Homework 4

Tania

2023-10-24

```r
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(moderndive)
library(plotly)
```

```
##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout
```

```r
library(rsm)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(rockchalk)
```

```
##
## Attaching package: 'rockchalk'

## The following object is masked from 'package:dplyr':
##
##     summarize

## The following object is masked from 'package:rsm':
##
##     model.data
```

```r
height <- read.csv("galton_height.csv")
height$Midparent <- (height$Father + (1.08*height$Mother))/2
```

Part I: Regression of child's height on mid-height of parent and gender

1. Have a scatterplot of child's height vs. mid height of parents, with different colors for male and female of the child. Add smooth lines for male and female separately in the scatterplot. First, plot the smooth lines with same slope. Secondly with different slopes. By looking at the plots, does it look like that the slopes are different even if you didn't force them to be parallel?

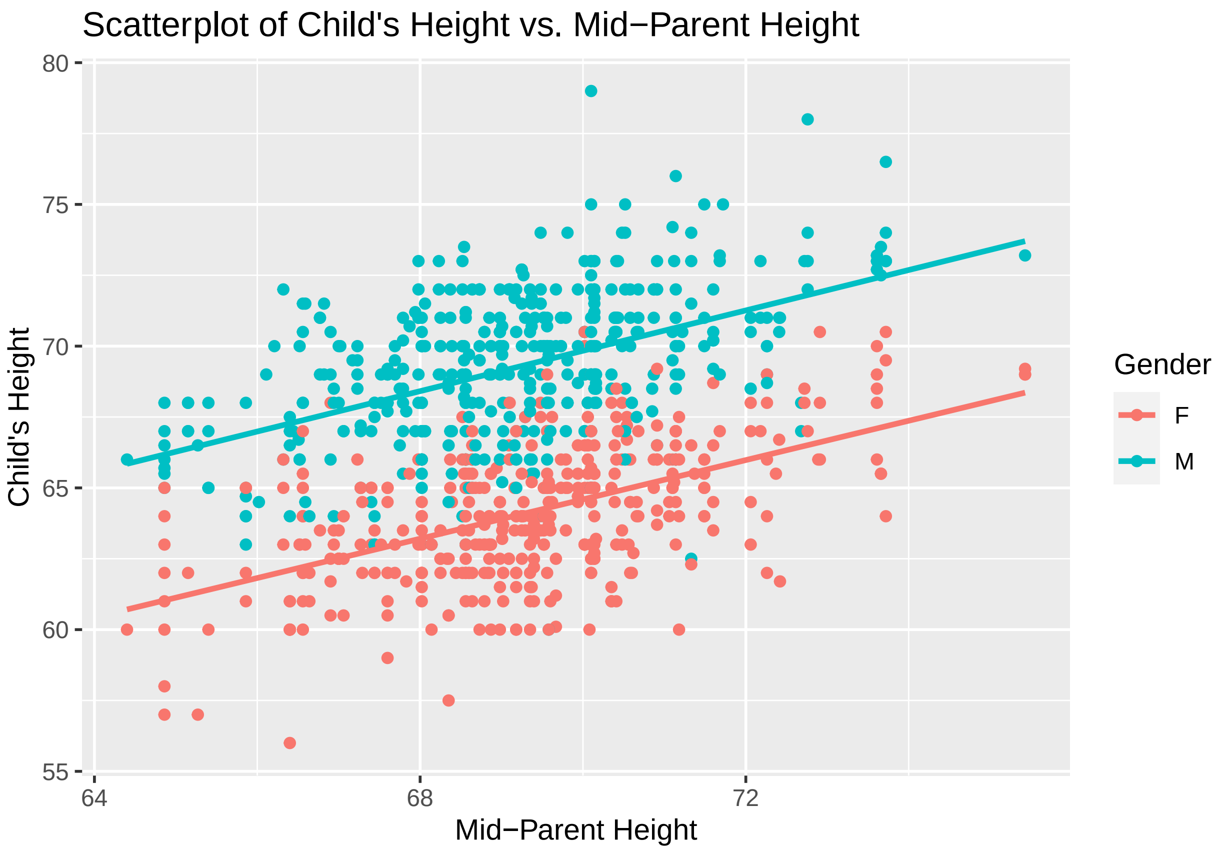
```r
#same slope
plot1 <- ggplot(data = height, aes(x = Midparent, y = Height, color = Gender)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Scatterplot of Child's Height vs. Mid-Parent Height",
    x = "Mid-Parent Height",
    y = "Child's Height"
  )

plot2 <- ggplot(data = height, aes(x = Midparent, y = Height, color = Gender)) +
  geom_point() + geom_parallel_slopes(method="lm", se=FALSE) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Scatterplot of Child's Height vs. Mid-Parent Height (parallel slope)",
    x = "Mid-Parent Height",
    y = "Child's Height"
  )
```

```
## Warning: `geom_parallel_slopes()` doesn't need a `method` argument ("lm" is
## used).
```

