

Exercises

Eleanor Howe

March 15, 2016

1. What is the average height of the sons (don't round off)?

```
mean(father.son$sheight)
```

```
## [1] 68.68407
```

2. One of the defining features of regression is that we stratify one variable based on others. In Statistics, we use the verb “condition”. For example, the linear model for son and father heights answers the question: how tall do I expect a son to be if I condition on his father being x inches? The regression line answers this question for any x .

Using the father.son dataset described above, we want to know the expected height of sons, if we condition on the father being 71 inches. Create a list of son heights for sons that have fathers with heights of 71 inches, rounding to the nearest inch.

What is the mean of the son heights for fathers that have a height of 71 inches (don't round off your answer)? Hint: use the function round on the fathers' heights.

```
father.son %>%  
  filter(round(fheight) == 71) %>%  
  subset(select=sheight, drop=T) %>%  
  mean
```

```
## [1] 70.54082
```

3. We say a statistical model is a linear model when we can write it as a linear combination of parameters and known covariates, plus random error terms. In the choices below, Y represents our observations, time t is our only covariate, unknown parameters are represented with letters a, b, c, d and measurement error is represented by ε . If t is known, then any transformation of t is also known. So, for example, both $Y = a + bt + \varepsilon$ and $Y = a + bf(t) + \varepsilon$ are linear models. Which of the following cannot be written as a linear model?

- A) $Y = a + bt + \varepsilon$
- B) $Y = a + b\cos(t) + \varepsilon$
- C) $Y = a + b^t + \varepsilon$
- D) $Y = a + bt + ct^2 + dt^3 + \varepsilon$

Answer: C

4. Suppose you model the relationship between weight and height across individuals with a linear model. You assume that the height of individuals for a fixed weight x follows a linear model $Y = a + bx + \varepsilon$. Which of the following do you feel best describes what ε represents?

- A) Measurement error: scales are not perfect.
- B) Within individual random fluctuations: you don't weigh the same in the morning as in the afternoon.
- C) Round off error introduced by the computer we use to analyze the data.
- D) Between individual variability: people of the same height vary in their weight.

Answer: D