Week 5 Workbook

Odelia

3/24/2021

Import data

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(patchwork)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(sjPlot)
library(ggplot2)
library(gtsummary)
nz_0 <- readr::read_csv2(url("https://raw.githubusercontent.com/go-bayes/psych-447/main/data/nz/nz.csv"
```

```
## i Using ',' as decimal and '.' as grouping mark. Use `read_delim()` for more control.
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   Male = col_character(),
##   BigDoms = col_character(),
##   GenCohort = col_character(),
##   Religious = col_character(),
##   Believe.God = col_character(),
##   Believe.Spirit = col_character(),
##   Env.SacMade = col_logical(),
##   FeelHopeless = col_character(),
##   FeelDepressed = col_character(),
##   FeelRestless = col_character(),
##   EverythingIsEffort = col_character(),
##   FeelWorthless = col_character(),
##   FeelNervous = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```
f<-c("None Of The Time","A Little Of The Time","Some Of The Time",  "Most Of The Time", "All Of The Tim
nz <- nz_0 %>%
  dplyr::mutate_if(is.character, factor) %>%
  select(
    -c(
      SWB.Kessler01,
      SWB.Kessler02,
      SWB.Kessler03,
      SWB.Kessler04,
      SWB.Kessler05,
      SWB.Kessler06
    )
  ) %>%
  dplyr::mutate(Wave = as.factor(Wave)) %>%
  mutate(FeelHopeless = forcats::fct_relevel(FeelHopeless, f)) %>%
  mutate(FeelDepressed = forcats::fct_relevel(FeelDepressed, f)) %>%
  mutate(FeelRestless = forcats::fct_relevel(FeelRestless, f)) %>%
  mutate(EverythingIsEffort = forcats::fct_relevel(EverythingIsEffort, f)) %>%
  mutate(FeelWorthless = forcats::fct_relevel(FeelWorthless, f)) %>%
  mutate(FeelNervous = forcats::fct_relevel(FeelNervous, f)) %>%
  dplyr::mutate(Wave = as.factor(Wave)) %>%
  dplyr::mutate(male_id = as.factor(Male)) %>%
  dplyr::mutate(date = make_date(year = 2009, month = 6, day = 30) + TSCORE)
md_df <- data.frame(read.table(url("https://raw.githubusercontent.com/avehtari/ROS-Examples/master/Pears
md_df <- md_df %>%
  dplyr::mutate(mother_height_c = as.numeric(scale(mother_height, center = TRUE, scale = FALSE)))
dplyr::glimpse(md_df)
```

```
## Rows: 5,524
## Columns: 3
## $ daughter_height <dbl> 52.5, 52.5, 53.5, 53.5, 55.5, 55.5, 55.5, 55.5, 56.5, ~
```

```
## $ mother_height   <dbl> 59.5, 59.5, 59.5, 59.5, 59.5, 59.5, 59.5, 59.5, 58.5, ~
## $ mother_height_c <dbl> -2.998732802, -2.998732802, -2.998732802, -2.998732802~
```

Question 1: Create a descriptive table and a descriptive graph for the HLTH.Weight and HLTH.Height variables in the nz dataset