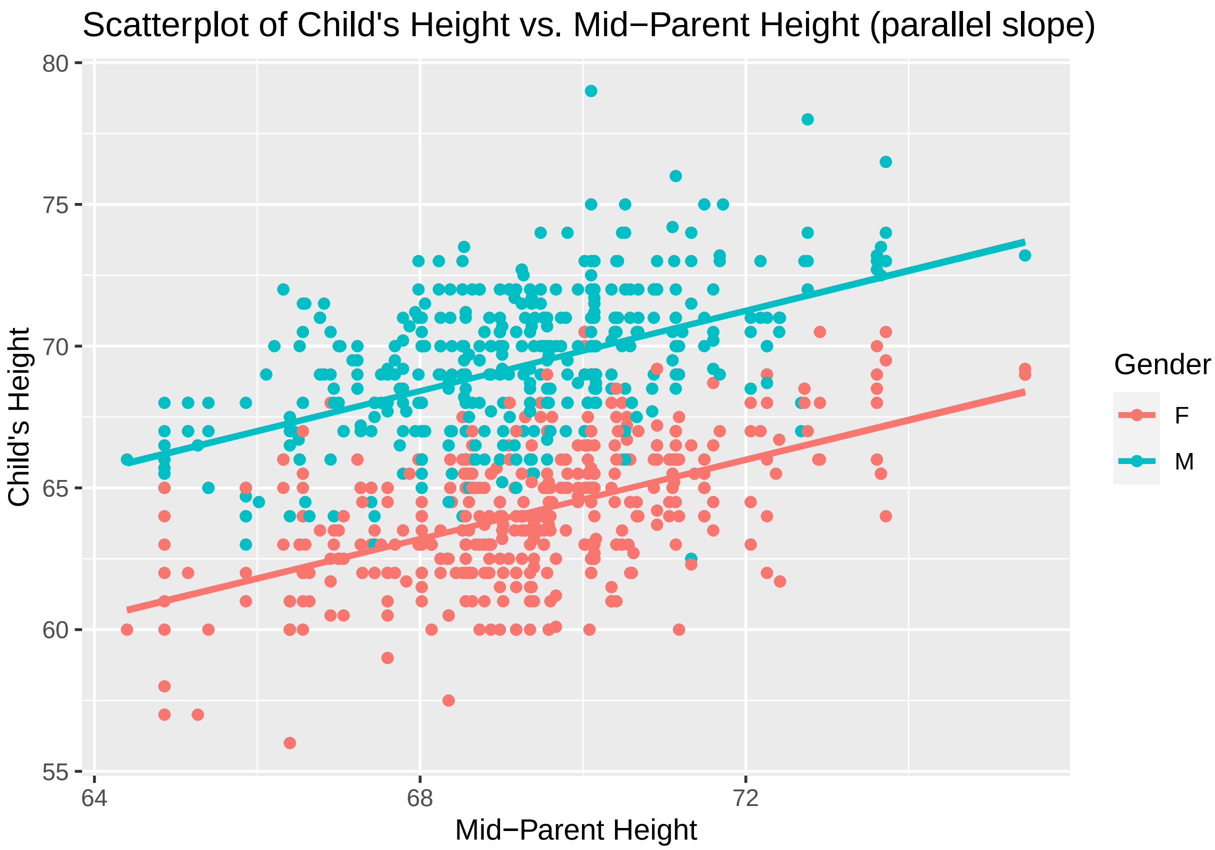
```r
plot1
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
plot2
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



The two plots appear to be the same, even though the slopes were forced to be parallel in the second one. Both these plots are the same, as the slopes for genders male and female form parallel lines on the plots. This shows that the impact of parents' mid-height on both female and male children is similar. The only difference in the smooth lines would be that the line for male children is higher than the line for females. This is due to the fact that, within the data, male children's average height is greater than that of females, so the male points and line is above the females. However, the slope, which shows the increase on children's height affected by parents' mid-height, is the same as they are parallel.

2. Run a regression with parallel slopes. Are both independent variables significant?

```r
#parallel slopes
model1 <- lm(Height ~ Midparent + Gender, data = height)
get_regression_table(model1)
```

```
## # A tibble: 3 x 7
##   term      estimate std_error statistic p_value lower_ci upper_ci
##   <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept   15.4       2.76       5.59       0   10.0     20.8
## 2 Midparent    0.703     0.04      17.7        0    0.625    0.781
## 3 Gender: M    5.23      0.144     36.2        0    4.94     5.51
```

```r
summary(model1)
```

```
##
## Call:
## lm(formula = Height ~ Midparent + Gender, data = height)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -9.5332 -1.4406  0.1003  1.4383  9.1013
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.40301    2.75523    5.59 3.01e-08 ***
## Midparent    0.70282    0.03973   17.69  < 2e-16 ***
## GenderM      5.22817    0.14444   36.20  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.161 on 895 degrees of freedom
## Multiple R-squared:  0.6371, Adjusted R-squared:  0.6363
## F-statistic: 785.5 on 2 and 895 DF,  p-value: < 2.2e-16
```

The p-value for the independent variable Midparent is 0, so it is significant, as the p-value is less than the default significance level of 0.05. The p-value for the indepedent variable Gender is also 0, which is less than 0.05, so it is statistically significant.

3. In the linear model in Q2, what is the expected increase in child's height if the mid-height of parents increases by 1 inch? Does this depend on the gender of the child?

The expected increase in child's height if the mid-height of parents increases by 1 inch is about 0.70282 inches, found in the linear model coefficients for the estimate value next to Midparent. This expected increase does not depend on the gender of the child as the coefficient for "Midparent" is the same for both male and female children (no gender effect).

4. Run a regression with interaction term. Is the interaction term significant?

