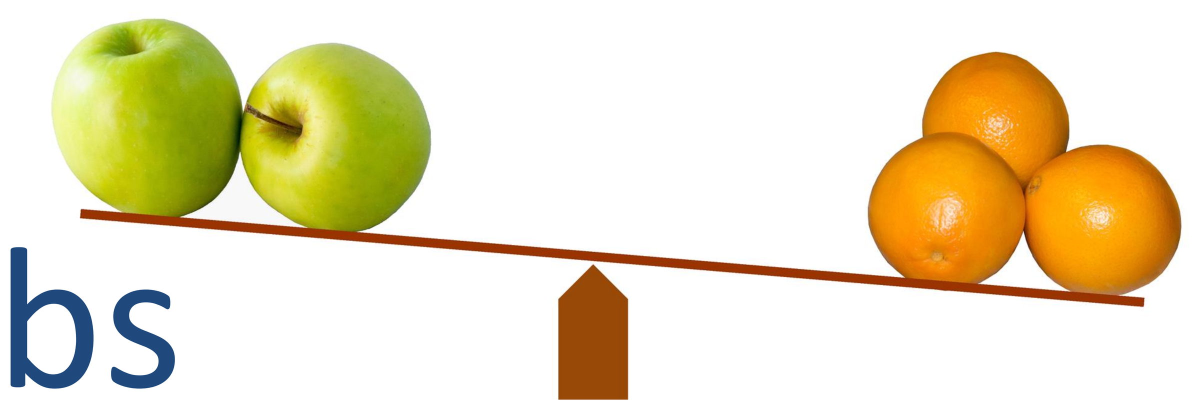
2018

Using Quantile-Quantile Plots to Compare Experimental Apples and Oranges in Physics Labs

Allen Andersen
Utah State University

JR Dennison
Utah State Univesity

Using Quantile-Quantile Plots to Compare Experimental Apples and Oranges in Physics Labs



Allen Andersen and JR Dennison

Materials Physics Group, *Utah State University*

I. Introduction

In experimental physics, measurements are made to test theoretical predictions. Ideally, diverse independent measurements will provide results consistent with either a confirmation or rejection of the theoretical model in question. However, in practice, laboratory measurements are rarely straightforward. Questions of measurement precision, accuracy, or inherently complex or stochastic systems frequently result in distributions of outcomes from repeated tests.

One way to represent such a distribution is the empirical cumulative distribution (ECD). For large sample size j the likelihood of occurrence P as a function of some variable V is

$$P(V) = \frac{1}{j}\sum_{i=1}^{j} \mathbf{1}\{x_i \leq V\} \text{ where } \mathbf{1}\{x_i \leq F\} = \begin{cases} 1 & \text{if } x_i \leq V \\ 0 & \text{otherwise} \end{cases}.$$

It can be useful to know if the ECD of a data set follows some known empirical or physics-based trend. Alternatively, it may be useful to know if two populations of observables follow the same trend.

