

Chapter 226

Analysis of Covariance (ANCOVA) with Two Groups

Introduction

This procedure performs analysis of covariance (ANCOVA) for a grouping variable with 2 groups and one covariate variable. This procedure uses multiple regression techniques to estimate model parameters and compute least squares means. This procedure also provides standard error estimates for least squares means and their differences and computes the T-test for the difference between group means adjusted for the covariate. The procedure also provides response vs covariate by group scatter plots and residuals for checking model assumptions.

This procedure will output results for a simple two-sample equal-variance T-test if no covariate is entered and simple linear regression if no group variable is entered. This allows you to complete the ANCOVA analysis if either the group variable or covariate is determined to be non-significant. For additional options related to the T-test and simple linear regression analyses, we suggest you use the corresponding procedures in **NCSS**.

The group variable in this procedure is restricted to two groups. If you want to perform ANCOVA with a group variable that has three or more groups, use the One-Way Analysis of Covariance (ANCOVA) procedure.

This procedure cannot be used to analyze models that include more than one covariate variable or more than one group variable. If the model you want to analyze includes more than one covariate variable and/or more than one group variable, use the General Linear Models (GLM) for Fixed Factors procedure instead.

Kinds of Research Questions

A large amount of research consists of studying the influence of a set of independent variables on a response (dependent) variable. Many experiments are designed to look at the influence of a single independent variable (factor or group) while holding other factors constant. These experiments are called single-factor or single-group experiments and are analyzed with the one-way analysis of variance (ANOVA) or a two-sample T-test. Analysis of covariance (ANCOVA) is useful when you want to improve precision by removing extraneous sources of variation from your study by including a covariate.

The ANCOVA Model

The analysis of covariance uses features from both analysis of variance and multiple regression. The usual one-way classification model in analysis of variance is

$$Y_{ij} = \mu_i + e_{1ij}$$

where Y_{ij} is the j^{th} observation in the i^{th} group, μ_i represents the true mean of the i^{th} group, and e_{1ij} are the residuals or errors in the above model (usually assumed to be normally distributed). Suppose you have measured a second variable with values X_{ij} that is linearly related to Y . Further suppose that the slope of the relationship between Y and X is constant from group to group. You could then write the analysis of covariance model assuming equal slopes as

$$Y_{ij} = \mu_i + \beta(X_{ij} - \bar{X}_{..}) + e_{2ij}$$

where $\bar{X}_{..}$ represents the overall mean of X . If X and Y are closely related, you would expect that the errors, e_{2ij} , would be much smaller than the errors, e_{1ij} , giving you more precise results.

The classical analysis of covariance is useful for many reasons, but it does have the (highly) restrictive assumption that the slope is constant over all the groups. This assumption is often violated, which limits the technique's usefulness. You will want to study more about this technique in statistical texts before you use it.

If it is not reasonable to conclude that the slopes are equal, then a covariate-by-group interaction term should be included in the model.

Assumptions

The following assumptions are made when using the F-test.

1. The response variable is continuous.
2. The treatments do not affect the value of the covariate, X_{ij} .
3. The e_{2ij} follow the normal probability distribution with mean equal to zero.
4. The variances of the e_{2ij} are equal for all values of i and j .
5. The individuals are independent.

Normality of Residuals

The residuals are assumed to follow the normal probability distribution with zero mean and constant variance. This can be evaluated using a normal probability plot of the residuals. Also, normality tests are used to evaluate this assumption. The most popular of the five normality tests provided is the Shapiro-Wilk test.

Unfortunately, a breakdown in any of the other assumptions results in a departure from this assumption as well. Hence, you should investigate the other assumptions first, leaving this assumption until last.

Limitations

There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can assume that you have met the continuous variable assumption. Perhaps the greatest restriction is that your data comes from a random sample of the population. If you do not have a random sample, the F-test will not work.

Representing Group Variables

Categorical group variables take on only a few unique values. For example, suppose a therapy variable has three possible values: A, B, and C. One question is how to include this variable in the regression model. At first glance, we can convert the letters to numbers by recoding A to 1, B to 2, and C to 3. Now we have numbers. Unfortunately, we will obtain completely different results if we recode A to 2, B to 3, and C to 1. Thus, a direct recode of letters to numbers will not work.

To convert a categorical variable to a form usable in regression analysis, we must create a new set of numeric variables. If a categorical variable has k values, $k - 1$ new binary variables must be generated.

Indicator (Binary) Variables

Indicator (dummy or binary) variables are created as follows. A *reference group* is selected. Usually, the most common value or the control is selected as the reference group. Next, a variable is generated for each of the groups other than the reference group. For example, suppose that B is selected as the reference group. An indicator variable is generated for the remaining value, A. The value of the indicator variable is one if the value of the original variable is equal to the value of interest, or zero otherwise. Here is how the original variable T and the two new indicator variable TA looks in a short example.

T	TA
A	1
A	1
B	0
B	0

The generated variable, TA , would be used as columns in the design matrix, X , in the model.

Representing Interactions of Numeric and Categorical Variables

When the interaction between a group variable and a covariate is to be included in the model, all proceeds as above, except that an interaction variable must be generated for each categorical variable. This can be accomplished automatically in **NCSS** based on the slopes assumption. When assuming that the slopes are unequal all applicable covariate-by-group interaction variables are automatically created.

In the following example, the interaction between the group variable T and the covariate variable X is created. When the variable XTA is added to the model, it will account for the interaction between T and X .

T	TA	X	XTA
A	1	1.2	1.2
A	1	1.4	1.4
B	0	2.3	0
B	0	4.7	0

Technical Details

This section presents the technical details of the analysis method (multiple regression) using a mixture of summation and matrix notation.

The Linear Model

The linear model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

where \mathbf{Y} is a vector of N responses, \mathbf{X} is an $N \times p$ design matrix, $\boldsymbol{\beta}$ is a vector of p fixed and unknown parameters, and \mathbf{e} is a vector of N unknown, random error values. Define the following vectors and matrices:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1j} & \cdots & x_{pj} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1N} & \cdots & x_{pN} \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_j \\ \vdots \\ e_N \end{bmatrix}, \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

\mathbf{X} is the design matrix that includes the covariate, a binary variable formed from the group variable, and a variable resulting from the covariate-by-group interaction (if included).

Least Squares

Using this notation, the least squares estimates of the model coefficients, \mathbf{b} , are found using the equation.

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

The vector of predicted values of the response variable is given by

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$$

The residuals are calculated using

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$

Estimated Variances

An estimate of the variance of the residuals is computed using

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{N - p - 1}$$

An estimate of the variance of the model coefficients is calculated using

$$\mathbf{v} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix} = s^2(\mathbf{X}'\mathbf{X})^{-1}$$

Analysis of Covariance (ANCOVA) with Two Groups

An estimate of the variance of the predicted mean of Y at a specific value of X , say X_0 , is given by

$$s_{Y_m|X_0}^2 = s^2(1, X_0)(\mathbf{X}'\mathbf{X})^{-1} \begin{pmatrix} 1 \\ X_0 \end{pmatrix}$$

An estimate of the variance of the predicted value of Y for an individual for a specific value of X , say X_0 , is given by

$$s_{Y_I|X_0}^2 = s^2 + s_{Y_m|X_0}^2$$

Hypothesis Tests of the Intercept and Coefficients

Using these variance estimates and assuming the residuals are normally distributed, hypothesis tests may be constructed using the Student's t distribution with $N - p - 1$ degrees of freedom using

$$t_{b_i} = \frac{b_i - B_i}{s_{b_i}}$$

Usually, the hypothesized value of B_i is zero, but this does not have to be the case.

Confidence Intervals of the Intercept and Coefficients

A $100(1 - \alpha)\%$ confidence interval for the true model coefficient, β_i , is given by

$$b_i \pm (t_{1-\alpha/2, N-p-1})s_{b_i}$$

Confidence Interval of Y for Given X

A $100(1 - \alpha)\%$ confidence interval for the mean of Y at a specific value of X , say X_0 , is given by

$$b'X_0 \pm (t_{1-\alpha/2, N-p-1})s_{Y_m|X_0}$$

A $100(1 - \alpha)\%$ prediction interval for the value of Y for an individual at a specific value of X , say X_0 , is given by

$$b'X_0 \pm (t_{1-\alpha/2, N-p-1})s_{Y_I|X_0}$$

R² (Percent of Variation Explained)

Several measures of the goodness-of-fit of the model to the data have been proposed, but by far the most popular is R^2 . R^2 is the square of the correlation coefficient between Y and \hat{Y} . It is the proportion of the variation in Y that is accounted by the variation in the independent variables. R^2 varies between zero (no linear relationship) and one (perfect linear relationship).

R^2 , officially known as the *coefficient of determination*, is defined as the sum of squares due to the linear regression model divided by the adjusted total sum of squares of Y . The formula for R^2 is

$$R^2 = 1 - \left(\frac{\mathbf{e}'\mathbf{e}}{\mathbf{Y}'\mathbf{Y} - \frac{(\mathbf{1}'\mathbf{Y})^2}{\mathbf{1}'\mathbf{1}}} \right)$$

$$= \frac{SS_{Model}}{SS_{Total}}$$

R^2 is probably the most popular measure of how well a model fits the data. R^2 may be defined either as a ratio or a percentage. Since we use the ratio form, its values range from zero to one. A value of R^2 near zero indicates no linear relationship, while a value near one indicates a perfect linear fit. Although popular, R^2 should not be used indiscriminately or interpreted without scatter plot support. Following are some qualifications on its interpretation:

1. *Additional independent variables.* It is possible to increase R^2 by adding more independent variables, but the additional independent variables may cause an increase in the mean square error, an unfavorable situation. This usually happens when the sample size is small.
2. *Range of the independent variables.* R^2 is influenced by the range of the independent variables. R^2 increases as the range of the X 's increases and decreases as the range of the X 's decreases.
3. *Slope magnitudes.* R^2 does not measure the magnitude of the slopes.
4. *Linearity.* R^2 does not measure the appropriateness of a linear model. It measures the strength of the linear component of the model. Suppose the relationship between X and Y was a perfect sphere. Although there is a perfect relationship between the variables, the R^2 value would be zero.
5. *Predictability.* A large R^2 does not necessarily mean high predictability, nor does a low R^2 necessarily mean poor predictability.
6. *Sample size.* R^2 is highly sensitive to the number of observations. The smaller the sample size, the larger its value.

\bar{R}^2 (Adjusted R^2)

R^2 varies directly with N , the sample size. In fact, when $N = p$, $R^2 = 1$. Because R^2 is so closely tied to the sample size, an adjusted R^2 value, called \bar{R}^2 , has been developed. \bar{R}^2 was developed to minimize the impact of sample size. The formula for \bar{R}^2 is

$$\bar{R}^2 = 1 - \frac{(N - 1)(1 - R^2)}{N - p - 1}$$

Least Squares Means

As opposed to raw or arithmetic means which are simply averages of the grouped raw data values, least squares means are adjusted for the other terms in the model, such as covariates. In balanced designs with no covariates, the least squares group means will be equal to the raw group means. In unbalanced designs or when covariates are present, the least squares means usually are different from the raw means.

