Inference

Inference is the process of drawing conclusions or making assumptions based on available evidence and reasoning. It's like reading between the lines to understand something that isn't explicitly stated.

Here's a breakdown:

- **Evidence:** This is the information you have, which can be facts, observations, or prior knowledge.
- **Reasoning:** This is the process of using logic and your understanding of the world to connect the evidence and reach a conclusion.
- **Conclusion:** This is the inferred idea or assumption you arrive at based on the evidence and reasoning.

Example:

If you see someone walking down the street with an umbrella on a sunny day, you might infer that they are expecting rain.

- **Evidence:** Person with an umbrella on a sunny day.
- **Reasoning:** People typically use umbrellas to protect themselves from rain.
- **Conclusion:** The person is expecting rain.

Key points about inference:

- **It's not always guaranteed:** Inferences are based on probabilities and assumptions, so they may not always be accurate.
- **It's a crucial skill:** Inference is essential for critical thinking, problem-solving, and understanding complex information.
- **It's used in many fields:** Inference is used in various fields, including science, law, literature, and everyday life.

Null and alternative hypothesis

A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allows us to make probabilistic statements about population parameters.

1. **Null hypothesis (H0):** ← starting point

   In simple terms, the null hypothesis is a statement that assumes there is no significant effect or relationship between the variables being studied. It serves as the starting point for hypothesis testing and represents the status quo or the assumption of no effect until proven otherwise. The purpose of hypothesis testing is to gather evidence (data) to either reject or fail to reject the null hypothesis in favour of the alternative hypothesis, which claims there is a significant effect or relationship.

2. **Alternative hypothesis (H1 or Ha):**

   The alternative hypothesis, is a statement that contradicts the null hypothesis and claims there is a significant effect or relationship between the variables being studied. It represents the research hypothesis or the claim that the researcher wants to support through statistical analysis.

both are mutually exclussive.

Imp points

- How to decide what null & alternative?
  [Typically the null hypothesis says nothing new is happening]
- We try to gather evidence to reject the null hypothesis.
- failing to reject null doesnt mean null is true
- If $H_0 : var = 1$, $H_a : var < 1$ } univariate distribution

here α will be same as taken

& if $H_0 : var = 1$, $H_a : var \neq 1$ } bivariate distribution

here α will be divided in 2.

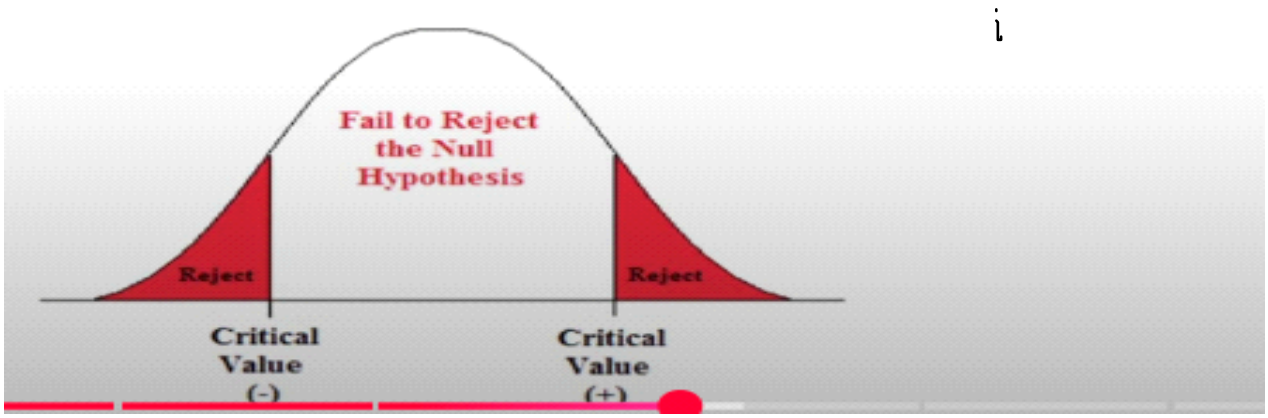
α - significance level is decided by domain experts

Significance level

It is denoted by α.
(alpha)

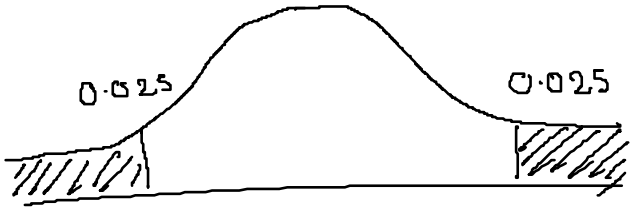
Significance level - denoted as α (alpha), is a predetermined threshold used in hypothesis testing to determine whether the null hypothesis should be rejected or not. It represents the probability of rejecting the null hypothesis when it is actually true, also known as Type 1 error.

The critical region is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level.



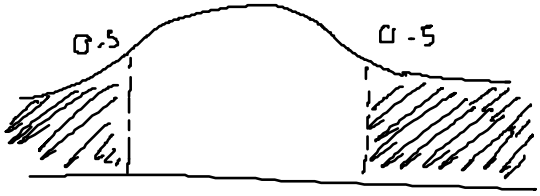
i

(1)



→ When $\alpha = 5\% = 0.05$

(2)



→ When $\alpha = 10\% = 0.1$

as we increase α the rejection region increases and it also doesn't give the strength of evidence.

