

Fox Module 3 Univariate displays

- Histograms
- Non-parametric density estimation
- Quantile comparison plots
- Box-plots

The introduction to Chapter 3, “Examining Data,” describes Anscombe’s quartet. This is a fascinating group of four data sets that differ one from the other but have similar regression characteristics.

Read Section 3.1.1, “Histograms,” on pages 28-30. Know how to interpret a stem and leaf display. Fox shows an example on page 29.

Read Section 3.1.2, “Non-parametric density estimation,” on pages 30-34. The formulas for the optimal bin width are not tested on the final exam. Know equation 3.3, and Fox’s comment that “the factor 1.349 is the interquartile range of the standard normal distribution, making $(\text{interquartile range})/1.349$ a robust estimator of σ in the normal setting.”

This robust estimator is useful for much actuarial work. If you have a sample of data points with suspected data errors and outliers, the usual estimators for σ may not work well. Use instead this robust estimator, which is less affected by data errors and outliers.

Read Section 3.1.3, “Quantile comparison plots,” on pages 34-37. Focus on Figures 3.8 and 3.9. The final exam may give a quantile comparison plot and ask if the distribution is heavy or light-tailed (compared to a normal distribution) and if it is positively or negatively skewed.

Read Section 3.1.4, “Box-plots,” on pages 37-40. Know what the hinges represent in a box plot. The box on page 40 summarizes univariate displays.

- The final exam asks you to choose transformations based on the quartile hinges.
- The homework assignment for a later module shows the type of exam problem.
- Understand the hinges and box-plots in this module to help with these problems.

Quantile comparison plots show if a sample is normally distributed or another distribution. The textbook explains the theory; both Excel and R draw quantile comparison plots.