Homework-5-Q2

BIOS507 Spring 2025 | Dr Lukemire | Elizabeth Nemeti Due: March 24 2025

Problem 2.

Public health researchers are investigating how lifestyle factors like sleep duration and physical activity influence cholesterol levels, and whether this relationship differs based on dietary habits. Cholesterol levels are a key indicator of cardiovascular health, with high levels increasing the risk of heart disease.

The study categorizes participants into three dietary patterns: **Plant-Based Diets, Balanced Diets, and High-Meat Diet**. The hypothesis is that increased sleep and physical activity are generally associated with lower cholesterol levels, but the magnitude of these effects differs based on dietary habits, with plant-based eaters potentially benefiting more due to better metabolic profiles, while high-meat consumers may show a weaker response.

The dataset is diet_sleep_exercise_cholesterol.csv. Note that this is simulated data for this example homework problem and does not represent a real study.

- **Y (response)**: cholesterol (high levels increasing the risk of heart disease)
- **X1 (predictor)**: sleepHours (sleep duration)
- **X2 (predictor)**: activity (physical activity)
- **X3 (predictor)**: diet (three dietary patterns: High-Meat; Plant-Based; Balanced)

```
data_path = "/Users/elizabethnemeti/Documents/GitHub/BIOS507-Coursework/Homeworks/Homework-5/"
data_file <- file.path(data_path, "diet_sleep_exercise_cholesterol.csv")
health_data <- read.csv(data_file, header = TRUE)

head(health_data)
```

```
##   sleepHours activity        diet cholesterol
## 1   6.914479 5.437672   High-Meat    170.7072
## 2   7.404992 6.670029 Plant-Based    104.9069
## 3   8.311707 7.165005    Balanced    156.0398
## 4   6.714480 4.695673    Balanced    168.4370
## 5   8.651047 5.691981 Plant-Based    120.8821
## 6   5.659362 1.454399    Balanced    197.4333
```

```
str(health_data)
```

```
## 'data.frame':    300 obs. of  4 variables:
##  $ sleepHours : num  6.91 7.4 8.31 6.71 8.65 ...
##  $ activity   : num  5.44 6.67 7.17 4.7 5.69 ...
##  $ diet       : chr  "High-Meat" "Plant-Based" "Balanced" "Balanced" ...
##  $ cholesterol: num  171 105 156 168 121 ...
```

1. Fit a model (Model 1) that includes all main effects and two-factor interactions. Conduct a test for the presence of the sleep duration $\times$ physical activity interaction. Conduct a test for the presence of the diet $\times$ physical activity interaction.

```
health_data$diet <- as.factor(health_data$diet)
```

```r
health_model <- lm(cholesterol ~ sleepHours * activity * diet, data = health_data)
summary(health_model)
```

```
##
## Call:
## lm(formula = cholesterol ~ sleepHours * activity * diet, data = health_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.7230  -6.6219   0.1621   6.6637  28.5740
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       202.85217    8.74332  23.201   <2e-16 ***
## sleepHours                         -2.32133    1.26013  -1.842   0.0665 .
## activity                           -4.60181    2.05837  -2.236   0.0261 *
## dietHigh-Meat                       3.25002   15.89525   0.204   0.8381
## dietPlant-Based                   -12.57792   13.64956  -0.921   0.3576
## sleepHours:activity                -0.05058    0.28794  -0.176   0.8607
## sleepHours:dietHigh-Meat            1.06540    2.28890   0.465   0.6420
## sleepHours:dietPlant-Based         -0.70631    1.91071  -0.370   0.7119
## activity:dietHigh-Meat              0.71565    3.54602   0.202   0.8402
## activity:dietPlant-Based           -4.88782    3.04652  -1.604   0.1097
## sleepHours:activity:dietHigh-Meat  -0.14111    0.50829  -0.278   0.7815
## sleepHours:activity:dietPlant-Based  0.35163    0.42164   0.834   0.4050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.694 on 288 degrees of freedom
## Multiple R-squared:  0.8122, Adjusted R-squared:  0.8051
## F-statistic: 113.3 on 11 and 288 DF,  p-value: < 2.2e-16
```

