Measures of Relative Dispersion (Coefficients of Variation)

Coefficient of variation "CV"

$$CV = \frac{\sigma}{\mu} \times 100 \text{ or } \frac{S}{\bar{x}} \times 100$$

- ✓ To compare the dispersion of <u>two (or more) groups</u>, we take the ratio of a measure of absolute dispersion to the corresponding measure of central tendency.
- ✓ Note that coefficients of variation are <u>**unitless.**</u>

Example

If the mean annual salaries of country(A) is $20,000 with a standard deviation of $2000 while the mean annual salaries of country(B) is $4,000 with a standard deviation of $1000. Which country has less inequality in salaries distribution?

	A	B
μ	**20,000**	**4,000**
σ	**2,000**	**1,000**
$CV = \frac{\sigma}{\mu} \times 100$	$\frac{2000}{20000} \times 100 = 10\ \%$	$\frac{1000}{4000} \times 100 = 25\ \%$

As country (A) is less varied than country (B), then country (A) has less inequality in salaries distribution than country (B).

كل لما ال **cv** <u>تطلع كبيرة</u> يبقى

<u>Higher</u> Variation/spread / dispersion/ Inequality/Instability/ Heterogeneity

❑ كل لما ال **cv** **تقل** يبقى **<u>Higher</u>** Homogeneity/Stability/Equality

Example

If you know that the arithmetical mean and the sample standard deviation of the production cost of factor "A" were respectively (60,0.5) and in factory " B" were respectively (88,0.5) , Determine which factory is more dispersed

	A	B
$\bar{x}$	**60**	**88**
S	**0.5**	**0.5**
$CV = \frac{S}{\bar{x}} \times 100$	$\frac{0.5}{60} \times 100 = 0.83\ \%$	$\frac{0.5}{88} \times 100 = 0.57\ \%$

Factory A more dispersed

The standardized values (z)

The standardized values (Z)

$$Z = \frac{Value - Mean}{Standard\ deviation} = \frac{x - \bar{x}}{S}$$

In certain occasions we need to compare two (or more) values belonging to two different data sets to determine the relative position of each value inside its own group.

An important fact about standardized units is that their average is always zero and their standard deviation is always one.

Example

You may be offered two job opportunities in two different countries. The annual salary of the first is \$36000 where the average annual salaries is \$30,000 with standard deviation \$4000. The annual salary of the second is \$28,000 where the average annual salaries is \$22000 with standard deviation \$2000. Which offer is relatively better?

	A	**B**
Annual salary	**36000**	**28000**
Average annual salary	**30000**	**22000**
Standard deviation	4000	2000
$Z = \frac{x - \bar{x}}{s}$	$\frac{36000-30000}{4000}$ = 1.5 units	$\frac{28000-22000}{2000}$ = 3 units

The second is better

Example

Given the following data find z score to determine which is better if :

1) A and B refers to profits
2) A and B refers to costs

	A	B
x	**70**	**90**
$\bar{x}$	**60**	**88**
S	0.5	0.5
$Z = \frac{x - \bar{x}}{s}$	$\frac{70-60}{0.5}$= 20 units	$\frac{90-88}{0.5}$= 4 units

- **Thus, if A and B refers to profits the first offer (20) may considered better**
- **Thus, if A and B refers to costs the second offer (4) may considered better**

Measures of skewness (β)

Measures of skewness (β)

$$\beta = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

$$\beta = \frac{\sum \left(\frac{x - \bar{x}}{s}\right)^3}{n - 1}$$

- If one tail of the distribution curve is longer than the other, the distribution is said to be asymmetric or skewed.
- When the right tail is the longer one, the distribution is said to be skewed to the right (positively skewed).
- If the left tail is longer, the distribution is said to be skewed to the left (negatively skewed).

