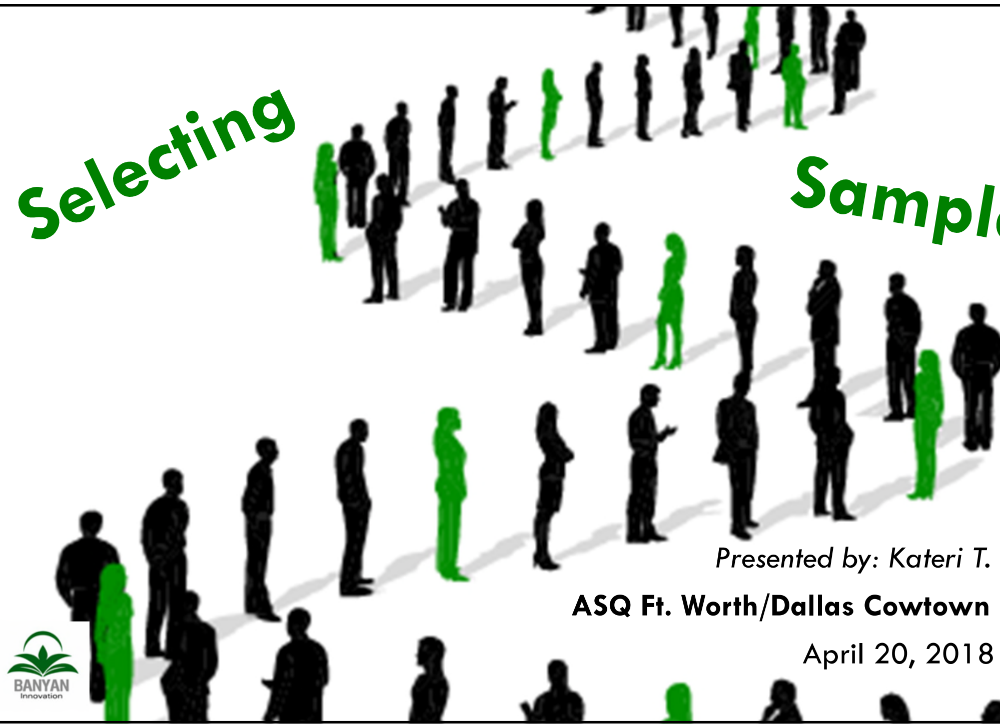
Selecting Sample Sizes



Presented by: Kateri T. Brunell

ASQ Ft. Worth/Dallas Cowtown Quality Roundup

April 20, 2018

BANYAN Innovation

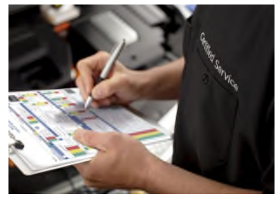
Learning Objectives

- Discuss purposes for sampling
- Review concepts of population, sample, randomness, and sample statistics
- Describe the 5 factors to consider when selecting a sample
- Distinguish between attribute and variables data types
- Explain the formulae for calculating sample sizes for both attribute and variables data
- Discuss finite vs. infinite populations (and effects on sample size)
- Demonstrate and practice using a sample size calculator tool

BANYAN Innovation

Purposes for Sampling

- Surveys

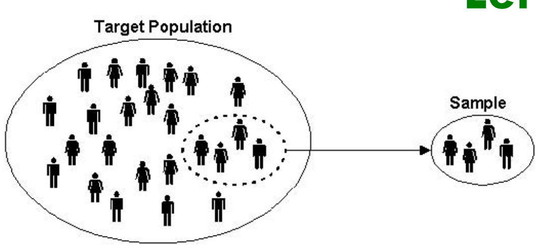


- Baseline measurements

- Process improvements

- Others?

