Hypothesis Testing

Tom Bruning

2018-01-30

In this chapter, you learn:

- The basic principles of hypothesis testing
- How to use hypothesis testing to test a mean or proportion
- The assumptions of each hypothesis-testing procedure, how to evaluate them, and the consequences if they are seriously violated
- Pitfalls & ethical issues involved in hypothesis testing
- How to avoid the pitfalls involved in hypothesis testing

What is a Hypothesis?

A *hypothesis* is a claim (assertion) about a population parameter:

- population mean
- population proportion

Example: The mean monthly cell phone bill in this city is $\mu = \$42$
Example: The proportion of adults in this city with cell phones is $\pi = 0.68$

The Null Hypothesis, H_0

States the claim or assertion to be tested for example:

- The mean diameter of a manufactured bolt is 30mm ($H_0 : \mu = 30$)

It is **always** about a population parameter, not about a sample statistic.

- Use $H_0 : \mu = 30$ **not** $H_0 : \bar{x} = 30$

Begin with the assumption that the null hypothesis is true.

- Similar to the notion of innocent until proven guilty

Refers to the status quo, specification, or historical value.
Always contains "=", or "≤", or "≥" sign.
The null hypotheis may or may not be rejected.

The Alternative Hypothesis, H_1

- It is the opposite of the null hypothesis, the two hypothesises are *mutually exclusive* and *collectively exhaustive.* e.g., The average diameter of a manufactured bolt is not equal to 30mm ($H_1 : \mu \neq 30$)

- Challenges the status quo
- Never contains "=", or "≤", or "≥" sign.
- May or may not be proven.
- Is generally the hypothesis that the researcher is trying to prove.

The Hypothesis Testing Process

Claim: The population mean age is 50.

$H_0 : \mu = 50$,

$H_1 : \mu \neq 50$

Sample the population and find the sample mean.

Suppose the sample mean age was $\bar{x} = 20$

This is significantly lower than the claimed mean population age of 50.

If the null hypothesis were true, the probability of getting such a different sample mean would be very small, so you reject the null hypothesis .

In other words, getting a sample mean of 20 is so unlikely if the population mean was 50, you conclude that the population mean must not be 50.

The Test Statistic and Critical Values

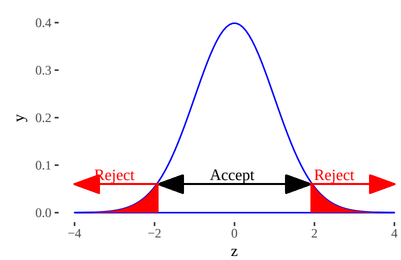
- If the sample mean is close to the stated population mean, the null hypothesis **is not** rejected.

- If the sample mean is far from the stated population mean, the null hypothesis **is** rejected.

- How far is "far enough" to reject H_0?

Important: The critical value of a test statistic creates a "line in the sand" for decision making – it answers the question of how far is far enough.

```
## Loading required package: grid
```



Possible Errors in Hypothesis Test Decision Making

- Type I Error
  - Reject a true null hypothesis
  - Considered a serious type of error
  - The probability of a Type I Error is α
  - Called level of significance of the test
  - Set by researcher in advance

- Type II Error

  - Failure to reject a false null hypothesis
  - The probability of a Type II Error is β

Table 1: Possible Hypothesis Test Outcomes

Decision	Null is True	Null is False
Do not reject Null	No Error	Type II Error
Reject Null	Type I Error	No Error

6 Steps in Hypothesis Testing

Doing hypothesis testing involves two phases. The first phase *Planning the test* should be completed before you begin collecting data. In this phase you determine all the testing parameters. This is where the management decisions are made. The second phase is running the test.

- Planning the test

  1. State the null hypothesis, H_0 and the alternative hypothesis, H_1
  2. Choose the level of significance, α, and the sample size, n
  3. Determine the appropriate test statistic and sampling distribution
  4. Determine the critical values that divide the rejection and nonrejection regions

- Run the test

  5. Collect data and compute the value of the test statistic
  6. Make the statistical decision and state the managerial conclusion. If the test statistic falls into the nonrejection region, do not reject the null hypothesis H0. If the test statistic falls into the rejection region, reject the null hypothesis. Express the managerial conclusion in the context of the problem.

Note: This is *not* and iterative process. The appropriate process is to design the test, set the parameters and then run the test. The conclusions you reach by adhering to this process is unaffected by the results of the math. If you start ‘monkeying around with the test’ you are not being honest about the result.

Questions To Address In The Planning Stage

- What is the goal of the survey, study, or experiment?
- How can you translate this goal into a null and an alternative hypothesis?
- Is the hypothesis test one or two tailed?

- Can a random sample be selected?
- What types of data will be collected? Numerical? Categorical?
- What level of significance should be used?
- Is the intended sample size large enough to achieve the desired power?
- What statistical test procedure should be used?
- What conclusions & interpretations can you reach from the results of the planned hypothesis test?

