**Linear Regression – in-class example**

For this exercise, we will explore the HeightWeight dataset. This dataset includes the height (inches), weight (pounds), age (years), and sex as classified at birth (female=1 if female; female=0 if male) of 237 adolescents (up to age 21). This exercise will explore the relationship between age and height before and after accounting for sex in this population. You may complete this exercise in R, Stata, or SAS.

1. Open the HeightWeight dataset
2. Inspect the data by preparing a two-way scatterplot of height as predicted by age. Do you notice any outliers that are of concern?

Chart, scatter chart

Description automatically generated

1. Delete any observations outside of the intended age range of the study.
2. Prepare a Table 1. Stratify by sex in the columns and show the data for age (continuous), age in 3 categories (<15, 15-18, >=18). Show the relevant p-values.

Table

Description automatically generated

1. Write out the linear regression model estimating the expected value of height as a function of age using standard notation and interpret each term in the model.
   1. What is the interpretation of β0?
   2. What is the interpretation of β1?
   3. What is the expected distribution of the residuals?

A picture containing table

Description automatically generated

1. Use software to fit the linear regression model specified in your answer to question 5.
2. Based on the fitted model, what is the expected height for a 16-year-old?

Graphical user interface

Description automatically generated

1. Inspect the data by preparing a two-way scatterplot of height as predicted by sex.

Chart, histogram

Description automatically generated

1. Use software to fit a linear regression model estimating the expected value of height as a function of sex (consider female as exposed).
   1. What is the interpretation of β0?
   2. What is the interpretation of β1?
   3. State the null and alternative hypothesis (in terms of the model coefficients), testing the hypothesis that mean height does not vary according to sex.
   4. Interpret the p-value for the test described in part 9c.
2. Use software to fit a linear regression model for the association between age and height, adjusting for sex. Interpret each term in the model.

A picture containing text

Description automatically generated

1. Using your expert subject matter knowledge, do you think the relationship between age and height might be different for adolescent males and females? Note that this asks if you hypothesize that there might be effect measure modification of the age-height relationship by sex.

Text

Description automatically generated with low confidence

1. Write out the linear regression model estimating the expected value of height as a function of age, sex, and the interaction between age and sex using standard notation and interpret each term in the model.
2. Use software to fit a linear regression model for the association between age and height, adjusting for female and allowing for the relationship between age and height to be different for males and females. **HINT:** You will need to create a multiplicative interaction term between age and sex to complete this analysis.
3. State the null and alternative hypotheses for the test of whether the association between age and height varies according to sex. What is the p-value for this test?

A screenshot of a computer

Description automatically generated with medium confidence

1. Estimate the association (including the point estimate and 95% confidence interval) between age and height among males. What is the expected height for a 16-year-old male?
2. Estimate the association (including the point estimate and 95% confidence interval) between age and height among females. What is the expected height for a 16-year-old female?