

First, Look at the Data

What can we learn from the historical record?

Donald J. Wheeler

I recently received a data set consisting of the number of major hurricanes in the North Atlantic from 1940 to 2007. (Major hurricanes are those that reach Category 3 status or higher at some point during their existence.)

The first step in analyzing any data set is to look at the data. This means plotting the data in some meaningful format. With data that form a time series the simplest and best format will be the running record where the data are plotted in time order. The running record for the number of major hurricanes is shown in Figure 1.

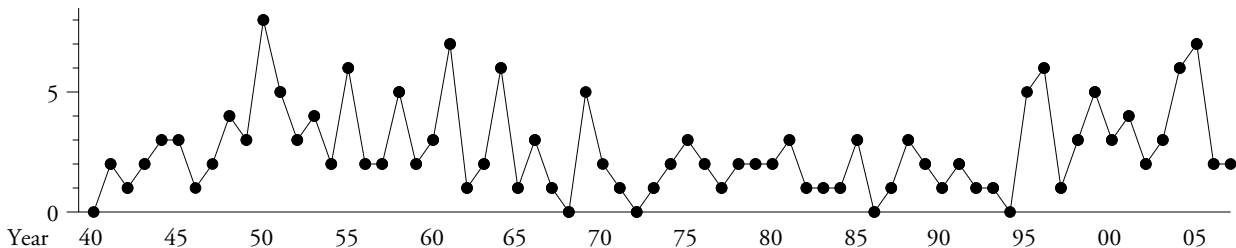


Figure 1: Number of Major Hurricanes in the North Atlantic

The first question of data analysis must always be the question of whether or not the data are homogeneous. If the data are homogeneous, then we can use them with the various computations commonly taught in classes in statistics. However, if the data are not homogeneous, then the question becomes, "Why or when did the changes occur?"

As the first step in answering this question of homogeneity, I look at the running record and visualize a horizontal band that encloses the data. Does this band stay the same for the whole running record, or does it change as time goes by? If there are no changes apparent in the running record, then we may have a homogeneous data set. However, if there are changes apparent in the running record the points where these changes appear to have happened become the focal points for our analysis.

So look at Figure 1. Visualize the horizontal band. Changes appear to have taken place around 1947, 1970, and 1995. These three points divide our data into four segments, which we will use for our analysis.

The primary tool to use in checking any data set for homogeneity is the process behavior chart (also known as a control chart). Here we shall use a chart for individual values, with limits based on the two-point moving ranges. Computing limits for each of our four segments we end up with Figure 2.

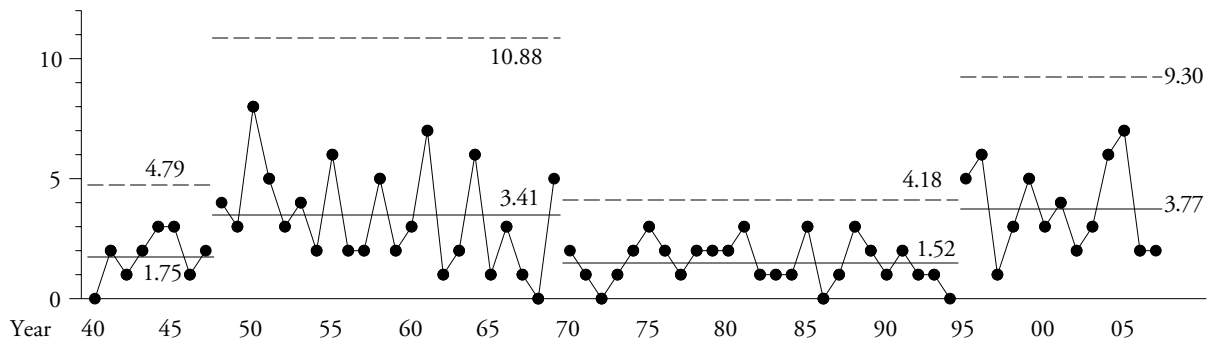


Figure 2: X Chart for Number of Major Hurricanes in the North Atlantic

Based on this analysis it would appear that the hurricane season is switching back and forth between periods of low activity and periods of high activity. In the low periods the average number of major hurricanes is about 1.7 per year, while in the high periods the average is roughly twice that. Moreover, each of these periods tends to last around 25 years. At this point in our analysis we have no theoretical basis for this conclusion, but we do have a very strong empirical case that this is what is happening.

Having found these cycles in the data, I then went to the NOAA Web site and found an article entitled “NOAA Attributes Recent Increase in Hurricane Activity to Naturally Occurring Multi-Decadal Climate Variability” which confirms these cycles (Story 184). In that article we find:

“Nov. 29, 2005—The nation is now wrapping up the 11th year of a new era of heightened Atlantic hurricane activity. This era has been unfolding in the Atlantic since 1995, and is expected to continue for the next decade or perhaps longer.”

“The tropical climate patterns producing the increased activity since 1995 are similar to those during the previous active hurricane era [ending in] the late 1960s. These patterns are opposite to the below-normal hurricane era which ran from 1970 to 1994.”

While NOAA is concerned with the how and the why, the oscillations are clear from the data themselves, and by analyzing the data in context we can demonstrate these patterns beyond a shadow of a doubt.