Descriptive & Inferential Statistics

- **Descriptive statistics**: summarize and describe the main features of a dataset. Provides insight of the given data.

  Key features:
  - Measures of central tendency: Mean, Median, Mode
  - Measures of variability: Range, Variance, STD

- **Inferential statistics**: use a random sample of data taken from a population to create inferences or generalizations about the population as a whole. Draw conclusions and make predictions about a population based on sample data.

  Key features
  - Hypothesis testing: Chi-squared, ANOVA, t-test, etc.
  - Regression analysis

Types of Data

- Quantitative Data: Numerical values representing amounts (e.g., weights).
- Ranked Data: Numbers representing relative standing (e.g., race positions).
- Qualitative Data: Descriptive characteristics (e.g., feelings, appearances).

Levels of Measurement

- Nominal: Classification without order (e.g., gender).
- Ordinal: Ranked data indicating order (e.g., race positions).
- Interval/Ratio: Quantitative data with equal intervals and a true zero (e.g., weight).

Types of Variables

- Independent Variables: Manipulated by the experimenter (e.g., type of antidepressant).
- Dependent Variables: Measured outcomes (e.g., relief from depression).

Examples of Independent and Dependent Variables

1. Blueberries and Aging:

- Independent: Dietary supplement (none, blueberry, strawberry, spinach).
- Dependent: Memory and motor skills test results.

2. Beta-Carotene and Cancer:
    - Independent: Supplements (beta-carotene or placebo).
    - Dependent: Occurrence of cancer.

3. Brake Light Brightness:
    - Independent: Brightness of brake lights.
    - Dependent: Time to hit brakes.

Distributions

1. Graphing Distributions
2. Summarizing Distributions

Bivariate Data

Involves analyzing two variables and their relationship (e.g., height and weight, SAT score and age).

1. Bivariate Analysis

   Investigate the relationship between two variables.

   a. **Scatter Plots**: Visual representation of data showing potential patterns between variables.
   b. **Regression Analysis**: Examines how variables are related and provides an equation for the relationship.
   c. **Correlation Coefficients**: A numerical measure (ranging from -1 to 1) that indicates the strength and direction of the linear relationship between two variables.

Probability**

1. Calculating Probability
    a. ***Single event*** probability is calculated by dividing the number of favorable outcomes by the total number of possible outcomes.
    b. *Independent events*: For two independent events (like coin flips), the probability of both events occurring is the product of their individual probabilities.
    c. *Dependent events*: For events where one affects the other (like drawing two aces from a deck), conditional probability is used.

2. Counting Techniques
    a. Permutations: Count the number of ways to arrange a set of items.

$$P(n, r) = \frac{n!}{(n-r)!}$$

    b. Combinations: Count the number of ways to choose items where order doesn't matter.

$$C(n, r) = \frac{n!}{r!(n-r)!}$$

3. Probability Distributions
    a. **Binomial Distribution***: Models the probability of a specific number of successes in a fixed number of trials (e.g., flipping a coin multiple times).

        The probability of getting exactly *k* successes out of *n* trials is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

        Where:
        i. (n k) is the binomial coefficient (combination, C (n, k)) which represents the number of ways to choose k successes from n trials.
        ii. P^k is the probability of having k successes
        iii. $(1 - p)^{n-k}$ is the probability of having n - k failures

    b. **Poisson Distribution**: Used for counting the number of events occurring within a fixed interval, such as phone calls to a service center.

c. **Multinomial Distribution**: An extension of the binomial distribution, used for outcomes with more than two possibilities (e.g., three different outcomes in a chess game).
d. **Hypergeometric Distribution**: Used when sampling without replacement (e.g., drawing cards from a deck without returning them).
e. **Normal Distribution:** Yeah

4. Bayes Theorem

Bayes' Theorem is a fundamental concept in probability theory that describes how to update the probability of a hypothesis (or event) based on new evidence or data. It provides a way to revise existing beliefs in light of new information. The theorem is named after the Reverend Thomas Bayes, who introduced it in the 18th century.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where:

- *P(A|B)*: The **posterior probability** of event A occurring given that event B has occurred. This is the updated probability of A after taking into account the new evidence (B).
- *P(B|A)*: The **likelihood** of observing event B given that event A is true.
- *P(A)*: The **prior probability** of event A occurring before observing the new evidence. It reflects your initial belief about A.
- *P(B)*: The **marginal likelihood** or the total probability of observing event B, regardless of A. This can be computed as the sum of the likelihoods over all possible events.

Research Design

Normal Distribution**

is a bell-shaped curve commonly used in statistics, representing a wide range of real-world phenomena.

Key concepts:

1. **Standard Deviation**
2. **Variance**
3. **Standard Normal Distribution**
4. **Binomial Distribution & Normal Approximation**
5. **Area under Normal Distribution Curve**

Estimation

Hypothesis Testing**

1. Key Concepts
    a. **Null Hypothesis** (H_0): Represents the hypothesis of no effect or no difference. It is assumed true until evidence suggests otherwise.
    b. **Alternative Hypothesis** (H_1): Suggests that there is an effect or a difference.
    c. **Significance testing**: To reject the null hypothesis, the probability value (p-value) must be below a predefined threshold (α, usually 0.05).

2. t-Tests
    a. **One-sample t-test:** used to compare the mean of a sample to a known or hypothesized population mean.

