Status and OSR Project: Study 1 and 2 Data Analysis

Benjamin J. Zubaly

2024-07-19

Table of contents

# Introduction

This document tracks the data analysis for both Study 1 and 2 of our 2022 USRA project “Sexual selection and offspring sex ratio: Facial width-to-height ratio, dominance, and prestige”.

Currently, the important files in this project are this quarto document (which is where all of the code and output will be tracked) and the data folder. The data directory contains the clean versions of the individual participant data (from Study 1; Clean\_Individual\_Data.xlsx) and the facial ratings data (Clean\_Ratings\_Data.xlsx). It also contains a rich text file where I took notes on the preparation of these data sets (preparation\_of\_datasets.rtf).

## Installing Packages

I am using the groundhog package here to ensure that packages and dependencies are all installed with their versions on April 1, 2024, which will enhance the reproducibility of this code. In the following chunk, I have suppressed the output of the groundhog.library() call because rendering to a PDF was not working due to an invalid character in the output.

# Install and load the groundhog package to allow for standardized installation of package versions  
install.packages("groundhog")

Installing package into '/Users/benjaminzubaly/Library/R/arm64/4.4/library'  
(as 'lib' is unspecified)

The downloaded binary packages are in  
 /var/folders/38/1ybnplc53zdb089bn6drqfn00000gn/T//Rtmp1lfzQM/downloaded\_packages

library(groundhog)

Attached: 'Groundhog' (Version: 3.2.0)

Tips and troubleshooting: https://groundhogR.com

# Creating vector of packages to be installed  
pkg <- c("interactions", "effsize", "car", "psych", "readxl", "stats", "lmtest", "dplyr", "sandwich")  
  
# Use groundhog to install and load other packages  
suppressMessages({  
 groundhog.library(pkg, "2024-06-01")  
})

## Initial Variables in Data Sets

For reference, I will define all of the variables in the datasets, because I have created shorter names to make things easier in this analysis.

* Clean\_Individual\_Data.xlsx:
  + ID = The randomly generated ID of the participant.
  + rec\_meth = The recruitment method for the participant (0 = Prolific Recruitment, 1 = School Recruitment).
  + study1\_part = The part of Study 1 of which the participant completed the main survey (2 = Part 2, 3 = Part 3, blank = Not applicable because they were a school recruit).
  + part\_round = The round of recruitment, within the Study 1 part, in which the participant was recruited (blank for school recruits).
  + partner\_completed = Whether or not the partner of the the participant also completed the study (0 = Partner did not complete, 1 = Partner did complete).
  + partner\_ID = The ID of the partner of the participant (blank if missing).
  + study2\_survey = The survey number the participant was included in for the facial ratings procedure of Study 2.
  + fWHR\_ben = Benjamin’s measurements of facial width-to-height ratio (fWHR) for facial images.
  + fWHR\_mad = Madison’s measurements of fWHR for facial images.
  + expression\_not\_neutral = Whether the participant had a neutral facial expression in their facial image, as determined by Benjamin (0 = Neutral, 1 = Not Neutral).
  + angled\_face = Whether the participant’s face was angled in their facial image enough to alter the relationship between fWHR markers, as determined by Benjamin (0 = Not Angled, 1 = Angled).
  + obstructed\_face = Whether the person’s facial markers of fWHR were obstructed (0 = No Obstruction, 1 = Obstruction).
  + useable\_for\_ratings = Whether the photograph was deemed eligible for the facial ratings procedure as per our criteria.
  + photo\_resubmission = Whether the person resubmitted a facial photograph because their first photograph submission did not match our criteria.
  + sex = Sex of participant (0 = Female, 1 = Male).
  + ethnicity = Ethnicity of participant (1 = African, 2 = Black/African American, 3 = Caribbean, 4 = East Asian, 5 = Latino/Hispanic, 6 = Middle Eastern, 7 = Mixed, 8 = Native American or Alaskan Native, 9 = South Asian, 10 = White/Caucasian, 11 = White/Sepharic Jew, 12 = Black/British, 13 = White Mexican, 14 = Romani/Traveller, 15 = South East Asian, 16 = Indian, 17 = Filipino, 0 = Other)
  + nationality = Nationality of participant (1 = United Kingdom, 2 = United States, 3 = Ireland, 4 = Germany, 5 = France, 6 = Spain, 7 = Canada, 8 = Mexico, 9 = Italy, 10 = South Africa, 11 = Hungary, 12 = Zimbabwe, 13 = Portugal, 14 = Poland, 15 = New Zealand, 16 = Austria, 17 = Bulgaria, 18 = Australia, 19 = Nigeria, 20 = Phillipines, 21 = Pakistan, 22 = Brazil, 23 = Turkey, 24 = Netherlands).
  + age = Age of participant in years.
  + age\_first\_bio\_child = Age of the first biological child of the participant in years.
  + sex\_first\_bio\_child = Sex of the first biological child of the participant (0 = Female, 1 = Male).
  + num\_bio\_child = Number of children of the participant.
  + has\_adopted\_child = Whether the participant has an adopted child (0 = No, 1 = Yes).
  + ppp\_adjusted\_income\_usd = Income adjusted for purchasing power parity (PPP) in USD, to the nearest cent.
  + ISCO-08\_code = The code assigned to participant’s occupation, according to the International Standard Classification of Occupations (International Labor Office, 2012).
  + ISEI-08\_occ\_status = The occupational status score according to the International Socio-Economic Index (ISEI-08) (Ganzeboom, 2010) based on the ISCO-08 code.
  + SSS = Subjective social status.
  + SR\_dom\_cheng = Self-reported dominance on the Dominance-Prestige Scales (Cheng et al., 2010).
  + PR\_dom\_cheng = Partner-reported dominance via our adaptation of the Dominance-Prestige Scales (Cheng et al., 2010).
  + SR\_pres\_cheng = Self-reported prestige on Dominance-Prestige Scales (Cheng et al., 2010).
  + PR\_pres\_cheng = Partner-reported prestige via our adaptation of the Dominance-Prestige Scales (Cheng et al., 2010).
  + SAT = The simple adjectives test (Grant, 1992).
  + IPIP\_dom = The International Personality Item Pool (IPIP) dominance questionnaire (Goldberg et al., 2006).
  + intra\_comp = The Intrasexual Competition Scale (Buunk & Fisher, 2009).
  + ind\_agg\_full = the full score for the Indirect Aggression Scale Aggressor Version (IAS-A) (Forrest et al., 2005).
  + ind\_agg\_soc\_excl = The social exclusion subscale of the IAS-A (Forrest et al., 2005).
  + ind\_agg\_mal\_hum = The malicious humor subscale of the IAS-A (Forrest et al., 2005).
  + ind\_agg\_guilt = The guilt induction subscale of the IAS-A (Forrest et al., 2005).
  + photo\_method = The method with which the facial photograph was taken (0 = Computer, 1 = Phone).
* Clean\_Ratings\_Data.xlsx:
  + Start Date = Date and time the rater began the survey.
  + End Date = Date and time the rater finished the survey.
  + Custom Data = Identifier for the rater.
  + Survey # = The Study 2 survey number which the rater participated in.
  + Sex = The sex of the rater (“FEMALE” or “MALE”).
  + Age = The age of the rater in years.
  + typeofitem\_survey#\_page#\_ID = This is how the rest of the columns are structured, and they represent the ratings of a facial characteristic for a particular facial image where typeofitem is either dominance, masculinity/feminity, or attractiveness, survey# is the number of the Study 2 survey that they facial images was included in, page# is the page of the survey that the facial image was presented on, and ID is the ID of the participant associated with the facial image. In total, there are 801 of these columns, three for each facial image rated and one for each characteristic of each image rated. They are arranged with dominance ratings first, then masculinity/feminity ratings, then attractiveness ratings.

# Study 1 Analysis

## Data Exploration

### Descriptive Statistics

#### Continuous Variables

First, we will read in the data as a data frame called individualsdata and calculate descriptive statistics for the continuous variables in the sample.

# Reading in the individuals dataset as a data frame called individualsdata   
individualsdata <- read\_excel("./data/Clean\_Individual\_Data.xlsx", sheet = 1)  
  
# Defining continuous variables to describe  
cont.variables <- individualsdata[c("age", "age\_first\_bio\_child", "num\_bio\_child", "ppp\_adjusted\_income\_usd", "ISEI-08\_occ\_status", "SSS", "SR\_dom\_cheng", "PR\_dom\_cheng", "SR\_pres\_cheng", "PR\_pres\_cheng", "SAT", "IPIP\_dom")]  
  
# Calculating descriptive statistics with the psych package's describe() function  
d.stats.cont.ind <- describe(cont.variables, na.rm = TRUE)  
  
# Displaying the result  
print(d.stats.cont.ind)

vars n mean sd median trimmed mad  
age 1 265 33.79 5.06 34.00 33.61 4.45  
age\_first\_bio\_child 2 265 4.18 2.18 4.00 4.11 2.97  
num\_bio\_child 3 265 1.68 0.69 2.00 1.59 1.48  
ppp\_adjusted\_income\_usd 4 212 46720.56 36994.16 43047.78 42333.36 31547.54  
ISEI-08\_occ\_status 5 247 60.24 19.05 65.01 62.00 17.26  
SSS 6 265 50.71 14.85 50.00 51.77 11.86  
SR\_dom\_cheng 7 265 2.92 1.06 2.88 2.86 1.11  
PR\_dom\_cheng 8 230 2.58 0.99 2.44 2.53 0.99  
SR\_pres\_cheng 9 265 5.04 0.97 5.00 5.06 1.15  
PR\_pres\_cheng 10 230 5.41 0.92 5.44 5.44 0.99  
SAT 11 265 2.50 2.21 2.00 2.23 1.48  
IPIP\_dom 12 265 27.68 7.99 28.00 27.57 8.90  
 min max range skew kurtosis se  
age 22.00 50.00 28.00 0.37 0.12 0.31  
age\_first\_bio\_child 1.00 9.00 8.00 0.25 -0.92 0.13  
num\_bio\_child 1.00 4.00 3.00 0.58 -0.48 0.04  
ppp\_adjusted\_income\_usd 0.00 203252.03 203252.03 1.39 2.96 2540.77  
ISEI-08\_occ\_status 11.56 86.72 75.16 -0.71 -0.64 1.21  
SSS 0.00 85.00 85.00 -0.85 1.51 0.91  
SR\_dom\_cheng 1.00 7.00 6.00 0.58 0.36 0.07  
PR\_dom\_cheng 0.89 5.33 4.44 0.42 -0.47 0.07  
SR\_pres\_cheng 1.89 7.00 5.11 -0.22 -0.35 0.06  
PR\_pres\_cheng 2.78 7.00 4.22 -0.32 -0.55 0.06  
SAT 0.00 11.00 11.00 1.05 0.80 0.14  
IPIP\_dom 11.00 48.00 37.00 0.09 -0.55 0.49

* Participant age in years:
  + Mean (SD): 33.8 (±5.06)
  + Median: 34
  + Range: 28 (22-50)
* First biological child age in years:
  + Mean (SD): 4.16 (±2.18)
  + Median: 4
  + Range: 8 (1-9)
* Number of Biological Children:
  + Mean (SD): 1.68 (±.69)
  + Median: 2
  + Range: 3 (1-4)
* Income Adjusted for Purchasing Power Parity (in USD):
  + Mean (SD): 46,720.56 (36,994.16)
  + Median: 43,047.78
  + Range: 203,252.03 (0-203,252.03)
* Occupational Status (ISEI-08)
  + Mean (SD): 60.23 (19.025)
  + Median: 65.01
  + Range: 75.16 (11.56-86.72)
* Subjective Social Status:
  + Mean (SD): 50.63 (14.895)
  + Median: 50
  + Range: 85 (0-85)
* Self-Reported Dominance:
  + Mean (SD): 2.92 (±1.06)
  + Median: 2.88
  + Range: 6 (1-7)
* Partner-Reported Dominance:
  + Mean (SD): 2.58 (±.99)
  + Median: 2.44
  + Range: 4.44 (.888-5.333)
* Self-reported Prestige:
  + Mean (SD): 5.04 (±.97)
  + Median: 5
  + Range: 5.111 (1.889-7)
* Partner-Reported Prestige:
  + Mean (SD): 5.41 (±.92)
  + Median: 5.444
  + Range: 4.222 (2.777-7)
* Simple Adjectives Test (SAT):
  + Mean (SD): 2.50 (±2.21)
  + Median: 2
  + Range: 11 (0-11)
* IPIP Dominance:
  + Mean (SD): 27.57 (±8.90)
  + Median: 28
  + Range: 37 (11-48)

#### Categorical Variables

Now we will make our categorical variables into factor variables and calculate frequencies for each.

# Recoding variables currently numeric as factors  
 # Recruitment method to prolific or school  
individualsdata$rec\_meth <- factor(individualsdata$rec\_meth, levels = c(0, 1), labels = c("prolific", "school"))  
 # Photo Resubmission to no\_submission or had\_submission  
individualsdata$photo\_resubmission <- factor(individualsdata$photo\_resubmission, levels = c(0, 1), labels = c("no\_resubmission", "had\_resubmission"))  
 # Ethnicity  
individualsdata$ethnicity <- factor(individualsdata$ethnicity, levels = c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17), labels = c("Other", "African", "Black/African\_American","Caribbean", "East Asian", "Latino/Hispanic", "Middle Eastern", "Mixed", "Native American or Alaskan Native", "South Asian", "White/Caucasian", "White/Sepharic Jew", "Black/British", "White Mexican", "Romani/Traveller", "South East Asian", "Indian", "Filipino"))  
 # Nationality  
individualsdata$nationality <- factor(individualsdata$nationality, levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24), labels = c("United Kingdom", "United States", "Ireland","Germany", "France", "Spain", "Canada", "Mexico", "Italy", "South Africa", "Hungary", "Zimbabwe", "Portugal", "Poland", "New Zealand", "Austria", "Bulgaria", "Australia", "Nigeria", "Phillipines", "Pakistan", "Brazil", "Turkey", "Netherlands"))  
 # Expression not neutral to neutral or not\_neutral  
individualsdata$expression\_not\_neutral <- factor(individualsdata$expression\_not\_neutral, levels = c(0, 1), labels = c("neutral", "not\_neutral"))  
 # Sex to female or male  
individualsdata$sex <- factor(individualsdata$sex, levels = c(0, 1), labels = c("female", "male"))  
 # Sex of first biological child to female or male  
individualsdata$sex\_first\_bio\_child <- factor(individualsdata$sex\_first\_bio\_child, levels = c(0, 1), labels = c("female", "male"))  
 # Has adopted child to no or yes  
individualsdata$has\_adopted\_child <- factor(individualsdata$has\_adopted\_child, levels = c(0, 1), labels = c("no", "yes"))  
 # Photograph method to computer or phone  
individualsdata$photo\_method <- factor(individualsdata$photo\_method, levels = c(0, 1), labels = c("computer", "phone"))  
  
# Frequencies of categorical variables  
rec\_meth\_table <- table(individualsdata$rec\_meth) # Calculating frequencies  
rec\_meth\_table # Displaying the frequencies

prolific school   
 257 8

photo\_resub\_table <- table(individualsdata$photo\_resubmission) # Calculating frequencies  
photo\_resub\_table # Displaying the frequencies

no\_resubmission had\_resubmission   
 231 34

ethnicity\_table <- table(individualsdata$ethnicity) # Calculating frequencies  
ethnicity\_table # Displaying the frequencies

Other African   
 0 52   
 Black/African\_American Caribbean   
 8 2   
 East Asian Latino/Hispanic   
 10 4   
 Middle Eastern Mixed   
 2 7   
Native American or Alaskan Native South Asian   
 0 7   
 White/Caucasian White/Sepharic Jew   
 165 0   
 Black/British White Mexican   
 3 0   
 Romani/Traveller South East Asian   
 0 4   
 Indian Filipino   
 1 0

nationality\_table <- table(individualsdata$nationality) # Calculating frequencies  
nationality\_table # Displaying the frequencies

United Kingdom United States Ireland Germany France   
 120 32 0 0 0   
 Spain Canada Mexico Italy South Africa   
 6 8 0 5 55   
 Hungary Zimbabwe Portugal Poland New Zealand   
 3 7 9 5 2   
 Austria Bulgaria Australia Nigeria Phillipines   
 1 1 2 2 1   
 Pakistan Brazil Turkey Netherlands   
 1 1 2 2

expression\_neutral\_table <- table(individualsdata$expression\_not\_neutral) # Calculating frequencies  
expression\_neutral\_table # Displaying the frequencies

neutral not\_neutral   
 230 35

sex\_table <- table(individualsdata$sex) # Calculating frequencies  
sex\_table # Displaying the frequencies

female male   
 141 124

child\_sex\_table <- table(individualsdata$sex\_first\_bio\_child) # Calculating frequencies  
child\_sex\_table # Displaying the frequencies

female male   
 108 157

adopted\_child\_table <- table(individualsdata$has\_adopted\_child) # Calculating frequencies  
adopted\_child\_table # Displaying the frequencies

no yes   
264 1

photo\_meth\_table <- table(individualsdata$photo\_method) # Calculating frequencies  
photo\_meth\_table # Displaying the frequencies

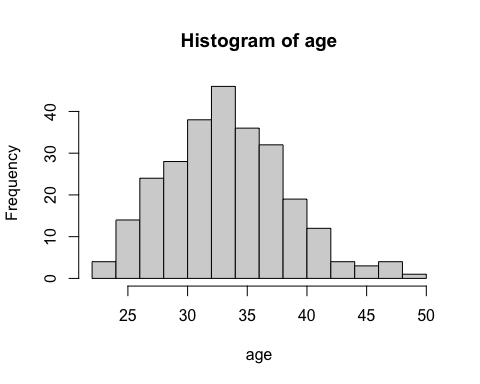
computer phone   
 48 215

* Sex of participant (parent):
  + Females: n=141, 53.2%
  + Males: n=124, 46.8%
* Sex of first biological child:
  + Female: n=108, 40.8%
  + Male: n=157, 59.2%
* Ethnicity:
  + For ethnicity, there were a higher proportion of White/Caucasian (n=165) and African (n=52) respondents, with other ethnicities represented at n≤10.
* Nationality:
  + For nationality, the most represented groups were participants from the United Kingdom (n=120), South Africa (n=55), and the United States (n=32), with all other nationalities at n≤9.
* Has Adopted Children:
  + No: n=264, 99.6%
  + Yes: n=1, 0.4%

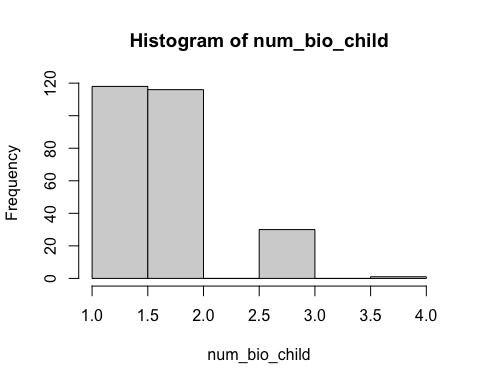
### Data Visualization

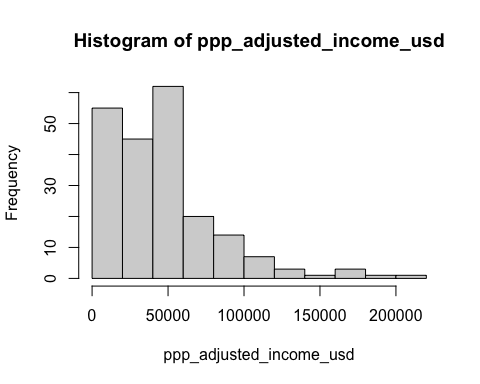
Now I will create histograms of each continuous variable to visualize their distributions.

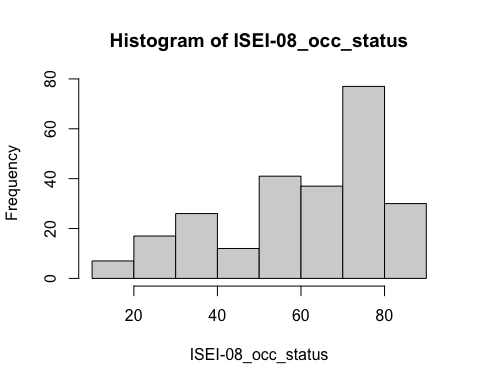
# Create histograms for each variable using the base hist function  
for (variable in names(cont.variables)) {  
 # Create histogram  
 hist(cont.variables[[variable]], main = paste("Histogram of", variable), xlab = variable)  
}

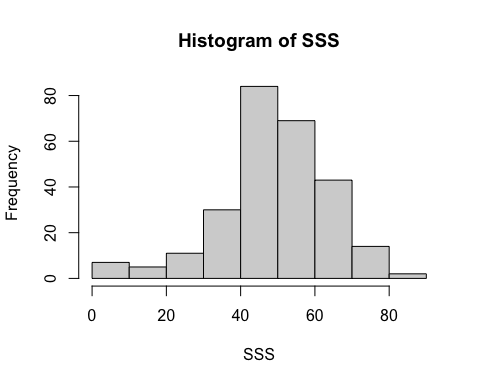


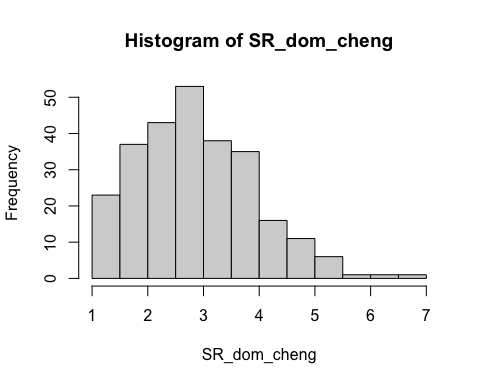


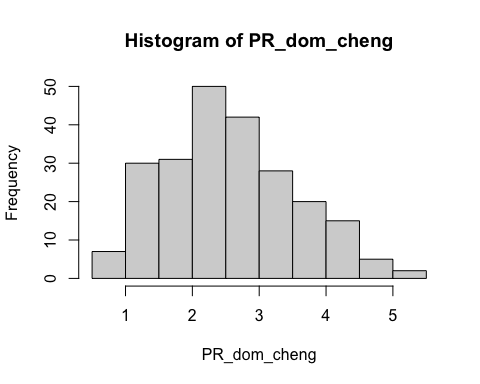


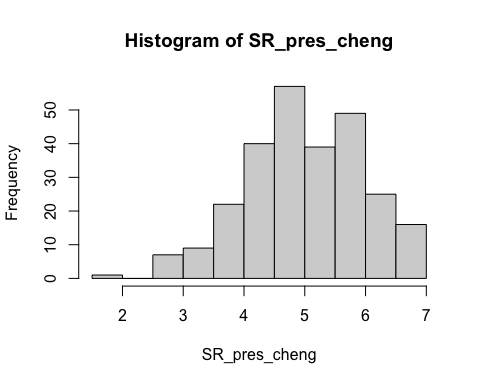


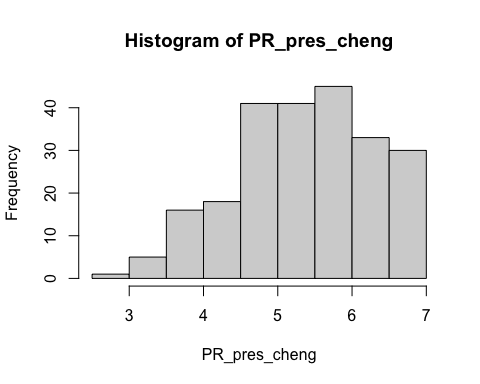


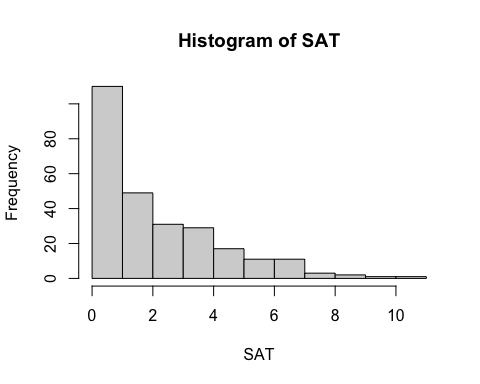


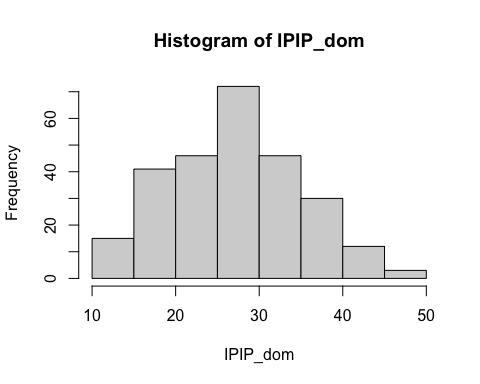












* Each of the variables seems to tend toward a normal distribution except for SAT (which, being a count variable, seems to follow a poisson distribution), the number of biological children (which is likely influenced by it being our selection criteria and stopping rules), and income (which is commonly positively skewed).

## Assessing Whether to Create Behavioral Dominance Composite

To determine whether self-and partner-reported dominance and self- and partner-reported prestige will be averaged into a composite measure, zero-order Pearson correlations will be calculated.

# Running the correlation between self-and partner-reported dominance  
dom.corr <- corr.test(individualsdata$SR\_dom\_cheng, individualsdata$PR\_dom\_cheng)  
print(dom.corr)

Call:corr.test(x = individualsdata$SR\_dom\_cheng, y = individualsdata$PR\_dom\_cheng)  
Correlation matrix   
[1] 0.48  
Sample Size   
[1] 230  
These are the unadjusted probability values.  
 The probability values adjusted for multiple tests are in the p.adj object.   
[1] 0  
  
 To see confidence intervals of the correlations, print with the short=FALSE option

# Runnign the correlation between self-and partner-reported prestige  
pres.corr <- corr.test(individualsdata$SR\_pres\_cheng, individualsdata$PR\_pres\_cheng)  
print(pres.corr)

Call:corr.test(x = individualsdata$SR\_pres\_cheng, y = individualsdata$PR\_pres\_cheng)  
Correlation matrix   
[1] 0.35  
Sample Size   
[1] 230  
These are the unadjusted probability values.  
 The probability values adjusted for multiple tests are in the p.adj object.   
[1] 0  
  
 To see confidence intervals of the correlations, print with the short=FALSE option

Although the correlations are significant, they are not as high as I would like them to be. We will first conduct the analysis with only the self-reported dominance and prestige, and then we will assess them as composite predictors.

* Self- and partner-reported dominance: r=.48 (p<.01)
* Self- and partner-reported prestige: r=.35 (p<.01)

## Assessing Demographic Differences Between Parents of First-Born Sons and First-Born Daughters

To assess for biased demographic characteristics (sex and age) between parents of first-born sons and first-born daughters, we will conduct a chi-square test and t-test.

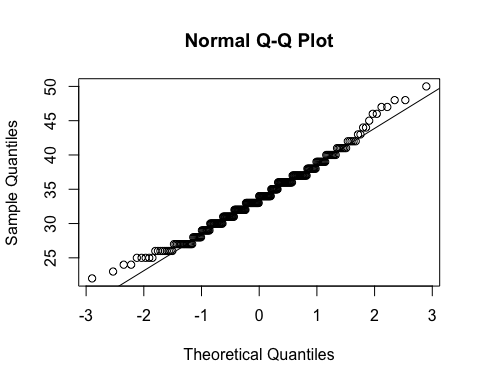
# Chi-square test to assess whether parental sex is associated with sex of first-born in our sample  
 # Creating the frequencies table  
par.dem.freq.table <- table(individualsdata$sex\_first\_bio\_child, individualsdata$sex)  
 # Conducting the chi-square test  
chi.parent.dem <- chisq.test(par.dem.freq.table, correct = FALSE)  
 # Displaying the result  
print(par.dem.freq.table)

female male  
 female 59 49  
 male 82 75

print(chi.parent.dem)

Pearson's Chi-squared test  
  
data: par.dem.freq.table  
X-squared = 0.14807, df = 1, p-value = 0.7004

# t-test to assess whether parental age differs between sex of child  
 # QQ-plot and Shapir-Wilk test to assess normality  
qqnorm(individualsdata$age); qqline(individualsdata$age)



shapiro.test(individualsdata$age)

Shapiro-Wilk normality test  
  
data: individualsdata$age  
W = 0.98512, p-value = 0.007436

# Levene's test for equality of variances  
leveneTest(y = individualsdata$age, group = individualsdata$sex\_first\_bio\_child)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.3623 0.5477  
 263

# Conducting the t.test  
t.parent.dem <- t.test(data = individualsdata, age ~ sex\_first\_bio\_child, var.equal = TRUE)  
 # Displaying the result  
print(t.parent.dem)

Two Sample t-test  
  
data: age by sex\_first\_bio\_child  
t = -0.64581, df = 263, p-value = 0.519  
alternative hypothesis: true difference in means between group female and group male is not equal to 0  
95 percent confidence interval:  
 -1.656494 0.838259  
sample estimates:  
mean in group female mean in group male   
 33.54630 33.95541

* Chi-Square Test:
  + The chi-square test indicates that there is not a significant relationship between parental sex and sex of first-born (χ²(1) = .15, p = .7).
* T-test:
  + The t-test assumption of normality was not supported by the Shapiro-Wilk test, but a visual inspection of the QQ-plot (as wella as the histogram above) indicates that it does not deviate substantially from normality. The significant Shapir-Wilk test is likely overpowered with our sample size. The Levene’s test is not significant, indicating that we should accept the assumption of homogeneity of variances. Therefore, the t-test is ran with equal variances assumed. There is no significant difference between the mean age of parents of each sex (t(263) = -.65, p = .519).

## Testing For Trivers-Willard Effects with Social Status Operationalizations:

To test for Trivers-Willard effects of social status, we will test three binary logistic regression models with the social status indicator, sex, and their interaction as predictors. The three social status indicators are income (ppp\_adjusted\_income\_usd), occupational status (ISEI-08\_occ\_status), and subjective social status (SSS). Before fitting the models, we will first create standardized versions of those variables as z\_income, z\_occ\_status, and z\_SSS. In addition, I will recode sex so that female is the moderator (i.e., female = 1).

# Creating standardized scores for status indicators  
individualsdata$z\_income <- scale(individualsdata$ppp\_adjusted\_income\_usd)  
individualsdata$z\_occ\_status <- scale(individualsdata$`ISEI-08\_occ\_status`)  
individualsdata$z\_SSS <- scale(individualsdata$SSS)  
# Creating a new variable sex\_2, which represents the participants' sex as 0 = male and 1 = female  
individualsdata$sex <- ifelse(individualsdata$sex == "female", 1, 0)  
# Converting sex back to a factor variable with labels  
individualsdata$sex <- factor(individualsdata$sex, levels = c(0, 1), labels = c("male", "female"))

### Model 1: Income

First, we will fit our model.

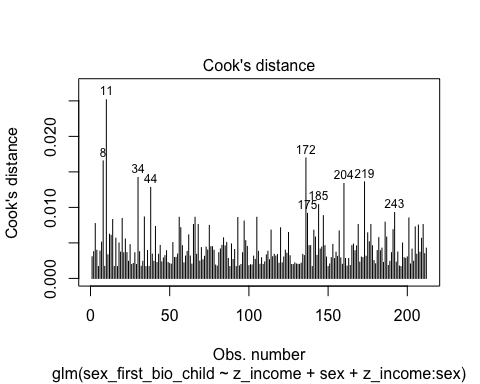
# Fitting the model with income, sex, and their interaction  
income\_status\_model <- glm(sex\_first\_bio\_child ~ z\_income + sex + z\_income:sex, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(income\_status\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating c\_z\_income, which represents standardized income after adding a constant of 10  
individualsdata$c\_z\_income <- individualsdata$z\_income + 10  
  
# Creating ln\_c\_z\_income, which represents the natural log of the standardized income scores after adding a constant of 10  
individualsdata$ln\_c\_z\_income <- log(individualsdata$c\_z\_income)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_income\_model <- glm(sex\_first\_bio\_child ~ c\_z\_income + sex + c\_z\_income:ln\_c\_z\_income, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_income\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_income + sex + c\_z\_income:ln\_c\_z\_income,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 36.77277 29.37186 1.252 0.211  
c\_z\_income -12.66782 9.60823 -1.318 0.187  
sexfemale -0.04325 0.29637 -0.146 0.884  
c\_z\_income:ln\_c\_z\_income 3.91271 2.89506 1.352 0.177  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 289.05 on 211 degrees of freedom  
Residual deviance: 280.56 on 208 degrees of freedom  
 (53 observations deleted due to missingness)  
AIC: 288.56  
  
Number of Fisher Scoring iterations: 5

* Because the interaction term is not significant (*p* = .18), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(income\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_income + sex + z\_income:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.33125 0.20536 1.613 0.107  
z\_income 0.27414 0.22068 1.242 0.214  
sexfemale 0.02617 0.29689 0.088 0.930  
z\_income:sexfemale 0.20572 0.33187 0.620 0.535  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 289.05 on 211 degrees of freedom  
Residual deviance: 282.50 on 208 degrees of freedom  
 (53 observations deleted due to missingness)  
AIC: 290.5  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_income\_model <- income\_status\_model$null.deviance - income\_status\_model$deviance  
Chi\_income\_model

[1] 6.541723

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_income\_model <- income\_status\_model$df.null - income\_status\_model$df.residual  
df\_income\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_income\_model <- 1 - pchisq(Chi\_income\_model, df\_income\_model)  
prob\_Chi\_income\_model

[1] 0.08803148

* Looking at the chi-square test, the model is marginally significant (χ²(3) = 6.54, p = .09). However, none of the other variables, including income, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(income\_status\_model$coefficients)

(Intercept) z\_income sexfemale z\_income:sexfemale   
 1.392707 1.315395 1.026514 1.228413

### Model 2: Occupational Status

First, we will fit our model.

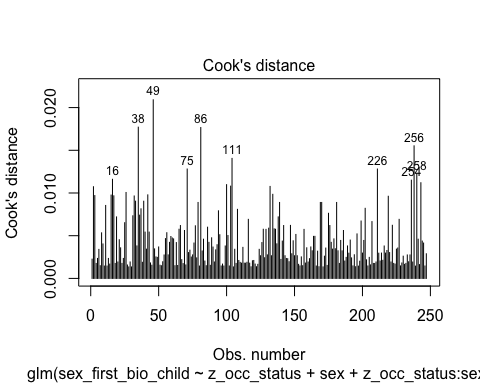
# Fitting the model with occupational status, sex, and their interaction  
occupational\_status\_model <- glm(sex\_first\_bio\_child ~ z\_occ\_status + sex + z\_occ\_status:sex, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(occupational\_status\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating c\_z\_occ\_status, which represents standardized income after adding a constant of 10  
individualsdata$c\_z\_occ\_status <- individualsdata$z\_occ\_status + 10  
  
# Creating ln\_c\_z\_occ\_status, which represents the natural log of the standardized occupational status scores after adding a constant of 10  
individualsdata$ln\_c\_z\_occ\_status <- log(individualsdata$c\_z\_occ\_status)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_occ\_status\_model <- glm(sex\_first\_bio\_child ~ c\_z\_occ\_status + sex + c\_z\_occ\_status:ln\_c\_z\_occ\_status, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_occ\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_occ\_status + sex + c\_z\_occ\_status:ln\_c\_z\_occ\_status,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -15.0458 25.3360 -0.594 0.553  
c\_z\_occ\_status 5.0833 8.6475 0.588 0.557  
sexfemale -0.1340 0.2606 -0.514 0.607  
c\_z\_occ\_status:ln\_c\_z\_occ\_status -1.5310 2.6506 -0.578 0.564  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 333.42 on 246 degrees of freedom  
Residual deviance: 332.29 on 243 degrees of freedom  
 (18 observations deleted due to missingness)  
AIC: 340.29  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .564), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(occupational\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_occ\_status + sex + z\_occ\_status:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4597 0.1897 2.423 0.0154 \*  
z\_occ\_status 0.2597 0.1955 1.329 0.1840   
sexfemale -0.1458 0.2609 -0.559 0.5762   
z\_occ\_status:sexfemale -0.3061 0.2620 -1.169 0.2426   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 333.42 on 246 degrees of freedom  
Residual deviance: 331.25 on 243 degrees of freedom  
 (18 observations deleted due to missingness)  
AIC: 339.25  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_occ\_status\_model <- occupational\_status\_model$null.deviance - occupational\_status\_model$deviance  
Chi\_occ\_status\_model

[1] 2.165361

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_occ\_status\_model <- occupational\_status\_model$df.null - occupational\_status\_model$df.residual  
df\_occ\_status\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_occ\_status\_model <- 1 - pchisq(Chi\_occ\_status\_model, df\_occ\_status\_model)  
prob\_Chi\_occ\_status\_model

[1] 0.5388033

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 2.165, *p* = .539). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(occupational\_status\_model$coefficients)

(Intercept) z\_occ\_status sexfemale   
 1.5835288 1.2965480 0.8643407   
z\_occ\_status:sexfemale   
 0.7362890

### Model 3: Subjective Social Status

First, we will fit our model.

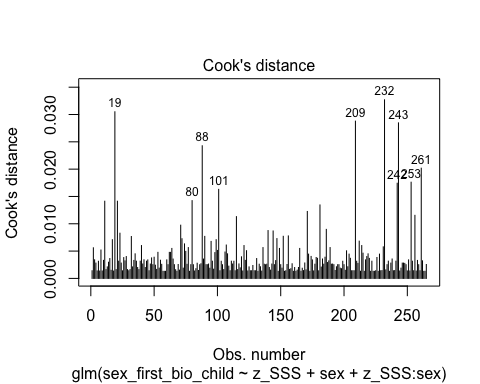
# Fitting the model with subjective social status, sex, and their interaction  
SSS\_status\_model <- glm(sex\_first\_bio\_child ~ z\_SSS + sex + z\_SSS:sex, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(SSS\_status\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating c\_z\_SSS, which represents standardized SSS after adding a constant of 10  
individualsdata$c\_z\_SSS <- individualsdata$z\_SSS + 10  
  
# Creating ln\_c\_z\_SSS, which represents the natural log of the standardized SSS scores after adding a constant of 10  
individualsdata$ln\_c\_z\_SSS <- log(individualsdata$c\_z\_SSS)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_SSS\_status\_model <- glm(sex\_first\_bio\_child ~ c\_z\_SSS + sex + c\_z\_SSS:ln\_c\_z\_SSS, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_SSS\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_SSS + sex + c\_z\_SSS:ln\_c\_z\_SSS,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -2.12274 13.19672 -0.161 0.872  
c\_z\_SSS 0.63992 4.53648 0.141 0.888  
sexfemale -0.07983 0.25213 -0.317 0.752  
c\_z\_SSS:ln\_c\_z\_SSS -0.16722 1.39572 -0.120 0.905  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 357.50 on 261 degrees of freedom  
AIC: 365.5  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .905), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(SSS\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_SSS + sex + z\_SSS:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.41245 0.18471 2.233 0.0256 \*  
z\_SSS 0.16039 0.18929 0.847 0.3968   
sexfemale -0.07916 0.25203 -0.314 0.7534   
z\_SSS:sexfemale -0.11409 0.25300 -0.451 0.6520   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 357.31 on 261 degrees of freedom  
AIC: 365.31  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_SSS\_status\_model <- SSS\_status\_model$null.deviance - SSS\_status\_model$deviance  
Chi\_SSS\_status\_model

[1] 0.9452272

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_SSS\_status\_model <- SSS\_status\_model$df.null - SSS\_status\_model$df.residual  
df\_SSS\_status\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_SSS\_status\_model <- 1 - pchisq(Chi\_SSS\_status\_model, df\_SSS\_status\_model)  
prob\_Chi\_SSS\_status\_model

[1] 0.814502

* Looking at the chi-square test, the entire model is not significant (χ²(3) = .945, *p* = .815). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(SSS\_status\_model$coefficients)

(Intercept) z\_SSS sexfemale z\_SSS:sexfemale   
 1.5105073 1.1739718 0.9238889 0.8921793

## Testing Hypothesis 1

Hypothesis 1 is that higher parental dominance increases the probability of having a son, and higher parental prestige neither increases nor decreases the probability of having a son. This hypothesis will be tested through three models. The first will have dominance and prestige as predictor variables of offspring sex. The second will have dominance, sex, and the interaction between dominance and sex as predictors of offspring sex. The third model will have prestige, sex, and the interaction between prestige and sex as predictors of offspring sex.

Before constructing the models and testing them, we will first standardize the predictor variables.

# Standardizing the two predictors  
individualsdata$z\_SR\_dom\_cheng <- scale(individualsdata$SR\_dom\_cheng)  
individualsdata$z\_SR\_pres\_cheng <- scale(individualsdata$SR\_pres\_cheng)

### Model 1: Dominance and Prestige Main Effects

First, we will fit our model.

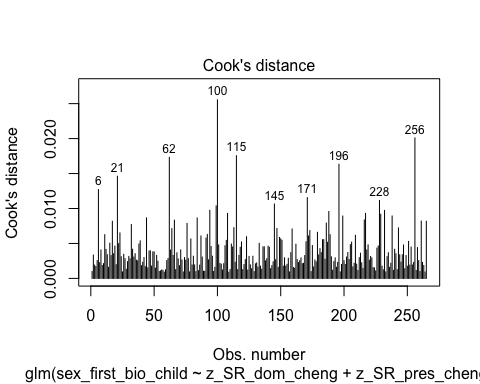
# Fitting the model with self-reported dominance and prestige  
hyp\_1\_main\_effects\_model <- glm(sex\_first\_bio\_child ~ z\_SR\_dom\_cheng + z\_SR\_pres\_cheng, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_main\_effects\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating new variables that add a constant to the standardized variables  
individualsdata$c\_z\_SR\_dom\_cheng <- individualsdata$z\_SR\_dom\_cheng + 10  
individualsdata$c\_z\_SR\_pres\_cheng <- individualsdata$z\_SR\_pres\_cheng + 10  
  
# Creating the natural log transformations of the variables  
individualsdata$ln\_c\_z\_SR\_dom\_cheng <- log(individualsdata$c\_z\_SR\_dom\_cheng)  
individualsdata$ln\_c\_z\_SR\_pres\_cheng <- log(individualsdata$c\_z\_SR\_pres\_cheng)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_main\_effects\_model <- glm(sex\_first\_bio\_child ~ c\_z\_SR\_dom\_cheng + c\_z\_SR\_pres\_cheng + c\_z\_SR\_dom\_cheng:ln\_c\_z\_SR\_dom\_cheng + c\_z\_SR\_pres\_cheng:ln\_c\_z\_SR\_pres\_cheng, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_main\_effects\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_SR\_dom\_cheng + c\_z\_SR\_pres\_cheng +   
 c\_z\_SR\_dom\_cheng:ln\_c\_z\_SR\_dom\_cheng + c\_z\_SR\_pres\_cheng:ln\_c\_z\_SR\_pres\_cheng,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -0.6553 26.9976 -0.024 0.981  
c\_z\_SR\_dom\_cheng 1.0101 6.4042 0.158 0.875  
c\_z\_SR\_pres\_cheng -1.4908 6.2775 -0.237 0.812  
c\_z\_SR\_dom\_cheng:ln\_c\_z\_SR\_dom\_cheng -0.2459 1.9275 -0.128 0.898  
c\_z\_SR\_pres\_cheng:ln\_c\_z\_SR\_pres\_cheng 0.4990 1.9104 0.261 0.794  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 354.58 on 260 degrees of freedom  
AIC: 364.58  
  
Number of Fisher Scoring iterations: 4

* Because neither of the interaction terms are significant (*p* = .898 and *p* = .794), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_main\_effects\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_SR\_dom\_cheng + z\_SR\_pres\_cheng,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.3796 0.1260 3.012 0.00259 \*\*  
z\_SR\_dom\_cheng 0.1885 0.1289 1.462 0.14375   
z\_SR\_pres\_cheng 0.1476 0.1259 1.173 0.24078   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 354.66 on 262 degrees of freedom  
AIC: 360.66  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_main\_effects\_model <- hyp\_1\_main\_effects\_model$null.deviance - hyp\_1\_main\_effects\_model$deviance  
Chi\_hyp\_1\_main\_effects\_model

[1] 3.595436

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_main\_effects\_model <- hyp\_1\_main\_effects\_model$df.null - hyp\_1\_main\_effects\_model$df.residual  
df\_hyp\_1\_main\_effects\_model

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_main\_effects\_model <- 1 - pchisq(Chi\_hyp\_1\_main\_effects\_model, df\_hyp\_1\_main\_effects\_model)  
prob\_Chi\_hyp\_1\_main\_effects\_model

[1] 0.1656766

* Looking at the chi-square test, the entire model is not significant (χ²(2) = 3.595, p = .166). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_main\_effects\_model$coefficients)

(Intercept) z\_SR\_dom\_cheng z\_SR\_pres\_cheng   
 1.461633 1.207414 1.159106

### Model 2: Dominance, Sex, and their Interaction

First, we will fit our model.

