Hypothesis testing

Anastasiia Kim

May 1, 2020

Idea

- Need to decide which of two competing claims or statements (hypotheses) about some parameter is true
- The decision-making procedure is called hypothesis testing
- H_0: the hypothesis that is initially assumed to be true
- H_1: the statement contradictory to H_0

Ex. a pharmaceutical company might be interested in knowing if a new drug is effective in treating a disease
Null Hypothesis H_0 (no difference): the drug is not effective
Alternative Hypothesis H_1: the drug is effective
Ex. a radar system that uses radio waves to detect aircraft. The system receives a signal and, based on the received signal, it needs to decide whether an aircraft is present or not.
Null Hypothesis H_0 (no difference): No aircraft is present.
Alternative Hypothesis H_1: An aircraft is present.

Idea

- a statistical hypothesis is a statement about the parameters of one or more populations
- a statistical hypothesis may also be thought of as a statement about the probability distribution of a random variable
- the hypothesis will usually involve one or more parameters of this distribution
- the hypotheses are always statements about the population or distribution under study, not statements about the sample

Idea

Ex. consider the air crew escape system. Suppose that we are interested in the burning rate of the solid propellant.

- Burning rate is a random variable that can be described by a probability distribution.
- Let's focus on the mean burning rate (a parameter of this distribution).
- We are interested in deciding whether or not the mean burning rate is 50 centimeters per second.

Null Hypothesis H_0 (no difference): $\mu = 50$ centimeters per second
Alternative Hypothesis H_1 (two-sided hypothesis): $\mu \neq 50$ centimeters per second
Alternative Hypothesis H_1 (one-sided hypothesis): $\mu < 50$ centimeters per second
Alternative Hypothesis H_1 (one-sided hypothesis): $\mu > 50$ centimeters per second
Alternative testing:
Null Hypothesis H_0: $\mu < 50$ centimeters per second
Alternative Hypothesis H_1: $\mu \geq 50$ centimeters per second

Idea

Null Hypothesis H_0 (no difference): $\mu = 50$ centimeters per second
Alternative Hypothesis H_1 (two-sided hypothesis): $\mu \neq 50$ centimeters per second
The value of the population parameter (50 in this case) specified in the null hypothesis is usually determined in one of three ways:

- From past experience or knowledge of the process or even from previous tests or experiments. The objective of hypothesis testing to determine whether the parameter value has changed.
- May be determined from some theory or model regarding the process under study. The objective of hypothesis testing is to verify the theory or model.
- The value of the population parameter results from external considerations, such as design or engineering specifications. The usual objective of hypothesis testing is conformance testing (whether a system meets a defined set of standards).

Idea

- A procedure leading to a decision about the null hypothesis is called a test of a hypothesis.
- Hypothesis-testing procedures rely on using the information in a random sample
- If this information is consistent with the null hypothesis, we will not reject it
- if this information is inconsistent with the null hypothesis, we will reject it in favor of the alternative

Testing Procedure

- State your hypotheses
- Compute a test statistic from the sample data
- Use the test statistic to make a decision about the null hypothesis

Important: the truth or falsity of a particular hypothesis can never be known with certainty unless we can examine the entire population. A null hypothesis is not accepted just because it is not rejected. A hypothesis test does not determine which hypothesis is true, it only assesses whether available evidence exists to reject the H_0.

- Reject H_0 and conclude that the H_1 is true at the 95% (or 90, 99, etc.) confidence level
- Fail to reject the H_0 and conclude that not enough evidence is available to suggest H_0 is false at the 95% (or 90, 99, etc.) confidence level.

Example. The judge in a courtroom

H_0: Not guilty. H_1: Guilty
Say that you fail to reject H_0. Does it mean that the defendant is innocent? Based on the evidence, you can't reject that possibility. The judge will say 'Not guilty' because the prosecution failed to convince the judge to abandon the assumption of innocence.

Therefore, a hypothesis-testing procedure should be developed with the probability of reaching a wrong conclusion in mind.

- Rejecting the null hypothesis H_0 when it is true is defined as a type I error

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

- Failing to reject the null hypothesis when it is false is defined as a type II error

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

- Type I and type II errors are related. A decrease in the probability of one type of error always results in an increase in the probability of the other provided that the sample size n does not change.

Type I and Type II errors

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false})$$

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	No error	Type II error
Reject H_0	Type I error	No error

P-values

In the above discussions, we only reported an "accept" or a "reject" decision as the conclusion of a hypothesis test. We can provide more information using what we call P-values.

- Reject H_0 and conclude that the H_1 is true at the $100(1-\alpha)\%$ confidence level (90, 95, 99, etc.)
- Fail to reject the H_0 and conclude that not enough evidence is available to suggest H_0 is false at the 95% (or 90, 99, etc.) confidence level.

What is the lowest significance level α that results in rejecting the null hypothesis? Can we reject at level $\alpha = 0.05$? Or at 0.01?

The probability, computed assuming H_0 is true, that the statistic would take a value as extreme as or more extreme than the one actually observed is called the P-value of the test.

