Part 2—Fisher One-Sample Sign Test

This two-part article of this column presents two well-established statistical tools that are easy to remember. Part 1 discussed Tukey's end count test, which identifies significant differences in the distribution of two samples. In Part 2, you will see how Fisher's one-sample sign test can be used to determine if the median of a population is different from any given value.

Here is an example one-sample problem:

Example 1: *The mortgage department at a local bank has started a project to improve loan processing time. The objective is for 50 percent of loans to be processed from application to closing in 10 days or less. In other words, the median time should be 10 days or less. Baseline measurements of processing time on 20 loan applications are listed below:*

13	*20*	*9*	*12*	*15*	*9*	*12*	*11*	*10*	*14*
20	*19*	*12*	*19*	*9*	*28*	*16*	*11*	*27*	*9*

In this sample, five loans took 10 days or less, and 15 did not. Does this data indicate that the median loan processing time is longer than 10 days, or is this just a string of bad luck?

Fisher's One-Sample Sign Test

Suppose you took a coin out of your pocket and flipped it. You see "heads" and make a note of it. You flip it again and see "heads" again. Then you see three "heads" in a row, and then four, and then five. How many "heads" in a row does it take before you become convinced that the coin is not fair? How confident are you in that conclusion?

Or, suppose you flip the coin 20 times. You observe five "heads" and 15 "tails." Is this series of outcomes unusual enough to convince you that the coin is not fair? How confident are you? This question is identical to Example 1. Example 1 is about loan processing times, not coins. We measured those times in days, but we want to see whether half the times are 10 days or less. To do this, we look at each loan as a trial with two outcomes, either "10 or less" or "more than 10." In the example, we have five "10 or less" outcomes and 15 "more than 10" outcomes, just like the coin experiment.

All these questions can be answered by Fisher's one-sample sign test. This tool is a well-established method available in MINITAB and other software. This article presents an approximate Pocket Stats version of the same test that is simple enough to remember. Often it is simple enough to calculate in your head.

In general, Fisher's one-sample sign test applies to any experiment with N trials, where each trial has two possible outcomes. If each of the two outcomes is equally likely, you would expect to see N/2 outcomes of each type. If the count of either outcome is much higher or lower than N/2, this is good reason to conclude that the outcomes are not equally likely. How much higher or lower is enough? If the count of either outcome is

$\frac{N}{2} - \sqrt{N}$ or less, than you can say with approximately 95 percent confidence that the outcomes are not equally likely.

To make this calculation easier to do in your head, we will add rounding rules so you only have to think about integers. If you can remember a few perfect squares and their square roots listed below, you will not even need a calculator:

N	$\sqrt{N}$
9	3
16	4
25	5
36	6
49	7
64	8
81	9
100	10
121	11
144	12

Here is the Pocket Stats recipe for Fisher's one-sample sign test:

- Find N, the total count of trials, and X, the count of outcomes with the smallest count. It is important that X be less than $\frac{N}{2}$.
- Calculate $\frac{N}{2}$. If N is odd, ignore the .5, rounding down. Using symbols, this is $\left\lfloor \frac{N}{2} \right\rfloor$, where $\lfloor\,\rfloor$ means "round down." Remember, or make a note of $\left\lfloor \frac{N}{2} \right\rfloor$.
- Round N up to the next perfect square, and take the square root $\lceil \sqrt{N} \rceil$. In this formula, the $\lceil\,\rceil$ brackets mean "round up."
- Subtract: $\left\lfloor \frac{N}{2} \right\rfloor - \lceil \sqrt{N} \rceil$. This is the critical value for X.
- If $X \le \left\lfloor \frac{N}{2} \right\rfloor - \lceil \sqrt{N} \rceil$, then you have approximately 95 percent confidence that the two outcomes are not equally likely.

To help remember the rounding rules, note that each rounding rule makes the critical value smaller and farther away from N/2.

Now let's apply this recipe to Example 1:

- $N = 20$ and $X = 5$.

