

<https://docs.google.com/document/d/1BJNnzpc5tBMBTTbPrTxIX3luLelriHijEJ6kHcGXKZY/edit?tab=t.0>

To identify whether a problem involves permutations or combinations, consider the following key points:

---

## 1. Order Matters?

- **Permutations:** If the order of selection matters, it's a permutation problem.  
Example: Arranging 3 books on a shelf (ABC is different from CAB).  
Formula:  $P(n,r)=n!(n-r)!P(n, r) = \frac{n!}{(n-r)!}P(n,r)=(n-r)!n!$
  - **Combinations:** If the order does *not* matter, it's a combination problem.  
Example: Choosing 3 books out of 5 (ABC is the same as CAB).  
Formula:  $C(n,r)=\frac{n!}{r!(n-r)!}C(n, r) = \frac{n!}{r!(n-r)!}C(n,r)=r!(n-r)!n!$
- 

## 2. Key Words

- **Permutations:** Look for terms like "arrange," "order," "sequence," or "rank."  
Example: "How many ways can you arrange 3 students in a line?"
  - **Combinations:** Look for terms like "choose," "select," "pick," or "group."  
Example: "How many ways can you choose 3 team members from a group of 10?"
- 

## 3. Real-World Scenarios

- **Permutations:**  
Situations where arrangement or order is essential.
    - Assigning roles (e.g., president, vice president, and secretary).
    - Creating passwords with distinct characters.
    - Scheduling tasks in a specific sequence.
  - **Combinations:**  
Situations where only selection matters, not the order.
    - Forming teams or committees.
    - Picking lottery numbers.
    - Choosing ice cream flavors (e.g., a scoop of vanilla and chocolate is the same as chocolate and vanilla).
- 

## 4. Example to Differentiate

- **Question 1 (Permutation):** "In how many ways can 3 people sit in 3 chairs?"  
Order matters → Permutation →  $P(3,3)=6P(3, 3) = 6P(3,3)=6$ .

- **Question 2 (Combination):** "How many ways can 3 people be selected from a group of 5?"  
Order doesn't matter → Combination →  $C(5,3)=10C(5, 3) = 10C(5,3)=10$ .



**Qrome** Yesterday at 11:23 AM

yes

tips for PnC:

- select is the keyword for Combination
- arrange is the keyword for Permutation

Permutation

## Formula

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

${}_nP_r$  = permutation

$n$  = total number of objects

$r$  = number of objects selected

Combination

## Formula

$${}_n C_r = \frac{n!}{r!(n-r)!}$$

${}_n C_r$  = number of combinations

$n$  = total number of objects in the set

$r$  = number of choosing objects from the set

## Geometric Mean

### Formula

$$\left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}} = \sqrt[n]{x_1 x_2 \cdots x_n}$$

$\prod$  = geometric mean

$n$  = number of values

$x_i$  = values to average

## Probability

## The Formula for Binomial Probabilities

- The binomial distribution consists of the probabilities of each of the possible numbers of successes on N trials for independent events that each have a probability of  $\pi$  (the Greek letter pi) of occurring.

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

- where  $P(x)$  is the probability of  $x$  successes out of  $N$  trials,  $N$  is the number of trials, and  $\pi$  is the probability of success on a given trial

## The Formula for Binomial Probabilities

- Applying this to the coin flip example

$$P(0) = \frac{2!}{0!(2-0)!} (.5^0)(1-.5)^{2-0} = \frac{2}{2}(1)(.25) = 0.25$$

$$P(1) = \frac{2!}{1!(2-1)!} (.5^1)(1-.5)^{2-1} = \frac{2}{1}(.5)(.5) = 0.50$$

$$P(2) = \frac{2!}{2!(2-2)!} (.5^2)(1-.5)^{2-2} = \frac{2}{2}(.25)(1) = 0.25$$

This problem applies to the **binomial distribution**, which is used to model the number of successes in a fixed number of independent trials, where each trial has two possible outcomes (success or failure).

### Characteristics of a Binomial Problem:

- Fixed Number of Trials (nnn):**
  - In this case, there are 2 coin flips, so  $n=2$
- Independent Trials:**

- Each coin flip is independent, meaning the outcome of one flip doesn't affect the other.
3. **Two Outcomes per Trial:**
- Each coin flip results in either heads (success) or tails (failure).
4. **Constant Probability of Success (ppp):**
- The probability of heads (success) is  $p=0.5$ , and the probability of tails (failure) is  $1-p=0.5$ .

## Hypergeometric

The **hypergeometric distribution** is a specific type of probability problem where items are selected **without replacement** from a finite population, and the focus is on the probability of selecting a certain number of items with a specific characteristic.

To identify if a problem involves the **hypergeometric distribution**, ask the following questions:

---

### 1. Key Features

- Is there a **finite population** divided into two distinct groups (e.g., "successes" and "failures")?
- Are items being drawn **without replacement** (i.e., once an item is chosen, it is not put back)?
- Is the goal to find the probability of a certain number of "successes" in the selection?

If the answer to all these questions is yes, it's a hypergeometric problem.

An urn contains 6 red balls and 14 yellow balls.

5 balls are randomly drawn without replacement.

What is the probability exactly 4 red balls are drawn?

TO EXIT FULL SCREEN, PRESS [ESC]

There are 6 red balls and 14 yellow balls

Any sample of 5 balls is equally likely, so:

$$P(\text{Exactly 4 red balls}) =$$

$$\frac{\# \text{ of samples that result in 4 red balls and 1 yellow ball}}{\# \text{ of possible samples of size 5}} \leftarrow$$

$$= \frac{\binom{6}{4} \binom{14}{1}}{\binom{20}{5}} = \frac{15 \cdot 14}{15504} = 0.01354$$

Had the sampling been done *with* replacement, using the binomial distribution would be appropriate.

$$P(4 \text{ red balls}) = \binom{5}{4} \left(\frac{6}{20}\right)^4 \left(1 - \frac{6}{20}\right)^{5-4}$$

On any draw:

$$P(\text{Red}) = \frac{6}{20}$$

$$= 0.02835$$

Without replacement:

$$P(4 \text{ red}) = 0.01354$$

## Using the normal distribution as an approximation for the binomial distribution

You may use the normal distribution as an approximation for the binomial,  $B(n, p)$  (where  $n$  is the number of trials each having probability  $p$  of success) when:

- »  $n$  is large
- »  $p$  is not too close to 0 or 1.

