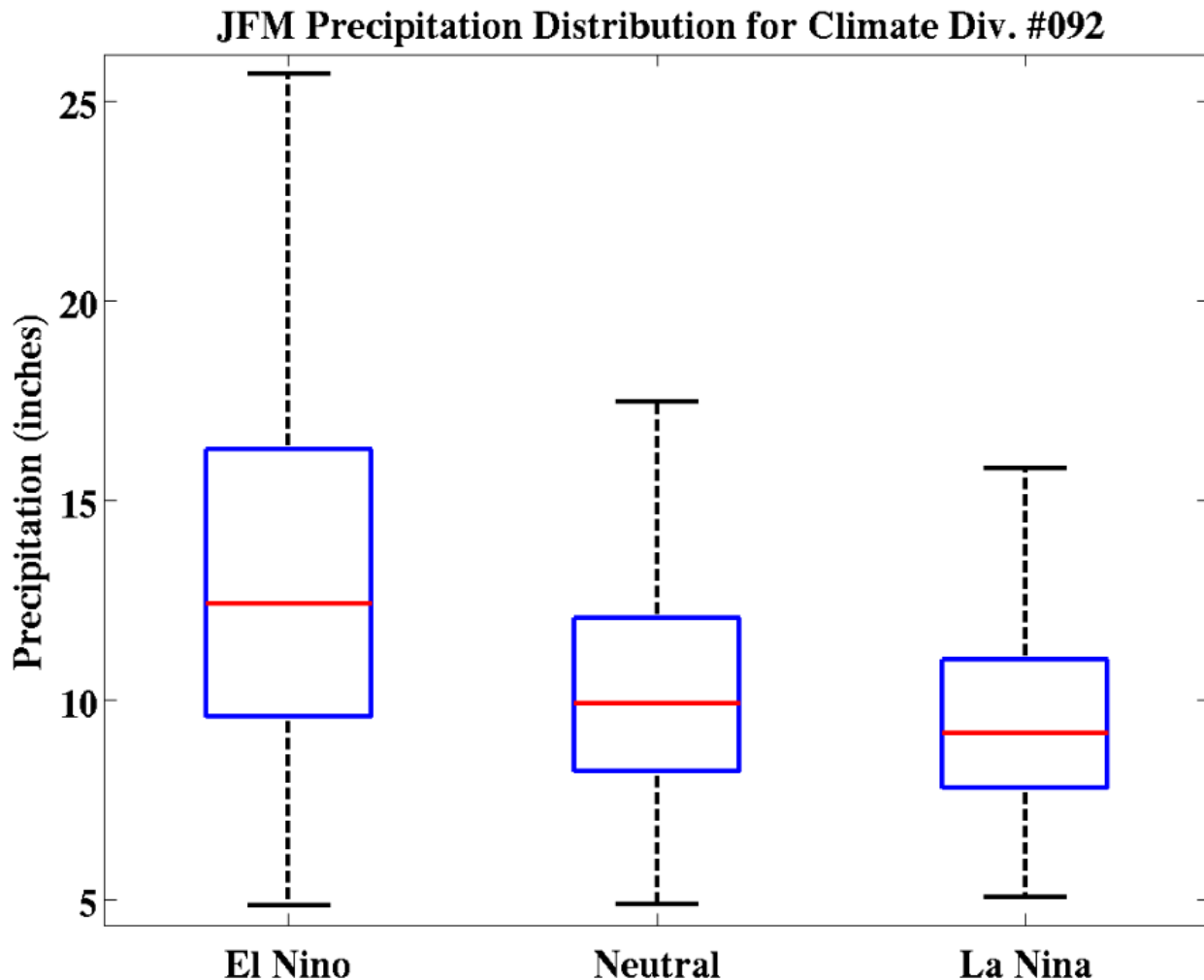


## Making Sense of Biostatistics: Boxplots

By Norman M. Goldfarb

A boxplot (also known as a box-and-whisker plot or diagram) is an efficient way to graphically depict numerical data sets. Because of their compact size and obvious information about key statistical parameters, boxplots are especially useful in side-to-side comparisons of multiple sets of data. The boxplot below, for example, shows winter precipitation in the coastal region south of San Francisco, California, during El Niño, normal and La Niña years.



Source: U.S. National Weather Service,  
[http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/box\\_whiskers/plots/precip/precip.seas.jfm.b50.e08.div.092.png](http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/box_whiskers/plots/precip/precip.seas.jfm.b50.e08.div.092.png)

Boxplots typically display five points in a data distribution (from top to bottom):

- The end of the top whisker is the largest observation (sample maximum, the top of the largest quartile)
- The top of the box is the top of the second-highest quartile
- The line near the middle of the box is the median
- The bottom of the box is the bottom of the third-highest quartile
- The end of the bottom whisker is the smallest observation (sample minimum, the bottom of the lowest quartile)

The box part of a boxplot consists of the two middle quartiles, which is the interquartile range (ICR). The whisker parts of a boxplot consist of the largest and smallest quartiles.

The distances between the different elements in a boxplot indicate the general shape of the data's distribution: the degree of dispersion and skewness, as well as the range of the data. However, boxplots are indifferent to the underlying statistical distribution of the data (i.e., they are non-parametric) and say nothing about the distribution of data within each quartile. The boxplot of a bimodal distribution thus might be identical to that of a normal distribution.

Boxplots can be drawn either horizontally or vertically.

### **Variations**

A boxplot may also indicate with small dots or circles which data points, if any, are outliers and thus not considered part of the quartiles.

The box part of a boxplot always represents the same points on the data distribution, but the ends of the whiskers can represent non-standard values, for example the following:

- One standard deviation above and below the mean of the data
- The 9<sup>th</sup> percentile and the 91<sup>st</sup> percentile
- The 2<sup>nd</sup> percentile and the 98<sup>th</sup> percentile

There are numerous other variations. Any non-standard representation should be explained.

### **Author**

Norman M. Goldfarb is Managing Director of First Clinical Research LLC, a provider of clinical research best practices information, consulting and training services. Contact him at 1.650.465.0119 or [ngoldfarb@firstclinical.com](mailto:ngoldfarb@firstclinical.com).