Stat 400 Sep 5, 2003

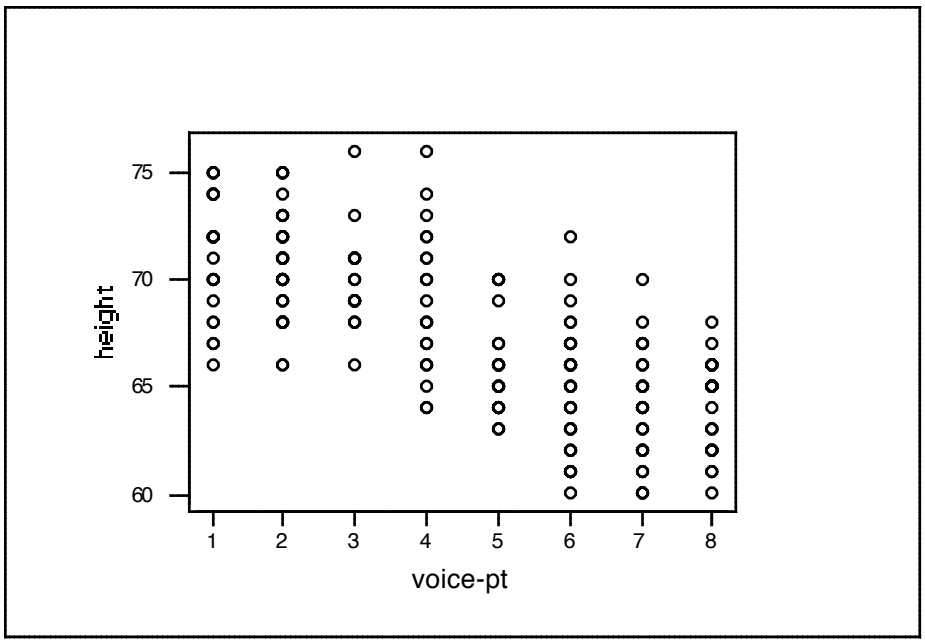
Today: Tutorial Scheduling
Discussion of Ubi Data Analysis
Chapter 2 from Cleveland (pp 16-30)

Discussion of Ubi Data Analysis
Preliminary graphics?
Transformations?
Indices?
Adjustment for Gradient?
Regression or Anova?
Graphical Output?

What's new in Chapter 2?

P 17 Multiple panels are usually ordered from bottom left to top right, horizontally first.

A traditional graph of the relationship between voice part and height would likely have voice part as 8 equally spaced "values" along the x axis and a dotplot or box plot arrayed vertically over the corresponding voice part codes.



Is the Cleveland display better? Objective is to compare the eight distributiions.
How important is it to compare quantiles?

Does the varying sample size for each distribution have a bearing ….?

Note that {q(f): $0 \le f \le 1$} for the various distributions can be compared for each f

Actually, we never know q(f) but only an estimate of it. In fact, in practice we estimate it only at the values of f which correspond to the sample values themselves – in other words at n value of $x_{(i)}$ at which corresponding $f_i = (i-0.5)/n$. (Here $x_{(i)}$ is the ith order statistic – the ith value in the sample when the values are ranked from smallest to largest.)

[q(.50) is the population median, and q(.25) is the population first quartile.]
See graph 2.2 on p 19 – note interpolation between sample values – but not in p16.

Note: Important in understanding the role of quantile plots is the fact that the visual impact of the range of a sample distribution can be misleading unless sample size is accounted for. Range will increase with sample size even when the underlying SD is constant.

Q-Q plots p 21

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.

