Samia Ashraf |(22011513-022)

1) Dataset Selection

- Choose a dataset from Kaggle that aligns with the research focus global health statistics data,).

2) Data Pre-Processing

Data pre-processing means cleaning data. It is different in both type of variables. In quantitative variables it is done by removing all errors like missing values and outliers . In qualitative data it is done by converting data from nominal to ordinal if needed

i) Checking for Missing Values

Frequencies

Command

Analyze > Descriptive Statistics>Frequencies

Statistics

	Prevalence Rate (%)	Incidence Rate (%)	Mortality Rate (%)	Population Affected	Healthcare Access (%)	Hospital Beds per 1000	DALYs	Average Treatment Cost (USD)	Recovery Ra (%)
N	3685	3685	3685	3685	3685	3685	3685	3685	3
	0	0	0	0	0	0	0	0	
Mean	10.07	7.50	5.01	498832.04	75.04	5.20	2485.69	25092.73	74
Std. Error of Mean	.094	.071	.047	4744.222	.239	.044	23.651	237.469	.2
Median	10.16	7.45	4.95	500187.00	75.55	5.12	2499.00	24943.00	74
Mode	3	11	3[a]	123560[a]	94	5	23[a]	35639[a]	
Std. Deviation	5.699	4.304	2.846	287994.230	14.533	2.699	1435.738	14415.366	14.3

Variance	32.480	18.522	8.102	82940676424.522	211.202	7.282	2061344.187	207802767.863	204.4
Skewness	-.018	.033	.023	.012	-.030	.026	-.015	.006	.0
Std. Error of Skewness	.040	.040	.040	.040	.040	.040	.040	.040	.0
Kurtosis	-1.171	-1.225	-1.195	-1.200	-1.209	-1.166	-1.208	-1.201	-1.2
Std. Error of Kurtosis	.081	.081	.081	.081	.081	.081	.081	.081	.0
Range	20	15	10	998583	50	10	4999	49877	
Minimum	0	0	0	1372	50	1	1	117	
Maximum	20	15	10	999955	100	10	5000	49994	
Sum	37097	27628	18457	1838196075	276511	19178	9159775	92466724	2739
Percentiles	5.32	3.79	2.52	244909.00	62.42	2.91	1230.00	12573.50	61
	10.16	7.45	4.95	500187.00	75.55	5.12	2499.00	24943.00	74
	14.88	11.24	7.48	750772.50	87.62	7.51	3718.50	37562.50	86

a. Multiple modes exist. The smallest value is shown

Interpretation:

The dataset provides descriptive statistics for health-related variables across 3,685 records. Key metrics include a **mean prevalence rate of 10.07%**, **average treatment cost of $25,092.73**, and **mean healthcare access at 75.04%**. Variance is high in variables like **population affected** and **treatment costs**, reflecting substantial variability across cases. Most variables exhibit normal distribution characteristics, with minimal skewness and kurtosis deviations.

ii) Outliers

Outliers are extreme large or extreme small values which don't fit in our data . before working further more on the data we have to check outliers . there are two basic methods for detection of outliers
i) Trimmed means

ii) Boxplot / Box and Whisker plot

Commands

Graphs > Legacy dialogue > Boxplot

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Prevalence Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
Incidence Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
Mortality Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
Population Affected	3685	100.0%	0	0.0%	3685	100.0%
Healthcare Access (%)	3685	100.0%	0	0.0%	3685	100.0%
Doctors per 1000	3685	100.0%	0	0.0%	3685	100.0%
Hospital Beds per 1000	3685	100.0%	0	0.0%	3685	100.0%
Average Treatment Cost (USD)	3685	100.0%	0	0.0%	3685	100.0%
Recovery Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
DALYs	3685	100.0%	0	0.0%	3685	100.0%
Improvement in 5 Years (%)	3685	100.0%	0	0.0%	3685	100.0%
Per Capita Income (USD)	3685	100.0%	0	0.0%	3685	100.0%
Education Index	3685	100.0%	0	0.0%	3685	100.0%
Urbanization Rate (%)	3685	100.0%	0	0.0%	3685	100.0%

I can make a box plot but they have no outliers in my data

Interpretation:

The dataset contains 3,685 valid records for all variables, with no missing data across health-related indicators such as prevalence rate, incidence rate, mortality rate, and healthcare access.

iii) Transform Variables

All the qualitative variables are transformed into categorical variable if they needed . We replaced some of the qualitative variables into categorical variables with new names such as ;

DiseaseName into DieseaseName_1

Disease_Category into Disease_Category_1

AgeGroup into AgeGroup_1

Gender into Gender_1

TreatmentType into TreatmentType_1

AvailabilityofVccinesTreatment into AvailabilityofVccinesTreatment_1

- **Commands**
  We can use three different commands for transformation
  - **Transform >Auto Recode**
  - **Transform > recode into same variable**
  - **Transform > recode into different variable**

3) Descriptive statistics & Visualizations

In descriptive statistics we are going to check mean , median , mode and all other values like skewness and kurtosis etc.

Central Tendency

Measure of mean , median , and mode

SPSS calculations

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. D
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Sta
Prevalence Rate (%)	3685	20	0	20	10.07	.094	
Incidence Rate (%)	3685	15	0	15	7.50	.071	
Mortality Rate (%)	3685	10	0	10	5.01	.047	
Population Affected	3685	998583	1372	999955	498832.04	4744.222	28
Healthcare Access (%)	3685	50	50	100	75.04	.239	
Doctors per 1000	3685	5	1	5	2.74	.022	
Hospital Beds per 1000	3685	10	1	10	5.20	.044	
Average Treatment Cost (USD)	3685	49877	117	49994	25092.73	237.469	1
Recovery Rate (%)	3685	49	50	99	74.34	.236	
DALYs	3685	4999	1	5000	2485.69	23.651	
Improvement in 5 Years (%)	3685	10	0	10	5.12	.047	
Per Capita Income (USD)	3685	99493	503	99996	50778.51	469.340	2
Education Index	3685	1	0	1	.65	.002	
Urbanization Rate (%)	3685	70	20	90	54.57	.331	
Valid N (listwise)	3685						

❖ **Visualization**

Bar Chart

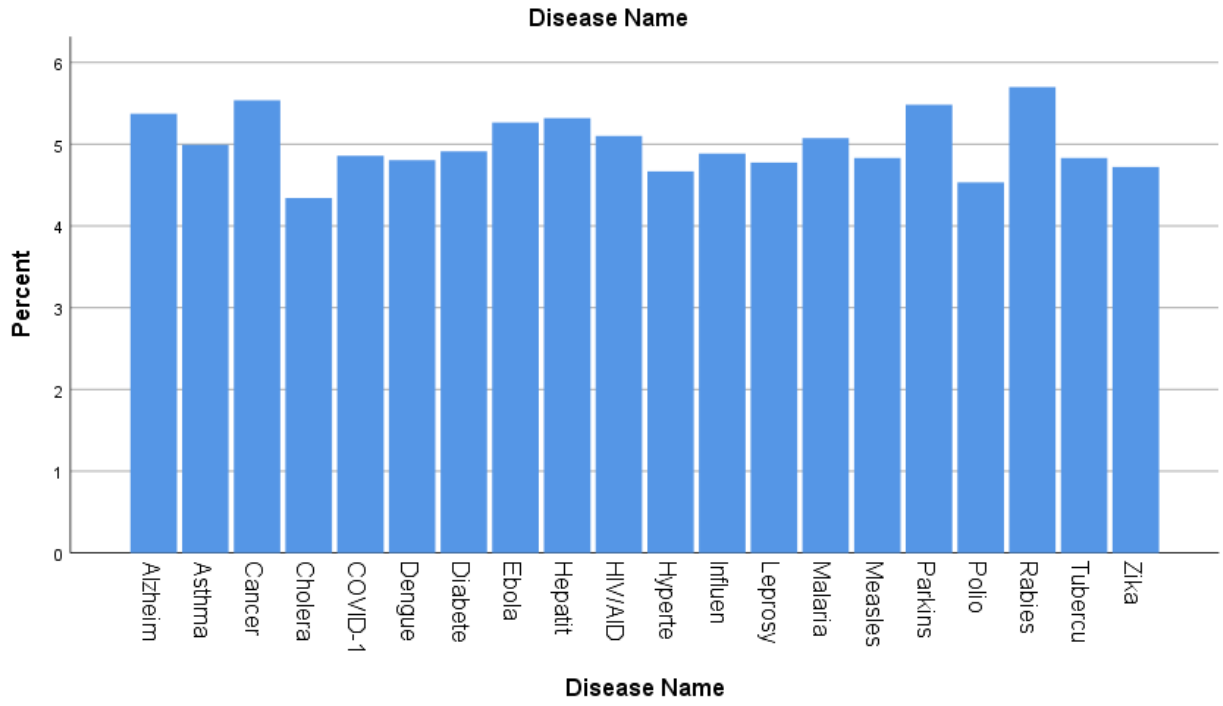
Bar chart represents data in qualitative variables . We select some categorical variable to be represented in Bar Chart .

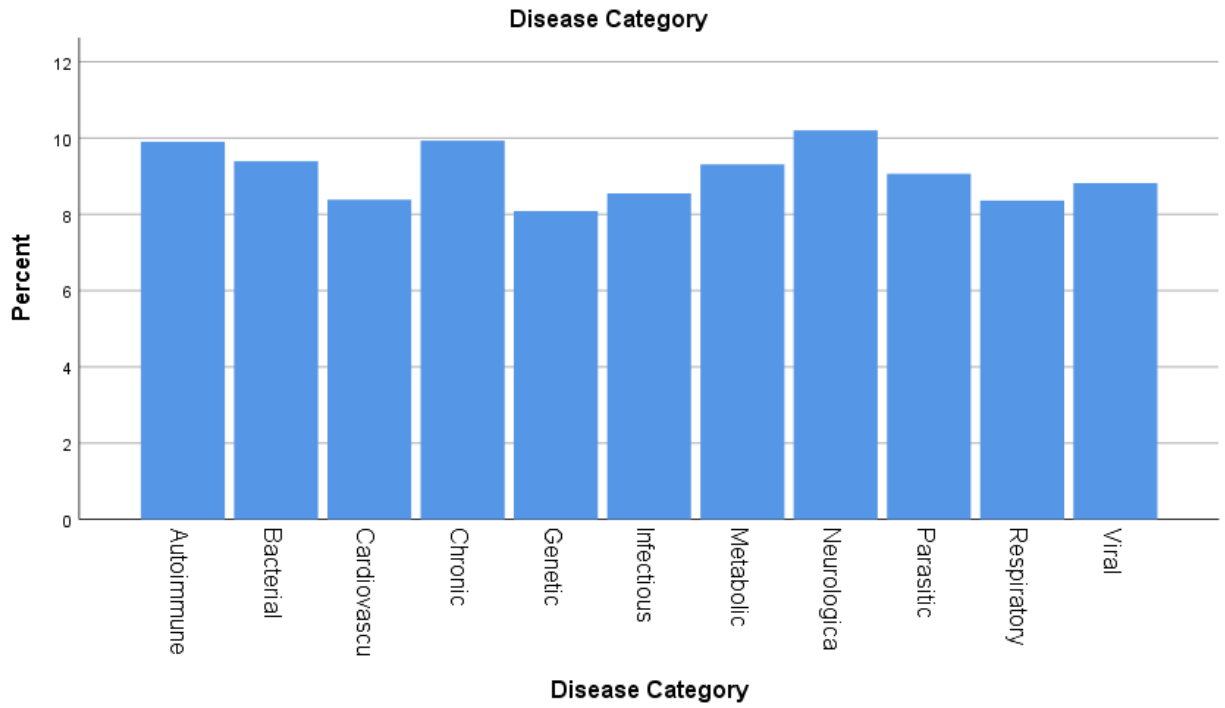
Commands

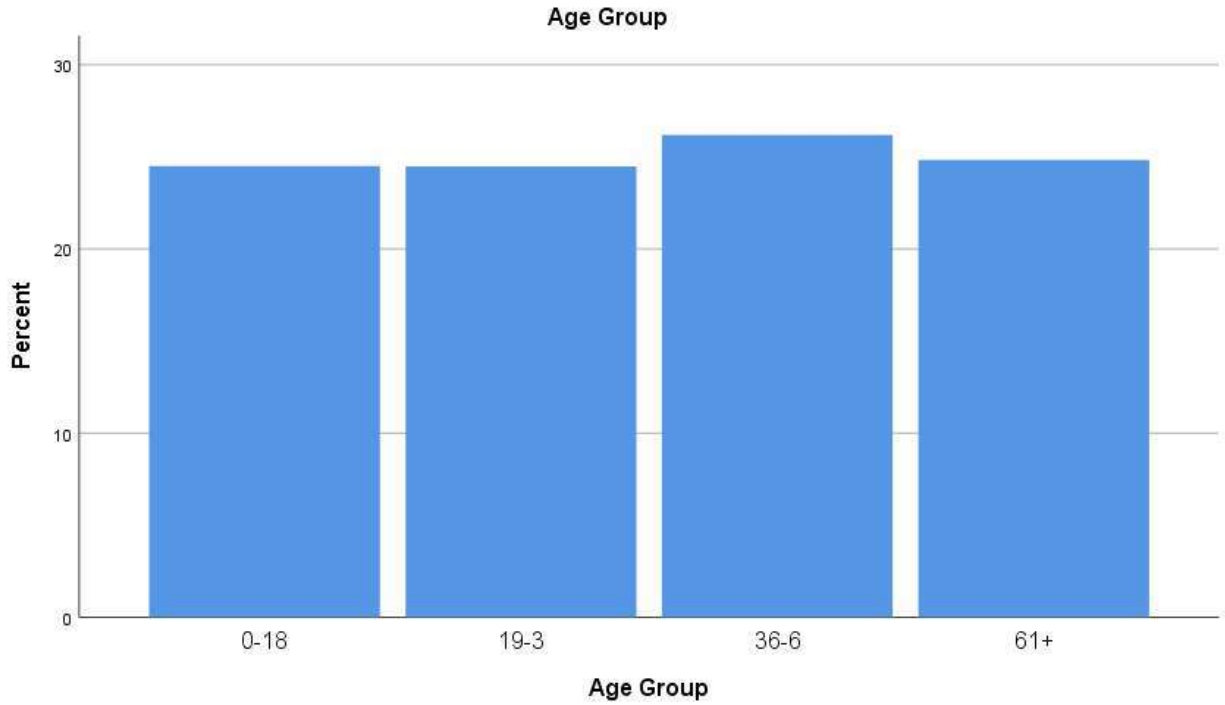
Graphs>Legacy dialog>Barchart

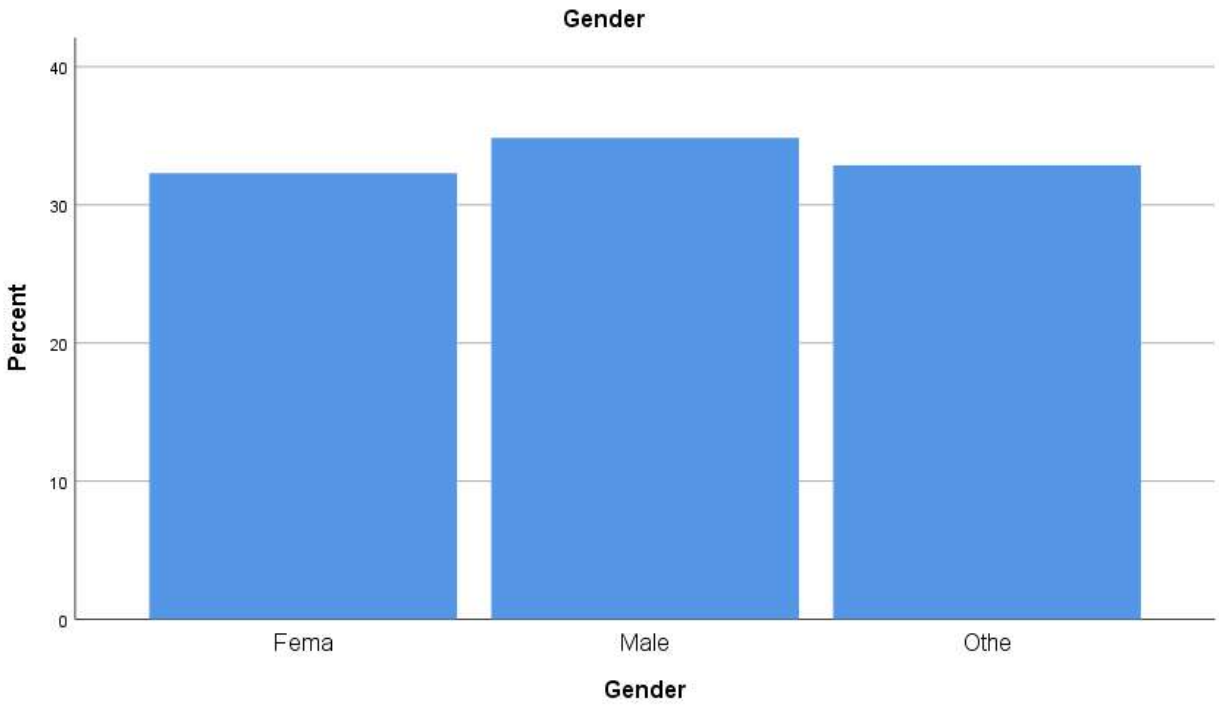
Statistics

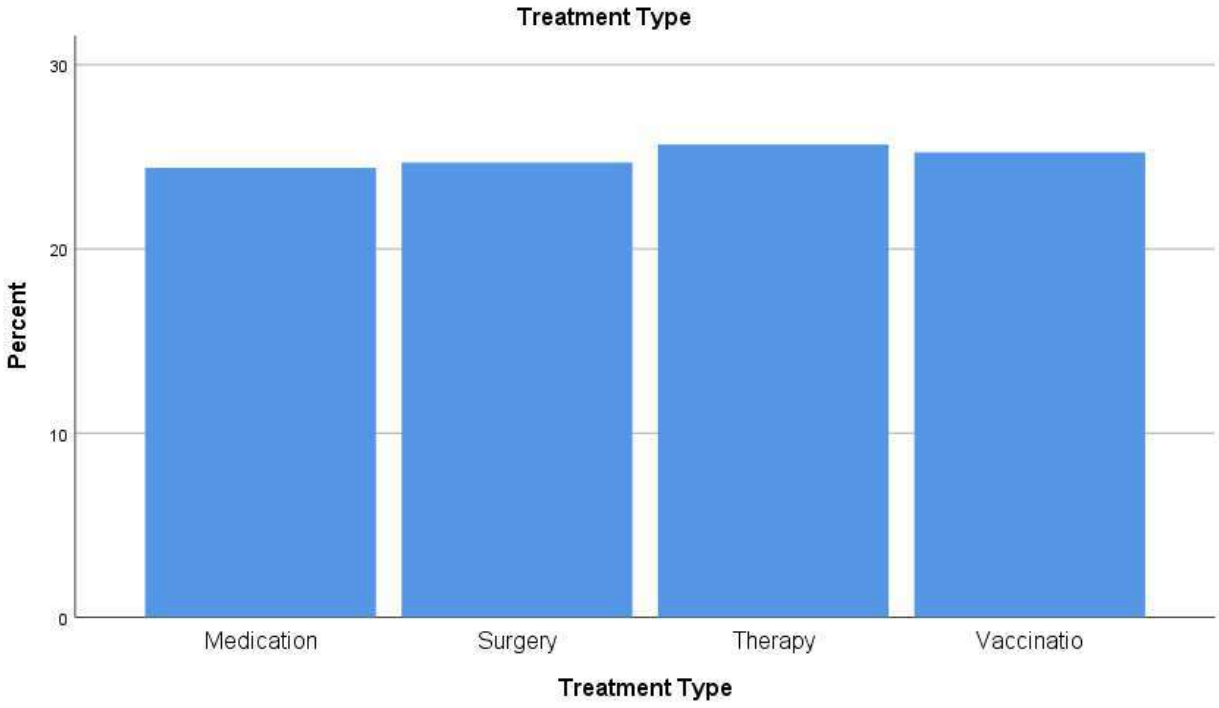
		Disease Name	Disease Category	Age Group	Gender	Treatment Type	Availability of Vaccines/Treatment
N	Valid	3685	3685	3685	3685	3685	3685
	Missing	0	0	0	0	0	0

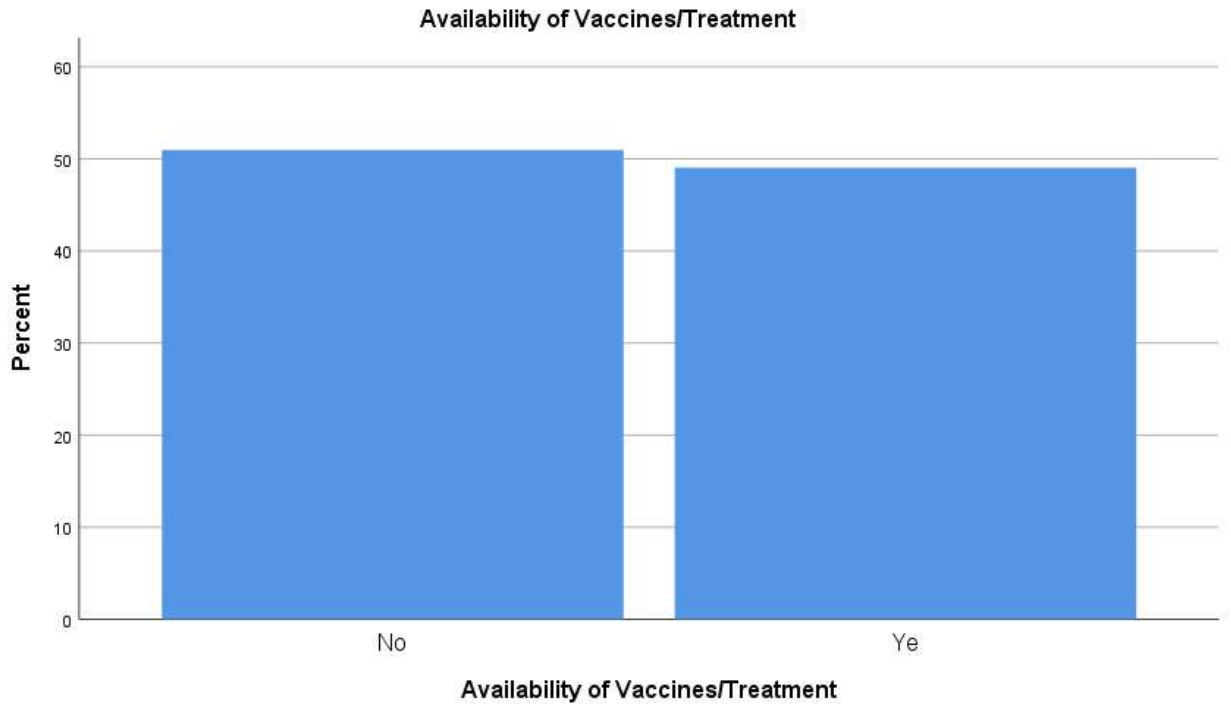












Pie Chart

Pie chart is another method of representing qualitative data . Let's represent our data in pie chart using previous variables .

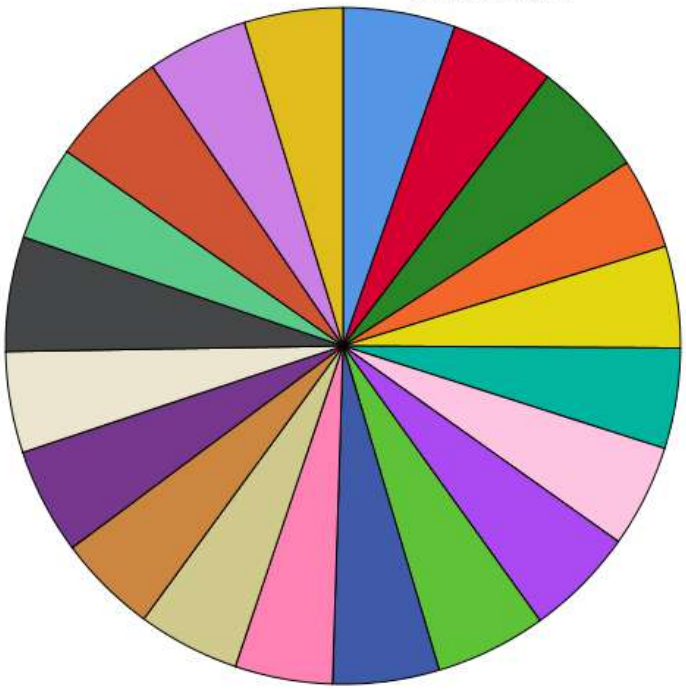
Commands

Graphs>Legacy Dialog>pie chart

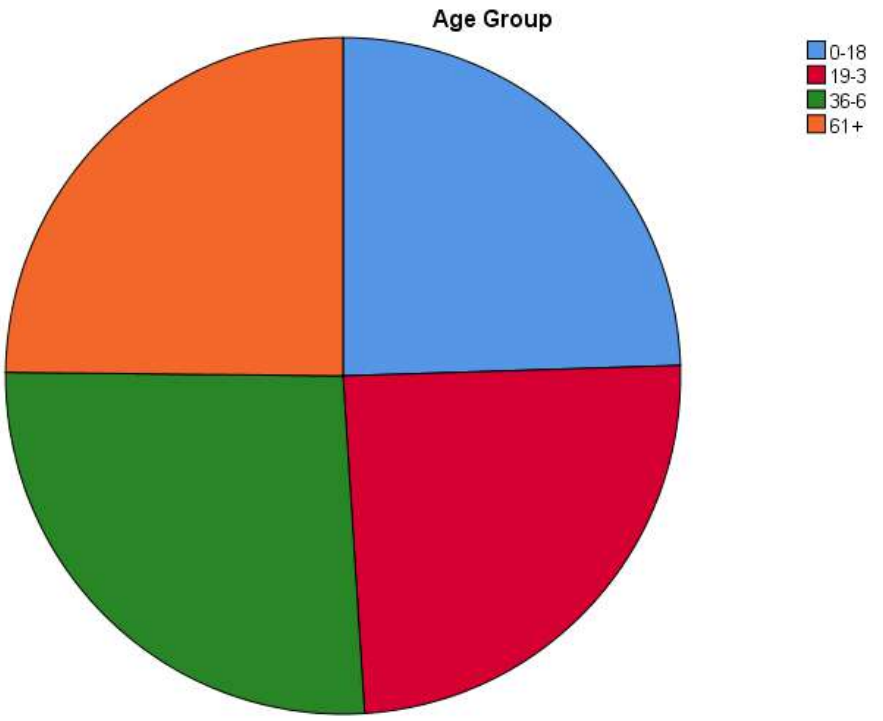
Statistics

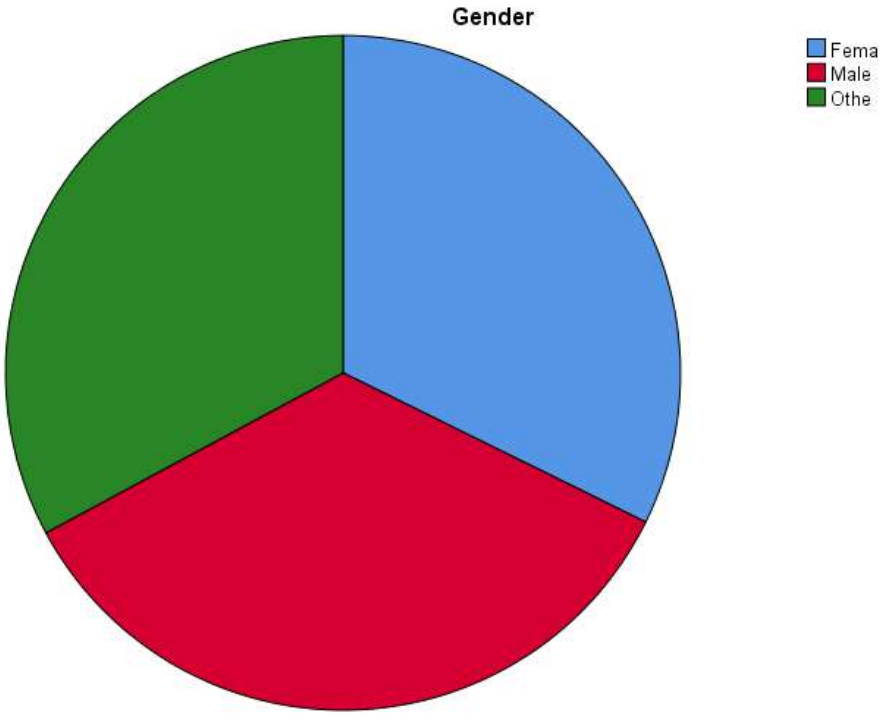
		Disease Name	Disease Category	Age Group	Gender	Treatment Type	Availability of Vaccines/Treatment
N	Valid	3685	3685	3685	3685	3685	3685
	Missing	0	0	0	0	0	0

