

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

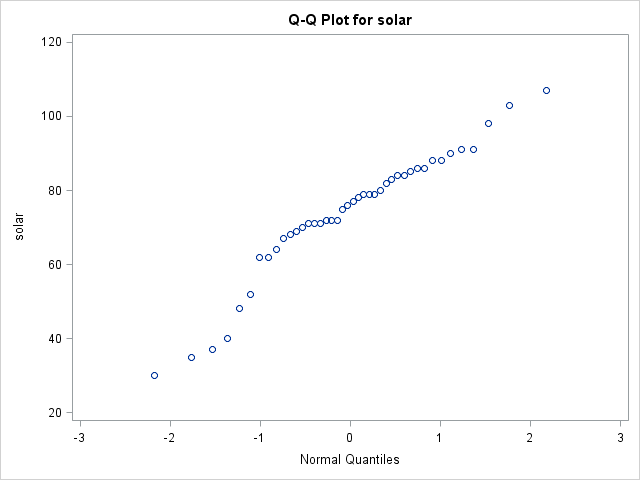
n = 42 subjects

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.

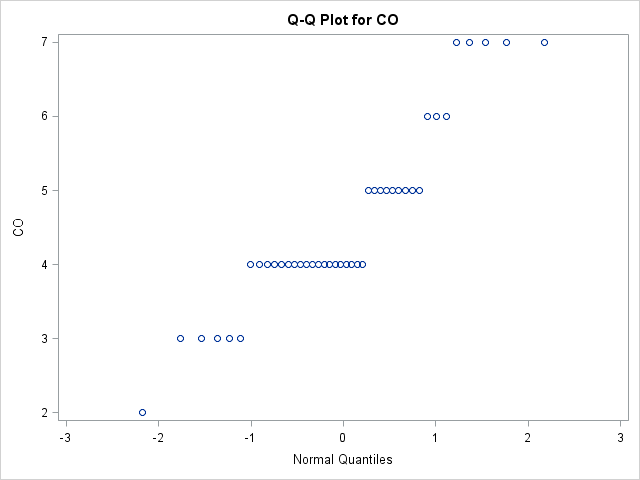
1. The various plots are below:



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.



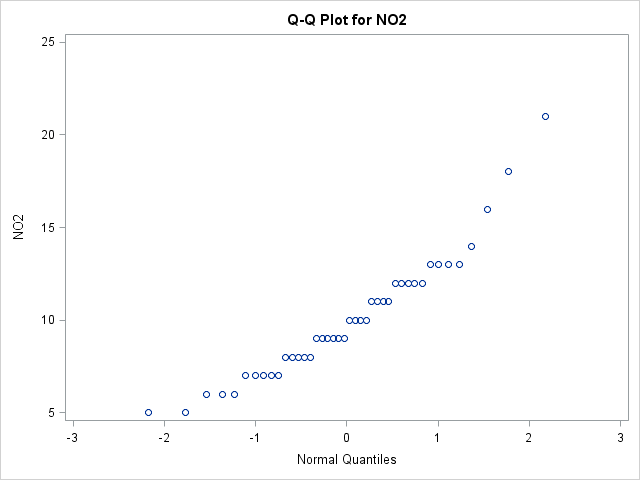
We also see that the p-value of the ryan joiner test < alpha (assume 0.05) and therefore we reject the null hypothesis of population normality.



