Autistic Spectrum Disorder Screening Data for Adult

Technical Report · November 2019

DOI: 10.13140/RG.2.2.33192.93448

1 author:

Rui Sarmento

79 PUBLICATIONS **185** CITATIONS

SEE PROFILE

NOTE: DATASET INFO from following UCI webpage:
https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult

Data Set Name: **Autistic Spectrum Disorder Screening Data for Adult**

Abstract: Autistic Spectrum Disorder (ASD) is a neurodevelopment condition associated with significant healthcare costs, and early diagnosis can significantly reduce these. Unfortunately, waiting times for an ASD diagnosis are lengthy and procedures are not cost effective. The economic impact of autism and the increase in the number of ASD cases across the world reveals an urgent need for the development of easily implemented and effective screening methods. Therefore, a time-efficient and accessible ASD screening is imminent to help health professionals and inform individuals whether they should pursue formal clinical diagnosis. The rapid growth in the number of ASD cases worldwide necessitates datasets related to behaviour traits. However, such datasets are rare making it difficult to perform thorough analyses to improve the efficiency, sensitivity, specificity and predictive accuracy of the ASD screening process. Presently, very limited autism datasets associated with clinical or screening are available and most of them are genetic in nature. Hence, we propose a new dataset related to autism screening of adults that contained 20 features to be utilised for further analysis especially in determining influential autistic traits and improving the classification of ASD cases. In this dataset, we record ten behavioural features (AQ-10-Adult) plus ten individuals characteristics that have proved to be effective in detecting the ASD cases from controls in behaviour science.

Source: Fadi Fayez Thabtah
Department of Digital Technology
Manukau Institute of Technology,
Auckland, New Zealand
fadi.fayez@manukau.ac.nz

Data Type: Multivariate OR Univariate OR Sequential OR Time-Series OR Text OR Domain-Theory
Nominal / categorical, binary and continuous

Task: Classification

Attribute Type: Categorical, continuous and binary

Area: Medical, health and social science

Format Type: Non-Matrix

Does your data set contain missing values? Yes

Number of Instances (records in your data set): 704

Number of Attributes (fields within each record): 21

Relevant Information: For Further information about the attributes/feature see below table.

Attribute Information:

Table 1: Features and their descriptions

Attribute	Type	Description
Age	Number	Age in years
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with PDD	Boolean (yes or no)	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician ,etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean (yes or no)	Whether the user has used a screening app
Screening Method Type	Integer (0,1,2,3)	The type of screening methods chosen based on age category (0=toddler, 1=child, 2= adolescent, 3= adult)
Question 1 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 2 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 3 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 4 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 5 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 6 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Question 7 Answer	Binary (0, 1)	The answer code of the question based on the screening method used

Relevant Papers:

1) Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. Proceedings of the 1st International Conference on Medical and Health Informatics 2017, pp.1-6. Taichung City, Taiwan, ACM.
2) Thabtah, F. (2017). ASDTests. A mobile app for ASD screening. www.asdtests.com [accessed December 20th, 2017].
3) Thabtah, F. (2017). Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review. To Appear in Informatics for Health and Social Care Journal. December, 2017 (in press)

A short example of your data to analyze:

V1	V2	...	V20	V21
1	1	...	Self	NO
1	1	...	Self	NO
1	1	...	Parent	YES
1	1	...	Self	NO
1	0	...		NO
1	1	...	Self	YES

1. Descriptive Analysis for All Variables:

1.1. Variable "V1" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V1"

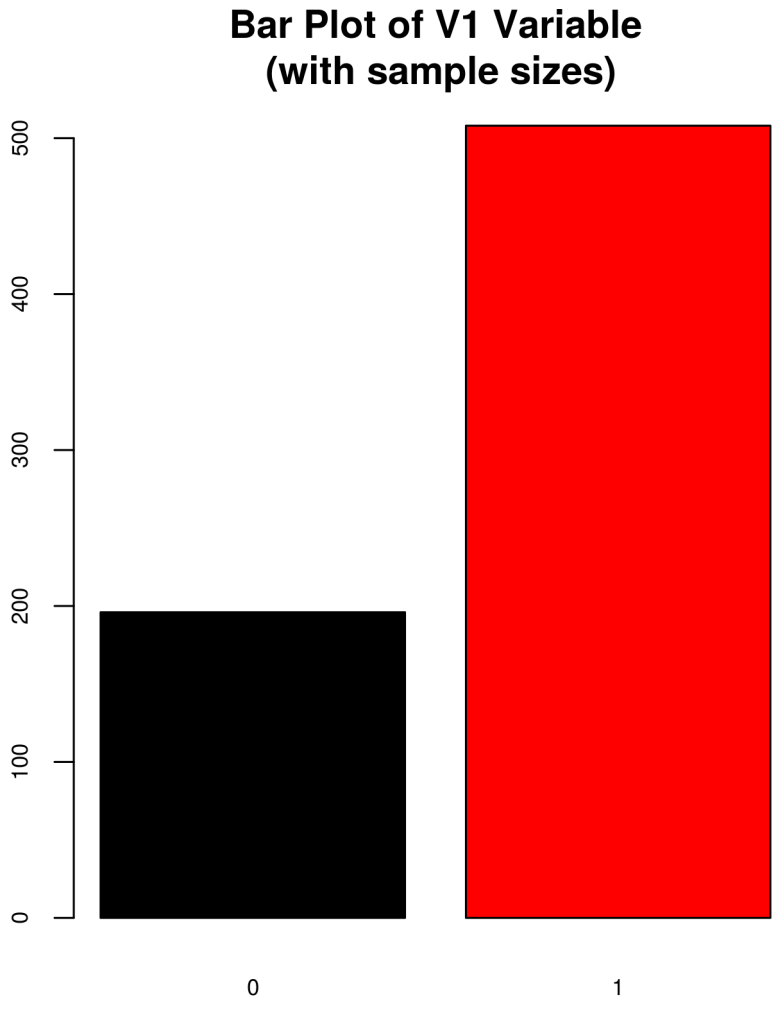
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 196 counts of category "0", corresponding to 27.8%. In total, there are 704 elements in the study.

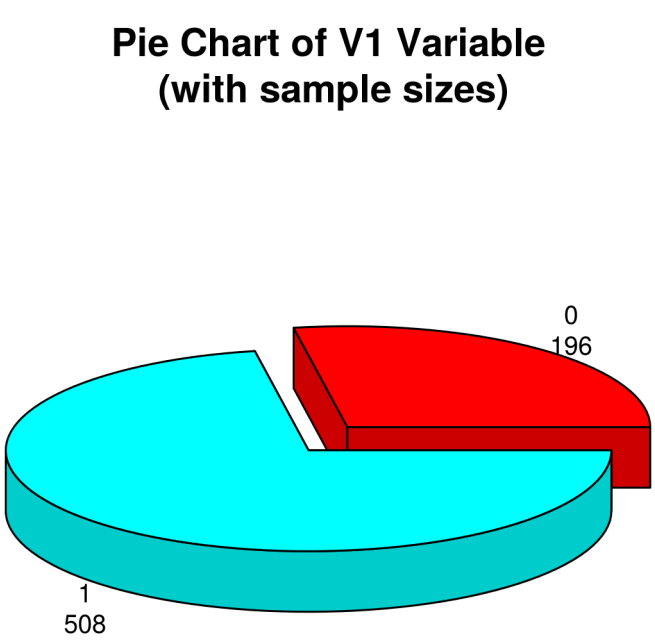
Analyses show that there are 508 counts of category "1", corresponding to 72.2%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 1, with 508 counts.

Bar Plot is:


Bar Plot of V1 Variable
(with sample sizes)
0
100
200
300
400
500
0
1

Pie Plot is:



1.2. Variable "V2" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V2"

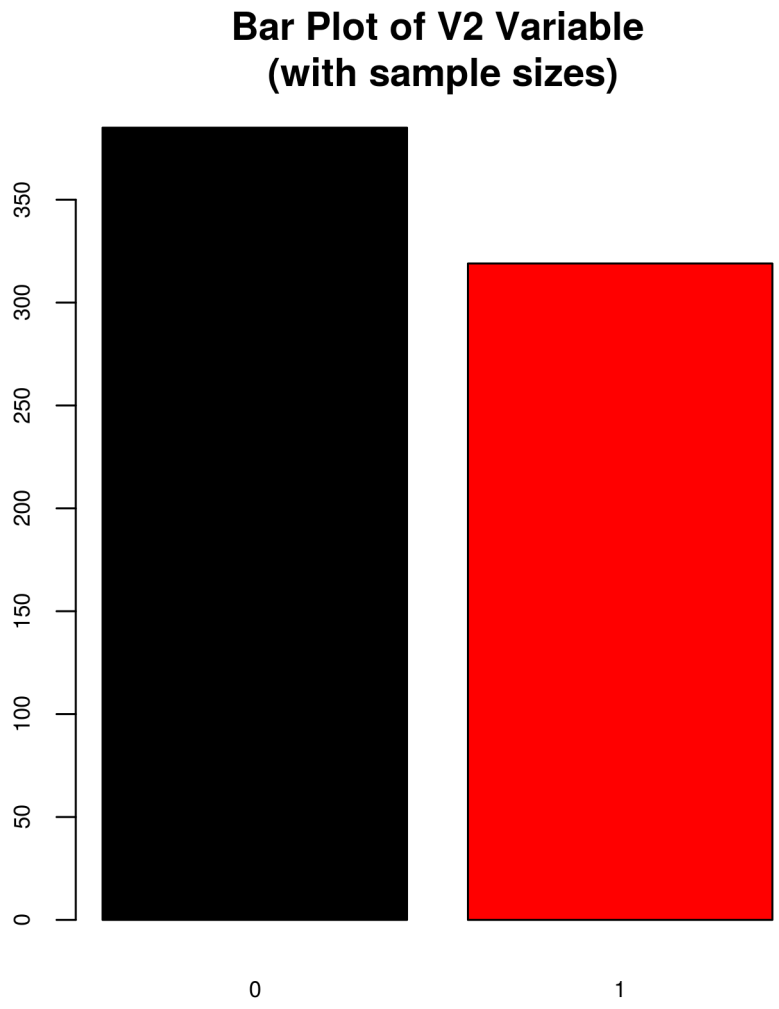
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 385 counts of category "0", corresponding to 54.7%. In total, there are 704 elements in the study.

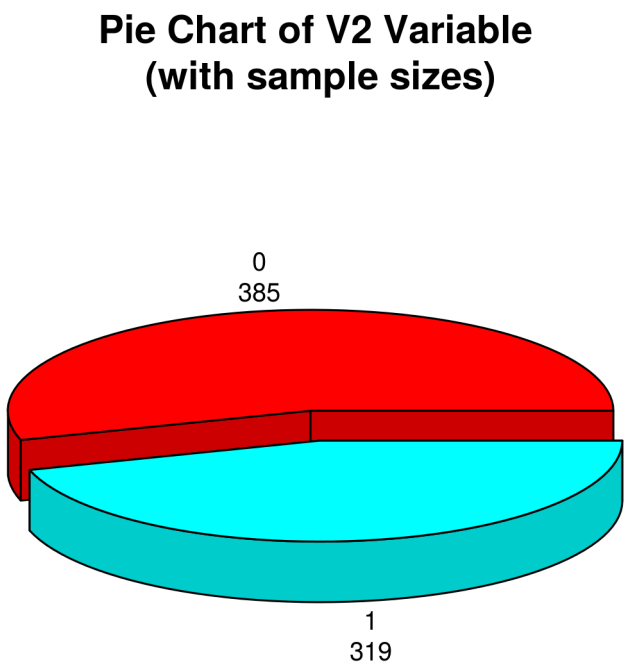
Analyses show that there are 319 counts of category "1", corresponding to 45.3%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 0, with 385 counts.

Bar Plot is:



Pie Plot is:



1.3. Variable "V3" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V3"

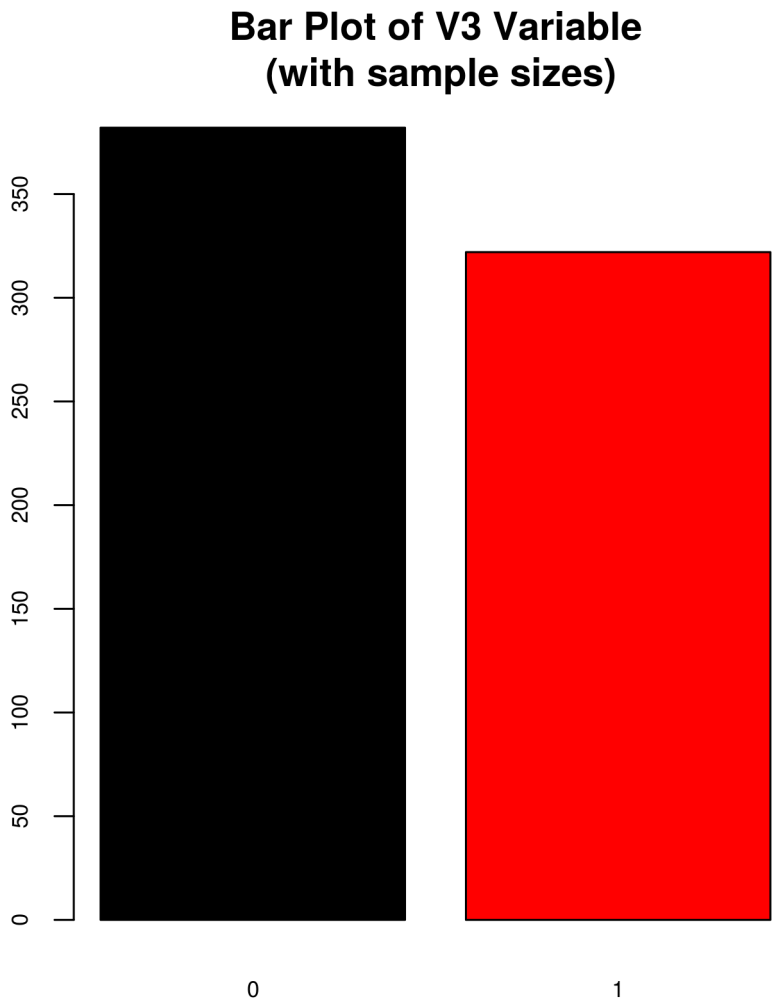
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 382 counts of category "0", corresponding to 54.3%. In total, there are 704 elements in the study.

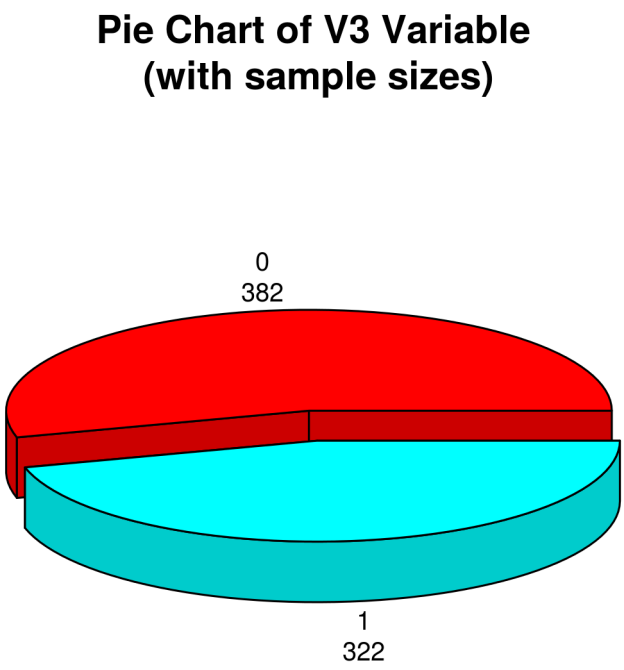
Analyses show that there are 322 counts of category "1", corresponding to 45.7%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 0, with 382 counts.

Bar Plot is:



Pie Plot is:



1.4. Variable "V4" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V4"

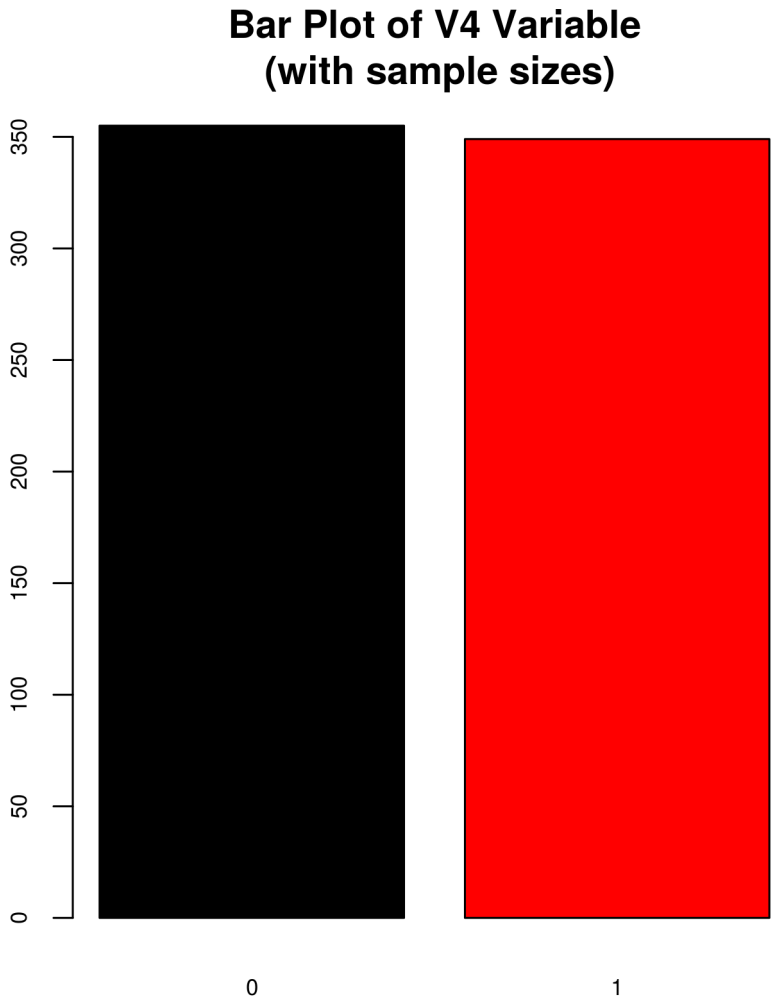
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 355 counts of category "0", corresponding to 50.4%. In total, there are 704 elements in the study.

