Sample Size and Power Considerations for Longitudinal Studies

Outline

- Quantities required to determine the sample size in longitudinal studies
- Review of type I error, type II error, and power
- For continuous data
  - Comparison of two groups for univariate data
  - Comparison of two groups for the rate of change
  - Comparison of two groups for the time-averaged difference
  - Sample size for three-level hierarchical data
- For dichotomous outcome
  - Comparison of two groups for univariate data
  - Comparison of two groups for the difference in proportions
- Sample size for longitudinal data based on GEE

Sample Size and Power Considerations for Longitudinal Studies

Quantities required to determine the sample size in longitudinal studies

- Type I error rate (α).
- Type II error rate (β), power ($P = 1 - \beta$).
- Smallest meaningful difference to be detected (d) .
- Measurement variation (σ^2).
- Number of repeated observations per person (n).
- Correlation among the repeated observations (ρ) or a general correlation matrix (R).

Hypothesis testing, type I and II error, and power

Let H_0 denote the null hypothesis, and H_1 denote the alternative hypothesis,

Conclusion	H_0 is true	H_1 is true
Reject H_0	Type I error (α)	Power ($1-\beta$)
Fail to reject H_0		Type II error (β)

- Choice of α: often α is specified at 0.05; $z_{\alpha/2} = 1.96$ for a two-sided test and $z_\alpha = 1.64$ for a one-sided test.

- Choice of β: power of a test is the probability of detecting a true underlying difference and depends on the alternative hypothesis. The power $(1-\beta)$ is often set to 0.8, i.e., $z_\beta = .842$. We want to choose the sample size to ensure the desired power for detecting the smallest meaningful difference.

Comparison of two groups for univariate continuous outcome

Number of subjects (m) in each of two groups:

$$m = \frac{2(z_{\alpha/2} + z_\beta)^2 \sigma^2}{d^2} = \frac{2(z_{\alpha/2} + z_\beta)^2}{\Delta^2}, \tag{1}$$

where $\Delta = \frac{d}{\sigma}$ is referred to as the effect size and σ^2 is the assumed common variance in the two groups.

- $m \uparrow$ if $\sigma^2 \uparrow$;

  $m \uparrow$ if $1 - \beta \uparrow$ (or $\beta \downarrow$);

  $m \uparrow$ if $\alpha \downarrow$;

  $m \uparrow$ if $d \downarrow$.

- Example: with type I error of 0.05, type II error of 0.2 (power of 80%), two-sided, effect size of 0.7, the required sample size is

$$m = \frac{2(1.96 + .842)^2}{.7^2} = \frac{15.7}{.49} = 32.$$

  Note that $m \approx (4/\Delta)^2$.

- The sample size formula can also be manipulated to determine the power for a given sample size.

- The sample size formula can also be modified to allow groups of unequal size.

Comparison of two groups for the rate of change for continuous outcomes

Consider a simple problem of comparing two groups, A and B with continuous outcomes. Assuming the responses depend on a single covariate as follows:

$$Y_{kij} = \beta_{0k} + \beta_{1k} x_{kij} + \epsilon_{kij}, \quad j = 1, \ldots, n; \quad i = 1, \ldots, m; \quad k = A \text{ or } B.$$

Both groups have the same number of subjects, m. We assume that $\mathrm{Var}(\epsilon_{kij}) = \sigma^2$ and $\mathrm{Cor}(Y_{kij}, Y_{kij'}) = \rho$ for all $j \neq j'$. We also assume that each person has the same set of covariate so that $x_{kij} = x_j$. The regression coefficients β_{1A} and β_{1B} are the rate of changes in Y for groups A and B, respectively.

With n fixed and known, the number of subjects (m) in each of two groups are needed to achieve type I error rate α and power $1-\beta$ for comparing the rate of change in a continuous response between two groups, is

$$m = \frac{2(z_{\alpha/2}+z_\beta)^2(1-\rho)\sigma^2}{ns_x^2 d^2} = \frac{2(z_{\alpha/2}+z_\beta)^2(1-\rho)}{ns_x^2\Delta^2}, \tag{2}$$

where $d = \beta_{1B} - \beta_{1A}$ and $s_x^2 = \Sigma_j(x_j - \bar{x})^2/n$ is the within-subject variance of the x_j.

