Chapter 538

GEE Tests for Two Correlated Proportions with Dropout

Introduction

This procedure provides power analysis and sample size calculation for studies that use a paired design that yield two binary outcomes, one of which may be incomplete. That is, in some pairs, the second observation is missing. The data analysis will use a mixed logistic regression model that is solved with GEE.

With complete data, the standard analysis is McNemar's Test (see McNemar (1947)), and **PASS** includes several procedures that analyze that test statistic. McNemar's Test requires that observations with one or two missing observations must be discarded. Zhang, Cao, and Ahn (2014) present a closed-form sample size formula for the case when some data pairs include missing values in the second observation. This is often referred to as *dropout*.

Another method, also available for sample size calculation in **PASS**, deals with the important case in which all missing values occur in the second observation. We refer to this as *dropout*. We refer to that procedure for further details.

Technical Details

Consider the following table that summarizes the results of a paired design in which one observation of the pair is designated as a treatment and the other is designated as a standard.

	Standard		
Treatment	**Yes**	**No**	**Total**
Yes	P11	P10	Pt
No	P01	P11	1 - Pt
Total	Ps	1 - Ps	1

Our formulation comes from Zhang, Cao, and Ahn (2014). Denote a binary observation by Y_{ij}, where $j = 0$ (for s) and 1 (for t) t, s gives the group and *i* = 1, 2, ..., *N* gives the subject. A "success" is represented by $Y_{ij} = 1$ and a "failure" by $Y_{ij} = 0$.

GEE Approach

Let $P(Y_{ij} = 1) = p_{ij}$. The GEE method models p_{ij} by the mixed logistic regression

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_1 + \beta_2 j$$

The regression coefficient β_2 is equal to the log odds ratio of the two response rates. It is given by

$$\beta_2 = \log\left(\frac{p_{it}}{1-p_{it}}\right) - \log\left(\frac{p_{is}}{1-p_{is}}\right)$$

The null hypothesis $H_0{:}\,\beta_2 = 0$ corresponds to $H_0{:}\,p_{is} = p_{it}$. The alternative hypothesis is $H_1{:}\,\beta_2 \neq 0$ or $H_1{:}\,p_{is} \neq p_{it}$.

Zhang, Cao, and Ahn (2014) provide the following formula for the overall sample size for a two-sided test

$$N = \frac{\sigma^2\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{\beta_2^2}$$

where and α is the probability of a type-I error and β is the probability of a type-II error.

The variance is given by

$$\sigma^2 = \frac{Vs + (Pmt)Vt - 2(Pmt)\rho\sqrt{VsVt}}{(Pmt)VsVt}$$

where *Pmt* is the proportion of the second observation missing, $Vs = Ps(1-Ps)$, $Vt = Pt(1-Pt)$, and ρ is the within-subject correlation coefficient.

Estimating ρ

As outlined in Zhang, Cao, and Ahn (2017), the relationship between *P*11 and ρ is

$$\rho = \frac{P11 - P_sP_t}{\sqrt{P_sP_t(1-P_s)(1-P_t)}}$$

where *P*11 is the joint probability that both observations are equal to one.

Using this relationship, values of ρ can be entered and transformed to the corresponding value of *P*11. The only concern is that values of ρ be used that limit *P*11 to be between 0 and 1.

The lower and upper limits of the correlation are

$$\rho_L = \max\left\{-\sqrt{\frac{P_sP_t}{(1-P_s)(1-P_t)}}, -\sqrt{\frac{(1-P_s)(1-P_t)}{P_sP_t}}\right\}$$

$$\rho_U = \min\left\{\sqrt{\frac{P_s(1-P_t)}{P_t(1-P_s)}}, \sqrt{\frac{P_t(1-P_s)}{P_s(1-P_t)}}\right\}$$

Example 1 – Calculating Sample Size

Suppose a dental clinical trial is being planned in which two sites are selected in each subject's mouth. One site is randomly assigned to receive the treatment intervention and the other is assigned the standard intervention. The trial is being conducted to compare two treatments for gingivitis. In the study, suppose *Ps* = 0.5; *Pt* = 0.6, 0.65, 0.7; ρ = 0, 0.2, 0.4, 0.6, 0.8; *alpha* = 0.05; and *power* = 0.9. Similar studies have had *Pmt* = 0.1. Sample size is to be calculated for a two-sided test.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Sample Size**
Alternative Hypothesis	**Two-Sided (Pt ≠ Ps)**
Power	**0.90**
Alpha	**0.05**
Pt Input Type	**Pt**
Pt (Probability (Yt = 1))	**0.6 0.65 0.7**
Ps (Probability (Ys = 1))	**0.5**
P11 Input Type	**ρ (Within-Subject Correlation)**
ρ (Within-Subject Correlation)	**0 0.2 0.4 0.6 0.8**
Pmt (Probability (Yt = Missing))	**0.1**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Reports