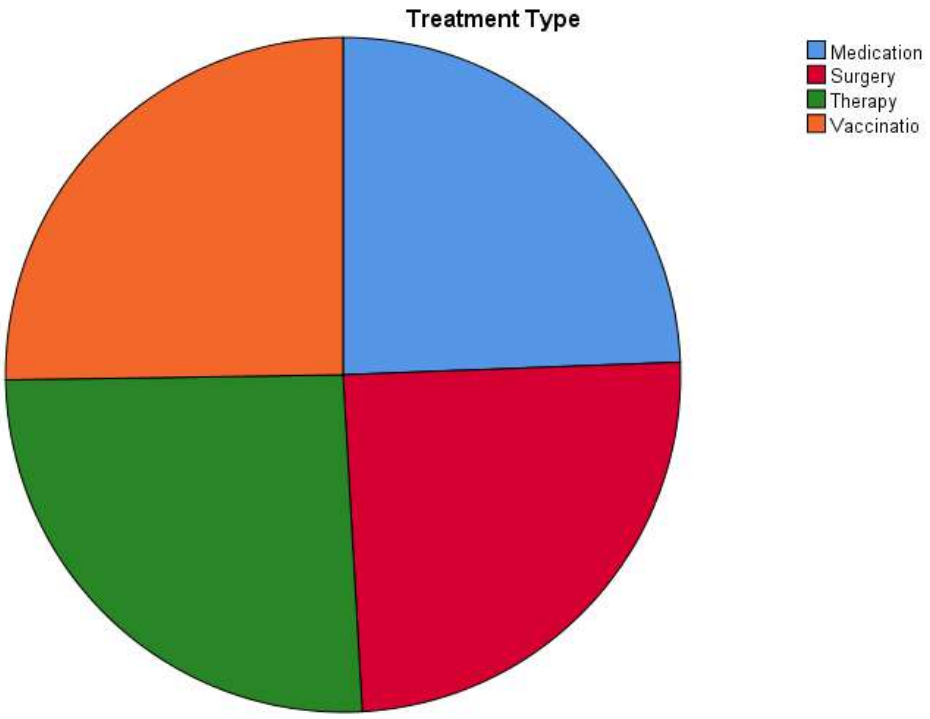
Disease Name

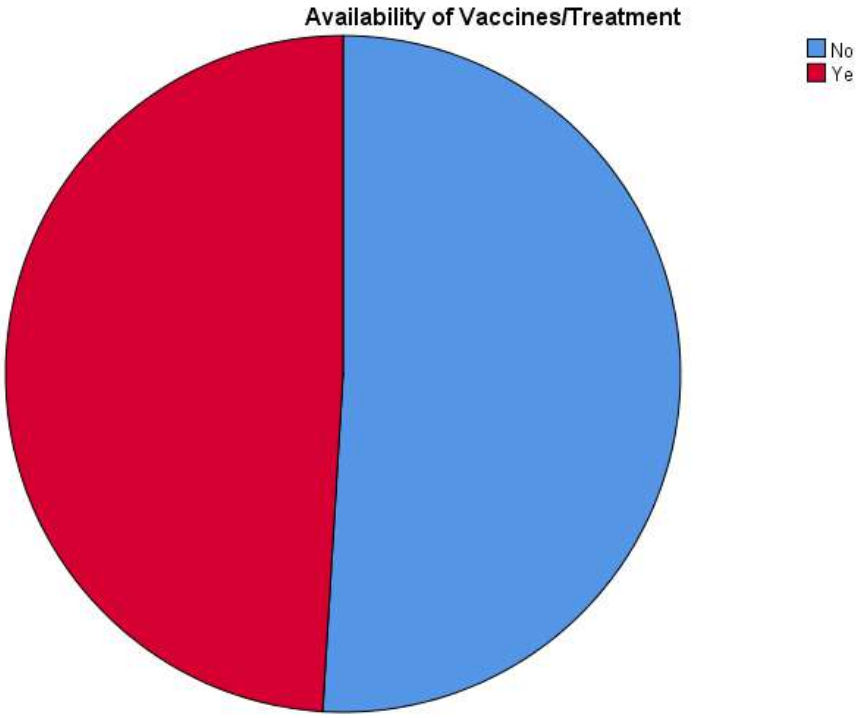


Disease Category









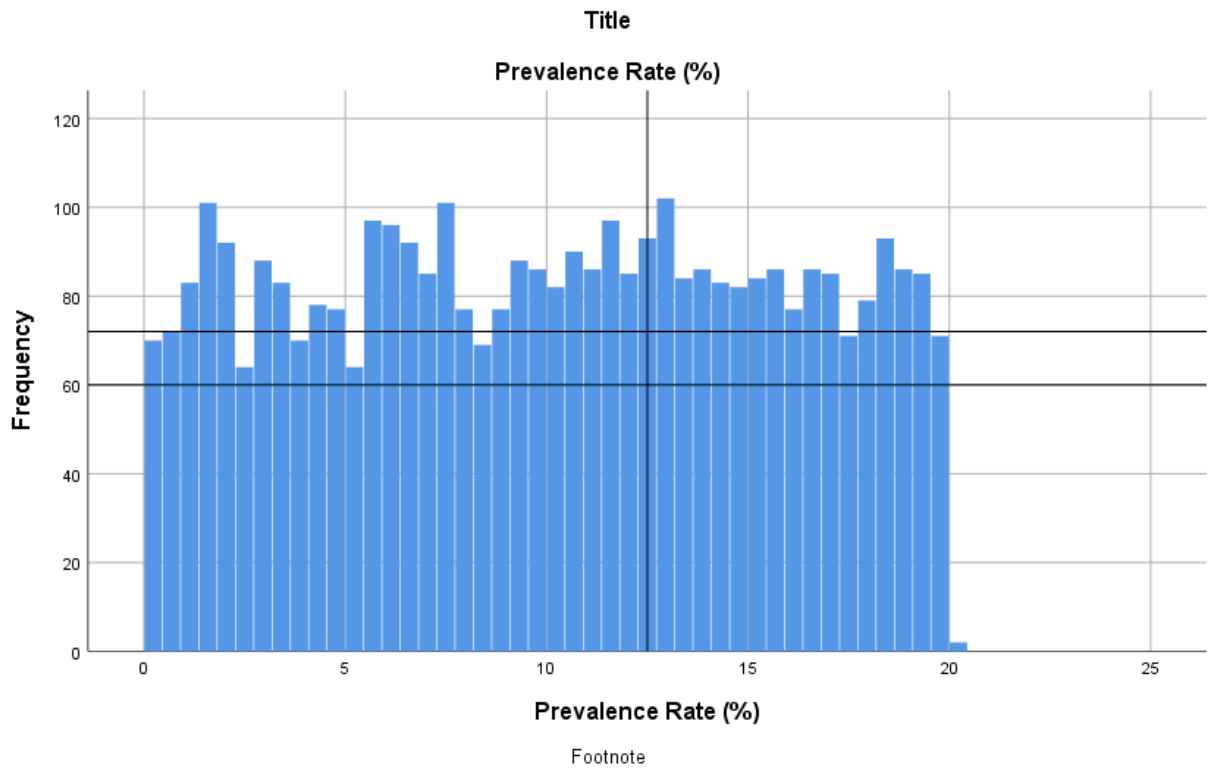
Histogram

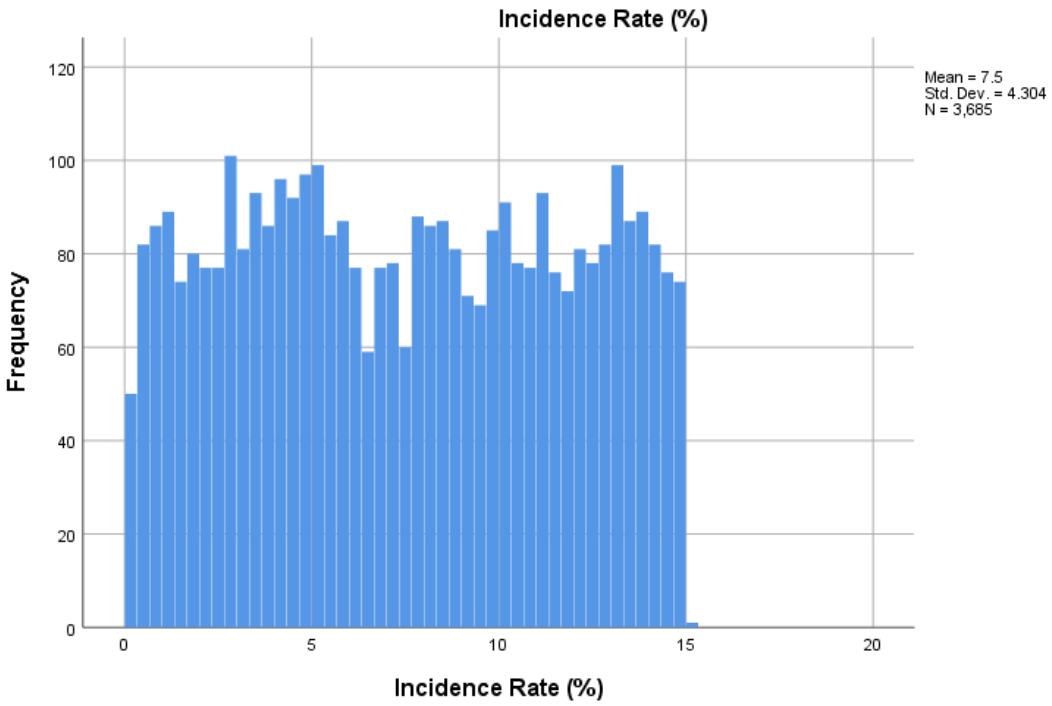
histogram is one of the most reliable way of representing quantitative data . We have here only 3 quantitative variables .So we represent each of them in histogram . There are two ways for making a histogram in SPSS. We will discuss both .

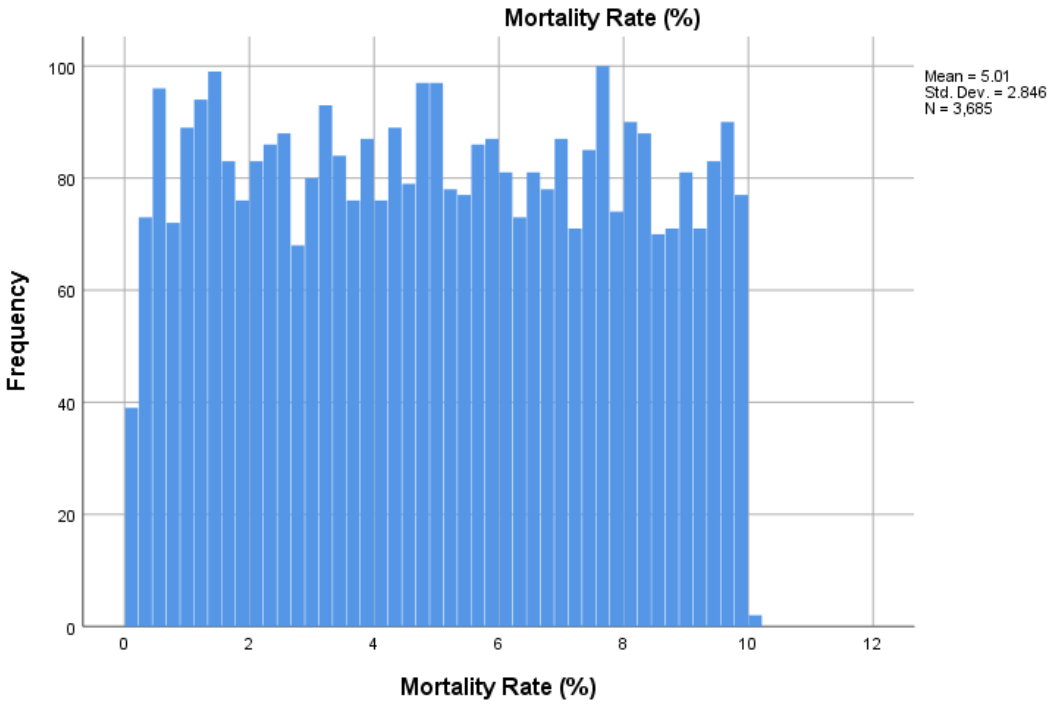
Commands

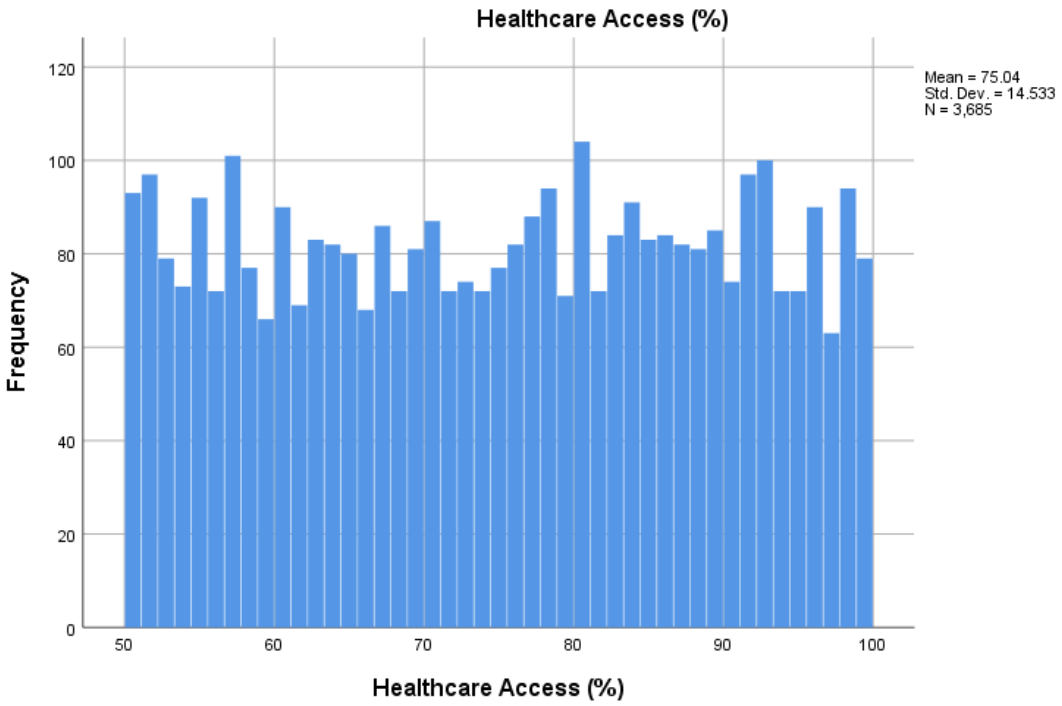
- **Graph>Legacy Dialog> Histogram**
- **Analyze>Descriptive Statistics>Explore (**Here we select histogram in plots **)**

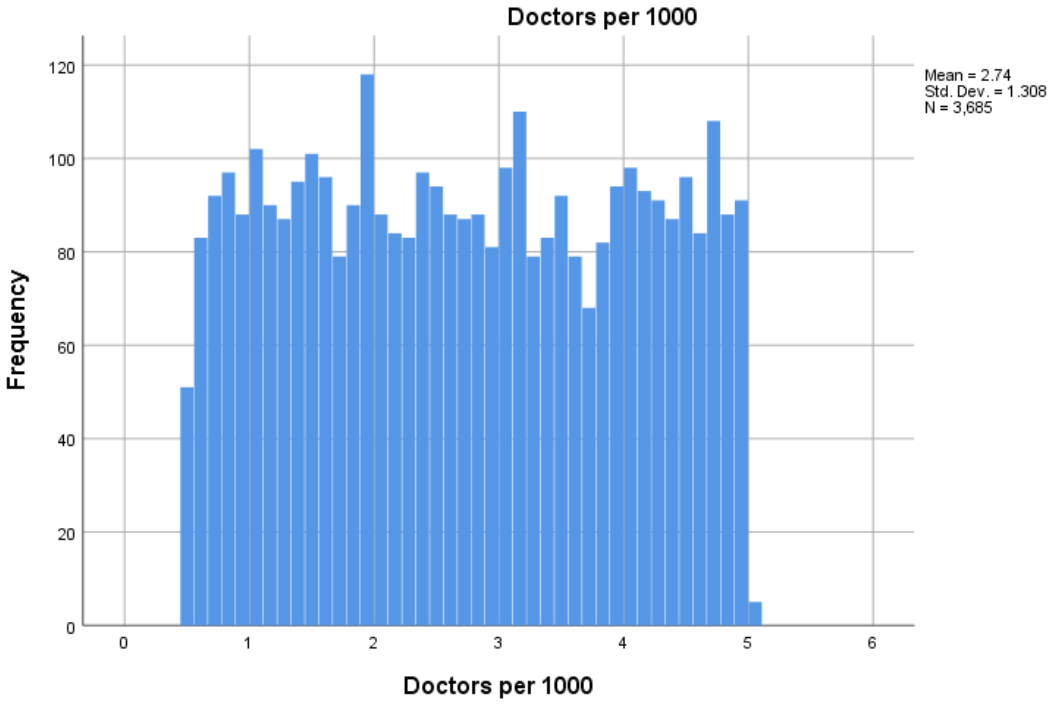
		Prevalence Rate (%)	Incidence Rate (%)	Mortality Rate (%)	Population Affected	Healthcare Access (%)	D
N	Valid	3685	3685	3685	3685	3685	
	Missing	0	0	0	0	0	

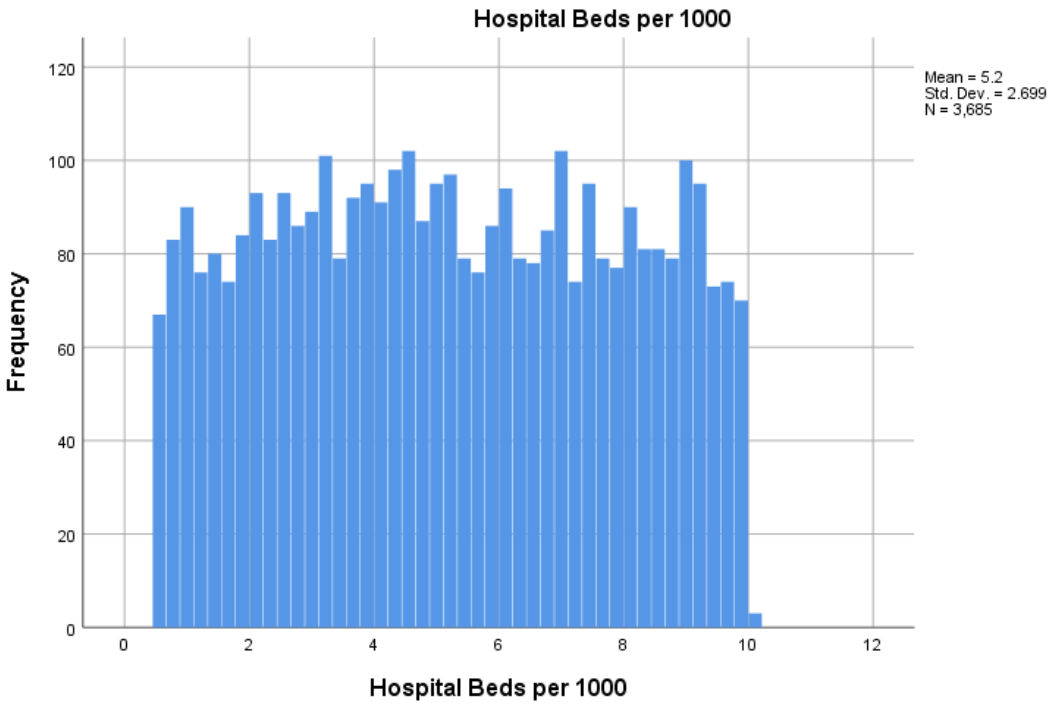


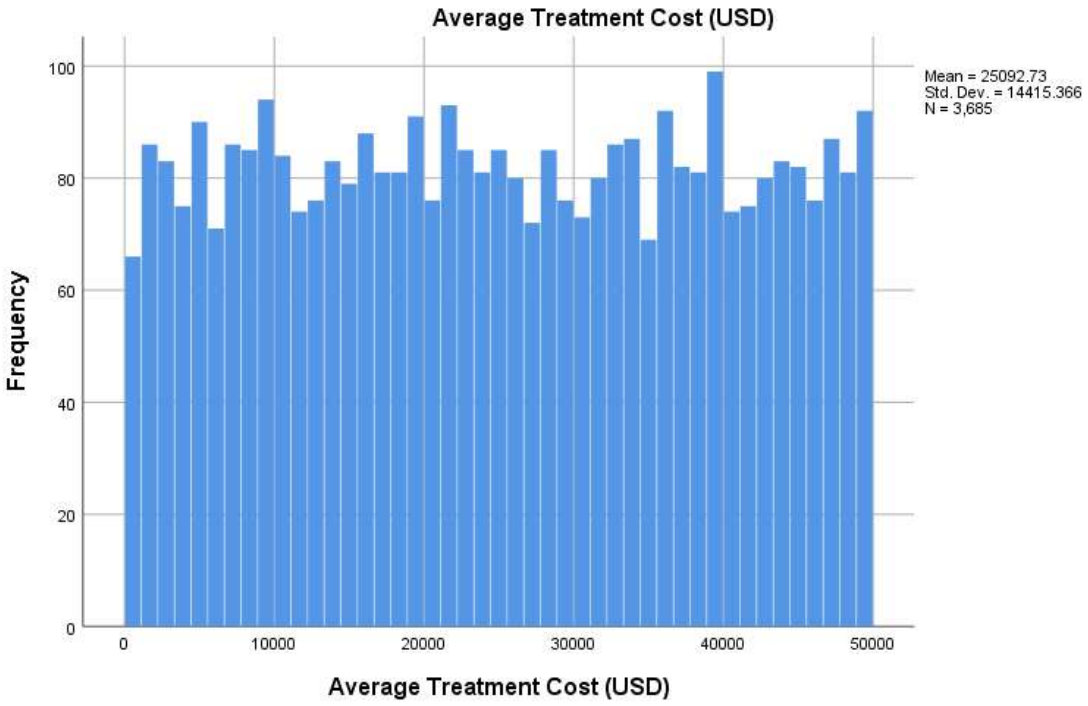


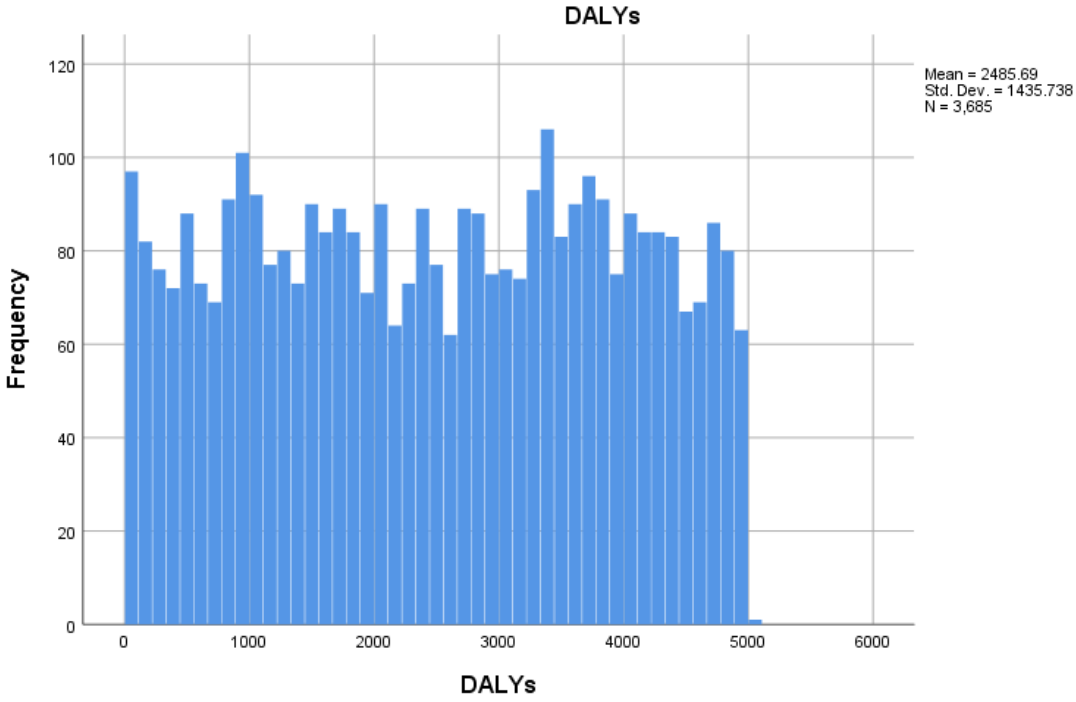


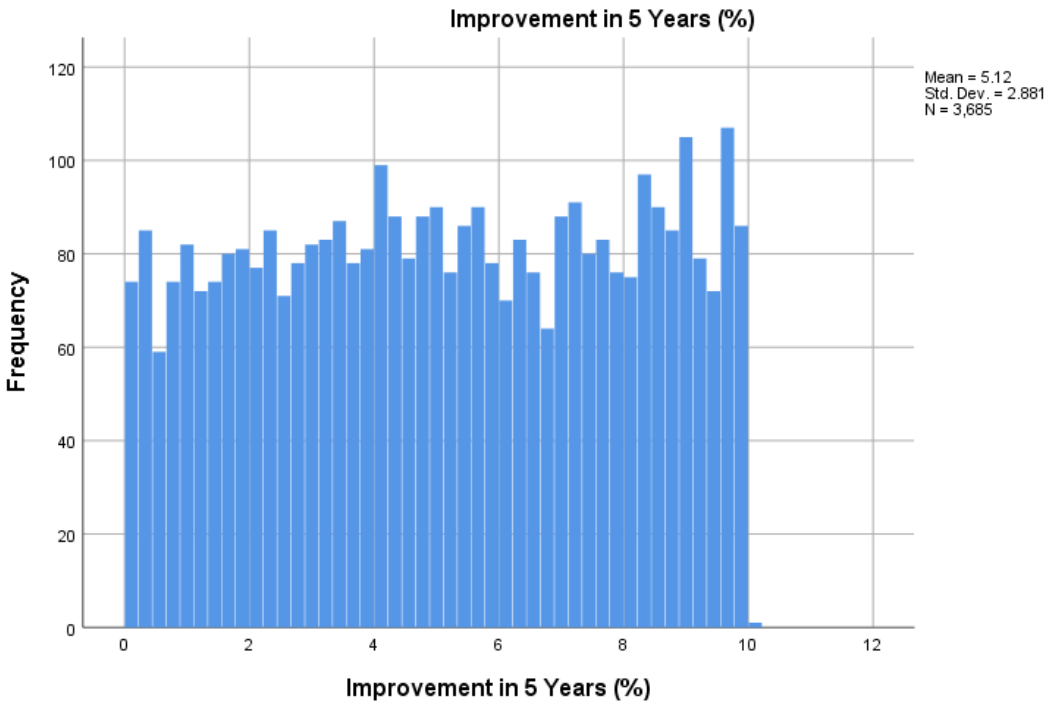


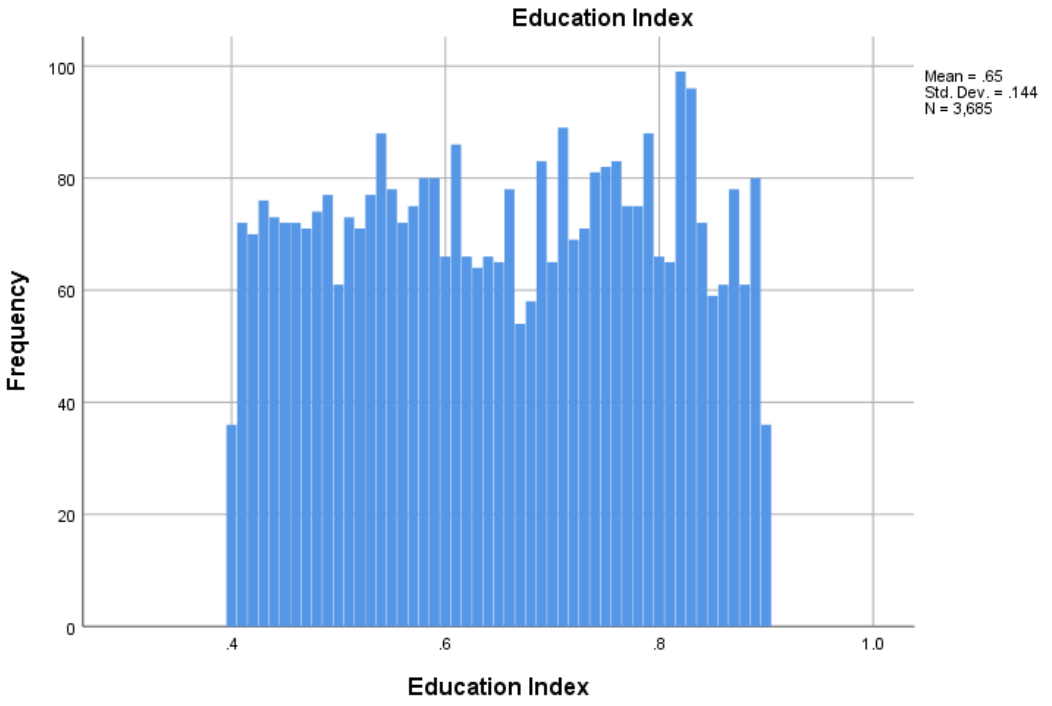


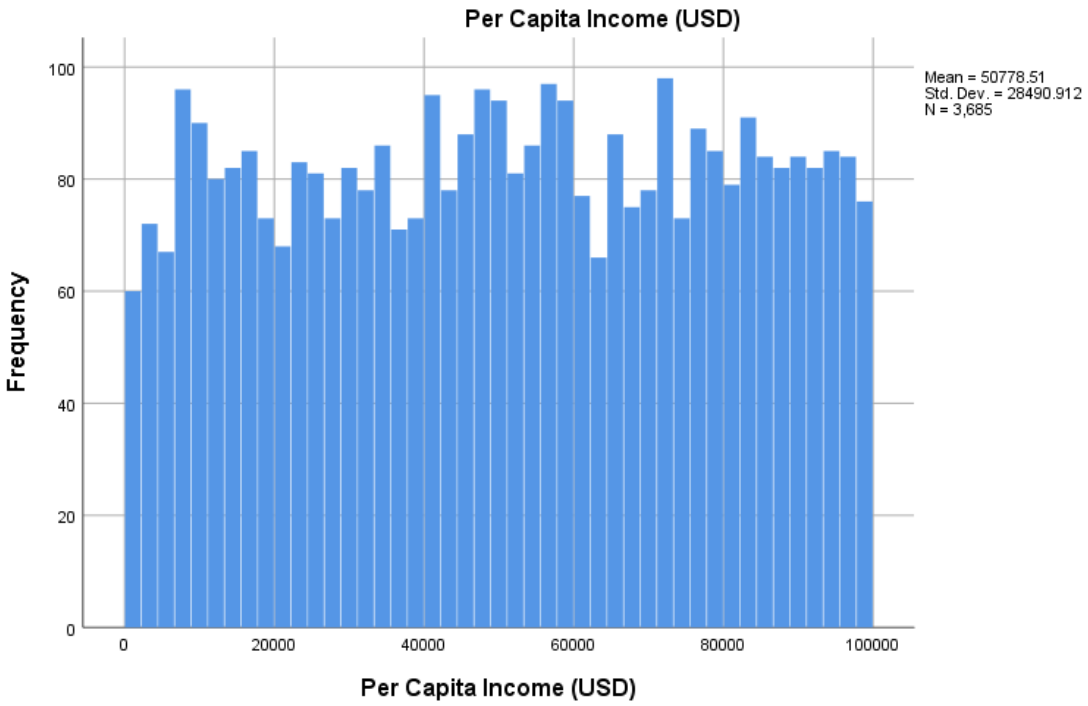


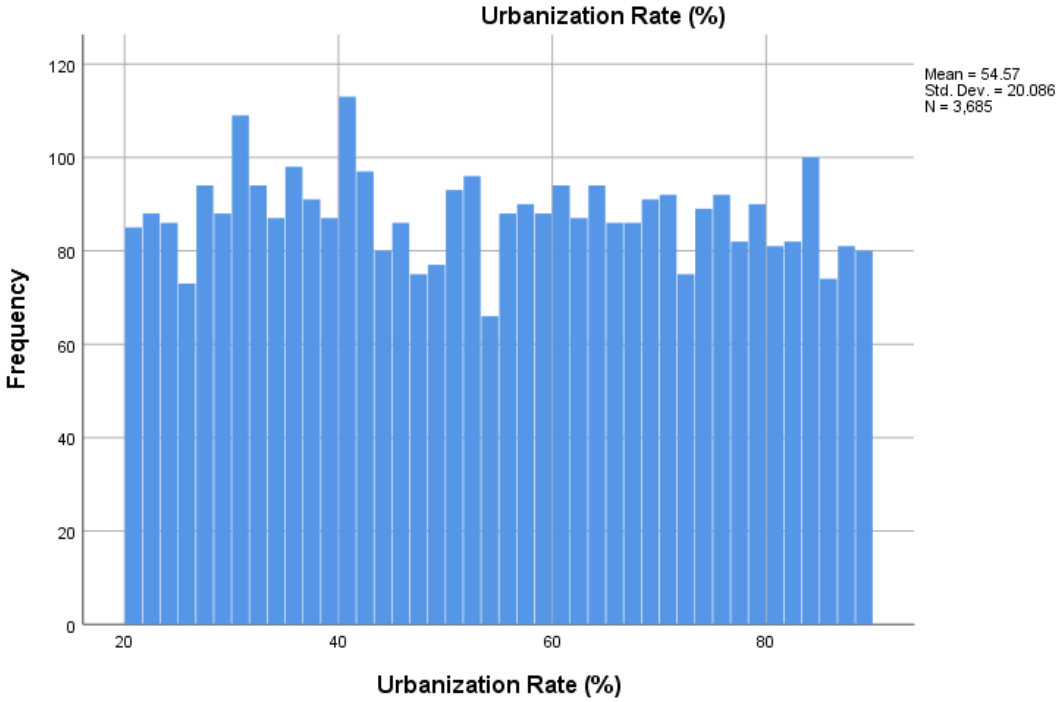












Box Plot

Box plot is usually termed as Box and Whisker plot . We use it to assess normality of the data . It's Box and Whiskers help us to detect normality of data. In SPSS ,Box and Whisker Plot ,can be drawn by two ways .

