Chapter 853

Tests for the Difference Between Two Linear Regression Intercepts

Introduction

Linear regression is a commonly used procedure in statistical analysis. One of the main objectives in linear regression analysis is to test hypotheses about the slope and intercept of the regression equation. This module calculates power and sample size for testing whether two intercepts computed from two groups are significantly different.

Technical Details

Suppose that the dependence of a variable *Y* on another variable *X* can be modeled using the simple linear regression equation

$$Y = \alpha + \beta X + \varepsilon$$

In this equation, α is the *Y*-intercept parameter, β is the slope parameter, *Y* is the dependent variable, *X* is the independent variable, and ε is the error. The nature of the relationship between *Y* and *X* is studied using a sample of *n* observations. Each observation consists of a data pair: the *X* value and the *Y* value. The parameters α and β are estimated using simple linear regression. We will call these estimates *a* and *b*.

Since the linear equation will not fit the observations exactly, estimated values must be used. These estimates are found using the method of least squares. Using these estimated values, each data pair may be modeled using the equation

$$Y = a + bX + e$$

Note that *a* and *b* are the estimates of the population parameters α and β. The *e* values represent the discrepancies between the estimated values ($a + bX$) and the actual values *Y*. They are called the errors or residuals.

Two Groups

Suppose there are two groups, and a separate regression equation is calculated for each group. If it is assumed that these *e* values are normally distributed, a test of the hypothesis that $\alpha 1 = \alpha 2$ versus the alternative that they are unequal can be constructed. Dupont and Plummer (1998) state that the test statistic

$$D = \frac{(\hat{\alpha}_2 - \hat{\alpha}_1)\sqrt{n_2}}{S_R}$$

follows the Student's t distribution with v degrees of freedom where

$$v = n_1 + n_2 - 4$$

$$S_R^2 = \frac{s^2}{m}\left[1 + \frac{\bar{X}_1^2}{\sigma_{X_1}^2} + m\left\{1 + \frac{\bar{X}_2^2}{\sigma_{X_2}^2}\right\}\right]$$

$$m = \frac{n_1}{n_2}$$

$$\sigma_{X_1}^2 = \frac{1}{n_1}\sum\nolimits_j \left(X_{1j} - \bar{X}_1\right)^2$$

$$\sigma_{X_2}^2 = \frac{1}{n_2}\sum\nolimits_j \left(X_{2j} - \bar{X}_2\right)^2$$

$$s^2 = \frac{1}{n_1 + n_2 - 4}\sum\nolimits_{ij} \left(Y_{ij} - \hat{Y}_{ij}\right)^2$$

$$\hat{Y}_{ij} = a_i + b_i X_{ij}$$

The power function of difference in intercepts in a two-sided test is (see Dupont and Plummer, 1998) given by

$$\text{Power} = T_v\left[\delta\sqrt{n_2} - t_{v,\alpha/2}\right] + T_v\left[-\delta\sqrt{n_2} - t_{v,\alpha/2}\right]$$

where

$$N = n_1 + n_2$$

$$v = n_1 + n_2 - 4$$

$$m = \frac{n_1}{n_2}$$

$$\delta = \frac{\alpha_2 - \alpha_1}{\sigma_R}$$

$$\sigma_R^2 = \frac{\sigma^2}{m}\left[1 + \frac{\bar{X}_1^2}{\sigma_{X_1}^2} + m\left\{1 + \frac{\bar{X}_2^2}{\sigma_{X_2}^2}\right\}\right]$$

$$\sigma_{X_1}^2 = \frac{1}{n_1}\sum_j \left(X_{1j} - \bar{X}_1\right)^2$$

$$\sigma_{X_2}^2 = \frac{1}{n_2}\sum_j \left(X_{2j} - \bar{X}_2\right)^2$$

$$\sigma^2 = \mathrm{Var}(\varepsilon)$$

Note that σ^2 is estimated by s^2.

Example 1 – Finding Sample Size

Suppose a sample size needs to be found for a study to compare two intercepts. The basic design is to measure the value of the dependent variable, Y, at five values of the independent variable, X. These values are 10, 20, 30, 40, and 50. The same values of X will be used in both groups. Hence the value of both μX1 and μX2 is 30, and the value of both σX1 and σX2 is 14.1421. The other parameters of the study are a two-sided alpha of 0.05, power of 0.90, equal subject allocation to both groups, δ of 1, σ of 0.5, 0.7, and 0.9.

