MDM 4U

REPEATED SAMPLING & HYPOTHESIS TESTING

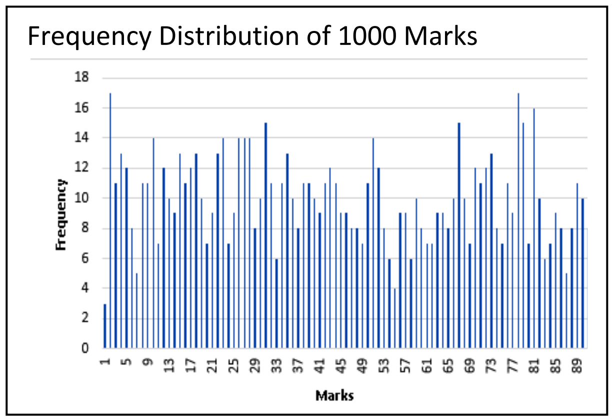
PURPOSE of REPEATED SAMPLING:

- to express a large set of data as a ***normal*** set of data
- to use a ***normal curve*** to represent the data analyzed
- to perform probability tests on the normal data

① ***REPEATED SAMPLING***

Central Limit Theorem: When independent random variables are added, their sum tends toward a normal distribution even if the original variables themselves are not normally distributed. The **CLT** states that given a sufficiently large sample size from a population, the ***mean of all samples*** from the same population will be approximately equal to the ***mean of the population***.

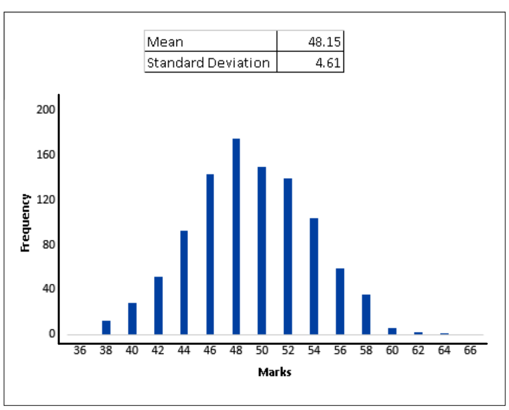
Consider the following case: Marks of 1000 Students



Recall a frequency diagram is often used to illustrate a single variable, in this case, the frequency of marks.

How many random samples of 40 students can be taken from this population? Statistically, 25 different samples: (1000/40 = 25). [It is possible to take even more 40-student samples, some groups will contain an overlap of randomly chosen students.] Will every sample have the same average as the population average (48.4)? Ideally, it is desirable, but practically every sample is unlikely to have the same average as the population average.

A random sample generator was used to take 1000 samples of 40 students. The frequency distribution of the ***sample averages*** of the 1000 samples is illustrated below:



The following properties are observed:

1. Mean of sample means is very close to the population mean.

2. The distribution of sample means is ***normal*** regardless of the distribution of the actual population. This is the ***Central Limit theorem***.

3. ***Standard deviation of the sample distribution*** (standard error of means) can be determined from the population standard deviation:

$$\sigma_{\bar{x}} = \frac{\sigma_{pop}}{\sqrt{n}}$$

② HYPOTHESIS TESTING

PURPOSE of HYPOTHESIS TESTING:

- to test a claim that a sample taken from a population is different from its population
- to accept the claim as a new version of the population or to accept the sample as being a random chance

Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. Hypothesis testing is used to infer the result of a hypothesis performed on sample data from a larger population.

Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed. All analysts use a random sample to test two different hypotheses: the null hypothesis and the alternative hypothesis. The ***null hypothesis*** is the hypothesis the analyst believes to be true. Analysts believe the ***alternative hypothesis*** to be untrue, making it effectively the opposite of a null hypothesis. Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true. Since the testing is based on probabilities, the decision to accept or reject a claim cannot be determined **with certainty**, rather we can test the strength of the statement, based on a sample.

CONSIDER the case: The average of grade 12 math students at ABC School is 85%. A random selection of 30 math students has an average of 95%. Two hypotheses can be made:

- The academic behavior of the random sample of 30 students is significantly different from ABC School's total population of grade 12 math students.
- There is no difference at all. The result is due to random chance only. The random sample could have been higher/lower than 85 since there are students having an average score more or less than 85.

How should we decide which explanation is correct? There are various methods to help you to decide this. Here are some options:

1. Increase sample size
2. Test other samples
3. **Calculate random chance probability**

The first two methods require more time & budget; hence, they are not desirable when time or budget are constraints. So, in such cases, a convenient method is to calculate the random chance probability for that sample; that is, what is the probability that the sample would have an average of 95%? It will help you to draw a conclusion from the two options given above. Now the question is, "*How should we calculate the random chance probability?*"
To answer the question, we review some underlying concepts of statistics.

