Dec 15, 2022

Engineering Statistics

Week 8: Hypothesis testing-I

©Mustafa Cavus, Ph.D.

Basic concepts

Introduction

Example 1: A soft-drink company may claim that, on average, its cans contain 12 ounce of soda. A government agency may want to test whether or not such cans do contain, on average, 12 ounce of soda.

Example 2: According to the Giving USA Foundation, 75% of the total charitable contributions in 2019 were given by individuals. An economist may want to check if this percentage is still true for this year.

Hypothesis

A hypothesis is **a statement** about the population parameter.

This can be a value of a population probability distribution parameter such as the mean (μ), the variance (σ^2), or the proportion (p).

The goal of testing statistical hypotheses is to determine if a claim or conjecture about some feature of the population, a parameter, is strongly supported by the information obtained from the sample data.

Hypothesis

The formulation of a hypotheses testing problem and then the steps for solving it require a number of definitions and concepts. These key statistical concepts are:

- Null hypothesis and the alternative hypothesis
- Type I and Type II errors
- Level of significance
- Power of the test
- Rejection region
- p-value

Formulating the hypotheses

Consider as a non statistical example a person who has been indicted for committing a crime and is being tried in a court. Based on the available evidence, the judge or jury will make one of two possible decision:

1. The person is not guilty.

2. The person is guilty.

At the outset of the trial, the person is presumed not guilty. The prosecutors efforts are to prove that the person has committed the crime and, hence, is guilty.

Formulating the hypotheses

Null hypothesis

A null hypothesis is a claim (or statement) about a population parameter that is assumed to be true until it is declared false. The null hypothesis is denoted by H_0.

Alternative hypothesis

An alternative hypothesis is a claim about a population parameter that will be true if the null hypothesis is false. The alternative hypothesis is denoted by H_1 or H_A.

Example

At the outset of the trial, the person is presumed not guilty. The prosecutors efforts are to prove that the person has committed the crime and, hence, is guilty.

H_0: The person is not guilty.

H_1: The person is guilty.

The trials begins with the assumption that the null hypothesis is true that is, the person is not guilty. The prosecutor assembles all the possible evidence and presents it in the court to prove that the null hypothesis is false and the alternative hypothesis is true (that is, the person is guilty).

Decisions

Once H_0 and H_1 are formulated, our goal is to analyze the sample data in order to choose between them.

Test statistic

A test statistic is a decision rule which is formulated to lead the investigator to either reject or fail to reject the null hypothesis on the basis of sample evidence.

Decisions

either <u>REJECT H_0</u> or <u>NOT REJECT H_0</u>.

For good reasons many statisticians prefer not say “accept the null hypothesis” instead, they say, “fail to reject the null hypothesis”.

Basic concepts of hypothesis testing by example

- We now introduce the key statistical concepts in the context of a specific problem to help integrate them with intuitive reasoning.

- Here we illustrate the testing of hypotheses concerning a population mean μ. The available data will be assumed to be a random sample of size n from a population of interest.

Example

Can an upgrade reduce the mean transaction time at automated teller machines?

At peak periods, customers are subject to unreasonably long waits before receiving cash. To help alleviate this difficulty, the bank wants to reduce the time it takes a customer to complete a transaction. From extensive records, it is found that the transaction times have a normal distribution with mean 270 and standard deviation 24 seconds. The teller machine vendor suggests that a new software and hardware upgrade will reduce the mean time for a customer to complete a transaction. For experimental verification, a random sample of 38 transaction times will be taken at a machine with the upgrade and the sample mean $\bar{x}$ calculated. How should the result be used toward a statistical validation of the claim that the true (population) mean transaction time is less that 270 seconds?

Example: formulation hypotheses

From extensive records, it is found that the transaction times have a normal distribution with mean 270 and standard deviation 24 seconds.

$$H_0 : \mu = 270$$

The teller machine vendor suggest that a new software and hardware upgrade will reduce the mean time for a customer to complete a transaction.

$$H_1 : \mu < 270$$

Example: decision

Naturally, the sample mean $\bar{x}$, calculated from the measurements of $n = 38$ randomly selected transaction times, should be the basis for rejecting H_0 or not. The question now is:

For what sort of values of $\bar{x}$ should we reject H_0 ?

