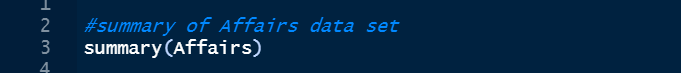
Affairs data set describes the division of a person’s time between work, time spent with a spouse and time spent with a paramour. This data set has 601 rows and 10 total columns. There are 9 variables in this data set. Those variables are age, affairs, gender, yearsmarried, children, religiousness, education, rating. Gender and children variables are qualitative while other variables are quantitative.

Graphical user interface, funnel chart

Description automatically generatedTable

Description automatically generated with low confidence  
Data about numerical variables is known as quantitative data. These type of data can be counted, measured and expressed by using numbers. Qualitative data are data about categorical variables. They are descriptive and conceptual.

Summary statistics about Affairs data set are as follows:



Console output:

A screen shot of a computer

Description automatically generated with low confidence

This summary statistics of affairs data set shows the central tendency(mean, median, mode), Minimum value, Maximum value, first quartile and the third quartile of each and every numerical variables in the data set.

Exploratory Data Analysis(EDA) is used to summarize the data to develop the hypothesis about the data set and identify the patterns in the data set. Central tendency and the measure of dispersion is used to describe the data’s distribution statistically. (Exploratory Data Analysis | US EPA, 2022)

Central tendency represents the numeric center of the data. Dispersion represents the spread of data around the data’s center. These two measurements provide overall description of a set of data.

The measure of dispersion gives an idea about how the mean is represents the data. When the spread of the values of data set is large then the mean is not representative of the data. This is due to the large spread indicates that there are likely to be large differences between individual estimates. Measure of dispersion is related to the range, variance and the standard deviation of the data set. (Measures of Spread | How and when to use measures of spread | Laerd Statistics, 2022)

Range is the difference between maximum value and the minimum value. It is the simplest measure of spread. Range shows how much numbers in a set vary. This can be useful when measuring the variables that have a critical lower or upper threshold (or both) that cannot be crossed. The range will notify if at least one value has passed this critical threshold. In addition to that, range can be used to identify any errors when entering the data.(What are Mean, Median, Mode and Range?, 2022)

Range=Highest value - Lowest value

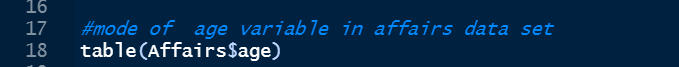
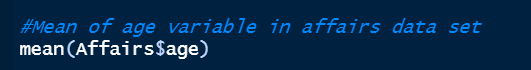
Variance is used to get more representative idea about spread of the data set.

Variance=

If the values are spread closely around the mean, then the variance will be small value. If the values are spread out, then the variance will be large value. Variance is measured in the units of squared. Therefore, it can not include to the frequency distribution. Standard deviation is calculated to avoid this problem.(What Is Data Variability? Data Defined - Indicative, 2022)

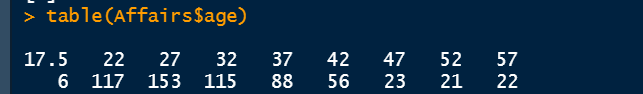
Standard deviation= (2. Mean and standard deviation | The BMJ, 2022)

Central tendency of age variable are as follows:



Console outputs:





By using this results, the mean age of a person who is participated to this survey is 32.48752 years.The middle age of a person is 32 years.Most of people are 27 years of age. (Practical Statistics for Data Scientists, 2022)

Histogram for age variable:

Histogram is a value distribution plot, which is drawn to a numerical variable. The distribution of data can be easily identify using the histograms. Whether the more values are lie in positive, negative or at the center can be identify using the histogram.

Graphical user interface, text

Description automatically generatedChart, histogram

Description automatically generated

Figure 01:Histogram for age variable in Affairs data set

The data distribution is right skewed(positive skewed). This situation occurs because the probabilities for higher values slowly decreases. Which means the number of persons who are older than 32 years of age are decreases. (Frost, 2022)

Central tendency of yearsmarried variable are as follows:



A blue screen with white text

Description automatically generated with medium confidence

Console outputs:

Text

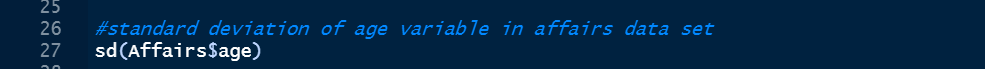
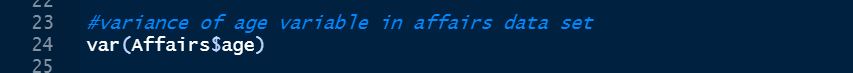
Description automatically generatedText

Description automatically generatedGraphical user interface, text, application

Description automatically generated

The mean married years of a person is 8.177696. Most of the persons are married 4 years ago. The middle-married years is 7. The middle value of the data distribution is 7.

