

1. Xij = Observation for variable j in subject i

Here we have data on 42 measurements (subjects) and 7 different gases (variables). So,

p = 7 variables

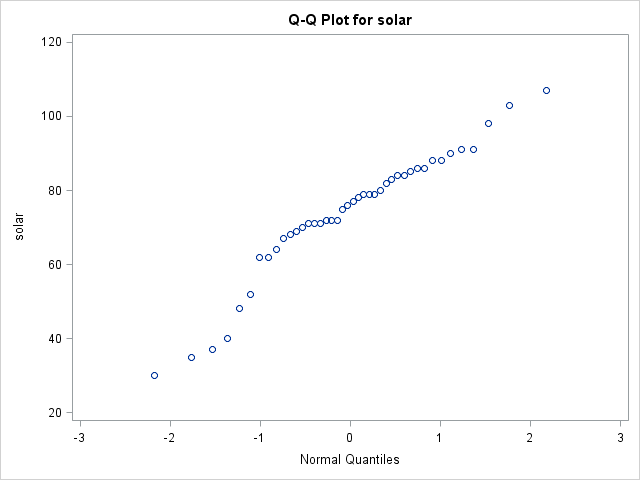
n = 42 subjects

µj = population mean for variable j. For instance the population mean for variable X1 wind

σjk = population covariance i.e. a measure of the association between pairs of variables in a population. Here, the population covariance between variables j and k (two different types of gases). For instance the population covariance between wind and CO

σjk = 0 implies that the two variables are uncorrelated.

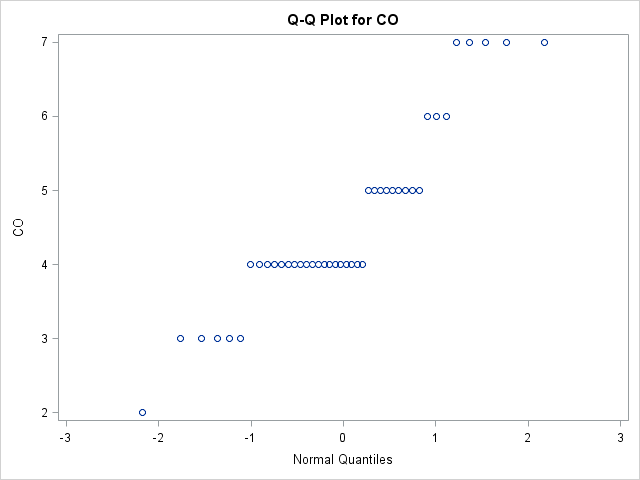
1. The various plots are below:



The plot indicates that we can’t assume normality of the 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.



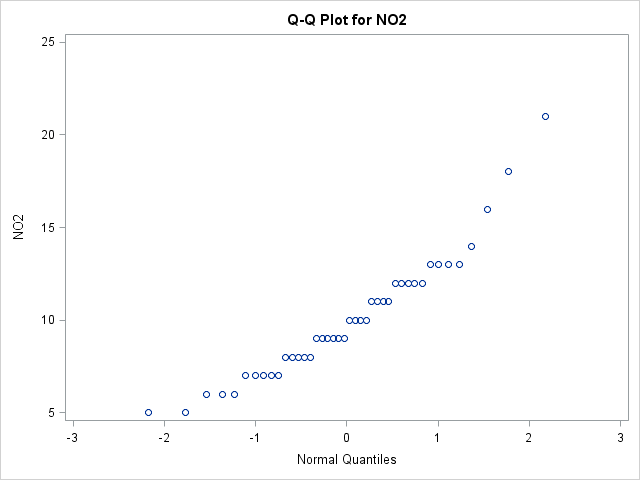
We also see that the p-value of the ryan joiner test < alpha (assume 0.05) and therefore we reject the null hypothesis of population normality.



The plot indicates that we can assume normality of the data.