```r
model2 <- lm(Height ~ Midparent * Gender, data = height)
get_regression_table(model2)
```

```
## # A tibble: 4 x 7
##   term              estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept           16.1       3.92      4.1     0         8.37    23.7
## 2 Midparent            0.693     0.056    12.3     0         0.582    0.804
## 3 Gender: M            3.93      5.50      0.714   0.475    -6.87    14.7
## 4 Midparent:GenderM    0.019     0.08      0.235   0.814    -0.137    0.175
```

```r
summary(model2)
```

```
##
## Call:
## lm(formula = Height ~ Midparent * Gender, data = height)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5372 -1.4502  0.0925  1.4435  9.0925
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       16.05747    3.91660   4.100 4.51e-05 ***
## Midparent          0.69337    0.05649  12.273  < 2e-16 ***
## GenderM            3.93353    5.50550   0.714    0.475
## Midparent:GenderM  0.01870    0.07950   0.235    0.814
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.162 on 894 degrees of freedom
## Multiple R-squared:  0.6371, Adjusted R-squared:  0.6359
## F-statistic: 523.1 on 3 and 894 DF,  p-value: < 2.2e-16
```

The interaction term "Midparent:GenderM" is not significant, as it has a p-value of 0.814 which is greater than the significance level of 0.05. Therefore, the interaction between Midparent and Gender is not statistically significant in this model, suggesting that the relationship between parents' mid-height and child's height does not differ between male and female children.

5. In the linear model in Q4, what is the expected increase in adult daughter's height if the mid-height of parent increases by 1 inch? What is the expected increase in adult son's height if the mid-height of parent increases by 1 inch? Is there significant difference between these two values? Why?

The expected increase in adult daughter's height if the mid-height of parents increases by 1 inch is 0.69337 inches (determined by coefficients for "Midparent"). The expected increase in adult son's height if the mid-height of parent increases by 1 inch is $0.69337 + 0.01870 = 0.71207$. inches (determined by the sum of the coefficients for "Midparent" and "Midparent:GenderM"). From these observed values, we can see that since p-value is greater than 0.05, there a significant difference (0.0187 inches) which could be because of random variation in the data set.

6. Run two regressions for child's height on mid-height of parents, respectively for adult daughters and sons. Compare the two slopes from the two regressions to the slopes from the interaction model.

```r
model_daughters <- lm(Height ~ Midparent, data = subset(height, Gender == "F"))
summary(model_daughters)
```

```
##
## Call:
## lm(formula = Height ~ Midparent, data = subset(height, Gender ==
##     "F"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0975 -1.3864 -0.0251  1.3771  5.8925
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.05747    3.62206   4.433 1.18e-05 ***
## Midparent    0.69337    0.05225  13.271  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.999 on 431 degrees of freedom
## Multiple R-squared:  0.2901, Adjusted R-squared:  0.2885
## F-statistic: 176.1 on 1 and 431 DF,  p-value: < 2.2e-16
```

```r
get_regression_table(model_daughters)
```

```
## # A tibble: 2 x 7
##   term      estimate std_error statistic p_value lower_ci upper_ci
##   <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept   16.1       3.62       4.43       0    8.94    23.2
## 2 Midparent    0.693     0.052     13.3        0    0.591    0.796
```

```r
model_sons <- lm(Height ~ Midparent, data = subset(height, Gender == "M"))
summary(model_sons)
```

```
##
## Call:
## lm(formula = Height ~ Midparent, data = subset(height, Gender ==
##     "M"))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5372 -1.5230  0.1891  1.5157  9.0925
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.99100    4.12165    4.85 1.69e-06 ***
## Midparent    0.71208    0.05959   11.95  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.303 on 463 degrees of freedom
## Multiple R-squared:  0.2357, Adjusted R-squared:  0.2341
## F-statistic: 142.8 on 1 and 463 DF,  p-value: < 2.2e-16
```