Let's Review Some Concepts



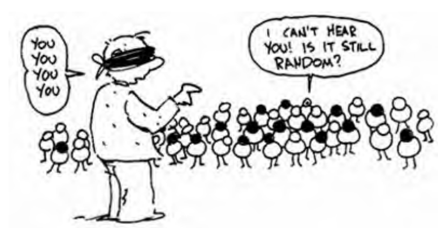
Sample statistics

μ = population mean

σ = population standard deviation

$\bar{x}$ = sample mean

s = sample standard deviation



Random

Independent (Uncorrelated)

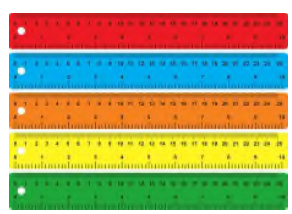
Without Replacement

5 Factors in Sample Size Calculation

1. Population Size (**N**)
2. Proportion of Sample to Population (**n**/**N**)
3. Estimated Population Variance (**s**) or Proportion Occurrence (**p**)
4. Desired Precision (**E**)
5. Confidence Level (**Z**)

Two Different Types of Data

Attribute (Counted)
- Counted
- Proportion of occurrence (p)
- Use Z-distribution
- *Examples: defectives*



Variables (Measured)
- Measured
- Variance or Standard deviation (s)
- Uses t- or Z distribution
- *Example: Average cost of defects*

Formulae for Attribute Data

p (estimated proportion of occurrence in the population)

Z (confidence level)

n (sample size)

E (desired precision)

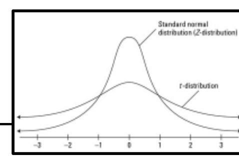
$$E = Z\sqrt{\frac{p(1-p)}{n}}$$

$$n = \frac{Z^2 p(1-p)}{E^2}$$

Need to optimize for *E*

BANYAN Innovation

T-Table and Z



***t* Table**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

BANYAN Innovation

Calculating Sample Size for Attribute Data

A company piloted a new product with 10,000 customers. After the pilot, they wanted to determine the customers' likelihood of buying the product again within the next month ("Yes or No" checked on a brief survey).

They got results from an initial sample of 100 customers. How many more customers should they sample if they want to be 95% confident of the results, within 3%, and their conservative estimate of likely proportion is 50%?

$$E = Z\sqrt{\frac{p(1-p)}{n}}$$

$$E = 1.96\sqrt{\frac{.50(1-.50)}{100}}$$

Z =1.96 p = 0.50 n = 100

p(1-p) = 0.50 (0.50) = 0.2500

0.2500 /n = 0.2500 / 100 = 0.0025

$\sqrt{0.0025}$ = -0.05 -0.05 (Z) = -0.05(1.96) = -0.098

E = -0.098

*Is this **E** precise enough? (NO)*

Calculating Sample for Attribute (cont'd)

A company piloted a new product with 10,000 customers. After the pilot, they wanted to determine the customers' likelihood of buying the product again within the next month ("Yes or No" checked on a brief survey).

They got results from an initial sample of 100 customers. How many more customers should they sample if they want to be 95% confident of the results, within 3%, and their conservative estimate of likely proportion is 50%?

$$n = \frac{Z^2 p(1-p)}{E^2}$$

$$n = \frac{(1.96^2).50(1-.50)}{0.03^2}$$

Z =1.96 p = 0.50 n = 100 E = 0.03

p(1-p) = 0.50 (0.50) = 0.2500

Z^2 = (1.96)(1.96) = 3.84 (3.84)(0.2500) = 0.96

E^2 = (0.03)(0.03) = 0.0009

n= 0.96 / 0.0009 = 1,066.67 ~ **1,067**

n (added) = 1,067 – 100 = **967**

Do We Have a Correct Sample? (NO)

Finite and Infinite Populations

Analytic studies – infinite populations
n/N < 5% of the population

Ongoing, changing process is considered infinite

If population is infinite, use the large population formulae (no correction factor

Enumerative studies – finite populations
n/N => 5% of the population

Ongoing, but unchanging process is finite

Use a **finite population correction factor (fpc)**

Small Population Formulae - Attribute Data

- Finite population correction factor also called "small population formulae
- Must iterate until a sample size is found that optimizes precision, confidence level, estimated proportion, and n/N dynamics

$$E = Z\left(\sqrt{\frac{p(1-p)}{n}}\right) * \left(\sqrt{1-\frac{n}{N}}\right)$$

$$n = \frac{Z^2 p(1-p)}{E^2 + \frac{Z^2 p(1-p)}{N}}$$

Calculating Sample Size -Attribute (small)