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



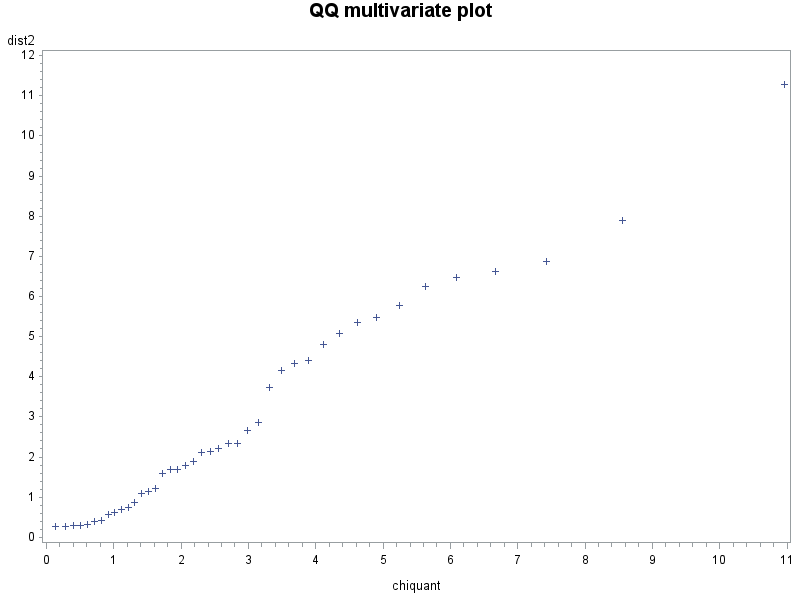
We also see that the p-value of the ryan joiner test > alpha (assume 0.05) and therefore we fail to reject the null hypothesis of population normality. We can assume normality



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.



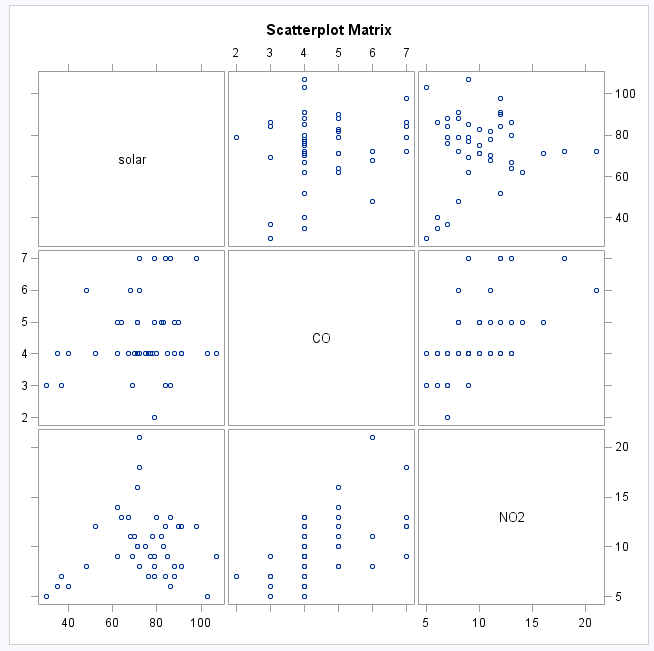
We also see that the p-value of the ryan joiner test < alpha (assume 0.05) and therefore we reject the null hypothesis of population normality.

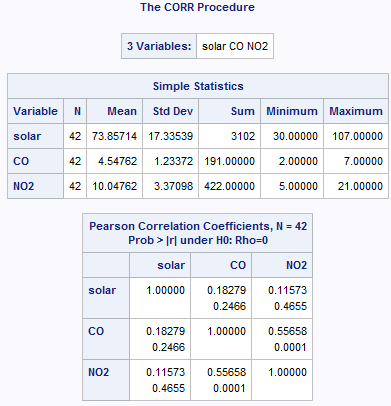


For multivariate data, we plot the ordered Mahalanobis distances versus estimated quantiles (percentiles) for a sample of size n from a chi-squared distribution with p degrees of freedom. Ideally this should resemble a straight-line for data from a multivariate normal distribution.

