Statistically-Based Investigations of Construction Defects

Presentation to

WestCon Consultants Association

By

Fred Webster, Ph.D., P.E.

Were the Anchor Bolts Installed as Specified?

Take the following construction investigation example:

> *Twenty-five (25) wood frame structures have been constructed by XYZ Construction Company. Each building was designed to include 1000 anchor bolts to provide base shear transfer between the shear panels and the foundation. The bolts are within the walls (at the sill level) and have been covered by plywood, siding, and gyp-board, so they can no longer be observed without destructively opening up portions of the walls. However, questions of quality control have been raised regarding their installation:*
>
> - *Were all the anchor bolts installed?*
> - *Were they installed as specified?*
> - *Will the anchor bolts, as installed, perform their function adequately?*

These types of questions are often the issue of construction disputes. One way of determining the answers is to open up <u>all</u> the shear-wall panels and look for <u>all</u> the anchor bolts that were supposed to be installed, verify their existence, check their installation quality against the installation specification, and re-analyze the structures, if necessary. This is neither practical in larger projects, nor very economical. If these questions are to be addressed in a practical and rational manner, the investigation must include a scope that is less than "look at every bolt", but still should have the potential to give the entire installation a "clean bill of health".

Statistically-Based Investigations

This brings us to what I call "statistically-based investigations". Only a sample of the 25,000 bolt installations is selected for examination, and the individually selected anchor bolts are observed by re-inspection and/or tested in a manner that will address the specific question of concern. On the basis of the sample result a decision is made to either "accept" or "reject" the hypothesis that the installation process was not flawed and that all the anchor bolts were installed as specified (within an acceptably small percentage of possible error), thus addressing one or both of the first two questions.

The third question, "will the anchor bolts, as installed, perform adequately," would usually require that the sampling unit not be individual anchor bolts, but possibly entire shear panels. Since anchor bolts act structurally in a group, one missing or incorrectly installed anchor bolt may not keep an entire shear panel from performing its function adequately for the lateral loads imposed. If it is found after re-evaluation that all the sampled shear panels are capable of performing adequately, in spite of any observed installation flaws, the answer to the third question is "yes" (again, within an acceptably small percentage of possible error).

This third type of question may be asked as a result of answering either of the first two types of questions with a rejection of the hypothesis. Or more likely, it is already known that there are missing and/or incorrectly installed anchor bolts and it is intended to find out whether the sampled walls all meet the test of structural adequacy.

Note that it is very important before a statistically-based investigation is undertaken to formulate the appropriate question or questions to be addressed, determine how they can be addressed by selecting an appropriate sampling unit to be inspected, tested and/or analyzed, and then to identify the population of those sampling units from which the samples may be drawn for the statistical evaluation.

Acceptance Sampling

This type of statistically based sampling investigation is known as "acceptance" sampling, or sampling inspection by attributes. It is concerned only with classifying items (e.g., anchor bolts or shear panels) as "defective" or "not defective" (e.g., the bolt is missing, the bolt is not missing, the shear panel is structurally adequate, the shear panel is not structurally adequate). The intent is to accept or reject the entire population of items based on an agreed upon upper limit for the fraction or rate of "defective" items in the population. If the population being sampled is homogeneous with respect to the installation procedure, then an acceptance sampling plan tells a story about the process that produces the installations. However, for any statistically-based sampling investigation to be rigorously valid, it is essential that the sample be drawn in a random fashion.

Focus on Consumer Risk

Whenever a sampling plan is carried out, rather than a one hundred percent re-inspection, a certain level of risk is assumed along with the conclusion reached. If the actual fraction of "defective" anchor bolts (missing or non-conforming to specifications) in the population of 25,000 anchor bolts is small, then we want to have a high probability of accepting the population of anchor bolts as "all being installed" or "being installed according to specifications" (i.e., a low probability of rejection). On the other hand, if the actual fraction of "defective" anchor bolts in the population is large, we want to have a low probability of accepting the population as being "installed as specified" (i.e., a high probability of rejecting the population).

These two types of risk associated with acceptance sampling are known as producer's risk and consumer's risk, respectively. In doing investigations for "defective" items, we are generally interested in the consumer's risk, the concern being that the number of randomly sampled items is large enough to result in a low probability of accepting a population of anchor bolts as being "installed adequately", when in actuality the fraction of "defective" anchor bolt installations in the population is high. Another way of saying this is that we want a high probability of rejecting a "bad" population that includes a large percentage of "defective" installations.

