Testing of Hypothesis

Course Taught at SUAD

Dr. Tanujit Chakraborty

Faculty @ Sorbonne

tanujitisi@gmail.com

Quote of the day..

Today's Topics…

- Hypotheses testing procedures
- Summary: Hypothesis Tests and P-values
- Standard Statistical Tests
- Nonparametric Tests (Not in Syllabus)

Hypothesis Testing

What is Hypothesis?

- "A hypothesis is an educated prediction that can be tested" (study.com).
- "A hypothesis is a proposed explanation for a phenomenon" (Wikipedia).
- "A hypothesis is used to define the relationship between two variables" (Oxford dictionary).
- "A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation" (Walpole).

- **Example: Avogadro's Hypothesis(1811)**

"The volume of a gas is directly proportional to the number of molecules of the gas."

$$V = aN$$

Basic Approach: Hypothesis Testing

The approach:

- Conduct a **test on hypothesis**.
  - Hypothesize that one (or more) parameter(s) has (have) some specific value(s) or relationship.
- Make your decision about the parameter(s) based on one (or more) sample statistic(s)
- Accuracy of the decision is expressed as the probability that the decision is incorrect.

Introduction to Hypothesis Testing

- Sometimes, we are not interested in precise estimates of a parameter, but we only want to examine whether a statement about a parameter of interest or the research hypothesis is true or not.

- Another related issue is that once an analyst estimates the parameters on the basis of a random sample, he would like to infer something about the value of the parameter in the population.

- Statistical hypothesis tests facilitates the comparison of estimated values with hypothetical values.

Hypothesis Testing

Statistical inference

Null hypothesis

Sample

Alternative hypothesis

A **hypothesis** about the value of a population parameter is an **assertion** about its value.

Hypothesis Testing

- A useful analogy for statistical decision making is the decision process used by a jury in a legal proceeding.
- In a criminal trial, the null hypothesis is that the defendant is "not guilty" ("not" in the null hypothesis).
- The alternative is, of course, that the defendant is guilty. The "benefit of the doubt" goes to the null hypothesis (i.e., the evidence must be evaluated by beginning with the assumption that the defendant is not guilty).
- The "burden of proof" is on the prosecution to demonstrate that the assumption of not guilty is incorrect (i.e., evidence must be gathered to support the claim of guilty. The defendant need not prove innocence).
- Jurors must only vote for guilty if they are convinced "beyond reasonable doubt" that the defendant is guilty. Basically, the decision process for a juror goes something like this.

Hypothesis Testing

- Assume the defendant is not guilty. Evaluate the evidence in light of this assumption. Could all the evidence presented simply be a coincidence? (That is, could the degree that the evidence points to guilt of the defendant just be explained as coincidence?)

- What is the chance of that? If that chance is small, then vote for guilty. Otherwise, don't vote for guilty. How small is a small chance? Whatever the juror interprets as "beyond reasonable doubt."

- In statistical hypothesis testing, the principles of the trial carry through quite similarly. The null hypothesis gets the "benefit of the doubt."

- The "burden of proof" is on the alternative hypothesis. The experimenter must have a test statistic (evidence) which convinces us that the alternative is true rather than the null hypothesis.

- The same question must be asked as we make a statistical decision as that asked above by the juror.

Statistical Hypothesis

- If the hypothesis is stated in terms of population parameters (such as mean and variance), the hypothesis is called statistical hypothesis.
- Data from a sample (which may be an experiment) are used to test the validity of the hypothesis.
- A procedure that enables us to agree (or disagree) with the statistical hypothesis is called a **test of the hypothesis.**

Example :

1. To determine whether the wages of men and women are equal.
2. A product in the market is of standard quality.
3. Whether a particular medicine is effective to cure a disease.

The Hypotheses

- The main purpose of statistical hypothesis testing is to choose between two competing hypotheses.

Example : One hypothesis might claim that wages of men and women are equal, while the **alternative** might claim that men make more than women.

- Hypothesis testing start by making a set of two statements about the parameter(s) in question.

- The hypothesis actually to be tested is usually given the symbol H_0 and is commonly referred as the **null hypothesis.**

- The other hypothesis, which is assumed to be true when null hypothesis is false, is referred as the **alternate hypothesis** and is often symbolized by H_1

- The two hypotheses are exclusive and exhaustive.

The Hypotheses

Example:

Ministry of Education takes an initiative to improve the country's human resources.

To measure the engineering aptitudes of graduates, they conducts an examination for a mark of 1000 in every year. A sample of 300 students who gave the examination in 2020 were collected and the mean is observed as 220.

In this context, statistical hypothesis testing is to determine the mean mark of all the examinee.

The two hypotheses in this context are:

$$H_0 : \mu = 220$$

$$H_1 : \mu < 220$$

The Hypotheses

Note:

1. As null hypothesis, we could choose $H_0\!: \mu \leq 220$ or $H_0\!: \mu \geq 220$
2. It is customary to always have the null hypothesis with an equal sign.
3. As an alternative hypothesis there are many options available with us.

Examples:

I. $H_1\!: \mu > 220$
II. $H_1\!: \mu < 220$
III. $H_1\!: \mu \neq 220$

4. The two hypothesis should be chosen in such a way that they are **exclusive** and **exhaustive**.
   - One or other must be true, but they cannot both be true.

The Hypotheses

One-tailed test

- A statistical test in which the alternative hypothesis specifies that the population parameter lies entirely above or below the value specified in H_0 is called a one-sided (or one-tailed) test.

Example: $H_0: \mu = 100$ $H_1: \mu > 100$

Two-tailed test

- An alternative hypothesis that specifies that the parameter can lie on their sides of the value specified by H_0 is called a two-sided (or two-tailed) test.

Example: $H_0: \mu = 100$ $H_1: \mu <> 100$

One- and Two-Tails Tests

Case	Null Hypothesis	Alternative Hypothesis	Tails of Test
(a)	$\theta = \theta_0$	$\theta \neq \theta_0$	Two tail test
(b)	$\theta = \theta_0$	$\theta < \theta_0$	Left tail test
(c)	$\theta = \theta_0$	$\theta > \theta_0$	Right tail test

The Hypotheses

- Simple and Composite Hypothesis:
  - A statistical hypothesis is some statement about a population, which we want to verify on the basis of information available from a sample.
  - If the statistical hypothesis specifies the population completely then it is termed as a simple hypothesis otherwise it is called composite hypothesis.
  - Example: If $X_1, X_2, \dots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$ population, then the hypothesis

$$H_0: \quad \mu = \mu_0, \qquad \sigma^2 = \sigma_0^2$$

Example of Composite Hypothesis:

i. $\mu = \mu_0$

ii. $\mu < \mu_0, \quad \sigma^2 = \sigma_0^2$

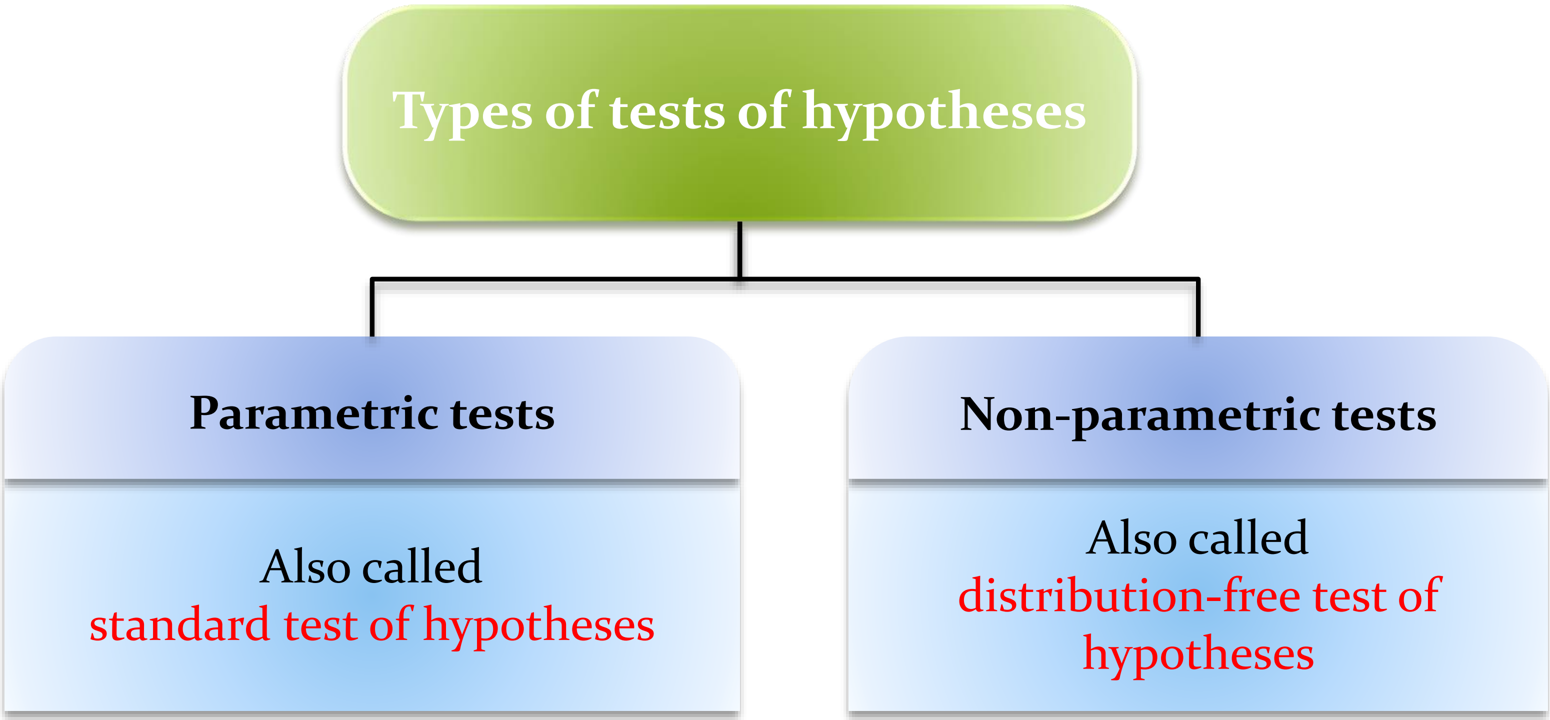
iii. $\mu < \mu_0, \quad \sigma^2 > \sigma_0^2$ etc.