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 .... **δ ≠ 0**
Power.. **0.90**
Alpha... **0.05**
Group Allocation .... **Equal (N1 = N2)**
δ (α1-α2, Intercept Difference)....................... **1**
μX1 (Mean of X in Group 1)........................... **30**
μX2 (Mean of X in Group 2)........................... **30**
σ (SD of Residuals) **0.5 0.7 0.9**
σX1 (SD of X in Group 1)............................... **14.1421**
σX2 (SD of X in Group 2)............................... **14.1421**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Ha: δ = α1 - α2 ≠ 0

Power		Sample Size			Intercept Difference	Mean of X Values		Standard Deviation			
								Residuals	X Values		
Target	Actual	N1	N2	N	δ	μX1	μX2	σ	σX1	σX2	Alpha
0.9	0.9005	30	30	60	1	30	30	0.5	14.142	14.142	0.05
0.9	0.9017	58	58	116	1	30	30	0.7	14.142	14.142	0.05
0.9	0.9011	95	95	190	1	30	30	0.9	14.142	14.142	0.05

Target Power The desired power value (or values) entered in the procedure. Power is the probability of rejecting a false null hypothesis.
Actual Power The power obtained in this scenario. Because N1 and N2 are discrete, this value is often (slightly) larger than the target power.
N1 and N2 The number of items sampled from each group.
N The total sample size. N = N1 + N2.
δ The difference between population intercepts at which power and sample size calculations are made. δ = α1 - α2.
μX1 and μX2 The means of the basic X values in groups 1 and 2 of your design, respectively.
σ The standard deviation of the residuals.
σX1 and σX2 The standard deviations of the basic X values in groups 1 and 2 of your design, respectively. Note that the divisor is n, not n - 1.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A two-group simple linear regression (Y versus X) design will be used to test whether the Group 1 intercept (α1) is different from the Group 2 intercept (α2) (H0: δ = 0 versus H1: δ ≠ 0, δ = α1 - α2). The comparison will be made using a two-sided intercept-difference t-test, with a Type I error rate (α) of 0.05. The mean of X for Group 1 is assumed to be 30 and the mean of X for Group 2 is assumed to be 30. The standard deviation of X for Group 1 is assumed to be 14.142 and the standard deviation of X for Group 2 is assumed to be 14.142. The common standard deviation of residuals for both groups is assumed to be 0.5. To detect an intercept difference (δ = α1 - α2) of 1 with 90% power, the number of needed subjects will be 30 in Group 1 and 30 in Group 2.

Dropout-Inflated Sample Size

	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
Dropout Rate	**N1**	**N2**	**N**	**N1'**	**N2'**	**N'**	**D1**	**D2**	**D**
20%	30	30	60	38	38	76	8	8	16
20%	58	58	116	73	73	146	15	15	30
20%	95	95	190	119	119	238	24	24	48

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N1, N2, and N The evaluable sample sizes at which power is computed. If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N' The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. After solving for N1 and N2, N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D1, D2, and D The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2.

Dropout Summary Statements

Anticipating a 20% dropout rate, 38 subjects should be enrolled in Group 1, and 38 in Group 2, to obtain final group sample sizes of 30 and 30, respectively.