Sample Size, Confidence, and the "Defective" Rate

More often than not the attraction of statistically-based investigations is economics. That is, we want to randomly select the least number of sample items that will result in a specified confidence level of detecting a "defective" rate in the population which is greater than some agreed upon upper limit. For example, we may want to determine how many sample items to draw in order to result in a 90 percent confidence that a "defective" percentage of 5 percent or greater will be detected by our sample. In other words, if the population percentage of inadequately installed anchor bolts is greater than or equal to 5 percent (i.e., 1,250), what is the smallest sample size that will detect this with a 90 percent confidence level (i.e., the smallest 90/5 plan)? For a single sampling plan, this requires a minimum of 46 anchor bolts randomly selected from the total population of 25,000. Then, if we select 46 anchor bolts randomly, inspect them and find they were all installed, or installed as specified, we would conclude at a 90 percent confidence level that the percentage of missing or improperly installed anchor bolts is less than 5 percent and the population would be accepted as-is. If, on the other hand, even one of the 46 bolts is missing or found not to be installed according to specification, then the above conclusion is not valid. In the strict sense, the hypothesis that the anchor bolts were installed to specifications would be rejected.

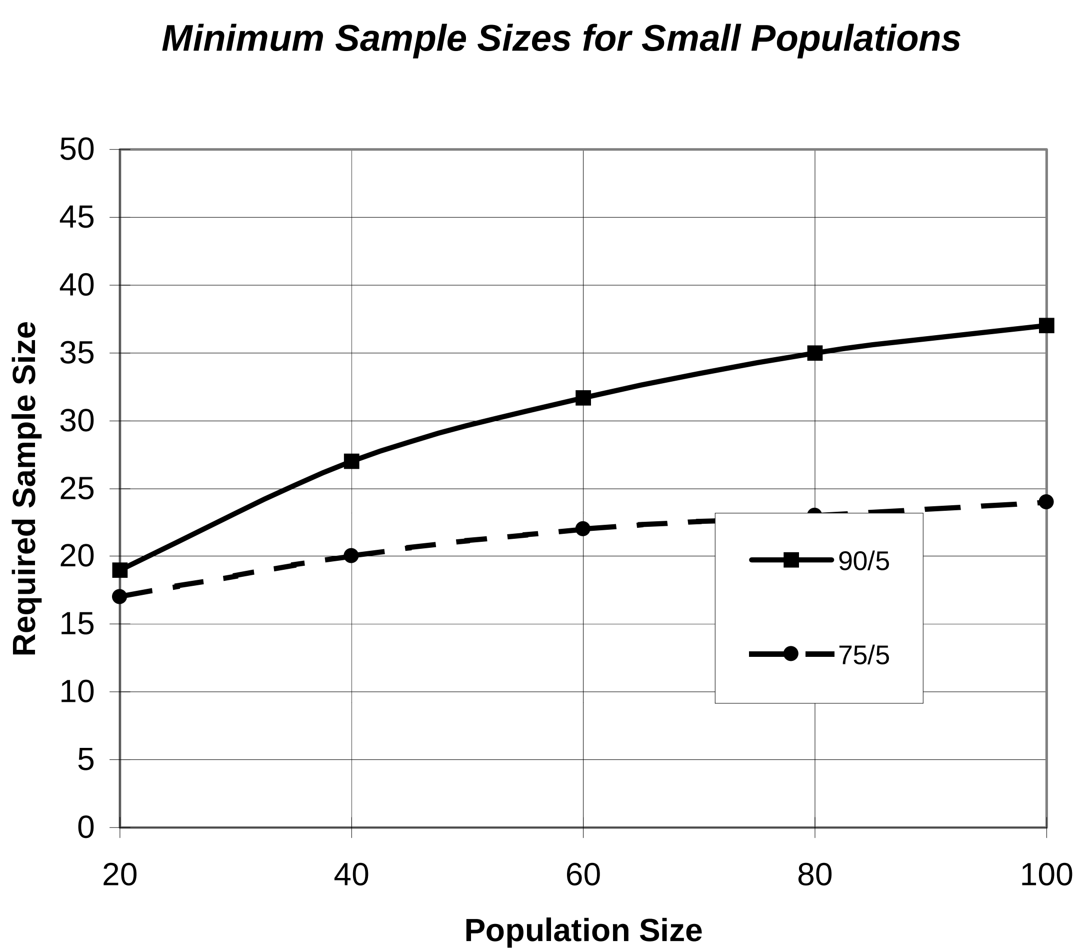
Other sample plans will allow you to find some "defective" items in the sample and still reach an acceptance conclusion. However, these do not represent the least sample size. For example, by allowing a single "defective" bolt installation in our sample, but still arriving at the acceptance conclusion for the 90/5 plan would require drawing a sample of 78 anchor bolts. To allow for the possibility of finding two "defective" items, the sample size would have to be 106, and so on.