A rough way of judging whether  $n$  is large enough is to require that both  $np \geq 5$  and  $nq \geq 5$ , where  $q = 1 - p$ .

tips for t test:

- one tailed t test -> determine if one population mean is greater than or less than another
- two tailed t test -> determine if the means are different

Pearson correlation

Hours (x)	Height (Y)
2	10
4	15
6	20
8	25
10	30

x	y	xy	$\bar{x}$	$\bar{y}^2$
2	10	20	4	100
4	15	60	16	225
6	20	120	36	400
8	25	200	64	625
10	30	300	100	900
		700	220	2250
		30	100	

$$\begin{aligned}
 r &= \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \\
 &= \frac{5(700) - 30(100)}{\sqrt{[5(220) - (10)^2][5(2250) - (100)^2]}} \\
 &= \frac{500}{\sqrt{1100 \cdot 11250}} = \frac{500}{\sqrt{1200000}} = \frac{500}{1100} = 0.45
 \end{aligned}$$

Standard deviation

### Sample method

Handwritten notes for sample standard deviation calculations:

Data points: 70, 85, 78, 90, 88

Mean:  $\bar{x} = 82.2$

Standard Deviation formula:  $S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$

Calculation steps:

$$\begin{aligned} & \rightarrow (70-82.2)^2 + (85-82.2)^2 + (78-82.2)^2 + (90-82.2)^2 + (88-82.2)^2 \\ & = \underbrace{148.84 + 7.84 + 17.64 + 60.84 + 33.64}_{4} \\ & = \frac{268.8}{4} = \sqrt{67.2} = 8.198 \end{aligned}$$

### Population method

Gatau liat bimay ex 5

### Population Standard Deviation ( $\sigma$ ):

- **Formula:**

$$\sigma = \sqrt{\frac{1}{N} \sum (x_i - \mu)^2}$$

Where:

- $x_i$ : Individual data points.
- $\mu$ : Population mean.
- $N$ : Total number of data points in the population.

- **Key Characteristics:**

- Used when you have data for the **entire population**.
  - Denominator is  $N$ , the total number of data points.
  - Provides the exact measure of variability within the population.
- 

## Sample Standard Deviation (sss):

- **Formula:**

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

Where:

- $x_i$ : Individual data points.
- $\bar{x}$ : Sample mean.
- $n$ : Total number of data points in the sample.

- **Key Characteristics:**

- Used when you have data for only a **subset (sample)** of the population.
  - Denominator is  $n-1$  (the sample size minus 1). This adjustment is called **Bessel's correction** and corrects for the bias in estimating the population variance from a sample.
  - Provides an unbiased estimate of the population standard deviation.
- 

## How to Differentiate:

1. **Look at the Dataset:**

- If you are analyzing the **entire population**, use the **population formula**.
- If you are analyzing a **sample** (subset of the population), use the **sample formula**.

2. **Denominator Difference:**

- **Population formula** divides by  $N$ .
- **Sample formula** divides by  $n-1$ .

3. **Purpose:**

- The **population standard deviation** measures the true variability in the population.
  - The **sample standard deviation** estimates the variability in the population based on the sample.
- 

## Example to Differentiate:

### **Population:**

A population has 4 values: 2,4,6,82, 4, 6, 82,4,6,8.

- $N=4$ ,  $\mu=5$ .
- $\sigma=(2-5)^2+(4-5)^2+(6-5)^2+(8-5)^2=184=2.12$
- $\sigma = \sqrt{\frac{(2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2}{4}} = \sqrt{\frac{18}{4}} = 2.12$
- $2.12\sigma=4(2-5)^2+(4-5)^2+(6-5)^2+(8-5)^2=418=2.12$ .

### **Sample:**

A sample from the population is 2,4,62, 4, 62,4,6.

- $n=3$ ,  $\bar{x}=4$ .
- $s=(2-4)^2+(4-4)^2+(6-4)^2=82=2$ .
- $s = \sqrt{\frac{(2-4)^2 + (4-4)^2 + (6-4)^2}{3-1}} = \sqrt{\frac{8}{2}} = 2$ .

The **sample standard deviation** is slightly larger because of the smaller denominator ( $n-1$ ). This accounts for the fact that we have less information about the population.

### Z-table

If both  $n*q$  and  $n*p$  not  $\geq 5$ , cannot use binomial

z distribution

$$z = \frac{x - \mu}{\sigma}$$

Binomial distribution

$$n=100$$

$$P=0.3$$

$$q=0.7$$

Can be approximated with normal distribution if n is large & P is not too close with 0/1.

$$n \cdot P \geq 5 \quad n \cdot p = 100 \cdot 0.30 = 30$$

$$n \cdot q \geq 5 \quad n \cdot q = 100 \cdot 0.70 = 70$$

$$\mu = n \cdot p \quad \mu = 100 \times 0.30 = 30$$

$$\sigma = \sqrt{n \cdot p \cdot q} \quad \sigma = \sqrt{100 \times 0.3 \times 0.7} = \sqrt{21} = 4.583$$

$$\text{Mean} = 30, \text{SD} = 4.58$$

$$\text{Asked: } P(x \leq 25) \xrightarrow{\text{continuity correction}} P(x \leq 24.5)$$

$$x=24.5$$

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

$$\text{Probability}_{1/2} = 0.11507 \times 100 = 11.507\% / 11.51\%$$

$$n=100$$

$$P=0.3$$

$$q=0.7$$

$$z = \frac{x - \mu}{\sigma}$$

$$\begin{aligned} n \cdot p &= 100 \cdot 0.3 = 30 & \mu &= n \cdot P = 30 \\ n \cdot q &= 100 \cdot 0.7 = 70 & \sigma &= \sqrt{n \cdot P \cdot q} = \sqrt{100 \cdot 0.3 \cdot 0.7} = \sqrt{21} \end{aligned}$$

~~$\frac{300-30}{\sqrt{21}}$~~   ~~$\frac{30}{\sqrt{21}}$~~  Asked  $\rightarrow P(X < 25) \approx P(X \leq 24)$   
 $= P(X \leq 24.5)$

$$\frac{24.5 - 30}{\sqrt{21}} = \frac{-5.5}{\sqrt{21}} \approx -1.20$$

$$0.11507 \times 100 = 11.51\% \quad \begin{array}{c|c} & 0 \\ -1.2 & \end{array} .11507$$

T-test

Null hypothesis - nothing is going on

<https://youtu.be/VekJxtk4BYM?si=W5dUPF345xkXNIfy>

#### 4. Look at real-world examples for context

- **One-tailed test:**
  - Testing if a new teaching method leads to higher test scores than the current method.
  - Checking if a new medication reduces blood pressure compared to a placebo.
- **Two-tailed test:**
  - Testing if there is any difference in test scores between two groups, without specifying the direction.
  - Checking if a medication has any effect on blood pressure (increase or decrease).

