Data Analysis #2

Dustin Leatherman

March 2, 2019

Introduction

The effect of Diets on Weight Loss is a popular discussion topic. Which Diet provides better weight loss results? There doesn't appear to be any correlation between Age and Weight Loss or BMI and Weight Loss. All correlation coefficients for Weight Loss and Age/BMI are below 0.22 indicating a weak relationship between Age, Sex, BMI, and Weight Loss[1]. The boxplots are centered around similar points but the spread appears variable depending on diet and gender.

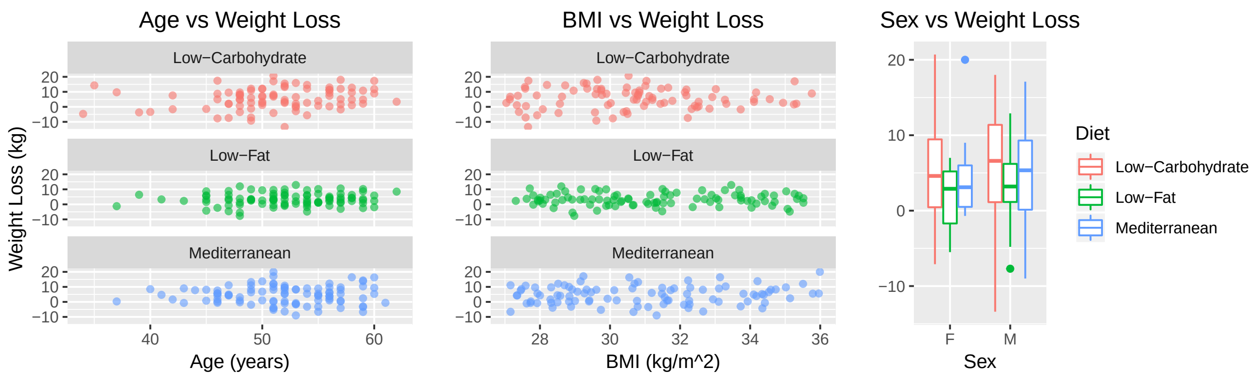
```
# Summary plots
plot.age <-
  ggplot(ex1420, aes(x = Age, y = WtLoss24, color = Diet)) +
    geom_point(alpha = 0.6, show.legend = FALSE) +
    facet_wrap(~Diet, ncol = 1) +
      ylab("Weight Loss (kg)") +
      xlab("Age (years)") +
      ggtitle("Age vs Weight Loss") +
        theme(plot.title = element_text(hjust = 0.5))

plot.bmi <-
  ggplot(ex1420, aes(x = BMI, y = WtLoss24, color = Diet)) +
    geom_point(alpha = 0.6, show.legend = FALSE) +
    facet_wrap(~Diet, ncol = 1) +
      ylab("") +
      xlab("BMI (kg/m^2)") +
      ggtitle("BMI vs Weight Loss") +
        theme(plot.title = element_text(hjust = 0.5))

plot.sex <-
  ggplot(ex1420, aes(x = Sex, y = WtLoss24, color = Diet)) +
    geom_boxplot() +
      ylab("") +
      xlab("Sex") +
      ggtitle("Sex vs Weight Loss") +
        theme(plot.title = element_text(hjust = 0.5))

grid.arrange(plot.age, plot.bmi, plot.sex,
               widths = c(1,1,1),
               ncol = 3)
```

[1] The correlation coefficient values are not depicted.



Experiment Design

The experiment cannot be considered balanced. There are differing counts for each Diet, Sex, and Initial Age. There are some ages that only contain one observation so there is a lack of replication for these ages. Otherwise, there is replication for the other fields.

Methods

The first inclincation is to use a full model with complete interaction terms. Given the high sample size, there are enough degrees of freedom to use 4 way interaction terms.