```
library(table1)
```

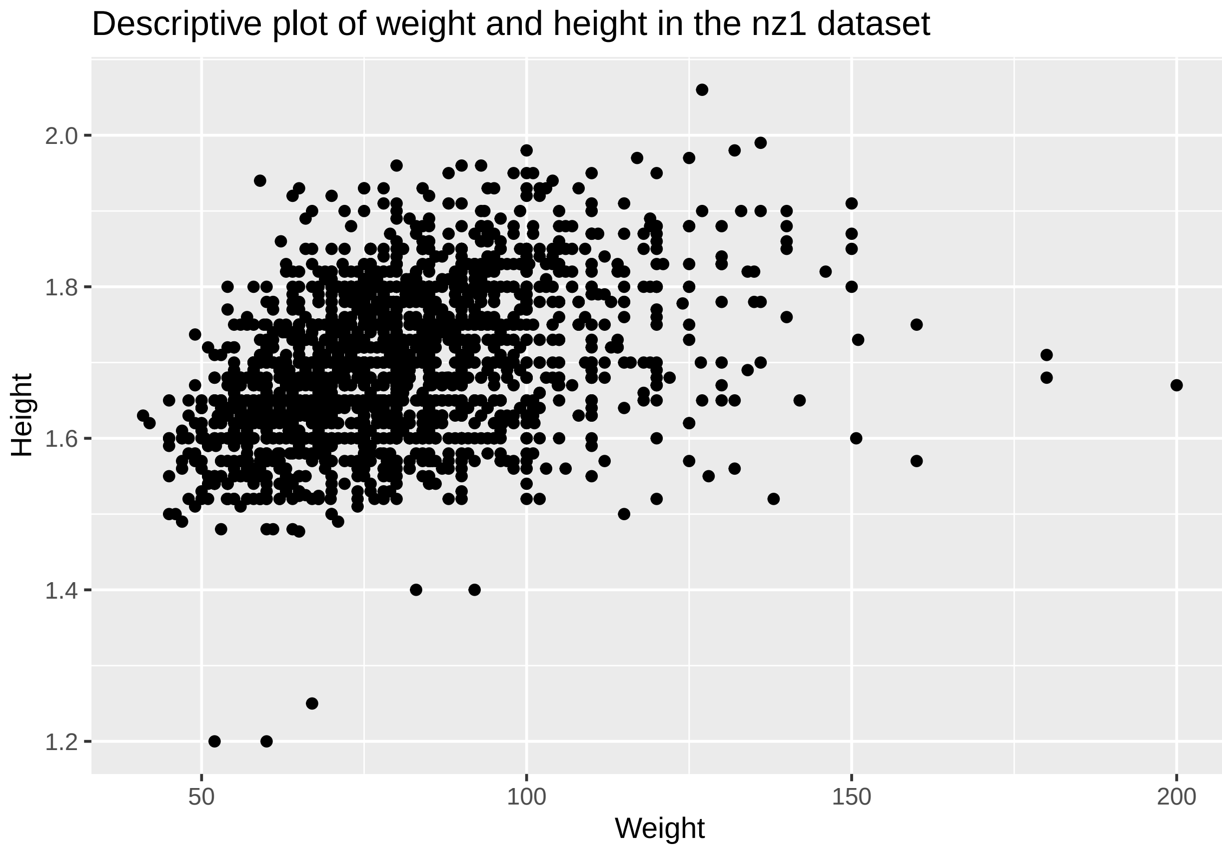
```
##
## Attaching package: 'table1'

## The following objects are masked from 'package:base':
##
##     units, units<-
```

```
nz1<-nz%>%
 dplyr::filter(Wave==2019) #to filter only 2019 data in the Wave column so that only 2019 data will be
label(nz1$HLTH.Weight) <-"Weight" #to label the Weight data so that we know what the values represent
label(nz1$HLTH.Height) <-"Height" #to label the Height data
table1(~HLTH.Weight+HLTH.Height, data=nz1) #to create table of descriptive stats for weight and height
```

```
## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th clas
```

```
ggplot(data = nz1) + #to tell R that we want to plot a graph using the nz1 dataset
    geom_point(mapping = aes(x = HLTH.Weight, y = HLTH.Height)) + #to add the layer of a scatterplot an
    labs(title = "Descriptive plot of weight and height in the nz1 dataset") +
    xlab("Weight") +
    ylab("Height") #labelling the title and axes so that we know what the graph is about
```

```
## Warning: Removed 32 rows containing missing values (geom_point).
```



Question 2: Write up a sample summary of the HLTH.Weight and HLTH.Height variables in the nz dataset in APA style

According to the analyses, the mean weight and standard deviation of the nz dataset for the sample in 2019 are 79.7 and 18.8 respectively. The median weight of the sample is 78.0, and the minimum and maximum values are 41.0 and 200. The mean height and standard deviation of the nz dataset for the sample in 2019 are 1.70 and 0.0984 respectively. The median height of the sample is 1.69, and the minimum and maximum values are 1.20 and 2.06.

Sample summary: Data of the weights and heights of a sample of 2063 people, aged 19 to 92, were collected for the nz dataset in 2019. More than half of the participants were women (63.3%). The data were collected as part of the New Zealand Attitudes and Values Study (NZAVS). Participants were recruited via online or post of a paper copy of the questionnaire. Data were collected and manipulated in R. A descriptive table was plotted using the "table1" function and a scatterplot was plotted to determine the relationship between weight and height using the "ggplot" package.

Question 3: Regression height ~ weight and report results

Regression model for height predicted by weight:

```
lm(HLTH.Height~HLTH.Weight, data = nz1)
```

```
##
## Call:
## lm(formula = HLTH.Height ~ HLTH.Weight, data = nz1)
##
## Coefficients:
## (Intercept)  HLTH.Weight
```

```
##    1.525450     0.002165
```

Height = 0.002(Weight) + 1.525

Table of regression model:

```
model <- lm(HLTH.Height~HLTH.Weight, data = nz1)
tab_model(model)
```

Height

Predictors

Estimates

CI

p

(Intercept)

1.53

1.51 – 1.54

<0.001

Weight

0.00

0.00 – 0.00

<0.001

Observations

2031

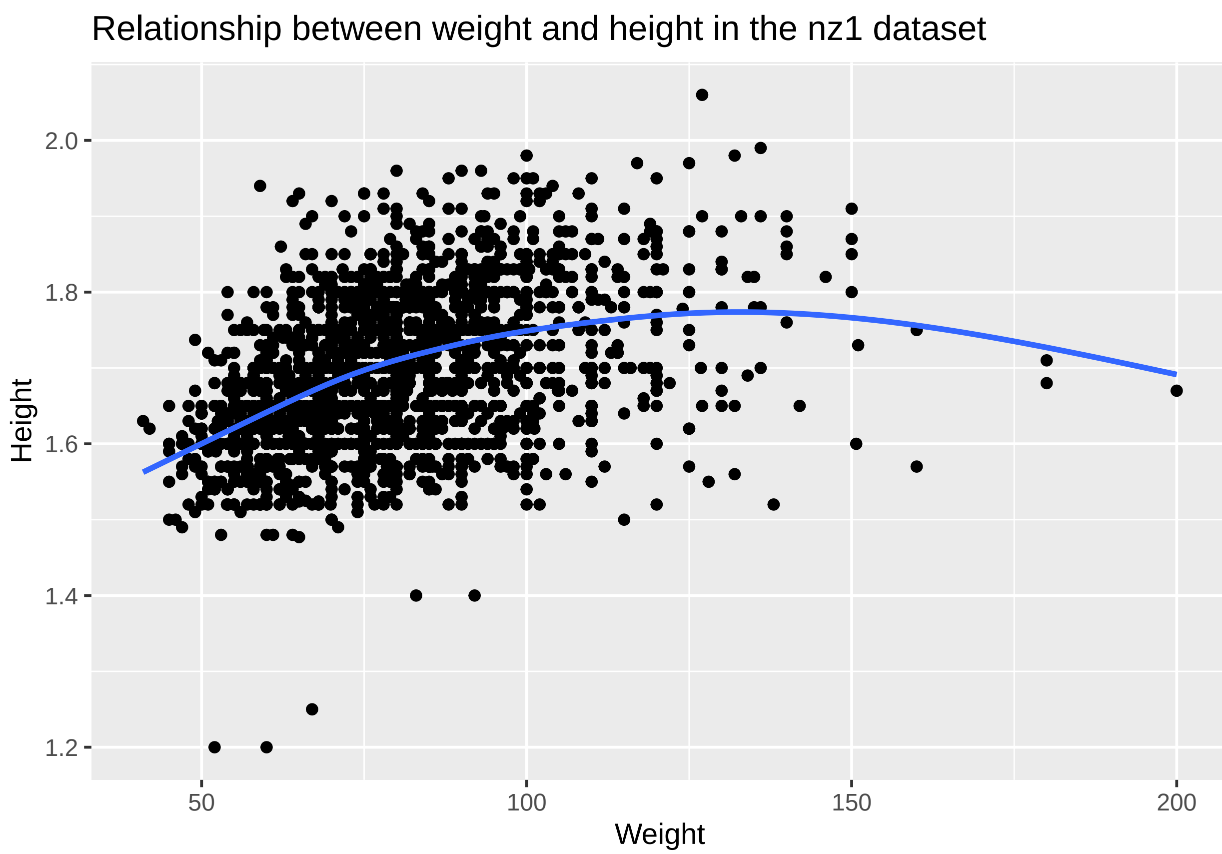
R2 / R2 adjusted

0.169 / 0.169

Graph of regression model:

```
ggplot(data = nz1) +
    geom_point(mapping = aes(x = HLTH.Weight, y = HLTH.Height)) +
    labs(title = "Relationship between weight and height in the nz1 dataset") +
    xlab("Weight") +
    ylab("Height") + geom_smooth(mapping = aes(x = HLTH.Weight, y = HLTH.Height), se = FALSE)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

## Warning: Removed 32 rows containing non-finite values (stat_smooth).

## Warning: Removed 32 rows containing missing values (geom_point).
```



The regression relationship between weight and height in the nz1 dataset was found to be statistically significant at $p<0.001$, as represented by "Height = 0.002(Weight) + 1.525". From the regression model, this relationship is a positive but subtle one as evident in the slope ("weight" coefficient). However, as depicted by the true line of best fit, this relationship appears to be a non-linear one. Nevertheless, despite several extreme outliers that skew the data away from a linear relationship, the general trend is that as weight increases, height also increases.

Question 4: Regress height ~ male_id and report results

Regression model for height predicted by male_id:

```
lm(HLTH.Height~male_id, data = nz1)
```

```
##
## Call:
## lm(formula = HLTH.Height ~ male_id, data = nz1)
##
## Coefficients:
##     (Intercept)  male_idNot_Male
##          1.7777          -0.1265
```

Height = -0.1265(male_id) + 1.7777

Table of regression model:

```r
model2 <- lm(HLTH.Height~male_id, data = nz1)
tab_model(model2)
```

Height

Predictors

Estimates

CI

p

(Intercept)

1.78

1.77 – 1.78

<0.001

male_id [Not_Male]

-0.13

-0.13 – -0.12

<0.001

Observations

2034

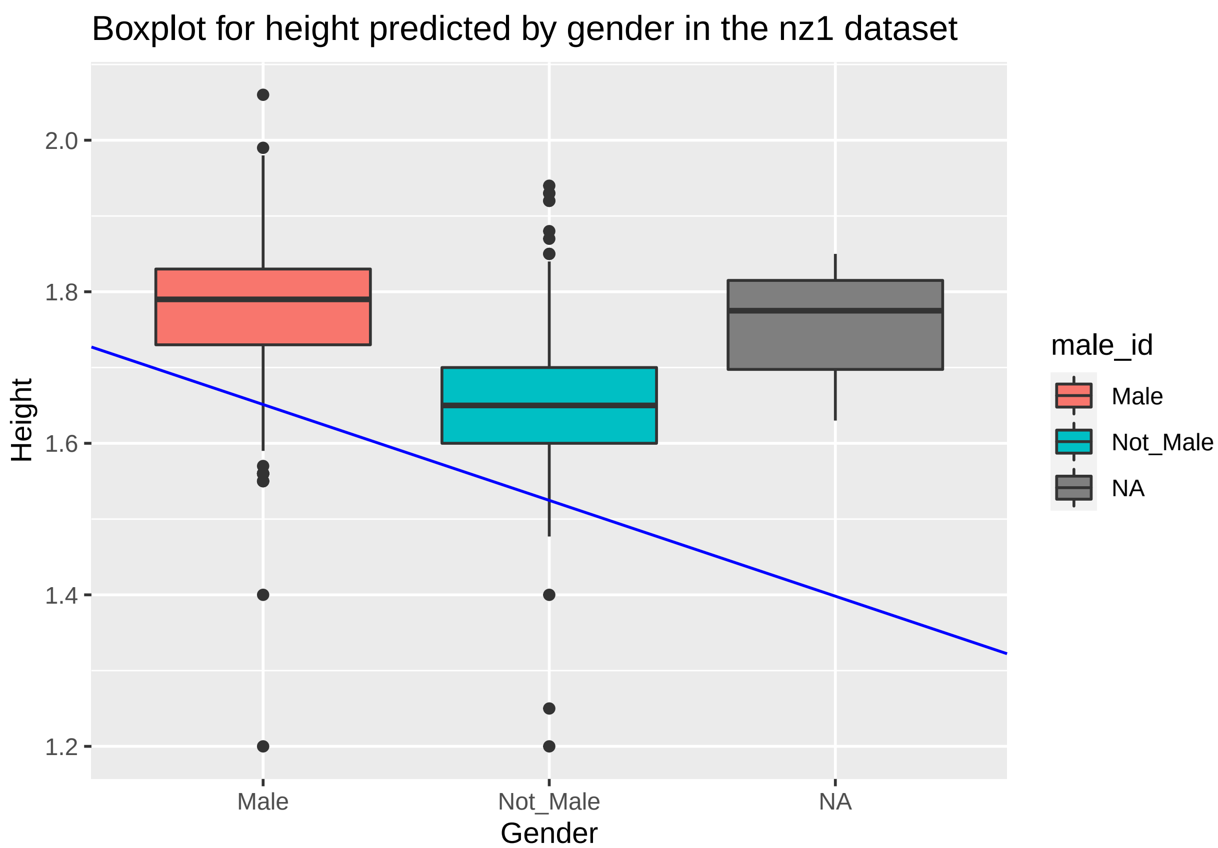
R2 / R2 adjusted

0.384 / 0.383

Graph of regression model:

```r
ggplot(data = nz1) +
    geom_boxplot(aes(x = male_id, y = HLTH.Height, fill = male_id)) +
    labs(title = "Boxplot for height predicted by gender in the nz1 dataset") +
    xlab("Gender") +
    ylab("Height") +
    geom_abline(aes(intercept = 1.7777, slope = -0.1265), col="blue")
```

```
## Warning: Removed 23 rows containing non-finite values (stat_boxplot).
```



The regression relationship between gender and height in the nz1 dataset was found to be statistically significant at $p<0.001$, as represented by "Height = -0.1265(male_id) + 1.7777". The boxplot depicts that females have a lower overall, and mean height as compared to males. Furthermore, the downward regression line depicts a negative relationship as expressed by the negative slope in the regression model, i.e. that females tend to be shorter than males.

Question 5: Using the regression coefficients from the Pearson and Lee 1903 dataset, predict the heights of daughters of women in the nz dataset.

load data:

```
md_df <- data.frame(read.table(url("https://raw.githubusercontent.com/avehtari/ROS-Examples/master/Pear
md_df <- md_df %>%
  dplyr::mutate(mother_height_c = as.numeric(scale(mother_height, center = TRUE, scale = FALSE)))
dplyr::glimpse(md_df)
```

```
## Rows: 5,524
## Columns: 3
## $ daughter_height <dbl> 52.5, 52.5, 53.5, 53.5, 55.5, 55.5, 55.5, 55.5, 56.5, ~
## $ mother_height   <dbl> 59.5, 59.5, 59.5, 59.5, 59.5, 59.5, 59.5, 59.5, 58.5, ~
## $ mother_height_c <dbl> -2.998732802, -2.998732802, -2.998732802, -2.998732802~
```

Regression coefficients for daughter_height ~ mother_height in the Pearson and Lee 1903 dataset:

```
m1 <- lm(daughter_height ~ mother_height, data = md_df)
```

To predict for the nz dataset:

```r
nz2<-nz1%>%
  dplyr::filter(male_id=="Not_Male") #To get only the population of interest -> women
table1(~male_id+HLTH.Height, data=nz2) #Min.Height = 1.20, Max.Height = 1.94
```