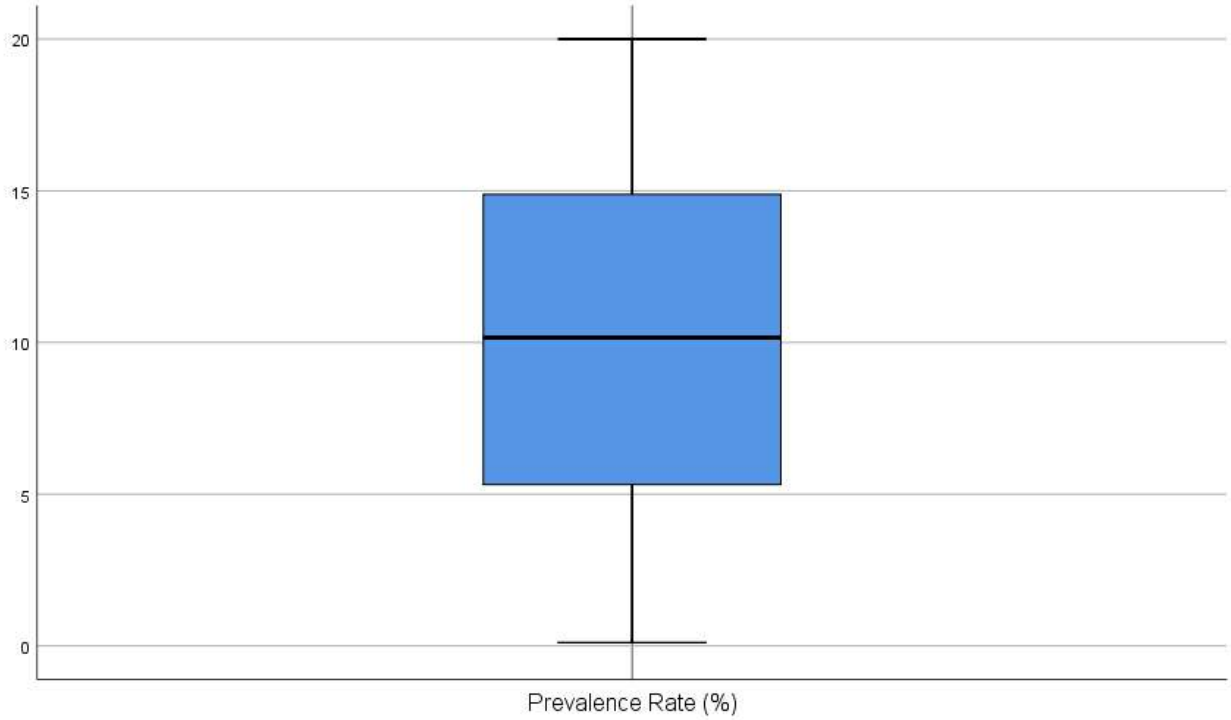
Commands

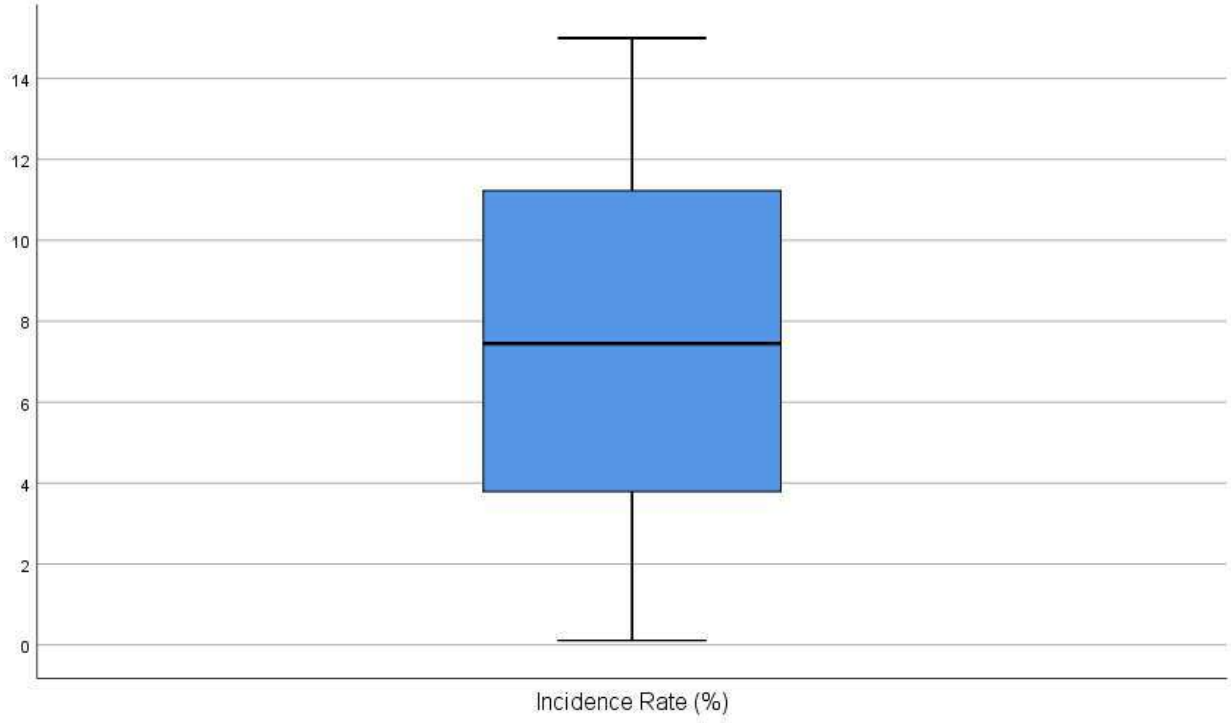
- **Graph>Legacy Dialog> Boxplot**
- **Analyze>Descriptive Statistics>Explore** (Here we select Box and whisker in plots)

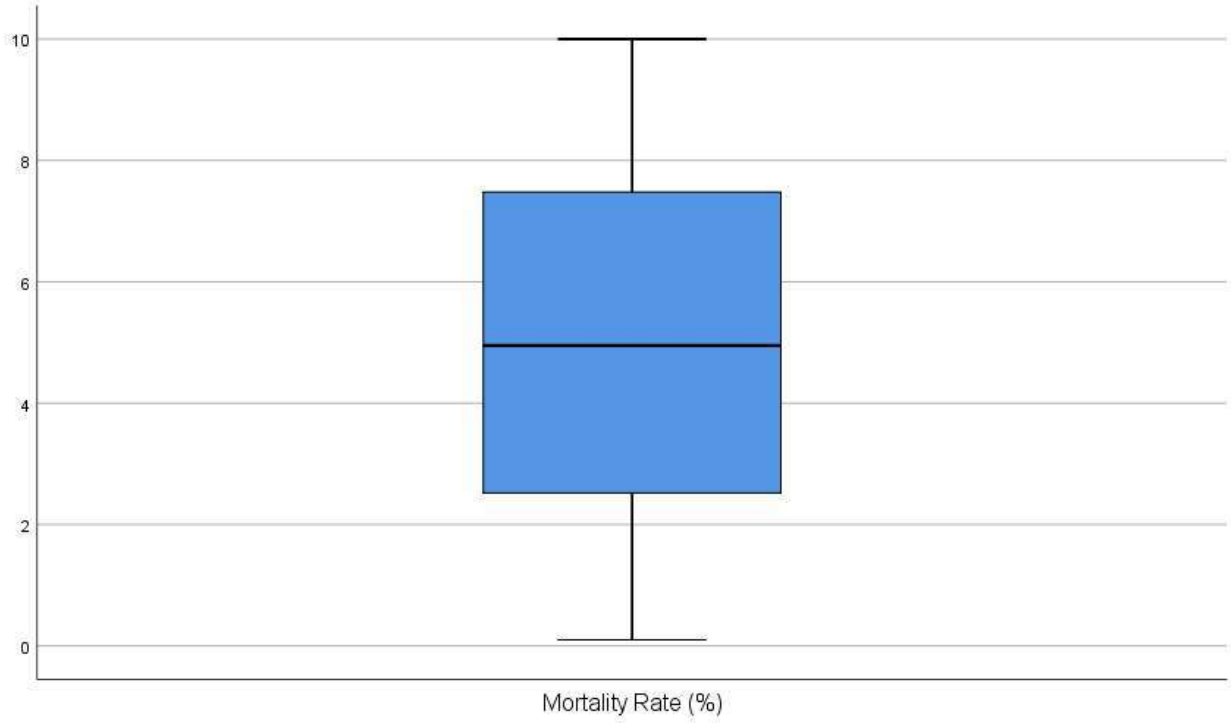
Case Processing Summary

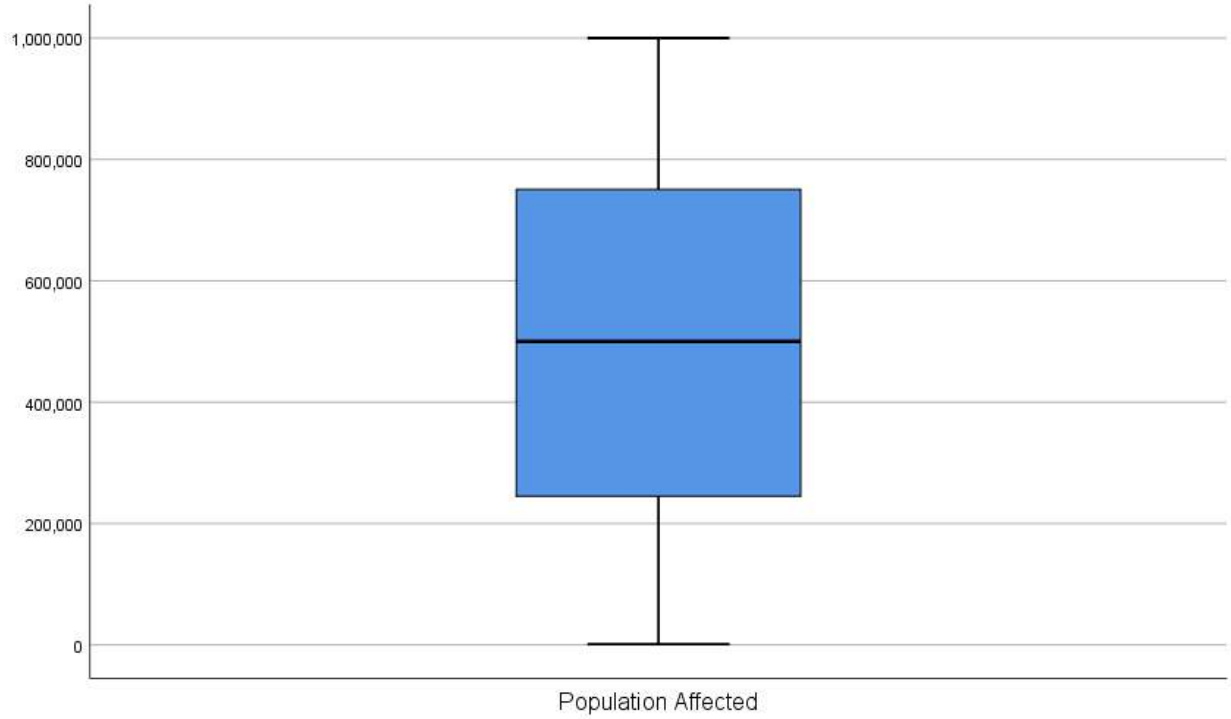
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Prevalence Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
Incidence Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
Mortality Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
Population Affected	3685	100.0%	0	0.0%	3685	100.0%
Healthcare Access (%)	3685	100.0%	0	0.0%	3685	100.0%
Doctors per 1000	3685	100.0%	0	0.0%	3685	100.0%
Hospital Beds per 1000	3685	100.0%	0	0.0%	3685	100.0%
Average Treatment Cost (USD)	3685	100.0%	0	0.0%	3685	100.0%
Recovery Rate (%)	3685	100.0%	0	0.0%	3685	100.0%
DALYs	3685	100.0%	0	0.0%	3685	100.0%
Improvement in 5 Years (%)	3685	100.0%	0	0.0%	3685	100.0%
Per Capita Income (USD)	3685	100.0%	0	0.0%	3685	100.0%
Education Index	3685	100.0%	0	0.0%	3685	100.0%
Urbanization Rate (%)	3685	100.0%	0	0.0%	3685	100.0%

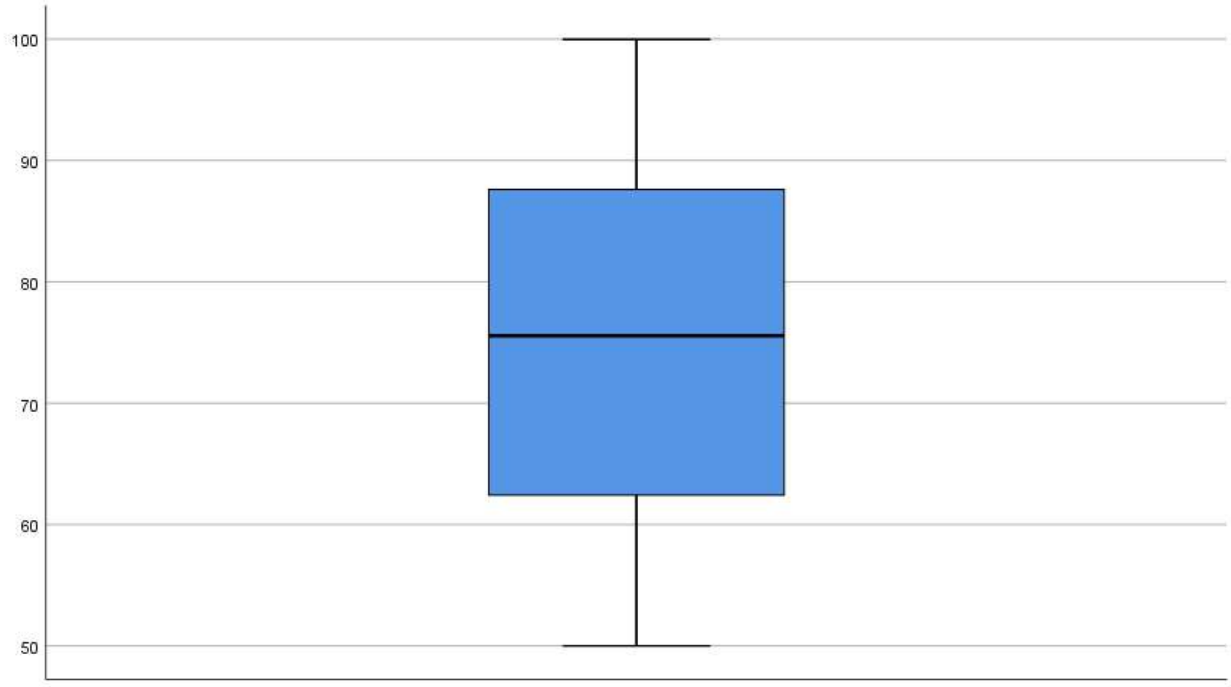
Prevalence Rate (%)