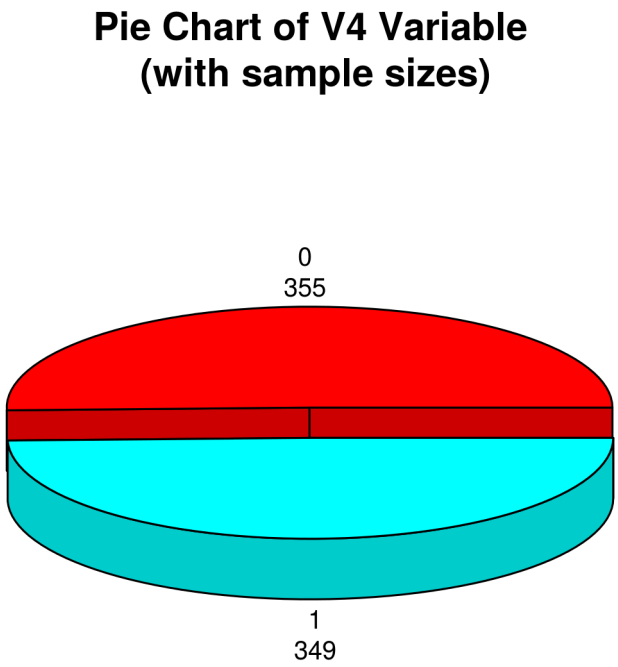
Analyses show that there are 349 counts of category "1", corresponding to 49.6%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 0, with 355 counts.

Bar Plot is:



Pie Plot is:



1.5. Variable "V5" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V5"

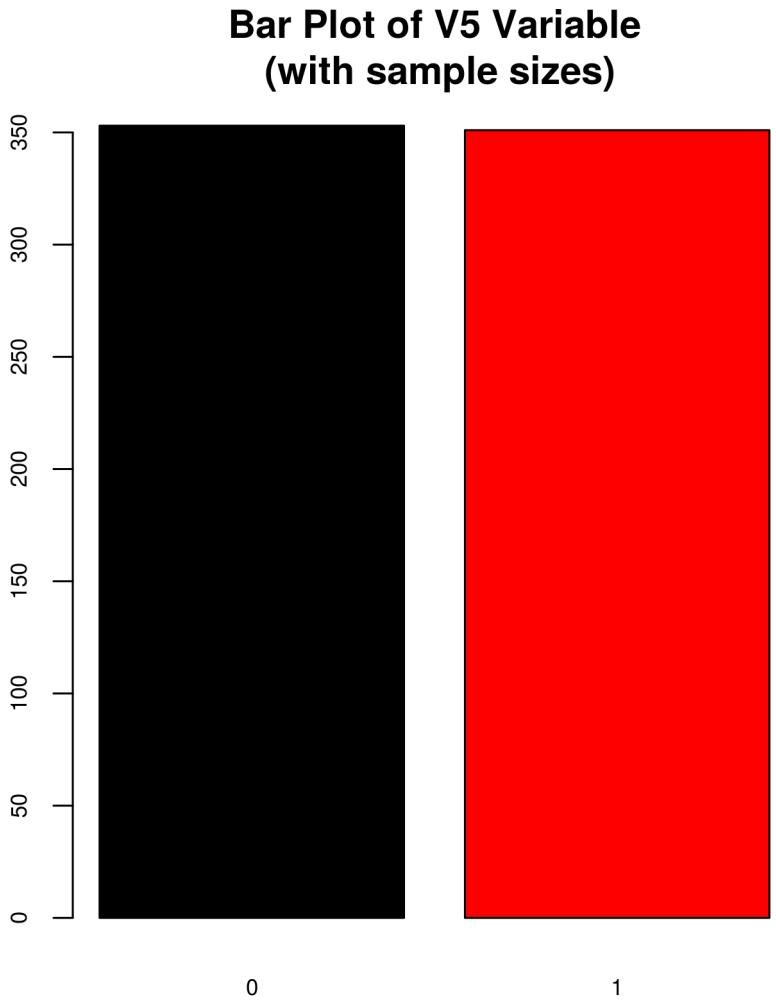
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 353 counts of category "0", corresponding to 50.1%. In total, there are 704 elements in the study.

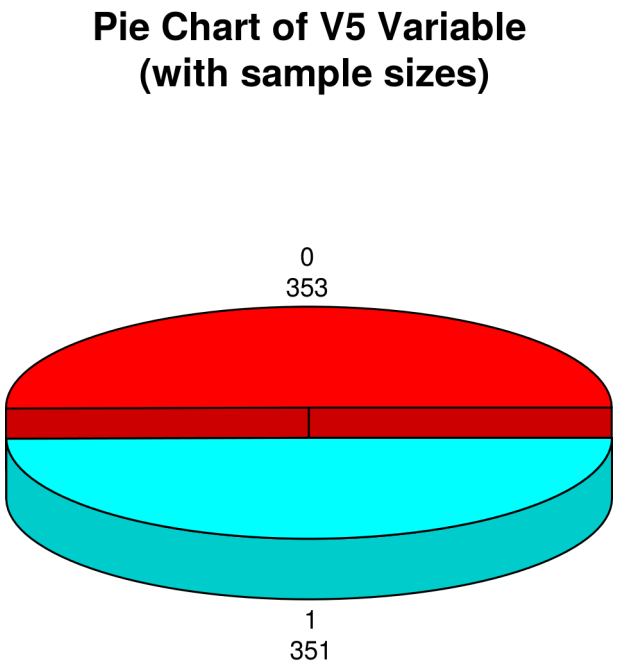
Analyses show that there are 351 counts of category "1", corresponding to 49.9%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 0, with 353 counts.

Bar Plot is:



Pie Plot is:



1.6. Variable "V6" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V6"

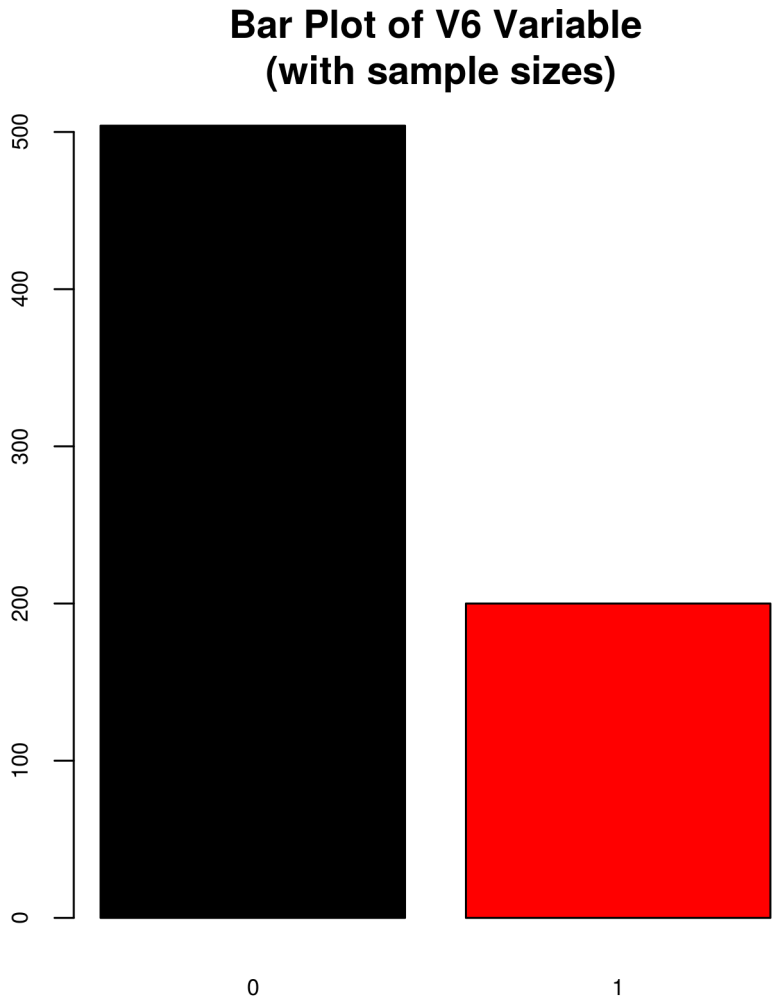
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 504 counts of category "0", corresponding to 71.6%. In total, there are 704 elements in the study.

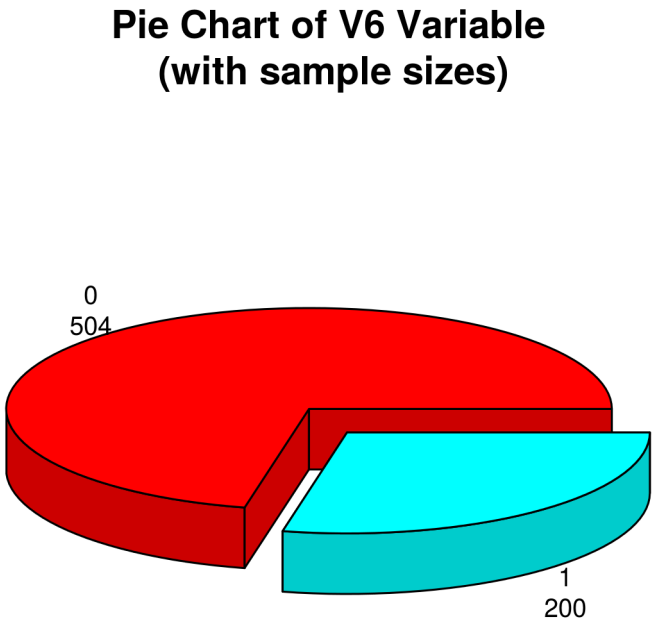
Analyses show that there are 200 counts of category "1", corresponding to 28.4%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 0, with 504 counts.

Bar Plot is:



Pie Plot is:



1.7. Variable "V7" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V7"

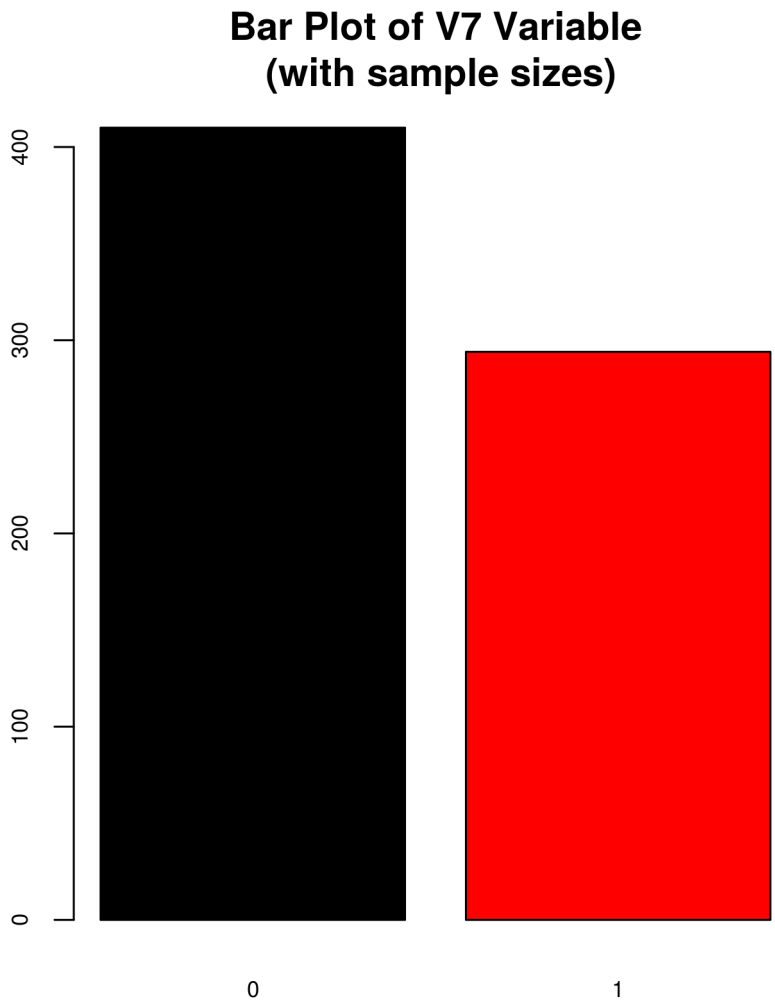
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 410 counts of category "0", corresponding to 58.2%. In total, there are 704 elements in the study.

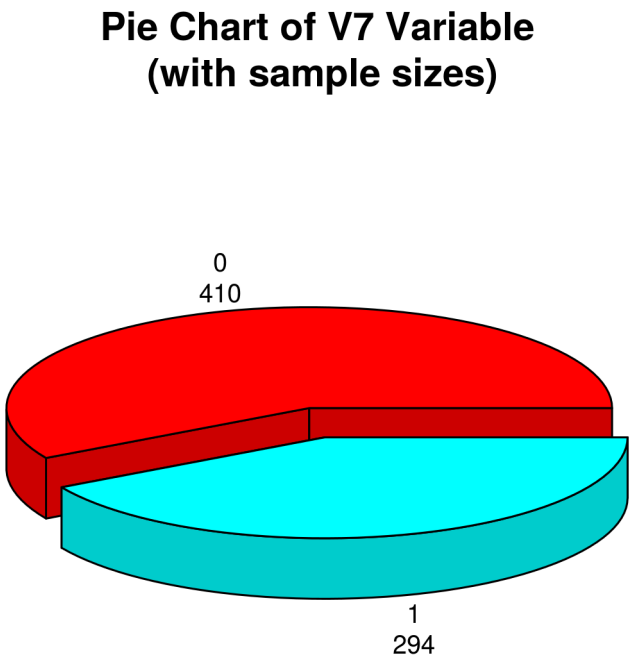
Analyses show that there are 294 counts of category "1", corresponding to 41.8%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 0, with 410 counts.

Bar Plot is:



Pie Plot is:



1.8. Variable "V8" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V8"

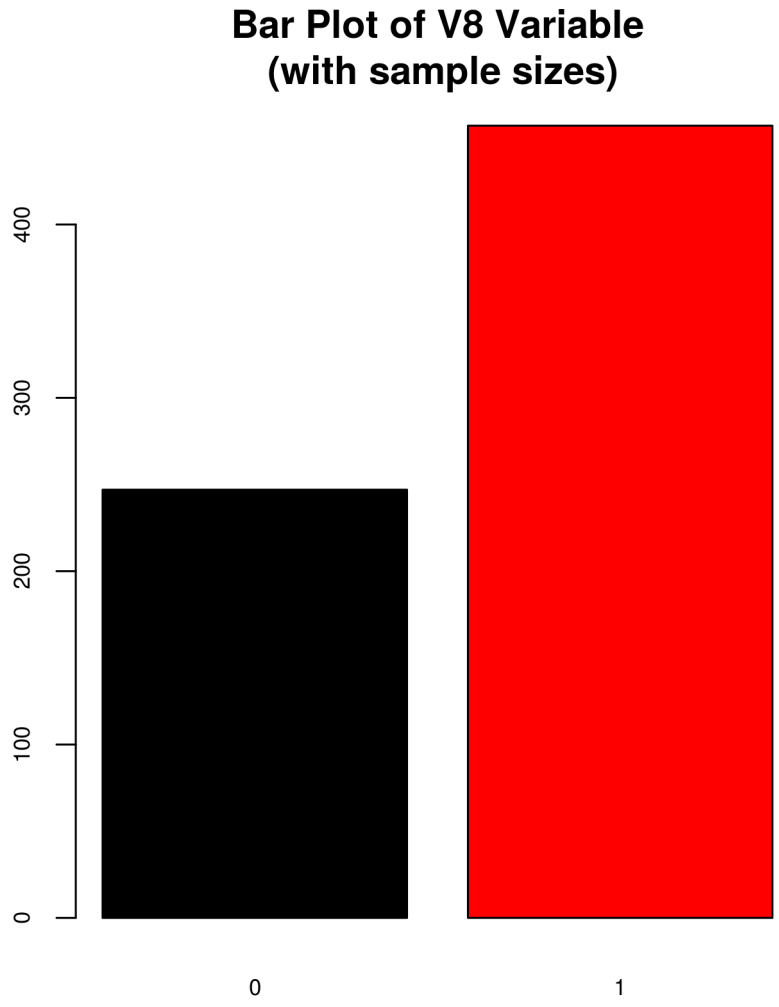
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 247 counts of category "0", corresponding to 35.1%. In total, there are 704 elements in the study.

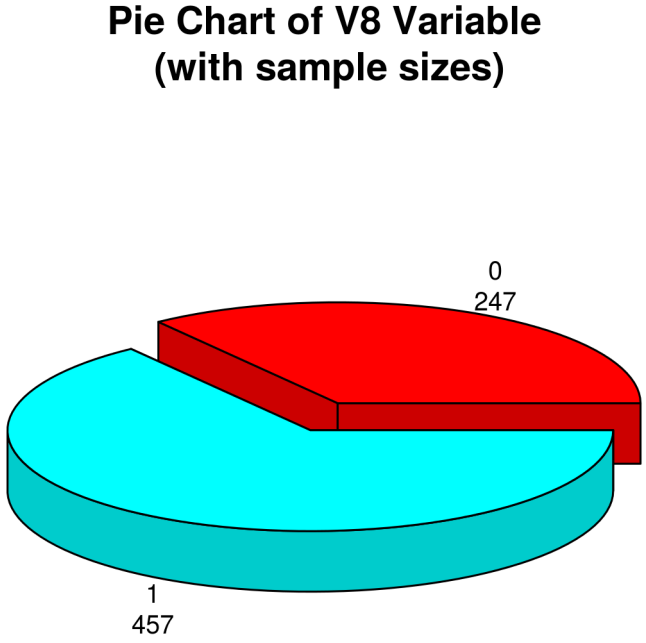
Analyses show that there are 457 counts of category "1", corresponding to 64.9%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 1, with 457 counts.

Bar Plot is:



Pie Plot is:



1.9. Variable "V9" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V9"

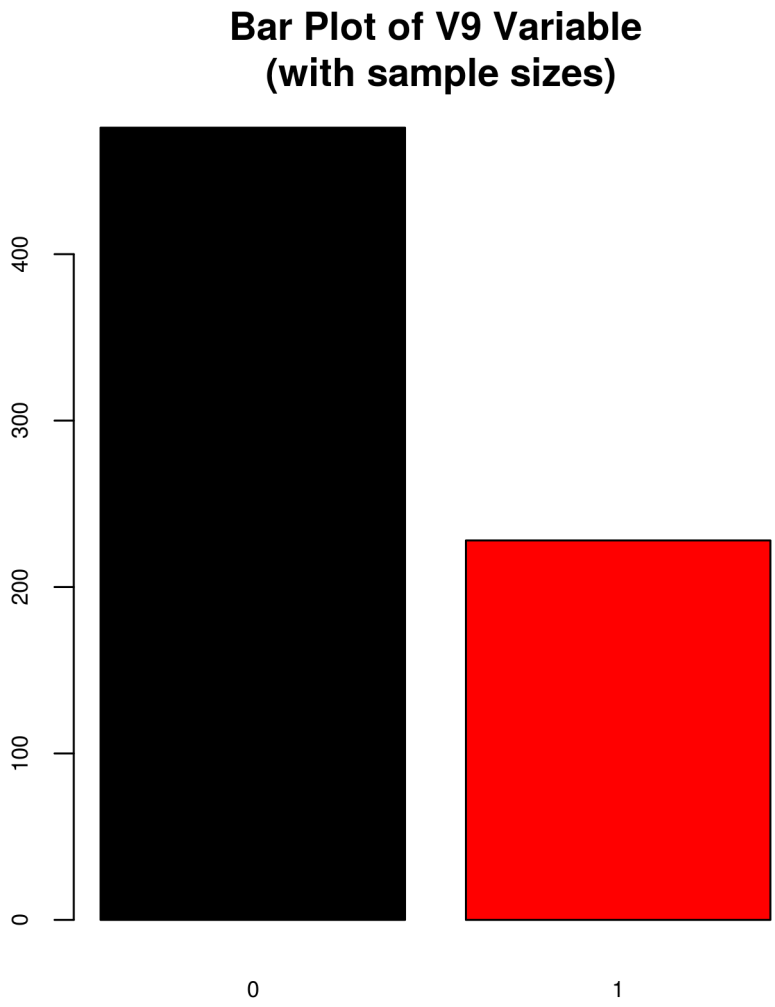
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 476 counts of category "0", corresponding to 67.6%. In total, there are 704 elements in the study.

