STAT 400 Data Analysis September 12, 2005

Today: Q-Q plots p 21
Box Plots p 25
Normal Q-Q Plots p 28
Fits and Residuals p 34
Stereogram Example p 42

Q-Q Plots

Usually for comparing commensurate distributions.

With two variables, just plot as a scatter plot of (x,y) the x and y that correspond to approximately the same quantile. If the sample sizes are unequal, find the number of (x,y) values equal to the smaller sample size.

First try to compare the quantile plots for Tenor 1 and Bass 2 (p 16). Then look at Fig 2.3 on p 22. Is it helpful?

How would Fig 2.3 compare with a scatter plot of bass2 vs tenor1?

Important to understand this to appreciate the Q-Q plot!

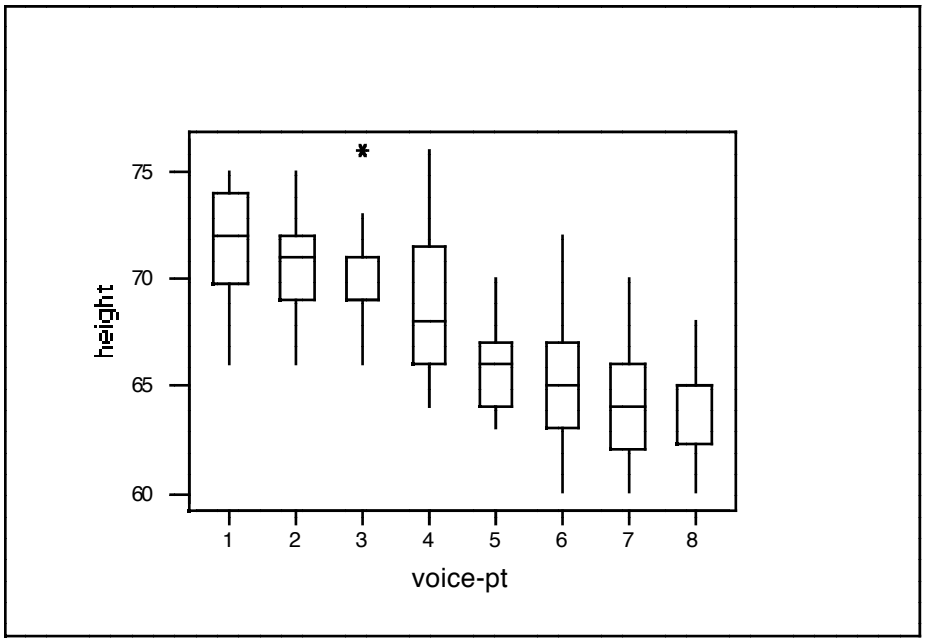
m-d plot

This is an estimate of how much the two distributions are shifted at various quantiles. (Actually, at the means of equal quantiles of the two distributions.)
Fig 2.4 shows that the shift is smaller at the higher quantiles. Cf last para p 23.

Q-Q plots playing role of matrix plots when the rows are not matched. p 24.
How does one view additivity (mentioned on top p 25)?
Is the redundancy useful?

