CONCEPTS OF STATISTICS II
PROJECT WEEK #2
SETH HOWELLS
07/19/2020

TOPIC: UNIVARIATE PROFILING: EXAMINING THE SHAPE OF THE DISTRIBUTION

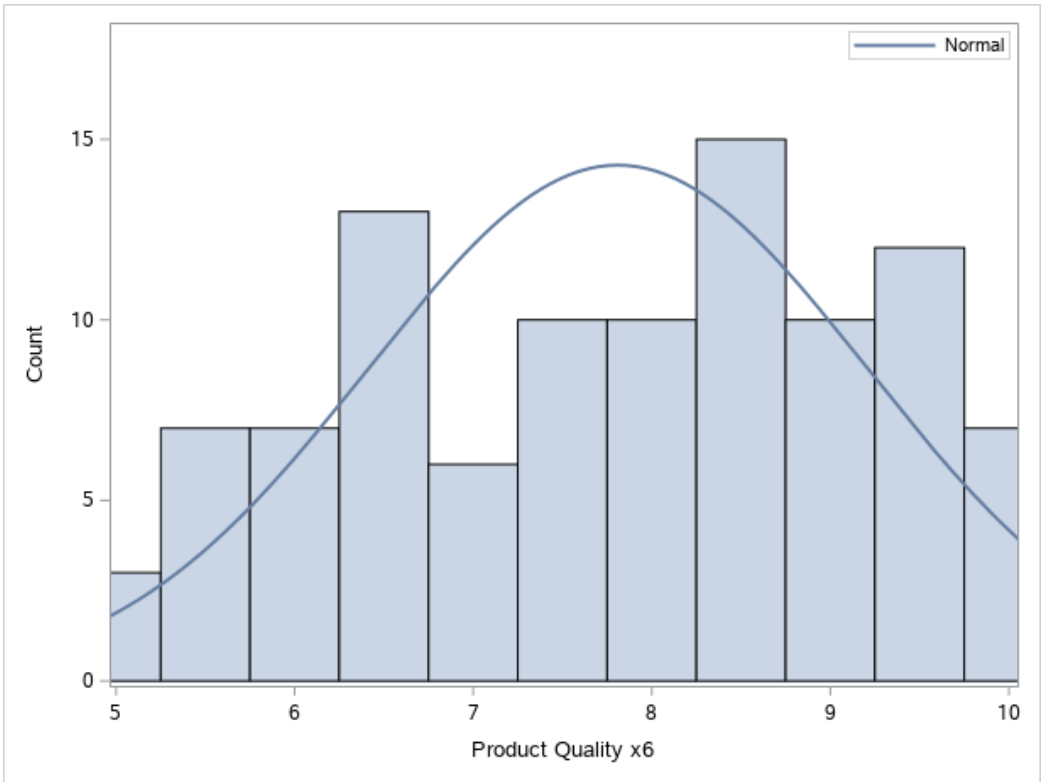
Data: HBAT.xls
Type: Histogram with normal curve overlay

The first analysis on the HBAT file examines the Product Quality (x6) distribution and compares the distribution to a normal curve, as shown in the line overlay. The y-axis (Count) refers to the frequency of values that fall between the x-axis bins.

The histogram infers that the distribution is fairly normal and does not present skewness despite the high frequency at 8.5 because the distribution falls short of the normal curve while the tails present higher frequencies than the curve. The interpretation is that there is a shortage of obversions toward the center of the distribution from x = 7 to x = 8. This is further supported by the summary statistics of the Product Quality vector. Skewness shows that it is between the range of -1 to 1 while Kurtosis computes less than -1, which indicates that the distribution is too flat.

Analysis Variable : x6 Product Quality								
Mean	**Std Dev**	**Min**	**Max**	**Median**	**N**	**N Miss**	**Skewness**	**Kurtosis**
7.81	1.3962793	5	10	8	100	0	-0.2445019	-1.1318375

FIG 2-1



TOPIC: BIVARIATE PROFILING: EXAMINING GROUP DIFFERENCES

Data: HBAT.xls
Type: Scatter plot matrix with histograms and correlation coefficients table

Bivariate analysis was performed on Product Quality (x6), E-Commerce Activities/Website (x7), Technical Support (x8), Salesforce Image (x12), and Competitive Pricing. All variables in the analysis are metric. A correlation matrix and scatter plot matrix were performed in efforts to examine all bivariate relationship combinations.

Pearson Correlation Coefficients Matrix: measures the strength of the association between two variables.

Rule of thumb:

- coefficients between ± 0.50 and ± 1 = strong correlation
- coefficients between ± 0.30 and ± 0.49 = moderate correlation
- coefficients between ± 0 and ± 0.29 = small correlation

Interpretation: E-Commerce Activity (x7) and Salesforce Image (x12) have the strongest correlation while E-Commerce Activities (x7) and Technical Support (x8) have the smallest correlation.

