Sample Size Determination in Health Sciences Research

Dr. Wan Nor Arifin

Biostatistics and Research Methodology Unit Universiti Sains Malaysia



wnarifin@usm.my / wnarifin.github.io

Outlines

- Introduction
- General health research:
 - One sample
 - Two samples (independent, paired)
 - Correlation, Regression
- Validation:
 - Exploratory and Confirmatory Factor Analyses
 - Reliability
 - Agreeement
 - Sensitivity & Specificity

Software used

• Sample Size Calculator (web)

https://wnarifin.github.io/ssc_web.html



Introduction

Why calculate?

- An important part of a research design & plan
- Determines:
 - Number of participants
 - Funding
 - Feasibility
 - Validity

Steps toward calculation

- (1) Define the research **objective**
- (2) Decide on the suitable statistical analysis
- (3) Determine the sample size

Define objective

- Clearly define objective
- Outcome (dependent variables)?
- Predictors (independent variables)?
- Measurable

Statistical analysis

- Choose suitable statistical analysis based on:
 - objective itself compare, estimate, correlate, validate ...
 - the scale of outcome & predictors
 - sample independent / dependent
- Examples:
 - Independent *t*-test for ...
 - Chi-squared test for ...
 - Linear regression for ...
 - Logistic regression for ...

Sample size formula

- Choose suitable sample size formula / method
- Based on objective & chosen statistical analysis
- Estimation vs Hypothesis Testing
- Examples:
 - Confidence interval estimation
 - Comparison of means hypothesis testing

General

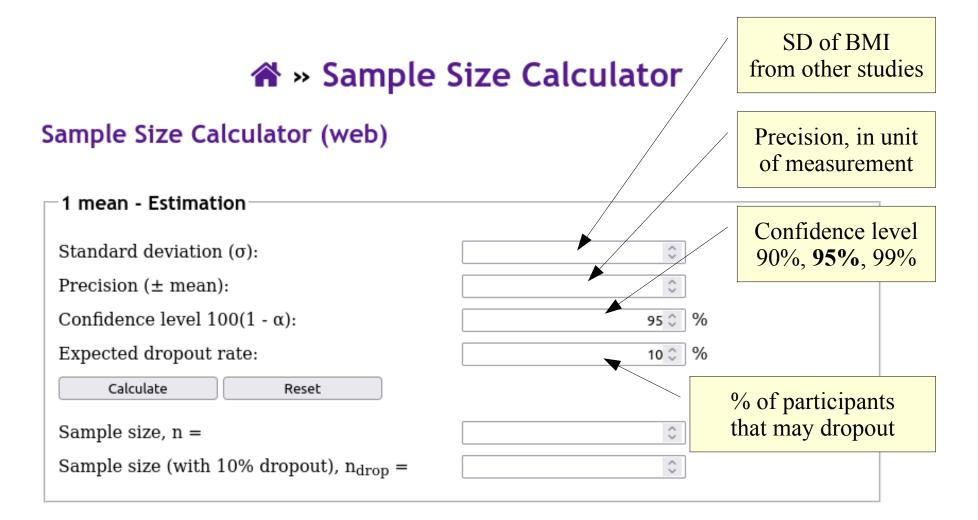
(1) Objective:

- Estimate mean of numerical outcome in a population e.g. BMI, weight etc
- "This study aims to estimate mean BMI among UiTM students."

- (2) Statistical analysis:
- Mean, 95% Confidence Interval (CI)

- (3) Sample size method:
- Single mean Estimation

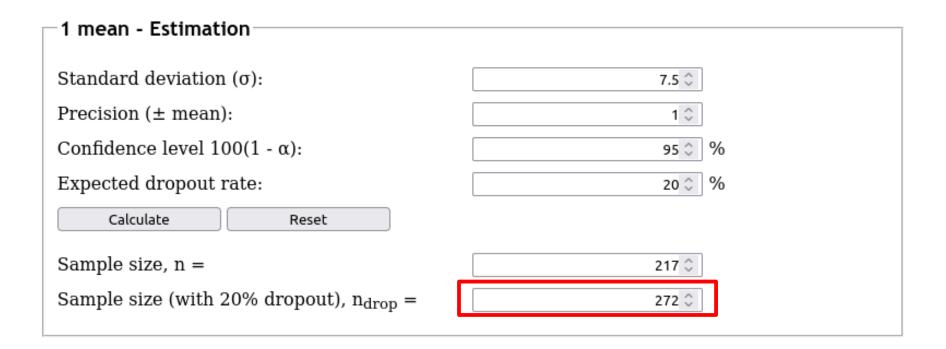
1. Means \rightarrow Single Mean \rightarrow 1 mean – Estimation



- Let say:
 - SD of BMI = 7.5kg/m²
 - Precision = 1kg/m^2
 - 95% Confidence level
 - 20% dropout
- How many UiTM students should we sample?

***** » Sample Size Calculator

Sample Size Calculator (web)



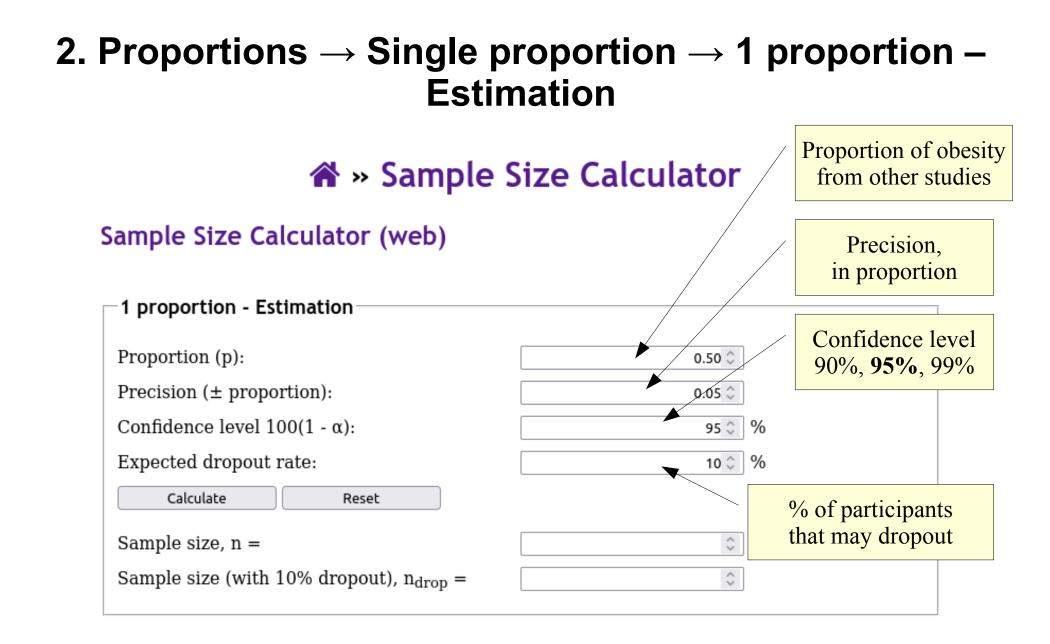
 We have to sample 272 students to estimate mean BMI among UiTM students, with mean BMI ± 1kg/m², taking into account a 20% dropout rate.

(1) Objective:

- Estimate % or proportion of categorical outcome in a population e.g. obesity, diabetes etc
- "This study aims to estimate prevalence of obesity among UiTM students."

