Sampling Distributions

Sampling Distribution of the Mean & Hypothesis Testing

Sampling

• Remember sampling?
  – Part 1 of definition
    • Selecting a subset of the population to create a sample
    • Generally random sampling—using randomization to identify a sample
  – Part 2 of definition
    • Sample used to infer qualities or characteristics of the population
    • How do we do that?
Sampling: Hows & Whys

• First, it’s important to realize that simple comparisons of sample values to populations values is meaningless
  – $\mu = 3$ compared to $x-bar = 4$
    • Measure of course satisfaction
  – Different values, yes… but are they different from one another in a meaningful way?

Sampling: Hows & Whys

• Second, we know that we can identify infrequent or exceptional values for any normally distributed variable
  – Thus, for any given mean value with a given standard deviation, we can identify values that fall outside of the 95% CI
Applied Example

• Imagine we’re interested in student attitudes regarding the General Psychology Course
  – Score of general satisfaction with course
    • 0 = poor to 6 = excellent
  – Collect from every student in the fall semester
    • 1080 students questioned
    • \( \mu = 3 \)
    • \( \sigma = 1.34 \)

Applied Example

• Population of 1080 students
  \[ \mu = 3.00 \]
  \[ \sigma = 1.34 \]
Caveat

• Generally, we would NOT know population values…
  – Depending on the population of interest, it may be impossible to determine population values
  – Contrived example to illustrate a concept

Applied Example

• Sample of 10 students from population of 1080 students
  \( x-bar = 3.00 \)
  \( \sigma = 1.25 \)
Applied Example

• Sample of 10 students from population of 1080 students
  \[ \bar{x} = 3.90 \]
  \[ \sigma = 1.37 \]

Applied Example

• Sample of 50 students from population of 1080 students
  \[ \bar{x} = 3.02 \]
  \[ \sigma = 1.41 \]
Applied Example

- Sample of 50 students from population of 1080 students
  \[ x-bar = 3.34 \]
  \[ \sigma = 1.45 \]

- Sample of 100 students from population of 1080 students
  \[ x-bar = 2.94 \]
  \[ \sigma = 1.24 \]
Applied Example

- Sample of 100 students from population of 1080 students
  - $x\bar{=}=3.02$
  - $\sigma=1.43$

Applied Example

- Sample of 540 students from population of 1080 students
  - $x\bar{=}=2.99$
  - $\sigma=1.36$
Applied Example

- 540 Students
- 1080 Students

Sampling Distributions

- For small samples:
  - Shape of sample distribution differed greatly from that of the population
  - Values of $x$-bar differed from $\mu$
  - Values of $s$ differed from $\sigma$
- For large samples ($n > 100$):
  - Shape of sample distribution and values of $x$-bar and $s$ similar to population values
Sampling Error

• But why do small samples look different from the population of origin?
  – Sampling error
    • Defined as variability due to chance differences between samples
    • Reflects degree to which chance variability between samples influences statistics, changing them from “expected” population values

Sampling Error (cont.)

• Sampling Error (cont.)
  – RANDOM variance—can only be controlled through the collection of large samples (reduce chance error)
  – NOT due to experimenter mistakes, confounded variables, or design flaws—outside of our control
    • …excepting, of course, sample size

• The take home lesson is…
Sampling Distributions

• Probably the most important implication of the sampling process is the concept of the *sampling distribution*
  – Sampling distributions tell us:
    • Degree of variability we should expect from repeated samplings of a population as a function of sampling error
    • Tells us the values we should and should not expect to find for a particular statistic under a particular set of conditions

Sampling Distributions

• Typically derived mathematically, you won’t normally need to produce your own sampling distributions
  – Sampling distribution of the mean
    • The distribution of obtained means obtained from repeated samplings
Sampling Dist. Of the Mean

Population

Sample 1   Sample 2   Sample 3   Sample 4   Sample n

\[ \bar{x} \quad \bar{x} \quad \bar{x} \quad \bar{x} \quad \bar{x} \]

Plot of Sample Means

Sampling Dist. Of the Mean: Example

- Population of 1080 students
- Draw 50 samples of 50
- Obtain the mean for each sample
- Plot the distribution of means
- Expect a fairly normal distribution of means
Sampling Dist. Of the Mean: Example

