MANIPAL
ACADEMY *of* HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)

Module (Lecture 7)

Assessing Multivariate Normality

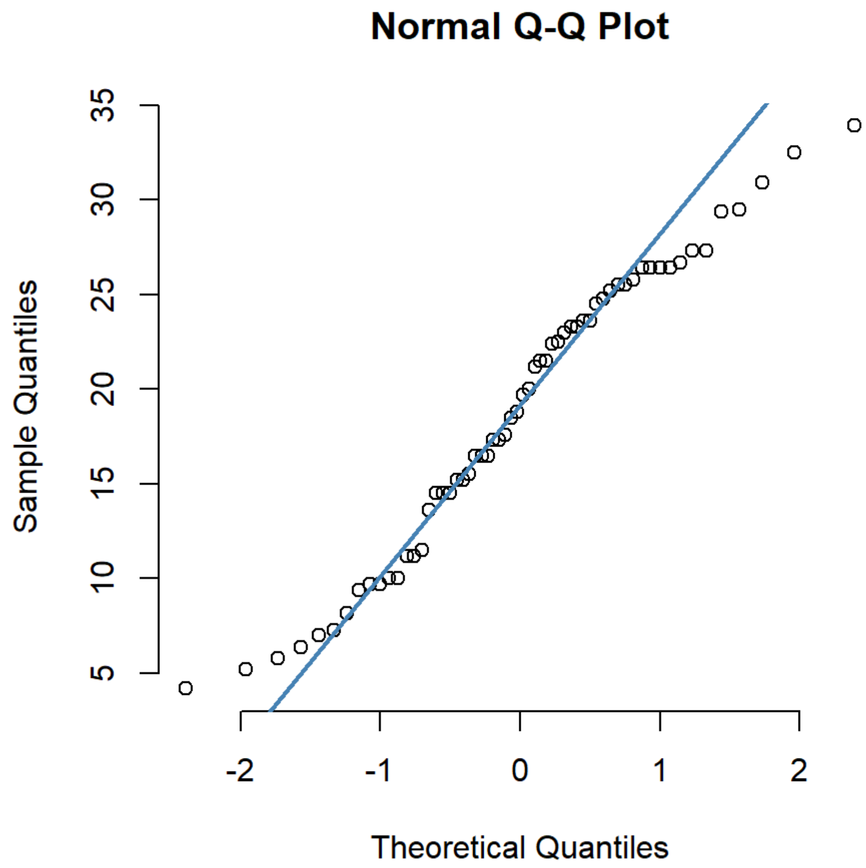
David Raj Micheal

August, 2018

Assessing Multivariate Normallity

Q-Q plots

- Plots are always useful devices in any data analysis.
- Special plots called Q-Q plots can be used to assess theassumption of normality.



Example of a Q-Q plot

MANIPAL ACADEMY of HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)
INSPIRED BY LIFE

Constructing a Q-Q plot (univariate case)

Let $y_1, y_2, \ldots, y_n$ represent n observations on a random variable x.

- Order the observations so that

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$$

- $y_{(j)}$'s are the sample quantiles
- When $y_{(j)}$'s are distinct, exactly j observations are less than or equal to $y_{(j)}$.
- The proportion j/n of the sample at or to the left of $y_{(j)}$ is often approximated by $\frac{\left(j-\frac{1}{2}\right)}{n}$ for analytical convenience.

MANIPAL
ACADEMY *of* HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)
INSPIRED BY LIFE

- For the standard noraml distribtuion, the quantiles $q_{(j)}$ are defined by the relation

$$P[z \leq q_{(j)}] = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dz = p_{(j)} = \frac{j - \frac{1}{2}}{n}$$

where $P_{(j)}$ is the probablity of getting a value less than or equal to $q_{(j)}$ in a single drawing from a standard normal population.

- The idea is to look at the pairs of quantiles $(q_{(j)}, y_{(j)})$ with the same associated cumulative probability $\frac{j-\frac{1}{2}}{n}$.
- If the data arise from a normal population,the pairs $(q_{(j)}, y_{(j)})$ will be approximately linearly related, since $\sigma q_{(j)} + \mu$ is nearly the expected sample quantile.

