A primer on statistical inferences for finite populations

Thomas R. Knapp
University of Rochester, tomknapp829@gmail.com

Recommended Citation

INVITED ARTICLE

A Primer on Statistical Inference for Finite Populations

Thomas R. Knapp
University of Rochester
Rochester, NY

This primer is intended to provide the basic information for sampling without replacement from finite populations.

Keywords: Finite populations, sampling without replacement

Introduction

The traditional approach to statistical inference based on simple random sampling with replacement from infinite normal population distributions, and employing corrections for finite populations when necessary, is backwards. Instead, concentrate on sampling without replacement from any finite population distribution and then see what happens when the sampling is with replacement and the population is infinite and normally distributed. This primer is an attempt to make sampling without replacement from finite populations both understandable and convincing.

Note that sampling without replacement is carried out within each sample. Sampling between samples must be with replacement; otherwise many finite populations would soon be depleted in the sampling process.

Are most populations infinite or finite? Real world populations, whether of people or of objects, are all finite, no matter how small or how large. How are samples drawn? Real world samples are all drawn without replacement.

Necessary Background

In order to understand what follows, some familiarity with permutations, combinations, and probability, and with the content of a traditional first course in statistics, should be sufficient.

A "Bare Bones" Example

Consider a population consisting of the observations 3,6,6,9,12, and 15.

a. It has a frequency distribution. Here it is:

Observation	Frequency
3	1
6	2
9	1
12	1
15	1

b. It has a mean of $(3 + 6 + 6 + 9 + 12 + 15) / 6 = 51/6 = 8.50$.
c. It has a median of 7.50 (if we "split the difference" between the middle two values).
d. It has a mode of 6 (there are more 6s than anything else).
e. It has a range of $15 - 3 = 12$.
f. It has a variance of $[(3 - 8.5)^2 + 2(6 - 8.5)^2 + (9 - 8.5)^2 + (12 - 8.5)^2 + (15 - 8.5)^2] / 6 = 97.50 / 6 = 16.25$.
g. It has a standard deviation of $\sqrt{16.25} = 4.03$.

It has other interesting summary measures, but those should suffice for now.

Consider taking all possible samples of size three from a population of six observations, without replacing an observation once it is drawn. For the 3, 6, 6, 9, 12, 15 population they are:

(3,6,6); (3,6,9); another (3,6,9); (3,6,12); another (3,6,12); (3,6,15); another (3,6,15); (3,9,12); (3,9,15); (3,12,15); (6,6,9); (6,6,12); (6,6,15); 6,9,12; another (6,9,12); (6,9,15); another (6,9,15); (6,12,15); another (6,12,15); and (9,12 15).

There are 20 such samples. Suppose you would like to estimate the mean of that population by using one of those samples. The population mean (see above) is 8.50.

The mean of (3,6,6) is 5; the mean of (3,6,9) is 6; ...the mean of (9,12,15) is 12.

The possible sample means are 5, 6, 6, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10, 11, 11, and 12 (trust me). The frequency distribution of those means is the sampling distribution for samples of size three taken from the 3, 6, 6, 9, 12, 15 population. Here it is:

Observation	Frequency
3	1
6	2
9	1
12	1
15	1

Ten are under-estimates, by various amounts; ten are over-estimates, also by various amounts. But the mean of those means (do you follow that?) is 8.50 (the population mean). Nice. The problem is that in real life if you have just one of those samples (the usual case) you could be lucky and come close to the population mean or you could be "way off." That's what sampling is all about.

Note: If we were interested in the range instead of, or in addition to, the mean, the possible sample ranges are 3, 6, 6, 9, 9, 12, 12, 9, 12, 12, 3, 6, 9, 6, 6, 9, 9, 9, 9, and 6. The population range is 12. A sample range could be less than or equal to 12 but could never be greater than 12.

Williams (1978) had also used a very small population (nine taxpayers) to introduce the concept of sampling without replacement from finite populations. Here is one of his examples:

What is Known and what is Unknown

It is important to understand that in sampling without replacement with respect to a particular variable (e.g., height), the population size, some parameter of interest (e.g., the population mean height), and the sample size, are all known or easily calculated, with an optimum sample size sometimes determined. What are

unknown, and are usually the focus of the inference from sample to population, are the various possible values for a statistic such as the sample mean height.

In an interesting conference presentation, Petocz (1990) used an example of stars in the sky to illustrate sampling without replacement from a finite population (although very large, the population of stars is finite). The device employed was a photograph of part of a night sky superimposed on a 10×10 grid. For that example, the population size was not only unknown but of principal concern. [It is theoretically possible for someone with particularly good eyesight to come close to counting all of the stars in the photograph, but I wouldn't want to try it!]

Inference for One Mean

The inference problem that is usually considered first in an introductory course in statistics is for a single arithmetic mean, where the sample has been randomly drawn with replacement from an infinite population in which the variable is normally distributed. I have already considered the single-mean case first in this primer (see the example above), but for the case of sampling without replacement from a finite population where the variable has no specified distribution.

Little's Approach

Little (2004) formulated the sample-to-population inference for one mean as a Bayesian type of stratified random sampling problem rather than a simple random sampling problem. Basu's (1971) total-weight-of-elephants example was used to illustrate the approach. Here is the elephant example in my own words and in terms of mean weight rather than total weight:

A circus owner would like to estimate the mean weight of a population of 50 elephants. He has resources sufficient to weigh just one elephant, but he also has records of the weights of all 50 elephants three years ago. The weight of one of the elephants, S, at that time was exactly equal to the mean weight. The circus trainer claims that S's weight is still equal to the mean weight of all the elephants. The owner designates S as the elephant to be weighed now, much to the horror of the circus statistician who insists on the choice being made at random. The owner and the statistician arrived at a compromise: Allot a selection probability of 99/100 to S and a selection probability of 1/4900 to each of the other 49 elephants. They then drew a sample of one elephant. It turned out to be S (no surprise there). The owner was happy, but the statistician was fired and became a teacher of statistics.

The inference from a sample mean to a population mean doesn't come up very often, but the inference from the difference between two sample means to the difference between two population means comes up a lot. We shall consider the latter problem in the next section.

Inference for the Difference Between Two Means

Correlated Samples

Consider first the case of two correlated means for small finite human populations, with the same people measured twice on the same variable or matched pairs measured once on the same variable, and interest in the difference between the two means. For example, we might have the following population data for husband and wife heights, with wife's height subtracted from husband's height:

Pair	Husband's height	Wife's height	Difference ($= H - W$)
A	71 inches	68 inches	+ 3 inches
B	70"	65"	+ 5"
C	69"	62"	+ 7"
D	68"	66"	+ 2"
E	67"	68"	− 1"
F	66"	70"	− 4"
Mean	68.5	66.5	+ 2"

Suppose you were to take a random sample of three out of the six pairs. Just as in the earlier section of this primer, there are 20 such samples. They are ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF, and DEF. You would like to make an inference from the sample to the population.