- (2) Statistical analysis:
- Proportion, 95% Confidence Interval (CI)

- (3) Sample size method:
- Single proportion Estimation



- Let say:
 - Prevalence of obesity = 25% = 0.25 (in proportion).
 - Precision = $\pm 5\%$ = 0.05 (in proportion).
 - 95% Confidence level.
 - 10% dropout (i.e. those who won't let you measure their weight and height).
- How many UiTM students should we sample?

***** » Sample Size Calculator

Sample Size Calculator (web)

| 1 proportion - Estimation | |
|--|--------|
| Proportion (p): | 0.25 🗘 |
| Precision (± proportion): | 0.05 🗘 |
| Confidence level $100(1 - \alpha)$: | 95 🗘 % |
| Expected dropout rate: | 10 🗘 % |
| Calculate Reset | |
| Sample size, n = | 289 🗘 |
| Sample size (with 10% dropout), $n_{\rm drop}$ = | 322 🗘 |

 We have to sample 322 students to estimate prevalence of obesity among UiTM students, with % of obesity ± 5%, taking into account a 10% dropout rate.

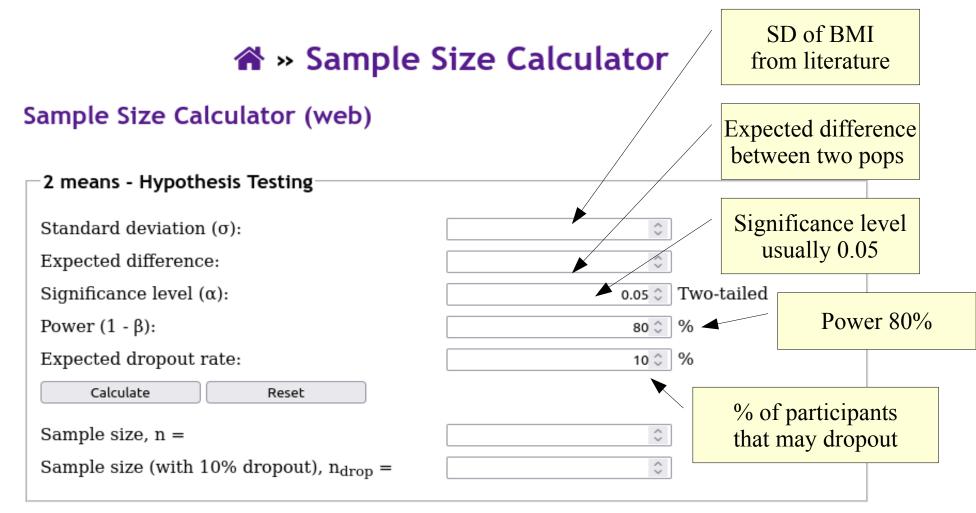
(1) Objective:

- Compare means of a numerical outcome in two populations e.g. BMI, weight etc
- "This study aims to compare mean BMI between 1st and final year students."

- (2) Statistical analysis:
- Independent t-test

- (3) Sample size method:
- Two means Hypothesis testing

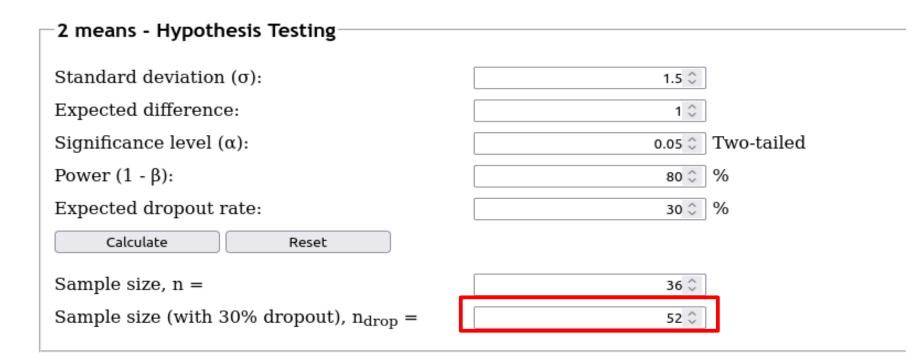
1. Means \rightarrow Two-mean comparison (independent) \rightarrow 2 means – Hypothesis Testing



- Let say:
 - Largest SD you could find from literature = 1.5
 - Expected Difference = 1 unit.
 - Significance level = 5% (0.05)
 - Leave Power = 80% default value.
 - 30% dropout (i.e. as some weight themselves while only one foot was on the scale).
- How many students per group should we sample?

Sample Size Calculator

Sample Size Calculator (web)



• We have to sample 52 1st year students and 52 final year students to make the comparison, expecting a difference of 1 unit BMI between the two, taking into account a 30% dropout rate.

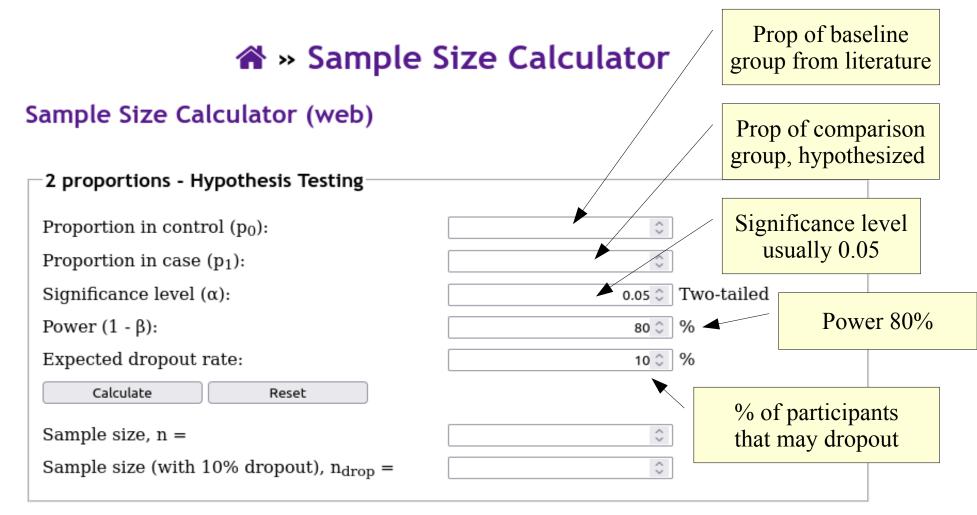
(1) Objective:

- Compare % or proportion of categorical outcome in two populations e.g. obesity, diabetes etc
- "This study aims to compare prevalence of obesity between 1st and final year students."

- (2) Statistical analysis:
- Chi-squared test

- (3) Sample size method:
- Two proportions Hypothesis testing

2. Proportions \rightarrow Two-proportion comparison (independent) \rightarrow 2 proportions – Hypothesis Testing



- Let's say:
 - $p_0 = 35\% = 0.35$ (in proportion) $\rightarrow 1$ st year students as control.
 - p₁ = 50% = 0.5 (in proportion) → Hypothesized % for final year students.
 - Significance level = 5% (0.05)
 - Power = 80%
 - 10% dropout
- How many students per group should we sample?

***** » Sample Size Calculator

Sample Size Calculator (web)

| 2 proportions - Hypothesis Testing | | |
|--|-------------------|--|
| Proportion in control (p ₀): | 0.35 🗘 | |
| Proportion in case (p ₁): | 0.5 🗘 | |
| Significance level (α): | 0.05 🗘 Two-tailed | |
| Power (1 - β): | 80 🗘 % | |
| Expected dropout rate: | 10 🗘 % | |
| Calculate Reset | | |
| Sample size, n = | 170 🗘 | |
| Sample size (with 10% dropout), $n_{drop} =$ | 189 🗘 | |
| | | |

• We have to sample 189 1st year students and 189 final year students to make the comparison, expecting a difference of 15% for % of obesity between the two, taking into account a 10% dropout rate.

