TABLE OF CONTENTS

Tables and Figures

- **Frequency Table**

Previous Ownership	Frequency	Relative Frequency
None	85	$\frac{85}{500} = 0.17$
Windows	60	$\frac{60}{500} = 0.12$
Macintosh	355	$\frac{355}{500} = 0.71$
Total	500	$\frac{500}{500} = 1$

- **Stem and Leaf**

```
3|2337
2|001112223889
1|2244456888899
0|69
```

```
3|7
3|233
2|889
2|001112223
1|56888899
1|22444
0|69
```

11	4	
	3	7
332	3	233
8865	2	889
44331110	2	001112223
987776665	1	56888899
321	1	22444
7	0	69

Formulas

1. **Box and Plot**

Name	**Formula**
25th Quartile / Lower Hinge	$\frac{(n+1)}{4}$th term
50th Quartile / Median	When n is odd: $\frac{(n+1)}{2}$th term When n is even: $\frac{(\frac{n}{2})th\ term + ((\frac{n}{2})+1)th\ term}{2}$
75th Quartile / Upper Hinge	$\frac{3(n+1)}{4}$th term
IQR / H-spread	Upper Hinge - Lower Hinge
Step	$1.5 \times HSpread$
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

2. **Trimean**

- $Trimean = \frac{Q_1 + 2Q_2 + Q_3}{4}$

3. **Geometric Mean**
    - $\overline{x}\ geom = \sqrt[n]{x_1 \times x_2 \times \ \dots \ \times x_n}$
    - 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)
4. **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.
5. **Pearson's correlation**
    - $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}}$, where x is $X - \overline{X}$ and y is $Y - \overline{Y}$
6. **Probability of a Single Event**
    - $probability = \frac{possible\ outcomes}{total\ outcomes}$
7. **Probability of 2 or more independent events**
    - $P(A\ and\ B) = P(A) \times P(B)$
8. **Probability of 2 or more dependent events**
    - $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$

- 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

9. Permutations

- Used when the order matters
- ${}_nP_r = \frac{n!}{(n-r)!}$

10. Combinations

- Used when the order doesn't matter
- ${}_nC_r = \frac{n!}{r!(n-r)!}$

11. Binomial Distribution

- $P(x) = \frac{N!}{k!(N-k)!} p^k (1 - p)^{N-k}$
- N = number of trials
- k = number of successes
- p = probability of success
- q= probability of failure
- $mean = N \times p$
- $\sigma^2 = Np(1 - \Pi)$ → variance
- STD = sqrt of n*p*q

12. Poisson Distribution

- A discrete (countable) probability distribution, used to predict the number of times an event occurs.
- $p = \frac{e^{-\mu} \mu^x}{x!}$
- e = base of natural logarithms (2.7183)
- μ = 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:

- $p = \frac{e^{-8}8^{11}}{11!} = 0.072$

13. Multinomial Distribution

- Obtains a specific set of outcomes when there are k possible outcomes for every event.
- $p = \frac{n!}{(n_1!)(n_2!)...(n_k!)} p_1^{n1} p_2^{n2} \cdots p_k^{nk}$
- p = probability
- n = total number of events
- n_1 = number of times Outcome 1 occurs
- n_k = number of times Outcome k occurs
- p_1 = probability of Outcome 1
- p_k = probability of Outcome k

14. Hypergeometric Distribution

- Used to calculate probabilities when sampling without replacement
- $p = \frac{{}_kC_x \; {}_{(N-k)}C_{(n-x)}}{{}_NC_n}$
- 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?)
- ${}_kC_x$ = number of combinations of k things taken x at a time
- $mean = \frac{nk}{N}$
- $sd = \sqrt{\frac{(n)(k)(N-k)(N-n)}{N^2(N-1)}}$

15. Variance and Standard Deviation

- $Variance\ =\ \sigma^2 = \frac{\Sigma(x-\bar{x})^2}{N}$
- $SD\ =\ \sigma\ =\sqrt{npq}$ → In normal distribution
- $SD\ =\ \sigma\ =\sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$ → population
- $SD\ =\ \sigma\ =\sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}}$ → 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

16. t-test for one sample (independent t test)

- $t\ =\frac{\bar{x}-\mu}{\frac{Std}{\sqrt{n}}}$
- $\bar{x}$ = calculated mean or sample mean
- μ = mean given in the question or population mean
- n = number of samples
- Std = Standard deviation
- df = n - 1
- Reject the H_0 if the calculated t-value is beyond the t-critical value

17. t-test for one sample (paired t test)