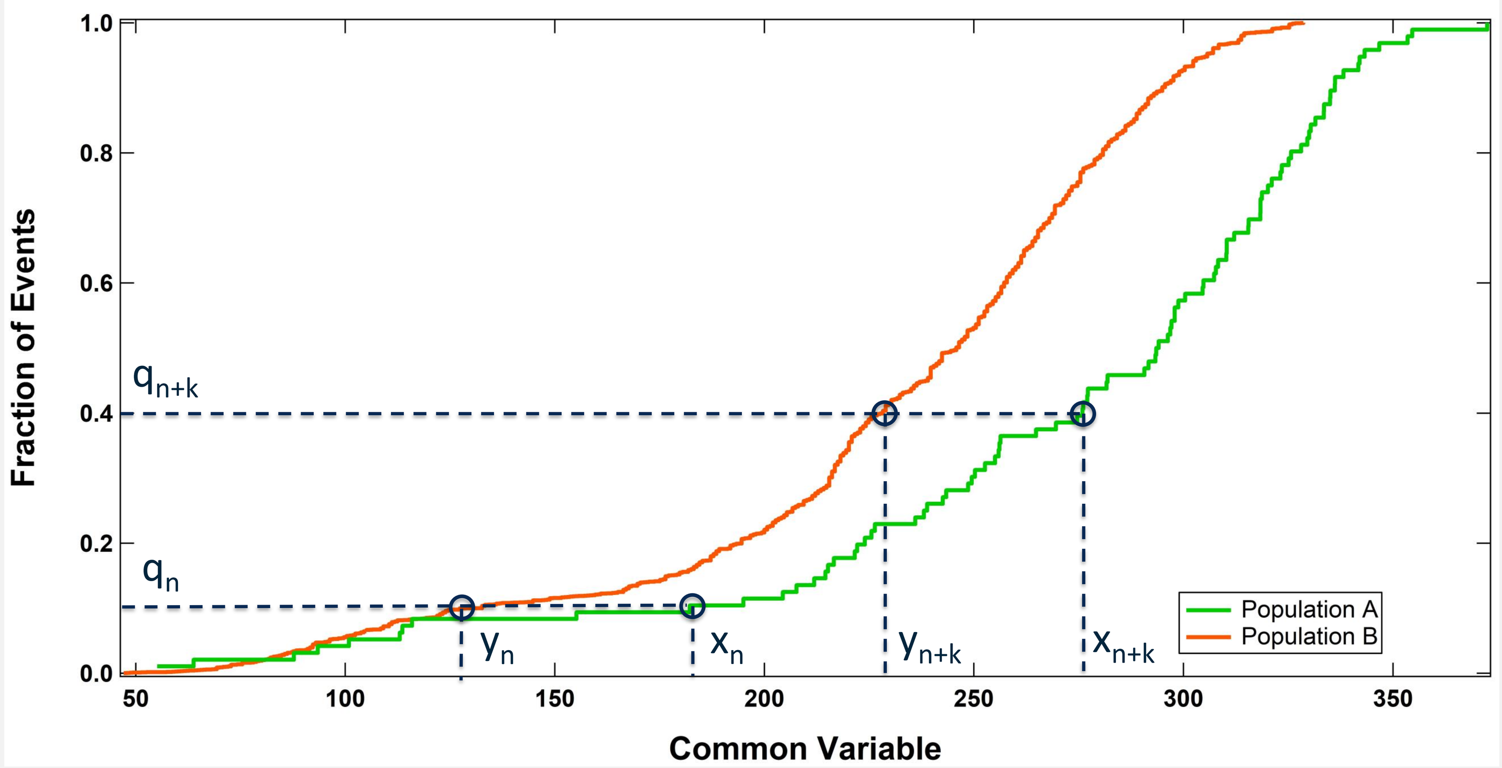


Figure 1. *Empirical cumulative distribution (ECD) plots of measurements of two different observable phenomena. Dashed lines show examples of matching quantiles from the two ECDs. For two ECDs plotted together, one quantile—the vertical axis value—correspond to two horizontal axis values which become the (x, y) pairs on a q-q plot. These ECDs yield the q-q plot in Fig. 2 (a).*

Utah State University
Physics Department
allen.andersen@aggiemail.usu.edu

Scan QR code to access additional publications.

II. 2-Sample Q-Q Plots

2-sample q-q plots compare the quantiles of two populations. For populations of different sample size, interpolation is required.

- Linear q-q plots indicate that the two populations are correlated. See Fig. 2 (a).
  - The distributions are identical if the q-q plot follows $y = x$.
  - Nonzero intercepts indicate that one population is shifted by a constant relative to the other.
  - Slopes other than unity reflect a relative scaling factor between the distributions..
- If the q-q plot is not linear, the distributions follow different trends. See Fig. 2 (b).