Box Plots: p 25

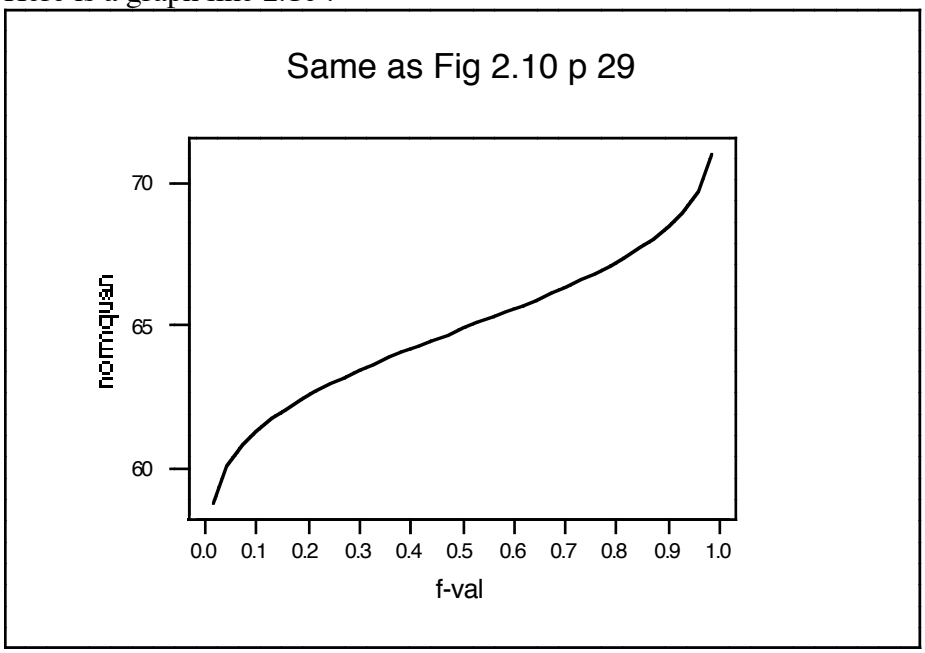
Useful for comparing distributions – like q-q- plots but just use a few quantiles.
Note use of median and IQR. Fig 2.7 shows other boxplot details.



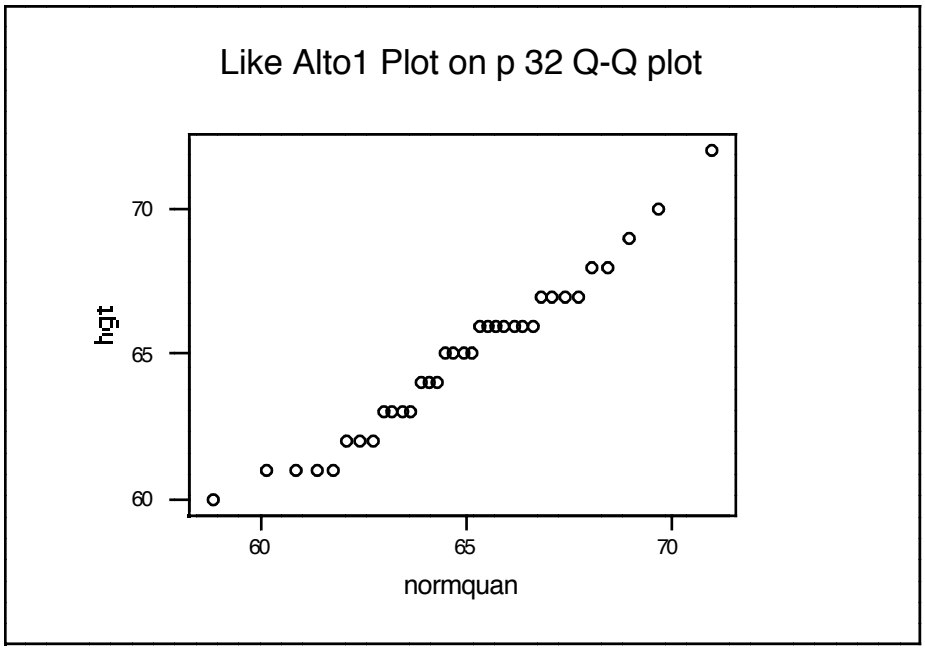
Cf Fig 2.8 Which is better?

Normal Q-Q plots: p 29.

Here is a graph like 2.10 .



How is it generated with software?
Here is the “Normal Q-Q Plot” for Alto 1 heights



Fits and Residuals: Fits need not be parametric (like lines, polynomials, or other such functions). Any smooth could be the basis of a residual calculation. For example, a moving average could be the "fit". (Moving average of order m is just a sequence of averages of a time series determined from adjacent sets of m original time series values.)

Any smooth will produce fits and residuals:

Response = fit + residual

Analysis of the model means analysis of the residual.

If the residual can be "explained" the model can be improved.

Application in text: Model = additive shifts (e.g. for singers heights - Q-Q plot p 22 shows this.)

See Fig 2.14 - a sort of Q-Q plot – not exactly – horizontal axis is from pooled residuals. Pooling residuals is not always a good idea – need to check first – see text p 38

r-f plot p 41 - Fig 2.17 should be suggestive of the R-squared calculation, but more informative.

The Stereogram p 42.