Because the three of the variables are described as "Performance Perceptions Variables," these highlighted correlations below suggest two key points:

1. E-Commerce Activity (x7) and Salesforce Image (x12): as E-Commerce Activities performance increases, the Salesforce Image performance increases, and vice-versa.
2. E-Commerce Activities (x7) and Technical Support (x8): E-Commerce Activities has no association with Technical Support and therefore do not impact another's performance.

Scatter Plot Matrix: graphical display comprised of multiple scatter plots which display the bivariate relationship between two variables. This provides a quick visual que to examine linearity and association. Data points forming a line on the scatter plot define linearity. The importance of linearity is similar to the overarching theme of correlation analysis: X relationship on Y.

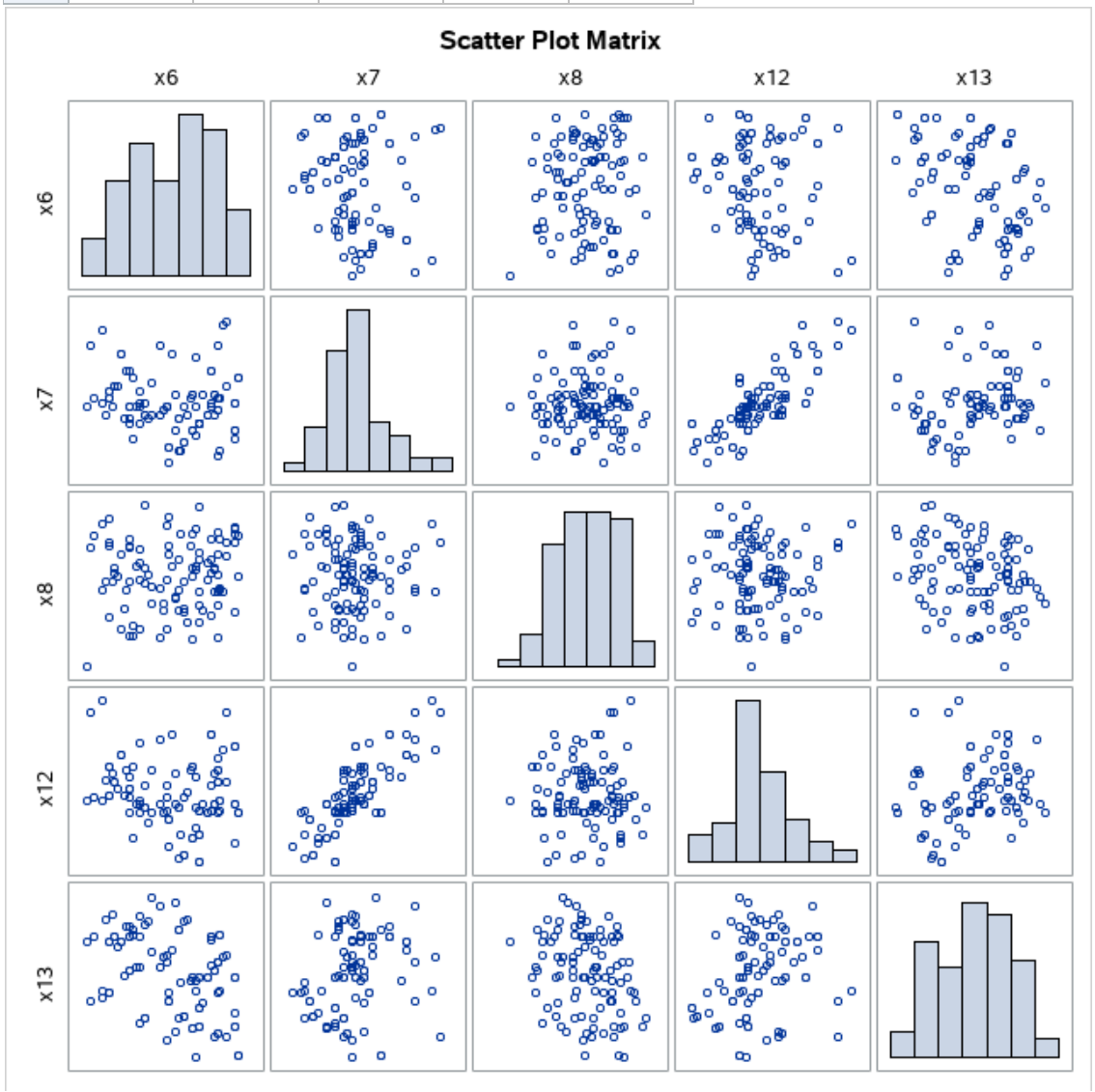
Histograms displayed show the distribution of corresponding variable.

Interpretation: the same variables from correlation will be under investigation.

1. E-Commerce Activity (x7) and Salesforce Image (x12): data points clearly form a line which displays linearity. The high correlation matches the scatter plot.
2. E-Commerce Activities (x7) and Technical Support (x8): data points clearly display no defined line. The extremely low correlation value matches the scatter plot.