The least squares means and associated comparisons (i.e., differences) can be calculated using a linear contrast vector, \mathbf{c}_i . Means and differences are estimated as

$$\mathbf{c}_i' \mathbf{b},$$

with estimated standard error,

$$SE(\mathbf{c}_i' \mathbf{b}) = s \sqrt{\mathbf{c}_i' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{c}_i}$$

where s is the square root of the estimated mean square error (MSE) from the model based on ν degrees of freedom.

For an ANCOVA model with a group variable with 2 levels and a covariate, and if level 2 were the reference value, the components of the contrast vector would take the form

$$\mathbf{c}_i = (I, \mu_1, X)$$

where I represents the indicator for the intercept and X is the value of the covariate which the mean or difference is evaluated. The contrast vector used to estimate μ_2 would be

$$\mathbf{c}_i = (1, 0, X)$$

The contrast vector used to estimate $\mu_1 - \mu_2$ would be

$$\mathbf{c}_i = (0, 1, 0)$$

Confidence intervals for the estimable functions are of the form

$$\mathbf{c}_i' \mathbf{b} \pm c_\alpha SE(\mathbf{c}_i' \mathbf{b})$$

where c_α is the critical value, usually selected to keep the experimentwise error rate equal to α for the collection of all comparisons (see Multiple Comparisons).

Two-sided significance tests for the mean and the difference (against a null value of zero) use the test statistic

$$T_i = \frac{|\mathbf{c}_i' \mathbf{b}|}{SE(\mathbf{c}_i' \mathbf{b})} \geq c_\alpha$$

Data Structure

The data must be entered in a format that puts the responses in one column, the group values in a second column, and the covariate values in a third column. An example of data that might be analyzed using this procedure is shown next. The data contains a response variable, a group variable with two groups (State), and a covariate (Age).

ANCOVA2Grp Dataset

State	Age	Response
Iowa	12	100
Iowa	13	102
Iowa	12	97
.	.	.
.	.	.
.	.	.
Utah	14	104
Utah	11	105
Utah	12	106
.	.	.
.	.	.
.	.	.

Example 1 – ANCOVA Model Assuming Unequal Slopes (with Covariate-by-Group Interaction)

This section presents an example of how to run an analysis of the data presented above. These data are contained in the ANCOVA2Grp dataset. In this example, the responses from two states, Iowa and Utah, are compared with an adjustment for the age of the respondent. The two groups will be compared at three different covariate values.

This example will run all reports and plots so that they may be documented.

Setup

To run this example, complete the following steps:

1 Open the ANCOVA2Grp example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **ANCOVA2Grp** and click **OK**.

2 Specify the Analysis of Covariance (ANCOVA) with Two Groups procedure options

- Find and open the **Analysis of Covariance (ANCOVA) with Two Groups** procedure using the menus or the Procedure Navigator.
- The settings 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.

Variables Tab

Response Variable(s)	Response
Group Variable.....	State
Covariate Variable	Age
Calc Group Means at.....	12 Mean 18

Reports Tab

Run Summary.....	Checked
Descriptive Statistics.....	Checked
ANOVA Table	Checked
Coefficient T-Tests.....	Checked
Coefficient C.I.'s.....	Checked
Least Squares Means.....	Checked
Least Squares Means with Hypothesis Tests	Checked
of H0 Mean = 0	
T-Tests for Group Least Squares Mean	Checked
Differences	
Show Confidence Intervals	Checked
Residual Normality Tests.....	Checked
Residuals.....	Checked

Analysis of Covariance (ANCOVA) with Two Groups

Predicted Values for Means..... **Checked**
 Predicted Values for Individuals..... **Checked**

Plots Tab

Response vs Covariate by Group Scatter Plot..... **Checked**
 Means Plots..... **Checked**
 Group Comparison Plots **Checked**
 Histogram **Checked**
 Probability Plot..... **Checked**
 Resids vs Yhat..... **Checked**
 Resids vs X..... **Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Run Summary**Run Summary**

Response Variable: Response
 Group Variable: State
 Reference Group: "Iowa"
 Covariate Variable: Age
 Slopes Assumed to be: Unequal
 Model: Age + State + Age*State

Parameter	Value	Rows	Value
R ²	0.3947	Rows Processed	20
Adj R ²	0.2812	Rows Filtered Out	0
Coefficient of Variation	0.0266	Rows with Response Missing	0
Mean Square Error	7.445498	Rows with Group or Covariate Missing	0
Square Root of MSE	2.728644	Rows Used in Estimation	20
Ave Abs Pct Error	1.711	Completion Status	Normal Completion
Error Degrees of Freedom	16		

This report summarizes the results. It presents the variables used, the model, the number of rows used, and basic summary results.

R²

R^2 , officially known as the *coefficient of determination*, is defined as

$$R^2 = \frac{SS_{Model}}{SS_{Total(Adjusted)}}$$

R^2 is probably the most popular measure of how well a model fits the data. R^2 may be defined either as a ratio or a percentage. Since we use the ratio form, its values range from zero to one. A value of R^2 near zero indicates no linear relationship, while a value near one indicates a perfect linear fit. Although popular, R^2

Analysis of Covariance (ANCOVA) with Two Groups

should not be used indiscriminately or interpreted without scatter plot support. Following are some qualifications on its interpretation:

1. *Additional independent variables.* It is possible to increase R^2 by adding more independent variables, but the additional independent variables may cause an increase in the mean square error, an unfavorable situation. This usually happens when the sample size is small.
2. *Range of the independent variables.* R^2 is influenced by the range of the independent variables. R^2 increases as the range of the X 's increases and decreases as the range of the X 's decreases.
3. *Slope magnitudes.* R^2 does not measure the magnitude of the slopes.
4. *Linearity.* R^2 does not measure the appropriateness of a linear model. It measures the strength of the linear component of the model. Suppose the relationship between X and Y was a perfect sphere. Although there is a perfect relationship between the variables, the R^2 value would be zero.
5. *Predictability.* A large R^2 does not necessarily mean high predictability, nor does a low R^2 necessarily mean poor predictability.
6. *Sample size.* R^2 is highly sensitive to the number of observations. The smaller the sample size, the larger its value.

Adjusted R^2

This is an adjusted version of R^2 . The adjustment seeks to remove the distortion due to a small sample size. The formula for adjusted R^2 is

$$\bar{R}^2 = 1 - \frac{(N - 1)(1 - R^2)}{N - p - 1}$$

Coefficient of Variation

The coefficient of variation is a relative measure of dispersion, computed by dividing root mean square error by the mean of the response variable. By itself, it has little value, but it can be useful in comparative studies.

$$CV = \frac{\sqrt{MSE}}{\bar{y}}$$

Ave Abs Pct Error

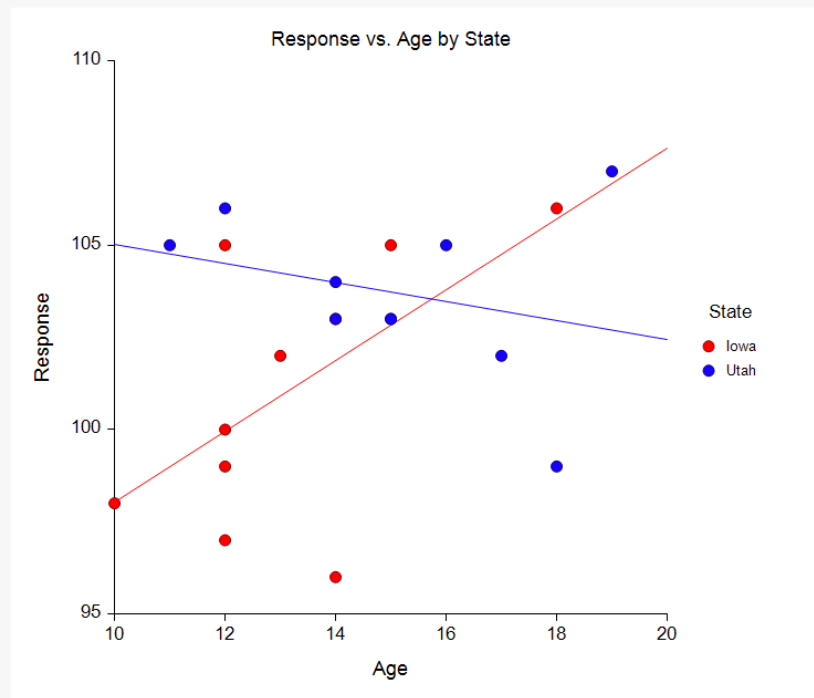
This is the average of the absolute percent errors. It is another measure of the goodness of fit of the linear model to the data. It is calculated using the formula

$$AAPE = \frac{100 \sum_{j=1}^N \left| \frac{y_j - \hat{y}_j}{y_j} \right|}{N}$$

Note that when the response variable is zero, its predicted value is used in the denominator.

Response vs Covariate by Group Scatter Plot

Response vs Covariate by Group Scatter Plot



This is a scatter plot with the response variable on the Y-axis, the covariate variable, Age, on the X-axis, and the group variable, State, in the legend. The slopes appear to be quite different, one positive and the other negative.

Descriptive Statistics

Descriptive Statistics

Variable	Count	Mean	Standard Deviation	Minimum	Maximum
Age	20	14.15	2.497894	10	19
(State="Utah")	20	0.5	0.5129892	0	1
Age*(State="Utah")	20	7.55	7.937088	0	19
Response	20	102.4	3.218368	96	107

For each variable, the count, arithmetic mean, standard deviation, minimum, and maximum are computed. This report is particularly useful for checking that the correct variables were selected. Recall that the group variable with two levels is represented by one binary indicator variable. The reference group is not listed.

Analysis of Variance

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F-Ratio	P-Value	Significant at 5%?
Model	3	77.67203	25.89068	3.477	0.0408	Yes
Age	1	12.10835	12.10835	1.626	0.2204	No
State	1	44.73146	44.73146	6.008	0.0261	Yes
Age*State	1	36.56446	36.56446	4.911	0.0415	Yes
Error	16	119.128	7.445498			
Total(Adjusted)	19	196.8	10.35789			

An analysis of variance (ANOVA) table summarizes the information related to the variation in data. The Age*State interaction is significant, indicating that the slopes are not equal. This being the case, you may want to perform comparisons at several covariate values.

Source

This represents a partition of the variation in Y .

DF

The degrees of freedom are the number of dimensions associated with this term. Note that each observation can be interpreted as a dimension in n -dimensional space. The degrees of freedom for the intercept, model, error, and adjusted total are 1, p , $n-p-1$, and $n-1$, respectively.