Example

$(Q_3 - Q_2) < (Q_2 - Q_1)$

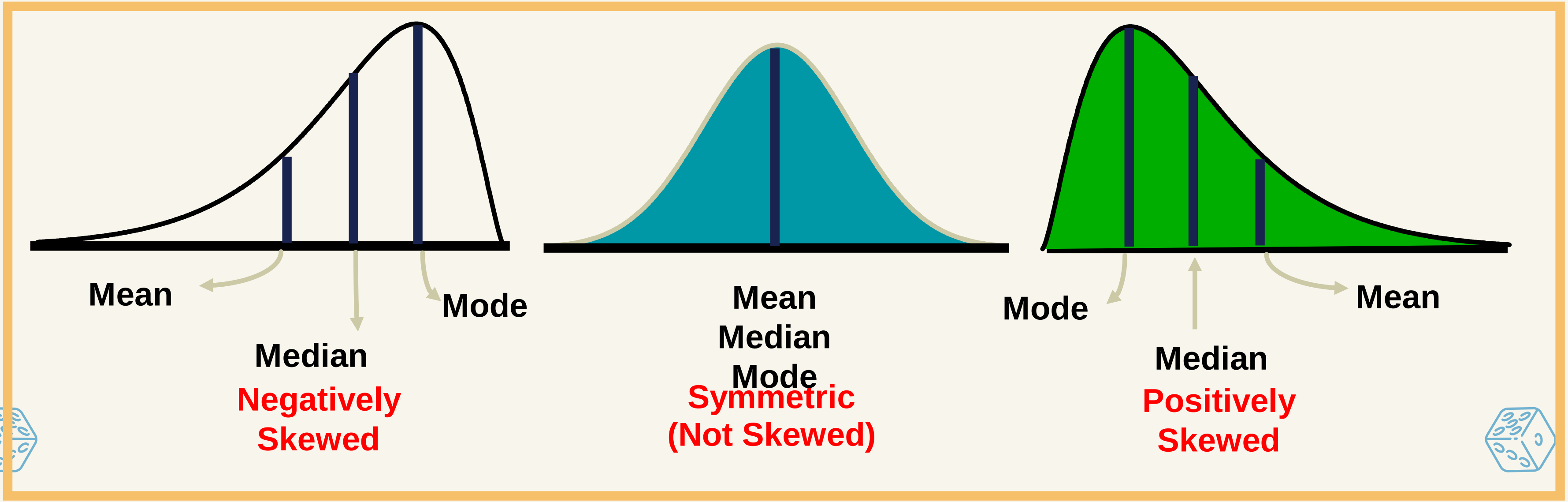
Mean $<$ median $<$ Mode

$(Q_3 - Q_2) = (Q_2 - Q_1)$

Mean $=$ median $=$ Mode

$(Q_3 - Q_2) > (Q_2 - Q_1)$

Mean $>$ median $>$ Mode



Box-and-Whisker Plot

Box-and-Whisker Plot Vocabulary

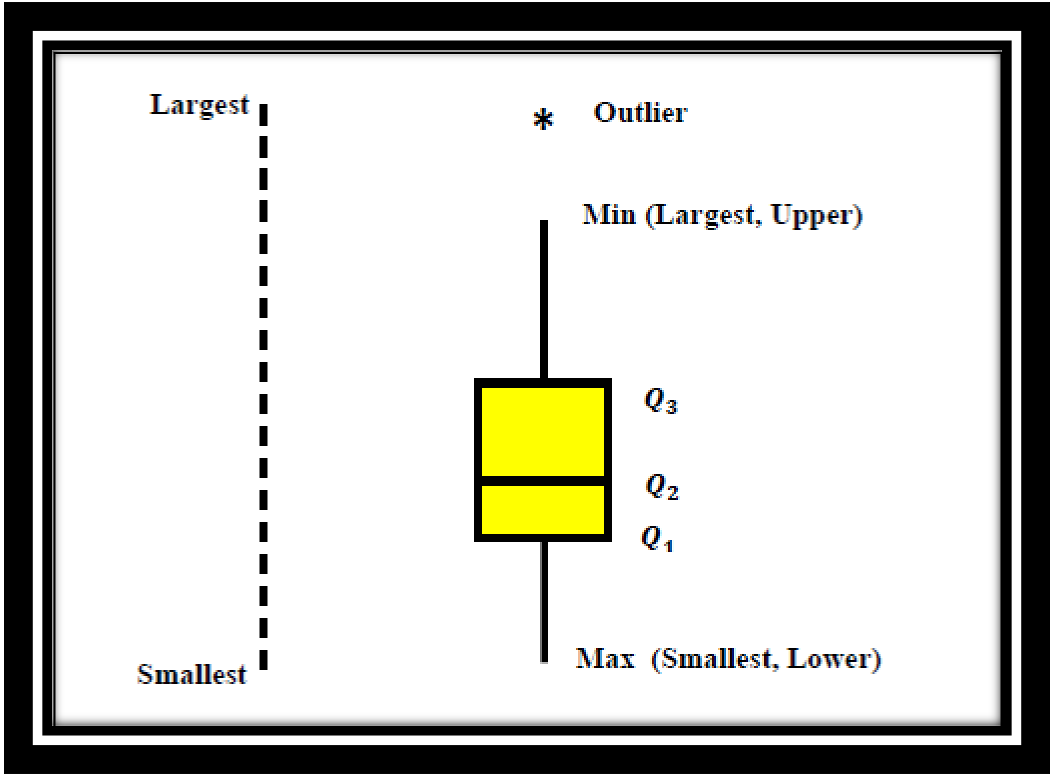
- **A box plot (also called box- whiskers plot) is a figure that is built up using 5 values. Yet, it reflects all the important information and features of the data distribution. To draw a box plot,**
- **1- Find the five values**

$Smallest\ observation$ $Q1$ $Q2$ $Q3$ largest

- **Median – The middle number of a set of data**
- **Upper Quartile – The median of the data values that are greater than the median of the total data set.**
- **Lower Quartile – The median of the data values that are less than the median of the total data set.**
- **Upper/Lower Extreme – The highest and lowest numbers in a set of data.**

- **2- Calculate two limits** (fences):

- **Upper Limit** (fence) = $Q3 + 1.5 \times (Q3 - Q1)$
- **Lower Limit** (fence) = $Q1 - 1.5 \times (Q3 - Q1)$


Largest
Smallest
* Outlier
Min (Largest, Upper)
Q_3
Q_2
Q_1
Max (Smallest, Lower)

Box-and-Whisker Plot Practice

Make a box-and-whisker plot using the following set of data:

1, 6, 10, 4, 2, 8, 15, 6, 3

Step 1: Put the numbers in order from least to greatest:

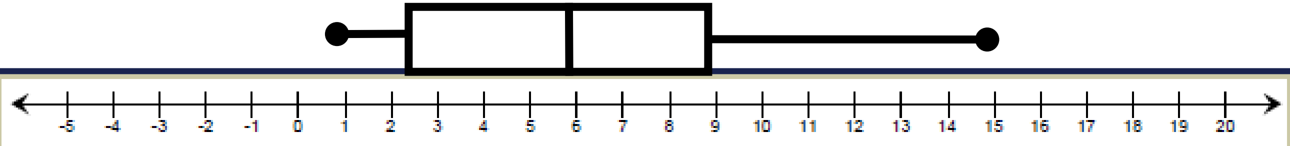
1, 2, 3, 4, 6, 6, 8, 10, 15

Step 2: Create a number line…that extends beyond the upper/lower extremes.