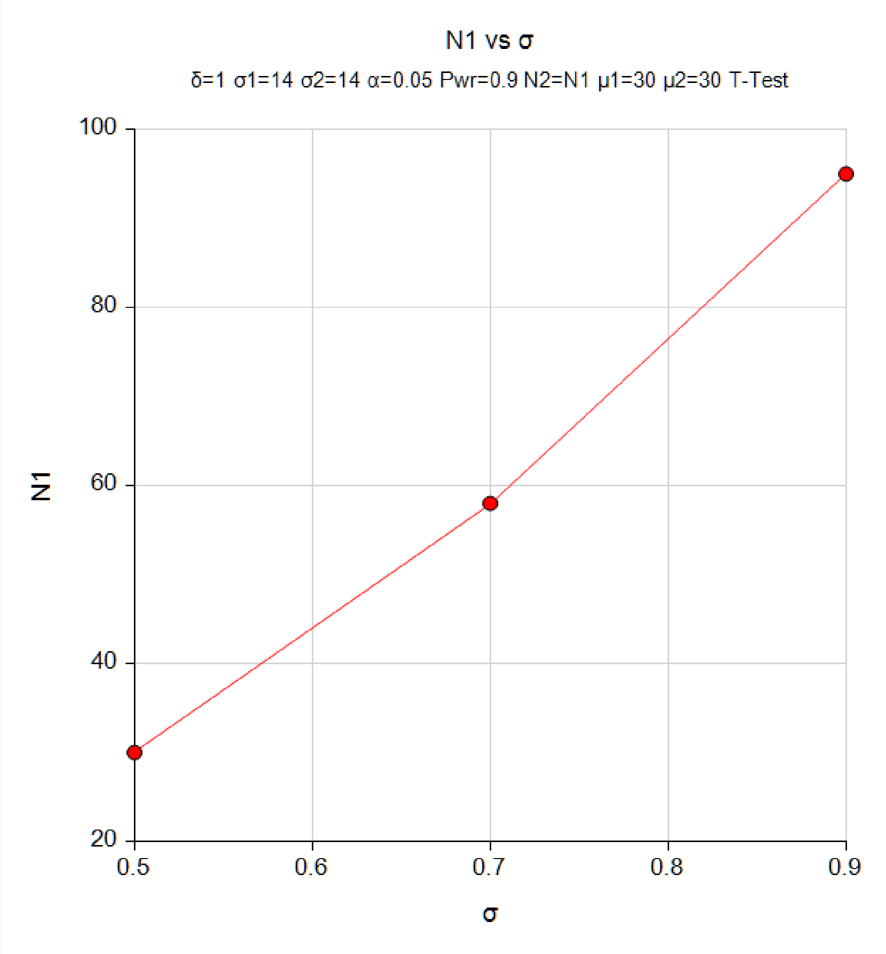
References

Dupont, W.D. and Plummer, W.D. Jr. 1998. Power and Sample Size Calculations for Studies Involving Linear Regression. Controlled Clinical Trials. Vol 19. Pages 589-601.

This report shows the calculated sample size for each of the scenarios. Note that since the design has five X values, the sample sizes would be rounded up to the first multiple of 5 above the designated value. For example, the value of 58 would be rounded up to 60.

Plots Section

Plots



This plot shows the sample size required for each value of σ.

Example 2 – Validation by Manual Calculation

We could not find a published validation example, so we will calculate the first row of the above report manually. The first row of Example 1 is

Numeric Results

Solve For: Power
Alternative Hypothesis: Ha: δ = α1 - α2 ≠ 0

	Sample Size			Intercept Difference	Mean of X Values		Standard Deviation Residuals	Standard Deviation X Values		
Power	N1	N2	N	δ	μX1	μX2	σ	σX1	σX2	Alpha
0.9005	30	30	60	1	30	30	0.5	14.142	14.142	0.05

The power function of difference in intercepts in a two-sided test is (see Dupont and Plummer, 1998) given by

$$\text{Power} = T_v\left[\delta\sqrt{n_2} - t_{v,\alpha/2}\right] + T_v\left[-\delta\sqrt{n_2} - t_{v,\alpha/2}\right]$$

Plugging in the values, including results from **PASS's** Probability Calculator, we obtain

$$N = n_1 + n_2 = 60$$

$$v = n_1 + n_2 - 4 = 56$$

$$m = \frac{n_1}{n_2} = 1$$

$$\sigma_R^2 = \frac{\sigma^2}{m}\left[1 + \frac{\bar{X}_1^2}{\sigma_{X_1}^2} + m\left\{1 + \frac{\bar{X}_2^2}{\sigma_{X_2}^2}\right\}\right] = \frac{0.25}{1}\left[1 + \frac{900}{200} + m\left\{1 + \frac{900}{200}\right\}\right] = \frac{11}{4} = 2.75$$

$$\delta = \frac{\alpha_2 - \alpha_1}{\sigma_R} = \frac{1}{\sqrt{2.75}} = 0.603022689$$

$$t_{v,\alpha/2} = t_{56,0.975} = 2.0032407188$$

$$\text{Power} = T_v\left[\delta\sqrt{n_2} - t_{v,\frac{\alpha}{2}}\right] + T_v\left[-\delta\sqrt{n_2} - t_{v,\frac{\alpha}{2}}\right]$$

$$= T_{56}\left[0.603\sqrt{30} - 2.0032407188\right] + T_{56}\left[-0.603\sqrt{30} - 2.0032407188\right]$$

$$= T_{56}[1.29965] + T_{56}[-4.60254187] = 0.90047$$

The 0.90047 matches the 0.9005 on the report.