- When $\rho \uparrow$, m?
- When $s_x^2 = \Sigma_j(x_j - \bar{x})^2/n \uparrow$, m?
  - For fixed length of study, τ and no requirement for the spacing between repeated measurements:
  - For equally spaced measurements:

Example:

Consider a hypothetical clinical trial on the effect of a new treatment in reducing blood pressure. There are three visits, including the baseline, are planned at years 0, 2, and 5. Thus, $n = 3$ and $s_x^2 = 4.22$. With type I error of 0.05, type II error of 0.2 (power of 80%), one-sided test, testing smallest meaningful difference $d = .5$ mmHg/year.

We listed the number of subjects required for both treated and control groups for some selected values of ρ and σ^2, are

	σ^2		
ρ	100	200	300
0	391	781	1172
0.2	313	625	938
0.5	196	391	586
0.8	79	157	235

Note for each value of σ^2, the required sample size decreases as the correlation, ρ, increases.

Extension to a general correlation structure

The sample size formula in (**??**) is for the longitudinal data assuming a exchangeable correlation for the responses with correlation of ρ between pairs of responses of a subject. Let R, a $n \times n$ matrix denote a common correlation matrix for each subject. Then the sample size formula for testing the rate of change becomes

$$m = \frac{2(z_{\alpha/2} + z_\beta)^2 \rho_R}{d^2} \tag{3}$$

where ρ_R is the lower right entry of the following 2×2 matrix

$$\left[\begin{pmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \end{pmatrix} R^{-1} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \right]^{-1}.$$

Example:

Consider the above hypothetical clinical trial on the effect of a new treatment in reducing blood pressure.

Now we assume an exponential correlation structure rather than an exchangeable correlation for the responses, so that $R_{jk} = \rho^{|j-k|}$, where $j, k = 1, 2, 3$.

With type I error of 0.05, type II error of 0.2 (power of 80%), one-sided test, testing smallest meaningful difference $d = .5$ mmHg/year.

We listed the number of subjects required for both treated and control groups for some selected values of ρ and σ^2, are

	σ^2		
ρ	100	200	300
0.2	125	249	374
0.5	97	194	290
0.8	46	92	138

Comparison of two groups for the time-averaged difference for continuous outcomes

The model is written as

$$Y_{ij} = \beta_0 + \beta_1 x_i + \epsilon_{ij}, \quad j = 1, \ldots, n; \quad i = 1, \ldots, 2m.$$

where x_i is the treatment indicator variable for the ith subject.

Number of subjects (m) in each of two groups are needed to achieve type I error rate α and power $1 - \beta$ to compare the time-average difference in a continuous response between two groups, is

$$m = \frac{2(z_{\alpha/2} + z_\beta)^2(1 + (n-1)\rho)\sigma^2}{nd^2} = \frac{2(z_{\alpha/2} + z_\beta)^2(1 + (n-1)\rho)}{n\Delta^2}, \tag{4}$$

- $1 + (n-1)\rho$ is called *variance inflation factor*. When $\rho \uparrow$, $m \uparrow$.
- For a general correlation matrix, R, the number of subjects needed per group in (**??**) can be modified by replacing $(1 + (n-1)\rho)/n$ by $(\mathbf{1}'R^{-1}\mathbf{1})^{-1}$, where $\mathbf{1}$ is a $n \times 1$ vector of ones.

Example:

With type I error of 0.05, type II error of 0.2 (power of 80%), one-sided test, testing smallest meaningful effect size $\Delta = .2, .3, .4, .5$.

We listed the number of subjects required for both treated and control groups for some selected values of ρ and Δ, are

	Δ (%)			
ρ	20	30	40	50
0	104	46	26	17
0.2	145	65	37	24
0.5	207	92	52	33
0.8	268	120	67	43

Note for each value of σ^2, the required sample size increases as the correlation, ρ, increases.