Step 3: Find the Median…draw a line above the median on the number line.

Step 4: Find the Upper Quartile…draw a line above the upper quartile.
Step 5: Find the Lower Quartile…draw a line above the lower quartile.

Step 6: Create a **<u>box</u>** (a rectangle) above the number line with the upper and lower quartiles as two of the sides.
Step 7: Create **<u>whiskers</u>** (straight lines) that extend from the box to the upper and lower extremes.



Box-and-Whisker Plot Practice

There are 12 people are in a Habanera Pepper eating contest who ate the following amount of peppers:

4, 4, 4, 9, 15, 2, 5, 0, 10, 12, 1, 18

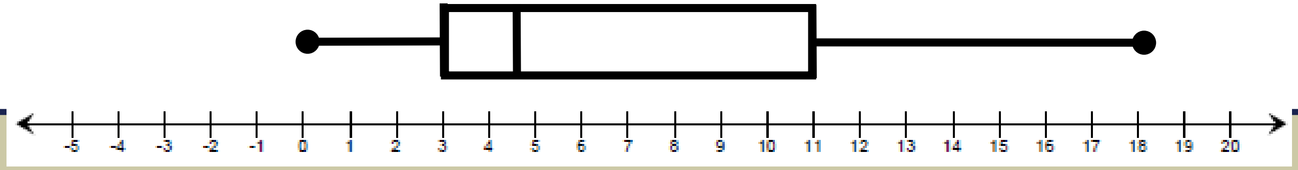
Put the numbers in order from least to greatest

0, 1, 2, 4, 4, 4, 5, 9, 10, 12, 15, 18

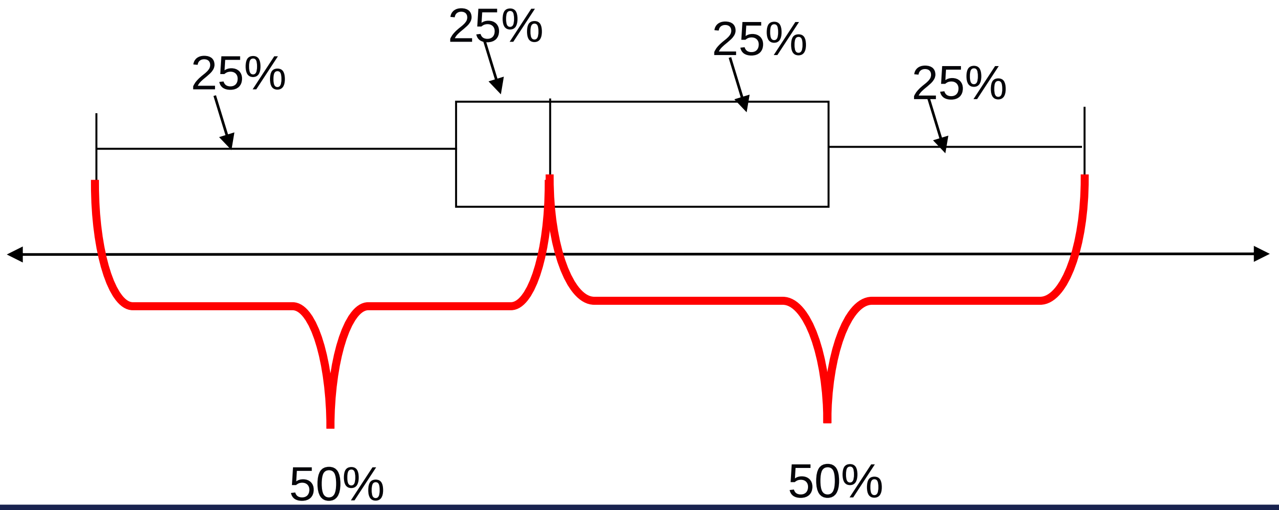
BELOW THE MEDIAN (4.5) **ABOVE THE MEDIAN (4.5)**

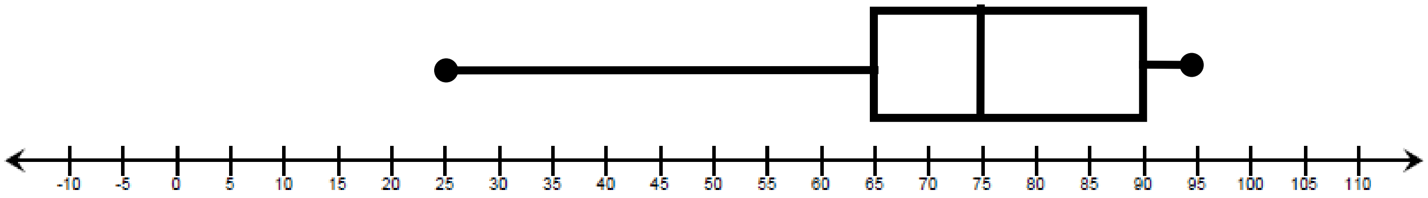
Discussion Question:

Why are all of the Boxes and Whiskers different sizes?