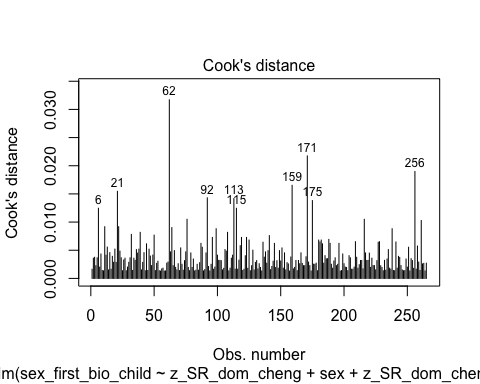
# Fitting the model with self-reported dominance, sex, and their interaction  
hyp\_1\_dom\_model <- glm(sex\_first\_bio\_child ~ z\_SR\_dom\_cheng + sex + z\_SR\_dom\_cheng:sex, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_dom\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_dom\_model <- glm(sex\_first\_bio\_child ~ c\_z\_SR\_dom\_cheng + sex + c\_z\_SR\_dom\_cheng:ln\_c\_z\_SR\_dom\_cheng, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_dom\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_SR\_dom\_cheng + sex +   
 c\_z\_SR\_dom\_cheng:ln\_c\_z\_SR\_dom\_cheng, family = binomial(link = logit),   
 data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -2.27409 19.43070 -0.117 0.907  
c\_z\_SR\_dom\_cheng 0.45174 6.34746 0.071 0.943  
sexfemale -0.03523 0.25528 -0.138 0.890  
c\_z\_SR\_dom\_cheng:ln\_c\_z\_SR\_dom\_cheng -0.08004 1.91027 -0.042 0.967  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 356.02 on 261 degrees of freedom  
AIC: 364.02  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .967), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_dom\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_SR\_dom\_cheng + sex + z\_SR\_dom\_cheng:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.42249 0.18638 2.267 0.0234 \*  
z\_SR\_dom\_cheng 0.01806 0.18104 0.100 0.9205   
sexfemale -0.02476 0.25789 -0.096 0.9235   
z\_SR\_dom\_cheng:sexfemale 0.34576 0.26435 1.308 0.1909   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 354.29 on 261 degrees of freedom  
AIC: 362.29  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_dom\_model <- hyp\_1\_dom\_model$null.deviance - hyp\_1\_dom\_model$deviance  
Chi\_hyp\_1\_dom\_model

[1] 3.9667

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_dom\_model <- hyp\_1\_dom\_model$df.null - hyp\_1\_dom\_model$df.residual  
df\_hyp\_1\_dom\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_dom\_model <- 1 - pchisq(Chi\_hyp\_1\_dom\_model, df\_hyp\_1\_dom\_model)  
prob\_Chi\_hyp\_1\_dom\_model

[1] 0.2650824

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 3.967, *p* = .265). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_dom\_model$coefficients)

(Intercept) z\_SR\_dom\_cheng sexfemale   
 1.5257551 1.0182276 0.9755472   
z\_SR\_dom\_cheng:sexfemale   
 1.4130684

### Model 3: Prestige, Sex, and their Interaction

First, we will fit our model.

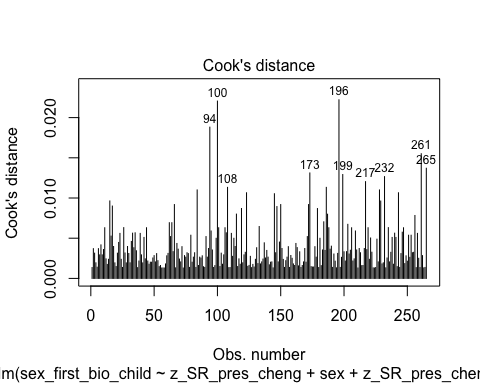
# Fitting the model with self-reported prestige, sex, and their interaction  
hyp\_1\_pres\_model <- glm(sex\_first\_bio\_child ~ z\_SR\_pres\_cheng + sex + z\_SR\_pres\_cheng:sex, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_pres\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_pres\_model <- glm(sex\_first\_bio\_child ~ c\_z\_SR\_pres\_cheng + sex + c\_z\_SR\_pres\_cheng:ln\_c\_z\_SR\_pres\_cheng, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_pres\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_SR\_pres\_cheng + sex +   
 c\_z\_SR\_pres\_cheng:ln\_c\_z\_SR\_pres\_cheng, family = binomial(link = logit),   
 data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -0.643435 18.529248 -0.035 0.972  
c\_z\_SR\_pres\_cheng 0.009025 6.217178 0.001 0.999  
sexfemale -0.088859 0.252152 -0.352 0.725  
c\_z\_SR\_pres\_cheng:ln\_c\_z\_SR\_pres\_cheng 0.042329 1.892193 0.022 0.982  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 356.72 on 261 degrees of freedom  
AIC: 364.72  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .982), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_pres\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_SR\_pres\_cheng + sex + z\_SR\_pres\_cheng:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.42429 0.18380 2.308 0.021 \*  
z\_SR\_pres\_cheng 0.06392 0.17658 0.362 0.717   
sexfemale -0.08512 0.25180 -0.338 0.735   
z\_SR\_pres\_cheng:sexfemale 0.17117 0.25265 0.678 0.498   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 356.26 on 261 degrees of freedom  
AIC: 364.26  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_pres\_model <- hyp\_1\_pres\_model$null.deviance - hyp\_1\_pres\_model$deviance  
Chi\_hyp\_1\_pres\_model

[1] 1.999807

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_pres\_model <- hyp\_1\_pres\_model$df.null - hyp\_1\_pres\_model$df.residual  
df\_hyp\_1\_pres\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_pres\_model <- 1 - pchisq(Chi\_hyp\_1\_pres\_model, df\_hyp\_1\_pres\_model)  
prob\_Chi\_hyp\_1\_pres\_model

[1] 0.5724468

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 2, *p* = .572). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_pres\_model$coefficients)

(Intercept) z\_SR\_pres\_cheng sexfemale   
 1.5285111 1.0660090 0.9184051   
z\_SR\_pres\_cheng:sexfemale   
 1.1866947

### Hypothesis 1 Summary

Overall, this data is inconsistent with Hypothesis 1. While one measure of social status (income) approached significance as a predictor of sex of first born son, dominance was not a significant predictor.

## Study 1 Exploratory Analyses

After our planned analysis of data, the results revealed other potential avenues of fruitful analysis, and there were variables in the data set that were not analyzed due to other variables being more pertinent to tests of our specific Hypothesis 1. Nevertheless, these follow up analyses and inclusion of previously neglected variables may be relevant to testing the broader Trivers-Willard hypothesis (TWH) and maternal dominance hypothesis (MDH). Specifically, this analysis will include the following:

1. A correlation analysis to determine what potential covariates may be controlled for in previous statistical models (binomial logistic regressions including status, sex, and status\*sex) to yield clearer results.
2. Inclusion of these covariate(s) (X) in relevant statistical models (mentioned in 1), to determine whether status predicts OSR (either directly or indirectly through sex as a moderator) after holding X constant.
3. A new binomial logistic regression model including partner-reported dominance and partner-reported prestige with offspring sex as the outcome variable.
4. New binomial logistic regression models (with the same status, sex, and status\*sex format) including previously neglected measures of status or dominance (i.e., partner-reported measures of dominance and prestige, SAT, and IPIP dominance).
5. New binomial logistic regression models to test Hypothesis 1 with composite dominance and prestige based on self-and peer-reported dominance and prestige together.

### Correlational Analysis to Determine Potential Control Variables

To determine whether there are possible covariates to add to bivariate logistic regression models predicting offspring sex, we will look for relationships among the predictor variables we have already assessed. These variables are already saved in a data frame called cont.variables which we used to generate descriptive statistics, so we will use that data frame to create the correlation matrix.

# Using the psych package to generate the correlation table  
study\_1\_cor\_matrix\_covariates <- corr.test(cont.variables)  
  
# Displaying the results  
print(study\_1\_cor\_matrix\_covariates)

Call:corr.test(x = cont.variables)  
Correlation matrix   
 age age\_first\_bio\_child num\_bio\_child  
age 1.00 0.25 0.08  
age\_first\_bio\_child 0.25 1.00 0.47  
num\_bio\_child 0.08 0.47 1.00  
ppp\_adjusted\_income\_usd 0.30 0.08 0.03  
ISEI-08\_occ\_status 0.02 -0.06 -0.06  
SSS 0.16 0.01 0.02  
SR\_dom\_cheng -0.11 -0.03 -0.09  
PR\_dom\_cheng -0.13 0.06 -0.03  
SR\_pres\_cheng -0.13 -0.11 -0.08  
PR\_pres\_cheng -0.06 -0.11 -0.07  
SAT -0.20 0.00 -0.07  
IPIP\_dom -0.02 -0.06 -0.04  
 ppp\_adjusted\_income\_usd ISEI-08\_occ\_status SSS  
age 0.30 0.02 0.16  
age\_first\_bio\_child 0.08 -0.06 0.01  
num\_bio\_child 0.03 -0.06 0.02  
ppp\_adjusted\_income\_usd 1.00 0.37 0.43  
ISEI-08\_occ\_status 0.37 1.00 0.33  
SSS 0.43 0.33 1.00  
SR\_dom\_cheng 0.03 0.01 0.04  
PR\_dom\_cheng -0.14 -0.06 -0.06  
SR\_pres\_cheng 0.13 0.20 0.31  
PR\_pres\_cheng 0.20 0.14 0.20  
SAT 0.03 -0.01 0.00  
IPIP\_dom 0.03 0.06 -0.02  
 SR\_dom\_cheng PR\_dom\_cheng SR\_pres\_cheng PR\_pres\_cheng  
age -0.11 -0.13 -0.13 -0.06  
age\_first\_bio\_child -0.03 0.06 -0.11 -0.11  
num\_bio\_child -0.09 -0.03 -0.08 -0.07  
ppp\_adjusted\_income\_usd 0.03 -0.14 0.13 0.20  
ISEI-08\_occ\_status 0.01 -0.06 0.20 0.14  
SSS 0.04 -0.06 0.31 0.20  
SR\_dom\_cheng 1.00 0.48 0.01 -0.03  
PR\_dom\_cheng 0.48 1.00 0.01 -0.23  
SR\_pres\_cheng 0.01 0.01 1.00 0.35  
PR\_pres\_cheng -0.03 -0.23 0.35 1.00  
SAT 0.22 0.09 0.38 0.18  
IPIP\_dom 0.57 0.32 -0.01 -0.01  
 SAT IPIP\_dom  
age -0.20 -0.02  
age\_first\_bio\_child 0.00 -0.06  
num\_bio\_child -0.07 -0.04  
ppp\_adjusted\_income\_usd 0.03 0.03  
ISEI-08\_occ\_status -0.01 0.06  
SSS 0.00 -0.02  
SR\_dom\_cheng 0.22 0.57  
PR\_dom\_cheng 0.09 0.32  
SR\_pres\_cheng 0.38 -0.01  
PR\_pres\_cheng 0.18 -0.01  
SAT 1.00 0.11  
IPIP\_dom 0.11 1.00  
Sample Size   
 age age\_first\_bio\_child num\_bio\_child  
age 265 265 265  
age\_first\_bio\_child 265 265 265  
num\_bio\_child 265 265 265  
ppp\_adjusted\_income\_usd 212 212 212  
ISEI-08\_occ\_status 247 247 247  
SSS 265 265 265  
SR\_dom\_cheng 265 265 265  
PR\_dom\_cheng 230 230 230  
SR\_pres\_cheng 265 265 265  
PR\_pres\_cheng 230 230 230  
SAT 265 265 265  
IPIP\_dom 265 265 265  
 ppp\_adjusted\_income\_usd ISEI-08\_occ\_status SSS  
age 212 247 265  
age\_first\_bio\_child 212 247 265  
num\_bio\_child 212 247 265  
ppp\_adjusted\_income\_usd 212 200 212  
ISEI-08\_occ\_status 200 247 247  
SSS 212 247 265  
SR\_dom\_cheng 212 247 265  
PR\_dom\_cheng 185 218 230  
SR\_pres\_cheng 212 247 265  
PR\_pres\_cheng 185 218 230  
SAT 212 247 265  
IPIP\_dom 212 247 265  
 SR\_dom\_cheng PR\_dom\_cheng SR\_pres\_cheng PR\_pres\_cheng  
age 265 230 265 230  
age\_first\_bio\_child 265 230 265 230  
num\_bio\_child 265 230 265 230  
ppp\_adjusted\_income\_usd 212 185 212 185  
ISEI-08\_occ\_status 247 218 247 218  
SSS 265 230 265 230  
SR\_dom\_cheng 265 230 265 230  
PR\_dom\_cheng 230 230 230 230  
SR\_pres\_cheng 265 230 265 230  
PR\_pres\_cheng 230 230 230 230  
SAT 265 230 265 230  
IPIP\_dom 265 230 265 230  
 SAT IPIP\_dom  
age 265 265  
age\_first\_bio\_child 265 265  
num\_bio\_child 265 265  
ppp\_adjusted\_income\_usd 212 212  
ISEI-08\_occ\_status 247 247  
SSS 265 265  
SR\_dom\_cheng 265 265  
PR\_dom\_cheng 230 230  
SR\_pres\_cheng 265 265  
PR\_pres\_cheng 230 230  
SAT 265 265  
IPIP\_dom 265 265  
Probability values (Entries above the diagonal are adjusted for multiple tests.)   
 age age\_first\_bio\_child num\_bio\_child  
age 0.00 0.00 1.00  
age\_first\_bio\_child 0.00 0.00 0.00  
num\_bio\_child 0.17 0.00 0.00  
ppp\_adjusted\_income\_usd 0.00 0.27 0.69  
ISEI-08\_occ\_status 0.73 0.38 0.35  
SSS 0.01 0.83 0.70  
SR\_dom\_cheng 0.09 0.60 0.13  
PR\_dom\_cheng 0.06 0.36 0.68  
SR\_pres\_cheng 0.03 0.07 0.20  
PR\_pres\_cheng 0.37 0.09 0.27  
SAT 0.00 0.96 0.27  
IPIP\_dom 0.74 0.36 0.55  
 ppp\_adjusted\_income\_usd ISEI-08\_occ\_status SSS  
age 0.00 1.00 0.39  
age\_first\_bio\_child 1.00 1.00 1.00  
num\_bio\_child 1.00 1.00 1.00  
ppp\_adjusted\_income\_usd 0.00 0.00 0.00  
ISEI-08\_occ\_status 0.00 0.00 0.00  
SSS 0.00 0.00 0.00  
SR\_dom\_cheng 0.66 0.83 0.56  
PR\_dom\_cheng 0.06 0.36 0.34  
SR\_pres\_cheng 0.07 0.00 0.00  
PR\_pres\_cheng 0.01 0.04 0.00  
SAT 0.70 0.84 0.94  
IPIP\_dom 0.63 0.36 0.79  
 SR\_dom\_cheng PR\_dom\_cheng SR\_pres\_cheng PR\_pres\_cheng  
age 1.00 1.00 1.00 1.00  
age\_first\_bio\_child 1.00 1.00 1.00 1.00  
num\_bio\_child 1.00 1.00 1.00 1.00  
ppp\_adjusted\_income\_usd 1.00 1.00 1.00 0.34  
ISEI-08\_occ\_status 1.00 1.00 0.08 1.00  
SSS 1.00 1.00 0.00 0.11  
SR\_dom\_cheng 0.00 0.00 1.00 1.00  
PR\_dom\_cheng 0.00 0.00 1.00 0.02  
SR\_pres\_cheng 0.85 0.87 0.00 0.00  
PR\_pres\_cheng 0.62 0.00 0.00 0.00  
SAT 0.00 0.18 0.00 0.01  
IPIP\_dom 0.00 0.00 0.93 0.87  
 SAT IPIP\_dom  
age 0.06 1  
age\_first\_bio\_child 1.00 1  
num\_bio\_child 1.00 1  
ppp\_adjusted\_income\_usd 1.00 1  
ISEI-08\_occ\_status 1.00 1  
SSS 1.00 1  
SR\_dom\_cheng 0.01 0  
PR\_dom\_cheng 1.00 0  
SR\_pres\_cheng 0.00 1  
PR\_pres\_cheng 0.35 1  
SAT 0.00 1  
IPIP\_dom 0.06 0  
  
 To see confidence intervals of the correlations, print with the short=FALSE option

In the correlation table above, there is one relationship that suggests controlling for this variable may yield clearer results from a previous analysis. Age is moderately and significantly correlated with income (*r* = .30, *p* < .01). Because age could potentially be a negative indicator of condition (with respect to the TWH), by controlling for age in our model using income, sex, and income\*sex as predictors of offspring sex ([Model 1: Income](#model-1-income)), we may be able to further isolate the main effect of income, which was the closest to being a significant predictor in prior analyses despite reduced power (*p* = .18; *n* = 212). I see no other correlations in this table that indicate that inclusion of other covariates would, theoretically, help to clarify our results.

### Rerunning Income Model With Age As Covariate

First, we will fit our model.

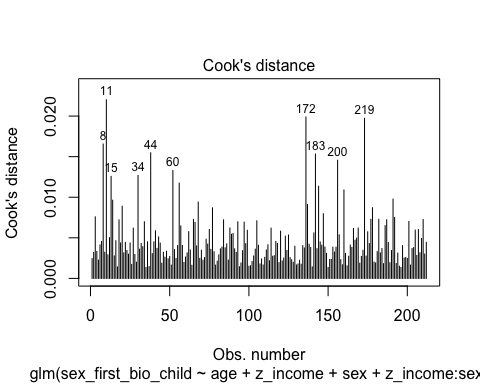
# Fitting the model with age, income, sex, and income\*sex  
income\_status\_model\_2 <- glm(sex\_first\_bio\_child ~ age + z\_income + sex + z\_income:sex, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(income\_status\_model\_2, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because we have already done this for income we only need to make the transformation for age. The code below takes the natural log of age for input into the Box-Tidwell model.

# Creating ln\_age, which represents the natural log of the standardized income scores after adding a constant of 10  
individualsdata$ln\_age <- log(individualsdata$age)

* Now we need to fit and summarize a model with the main effects of the continuous variables with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction terms are not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_income\_model\_2 <- glm(sex\_first\_bio\_child ~ age + c\_z\_income + sex + age:ln\_age + c\_z\_income:ln\_c\_z\_income, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_income\_model\_2)

Call:  
glm(formula = sex\_first\_bio\_child ~ age + c\_z\_income + sex +   
 age:ln\_age + c\_z\_income:ln\_c\_z\_income, family = binomial(link = logit),   
 data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 36.35251 30.10919 1.207 0.227  
age 0.22950 1.21323 0.189 0.850  
c\_z\_income -13.11202 9.85001 -1.331 0.183  
sexfemale -0.03527 0.30111 -0.117 0.907  
age:ln\_age -0.04955 0.26697 -0.186 0.853  
c\_z\_income:ln\_c\_z\_income 4.04351 2.96578 1.363 0.173  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 289.05 on 211 degrees of freedom  
Residual deviance: 280.51 on 206 degrees of freedom  
 (53 observations deleted due to missingness)  
AIC: 292.51  
  
Number of Fisher Scoring iterations: 5

* Because neither interaction term is significant (*p* = .853 and *p* = .173), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(income\_status\_model\_2)

Call:  
glm(formula = sex\_first\_bio\_child ~ age + z\_income + sex + z\_income:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.416495 1.031494 0.404 0.686  
age -0.002445 0.028986 -0.084 0.933  
z\_income 0.277400 0.224029 1.238 0.216  
sexfemale 0.021439 0.302231 0.071 0.943  
z\_income:sexfemale 0.206868 0.332478 0.622 0.534  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 289.05 on 211 degrees of freedom  
Residual deviance: 282.50 on 207 degrees of freedom  
 (53 observations deleted due to missingness)  
AIC: 292.5  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_income\_model\_2 <- income\_status\_model\_2$null.deviance - income\_status\_model\_2$deviance  
Chi\_income\_model\_2

[1] 6.548835

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_income\_model\_2 <- income\_status\_model\_2$df.null - income\_status\_model\_2$df.residual  
df\_income\_model\_2

[1] 4

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_income\_model\_2 <- 1 - pchisq(Chi\_income\_model\_2, df\_income\_model\_2)  
prob\_Chi\_income\_model\_2

[1] 0.1617392

* Looking at the chi-square test, the model is marginally significant (χ²(4) = 6.549, *p* = .162). However, none of the other variables, including income, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(income\_status\_model\_2$coefficients)

(Intercept) age z\_income sexfemale   
 1.5166370 0.9975582 1.3196945 1.0216700   
z\_income:sexfemale   
 1.2298200

### Rerunning Dominance and Prestige Model with Partner-Reported Measures

Before running the dominance and prestige model again with partner-reported measures instead of individual-reported measures, we need to transform our variables to standardize them.

# Standardizing the two predictors  
individualsdata$z\_PR\_dom\_cheng <- scale(individualsdata$PR\_dom\_cheng)  
individualsdata$z\_PR\_pres\_cheng <- scale(individualsdata$PR\_pres\_cheng)

Now, we will fit our model.

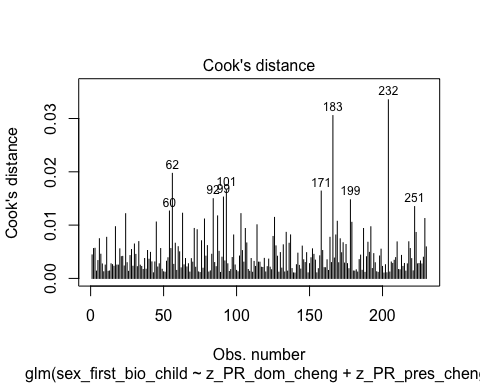
# Fitting the model with self-reported dominance and prestige  
hyp\_1\_main\_effects\_model\_PR <- glm(sex\_first\_bio\_child ~ z\_PR\_dom\_cheng + z\_PR\_pres\_cheng, family = binomial(link = logit), data = individualsdata)

#### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_main\_effects\_model\_PR, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating new variables that add a constant to the standardized variables  
individualsdata$c\_z\_PR\_dom\_cheng <- individualsdata$z\_PR\_dom\_cheng + 10  
individualsdata$c\_z\_PR\_pres\_cheng <- individualsdata$z\_PR\_pres\_cheng + 10  
  
# Creating the natural log transformations of the variables  
individualsdata$ln\_c\_z\_PR\_dom\_cheng <- log(individualsdata$c\_z\_PR\_dom\_cheng)  
individualsdata$ln\_c\_z\_PR\_pres\_cheng <- log(individualsdata$c\_z\_PR\_pres\_cheng)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_main\_effects\_model\_PR <- glm(sex\_first\_bio\_child ~ c\_z\_PR\_dom\_cheng + c\_z\_PR\_pres\_cheng + c\_z\_PR\_dom\_cheng:ln\_c\_z\_PR\_dom\_cheng + c\_z\_PR\_pres\_cheng:ln\_c\_z\_PR\_pres\_cheng, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_main\_effects\_model\_PR)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_PR\_dom\_cheng + c\_z\_PR\_pres\_cheng +   
 c\_z\_PR\_dom\_cheng:ln\_c\_z\_PR\_dom\_cheng + c\_z\_PR\_pres\_cheng:ln\_c\_z\_PR\_pres\_cheng,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -36.260 32.088 -1.130 0.258  
c\_z\_PR\_dom\_cheng 3.732 7.996 0.467 0.641  
c\_z\_PR\_pres\_cheng 7.675 7.583 1.012 0.312  
c\_z\_PR\_dom\_cheng:ln\_c\_z\_PR\_dom\_cheng -1.065 2.409 -0.442 0.658  
c\_z\_PR\_pres\_cheng:ln\_c\_z\_PR\_pres\_cheng -2.289 2.312 -0.990 0.322  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 304.95 on 225 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 314.95  
  
Number of Fisher Scoring iterations: 4

* Because neither of the interaction terms are significant (*p* = .658 and *p* = .322), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

#### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_main\_effects\_model\_PR)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_PR\_dom\_cheng + z\_PR\_pres\_cheng,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4116 0.1358 3.032 0.00243 \*\*  
z\_PR\_dom\_cheng 0.2300 0.1425 1.615 0.10638   
z\_PR\_pres\_cheng 0.1687 0.1395 1.209 0.22651   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 306.19 on 227 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 312.19  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_main\_effects\_model\_PR <- hyp\_1\_main\_effects\_model\_PR$null.deviance - hyp\_1\_main\_effects\_model\_PR$deviance  
Chi\_hyp\_1\_main\_effects\_model\_PR

[1] 3.393547

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_main\_effects\_model\_PR <- hyp\_1\_main\_effects\_model\_PR$df.null - hyp\_1\_main\_effects\_model\_PR$df.residual  
df\_hyp\_1\_main\_effects\_model\_PR

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_main\_effects\_model\_PR <- 1 - pchisq(Chi\_hyp\_1\_main\_effects\_model\_PR, df\_hyp\_1\_main\_effects\_model\_PR)  
prob\_Chi\_hyp\_1\_main\_effects\_model\_PR

[1] 0.1832739

* Looking at the chi-square test, the entire model is not significant (χ²(2) = 3.394, *p* = .183). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_main\_effects\_model\_PR$coefficients)

(Intercept) z\_PR\_dom\_cheng z\_PR\_pres\_cheng   
 1.509201 1.258619 1.183724

### Rerunning Status Models with Previously Neglected Measures of Status or Dominance

Now we will run new binomial logistic regression models (with the same status, sex, and status\*sex format) while including previously neglected measures of status or dominance (i.e., partner-reported measures of dominance and prestige, the SAT, and the IPIP dominance).

#### Partner-Reported Dominance, Sex, and their Interaction

First, we will fit our model.

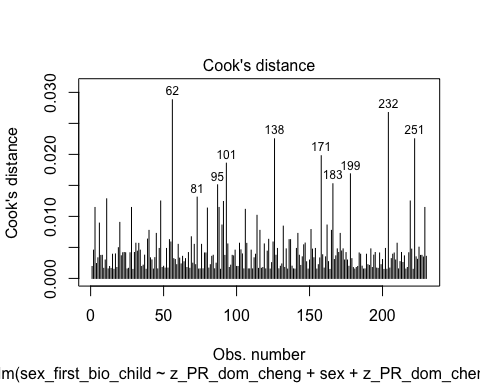
# Fitting the model with self-reported dominance, sex, and their interaction  
hyp\_1\_dom\_model\_PR <- glm(sex\_first\_bio\_child ~ z\_PR\_dom\_cheng + sex + z\_PR\_dom\_cheng:sex, family = binomial(link = logit), data = individualsdata)

##### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_dom\_model\_PR, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_dom\_model\_PR <- glm(sex\_first\_bio\_child ~ c\_z\_PR\_dom\_cheng + sex + c\_z\_PR\_dom\_cheng:ln\_c\_z\_PR\_dom\_cheng, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_dom\_model\_PR)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_PR\_dom\_cheng + sex +   
 c\_z\_PR\_dom\_cheng:ln\_c\_z\_PR\_dom\_cheng, family = binomial(link = logit),   
 data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -11.9067 24.1208 -0.494 0.622  
c\_z\_PR\_dom\_cheng 3.6315 7.9103 0.459 0.646  
sexfemale -0.1010 0.2708 -0.373 0.709  
c\_z\_PR\_dom\_cheng:ln\_c\_z\_PR\_dom\_cheng -1.0378 2.3840 -0.435 0.663  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 307.33 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 315.33  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .663), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

##### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_dom\_model\_PR)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_PR\_dom\_cheng + sex + z\_PR\_dom\_cheng:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.45847 0.19488 2.353 0.0186 \*  
z\_PR\_dom\_cheng -0.03676 0.19625 -0.187 0.8514   
sexfemale -0.08571 0.27247 -0.315 0.7531   
z\_PR\_dom\_cheng:sexfemale 0.44030 0.27840 1.582 0.1138   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 304.99 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 312.99  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_dom\_model\_PR <- hyp\_1\_dom\_model\_PR$null.deviance - hyp\_1\_dom\_model\_PR$deviance  
Chi\_hyp\_1\_dom\_model\_PR

[1] 4.59687

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_dom\_model\_PR <- hyp\_1\_dom\_model\_PR$df.null - hyp\_1\_dom\_model\_PR$df.residual  
df\_hyp\_1\_dom\_model\_PR

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_dom\_model\_PR <- 1 - pchisq(Chi\_hyp\_1\_dom\_model\_PR, df\_hyp\_1\_dom\_model\_PR)  
prob\_Chi\_hyp\_1\_dom\_model\_PR

[1] 0.2038108

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 4.597, *p* = .204). In addition, none of the other variables, except for the intercept, are significant, although the interaction between partner-reported dominance and sex approaches significance. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_dom\_model\_PR$coefficients)

(Intercept) z\_PR\_dom\_cheng sexfemale   
 1.5816461 0.9639096 0.9178589   
z\_PR\_dom\_cheng:sexfemale   
 1.5531697

* Although the odds ratio for the interaction between partner-reported dominance and sex suggests a strong effect size, this term is still not significant (p = .085).

#### Partner-Reported Prestige, Sex, and their Interaction

First, we will fit our model.

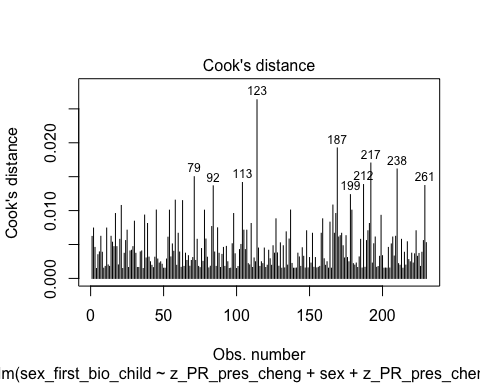
# Fitting the model with self-reported prestige, sex, and their interaction  
hyp\_1\_pres\_model\_PR <- glm(sex\_first\_bio\_child ~ z\_PR\_pres\_cheng + sex + z\_PR\_pres\_cheng:sex, family = binomial(link = logit), data = individualsdata)

##### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_pres\_model\_PR, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_pres\_model\_PR <- glm(sex\_first\_bio\_child ~ c\_z\_PR\_pres\_cheng + sex + c\_z\_PR\_pres\_cheng:ln\_c\_z\_PR\_pres\_cheng, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_pres\_model\_PR)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_PR\_pres\_cheng + sex +   
 c\_z\_PR\_pres\_cheng:ln\_c\_z\_PR\_pres\_cheng, family = binomial(link = logit),   
 data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -30.09061 22.07674 -1.363 0.173  
c\_z\_PR\_pres\_cheng 10.01869 7.41209 1.352 0.176  
sexfemale -0.04448 0.27393 -0.162 0.871  
c\_z\_PR\_pres\_cheng:ln\_c\_z\_PR\_pres\_cheng -3.01890 2.25705 -1.338 0.181  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 306.96 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 314.96  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .181), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

##### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_pres\_model\_PR)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_PR\_pres\_cheng + sex + z\_PR\_pres\_cheng:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.46977 0.19746 2.379 0.0174 \*  
z\_PR\_pres\_cheng -0.08243 0.21256 -0.388 0.6982   
sexfemale -0.07759 0.27407 -0.283 0.7771   
z\_PR\_pres\_cheng:sexfemale 0.32853 0.27808 1.181 0.2374   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 307.38 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 315.38  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_pres\_model\_PR <- hyp\_1\_pres\_model\_PR$null.deviance - hyp\_1\_pres\_model\_PR$deviance  
Chi\_hyp\_1\_pres\_model\_PR

[1] 2.204522

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_pres\_model\_PR <- hyp\_1\_pres\_model\_PR$df.null - hyp\_1\_pres\_model\_PR$df.residual  
df\_hyp\_1\_pres\_model\_PR

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_pres\_model\_PR <- 1 - pchisq(Chi\_hyp\_1\_pres\_model\_PR, df\_hyp\_1\_pres\_model\_PR)  
prob\_Chi\_hyp\_1\_pres\_model\_PR

[1] 0.5310582

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 2.205, *p* = .531). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_pres\_model\_PR$coefficients)

(Intercept) z\_PR\_pres\_cheng sexfemale   
 1.5996207 0.9208733 0.9253400   
z\_PR\_pres\_cheng:sexfemale   
 1.3889277

#### SAT, Sex, and their Interaction

Now we will look for Trivers-Willard effects with the SAT as our measure of dominance. First, we need to standardize the SAT variable.

# Creating standardized scores for SAT  
individualsdata$z\_SAT <- scale(individualsdata$SAT)

Now, we will fit our model.

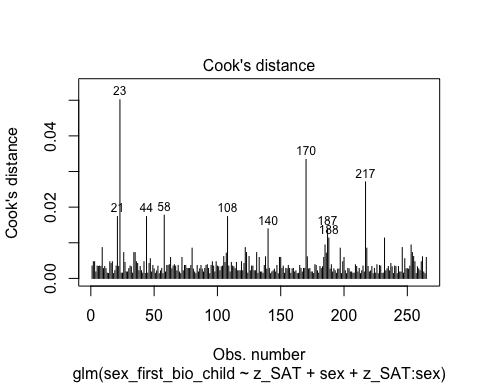
# Fitting the model with standardized SAT, sex, and their interaction  
SAT\_status\_model <- glm(sex\_first\_bio\_child ~ z\_SAT + sex + z\_SAT:sex, family = binomial(link = logit), data = individualsdata)

##### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(SAT\_status\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating c\_z\_SAT, which represents standardized SAT after adding a constant of 10  
individualsdata$c\_z\_SAT <- individualsdata$z\_SAT + 10  
  
# Creating ln\_c\_z\_SAT, which represents the natural log of the standardized SAT scores after adding a constant of 10  
individualsdata$ln\_c\_z\_SAT <- log(individualsdata$c\_z\_SAT)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_SAT\_status\_model <- glm(sex\_first\_bio\_child ~ c\_z\_SAT + sex + c\_z\_SAT:ln\_c\_z\_SAT, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_SAT\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_SAT + sex + c\_z\_SAT:ln\_c\_z\_SAT,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 26.39562 22.98298 1.148 0.251  
c\_z\_SAT -8.36190 7.39522 -1.131 0.258  
sexfemale -0.09818 0.25188 -0.390 0.697  
c\_z\_SAT:ln\_c\_z\_SAT 2.49858 2.21030 1.130 0.258  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 356.76 on 261 degrees of freedom  
AIC: 364.76  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .258), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

##### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(SAT\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_SAT + sex + z\_SAT:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.42293 0.18398 2.299 0.0215 \*  
z\_SAT 0.04748 0.18091 0.262 0.7930   
sexfemale -0.09580 0.25116 -0.381 0.7029   
z\_SAT:sexfemale -0.08664 0.25150 -0.345 0.7305   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 357.99 on 261 degrees of freedom  
AIC: 365.99  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_SAT\_status\_model <- SAT\_status\_model$null.deviance - SAT\_status\_model$deviance  
Chi\_SAT\_status\_model

[1] 0.2674562

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_SAT\_status\_model <- SAT\_status\_model$df.null - SAT\_status\_model$df.residual  
df\_SAT\_status\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_SAT\_status\_model <- 1 - pchisq(Chi\_SAT\_status\_model, df\_SAT\_status\_model)  
prob\_Chi\_SAT\_status\_model

[1] 0.9660282

* Looking at the chi-square test, the entire model is not significant (χ²(3) = .267, *p* = .966). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(SAT\_status\_model$coefficients)

(Intercept) z\_SAT sexfemale z\_SAT:sexfemale   
 1.5264269 1.0486207 0.9086413 0.9170046

#### IPIP Dominance, Sex, and their Interaction

Now we will look for Trivers-Willard effects with the IPIP\_dom as our measure of dominance. First, we need to standardize the IPIP\_dom variable.

# Creating standardized scores for IPIP\_dom  
individualsdata$z\_IPIP <- scale(individualsdata$IPIP\_dom)

Now, we will fit our model.

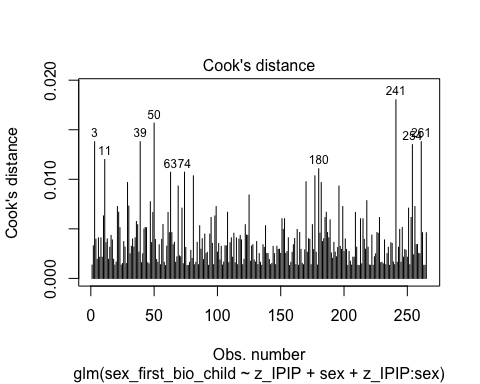
# Fitting the model with standardized IPIP, sex, and their interaction  
IPIP\_status\_model <- glm(sex\_first\_bio\_child ~ z\_IPIP + sex + z\_IPIP:sex, family = binomial(link = logit), data = individualsdata)

##### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(IPIP\_status\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating c\_z\_IPIP, which represents standardized IPIP after adding a constant of 10  
individualsdata$c\_z\_IPIP <- individualsdata$z\_IPIP + 10  
  
# Creating ln\_c\_z\_IPIP, which represents the natural log of the standardized IPIP scores after adding a constant of 10  
individualsdata$ln\_c\_z\_IPIP <- log(individualsdata$c\_z\_IPIP)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_IPIP\_status\_model <- glm(sex\_first\_bio\_child ~ c\_z\_IPIP + sex + c\_z\_IPIP:ln\_c\_z\_IPIP, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_IPIP\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_IPIP + sex + c\_z\_IPIP:ln\_c\_z\_IPIP,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 32.0576 21.8271 1.469 0.142  
c\_z\_IPIP -10.4439 7.2274 -1.445 0.148  
sexfemale -0.1106 0.2604 -0.425 0.671  
c\_z\_IPIP:ln\_c\_z\_IPIP 3.1557 2.1878 1.442 0.149  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 355.92 on 261 degrees of freedom  
AIC: 363.92  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .149), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

##### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(IPIP\_status\_model)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_IPIP + sex + z\_IPIP:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4738 0.1936 2.447 0.0144 \*  
z\_IPIP -0.1633 0.1920 -0.850 0.3951   
sexfemale -0.1183 0.2625 -0.451 0.6522   
z\_IPIP:sexfemale 0.2672 0.2620 1.020 0.3079   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 358.26 on 264 degrees of freedom  
Residual deviance: 357.04 on 261 degrees of freedom  
AIC: 365.04  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_IPIP\_status\_model <- IPIP\_status\_model$null.deviance - IPIP\_status\_model$deviance  
Chi\_IPIP\_status\_model

[1] 1.217753

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_IPIP\_status\_model <- IPIP\_status\_model$df.null - IPIP\_status\_model$df.residual  
df\_IPIP\_status\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_IPIP\_status\_model <- 1 - pchisq(Chi\_IPIP\_status\_model, df\_IPIP\_status\_model)  
prob\_Chi\_IPIP\_status\_model

[1] 0.7487497

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 1.21, *p* = .749). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(IPIP\_status\_model$coefficients)

(Intercept) z\_IPIP sexfemale z\_IPIP:sexfemale   
 1.6060863 0.8493766 0.8884259 1.3062380

### Rerun Models Testing Hypothesis 1 with Composite Dominance and Prestige

Originally, we planned to only make composite measures of dominance and prestige if the self- and partner-reported measures correlated at r = .65 or higher ([Assessing Whether to Create Behavioral Dominance Composite](#Xd18e483d5adc6c333f9c13ac3b091d94792667f)). However, this may have been a more stringent and arbitrary threshold than necessary. Here, we will rerun the models from [Testing Hypothesis 1](#testing-hypothesis-1) with composite versions of dominance and prestige.

#### Creating the Composite Variables

First, we need to create the composite variables. To do this, we will simply take the mean of self- and partner-reported dominance and prestige, respectively.

# Taking the mean of self- and partner-reported measures as the composite measures  
individualsdata$comp\_dom <- (individualsdata$SR\_dom\_cheng + individualsdata$PR\_dom\_cheng)/2  
individualsdata$comp\_pres <- (individualsdata$SR\_pres\_cheng + individualsdata$PR\_pres\_cheng)/2

Next, we will standardize both of the variables.

# Standardizing the composite variables  
individualsdata$z\_comp\_dom <- scale(individualsdata$comp\_dom)  
individualsdata$z\_comp\_pres <- scale(individualsdata$comp\_pres)

#### Rerunning the Models

##### Model 1: Dominance and Prestige Main Effects Model

First, we will fit our model.

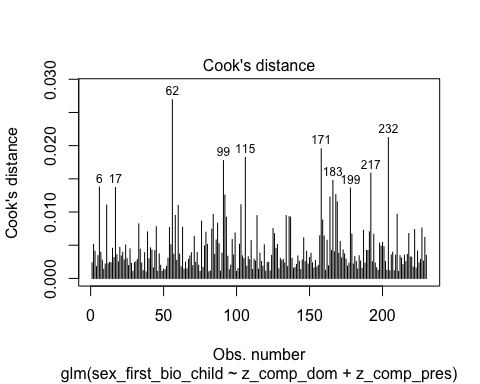
# Fitting the model with composite dominance and prestige  
hyp\_1\_main\_effects\_model\_comp <- glm(sex\_first\_bio\_child ~ z\_comp\_dom + z\_comp\_pres, family = binomial(link = logit), data = individualsdata)

###### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_main\_effects\_model\_comp, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our continuous predictor variable to be positive by adding a constant. The code below adds a constant of 10 that makes all values for the variable positive before making the natural log transformation.

# Creating new variables that add a constant to the standardized variables  
individualsdata$c\_z\_comp\_dom <- individualsdata$z\_comp\_dom + 10  
individualsdata$c\_z\_comp\_pres <- individualsdata$z\_comp\_pres + 10  
  
# Creating the natural log transformations of the variables  
individualsdata$ln\_c\_z\_comp\_dom <- log(individualsdata$c\_z\_comp\_dom)  
individualsdata$ln\_c\_z\_comp\_pres <- log(individualsdata$c\_z\_comp\_pres)

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_main\_effects\_model\_comp <- glm(sex\_first\_bio\_child ~ c\_z\_comp\_dom + c\_z\_comp\_pres + c\_z\_comp\_dom:ln\_c\_z\_comp\_dom + c\_z\_comp\_pres:ln\_c\_z\_comp\_pres, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_main\_effects\_model\_comp)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_comp\_dom + c\_z\_comp\_pres +   
 c\_z\_comp\_dom:ln\_c\_z\_comp\_dom + c\_z\_comp\_pres:ln\_c\_z\_comp\_pres,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -10.14581 31.78512 -0.319 0.750  
c\_z\_comp\_dom 2.04702 7.09859 0.288 0.773  
c\_z\_comp\_pres 0.51831 7.67933 0.067 0.946  
c\_z\_comp\_dom:ln\_c\_z\_comp\_dom -0.56063 2.13394 -0.263 0.793  
c\_z\_comp\_pres:ln\_c\_z\_comp\_pres -0.09356 2.33369 -0.040 0.968  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 305.74 on 225 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 315.74  
  
Number of Fisher Scoring iterations: 4

* Because neither of the interaction terms are significant (*p* = .793 and *p* = .968), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

###### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_main\_effects\_model\_comp)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_comp\_dom + z\_comp\_pres,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4124 0.1359 3.035 0.00241 \*\*  
z\_comp\_dom 0.1837 0.1396 1.316 0.18817   
z\_comp\_pres 0.2068 0.1367 1.513 0.13018   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 305.81 on 227 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 311.81  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_main\_effects\_model\_comp <- hyp\_1\_main\_effects\_model\_comp$null.deviance - hyp\_1\_main\_effects\_model\_comp$deviance  
Chi\_hyp\_1\_main\_effects\_model\_comp

[1] 3.775693

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_main\_effects\_model\_comp <- hyp\_1\_main\_effects\_model\_comp$df.null - hyp\_1\_main\_effects\_model\_comp$df.residual  
df\_hyp\_1\_main\_effects\_model\_comp

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_main\_effects\_model\_comp <- 1 - pchisq(Chi\_hyp\_1\_main\_effects\_model\_comp, df\_hyp\_1\_main\_effects\_model\_comp)  
prob\_Chi\_hyp\_1\_main\_effects\_model\_comp

[1] 0.1513975

* Looking at the chi-square test, the entire model is not significant (χ²(2) = 3.776, *p* = .151). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_main\_effects\_model\_comp$coefficients)

(Intercept) z\_comp\_dom z\_comp\_pres   
 1.510399 1.201669 1.229773

* In contrast to our hypothesis, the odds-ratio for prestige is higher than the odds-ratio for dominance, although neither are significant predictors.