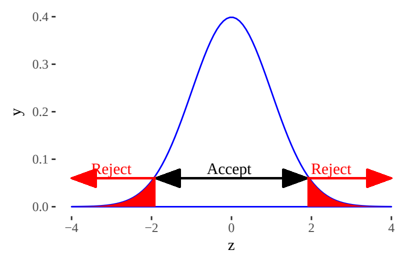
Critical Value Approach to Testing

For a two-tail test for the mean, σ known:

- Convert sample statistic ($\bar{x}$) to test statistic (Z_{STAT})
- Determine the critical Z values for a specified level of significance α from a table or computer
- Decision Rule: If the test statistic falls in the rejection region, reject H_0 ; otherwise do not reject H_0.

Two-Tail Tests

There are two cutoff values (critical values), defining the regions of rejection for *two tails test*.



- Reject H_0 if the test statistic is below the $-Z_{STAT}$
- Do not reject H_0 is higher than the $-Z_{STAT}$ and less than the $+Z_{STAT}$
- Reject H_0 if the test statistic is more than $+Z_{STAT}$
  Lower critical value is $-Z_{STAT}$
  Upper critical value is $+Z_{STAT}$

Hypothesis Testing Example

Test the claim that the true mean diameter of a manufactured bolt is 30mm. (Assume $\sigma = 0.8$)

1. State the appropriate null and alternative hypotheses $H_0 : \mu = 30 \quad H_1 : \neq 30$ (This is a two-tail test)

2. Specify the desired level of significance and the sample size

   Suppose that:

   - $\alpha = 0.05$ and
   - $n = 100$
     are chosen for this test

3. Determine the appropriate technique σ is assumed known so this is a Z test.

4. Determine the critical values
   For $\alpha = 0.05$ the critical Z values are ± 1.96

5. Collect the data and compute the test statistic Suppose the sample results are:

- $n = 100$, $\bar{X} = 29.84$ ($\sigma = 0.8$ is assumed known)
  So the test statistic is:

$$Z_{STAT} = \frac{X \bar{-} \mu}{\frac{\sigma}{\sqrt{100}}} = \frac{29.84 - 30}{\frac{0{,}8}{\sqrt{100}}} = \frac{-0.16}{0.08} = -2 \tag{1}$$

6. Is the test statistic in the rejection region?

- Reject H_0 if $Z_{STAT} < -1.96$ *or* $Z_{STAT} > 1.96$; otherwise do not reject H_0.
- Here, Z_{STAT} = -2.0 < -1.96, so the test statistic is in the rejection region.

  Reach a decision and interpret the result:

- Since Z_{STAT} = -2.0 < -1.96, *reject the null hypothesis* and conclude there is sufficient evidence that the mean diameter of a manufactured bolt is not equal to 30.

Connection Between Two Tail Tests and Confidence Intervals

- For $\bar{X} = 29.84$, $\sigma = 0.8$ and $n = 100$, the 95% confidence interval is:

$$\bar{X} - (Z_{score})(\frac{\sigma}{\sqrt{n}}) \ \ to \ \ \bar{X} + (Z_{score})(\frac{\sigma}{\sqrt{n}}) = 29.84 = 1.96(\frac{.08}{\sqrt{100}}) \ \ to \ \ 29.84 + 1.96(\frac{.08}{\sqrt{100}}) \tag{2}$$

$29.6832 \le \mu \le 29.9968$

- Since this interval does not contain the hypothesized mean (30), we reject the null hypothesis at $\sigma = 0.05$

Do You Ever Truly Know σ?

- Probably not!
- In virtually all real world business situations, σ is not known.
- If there is a situation where σ is known then μ is also known (since to calculate   you need to know μ.)
- If you truly know μ there would be no need to gather a sample to estimate it.

Hypothesis Testing: σ Unknown

- If the population standard deviation is unknown, you instead use the sample standard deviation S.
- Because of this change, you use the t distribution instead of the Z distribution to test the null hypothesis about the mean.
- When using the t distribution you must assume the population you are sampling from follows a normal distribution.
- All other steps, concepts, and conclusions are the same.

Example: Two-Tail Test (σ Unknown)

The average cost of a hotel room in New York is said to be \$168 per night. To determine if this is true, a random sample of 25 hotels is taken and resulted in an X of \$172.50 and an S of \$15.40. Test the appropriate hypotheses at $\alpha = 0.05$.

(Assume the population distribution is normal)

$H_0 : \mu = 168$

$H_1 : \mu \neq 168$

$\alpha = 0.05$ $n = 25,$ $df = 25 - 1 = 24$

σ is unknown, so use a t statistic

Critical Value: $\pm t_{24,0.025} = \pm 2.064$

$$t_{stat} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{172.5 - 168}{\frac{15.4}{\sqrt{25}}} = 1.46 \tag{3}$$

Do not reject H_0 : insufficient evidence that true mean cost is different from \$168

Connection of Two Tail Tests to Confidence Intervals

For $X = 172.5, S = 15.40$ *and* $n = 25$, the 95% confidence interval for μ is:

$$172.5 - (2.064)\frac{15.4}{\sqrt{25}} \quad to \quad 172.5 + (2.064)\frac{15.4}{\sqrt{25}} \tag{4}$$

$166.14 \leq \mu \leq 178.86$

Since this interval contains the Hypothesized mean (**168**), we do not reject the null hypothesis at $\alpha = 0.05$.