Rejection region exp1

Rejection Region Approach

1. Formulate a Null and Alternate hypothesis
2. Select a significance level(This is the probability of rejecting the null hypothesis when it is actually true, usually set at 0.05 or 0.01)
3. Check assumptions (example distribution)
4. Decide which test is appropriate(Z-test, T-test, Chi-square test, ANOVA)
5. State the relevant test statistic
6. Conduct the test
7. Reject or not reject the Null Hypothesis.
8. Interpret the result

eg :

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day with a known population standard deviation of 5 units. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day. The company wants to know if the new training program has significantly increased productivity.

Ans: Population mean $\mu = 50$

pop Standard deviation $\sigma = 5$

n = Sample = 30 , Sample mean $\bar{x} = 53$

1) $H_0 : \mu = 50$ $H_a : \mu > 50$

2) Significance level $\alpha = 0.05 \rightarrow 5\%$

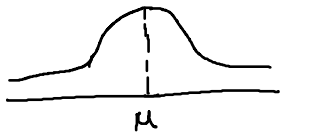
Assumption 3) Normality valid / pop std (σ) known

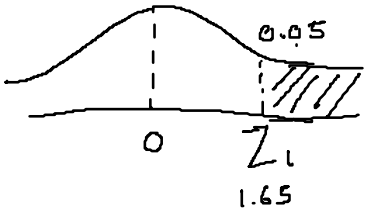
4) Which test - pop std known then z-test

5) z -value

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{53 - 50}{5/\sqrt{30}} = 3.28$$

normality valid means normal distribution is assumed.





$Z_l < Z$

$1.65 < 3.28$

We reject the null hypothesis

Rejection region exp2

eg : 2

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a known population standard deviation of 4 grams.

Ans: given: population mean $\mu = 50$ gms

Sample $(n) = 40$, $\bar{X} = 49$ gms, $\sigma = 4$

1) $H_0 : \mu = 50$ (exact 50 grams)

$H_a : \mu \neq 50$ (not exact 50 grams)

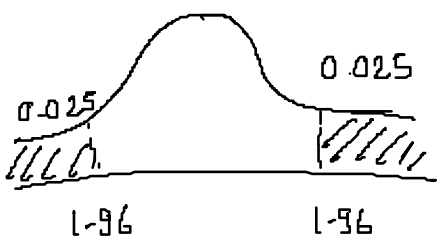
2) $\alpha = 0.05$

3) Normality valid, σ given $\rightarrow$ Z-test

4) Z-test selected

5) Z

6) $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{49 - 50}{4/\sqrt{40}} = -1.58$$



here, -1.58 lies within the acceptance region.

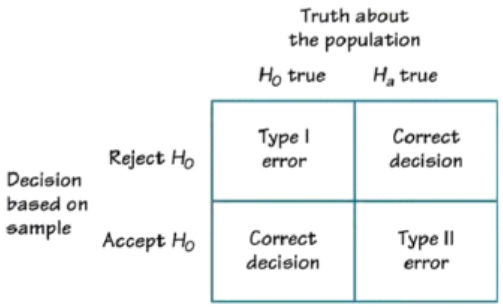
Can't reject null hypothesis. so, $\mu = 50$ gms

Type 1 and type 2 error

In hypothesis testing, there are two types of errors that can occur when making a decision about the null hypothesis: Type I error and Type II error.

Type-I (False Positive) error occurs when the sample results, lead to the rejection of the null hypothesis when it is in fact true.

In other words, it's the mistake of finding a significant effect or relationship when there is none. The probability of committing a Type I error is denoted by α (alpha), which is also known as the significance level. By choosing a significance level, researchers can control the risk of making a Type I error.



Type-II (False Negative) error occurs when based on the sample results, the null hypothesis is not rejected when it is in fact false.

This means that the researcher fails to detect a significant effect or relationship when one actually exists. The probability of committing a Type II error is denoted by β (beta).

Trade-off between Type 1 and Type 2 errors

Type 1 error :- when we reject the null because of α even if it true
- if we increase α-value the chances of rejecting increases

Type 2 error:- when we reject alternate hypo even if it's true.

Probability of type 1 error is denoted by α.

Probability of type 2 error is denoted by β.

One sided vs two sided test

One sided vs two sided test

04 April 2023 13:29

One-sided (one-tailed) test: A one-sided test is used when the researcher is interested in testing the effect in a specific direction (either greater than or less than the value specified in the null hypothesis). The alternative hypothesis in a one-sided test contains an inequality (either ">" or "<").

Example: A researcher wants to test whether a new medication increases the average recovery rate compared to the existing medication.

Two-sided (two-tailed) test: A two-sided test is used when the researcher is interested in testing the effect in both directions (i.e., whether the value specified in the null hypothesis is different, either greater or lesser). The alternative hypothesis in a two-sided test contains a "not equal to" sign (≠).

Example: A researcher wants to test whether a new medication has a different average recovery rate compared to the existing medication.

The main difference between them lies in the directionality of the alternative hypothesis and how the significance level is distributed in the critical regions.

