



11-1993

Attribute Sampling: A Library Management Tool

Ken Wise

University of Tennessee - Knoxville, kwise@utk.edu

Jack Kiger

University of Tennessee

Follow this and additional works at: https://trace.tennessee.edu/utk_libpub

Recommended Citation

Kiger, Jack E., and Kenneth Wise. "Attribute Sampling: A Library Management Tool." *College & Research Libraries* 54, no. 6 (1993): 537-549.

This Article is brought to you for free and open access by the University Libraries at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in UT Libraries Faculty: Peer-Reviewed Publications by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

Research Notes

Attribute Sampling: A Library Management Tool

Jack E. Kiger and Kenneth Wise

Attribute sampling is a tool that librarians may use to estimate characteristics of their collection, such as the portion of books needing repair, the accuracy of the circulation records, or the accuracy of cataloging activities. Because sampling always results in risk that the sample is not an accurate indicator of true conditions, one can establish the risk of an incorrect inference. This article describes the nature of attribute sampling and presents the process a librarian might use to make a defensible inference.



Librarians may need to estimate the maximum rate of occurrence of some specific quality or attribute for a particular function within their library. Making these estimates can be difficult since libraries tend to be rather large operations having some functions that are cumbersome to analyze. Making inferences about the number of books missing from the collection, or the accuracy of the circulation system, or the percentage of items in the collection that is not properly bar-coded would be intimidating tasks indeed if the librarian had to review all items or records before drawing any conclusions.

Library management literature (Drott, 1969; Dougherty, et al., 1982; Simpson, 1988; and Powell, 1991) has discussed the use of attribute sampling to estimate

attributes such as the average number of patrons served per day or the average age of patrons. These approaches involve the use of equations, which makes the process unnecessarily complex. Certified public accountants often employ the techniques discussed in this article to estimate the maximum occurrence rate of a phenomenon, such as the maximum portion of the books reflected in the records as being on the shelf that are not. Advantages of the technique are that by using a table to determine sample size and another table to evaluate results, one can draw measurably precise conclusions based on an examination of relatively few items. In the following discussion, we will describe the nature of attribute sampling and illustrate how the librarian may use attribute sampling techniques in managing library operations.

Jack E. Kiger is Warren L. Slagle Professor of Accounting, College of Business Administration, and Kenneth Wise is Business Manager of the University Libraries at the University of Tennessee, Knoxville, Tennessee 37996.

		True State of Population Occurrence Rate is	
		4% or less	Higher than 4%
Examination of sample indicates that the occurrence rate is	4% or less	Correct Decision	Type II Error Risk of concluding the occurrence rate is lower than it actually is
	Higher than 4%	Type I Error Risk of concluding the occurrence rate is higher than it actually is	Correct Decision

FIGURE 1
Sampling Risk in Attribute Testing

THE NATURE OF STATISTICAL SAMPLING

Statistical sampling involves applying procedures to fewer than all items composing a population. A population is all items about which one wishes to make an inference, such as all the books on reserve, all rare books, all books currently circulating, or all bound volumes. Sampling is based on the premise that a sample will be representative of the population. After examining the sample, one makes an inference about the population.

Attribute sampling, a statistical technique, estimates the rate or percentage of occurrence of a specific characteristic or attribute in a population. Attribute sampling is concerned with a rate of occurrence. For example, attribute sampling may be used to estimate the maximum percentage of books not on the shelf that the catalog record indicates are on the shelf. When using such sampling, one evaluates whether a characteristic or attribute is present with a yes or no answer.

Sampling Risk

When selecting a statistical sample from a population, the objective is to obtain a sample that has the same characteristics as

the entire population. For example, if an examination of a sample indicates that 2 percent of the books that should have been on the shelf were not there, one would expect that 2 percent of all the books in the population would not be on the shelf. However, one must accept the risk that the conclusion based on examining a sample will be different from the conclusion if the entire population were examined. This risk is referred to as *sampling risk*. Sampling risk is the risk that the projected characteristics will be different from the true characteristics of the population because all items in the population were not examined. When one is unwilling to accept any sampling risk, one must examine every item in the population.

Considering the relationship between risk and reliability makes the nature of risk clearer. Reliability, a measure of the dependability of an estimate based on a sample, is the complement of risk (1 minus risk). One can specify a level of reliability and determine the number of items that must be examined to achieve it. The degree of reliability of an estimate based on a sample increases as the portion of the items in the population is examined increases. Examining a relatively small number of items can provide

a high degree of reliability beyond which additional testing will improve reliability only in very small increments. Also one can have complete confidence in an estimate by examining the whole population.