Sum of Squares

These are the sums of squares associated with the corresponding sources of variation. Note that these values are in terms of the dependent variable. The formulas for each are

$$SS_{Model} = \sum (\hat{y}_j - \bar{y})^2$$

$$SS_{Error} = \sum (y_j - \hat{y}_j)^2$$

$$SS_{Total} = \sum (y_j - \bar{y})^2$$

Mean Square

The mean square is the sum of squares divided by the degrees of freedom. This mean square is an estimated variance. For example, the mean square error is the estimated variance of the residuals.

F-Ratio

This is the F-statistic for testing the null hypothesis that all $\beta_j = 0$. This F-statistic has p degrees of freedom for the numerator variance and $n-p-1$ degrees of freedom for the denominator variance.

Analysis of Covariance (ANCOVA) with Two Groups

P-Value

This is the p-value for the above F-test. The p-value is the probability that the test statistic will take on a value at least as extreme as the observed value, if the null hypothesis is true. If the p-value is less than α , say 0.05, the null hypothesis is rejected. If the p-value is greater than α , then the null hypothesis is accepted.

Significant at [5%]?

This is the decision based on the p-value and the user-entered Tests Alpha value. The default is Tests Alpha = 0.05.

Model Coefficient T-Tests**Model Coefficient T-Tests**

Independent Variable	Model Coefficient b(i)	Standard Error Sb(i)	T-Statistic to Test H0: $\beta(i)=0$	P-Value	Reject H0 at 5%?
Intercept	88.44495	5.52261	16.015	0.0000	Yes
Age	0.9587156	0.4132412	2.320	0.0339	Yes
(State="Utah")	19.1561	7.815332	2.451	0.0261	Yes
Age*(State="Utah")	-1.217064	0.5492	-2.216	0.0415	Yes

This section reports the values and significance tests of the model coefficients.

Independent Variable

The names of the independent variables are listed here. The intercept is the value of the Y intercept.

Note that the name may become very long, especially for interaction terms. These long names may misalign the report. You can force the rest of the items to be printed on the next line by using the Stagger label ... option on the Report Options tab. This should create a better-looking report when the names are extra-long.

Model Coefficient b(i)

The coefficients are the least squares estimates of the parameters. The value indicates how much change in Y occurs for a one-unit change in a particular X when the remaining X's are held constant. These coefficients are often called partial-regression coefficients since the effect of the other X's is removed. These coefficients are the values of b_0, b_1, \dots, b_p .

Standard Error Sb(i)

The standard error of the coefficient, s_{b_j} , is the standard deviation of the estimate. It is used in hypothesis tests and confidence limits.

Analysis of Covariance (ANCOVA) with Two Groups

T-Statistic to Test H0: $\beta(i)=0$

This is the t-test value for testing the hypothesis that $\beta_j = 0$ versus the alternative that $\beta_j \neq 0$ after removing the influence of all other X 's. This t -value has $n-p-1$ degrees of freedom.

To test for a value other than zero, use the formula below. There is an easier way to test hypothesized values using confidence limits. See the discussion below under Confidence Limits. The formula for the t -test is

$$t_j = \frac{b_j - \beta_j^*}{s_{b_j}}$$

P-Value

This is the p -value for the significance test of the coefficient. The p -value is the probability that this t -statistic will take on a value at least as extreme as the observed value, assuming that the null hypothesis is true (i.e., the coefficient estimate is equal to zero). If the p -value is less than alpha, say 0.05, the null hypothesis of equality is rejected. This p -value is for a two-tail test.

Reject H0 at [5%]?

This is the decision based on the p -value and the user-entered Tests Alpha value. The default is Tests Alpha = 0.05.

Model Coefficient Confidence Intervals**Model Coefficient Confidence Intervals**

Independent Variable	Model Coefficient b(i)	Standard Error Sb(i)	Lower 95% Conf. Limit of $\beta(i)$	Upper 95% Conf. Limit of $\beta(i)$
Intercept	88.44495	5.52261	76.73754	100.1524
Age	0.9587156	0.4132412	0.08268333	1.834748
(State="Utah")	19.1561	7.815332	2.588337	35.72387
Age*(State="Utah")	-1.217064	0.5492	-2.381316	-0.05281156

Note: The T-Value used to calculate these confidence limits was 2.120.

This section reports the values and confidence intervals of the model coefficients.

Independent Variable

The names of the independent variables are listed here. The intercept is the value of the Y intercept.

Note that the name may become very long, especially for interaction terms. These long names may misalign the report. You can force the rest of the items to be printed on the next line by using the Stagger label ... option on the Report Options tab. This should create a better-looking report when the names are extra-long.

Analysis of Covariance (ANCOVA) with Two Groups

Model Coefficient

The coefficients are the least squares estimates of the parameters. The value indicates how much change in Y occurs for a one-unit change in a particular X when the remaining X 's are held constant. These coefficients are often called partial-regression coefficients since the effect of the other X 's is removed. These coefficients are the values of b_0, b_1, \dots, b_p .

Standard Error

The standard error of the coefficient, s_{b_j} , is the standard deviation of the estimate. It is used in hypothesis tests and confidence limits.

Lower and Upper 95% Conf. Limit of β_j

These are the lower and upper values of a $100(1 - \alpha)\%$ interval estimate for β_j based on a t -distribution with $n-p-1$ degrees of freedom. This interval estimate assumes that the residuals for the regression model are normally distributed.

These confidence limits may be used for significance testing values of β_j other than zero. If a specific value is not within this interval, it is significantly different from that value. Note that these confidence limits are set up as if you are interested in each regression coefficient separately.

The formulas for the lower and upper confidence limits are:

$$b_j \pm t_{1-\alpha/2, n-p-1} s_{b_j}$$

Note: The T-Value ...

This is the value of $t_{1-\alpha/2, n-p-1}$ used to construct the confidence limits.

Least Squares Means**Least Squares Means**

Error Degrees of Freedom (DF): 16
Means Calculated at: Covariate Values 1 to 3 (See below)

Name	Count	Least Squares Mean	Standard Error	Lower 95% Conf. Limit for Mean	Upper 95% Conf. Limit for Mean
Covariate Value 1: Age = 12					
Intercept					
All	20	102.2252	0.8649418	100.3916	104.0588
State					
Iowa	10	99.94954	0.9952167	97.83978	102.0593
Utah	10	104.5009	1.414935	101.5014	107.5004

Analysis of Covariance (ANCOVA) with Two Groups

Covariate Value 2: Age = 14.15 (Mean)**Intercept**

All	20	102.9781	0.6635722	101.5714	104.3848
-----	----	----------	-----------	----------	----------

State

Iowa	10	102.0108	0.9479811	100.0012	104.0204
------	----	----------	-----------	----------	----------

Utah	10	103.9454	0.9287863	101.9765	105.9144
------	----	----------	-----------	----------	----------

Covariate Value 3: Age = 18**Intercept**

All	20	104.3263	1.277113	101.619	107.0337
-----	----	----------	----------	---------	----------

State

Iowa	10	105.7018	2.163112	101.1162	110.2874
------	----	----------	----------	----------	----------

Utah	10	102.9508	1.358314	100.0713	105.8303
------	----	----------	----------	----------	----------

Note: When the model includes a significant covariate-by-group interaction, you may want to calculate and compare means at various values of the covariate and consider the results collectively. If you calculate and compare means at only one covariate value, the results may be misleading.

This section reports the least squares means and associated confidence intervals. In this example, the least squares means are calculated at Age = 12, 14.15 (the mean), and 18. The results are based on $n-p-1 = 16$ degrees of freedom for error.

Name

The name of the group variable and its individual group names are listed here. The intercept is the value of the Y intercept.

Note that the name may become very long, especially for interaction terms. These long names may misalign the report. You can force the rest of the items to be printed on the next line by using the Stagger label ... option on the Report Options tab. This should create a better-looking report when the names are extra-long.

Count

This column specifies the number of observations in each group.

Least Squares Mean

This is the least squares mean estimate, $\hat{\mu}_j$. The least squares means are adjusted based on the model. In balanced designs with no covariates, the least squares group means will be equal to the raw group means. In unbalanced designs or when covariates are present, the least squares means usually are different from the raw means.

Standard Error

The standard error of the mean, $SE(\hat{\mu}_j)$, is the standard deviation of the estimate. It is used in hypothesis tests and confidence limits.

Analysis of Covariance (ANCOVA) with Two Groups

Lower and Upper 95% Conf. Limits for Mean

These are the lower and upper values of a $100(1 - \alpha)\%$ interval estimate for the mean, μ_j , based on a t -distribution with $n-p-1$ degrees of freedom.

The formulas for the lower and upper confidence limits are:

$$\hat{\mu}_j \pm t_{1-\frac{\alpha}{2}, n-p-1} \times SE(\hat{\mu}_j)$$

Least Squares Means with Hypothesis Tests of H0: Mean = 0

Least Squares Means with Hypothesis Tests of H0: Mean = 0

Error Degrees of Freedom (DF): 16
 Means Calculated at: Covariate Values 1 to 3 (See below)
 Hypotheses Tested: H0: Mean = 0 vs. H1: Mean \neq 0

Name	Count	Least Squares Mean	Standard Error	T-Statistic to Test H0: Mean=0	P-Value	Reject H0 at 5%?
Covariate Value 1: Age = 12						
Intercept						
All	20	102.2252	0.8649418	118.187	0.0000	Yes
State						
Iowa	10	99.94954	0.9952167	100.430	0.0000	Yes
Utah	10	104.5009	1.414935	73.856	0.0000	Yes
Covariate Value 2: Age = 14.15 (Mean)						
Intercept						
All	20	102.9781	0.6635722	155.187	0.0000	Yes
State						
Iowa	10	102.0108	0.9479811	107.608	0.0000	Yes
Utah	10	103.9454	0.9287863	111.915	0.0000	Yes
Covariate Value 3: Age = 18						
Intercept						
All	20	104.3263	1.277113	81.689	0.0000	Yes
State						
Iowa	10	105.7018	2.163112	48.866	0.0000	Yes
Utah	10	102.9508	1.358314	75.793	0.0000	Yes

Note: When the model includes a significant covariate-by-group interaction, you may want to calculate and compare means at various values of the covariate and consider the results collectively. If you calculate and compare means at only one covariate value, the results may be misleading.

This section reports the least squares means and associated hypothesis tests. In this example, the least squares means are calculated at Age = 12, 14.15 (the mean), and 18. The results are based on $n-p-1 = 16$ degrees of freedom for error.

Analysis of Covariance (ANCOVA) with Two Groups

Name

The name of the group variable and its individual group names are listed here. The intercept is the value of the Y intercept.

Note that the name may become very long, especially for interaction terms. These long names may misalign the report. You can force the rest of the items to be printed on the next line by using the Stagger label ... option on the Report Options tab. This should create a better-looking report when the names are extra-long.