Because the claim states that μ is low (a left-sided alternative), only low values of $\bar{x}$ can contradict H_0 in favor of H_1. Therefore, a reasonable decision rule should of the form:

Reject H_0 if $\bar{x} \leq c$

Not reject H_0 if $\bar{x} > c$

Test criterion and rejection region

- This decision rule is conveniently expressed (for our example) as

$$\text{Reject } H_0 \text{ if } \bar{x} \leq c$$

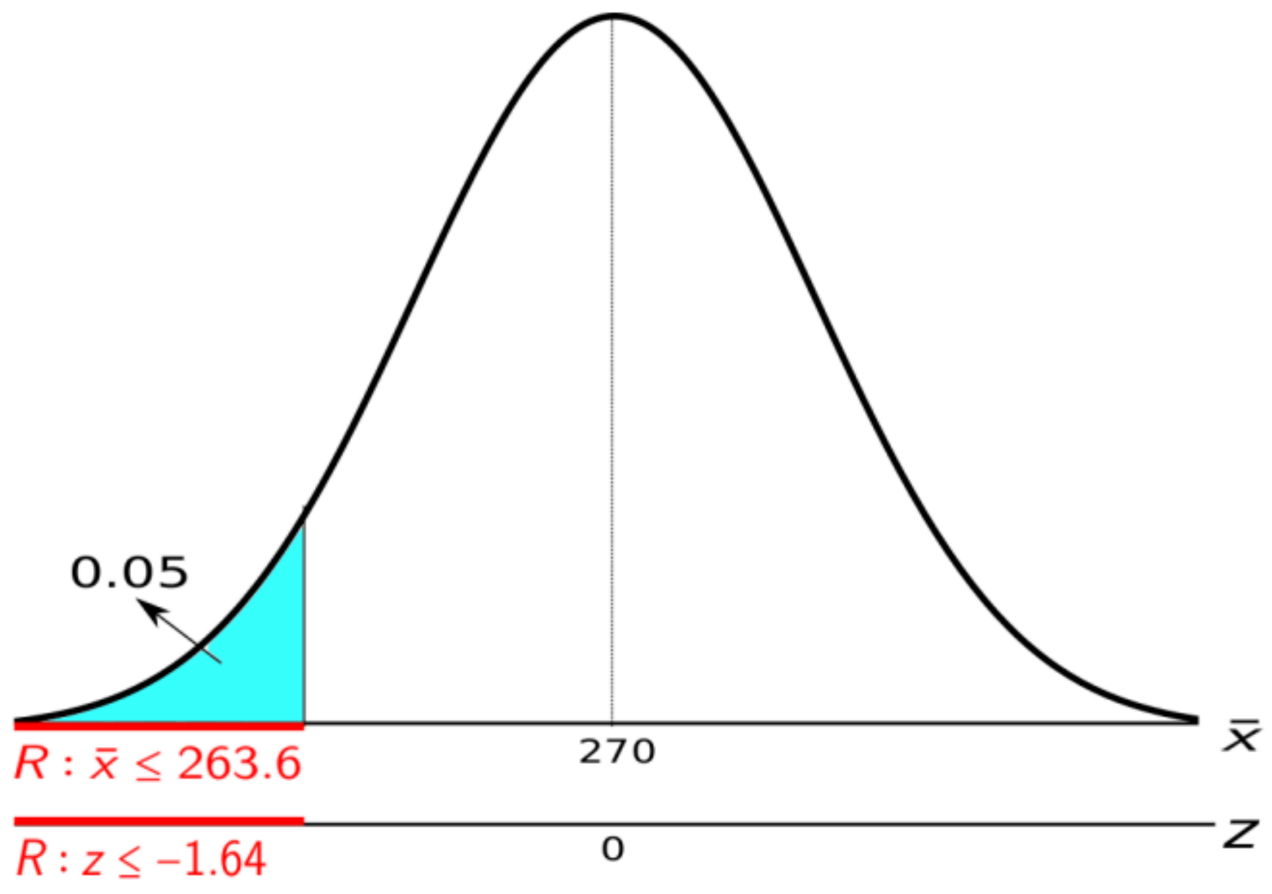
- The cut-off point c is called the critical value.

