

PS 211: Introduction to Experimental Design

Spring 2025 · Section C1

Lecture 8: Hypothesis Testing with Z Tests (Continued)

Updates and Reminders

- **Exam 1** grades have been posted.
- People did well! Median = 29.
- If you would like to go over your exam, please come to office hours next week.
- The next exam is somewhat soon, on **Thursday, October 16**.
- Coming up:
 - Tuesday (10/7): Lecture 9
 - Thursday (10/9): Exam 2 Review
 - Tuesday (10/14): No class (Monday schedule)
 - Thursday (10/16): Exam 2

Z Statistics for Distribution of *Means*

If we want to think about the z statistic for a **group**, we need to change a few things:

1. We use **means** instead of raw scores.
2. We calculate the mean and **standard error** for the distribution of means.
3. Then we calculate a z statistic for the sample mean.

When might this be useful?

Answer: All the time!

We often want to know if a sample is different from a population.

Example: Dating Profiles

Researchers are studying online dating profile ratings.

- They have a sample of 30 profiles from Rhode Island (RI).
- RI sample ($n=30$): $M = 2.84$
- U.S. population: $\mu = 2.5$, $SD = 0.833$
- **Is the RI sample mean different from the U.S. mean?**

What is the standard error for the distribution of means?

Answer: $SE = SD/\sqrt{n} = 0.833/\sqrt{30} \approx 0.152$

What does this standard error tell us?

Answer: The average sample mean computed from samples of size $n = 30$ will be about 0.152 away from the population mean (μ) of 2.5.

Example: Dating Profiles (Continued)

Review: Hypotheses

- **Null hypothesis (H_0):** There is no real difference. The populations from which samples are drawn are the same / equal. Any observed difference is due to chance.

The goal of hypothesis testing is to determine how likely our sample data would be if the null hypothesis were true.

We do statistics to either: reject or fail to reject the null hypothesis.

We do not "prove" that the null hypothesis is true. We can only fail to reject it if we don't have enough evidence against it.

Absence of evidence is not evidence of absence. There could still be a difference that we just didn't detect.

- **Research hypothesis or alternative hypothesis (H_1):** What the researcher expects to find. Sometimes states the direction of the effect (e.g., group A will have a higher mean than group B).

The research hypothesis is the hypothesis that would be true if the null hypothesis is false.

Hypothesis Testing

- We use statistics to formally test hypotheses.
- Statistical analyses are based on certain assumptions about the dataset.
- **Statistical assumptions** describe the ideal conditions for hypothesis testing. (More on this later!)
- We want to ensure these assumptions are (mostly) met so that we can make accurate inferences.

Six Steps of Hypothesis Testing

1. Identify the populations, comparison distribution, and assumptions
 - What populations are represented by the sample(s)?
 - What is the comparison distribution? Is it a distribution of means? Raw scores?
 - What assumptions, if any, do our data meet? (More on this later!)
 - This helps us choose the right statistical test!

Six Steps of Hypothesis Testing (Continued)

2. State null and research hypotheses

- Your hypotheses should be about the **population(s)**, not the **sample(s)**. Remember, we want to make inferences about populations!

3. Determine characteristics of **comparison distribution**.

- The comparison distribution = the distribution based on the null hypothesis.
- For z tests, we determine the mean and standard error of the comparison distribution and use this to calculate our test statistic.

Six Steps of Hypothesis Testing (Continued)

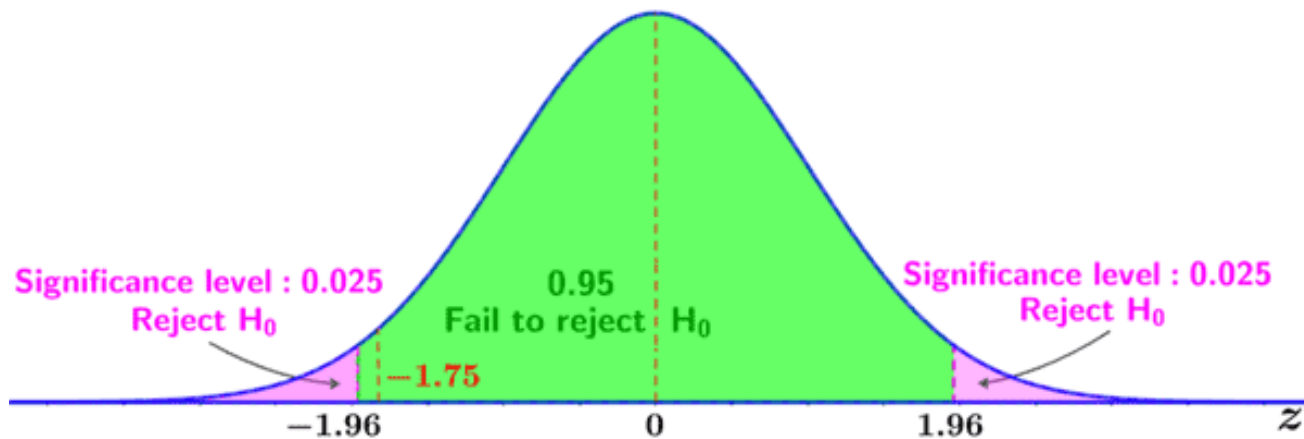
4. Determine **critical values**, or cutoffs

- The critical value defines the boundaries of the "extreme" scores. They determine how extreme the data must be (e.g., how large the z statistic must be) to reject the null hypothesis.
- The standard in psychological research is typically .05 or 5%.
- For a "two-tailed" test, this means we reject the null hypothesis if the test statistic falls in the upper or lower 2.5% of our distribution.