Just as in traditional statistics, the difference between the mean is the same as the mean of the differences. Also just as in traditional statistics, the problem can be conceptualized as one for a single-mean (of the differences) rather than one for the difference between two means. Here are the differences for all of the possible samples:

Sample	Husband's ht	Wife's ht	Difference	Mean of Differences
ABC	A: 71	A: 68	+ 3	+ 5.00×
	B: 70	B: 65	+ 5	
	C: 69	C: 62	+ 7	
ABD	A: 71	A: 68	+ 3	+ 3.33×
	B: 70	B: 65	+ 5	
	D: 68	D: 66	+ 2	
ABE	A: 71	A: 68	+ 3	+ 2.33×
	B: 70	B: 65	+ 5	
	E: 67	E: 68	− 1	
ABF	A: 71	A: 68	+ 3	+ 1.33×
	B: 70	B: 65	+ 5	
	F: 66	F: 70	− 4	
ACD	A: 71	A: 68	+ 3	+ 4.00×
	C: 69	C: 62	+ 7	
	D: 68	D: 66	+ 2	
ACE	A: 71	A: 68	+ 3	+ 3.00×
	C: 69	C: 62	+ 7	
	E: 67	E: 68	− 1	
ACF	A: 71	A: 68	+ 3	+ 2.00×
	C: 69	C: 62	+ 7	
	F: 66	F: 70	− 4	
ADE	A: 71	A: 68	+ 3	+ 1.33×
	D: 68	D: 66	+ 2	
	E: 67	E: 68	− 1	
ADF	A: 71	A: 68	+ 3	+ 0.33×
	D: 68	D: 66	+ 2	
	F: 66	F: 70	− 4	
AEF	A: 71	A: 68	+ 3	− 0.67×
	E: 67	E: 68	− 1	
	F: 66	F: 70	− 4	

Sample	Husband's ht	Wife's ht	Difference	Mean of Differences
BCD	B: 70	B: 65	+ 5	+ 4.67×
	C: 69	C: 62	+ 7	
	D: 68	D: 66	+ 2	
BCE	B: 70	B: 65	+ 5	+ 3.67×
	C: 69	C: 62	+ 7	
	E: 67	E: 68	− 1	
BCF	B: 70	B: 65	+ 5	+ 2.67×
	C: 69	C: 62	+ 7	
	F: 66	F: 70	− 4	
BDE	B: 70	B: 65	+ 5	+ 2.00
	D: 68	D: 66	+ 2	
	E: 67	E: 68	− 1	
BDF	B: 70	B: 65	+ 5	+ 1.00
	D: 68	D: 66	+ 2	
	F: 66	F: 70	− 4	
BEF	B: 70	B: 65	+ 5	0.00
	E: 67	E: 68	− 1	
	F: 66	F: 70	− 4	
CDE	C: 69	C: 62	+ 7	+ 2.67
	D: 68	D: 66	+ 2	
	E: 67	E: 68	− 1	
CDF	C: 69	C: 62	+ 7	+ 1.67
	D: 68	D: 66	+ 2	
	F: 66	F: 70	− 4	
CEF	C: 69	C: 62	+ 7	+ 0.67
	E: 67	E: 68	− 1	
	F: 66	F: 70	− 4	
DEF	D: 68	D: 66	+ 2	− 1.00
	E: 67	E: 68	− 1	
	F: 66	F: 70	− 4	

8

Here is a list, in decreasing order, of the 20 mean differences:

$+5.00$, $+4.67$, $+4.00$, $+3.67$, $+3.33$, $+3.00$, $+2.67$, $+2.67$, $+2.33$, $+2.00$, $+2.00$, $+1.67$, $+1.33$, $+1.33$, $+1.00$, $+0.67$, $+0.33$, 0.00, -0.67, -1.00

The mean of those means is $+2.00$. The population mean difference is also 2.00. So all is well, except if we drew only one sample (the usual eventuality in real-world research) we might get one of those means that are quite far away from 2.00.

Independent Samples

The independent samples case (the more common situation) was addressed by Pitman (1937) through his "separation test." I shall use an example from the Pitman article to try to explain the process. [The title of his article, "Significance tests which may be applied to samples from any populations" suggests that he makes a very strong claim. He does; and, fortunately, he's correct.]

Pitman's Test

Consider two samples, one of size n_1 and the other of size n_2, where n_1 is less than or equal to n_2. In order to test the statistical significance of the difference between their means you need to determine all of the possible "separations" between the two sets of observations in the total population of observations.

Example (page 122 of Pitman's article): Sample 1 consists of 1.2, 2.3, 2.4, and 3.2, with a mean of 2.275; Sample 2 consists of 2.8, 3.1, 3.4, 3.6, and 4.1, with a mean of 3.400. Are those two means significantly different from one another?

In order to simplify things a bit and without loss of generality, Pitman suggests subtracting 1.2 from each sample value and then multiplying each by 10 in order to have the smallest value equal to 0 and to get rid of the decimal points. We then have 0, 11, 12, and 20 in Sample 1; and 16, 19, 22, 24, and 29 in Sample 2. Putting these in a sequence from smallest to largest we get nine observations 0, 11, 12, 16, 19, 20, 22, 24, and 29, for which the sum is 153 and the mean is 17.

We now consider the means for all of the ways to divide those observations into two groups, with four observations in one of the groups and with the five other observations in the other group. One way is to have the four smallest observations (0, 11, 12, 16) in one of the groups and the five largest observations (19, 20, 22, 24, 29) in the other group. Another way is to have the three smallest observations and

the next smallest observation (0, 11, 12, 19) in one of the groups and the other observations in the other group. A third way is the way things actually turned out in the study itself (0, 11, 12, 20 vs. 16, 19, 22, 24, 29), indicated by an asterisk in the table below. And so forth. Each of those ways is referred to as a "separation." Here are all of the possible separations for the smaller sample; that's all you need, because the corresponding larger sample consists of the remaining observations, and its mean necessarily follows by comparison with the "grand mean" of 17:

Separation	Group 1 (smaller sample)	Sum	Mean
1	0, 11, 12, 16	39	9.75
2	0, 11, 12, 19	42	10.50
*3	0, 11, 12, 20	43	10.75
4	0, 11, 12, 22	45	11.25
5	0, 11, 12, 24	47	11.75
6	0, 11, 12, 29	52	13.00
7	0, 11, 16, 19	46	11.50
8	0, 11, 16, 20	47	11.75
9	0, 11, 16, 22	49	12.25
10	0, 11, 16, 24	51	12.75
11	0, 11, 16, 29	56	14.00
12	0, 11, 19, 20	50	12.50
13	0, 11, 19, 22	52	13.00
14	0, 11, 19, 24	54	13.50
15	0, 11, 19, 29	59	14.75
16	0, 11, 20, 22	53	13.25
17	0, 11, 20, 24	55	13.75
18	0, 11, 20, 29	60	15.00
19	0, 11, 22, 24	57	14.25
20	0, 11, 22, 29	62	15.50
21	0, 11, 24, 29	64	16.00
22	0, 12, 16, 19	47	11.75
23	0, 12, 16, 20	48	12.00
24	0, 12, 16, 22	50	12.50
25	0, 12, 16, 24	52	13.00
26	0, 12, 16, 29	57	14.25
27	0, 12, 19, 20	51	12.75
28	0, 12, 19, 22	53	13.25
29	0, 12, 19, 24	55	13.75

Separation	Group 1 (smaller sample)	Sum	Mean
30	0, 12, 19, 29	60	15.00
31	0, 12, 20, 22	54	13.50
32	0, 12, 20, 24	56	14.00
33	0, 12, 20, 29	61	15.25
34	0, 12, 22, 24	58	14.50
35	0, 12, 22, 29	63	15.75
36	0, 12, 24, 29	65	16.25
37	0, 16, 19, 20	55	13.75
38	0, 16, 19, 22	57	14.25
39	0, 16, 19, 24	59	14.75
40	0, 16, 19, 29	64	16.00
41	0, 16, 20, 22	58	14.50
42	0, 16, 20, 24	60	15.00
43	0, 16, 20, 29	65	16.25
44	0, 16, 22, 24	62	15.50
45	0, 16, 22, 29	67	16.75
46	0, 16, 24, 29	69	17.25
47	0, 19, 20, 22	61	15.25
48	0, 19, 20, 24	63	15.75
49	0, 1920, 29	68	17.00
50	0, 19, 22, 24	65	16.25
51	0, 19, 22, 29	70	17.50
52	0, 19, 24, 29	72	18.00
53	0, 20, 22, 24	66	16.50
54	0, 20, 22, 29	71	17.75
55	0, 20, 24, 29	73	18.25
56	0, 22, 24, 29	75	18.75
57	11, 12, 16, 19	58	14.50
58	11, 12, 16, 20	59	14.75
59	11, 12, 16, 22	61	15.25
60	11, 12, 16, 24	63	15.75
61	11, 12, 16, 29	68	17.00
62	11, 12, 19, 20	62	15.50
63	11, 12, 19, 22	64	16.00
64	11, 12, 19, 24	66	16.50
65	11, 12, 19, 29	71	17.75
66	11, 12, 20, 22	75	18.75

Separation	Group 1 (smaller sample)	Sum	Mean
67	11, 12, 20, 24	77	19.25
68	11, 12, 20, 29	82	20.50
69	11, 12, 22, 24	69	17.25
70	11, 12, 22, 29	74	18.50
71	11, 12, 24, 29	76	19.00
72	11, 16, 19, 20	66	16.50
73	11, 16, 19, 22	68	17.00
74	11, 16, 19, 24	70	17.50
75	11, 16, 19, 29	75	18.75
76	11, 16, 20, 22	69	17.25
77	11, 16, 20, 24	71	17.75
78	11, 16, 20, 29	76	19.00
79	11, 16, 22, 24	73	18.25
80	11, 16, 22, 29	78	19.50
81	11, 16, 24, 29	80	20.00
82	11, 19, 20, 22	72	18.00
83	11, 19, 20, 24	74	18.50
84	11, 19, 20, 29	79	19.75
85	11, 19, 22, 24	76	19.00
86	11, 19, 22, 29	81	20.25
87	11, 19, 24, 29	83	20.75
88	11, 20, 22, 24	77	19.25
89	11, 20, 22, 29	82	20.50
90	11, 20, 24, 29	84	21.00
91	11, 22, 24, 29	86	21.50
92	12, 16, 19, 20	67	16.75
93	12, 16, 19, 22	69	17.25
94	12, 16, 19, 24	71	17.75
95	12, 16, 19, 29	76	19.00
96	12, 16, 20, 22	70	17.50
97	12, 16, 20, 24	72	18.00
98	12, 16, 20, 29	77	19.25
99	12, 16, 22, 24	74	18.50
100	12, 16, 22, 29	79	19.75
101	12, 16, 24, 29	81	20.25
102	12, 19, 20, 22	73	18.25
103	12, 19, 20, 24	75	18.75

Separation	Group 1 (smaller sample)	Sum	Mean
104	12, 19, 20, 29	80	20.00
105	12, 19, 22, 24	77	19.25
106	12, 19, 22, 29	82	20.50
107	12, 19, 24, 29	84	21.00
108	12, 20, 22, 24	78	19.50
109	12, 20, 22, 29	83	20.75
110	12, 20, 24, 29	85	21.25
111	12, 22, 24, 29	87	21.75
112	16, 19, 20, 22	77	19.25
113	16, 19, 20, 24	79	19.75
114	16, 19, 20, 29	84	21.00
115	16, 19, 22, 24	81	20.25
116	16, 19, 22, 29	86	21.50
117	16, 19, 24, 29	88	22.00
118	16, 20, 22, 24	82	20.50
119	16, 20, 22, 29	87	21.75
120	16, 20, 24, 29	89	22.25
121	16, 22, 24, 29	91	22.75
122	19, 20, 22, 24	85	21.25
123	19, 20, 22, 29	90	22.50
124	19, 20, 24, 29	92	23.00
125	19, 22, 24, 29	94	23.50
126	20, 22, 24, 29	95	23.75

The next step is to make a frequency distribution of all of the means.

Mean	Frequency	Relative frequency
9.75	1	1/126 = .008
10.50	1	1/126 = .008
10.75	1	1/126 = .008
11.25	1	1/126 = .008
11.50	1	1/126 = .008
11.75	3	3/126 = .024
12.00	1	1/126 = .008
12.25	1	1/126 = .008
12.50	2	2/126 = .016
12.75	2	2/126 = .016
13.00	3	3/126 = .024