Advantages and Disadvantages?

Two-tailed test (two-sided):

Advantages:

1. **Detects effects in both directions**: Two-tailed tests can detect effects in both directions, which makes them suitable for situations where the direction of the effect is uncertain or when researchers want to test for any difference between the groups or variables.
2. **More conservative**: Two-tailed tests are more conservative because the significance level (α) is split between both tails of the distribution. This reduces the risk of Type I errors in cases where the direction of the effect is uncertain.

Disadvantages:

1. **Less powerful**: Two-tailed tests are generally less powerful than one-tailed tests because the significance level (α) is divided between both tails of the distribution. This means the test requires a larger effect size to reject the null hypothesis, which could lead to a higher risk of Type II errors (failing to reject the null hypothesis when it is false).
2. **Not appropriate for directional hypotheses**: Two-tailed tests are not ideal for cases where the research question or hypothesis is directional, as they test for differences in both directions, which may not be of interest or relevance.

One sided vs two sided test

One-tailed test (one-sided):

Advantages:

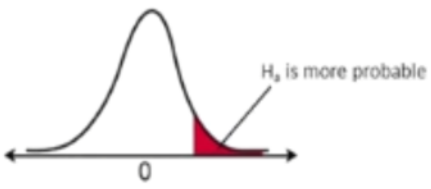
1. **More powerful**: One-tailed tests are generally more powerful than two-tailed tests, as the entire significance level (α) is allocated to one tail of the distribution. This means that the test is more likely to detect an effect in the specified direction, assuming the effect exists.

2. **Directional hypothesis**: One-tailed tests are appropriate when there is a strong theoretical or practical reason to test for an effect in a specific direction.

Disadvantages:

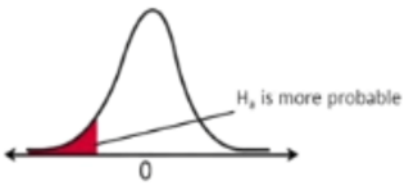
1. **Missed effects**: One-tailed tests can miss effects in the opposite direction of the specified alternative hypothesis. If an effect exists in the opposite direction, the test will not be able to detect it, which could lead to incorrect conclusions.

2. **Increased risk of Type I error**: One-tailed tests can be more prone to Type I errors if the effect is actually in the opposite direction than the one specified in the alternative hypothesis.



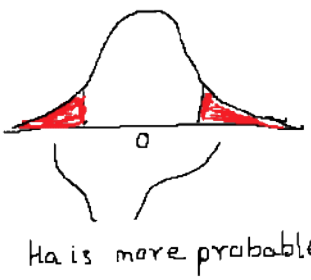
Right-tail test

$H_a: \mu > \text{value}$



Left-tail test

$H_a: \mu < \text{value}$



two-sided test

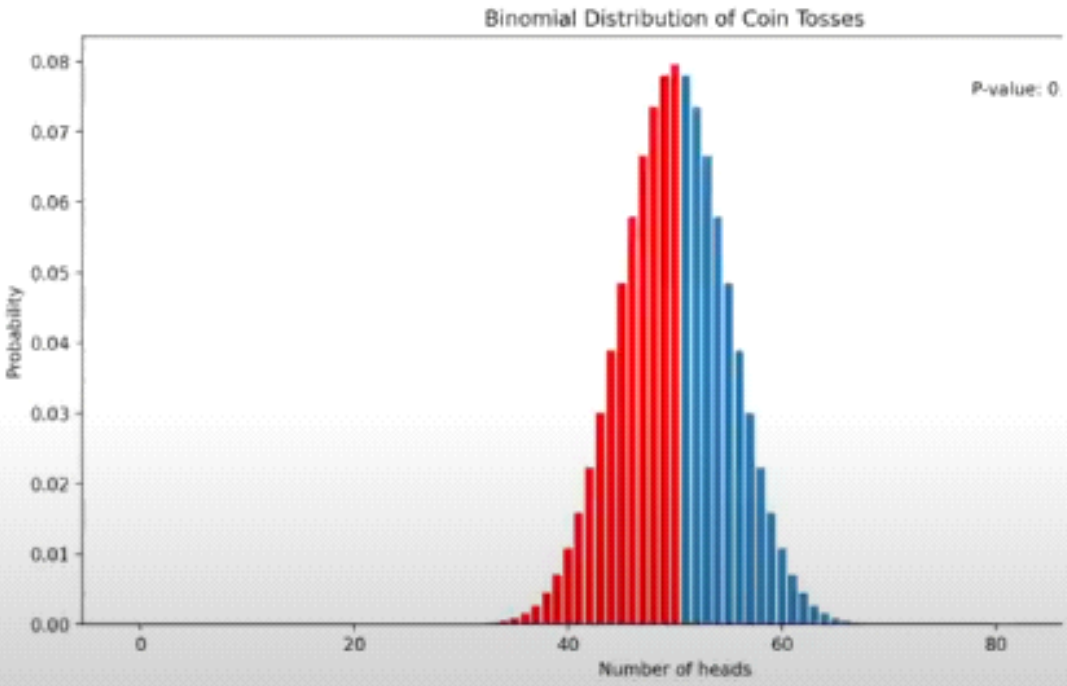
Ha: M ≠ value

P-value

P-value

06 April 2023 06:48

P-value is the probability of getting a sample as or more extreme(having more evidence against H0) than our own sample given the Null Hypothesis(H0) is true.