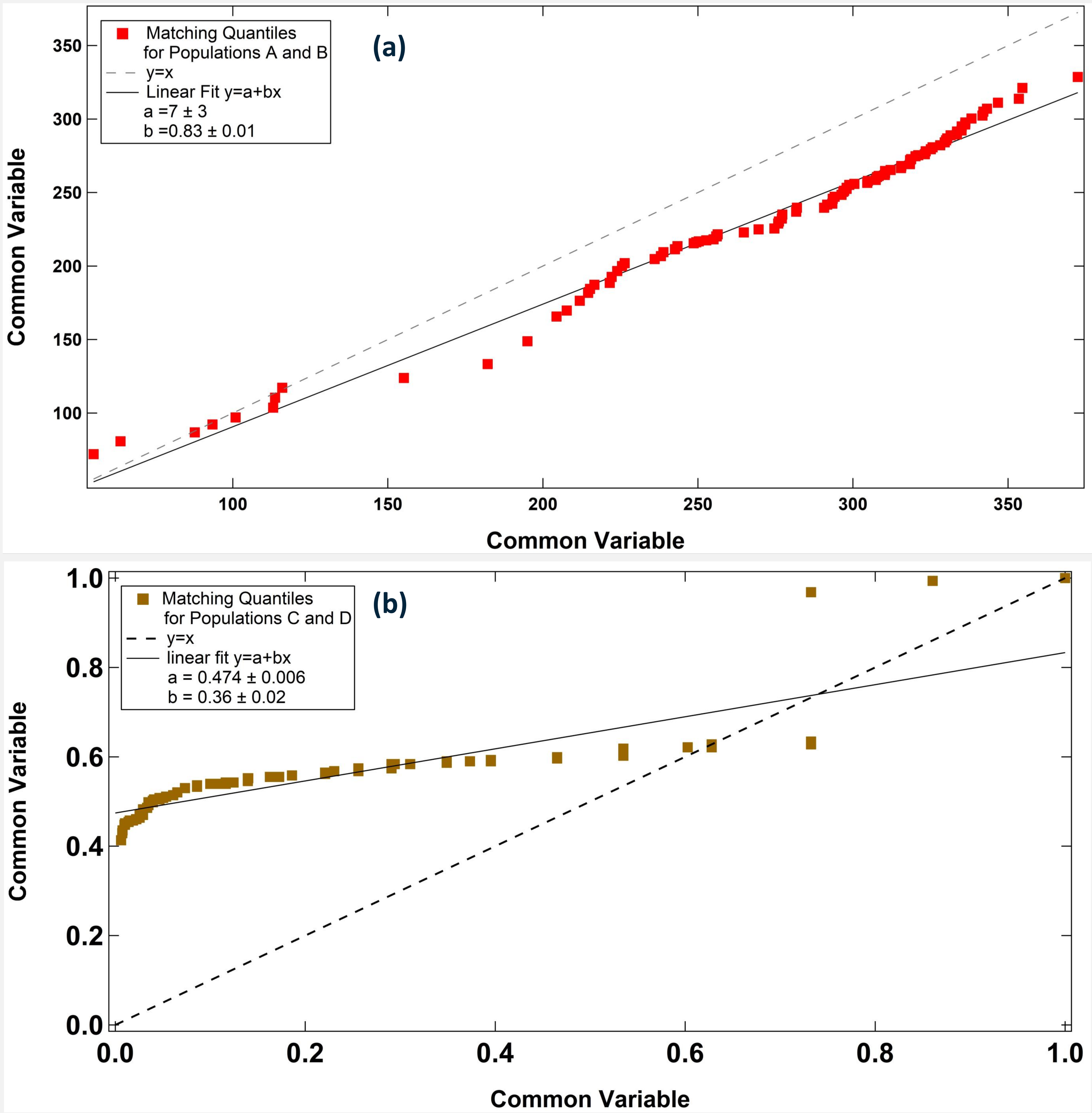


Figure 2. *Examples of quantile-quantile (q-q) plots with linear fits. Dashed lines indicate $y = x$ for reference. (a) q-q plot of two correlated populations of data. (b) q-q plot of uncorrelated data.*

III. 1-Sample Q-Q Plots

Q-q plots can also be used to compare data to fitting functions or two mathematical functions.

Consider a distribution of data fit with both a Gaussian function

$$f(x) = Ae^{-(x-\bar{x})^2/2\sigma^2}$$

and a Lorentzian function

$$f(x) = B\left[\frac{\Gamma^2}{(x - X)^2 + \Gamma^2}\right]$$

as shown in Fig. 3 (a).

Fig. 3 (b) is a q-q plot comparing the two fitting functions. The nonlinearity indicates that they are clearly different.