- $\frac{N}{2} = 10$.
- N is between two perfect squares, 16 and 25. Round up to 25, and take the square root, so $\lceil\sqrt{N}\rceil = 5$.
- The critical value is 10 – 5 = 5
- Since X = 5, X $\leq$ the critical value of 5. You can say with 95 percent confidence that the median loan processing time is greater than 10 days.

Consider another example, presented in Part 1 of this two-part article:

Example 2: *A coil supplier and customer disagree about whether certain parts conform to specifications. To investigate, one lot of 10 parts is measured first at the supplier and then at the customer. Here are the inductance measurements:*

Coil ID:	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Supplier:	*220*	*216*	*221*	*215*	*224*	*213*	*219*	*223*	*221*	*224*
Customer:	*218*	*215*	*222*	*212*	*223*	*210*	*218*	*221*	*221*	*222*

Is there a significant difference between the supplier's measurements and the customer's measurements?

Example 2 might appear to be a two-sample problem, because there are two sets of numbers, but it is actually a one-sample problem. (Often, this sort of problem is called a paired-sample or repeated measures problem.) Here, there is only one set of parts, and each part is measured twice. Here are the differences between measurements:

Coil ID:	1	2	3	4	5	6	7	8	9	10
S – C:	+2	+1	-1	+3	+1	+3	+1	+2	0	+2

Notice that there are eight parts with a positive difference, one part with zero difference, and one part with negative difference. If the two measurement systems are the same, then positive and negative differences should be equally likely.

Figure 1 below is a special type of graph for paired-sample problems. This is Tukey's mean-difference plot, which plots the mean value on one axis and the difference between values on the other axis. This graph is simple enough to sketch by hand, when the sample size is small.

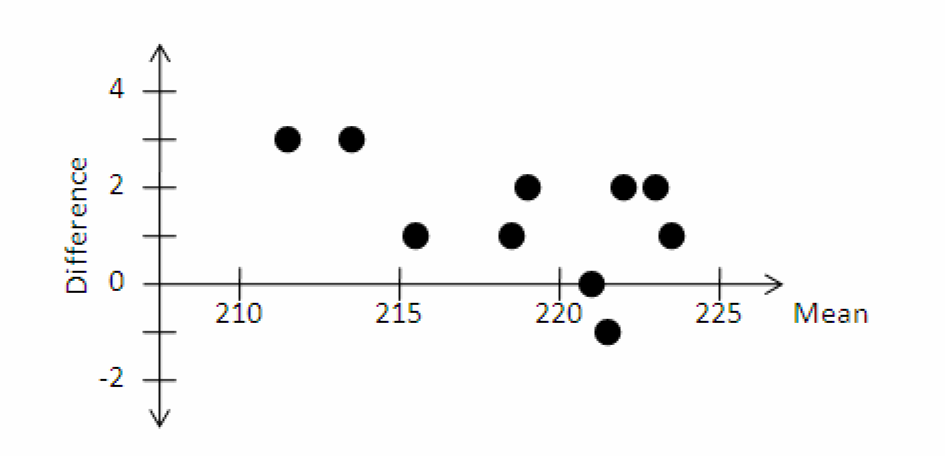

Figure 1—Tukey Mean-Difference Plot

The next tricky part of Example 2 is to ignore the one part with zero difference. When we are looking for a positive or negative difference, zero values provide no useful information, so we have to ignore them. Here is the Pocket Stats solution:

- N = 9 (ignoring the zero value) and X = 1
- $\frac{N}{2} = 4.5$. Ignore the fractional part and remember 4
- $\sqrt{N} = 3$.
- The critical value is 4 – 3 = 1.
- Since X ≤ 1, we can say with 95 percent confidence that the Supplier's gage reads higher than the Customer's gage.