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:**

For significance of correlation coefficients, since the number of correlations tested is 3, and we want a 95% confidence level (alpha = 0.025 for 2-tailed test), the alpha-level for the Bonferroni-adjusted confidence interval is

alpha` = alpha / 3 = 0.025 / 3 = 0.0083

Now given cumulative probability p = 1-0.0083 = 0.9917, the z value = 2.394

**Between solar and CO, r = 0.****1828**

Step 1: Compute the Fisher transform:

= 1/2 \* LN((1+0.1828) / (1-0.1828)) = 0.1849

Step 2: Next, compute the 95% confidence interval for the Fisher transform

Zl = 0.1849 – (2.394 / sqrt(42-3)) = -0.1984

Zl = 0.1849 + (2.394 / sqrt(42-3)) = 0.5682

Step 3: Carry out the back-transform to obtain the 95% confidence interval for ρ12. This is shown in the expression below:

= ((exp(2\*-0.1984)-1) / (exp(2\*-0.1984)+1), (exp(2\*0.5682)-1)/(exp(2\*0.5682)+1))

= -0.1958, 0.5140

This yields the interval from -0.1958 to 0.5140

**Between solar and NO2, r = 0.1157**

Step 1: Compute the Fisher transform:

= 1/2 \* LN((1+0.1157) / (1-0.1157)) = 0.1162

Step 2: Next, compute the 95% confidence interval for the Fisher transform

Zl = 0.1162 – (2.394 / sqrt(42-3)) = -0.2671

Zl = 0.1162 + (2.394 / sqrt(42-3)) = 0.4995

Step 3: Carry out the back-transform to obtain the 95% confidence interval for ρ12. This is shown in the expression below:

= ((exp(2\*-0.2671)-1) / (exp(2\*-0.2671)+1), (exp(2\*0.4995)-1)/(exp(2\*0.4995)+1))

= -0.2609, 0.4617

This yields the interval from -0.2609 to 0.4617

**Between CO and NO2, r = 0.****5566**

Step 1: Compute the Fisher transform:

= 1/2 \* LN((1+0.5566) / (1-0.5566)) = 0.6279

Step 2: Next, compute the 95% confidence interval for the Fisher transform

Zl = 0.6279 – (2.394 / sqrt(42-3)) = 0.2446

Zl = 0.6279 + (2.394 / sqrt(42-3)) = 1.0112

Step 3: Carry out the back-transform to obtain the 95% confidence interval for ρ12. This is shown in the expression below:

= ((exp(2\*0.2446)-1) / (exp(2\*0.2446)+1), (exp(2\*1.0112)-1)/(exp(2\*1.0112)+1))

= 0.2398, 0.7663

This yields the interval from 0.2398 to 0.7663



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

1. The sample size is n = 11 and we want a 95% confidence interval for the population mean. Thus α = 0.05.
2. One at a Time Multiplier:

In Excel, the command =TINV(.05,10) will give the multiplier (value = 2.228). In SAS a command such as command t1=tinv(.975, 10)

With this notation a confidence interval for μj is computed as:



For oxygen, 34.636 ± (2.228 \* 10.452 / sqrt(11)) = 27.615 to 41.657

For solids, 33.182 ± (2.228 \* 19.073 / sqrt(11)) = 20.369 to 45.995

In this case, we can conclude that we are 95% confident that the interval (27.615 to 41.657) and (20.369 to 45.995) contains the means for oxygen and solids respectively. This strategy essentially considers each mean separately and uses the desired confidence level (usually 95%) for each single interval.

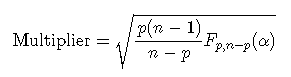
1. Bonferroni Method Multiplier: n = 11, we want family wide error = 5% (so family confidence = 95%) and we are computing intervals for p = 2 means. The error rate for each interval will be .05/2 = 2.5%. We might use the Excel command = TINV(.025, 10) to find that the multiplier = 2.634. In SAS, we use the cumulative probability = 1- (α /2p) so the command for finding the t-multiplier in this instance is something like t1=tinv(.9875, 10).

For oxygen, 34.636 ± (2.634 \* 10.452 / sqrt(11)) = 26.335 to 42.937

For solids, 33.182 ± (2.634 \* 19.073 / sqrt(11)) = 18.035 to 48.329

In this case, we can conclude that we are 95% confident that the interval (26.335 to 42.937) and (18.035 to 48.329) contains the means for oxygen and solids respectively. Here we set a family wide error rate and then divide this family error rate by the number of intervals to be computed to determine the error rate (and hence confidence level) for each individual interval.