One- and Two-sample Problems

- In one-sample problems, the data is usually assumed to arise as one sample from a defined population.

- In two-sample problem, the data originates in the form of two sample possibly from two different populations.

Hypothesis Testing Strategies



Note: Hypothesis testing determines the validity of an assumption (technically described as null hypothesis), with a view to choose between two conflicting hypothesis about the value of a population parameter.

Parametric Tests : Assumptions

Usually assume certain properties of the population from which we draw samples.

- Observation come from a normal population
- Sample size is small
- Population parameters like mean, variance, etc. hold good.
- Requires measurement dealing with interval scaled data.

Hypothesis Testing: Non-Parametric Test

Non-parametric tests

- Does not come under any assumption.
- Assumes nominal, ordinal as well as interval scale data.

Note:

Non-parametric tests need entire population (or sample of very large size)

Hypothesis Testing Procedures

The following five steps are followed when testing hypothesis

1. Specify H_0 and H_1, the null and alternate hypothesis, and an **acceptable level of α.**

2. Determine an appropriate sample-based test statistics and the **rejection region** for the specified H_0.

3. Collect the sample data and calculate the test statistics.

4. Make a decision to either reject or fail to reject H_0.

5. Interpret the result in common language suitable for practitioners.

Hypothesis Testing Procedure

- In summary, we must choose between H_0 and H_1
- The standard procedure is to assume H_0 is true.
  (Just we presume innocent until proven guilty)
- Using statistical test, we try to determine whether there is sufficient evidence to declare H_0 false.
- We reject H_0 only when the **chance is small** that H_0 is true.
- The procedure is based on probability theory, that is, there is a chance that we can make errors.

Errors in Hypothesis Testing

In hypothesis testing, there are two types of errors.

Type I error: A type I error occurs when we incorrectly reject H_0 (i.e., we reject the null hypothesis, when H_0 is true).

Type II error: A type II error occurs when we incorrectly fail to reject H_0 (i.e., we accept H_0 when it is not true).

Decision	Observation: H_0 is true	Observation: H_0 is false
H_0 is accepted	Decision is correct	Type II error
H_0 is rejected	Type I error	Decision is correct

Probabilities of Making Errors

Type I error calculation

$\boldsymbol{\alpha}$**:** denotes the probability of making a Type I error

$$\boldsymbol{\alpha} = \mathbf{P}(\text{Rejecting } H_0 | H_0 \text{ is true})$$

Type II error calculation

$\boldsymbol{\beta}$**:** denotes the probability of making a Type II error

$$\boldsymbol{\beta} = \mathbf{P}(\text{Accepting } H_0 | H_0 \text{ is false})$$

Note:

- $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not independent of each other as one increases, the other decreases
- When the sample size increases, both to decrease since sampling error is reduced.
- In general, we focus on Type I error, but Type II error is also important, particularly when sample size is small.
- So the test statistics are obtained by fixing α and then minimizing β.
- The probability $1 - \boldsymbol{\beta} = \mathbf{P}(\mathrm{H}_0 \text{ is rejected} | \mathrm{H}_0 \text{ is Not True})$ is called the power of the test.

Calculating α

Assuming that we have the results of random sample. Hence, we use the characteristics of sampling distribution to calculate the probabilities of making either Type I or Type II error.

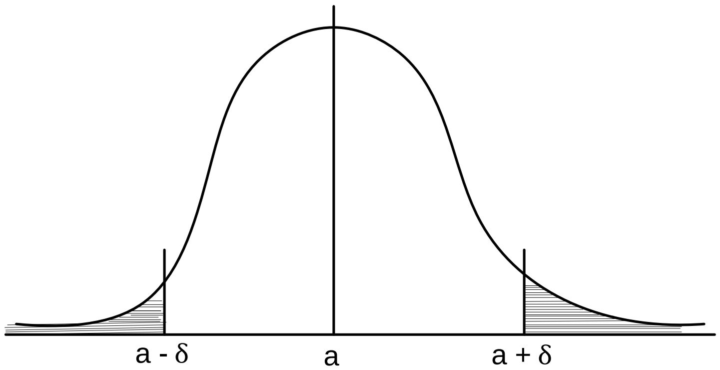
Example :

Suppose, two hypotheses in a statistical testing are:

$$H_0 : \mu = a$$
$$H_1 : \mu \neq a$$

Also, assume that for a given sample, population obeys normal distribution. A threshold limit say $a \pm \delta$ is used to say that they are significantly different from a.

Calculating α



Here, shaded region implies the probability that, $\overline{\mathrm{X}} < a - \delta \ \ or \ \ \overline{\mathrm{X}} > a + \delta$

Thus the null hypothesis is to be rejected if the mean value is less than $a - \delta$ or greater than $a + \delta$.

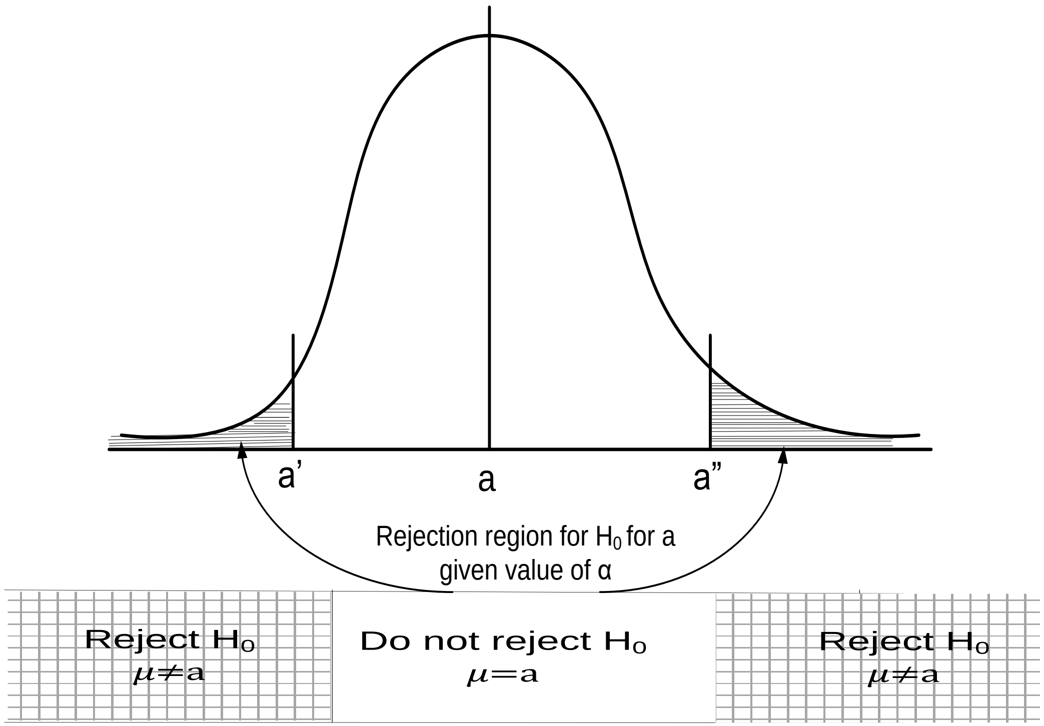
If $\overline{\mathrm{X}}$ denotes the sample mean, then the Type I error is

$$\alpha = P(\overline{\mathrm{X}} < a - \delta \ or \ \overline{\mathrm{X}} > a + \delta, \qquad when \ \mu = a\ , \qquad \text{i.e.,}\ H_0 \text{ is true})$$

The Rejection Region

The rejection region comprises of value of the test statistics for which

1. The probability when the null hypothesis is true is less than or equal to the specified α.
2. Probability when H_1 is true are greater than they are under H_0.



