

1. The histograms are:

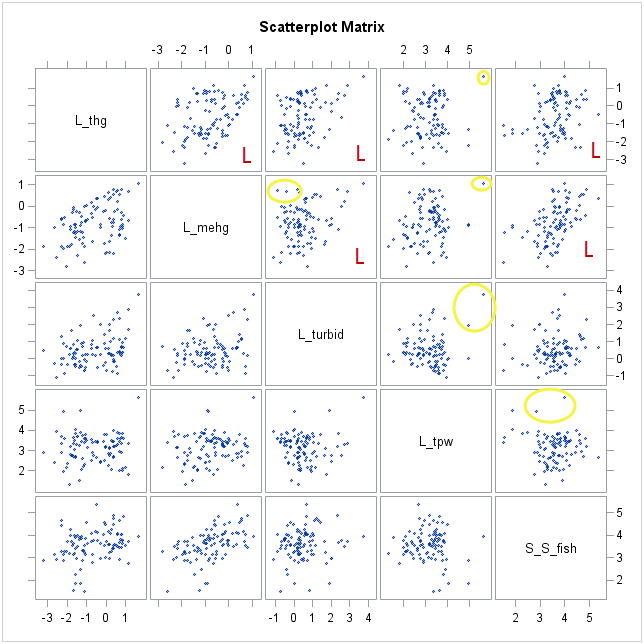
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Original** | **Log** | **Square Root** | **Quarter Root** |
| **Total Mercury thg** | Histogram for thg | Histogram for L_thg | Histogram for S_thg | Histogram for S_thg |
| **Methyl Mercury mehg** | Histogram for mehg | Histogram for L_mehg | Histogram for S_mehg | Histogram for S_S_mehg |
| **Turbidity** | Histogram for turbid | Histogram for L_turbid | Histogram for S_turbid | Histogram for S_S_turbid |
| **Phosphorous** | Histogram for tpw | Histogram for L_tpw | Histogram for S_tpw | Histogram for S_S_tpw |
| **Fish** | Histogram for fish | Histogram for L_fish | Histogram for L_fish | Histogram for S_S_fish |

1. The transformations should be applied and the following table recommends the transformation as applicable:

|  |  |
| --- | --- |
| **Variable** | **Recommended Transformation** |
| Total Mercury thg | Log |
| Methyl Mercury mehg | Log or Quarter Root |
| Turbidity | Log |
| Phosphorous | Log or Quarter Root |
| Fish | Quarter Root |

1. See the above table that has all the histograms
2. We will use the following variables in the scatter plot:

L\_thg, L\_mehg, L\_turbid, L\_tpw, S\_S\_fish



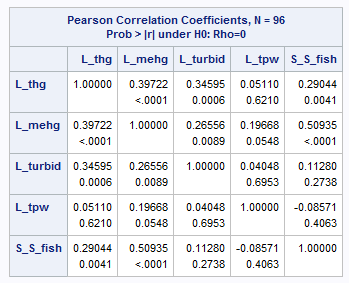
The above scatter plot has the upper diagonal annotated to indicate the following:

 : Wherever the variables show atleast an approximate linear relationship

: Wherever there is an indication of possible outliers.

It is also clear from the diagram that the annotations are indicative and further statistical tests will be required to confirm the nature of relationship and the existence of outliers.

The correlation matrix for the above variables is:

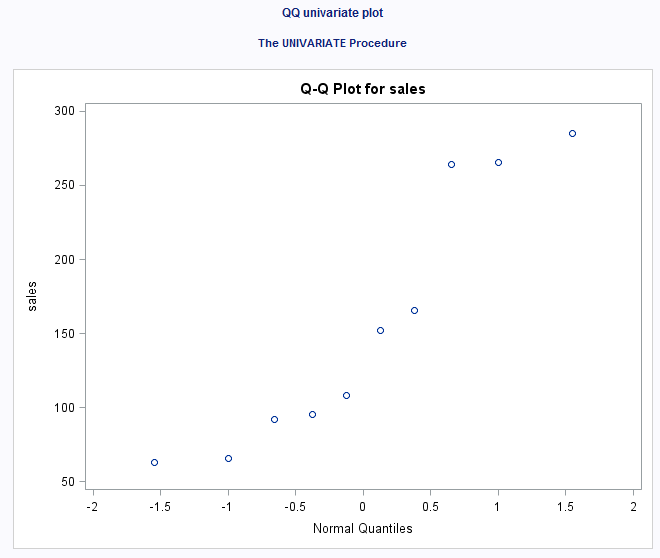


The scatter plot and the correlation indicates that we have a positive relation ship wherever we do have any linear relationship. The only negative correlation is between L\_tpw and S\_S\_fish but the relationship is weak.