##### Model 2: Dominance, Sex, and their Interaction

First, we will fit our model.

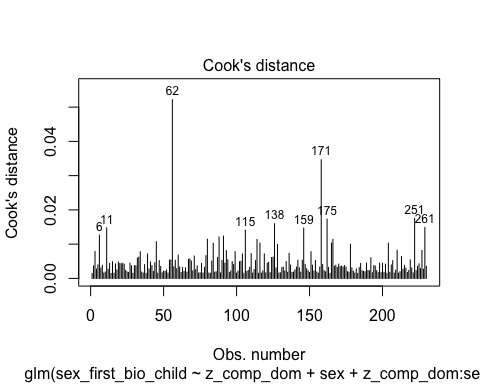
# Fitting the model with composite dominance, sex, and their interaction  
hyp\_1\_dom\_model\_comp <- glm(sex\_first\_bio\_child ~ z\_comp\_dom + sex + z\_comp\_dom:sex, family = binomial(link = logit), data = individualsdata)

###### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_dom\_model\_comp, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_dom\_model\_comp <- glm(sex\_first\_bio\_child ~ c\_z\_comp\_dom + sex + c\_z\_comp\_dom:ln\_c\_z\_comp\_dom, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_dom\_model\_comp)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_comp\_dom + sex + c\_z\_comp\_dom:ln\_c\_z\_comp\_dom,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -3.70595 21.52091 -0.172 0.863  
c\_z\_comp\_dom 0.99028 7.01846 0.141 0.888  
sexfemale -0.07188 0.27167 -0.265 0.791  
c\_z\_comp\_dom:ln\_c\_z\_comp\_dom -0.24924 2.11020 -0.118 0.906  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 308.04 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 316.04  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .906), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

###### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_dom\_model\_comp)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_comp\_dom + sex + z\_comp\_dom:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.46374 0.19606 2.365 0.018 \*  
z\_comp\_dom -0.05059 0.18774 -0.269 0.788   
sexfemale -0.05120 0.27525 -0.186 0.852   
z\_comp\_dom:sexfemale 0.46692 0.28549 1.636 0.102   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 305.31 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 313.31  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_dom\_model\_comp <- hyp\_1\_dom\_model\_comp$null.deviance - hyp\_1\_dom\_model\_comp$deviance  
Chi\_hyp\_1\_dom\_model\_comp

[1] 4.275591

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_dom\_model\_comp <- hyp\_1\_dom\_model\_comp$df.null - hyp\_1\_dom\_model\_comp$df.residual  
df\_hyp\_1\_dom\_model\_comp

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_dom\_model\_comp <- 1 - pchisq(Chi\_hyp\_1\_dom\_model\_comp, df\_hyp\_1\_dom\_model\_comp)  
prob\_Chi\_hyp\_1\_dom\_model\_comp

[1] 0.233202

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 4.276, *p* = .102). In addition, none of the other variables, except for the intercept, are significant, although the interaction with sex is marginally significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_dom\_model\_comp$coefficients)

(Intercept) z\_comp\_dom sexfemale   
 1.5900127 0.9506681 0.9500882   
z\_comp\_dom:sexfemale   
 1.5950722

* Although the odds-ratio for the interaction term is a pretty large effect size, it is still not a significant predictor.

##### Model 3: Prestige, Sex, and their Interaction

First, we will fit our model.

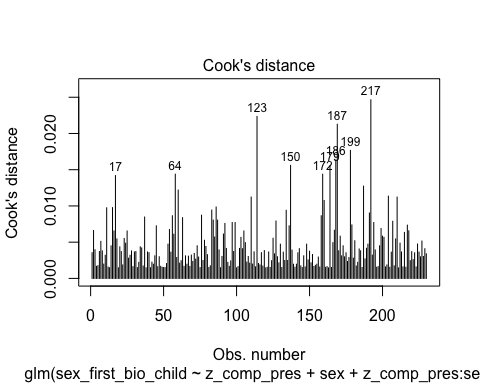
# Fitting the model with self-reported prestige, sex, and their interaction  
hyp\_1\_pres\_model\_comp <- glm(sex\_first\_bio\_child ~ z\_comp\_pres + sex + z\_comp\_pres:sex, family = binomial(link = logit), data = individualsdata)

###### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_1\_pres\_model\_comp, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* Now we need to fit and summarize a model with the main effects of the continuous variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that the interaction term is not significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the Box-Tiddwell procedure  
BT\_test\_hyp\_1\_pres\_model\_comp <- glm(sex\_first\_bio\_child ~ c\_z\_comp\_pres + sex + c\_z\_comp\_pres:ln\_c\_z\_comp\_pres, family = binomial(link = logit), data = individualsdata)  
# Summarizing the model  
summary(BT\_test\_hyp\_1\_pres\_model\_comp)

Call:  
glm(formula = sex\_first\_bio\_child ~ c\_z\_comp\_pres + sex + c\_z\_comp\_pres:ln\_c\_z\_comp\_pres,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -7.31548 22.54720 -0.324 0.746  
c\_z\_comp\_pres 2.15340 7.53004 0.286 0.775  
sexfemale -0.06347 0.27207 -0.233 0.816  
c\_z\_comp\_pres:ln\_c\_z\_comp\_pres -0.59702 2.28756 -0.261 0.794  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 307.45 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 315.45  
  
Number of Fisher Scoring iterations: 4

* Because the interaction term is not significant (*p* = .794), the assumption of a linear relationship between continuous IVs and the logit of the DV is satisfied.

###### Summary of the Model

Now we will summarize the model, along with the Chi-square test for whether the model fits significantly better than the model with only the intercept.

# Producing the summary of the model  
summary(hyp\_1\_pres\_model\_comp)

Call:  
glm(formula = sex\_first\_bio\_child ~ z\_comp\_pres + sex + z\_comp\_pres:sex,   
 family = binomial(link = logit), data = individualsdata)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.45309 0.19573 2.315 0.0206 \*  
z\_comp\_pres 0.05463 0.20021 0.273 0.7850   
sexfemale -0.05960 0.27304 -0.218 0.8272   
z\_comp\_pres:sexfemale 0.25000 0.27487 0.910 0.3631   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 309.59 on 229 degrees of freedom  
Residual deviance: 306.69 on 226 degrees of freedom  
 (35 observations deleted due to missingness)  
AIC: 314.69  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_1\_pres\_model\_comp <- hyp\_1\_pres\_model\_comp$null.deviance - hyp\_1\_pres\_model\_comp$deviance  
Chi\_hyp\_1\_pres\_model\_comp

[1] 2.893484

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_1\_pres\_model\_comp <- hyp\_1\_pres\_model\_comp$df.null - hyp\_1\_pres\_model\_comp$df.residual  
df\_hyp\_1\_pres\_model\_comp

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_1\_pres\_model\_comp <- 1 - pchisq(Chi\_hyp\_1\_pres\_model\_comp, df\_hyp\_1\_pres\_model\_comp)  
prob\_Chi\_hyp\_1\_pres\_model\_comp

[1] 0.4083411

* Looking at the chi-square test, the entire model is not significant (χ²(3) = 2.893, p = .408). In addition, none of the other variables, except for the intercept, are significant. I will exponentiate the coefficients to reveal the odds-ratio of the coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_1\_pres\_model\_comp$coefficients)

(Intercept) z\_comp\_pres sexfemale   
 1.5731716 1.0561515 0.9421389   
z\_comp\_pres:sexfemale   
 1.2840305

#### Summary of the Composite Models

Like the models that used only self-reported dominance and prestige, these models do not provide support for Hypothesis 1.

## Clean-Up From Study 1 Analysis

Before going on to Study 2, we will remove unnecessary objects in the environment, write the individualsdata data frame to the hard drive as data\_after\_study\_1\_analysis.csv, and delete unnecessary variables we have created from the individualsdata data frame.

# Creating a vector containing the names of all objects in the environment  
objects <- ls()  
  
# Removing all objects except for the "individualsdata" data frame  
rm(list = objects[objects != "individualsdata"])  
rm("objects")  
  
# Writing the individualsdata data frame to the data folder  
write.csv(individualsdata, file = "./data/data\_after\_study\_1\_analysis.csv")  
  
# Remove the variables we have created from the individualsdata data frame  
individualsdata <- subset(individualsdata, select = ID:photo\_method)

# Study 2 Analysis

## Reliability of Facial Ratings and Facial Width-to-Height Ratio

### Data Cleaning for Ratings

Before we can proceed with analyses, we must calculate the average facial ratings for dominance, masculinity/femininity, and attractiveness, but, first, we will check the reliability of the facial ratings. First, let’s read in the data for the facial ratings and take a look at it.

# Reading in the data as the data frame ratingsdata  
ratingsdata <- read\_excel("./data/Clean\_Ratings\_Data.xlsx", sheet = 1)

The raters are in rows, and the facial images identifiers are in the columns (after the demographic data). Also, the dominance ratings, masculinity/femininity ratings, and attractiveness ratings are all in the same data frame here. To allow for easier reliability analysis, I will clean this up a bit. In the following chunks, I will create a data frame that isolates each of the different types of ratings with the ID as the row identifier. I will save the resultant data frames as CSV files in the data folder.

* For the dominance ratings:
* # Extract the relevant columns for dominance ratings using grep and transpose the data as transposed\_dominance\_data  
  transposed\_dominance\_data <- as.data.frame(  
   t(  
   ratingsdata[, grep("dominance", names(ratingsdata))]  
   )  
  )  
    
  # Remove extraneous features of the row names of the transposed\_dominance\_data data frame to only include the ID  
  row.names(transposed\_dominance\_data) <- gsub("dominance\_[0-9]+\_[0-9]+\_", "", row.names(transposed\_dominance\_data))  
    
  # Write the transposed data to a CSV file called transposed\_dominance\_data.csv with empty strings for missing values  
  write.csv(transposed\_dominance\_data, "./data/transposed\_dominance\_data.csv", row.names = TRUE, na = "")
* For the masculinity/femininity ratings:
* # Extract the relevant columns for masculinity/femininity ratings using grep and transpose the data as transposed\_masculinityfemininity\_data (note that I accidently spelled "femininity" as "feminity" in the Python script that cleaned up this dataset before importing the ratings data, which is why the grep function has "masculinity/feminity" in it here)  
  transposed\_masculinityfemininity\_data <- as.data.frame(  
   t(  
   ratingsdata[, grep("masculinity/feminity", names(ratingsdata))]  
   )  
  )  
    
  # Remove extraneous features of the row names of the transposed\_masculinityfemininity\_data data frame to only include the ID (note that I accidently spelled "femininity" as "feminity" in the Python script that cleaned up this dataset before importing the ratings data, which is why the gsub function has "masculinity/feminity" in it here)  
  row.names(transposed\_masculinityfemininity\_data) <- gsub("masculinity/feminity\_[0-9]+\_[0-9]+\_", "", row.names(transposed\_masculinityfemininity\_data))  
    
  # Write the transposed data to a CSV file called transposed\_masculinityfemininity\_data.csv with empty strings for missing values  
  write.csv(transposed\_masculinityfemininity\_data, "./data/transposed\_masculinityfemininity\_data.csv", row.names = TRUE, na = "")
* For the attractiveness ratings:
* # Extract the relevant columns for attractiveness ratings using grep and transpose the data as transposed\_attractiveness\_data  
  transposed\_attractiveness\_data <- as.data.frame(  
   t(  
   ratingsdata[, grep("attractiveness", names(ratingsdata))]  
   )  
  )  
    
  # Remove extraneous features of the row names of the transposed\_attractiveness\_data data frame to only include the ID  
  row.names(transposed\_attractiveness\_data) <- gsub("attractiveness\_[0-9]+\_[0-9]+\_", "", row.names(transposed\_attractiveness\_data))  
    
  # Write the transposed data to a CSV file called transposed\_attractiveness\_data.csv with empty strings for missing values  
  write.csv(transposed\_attractiveness\_data, "./data/transposed\_attractiveness\_data.csv", row.names = TRUE, na = "")

### Reliability of Facial Ratings Within Types:

Now that we have separated our data out into the different types of ratings, we can assess the degree of inter-rater agreement within rating types using intraclass correlation coefficients (ICCs) with the ‘psych’ package (Revelle & Revelle, 2015). The documentation of the psych package indicates that it uses the six varieties of ICCs outlined by Shrout & Fleiss (1979). For our purposes, we are (1) not interested in generalizing to the population of raters but rather the population of faces, and we are looking for (2) the consistency of ratings to calculate (3) average scores. Therefore, we will use the ICC(3,k), which is based on a two-way mixed model. The ICC(3,k) is actually identical to Chronbach’s alpha.

* For the dominance ratings:
* # Calculate ICC using a two-way mixed-effects model for average ratings and consistency  
  icc\_result\_dominance <- ICC(transposed\_dominance\_data)  
    
  # View the result  
  print(icc\_result\_dominance)
* Call: ICC(x = transposed\_dominance\_data)  
    
  Intraclass correlation coefficients   
   type ICC F df1 df2 p lower bound upper bound  
  Single\_raters\_absolute ICC1 0.12 43 255 76288 0 0.11 0.15  
  Single\_random\_raters ICC2 0.12 56 255 75990 0 0.11 0.15  
  Single\_fixed\_raters ICC3 0.16 56 255 75990 0 0.13 0.18  
  Average\_raters\_absolute ICC1k 0.98 43 255 76288 0 0.97 0.98  
  Average\_random\_raters ICC2k 0.98 56 255 75990 0 0.97 0.98  
  Average\_fixed\_raters ICC3k 0.98 56 255 75990 0 0.98 0.99  
    
   Number of subjects = 256 Number of Judges = 299  
  See the help file for a discussion of the other 4 McGraw and Wong estimates,
* For masculinity/femininity ratings:
* # Calculate ICC using a two-way mixed-effects model for average ratings and consistency  
  icc\_result\_masculinityfemininity <- ICC(transposed\_masculinityfemininity\_data)  
    
  # View the result  
  print(icc\_result\_masculinityfemininity)
* Call: ICC(x = transposed\_masculinityfemininity\_data)  
    
  Intraclass correlation coefficients   
   type ICC F df1 df2 p lower bound upper bound  
  Single\_raters\_absolute ICC1 0.28 119 255 76288 0 0.25 0.32  
  Single\_random\_raters ICC2 0.28 235 255 75990 0 0.25 0.33  
  Single\_fixed\_raters ICC3 0.44 235 255 75990 0 0.40 0.48  
  Average\_raters\_absolute ICC1k 0.99 119 255 76288 0 0.99 0.99  
  Average\_random\_raters ICC2k 0.99 235 255 75990 0 0.99 0.99  
  Average\_fixed\_raters ICC3k 1.00 235 255 75990 0 0.99 1.00  
    
   Number of subjects = 256 Number of Judges = 299  
  See the help file for a discussion of the other 4 McGraw and Wong estimates,
* For attractiveness ratings:
* # Calculate ICC using a two-way mixed-effects model for average ratings and consistency  
  icc\_result\_attractiveness <- ICC(transposed\_attractiveness\_data)  
    
  # View the result  
  print(icc\_result\_attractiveness)
* Call: ICC(x = transposed\_attractiveness\_data)  
    
  Intraclass correlation coefficients   
   type ICC F df1 df2 p lower bound upper bound  
  Single\_raters\_absolute ICC1 0.16 59 255 76288 0 0.14 0.19  
  Single\_random\_raters ICC2 0.16 104 255 75990 0 0.14 0.19  
  Single\_fixed\_raters ICC3 0.26 104 255 75990 0 0.22 0.29  
  Average\_raters\_absolute ICC1k 0.98 59 255 76288 0 0.98 0.99  
  Average\_random\_raters ICC2k 0.98 104 255 75990 0 0.98 0.99  
  Average\_fixed\_raters ICC3k 0.99 104 255 75990 0 0.99 0.99  
    
   Number of subjects = 256 Number of Judges = 299  
  See the help file for a discussion of the other 4 McGraw and Wong estimates,

These ICCs (all > .98) are very high, indicating strong inter-rater reliability of facial ratings. Thus, we will calculate the average rating for each face for later analyses.

### Calculating Average Facial Ratings

The following will calculate average facial ratings for each type of rating and add them to the main dataset structured with individuals, individualsdata.

# Calculate row means for all columns from the transposed datasets and ignore missing values (the IDs are in the row number positions, so we do not need to worry about starting with the second column for calculating the means)  
row\_meansd <- rowMeans(transposed\_dominance\_data, na.rm = TRUE)  
row\_meansmf <- rowMeans(transposed\_masculinityfemininity\_data, na.rm = TRUE)  
row\_meansa <- rowMeans(transposed\_attractiveness\_data, na.rm = TRUE)  
  
# Create means ratings data frames with IDs (from the transposed data row names) and their corresponding means  
mean\_ratings\_df\_d <- data.frame(ID = rownames(transposed\_dominance\_data), facial\_dominance = row\_meansd)  
mean\_ratings\_df\_mf <- data.frame(ID = rownames(transposed\_masculinityfemininity\_data), facial\_masculinityfemininity = row\_meansmf)  
mean\_ratings\_df\_a <- data.frame(ID = rownames(transposed\_attractiveness\_data), facial\_attractiveness = row\_meansa)  
  
# Merges the mean dominance ratings data frame with the individualsdata data frame to create the new data frame individualsdata, then merges the mean masculinity/femininity and attractiveness ratings with individualsdata as well  
individualsdata <- merge(individualsdata, mean\_ratings\_df\_d, by.x = "ID", by.y = "ID", all.x = TRUE)  
individualsdata <- merge(individualsdata, mean\_ratings\_df\_mf, by.x = "ID", by.y = "ID", all.x = TRUE)  
individualsdata <- merge(individualsdata, mean\_ratings\_df\_a, by.x = "ID", by.y = "ID", all.x = TRUE)

I have verified that the first couple ratings are correctly calculated by checking out the individualsdata data frame directly. The average facial ratings are now the last three columns and are called facial\_dominance, facial\_masculinityfemininity, and facial\_attractiveness.

# Writing the current individualsdata data frame to the data folder  
write.csv(individualsdata, file = "./data/individuals\_data\_with\_ratings.csv")

## Reliability of Facial Width-to-Height Ratio

To allow for reliability analyses of facial width-to-height ratio (fWHR) measures, two independent raters measured fWHR using the aspect ratio produced by Fiji (Schindelin et al., 2012). Reliability will be assessed using ICCs before averaging them together for a final fWHR estimate. Because we are not interested in generalizing to the rater population and raters were not selected, this will also be a two-way mixed effects model where we are interested in consistency of ratings and computing average scores. Therefore, the ICC which will be used to assess inter-rater agreement will be the ICC(3,k) (Shrout & Fleiss, 1979).

# Calculating ICCs from fWHR\_ben and fWHR\_mad  
icc\_result\_fWHR <- ICC(individualsdata[, c("fWHR\_ben", "fWHR\_mad")])  
  
# Viewing the results  
icc\_result\_fWHR

Call: ICC(x = individualsdata[, c("fWHR\_ben", "fWHR\_mad")])  
  
Intraclass correlation coefficients   
 type ICC F df1 df2 p lower bound upper bound  
Single\_raters\_absolute ICC1 0.97 63 264 265 1.5e-162 0.96 0.98  
Single\_random\_raters ICC2 0.97 64 264 264 4.7e-163 0.96 0.98  
Single\_fixed\_raters ICC3 0.97 64 264 264 4.7e-163 0.96 0.98  
Average\_raters\_absolute ICC1k 0.98 63 264 265 1.5e-162 0.98 0.99  
Average\_random\_raters ICC2k 0.98 64 264 264 4.7e-163 0.98 0.99  
Average\_fixed\_raters ICC3k 0.98 64 264 264 4.7e-163 0.98 0.99  
  
 Number of subjects = 265 Number of Judges = 2  
See the help file for a discussion of the other 4 McGraw and Wong estimates,

The ICC(3,k) shows excellent reliability of fWHR measurements between raters (.984), so we will average them together to get a final estimate.

### Calculating Average fWHR Estimates

To calculate average fWHR estimates we will create a new data frame first—to retain the data at each stage of transformation—and then add the new fWHR variable to the copied data frame, individualsdata.

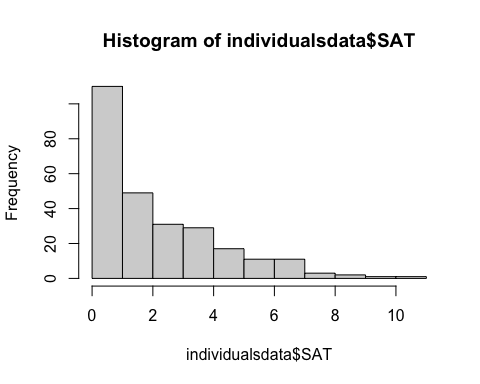
# Calculate the average fWHR and add it as a new variable "fWHR" in the data frame  
individualsdata$fWHR <- with(individualsdata, (fWHR\_ben + fWHR\_mad) / 2)

The last column now contains the fWHR variable, and rows where there were no values for fWHR\_ben or fWHR\_mad have been marked as NA.

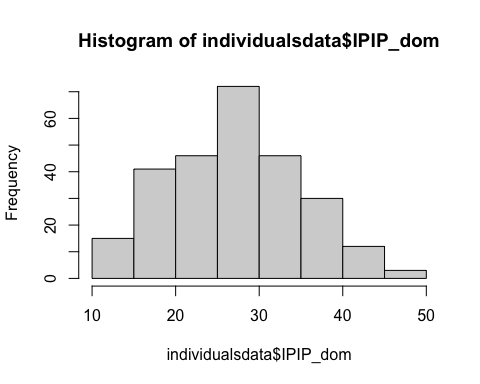
## Analysis to Determine whether to Create Composite Behavioral Dominance Measure:

In order to assess whether to create a composite behavioral dominance measure for use in later analyses, we will assess whether the SAT and the IPIP\_dom are correlated. The following code checks the assumptions of a Pearson correlation and analyzes the bivariate correlation between these variables. The SAT is actually a count variable, but I will see if I can treat it as a continuous variable by assessing these assumptions.

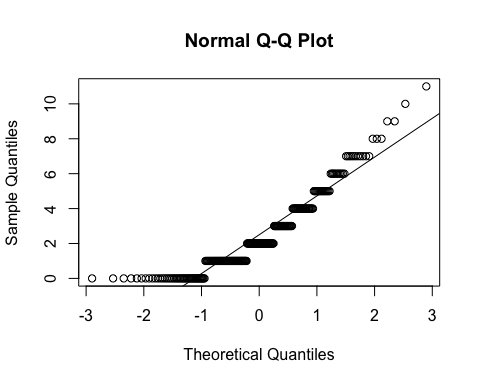
# Histograms and QQ-plots to assess normality  
hist(individualsdata$SAT)



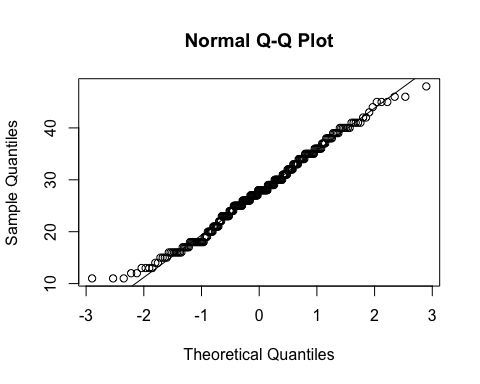
hist(individualsdata$IPIP\_dom)



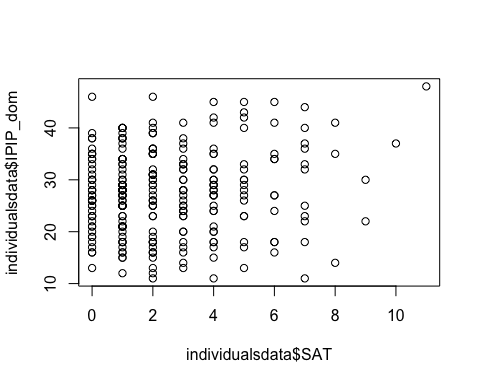
qqnorm(individualsdata$SAT); qqline(individualsdata$SAT)



qqnorm(individualsdata$IPIP\_dom); qqline(individualsdata$IPIP\_dom)



# Scatter plot to assess linearity and homoscedasticity  
plot(individualsdata$SAT, individualsdata$IPIP\_dom)

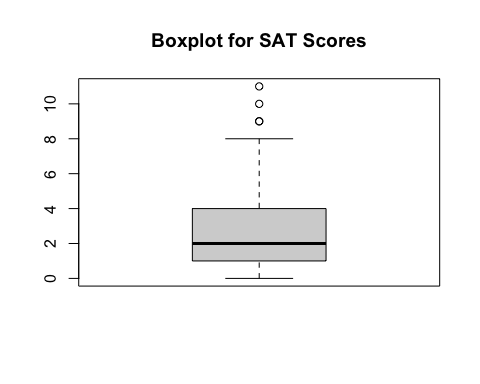


# Pearson Correlation with the variables  
corr1 <- cor.test(individualsdata$SAT, individualsdata$IPIP\_dom, method="pearson")  
  
# Display the result  
print(corr1, short = TRUE)

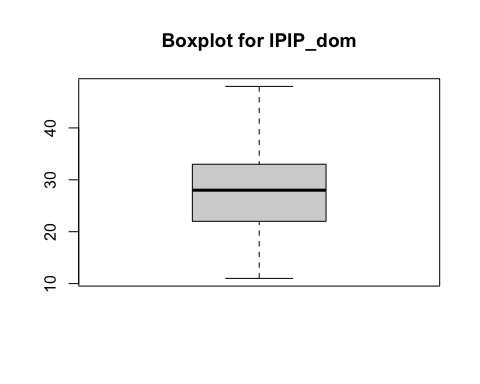
Pearson's product-moment correlation  
  
data: individualsdata$SAT and individualsdata$IPIP\_dom  
t = 1.8667, df = 263, p-value = 0.06306  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 -0.00623599 0.23165526  
sample estimates:  
 cor   
0.1143486

The histogram and QQ-plot of SAT indicates that it is far form normally distributed, and, if it is, it is highly right skewed. Although the histogram and QQ-plot indicate normality for IPIP\_dom, the assumptions have already been violated. This makes the scatterplot and correlation uninterpretable, but either way neither of them indicate a relationship between the variables. Just to be sure, I would like to use a correlation coefficient that does not assume normality. Perhaps with a robust test there will be a correlation. In the following chunk, I will test the assumptions and run the analysis for a Spearman rank correlation.

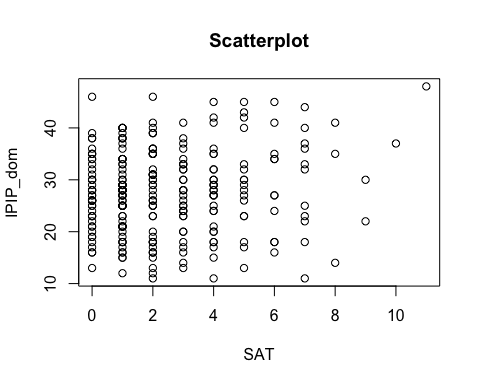
# Checking for outliers with boxplots  
boxplot(individualsdata$SAT, main="Boxplot for SAT Scores")



boxplot(individualsdata$IPIP\_dom, main="Boxplot for IPIP\_dom")



# Scatterplot to assess monotonic relationship  
plot(individualsdata$SAT, individualsdata$IPIP\_dom, main="Scatterplot", xlab="SAT", ylab="IPIP\_dom")



# Spearman's rank correlation test  
corr1.5 <- corr.test(individualsdata$SAT, individualsdata$IPIP\_dom, method = "spearman")  
  
# Displaying the Spearman correlation result  
print(corr1.5, short = FALSE)

Call:corr.test(x = individualsdata$SAT, y = individualsdata$IPIP\_dom,   
 method = "spearman")  
Correlation matrix   
[1] 0.08  
Sample Size   
[1] 265  
These are the unadjusted probability values.  
 The probability values adjusted for multiple tests are in the p.adj object.   
[1] 0.22  
  
 Confidence intervals based upon normal theory. To get bootstrapped values, try cor.ci  
 raw.lower raw.r raw.upper raw.p lower.adj upper.adj  
NA-NA -0.05 0.08 0.19 0.22 -0.05 0.19

There seems to be a few outliers in the SAT scores, but I took a look at them, and they are simply four responses that reach the upper end of the scale. Given that it is completely plausible that a person considers themselves to “often feel” 9, 10, or 11 of the adjectives included when calculating the scores for this variable, I do not think it is helpful to remove these values. The scatterplot seems to indicate monotonicity, with no strong indications of a changing relationship at high or low levels of either variable. Ultimately, the correlation is not significant, and is in any case very low, so it does not make sense to compile these variables into a composite behavioral dominance measure. This result was surprising to me, because I looked back at Palmer-Hague & Watson (2016) and found that they were moderately correlated in that sample (*r* = .28). So I looked at the descriptive statistics for the SAT there, and they were M = 3.4, SD = 2.45. The mean and standard deviation of the SAT in this sample are as follows.

mean(individualsdata$SAT, na.rm = TRUE)

[1] 2.498113

sd(individualsdata$SAT, na.rm = TRUE)

[1] 2.205579

Clearly we have a lower mean and SD for the variable here, which I am guessing is due to the preponderance of 0-2 responses in our sample (based on the histogram). Perhaps this is due to simply lower quality responses due to online data collection, rather than having parents complete packets and have their child bring it back to the lab.

## Data Cleaning for Testing Hypotheses:

### Creating Dyadic Data Set

Now I am going to prepare the dyadic data set so that we can use it to test our hypotheses. First, I need to export the current data frame as a csv file.

# Create the csv file in the "data" folder  
write.csv(individualsdata, file = "./data/individuals\_data\_with\_fWHR.csv")

In order to create the dyadic data set, I first filter out all of the individuals for which their partner did not complete the study. Then I separate the individual data set into males and females and rename the variables with m\_ for the mother variables and f\_ for the father variables, except I make the ID columns have suffixes of \_mother and \_father. I then merge them using the mother ID\_mother and father partner\_ID\_father columns.

# Filter out participants without a partner  
individualsdata\_with\_partners <- individualsdata[individualsdata$partner\_completed == 1,]  
  
# Separate the dataset into two data frames, one for mothers and one for fathers  
mothers <- individualsdata\_with\_partners[individualsdata\_with\_partners$sex == "female", ]  
fathers <- individualsdata\_with\_partners[individualsdata\_with\_partners$sex == "male", ]  
  
# Rename the columns for mothers (to include prefix m\_ to indicate mothers) and fathers (to include prefix f\_ to indicate fathers) except for the columns with IDs  
names(mothers) <- ifelse(names(mothers) %in% c('ID', 'partner\_ID'), paste0(names(mothers), "\_mother"), paste0("m\_", names(mothers)))  
names(fathers) <- ifelse(names(fathers) %in% c('ID', 'partner\_ID'), paste0(names(fathers), "\_father"), paste0("f\_", names(fathers)))  
  
# Merge the datasets using the mothers' (ID) and the fathers' (partner\_ID) columns  
dyadic\_data <- merge(mothers, fathers, by.x='ID\_mother', by.y='partner\_ID\_father')  
  
# Save the combined dataset to a new csv file called dyadic\_data.csv in the data directory without including the row names  
write.csv(dyadic\_data, file = "./data/dyadic\_data.csv", row.names = FALSE)

We have now created the dyadic\_data data frame and saved it as dyadic\_data.csv in the data directory. ID\_mother is the ID of the mother, and all of the maternal variables have the prefix m\_ to indicate that this is so. partner\_ID\_mother is the ID of the mother’s partner (i.e., the father). Similarly, ID\_father is the ID of the father, and all of the paternal variables have the prefix f\_ to indicate that this is so. partner\_ID\_father was used to merge the data frames but it is not retained in the dyadic data set. There are now 110 dyads in dyadic\_data, corresponding to the 220 participants in individualsdata who’s partner validly completed the survey.

### Adding Sex of First Shared Biological Child to Dyadic Dataset

Now that the dyadic dataset is created, I must add our primary outcome variable, shared\_child\_sex. This is the eldest (closes to first born) biological child shared by the couple. To add this variable, I could not simply use a function in R, because the sex\_first\_bio\_child variable of each dyad member may not be the same sex for the other member of the dyad (e.g., if the person had children with another person before beginning their romantic relationship at the time of data collection). Therefore, I did this outside of R by (1) using an Excel function to determine the shared child’s sex if the sex\_first\_bio\_child was matching for the members of the dyad and (2) for the cases where sex\_first\_bio\_child did not match (*n* = 10) checking the raw data to try to find the shared child. In *n* = 4 of these cases, I was able to determine the sex of the shared child by matching the sex and age for each member of the dyad, and for *n* = 6 of these cases, there was no matching sex and age between the members of the dyad. It is not clear why these remaining *n* = 6 cases do not have a matching sex and age, because I have checked each case and it is not plausible for any of them that the child had their birthday between when the parents’ completed the study. In any case, there is now a total of *n* = 104 dyads with data for shared\_child\_sex. The variable shared\_child\_age was also added by either using the existing ages when they, and the sex, matched, or by going to the raw data to fish out the correct age. In *n* = 10 cases, the partner who completed the study first reported the shared child as one year younger than the partner that completed the study second, due to the time lapse between partners completing the study. In these cases, the first response was recorded for the shared\_child\_age.

The following code reads in the new dataset dyadic\_data\_with\_shared\_child\_sex.csv to overwrite the data frame dyadic\_data.

# Read in dyadic dataset with shared child sex included to overwrite current data frame  
dyadic\_data <- read.csv("./data/dyadic\_data\_with\_shared\_child\_data.csv")  
  
# Remove participants that don't have shared\_child\_sex data (n = 6 from above)  
dyadic\_data <- dyadic\_data[!is.na(dyadic\_data$shared\_child\_sex), ]

The new variable shared\_child\_sex is now in the dyadic\_data data frame (with 0 as female and 1 as male). We will do some quick descriptive statistics for these children.

# The number of male and female shared first-borns  
table(dyadic\_data$shared\_child\_sex)

0 1   
41 63

# Descriptives for the shared\_child\_age  
describe(dyadic\_data$shared\_child\_age)

vars n mean sd median trimmed mad min max range skew kurtosis se  
X1 1 104 4 2.15 4 3.94 2.97 1 9 8 0.22 -1.06 0.21

## Seperating the Effect of Other Facial Characteristics and Age from Facial Ratings of Dominance Within Sexes

Although we could use the data frame organized by individuals (as we did when we calculated standardized residuals before in the whole sample) by creating new data frames from it that filter by sex, I am actually going to just use the dyadic dataset that we have already created for two reasons: (1) it will then not require me to recreate the dyadic dataset after the current procedure and (2) it will only include the individuals that are a part of a dyad, thereby removing the potential for individuals who’s romantic partner did not also complete the study to bias the parameters of the models. Each model will include facial\_dominance regressed on facial\_masculinityfemininity, facial\_attractiveness, and age, but first we will look at the bivariate relationships between these variables within each sex.

### Bivariate Correlations Between Facial Characteristics and Age within Sexes

# Calculate pairwise correlations with p-values and confidence intervals for mothers and store it in corr2  
corr2 <- corr.test(dyadic\_data[, c("m\_facial\_dominance", "m\_facial\_attractiveness", "m\_facial\_masculinityfemininity", "m\_age")], use="pairwise.complete.obs")  
  
# Calculate pairwise correlations with p-values and confidence intervals for fathers and store it in corr3  
corr3 <- corr.test(dyadic\_data[, c("f\_facial\_dominance", "f\_facial\_attractiveness", "f\_facial\_masculinityfemininity", "f\_age")], use="pairwise.complete.obs")  
  
print(corr2, short = FALSE)

Call:corr.test(x = dyadic\_data[, c("m\_facial\_dominance", "m\_facial\_attractiveness",   
 "m\_facial\_masculinityfemininity", "m\_age")], use = "pairwise.complete.obs")  
Correlation matrix   
 m\_facial\_dominance m\_facial\_attractiveness  
m\_facial\_dominance 1.0 0.30  
m\_facial\_attractiveness 0.3 1.00  
m\_facial\_masculinityfemininity 0.0 -0.75  
m\_age 0.0 -0.31  
 m\_facial\_masculinityfemininity m\_age  
m\_facial\_dominance 0.00 0.00  
m\_facial\_attractiveness -0.75 -0.31  
m\_facial\_masculinityfemininity 1.00 0.16  
m\_age 0.16 1.00  
Sample Size   
[1] 104  
Probability values (Entries above the diagonal are adjusted for multiple tests.)   
 m\_facial\_dominance m\_facial\_attractiveness  
m\_facial\_dominance 0.00 0.01  
m\_facial\_attractiveness 0.00 0.00  
m\_facial\_masculinityfemininity 1.00 0.00  
m\_age 0.96 0.00  
 m\_facial\_masculinityfemininity m\_age  
m\_facial\_dominance 1.00 1.00  
m\_facial\_attractiveness 0.00 0.01  
m\_facial\_masculinityfemininity 0.00 0.35  
m\_age 0.12 0.00  
  
 Confidence intervals based upon normal theory. To get bootstrapped values, try cor.ci  
 raw.lower raw.r raw.upper raw.p lower.adj upper.adj  
m\_fcl\_d-m\_fcl\_t 0.11 0.30 0.46 0.00 0.06 0.50  
m\_fcl\_d-m\_fcl\_m -0.19 0.00 0.19 1.00 -0.19 0.19  
m\_fcl\_d-m\_age -0.20 0.00 0.19 0.96 -0.22 0.21  
m\_fcl\_t-m\_fcl\_m -0.83 -0.75 -0.66 0.00 -0.85 -0.62  
m\_fcl\_t-m\_age -0.47 -0.31 -0.12 0.00 -0.52 -0.06  
m\_fcl\_m-m\_age -0.04 0.16 0.34 0.12 -0.08 0.38

print(corr3, short = FALSE)

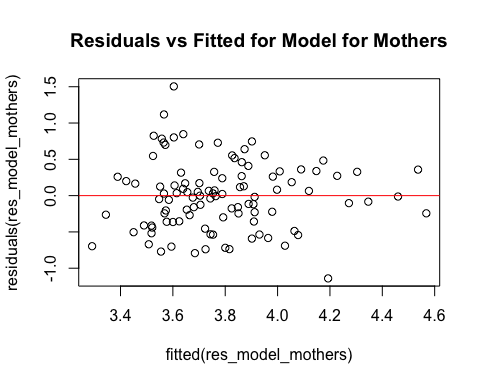
Call:corr.test(x = dyadic\_data[, c("f\_facial\_dominance", "f\_facial\_attractiveness",   
 "f\_facial\_masculinityfemininity", "f\_age")], use = "pairwise.complete.obs")  
Correlation matrix   
 f\_facial\_dominance f\_facial\_attractiveness  
f\_facial\_dominance 1.00 0.35  
f\_facial\_attractiveness 0.35 1.00  
f\_facial\_masculinityfemininity 0.77 0.25  
f\_age -0.01 -0.40  
 f\_facial\_masculinityfemininity f\_age  
f\_facial\_dominance 0.77 -0.01  
f\_facial\_attractiveness 0.25 -0.40  
f\_facial\_masculinityfemininity 1.00 0.02  
f\_age 0.02 1.00  
Sample Size   
 f\_facial\_dominance f\_facial\_attractiveness  
f\_facial\_dominance 103 103  
f\_facial\_attractiveness 103 103  
f\_facial\_masculinityfemininity 103 103  
f\_age 103 103  
 f\_facial\_masculinityfemininity f\_age  
f\_facial\_dominance 103 103  
f\_facial\_attractiveness 103 103  
f\_facial\_masculinityfemininity 103 103  
f\_age 103 104  
Probability values (Entries above the diagonal are adjusted for multiple tests.)   
 f\_facial\_dominance f\_facial\_attractiveness  
f\_facial\_dominance 0.0 0.00  
f\_facial\_attractiveness 0.0 0.00  
f\_facial\_masculinityfemininity 0.0 0.01  
f\_age 0.9 0.00  
 f\_facial\_masculinityfemininity f\_age  
f\_facial\_dominance 0.00 1  
f\_facial\_attractiveness 0.04 0  
f\_facial\_masculinityfemininity 0.00 1  
f\_age 0.86 0  
  
 Confidence intervals based upon normal theory. To get bootstrapped values, try cor.ci  
 raw.lower raw.r raw.upper raw.p lower.adj upper.adj  
f\_fcl\_d-f\_fcl\_t 0.17 0.35 0.51 0.00 0.11 0.55  
f\_fcl\_d-f\_fcl\_m 0.67 0.77 0.84 0.00 0.63 0.86  
f\_fcl\_d-f\_age -0.21 -0.01 0.18 0.90 -0.21 0.18  
f\_fcl\_t-f\_fcl\_m 0.06 0.25 0.42 0.01 0.01 0.45  
f\_fcl\_t-f\_age -0.55 -0.40 -0.22 0.00 -0.59 -0.16  
f\_fcl\_m-f\_age -0.18 0.02 0.21 0.86 -0.20 0.24

Looking at the output for mothers (corr2), we can see that facial dominance is moderately positively correlated with facial attractiveness, but it is not correlated with facial masculinity/femininity or age. For fathers, there was a similar moderate positive correlation between facial dominance and no correlation between facial dominance and facial attractiveness, but, in contrast, there was a very strong positive correlation between facial dominance and facial masculinity/femininity.

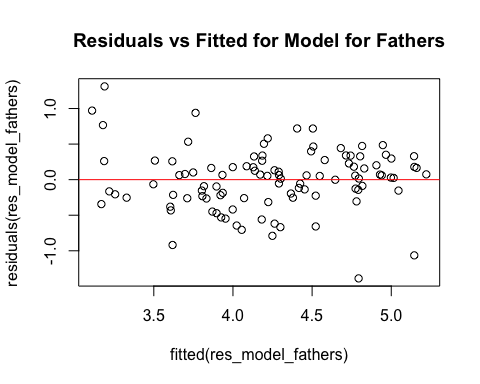
### Constructing the Models for Residual Extraction

The following will construct the models for both sexes, create scatter plots of residual vs. fitted values to visually check for homoscedasticity, create QQ-plots to visually check for normality of residuals, and display the summary statistics for the models.

