From “Probability plotting methods for the analysis of data”

* ECDF
  + Order of observations is immaterial
  + No classification of observations, based on extraneous considerations
  + If sample is non-random, appropriate weights are specified
  + Does not depend on any assumption of a parametric distributional specification
* Q-Q Plots (Quantile Plots)
  + If x & y are identically distributed variables, then the plot of x-quantiles vs y-quantiles will be a straight line configuration with slope 1
  + If y is a linear function of x, the Q-Q plot will still be linear but with possibly changed location and slope.
  + Can serve as a valuable stimulus for a variety of statistical procedures.
    - Ex. K-S stat may be viewed, after the appropriate probability transform of the sample, as the maximum deviation from the 45 degree line on a uniform P-P plot (which for the uniform distribution on (0,1) is equivalent to a Q-Q plot)
    - Ex. regression of order statistics on expected values of standard order statistics in a Q-Q plot to generate test procedures for composite distributional hypotheses.
  + Probability plots as informal aids to inference
    - Procedures of probability plotting may also be uservully employed in connexion with the complex objectives generally associated w/analysis of variance
    - Example of non-obvious, interesting indications concerning the structure of the data are the presence of possible real effects, existence of outliers, distributional peculiarities, etc. The real applied value of the analysis of variance table is as a summary of patterns of variability. It provides a collection of mean squares each associated with an identifiable facet of the experiment, and it will often be true that appropriate subsets of these are meaningfully comparable.
    - Some probability plotting techniques which have been proposed as internal comparison methods for the analysis of variance will be discussed in the sequel.
    - The presence of real effects, the existence of distributional peculiarities, of outliers, and of heterogeneities of variance result in distortions of the linear configuration of the plot.
    - The graphical summary provided by the plot focuses attention on the large effects and groupings among them, if any, and moreover does this in a simple and palatable fashion.
  + Specific internal comparisons probability plotting techniques
  + Plotting residuals from regression analysis
    - May be used to check the adequacy of the [regression] model, the appropriateness of the independent variables, the existence of outliers, the relevance of extraneous variables, and on distributional peculiarities.
  + Conclusion
    - Informal tools for the statistical analysis of data; may be employed for describing and summarizing as well as for uncovering and understanding the structure underlying a body of data.

From “Quantifying the effects of social influence”

* we check normality of errors by plotting the quantiles of the residual distribution against the quantiles of a normal distribution. The off-diagonal points in all questions clearly indicate the presence of a few large outliers, as expected for skewed data

From “STATISTICAL METHODS FOR IDENTIFYING

DIFFERENTIALLY EXPRESSED GENES IN REPLICATED

cDNA MICROARRAY EXPERIMENTS”

* In general, Q-Q plots are used to assess whether data have a particular distribution or whether two datasets have the same distribution. In our application, we are not so much interested in testing whether the test statistics follow a particular distribution, but in using the Q-Q plots as a visual aid for identifying genes with “unusual” test statistics.
* Q-Q plots informally correct for the large number of comparisons and the points which deviate markedly from an otherwise linear relationship are likely to correspond to those genes whose expression levels differ between the control and treatment groups.
* In a normal Q-Q plot, the ordered test statistics are plotted against the quantiles of a standard normal distribution. Alternatively, Q-Q plots may be obtained by plotting the ordered tests statistics against quantiles estimated from the permutation distribution of these test statistics.