$\mu\{\text{Weight Loss} \mid \text{MALE, Age, BMI, Diet}\} = \beta_0 + \beta_1\text{MALE} + \beta_2\text{Age} + \beta_3\text{BMI} + \beta_4\text{LOWCARB} + \beta_5\text{LOWFAT} + \beta_6(\text{LOWCARB} \times \text{MALE}) + \beta_7(\text{LOWFAT} \times \text{MALE}) + \beta_8(\text{Age} \times \text{MALE}) + \beta_9(\text{BMI} \times \text{MALE}) + \beta_{10}(\text{Age} \times \text{BMI}) + \beta_{11}(\text{LOWCARB} \times \text{Age}) + \beta_{12}(\text{LOWFAT} \times \text{Age}) + \beta_{13}(\text{LOWCARB} \times \text{BMI}) + \beta_{14}(\text{LOWFAT} \times \text{BMI}) + \beta_{15}(\text{MALE} \times \text{BMI} \times \text{Age}) + \beta_{16}(\text{MALE} \times \text{BMI} \times \text{LOWFAT}) + \beta_{17}(\text{MALE} \times \text{BMI} \times \text{LOWCARB}) + \beta_{18}(\text{MALE} \times \text{Age} \times \text{LOWFAT}) + \beta_{19}(\text{MALE} \times \text{Age} \times \text{LOWCARB}) + \beta_{20}(\text{MALE} \times \text{BMI} \times \text{LOWFAT}) + \beta_{21}(\text{MALE} \times \text{BMI} \times \text{LOWCARB}) + \beta_{22}(\text{Age} \times \text{BMI} \times \text{LOWFAT}) + \beta_{23}(\text{Age} \times \text{BMI} \times \text{LOWCARB}) + \beta_{24}(\text{MALE} \times \text{Age} \times \text{BMI} \times \text{LOWCARB}) + \beta_{25}(\text{MALE} \times \text{Age} \times \text{BMI} \times \text{LOWFAT})$

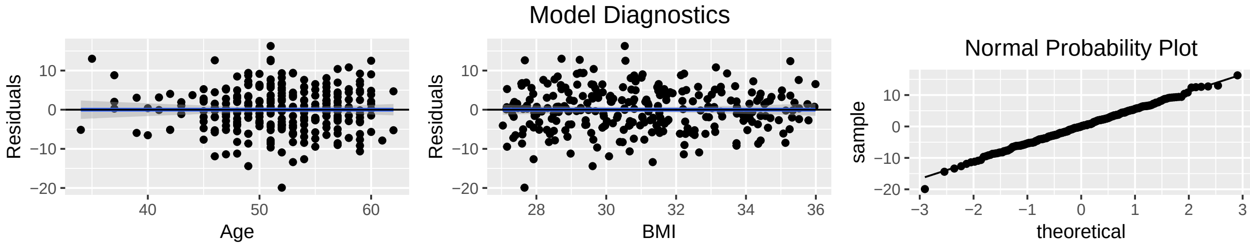
```
ex1420$DietRelevel <- relevel(ex1420$Diet, ref = "Mediterranean")
full.model <- lm(WtLoss24 ~ Sex * Age * BMI * Diet, data = ex1420)

model.plot.res.age <-
  ggplot(full.model, aes(x = Age, y = .resid)) +
  geom_point() +
  geom_smooth(method = "lm") +
  geom_hline(yintercept = 0) +
    ylab("Residuals")

model.plot.res.bmi <-
  ggplot(full.model, aes(x = BMI, y = .resid)) +
  geom_point() +
  geom_smooth(method = "lm") +
  geom_hline(yintercept = 0) +
    ylab("Residuals")

model.plot.norm <-
  ggplot(full.model, aes(sample = .resid)) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("Normal Probability Plot") +
        theme(plot.title = element_text(hjust = 0.5))
```

```r
grid.arrange(model.plot.res.age, model.plot.res.bmi, model.plot.norm,
               widths = c(1,1,1),
               ncol = 3,
               top = textGrob("Model Diagnostics",
                            gp=gpar(fontsize=14,font=1),just=c("center"))
             )
```



The residual plots for Age and BMI appear to be symmetrical and random. The normal probability plot indicates that the normality assumption is met. Given that the full model meets these aassumptions, it is assumed that derivative models also meet the assumptions. The additional models considered are:

1. Two-Way Interactions Only Model

$\mu\{\text{Weight Loss} \mid \text{MALE, Age, BMI, Diet}\} = \beta_0 + \beta_1\text{MALE} + \beta_2\text{Age} + \beta_3\text{BMI} + \beta_4\text{LOWCARB} + \beta_5\text{LOWFAT} + \beta_6(\text{LOWCARB} \times \text{MALE}) + \beta_7(\text{LOWFAT} \times \text{MALE}) + \beta_8(\text{Age} \times \text{MALE}) + \beta_9(\text{BMI} \times \text{MALE}) + \beta_{10}(\text{Age} \times \text{BMI}) + \beta_{11}(\text{LOWCARB} \times \text{Age}) + \beta_{12}(\text{LOWFAT} \times \text{Age}) + \beta_{13}(\text{LOWCARB} \times \text{BMI}) + \beta_{14}(\text{LOWFAT} \times \text{BMI})$

2. Parallel lines Model (No Interaction)

$\mu\{\text{Weight Loss} \mid \text{MALE, Age, BMI, Diet}\} = \beta_0 + \beta_1\text{MALE} + \beta_2\text{Age} + \beta_3\text{BMI} + \beta_4\text{LOWCARB} + \beta_5\text{LOWFAT}$