A librarian who examines a sample and concludes that the occurrence rate of a specific characteristic is 4 percent or less when the population's occurrence rate is 4 percent or less makes a correct decision (see figure 1). Similarly, when a librarian examines a sample containing a 4 percent or greater occurrence rate and the occurrence rate in the population is 4 percent or greater, the librarian makes a correct decision.

Attribute sampling, a statistical technique, estimates the rate or percentage of occurrence of a specific characteristic or attribute in a population.

However, when sampling, one may make either of two mistakes. One may conclude that the occurrence rate is higher than 4 percent when it is not. In such a situation, the occurrence rate is estimated to be higher than it actually is. Making such an error is generally referred to as a *Type I error* and the risk of making such an error is referred to as *Alpha risk*. Alternatively, one may examine a sample and conclude that the occurrence rate is 4 percent or less when it actually is not. This type of error is referred to as a *Type II error* and the risk of making such an error is referred to as a *Beta risk*. In such a situation, the librarian concludes the occurrence rate to be lower than it actually is.

When sampling results cause a librarian to conclude that the occurrence rate is higher than it actually is, the librarian may incur additional costs by assigning staff to check and correct all of the records. When a librarian concludes that the occurrence rate is lower than it is, the librarian accepts the occurrence rate as satisfactory when it is really not. Hence, the records continue to be in error.

Should a librarian conclude that the occurrence rate is less than it actually is, the librarian would assume that the records are correct. Hence, when sampling, librarians are concerned with the risk of concluding the occurrence rate is lower than it actually is. This risk (Beta) may be reduced by increasing the sample size. Sampling risk varies inversely with the sample size: the greater the sample size, the smaller the sampling risk. Increasing the sample size to include all items in the population would eliminate all sampling risk.

Nonsampling Risk

In addition to the sampling risk, librarians incur the nonsampling risk, which results from human error such as failure to recognize an occurrence when performing a procedure or use of an ineffective procedure. For example, an exhausted or inadequately trained person might misread a call number when examining a book. An example of using an ineffective procedure is comparing only the title of a book to the catalog record rather than comparing all of the details. Nonsampling risk does not result from failure to examine all items in the population. Nonsampling risk is not ordinarily quantified. Librarians may minimize nonsampling risk by providing adequate training and supervision to persons examining sample items.

Statistical versus Nonstatistical Sampling

Statistical sampling requires using random techniques for selecting a sample and using the laws of probability to evaluate results of the sampling process. Nonstatistical sampling refers to selecting a sample without using random selection techniques or making an inference from a sample without using the laws of probability. When using statistical sampling, a librarian can use probability theory to make statements or generalizations about a population and to measure the risk that the sample is not representative of the population (sampling risk). Statistical sampling also assists the librarian in setting an efficient sample size and in evaluating sample results.

Sampling is not always appropriate. For example, a librarian wishing to correct all of the errors in the circulation system may identify errors but would not detect all of them by sampling. The rare book librarian may choose not to verify every item in the listing while the newspaper librarian may verify no items.

MAKING A STATISTICAL INFERENCE

Statistical sampling provides a framework for making a statistical inference. Below, we will discuss the steps in making a statistical inference:

1. Determine the objective of the statistical inference.
2. Define the population and the sampling unit.
3. Determine the acceptable risk of concluding that the occurrence rate is lower than it actually is.
4. Set the tolerable occurrence rate.
5. Determine the expected population occurrence rate.
6. Using a statistical sample size table, determine the initial sample size.
7. Using random sampling techniques, identify the actual items to examine.
8. Examine the selected items and identify occurrences of deviations.
9. Make conclusions about the frequency of occurrences.

To illustrate these steps, we will assume that a librarian wants to estimate the portion of books missing from a specified range of the collection; for example, all books in the LC classification PR, generally the English literature collection. The collection consists of 15,000 items readily identifiable in the catalog record. The librarian is willing to accept a 5 percent risk of concluding that the occurrence rate is lower than it actually is and that the records are accurate enough if they contain a 4 percent error rate. The librarian expects that only .5 percent of the books are missing.

