Testing for normality

*Owen Petchey*

1/19/2017

Introduction

We often want to know if our data or our residuals are not normally distributed. You've now done a lot with QQ-plots to examine how normally distributed are your data. Some people like to also have a statistical test for normality. Here I'll show some tests.

Shapiro-Wilk test on normally distributed data

The Shapiro-Wilk test is frequently used. The null hypothesis is that the data are normally normally distributed; the alternate is that the data are not normally distributed.

Lets make some data and look at its distribution and QQ-plot:

```
x <- rnorm(100)
qplot(x, bins=10)
qqnorm(x); qqline(x)
```
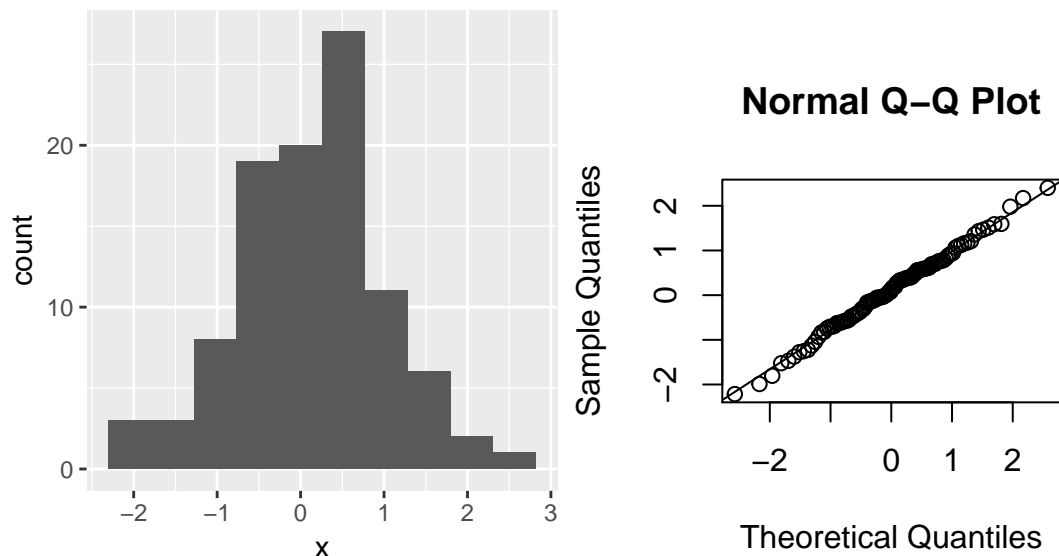

Looks pretty normal. Lets see what the Shapiro-Wilk test tells us:

```
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.9956, p-value = 0.9876
```

The p-value for the test is 0.9876. When we get a large p-value like this, it means we cannot reject the null hypothesis. (A small p-value, say p < 0.05, would cause us to reject the null hypothesis). Hence, here, we

cannot reject the null hypothesis that the data are normally distributed. (Note that, as usual, technically we cannot accept the null hypothesis.) So we can happily continue with any work that assumes normally distributed residuals.

Shapiro-Wilk test on non-normally distributed data

```r
x <- runif(100)
qplot(x, bins=10)
qqnorm(x); qqline(x)
```



```r
shapiro.test(x)
```
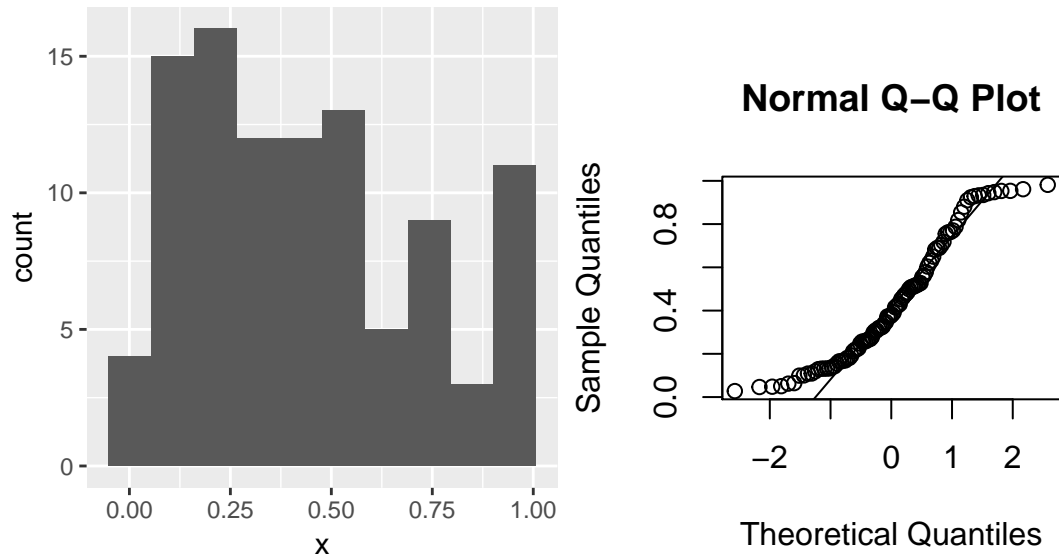
```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.93092, p-value = 5.616e-05
```

Now the p-value is small, so we reject the null hypothesis, that the data is normally distributed. Now we can't continue with any methods, such as linear models, that assume normally distributed residuals.

