

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

$$Y = a + bX + \varepsilon$$

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

Tests for the Difference Between Two Linear Regression Intercepts

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 ν degrees of freedom where

$$\nu = n_1 + n_2 - 4$$

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

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

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

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

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

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

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

$$\text{Power} = T_\nu[\delta\sqrt{n_2} - t_{\nu,\alpha/2}] + T_\nu[-\delta\sqrt{n_2} - t_{\nu,\alpha/2}]$$

where

$$N = n_1 + n_2$$

$$\nu = n_1 + n_2 - 4$$

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

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

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

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

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

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

Note that σ^2 is estimated by s^2 .

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

Test Direction

Alternative Hypothesis

Specify the alternative hypothesis of the test. This option implicitly determines if the test is one-sided ($\delta > 0$ or $\delta < 0$) or two-sided ($\delta \neq 0$).

When you choose a one-sided alternative, you should be sure that you enter a δ value that is consistent with your choice. For example, if you select $\delta > 0$, all values of δ should be greater than zero.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on *N1* or *N2* or both. In most cases, the sample sizes *N1* and *N2* should be multiples of the number of unique X's to allow for a complete design set.

Tests for the Difference Between Two Linear Regression Intercepts

The options are

- Equal ($N1 = N2$)**
 This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.
- Enter $N1$, solve for $N2$**
 Select this option when you wish to fix $N1$ at some value (or values), and then solve only for $N2$. Please note that for some values of $N1$, there may not be a value of $N2$ that is large enough to obtain the desired power.
- Enter $N2$, solve for $N1$**
 Select this option when you wish to fix $N2$ at some value (or values), and then solve only for $N1$. Please note that for some values of $N2$, there may not be a value of $N1$ that is large enough to obtain the desired power.
- Enter $R = N2/N1$, solve for $N1$ and $N2$**
 For this choice, you set a value for the ratio of $N2$ to $N1$, and then PASS determines the needed $N1$ and $N2$, with this ratio, to obtain the desired power. An equivalent representation of the ratio, R , is

$$N2 = R * N1.$$
- Enter percentage in Group 1, solve for $N1$ and $N2$**
 For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed $N1$ and $N2$ with this percentage to obtain the desired power.

$N1$ (Sample Size, Group 1)

This option is displayed if Group Allocation = "Enter $N1$, solve for $N2$ "

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2 . You can enter a single value or a series of values.

$N2$ (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter $N2$, solve for $N1$ "

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter $R = N2/N1$, solve for $N1$ and $N2$."

R is the ratio of $N2$ to $N1$. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of $N2$ to $N1$ while solving for $N1$ and $N2$. Only sample size combinations with this ratio are considered.

$N2$ is related to $N1$ by the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., $N1 = 10$ and $N2 = 20$, or $N1 = 50$ and $N2 = 100$).

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single or a series of values.

Tests for the Difference Between Two Linear Regression Intercepts

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for $N1$ and $N2$. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When Not Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal ($N1 = N2$)**
This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.
- **Enter N1 and N2 individually**
This choice permits you to enter different values for $N1$ and $N2$.
- **Enter N1 and R, where $N2 = R * N1$**
Choose this option to specify a value (or values) for $N1$, and obtain $N2$ as a ratio (multiple) of $N1$.
- **Enter total sample size and percentage in Group 1**
Choose this option to specify a value (or values) for the total sample size (N), obtain $N1$ as a percentage of N , and then $N2$ as $N - N1$.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal ($N1 = N2$)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for $N1$, and also the value for $N2$.

The Sample Size Per Group must be ≥ 2 . You can enter a single value or a series of values.

N1 (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter N1 and N2 individually" or "Enter N1 and R, where $N2 = R * N1$."*

$N1$ is the number of items or individuals sampled from the Group 1 population.

$N1$ must be ≥ 2 . You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter N1 and N2 individually."

$N2$ is the number of items or individuals sampled from the Group 2 population.

$N2$ must be ≥ 2 . You can enter a single value or a series of values.

Tests for the Difference Between Two Linear Regression Intercepts

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter N1 and R, where $N2 = R * N1$."*

R is the ratio of $N2$ to $N1$. That is,

$$R = N2/N1$$

Use this value to obtain $N2$ as a multiple (or proportion) of $N1$.

$N2$ is calculated from $N1$ using the formula:

$$N2 = [R \times N1],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If $R < 1$, then $N2$ will be less than $N1$; if $R > 1$, then $N2$ will be greater than $N1$. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines $N1$ and $N2$.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Difference Between Intercepts

δ ($\alpha_1 - \alpha_2$, Intercept Difference)

Enter a value for the assumed difference between the intercepts of Groups 1 and 2. This difference is the difference at which the design is powered to reject equal intercepts. δ can be any non-zero value (positive or negative), but it should be consistent with the choice of the alternative hypothesis.