In simple words p-value is a measure of the strength of the evidence against the Null Hypothesis that is provided by our sample data.

Example: we are doing a experiment where we toss 100 coin (H,T). A predict how many times Head occurs out of 100 time. (100 coin 100 time toss)

→ $H_0 : p(H) = p(T)$

$H_a : p(H) \neq p(T)$

One sided → Senareo 1: I got 53 heads. The probability is 0.066.

p value = 0.3086, it says out of 100 times of toss 30 times i can get 53 coins.

Scenario 2: I got 60 heads. The probability of getting 60 H is 0.0108

p value = 0.0284, it says out of 100 time of toss 3 times. i can get 60 coins

P value – the power of alternative hypothesis

– maller P-value is better for alternative hypo

T- test

T-tests

06 April 2023 14:14

A t-test is a statistical test used in hypothesis testing to compare the means of two samples or to compare a sample mean to a known population mean. The t-test is based on the t-distribution, which is used when the population standard deviation is unknown and the sample size is small.

There are three main types of t-tests:

One-sample t-test: The one-sample t-test is used to compare the mean of a single sample to a known population mean. The null hypothesis states that there is no significant difference between the sample mean and the population mean, while the alternative hypothesis states that there is a significant difference.

Independent two-sample t-test: The independent two-sample t-test is used to compare the means of two independent samples. The null hypothesis states that there is no significant difference between the means of the two samples, while the alternative hypothesis states that there is a significant difference.

Paired t-test (dependent two-sample t-test): The paired t-test is used to compare the means of two samples that are dependent or paired, such as pre-test and post-test scores for the same group of subjects or measurements taken on the same subjects under two different conditions. The null hypothesis states that there is no significant difference between the means of the paired differences, while the alternative hypothesis states that there is a significant difference.

For calculating one sample T-test

① $df = n - 1$

② $t\,stat = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

* Note

formula changes as per type of t-test.

③ T table link:
t table link

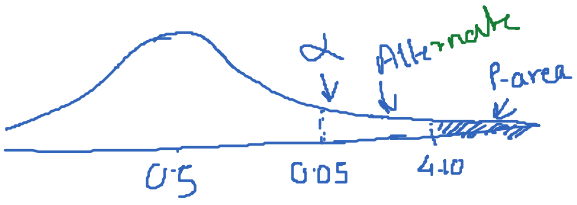
Z-test p-value exp1

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day and the sample std is 4. The company wants to know if the new training program has significantly increased productivity.

Given : $\mu = 50, n = 30, \bar{x} = 53, \sigma = 4, \alpha = 0.05$

$H_0 : \mu = 50$

$H_a : \mu > 50$



$$\text{Z-test} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{53 - 50}{4/\sqrt{30}} = \frac{3}{4}\sqrt{30} = 4.10$$

$$\text{P-value} = 1 - Z(4.10) = 0.001$$

here, P value $< \alpha$

$0.001 < 0.05$

here we reject null hypothesis. $\mu > 50$

Z-test p-value exp2

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a standard deviation of 5 grams.

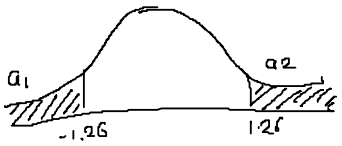
→ given: $\mu = 50$, $n = 40$, $\bar{x} = 49$, $\sigma = 5$

$$Z = \frac{\bar{x} - \mu}{\sigma\sqrt{n}} = \frac{49 - 50}{5/\sqrt{40}} = \frac{-1}{\frac{5}{\sqrt{40}}} = \frac{-\sqrt{40}}{5} = -1.265$$

$H_0 : \mu = 50$

$H_a : \mu \neq 50$

For 2-Tailed distribution



$$P\text{ value} = a_1 + a_2$$
$$= 0.103 + 0.103$$
$$= 0.206$$

here, $P\text{ value} = 0.206$, $\alpha = 0.05$ ⟹ $P > \alpha$

Here P is greater so we failed to reject null hypothesis.

Z-Test vs T-Test

20 January 2025 13:13

The **Z-test** and **T-test** are both statistical methods used to determine whether there is a significant difference between the means of two groups or a population mean and a sample mean. However, they differ in terms of assumptions and applications. Here's a breakdown:

1. Definition

- **Z-test**: Used when the population variance (σ2\sigma^2) is known or the sample size is large (n>30n > 30).
- **T-test**: Used when the population variance is unknown, and the sample size is small (n≤30n \leq 30).

2. Key Differences

Aspect	Z-Test	T-Test
Use Case	Large sample size (n>30n > 30) or known population variance.	Small sample size (n≤30n \leq 30) or unknown population variance.
Distribution Assumption	Assumes data follows a normal distribution (especially for small samples).	Assumes data follows a normal distribution or approximately normal for small samples.
Population Variance (σ2 \sigma^2)	Known or large enough sample to approximate it.	Unknown (uses sample standard deviation instead).
Test Statistic	$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$	$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$
Critical Values	Derived from the **Z-distribution**.	Derived from the **T-distribution** (adjusted for degrees of freedom).
Accuracy	More accurate with large samples.	Better suited for small samples due to the flexibility of the T-distribution.