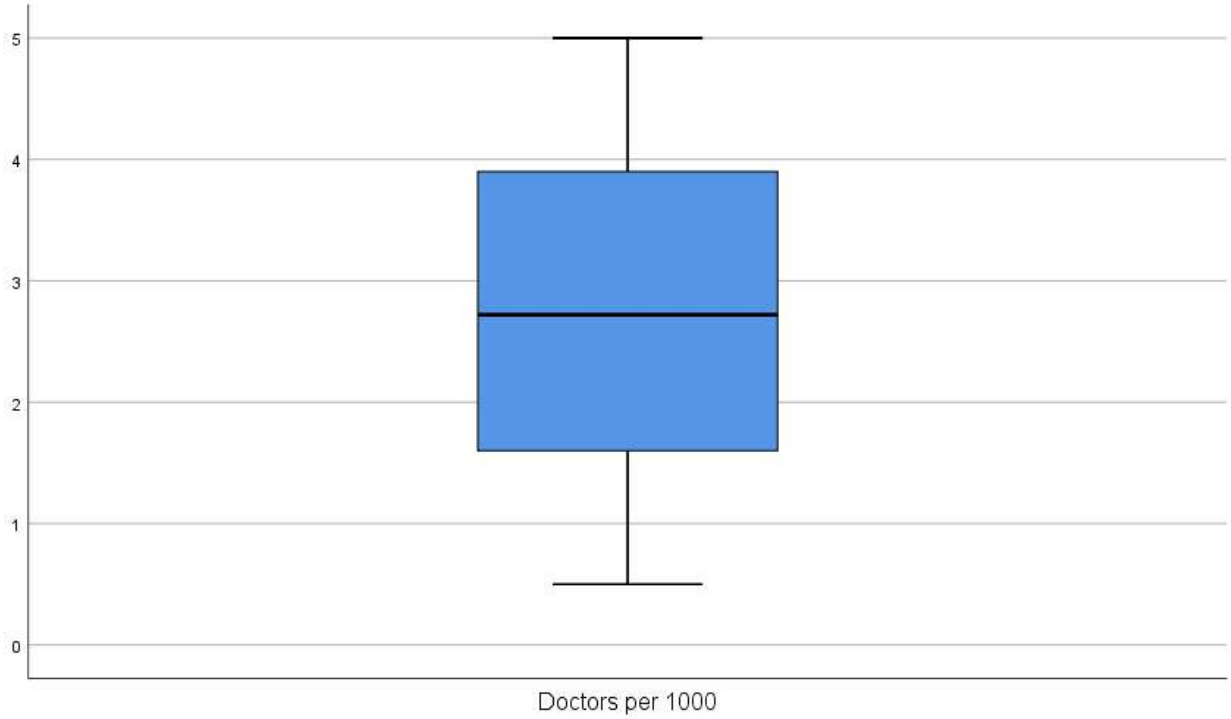


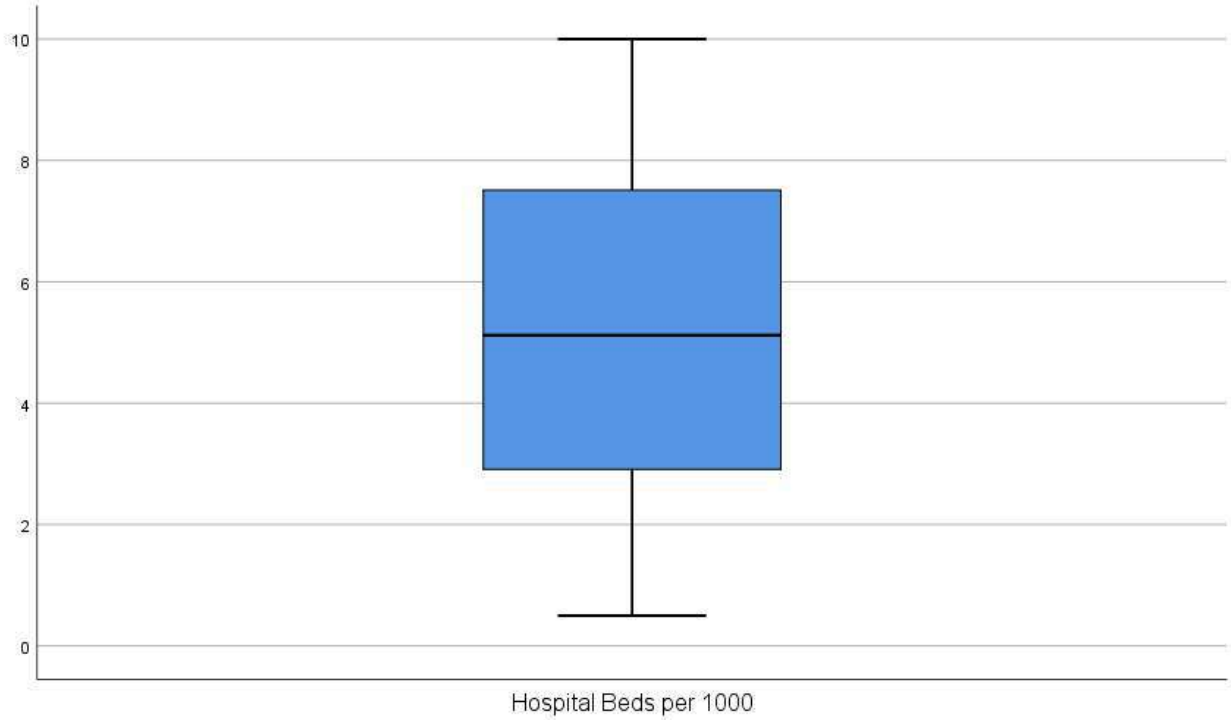


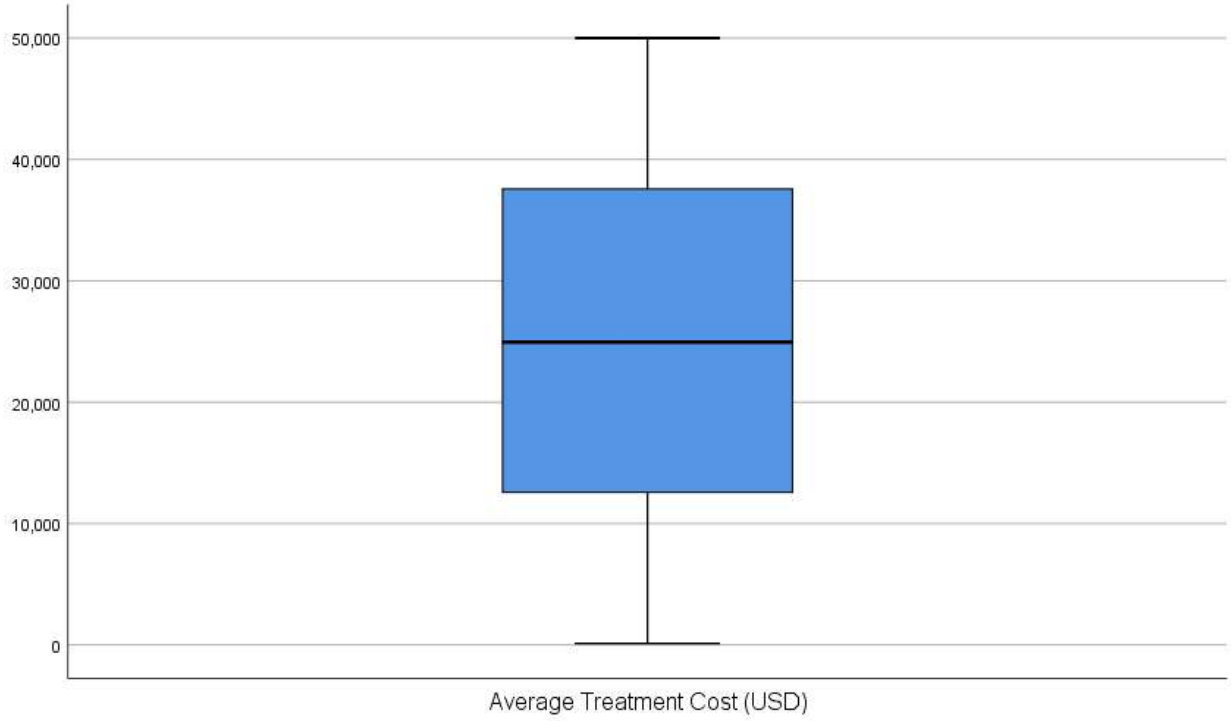


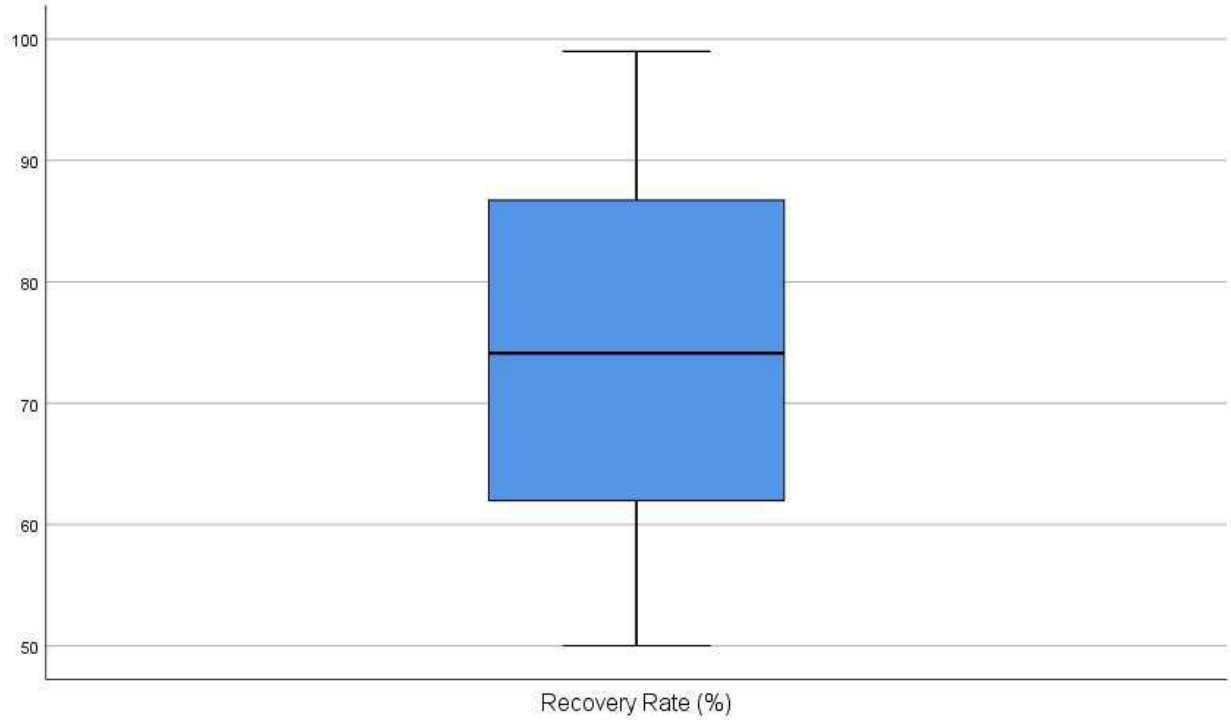


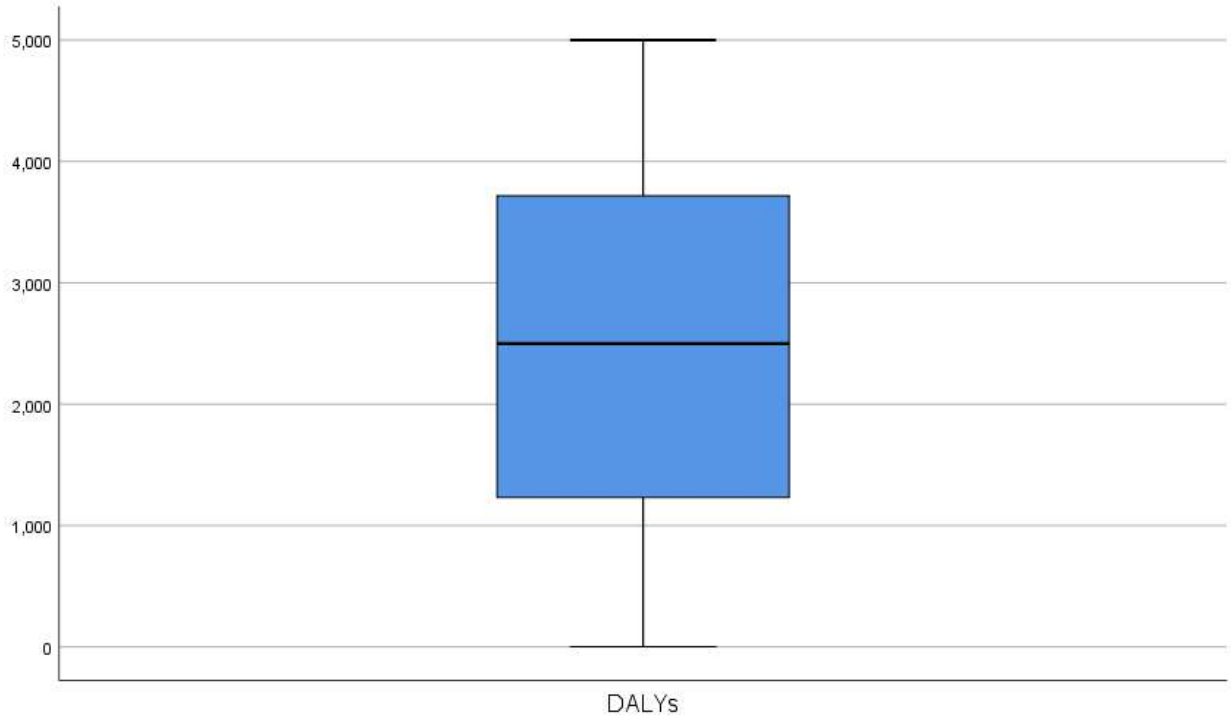


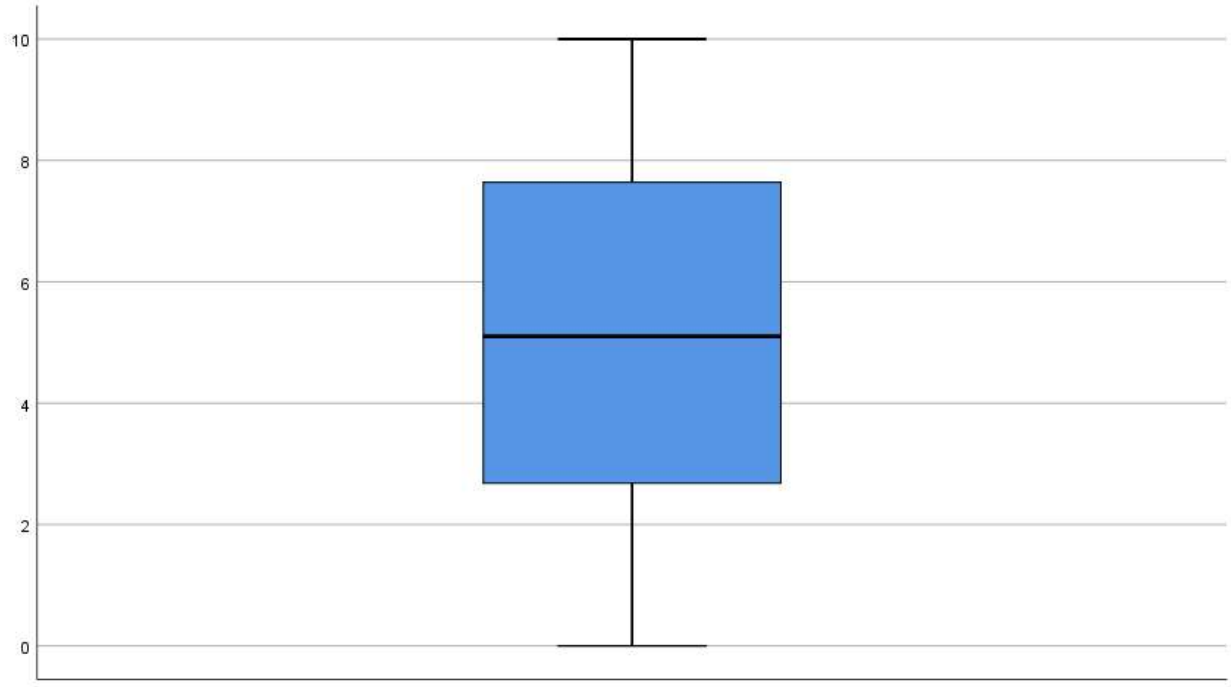


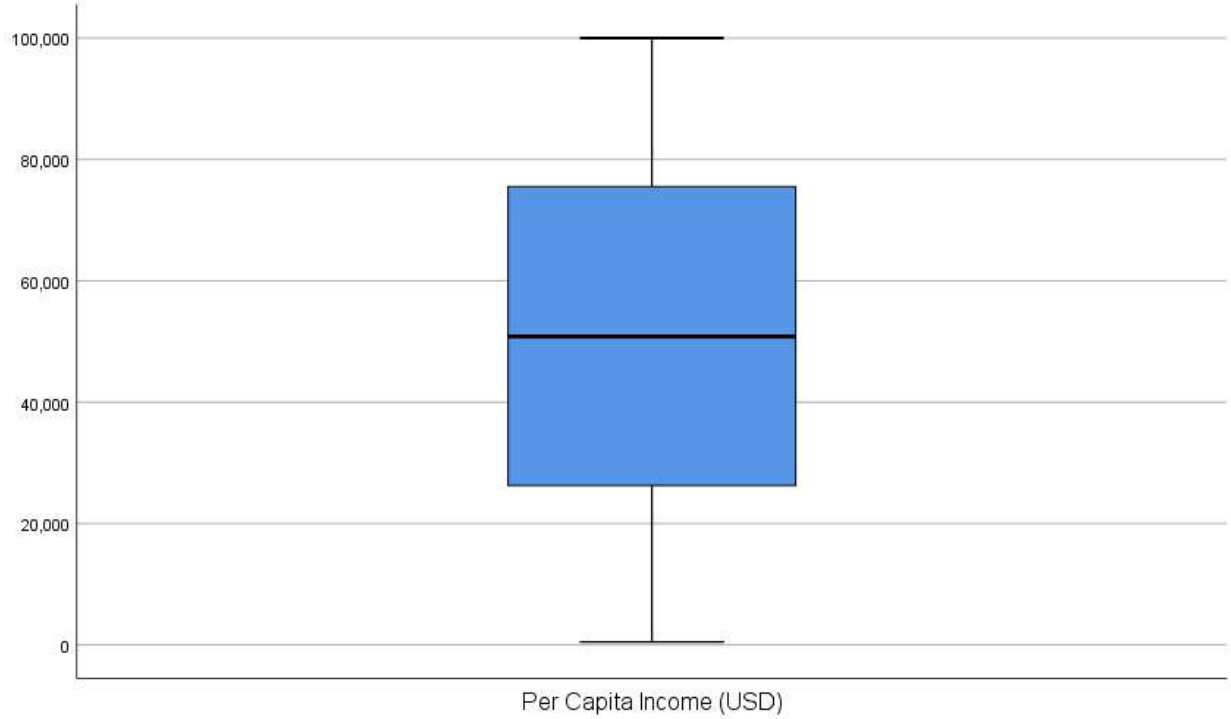


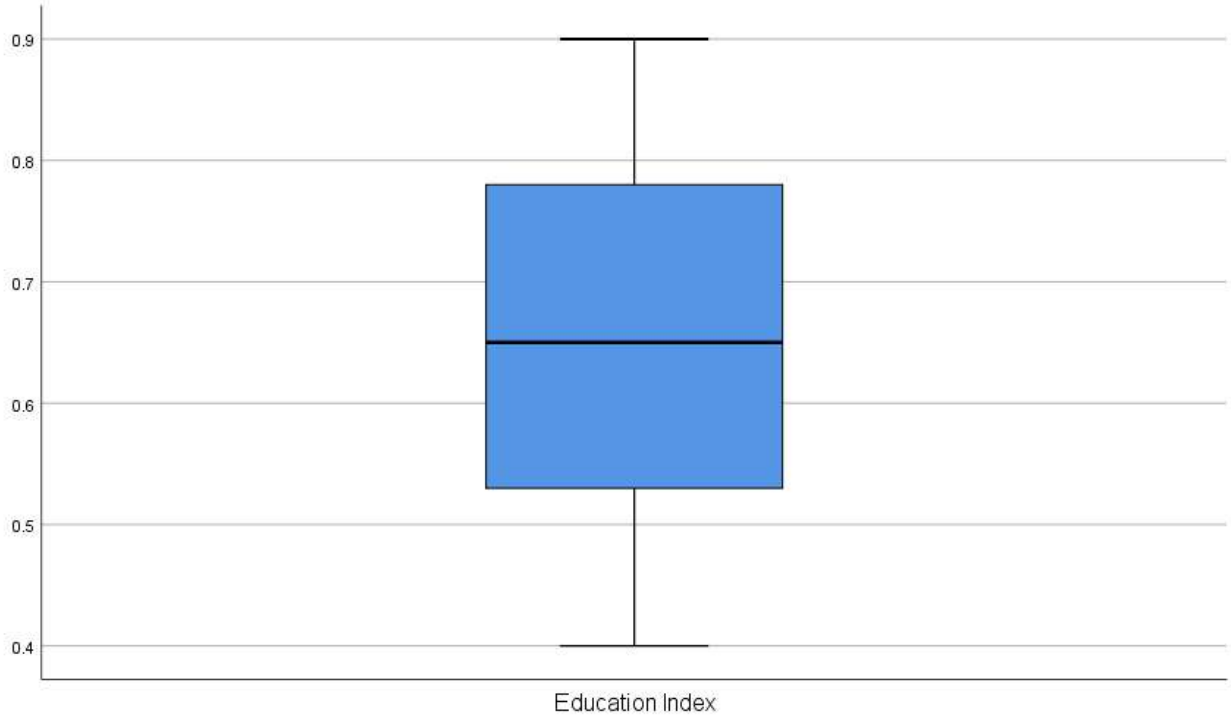


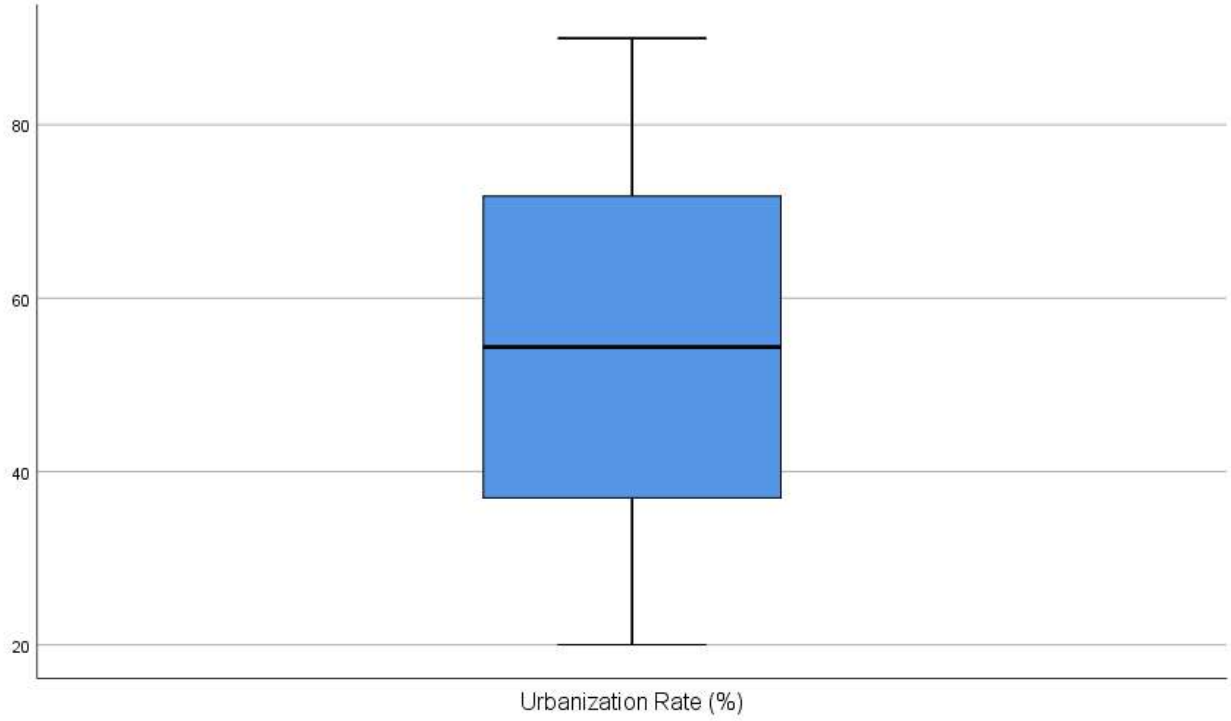












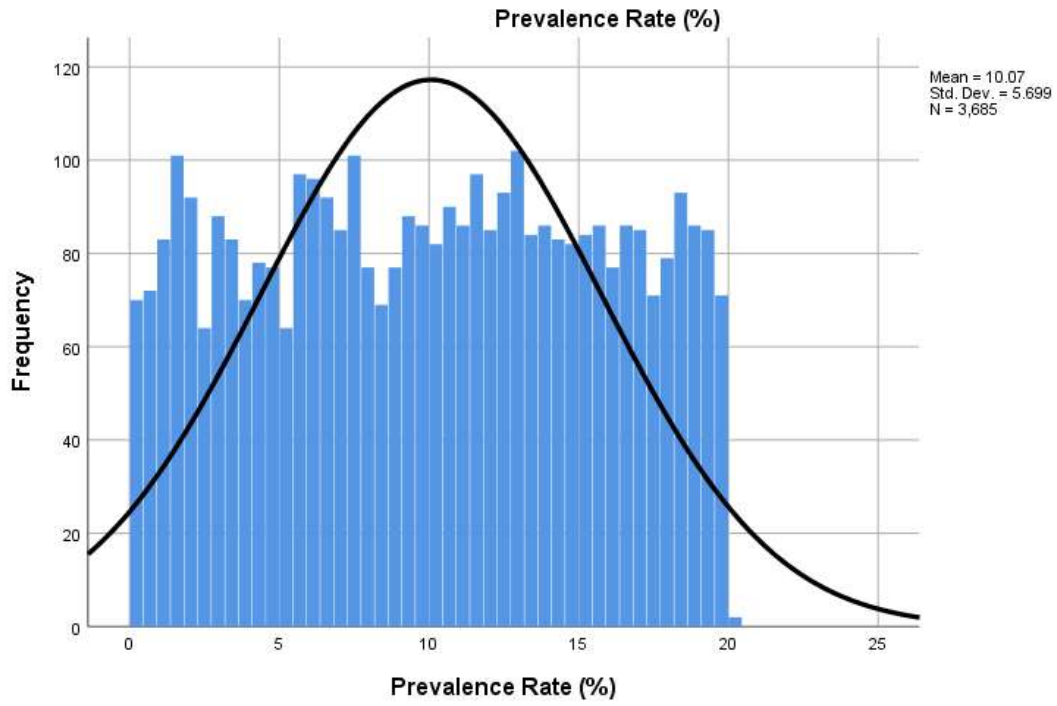
Normality Assessment

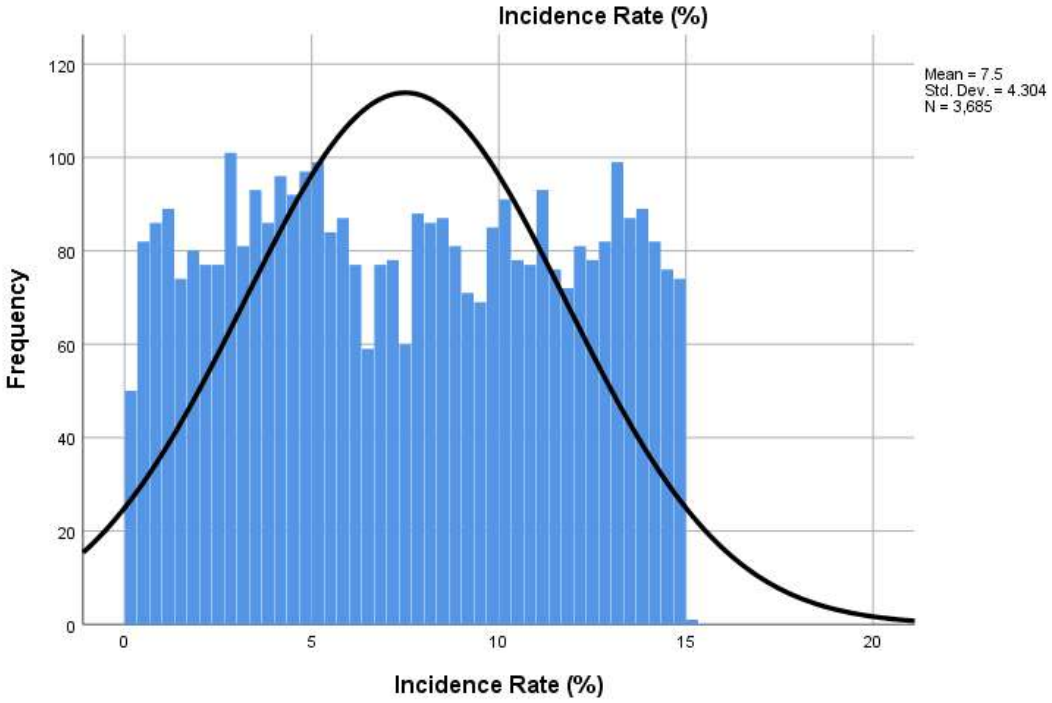
normality assessments are basically tests that show either data is normal or not. For assessing normality of data we perform Kolmogorov Smirnov and Shapiro Wilk tests

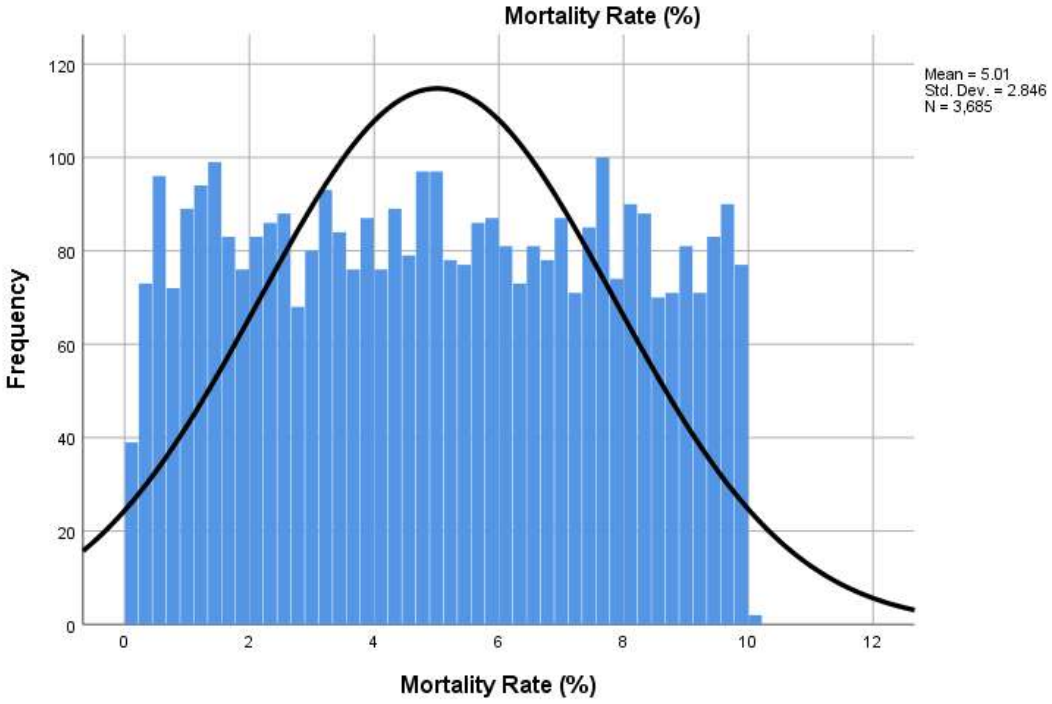
- **Commands**

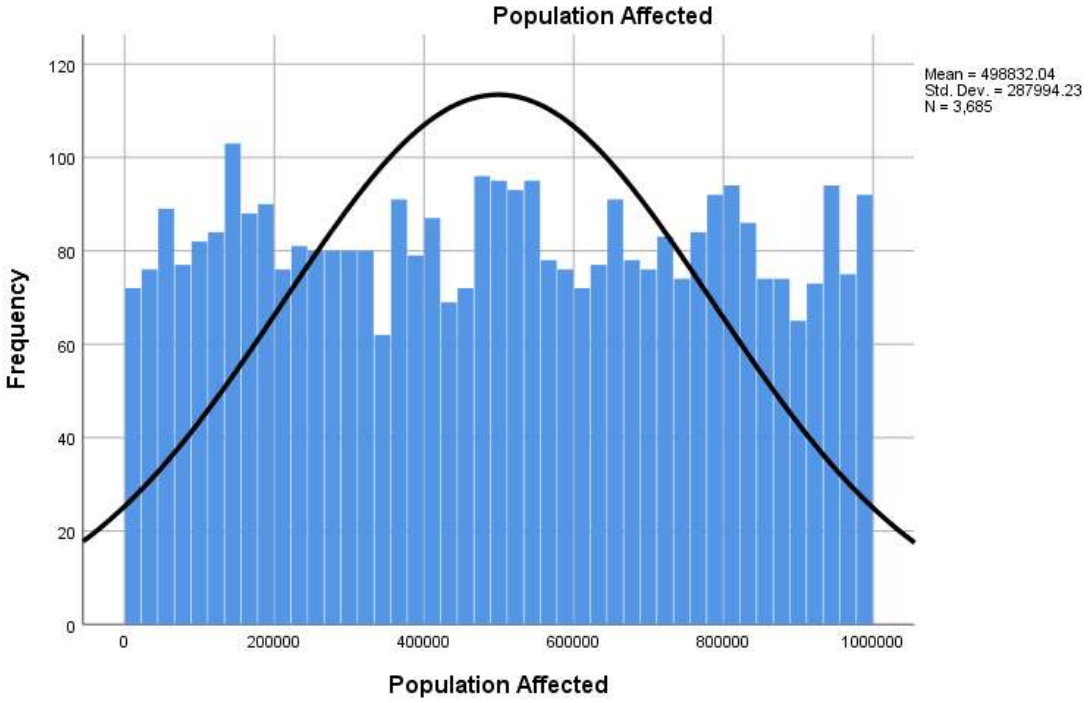
Analyze>Descriptive Statistics>Explore(in option check normality**)**

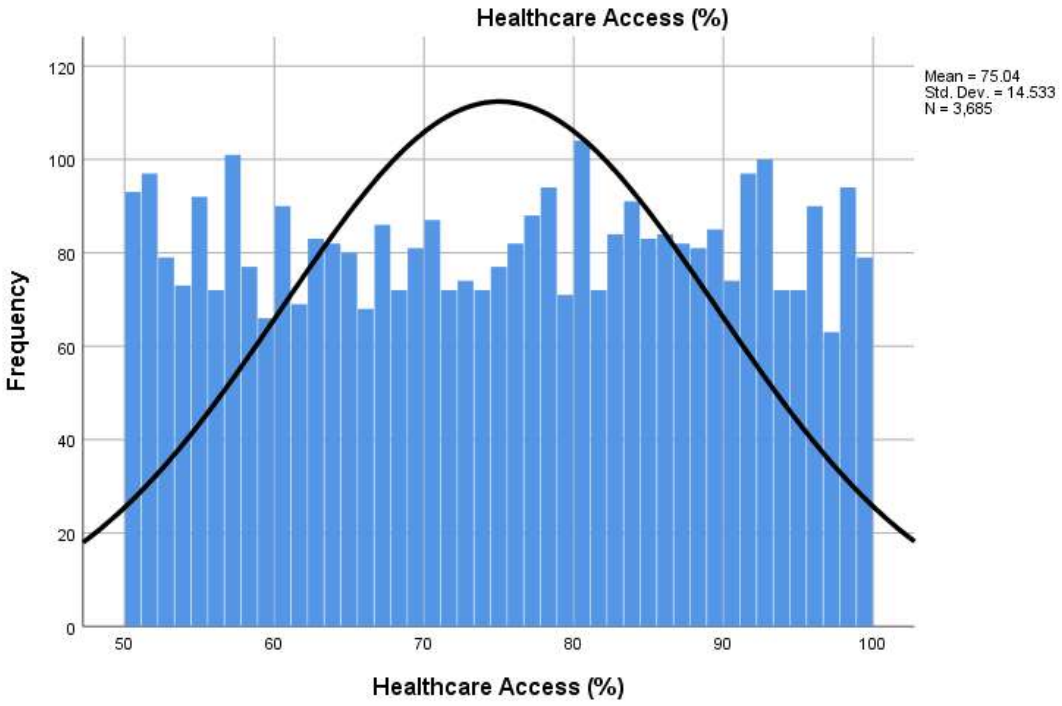
		Prevalence Rate (%)	Incidence Rate (%)	Mortality Rate (%)	Population Affected	Healthcare Access (%)	D
N	Valid	3685	3685	3685	3685	3685	
	Missing	0	0	0	0	0	