Sample size for continuous outcomes from three level hierarchical data

A three level mixed-effects linear model to test an intervention effect on continuous outcomes Y can be written (Heo and Leon, 2008)

$$Y_{ijk} = \beta_0 + \delta X_i + \mu_i + \mu_{j(i)} + \epsilon_{ijk},$$

- X_i is the intervention indicator for all the subjects within the ith level one unit, e.g. the conventional ($X = 0$) vs. the new medical chart system ($X = 1$) that the hospitals are using.
- $i = 1, \ldots, N_3$ is the index for the level three unit (e.g, hospital), where N_3 is the total number hospitals in each of the two groups of X,
- $j = 1, \ldots, N_2$ is the index for the level two unit (e.g., physician) nested within each i,
- $k = 1, \ldots, N_1$ is the index for the level one unit (e.g., subjects) nested within each j,
- δ is the intervention effect, the parameter of interest,

- μ_i and $\mu_{j(i)}$ are level 3 and level 2 random effects.

- The null hypothesis: $H_0 : \delta = 0$.

Number of subjects (N_3) per group is needed for level three units:

$$N_3 = \frac{2(z_{\alpha/2} + z_\beta)^2(1 + N_1(N_2 - 1)\rho_2 + (N_1 - 1)\rho_1)\sigma^2}{N_2 N_1 \delta^2}, \tag{5}$$

where $\sigma^2 = \mathrm{Var}(Y_{ijk})$, $\rho_1 = \mathrm{Cor}(Y_{ijk}, Y_{ijk'})$ is the correlation among level one data, and $\rho_2 = \mathrm{Cor}(Y_{ijk}, Y_{ij'k'})$ is the correlation among level two data.

- $1 + N_1(N_2 - 1)\rho_2 + (N_1 - 1)\rho_1$ is the variance inflation factor.

- Example: with type I error of 0.05, type II error of 0.2 (power of 80%), $N_2 = 5$, $N_1 = 6$, level one correlation $\rho_1 = .6$, level two correlation $\rho_2 = .05$, two-sided test, testing a smallest meaningful effect size $\Delta = \delta/\sigma = .3$ requires sample size for the level three unit

$$N_3 = \frac{2(1.96 + .842)^2(1 + 6(5 - 1).05 + (6 - 1).6)}{5 \times 6 \times .3^2} = 31.$$

Comparison of two groups for univariate dichotomous outcome

Number of subjects (m) in each of two groups:

$$m = \frac{(z_{\alpha/2}(2\bar{p}\bar{q})^{1/2} + z_{\beta}(p_1 q_1 + p_2 q_2)^{1/2})^2}{(p_1 - p_2)^2}, \tag{6}$$

where p_1 is the response proportion in group A and $q_1 = 1 - p_1$; p_2 is the response proportion in group B and $q_2 = 1 - p_2$; $\bar{p} = (p_1 + p_2)/2$ and $\bar{q} = 1 - \bar{p}$.

Example: with type I error of 0.05, type II error of 0.2 (power of 80%), two-sided test, testing two proportions of $p_1 = 0.15$ and $p_2 = 0.30$ requires sample size

$$m = \frac{(1.96(2 \times .225 \times .775)^{1/2} + .842(.15 \times .85 + .3 \times .7)^{1/2})^2}{(.15 - .30)^2} = 76.$$

Comparison of two groups for the difference in proportions for dichotomous outcomes

Number of subjects (m) in each of two groups:

$$m = \frac{(z_{\alpha/2}(2\bar{p}\bar{q})^{1/2} + z_{\beta}(p_1 q_1 + p_2 q_2)^{1/2})^2 (1 + (n-1)\rho)}{n(p_1 - p_2)^2}, \tag{7}$$

where n is the number of repeated measurements and ρ is the correlation between pairs of responses of a subject; p_1 is the response proportion in group A and $q_1 = 1 - p_1$; p_2 is the response proportion in group B and $q_2 = 1 - p_2$; $\bar{p} = (p_1 + p_2)/2$ for 2 equal sized groups and $\bar{q} = 1 - \bar{p}$.