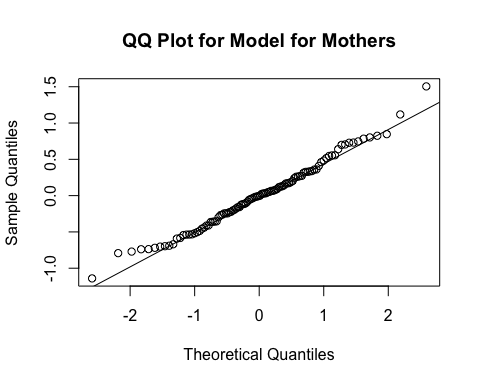
# Defining the models  
 # Model for mothers  
res\_model\_mothers <- lm(m\_facial\_dominance ~ m\_facial\_masculinityfemininity + m\_facial\_attractiveness + m\_age, data = dyadic\_data)  
 # Model for fathers  
res\_model\_fathers <- lm(f\_facial\_dominance ~ f\_facial\_masculinityfemininity + f\_facial\_attractiveness + f\_age, data = dyadic\_data)  
  
# Scatter plots to check for linearity and homoscedasticity  
 # Model for mothers  
plot(residuals(res\_model\_mothers) ~ fitted(res\_model\_mothers))  
abline(h=0, col="red")  
title("Residuals vs Fitted for Model for Mothers")



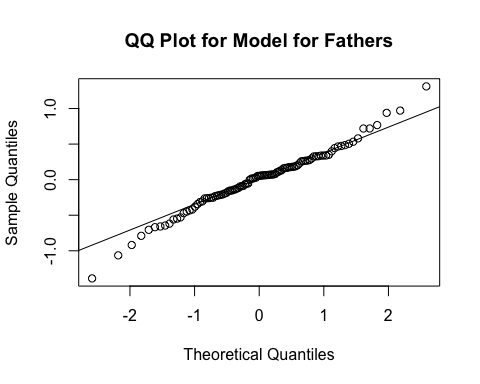
# Model for fathers  
plot(residuals(res\_model\_fathers) ~ fitted(res\_model\_fathers))  
abline(h=0, col="red")  
title("Residuals vs Fitted for Model for Fathers")



# Check for normality of residuals  
 # Model for mothers  
qqnorm(residuals(res\_model\_mothers), main = "QQ Plot for Model for Mothers")  
qqline(residuals(res\_model\_mothers))



# Model for fathers  
qqnorm(residuals(res\_model\_fathers), main = "QQ Plot for Model for Fathers")  
qqline(residuals(res\_model\_fathers))



# Summary statistics for both models  
 # Model for mothers  
summary\_res\_model\_mothers <- summary(res\_model\_mothers)  
print(summary\_res\_model\_mothers)

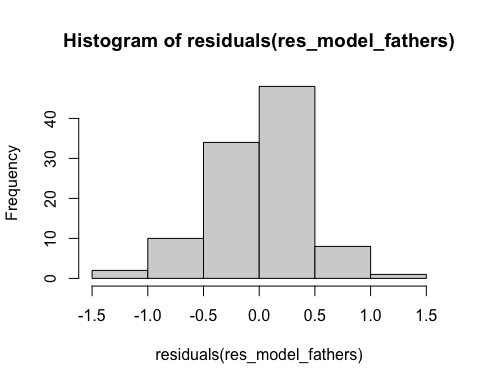
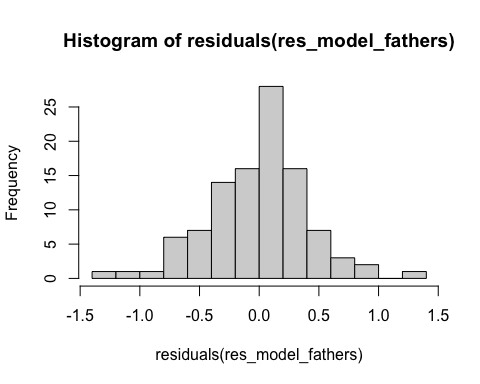
Call:  
lm(formula = m\_facial\_dominance ~ m\_facial\_masculinityfemininity +   
 m\_facial\_attractiveness + m\_age, data = dyadic\_data)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.14036 -0.35473 -0.00397 0.28335 1.50451   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 0.19873 0.77469 0.257 0.798073   
m\_facial\_masculinityfemininity 0.42356 0.10605 3.994 0.000124 \*\*\*  
m\_facial\_attractiveness 0.52803 0.09940 5.312 6.58e-07 \*\*\*  
m\_age 0.01610 0.01081 1.490 0.139343   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.4798 on 100 degrees of freedom  
Multiple R-squared: 0.2201, Adjusted R-squared: 0.1967   
F-statistic: 9.407 on 3 and 100 DF, p-value: 1.56e-05

# Model for fathers  
summary\_res\_model\_fathers <- summary(res\_model\_fathers)  
print(summary\_res\_model\_fathers)

Call:  
lm(formula = f\_facial\_dominance ~ f\_facial\_masculinityfemininity +   
 f\_facial\_attractiveness + f\_age, data = dyadic\_data)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.38825 -0.22864 0.05321 0.25948 1.31102   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) -1.673267 0.577199 -2.899 0.00461 \*\*   
f\_facial\_masculinityfemininity 0.927528 0.083566 11.099 < 2e-16 \*\*\*  
f\_facial\_attractiveness 0.224574 0.081895 2.742 0.00724 \*\*   
f\_age 0.006773 0.008948 0.757 0.45084   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.4313 on 99 degrees of freedom  
 (1 observation deleted due to missingness)  
Multiple R-squared: 0.6173, Adjusted R-squared: 0.6057   
F-statistic: 53.23 on 3 and 99 DF, p-value: < 2.2e-16

Taking a look at the residual vs. fitted values, there is not a perfectly symmetrical range of residual values for mothers across the distribution of fitted values of facial dominance. This may also be the case for fathers to a lesser degree, although it is harder to tell due to fewer values at the lower end of the distribution of fitted values. For both plots, it is clear that the assumption of linearity is reasonable, but the assumption of homogeneity of variance is not so obviously supported. Still, I think that it would be very stringent—particularly considering these models will not be used to test our hypotheses—to reject the assumption of homogeneity of variance based on these plots.

Next, for the QQ plots, there is some indication that the residuals are not normally distributed at the tails of the distribution for fathers, but the plot for mothers looks pretty good. Before moving on to the summary statistics for the models, I will construct a histogram of the residuals for fathers to take a look at why the QQ plot might deviate at the ends.

* Histogram of residual values for fathers:
* # Histogram for the residuals for the model with fathers data  
  hist(residuals(res\_model\_fathers))
* 
  + The histogram for residuals of the model for fathers seems to have some deviations from normality. Given that the mean for the residual values must be zero, we can see that the most frequent bin of values is just higher than the mean. That is, using these bins, the mean seems to be lower than the mode, indicating that there is some negative skew influencing the mean, which can also simply be seen visually. However, I would like to see this histogram with smaller bins of values to get a better sense of what is going on.
* hist(residuals(res\_model\_fathers), breaks = 10)
* 
  + This more detailed histogram shows the same pattern, where the most frequent bin is just above zero with more values at the lower end of the distribution than the higher end of the distribution, indicating a slight degree of negative skew.
  + I am getting into the weeds a bit here, because we are not trying to make inferences about the significance of the model or it’s parameters here, but I wanted to look at the normality of these residuals because they will be important for later analyses.

Finally, looking at the summaries of the models, it seems that the model for fathers explains far more variance in facial dominance than the model for mothers does. This is likely drive by the massive beta value for facial masculinity and femininity, especially considering their zero-order relationship we found doing bivariate correlations.

### Calculating Standardized Residuals for Facial Dominance within Sex

Now we will calculate the standardized residuals for facial dominance using the regression models within each sex. To do this, we will create the variables m\_res\_facial\_dominance and f\_res\_facial\_dominance by initializing them with NAs for all cases, and we will fill these variables with the standardized residuals for their respective models.

# Identify rows without missing values (because cases with missing values were excluded from both models during their construction)  
full\_rows\_res\_model\_mothers <- with(dyadic\_data, complete.cases(m\_facial\_dominance, m\_facial\_masculinityfemininity, m\_facial\_attractiveness, m\_age))  
full\_rows\_res\_model\_fathers <- with(dyadic\_data, complete.cases(f\_facial\_dominance, f\_facial\_masculinityfemininity, f\_facial\_attractiveness, f\_age))  
  
# Add the standardized residuals to the new data frame, aligning them with the full rows  
 # Initializing new variables with NA values first  
dyadic\_data$m\_res\_facial\_dominance <- NA # Initialize new variable with NAs for mothers  
dyadic\_data$f\_res\_facial\_dominance <- NA # Initialize new variable with NAs for fathers  
 # Add the standardized residuals to the mothers variable  
dyadic\_data$m\_res\_facial\_dominance[full\_rows\_res\_model\_mothers] <- rstandard(res\_model\_mothers)  
 # Add the standardized residuals to the fathers variable  
dyadic\_data$f\_res\_facial\_dominance[full\_rows\_res\_model\_fathers] <- rstandard(res\_model\_fathers)

The dyadic\_data data frame now contains the m\_res\_facial\_dominance and f\_res\_facial\_dominance variables in the last columns.

Before moving on, I am going to write this new dataframe as a csv file to ensure it is saved.

write.csv(dyadic\_data, file = "./data/dyadic\_data\_with\_res\_facial\_dominance.csv")

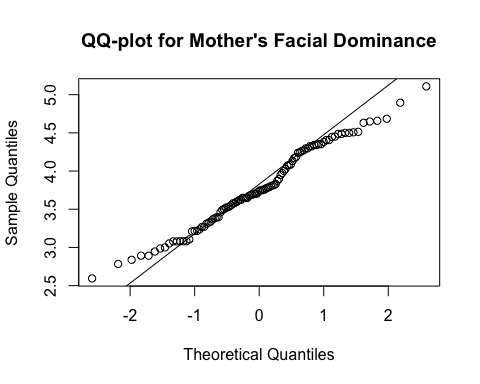
## Assessing Whether Facial Ratings and fWHR Vary by Facial Expression Within Sexes

Now we will take a look within the sexes at whether facial ratings differ between those that had a neutral facial expression—as we asked for—or had a non-neutral facial expression. To do this, we will test for group differences using an independent samples t-test. We will do this for mothers first, and we will test assumptions first.

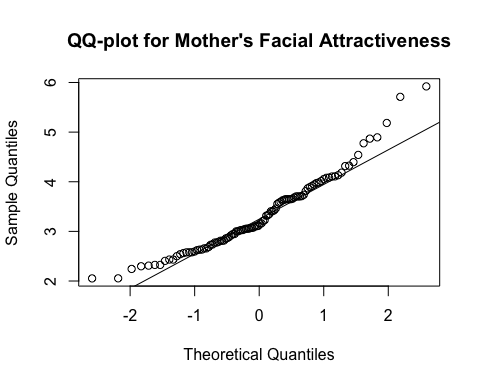
### For Mothers

The following code tests the assumptions of the independent-samples t-test to determine whether a regular or Welch’s t-test is most appropriate. QQ-plots are produced to assess normality, and Levene’s test is employed to assess whether homogeneity of variance can be assumed. Finally, we count the number of cases that have neutral and non-neutral facial expressions to see whether there are enough cases in both groups to make valid inferences.

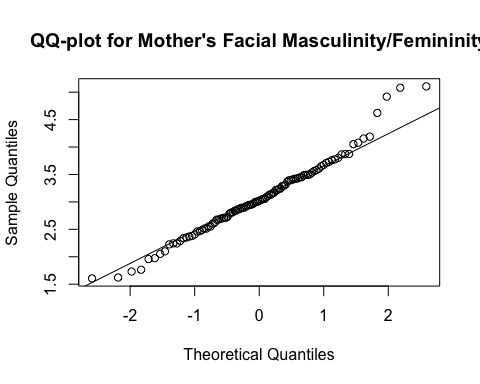
# Q-Q Plots to assess normality  
qqnorm(dyadic\_data$m\_facial\_dominance, main = "QQ-plot for Mother's Facial Dominance")  
qqline(dyadic\_data$m\_facial\_dominance)



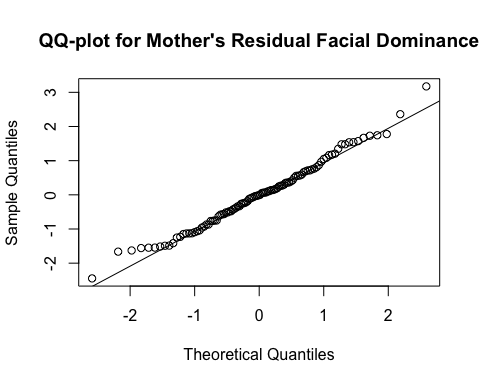
qqnorm(dyadic\_data$m\_facial\_attractiveness, main = "QQ-plot for Mother's Facial Attractiveness")  
qqline(dyadic\_data$m\_facial\_attractiveness)



qqnorm(dyadic\_data$m\_facial\_masculinityfemininity, main = "QQ-plot for Mother's Facial Masculinity/Femininity")  
qqline(dyadic\_data$m\_facial\_masculinityfemininity)



qqnorm(dyadic\_data$m\_res\_facial\_dominance, main = "QQ-plot for Mother's Residual Facial Dominance")  
qqline(dyadic\_data$m\_res\_facial\_dominance)



# Convert 'm\_expression\_not\_neutral' to a factor  
dyadic\_data$m\_expression\_not\_neutral <- as.factor(dyadic\_data$m\_expression\_not\_neutral)  
  
# Levene's Test to assess homogeneity of variances  
leveneTest(m\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.4312 0.5129  
 102

leveneTest(m\_facial\_attractiveness ~ m\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 1.0618 0.3052  
 102

leveneTest(m\_facial\_masculinityfemininity ~ m\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.1374 0.7116  
 102

leveneTest(m\_res\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.0075 0.931  
 102

# Check the frequency of the facial expressions that are not neutral  
summary(dyadic\_data$m\_expression\_not\_neutral)

neutral not\_neutral   
 92 12

The QQ-plots for the mothers facial dominance, facial attractiveness, and facial masculinity/femininity do not look great—although residual facial dominance, which we are principally interested in here, looks pretty good—but with *n* = 104 cases, I think we can rely on the central limit theorum. For each of the Levene’s test, the p-value does not fall below *p* = .305, so we can assume homogeneity of variance for between the groups. However, there are *n* = 92 cases in the neutral facial expression group and only *n* = 12 cases in the non-neutral facial expression group. Because of this, I will do both a normal independent samples t-test and a Welch’s t-test for each comparison, just to make sure inferences are appropriate.

# Performing the regular t-tests (two-way)  
t\_test\_dom\_expression <- t.test(m\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
t\_test\_att\_expression <- t.test(m\_facial\_attractiveness ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
t\_test\_masfem\_expression <- t.test(m\_facial\_masculinityfemininity ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
t\_test\_res\_dom\_expression <- t.test(m\_res\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
  
# Performing the Welch's t-tests (two-way)  
w\_t\_test\_dom\_expression <- t.test(m\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
w\_t\_test\_att\_expression <- t.test(m\_facial\_attractiveness ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
w\_t\_test\_masfem\_expression <- t.test(m\_facial\_masculinityfemininity ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
w\_t\_test\_res\_dom\_expression <- t.test(m\_res\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
  
# Printing the results  
t\_test\_dom\_expression

Two Sample t-test  
  
data: m\_facial\_dominance by m\_expression\_not\_neutral  
t = 2.1763, df = 102, p-value = 0.03184  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.0311239 0.6713665  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 3.824159 3.472914

w\_t\_test\_dom\_expression

Welch Two Sample t-test  
  
data: m\_facial\_dominance by m\_expression\_not\_neutral  
t = 2.2758, df = 14.418, p-value = 0.0386  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.02111682 0.68137363  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 3.824159 3.472914

t\_test\_att\_expression

Two Sample t-test  
  
data: m\_facial\_attractiveness by m\_expression\_not\_neutral  
t = -1.7386, df = 102, p-value = 0.08512  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 -0.85526641 0.05627268  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 3.282658 3.682154

w\_t\_test\_att\_expression

Welch Two Sample t-test  
  
data: m\_facial\_attractiveness by m\_expression\_not\_neutral  
t = -1.5254, df = 13.129, p-value = 0.1509  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 -0.9647153 0.1657216  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 3.282658 3.682154

t\_test\_masfem\_expression

Two Sample t-test  
  
data: m\_facial\_masculinityfemininity by m\_expression\_not\_neutral  
t = 2.2341, df = 102, p-value = 0.02766  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.05155174 0.86744362  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 3.121890 2.662393

w\_t\_test\_masfem\_expression

Welch Two Sample t-test  
  
data: m\_facial\_masculinityfemininity by m\_expression\_not\_neutral  
t = 2.5291, df = 15.251, p-value = 0.02292  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.07280782 0.84618754  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 3.121890 2.662393

t\_test\_res\_dom\_expression

Two Sample t-test  
  
data: m\_res\_facial\_dominance by m\_expression\_not\_neutral  
t = 2.8848, df = 102, p-value = 0.004779  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.2680257 1.4476950  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 0.09918414 -0.75867621

w\_t\_test\_res\_dom\_expression

Welch Two Sample t-test  
  
data: m\_res\_facial\_dominance by m\_expression\_not\_neutral  
t = 2.6666, df = 13.45, p-value = 0.01893  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.165223 1.550498  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 0.09918414 -0.75867621

* Mother’s Facial Dominance:
  + Both the regular and Welch’s t-test indicate that there is a group difference in dominance ratings between the neutral and non-neutral facial expression groups, with the neutral group (M = 3.824) rated as more dominant than the non-neutral group (M = 3.473). This is presumably because most of the non-neutral faces were smiling.
* Mother’s Facial Attractiveness:
  + The both versions of the t-test indicate that there is a marginal group difference in attractiveness ratings between the neutral and non-neutral facial expression groups, with the neutral group (M = 3.283) rated as less attractive than the non-neutral group (M = 3.682) (again, I presume due to smiling).
* Mother’s Facial Masculinity/Femininity:
  + Both the regular and Welch’s t-test indicate that there is a group difference in masculinity/femininity ratings between the neutral and non-neutral facial expression groups, with the neutral group (M = 3.122) rated as more masculine (less feminine) than the non-neutral group (M = 2.662). If I am not mistaken, there are sex differences in the frequency of smiling such that women tend to smile more, which may be why neutral facial expressions were rated as less feminine (more masculine) for these women.
* Mother’s Residual Facial Dominance:
  + Both the regular and Welch’s t-test indicate that there is a group difference in residual dominance scores between the neutral and non-neutral facial expression groups, with the neutral group (M = 0.099) higher on residual dominance than the non-neutral group (M = -0.759). This is presumably because most of the non-neutral faces were smiling.

This last result indicates that when testing hypotheses later in the analysis, we would construct the model both with and without the non-neutral facial expressions to ensure that this is not affecting the results. I am curious how large these group differences are, so the following code will calculate the Cohen’s d for each of these group differences.

cohen.d(m\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data)

Call: cohen.d(x = m\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
m\_facial\_dominance -1.28 -0.67 -0.06  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.67  
r equivalent of difference between two means  
m\_facial\_dominance   
 -0.21

cohen.d(m\_facial\_attractiveness ~ m\_expression\_not\_neutral, data = dyadic\_data)

Call: cohen.d(x = m\_facial\_attractiveness ~ m\_expression\_not\_neutral,   
 data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
m\_facial\_attractiveness -0.07 0.54 1.14  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.54  
r equivalent of difference between two means  
m\_facial\_attractiveness   
 0.17

cohen.d(m\_facial\_masculinityfemininity ~ m\_expression\_not\_neutral, data = dyadic\_data)

Call: cohen.d(x = m\_facial\_masculinityfemininity ~ m\_expression\_not\_neutral,   
 data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
m\_facial\_masculinityfemininity -1.3 -0.69 -0.08  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.69  
r equivalent of difference between two means  
m\_facial\_masculinityfemininity   
 -0.22

cohen.d(m\_res\_facial\_dominance ~ m\_expression\_not\_neutral, data = dyadic\_data)

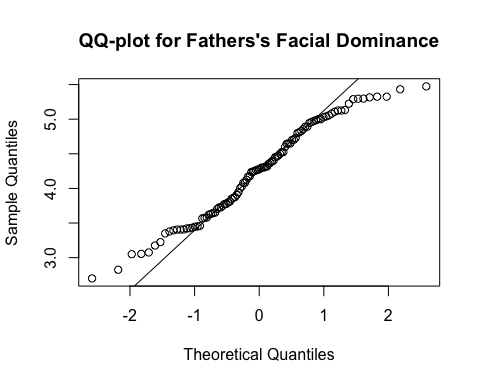
Call: cohen.d(x = m\_res\_facial\_dominance ~ m\_expression\_not\_neutral,   
 data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
m\_res\_facial\_dominance -1.51 -0.89 -0.28  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.89  
r equivalent of difference between two means  
m\_res\_facial\_dominance   
 -0.27

Given the conventional standards for Cohen’s d, facial dominance, facial attractiveness, and facial masculinity/femininity each show medium group differences. In addition, the estimate of Cohen’s d for residual facial dominance scores is fairly large (d = -.89).

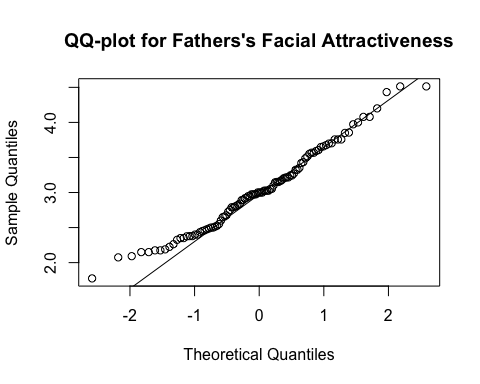
### For Fathers

The following code tests the assumptions of the independent-samples t-test to determine whether a regular or Welch’s t-test is most appropriate. QQ-plots are produced to assess normality, and Levene’s test is employed to assess whether homogeneity of variance can be assumed. Finally, we count the number of cases that have neutral and non-neutral facial expressions to see whether there are enough cases in both groups to make valid inferences.

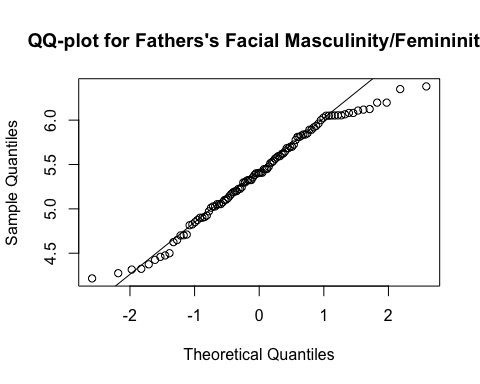
# Q-Q Plots to assess normality  
qqnorm(dyadic\_data$f\_facial\_dominance, main = "QQ-plot for Fathers's Facial Dominance")  
qqline(dyadic\_data$f\_facial\_dominance)



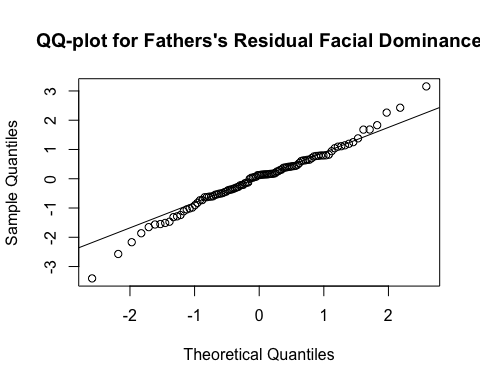
qqnorm(dyadic\_data$f\_facial\_attractiveness, main = "QQ-plot for Fathers's Facial Attractiveness")  
qqline(dyadic\_data$f\_facial\_attractiveness)



qqnorm(dyadic\_data$f\_facial\_masculinityfemininity, main = "QQ-plot for Fathers's Facial Masculinity/Femininity")  
qqline(dyadic\_data$f\_facial\_masculinityfemininity)



qqnorm(dyadic\_data$f\_res\_facial\_dominance, main = "QQ-plot for Fathers's Residual Facial Dominance")  
qqline(dyadic\_data$f\_res\_facial\_dominance)



# Convert 'f\_expression\_not\_neutral' to a factor variable  
dyadic\_data$f\_expression\_not\_neutral <- as.factor(dyadic\_data$f\_expression\_not\_neutral)  
  
# Levene's Test to assess homogeneity of variances  
leveneTest(f\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 1 6.1346 0.01492 \*  
 101   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

leveneTest(f\_facial\_attractiveness ~ f\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.0193 0.8898  
 101

leveneTest(f\_facial\_masculinityfemininity ~ f\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.0018 0.9658  
 101

leveneTest(f\_res\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 1 4.0448 0.04697 \*  
 101   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Check the frequency of the facial expressions that are not neutral  
summary(dyadic\_data$f\_expression\_not\_neutral)

neutral not\_neutral   
 91 13

The QQ-plots for the fathers facial dominance, facial attractiveness, facial masculinity/femininity, *and* residual facial dominance are not as uniform as I would like them to be, but, as with above, having *n* = 104 cases should allow us to be liberal with the assumption of normality. However, Levene’s test indicates that we should reject the null hypothesis that variances are equal for facial dominance and, almost, for residual facial dominance. Further, there is a large difference is the number of cases in each group (neutral *n* = 91 and non-neutral *n* = 13). Taken together, this indicates that we should make our inferences from the results for Welch’s t-tests below.

# Performing the regular t-tests (two-way)  
t\_test\_dom\_expression\_f <- t.test(f\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
t\_test\_att\_expression\_f <- t.test(f\_facial\_attractiveness ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
t\_test\_masfem\_expression\_f <- t.test(f\_facial\_masculinityfemininity ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
t\_test\_res\_dom\_expression\_f <- t.test(f\_res\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = TRUE)  
  
# Performing the Welch's t-tests (two-way)  
w\_t\_test\_dom\_expression\_f <- t.test(f\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
w\_t\_test\_att\_expression\_f <- t.test(f\_facial\_attractiveness ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
w\_t\_test\_masfem\_expression\_f <- t.test(f\_facial\_masculinityfemininity ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
w\_t\_test\_res\_dom\_expression\_f <- t.test(f\_res\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data, var.equal = FALSE)  
  
# Printing the results  
t\_test\_dom\_expression\_f

Two Sample t-test  
  
data: f\_facial\_dominance by f\_expression\_not\_neutral  
t = 2.9647, df = 101, p-value = 0.003781  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.199431 1.006044  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 4.316144 3.713407

w\_t\_test\_dom\_expression\_f

Welch Two Sample t-test  
  
data: f\_facial\_dominance by f\_expression\_not\_neutral  
t = 4.7533, df = 23.491, p-value = 8.184e-05  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.3407260 0.8647491  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 4.316144 3.713407

t\_test\_att\_expression\_f

Two Sample t-test  
  
data: f\_facial\_attractiveness by f\_expression\_not\_neutral  
t = -2.6664, df = 101, p-value = 0.008929  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 -0.8194121 -0.1202965  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 2.974734 3.444589

w\_t\_test\_att\_expression\_f

Welch Two Sample t-test  
  
data: f\_facial\_attractiveness by f\_expression\_not\_neutral  
t = -2.5114, df = 13.604, p-value = 0.02533  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 -0.8722178 -0.0674908  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 2.974734 3.444589

t\_test\_masfem\_expression\_f

Two Sample t-test  
  
data: f\_facial\_masculinityfemininity by f\_expression\_not\_neutral  
t = 2.0491, df = 101, p-value = 0.04304  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.01051675 0.64843906  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 5.427786 5.098308

w\_t\_test\_masfem\_expression\_f

Welch Two Sample t-test  
  
data: f\_facial\_masculinityfemininity by f\_expression\_not\_neutral  
t = 2.0051, df = 13.888, p-value = 0.06484  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 -0.02321507 0.68217089  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 5.427786 5.098308

t\_test\_res\_dom\_expression\_f

Two Sample t-test  
  
data: f\_res\_facial\_dominance by f\_expression\_not\_neutral  
t = 3.1036, df = 101, p-value = 0.00248  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.3346447 1.5202643  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 0.1086182 -0.8188363

w\_t\_test\_res\_dom\_expression\_f

Welch Two Sample t-test  
  
data: f\_res\_facial\_dominance by f\_expression\_not\_neutral  
t = 2.2173, df = 12.221, p-value = 0.04628  
alternative hypothesis: true difference in means between group neutral and group not\_neutral is not equal to 0  
95 percent confidence interval:  
 0.0179447 1.8369642  
sample estimates:  
 mean in group neutral mean in group not\_neutral   
 0.1086182 -0.8188363

* Father’s Facial Dominance:
  + The Welch’s t-test indicates that there is a difference between the neutral and non-neutral group such that the neutral group (M = 4.316) was rated significantly more dominant than the non-neutral group (M = 3.713). I suspect this is due to most non-neutral expressions smiling.
* Father’s Facial Attractiveness:
  + The Welch’s t-test indicates that there is a difference between the neutral and non-neutral group such that the neutral group (M = 2.975) was rated significantly less attractive than the non-neutral group (M = 3.445) (probably because they were mostly smiling).
* Father’s Facial Masculinity/Femininity:
  + The Welch’s t-test (and regular t-test) was close to our level for significance here (*p* = .06), suggesting there may be a difference between the neutral and non-neutral group such that the neutral group (M = 5.428) was rated significantly more masculine (less feminine) than the non-neutral group (M = 5.098). Again, I suspect that because women tend to smile more than men, the smiling can account for the difference in ratings here, but I would need to look into this.
* Father’s Residual Facial Dominance:
  + The Welch’s t-test indicates that there is a difference between the neutral and non-neutral group such that the neutral group (M = 0.109) has significantly higher residual facial dominance scores than the non-neutral group (M = -0.819). I suspect this is due to most non-neutral expressions smiling.

This last result indicates that when testing hypotheses later in the analysis, we would construct the model both with and without the non-neutral facial expressions to ensure that this is not affecting the results. And, again, the following code will calculate the Cohen’s d for each of these group differences.

cohen.d(f\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data)

Call: cohen.d(x = f\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
f\_facial\_dominance -1.53 -0.92 -0.3  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.92  
r equivalent of difference between two means  
f\_facial\_dominance   
 -0.28

cohen.d(f\_facial\_attractiveness ~ f\_expression\_not\_neutral, data = dyadic\_data)

Call: cohen.d(x = f\_facial\_attractiveness ~ f\_expression\_not\_neutral,   
 data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
f\_facial\_attractiveness 0.21 0.83 1.44  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.83  
r equivalent of difference between two means  
f\_facial\_attractiveness   
 0.26

cohen.d(f\_facial\_masculinityfemininity ~ f\_expression\_not\_neutral, data = dyadic\_data)

Call: cohen.d(x = f\_facial\_masculinityfemininity ~ f\_expression\_not\_neutral,   
 data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
f\_facial\_masculinityfemininity -1.24 -0.64 -0.03  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.64  
r equivalent of difference between two means  
f\_facial\_masculinityfemininity   
 -0.2

cohen.d(f\_res\_facial\_dominance ~ f\_expression\_not\_neutral, data = dyadic\_data)

Call: cohen.d(x = f\_res\_facial\_dominance ~ f\_expression\_not\_neutral,   
 data = dyadic\_data)  
Cohen d statistic of difference between two means  
 lower effect upper  
f\_res\_facial\_dominance -1.58 -0.96 -0.34  
  
Multivariate (Mahalanobis) distance between groups  
[1] 0.96  
r equivalent of difference between two means  
f\_res\_facial\_dominance   
 -0.3

Interestingly, it seems that having a non-neutral facial expression had a slightly larger effect on rated facial dominance and residual facial dominance in men than in the previous analyses for women (both large effects; d = -.92 and d = -.96, respectively). For facial masculinity/femininity and facial attractiveness, effect sizes were similar to those of the mothers group.

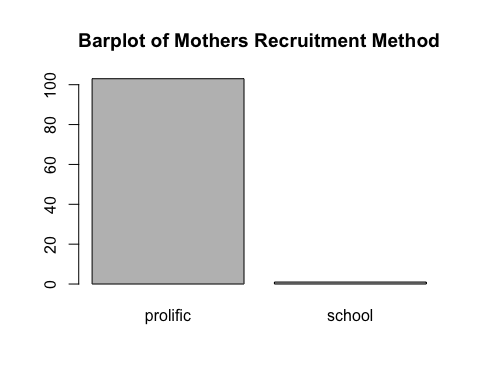
## Demographic and Recruitment Characteristics

Before moving on, I want to quickly take a look at some demographic and recruitment characteristics within the mothers and fathers in the dyadic dataset. That is, we will look at the method of recruitment (m\_rec\_method and f\_rec\_method), whether the participant resubmitted their photograph at our request (m\_photo\_resubmission and f\_photo\_resubmission), ethnicity (m\_ethnicity and f\_ethnicity), nationality (m\_nationality and f\_nationality), and age (m\_age and f\_age). Some of these variables need to be recoded as factor variables—in accordance with the labels specified in the [Introduction](#introduction) above—which will be done before producing the descriptive statistics.

# Recoding variables currently numeric as factors  
dyadic\_data$m\_rec\_meth <- as.factor(dyadic\_data$m\_rec\_meth)  
dyadic\_data$f\_rec\_meth <- as.factor(dyadic\_data$f\_rec\_meth)  
dyadic\_data$m\_photo\_resubmission <- as.factor(dyadic\_data$m\_photo\_resubmission)  
dyadic\_data$f\_photo\_resubmission <- as.factor(dyadic\_data$f\_photo\_resubmission)  
dyadic\_data$m\_ethnicity <- as.factor(dyadic\_data$m\_ethnicity)  
dyadic\_data$f\_ethnicity <- as.factor(dyadic\_data$f\_ethnicity)  
dyadic\_data$m\_nationality <- as.factor(dyadic\_data$m\_nationality)  
dyadic\_data$f\_nationality <- as.factor(dyadic\_data$f\_nationality)  
  
# Frequencies and barplot for mothers' recruitment method  
table\_m\_rec\_meth <- table(dyadic\_data$m\_rec\_meth)  
print(table\_m\_rec\_meth)

prolific school   
 103 1

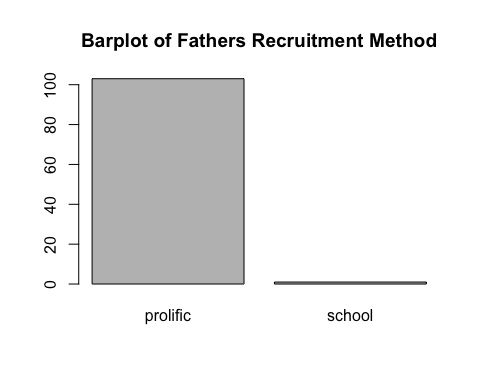
barplot(table\_m\_rec\_meth, main = "Barplot of Mothers Recruitment Method")



# Frequencies and barplot for fathers' recruitment method  
table\_f\_rec\_meth <- table(dyadic\_data$f\_rec\_meth)  
print(table\_f\_rec\_meth)

prolific school   
 103 1

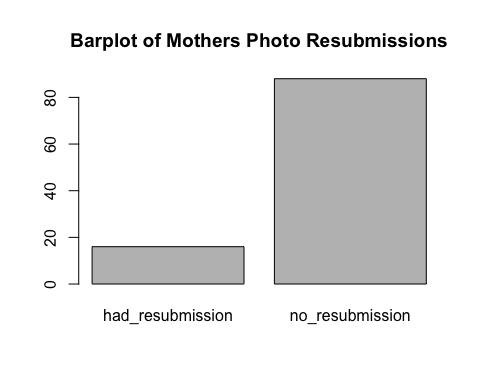
barplot(table\_f\_rec\_meth, main = "Barplot of Fathers Recruitment Method")



# Frequencies and barplot for mothers' photograph resubmissions  
table\_m\_photo\_resubmission <- table(dyadic\_data$m\_photo\_resubmission)  
print(table\_m\_photo\_resubmission)

had\_resubmission no\_resubmission   
 16 88

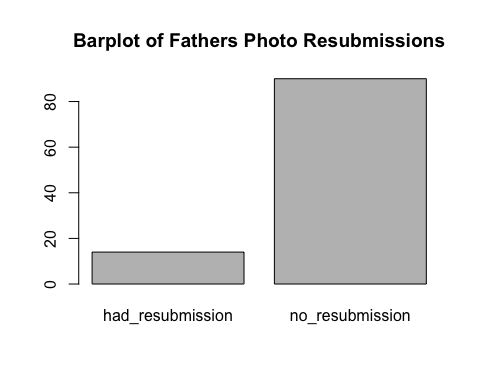
barplot(table\_m\_photo\_resubmission, main = "Barplot of Mothers Photo Resubmissions")



# Frequencies and barplot for fathers' photograph resubmissions  
table\_f\_photo\_resubmission <- table(dyadic\_data$f\_photo\_resubmission)  
print(table\_f\_photo\_resubmission)

had\_resubmission no\_resubmission   
 14 90

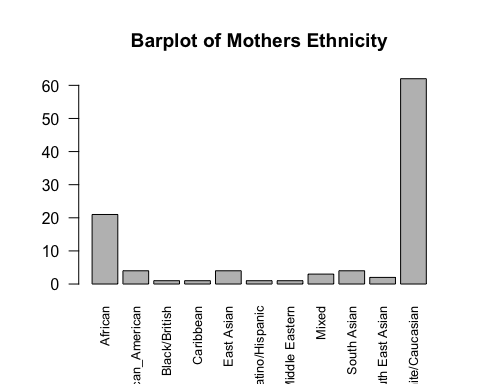
barplot(table\_f\_photo\_resubmission, main = "Barplot of Fathers Photo Resubmissions")



# Frequencies and barplot for mothers' ethnicity  
table\_m\_ethnicity <- table(dyadic\_data$m\_ethnicity)  
print(table\_m\_ethnicity)

African Black/African\_American Black/British   
 21 4 1   
 Caribbean East Asian Latino/Hispanic   
 1 4 1   
 Middle Eastern Mixed South Asian   
 1 3 4   
 South East Asian White/Caucasian   
 2 62

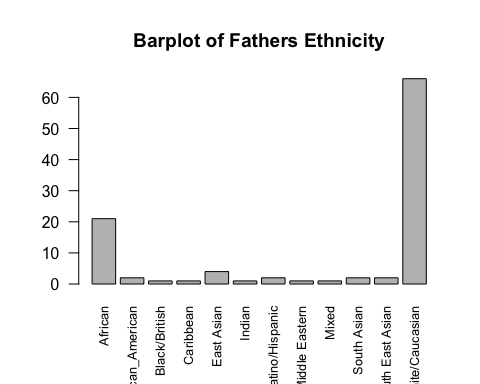
barplot(table\_m\_ethnicity, las = 2, cex.names = .8, main = "Barplot of Mothers Ethnicity")



# Frequencies and barplot for fathers' ethnicity  
table\_f\_ethnicity <- table(dyadic\_data$f\_ethnicity)  
print(table\_f\_ethnicity)

African Black/African\_American Black/British   
 21 2 1   
 Caribbean East Asian Indian   
 1 4 1   
 Latino/Hispanic Middle Eastern Mixed   
 2 1 1   
 South Asian South East Asian White/Caucasian   
 2 2 66

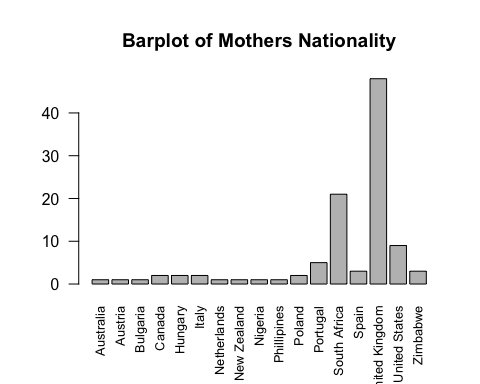
barplot(table\_f\_ethnicity, las = 2, cex.names = .8, main = "Barplot of Fathers Ethnicity")



# Frequencies and barplot for mothers' nationality  
table\_m\_nationality <- table(dyadic\_data$m\_nationality)  
print(table\_m\_nationality)

Australia Austria Bulgaria Canada Hungary   
 1 1 1 2 2   
 Italy Netherlands New Zealand Nigeria Phillipines   
 2 1 1 1 1   
 Poland Portugal South Africa Spain United Kingdom   
 2 5 21 3 48   
 United States Zimbabwe   
 9 3

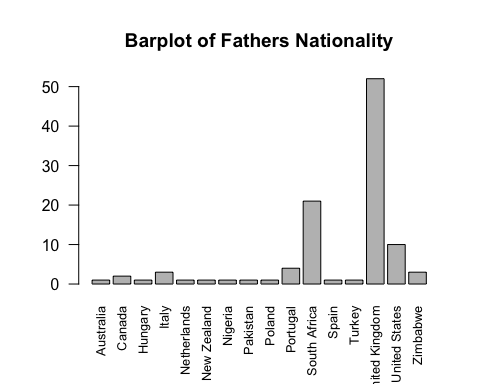
barplot(table\_m\_nationality, las = 2, cex.names = .8, main = "Barplot of Mothers Nationality")



# Frequencies and barplot for fathers' ethnicity  
table\_f\_nationality <- table(dyadic\_data$f\_nationality)  
print(table\_f\_nationality)

Australia Canada Hungary Italy Netherlands   
 1 2 1 3 1   
 New Zealand Nigeria Pakistan Poland Portugal   
 1 1 1 1 4   
 South Africa Spain Turkey United Kingdom United States   
 21 1 1 52 10   
 Zimbabwe   
 3

barplot(table\_f\_nationality, las = 2, cex.names = .8, main = "Barplot of Fathers Nationality")

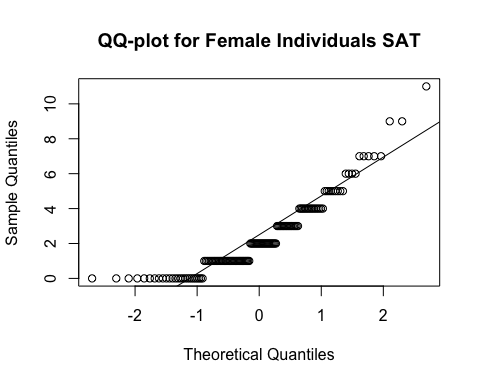


* Recruitment Method:
  + As expected, only one participant from recruitment from schools was included, because only one couple submitted both facial photographs.
* Photograph Resubmissions:
  + The vast majority of people did not need to resubmit their photograph, with *n* = 16 in the mothers’ group and *n* = 14 in the fathers’ group having resubmissions.
* Ethnicity:
  + Regarding ethnicity, the bar plots for mothers and fathers both indicate that the majority of mothers and fathers were white/caucasian (*n* = 62 and *n* = 66, respectively), and the second most common ethnic category for each group was African (*n* = 21 and *n* = 21, respectively). No other ethnic category had more than *n* = 5 cases.
* Nationality:
  + For nationality, the bar plots for mothers’ and fathers’ nationality indicate that the most common nationality was United Kingdom (*n* = 48 and *n* = 52, respectively), followed by South Africa (*n* = 21 and *n* = 21, respectively) and the United States (*n* = 9 and *n* = 10, respectively). No other nationality had more than *n* = 5 cases.

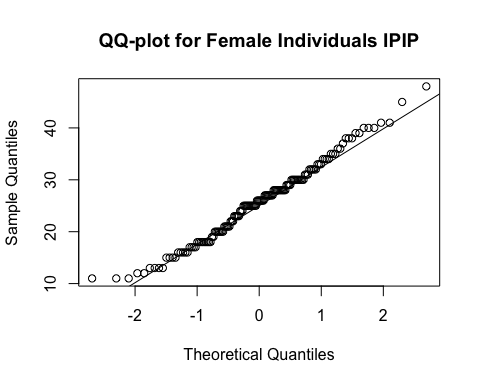
## A Quick Look at the Maternal Dominance Hypothesis for Individual Differences in Condition

Before moving on to test our Hypothesis 2 or Hypothesis 3, I am curious if we can replicate the maternal dominance hypothesis (Grant, 1990), so I will run some simple t-tests with sex of first born child as the grouping variable and the SAT dominance and IPIP dominance measures as the dependent variable. To retain all of the female participants, I will run this analysis by creating a data frame called female\_individuals with only female individuals from the individualsdata data frame. First, I will test the assumptions, then I will do the t-tests.

# Create a data frame with only female individuals from the individualsdata data frame  
female\_individuals <- individualsdata[individualsdata$sex == "female", ]  
  
# Q-Q Plots to assess normality  
qqnorm(female\_individuals$SAT, main = "QQ-plot for Female Individuals SAT")  
qqline(female\_individuals$SAT)



qqnorm(female\_individuals$IPIP\_dom, main = "QQ-plot for Female Individuals IPIP")  
qqline(female\_individuals$IPIP\_dom)



# Levene's Test to assess homogeneity of variances  
leveneTest(IPIP\_dom ~ sex\_first\_bio\_child, data = female\_individuals)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.2626 0.6092  
 139

leveneTest(SAT ~ sex\_first\_bio\_child, data = female\_individuals)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.3158 0.5751  
 139

The QQ-plot for the SAT dominance has some deviation from predicted values at the lower end of the distribution, but the IPIP dominance looks good. Also, there are *n* = 140 observations for each variable, so I am not worried about violations of the assumption of normality. Based on Levene’s test, we will accept the null hypothesis that there is no difference between the variances of the two groups.

Now for Student’s t-test.

