Contents

1 Quantile-quantile plots

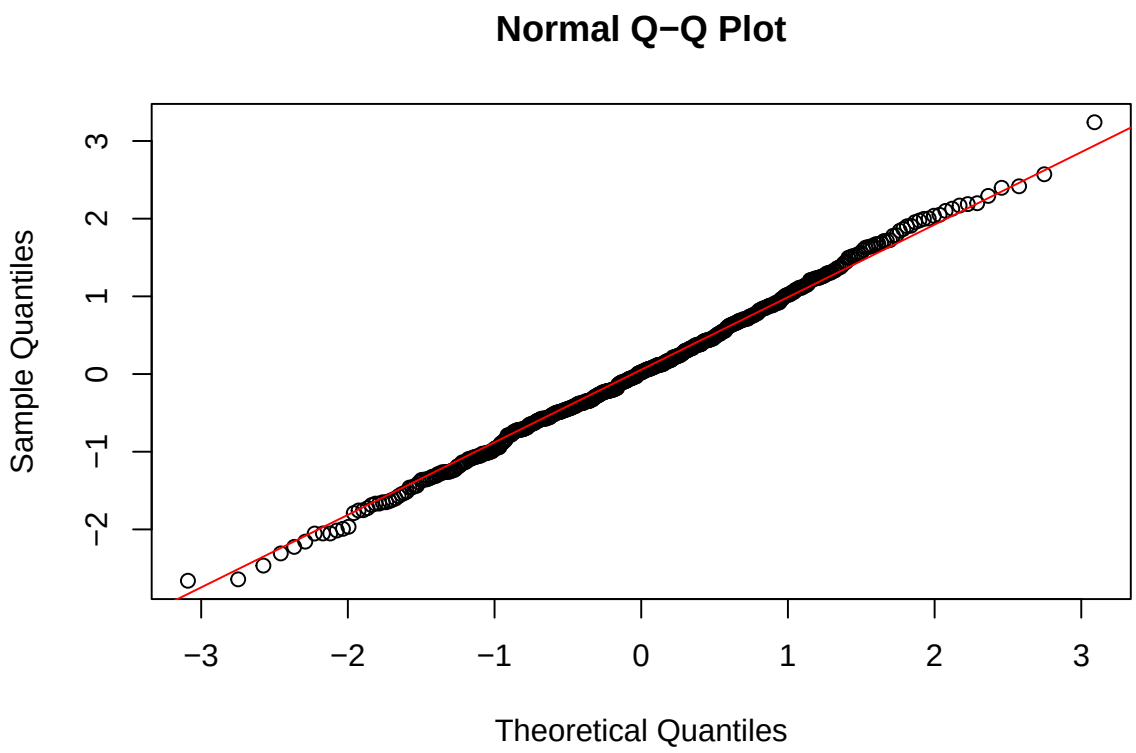
It might be helpful to understand how quantile-quantile plots (qqplots) are created, and their purpose.

Generally, the purpose of a qqplot is to compare the distribution of two variables, *usually* with reference to a theoretical distribution. When a theoretical distribution is not available, one may rely on an empirical distribution (based on data).

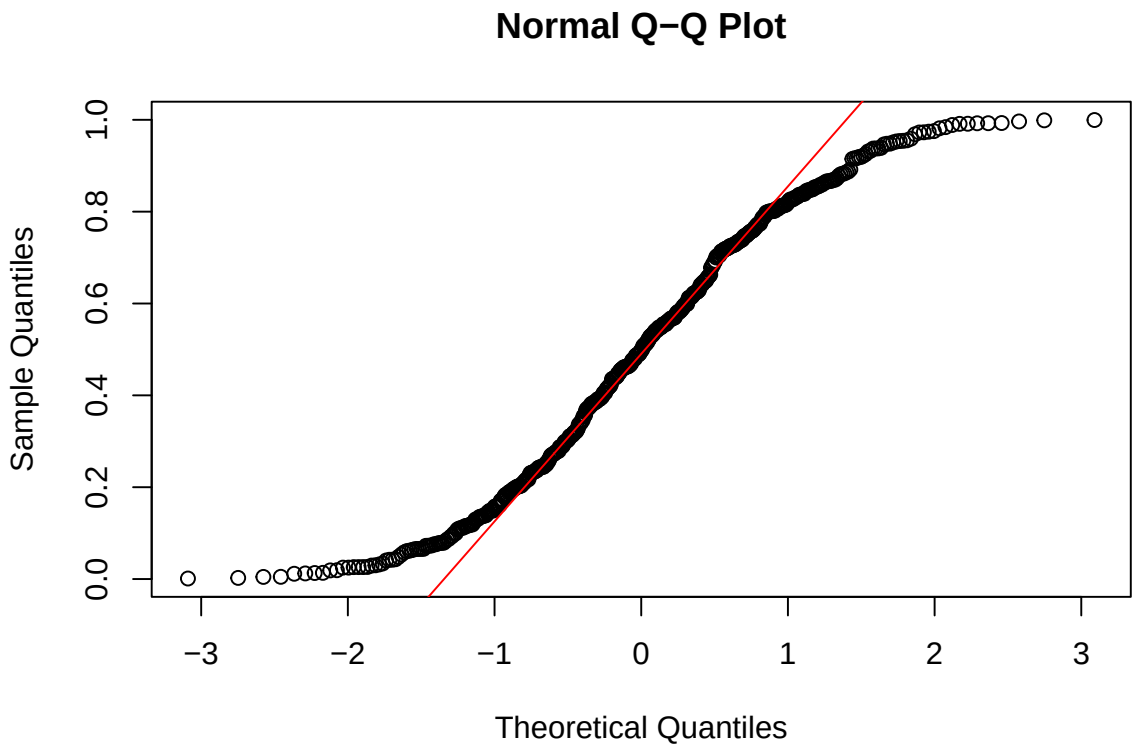
An example: we have a set of data, which we want to check follows the normal distribution. The `qqnorm()` function helps us do that. It automatically arranges the data for plotting.