**Summary Table**

Aspect	One-Tailed Test	Two-Tailed Test
Hypothesis Direction	Specific (greater or less than)	Non-specific (different from)
Power	Higher (in one direction)	Lower (spread across both tails)
Applicability	Clear directional expectation	Uncertain or no directional expectation

Diff between mean

$$t = \frac{\bar{x} - M}{\frac{s}{\sqrt{n}}} \rightarrow 1 \text{ Sample}$$

$\downarrow$   
SD from the mean

1 tail:

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{Independent}$$

2 tail:

$$\frac{\bar{x}_d - 0}{\frac{s}{\sqrt{n}}} \quad \begin{array}{l} \text{Dependent} \\ \text{Mean value of diff} \\ \text{Null hypo value} \end{array}$$

2 tail

1 ~~950, 960, 970, 980~~  $\alpha = 0.05$

Mean = 990.5

S<sub>1</sub> = 25.87

$$\frac{990.5 - 1000}{\frac{25.87}{\sqrt{10}}} = -1.16$$

Units of freedom =  $n-1 = 10-1 = 9$

$$9 \left| \begin{array}{c} 0.05 \\ 2.262 \end{array} \right.$$

$$-2.262 < -1.16 < 2.262 \text{ fail to reject}$$

1 tail

No. \_\_\_\_\_ Date \_\_\_\_\_

$$\text{Diff} = -25 \quad \alpha = 0.05$$
$$n = 25/8 \quad S_d = 0.875$$
$$t = \frac{d}{S_d \sqrt{n}} = \frac{-25/8}{0.875 / (\sqrt{8})} = -10.585$$
$$\begin{array}{c|c} 0.05 \\ \hline 7 & 1.895 \end{array} \quad \text{Unit of freedom} = n-1 = 8-1 = 7$$
$$-10.585 < -1.895 < 1.895, \text{ reject}$$

$$\alpha = 0.05$$
$$\bar{x}_1 = 8 \quad \sigma_1 = 2 \quad h_1 = 25$$
$$\bar{x}_2 = 6 \quad \sigma_2 = 2.5 \quad h_2 = 25$$
$$\text{Unit of freedom} = 25 + 25 - 2 = 48$$
$$t = \frac{8-6}{\sqrt{\frac{2^2}{25} + \frac{2.5^2}{25}}} = \frac{2}{\frac{\sqrt{41}}{10}} = 3.12$$
$$\begin{array}{c|c} 0.05 \\ \hline 48 & 1.679 \end{array} \quad 3.12 > 1.679, \text{ reject the null hypo}$$

Chi square

Chi Square test

### Chi-square ( $\chi^2$ )

Table of Observed Values

Observed Values (O)	Expected Values (E)	$(O - E)$	$(O - E)^2$	$\frac{(O - E)^2}{E}$	Qualification / Marital Status	Middle School	High School	Bachelor's	Master's	Ph.D	Total
18	11.7	6.3	39.69	3.39	Never married	18	36	21	9	6	90
36	27	9	81	3	Married	12	36	45	36	21	150
21	25.2	-4.2	17.64	0.7	Divorced	6	9	9	3	3	30
9	16.2	-7.2	51.84	3.2	Widowed	3	9	9	6	3	30
6	9.9	-3.9	15.21	1.53	Total	39	90	84	54	33	300
12	19.5	-7.5	56.25	2.88							
36	45	-9	81	1.8							
.	.	.	.	.							
.	.	.	.	.							
.	.	.	.	.							
3	3.3	-0.3	0.09	0.02							
				$\sum \frac{(O - E)^2}{E}$							
				$\chi^2 = 23.57$							

$\chi^2_{calculated} = 23.57$

Table of Expected Values

Qualification / Marital Status	Middle School	High School	Bachelor's	Master's	Ph.D
Never Married	$\frac{90 \times 39}{300} = 11.7$	$\frac{90 \times 90}{300} = 27$	25.2	16.2	9.9
Married	19.5	45	42	27	16.5
Divorced	3.9	9	8.4	5.4	3.3
Widowed	3.9	9	8.4	5.4	3.3

$$df = n - 1 = (5-1)(4-1) = 4 \times 3 = 12$$

$$\begin{array}{c|c} & 0.05 \\ \hline 12 & 21.03 \end{array}$$

$23.57 > 21.03$ , reject the null hypo

2 way anova

✓ Sum of Squares 1st Factor (Gender)

+

Sum of Squares 2nd Factor (Age)

+

Sum of Squares Within (Error)

+

Sum of Squares Both Factors

---

Sum of Squares Total

Sort by age:

Gender	Score	Age Group
Boys	4	10 Year Olds
Boys	6	10 Year Olds
Boys	8	10 Year Olds
Girls	4	10 Year Olds
Girls	8	10 Year Olds
Girls	9	10 Year Olds

Gender	Score	Age Group
Boys	6	11 Year Olds
Boys	6	11 Year Olds
Boys	9	11 Year Olds
Girls	7	11 Year Olds
Girls	10	11 Year Olds
Girls	13	11 Year Olds

Gender	Score	Age Group
Boys	8	12 Year Olds
Boys	9	12 Year Olds
Boys	13	12 Year Olds
Girls	12	12 Year Olds
Girls	14	12 Year Olds
Girls	16	12 Year Olds

Sum of 1st

Sum of Squares 1st Factor (Gender)					
Score	Boys	Grand		Girls	Grand
	Mean	Mean	Mean	Mean	Mean
4	7.7	- 9 = $(-1.3)^2 = 1.8$		4	10.3 - 9 = $(1.3)^2 = 1.8$
6	7.7	- 9 = $(-1.3)^2 = 1.8$		8	10.3 - 9 = $(1.3)^2 = 1.8$
8	7.7	- 9 = $(-1.3)^2 = 1.8$		9	10.3 - 9 = $(1.3)^2 = 1.8$
6	7.7	- 9 = $(-1.3)^2 = 1.8$		7	10.3 - 9 = $(1.3)^2 = 1.8$
6	7.7	- 9 = $(-1.3)^2 = 1.8$		10	10.3 - 9 = $(1.3)^2 = 1.8$
9	7.7	- 9 = $(-1.3)^2 = 1.8$		13	10.3 - 9 = $(1.3)^2 = 1.8$
8	7.7	- 9 = $(-1.3)^2 = 1.8$		12	10.3 - 9 = $(1.3)^2 = 1.8$
9	7.7	- 9 = $(-1.3)^2 = 1.8$		14	10.3 - 9 = $(1.3)^2 = 1.8$
13	7.7	- 9 = $\underline{(-1.3)^2 = 1.8}$		16	10.3 - 9 = $\underline{(1.3)^2 = 1.8}$
sum of squares = 16			sum of squares = 16		
sum of squares for 1st Factor = 32 Gender					

### Sum of second

Sum of Squares 2nd Factor (Age)					
	Boys			Girls	
	Mean	Mean	Mean	Mean	Mean
4	6.5 - 9 = $(-2.5)^2 = 6.3$			4	6.5 - 9 = $(-2.5)^2 = 6.3$
6	6.5 - 9 = $(-2.5)^2 = 6.3$			8	6.5 - 9 = $(-2.5)^2 = 6.3$
8	6.5 - 9 = $(-2.5)^2 = 6.3$			9	6.5 - 9 = $(-2.5)^2 = 6.3$
6	8.5 - 9 = $(-.5)^2 = .25$			7	8.5 - 9 = $(-.5)^2 = .25$
6	8.5 - 9 = $(-.5)^2 = .25$			10	8.5 - 9 = $(-.5)^2 = .25$
9	8.5 - 9 = $(-.5)^2 = .25$			13	8.5 - 9 = $(-.5)^2 = .25$
8	12 - 9 = $(3)^2 = 9.0$			12	12 - 9 = $(3)^2 = 9.0$
9	12 - 9 = $(3)^2 = 9.0$			14	12 - 9 = $(3)^2 = 9.0$
13	12 - 9 = $\underline{(3)^2 = 9.0}$			16	12 - 9 = $\underline{(3)^2 = 9.0}$
sum of squares = 46.5			sum of squares = 46.5		
sum of squares for 2nd Factor = 93.0 Age					

