PROBABILITY PLOTS

How to judge whether data come from a given distribution?

Histograms have problems.

Probability plots (AKA *Q-Q plots* or *quantile plots*):

- Order the data $y_1 \le y_2 \le \ldots \le y_n$.
- Graph them vs. $q_1, \le q_2 \le \ldots \le q_n,$

where

q_k = the expected value (as approximated by computer) of the kth smallest member of a simple random sample of size n from the "test distribution".
(i.e., q_k = the expected value of Y_k, if Y has the conjectured distribution)

If the data come from this distribution, we expect $y_k \approx q_k$, so the graph will lie approximately along the line y = x.

Variation often used to test for normality:

Take the q_k's from the *standard normal* distribution.

If the y_k's are sampled from an $N(\mu,\sigma)$ distribution, then the transformed (standardized) data $\frac{y_k - \mu}{\sigma}$ come from a standard normal distribution, so we expect

$$\frac{y_k - \mu}{\sigma} \approx q_k$$

i.e., if the y_k's are sampled from an $N(\mu,\sigma)$ distribution, then

$$y_k \approx \sigma\, q_k + \mu,$$

so the graph should lie approximately on a straight line with slope and intercept σ and μ, respectively.