3. When to Use

Scenario	Test to Use
Sample size is **large** (n>30n > 30) and population variance is **known**.	Z-test
Sample size is **small** (n≤30n \leq 30) and population variance is **unknown**.	T-test
Comparing two sample means when variances are unknown and sample sizes are small.	T-test
Testing proportions for large samples.	Z-test

4. Examples

- **Z-Test Example**: A company claims that the average weight of a product is 500 grams. You take a sample of 50 products (n>30n > 30), and the population standard deviation is known to be 10 grams. Use the Z-test to check the claim.
- **T-Test Example**: A researcher measures the height of 15 students (n<30n < 30) and compares the sample mean to a population mean, but the population standard deviation is unknown. Use the T-test.

5. Conclusion

- **Z-test**: Best for large samples and when the population variance is known.
- **T-test**: Best for small samples or when the population variance is unknown.

Single sample T-test exp1

Single Sample t-test

06 April 2023 14:14

A one-sample t-test checks whether a sample mean differs from the population mean.

Assumptions for a single sample t-test

1. **Normality** - Population from which the sample is drawn is normally distributed
2. **Independence** - The observations in the sample must be independent, which means that the value of one observation should not influence the value of another observation.
3. **Random Sampling** - The sample must be a random and representative subset of the population.
4. **Unknown population std** - The population std is not known.

Suppose a manufacturer claims that the average weight of their new chocolate bars is 50 grams, we highly doubt that and want to check this so we drew out a sample of 25 chocolate bars and measured their weight, the sample mean came out to be 49.7 grams and the sample std deviation was 1.2 grams. Consider the significance level to be 0.05

$\rightarrow$ $N=50$, $n=25$, $\bar{x}=49.7$, $S=1.2$, $\alpha=0.05$

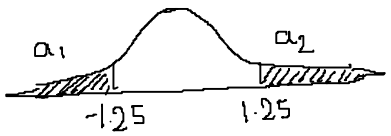
$H_0 : N = 50$

$H_a : N \neq 50$

Assume normality

$$t = \frac{\bar{x}-N}{S/\sqrt{n}} = \frac{49.7-50}{1.2\sqrt{25}} = -\frac{0.3\times5}{1.2} = \frac{-1.5}{1.2} = -1.25$$

$\neq$ not equal



We can find P-value, with t-table and Stats lib

t-cds value $(a_1) = 0.11$

P value $= a_1 + a_2 = 0.22$

P value $= 0.22$ } $P > \alpha$

$\alpha = 0.05$

We fail to reject null hypothesis, chocolate's average weight is 50 grams.

Chi squared test

Uses of Chi-squared test

(1) Goodness of fit test:

- finding uniform distrubution
  eg: Determine die is fair or not.

- finding specific distribution
  eg: Web site visits follows poisson distribution.

(2) Test of Independence:

- Analyzing relationships betn categorical variables
  eg:- Actor/watch relationship
  - Gender/transport preference relationship.
  - In customer churn, outcome related to age, location?

(3) Feature selection in machine learning.

- Relationship betn categorical column/feature and target variable in classification problems.

☐ Assumptions in Chi-squared test.

1. Both variables must be categorical values
2. Independence of observations
   ↳ each observation in data should be independent of the other
   (No double testamony)
3. Sufficient sample size: Minimum frequency in each cell should be 5.

☐ Interpretation:

Chi-squared test indicates whether there is a statistically significant association betn variables. but it doesn't necessary imply causation.

Chi squared test

20 January 2025 13:10

The **chi-square test** is specifically designed to test the relationship between **two categorical variables**, not numerical variables. Here's why:

1. **Chi-Square Test Explanation**:
   - The chi-square test assesses whether there is a significant association between the categories of one variable and the categories of another.
   - It uses a **contingency table** that shows the frequency distribution of categorical variables.
2. **Numerical Variables**:
   - Chi-square cannot directly handle numerical variables because it requires categorical input.
   - To apply the chi-square test to numerical data, you must first convert the numerical variables into categories (e.g., using binning or discretization).
3. **When to Use Chi-Square**:
   - **Two Categorical Variables**: Directly applicable.
   - **Numerical and Categorical Variables**: Convert numerical data into categories (e.g., using equal-width or quantile binning) before applying chi-square.

Example Scenarios:

Valid Chi-Square Use Case:

- **Dataset**: A survey dataset with:
  - Variable 1: Gender (Male/Female)
  - Variable 2: Preference (Likes product/Dislikes product)
- Use chi-square to test if product preference depends on gender.

Invalid Use Case (Numerical Data):

- **Dataset**: A dataset with:
  - Variable 1: Age (e.g., 20, 25, 30)
  - Variable 2: Income (e.g., 30000, 40000, 50000)
- Chi-square cannot be applied directly because both variables are numerical. You would need to bin these variables into categories like "Young", "Middle-aged", "Old" for age and "Low", "Medium", "High" for income.

Summary:

- Use chi-square for **two categorical variables**.
- Convert numerical variables into categorical form if you want to use chi-square.