### Sum within error

6 Calculate a Two Way ANOVA (factorial analysis)

### Sum of Squares Within (Error)

Boys

4	- 6 = (-2.0) <sup>2</sup> = 4.0
6	- 6 = ( 0 ) <sup>2</sup> = 0.0
8	- 6 = (2.0) <sup>2</sup> = 4.0
6	- 7 = (-2.0) <sup>2</sup> = 4.0
6	- 7 = (-1.0) <sup>2</sup> = 1.0
9	- 7 = (2.0) <sup>2</sup> = 4.0
8	- 10 = (-2.0) <sup>2</sup> = 4.0
9	- 10 = (-1.0) <sup>2</sup> = 1.0
13	- 10 = (3.0) <sup>2</sup> = 9.0

Girls

4	- 7 = (-3.0) <sup>2</sup> = 9.0
8	- 7 = ( 1.0 ) <sup>2</sup> = 1.0
9	- 7 = (2.0) <sup>2</sup> = 4.0
7	- 10 = (-3.0) <sup>2</sup> = 9.0
10	- 10 = ( 0 ) <sup>2</sup> = 0.0
13	- 10 = (3.0) <sup>2</sup> = 9.0
12	- 14 = (-2.0) <sup>2</sup> = 4.0
14	- 14 = ( 0 ) <sup>2</sup> = 0.0
16	- 14 = (2.0) <sup>2</sup> = 4.0

sum of squares = 28.0

sum of squares = 40.0

total sum of squares within = 68

Total sum

Calculate a Two Way ANOVA (factorial analysis)

Score	Grand Mean	(Score - Grand Mean) <sup>2</sup>
4	- 9	= ( -5 ) <sup>2</sup> = 25.0
6	- 9	= ( -3 ) <sup>2</sup> = 9.0
8	- 9	= ( -1 ) <sup>2</sup> = 1.0
6	- 9	= ( -3 ) <sup>2</sup> = 9.0
6	- 9	= ( -3 ) <sup>2</sup> = 9.0
9	- 9	= ( 0 ) <sup>2</sup> = 0.0
8	- 9	= ( -1 ) <sup>2</sup> = 1.0
9	- 9	= ( 0 ) <sup>2</sup> = 0.0
13	- 9	= ( 4 ) <sup>2</sup> = 16.0
4	- 9	= ( -5 ) <sup>2</sup> = 25.0
8	- 9	= ( 1 ) <sup>2</sup> = 1.0
9	- 9	= ( 0 ) <sup>2</sup> = 0.0
7	- 9	= ( -2 ) <sup>2</sup> = 4.0
10	- 9	= ( 1 ) <sup>2</sup> = 1.0
13	- 9	= ( 4 ) <sup>2</sup> = 16.0
12	- 9	= ( -3 ) <sup>2</sup> = 9.0
14	- 9	= ( 5 ) <sup>2</sup> = 25.0
16	- 9	= ( 7 ) <sup>2</sup> = 49.0
		<hr/>
		200

Degrees of Freedom				
	Sum of Squares	d.f.	Mean Square	F Score
Sum of Squares 1st Factor (Gender)	32	1		
Sum of Squares 2nd Factor (Age)	93	2		
Sum of Squares Within (Error)	68	12		
Sum of Square Both Factors	7	2		
Sum of Squares Total	200	17		

Sum of Squares Within (Error) degrees of freedom					
Boys			Girls		
10 Year Olds	11 Year Olds	12 Year Olds	10 Year Olds	11 Year Olds	12 Year Olds
4	6	8	4	7	12
6	6	9	8	10	14
8	9	13	9	13	16
n - 1	n - 1	n - 1	n - 1	n - 1	n - 1
3 - 1	3 - 1	3 - 1	3 - 1	3 - 1	3 - 1
2	2	2	2	2	2
2 + 2 + 2 + 2 + 2 + 2 = 12					

Sum of both square factors = df1 x df2 = 1 x 2 = 2

Degrees of Freedom				
	Sum of Squares	d.f.	Mean Square	F Score
Sum of Squares 1st Factor (Gender)	32	1	$\frac{32}{1} = 32$	$\frac{32}{5.67} = 5.64$
Sum of Squares 2nd Factor (Age)	93	2	$\frac{93}{2} = 46.50$	$\frac{46.50}{5.67} = 8.20$
Sum of Square Both Factors	7	2	$\frac{7}{2} = 3.5$	$\frac{3.5}{5.67} = .62$
Sum of Squares Within (Error)	68	12	$\frac{68}{12} = 5.67$	
Sum of Squares Total	200	17		

$$F(1,12) = 5.64 \quad p < .05 \quad 95\% \text{ confidence level}$$

F Distribution $F(1,12) = 5.64, p < .05$																
degrees of freedom numerator																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.1	250.1	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	

$$F(1,12) = 1 \text{ column, } 12 \text{ row (4.75)}$$

$F(1,12) = 5.64 > 4.75$ , reject null hypothesis

For Age,  $F(2,12) = 8.2 \quad p < .05$

$F(2,12) = 8.2 > 3.89$ , reject null hypothesis

For sum of square both factors,  $F(2,12) = .62 \quad p < .05$

$F(2,12) = .62 < 3.89$ , accept null hypothesis

Gender	Score	Age
B	4	16
B	6	10
B	8	10
B	4	10
G	8	10
G	9	10
G	9	10
B	6	11
B	6	11
B	9	11
B	7	11
G	10	11
G	13	11
B	8	13
B	9	12
B	13	12
G	12	12
G	14	12
G	16	12

Factors  
 ↙ ↘  
 Gender Age

Boys		
10 yo	11 yo	12 yo
9	6	8
6	6	9
8	9	13

Girls		
10 yo	11 yo	12 yo
9	7	12
8	10	14
9	13	16

Mean table

	10 yo	11 yo	12 yo	Average
B	6	7	16	7.7
G Average	7	10	14	10.3
Average	6.5	8.5	12	9

Gender	Score	Age	Gender	Score	Age
B	4	10	G	4	9
B	6	10	G	6	8
B	8	10	G	6	9
b	6	11	b	6	7
b	6	11	b	6	16
b	9	11	b	6	13
b	8	12	b	6	12
b	9	12	b	6	14
b	13	12	b	6	16

