Inferential Statistics

Pablo E. Gutierrez-Fonseca

Fall 2024

University of Vermont

Expanding on Hypothesis Testing

- **Tails in hypothesis testing** represent the critical regions where **extreme values** of the test statistic indicate significant differences. If the test statistic falls in one of the tails, we reject the null hypothesis in favor of the alternative hypothesis.

$$H_0 : \mu = value \quad \text{vs.} \quad H_1 : \mu \neq value \quad H_1 : \mu < value \quad H_1 : \mu < value$$

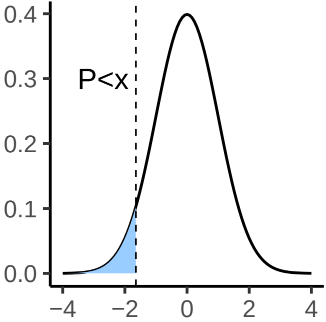
- To test this hypothesis, we define both rejection and acceptance regions. These regions are determined by our chosen significance level, typically set at **0.05**.

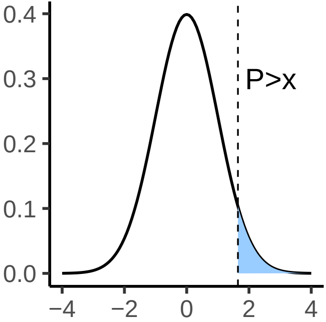
Types of Hypothesis Tests:

- **Left-tailed** test: Tests if the sample mean is significantly **less than** the population mean.
- **Right-tailed** test: Tests if the sample mean is significantly **greater than** the population mean.
- **Two-tailed** test: Tests if the sample mean is significantly different from the population mean, in either direction.

Expanding on Hypothesis Testing

- 1-tailed
  - Hypothesis includes an **expected direction**.



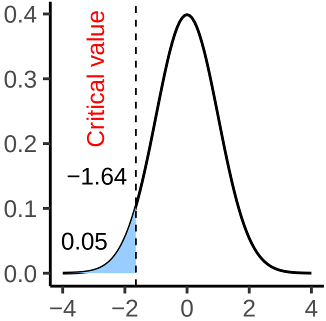


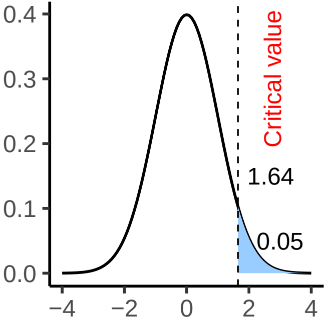
- Decrease
- Cooler
- Smaller
- Lower

- Increase
- Warmer
- Higher
- Expand

Expanding on Hypothesis Testing

- 1-tailed - hypothesis includes an **expected direction**.

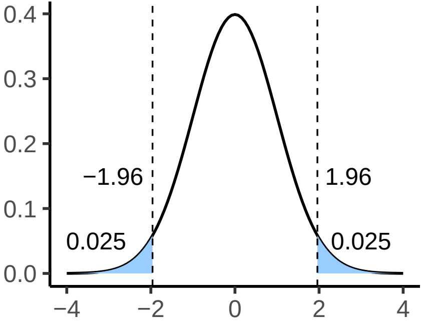




- If your obtained test statistic falls beyond the critical value (lightblue) for your given Alpha threshold = Significant result, **reject the null**.

Expanding on Hypothesis Testing

- 2-tailed tests:
  - Have **no expected directionality** hypothesized.
  - Splits the 5% of the area under the curve that would be considered significant between both tails of the normal distribution curve.
  - Are therefore less powerful tests (more likely to find a significant result).



Significant or Not?

- **Not significant**:
- Accept the null hypothesis.
- There is **no** difference between the sample and population mean.
- Obtained test statistic $<$ critical value threshold.
- p-value $>$ alpha threshold (usually 0.05).

Significant:

- Reject the null hypothesis,
- There **is** a difference between the sample and population mean.
- Obtained test statistic $>$ critical value threshold.
- p-value $<$ alpha threshold (usually 0.05).

Basic steps for an **Inferential Test**

- A statement of null hypothesis:

  Know the research Hypothesis

- Choose the appropriate test:

  Including if a 1 or 2-tailed test

- Set the level of Type I error risk (alpha):

  Usually < 0.05

- Analyze data distribution:

  do you meet assumptions like normality and independence?