# Saving the results of the t-tests  
mdh\_t\_test\_SAT <- t.test(SAT ~ sex\_first\_bio\_child, data = female\_individuals, var.equal = TRUE)  
mdh\_t\_test\_IPIP <- t.test(IPIP\_dom ~ sex\_first\_bio\_child, data = female\_individuals, var.equal = TRUE)  
  
# Displaying the results  
mdh\_t\_test\_SAT

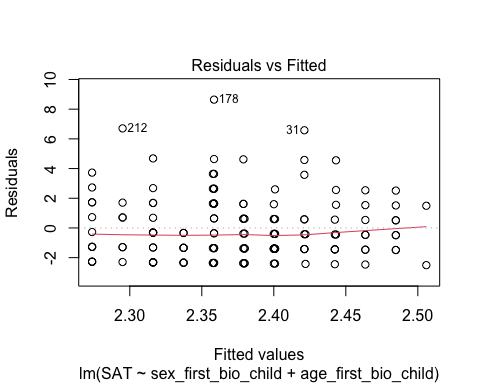
Two Sample t-test  
  
data: SAT by sex\_first\_bio\_child  
t = 0.22269, df = 139, p-value = 0.8241  
alternative hypothesis: true difference in means between group female and group male is not equal to 0  
95 percent confidence interval:  
 -0.6481232 0.8126540  
sample estimates:  
mean in group female mean in group male   
 2.423729 2.341463

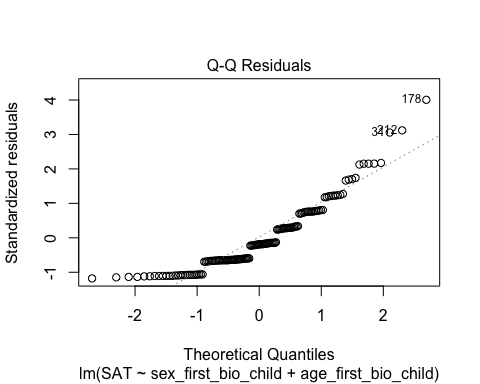
mdh\_t\_test\_IPIP

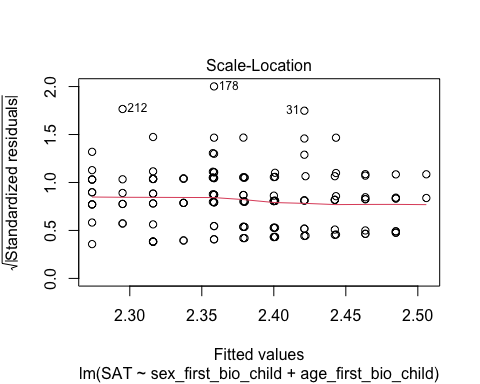
Two Sample t-test  
  
data: IPIP\_dom by sex\_first\_bio\_child  
t = -0.57983, df = 139, p-value = 0.563  
alternative hypothesis: true difference in means between group female and group male is not equal to 0  
95 percent confidence interval:  
 -3.375335 1.844537  
sample estimates:  
mean in group female mean in group male   
 25.27119 26.03659

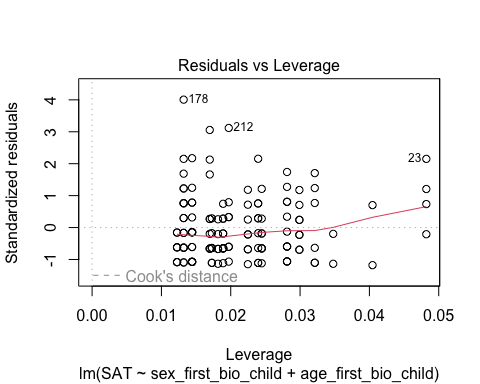
In this dataset, we are unable to conceptually replicate the individual differences version of (Grant, 1990) using a simple t-test, although it is worth noting that her behavioral dominance data came from just before the children were born. By adding in age\_first\_bio\_child as a covariate in a general linear model with sex\_first\_bio\_child as a predictor and the behavioral dominance measures as outcome variables, we could see whether the time since birth of the child is potentially confounding the results here.

# Constructing the models  
MDH\_glm\_SAT <- lm(SAT ~ sex\_first\_bio\_child + age\_first\_bio\_child, data = female\_individuals)  
MDH\_glm\_IPIP <- lm(SAT ~ sex\_first\_bio\_child + age\_first\_bio\_child, data = female\_individuals)  
  
# Plotting residuals vs. fitted values for linearity and homoscedasticity and producing QQ plots for normality of residuals  
plot(MDH\_glm\_SAT)

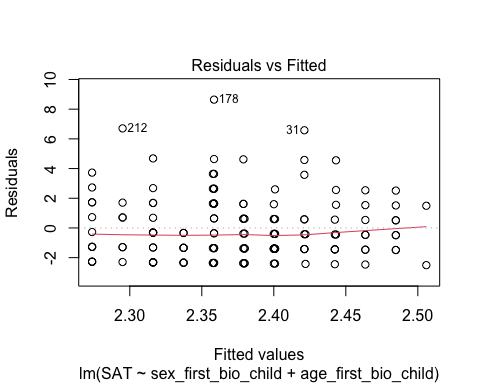


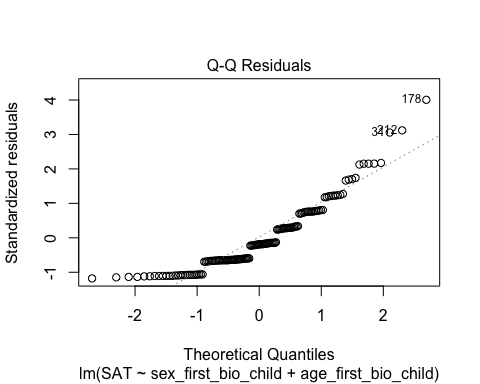


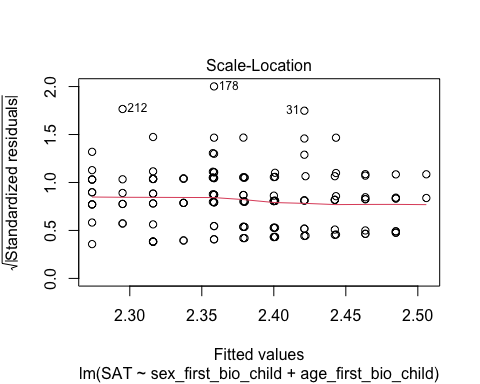


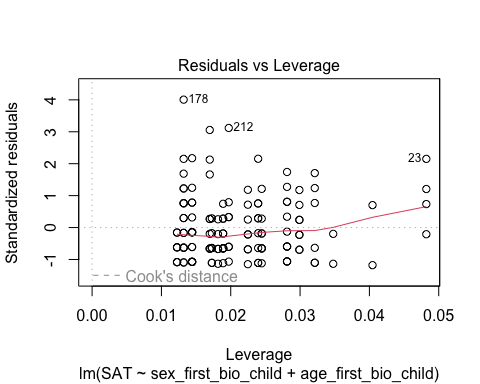


plot(MDH\_glm\_IPIP)









# Durbin-Watson test for independence of errors  
dwtest(MDH\_glm\_SAT)

Durbin-Watson test  
  
data: MDH\_glm\_SAT  
DW = 2.1188, p-value = 0.7698  
alternative hypothesis: true autocorrelation is greater than 0

dwtest(MDH\_glm\_IPIP)

Durbin-Watson test  
  
data: MDH\_glm\_IPIP  
DW = 2.1188, p-value = 0.7698  
alternative hypothesis: true autocorrelation is greater than 0

# Displaying the results  
summary(MDH\_glm\_SAT)

Call:  
lm(formula = SAT ~ sex\_first\_bio\_child + age\_first\_bio\_child,   
 data = female\_individuals)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-2.5059 -1.4212 -0.4008 1.5153 8.6415   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 2.33654 0.45238 5.165 8.23e-07 \*\*\*  
sex\_first\_bio\_childmale -0.08388 0.37072 -0.226 0.821   
age\_first\_bio\_child 0.02117 0.08575 0.247 0.805   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 2.171 on 138 degrees of freedom  
Multiple R-squared: 0.0007979, Adjusted R-squared: -0.01368   
F-statistic: 0.0551 on 2 and 138 DF, p-value: 0.9464

summary(MDH\_glm\_IPIP)

Call:  
lm(formula = SAT ~ sex\_first\_bio\_child + age\_first\_bio\_child,   
 data = female\_individuals)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-2.5059 -1.4212 -0.4008 1.5153 8.6415   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 2.33654 0.45238 5.165 8.23e-07 \*\*\*  
sex\_first\_bio\_childmale -0.08388 0.37072 -0.226 0.821   
age\_first\_bio\_child 0.02117 0.08575 0.247 0.805   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 2.171 on 138 degrees of freedom  
Multiple R-squared: 0.0007979, Adjusted R-squared: -0.01368   
F-statistic: 0.0551 on 2 and 138 DF, p-value: 0.9464

Most of the assumptions are met here, but the residuals vs. fitted values plots for both models look like there may be heteroscedasticity. Because of this, I will more formally test this with a Breusch-Pagan test.

# Conducting the Breusch-Pagan test for homogeneity of variance  
bptest(MDH\_glm\_SAT)

studentized Breusch-Pagan test  
  
data: MDH\_glm\_SAT  
BP = 0.72161, df = 2, p-value = 0.6971

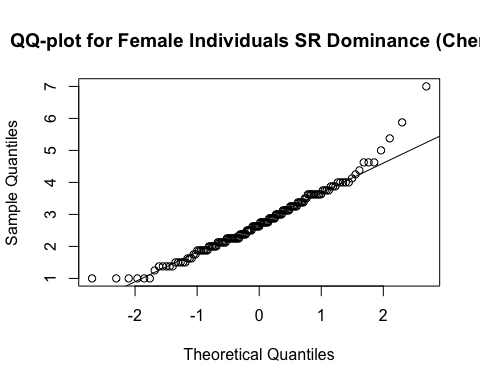
bptest(MDH\_glm\_IPIP)

studentized Breusch-Pagan test  
  
data: MDH\_glm\_IPIP  
BP = 0.72161, df = 2, p-value = 0.6971

The Breusch-Pagan test is not significant, so we will accept the null hypothesis that the variance of the residuals is constant across levels of the independent variable. Taking a look at the summaries of the models, it does not seem that adding in age\_first\_bio\_child as a covariate affected the way that sex\_first\_bio\_child relates (or does not relate, rather) to SAT or to IPIP dominance. Furthermore, neither model explains much of the dependent variable as determined by the F-tests.

Now I am going to do the same analysis for the self-reported dominance version of the dominance and prestige scale (Cheng et al., 2010) (honestly, because I forgot to do it the first time). First, to check the assumptions of the t-test we will do a QQ plot and a Levene’s test.

# QQ-plot for normality  
qqnorm(female\_individuals$SR\_dom\_cheng, main = "QQ-plot for Female Individuals SR Dominance (Cheng)")  
qqline(female\_individuals$SR\_dom\_cheng)



# Levene's Test to assess homogeneity of variances  
leveneTest(IPIP\_dom ~ sex\_first\_bio\_child, data = female\_individuals)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.2626 0.6092  
 139

The QQ-plot deviates from normality at the ends a bit, but I think with *n* = 140 observations we will be OK to assume normality for this distribution. Just in case, however, I will run the t-test as Students and as Welsch’s (which is more rhobust to violations of normality) to be sure. We can also accept the null hypothesis of homogeneity of variances as per the Levene’s test.

Now for the t-tests.

# Saving the results of the t-test  
mdh\_t\_test\_srcheng <- t.test(SR\_dom\_cheng ~ sex\_first\_bio\_child, data = female\_individuals, var.equal = TRUE)  
mdh\_w\_t\_test\_srcheng <- t.test(SR\_dom\_cheng ~ sex\_first\_bio\_child, data = female\_individuals, var.equal = FALSE)  
  
# Displaying the results  
mdh\_t\_test\_srcheng

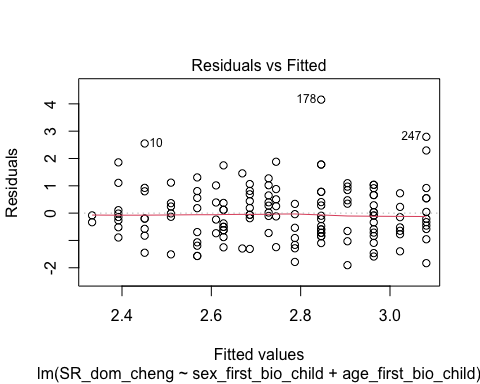
Two Sample t-test  
  
data: SR\_dom\_cheng by sex\_first\_bio\_child  
t = -1.9308, df = 139, p-value = 0.05555  
alternative hypothesis: true difference in means between group female and group male is not equal to 0  
95 percent confidence interval:  
 -0.671677323 0.007973313  
sample estimates:  
mean in group female mean in group male   
 2.561441 2.893293

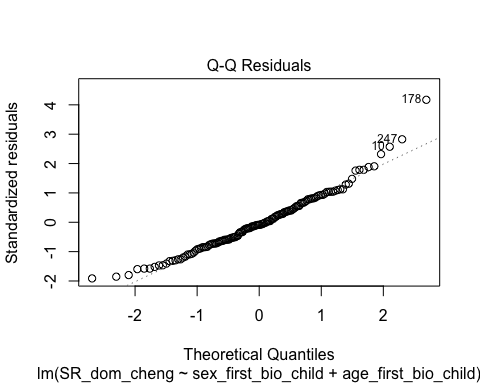
mdh\_w\_t\_test\_srcheng

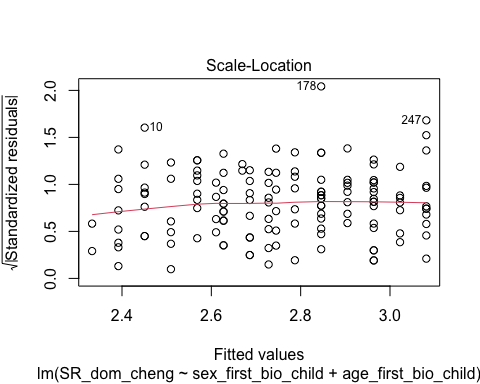
Welch Two Sample t-test  
  
data: SR\_dom\_cheng by sex\_first\_bio\_child  
t = -1.9718, df = 133.44, p-value = 0.0507  
alternative hypothesis: true difference in means between group female and group male is not equal to 0  
95 percent confidence interval:  
 -0.664726135 0.001022125  
sample estimates:  
mean in group female mean in group male   
 2.561441 2.893293

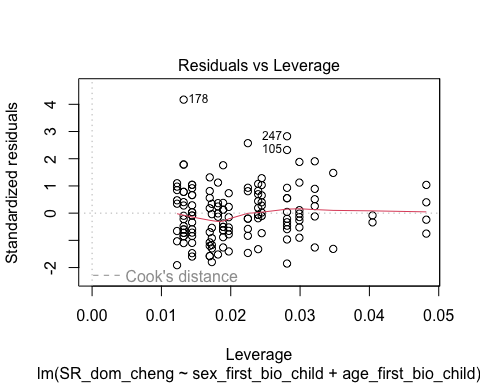
Interestingly, the results of this t-test—although not significant at our alpha level—indicate that the mothers of female first-born children are lower in self-reported behavioral dominance than the mothers of first-born children, as would be predicted by the MDH. I would like to see whether this result becomes more robust if we put it into the general linear model with age of first-born child as a coveriate (as done above).

# Constructing the model  
MDH\_glm\_srcheng <- lm(SR\_dom\_cheng ~ sex\_first\_bio\_child + age\_first\_bio\_child, data = female\_individuals)  
  
# Plotting residuals vs. fitted values for linearity and homoscedasticity and producing QQ plots for normality of residuals  
plot(MDH\_glm\_srcheng)









# Durbin-Watson test for independence of errors  
dwtest(MDH\_glm\_srcheng)

Durbin-Watson test  
  
data: MDH\_glm\_srcheng  
DW = 1.7273, p-value = 0.05483  
alternative hypothesis: true autocorrelation is greater than 0

# Displaying the results  
summary(MDH\_glm\_srcheng)

Call:  
lm(formula = SR\_dom\_cheng ~ sex\_first\_bio\_child + age\_first\_bio\_child,   
 data = female\_individuals)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-1.9048 -0.6934 -0.0886 0.6468 4.1541   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 2.80380 0.20886 13.424 <2e-16 \*\*\*  
sex\_first\_bio\_childmale 0.33635 0.17116 1.965 0.0514 .   
age\_first\_bio\_child -0.05885 0.03959 -1.486 0.1395   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 1.002 on 138 degrees of freedom  
Multiple R-squared: 0.04146, Adjusted R-squared: 0.02757   
F-statistic: 2.985 on 2 and 138 DF, p-value: 0.05383

Like above, I want to ensure that the residuals for the model are have constant variance, so I will formally test this using the Breusch-Pagan test.

bptest(MDH\_glm\_srcheng)

studentized Breusch-Pagan test  
  
data: MDH\_glm\_srcheng  
BP = 0.75531, df = 2, p-value = 0.6855

The Breusch-Pagan test indicates we should accept the null of homogeneity of variance.

Looking at the output of the summary of the model, we can see that the F-test was not significant, and the predictive ability of sex of first born child did not flip significance.

In sum, it does not look like there is evidence here for the individual differences version of the maternal dominance hypothesis. Although the t-test using self-reported dominance was close to significant, I suspect that this is due to sampling variability, given the lack of evidence (and, indeed, a reversed effect) for the other dominance measures, including the SAT, which showed the original effect.

## Testing Hypothesis 2

### Notes About Subsequent Analyses

* Because the models involved in testing our Hypothesis 2 are moderation models and require probing interactions for effects of the focal predictor at different levels of the moderator, I wanted to find a statistical package that would be able to automate the simple-slopes analysis, becuase it can be quite tedious, especially when working with multiple different models, requiring the creation of many different variables. I was also interested in finding a package that can do the Johnson-Neyman procedure to probe for “regions of significance”—the exact range of the moderator for which the slope of the focal predictor is significant (Johnson & Neyman, 1936). This approach is especially advantageous in cases such as ours where the choice of the exact high and low values of the moderator is relatively arbitrary and this choice may not represent the nature of the interaction at different values than we have chosen. Because our hypothesis actually predicts an interaction effect at high levels of the moderator *a priori* (Palmer-Hague & Watson, 2016), I needed to find something that would probe the interaction at different levels of the moderator without requiring the interaction in the model to be significant. ‘interactions’ allows for probing of interactions using both simple-slopes analysis and Johnson-Neyman intervals of significance in either general linear models or generalized linear models (Long, 2021). Therefore, in the following section of the analysis we will be using the base R package GLM that fits generalized linear models to fit our binary logistic regression models and assess the model summary statistics, then we will use the ‘interactions’ package to probe interaction effects using both simple-slopes and Johnson-Neyman output.
* As we saw in the [Assessing Whether Facial Ratings and fWHR Vary by Facial Expression Within Sexes](#Xe5fdb858c19e1003ef84cf3ba27fc6f550f88a5) section, our non-neutral facial expression and neutral facial expression groups differ in their aggregate levels of facial rating characteristics. Because of this, when we do analyses with facial dominance as a predictor we will repeat the analysis within a dataframe that only contains the neutral facial expressions to check if the result is still robust.
* To simplify the interpretation of our models, each of the predictor variables will be standardized before entering them in as predictors. This will also make the interaction effects more easily interpretable.
* Each model will be constructed with only the main effects first, then the interaction effect will be added to determine whether it significantly improves the fit of the model (using a likelihood ratio test).
* When predictors significantly add to the model fit, their coefficients will be converted to odds-ratios through exponentiation.
* shared\_child\_sex is coded as female = 0 and male = 1, so in all analyses we are modeling the probability of having a first born son.

### Hypothesis 2: Behavioral Dominance Measure Operationalizations

We will begin by assessing whether maternal and paternal behavioral dominance (the SAT and the IPIP) interact to predict offspring sex.

#### SAT

First, we will standardize our independent variables, m\_SAT and f\_SAT.

# Standardizing the m\_SAT and f\_SAT variables as z\_m\_SAT and z\_f\_SAT  
dyadic\_data$z\_m\_SAT <- scale(dyadic\_data$m\_SAT, center = TRUE, scale = TRUE)  
dyadic\_data$z\_f\_SAT <- scale(dyadic\_data$f\_SAT, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

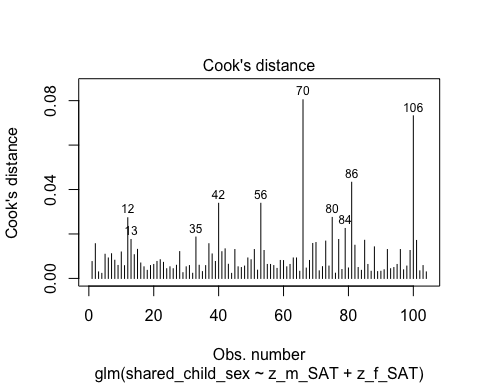
# Fitting the model with only the main effects of SAT for mothers and fathers  
SAT\_main\_effects\_model <- glm(shared\_child\_sex ~ z\_m\_SAT + z\_f\_SAT, family = binomial(link = logit), data = dyadic\_data)  
  
# Fitting the model with the main effects of SAT for mothers and fathers as well as their interaction  
SAT\_interaction\_model <- glm(shared\_child\_sex ~ z\_m\_SAT + z\_f\_SAT + z\_m\_SAT:z\_f\_SAT, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

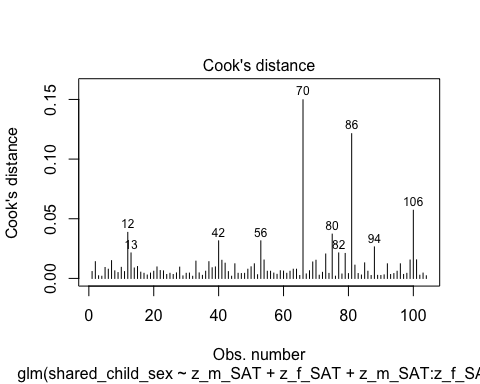
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(SAT\_main\_effects\_model, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(SAT\_interaction\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(SAT\_main\_effects\_model)

z\_m\_SAT z\_f\_SAT   
1.044477 1.044477

1/vif(SAT\_main\_effects\_model)

z\_m\_SAT z\_f\_SAT   
0.9574171 0.9574171

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. The below code determines the lowest value of z\_m\_SAT and z\_f\_sat so that we can add a constant that makes all values for the variable positive before making the natural log transformation.

min(dyadic\_data$z\_m\_SAT)

[1] -1.08544

min(dyadic\_data$z\_f\_SAT)

[1] -1.125062

* We can see that the lowest value for each variable is between -1 and -2, so I will add 2 to each variable then take the natural log of the resulting variable.

# Creating c\_z\_m\_SAT, which represents mothers' SAT scores after adding a constant of 2  
dyadic\_data$c\_z\_m\_SAT <- dyadic\_data$z\_m\_SAT + 2  
  
# Creating ln\_c\_z\_m\_SAT, which represents the natural log of mothers' SAT scores after adding a constant of 2  
dyadic\_data$ln\_c\_z\_m\_SAT <- log(dyadic\_data$c\_z\_m\_SAT)  
  
# Creating c\_z\_f\_SAT, which represents fathers' SAT scores after adding a constant of 2  
dyadic\_data$c\_z\_f\_SAT <- dyadic\_data$z\_f\_SAT + 2  
  
# Creating ln\_c\_z\_f\_SAT, which represents the natural log of fathers' SAT scores after adding a constant of 2  
dyadic\_data$ln\_c\_z\_f\_SAT <- log(dyadic\_data$c\_z\_f\_SAT)

* Through manually doing the transformations for a couple numbers, I believe that this transformation was done correctly.
* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_SAT\_model <- glm(shared\_child\_sex ~ c\_z\_m\_SAT + c\_z\_f\_SAT + c\_z\_m\_SAT:ln\_c\_z\_m\_SAT + c\_z\_f\_SAT:ln\_c\_z\_f\_SAT, family = binomial(link = logit), data = dyadic\_data)  
  
summary(BT\_test\_SAT\_model)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_SAT + c\_z\_f\_SAT + c\_z\_m\_SAT:ln\_c\_z\_m\_SAT +   
 c\_z\_f\_SAT:ln\_c\_z\_f\_SAT, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 4.6833 2.6238 1.785 0.0743 .  
c\_z\_m\_SAT -4.4486 1.9238 -2.312 0.0208 \*  
c\_z\_f\_SAT 0.7132 1.6524 0.432 0.6660   
c\_z\_m\_SAT:ln\_c\_z\_m\_SAT 2.3856 1.0679 2.234 0.0255 \*  
c\_z\_f\_SAT:ln\_c\_z\_f\_SAT -0.3565 0.8972 -0.397 0.6911   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 132.62 on 99 degrees of freedom  
AIC: 142.62  
  
Number of Fisher Scoring iterations: 4

* Unfortunately, the significant p-value (*p* = .0256) for the interaction term for mothers’ SAT indicates that the relationship between mothers’ SAT scores and the logit is not linear, which violates our assumption of linearity of the logit. The positive coefficient suggests that as mothers’ SAT increases, it’s effect on the log odds of having a first born son increase in a non-linear fashion.
  + I will try to do a square root transformation to the standardized mothers’ SAT scores to see if this makes the relationship between the predictor and the logit linear.

# Square root transformation of the standardized mothers' SAT + the constant  
dyadic\_data$sqrt\_c\_z\_m\_SAT <- sqrt(dyadic\_data$c\_z\_m\_SAT)  
  
# Taking the natural log of the square root transformed variable for input into another Box-Tidwell procedure  
dyadic\_data$ln\_sqrt\_c\_z\_m\_SAT <- log(dyadic\_data$sqrt\_c\_z\_m\_SAT)  
  
# Re-running the box-tidwell procedure with the square-root transformation  
sqrt\_m\_SAT\_BT\_model <- glm(shared\_child\_sex ~ sqrt\_c\_z\_m\_SAT + c\_z\_f\_SAT + sqrt\_c\_z\_m\_SAT:ln\_sqrt\_c\_z\_m\_SAT + c\_z\_f\_SAT:ln\_c\_z\_f\_SAT, family = binomial(link = logit), data = dyadic\_data)  
  
summary(sqrt\_m\_SAT\_BT\_model)

Call:  
glm(formula = shared\_child\_sex ~ sqrt\_c\_z\_m\_SAT + c\_z\_f\_SAT +   
 sqrt\_c\_z\_m\_SAT:ln\_sqrt\_c\_z\_m\_SAT + c\_z\_f\_SAT:ln\_c\_z\_f\_SAT,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 18.8874 8.1604 2.315 0.0206 \*  
sqrt\_c\_z\_m\_SAT -18.5765 7.9706 -2.331 0.0198 \*  
c\_z\_f\_SAT 0.6690 1.6497 0.406 0.6851   
sqrt\_c\_z\_m\_SAT:ln\_sqrt\_c\_z\_m\_SAT 13.2559 5.8556 2.264 0.0236 \*  
c\_z\_f\_SAT:ln\_c\_z\_f\_SAT -0.3284 0.8944 -0.367 0.7135   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 132.57 on 99 degrees of freedom  
AIC: 142.57  
  
Number of Fisher Scoring iterations: 4

* The square root transformation did not seem to make a difference, as the p-value is still significant, indicating a non-linear relationship between the predictor and the logit.
* I will now try an inverse transformation to see if that makes a difference.

# Taking the inverse of the standardized mothers' SAT + the constant  
dyadic\_data$inv\_c\_z\_m\_SAT <- 1/(dyadic\_data$c\_z\_m\_SAT)  
  
# Taking the natural log of the inverse transformed variable for input into another Box-Tidwell procedure  
dyadic\_data$ln\_inv\_c\_z\_m\_SAT <- log(dyadic\_data$inv\_c\_z\_m\_SAT)  
  
# Re-running the Box-Tidwell procedure with the new transformations  
inv\_m\_SAT\_BT\_model <- glm(shared\_child\_sex ~ inv\_c\_z\_m\_SAT + c\_z\_f\_SAT + inv\_c\_z\_m\_SAT:ln\_inv\_c\_z\_m\_SAT + c\_z\_f\_SAT:ln\_c\_z\_f\_SAT, family = binomial(link = logit), data = dyadic\_data)  
  
summary(inv\_m\_SAT\_BT\_model)

Call:  
glm(formula = shared\_child\_sex ~ inv\_c\_z\_m\_SAT + c\_z\_f\_SAT +   
 inv\_c\_z\_m\_SAT:ln\_inv\_c\_z\_m\_SAT + c\_z\_f\_SAT:ln\_c\_z\_f\_SAT,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 3.2593 2.9230 1.115 0.2648   
inv\_c\_z\_m\_SAT -2.7209 2.1783 -1.249 0.2116   
c\_z\_f\_SAT 0.4155 1.6368 0.254 0.7996   
inv\_c\_z\_m\_SAT:ln\_inv\_c\_z\_m\_SAT 7.0060 3.6776 1.905 0.0568 .  
c\_z\_f\_SAT:ln\_c\_z\_f\_SAT -0.1880 0.8852 -0.212 0.8318   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 133.28 on 99 degrees of freedom  
AIC: 143.28  
  
Number of Fisher Scoring iterations: 4

* This transformation barely flipped the significance of the p-value for the interaction term.
* With these transformations not helping very much, I do not know what else to do. I will therefore finish the analysis using the standardized variables as predictors like we fit in the beginning and avoid over-interpreting confidence intervals or p-values.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ SAT and fathers’ SAT only, along with the Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(SAT\_main\_effects\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_SAT + z\_f\_SAT, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.43013 0.20082 2.142 0.0322 \*  
z\_m\_SAT -0.07761 0.20481 -0.379 0.7047   
z\_f\_SAT 0.02009 0.20642 0.097 0.9225   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 139.34 on 101 degrees of freedom  
AIC: 145.34  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_SAT\_main\_effects\_model <- SAT\_main\_effects\_model$null.deviance - SAT\_main\_effects\_model$deviance  
Chi\_SAT\_main\_effects\_model

[1] 0.1435668

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_SAT\_main\_effects\_model <- SAT\_main\_effects\_model$df.null - SAT\_main\_effects\_model$df.residual  
df\_SAT\_main\_effects\_model

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_SAT\_main\_effects\_model <- 1 - pchisq(Chi\_SAT\_main\_effects\_model, df\_SAT\_main\_effects\_model)  
prob\_Chi\_SAT\_main\_effects\_model

[1] 0.9307325

* The model with main effects only seems to be a poor fit for the data, and, although the p-value may be biased due to the violation of the linearity assumption, it is very close to one, indicating that the predictors added virtually nothing to the model with only the intercept included.

Now we will take a quick look at the model with the interaction included, and we will compare it’s fit to the model with only the main effects included.

# Producing the summary of the interaction model  
summary(SAT\_interaction\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_SAT + z\_f\_SAT + z\_m\_SAT:z\_f\_SAT,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.41083 0.20398 2.014 0.044 \*  
z\_m\_SAT -0.08591 0.20638 -0.416 0.677   
z\_f\_SAT 0.01025 0.20723 0.049 0.961   
z\_m\_SAT:z\_f\_SAT 0.09996 0.18963 0.527 0.598   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 139.06 on 100 degrees of freedom  
AIC: 147.06  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the interaction model compared to the model with only main effects included  
Chi\_SAT\_interaction\_model\_v\_main <- SAT\_main\_effects\_model$deviance - SAT\_interaction\_model$deviance  
Chi\_SAT\_interaction\_model\_v\_main

[1] 0.2845532

# Calculating the degrees of freedom for the chi-square statistic comparing the model with the interaction to the main-effects model  
df\_SAT\_interaction\_model\_v\_main <- SAT\_main\_effects\_model$df.residual - SAT\_interaction\_model$df.residual  
df\_SAT\_interaction\_model\_v\_main

[1] 1

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_SAT\_interaction\_model\_v\_main <- 1 - pchisq(Chi\_SAT\_interaction\_model\_v\_main, df\_SAT\_interaction\_model\_v\_main)  
prob\_Chi\_SAT\_interaction\_model\_v\_main

[1] 0.5937323

* The coefficients for each predictor are very small, and they are not close to significant. Similarly, the model with the interaction included does not fit the data better than the main-effects only model.
* We will look for whether the interaction model is significant compared to the baseline with only the intercept.

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_SAT\_interaction\_model <- SAT\_interaction\_model$null.deviance - SAT\_interaction\_model$deviance  
Chi\_SAT\_interaction\_model

[1] 0.42812

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_SAT\_interaction\_model <- SAT\_interaction\_model$df.null - SAT\_interaction\_model$df.residual  
df\_SAT\_interaction\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_SAT\_interaction\_model <- 1 - pchisq(Chi\_SAT\_interaction\_model, df\_SAT\_interaction\_model)  
prob\_Chi\_SAT\_interaction\_model

[1] 0.9343742

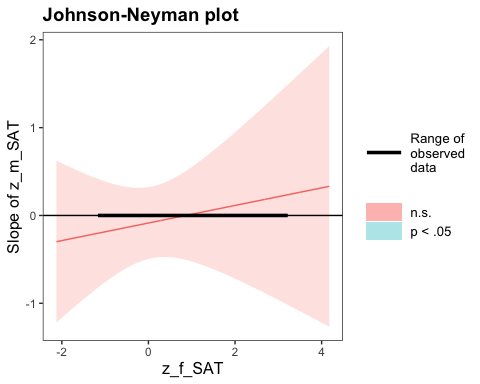
* The model is not significant.

##### Probing for an Interaction

For completion’s sake—although I am almost sure we will not find anything—I figure I may as well take a look at potential differences in slope of the focal predictor (mother’s standardized SAT) at different levels of the moderator (fathers’ standardized SAT). The following is the simple slopes analysis and Johnson-Neyman plot for the interaction model.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(SAT\_interaction\_model, pred = z\_m\_SAT, modx = z\_f\_SAT, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_SAT when z\_f\_SAT = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------- ------ -------- ------  
 -0.19 0.29 -0.64 0.52  
  
Slope of z\_m\_SAT when z\_f\_SAT = -1.494531e-17 (Mean):   
  
 Est. S.E. z val. p  
------- ------ -------- ------  
 -0.09 0.21 -0.42 0.68  
  
Slope of z\_m\_SAT when z\_f\_SAT = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.01 0.27 0.05 0.96

* As suspected, Johnson-Neyman result indicates that there are no values of fathers’ SAT where mothers’ SAT significantly predicts the probability of having a first-born son.

Although it may be that the estimate of the coefficient for mothers’ SAT scores is attenuated due to a truly non-linear relationship such as a growth curve, with the current modeling framework there does not seem to be evidence that mothers’ SAT positively predicts the probability of having a first born child when fathers’ SAT score is high.

#### IPIP Dominance

First, we will standardize our independent variables, m\_IPIP\_dom and f\_IPIP\_dom.

# Standardizing the m\_IPIP\_dom and f\_IPIP\_dom variables as z\_m\_IPIP\_dom and z\_f\_IPIP\_dom  
dyadic\_data$z\_m\_IPIP\_dom <- scale(dyadic\_data$m\_IPIP\_dom, center = TRUE, scale = TRUE)  
dyadic\_data$z\_f\_IPIP\_dom <- scale(dyadic\_data$f\_IPIP\_dom, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

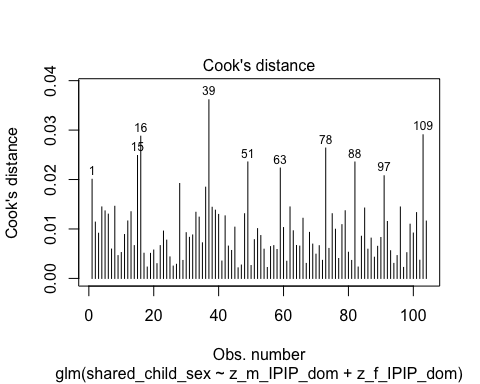
# Fitting the model with only the main effects of IPIP dominance for mothers and fathers  
IPIP\_main\_effects\_model <- glm(shared\_child\_sex ~ z\_m\_IPIP\_dom + z\_f\_IPIP\_dom, family = binomial(link = logit), data = dyadic\_data)  
  
# Fitting the model with the main effects of IPIP dominance for mothers and fathers as well as their interaction  
IPIP\_interaction\_model <- glm(shared\_child\_sex ~ z\_m\_IPIP\_dom + z\_f\_IPIP\_dom + z\_m\_IPIP\_dom:z\_f\_IPIP\_dom, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

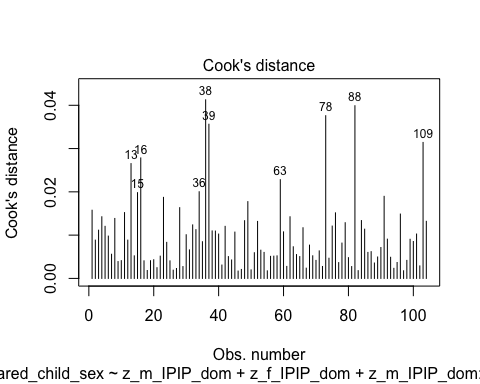
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(IPIP\_main\_effects\_model, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(IPIP\_interaction\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(IPIP\_main\_effects\_model)

z\_m\_IPIP\_dom z\_f\_IPIP\_dom   
 1.063057 1.063057

1/vif(IPIP\_main\_effects\_model)

z\_m\_IPIP\_dom z\_f\_IPIP\_dom   
 0.9406837 0.9406837

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, and in this case I am pretty sure that the raw IPIP variable does not contain zero or negative numbers. I will check below by looking at the minimum value for the raw IPIP dominance measure for both mothers and fathers and if this is the case I will simply do the log transformation to this variable for use in the Box-Tidwell procedure.

# Checking the minimum value for IPIP dominance for mothers and fathers  
min(dyadic\_data$m\_IPIP\_dom, na.rm = TRUE)

[1] 11

min(dyadic\_data$f\_IPIP\_dom, na.rm = TRUE)

[1] 14

* Indeed, the mimimum value for the raw variables is positive for both, so I will do the natural log transformation of each of these variables for input as interaction terms to complete the Box-Tidwell procedure.

# Creating ln\_m\_IPIP\_dom, which represents the natural log of mothers' IPIP dominance scores  
dyadic\_data$ln\_m\_IPIP\_dom <- log(dyadic\_data$m\_IPIP\_dom)  
  
# Creating ln\_f\_IPIP\_dom, which represents the natural log of fathers' IPIP dominance scores  
dyadic\_data$ln\_f\_IPIP\_dom <- log(dyadic\_data$f\_IPIP\_dom)

* Through manually doing the transformations for a couple numbers, I believe that this transformation was done correctly.
* Now we need to fit and summarize a model with the main effects of each variable plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the BT procedure  
BT\_test\_IPIP\_model <- glm(shared\_child\_sex ~ m\_IPIP\_dom + f\_IPIP\_dom + m\_IPIP\_dom:ln\_m\_IPIP\_dom + f\_IPIP\_dom:ln\_f\_IPIP\_dom, family = binomial(link = logit), data = dyadic\_data)  
  
# Displaying the results  
summary(BT\_test\_IPIP\_model)

Call:  
glm(formula = shared\_child\_sex ~ m\_IPIP\_dom + f\_IPIP\_dom + m\_IPIP\_dom:ln\_m\_IPIP\_dom +   
 f\_IPIP\_dom:ln\_f\_IPIP\_dom, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 12.0595 6.7964 1.774 0.076 .  
m\_IPIP\_dom -0.8261 0.6574 -1.257 0.209   
f\_IPIP\_dom -1.0217 0.8494 -1.203 0.229   
m\_IPIP\_dom:ln\_m\_IPIP\_dom 0.1980 0.1544 1.282 0.200   
f\_IPIP\_dom:ln\_f\_IPIP\_dom 0.2281 0.1939 1.176 0.239   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 134.84 on 99 degrees of freedom  
AIC: 144.84  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ IPIP dominance and fathers’ IPIP dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(IPIP\_main\_effects\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_IPIP\_dom + z\_f\_IPIP\_dom,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4357 0.2022 2.155 0.0312 \*  
z\_m\_IPIP\_dom 0.1669 0.2114 0.789 0.4298   
z\_f\_IPIP\_dom -0.2094 0.2107 -0.994 0.3202   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 138.14 on 101 degrees of freedom  
AIC: 144.14  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_IPIP\_main\_effects\_model <- IPIP\_main\_effects\_model$null.deviance - IPIP\_main\_effects\_model$deviance  
Chi\_IPIP\_main\_effects\_model

[1] 1.340487

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_IPIP\_main\_effects\_model <- IPIP\_main\_effects\_model$df.null - IPIP\_main\_effects\_model$df.residual  
df\_IPIP\_main\_effects\_model

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_IPIP\_main\_effects\_model <- 1 - pchisq(Chi\_IPIP\_main\_effects\_model, df\_IPIP\_main\_effects\_model)  
prob\_Chi\_IPIP\_main\_effects\_model

[1] 0.511584

* Looking at the summary of the model, we can see that neither predictor is significant at the .05 level, and the Chi square comparison to test whether the main effects of mother and father IPIP dominance improve the fit of the model is also not significant (χ²(2) = 1.340, *p* = .512).

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(IPIP\_interaction\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_IPIP\_dom + z\_f\_IPIP\_dom +   
 z\_m\_IPIP\_dom:z\_f\_IPIP\_dom, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.43231 0.20785 2.080 0.0375 \*  
z\_m\_IPIP\_dom 0.16751 0.21142 0.792 0.4282   
z\_f\_IPIP\_dom -0.21086 0.21158 -0.997 0.3190   
z\_m\_IPIP\_dom:z\_f\_IPIP\_dom 0.01492 0.21056 0.071 0.9435   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 138.14 on 100 degrees of freedom  
AIC: 146.14  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_IPIP\_interaction\_model <- IPIP\_interaction\_model$null.deviance - IPIP\_interaction\_model$deviance  
Chi\_IPIP\_interaction\_model

[1] 1.34551

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_IPIP\_interaction\_model <- IPIP\_interaction\_model$df.null - IPIP\_interaction\_model$df.residual  
df\_IPIP\_interaction\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_IPIP\_interaction\_model <- 1 - pchisq(Chi\_IPIP\_interaction\_model, df\_IPIP\_interaction\_model)  
prob\_Chi\_IPIP\_interaction\_model

[1] 0.7183556

* The model summary indicates that none of the predictors are significant, including the interaction term. In addition, the model does not fit the data better than the intercept-only version of the model.

Looking at the summary of the interaction model, it would be very surprising that the interaction model would fit any better than the main-effects model, but I will go ahead and do it anyways.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_IPIP\_interaction\_model\_v\_main <- IPIP\_main\_effects\_model$deviance - IPIP\_interaction\_model$deviance  
Chi\_IPIP\_interaction\_model\_v\_main

[1] 0.005023363

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_IPIP\_interaction\_model\_v\_main <- IPIP\_main\_effects\_model$df.residual - IPIP\_interaction\_model$df.residual  
df\_IPIP\_interaction\_model\_v\_main

[1] 1

prob\_Chi\_IPIP\_interaction\_model\_v\_main <- 1 - pchisq(Chi\_IPIP\_interaction\_model\_v\_main, df\_IPIP\_interaction\_model\_v\_main)  
prob\_Chi\_IPIP\_interaction\_model\_v\_main

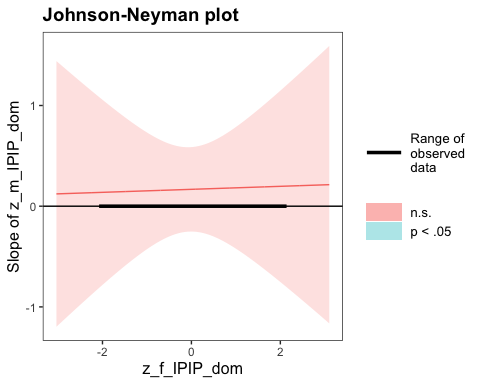
[1] 0.9434967

* As suspected, the Chi-squared test does not indicate that the interaction model fits better than the main effects model (χ²(1) = .005, *p* = .943).

Again, for completion’s sake, I will run the simple slopes and Johnson-Neyman analysis for the interaction model.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(IPIP\_interaction\_model, pred = z\_m\_IPIP\_dom, modx = z\_f\_IPIP\_dom, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_IPIP\_dom when z\_f\_IPIP\_dom = -1.00 (- 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.15 0.29 0.52 0.60  
  
Slope of z\_m\_IPIP\_dom when z\_f\_IPIP\_dom = 0.00 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.17 0.21 0.79 0.43  
  
Slope of z\_m\_IPIP\_dom when z\_f\_IPIP\_dom = 1.00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.18 0.30 0.60 0.55

* As suspected, Johnson-Neyman result indicates that there are no values of fathers’ SAT where mothers’ SAT significantly predicts the probability of having a first-born son.

### Hypothesis 2: Dominance Status Measure Operationalization

Now we will test the hypothesis with self-reported dominance status (Cheng et al., 2010).

#### Self-Reported Dominance Status

First we will standardize our independent variables, m\_SR\_dom\_cheng and f\_SR\_dom\_cheng.

