# TABLE OF CONTENTS

[**TABLE OF CONTENTS 1**](#_9xjohjrjdxr8)

[**Tables and Figures 2**](#_mzti8dqkf1xu)

[- Frequency Table 2](#_p232cyjt6lny)

[- Stem and Leaf 2](#_sijs6zfliojc)

[**Formulas 4**](#_uwqr7tedhk78)

[1. Box and Plot 4](#_3ch9wvyx5rf5)

[2. Trimean 4](#_i0jfzssr2dca)

[3. Geometric Mean 5](#_3kw8a78hueqg)

[4. Trimmed Mean 5](#_i2fewcqn6tsu)

[5. Pearson’s correlation 5](#_hu56nlab00f9)

[6. Probability of a Single Event 5](#_sr3dnn8575y0)

[7. Probability of 2 or more independent events 5](#_j0oo0ng3wemg)

[8. Probability of 2 or more dependent events 5](#_24s3airs88ds)

[9. Permutations 6](#_fh0csggibd2p)

[10. Combinations 6](#_s594ajw5m8x)

[11. Binomial Distribution 6](#_yd5lq0eimk7k)

[12. Poisson Distribution 6](#_en82m0g97jmy)

[13. Multinomial Distribution 7](#_po7m8zkqknb9)

[14. Hypergeometric Distribution 7](#_rc916yy1vp0x)

[15. Variance and Standard Deviation 7](#_jb12u48a154g)

[16. t-test for one sample 8](#_sdgcy6ixxk8o)

[17. t-test for two independent samples 8](#_vocgzq87swc)

[18. t-test for two related samples 9](#_e6bdw74u4iey)

[**Step-by-Step Tutorials 10**](#_d8fthhn7k7jk)

[- One Way ANOVA - Independent Measures 10](#_ja2bk2st9ej0)

[- Two Way ANOVA 13](#_hqeitrfj7k86)

[**Links 17**](#_pby59en2lj6e)

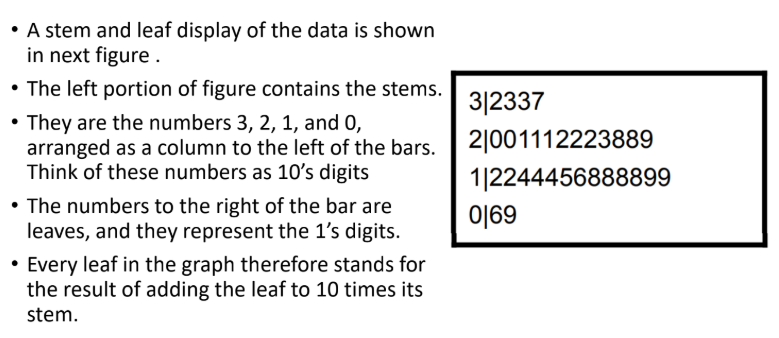
[**Exercises 18**](#_5zr2dl7ggvod)

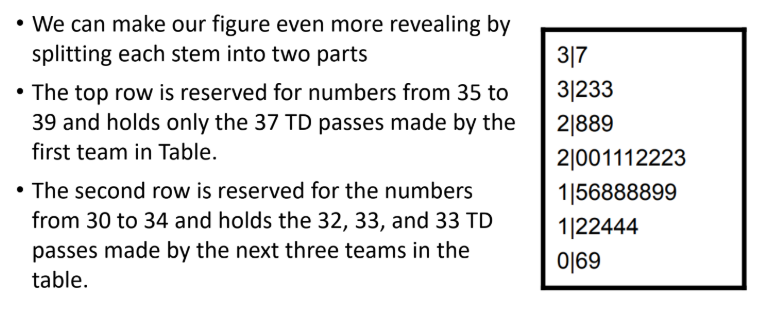
# Tables and Figures

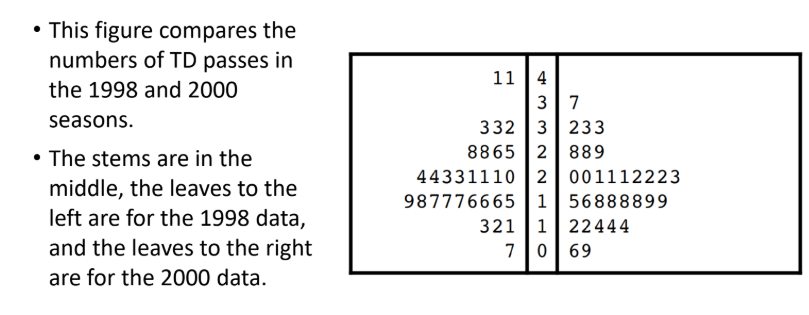
## Frequency Table

| **Previous Ownership** | **Frequency** | **Relative Frequency** |
| --- | --- | --- |
| None | 85 |  |
| Windows | 60 |  |
| Macintosh | 355 |  |
| Total | 500 |  |