Conduct a test for the presence of the sleep duration × physical activity interaction. Full model: E[cholesterol] = ß0 + ß1(sleepHours) + ß2(activity) + ß3(diet) + ß4(sleepHours)(activity) + ß5(diet)(activity) + ß6(sleepHours)(diet)

Reduced model: E[cholesterol] = ß0 + ß1(sleepHours) + ß2(activity) + ß3(diet) + ß5(diet)(activity) + ß6(sleepHours)(diet)

```r
full_model_a <- lm(cholesterol ~ sleepHours + activity + diet + sleepHours:activity + diet:activity + sl
reduced_model_a <- lm(cholesterol ~ sleepHours + activity + diet + diet:activity + sleepHours:diet, data

anova_result_a <- anova(reduced_model_a, full_model_a)
print(anova_result_a)
```

```
## Analysis of Variance Table
##
## Model 1: cholesterol ~ sleepHours + activity + diet + diet:activity +
##     sleepHours:diet
## Model 2: cholesterol ~ sleepHours + activity + diet + sleepHours:activity +
##     diet:activity + sleepHours:diet
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    291 27174
## 2    290 27167  1    7.1776 0.0766 0.7821
```

Conduct a test for the presence of the diet × physical activity interaction. Full

model: E[cholesterol] = ß0 + ß1(sleepHours) + ß2(activity) + ß3(diet) + ß4(sleepHours)(activity) + ß5(diet)(activity) + ß6(sleepHours)(diet)