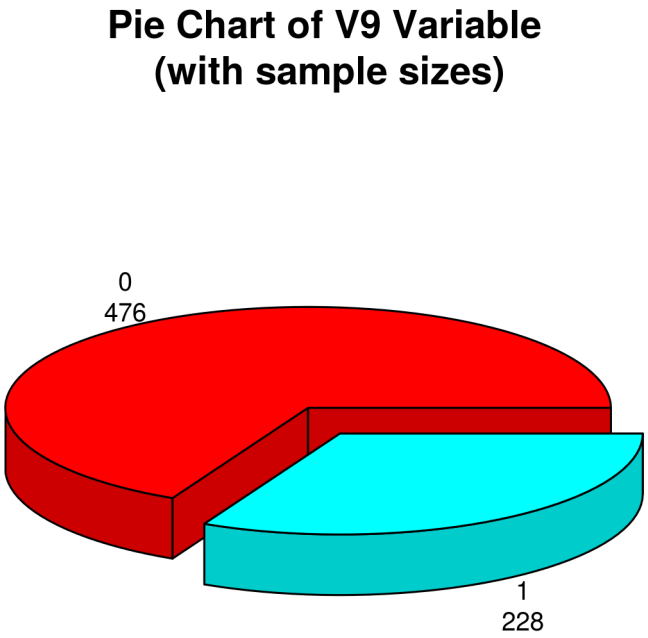
Analyses show that there are 228 counts of category "1", corresponding to 32.4%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 0, with 476 counts.

Bar Plot is:



Pie Plot is:



1.10. Variable "V10" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V10"

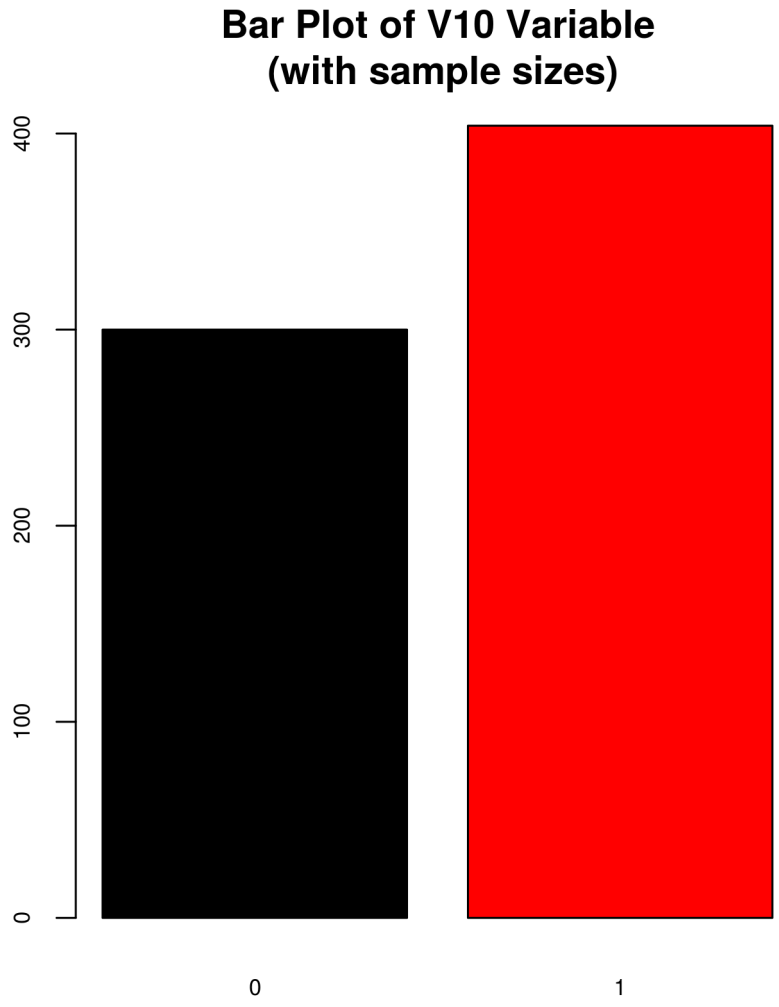
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 300 counts of category "0", corresponding to 42.6%. In total, there are 704 elements in the study.

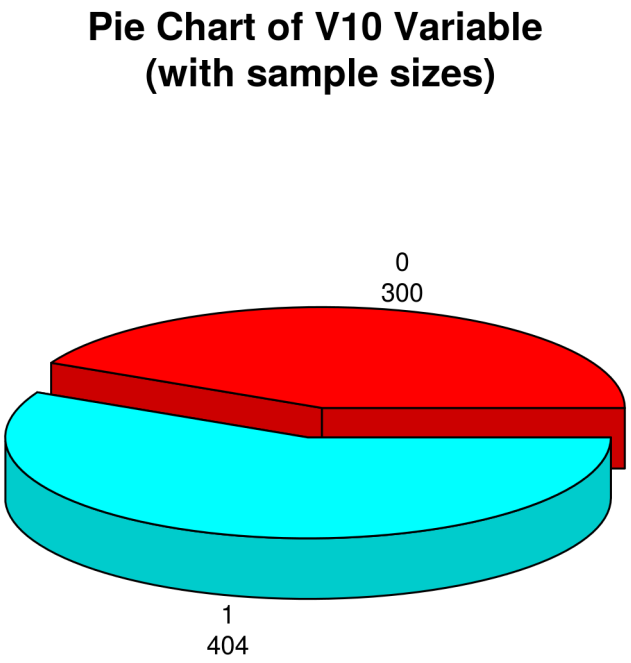
Analyses show that there are 404 counts of category "1", corresponding to 57.4%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is 1, with 404 counts.

Bar Plot is:



Pie Plot is:

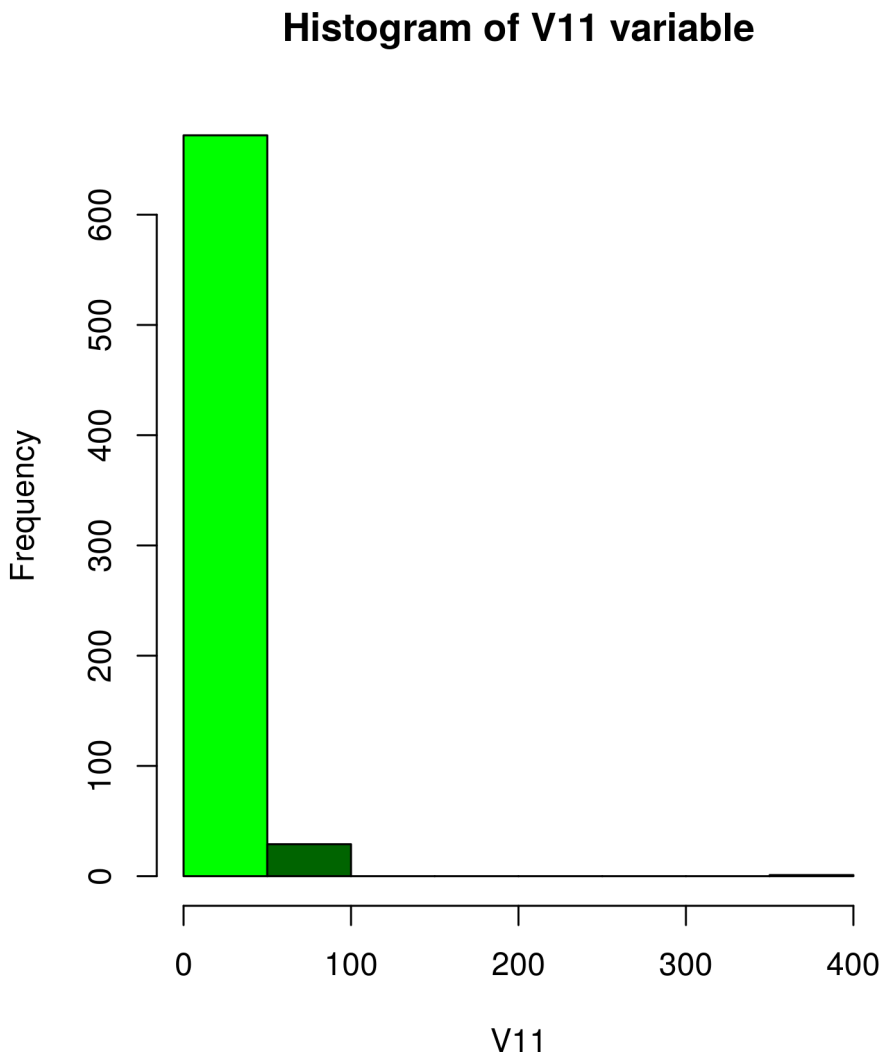


1.11. Variable "V11" - This is the Descriptive Analysis:

Results show that there are 2 missing values in variable "V11"

"V11" variable varies between 17 and 383. The mean is 29.698. The standard deviation is 16.507, that is, on average, the "V11" varies about 16.507 of the mean. The first and third quartiles are 21 and 35 respectively. This means that 50% (half) of elements of the sample have "V11" between 21 and 35.

Histogram Plot is:



1.12. Variable "V12" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V12"

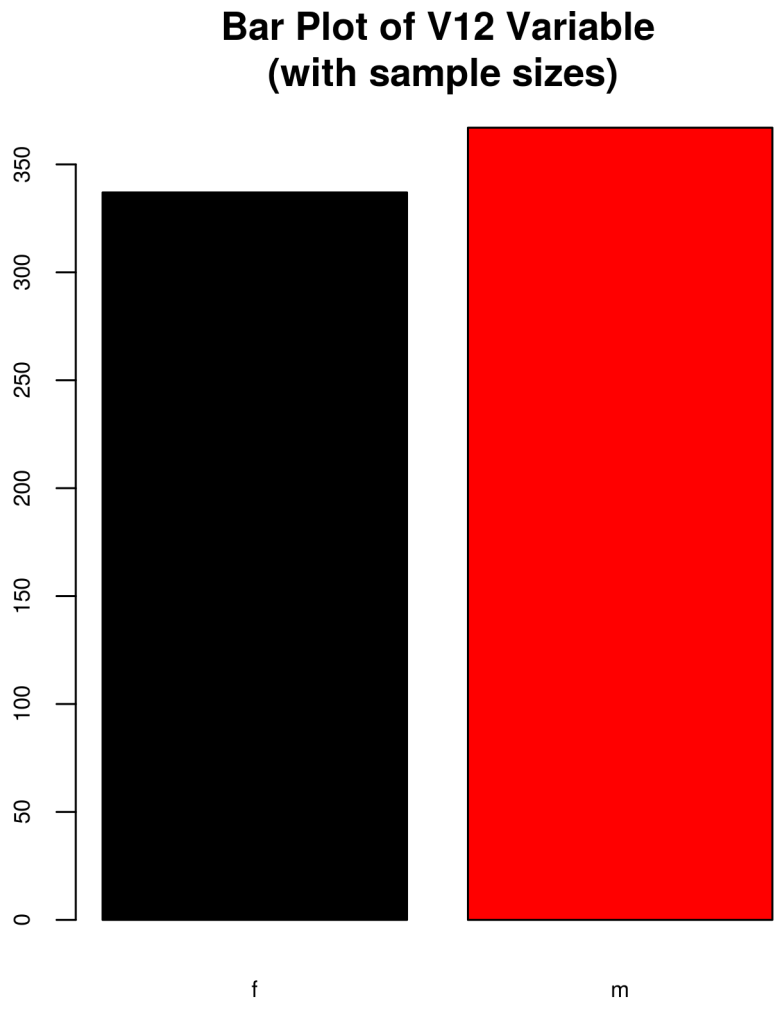
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 337 counts of category "f", corresponding to 47.9%. In total, there are 704 elements in the study.

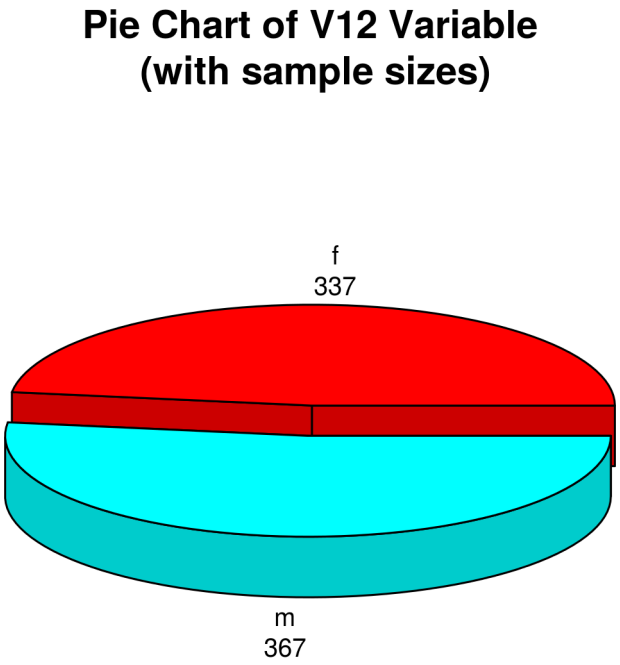
Analyses show that there are 367 counts of category "m", corresponding to 52.1%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is m, with 367 counts.

Bar Plot is:



Pie Plot is:



1.13. Variable "V13" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V13"

Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 95 counts of category "", corresponding to 13.5%. In total, there are 704 elements in the study.

Analyses show that there are 123 counts of category "Asian", corresponding to 17.5%. In total, there are 704 elements in the study.

Analyses show that there are 43 counts of category "Black", corresponding to 6.1%. In total, there are 704 elements in the study.

Analyses show that there are 13 counts of category "Hispanic", corresponding to 1.8%. In total, there are 704 elements in the study.

Analyses show that there are 20 counts of category "Latino", corresponding to 2.8%. In total, there are 704 elements in the study.

Analyses show that there are 92 counts of category "Middle Eastern ", corresponding to 13.1%. In total, there are 704 elements in the study.

Analyses show that there are 1 counts of category "others", corresponding to 0.1%. In total, there are 704 elements in the study.

Analyses show that there are 30 counts of category "Others", corresponding to 4.3%. In total, there are 704 elements in the study.

Analyses show that there are 12 counts of category "Pasifika", corresponding to 1.7%. In total, there are 704 elements in the study.

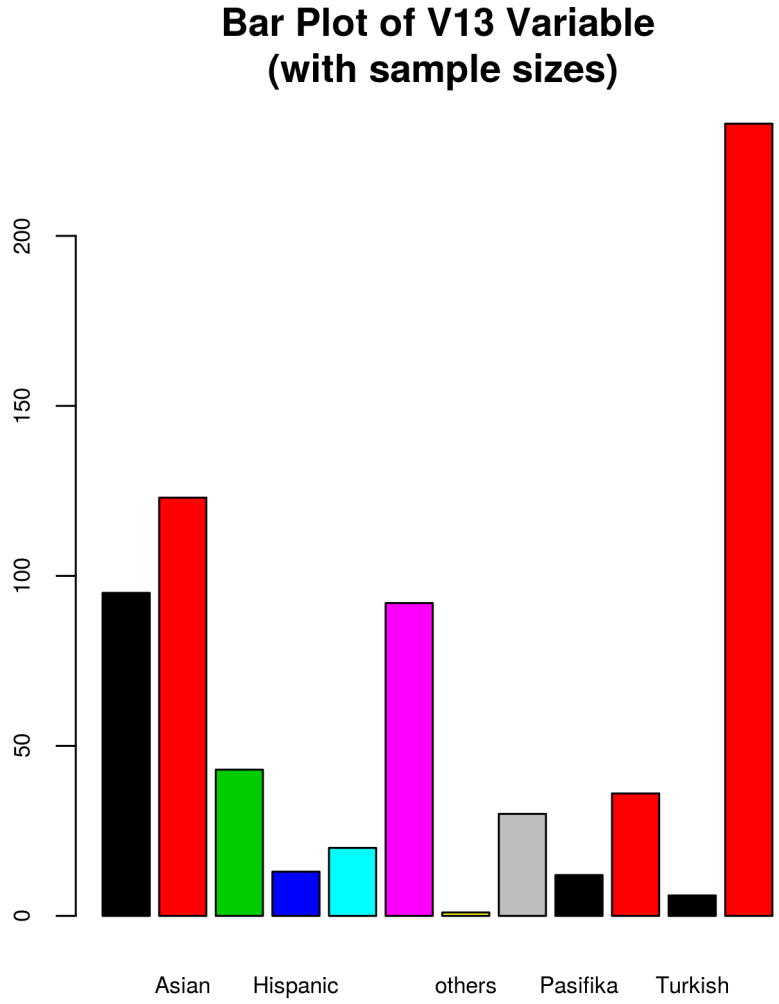
Analyses show that there are 36 counts of category "South Asian", corresponding to 5.1%. In total, there are 704 elements in the study.

Analyses show that there are 6 counts of category "Turkish", corresponding to 0.9%. In total, there are 704 elements in the study.

