Methods and equations for estimating statistical sample size in for data analysis using statistical software

Dr.Hissa Alzahrani

Determining the proper sample size for a statistical investigation is contingent on various elements, such as the investigation's objectives, analytical approach, desired statistical power, alpha level (α), magnitude of effect, and data variability. Commonly employed techniques and formulas for calculating sample size across diverse scenarios utilize practical software like SAS.

1. Estimating Sample Size for Means

Method: Confidence Interval Method

- **Objective**: Determine the sample size needed to estimate a population mean within a specified margin of error.
- **Method:** To estimate the mean (μ) of a population within a specified margin of error (E):

Equation:

$$n = \left(\frac{Z \cdot \sigma}{E}\right)^2$$

Where:

- n: Required sample size.
- Z: Z-value corresponding to the desired confidence level (e.g., 1.96 for 95% confidence).
- σ: Estimated population standard deviation.
- E: Margin of error (desired precision).

Example:

If σ = 10, E=2, and confidence level is 95% (Z=1.96):

$$n = \left(\frac{1.96 \cdot 10}{2}\right)^2 = 96.04 \approx 97$$

---

2. Estimating Sample Size for Proportions

Method: Confidence Interval for Proportions

- **Objective**: Calculate the sample size needed to estimate a proportion within a specified margin of error.

Method: For estimating a population proportion (p) within a margin of error (E):

Equation:

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

Where:

- p: Estimated population proportion (use p=0.5) if unknown for maximum variability).
- E: Margin of error.

Example:

If p=0.5, E=0.05, and Z=1.96:

$$n = \frac{1.96^2 \cdot 0.5 \cdot (1-0.5)}{0.05^2} = 384.16 \approx 385$$

---

3. Sample Size for Comparing Two Means

Method: Two-Sample t-Test

- **Objective**: Estimate the sample size needed to detect a difference between two group means.

Method: For comparing two means with equal sample sizes:

Equation:

$$n = \frac{2 \cdot (Z_{\alpha/2} + Z_{\beta})^2 \cdot \sigma^2}{\Delta^2}$$

Where:

- $Z_{\alpha/2}$: Z-value for the desired significance level (e.g., 1.96 for 95% confidence).
- Z_{β}: Z-value for desired power (e.g., 0.84 for 80% power).
- σ: Standard deviation of the population.
- Δ: Minimum detectable difference between the two groups.

---

4. Sample Size for Comparing Two Proportions

Method: Two-Sample Test of Proportions

- **Objective**: Calculate the sample size to detect a difference in proportions between two groups.
- **Method**: For comparing two proportions ($p_1$and p_2):

Equation:

$$n = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \cdot (p_1(1-p_1) + p_2(1-p_2))}{(p_1 - p_2)^2}$$

Example:

If $p_1 = 0.6, p_2 = 0.5, \alpha = 0.05$ ($Z_{\alpha/2} = 1.96$), and power = 80% ($Z_{\beta} = 0.84$):

$$n = \frac{(1.96 + 0.84)^2 \cdot (0.6 \cdot 0.4 + 0.5 \cdot 0.5)}{(0.6 - 0.5)^2} \approx 199$$

---

5. Sample Size for Regression

Method: Rule of Thumb or Statistical Power for F-Test

- **Objective**: Ensure adequate sample size for multiple linear regression models, considering the number of predictors.
- **Rule of Thumb**: n=10×number of predictors n=10×number of predictors.

---

6. Sample Size for Non-Parametric Tests

Non-parametric tests often require slightly larger sample sizes than parametric tests due to lower power.

Example for Wilcoxon Rank-Sum Test:

Sample size tables or software like SAS or R can be used based on the desired power, effect size, and significance level.

---

7. Sample Size for Repeated Measures

Method: Mixed Models or Repeated ANOVA

- **Objective**: Determine the sample size for studies with repeated measurements over time.
- **Method**: Use custom simulation in SAS with PROC MIXED or power analysis based on variances and correlations.

---

8. Bayesian Sample Size

Method: Credible Intervals for Bayesian Analysis

- **Objective**: Estimate sample size for Bayesian models by ensuring posterior precision or probability of certain outcomes.
- **SAS Application**: Requires simulation in PROC MCMC to assess sample size.

---

9. Tools and Software for Sample Size Calculation

- **Software**: SAS (`PROC POWER`), R (`pwr` package), G*Power.
- **Online Calculators**: Many websites provide user-friendly tools for estimating sample size.

References

1. Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). Wiley.
2. Charan, J., & Biswas, T. (2013). How to calculate sample size for different study designs in medical research? *Indian Journal of Psychological Medicine*, *35*(2), 121–126. DOI:10.4103/0253-7176.116232
3. Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, *55*(3), 187–193. DOI:10.1198/000313001317098149