**MKTG5883 Exercise 1: Understanding Data Through Statistics**

1. **Data distribution**
   1. **Issue 1**

**A graph of height and height

AI-generated content may be incorrect.**

**Mean (Average):**

**Mean= Number of values Sum of all values ​**

**Sum of all values:** 110+128+129+129+130+130+130+131+132+132+132+132+133+133+133+133+133+133+134+134+134+134+135+135+135+135+136+136+138+153=3982

**Number of values (n):** n=30

**Mean** = 4032/30 ​ ~ 132.73cm

**Median:**

**Sorted data:**

110,128,129,129,130,130,130,131,132,132,132,132,133,133,133,133,133,133,134,134,134,134,135,135,135,135,136,136,138,153110, 128, 129, 129, 130, 130, 130, 131, 132, 132, 132, 132, 133, 133, 133, 133, 133, 133, 134, 134, 134, 134, 135, 135, 135, 135, 136, 136, 138, 153

**Number of values:** n=30

**Find the median position:**  
Since n is even, the median is the average of the 15th and 16th values.

Median=(133+133)/2 = 133

**Mode:**

| **Number** | **110** | **128** | **129** | **130** | **131** | **132** | **133** | **134** | **135** | **136** | **138** | **153** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Frequency | 1 | 1 | 2 | 3 | 1 | 4 | 6 | 4 | 4 | 2 | 1 | 1 |

**Frequency count:**

**Most frequent number:**

* + **133** appears **6 times**, making it the mode.

**Range Spread:**

The range is the difference between the maximum and minimum values.

**Range** = Maximum − Minimum= 153 – 110 = 43

**Conclusion on Data Distribution**

**Shape of Distribution:**

* + - The histogram shows a **roughly symmetric** distribution with a slight right skew due to the presence of a high value (153 cm).
    - The boxplot shows no extreme outliers except for 153 cm, which might be an outlier.

**Central Tendency:**

* The mean, median, and mode are close to each other (around 133 cm), which suggests that the data is approximately **normally distributed**.
  1. **Issue 2**

A diagram of a weight distribution

AI-generated content may be incorrect.

* **Given Data (Weights in kg):** 37 , 40 , 39 , 51 , 41 , 30 , 39 , 38 , 41 , 37 , 39 , 38 , 37 , 40 , 41 , 40, 37 , 39 , 40 , 41 , 38 , 39 , 40 , 41 , 39 , 40 , 40 , 38 , 39 , 41
* **Organize the Data in Ascending Order:** 30 , 37 , 37 , 37 , 37 , 38 , 38 , 38 , 39 , 39 , 39 , 39 , 39 , 39 , 39 , 40 , 40 , 40 , 40 , 40 , 40 , 40 , 41 , 41 , 41 , 41 , 41 , 41 , 41 , 51
* **Mean (Average)**

Sum = 37 + 40 + 39 + 51 + 41 + 30 + 39 + 38 + 41 + 37 + 39 + 38 + 37 + 40 + 41 + 40 + 37 + 39 + 40 + 41 + 38 + 39 + 40 + 41 + 39 + 40 + 40 + 38 + 39 + 41 = 1180

n = 30

**Mean** = 1180/30 ~ 39.33333

|  |  |  |  |
| --- | --- | --- | --- |
| n | = | 30 | =15 |
| 2 | 2 |

|  |  |  |  |
| --- | --- | --- | --- |
| n | + 1 = | 30 | + 1 =16 |
| 2 | 2 |

* **Median****n = 30**.
* The 15'th number is 39, the 16'th number is 39  
  **30 , 37 , 37 , 37 , 37 , 38 , 38 , 38 , 38 , 39 , 39 , 39 , 39 , 39 , 39 , 39 , 40 , 40 , 40 , 40 , 40 , 40 , 40 , 41 , 41 , 41 , 41 , 41 , 41 , 51**
* Take the average of the two middle numbers:  
  **(39+39)/2 = 39.**
* **Mode**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Number** | **30** | **37** | **38** | **39** | **40** | **41** | **51** |
| **Frequency** | **1** | **4** | **4** | **7** | **7** | **6** | **1** |

**39, 40** appear most frequently (7 times), these are the modes

* **Spread**

**=** Max – Min = 51 – 30 = 21

**Conclusion:** The distribution of student weights is approximately **symmetrical (normal distribution)**. There are no significant outliers, and the spread is moderate. The data is right-skewed (positively skewed).