## Stem and Leaf

****

****

****

## 

# Formulas

## Box and Plot

| **Name** | **Formula** |
| --- | --- |
| 25th Quartile / Lower Hinge | th term |
| 50th Quartile / Median | When n is odd:  th term  When n is even: |
| 75th Quartile / Upper Hinge | th term |
| IQR / H-spread | Upper Hinge - Lower Hinge |
| Step |  |
| Upper Inner Fence | Upper Hinge + 1 Step |
| Lower Inner Fence | Lower Hinge - 1 Step |
| Upper Outer Fence | Upper Hinge + 2 Step |
| Lower Outer Fence | Lower Hinge - 2 Step |
| Upper Adjacent | Largest value below Upper Inner Fence |
| Lower Adjacent | Smallest value below Lower Inner Fence |
| Outlier | Values beyond Upper and Inner Fence |
| Extreme outlier | Values beyond Upper and Outer Fence |
| Outside Value | A value beyond an Inner fence but not beyond an Outer Fence |
| Far Out Value | A value beyond an Outer Fence |

## Trimean

## Geometric Mean

* Usually questions will be per year/certain time frame with percentages of growth. The growth will be in positive/negative percentages.
* Remember to change these into decimals then add 1.
* Ex: 10% = 0.1 = 1+0.1 = 1.1 (this is the growth)
* THEN you can put it in the formula
* Even after you get the result, subtract it by 1 again and convert to percentage.
* Ex: Geom = 1,045 = 1,045 - 1 = 0,045 = 4,5% (this is the average population grown based on the time frame given)

## Trimmed Mean

* Remove the values in the bottom X% and top X% of the dataset. Then, calculate the mean of the remaining values.
* For these trimmed means, don’t be confused if the trim given is the overall trim or the trim for each side.
* Make it make sense, if the number of elements matches the amount to trim, then follow suit.
* Example: 10 total elements, but it says trim by 10%. 5% off the front and back doesn’t make sense, since 5% of 10 it’s barely one element, so just do 10% off front and back.

## Pearson’s correlation

* , where x is and y is

## Probability of a Single Event

## Probability of 2 or more independent events

## Probability of 2 or more dependent events

* It covers 3 possibilities:

1. A occurs and B doesn’t occur
2. B occurs and A doesn’t occur
3. Both A and B occur

## Permutations

* Used when the order matters
* nPr =

## Combinations

* Used when the order doesn’t matter
* nCr =

## Binomial Distribution

* N = number of trials
* k = number of successes
* p = probability of success
* q= probability of failure
* → variance
* STD = sqrt of n\*p\*q

## Poisson Distribution

* A discrete (countable) probability distribution, used to predict the number of times an event occurs.
* e = base of natural logarithms (2.7183)
* = mean number of successes
* x = number of successes in question
* example:

The mean number of calls to a fire station on a weekday is 8.

The probability that on a given weekday there would be 11 calls:

## Multinomial Distribution

* Obtains a specific set of outcomes when there are k possible outcomes for every event.
* p = probability
* n = total number of events
* n1 = number of times Outcome 1 occurs
* nk = number of times Outcome k occurs
* p1 = probability of Outcome 1
* pk = probability of Outcome k

## Hypergeometric Distribution

* Used to calculate probabilities when sampling without replacement
* k = total number of the group for success in the population (if your success is a red ball, how many red balls are there in total?)
* x = number of successes in the sample (what probability is it asking for?)
* N = total number of population (all the objects together without category)
* n = number of selections or samples (how many are we picking from the total?)
* = number of combinations of k things taken x at a time

## Variance and Standard Deviation

* **→** In normal distribution
* **→** population
* **→** sample (for t-test)

## Z score

* N = sample
* P = success
* Q = fail
* Mean = N\*P
* STD = sqrt of n\*p\*q
* Corrected X (i.e. if probability of X<25, then 25-0.5 = 24.5)
* Z-score = (24.5-Mean)/STD
* If looking for less than find z table value, if looking for more then 1-ztable vlaue