Fig. 3 (c) show the q-q plots comparing the data to each fit. The Lorentzian fit is shown to be the better fit, since it is clearly more linear with a slope near unity and with an intercept closer to zero compared to the Gaussian fit.

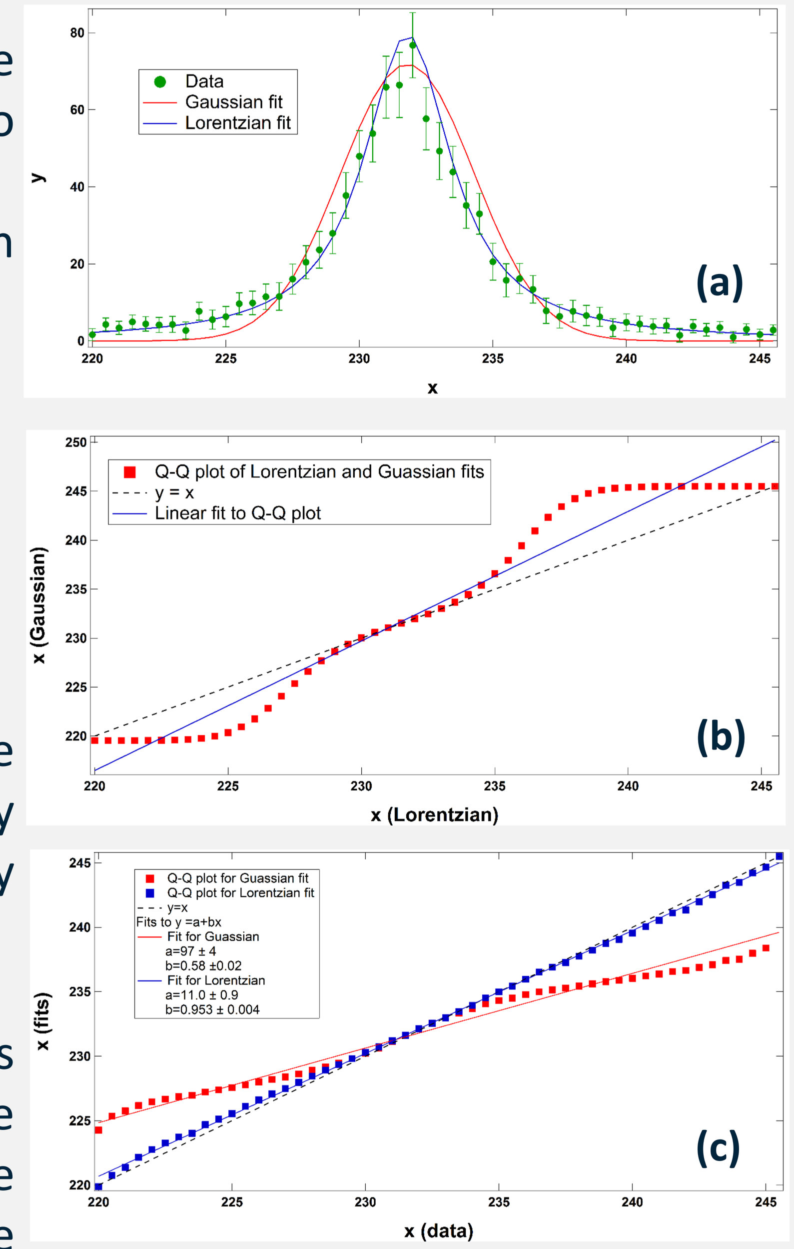


Figure 3. *(a) Data together with Gaussian and Lorentzian fits (b) q-q plot comparing the fitting functions to each other (c) q-q plots comparing the data to each fitting function.*

IV. Conclusions

Distribution functions are found in many branches of physics including quantum mechanics, statistical mechanics, plasma physics, etc. Quantile-quantile (q-q) analysis is not part of the standard physics curriculum, however, it is a useful statistical tool for comparing any two distributions.

- Q-q plots are an easy-to-visualize representation of the relationship between any two distributions.
- For the experimental physicist, q-q plots are especially useful for comparing different populations of measurements. Q-q plots also can compare data to a fitting function.

This work was supported by a NASA Space Technology Research Fellowship.