- $t = \frac{\bar{d}}{\frac{Std}{\sqrt{n}}}$
- $\bar{d}$ = mean of differences
- n = number of pairs
- Std = Standard deviation
- df = n - 1
- Reject the H_0 if the calculated t-value is beyond the t-critical value

18. t-test for two independent samples (paired t test)

- To test if the mean of two independent groups are different.
- H0 = $\mu_1 = \mu_2$

Nondirectional:	$H_0 : \mu_1 - \mu_2 = 0$ $H_1 : \mu_1 - \mu_2 \neq 0$
Directional, lower tail critical:	$H_0 : \mu_1 - \mu_2 \geq 0$ $H_1 : \mu_1 - \mu_2 < 0$
Directional, upper tail critical:	$H_0 : \mu_1 - \mu_2 \leq 0$ $H_1 : \mu_1 - \mu_2 > 0$

- "increases" or "improves" → right tailed. Reject H_0 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:$\mu > \mu 0$
  - Example: Testing if the new teaching method leads to higher test scores than the traditional method.
- "decreases" or "reduces" → left tailed. Reject H_0 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:$\mu < \mu 0$
  - Example: Testing if a new drug decreases blood pressure compared to the existing standard.

- $t = \frac{\overline{xA} - \overline{xB}}{\sqrt{\frac{SDA^2}{nA} + \frac{SDB^2}{nB}}}$
- Df = (nA + nB)/2

19. t-test for two related samples

Nondirectional:

$$H_0: \mu_D = 0$$
$$H_1: \mu_D \neq 0$$

Directional, lower tail critical:

$$H_0: \mu_D \geq 0$$
$$H_1: \mu_D < 0$$

Directional, upper tail critical:

$$H_0: \mu_D \leq 0$$
$$H_1: \mu_D > 0$$

- $t = \frac{\overline{x} - \mu}{\frac{SD}{\sqrt{n}}}$

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 / two tailed)
     - One-tailed t-table critical values only cover left side, so if hypothesis asks for something increasing/involves using $\geq$, then it's 1 - P(X)
     - Two-tailed critical value is $\pm$, reject null hypothesis if t-value is within the crit value
   - z-table (negative / positive)
     - 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 (α = 0.10 / 0.05 / 0.01)
   - Chi squared (table, calculator)
2. 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**
- $F = \frac{MSB}{MSW}$
- MSB = Mean squared between groups
- MSW mean squared within groups
- $MSB = \frac{SSB}{k-1}$
- $MSW = \frac{SSW}{N-k}$

  K is total groups, N is total observations

2. Between-Group Sum of Squares (SSB)

This measures the variability between group means:

$$SSB = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$$

Where:

- n_i: Number of observations in group i.
- $\bar{x}_i$: Mean of group i.

3. Within-Group Sum of Squares (SSW)

This measures the variability within each group:

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$$

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 = $\mu_1 = \mu_2 = \ldots = \mu_n$.
   - H1 = Not all μ's are the same.
2. Find degrees of freedom
   - $df_{between}$ = a - 1 (df numerator)
   - df_{within} = N - a (df denominator)
   - df_{total} = N - 1
3. Find critical value with the 2 df's calculated.
   - Calculator / Table
4. Calculate F-Statistic Value:

   SS Between:

   - $SS_{between} = \frac{\Sigma(\Sigma a_i)^2}{n} - \frac{T^2}{N}$ where:

i. $\Sigma(\Sigma a_i)^2$ is the sum of squared sums of all groups. $[(\Sigma A)^2 + (\Sigma B)^2 + \ldots + (\Sigma Z)^2]$

- Find total for Group A and square it, find total for Group B and square it, and so on for all groups.

ii. T^2 is the sum of all elements and then squared$[(\Sigma A) + (\Sigma B) + \ldots + (\Sigma Z)]^2$

- Add all elements together, then square it.

OR USE THIS:

Stupid Language Formula:

- $SS_{between}$ = n [(mean of every group - grand mean)^2 + … + (mean of every group - grand mean)^2]

SS Within:

- $SS_{within} = \Sigma Y^2 - \frac{\Sigma(\Sigma a_i)^2}{n}$ where:

i. ΣY^2 is the sum of each value squared $[a^2 + b^2 + c^2 + ... + z^2]$

OR USE THIS:

Stupid Language Formula:

- Calculate for every group:
- SS_{within} (group A) = (every value in group A - mean of group A)^2 + … + (every value in group A - mean of group A)^2
- SS_{within} (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
- SS_{within} = Sum of all SS_{within} groups

SS Total:

- $SS_{total} = SS_{between} + SS_{within}$

5. Find Mean Squares:
   - $MS_{between} = \frac{SS_{between}}{df_{between}}$
   - $MS_{within} = \frac{SS_{within}}{df_{within}}$

6. Calculate F Statistic:
    - $F = \frac{MS_{between}}{MS_{within}}$
7. 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**

The F-statistics for Two-Way ANOVA are calculated for three components:

1. **Main Effect of Factor A**:

$$F_A = \frac{\text{Mean Square for Factor A (MSA)}}{\text{Mean Square for Error (MSE)}}$$

2. **Main Effect of Factor B**:

$$F_B = \frac{\text{Mean Square for Factor B (MSB)}}{\text{Mean Square for Error (MSE)}}$$

3. **Interaction Effect (Factor A × Factor B)**:

$$F_{AB} = \frac{\text{Mean Square for Interaction (MSAB)}}{\text{Mean Square for Error (MSE)}}$$

2. Calculate the Sum of Squares for Factor A (SSA)

$$SSA = n_B \cdot n_C \cdot \sum(\bar{X}_A - \bar{X})^2$$

Where:

- $\bar{X}_A$: Mean of each level of Factor A.
- n_B: Number of levels in Factor B.
- n_C: Number of replicates for each combination of A and B.

3. Calculate the Sum of Squares for Factor B (SSB)

$$SSB = n_A \cdot n_C \cdot \sum(\bar{X}_B - \bar{X})^2$$

Where:

- $\bar{X}_B$: Mean of each level of Factor B.
- n_A: Number of levels in Factor A.

4. Calculate the Interaction Sum of Squares ($SSAB$)

$$SSAB = n_C \cdot \sum(\bar{X}_{AB} - \bar{X}_A - \bar{X}_B + \bar{X})^2$$

Where:

- $\bar{X}_{AB}$: Mean for each combination of Factors A and B.

5. Calculate the Error Sum of Squares (SSE)

$$SSE = SST - SSA - SSB - SSAB$$

Mean Squares

For each component, calculate the mean square by dividing the sum of squares by the respective degrees of freedom:

1. **Mean Square for Factor A (MSA):**

$$MSA = \frac{SSA}{df_A}$$

$df_A = a - 1$, where a is the number of levels in Factor A.

2. **Mean Square for Factor B (MSB):**

$$MSB = \frac{SSB}{df_B}$$

$df_B = b - 1$, where b is the number of levels in Factor B.

3. **Mean Square for Interaction ($MSAB$):**

$$MSAB = \frac{SSAB}{df_{AB}}$$

$df_{AB} = (a - 1)(b - 1)$.

4. **Mean Square for Error (MSE):**

$$MSE = \frac{SSE}{df_E}$$

$df_E = N - (a \cdot b)$, where N is the total number of observations.

Degrees of Freedom

1. Factor A (df_A): $a - 1$
2. Factor B (df_B): $b - 1$
3. Interaction (df_{AB}): $(a - 1)(b - 1)$
4. Error (df_E): $N - a \cdot b$

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:
    - H_0 = Mean test score across Group 1 is the same
    - H_0 = Mean test score across Group 2 is the same
    - H_0 = There is no interaction between Group 1 and Group 2
2. Calculate:
    - Grand Mean = $\frac{sum\ of\ all\ values}{N}$
    - Group 1 Mean = $\frac{sum\ of\ values\ in\ group\ 1}{n}$
    - Group 2 Mean = $\frac{sum\ of\ values\ in\ group\ 2}{n}$
    - 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)
3. Total:
    - SS_{total} = $(x_1$ - grand mean$)^2$ + … + $(x_n$ - grand mean$)^2$
        - x here refers to each and every individual recorded value in the dataset.
    - $df_{total} = n \times p \times q - 1$
    - Mean Square $_{total}$ = $\frac{SS_{total}}{df_{total}}$

4. Between:
    - $SS_{between} = n \times$ [(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$)^2$ + ...
    - $df_{between} = p \times q - 1$

