

Statistical concepts for proposal reviewers

By

Abdulhakeem Alrawahi

OMSB-Research Section

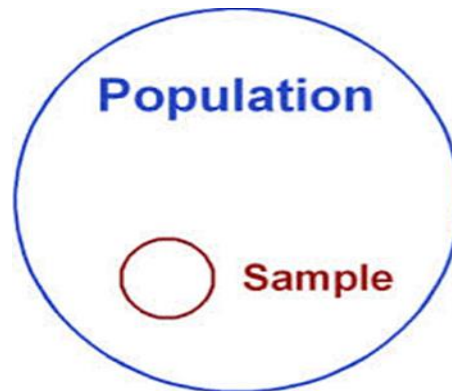
- Introduction
- Sample size calculation
- Descriptive measures
- Measures of occurrence
- Measures of association
- Measures of diagnostic accuracy
- Measures of agreement

- **Statistical concepts in a research proposal:**
- Sample size calculation
- Statistical analysis

Sample size

Study sample

- A subset (part) of a population, selected either by probability (random) or non-probability (non-random) methods



Good sample

- **Represents** the population -> sampling method

- **Adequate** in number, not small, not large -> **sample size calculation**

Sample size in qualitative research

- The issue is to understand the problem from all aspects and not to get figures and associations
- Start with small number of participants (e.g. 10-20) and then further recruitment left open until saturation
- saturation is when researcher feels that he/she has understood the problem from all aspects (no more ideas)

Quantitative studies

- **Different formulae for different studies**
- **Depending on study design and study objective**


1- Descriptive cross-sectional (prevalence) or longitudinal (incidence) studies, assumptions to be reported in the proposal:

- Expected prevalence / incidence = P (close to study setting)
- Confidence level \Rightarrow 95 % \Rightarrow $z = 1.96$
- Permissible error around the estimate = d

Prevalence/ incidence assumptions

OpenEpi Menu Sample Size Calculator by Ra Google

www.raosoft.com/samplesize.html Search



Sample size calculator

What margin of error can you accept? %
5% is a common choice
 The margin of error is the amount of error that you can tolerate. If 90% of respondents answer *yes*, while 10% answer *no*, you may be able to tolerate a larger amount of error than if the respondents are split 50-50 or 45-55.
 Lower margin of error requires a larger sample size.

What confidence level do you need? %
Typical choices are 90%, 95%, or 99%
 The confidence level is the amount of uncertainty you can tolerate. Suppose that you have 20 yes-no questions in your survey. With a confidence level of 95%, you would expect that for one of the questions (1 in 20), the percentage of people who answer *yes* would be more than the margin of error away from the true answer. The true answer is the percentage you would get if you exhaustively interviewed everyone.
 Higher confidence level requires a larger sample size.

What is the population size?
If you don't know, use 20000
 How many people are there to choose your random sample from? The sample size doesn't change much for populations larger than 20,000.

What is the response distribution? %
Leave this as 50%
 For each question, what do you expect the results will be? If the sample is skewed highly one way or the other, the population probably is, too. If you don't know, use 50%, which gives the largest sample size. See below under **More information** if this is confusing.

Your recommended sample size is **377**
 This is the minimum recommended size of your survey. If you create a sample of this many people and get responses from everyone, you're more likely to get a correct answer than you would from a large sample where only a small percentage of the sample responds to your survey.

Online surveys with **Vovici** have completion rates of 66%!

Alternate scenarios

With a sample size of	<input type="text" value="100"/>	<input type="text" value="200"/>	<input type="text" value="300"/>	With a confidence level of	<input type="text" value="90"/>	<input type="text" value="95"/>	<input type="text" value="99"/>
Your margin of error would be	9.78%	6.89%	5.62%	Your sample size would need to be	267	377	643

Save effort, save time. **Conduct your survey online with Vovici.**

More information

Analytic studies

- **Testing associations /difference**
- Depends on the selected design
- But always the confidence level and the power are to be reported (beside other needed parameters)

Case control studies assumptions

- Confidence level
- Power
- % exposure among controls (population)
- OR or % exposure among cases

Case control studies

The screenshot shows a web browser window with the URL www.openepi.com/SampleSize/SSCC.htm. The page has a navigation menu with tabs for 'Start', 'Enter', 'Results', 'Examples', and 'Help'. Below the navigation are 'Clear' and 'Calculate' buttons. The main content area is titled 'Sample Size for Unmatched Case Control Study' and contains a table of input fields. The 'Two-sided confidence level' is set to 95, 'Power(% chance of detecting)' is 80, 'Ratio of Controls to Cases' is 1.0, and 'Percent of controls exposed' is 40. A yellow highlighted row contains the instruction: 'Please fill in one of the following. The other will be calculated.' Below this, the 'Odds ratio' and 'Percent of cases with exposure' fields are empty.

