Chapter 123

Tests for Two Means Assuming Equal Variances using a Bayesian Approach

Introduction

This procedure provides sample size and power calculations for one- or two-sided two-sample z-tests when the variances of the two groups (populations) are assumed to be known and equal. If the variances are unknown, estimates from previous studies can be used if the anticipated group sample sizes are at least 30.

Most of the procedures in PASS adopt the frequentist approach to power analysis. The methodology in this procedure uses what is called “hybrid classical and Bayesian” techniques. A prior distribution of the treatment effect parameter is included in the methodology. This allows uncertainty about the value of the treatment effect to be included that cannot be added in the classical frequentist approach.

This procedure is based on Kieser (2020), Ciarleglio *et al*. (2015), and Ciarleglio *et al*. (2016).

Test Assumptions

When running a two-sample equal-variance z-test, the basic assumptions are that the distributions of the two populations are normal, and that the variances of the two distributions are known and are the same. If those assumptions are not likely to be met (even approximately), another testing procedure should be used, and the corresponding procedure in **PASS** should be used for sample size or power calculations.

Technical Details

We will summarize the results shown in Kieser (2020).

Assume that we will have two groups of normally distributed responses with means μ_1 and μ_2 and common variance σ^2. We assume that group 1 is the treatment group and group 2 is the control group. The usual test statistic to compare the group means is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

A one-sided set of hypotheses is H_0: $\delta \leq 0$ versus H_1: $\delta > 0$. This tests whether the treatment mean is higher than the control mean.

If we let $N = N_1 + N_2$ and $R = N_2/N_1$, the traditional frequentist power is given by

$$TP(\delta) = \Pr(Z > z_{1-\alpha}|\delta) = 1 - \Phi\left(z_{1-\alpha} - \left(\frac{\delta}{\sigma}\right)\sqrt{N(R/(1+R)^2)}\right)$$

where $\Phi(X)$ is the standard normal CDF.

Here the value of δ is selected as mean difference that is large enough to be important and meaningful.

Suppose that during the planning phase of the study, there is some doubt about the appropriate value of δ to adopt. Instead, the research team is able to specify a reasonable distribution for δ as being normal with mean δ and variance σ_P^2. This distribution is called the *prior distribution* of δ.

The expected (average) power can be calculated as

$$EP = \int_{-\infty}^{\infty} \Pr(Z > z_{1-\alpha}|\delta) f(\delta) d\delta = 1 - \int_{-\infty}^{\infty} \Phi\left(z_{1-\alpha} - \left(\frac{\delta}{\sigma}\right)\sqrt{N(R/(1+R)^2)}\right) f(\delta) d\delta$$

were $f(\delta)$ is the normal density.

Note that *EP* provides the unconditional probability of rejecting the null hypothesis, assuming that the prior distribution is correct.

The expression for *EP* can be simplified to

$$EP = 1 - \Phi\left(\frac{z_{1-\alpha}\theta - \delta}{\sqrt{\theta^2 + \sigma_P^2}}\right)$$

where

$$\theta = \frac{(1+R)^2}{RN}\sigma^2$$

With this expression for the expected power, the minimum required sample size can be determined using a binary search.

Finding the Value of σ_P^2

A difficult step in the above is to find the appropriate value for the variance of the prior distribution. A useful tool during the planning process is to use the following relationship.

$$\sigma_P = \frac{-\delta}{\Phi^{-1}\left(\Pr(y < 0)\right)}$$

where $\Pr(y < 0)$ is a stated value such as 0.001, 0.01, or 0.02.

Using this tool, appropriate values σ_P can be found when other methods are not available.

Example 1 – Finding the Sample Size

Researchers wish to compare two types exercise regimens to determine whether there is a difference in strength. Subjects will be randomized to the treatment or control group, the corresponding exercise regimen will be used, and the strength will be measured using a standard test. The strength index in the control group has been measured in previous research at 22.8 with a standard deviation of 8.3.

The researchers determine that a gain in strength of 20% will be required before the new exercise regimen is of interest to others. This equates to a strength increase of 4.56.

The researchers want to base the required sample size on reaching 90% expected power. They want to see the impact of using three values for the standard deviation of the prior distribution: 1.5, 2, and 2.5.

