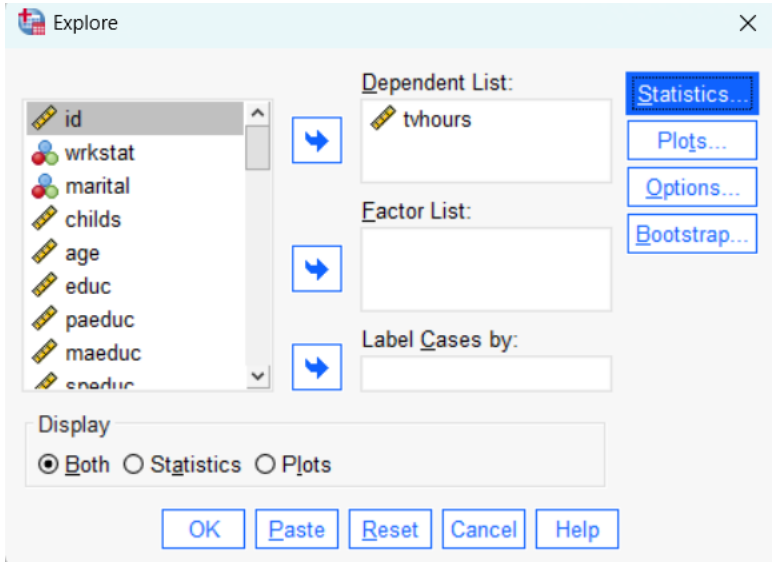
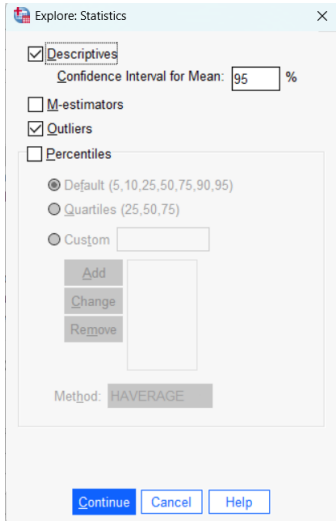
VERSION 1

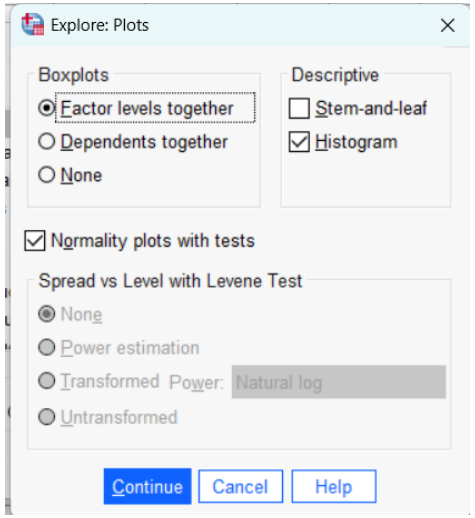
How to Quickly Detect Outliers in IBM SPSS Statistics

Detecting Outliers

1. Open IBM SPSS Statistics.
2. Load your dataset.
3. Navigate to **Analyze > Descriptive Statistics > Explore**.
4. Select the variable(s) you want to analyze.
5. Choose the **statistics** and **plots** you want to generate.
6. Click OK to run the procedure.







VERSION 1

Output

- Check Descriptives - Look at the difference between the mean and the 5% trimmed mean, consider data distribution and minimum and maximum values.
- Review the table of outliers.
- Review Boxplots - Outliers and extreme cases are displayed as individual points beyond the whiskers.
- Examine Stem and Leaf Plots to find outliers in the data distribution.

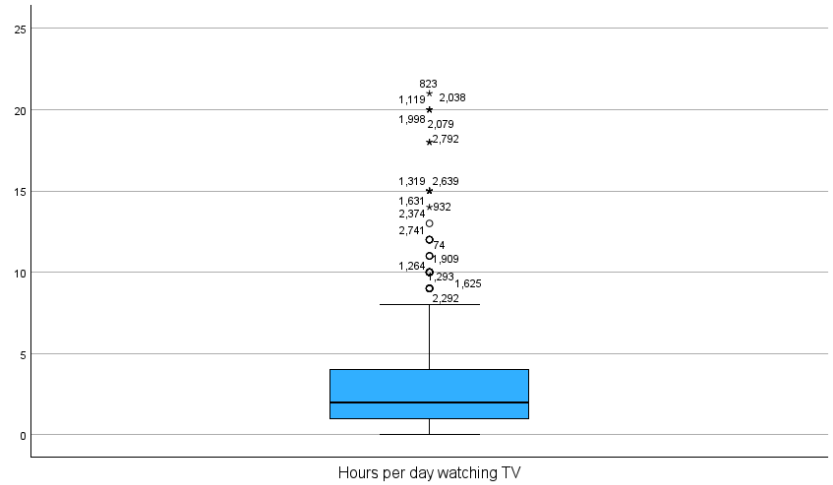
Descriptives

			Statistic	Std. Error
Hours per day watching TV	Mean		2.86	.046
	95% Confidence Interval for Mean	Lower Bound	2.77	
		Upper Bound	2.95	
	5% Trimmed Mean		2.64	
	Median		2.00	
	Variance		5.049	
	Std. Deviation		2.247	
	Minimum		0	
	Maximum		21	
	Range		21	
	Interquartile Range		3	
	Skewness		2.572	.051
	Kurtosis		12.336	.101

Extreme Values

			Case Number	Value
Hours per day watching TV	Highest	1	1688	21
		2	823	20
		3	1119	20
		4	1998	20
		5	2038	20
	Lowest	1	2791	0
		2	2788	0
		3	2695	0
		4	2674	0
		5	2631	0[a]

a. Only a partial list of cases with the value 0 are shown in the table of lower extremes.