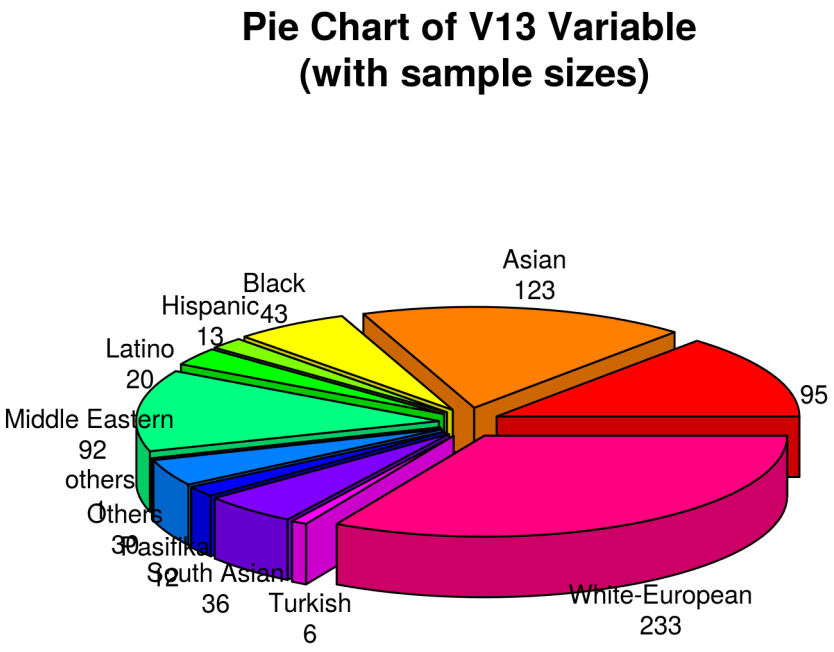
Analyses show that there are 233 counts of category "White-European", corresponding to 33.1%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is White-European, with 233 counts.

Bar Plot is:


Bar Plot of V13 Variable
(with sample sizes)
0
50
100
150
200
Asian
Hispanic
others
Pasifika
Turkish

Pie Plot is:



1.14. Variable "V14" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V14"

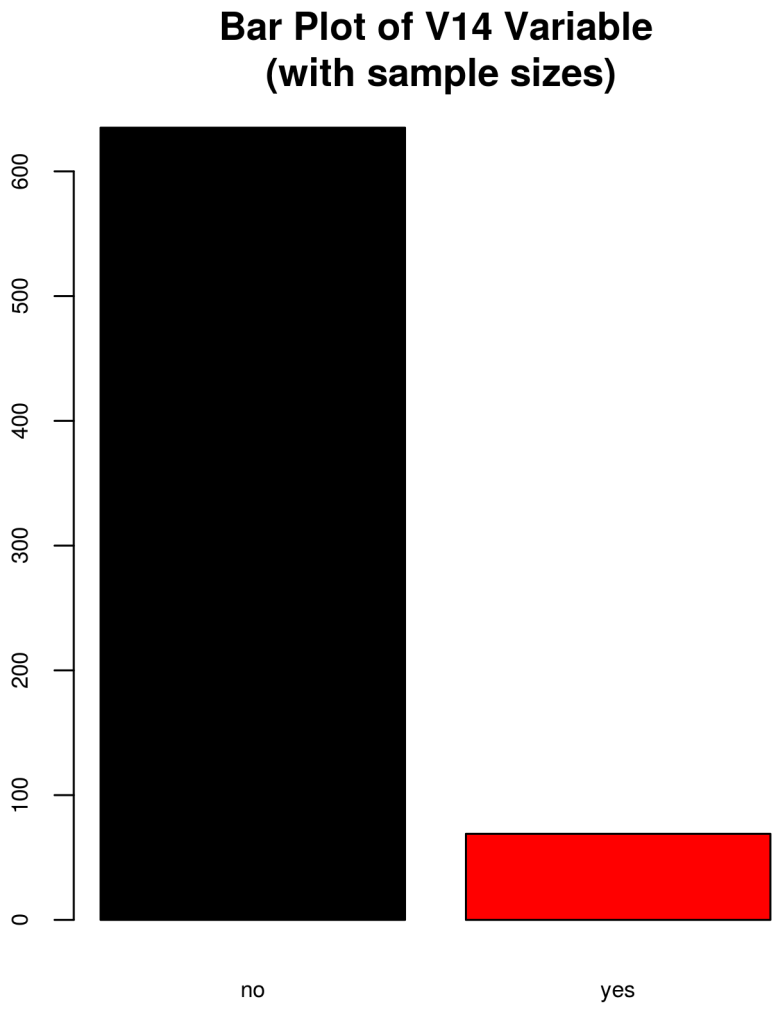
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 635 counts of category "no", corresponding to 90.2%. In total, there are 704 elements in the study.

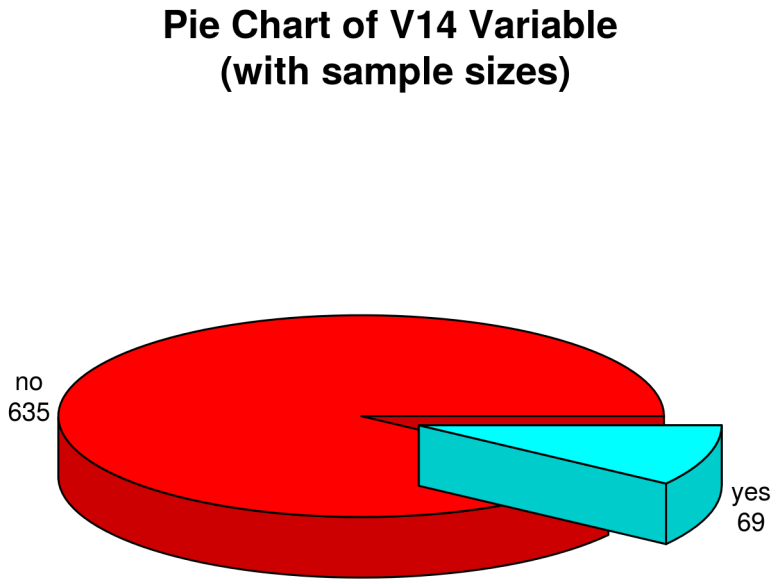
Analyses show that there are 69 counts of category "yes", corresponding to 9.8%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is no, with 635 counts.

Bar Plot is:



Pie Plot is:



1.15. Variable "V15" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V15"

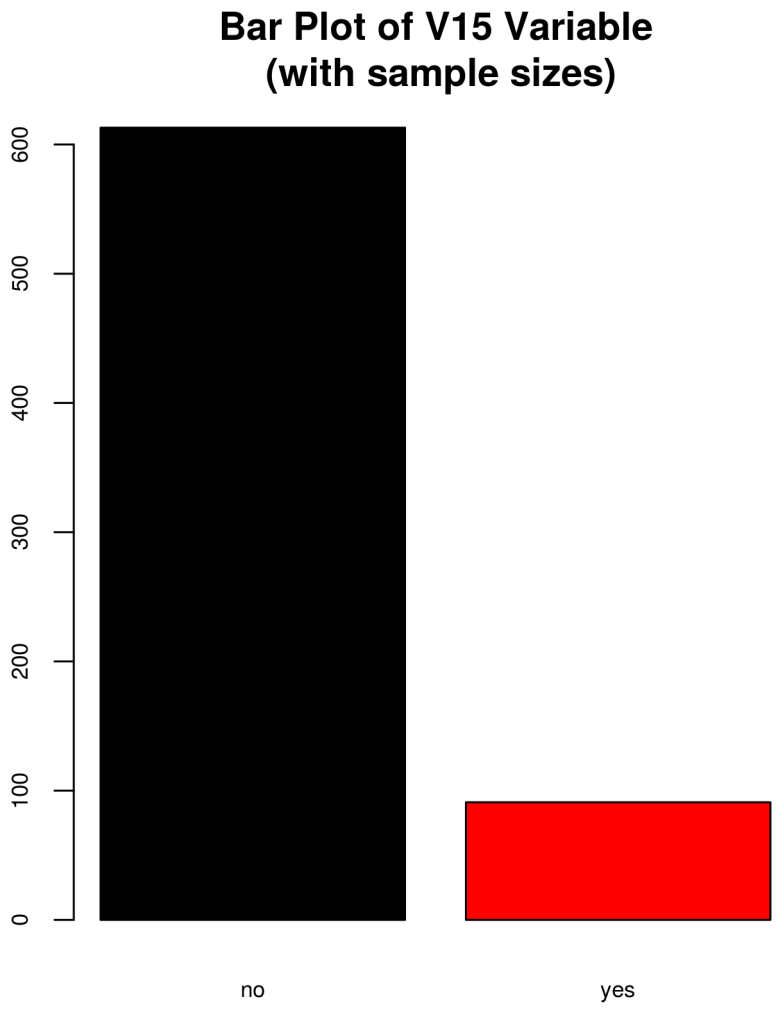
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 613 counts of category "no", corresponding to 87.1%. In total, there are 704 elements in the study.

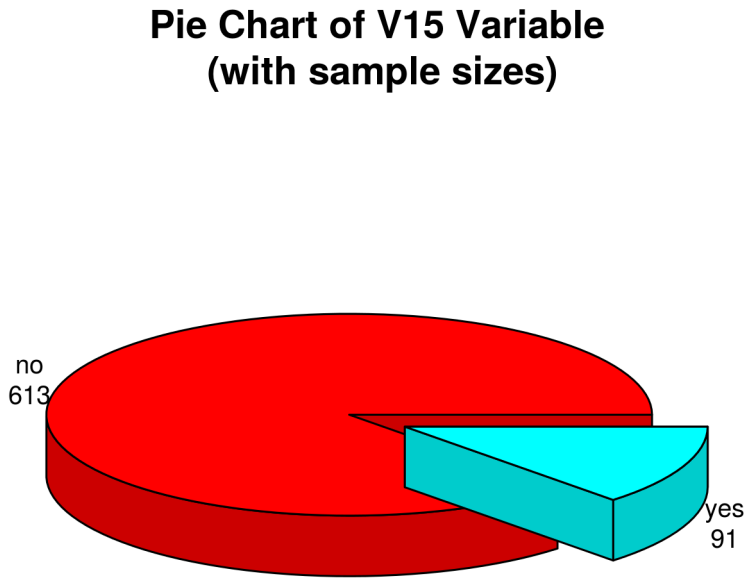
Analyses show that there are 91 counts of category "yes", corresponding to 12.9%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is no, with 613 counts.

Bar Plot is:



Pie Plot is:



1.16. Variable "V16" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V16"

Analysis not possible due to data constraints!

1.17. Variable "V17" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V17"

Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In

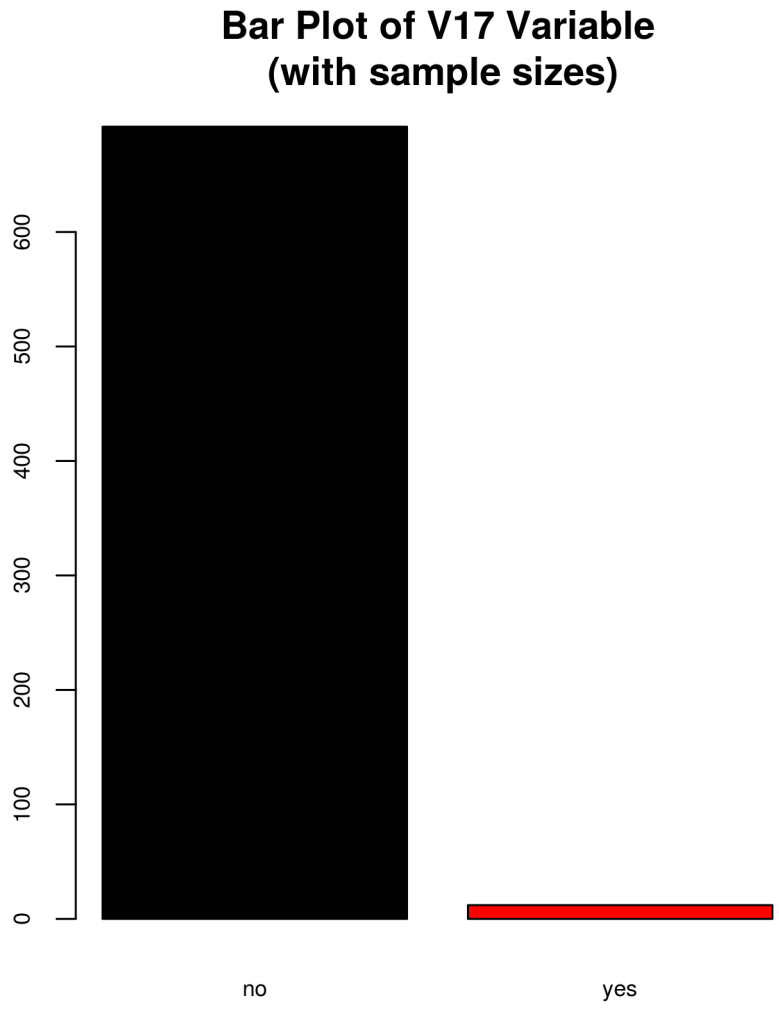
case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 692 counts of category "no", corresponding to 98.3%. In total, there are 704 elements in the study.

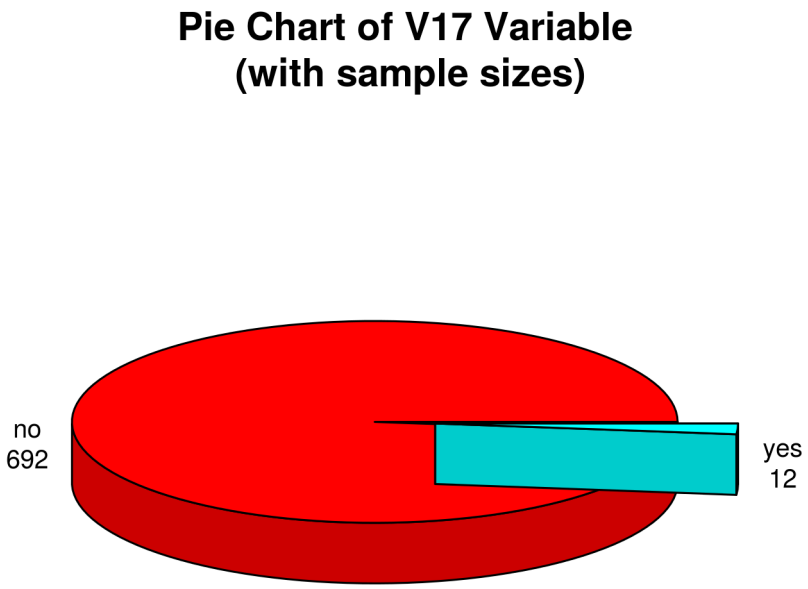
Analyses show that there are 12 counts of category "yes", corresponding to 1.7%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is no, with 692 counts.

Bar Plot is:



Pie Plot is:

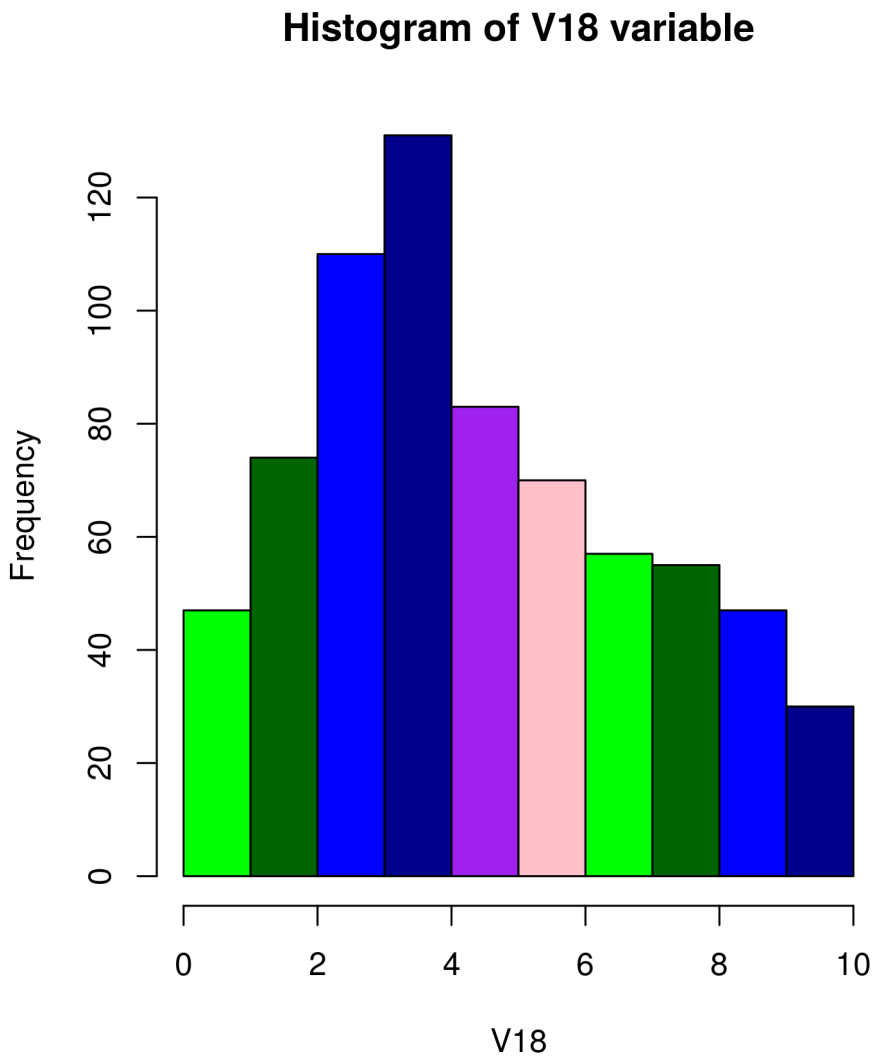


1.18. Variable "V18" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V18"

"V18" variable varies between 0 and 10. The mean is 4.875. The standard deviation is 2.501, that is, on average, the "V18" varies about 2.501 of the mean. The first and third quartiles are 3 and 7 respectively. This means that 50% (half) of elements of the sample have "V18" between 3 and 7.