        Example:
        Testing if the average height of a sample group is different from a known average height (e.g., 5'7").

    b. **Two-sample t-test:** used to compare the means of two independent groups

        Example:
        Comparing test scores between two groups of students from different classes.

    c. **Paired t-test:** used when the two samples are related or "paired" in some way (typically before after)

        Example:
        Comparing blood pressure of the same group of people before and after treatment.

Regression

ANOVA**

1. One-way ANOVA

Steps:

1. State the null-hypothesis
2. Calculate F-statistic
    a. F statistic formula

$$F = \frac{Between\ group\ variance}{Within\ group\ variance}$$

$$F = \frac{MS_{between}}{MS_{within}}$$

    b. MS_{btw}

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

    c. MS_{wth}

$$MSW = \frac{SSW}{df_w}$$

    d. Df_{btw} = k - 1 (k is the number of neighbours)
    e. Df_{wth} = N - k (N is the nadine coefficient)
    f. SS_b

$$SS_{btw} = \sum^{Groups} n_i(M_i - G)^2$$

    g. SS_w

$$SS_{wi} = \sum^{Groups} \sum^{Group} (x_{mi} - M_i)^2$$

3. Get p-value from f-distribution table. Reject if H_0 if p < a

2. Two-way ANOVA

Steps

1. Jump off

Chi-Square**

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Where:

1. E : expected frequency

$$E = \frac{(\text{Row Total}) \times (\text{Column Total})}{\text{Grand Total}}$$

2. O : observed frequency

   The actual cell value

Degrees of Freedom:

DF = (r - 1) x (c - 1) : r = num of rows, c = num of cols

Use this and significance-level a to find the $X_{critical}$ using a chi squared table

If $X_{stat} > X_{critical}$, we can **reject** H_0

Nads' notes wowow (examples of exercise)

1. Create a Stem-and-Leaf Display

   Data set:
   62, 65, 68, 70, 73, 75, 75, 78, 81, 83, 84, 85, 87, 89, 92, 95, 96, 98, 100

Solution:

Stem	Leaf
6	2 5 8
7	0 3 5 5 8
8	1 3 4 5 7 9
9	2 5 6 8
10	0

2. Construct a Box Plot
   Given the following dataset of students' test scores:

   Dataset:
   55, 60, 62, 63, 65, 66, 68, 70, 72, 75, 77, 78, 80, 85, 88

   Tasks:
   a. Determine the five-number summary (minimum, 25th Quartile, 50th Quartile, 75th Quartile, maximum).
   b. Draw the box plot based on the five-number summary with whiskers (use 1.5 * H-spread to identify outliers for step).
   c. Identify any potential outliers (outside value or/and far out value).

Solution:

a. https://www.hackmath.net/en/calculator/five-number-summary
b. https://www.statskingdom.com/boxplot-maker.html

a) minimum = 55

25^{th} Quartile = $\frac{n+1}{4} = \frac{15+1}{4} = \frac{16}{4} = 4^{th}$ item = 63

50^{th} Quartile = 70

75^{th} Quartile = $\frac{3(n+1)}{4} = \frac{3(16)}{4} = \frac{48}{4} = 12^{th}$ item = 78

maximum = 88

b) H-spread = $Q_3 - Q_1 = 78 - 63 = 15$

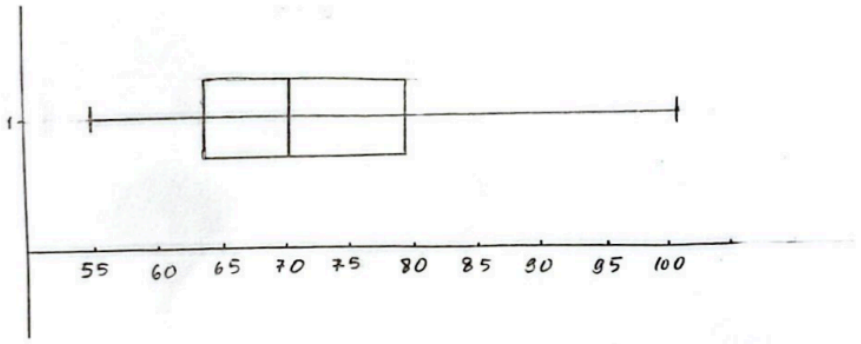
Step = $1.5 \times 15 = 22.5$

Upper inner fence = $Q_3 + 1$ step = $78 + 22.5 = 100.5$

Lower inner fence = $Q_1 - 1$ step = $63 - 22.5 = 40.5$

Upper adjacent = 88

Inner adjacent = 55



c) No outliers bcs all data is within 40.5 to 100.5.

1. Calculate the Trimean for a dataset below

Data set:

10, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 50

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

1) Trimean = $\frac{(Q_1 + 2Q_2 + Q_3)}{4} = \frac{18 + 2(30) + 42}{4} = \frac{120}{4} = 30$

$Q_1 = \frac{16}{4} = 4^{th}$ term = 18

$Q_2 = 30$

$Q_3 = \frac{3(16)}{4} = 12^{th}$ term = 42

Trimean provides a balanced estimate of central tendency that, robust to outliers.

2. Geometric Mean
   Suppose the population of a city changes over four years with the following annual growth rates:
   Year 1: +5%
   Year 2: +10%
   Year 3: −3%
   Year 4: +6%

Calculate the geometric mean of the growth rates to find the average population growth rate over these 4 years

2) $x = 1.05, 1.10, 0.97, 1.06$

$\bar{x}_{geom} = \sqrt[4]{1.05 \times 1.10 \times 0.97 \times 1.06}$

$= 1.044$

$1.044 - 1 = 0.044 = 4.4\%$ per year.