INFERENCE FOR FINITE POPULATIONS

Mean	Frequency	Relative frequency
13.25	2	2/126 = .016
13.50	2	2/126 = .016
13.75	3	3/126 = .024
14.00	2	2/126 = .016
14.25	3	3/126 = .024
14.50	3	3/126 = .024
14.75	3	3/126 = .024
15.00	3	3/126 = .024
15.25	3	3/126 = .024
15.50	3	3/126 = .024
15.75	3	3/126 = .024
16.00	3	3/126 = .024
16.25	3	3/126 = .024
16.50	3	3/126 = .024
16.75	2	2/126 = .016
17.00	3	3/126 = .024
17.25	4	4/126 = .032
17.50	3	3/126 = .024
17.75	4	4/126 = .032
18.00	3	3/126 = .024
18.25	3	3/126 = .024
18.50	3	3/126 = .024
18.75	4	4/126 = .032
19.00	4	4/126 = .032
19.25	5	5/126 = .040
19.50	2	2/126 = .016
19.75	3	3/126 = .024
20.00	2	2/126 = .016
20.25	3	3/126 = .024
20.50	4	4/126 = .032
20.75	2	2/126 = .016
21.00	3	3/126 = .024
21.25	2	2/126 = .016
21.50	2	2/126 = .016
21.75	2	2/126 = .016
22.00	1	1/126 = .008
22.25	1	1/126 = .008
22.50	1	1/126 = .008
22.75	1	1/126 = .008
23.00	1	1/126 = .008

Mean	Frequency	Relative frequency
23.50	1	1/126 = .008
23.75	1	1/126 = .008
	126	1.000

The way the test works is to see where the result for the actual separation (the starred row), which resulted in a mean of 10.75, falls in the distribution of all of the means. The 10.75 is one of the three smallest means, out of 126, which is .024. It is unlikely to have happened by chance when taking two samples of size 4 and size 5 from a population that has a mean of 17. So, the difference between those two means is statistically significant beyond the .05 level.

A few comments regarding Pitman's test:

1. If the two samples are of equal size, it doesn't matter which one is referred to as the smaller one.
2. If there are any ties in the population of observations, i.e., a particular value appears more than once, all of the tied observations must be distinguished from one another and each must be capable of being sampled. In Pitman's delightful way of putting it: "Numbers which are equal in value are supposed to be distinguishable from one another-we may think of the m + n numbers as painted on $m + n$ different marbles." (Pitman, 1937, p. 119)
3. [Most importantly] There is nothing special about two means. The test is sensitive to other differences between the samples, just as the better-known Kolmogorov-Smirnov test is.
4. There's a great website called "Combination N choose K" ("N choose n" in our notation) that generates all of the combinations for you (https://www.dcode.fr/combinations). And approximations to the exact test are available if the calculations get too complicated even for computers.

Inference for One Percentage

You know what a percentage is. Two out of four is 50%; one out of 5 is 20%; etc. A percentage is easily converted into a proportion by removing the % symbol and moving the decimal point two places to the left. A proportion is easily converted into a percentage by multiplying by 100 and affixing a % symbol. For example, 25% is the same as .25. [Note that a proportion and a percentage are both special cases of means. If the observations for a variable consist of 0s and 1s (a so-called "dummy" variable), the proportion of 1s is the sum of all of the observations

divided by the number of them. If the observations consist of 0s and 100s (an admittedly unusual situation), the percentage of 100s is the sum of all of those observations divided by the number of them.]

Two Classic Examples of a Confidence Interval for a Percentage

People in general, and researchers in particular, are often interested in estimating a population percentage from a sample percentage. The quality control expert in a factory that manufactures "widgets" wants to estimate the percentage of defectives in an entire lot by inspecting a relatively small sample of widgets. If the widgets are tiny objects such as thumbtacks, it would be too expensive to inspect every thumbtack in a lot of a thousand or more thumbtacks. So, they draw a sample of, say, 20 thumbtacks, carefully inspects each of those, and determines the number, a, of defectives in that sample. Suppose a turns out to be equal to one, i.e., 5% of the sample. Can they conclude that there are 5% defectives in the entire lot? No. That might the best guess, but it is subject to sampling error because it is based upon a sample and not a population. What needs to be done is to determine how confident they can be in the 5%.

A person who takes an exit poll as voters emerge from a precinct would like to know, before the official results are posted, who voted for which candidate. Suppose 18 out of a sample of 30 voters (60%) say they voted for Smith. Does that mean 60% of all voters at that precinct voted for Smith? No; it's a sample and not a population. Again, what needs to be done is to determine a range of values around the 60% for which the pollster can be highly confident of "capturing" the true population percentage. Such a range is called, naturally enough, a confidence interval.

Other Examples of Confidence Intervals for a Finite Population Percentage

In his one-of-a-kind book, Tommy Wright (1991) provided extensive tables for estimating the number of units, A, in a population of size N that have a particular attribute from the number of units, a, in a sample of size n that have the same attribute. Wright's tables cover possibilities for N of 2 to 2000 and for a from 0 to 200. He gives an early example (p. 8) of $a = 28$ out of $n = 154$, i.e., 18.2%, for an N of 1600. The 95% confidence interval for A ranges from 204 to 397, i.e., from 12.8% to 24.8%. Not bad for a sample size of 154 that is only 9.6% of a population size of 1600.

16

Wright also explains how to use the confidence interval tables to test a hypothesis and to determine an optimum sample size. Just like sampling with replacement from infinite populations, if the hypothesized parameter is inside the 95% confidence interval, for example, it can't be rejected at the .05 significance level. If the hypothesized parameter is outside the interval, it can. In the example in the previous paragraph any hypothesized value for A between 204 and 397 would not be rejected.

His discussion of the determination of an optimum sample size when inferring from a sample percentage to a population percentage (he does it all in terms of proportions, not percentages) is based upon the tolerable width of a confidence interval rather than upon desired power. [That makes sense, given the title of his book.] In the second of two examples he derives an optimum sample size of 80 for the following specifications: $N = 480$; 95% confidence; and an "initial feeling" that A is approximately 25% [sounds Bayesian]. After trying various values of n while keeping in mind that a/n should be somewhere around 25%, the optimum value of n is found to be around 80, with an interval half-width of about 44 for A and about 9% for the population percentage.

In an earlier article, Buonaccorsi (1987) had provided a comparison of two competing methods for establishing a confidence interval for a proportion and gave as a simple example the 90% confidence interval for $N = 10$, $n = 4$, and $a = 0$ through 4. [95% confidence is conventional, but other levels can be chosen, depending upon the seriousness of the inference.]

In a much earlier article, Katz (1953) pointed out that the maximum likelihood point estimate for A is the largest integer less than $a / n (N + 1)$. [Katz actually used m and M rather than a and A.] He then went on to show how to construct a confidence interval around that quantity. Here was one of his examples:

> In a very small sample inquiry, we ask nine persons, randomly selected from a group of 100, whether they are in favor of a certain proposal and we find three in favor six opposed. We wish to construct a 95 per cent confidence interval for the number, M, in the whole group, in favor of the proposal. (p. 259)

He obtained the following confidence interval, using the hypergeometric distribution (see a later section of this primer): $9 < M < 68$. In terms of percentages, the sample percentage of $3/9 = 33.3\%$ yielded a 95% confidence interval whose lower limit was $9/100 = 9\%$ and whose upper limit was $68/100 = 68\%$. That's a fairly wide interval, but n was only 9.

