ECON 340
Economic Research Methods

Div Bhagia

Lecture 3
Variance, Standard Deviation, Z-Score

NYT Article: 2016 Election Predictions

- Summarize the main issue being discussed in the article.
- What were the three types of errors identified in the article? What is the common thread across these errors?
- One of the fixes suggested in the article was “education weighting”. Which of the three errors would this fix and how?

NYT Article: 2016 Election Predictions

- Summarize the main issue being discussed in the article.
- What were the three types of errors identified in the article? What is the common thread across these errors?
- One of the fixes suggested in the article was “education weighting”. Which of the three errors would this fix and how?
- In general, how can we pick a sample that is representative of the population to avoid having to reweight?

Another Example

- We want to estimate the average starting salary of students at a university that has only two majors
- Half of the students are *Business* majors, while the other half are *Engineering* majors
- Randomly select 100 Business students and 100 Engineering for a survey
- Response rate among Business students is 100%, while it 50% for engineering students

How can we use weighting to adjust for this?

Last Class

How to describe variables?

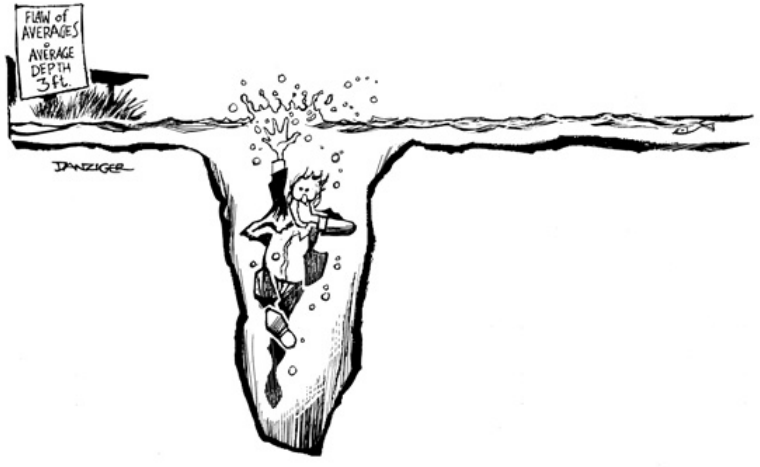
- Empirical Distribution
- Measures of central tendency: mean and median

μ : population mean, $\bar{X}$: sample mean

Two equivalent formulas:

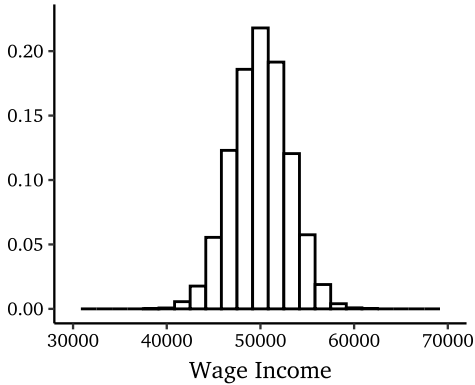
$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \bar{X} = \sum_{k=1}^{K} f_k X_k$$

Measures of central tendency are not enough!



Where would you want to live?

Mushroom Kingdom



Mean = Median= $50,000

Bowser's Kingdom

Mean = Median= $50,000

Deviations from the Mean

- Even with identical mean and median, the two countries are not identical.
- There is certainly more *dispersion* or *variability* in income in Bowser's Kingdom.
- More observations are *further from the mean* in Bowser's Kingdom.
- *What could be a potential statistic that could capture this?*

Deviations from the Mean

One option: average deviations from the mean. *Will this work?*

X_i	$X_i - \mu$
5	
5	
10	
10	
20	

Deviations from the Mean

Why does this not work? Remember from the last class:

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \bar{X}) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} \qquad (Why?)\\ &= \sum_{i=1}^{n} X_i - n\bar{X}\\ &= n\bar{X} - n\bar{X} = 0 \qquad (Why?)\end{aligned}$$

Deviations from the Mean

Why does this not work? Remember from the last class:

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \bar{X}) &= \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} \qquad (Why?) \\ &= \sum_{i=1}^{n} X_i - n\bar{X} \\ &= n\bar{X} - n\bar{X} = 0 \qquad (Why?)\end{aligned}$$

Can you think of a way to construct a statistic that would capture variation around the mean?

Variance and Standard Deviation

Population Variance

$$\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_x)^2$$

Sample Variance

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Standard Deviation

$$\sigma_X = \sqrt{\sigma_X^2} \qquad S_X = \sqrt{S_X^2}$$

Variance and Standard Deviation

Back to our example.