What about if we have fewer data point, say 30? And lets do the test lots of times and count how often we reject the null hypothesis:

```r
x <- replicate(1000, shapiro.test(runif(30))$p.value)
sum(x<0.05)
```

```
## [1] 372
```

So we reject the null in 38.8% of the trials. That is, we don't reject the hypothesis that the data is normally distributed in nearly 40% of cases, even though it came from a uniform distribution!

Shapiro-Wilk on real data

How useful this (or any) test of normality is, depends on our data. If we have a small sample the we could easily fail to reject the null hypothesis even when the generating process does not create normally distributed data. The test has low power for small samples.

If we have large samples, on the other hand, the test has very large power, and we can easily reject the null hypothesis due to tiny deviations from normality.

Here's an example of the last point. We make data that has 5000 numbers sampled from a normal distribution, and add a bit of non-normality:

```r
x <- rnorm(5000)+c(1,0,2,0,1)
```

Now lets look at the distribution and QQ-plot:

```r
qplot(x, bins=30)
qqnorm(x); qqline(x)
```


Looks fine.
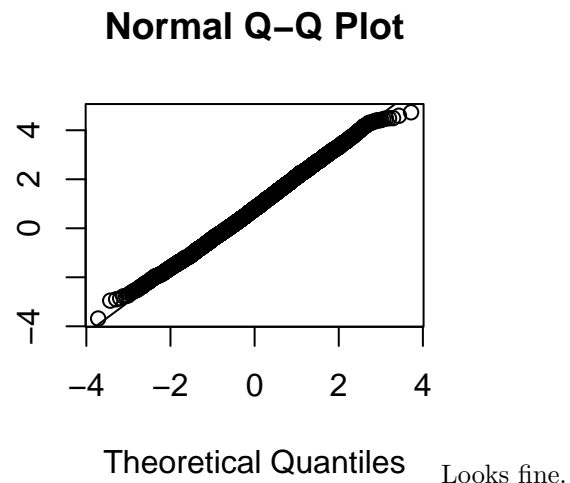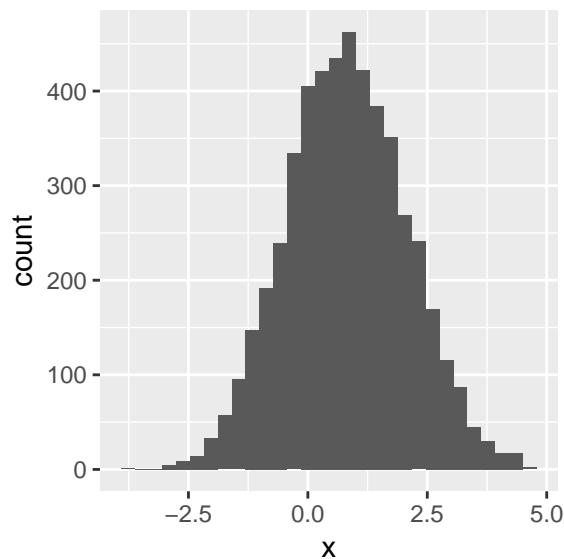
And what does the test tell us?

```r
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.99909, p-value = 0.008756
```

Wow! The test soundly rejects the null hypothesis. We can do this 1000 times and see how often the test rejects the null.

```r
x <- replicate(1000, shapiro.test( rnorm(5000)+c(1,0,2,0,1) )$p.value)
sum(x<0.05)
```

```
## [1] 786
```

So in about 78% of trials, the test rejects the null hypothesis that the data is normally distributed, even though its really very close to normally distributed.

So what do we do?

Actually, we might be focusing on the wrong question. We have been asking how close to normally distributed is our data. Actually, we should probably be more concerned with how robust (or sensitive) is our test to deviations from normality, and what types of deviations are more important. Generally speaking, linear models are really quite robust to deviations from normality, so we can be relatively happy that the Shapiro-Wilk test lacks power with small samples (and so will allow us to assume normality even when in reality this might not be justified) and need to be cautious about its rejection of normality when sample size is large.

There is lots of information online regarding this, and surrounding issues. This web page has a particularly interesting discussion: Is normality testing 'essentially useless'?. Enjoy, and of course discuss as you please on the Discussion forum.