Two means (paired)

(1) Objective:

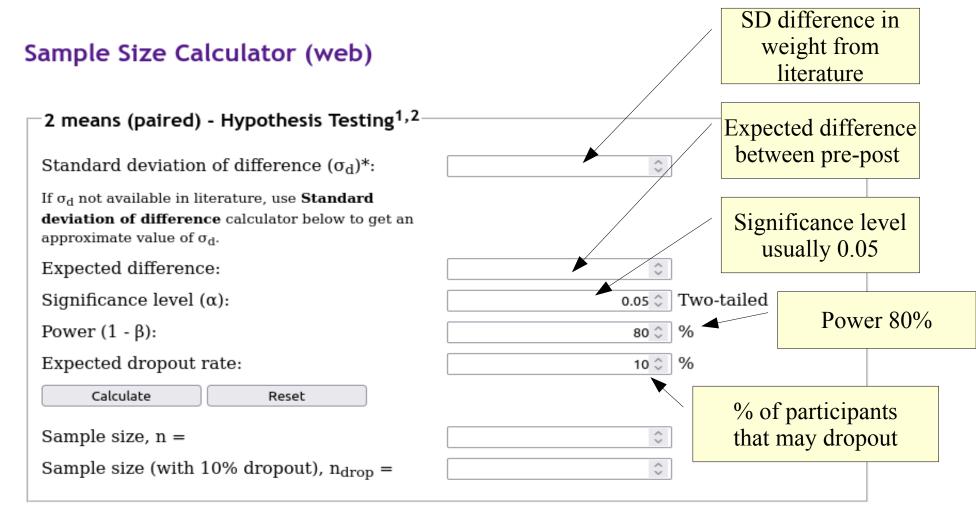
- Compare means of a numerical outcome in a population, pre-post design e.g. weight before and after intervention
- "This study aims to compare mean weight before and after a weight loss program among students."

Two means (paired)

- (2) Statistical analysis:
- Paired t-test

- (3) Sample size method:
- Two means (paired) Hypothesis testing

1. Means \rightarrow Two-mean comparison (paired) \rightarrow 2 means (paired) – Hypothesis Testing



Two means (paired)

- Let say:
 - SD of difference from literature = 2.5kg
 - Expected Difference = 5kg (post pre weight)
 - Significance level = 5% (0.05)
 - Power = 80%
 - 20% dropout
- How many students should we sample?

Two means (paired)

***** » Sample Size Calculator

Sample Size Calculator (web)

| 2 means (paired) - Hypothesis Testing ^{1,2} | |
|--|-------------------|
| Standard deviation of difference $(\sigma_d)^*$: | 7.5 🗘 |
| If σ_d not available in literature, use Standard deviation of difference calculator below to get an approximate value of σ_d . | |
| Expected difference: | 5 🗘 |
| Significance level (α): | 0.05 🗘 Two-tailed |
| Power (1 - β): | 80 🗘 % |
| Expected dropout rate: | 20 🗘 % |
| Calculate Reset | |
| Sample size, n = | 18 🗘 |
| Sample size (with 20% dropout), $n_{drop} =$ | 23 🗘 |

Two means (paired)

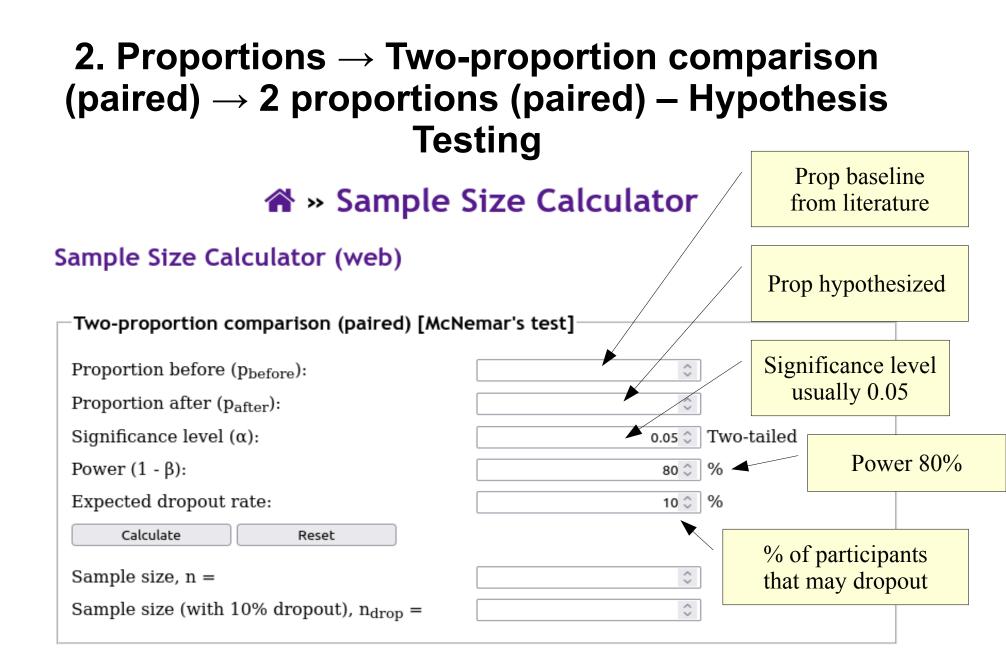
• We have to sample 23 students to make the comparison, expecting 5kg change following the weight reduction program, taking into account a 20% dropout rate.

(1) Objective:

- Compare % or proportion of a categorical outcome in a population, pre-post design e.g. % vaccine uptake after before and after health promotion
- "This study aims to compare percentage of vaccine uptake among villagers before and after a vaccine awareness campaign."

- (2) Statistical analysis:
- McNemar's test

- (3) Sample size method:
- Two proportions (paired) Hypothesis testing



- Let's say:
 - $p_{before} = 50\% = 0.5$ (in proportion) \rightarrow before campaign
 - $p_{after} = 80\% = 0.8$ (in proportion) \rightarrow after campaign
 - Significance level = 5% (0.05)
 - Power = 80% (0.8)
 - 10% dropout
- How many villagers should we sample?

***** » Sample Size Calculator

Sample Size Calculator (web)

| Two-proportion comparison (paired) [McNemar's test] | | | |
|---|-------------------|--|--|
| Proportion before (p _{before}): | 0.5 🗘 | | |
| Proportion after (p _{after}): | 0.8 🗘 | | |
| Significance level (α): | 0.05 🗘 Two-tailed | | |
| Power (1 - β): | 80 🗘 % | | |
| Expected dropout rate: | 10 🗘 % | | |
| Calculate Reset | | | |
| Sample size, n = | 42 🗘 | | |
| Sample size (with 10% dropout), $n_{\rm drop}$ = | 47 🗘 | | |

• We have to sample 47 villagers to make the comparison, expecting 30% change in vaccine uptake following the vaccine awareness campaign, taking into account a 10% dropout rate.