Basic steps for an **Inferential Test**

- Compute the test statistic (obtained) value:

Calculated from formula

- Assess significance:
- Determine the critical value needed to reject the null hypothesis and compare it to your:

$$|\text{calculated}| > \text{critical value} = \text{significant}$$

- Determine the p-value associated with your calculated test statistic

$$\text{P-value} < \text{alpha threshold} = \text{significant}$$

- Summarize:

Clear, succinct paragraph with all pertinent information and interpretation

Basic steps for an **Inferential Test**

- We can select an appropriate test simply by answering some questions.
- What type of data do we have?
  - If we have **frequency data**, we select the Chi-square family.
  - **Continuous and categorical variables** will have a variety of different tests depending on the research question:
- What type of research question are we considering?
  - If **comparing one sample mean** to a **population** = z-test.
  - If focus is on **differences** we go to the family of tests concerned with comparing groups or treatments (i.e. t-tests, ANOVA).
  - If the focus is on **relationships** between variables, we go to the correlation tests.
  - If we want to **predict outcomes** we go to the regression family.

$$y = \beta_0 + \beta_1 x + \epsilon$$

where:

- y is the dependent variable,
- x is the independent variable,
- β_0 is the intercept,
- β_1 is the slope,
- ϵ is the error term.

Basic steps for an **Inferential Test**

- We can select an appropriate test simply by answering some questions.
  - How many variables, how many groups are we including? (multivariate)
  - Are observations independent from each other or purposely paired?
  - Is my data not normally distributed? (non-parametric tests)

Statistical shorthand **z-test** example

$$z_{(100)} = 1.4, p = 0.16$$

Statistical shorthand **z-test** example

$$z_{(100)} = 1.4,\ p = 0.16$$

$$\downarrow$$

$$\underbrace{z\underbrace{(100)}_{\text{Sample size}}}_{\text{Test statistic}} = \underbrace{1.4}_{\text{z-value}},\quad \underbrace{p = 0.16}_{\text{p-value}}$$

- Every statistical test has a letter designating the type of test.
- Because power is so closely tied to sample size, we report sample size.
- For clarity (and to confirm the correct-tailed test is used to assess the p-value), include the calculated test statistic value.
- Always report the p-value.
- Some tests will also include a secondary metric to assess how meaningful results are (if significant).

University of Vermont

How to summarize your analysis: Key Components of a Statistical Summary

Before installing a new air quality monitoring instrument, we tested to see if a sample of measurements taken at the testing lab differed significantly different from the long-term mean for the larger network population.

Using a 2-tailed, one-sample z-test for on our normally distributed samples (W = 0.78).

While the mean of the new instrument sample was slightly higher (sample mean = 117 vs. population mean = 108), we found no significant difference ($z_{(100)} = 1.4$, $p = 0.16$).

Based on these results we approve the installation of this instrument at the new site. However, we would suggest statistical comparisons of the current and new unit side-by-side, prior to decommissioning the current instrument.

How to summarize your analysis: Key Components of a Statistical Summary

Introduction/Background:

Before installing a new air quality monitoring instrument, we tested to see if a sample of measurements taken at the testing lab differed significantly different from the long-term mean for the larger network population.

Methods:

Using a 2-tailed, one-sample z-test for on our normally distributed samples (W = 0.78).

Results:

While the mean of the new instrument sample was slightly higher (sample mean = 117 vs. population mean = 108), we found no significant difference ($z_{(100)} = 1.4$, $p = 0.16$).

Implications:

Based on these results we approve the installation of this instrument at the new site. However, we would suggest statistical comparisons of the current and new unit side-by-side, prior to decommissioning the current

Our First Inferential Tests!
One sample z-test

When would you run a one-sample z-test?

- Use the One Sample z-test when:
  - You want to test for a **difference** between one **sample mean** and a larger **population mean**.
  - There is only **one group (sample)** being tested against the larger **population**.
  - You know (or can estimate) the mean and standard deviation of the population.
  - Data is normally distributed.

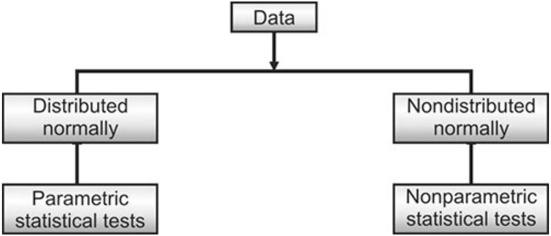
$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where:
- $\bar{X}$: Sample mean.
- μ: Population mean.
- σ: Population standard deviation. - n: Sample size.