Gender

B

7.7

G

10.3

9

Boys Score

4

6

8

6

6

9

8

9

13

Mean

7.7

~ 9

~ 9

~ 9

~ 9

~ 9

~ 9

~ 9

~ 9

Grand  
Mean

$$7.7 - 9 = (-1.3)^2 = 1.6$$

16.2

Girls Score

4

8

9

7

16

13

12

14

16

Mean

10.3

~ 9

~ 9

~ 9

~ 9

~ 9

~ 9

~ 9

~ 9

Grand  
Mean

$$10.3 - 9 = (-1.3)^2 = 1.6$$

16.2

Sum of Squares for 1st Gender = 32

Age

6.5

8.5

12

9

Boys

$$\begin{array}{r|l} 4 & 6.5 - 9 = (-2.5)^2 = 6.25 \\ 6 & 6.5 \\ 8 & 6.5 \\ 6 & 8.5 \\ 6 & 8.5 \\ 9 & 8.5 \\ 8 & 12 \\ 9 & 12 \\ 13 & 12 \end{array}$$

$\frac{46.5}{}$

Girls

$$\begin{array}{r|l} 4 & 6.5 - 9 = \\ 8 & 6.5 - \\ 9 & 6.5 - \\ 7 & 8.5 - \\ 10 & 8.5 - \\ 13 & 8.5 - \\ 12 & 12 - \\ 14 & 12 - \\ 16 & 12 - \end{array}$$

$\frac{46.5}{}$

Sum of Squares Of 2nd Age = 93

Sum of Squares within Error

BOYS

4	- 6	$= (-2)^2$	=	4
6	- 6	$= 0^2$	=	0
8	- 6	$= 2^2$	=	4
6	- 7	$= (-1)^2$	=	1
6	- 7	$= (-1)^2$	=	1
9	- 7	$= 2^2$	=	4
8	- 10	$= (-2)^2$	=	4
9	- 10	$= (-1)^2$	=	1
13	- 10	$= 3^2$	=	9

$\underline{28}$

Girls

4	- 7	$= -3^2$	=	9
8	- 7	$= 1^2$	=	1
9	- 7	$= 2^2$	=	4
7	- 10	$= -3^2$	=	9
16	- 10	$= 0^2$	=	0
13	- 10	$= 3^2$	=	9
12	- 14	$= -2^2$	=	4
14	- 14	$= 0^2$	=	0
16	- 14	$= 2^2$	=	4

$\underline{40}$

Total Sum = 68

Sum of Squares total		
Score	Grand Mean	$(Score - GrandMean)^2$
4	- . 9	$= (-5)^2 = 25$
6		$= -3^2 = 9$
8		$= -1^2 = 1$
6		$= -3^2 = 9$
6		$= -3^2 = 9$
9		$= 0^2 = 0$
8		$= -1^2 = 1$
9		$= 0^2 = 0$
13		$= 4^2 = 16$
4		$= -5^2 = 25$
8		$= 1^2 = 1$
9		$= 6^2 = 0$
7		$= -2^2 = 4$
10		$= 1^2 = 1$
13		$= 4^2 = 16$
12		$= -3^2 = 9$
14		$= 5^2 = 25$
16		$= 7^2 = 49$
		<u>200</u>