You can enter a single value such as 2 or a series of values such as 2 3 4 or 1 to 5 by 1. When a series of values is entered, PASS will generate a separate report line for each value of the series.

Means of X

μ_{Xi} (Mean of X in Group i)

Enter values for the assumed means of groups 1 and 2, respectively. These means are only used in the calculation of the variation in the test statistic. They are NOT used in the specification of the values of the intercepts. These are the means of the same X values that are used to calculate σ_{X1} and σ_{X2} .

These means can be any value (positive, negative, or zero).

Tests for the Difference Between Two Linear Regression Intercepts

You can enter a single value such as 10 or a series of values such as 10 20 30. When a series of values is entered, PASS will generate a separate calculation result for each value of the series.

Example

Suppose a study is planned to study the linear relationship between X and Y. Further suppose that multiple observations will be made at each of five X values: 10 20 30 40 50. For this design, $\mu_X = 30$ and $\sigma_X = 14.1421$. Since there are five unique values, the final sample size should be a multiple of 5 so that an equal number of subjects are available for each X value.

Standard Deviations of Residuals and X

σ (SD of Residuals)

Enter one or more values of the standard deviation of the residuals from the regression of Y on X. It can be obtained from a prior study as the square root of the residual mean square (often called MSE). The value (or values) must be greater than zero.

σ_{Xi} (SD of X in Group i)

This is the standard deviation of the X values in the sample calculated using the population formula (divide sum of squares by n , not $n-1$). It is not necessarily the standard deviation of the X variable, but it can be. This depends on whether X is a random variable or fixed variable. For example, suppose the X values are five -1's and five 1's. The standard deviation of these values (dividing by n rather than $n-1$) is 1.0. This is the value of σ_{Xi} .

The value must be greater than zero.

σ Button

An easy way to determine σ_{Xi} for a particular configuration of X's is to use the *Standard Deviation Estimator* tool. This tool can be quickly accessed by pressing the σ button on the right side of the panel.

Once this tool is open, click the Data tab. Set *Solve For* to 'Population Standard Deviation'. Next, enter the set of X values in the *Data Values* column. The value of σ_{Xi} will be displayed in the *Population Standard Deviation* box.

Example

Suppose the following fixed set of X values have been selected for your design: 1, 2, 3, 7.

The standard deviation of these values using the population formula is 2.22776. This is the value of σ_{Xi} . Assuming you use the same values for both groups (the usual case), you will set but σ_{X1} and σ_{X2} to 2.22776.

Tests for the Difference Between Two Linear Regression Intercepts

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 μ_{X1} and μ_{X2} is 30, and the value of both σ_{X1} and σ_{X2} 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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for the Difference Between Two Linear Regression Intercepts** procedure window by clicking on **Regression**, and then clicking on **Tests for the Difference Between Two Linear Regression Intercepts**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	$\delta \neq 0$
Power.....	0.90
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
δ ($\alpha_1 - \alpha_2$, Intercept Difference)	1
μ_{X1} (Mean of X in Group 1).....	30
μ_{X2} (Mean of X in Group 2).....	30
σ (SD of Residuals)	0.5 0.7 0.9
σ_{X1} (SD of X in Group 1)	14.1421
σ_{X2} (SD of X in Group 2)	14.1421

Annotated Output

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

Numeric Results

Numeric Results for Testing the Difference Between Two Intercepts

Alternative Hypothesis: $\delta = \alpha_1 - \alpha_2 \neq 0$

Power	N1	N2	Total N	Intcpt	SD of	SD	SD	Mean	Mean	Alpha
				$\alpha_1 - \alpha_2$ δ		Resid σ	of X in Grp 1 σ_{X1}	of X in Grp 2 σ_{X2}	of X in Grp 1 μ_{X1}	
0.9005	30	30	60	1.000	0.500	14.142	14.142	30.000	30.000	0.050
0.9017	58	58	116	1.000	0.700	14.142	14.142	30.000	30.000	0.050
0.9011	95	95	190	1.000	0.900	14.142	14.142	30.000	30.000	0.050

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.

Tests for the Difference Between Two Linear Regression Intercepts

Report Definitions

Power is the probability of rejecting a false null hypothesis. Because $N1$ and $N2$ are integers, this value is often (slightly) larger than the target (input) power.