Six Steps of Hypothesis Testing (Continued)

4. Determine **critical values**, or cutoffs

- The critical region is the area in the tails of the distribution beyond the critical values. Values that fall in the critical region are considered extreme enough to reject the null hypothesis.
- The probabilities used to determine the critical values in hypothesis testing are called alpha levels.



Six Steps of Hypothesis Testing (Continued)

5. Calculate test statistic

- All the information from the previous steps is used to calculate the test statistic.
- We will focus on the z statistic for now, but these same steps apply for other test statistics (to be discussed later in the course!)
- Once we have our test statistic, we can compare it to the critical values from step 4 to determine whether the sample data is extreme enough to reject the null hypothesis.

Six Steps of Hypothesis Testing (Continued)

6. Determine whether you can reject the null hypothesis.

- Reject if: The test statistic is beyond the cutoff
- Fail to reject if: The test statistic is not beyond the cutoff
- This usually involves comparing p values (obtained probabilities) to α values (predetermined cutoffs).
- If we reject the null hypothesis, we say our results are statistically significant.

Statistically significant: Data are more extreme than what we would expect by chance if there truly were no actual difference.

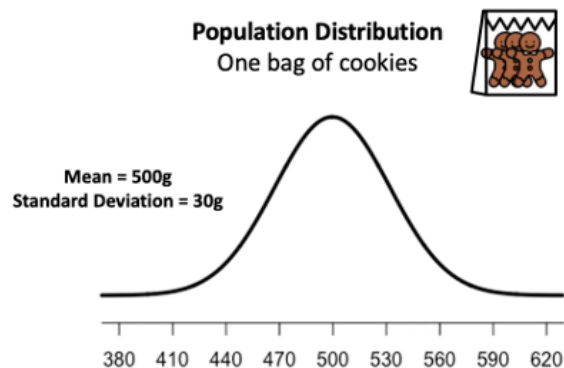
P values

- The p value is the probability of obtaining the observed results if the null hypothesis is true.
- A smaller p value indicates stronger evidence against the null hypothesis.



Example: Santa's Cookies

- Santa Claus claims that the average bag of gingerbread cookies weighs 500g, and the standard deviation across cookie bags is 30g.



You are not sure whether what Santa is saying is true. You decide to test a hypothesis that the average weight of one bag of cookies is LESS than 500g.

What are your null and research hypotheses?

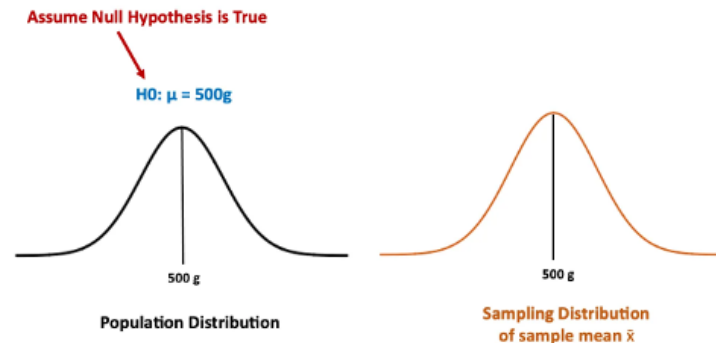
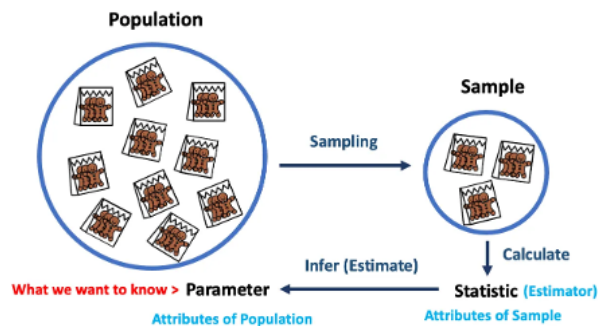
Answer:

H_0 : Average weight of one bag of cookies = 500g.

H_1 : Average weight of one bag of cookies < 500g.

Example: Santa's Cookies (Continued)

- You can't test every single bag of cookies, so you weigh a sample of Santa's bags.
- You collect 25 bags of cookies.
- The mean weight of the 25 bags is 485 g.
- If the null hypothesis is true, your sampling distribution should look like this:



- But your sample mean is only 485 g, 15 g below the mean expected value.