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

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

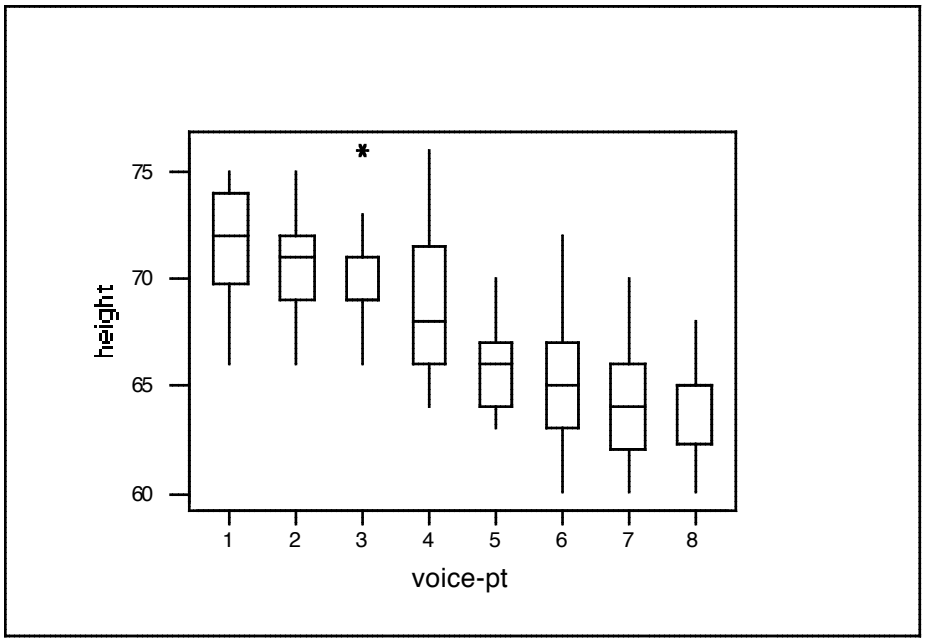
m-d plot

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

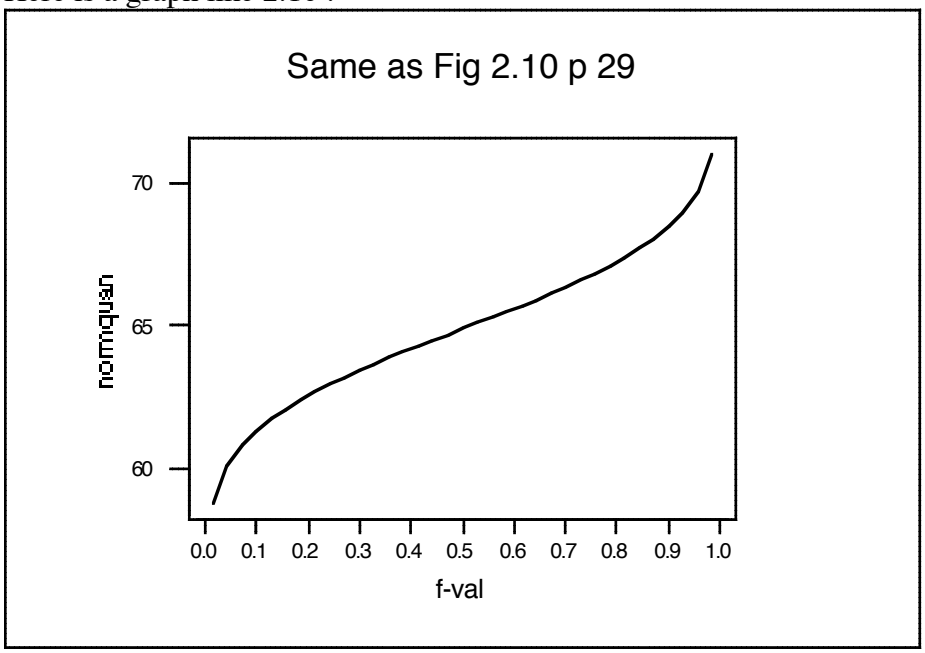
Useful for comparing distributions – like q-q- plots but just use a few quantiles.
Note use of median and IQR.



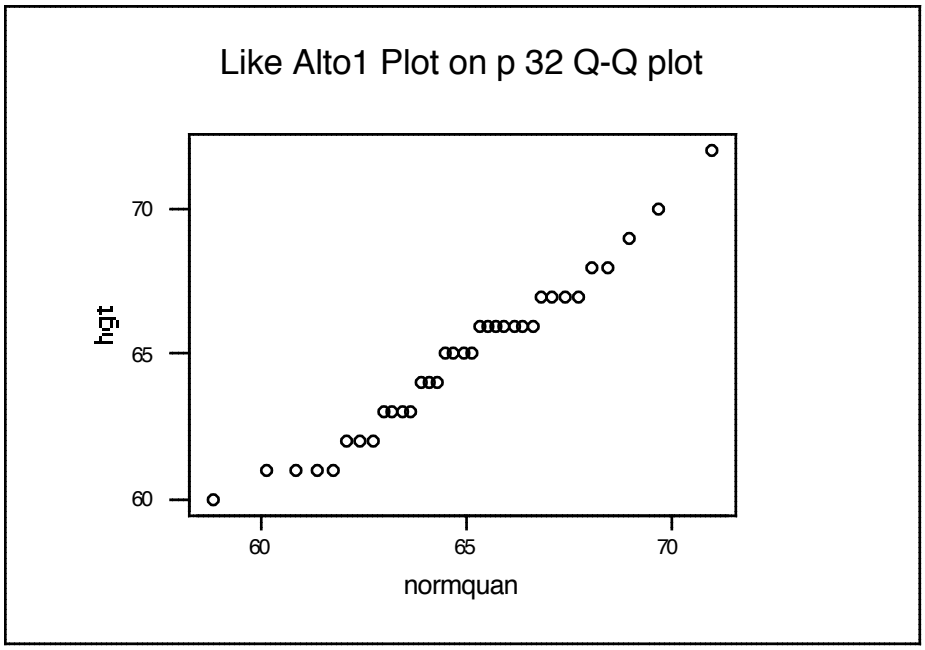
Cf Fig 2.8 Which is better?

Normal Q-Q plots: p 29. Usually for comparing commensurate distributions.

Here is a graph like 2.10 .



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



Homework for Monday – 1. Read ahead in Ch 2 (to p 43)

2. Analyze Ubi Data – to hand in on Wednesday, Sept 11, at class. You submission should include **1 page** only containing a description of what analysis you did, and your findings (at least the ordination of the 10 species). (Mention any transformations you used, indices formed, and regression equations used.) If you have room for a pictorial representation of your ordination, that should be included.

Also, not necessarily for Monday, but try to reproduce the plots that I show you. Ask if you need help.