- Sampling Distribution of Mean Course Satisfaction Scores for \( n = 50 \)

- Histagram
  - \( n = 50 \)
  - \( x-bar = 2.995 \)
  - \( s = .15 \)
  - range = 2.70 → 3.32
  - 95% CI = 2.70 → 3.29

Sampling Dist. Of the Mean: Example

- For this case, 1 score falls outside the expected boundaries
  - 3.32 → not expected in this sampling distribution
  - We might conclude then, if we have a sample with this mean, that the sample we drew does not come from the population of interest
Sampling Dist. Of the Mean: Reality

• Statistical tests use a similar process that I’ve described to produce sampling distributions of the mean
  – Larger sample sizes (essentially infinite)
  – Closer $n$ comes to $\infty$, closer sampling distribution will be to normal

Sampling Dist. Of the Mean: Reality

• When conducting statistical tests:
  – Compare our obtained value from our sample to the sampling distribution of the mean for the population
  – Look for extreme scores
• But where do these sampling distributions of the mean for the population come from?
  – Stay tuned to Chapter 7
Hypothesis Testing

• Sampling distributions inform the way in which we test our hypotheses
• Care only about sampling distributions because they allow us to test hypotheses
• Before exploring the process of hypothesis testing, need to understand types of hypotheses

Types of Hypotheses

• Hypothesis
  – Defined as an informed belief regarding the relationships between two or more variables
    • Social support → depression
    • Subliminal advertising → product sales
  – Must be an informed belief—guessing the relationships between variables is not a hypothesis
Types of Hypotheses

- **Research Hypothesis** (H₁)
  - *The hypothesis that we’re interested in testing with our experiment or study*
    - The beta blocker Atenolol reduces blood pressure
    - Group therapy reduces violent urges

- **Null Hypothesis** (H₀)
  - *The starting hypothesis, generally specifying no relationship between variables*
    - Atenolol has no effect on blood pressure
    - Group therapy has no effect on violent urges

---

The Null Hypothesis

- *Opposite* of what we’re trying to test!
- Why expect no differences?
  - Practical reason
    - Gives us a starting point
    - A place of comparison
    - Construct sampling distribution based on no effect (or difference) between groups of interest
The Null Hypothesis

• Why expect no differences?
  – Philosophical reason (Fisher)
    • We can never prove the truth of any proposition, only if it is false
      – “All swans are white”
      – “All squirrels are grey or red”
      – “Depressed individuals lack social supports”
    • 10,000:1
    • “Fail to reject” null hypothesis
      – Falsifying evidence may be right around the corner

The Null Hypothesis: Dissent

• Not all statisticians and mathematicians agree that we can’t accept the null hypothesis
  – Debate continues
  – “Benefit of the doubt”
  – Does “benefit of doubt” differ from “fail to reject”
Process of Hypothesis Testing

1. Identify a research hypothesis (H₁)
   • Specify hypothesis in quantitative terms
2. Identify null hypothesis (H₀)
   • Specify hypothesis in quantitative terms
3. Collect random sample of participants or events that can answer H₁

Process of Hypothesis Testing

4. Select rejection region and tail of the test
   • Rejection region (α)
     • The probability associated with rejecting H₀ when it is, in fact, false
     • Typically a low frequency value is selected
     • For Psychology, α = .05 for most situations
       • The 5% least frequent scores (-1.96 < z > 1.96)
   • “Tail” of test
     • Directionality: do we look at both ends of the distribution or only one end?
Tail of Test, $\alpha = .05$:

**Two-tailed**

Tail of Test, $\alpha = .05$:

**One-tailed**
Process of Hypothesis Testing

5. Generate sampling distribution of the mean assuming $H_0$ is true
   • This is done for us by the statistical test we choose to employ for the analysis
   • Essentially, we choose the test to use at this point

6. Given our sampling distribution:
   • What is the probability of finding a sample mean outside of our rejection region?
   • Conduct the statistical test
Process of Hypothesis Testing

7. On the basis of that probability:
   1. Reject $H_0$ when our sample mean falls within the boundaries of the rejection region
      • Supports $H_1$, but does not prove it
      • Remember, we can’t prove anything
   2. Fail to reject $H_0$ when our sample mean falls outside the boundaries of the rejection region
      • Supports $H_0$, but does not mean that $H_1$ is wrong...