Back to our previous example….How many more customers should they sample if they want to be 95% confident of the results, within 3%, and their conservative estimate of likely proportion is 50%?

$$E = Z(\sqrt{\frac{p(1-p)}{n}})*(\sqrt{1-\frac{n}{N}})$$

$$n = \frac{Z^2 p(1-p)}{E^2 + \frac{Z^2 p(1-p)}{N}}$$

Z =1.96   p = 0.50   n = 100

p(1-p) = 0.50 (0.50) = 0.2500

0.2500 /n = 0.2500 / 100 = 0.0025

$\sqrt{0.0025}$ = -0.05

-0.05 (Z) = -0.05(1.96) = -0.098

n/N = 100/10,000 = 0.01

$\sqrt{(1-0.01 = .99}$ = 0.995

E = (-0.098)(0.995) = **-0.098**

BANYAN Innovation

Sample Size -Attribute (small) – cont'd

Back to our previous example….How many more customers should they sample if they want to be 95% confident of the results, within 3%, and their conservative estimate of likely proportion is 50%?

$$n = \frac{Z^2 p(1-p)}{E^2 + \frac{Z^2 p(1-p)}{N}}$$

Z =1.96   p = 0.50   n = 100   E = .03

p(1-p) = 0.50 (0.50) = 0.2500

Z^2 = (1.96)(1.96) = 3.84

E^2 = (0.03)(0.03) = 0.0009

(3.84)(0.2500) = .96   .96/N = .96/10,000 = 0.000096

E^2 + (0.000096) = (0.0009 + 0.000096 = 0.000996

n = (0.96) / (0.000996) = 963.86 ~ 964

n (added) = 964 – 100 = **864**

BANYAN Innovation

Formulae for Variables Data

$$E = Z * \frac{s}{\sqrt{n}}$$

s (estimated population variance)

Z (confidence level)

n (sample size)

E (desired precision)

$$n = \frac{Z^2 s^2}{E^2}$$

Calculating Sample Size for Variables Data

A company piloted a new product with 10,000 customers. After the pilot, they wanted to determine the average time (in hours) from order to delivery.

They randomly sampled 30 customer orders. How many more orders should they sample if they want to be 95% confident of the results, within 3%, and their estimated population variation is 1.0 (hours)?

$$E = Z * \frac{s}{\sqrt{n}}$$

Z =1.96 s = 1.0 n = 30

$\sqrt{30}$ = 5.48 1/5.48 = 0.1825

$$E = 1.96 * \frac{1.0}{\sqrt{30}}$$

1.96 * 0.1825 = 0.358

E = 0.358

*Is this **E** precise enough? (NO)*

Calculating Sample - Variables Data (cont'd)

A company piloted a new product with 10,000 customers. After the pilot, they wanted to determine the average time (in hours) from order to delivery.

They randomly sampled 30 customer orders. How many more orders should they sample if they want to be 95% confident of the results, within 3%, and their estimated population variation is 1.0 (hours)?

$$\mathrm{n} = \frac{Z^2 s^2}{E^2}$$

Z =1.96 s = 1.0 n = 30 E = 0.03

(1.96)(1.96) = 3.84 (1)(1) = 1

E^2 = (0.03)(0.03) = 0.0009

$$n = \frac{(1.96^2)(1^2)}{0.03^2}$$

3.84/0.0009 = 4,266.67

Do We Have a Correct Sample (????)