```
get_regression_table(model_sons)
```

```
## # A tibble: 2 x 7
##   term      estimate std_error statistic p_value lower_ci upper_ci
##   <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept   20.0        4.12      4.85       0   11.9     28.1
## 2 Midparent    0.712      0.06     12.0        0    0.595    0.829
```

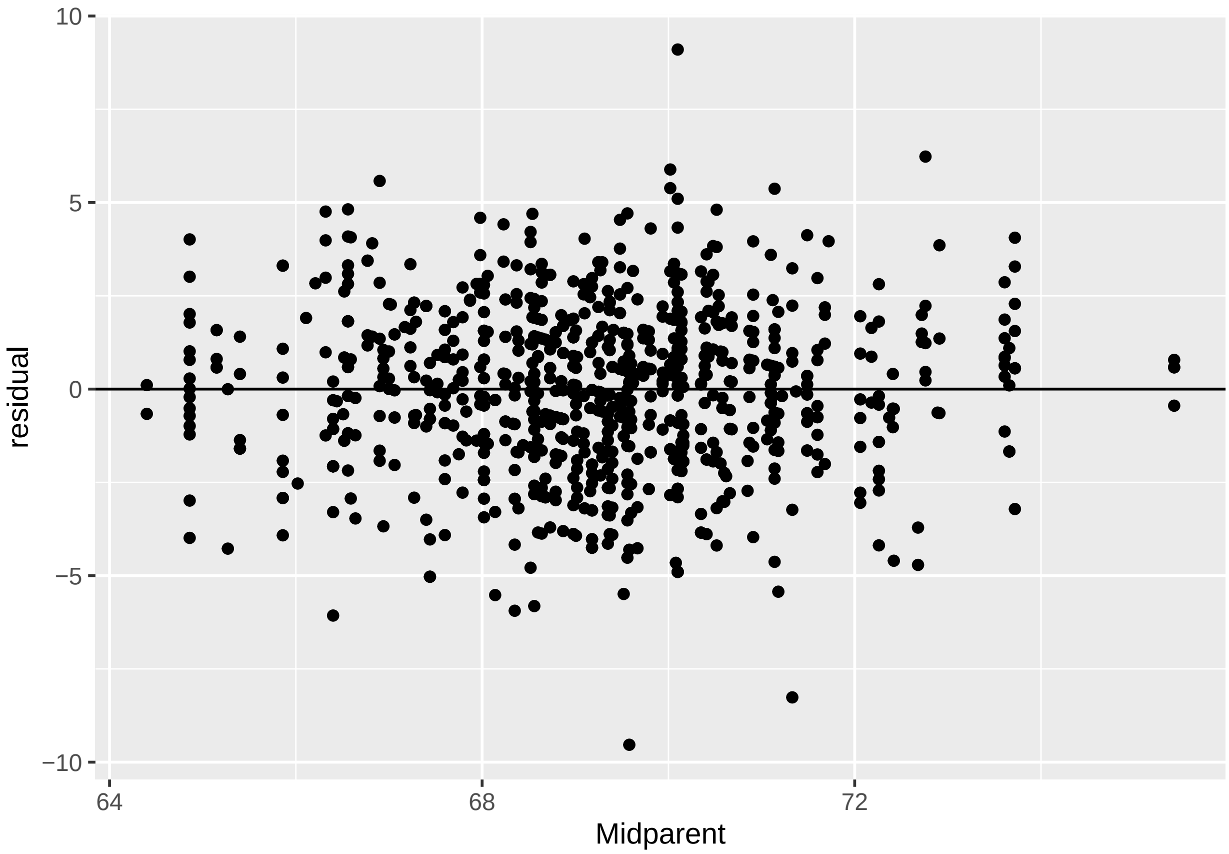
In the regression model for child's height on mid-height of parents for adult daughters, the slope for "Midparent" is about 0.69337. In the regression model for child's height on mid-height of parents for adult sons, the slope for "Midparent" is about 0.71208. In the regression model with the interaction term, the slope for "Midparent" in the interaction model is about 0.69337. The slopes in the daughters' model and the interaction model are the same, while the slope in the sons model is the same as the slope in the interaction model (expected increase in son's height). The slope for adult sons is slightly larger. This shows that the relationship between the mid-height of parents and child's height is very similar for daughters in both models (daughters and interaction), but there is a slight difference for sons between the seprate and interaction models.

7. If you are going to choose the final model between Q2 and Q4, which one will you choose? Why?

I would choose the model from Q2 as my final model. The Q2 model explains the main effects of Midparent height and gender on child height, which is statistically significant. You can directly interpret the coefficients for Midparent height and Gender, and their significance. In the model from Q4, the relationship between parents' mid-height and child's height does not differ significantly between male and female children, as the interaction between Midparent and Gender is not statistically significant. This suggests that gender might not be a crucial factor in explaining the variation in child height. Therefore, I picked the model from Q2.

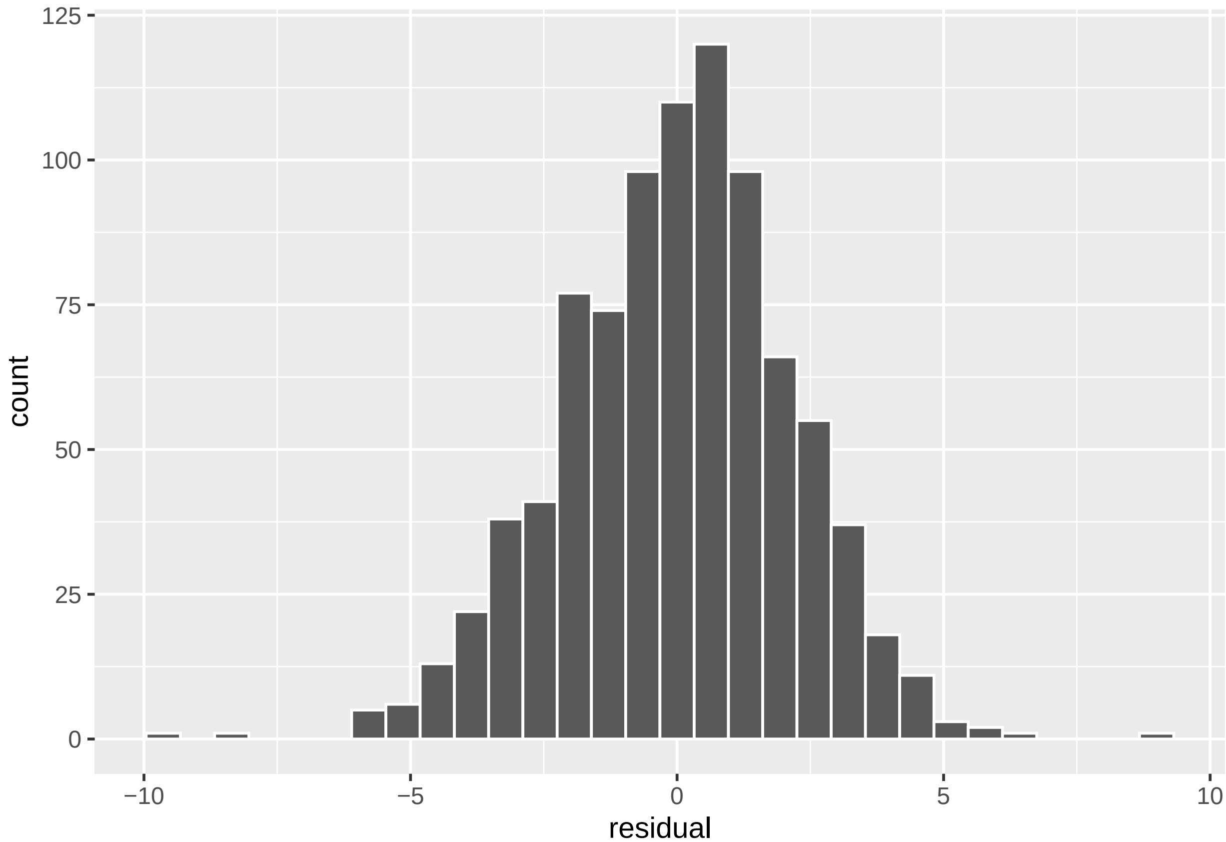
8. Have the scatterplot, histogram and QQ plot of the residual from the model you choose in Q7, and comment on the assumptions of the multiple regression. What is the R^2 from the model?