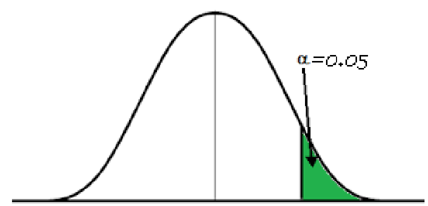
SIGNIFICANCE LEVEL

DEFINITION: Statistical significance is the likelihood that a relationship between variables is caused by something other than chance. The significance level is used to accept or reject the null hypothesis; this level is subjective in nature and is often set by the analyzer. Some business models require a 10% significance level of probability, whereas medical models require a significance level of 1%. In general, a cut off of 5% is often used. The 5% is called the **Significance Level,** also known as **alpha level (α)**.

A sample of data is statistically significant when the set is large enough (***<u>at least 30 data</u>***) to accurately represent the population being studied. The data set is deemed to be statistically significant if the probability of the phenomenon being random is less than 5%. When the test result exceeds the p-value, the null hypothesis is accepted. When the test result is less than the p-value, the null hypothesis is rejected. In other words, a high probability leads to acceptance of randomness and a low probability leads to behavior difference.

How do we decide between high and low probability? The answer is subjective in nature.

If the random chance probability is less than 5%, then we can conclude that there is a difference in the behavior of two different populations. (1 – α) is also known as **Confidence Level**; that is, we can say that we are 95% confident the difference is not driven by randomness, that it is within the boundaries of accepted values, although not with absolute certainty.



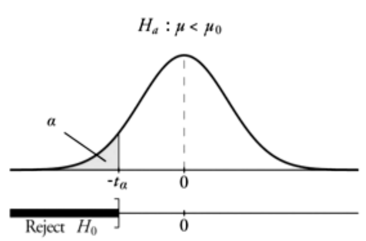
<u>DIRECTIONAL/ NON-DIRECTIONAL Significance Levels</u>

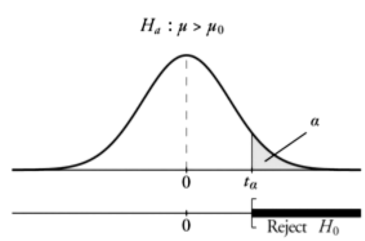
ONE-TAIL (DIRECTIONAL) test: the alternate hypothesis tests a sample mean that is greater than a test value or is less than a test value.

- LEFT-TAIL test – testing a value that is less than our null hypothesis
- RIGHT-TAIL test – testing a value that more than our null hypothesis

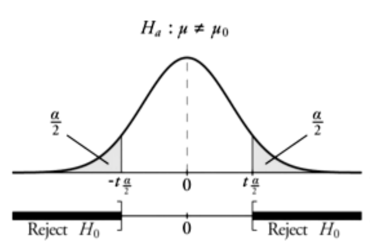
TWO-TAIL (NON-DIRECTIONAL) test: In some cases, we may not know whether the impact of a test (or sample) has a positive or negative impact. So, in these cases, the significance level is split in half.

In a one tail test, we reject the Null hypothesis if the sample mean is either in a left rejection region or in a right rejection region.





However, in the case of a two tail test, we can reject the Null hypothesis in a left or right rejection region. Two-tailed tests allot half of the alpha to testing the statistical significance in one direction and half of your alpha in the other direction.



STEPS to HYPOTHESIS TESTING:

① State the **Null (H_0) Hypothesis** & the **Alternative (H_1 or H_a) Hypothesis**:

$\boldsymbol{H_0}$: $\boldsymbol{\mu = recognized\ value}$

- claim made by a company about its product

$\boldsymbol{H_1}$: $\boldsymbol{\mu} < value$ or $\boldsymbol{\mu} > value$ or $\boldsymbol{\mu \neq value}$

- reason for testing the null hypothesis is that we think it is wrong
- state what we think is wrong about the null hypothesis
- testing for a significant difference between a sample and its population
- if test fails to prove a significant difference, then observed difference is due to random chance.

Assume null hypothesis is true, placing the burden on the researcher to conduct a study to show evidence that the null hypothesis is unlikely to be true.

② Choose the **SIGNIFICANCE LEVEL** $(\boldsymbol{\alpha})$

- How strong must the evidence be to **reject** H_0? *Probability threshold* chosen for deciding whether the observed results are rare enough to justify rejecting H_0
- significance level $(\boldsymbol{\alpha})$ = 10%, 5%, 1% or 0.5%.
- **SIGNIFICANCE LEVEL =** $\boldsymbol{1 - CONFIDENCE\ LEVEL}$
- If $\alpha = 0.05$, we are willing to be wrong 5% of the time = 95% confident to accept H_0.

③ Find the **CRITICAL VALUES**

- Determine z –scores related to significance (confidence) level
- Label a normal curve illustrating the zones of rejection/acceptance.