Chi square test exp1

Link = [chi square test in python]

Watch/Actor	y	n	®total
m	140	44	184
f	178	38	216
©total	318	82	400

Ho: No Connection betⁿ watch/actor.

Ha: full Connection betⁿ watch/actor.

Actor/Watch	O	Execpted (E)	$(O-E)^2$	$(O-E)^2/E$
M/Y	140	$\frac{184 \times 318}{400} = 146$	36	0.246
M/N	44	$\frac{184 \times 82}{400} = 38$	36	0.947
F/Y	178	$\frac{216 \times 318}{400} = 172$	36	0.209
F/N	38	$\frac{216 \times 82}{400} = 44$	36	0.818
				2.220

$\bar{x}^2 = 2.22$

$df = (\text{rows} - 1)(\text{column} - 1) = 1$

$\alpha = 0.05$

$\bar{X}^2$ stat for 0.05 $= 3.84$

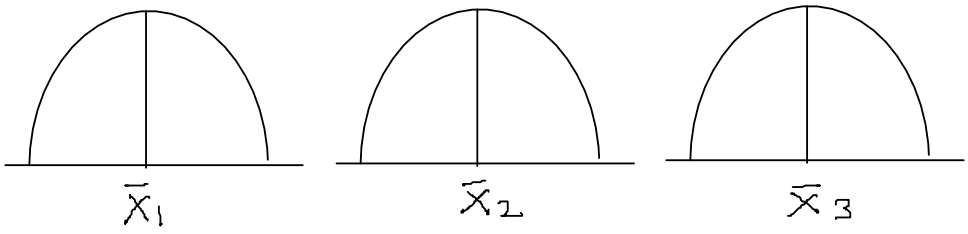
$3.84 > 2.22$

∴ We fail to reject null hypothesis

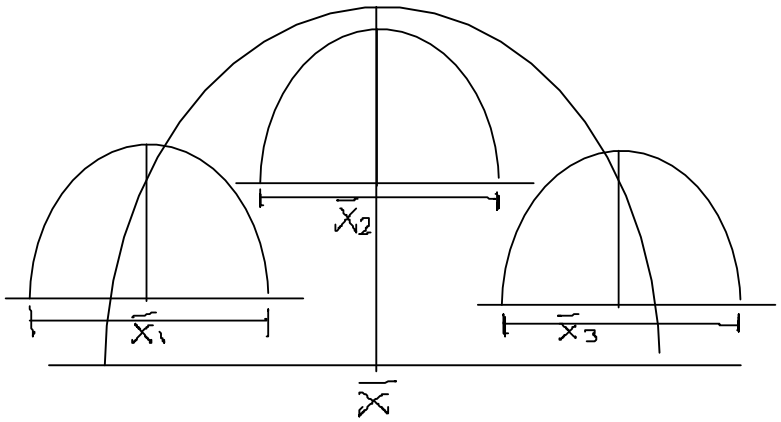
So, No connection between watch/actor.

Anova test

- Anova is analysis of variance. (σ^2)
- It used to comparison of more than two population or population having more than two subgroups. we will use Anova technique.



Do all these 3 means are comming from same population?



$$\text{Anova} = \frac{\text{Variability between the means}}{\text{Variability within the means}}$$

Total variance = Variability between the means + Variability within the distribution

Anova test assumptions

Assumptions:

1. Each population is having normal distribution.
2. The population from which the sample are drawn have the equal variance. ie. $S_1^2 = S_2^2 = S_3^2 \ldots S_n^2$ for n samples.
3. Each sample is drawn randomly & they are independent.

