SAMPLE INFERENCES ON THE QUALITY OF A LARGE BATCH OF RADIOACTIVE TRAINING CAPSULES

David W. Chapman,[1] Bureau of the Census [2]
Washington, DC 20233

Key Words: Quality control, confidence intervals, population estimators

1. Introduction

In 1987 about 8,600 radioactive training capsules were produced for the Federal Emergency Management Agency (FEMA) by a contractor. These capsules are used by various state and local agencies (e.g., fire and police departments) to train personnel on certain emergency procedures. The capsules were manufactured according to very specific procedures designed to prevent any leakage of radioactivity from the capsules. As a precaution, each capsule was sealed by two casings: an inner casing and an outer casing.

In spite of these precautions, about 50 defective capsules (leakers) were detected prior to shipping them from storage locations to state and other local agencies. This was quite disturbing since leakers constitute a serious hazard. The contractor suggested that all the leakers had been discovered and that the remaining capsules were safe. However, FEMA felt that it was quite possible that there were more defective capsules. In order to assess the quality of the remaining 8591 capsules, FEMA and the contractor agreed to carefully examine a simple random sample of 400 of the remaining radioactive capsules produced. (Thorough testing of capsules is expensive and destructive.) It was decided that the entire batch of 8591 capsules would be accepted if no defective capsules were detected in the test sample of 400 capsules.

At that point, FEMA staff requested statistical assistance from the Census Bureau. In response, I was brought in as a consultant to FEMA to help determine what could be said about the quality of the entire batch, based on the sample results. In particular, FEMA staff wondered whether one could determine the probability that the entire batch (minus the original "leakers") was good, given that no defectives were detected among the 400 examined. This paper addresses the types of inferences that can be made in this situation.

2. The Probability of No Remaining Defectives

It is not difficult to write an expression for the desired probability requested by FEMA staff. Let S and Y be random variables that represent the number of defective capsules in the sample of 400 and the universe of 8591, respectively. Although Y is a population parameter, it can be viewed as a random variable since its value is the result of a manufacturing process.

From Bayes Theorem (see, for example, Parzen (1960), p.119), the probability that there are no defective capsules among the remaining 8591, given no sample defects, is:

$$P(Y=0 \mid S=0) = P(Y=0) \Big/ \sum_{j=0}^{8591} P(Y=j)\, P(S=0 \mid Y=j) . \qquad (1)$$

Unfortunately, the probability given in equation (1) cannot be computed unless the production process could be characterized probabilistically. Specifically, the exact probability distribution of the number of defective capsules produced per "batch" would have to be known and assumed to have remained fixed throughout the production process. Since this probability distribution is unknown, and could not be adequately approximated, the desired probability cannot be calculated.

However, some useful inferences can still be made about the entire batch of capsules based on the test sample results.

3. Inferences Based on Confidence Intervals

A confidence interval of any desired level (e.g., 95% or 90%) can be computed for the number of defective capsules in the entire batch, based on the number of defective capsules identified in the sample. In particular, one-sided confidence intervals of the form $Y \le k$ are most appropriate in this situation, where Y is the number of defects in the entire batch of capsules and k is the upper bound of the confidence interval. For example, if no defectives are identified in a random sample of 400 from the entire batch of 8591 capsules, then, with 95% confidence, the number of defectives in the entire batch is 62 or less. This confidence interval was computed using the basic relationship between confidence intervals and hypothesis tests and recognizing that S, the number of sample defects, has a hypergeometric distribution.

Specifically, consider a test of the null hypothesis that Y=k against the alternative hypothesis that Y<k at the 5% level of significance. The critical region for this test has the form $S \le c$, where c (the critical value) is the highest nonnegative integer having the property that $P(S \le c \mid Y = k) \le .05$. Then, a 95% confidence set for Y consists of all values, k, such that the null hypothesis would be accepted, given the sample results. In this case (i.e., S = 0), this amounts to finding the maximum value of k such that the null hypothesis would be accepted, given that there were no sample defects. Using the hypergeometric distribution with parameters N=8591 and k, the largest value of k such that the $P(S \le 0) \mid Y = k) > .05$ is k=62. Hence, if there are no defects in the sample of 400 capsules, the 95% confidence interval is $Y \le 62$. This type of inference is discussed by Wright (1990).