These are relatively large samples, and may make us think that our confidence level is too high for what we are looking for. The lower the confidence level desired, or the greater the "defective" rate to be detected, the smaller the sample size can be. For example, a 75/5 sampling plan requires a minimum sample of 28 items. A 75/10 plan requires a minimum of 14 sample items, and so on. On the other hand, if the "defective" rate that is acceptable is smaller, or the required confidence level is higher, the sample size must be larger. A 95/5 sample plan requires a minimum sample size of 60. A 95/2.5 sample plan requires a minimum of 120.

Obviously, if we start our sampling effort and find that the "defective" portion to be significant, there is no point in continuing the sampling effort. There is no question that corrective action should be stated directly. For example, if in the first 10 randomly selected anchor bolts, we observe four to be "defective", then we could conclude right away at a 75 percent confidence level that as much as 63 percent of the anchor bolt installations are “defective”. It is time to stop sampling at this point and start systematically opening up all the shear walls and correcting the condition.

Sampling from Smaller Populations

Sample plans, as outlined above for investigating specific attributes, are independent of the population size. However, it is assumed the number of items or units in the population is large. The sample sizes based on the “large population” assumption are conservative with respect to smaller populations. For population sizes less than 100, it may be desirable to choose a more exact sample size to accomplish the desired “defective” rate detection and confidence level. The attached chart gives the minimum sample sizes for detecting a minimum 5 percent “defective” rate at the 75 and 90 percent confidence levels. Note that as the population size increases, the minimum sample sizes approach their “large population” sizes, asymptotically.



Definitions

Population: The total number items or units that encompasses a work activity or specific process to be investigated.

Random Sample: A sample is random if every item or unit in the population of items or units has an equal chance of being selected.

Sample Size: The number of individual items or units to be observed, inspected, or evaluated from the total number that makes up the population.

Defective Rate, Defective Percentage, or Fraction Defective: The actual number of items or units in the population that are "defective", divided by the total number of items or units in the population, expressed as a fraction or percentage of the population.

Confidence Level: One minus the probability of acceptance. This is the same as one minus the probability of the sample outcome, given the defective percentage in the population.

Follow Up Questions

Following up on questions raised at the WestCon meeting regarding the inspection of fire walls for several attributes at a time and the fact that there may only be a few walls with many items within each. In such cases it is not practical to identify each and every type of item, such as gyp-board or plywood nails, shear bolts, hold downs, etc., and randomly sample from these populations separately. In this case it is more convenient and meaningful to look at entire walls and all their attendant construction elements.

So how do you use statistical sampling and levels of confidence in such a case when the population of walls is small, say for example13 walls?

First, by obtaining a complete picture of the construction elements included in each wall that is destructively investigated, it should be a straight forward exercise to determine a criterion for the adequacy of the entire wall relative to each of these different construction elements. Thus, different adequacy conclusions can be drawn for the construction elements separately for each wall sampled for destructive testing. A sampled wall may be defective in terms of one type of construction element, say gyp-board thickness, but not defective in another, say gyp-board nailing. Based on this approach, the population to be sampled is the population of individual walls, and not a separate population for gyp-board panels and another for gyp-board nails.

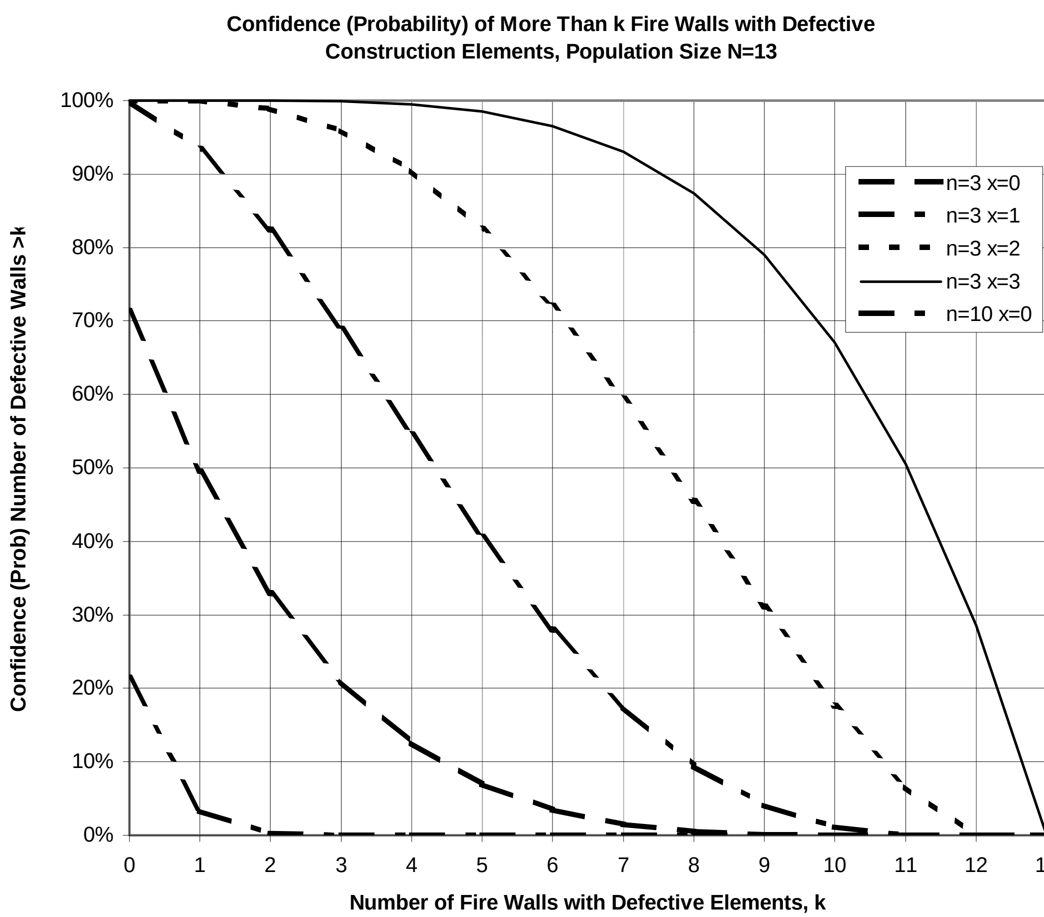
Now, regarding the second part of the question, "How many of the 13 walls do we have to sample for our destructive investigation?"