Count

This column specifies the number of observations in each group.

Least Squares Mean

This is the least squares mean estimate, $\hat{\mu}_j$. The least squares means are adjusted based on the model. In balanced designs with no covariates, the least squares group means will be equal to the raw group means. In unbalanced designs or when covariates are present, the least squares means usually are different from the raw means.

Standard Error

The standard error of the mean, $SE(\hat{\mu}_j)$, is the standard deviation of the estimate. It is used in hypothesis tests and confidence limits.

T-Statistic to Test H0: Mean=0

This is the t-test value for testing the hypothesis that the mean is equal to 0 versus the alternative that it is not equal to 0. This t -value has $n-p-1$ degrees of freedom and is calculated as

$$t_j = \frac{\hat{\mu}_j}{SE(\hat{\mu}_j)}$$

P-Value

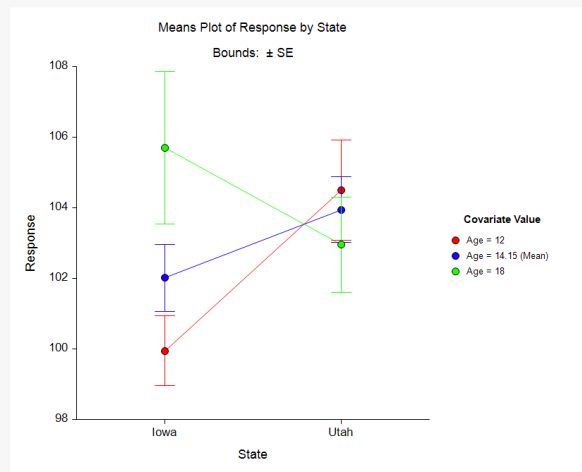
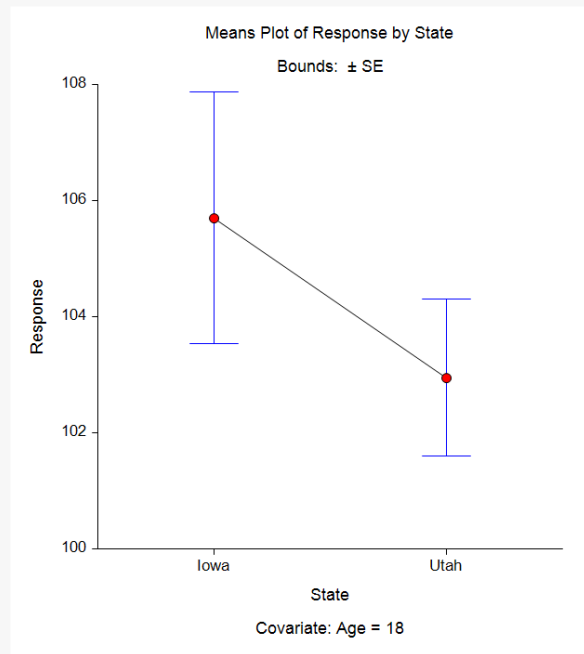
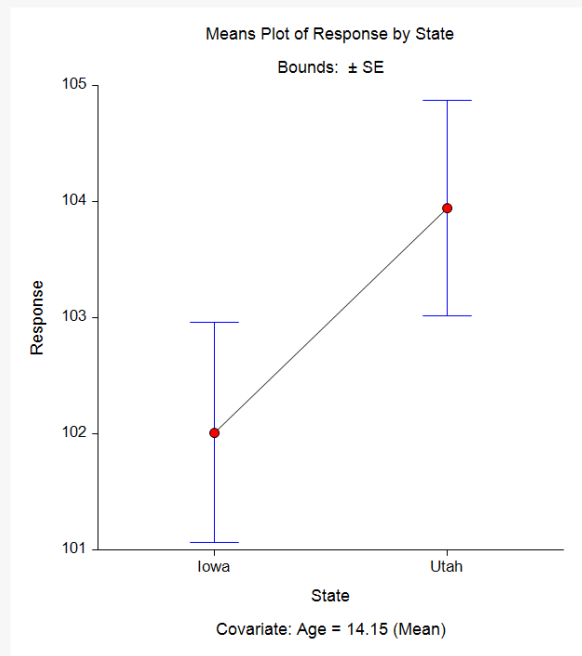
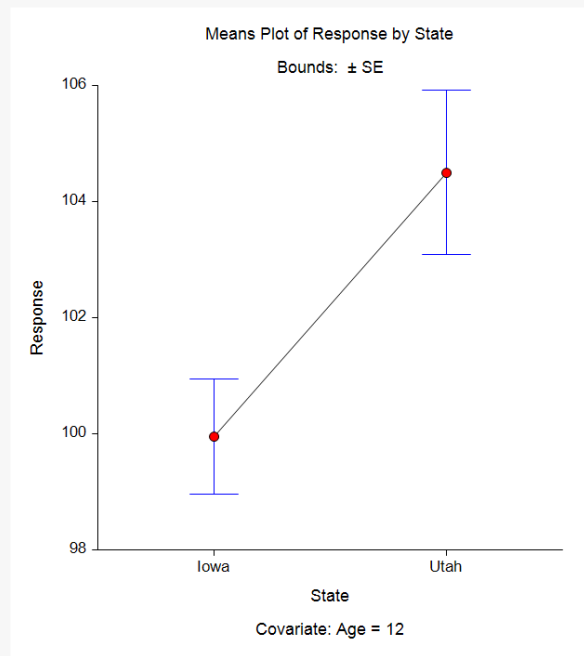
This is the p -value for the significance test of the mean. The p -value is the probability that this t -statistic will take on a value at least as extreme as the observed value, if the null hypothesis is true (i.e., the mean estimate is equal to zero). If the p -value is less than alpha, say 0.05, the null hypothesis of equality is rejected. This p -value is for a two-tail test.

Reject H0 at [5%]?

This is the decision based on the p -value and the user-entered Tests Alpha value. The default is Tests Alpha = 0.05.

Means Plots

Means Plots



The means plots displays the least squares means along with user-selected variability lines, in this case \pm SE. Individual plots and a combined plot are created for the various covariate values evaluated. Notice how the means change quite drastically depending on the covariate value used for the evaluation. This happens because the covariate-by-group interaction term is significant.

T-Tests for Group Least Squares Mean Differences

T-Tests for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 16
 Means Calculated at: Covariate Values 1 to 3 (See below)
 Hypotheses Tested: H0: Diff = 0 vs. H1: Diff \neq 0

Comparison	Least Squares Mean Difference	Standard Error	T-Statistic to Test H0: Diff=0	P-Value	Reject H0 at 5%?
Covariate Value 1: Age = 12					
State					
Utah - Iowa	4.551337	1.729884	2.631	0.0182	Yes
Covariate Value 2: Age = 14.15 (Mean)					
State					
Utah - Iowa	1.934651	1.327144	1.458	0.1643	No
Covariate Value 3: Age = 18					
State					
Utah - Iowa	-2.751044	2.554226	-1.077	0.2974	No

Note: When the model includes a significant covariate-by-group interaction, you may want to calculate and compare means at various values of the covariate and consider the results collectively. If you calculate and compare means at only one covariate value, the results may be misleading.

This section reports the least squares mean difference and associated hypothesis test for each covariate value. In this example, the least squares means are calculated at Age = 12, 14.15 (the mean), and 18. These tests indicate that the difference is only significant for Age = 12. The results are based on $n-p-1 = 16$ degrees of freedom for error.

Confidence Intervals for Group Least Squares Mean Differences

Confidence Intervals for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 16
 Means Calculated at: Covariate Values 1 to 3 (See below)

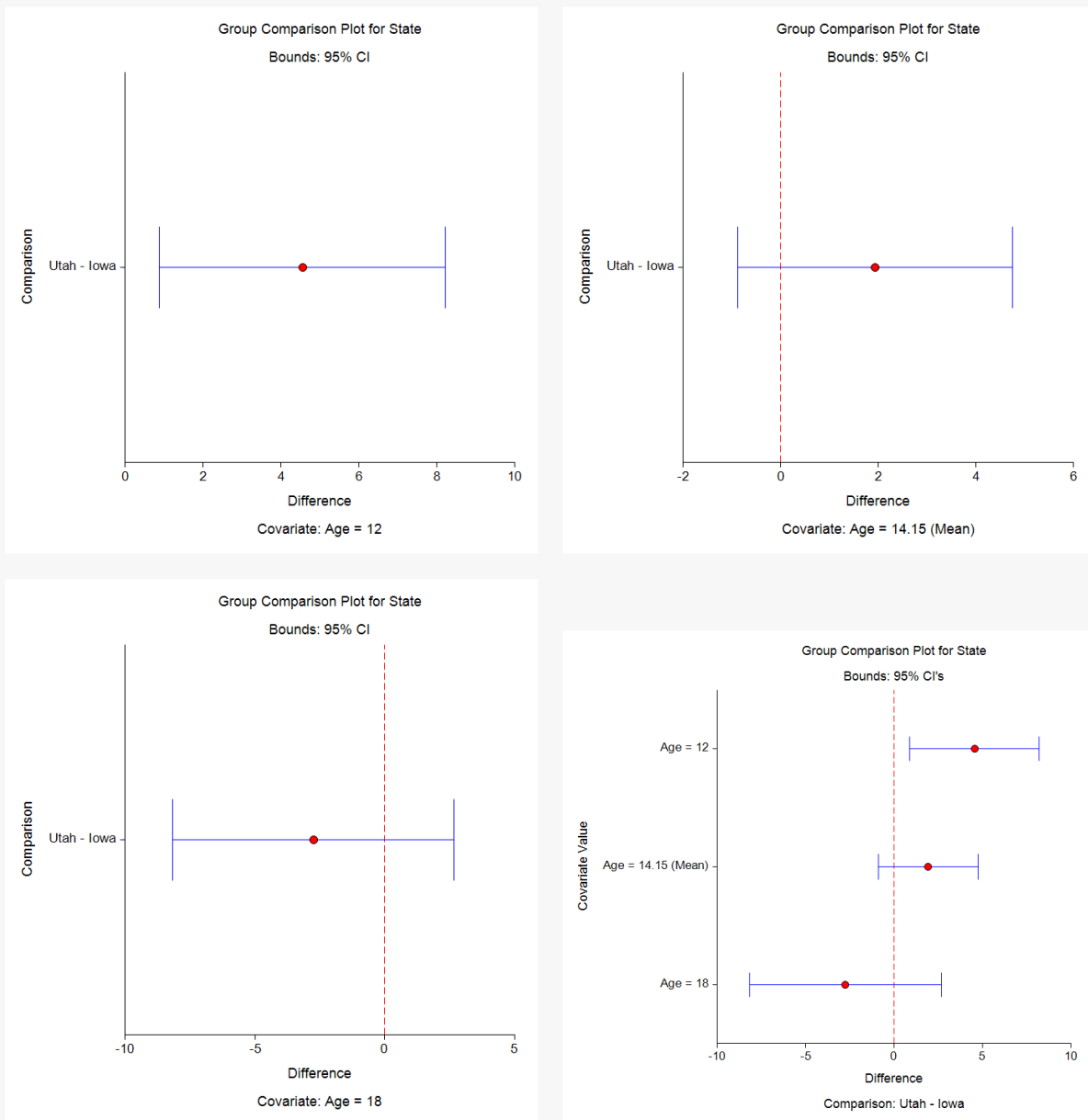
Comparison	Least Squares Mean Difference	Standard Error	Lower 95% Conf. Limit for Difference	Upper 95% Conf. Limit for Difference
Covariate Value 1: Age = 12				
State				
Utah - Iowa	4.551337	1.729884	0.8841482	8.218527
Covariate Value 2: Age = 14.15 (Mean)				
State				
Utah - Iowa	1.934651	1.327144	-0.8787697	4.748071
Covariate Value 3: Age = 18				
State				
Utah - Iowa	-2.751044	2.554226	-8.165761	2.663673

Note: When the model includes a significant covariate-by-group interaction, you may want to calculate and compare means at various values of the covariate and consider the results collectively. If you calculate and compare means at only one covariate value, the results may be misleading.