Numeric Results

Solve For: Sample Size
Test: Mixed Logistic Regression using GEE
Alternative Hypothesis: Two-Sided (Pt ≠ Ps)

Power	Total Sample Size N	Marginal Probabilities Pt	Ps	Difference Pt - Ps	Within-Subject Correlation ρ	Joint Probability that Yt = 1 and Ys = 1 P11	Proportion Discordant P01 + P10	Probability Treatment is Missing Pmt	Alpha
0.9002	552	0.60	0.5	0.10	0.0	0.3000	0.5000	0.1	0.05
0.9005	448	0.60	0.5	0.10	0.2	0.3490	0.4020	0.1	0.05
0.9002	343	0.60	0.5	0.10	0.4	0.3980	0.3040	0.1	0.05
0.9007	239	0.60	0.5	0.10	0.6	0.4470	0.2061	0.1	0.05
0.9020	135	0.60	0.5	0.10	0.8	0.4960	0.1081	0.1	0.05
0.9005	244	0.65	0.5	0.15	0.0	0.3250	0.5000	0.1	0.05
0.9006	198	0.65	0.5	0.15	0.2	0.3727	0.4046	0.1	0.05
0.9007	152	0.65	0.5	0.15	0.4	0.4204	0.3092	0.1	0.05
0.9010	106	0.65	0.5	0.15	0.6	0.4681	0.2138	0.1	0.05
		0.65	0.5	0.15	0.8			0.1	0.05
0.9000	136	0.70	0.5	0.20	0.0	0.3500	0.5000	0.1	0.05
0.9015	111	0.70	0.5	0.20	0.2	0.3958	0.4083	0.1	0.05
0.9004	85	0.70	0.5	0.20	0.4	0.4417	0.3167	0.1	0.05
0.9032	60	0.70	0.5	0.20	0.6	0.4875	0.2250	0.1	0.05
		0.70	0.5	0.20	0.8			0.1	0.05

Warning: One or more input parameter combinations resulted in a table (possibly undefined) for which no calculations are possible.

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The total number of subjects in the study.
Pt The marginal probability of a "true" response in the treatment observation.
Ps The marginal probability of a "true" response in the standard observation.
Pt - Ps The difference between the two marginal probabilities.
ρ The correlation between the two observations within a subject.
P11 The joint probability that both observations in a pair are true (equal to 1).
P01 + P10 Proportion Discordant. The sum of the two off-diagonal elements, P01 and P10. It is provided to allow easy comparison of these results with McNemar's test.
Pmt The probability that the treatment observation is missing and the standard observation is observed.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A paired design (where it is anticipated that some of only the treatment observations will be missing) will be used to test whether the treatment proportion (Pt) is different from the standard proportion (Ps) (H0: Pt = Ps versus H1: Pt ≠ Ps). The comparison will be made using a two-sided, incomplete data, paired-sample, mixed logistic regression model test, performed using generalized estimating equations (GEE) methods. The Type I error rate (α) will be 0.05. The joint probabilities are calculated from Pt, Ps, and a within-subject correlation of 0. The proportion of missing treatment observations is anticipated to be 0.1. To detect a difference of 0.1 (treatment response probability = 0.6, standard response probability = 0.5), with 90% power, 552 subject pairs will be needed.