One-Tail Tests

- In many cases, the alternative hypothesis focuses on a particular direction.

$H_0 : \mu \geq 3$
$H_1 : \mu < 3$

This is a lower-tail test since the alternative hypothesis is focused on the lower tail below the mean of 3. There is only one critical value, since the rejection area is in only one tail. $-Z_\alpha$ or $-t_\alpha$

$H_0 : \mu \leq 3$
$H_1 : \mu > 3$

This is an upper-tail test since the alternative hypothesis is focused on the upper tail above the mean of 3. There is only one critical value, since the rejection area is in only one tail. Z_α or t_α

Example: Upper-Tail t Test for Mean (σ unknown)

A phone industry manager thinks that customer monthly cell phone bills have increased, and now average over \$52 per month. The company wishes to test this claim. (Assume a normal population) Form hypothesis test:

$H_0 : \mu \leq 52$ the average is not over \$52 per month
$H_1 : \mu > 52$ the average is greater than \$52 per month
(i.e., sufficient evidence exists to support the manager's claim)

Suppose that $\alpha = 0.10$ is chosen for this test and n = 25. Find the rejection region: Using the t distribution table for $\alpha = .10 \;\; and \;\; n = 25 \;\; (d.f = 24)$:

Reject H_0 if $t_{STAT} > 1.318$.

- Obtain sample and compute the test statistic

Suppose a sample is taken with the following results: n = 25, $\bar{X} = 53.1$, and S = 10.

Then the test statistic is:

$$t_{stat} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{53.1 - 52}{\frac{10}{\sqrt{25}}} = 0.55 \qquad (5)$$

Do not reject H_0 since $t_{STAT} = 0.55 \leq 1.318$ there is not sufficient evidence that the mean bill is over \$52

Hypothesis Tests for Proportions

- Involves categorical variables
- Two possible outcomes
- Possesses characteristic of interest
- Does not possess characteristic of interest
- Fraction or proportion of the population in the category of interest is denoted by π

Sample proportion in the category of interest is denoted by p:

$$p = \frac{\bar{X}}{n} = \frac{Number\ in\ category\ of\ interest}{sample\ size} \tag{6}$$

When both nπ and n(1-π) are at least 5, p can be approximated by a normal distribution with mean and standard deviation.

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} \tag{7}$$

The sampling distribution of p is approximately normal, so the test statistic is a ZSTAT value:

$$Z_{stat} = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \tag{8}$$

An equivalent form to the last equation, but in terms of the number in the category of interest, $\bar{X}$:

$$Z_{stat} = \frac{X \bar{-} np}{\sqrt{np(1-p)}} \tag{9}$$

Example: Z Test for Proportion

A marketing company claims that it receives 8% responses from its mailing. To test this claim, a random sample of 500 were surveyed with 25 responses. Test at the $\alpha = 0.05$ significance level.

Check:

- nπ = (500)(.08) = 40
- n(1-π) = (500)(.92) = 460

Z Test for Proportion: Solution

Test Statistic: $H_0 : \pi = 0.08 \quad H_1 : \pi \neq 0.08$

$\alpha = 0.05$

n = 500, p = 0.05

Decision: Critical Values: ± 1.96

$$Z_{stat} = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{.05 - .08}{\sqrt{\frac{08(1-.08)}{500}}} = -2.47 \tag{10}$$

Reject H_0 at $\alpha = 0.05$

Conclusion:

There is sufficient evidence to reject the company's claim of 8% response rate.

Example: Confidence Interval for Proportion

$$p \pm (Z_{score})\sqrt{\frac{\pi(1-\pi)}{n}} = .05 \pm (1,96)\sqrt{\frac{.08(1-.08)}{500}} \tag{11}$$

$$0.026 \leq p \leq .074 \tag{12}$$

Reject H_0 at $\alpha = 0.05$
Conclusion:
Since the interval does not contain π there is sufficient evidence to reject the company's claim of 8% response rate.

Statistical Significance vs Practical Significance

Statistically significant results (rejecting the null hypothesis) are not always of practical significance.

- This is more likely to happen when the sample size gets very large

Practically significant results might be found to be statistically insignificant (failing to reject the null hypothesis)

- This is more likely to happen when the sample size is relatively small

Note: Failing to consider these questions can lead to bias or incomplete results.

Reporting Findings & Ethical Issues

- Should document & report both good & bad results.
- Should not only report statistically significant results
- Reports should distinguish between poor research methodology and unethical behavior
- Ethical issues can arise in:
  - The use of human subjects
  - The data collection method
  - The type of test being used
  - The level of significance being used
  - The cleansing and discarding of data
  - The failure to report pertinent findings