## t-test for one sample (independent t test)

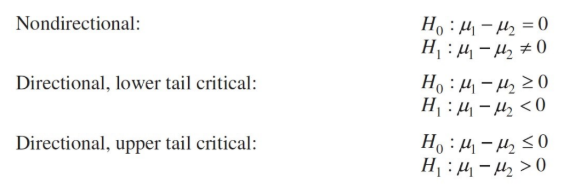
* = calculated mean or sample mean
* = mean given in the question or population mean
* n = number of samples
* Std = Standard deviation
* df = n - 1
* Reject the H0 if the calculated t-value is beyond the t-critical value

## t-test for one sample (paired t test)

* = mean of differences
* n = number of pairs
* Std = Standard deviation
* df = n - 1
* Reject the H0 if the calculated t-value is beyond the t-critical value

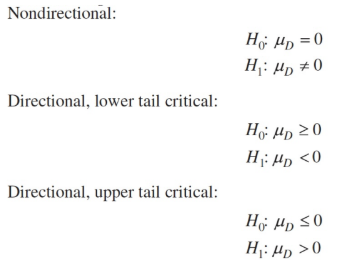
## t-test for two independent samples (paired t test)

* To test if the mean of two independent groups are different.
* H0 =



* “increases” or “improves” → right tailed. Reject H0 if t-value > t-critical value
* **Right-tailed test**: Used if the alternative hypothesis states that the population mean is **greater than** a certain value.
* Ha​:μ>μ0​
* Example: Testing if the new teaching method leads to higher test scores than the traditional method.
* “decreases” or “reduces” → left tailed. Reject H0 if t-value < t-critical value
* **Left-tailed test**: Used if the alternative hypothesis states that the population mean is **less than** a certain value.
* Ha​:μ<μ0​
* Example: Testing if a new drug decreases blood pressure compared to the existing standard.
* Df = (nA + nB)/2

## t-test for two related samples



# Step-by-Step Tutorials

Reminder: Critical Values and p-values are different variables.

1. Critical Values are obtained from the distribution tables for certain tests:

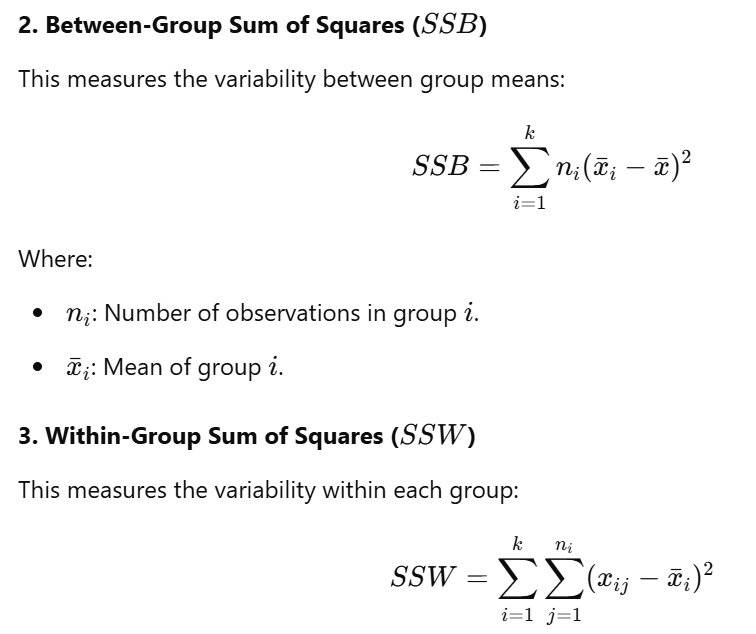
* t-table ([one-tailed](https://www.statisticshowto.com/tables/t-distribution-table/) / [two tailed](https://www.statisticshowto.com/tables/t-distribution-table/#two))
  + One-tailed t-table critical values only cover left side, so if hypothesis asks for something increasing/involves using >, then it’s 1 - P(X)
  + Two-tailed critical value is +, reject null hypothesis if t-value is within the crit value
* z-table ([negative](https://z-table.com/uploads/2/1/7/9/21795380/9340559_orig.png?ezimgfmt=rs:657x781/rscb1/ng:webp/ngcb1) / [positive](https://z-table.com/uploads/2/1/7/9/21795380/8573955.png?ezimgfmt=rs:657x793/rscb1/ng:webp/ngcb1))
  + Left side describes the first two parts of the decimal of your calculated Z value.
  + Top side describes the hundredths place of your Z value.
* f-table ( ⍺ = [0.10](https://i0.wp.com/statisticsbyjim.com/wp-content/uploads/2022/02/F-table_Alpha10.png?w=814&ssl=1) / [0.05](https://i0.wp.com/statisticsbyjim.com/wp-content/uploads/2022/02/F-table_Alpha05.png?w=817&ssl=1) / [0.01](https://statisticsbyjim.com/wp-content/uploads/2022/02/F-table_Alpha01.png))
* Chi squared ([table](https://www.chisquaretable.net/), [calculator](https://www.socscistatistics.com/tests/chisquare2/default2.aspx))

1. P-values are more of the probability, which you need to use a calculator to get. Usually, the p-value you can get from the calculated statistic and degree of freedom. After getting the p-value, compare it with the significance level which is the 𝝰. Reject H0 if the p-value is less than the 𝝰.

