Introduction to Statistics - Young Researchers Fellowship Program

Lecture 6 - Foundations of Hypothesis Testing

Daniel Sánchez Pazmiño

Laboratorio de Investigación para el Desarrollo del Ecuador

October 2024

Introduction to Hypothesis Testing

Null and Alternative Hypotheses

- We use a sample to make inferences about the population
- A hypothesis test helps us determine if there's enough evidence in the sample to support a statement about the population
- **Null hypothesis**: the baseline assumption
- **Alternative hypothesis**: the opposite of the null
  - Typically what we "want", the phenomenon being studied

Defining the null

- We define the null by setting our baseline scenario: what we believe might be true in "normal circumstances"
- For instance, we may think that typically the sample mean is of a certain historical value:

$$H_0 : \mu = \mu_0$$

- μ_0 is typically called "mu naught" or hypothesized mean
  - The hypothetical baseline value

Defining the alternative

- We define the alternative as the contrary to the null
- Usually what we want to look out for (i.e. is the treatment effective?)
- For example, that the historical value is no longer the baseline and the population mean changed.

$$H_1 : \mu \neq \mu_0$$

Errors

Type I and Type II Errors

Decision	H_0 is True	H_0 is False
Reject H_0	Type I Error	Correct Decision
Fail to reject H_0	Correct Decision	Type II Error

Type I and Type II Errors

- **Type I Error (α)**: Rejecting the null hypothesis when it's true
  - Example: Null: $\mu \geq 3.0$, we conclude $\mu < 3.0$
  - The probability of making this error is the significance level α
- **Type II Error (β)**: Failing to reject the null hypothesis when it's false

Hypothesis testing

General idea of a hypothesis test

- We are asking: "if the sampling distribution is truly distributed according to the null, what is the most extreme value of a sampling exercise I can get to not doubt the null?"
- Means computing a number of standard errors that the sample statistic is located relative to the hypothesized mean.
    - The number of standard errors is called the "test statistic"
- If the null is true, this shouldn't be too far from μ_0
    - If it is far, we have to reject the null
    - Otherwise, "fail to reject" H_0.
- How far will we tolerate? Based on α, our significance level (related to confidence we want).

p-values

- The p-value is the probability associated to the test statistic.
- Answers the question: how likely is it to get this test statistic if the null were true?
- The general rejection rule is:

$$p < \alpha$$

because it means that the p-value shows that it is more likely that we have gotten an extreme value, which does not come from the H_0 scenario, than the tolerance we've set to commit a type-I error.

General type of tests

Case 1: Left-tailed Test (One-side)

- Null hypothesis: population parameter $\mu \geq \mu_0$
- Alternative hypothesis: $\mu < \mu_0$
- Example: Average GPA of students is less than 3.0

Case 2: Right-tailed Test (One-side)

- Null hypothesis: population parameter $\mu \leq \mu_0$
- Alternative hypothesis: $\mu > \mu_0$
- Example: Average GPA of students is greater than 3.0

Case 3: Two-tailed Test (Two side)

- Null hypothesis: population parameter $\mu = \mu_0$
- Alternative hypothesis: $\mu \neq \mu_0$
- Example: Average GPA is different from 3.0

One sample tests about the mean

One Sample Tests?

- This means we're only working with one sample, not several.
- Comparing a mean against a numerical value.
- Later we will work with many samples.

The z-test (one sample)

- Tests whether the population mean μ is equal to a given value
- Used when population standard deviation is known
- We can work with a normal distribution for computing probabilities

Z-Test Types

- **Left-tailed test**: $H_0 : \mu \geq \mu_0$, $H_1 : \mu < \mu_0$
- **Right-tailed test**: $H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$
- **Two-tailed test**: $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

General procedure to do a hypothesis test

1. Compute sample mean to be used or use the given one.
2. Compute the *test statistic*. In the case of a Z-test, the test statistic is Z:

$$Z = \frac{x - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

3. Compute the p-value associated with the test statistic and given α
4. If the p-value is smaller than α, reject H_0.

Calculating p-values

- We need to know how to calculate p-values based on each type of test, to accurately reject based on available information.

