Normal distribution assumption review

What is the assumption of normal data?

Why is it important?

How do you asses whether or not data are normal?

Statistical tests

Generally not good because we can't prove H0.

Graphical methods:

Histograms
Boxplots

But by far the best is a QQ plot (or normal probability plot).

Make sure you know how to construct a QQ plot.

Make sure you know all the steps - if you're given a small data set of 3 or 4 points, you should be able to construct a QQ plot!

(Note that the exact steps are a little bit different depending on whether or not you use the fourth edition - just be consistent!)

Be able to interpret a QQ plot and figure out:

1) If the data are approximately normal

2) If the data are not normal, what the problem is:

Long tails? Short tails? Skewed right? Left? Etc.

Also be able to determine how serious the problem is.

(E.g., long tails are really bad!).

Understand all the assumptions of a *t*-test:

1) Random data

Independence between columns

2) Normally distributed data (in each sample).

3) Variances equal **or** not equal ($\sigma^2_1 \neq \sigma^2_2$ **or** $\sigma^2_1 = \sigma^2_2$).