# Standardizing the m\_SR\_dom\_cheng and f\_SR\_dom\_cheng variables as z\_m\_SR\_dom\_cheng and z\_f\_SR\_dom\_cheng  
dyadic\_data$z\_m\_SR\_dom\_cheng <- scale(dyadic\_data$m\_SR\_dom\_cheng, center = TRUE, scale = TRUE)  
dyadic\_data$z\_f\_SR\_dom\_cheng <- scale(dyadic\_data$f\_SR\_dom\_cheng, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

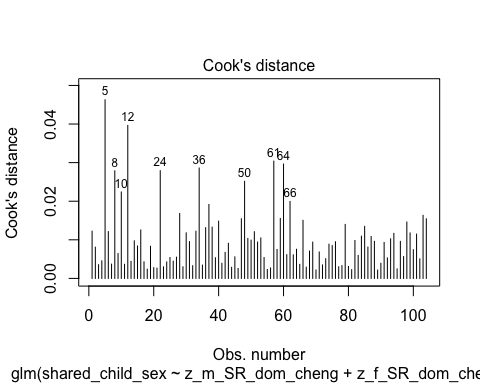
# Fitting the model with only the main effects of IPIP dominance for mothers and fathers  
sr\_cheng\_main\_effects\_model <- glm(shared\_child\_sex ~ z\_m\_SR\_dom\_cheng + z\_f\_SR\_dom\_cheng, family = binomial(link = logit), data = dyadic\_data)  
  
# Fitting the model with the main effects of IPIP dominance for mothers and fathers as well as their interaction  
sr\_cheng\_interaction\_model <- glm(shared\_child\_sex ~ z\_m\_SR\_dom\_cheng + z\_f\_SR\_dom\_cheng + z\_m\_SR\_dom\_cheng:z\_f\_SR\_dom\_cheng, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

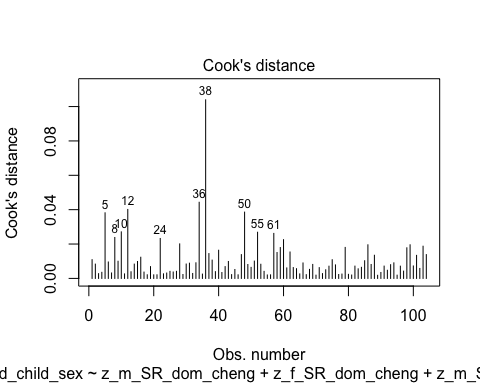
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(sr\_cheng\_main\_effects\_model, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(sr\_cheng\_interaction\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(sr\_cheng\_main\_effects\_model)

z\_m\_SR\_dom\_cheng z\_f\_SR\_dom\_cheng   
 1.208507 1.208507

1/vif(sr\_cheng\_main\_effects\_model)

z\_m\_SR\_dom\_cheng z\_f\_SR\_dom\_cheng   
 0.8274674 0.8274674

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, and in this case I am pretty sure that the raw SR dominance variable does not contain zero or negative numbers. I will check below by looking at the minimum value for the raw SR dominance measure for both mothers and fathers and if this is the case I will simply do the log transformation to this variable for use in the Box-Tidwell procedure.

# Checking the minimum value for IPIP dominance for mothers and fathers  
min(dyadic\_data$m\_SR\_dom\_cheng, na.rm = TRUE)

[1] 1

min(dyadic\_data$f\_SR\_dom\_cheng, na.rm = TRUE)

[1] 1.125

* Indeed, the mimimum value for the raw variables is positive for both, so I will do the natural log transformation of each of these variables for input as interaction terms to complete the Box-Tidwell procedure.

# Creating ln\_m\_SR\_dom\_cheng, which represents the natural log of mothers' SR dominance scores  
dyadic\_data$ln\_m\_SR\_dom\_cheng <- log(dyadic\_data$m\_SR\_dom\_cheng)  
  
# Creating ln\_f\_SR\_dom\_cheng, which represents the natural log of fathers' SR dominance scores  
dyadic\_data$ln\_f\_SR\_dom\_cheng <- log(dyadic\_data$f\_SR\_dom\_cheng)

* Through manually doing the transformations for a couple numbers, I believe that this transformation was done correctly.
* Now we need to fit and summarize a model with the main effects of each variable plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

# Fitting the model for the BT procedure  
BT\_test\_SR\_cheng\_model <- glm(shared\_child\_sex ~ m\_SR\_dom\_cheng + f\_SR\_dom\_cheng + m\_SR\_dom\_cheng:ln\_m\_SR\_dom\_cheng + f\_SR\_dom\_cheng:ln\_f\_SR\_dom\_cheng, family = binomial(link = logit), data = dyadic\_data)  
  
# Displaying the results  
summary(BT\_test\_SR\_cheng\_model)

Call:  
glm(formula = shared\_child\_sex ~ m\_SR\_dom\_cheng + f\_SR\_dom\_cheng +   
 m\_SR\_dom\_cheng:ln\_m\_SR\_dom\_cheng + f\_SR\_dom\_cheng:ln\_f\_SR\_dom\_cheng,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -0.3412 3.5538 -0.096 0.924  
m\_SR\_dom\_cheng -0.1285 1.5976 -0.080 0.936  
f\_SR\_dom\_cheng 0.5728 1.9817 0.289 0.773  
m\_SR\_dom\_cheng:ln\_m\_SR\_dom\_cheng 0.1847 0.7715 0.239 0.811  
f\_SR\_dom\_cheng:ln\_f\_SR\_dom\_cheng -0.3223 0.9223 -0.349 0.727  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 138.05 on 99 degrees of freedom  
AIC: 148.05  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ SR dominance status and fathers’ SR dominance status only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(sr\_cheng\_main\_effects\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_SR\_dom\_cheng + z\_f\_SR\_dom\_cheng,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4356 0.2022 2.155 0.0312 \*  
z\_m\_SR\_dom\_cheng 0.2541 0.2321 1.095 0.2735   
z\_f\_SR\_dom\_cheng -0.1218 0.2235 -0.545 0.5857   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 138.23 on 101 degrees of freedom  
AIC: 144.23  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_sr\_cheng\_main\_effects\_model <- sr\_cheng\_main\_effects\_model$null.deviance - sr\_cheng\_main\_effects\_model$deviance  
Chi\_sr\_cheng\_main\_effects\_model

[1] 1.257427

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_sr\_cheng\_main\_effects\_model <- sr\_cheng\_main\_effects\_model$df.null - sr\_cheng\_main\_effects\_model$df.residual  
df\_sr\_cheng\_main\_effects\_model

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_sr\_cheng\_main\_effects\_model <- 1 - pchisq(Chi\_sr\_cheng\_main\_effects\_model, df\_sr\_cheng\_main\_effects\_model)  
prob\_Chi\_sr\_cheng\_main\_effects\_model

[1] 0.5332774

* Looking at the summary of the model, we can see that neither predictor is significant at the .05 level, and the Chi square comparison to test whether the main effects of mother and father SR dominance status improve the fit of the model is also not significant (χ²(2) = 1.257, *p* = .533).

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(sr\_cheng\_interaction\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_SR\_dom\_cheng + z\_f\_SR\_dom\_cheng +   
 z\_m\_SR\_dom\_cheng:z\_f\_SR\_dom\_cheng, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.5510 0.2297 2.399 0.0164 \*  
z\_m\_SR\_dom\_cheng 0.3576 0.2548 1.403 0.1605   
z\_f\_SR\_dom\_cheng -0.1062 0.2280 -0.466 0.6412   
z\_m\_SR\_dom\_cheng:z\_f\_SR\_dom\_cheng -0.2644 0.2323 -1.138 0.2552   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 139.49 on 103 degrees of freedom  
Residual deviance: 136.91 on 100 degrees of freedom  
AIC: 144.91  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_sr\_cheng\_interaction\_model <- sr\_cheng\_interaction\_model$null.deviance - sr\_cheng\_interaction\_model$deviance  
Chi\_sr\_cheng\_interaction\_model

[1] 2.572065

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_sr\_cheng\_interaction\_model <- sr\_cheng\_interaction\_model$df.null - sr\_cheng\_interaction\_model$df.residual  
df\_sr\_cheng\_interaction\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_sr\_cheng\_interaction\_model <- 1 - pchisq(Chi\_sr\_cheng\_interaction\_model, df\_sr\_cheng\_interaction\_model)  
prob\_Chi\_sr\_cheng\_interaction\_model

[1] 0.462408

* The model summary indicates that none of the predictors are significant, including the interaction term. In addition, the model does not fit the data better than the intercept-only version of the model.

Looking at the estimate and p-value for the interaction term, it would be very unlikely that the interaction model would fit any better than the main-effects model, but I will go ahead and do the comparison anyways.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_sr\_cheng\_interaction\_model\_v\_main <- sr\_cheng\_main\_effects\_model$deviance - sr\_cheng\_interaction\_model$deviance  
Chi\_sr\_cheng\_interaction\_model\_v\_main

[1] 1.314638

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_sr\_cheng\_interaction\_model\_v\_main <- sr\_cheng\_main\_effects\_model$df.residual - sr\_cheng\_interaction\_model$df.residual  
df\_sr\_cheng\_interaction\_model\_v\_main

[1] 1

prob\_Chi\_sr\_cheng\_interaction\_model\_v\_main <- 1 - pchisq(Chi\_sr\_cheng\_interaction\_model\_v\_main, df\_sr\_cheng\_interaction\_model\_v\_main)  
prob\_Chi\_sr\_cheng\_interaction\_model\_v\_main

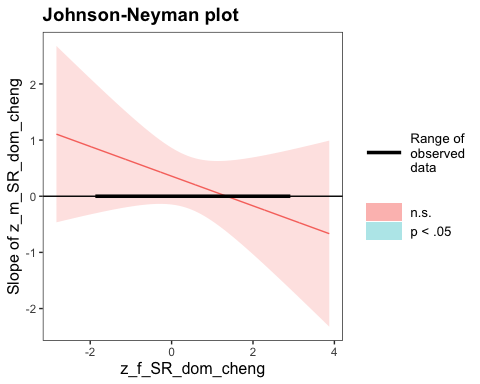
[1] 0.2515567

* As suspected, the Chi-squared test does not indicate that the interaction model fits better than the main effects model (χ²(1) = 1.315, *p* = .252).

Again, for completion’s sake, I will run the simple slopes and Johnson-Neyman analysis for the interaction model.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(sr\_cheng\_interaction\_model, pred = z\_m\_SR\_dom\_cheng, modx = z\_f\_SR\_dom\_cheng, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_SR\_dom\_cheng when z\_f\_SR\_dom\_cheng = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.62 0.41 1.53 0.13  
  
Slope of z\_m\_SR\_dom\_cheng when z\_f\_SR\_dom\_cheng = -1.561251e-17 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.36 0.25 1.40 0.16  
  
Slope of z\_m\_SR\_dom\_cheng when z\_f\_SR\_dom\_cheng = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.09 0.27 0.34 0.73

* As suspected, Johnson-Neyman result indicates that there are no values of fathers’ SR dominance status where mothers’ SR dominance significantly predicts the probability of having a first-born son.

### Hypothesis 2: Facial Dominance Operationalization

Now we will test the hypothesis with the standardized residuals for facial dominance that we created within sexes. Because there are group differences in facial dominance between the neutral and non-neutral facial expression groups, we will conduct this analysis on both the full dyadic dataset and on the neutral faces only.

#### Full Dataset

First we will standardize our independent variables, m\_res\_facial\_dominance and f\_res\_facial\_dominance.

# Standardizing the m\_res\_facial\_dominance and f\_res\_facial\_dominance variables as z\_m\_res\_facial\_dominance and z\_f\_res\_facial\_dominance  
dyadic\_data$z\_m\_res\_facial\_dominance <- scale(dyadic\_data$m\_res\_facial\_dominance, center = TRUE, scale = TRUE)  
dyadic\_data$z\_f\_res\_facial\_dominance <- scale(dyadic\_data$f\_res\_facial\_dominance, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

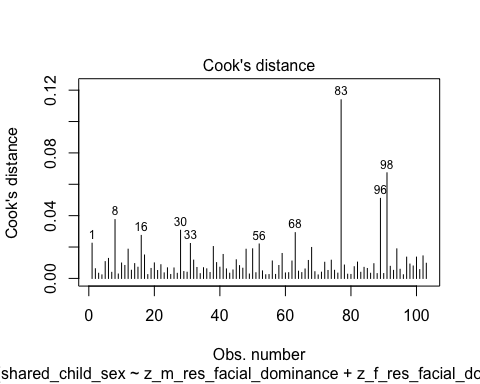
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers  
res\_fac\_dom\_main\_effects\_model <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)  
  
# Fitting the model with the main effects of residual facial dominance for mothers and fathers as well as their interaction  
res\_fac\_dom\_interaction\_model <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance + z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

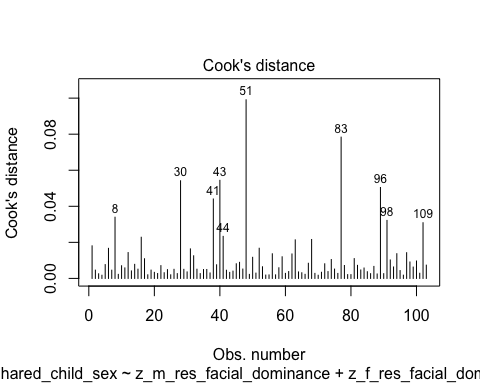
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(res\_fac\_dom\_main\_effects\_model, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(res\_fac\_dom\_interaction\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(res\_fac\_dom\_main\_effects\_model)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 1.014754 1.014754

1/vif(res\_fac\_dom\_main\_effects\_model)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 0.9854607 0.9854607

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. The below code determines the lowest value of z\_m\_res\_facial\_dominance and z\_f\_res\_facial\_dominance so that we can add a constant that makes all values for the variable positive before making the natural log transformation.

min(dyadic\_data$z\_m\_res\_facial\_dominance)

[1] -2.439434

min(dyadic\_data$z\_f\_res\_facial\_dominance, na.rm = TRUE) # Because there is one NA value in the father data we specify that the algorithm should ignore this value

[1] -3.361094

* Given these minimum values, I will add 4 to each variable to make them positive, then I will make a natural log transformation to both.

# Creating c\_z\_m\_res\_facial\_dominance, which represents mothers' residual facial dominance after adding a constant of 4  
dyadic\_data$c\_z\_m\_res\_facial\_dominance <- dyadic\_data$z\_m\_res\_facial\_dominance + 4  
  
# Creating ln\_c\_z\_m\_res\_facial\_dominance, which represents the natural log of mothers' residual facial dominance after adding a constant of 4  
dyadic\_data$ln\_c\_z\_m\_res\_facial\_dominance <- log(dyadic\_data$c\_z\_m\_res\_facial\_dominance)  
  
# Creating c\_z\_f\_res\_facial\_dominance, which represents fathers' residual facial dominance after adding a constant of 4  
dyadic\_data$c\_z\_f\_res\_facial\_dominance <- dyadic\_data$z\_f\_res\_facial\_dominance + 4  
  
# Creating ln\_c\_z\_f\_res\_facial\_dominance, which represents the natural log of fathers' residual facial dominance after adding a constant of 4  
dyadic\_data$ln\_c\_z\_f\_res\_facial\_dominance <- log(dyadic\_data$c\_z\_f\_res\_facial\_dominance)

* Through manually doing the transformations for a couple numbers, I believe that this transformation was done correctly.
* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_res\_facial\_dom\_model <- glm(shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)  
  
summary(BT\_test\_res\_facial\_dom\_model)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error  
(Intercept) -0.2428 6.4165  
c\_z\_m\_res\_facial\_dominance 2.9280 3.0059  
c\_z\_f\_res\_facial\_dominance -3.3793 2.5450  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -1.2074 1.2521  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 1.6497 1.1157  
 z value Pr(>|z|)  
(Intercept) -0.038 0.970  
c\_z\_m\_res\_facial\_dominance 0.974 0.330  
c\_z\_f\_res\_facial\_dominance -1.328 0.184  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -0.964 0.335  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 1.479 0.139  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 130.80 on 98 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 140.8  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(res\_fac\_dom\_main\_effects\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.42678 0.20505 2.081 0.0374 \*  
z\_m\_res\_facial\_dominance 0.02287 0.20489 0.112 0.9111   
z\_f\_res\_facial\_dominance 0.37699 0.21672 1.739 0.0820 .  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 135.15 on 100 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 141.15  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_main\_effects\_model <- res\_fac\_dom\_main\_effects\_model$null.deviance - res\_fac\_dom\_main\_effects\_model$deviance  
Chi\_res\_fac\_dom\_main\_effects\_model

[1] 3.3311

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_main\_effects\_model <- res\_fac\_dom\_main\_effects\_model$df.null - res\_fac\_dom\_main\_effects\_model$df.residual  
df\_res\_fac\_dom\_main\_effects\_model

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model <- 1 - pchisq(Chi\_res\_fac\_dom\_main\_effects\_model, df\_res\_fac\_dom\_main\_effects\_model)  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model

[1] 0.1890866

* Although the whole model is not a better fit than the baseline model with just the intercept (x2(2) = 3.331, *p* = .189), the residual facial dominance predictor for fathers was close to significant (b = .377, *z* = 1.739, *p* = .082), with a one standard deviation increase in father’s residual facial dominance leading to a .377 unit increase in the log-odds of having a first born son. I will exponentiate this coefficient to make it more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_main\_effects\_model$coefficients)

(Intercept) z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 1.532313 1.023137 1.457884

* The odds ratio for father’s residual facial dominance (although not significant) is OR = 1.457, indicating that while holding mother’s residual facial dominance constant a one standard deviation increase in father’s residual facial dominance is associated with a 45.7% higher odds of having a first born son.

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(res\_fac\_dom\_interaction\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance +   
 z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value  
(Intercept) 0.39794 0.20835 1.910  
z\_m\_res\_facial\_dominance 0.09678 0.21849 0.443  
z\_f\_res\_facial\_dominance 0.45865 0.23069 1.988  
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.40692 0.26321 1.546  
 Pr(>|z|)   
(Intercept) 0.0561 .  
z\_m\_res\_facial\_dominance 0.6578   
z\_f\_res\_facial\_dominance 0.0468 \*  
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.1221   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 132.53 on 99 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 140.53  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_interaction\_model <- res\_fac\_dom\_interaction\_model$null.deviance - res\_fac\_dom\_interaction\_model$deviance  
Chi\_res\_fac\_dom\_interaction\_model

[1] 5.944714

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_interaction\_model <- res\_fac\_dom\_interaction\_model$df.null - res\_fac\_dom\_interaction\_model$df.residual  
df\_res\_fac\_dom\_interaction\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_interaction\_model <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model, df\_res\_fac\_dom\_interaction\_model)  
prob\_Chi\_res\_fac\_dom\_interaction\_model

[1] 0.1143312

* Interestingly—although the model still does not fit significantly better than the intercept-only model (x2(3) = 5.945, *p* = .114)—when the interaction between mothers’ and fathers’ residual facial dominance is added to the model fathers’ residual facial dominance becomes significance (*b* = .459, z = 1.988, *p* = .047). Also of note, but not significant, the interaction between mothers’ and fathers’ residual facial dominance has a positive coefficient (*b* = .407, *z* = 1.546, *p* = .122). I will exponentiate these coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_interaction\_model$coefficients)

(Intercept)   
 1.488749   
 z\_m\_res\_facial\_dominance   
 1.101617   
 z\_f\_res\_facial\_dominance   
 1.581937   
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance   
 1.502177

* For father’s residual facial dominance, the OR = 1.582, and for the interaction between mothers’ and fathers’ residual facial dominance the OR = 1.502. This indicates that, with other variables held constant, a one standard deviation unit increase in fathers’ residual facial dominance is associated with a 58.2% increase in the odds of having a first born son, and a one unit increase in the product of mothers’ and fathers’ residual facial dominance is associated with a 50.2% increase in the odds of having a first born son.

Because the whole interaction model is not significant compared to the intercept-only model, it would be very surprising if the interaction model would fit better than the main-effects only model, but I will go ahead and do the comparison anyways.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_res\_fac\_dom\_interaction\_model\_v\_main <- res\_fac\_dom\_main\_effects\_model$deviance - res\_fac\_dom\_interaction\_model$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_v\_main

[1] 2.613613

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_res\_fac\_dom\_interaction\_model\_v\_main <- res\_fac\_dom\_main\_effects\_model$df.residual - res\_fac\_dom\_interaction\_model$df.residual  
df\_res\_fac\_dom\_interaction\_model\_v\_main

[1] 1

prob\_Chi\_res\_fac\_dom\_interaction\_model\_v\_main <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_v\_main, df\_res\_fac\_dom\_interaction\_model\_v\_main)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_v\_main

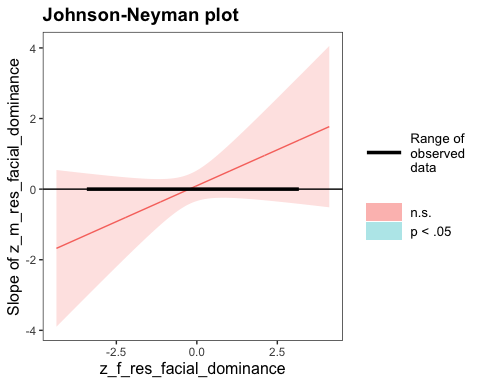
[1] 0.1059501

* As suspected, the Chi-squared test does not indicate that the interaction model fits better than the main effects model (x2(1) = 2.614, *p* = .106).

Although the interaction is not significant in our model, our hypothesis predicts that at high levels of fathers’ facial dominance mothers’ dominance predicts the probability of a first born son, so I will run a simple slopes and Johnson-Neyman analysis with mothers’ residual facial dominance as the focal predictor and fathers’ residual facial dominance as the moderator.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(res\_fac\_dom\_interaction\_model, pred = z\_m\_res\_facial\_dominance, modx = z\_f\_res\_facial\_dominance, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------- ------ -------- ------  
 -0.31 0.30 -1.03 0.30  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = 1.293464e-17 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.10 0.22 0.44 0.66  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.50 0.38 1.33 0.18

* Although nothing here is significant, I will exponentiate the coefficients here to make them more interpretable.

# Producing odds-ratios for each value of the slope for mothers' residual facial dominance for each level of the moderator in the simple slopes analysis  
exp(-.30) # -1 SD

[1] 0.7408182

exp(.09) # Mean

[1] 1.094174

exp(.49) # +1 SD

[1] 1.632316

# Taking the inverse of the odds-ratio for the first coefficient  
1 - 0.7408182

[1] 0.2591818

* The Johnson-Neyman analysis indicates that there are no values of fathers’ residual facial dominance for which mothers’ residual facial dominance is a significant predictor of the probability of having a first born child. Worthy of note, however, the simple slopes analysis indicates that at low levels of fathers’ residual facial dominance (-1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 25.9% *decrease* in the odds of having a first born son and at high levels of fathers’ residual facial dominance (+1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 63.2% *increase* in the odds of having a first born son.
  + Although not significant, the direction of this moderation effect is consistent with our hypothesis.
  + It is possible that in the population there exists such an effect, whereas we simply do not have enough power to detect it.

#### Neutral Faces Only

Now we will repeat the same analysis as just above but with the neutral faces only. To do this, I will create a new data frame that only contains the neutral faces.

# Creating a new dyadic data frame called neutral\_face\_dyadic\_data where only cases where mothers' and fathers' face was neutral in the photograph are included  
neutral\_face\_dyadic\_data <- subset(dyadic\_data, m\_expression\_not\_neutral == "neutral" & f\_expression\_not\_neutral == "neutral")

* I have checked the new data frame and it seems to have been created correctly, resulting in *n* = 87 cases.

Now we will standardize our independent variables in the new data frame, m\_res\_facial\_dominance and f\_res\_facial\_dominance.

# Standardizing the m\_res\_facial\_dominance and f\_res\_facial\_dominance variables as z\_m\_res\_facial\_dominance and z\_f\_res\_facial\_dominance (because the mean and SD may be slightly different in this new data frame)  
neutral\_face\_dyadic\_data$z\_m\_res\_facial\_dominance <- scale(neutral\_face\_dyadic\_data$m\_res\_facial\_dominance, center = TRUE, scale = TRUE)  
neutral\_face\_dyadic\_data$z\_f\_res\_facial\_dominance <- scale(neutral\_face\_dyadic\_data$f\_res\_facial\_dominance, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

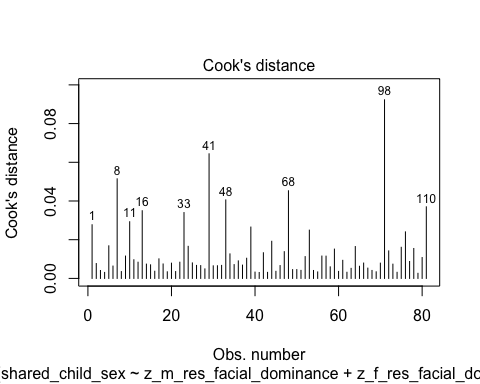
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers  
res\_fac\_dom\_main\_effects\_model\_neutral <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
# Fitting the model with the main effects of residual facial dominance for mothers and fathers as well as their interaction  
res\_fac\_dom\_interaction\_model\_neutral <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance + z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)

##### Assumptions

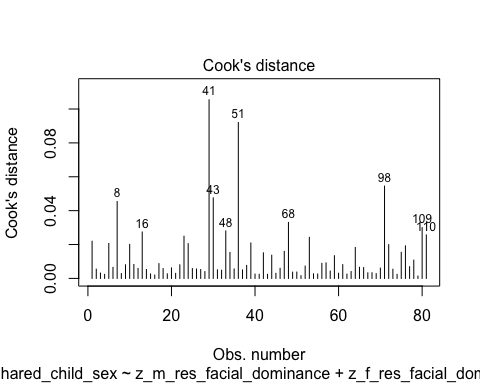
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(res\_fac\_dom\_main\_effects\_model\_neutral, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(res\_fac\_dom\_interaction\_model\_neutral, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(res\_fac\_dom\_main\_effects\_model\_neutral)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 1.000339 1.000339

1/vif(res\_fac\_dom\_main\_effects\_model\_neutral)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 0.9996615 0.9996615

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. The below code determines the lowest value of z\_m\_res\_facial\_dominance and z\_f\_res\_facial\_dominance so that we can add a constant that makes all values for the variable positive before making the natural log transformation.

min(neutral\_face\_dyadic\_data$z\_m\_res\_facial\_dominance)

[1] -1.842716

min(neutral\_face\_dyadic\_data$z\_f\_res\_facial\_dominance)

[1] -2.616872

* Given these minimum values, I will add 3 to each variable to make them positive, then I will make a natural log transformation to both.

# Creating c\_z\_m\_res\_facial\_dominance, which represents mothers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$c\_z\_m\_res\_facial\_dominance <- neutral\_face\_dyadic\_data$z\_m\_res\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_m\_res\_facial\_dominance, which represents the natural log of mothers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$ln\_c\_z\_m\_res\_facial\_dominance <- log(neutral\_face\_dyadic\_data$c\_z\_m\_res\_facial\_dominance)  
  
# Creating c\_z\_f\_res\_facial\_dominance, which represents fathers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$c\_z\_f\_res\_facial\_dominance <- neutral\_face\_dyadic\_data$z\_f\_res\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_f\_res\_facial\_dominance, which represents the natural log of fathers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$ln\_c\_z\_f\_res\_facial\_dominance <- log(neutral\_face\_dyadic\_data$c\_z\_f\_res\_facial\_dominance)

* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_res\_facial\_dom\_model\_neutral <- glm(shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
summary(BT\_test\_res\_facial\_dom\_model\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error  
(Intercept) -2.9930 3.6877  
c\_z\_m\_res\_facial\_dominance 1.3959 2.2373  
c\_z\_f\_res\_facial\_dominance 0.1359 1.9328  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -0.5843 1.0558  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 0.2251 0.9547  
 z value Pr(>|z|)  
(Intercept) -0.812 0.417  
c\_z\_m\_res\_facial\_dominance 0.624 0.533  
c\_z\_f\_res\_facial\_dominance 0.070 0.944  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -0.553 0.580  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 0.236 0.814  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 103.27 on 76 degrees of freedom  
AIC: 113.27  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(res\_fac\_dom\_main\_effects\_model\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.3523 0.2355 1.496 0.1346   
z\_m\_res\_facial\_dominance 0.1596 0.2367 0.674 0.5000   
z\_f\_res\_facial\_dominance 0.6038 0.2668 2.263 0.0236 \*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 103.61 on 78 degrees of freedom  
AIC: 109.61  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_main\_effects\_model\_neutral <- res\_fac\_dom\_main\_effects\_model\_neutral$null.deviance - res\_fac\_dom\_main\_effects\_model\_neutral$deviance  
Chi\_res\_fac\_dom\_main\_effects\_model\_neutral

[1] 6.589181

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_main\_effects\_model\_neutral <- res\_fac\_dom\_main\_effects\_model\_neutral$df.null - res\_fac\_dom\_main\_effects\_model\_neutral$df.residual  
df\_res\_fac\_dom\_main\_effects\_model\_neutral

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model\_neutral <- 1 - pchisq(Chi\_res\_fac\_dom\_main\_effects\_model\_neutral, df\_res\_fac\_dom\_main\_effects\_model\_neutral)  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model\_neutral

[1] 0.03708323

* Despite the loss of power, the full main effects model is significant (x2(2) = 6.589, *p* = .037). The predictor for fathers’ residual facial dominance is also significant (*b* = .604, *z* = 2.263, *p* = .024). I will exponentiate the coefficients to make this more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_main\_effects\_model\_neutral$coefficients)

(Intercept) z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 1.422374 1.173076 1.828981

* The odds ratio for father’s residual facial dominance is OR = 1.829, indicating that while holding mother’s residual facial dominance constant a one standard deviation increase in father’s residual facial dominance is associated with a 83% higher odds of having a first born son.

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(res\_fac\_dom\_interaction\_model\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance +   
 z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value  
(Intercept) 0.3620 0.2395 1.511  
z\_m\_res\_facial\_dominance 0.2463 0.2549 0.966  
z\_f\_res\_facial\_dominance 0.7024 0.2879 2.440  
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.3274 0.2669 1.227  
 Pr(>|z|)   
(Intercept) 0.1307   
z\_m\_res\_facial\_dominance 0.3338   
z\_f\_res\_facial\_dominance 0.0147 \*  
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.2199   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 102.00 on 77 degrees of freedom  
AIC: 110  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_interaction\_model\_neutral <- res\_fac\_dom\_interaction\_model\_neutral$null.deviance - res\_fac\_dom\_interaction\_model\_neutral$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_neutral

[1] 8.189911

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_interaction\_model\_neutral <- res\_fac\_dom\_interaction\_model\_neutral$df.null - res\_fac\_dom\_interaction\_model\_neutral$df.residual  
df\_res\_fac\_dom\_interaction\_model\_neutral

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_neutral <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_neutral, df\_res\_fac\_dom\_interaction\_model\_neutral)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_neutral

[1] 0.04224562

* The full interaction model was significant (x2(3) = 8.19, *p* = .042), and although the interaction term is not significant, when the interaction term was added to the model the coefficient became larger for fathers’ residual facial dominance (*b* = .702, *z* = 2.440, *p* = .015). Again, I will exponentiate this coefficient to understand it better.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_interaction\_model\_neutral$coefficients)

(Intercept)   
 1.436188   
 z\_m\_res\_facial\_dominance   
 1.279336   
 z\_f\_res\_facial\_dominance   
 2.018679   
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance   
 1.387320

* The odds ratio for father’s residual facial dominance is now 2.02, indicating that while holding mother’s residual facial dominance constant a one standard deviation increase in father’s residual facial dominance is associated with a 102% higher odds of having a first born son.

Now to compare the main-effects model with the main-effects and interaction model.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral <- res\_fac\_dom\_main\_effects\_model\_neutral$deviance - res\_fac\_dom\_interaction\_model\_neutral$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral

[1] 1.60073

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral <- res\_fac\_dom\_main\_effects\_model\_neutral$df.residual - res\_fac\_dom\_interaction\_model\_neutral$df.residual  
df\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral

[1] 1

prob\_Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral, df\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral

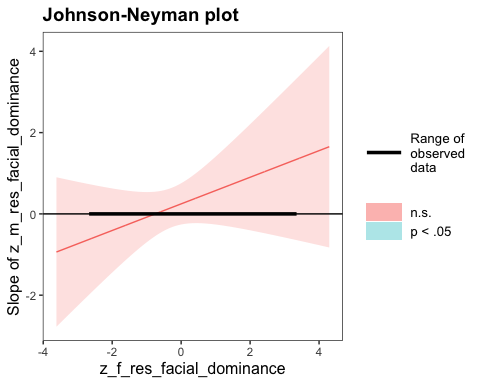
[1] 0.2057997

* The Chi-squared test does not indicate that the interaction model fits better than the main effects model (x2(1) = 1.601, *p* = .206).

Although the interaction is not significant in our model, our hypothesis predicts that at high levels of fathers’ facial dominance mothers’ dominance predicts the probability of a first born son, so I will run a simple slopes and Johnson-Neyman analysis with mothers’ residual facial dominance as the focal predictor and fathers’ residual facial dominance as the moderator.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(res\_fac\_dom\_interaction\_model\_neutral, pred = z\_m\_res\_facial\_dominance, modx = z\_f\_res\_facial\_dominance, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------- ------ -------- ------  
 -0.08 0.31 -0.26 0.79  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = -2.193033e-17 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.25 0.25 0.97 0.33  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.57 0.42 1.37 0.17

* Although nothing here is significant, I will exponentiate the coefficients here to make them more interpretable.

# Producing odds-ratios for each value of the slope for mothers' residual facial dominance for each level of the moderator in the simple slopes analysis  
exp(-.08) # -1 SD

[1] 0.9231163

exp(.24) # Mean

[1] 1.271249

exp(.56) # +1 SD

[1] 1.750673

# Taking the inverse of the odds-ratio for the first coefficient  
1 - 0.9231163

[1] 0.0768837

* The Johnson-Neyman analysis indicates that there are no values of fathers’ residual facial dominance for which mothers’ residual facial dominance is a significant predictor of the probability of having a first born child. Similar to the analysis with all facial expression types, however, the simple slopes analysis indicates that at low levels of fathers’ residual facial dominance (-1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 7.7% *decrease* in the odds of having a first born son; at the mean for fathers’ residual facial dominance a one standard deviation increase in mothers’ residual facial dominance is associated with an 27.1% *increase* in the odds of having a first born son; and at high levels of fathers’ residual facial dominance (+1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 75.1% *increase* in the odds of having a first born son.
  + Although not significant, the direction of this moderation effect is consistent with our hypothesis, like in the model with all facial expressions included.
  + Again, it is possible that in the population there exists such an effect, whereas we simply do not have enough power to detect it.

## Testing Hypothesis 3

To test hypothesis three, we will (1) look at whether fathers’ fWHR is significantly correlated with their residual facial dominance (and I will also include the other facial rating characteristics) either in the full dyadic dataset or in the neutral facial expression dyadic dataset, then, if so, (2) test whether fathers’ fWHR significantly differs by offspring sex, and (3) substitute fathers’ fWHR for residual facial dominance in the same binary logistic regression models as the previous section. In these latter two analyses, we should expect to see that as fathers’ fWHR increases the probability of having a first-born son also increases.

### Correlating fWHR with Facial Dominance

Here are the bivariate correlations between fathers’ fWHR, ratings of attractiveness, ratings of masculinity/femininity, ratings of dominance, and residual facial dominance.

# Calculate pairwise correlations with p-values and confidence intervals for both datasets  
corr4 <- corr.test(dyadic\_data[, c("f\_fWHR", "f\_res\_facial\_dominance", "f\_facial\_masculinityfemininity", "f\_facial\_dominance", "f\_facial\_attractiveness")], use="pairwise.complete.obs")  
# Calculate pairwise correlations with p-values and confidence intervals for both datasets  
corr5 <- corr.test(dyadic\_data[, c("m\_fWHR", "m\_res\_facial\_dominance", "m\_facial\_masculinityfemininity", "m\_facial\_dominance", "m\_facial\_attractiveness")], use="pairwise.complete.obs")  
  
print(corr4, short=FALSE)

Call:corr.test(x = dyadic\_data[, c("f\_fWHR", "f\_res\_facial\_dominance",   
 "f\_facial\_masculinityfemininity", "f\_facial\_dominance", "f\_facial\_attractiveness")],   
 use = "pairwise.complete.obs")  
Correlation matrix   
 f\_fWHR f\_res\_facial\_dominance  
f\_fWHR 1.00 0.21  
f\_res\_facial\_dominance 0.21 1.00  
f\_facial\_masculinityfemininity 0.11 -0.01  
f\_facial\_dominance 0.21 0.61  
f\_facial\_attractiveness 0.12 0.00  
 f\_facial\_masculinityfemininity  
f\_fWHR 0.11  
f\_res\_facial\_dominance -0.01  
f\_facial\_masculinityfemininity 1.00  
f\_facial\_dominance 0.77  
f\_facial\_attractiveness 0.25  
 f\_facial\_dominance f\_facial\_attractiveness  
f\_fWHR 0.21 0.12  
f\_res\_facial\_dominance 0.61 0.00  
f\_facial\_masculinityfemininity 0.77 0.25  
f\_facial\_dominance 1.00 0.35  
f\_facial\_attractiveness 0.35 1.00  
Sample Size   
 f\_fWHR f\_res\_facial\_dominance  
f\_fWHR 93 93  
f\_res\_facial\_dominance 93 103  
f\_facial\_masculinityfemininity 93 103  
f\_facial\_dominance 93 103  
f\_facial\_attractiveness 93 103  
 f\_facial\_masculinityfemininity  
f\_fWHR 93  
f\_res\_facial\_dominance 103  
f\_facial\_masculinityfemininity 103  
f\_facial\_dominance 103  
f\_facial\_attractiveness 103  
 f\_facial\_dominance f\_facial\_attractiveness  
f\_fWHR 93 93  
f\_res\_facial\_dominance 103 103  
f\_facial\_masculinityfemininity 103 103  
f\_facial\_dominance 103 103  
f\_facial\_attractiveness 103 103  
Probability values (Entries above the diagonal are adjusted for multiple tests.)   
 f\_fWHR f\_res\_facial\_dominance  
f\_fWHR 0.00 0.24  
f\_res\_facial\_dominance 0.04 0.00  
f\_facial\_masculinityfemininity 0.31 0.94  
f\_facial\_dominance 0.04 0.00  
f\_facial\_attractiveness 0.26 0.96  
 f\_facial\_masculinityfemininity  
f\_fWHR 1.00  
f\_res\_facial\_dominance 1.00  
f\_facial\_masculinityfemininity 0.00  
f\_facial\_dominance 0.00  
f\_facial\_attractiveness 0.01  
 f\_facial\_dominance f\_facial\_attractiveness  
f\_fWHR 0.24 1.00  
f\_res\_facial\_dominance 0.00 1.00  
f\_facial\_masculinityfemininity 0.00 0.08  
f\_facial\_dominance 0.00 0.00  
f\_facial\_attractiveness 0.00 0.00  
  
 Confidence intervals based upon normal theory. To get bootstrapped values, try cor.ci  
 raw.lower raw.r raw.upper raw.p lower.adj upper.adj  
f\_WHR-f\_r\_\_ 0.01 0.21 0.40 0.04 -0.06 0.46  
f\_WHR-f\_fcl\_m -0.10 0.11 0.30 0.31 -0.14 0.35  
f\_WHR-f\_fcl\_d 0.01 0.21 0.40 0.04 -0.06 0.45  
f\_WHR-f\_fcl\_t -0.09 0.12 0.31 0.26 -0.14 0.36  
f\_r\_\_-f\_fcl\_m -0.20 -0.01 0.19 0.94 -0.23 0.21  
f\_r\_\_-f\_fcl\_d 0.48 0.61 0.72 0.00 0.41 0.76  
f\_r\_\_-f\_fcl\_t -0.20 0.00 0.19 0.96 -0.20 0.19  
f\_fcl\_m-f\_fcl\_d 0.67 0.77 0.84 0.00 0.62 0.86  
f\_fcl\_m-f\_fcl\_t 0.06 0.25 0.42 0.01 -0.02 0.48  
f\_fcl\_d-f\_fcl\_t 0.17 0.35 0.51 0.00 0.09 0.56

print(corr5, short=FALSE)

Call:corr.test(x = dyadic\_data[, c("m\_fWHR", "m\_res\_facial\_dominance",   
 "m\_facial\_masculinityfemininity", "m\_facial\_dominance", "m\_facial\_attractiveness")],   
 use = "pairwise.complete.obs")  
Correlation matrix   
 m\_fWHR m\_res\_facial\_dominance  
m\_fWHR 1.00 0.09  
m\_res\_facial\_dominance 0.09 1.00  
m\_facial\_masculinityfemininity 0.11 0.00  
m\_facial\_dominance 0.22 0.88  
m\_facial\_attractiveness 0.15 0.00  
 m\_facial\_masculinityfemininity  
m\_fWHR 0.11  
m\_res\_facial\_dominance 0.00  
m\_facial\_masculinityfemininity 1.00  
m\_facial\_dominance 0.00  
m\_facial\_attractiveness -0.75  
 m\_facial\_dominance m\_facial\_attractiveness  
m\_fWHR 0.22 0.15  
m\_res\_facial\_dominance 0.88 0.00  
m\_facial\_masculinityfemininity 0.00 -0.75  
m\_facial\_dominance 1.00 0.30  
m\_facial\_attractiveness 0.30 1.00  
Sample Size   
 m\_fWHR m\_res\_facial\_dominance  
m\_fWHR 90 90  
m\_res\_facial\_dominance 90 104  
m\_facial\_masculinityfemininity 90 104  
m\_facial\_dominance 90 104  
m\_facial\_attractiveness 90 104  
 m\_facial\_masculinityfemininity  
m\_fWHR 90  
m\_res\_facial\_dominance 104  
m\_facial\_masculinityfemininity 104  
m\_facial\_dominance 104  
m\_facial\_attractiveness 104  
 m\_facial\_dominance m\_facial\_attractiveness  
m\_fWHR 90 90  
m\_res\_facial\_dominance 104 104  
m\_facial\_masculinityfemininity 104 104  
m\_facial\_dominance 104 104  
m\_facial\_attractiveness 104 104  
Probability values (Entries above the diagonal are adjusted for multiple tests.)   
 m\_fWHR m\_res\_facial\_dominance  
m\_fWHR 0.00 1.00  
m\_res\_facial\_dominance 0.41 0.00  
m\_facial\_masculinityfemininity 0.30 0.98  
m\_facial\_dominance 0.04 0.00  
m\_facial\_attractiveness 0.16 0.98  
 m\_facial\_masculinityfemininity  
m\_fWHR 1  
m\_res\_facial\_dominance 1  
m\_facial\_masculinityfemininity 0  
m\_facial\_dominance 1  
m\_facial\_attractiveness 0  
 m\_facial\_dominance m\_facial\_attractiveness  
m\_fWHR 0.28 0.94  
m\_res\_facial\_dominance 0.00 1.00  
m\_facial\_masculinityfemininity 1.00 0.00  
m\_facial\_dominance 0.00 0.02  
m\_facial\_attractiveness 0.00 0.00  
  
 Confidence intervals based upon normal theory. To get bootstrapped values, try cor.ci  
 raw.lower raw.r raw.upper raw.p lower.adj upper.adj  
m\_WHR-m\_r\_\_ -0.12 0.09 0.29 0.41 -0.18 0.34  
m\_WHR-m\_fcl\_m -0.10 0.11 0.31 0.30 -0.16 0.37  
m\_WHR-m\_fcl\_d 0.01 0.22 0.41 0.04 -0.07 0.47  
m\_WHR-m\_fcl\_t -0.06 0.15 0.35 0.16 -0.13 0.41  
m\_r\_\_-m\_fcl\_m -0.19 0.00 0.19 0.98 -0.22 0.22  
m\_r\_\_-m\_fcl\_d 0.83 0.88 0.92 0.00 0.80 0.93  
m\_r\_\_-m\_fcl\_t -0.19 0.00 0.19 0.98 -0.23 0.24  
m\_fcl\_m-m\_fcl\_d -0.19 0.00 0.19 1.00 -0.19 0.19  
m\_fcl\_m-m\_fcl\_t -0.83 -0.75 -0.66 0.00 -0.85 -0.61  
m\_fcl\_d-m\_fcl\_t 0.11 0.30 0.46 0.00 0.03 0.52

### Substituting Fathers’ fHWR for Fathers’ Facial Dominance in Previous Models

To test Hypothesis 3 more rigorously, we will replace fathers’ facial dominance in the previous binary logistic regression models predicting shared child sex with fathers’ fWHR.

#### Full Dataset

First we will standardize our new independent variablee, f\_fWHR.