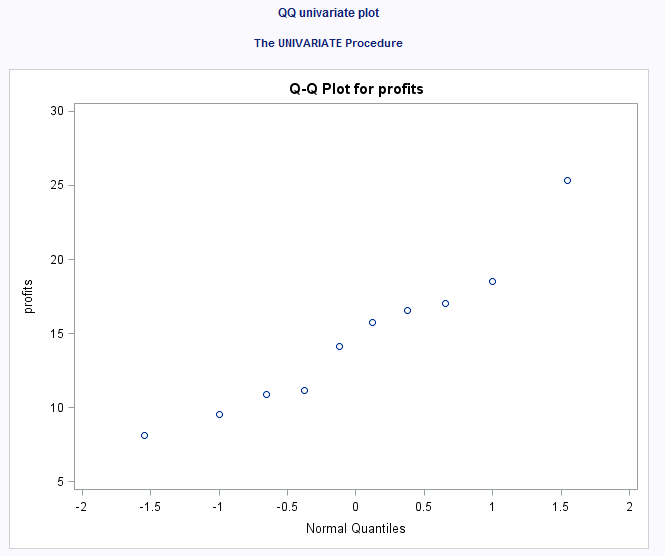
The strongest relationship is between: L\_mehg and S\_S\_fish followed by L\_thg and L\_mehg



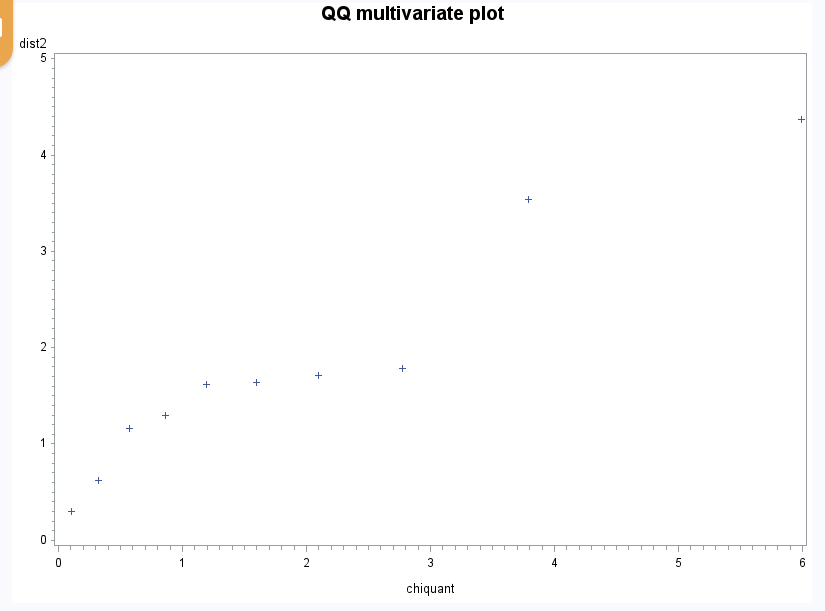
The various plots are below:



The plot indicates that we can’t assume normality of the sales data. For safe assumption of normality we should see a linear relationship in the Q-Q plot. Since we have far from a linear relationship therefore normality is in doubt.



The plot indicates that we can’t assume normality of the profits data. For safe assumption of normality we should see a linear relationship in the Q-Q plot. Since we have far from a linear relationship therefore normality is in doubt.



For multivariate data, we plot the ordered Mahalanobis distances versus estimated quantiles (percentiles) for a sample of size n from a chi-squared distribution with p degrees of freedom. Ideally this should resemble a straight-line for data from a multivariate normal distribution.