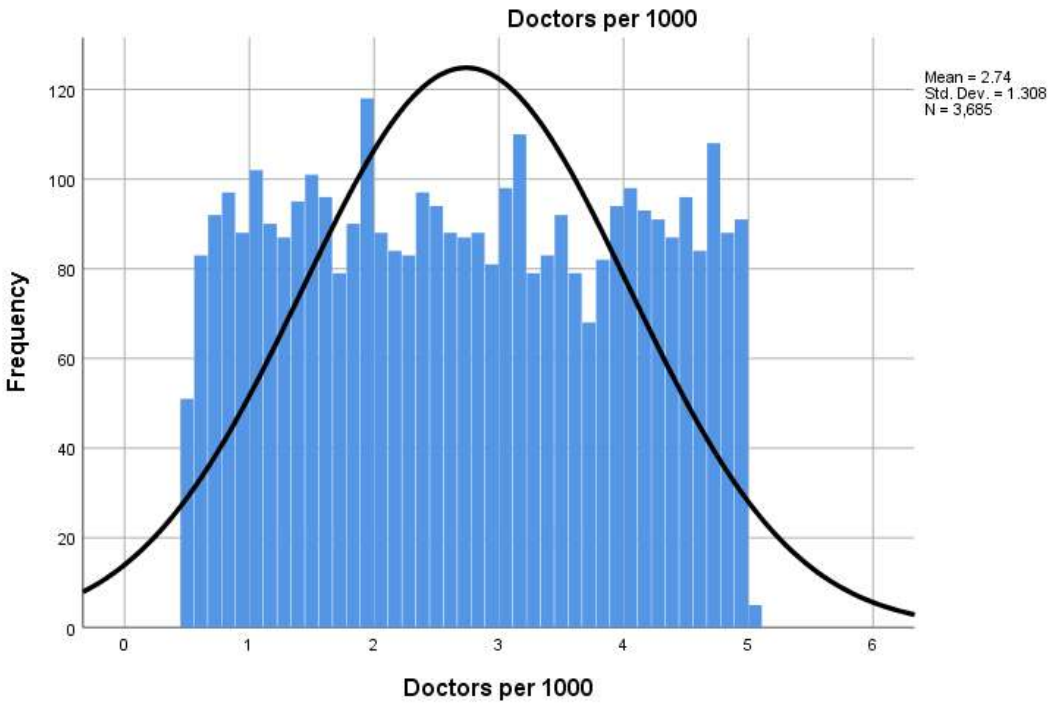


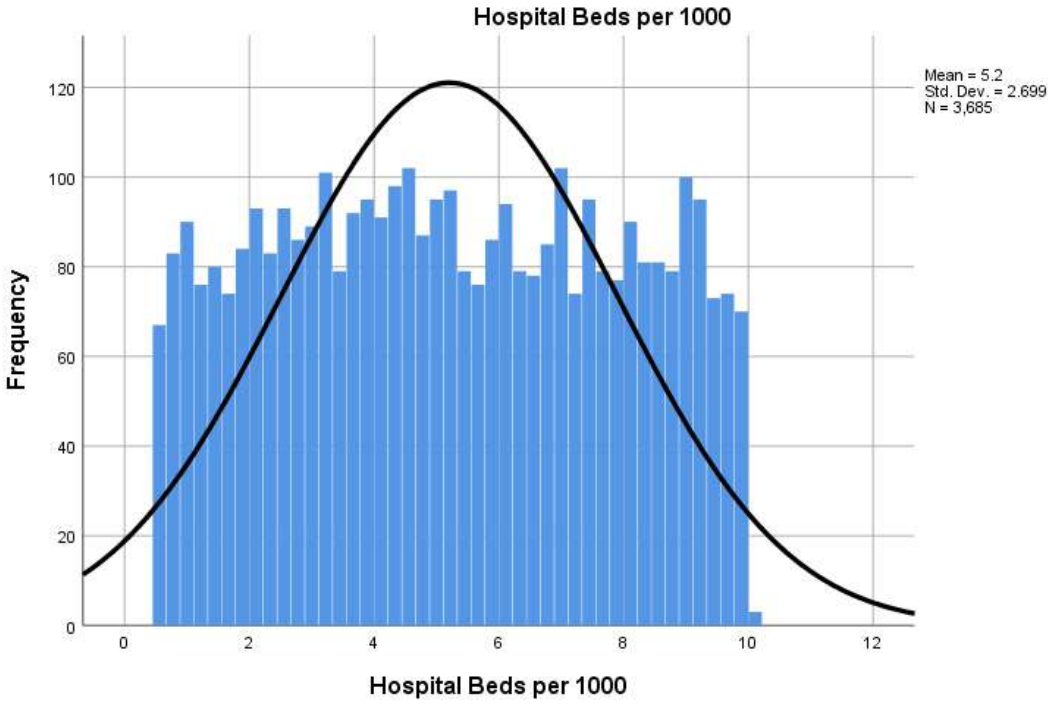


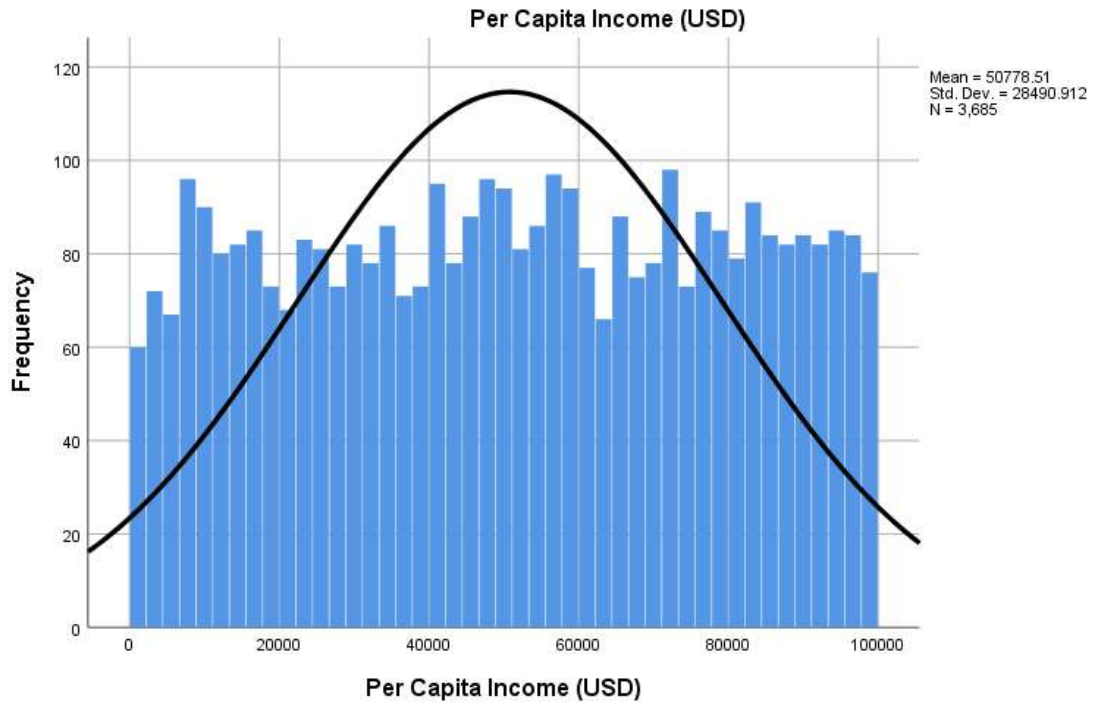


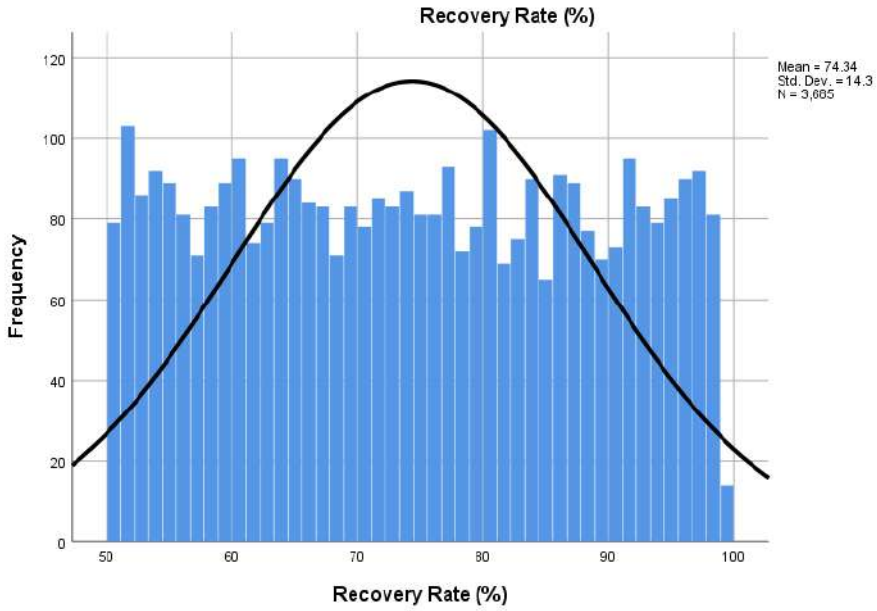


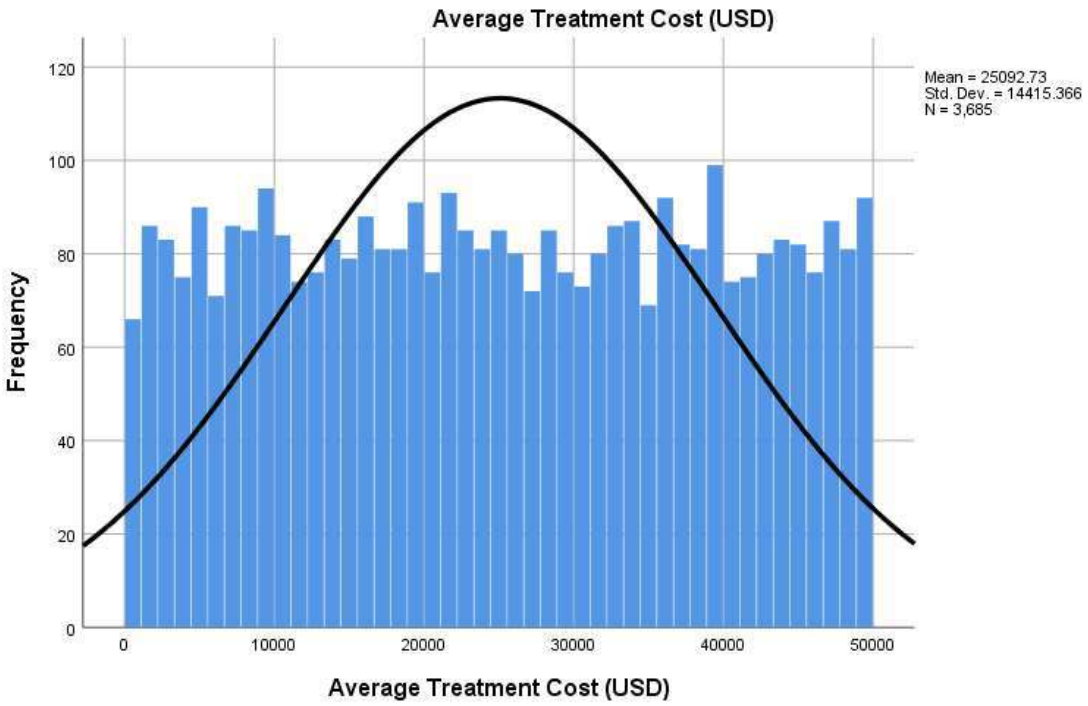


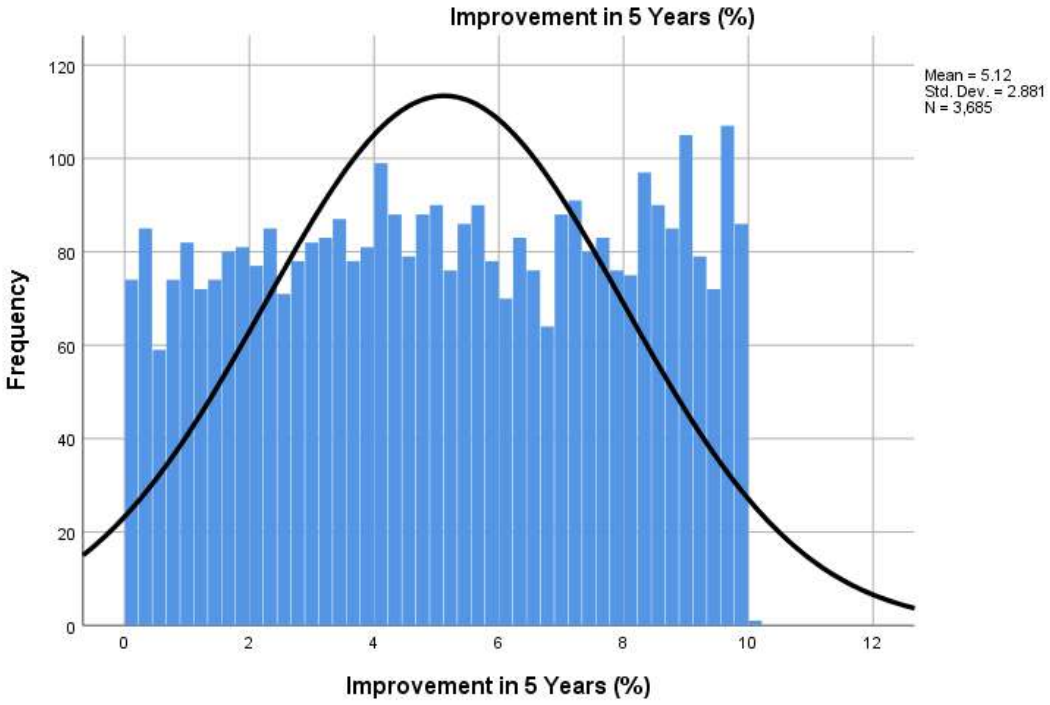


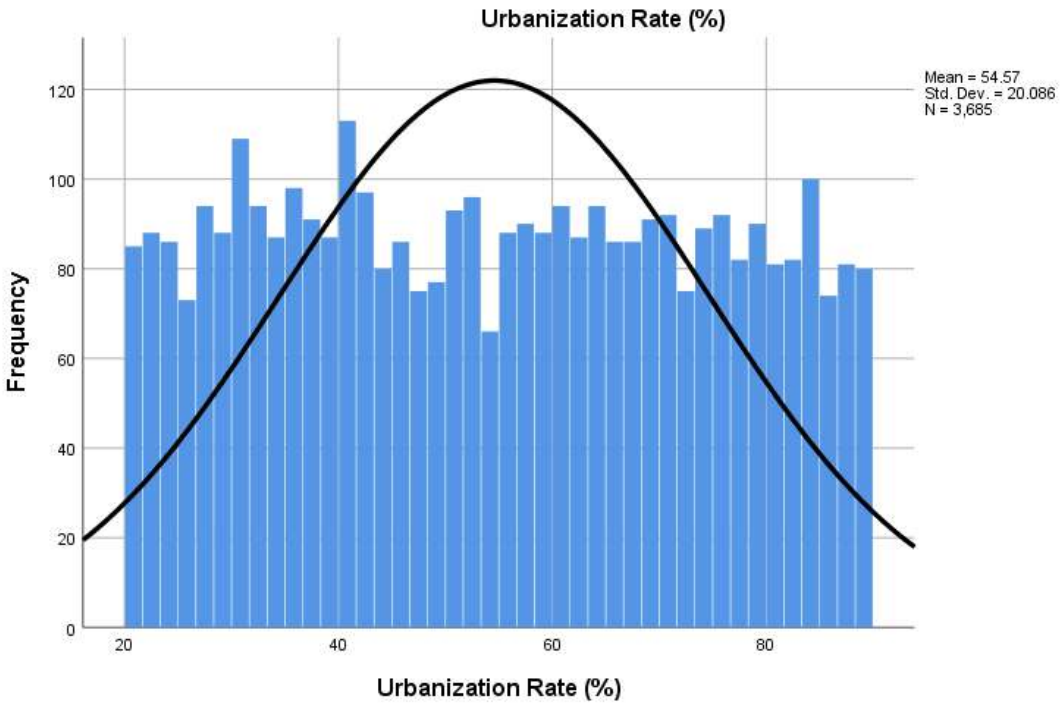


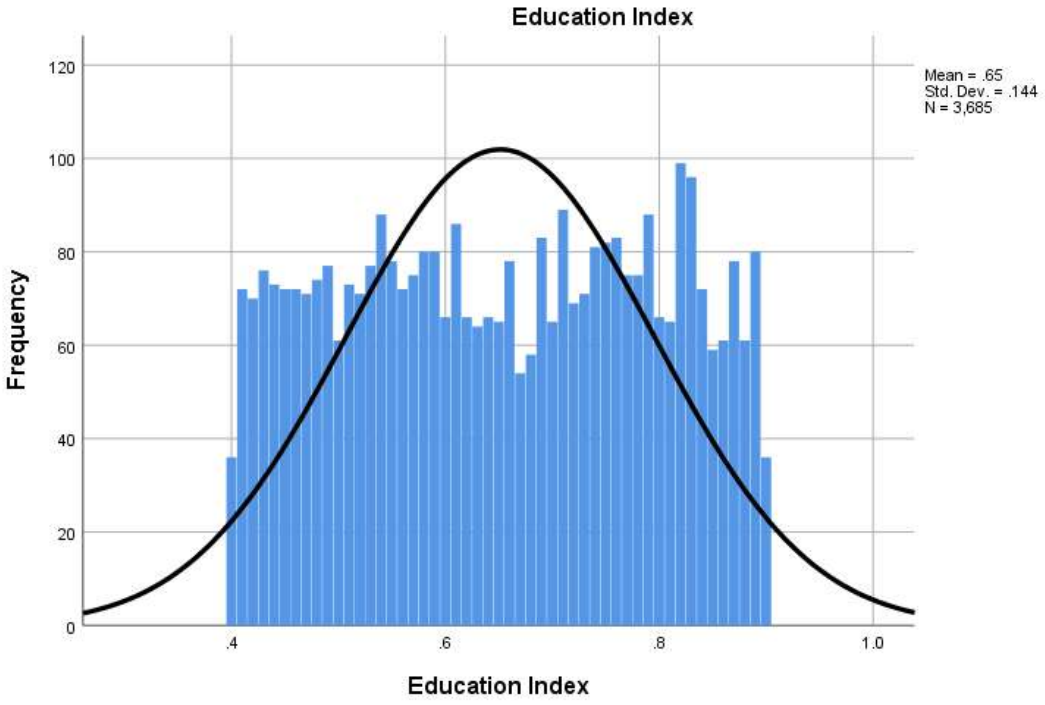












Hypothesis Testing

- Parametric Tests (Assume data is normal)

One Sample T-test

In one sample t test we use one specific data with a test value to compare it with mean of the data . In this test we test a claim that a company or person makes for the product or about anything . In our this dataset we don't have such a variable that makes any claim or which have specific value that is to be tested . So, we assume **amount** variable as claiming variable and use it in our One Sample T-test.

- **Commands**

  Analyze>Compare means >One sample t test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Prevalence Rate (%)	3685	10.07	5.699	.094

One-Sample Test

	Test Value = 0					
					90% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Prevalence Rate (%)	107.230	3684	.000	10.067	9.91	10.22

Interpretation:

The sample prevalence rate (%) is significantly greater than 0 ($p<.001p < .001p<.001$). The observed mean prevalence rate is 10.067%, and the confidence interval confirms this is a reliable estimate. This finding strongly supports the conclusion that the prevalence rate in the population is greater than 0.

Paired Sample T-Test

Our data don't meet the assumptions of the paired sample t-test i.e we don't have pairing variables.

Independent Sample T-test

- **Assumptions**

a. Independent sample t-test takes two variables in which one is categorical and the other one is numeric .

Data should follow normal distribution

- **Commands**

  Analyze > Compare Means >independent sample t test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Recovery Rate (%)	Fema	1190	75.04	14.498	.420
	Male	1284	74.45	13.875	.387

Independent Samples Test

		Levene's Test for Equality of Variances			
		F	Sig.	t	df
Recovery Rate (%)	Equal variances assumed	4.010	.045	1.037	2472
	Equal variances not assumed			1.036	2436.998

Interpretation:

The t-test shows no significant difference in recovery rates between groups (p=.300p = .300p=.300); the small mean difference (.592%) is within a 90% confidence interval including 0.

❖ **One Way ANOVA**

One way ANOVA is extension of one sample t-test . We are going to use two variables in one way ANOVA.

➤ **Commands**

Analyze>compare means > one Way ANOVA

ANOVA

Prevalence Rate (%)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	179.639	10	17.964	.552	.853
Within Groups	119477.416	3674	32.520		
Total	119657.055	3684			

Interpretation:

The one-way ANOVA does not find significant differences in prevalence rates (%) among the 11 groups (F=0.552,p=0.853F = 0.552, p = 0.853F=0.552,p=0.853). The observed variation in prevalence rates is likely due to random chance rather than actual differences between group means.

❖ **Two Way ANOVA**

In two way ANOVA we use 2 categorical and one continues variable . In our data set we are going to create a two way ANOVA for variables defaulted, purpose and amount .We don't have any continuous

variable so we compute a new variable

Commands

- :- Analyze>general linear model> univariate

post-hoc tests (e.g., LSD, Tukey's etc)

Multiple Comparisons

Dependent Variable: Prevalence Rate (%)

	(I) Disease Category	(J) Disease Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confic Lower Bound
Tukey HSD	Autoimmune	Bacterial	-.051	.428	1.000	-1.43
		Cardiovascu	-.209	.441	1.000	-1.63
		Chronic	-.370	.422	.999	-1.73
		Genetic	.206	.445	1.000	-1.23
		Infectious	-.140	.439	1.000	-1.55
		Metabolic	.024	.429	1.000	-1.36
		Neurologica	-.258	.419	1.000	-1.61
		Parasitic	.033	.432	1.000	-1.36
		Respiratory	-.108	.441	1.000	-1.53
		Viral	-.662	.435	.913	-2.06
	Bacterial	Autoimmune	.051	.428	1.000	-1.33
		Cardiovascu	-.158	.446	1.000	-1.60
		Chronic	-.319	.428	1.000	-1.70
		Genetic	.257	.451	1.000	-1.19
		Infectious	-.089	.444	1.000	-1.52
		Metabolic	.075	.435	1.000	-1.32
		Neurologica	-.207	.425	1.000	-1.57
		Parasitic	.084	.437	1.000	-1.32
		Respiratory	-.057	.447	1.000	-1.50
		Viral	-.611	.441	.952	-2.03
	Cardiovascu	Autoimmune	.209	.441	1.000	-1.21
		Bacterial	.158	.446	1.000	-1.28
		Chronic	-.161	.441	1.000	-1.58
		Genetic	.415	.463	.998	-1.08
		Infectious	.069	.457	1.000	-1.40
		Metabolic	.233	.447	1.000	-1.21
		Neurologica	-.048	.438	1.000	-1.46
		Parasitic	.242	.450	1.000	-1.21
		Respiratory	.101	.459	1.000	-1.38

	Viral	-.453	.453	.996	-1.91
Chronic	Autoimmune	.370	.422	.999	-.99
	Bacterial	.319	.428	1.000	-1.06
	Cardiovascu	.161	.441	1.000	-1.26
	Genetic	.577	.445	.970	-.86
	Infectious	.230	.438	1.000	-1.18
	Metabolic	.394	.429	.998	-.99
	Neurologica	.113	.419	1.000	-1.24
	Parasitic	.404	.432	.998	-.99
	Respiratory	.262	.441	1.000	-1.16
	Viral	-.291	.435	1.000	-1.69
Genetic	Autoimmune	-.206	.445	1.000	-1.64
	Bacterial	-.257	.451	1.000	-1.71
	Cardiovascu	-.415	.463	.998	-1.91
	Chronic	-.577	.445	.970	-2.01
	Infectious	-.346	.461	1.000	-1.83
	Metabolic	-.183	.452	1.000	-1.64
	Neurologica	-.464	.442	.994	-1.89
	Parasitic	-.173	.454	1.000	-1.64
	Respiratory	-.315	.463	1.000	-1.81
	Viral	-.868	.457	.719	-2.34
Infectious	Autoimmune	.140	.439	1.000	-1.27
	Bacterial	.089	.444	1.000	-1.34
	Cardiovascu	-.069	.457	1.000	-1.54
	Chronic	-.230	.438	1.000	-1.64
	Genetic	.346	.461	1.000	-1.14
	Metabolic	.164	.445	1.000	-1.27
	Neurologica	-.117	.436	1.000	-1.52
	Parasitic	.173	.448	1.000	-1.27
	Respiratory	.032	.457	1.000	-1.44
	Viral	-.522	.451	.987	-1.97
Metabolic	Autoimmune	-.024	.429	1.000	-1.40
	Bacterial	-.075	.435	1.000	-1.47
	Cardiovascu	-.233	.447	1.000	-1.67
	Chronic	-.394	.429	.998	-1.77
	Genetic	.183	.452	1.000	-1.27
	Infectious	-.164	.445	1.000	-1.60
	Neurologica	-.281	.426	1.000	-1.65

	Parasitic	.010	.438	1.000	-1.40
	Respiratory	-.132	.448	1.000	-1.57
	Viral	-.685	.441	.902	-2.11
Neurologica	Autoimmune	.258	.419	1.000	-1.09
	Bacterial	.207	.425	1.000	-1.16
	Cardiovascu	.048	.438	1.000	-1.36
	Chronic	-.113	.419	1.000	-1.46
	Genetic	.464	.442	.994	-.96
	Infectious	.117	.436	1.000	-1.29
	Metabolic	.281	.426	1.000	-1.09
	Parasitic	.291	.429	1.000	-1.09
	Respiratory	.149	.438	1.000	-1.26
	Viral	-.404	.432	.998	-1.80
Parasitic	Autoimmune	-.033	.432	1.000	-1.42
	Bacterial	-.084	.437	1.000	-1.49
	Cardiovascu	-.242	.450	1.000	-1.69
	Chronic	-.404	.432	.998	-1.79
	Genetic	.173	.454	1.000	-1.29
	Infectious	-.173	.448	1.000	-1.62
	Metabolic	-.010	.438	1.000	-1.42
	Neurologica	-.291	.429	1.000	-1.67
	Respiratory	-.142	.450	1.000	-1.59
	Viral	-.695	.444	.897	-2.13
Respiratory	Autoimmune	.108	.441	1.000	-1.31
	Bacterial	.057	.447	1.000	-1.38
	Cardiovascu	-.101	.459	1.000	-1.58
	Chronic	-.262	.441	1.000	-1.68
	Genetic	.315	.463	1.000	-1.18
	Infectious	-.032	.457	1.000	-1.50
	Metabolic	.132	.448	1.000	-1.31
	Neurologica	-.149	.438	1.000	-1.56
	Parasitic	.142	.450	1.000	-1.31
	Viral	-.553	.453	.980	-2.01
Viral	Autoimmune	.662	.435	.913	-.74
	Bacterial	.611	.441	.952	-.81
	Cardiovascu	.453	.453	.996	-1.01
	Chronic	.291	.435	1.000	-1.11
	Genetic	.868	.457	.719	-.61

		Infectious	.522	.451	.987	-.93
		Metabolic	.685	.441	.902	-.74
		Neurologica	.404	.432	.998	-.99
		Parasitic	.695	.444	.897	-.74
		Respiratory	.553	.453	.980	-.91
LSD	Autoimmune	Bacterial	-.051	.428	.905	-.89
		Cardiovascu	-.209	.441	.635	-1.07
		Chronic	-.370	.422	.380	-1.20
		Genetic	.206	.445	.643	-.67
		Infectious	-.140	.439	.749	-1.00
		Metabolic	.024	.429	.956	-.82
		Neurologica	-.258	.419	.539	-1.08
		Parasitic	.033	.432	.939	-.81
		Respiratory	-.108	.441	.806	-.97
		Viral	-.662	.435	.128	-1.51
	Bacterial	Autoimmune	.051	.428	.905	-.79
		Cardiovascu	-.158	.446	.723	-1.03
		Chronic	-.319	.428	.455	-1.16
		Genetic	.257	.451	.568	-.63
		Infectious	-.089	.444	.841	-.96
		Metabolic	.075	.435	.863	-.78
		Neurologica	-.207	.425	.627	-1.04
		Parasitic	.084	.437	.847	-.77
		Respiratory	-.057	.447	.898	-.93
		Viral	-.611	.441	.166	-1.47
	Cardiovascu	Autoimmune	.209	.441	.635	-.66
		Bacterial	.158	.446	.723	-.72
		Chronic	-.161	.441	.714	-1.02
		Genetic	.415	.463	.370	-.49
		Infectious	.069	.457	.880	-.83
		Metabolic	.233	.447	.603	-.64
		Neurologica	-.048	.438	.912	-.91
		Parasitic	.242	.450	.590	-.64
		Respiratory	.101	.459	.826	-.80
		Viral	-.453	.453	.318	-1.34
	Chronic	Autoimmune	.370	.422	.380	-.46
		Bacterial	.319	.428	.455	-.52
		Cardiovascu	.161	.441	.714	-.70

		Genetic	.577	.445	.195	-.30
		Infectious	.230	.438	.599	-.63
		Metabolic	.394	.429	.358	-.45
		Neurologica	.113	.419	.788	-.71
		Parasitic	.404	.432	.350	-.44
		Respiratory	.262	.441	.552	-.60
		Viral	-.291	.435	.503	-1.14
	Genetic	Autoimmune	-.206	.445	.643	-1.08
		Bacterial	-.257	.451	.568	-1.14
		Cardiovascu	-.415	.463	.370	-1.32
		Chronic	-.577	.445	.195	-1.45
		Infectious	-.346	.461	.452	-1.25
		Metabolic	-.183	.452	.686	-1.07
		Neurologica	-.464	.442	.294	-1.33
		Parasitic	-.173	.454	.703	-1.06
		Respiratory	-.315	.463	.497	-1.22
		Viral	-.868	.457	.058	-1.76
	Infectious	Autoimmune	.140	.439	.749	-.72
		Bacterial	.089	.444	.841	-.78
		Cardiovascu	-.069	.457	.880	-.96
		Chronic	-.230	.438	.599	-1.09
		Genetic	.346	.461	.452	-.56
		Metabolic	.164	.445	.713	-.71
		Neurologica	-.117	.436	.787	-.97
		Parasitic	.173	.448	.699	-.70
		Respiratory	.032	.457	.945	-.86
		Viral	-.522	.451	.247	-1.41
	Metabolic	Autoimmune	-.024	.429	.956	-.86
		Bacterial	-.075	.435	.863	-.93
		Cardiovascu	-.233	.447	.603	-1.11
		Chronic	-.394	.429	.358	-1.23
		Genetic	.183	.452	.686	-.70
		Infectious	-.164	.445	.713	-1.04
		Neurologica	-.281	.426	.509	-1.12
		Parasitic	.010	.438	.983	-.85
		Respiratory	-.132	.448	.768	-1.01
		Viral	-.685	.441	.121	-1.55
	Neurologica	Autoimmune	.258	.419	.539	-.56

		Bacterial	.207	.425	.627	-.63
		Cardiovascu	.048	.438	.912	-.81
		Chronic	-.113	.419	.788	-.93
		Genetic	.464	.442	.294	-.40
		Infectious	.117	.436	.787	-.74
		Metabolic	.281	.426	.509	-.55
		Parasitic	.291	.429	.498	-.55
		Respiratory	.149	.438	.733	-.71
		Viral	-.404	.432	.350	-1.25
	Parasitic	Autoimmune	-.033	.432	.939	-.88
		Bacterial	-.084	.437	.847	-.94
		Cardiovascu	-.242	.450	.590	-1.12
		Chronic	-.404	.432	.350	-1.25
		Genetic	.173	.454	.703	-.72
		Infectious	-.173	.448	.699	-1.05
		Metabolic	-.010	.438	.983	-.87
		Neurologica	-.291	.429	.498	-1.13
		Respiratory	-.142	.450	.753	-1.02
		Viral	-.695	.444	.118	-1.57
	Respiratory	Autoimmune	.108	.441	.806	-.76
		Bacterial	.057	.447	.898	-.82
		Cardiovascu	-.101	.459	.826	-1.00
		Chronic	-.262	.441	.552	-1.13
		Genetic	.315	.463	.497	-.59
		Infectious	-.032	.457	.945	-.93
		Metabolic	.132	.448	.768	-.75
		Neurologica	-.149	.438	.733	-1.01
		Parasitic	.142	.450	.753	-.74
		Viral	-.553	.453	.222	-1.44
	Viral	Autoimmune	.662	.435	.128	-.19
		Bacterial	.611	.441	.166	-.25
		Cardiovascu	.453	.453	.318	-.44
		Chronic	.291	.435	.503	-.56
		Genetic	.868	.457	.058	-.03
		Infectious	.522	.451	.247	-.36
		Metabolic	.685	.441	.121	-.18
		Neurologica	.404	.432	.350	-.44
		Parasitic	.695	.444	.118	-.18

	Respiratory	.553	.453	.222	-.34

Homogeneous Subsets:

Prevalence Rate (%)

	Disease Category	N	Subset for alpha = 0.05 1
Tukey HSD[a,b]	Genetic	298	9.72
	Parasitic	334	9.89
	Metabolic	343	9.90
	Autoimmune	365	9.92
	Bacterial	346	9.98
	Respiratory	308	10.03
	Infectious	315	10.06
	Cardiovascu	309	10.13
	Neurologica	376	10.18
	Chronic	366	10.29
	Viral	325	10.59
	Sig.		.674

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 333.117.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Non parametric test:

Sign test:

Univariate Analysis of Variance

Between-Subjects Factors

		Value Label	N
Gender	1	Fema	1190
	2	Male	1284
	3	Othe	1211

Tests of Between-Subjects Effects

Dependent Variable: Average Treatment Cost (USD)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.

Corrected Model	139965566.421[a]	2	69982783.211	.337	.714
Intercept	2318957166897.716	1	2318957166897.716	11155.395	.000
Gender_1	139965566.421	2	69982783.211	.337	.714
Error	765405431241.422	3682	207877629.343		
Total	3085788286168.000	3685			
Corrected Total	765545396807.843	3684			

a. R Squared = .000 (Adjusted R Squared = .000)

Mann-Whitney Test

Ranks

	Gender	N	Mean Rank	Sum of Ranks
Average Treatment Cost (USD)	Fema	1190	1248.67	1485913.50
	Male	1284	1227.15	1575661.50
	Total	2474		

Test Statistics[a]

	Average Treatment Cost (USD)
Mann-Whitney U	750691.500
Wilcoxon W	1575661.500
Z	-.749
Asymp. Sig. (2-tailed)	.454

a. Grouping Variable: Gender

Interpretation:

The Mann-Whitney U test shows no significant difference in average treatment costs between genders (p=0.454p = 0.454p=0.454), indicating similar costs for males and females.

Kruskal-Wallis Test

Ranks

	Treatment Type	N	Mean Rank
Average Treatment Cost	Medication	899	1877.27

(USD)	Surgery	910	1839.79
	Therapy	946	1796.29
	Vaccinatio	930	1860.52
	Total	3685	

Test Statistics[a,b]

	Average Treatment Cost (USD)
Kruskal-Wallis H	3.017
df	3
Asymp. Sig.	.389

a. Kruskal Wallis Test

b. Grouping Variable: Treatment Type

Interpretation:

The Kruskal-Wallis H test shows no significant difference in **Average Treatment Cost (USD)** among the different **Treatment Types** (p=0.389p = 0.389p=0.389), suggesting that treatment costs are similar across the treatment types in this dataset.

Measure of Association

- **Chi-Square Test**

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Treatment Type	3685	100.0%	0	0.0%	3685	100.0%

Gender * Treatment Type Crosstabulation

Count

		Treatment Type				
		Medication	Surgery	Therapy	Vaccinatio	Total
Gender	Fema	271	302	314	303	1190
	Male	315	296	331	342	1284
	Othe	313	312	301	285	1211

Total	899	910	946	930	3685

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	7.493[a]	6	.278
Likelihood Ratio	7.540	6	.274
Linear-by-Linear Association	3.478	1	.062
N of Valid Cases	3685		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 290.31.

Interpretation:

The Chi-Square test shows no significant association between **Gender** and **Treatment Type** (p=0.278p = 0.278p=0.278), with a weak trend observed in the linear association (p=0.062p = 0.062p=0.062), but it is not statistically significant.

Additional Measures:

Kendall's Tau-b

Correlations

			Healthcare Access (%)	Recovery Rate (%)
Kendall's tau_b	Healthcare Access (%)	Correlation Coefficient	1.000	.008
		Sig. (2-tailed)	.	.469
		N	3685	3685
	Recovery Rate (%)	Correlation Coefficient	.008	1.000
		Sig. (2-tailed)	.469	.
		N	3685	3685

Interpretation:

The correlation between **Healthcare Access (%)** and **Recovery Rate (%)** is very weak (τb=0.008\tau_b = 0.008τb=0.008) and not statistically significant (p=0.469p = 0.469p=0.469), indicating no meaningful relationship between the two variables.

Gamma:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Treatment Type * Gender	3685	100.0%	0	0.0%	3685	100.0%

Treatment Type * Gender Crosstabulation

Count

		Gender			
		Fema	Male	Othe	Total
Treatment Type	Medication	271	315	313	899
	Surgery	302	296	312	910
	Therapy	314	331	301	946
	Vaccinatio	303	342	285	930
Total		1190	1284	1211	3685

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Ordinal by Ordinal	Gamma	-.037	.020	-1.875	.061
N of Valid Cases		3685			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Interpretation

The Gamma correlation between the ordinal variables is −0.037-0.037−0.037, with a p-value of 0.0610.0610.061, indicating a very weak and marginally non-significant negative association.

Somer's D:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Treatment Type * Gender	3685	100.0%	0	0.0%	3685	100.0%

Treatment Type * Gender Crosstabulation

Count

		Gender			
		Fema	Male	Othe	Total
Treatment Type	Medication	271	315	313	899
	Surgery	302	296	312	910
	Therapy	314	331	301	946
	Vaccinatio	303	342	285	930
Total		1190	1284	1211	3685

Interpretation:

The crosstabulation shows a relatively even distribution of treatment types across genders, with males slightly preferring vaccination and therapy, while females and others are more evenly distributed across all treatment types.

Tau-c

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Treatment Type * Gender	3685	100.0%	0	0.0%	3685	100.0%

Treatment Type * Gender Crosstabulation

Count

		Gender			
		Fema	Male	Othe	Total
Treatment Type	Medication	271	315	313	899
	Surgery	302	296	312	910
	Therapy	314	331	301	946
	Vaccinatio	303	342	285	930
Total		1190	1284	1211	3685

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Ordinal by Ordinal	Kendall's tau-c	-.028	.015	-1.875	.061
N of Valid Cases		3685			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Interpretation:

Kendall's tau-c is −0.028-0.028−0.028 with a p-value of 0.0610.0610.061, indicating a very weak and marginally non-significant negative association between the ordinal variables

Yate's corrected Chisquare:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Treatment Type * Gender	3685	100.0%	0	0.0%	3685	100.0%

Treatment Type * Gender Crosstabulation

Count

		Gender			
		Fema	Male	Othe	Total
Treatment Type	Medication	271	315	313	899
	Surgery	302	296	312	910
	Therapy	314	331	301	946
	Vaccinatio	303	342	285	930
Total		1190	1284	1211	3685

Directional Measures

			Value	Asymptotic Standard Error[a]	Approx
Nominal by Nominal	Lambda	Symmetric	.008	.008	
		Treatment Type Dependent	.008	.013	
		Gender Dependent	.007	.010	
	Goodman and Kruskal tau	Treatment Type Dependent	.001	.000	
		Gender Dependent	.001	.001	
	Uncertainty Coefficient	Symmetric	.001	.001	
		Treatment Type Dependent	.001	.001	
		Gender Dependent	.001	.001	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on chi-square approximation

d. Likelihood ratio chi-square probability.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.045	.278
	Cramer's V	.032	.278
	Contingency Coefficient	.045	.278
N of Valid Cases		3685	

Interpretation:

The Phi, Cramer's V, and Contingency Coefficient values are all very small (0.0450.0450.045 and 0.0320.0320.032) with a p-value of 0.2780.2780.278, indicating no significant association between the nominal variables.