The `qqline()` function is used to draw a line through the first and third quantiles, which uses the `qnorm()` distribution by default. It can take any other quantile function (e.g. `qunif()`, `qchisq()`, `qexp()`) with a change in the `distribution` parameter.

```r
set.seed(123)
x.sample <- rnorm(500)
qqnorm(x.sample)
qqline(x.sample, col="red")
```



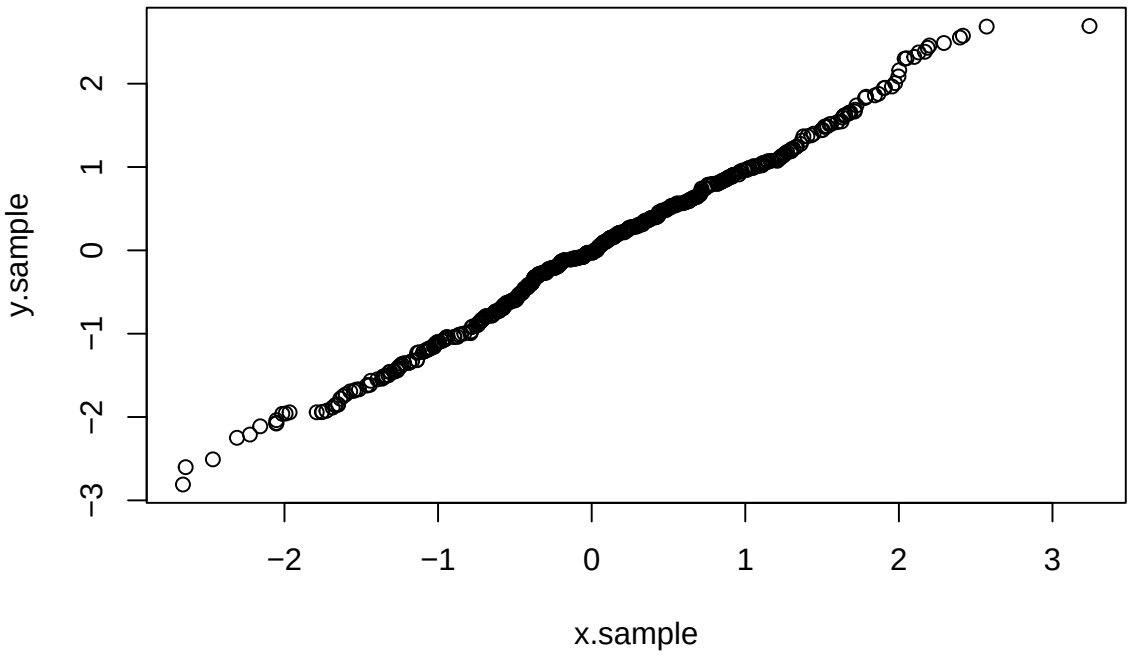

```r
x.sample <- runif(500)
qqnorm(x.sample)
qqline(x.sample, col="red")
```