④ Find the **TEST STATISTIC**

- Compute the random chance probability
- Calculate $\boldsymbol{z - scores}$ or $\boldsymbol{p - values}$ (probabilities) – test the sample measure (value given in H_1) with the population standard (value given in H_0).

⑤ **MAKE DECISION**

- Compare $\boldsymbol{p - value}$ (or $\boldsymbol{z - score}$) with significance level – if p-value is less than significance level, reject null hypothesis; otherwise, accept it.
- Accept or reject the null hypothesis
- State a conclusion – it is possible that while making a decision to accept or reject the null hypothesis, we might go wrong because we are observing a sample and not an entire population.

EXAMPLE 1:

Blood glucose levels for obese patients have a mean of 100 with a standard deviation of 15. A researcher thinks that a diet high in raw cornstarch will have a positive effect on blood glucose levels. A sample of 36 patients who have tried the raw cornstarch diet have a mean glucose level of 106. Test the hypothesis that the raw cornstarch had an effect or not.

***Solution*:**

Follow the above discussed steps to test this hypothesis:

Step ①: State hypotheses. The population mean is 100.

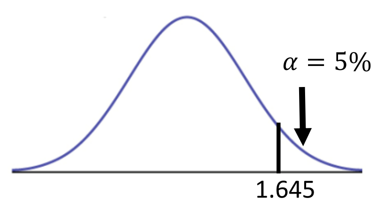
H_0: μ= 100

H_1: μ > 100

Step ②: State significance level. It is not given in the problem so let's assume it as 5% (0.05).

Step ③: Label a normal distribution curve illustrating the critical values for decision making.

Since H_1: μ > 100, we will perform a right-tailed test and using $\alpha = 0.05$, we shade the right side of a normal curve corresponding with 5%.



Step ④: Compute the random chance probability using z score and z-table.

For this set of data: $P(X > 106) = P\left(z > \frac{106-100}{15/\sqrt{36}}\right)$

$$= 1 - P(z > 2.40)$$

$$= 1 - 0.9918 = 0.0082$$

OR compare the z-score with the diagram. Notice that z=2.40 is in the rejection zone.

Step ⑤: Since the p-value is less than $\alpha = 0.05$ (or the z-score is located in the rejection zone, we reject H_0. In conclusion, raw cornstarch has an effect on the blood glucose levels.

EXAMPLE 2:

Templer and Tomeo (2002) reported that the population mean score on the quantitative portion of the Graduate Record Examination (GRE) General Test for students taking the exam between 1994 and 1997 was 558 ± 139 (μ ± σ). Suppose we select a sample of 100 participants (*n* = 100). We record a sample mean equal to 585 (*M* = 585). Compute the p-value to check whether or not we will retain the null hypothesis (μ = 558) at a significance level of 0.05.

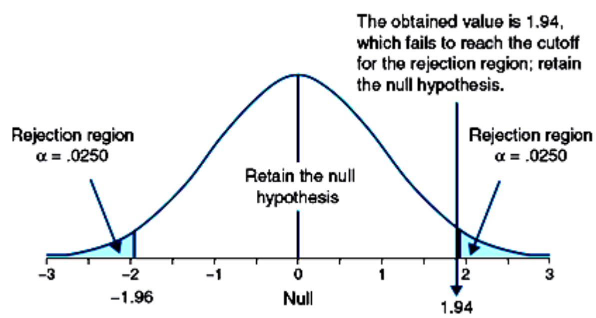
***Solution*:**

Step ①: State hypotheses. The population mean is 558.

H_0: μ= 558
H_1: μ ≠ 558 (two tail test)

Step ②: Significance level = 5% (0.05). In a non-directional two-tailed test, we divide the alpha value by 2, so that an equal proportion of area is placed in the upper and lower tails. So, the significance level on either side is calculated as: α/2 = 0.025. The z-score associated with this (1-0.025=0.975) is 1.96. As this is a two-tailed test, z-score(observed) which is less than -1.96 or greater than 1.96 is evidence to reject the Null hypothesis.

Step ③: Label a normal distribution curve with critical values.



Step ④: Compute the random chance probability OR z-score

For this set of data: $z = \frac{585-558}{139/\sqrt{100}} = 1.94$

$$P(z < 1.94) = 0.9738$$

$$P(z > 1.94) = 1 - 0.9738 = 0.0262$$

Step ⑤: Comparing the obtained z-value to the critical values (+/- 1.96), we reject the null hypothesis if value exceeds either of the critical values. Since Z_{obt}= 1.94 is less than the critical value, it does not fall in the rejection region. The decision is to retain the null hypothesis.
If we consider the p-value, since it is 2.62%, it is greater than the significance level of 2.5% for the right-tail. Based on this comparison, we retain the null hypothesis.

