ISS 305:002
Evaluating Evidence:
Becoming a Smart Research Consumer

6. Problems of description

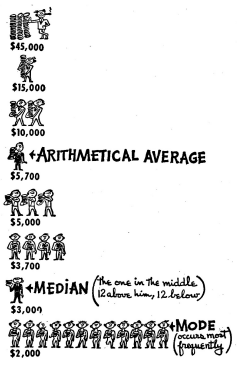
Reminder: Turn on your I<CLICKER

Slippery "averages"

- **People often choose one way rather than another**
  - **Carelessly**
  - **Arbitrarily**
  - **To "grind a particular ax" (to mislead you)...**

Slippery "averages"

- Q: Why would one choose the mean vs. the median?
  - **To create an impression that scores are especially high or low.**
- Who might prefer to use the mean?
  - **Recruiters for the company. The company during labor negotiations.**
- Who might prefer to use the mode?
  - **The union during labor negotiations.**
- Who might prefer to use the median?
  - **Someone without an ax to grind.**
- Again, remember that if the distribution were symmetric, Mean = Median
  - Is the distribution symmetric here?



Slippery "Averages" – NFL Salary Example

- Are NFL players paid a lot?
- What's the "average" salary in the NFL?
- For the 2013-2014 season, here were the figures (using base salary):
  - N = 2454 players (Included players cut before the season).
  - Mean = arithmetic average = $1,064,704
  - Median = $555,000
  - Mode = $405,000
  - SD = $1,500,218

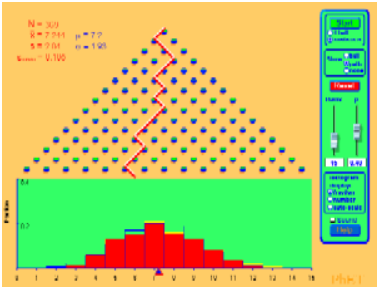
Slippery "Averages" – NFL Salary Example

- Why so different?
  - **There are a few highly paid superstars.**
  - **Top 15 make $167,000,000 combined.**
  - **But many players make at or near the League's minimum ($405,000)**
    - **N = 643 (26% make minimum)**
    - **42% make $500,000 or less**

- **You get very different answers, depending on how you summarize the salaries**
- MORAL: **To interpret any "average", you need to know what it really is – mean, median, or mode.**

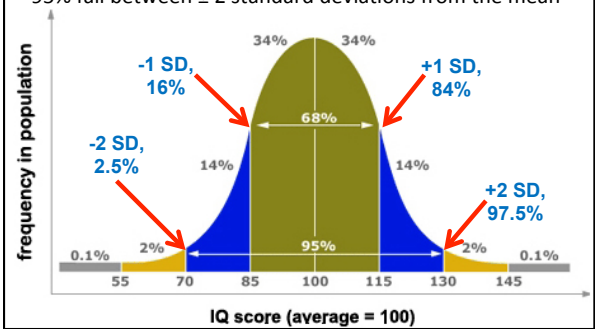
The Normal Distribution

- Many attributes (IQ, height, etc.) distributed "normally"= bell shaped curve
- Why?
- **If there are lots of small, independent effects that combine to determine the variable, it will be distributed normally**



The Normal Distribution

- 68% fall between ±1 standard deviation from the mean
- 95% fall between ± 2 standard deviations from the mean

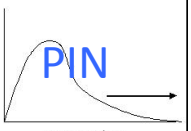


The Normal Distribution

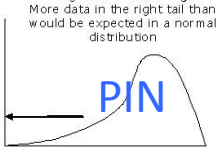
- So, if one can assume a normal distribution, and knows the mean and standard deviation, one can make rough percentile estimates
  - e.g., assume Normal distribution with Mean = 12, σ = standard deviation = 3
    - what is percentile rank of a score of 9?
    - of 18? **+2 SD; 97.5%**
    - of 16? **+1.33 SD, 88.5%**
    - **97.5 – 84 = 13.5 ; 13.5*.33 = 4.5 ; 84 + 4.5 = 88.5**
- If for an Exam. Mean=42.61 and Standard deviation=9.8, so if normally distributed
  - -2 sd ~ **22.9**
  - -1 sd ~ **32.8**
  - 0 sd = **mean= 42.6**
  - +1sd ~ **52.4**
  - +2sd ~ **62.3**
  - knowing this, one can approximately estimate the percentile rank of <u>any</u> score
    - e.g., score of 53 is near +1 sd, so something just over 84% (e.g., let's guesstimate around 85%. This student did better than ~85% of the students in the class.

Other useful distributional features

- Skewness (or skew)
  - Summarizes the degree of asymmetry
  - Skewness = 0 means that the distribution is symmetric
  - Skewness > 0 means that the distribution has a "tail" in the positive direction
    - Or, majority of the scores are at the lower end of the distribution
    - Floor Effect
  - Skewness < 0 means that the distribution has a "tail" in the negative direction
    - Or, majority of the scores are at the higher end of the distribution
    - Ceiling Effect



Positive Skew
Elongated tail at the **right**
More data in the right tail than would be expected in a normal distribution



Negative Skew
Elongated tail at the **left**
More data in the left tail than would be expected in a normal distribution

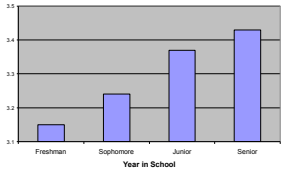
Problems with Proportions or Percentages

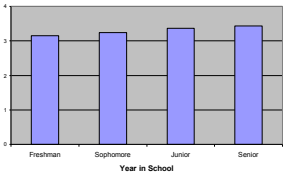
1. When the base is small.

- E.g., survey finds that 50% of those interviewed were in favor of a tax increase
- **Same meaning when base is 2 interviews vs. 200 interviews?**
- Small changes in numbers make big changes in the % when the base is small
  - **Misrecording just one person [as anti-tax-increase] changes the estimate from 50% to 100% for the 2 person base;**
  - **from 50% to 50.5% for the 200 person base**
- If base is small, usually more informative to see the numbers themselves rather than the %s
  - **how many interviewed and how many for tax increase**

Graphical illusions/deceptions

- Suppose I'm a freshman at MSU, and I don't do well my 1st year.
- My parents are concerned, but I tell them, "don't worry, everyone struggles their first year, but things get better later"
- And I present them with graphical evidence on mean GPA at MSU by year of study
- What's wrong with my argument?
  - **Results may have more to do with who drops out than with real improvements**
  - **May reflect learning to navigate the system (e.g., find easy courses) than in better real learning**
  - **But graphical evidence is also deceptive.**
    - **Here's same data, graphed less deceptively**
    - **What's different and what difference does it make in the impression created?**





Graphical illusions/deceptions: The "One-dimensional Picture"

Using two (or even three) dimensional objects comparing two groups or conditions which are actually only being compared on a single dimension.

- Illustration
  - simple bar graph accurately conveys the difference in incomes in the two countries
  - but graphic does not. Why?
    - **bag on left is exactly twice as high as bag on the right, but**
      - **doubling on one dimension**
      - **quadruples on area, and**
      - **increases by 8 times for volume.**