Note: The **5% trimmed mean** sorts data removing lowest and highest 5% of values – averages the remaining values.

VERSION 1

Flagging Outliers in the Data

- You can use **Data Validation** to identify the number of rule violations and flag cases that violate the rule.
- Below, the rule was if TV Hours per day > 7 then flag case in the data.

Variable Summary

	Rule	Number of Violations
Hours per day watching TV	TV_Rule	96
	Total	96

[Data Validation - SPSS Analytics Partner]

VERSION 1

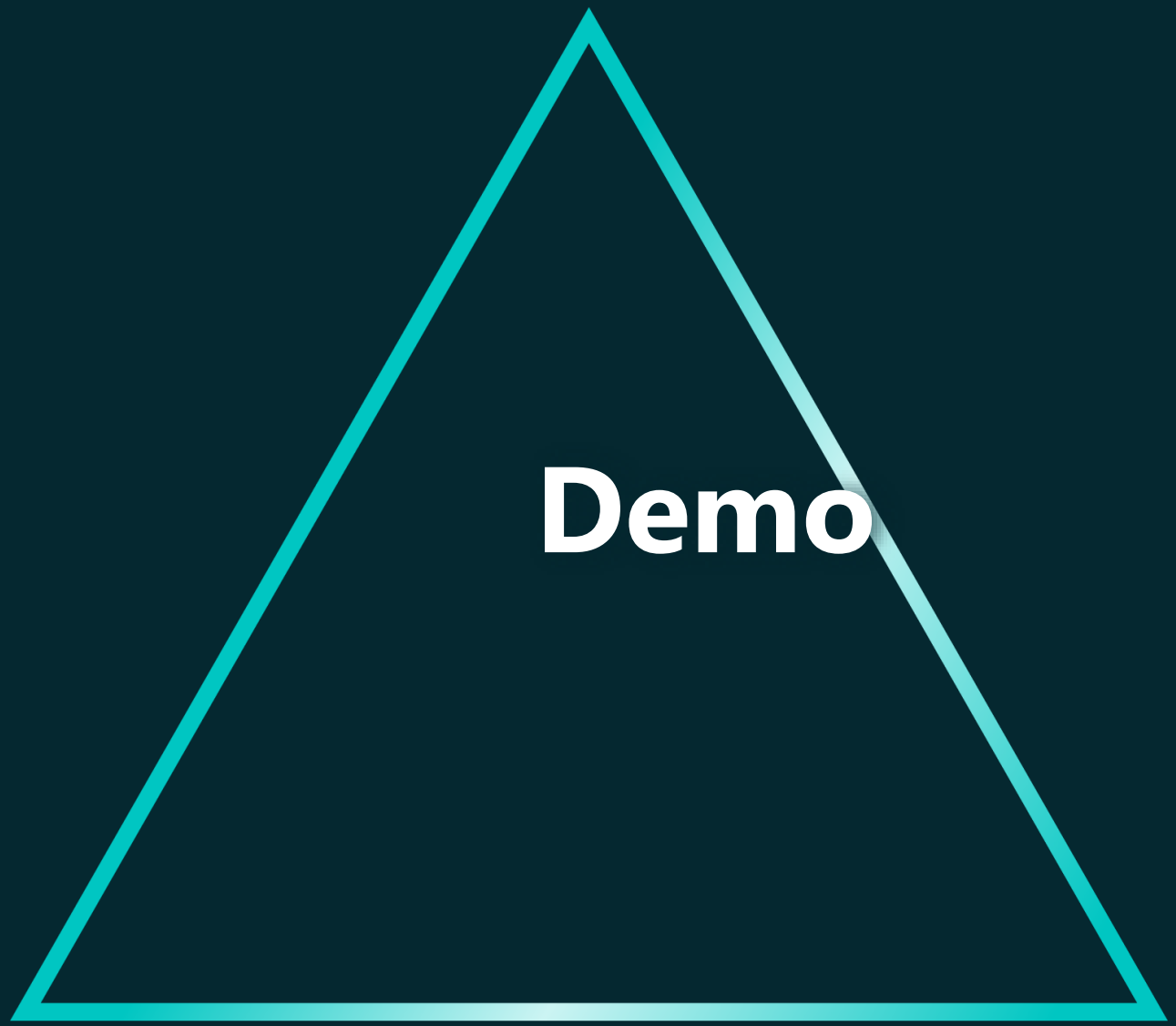
Tips and Best Practices

- Visualise data to spot outliers.
- Recognize the effect of outliers on your analysis.
- Verify outliers using methods like Boxplots and Data Validation.
- Document and explain your approach to handling outliers.

- **Outlier**: A data point far outside the normal range, potentially valid or erroneous.
- **Anomalous case**: Any data point that deviates from expected patterns, including outliers and errors.

- *Data understanding and examination is crucial!*

VERSION 1



VERSION 1