Analyzing a Box-and-Whisker Plot





Which statement is true?

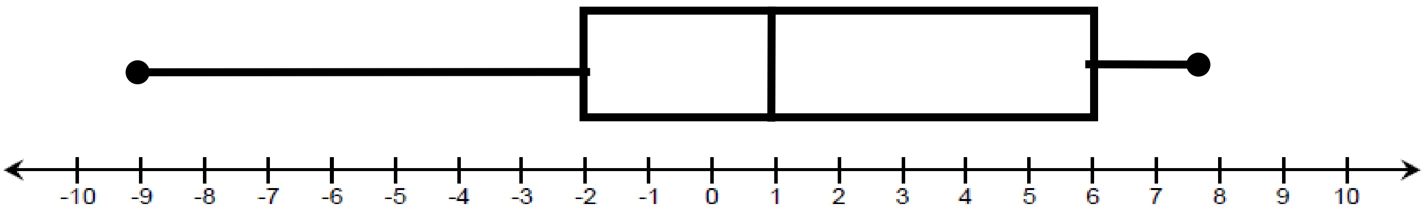
A. The range of the data is 25.

B. One-half of the data is below 65.

C. The median of the data is 60.

D. Three-fourths of the data is below 90.

Which of the following is not true about the box-and-whisker plot shown below?



A. The inter-quartile range is 8

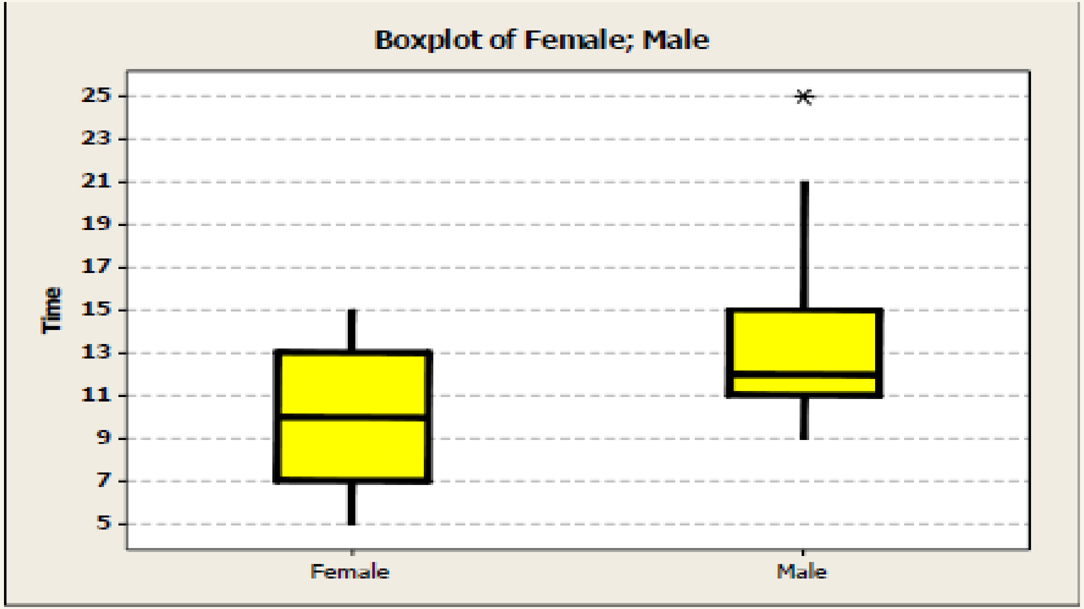
B. The upper quartile is -2.

C. The median of the data is 1.

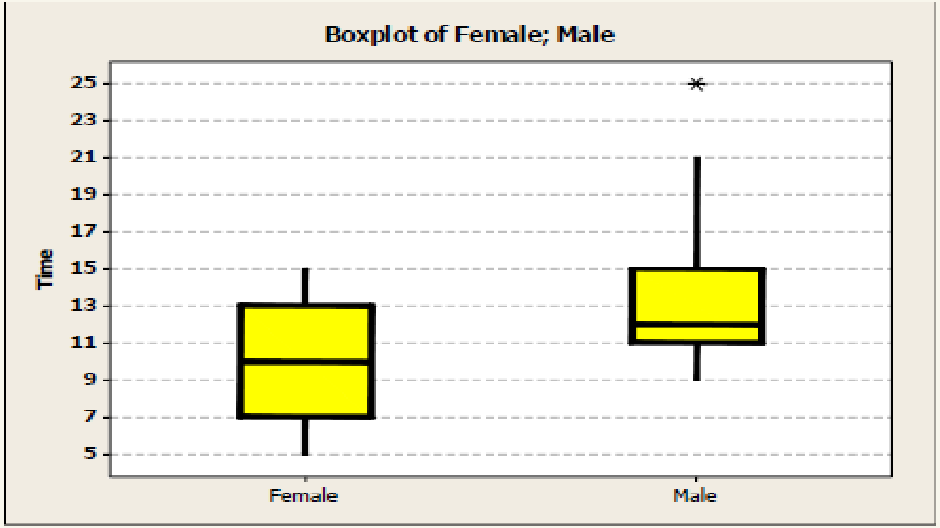
D. The lower extreme is -9.