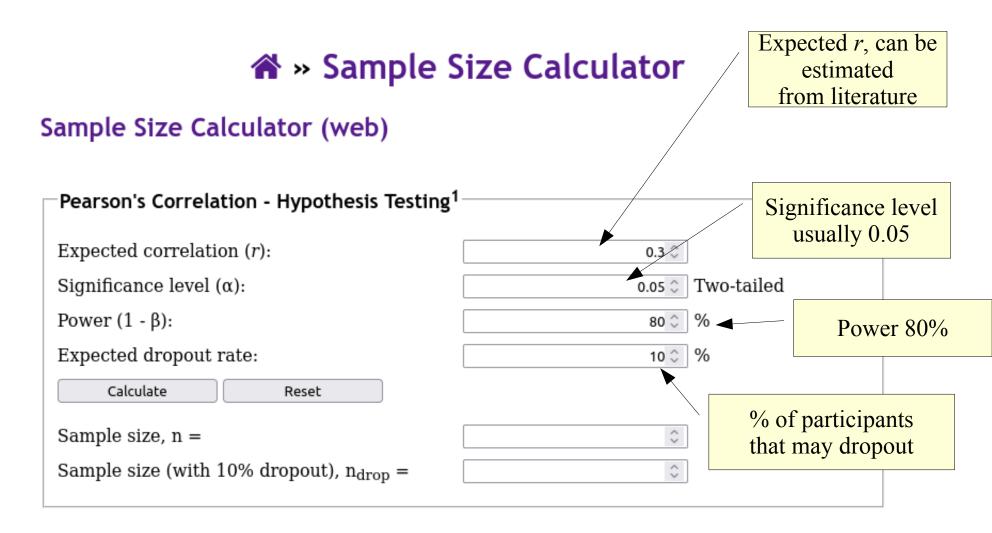
(1) Objective:

- Determine correlation between two numerical outcomes e.g. between age and cholesterol level etc
- "This study aims to determine the correlation between age and cholesterol level among lecturers."

- (2) Statistical analysis:
- Pearson's correlation coefficient

- (3) Sample size method:
- Pearson's correlation Hypothesis testing ($H_0: r = 0$)

3. Pearson's correlation \rightarrow Hypothesis Testing



- Let say:
 - Correlation *r* that you expect, educated guess from literature = 0.55
 - Significance level = 5% (0.05)
 - Power = 80%
 - 30% dropout
- How many lecturers should we sample?

***** » Sample Size Calculator

Sample Size Calculator (web)

| Pearson's Correlation - Hypothesis Testing ¹ | | |
|---|-------------------|--|
| Expected correlation (r): | 0.55 🗘 | |
| Significance level (α): | 0.05 🗘 Two-tailed | |
| Power (1 - β): | 80 🗘 % | |
| Expected dropout rate: | 30 🗘 % | |
| Calculate Reset | | |
| Sample size, n = | 23 🗘 | |
| Sample size (with 30% dropout), $\rm n_{drop}$ = | 33 🗘 | |

• We have to sample 33 lecturers to determine the correlation between age and cholesterol level, taking into account a 30% dropout rate.

(1) Objective:

- Determine associated factors of a numerical outcome, e.g. predictors of cholesterol level etc
- "This study aims to determine the associated factors of cholesterol level among lecturers."

- (2) Statistical analysis:
- Multiple linear regression

- (3) Sample size method:
- Rule-of-thumb 10 subjects per independent variable*

 $- n = k \ge 10$

*Norman, G. R., & Streiner, D. L. (2008). Biostatistics: the bare essentials. Ontario, Canada: BC Decker Inc.

- Let's say:
 - 1 numerical outcome (cholesterol)
 - 3 numerical independent variables:
 - Age in years, BMI in kg/m^2 , weekly physical activity in hours
 - 3 categorical independent variables:
 - Male = yes, no (binary)
 - Smoking = yes, no (binary)
 - Race = Malay, Chinese, Indian (3 levels)
 - 20% dropout
- How many lecturers should we sample?

| Independent Variables, k | Count |
|--|-------|
| Age (numerical) | 1 |
| BMI (numerical) | 1 |
| Weekly physical activity (numerical) | 1 |
| Male (binary categorical) | 1 |
| Smoking (binary categorical) | 1 |
| Race (categorical, 3 levels \rightarrow 2 dummy variables) | 2 |
| Total | 7 |

Formula:

 $n = k \ge 10$

Independent variable count, k = 7

Sample size, n?

 $n = k \ge 10 = 7 \ge 10 = 70$

n + 20% dropout?

 $n_{drop} = n / (1 - dropout proportion)$ = n / (1 - 0.2) = 70 / 0.8 = 87.5

Sample size is 88 (rounded up).

• We have to sample 88 lecturers to investigate the associated factors of cholesterol level among lecturers, taking into account a 20% dropout rate.

Multivariable, binary outcome

(1) Objective:

- Determine associated factors of a binary outcome, e.g. predictors of hypertension (HPT)
- "This study aims to determine the associated factors of HPT among workers in ABC office."

Multivariable, binary outcome

- (2) Statistical analysis:
- Multiple logistic regression
- (3) Sample size method:
- Rule-of-thumb 10 events-per-parameter*
 - Event, $n_1 =$ number with outcome e.g. CAD = yes
 - Parameters = Indep. Variables + (Intercept + Interaction Terms)
 - Sample size, $n = n_1 + n_0$ (w/out outcome) = n_1 / p (prevalence)

*Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). Applied logistic regression (3rd ed.). Hoboken, New Jersey: John Wiley & Sons Inc

- Let's say:
 - 1 binary outcome (HPT)
 - Prevalence of HPT, p = 0.3
 - 3 numerical independent variables:
 - Age in years, BMI in kg/m^2 , weekly physical activity in hours
 - 3 categorical independent variables:
 - Male = yes, no (binary)
 - Smoking = yes, no (binary)
 - Race = Malay, Chinese, Indian (3 levels)
 - 30% dropout
- How many workers should we sample?

| Independent Variables, k | Count |
|--|-------|
| Age (numerical) | 1 |
| BMI (numerical) | 1 |
| Weekly physical activity (numerical) | 1 |
| Male (binary categorical) | 1 |
| Smoking (binary categorical) | 1 |
| Race (categorical, 3 levels \rightarrow 2 dummy variables) | 2 |
| Total | 7 |

Formula:

$$n_{1} = \text{parameters x 10} \\ n_{1} = (k + 1) \times 10 \\ n = n_{1} / p$$
1 intercept
0 interaction

Events, n_1 ? $n_1 = (k + 1) \times 10 = (7 + 1) \times 10 = 80$

Sample size, n? n = $n_1 / p = 80 / 0.3 = 266.7 \approx 267$

 $n_{drop} = n / (1 - dropout proportion)$ = n / (1 - 0.3) = 267 / 0.7 = 381.4 \approx 382

Sample size is 382.

• We have to sample 334 workers to investigate the associated factors of hypertension among workers in ABC office, taking into account a 20% dropout rate.

Validation

(1) Objective:

- Explore internal structure validity / construct validity of a psychometric tool / questionnaire
- "This study aims to explore the internal structure validity of ABC-Q among students."

- (2) Statistical analysis:
- EFA

- (3) Sample size method:
- Rule-of-thumb 5 respondents / item*

*Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10(7).

- Let's say a researcher wants to validate a questionnaire with:
 - 40 items (variables)
 - 30% dropout
- How many respondents should he sample?

Formula:

$$n = k \ge 5$$

 $n = k \ge 5 = 40 \ge 5 = 200$

 $n_{drop} = n / (1 - dropout proportion) = 200 / (1 - 0.3) = 287.7 \approx 288$

Sample size is 288.

• We have to sample 288 students to explore the internal structure validity of ABC-Q among the students, taking into account a 30% dropout rate.

Confirmatory factor analysis

(1) Objective:

- Confirm internal structure validity / construct validity of a psychometric tool / questionnaire
- "This study aims to confirm the internal structure validity of ABC-Q among students."