This section reports the least squares mean difference and associated confidence interval for each covariate value. In this example, the least squares means are calculated at Age = 12, 14.15 (the mean), and 18. These intervals indicate that the difference is only significant for Age = 12. The results are based on $n-p-1 = 16$ degrees of freedom for error.

Group Comparison Plots

Group Comparison Plots



These comparison plots display the mean differences along with 95% confidence intervals. Comparisons for which the interval does not contain zero are significant (i.e., for Age = 12).

Residual Normality Assumption Tests

Residual Normality Assumption Tests

Test Name	Test Statistic	P-Value	Reject Residual Normality at 20%?
Shapiro-Wilk	0.956	0.4700	No
Anderson-Darling	0.486	0.2264	No
D'Agostino Skewness	-0.499	0.6180	No
D'Agostino Kurtosis	1.205	0.2282	No
D'Agostino Omnibus (Skewness and Kurtosis)	1.701	0.4272	No

This report gives the results of applying several normality tests to the residuals. The Shapiro-Wilk test is probably the most popular, so it is given first. These tests are discussed in detail in the Normality Tests section of the Descriptive Statistics procedure.

Graphic Residual Analysis

The residuals can be graphically analyzed in numerous ways. You should examine all of the basic residual graphs: the histogram, the density trace, the normal probability plot, the scatter plot of the residuals versus the predicted value of the dependent variable, and the scatter plot of the residuals versus each of the independent variables.

For the basic scatter plots of residuals versus either the predicted values of Y or the independent variables, Hoaglin (1983) explains that there are several patterns to look for. You should note that these patterns are very difficult, if not impossible, to recognize for small data sets.

Point Cloud

A point cloud, basically in the shape of a rectangle or a horizontal band, would indicate no relationship between the residuals and the variable plotted against them. This is the preferred condition.

Wedge

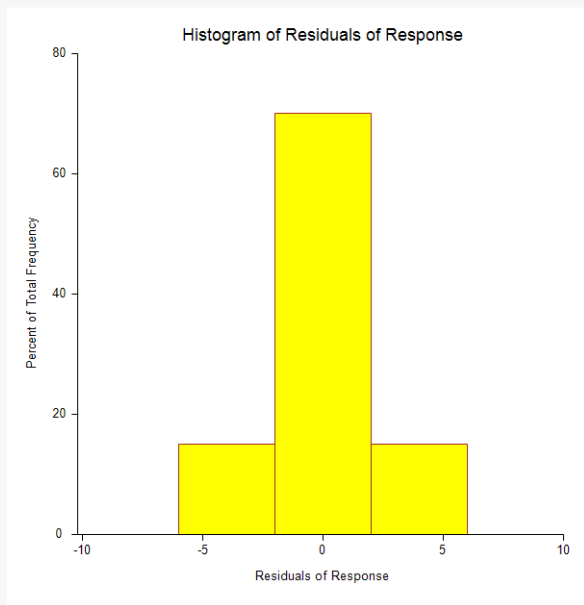
An increasing or decreasing wedge would be evidence that there is increasing or decreasing (non-constant) variation. A transformation of Y may correct the problem.

Bowtie

This is similar to the wedge above in that the residual plot shows a decreasing wedge in one direction while simultaneously having an increasing wedge in the other direction. A transformation of Y may correct the problem.

Histogram of Residuals

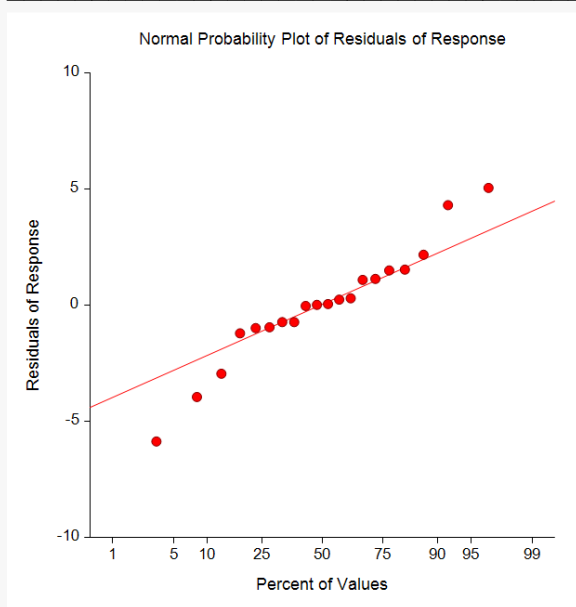
Residual Analysis Plots



The purpose of the histogram and density trace of the residuals is to evaluate whether they are normally distributed. Unless you have a large sample size, it is best not to rely on the histogram for visually evaluating normality of the residuals. The better choice would be the normal probability plot.

Probability Plot of Residuals

Residual Analysis Plots

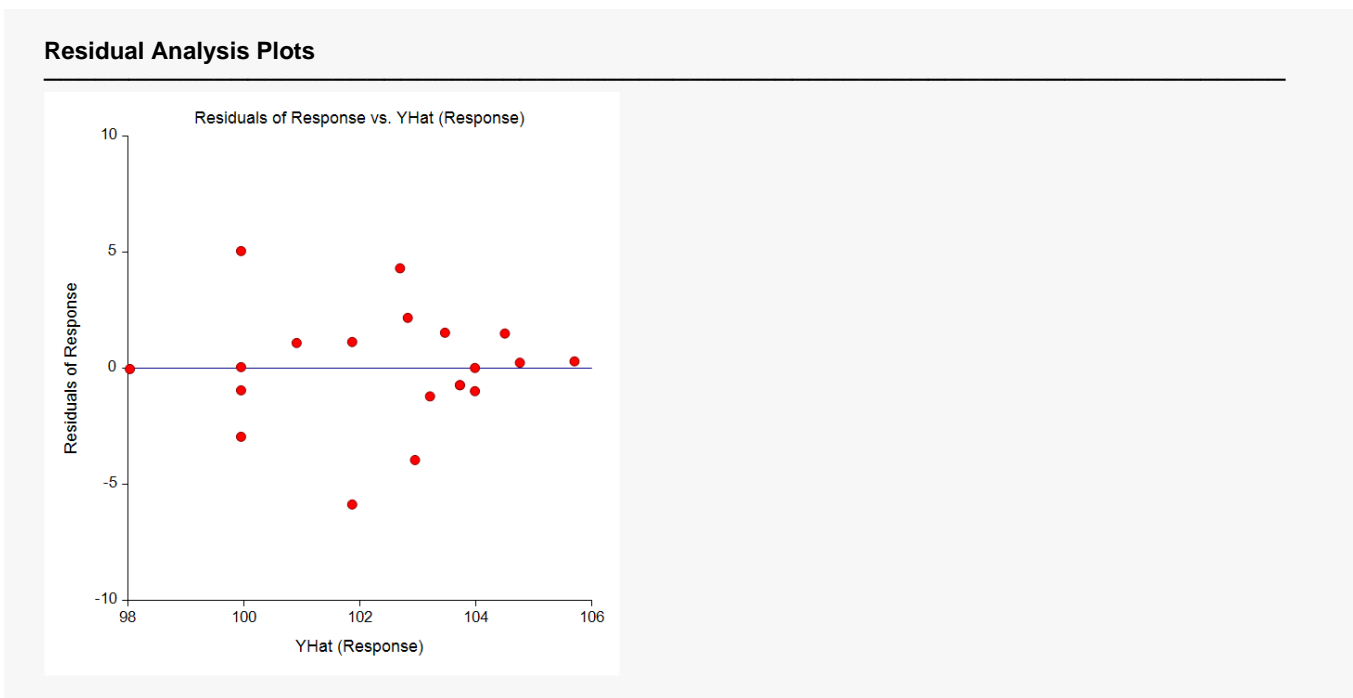


Analysis of Covariance (ANCOVA) with Two Groups

If the residuals are normally distributed, the data points of the normal probability plot will fall along a straight line through the origin with a slope of 1.0. Major deviations from this ideal picture reflect departures from normality. Stragglers at either end of the normal probability plot indicate outliers, curvature at both ends of the plot indicates long or short distributional tails, convex or concave curvature indicates a lack of symmetry, and gaps or plateaus or segmentation in the normal probability plot may require a closer examination of the data or model. Of course, use of this graphic tool with very small sample sizes is not recommended.

If the residuals are not normally distributed, then the t-tests on regression coefficients, the F-tests, and any interval estimates are not valid. This is a critical assumption to check.