- Mean Square $_{\text{between}} = \frac{SS_{between}}{df_{between}}$

5. Group 1:
   - $SS_{\text{group 1}} = n \times q \times$ [(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.
   - $df_{\text{group 1}} = p - 1$
   - Mean Square $_{\text{group 1}} = \frac{SS_{group\ 1}}{df_{group\ 1}}$
6. Group 2:
   - $SS_{\text{group 2}} = n \times p \times$ [(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.
   - $df_{\text{group 2}} = q - 1$
   - Mean Square $_{\text{group 2}} = \frac{SS_{group\ 2}}{df_{group\ 2}}$
7. Interaction:
   - $SS_{\text{interaction}} = SS_{\text{between}} - SS_{\text{group 1}} - SS_{\text{group 2}}$
   - $Df_{\text{interaction}} = (p - 1)(q - 1)$
   - Mean Square $_{\text{interaction}} = \frac{SS_{interaction}}{df_{interaction}}$
8. Error:
   - SS_{error} = (every sample - the mean of its respective category)^2 + … + (every sample - the mean of its respective category)^2
   - $Df_{\text{error}} = (n - 1) \times p \times q$
   - Mean Squares $_{\text{error}} = \frac{SS_{error}}{df_{error}}$
9. F-Values:
   - F Value Group 1: $\frac{Mean\ Square_{Group\ 1}}{Mean\ Square_{Error}}$

- F Value Group 2: $\frac{Mean\ Square_{Group\ 2}}{Mean\ Square_{Error}}$
- F Value Interaction: $\frac{Mean\ Square_{Interaction}}{Mean\ Square_{Error}}$

10. 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, $df_{group\ 1}$ (df numerator), and Df_{error} (df denominator)
        - Continue for F values of Group 2 and Interaction (df denominator is df_{error} 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), $Df_{group\ 1}$ (df numerator), and Df_{error} (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

Chi-Square Formula

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Explanation of Terms

1. χ^2:
   - The Chi-Square test statistic.
2. O_i:
   - The observed frequency in each category or cell.
3. E_i:
   - The expected frequency in each category or cell.

1. **Calculate the Expected Frequencies (E_i):**
   - For tests of independence in contingency tables:
   $$E_{ij} = \frac{(\text{Row Total}) \times (\text{Column Total})}{(\text{Grand Total})}$$
2. **Compute the Chi-Square Statistic (χ^2):**
   - For each category or cell, compute:
   $$\frac{(O_i - E_i)^2}{E_i}$$
   - Add these values across all categories.

1) H_0 = The 2 groups are independent

   H_1 = The 2 groups are not independent.

2) Calculate expected Frequency Table:

   $$E_{ij} = \frac{Row\ Total \times Column\ Total}{Grand\ Total}$$

3) Calculate the Chi-Square

   For every cell:

   $$X = \frac{(original\ value - expected\ frequency)^2}{expected\ frequency}$$

   $$X^2 = Sum\ of\ all\ X\ from\ all\ cells$$

4) Calculate Degree of Freedom

$$df = (no\ of\ rows - 1) \times (no\ of\ columns - 1)$$

5) Get the critical value from the table or get p value with a calculator.

6) Reject H_0 if X^2 > critical value.

   Reject H_0 if p value < significance level.

Links

1. Casio Scientific Calculator:
   - https://mathda.com/calculator/
2. 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
3. Trimmed Mean Calculator:
   - Trimmed Mean Calculator
4. Permutation Combination Calculator:
   - https://www.calculator.net/permutation-and-combination-calculator.html
5. Binomial Distribution (Singular and Cumulative):
   - https://stattrek.com/online-calculator/binomial
6. Pearson's Correlation Coefficient:
   - https://www.socscistatistics.com/tests/pearson/default2.aspx
7. One Way ANOVA - Independent Measures
   - https://www.socscistatistics.com/tests/anova/default2.aspx
8. Two Way ANOVA
   - Two-Way-ANOVA Calculator - With AI Interpretation - DATAtab
9. Single Sample T-Test:
   - https://www.socscistatistics.com/tests/tsinglesample/default.aspx
10. Chi-Square Test:
    - https://www.socscistatistics.com/tests/chisquare2/default2.aspx
11. Critical Value for Multiple Tests:
    - https://www.socscistatistics.com/tests/criticalvalues/default.aspx
12. P Value Calculator and F Critical Table for ANOVA Test:
    - https://datatab.net/tutorial/f-distribution
13. 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
- Two Way ANOVA Practice.png
- Two Way ANOVA Exercise 7.jpg
- t-test exercise 6 numb 3.jpg
- t-test exercise 6 numb 2.jpg
- standard deviation - exercise 5 numb 1.jpg
- probability via binomial distribution exercise 4 num 5&6.jpg
- pearson coefficient exercise 4 num 7.jpg
- One Way ANOVA Exercise 7.png
- normal distribution and z-statistic exercise 5 numb 3.jpg
- normal distribution and z-statistic exercise 5 numb 2.jpg

Exercise 1 - Answers

Exercise 2 - Answers

Exercise 3 - Answers

Exercise 4 - Answers

Exercise 5 - Answers

Exercise 6 - Answers

Exercise 7 - Answers