Outliers will show up as points on the upper right side of the plot for which the Mahalanobis distance is notably greater than the chi-square quantile value.



In this case, since we have two variables, this should be chi-square with 2 degrees of freedom. In this case, if we are going to consider a 95% prediction ellipse, the critical value for chi-square with 2 degrees of freedom is equal to 5.99 from the statistical table.

The eigenvalues and eigenvectors and other computations of ∑1 - ∑4 are:

∑1

$values

[1] 1.5 0.5

$vectors

[,1] [,2]

[1,] 0.7071068 -0.7071068

[2,] 0.7071068 0.7071068

Half length l1 = sqrt(1.5 \* 5.99) = 2.997

The direction of the axis is given by the first eigenvector [0.7071 0.7071]. Looking at this first eigenvector we can see large positive elements corresponding to both of the variables. This suggests that this particular axis points in the direction specified by e1; that is, increasing values of x1 and x2. The values are equal indicating that the angle is 45 degrees.

Half length l2 = sqrt(0.5 \* 5.99) = 1.731

The direction of the axis is given by the second eigenvector [-0.7071 0.7071]. Looking at this eigenvector we can see large negative elements corresponding to first and positive element corresponding to second variable. This suggests that this particular axis points in the direction specified by e2; that is, decreasing values of x1 and increasing x2.

∑2

$values

[1] 1.9 0.1

$vectors

[,1] [,2]

[1,] 0.7071068 -0.7071068

[2,] 0.7071068 0.7071068

Half length l1 = sqrt(1.9 \* 5.99) = 3.374

The direction of the axis is given by the first eigenvector [0.7071 0.7071]. Looking at this first eigenvector we can see large positive elements corresponding to both of the variables. This suggests that this particular axis points in the direction specified by e1; that is, increasing values of x1 and x2. The values are equal indicating that the angle is 45 degrees.

Half length l2 = sqrt(0.1 \* 5.99) = 0.774

The direction of the axis is given by the second eigenvector [-0.7071 0.7071]. Looking at this eigenvector we can see large negative elements corresponding to first and positive element corresponding to second variable. This suggests that this particular axis points in the direction specified by e2; that is, decreasing values of x1 and increasing x2.

Comparison with ∑1: The first halflength is longer and the second is shorter. The directions are the same.

∑3

$values

[1] 2.207 0.793

$vectors

[,1] [,2]

[1,] 0.3826834 -0.9238795

[2,] 0.9238795 0.3826834

Half length l1 = sqrt(2.207 \* 5.99) = 3.636

The direction of the axis is given by the first eigenvector [0.3827 0.9239]. Looking at this first eigenvector we can see large positive elements corresponding to both of the variables. This suggests that this particular axis points in the direction specified by e1; that is, increasing values of x1 and x2 with higher values of x2.

Half length l2 = sqrt(0.793 \* 5.99) = 2.179

The direction of the axis is given by the second eigenvector [-0.9239 0.3827]. Looking at this eigenvector we can see large negative elements corresponding to first and positive element corresponding to second variable. This suggests that this particular axis points in the direction specified by e2; that is, decreasing values of x1 and increasing x2.

Comparison with ∑1: The halflengths are longer. The quadrants are the same but the angles are different.

∑4

$values

[1] 1.5 0.5

$vectors

[,1] [,2]

[1,] -0.7071068 -0.7071068

[2,] 0.7071068 -0.7071068

Half length l1 = sqrt(1.5 \* 5.99) = 2.997

The direction of the axis is given by the eigenvector [-0.7071 0.7071]. Looking at this eigenvector we can see large negative elements corresponding to first and positive element corresponding to second variable. This suggests that this particular axis points in the direction specified by e2; that is, decreasing values of x1 and increasing x2.

Half length l2 = sqrt(0.5 \* 5.99) = 1.731

The direction of the axis is given by the eigenvector [-0.7071 -0.7071]. Looking at this eigenvector we can see large negative elements corresponding to both variables. This suggests that this particular axis points in the direction specified by e2; that is, decreasing values of x1 and x2.

Comparison with ∑1: The halflengths are same. The quadrants are different but they both bisect their respective quadrants.