17

One of the Most Interesting "Real World" Finite Populations: the USA

Consider the following data:

The United States (ordered by admission to the union, and with geographical location indicated by $1 =$ east of the Mississippi River and $0 =$ west of the Mississippi River):

1. Delaware (1)
2. Pennsylvania (1)
3. New Jersey (1)
4. Georgia (1)
5. Connecticut (1)
6. Massachusetts (1)
7. Maryland (1)
8. South Carolina (1)
9. New Hampshire (1)
10. Virginia (1)
11. New York (1)
12. North Carolina (1)
13. Rhode Island (1)
14. Vermont (1)
15. Kentucky (1)
16. Tennessee (1)
17. Ohio (1)
18. Louisiana (0)
19. Indiana (1)
20. Mississippi (1)
21. Illinois (1)
22. Alabama (1)
23. Maine (1)
24. Missouri (0)
25. Arkansas (0)
26. Michigan (1)
27. Florida (1)
28. Texas (0)
29. Iowa (0)
30. Wisconsin (1)
31. California (0)

32. Minnesota (0)
33. Oregon (0)
34. Kansas (0)
35. West Virginia (1)
36. Nevada (0)
37. Nebraska (0)
38. Colorado (0)
39. North Dakota (0)
40. South Dakota (0)
41. Montana (0)
42. Washington (0)
43. Idaho (0)
44. Wyoming (0)
45. Utah (0)
46. Oklahoma (0)
47. New Mexico (0)
48. Arizona (0)
49. Alaska (0)
50. Hawaii (0)

A quick count indicates there are 26 states east of the Mississippi River and 24 states west of the Mississippi River. The percentage of states east is therefore $26/50 = 52\%$. The percentage west is, necessarily, 48%.

Suppose the interest is to draw a random sample of five of the fifty states.

There are (trust me) 2,118,760 different samples of size 5 that could be drawn. How are they enumerated to draw some of them? There is a neat website called the Research Randomizer (https://www.randomizer.org/) that does most of the work for you. I got on the site to see how it worked, gave it the numbers 1 through 50, told it I wanted one such example, and it returned to me the following ID numbers: 8, 10, 26, 27, 43, i.e., South Carolina (SC), Virginia (VI), Michigan (MI), Florida (FL), and Idaho (ID). Four of those five states (80%) are east of the Mississippi River. That is an over-estimate, because 52% of the states are east, but SC, VI, MI, FL, and ID are a sample, not the entire population of states.

I then turned to Wright's (1991) tables for $N = 50$, $n = 5$, and $a = 4$, and I found the 95% confidence interval around the 80% to extend from 68% to 99%. The true population percentage of 52% falls outside of that interval, so I had a "bad" sample. In other words, if a finite population of size 50 has 52% of the observations of a particular type, a sample of size 5 is unlikely to yield a sample percentage of 80.

[Did you follow that? If so, congratulations! If not, the other examples to follow should make things clearer.]

More [You Can Tell I Love Percentages and Proportions]

Zieliński (2016) was concerned with the shortest (narrowest) confidence interval for estimating a proportion. He provided the following example:

> Let the size of a population be $N = 1000$. We took a sample of size $n = 100$ and we observed $\xi = 2$ objects with a given property. Let the confidence level be $\delta = 0.95$. ...The shortest confidence interval is (0.0043349, 0.0788678). Its length is 0.0745329. (p. 181)

A sample of size 100 takes a 10% "bite" out of the population of 1000. His ξ is equivalent to Wright's a. There weren't many "successes" (the word "success" as used in statistics can refer to either category of a dichotomous dependent variable), and that's a very tight confidence interval.

Zieliński (2011) had previously been interested in the approximations of the binomial and the normal to the "exact" hypergeometric-based confidence intervals. [See a later section of this primer for a discussion of the hypergeometric distribution.] Here is a segment of the Abstract for that 2011 article:

> Consider a finite population. Let $(0, 1)$ $\theta \in$ denotes the fraction of units with a given property. The problem is in interval estimation of θ on the basis of a sample drawn due to the simple random sampling without replacement. In the paper three confidence intervals are compared: exact based on hypergeometric distribution and two other based on approximations to hypergeometric distribution: Binomial and Normal. It appeared that Binomial based confidence interval is too conservative while the Normal based one does not keep the prescribed confidence level. (p. 177)

The English translation of that abstract is a bit stilted, but I think you get the idea. [The $(0, 1)$ $\theta \in$ notation is equivalent to Wright's A / N.]

The following example is based upon some clever, albeit artificial, data in *Primer of biostatistics* by Stanton A. Glantz (2012). In one section of that book he discusses a number of examples of fairly large, but not infinite, populations, and implicitly treats them all as infinite by not employing finite population corrections.

The examples are for Martians ($N = 200$), with a mean height of 40 cm and a standard deviation of 5 cm; Venusians ($N = 150$), with a mean height of 15 cm and a standard deviation of 2.5 cm; and Jovians ($N = 100$), with a mean height of 37.6 cm and a standard deviation of 4.5 cm. They are all very short creatures!

Let's just consider the Martians. For the entire population of 200 Martians, 50 are left-footed (and 150 are right-footed), i.e., 25% are left-footed, but in real life [and even in Martian life] that is unknown and needs to be estimated. Suppose we draw a random sample of 20 Martians and determine that 6 of them (30%) are left-footed. Using the table on page 184 of Wright's book for $N = 200$, $n = 20$, and $a = 6$ we find that the lower limit of the 95% confidence interval for A is 30 and the upper limit is 99. Since 30 out of 200 is 6% and 99 out of 200 is just under 50%, our "best guess" of 30% is not very precise. But what can you expect for a small sample that takes a small "bite" out of the population? [In his text Glantz doesn't carry out a confidence interval for that example. For all other examples regarding the population of Martians he uses the traditional formulas for sampling with replacement from infinite populations. He shouldn't.]

Inference for the Difference Between Two Percentages

I'm especially fond of inferences from sample percentage differences to population percentage differences, such as the difference between males and females for some dichotomy, e.g., belief in God (yes or no), or the difference between Democrats and Republicans for that same dichotomy. Here are three examples for other differences between two percentages:

Example #1

Krishnamoorthy and Thomson (2002) gave the artificial but realistic quality control example of the percentages of non-acceptable cans produced by two canning machines. [They actually do everything in terms of proportions, but I prefer percentages.] "Non-acceptable" was defined as containing less than 95% of the purported weight on the can. Each machine produced 250 cans. One machine was expected to have an approximate 6% non-acceptable rate and the other machine was expected to have an approximate 2% non-acceptable rate. A sample of size n is to be drawn from each machine. The authors provide tables for determining the appropriate sample size for each machine, depending upon the tolerance for Type I errors (rejecting a true hypothesis) and Type II errors (not rejecting a false hypothesis). For their specifications the appropriate sample size was 136 cans from

each machine for what they called "the Z-test," which was one of three tests discussed in their article and for which the normal sampling distribution is relevant.