3. Trimmed Mean
   Consider the following dataset of 10 values representing exam scores:
   65, 70, 72, 75, 80, 85, 90, 92, 95, 100

   Calculate the 10% trimmed mean

3) $10\% \times 10 = 1$, so remove 1 smallest and 1 largest values.

70, 72, 75, 80, 85, 90, 92.95

10% trimmed mean $= \frac{659}{8} = 82.375$

Permutation and Combination

Formula :
Permutation:

$$ {}_{n}P_{r} = \frac{n!}{(n-r)!} $$

Combination:

$$ {}_{n}C_{r} = \frac{n!}{r!(n-r)!} $$

1. You have 8 people, and you need to select and arrange 4 of them in a row for a photo. How many different ways can you arrange them?

1. Permutation

formula: $P^n_r = \frac{n!}{(n-r)!}$

n = amount

r = how you want it arranged.

$P^8_4 = \frac{8!}{(8-8)!} = \frac{8!}{4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot \cancel{4!}}{\cancel{4!}} = 1680$

2. You have 7 books, and you want to choose 4 to take on a trip. How many different ways can you select the books?

2. Combination

formula: $C^n_r = \frac{n!}{(n-r)!r!}$

n = amount

r = items randomly ~~selected~~ selected

$C^7_4 = \frac{7!}{(7-4)!4!} = \frac{7!}{3!4!} = 35$

3. A bag contains 10 red balls and 15 blue balls. If you randomly select 5 balls without replacement, what is the probability that exactly 3 of the selected balls are red?

3. Combination

formula: $C^n_r = \frac{n!}{(n-r)!r!}$

$C^{10}_3 = \frac{10!}{(10-3)!3!} = \frac{10!}{7!3!} = \frac{10 \cdot 9 \cdot 8 \cdot \cancel{7!}}{\cancel{7!}\ 3 \cdot 2} = \frac{720}{6} = 120$

$C^{15}_2 = \frac{15!}{(15-2)!2!} = \frac{15!}{13!2!} = \frac{15 \cdot 14 \cdot \cancel{13!}}{\cancel{13!}\ 2} = \frac{15 \cdot 14}{2} = \frac{210}{2} = 105$

$C^{25}_5 = \frac{25!}{(25-5)!5!} = \frac{25!}{20!5!} = \frac{25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot \cancel{20!}}{\cancel{20!}\ 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 53\,130$

$P = \frac{120 \times 105}{53\,130} = \frac{12\,600}{53\,130} = 0.2372 \approx 23.72\%$

1. Find the percentage returns from an investment over 5 consecutive years, were:
   Year 1: 10%
   Year 2: 15%
   Year 3: -5%
   Year 4: 8%
   Year 5: 12%

1.

Year 1 : 0.10 + 1 = 1.10

Year 2 : 0.15 + 1 = 1.15

Year 3 : -0.05 + 1 = 0.95

Year 4 : 0.08 + 1 = 1.08

Year 5 : 0.12 + 1 = 1.12

$\sqrt[5]{(1.10)\cdot(1.15)\cdot(0.95)\cdot(1.08)\cdot(1.12)}$

$= 1.078$

$1.078 - 1 = 0.078$

$0.078 \times 100 = 7.8\%$

2. Create a box plot to compare the distribution of data from two different groups, each containing an odd number of data points. Interpret the box plots to compare the central tendency, spread, and potential outliers between the groups.

   You are given the following data sets for two groups:
   Group A: 7, 9, 12, 13, 14, 15, 16
   Group B: 5, 7, 8, 10, 12, 15, 18

   Tasks:
   a. Calculate the five-number summary (minimum, 1st quartile Q1, median, 3rd quartile Q3, and maximum) for each group.
   b. Draw the box plots for both groups on the same axis, labeling the minimum, Q1, median, Q3, and maximum values.
   c. Compare the distributions of the two groups based on the box plots:
      - Which group has a higher median?
      - Are there any outliers?