```
points <- get_regression_points(model1)
ggplot(data = points, aes(x = Midparent, y = residual)) + geom_point()+ geom_hline(yintercept = 0)
```

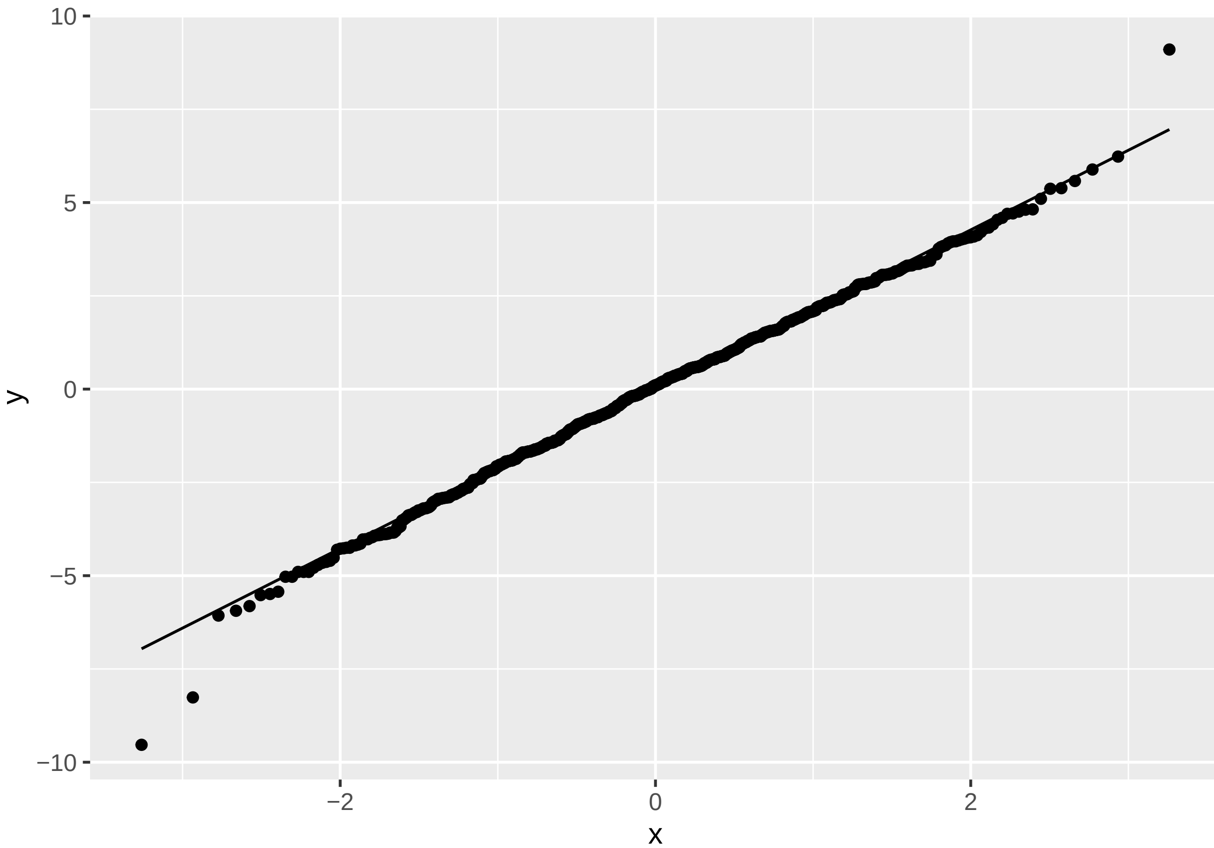



```
ggplot(data = points,aes(x = residual))+ geom_histogram( color = "white")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
ggplot(data = points,aes(sample=residual))+stat_qq()+ stat_qq_line()
```




```
cat("R-squared from model: " , summary(model1)$r.squared)
```

```
## R-squared from model:  0.637071
```

Linearity: The scatterplot of the residuals shows a slight pattern, with some of the points plotted directly in a line. However, for the most part the points are placed randomly and somewhat evenly scattered around zero. Normality: The histogram of the residuals approximates a normal distribution. It is mostly symmetric and creates a bell curve shape. The QQ plot of the residuals has points plotted following a mostly straight line, approximating a normal distribution. Homoscedasticity: The scatterplot of the residuals has relatively consistent spread throughout the range of fitted values. Independence: The residuals are mostly independent of each other. The points forming a straight line in the QQ plot and the points being mostly randomly scattered in the scatterplot verify this assumption.

Part II: Regression of child's height on father and mother's height

```
height$Height[height$Gender == "F"] <- height$Height[height$Gender == "F"] * 1.08
```

9. Run a regression without interaction term. Are both independent variables significant?