- Bowker's Test & McNemar Test

We can't use these two rest in our dataset because data don't meet assumptions of data i.e Data don't consist of any paired variables

Risk Measurement: Odds Ratio Risk Ratios

Risk Estimate

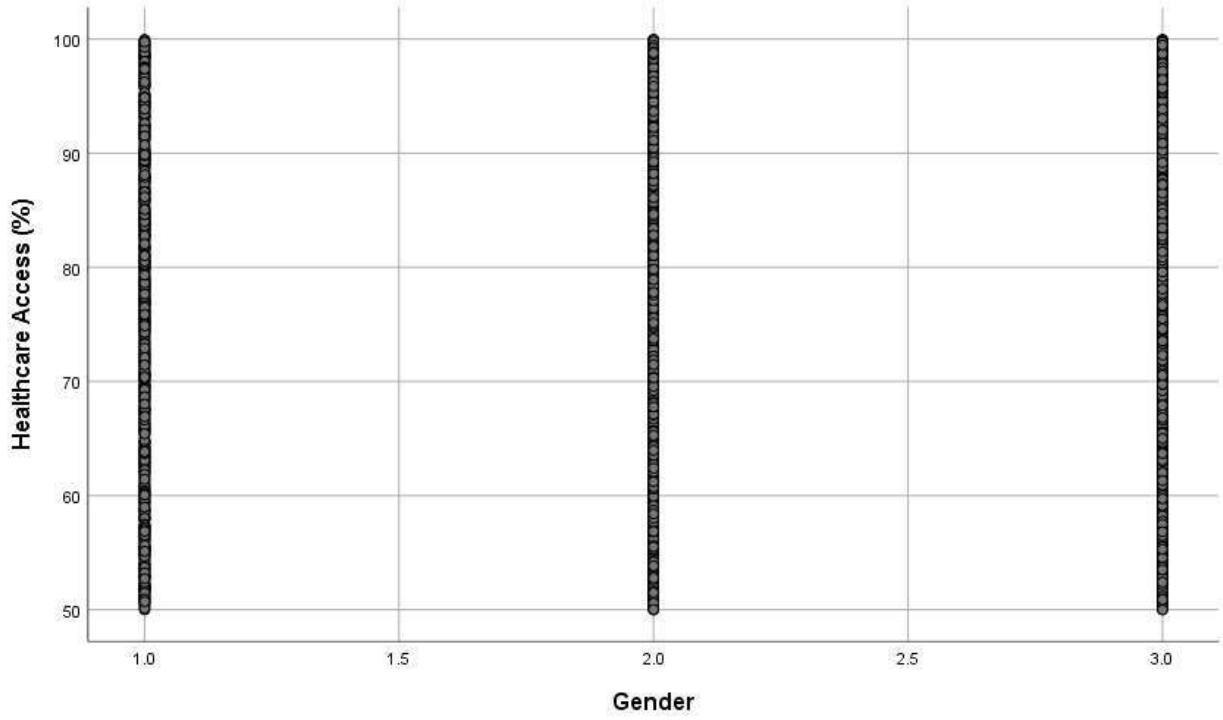
	Value
Odds Ratio for Treatment Type (Medication / Surgery)	[a]

a. Risk Estimate statistics cannot be computed. They are only computed for a 2*2 table without empty cells.

Interpretation:

The odds ratio for **Medication** vs. **Surgery** cannot be computed due to the need for a 2x2 table without empty cells.

Regression Analysis
Linearity



Linear Regression:

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Gender[b]	.	Enter

a. Dependent Variable: Healthcare Access (%)

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.008[a]	.000	.000	14.534

a. Predictors: (Constant), Gender

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	52.386	1	52.386	.248	.619[b]
	Residual	778014.599	3683	211.245		

	Total	778066.985	3684			

a. Dependent Variable: Healthcare Access (%)

b. Predictors: (Constant), Gender

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	74.741	.641		116.544	.000
	Gender	.148	.297	.008	.498	.619

a. Dependent Variable: Healthcare Access (%)

Interpretation:

The model with **Gender** as the predictor explains none of the variance in the dependent variable, as indicated by an **R^2 of 0.000** and a very weak correlation (**R = 0.008**). The model does not fit the data well.

4o mini

Logistic Regression:

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	3685	100.0
	Missing Cases	0	.0
	Total	3685	100.0
Unselected Cases		0	.0
Total		3685	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
No	0
Ye	1

Block 0: Beginning Block

Classification Table[a,b]

	Observed		Predicted: Availability of Vaccines/Treatment		Percentage Correct
			No	Ye	
Step 0	Availability of Vaccines/Treatment	No	1878	0	100.0
		Ye	1807	0	.0
	Overall Percentage				51.0

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.039	.033	1.368	1	.242	.962

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Healthcare Access (%)	.940	1	.332
	Overall Statistics		.940	1	.332

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	.940	1	.332
	Block	.940	1	.332
	Model	.940	1	.332

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5106.187[a]	.000	.000

a. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted: Availability of Vaccines/Treatment		Percentage Correct
			No	Ye	
Step 1	Availability of Vaccines/Treatment	No	1592	286	84.8
		Ye	1511	296	16.4
	Overall Percentage				51.2

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Healthcare Access (%)	-.002	.002	.940	1	.332	.998
	Constant	.126	.173	.532	1	.466	1.135

a. Variable(s) entered on step 1: Healthcare Access (%).

Interpretation:

The model shows a **Cox & Snell R^2** and **Nagelkerke R^2** of **0.000**, indicating that the predictors explain none of the variance in the dependent variable. The **-2 Log Likelihood** value of **5106.187** suggests a poor model fit.

Exploratory Factor Analysis (EFA)

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.491
Bartlett's Test of Sphericity	Approx. Chi-Square	4.085
	df	3
	Sig.	.252

Anti-image Matrices

		Healthcare Access (%)	Doctors per 1000	Hospital Beds per 1000
Anti-image Covariance	Healthcare Access (%)	.999	.021	-.020
	Doctors per 1000	.021	.999	-.017
	Hospital Beds per 1000	-.020	-.017	.999
Anti-image Correlation	Healthcare Access (%)	.492[a]	.021	-.020

	Doctors per 1000	.021	.490[a]	-.017
	Hospital Beds per 1000	-.020	-.017	.490[a]

a. Measures of Sampling Adequacy(MSA)

Communalities

	Initial
Healthcare Access (%)	1.000
Doctors per 1000	1.000
Hospital Beds per 1000	1.000

Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	1.022	34.054	34.054
2	1.016	33.882	67.937
3	.962	32.063	100.000

Extraction Method: Principal Component Analysis.

Component Matrix[a]

a. 2 components extracted.

Interpretation:

The Principal Component Analysis (PCA) shows three components explaining the total variance. The first two components explain **34.05%** and **33.88%** of the variance, respectively, with the third component explaining **32.06%**, cumulatively accounting for **100%** of the variance.

In Minitab

Outliers:

Outlier Test: 1, 2, 8, 471007, 58, 3, 8_1, 21064, 92, 4493, 2_1, 16886, 1_1, 86

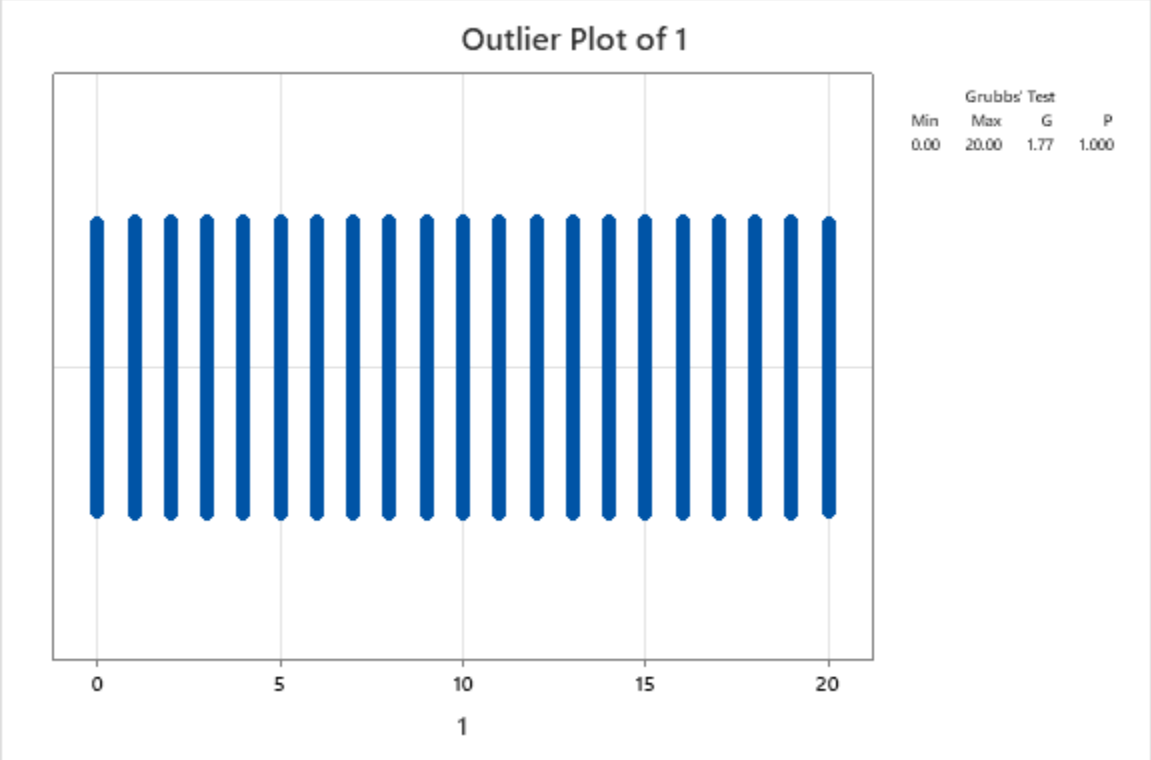
Method

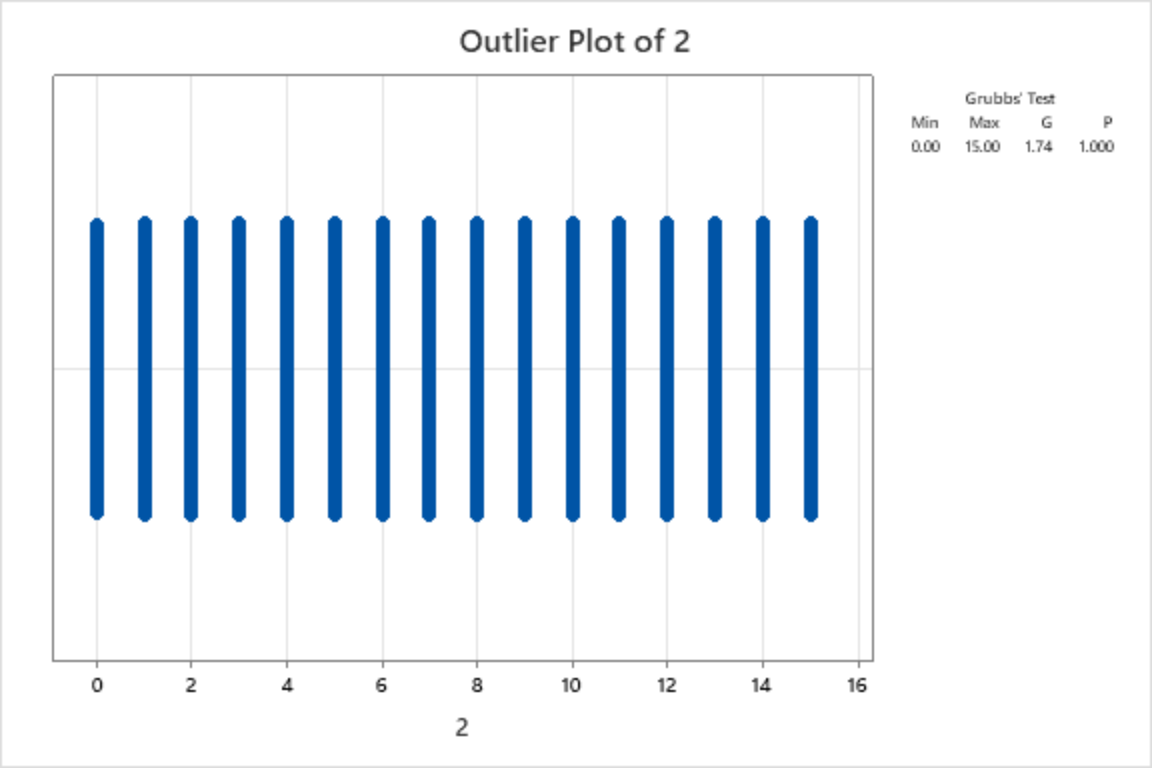
Null hypothesis	All data values come from the same normal population
Alternative hypothesis	Smallest or largest data value is an outlier
Significance level	α = 0.05

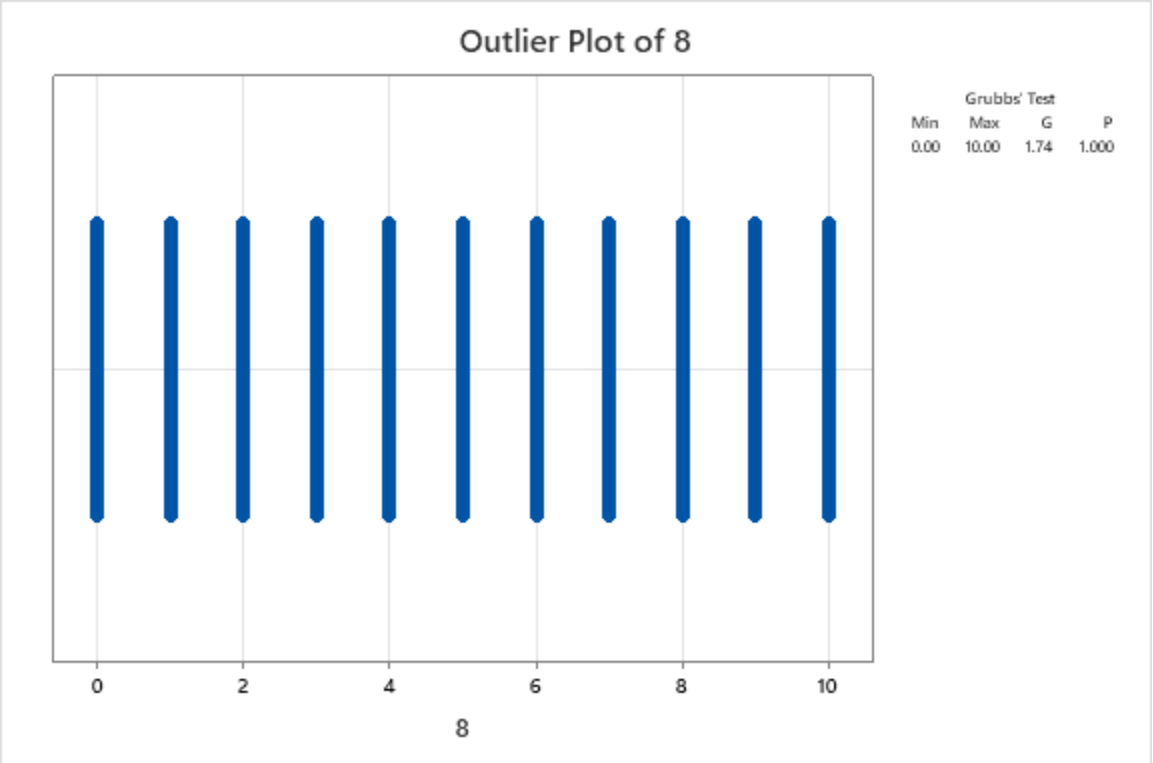
Grubbs' Test

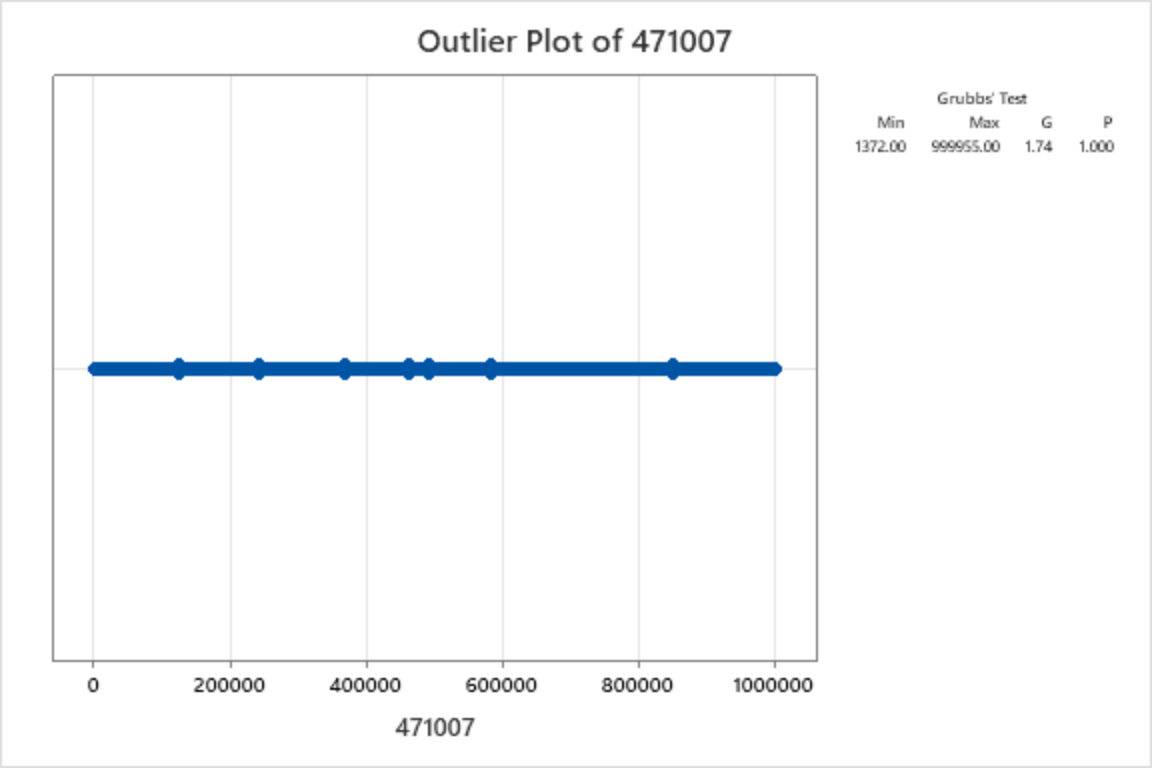
Variable	N	Mean	StDev	Min	Max	G	P
1	3684	10.078	5.709	0.0000	20.000	1.77	1.000
2	3684	7.5057	4.3165	0.0000	15.0000	1.74	1.000
8	3684	5.0109	2.8747	0.0000	10.0000	1.74	1.000
471007	3684	498840	288033	1372	999955	1.74	1.000
58	3684	75.050	14.540	50.000	100.000	1.72	1.000
3	3684	2.7693	1.3223	1.0000	5.0000	1.69	1.000
8_1	3684	5.2153	2.7068	1.0000	10.0000	1.77	1.000
21064	3684	25094	14417	117	49994	1.73	1.000
92	3684	74.343	14.304	50.000	99.000	1.72	1.000
4493	3684	2485.1	1435.6	1.0	5000.0	1.75	1.000
2_1	3684	5.1306	2.9145	0.0000	10.0000	1.76	1.000
16886	3684	50788	28489	503	99996	1.77	1.000
1_1	3684	0.81189	0.39085	0.00000	1.00000	2.08	1.000
86	3684	54.564	20.077	20.000	90.000	1.77	1.000

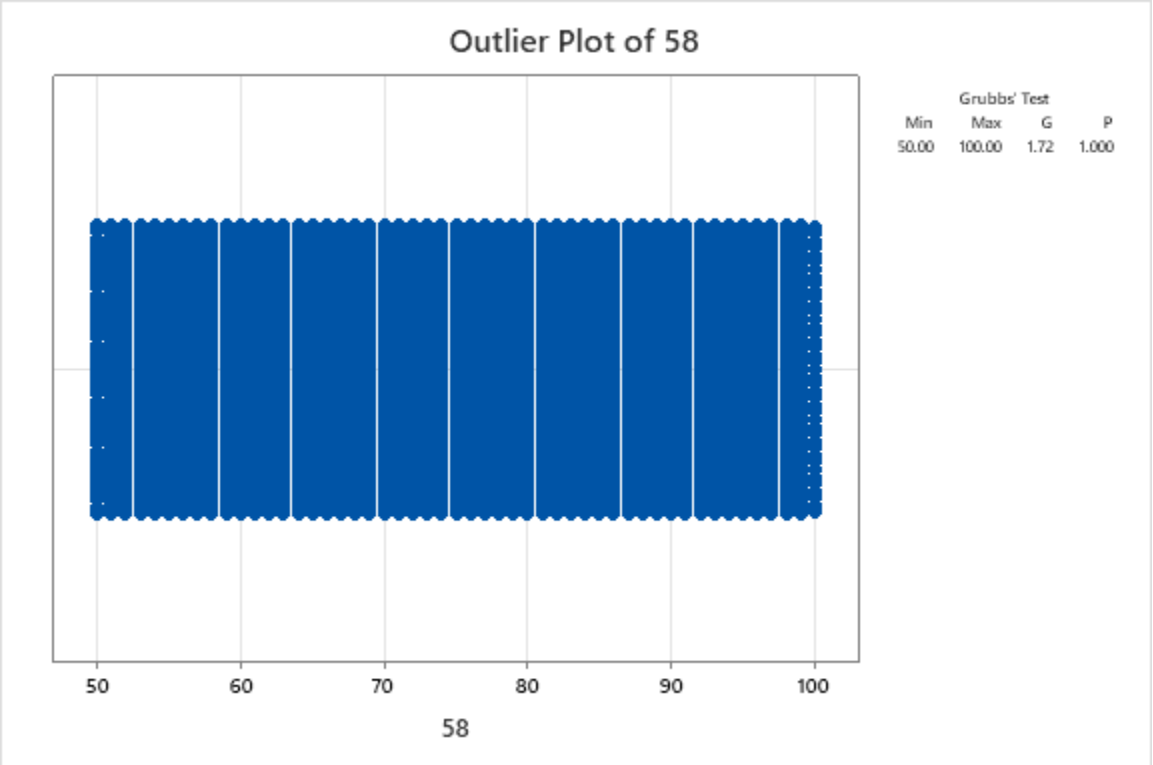
* NOTE * No outlier at the 5% level of significance