## One Way ANOVA - Independent Measures

* MSB = Mean squared between groups
* MSW mean squared within groups

K is total groups, N is total observations



N = Total no. of ppl/objects in the experiment

n = No. of ppl/objects per group

a = No. of experimental groups/conditions

μ = Sample mean

SS = Sum of Squares

1. State Null and Alternate Hypothesis
   * H0 = μ1 = μ2 = … = μn.
   * H1 = Not all μ’s are the same.
2. Find degrees of freedom
   * dfbetween = a - 1 (df numerator)
   * dfwithin = N - a (df denominator)
   * dftotal = N - 1
3. Find critical value with the 2 df’s calculated.
   * [**Calculator**](https://www.socscistatistics.com/tests/criticalvalues/default.aspx) **/** [**Table**](https://statisticsbyjim.com/wp-content/uploads/2022/02/F-table_Alpha05.png)
4. Calculate F-Statistic Value:

**SS Between:**

* + SSbetween = where:
    1. is the sum of squared sums of all groups. [(ΣA)2 + (ΣB)2 + … + (ΣZ)2]
* Find total for Group A and square it, find total for Group B and square it, and so on for all groups.
  + 1. T2 is the sum of all elements and then squared[(ΣA) + (ΣB) + … + (ΣZ)]2
* Add all elements together, then square it.

**OR USE THIS:**

Stupid Language Formula:

* SSbetween = n [(mean of every group - grand mean)2  + … + (mean of every group - grand mean)2 ]

**SS Within:**

* + SSwithin = where:
    1. ΣY2 is the sum of each value squared [a2 + b2 + c2+...+z2]

**OR USE THIS:**

Stupid Language Formula:

* Calculate for every group:
* SSwithin (group A) = (every value in group A - mean of group A)2 + … + (every value in group A - mean of group A)2
* SSwithin (group B) = (every value in group B - mean of group B)2 + … + (every value in group B - mean of group B)2
* Continue for all groups
* SSwithin = Sum of all SSwithin groups