Determine the Objective of the Statistical Inference

Attribute sampling is used to estimate the rate of occurrence or percentage of

items with a specific quality or attribute within a population. An attribute is a characteristic of an item being examined. When the characteristic is not present, a deviation exists. For example, a librarian may be concerned about the reliability of the catalog records of the English literature collection. On the basis of professional judgment, the librarian would like to be able to determine that 4 percent or fewer of the books which the catalog

One must identify the population in such a way as to ascertain that all items in the population are subject to being included in the sample.

records show as being on the shelf are missing. Rather than determining the accuracy of every item in the catalog record the librarian establishes the hypothesis that the occurrence rate of the deviation is 4 percent or less. The attribute being examined is whether a book that should be on the shelf according to the catalog record is in fact on the shelf. When the librarian examines the shelf, the book is either there or it is not. When the librarian looks on the shelf for a book and it is not there, a deviation exists.

Define the Population and Sampling Unit

A population is all the items about which one wishes to make an inference. The population one examines is generally dictated by the objective of using attribute sampling. For example, if one is evaluating whether the catalog records include all English literature books actually on the shelf, the population is all of the English literature books on the shelf at a particular point in time. If one is evaluating the accuracy of the catalog records, the population is all the catalog records of English literature books at a particular point in time. A population consists of sampling units. A sampling unit is an individual item such as a book or an entry in a record of the population (such as the catalog record) that is examined.

One must identify the population in such a way as to ascertain that all items in the population are subject to being included in the sample. For example, if one wants to make an inference about the entire collection of English literature books, all books in the collection must be subject to selection, not just those currently on the shelves.

Population size has little impact on the sample size for populations of less than 5,000 items, and no effect on sample size for populations of 5,000 or more items.

Specify Acceptable Risk of Concluding that the Occurrence Rate Is Lower Than It Actually Is

When sampling, one must accept some risk that his or her conclusion about the population occurrence rate of a characteristic is incorrect. The risk of concluding that the occurrence rate is lower than it actually is refers to the probability of accepting an attribute as satisfactory because of the tolerable occurrence rate specified when the occurrence rate actually is higher. When specifying a 5 percent risk of concluding that the occurrence rate is lower than it actually is, one accepts a 5 percent chance of concluding that an occurrence rate is lower than the tolerable rate when it is not. Looking at it another way, one has a 95 percent reliability level or a 95 percent chance of being right.

Prior to selecting a sample and performing procedures to determine the presence or absence of an attribute, one must specify the acceptable risk of concluding that the occurrence rate is lower than it actually is. The higher the risk of concluding that the occurrence rate is lower than it actually is, the smaller the required sample size. This is logical, because the higher the risk, the smaller the likelihood that the sample will be representative of the population. In other words, the less evidence gathered, the higher the risk of concluding that the occurrence rate is lower than it actually is. As discussed below, the acceptable risk of concluding that the occurrence rate is lower than it actually is determines which sample size table to use.

Set the Tolerable Occurrence Rate

The tolerable occurrence rate is the maximum occurrence rate for a specific attribute that the librarian will permit. For example, consider a librarian's test of the catalog records for English literature books. When setting the tolerable occurrence rate at 4 percent, the librarian has decided that even if 4 percent of the English literature books included in the catalog records are not on the shelf, the assessment of the occurrence rate would not change.

Setting the tolerable occurrence rate involves judgment. Tolerable occurrence rates vary with the importance of a particular attribute. The more critical the attribute, the lower the tolerable occurrence rate should be.

Estimate the Expected Population Occurrence Rate

The expected population occurrence rate or the frequency of the attribute also affects the initial sample size. An estimate of the rate can be made from the previous year's occurrence rate, the occurrence rate in a preliminary random sample of the population being examined, or an estimate based on one's experience with occurrence rates in similar situations. Estimating the occurrence rate incorrectly may cause the initial sample size to be incorrect and require selecting an additional sample. Fortunately, an incorrect estimate does not increase the risk of concluding that the occurrence rate is lower than it actually is.

The smaller the expected occurrence rate in relation to the tolerable occurrence rate, the smaller the required sample size. In other words, when the maximum tolerable occurrence rate is 4 percent, and the estimate of the actual occurrence rate is .5 percent, the sample size will be smaller than if the estimate of the actual occurrence rate were 2 percent. The closer the expected population occurrence rate is to the tolerable occurrence rate, the larger the required sample size.