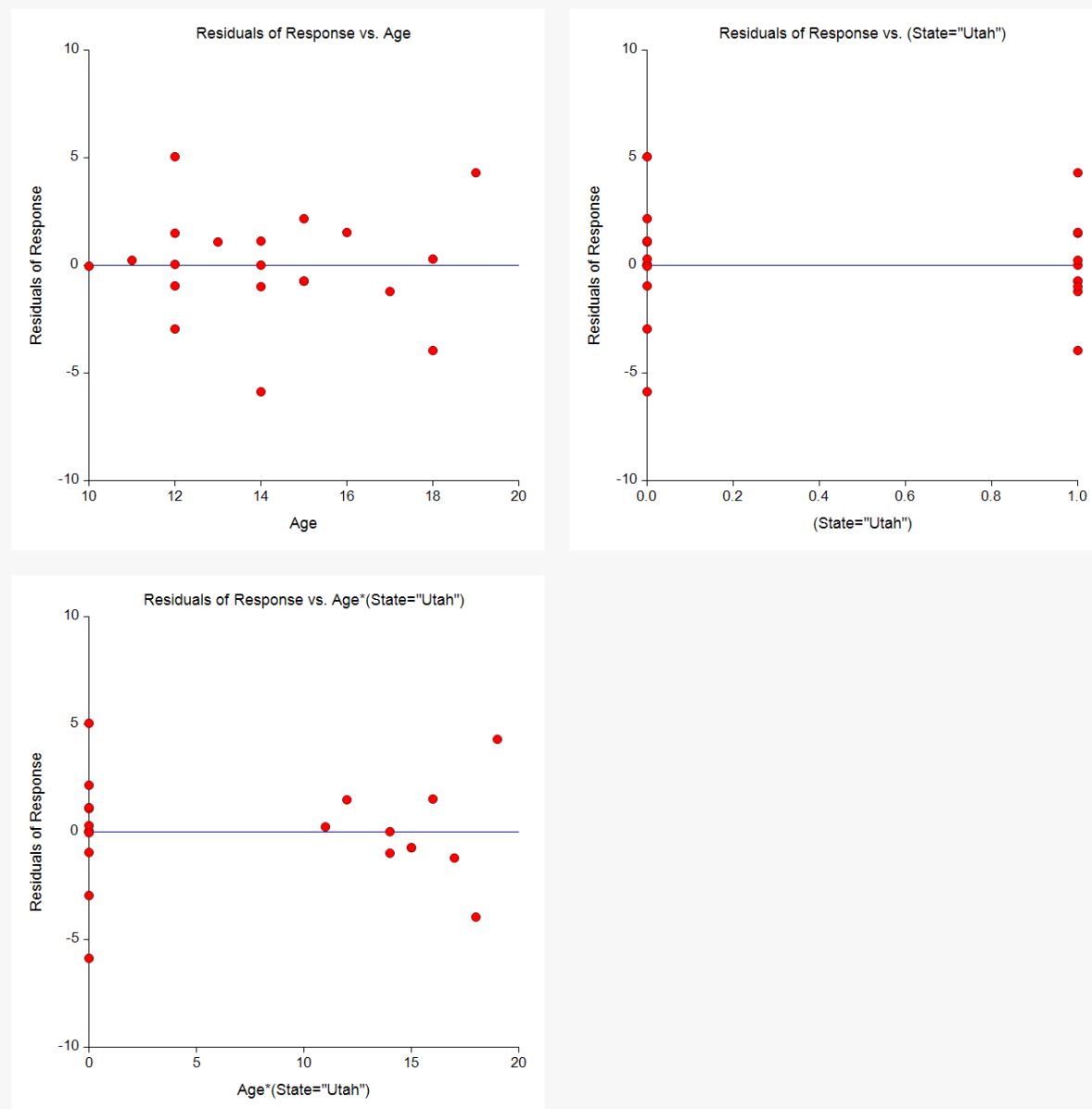
Residuals vs Yhat (Predicted) Plot



This plot should always be examined. The preferred pattern to look for is a point cloud or a horizontal band. A wedge or bowtie pattern is an indicator of non-constant variance, a violation of a critical assumption. The sloping or curved band signifies inadequate specification of the model. The sloping band with increasing or decreasing variability suggests non-constant variance and inadequate specification of the model.

Residuals vs X Plots

Residual Analysis Plots



These are scatter plots of the residuals versus each independent variable. Again, the preferred pattern is a rectangular shape or point cloud. Any other nonrandom pattern may require a redefining of the model.

Residuals List Report

Residuals List

Row	Actual Response	Predicted Response	Residual	Absolute Percent Error	Sqrt(MSE) Without This Row
1	100	99.94954	0.05045871	0.050	2.818097
2	102	100.9083	1.091743	1.070	2.802407
3	97	99.94954	-2.949541	3.041	2.696829
4	96	101.867	-5.866972	6.111	2.312976
5	105	102.8257	2.174312	2.071	2.749573
6	106	105.7018	0.2981651	0.281	2.8153
7	105	99.94954	5.050459	4.810	2.4455
8	103	101.867	1.133028	1.100	2.800928
9	99	99.94954	-0.9495413	0.959	2.805804
10	98	98.03211	-0.03211009	0.033	2.818113
11	104	103.9842	0.01581722	0.015	2.818128
12	105	104.7592	0.2407733	0.229	2.816997
13	106	104.5009	1.499121	1.414	2.781535
14	103	103.7258	-0.7258348	0.705	2.811198
15	102	103.2091	-1.209139	1.185	2.797383
16	99	102.9508	-3.950791	3.991	2.560951
17	107	102.6924	4.307557	4.026	2.446776
18	105	103.4675	1.532513	1.460	2.786593
19	103	103.7258	-0.7258348	0.705	2.811198
20	103	103.9842	-0.9841828	0.956	2.805063

This section reports on the sample residuals, or e_i 's.

Actual

This is the actual value of Y .

Predicted

The predicted value of Y using the values of the independent variables given on this row.

Residual

This is the error in the predicted value. It is equal to the *Actual* minus the *Predicted*.

Absolute Percent Error

This is percentage that the absolute value of the *Residual* is of the *Actual* value. Scrutinize rows with the large percent errors.

Sqrt(MSE) Without This Row

This is the value of the square root of the mean square error that is obtained if this row is deleted. A perusal of this statistic for all observations will highlight observations that have an inflationary impact on mean square error and could be outliers.

Predicted Values with Confidence Limits of Means

Predicted Values with Confidence Limits of Means

Row	Actual Response	Predicted Response	Standard Error of Predicted	Lower 95% Conf. Limit of Mean	Upper 95% Conf. Limit of Mean
1	100	99.94954	0.9952167	97.83978	102.0593
2	102	100.9083	0.8668221	99.07068	102.7458
3	97	99.94954	0.9952167	97.83978	102.0593
4	96	101.867	0.9240355	99.9081	103.8258
5	105	102.8257	1.139227	100.4106	105.2407
6	106	105.7018	2.163112	101.1162	110.2874
7	105	99.94954	0.9952167	97.83978	102.0593
8	103	101.867	0.9240355	99.9081	103.8258
9	99	99.94954	0.9952167	97.83978	102.0593
10	98	98.03211	1.578992	94.6848	101.3794
11	104	103.9842	0.9502006	101.9698	105.9985
12	105	104.7592	1.715861	101.1218	108.3967
13	106	104.5009	1.414935	101.5014	107.5004
14	103	103.7258	0.8636309	101.895	105.5566
15	102	103.2091	1.103144	100.8706	105.5477
16	99	102.9508	1.358314	100.0713	105.8303
17	107	102.6924	1.653727	99.1867	106.1982
18	105	103.4675	0.9222474	101.5124	105.4226
19	103	103.7258	0.8636309	101.895	105.5566
20	103	103.9842	0.9502006	101.9698	105.9985

Confidence intervals for the mean response of Y given specific levels for the group and covariate variables are provided here. It is important to note that violations of any assumptions will invalidate these interval estimates.

Actual

This is the actual value of Y .

Predicted

The predicted value of Y . It is predicted using the values of the group and covariate variables for this row. If the input data had both group and covariate values but no value for Y , the predicted value is still provided.

Standard Error of Predicted

This is the standard error of the mean response for the specified values of the group and covariate variables. Note that this value is not constant for all variable values. In fact, it is a minimum at the average value of each group and covariate variable.

Lower 95% C.L. of Mean

This is the lower limit of a 95% confidence interval estimate of the mean of Y for this observation.

Upper 95% C.L. of Mean

This is the upper limit of a 95% confidence interval estimate of the mean of Y for this observation.

Predicted Values with Prediction Limits of Individuals

Predicted Values with Prediction Limits of Individuals

Row	Actual Response	Predicted Response	Standard Error of Predicted	Lower 95% Pred. Limit of Individual	Upper 95% Pred. Limit of Individual
1	100	99.94954	2.904471	93.79234	106.1067
2	102	100.9083	2.863019	94.83893	106.9776
3	97	99.94954	2.904471	93.79234	106.1067
4	96	101.867	2.880857	95.75983	107.9741
5	105	102.8257	2.956913	96.55731	109.0941
6	106	105.7018	3.482033	98.32026	113.0834
7	105	99.94954	2.904471	93.79234	106.1067
8	103	101.867	2.880857	95.75983	107.9741
9	99	99.94954	2.904471	93.79234	106.1067
10	98	98.03211	3.152573	91.34895	104.7153
11	104	103.9842	2.889356	97.85902	110.1093
12	105	104.7592	3.223302	97.92613	111.5923
13	106	104.5009	3.073685	97.98495	111.0168
14	103	103.7258	2.862055	97.65855	109.7931
15	102	103.2091	2.9432	96.96983	109.4484
16	99	102.9508	3.048035	96.48924	109.4123
17	107	102.6924	3.19066	95.92854	109.4563
18	105	103.4675	2.880285	97.36156	109.5734
19	103	103.7258	2.862055	97.65855	109.7931
20	103	103.9842	2.889356	97.85902	110.1093

A prediction interval for the individual response of Y given specific values of the group and covariate variables is provided here for each row.

Actual

This is the actual value of Y .

Predicted

The predicted value of Y . It is predicted using the values of the group and covariate variables for this row. If the input data had both group and covariate values but no value for Y , the predicted value is still provided.

Standard Error of Predicted

This is the standard error of the mean response for the specified values of the group and covariate variables. Note that this value is not constant for all variable values. In fact, it is a minimum at the average value of the group and covariate variable.

Lower 95% Pred. Limit of Individual

This is the lower limit of a 95% prediction interval of the individual value of Y for this observation.

Upper 95% Pred. Limit of Individual

This is the upper limit of a 95% prediction interval of the individual value of Y for this observation.

Example 2 – ANCOVA Model Assuming Equal Slopes (No Covariate-by-Group Interaction)

In this example, the responses from two states, Iowa and Utah, are compared with an adjustment for the age of the respondent. This section presents an example of how to run an analysis on the same data assuming equal slopes. (Note: You would not normally analyze data assuming equal slopes if the interaction is found to be significant. This is done here for demonstration purposes only.)

Setup

To run this example, complete the following steps:

1 Open the ANCOVA2Grp example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **ANCOVA2Grp** and click **OK**.

2 Specify the Analysis of Covariance (ANCOVA) with Two Groups procedure options

- Find and open the **Analysis of Covariance (ANCOVA) with Two Groups** procedure using the menus or the Procedure Navigator.
- The settings 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.