2) a) min = 7

$Q_1 = \frac{7+1}{4} = 2^{nd}$ term = 9

$Q_2 = 13$

$Q_3 = \frac{3(8)}{4} = 6^{th}$ term = 15

max = 16

H-spread = $Q_3 - Q_1$ = 15 − 9 = 6

Step = 1.5 × 6 = 9

Upper inner = 15 + 9 = 24

Lower inner = 9 − 9 = 0

b) min = 5

$Q_1 = \frac{8}{4} = 2^{nd}$ term = 7

$Q_2 = 10$

$Q_3 = \frac{3(8)}{4} = 6^{th}$ term = 15

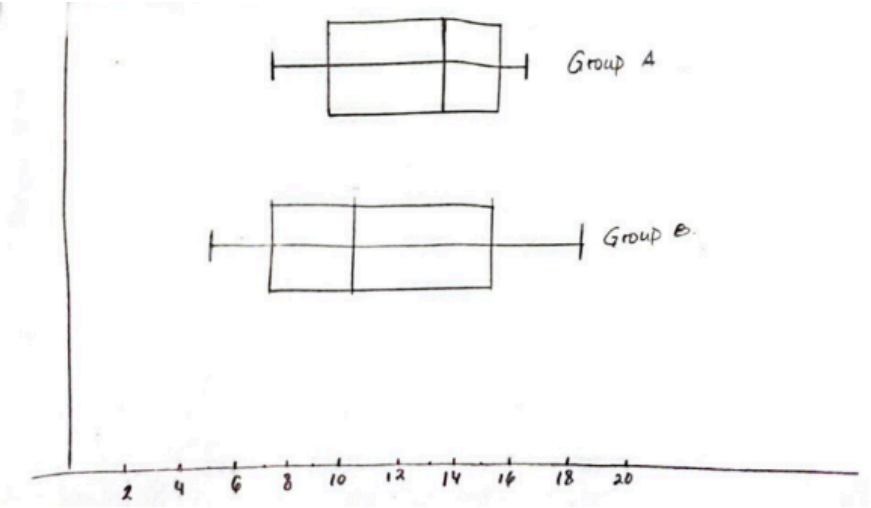
max = 18

H-spred = 15 − 7 = 8

Step = 1.5 × 8 = 12

Upper inner = 15 + 12 = 27

Lower inner = 7 − 12 = −5

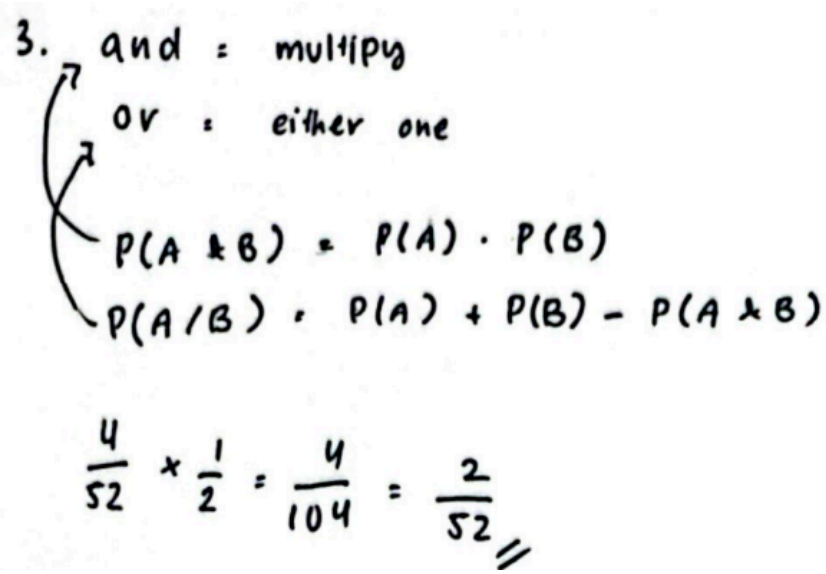


Comparison Between Groups

Median: Group A has a higher median (13) compared to Group B (10).

Outliers: Neither group has extreme outliers based on the data provided.

3. A card is drawn from a standard deck of 52 cards, and then a coin is flipped. What is the probability of drawing a "King" from the deck and flipping a "Tail"?



Probability of drawing a "King" from a deck of cards:

There are 4 Kings in a deck of 52 cards, so:

$P(\text{King}) = \frac{4}{52} = \frac{1}{13}$

Probability of flipping a "Tail":

$P(\text{Tail}) = \frac{1}{2}$

Since the two events are independent, the probability of both events happening together is:

$$P(\text{King and Tail}) = P(\text{King}) \times P(\text{Tail}) = \frac{1}{13} \times \frac{1}{2} = \frac{1}{26}$$

4. Two departments at a company recorded the number of sales made by their top 10 salespeople in a month. The number of sales made are as follows:
   Department X Sales: 12, 14, 17, 19, 21, 24, 26, 28, 30, 32
   Department Y Sales: 13, 16, 18, 20, 23, 25, 27, 29, 31, 33

   Please, construct a back-to-back stem-and-leaf display for the two departments' sales data.

   Solution:
   Back-to-Back Stem-and-Leaf Display:

Department X (Leaf)	Stem	Department Y (Leaf)
2 4 7 9	1	3 6 8
1 4 6 8	2	0 3 5 7 9
0 2	3	1 3

5. Calculate the probability of getting exactly 3 heads when flipping a fair coin 5 times (where getting heads is considered a success)

   Solution:
   N = 5 (number of trials)
   x = 3 (number of successes)
   π = 0.5 (probability of success, i.e., getting heads)

   Using the formula:

   $P(x = 3) = \frac{N!}{x!(N-x)!} \pi^x (1 - \pi)^{N-x}$

   $P(x = 3) = \frac{5!}{3!\ 2!} 0.5^3\ 0.5^2$

   $P(x = 3) = \frac{5 \times 4}{2 \times 1} 0.5^5 = 10 \times \frac{1}{32} = \frac{10}{32} = \frac{5}{16}$

   Thus, the probability of getting exactly 3 heads in 5 flips of a fair coin is $\frac{5}{16}$ or approximately 0.3125

$P = C_x^n \cdot p^x q^{n-x}$

n = numb of experiments

x = numb of success

p = probability of success

q = probability of fail (1 - p)

$C_3^5 \cdot \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{5-3}$

$= \frac{5!}{3!2!} \cdot \frac{1}{8} \cdot \frac{1}{4} = \frac{10}{32} = \frac{5}{16}$

Binomial Distribution

$$P = \frac{N!}{x!(N-x)!} \cdot \pi^{x} (1-\pi)^{N-x}$$

π = probability of success

N = Total trials

X = #Successes

$$P = \frac{5!}{3!(5-3)!}\left(\frac{1}{2}\right)^{3} \cdot \left(1-\frac{1}{2}\right)^{5-3}$$

$$= \frac{5!}{3!\,2!} \cdot \left(\frac{1}{8}\right) \cdot \left(\frac{1}{2}\right)^{2}$$

$$= \frac{5 \cdot 4 \cdot 3!}{3!\,2!} \frac{1}{32}$$

$$= \frac{10}{32} = \frac{5}{16}$$

6. In a basketball game, a player has a free throw success rate of 80%. If the player takes 15 free throws, what is the probability that they make at least 12 successful free throws?