Confirmatory factor analysis

- (2) Statistical analysis:
- CFA

- (3) Sample size method:
- Rule-of-thumb at least 200 respondents*
- RMSEA, CFI model fit indices Hypothesis testing

*Kline, R. B. (2016). Principles and practice of structural equation modeling. 4th ed. New York: Guilford Publications.

5. Structural Equation Modeling \rightarrow **RMSEA**

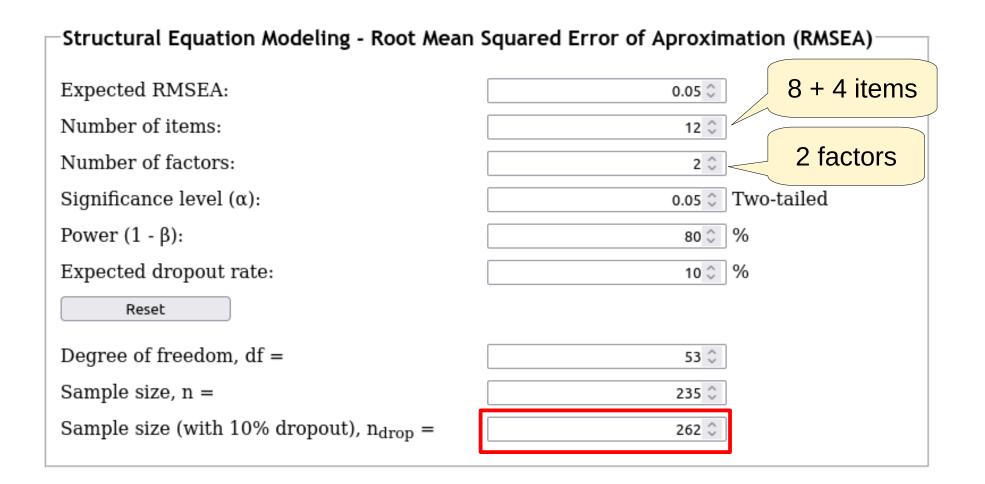
***** » Sample Size Calculator

Usually aim < 0.05Sample Size Calculator (web) Can expect based on literature Structural Equation Modeling - Root Mean Squared Error of Aproximation (RMSEA) Expected RMSEA: 0.05 🗘 Total number of items. Can be unequal per Number of items: $\hat{\mathbf{v}}$ factor. Number of factors: $\hat{\mathbf{v}}$ Significance level (α) : Two-tailed 0.05 🗘 Power $(1 - \beta)$: 80 🗘 % Expected dropout rate: 10 0 % Reset Please input number of item. Please input number of factor.

Confirmatory factor analysis

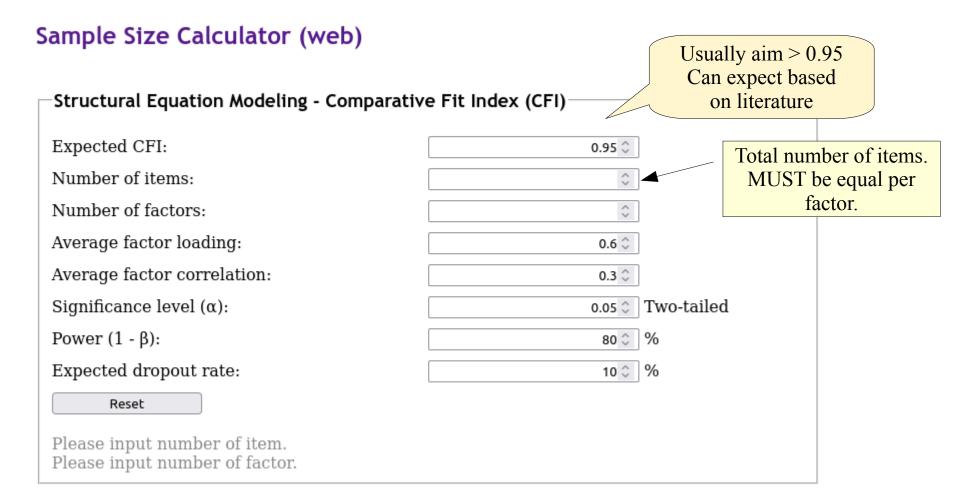
- Let's say a researcher wants to validate a questionnaire with:
 - 2 factors
 - Factor 1: 8 items, Factor 2: 4 items
 - Acceptable RMSEA is set < 0.05.
 - 10% dropout
- How many respondents should he sample?

Confirmatory factor analysis

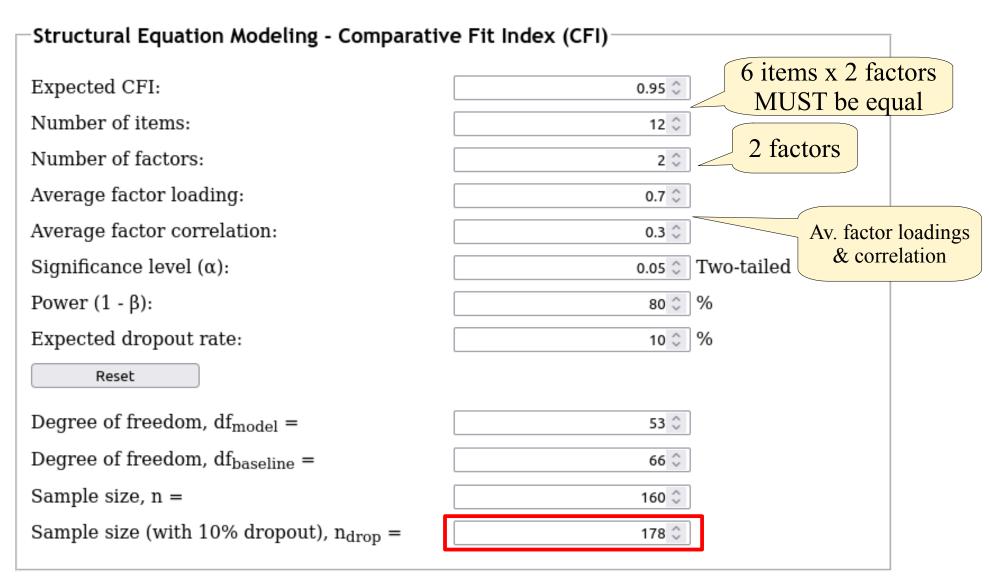


• We have to sample 262 students to confirm the internal structure validity of ABC-Q among the students, taking into account a 10% dropout rate.

5. Structural Equation Modeling → CFI Requires equal number of items per factor ☆ » Sample Size Calculator



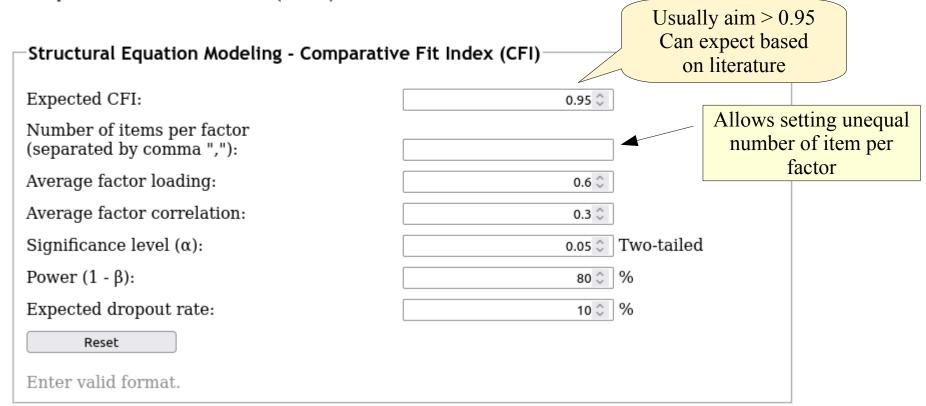
- Let's say a researcher wants to validate a questionnaire with:
 - 2 factors, 6 items each
 - Based on literature, average factor loading is around 0.7
 - Average factor-factor correlation around 0.3
 - Acceptable CFI > 0.95
 - 10% dropout
- How many respondents should he sample?