```
model_no_interaction <- lm(Height ~ Father + Mother, data = height)
summary(model_no_interaction)
```

```
##
## Call:
## lm(formula = Height ~ Father + Mother, data = height)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -9.4845 -1.5119  0.1268  1.5175  9.1382
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.70154    2.83140   6.605 6.81e-11 ***
## Father       0.42288    0.03017  14.017  < 2e-16 ***
## Mother       0.33167    0.03230  10.267  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.226 on 895 degrees of freedom
## Multiple R-squared:  0.2663, Adjusted R-squared:  0.2647
## F-statistic: 162.4 on 2 and 895 DF,  p-value: < 2.2e-16
```

```
get_regression_table(model_no_interaction)
```

```
## # A tibble: 3 x 7
##   term      estimate std_error statistic p_value lower_ci upper_ci
##   <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept   18.7       2.83       6.60       0   13.1     24.3
## 2 Father       0.423     0.03      14.0        0    0.364    0.482
## 3 Mother       0.332     0.032     10.3        0    0.268    0.395
```

Yes, both indepedent variables are significant. The p-value for "Father" and "Mother" are both 0 which is less than the default significance level of 0.05, so both variables are significant. The coefficient for "Father" is about 0.42288, with a very low standard error, and a highly significant t-value of 14.017. This indicates that the height of the father is a highly significant predictor of the child's height. For every unit increase in the father's height, we expect an increase of approximately 0.42288 units in the child's height. The coefficient for "Mother" is about 0.30710, with a very low standard error, and a highly significant t-value of 10.267. This indicates that the height of the mother is also a highly significant predictor of the child's height. For every unit increase in the mother's height, we expect an increase of approximately 0.30710 units in the child's height.

10. In the linear model in Q9, what is the expected increase in child's height if the mother's height increases by 1 inch, while father's height is 68 inches? While father's height is 70 inches? While the father's height is 72 inches? Does the increment depend on the height of the father?

```
coefs <- coef(model_no_interaction)

change_in_mother_height <- 1
father_heights <- c(68, 70, 72)

# expected increases in child's height for each father's height
expected_increases <- coefs["Mother"] * change_in_mother_height

results <- data.frame(Father_Height = father_heights, Expected_Increase = expected_increases)
```

```
## Warning in data.frame(Father_Height = father_heights, Expected_Increase =
## expected_increases): row names were found from a short variable and have been
## discarded
```

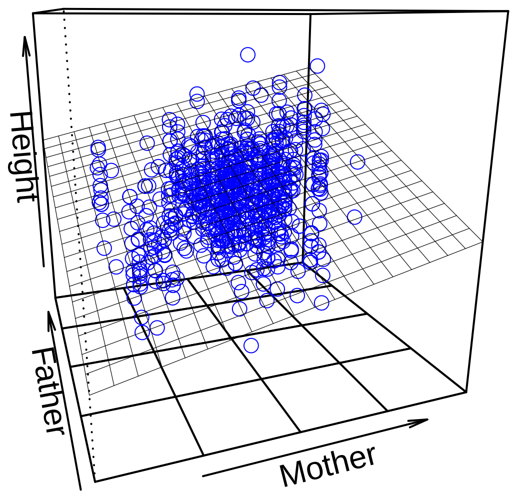
```
print(results)
```

```
##   Father_Height Expected_Increase
## 1            68         0.3316683
## 2            70         0.3316683
## 3            72         0.3316683
```

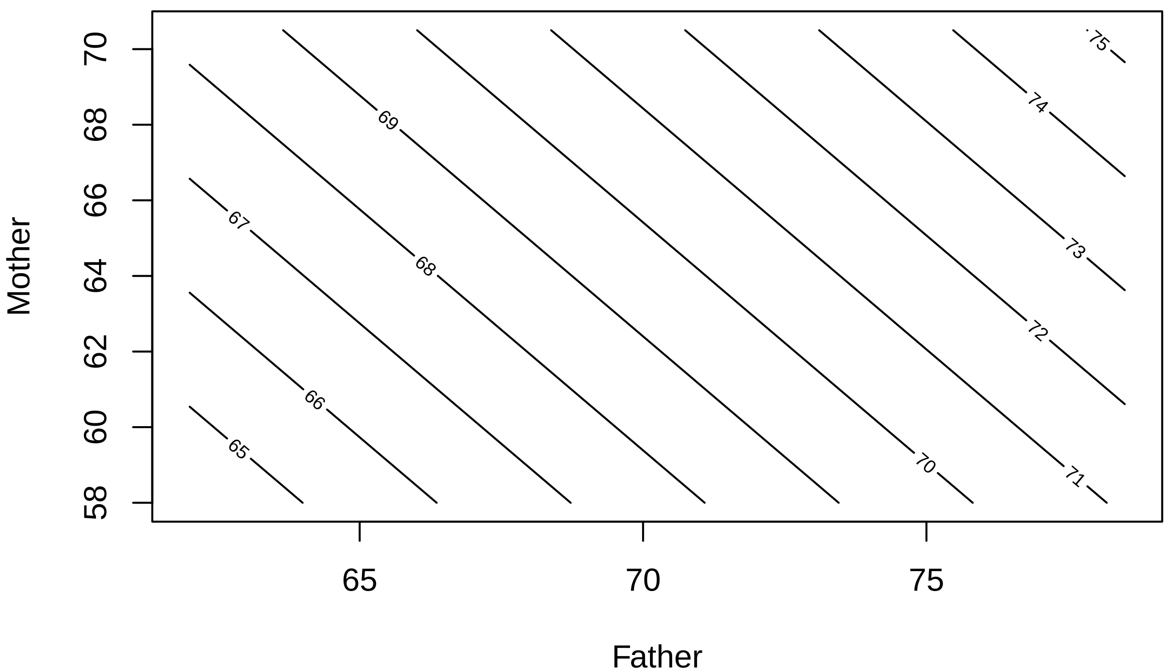
The expected increase in the child's height when the mother's height increases by 1 inch does not depend on the father's height in this model. The coefficient for "Mother" represents a constant effect, and it is the same regardless of the father's height. The increment in the child's height is the same for different father's heights when the mother's height increases by 1 inch.

11. Get the regression plane and contour plot from the model in Q9.

```
plotPlane(model_no_interaction,plotx1 = "Mother", plotx2 = "Father")
```




```
contour(model_no_interaction,Mother~Father)
```



12. Run a regression with interaction term. Is the interaction term significant?