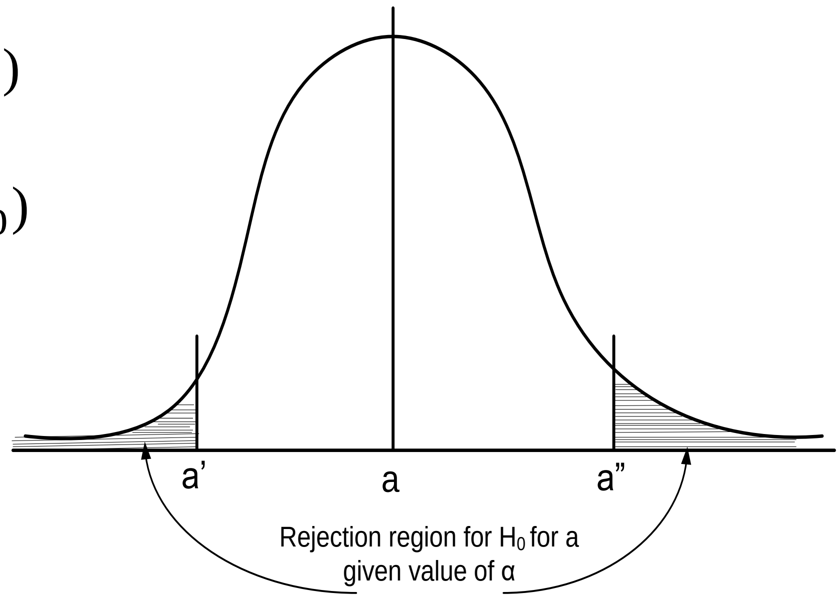
The Significance Level α

- $\alpha = 0\%$
  - Always **accepts** the null hypothesis (Ultra liberal test)

- $\alpha = 100\%$
  - Always **rejects** the null hypothesis (Ultra conservative test)

- $\alpha = 1\%$ (Lesser probability to reject a H_0)

- $\alpha = 5\%$ (Higher probability to reject a H_0)



Note:
α is the maximum acceptable probability of rejecting a true null hypothesis

Two-Tailed Test

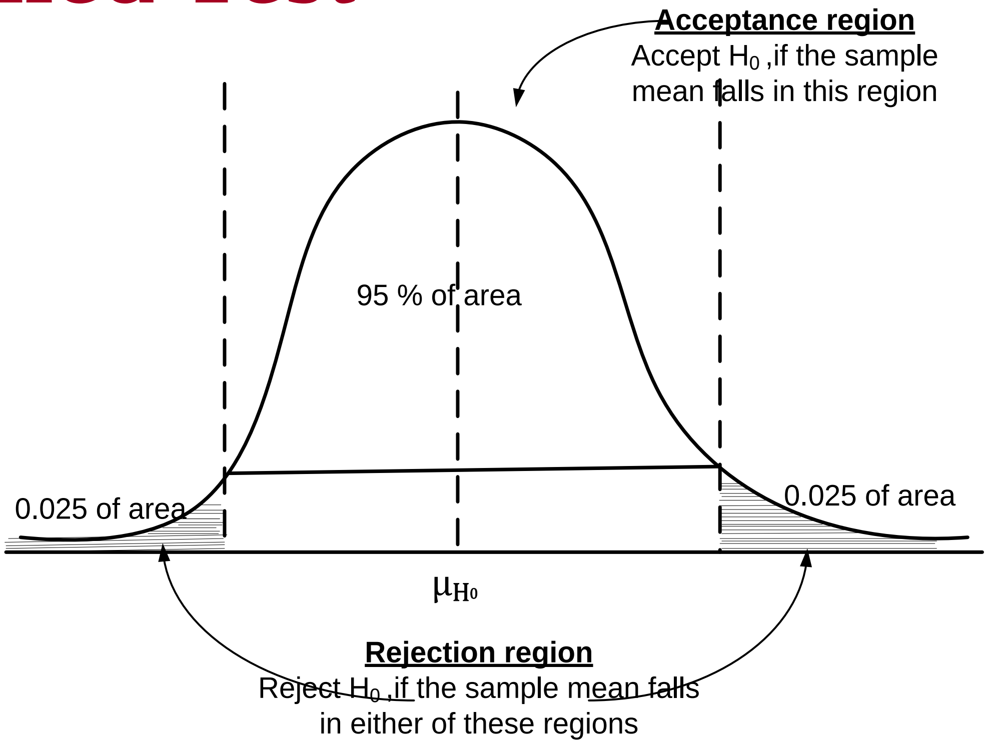
For two-tailed hypothesis test, hypotheses take the form

$$H_0: \mu = \mu_{H_0}$$
$$H_1: \mu \neq \mu_{H_0}$$

In other words, to reject a null hypothesis, sample mean $\mu > \mu_{H_0}$ or $\mu < \mu_{H_0}$ under a given α.

Thus, in a two-tailed test, there are two rejection regions (also known as critical region), one on each tail of the sampling distribution curve.

Two-Tailed Test



Acceptance and rejection regions in case of a two-tailed test with 5% significance level.

One-Tailed Test

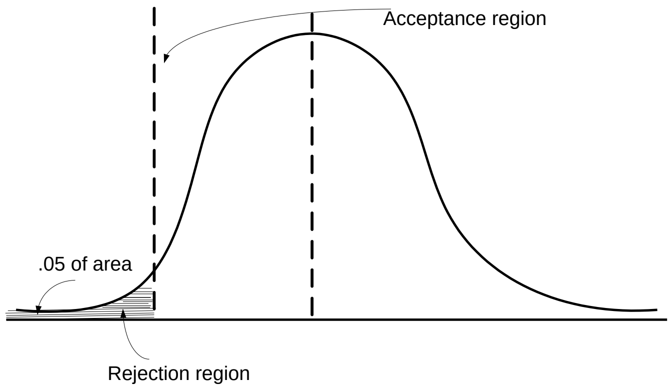
A one-tailed test would be used when we are to test, say, whether the population mean is either lower or higher than the hypothesis test value.

Symbolically,

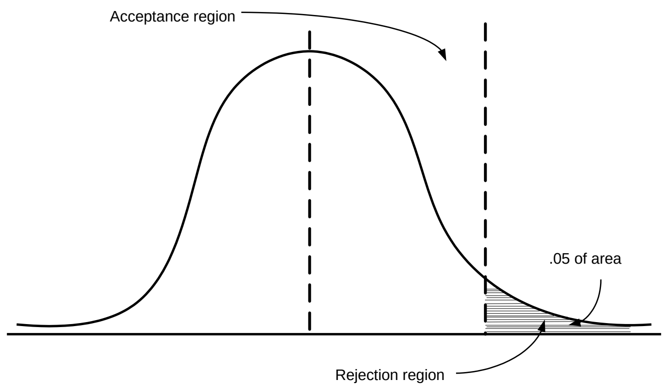
$$H_0 : \mu = \mu_{H_0}$$

$$H_1 : \mu < \mu_{H_0} \qquad [or\ \mu > \mu_{H_0}]$$

Wherein there is one rejection region only on the left-tail (or right-tail).



Left – tailed test



Right – tailed test

Example : Calculating α

Consider the two hypotheses are

The null hypothesis is

$$H_0: \mu = 8$$

The alternative hypothesis is

$$H_1: \mu \neq 8$$

Assume that given a sample of size 16 and standard deviation is 0.2 and sample follows normal distribution.

Example : Calculating α

We can decide the rejection region as follows.

Suppose, the null hypothesis is to be rejected if the mean value is less than 7.9 or greater than 8.1. If $\overline{\mathrm{X}}$ is the sample mean, then the probability of Type I error is

$$\alpha = P(\overline{\mathrm{X}} < 7.9 \ or\ \overline{\mathrm{X}} > 8.1, \text{when } \mu = 8)$$

Given σ, the standard deviation of the sample is 0.2 and that the distribution follows normal distribution.

Thus,

$$P(\overline{\mathrm{X}} < 7.9) = P\left[Z < \frac{7.9 - 8}{0.2/\sqrt{16}}\right] = P[Z < -2.0] = 0.0228$$

and

$$P(\overline{\mathrm{X}} > 8.1) = P\left[Z > \frac{8.1 - 8}{0.2/\sqrt{16}}\right] = P[Z > 2.0] = 0.0228$$

Hence, $\alpha = 0.0228 + 0.0228 = 0.0456$

Example : Calculating α

There are two identically appearing boxes of chocolates. Box A contains 60 red and 40 black chocolates whereas box B contains 40 red and 60 black chocolates. There is no label on the either box. One box is placed on the table. We are to test the hypothesis that "Box B is on the table".

To test the hypothesis an experiment is planned, which is as follows:

- Draw at random five chocolates from the box.
- We replace each chocolates before selecting a new one.
- The number of red chocolates in an experiment is considered as the sample statistics.

Note: Since each draw is independent to each other, we can assume the sample distribution follows binomial probability distribution.

Example : Calculating α

Let us express the population parameter as $p =$ the number of red chocolates in Box B.

The hypotheses of the problem can be stated as:

$$H_0 : p = 0.4 \qquad \text{// Box B is on the table}$$

$$H_1 : p = 0.6 \qquad \text{// Box A is on the table}$$

Calculating α:

In this example, the null hypothesis (H_0) specifies that the probability of drawing a red chocolate is 0.4. This means that, lower proportion of red chocolates in observations $(i.e., sample)$ favors the null hypothesis. In other words, drawing all red chocolates provides sufficient evidence to reject the null hypothesis. Then, the probability of making a $Type\ I$ error is the probability of getting five red chocolates in a sample of five from Box B. That is,

$$\alpha = P(X = 5 \quad when\, p = 0.4)$$

Using the binomial distribution

$$= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \; where\, n = 5, x = 5$$

$$= (0.4)^5 = 0.01024$$

Thus, the probability of rejecting a true null hypothesis is ≈ 0.01. That is, there is approximately 1 in 100 chance that the box B will be mislabeled as box A.

Example : Calculating $\boldsymbol{\beta}$

The $Type\ II$ error occurs if we fail to reject the null hypothesis when it is not true. For the current illustration, such a situation occurs, if Box A is on the table but we did not get the five red chocolates required to reject the hypothesis that Box B is on the table.