- The set of all sample points for which H_0 is rejected is called as rejection region or critical region, i.e. R = set of all sample points for which H_0 is rejected = $[\bar{X} \leq c]$

Test criterion and rejection region

- The cut-off point c must be specified in order to fully describe a decision rule.
- For example, suppose that we wish to hold a low probability of $\alpha = 0.05$ for a wrong rejection of H_0. Then, for our problem the task is to find the c that makes

$$P[\bar{X} \leq c] = 0.05 \text{ when } \mu = 270$$



Two types of errors and their probabilities

Decision	Unknown true situation	
	H_0 true	H_0 false
Reject H_0	Type I error	Correct decision
Not reject H_0	Correct decision	Type II error

Two types of errors and their probabilities

Type I error and level of significance

- Type I error is the rejection of a true null hypothesis.
- The probability of the Type I error is called as level of significance and denoted by α.

$$\alpha = P(\text{Reject } H_0 \,|\, H_0 \text{ true})$$

Two types of errors and their probabilities

Type II error

- Type II error is the failure to reject a false null hypothesis.
- The probability of the Type II error is denoted by β.

$$\beta = P(\text{Not reject } H_0 \,|\, H_0 \text{ false})$$

Two types of errors and their probabilities

Power of the test

- Power of the test is the probability of rejecting a null hypothesis that is false.
- Power of the test is denoted by $1 - \beta$.

$$1 - \beta = P(\text{Reject } H_0 \,|\, H_0 \text{ false})$$

Example (cont.)

In our problem of evaluating the upgraded teller machine, the rejection region is of the form $R : \bar{X} \leq c$; so that,

$$\alpha = P(\bar{X} \leq c) \text{ when } \mu = 270 \text{ } (H_0 \text{ is true})$$

$$\beta = P(\bar{X} > c) \text{ when } \mu < 270 \text{ } (H_0 \text{ is false})$$

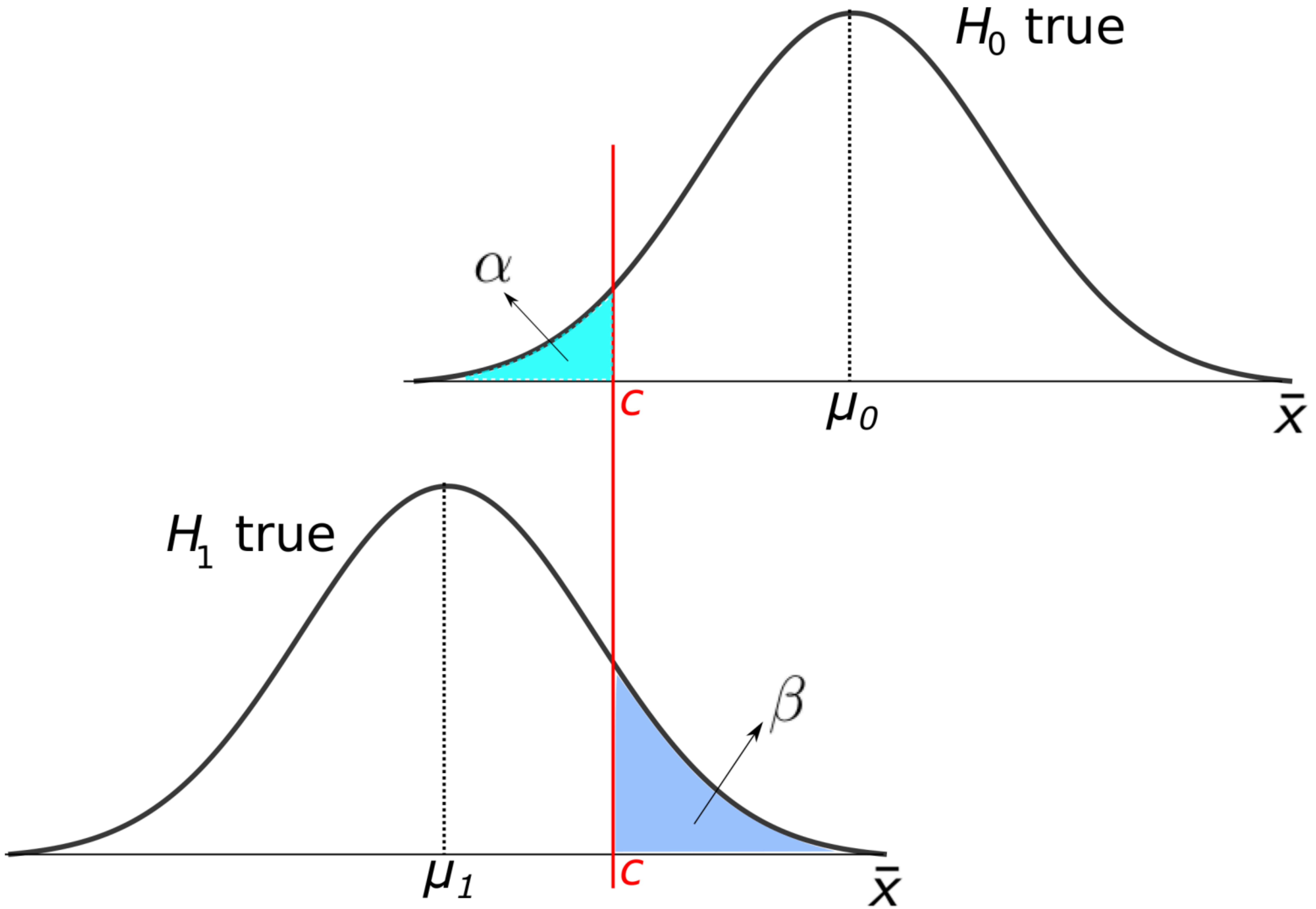
Example (cont.)