Variables Tab

Response Variable(s)	Response
Group Variable.....	State
Covariate Variable	Age
Assume Slopes are.....	Equal (No Covariate-by-Group Interaction)

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

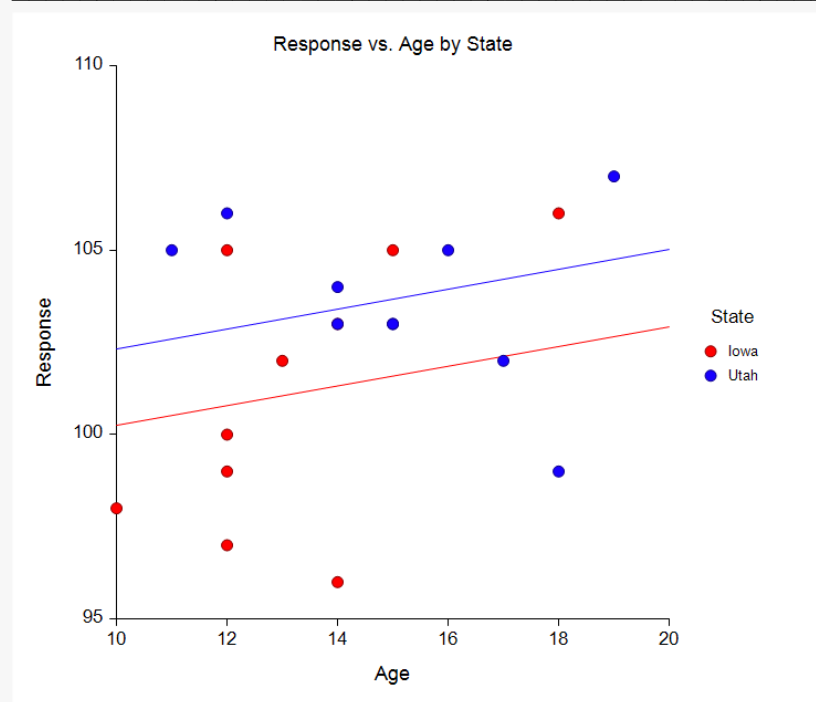
Output

Run Summary

Response Variable: Response
 Group Variable: State
 Reference Group: "Iowa"
 Covariate Variable: Age
 Slopes Assumed to be: Equal
 Model: Age + State

Parameter	Value	Rows	Value
R ²	0.2089	Rows Processed	20
Adj R ²	0.1158	Rows Filtered Out	0
Coefficient of Variation	0.0296	Rows with Response Missing	0
Mean Square Error	9.158379	Rows with Group or Covariate Missing	0
Square Root of MSE	3.026281	Rows Used in Estimation	20
Ave Abs Pct Error	2.293	Completion Status	Normal Completion
Error Degrees of Freedom	17		

Response vs Covariate by Group Scatter Plot



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F-Ratio	P-Value	Significant at 5%?
Model	2	41.10756	20.55378	2.244	0.1365	No
Age	1	7.307562	7.307562	0.798	0.3842	No
State	1	18.47374	18.47374	2.017	0.1736	No
Error	17	155.6924	9.158379			
Total(Adjusted)	19	196.8	10.35789			

Analysis of Covariance (ANCOVA) with Two Groups

Model Coefficient T-Tests

Independent Variable	Model Coefficient b(i)	Standard Error Sb(i)	T-Statistic to Test H0: $\beta(i)=0$	P-Value	Reject H0 at 5%?
Intercept	97.5406	4.098049	23.802	0.0000	Yes
Age	0.2696517	0.3018744	0.893	0.3842	No
(State="Utah")	2.087662	1.469914	1.420	0.1736	No

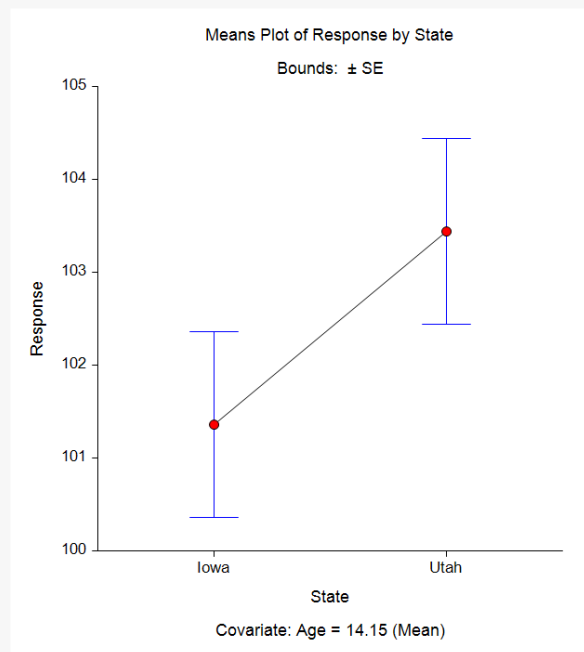
Least Squares Means

Error Degrees of Freedom (DF): 17
 Means Calculated at: Age = 14.15 (Mean)

Name	Count	Least Squares Mean	Standard Error	Lower 95% Conf. Limit for Mean	Upper 95% Conf. Limit for Mean
Intercept					
All	20	102.4	0.6766971	100.9723	103.8277
State					
Iowa	10	101.3562	0.9990401	99.24838	103.464
Utah	10	103.4438	0.9990401	101.336	105.5516

Note: These results assume that the slopes for both groups are equal (i.e., the covariate-by-group interaction is not significant). To check this assumption, run the model with unequal slopes that includes the covariate-by-group interaction and review the test of the interaction term in the Analysis of Variance report.

Means Plots



Analysis of Covariance (ANCOVA) with Two Groups

T-Tests for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 17
 Means Calculated at: Age = 14.15 (Mean)
 Hypotheses Tested: H0: Diff = 0 vs. H1: Diff ≠ 0

Comparison	Least Squares Mean Difference	Standard Error	T-Statistic to Test H0: Diff=0	P-Value	Reject H0 at 5%?
State Utah - Iowa	2.087662	1.469914	1.420	0.1736	No

Note: These results assume that the slopes for both groups are equal (i.e., the covariate-by-group interaction is not significant). To check this assumption, run the model with unequal slopes that includes the covariate-by-group interaction and review the test of the interaction term in the Analysis of Variance report.

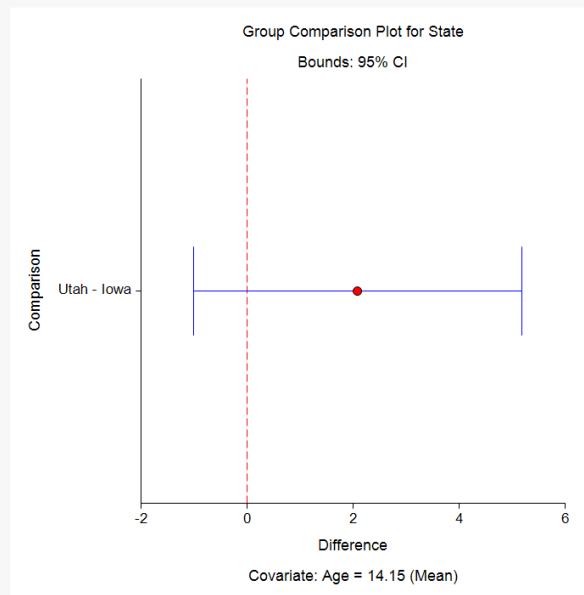
Confidence Intervals for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 17
 Means Calculated at: Age = 14.15 (Mean)

Comparison	Least Squares Mean Difference	Standard Error	Lower 95% Conf. Limit for Difference	Upper 95% Conf. Limit for Difference
State Utah - Iowa	2.087662	1.469914	-1.013587	5.18891

Note: These results assume that the slopes for both groups are equal (i.e., the covariate-by-group interaction is not significant). To check this assumption, run the model with unequal slopes that includes the covariate-by-group interaction and review the test of the interaction term in the Analysis of Variance report.

Group Comparison Plots

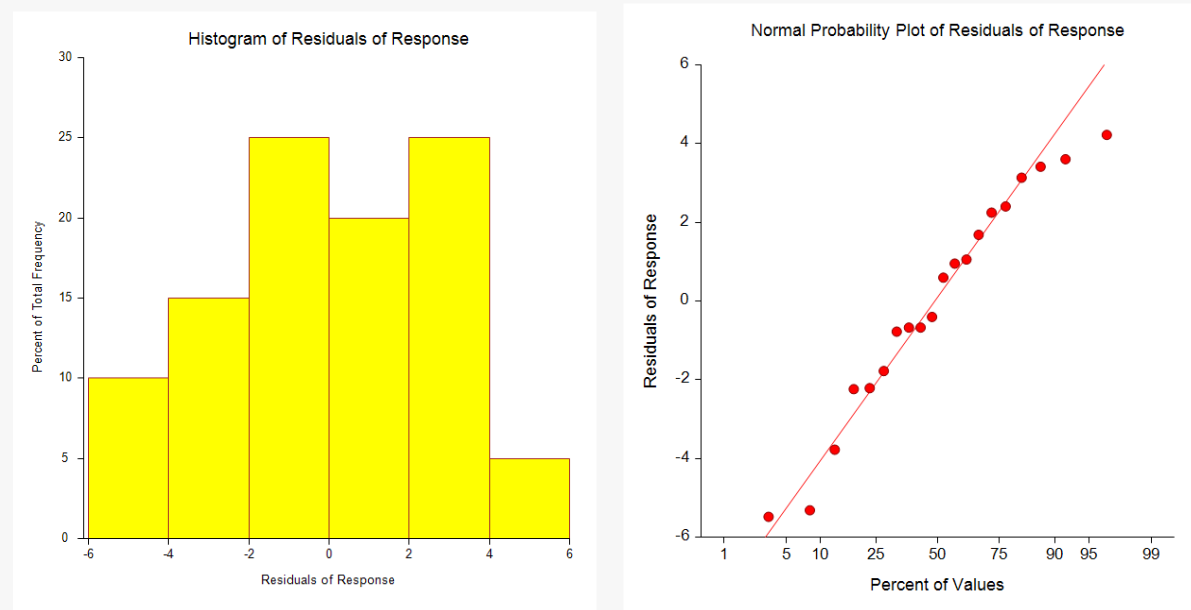


Analysis of Covariance (ANCOVA) with Two Groups

Residual Normality Assumption Tests

Test Name	Test Statistic	P-Value	Reject Residual Normality at 20%?
Shapiro-Wilk	0.955	0.4523	No
Anderson-Darling	0.277	0.6560	No
D'Agostino Skewness	-0.878	0.3797	No
D'Agostino Kurtosis	-0.509	0.6104	No
D'Agostino Omnibus (Skewness and Kurtosis)	1.031	0.5971	No

Residual Analysis Plots



The scatter plot shows the two regression lines with equal slopes. The lines don't appear to fit the data very well as is expected when the interaction is significant. Neither State nor Age is significant in these results.

Example 3 – ANCOVA Model with One-Sided Multiple Comparison Tests and Confidence Intervals

In this example, we'll run the same analysis as in Example 1, except that we will perform one-sided tests and create one-sided confidence intervals for the comparisons. Only reports that are different from Example 1 will be shown.