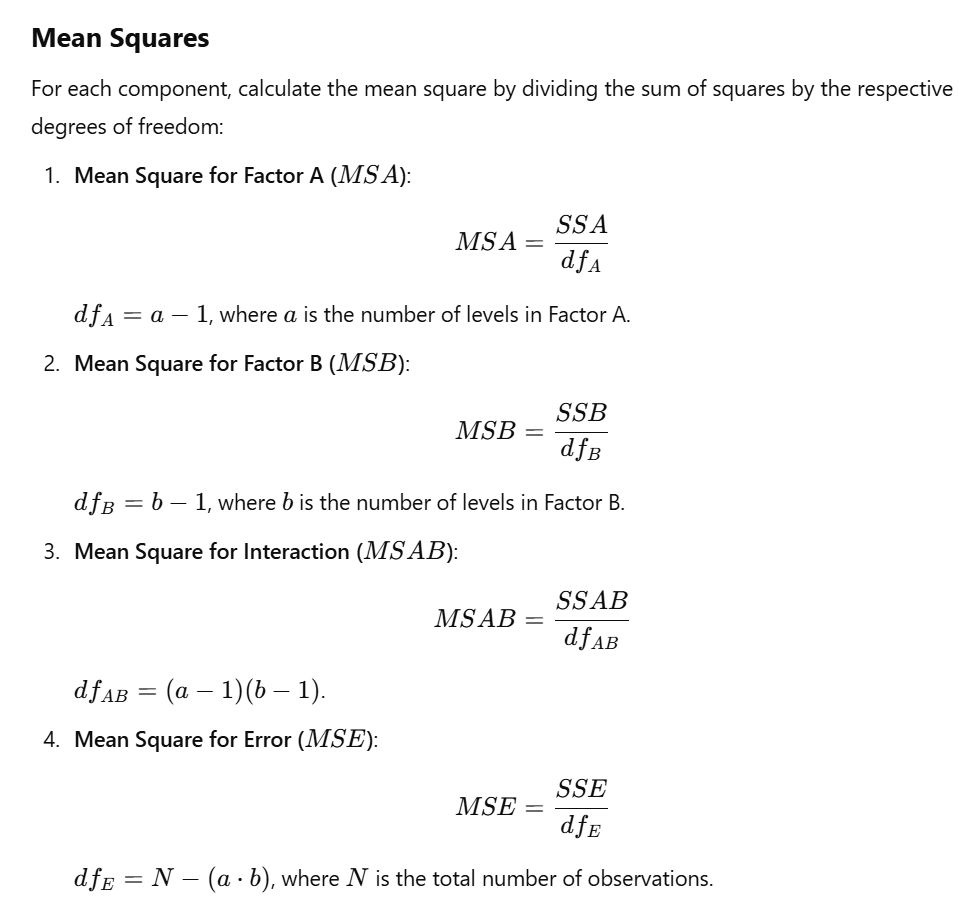
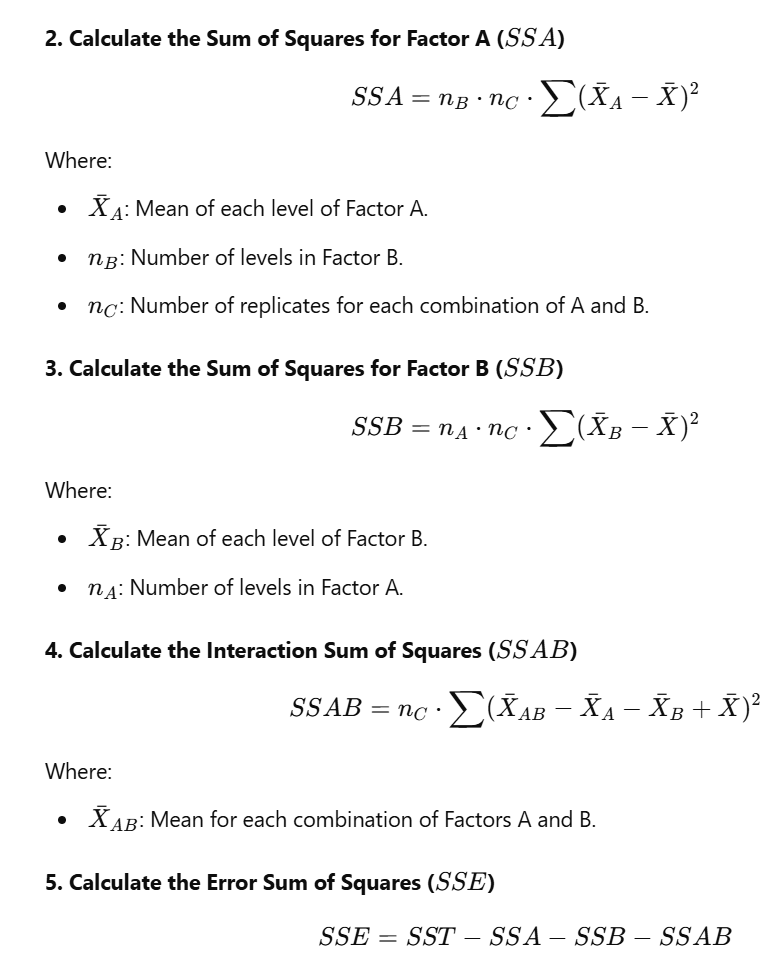
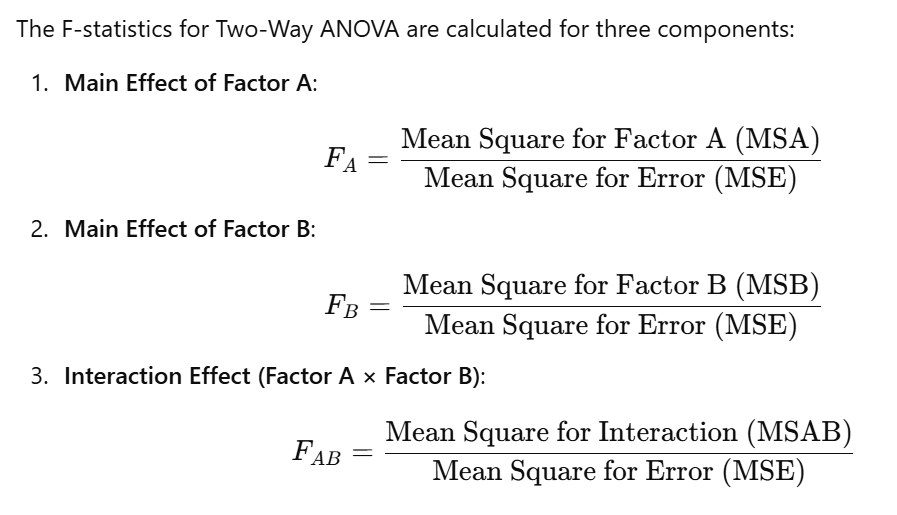
**SS Total:**