1. **Data correlation**

**A graph with blue dots

AI-generated content may be incorrect.**

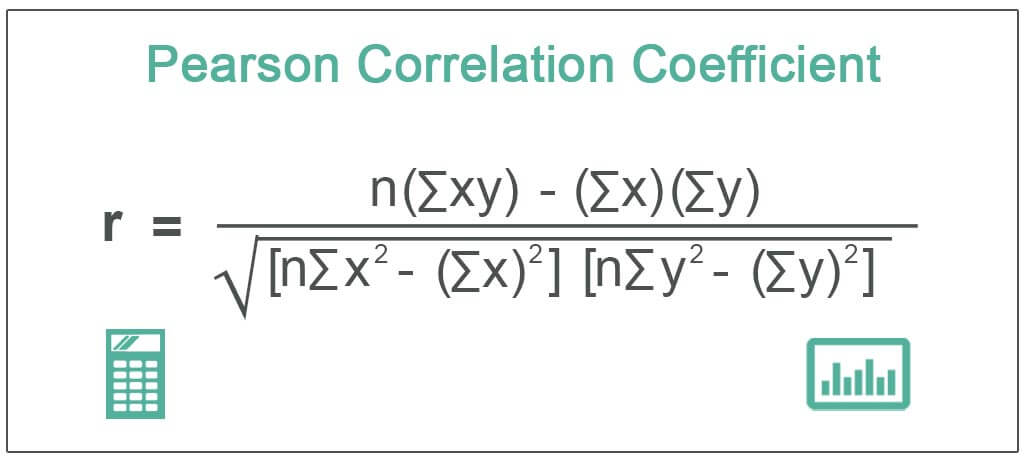
**Is there any correlation between height and weight of students?**

Likely, yes. Generally, taller individuals tend to weigh more, suggesting a positive correlation. However, this isn't always the case, and individual variations exist. The presence of outliers in both datasets (110cm/153cm for height and 30kg/51kg for weight) could influence the correlation.

Calculate the Pearson correlation coefficient.

**Steps to Quantify Correlation**

1. Pair the Data:  
   Combine height and weight data into pairs.
2. Calculate Pearson Correlation Coefficient (r):



where:

* + n = number of pairs
  + x = height
  + y = weight

**Conclusion on Correlation**

* If r is close to +1, there is a strong positive correlation.
* If r is close to -1, there is a strong negative correlation.
* If r is around 0, there is no correlation.

**Solve the problems:**

Σ(xi - x̄)2 = (130-132.73)2+(132-132.73)2+(138-132.73)2+(153-132.73)2+(133-132.73)2+(110-132.73)2+(132-132.73)2+(129-132.73)2+(135-132.73)2+(134-132.73)2+(136-132.73)2+(133-132.73)2+(133-132.73)2+(134-132.73)2+(135-132.73)2+(132-132.73)2+(135-132.73)2+(134-132.73)2+(133-132.73)2+(132-132.73)2+(130-132.73)2+(131-132.73)2+(135-132.73)2+(134-132.73)2+(136-132.73)2+(133-132.73)2+(133-132.73)2+(130-132.73)2+(129-132.73)2+(128-132.73)2 = 1081.8667

Σ(yi - ȳ)2 = (37-39.33)2+(40-39.33)2+(39-39.33)2+(51-39.33)2+(41-39.33)2+(30-39.33)2+(39-39.33)2+(38-39.33)2+(41-39.33)2+(37-39.33)2+(39-39.33)2+(38-39.33)2+(37-39.33)2+(40-39.33)2+(41-39.33)2+(40-39.33)2+(37-39.33)2+(39-39.33)2+(40-39.33)2+(41-39.33)2+(38-39.33)2+(39-39.33)2+(40-39.33)2+(41-39.33)2+(39-39.33)2+(40-39.33)2+(40-39.33)2+(38-39.33)2+(39-39.33)2+(41-39.33)2 = 272.6667

Σ(xi - x̄)(yi - ȳ) = (130-132.73)\*(37-39.33)+(132-132.73)\*(40-39.33)+(138-132.73)\*(39-39.33)+(153-132.73)\*(51-39.33)+(133-132.73)\*(41-39.33)+(110-132.73)\*(30-39.33)+(132-132.73)\*(39-39.33)+(129-132.73)\*(38-39.33)+(135-132.73)\*(41-39.33)+(134-132.73)\*(37-39.33)+(136-132.73)\*(39-39.33)+(133-132.73)\*(38-39.33)+(133-132.73)\*(37-39.33)+(134-132.73)\*(40-39.33)+(135-132.73)\*(41-39.33)+(132-132.73)\*(40-39.33)+(135-132.73)\*(37-39.33)+(134-132.73)\*(39-39.33)+(133-132.73)\*(40-39.33)+(132-132.73)\*(41-39.33)+(130-132.73)\*(38-39.33)+(131-132.73)\*(39-39.33)+(135-132.73)\*(40-39.33)+(134-132.73)\*(41-39.33)+(136-132.73)\*(39-39.33)+(133-132.73)\*(40-39.33)+(133-132.73)\*(40-39.33)+(130-132.73)\*(38-39.33)+(129-132.73)\*(39-39.33)+(128-132.73)\*(41-39.33) = 458.6667

|  |  |
| --- | --- |
| SXY = | Σ(xi - x̄)(yi - ȳ) |
| n - 1 |

|  |  |  |
| --- | --- | --- |
| SXY = | 458.6667 | = 15.8161 |
| 30 - 1 |

|  |  |
| --- | --- |
| r = | Σ(xi - x̄)(yi - ȳ) |
| √(Σ( xi - x̄)2Σ(yi - ȳ)2 ) |

|  |  |  |
| --- | --- | --- |
| r = | 458.6667 | = **0.8445** |
| √(1081.8667\*272.6667) |

* + - The Pearson correlation coefficient (r=0.8445) indicates a **strong positive correlation** between the two variables x and y. This means that as x increases, y tends to increase as well.