Example: Constructing Q-Q plot

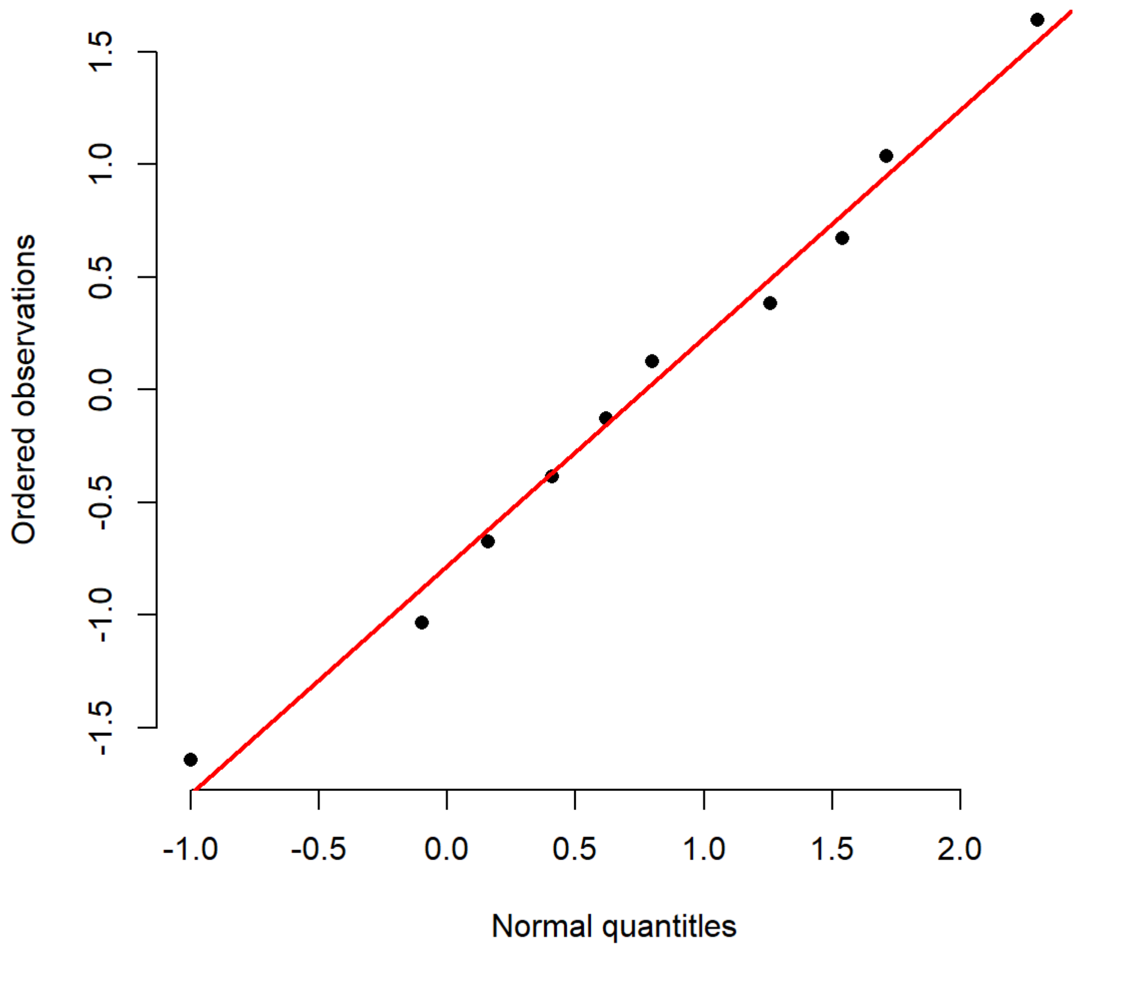
A sample of $n = 10$ observations gives the values in the following table:

Ordered Observations $y_{(j)}$	Probability Levels $\frac{j-1/2}{n}$	Standard Normal quantiles $q_{(j)}$
-1.00	0.05	-1.64
-0.10	0.15	-1.04
0.16	0.25	-0.67
0.41	0.35	-0.39
0.62	0.45	-0.13
0.80	0.55	0.13
1.26	0.65	0.39
1.54	0.75	0.67
1.71	0.85	1.04
2.30	0.95	1.64

- The Q-Q plot for the forgoing data, which is a plot of the ordered data $x_{(j)}$ against the noramal quantile $q_{(j)}$.

```
plot(qqplottab$qnorm.problevel.~qqplottab$ordered,
     pch = 16, frame.plot = F,
     xlab = "$x_{(j)}$",
     ylab = "Normal quantitles")
abline(lm(qqplottab$qnorm.problevel.~qqplottab$ordered), col='red', lwd = 2)
```

Q-Q Plot for the foregoing data



MANIPAL ACADEMY of HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)
INSPIRED BY LIFE

Q-Q Plots for multivariate data

- Find the squared mahalanobis distances

$$d_j^2 = (y_j - \overline{x})^T S^{-1} (y_j - \overline{x}), \quad j = 1, 2, \ldots, n$$

  where y_j is the jth oberservation vector.