```
model_with_interaction <- lm(Height ~ Father * Mother, data = height)
summary(model_with_interaction)
```

```
##
## Call:
## lm(formula = Height ~ Father * Mother, data = height)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.4791 -1.5025  0.1189  1.5168  9.1381
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   76.42313   54.72131   1.397    0.163
## Father        -0.40904    0.78820  -0.519    0.604
## Mother        -0.56669    0.85113  -0.666    0.506
## Father:Mother  0.01295    0.01226   1.056    0.291
##
## Residual standard error: 2.226 on 894 degrees of freedom
## Multiple R-squared:  0.2672, Adjusted R-squared:  0.2648
## F-statistic: 108.7 on 3 and 894 DF,  p-value: < 2.2e-16
```

```
get_regression_table(model_with_interaction)
```

```
## # A tibble: 4 x 7
##   term          estimate std_error statistic p_value lower_ci upper_ci
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept       76.4      54.7        1.40   0.163  -31.0    184.
## 2 Father          -0.409     0.788    -0.519   0.604   -1.96     1.14
## 3 Mother          -0.567     0.851    -0.666   0.506   -2.24     1.10
## 4 Father:Mother    0.013     0.012     1.06    0.291   -0.011    0.037
```

The interaction term "Father:Mother" has a p-value of 0.291, which is greater than the signiifcance level of 0.05, so the intraction term is not significant. Therefore, the interaction between the father's height and mother's height does not have a significant effect on the child's height in this regression model.

13. In the linear model in Q12, what is the expected increase in child's height if the mother's height increases by 1 inch, while father's height is 68 inches? While father's height is 70 inches? While the father's height is 72 inches? Are there any differences between these increments? Are the differences significant?

```
coefs_interaction <- coef(model_with_interaction)
#Expected Increase = B_mother + B_interaction * Father
#Expected Increase = -0.567 + ( 0.013 * Fathers height)

Expectedincrease1 = (-0.567 + (0.013 * 68))
cat("Expected increase when father's height is 68 inches: " , Expectedincrease1, "\n")
```

```
## Expected increase when father's height is 68 inches:  0.317
```

```
Expectedincrease2 = (-0.567 + (0.013 * 70))
cat("Expected increase when father's height is 70 inches: " , Expectedincrease2, "\n")
```

```
## Expected increase when father's height is 70 inches:  0.343
```

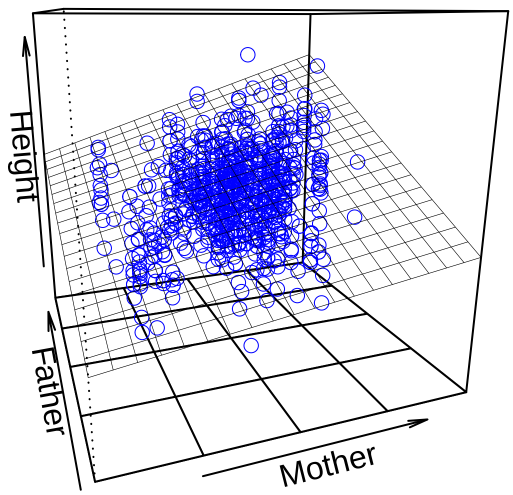
```
Expectedincrease3 = (-0.567 + (0.013 * 72))
cat("Expected increase when father's height is 72 inches: " , Expectedincrease3)
```

```
## Expected increase when father's height is 72 inches:  0.369
```

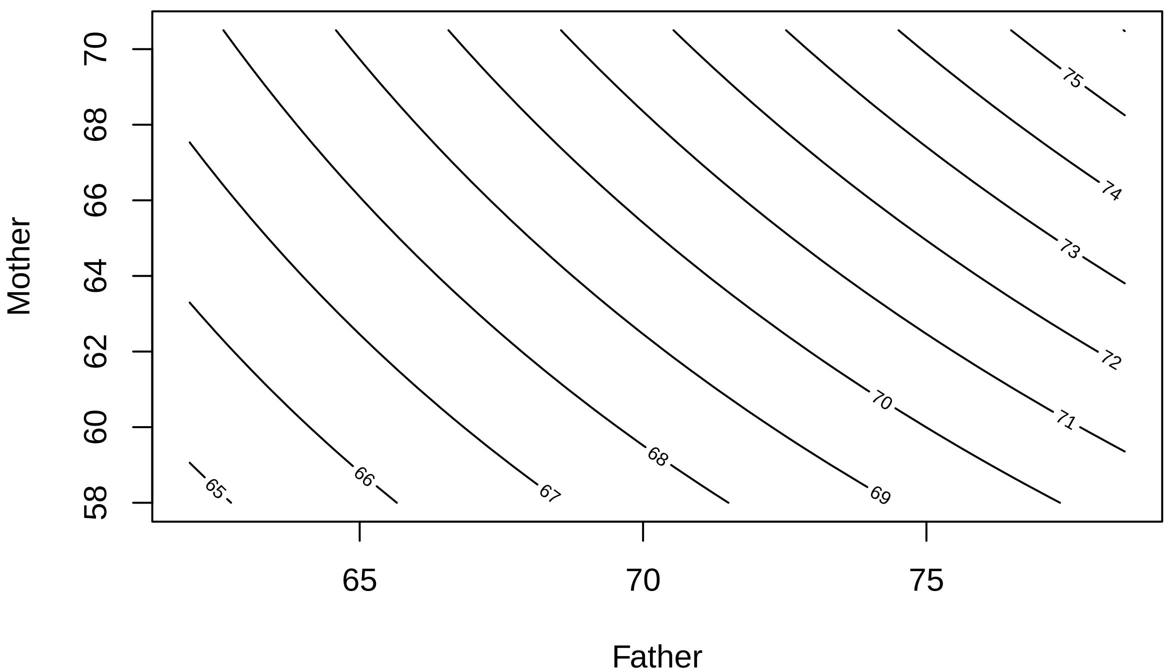
The expected increase in child's height if the mothers height increases by 1 inch while the father's height is 68 inches is 0.317 inches. The expected increase in child's height if the mothers height increases by 1 inch while the father's height is 70 inches is 0.343 inches. The expected increase in child's height if the mothers height increases by 1 inch while the father's height is 72 inches is 0.369 inches. The difference between each of these increments is 0.026 inches. These increments suggest that the effect of mother's height on child's height is consistent regardless of father's height. There is no significant difference between these increments, as the interaction term "Father:Mother" is not significant. This indicates that the interaction between father's and mother's heights doesn't significantly impact the child's height, and the effect of mother's height is similar across different values of father's height.

14. Get the regression plane and contour plot from the model in Q12.