When would you run a one-sample z-test?

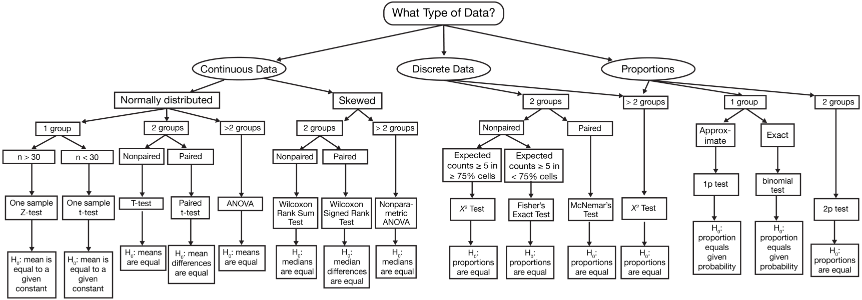
- Use the One Sample z-test when:
  - You want to test for a **difference** between one **sample mean** and a larger **population mean**.
  - There is only **one group (sample)** being tested against the larger **population**.
  - You know (or can estimate) the mean and standard deviation of the population.
    - Expert Information.
    - Scientific Literature.
    - Data archives.
    - Meaningful (hypothesized value).
  - Data is normally distributed.

Flowcharts:

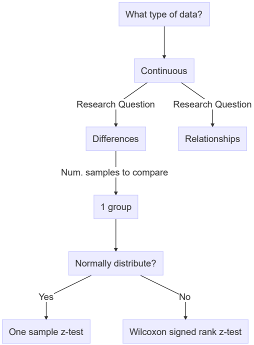

Data
Distributed normally
Nondistributed normally
Parametric statistical tests
Nonparametric statistical tests

Flowcharts

Flow chart: which test statistic should you use?



Flowcharts


What type of data?
Continuous
Research Question
Research Question
Differences
Relationships
Num. samples to compare
1 group
Normally distribute?
Yes
No
One sample z-test
Wilcoxon signed rank z-test

One sample z-test

Assume you have a sample of **20 observations**. The mean of this **sample is 150**, and you want to determine whether there is a **difference between the sample mean and the larger population mean**. Since you are not specifying a direction for this difference, you will conduct a two-tailed test.

The **population mean is 164**, and the **population standard deviation is 33**.

Next, calculate the obtained value for this **one-sample z-test.**

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

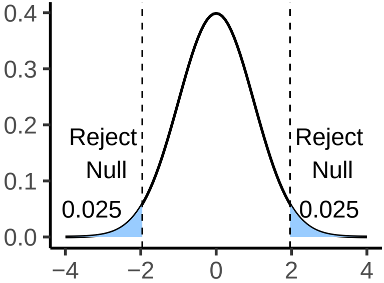
$$z = \frac{150 - 164}{\frac{33}{\sqrt{20}}} = -1.897$$

One sample z-test

Assume you have a sample of **20 observations**. The mean of this **sample is 150**, and you want to determine whether there is a **difference between the sample mean and the larger population mean**. Since you are not specifying a direction for this difference, you will conduct a two-tailed test.

The **population mean is 164**, and the **population standard deviation is 33**.

Next, calculate the obtained value for this **one-sample z-test.**



$$z = \frac{150 - 164}{\frac{33}{\sqrt{20}}} = -1.897$$

Critical Value Approach

- In the critical value approach, we compare the calculated **test statistic** to the **critical value** from the standard normal distribution. For a **two-tailed test** at the $\alpha = 0.05$ significance level, the critical values are approximately ± 1.96, which corresponds to the **0.025** in each tail.

- Decision Rule
- If $|\text{calculated test statistic}| >$ critical value = **Significant**
- If $|\text{calculated test statistic}| \leq$ critical value = **Not Significant**

Critical Value Approach

In this case:

$$| - 1.897| < 1.96$$

- Thus, we fail to reject the null hypothesis, indicating that there is **no significant** difference between the sample mean and the population mean.

Finding the p-value

Now we need to determine the probability (p-value) associated with the calculated test statistic. You have several options to achieve this:

```
z_value <- -1.897
p_value <- 2 * pnorm(z_value)  # Two-tailed p-value
p_value  # This will give the result
```

```
## [1] 0.05782794
```

- Since this p-value (0.0578) is greater than the alpha threshold of 0.05, we conclude that the result is **not significant**.