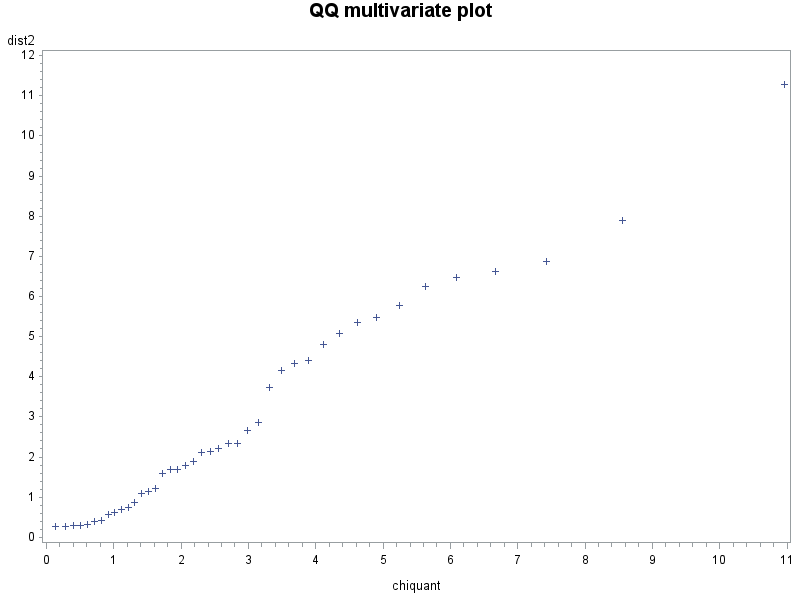
We also see that the p-value of the ryan joiner test > alpha (assume 0.05) and therefore we fail to reject the null hypothesis of population normality. We can assume normality



The plot indicates that we can’t assume normality of the 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.



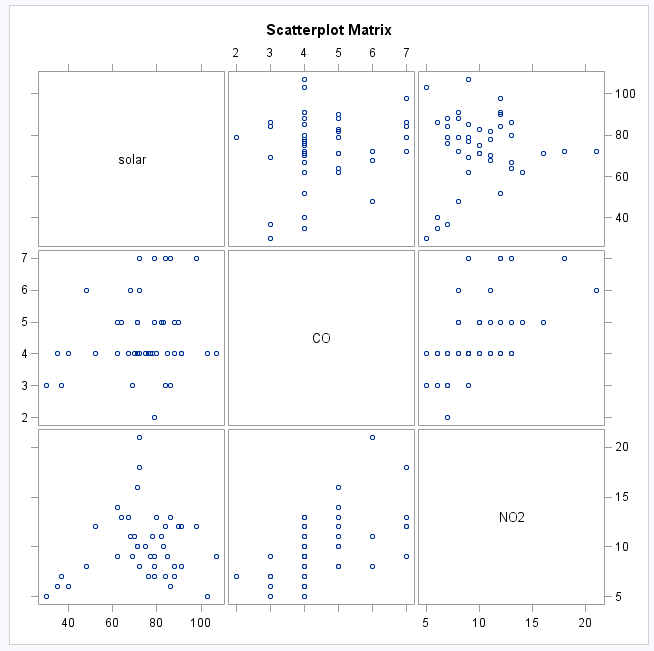
We also see that the p-value of the ryan joiner test < alpha (assume 0.05) and therefore we reject the null hypothesis of population normality.

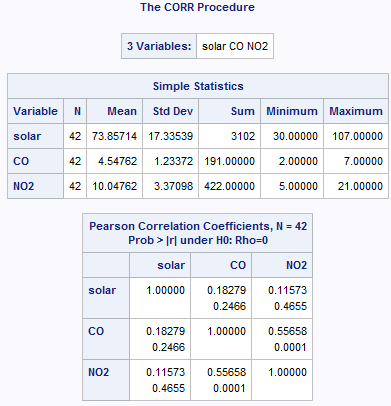


For multivariate data, we plot the ordered Mahalanobis distances versus estimated quantiles (percentiles) for a sample of size n from a chi-squared distribution with p degrees of freedom. Ideally this should resemble a straight-line for data from a multivariate normal distribution.

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.

1. The scatter plot is:





Confidence Interval



1. The sample means are:

x1bar (oxygen) = 34.64 and x2bar (solids) = 33.18

Therefore the sample mean vector x¯=

We have the following variances and covariance:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | plant | oxygen | solids | O\*S |
|  | 1 | 25 | 15 | 375 |
|  | 2 | 28 | 13 | 364 |
|  | 3 | 36 | 22 | 792 |
|  | 4 | 35 | 29 | 1015 |
|  | 5 | 15 | 31 | 465 |
|  | 6 | 44 | 64 | 2816 |
|  | 7 | 42 | 30 | 1260 |
|  | 8 | 54 | 64 | 3456 |
|  | 9 | 34 | 56 | 1904 |
|  | 10 | 29 | 20 | 580 |
|  | 11 | 39 | 21 | 819 |
| Avg |  | 34.63636 | 33.18181818 | 1258.72727 |
| Sum |  | 381 | 365 | 13846 |
| Covariance |  |  |  | 120.372727 |
| Variance |  | 109.2545 | 363.7636364 |  |
| SD |  | 10.45249 | 19.07258861 |  |

Therefore sample covariance matrix S =

The correlation between X1 and X2 is = 120.37/(sqrt(109.25) \* sqrt(363.76)) = 0.6038