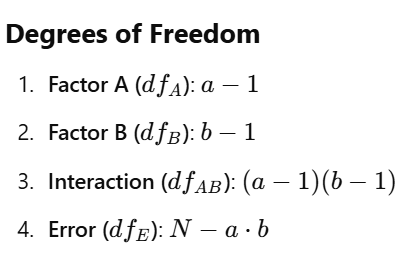
* + SStotal = SSbetween + SSwithin

1. Find Mean Squares:
   * MSbetween =
   * MSwithin =
2. Calculate F Statistic:
   * F =
3. Compare F with the Critical Value:
   * If calculated F-statistic > critical value, reject the null hypothesis.
   * If calculated p-value < significance level, reject null hypothesis

* There is **no accepting** of the null hypothesis, only rejecting and failure to reject it.

## Two Way ANOVA





N = Total no. of ppl/objects in the experiment

n = No. of ppl/objects per group

a = No. of experimental groups/conditions

μ = Sample mean

SS = Sum of Squares

p = number of categories in Group 1

q = number of categories in Group 2

σ = Mean Square

1. State Null and Alternate Hypothesis:
   * H0 = Mean test score across Group 1 is the same
   * H0 = Mean test score across Group 2 is the same
   * H0 = There is no interaction between Group 1 and Group 2
2. Calculate:

* Grand Mean =
* Group 1 Mean =
* Group 2 Mean =
* Mean for every Group 1 x Group 2 combination (category)
  + If Group 1 has m categories, and Group 2 has n categories, you need to find m\*n different means.
  + Example, Group 1: Python, Java, C++ | Group 2: Self-learn, Instructed
    - Find mean of Python and Self-learn, Python and Instructed, Java and Self-learn, etc. (in this case you’ll have 6 different means)

1. Total:

* SStotal = (x1 - grand mean)2 + … + (xn - grand mean)2
  + x here refers to each and every individual recorded value in the dataset.
* dftotal =
* Mean Square total =

1. Between:

* SSbetween = [(mean of every category - grand mean)2 + … + (mean of every category - grand mean)2]
  + Get the sum first THEN multiply with the n.
  + Mean of every category is the one from the previous example where its the mean of the combination of Group 1 and Group 2
    - Example, Mean of Python and Self-learn is 6.4:
      * In formula, that would be (6.4 -Grand Mean)2 + ...
* dfbetween =
* Mean Square between =

1. Group 1:

* SSgroup 1 = [(mean of every category for group 1- grand mean)2 + … + (mean of every category for group 1 - grand mean)2 ]
  + Example, if the categories in group 1 are Python, Java and C++, only use the means of those values.
* dfgroup 1 =
* Mean Square group 1 =

1. Group 2:

* SSgroup 2 = [(mean of every category for group 2- grand mean)2 + … + (mean of every category for group 2 - grand mean)2]
* Example, if the categories in Group 2 are Self-learn and Instructed, only use the means of those values.
* dfgroup 2 =
* Mean Square group 2 =

1. Interaction:

* SSinteraction = SSbetween - SSgroup 1 - SSgroup 2
* Dfinteraction = (p - 1) (q - 1)
* Mean Square interaction =

1. Error:

* SSerror = (every sample - the mean of its respective category)2 + … + (every sample - the mean of its respective category)2
* Dferror = (n - 1)
* Mean Squares error =

