# Applied Statistical Analysis

## EDUC 6050

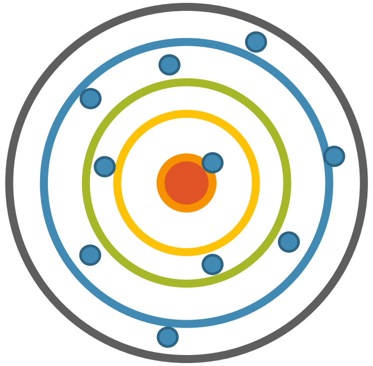
### Week 4

Finding clarity using data

# Today

1. Statistics terminology continued (hypothesis testing, descriptive and inferential statistics, effect sizes, confidence intervals, Type I and II errors)
2. Chapters 4, 5, and 6 in Book
3. Statistical Organizer due

# Review



1. The figure to the right is reliable/unreliable and valid/invalid.
2. When should you use the mean? What about the median?
3. What does the standard deviation tell us?
4. Can we obtain a standard deviation with nominal data?

# Reliability and Validity

**Reliability**: the consistency of the measure

**Validity**: does it measure what we think it measures?

Reliable
Valid

Reliable
Not Valid

Not Reliable
Valid

Not Reliable
Not Valid

# Correlation and Experimentation

| Correlation | Experimentation |
| --- | --- |
| observational, no treatment/intervention | treatment/intervention (best if groups are randomized) |

**Correlation does not imply causation**

**AND**

**Correlation does not imply it isn’t causal**

# Central Tendency

| Measure | When to use it |
| --- | --- |
| **Mean** | With interval/ratio data that are ~normally distributed |
| **Median** | With ordinal data<br>With interval/ratio data that are skewed or have outliers |
| **Mode** | With nominal data |

**Outliers** = points far from the other points

# Variability

| Measure | When to Use | Possible Values |
|---|---|---|
| Range | Ordinal, Interval, Ratio | 0+ |
| Standard Deviation | Interval, Ratio | 0+ |

# Hypothesis Testing

| Null Hypothesis | Alternative Hypothesis |
| --- | --- |
| No effect | Effect exists |

Essentially, analyze data and see if null hypothesis seems plausible

- If not plausible, we believe the alternative
- If plausible, we assume there is no effect

# Hypothesis Testing

## “P-Values”

- The probability of observing an effect that large or larger, given the null hypothesis is true.
- It is trying to tell us if an effect exists in the population

**Usually a p-value < .05 is considered “statistically significant”**

# Hypothesis Testing

## P-Values

- Researchers rely on them too much (Cumming, 2014)
- Effect sizes should be used with them
  - We need to highlight that effect sizes are uncertain
  - A “significant” finding may not be meaningful or reproducible

# Reading

## Questions from Chapter 4, 5, and 6 that you’d like us to cover today?

# Z-Scores

## Important Point:

- There are distributions of single scores
- There are distributions of statistics
    - This is generally in reference to the sample mean

## Chapter 4 is about single scores

# Z-Scores for an Individual Point

$$z = \frac{X - \mu}{\sigma}$$

**Tells us:**

- If the score is above or below the mean
- How large (the magnitude) the deviation from the mean is to other data points

# Z-Score Examples

$$z = \frac{X - \mu}{\sigma}$$

1. M = 20, Score = 10, SD = 10, z = ?
2. M = 5, Score = 5, SD = 1, z = ?
3. M = 5, Score = 6, SD = 1, z = ?
4. Z = 1, Mean = 1, SD = 1, M = ?
5. Z = -1, Mean = 0, SD = 0.5, M = ?

# Z-Score Interpretations

- If the score is + then above the mean
- If the score is - then below the mean
- If score is more than ± 1 then score is considered "atypical"
- If score is less than ± 1 then score is considered "typical"

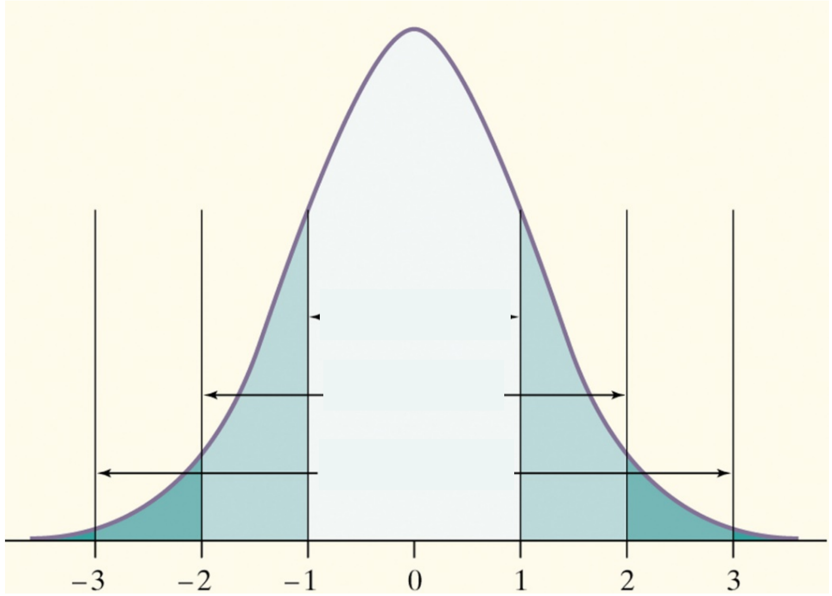
The z tells us more information than just a score. Why?