The probability of $Type\ II$ error is then the probability of getting four or fewer red chocolates in a sample of five from Box A.

That is,

$$\beta = P(X \leq 4 \qquad \text{when } p = 0.6)$$

Using the probability rule:

$$P(X \leq 4) + P(X = 5) = 1$$

$$\text{That is, } P(X \leq 4) = 1 - P(X = 5)$$

$$\text{Now, } P(X = 5) = (0.6)^5$$

$$\text{Hence, } \beta = 1 - (0.6)^5$$

$$= 1 - 0.07776 = 0.92224$$

That is, the probability of making $Type\ II$ error is over 92%. This means that, if Box A is on the table, the probability that we will be unable to detect it is 0.92.

How to Conduct a Statistical Test?

- Define the distributional assumption for the random variables of interest, and specify them in terms of population parameters.
- Formulate H_0 and H_1.
- Fix a significance value (Type I error) α (say 0.05).
- Construct a test statistic $T(\boldsymbol{X}) = T(X_1, X_2, \dots, X_n)$. The distribution of T has to be known under the null hypothesis H_0.
- Construct a critical region W for the statistic T, i.e. a region where – if T falls in this region- H_0 is rejected, such that
$$P_{H_0}(T(\boldsymbol{X}) \in W) \leq \alpha.$$
- Calculate $t(x) = T(x_1, x_2, \dots, x_n)$ based on the realized sample values
$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n.$$
- Decision Rule:
  i. $t(x) \in W$: H_0 *is rejected* $\Rightarrow$ H_1 *is statistically significant,*
  ii. $t(x) \notin W$: H_0 *is not rejected*

P-values…

- P-value is a measure of consistency between the null hypothesis and the observed data.

- Hence, we will be more inclined to believe the alternative hypothesis when the P-value is small and less likely to believe the alternative hypothesis when the P-value is not so small.

- We make a rule for statistical decisions as follows. We establish a cut-off value called the level of significance.

- Then we reject the null hypothesis in favor of the alternative hypothesis when the P-value is less than the level of significance.

- If the P-value is not less than the level of significance, then we fail to reject the null hypothesis in favor of the alternative hypothesis.

P-values…

- If the P-value is small, the data is inconsistent with H_0, so we reject H_0 in favor of H_a.

- If the P-value is not small, the data is not inconsistent with H_0, so we fail to reject H_0 in favour of H_a.

- The natural question at this point is "How small does the P-value have to be to be small?" (or, "What amount of evidence is beyond reasonable doubt?")

- If the P-value is less than α, then we reject H_0 in favor of H_a.

- **Remember :** P-value is NOT the probability that the null hypothesis is true.

α	**Interpretation** (if p-value is less than the given α)	**Strength of Evidence** (if p-value is less than the given α)
0.10	Approaching statistical significance	Potential Evidence
0.05	Statistically significant	Sufficient Evidence
0.01	Highly statistically significant	Strong Evidence
0.001	Extremely statistically significant	Very Strong Evidence

Test Decision Using the p-value

- It is possible to use the p-value instead of critical regions for making test decisions.

- The p-value of the test statistic $T(\boldsymbol{X})$ is defined as follows:

Right-sided case: $P(T \geq t(x)|H_0 \; is \; true) = p - value$

Left-sided case: $P(T \leq t(x)|H_0 \; is \; true) = p - value$

Two sided case:

$$2\min\{P(T \leq t(x)|H_0 \; is \; true), P(T \geq t(x)|H_0 \; is \; true)\} = p \; value$$

- If $p - value < 0.01$: very strong evidence against H_0, i.e., very strongly Reject H_0.
- If $0.01 < p - value < 0.05$: strong evidence against H_0, i.e., strongly Reject H_0.
- If $0.05 < p - value < 0.10$: some weak evidence against H_0, i.e., Not Reject H_0.
- If $p - value > 0.10$: no evidence against H_0, i.e. Not Reject H_0.

Case Study 1: **Coffee Sale**

A coffee vendor nearby Howrah railway station has been having average sales of 500 cups per day. Because of the development of a bus stand nearby, it expects to increase its sales. During the first 12 days, after the inauguration of the bus stand, the daily sales were as under:

550 570 490 615 505 580 570 460 600 580 530 526

On the basis of this sample information, can we conclude that the sales of coffee have increased?

Consider 5% level of confidence.

Hypothesis Testing : 5 Steps

The following five steps are followed when testing hypothesis

1. Specify H_0 and H_1, the null and alternate hypothesis, and an **acceptable level of α.**

2. Determine an appropriate sample-based test statistics and the **rejection region** for the specified H_0.

3. Collect the sample data and calculate the test statistic.

4. Make a decision to either reject or fail to reject H_0.

5. Interpret the result in common language suitable for practitioner.

Case Study 1: **Step 1**

Step 1: Specification of hypothesis and acceptable level of α

Let us consider the hypotheses for the given problem as follows.

$H_0: \mu = 500$ cups per day

The null hypothesis that sales average 500 cups per day and they have not increased.

$H_1: \mu > 500$

The alternative hypothesis is that the sales have increased.

Given the acceptance level of $\alpha = 0.05$ $(i.e., 5\%\ level\ of\ significance)$

Case Study 1: **Step 2**

Step 2: Sample-based test statistics and the rejection region for specified H_0

Given the sample as

550 570 490 615 505 580 570 460 580 530 526

Since the sample size is small and the population standard deviation is not known, we shall use $t - test$ assuming normal population. The test statistics t is

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

To find $\bar{X}$ and S, we make the following computations.

$$\bar{X} = \frac{\sum x_i}{n} = \frac{6576}{12} = 548$$

Case Study 1: **Step 2**

Sample #	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	550	2	4
2	570	22	484
3	490	-58	3364
4	615	67	4489
5	505	-43	1849
6	580	32	1024
7	570	22	484
8	460	-88	7744
9	600	52	2704
10	580	32	1024
11	530	-18	324
12	526	-22	484
$n = 12$	$\sum X_i = 6576$		$\sum (X_i - \bar{X})^2 = 23978$

Case Study 1: **Step 2**

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{23978}{12-1}} = 46.68$$

Hence, $t = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{48}{46.68/\sqrt{12}} = \frac{48}{13.49} = 3.558$

Note:

Statistical table for t-distributions gives a *t*-value given *n*, the degrees of freedom and α, the level of significance and vice-versa.

Case Study 1: **Step 3**

Step 3: Collect the sample data and calculate the test statistics

$$Degree\ of\ freedom = n - 1 = 12 - 1 = 11$$

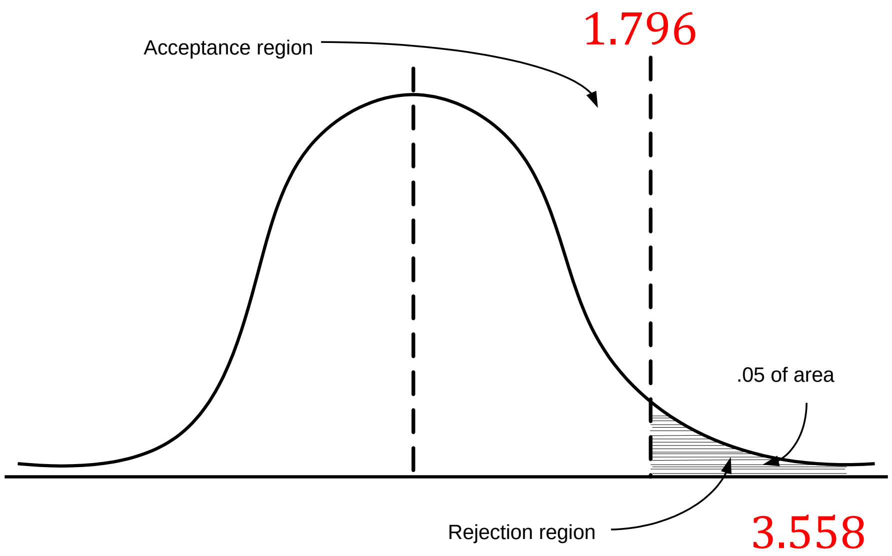
As H_1 is one-tailed, we shall determine the rejection region applying one-tailed in the right tail because H_1 is more than type) at 5% level of significance.

Using table of $t - distribution$ for 11 degrees of freedom and with 5% level of significance,

$$R: t > 1.796$$

Case Study 1: **Step 4**

Step 4: Make a decision to either reject or fail to reject H_0



The observed value of $t = 3.558$ which is in the rejection region and thus H_0 is rejected at 5% level of significance.

Case Study 1: **Step 5**

Step 5: Final comment and interpret the result

We can conclude that the sample data indicate that coffee sales have increased.

Comments on Case Study 1

Step 1: Specification of hypotheses and significance level α

Let us consider the hypotheses for the given problem as follows.

$H_0: \mu = 500$ cups per day

The null hypothesis that sales average 500 cups per day and they have not increased.

$H_1: \mu > 500$

The alternative hypothesis is that the sales have increased.

Given the value of $\alpha = 0.01$ $(i.e., 1\%\ significance\ level\ of\ the\ test)$

Case Study 2: **Machine Testing**

A medicine production company packages medicine in a tube of 8 ml. In maintaining the control of the amount of medicine in tubes, they use a machine. To monitor this control a sample of 16 tubes is taken from the production line at random time interval and their contents are measured precisely. The mean amount of medicine in these 16 tubes will be used to test the hypothesis that the machine is indeed working properly. Maximum variance that can be allowed is 0.2.