Sum of 1st (Gender) 32

Sum of 2nd (Age) 93

Sum of Squares within (Error) 68

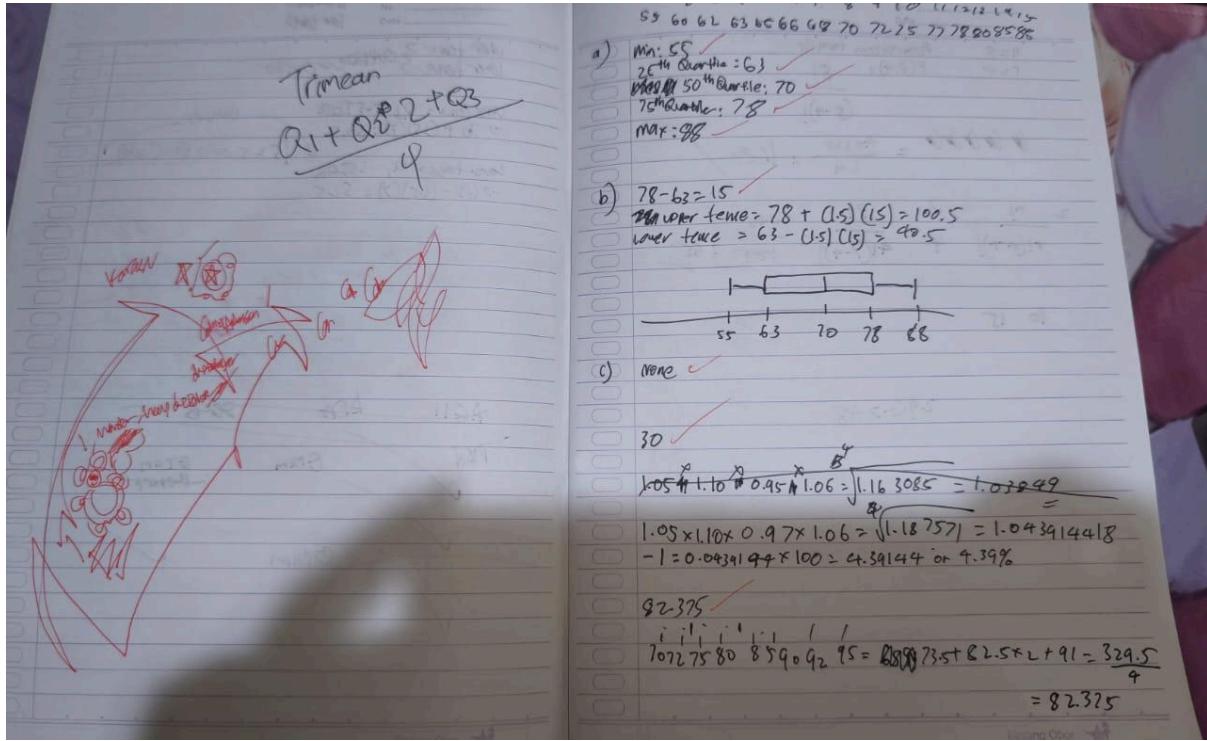
Sum of Squares both factor 7 C

Sum of Squares total 200

$$200 - 32 - 93 - 68 = 7$$



Exercise 1 & Exercise 2



### Exercise 3

2  
n = 8  
r = 4

Permutation formula  
 $P(8,4) = \frac{8!}{(8-4)!}$

$\rightarrow (8 \times 7 \times 6 \times 5) = \frac{40320}{24} = 1680$  ✓

2  $\frac{n!}{r!(n-r)!}$  7  $\frac{7!}{4!(7-4)!} = \frac{5040}{168} = 30$  ✓

10 15

29  $\leftrightarrow$  48

3 | 25

15

10

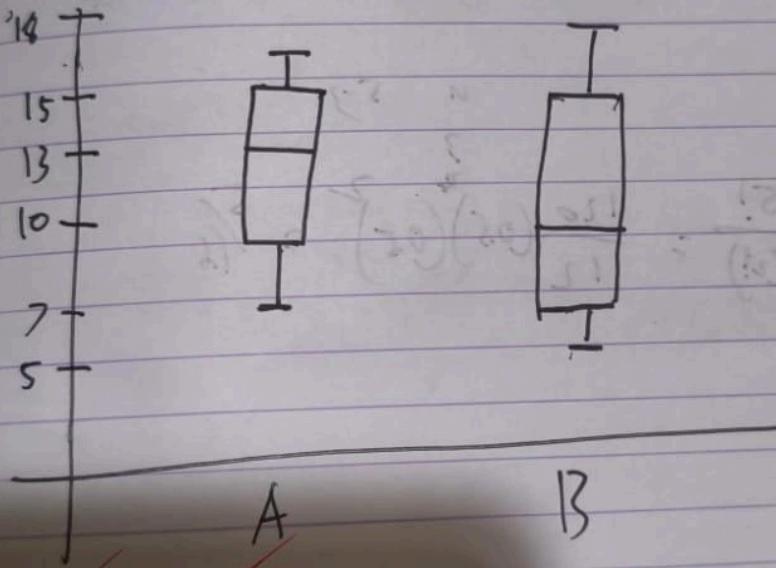
5

3

2 10  
15  
-5  
8  
12

$1.10 \times 1.15 \times 0.95 \times 1.08 \times 1.12 = \sqrt{1.9516168} = 1.077688$   
 $-1 = 0.077688 \times 100 > 7.7688\%$

A	B	C	D
Min: 7	Min: 5	6	8
25: 9	25: 7	$15+9=24$	$15+12=27$
50: 13	50: 10	No	No
75: 15	75: 15	$9-9=0$	$7-12=-5$
Max: 16	Max: 18		



A  
NO

$$\frac{1}{5_2} = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

✓

X

1
2
3

Y

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

$$\frac{5!}{3!2!} = \frac{120}{12} = 10 \cdot (0.5)^3 (0.5)^2 = \frac{10}{16}$$

$$\frac{5!}{3!(2!)^2} = \frac{120}{12} (0.5)^3 (0.5)^2 = \frac{10}{16} = \frac{5}{8}$$



$$\frac{15!}{12!(3!)}$$

$$(0.8)^{12}(0.2)^3 = 0.25$$

$$\frac{15!}{13!(2!)} (0.8)^{13} (0.2)^2 = 0.23$$

$$0.647$$

$$\frac{15!}{14!(1!)} (0.8)^{14} (0.2)^1 = 0.132$$

$$\frac{15!}{15!(0!)} (0.8)^{15} (0.2)^0 = 0.035$$

Hours (x)	Height (Y)
2	10
4	15
6	20
8	25
10	30

x	y	xy	x <sup>2</sup>	y <sup>2</sup>
2	10	20	4	100
4	15	60	16	225
6	20	120	36	400
8	25	200	64	625
10	30	300	100	900

$$30 \quad 100 \quad 700 \quad 220 \quad 2250$$

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

$$= \frac{5(700) - 30(100)}{\sqrt{[5(220) - (10)^2][5(2250) - (100)^2]}}$$

$$= \frac{500}{\sqrt{(1000 - 100)(11250 - 10000)}} = \frac{500}{\sqrt{200(1250)}} = \frac{500}{\sqrt{250000}} = \frac{500}{500} = 1$$

Bintang Obo



### Exercise 5

70 85 78 90 88

$$\bar{x}/\text{Mean} = 82.2$$

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

$$\rightarrow (70-82.2)^2 + (85-82.2)^2 + (78-82.2)^2 + (90-82.2)^2 + (88-82.2)^2$$

$$= \underbrace{148.84 + 7.84 + 17.64 + 60.84 + 33.64}_4$$

$$= \frac{268.8}{4} = \sqrt{67.2} = 8.198$$

No. \_\_\_\_\_  
Date \_\_\_\_\_

Z distribution

$$Z = \frac{x - \mu}{\sigma}$$

Binomial distribution

$$n=100$$

$$P=0.3$$

$$q=0.7$$

Can be approximated with normal distribution if  $n$  is large &  $P$  is not too close with 0/1.

$$n \cdot P \geq 5$$

$$n \cdot p = 100 \cdot 0.30 = 30$$

$$n \cdot q \geq 5$$

$$n \cdot q = 100 \cdot 0.70 = 70$$

$$\mu = n \cdot p$$

$$\mu = 100 \cdot 0.30 = 30$$

$$\sigma = \sqrt{n \cdot p \cdot q}$$

$$\sigma = \sqrt{100 \cdot 0.30 \cdot 0.70} = \sqrt{21} = 4.583$$

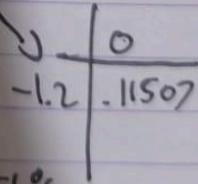
$$\text{Mean} = 30, \text{SD} = 4.58$$

Asked:  $P(x \leq 25) \xrightarrow{\text{continuity correction}} P(x \leq 24.5)$   $\longrightarrow P(x \leq 24.5)$

$$x=24.5$$

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

Probability  $\approx 0.11507 \times 100 = 11.507\% / 11.51\%$



$$n = 100$$

$$p = 0.4$$

$$q = 0.6$$

$$z = \frac{x - \mu}{\sigma}$$

$$N = n \cdot p = 100 \cdot 0.4 = 40$$

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{100 \cdot 0.4 \cdot 0.6} = \sqrt{24} = 2\sqrt{6}$$

$$P(X > 44) \approx P(X \geq 45) = P(X \geq 44.5)$$

$$z = \frac{44.5 - 40}{2\sqrt{6}} = \frac{4.5}{2\sqrt{6}} = 0.9148$$

$$\begin{array}{|c|c|} \hline & 0.02 \\ \hline 0.9 & , 82121 \\ \hline \end{array}$$

$$0.82121 \times 100 = 82.12\%$$

$$1 - 0.82121 = 0.17879 \xrightarrow{\times 100} 17.88\%$$

at least 45 success  $\rightarrow$  not  $\leftarrow$

At

~~0.82121~~

### Exercise 6

1 ~~950, 960, 970, 980~~  $\alpha: 0.05$

$$\text{Mean} = 990.5$$

$$S_1 = 25.87$$

$$\frac{990.5 - 1000}{\frac{25.87}{\sqrt{10}}} = -1.16$$

$$\text{units of freedom} = n-1 = 10-1 = 9$$

$$9 \left| \begin{array}{c} 0.05 \\ 2.262 \end{array} \right.$$

$-2.262 < -1.16 < 2.262$ , fail to reject

Date \_\_\_\_\_

$$\text{Diff} = -25$$

$$\mu = -25/8$$

$$\alpha = 0.05$$

$$S_d = 0.835$$

$$t = \frac{d}{S_d \sqrt{n}} = \frac{-25/8}{0.835 / \sqrt{8}} = \cancel{-1.332} -10.585$$