- Order the squared mahalabis distances from smallest to largest as

$$d_{(1)}^2 \leq d_{(2)}^2 \leq \cdots \leq d_{(n)}^2.$$

- Graph the pairs $\left(\chi_p^2\left((n - j + \frac{1}{2})/n\right), d_{(j)}^2\right)$

Conclusion: The plot should resemble a straight line through the origin having slope 1.

Example: Constructing Q-Q plot

```r
set.seed(100)
data = data.frame("Height" = sample(45:65, size = 100, replace = TRUE),
                  "Weight" = sample(145:175, size = 100, replace = TRUE),
                  "SBP"    = rnorm(100, 130, 10),
                  "DBP"    = rnorm(100, 90, 10)
                  )
head(data)
```

Height	Weight	SBP	DBP
51	155	127	90
50	157	144	86
56	146	125	99
46	156	138	95
54	162	115	100
55	166	126	80

Find Mahalanobis Distance (Sample Quantile)

```
Stat.dist = mahalanobis(data, center = colMeans(data), cov = cov(data))
Stat.dist = sort(Stat.dist)
Stat.dist
```

```
##   [1]  0.41  0.60  0.78  0.89  0.92  0.94  1.02  1.02  1.05  1.23  1.32
##  [12]  1.33  1.37  1.38  1.55  1.65  1.70  1.74  1.74  1.83  2.10  2.11
##  [23]  2.12  2.14  2.15  2.20  2.23  2.24  2.32  2.40  2.42  2.46  2.52
##  [34]  2.56  2.63  2.69  2.73  2.78  2.87  2.88  2.91  2.98  2.99  3.02
##  [45]  3.05  3.08  3.21  3.22  3.23  3.28  3.36  3.37  3.51  3.56  3.59
##  [56]  3.60  3.65  3.65  3.81  4.10  4.12  4.20  4.26  4.30  4.38  4.40
##  [67]  4.44  4.47  4.62  4.71  4.85  4.88  5.02  5.14  5.17  5.19  5.24
##  [78]  5.24  5.27  5.39  5.40  5.63  5.75  6.11  6.16  6.44  6.50  6.55
##  [89]  6.61  6.93  7.46  7.48  7.79  8.49  8.49  8.91  9.23 12.81 13.73
## [100] 14.07
```

MANIPAL ACADEMY of HIGHER EDUCATION
(Deemed to be University under Section 3 of the UGC Act, 1956)
INSPIRED BY LIFE