3. Parallel Lines Model (No Diet)

$\mu\{\text{Weight Loss} \mid \text{MALE, Age, BMI}\} = \beta_0 + \beta_1\text{MALE} + \beta_2\text{BMI} + \beta_3\text{Age}$

```r
two.way.interactions.model <- lm(WtLoss24 ~ Sex * Age + Sex * DietRelevel + DietRelevel * BMI + BMI * Ag
plines.model <- lm(WtLoss24 ~ Sex + Age + DietRelevel + BMI, data = ex1420)
plines.noDiet.model <- lm(WtLoss24 ~ Sex + Age + BMI, data = ex1420)
# The following lines are commented out to prevent them from printing on the page
# These can be uncommented for execution
#Anova(plines.model, contrasts=list(topic=contr.sum, sys=contr.sum), type=3)
# anova(two.way.interactions.model, full.model) %>%
#   as_tibble %>%
#   kable(
#     digits = 4,
#     col.names = c("Residual DF", "RSS", "DF", "Sum of Squares", "F statistic", "p-value"),
#     caption = "Comparing the Two-Way Interaction model with the Full Model"
#   ) %>%
#   kable_styling(full_width = T, bootstrap_options = "striped", latex_options = "hold_position")
#
# anova(plines.model, two.way.interactions.model) %>%
#   as_tibble %>%
#   kable(
#     digits = 4,
#     col.names = c("Residual DF", "RSS", "DF", "Sum of Squares", "F statistic", "p-value"),
#     caption = "Comparing the Parallel Lines Model with the Two-Way Interaction Model"
#   ) %>%
#   kable_styling(full_width = T, bootstrap_options = "striped", latex_options = "hold_position")
```

```r
#
# anova(plines.noDiet.model, plines.model) %>%
#   as_tibble %>%
#   kable(
#     digits = 4,
#     col.names = c("Residual DF", "RSS", "DF", "Sum of Squares", "F statistic", "p-value"),
#     caption = "Comparing the Parallel Lines Model (Without Diet) with the Parallel Lines Model (With 
#   ) %>%
#   kable_styling(full_width = T, bootstrap_options = "striped", latex_options = "hold_position")

ex1420$DietRelevelCarb <- relevel(ex1420$Diet, ref = "Low-Carbohydrate")
ex1420$DietRelevelFat <- relevel(ex1420$Diet, ref = "Low-Fat")
plines.model.carb <- lm(WtLoss24 ~ Sex + Age + DietRelevelCarb + BMI, data = ex1420)
plines.model.fat <- lm(WtLoss24 ~ Sex + Age + DietRelevelFat + BMI, data = ex1420)
# summary(plines.model.carb)
# summary(plines.model.fat)
```

There is not enough evidence to suggest that the three and four way interactions have a significant impact on the outcome of the model when compared to the two-way interaction model. (SS F-Test. p-value = 0.6698)

There is not enough evidence to suggest that the two-way interactions have a signicant impact on the outcome of the model when compared to the parallel lines model (SS F-Test. p-value = 0.434).

There is moderate evidence to suggest that Diet has a significant impact on the outcome of a model when compared to the parallel lines model (SS F-Test. p-value = 0.025).

Given that there is moderate evidence for the parallel lines model including Diet, the parallel lines model is the best choice among these models for determining the effect of diet on weight loss. Additionally, since this design is unbalanced, the Type III Sum of Squares must be analyzed for interpretation. The adjusted R^2 value is 0.02355 which indicates that much of this model is left unexplained. Additional data points should be collected to bridge this gap.

Conclusion

There is convincing evidence that the Low-Carbohydrate Diet has an impact on average weight loss compared to the Mediterranean Diet after accounting for Sex, Age, and BMI (Two-Way ANOVA. p-value = 0.00703). It is estimated that the Low-Carbohydrate Diet increases average weight loss by 2.361kg compared to the participants on the Mediterannean Diet. With 95% confidence, average weight loss is be between 0.6497 and 4.0716 kilograms higher for participants on a Low-Carb diet compared to participants on the Mediterranean Diet.

There is convincing evidence that the average weight gain increases for participants on the Low-Fat diet compared to the Low-Carb Diet after accounting for Age, BMI, and Sex (Two-Way ANOVA. p-value = 0.00703). It is estimated that the average weight gain is 2.36kg higher for participants on the Low-Fat Diet compared to the Low-Carb Diet. With 95% confidence, average weight gain is between 0.6497 and 4.0716 kilograms higher compared to participants on the Low-Carb Diet.

There is no evidence that the average weight loss increases for participants on the Mediterranean Diet compared to the Low-Carb or Low-Fat diet (Two-Way ANOVA. p-values = 0.249, 0.109 respectively).