Dispersion of age variable:



Console output:

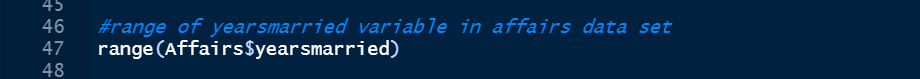
Text

Description automatically generated

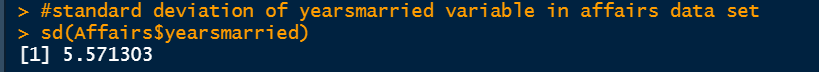
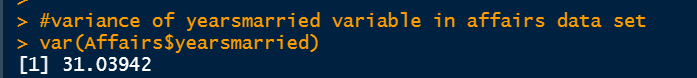
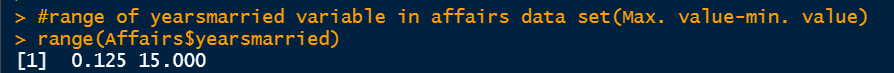
Range of age variable in affairs data set is 17.5 to 57 years. Which means ,the youngest person’s age of this data set is 17.5 years and the eldest person’s age is 57 years.

Variance of the age variable is 86.28109.Standard deviation is 9.288762.

Dispersion of yearsmarried variable:



Console outputs:



The minimum number of years married of a person is 0.125 and the maximum number of years married of a person is 15 years. Variance of years married of a person is 31.03942. Standard deviation of years married is 5.571303.

Data visualization:

Density plot for age variable:

The density plot shows the distribution of data over a continues time period. The values which are concentrated over the interval are shown by the peaks of the density plot. Density plots are better than histogram to identify the shape of the data distribution. Because, the density plot is not affected by the number of bins used.

Chart, histogram

Description automatically generated

Figure 02: Density plot for age variable in affairs data set

Above density plot for age variable shows that the data distribution of age variable is right skewed.

Density plot for yearsmarried variable:

Chart, line chart, histogram

Description automatically generated

Figure 03: Density plot for yearsmarried variable in affairs data set

Above density plot for yearsmarried variable shows that the data distribution of yearsmarried is left skewed.

References:

1. US EPA. 2022. *Exploratory Data Analysis | US EPA*. [online] Available at: <[https://www.epa.gov/caddis-vol4/exploratory-data-analysis#:~:text=References-,Overview,step%20in%20any%20data%20analysis](https://www.epa.gov/caddis-vol4/exploratory-data-analysis%23:~:text=References-,Overview,step%20in%20any%20data%20analysis).> [Accessed 11 April 2022].
2. O’Reilly Online Learning. 2022. *Practical Statistics for Data Scientists*. [online] Available at: <<https://www.oreilly.com/library/view/practical-statistics-for/9781491952955/ch01.html>> [Accessed 11 April 2022].
3. Frost, J., 2022. *Skewed Distribution*. [online] Statistics By Jim. Available at: <<https://statisticsbyjim.com/basics/skewed-distribution/>> [Accessed 11 April 2022].
4. Statistics.laerd.com. 2022. *Measures of Spread | How and when to use measures of spread | Laerd Statistics*. [online] Available at: <[https://statistics.laerd.com/statistical-guides/measures-of-spread-range-quartiles.php#:~:text=A%20measure%20of%20spread%2C%20sometimes,of%20a%20set%20of%20data.](https://statistics.laerd.com/statistical-guides/measures-of-spread-range-quartiles.php%23:~:text=A%20measure%20of%20spread%2C%20sometimes,of%20a%20set%20of%20data.)> [Accessed 11 April 2022].
5. SearchDataCenter. 2022. *What are Mean, Median, Mode and Range?*. [online] Available at: <[https://www.techtarget.com/searchdatacenter/definition/statistical-mean-median-mode-and-range#:~:text=The%20range%20is%20the%20difference%20between%20the%20highest%20and%20lowest,largest%20number%20in%20the%20set.](https://www.techtarget.com/searchdatacenter/definition/statistical-mean-median-mode-and-range%23:~:text=The%20range%20is%20the%20difference%20between%20the%20highest%20and%20lowest,largest%20number%20in%20the%20set.)> [Accessed 11 April 2022].
6. Indicative. 2022. *What Is Data Variability? Data Defined - Indicative*. [online] Available at: <[https://www.indicative.com/resource/data-variability/#:~:text=Data%20variability%20also%20known%20as,to%20other%20sets%20of%20data](https://www.indicative.com/resource/data-variability/%23:~:text=Data%20variability%20also%20known%20as,to%20other%20sets%20of%20data).> [Accessed 11 April 2022].
7. The BMJ | The BMJ: leading general medical journal. Research. Education. Comment. 2022. *2. Mean and standard deviation | The BMJ*. [online] Available at: <<https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/2-mean-and-standard-deviation>> [Accessed 11 April 2022].