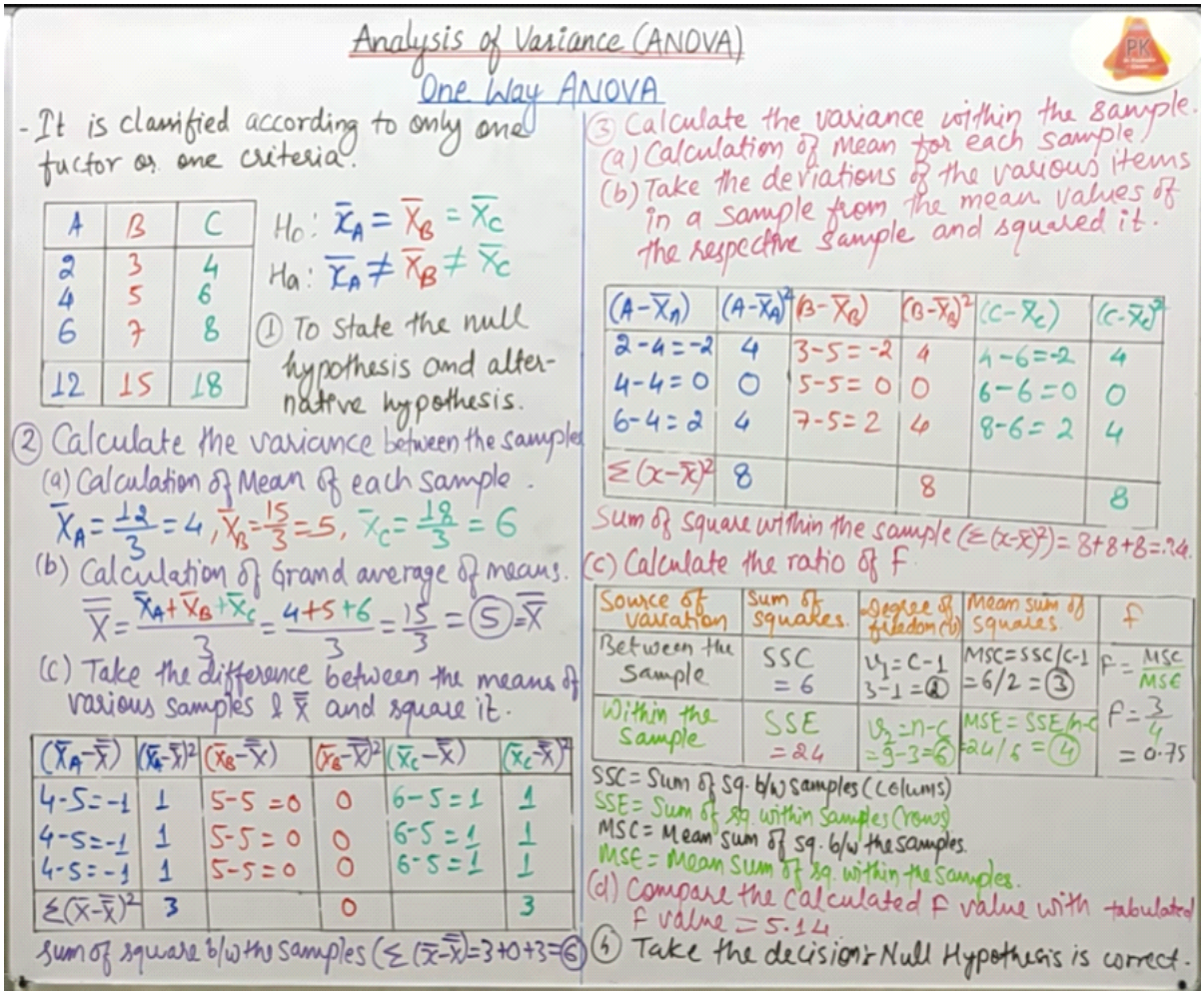
Hypothesis:

$H_0: \mu_1 = \mu_2 = \mu_3 \ldots = \mu_n$

$H_a: \mu_1 \neq \mu_2 \neq \mu_3 \ldots \neq \mu_n$

Classification
- One factor — one way anova.
- two factor — two way anova.

Anova test one way exp 1

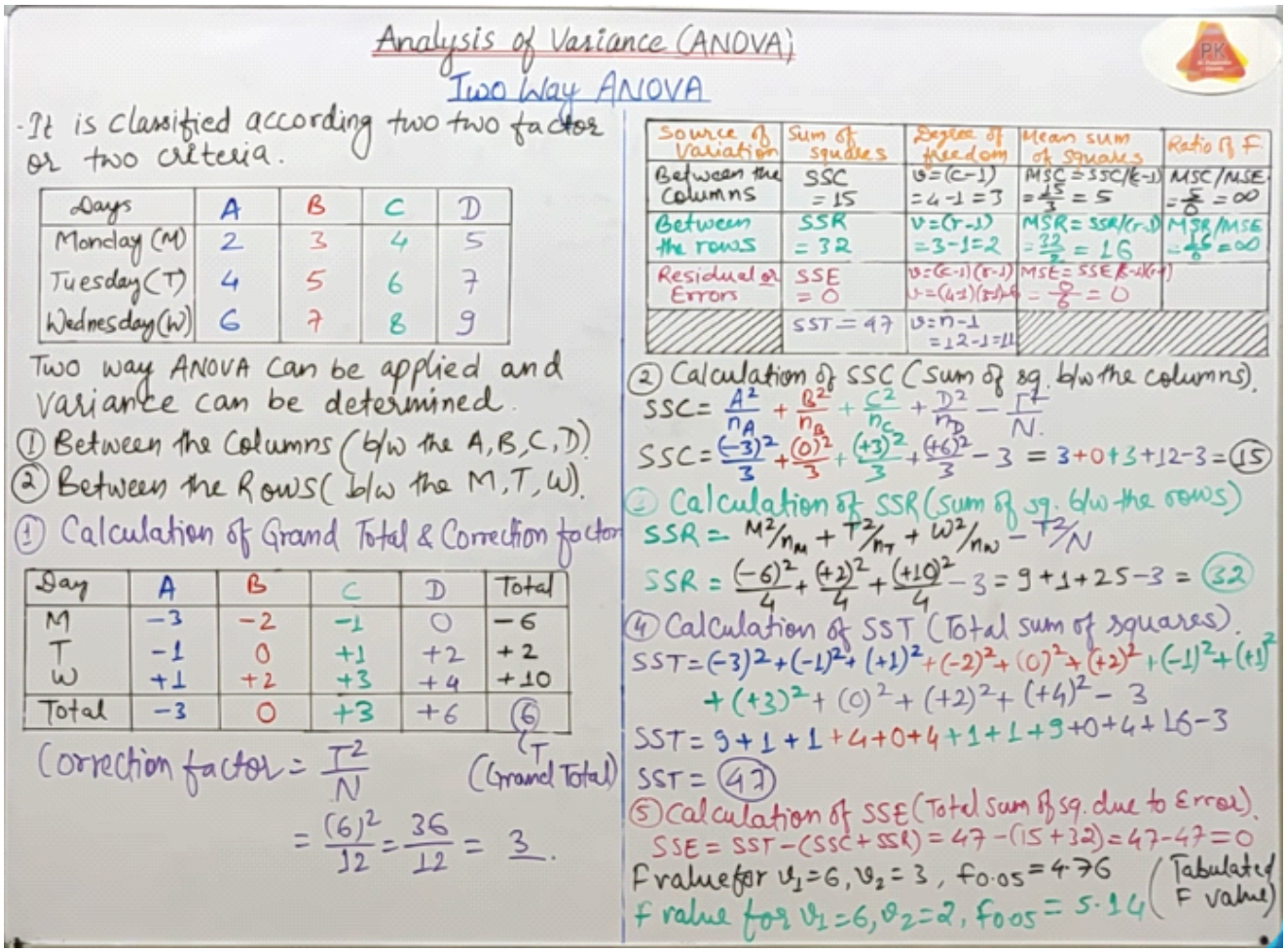



```
data = {
    'A':[2,4,6],
    'B':[3,5,7],
    'C':[4,6,8]
}

# Extract the groups
group_A = data['A']
group_B = data['B']
group_C = data['C']


# Perform the ANOVA test
f_statistic, p_value = f_oneway(group_A, group_B, group_C)

print("F-statistic:", f_statistic)
print("P-value:", p_value)

# Interpret the results
if p_value < 0.05:
    print("There is a significant difference between the groups.")
else:
    print("No significant difference between the groups.")
```

```
F-statistic: 0.75
P-value: 0.5120000000000001
No significant difference between the groups.
```