Example: Car rental

Example: Car rental

- A rental car company claims the mean time to rent a car on their **website is 60 seconds with a standard deviation of 30 seconds**. A random sample **36 customers** attempted to rent a car on the website. The **mean time to rent was 75 seconds**. Is this enough evidence to contradict the company's claim?

- Step 1: Set up the hypotheses

- **Null hypothesis (H_0)**: The mean rental time is 60 seconds.

$$H_0 : \mu = 60$$

- **Alternative hypothesis (H_1)**: The mean rental time is greater than 60 seconds.

$$H_1 : \mu > 60$$

Example: Car rental

- A rental car company claims the mean time to rent a car on their **website is 60 seconds with a standard deviation of 30 seconds**. A random sample **36 customers** attempted to rent a car on the website. The **mean time to rent was 75 seconds**. Is this enough evidence to contradict the company's claim?

- Step 2: Test statistic formula

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad z = \frac{75 - 60}{\frac{30}{\sqrt{36}}} = \frac{15}{5} = 3$$

Example: Car rental

- A rental car company claims the mean time to rent a car on their **website is 60 seconds with a standard deviation of 30 seconds**. A random sample **36 customers** attempted to rent a car on the website. The **mean time to rent was 75 seconds**. Is this enough evidence to contradict the company's claim?

- Step 2: Test statistic formula

$$z = \frac{75 - 60}{\frac{30}{\sqrt{36}}} = \frac{15}{5} = 3$$

$$|3| > 1.645$$

Example: Car rental

- A rental car company claims the mean time to rent a car on their **website is 60 seconds with a standard deviation of 30 seconds**. A random sample **36 customers** attempted to rent a car on the website. The **mean time to rent was 75 seconds**. Is this enough evidence to contradict the company's claim?

- Step 3: Set up the hypotheses

For a **one-tailed test** at a significance level of $\alpha = 0.05$, the critical z-value is approximately **1.645** (since we are only looking at one tail).

```r
# Given z-value
z_value <- 3
# Calculate the one-tailed p-value
p_value <- 1 - pnorm(z_value)
# Print the p-value
p_value
```

```
## [1] 0.001349898
```

Example: Car rental

We tested a rental car company's claim that reservations could be made online in 60 seconds on average. Using a random sample of 36 online rentals in a one-tailed, one-sample z-test we found that mean reservation time is actually significantly higher than the company claims ($z_{(36)} = 3$, $p = 0.001$). This result indicates that it is highly unlikely customers can rent a car in 60 seconds as claimed. We suggest that the rental company adjust its claim to match the 75 second mean encountered for our test sample to avoid misleading potential customers.

Example: Another example

Example: Another example

A water quality regulation specifies that the lead level in drinking water should be no more than **15 ppb** (μ) with a standard deviation $\sigma = 0.25$. A random sample of three lead measurements yielded values of 15.84, 15.33, and 15.58. Perform a one-sided z-test to determine if the mean lead level exceeds 15 ppb.

$$H_0 : \mu = 15 \qquad\qquad H_1 : \mu > 15$$

Example: Another example

library(BSDA)

z.test(x, alternative='greater', mu=, sigma.x=)

where:

- x: values for the first sample
- alternative: the alternative hypothesis ('greater', 'less', 'two.sided')
- mu: mean population.
- sigma.x: population standard deviation,

Example: Another example

```
##
##  One-sample z-Test
##
## data:  water_sample
## z = 4.0415, p-value = 2.656e-05
## alternative hypothesis: true mean is greater than 15
## 95 percent confidence interval:
##  15.34592       NA
## sample estimates:
## mean of x
##  15.58333
```

The test statistic for the one-sample **z-test is 4.0415** and the corresponding **p-value is** $p = 2.66 \times 10^{-5}$.

Example: Another example

We tested a water sample to determine whether the mean lead level exceeds the regulatory limit of 15 ppb. Using a random sample of three lead measurements in a one-tailed, one-sample z-test, we found that the mean lead level is significantly greater than the regulatory limit ($z_{(3)} = 4.04$, $p = 2.66 \times 10^{-5}$). This result indicates that it is highly unlikely that the true mean lead level is 15 ppb or lower. We suggest further investigation and corrective actions to reduce the lead concentration, as the sample mean was 15.58 ppb.