1. F-Values:

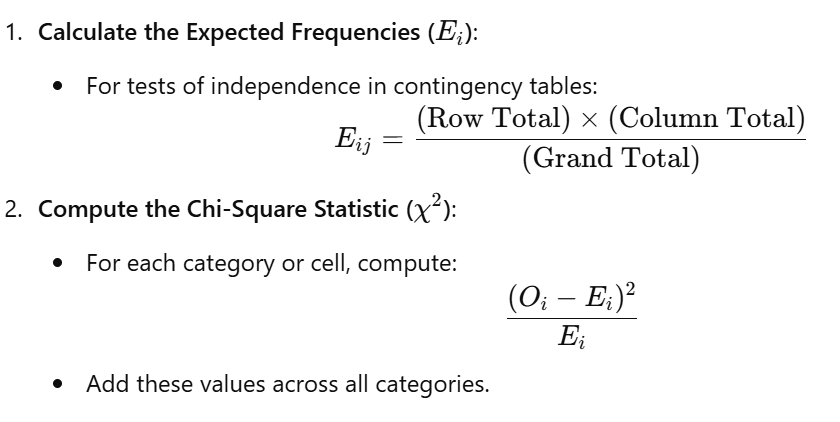
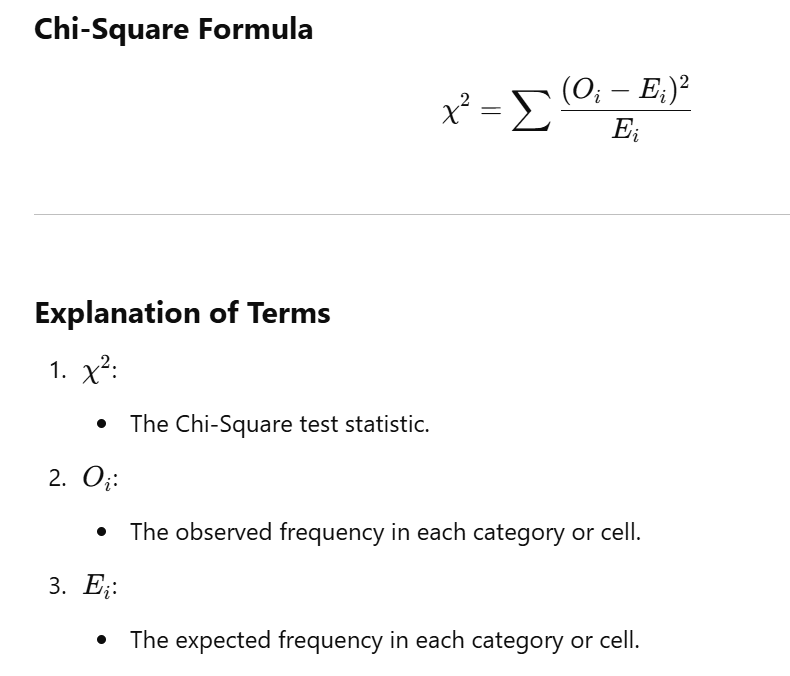
* F Value Group 1:
* F Value Group 2:
* F Value Interaction:

1. Null Hypothesis:

* Pick either method depending on what ur told to do in the question
* P-Value Method
  + Input into P Value Calculator, find P value.
  + Input F-Value Group 1, dfgroup 1  (df numerator), and Dferror (df denominator)
  + Continue for F values of Group 2 and Interaction (df denominator is dferror for all)
  + If value is less than the alpha given in question (in this case 0.05), reject null hypothesis.
  + If any of the f values are greater than the null hypothesis, then fail to reject hypothesis.
* F-Critical Method
  + Find F-Critical Value with calculator/table.
  + Input alpha (a), Dfgroup 1  (df numerator), and Dferror (df denominator)
  + Find F-Critical Value for Group 2 and Interaction as well.
  + If F value greater than F-Critical, reject null hypothesis. If smaller than, fail to reject null hypothesis.

## Chi Square

<https://www.socscistatistics.com/tests/chisquare2/default2.aspx>



1. H0 = The 2 groups are independent

H1 = The 2 groups are not independent.

1. Calculate expected Frequency Table:

Eij =

1. Calculate the Chi-Square

For every cell:

X =

1. Calculate Degree of Freedom
2. Get the critical value from the [table](https://www.chisquaretable.net/) or get p value with a [calculator](https://www.socscistatistics.com/tests/criticalvalues/default.aspx).
3. Reject H0 if > critical value.

Reject H0 if p value < significance level.

# Links

1. Casio Scientific Calculator:

* <https://mathda.com/calculator/>

1. 5 Number Summary + IQR + Inner Outer Fence + Outliers + Geometric Mean + Sum of Squares + Standard Deviation (Sample/Population) + Variance Calculator:

* <https://www.hackmath.net/en/calculator/five-number-summary>

1. Trimmed Mean Calculator:

* [Trimmed Mean Calculator](https://www.statology.org/trimmed-mean-calculator/)