Histogram Plot is:



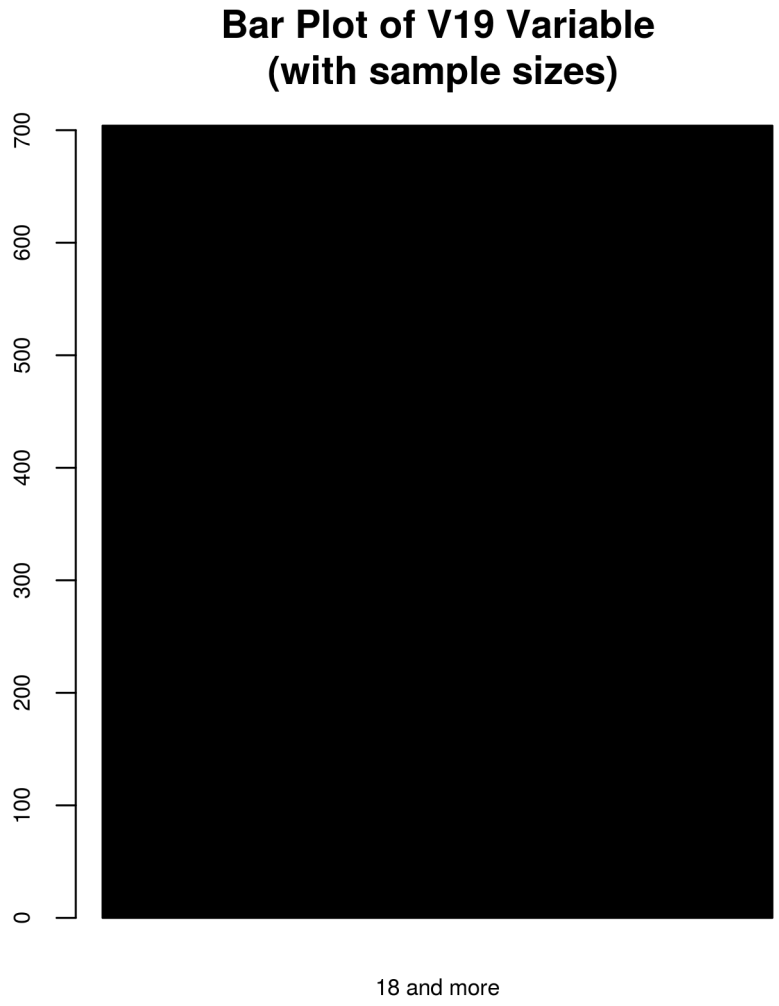
1.19. Variable "V19" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V19"

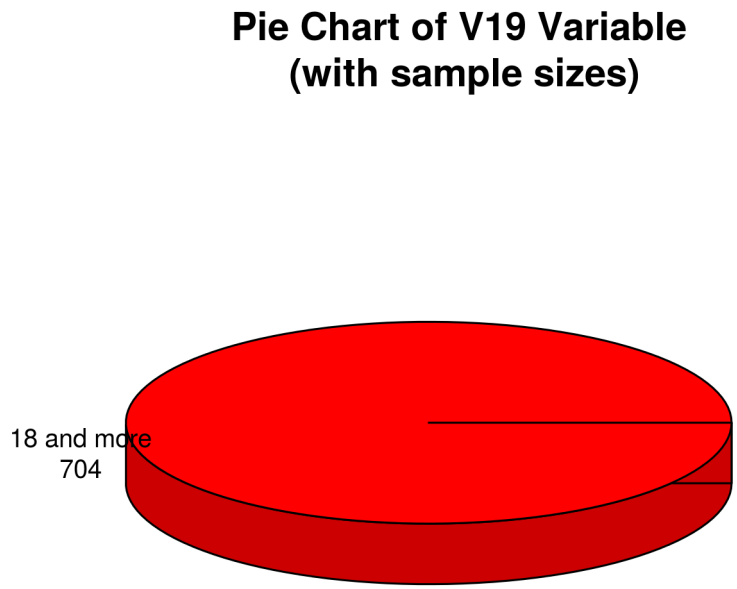
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 704 counts of category "18 and more", corresponding to 100%. In total, there are 704 elements in the study.

Bar Plot is:


Bar Plot of V19 Variable
(with sample sizes)
0
100
200
300
400
500
600
700
18 and more

Pie Plot is:



1.20. Variable "V20" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V20"

Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 95 counts of category "", corresponding to 13.5%. In total, there are 704 elements in the study.

Analyses show that there are 4 counts of category "Health care professional", corresponding to 0.6%. In total, there are 704 elements in the study.

Analyses show that there are 5 counts of category "Others", corresponding to 0.7%. In total, there are 704 elements in the study.

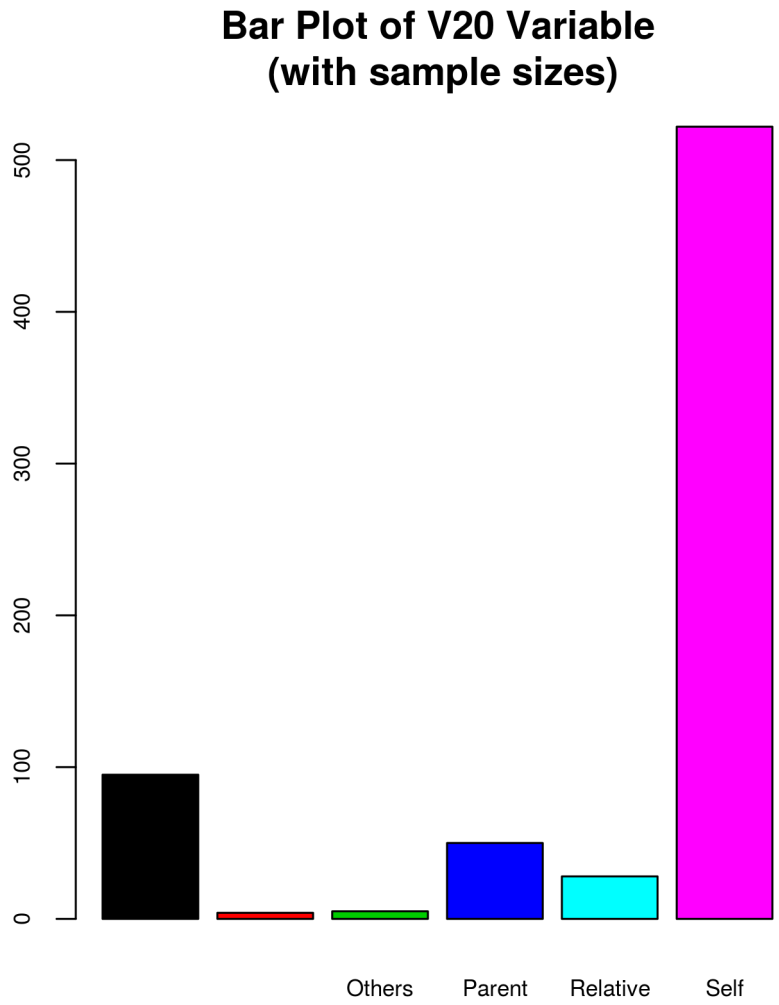
Analyses show that there are 50 counts of category "Parent", corresponding to 7.1%. In total, there are 704 elements in the study.

Analyses show that there are 28 counts of category "Relative", corresponding to 4%. In total, there are 704 elements in the study.

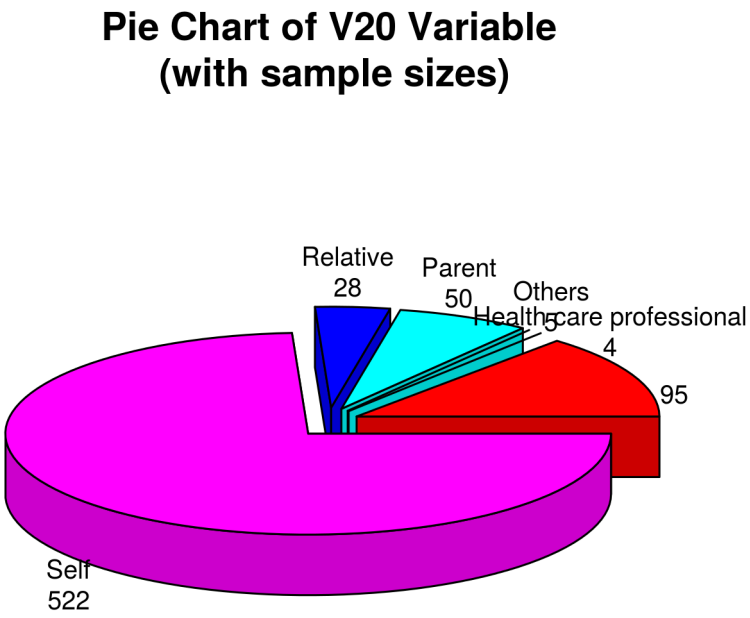
Analyses show that there are 522 counts of category "Self", corresponding to 74.1%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is Self, with 522 counts.

Bar Plot is:



Pie Plot is:



1.21. Variable "V21" - This is the Descriptive Analysis:

Results show that there are 0 missing values in variable "V21"

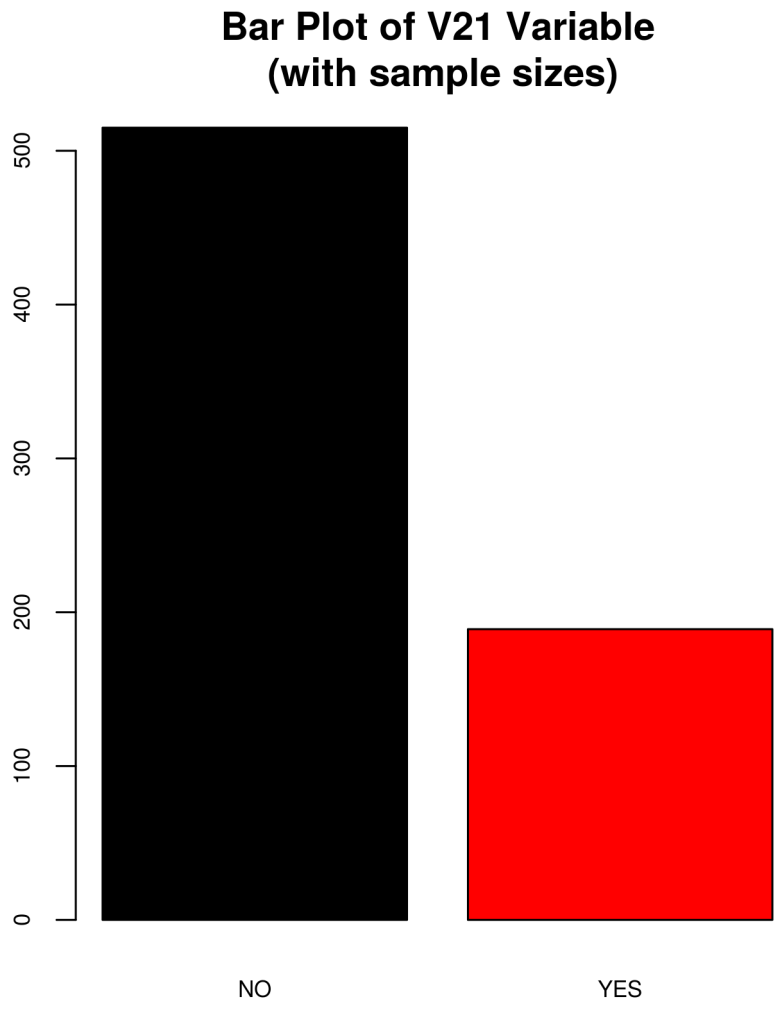
Categorical variables are qualitative variables and cannot be represented by a number. Categorical variables could be nominal and ordinal categorical variables. The difference is that, regarding their presentation, while nominal variables may be presented randomly or in the preferred order of the analyst, the ordinal variables must be presented in the order that is more easily understood (lowest to high, for example). In case of categorical variables, the analysis that can be done is the frequency of each category.

Analyses show that there are 515 counts of category "NO", corresponding to 73.2%. In total, there are 704 elements in the study.

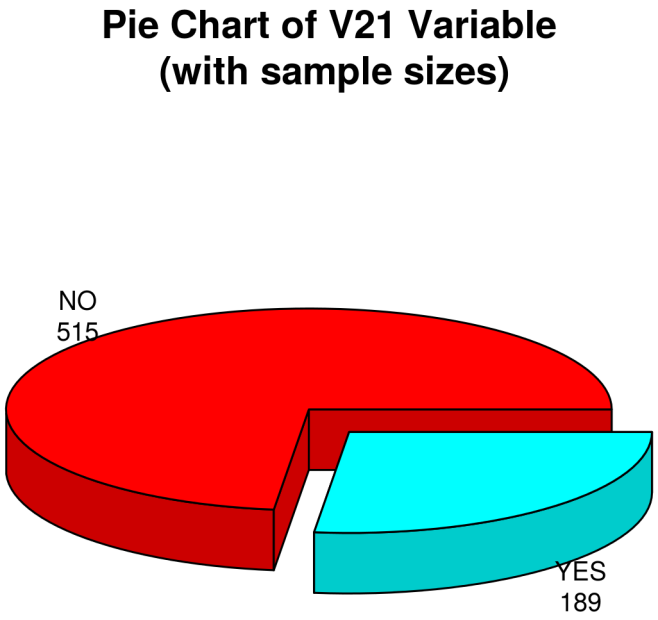
Analyses show that there are 189 counts of category "YES", corresponding to 26.8%. In total, there are 704 elements in the study.

The mode (most common element) of this variable is NO, with 515 counts.

Bar Plot is:



Pie Plot is:



2. Inference Analysis for All Variables Combination in Dataset:

An assessment of the normality of data is a prerequisite for many statistical tests because normal data is an underlying assumption in parametric tests. The variables used to test the normality need to be numerical. Additionally, to numerically test the normality, the length of the sample should be taken into account. If the length is smaller than 50 records we use “Shapiro-Wilk test” or else we use “Kolmogorov-Smirnov test”.

Thus, normality test was done for the numerical variables. The results of the tests used -- Shapiro-Wilk or Kolmogorov-Smirnov test - allows us concluding that, with a 95% of confidence, the null hypothesis is rejected if $p<0.05$/non-rejected if $p>0.05$, i.e., there is evidence/there is no evidence to reject the null hypothesis and it may not be/ be considered the existence of normality, respectively.

In this sense, the numeric variable has not/has normal distribution and, consequently, non-parametric/parametric tests should be used with that variable.

To compare two or more categorical variables, a cross-tabulation (also called the contingency table) is the most adequate option. However, to analyze the statistical differences, the chi-squared test or fisher test, for independence, should be applied to the crosstab.

To analyze two numerical variables, correlations analysis is applied. If variables have/have not normal distribution, Pearson's / Spearman's correlation is more appropriate, respectively.

2.1. Inference Analysis between Variable "V1" and "V2":

estimate	statistic	p.value	method	alternative	call
0.012	57481147.011	0.760	Spearman's rank correlation rho	two.sided	V1 and V2

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.76. This means that the independence of both variables exists. The value of the correlation is 0.012. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V2 increases.

2.2. Inference Analysis between Variable "V1" and "V3":

estimate	statistic	p.value	method	alternative	call
0.074	53843321.724	0.049	Spearman's rank correlation rho	two.sided	V1 and V3

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.049. This means that a significant correlation between both variables exists. The value of the correlation is 0.074. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V3 increases.

2.3. Inference Analysis between Variable "V1" and "V4":

estimate	statistic	p.value	method	alternative	call
0.128	50719486.433	0.001	Spearman's rank correlation rho	two.sided	V1 and V4

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.001. This means that a significant correlation between both variables exists. The value of the correlation is 0.128. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V4 increases.

2.4. Inference Analysis between Variable "V1" and "V5":

estimate	statistic	p.value	method	alternative	call
0.169	48302981.280	0.000	Spearman's rank correlation rho	two.sided	V1 and V5

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.169. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V5 increases.

2.5. Inference Analysis between Variable "V1" and "V6":

estimate	statistic	p.value	method	alternative	call
0.110	51743831.499	0.003	Spearman's rank correlation rho	two.sided	V1 and V6

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.003. This means that a significant correlation between both variables exists. The value of the correlation is 0.11. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V6 increases.

2.6. Inference Analysis between Variable "V1" and "V7":

estimate	statistic	p.value	method	alternative	call
0.218	45501859.105	0.000	Spearman's rank correlation rho	two.sided	V1 and V7

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.218. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V7 increases.

2.7. Inference Analysis between Variable "V1" and "V8":

estimate	statistic	p.value	method	alternative	call
0.148	49566588.259	0.000	Spearman's rank correlation rho	two.sided	V1 and V8

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.148. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V8 increases.

2.8. Inference Analysis between Variable "V1" and "V9":

estimate	statistic	p.value	method	alternative	call
0.145	49693786.662	0.000	Spearman's rank correlation rho	two.sided	V1 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.145. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V9 increases.

2.9. Inference Analysis between Variable "V1" and "V10":

estimate	statistic	p.value	method	alternative	call
0.118	51266189.191	0.002	Spearman's rank correlation rho	two.sided	V1 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.002. This means that a significant correlation between both variables exists. The value of the correlation is 0.118. Since the

correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V10 increases.

2.10. Inference Analysis between Variable "V1" and "V11":

estimate	statistic	p.value	method	alternative	call
-0.001	57688339.522	0.989	Spearman's rank correlation rho	two.sided	V1 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.989. This means that the independence of both variables exists. The value of the correlation is -0.001. Since the correlation result is negative, both variables are negatively correlated, i.e., when V1 increases, V11 decreases.

2.11. Inference Analysis between Variable "V1" and "V12":

statistic	p.value	method	alternative	data.name
66001.500	0.047	Wilcoxon rank sum test with continuity correction	two.sided	V1 by V12

Analyses show that p-value is 0.047 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V12" are very different regarding "V1" variable.

2.12. Inference Analysis between Variable "V1" and "V13":

statistic	p.value	parameter	method	data.name
33.291	0.000	11	Kruskal-Wallis rank sum test	V1 by V13

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V1" variable.

2.13. Inference Analysis between Variable "V1" and "V14":

statistic	p.value	method	alternative	data.name
22537.500	0.613	Wilcoxon rank sum test with continuity correction	two.sided	V1 by V14

Analyses show that p-value is 0.613 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V1" variable.

2.14. Inference Analysis between Variable "V1" and "V15":

statistic	p.value	method	alternative	data.name
24253.500	0.010	Wilcoxon rank sum test with continuity correction	two.sided	V1 by V15

Analyses show that p-value is 0.01 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V1" variable.

2.15. Inference Analysis between Variable "V1" and "V16":

statistic	p.value	parameter	method	data.name
121.646	0.000	66	Kruskal-Wallis rank sum test	V1 by V16

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V1" variable.

2.16. Inference Analysis between Variable "V1" and "V17":

statistic	p.value	method	alternative	data.name
4736.000	0.282	Wilcoxon rank sum test with continuity correction	two.sided	V1 by V17

Analyses show that p-value is 0.282 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V1" variable.

2.17. Inference Analysis between Variable "V1" and "V18":

estimate	statistic	p.value	method	alternative	call
0.395	35208294.796	0.000	Spearman's rank correlation rho	two.sided	V1 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.395. Since the correlation result is positive, both variables are positively correlated, i.e., when V1 increases, V18 increases.

2.18. Inference Analysis between Variable "V1" and "V19":

Analysis not possible due to data constraints!

2.19. Inference Analysis between Variable "V1" and "V20":

statistic	p.value	parameter	method	data.name
8.820	0.116	5	Kruskal-Wallis rank sum test	V1 by V20

Analyses show that p-value is 0.116 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, there are no statistical differences. Thus, it is possible to assume that groups have similar distribution regarding "V1" variable.

2.20. Inference Analysis between Variable "V1" and "V21":

statistic	p.value	method	alternative	data.name
34017.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V1 by V21

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V1" variable.

2.21. Inference Analysis between Variable "V2" and "V3":

estimate	statistic	p.value	method	alternative	call
0.224	45130674.919	0.000	Spearman's rank correlation rho	two.sided	V2 and V3

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.224. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V3 increases.