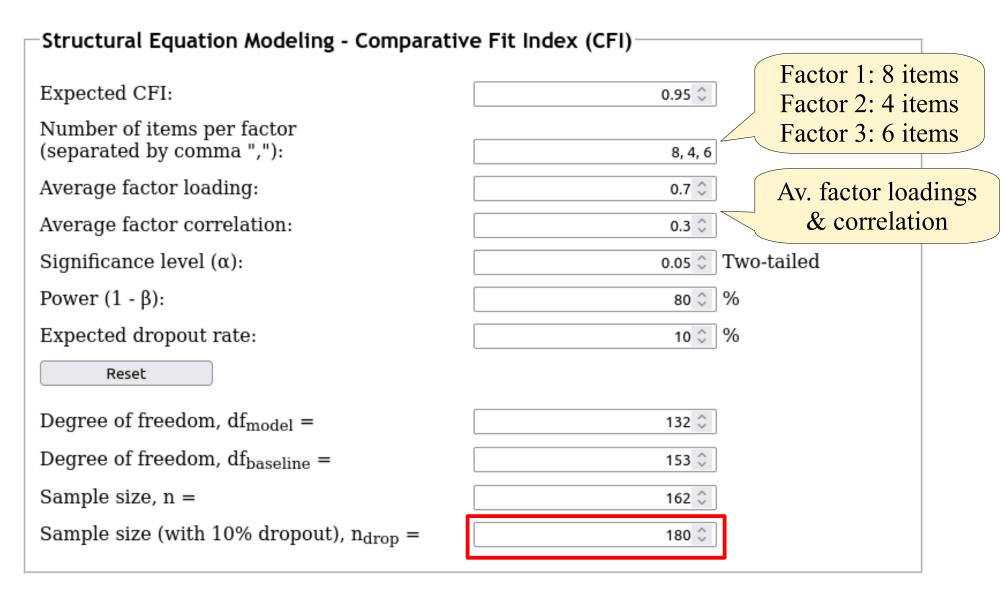
• We have to sample 178 students to confirm the internal structure validity of ABC-Q among the students, taking into account a 10% dropout rate.

5. Structural Equation Modeling → CFI (advanced) Allows unequal number of items per factor ☆ » Sample Size Calculator

Sample Size Calculator (web)



- Let's say a researcher wants to validate a questionnaire with:
 - 3 factors
 - Factor 1: 8 items, Factor 2: 4 items, Factor 3: 6 items
 - Based on literature, average factor loadings around 0.7
 - Average factor-factor correlation around 0.3
 - Acceptable CFI > 0.95
 - 10% dropout
- How many respondents should he sample?



• We have to sample 180 students to confirm the internal structure validity of ABC-Q among the students, taking into account a 10% dropout rate.

(1) Objective:

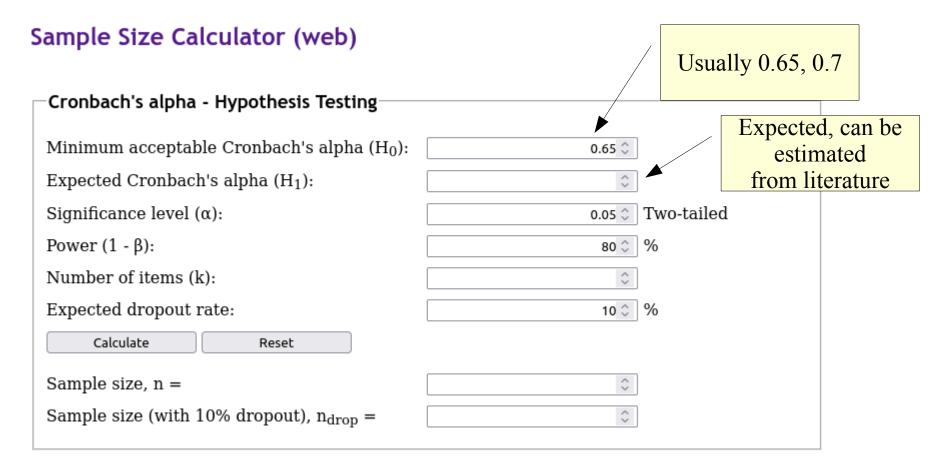
- Determine reliability of a psychometric tool / questionnaire
- "This study aims to determine the reliability of ABC-Q among students."

- (2) Statistical analysis:
- Cronbach's alpha coefficient

- (3) Sample size method:
- Cronbach's alpha coefficient Hypothesis testing

4. Reliability \rightarrow Cronbach's alpha coefficient

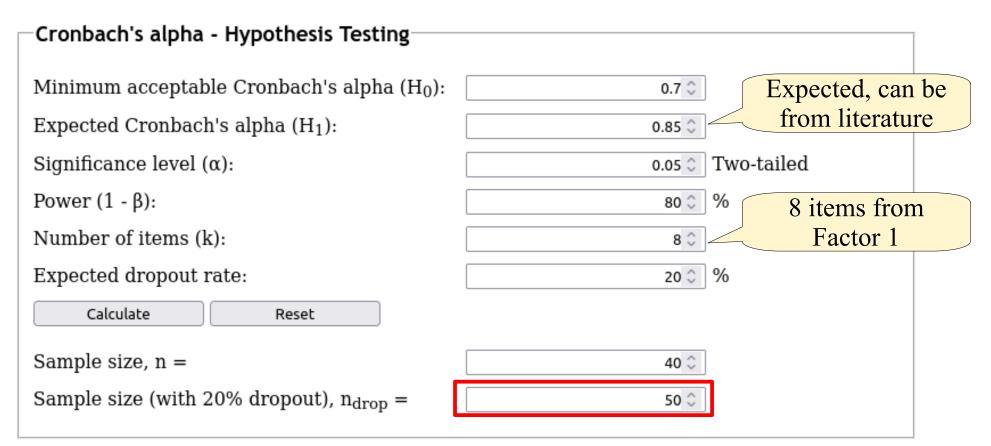
Sample Size Calculator



- Let's say a researcher wants to determine reliability of a questionnaire with:
 - Minimum acceptable Cronbach's alpha = 0.7
 - Expected Cronbach's alpha = 0.85
 - 2 factors
 - Factor 1: 8 items, Factor 2: 6 items
 - 20% dropout
- How many respondents should he sample?
 - Calculate for each factor, and take the largest.