Setup

To run this example, complete the following steps:

1 Open the ANCOVA2Grp example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **ANCOVA2Grp** and click **OK**.

2 Specify the Analysis of Covariance (ANCOVA) with Two Groups procedure options

- Find and open the **Analysis of Covariance (ANCOVA) with Two Groups** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variable(s) **Response**
 Group Variable..... **State**
 Covariate Variable **Age**
 Calc Group Means at..... **12 Mean 18**

Reports Tab

Hypothesis Test and Confidence Interval **One-Sided Upper**
 Direction
 T-Tests for Group Least Squares Mean **Checked**
 Differences
 Show Confidence Intervals **Checked**

Plots Tab

Group Comparison Plots **Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Output

T-Tests for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 16
 Means Calculated at: Covariate Values 1 to 3 (See below)
 Hypotheses Tested: H0: Diff ≤ 0 vs. H1: Diff > 0

Comparison	Least Squares Mean Difference	Standard Error	T-Statistic to Test H0: Diff ≤ 0	P-Value	Reject H0 at 5%?
------------	-------------------------------	----------------	----------------------------------	---------	------------------

Covariate Value 1: Age = 12

State	Least Squares Mean Difference	Standard Error	T-Statistic to Test H0: Diff ≤ 0	P-Value	Reject H0 at 5%?
Utah - Iowa	4.551337	1.729884	2.631	0.0091	Yes

Covariate Value 2: Age = 14.15 (Mean)

State	Least Squares Mean Difference	Standard Error	T-Statistic to Test H0: Diff ≤ 0	P-Value	Reject H0 at 5%?
Utah - Iowa	1.934651	1.327144	1.458	0.0821	No

Covariate Value 3: Age = 18

State	Least Squares Mean Difference	Standard Error	T-Statistic to Test H0: Diff ≤ 0	P-Value	Reject H0 at 5%?
Utah - Iowa	-2.751044	2.554226	-1.077	0.8513	No

Note: When the model includes a significant covariate-by-group interaction, you may want to calculate and compare means at various values of the covariate and consider the results collectively. If you calculate and compare means at only one covariate value, the results may be misleading.

Confidence Intervals for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 16
 Means Calculated at: Covariate Values 1 to 3 (See below)

Comparison	Least Squares Mean Difference	Standard Error	Lower 95% Conf. Limit for Difference	Upper 95% Conf. Limit for Difference
------------	-------------------------------	----------------	--------------------------------------	--------------------------------------

Covariate Value 1: Age = 12

State	Least Squares Mean Difference	Standard Error	Lower 95% Conf. Limit for Difference	Upper 95% Conf. Limit for Difference
Utah - Iowa	4.551337	1.729884	1.531162	Infinity

Covariate Value 2: Age = 14.15 (Mean)

State	Least Squares Mean Difference	Standard Error	Lower 95% Conf. Limit for Difference	Upper 95% Conf. Limit for Difference
Utah - Iowa	1.934651	1.327144	-0.382389	Infinity

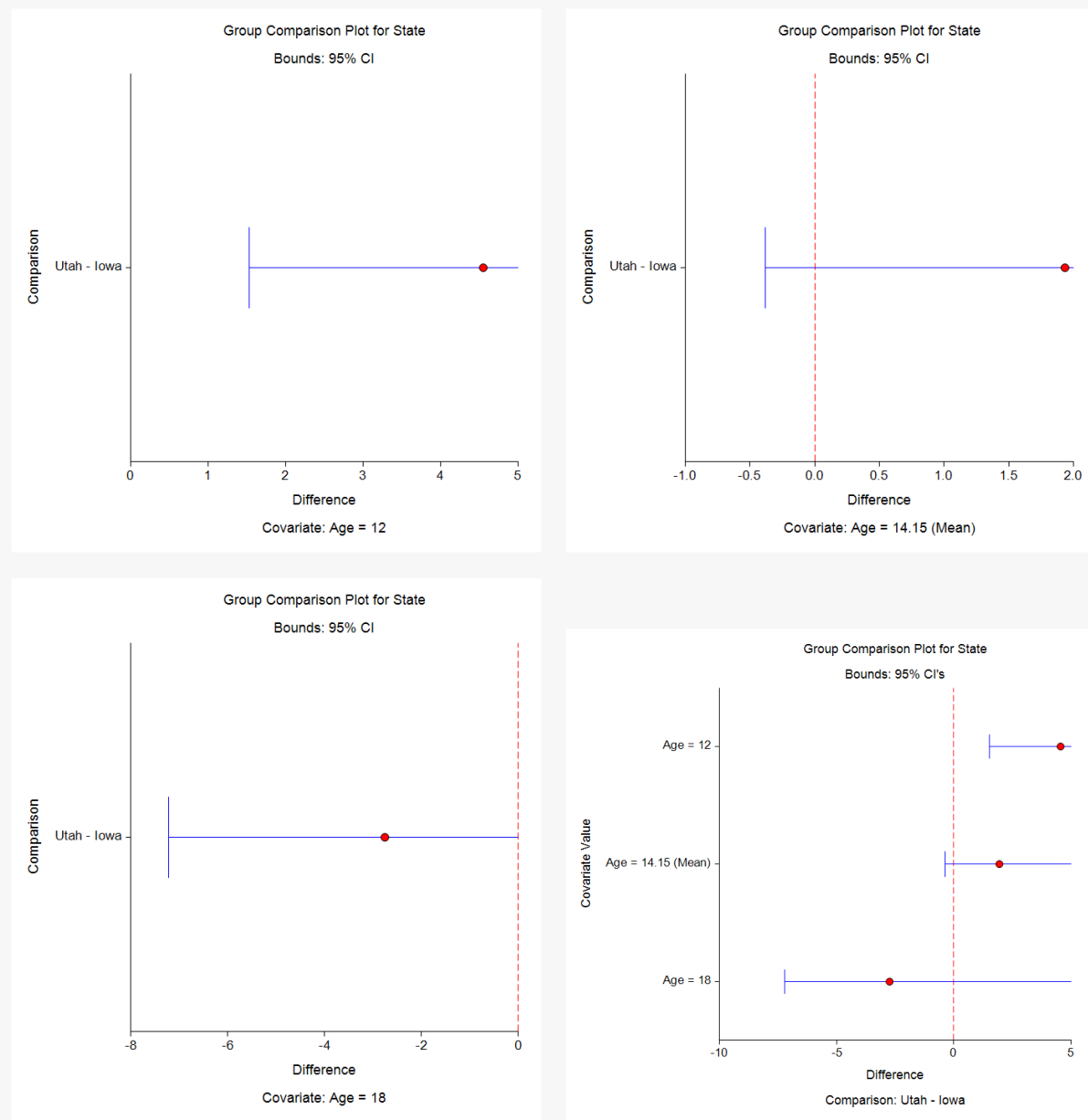
Covariate Value 3: Age = 18

State	Least Squares Mean Difference	Standard Error	Lower 95% Conf. Limit for Difference	Upper 95% Conf. Limit for Difference
Utah - Iowa	-2.751044	2.554226	-7.210425	Infinity

Analysis of Covariance (ANCOVA) with Two Groups

Note: When the model includes a significant covariate-by-group interaction, you may want to calculate and compare means at various values of the covariate and consider the results collectively. If you calculate and compare means at only one covariate value, the results may be misleading.

Group Comparison Plots



These results indicate that for one-sided tests and confidence intervals, only the comparison with Age = 12 is significant.

Example 4 – Two-Sample Equal-Variance T-Test

If you run an ANCOVA analysis with two groups and find the covariate to be non-significant, you may want to remove the covariate from the analysis and run a simple two-sample equal-variance T-test. This example will show you how to perform a two-sample T-test using this procedure.

Note: The two-sample T-test options are limited in this procedure. For additional options specifically related to the two-sample T-test scenario, we suggest you use the Two-Sample T-Test procedure in **NCSS** instead.

Setup

To run this example, complete the following steps:

1 Open the ANCOVA2Grp example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **ANCOVA2Grp** and click **OK**.

2 Specify the Analysis of Covariance (ANCOVA) with Two Groups procedure options

- Find and open the **Analysis of Covariance (ANCOVA) with Two Groups** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variable(s) **Response**

Group Variable..... **State**

Reports Tab

Run Summary..... **Checked**

T-Tests for Group Least Squares Mean **Checked**

Differences

Show Confidence Intervals **Checked**

Plots Tab

Group Comparison Plots **Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Output

Run Summary

Response Variable: Response
 Group Variable: State
 Reference Group: "Iowa"
 Covariate Variable: [None]
 Slopes Assumed to be: Unequal
 Model: State

Parameter	Value	Rows	Value
R ²	0.1717	Rows Processed	20
Adj R ²	0.1257	Rows Filtered Out	0
Coefficient of Variation	0.0294	Rows with Response Missing	0
Mean Square Error	9.055555	Rows with Group or Covariate Missing	0
Square Root of MSE	3.009245	Rows Used in Estimation	20
Ave Abs Pct Error	2.359	Completion Status	Normal Completion
Error Degrees of Freedom	18		

T-Tests for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 18
 Hypotheses Tested: H0: Diff = 0 vs. H1: Diff ≠ 0

Comparison	Least Squares Mean Difference	Standard Error	T-Statistic to Test H0: Diff=0	P-Value	Reject H0 at 5%?
State					
Utah - Iowa	2.6	1.345775	1.932	0.0693	No

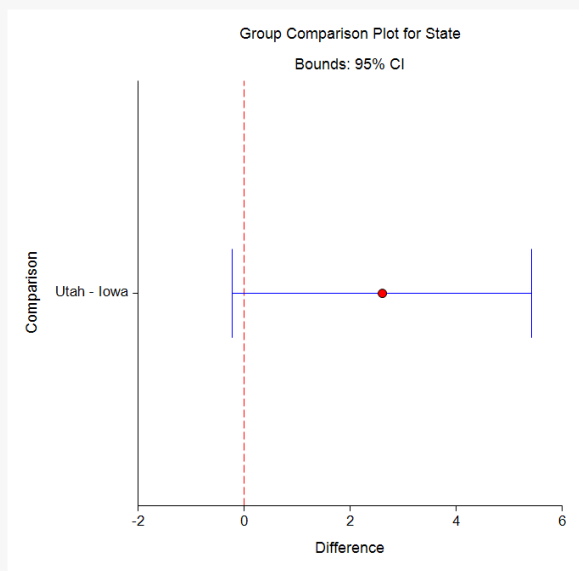
Confidence Intervals for Group Least Squares Mean Differences

Error Degrees of Freedom (DF): 18

Comparison	Least Squares Mean Difference	Standard Error	Lower 95% Conf. Limit for Difference	Upper 95% Conf. Limit for Difference
State				
Utah - Iowa	2.6	1.345775	-0.227369	5.427369

Analysis of Covariance (ANCOVA) with Two Groups

Group Comparison Plots



These results are equivalent to those that would be obtained for the equal-variance T-test in the Two-Sample T-Test procedure in **NCSS**.