5 Variables:	x6 x7 x8 x12 x13

Pearson Correlation Coefficients, N = 100					
	x6	**x7**	**x8**	**x12**	**x13**
x6	1.00000	-0.13716	0.09560	-0.15181	-0.40128
x7	-0.13716	1.00000	**0.00087**	**0.79154**	0.22946
x8	0.09560	0.00087	1.00000	0.01699	-0.27079
x12	-0.15181	0.79154	0.01699	1.00000	0.26460
x13	-0.40128	0.22946	-0.27079	0.26460	1.00000



TOPIC: BIVARIATE PROFILING: EXAMINING GROUP DIFFERENCES

Data: HBAT.xls
Type: Box Plot

Box Plot (or Box and Whiskers): bivariate graphical display consisting of a single metric data and a single categorical data that provides details on the analysis variable's distribution (Product Quality x6 & E-Commerce Activities x7) grouped by the categorical variable (Customer Type x3).

- **Lower Whisker**: represents the range of the 25th percentile and lower.
- **Box**: represents the range between the 25th and 75th percentile.
- **Line in Box**: represents the median.
- **Upper Whisker**: represents the range of the 75th percentile and higher.
- **Outliers**: represented by the open circle above and below the boundaries.

Interpretation: The length of the whiskers and positioning of the box is representative of the variables distribution when grouped by the categorical data. Equal whisker lengths with the median being directly center on the box would indicate a normal distribution. Skewed data will be easy to find if the whisker lengths are severely unequal.

FIG 2-3A

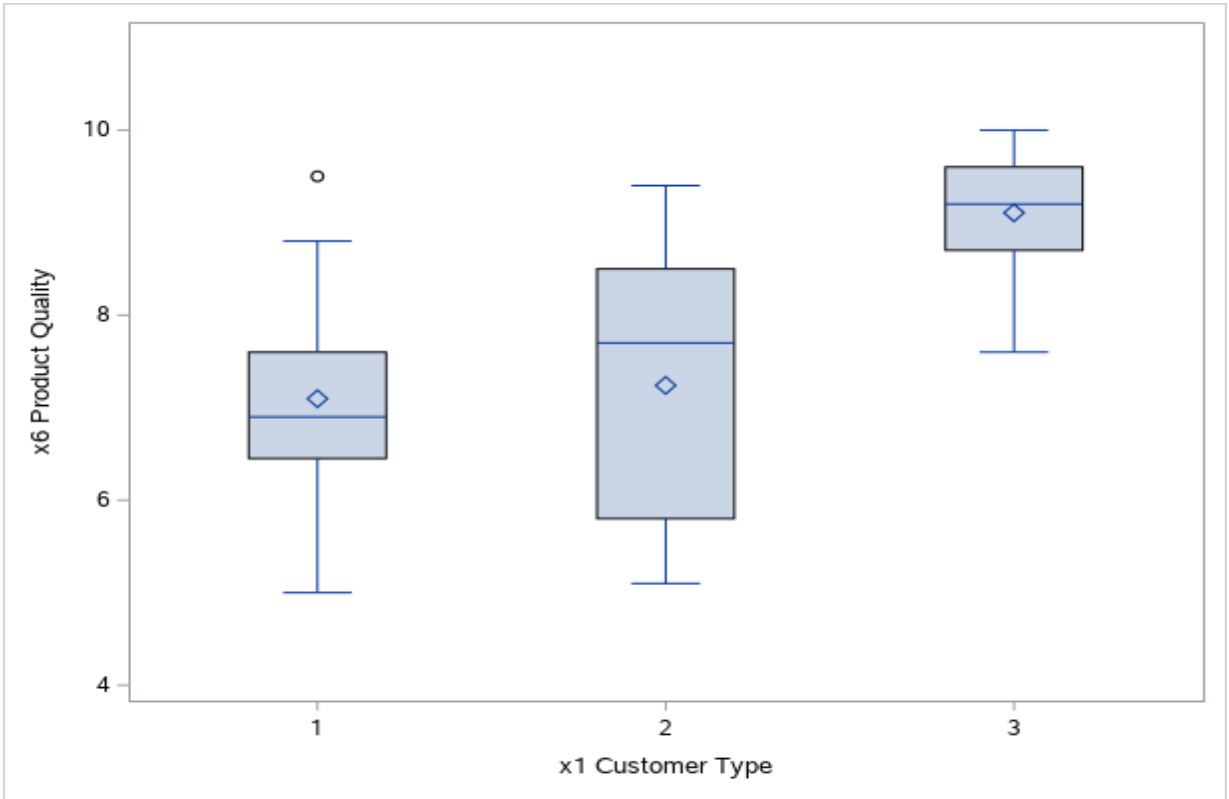


FIG 2-3B