Anova test two way exp 1

09 January 2025 16:21



Analysis of Variance (ANOVA)

Two Way ANOVA

- It is classified according two two factor or two criteria.

Days	A	B	C	D
Monday (M)	2	3	4	5
Tuesday (T)	4	5	6	7
Wednesday (W)	6	7	8	9

Two way ANOVA can be applied and Variance can be determined.

① Between the Columns (b/w the A, B, C, D)

② Between the Rows (b/w the M, T, W).

① Calculation of Grand Total & Correction factor

Day	A	B	C	D	Total
M	−3	−2	−1	0	−6
T	−1	0	+1	+2	+2
W	+1	+2	+3	+4	+10
Total	−3	0	+3	+6	6 (T)

Correction factor $= \frac{T^2}{N}$ (Grand Total)

$= \frac{(6)^2}{12} = \frac{36}{12} = 3.$

Source of Variation	Sum of squares	Degree of freedom	Mean sum of Squares	Ratio of F
Between the Columns	SSC = 15	$v = (c-1)$ $= 4-1 = 3$	$MSC = SSC/(c-1)$ $= \frac{15}{3} = 5$	MSC/MSE $= \frac{5}{0} = \infty$
Between the rows	SSR = 32	$v = (r-1)$ $= 3-1 = 2$	$MSR = SSR/(r-1)$ $= \frac{32}{2} = 16$	MSR/MSE $= \frac{16}{0} = \infty$
Residual or Errors	SSE = 0	$v = (c-1)(r-1)$ $= (4-1)(3-1) = 6$	$MSE = SSE/(c-1)(r-1)$ $= \frac{0}{6} = 0$	
	SST = 47	$v = n-1$ $= 12-1 = 11$		

② Calculation of SSC (Sum of sq. b/w the columns).

$SSC = \frac{A^2}{n_A} + \frac{B^2}{n_B} + \frac{C^2}{n_C} + \frac{D^2}{n_D} - \frac{T^2}{N}$

$SSC = \frac{(-3)^2}{3} + \frac{(0)^2}{3} + \frac{(+3)^2}{3} + \frac{(+6)^2}{3} - 3 = 3+0+3+12-3 = 15$

③ Calculation of SSR (Sum of sq. b/w the rows)

$SSR = M^2/n_M + T^2/n_T + W^2/n_W - T^2/N$

$SSR = \frac{(-6)^2}{4} + \frac{(+2)^2}{4} + \frac{(+10)^2}{4} - 3 = 9+1+25-3 = 32$

④ Calculation of SST (Total sum of squares).

$SST = (-3)^2 + (-1)^2 + (+1)^2 + (-2)^2 + (0)^2 + (+2)^2 + (-1)^2 + (+1)^2 + (+3)^2 + (0)^2 + (+2)^2 + (+4)^2 - 3$

$SST = 9+1+1+4+0+4+1+1+9+0+4+16-3$

$SST = 47$

⑤ Calculation of SSE (Total sum of sq. due to Error).

$SSE = SST - (SSC + SSR) = 47 - (15+32) = 47-47 = 0$

F value for $v_1 = 6, v_2 = 3$, $f_{0.05} = 4.76$

f value for $v_1 = 6, v_2 = 2$, $f_{0.05} = 5.14$ (Tabulated F value)