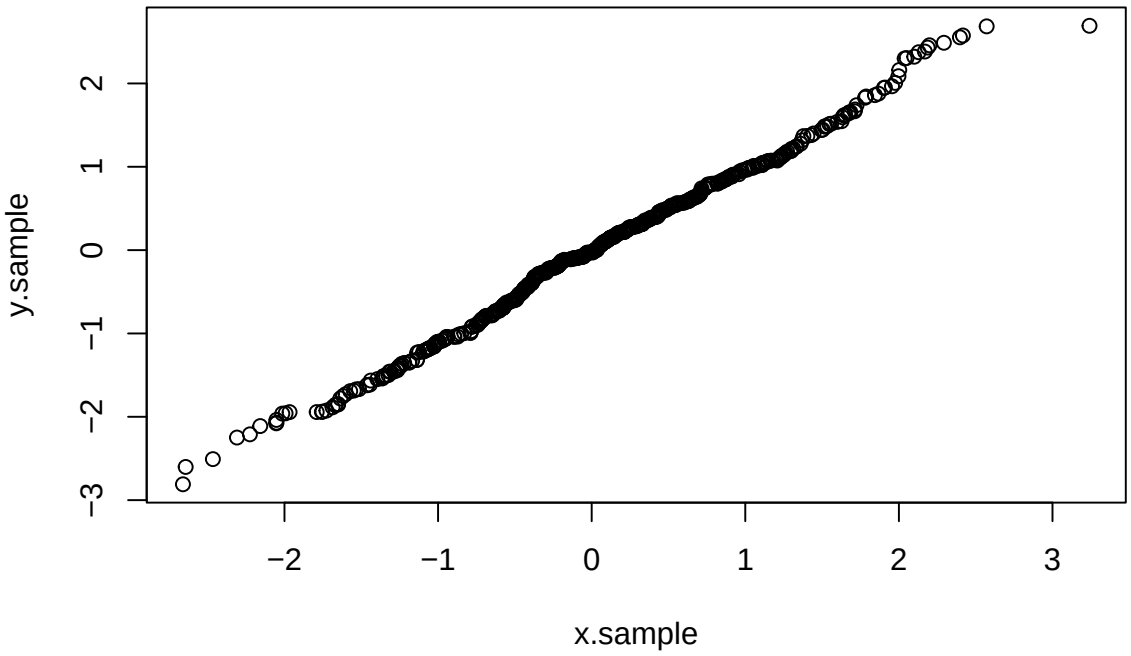
Second example: we have two sets of data, which we want to compare have the same distribution. Assume we don't have a theoretical distribution to compare it with. In this case, we use the `qqplot()` function. The base `plot()` function also works. The data must be arranged in ascending order, here it is done with `sort()`.

`qqline()` is harder to use in this case, as it requires a distribution function. The qq-line just draws a line between the 25th and 75th quantiles, though, so it can be manually calculated and implemented.

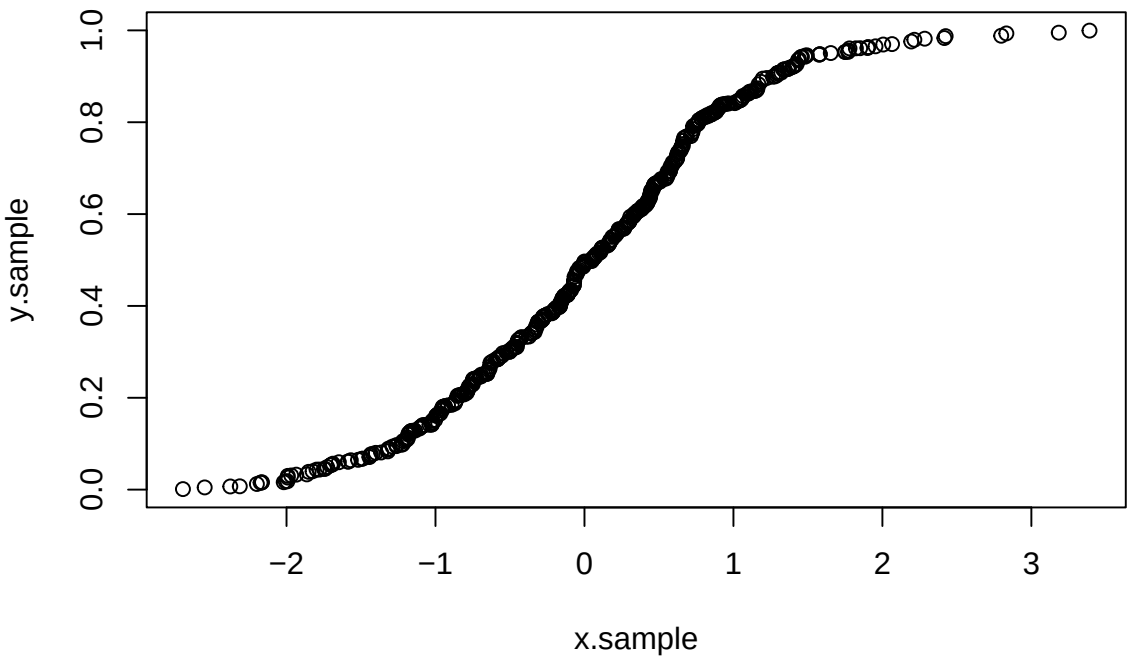
```r
set.seed(123)
x.sample <- sort(rnorm(500))
y.sample <- sort(rnorm(500))
plot(x.sample, y.sample)
```