Determine the Initial Sample Size

After estimating the expected population occurrence rate, specifying a risk of

TABLE 1
DETERMINATION OF SAMPLE SIZE TABLE—10% RISK OF CONCLUDING THE
OCCURRENCE RATE IS LOWER THAN IT ACTUALLY IS

Expected Percent Rate of Occurrence	Tolerable Rate: % Rate of Occurrence									
	1	2	3	4	5	6	7	8	9	10
0.25	400	200	140	100	80	70	60	50	50	40
0.50	800	200	140	100	80	70	60	50	50	40
1.0		400	180	100	80	70	60	50	50	40
1.5		*	320	180	120	90	60	50	50	40
2.0			600	200	140	90	80	50	50	40
2.5			*	360	160	120	80	70	60	40
3.0				800	260	160	100	90	60	60
3.5				*	400	200	140	100	80	70
4.0					900	300	200	100	90	70
4.5					*	550	220	160	120	80
5.0						*	320	160	120	80
5.5						*	600	280	160	120
6.0							*	380	200	160
6.5							*	600	260	180
7.0								*	400	200
7.5								*	800	280
8.0									*	460
8.5									*	800
9.0										*
9.5										*

* Sample size more than 1,000.

concluding the occurrence rate is lower than it actually is, and setting a tolerable occurrence rate, one may use a sample size table such as that in table 1, 2, or 3 to make an initial estimate of sample size. The sample size is called the initial sample size because the occurrence rate in the actual sample determines whether the sample size is large enough to reach the desired conclusion.

The specified risk of concluding the occurrence rate is lower than it actually is determines which table to use. The sample size tables are one-sided tables because they present an upper occurrence rate (not an upper and lower) for a given risk of concluding that the occurrence rate is lower than it actually is. One-sided tables are used because of concern with knowing the maximum, not the minimum occurrence rate.

The librarian is willing to accept a 5 percent risk of concluding that the occur-

rence rate is lower than it actually is, and is willing to assume that the catalog records are accurate enough if they contain a 4 percent error rate (tolerable occurrence rate), and expects the population occurrence rate to be .5 percent. To determine the initial sample size, follow these steps:

1. Locate the table that corresponds to the acceptable risk of concluding that the error rate is lower than it actually is.
2. Locate at the top of the table the tolerable occurrence rate.
3. Locate the expected occurrence rate at the left of the table.
4. Read the initial sample size from the intersection of the column (determined in 2 above) and row (determined in 3 above).

Using table 2, the initial sample size is 120. Table 4 shows the effect of the risk of concluding the occurrence rate is

TABLE 2
DETERMINATION OF SAMPLE SIZE TABLE—5% RISK OF CONCLUDING THE OCCURRENCE RATE IS LOWER THAN IT ACTUALLY IS

Expected Percent Rate of Occurrence	Tolerable Rate: % Rate of Occurrence									
	1	2	3	4	5	6	7	8	9	10
0.25	650	240	160	120	100	80	70	60	60	50
0.50	*	320	160	120	100	80	70	60	60	50
1.0		600	260	160	100	80	70	60	60	50
1.5		*	400	200	160	120	90	60	60	50
2.0			900	300	200	140	90	80	70	50
2.5			*	550	240	160	120	80	70	70
3.0				*	400	200	160	100	90	80
3.5				*	650	280	200	140	100	80
4.0					*	500	240	180	100	90
4.5					*	800	360	200	160	120
5.0						*	500	240	160	120
5.5						*	900	360	200	160
6.0							*	550	280	180
6.5							*	1000	400	240
7.0								*	600	300
7.5								*	*	460
8.0									*	650
8.5									*	*
9.0										*
9.5										*

* Sample size more than 1,000.

lower than it actually is, tolerable occurrence rate, and expected population occurrence rate on the initial sample size. The relationship between sample size and these factors can be summarized in table 5.

Select the Sample

After determining the initial sample size, select a random sample from the population. A random sample is a sample in which every sampling unit in the population has an equal chance of being included in the sample. For the English literature collection example, a librarian would select 120 entries in the catalog record. While books including random number tables are available, using computer software to generate a listing of random numbers is much more efficient. Both Lotus and Excel have a feature for generating random numbers. Also, most computer programmers can easily incorporate random number

generators into programs they routinely run to select a random sample.

Examine the Items in the Sample

Next, the librarian performs procedures to determine whether deviations occur. The procedures are the same whether one uses statistical sampling or examines all items in the population. As stated above, a deviation exists for any book that the record indicates is on the shelf if the book is not on the shelf. Assume that only one English literature book was not found on the shelf.