In our problem of evaluating the upgraded teller machine, the rejection region is of the form $R : \bar{X} \leq c$; so that,

$$\alpha = P(\bar{X} \leq c) \text{ when } \mu = 270 \text{ (}H_0\text{ is true)}$$

$$\beta = P(\bar{X} > c) \text{ when } \mu < 270 \text{ (}H_0\text{ is false)}$$

Example (cont.)



Example (cont.)

Ideally, we would like to have the probabilities of both types of errors be as small as possible.

However, there is a trade-off between the probabilities of two types of errors.

Given a particular sample, any reduction in the probability of Type I error, α, will result in an increase in the probability of Type II error, β, and vice versa.

Investigator chooses a constant level of significance (probability of Type I error = α) at beginning of the study. It is therefore chosen one of the following values in most of the studies:

$$\alpha = 0.01, 0.05, 0.10$$

p-value

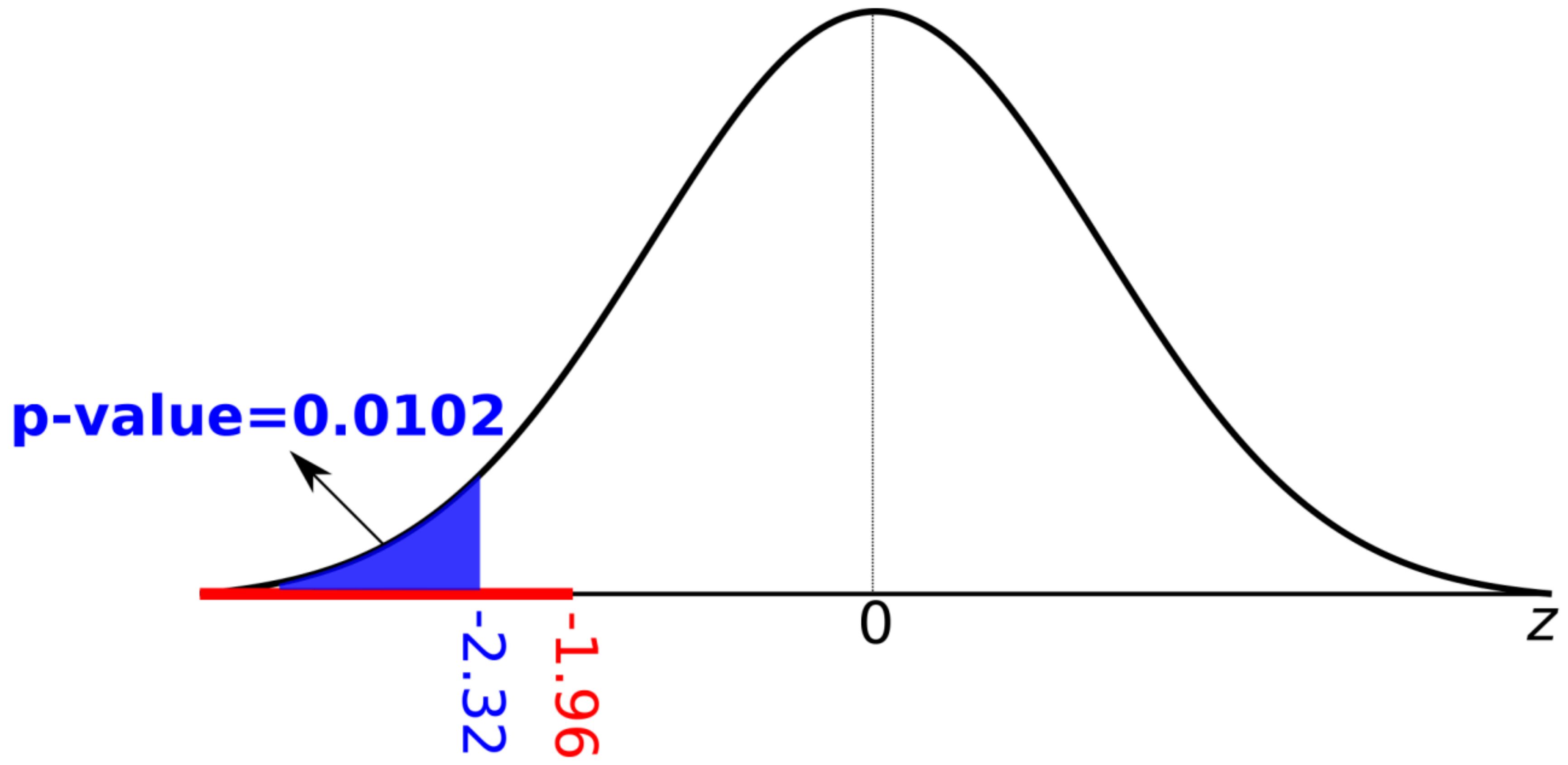
- The p-value is the probability, calculated under H_0, that the test statistic takes a value equal to or more extreme than the value actually observed.