n = ~ 4,267

Small Population Formulae - Variables Data

$$\sqrt{1 - \frac{30}{10{,}000}} \quad \sqrt{1 - 0.003} \quad \sqrt{0.997}$$

0.9985

$$E = Z * \frac{s}{\sqrt{n}} * \sqrt{1 - \left(\frac{n}{N}\right)}$$

E = 1.96 * 0.1825 * 0.9985 = 0.3572

(1.96)(1.96) = 3.84 (1)(1) = 1

E^2 = (0.03)(0.03) = 0.0009

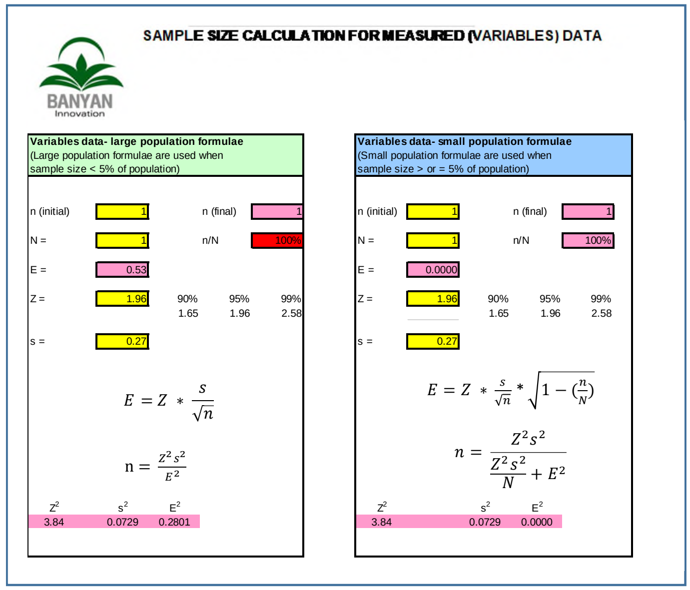
(3.84/10,0000) + .0009 = 0.001284

n = 3.84 / 0.001284 = 2,990.7 ~ **2,991**

$$n = \frac{Z^2 s^2}{\frac{Z^2 s^2}{N} + E^2}$$

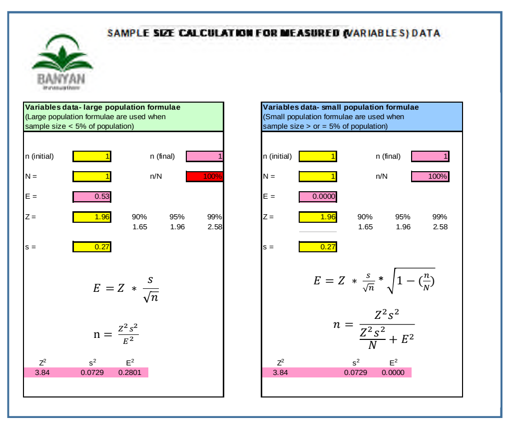
Sample Size Calculator

- Excel workbook
- Formulae programmed into pink cells
- Inputs are in yellow cells
- Use Z/t table to select confidence level if different from 90%, 95% or 99%



Example Using Calculator

- Jump to Excel sample size calculator



Now You Try It

1.
Population (N) = 10,000
Initial sample (n) = 50
Estimated proportion (p) = 0.50
Desired precision (E) = 0.03
Confidence level = 95%

2.
Population (N) = 5,000
Initial sample (n) = 500
Estimated proportion (p) = 0.25
Desired precision (E) = 0.03
Confidence level = 95%

3.
Population (N) = 1000
Initial sample (n) = 50
Estimated variation (s) = .25
Desired precision (E) = 0.03
Confidence level = 95%

4.
Population (N) = 1000
Initial sample (n) = 30
Estimated variation (s) = .05
Desired precision (E) = 0.03
Confidence level = 99%

Summary

- ✓Discussed purposes for sampling
- ✓Reviewed concepts of population, sample, randomness, and sample statistics
- ✓Described the 5 factors to consider when selecting a sample
- ✓Distinguished between attribute and variables data types
- ✓Explained the formulae for calculating sample sizes for both attribute and variables data
- ✓Discussed finite vs. infinite populations (and effects on sample size)
- ✓Demonstrated and practiced using a sample size calculator tool

Questions?

For more information and additional resources:

www.banyan-innovation.com
kateri@banyan-innovation.com
(561) 352-5070