# Standardizing the f\_fWHR variable as z\_f\_fWHR  
dyadic\_data$z\_f\_fWHR <- scale(dyadic\_data$f\_fWHR, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

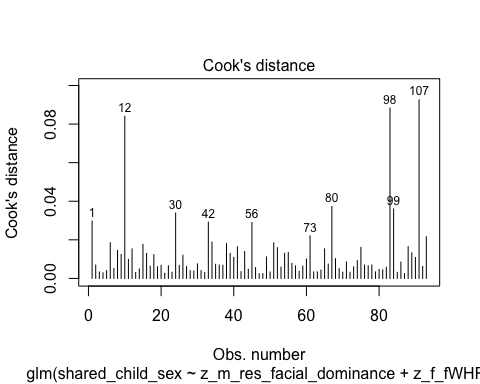
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers' fWHR  
hyp\_3\_main\_all <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR, family = binomial(link = logit), data = dyadic\_data)  
  
# Fitting the model with the main effects of residual facial dominance for mothers and fathers' fWHR as well as their interaction  
hyp\_3\_interaction\_all <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR + z\_m\_res\_facial\_dominance:z\_f\_fWHR, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

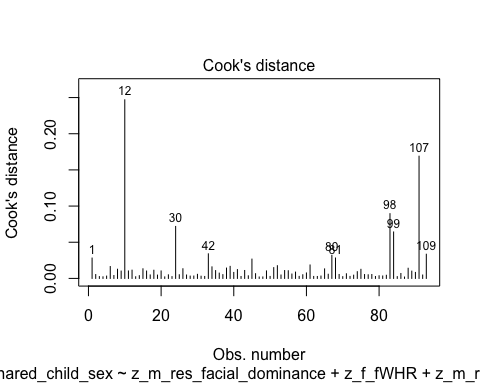
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(hyp\_3\_main\_all, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(hyp\_3\_interaction\_all, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(hyp\_3\_main\_all)

z\_m\_res\_facial\_dominance z\_f\_fWHR   
 1.021395 1.021395

1/vif(hyp\_3\_main\_all)

z\_m\_res\_facial\_dominance z\_f\_fWHR   
 0.9790534 0.9790534

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. We will add 10 to make z\_f\_fWHR all positive before fitting the Box-Tidwell model.

# Creating c\_z\_f\_fWHR, which represents standardized fathers' fWHR after adding a constant of 10  
dyadic\_data$c\_z\_f\_fWHR <- dyadic\_data$z\_f\_fWHR + 10  
  
# Creating ln\_c\_z\_m\_res\_facial\_dominance, which represents the natural log of standardized fathers' fWHR after adding a constant of 10  
dyadic\_data$ln\_c\_z\_f\_fWHR <- log(dyadic\_data$c\_z\_f\_fWHR)

* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_hyp\_3\_main\_all <- glm(shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance + c\_z\_f\_fWHR + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance + c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit), data = dyadic\_data)  
  
summary(BT\_test\_hyp\_3\_main\_all)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_fWHR + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error  
(Intercept) -25.770 36.592  
c\_z\_m\_res\_facial\_dominance 4.725 3.249  
c\_z\_f\_fWHR 5.368 11.919  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -1.909 1.350  
c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR -1.539 3.583  
 z value Pr(>|z|)  
(Intercept) -0.704 0.481  
c\_z\_m\_res\_facial\_dominance 1.454 0.146  
c\_z\_f\_fWHR 0.450 0.652  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -1.414 0.157  
c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR -0.429 0.668  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 124.14 on 92 degrees of freedom  
Residual deviance: 119.44 on 88 degrees of freedom  
 (11 observations deleted due to missingness)  
AIC: 129.44  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ fWHR only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(hyp\_3\_main\_all)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4668 0.2160 2.162 0.0307 \*  
z\_m\_res\_facial\_dominance 0.1432 0.2118 0.676 0.4989   
z\_f\_fWHR 0.3086 0.2272 1.358 0.1744   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 124.14 on 92 degrees of freedom  
Residual deviance: 121.97 on 90 degrees of freedom  
 (11 observations deleted due to missingness)  
AIC: 127.97  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_main\_all <- hyp\_3\_main\_all$null.deviance - hyp\_3\_main\_all$deviance  
Chi\_hyp\_3\_main\_all

[1] 2.175194

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_main\_all <- hyp\_3\_main\_all$df.null - hyp\_3\_main\_all$df.residual  
df\_hyp\_3\_main\_all

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_main\_all <- 1 - pchisq(Chi\_hyp\_3\_main\_all, df\_hyp\_3\_main\_all)  
prob\_Chi\_hyp\_3\_main\_all

[1] 0.3370255

* The model is not a better fit than the baseline model with just the intercept (x2(2) = 2.175, *p* = .337), and the fathers’ fWHR predictor is also not significant (*b* = .309, *z* = 1.358, *p* = .174). Still, I will exponentiate the coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_3\_main\_all$coefficients)

(Intercept) z\_m\_res\_facial\_dominance z\_f\_fWHR   
 1.594920 1.154010 1.361524

* The odds ratio for father’s residual facial dominance (although not significant) is OR = 1.362, indicating that while holding mother’s residual facial dominance constant a one standard deviation increase in father’s residual facial dominance is associated with a 36.2% higher odds of having a first born son.

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(hyp\_3\_interaction\_all)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR +   
 z\_m\_res\_facial\_dominance:z\_f\_fWHR, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.48170 0.21971 2.192 0.0283 \*  
z\_m\_res\_facial\_dominance 0.17475 0.22623 0.772 0.4399   
z\_f\_fWHR 0.31392 0.22902 1.371 0.1705   
z\_m\_res\_facial\_dominance:z\_f\_fWHR 0.09505 0.22659 0.419 0.6749   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 124.14 on 92 degrees of freedom  
Residual deviance: 121.79 on 89 degrees of freedom  
 (11 observations deleted due to missingness)  
AIC: 129.79  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_interaction\_all <- hyp\_3\_interaction\_all$null.deviance - hyp\_3\_interaction\_all$deviance  
Chi\_hyp\_3\_interaction\_all

[1] 2.353127

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_interaction\_all <- hyp\_3\_interaction\_all$df.null - hyp\_3\_interaction\_all$df.residual  
df\_hyp\_3\_interaction\_all

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_interaction\_all <- 1 - pchisq(Chi\_hyp\_3\_interaction\_all, df\_hyp\_3\_interaction\_all)  
prob\_Chi\_hyp\_3\_interaction\_all

[1] 0.5024196

* The model still does not fit significantly better than the intercept-only model (x2(3) = 2.353, *p* = .502) when the interaction between mothers’ residual facial dominance and fathers’ fWHR is added to the model. Similarly, fathers’ fWHR remains not significant (*b* = .314, z = 1.371, *p* = .171). I will exponentiate these coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_3\_interaction\_all$coefficients)

(Intercept) z\_m\_res\_facial\_dominance   
 1.618827 1.190944   
 z\_f\_fWHR z\_m\_res\_facial\_dominance:z\_f\_fWHR   
 1.368778 1.099714

* For father’s fWHR, the odds-ratio is virtually the same as in the main effects model

Because the whole interaction model is not significant compared to the intercept-only model, it would be very surprising if the interaction model would fit better than the main-effects only model, but I will go ahead and do the comparison anyways.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_hyp\_3\_interaction\_v\_main\_all <- hyp\_3\_main\_all$deviance - hyp\_3\_interaction\_all$deviance  
Chi\_hyp\_3\_interaction\_v\_main\_all

[1] 0.1779338

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_hyp\_3\_interaction\_v\_main\_all <- hyp\_3\_main\_all$df.residual - hyp\_3\_interaction\_all$df.residual  
df\_hyp\_3\_interaction\_v\_main\_all

[1] 1

prob\_Chi\_hyp\_3\_interaction\_v\_main\_all <- 1 - pchisq(Chi\_hyp\_3\_interaction\_v\_main\_all, df\_hyp\_3\_interaction\_v\_main\_all)  
prob\_Chi\_hyp\_3\_interaction\_v\_main\_all

[1] 0.6731549

* As suspected, the Chi-squared test does not indicate that the interaction model fits better than the main effects model (x2(1) = 0.178, *p* = .673).

#### Neutral Faces Only

Now we will repeat the same analysis as just above but with the neutral faces only.

Now we will standardize our new independent variables in the neutral faces data frame, f\_fWHR.

# Standardizing the f\_fWHR variable as z\_f\_fWHR  
neutral\_face\_dyadic\_data$z\_f\_fWHR <- scale(neutral\_face\_dyadic\_data$f\_fWHR, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

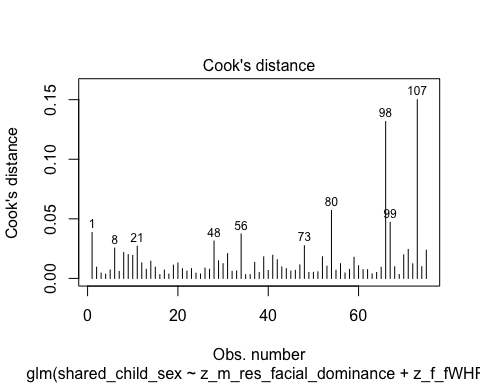
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers fWHR  
hyp\_3\_main\_neutral <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers fWHR plus their interaction  
hyp\_3\_interaction\_neutral <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR + z\_m\_res\_facial\_dominance:z\_f\_fWHR, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)

##### Assumptions

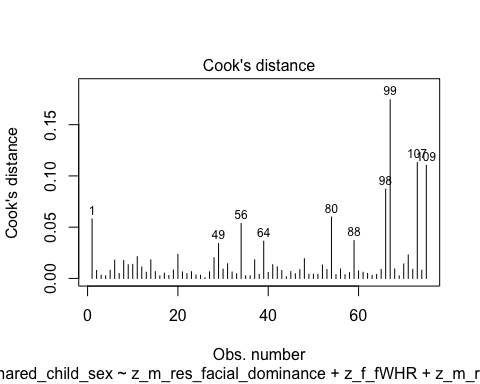
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(hyp\_3\_main\_neutral, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(hyp\_3\_interaction\_neutral, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(hyp\_3\_main\_neutral)

z\_m\_res\_facial\_dominance z\_f\_fWHR   
 1.013037 1.013037

1/vif(hyp\_3\_main\_neutral)

z\_m\_res\_facial\_dominance z\_f\_fWHR   
 0.9871307 0.9871307

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. I will add 10 to the f\_fWHR variable to make it positive, then I will make a natural log transformation to it.

# Creating c\_z\_f\_fWHR, which represents standardized fathers' fWHR after adding a constant of 10  
neutral\_face\_dyadic\_data$c\_z\_f\_fWHR <- neutral\_face\_dyadic\_data$z\_f\_fWHR + 10  
  
# Creating ln\_c\_z\_f\_fWHR, which represents the natural log of standardized fathers' fWHR after adding a constant of 10  
neutral\_face\_dyadic\_data$ln\_c\_z\_f\_fWHR <- log(neutral\_face\_dyadic\_data$c\_z\_f\_fWHR)

* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_hyp\_3\_main\_neutral <- glm(shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance + c\_z\_f\_fWHR + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance + c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
summary(BT\_test\_hyp\_3\_main\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_fWHR + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error  
(Intercept) -20.054 37.250  
c\_z\_m\_res\_facial\_dominance 2.360 2.185  
c\_z\_f\_fWHR 4.890 12.072  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -1.016 1.034  
c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR -1.386 3.625  
 z value Pr(>|z|)  
(Intercept) -0.538 0.590  
c\_z\_m\_res\_facial\_dominance 1.080 0.280  
c\_z\_f\_fWHR 0.405 0.685  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -0.982 0.326  
c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR -0.382 0.702  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 100.952 on 74 degrees of freedom  
Residual deviance: 97.606 on 70 degrees of freedom  
 (6 observations deleted due to missingness)  
AIC: 107.61  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Model

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(hyp\_3\_main\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR,   
 family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4103 0.2397 1.712 0.087 .  
z\_m\_res\_facial\_dominance 0.2245 0.2394 0.938 0.348   
z\_f\_fWHR 0.3056 0.2529 1.208 0.227   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 100.952 on 74 degrees of freedom  
Residual deviance: 98.734 on 72 degrees of freedom  
 (6 observations deleted due to missingness)  
AIC: 104.73  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_main\_neutral <- hyp\_3\_main\_neutral$null.deviance - hyp\_3\_main\_neutral$deviance  
Chi\_hyp\_3\_main\_neutral

[1] 2.217739

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_main\_neutral <- hyp\_3\_main\_neutral$df.null - hyp\_3\_main\_neutral$df.residual  
df\_hyp\_3\_main\_neutral

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_main\_neutral <- 1 - pchisq(Chi\_hyp\_3\_main\_neutral, df\_hyp\_3\_main\_neutral)  
prob\_Chi\_hyp\_3\_main\_neutral

[1] 0.3299318

* The full main effects model is not significant (x2(2) = 2.218, *p* = .330). The predictor for fathers’ fWHR is also not significant (*b* = .305, *z* = 1.208, *p* = .227). I will exponentiate the coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_3\_main\_neutral$coefficients)

(Intercept) z\_m\_res\_facial\_dominance z\_f\_fWHR   
 1.507319 1.251664 1.357377

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(hyp\_3\_interaction\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_fWHR +   
 z\_m\_res\_facial\_dominance:z\_f\_fWHR, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4891 0.2522 1.940 0.0524 .  
z\_m\_res\_facial\_dominance 0.3745 0.2681 1.397 0.1625   
z\_f\_fWHR 0.3551 0.2772 1.281 0.2001   
z\_m\_res\_facial\_dominance:z\_f\_fWHR 0.4956 0.3086 1.606 0.1083   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 100.952 on 74 degrees of freedom  
Residual deviance: 95.755 on 71 degrees of freedom  
 (6 observations deleted due to missingness)  
AIC: 103.76  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_interaction\_neutral <- hyp\_3\_interaction\_neutral$null.deviance - hyp\_3\_interaction\_neutral$deviance  
Chi\_hyp\_3\_interaction\_neutral

[1] 5.196411

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_interaction\_neutral <- hyp\_3\_interaction\_neutral$df.null - hyp\_3\_interaction\_neutral$df.residual  
df\_hyp\_3\_interaction\_neutral

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_interaction\_neutral <- 1 - pchisq(Chi\_hyp\_3\_interaction\_neutral, df\_hyp\_3\_interaction\_neutral)  
prob\_Chi\_hyp\_3\_interaction\_neutral

[1] 0.1579672

* The full interaction model was not significant (x2(3) = 5.196, *p* = .157), and neither is the predictor for fathers’ fWHR (*b* = .355, *z* = 1.281, *p* = .20). Again, I will exponentiate these coefficients.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_3\_interaction\_neutral$coefficients)

(Intercept) z\_m\_res\_facial\_dominance   
 1.630824 1.454197   
 z\_f\_fWHR z\_m\_res\_facial\_dominance:z\_f\_fWHR   
 1.426288 1.641555

Now to compare the main-effects model with the main-effects and interaction model.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral <- hyp\_3\_main\_neutral$deviance - hyp\_3\_interaction\_neutral$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral

[1] 2.978672

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral <- hyp\_3\_main\_neutral$df.residual - hyp\_3\_interaction\_neutral$df.residual  
df\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral

[1] 1

prob\_Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral, df\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_v\_main\_neutral

[1] 0.08436846

* The Chi-squared test, although close, indicates that the interaction model does not fit significantly better than the main effects model (x2(1) = 2.979, *p* = .084).

### Predicting Offspring Sex from Mother and Father fWHR:

#### Full Dataset

First we will standardize our new independent variables, m\_fWHR and f\_fWHR.

# Standardizing the f\_fWHR variable as z\_f\_fWHR and m\_fWHR as z\_m\_fWHR  
dyadic\_data$z\_f\_fWHR <- scale(dyadic\_data$f\_fWHR, center = TRUE, scale = TRUE)  
dyadic\_data$z\_m\_fWHR <- scale(dyadic\_data$m\_fWHR, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

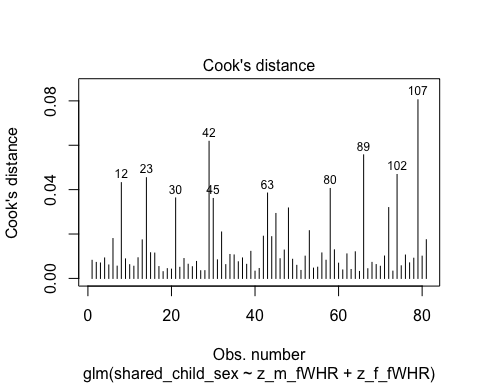
# Fitting the model with only the main effects of mothers' fWHR and fathers' fWHR  
hyp\_3\_pred\_OSR\_all <- glm(shared\_child\_sex ~ z\_m\_fWHR + z\_f\_fWHR, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the model  
plot(hyp\_3\_pred\_OSR\_all, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(hyp\_3\_pred\_OSR\_all)

z\_m\_fWHR z\_f\_fWHR   
1.031947 1.031947

1/vif(hyp\_3\_pred\_OSR\_all)

z\_m\_fWHR z\_f\_fWHR   
0.9690425 0.9690425

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. We will add 10 to make z\_f\_fWHR all positive before fitting the Box-Tidwell model.

# Creating c\_z\_f\_fWHR, which represents standardized fathers' fWHR after adding a constant of 10  
dyadic\_data$c\_z\_f\_fWHR <- dyadic\_data$z\_f\_fWHR + 10  
# Creating c\_z\_m\_fWHR, which represents standardized mothers' fWHR after adding a constant of 10  
dyadic\_data$c\_z\_m\_fWHR <- dyadic\_data$z\_m\_fWHR + 10  
  
# Creating ln\_c\_z\_f\_fWHR, which represents the natural log of standardized fathers' fWHR after adding a constant of 10  
dyadic\_data$ln\_c\_z\_f\_fWHR <- log(dyadic\_data$c\_z\_f\_fWHR)  
# Creating ln\_c\_z\_m\_fWHR, which represents the natural log of standardized fathers' fWHR after adding a constant of 10  
dyadic\_data$ln\_c\_z\_m\_fWHR <- log(dyadic\_data$c\_z\_m\_fWHR)

* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_hyp\_3\_pred\_OSR\_all <- glm(shared\_child\_sex ~ c\_z\_m\_fWHR + c\_z\_f\_fWHR + c\_z\_m\_fWHR:ln\_c\_z\_m\_fWHR + c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit), data = dyadic\_data)  
  
summary(BT\_test\_hyp\_3\_pred\_OSR\_all)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_fWHR + c\_z\_f\_fWHR + c\_z\_m\_fWHR:ln\_c\_z\_m\_fWHR +   
 c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) -16.3126 50.7747 -0.321 0.748  
c\_z\_m\_fWHR 1.7528 11.6365 0.151 0.880  
c\_z\_f\_fWHR 3.6840 12.1441 0.303 0.762  
c\_z\_m\_fWHR:ln\_c\_z\_m\_fWHR -0.5963 3.4874 -0.171 0.864  
c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR -1.0348 3.6543 -0.283 0.777  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 109.5 on 80 degrees of freedom  
Residual deviance: 107.4 on 76 degrees of freedom  
 (23 observations deleted due to missingness)  
AIC: 117.4  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ fWHR only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(hyp\_3\_pred\_OSR\_all)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_fWHR + z\_f\_fWHR, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4184 0.2320 1.804 0.0712 .  
z\_m\_fWHR -0.2458 0.2370 -1.037 0.2996   
z\_f\_fWHR 0.2502 0.2307 1.085 0.2780   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 109.50 on 80 degrees of freedom  
Residual deviance: 107.52 on 78 degrees of freedom  
 (23 observations deleted due to missingness)  
AIC: 113.52  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_pred\_OSR\_all <- hyp\_3\_pred\_OSR\_all$null.deviance - hyp\_3\_pred\_OSR\_all$deviance  
Chi\_hyp\_3\_pred\_OSR\_all

[1] 1.979839

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_pred\_OSR\_all <- hyp\_3\_pred\_OSR\_all$df.null - hyp\_3\_pred\_OSR\_all$df.residual  
df\_hyp\_3\_pred\_OSR\_all

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_pred\_OSR\_all <- 1 - pchisq(Chi\_hyp\_3\_pred\_OSR\_all, df\_hyp\_3\_pred\_OSR\_all)  
prob\_Chi\_hyp\_3\_pred\_OSR\_all

[1] 0.3716066

* The model is not a better fit than the baseline model with just the intercept (x2(2) = 1.98, *p* = .372), and neither of the predictors are significant. Still, I will exponentiate the coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_3\_pred\_OSR\_all$coefficients)

(Intercept) z\_m\_fWHR z\_f\_fWHR   
 1.5195626 0.7820783 1.2843112

##### Fathers’ fWHR Only

Now we will look at whether fathers’ fWHR predicts offspring sex without controlling for mothers’ facial dominance.

* Fitting the model:

hyp\_3\_f\_pred\_OSR\_all <- glm(shared\_child\_sex ~ z\_f\_fWHR, family = binomial(link = logit), data = dyadic\_data)

* Summarizing the model and testing it against the intercept-only model:

# Producing the summary of the model  
summary(hyp\_3\_f\_pred\_OSR\_all)

Call:  
glm(formula = shared\_child\_sex ~ z\_f\_fWHR, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4695 0.2154 2.179 0.0293 \*  
z\_f\_fWHR 0.2873 0.2238 1.284 0.1992   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 124.14 on 92 degrees of freedom  
Residual deviance: 122.43 on 91 degrees of freedom  
 (11 observations deleted due to missingness)  
AIC: 126.43  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_f\_pred\_OSR\_all <- hyp\_3\_f\_pred\_OSR\_all$null.deviance - hyp\_3\_f\_pred\_OSR\_all$deviance  
Chi\_hyp\_3\_f\_pred\_OSR\_all

[1] 1.713106

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_f\_pred\_OSR\_all <- hyp\_3\_f\_pred\_OSR\_all$df.null - hyp\_3\_f\_pred\_OSR\_all$df.residual  
df\_hyp\_3\_f\_pred\_OSR\_all

[1] 1

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_f\_pred\_OSR\_all <- 1 - pchisq(Chi\_hyp\_3\_f\_pred\_OSR\_all, df\_hyp\_3\_f\_pred\_OSR\_all)  
prob\_Chi\_hyp\_3\_f\_pred\_OSR\_all

[1] 0.1905829

* Neither the model, nor the predictor, are significant at the .05 level.

##### Mothers’ fWHR Only

Now we will look at whether fathers’ fWHR predicts offspring sex without controlling for mothers’ facial dominance.

* Fitting the model:

hyp\_3\_m\_pred\_OSR\_all <- glm(shared\_child\_sex ~ z\_m\_fWHR, family = binomial(link = logit), data = dyadic\_data)

* Summarizing the model and testing it against the intercept-only model:

# Producing the summary of the model  
summary(hyp\_3\_m\_pred\_OSR\_all)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_fWHR, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.31422 0.21363 1.471 0.141  
z\_m\_fWHR -0.08784 0.21442 -0.410 0.682  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 122.58 on 89 degrees of freedom  
Residual deviance: 122.41 on 88 degrees of freedom  
 (14 observations deleted due to missingness)  
AIC: 126.41  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_m\_pred\_OSR\_all <- hyp\_3\_m\_pred\_OSR\_all$null.deviance - hyp\_3\_m\_pred\_OSR\_all$deviance  
Chi\_hyp\_3\_m\_pred\_OSR\_all

[1] 0.1679021

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_m\_pred\_OSR\_all <- hyp\_3\_m\_pred\_OSR\_all$df.null - hyp\_3\_m\_pred\_OSR\_all$df.residual  
df\_hyp\_3\_m\_pred\_OSR\_all

[1] 1

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_m\_pred\_OSR\_all <- 1 - pchisq(Chi\_hyp\_3\_m\_pred\_OSR\_all, df\_hyp\_3\_m\_pred\_OSR\_all)  
prob\_Chi\_hyp\_3\_m\_pred\_OSR\_all

[1] 0.681983

* Neither the model, nor the predictor, are significant at the .05 level.

#### Neutral Faces Only

Now we will repeat the same analysis as just above but with the neutral faces only.

We will first standardize our new independent variables in the neutral faces data frame, f\_fWHR.

# Standardizing the f\_fWHR variable as z\_f\_fWHR and m\_fWHR as z\_m\_fWHR  
neutral\_face\_dyadic\_data$z\_f\_fWHR <- scale(neutral\_face\_dyadic\_data$f\_fWHR, center = TRUE, scale = TRUE)  
neutral\_face\_dyadic\_data$z\_m\_fWHR <- scale(neutral\_face\_dyadic\_data$m\_fWHR, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

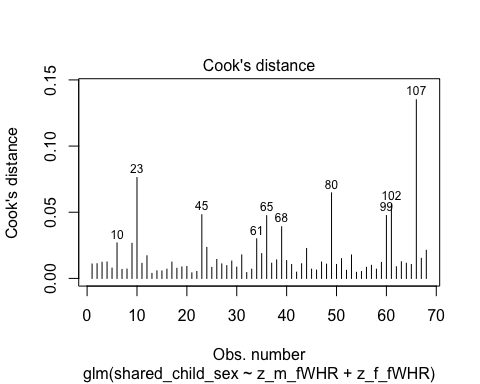
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers fWHR  
hyp\_3\_pred\_OSR\_neutral <- glm(shared\_child\_sex ~ z\_m\_fWHR + z\_f\_fWHR, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)

##### Assumptions

Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(hyp\_3\_pred\_OSR\_neutral, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(hyp\_3\_pred\_OSR\_neutral)

z\_m\_fWHR z\_f\_fWHR   
1.081309 1.081309

1/vif(hyp\_3\_pred\_OSR\_neutral)

z\_m\_fWHR z\_f\_fWHR   
0.9248049 0.9248049

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. I will add 10 to the f\_fWHR variable to make it positive, then I will make a natural log transformation to it.

# Creating c\_z\_f\_fWHR, which represents standardized fathers' fWHR after adding a constant of 10  
neutral\_face\_dyadic\_data$c\_z\_f\_fWHR <- neutral\_face\_dyadic\_data$z\_f\_fWHR + 10  
# Creating c\_z\_m\_fWHR, which represents standardized mothers' fWHR after adding a constant of 10  
neutral\_face\_dyadic\_data$c\_z\_m\_fWHR <- neutral\_face\_dyadic\_data$z\_m\_fWHR + 10  
  
# Creating ln\_c\_z\_f\_fWHR, which represents the natural log of standardized fathers' fWHR after adding a constant of 10  
neutral\_face\_dyadic\_data$ln\_c\_z\_f\_fWHR <- log(neutral\_face\_dyadic\_data$c\_z\_f\_fWHR)  
# Creating ln\_c\_z\_m\_fWHR, which represents the natural log of standardized fathers' fWHR after adding a constant of 10  
neutral\_face\_dyadic\_data$ln\_c\_z\_m\_fWHR <- log(neutral\_face\_dyadic\_data$c\_z\_m\_fWHR)

* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_hyp\_3\_pred\_OSR\_neutral <- glm(shared\_child\_sex ~ c\_z\_m\_fWHR + c\_z\_f\_fWHR + c\_z\_m\_fWHR:ln\_c\_z\_m\_fWHR + c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
summary(BT\_test\_hyp\_3\_pred\_OSR\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_fWHR + c\_z\_f\_fWHR + c\_z\_m\_fWHR:ln\_c\_z\_m\_fWHR +   
 c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 5.9371 51.7111 0.115 0.909  
c\_z\_m\_fWHR -5.4289 12.8844 -0.421 0.673  
c\_z\_f\_fWHR 3.3566 12.3263 0.272 0.785  
c\_z\_m\_fWHR:ln\_c\_z\_m\_fWHR 1.5864 3.8894 0.408 0.683  
c\_z\_f\_fWHR:ln\_c\_z\_f\_fWHR -0.9313 3.7075 -0.251 0.802  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 92.792 on 67 degrees of freedom  
Residual deviance: 91.309 on 63 degrees of freedom  
 (13 observations deleted due to missingness)  
AIC: 101.31  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Model

Here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(hyp\_3\_pred\_OSR\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_fWHR + z\_f\_fWHR, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.3288 0.2499 1.315 0.188  
z\_m\_fWHR -0.1840 0.2626 -0.701 0.483  
z\_f\_fWHR 0.2708 0.2638 1.027 0.305  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 92.792 on 67 degrees of freedom  
Residual deviance: 91.520 on 65 degrees of freedom  
 (13 observations deleted due to missingness)  
AIC: 97.52  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_pred\_OSR\_neutral <- hyp\_3\_pred\_OSR\_neutral$null.deviance - hyp\_3\_pred\_OSR\_neutral$deviance  
Chi\_hyp\_3\_pred\_OSR\_neutral

[1] 1.272484

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_pred\_OSR\_neutral <- hyp\_3\_pred\_OSR\_neutral$df.null - hyp\_3\_pred\_OSR\_neutral$df.residual  
df\_hyp\_3\_pred\_OSR\_neutral

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_pred\_OSR\_neutral <- 1 - pchisq(Chi\_hyp\_3\_pred\_OSR\_neutral, df\_hyp\_3\_pred\_OSR\_neutral)  
prob\_Chi\_hyp\_3\_pred\_OSR\_neutral

[1] 0.5292776

* The full model is not significant (x2(2) = 1.27, *p* = .529). Neither of the predictors are significant. I will still exponentiate the coefficients to check them.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(hyp\_3\_pred\_OSR\_neutral$coefficients)

(Intercept) z\_m\_fWHR z\_f\_fWHR   
 1.3892589 0.8319212 1.3110644

##### Fathers’ fWHR Only

Now we will look at whether fathers’ fWHR predicts offspring sex without controlling for mothers’ facial dominance.

* Fitting the model:

hyp\_3\_f\_pred\_OSR\_neutral <- glm(shared\_child\_sex ~ z\_f\_fWHR, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)

* Summarizing the model and testing it against the intercept-only model:

# Producing the summary of the model  
summary(hyp\_3\_f\_pred\_OSR\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_f\_fWHR, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.4141 0.2383 1.738 0.0823 .  
z\_f\_fWHR 0.2803 0.2491 1.125 0.2605   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 100.952 on 74 degrees of freedom  
Residual deviance: 99.631 on 73 degrees of freedom  
 (6 observations deleted due to missingness)  
AIC: 103.63  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_f\_pred\_OSR\_neutral <- hyp\_3\_f\_pred\_OSR\_neutral$null.deviance - hyp\_3\_f\_pred\_OSR\_neutral$deviance  
Chi\_hyp\_3\_f\_pred\_OSR\_neutral

[1] 1.320907

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_f\_pred\_OSR\_neutral <- hyp\_3\_f\_pred\_OSR\_neutral$df.null - hyp\_3\_f\_pred\_OSR\_neutral$df.residual  
df\_hyp\_3\_f\_pred\_OSR\_neutral

[1] 1

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_f\_pred\_OSR\_neutral <- 1 - pchisq(Chi\_hyp\_3\_f\_pred\_OSR\_neutral, df\_hyp\_3\_f\_pred\_OSR\_neutral)  
prob\_Chi\_hyp\_3\_f\_pred\_OSR\_neutral

[1] 0.2504293

* Neither the model, nor the predictor, are significant at the .05 level.

##### Mothers’ fWHR Only

Now we will look at whether fathers’ fWHR predicts offspring sex without controlling for mothers’ facial dominance.

* Fitting the model:

hyp\_3\_m\_pred\_OSR\_neutral <- glm(shared\_child\_sex ~ z\_m\_fWHR, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)

* Summarizing the model and testing it against the intercept-only model:

# Producing the summary of the model  
summary(hyp\_3\_m\_pred\_OSR\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_fWHR, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.19240 0.23518 0.818 0.413  
z\_m\_fWHR 0.02353 0.23689 0.099 0.921  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 100.53 on 72 degrees of freedom  
Residual deviance: 100.52 on 71 degrees of freedom  
 (8 observations deleted due to missingness)  
AIC: 104.52  
  
Number of Fisher Scoring iterations: 3

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_hyp\_3\_m\_pred\_OSR\_neutral <- hyp\_3\_m\_pred\_OSR\_neutral$null.deviance - hyp\_3\_m\_pred\_OSR\_neutral$deviance  
Chi\_hyp\_3\_m\_pred\_OSR\_neutral

[1] 0.009870298

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_hyp\_3\_m\_pred\_OSR\_neutral <- hyp\_3\_m\_pred\_OSR\_neutral$df.null - hyp\_3\_m\_pred\_OSR\_neutral$df.residual  
df\_hyp\_3\_m\_pred\_OSR\_neutral

[1] 1

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_hyp\_3\_m\_pred\_OSR\_neutral <- 1 - pchisq(Chi\_hyp\_3\_m\_pred\_OSR\_neutral, df\_hyp\_3\_m\_pred\_OSR\_neutral)  
prob\_Chi\_hyp\_3\_m\_pred\_OSR\_neutral

[1] 0.9208609

* Neither the model, nor the predictor, are significant at the .05 level.

### Summary for Hypothesis 3

Although the direction of the effects here are as predicted, there are no significant relationships between fathers’ fWHR and sex of first born child in these models. It is worth noting that we have a loss of power due to some faces not having fWHR measurements. However, overall these analyses do not provide strong evidence for Hypothesis 3.

# Addressing Critiques from EHB Editor

An editor from EHB brought to our attention a few critiques that we may be able to address with additional analyses:

1. By removing the effects of facial masculinity/femininity, attractiveness, and age from facial dominance ratings, interpretation becomes more complicated, so we should run the analyses with raw facial dominance as well.
2. We are limited in our conclusions because data collection occurred around 4 years after the birth of the first child on average. We could potentially address this by controlling for age.

In the following, I will run analyses that address these critiques.

## Raw Facial Dominance Model

### Hypothesis 2: Raw Facial Dominance

Now we will retest the hypothesis with facial dominance as such. Because there are group differences in facial dominance between the neutral and non-neutral facial expression groups, we will conduct this analysis on both the full dyadic dataset and on the neutral faces only.

#### Full Dataset

First we will standardize our independent variables, m\_facial\_dominance and f\_facial\_dominance.

# Standardizing the m\_facial\_dominance and f\_facial\_dominance variables as z\_m\_facial\_dominance and z\_f\_facial\_dominance  
dyadic\_data$z\_m\_facial\_dominance <- scale(dyadic\_data$m\_facial\_dominance, center = TRUE, scale = TRUE)  
dyadic\_data$z\_f\_facial\_dominance <- scale(dyadic\_data$f\_facial\_dominance, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

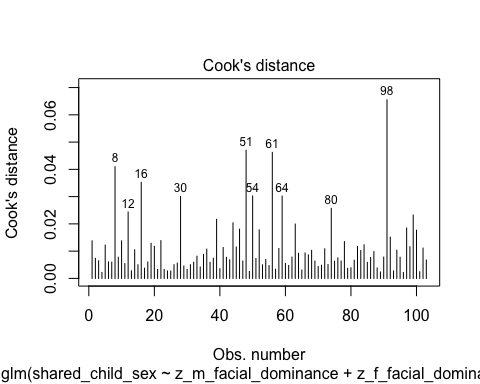
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers  
fac\_dom\_main\_effects\_model <- glm(shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)  
  
# Fitting the model with the main effects of residual facial dominance for mothers and fathers as well as their interaction  
fac\_dom\_interaction\_model <- glm(shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance + z\_m\_facial\_dominance:z\_f\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

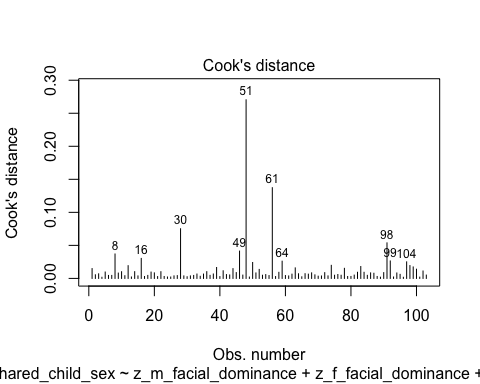
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(fac\_dom\_main\_effects\_model, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(fac\_dom\_interaction\_model, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(fac\_dom\_main\_effects\_model)

z\_m\_facial\_dominance z\_f\_facial\_dominance   
 1.02046 1.02046

1/vif(fac\_dom\_main\_effects\_model)

z\_m\_facial\_dominance z\_f\_facial\_dominance   
 0.9799504 0.9799504

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. The below code determines the lowest value of z\_m\_facial\_dominance and z\_f\_facial\_dominance so that we can add a constant that makes all values for the variable positive before making the natural log transformation.

min(dyadic\_data$z\_m\_facial\_dominance)

[1] -2.221293

min(dyadic\_data$z\_f\_facial\_dominance, na.rm = TRUE) # Because there is one NA value in the father data we specify that the algorithm should ignore this value

[1] -2.250934

* Given these minimum values, I will add 3 to each variable to make them positive, then I will make a natural log transformation to both.

# Creating c\_z\_m\_facial\_dominance, which represents mothers' facial dominance after adding a constant of 3  
dyadic\_data$c\_z\_m\_facial\_dominance <- dyadic\_data$z\_m\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_m\_facial\_dominance, which represents the natural log of mothers' facial dominance after adding a constant of 3  
dyadic\_data$ln\_c\_z\_m\_facial\_dominance <- log(dyadic\_data$c\_z\_m\_facial\_dominance)  
  
# Creating c\_z\_f\_facial\_dominance, which represents fathers' facial dominance after adding a constant of 3  
dyadic\_data$c\_z\_f\_facial\_dominance <- dyadic\_data$z\_f\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_f\_facial\_dominance, which represents the natural log of fathers' facial dominance after adding a constant of 3  
dyadic\_data$ln\_c\_z\_f\_facial\_dominance <- log(dyadic\_data$c\_z\_f\_facial\_dominance)

* Through manually doing the transformations for a couple numbers, I believe that this transformation was done correctly.
* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_facial\_dom\_model <- glm(shared\_child\_sex ~ c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance + c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)  
  
summary(BT\_test\_facial\_dom\_model)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance +   
 c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value  
(Intercept) -5.65315 3.75784 -1.504  
c\_z\_m\_facial\_dominance 4.30284 2.38592 1.803  
c\_z\_f\_facial\_dominance 0.17066 2.48779 0.069  
c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance -2.06279 1.14918 -1.795  
c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance -0.05533 1.21132 -0.046  
 Pr(>|z|)   
(Intercept) 0.1325   
c\_z\_m\_facial\_dominance 0.0713 .  
c\_z\_f\_facial\_dominance 0.9453   
c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance 0.0727 .  
c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance 0.9636   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 134.17 on 98 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 144.17  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(fac\_dom\_main\_effects\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.41452 0.20153 2.057 0.0397 \*  
z\_m\_facial\_dominance 0.03685 0.20376 0.181 0.8565   
z\_f\_facial\_dominance 0.07664 0.20461 0.375 0.7080   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 138.28 on 100 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 144.28  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_fac\_dom\_main\_effects\_model <- fac\_dom\_main\_effects\_model$null.deviance - fac\_dom\_main\_effects\_model$deviance  
Chi\_fac\_dom\_main\_effects\_model

[1] 0.1965799

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_fac\_dom\_main\_effects\_model <- fac\_dom\_main\_effects\_model$df.null - fac\_dom\_main\_effects\_model$df.residual  
df\_fac\_dom\_main\_effects\_model

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_fac\_dom\_main\_effects\_model <- 1 - pchisq(Chi\_fac\_dom\_main\_effects\_model, df\_fac\_dom\_main\_effects\_model)  
prob\_Chi\_fac\_dom\_main\_effects\_model

[1] 0.906386

* Although the whole model is not a better fit than the baseline model with just the intercept (x2(2) = .197, *p* = .906). In addition, neither of the predictors are close to significant. I will exponentiate these coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(fac\_dom\_main\_effects\_model$coefficients)

(Intercept) z\_m\_facial\_dominance z\_f\_facial\_dominance   
 1.513638 1.037542 1.079657

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(fac\_dom\_interaction\_model)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance +   
 z\_m\_facial\_dominance:z\_f\_facial\_dominance, family = binomial(link = logit),   
 data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.39030 0.20363 1.917 0.0553  
z\_m\_facial\_dominance 0.05891 0.20767 0.284 0.7767  
z\_f\_facial\_dominance 0.07431 0.20792 0.357 0.7208  
z\_m\_facial\_dominance:z\_f\_facial\_dominance 0.22360 0.18020 1.241 0.2147  
   
(Intercept) .  
z\_m\_facial\_dominance   
z\_f\_facial\_dominance   
z\_m\_facial\_dominance:z\_f\_facial\_dominance   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 136.66 on 99 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 144.66  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_fac\_dom\_interaction\_model <- fac\_dom\_interaction\_model$null.deviance - fac\_dom\_interaction\_model$deviance  
Chi\_fac\_dom\_interaction\_model

[1] 1.812384

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_fac\_dom\_interaction\_model <- fac\_dom\_interaction\_model$df.null - fac\_dom\_interaction\_model$df.residual  
df\_fac\_dom\_interaction\_model

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_fac\_dom\_interaction\_model <- 1 - pchisq(Chi\_fac\_dom\_interaction\_model, df\_fac\_dom\_interaction\_model)  
prob\_Chi\_fac\_dom\_interaction\_model

[1] 0.6122438

* The model is still not significant (x2(3) = 1.812, *p* = .612). None of the predictors, including the interaction, are significant. I will exponentiate these coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(fac\_dom\_interaction\_model$coefficients)

(Intercept)   
 1.477430   
 z\_m\_facial\_dominance   
 1.060683   
 z\_f\_facial\_dominance   
 1.077138   
z\_m\_facial\_dominance:z\_f\_facial\_dominance   
 1.250575

Because the whole interaction model is not significant compared to the intercept-only model, it would be very surprising if the interaction model would fit better than the main-effects only model, but I will go ahead and do the comparison anyways.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_fac\_dom\_interaction\_model\_v\_main <- fac\_dom\_main\_effects\_model$deviance - fac\_dom\_interaction\_model$deviance  
Chi\_fac\_dom\_interaction\_model\_v\_main

[1] 1.615804

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_fac\_dom\_interaction\_model\_v\_main <- fac\_dom\_main\_effects\_model$df.residual - fac\_dom\_interaction\_model$df.residual  
df\_fac\_dom\_interaction\_model\_v\_main

[1] 1

prob\_Chi\_fac\_dom\_interaction\_model\_v\_main <- 1 - pchisq(Chi\_fac\_dom\_interaction\_model\_v\_main, df\_fac\_dom\_interaction\_model\_v\_main)  
prob\_Chi\_fac\_dom\_interaction\_model\_v\_main

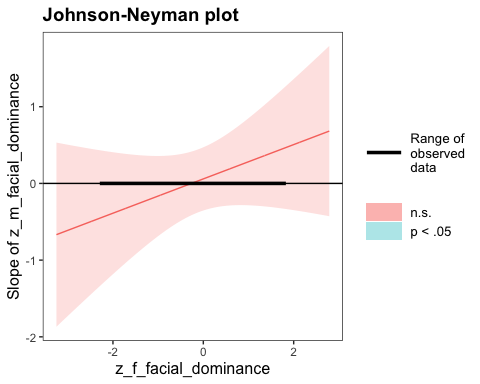
[1] 0.2036779

* As suspected, the Chi-squared test does not indicate that the interaction model fits better than the main effects model (x2(1) = 1.616, *p* = .204).

Although the interaction is not significant in our model, our hypothesis predicts that at high levels of fathers’ facial dominance mothers’ dominance predicts the probability of a first born son, so I will run a simple slopes and Johnson-Neyman analysis with mothers’ residual facial dominance as the focal predictor and fathers’ residual facial dominance as the moderator.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(fac\_dom\_interaction\_model, pred = z\_m\_facial\_dominance, modx = z\_f\_facial\_dominance, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_facial\_dominance when z\_f\_facial\_dominance = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------- ------ -------- ------  
 -0.16 0.26 -0.63 0.53  
  
Slope of z\_m\_facial\_dominance when z\_f\_facial\_dominance = 1.567247e-15 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.06 0.21 0.28 0.78  
  
Slope of z\_m\_facial\_dominance when z\_f\_facial\_dominance = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.28 0.29 0.99 0.32

* Although nothing here is significant, I will exponentiate the coefficients here to make them more interpretable.

# Producing odds-ratios for each value of the slope for mothers' residual facial dominance for each level of the moderator in the simple slopes analysis  
exp(-.16) # -1 SD

[1] 0.8521438

exp(.06) # Mean

[1] 1.061837

exp(.28) # +1 SD

[1] 1.32313

# Taking the inverse of the odds-ratio for the first coefficient  
1 - 0.8521438

[1] 0.1478562

* The Johnson-Neyman analysis indicates that there are no values of fathers’ facial dominance for which mothers’ facial dominance is a significant predictor of the probability of having a first born son. Worthy of note, however, the simple slopes analysis indicates that at low levels of fathers’ residual facial dominance (-1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 14.8% *decrease* in the odds of having a first born son and at high levels of fathers’ residual facial dominance (+1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 32.3% *increase* in the odds of having a first born son.
  + Although not significant, the direction of this moderation effect is consistent with our hypothesis.
  + It is possible that in the population there exists such an effect, whereas we simply do not have enough power to detect it.