- The p-value serves as a measure of the strength of evidence against H_0.

- A small p-value means that the null hypothesis is strongly rejected or the result is highly statistically significant.

Example (cont.)

Suppose that, from the measurements of a random sample of 38 transaction times, the sample mean is found to be 261 seconds. Calculate the p-value.



$$p - value = P[Z \leq -2.32] = 0.0102 \quad < \quad \alpha = P[\underbrace{Z \leq -1.96}_{R}] = 0.05$$

$$\Downarrow$$

reject H_0

The steps for testing hypotheses

1. State the null hypothesis H_0 and the alternative hypothesis H_1.
2. Test criterion: State the test statistic and the form of the rejection region.
3. With a specified α, determine the rejection region.
4. Calculate the test statistic from the data.
5. Draw a conclusion: State whether or not H_0 is rejected at the specified and interpret the conclusion in the context of the problem. Also, it is a good statistical practice to calculate the p-value and strengthen the conclusion.

Notes

- Unless otherwise specified, H_0 is mostly stated as a simple hypothesis. In other words, it specifies a single value of the parameter, i.e.

$$H_0 : \mu = \mu_0$$

  where μ_0 is a known real number.

- On the other hand, H_1 is usually stated as a composite hypothesis. In other words, it specifies a range of value for the parameter, i.e.

$$H_1 : \mu < \mu_0,\ H_1 : \mu > \mu_0,\ H_1 : \mu \neq \mu_0$$

Notes

- $H_1 : \mu < \mu_0$ and $H_1 : \mu > \mu_0$ are called **one-sided hypotheses**, because the values of the parameter μ under the alternative hypothesis lie on one side of those under the null hypothesis. The corresponding tests are called **one-sided tests** or **one-tailed tests**.

- $H_1 : \mu \neq \mu_0$ is called as **two-sided alternative** or **two-sided hypothesis**. The corresponding test is called **two-sided test** or **two-tailed test**.

Tests for population mean μ

Z-test for testing population mean μ

<u>Assumptions:</u>

(i) $X_1, X_2, \ldots, X_n$ is a random sample drawn from a normal distribution with mean μ and variance σ^2.

(ii) Population mean μ is unknown but population variance σ^2 is known.

We want to test $H_0 : \mu = \mu_0$ versus one of the following alternatives at given α level of significance:

$$H_1 : \mu < \mu_0,\ H_1 : \mu > \mu_0,\ H_1 : \mu \neq \mu_0$$

Z-test for testing population mean μ

The following Z-test statistic is used to test H_0:

$$T_Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

It is well-known that $T_Z \sim N(0,1)$.

