BIOST 515 Homework 1

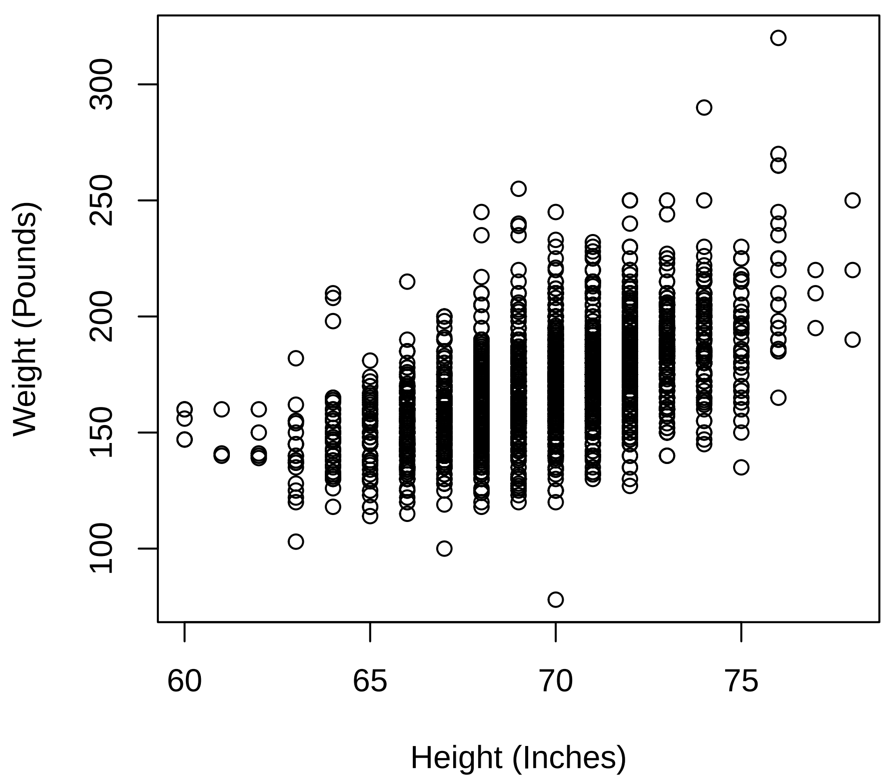
Katie McFarlane

January 14, 2022

Responses

1. Make a scatterplot of weight on the vertical axis and height on the horizontal axis. Based on this data, do you believe there is a linear relationship between weight and height in this population? Why or why not?

Yes, I do believe there is a linear relationship between weight and height within this population. Visually one can see that on average, those in this study who are shorter appear to weigh less, while those who are taller appear to weigh more. With a sample size of over 3,000 individuals, we expect this sample to have a similar distribution to the greater population of 39 – 59 year old males, where we also expect a linear relationship between average weight and height.



2. Perform a simple linear regression of weight on height using the WCGS data. In 1-3 carefully written sentences, summarize the results in language suitable for a scientific publication.

The estimated difference in the expected value of weight between the populations of middle-aged men whose heights differ by one inch is 4.45 pounds. The simple linear regression is fitted through the point representing both mean height and weight in the study, which is observed at 69.78 inches, 169.95 pounds.

3. Contrast the heteroscedasticity-robust standard error in the estimate of the slope coefficient to a standard error obtained assuming homoscedasticity, and comment. Which do you prefer and why?

The robust standard error of the slope parameter is 0.14, while the non-robust standard error is 0.13. In practice, this difference is small enough that it is likely negligible. Regardless, I would prefer to use the robust standard error: while it is larger and may slightly overestimate the variance, using the non-robust standard error can inflate type-I errors in the case of heteroskedasticity, which is preferable to avoid.

4. Comment on the reasonableness of using the fitted linear regression to estimate the mean weight of men from this population who are 70 inches tall.

Our data includes many participants who are 70 inches tall, meaning that the use of this data to talk about men in the greater population who are 70 inches tall is not extrapolation. We also have a large sample size, and the data collected in this study appears to be reasonably close to the distribution we would expect from the greater population. Therefore, it is reasonable to estimate the mean weight of men from this population who are 70 inches tall using our fitted linear regression.

5. Comment on the reasonableness of using the fitted linear regression to estimate the mean weight of an adolescent boy who is 42 inches tall.

Subjects in the study are between the ages of 39 and 59, and the minimum height recorded is 60 inches, so any estimations we could calculate about a 42-inch child is extrapolation. I do not expect this linear relationship to generalize to children, who have heights and weights that are inconsistent during devlopment. Thus, it is unreasonable to use this fitted linear regression to estimate the mean weight of adolescents who are 42 inches tall.

6. Convert height from inches to centimeters and fit a new simple linear regression model. Interpret the fitted model parameters and compare/contrast your results to those from Question 2.

The estimated difference in the expected value of weight between the populations of middle-aged men whose heights differ by one centimeter is 1.75 pounds. This line is fitted through the point representing both mean height and weight in this study, which is observed at 177.24 centimeters, 169.95 pounds.

The y-intercept of these two fitted lines is the same at -140.3 pounds, which makes sense, as we did not alter the weight variable, and this intercept corresponds to a height of zero, which is equivalent across inches and centimeters. Our mean weight and height are the equivalent (with the mean height having been converted from 69.78 inches to 177.24 centimeters). Our slope parameter did change, though; the parameter is 2.54 times smaller in our second linear regression. This corresponds directly to the conversion between inches and centimeters (1 inch: 2.54 centimeters).

7. Participant 2001 (aged 49 y.o.) weighed 150 pounds and was 73 inches tall. Which do you think is a better prediction of the weight of another 73 inch tall man from this population: the observed weight of Participant 2001 of 150 pounds, or the fitted value from the simple linear regression that you fit? Why?

I believe the fitted value from the linear regression will be a better predictor – the linear regression is based on more data, including data on other participants who are 73 inches tall, and therefore can be used as a better indicator for the population than one single participant might be. Any one participant could be extremely tall or extremely short for his weight compared to the average and would therefore be an unhelpful predictor on average.

Code Appendix

```
### Setting up the packages, options we'll need:
library(knitr)
library(rigr)
knitr::opts_chunk$set(echo = FALSE)
### -----------------------------------------------------------
### Reading in the data.
wcgs <- read.csv("~/Graduate School Work/Winter 2022 - BIOST 515/wcgs.csv")

### -----------------------------------------------------------
### Q1
plot(wcgs$weight0~wcgs$height0, xlab="Height (Inches)", ylab="Weight (Pounds)")
### -----------------------------------------------------------
### Q2
lm(wcgs$weight0~wcgs$height0)
### -----------------------------------------------------------
### Q3
regress("mean", weight0~height0, data=wcgs)
### -----------------------------------------------------------
### Q6
wcgs$heightcm <- (2.54)*wcgs$height0
regress("mean", weight0~heightcm, data=wcgs)
mean(wcgs$heightcm)
```