Eight non-acceptable cans were produced by Machine 1 in a sample of 137 cans (5.84%). Three non-acceptable cans were produced by Machine 2 in a sample of 110 cans (2.73%). Therefore, $N_1 = N_2 = 250$, $n_1 = n_2 = 110$, $a_1 = 8$, and $a_2 = 3$. The E (for exact hypergeometric) test produced a p-value of .0365. The p-value for the binomial test was .1378. The p-value for the normal approximation was .0224. Therefore, the E-test and the Z-test rejected the null hypothesis at the .05 level of significance, but the binomial test did not reject the null hypothesis.

Example #2

On page 25 of his book, Wright (1991) gives the example of a "conservative" confidence interval for the difference between two As. The number of observations N_1 in Population 1 is 185; the number of observations N_2 in Population 2 is 440; the sample size n_1 for the first population is 35; the sample size n_2 for the second population is 40; there are 11 "successes" a_1 out of 35, i.e. 31.43%, in Sample 1; and there are 19 "successes" a_2 out of 40 in Sample 2, i.e. 47.50%. The lower limit of the confidence interval for $A_1 - A_2$ was found to be -241 and the upper limit was found to be -59. The lower limit for the difference between the corresponding percentages is -43% and the upper limit is 13%.

Example #3

In Chapter 16 of his book, used in his statistics course, Wardrop (2015) provided a hypothetical example of the difference between the percentage of female students at a small college who wore corrective lenses and the percentage of male students at that same college who wore corrective lenses. In the population of 1000 students 600 were females and 400 were males. 140 of the males (35%) wore corrective lenses and 360 of the females (60%) wore corrective lenses, a difference of 25%. A random sample of 10 out of the 600 females was taken and all 400 of the males were sampled. The sample percentages of wearers of corrective lenses were not reported, but Wardrop claimed, rightly so, that it was a bad sampling plan, and the narrative ended there!

We could have tested the difference between two independent percentages by using Pitman's test. but it would have been very difficult to "paint" all of those 0s and 1s.

Inference for the Relationship Between Two Variables

Consider the relationship between two variables X and Y, such as height and weight, education and income, and other interesting pairs. The statistic most commonly employed for investigating relationships between variables is the Pearson product-moment correlation coefficient r, which is a measure of the strength and the direction of linear relationship. It can go from -1 to $+1$, with -1 indicative of a perfect inverse relationship and $+1$ indicative of a perfect direct relationship, but almost all relationships fall between the two endpoints.

Here is a simple, artificial example for a population of five observations:

Observation	X	Y
A	1	2
B	2	5
C	3	3
D	4	1
E	5	4

The Pearson correlation and the Spearman rank correlation in the population are both equal to 0.

Suppose you would like to sample three of those observations from the population of five observations. The number of such samples is equal to the number of combinations of five things taken three at a time, which is 10. They are: ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, and CDE. The samples, the corresponding X, Y data, and the correlations are:

Sample	Data		Correlation
	X	Y	
ABC	1	2	.50
	2	5	
	3	3	
ABD	1	2	$-.50$
	2	5	
	4	1	

Sample	Data		Correlation
	X	*Y*	
ABE	1	2	.50
	2	5	
	5	4	
ACD	1	2	−.50
	3	3	
	4	1	
ACE	1	2	1.00
	3	3	
	5	4	
ADE	1	2	.50
	4	1	
	5	4	
BCD	2	5	−1.00
	3	3	
	4	1	
BCE	2	5	−.50
	3	3	
	5	4	
BDE	2	5	−.50
	4	1	
	5	4	
CDE	3	3	−.50
	4	1	
	5	4	

Four of the sample correlations are .50, four are −.50, one is 1.00, and one is −1.00. The population correlation is 0. But none of the samples got 0.

Here is a more complicated example, for three variables and seven observations:

Observation	X	Y	Z
A	1	3	7
B	2	6	1
C	3	2	2
D	4	7	5
E	5	1	4
F	6	5	6
G	7	4	3

All of the correlations for pairs of variables (X, Y), (X, Z), and (Y, Z) are equal to 0 in this population.

Now suppose you were to take a simple random sample of three observations from the population of seven observations. The number of possible such samples is equal to the number of combinations of seven things taken three at a time, which is 35. Here they are, with the sample correlations for each triplet:

Sample	(X, Y) correlation	(X, Z) correlation	(Y, Z) correlation
ABC	−.240	−.778	−.423
ABD	.891	−.143	−.577
ABE	−.636	−.240	−.596
ABF	.371	.176	−.849
ABG	−.034	−.339	−.929
ACD	.619	−.564	.300
ACE	−1.000	−.596	.596
ACF	.737	−.075	.619
ACG	.655	−.619	.189
ADE	−.052	−.996	.143
ADF	.596	−.596	−1.000
ADG	.240	−1.000	−.240
AEF	.189	−.619	.655
AEG	.143	−.996	−.052
AFG	.778	−.797	−.240
BCD	.189	.961	.454
BCE	−.866	1.000	−.866
BCF	.038	.999	.091
BCG	−.189	.945	−.500
BDE	−.645	.839	−.125
BDF	−.500	.945	−.189

Sample	(X, Y) correlation	(X, Z) correlation	(Y, Z) correlation
BDG	−.737	.397	.327
BEF	−.454	.986	−.300
BEG	−.500	.737	−.954
BFG	−.945	.676	−.397
CDE	−.156	.655	.645
CDF	.434	.891	.795
CDG	.127	.052	.997
CEF	.577	.982	.721
CEG	.655	.500	−.327
CFG	.839	.500	.891
DEF	−.327	.500	.655
DEG	−.327	−.982	.500
DFG	−1.000	−.500	.500
EFG	.721	−.327	.419

As you can see, those correlations are all over the place, but just as for the preceding example, not one of them was equal to the population correlation of 0.

Covariance vs. Correlation

One of the reasons for the popularity of the Pearson correlation is it is "dimensionless," i.e., you don't have to worry about the units of measurement for the variables. This can be seen from one of its many formulas [Rodgers and Nicewander (1988) claimed there were thirteen of them], the average product of standard scores on X and on Y, which are themselves dimensionless. The covariance between two variables X and Y is equally defensible as a measure of their relationship (it is the correlation multiplied by the product of the standard deviation of X and the standard deviation of Y), but it comes out in the units of X and Y. For example, if X is height in inches and Y is weight in pounds, the covariance is in inch-pounds. Sounds strange, doesn't it? [The situation is similar for the standard deviation and the variance. If X is height in inches and Y is weight in pounds, the standard deviation of X is in inches and the standard deviation of Y is in pounds, but the variance of X is in squared inches and the variance of Y is in squared pounds.]

