Final Project

by

Priya Subramanian

Topic :   Statistik mit R

Lecturer :   Julianne Wawerda

Project Period :   28.07.2021 – 30.07.2021

Project Partners : Ghazaleh Erfanianfar

                   Dan-Ovidiu Melinte

                   Serkan Tezgel

                   Srilakshmi Velusamy

Table of Contents

Task 1: Basics

SAP before	2	5	2	7	5	6	1	3	7	3
SAP after	10	10	8	6	4	9	4	8	7	5

1) Calculate the Means, Mode/Modes and Median (SAP_before und SAP_after)

Mode:

SAP_before:

2, 5,2, 7, 5, 6, 1, 3, 7, 3

Value	Frequency
1	1
2	2
3	2
5	2
6	1
7	2

(Has **4 Modes**)
Modes: **2,3,5,7**
Type: **Multimodal**

SAP_after:

10, 10, 8, 6, 4, 9, 4, 8, 7, 5

Value	Frequency
4	2
5	1
6	1
7	1
8	2
9	1
10	2

(Has **3 Modes**)
 Modes: **4,8,10**
Type: **Trimodal**

Mean:

Symbol: $\bar{x}$

SAP_before:

$\sum (2+5+2+7+5+6+1+3+7+3) / 10 = \bar{x}\ (4.1)$

Mean = 4.1

SAP_after:

$\sum (10+10+8+6+4+9+4+8+7+5) / 10 = \bar{x}\ (7.1)$

Mean = 7.1

Median:

SAP_before:

2, 5, 2, 7, 5, 6, 1, 3, 7, 3
Sorted:
1, 2, 2, 3, 3, 5, 5, 6, 7, 7

Median = 4

SAP_after:

10, 10, 8, 6, 4, 9, 4, 8, 7, 5
Sorted:
4, 4, 5, 6, 7, 8, 8, 9, 10,10

Median = 7.5

2) Calculate the variances and Standard Deviations (SAP_before and SAP_after)

<u>Variance:</u>

Formula:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

SAP_before:	SAP_after:
Variance = 4.76	Variance = 5.21

<u>Standard Deviation:</u>

Formula:

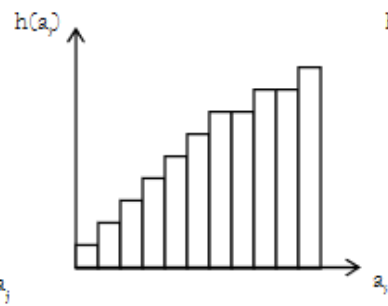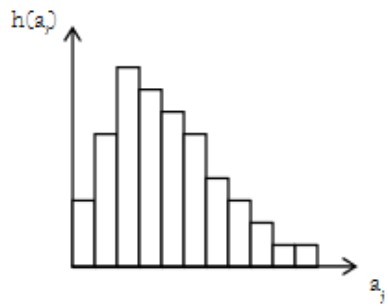$$SD = \sqrt{\frac{\sum(x - \overline{x})^2}{n}}$$

SAP_before:	SAP_after:
Standard deviation = 2.18	Standard deviation = 2.28

3) Is the Graph right-skewed, left-skewed, or symmetrical?

$h(a_j)$ $h(a_j)$ $h(a_j)$
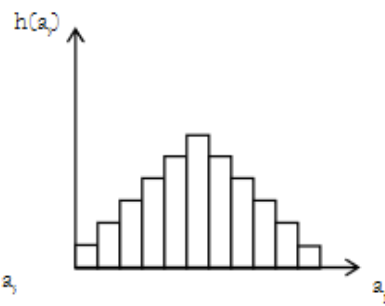
Right-skew scaled   Left-skew scaled   Symmetrical scaled

4) Assign the scale level to the following data: (Nominal, Ordinal, Interval, Ratio, Absolut), and which operation is allowed.

Variable type	Scale Level	Operation
Military Rank	Ordinal	= / ≠ < / >
Age	Ratio	= / ≠ < / > + / - ÷ / *
Traffic density	Ratio	= / ≠ < / > + / - ÷ / *
Gender	Nominal (Dichotomous)	= / ≠
Fares	Ratio	= / ≠ < / > + / - ÷ / *
Nationality	Nominal	= / ≠
School Education	Ordinal	= / ≠ < / >
Intelligence Quotient	Interval	= / ≠ < / > + / -
Subject	Nominal	= / ≠
Semester number	Absolut	= / ≠

		< / > + / - ÷ / *
Exam Points	Ratio	= / ≠ < / > + / - ÷ / *
Rate classes for Motor Vehicle liability	Ordinal	= / ≠ < / >

5)   Assign the following variables to their variable level (continuous vs. discrete).

Nr.	Wert	Variable	
		discrete	continuous
1	Tax class	◉	○
2	Gender	◉	○
3	Social Class	◉	○
4	Income Tax	○	◉
5	Temperatur in Kelvin	○	◉
6	Windforce in Meter/Sekunde	○	◉
7	Body weight	○	◉

8	School grade (1-6)	⦿	○
9	Exam points	○	⦿
10	Population	⦿	○
11	Semester number	⦿	○
12	Commercial class (Fruit)	⦿	○

6) Describe in sentences what is the difference and similarities between standard normal distribution and normal distribution. Use the formulas.

Similarity:

* The basic properties for the both are same. Such as , Bell-shaped, Symmetrical is similar to Empirical Rule

Difference:

Standard Normal Distribution	Normal Distribution
* It is a special case of the normal distribution	* It is the most commonly used probability distribution in statistics.
* Mean is equal to 0 and the standard deviation is equal to 1	* Mean and median are equal; both located at the center of the distribution
	* The mean of the normal distribution determines its location
	* The standard deviation determines its spread.

Formula:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{1 \times \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Task 2: Multiple Choice

The special point is valid only within this task. Maximum you can reach is 10 points.

1) A Bravais-Pearson correlation coefficient of 0.85 indicates a weak linear correlation.	
Richtig	**Falsch**
2) The interquartile range (IQR) is twice the distance between median and mode.	
Richtig	**Falsch**
3) The modes can only be determined if there is a unimodal distribution.	
Richtig	**Falsch**
4) Nominal scaled data can be put in a natural order.	
Richtig	**Falsch**
5) Outliers have a particularly strong effect on the results of non-robust analysis methods.	
Richtig	Falsch
6) The standard deviation is not calculated as a positive square root of the variance.	
Richtig	**Falsch**
7) The kurtosis is a measure of the curvature of a distribution.	
Richtig	Falsch
8) The calculation of the variance requires at least metrically scaled data.	
Richtig	Falsch

9) The range is the absolute distance between the smallest and the largest value.	
Richtig	Falsch

10) The Bravais-Pearson correlation coefficient can only take values between 0 and 1.	
Richtig	Falsch

11) The statistical replacement of missing values requires at least metrically scaled data.	
Richtig	Falsch