Hypothesis Testing: Example

• You are a researcher testing the efficacy of a new antidepressant medication
• This is the first test of the new drug
• You decide to use two groups of depressed participants, 1 who receive the drug, 1 who receive no medication
• What is the process of hypothesis testing involved?
Hypothesis Testing: Example

1. H₁: The antidepressant medication will reduce the symptoms of depression
   • H₁: μₐ ≠ μₜ
   • H₁: μₐ < μₜ

2. H₀: The antidepressant medication will have no effect
   • H₀: μₐ = μₜ

3. Collect random sample of depressed individuals, assign randomly to 2 groups

4. Select:
   • Rejection region
     • α = .05
   • “Tail” or directionality
     • Probably want two-tailed
     • Uncertain of how the medication will work
     • Might be able to argue one-tailed
Hypothesis Testing: Example

5. Generate sampling distribution of the mean assuming $H_0$ is true
   • Select confidence intervals on $z$-distribution
6. Given our sampling distribution:
   • Conduct the statistical test

Sampling distribution of the mean:
   $\mu = 5$
   $\sigma = .5$
Sample of patients taking antidepressant:
   $\bar{x} = 6$
Hypothesis Testing: Example

• For depression scores:

\[ x = \mu \pm 1.96\sigma \]
\[ x = 5 - 1.96(.5) \]
\[ x = 5 - .98 \]
\[ x = 4.02 \]

\[ x = \mu \pm 1.96\sigma \]
\[ x = 5 + 1.96(.5) \]
\[ x = 5 + .98 \]
\[ x = 5.98 \]

• Thus, the 95% CI for depression scores is 4.02 to 5.98

Hypothesis Testing: Example

7. On the basis of that probability:
• 95% CI for depression = 4.02 to 5.98
• Obtained sample score = 6.00
• REJECT \( H_o \)!
  • Reject any value < 4.02
  • Reject any value > 5.98
  • Fail to reject values between 4.02 and 5.98
Hypothesis Testing: Example

\[ x = 4.02 \quad \text{and} \quad x = 5.98 \]

\[ \bar{x} = 6 \]

\[ \mu = 5 \]

Reject \quad \text{Fail to reject} \quad \text{Reject}

Alpha (\( \alpha \))

- Represents the rejection region—where we are correct to reject the \( H_0 \)
- In Psychology, the convention is to use .05
  - A 1 in 20 chance of rejecting \( H_0 \)
- Sometimes, we want to be really conservative about the conclusions we draw—reduce errors
  - Might select \( \alpha = .01 \)
  - A 1 in 100 chance of rejection \( H_0 \)
- .05 is a convention, not an absolute rule
## Error in Hypothesis Testing

- Hypothesis testing is not a perfect science
  - Errors occur
- Two types of errors can be made
- The probability of making an error is related to the probability of rejecting the $H_0$
- For example:

## Error in Hypothesis Testing: Example

- You are a researcher attempting to determine the intelligence of a struggling student
- You administer a test of IQ to the student
- What is the process of hypothesis testing involved?
Error in Hypothesis Testing: Example

1. \( H_1 \): The student’s IQ is below the population mean
   - \( H_1: \mu_s < \mu_p \)
2. \( H_0 \): The student’s IQ is not different from the population
   - \( H_0: \mu_s = \mu_p \)
3. Test the student

Error in Hypothesis Testing: Example

4. Select:
   - Rejection region
     - \( \alpha = .05 \)
   - “Tail” or directionality
     - Two-tailed—poor performance could be a matter of low IQ, poor motivation, chaotic household, or a number of factors unrelated to IQ
Error in Hypothesis Testing: Example

5. Generate sampling distribution of the mean assuming $H_0$ is true
   • Select confidence intervals on $z$-distribution

6. Given our sampling distribution:
   • Conduct the statistical test

---

Error in Hypothesis Testing: Example

Sampling distribution of the mean:
\[ \mu = 100 \]
\[ \sigma = 15 \]

Sample of patients taking antidepressant:
\[ \bar{x} = 63 \]
Error in Hypothesis Testing: Example