X_i	$(X_i - \mu)$	$(X_i - \mu)^2$
5	-5	
5	-5	
10	0	
10	0	
20	10	
50	**0**	

Variance with Grouped Data

Population Variance

$$\sigma_X^2 = \sum_{k=1}^{K} f_k (X_k - \mu_X)^2$$

Sample Variance

$$S_X^2 = \frac{n}{n-1} \sum_{k=1}^{K} f_k (X_k - \bar{X})^2$$

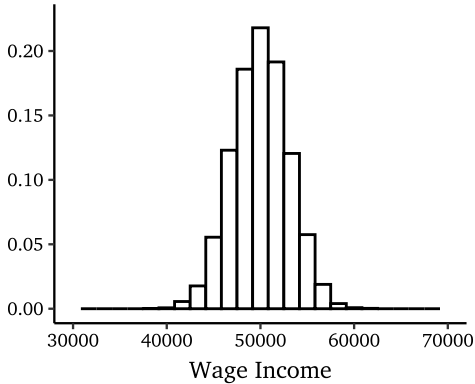
Variance with Grouped Data

In our example: 5, 5, 10, 10, 20. Present this as:

X_k	f_k	$f_k X_k$	$(X_k - \mu)^2$	$f_k(X_k - \mu)^2$
5	2/5			
10	2/5			
20	1/5			
Total				

Where would you want to live?

Mushroom Kingdom



Mean = Median= $50,000
SD= $3,000

Bowser's Kingdom

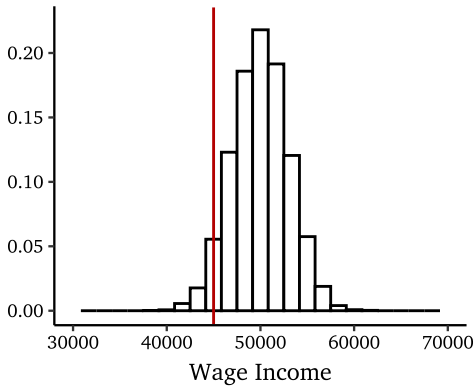
Mean = Median= $50,000
SD= $5,000

Where would you want to live?

- If we don't know where we will end up in the income distribution, some of us might prefer the Mushroom Kingdom since it is unlikely we would earn very little.
- For the same reason, some of us might like Bowser, as it is more likely that one could make a lot.
- But what if Luigi has a job for you as a plumber in both locations, and you will earn $45,000 regardless of where you end up? Are you now indifferent between the two?

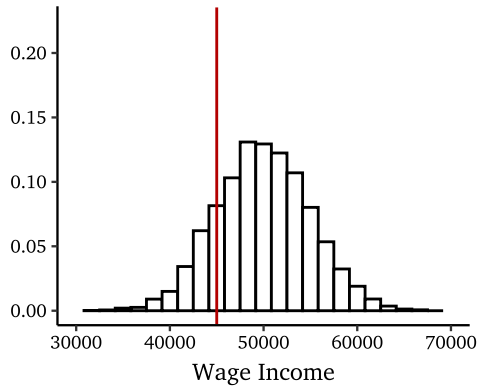
Where would you want to live?

Mushroom Kingdom



Mean = Median= $50,000
SD= $3,000

Bowser's Kingdom



Mean = Median= $50,000
SD= $5,000

Z-Score

We can calculate the Z-Score to capture how many standard deviations (σ) away from the mean (μ) a specific observation is.

$$Z = \frac{X - \mu}{\sigma} \quad \rightarrow \quad X = \mu + Z.\sigma$$

Example: $\sigma_{MK} = 3000$, $\sigma_{BK} = 5000$

$$Z_{MK} = \frac{45000 - 50000}{3000} = -1.66 \qquad Z_{BK} = \frac{45000 - 50000}{5000} = -1$$

Z-Score

- Someone who earns $45,000 in the Mushroom Kingdom is 1.66 *standard deviations* below the mean.
- While someone who earns $45,000 in the Bowser's Kingdom is 1 *standard deviation* below the mean.
- Here, Z-score is informative about how many people are there between someone who earns $45,000 and the average person
- More generally, Z-score tells us the relative position of an observation in the distribution

Things to do next

- Make sure you are staying up to date with the class; notes complement the slides
- Please utilize my office hours
- Let me know your research partner by this Thursday
  - You can self-sign up on Canvas by going to *People* and then clicking on the *Research Project Group* tab or email me
  - If you do not inform me of your partner or indicate that you will be working alone by Thursday, I will assign a partner for you.
- Coming up: Problem Set 1 (due next week on Tues)