Homework-4-Q2

BIOS507 Spring 2025 | Dr Lukemire | Elizabeth Nemeti Due: March 17 2025

Problem 2.

A researcher is investigating how BMI (Y) depends on age (X1, years) and healthy diet score (X2). The higher the diet score, the healthier the daily diet. The researcher has developed the following 2 regression models: Model 1: Yhat = 18 + 0.2(X1) - 0.1(X2) Model 2: Yhat = 26 + 0.15(X1) - 0.07(X2) − 0.02(X1)(X2)

Both models have been built using a data set with data ranges: 20 X1 50 and 1 X2 10.

- **Y (response variable)**: BMI
- **X1 (predictor variable)** : age in years
- **X2 (predictor variable)**: healthy diet score (higher = better)

1. Using both models, write down the predicted value of BMI when X2 = 1. Note that this will be a function of X1 - you will not have a single answer. Do the same thing for X2 = 5. Comment on the effect of the interaction term in model 2. Bonus: Generate plots under both models for the predicted value of BMI when X2 = 1 as a function of age.

Predicted value of BMI when X2 = 1:

Model 1: Yhat = 18 + 0.2(X1) - 0.1(X2)
= 18 + 0.2(X1) - 0.1(1)
= 18 + 0.2(X1) - 0.1
= 17.9 + 0.2(X1)

Model 2: Yhat = 26 + 0.15(X1) - 0.07(X2) - 0.02(X1)(X2)
= 26 + 0.15(X1) - 0.07(1) - 0.02(X1)(1)
= 26 + 0.15(X1) - 0.07 - 0.02(X1)
= 25.93 + 0.15(X1) - 0.02(X1)
= 25.93 + (0.15 - 0.02)X1
= 25.93 + 0.13(X1)

Predicted value of BMI when X2 = 5:

Model 1: Yhat = 18 + 0.2(X1) - 0.1(X2) = 18 + 0.2(X1) - 0.1(5)
= 18 + 0.2(X1) - 0.5
= 17.5 + 0.2(X1)

Model 2: Yhat = 26 + 0.15(X1) - 0.07(X2) - 0.02(X1)(X2)
= 26 + 0.15(X1) - 0.07(5) - 0.02(X1)(5)
= 26 + 0.15(X1) - 0.35 - 0.1(X1)
= 25.65 + 0.15(X1) - 0.1(X1)
= 25.65 + (0.15 - 0.1)X1
= 25.65 + 0.05(X1)

The interaction term in Model 2 is **-0.02(X1)(X2)**. This term shows how the effect of age (X1) on BMI (Y) is not constant when we have the healthy diet score (X2) to consider. As the healthy diet score (X2) increases (e.g. X2 = 1 example vs X2 = 5 example), the slope for age (X1) decreases from 0.13 to 0.05, indicating that the better the diet, the less of an effect age has on BMI. This makes intuitive sense, as a healthy diet is a core component of maintaining a healthy and fit BMI as one ages.