Reduced model: E[cholesterol] = ß0 + ß1(sleepHours) + ß2(activity) + ß3(diet) + ß4(sleepHours)(activity) + ß6(sleepHours)(diet)

```
full_model_b <- lm(cholesterol ~ sleepHours + activity + diet + sleepHours:activity + diet:activity + sl
reduced_model_b <- lm(cholesterol ~ sleepHours + activity + diet + sleepHours:activity + sleepHours:diet

anova_result_b <- anova(reduced_model_b, full_model_b)
print(anova_result_b)
```

```
## Analysis of Variance Table
##
## Model 1: cholesterol ~ sleepHours + activity + diet + sleepHours:activity +
##     sleepHours:diet
## Model 2: cholesterol ~ sleepHours + activity + diet + sleepHours:activity +
##     diet:activity + sleepHours:diet
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    292 28808
## 2    290 27167  2    1640.9 8.7583 0.0002027 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2. Fit a model (Model 2) that includes all main effects, but only the diet × physical activity interaction. Create a conditional effects plot based on this model that demonstrates the interaction between physical activity and diet on cholesterol. Be sure to provide a written description of the pattern that you observe.

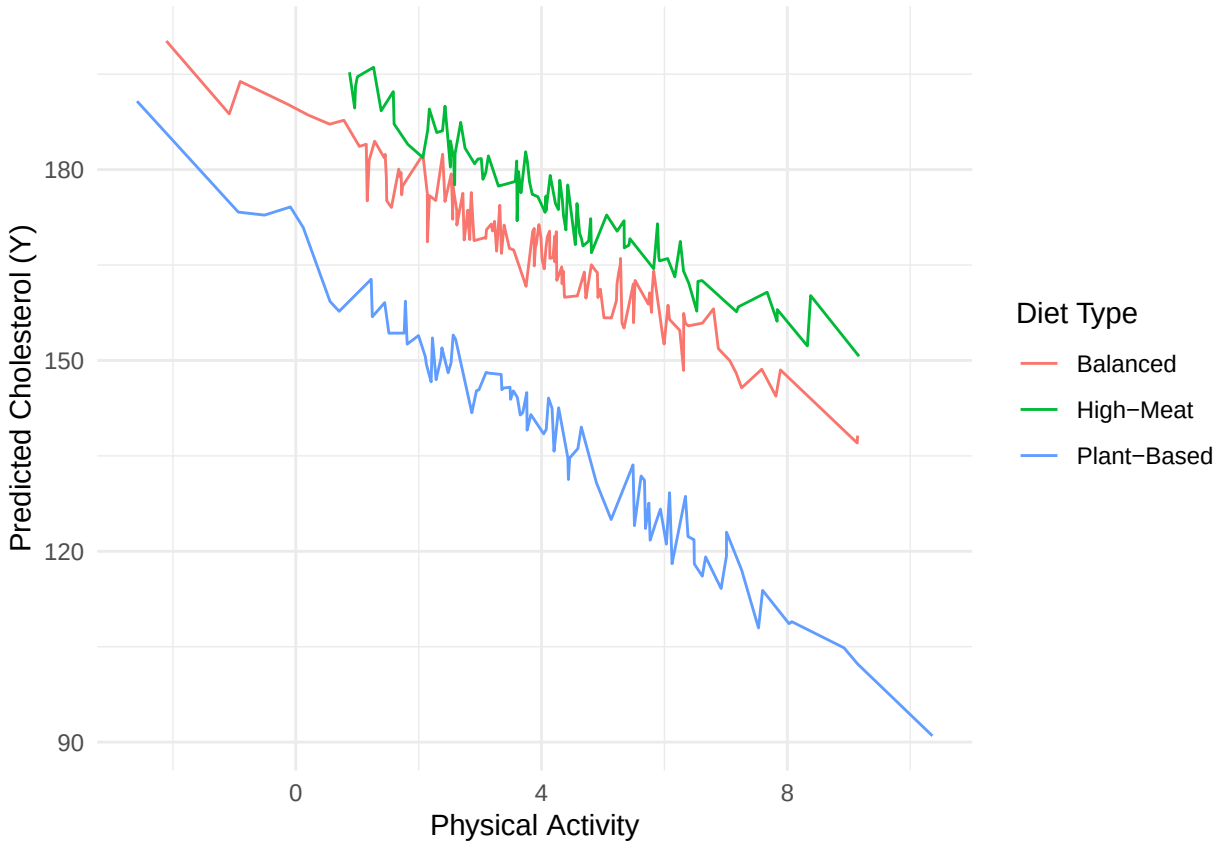
```
# just the diet × physical activity interaction
model_2 <- lm(cholesterol ~ sleepHours + activity + diet + diet:activity, data = health_data)

model_2_preds <- predict(model_2, newdata = health_data) # getting cholesterol predictions

# create new df for plotting, so we can better group by diet
ggplot_df <- data.frame(
  activity = health_data$activity,
  diet = health_data$diet,
  predicted_cholesterol = model_2_preds
)

ggplot(ggplot_df, aes(x = activity, y = predicted_cholesterol, color = diet)) +
  geom_line(size = 0.5) +
  labs(x = "Physical Activity",
       y = "Predicted Cholesterol (Y)",
       color = "Diet Type") +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



In the above plot, we can distinguish three distinct slopes for each of the 3 diet types: balanced, high-meat, and plant-based. This pattern indicates a significant interaction between diet and physical activity. The plant-based diet (blue slope) shows the steepest decline in predicted cholesterol levels as physical activity increases, suggesting that physical activity has the strongest effect on cholesterol reduction for those following a plant-based diet. In contrast, individuals following a high-meat diet (green slope) or balanced diet (orange slope) experience a weaker effect of physical activity on cholesterol reduction, as their slopes decline less sharply. All three slopes show fluctuations, which could be expected in a system as complex as cholesterol regulation. The intercept for the high-meat diet is the highest, suggesting that at zero physical activity, individuals on a high-meat diet are predicted to have the highest cholesterol levels of the three diets. Conversely, the plant-based diet starts with the lowest predicted cholesterol at zero physical activity, suggesting it may contribute to a better metabolic profile.