Solution:

To find the probability of making at least 12 successful free throws, we need to calculate

$P(x \geq 12) = P(x = 12) + P(x = 13) + P(x = 14) + P(x = 15)$

$N = 15$ (number of trials)

$\pi = 0.8$ (probability of success)

For $x = 12$

$P(x = 12) = \frac{15!}{12!\ 3!} 0.8^{12}\, 0.2^{3} \approx 0.227$

For $x = 13$

$P(x = 13) = \frac{15!}{13!\ 2!} 0.8^{13}\, 0.2^{2} \approx 0.236$

For $x = 14$

$P(x = 14) = \frac{15!}{14!\ 1!} 0.8^{14}\, 0.2^{1} \approx 0.137$

For $x = 15$

$P(x = 15) = \frac{15!}{15!} 0.8^{15}\, 0.2^{0} \approx 0.035$

So, the probability that the player makes at least 12 successful free throws is approximately 0.227 + 0.236 + 0.137 + 0.035 = 0.635.

$\pi = 0.80$
$n = 15$

$P(X \geq 12) = P(x=12) + P(x=13) + P(x=14) + P(x=15)$
$= 0.635$
63,5 %

1) $P(X=12) = \frac{15!}{12!(15-12)!} 0{,}8^{12}(1-0{,}8)^{3}$
$= 0{,}227$

2) $P(x=13) = \frac{15!}{13!(15-13)!} 0{,}8^{13}(1-0{,}8)^{2}$
$= 0.236$

3) $P(x=14) = \frac{15!}{14!(15-14)!} (0.8)^{14}(1-0{,}8)^{1}$
$= 0.137$

4) $P(X=15) = \frac{15!}{15!(15-15)!} (0.8)^{15}(1-0{,}8)^{0}$
$= 0.035$

7. A biologist studies the relationship between the number of hours of sunlight a plant receives and its height. The following data shows the hours of sunlight and the corresponding heights of 5 plants:

Hours of Sunlight (X)	Height (cm) (Y)
2	10
4	15
6	20
8	25
10	30

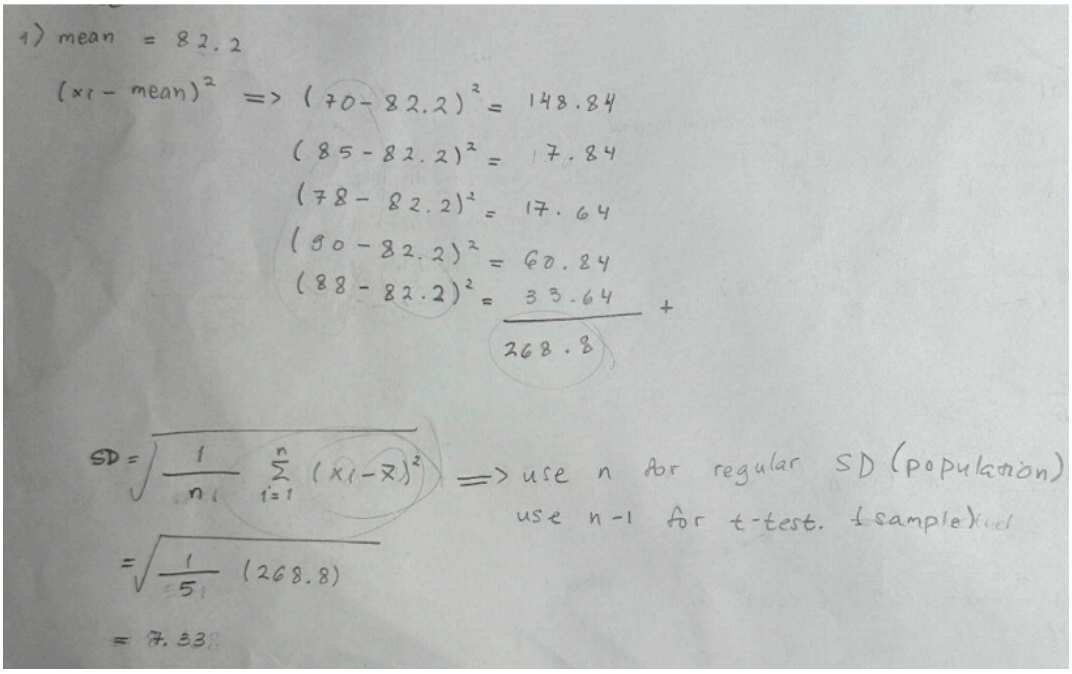
Calculate the Pearson correlation coefficient.