Evaluate the Sample Results

The librarian may calculate the occurrence rate of deviations in the sample and use it to estimate the population's upper occurrence rate. The sample occurrence rate is computed by dividing the occurrence rate for each attribute by the sample size. This rate is the best esti-

TABLE 3
DETERMINATION OF SAMPLE SIZE TABLE—1% RISK OF CONCLUDING THE OCCURRENCE RATE IS LOWER THAN IT ACTUALLY IS

Expected Percent Rate of Occurrence	Tolerable Rate: % Rate of Occurrence									
	1	2	3	4	5	6	7	8	9	10
0.25	*	340	240	180	140	120	100	90	80	70
0.50	*	500	280	180	140	120	100	90	80	70
1.0		*	400	260	180	140	100	90	80	70
1.5		*	800	360	200	180	120	120	100	90
2.0			*	500	300	200	140	140	100	90
2.5			*	1000	400	240	200	160	120	100
3.0				*	700	360	260	160	160	100
3.5				*	*	550	340	200	160	140
4.0					*	800	400	280	200	160
4.5					*	*	600	380	220	200
5.0						*	900	460	280	200
5.5						*	*	650	380	280
6.0							*	1000	500	300
6.5							*	*	800	400
7.0								*	*	600
7.5									*	800
8.0									*	*
8.5									*	*
9.0										*
9.5										*

* Sample size more than 1,000.

TABLE 4
EFFECT OF RISK, TOLERABLE OCCURRENCE RATE, AND EXPECTED OCCURRENCE RATE ON INITIAL SAMPLE SIZE ON

Risk of Concluding the Occurrence Rate Is Lower Than It Actually Is	Tolerable Occurrence Rate	Expected Population Occurrence Rate	Initial Sample Size
5%	6%	1.0%	78
5	6	2.0	129
5	5	2.0	181
10	5	2.0	132

TABLE 5
RELATIONSHIP BETWEEN OCCURRENCE RATE AND SAMPLE SIZE

Change In	Impact on Sample Size
Expected population occurrence rate	Direct
Tolerable occurrence rate	Inverse
Risk of concluding the occurrence rate is lower than it actually is	Inverse

mate of the occurrence rate for the population. A statistical table similar to the ones in tables 6, 7, or 8 may be used to estimate the population's upper occurrence rate at the level of risk specified earlier. The initial sample size was determined to be 120. When examining the sample, one occurrence was found. The librarian would estimate that approximately 1 percent (computed by dividing 1 by 120) of English literature books are missing. To estimate the population's

TABLE 6
TABLE FOR EVALUATING SAMPLE RESULTS—10% RISK OF CONCLUDING
OCCURRENCE RATE IS LOWER THAN IT ACTUALLY IS

Sample Size	Number of Observed Occurrences																			
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40	45
10																0		1		2
20											0				1	2		3	4	5
30								0				1		2		4	5	6	8	9
40						0			1			2	3		4	6	7	9	11	13
50					0		1			2	3	4	5		8	10	12	15	17	
60			0				1	2		3	4	5	6	7	10	13	15	18	21	
70			0		1		2	3	4	5	6	8	9	12	15	18	22	25		
80			0		1		2	3	4	5	6	8	9	10	14	18	22	25	29	
90			0		1	2		3	4		6	7	9	11	12	16	20	25	29	33
100			0	1		2	3	4		5	7	9	10	12	14	19	23	28	33	38
120		0		1	2	3	4	5	6	7	9	11	13	15	17	23	29	34	40	46
140		0	1	2	3	4	5	6	7	9	11	13	16	18	21	27	34	41	48	54
160		0	1	2	4	5	6	8	9	10	13	16	19	22	25	32	40	47	55	63
180		0	2	3	4	6	7	9	10	12	15	18	22	25	28	37	45	54	63	71
200	0	1	2	4	5	7	8	10	12	14	17	21	24	28	32	41	51	60	70	80
220		1	2	4	6	8	10	12	13	15	19	23	27	31	35	46	56	67	78	89
240	0	1	3	5	7	9	11	13	15	17	21	26	30	35	39	50	62	74	85	97
260	0	1	3	5	8	10	12	14	17	19	24	28	33	38	43	55	68	80	93	106
280	0	2	4	6	8	11	13	16	18	21	26	31	36	41	46	60	73	87	101	114
300	0	2	4	7	9	12	14	17	20	22	28	33	39	45	50	64	79	93	108	123
320	0	2	5	7	10	13	16	18	21	24	30	36	42	48	54	69	85	100	116	132
340	0	3	5	8	11	14	17	20	23	26	32	38	45	51	58	74	90	107	123	140
360	0	3	6	9	12	15	18	21	25	28	34	41	48	55	61	79	96	113	131	149
380	0	3	6	9	13	16	19	23	26	30	37	44	51	58	65	83	102	120	139	158
400	1	4	7	10	14	17	21	24	28	31	39	46	54	61	69	88	107	127	146	166
420	1	4	7	11	14	18	22	26	29	33	41	49	57	65	73	93	113	134	154	175
460	1	4	8	12	16	20	24	28	33	37	45	54	63	71	80	102	124	147	170	192
500	1	5	9	13	18	22	27	31	36	40	50	59	69	78	88	112	136	160	185	210
550	2	6	10	15	20	25	30	35	40	45	55	66	76	87	97	124	150	177	204	232
600	2	7	12	17	22	28	33	39	44	50	61	72	84	95	107	135	165	194	224	253
650	2	8	13	19	24	30	36	42	48	54	66	79	91	104	116	147	179	211	243	275
700	3	8	14	20	27	33	39	46	52	59	72	85	99	112	126	159	194	228	262	297
800	4	10	17	24	31	38	46	53	61	68	83	99	114	129	145	183	222	262	301	341
900	4	12	20	28	36	44	52	61	69	78	95	112	129	146	164	207	251	296	340	385
1000	5	13	22	31	40	49	59	68	77	87	106	125	144	164	183	232	280	330	379	429