Sample Size for Unmatched Case Control Study		
Two-sided confidence level	95	(1-alpha) usually 95%
Power(% chance of detecting)	80	Usually 80%
Ratio of Controls to Cases	1.0	For equal samples, use 1.0
Percent of controls exposed	40	Between 0.0 and 99.99
Please fill in one of the following. The other will be calculated.		
Odds ratio		
Percent of cases with exposure		Between 0.0 and 99.99

Cohort/RCT studies assumptions

- Confidence level
- Power
- % of unexposed with outcome
- OR or % exposed with outcome

Cohort/RCT studies

OpenEpi Menu OpenEpi:Sample Size for X-Sectional Google

www.openepi.com/SampleSize/SSCohort.htm

[Expand All](#) | [Collapse](#)

- Home
- Info and Help
- Language/Options/Settings
- Calculator
- Counts
 - Std.Mort.Ratio
 - Proportion
 - Two by Two Table
 - Dose-Response
 - R by C Table
 - Matched Case Control
 - Screening
- Person Time
 - 1 Rate
 - Compare 2 Rates
- Continuous Variables
 - Mean CI
 - Median/%ile CI
 - t test
 - ANOVA
- Sample Size
- Power
- Random numbers
- Searches
 - Google--Internet
 - PubMed--MEDLARS
- Internet Links
- Download OpenEpi
- Development

Start **Enter** **Results** **Examples** **Help**

Clear Calculate

Sample Size: X-Sectional, Cohort, & Randomized

Two-sided confidence level(%)	95	(1-alpha) usually 95%
Power (1-beta or % chance of detecting)	80	Usually 80%
Ratio of Unexposed to Exposed in sample	1.0	For equal samples, use 1.0
Percent of Unexposed with Outcome	5	Between 0.0 and 99.9
Please fill in 1 of the following. The others will be calculated.		
Odds ratio		
Percent of Exposed with Outcome		Between 0.0 and 99.9
Risk/Prevalence Ratio		
Risk/Prevalence difference		Between -99.99 and 99.99

Windows Taskbar: 8:52 AM, 10/8/2017

- **In prospective studies researcher should consider:**
- Non-response /non participation rate, loss of follow-up

- **In retrospective studies researcher should consider:**
- Missing info, loss to follow up

Statistical analysis

Descriptive statistics

- **Almost all types of study proposal should report this**
- **Categorized variables**
 - Proportions
- **Continuous variables:**
 - **Normally distributed variables: Mean \pm SD**
 - most widely used
 - uses all the data
 - susceptible to outliers
 - **Non-normally distributed: Median with range / IQR**
 - Does not use all the data
 - Resistant to outliers

Prevalence / incidence analysis

- Proportion with 95% confidence interval
- Incidence density can also be calculated (person-rears)

Assessing associations/ differences

- **Epidemiological measures:**
- RR/OD with confidence interval

- **Statistical tests:**
- Univariate (crude) analysis ->
- parametric tests -> for normally distributed data
- non-parametric tests-> for non-normal

Univariate parametric and nonparametric

Analysis of association between	Parametric Test	Non Parametric Test
Two categorized variables		Chi-sqaure
Continuous variable and categorized variable with two independent groups	Independent sample t-test	Mann Whitney U test
Continuous variable and categorized variable with more than two independent groups	Analysis of variance (ANOVA)	Kruskal-Wallis test
Two continuous variables	Pearson correlation	Spearman's rank correlation

Continue: assessing associations/differences

- **Multivariate analysis**

- **Stratification analysis**

- Potential confounding factors
- How potential confounders will be categorized

- **Regression**

- Linear regression: continuous outcome
- Logistic regression: categorized outcome
- Cox regression: time to outcome

Assessing diagnostic accuracy

- Sensitivity
-
- Specificity
- Predictive values
- Likelihood ratios
- Receiver operating curves (AUC)

Assessing agreement (reliability studies)

- **Categorical outcome:**
 - McNemar chi-square/ Cohen's Kappa coefficient
- **Continuous outcome:**
 - Paired T-test
 - Intra-class correlation coefficient
 - Bland-Altman method (95% CI for difference in raters results)

Thank uuuuu