Case Study 2: **Step 1**

Step 1: Specification of hypothesis and level of significance α

The hypotheses are given in terms of the population mean of medicine per tube.

The null hypothesis is

$$H_0: \mu = 8$$

The alternative hypothesis is

$$H_1: \mu \neq 8$$

We assume α, the significance level in our hypothesis testing ≈ 0.05.
(This signifies the probability that the machine needs to be adjusted less than 5%).

Case Study 2: **Step 2**

Step 2: Collect the sample data and calculate the test statistics

Sample results: $n = 16,\ \bar{\mathrm{x}} = 7.89, \sigma = 0.2$

With the sample, the test statistics is

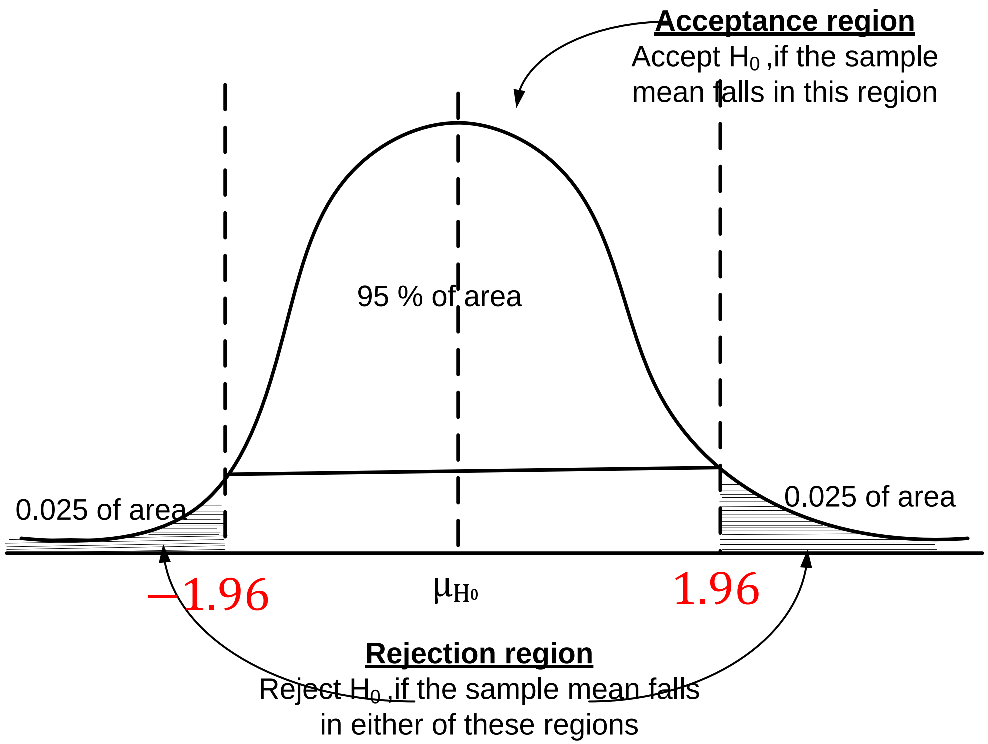
$$z = \frac{7.89-8}{{0.2}/{\sqrt{16}}} = -2.20$$

Hence, $|Z| = 2.20$

Case Study 2: **Step 3**

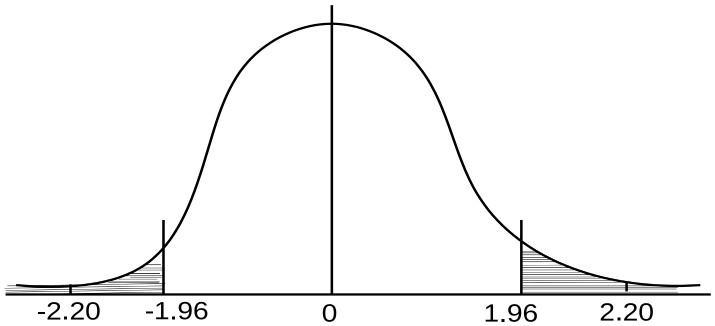
Step 3: To decide the critical region for specified H_0

Rejection region: Given $\alpha = 0.05$**,** which gives $(\mathrm{P}|Z| > 1.96)$ (obtained from standard normal calculation for $n(Z: 0,1) = 0.025$ for a rejection region with two-tailed test).



Case Study 2: **Step 4**

Step 4: Make a decision to either reject or fail to reject H_0



Since $Z > 1.96$, we reject H_0

Case Study 2: **Step 5**

Step 5: Final comment and interpret the result

We conclude $\mu \neq 8$ and recommend that the machine be adjusted.

Case Study 2: **Comment 1 (α = 1%)**

Suppose that in our initial setup of hypothesis test, if we choose $\alpha = 0.01$ instead of 0.05, then the test can be summarized as:

1. $H_0 : \mu = 8$, $H_1 : \mu \neq 8$ $\alpha = 0.01$
2. Reject H_0 if $Z > 2.576$
3. Sample result n =16, σ = 0.2, $\overline{X}$=7.89, Z $= \frac{7.89-8}{0.2/\sqrt{16}} = -2.20$, $|Z| = 2.20$
4. $|Z| < 2.20$, we fail to reject H_0= 8
5. We do not recommend that the machine be adjusted.

Case Study 2: **Comment 1 ($\bar{X}$= 7.91)**

Suppose that in our initial setup of hypothesis test, we choose $\alpha = 0.05$, and the collected sample is $\bar{X}$=7.91 of size 16 with σ = 0.2. In this case, the test can be summarized as:

1. $H_0 : \mu = 8$, $H_1 : \mu \neq 8$ $\alpha = 0.05$
2. Reject H_0 if $Z > 1.96$
3. Sample result n =16, σ = 0.2, $\bar{X}$=7.91, Z $= \frac{7.91-8}{0.2/\sqrt{16}} = -1.80$, $|Z| = 1.80$
4. $|Z| < 1.96$, we fail to reject H_0= 8
5. We do not recommend that the machine be adjusted.

Case Study 2: **Observation**

For = 0.05

- H_0 is rejected with $\overline{X} = 7.89$
- H_0 is not rejected with $\overline{X} = 7.91$

For $\overline{X} = 7.89$

- H_0 is rejected with $\alpha = 0.05$
- H_0 is not rejected with $\alpha = 0.01$

What about with = 0.02, 0.03, … ?

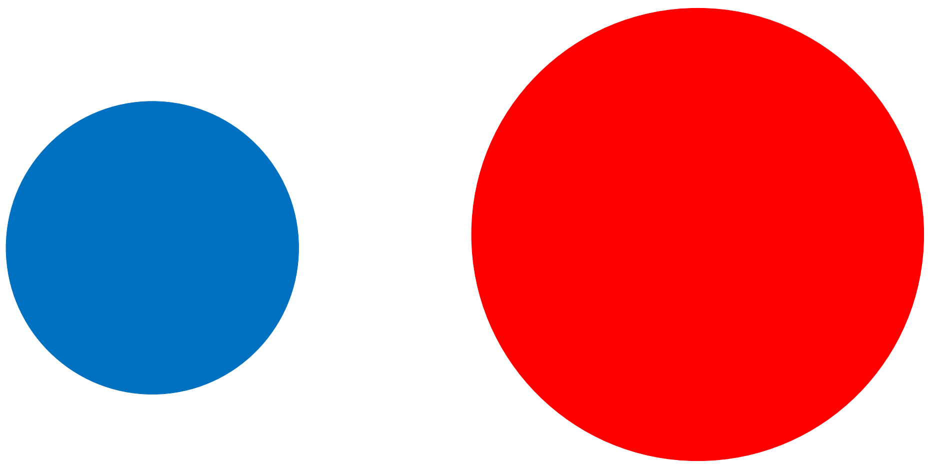
Case Study 3: χ^2-Test

Case Study 3: **Perceptual Psychology**

In perceptual psychology, a person is asked to judge the relative areas of circles of varying sizes. A person typically judges the areas on a perceptual scale that can be approximated by

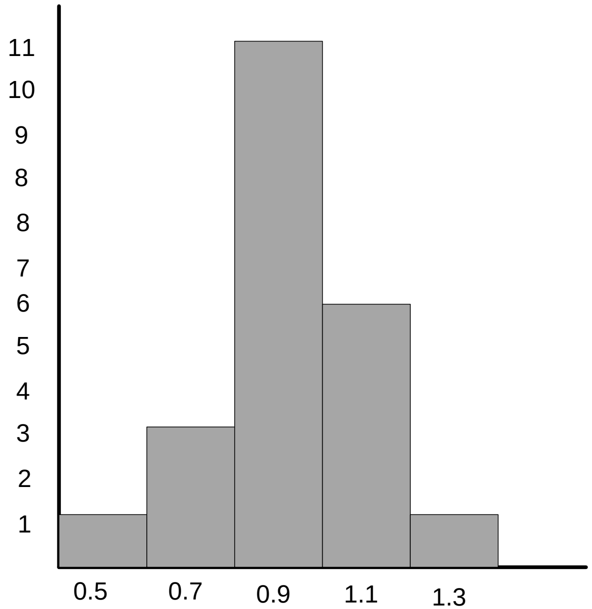
$$Judged\ area = a.(True\ area)^b$$

For most of the people, the exponent b is between 0.6 and 1. That is a person with an exponent of 0.8, who sees two circles, one twice the area of the other, would judge the larger one to be only $2^{0.8} = 1.74$. If the exponent is less than one then the person underestimate the area; if larger than 1, he overestimate the area. Based on an experiment with 24 people, a data on the perceptual psychology is observed.