```
plotPlane(model_with_interaction,plotx1 = "Mother", plotx2 = "Father")
```




```
contour(model_with_interaction,Mother~Father)
```



15. If you are going to choose a model between Q9 and Q12, which one will you choose? Why?

The model from Q9 is regression without interaction while the model from Q12 is regression with interaction. I would choose the regression model without interaction from Q9. This model accounts for the individual effects of Mother and Father's heights on the child's height, where both of these variables are significant. I would not choose the model in Q12, as the interaction term is not significant, suggesting that the interaction between Father's and Mother's heights doesn't significantly impact the child's height.

Part III: Regression of child’s height on father and mother’s height and gender

```
height$Height[height$Gender == "F"] <- height$Height[height$Gender == "F"] / 1.08
```

16. Run the model without any interaction. Are all independent variables significant? What are the coefficients for each variable?

```
model3 <- lm(Height ~ Mother + Father + Gender, data = height)
summary(model3)
```

```
##
## Call:
## lm(formula = Height ~ Mother + Father + Gender, data = height)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.523 -1.440  0.117  1.473  9.114
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.34476    2.74696   5.586 3.08e-08 ***
## Mother       0.32150    0.03128  10.277  < 2e-16 ***
## Father       0.40598    0.02921  13.900  < 2e-16 ***
## GenderM      5.22595    0.14401  36.289  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.154 on 894 degrees of freedom
## Multiple R-squared:  0.6397, Adjusted R-squared:  0.6385
## F-statistic:   529 on 3 and 894 DF,  p-value: < 2.2e-16
```

```
get_regression_table(model3)
```

```
## # A tibble: 4 x 7
##   term      estimate std_error statistic p_value lower_ci upper_ci
##   <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept   15.3       2.75       5.59       0    9.95    20.7
## 2 Mother       0.321     0.031     10.3        0    0.26     0.383
## 3 Father       0.406     0.029     13.9        0    0.349    0.463
## 4 Gender: M    5.23      0.144     36.3        0    4.94     5.51
```

Yes, all the indepedent variables (“Mother”, “Father”, “Gender:M”) are statistically significant, as they all have p-values of 0 which is less than the significance level of 0.05. The coefficients represent the estimated effect of each independent variable on the child’s height. The coefficient for “Mother” (mother’s raw height) is 0.32150. The coefficient for “Father” (father’s raw height) is 0.40598. The coefficient for “GenderM” is 5.22595.

17. From the final model you got from Q7, replacing mid parent height by the formula (Father + Mother * 1.08)/2, you can also get the coefficients for Father’s height, Mother’s height, and Gender. Compare these 3 coefficients to those from Q16. Are they the same? Why?

```
#b1 = 0.703
#formula B1/2 * fathers + B1 * mothers *1.08/2

height$new_mother = 0.703 * height$Mother *1.08/2
```

```r
height$new_father = 0.703/2 * height$Father

final_model <- lm(Height ~ new_father + new_mother + Gender, data = height)
get_regression_table(final_model)
```

```
## # A tibble: 4 x 7
##   term       estimate std_error statistic p_value lower_ci upper_ci
##   <chr>         <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept    15.3       2.75       5.59       0    9.95     20.7
## 2 new_father    1.16      0.083     13.9        0    0.992     1.32
## 3 new_mother    0.847     0.082     10.3        0    0.685     1.01
## 4 Gender: M     5.23      0.144     36.3        0    4.94      5.51
```

```r
summary(final_model)
```

```
##
## Call:
## lm(formula = Height ~ new_father + new_mother + Gender, data = height)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.523 -1.440  0.117  1.473  9.114
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.34476    2.74696   5.586 3.08e-08 ***
## new_father   1.15499    0.08309  13.900  < 2e-16 ***
## new_mother   0.84689    0.08240  10.277  < 2e-16 ***
## GenderM      5.22595    0.14401  36.289  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.154 on 894 degrees of freedom
## Multiple R-squared:  0.6397, Adjusted R-squared:  0.6385
## F-statistic:   529 on 3 and 894 DF,  p-value: < 2.2e-16
```

The coefficient for "Mother" is 0.84689 The coefficient for "Father" is 1.15499 The coefficient for "GenderM" is 5.22595. The coefficients for "Mother" and "Father" are not the same because we used the newly calculated variables "new_mother" and "new_father". The coefficient for "GenderM' is the same in this new model when compared to its value in Q16, as the Gender variable was not changed in the new model. Calculations: The alpha1 in the new model is same as beta1/2* for father. alpha 2 is beta1*1.08*Mother/2. alpha3 remains the same which is the Gender male interaction term. The beta1 is derived from the model we created in Q2 which is the final model I chose in Q7.