# Open-ended Practice Test (Ch 1-5)

Here is the code to get the data:

PRAISESTUDY <- read.csv("http://bit.ly/muellerdweck\_study1", header=TRUE)

Story time! Mueller and Dweck did a study with kids, giving them IQ test questions (which were framed as “puzzles”) and praising them differently (that is, giving them different feedback).

At the very beginning of the study, kids did puzzle set #1 (PS1). After that (no matter what their score), kids got either EFFORT feedback (e.g., “hey, you did great; you must have tried hard!”) or INTELLIGENCE feedback (e.g., “hey, you did great; you must be so smart!”). After getting feedback, students did a very difficult puzzle set (PS2) and then they got a third puzzle set (PS3) that was as difficult as the first one.

1. What is the observational unit in this data frame? Could we put the average school’s PS3 score into this data frame? Why or why not?
2. Let’s say we want to create a new variable called **PSDIFF**. We want to figure out how many more problems a kid solved in **PS3** than in **PS1** and save that number as a variable called **PSDIFF**. Write code to create such a variable in R. (Hint: Should you save in a data frame? Which data frame?)
3. Let’s say we wanted to write a word equation to explain the variation in **PSDIFF**. Maybe **FEEDBACK** is an important explanatory variable. What would the word equation look like?
4. What percentage of kids improved over the course of the study (answered more questions right on PS3 than PS1)? Write R code that would give you the answer to this question.
5. How would you create a plot to look at the distribution of **PSDIFF**?
6. How would you create a plot to explore whether **FEEDBACK** explains some of the variation in **PSDIFF**?
7. Why is the mean a good model for the distribution of **PSDIFF**?
8. The median of **PSDIFF** is \_\_\_\_\_. In what way does that number represent the center of the data?
9. If you create an empty model of **PSDIFF**, what would it mean to have an “empty model”?
10. What R code would you use to fit the empty model for **PSDIFF**?
11. If you create a histogram for **PSDIFF**, what R code would create a blue line that indicates the mean?
12. If the mean of **PSDIFF** is \_\_\_ and a given observation has a **PSDIFF** of \_\_\_, what is the data? What part of the notation , would you use to represent the mean? Which part of the notation represents a particular observation?
13. If the mean of **PSDIFF** is \_\_\_ and a given observation has a **PSDIFF** of \_\_\_, what is the residual?
14. What will the empty model predict for each observation’s **PSDIFF**?
15. Imagine you make three histograms: one for **PSDIFF**, one for the predicted values based on the empty model for **PSDIFF**, and one for the residuals. Which two distributions will have a similar shape?
16. If we’ve used our **PSDIFF** data frame to calculate the exact mean and the exact size of the errors, what have we NOT calculated?
17. What are the three different symbols (notation) to represent the mean of the sample?
18. What notation can be used to represent the mean of the population?
19. What is the “population” that the **PSDIFF** empty model is trying to understand?
20. Explain what you would look for in your plot that would show that **FEEDBACK** does indeed explain some of the variation in **PSDIFF**.
21. If you look at the histogram for **PSDIFF**, you’ll see there is just one kid who got 6 more correct on the third set of problems than the first set. If you decided to ignore that kid’s data, which will be most affected by this change: the mean, the median, or the minimum? Will the maximum be affected?
22. Print out the favstats for **PSDIFF**. If you wanted to predict what the next randomly chosen observation and minimize the SS between your predictions and these data, what would be your prediction?
23. A researcher has an intuition that how much the kid enjoys these puzzles (**ENJOY**) is associated with **PSDIFF**. Which visualization will help use explore this idea?
24. A researcher has an intuition that how much the kid enjoys these puzzles (**ENJOY**) is associated with **PSDIFF**. Write a word equation to represent this idea.
25. Explain what you would look for in your plot that would show that **ENJOY** does indeed explain some of the variation in **PSDIFF**.
26. If the mean of **PSDIFF** is .13. what will the empty model predict for a 5th grader who reported very high levels of enjoyment?
27. If the mean of **PSDIFF** is .13 and a given 5th grader has got 3 more right on the third set, what number is the data? What number is the model? What number is the residual?
28. If the mean of **PSDIFF** is .13 and a given 5th grader has got 3 more right on the third set, what notation could be used to represent the data?
29. What notation represents the model? There are 3 ways of representing this simple model -- the mean.
30. What notation represents is the residual? There are 4 ways of representing the residual from the simple model.
31. What notation can be used to represent the mean of the population? There are 2 ways.