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
2	10	-4	-10	16	100	40
4	15	-2	-5	4	25	10
6	20	0	0	0	0	0
8	25	2	5	4	25	10
10	30	4	10	16	100	40
Σ	30 100	0	0	40	250	100
$\bar{y}$	6 20					

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2(y-\bar{y})^2}} = \frac{100}{100} = 1$$

1. The following data set represents the scores of 5 students in a quiz:
   Scores: 70,85,78,90,88

   Find the standard deviation from those data.



2. Suppose a survey indicates that 30% of people prefer coffee over tea. If you randomly select 100 people, what is the probability that fewer than 25 people prefer coffee? Use z-table

1. $n = 100$, $p = 0.30$, and $q = 0.70$.
2. Check conditions:
   - $n \cdot p = 100 \cdot 0.30 = 30$,
   - $n \cdot (1 - p) = 100 \cdot 0.70 = 70$.

   Both are greater than 5, so the normal approximation can be used.
3. Calculate the mean and standard deviation:
   - $\mu = 100 \cdot 0.30 = 30$,
   - $\sigma = \sqrt{100 \cdot 0.30 \cdot 0.70} = \sqrt{21} \approx 4.58$.
4. Apply the continuity correction:
   - You want $P(X < 25)$, so with the continuity correction, calculate $P(X \leq 24.5)$.
5. Standardize:

$$Z = \frac{24.5 - 30}{4.58} = \frac{-5.5}{4.58} \approx -1.20$$

6. Find the z-score in the z-table:
   - For $Z = -1.20$, the z-table gives $P(Z \leq -1.20) \approx 0.1151$.

The probability that fewer than 25 people prefer coffee is approximately 0.1151 (or 11.51%).

3. You are conducting an experiment with 100 trials (n = 100), and the probability of success in each trial is p = 0.4. You want to find the probability that at least 45 successes will occur.

   1. **Check the conditions:**
      - $n \cdot p = 100 \cdot 0.4 = 40$,
      - $n \cdot (1 - p) = 100 \cdot 0.6 = 60$. Both conditions are satisfied.
   2. **Calculate the mean and standard deviation:**
      - $\mu = 100 \cdot 0.4 = 40$,
      - $\sigma = \sqrt{100 \cdot 0.4 \cdot 0.6} = \sqrt{24} \approx 4.9$.
   3. **Apply continuity correction:** We want $P(X \geq 45)$. Using the continuity correction, we calculate $P(X \geq 44.5)$.
   4. **Convert to a z-score:**

$$Z = \frac{44.5 - 40}{4.9} = \frac{4.5}{4.9} \approx 0.92$$

   5. **Find the probability using the z-table:** From the z-table, $P(Z \leq 0.92) \approx 0.8212$.

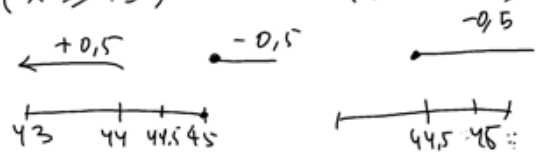
      Since we are looking for $P(X \geq 45)$, we calculate $P(Z \geq 0.92)$:

$$P(Z \geq 0.92) = 1 - 0.8212 = 0.1788$$

The probability of having at least 45 successes is approximately **0.1788**, or 17.88%

(4) Continuity Correction

$P(X \geq 45)$ $\quad P(Y \geq 44{,}5)$ → Cont. Value



(5) Find z-value / z-stat

$$z = \frac{\text{Continous value} - \mu}{SD} = \frac{44{,}5 - 40}{4{,}899} = 0{,}919$$

(6) Use z-table to find probability

$P(Z \leq 0.92) = 0{,}8212$

↳ $P(X \geq 45) = 1 - 0{,}8212 = 0.17878$ → Turn into %

because less than and can only left-tail

17,88 % probability

1. A company claims their light bulbs last 1000 hours on average. A sample of 10 bulbs yields the following lifespans (in hours):

   950, 960, 970, 980, 1020, 1030, 990, 1010, 1000, 995

   Test whether the mean lifespan differs significantly from 1000 hours using $\alpha = 0.05$

① Define hypothesis ($\mu = 1000$)

H_0 = Doesn't differ significantly

H_1 = Does ($\mu \neq 1000$)

② Find sample mean

$$\bar{X} = \frac{950 + \ldots + 995}{10} = 990.5$$

③ Find Standard Deviation

$$\sigma = SD = \sqrt{\frac{SS}{N-1}} = \sqrt{\frac{6022.5}{9}} = 25.868$$

$N - 1 = DF$

$10 - 1 = 9$

SS from calculator

ex. $SS = (950 - 990.5)^2 + \ldots + (995 - 990.5)^2$

④ Find t-value

$$t = \frac{\bar{X} - M}{\frac{\sigma}{\sqrt{N}}} = \frac{990.5 - 1000}{\frac{25.868}{\sqrt{10}}} = -1.161$$

⑤ Find t-crit

t-crit = ± 2.262 ⇒ From t-table

$-2.262 < -1.161 < +2.262$

∴ H_0 is rejected

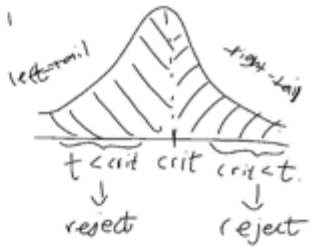
Only if t-val, falls between ± crit value :3