To calculate the critical value for higher confidence levels, use these simple adjustments:

Confidence level percentage	Critical value
95	$\left\lfloor \frac{N}{2} \right\rfloor - \left\lceil \sqrt{N} \right\rceil$
99	$\left\lfloor \frac{N}{2} \right\rfloor - \left\lceil \sqrt{2N} \right\rceil$
99.9	$\left\lfloor \frac{N}{2} \right\rfloor - \left\lceil \sqrt{3N} \right\rceil$

For example, when $N = 30$, the critical values are 9, 7, and 5 for 95 percent, 99 percent, and 99.9 percent confidence, respectively.

Example 3*: In a quality control application, you might be looking at a control chart to detect shifts or changes in a process. Do you see any out of control situations in the control chart shown in Figure 2 below?*

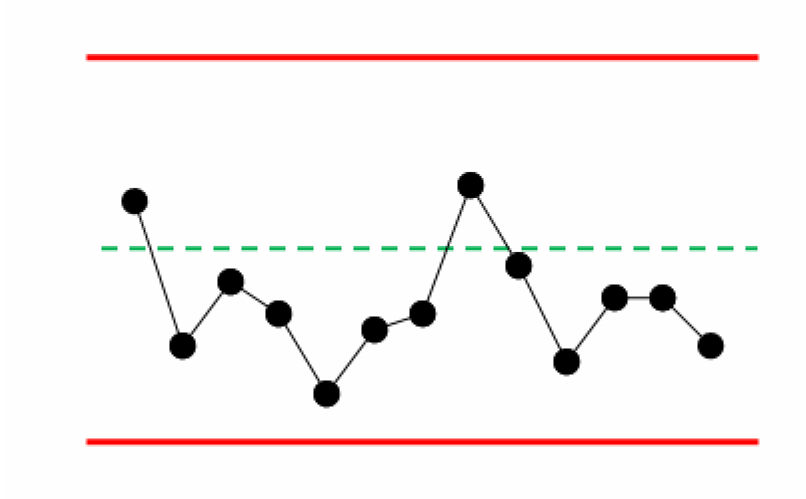

Figure 2—Control Chart

It is interesting that out of the last 12 points plotted on this chart, only one was above the center line, while 11 were below the center line. Does this indicate an out of control condition, or a special cause of variation? There are rules for this, but can you remember them?

It is worth remembering that standard control chart rules require 99 percent confidence before identifying out of control conditions. When $N = 12$, $\frac{N}{2} = 6$ and $\lceil\sqrt{2N}\rceil = 5$, so the 99 percent critical value is 6 – 5 = 1. Since only one out of the last 12 points is above the center line, this is an out of control condition.

Fisher's one-sample sign test is an established statistical tool for testing the median of a continuous distribution, without assuming normality or any other distribution shape. The Pocket Stats recipe described here is based on an approximation, and the critical values may not always be correct. You can calculate the exact critical value in Microsoft Excel. When confidence *C* is a number between 0 and 1, the Excel formula is =CRITBINOM(*N*,0.5,(1-*C*)/2)-1. For marginal cases, or where the impact of a decision is great, Excel can be used to verify the decision.

If you have statistical training, you may know about the one-sample *t*-test. The table below summarizes differences between Fisher's one-sample sign test and the one-sample *t*-test:

	Fisher's one-sample sign test	One-sample *t*-test
Null hypothesis:	The population median is $\tilde{\mu}_0$, a given value	The population mean is μ_0, a given value
Alternative: The test might prove …	… that the population median is higher or lower than $\tilde{\mu}_0$	… that the population mean is higher or lower than μ_0
Assumed population distribution family	No assumption	Normal
Additional assumptions	None	The standard deviation is

		estimated from the same sample

Because of these reasons, applying both tests to the same data may give different results.

Regardless of your job title, whether you are a mechanic, a Black Belt, an engineer or a CEO, you need to make fast, correct decisions from data. When you can make these decisions using simple sketches and easy math with Pocket Stats, you will be more effective and more productive.

References:

Sleeper, A. D. (2006) *Design for Six Sigma Statistics: 59 Tools for Diagnosing and Solving Problems in DFSS Initiatives*, McGraw-Hill