2.22. Inference Analysis between Variable "V2" and "V4":

estimate	statistic	p.value	method	alternative	call
0.159	48906087.749	0.000	Spearman's rank correlation rho	two.sided	V2 and V4

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.159. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V4 increases.

2.23. Inference Analysis between Variable "V2" and "V5":

estimate	statistic	p.value	method	alternative	call
0.154	49207146.134	0.000	Spearman's rank correlation rho	two.sided	V2 and V5

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.154. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V5 increases.

2.24. Inference Analysis between Variable "V2" and "V6":

estimate	statistic	p.value	method	alternative	call
0.186	47343784.786	0.000	Spearman's rank correlation rho	two.sided	V2 and V6

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.186. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V6 increases.

2.25. Inference Analysis between Variable "V2" and "V7":

estimate	statistic	p.value	method	alternative	call
-0.042	60581057.633	0.268	Spearman's rank correlation rho	two.sided	V2 and V7

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.268. This means that the independence of both variables exists. The value of the correlation is -0.042. Since the correlation result is negative, both variables are negatively correlated, i.e., when V2 increases, V7 decreases.

2.26. Inference Analysis between Variable "V2" and "V8":

estimate	statistic	p.value	method	alternative	call
0.035	56093116.779	0.348	Spearman's rank correlation rho	two.sided	V2 and V8

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.348. This means that the independence of both variables exists. The value of the correlation is 0.035. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V8 increases.

2.27. Inference Analysis between Variable "V2" and "V9":

estimate	statistic	p.value	method	alternative	call
0.205	46206485.238	0.000	Spearman's rank correlation rho	two.sided	V2 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.205. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V9 increases.

2.28. Inference Analysis between Variable "V2" and "V10":

estimate	statistic	p.value	method	alternative	call
0.069	54146492.671	0.068	Spearman's rank correlation rho	two.sided	V2 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.068. This means that the independence of both variables exists. The value of the correlation is 0.069. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V10 increases.

2.29. Inference Analysis between Variable "V2" and "V11":

estimate	statistic	p.value	method	alternative	call
0.059	54279266.651	0.121	Spearman's rank correlation rho	two.sided	V2 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.121. This means that the independence of both variables exists. The value of the correlation is 0.059. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V11 increases.

2.30. Inference Analysis between Variable "V2" and "V12":

statistic	p.value	method	alternative	data.name
64760.000	0.209	Wilcoxon rank sum test with continuity correction	two.sided	V2 by V12

Analyses show that p-value is 0.209 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V2" variable.

2.31. Inference Analysis between Variable "V2" and "V13":

statistic	p.value	parameter	method	data.name
49.637	0.000	11	Kruskal-Wallis rank sum test	V2 by V13

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V2" variable.

2.32. Inference Analysis between Variable "V2" and "V14":

statistic	p.value	method	alternative	data.name
17777.000	0.003	Wilcoxon rank sum test with continuity correction	two.sided	V2 by V14

Analyses show that p-value is 0.003 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V14" are very different regarding "V2" variable.

2.33. Inference Analysis between Variable "V2" and "V15":

statistic	p.value	method	alternative	data.name
24806.000	0.048	Wilcoxon rank sum test with continuity correction	two.sided	V2 by V15

Analyses show that p-value is 0.048 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V2" variable.

2.34. Inference Analysis between Variable "V2" and "V16":

statistic	p.value	parameter	method	data.name
124.955	0.000	66	Kruskal-Wallis rank sum test	V2 by V16

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V2" variable.

2.35. Inference Analysis between Variable "V2" and "V17":

statistic	p.value	method	alternative	data.name
4306.000	0.799	Wilcoxon rank sum test with continuity correction	two.sided	V2 by V17

Analyses show that p-value is 0.799 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V2" variable.

2.36. Inference Analysis between Variable "V2" and "V18":

estimate	statistic	p.value	method	alternative	call
0.374	36418831.572	0.000	Spearman's rank correlation rho	two.sided	V2 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.374. Since the correlation result is positive, both variables are positively correlated, i.e., when V2 increases, V18 increases.

2.37. Inference Analysis between Variable "V2" and "V19":

Analysis not possible due to data constraints!

2.38. Inference Analysis between Variable "V2" and "V20":

statistic	p.value	parameter	method	data.name
9.520	0.090	5	Kruskal-Wallis rank sum test	V2 by V20

Analyses show that p-value is 0.09 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, there are no statistical differences. Thus, it is possible to assume that groups have similar distribution regarding "V2" variable.

2.39. Inference Analysis between Variable "V2" and "V21":

statistic	p.value	method	alternative	data.name
31645.000	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V2 by V21

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V2" variable.

2.40. Inference Analysis between Variable "V3" and "V4":

estimate	statistic	p.value	method	alternative	call
0.413	34151480.815	0.000	Spearman's rank correlation rho	two.sided	V3 and V4

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.413. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V4 increases.

2.41. Inference Analysis between Variable "V3" and "V5":

estimate	statistic	p.value	method	alternative	call
0.265	42746058.495	0.000	Spearman's rank correlation rho	two.sided	V3 and V5

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.265. Since the correlation

result is positive, both variables are positively correlated, i.e., when V3 increases, V5 increases.

2.42. Inference Analysis between Variable "V3" and "V6":

estimate	statistic	p.value	method	alternative	call
0.269	42518173.321	0.000	Spearman's rank correlation rho	two.sided	V3 and V6

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.269. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V6 increases.

2.43. Inference Analysis between Variable "V3" and "V7":

estimate	statistic	p.value	method	alternative	call
0.078	53603745.591	0.038	Spearman's rank correlation rho	two.sided	V3 and V7

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.038. This means that a significant correlation between both variables exists. The value of the correlation is 0.078. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V7 increases.

2.44. Inference Analysis between Variable "V3" and "V8":

estimate	statistic	p.value	method	alternative	call
0.018	57118741.257	0.638	Spearman's rank correlation rho	two.sided	V3 and V8

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.638. This means that the independence of both variables exists. The value of the correlation is 0.018. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V8 increases.

2.45. Inference Analysis between Variable "V3" and "V9":

estimate	statistic	p.value	method	alternative	call
0.315	39827659.041	0.000	Spearman's rank correlation rho	two.sided	V3 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.315. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V9 increases.

2.46. Inference Analysis between Variable "V3" and "V10":

estimate	statistic	p.value	method	alternative	call
0.168	48356200.359	0.000	Spearman's rank correlation rho	two.sided	V3 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.168. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V10 increases.

2.47. Inference Analysis between Variable "V3" and "V11":

estimate	statistic	p.value	method	alternative	call
0.070	53611873.949	0.063	Spearman's rank correlation rho	two.sided	V3 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.063. This means that the independence of both variables exists. The value of the correlation is 0.07. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V11 increases.

2.48. Inference Analysis between Variable "V3" and "V12":

statistic	p.value	method	alternative	data.name
61790.500	0.983	Wilcoxon rank sum test with continuity correction	two.sided	V3 by V12

Analyses show that p-value is 0.983 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V3" variable.

2.49. Inference Analysis between Variable "V3" and "V13":

statistic	p.value	parameter	method	data.name
51.973	0.000	11	Kruskal-Wallis rank sum test	V3 by V13

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V3" variable.

2.50. Inference Analysis between Variable "V3" and "V14":

statistic	p.value	method	alternative	data.name
19640.500	0.102	Wilcoxon rank sum test with continuity correction	two.sided	V3 by V14

Analyses show that p-value is 0.102 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V3" variable.

2.51. Inference Analysis between Variable "V3" and "V15":

statistic	p.value	method	alternative	data.name
23182.500	0.003	Wilcoxon rank sum test with continuity correction	two.sided	V3 by V15

Analyses show that p-value is 0.003 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V3" variable.

2.52. Inference Analysis between Variable "V3" and "V16":

statistic	p.value	parameter	method	data.name
113.182	0.000	66	Kruskal-Wallis rank sum test	V3 by V16

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V3" variable.

2.53. Inference Analysis between Variable "V3" and "V17":

statistic	p.value	method	alternative	data.name
3268.000	0.143	Wilcoxon rank sum test with continuity correction	two.sided	V3 by V17

Analyses show that p-value is 0.143 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V3" variable.

2.54. Inference Analysis between Variable "V3" and "V18":

estimate	statistic	p.value	method	alternative	call
0.547	26321393.110	0.000	Spearman's rank correlation rho	two.sided	V3 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.547. Since the correlation result is positive, both variables are positively correlated, i.e., when V3 increases, V18 increases.

2.55. Inference Analysis between Variable "V3" and "V19":

Analysis not possible due to data constraints!

2.56. Inference Analysis between Variable "V3" and "V20":

statistic	p.value	parameter	method	data.name
21.737	0.001	5	Kruskal-Wallis rank sum test	V3 by V20

Analyses show that p-value is 0.001 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V3" variable.

2.57. Inference Analysis between Variable "V3" and "V21":

statistic	p.value	method	alternative	data.name
24536.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V3 by V21

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V3" variable.

2.58. Inference Analysis between Variable "V4" and "V5":

estimate	statistic	p.value	method	alternative	call
0.307	40310708.017	0.000	Spearman's rank correlation rho	two.sided	V4 and V5

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.307. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V5 increases.

2.59. Inference Analysis between Variable "V4" and "V6":

estimate	statistic	p.value	method	alternative	call
0.295	40988416.468	0.000	Spearman's rank correlation rho	two.sided	V4 and V6

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.295. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V6 increases.

2.60. Inference Analysis between Variable "V4" and "V7":

estimate	statistic	p.value	method	alternative	call
0.151	49357433.604	0.000	Spearman's rank correlation rho	two.sided	V4 and V7

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.151. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V7 increases.

2.61. Inference Analysis between Variable "V4" and "V8":

estimate	statistic	p.value	method	alternative	call
0.009	57651080.578	0.819	Spearman's rank correlation rho	two.sided	V4 and V8

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.819. This means that the independence of both variables exists. The value of the correlation is 0.009. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V8 increases.

2.62. Inference Analysis between Variable "V4" and "V9":

estimate	statistic	p.value	method	alternative	call
0.328	39097292.206	0.000	Spearman's rank correlation rho	two.sided	V4 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.328. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V9 increases.

2.63. Inference Analysis between Variable "V4" and "V10":

estimate	statistic	p.value	method	alternative	call
0.211	45883933.143	0.000	Spearman's rank correlation rho	two.sided	V4 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.211. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V10 increases.

2.64. Inference Analysis between Variable "V4" and "V11":

estimate	statistic	p.value	method	alternative	call
0.083	52890982.756	0.028	Spearman's rank correlation rho	two.sided	V4 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.028. This means that a significant correlation between both variables exists. The value of the correlation is 0.083. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V11 increases.

2.65. Inference Analysis between Variable "V4" and "V12":

statistic	p.value	method	alternative	data.name
65337.000	0.134	Wilcoxon rank sum test with continuity correction	two.sided	V4 by V12

Analyses show that p-value is 0.134 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V4" variable.

2.66. Inference Analysis between Variable "V4" and "V13":

statistic	p.value	parameter	method	data.name
67.765	0.000	11	Kruskal-Wallis rank sum test	V4 by V13

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V4" variable.

2.67. Inference Analysis between Variable "V4" and "V14":

statistic	p.value	method	alternative	data.name
19516.000	0.085	Wilcoxon rank sum test with continuity correction	two.sided	V4 by V14

Analyses show that p-value is 0.085 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V4" variable.

2.68. Inference Analysis between Variable "V4" and "V15":

statistic	p.value	method	alternative	data.name
19835.000	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V4 by V15

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V4" variable.

2.69. Inference Analysis between Variable "V4" and "V16":

statistic	p.value	parameter	method	data.name
130.946	0.000	66	Kruskal-Wallis rank sum test	V4 by V16

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V4" variable.

2.70. Inference Analysis between Variable "V4" and "V17":

statistic	p.value	method	alternative	data.name
3782.000	0.541	Wilcoxon rank sum test with continuity correction	two.sided	V4 by V17

Analyses show that p-value is 0.541 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V4" variable.

2.71. Inference Analysis between Variable "V4" and "V18":

estimate	statistic	p.value	method	alternative	call
0.591	23767921.648	0.000	Spearman's rank correlation rho	two.sided	V4 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.591. Since the correlation result is positive, both variables are positively correlated, i.e., when V4 increases, V18 increases.

2.72. Inference Analysis between Variable "V4" and "V19":

Analysis not possible due to data constraints!

2.73. Inference Analysis between Variable "V4" and "V20":

statistic	p.value	parameter	method	data.name
11.319	0.045	5	Kruskal-Wallis rank sum test	V4 by V20

Analyses show that p-value is 0.045 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V4" variable.

2.74. Inference Analysis between Variable "V4" and "V21":

statistic	p.value	method	alternative	data.name
22864.000	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V4 by V21

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V4" variable.

2.75. Inference Analysis between Variable "V5" and "V6":

estimate	statistic	p.value	method	alternative	call
0.392	35335899.701	0.000	Spearman's rank correlation rho	two.sided	V5 and V6

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.392. Since the correlation result is positive, both variables are positively correlated, i.e., when V5 increases, V6 increases.

2.76. Inference Analysis between Variable "V5" and "V7":

estimate	statistic	p.value	method	alternative	call
0.239	44277668.414	0.000	Spearman's rank correlation rho	two.sided	V5 and V7

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.239. Since the correlation

result is positive, both variables are positively correlated, i.e., when V5 increases, V7 increases.

2.77. Inference Analysis between Variable "V5" and "V8":

estimate	statistic	p.value	method	alternative	call
0.102	52215617.739	0.007	Spearman's rank correlation rho	two.sided	V5 and V8

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.007. This means that a significant correlation between both variables exists. The value of the correlation is 0.102. Since the correlation result is positive, both variables are positively correlated, i.e., when V5 increases, V8 increases.

2.78. Inference Analysis between Variable "V5" and "V9":

estimate	statistic	p.value	method	alternative	call
0.397	35090075.309	0.000	Spearman's rank correlation rho	two.sided	V5 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.397. Since the correlation result is positive, both variables are positively correlated, i.e., when V5 increases, V9 increases.

2.79. Inference Analysis between Variable "V5" and "V10":

estimate	statistic	p.value	method	alternative	call
0.268	42592913.675	0.000	Spearman's rank correlation rho	two.sided	V5 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.268. Since the correlation result is positive, both variables are positively correlated, i.e., when V5 increases, V10 increases.

2.80. Inference Analysis between Variable "V5" and "V11":

estimate	statistic	p.value	method	alternative	call
0.018	56644612.417	0.642	Spearman's rank correlation rho	two.sided	V5 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.642. This means that the independence of both variables exists. The value of the correlation is 0.018. Since the correlation result is positive, both variables are positively correlated, i.e., when V5 increases, V11 increases.

2.81. Inference Analysis between Variable "V5" and "V12":

statistic	p.value	method	alternative	data.name
64296.000	0.293	Wilcoxon rank sum test with continuity correction	two.sided	V5 by V12

Analyses show that p-value is 0.293 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V5" variable.

2.82. Inference Analysis between Variable "V5" and "V13":

statistic	p.value	parameter	method	data.name
34.010	0.000	11	Kruskal-Wallis rank sum test	V5 by V13

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V5" variable.

2.83. Inference Analysis between Variable "V5" and "V14":

statistic	p.value	method	alternative	data.name
20641.000	0.362	Wilcoxon rank sum test with continuity correction	two.sided	V5 by V14

Analyses show that p-value is 0.362 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V5" variable.

2.84. Inference Analysis between Variable "V5" and "V15":

statistic	p.value	method	alternative	data.name
24150.000	0.017	Wilcoxon rank sum test with continuity correction	two.sided	V5 by V15

Analyses show that p-value is 0.017 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V5" variable.

2.85. Inference Analysis between Variable "V5" and "V16":

statistic	p.value	parameter	method	data.name
110.765	0.000	66	Kruskal-Wallis rank sum test	V5 by V16

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V5" variable.

2.86. Inference Analysis between Variable "V5" and "V17":

statistic	p.value	method	alternative	data.name
4146.000	0.993	Wilcoxon rank sum test with continuity correction	two.sided	V5 by V17

Analyses show that p-value is 0.993 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V5" variable.

2.87. Inference Analysis between Variable "V5" and "V18":

estimate	statistic	p.value	method	alternative	call
0.649	20431205.945	0.000	Spearman's rank correlation rho	two.sided	V5 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.649. Since the correlation result is positive, both variables are positively correlated, i.e., when V5 increases, V18 increases.

2.88. Inference Analysis between Variable "V5" and "V19":

Analysis not possible due to data constraints!

2.89. Inference Analysis between Variable "V5" and "V20":

statistic	p.value	parameter	method	data.name
18.719	0.002	5	Kruskal-Wallis rank sum test	V5 by V20

Analyses show that p-value is 0.002 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V5" variable.

2.90. Inference Analysis between Variable "V5" and "V21":

statistic	p.value	method	alternative	data.name
19181.000	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V5 by V21

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V5" variable.

2.91. Inference Analysis between Variable "V6" and "V7":

estimate	statistic	p.value	method	alternative	call
0.175	47947083.265	0.000	Spearman's rank correlation rho	two.sided	V6 and V7

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.175. Since the correlation result is positive, both variables are positively correlated, i.e., when V6 increases, V7 increases.

2.92. Inference Analysis between Variable "V6" and "V8":

estimate	statistic	p.value	method	alternative	call
0.100	52329783.830	0.008	Spearman's rank correlation rho	two.sided	V6 and V8

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.008. This means that a significant correlation between both variables exists. The value of the correlation is 0.1. Since the

correlation result is positive, both variables are positively correlated, i.e., when V6 increases, V8 increases.

2.93. Inference Analysis between Variable "V6" and "V9":

estimate	statistic	p.value	method	alternative	call
0.479	30272744.806	0.000	Spearman's rank correlation rho	two.sided	V6 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.479. Since the correlation result is positive, both variables are positively correlated, i.e., when V6 increases, V9 increases.

2.94. Inference Analysis between Variable "V6" and "V10":

estimate	statistic	p.value	method	alternative	call
0.294	41030131.877	0.000	Spearman's rank correlation rho	two.sided	V6 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.294. Since the correlation result is positive, both variables are positively correlated, i.e., when V6 increases, V10 increases.

2.95. Inference Analysis between Variable "V6" and "V11":

estimate	statistic	p.value	method	alternative	call
0.066	53844627.325	0.080	Spearman's rank correlation rho	two.sided	V6 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.08. This means that the independence of both variables exists. The value of the correlation is 0.066. Since the correlation result is positive, both variables are positively correlated, i.e., when V6 increases, V11 increases.

2.96. Inference Analysis between Variable "V6" and "V12":

statistic	p.value	method	alternative	data.name
66507.500	0.027	Wilcoxon rank sum test with continuity correction	two.sided	V6 by V12

Analyses show that p-value is 0.027 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V12" are very different regarding "V6" variable.