TABLE 7
 TABLE FOR EVALUATING SAMPLE RESULTS—5% RISK OF CONCLUDING
 OCCURRENCE RATE IS LOWER THAN IT ACTUALLY IS

Sample Size	Number of Observed Occurrences Achieved Upper Precision Limit: % Rate of Occurrence																																												
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40	45																									
10																			0	1																									
20												0							1	2	3	4																							
30										0			1						2	3	4	5	7	8																					
40								0			1			2					3	5	6	8	10	12																					
50						0				1			2	3	4	5	7	9	11	13	16																								
60					0			1			2	3	4	5	6	9	11	14	17	20																									
70					0		1		2		3	4	5	7	8	11	14	17	20	24																									
80				0		1		2		3	4	5	7	8	9	13	16	20	24	28																									
90				0		1	2		3	4	5	6	8	9	11	15	19	23	27	32																									
100			0		1		2	3	4		6	8	9	11	13	17	22	26	31	36																									
120			0	1		2	3	4	5	6	8	10	12	14	16	21	27	33	38	44																									
140			0	1	2	3	4	5	6	7	10	12	14	17	19	26	32	39	46	52																									
160		0	1	2	3	4	5	6	8	9	12	14	17	20	23	30	38	45	53	61																									
180		0	1	2	3	5	6	8	9	11	14	17	20	23	26	35	43	52	60	69																									
200		0	1	3	4	6	7	9	11	12	16	19	23	26	30	39	48	58	68	77																									
220		0	2	3	5	7	8	10	12	14	18	22	25	29	33	44	54	64	75	86																									
240		1	2	4	6	8	10	12	14	16	20	24	28	33	37	48	59	71	83	94																									
260		1	3	4	7	9	11	13	15	17	22	26	31	36	41	53	65	77	90	103																									
280		1	3	5	7	10	12	14	17	19	24	29	34	39	44	57	71	84	98	111																									
300	0	1	3	6	8	11	13	16	18	21	26	31	37	42	48	62	76	91	105	120																									
320	0	2	4	6	9	11	14	17	20	22	28	34	40	45	51	66	82	97	113	128																									
340	0	2	4	7	10	12	15	18	21	24	30	36	42	49	55	71	87	104	120	137																									
360	0	2	5	8	10	13	17	20	23	26	32	39	45	52	59	76	93	110	128	146																									
380	0	2	5	8	11	14	18	21	24	28	34	41	48	55	62	80	98	117	135	154																									
400	0	3	6	9	12	15	19	22	26	29	37	44	51	59	66	85	104	123	143	163																									
420	0	3	6	9	13	16	20	24	27	31	39	46	54	62	70	90	110	130	151	171																									
460	0	4	7	11	15	18	22	26	31	35	43	51	60	68	77	99	121	143	166	188																									
500	1	4	8	12	16	21	25	29	34	38	47	56	66	75	84	108	132	157	181	197																									
550	1	5	9	14	18	23	28	33	38	43	53	63	73	83	94	120	146	173	200	227																									
600	1	6	10	15	20	26	31	36	42	47	58	69	80	92	103	132	161	190	219	249																									
650	2	6	12	17	23	28	34	40	46	52	64	76	88	100	112	143	175	207	239	271																									
700	2	7	13	19	25	31	37	43	50	56	69	82	95	108	122	155	189	223	258	292																									
800	3	9	15	22	29	36	43	51	58	65	80	95	110	125	141	179	218	257	296	336																									
900	4	10	18	26	34	42	50	58	66	74	91	108	125	142	159	203	247	291	335	379																									
1000	4	12	20	29	38	47	56	65	74	84	102	121	140	159	178	227	275	324	374	423																									