Test Type	Null Hypothesis (H_0)	Alternative Hypothesis (H_1)	Formula for p-value	R Code Example
Two-tailed	$H_0 : \mu = \mu_0$	$H_1 : \mu \neq \mu_0$	$2 \times P(Z \geq \lvert z_{\text{score}} \rvert)$	`2 * (1 - pnorm(abs(z_score))`
Left-tailed	$H_0 : \mu \geq \mu_0$	$H_1 : \mu < \mu_0$	$P(Z \leq z_{\text{score}})$	`pnorm(z_score)`
Right-tailed	$H_0 : \mu \leq \mu_0$	$H_1 : \mu > \mu_0$	$P(Z \geq z_{\text{score}})$	`1 - pnorm(z_score)`

Example of Left-Tailed Z-Test

- Test whether population mean is **greater than or equal** to 3.0
  - $H_0 : \mu \geq 3.0$
  - $H_1 : \mu < 3.0$
  - Sample size = 100, sample mean = 2.8, $\sigma = 0.3$
  - $z = \frac{2.8-3.0}{0.03} = -6.67$
- Use p-value approach or critical value approach for rejecting

R implementation

- Base R does not offer a built-in package for this test, however, we may easily calculate probabilities using `pnorm()`
- For this, we should know how to calculate the p-value as per the table above.

R Implementation for the Example

```r
# Given values
mu_0 <- 3.0      # Hypothesized population mean
x_bar <- 2.8     # Sample mean
sigma <- 0.3     # Population standard deviation
n <- 100         # Sample size

# Calculate the standard error
se <- sigma / sqrt(n)

# Calculate the Z-score
z_score <- (x_bar - mu_0) / se

# Calculate the p-value for a left-tailed test
p_value <- pnorm(z_score)
```

Proportions

- Proportions work just like means.
- Need only to redefine the Z test statistic to follow the modified standard error for $\hat{p}$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where p_0 is the hypothesized population proportion.

- Apply the same rules you would for Z-tests.

R Implementation for Proportion Z-tests

```
# Given values
p_hat <- 0.55  # Sample proportion
p_0 <- 0.50    # Hypothesized population proportion (null hypoth
n <- 100       # Sample size

# Calculate the Z-score
z_score <- (p_hat - p_0) / sqrt((p_0 * (1 - p_0)) / n)

# P-value
p_value <- 2 * (1 - pnorm(abs(z_score)))
```

Errors and Significance Levels, different rejection rules

- **Significance level (α)**: Probability of making a Type I error
  - Example: $\alpha = 0.05$ means 95% confidence level
- Compare p-value to α or use critical value
- Critical values are the number of standard errors associated with the α probability under a specific type of test.
  - Rejection rules with these vary by type of test

Common Critical Values

α	Left-Tailed	Right-Tailed	Two-Tailed
0.05	-1.645	1.645	$\pm$ 1.96
0.01	-2.33	2.33	$\pm$ 2.58
0.10	-1.28	1.28	$\pm$ 1.64

Rejection rules for critical values, one sample tests

Test Type	Hypothesis (H_0, H_1)	Critical Value (Z)	Rejection Rule
Left-tailed	$H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$	z_α (negative)	Reject H_0 if $z < z_\alpha$
Right-tailed	$H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$	z_α (positive)	Reject H_0 if $z > z_\alpha$
Two-tailed	$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$	$z_{\alpha/2}$ (positive and negative)	Reject H_0 if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

T-tests (when we don't know the population standard deviation)

As you might remember, when we cannot obtain enough information for a reliable estimate of the population standard deviation, we use the sample standard deviation as an estimate. This is common in practice when we have small samples or lack population standard deviation.

T-statistic:

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

- μ_0: hypothesized population mean
- Denominator: standard error of the sample mean.

The t distribution is used, which has heavier tails than the normal distribution and depends on degrees of freedom $(n - 1)$.

Types of T-Tests

1. **Left-tailed test**: Test for means less than the hypothesized value.
2. **Right-tailed test**: Test for means greater than the hypothesized value.
3. **Two-tailed test**: Test for means different from the hypothesized value.

Example: Left-tailed Test

- Sample size: 25
- Sample mean: 9.5
- Sample standard deviation: 2.5
- Hypothesized mean: 10
- Significance level: 5%

Example: Left-tailed test

Null and Alternative Hypotheses:

$$H_0 : \mu \geq 10$$
$$H_1 : \mu < 10$$

T-statistic:

$$t = \frac{9.5 - 10}{2.5/\sqrt{25}} = -2$$

Critical value at 24 degrees of freedom (5% significance):

$$t_{critical} = -1.711$$

Since $t = -2$ is less than the critical value, we reject the null hypothesis.

Example: Right-tailed Test

- Sample size: 25
- Sample mean: 9.5
- Sample standard deviation: 2.5
- Hypothesized mean: 10

Example: Right-tailed Test

Null and Alternative Hypotheses:

$$H_0 : \mu \leq 10$$
$$H_1 : \mu > 10$$

T-statistic:

$$t = \frac{9.5 - 10}{2.5/\sqrt{25}} = -2$$

For a right-tailed test, the critical value is $t_{critical} = 1.711$. Since $t = -2$, we fail to reject the null hypothesis.