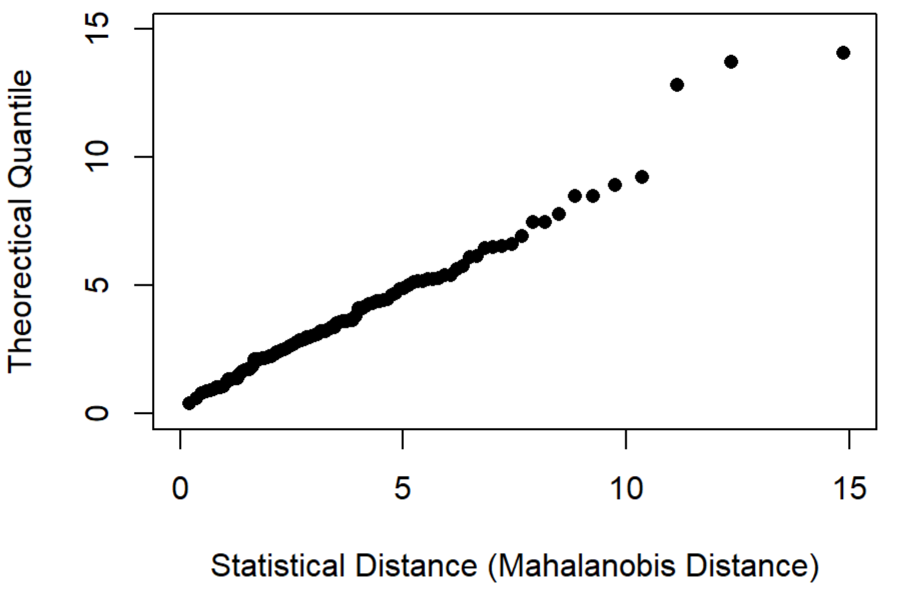
Find the theoretical quantile value

```r
theo.quant = NULL
for (j in 1:nrow(data)){
n = nrow(data)
k = (n-j+1/2)/n
theo.quant[j] = qchisq(k,df = 4,lower.tail = FALSE)
}
theo.quant
```

```
##   [1]  0.21  0.37  0.48  0.58  0.67  0.75  0.83  0.90  0.97  1.03  1.10
##  [12]  1.16  1.22  1.28  1.34  1.40  1.45  1.51  1.57  1.62  1.68  1.73
##  [23]  1.79  1.84  1.90  1.95  2.00  2.06  2.11  2.17  2.22  2.28  2.33
##  [34]  2.39  2.44  2.50  2.55  2.61  2.67  2.72  2.78  2.84  2.90  2.96
##  [45]  3.02  3.08  3.14  3.20  3.26  3.32  3.39  3.45  3.52  3.59  3.65
##  [56]  3.72  3.79  3.86  3.93  4.01  4.08  4.16  4.24  4.32  4.40  4.48
##  [67]  4.56  4.65  4.74  4.83  4.93  5.02  5.12  5.22  5.33  5.44  5.55
##  [78]  5.67  5.79  5.92  6.06  6.20  6.34  6.50  6.66  6.83  7.02  7.21
##  [89]  7.43  7.66  7.91  8.19  8.50  8.85  9.26  9.74 10.35 11.14 12.34
## [100] 14.86
```

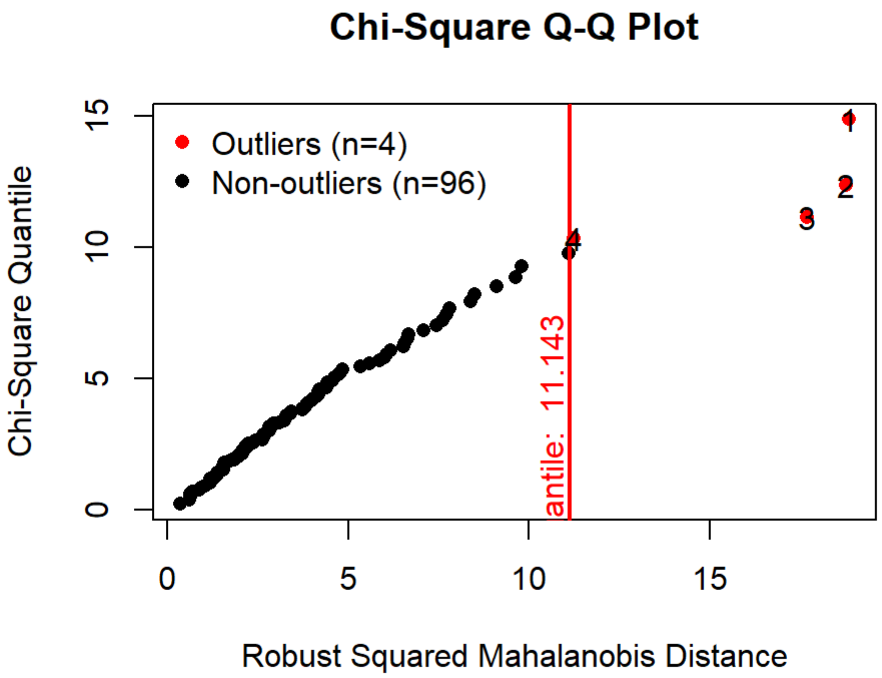
Plot Statistical Distance Vs Theoretical Quantile

```
plot(Stat.dist~theo.quant,
     xlim = c(0,15), ylim = c(0,15),
     ylab = "Theorectical Quantile",
     xlab = "Statistical Distance (Mahalanobis Distance)",
     pch = 16)
```



How to plot Q-Q plot using R?

```
library(MVN)
mvn(data = data,mvnTest = "hz", multivariateOutlierMethod = "quan")
```




```
## $multivariateNormality
##            Test  HZ p value MVN
## 1 Henze-Zirkler 1.2   0.001  NO
##
```

MANIPAL ACADEMY of HIGHER EDUCATION (Deemed to be University under Section 3 of the UGC Act, 1956)