Example

An experiment was conducted to compare the efficiency of males and females in executing a certain task. The following Box-Whiskers plots represent the distribution of the execution time (in minutes) of a random sample from each gender:



	Smallest observation	Q_1	Median	Q_3	**Largest observation**
Female					
Male					



b) Describe the type of skewness in each group. Justify your answer.
c) Determine the outlying observations. Justify your answer.
d) Compare the general level of observations of the males and females.
e) Compare the dispersion (or homogeneity) of the two groups. Justify your answer.
f) Which gender is more efficient in executing the task? Justify your answer.
g) Explain why the upper whisker of the males' plot is not extended to reach 25.
h) Explain why we can use the figure to determine the female mean execution time, while we cannot determine it for the male group.
i) It is more achievable for a female or a male to execute the task in just one minute?
j) Find the percentage of females who execute the task in more than 10 minutes.

Solution:

a)

	Min	Q_1	Median	Q_3	Max
Female	5	7	10	13	15
Male	9	11	12	15	25

b)

Female	Male
Symmetric $(Q_3 - Q_2) = (Q_2 - Q_1)$ 3=3	Skewed to right $(Q_3 - Q_2) > (Q_2 - Q_1)$ 3>1

c)

Female	Male
$\text{Upper Limit} = Q_3 + 1.5\times(Q_3 - Q_1) = 13 + 1.5(6) = 13 + 9 = 22$ $\text{Lower Limit} = Q_1 - 1.5\times(Q_3 - Q_1) = 7 - 1.5(6) = 7 - 9 = -2$ 15 not > 22 5 not < -2 No outliers.	$\text{Upper Limit} = Q_3 + 1.5\times(Q_3 - Q_1) = 15 + 1.5(4) = 15 + 6 = 21$ $\text{Lower Limit} = Q_1 - 1.5\times(Q_3 - Q_1) = 11 - 1.5(4) = 11 - 6 = 5$ 25 > 21. 25 is an outlier. 9 not < 5 9 is not outliers.

d)

Female	Male
The median Q_2 =10, less than	The median Q_2 =12

e)

Female	Male
$CV = \frac{IQR}{Q_2} \times 100 = \frac{6}{10} \times 100 = 60\%$ More dispersion.	$CV = \frac{IQR}{Q_2} \times 100 = \frac{4}{12} \times 100 = 33.3\%$

f)

Female is more efficient, because the median of time for female less than the median of male.

g)

The upper “whisker” connected to: Min. (Largest, Upper Limit) =Min. (25, 21)= **21**

h)

If the distribution is symmetric, then: Mean = Median = 10 (female),

While the distribution of male is skewed to right, and then : Mean > Median.

i)

Female	Male
1 is not less than -2 (lower limit), so 1 is not outlier. Then a female is more achievable.	1<5 (lower limit), so 1 is an outlier.

j) 50%.

Discrete Random Variables and Some Probability Distributions

1) Random variables

A random variable is a variable whose values are determined by the outcomes of a random experiment.

2) Discrete Random Variables

A discrete random variable assumes countable values. In other words, the consecutive values of a discrete random variable are separated by a certain gap.

3) Probability Distribution of a Discrete Random Variable

The probability distribution of a discrete random variable lists all the possible values that the random variable can assume and their corresponding probabilities.

- Two characteristics of a Probability Distribution of a Discrete Random Variable

(1) $0 \leq p(X = x) \leq 1$

(2) $\sum P(X = x) = \sum P(x) = 1$

Example

In a random experiment of tossing 3 coins, let the random variable (X) represents the number of Heads. Find The sample space , The probability distribution of the random variable (X) , $P(X \leq 1.7)$, $P(X \leq 2.99)$, $P(X \leq 3.08)$.

Answer:

1- The sample space (S) is given as:

S={HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

2- The possible values that the number of heads (X) can take are (0, 1, 2, and 3). Each of these numbers corresponds to an event in the sample space (S) of equally likely outcomes for this experiment:

S={HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

3 , 2 , 2 , 1 , 2 , 1 , 1 , 0

Range = { 0 , 1 , 2 ,3 }

$$P(0) = \frac{1}{8} = 0.125$$

$$P(1) = \frac{3}{8} = 0.375$$

$$P(2) = \frac{3}{8} = 0.375$$

$$P(3) = \frac{1}{9} = 0.125$$

Number of Heads (X = x)	0	1	2	3
Probability P(X= x)	0.125	0.375	0.375	0.125

Example

The probability distribution of the discrete random variable (X) can be summarized in the following table:

Number of Heads (X = x)	0	1	2	3
Probability P(X= x)	0.125	0.375	0.375	0.125

3- $P(X \leq 1.7) = P(X \leq 1) = P(X = 1) + P(X = 0) = 0.375 + 0.125 = 0.5$

4- $P(X \leq 2.99) = P(X \leq 2) = P(X = 2) + P(X = 1) + P(X = 0) = 0.375 + 0.375 + 0.125 = 0.875$

5- $P(X \leq 3.08) = P(X \leq 3) = P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0) = 0.125 + 0.375 + 0.375 + 0.125 = 1$

The Mean and Standard Deviation of a Discrete Random Variable

1- Mean (Expectation) : $\mu = E(X) = \sum x \, . P(x)$

2- Variance = $\sigma^2 = \sum x^2 . p(x) - \mu^2$ Or $= \sum (x - \mu)^2 . p(x)$

Example :

Given the following probability distribution of the random variable (X):

X=x	0	1	2	3	Total
P(x)	0.5	0.3	0.1	0.1	1

(1) Calculate the mean, standard deviation, and the coefficient of variation of the random variable (X).