EXAMPLE 3:

The mayor of a town read an article that claimed the national unemployment rate is 8%.
She wondered if this held true in her own town, so she took a sample of 200 residents to test.
The sample included 22 residents who were unemployed. [significance level of 5%]

Solution:

Step ①: State hypotheses. This example will illustrate a hypothesis test using ***proportions***.

H_0: p= 0.08 where p is the proportion of residents that are unemployed
H_1: $p \neq 0.08$ (two tail test)

Step ②: Significance level = 5% (0.05). In this two-tailed test, we divide the alpha value by 2 again. So, the significance level on either side is: $\alpha/2 = 0.025$. The test will be compared with a z-score less than -1.96 or greater than 1.96.

Step ③: Label a normal distribution curve with critical values. [see example 2]

Step ④: Compute the random chance probability OR z-score

In this case, we use proportional sampling (which uses calculations similar to binomial distribution sampling), as follows:

NOTE: $\boldsymbol{np} > 5$ and $\boldsymbol{nq} > 5$

$\boldsymbol{\mu} = \boldsymbol{np} = 200(0.08) = 16$ In our sample of 200 residents, it follows that according to the national unemployment rate, 16 residents would be unemployed. We will compare the 22 residents from the sample in this step.

$$\boldsymbol{\sigma} = \sqrt{\boldsymbol{npq}} = \sqrt{16(0.92)} = 3.836665$$

In terms of the proportion sampling, we will represent the population with the above calculations of μ and σ.

For this set of data: $z = \frac{22-16}{3.836665} = 1.564$

Step ⑤: Comparing the z-value to the critical values (+/- 1.96), since Z_{obt}= 1.564 does not fall in the rejection region, we retain the null hypothesis. The mayor of the town does not have sufficient evidence to reject the null hypothesis. The town's unemployment rate likely falls within the national average of 8%.

<u>EXERCISE 1</u>: For each scenario, state the null and alternative hypotheses and the type of test.

A) We have a medicine that is being manufactured and each pill is supposed to have 14 milligrams of the active ingredient.

B) The school principal wants to test if it is true what teachers say – that high school juniors use the computer an average 3.2 hours a day.

C) A package of gum claims that the flavor lasts more than 39 minutes.

D) The average miniature horse is less than 34 inches at the shoulder. Tiny Equines claims that their miniature horses are an average of 3 inches less than the breed average.

E) Speedy Solutions is a delivery company that claims all deliveries average less than 72 h.

F) A company claims that 70% or more of its cars on the lot average more than 40 mpg.

G) The 2011 average service time for Fast Fatz Burgers was 56 seconds. The store manager claims that the 2012 average is more than 5 seconds faster.

H) Aaron works for an electronics company, quality-testing batteries. The company claims a minimum of 5 hrs battery life.

I) The outside temperature in Maui averages 33° year-round.

N) At least 72% of elementary school teachers are female.

EXERCISE 2: For each scenario, perform a hypothesis testing.

1. The average IQ for the adult population is 100 with a standard deviation of 15. A researcher believes this value has changed and decides to test the IQ of 75 random adults. The average IQ of the sample is 105. Has the average IQ changed?

2. The NFL believes the chance of the New England Patriots winning the coin toss is greater than 50%. In a random sample of 200 coin tosses, the Patriots won 118 times. Is there enough evidence to suggest the Patriots are cheaters? ($\alpha = 1\%$)

3. A quality control engineer finds that a sample of 100 light bulbs had an average life-time of 470 hours. Assuming a population standard deviation of 25 hours, test whether the population mean is 480 hours at a significance level of $\alpha = 0.05$.

4. A farmer is trying out a planting technique that he hopes will increase the yield on his pea plants. The average number of pods on one of his pea plants is 145 pods with a standard deviation of 100 pods. This year, after trying his new planting technique, he takes a random sample of 144 of his plants and finds the average number of pods to be 147. He wonders whether or not this is a statistically significant increase.

5. Duracell manufactures batteries that the CEO claims will last an average of 300 hours under normal use. A researcher randomly selected 30 batteries from the production line and tested these batteries. The tested batteries had a mean life span of 270 hours with a standard deviation of 60 hours. Do we have enough evidence to suggest that the claim of an average lifetime of 300 hours is false?

6. It has been reported that the average credit card debt for college seniors is $3262. The student senate at a large university feels that their seniors have a debt much less than this, so it conducts a study of 50 randomly selected seniors and finds that the average debt is $2995, and the population standard deviation is $1100. Use $\alpha=0.05$.

7. An educator estimates that the dropout rate for seniors at high schools in Toronto is 12%. Last year in a random sample of 300 Toronto senior students, 27 withdrew from school. At $\alpha = 0.05$, is there enough evidence to reject the educator's claim?