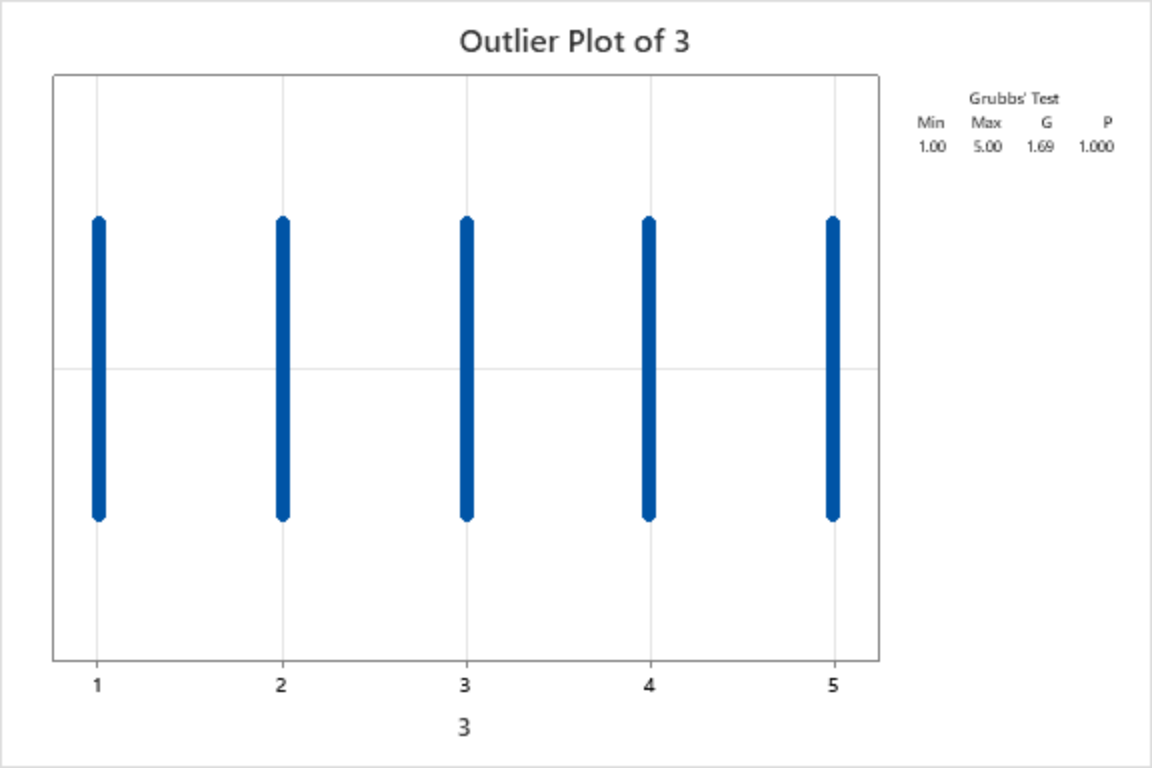


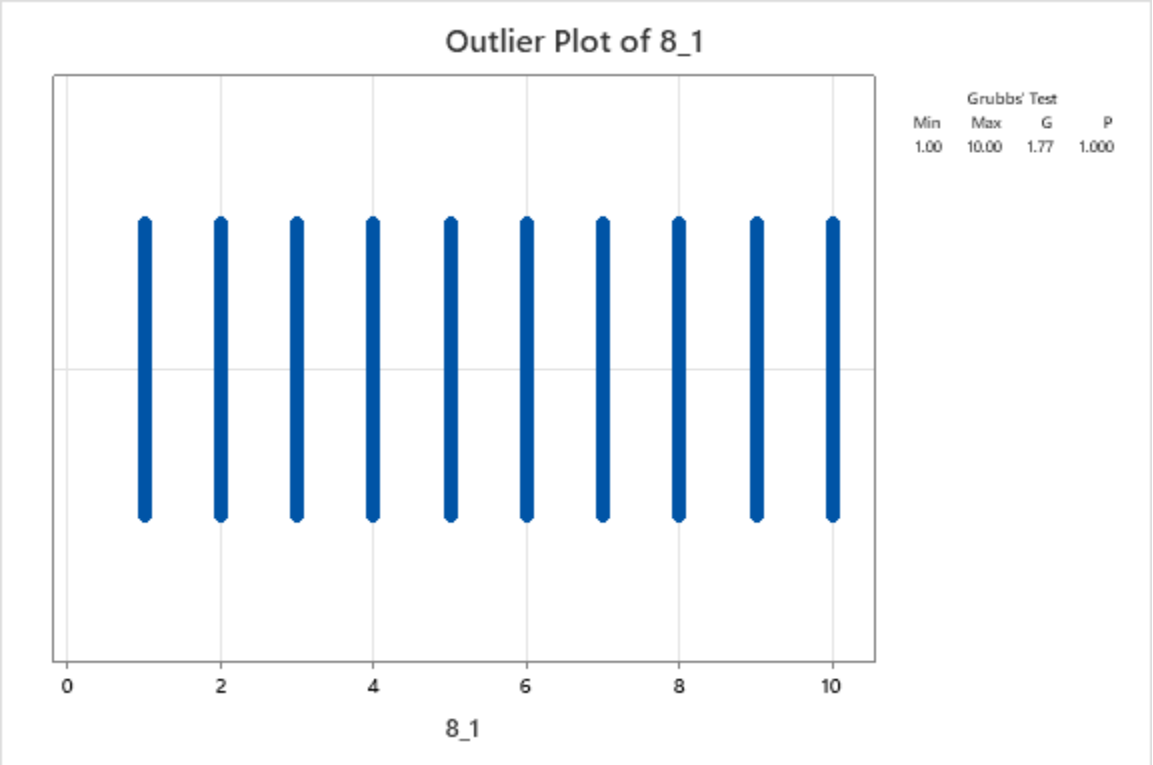


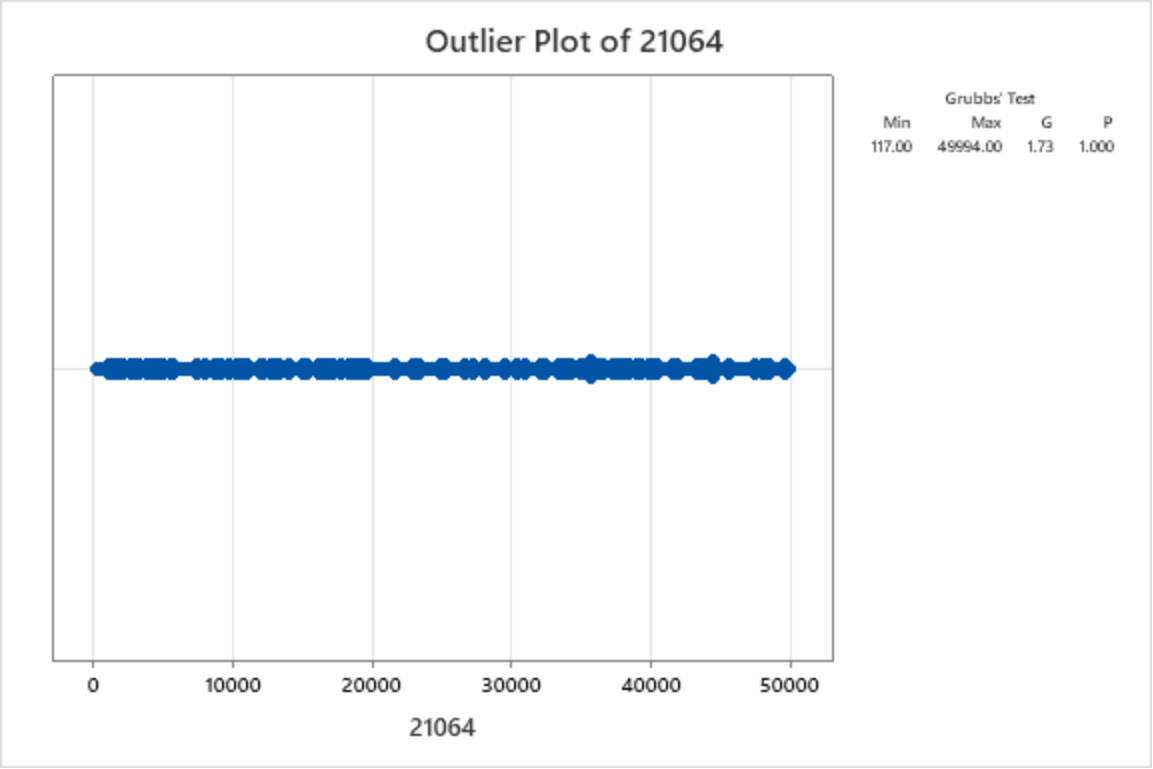


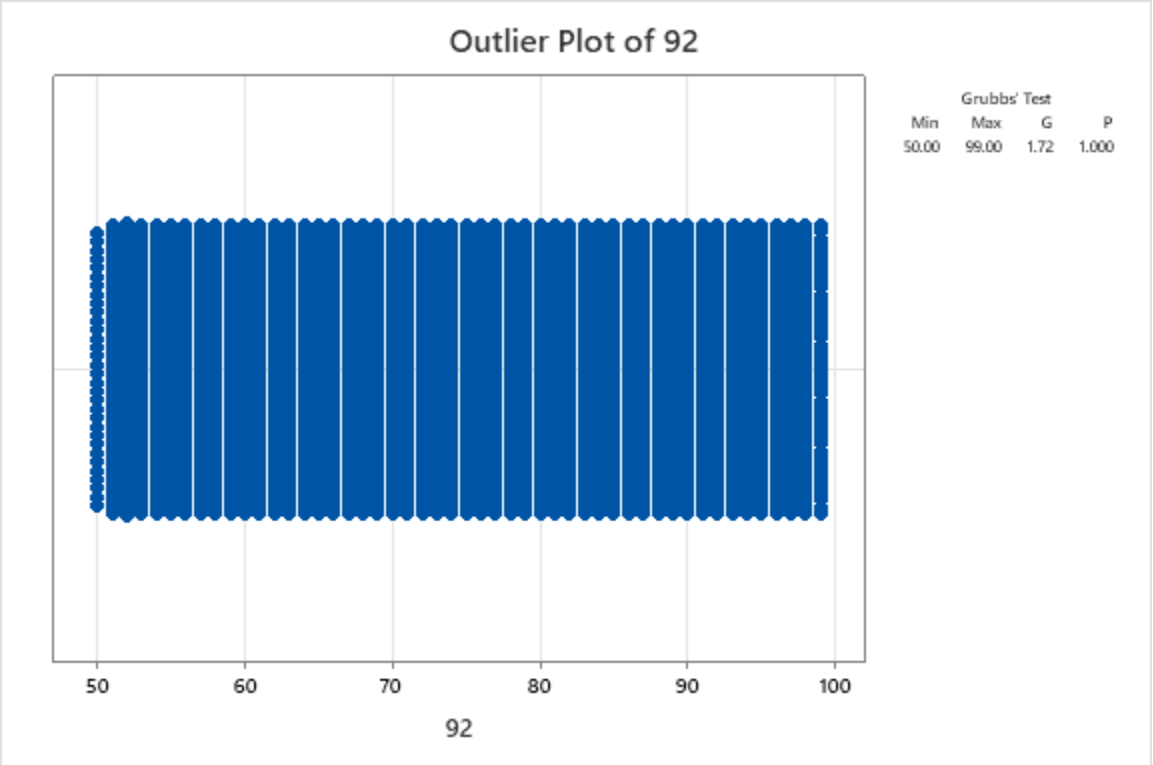


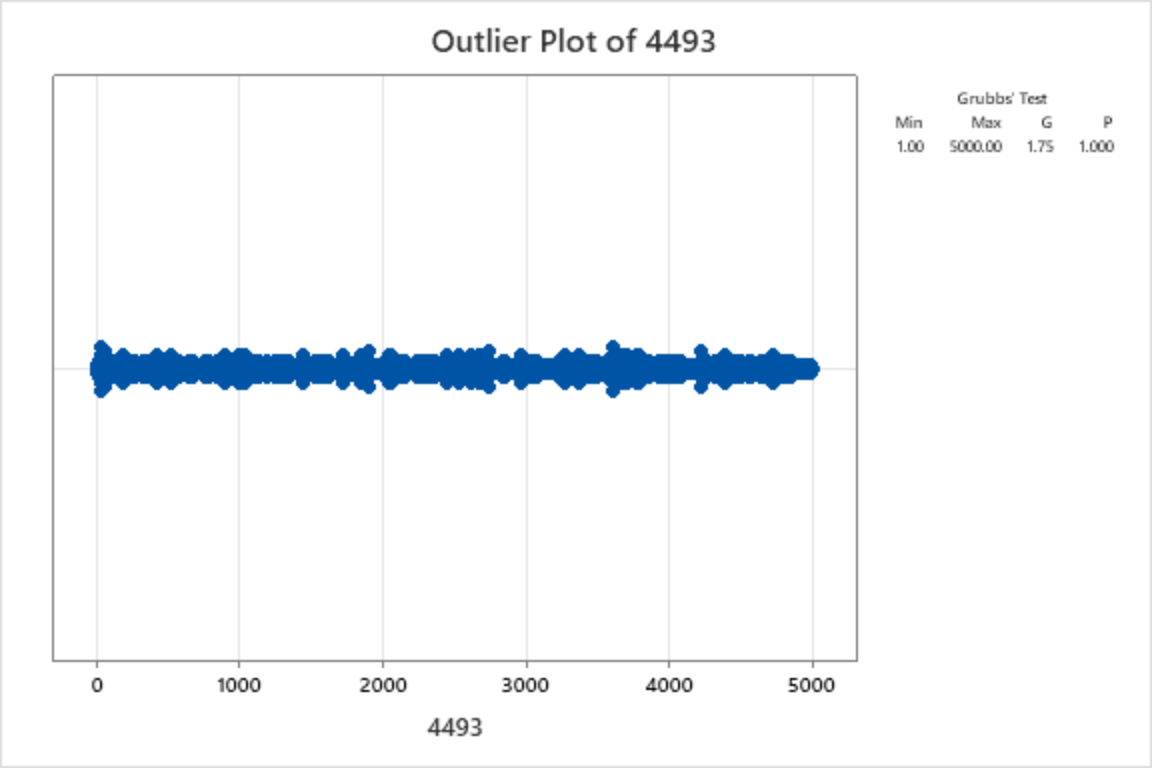


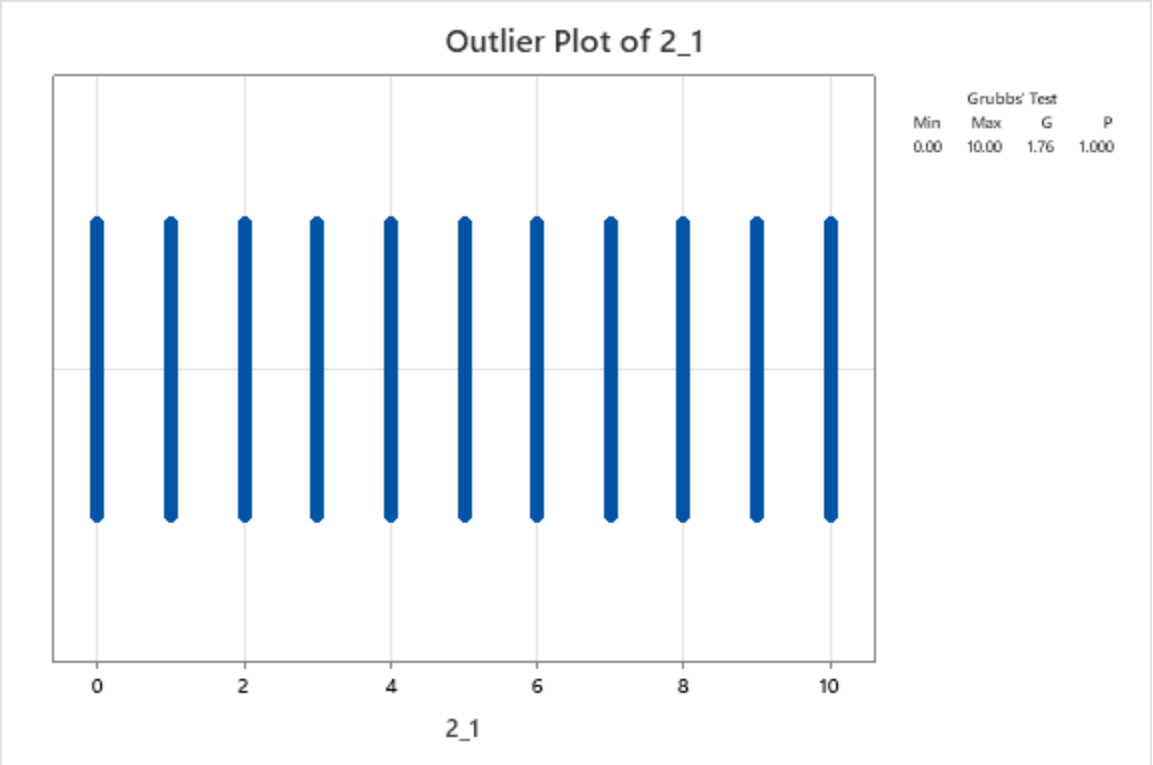


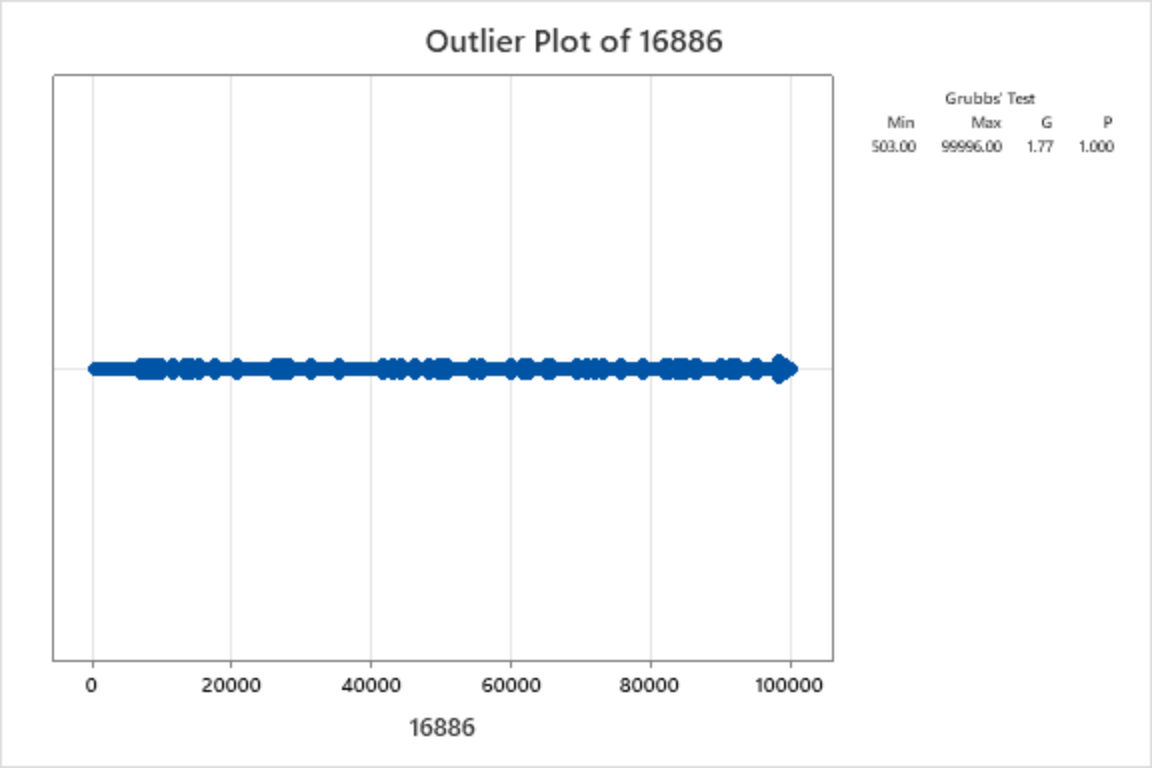


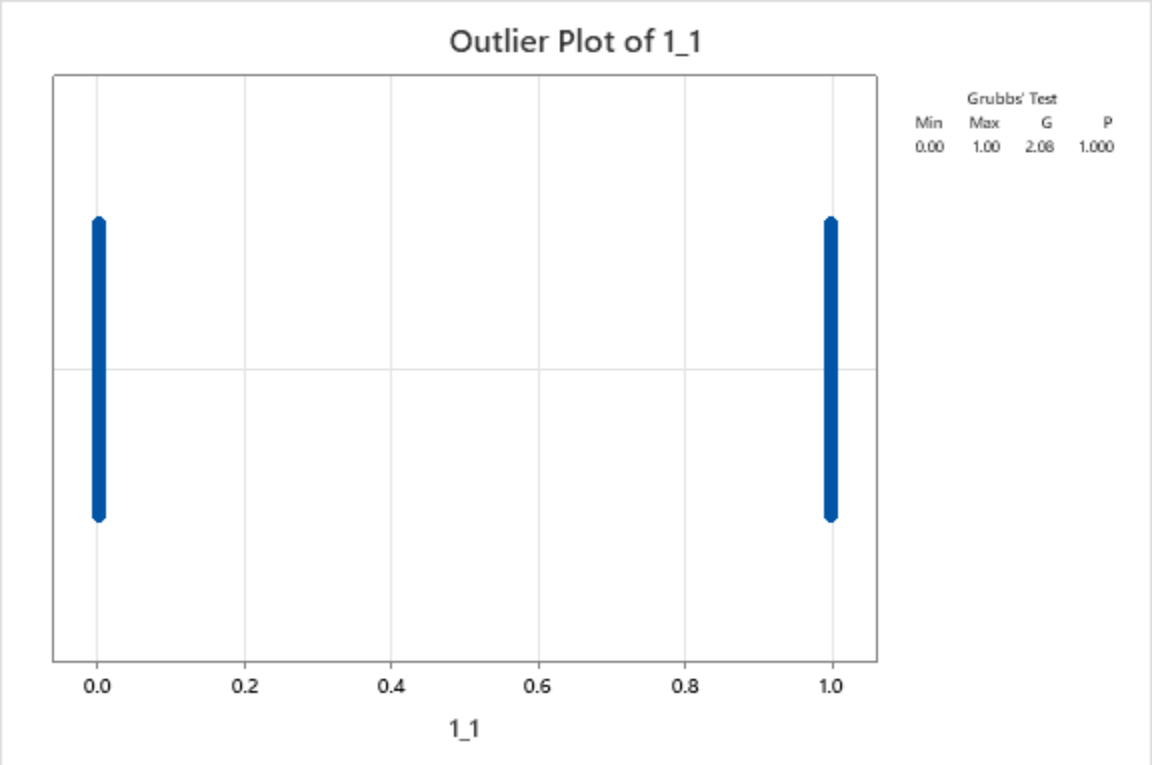


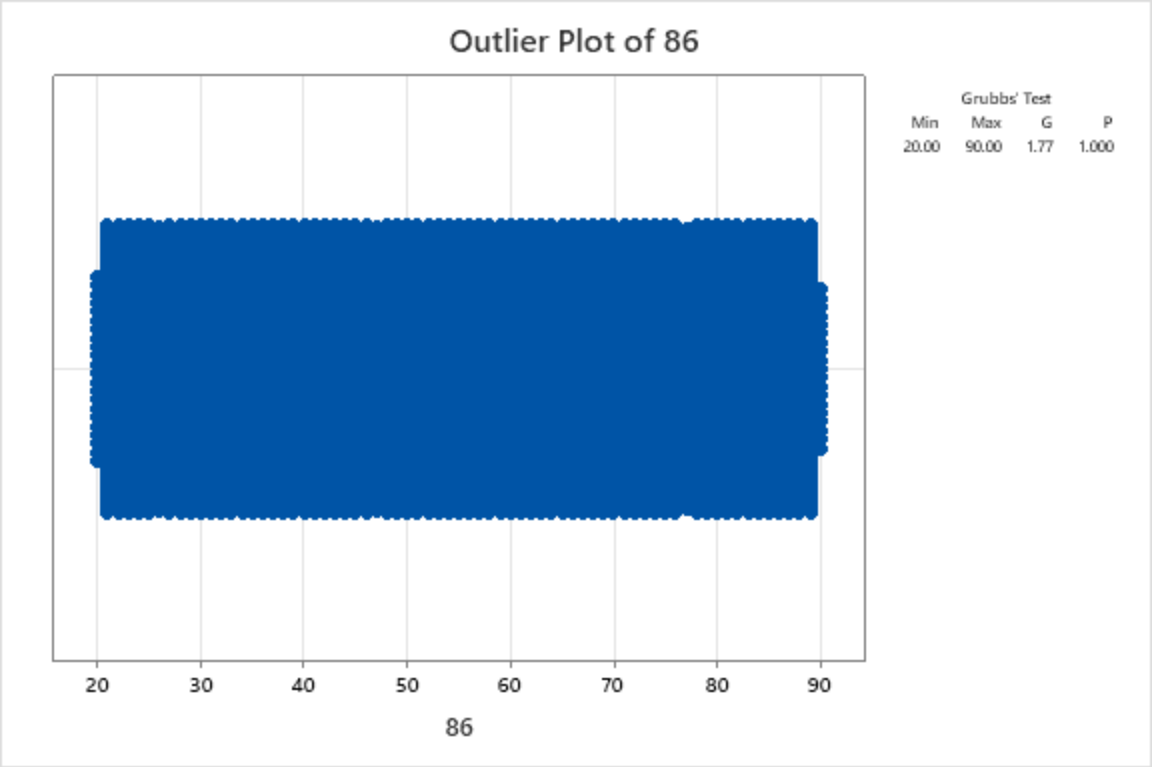






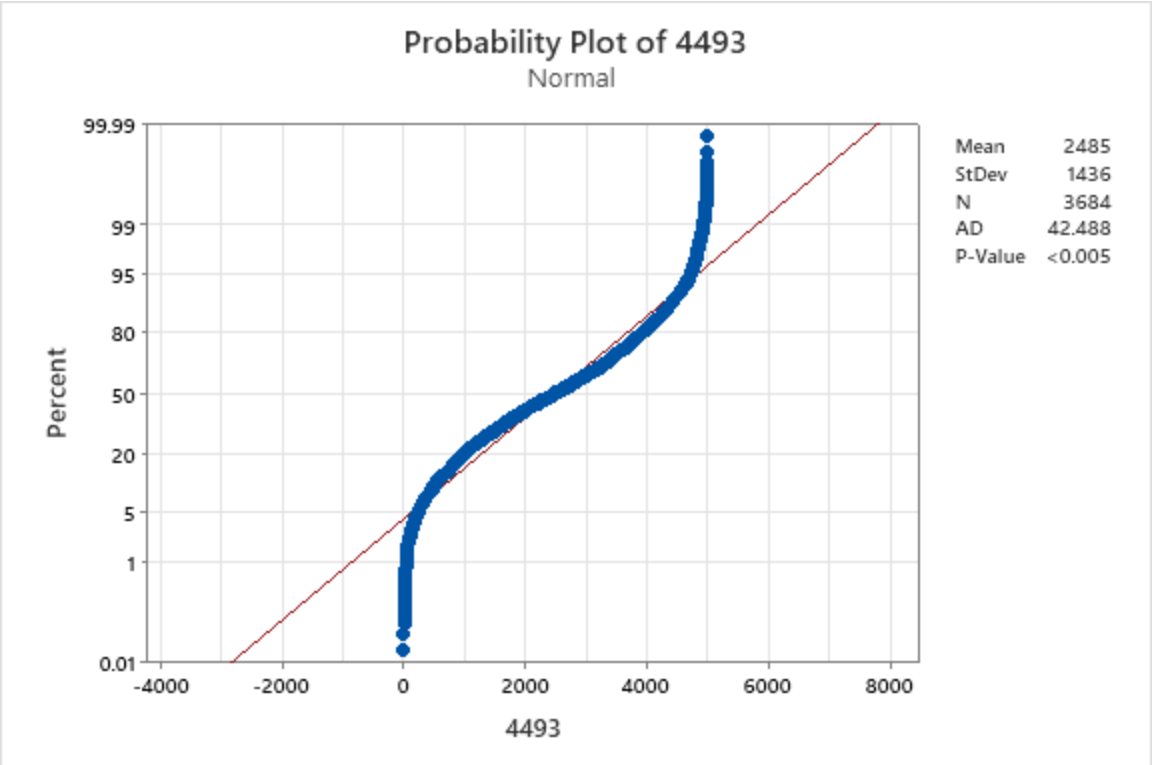




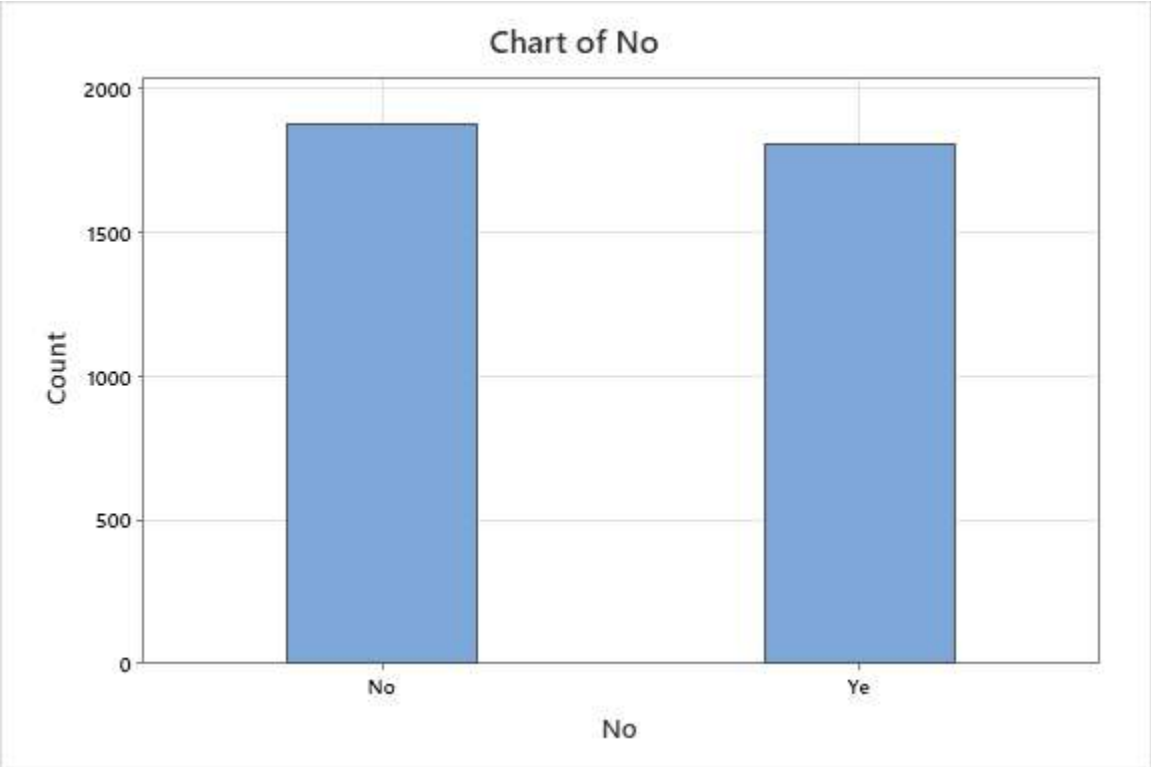


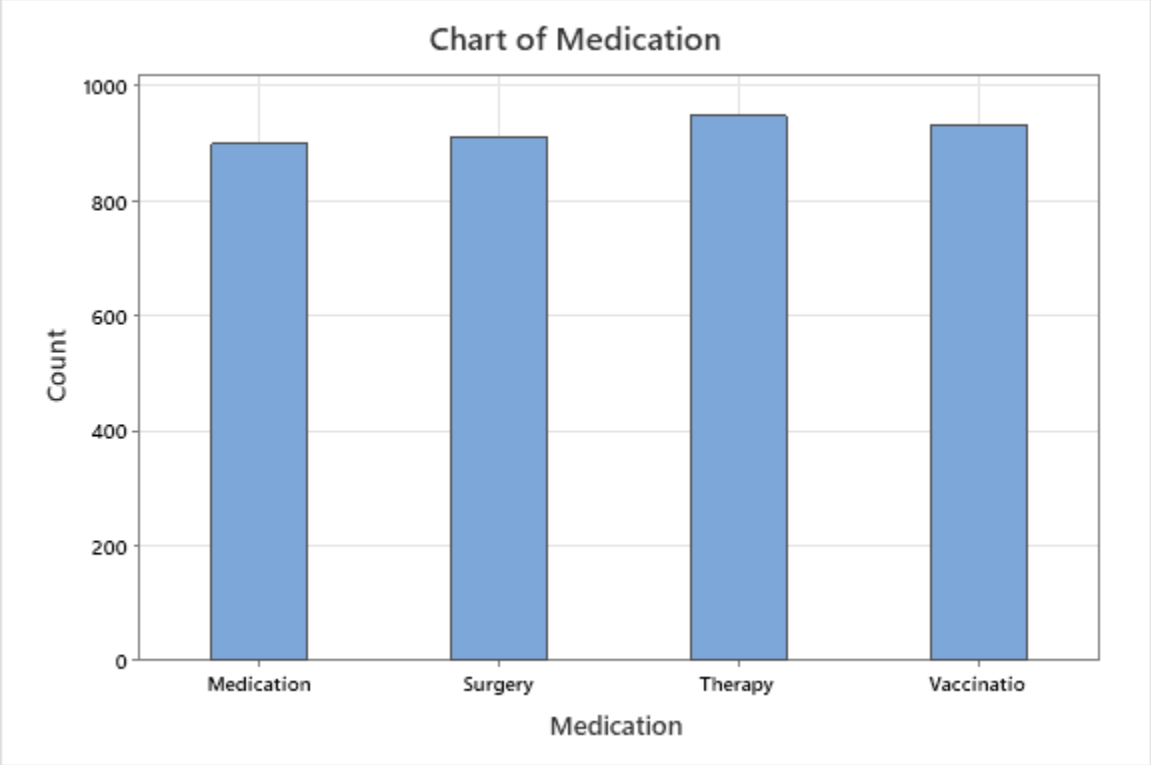
TO check Normality

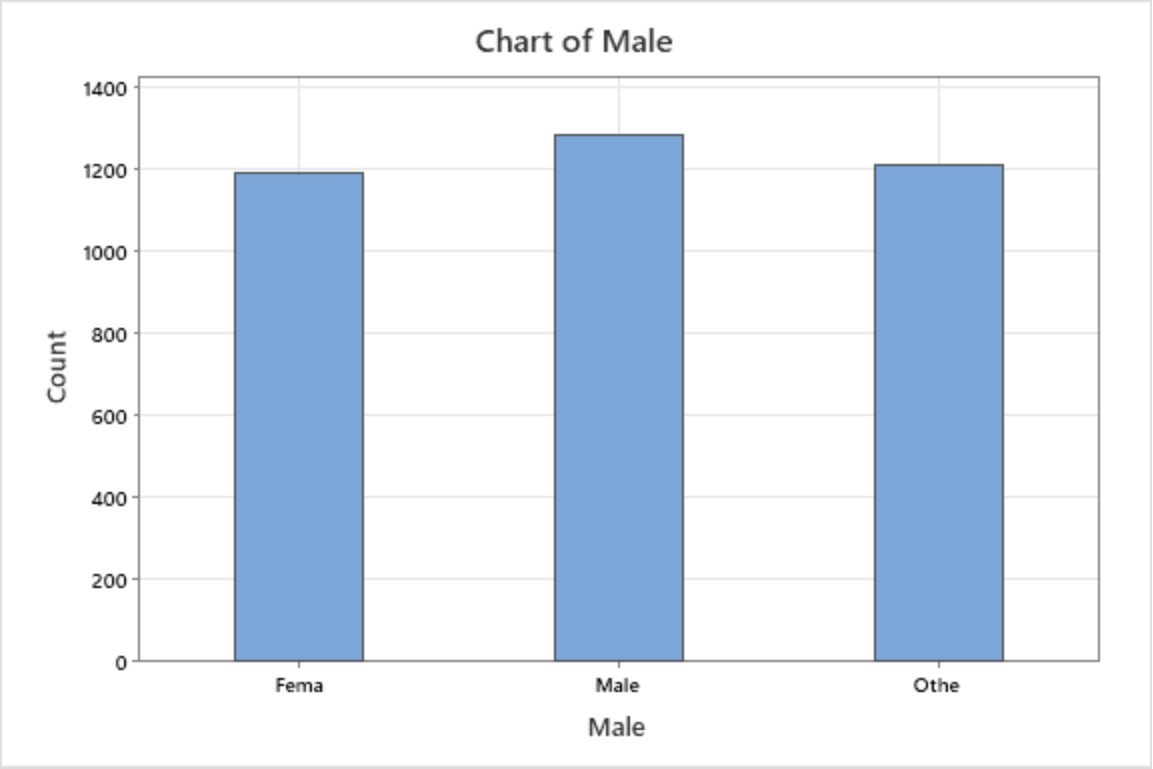
Probability Plot of 4493

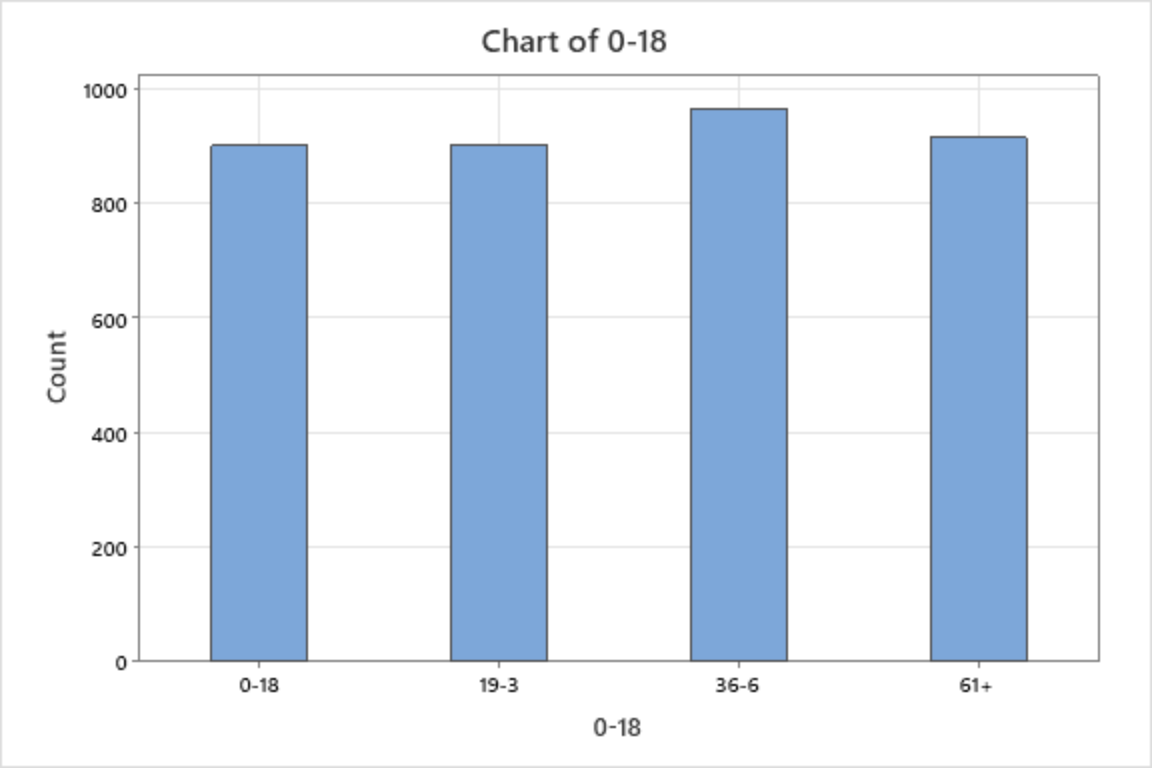


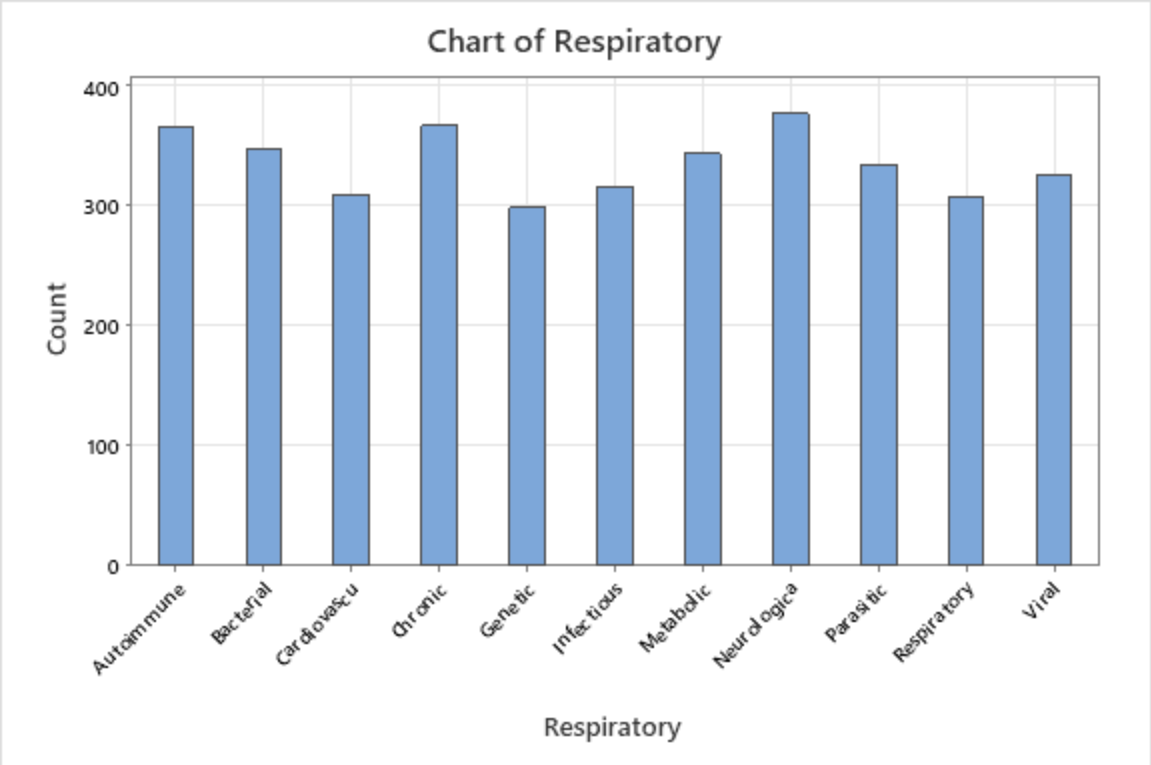
Bar charts

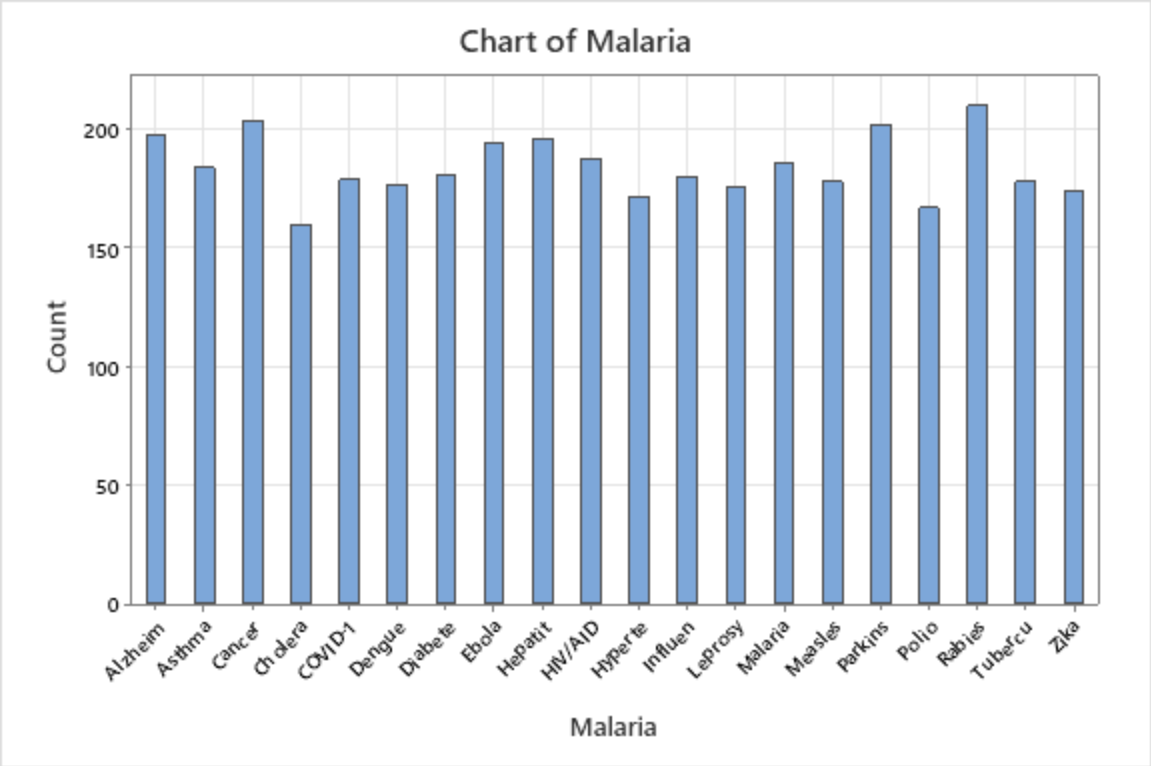




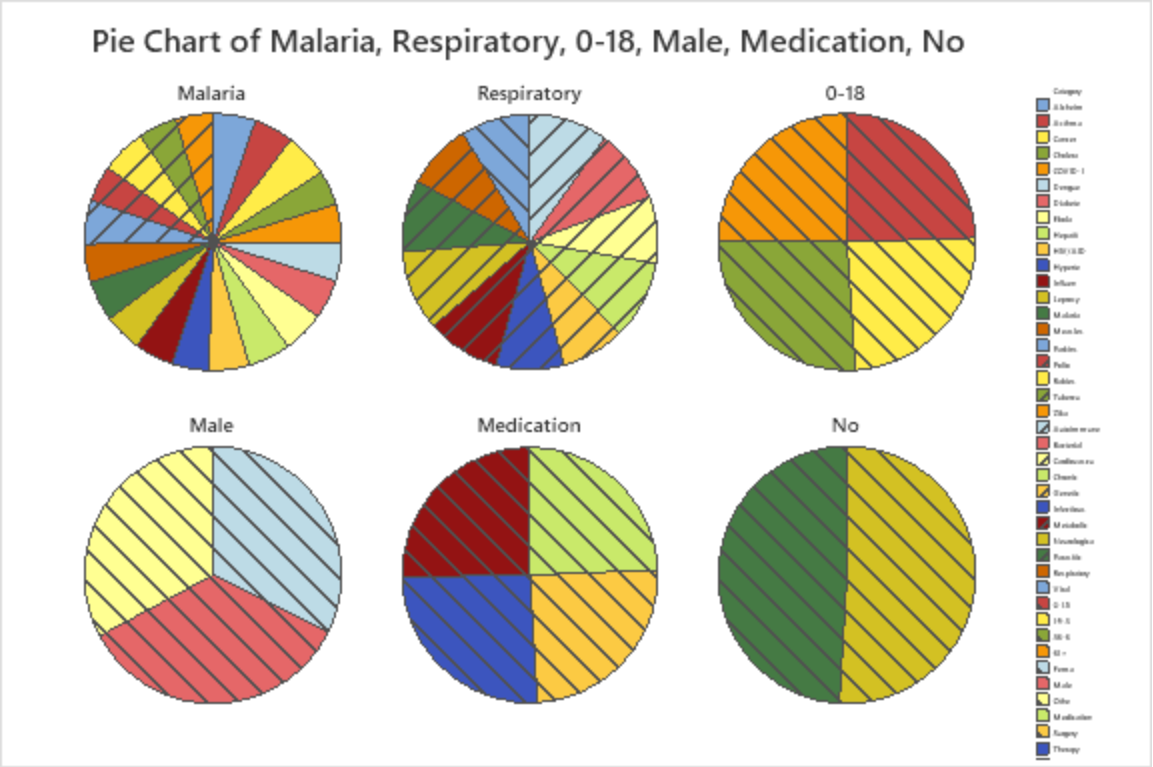




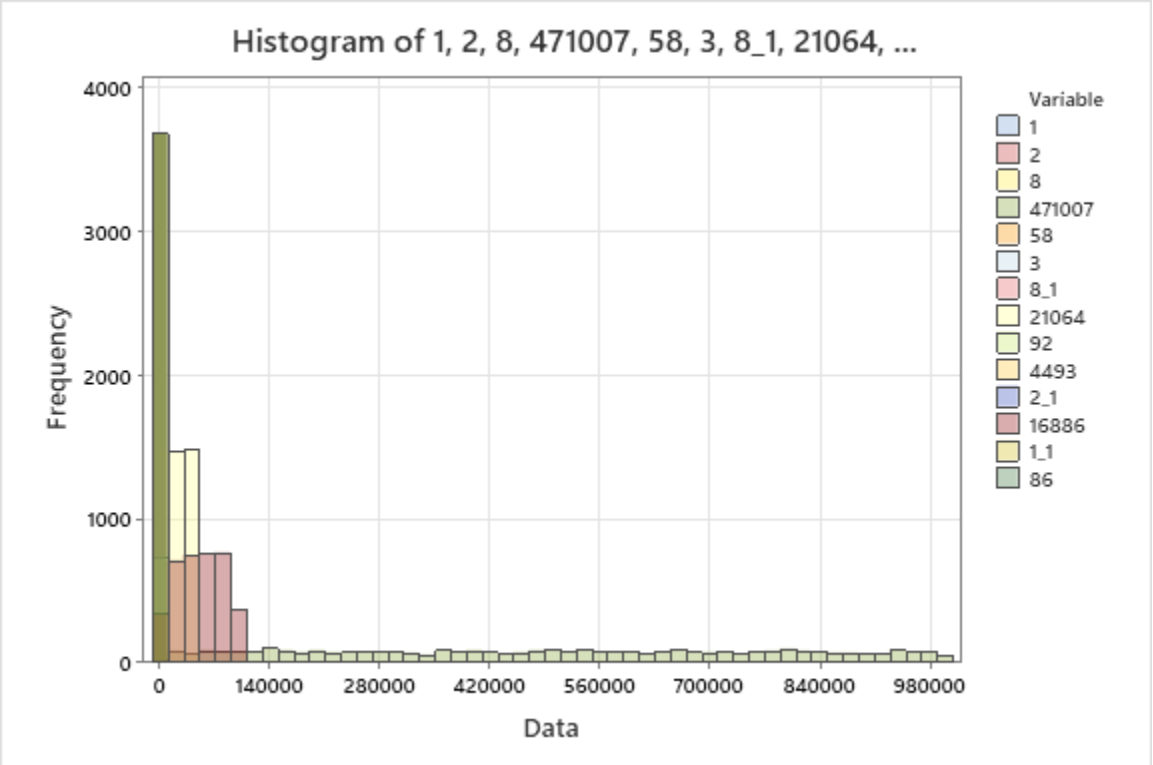




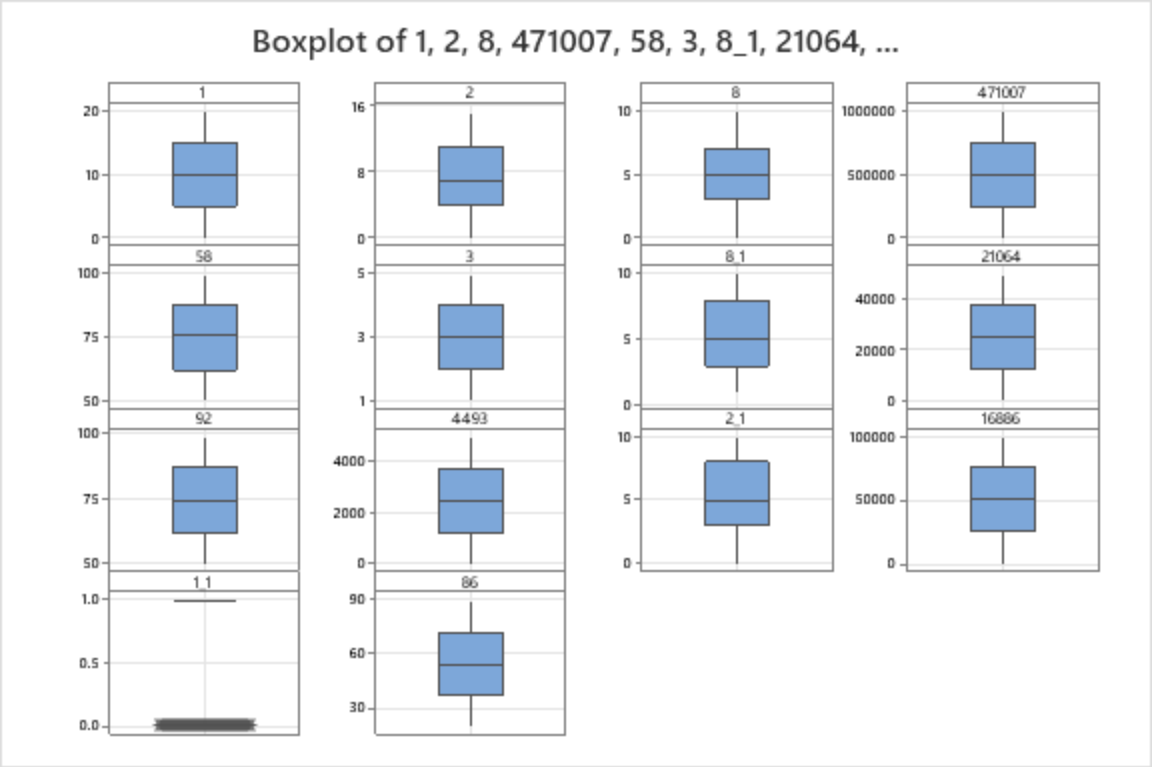
Piechart:



Histogram:



Box plot



One way ANOVA:

Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	α = 0.1

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Male	3	Fema, Male, Othe

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Male	2	27.5	13.732	1.66	0.190
Error	3681	30408.1	8.261		
Total	3683	30435.6			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.87417	0.09%	0.04%	0.00%

Means

Male	N	Mean	StDev	90% CI
Fema	1190	5.0529	2.8397	(4.9159, 5.1900)
Male	1283	4.8948	2.9213	(4.7628, 5.0268)
Othe	1211	5.0925	2.8574	(4.9566, 5.2284)

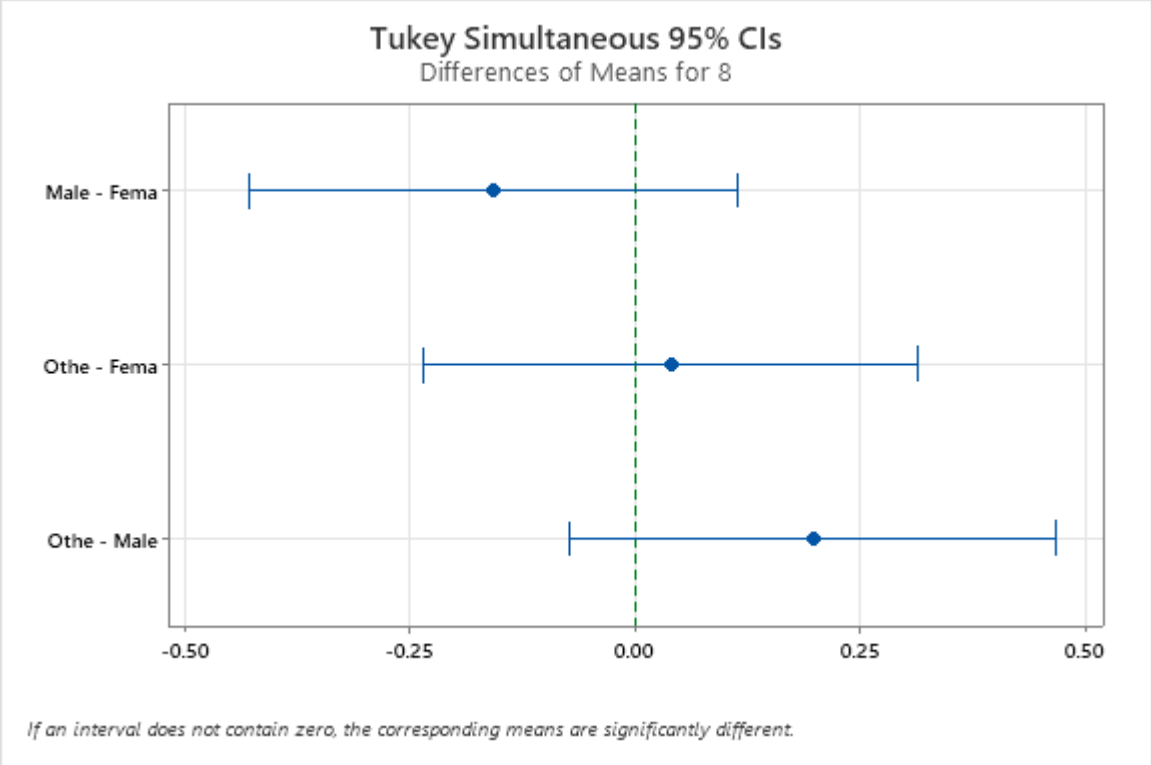
Pooled StDev = 2.87417

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

Male	N	Mean	Grouping
Othe	1211	5.0925	A
Fema	1190	5.0529	A
Male	1283	4.8948	A

Means that do not share a letter are significantly different.

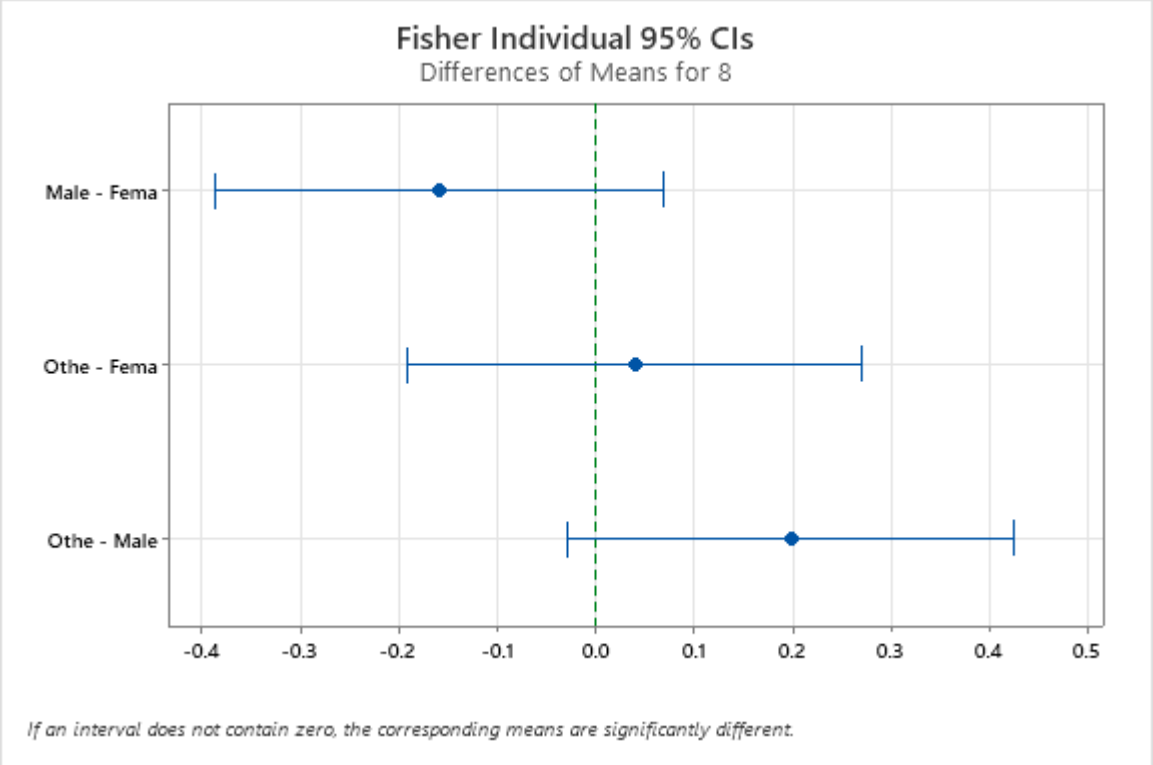


Fisher Pairwise Comparisons

Grouping Information Using the Fisher LSD Method and 95% Confidence

Male	N	Mean	Grouping
Othe	1211	5.0925	A
Fema	1190	5.0529	A
Male	1283	4.8948	A

Means that do not share a letter are significantly different.

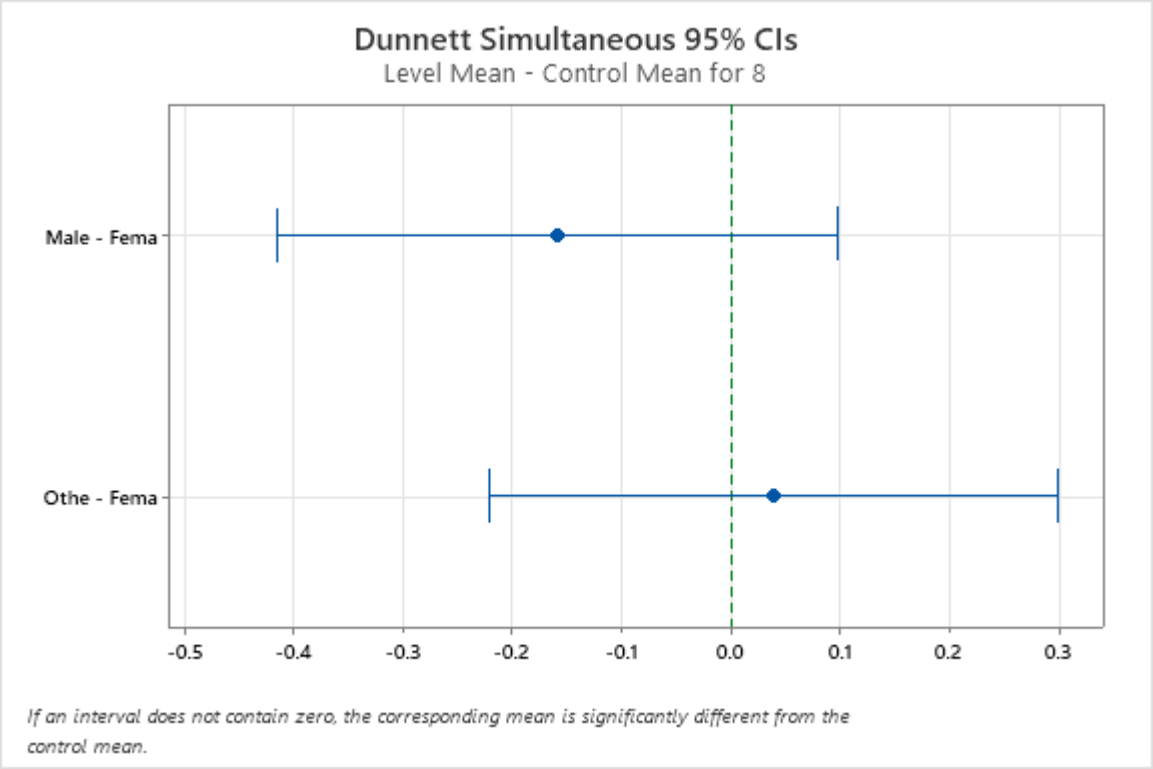


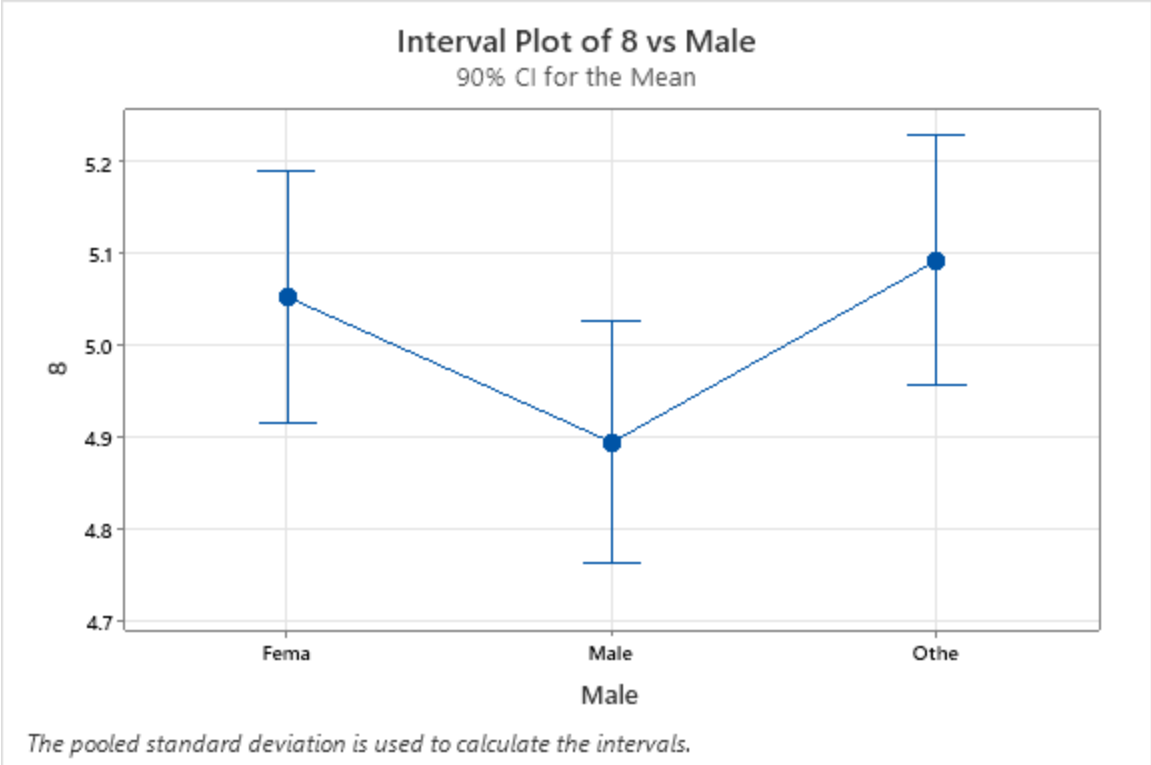
Dunnett Multiple Comparisons with a Control

Grouping Information Using the Dunnett Method and 95% Confidence

Male	N	Mean	Grouping
Fema (control)	1190	5.0529	A
Othe	1211	5.0925	A
Male	1283	4.8948	A

Means not labeled with the letter A are significantly different from the control level mean.





General linear model:

General Linear Model: 471007 versus 2, 4493, Medication

Method

Factor coding (-1, 0, +1)

Factor Information

Factor	Type	Levels	Values
Medication	Fixed	4	Medication, Surgery, Therapy, Vaccinatio