$$\begin{array}{c|c} 0.05 \\ \hline 1.895 \end{array}$$

$$\text{Unit of freedom} = n - 1 = 8 - 1 = 7$$

$$-10.585 < -1.895 < 1.895, \text{ Reject}$$

$$\alpha = 0.05$$

$$\bar{x}_1 = 8 \quad \sigma_1 = 2 \quad n_1 = 25$$

$$\bar{x}_2 = 6 \quad \sigma_2 = 2.5 \quad n_2 = 25$$

$$\text{Unit of freedom} = 25 + 25 - 2 = 48$$

$$t = \frac{8 - 6}{\sqrt{\frac{2^2}{25} + \frac{2.5^2}{25}}} = \frac{2}{\frac{\sqrt{41}}{10}} = 3.12$$

$$\begin{array}{c|c} 0.05 \\ \hline 1.679 \end{array}$$

$$3.12 > 1.679, \text{ Reject the null hypo}$$

Exercise 7

Answer

No.  
Date

①	②	③
1	2	2
2	4	3
5	2	4

$$\textcircled{1} H_0: M_1 = M_2 = M_3$$

$$H_A: M_1 \neq M_2 \neq M_3$$

$$\alpha = 0.05$$

$$\textcircled{3} \bar{x}_1 = \frac{8}{3} \text{ or } 2.67$$

$$\bar{x}_2 = 2.67$$

$$\bar{x}_3 = 3$$

$$G = \frac{25}{9} = 2.78$$

$$SS_{\text{Total}} = \sum (x - G)^2$$

$$= (1 - 2.78)^2 + (2 - 2.78)^2 + (5 - 2.78)^2 + (2 - 2.78)^2 + (4 - 2.78)^2$$

$$+ (2 - 2.78)^2 + (2 - 2.78)^2 + (3 - 2.78)^2 + (4 - 2.78)^2 = 13.6$$

$$SS_{\text{within}} = \sum (x_i - \bar{x}_i)^2 + (x_2 - \bar{x}_2)^2 + (x_3 - \bar{x}_3)^2$$

$$= (1 - 2.67)^2 + (2 - 2.67)^2 + (5 - 2.67)^2 + (2 - 2.67)^2 + (4 - 2.67)^2 \\ + (2 - 2.67)^2 + (2 - 3)^2 + (3 - 3)^2 + (4 - 3)^2 = 13.3$$

$$SS_{\text{between}} = 13.6 - 13.3 = 0.23$$

$$\textcircled{2} \text{ Degrees of freedom between} \\ = k (\text{groups}) - 1 = 3 - 1 = 2$$

$$\text{Degrees of freedom within} \\ = N - k = 9 - 3 = 6$$

$$df_{\text{total}} = 6 + 2 = 8$$

$$F_{\text{crit}} = 5.14$$

$$\begin{array}{r} 2 \\ 6 \mid 5.14 \\ \hline 10.92 \end{array}$$

A	B	C	$\alpha=0.05$
15	20	20	
16	22	27	
14	19	26	
15	21	28	
17	20	29	

$$df_{\text{between}} = 3 - 1 = 2$$

$$df_{\text{within}} = 15 - 3 = 12$$

$$df_{\text{total}} = 12 + 2 = 14$$

$$F_{\text{crit}} = 3.89 \quad | \quad 12 \quad | \quad 3.89$$

$$\bar{x}_1 = 15.4$$

$$\bar{x}_2 = 20.4$$

$$\bar{x}_3 = 26$$

$$G = 20.6$$

$$\begin{aligned} SS_{\text{total}} &= \sum (x - G)^2 = (15 - 20.6)^2 + (16 - 20.6)^2 + (14 - 20.6)^2 + (15 - 20.6)^2 \\ &+ (17 - 20.6)^2 + (20 - 20.6)^2 + (22 - 20.6)^2 + (19 - 20.6)^2 + (21 - 20.6)^2 + \\ &(20 - 20.6)^2 + (25 - 20.6)^2 + (27 - 20.6)^2 + (26 - 20.6)^2 + (28 - 20.6)^2 + \\ &(29 - 20.6)^2 = 301.6 \end{aligned}$$

$$\begin{aligned} SS_{\text{within}} &= (15 - 15.4)^2 + (16 - 15.4)^2 + (14 - 15.4)^2 + (15 - 15.4)^2 + \\ &(17 - 15.4)^2 + (20 - 20.4)^2 + (22 - 20.4)^2 + (19 - 20.4)^2 + (21 - 20.4)^2 + \\ &(20 - 20.4)^2 + (25 - 26)^2 + (27 - 26)^2 + (26 - 26)^2 + (28 - 26)^2 + \\ &(29 - 26)^2 = 120.4 \end{aligned}$$

$$SS_{\text{between}} = 301.6 - 120.4 = 181.2$$

$$\begin{aligned} MS_{\text{between}} &= \frac{181.2}{2} = 90.6 \\ MS_{\text{within}} &= \frac{120.4}{12} = 10.03 \\ F &= \frac{90.6}{10.03} = 8.99 \end{aligned}$$

## Observed

Fertilizer	Plant A	Plant B	Plant C	Total	$\alpha = 0.05$
X	10	20	10	40	
Y	15	10	7	32	
Z	5	5	10	20	
Total	30	35	27	92	

$$df = n - 1 = 2 \times 2 = 4$$

$$\begin{array}{r} .05 \\ 4 \quad | \quad 9.49 \rightarrow \text{crit} \end{array}$$

## Expected

Fertilizer	Plant A	Plant B	Plant C
X	13.3	15.6	11.1
Y	10	11.7	8.3
Z	6.7	7.8	5.6

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
10	13.3	-3.3	10.89	0.82
15	16	5	25	2.5
5	6.7	-1.7	2.89	0.43
20	15.6	4.4	19.36	1.24
10	11.7	-1.7	2.89	0.25
5	7.8	-2.8	7.84	1.01
10	11.1	-1.1	1.21	0.11
5	8.3	-3.3	10.89	1.31
10	5.6	4.4	19.36	3.46
				11.13

$11.13 > 9.49$ , Reject

1. Create a Stem-and-Leaf Display

Data set:  
62, 65, 66, 70, 73, 75, 75, 76, 81, 83, 84, 85, 87, 89, 92, 95, 96, 98, 100

C	2	5	6		
2	0	3	5	6	
5	1	3	4	5	7
9	2	5	6		
10	0				

2. Geometric Mean

Show the population of a city changes over four years with the following annual growth rates:  
Year 1: +10%  
Year 2: -5%  
Year 3: +4%  
Year 4: -6%

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

year	return	rate
1	5%	1.05
2	-3%	0.97
3	+4%	1.04
4	-6%	0.94

Geometric Mean =  $\sqrt[4]{1.05 \cdot 0.97 \cdot 1.04 \cdot 0.94} = 1.02\% \text{ growth}$

growth factor:  $1.05 \times 1.02 \times 1.04 \times 1.02 = 1.08274$

3. Truncated Mean

The following list shows the population of 10 states representing income scores:  
\$55, \$60, \$62, \$65, \$65, \$66, \$68, \$70, \$72, \$75, \$80, \$85, \$88