```
## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th clas
```

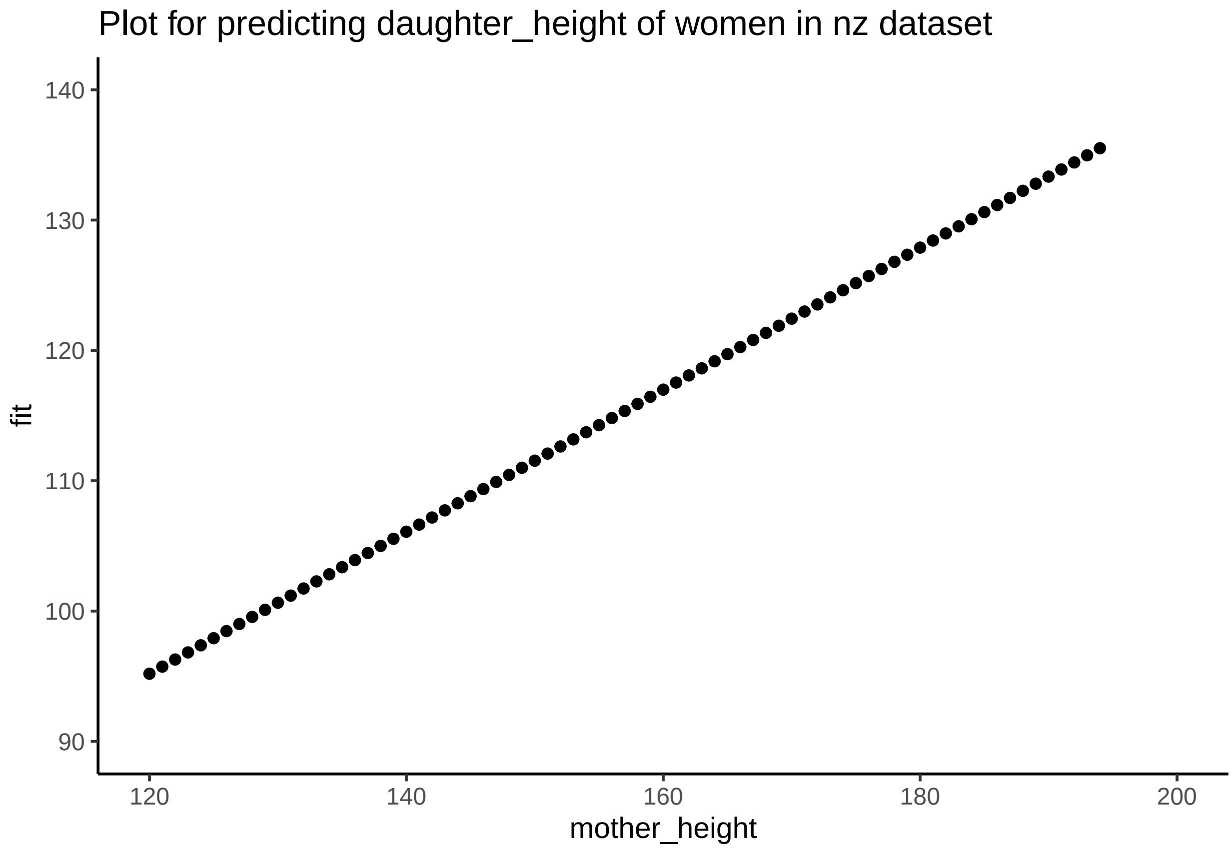
```r
nd<-expand.grid(mother_height = c(120:194))
pr<-predict(m1, type = "response", interval = "confidence", newdata =nd)
newdata<-data.frame(nd,pr)
newdata
```