• For IQ:
\[ x = \mu \pm 1.96\sigma \]
\[ x = 100 - 1.96(15) \]
\[ x = 100 - 29.4 \]
\[ x = 70.6 \]
\[ x = 100 + 1.96(15) \]
\[ x = 100 + 29.4 \]
\[ x = 129.4 \]

• Thus, the 95% CI for IQ is 70.6 to 129.4

Error in Hypothesis Testing: Example

7. On the basis of that probability:
• 95% CI for population = 70.60 to 129.40
• Obtained sample score = 63.00
• REJECT H₀!
  • Reject any value < 70.60
  • Reject any value > 129.40
  • Fail to reject values between 70.60 and 129.40
Error in Hypothesis Testing:
Example

\[
x = 70.6 \quad x = 63 \quad x = 129.40
\]

Error in Hypothesis Testing

- So we conclude that the child has a lower than average IQ…
  - …but are we correct in drawing this conclusion?
Error in Hypothesis Testing

• Remember: we defined the rejection region, $\alpha$, as a probability
  – That means, sometimes our conclusions will be right, sometimes they will be wrong
  – In this case, the probability of an IQ score of 63 is low… but still possible, even in a sample of normal individuals!
    • Error associated with the test, sampling, etc.

Rationale of Hypothesis Testing

• Not testing extreme scores against the general population
• Testing if the sample score is so infrequent that we might conclude it comes from ANOTHER population
  – Population of normal IQ to low IQ individuals
Rationale of Hypothesis Testing

Population of normal IQ scores

Population of low IQ scores

Rationale of Hypothesis Testing

• Thus, the rationale isn’t that we look for extreme scores to conclude that the child’s IQ is outside the range of normal IQ
• We look at extreme scores to determine if the obtained value is so low that it probably comes from a population of individuals with low IQ
  – Note: probably—individuals from the population of normal IQ can score 63s as well!
Error in Hypothesis Testing

• In order to identify the types of errors and correct decisions we can make, we must look at two categories of behavior:
  – The decisions we make about $H_0$
  – The true state of $H_0$

• Note: we never really know the true state of $H_0$, this example is simply a theoretical way of looking at the quandary of error

Error in Hypothesis Testing

<table>
<thead>
<tr>
<th>Decision</th>
<th>$H_0$ True</th>
<th>$H_0$ False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td>Type I Error $p = \alpha$</td>
<td>Correct Decision $p = 1 - \beta = \text{Power}$</td>
</tr>
<tr>
<td>Fail to Reject $H_0$</td>
<td>Correct Decision $p = 1 - \alpha$</td>
<td>Type II Error $p = \beta$</td>
</tr>
</tbody>
</table>

True State of The World

- Type I Error “embarrassing type”
- Type II Error “unknown type”
Type I and Type II Errors

Trading Error for Error

• So, we might conclude that in order to avoid making “embarassing” Type I errors, we keep $\alpha$ as low as possible
  – $\alpha = .00000000000000000001$, anyone?
• Doing so, however, leads to a reduction in the power of the test—we may not make an error, but we won’t be right either!
  – Type II error increases—we lose the ability to find real differences when they occur
  – This is one reason we set $\alpha = .05$ (trade-off)
Trading Error for Error

- The Cliff’s Notes Version:
  - $\downarrow$ Type I error lead to $\uparrow$ in Type II error
  - $\downarrow$ Type II error lead to $\uparrow$ in Type I error

- Errors are inescapable
  - Seek to minimize error by using a compromise value of $\alpha = .05$

Power

- Power
  - An extremely important concept, defined as the ability to actually detect whatever you want to detect
  - Power is a measure of our tests ability to answer our hypotheses
    - Typically .80 or above is desirable (80% chance)
  - Outside the scope of this class, power has become a major consideration for psychologists in the past decade
    - See Cohen’s “A Power Primer” (1998)
Effect Size

- Since we can make errors in our hypothesis testing, statistical significance is often not sufficient
  - Effect size reflects, simply, the size of the effect or difference observed in our analysis
    - Standard deviation
    - Different measure for each test (stay tuned…)
- Statistical significance without meaningful effect size
- Meaningful effect size without statistical significance