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
2	1	23781276272	23781276272	0.29	0.592
4493	1	59277738258	59277738258	0.72	0.398
	3	6.49274E+11	2.16425E+11	2.61	0.050
Medication					
Error	3678	3.04818E+14	82875913693		
Lack-of-Fit	3658	3.02958E+14	82820605621	0.89	0.683
Pure Error	20	1.85984E+12	92991760071		
Total	3683	3.05553E+14			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
287882	0.24%	0.10%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	501198	12640	39.65	0.000	
2	589	1099	0.54	0.592	1.00
4493	-2.80	3.30	-0.85	0.398	1.00
Medication					
Medication	-14987	8286	-1.81	0.071	1.51
Surgery	-6566	8250	-0.80	0.426	1.51
Therapy	605	8143	0.07	0.941	1.50

Regression Equation

Medication	
Medication	471007 = 486210 + 589 2 - 2.80 4493
Surgery	471007 = 494632 + 589 2 - 2.80 4493
Therapy	471007 = 501803 + 589 2 - 2.80 4493
Vaccinatio	471007 = 522146 + 589 2 - 2.80 4493

Non parametric test

Sign CI: 471007, 4493

Method

η: median of 471007, 4493

Descriptive Statistics

Sample	N	Median
471007	3684	500263
4493	3684	2499

90% Confidence Interval for η

Sample	CI for η	Achieved Confidence	Position
471007	(488321, 511445)	89.71%	(1793, 1892)
	(488259, 511531)	90.00%	Interpolation
	(488170, 511655)	90.39%	(1792, 1893)
4493	(2429, 2571)	89.71%	(1793, 1892)
	(2428.18, 2571.41)	90.00%	Interpolation
	(2427, 2572)	90.39%	(1792, 1893)

Wilcoxon Signed Rank CI: 471007, 4493

Method

η: median of 471007, 4493

Descriptive Statistics

Sample	N	Median	CI for η	Achieved Confidence
471007	3684	498810	(490861, 506651)	90.00%
4493	3684	2485	(2445.5, 2524.5)	90.00%

Mann-Whitney: 471007, 4493

Method

η_1: median of 471007
η_2: median of 4493
Difference: η_1 - η_2

Descriptive Statistics

Sample	N	Median
471007	3684	500263
4493	3684	2499

Estimation for Difference

Difference	CI for Difference	Achieved Confidence
497865	(488277, 507157)	90.00%

Test

Null hypothesis	H_0: $\eta_1 - \eta_2 = 0$
Alternative hypothesis	H_1: $\eta_1 - \eta_2 \neq 0$

Method	W-Value	P-Value
Not adjusted for ties	20350850.00	0.000
Adjusted for ties	20350850.00	0.000

Kruskal-Wallis Test: 471007 versus Medication

Descriptive Statistics

Medication	N	Median	Mean Rank	Z-Value
Medication	898	482497	1786.4	-1.82
Surgery	910	496694	1817.8	-0.81
Therapy	946	492924	1844.0	0.05
Vaccinatio	930	530168	1919.2	2.54
Overall	3684		1842.5	

Test

Null hypothesis	H_0: All medians are equal
Alternative hypothesis	H_1: At least one median is different

Method	DF	H-Value	P-Value
Not adjusted for ties	3	7.83	0.050
Adjusted for ties	3	7.83	0.050