These one-sided confidence intervals for Y are given in Table 1 for both the 90% and 95% confidence levels; for sample sizes of 400, 600, and 800; and for zero or one defect in the sample. Although FEMA and the contractor had already decided to select and examine a sample of 400 capsules, the larger sample sizes were included to illustrate the potential gains associated with them. Table 1 could be used to derive confidence intervals for (1) the total number of capsules that are defective, (2) the total number of outer casings that are defective, or (3) the total number of inner casings that are defective. (That is, Y can represent any of these three quantities.)

The validity of a confidence interval computed from Table 1, as for any statistical inference, depends on the accuracy of the sample measurement process. In this application, the measurement process involves a careful evaluation of the safety of each sample capsule, including a close examination of both the inner and outer casings. FEMA staff were confident that the testing procedure for the sample capsules is almost certain to detect defective outer casings, but may not be as likely to detect defective inner casings since the outer, but not the inner, casings will be cut open during testing. If any sample defects are not detected, the confidence interval based on the number that are detected would be invalid and misleading. Consequently, if there are any doubts regarding the capability of identifying defective inner casings in the sample, confidence intervals should only be calculated for the outer casings.

An approach which would address the FEMA request directly would be to determine the maximum confidence level that could be associated with the confidence interval, $Y \leq 0$ (i.e., no population defects) given that $S = 0$ (i.e., no sample defects). This approach was investigated by Wright (1990) who determined that the maximum confidence level in this case would be the sampling rate, n/N. Therefore, for this application, the maximum confidence that there are no population defective capsules, given that there were no sample defectives, is only 4.7% (i.e., 400/8591). For a sample of 800 capsules this maximum confidence would still only be about 9%. These confidence levels are, of course, too low to be of practical value.

4. Minimum Probability of No Leakers Based on a Joint Confidence Interval

If the testing of inner casings is thought to be reliable, confidence intervals based on Table 1 for the total number of defective inner casings or capsules, as well as for outer casings, should be valid. In addition, a rough lower bound could be placed on the probability that there are no leakers--i.e., no capsules with inner and outer casings that are both defective--based on a joint confidence interval for the inner and outer casings. A crucial assumption underlying this procedure is that capsules were produced in such a way that inner and outer casings were randomly paired. In discussions with FEMA staff, they stated that the production process did put together inner and outer casings in essentially a random way.

If these two assumptions are valid--i.e., that the testing of both inner and outer casings in the sample capsules is reliable and the pairing of inner and outer casings is random--then for a given level of confidence it can be said that, based on the sample results, the probability is at least some specific amount that there are no leakers in the entire batch. For example, if there are no defective inner or outer casings in a random sample of 400 capsules, it can be said that, with 90% confidence, the probability is at least 0.64 that there are no leakers in the entire batch. This lower probability bound was calculated by assuming that the number of defective outer and inner casings were each equal to 62, the upper confidence limit for the 95% confidence interval based on zero sample defectives. Then, the probability that, with a random pairing of inner and outer casings, none of the 62 defective outer casings would be paired with any of the 62 defective inner casings, is .64. The confidence level of 90% is the product of 95% and 95%, which is appropriate as long as the inner and outer casings are paired at random.

This lower probability bound is given in Table 2 for both the 90% and 81% (i.e., 90% squared) confidence levels and for sample sizes of 400, 600, and 800. Note that the first row of the table corresponds to the example discussed in the previous paragraph. In all cases, the calculations are based on a sample having no defective inner or outer casings. It would be straightforward to extend the results to samples that involve one or more defective outer or inner casings.

The probability bounds in Table 2 are rough because they are based on rather conservative numbers of defective inner and outer casings in the entire batch--i.e., the highest value in the one-sided confidence interval. However, there does not appear to be a statistically sound way to improve on these bounds. It is interesting to note that considerable gains are made in these bounds when the sample size is raised, especially from 400 to 600.