X	P (X)	X.P(X)	$X^2.p(X)$
0	0.5	0	0
1	0.3	0.3	0.3
2	0.1	0.2	0.4
3	0.1	0.3	0.9
Total	1	0.8	1.6

X - μ	$(X-\mu)^2$	$(X-\mu)^2.p(x)$
0-0.8=-0.8	0.64	0.32
1-0.8=0.2	0.04	0.012
2-0.8=1.2	1.44	0.144
3-0.8=2.2	4.84	0.484
		0.96

1) Mean of x = μ_x = E(X) = $\sum X.p(x) = 0.8$

variance of x = $\sigma^2 = \sum x^2.p(x) - \mu^2$ = 1.6 - 0.8^2 = 0.96

Or = $\sum(x - \mu)^2.p(x)$ = 0.96 (طريقة ثانية)

Standard deviation = σ = $\sqrt{0.96} = 0.979$

Coefficient of variation = $\frac{\sigma}{\mu} \times 100 = \frac{0.979}{0.8} \times 100$ = 122.4 %

Binomial Distribution

The binomial probability distribution is one of the most widely used discrete probability distributions. It is applied to find the probability that an outcome will occur “x” times in “n” performances of an experiment .

For example given that 75% of the students at a college use Instagram , we want to find the probability that in a random sample of five students at this college , exactly three use Instagram.

An experiment that satisfies the following four conditions is called a binomial experiment.

1) There are "n" identical trials. In other words, the given experiment is repeated “n” times, where "n" is a positive integer. All of these repetitions are performed under identical conditions.

2) Each trial has two and only two outcomes (or can be classified into two mutually exclusive categories of outcomes). These outcomes are usually called a success and a failure.

3) The probability of success is denoted by “p " and that of failure by “1-p". The probability “p" remains constant for each trial

4) The trials are independent . In other words, the outcome of one trial does not affect the outcome of another trial

The Binomial Probability Mass Function (p.m.f.):

$P(X = x) = \binom{n}{x} p^x . (1-p)^{n-x}$; x = 0,1,2 ….. , n

Or = $(C_x^n)\, p^x . (1-p)^{n-x}$

Binomial Rules :

The Mean of "X" = E(X) = μ = n P

Variance = σ^2 = n.P.(1-p)

C.V = $\frac{\sigma}{\mu} \times 100 = \sqrt{\frac{1-P}{NP}} \times 100$

Where

n	= total number of trails
P	= probability of success
1 - P	= probability of failure
x	= number of successes in "n" trails
N −x	= number of Failure in "n" trails

The only values needed are those of “n " and “P",. These are called the **binomial parameters.**

Important Notes:

- If P > 0. 5, the distribution is skewed to the left.
- If P = 0. 5, the distribution is symmetric.
- If p ≠ 0. 5 but "n" is sufficiently large, the distribution is near symmetric.
- If p < 0. 5, the distribution is skewed to the right.
- $\mu > \sigma^2$ (Mean bigger than Variance)

Example

If the probability that a graduate will be employed is 0.60. In a sample of 5 students, find:

1. The probability that exactly two students will be employed.
2. The probability that exactly one of them will be employed.
3. The probability that none of the 5 students will be employed.
4. The probability that at least two will be employed.

Solution:

Let X be the number of employed graduate students then P = 0. 60, 1 - p = 1 - 0. 60 = 0. 40, n = 5.
The distribution is skewed to the left (or negatively skewed),
because p = 0. 6 > 0. 50

P(X = x) = $(C_x^n)\, p^x . (1-p)^{n-x}$

1) P(X = 2) = $C_2^5 \,. (0.6)^2 . (0.4)^{5-2}$ = 10 x 0. 36 x 0. 064 = o. 2304

2) P(X = 1) = $C_1^5 \,. (0.6)^1 . (0.4)^{5-1}$ = 5 x 0. 60 x 0. 0256 = 0. 0768

3) P(X = O) = $C_0^5 \,. (0.6)^0 . (0.4)^{5-0}$ = 1 x 1 x 0. 01024 = 0. 01024

4) P(X ≥ 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)

or 1 - P(X < 2) = 1 - [P(X = 1) + P(X = O)] = 1 - [O. 0768 + 0. 01024]

= 1 - [O. 08704] = **0. 912961**

Example

If the random variable X has a binomial distribution, with the parameters “P = 0. 5" and "n = 20", find the following probabilities:

P(5 < X < 7)   2) P(X > 0)

Solution:

1) $P(5 < X < 7) = p(X = 6) = C_6^{20} \,.(0.5)^6.(1 - 0.5)^{20-6} = 38760\ X\ 0.015625\ X\ 0.000061 = 0.03694$

2) $P(X > 0) = 1 - P(X \leq 0) = 1 - P(X = 0) = 1 - C_0^{20} \,.(0.5)^0.(1 - 0.5)^{20-0} \cong 1$

Poisson Distribution

The Poisson probability distribution is another important probability distribution of a discrete random variable that has a large number of applications .