Since they will use a one-sided z test, the significance level is set to 0.025.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Alternative Hypothesis	**One-Sided**
Expected Power	**0.90**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
δ (Diff of Group Means, μ1 - μ2)	**4.56**
σ (Std Dev of Responses)	**8.3**
σ_P Input Type	**Enter σ_P**
σ_P (Std Dev of Prior Distribution)	**1.5 2 2.5**

Output

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

Numeric Results

Numeric Results
Alternative Hypothesis: One-Sided

Expected Power					Mean Diff	Std Deviations			
Target	Actual	N1	N2	N	δ	Resp σ	Prior σ_P	P(D<0)	Alpha
0.9	0.90082	113	113	226	4.56	8.3	1.5	0.00118	0.025
0.9	0.90034	173	173	346	4.56	8.3	2.0	0.01130	0.025
0.9	0.90013	338	338	676	4.56	8.3	2.5	0.03408	0.025

References

Kieser, Meinhard. 2020. Methods and Applications of Sample Size Calculation and Recalculation in Clinical Trials. Springer. Cham, Switzerland.
Ciarleglio, M.M., Arendt, C.D., and Peduzzi, P.N. 2016. 'Selection of the effect size for sample size determination for a continuous response in a superiority clinical trial using a hybrid classical and Bayesian procedure'. Clinical Trials, Volume 13(3), pages 275-285. DOI: 10.1177.1740774516628825
Ciarleglio, M.M., Arendt, C.D., Makuch, R.W., and Peduzzi, P.N. 2015. 'Selection of the treatment effect for sample size determination in a superiority clinical trial using a hybrid classical and Bayesian procedure'. Contemporary Clinical Trials, Volume 41, pages 160-171. DOI: 10.1016/j.cct.2015.01.002

Report Definitions

Target Expected Power is the average probability of rejecting a false null hypothesis that is desired. This is a Bayesian-type definition of power. It is referred to as Bayesian predictive power, assurance, or probability of success.
Actual Expected Power is the average power obtained in this scenario. Because N1 and N2 are discrete, this value is often (slightly) larger than the target expected power.
N1 and N2 are the number of subjects sampled from the treatment and control populations, respectively.
N is the total sample size, N1 + N2.
δ is the difference between means ($\mu 1 - \mu 2$) assumed by the alternative hypothesis. It is also the mean of the prior distribution.
σ is the standard deviation of responses.
σ_P is the standard deviation of the normal prior distribution of difference.
P(D < 0) is the probability that the hypothesized difference is actually negative according to the prior distribution. This value is used to calculate σ_P.
Alpha is the probability of rejecting a true null hypothesis.

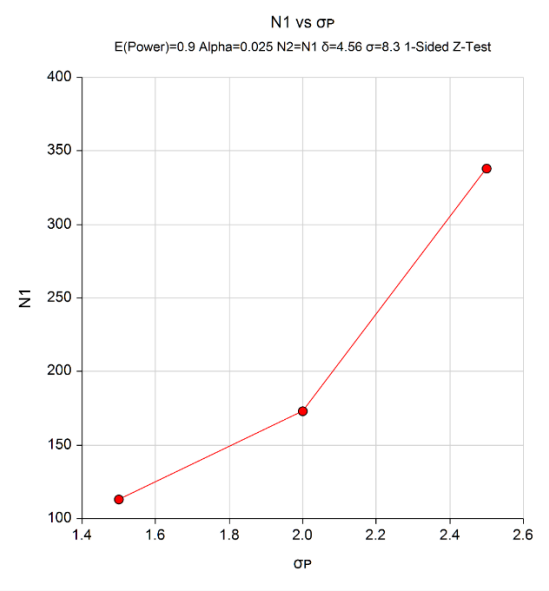
Summary Statements

Group sample sizes of 113 and 113 achieve an expected power of 90% to reject the null hypothesis of equal means. The expected power is calculated using the 'hybrid classical and Bayesian' technique (see Kieser, 2020). The population mean difference is 4.56 with a 'known' standard deviation for both groups of 8.3 and with a one-sided significance level (alpha) of 0.025 using a two-sample equal-variance z-test. The expected power is calculated by forming a weighted average of power values from all possible mean differences in which the weights are proportional to the normal distribution. This distribution is called the prior distribution and has a mean of 4.56 and a standard deviation of 1.5.

This report shows the values of each of the parameters, one scenario per row.

Chart Section

Chart Section



This plot shows the relationship between the standard deviation of the prior distribution and the group sample sizes.

Example 2 – Validation using Kieser (2020)

Kieser (2020) page 36 presents Table 3.3b which we will use as a validation example.

Subjects will be randomized equally to either a treatment or a control group. The value of the difference is 10 and the standard deviation is 20. The one-sided significance level is 0.025. The expect power is set to 0.90. The value of P(D < 0) is 0.01. The resulting group sample sizes are 200.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Alternative Hypothesis	**One-Sided**
Expected Power	**0.90**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
δ (Diff of Group Means, μ1 - μ2)	**10**
σ (Std Dev of Responses)	**20**
σ_P Input Type	**Enter P(D < 0)**
P(D < 0)	**0.01**

Output

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

Numeric Results

Numeric Results

Alternative Hypothesis: One-Sided

Expected Power					Mean Diff	Std Deviations			
Target	Actual	N1	N2	N	δ	Resp σ	Prior σ_P	P(D<0)	Alpha
0.9	0.90015	200	200	400	10	20	4.29858	0.01	0.025

PASS has also calculated the group sample size to be 200. The procedure is validated.