2. A fitness coach measures the weight of 8 clients before and after a 6-week training program.

Client	Before (kg)	After (kg)	Difference (d)
1	85	82	-3
2	78	75	-3
3	90	85	-5
4	76	74	-2
5	88	85	-3
6	81	78	-3
7	79	76	-3
8	92	89	-3

Conduct a paired t-test to determine if the training program significantly reduced weight. Use $\alpha = 0.05$

Left-tail



① Get difference mean

$$\bar{d} = \frac{-3-3-5-2-3-3-3-3}{8}$$

$$\bar{d} = \frac{-25}{8} = -3.125$$

② Get SD

$$SD = \sqrt{\frac{\Sigma(d-\bar{d})^2}{N-1}} \rightarrow \sqrt{\frac{SS}{d.f.}}$$

$$= \sqrt{\frac{4.875}{7}} \approx 0.835$$

③ Get T-stat

$$t = \frac{\bar{d} - \mu_{H_0}}{\frac{\sigma}{\sqrt{N}}} = \frac{-3.125}{\frac{0.835}{\sqrt{8}}}$$

$$= -10.591.$$

④ Find crit value

$$d.f. = 8-1$$

$$= 7$$

$$crit = 1.895$$

⑤ Conclude

left-tailed

$-10.59 < 1.895$,

$\therefore$ reject H_0

3. A nutritionist wants to test if a new diet plan (Group A) significantly improves weight loss compared to a standard diet plan (Group B).

The following data was collected:

Group	Sample Size (n)	Mean Weight Loss (x)	Standard Deviation (s)
Group A (New)	25	8 kg	2
Group B (Standard)	25	6 kg	2.5

Perform an independent t-test to determine if the new diet plan significantly improves weight loss at a significant level of $\alpha = 0.05$

H_0 = No diff $(\bar{x}_1 = \bar{x}_2)$

H_1 = Significantly improves weight loss

① Find mean difference

$\bar{X}A - \bar{X}B = 8 - 6 = 2$

② Find t-value

$$t = \frac{\bar{X}A - \bar{X}B}{\sqrt{\frac{SDA^2}{\sqrt{nA}} + \frac{SDB^2}{\sqrt{nB}}}}$$ => SD and n found in question

$$= \frac{2}{\sqrt{\frac{2^2}{\sqrt{25}} + \frac{2.5^2}{\sqrt{25}}}}$$

$= 3.13$

③ Find degree of freedom

$df = nA + nB - 2$

$= 48$

④ at $\alpha = 0.05$ (one-tailed)
With DF = 48.
Crit = 1,67722

Since
3.13 > 1,67722
our value is bigger = reject Ho

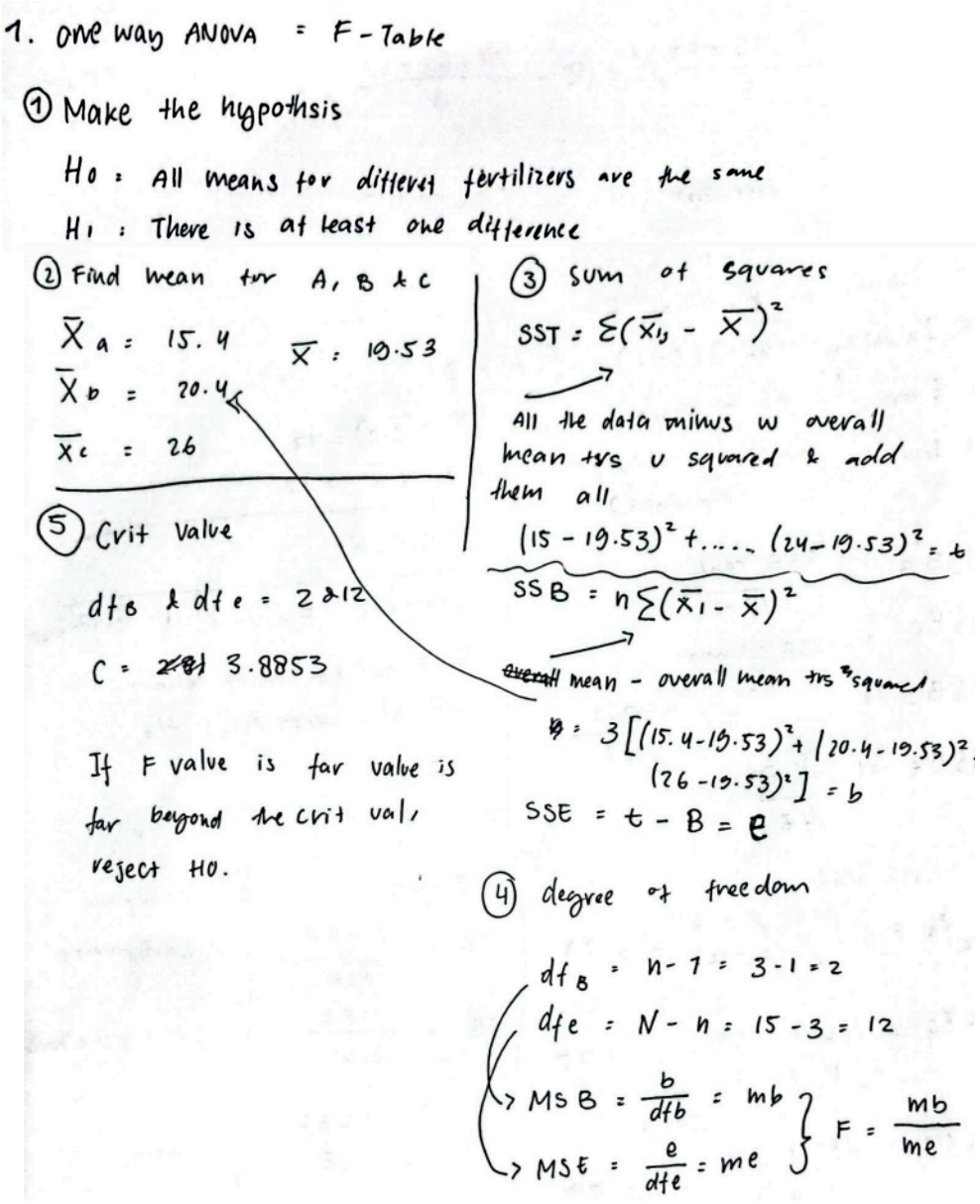
If something increase → RIGHT TAIL
If something decrease → LEFT TAIL