Suppose a washing machine in a laundromat breaks down an average of three times a month. We may want to find the probability of exactly two breakdowns during the next month

- <u>For Three conditions must be satisfied to apply the Poisson probability distribution:</u>

1) X is a discrete random variable
2) The occurrences are random
3) The occurrences are independent.

- Notes :

• The Poisson distribution is used in experiments of rare events
• Poisson random quantities are used in place of Binomial in situations where : n is large and P is small
• The distribution is Skewed to right but will be Symmetric for large λ

- The Poisson probability Mass function (p.m.f) is given as : $P(X) = \frac{e^{-\lambda} . \lambda^{x}}{x\,!}$, $\lambda > 0$, x = 0,1,2,3,........
- Where e ≈ = 2.71828 ,
- λ is called the Poisson parameter and average number of occurrences in an interval (the mean , the expected number)

The mean = E (x) = $\mu = \lambda$
The variance = Var (x) = $\sigma^2 = \lambda$
$\mu = \sigma^2 = \lambda$

λ = n p

Example

If the number of warranty calls for a unit sold during its first year is a Poisson random variable. The expected number of warranty calls is 4.
Find:
1) The probability of receiving 2 calls.
2) The probability of receiving at least 2 calls.
3) Find the mean and the variance.

Solution:

$\lambda = 4$

1) $P(X = 2) = \frac{e^{-4} . 4^2}{2!} = 0.1465$

2) $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 1) + P(X = 0)]$

$= 1 - \left[\frac{e^{-4} . 4^1}{1!} + \frac{e^{-4} . 4^0}{0!}\right] = 0.9084$

3) The mean = the variance = λ = 4

Example

A university police department receives an average of 3.7 reports per week of lost students ID cards. Find the probability that during a given week the number of such reports received by this police department is:

At most 2

At least 2

Solution:

λ = 3.7

a. P(at most 2) = P(X ≤ 2) = P(X = 2) + P(X = 1) + P(X = 0)

$$= \left[\frac{e^{-3.7}.\ 3.7^2}{2\,!} + \frac{e^{-3.7}.\ 3.7^1}{1!} + \frac{e^{-3.7}.\ 3.7^0}{0\,!}\right] = 0.1692 + 0.0914 + 0.0247 = 0.2853$$

b) P(at least 2) = P(X ≥ 2) = P(X = 2) + P(X = 3) + P(X = 4) + …… .. ..

= 1 - [$P(X = 1) + P(X = O)$] = 1 - [0.0914 + 0.0247] = 0.8839

Normal Distribution

- The normal probability distribution is the **most important** and **most widely** used of all probability distributions

- A large number of phenomena in real world are normally distributed

- The normal probability distribution is a **bell shaped (symmetric)** curve.

- It has **2 parameters** population mean (μ) and population standard deviation (σ)

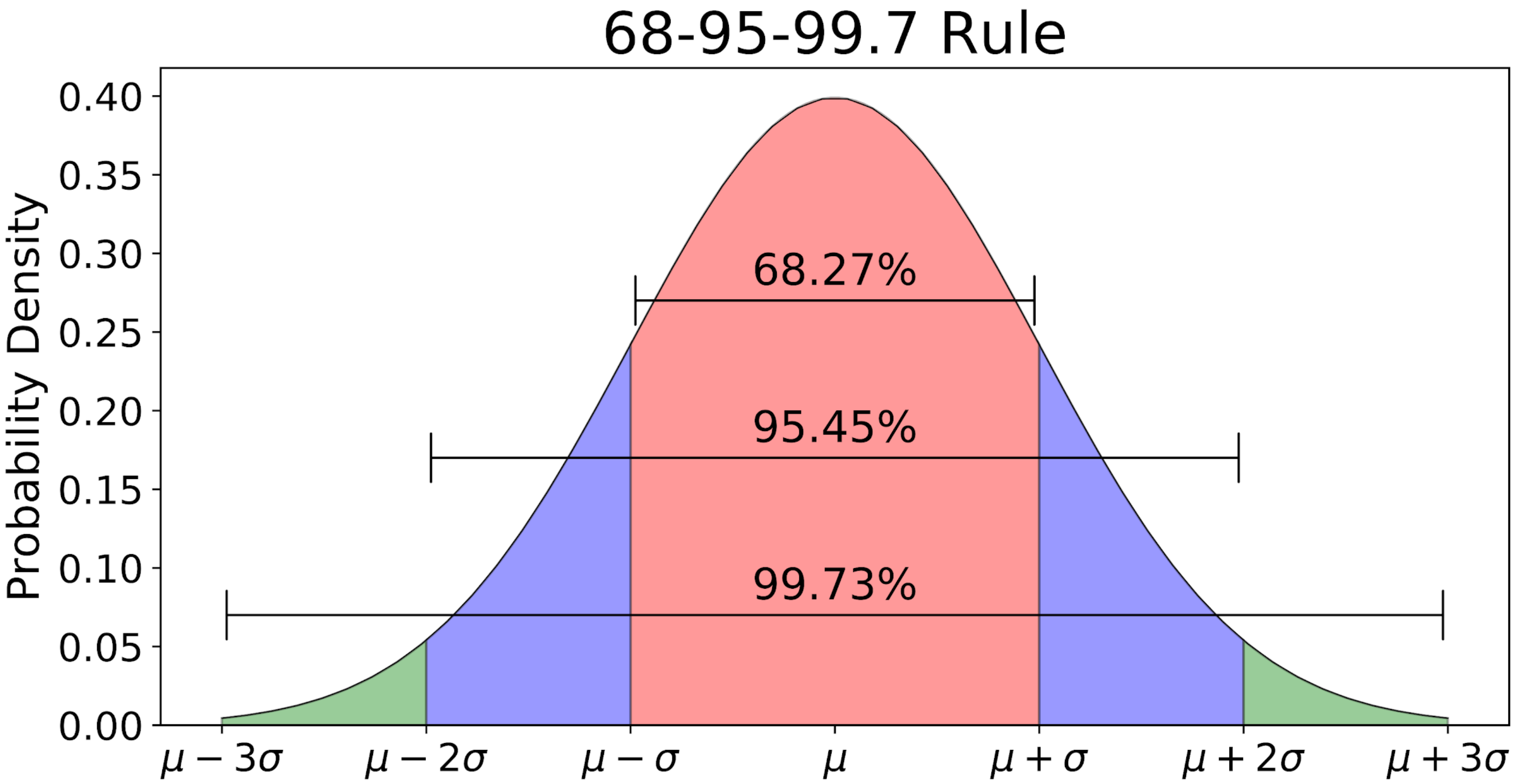
Characteristics of Normal Distribution :