#### Neutral Faces Only

Now we will repeat the same analysis as just above but with the neutral faces only. I will first standardize our independent variables in the new data frame, m\_facial\_dominance and f\_facial\_dominance.

# Standardizing the m\_facial\_dominance and f\_facial\_dominance variables as z\_m\_facial\_dominance and z\_f\_facial\_dominance (because the mean and SD may be slightly different in this new data frame)  
neutral\_face\_dyadic\_data$z\_m\_facial\_dominance <- scale(neutral\_face\_dyadic\_data$m\_facial\_dominance, center = TRUE, scale = TRUE)  
neutral\_face\_dyadic\_data$z\_f\_facial\_dominance <- scale(neutral\_face\_dyadic\_data$f\_facial\_dominance, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

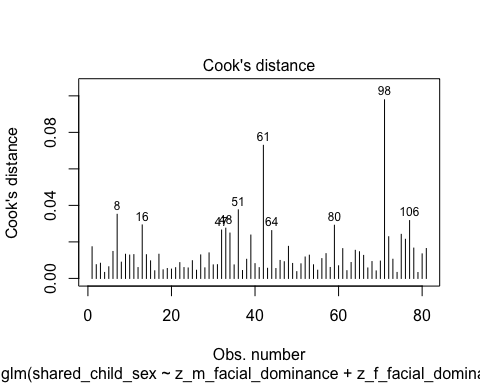
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers  
fac\_dom\_main\_effects\_model\_neutral <- glm(shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
# Fitting the model with the main effects of residual facial dominance for mothers and fathers as well as their interaction  
fac\_dom\_interaction\_model\_neutral <- glm(shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance + z\_m\_facial\_dominance:z\_f\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)

##### Assumptions

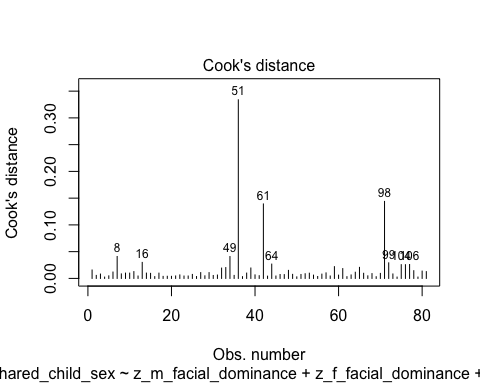
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(fac\_dom\_main\_effects\_model\_neutral, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(fac\_dom\_interaction\_model\_neutral, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(fac\_dom\_main\_effects\_model\_neutral)

z\_m\_facial\_dominance z\_f\_facial\_dominance   
 1.001747 1.001747

1/vif(fac\_dom\_main\_effects\_model\_neutral)

z\_m\_facial\_dominance z\_f\_facial\_dominance   
 0.998256 0.998256

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. The below code determines the lowest value of z\_m\_facial\_dominance and z\_f\_facial\_dominance so that we can add a constant that makes all values for the variable positive before making the natural log transformation.

min(neutral\_face\_dyadic\_data$z\_m\_facial\_dominance)

[1] -2.278624

min(neutral\_face\_dyadic\_data$z\_f\_facial\_dominance)

[1] -2.523205

* Given these minimum values, I will add 3 to each variable to make them positive, then I will make a natural log transformation to both.

# Creating c\_z\_m\_facial\_dominance, which represents mothers' facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$c\_z\_m\_facial\_dominance <- neutral\_face\_dyadic\_data$z\_m\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_m\_facial\_dominance, which represents the natural log of mothers' facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$ln\_c\_z\_m\_facial\_dominance <- log(neutral\_face\_dyadic\_data$c\_z\_m\_facial\_dominance)  
  
# Creating c\_z\_f\_facial\_dominance, which represents fathers' facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$c\_z\_f\_facial\_dominance <- neutral\_face\_dyadic\_data$z\_f\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_f\_facial\_dominance, which represents the natural log of fathers' facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$ln\_c\_z\_f\_facial\_dominance <- log(neutral\_face\_dyadic\_data$c\_z\_f\_facial\_dominance)

* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_facial\_dom\_model\_neutral <- glm(shared\_child\_sex ~ c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance + c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
summary(BT\_test\_facial\_dom\_model\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance +   
 c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance + c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance,   
 family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value  
(Intercept) -4.33791 3.73137 -1.163  
c\_z\_m\_facial\_dominance 3.15706 2.53723 1.244  
c\_z\_f\_facial\_dominance -0.03659 2.42403 -0.015  
c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance -1.43669 1.23970 -1.159  
c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance 0.08294 1.20964 0.069  
 Pr(>|z|)  
(Intercept) 0.245  
c\_z\_m\_facial\_dominance 0.213  
c\_z\_f\_facial\_dominance 0.988  
c\_z\_m\_facial\_dominance:ln\_c\_z\_m\_facial\_dominance 0.246  
c\_z\_f\_facial\_dominance:ln\_c\_z\_f\_facial\_dominance 0.945  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 107.11 on 76 degrees of freedom  
AIC: 117.11  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(fac\_dom\_main\_effects\_model\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance,   
 family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.3297 0.2274 1.450 0.147  
z\_m\_facial\_dominance 0.2309 0.2306 1.002 0.317  
z\_f\_facial\_dominance 0.1446 0.2300 0.629 0.529  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 108.70 on 78 degrees of freedom  
AIC: 114.7  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_fac\_dom\_main\_effects\_model\_neutral <- fac\_dom\_main\_effects\_model\_neutral$null.deviance - fac\_dom\_main\_effects\_model\_neutral$deviance  
Chi\_fac\_dom\_main\_effects\_model\_neutral

[1] 1.493213

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_fac\_dom\_main\_effects\_model\_neutral <- fac\_dom\_main\_effects\_model\_neutral$df.null - fac\_dom\_main\_effects\_model\_neutral$df.residual  
df\_fac\_dom\_main\_effects\_model\_neutral

[1] 2

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_fac\_dom\_main\_effects\_model\_neutral <- 1 - pchisq(Chi\_fac\_dom\_main\_effects\_model\_neutral, df\_fac\_dom\_main\_effects\_model\_neutral)  
prob\_Chi\_fac\_dom\_main\_effects\_model\_neutral

[1] 0.4739722

* The full main effects model is not significant compared to the intercept-only model (x2(2) = 1.493, *p* = .474). Neither of the predictors are significant. I will exponentiate the coefficients to make this more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(fac\_dom\_main\_effects\_model\_neutral$coefficients)

(Intercept) z\_m\_facial\_dominance z\_f\_facial\_dominance   
 1.390487 1.259788 1.155602

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(fac\_dom\_interaction\_model\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_facial\_dominance + z\_f\_facial\_dominance +   
 z\_m\_facial\_dominance:z\_f\_facial\_dominance, family = binomial(link = logit),   
 data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)  
(Intercept) 0.3264 0.2281 1.431 0.152  
z\_m\_facial\_dominance 0.2473 0.2331 1.061 0.289  
z\_f\_facial\_dominance 0.1409 0.2299 0.613 0.540  
z\_m\_facial\_dominance:z\_f\_facial\_dominance 0.1123 0.1905 0.589 0.556  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 108.35 on 77 degrees of freedom  
AIC: 116.35  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_fac\_dom\_interaction\_model\_neutral <- fac\_dom\_interaction\_model\_neutral$null.deviance - fac\_dom\_interaction\_model\_neutral$deviance  
Chi\_fac\_dom\_interaction\_model\_neutral

[1] 1.841549

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_fac\_dom\_interaction\_model\_neutral <- fac\_dom\_interaction\_model\_neutral$df.null - fac\_dom\_interaction\_model\_neutral$df.residual  
df\_fac\_dom\_interaction\_model\_neutral

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_fac\_dom\_interaction\_model\_neutral <- 1 - pchisq(Chi\_fac\_dom\_interaction\_model\_neutral, df\_fac\_dom\_interaction\_model\_neutral)  
prob\_Chi\_fac\_dom\_interaction\_model\_neutral

[1] 0.6059354

* The full interaction model is not significant compared to the intercept only (x2(3) = 1.842, *p* = .606). None of the predictors are significant. Again, I will exponentiate these coefficients to understand it better.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(fac\_dom\_interaction\_model\_neutral$coefficients)

(Intercept)   
 1.385968   
 z\_m\_facial\_dominance   
 1.280619   
 z\_f\_facial\_dominance   
 1.151255   
z\_m\_facial\_dominance:z\_f\_facial\_dominance   
 1.118803

Now to compare the main-effects model with the main-effects and interaction model.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_fac\_dom\_interaction\_model\_v\_main\_neutral <- fac\_dom\_main\_effects\_model\_neutral$deviance - fac\_dom\_interaction\_model\_neutral$deviance  
Chi\_fac\_dom\_interaction\_model\_v\_main\_neutral

[1] 0.3483355

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_fac\_dom\_interaction\_model\_v\_main\_neutral <- fac\_dom\_main\_effects\_model\_neutral$df.residual - fac\_dom\_interaction\_model\_neutral$df.residual  
df\_fac\_dom\_interaction\_model\_v\_main\_neutral

[1] 1

prob\_Chi\_fac\_dom\_interaction\_model\_v\_main\_neutral <- 1 - pchisq(Chi\_fac\_dom\_interaction\_model\_v\_main\_neutral, df\_fac\_dom\_interaction\_model\_v\_main\_neutral)  
prob\_Chi\_fac\_dom\_interaction\_model\_v\_main\_neutral

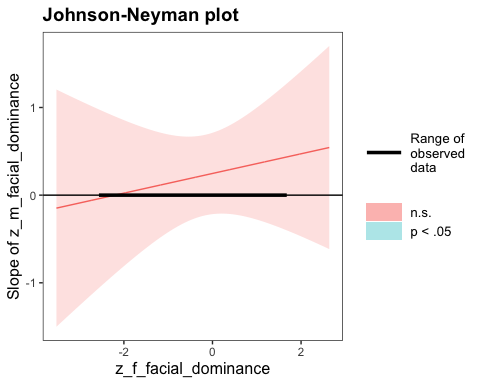
[1] 0.5550569

* The Chi-squared test does not indicate that the interaction model does not fit better than the main effects model (x2(1) = .348, *p* = .555).

Although the interaction is not significant in our model, our hypothesis predicts that at high levels of fathers’ facial dominance mothers’ dominance predicts the probability of a first born son, so I will run a simple slopes and Johnson-Neyman analysis with mothers’ residual facial dominance as the focal predictor and fathers’ residual facial dominance as the moderator.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(fac\_dom\_interaction\_model\_neutral, pred = z\_m\_facial\_dominance, modx = z\_f\_facial\_dominance, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_facial\_dominance when z\_f\_facial\_dominance = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.14 0.28 0.48 0.63  
  
Slope of z\_m\_facial\_dominance when z\_f\_facial\_dominance = 3.974873e-16 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.25 0.23 1.06 0.29  
  
Slope of z\_m\_facial\_dominance when z\_f\_facial\_dominance = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.36 0.32 1.12 0.26

* Although nothing here is significant, I will exponentiate the coefficients here to make them more interpretable.

# Producing odds-ratios for each value of the slope for mothers' residual facial dominance for each level of the moderator in the simple slopes analysis  
exp(-.14) # -1 SD

[1] 0.8693582

exp(.25) # Mean

[1] 1.284025

exp(.36) # +1 SD

[1] 1.433329

# Taking the inverse of the odds-ratio for the first coefficient  
1 - 0.8693582

[1] 0.1306418

* The Johnson-Neyman analysis indicates that there are no values of fathers’ facial dominance for which mothers’ facial dominance is a significant predictor of the probability of having a first born son. Similar to the analysis with all facial expression types, however, the simple slopes analysis indicates that at low levels of fathers’ residual facial dominance (-1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 13.1% *decrease* in the odds of having a first born son; at the mean for fathers’ residual facial dominance a one standard deviation increase in mothers’ residual facial dominance is associated with an 28.4% *increase* in the odds of having a first born son; and at high levels of fathers’ residual facial dominance (+1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 43.3% *increase* in the odds of having a first born son.
  + Although not significant, the direction of this moderation effect is consistent with our hypothesis, like in the model with all facial expressions included.
  + Again, it is possible that in the population there exists such an effect, whereas we simply do not have enough power to detect it.

## Controlling for Age of First-Born

### Hypothesis 2: Facial Dominance Operationalization

Now we will test the hypothesis with the standardized residuals for facial dominance that we created within sexes, while controlling shared\_child\_age. Because there are group differences in facial dominance between the neutral and non-neutral facial expression groups, we will conduct this analysis on both the full dyadic dataset and on the neutral faces only.

#### Full Dataset

Now we will fit our two models, one with the interaction term and one without (both with shared\_child\_age as a covariate).

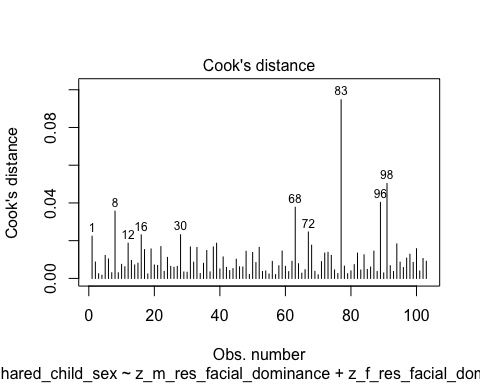
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers  
res\_fac\_dom\_main\_effects\_model\_a <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance + shared\_child\_age, family = binomial(link = logit), data = dyadic\_data)  
  
# Fitting the model with the main effects of residual facial dominance for mothers and fathers as well as their interaction  
res\_fac\_dom\_interaction\_model\_a <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance + shared\_child\_age + z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)

##### Assumptions

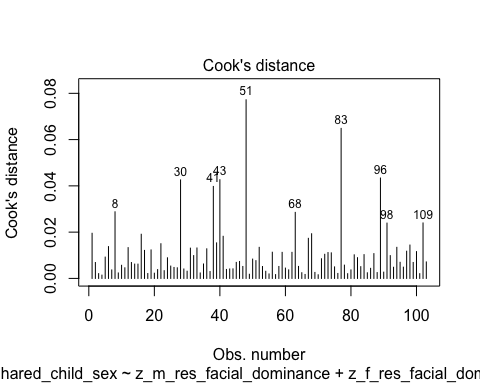
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(res\_fac\_dom\_main\_effects\_model\_a, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(res\_fac\_dom\_interaction\_model\_a, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(res\_fac\_dom\_main\_effects\_model\_a)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance shared\_child\_age   
 1.016152 1.015381 1.001956

1/vif(res\_fac\_dom\_main\_effects\_model\_a)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance shared\_child\_age   
 0.9841047 0.9848521 0.9980477

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. The below code determines the lowest value of z\_m\_res\_facial\_dominance and z\_f\_res\_facial\_dominance so that we can add a constant that makes all values for the variable positive before making the natural log transformation.

min(dyadic\_data$z\_m\_res\_facial\_dominance)

[1] -2.439434

min(dyadic\_data$z\_f\_res\_facial\_dominance, na.rm = TRUE) # Because there is one NA value in the father data we specify that the algorithm should ignore this value

[1] -3.361094

* Given these minimum values, I will add 4 to each variable to make them positive, then I will make a natural log transformation to both.

# Creating c\_z\_m\_res\_facial\_dominance, which represents mothers' residual facial dominance after adding a constant of 4  
dyadic\_data$c\_z\_m\_res\_facial\_dominance <- dyadic\_data$z\_m\_res\_facial\_dominance + 4  
  
# Creating ln\_c\_z\_m\_res\_facial\_dominance, which represents the natural log of mothers' residual facial dominance after adding a constant of 4  
dyadic\_data$ln\_c\_z\_m\_res\_facial\_dominance <- log(dyadic\_data$c\_z\_m\_res\_facial\_dominance)  
  
# Creating c\_z\_f\_res\_facial\_dominance, which represents fathers' residual facial dominance after adding a constant of 4  
dyadic\_data$c\_z\_f\_res\_facial\_dominance <- dyadic\_data$z\_f\_res\_facial\_dominance + 4  
  
# Creating ln\_c\_z\_f\_res\_facial\_dominance, which represents the natural log of fathers' residual facial dominance after adding a constant of 4  
dyadic\_data$ln\_c\_z\_f\_res\_facial\_dominance <- log(dyadic\_data$c\_z\_f\_res\_facial\_dominance)

* Through manually doing the transformations for a couple numbers, I believe that this transformation was done correctly.
* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_res\_facial\_dom\_model <- glm(shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = dyadic\_data)  
  
summary(BT\_test\_res\_facial\_dom\_model)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error  
(Intercept) -0.2428 6.4165  
c\_z\_m\_res\_facial\_dominance 2.9280 3.0059  
c\_z\_f\_res\_facial\_dominance -3.3793 2.5450  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -1.2074 1.2521  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 1.6497 1.1157  
 z value Pr(>|z|)  
(Intercept) -0.038 0.970  
c\_z\_m\_res\_facial\_dominance 0.974 0.330  
c\_z\_f\_res\_facial\_dominance -1.328 0.184  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -0.964 0.335  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 1.479 0.139  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 130.80 on 98 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 140.8  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(res\_fac\_dom\_main\_effects\_model\_a)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance +   
 shared\_child\_age, family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.18791 0.43067 0.436 0.663   
z\_m\_res\_facial\_dominance 0.01830 0.20554 0.089 0.929   
z\_f\_res\_facial\_dominance 0.37273 0.21639 1.723 0.085 .  
shared\_child\_age 0.06000 0.09576 0.627 0.531   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 134.75 on 99 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 142.75  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_main\_effects\_model\_a <- res\_fac\_dom\_main\_effects\_model\_a$null.deviance - res\_fac\_dom\_main\_effects\_model\_a$deviance  
Chi\_res\_fac\_dom\_main\_effects\_model\_a

[1] 3.725526

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_main\_effects\_model\_a <- res\_fac\_dom\_main\_effects\_model\_a$df.null - res\_fac\_dom\_main\_effects\_model\_a$df.residual  
df\_res\_fac\_dom\_main\_effects\_model\_a

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model\_a <- 1 - pchisq(Chi\_res\_fac\_dom\_main\_effects\_model\_a, df\_res\_fac\_dom\_main\_effects\_model\_a)  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model\_a

[1] 0.2926683

* Although the whole model is not a better fit than the baseline model with just the intercept (x2(3) = 3.726, *p* = .293), the residual facial dominance predictor for fathers was close to significant (b = .373, *z* = 1.723, *p* = .085), with a one standard deviation increase in father’s residual facial dominance leading to a .373 unit increase in the log-odds of having a first born son. I will exponentiate this coefficient to make it more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_main\_effects\_model\_a$coefficients)

(Intercept) z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 1.206720 1.018465 1.451695   
 shared\_child\_age   
 1.061839

* The odds ratio for father’s residual facial dominance (although not significant) is OR = 1.452, indicating that while holding mother’s residual facial dominance constant a one standard deviation increase in father’s residual facial dominance is associated with a 45.2% higher odds of having a first born son.

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(res\_fac\_dom\_interaction\_model\_a)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance +   
 shared\_child\_age + z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value  
(Intercept) 0.10565 0.43861 0.241  
z\_m\_res\_facial\_dominance 0.09536 0.21997 0.434  
z\_f\_res\_facial\_dominance 0.45774 0.23094 1.982  
shared\_child\_age 0.07295 0.09703 0.752  
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.42393 0.26611 1.593  
 Pr(>|z|)   
(Intercept) 0.8097   
z\_m\_res\_facial\_dominance 0.6646   
z\_f\_res\_facial\_dominance 0.0475 \*  
shared\_child\_age 0.4521   
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.1111   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 138.48 on 102 degrees of freedom  
Residual deviance: 131.96 on 98 degrees of freedom  
 (1 observation deleted due to missingness)  
AIC: 141.96  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_interaction\_model\_a <- res\_fac\_dom\_interaction\_model\_a$null.deviance - res\_fac\_dom\_interaction\_model\_a$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_a

[1] 6.514159

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_interaction\_model\_a <- res\_fac\_dom\_interaction\_model\_a$df.null - res\_fac\_dom\_interaction\_model\_a$df.residual  
df\_res\_fac\_dom\_interaction\_model\_a

[1] 4

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_a <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_a, df\_res\_fac\_dom\_interaction\_model\_a)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_a

[1] 0.1639004

* Interestingly—although the model still does not fit significantly better than the intercept-only model (x2(4) = 6.514, *p* = .164)—when the interaction between mothers’ and fathers’ residual facial dominance is added to the model fathers’ residual facial dominance becomes significance (*b* = .458, z = 1.982, *p* = .048). Also of note, but not significant, the interaction between mothers’ and fathers’ residual facial dominance has a positive coefficient (*b* = .424, *z* = 1.593, *p* = .111). I will exponentiate these coefficients to make them more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_interaction\_model\_a$coefficients)

(Intercept)   
 1.111427   
 z\_m\_res\_facial\_dominance   
 1.100056   
 z\_f\_res\_facial\_dominance   
 1.580504   
 shared\_child\_age   
 1.075678   
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance   
 1.527948

* For father’s residual facial dominance, the OR = 1.581, and for the interaction between mothers’ and fathers’ residual facial dominance the OR = 1.528. This indicates that, with other variables held constant, a one standard deviation unit increase in fathers’ residual facial dominance is associated with a 58.1% increase in the odds of having a first born son, and a one unit increase in the product of mothers’ and fathers’ residual facial dominance is associated with a 52.8% increase in the odds of having a first born son (while controlling for child’s’ age).

Because the whole interaction model is not significant compared to the intercept-only model, it would be very surprising if the interaction model would fit better than the main-effects only model, but I will go ahead and do the comparison anyways.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main <- res\_fac\_dom\_main\_effects\_model\_a$deviance - res\_fac\_dom\_interaction\_model\_a$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main

[1] 2.788633

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_res\_fac\_dom\_interaction\_model\_a\_v\_main <- res\_fac\_dom\_main\_effects\_model\_a$df.residual - res\_fac\_dom\_interaction\_model\_a$df.residual  
df\_res\_fac\_dom\_interaction\_model\_a\_v\_main

[1] 1

prob\_Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main, df\_res\_fac\_dom\_interaction\_model\_a\_v\_main)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main

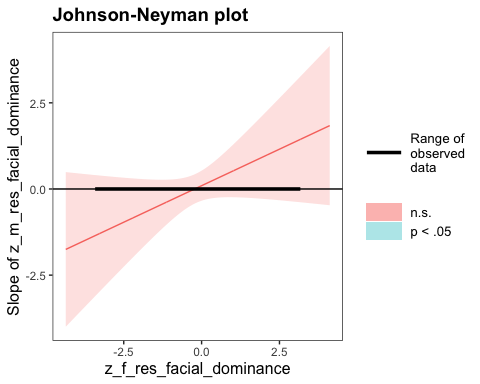
[1] 0.09493517

* As suspected, the Chi-squared test does not indicate that the interaction model fits better than the main effects model (x2(1) = 2.789, *p* = .09).

Although the interaction is not significant in our model, our hypothesis predicts that at high levels of fathers’ facial dominance mothers’ dominance predicts the probability of a first born son, so I will run a simple slopes and Johnson-Neyman analysis with mothers’ residual facial dominance as the focal predictor and fathers’ residual facial dominance as the moderator.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(res\_fac\_dom\_interaction\_model\_a, pred = z\_m\_res\_facial\_dominance, modx = z\_f\_res\_facial\_dominance, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------- ------ -------- ------  
 -0.33 0.30 -1.08 0.28  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = 1.293464e-17 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.10 0.22 0.43 0.66  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.52 0.38 1.36 0.17

* Although nothing here is significant, I will exponentiate the coefficients here to make them more interpretable.

# Producing odds-ratios for each value of the slope for mothers' residual facial dominance for each level of the moderator in the simple slopes analysis  
exp(-.33) # -1 SD

[1] 0.7189237

exp(.10) # Mean

[1] 1.105171

exp(.52) # +1 SD

[1] 1.682028

# Taking the inverse of the odds-ratio for the first coefficient  
1 - 0.7189237

[1] 0.2810763

* The Johnson-Neyman analysis indicates that there are no values of fathers’ residual facial dominance for which mothers’ residual facial dominance is a significant predictor of the probability of having a first born child. Worthy of note, however, the simple slopes analysis indicates that at low levels of fathers’ residual facial dominance (-1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 28.1% *decrease* in the odds of having a first born son and at high levels of fathers’ residual facial dominance (+1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 68.2% *increase* in the odds of having a first born son.
  + Although not significant, the direction of this moderation effect is consistent with our hypothesis.
  + It is possible that in the population there exists such an effect, whereas we simply do not have enough power to detect it.

#### Neutral Faces Only

Now we will repeat the same analysis as just above but with the neutral faces only.

First, we will standardize our independent variables in the new data frame, m\_res\_facial\_dominance and f\_res\_facial\_dominance.

# Standardizing the m\_res\_facial\_dominance and f\_res\_facial\_dominance variables as z\_m\_res\_facial\_dominance and z\_f\_res\_facial\_dominance (because the mean and SD may be slightly different in this new data frame)  
neutral\_face\_dyadic\_data$z\_m\_res\_facial\_dominance <- scale(neutral\_face\_dyadic\_data$m\_res\_facial\_dominance, center = TRUE, scale = TRUE)  
neutral\_face\_dyadic\_data$z\_f\_res\_facial\_dominance <- scale(neutral\_face\_dyadic\_data$f\_res\_facial\_dominance, center = TRUE, scale = TRUE)

Now we will fit our two models, one with the interaction term and one without.

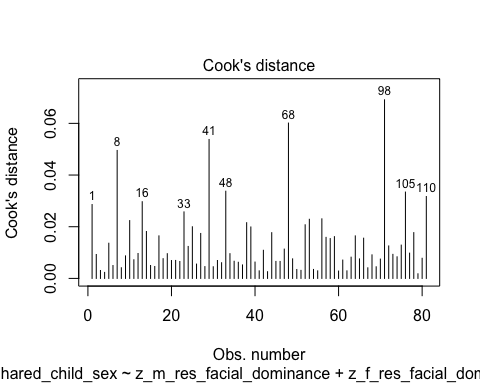
# Fitting the model with only the main effects of residual facial dominance for mothers and fathers  
res\_fac\_dom\_main\_effects\_model\_a\_neutral <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance + shared\_child\_age, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
# Fitting the model with the main effects of residual facial dominance for mothers and fathers as well as their interaction  
res\_fac\_dom\_interaction\_model\_a\_neutral <- glm(shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance + shared\_child\_age + z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)

##### Assumptions

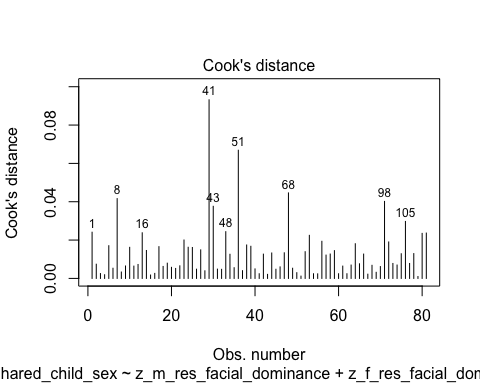
Now we need to check our assumptions.

1. Dichotomous DV (satisfied)
2. One or more continuous IVs, continuous or nominal (satisfied)
3. Independence of observations (satisfied)
4. Categories of dichotomous DV and nominal IVs are mutually exclusive and exhaustive (satisfied)
5. No outliers: Will assess by plotting Cooks distance for each case in each models below (and identify the top 10 cases by number).

# Plotting Cook's distance for the main effects model  
plot(res\_fac\_dom\_main\_effects\_model\_a\_neutral, which = 4, id.n = 10)



#Plotting Cook's distance for the interaction model  
plot(res\_fac\_dom\_interaction\_model\_a\_neutral, which = 4, id.n = 10)



* We can see that none of the cases come close to the common threshold of 1.

1. No excessive multicollinearity: Will assess with VIF and tolerance statistics for the main effects model only because it cannot handle interactions.

# VIF and tolerance for the main effects model  
vif(res\_fac\_dom\_main\_effects\_model\_a\_neutral)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance shared\_child\_age   
 1.000209 1.002538 1.002342

1/vif(res\_fac\_dom\_main\_effects\_model\_a\_neutral)

z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance shared\_child\_age   
 0.9997907 0.9974685 0.9976631

* The VIF and tolerance statistics are well within the reasonable range.

1. Linear relationship between continuous IVs and logit of DV: Box-Tidwell Approach

* We need to create the natural log transformations of each of the continuous IVs, but because the ln of a negative number is undefined we must first transform our predictor variables to all be positive by adding a constant. The below code determines the lowest value of z\_m\_res\_facial\_dominance and z\_f\_res\_facial\_dominance so that we can add a constant that makes all values for the variable positive before making the natural log transformation.

min(neutral\_face\_dyadic\_data$z\_m\_res\_facial\_dominance)

[1] -1.842716

min(neutral\_face\_dyadic\_data$z\_f\_res\_facial\_dominance)

[1] -2.616872

* Given these minimum values, I will add 3 to each variable to make them positive, then I will make a natural log transformation to both.

# Creating c\_z\_m\_res\_facial\_dominance, which represents mothers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$c\_z\_m\_res\_facial\_dominance <- neutral\_face\_dyadic\_data$z\_m\_res\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_m\_res\_facial\_dominance, which represents the natural log of mothers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$ln\_c\_z\_m\_res\_facial\_dominance <- log(neutral\_face\_dyadic\_data$c\_z\_m\_res\_facial\_dominance)  
  
# Creating c\_z\_f\_res\_facial\_dominance, which represents fathers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$c\_z\_f\_res\_facial\_dominance <- neutral\_face\_dyadic\_data$z\_f\_res\_facial\_dominance + 3  
  
# Creating ln\_c\_z\_f\_res\_facial\_dominance, which represents the natural log of fathers' residual facial dominance after adding a constant of 3  
neutral\_face\_dyadic\_data$ln\_c\_z\_f\_res\_facial\_dominance <- log(neutral\_face\_dyadic\_data$c\_z\_f\_res\_facial\_dominance)

* Now we need to fit and summarize a model with the main effects of each variable with the constant transformation plus the interaction of those variables with their natural log transformations. The summary should indicate that neither of the interaction terms are significant, thereby satisfying the assumption of linearity of the logit.

BT\_test\_res\_facial\_dom\_model\_neutral <- glm(shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance + c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
summary(BT\_test\_res\_facial\_dom\_model\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance + c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance +   
 c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error  
(Intercept) -2.9930 3.6877  
c\_z\_m\_res\_facial\_dominance 1.3959 2.2373  
c\_z\_f\_res\_facial\_dominance 0.1359 1.9328  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -0.5843 1.0558  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 0.2251 0.9547  
 z value Pr(>|z|)  
(Intercept) -0.812 0.417  
c\_z\_m\_res\_facial\_dominance 0.624 0.533  
c\_z\_f\_res\_facial\_dominance 0.070 0.944  
c\_z\_m\_res\_facial\_dominance:ln\_c\_z\_m\_res\_facial\_dominance -0.553 0.580  
c\_z\_f\_res\_facial\_dominance:ln\_c\_z\_f\_res\_facial\_dominance 0.236 0.814  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 103.27 on 76 degrees of freedom  
AIC: 113.27  
  
Number of Fisher Scoring iterations: 4

* Neither of the interaction terms are significant, indicating that we should not reject the null hypothesis that there is a linear relationship between our continuous predictors and the logit.

##### Summary of the Models

First, here is the summary of the model with the main effects of mothers’ residual facial dominance and fathers’ residual facial dominance only, along with a Chi-square test for whether the model with the main effects fits significantly better than the model with only the intercept.

# Producing the summary of the main effects model  
summary(res\_fac\_dom\_main\_effects\_model\_a\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance +   
 shared\_child\_age, family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) 0.12378 0.50792 0.244 0.8075   
z\_m\_res\_facial\_dominance 0.15950 0.23749 0.672 0.5018   
z\_f\_res\_facial\_dominance 0.60848 0.26840 2.267 0.0234 \*  
shared\_child\_age 0.05699 0.11271 0.506 0.6131   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 103.35 on 77 degrees of freedom  
AIC: 111.35  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_main\_effects\_model\_a\_neutral <- res\_fac\_dom\_main\_effects\_model\_a\_neutral$null.deviance - res\_fac\_dom\_main\_effects\_model\_a\_neutral$deviance  
Chi\_res\_fac\_dom\_main\_effects\_model\_a\_neutral

[1] 6.845574

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_main\_effects\_model\_a\_neutral <- res\_fac\_dom\_main\_effects\_model\_a\_neutral$df.null - res\_fac\_dom\_main\_effects\_model\_a\_neutral$df.residual  
df\_res\_fac\_dom\_main\_effects\_model\_a\_neutral

[1] 3

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model\_a\_neutral <- 1 - pchisq(Chi\_res\_fac\_dom\_main\_effects\_model\_a\_neutral, df\_res\_fac\_dom\_main\_effects\_model\_a\_neutral)  
prob\_Chi\_res\_fac\_dom\_main\_effects\_model\_a\_neutral

[1] 0.07698619

* With the loss of power, the full main effects model just not significant (x2(3) = 6.846, *p* = .077). However, the predictor for fathers’ residual facial dominance remains significant (*b* = .608, *z* = 2.267, *p* = .023). I will exponentiate the coefficients to make this more interpretable.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_main\_effects\_model\_a\_neutral$coefficients)

(Intercept) z\_m\_res\_facial\_dominance z\_f\_res\_facial\_dominance   
 1.131769 1.172920 1.837641   
 shared\_child\_age   
 1.058646

* The odds ratio for father’s residual facial dominance is OR = 1.838, indicating that while holding mother’s residual facial dominance constant a one standard deviation increase in father’s residual facial dominance is associated with a 83.8% higher odds of having a first born son, while controlling for time since birth (child age).

Now we will summarize the model with the main effects and their interaction, and we will compare it’s fit to the intercept only model. After this we will compare it to the main effects model.

# Summarizing the interaction model  
summary(res\_fac\_dom\_interaction\_model\_a\_neutral)

Call:  
glm(formula = shared\_child\_sex ~ z\_m\_res\_facial\_dominance + z\_f\_res\_facial\_dominance +   
 shared\_child\_age + z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance,   
 family = binomial(link = logit), data = neutral\_face\_dyadic\_data)  
  
Coefficients:  
 Estimate Std. Error z value  
(Intercept) 0.08871 0.51215 0.173  
z\_m\_res\_facial\_dominance 0.25021 0.25687 0.974  
z\_f\_res\_facial\_dominance 0.70917 0.28899 2.454  
shared\_child\_age 0.06841 0.11392 0.600  
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.34338 0.27175 1.264  
 Pr(>|z|)   
(Intercept) 0.8625   
z\_m\_res\_facial\_dominance 0.3300   
z\_f\_res\_facial\_dominance 0.0141 \*  
shared\_child\_age 0.5482   
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance 0.2064   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 110.19 on 80 degrees of freedom  
Residual deviance: 101.64 on 76 degrees of freedom  
AIC: 111.64  
  
Number of Fisher Scoring iterations: 4

# Calculating the chi-squared statistic for the model compared to the baseline model with only the intercept  
Chi\_res\_fac\_dom\_interaction\_model\_a\_neutral <- res\_fac\_dom\_interaction\_model\_a\_neutral$null.deviance - res\_fac\_dom\_interaction\_model\_a\_neutral$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_a\_neutral

[1] 8.552133

# Calculating the degrees of freedom for the chi-square statistic comparing the model to it's baseline  
df\_res\_fac\_dom\_interaction\_model\_a\_neutral <- res\_fac\_dom\_interaction\_model\_a\_neutral$df.null - res\_fac\_dom\_interaction\_model\_a\_neutral$df.residual  
df\_res\_fac\_dom\_interaction\_model\_a\_neutral

[1] 4

# Conducting a chi-square test for the p-value using the chi-square statistic and the degrees of freedom  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_a\_neutral <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_a\_neutral, df\_res\_fac\_dom\_interaction\_model\_a\_neutral)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_a\_neutral

[1] 0.07332266

* The full interaction model is not significant (x2(4) = 8.552, *p* = .073), but, although the interaction term is not significant, when the interaction term was added to the model the coefficient became larger for fathers’ residual facial dominance (*b* = .709, *z* = 2.454, *p* = .014). Again, I will exponentiate this coefficient to understand it better.

# Exponentiating the coefficients of the model to reveal the odds ratio conversion of the coefficients  
exp(res\_fac\_dom\_interaction\_model\_a\_neutral$coefficients)

(Intercept)   
 1.092767   
 z\_m\_res\_facial\_dominance   
 1.284301   
 z\_f\_res\_facial\_dominance   
 2.032298   
 shared\_child\_age   
 1.070801   
z\_m\_res\_facial\_dominance:z\_f\_res\_facial\_dominance   
 1.409702

* The odds ratio for father’s residual facial dominance is now 2.03, indicating that while holding mother’s residual facial dominance constant a one standard deviation increase in father’s residual facial dominance is associated with a 103% higher odds of having a first born son.

Now to compare the main-effects model with the main-effects and interaction model.

# Calculating the chi-square statistic to compare the interaction model with the main effects model  
Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral <- res\_fac\_dom\_main\_effects\_model\_a\_neutral$deviance - res\_fac\_dom\_interaction\_model\_a\_neutral$deviance  
Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral

[1] 1.706559

# Calculating the degrees of freedom to compare the interaction model with the main effects model  
df\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral <- res\_fac\_dom\_main\_effects\_model\_a\_neutral$df.residual - res\_fac\_dom\_interaction\_model\_a\_neutral$df.residual  
df\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral

[1] 1

prob\_Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral <- 1 - pchisq(Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral, df\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral)  
prob\_Chi\_res\_fac\_dom\_interaction\_model\_a\_v\_main\_neutral

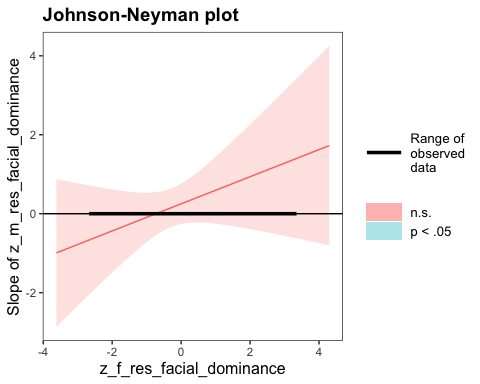
[1] 0.1914324

* The Chi-squared test does not indicate that the interaction model fits better than the main effects model (x2(1) = 1.707, *p* = .191).

Although the interaction is not significant in our model, our hypothesis predicts that at high levels of fathers’ facial dominance mothers’ dominance predicts the probability of a first born son, so I will run a simple slopes and Johnson-Neyman analysis with mothers’ residual facial dominance as the focal predictor and fathers’ residual facial dominance as the moderator.

# Conducting the simple slopes and Johnson-Neyman analysis  
sim\_slopes(res\_fac\_dom\_interaction\_model\_a\_neutral, pred = z\_m\_res\_facial\_dominance, modx = z\_f\_res\_facial\_dominance, jnplot = TRUE)

JOHNSON-NEYMAN INTERVAL   
  
The Johnson-Neyman interval could not be found. Is the p value for your  
interaction term below the specified alpha?



SIMPLE SLOPES ANALYSIS   
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = -1.000000e+00 (- 1 SD):   
  
 Est. S.E. z val. p  
------- ------ -------- ------  
 -0.09 0.31 -0.30 0.77  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = -2.193033e-17 (Mean):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.25 0.26 0.97 0.33  
  
Slope of z\_m\_res\_facial\_dominance when z\_f\_res\_facial\_dominance = 1.000000e+00 (+ 1 SD):   
  
 Est. S.E. z val. p  
------ ------ -------- ------  
 0.59 0.43 1.39 0.16

* Although nothing here is significant, I will exponentiate the coefficients here to make them more interpretable.

# Producing odds-ratios for each value of the slope for mothers' residual facial dominance for each level of the moderator in the simple slopes analysis  
exp(-.09) # -1 SD

[1] 0.9139312

exp(.25) # Mean

[1] 1.284025

exp(.59) # +1 SD

[1] 1.803988

# Taking the inverse of the odds-ratio for the first coefficient  
1 - 0.9139312

[1] 0.0860688

* The Johnson-Neyman analysis indicates that there are no values of fathers’ residual facial dominance for which mothers’ residual facial dominance is a significant predictor of the probability of having a first born child. Similar to the analysis with all facial expression types, however, the simple slopes analysis indicates that at low levels of fathers’ residual facial dominance (-1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 8.6% *decrease* in the odds of having a first born son; at the mean for fathers’ residual facial dominance a one standard deviation increase in mothers’ residual facial dominance is associated with an 28.4% *increase* in the odds of having a first born son; and at high levels of fathers’ residual facial dominance (+1 SD) a one standard deviation increase in mothers’ residual facial dominance is associated with a 80.4% *increase* in the odds of having a first born son.
  + Although not significant, the direction of this moderation effect is consistent with our hypothesis, like in the model with all facial expressions included.
  + Again, it is possible that in the population there exists such an effect, whereas we simply do not have enough power to detect it.

## Clean-Up From Study 2 Analysis

I will now write the final versions of the analysis files that I have used as csv files so that they are not lost.

# Writing the final version of the dyadic dataset as "post\_analysis\_full\_dyads.csv"  
write.csv(dyadic\_data, file = "./data/post\_analysis\_dyadic\_data\_full.csv")  
  
# Writing the final version of the female-only individual data as "post\_analysis\_individual\_data\_females\_only.csv"  
write.csv(female\_individuals, file = "./data/post\_analysis\_individual\_data\_females\_only.csv")  
  
# Writing the final version of the dyadic dataset with only neutral faces as "post\_analysis\_dyadic\_data\_neutral\_faces\_only"  
write.csv(neutral\_face\_dyadic\_data, file = "./data/post\_analysis\_dyadic\_data\_neutral\_faces\_only.csv")  
  
# Remove the objects from the environment  
rm(list = ls())

Buunk, A., & Fisher, M. (2009). Individual differences in intrasexual competition. *Journal of Evolutionary Psychology*, *7*, 37–48. <https://doi.org/10.1556/JEP.7.2009.1.5>

Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evolution and Human Behavior*, *31*(5), 334–347. <https://doi.org/10.1016/j.evolhumbehav.2010.02.004>

Forrest, S., Eatough, V., & Shevlin, M. (2005). Measuring Adult Indirect Aggression: The Development and Psychometric Assessment of the Indirect Aggression Scales. *Aggressive Behavior*, *31*(1), 84–97. <https://doi.org/10.1002/ab.20074>

Ganzeboom, H. (2010, May 1). *A new International Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007*.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>

Grant, V. J. (1990). Maternal personality and sex of infant. *The British Journal of Medical Psychology*, *63 ( Pt 3)*, 261–266. <https://doi.org/10.1111/j.2044-8341.1990.tb01618.x>

Grant, V. J. (1992). The measurement of dominance in pregnant women by use of the Simple Adjective Test. *Personality and Individual Differences*, *13*(1), 99–102. <https://doi.org/10.1016/0191-8869(92)90225-E>

International Labor Office. (2012). *International Standard Classification of Occupations: Structure, group definitions and correspondence tables*. International Labor Organization. <https://www.ilo.org/public/english/bureau/stat/isco/docs/publication08.pdf>

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, *1*, 57–93.

Long, J. A. (2021). *interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*. <https://cran.r-project.org/web/packages/interactions/index.html>

Palmer-Hague, J. L., & Watson, N. V. (2016). Effects of Mother and Father Dominance on Offspring Sex in Contemporary Humans. *Adaptive Human Behavior and Physiology*, *2*(1), 57–76. <https://doi.org/10.1007/s40750-015-0032-6>

Revelle, W., & Revelle, M. W. (2015). Package “psych.” *The Comprehensive R Archive Network*, *337*(338). <http://mirror.ibcp.fr/pub/CRAN/web/packages/psych/psych.pdf>

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji - an Open Source platform for biological image analysis. *Nature Methods*, *9*(7), 10.1038/nmeth.2019. <https://doi.org/10.1038/nmeth.2019>

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420. <https://psycnet.apa.org/record/1979-25169-001>