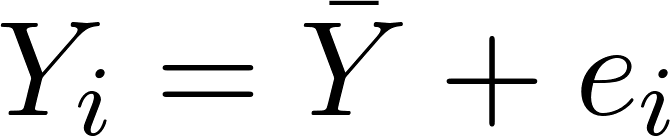
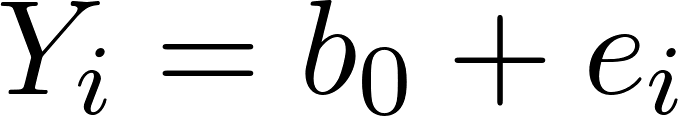
Classwork 5.1

* Diagrams can be written in any way… but I like the way I drew the distribution triad in week 1 better…
  + Either way you need three main parts of the distribution triad
    - DGP
      * Could be like our story/theory
      * Could be random
    - Data (or sample)
    - Simulated random process (sampling distribution)
* If I wanted to write my theory explaining variation in **FiveVegetables** with information about whether the state went for McCain or Obama in a word equation, how would I write it? (PE)
* Story time! Mueller and Dweck did a study with kids, giving them IQ test questions, praised them differently -- “hey, you did great; you must be so smart!” or “hey, you did great; you must have tried hard!” Now I’m not going to give you a lot of details about the study yet… but students did three sets (10 puzzles in each set).
  + Does **FEEDBACK** explain any of the variation in **PS1**? **PS2**? **PS3**?
    - Which one looks like **FEEDBACK** explains some of the variation?  (PE)
    - Review definition of “explaining variation”
    - What are we looking at when we make this decision? (PE)
    - How would we write a word equation for when **FEEDBACK** explains some of the variation? How would we write a word equation for when **FEEDBACK** does not explain any of the variation?
    - Do PE questions at the end of this sequence
  + Let’s learn a little more about the study. Study’s procedure: **PS1**, verbal feedback, **PS2** (difficult), **PS3**
    - Feedback CAN’T have influenced **PS1** because **PS1** was what they did BEFORE the researchers praised them! So why does it look like these two groups are slightly different?
    - R practice: Can you get rid of gaps between bars? (PE)
    - What do you think the boxplot of **PS1** split up by **FEEDBACK** should look like? (PE)

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

PRAISESTUDY <- filter(PRAISESTUDY, FEEDCODE < 2)

* Explaining variation in with feedback
  + Does FEEDBACK explain any of the variation in PS1? PS2? PS3?
  + Which one looks like FEEDBACK explains some of the variation?
  + Review definition of “explaining variation.”
  + What are we looking at when we make this decision? How would we write a word equation for when FEEDBACK explains some of the variation? How would we write a word equation for when FEEDBACK does not explain any of the variation?
* Modeling PS1 with a number
  + Review **data = model + error** with California metaphor
  + If we needed to model the distribution of PS1 with a number, what number should we pick?
  + What are the ways we can get median and mean in R? (favstats, mean, median, lm)
  + Putting mean and median into a histogram
    - Also into a faceted histogram
  + Mean versus Median
    - How is the mean the middle?
    - How is the median the middle?
    - Can you prove that the mean really is the “middle” (in it’s own special way) if you had 77 pieces of paper? What if you were using R?
    - Can you prove that the median really is the “middle” (in it’s own special way) if you had 77 pieces of paper? What if you were using R?
  + Using an "empty model"
    - predict()
    - resid()
    - What are these two functions doing?
  + Linking representations of our empty model
    - word equations & mathematical notation
    - x-bar versus b-sub-1
      * Data = model + error
      * Notation for data ([](https://www.codecogs.com/eqnedit.php?latex=Y_i%20%3D%20%5Cbar%7BY%7D%20%2B%20e_i%250) versus [](https://www.codecogs.com/eqnedit.php?latex=Y_i%20%3D%20b_0%20%2B%20e_i%250))
      * Notation for the DGP/population ([](https://www.codecogs.com/eqnedit.php?latex=Y_i%20%3D%20%5Cbeta_0%20%2B%20%5Cepsilon_i%250))

Extra Practice

* More R practice:
  + Filter out intelligence so we can just compare control versus intelligence
  + Let’s say FEEDCODE wasn’t already labeled in FEEDBACK variable. Can you create a column where FEEDCODE is labeled with the right categories?
* Without filter: Does **FEEDBACK** explain any of the variation in **PS1**? **PS2**? **PS3**?
  + Which one looks like **FEEDBACK** explains some of the variation?  (PE)
  + Review definition of “explaining variation”
  + What are we looking at when we make this decision? (PE)
  + How would we write a word equation for when **FEEDBACK** explains some of the variation? How would we write a word equation for when **FEEDBACK** does not explain any of the variation?
  + Do PE questions at the end of this sequence