Z-test for testing population mean μ

The rejection region is one or two-sided depending on the alternative hypothesis. Specifically,

Alternative hypothesis	Rejection region	p-value
$H_1 : \mu < \mu_0$	$T_Z < -Z_\alpha$	$p < \alpha$
$H_1 : \mu > \mu_0$	$T_Z > Z_\alpha$	$p < \alpha$
$H_1 : \mu \neq \mu_0$	$\lvert T_Z \rvert > Z_{\alpha/2}$	$p < \alpha/2$

Z-test for testing population mean μ

- A known σ^2 is not realistic in real life problems.
- If sample size n is large enough, we can use the following version of Z-test statistic to test H_0

$$T_Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

where s is sample standard deviation. In that case, the distribution of T_Z is approximately $N(0,1)$ even the population distribution is not needed to be normal. Therefore, level of rejection α regions remain same.

Example

An investigator at a large midwestern university wants to determine the typical weekly amount of time students work on part-time jobs. More particularly, he wants to test the null hypothesis that the mean time is 15 hours versus a two-sided alternative. A sample of 39 students who hold part-time jobs is summarized by the computer output:

Variable	n	mean	median	standard deviation
hours	39	16.69	15	7.61

(a) Perform the hypothesis test at the 1% level of significance.

(b) Perform the hypothesis test at the 5% level of significance.

(c) Calculate the p-value and interpret the results.

Solution

(a) Perform the hypothesis test at the 1% level of significance.

$H_0 : \mu = 15$

$H_1 : \mu \neq 15$

$$T_Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{16.69 - 15}{7.61/\sqrt{39}} = 1.386$$

$Z_{0.005} = -2.575$ or $Z_{0.995} = 2.575$ (We use two bounds for critical region because the alternative hypothesis H_1 is two-sided. <u>If it consists > or <, it is one-sided.</u>)

Because of $Z_{0.005} = -2.575 < T_Z = 1.386 < Z_{0.995} = 2.575$, we do not reject H_0. (The calculated T_Z is not in the critical regions, thus we do not reject H_0.)

Solution

(b) Perform the hypothesis test at the 5% level of significance.

$H_0 : \mu = 15$

$H_1 : \mu \neq 15$

$$T_Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{16.69 - 15}{7.61/\sqrt{39}} = 1.386$$

$Z_{0.025} = -1.96$ or $Z_{0.975} = 1.96$ (We use two bounds for critical region because the alternative hypothesis H_1 is two-sided. <u>If it consists > or <, it is one-sided.</u>)

Because of $Z_{0.025} = -1.96 < T_Z = 1.386 < Z_{0.975} = 1.96$, we do not reject H_0. (The calculated T_Z is not in the critical regions, thus we do not reject H_0.)

Solution

(b) Perform the hypothesis test at the 5% level of significance.

$H_0 : \mu = 15$

$H_1 : \mu \neq 15$

$$T_Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{16.69 - 15}{7.61/\sqrt{39}} = 1.386$$

p-value ~ 0.0868 > $\alpha = 0.05$ thus we do not reject H_0. (If p-value < α, we reject H_0.)

We calculated p-value by using the standard normal distribution table:

P(Z = 1.386) ~ 0.0868

t-test for testing population mean μ

<u>Assumptions:</u>

(i) $X_1, X_2, \ldots, X_n$ is a random sample drawn from a normal distribution with mean μ and variance σ^2.

(ii) Population mean μ and population variance σ^2 are both unknown.

We want to test $H_0 : \mu = \mu_0$ versus one of the following alternatives at given α level of significance:

$$H_1 : \mu < \mu_0,\ H_1 : \mu > \mu_0,\ H_1 : \mu \neq \mu_0$$

t-test for testing population mean μ

The following t-test statistic is used to test H_0:

$$T_t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

It is well-known that $T_t \sim t_{(n-1)}$.

t-test for testing population mean μ

The rejection region is one or two-sided depending on the alternative hypothesis. Specifically,

Alternative hypothesis	Rejection region	p-value
$H_1 : \mu < \mu_0$	$T_t < -t_{n-1,\alpha}$	$p < \alpha$
$H_1 : \mu > \mu_0$	$T_t > t_{n-1,\alpha}$	$p < \alpha$
$H_1 : \mu \neq \mu_0$	$\lvert T_t \rvert > t_{n-1,\alpha/2}$	$p < \alpha/2$

t-test for testing population mean μ

Z-test vs. t-test

- Both tests are essentially based on a normal population and testing the population mean.