2.97. Inference Analysis between Variable "V6" and "V13":

statistic	p.value	parameter	method	data.name
53.033	0.000	11	Kruskal-Wallis rank sum test	V6 by V13

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V6" variable.

2.98. Inference Analysis between Variable "V6" and "V14":

statistic	p.value	method	alternative	data.name
19303.500	0.038	Wilcoxon rank sum test with continuity correction	two.sided	V6 by V14

Analyses show that p-value is 0.038 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V14" are very different regarding "V6" variable.

2.99. Inference Analysis between Variable "V6" and "V15":

statistic	p.value	method	alternative	data.name
23615.500	0.002	Wilcoxon rank sum test with continuity correction	two.sided	V6 by V15

Analyses show that p-value is 0.002 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V6" variable.

2.100. Inference Analysis between Variable "V6" and "V16":

statistic	p.value	parameter	method	data.name
117.415	0.000	66	Kruskal-Wallis rank sum test	V6 by V16

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V6" variable.

2.101. Inference Analysis between Variable "V6" and "V17":

statistic	p.value	method	alternative	data.name
2888.000	0.021	Wilcoxon rank sum test with continuity correction	two.sided	V6 by V17

Analyses show that p-value is 0.021 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V17" are very different regarding "V6" variable.

2.102. Inference Analysis between Variable "V6" and "V18":

estimate	statistic	p.value	method	alternative	call
0.601	23203388.802	0.000	Spearman's rank correlation rho	two.sided	V6 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.601. Since the correlation result is positive, both variables are positively correlated, i.e., when V6 increases, V18 increases.

2.103. Inference Analysis between Variable "V6" and "V19":

Analysis not possible due to data constraints!

2.104. Inference Analysis between Variable "V6" and "V20":

statistic	p.value	parameter	method	data.name
19.251	0.002	5	Kruskal-Wallis rank sum test	V6 by V20

Analyses show that p-value is 0.002 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V6" variable.

2.105. Inference Analysis between Variable "V6" and "V21":

statistic	p.value	method	alternative	data.name
19343.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V6 by V21

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V6" variable.

2.106. Inference Analysis between Variable "V7" and "V8":

estimate	statistic	p.value	method	alternative	call
0.085	53185776.894	0.023	Spearman's rank correlation rho	two.sided	V7 and V8

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.023. This means that a significant correlation between both variables exists. The value of the correlation is 0.085. Since the correlation result is positive, both variables are positively correlated, i.e., when V7 increases, V8 increases.

2.107. Inference Analysis between Variable "V7" and "V9":

estimate	statistic	p.value	method	alternative	call
0.189	47133510.986	0.000	Spearman's rank correlation rho	two.sided	V7 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.189. Since the correlation result is positive, both variables are positively correlated, i.e., when V7 increases, V9 increases.

2.108. Inference Analysis between Variable "V7" and "V10":

estimate	statistic	p.value	method	alternative	call
0.252	43491622.035	0.000	Spearman's rank correlation rho	two.sided	V7 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.252. Since the correlation result is positive, both variables are positively correlated, i.e., when V7 increases, V10 increases.

2.109. Inference Analysis between Variable "V7" and "V11":

estimate	statistic	p.value	method	alternative	call
-0.004	57900625.446	0.911	Spearman's rank correlation rho	two.sided	V7 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.911. This means that the independence of both variables exists. The value of the correlation is -0.004. Since the correlation result is negative, both variables are negatively correlated, i.e., when V7 increases, V11 decreases.

2.110. Inference Analysis between Variable "V7" and "V12":

statistic	p.value	method	alternative	data.name
58060.500	0.101	Wilcoxon rank sum test with continuity correction	two.sided	V7 by V12

Analyses show that p-value is 0.101 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V7" variable.

2.111. Inference Analysis between Variable "V7" and "V13":

statistic	p.value	parameter	method	data.name
23.234	0.016	11	Kruskal-Wallis rank sum test	V7 by V13

Analyses show that p-value is 0.016 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V7" variable.

2.112. Inference Analysis between Variable "V7" and "V14":

statistic	p.value	method	alternative	data.name
20786.500	0.414	Wilcoxon rank sum test with continuity correction	two.sided	V7 by V14

Analyses show that p-value is 0.414 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V7" variable.

2.113. Inference Analysis between Variable "V7" and "V15":

statistic	p.value	method	alternative	data.name
28244.500	0.820	Wilcoxon rank sum test with continuity correction	two.sided	V7 by V15

Analyses show that p-value is 0.82 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V15" are similar regarding "V7" variable.

2.114. Inference Analysis between Variable "V7" and "V16":

statistic	p.value	parameter	method	data.name
83.310	0.074	66	Kruskal-Wallis rank sum test	V7 by V16

Analyses show that p-value is 0.074 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, there are no statistical differences. Thus, it is possible to assume that groups have similar distribution regarding "V7" variable.

2.115. Inference Analysis between Variable "V7" and "V17":

statistic	p.value	method	alternative	data.name
4508.000	0.551	Wilcoxon rank sum test with continuity correction	two.sided	V7 by V17

Analyses show that p-value is 0.551 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V7" variable.

2.116. Inference Analysis between Variable "V7" and "V18":

estimate	statistic	p.value	method	alternative	call
0.453	31836759.637	0.000	Spearman's rank correlation rho	two.sided	V7 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.453. Since the correlation

result is positive, both variables are positively correlated, i.e., when V7 increases, V18 increases.

2.117. Inference Analysis between Variable "V7" and "V19":

Analysis not possible due to data constraints!

2.118. Inference Analysis between Variable "V7" and "V20":

statistic	p.value	parameter	method	data.name
4.672	0.457	5	Kruskal-Wallis rank sum test	V7 by V20

Analyses show that p-value is 0.457 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, there are no statistical differences. Thus, it is possible to assume that groups have similar distribution regarding "V7" variable.

2.119. Inference Analysis between Variable "V7" and "V21":

statistic	p.value	method	alternative	data.name
29634.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V7 by V21

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V7" variable.

2.120. Inference Analysis between Variable "V8" and "V9":

estimate	statistic	p.value	method	alternative	call
0.102	52236179.079	0.007	Spearman's rank correlation rho	two.sided	V8 and V9

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.007. This means that a significant correlation between both variables exists. The value of the correlation is 0.102. Since the correlation result is positive, both variables are positively correlated, i.e., when V8 increases, V9 increases.

2.121. Inference Analysis between Variable "V8" and "V10":

estimate	statistic	p.value	method	alternative	call
0.101	52291478.889	0.007	Spearman's rank correlation rho	two.sided	V8 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.007. This means that a significant correlation between both variables exists. The value of the correlation is 0.101. Since the correlation result is positive, both variables are positively correlated, i.e., when V8 increases, V10 increases.

2.122. Inference Analysis between Variable "V8" and "V11":

estimate	statistic	p.value	method	alternative	call
-0.041	60043876.686	0.274	Spearman's rank correlation rho	two.sided	V8 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.274. This means that the independence of both variables exists. The value of the correlation is -0.041. Since the correlation result is negative, both variables are negatively correlated, i.e., when V8 increases, V11 decreases.

2.123. Inference Analysis between Variable "V8" and "V12":

statistic	p.value	method	alternative	data.name
58051.000	0.089	Wilcoxon rank sum test with continuity correction	two.sided	V8 by V12

Analyses show that p-value is 0.089 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V8" variable.

2.124. Inference Analysis between Variable "V8" and "V13":

statistic	p.value	parameter	method	data.name
21.851	0.026	11	Kruskal-Wallis rank sum test	V8 by V13

Analyses show that p-value is 0.026 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V8" variable.

2.125. Inference Analysis between Variable "V8" and "V14":

statistic	p.value	method	alternative	data.name
21482.000	0.749	Wilcoxon rank sum test with continuity correction	two.sided	V8 by V14

Analyses show that p-value is 0.749 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V8" variable.

2.126. Inference Analysis between Variable "V8" and "V15":

statistic	p.value	method	alternative	data.name
26509.000	0.356	Wilcoxon rank sum test with continuity correction	two.sided	V8 by V15

Analyses show that p-value is 0.356 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V15" are similar regarding "V8" variable.

2.127. Inference Analysis between Variable "V8" and "V16":

statistic	p.value	parameter	method	data.name
83.900	0.068	66	Kruskal-Wallis rank sum test	V8 by V16

Analyses show that p-value is 0.068 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, there are no statistical differences. Thus, it is possible to assume that groups have similar distribution regarding "V8" variable.

2.128. Inference Analysis between Variable "V8" and "V17":

statistic	p.value	method	alternative	data.name
4782.000	0.276	Wilcoxon rank sum test with continuity correction	two.sided	V8 by V17

Analyses show that p-value is 0.276 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V8" variable.

2.129. Inference Analysis between Variable "V8" and "V18":

estimate	statistic	p.value	method	alternative	call
0.317	39716414.715	0.000	Spearman's rank correlation rho	two.sided	V8 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.317. Since the correlation result is positive, both variables are positively correlated, i.e., when V8 increases, V18 increases.

2.130. Inference Analysis between Variable "V8" and "V19":

Analysis not possible due to data constraints!

2.131. Inference Analysis between Variable "V8" and "V20":

statistic	p.value	parameter	method	data.name
7.949	0.159	5	Kruskal-Wallis rank sum test	V8 by V20

Analyses show that p-value is 0.159 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, there are no statistical differences. Thus, it is possible to assume that groups have similar distribution regarding "V8" variable.

2.132. Inference Analysis between Variable "V8" and "V21":

statistic	p.value	method	alternative	data.name
36238.000	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V8 by V21

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V8" variable.

2.133. Inference Analysis between Variable "V9" and "V10":

estimate	statistic	p.value	method	alternative	call
0.283	41675334.319	0.000	Spearman's rank correlation rho	two.sided	V9 and V10

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.283. Since the correlation result is positive, both variables are positively correlated, i.e., when V9 increases, V10 increases.

2.134. Inference Analysis between Variable "V9" and "V11":

estimate	statistic	p.value	method	alternative	call
0.121	50708931.063	0.001	Spearman's rank correlation rho	two.sided	V9 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.001. This means that a significant correlation between both variables exists. The value of the correlation is 0.121. Since the correlation result is positive, both variables are positively correlated, i.e., when V9 increases, V11 increases.

2.135. Inference Analysis between Variable "V9" and "V12":

statistic	p.value	method	alternative	data.name
61437.500	0.854	Wilcoxon rank sum test with continuity correction	two.sided	V9 by V12

Analyses show that p-value is 0.854 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V9" variable.

2.136. Inference Analysis between Variable "V9" and "V13":

statistic	p.value	parameter	method	data.name
52.390	0.000	11	Kruskal-Wallis rank sum test	V9 by V13

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V9" variable.

2.137. Inference Analysis between Variable "V9" and "V14":

statistic	p.value	method	alternative	data.name
19917.500	0.126	Wilcoxon rank sum test with continuity correction	two.sided	V9 by V14

Analyses show that p-value is 0.126 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V9" variable.

2.138. Inference Analysis between Variable "V9" and "V15":

statistic	p.value	method	alternative	data.name
21369.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V9 by V15

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V9" variable.

2.139. Inference Analysis between Variable "V9" and "V16":

statistic	p.value	parameter	method	data.name
120.027	0.000	66	Kruskal-Wallis rank sum test	V9 by V16

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V9" variable.

2.140. Inference Analysis between Variable "V9" and "V17":

statistic	p.value	method	alternative	data.name
3408.000	0.189	Wilcoxon rank sum test with continuity correction	two.sided	V9 by V17

Analyses show that p-value is 0.189 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V9" variable.

2.141. Inference Analysis between Variable "V9" and "V18":

estimate	statistic	p.value	method	alternative	call
0.637	21081075.005	0.000	Spearman's rank correlation rho	two.sided	V9 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.637. Since the correlation result is positive, both variables are positively correlated, i.e., when V9 increases, V18 increases.

2.142. Inference Analysis between Variable "V9" and "V19":

Analysis not possible due to data constraints!

2.143. Inference Analysis between Variable "V9" and "V20":

statistic	p.value	parameter	method	data.name
8.517	0.130	5	Kruskal-Wallis rank sum test	V9 by V20

Analyses show that p-value is 0.13 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, there are no statistical differences. Thus, it is possible to assume that groups have similar distribution regarding "V9" variable.

2.144. Inference Analysis between Variable "V9" and "V21":

statistic	p.value	method	alternative	data.name
16005.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V9 by V21

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V9" variable.

2.145. Inference Analysis between Variable "V10" and "V11":

estimate	statistic	p.value	method	alternative	call
0.064	53992130.974	0.092	Spearman's rank correlation rho	two.sided	V10 and V11

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.092. This means that the independence of both variables exists. The value of the correlation is 0.064. Since the correlation result is positive, both variables are positively correlated, i.e., when V10 increases, V11 increases.

2.146. Inference Analysis between Variable "V10" and "V12":

statistic	p.value	method	alternative	data.name
65221.500	0.143	Wilcoxon rank sum test with continuity correction	two.sided	V10 by V12

Analyses show that p-value is 0.143 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V10" variable.

2.147. Inference Analysis between Variable "V10" and "V13":

statistic	p.value	parameter	method	data.name
42.068	0.000	11	Kruskal-Wallis rank sum test	V10 by V13

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V10" variable.

2.148. Inference Analysis between Variable "V10" and "V14":

statistic	p.value	method	alternative	data.name
20005.500	0.166	Wilcoxon rank sum test with continuity correction	two.sided	V10 by V14

Analyses show that p-value is 0.166 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V14" are similar regarding "V10" variable.

2.149. Inference Analysis between Variable "V10" and "V15":

statistic	p.value	method	alternative	data.name
23041.500	0.002	Wilcoxon rank sum test with continuity correction	two.sided	V10 by V15

Analyses show that p-value is 0.002 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V10" variable.

2.150. Inference Analysis between Variable "V10" and "V16":

statistic	p.value	parameter	method	data.name
94.160	0.013	66	Kruskal-Wallis rank sum test	V10 by V16

Analyses show that p-value is 0.013 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V10" variable.

2.151. Inference Analysis between Variable "V10" and "V17":

statistic	p.value	method	alternative	data.name
4816.000	0.267	Wilcoxon rank sum test with continuity correction	two.sided	V10 by V17

Analyses show that p-value is 0.267 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V10" variable.

2.152. Inference Analysis between Variable "V10" and "V18":

estimate	statistic	p.value	method	alternative	call
0.544	26540987.537	0.000	Spearman's rank correlation rho	two.sided	V10 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0. This means that a significant correlation between both variables exists. The value of the correlation is 0.544. Since the correlation result is positive, both variables are positively correlated, i.e., when V10 increases, V18 increases.

2.153. Inference Analysis between Variable "V10" and "V19":

Analysis not possible due to data constraints!

2.154. Inference Analysis between Variable "V10" and "V20":

statistic	p.value	parameter	method	data.name
13.797	0.017	5	Kruskal-Wallis rank sum test	V10 by V20

Analyses show that p-value is 0.017 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V10" variable.

2.155. Inference Analysis between Variable "V10" and "V21":

statistic	p.value	method	alternative	data.name
27709.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V10 by V21

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V10" variable.

2.156. Inference Analysis between Variable "V11" and "V12":

statistic	p.value	method	alternative	data.name
65718.500	0.115	Wilcoxon rank sum test with continuity correction	two.sided	V11 by V12

Analyses show that p-value is 0.115 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V12" are similar regarding "V11" variable.

2.157. Inference Analysis between Variable "V11" and "V13":

statistic	p.value	parameter	method	data.name
67.420	0.000	11	Kruskal-Wallis rank sum test	V11 by V13

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V11" variable.

2.158. Inference Analysis between Variable "V11" and "V14":

statistic	p.value	method	alternative	data.name
16900.500	0.002	Wilcoxon rank sum test with continuity correction	two.sided	V11 by V14

Analyses show that p-value is 0.002 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V14" are very different regarding "V11" variable.

2.159. Inference Analysis between Variable "V11" and "V15":

statistic	p.value	method	alternative	data.name
18387.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V11 by V15

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V15" are very different regarding "V11" variable.

2.160. Inference Analysis between Variable "V11" and "V16":

statistic	p.value	parameter	method	data.name
197.051	0.000	66	Kruskal-Wallis rank sum test	V11 by V16

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V11" variable.

2.161. Inference Analysis between Variable "V11" and "V17":

statistic	p.value	method	alternative	data.name
4703.000	0.419	Wilcoxon rank sum test with continuity correction	two.sided	V11 by V17

Analyses show that p-value is 0.419 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V17" are similar regarding "V11" variable.

2.162. Inference Analysis between Variable "V11" and "V18":

estimate	statistic	p.value	method	alternative	call
0.069	53652779.972	0.066	Spearman's rank correlation rho	two.sided	V11 and V18

To analyze two numerical variables (in this case, one of them is an integer), correlations analysis is applied. Since we have an integer variable, Spearman correlation is more appropriate.

Spearman correlation presents a p-value of 0.066. This means that the independence of both variables exists. The value of the correlation is 0.069. Since the correlation result is positive, both variables are positively correlated, i.e., when V11 increases, V18 increases.

2.163. Inference Analysis between Variable "V11" and "V19":

Analysis not possible due to data constraints!

2.164. Inference Analysis between Variable "V11" and "V20":

statistic	p.value	parameter	method	data.name
23.280	0.000	5	Kruskal-Wallis rank sum test	V11 by V20

Analyses show that p-value is 0 (p<0.05). This means that null hypothesis is rejected (p<0.05) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V11" variable.

2.165. Inference Analysis between Variable "V11" and "V21":

statistic	p.value	method	alternative	data.name
40857.000	0.001	Wilcoxon rank sum test with continuity correction	two.sided	V11 by V21

Analyses show that p-value is 0.001 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V11" variable.

2.166. Inference Analysis between Variable "V12" and "V13":

V12	V13	Freq
f		49
m		46
f	Asian	51
m	Asian	72
f	Black	23
m	Black	20
f	Hispanic	2
m	Hispanic	11
f	Latino	8
m	Latino	12
f	Middle Eastern	38
m	Middle Eastern	54
f	others	0
m	others	1
f	Others	14
m	Others	16
f	Pasifika	6
m	Pasifika	6
f	South Asian	20
m	South Asian	16
f	Turkish	2
m	Turkish	4
f	White-European	124
m	White-European	109

p.value	method	alternative
0.134	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0.134 ($p>0.05$). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 20.8333333333333%

2.167. Inference Analysis between Variable "V12" and "V14":

V12	V14	Freq
f	no	302
m	no	333
f	yes	35
m	yes	34

statistic	p.value	parameter	method
0.139	0.709	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0.709 (p>0.05). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

2.168. Inference Analysis between Variable "V12" and "V15":

V12	V15	Freq
f	no	283
m	no	330
f	yes	54
m	yes	37

statistic	p.value	parameter	method
4.996	0.025	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0.025 (p<0.05). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

2.169. Inference Analysis between Variable "V12" and "V16":

NOTE: TEST Crosstable of high dimension. Therefore, not available in this report!

