

One-way Anova

Go to the course web page, then the *Notes and Handouts* section.

One-way Anova

Go to the course web page, then the *Notes and Handouts* section.

Right click on the *onewayanova_download.r* link and select *copy link location*

One-way Anova

Go to the course web page, then the *Notes and Handouts* section.

Right click on the *onewayanova_download.r* link and select *copy link location*

This should copy the URL for the Rdata file for this exercise. Use `setwd()` to switch to a directory you can write to.

One-way Anova

Go to the course web page, then the *Notes and Handouts* section.

Right click on the *onewayanova_download.r* link and select *copy link location*

This should copy the URL for the Rdata file for this exercise. Use `setwd()` to switch to a directory you can write to.

Paste the URL (now in the clipboard) between the quotes in the command `source(" ")`

One-way Anova

If all goes well, R should download three data frames, `onewayanova1-3`, to your workspace. Use `ls()` to verify that they are there.

One-way Anova

If all goes well, R should download three data frames, `onewayanova1-3`, to your workspace. Use `ls()` to verify that they are there.

Enter `plot(y~x)` to graph the data.

One-way Anova

If all goes well, R should download three data frames, `onewayanova1-3`, to your workspace. Use `ls()` to verify that they are there.

Enter `plot(y~x)` to graph the data.

This should show three groups of measurements, indexed by the variable `x`

One-way Anova

If all goes well, R should download three data frames, `onewayanova1-3`, to your workspace. Use `ls()` to verify that they are there.

Enter `plot(y~x)` to graph the data.

This should show three groups of measurements, indexed by the variable `x`

You can get a boxplot by group with the command `boxplot(y~x)`

One-way Anova

If all goes well, R should download three data frames, `onewayanova1-3`, to your workspace. Use `ls()` to verify that they are there.

Enter `plot(y~x)` to graph the data.

This should show three groups of measurements, indexed by the variable `x`

You can get a boxplot by group with the command `boxplot(y~x)`

You summarize the data by group with the command `tapply(y, as.factor(x), summary)`

One-way Anova

The objective in a one-way ANOVA analysis is usually to determine whether there is any difference among the mean Y values for the three groups.

One-way Anova

The objective in a one-way ANOVA analysis is usually to determine whether there is any difference among the mean Y values for the three groups.

Although we estimate the coefficients the same way as we did in a regression analysis, the emphasis is on the overall F statistic which tests whether the group means are all the same.

One-way Anova

The objective in a one-way ANOVA analysis is usually to determine whether there is any difference among the mean Y values for the three groups.

Although we estimate the coefficients the same way as we did in a regression analysis, the emphasis is on the overall F statistic which tests whether the group means are all the same.

In fact, most statistical packages do not display the coefficients unless you ask for them.

One-way Anova

The objective in a one-way ANOVA analysis is usually to determine whether there is any difference among the mean Y values for the three groups.

Although we estimate the coefficients the same way as we did in a regression analysis, the emphasis is on the overall F statistic which tests whether the group means are all the same.

In fact, most statistical packages do not display the coefficients unless you ask for them.

To start, we will examine the structure of the data frame with the command `str(onewayanova1)`

One-way Anova

This should show us that `onewayanova1` is a data frame with 60 rows, and contains two numeric arrays called x and y .

One-way Anova

This should show us that `onewayanova1` is a data frame with 60 rows, and contains two numeric arrays called x and y .

Our classification variable, x , needs to be of type *factor* for the R ANOVA routine to handle it properly.

One-way Anova

This should show us that `onewayanova1` is a data frame with 60 rows, and contains two numeric arrays called x and y .

Our classification variable, x , needs to be of type *factor* for the R ANOVA routine to handle it properly.

To convert it to a factor, we can use the command
`as.factor(x)`

One-way Anova

This should show us that `onewayanova1` is a data frame with 60 rows, and contains two numeric arrays called x and y .

Our classification variable, x , needs to be of type *factor* for the R ANOVA routine to handle it properly.

To convert it to a factor, we can use the command
`as.factor(x)`

To run the ANOVA and save the results, enter
`aov1=aov(y~as.factor(x))`

One-way Anova

This should show us that `onewayanova1` is a data frame with 60 rows, and contains two numeric arrays called x and y .

Our classification variable, x , needs to be of type *factor* for the R ANOVA routine to handle it properly.

To convert it to a factor, we can use the command
`as.factor(x)`

To run the ANOVA and save the results, enter
`aov1=aov(y~as.factor(x))`

If we examine `aov1` with `str(aov1)`, we see that it is a list containing information about the linear model, much like the output from the `lm` function. `aov()` is customized for ANOVA while `lm()` is customized for regression.

One-way Anova

The overall F test is significant if $Pr(> F)$ is smaller than our chosen alpha value (usually 0.05).

One-way Anova

The overall F test is significant if $Pr(> F)$ is smaller than our chosen alpha value (usually 0.05).

The alpha value is the probability of getting a false positive when there is actually no difference between the groups.

One-way Anova

The overall F test is significant if $Pr(> F)$ is smaller than our chosen alpha value (usually 0.05).

The alpha value is the probability of getting a false positive when there is actually no difference between the groups.

This can always happen by chance if one group ends up with a disproportionate share of the larger e terms.

One-way Anova

The overall F test is significant if $Pr(> F)$ is smaller than our chosen alpha value (usually 0.05).

The alpha value is the probability of getting a false positive when there is actually no difference between the groups.

This can always happen by chance if one group ends up with a disproportionate share of the larger e terms.

Once the overall F test is determined to be significant, we can do a follow-up test to determine which groups are different. Enter `TukeyHSD(aov1)` to display this.

One-way Anova

The Tukey Honest Significant Difference is designed to protect the alpha level when many comparisons are done.

One-way Anova

The Tukey Honest Significant Difference is designed to protect the alpha level when many comparisons are done.

As the number of comparisons increases, the probability of at least one false positive increases rapidly. The Tukey test attempts to correct for this.

One-way Anova

The Tukey Honest Significant Difference is designed to protect the alpha level when many comparisons are done.

As the number of comparisons increases, the probability of at least one false positive increases rapidly. The Tukey test attempts to correct for this.

To interpret the output, note that there is one line for each distinct pair of levels of the factor x , and each line shows l_{wr} and u_{pr} limits for the difference.

One-way Anova

The Tukey Honest Significant Difference is designed to protect the alpha level when many comparisons are done.

As the number of comparisons increases, the probability of at least one false positive increases rapidly. The Tukey test attempts to correct for this.

To interpret the output, note that there is one line for each distinct pair of levels of the factor x , and each line shows l_{wr} and u_{pr} limits for the difference.

If zero does not lie between l_{wr} and u_{pr} , we declare the difference between those two levels to be significant. Otherwise, we say it is not significant.