5. Summary and Discussion

Because of the unexpected appearance of about 50 defective radioactive capsules (leakers) among the over 8600 training capsules produced under contract, FEMA staff became very concerned about the quality of the remaining 8591 capsules. Leaking capsules represent a critical hazard to those working with them.

The contractor and FEMA agreed to thoroughly test a simple random sample of 400 of the remaining 8591 capsules to assess their quality. It was decided to accept the remaining capsules if no sample defects were found. FEMA staff wanted to be able to determine the probability that the entire batch was good, given no sample defects. As described in Section 2, this probably depends on the distribution of the number of defectives produced by the process. Unfortunately, this distribution is not known. Even if it may be reasonable to assume a specific type of distribution (e.g., the Poisson) values of the parameters of the distribution are not known. It seems likely that the desired probability would depend substantially on the assumed values of the parameters. (However, this assertion has not been verified).

In order to provide some useful sample inferences, one-sided 90% and 95% confidence intervals for the maximum number of defectives in the batch were derived based on observing zero or one sample defective, for sample sizes of 400, 600, and 800. Although the one-sided confidence interval is statistically sound and easy to interpret, it was not the specific inference requested by FEMA staff. To address this, an approach by Wright (1990) was considered which provides the maximum confidence that one can have that there are no defective capsules in the population, given that there are no sample defects. Unfortunately, the maximum confidence level for a sample of 400 is only 4.7%. So this approach was not helpful for this application.

In order to provide a more useful inference that was close to that requested, a joint confidence interval for the outer and inner casings was developed. Using the upper bound of the confidence interval for the number of defective outer and inner casings, a minimum probability of no leakers was derived for a specific joint confidence level, making use of the fact that a leaking capsule must have both of its casings defective. These minimum probabilities are given in Table 2 for the 81% and 90% confidence levels.

A crucial assumption in the calculation of the minimum probability was that the inner and outer casings are paired randomly in producing capsules. If, on the other hand, a capsule with a defective inner casing is more likely to have a defective outer casing than is a capsule that has a good inner casing, the lower bounds of the probability of no leakers given in Table 2 would be invalid. Although FMEA staff felt that this assumption was reasonable, no special investigation of it was made.

A major issue regarding the usefulness of the probability bounds given in Table 2 is the understanding and interpretation of these bounds. To state that, with a certain level of confidence, the probability is greater than or equal to a specified amount that there are no leakers may not be straightforward enough in many applications. However, there does not appear to be a better approach to provide the type of inference requested in this situation.

Acknowledgements:

The author thanks Daniel Gillman of the Census Bureau for the high quality programming work he did in support of this research (see Tables 3 and 4). He also programmed the random selection of the sample of 400 capsules to be tested. In addition, Dan reviewed an earlier draft of this paper and provided useful comments. The author also thanks Ram Chakrabarty of the Census Bureau who also reviewed the paper and made helpful comments.

[1] Now with National Analysts, 1700 Market St., Philadelphia, PA. 19103.

[2] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

References

Parzen, Emanuel (1960), Modern Probability Theory and Its Applications, New York: John Wiley and Sons.

Wright, Tommy (1990), "When Zero Defectives Appear in a Sample: Upper Bounds on Confidence Coefficients of Upper Bounds," The American Statistician, Vol. 44, No. 1, 40-41.

Table 1. One-sided confidence intervals for number of defectives, Y, in a batch of 8591 items.

Sample Size	No. of Sample Defectives	Confidence Level	Confidence Interval
400	0	90	Y ≤ 48
400	0	95	Y ≤ 62
400	1	90	Y ≤ 81
400	1	95	Y ≤ 99
600	0	90	Y ≤ 31
600	0	95	Y ≤ 41
600	1	90	Y ≤ 54
600	1	95	Y ≤ 65
800	0	90	Y ≤ 23
800	0	95	Y ≤ 30
800	1	90	Y ≤ 40
800	1	95	Y ≤ 48

Table 2. Lower bounds for the probability of no leakers in the entire batch when no defective inner or outer casings are found in the sample.

Sample Size	Confidence Level	Minimum Probability of No Leakers
400	90%	0.64
400	81%	0.76
600	90%	0.82
600	81%	0.89
800	90%	0.90
800	81%	0.94