Analysis not possible due to data constraints!

2.170. Inference Analysis between Variable "V12" and "V17":

V12	V17	Freq
f	no	330
m	no	362
f	yes	7

V12	V17	Freq
m	yes	5

statistic	p.value	parameter	method
0.194	0.660	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0.66 (p>0.05). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

2.171. Inference Analysis between Variable "V12" and "V18":

statistic	p.value	method	alternative	data.name
64480.500	0.324	Wilcoxon rank sum test with continuity correction	two.sided	V18 by V12

Analyses show that p-value is 0.324 (p>0.05). This means that null hypothesis is non-rejected (p>0.05) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V18" are similar regarding "V12" variable.

2.172. Inference Analysis between Variable "V12" and "V19":

Analysis not possible due to data constraints!

2.173. Inference Analysis between Variable "V12" and "V20":

V12	V20	Freq
f		49
m		46
f	Health care professional	1
m	Health care professional	3
f	Others	0
m	Others	5
f	Parent	25
m	Parent	25
f	Relative	7
m	Relative	21
f	Self	255
m	Self	267

p.value	method	alternative
0.024	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0.024 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 25%

2.174. Inference Analysis between Variable "V12" and "V21":

V12	V21	Freq
f	NO	234
m	NO	281
f	YES	103
m	YES	86

statistic	p.value	parameter	method
4.192	0.041	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0.041 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

2.175. Inference Analysis between Variable "V13" and "V14":

V13	V14	Freq
	no	85
Asian	no	116
Black	no	40
Hispanic	no	12
Latino	no	16
Middle Eastern	no	83
others	no	1
Others	no	28
Pasifika	no	11
South Asian	no	36
Turkish	no	6
White-European	no	201
	yes	10
Asian	yes	7
Black	yes	3
Hispanic	yes	1

V13	V14	Freq
Latino	yes	4
Middle Eastern	yes	9
others	yes	0
Others	yes	2
Pasifika	yes	1
South Asian	yes	0
Turkish	yes	0
White-European	yes	32

p.value	method	alternative
0.172	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0.172 (p>0.05). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 37.5%

2.176. Inference Analysis between Variable "V13" and "V15":

V13	V15	Freq
	no	89
Asian	no	118
Black	no	38
Hispanic	no	12
Latino	no	12
Middle Eastern	no	83
others	no	1
Others	no	28
Pasifika	no	10
South Asian	no	34
Turkish	no	5
White-European	no	183
	yes	6
Asian	yes	5
Black	yes	5
Hispanic	yes	1
Latino	yes	8
Middle Eastern	yes	9

V13	V15	Freq
others	yes	0
Others	yes	2
Pasifika	yes	2
South Asian	yes	2
Turkish	yes	1
White-European	yes	50

p.value	method	alternative
0.000	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 29.1666666666667%

2.177. Inference Analysis between Variable "V13" and "V16":

NOTE: TEST Crosstable of high dimension. Therefore, not available in this report!

Analysis not possible due to data constraints!

2.178. Inference Analysis between Variable "V13" and "V17":

V13	V17	Freq
	no	93
Asian	no	122
Black	no	42
Hispanic	no	12
Latino	no	19
Middle Eastern	no	89
others	no	1
Others	no	29
Pasifika	no	12
South Asian	no	36
Turkish	no	6
White-European	no	231
	yes	2
Asian	yes	1

V13	V17	Freq
Black	yes	1
Hispanic	yes	1
Latino	yes	1
Middle Eastern	yes	3
others	yes	0
Others	yes	1
Pasifika	yes	0
South Asian	yes	0
Turkish	yes	0
White-European	yes	2

p.value	method	alternative
0.259	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0.259 ($p>0.05$). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 54.1666666666667%

2.179. Inference Analysis between Variable "V13" and "V18":

statistic	p.value	parameter	method	data.name
112.127	0.000	11	Kruskal-Wallis rank sum test	V18 by V13

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V13" variable.

2.180. Inference Analysis between Variable "V13" and "V19":

Analysis not possible due to data constraints!

2.181. Inference Analysis between Variable "V13" and "V20":

V13	V20	Freq
		95
Asian		0
Black		0

V13	V20	Freq
Hispanic		0
Latino		0
Middle Eastern		0
others		0
Others		0
Pasifika		0
South Asian		0
Turkish		0
White-European		0
	Health care professional	0
Asian	Health care professional	1
Black	Health care professional	1
Hispanic	Health care professional	0
Latino	Health care professional	0
Middle Eastern	Health care professional	0
others	Health care professional	0
Others	Health care professional	1
Pasifika	Health care professional	0
South Asian	Health care professional	0
Turkish	Health care professional	0
White-European	Health care professional	1
	Others	0
Asian	Others	1
Black	Others	0
Hispanic	Others	0
Latino	Others	0
Middle Eastern	Others	0
others	Others	0
Others	Others	0
Pasifika	Others	0
South Asian	Others	0
Turkish	Others	1
White-European	Others	3
	Parent	0
Asian	Parent	7
Black	Parent	9
Hispanic	Parent	1
Latino	Parent	3

V13	V20	Freq
Middle Eastern	Parent	15
others	Parent	0
Others	Parent	1
Pasifika	Parent	0
South Asian	Parent	1
Turkish	Parent	0
White-European	Parent	13
	Relative	0
Asian	Relative	1
Black	Relative	3
Hispanic	Relative	1
Latino	Relative	1
Middle Eastern	Relative	8
others	Relative	0
Others	Relative	0
Pasifika	Relative	2
South Asian	Relative	0
Turkish	Relative	0
White-European	Relative	12
	Self	0
Asian	Self	113
Black	Self	30
Hispanic	Self	11
Latino	Self	16
Middle Eastern	Self	69
others	Self	1
Others	Self	28
Pasifika	Self	10
South Asian	Self	35
Turkish	Self	5
White-European	Self	204

p.value	method	alternative
0.000	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 76.3888888888889%

2.182. Inference Analysis between Variable "V13" and "V21":

V13	V21	Freq
	NO	86
Asian	NO	107
Black	NO	25
Hispanic	NO	8
Latino	NO	10
Middle Eastern	NO	84
others	NO	1
Others	NO	21
Pasifika	NO	11
South Asian	NO	33
Turkish	NO	5
White-European	NO	124
	YES	9
Asian	YES	16
Black	YES	18
Hispanic	YES	5
Latino	YES	10
Middle Eastern	YES	8
others	YES	0
Others	YES	9
Pasifika	YES	1
South Asian	YES	3
Turkish	YES	1
White-European	YES	109

p.value	method	alternative
0.000	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 20.8333333333333%

2.183. Inference Analysis between Variable "V14" and "V15":

V14	V15	Freq
no	no	564
yes	no	49
no	yes	71
yes	yes	20

statistic	p.value	parameter	method
15.982	0.000	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

2.184. Inference Analysis between Variable "V14" and "V16":

NOTE: TEST Crosstable of high dimension. Therefore, not available in this report!

Analysis not possible due to data constraints!

2.185. Inference Analysis between Variable "V14" and "V17":

V14	V17	Freq
no	no	625
yes	no	67
no	yes	10
yes	yes	2

statistic	p.value	parameter	method
0.101	0.751	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0.751 ($p>0.05$). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 25%

2.186. Inference Analysis between Variable "V14" and "V18":

statistic	p.value	method	alternative	data.name
18422.000	0.029	Wilcoxon rank sum test with continuity correction	two.sided	V18 by V14

Analyses show that p-value is 0.029 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V18" are very different regarding "V14" variable.

2.187. Inference Analysis between Variable "V14" and "V19":

Analysis not possible due to data constraints!

2.188. Inference Analysis between Variable "V14" and "V20":

V14	V20	Freq
no		85
yes		10
no	Health care professional	3
yes	Health care professional	1
no	Others	5
yes	Others	0
no	Parent	40
yes	Parent	10
no	Relative	23
yes	Relative	5
no	Self	479
yes	Self	43

p.value	method	alternative
0.041	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0.041 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 25%

2.189. Inference Analysis between Variable "V14" and "V21":

V14	V21	Freq
no	NO	474
yes	NO	41
no	YES	161
yes	YES	28

statistic	p.value	parameter	method
6.591	0.010	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0.01 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

2.190. Inference Analysis between Variable "V15" and "V16":

NOTE: TEST Crosstable of high dimension. Therefore, not available in this report!

Analysis not possible due to data constraints!

2.191. Inference Analysis between Variable "V15" and "V17":

V15	V17	Freq
no	no	603
yes	no	89
no	yes	10
yes	yes	2

statistic	p.value	parameter	method
0.000	1.000	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 1 ($p>0.05$). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 25%

2.192. Inference Analysis between Variable "V15" and "V18":

statistic	p.value	method	alternative	data.name
19260.500	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V18 by V15

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V18" are very different regarding "V15" variable.

2.193. Inference Analysis between Variable "V15" and "V19":

Analysis not possible due to data constraints!

2.194. Inference Analysis between Variable "V15" and "V20":

V15	V20	Freq
no		89
yes		6
no	Health care professional	3
yes	Health care professional	1
no	Others	4
yes	Others	1
no	Parent	37
yes	Parent	13
no	Relative	22
yes	Relative	6
no	Self	458
yes	Self	64

statistic	p.value	parameter	method
14.026	0.015	5	Pearson's Chi-squared test

The results of the analysis shows that chi-squared test presents a p-value of 0.015 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 33.3333333333333%

2.195. Inference Analysis between Variable "V15" and "V21":

V15	V21	Freq
no	NO	467
yes	NO	48
no	YES	146
yes	YES	43

statistic	p.value	parameter	method
20.982	0.000	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

2.196. Inference Analysis between Variable "V16" and "V17":

NOTE: TEST Crosstable of high dimension. Therefore, not available in this report!

Analysis not possible due to data constraints!

2.197. Inference Analysis between Variable "V16" and "V18":

statistic	p.value	parameter	method	data.name
174.749	0.000	66	Kruskal-Wallis rank sum test	V18 by V16

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V16" variable.

2.198. Inference Analysis between Variable "V16" and "V19":

Analysis not possible due to data constraints!

2.199. Inference Analysis between Variable "V16" and "V20":

NOTE: TEST Crosstable of high dimension. Therefore, not available in this report!

Analysis not possible due to data constraints!

2.200. Inference Analysis between Variable "V16" and "V21":

NOTE: TEST Crosstable of high dimension. Therefore, not available in this report!

Analysis not possible due to data constraints!

2.201. Inference Analysis between Variable "V17" and "V18":

statistic	p.value	method	alternative	data.name
4048.500	0.882	Wilcoxon rank sum test with continuity correction	two.sided	V18 by V17

Analyses show that p-value is 0.882 ($p>0.05$). This means that null hypothesis is non-rejected ($p>0.05$) and, therefore, the distributions are similar. Thus, it is possible to assume that groups of "V18" are similar regarding "V17" variable.

2.202. Inference Analysis between Variable "V17" and "V19":

Analysis not possible due to data constraints!

2.203. Inference Analysis between Variable "V17" and "V20":

V17	V20	Freq
no		93
yes		2
no	Health care professional	4
yes	Health care professional	0
no	Others	5
yes	Others	0
no	Parent	49
yes	Parent	1
no	Relative	27
yes	Relative	1
no	Self	514
yes	Self	8

p.value	method	alternative
0.497	Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates)	two.sided

The results of the analysis shows that Fisher test presents a p-value of 0.497 (p>0.05). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 50%

2.204. Inference Analysis between Variable "V17" and "V21":

V17	V21	Freq
no	NO	508
yes	NO	7
no	YES	184
yes	YES	5

statistic	p.value	parameter	method
0.706	0.401	1	Pearson's Chi-squared test with Yates' continuity correction

The results of the analysis shows that chi-squared test presents a p-value of 0.401 (p>0.05). Thus, the null hipothesis is non-rejected with a confidence level of 95%. We can conclude that both variables are independent.

2.205. Inference Analysis between Variable "V18" and "V19":

Analysis not possible due to data constraints!

2.206. Inference Analysis between Variable "V18" and "V20":

statistic	p.value	parameter	method	data.name
34.464	0.000	5	Kruskal-Wallis rank sum test	V18 by V20

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups are very different regarding "V18" variable.

2.207. Inference Analysis between Variable "V18" and "V21":

statistic	p.value	method	alternative	data.name
0.000	0.000	Wilcoxon rank sum test with continuity correction	two.sided	V18 by V21

Analyses show that p-value is 0 ($p<0.05$). This means that null hypothesis is rejected ($p<0.05$) and, therefore, there are statistical differences. Thus, it is possible to assume that groups of "V21" are very different regarding "V18" variable.

2.208. Inference Analysis between Variable "V19" and "V20":

Analysis not possible due to data constraints!

2.209. Inference Analysis between Variable "V19" and "V21":

Analysis not possible due to data constraints!

2.210. Inference Analysis between Variable "V20" and "V21":

V20	V21	Freq
	NO	86
Health care professional	NO	3
Others	NO	4
Parent	NO	38
Relative	NO	19
Self	NO	365
	YES	9
Health care professional	YES	1

V20	V21	Freq
Others	YES	1
Parent	YES	12
Relative	YES	9
Self	YES	157

statistic	p.value	parameter	method
18.105	0.003	5	Pearson's Chi-squared test

The results of the analysis shows that chi-squared test presents a p-value of 0.003 ($p<0.05$). Thus, the null hipothesis is rejected with a confidence level of 95%. We can conclude that both variables are dependent.

NOTE: TEST not robust due to high percentage of values inferior to 5 in crosstable: 33.3333333333333%

3. Classifiers Analysis:

3.1. About the Classifiers:

Classification is a task of supervised learning that conveys one or more attributes in order to group subpopulations into different labels or classes. The function that maps the attribution of these classes is called a classifier and its output is therefore discrete. The output is what differentiates classification from the other supervised learning method – regression output is continuous and its mapping function is called estimator. Knowing the underlying assumptions of the most used classification algorithms may be very useful when the analyst wants to deepen its data understanding. Prediction accuracy for each problem is usually the most important feature. But it may also be useful to have a better computational efficiency. Transparency is another issue that may arise because sometimes algorithms that provide the most accurate models do not reveal how their models are generated. In this section some of the most popular classification algorithms available in the Weka software application and that will be used in this work are presented.

The simplest algorithm to be tested is the OneR method that stands for one rule, i.e. based on a unique attribute. Attributes are ranked based on the training set error rate (Holte, 1993). Even considering its simplicity, it can be a relatively accurate method. Rule sets have several advantages. They are easier to understand and may be used as a first order logic. But have some disadvantages too, like poor scaling and noisy data susceptibility.

JRip implements the propositional rule learner Repeated Incremental Pruning to Produce Error Reduction (RIPPER) proposed by (Cohen, 1995). It is an improved

method that grows rules to 100% accuracy and then prunes over fitting rules until accuracy starts to decrease.

In a decision tree, each node is either a decision node for an attribute or a leaf node corresponding to a classification. In contrast to rules setting based on error rates, decision trees rely on entropy-based measures (Holte, 1993). The decision stump is simply a one level tree resulting in the same configuration of the OneR algorithm over a single attribute.

Haykin (1999) defines an Artificial Neural Network (ANN) as a massively parallel processor, distributed, consisting of simple processing units, which have a natural propensity for storing experiential knowledge and making it available for use.

The Support Vector Machines (SVM) algorithm represents instances as points in space creating clear gaps between different classes (Chandan Kolvankar et al., 2012). http://www.svms.org/anns.html

Contrary to the previous algorithms, Instance-based learning (IBl) does not record abstractions from instances, but predicts based on specific instances (Aha, Kibler & Albert, 1991). This is why it is very easy to interpret its results.

From the assumption that combining classifiers has better results than each classifier isolated resulted different ensemble methods. Bagging is an acronym for bootstrap aggregation (Breiman, 1996). It generates multiple versions of a predictor in order to use them into an aggregated average to predict a class. Boosting “works by sequentially applying a classification algorithm to re-weighted versions of the training data and then taking a weighted majority vote of the sequence of classifiers thus produced” (Friedman et al., 2000). It improves the performance of any weak learning algorithm by running on various distributions over the training data and combining the resulting models in a composite classifier (Freund & Schapire, 1996).

Stacking differs from boosting because it combines different learning algorithms and tries to balance their strengths and weaknesses.

3.1.1. Classifiers Tests Results:

classifier	pctCorrect	pctIncorrect	pctUnclassified	kappa	meanAbsoluteError	rootMeanSquaredError	relativeAbsoluteError	rootRelativeSquaredError
AdaBoostM1	100	0	0	1	0	0	0	0
Bagging	100	0	0	1	0	0	0	0
DecisionStump	100	0	0	1	0	0	0	0
JRip	100	0	0	1	0	0	0	0
LMT	100	0	0	1	0.216	0.25	54.823	56.436
Logistic	96.439	3.561	0	0.91	0.036	0.184	9.02	41.539
OneR	100	0	0	1	0	0	0	0
PART	100	0	0	1	0	0	0	0
SMO	100	0	0	1	0	0	0	0
Stacking	73.077	26.923	0	0	0.394	0.444	100	100

classifier	pctCorrect	pctIncorrect	pctUnclassified	kappa	meanAbsoluteError	rootMeanSquaredError	relativeAbsoluteError	rootRelativeSquaredError
LogitBoost	100	0	0	1	0	0	0.003	0.003
J48	100	0	0	1	0	0	0	0
IBk	95.726	4.274	0	0.892	0.044	0.206	11.209	46.533

Now, the analyst should pay attention to these results metrics and decide which classifier makes the most sense to be used for their data.

3.2. Additional Readings

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-Based Learning Algorithms. Machine Learning, (6), 37–66.

Breiman, L. (1996). Bagging Predictors. Machine Learning, 24(2), 123–140.

Chandan Kolvankar, Trivedi, J., Mani, B., Ramanathan, R., & Kadam, S. (2012). Support Vector Machine for Learning in Artificial Intelligence Systems. International Journal of Engineering Research and Applications (IJERA), 409–412.

Cohen, W. W. (1995). Fast Effective Rule Induction. In In Proceedings of the Twelfth International Conference on Machine Learning (pp. 115–123). Morgan Kaufmann.

Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. In Proceedings of the Thirteenth International Conference (pp. 148–156).

Haykin, S. (1999). Neural Networks: A Comprehensive Foundation (2nd ed.). Prentice-Hall.

Holte, R. C. (1993). Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. In Machine Learning (pp. 63–91).

Jerome Friedman, Hastie, T., & Tibshirani, R. (2000). Additive Logistic Regression: a Statistical View of Boosting. Annals of Statistics, 28(2), 337–407.