Shapiro Wilk test can be too strict

Saket Choudhary

This notebook demonstrates that the ShapiroWilk test can give unreliable p values and suggests qqplots as a more reliable estimate.

`shapiro.test(x)` in R uses the following null and alternative hypothesis:

$$H_0 : \text{Input distribution x is normally distributed}$$

$$H_A : \text{Input distribution x is not normally distributed}$$

Ideally you would reject the null when the p-value of `shapiro.test` is say < 0.05. However, the following simulation demonstrates that it can be too strict.

Simulation

We will generate normal random variables for 100 iterations. In each iteration we generate 5000 normal random variables.

We introduce some noise in the data by adding 1 to 10% of the data points. This is done by `+c(1,0,0,2,1)` so that the vector `c(1,0,0,2,1)` gets added to every five entries.

```
set.seed(420)
n5000 <- replicate(1000, {
    c(shapiro.test(rnorm(5000)+c(1,0,0,2,1))$p.value)
    })
```

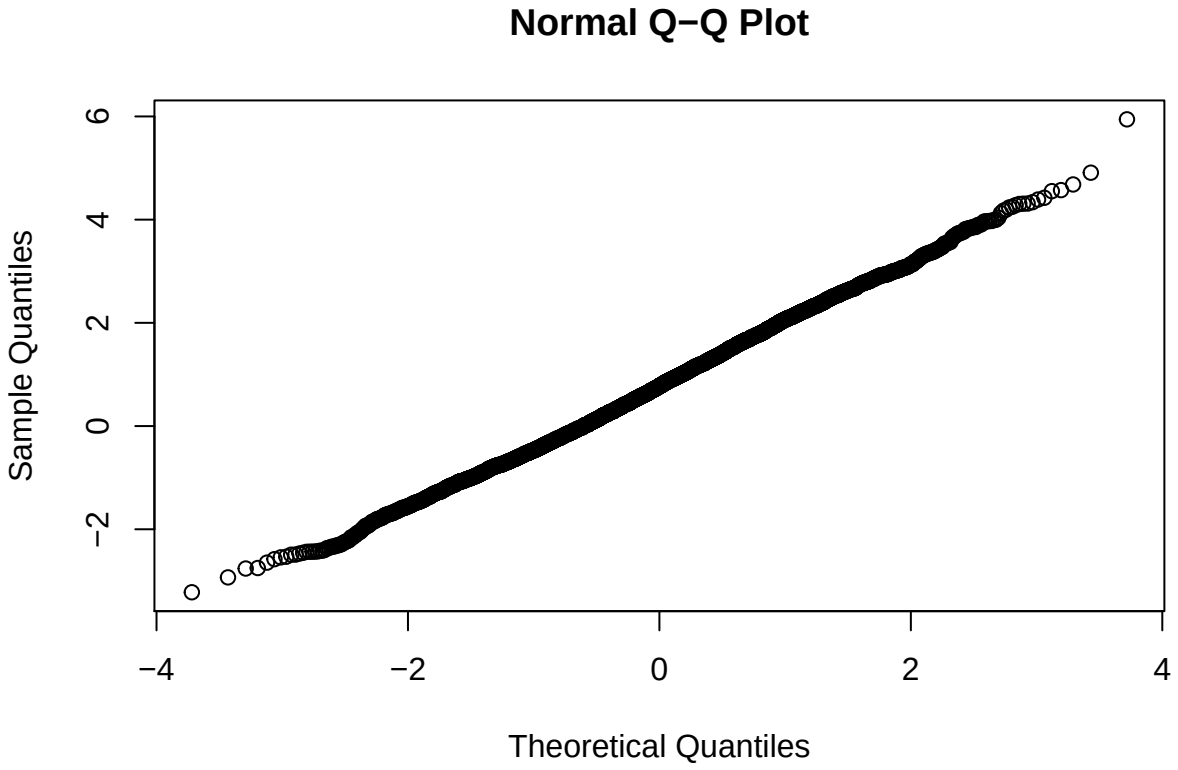
We now calculate the proportion of tests that were rejected on a threshold of 0.05:

```
sum(n5000<0.05)/5000
```

```
## [1] 0.1598
```

So around 15% of `rnorm(5000)` samples with just three entries slightly modified will cause the `shapiro.test` to fail while the qqplot looks normal:

```
qqnorm(rnorm(5000)+c(1,0,0,2,1))
```



A visual inspection of QQplot might often be taken as a proof for approximate normality. Approximate normality is sufficient for t-test.