Obviously, the larger the sample size the more confident we will be of detecting defects that occur in only a small number of walls. On the other hand, if the type of defect is systemic, occurring in many of the walls, it only takes a small number of samples to detect its presence with confidence.

Given that destructive testing is expensive, inconvenient, and destructive, a more practical question would be, “What can be concluded from the results of the sample that you are able to take?”

Suppose that due to time and budgetary constraints that you can only investigate three walls, to be randomly selected and destructively tested. Your investigation finds, for example, that all three of the sampled walls is defective with respect to gyp-board thickness. However, only one of the same three walls is defective with regard to gyp-board nailing. Separate sample results are found for each of the construction elements investigated. For any given sample result there will be a corresponding probability (confidence) distribution of the actual defective rate or number of walls with defective elements in the wall population.

Below is a chart that shows these confidence level distributions for the four possible sample outcomes based on a random sample of three walls from a population of 13. An additional sample outcome for 10 random samples is also shown. Assuming for the moment the sample size is three, the outcome of three defective walls out of three samples (n=3, x=3) shows the highest confidence levels for a systemic problem (e.g., a 93% confidence of more than 7 defective walls out of 13).



On the other hand, if you are trying to show that the walls do <u>not</u> contain systemic defects with regard to various construction elements, a sample size of three is not very persuasive. Even with a sample outcome of zero defective walls out of three samples (n=3, x=0), there is a 71% confidence that one or more walls out of the 13 are defective, although this confidence level drops to only 20% that there are four or more defective walls. Even by taking a sample of 10 walls with an outcome of zero defective walls found during destructive testing (n=10, x=0), there is still a 21% chance of one or more walls out of the 13 being defective.

This example provides a very powerful insight into the investigation of potential construction defects. First of all, construction defects usually occur in the absence of adequate design and/or construction quality control (i.e., lack of appropriate specifications, or lack of procedures to follow specifications). Whenever defect problems start to surface and an investigation is called for, it is anybody's guess as to the number or fraction of defective items. There is usually no way of knowing what to expect based on written specifications or inspection documentation. For the most part these documents simply do not exist.

At the beginning stage of a construction defect investigation the true state (i.e., the percentage or number of defective walls) is unknown and uncertain. However, it may be assumed without undue reservation that prior to sampling and destructive testing, the true condition is equally likely to be any percentage between zero and 100 percent, or any number between 0 and 13 walls (you're not saying there is a systemic problem, you're not saying there is not a systemic problem). Thus, the sampling results will weigh most heavily on the determination of the true state. If the problem is systemic (i.e., the percentage of defective walls is high), it is very likely that it will show up even if only a few random samples are tested. If, however, the problem is not significant (i.e., the percentage of defective walls is low or zero), it will take a large number of samples to show this with any great confidence, as was pointed out in the first part of this presentation. Having to sample a large number of items after the fact to show the quality of a construction process that has taken place in the past is equivalent to having and following specifications (i.e., a quality control process) during construction.

An investigation that is attempting to uncover a systemic problem in a population of fire walls, in the absence of a quality control program during construction, requires only a few samples to show this, if in deed a systemic problem exists. On the other hand, an investigation attempting to show a lack of a systemic problem, in the absence of a quality control program, requires a relatively large sample to show this, if in fact a systemic problem exists does not exist. The large sample becomes the de facto after the fact quality control program.