Case Study 3: **Perceptual Psychology**

Measured exponents in Perceptual Psychology					
0.58	0.63	0.69	0.72	0.74	0.79
0.88	0.88	0.90	0.91	0.93	0.94
0.97	0.97	0.99	0.99	0.99	1.00
1.03	1.04	1.05	1.07	1.18	1.27



Suppose, in the study, that variability of subjects is of concern. Researchers want to know whether the variance of exponents differ from 0.02.

Consider the acceptable level of confidence is 5%.

Case Study 2: **Step 1**

Step 1: Specification of hypothesis and significance level α

The hypotheses of interest is given by

The null hypothesis is

$$H_0:\ \sigma^2 = \ 0.02$$

The alternative hypothesis is

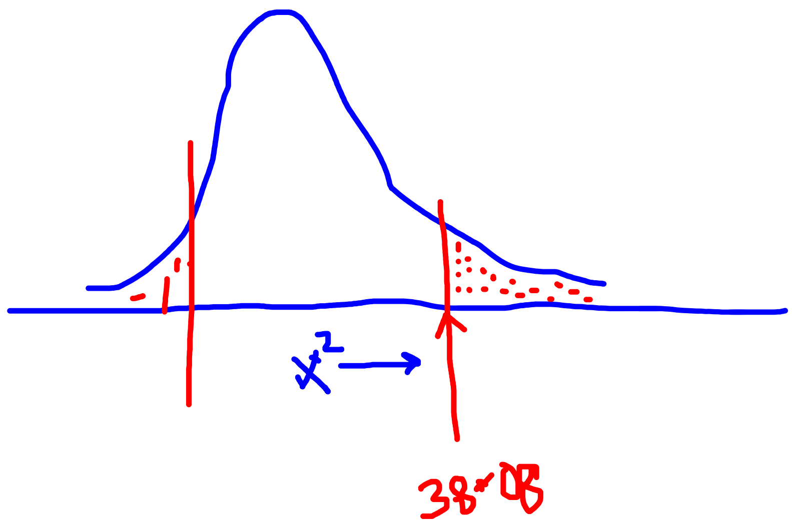
$$H_1:\ \sigma^2 \neq 0.02$$

We assume α, the significance level in our hypothesis testing ≈ 0.05.

Case Study 2: **Step 2**

Step 2: Decide the rejection region for specified hypothesis testing

Rejection region: Given $\alpha = 0.05$**,** and with degree of freedom = 24-1 = 23, the χ^2 value of the critical region is 38.08.



Case Study 2: **Step 3**

Step 3: Collect the sample data and calculate the test statistics

Sample results: $n = 24,\ SS = 0.628, \sigma^2 = 0.02$

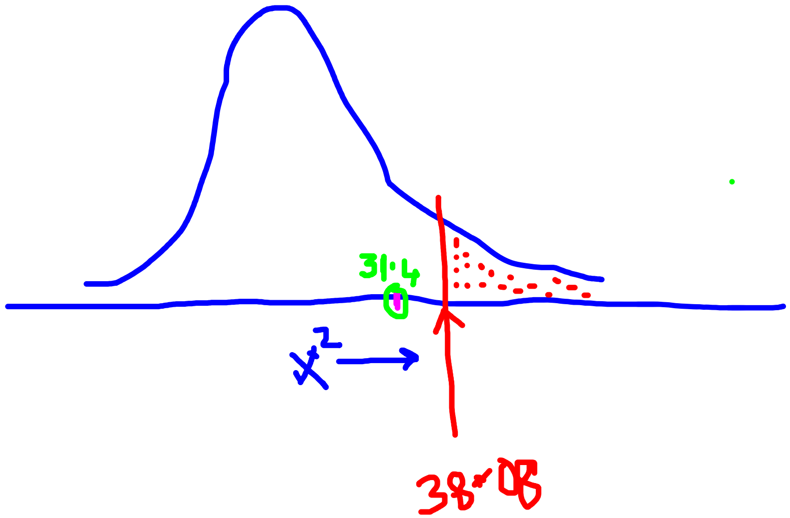
With the sample, the test statistics is

$$\chi^2 = \frac{SS}{\sigma^2} = \frac{0.628}{0.02} = 31.4$$

Case Study 2: **Step 4**

Step 4: Make a decision to either reject or fail to reject H_0

Since, $\chi^2 = 31.40 < 38.08$, we cannot reject the null hypothesis.



Case Study 2: **Step 5**

Step 5: Final comment and interpret the result

We conclude that the sample variance does not significantly differ from 0.02.

p Value Concept

p Value Concept

Issues

1. Many users do not have a fixed or definitive idea of what should be an appropriate value for α in hypothesis testing.
2. Using a specified level of significance, a decision differs even for a minor change in sample statistics.

Need

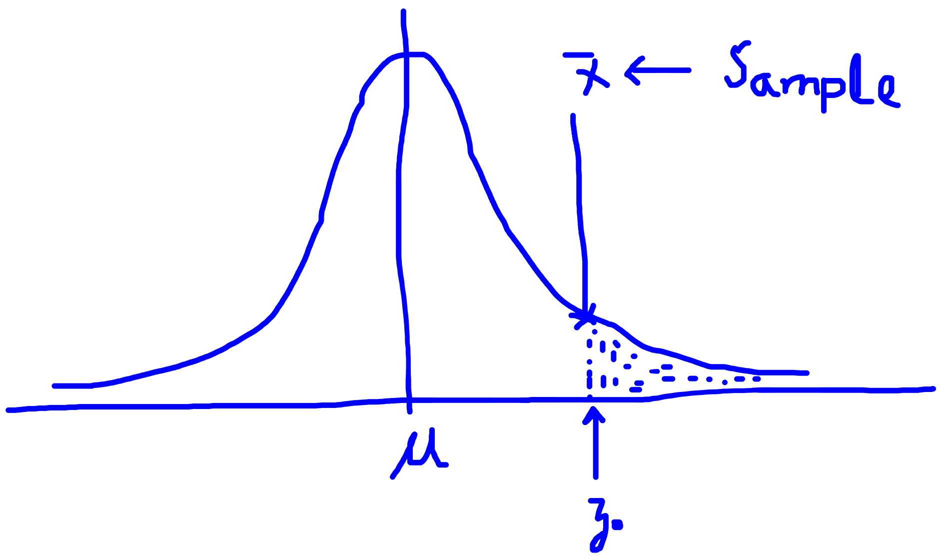
- There should be a method of reporting the results of a hypothesis without having to chose an exact value of level of significance.
  - It can be left to the decision maker who will use the test result.

The method of reporting results is referred to as report the p-value of a test.

p Value: Definition

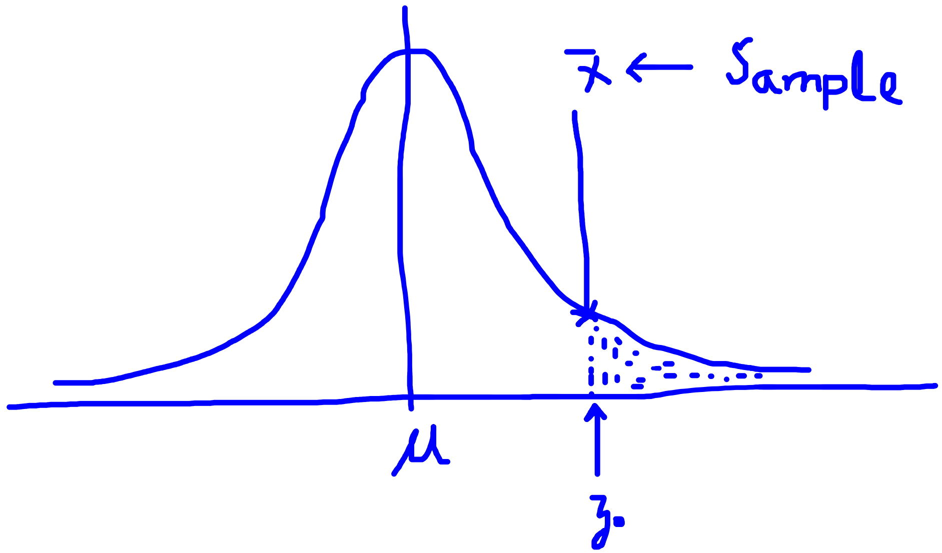
Definition

- p value is the probability of committing Type I Error if the actual sample value of the statistics is used as the boundary of the rejection region.
  - It is the smallest level of significance for which H_0 is to be rejected.



p Value: Definition

Definition



$$p = P(Z > z_0) \qquad \text{for one-tailed test}$$

$$p = 2 \times P(Z > z_0) \qquad \text{for two-tailed test}$$

p Value: Example

Example:

H_0: $\mu = 8.0$

H_0: $\mu \neq 8.0$

n =16, σ = 0.2, $\overline{X}$ = 7.89

$$z_0 = \frac{7.89-8.0}{0.2/\sqrt{16}} = -2.20$$

$$p = 2 \times P(|Z| > z_0)$$

$$= 2 \times 0.0139$$

$$= 0.0278$$

$$\approx 3\%$$

p Value: Interpretations

Many interpretation:

- The example implies that the probability of Type-I error is $\approx$ 3% with the considered sample.
- The null hypothesis is rejected with level of significance 0.0278 or higher.
- The inference of population mean $\mu = 8$ is acceptable with $\approx$ 3% error (or 97% test accuracy).
- Here, no need to specify significance level a priori.
- Reporting results with p-value is a better information for decision makers from data analysis.