- If population variance σ^2 is unknown, we can use either Z-test or t-test, but we should be careful that if:

  - n is large enough ($n \geq 30$), use Z-test.

  - n is large enough ($n < 30$), use t-test.

Example

A process that produces bottles of shampoo, when operating correctly, produces bottles whose contents weight, on average, 20 ounces. A random sample of nine bottles from a single production run yielded the following content weights (in ounces):

21.4, 19.7, 19.7, 20.6, 20.8, 20.1, 19.7, 20.3, 20.9

Assuming that the population distribution is normal, test the null hypothesis at the 10% level against alternative that the process is produces bottles whose weights are greater than 20.

Example

$H_0 : \mu \leq 20$

$H_1 : \mu > 20$

$\bar{x} = 20.35$ (calculated on the given sample in the question)

$s = 0.61$ (calculated on the given sample in the question)

$n = 9$

$\alpha = 0.10$

$$T_t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{20.35 - 20}{0.61/\sqrt{9}} = 1.22$$

$t_{8,0.10} = 1.397$ (We use a bound for critical region because the alternative hypothesis H_1 is one-sided. <u>If it consists $\neq$, it is called as two-sided.</u>)

Because of $T_t = 1.22 < t_{8,0.10} = 1.397$, we do not reject H_0. (The calculated T_t is not in the critical region, we do not reject H_0.)

Example

A city health department wishes to determine if the mean bacteria count per unit volume of water at a lake beach is within the safety level of 200. A researcher collected 10 water samples of unit volume and found the bacteria counts to be

175, 190, 205, 193, 184, 207, 204, 193, 196, 180

Do the data strongly indicate that there is no cause for concern? Test with $\alpha = 0.05$ (Assume that population distribution is normal.)

Solution

$H_0 : \mu \leq 200$

$H_1 : \mu > 200$

$\bar{x} = 192.7$ (calculated on the given sample in the question)

$s = 10.81$ (calculated on the given sample in the question)

$n = 10$

$\alpha = 0.05$

$$T_t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{192.7 - 200}{10.81/\sqrt{10}} = -2.135$$

$t_{9,0.05} = 1.833$ (We use a bound for critical region because the alternative hypothesis H_1 is one-sided. If it consists $\neq$, it is called as two-sided.)

Because of $T_t = -2.135 < t_{9,0.05} = 1.833$, we do not reject H_0. (The calculated T_t is not in the critical region, thus we do not reject H_0.)

Tests for population variance σ^2

χ^2 test for testing population variance σ^2

Assumptions:

(i) $X_1, X_2, \ldots, X_n$ is a random sample drawn from a normal distribution with mean μ and variance σ^2.

(ii) Population mean μ is unknown but population variance σ^2 is known.

We want to test $H_0 : \sigma^2 = \sigma_0^2$ versus one of the following alternatives at given α level of significance:

$$H_1 : \sigma < \sigma_0,\ H_1 : \sigma > \sigma_0,\ H_1 : \sigma \neq \sigma_0$$

χ^2 test for testing population variance σ^2

The following χ^2 test statistic is used to test H_0:

$$T_C = \frac{(n-1)s^2}{\sigma_0^2}$$

where s is the standard deviation of the sample. It is clear that $T_C \sim \chi^2_{n-1}$.

χ^2 test for testing population variance σ^2

The rejection region is one or two-sided depending on the alternative hypothesis. Specifically,

Alternative hypothesis	**Rejection region**	**p-value**
$H_1 : \sigma^2 < \sigma_0^2$	$T_C < \chi^2_{n-1,1-\alpha}$	$p < \alpha$
$H_1 : \sigma^2 > \sigma_0^2$	$T_C > \chi^2_{n-1,\alpha}$	$p < \alpha$
$H_1 : \sigma^2 \neq \sigma_0^2$	$T_C < \chi^2_{n-1,1-\alpha/2}$	$p < \alpha/2$

Example

One company, actively pursuing the making of green gasoline, starts with biomass in the form of sucrose and converts it into gasoline using catalytic reactions. At one step in a pilot plant process, the output includes carbon chains of length 3. Fifteen runs with same catalyst produced the product volumes (liter)

2.79, 2.88, 2.09, 2.32, 3.51, 3.31, 3.17, 3.62, 2.79, 3.94, 2.34, 3.62, 3.22, 2.80, 2.70

While mean product volume is the prime parameter, it is also important to control variation. Conduct a test with intent of showing that the population variance σ^2 is less than 0.64 liter. Use $\alpha = 0.05$.