Tallest: \$88  
Tallest 2: \$85, \$88  
a) Determine the five number summary (minimum, 25th Quartile, 50th Quartile, 75th Quartile, maximum)  
b) Calculate the box plot based on the five number summary with whiskers (use 1.5 \* IQR spread to identify outliers for step).  
c) Identify any potential outliers (outside value or stand out for value).

a) min: \$55  
max: \$88  
25th quartile: \$62  
50th quartile: \$70  
75th quartile: \$78  
Lower Bound:  $62 + 2(70 - 62) = 45.5$   
Upper Bound:  $78 + 2(70 - 62) = 100.5$

b)

c) No outliers, as upper and lower whiskers are within the inner fences.  
8 p 4

2. 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?  
7 c 4

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

4. 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?  
10 C 3 15 C 2  
25 C 5

5. Calculate the probability of getting exactly 3 heads when flipping a fair coin 5 times (where getting heads is considered a success)?  
 $P(X=3) = \binom{5}{3} (0.5)^3 (0.5)^2 = 0.3125$

6. In a basketball game, a player has a free throw success rate of 60%. If the player takes 15 free throws, what is the probability that they make at least 10 successful free throws?  
 $P(Z \geq 12) = P(Z \geq 12) = P(Z \geq 12) + P(Z \leq 12) + P(Z \leq 12) = 15C12 (0.6)^{12} (0.4)^3 + 15C13 (0.6)^{13} (0.4)^2 + 15C14 (0.6)^{14} (0.4)^1 + 15C15 (0.6)^{15} (0.4)^0$

7. Find the percentage change when an investment grows from \$100 to \$145 in 5 years.  
 $\text{Growth Factor} = \frac{145}{100} = 1.45$   
 $\text{GM} = \sqrt[5]{1.45} = 1.08668$   
 $= 1.08668 \times 100 = 8.668\%$

8. A straight line shows the relationship between the number of height of plant measured in cm 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)
1	1.10
2	1.15
3	0.95
4	1.08
5	1.12

Calculate the Pearson correlation coefficient.

x	y	$x^2$	$y^2$	$xy$
2	1.10	4	1.21	2.20
4	1.15	16	1.3225	5.60
5	0.95	25	0.9025	4.75
6	1.08	36	1.1664	6.48
Total	10	100	50.1774	28.83
Mean	6	20	10.0857	5.4165

$r = \frac{100}{\sqrt{100 \times 25.1774}} = 1 \leftarrow \text{positive correlation}$

9. The following data set represents the scores of 5 students in a quiz:  
Scores: 80, 70, 85, 80, 88

Find the standard deviation from these data.

1. Mean =  $\frac{80+70+85+80+88}{5} = 82.2$   
2.  $S = \sqrt{\frac{(12.2)^2 + (8.2)^2 + (-2.2)^2 + (7.8)^2 + (7.8)^2}{5}} = 8.83$

3. A survey shows that 30% of drivers drink coffee and 12% drink tea, and then a coin is flipped. What is the probability of drinking a "King" the next day and having "King" the next day?

4. Two experiments at a company record the number of sales made by their top 10 salespeople in a month. The number of sales made are as follows:  
Depositions: 15, 16, 17, 18, 19, 20, 21, 22, 23, 24  
Sales: 30, 31, 32, 33, 34, 35, 36, 37, 38, 39

5. 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 40 successes will occur.  
 $M = np = 100 \times 0.4 = 40$   
 $O^2 = np(1-p) = 40(0.6) = 24$   
 $Z = \frac{40 - 40}{\sqrt{24}} = 0.92$   
 $P(Z \geq 0.92) = 1 - 0.82121 = 0.1788$

10. A company claims their light bulbs last 1000 hours on average. A sample of 100 bulbs yields the following results:  
Data set:  
91, 940, 970, 980, 1000, 1030, 990, 1010, 995, 1005

Null Hypothesis: Mean Bulb is 1000 hours  
Alternative Hypothesis: Mean Bulb is not 1000 hours

$\bar{x} = 990.5$   
 $s = \frac{25.668}{\sqrt{100}} = 2.5668$   
 $S = 25.868$   
 $n = 10$   
 $Df = 10 - 1 = 9$

Since  $-2.62 < -1.6 < +2.62$ , we fail to reject null hypothesis.

11. Three coaches measure the weight of 10 clients before and after a 6 week training program.

Client	Weight (before)	Weight (after)	Difference (D)
1	85	82	-3
2	90	85	-5
3	94	93	-1
4	88	86	-2
5	81	78	-3
6	79	76	-3
7	82	80	-2
8	86	84	-2
9	80	78	-2
10	84	82	-2

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

1.  $H_0: \mu_D = 0$   
2.  $H_a: \mu_D < 0$   
3.  $t = \frac{2.125}{\sqrt{18}} / \sqrt{8} = -2.125$   
4.  $t = -2.125 > -1.645$   
 $P(D < 0) = 0.057$   
 $P(D < 0) = 0.057 > 0.05$   
 $\therefore H_0$  is rejected.

5.  $P(Z < -1.62)$  since  $-1.62$  is not between  $-1.645$  and  $0.057$ , we reject  $H_0$ .  
 $= -0.11507$  since  $-0.11507$  is between  $-6.34$  and  $6.34$ , we reject  $H_0$ .  
 $\approx -0.1151$  | the null hypothesis.

To perform a Two-Way ANOVA, we'll follow these steps:

## Step 1: Define the Null Hypotheses

1.  $H_{01}$ : There is no significant effect of the programming language on test scores.
  2.  $H_{02}$ : There is no significant effect of the study method on test scores.
  3.  $H_{03}$ : There is no significant interaction effect between the programming language and the study method on test scores.

## Step 2: Data Entry

We have the following data from the table:

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

## **Step 3: Perform the Two-Way ANOVA**

Let's compute the Two-Way ANOVA using this data.

## **Step 4: Interpret the Results**

From the ANOVA table:

- **Effect of Language:**  $F(2,12)=40.97$   $F(2, 12) = 40.97$   $F(2,12)=40.97$ ,  $p=0.000004$   $p = 0.000004$   $p=0.000004$ 
  - The p-value is much less than 0.05, so we reject the null hypothesis  $H_01$  $H_01$ . This indicates a significant effect of programming language on test scores.
- **Effect of Study Method:**  $F(1,12)=59.90$   $F(1, 12) = 59.90$   $F(1,12)=59.90$ ,  $p=0.000005$   $p = 0.000005$   $p=0.000005$ 
  - The p-value is much less than 0.05, so we reject the null hypothesis  $H_02$  $H_02$ . This indicates a significant effect of the study method on test scores.
- **Interaction Effect (Language \* Study Method):**  $F(2,12)=0.17$   $F(2, 12) = 0.17$   $F(2,12)=0.17$ ,  $p=0.849605$   $p = 0.849605$   $p=0.849605$ 
  - The p-value is greater than 0.05, so we fail to reject the null hypothesis  $H_03$  $H_03$ . This suggests there is no significant interaction effect between the programming language and the study method on test scores.

## Conclusion:

Both the programming language and the study method have significant main effects on test scores. However, there is no significant interaction effect between these two factors.