TABLE 8
TABLE FOR EVALUATING SAMPLE RESULTS—1% RISK OF CONCLUDING
OCCURRENCE RATE IS LOWER THAN IT ACTUALLY IS

Sample Size	Number of Observed Occurrences Achieved Upper Precision Limit: % Rate of Occurrence																							
	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30	35	40	45				
10																						0		
20																0	1			2	3			
30														0		1	3	4	5	6				
40											0		1		2	3	5	7	8	10				
50									0			1	2		3	5	7	9	11	13				
60							0			1	2	3			4	7	9	12	14	17				
70							0			1	2	3	4	5	6	9	11	14	18	21				
80						0			1		2	4	5	6	7	10	14	17	21	25				
90					0			1		2	3	5	6	7	9	12	16	20	24	29				
100					0		1		2	3	4	6	7	9	10	14	19	23	28	33				
120				0		1	2		3	4	6	8	9	11	13	18	24	29	35	40				
140				0	1	2	3		4	5	7	10	12	14	16	22	29	35	42	48				
160			0		1	2	3	5	6	7	9	12	14	17	20	27	34	41	49	56				
180			0	1	2	3	4	6	7	8	11	14	17	20	23	31	39	47	56	65				
200			0	1	3	4	5	7	8	10	13	16	19	23	26	35	44	54	63	73				
220			0	2	3	5	6	8	10	11	15	18	22	26	30	39	50	60	70	81				
240	0	1	2	4	6	7	9	11	13	17	21	25	29	33	44	55	66	78	89					
260	0	1	3	5	6	8	10	12	14	19	23	27	32	36	48	60	72	85	97					
280	0	2	3	4	7	9	12	14	16	21	25	30	35	40	53	65	79	92	106					
300	0	2	4	6	8	10	13	15	18	23	28	33	38	43	57	71	85	99	114					
320	0	2	4	7	9	11	14	17	19	24	30	35	41	47	61	76	91	107	122					
340	1	3	5	7	10	13	15	18	21	26	32	38	44	50	66	82	98	114	131					
360	1	3	6	8	11	14	16	19	22	28	35	41	47	54	70	87	104	122	139					
380	1	3	6	9	12	15	18	21	24	30	37	44	50	57	75	93	111	129	148					
400	1	4	7	10	13	16	19	22	26	32	39	46	54	61	79	98	117	136	156					
420	2	4	7	10	14	17	20	24	27	35	42	49	57	64	84	103	124	144	164					
460	0	2	5	8	12	15	19	23	27	31	39	47	55	63	72	93	114	136	159	181				
500	0	3	6	10	13	17	21	26	30	34	43	52	60	70	79	102	125	149	174	198				
550	0	3	7	11	15	20	24	29	34	38	48	58	68	78	88	113	139	166	192	219				
600	0	4	8	13	17	22	27	32	37	43	53	64	78	86	97	125	153	182	211	241				
650	0	4	9	14	19	25	30	36	41	47	58	70	82	94	106	136	167	198	230	262				
700	1	5	10	16	21	27	33	39	45	51	64	76	89	102	115	148	181	215	249	283				
800	1	7	13	19	25	32	39	46	53	60	74	89	103	118	133	171	209	248	287	326				
900	2	8	15	22	29	37	45	53	61	69	85	101	118	135	152	194	237	281	325	369				
1000	2	9	17	25	34	42	51	60	69	78	96	114	133	151	170	218	266	314	363	412				

TABLE 9
EXAMPLES SHOWING EFFECT ON COMPUTED UPPER OCCURRENCE RATE OF CHANGING RISK, SAMPLE SIZE, AND NUMBER OF DEVIATIONS FOUND