However, there is an important statistical advantage for the covariance. The sample covariance was shown [by Sirotnik and Wellington (1977) in one context, by myself (Knapp, 1979), and by others] to be an unbiased estimator of the

population covariance, but the sample correlation is not an unbiased estimator of the population correlation.

Relationship Between the Heights and the Weights of Martians

In his textbook, Glantz (2012) provided a detailed discussion of the relationship between height and weight for the entire Martian population and for random samples drawn from that population. The Pearson correlation for the population ($N = 200$) is .917. The correlation between height and weight for one sample ($n = 10$) was found to be .925. For another sample of the same size the correlation was found to be .880. So far, so good, no matter whether the population is infinite or finite, and whether the samples have been drawn with or without replacement. But when it comes to inference from sample to population it makes a big difference. [Once again, Glantz uses the traditional formulas for hypothesis testing and for interval estimation without the finite population correction.]

Rank Correlations

For statistical inferences regarding relationships between variables in finite populations, things are sometimes simpler for rank correlations than for Pearson correlations.

Here are some data for the population of our 50 states:

state	admrank	arearank
DE	1	49
PA	2	32
NJ	3	46
GA	4	21
CT	5	48
MA	6	45
MD	7	42
SC	8	40
NH	9	44
VA	10	37
NY	11	30
NC	12	29
RI	13	50
VT	14	43
KY	15	36

state	admrank	arearank
TN	16	34
OH	17	35
LA	18	33
IN	19	38
MS	20	31
IL	21	24
AL	22	28
ME	23	39
MO	24	18
AR	25	27
MI	26	22
FL	27	26
TX	28	2
IA	29	23
WI	30	25
CA	31	3
MN	32	14
OR	33	10
KS	34	13
WV	35	41
NV	36	7
NE	37	15
CO	38	8
ND	39	17
SD	40	16
MT	41	4
WA	42	20
ID	43	11
WY	44	9
UT	45	12
OK	46	19
NM	47	5
AZ	48	6
AK	49	1
HI	50	47

where:

1. state is the two-letter abbreviation for each of the 50 states.
2. admrank is the rank-order of their admission to the union (Delaware was first, Pennsylvania was second,…, Hawaii was fiftieth). [In a previous section I already listed the 50 states in order of admission to the union.]
3. arearank is the rank-order of land area (Alaska is largest, Texas is next largest,…, Rhode Island is the smallest).

Of considerable interest (at least to me) is the relationship between those two variables (admission to the union and land area). I (and I hope you) do not care about the means, variances, or standard deviations of those variables. [Hint: If you do care about such things for this example, you will find that they're the same for both variables.]

The relationship (Spearman's rank correlation) for the population is −.720. The correlation can go from −1 through 0 to +1, where a negative correlation is indicative of inverse relationship and a positive correlation is indicative of direct relationship. That correlation is inverse and rather strong. That makes sense if you think about it and call upon your knowledge of American history.

But what happens if you take samples from this population? I won't go through all possible samples of all possible sizes, but let's see what happens if you take, say, ten samples of ten observations each. And let's choose those samples randomly.

Table 1. Numbers of the states drawn with replacement

Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
1	6	2	1	1	2	2	2	10	12
5	11	7	7	12	3	6	12	20	15
9	21	11	8	19	10	11	17	22	23
17	23	13	10	20	13	14	21	28	24
21	26	23	13	36	17	26	27	29	35
23	27	29	18	39	20	28	34	31	36
25	34	31	19	43	27	30	38	32	37
29	36	34	27	46	35	44	43	42	39
33	39	40	40	48	41	46	48	43	45
48	40	48	46	49	45	50	50	47	47

I got on the internet and used the Research Randomizer. The numbers of the states that I drew for each of those sets of samples are presented in Table 1 (As indicated in a previous section, sampling within sample was without replacement, but sampling between samples was with replacement; otherwise I would have run out of states to sample after taking five samples!)

For the first set (DE, CT, NH, OH, IL, ME, AR, IA, OR, and AZ) the sample data are (the numbers in parentheses are the ranks for the ranks; they are needed because all ranks must go from 1 to the number of things being ranked, in this case 10):

state	admrank	arearank
DE	1(1)	49(10)
CT	5(2)	48(9)
NH	9(3)	44(8)
OH	17(4)	35(7)
IL	21(5)	24(4)
ME	23(6)	39(6)
AR	25(7)	27(5)
IA	29(8)	23(3)
OR	33(9)	10(2)
AZ	48(10)	6(1)

The rank correlation is (using the ranks of the ranks) is −.939. That's not bad (the population correlation is −.720.)

Relationship Between Two Dichotomies

For the remainder of this section I'll show you how to use the difference between two percentages to infer relationships between two dichotomies.

Consider the can example in the previous section taken from Krishnamoorthy and Thomson (2002). The percentage of non-acceptable cans produced by Machine 1 was 5.84. The percentage of non-acceptable cans produced by Machine 2 was 2.73. That difference was found to be statistically significant at the .05 level. Therefore, there must be a statistically significant relationship between Machine Number and Can Acceptability.

The basic principle is as follows: If there is a difference between the percentage of "successes" in Group A and the percentage of "successes" in Group B there must be a relationship between the group variable and the variable that

Table 2. Frequency table for phi

	Machine number		
	1	**2**	
Acceptable	129 (94.2 % of 137)	107 (97.3% of 110)	236
Not	8 (5.8 % of 137)	3 (2.7 % of 110)	11
	137	110	247

determines "success." The statistic that quantifies such a relationship is a variation of the Pearson *r* called a phi coefficient. For the can example, phi is found by setting up the 2 × 2 Table 2 (three cans are unaccounted for).

The formula for phi is

$$\phi = \frac{ad - bc}{\sqrt{efgh}},$$

where *a* is the frequency in the upper-left corner of the table; *d* is the frequency in the lower-right corner; *b* the upper-right; *c* the lower-left; *e* the first row total; *f* the second row total; *g* the first column total; and *h* the second column total. For this example we have 129(3) – 107(8) divided by the square root of 236(11)(137)110, i.e., −.075. That's a small relationship (the sign doesn't really matter), but the difference between the two percentages 2.7% and 5.8% is also small.

