University of Mosul
Lecture No.: 5
College of Veterinary Medicine
Date: 2024-2025
Unit of Scientific Affairs
Website: https://uomosul.edu.iq/veterinarymedicine/

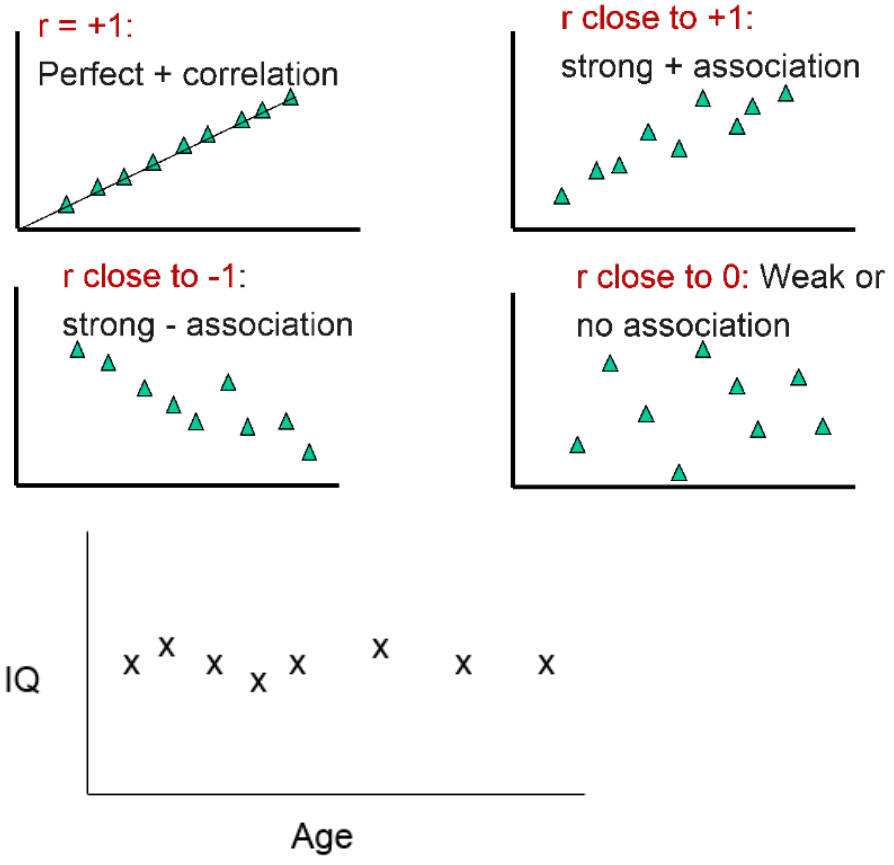
Lecture title: **Statistiscs: Measures of Variation**

Lecturer Affiliation: ***Iman Thannoon Sedeeq,*** *BSc, MSc, PhD*
Assistant Professor, Department of Public Health
College of Veterinary Medicine, University of Mosul, Mosul, Iraq
https://orcid.org/0000-0001-5903-3593

Summary: **Correlation Coefficients:** are used in statistics to measure how strong a relationship is between

two variables. There are several types of correlation coefficient:

Pearson's correlation (also called Pearson's *R*) is a **correlation coefficient** commonly used in linear regression.



The scatter plot suggests that measurement of IQ do not change with increasing age,

i.e., there is no evidence that IQ is associated with age.

Correlation coefficient describes the strength, or degree, of linear relationship. The Correlation coefficient formulas return a value between -1 and 1, where:

- 1 indicates a strong positive relationship that means for every positive increase in one variable, there is a positive increase of a fixed proportion in the other. For example, shoe sizes go up in (almost) perfect correlation with foot length.

- -1 indicates a strong negative relationship that means for every positive increase in one variable, there is a negative decrease of a fixed proportion in the other. For example, the amount of gas in a tank decreases in (almost) perfect correlation with speed.
- Zero indicates no relationship at all.

There are several types of correlation coefficient formulas.
One of the most commonly used formulas in stats is Pearson's correlation coefficient formula

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)}\sqrt{\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}}$$

Example: Find the value of the correlation coefficient from the following table:

SUBJECT	AGE X	GLUCOSE LEVEL Y
1	43	99

2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Step 1: *Make a tablet.* Use the given data, and add three more columns: xy, x^2, and y^2.

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	X^2	Y^2
1	43	99			
2	21	65			
3	25	79			
4	42	75			
5	57	87			
6	59	81			

Step 2: *Multiply x and y together to fill the xy column. For example, row 1 would be 43 × 99 = **4,257**.*

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	X^2	Y^2
1	43	99	4257		
2	21	65	1365		
3	25	79	1975		
4	42	75	3150		
5	57	87	4959		
6	59	81	4779		

Step 3: *Take the square of the numbers in the x column, and put the result in the x^2 column.*

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	X^2	Y^2
1	43	99	4257	1849	
2	21	65	1365	441	
3	25	79	1975	625	
4	42	75	3150	1764	

5	57	87	4959	3249	
6	59	81	4779	3481	

Step 4: *Take the square of the numbers in the y column, and put the result in the y^2 column.*

	AGE	GLUCOSE LEVEL			
SUBJECT	X	Y	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561

Step 5: *Add up all of the numbers in the columns and put the result at the bottom of the column.* The Greek letter sigma (Σ) is a short way of saying "sum of."

	AGE	GLUCOSE			
SUBJECT	X	LEVEL Y	XY	X^2	Y^2

1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

Step 6: *Use the following correlation coefficient formula.*

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right)}\sqrt{\left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}}$$

The answer is: **2868 / 5413.27 = 0.529809**
From the table:

- Σx = 247
- Σy = 486
- Σxy = 20,485
- Σx^2 = 11,409
- Σy^2 = 40,022
- n is the sample size, in our case = 6

The correlation coefficient = 6(20,485) – (247 × 486) / [√ [[6(11,409) – (247^2)] × [6(40,022) – 486^2]]]

= 0.5298