Solution

$H_0 : \sigma^2 \geq 0.64$

$H_1 : \sigma^2 < 0.64$

$n = 15$

$s^2 = 0.28$

$\alpha = 0.05$

$$T_C = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(15-1)0.28}{0.64} = 6.125$$

$\chi^2_{14,0.05} = 6.571$ (We use a bound for critical region because the alternative hypothesis H_1 is one-sided. <u>If it consists $\neq$, it is called as two-sided.</u>)

Because of $T_C = 6.125 < \chi^2_{14,0.05} = 6.571$, we reject H_0. (The calculated T_C is in the critical region, we reject H_0.)

Tests for population proportion p

Z-test for testing population proportion p

Let population proportion be p and X is the number having the characteristic in a random sample of size n where n is large enough.

We want to test $H_0 : p = p_0$ versus one of the following alternatives at given α level of significance:

$$H_1 : p < p_0, \ H_1 : p > p_0, \ H_1 : p \neq p_0$$

Z-test for testing population proportion p

The following Z-test statistic is used to test H_0:

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

where $\hat{p} = X/n$ is the sample proportion. It is known that the distribution of T_P is approximately $N(0,1)$ when n is large.

Z-test for testing population proportion p

The rejection region is one or two-sided depending on the alternative hypothesis. Specifically,

Alternative hypothesis	**Rejection region**	**p-value**
$H_1 : p < p_0$	$T_P < - Z_\alpha$	$p < \alpha$
$H_1 : p > p_0$	$T_P > Z_\alpha$	$p < \alpha$
$H_1 : p \neq p_0$	$\lvert T_C \rvert < Z_{\alpha/2}$	$p < \alpha/2$

Example #1

A five-year-old census recorded that 20% of the families in a large community lived below the poverty level. To determine if this percentage has changed, a random sample of 400 families is studied and 70 are found to be living below the poverty level. Does this finding indicate that the current percentage of families earning incomes below the poverty level has changed from what it was five years ago?

Solution #1

$H_0 : p = 0.20$

$H_1 : p \neq 0.20$

$n = 400$

$\hat{p} = 70/400 = 0.175$

$$T_P = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.175 - 0.2}{\sqrt{0.20(1 - 0.20)/400}} = -0.125$$

$Z_{0.975} = 1.96$ or $Z_{0.025} = -1.96$ (We use bounds for critical region because the alternative hypothesis H_1 is two-sided. <u>If it does not consist $\neq$, it is called as one-sided.</u>)

Because of $T_P = -0.125 > Z_{0.025} = -1.96$, we do not reject H_0. (The calculated T_P is not in the critical region, we do not reject H_0.)

Example #2

An independent bank concerned about its customer base decided to conduct a survey of bank customers. Out of 505 customers who returned the survey form, 258 rated the overall bank services as excellent. Test, at level $\alpha = 0.10$, the null hypothesis that the proportion of customers who would rate the overall bank services as excellent is 0.46 versus a two-sided alternative.

Solution #2

Test, at level $\alpha = 0.10$, the null hypothesis that the proportion of customers who would rate the overall bank services as excellent is 0.46 versus a two-sided alternative.

$H_0 : p = 0.46$

$H_1 : p \neq 0.46$

$n = 505$

$\hat{p} = 258/505 = 0.510$

$\alpha = 0.10$

$$T_P = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.510 - 0.46}{\sqrt{0.46(1 - 0.46)/505}} = 2.254$$

$Z_{0.05} = -1.645$ and $Z_{0.95} = 1.645$ (We use two bounds for critical region because the alternative hypothesis H_1 is two-sided. When it consists $\neq$, it is called as two-sided.)

Because of $T_P = 2.554 > Z_{0.95} = 1.645$, we reject H_0. (The calculated T_P is in the critical region, we reject H_0.)