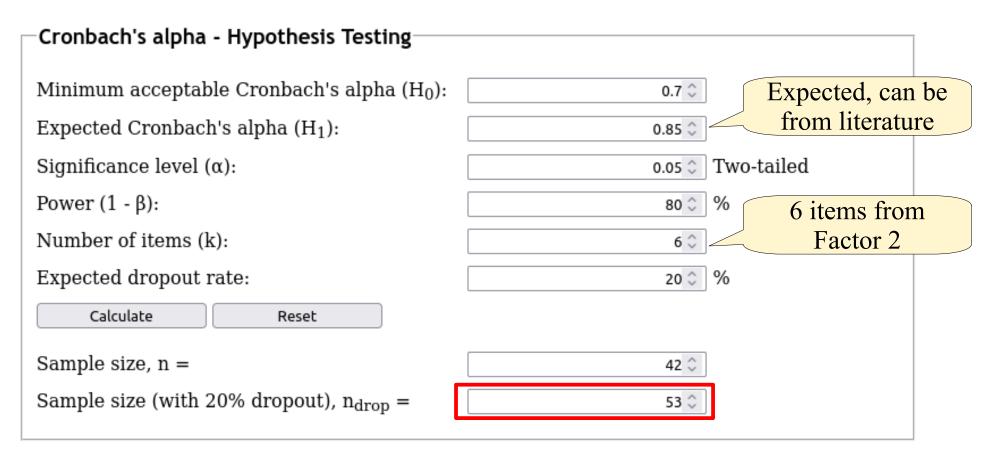
***** » Sample Size Calculator

Sample Size Calculator (web)



***** » Sample Size Calculator

Sample Size Calculator (web)



• We have to sample 53 students to determine the reliability of ABC-Q among the students, taking into account a 20% dropout rate.

Intraclass correlation

(1) Objective:

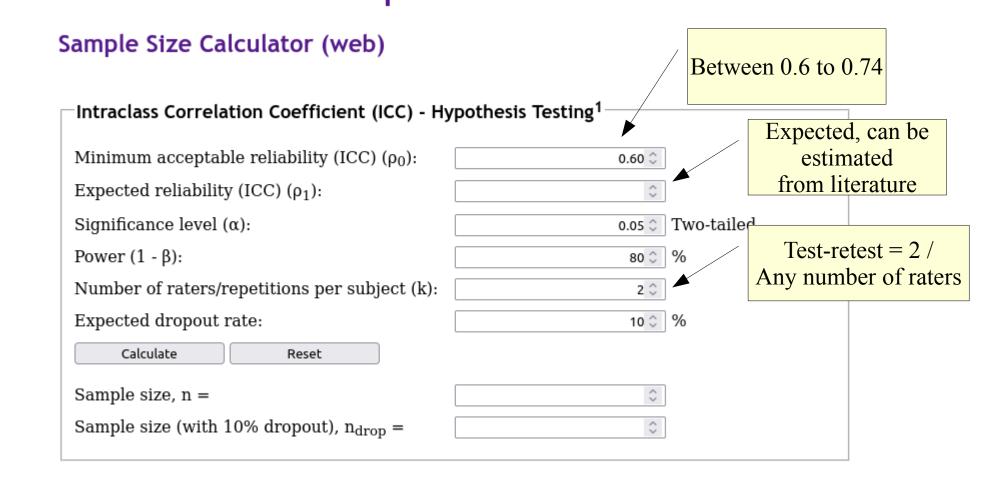
- Determine test-retest reliability / interrater reliability of a numerical measure, i.e. consistency / agreement.
- "This study aims to determine the test-retest reliability of ABC-Q among students."
- "This study aims to determine the interrater reliability of new doctors in measuring blood pressure."

Intraclass correlation

- (2) Statistical analysis:
- Intraclass correlation coefficient (ICC)

- (3) Sample size method:
- Intraclass Correlation Coefficient (ICC) -Hypothesis Testing

4. Reliability → Intraclass correlation coefficient (ICC) → ICC - Hypothesis Testing ☆ » Sample Size Calculator

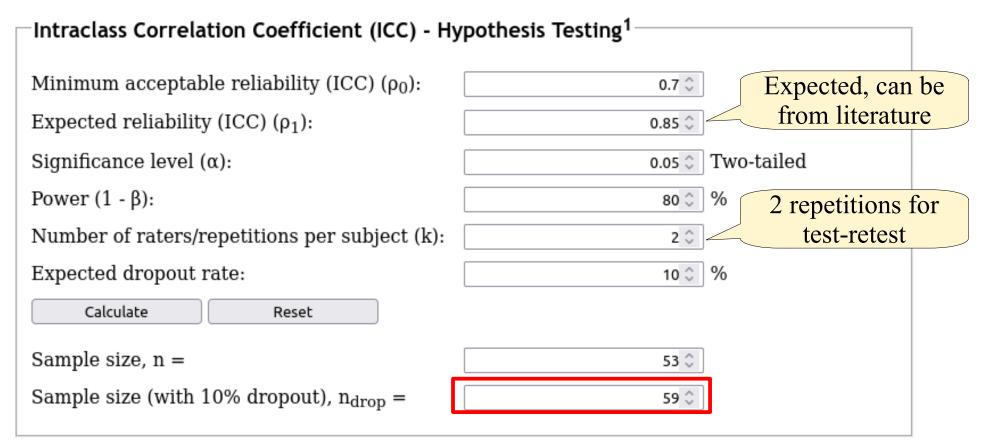


Intraclass correlation - test-retest

- Let's say for test-retest reliability:
 - Minimum acceptable ICC = 0.7
 - Expected ICC = 0.85
 - Two occasions
 - 10% dropout
- How many respondents should he sample?

Intraclass correlation ***** * Sample Size Calculator

Sample Size Calculator (web)



Intraclass correlation

• We have to sample 59 students to determine the test-retest reliability of ABC-Q among the students, taking into account a 10% dropout rate.

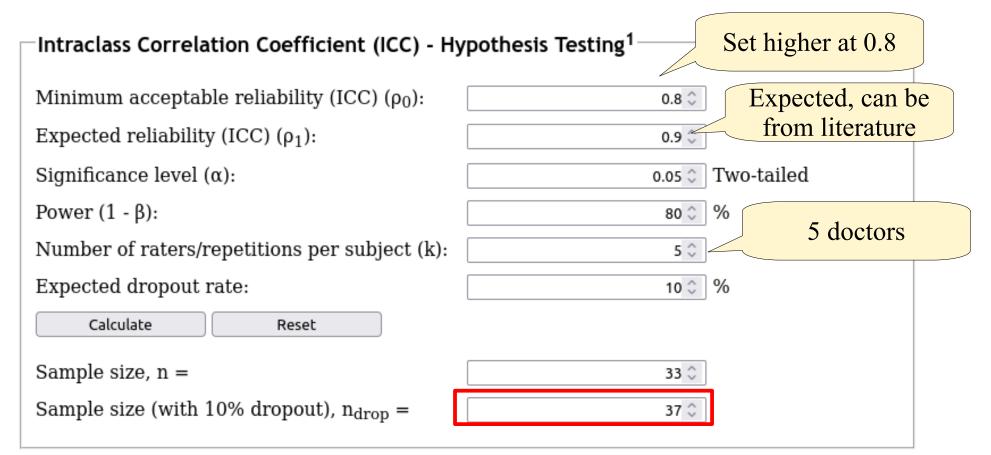
Intraclass correlation - interrater

- Let's say for interrater reliability:
 - Minimum acceptable ICC = 0.8 (set higher, measuring blood pressure correctly is a serious matter)
 - Expected ICC = 0.9
 - 5 doctors
 - 10% dropout
- How many doctors should he sample?

Intraclass correlation

***** » Sample Size Calculator

Sample Size Calculator (web)



Intraclass correlation

• We have to sample 37 <u>patients</u> to determine the interrater reliability of 5 <u>doctors</u> in measuring blood pressure, taking into account a 10% dropout rate.