With type I error of 0.05, type II error of 0.2 (power of 80%), one-sided test, $p_1 = .5$, we listed the sample size needed per group for some selected ρ and $d = p_1 - p_2$

	$d = p_1 - p_2$		
ρ	0.3	0.2	0.1
0	11	25	102
0.2	15	35	143
0.5	21	49	204
0.8	27	64	265

Sample size for longitudinal data based on GEE

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})'$ denote a vector of responses for the ith subject, $i = 1, \ldots, m$. The marginal model is (Liu and Liang, 1997)

$$g(\mu_{ij}) = x'_{ij}\beta + z'_{ij}\lambda$$

where $\mu_{ij} = E(y_{ij})$ and $g(\cdot)$ is a link function; β is a $p \times 1$ vector of parameter of interest and λ is a $q \times 1$ vector that are for the adjusting covariates in the above marginal model.

To test the null hypothesis of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, we consider the quasi-score test statistics based on GEE

$$T = \mathbf{S}'_\beta(\boldsymbol{\beta}_0, \hat{\lambda}_0, \alpha)\Sigma_0^{-1}\mathbf{S}_\beta(\boldsymbol{\beta}_0, \hat{\lambda}_0, \alpha)$$

where

$$\mathbf{S}_\beta(\boldsymbol{\beta}_0, \hat{\lambda}_0, \alpha) = \sum_{i=1}^{m}\left(\frac{\partial \mu_i(\boldsymbol{\beta}_0, \hat{\lambda}_0)}{\partial \boldsymbol{\beta}}\right)' V_i^{-1}(\boldsymbol{\beta}_0, \hat{\lambda}_0, \alpha)(y_i - \mu_i(\boldsymbol{\beta}_0, \hat{\lambda}_0))$$

$$\Sigma_0 = \mathrm{Cov}(\mathbf{S}_\beta(\boldsymbol{\beta}_0, \hat{\lambda}_0, \alpha)), \qquad V_i = \mathrm{Cov}(\mathbf{Y}_i; \boldsymbol{\beta}_0, \lambda, \alpha),$$

and $\hat{\lambda}_0$ is an estimator of λ under H_0 obtained by solving the equation of

$$\sum_{i=1}^{m}\left(\frac{\partial\mu_i(\boldsymbol{\beta}_0,\lambda)}{\partial\lambda}\right)' V_i^{-1}(\boldsymbol{\beta}_0,\lambda,\alpha)(y_i-\mu_i(\boldsymbol{\beta}_0,\lambda))=0$$

Under H_0, the test statistic T follows a χ^2_p when $m\to\infty$, and under $H_1:\boldsymbol{\beta}=\boldsymbol{\beta}_1,\lambda=\lambda_1$, T follows a non-central chi-square distribution asymptotically with the non-centrality parameter of

$$\nu=\zeta'\Sigma_1^{-1}\zeta$$

where $\zeta=E_{H_1}(\mathbf{S}_\beta(\boldsymbol{\beta}_0,\hat{\lambda}_0,\alpha))$ and $\Sigma_1=\mathrm{Cov}_{H_1}(\mathbf{S}_\beta(\boldsymbol{\beta}_0,\hat{\lambda}_0,\alpha))$.

If we further assume $n_i=n$ and the covariates (x_{ij},z_{ij}) are discrete, then $\zeta=m\tilde{\zeta}$ and $\Sigma_1=m\tilde{\Sigma}_1$ and the sample size, m, is

$$m=\nu/(\tilde{\zeta}'\tilde{\Sigma}_1^{-1}\tilde{\zeta}). \tag{8}$$

Missing data and sample size

- We have assumed no missing data or attrition so far.
- The impact of missing data is difficult to quantify precisely because it depends on the patterns of missingness.
- An *ad hoc* approach is to inflate the required sample size to account for the assumed attrition rate. For example, if the attrition rate is assumed to be 10%, then the target sample size should be m/0.9.

Further Reading

- Chapter 2 and 8.5 of DHLZ and Chapter 15 of Fitzmaurice, Laird and Ware (2004).

References

- Hedeker D, Gibbons R, and Waternaux C. (1999) Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, **24**(1):70-93.
- Heo M and Leon AC. (2008). Statistical power and sample size requirements for three level hierarchical cluster randomized trials. *Biometrics* **64**(4):1256-62.
- Liu G and Liang KY. (1997). Sample size calculations for studies with correlated observations. *Biometrics* **53**(3):937-47.