SAMPLE SIZE

1. Introduction:
   - Questions:
     - How large should my sample be?
   - Answer:
     - It depends…
       - …large enough to be an accurate representation of the population.
       - …large enough to achieve statistically significant results.

2. Determining sample size:
   - **What is sample size** that would be required to make reasonably precise generalizations with confidence?
   - A reliable and valid sample should enable us to generalize the findings from the sample to the population under investigation.
   - The sample statistic (statistic finding) should be reliable estimates and reflect the population parameter (actual finding) as closely as possible within a narrow margin of error.
   - **Precision**:
     - Precision refers to how close our estimate is to the true population characteristic.
     - Normally, the greater the precision required, the larger is the sample size needed.
   - **Confidence**:
     - Confidence denotes how certain we are that our estimate will really hold true for the population.

  - Confidence reflect the level of certainty with which we can state that our estimates of the population parameters, based on our sample statistics, hold true.
  - Level of confidence can range from 0 to 100%.
  - A level of confidence of 95% is conventionally acceptable.
- Sample size is function of…
  - Variability (heterogeneity) in the population
    - The more variance we find, the bigger the sample should be
  - Precision or accuracy needed
    - The more precise or accurate we want, the bigger the sample size should be
  - Confidence level desired
    - The higher the confidence level we want, the bigger the sample size should be
  - Type of sampling plan used
    - Different sampling approaches will require different sample size
- Trade-off between confidence and precision
  - If there is little variability in the population, a small sample size will be sufficient to obtain a high confidence and precision level.
  - The higher the precision, the lower will our confidence level be.
  - The higher the confidence level, the lower will our precision level be. → That is why, in both cases, we need bigger sample size to increase the precision and confidence.
- Roscoe proposes the following rules of thumb for determining sample size
  - Sample size larger than 30 and less than 500 are appropriate for most research
  - Where samples are to be broken into sub samples, a minimum sample size of 30 for each category is necessary.

- In multivariate research (including regression analyses) the sample size should be several times (preferably 10 times or more) as large as the number of variables in the study.
- For simple experimental research with tight experimental controls, successful research is possible with samples as small as 10 to 20 in size.

3. The term statistically significant ($p<.05$) is used merely as a way indicating the chances are at least 95 out of 100 that the findings obtained from the sample of people who participate in the study are similar to what the findings would be if one were actually able to carry out the study with the entire population.

4. Sample size for single mean

$n = (Z\ \sigma/\Delta)^2$

n = sample size
σ = population standard deviation
Δ = precision
Z = Z-score at significance level

If there is a possibility of response from 80% of sample population the sample size = n/0.8

5. Sample size for two means

$n = (A + B)^2 * 2\sigma^2 / \Delta^2$

n = sample size
σ = population standard deviation
Δ = expected difference of mean
A = significance level (usually 95%, equals 1.96)
B = power (usually 80%, equals 0.84)

Table of values for A and B:

Significant level A	
5%	1.98
1%	2.58
Power	**B**
80%	0.84
90%	1.28
95%	1.54

6. Sample size for single proportion:

$n = (Z/\Delta)^2 \, p \, (1-p)$

n = sample size

Δ = precision

Z = Z-score at significance level

p = population proportion

7. Sample size for two proportions:

$n = (A + B)^2 * [(p_1 (1-p_1)) + (p_2 (1-p_2))] / (p_1 - p_2)^2$

n = sample size

A = significance level (usually 95%, equals 1.96)

B = power (usually 80%, equals 0.84)

p_1 = first proportion

p_2 = second proportion

Power is the probability that the null hypothesis will be correctly rejected i.e. rejected when there is indeed a real difference or association. It can also be thought of as "100 minus the percentage chance of missing real effect" – therefore the higher the power, the lower the chance of missing a real effect.

Some definition:

Sampling error is the difference of statistically finding between actual parameter of population

Standard error is means of deviation values between two or more groups of sample or population.

Standard deviation is means of deviation values between two or more units of samples or population.