```
##    mother_height       fit       lwr       upr
## 1            120  95.19082  93.76514  96.61649
## 2            121  95.73575  94.28533  97.18618
## 3            122  96.28069  94.80552  97.75587
## 4            123  96.82563  95.32570  98.32556
## 5            124  97.37057  95.84588  98.89525
## 6            125  97.91550  96.36607  99.46494
## 7            126  98.46044  96.88625 100.03463
## 8            127  99.00538  97.40643 100.60432
## 9            128  99.55031  97.92661 101.17401
## 10           129 100.09525  98.44680 101.74370
## 11           130 100.64019  98.96698 102.31340
## 12           131 101.18512  99.48716 102.88309
## 13           132 101.73006 100.00734 103.45278
## 14           133 102.27500 100.52752 104.02248
## 15           134 102.81993 101.04769 104.59217
## 16           135 103.36487 101.56787 105.16187
## 17           136 103.90981 102.08805 105.73156
## 18           137 104.45474 102.60823 106.30126
## 19           138 104.99968 103.12841 106.87095
## 20           139 105.54462 103.64858 107.44065
## 21           140 106.08955 104.16876 108.01035
## 22           141 106.63449 104.68894 108.58004
## 23           142 107.17943 105.20911 109.14974
## 24           143 107.72436 105.72929 109.71944
## 25           144 108.26930 106.24947 110.28914
## 26           145 108.81424 106.76964 110.85883
## 27           146 109.35917 107.28982 111.42853
## 28           147 109.90411 107.80999 111.99823
## 29           148 110.44905 108.33017 112.56793
## 30           149 110.99398 108.85034 113.13763
## 31           150 111.53892 109.37051 113.70733
## 32           151 112.08386 109.89069 114.27703
## 33           152 112.62879 110.41086 114.84673
## 34           153 113.17373 110.93104 115.41643
## 35           154 113.71867 111.45121 115.98613
## 36           155 114.26361 111.97138 116.55583
## 37           156 114.80854 112.49155 117.12553
## 38           157 115.35348 113.01173 117.69523
## 39           158 115.89842 113.53190 118.26493
## 40           159 116.44335 114.05207 118.83463
## 41           160 116.98829 114.57225 119.40433
```

```
## 42            161 117.53323 115.09242 119.97403
## 43            162 118.07816 115.61259 120.54374
## 44            163 118.62310 116.13276 121.11344
## 45            164 119.16804 116.65293 121.68314
## 46            165 119.71297 117.17310 122.25284
## 47            166 120.25791 117.69328 122.82254
## 48            167 120.80285 118.21345 123.39224
## 49            168 121.34778 118.73362 123.96195
## 50            169 121.89272 119.25379 124.53165
## 51            170 122.43766 119.77396 125.10135
## 52            171 122.98259 120.29413 125.67105
## 53            172 123.52753 120.81430 126.24076
## 54            173 124.07247 121.33447 126.81046
## 55            174 124.61740 121.85464 127.38016
## 56            175 125.16234 122.37481 127.94987
## 57            176 125.70728 122.89498 128.51957
## 58            177 126.25221 123.41515 129.08927
## 59            178 126.79715 123.93532 129.65898
## 60            179 127.34209 124.45549 130.22868
## 61            180 127.88702 124.97566 130.79838
## 62            181 128.43196 125.49583 131.36809
## 63            182 128.97690 126.01600 131.93779
## 64            183 129.52183 126.53617 132.50749
## 65            184 130.06677 127.05634 133.07720
## 66            185 130.61171 127.57651 133.64690
## 67            186 131.15664 128.09668 134.21661
## 68            187 131.70158 128.61685 134.78631
## 69            188 132.24652 129.13702 135.35602
## 70            189 132.79146 129.65719 135.92572
## 71            190 133.33639 130.17736 136.49542
## 72            191 133.88133 130.69753 137.06513
## 73            192 134.42627 131.21770 137.63483
## 74            193 134.97120 131.73787 138.20454
## 75            194 135.51614 132.25804 138.77424
```

For visualisation:

```
ggplot(data = newdata,
       aes(x= mother_height, y = fit)) +
    geom_point() +
    expand_limits(x = c(120,200), y = c(90,140)) + theme_classic() +
    labs(title = "Plot for predicting daughter_height of women in nz dataset")
```



Question 6: On average, how much taller or shorter are women in New Zealand as sampled in 2019 nz dataset compared with women in 1903 as sampled in the Pearson and Lee dataset.

Height of women in New Zealand in 2019:

```
table1(~HLTH.Height, data=nz2)
```

```
## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th clas
```

Height of women in 1903:

```
table1(~daughter_height+mother_height, data = md_df)
```

```
## [1] "<table class=\"Rtable1\">\n<thead>\n<tr>\n<th class='rowlabel firstrow lastrow'></th>\n<th clas
```

Comparing the means, women sampled in New Zealand in 2019 are much taller than women sampled by Pearson and Lee in 1903 by around 1m.