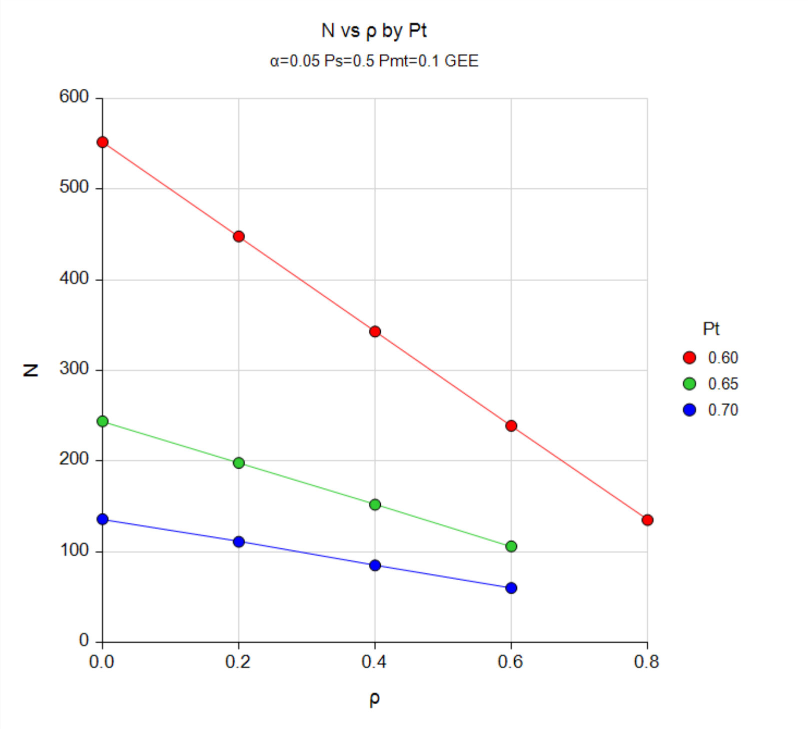
References

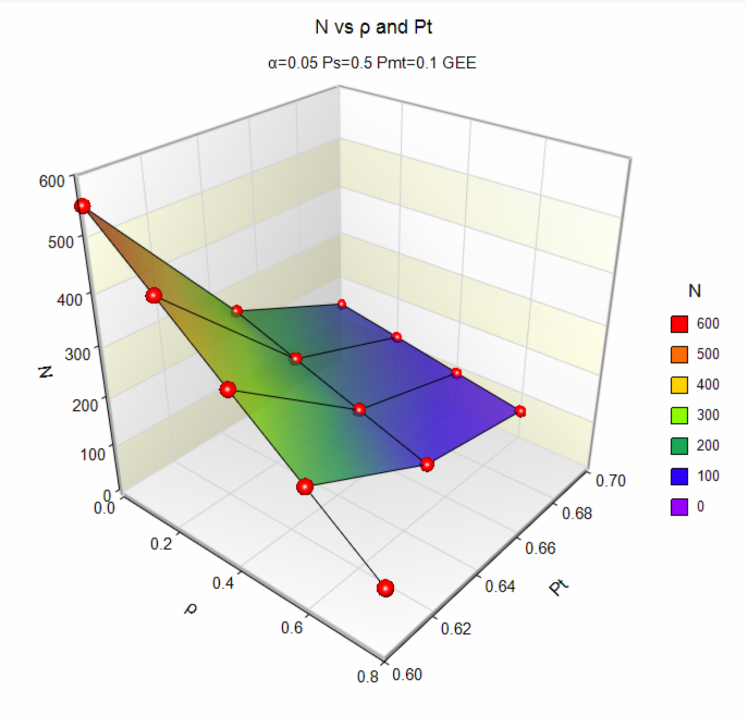
Zhang, S., Cao, J., Ahn, C. 2014. 'A GEE approach to determine sample size for pre- and post-intervention experiments with dropout'. Computational Statistics and Data Analysis. Volume 69. Pages 114-121.
Zhang, S., Cao, J., Ahn, C. 2017. 'Inference and sample size calculation for clinical trials with incomplete observations of paired binary outcomes'. Statistics in Medicine. Volume 36. Pages 581-591.

This report gives the sample size for each of the requested scenarios.

Plots Section

Plots





These plots show the sample size for the various combination of the other parameters.

Example 2 – Validation using Zhang, Cao, and Ahn (2014)

Zhang, Cao, and Ahn (2014) page 119 present Table 3 which provides examples that we can use to validate this procedure. The second set of three rows, last column, has the following settings: *Ps* = 0.1; *Pt* = 0.2; *ρ* = 0.0, 0.15, 0.3; *alpha* = 0.05; *power* = 0.8; and *Pmt* = 0.4. Sample size is calculated for a two-sided test. The resulting sample sizes are 257, 228, and 198.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Sample Size**
Alternative Hypothesis .... **Two-Sided (Pt ≠ Ps)**
Power... **0.8**
Alpha.. **0.05**
Pt Input Type .... **Pt**
Pt (Probability (Yt = 1)) **0.2**
Ps (Probability (Ys = 1))................................ **0.1**
P11 Input Type .... **ρ (Within-Subject Correlation)**
ρ (Within-Subject Correlation)....................... **0 0.15 0.3**
Pmt (Probability (Yt = Missing)) **0.4**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For: Sample Size
Test: Mixed Logistic Regression using GEE
Alternative Hypothesis: Two-Sided (Pt ≠ Ps)

Power	Total Sample Size N	Marginal Probabilities Pt	Marginal Probabilities Ps	Difference Pt - Ps	Within-Subject Correlation ρ	Joint Probability that Yt = 1 and Ys = 1 P11	Proportion Discordant P01 + P10	Probability Treatment is Missing Pmt	Alpha
0.8001	257	0.2	0.1	0.1	0.00	0.020	0.260	0.4	0.05
0.8015	228	0.2	0.1	0.1	0.15	0.038	0.224	0.4	0.05
0.8015	198	0.2	0.1	0.1	0.30	0.056	0.188	0.4	0.05

PASS matches the sample sizes of 257, 228, and 198. The procedure is validated.