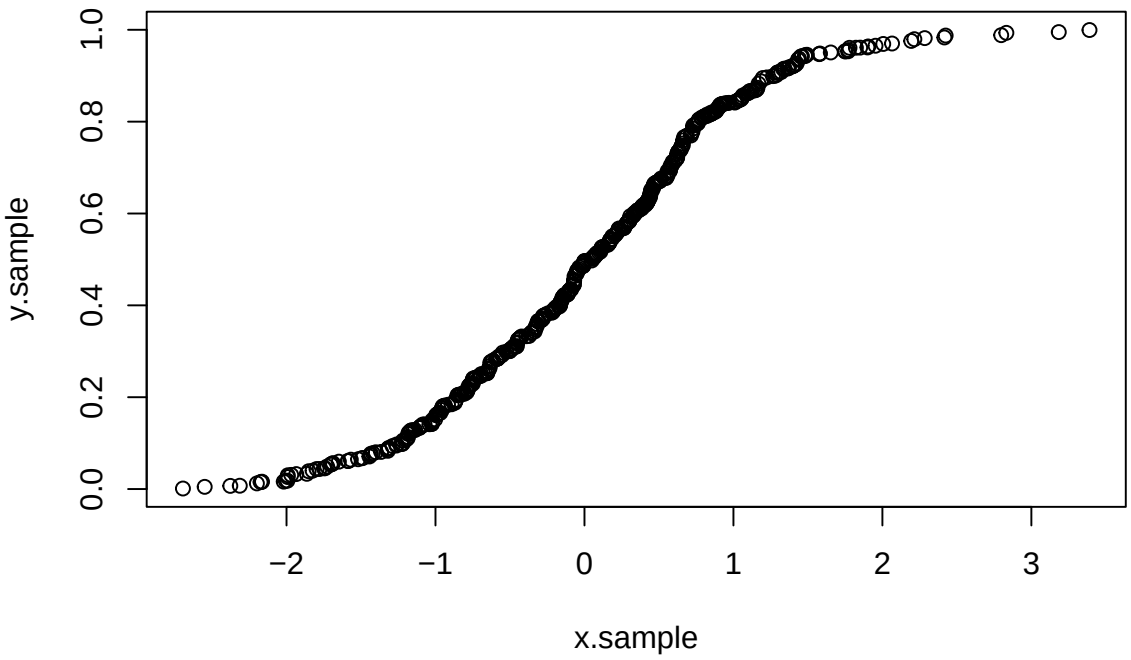

```r
qqplot(x.sample, y.sample)
```



```r
x.sample <- sort(rnorm(500))
y.sample <- sort(runif(500))
plot(x.sample, y.sample)
```




```r
qqplot(x.sample, y.sample)
```



Once you know how the quantile-quantile plot is created, interpreting it should come more naturally.

1.1 Interpretation

The quantile-quantile plot gives us a quick view of how a set of values is distributed with reference to another distribution.