Parametric Tests

Important Parametric Tests

The widely used sampling distribution for parametric tests are

- $Z - test$
- $t - test$
- $\chi^2 - test$
- $F - test$

Note:

All these tests are based on the **assumption of normality** (i.e., the source of data is considered to be normally distributed).

Parametric Tests : Z-test

$Z-test$**:** This is most frequently test in statistical analysis.

- It is based on the normal probability distribution.
- Used for judging the significance of several statistical measures particularly the mean.
- It is used even when $binomial\, distribution$ or $t-distribution$ is applicable with a condition that such a distribution tends to normal distribution when n becomes large.
- Typically, it is used for comparing the mean of a sample to some hypothesized mean for the population in case of large sample, or when **population variance** is known.

Single Sample: Test Concerning a Single Mean

- Null Hypothesis: H_0: $\mu = \mu_0$ (Variance known)

- Value of Test Statistic: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

- Significance Level: α

Alternative Hypothesis H_1	Critical Region
$\mu < \mu_0$ $\mu > \mu_0$ $\mu \neq \mu_0$	$z < -z_\alpha$ $z > z_\alpha$ $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

Exercise:

A company manufacturing automobile tyres finds that tyre-life is normally distributed with mean of 40,000km and standard deviation of 3000 km. It is believed that a change in the production process will result in a better product and the company has developed a new tyre. A sample of 100 new tyres has been selected. The company has found that the mean life of these new tyres is 40,900 km. Can it be concluded that the new tyre is significantly better than the old one, using the significant level of 0.01?

Parametric Tests : t-test

$\boldsymbol{t-test}$: It is based on the t-distribution.

- It is considered an appropriate test for judging the significance of a sample mean or for judging the significance of difference between the means of two samples in case of
  - small sample(s)
  - **population variance is not known** (in this case, we use the variance of the sample as an estimate of the population variance)

Single Sample: Test Concerning a Single Mean

- Null Hypothesis: $H_0: \ \mu = \mu_0$ (Variance unknown)

- Value of Test Statistic: $t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

- Significance Level: α

-

Alternative Hypothesis H_1	Critical Region
$\mu < \mu_0$	$t < -t_{\alpha,n-1}$
$\mu > \mu_0$	$t > t_{\alpha,n-1}$
$\mu \neq \mu_0$	$t < -\, t_{\frac{\alpha}{2},n-1}$ or $t > \ t_{\frac{\alpha}{2},n-1}$

Example

A manufacturer of electric batteries claims that the average capacity of a certain type of battery that the company produces is at least 140 ampere hour with a standard deviation of 2.66 ampere hour. An independent sample of 20 batteries gave a mean of 138.47 ampere hour. Test a 5% significance level $H_0 = 140$ against $H_1 < 140$.Can the manufacturer claim be sustained on the basis of this sample ?

Two Sample: Test on Two Mean

- Null Hypothesis: H_0: $\mu_1 - \mu_2 = d_0$ (σ_1 and σ_2 known)
- Value of Test Statistic: $z = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$
- Significance Level: α
-

Alternative Hypothesis H_1	Critical Region
$\mu_1 - \mu_2 < d_0$	$z < -z_\alpha$
$\mu_1 - \mu_2 > d_0$	$z > z_\alpha$
$\mu_1 - \mu_2 \neq d_0$	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

Example

The alkalinity, in milligrams per litre, of water in the upper reaches of rivers in a particular region is known to be normally distributed with a standard deviation of 10 mg/l. Alkalinity readings in the lower reaches of rivers in the same region are also known to be normally distributed, but with a standard deviation of 25 mg/l. Ten alkalinity readings are made in the upper reaches of a river in the region and fifteen in the lower reaches of the same river with the following results.

Upper reaches	91	75	91	88	94	63	86	77	71	69
Lower reaches	86	95	135	121	68	64	113	108	79	62
	143	108	121	85	97					

Investigate, at the 1% level of significance, the claim that the true mean alkalinity of water in the lower reaches of this river is greater than in the upper reaches.

Two Sample: Test on Two Mean

- Null Hypothesis: H_0: $\mu_1 - \mu_2 = d_0$ ($\sigma_1 = \sigma_2$ but unknown)
- Value of Test Statistic: $t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$
- $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
- Significance Level: α
-

Alternative Hypothesis H_1	Critical Region
$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_{\alpha,n_1+n_2-2}$ $t > t_{\alpha,n_1+n_2-2}$ $t < -\,t_{\frac{\alpha}{2},n_1+n_2-2}$ or $t > t_{\frac{\alpha}{2},n_1+n_2-2}$

Example

An experiment was performed to compare the abrasive wear of two different laminated materials. Twelve pieces of material 1 were tested by exposing each piece to a machine measuring wear. Ten pieces of material 2 were similarly tested. In each case, the depth of wear was observed. The samples of material 1 gave an average (coded) wear of 85 units with a sample standard deviation of 4, while the samples of material 2 gave an average of 81 with a sample standard deviation of 5. Can we conclude at the 0.05 level of significance that the abrasive wear of material 1 exceeds that of material 2 by more than 2 units? Assume the populations to be approximately normal with equal variances.

Two Sample: Test on Two Mean

- Null Hypothesis: H_0: $\mu_1 - \mu_2 = d_0$ ($\sigma_1 \neq \sigma_2$ but unknown)
- Value of Test Statistic: $t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_\nu$
- Degrees of Freedom: $\nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$
- Significance Level: α
-

Alternative Hypothesis H_1	Critical Region
$\mu_1 - \mu_2 < d_0$ $\mu_1 - \mu_2 > d_0$ $\mu_1 - \mu_2 \neq d_0$	$t < -t_{\alpha,\nu}$ $t > t_{\alpha,\nu}$ $t < -t_{\frac{\alpha}{2},\nu}$ or $t > t_{\frac{\alpha}{2},\nu}$

Parametric Tests : $\boldsymbol{\chi^2}$ -test

$\boldsymbol{\chi^2 - test}$: It is based on Chi-squared distribution.

- It is used for comparing a sample variance to a theoretical population variance.

Parametric Tests : $\boldsymbol{F}$ -test

$\boldsymbol{F - test}$: It is based on F-distribution.

- It is used to compare the variance of two independent samples.
- This test is also used in the context of analysis of variance (ANOVA) for judging the significance of more than two sample means.

Single Sample: Test Concerning Variance

- Null Hypothesis: H_0: $\sigma^2 = \sigma_0^2$

- Value of Test Statistic: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$

- Significance Level: α

-

Alternative Hypothesis H_1	Critical Region
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2{}_{1-\alpha,n-1}$
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2{}_{\alpha,n-1}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{1-\frac{\alpha}{2},n-1}$ or $\chi^2 > \chi^2_{\frac{\alpha}{2},n-1}$

Example

A manufacturer of car batteries claims that the life of the company's batteries is approximately normally distributed with a standard deviation equal to 0.9 year. If a random sample of 10 of these batteries has a standard deviation of 1.2 years, do you think that $\sigma > 0.9$ year? Use a 0.05 level of significance.

Two Sample: Test Concerning Variance

- Null Hypothesis: H_0: $\sigma_1^2 = \sigma_2^2$
- Value of Test Statistic: $\mathrm{F} = \frac{s_1^2}{s_2^2} \sim F_{v_1,v_2}$, where $v_1 = n_1 - 1,\ v_2 = n_2 - 1$
- Significance Level: α

Alternative Hypothesis H_1	Critical Region
$\sigma_1^2 < \sigma_2^2$	$F < F_{1-\alpha}(v_1, v_2)$
$\sigma_1^2 > \sigma_2^2$	$F > F_{\alpha}(v_1, v_2)$
$\sigma_1^2 \neq \sigma_2^2$	$F < F_{1-\frac{\alpha}{2}}(v_1, v_2) \quad or \quad F > F_{\alpha/2}(v_1, v_2)$

Example

In testing for the difference in the abrasive wear of the two materials in the previous example we assumed that the two unknown population variances were equal. Were we justified in making this assumption? Use a 0.10 level of significance.

Consultation of Statistical Tables

How to find $\mathbf{t-}, \mathbf{z-}$ and $\boldsymbol{\chi^2}$ values?

Method 1: Based on PDF of a distribution

$$f(z:0,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz$$

Method 2: To find a $\mathbf{t-}, \mathbf{z-}$ and $\boldsymbol{\chi^2}$ value using statistical tables

χ^2 Test for Goodness of Fit

- Suppose we are sampling from a distribution $F(x)$ which may depend upon a parameter $\boldsymbol{\theta}$.
- We want to test

$$H_0 : F(x) = F_0(x) \qquad \forall \ x$$
$$H_1 : F(x) \neq F_0(x) \text{ for at least some } x.$$

Where $F_0(x)$ is a known cumulative distribution function.

