due to accumulated experience or seniority.

Being male has an insignificant (p = 0.793) but negative effect on occupational prestige (-0.13). This suggests that males may be at a slight disadvantage in obtaining higher occupational prestige compared to the reference group that includes females, holding other variables constant, but there is no statistical evidence to back this up.

Being an immigrant has an insignificant (p = 0.526) but negative effect on occupational prestige as well (-0.55). This suggests that immigrants may be at a slight disadvantage in obtaining higher occupational prestige compared to the reference group that includes non-immigrants, holding other variables constant, but there is no statistical evidence to back this up.

The final model synthesizes the most theoretically relevant variables from both Model 3 and Model 4: it keeps essential demographic characteristics (age, race, sex, immigration), adds social class background (class origin, siblings, residential context), and emphasizes education as the key mediating variable. More importantly, it reveals that while higher education significantly boosts occupational prestige and mediates some class disadvantages, it does not eliminate racial inequality or fully override early class background. This is thus the best summary model considering this research question.