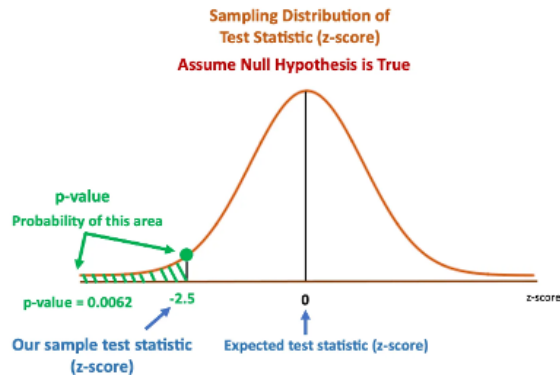
Example: Santa's Cookies (Continued)

Santa's Cookies: The Solution

- We can use the z statistic to address this question.
- First, we can use z scores to standardize our data. If the mean of our sample data equals the population mean (our null hypothesis), then our z score will be 0.
- Computing the z score:
 1. Compute difference in means: $485 - 500 = -15$
 2. Compute standard error: $30 / \sqrt{(25)} = 6$
 3. Compute z score for sample mean: $-15 / 6 = -2.5$.
- Next, we want to determine how extreme this z score is.
- We can use our z table to look up the percentile, which is $1 - .9938 = .0062$.

Santa's Cookies: The Solution (Continued)

- This means the probability of getting a value as extreme as ours is **.0062**. This is our *p value*.



- Assuming we had set our alpha level at .05, then we would reject the null hypothesis, because there is < 5% chance that we would observe this pattern of data if the null were true.

Be careful when moving back and forth between decimals and percents!

$$.05 = 5\%$$

$$.0062 = .62\%$$

Point estimates vs. interval estimates

- The best estimate of a population parameter is the corresponding sample statistic. This is called a point estimate.
- However, a point estimate is unlikely to be exactly equal to the population parameter.
- For example, the best estimate of the population mean is the sample mean, but the sample mean is unlikely to be exactly equal to the population mean.

Point estimates vs. interval estimates (Continued)

- To account for this uncertainty, we can compute a range of values that is likely to contain the population parameter. This is called an interval estimate.

Example: Point estimates vs. interval estimates

- *You ask parents of middle schoolers how many hours their child spends on homework each week. You collect data from a sample of 50 parents. On average, parents report that their child spends 10 hours on homework each week.*
- This is your **point estimate** of the population mean. For example, you might report, "The average middle schooler spends 10 hours on homework each week."
- But how confident are you that this is the true population mean? You can compute an **interval estimate** to provide a range of values that likely contains the population mean.
- Your interval estimate might be something like: 9 to 11 hours. You might report, "The average middle schooler spends between 9 and 11 hours on homework each week."

But how do we compute this interval estimate?

Confidence Intervals

- A common type of interval estimate is a confidence interval.
- A confidence interval is the range of values within which a population parameter is estimated to fall.
- A confidence interval is usually expressed in terms of a percentage, such as 95% or 99%. This percentage is called the confidence level.
- A 95% confidence interval means that if we were to take many samples and compute a 95% confidence interval for each sample, then approximately 95% of those intervals would contain the true population parameter.

Confidence Intervals (Continued)

- When we compute a 95% confidence interval, we are *not* saying that there is a 95% probability that the population parameter falls within our specific interval. The population parameter is fixed; it either falls within our interval or it does not.
- We **are** saying that we expect to find the population mean within this interval 95% of the time when we conduct the same study with the same sample size

Explanation: <https://www.youtube.com/watch?v=tFWsuO9f74o>



Calculating confidence intervals with z distributions

You have:

- Sample mean (M): E.g., 10 hours of homework per week
- Sample size (n): E.g., 50 middle schoolers
- Population standard deviation (SD): E.g., 2 hours

Note: This is rare in practice! We will learn how to handle unknown population SDs later in the course.

You want:

- 95% Confidence interval (CI) around the sample mean.

Calculating confidence intervals with z distributions (Continued)

1. Assume your **sample** mean lies in the center of a normal distribution.
2. Determine the area under the normal curve that corresponds to your desired confidence level (e.g., 95%).
 - For a 95% CI, this is .95. This leaves .05 in the tails, or .025 in each tail. From this, you can determine the bounds of the CI.
 - For a 95% CI, the upper bound is .975. The lower bound is .025.

Calculating confidence intervals with z distributions (Continued)

3. Find the corresponding z scores for these bounds using a z table or computer programming.
 - For the upper bound (.975), the z score is approximately 1.96.
 - For the lower bound (.025), the z score is approximately -1.96.
4. Convert the z statistic to raw scores.
 - Use the formula: $X = \mu + z \times SE$, where $SE = SD/\sqrt{n}$.
5. Now we have our 95% CI!

R Demo: Z Scores, Percentiles, and Confidence Intervals

- Let's do a demo in R to calculate z scores, percentiles, and confidence intervals.
- This will help prepare you for Homework 2, which involves some R coding.

That's all for today!

Next time: more on **t tests**