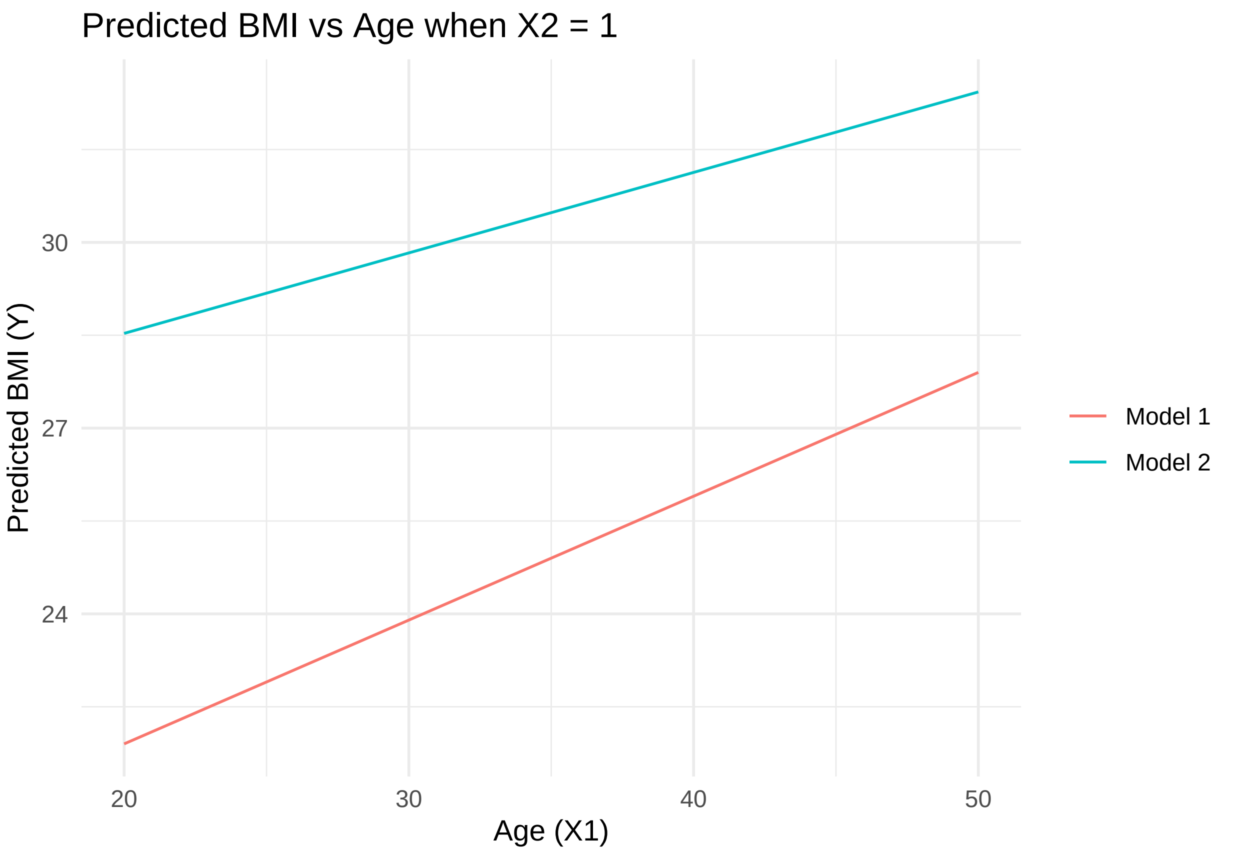
```r
# Generate plots under both models for the predicted value of BMI (Y) when X2=1 as a function of age (X

X2 <- 1
X1 <- seq(20, 50, by = 1) # refer back to our range 20   X1   50, make sure it's 1 step

model_1_pred <- 17.9 + 0.2*X1
model_2_pred <- 25.93 + 0.13*X1

# need to make a new df for ggplot
ggplot_df <- data.frame(
  age = X1,
  model1 = model_1_pred,
  model2 = model_2_pred)

ggplot(ggplot_df, aes(x = age)) +
  geom_line(aes(y = model1, color = "Model 1")) +
  geom_line(aes(y = model2, color = "Model 2")) +
  labs(x = "Age (X1)", y = "Predicted BMI (Y)",
       title = "Predicted BMI vs Age when X2 = 1",
       color = " ") +
  theme_minimal()
```



2. Find the expected change in BMI for a one year increase in age for model 1 when X2 = 5. Does this quantity depend on the specific value of X2? Why or why not?

For Model 1 when X2 = 5, the predicted BMI is: Yhat = 17.5 + 0.2(X1). This means the expected change in BMI for a one year increase in age would be 0.2 units, the coefficient for our age variable X1. Since there is no interaction term for X1 and X2 in Model 1 like there was in Model 2, the slope for age is constant at 0.2 units per year regardless of the value of X2. Therefore, the change in BMI does not depend on X2.

3. Find the expected change in BMI for a one year increase in age for model 2 when X2 = 5. Does this quantity depend on the specific value of X2? Why or why not?

For Model 2 when X2 = 5, the predicted BMI is: Yhat = 25.65 + 0.05(X1). This means the expected change in BMI for a one year increase in age would be 0.05 units. However, in Model 2, we do have an interaction term: -0.02(X1)(X2). Now this means that the expected change in BMI for a one year increase in age does depend on a specific value of X2. As X2 increases, the effect of age (X1) on BMI (Y) decreases.