Control Charts

Joe Mienko

Thursday, August 21, 2014

Monitoring child fatalities on opened child welfare cases as a statistical process

When monitoring a process, such as the percentage of abuse-related fatalities out of opened cases in a child welfare agency, it is often helpful to look at the process in terms of statistical process control. A common tool in statistical process control, is referred to as an "attribute control chart" or just "control chart".

The goal of a control chart is to measure a particular attribute (e.g. abuse-related fatalities) across a number of different cases (e.g. opened cases in a child welfare agency). A standard control chart is a two-dimensional graph with the x-axis representing time and the y-axis representing the outcome of whatever process we want to keep in control (in our case, the percentage of abuse-related fatalities out of opened cases in a child welfare agency). Three lines are plotted on the chart: 1) a top line which identifies the upper limit of "acceptable" values, 2) a center line which identifies the value we would expect to see (perhaps based on previous years' performance), and 3) a bottom line which identifies the lower limit of "acceptable" values.

POC recently released an op-ed in response to a KOMO news report demonstrating the use of a control chart in our attempt to better understand the percentage of abuse-related fatalities out of opened cases in a child welfare agency in Washington. This blog entry describes how we went about creating this chart and some of the assumptions that we have to make when doing so.

A little math

The proportion of fatalities over a given year of opened cases in a child welfare agency is referred to as p. Here, we are testing the hypothesis $H_0 : p = p_0$ against the hypothesis $H_1 : p \neq p_0$. Here, p_0 could either be an arbitrary target value or an estimated value based on the prior performance of the agency. In our op-ed, we have estimated p_0 and will describe this in more detail below.

If we knew the value of p_0, we would calculate the standard deviation (σ) of the proportion of fatalities in a given year ($\hat{p}$) as

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$$

where n is the count of Children's Administration cases opened in a given year. This formula follows naturally from the moments of the binomial distribution. If we capitalize on a feature of certain statistical distributions which statisticians call the "three sigma rule", we can multiply $\sigma_{\hat{p}}$ by three to derive control limits as

$$p_0 \pm 3 \cdot \sqrt{\frac{p_0(1-p_0)}{n}}.$$

By multiplying a standard deviation by 3, we are stating implicitly that we would expect over 99% of our observed values to fall within these limits. In other words, if we were able to observe this process for several years, over 99% of the years that we observe should fall within the upper and lower limits as defined by this formula.

As we stated above, however, we cannot specify p_0 precisely; we need to estimate it from our data. While there are many reasonable approaches to estimating p_0, the approach used in the current example makes use of the average proportion of abuse-related fatalities given as

$$\bar{p} = \frac{\sum_{i=1}^{t} a_i}{\sum_{i=1}^{t} n_i}$$

where a is the count of abuse-related fatalities in a given year (t).

When using $\bar{p}$, the upper and lower limits for our chart are given as

$$\bar{p} \pm 3 \cdot \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}.$$