The Hypergeometric Distribution

In previous sections of this book there was occasional reference to the word "hypergeometric." It turns out that the hypergeometric formula for probability is the foundation of sampling without replacement from finite populations. It's a bit complicated so I didn't want to introduce it earlier, but here it is:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}},$$

where P is probability, *X* is the number of units that have a particular attribute; *K* is the number of units in the population that have the attribute; *k* is the number of units in the sample that have the attribute; *N* is the population size; and *n* is the sample

31

size. The expressions within the parentheses are the number of combinations of *K* things taken *k* at a time; the number of combinations of *N* − *K* things taken *n* − *k* at a time; and the number of combinations of *N* things taken *n* at a time, respectively. In this primer and in Wright's (1991) tables, *A* is used instead of *K* and *a* is used instead of *k*. [It is quite common to find that different authors choose different symbols for the same things.]

Let's try a couple of examples:

Example #1. What is the probability of two aces in five draws from an ordinary deck of playing cards?

There are 52 cards in the deck, so *N* = 52. Since 5 cards are to be drawn, *n* = 5. Since there are 4 aces in the deck, *K* = 4. Since the desired outcome is 2 aces, *k* = 2. The number of combinations of 52 things taken 5 at a time (the denominator) is 52! / 5!47!, where the symbol ! stands for factorial and in the numerator requires starting with the number 52, multiplying it by 51, multiplying that by 50...all the way down to 1. The numbers in the denominator work the same way: first 5 × 4 × 3 × 2 × 1, then 47 × 46 × 45 ×…× 1. The 47! in the denominator cancels out all but 52 × 51 × 50 × 49 × 48 of the numerator, giving us 52 × 51 × 50 × 49 × 48 divided by 120, which works out to be 2,598,960. [I used the nice calculator in Windows 10.]

We're not done yet. We also need to determine the number of combinations of 4 things taken 2 at a time (that's easy; it's 6) and the number of combinations of 48 things taken 3 at a time (that turns out to be 17,296). Finally multiply the 6 by the 17,296 and divide by the 2,598,960, which gives.04.

For the five-card-stud poker players among you, a hand consisting of two aces and three other cards is a pretty good, but don't expect to get one because its probability is only .04.

Example #2. For the example in Glantz (2012), what is the probability of drawing all left-footed Martians in a sample of size five?

Recall from a previous section that there were 200 Martians altogether, so *N* = 200. 25% of them were left-footed, so *K* = 50. A sample of size 5 is to be drawn, so *n* = 5.

Plugging these numbers into the hypergeometric formula I get .000836. Therefore, if you find yourself on Mars and you take a random sample of five of its inhabitants, be prepared to get all right-footers.

Finite Population Correction Factor

Using the procedures for sampling with replacement from infinite normal populations for problems involving finite populations can be slightly improved by employing the finite population correction whenever a statistical inference is carried out. Its formula for one mean or one percentage or one correlation is

$$\sqrt{\frac{N-n}{N-1}},$$

where N is the population size and n is the sample size, and the formula for the standard error of the statistic is multiplied by it. The net result is a smaller standard error, since the fpc is less than 1, which makes the inference more precise.

A Final Note

It is ironic that the right way to handle real-world populations is often the most difficult, statistically speaking. I have only scratched the surface of the body of available literature on sampling without replacement from finite populations. I would like to believe, however, that I have covered the basics. The next time you have the opportunity to design a study that is concerned with sample-to-population inference, please at least consider using one of the approaches included in this primer.

Always keep in mind the difference between the role of the statistician and the role of the researcher. The statistician tells us what happens when you take samples, and you want to make statistical inferences from sample statistics to population parameters. The researcher (usually) has only one sample, and they are concerned only with the inference from their sample to the population from which it was drawn.

Oh; that reminds me. I said that we should start by using sampling without replacement from finite populations and then move on to sampling with replacement from infinite populations. How can we do that?

The simple answer is "with considerable difficulty." We can no longer specify N, since the population is taken to be of infinite size. One consolation is that for infinite populations it doesn't really matter if the sampling within sample is with or without replacement. If you draw an observation into your sample, put it back in the population, and then draw subsequent observations, it's most unlikely that you'll get the first observation again or any other "repeats." That's another consolation,

and a partial defense for using the traditional approach when you have a large finite population. A third consolation is that many large finite populations have frequency distributions very close to normal [but see the article by Micceri (1989) regarding such a claim].

References

Basu, D. (1971). An essay on the logical foundations of survey sampling, Part one. In V. P. Godambe, & D. A. Sprott (Eds.), *Foundations of statistical inference* (pp. 203-242). Toronto: Holt, Rinehart and Winston.

Buonaccorsi, J. P. (1987). A note on confidence intervals for proportions in finite populations. *The American Statistician, 41*(3), 215-218. doi: 10.1080/00031305.1987.10475484

Glantz, S. A. (2012). *Primer of biostatistics* (7th edition). New York: McGraw-Hill.

Katz, L. (1953). Confidence intervals for the number showing a certain characteristic in a population when sampling is without replacement. *Journal of the American Statistical Association, 48*(262), 256-261. doi: 10.1080/01621459.1953.10483471

Knapp, T. R. (1979). Using incidence sampling to estimate covariances. *Journal of Educational and Behavioral Statistics, 4*(1), 41-58.

Krishnamoorthy, K., & Thomson, J. (2002). Hypothesis testing about proportions in two finite populations. *The American Statistician, 56*(3), 215-222. doi: 10.1198/000313002164

Little, R. J. (2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association, 99*(466), 546-556. doi: 10.1198/016214504000000467

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156-166. doi: 10.1037/0033-2909.105.1.156

Petocz, P. (1990). *Sample space: Practical experiments for teaching statistics*. In the proceedings of the Third International Conference on the Teaching of Statistics, Dunedin, New Zealand. Retrieved from https://iase-web.org/documents/papers/icots3/BOOK1/A4-9.pdf?1402524943

Pitman, E. T. G. (1937). Significance tests which may be applied to samples from any populations. *Supplement to the Journal of the Royal Statistical Society 4*(1), 119-130. doi: 10.2307/2984124

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42*(1), 59-66. doi: 10.1080/00031305.1988.10475524

Sirotnik, K., & Wellington, R. (1977). Incidence sampling: An integrated theory for "matrix sampling". *Journal of Educational Measurement, 14*(4), 343-399. doi: 10.1111/j.1745-3984.1977.tb00050.x

Wardrop, R. L. (2015, May 23). *Statistics 371, blended: Course notes* (Unpublished manuscript). Retrieved from http://pages.stat.wisc.edu/~wardrop/courses/371chapter1-22sum15b.pdf

Wright, T. (1991). *Exact confidence bounds when sampling from small finite universes*. New York: Springer-Verlag. doi: 10.1007/978-1-4612-3140-0

Zieliński, W. (2011). Comparison of confidence intervals for fraction in finite populations. *Metody Ilościowe w Badaniach Ekonomicznych, 12*(1), 177-182.

Zieliński, W. (2016). The shortest confidence interval for proportions in finite populations. *Applicationes Mathematicae, 43*, 173-183. doi: 10.4064/am2297-7-2016