1. Permutation Combination Calculator:

* <https://www.calculator.net/permutation-and-combination-calculator.html>

1. Binomial Distribution (Singular and Cumulative):

* <https://stattrek.com/online-calculator/binomial>

1. Pearson's Correlation Coefficient:

* <https://www.socscistatistics.com/tests/pearson/default2.aspx>

1. One Way ANOVA - Independent Measures

* <https://www.socscistatistics.com/tests/anova/default2.aspx>

1. Two Way ANOVA

* [Two-Way-ANOVA Calculator - With AI Interpretation - DATAtab](https://datatab.net/statistics-calculator/hypothesis-test/two-way-anova-calculator)

1. Single Sample T-Test:

* <https://www.socscistatistics.com/tests/tsinglesample/default.aspx>

1. Chi-Square Test:

* <https://www.socscistatistics.com/tests/chisquare2/default2.aspx>

1. Critical Value for Multiple Tests:

* <https://www.socscistatistics.com/tests/criticalvalues/default.aspx>

1. P Value Calculator and F Critical Table for ANOVA Test:

* <https://datatab.net/tutorial/f-distribution>

1. Pearson Edexcel Formula Book, Statistics S1, S2 and S3 with complete Z table, Normal Distribution Table (z table), and Statistical Formula(Page 14 - 23)

* <https://qualifications.pearson.com/content/dam/pdf/International%20Advanced%20Level/Mathematics/2018/Specification-and-Sample-Assessment/IAL-Mathematics-Formula-Book.pdf>

# Exercises

[two-tailed t-test exercise 6 numb 1.jpg](https://drive.google.com/file/d/1UHZZIF8_cuoToCLWXAI_I0XkQecSCTtj/view?usp=drive_link)

[Two Way ANOVA Practice.png](https://drive.google.com/file/d/17K7IgkoP8zcHELrrk9V8n3jEfM1LCc_U/view?usp=drive_link)

[Two Way ANOVA Exercise 7.jpg](https://drive.google.com/file/d/1VCFdQxwr6l3nrGU6s3STrKLdh-DU79Se/view?usp=drive_link)

[t-test exercise 6 numb 3.jpg](https://drive.google.com/file/d/1Qel7IsWkGD23NAU7OMKGLvZRmKF6pj5G/view?usp=drive_link)

[t-test exercise 6 numb 2.jpg](https://drive.google.com/file/d/1wj25YMfKMuokvoTvzTVKD4ix5vAXV7_L/view?usp=drive_link)

[standard deviation - exercise 5 numb 1.jpg](https://drive.google.com/file/d/12iuHPfWYBf3svyOlufWc_2XjNCq5A-Zr/view?usp=drive_link)

[probability via binomial distribution exercise 4 num 5&6.jpg](https://drive.google.com/file/d/188wyG46jupWb0YyJSKZoWSo3T9_CxaRG/view?usp=drive_link)

[pearson coefficient exercise 4 num 7.jpg](https://drive.google.com/file/d/1QYorfBvPkUlzdynWsvxS74qfscDWVnLz/view?usp=drive_link)

[One Way ANOVA Exercise 7.png](https://drive.google.com/file/d/1PAEmvWAAc9R3o6mMhVXwW5NhUFyggASk/view?usp=drive_link)

[normal distribution and z-statistic exercise 5 numb 3.jpg](https://drive.google.com/file/d/1cLw-HkioXZ0nvP_dh3NzpYWKtgy8ektY/view?usp=drive_link)

[normal distribution and z-statistic exercise 5 numb 2.jpg](https://drive.google.com/file/d/1qAO-wM-HQMI5OMowXvs5G_wb4eiqQMc1/view?usp=drive_link)

[Exercise 1 - Answers](https://drive.google.com/file/d/1__CkfNej8wO2K-nATeC5DdAtRgqjJuXv/view?usp=drive_link)

[Exercise 2 - Answers](https://drive.google.com/file/d/1nWOSqB7J0M3rbrw0cJg3hND1xQHaX4Ma/view?usp=drive_link)

[Exercise 3 - Answers](https://drive.google.com/file/d/134y4ye4KONTuwXY7Q1LDgur2G2uBju2i/view?usp=drive_link)

[Exercise 4 - Answers](https://drive.google.com/file/d/1aoJW6-ctATh95PU_AG9S2pARFbvneNHU/view?usp=drive_link)

[Exercise 5 - Answers](https://drive.google.com/file/d/19ramraRx58qf7GlcfzJlms8UT1vrAkYq/view?usp=drive_link)

[Exercise 6 - Answers](https://drive.google.com/file/d/1uPi2Qhta4hONKv8HA39qjBxD48VRyePu/view?usp=drive_link)

[Exercise 7 - Answers](https://drive.google.com/file/d/1yzXewDJVbG4J8LdJ2R_4_Va1lpmSbLd9/view?usp=drive_link)