1. Simultaneous Confidence Region Multiplier: we have a sample size of n = 11 and we have p = 2 variables. With a 5% family error rate (and 95% family confidence), the F-value can be found in Excel using = FINV(.05, 2, 9) = 4.256. SAS uses cumulative probabilities so in this case, a command like f1= FINV(.95, 2, 9) would make f1 be the F-value. The multiplier in this example is:



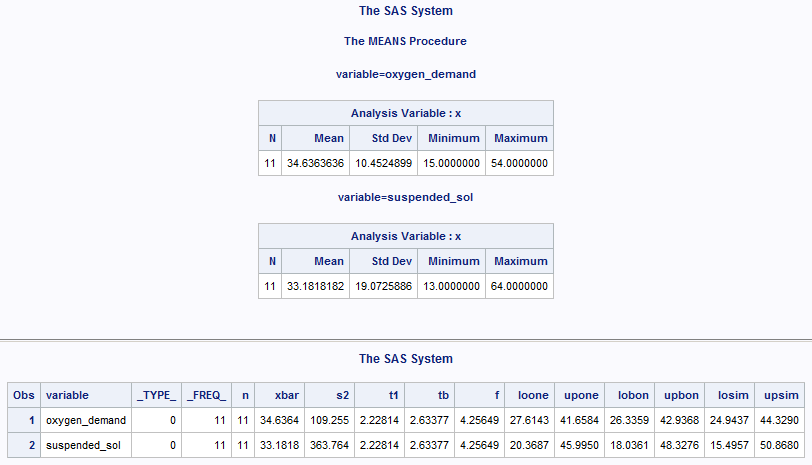
= sqrt(2\*10\*4.256 / 9) = 3.0754

For oxygen, 34.636 ± (3.0754 \* 10.452 / sqrt(11)) = 24.944 to 44.328

For solids, 33.182 ± (3.0754 \* 19.073 / sqrt(11)) = 15.496 to 50.868

In this case, we can conclude that we are 95% confident that the interval (24.944 to 44.328) and (15.496 to 50.868) contains the means for oxygen and solids respectively. Here we use properties of the multivariate normal distribution to define joint confidence intervals. The multiplier for this method is conservative.

Using SAS, we get the following matching outputs:



1. The correlation between X1 and X2 is = 120.37/(sqrt(109.25) \* sqrt(363.76)) = 0.6038 = r12
2. Test H0 : ρ = 0 against Ha : ρ ≠ 0 at the alpha = 0.01 level

Compute the test-statistic: 

t = 0.6038 \* sqrt((11-2) / (1-0.6038^2)) = 2.2724

Next we get a critical value of t(df, 1-α/2) = t(9, 0.995) = 3.250

Finally since 2.2724 < 3.250 we fail to reject the null hypothesis that Y1 and Y2 are uncorrelated at the α < 0.01 level.

0.0492

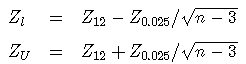
Our conclusion here is that: Oxygen and solids are uncorrelated (t = 2.2724; d.f. = 9; p > 0.01)

1. Step 1: Compute the Fisher transform:



= 1/2 \* LN((1+0.6038) / (1-0.6038)) = 0.6991

Step 2: Next, compute the 95% confidence interval for the Fisher transform



Zl = 0.6991 – (1.96 / sqrt(11-3)) = 0.0061

Zl = 0.6991 + (1.96 / sqrt(11-3)) = 1.3921

Step 3: Carry out the back-transform to obtain the 95% confidence interval for ρ12. This is shown in the expression below:



= ((exp(2\*0.0061)-1) / (exp(2\*0.0061)+1), (exp(2\*1.3921)-1)/(exp(2\*1.3921)+1))

= 0.006099924, 0.883632052

This yields the interval from 0.0061 to 0.8836

Conclusion: In this case, we can conclude that we are 95% confident that the interval (0.0061 to 0.8836) contains the correlation between oxygen and solids.