(1) Objective:

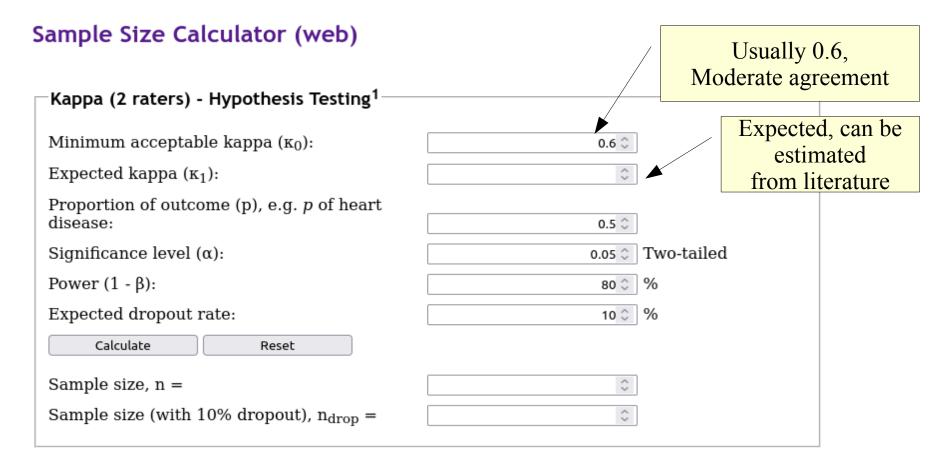
- Determine interrater reliability of a categorical measure, i.e. consistency / agreement.
- "This study aims to determine the interrater reliability of two radiologists in rating X-ray films for the presence of pneumonia among patients with shortness of breath."
- Rater = human rater / diagnostic test

- (2) Statistical analysis:
- Kappa coefficient

- (3) Sample size method:
- Kappa (2 raters) Hypothesis testing

4. Reliability \rightarrow Kappa coefficient \rightarrow Kappa (2 raters) - Hypothesis Testing

Sample Size Calculator



- Let's say a researcher wants to determine the agreement between two radiologists:
 - Minimum acceptable kappa = 0.6
 - Expected kappa = 0.9
 - Presence of pneumonia among patients with shortness of breath is 30% (p = 0.3)
 - 30% dropout
- How many patients should be sampled?

Sample Size Calculator (web)

| Kappa (2 raters) - Hypothesis Testing ¹ | Expected, can be from literature |
|--|-------------------------------------|
| Minimum acceptable kappa (ĸ ₀): | 0.6 🗘 |
| Expected kappa (κ_1): | 0.9 C 30% presence |
| Proportion of outcome (p), e.g. <i>p</i> of heart disease: | $0.3 \diamond p = 0.3$ |
| Significance level (α): | 0.05 🗘 Two-tailed |
| Power (1 - β): | 80 🗘 % |
| Expected dropout rate: | 30 🗘 % |
| Calculate Reset | |
| Sample size, n = | 66 🗘 |
| Sample size (with 30% dropout), $\rm n_{drop}$ = | 95 🗘 |

• We have to sample 95 X-ray films of patients with shortness of breath to determine the interrater reliability of two radiologists in rating Xray films for the presence of pneumonia, taking into account a 30% dropout rate.

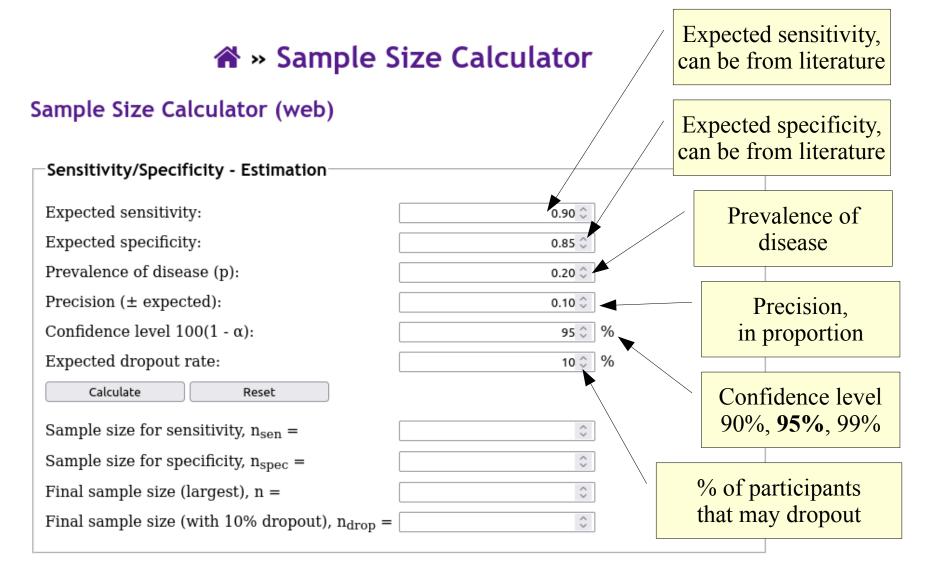
(1) Objective:

- Estimate sensitivity and specificity of a diagnostic test, e.g. RTK-Ag for COVID-19
- "This study aims to determine sensitivity and specificity of a new COVID-19 self-test kit among exposed adults."

- (2) Statistical analysis:
- Proportion, 95% Confidence Interval (CI)

- (3) Sample size method:
- Sensitivity/Specificity Estimation

2. Proportions \rightarrow Sensitivity and specificity \rightarrow Sensitivity/Specificity - Estimation

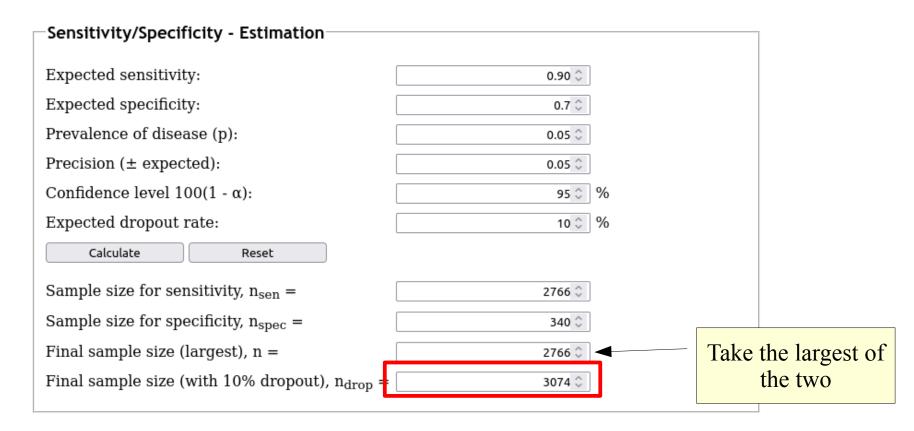


- Let say:
 - Expected Sensitivity = 0.9
 - Expected Specificity = 0.7
 - Prevalence of COVID-19* = 5% = 0.05 (in proportion)
 - Precision = $\pm 5\%$ = 0.05 (in proportion)
 - 95% Confidence level
 - 10% dropout
- How many exposed adults should we sample?

*based on positivity rate

Sample Size Calculator

Sample Size Calculator (web)



• We have to sample 3075 patients to estimate both sensitivity and specificity of the new test kit among exposed adults, taking into account a 10% dropout rate.

References

Arifin, W. N. (2013). *Introduction to sample size calculation*. *Education in Medicine Journal*, 5(2), e89-e96.

Arifin, W. N. (2014). Calculating standard deviation of difference for determination of sample size for planned paired t-test analysis. *Education in Medicine Journal*, 6(2).

Arifin, W. N. (2018). A Web-based Sample Size Calculator for Reliability Studies. *Education in medicine journal*, 10(3).

Arifin, W. N. (2021). A Web-based Sample Size Calculator for Structural Equation Modeling in Psychological Research. October 2021, 3rd Biennial International Psychology Conference, Effat University. Available on https://www.researchgate.net/profile/Wan-Nor-Arifin