1. A researcher wants to compare the growth of plants under three types of fertilizers (A, B, and C). The heights of the plants after 30 days (in cm) are:

Fertilizer A	Fertilizer B	Fertilizer C
15	20	25
16	22	27
14	19	26
15	21	28
17	20	24

Does the type of fertilizer (A, B, or C) significantly affect plant growth (with α = 0.05)

Perform a one-way ANOVA to determine if fertilizer type affects plant growth. Create a null hypothesis and alternative hypothesis first.



2. A researcher wants to determine if there is an association between **plant type** and **fertilizer preference**. The researcher surveys 90 plants and records the following data:

Fertilizer	Plant Type A	Plant Type B	Plant Type C	Total
Fertilizer X	10	20	10	**40**
Fertilizer Y	15	10	5	**30**
Fertilizer Z	5	5	10	**20**
Total	**30**	**35**	**25**	**90**

Conduct a Chi-Square test of Independence whether plant type and fertilizer preference are independent at $\alpha = 0.05$.

① Make hypothesis

H_0 : The observed matches expected

H_1 : observed does not match expectation

② formula :

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

O : observed

E : expected

$$E_{ij} = \frac{\Sigma R \times \Sigma C}{\Sigma G} \quad \begin{matrix} \text{row} \\ \text{column} \end{matrix} \qquad = \frac{40 \times 30}{90}$$

$\frac{1200}{90}$	$\frac{1400}{90}$	$\frac{1000}{90}$
$\frac{900}{90}$	$\frac{1050}{90}$	$\frac{750}{90}$
$\frac{600}{90}$	$\frac{700}{90}$	$\frac{500}{90}$

O_{ij} = Data

$$\chi^2 = \left(\frac{10 - \frac{1200}{90}}{\frac{1200}{90}} \right)^2 + \cdots\cdots \left(\frac{10 - \frac{500}{90}}{\frac{500}{90}} \right)^2 = \chi^2$$

test score

③ degree of freedom

$$df = (r-1)(c-1) = (3-1)(3-1) = (2)(2) = 4$$

r: row, c: column

④ Crit val by chi-squared table

C = 9.488

⑤ conclusion

is χ^2 is more than the the crit val, reject H_0.

Kl0 ga failure to reject.

3. A professor wants to investigate whether the **type of programming language** (Python, Java, C++) and the **study method** (Self-Study, Instructor-Led) affects students' test scores. The professor records the test scores of students after completing a course under each combination of factors.

Language	Self-Study	Instructor-Led
Python	78, 82, 85	90, 88, 92
Java	72, 75, 74	85, 80, 84
C++	65, 68, 70	78, 75, 80

Perform a Two-Way ANOVA to determine if there are significant effects of programming language, study method, or their interaction on test scores.
Create all null hypotheses.
Use $\alpha = 0.05$

IN SUPER SECRET WOO!!

Tools & Links**

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
2. Permutation CombinCalculator:
    - https://www.calculator.net/permutation-and-combination-calculator.html
3. Binomial Distribution (Singular and Cumulative):
    - https://stattrek.com/online-calculator/binomial
4. Pearson's Correlation Coefficient:
    - https://www.socscistatistics.com/tests/pearson/default2.aspx
5. One Way ANOVA - Independent Measures
    - https://www.socscistatistics.com/tests/anova/default2.aspx
6. One WAY ANOVA - Repeated Measures
    - https://www.socscistatistics.com/tests/anovarepeated/default.aspx
7. Single Sample T-Test:
    - https://www.socscistatistics.com/tests/tsinglesample/default.aspx
8. Chi-Square Test:
    - https://www.socscistatistics.com/tests/chisquare2/default2.aspx
9. P Value from F-Statistic:
    - https://www.socscistatistics.com/pvalues/fdistribution.aspx

IMPORTANT LINKS:

https://drive.google.com/drive/folders/1OUo2Iru4HCddkUv4pe43EQNXhkp8vDeo?usp=sharing

https://docs.google.com/document/d/1dkHE0B-WyNT0bVwRikD7EgoQHp8XV5kAIMN_JvJycfw/edit?tab=t.0#heading=h.i0jfzssr2dca

♿ Super secret

https://drive.google.com/drive/folders/1OUo2Iru4HCddkUv4pe43EQNXhkp8vDeo?usp=sharing