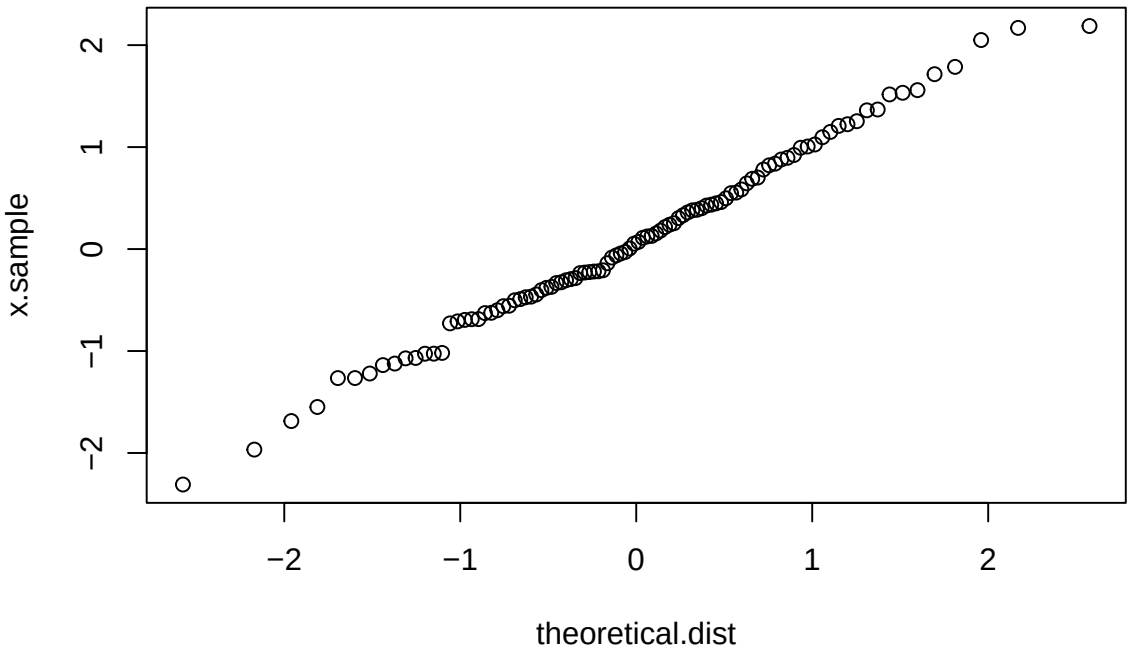
If the two distributions are similar, then the plot should have a nice line forming.

1.2 Optional: theoretical distribution without `qqplot()`/`qqnorm()`

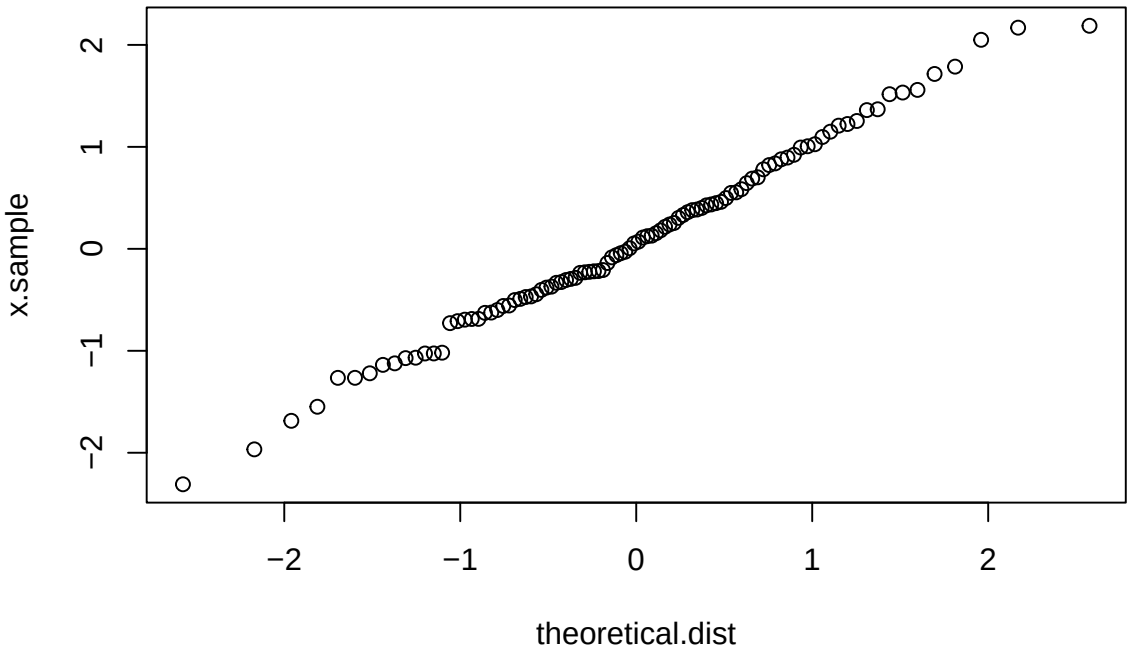
We need to get the quantiles of the theoretical distribution. Thus, we space out numbers evenly in the interval $(0, 1)$, acting as the probabilities. Importantly, we avoid $p = 0$ and $p = 1$ to avoid extremities, e.g. `qnorm(0) = -Inf` and `qnorm(1) = +Inf`.

```r
set.seed(123)
x.sample <- sort(rnorm(100))
p <- ppoints(100)
theoretical.dist <- qnorm(p)

# All three create the same plot.
plot(theoretical.dist, x.sample)
```




```r
qqplot(theoretical.dist, x.sample)
```



```r
qqnorm(x.sample)
```