- We follow the given procedure:
  - Divide the range of the distribution in k mutually exclusive and exhaustive intervals: $I_1, I_2, \ldots, I_k$.
  - Let us assume that $P(X \in I_i) = \pi_i, \ i = 1,2, \ldots, k$.(Theoretical Probability)
  - The observed frequencies $O_1, O_2, \ldots, O_k$ be the respective observed number of observations in the intervals $I_1, I_2, \ldots, I_k$.
  - The vector $\boldsymbol{O} = (O_1, O_2, \ldots, O_k)$ has a multinomial distribution.
  - $P(O_1 = o_1, O_2 = o_2, \ldots, O_k = o_k) = \frac{n!}{\prod_{i=1}^{k} o_i!} \prod_{i=1}^{k} (\pi_i)^{o_i}$, where $\sum o_i = n, \sum \pi_i = 1,$

π_i is the probability of the ith interval.

χ^2 Test for Goodness of fit

- $E(O_i) = n\,\pi_i = e_i$ (say), $Var(O_i) = n\pi_i(1-\pi_i), \quad i = 1,2,\dots,k.$
- For $k = 2$, the distribution of O_1 is Binomial (n, π_1).
- Thus $\frac{O_1 - n\pi_1}{\sqrt{n\pi_1(1-\pi_1)}}$ follows approximately N(0,1).
- Hence, $\frac{(O_1 - n\pi_1)^2}{n\pi_1(1-\pi_1)}$ follows χ^2 distribution with 1 degree of freedom.
- Using $O_2 = n - O_1$, $\pi_2 = 1 - \pi_1$ (for $k = 2$)

$$\frac{(O_1 - n\pi_1)^2}{n\pi_1} + \frac{(O_1 - n\pi_2)^2}{n\pi_2} = \frac{(O_1 - n\pi_1)^2}{n\pi_1(1-\pi_1)}.$$

- Hence, we observed that $\sum_{i=1}^{2} \frac{(O_i - e_i)^2}{e_i}$ follows χ^2 distribution with 1 degree of freedom.
- Thus for general k, the quantity

$$W = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{k-1}.$$

- H_0 is rejected if $W \geq \chi^2_{\alpha,k-1}$.

χ^2 Test for Goodness of Fit

- When F_0 is not completely specified. That means, it may includes certain parameters (say m parameters).

- In this case we can estimate (we may use MLE) the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$ from the sample and accordingly find $\hat{\pi}_i$ and $\hat{e}_i = n\hat{\pi}_i$.

- So we can use

$$W^* = \sum_{i=1}^{k} \frac{(O_i - \hat{e}_i)^2}{\hat{e}_i}.$$

- W^* follows χ^2_{k-m-1} distribution.

- H_0 is rejected if $W^* \geq \chi^2_{\alpha, k-m-1}$.

Examples

- In a sale of 300 units of an item the following preferences for colours are observed for customers:

Colour	Brown	Grey	Red	Blue	White	Total
Customers	88	65	52	40	55	300

Test the hypothesis that all colours are equally popular.

- One wants to investigate the distribution of the number of claims for medical treatments over a given period by families. A previous study suggested that the distribution may be Poisson. To investigate this a random sample of 200 families is taken with the following classifications.

No. of Claims	0	1	2	3	4	5	6	7	Total
Frequency	22	53	58	39	20	5	2	1	200

Test whether a Poisson distribution fits the data appropriately.

Important notes on Hypothesis Testing

Considerations in Hypothesis Testing

Hypothesis testing is sensitive to…

1. Acceptable level of significance, α = 10%, 5%, 3%, 2%, 1%, etc.
2. z-test, t-test of χ^2-test?
3. Selection of a sample and hence observed values of $\overline{X}$ and S.
4. Size of the sample and repetition of test with different samples.
5. Reporting results with p values without specification of level of significance.
6. Type-I Error of Type-II Error in testing?
7. ?????.

Hypothesis Testing : Assumptions

Case 1: Normal population, population infinite, sample size may be large or small, variance of the population is known.

$$z = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}}$$

Case 2: Population normal, population finite, sample size may large or small………variance is known.

$$z = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}[\sqrt{(N-n)/(N-1)}]}$$

Case 3: Population normal, population infinite, sample size is small and variance of the population is unknown.

$$t = \frac{\bar{X} - \mu_{H_0}}{S/\sqrt{n}} \qquad with\ degree\ of\ freedom = (n-1)$$

and

$$s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{(n-1)}}$$

Hypothesis Testing

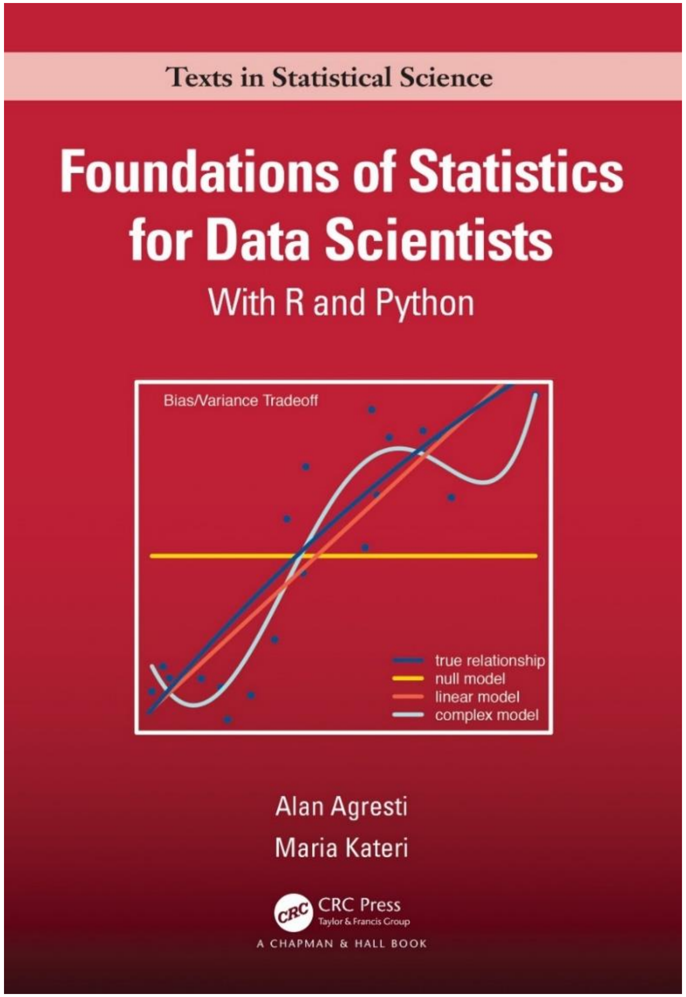
Case 4: Population is normal, finite, variance is known and sample with small size

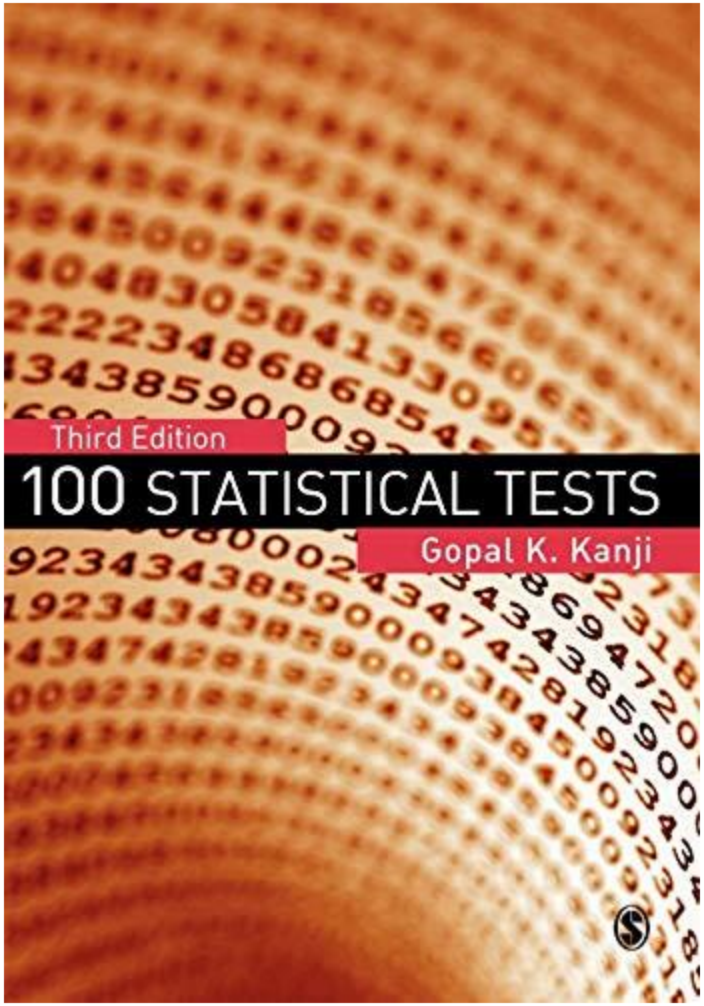
$$t = \frac{\bar{X} - \mu_{H_0}}{\sigma/\sqrt{n}[\sqrt{(N-n)/(N-1)}]} \quad with\ degree\ of\ freedom = (n-1)$$

Note:

If variance of population (σ) is known, replace S by σ.

Reference


Texts in Statistical Science
Foundations of Statistics for Data Scientists
With R and Python
Bias/Variance Tradeoff
true relationship
null model
linear model
complex model
Alan Agresti
Maria Kateri
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK


Third Edition
100 STATISTICAL TESTS
Gopal K. Kanji