$N1$ and $N2$ are the number of items sampled from each group.

N is the total sample size, $N1 + N2$.

δ ($\alpha_1 - \alpha_2$) is the difference between population intercepts at which power and sample size calculations are made.

σ is the standard deviation of the residuals.

$\sigma X1$ and $\sigma X2$ are the standard deviations of the basic X values in groups 1 and 2 of your design, respectively.

Note that the divisor is n , not $n-1$.

$\mu X1$ and $\mu X2$ are the means of the basic X values in groups 1 and 2 of your design, respectively.

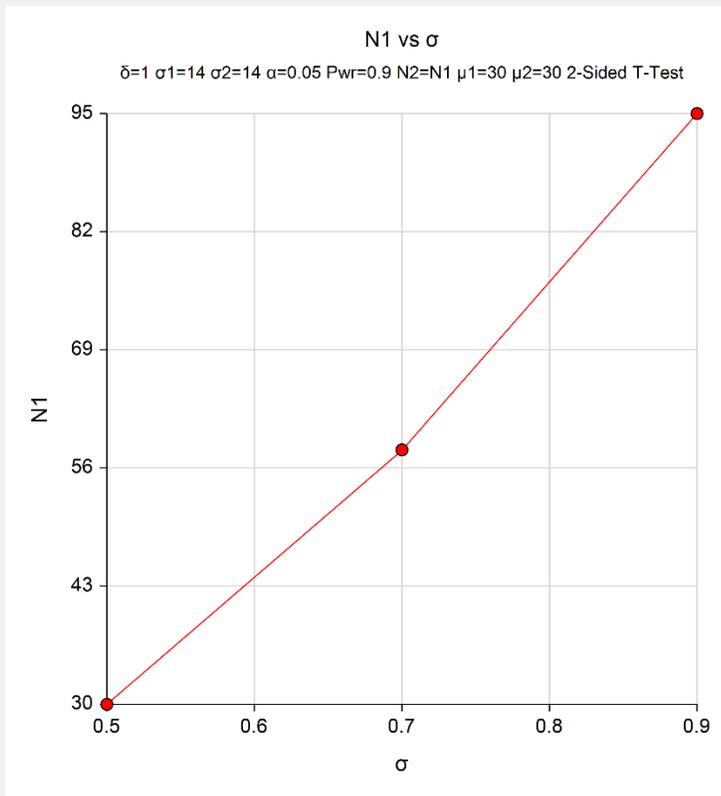
Alpha is the probability of rejecting a true null hypothesis.

Summary Statements

Group sample sizes of 30 and 30 achieve 90.05% power to reject the null hypothesis of equal intercepts when the actual difference in population intercepts is 30.000 with X means of 30.000 for group 1 and 30.000 for group 2, X standard deviations of 14.142 for group 1 and 14.142 for group 2, a standard deviation of residuals of 0.500, and a significance level (alpha) of 0.050 using a two-sided test.

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



This plot show 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 for Testing the Difference Between Two Intercepts										
Alternative Hypothesis: $\delta = \alpha_1 - \alpha_2 \neq 0$										
Power	N1	N2	Total N	Intcpt Diff $\alpha_1 - \alpha_2$ δ	SD of Resid σ	SD of X in Grp 1 σ_{X1}	SD of X in Grp 2 σ_{X2}	Mean of X in Grp 1 μ_{X1}	Mean of X in Grp 2 μ_{X2}	Alpha
0.9005	30	30	60	1.000	0.500	14.142	14.142	30.000	30.000	0.050

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

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

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

$$N = n_1 + n_2 = 60$$

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

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

$$\sigma_R^2 = \frac{\sigma^2}{m} \left[1 + \frac{\bar{X}_1^2}{\sigma_{X1}^2} + m \left\{ 1 + \frac{\bar{X}_2^2}{\sigma_{X2}^2} \right\} \right] = \frac{0.25}{1} \left[1 + \frac{900}{200} + m \left\{ 1 + \frac{900}{200} \right\} \right] = \frac{11}{4} = 2.75$$

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

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

$$\begin{aligned} \text{Power} &= T_v \left[\delta\sqrt{n_2} - t_{v,\alpha/2} \right] + T_v \left[-\delta\sqrt{n_2} - t_{v,\alpha/2} \right] \\ &= T_{56} [0.603\sqrt{30} - 2.0032407188] + T_{56} [-0.603\sqrt{30} - 2.0032407188] \\ &= T_{56} [1.29965] + T_{56} [-4.60254187] \\ &= 0.90047 \end{aligned}$$

The 0.90047 matches the 0.9005 on the report.