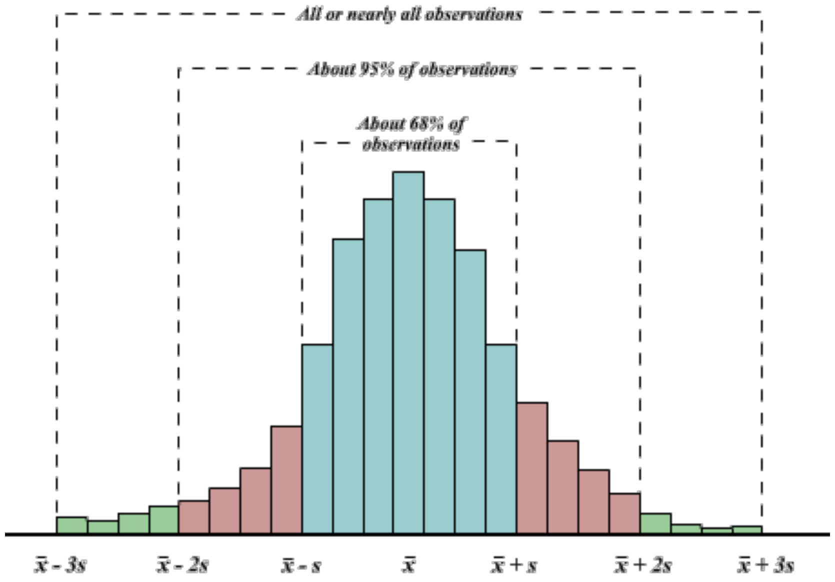
# Z-Score and the Standard Normal Curve

**The 68-95-99.7 Rule**

In the Normal distribution with mean $\mu$ and standard deviation $\sigma$:

- Approximately **68%** of the observations fall within $\sigma$ of $\mu$.
- Approximately **95%** of the observations fall within $2\sigma$ of $\mu$.
- Approximately **99.7%** of the observations fall within $3\sigma$ of $\mu$.



# Z-Score and the Standard Normal Curve

## So...

- We can use the same idea to estimate the probability of scoring higher or lower than a certain level

Example: If the scores on an exam have a mean of 70, an SD of 10, we know the distribution is normal, what is the probability of scoring 90 or higher.

# Normal Distributions and Shading

Page 101 of the book

# Review of Z-Scores (Chapter 4)

1. What does a z-score about an individual point tell us?
2. Is it possible to make a specific probability statement about a z-score if the distribution is normal?
3. What proportion of scores are between z-scores of 0 and 1? (hint: use shading and the appendix)

# Break Time

# Distribution of Sample Means

## Important Point:

- There are distributions of single scores
- There are distributions of statistics
    - This is generally in reference to the sample mean

## Chapter 5 is about distributions of statistics

# Distribution of Sample Means

Inferential statistics is all about using the sample to infer population parameters

But the sample is almost certainly going to differ from the population (at least a little)

So what if we took 5 different samples (or 10, or 50, etc.). Will each sample have the same mean?

# Distribution of Sample Means

Inferential statistics is all about using the sample to infer population parameters

But the sample is almost certainly going to differ from the population (at least a little)

So ... es

(o ... le

http://shiny.stat.calpoly.edu/Sampling_Distribution/

# Standard Error of the Mean

## “SEM” or “SE”

- Depends on sample size (bigger sample, smaller SEM)
- Tells us, *if we were to collect many samples*, how much the sample means would vary

$$SEM = \frac{\sigma}{\sqrt{N}}$$

# Since we don’t want to take lots of samples...

## We use statistical theory! (or “the magic of math”)

- Central Limit Theorem
  - Tells us the shape (normal), center ($\mu$) and spread (SEM) of the distribution of sampling means
- Law of Large Numbers
  - As N increases, the sample statistic is better and better at estimating the population parameter

# The Z for a Sample Mean

$$Z_{Mean} = \frac{Mean - \mu}{SEM}$$

**This is important because of what we will talk about in Chapter 6**

- Hypothesis Testing with Z Scores

# The Z for a Sample Mean

$$Z_{Mean} = \frac{Mean - \mu}{SEM}$$

1. N = 100, Mean = 10, $\mu$ = 5, $\sigma$ = 5, $Z_{Mean}$ = ?
2. N = 100, Mean = 2, $\mu$ = 0, $\sigma$ = 10, $Z_{Mean}$ = ?
3. What is the probability of having a mean greater than 10 for the first example?

# Break Time

# Hypothesis Testing with Z Scores

## Hypothesis Testing uses Inferential Statistics

- Is there evidence that this sample (maybe because of an intervention) is different than the population?

# Hypothesis Testing with Z Scores

## We’ll use a 6-step approach

**We’ll use this throughout the class so get familiar with it**

1. Examine Variables to Assess Statistical Assumptions
2. State the Null and Research Hypotheses (symbolically and verbally)
3. Define Critical Regions
4. Compute the Test Statistic
5. Compute an Effect Size and Describe it
6. Interpreting the results

# Hypothesis Testing with Z Scores

Because assessing z-scores and t-tests are so similar, we will talk about both next week

Read Chapter 7

# Review of Sample Mean Distributions (Chapter 5 and Intro to 6)

1. Why is understanding the distribution of sample means important?
2. What does the standard error of the mean tell us?
3. How would we get a smaller SEM?
4. What are the steps in the 6-step approach?

# Another look at Jamovi and Excel

## Standardizing (getting z-scores)

# Questions?

# Next week:

1. Hypothesis Testing with Z Scores (continued)
2. Chapters 6 and 7 in Book
3. Keep updating your Statistical Organizer