- P-value is the lowest significance level α that results in rejecting H_0
- if the P-value is small, it means that the observed data is very unlikely to have occurred under H_0, so we are more confident in rejecting the null hypothesis.

Z-test for a population mean. Normal distribution. Variance σ^2 is known

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

z TEST FOR A POPULATION MEAN

Draw an SRS of size n from a Normal population that has unknown mean μ and known standard deviation σ. To test the null hypothesis that μ has a specified value,

$$H_0: \mu = \mu_0$$

calculate the **one-sample z statistic**

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In terms of a variable Z having the standard Normal distribution, the P-value for a test of H_0 against

$H_a: \mu > \mu_0$ is $P(Z \geq z)$

$H_a: \mu < \mu_0$ is $P(Z \leq z)$

$H_a: \mu \neq \mu_0$ is $2P(Z \geq |z|)$

Here, we will reject H_0 if

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}$$

The image:The Basic Practice of Statistics (7th Edition), by Moore, Notz and Fligner.

Relation to Confidence Intervals

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

Here, we will fail to reject H_0 if

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \leq z_{\alpha/2}$$

We can rewrite the above condition as

$$\mu_0 \in [\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$$

It is the $(1-\alpha)100\%$ confidence interval for μ_0. There is a general relationship between confidence interval problems and hypothesis testing problems.

Z-test for a population mean. Normal distribution. Variance σ^2 is Unknown

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

The only difference is that we need to replace σ by s. The test statistic

$$\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

has a t-distribution with n-1 degrees of freedom. We will reject H_0 if

$$\left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right| > t_{\alpha/2, n-1}$$

Example. Normal distribution

The average adult male height in a certain country is 170 cm. We suspect that the men in a certain city in that country might have a different average height due to some environmental factors. We pick a random sample of size 9 from the adult males in the city and obtain the following values for their heights (in cm):

$$176.2, 157.9, 160.1, 180.9, 165.1, 167.2, 162.9, 155.7, 166.2$$

Assume that the height distribution in this population is normally distributed. Here, we need to decide between

$$H_0 : \mu = 170, \quad H_1 : \mu \neq 170$$

Based on the observed data, is there enough evidence to reject H_0 at significance level $\alpha = 0.05$?

From the data

$$\bar{x} = 165.8, \quad s^2 = \frac{1}{9-1}\sum_{i=1}^{9}(x_i - \bar{x})^2 = 68.01, \quad s = \sqrt{68.01} = 8.25$$

Example. Normal distribution

$$H_0 : \mu = 170, \quad H_1 : \mu \neq 170$$

Based on the observed data, is there enough evidence to reject H_0 at significance level $\alpha = 0.05$?

From the data

$$\bar{x} = 165.8, \quad s^2 = \frac{1}{9-1} \sum_{i=1}^{9} (x_i - \bar{x})^2 = 68.01, \quad s = \sqrt{68.01} = 8.25$$

Then we need to calculate the test statistic. Since sample is drawn from the normal distribution and σ is unknown, the test statistic is

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{165.8 - 170}{8.25/3} = -1.52$$

From R: qt(1-0.05/2, 9-1) = 2.31 which is $t_{\alpha/2,n-1} = t_{0.05/2,9-1} = t_{0.025,8}$

Example. Normal distribution

$$H_0 : \mu = 170, \quad H_1 : \mu \neq 170$$

Based on the observed data, is there enough evidence to reject H_0 at significance level $\alpha = 0.05$?

The test statistic is

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{165.8 - 170}{8.25/3} = -1.52$$

From R: qt(1-0.05/2, 9-1) = 2.31 which is $t_{\alpha/2,n-1} = t_{0.05/2,9-1} = t_{0.025,8}$
The p-value is the probability below $\bar{x} = 165.8$ plus the probability above $170 + (170 - 165.8) = 174.2$. This can be expressed as

$$P(Z > 1.52) + P(Z < -1.52) = 0.13 \quad or \quad 2P(Z > |1.52|) = 0.13 \quad R : 2 * pnorm(-1.52)$$

Since $|z| = |1.52| \leq t_{\alpha/2,n-1} = 2.31$ we fail to reject H_0. We do not have enough evidence to conclude that the average height in the city is different from the average height in the country at 5% significance level. $P - value = 0.13 > 0.05$.

Summary

- The null hypothesis H_0 states the claim that we are seeking evidence against
- The probability that measures the strength of the evidence against a null hypothesis is called a P-value
- A test statistic calculated from the sample data measures how far the data diverge from what we would expect if the null hypothesis H_0 were true. Large values of the statistic show that the data are not consistent with H_0.
- The smaller the P-value, the stronger the evidence against H_0 provided by the data.
- A null hypothesis is not accepted just because it is not rejected.
- If the H_1 is two sided use $z_{\alpha/2}, t_{\alpha/2,n-1}$
- If the H_1 is one sided use $z_{\alpha}, t_{\alpha,n-1}$