Two-tailed Test Example

- Sample size: 25
- Sample mean: 12
- Sample standard deviation: 2.5
- Hypothesized mean: 10

Two-tailed Test Example

Null and Alternative Hypotheses:

$$H_0 : \mu = 10$$

$$H_1 : \mu \neq 10$$

T-statistic:

$$t = \frac{12 - 10}{2.5/\sqrt{25}} = 4$$

Critical values are $t_{critical} = \pm 2.064$. Since $t = 4$ exceeds the critical values, we reject the null hypothesis.

Proportions and Sampling Distribution

For categorical data, we compute proportions instead of means. Proportions follow a normal distribution with large enough samples. The sample proportion is calculated as:

$$\hat{p} = \frac{x}{n}$$

Where x is the number of successes and n is the sample size.

Standard error of proportion:

$$SE(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

The Interval Approach for Hypothesis Testing

We can also use confidence intervals for hypothesis testing. If the null hypothesis value falls outside the confidence interval, we reject the null hypothesis.

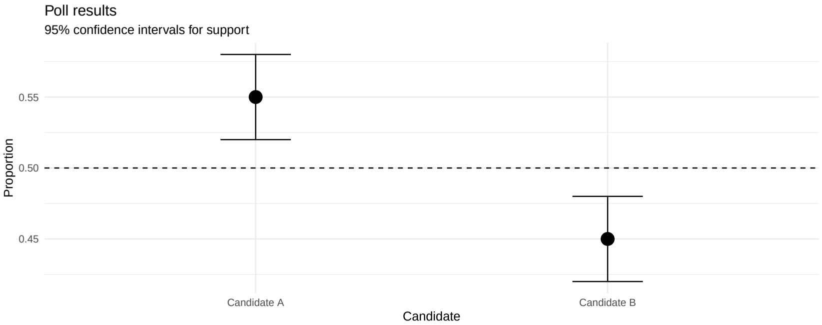

Figure 1: Confidence interval for the proportion of people who support each candidate

R Implementation: T-test

- We use the `t.test()` base R function for computing t-tests, either one sample or two sample (we will cover this soon).
- Define the alternative hypothesis for a right, left or two-tailed test using `alternative`
- May need to use `na.action` to not consider `NA` values.
- When we already have the summarised data (i.e. given the mean and std. deviation), we should calculate our own p-value or critical value using `pt()` and `qt()`.

Hypothesis Testing for Means of Two Populations

Two populations?

So far, we've been performing hypothesis tests about one population. However, we can also test hypotheses about two populations. In this case, we test whether the population mean of one group is equal to the mean of another group.

Difference Between Two Population Means

Known Population Standard Deviations (Independent Samples)

In this case, we want to know if the means of two populations are different.
We compute the difference between the two sample means:

$$\bar{x}_1 - \bar{x}_2$$

The standard error of this difference is:

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Where: - σ_1 and σ_2 are the population standard deviations - n_1 and n_2 are the sample sizes.

Hypothesis Testing for Two Means

The test statistic for the difference between two means is:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

We use this z statistic to test the null hypothesis $H_0 : \mu_1 = \mu_2$.

Example Hypotheses for Two Means

We can define three types of hypothesis tests for two means:

1. **Two-tailed test:**

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$

2. **Right-tailed test:**

$$H_0 : \mu_1 - \mu_2 \leq 0$$
$$H_1 : \mu_1 - \mu_2 > 0$$

3. **Left-tailed test:**

$$H_0 : \mu_1 - \mu_2 \geq 0$$
$$H_1 : \mu_1 - \mu_2 < 0$$

Unknown Population Standard Deviations (Independent Samples)

When the population standard deviations are unknown, we estimate them using the sample standard deviations, and use the t distribution.

The standard error becomes:

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The test statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Degrees of Freedom

To calculate the degrees of freedom, use the formula:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

The t statistic is compared with the critical value from the t distribution based on the calculated degrees of freedom.

Dependent Samples (Paired Samples)

For dependent samples, we work with the **difference** between paired observations. For example:

Student	Test score before	Test score after	Difference
1	80	90	10
2	70	85	15
...	...	...	...

We conduct a one-sample t-test on the mean difference between the groups.

Test Statistic for Paired Samples

For paired samples, the test statistic is calculated as:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

Where: - $\bar{d}$ is the mean difference between pairs. - μ_d is the hypothesized mean difference. - s_d is the sample standard deviation of the differences. - n is the sample size.

R Implementation: Two Sample t-tests

- `t.test()` is a wildly customizable function!
  - Allows for feeding two vectors of the samples to be used
- For paired samples, use the `paired` argument.
- Two sample proportion tests have the `prop.test()` function, which provides a chi-squared approximation of the test (better).
  - To actually use the z-test, calculate p-values or critical values yourself with the `_norm()` functions