1- Takes the bell shape and it's tails extend , theoretically from - ∞ to ∞
2- Symmetric distribution , Mean = Median = Mode
3- The are under the curve represents the probability

4- The total area under the curve is 1

5 – The normal distribution is always denoted as ~ N (μ , σ^2)

6- If X is a random variable having a normal distribution ~ ~ N (μ , σ^2) **, and**
If c is any **positive** number then P (X > μ + C) = p (X < μ – c)
parameter and average number of occurrences in an interval (the mean , the expected number)



68-95-99.7 Rule
Probability Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
68.27%
95.45%
99.73%
μ − 3σ
μ − 2σ
μ − σ
μ
μ + σ
μ + 2σ
μ + 3σ

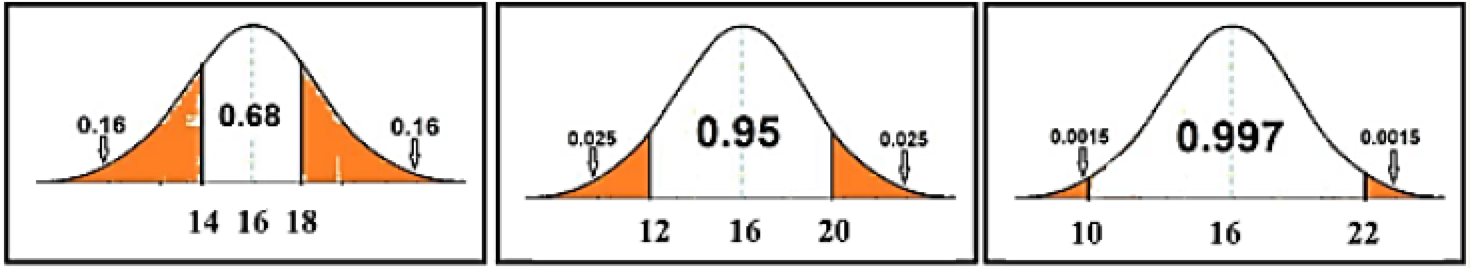
Example

If the scores of the Statistics mid-term exam follow normal distribution with mean of 16 and variance 4, find the probability that a student scored:

1) Less than 18 2) Greater than 22 3) Between 12 and 20

Solution:

Let $X \sim N(16,\ 4)$ where $\mu = 16$ and $\sigma^2 = 4$ and $\sigma = 2$



Less than 18 $= P(X < 18) = 0.68 + 0.16 = 0.84$ or $1 - 0.16 = 0.84$

Greater than 22 = $P(X > 22) = 0.0015$

Between 12 and 20 $= P(12 < X < 20) = 0.95$

Example

If If X has a normal distribution $N(50{,}25)$, find:

1- $P(X<45)$

2- $P(X>40)$

3- $P(X>50)$

4- The median and the mode.

Solution:

Let $X\sim N(50{,}25)$ where $\mu=50$ and $\sigma 2=25$ and $\sigma=5$

1- $P(X<45)=P(X<\mu-\sigma)=0.16$

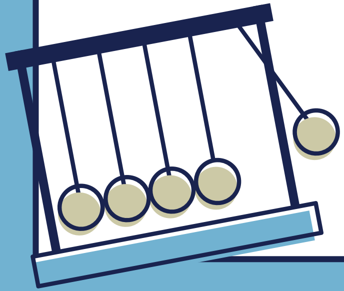
2- $P(X>40)=P(X>\mu-2\sigma)=0.95+0.025=0.975$

3- $P(X>50)=P(X>\mu)=0.5$

4- $mean=$ $median=$ $mode=50$

“الإحصاء و الرياضيات زي الدنيا في بعض الأحيان مبتتعلمش
غير لما تحاول تمشي صح بس تغلط , فتحاول تاني فتقع تاني
بعدها بقى تتعلم من غلطاتك و كل مزادتك تجاربك و محاولة
تعلمك من اللي فات كل لما إكتسبت خبرة”.

-أكمل العراقي



Thanks!

Do you have any questions?

Akmal.eleraky22@gmail.com
01280010848
YouTube: https://www.youtube.com/@akmaleleraky

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**