Case	Risk of Concluding the Occurrence Rate Is Lower Than It Actually Is	Sample Size	No. of Deviations	Computed Upper Occurrence Rate
1	5%	50	1	9.1%
2	5	100	2	6.2
3	5	150	3	5.1
4	5	200	4	4.5
5	10	100	2	5.2
6	10	100	3	6.6

upper occurrence rate (called the *computed upper occurrence rate*), these steps should be followed:

1. Locate the table that corresponds to the risk of concluding that the occurrence rate is lower than it actually is that was specified earlier.
2. Locate the actual sample size at the left of the table.
3. Look across the row (identified in step 2) to find the actual number of occurrences found when examining the sample
4. Look to the top of that column (identified in step 3) to read the computed upper occurrence rate.

Librarians, too, may find attribute sampling useful to make inferences about characteristics such as the portion of books misclassified, the error rate in the catalog record, or the portion of books missing.

Using table 7, the librarian would conclude that the maximum percent of books missing (computed upper occurrence rate) is 4 percent. Comparing the computed upper occurrence rate to the maximum tolerable rate indicates that the records meet the librarian's criteria. If the computed upper occurrence rate is less than or equal to the tolerable occurrence rate, the librarian may statistically conclude that the records are satisfactory. Earlier, the librarian specified a

tolerable occurrence rate of 4 percent. Hence, the librarian may conclude that the records are okay unless the qualitative aspects of the occurrence should be considered. In contrast, had the librarian found two occurrences, the librarian should have concluded that the maximum occurrence rate was 6 percent, which exceeds the tolerable rate. Two occurrences yield a significantly higher occurrence rate than the librarian originally expected. Hence, because the actual occurrence rate is much greater than anticipated, a librarian may choose to expand the sample. We will discuss qualitative aspects of occurrences later.

Table 9 presents a series of cases which show the effect on the computed upper deviation rate of changing the number of occurrences found and risk.

Consider the Qualitative Aspects of Deviations

Before drawing a conclusion about the results of the sample, one should consider the qualitative characteristics of any occurrence found. Sometimes deviations in the sample may signal that unexamined population items include many occurrences or deviations. For example, deviations may occur because an employee was untrained or an employee's personal problems have resulted in less than quality performance.

CONCLUSION

Attribute sampling is a technique widely used by CPAs when auditing to make inferences about a population they

want to know about but cannot afford the time or cost to examine all items in the population. Librarians, too, may find attribute sampling useful to make inferences about characteristics such as the portion of books misclassified, the error rate in the catalog record, or the portion of books missing. Attribute sampling techniques provide a basis for making defensible statements about an attribute of a population. This article describes

techniques auditors utilize that librarians may apply when sampling to estimate occurrence rates. These techniques also enable librarians to evaluate the risk of making an incorrect estimate. While that risk always exists, the tables used for determining initial sample size and evaluating results enable the sampler to control the risk of concluding that the occurrence rate is lower than it actually is.

REFERENCES

1. D. M. Guy and D. R. Carmichael. *Audit Sampling: An Introduction to Statistical Sampling in Auditing*. New York: Wiley, 1986.
2. Carl M. Drott. "Random Sampling: A Tool for Library Research." *College & Research Libraries* 30 (March 1969) 119-25.
3. Richard M. Dougherty and Fred J. Heinritz with the assistance of Neal K. Kaske. *Scientific Management of Library Operations*. Metuchen, N.J. and London: Scarecrow, 1982.
4. Ronald R. Powell. *Basic Research Methods for Librarians*, 2d ed. Norwood, N.J.: Ablex, 1991.
5. D. M. Roberts. *Statistical Auditing*. New York: American Institute of Certified Public Accountants, 1978.
6. I. S. Simpson. *Basic Statistics for Librarians*, 3d ed. London: Library Association, 1988.

Continuous Tone Filming

Like embrittled books, collections of photographs and negatives are disappearing from use or age. To preserve these collections, MAPS developed a continuous tone filming technique offering superior resolution (up to 200 lpm) while meeting all preservation standards. The high-quality images (15-19 gray scale steps) can be converted to Kodak Photo CD format for expanded access.

Is your collection
as well-preserved
as Aunt Vannis?

MAPS Highest Quality
Preservation Microfilming

(215) 758-8700
FAX (215) 758-9700

