

## Week Two

[Help Center](#)

# Foundations of Regression Analysis

This week we will start with linear regression for a single explanatory variable. This builds upon the basic statistical concepts required for regression analysis covered during week one.

In this week you will:

- learn to make a scatter plot and fit a regression line using STATA.
- be acquainted with the assumption of homoscedasticity for linear regression.
- learn to draw inference based on hypothesis testing and estimation of confidence interval.

We will end this week learning about prediction interval and also have a chance to see a demonstration on STATA that will be helpful in completing your first homework.

## Lectures

Please click on the links below to access the video lectures for this second week

- [Linear Regression I](#)
- [Linear Regression II](#)
- [Assumptions of Linear Regression](#)
- [Hypothesis Testing and Confidence Interval](#)
- [Confidence Interval and Homework](#)

## Lecture Material

Please click on the link below to download the slides of the second week

[Week Two "Fundamentals of Regression Analysis"](#)

## Conversations

Please join in the conversations around regression analysis in our [community forums](#) area. You can ask and answer questions and discover insights and help for yourself and others as we come together to encourage each other in our exploration.

## Key Terms

Below are definitions of some important terms covered this week:

- **Scatter Plot/Diagram:** Scatter plot is a plot to display values for two variables of the dataset. It is a plot of Y values versus X values. Making a scatter plot should be the first step in regression analysis.
- **Outlier:** An outlier is an observation point that is far from the other observations. Outlier can affect the slope of the regression line. Scatter plots are useful in detecting outliers.
- **Residuals:** Residual is the difference between observed value and the predicted value. Predicted values are obtained by using regression equation.
- **Homoscedasticity:** Homoscedasticity is also known as the homogeneity of variance. It means that the dependent variable exhibits similar amounts of variance across the range of values for an independent variable. This is an important assumption in linear regression analysis.
- **Naïve Model:** The Naïve Model is a model without any explanatory/ predictor variables. This model only contains an intercept value with no other covariates. Therefore, every observation in the dataset will be modeled with the mean outcome. The mean of the Independent variable is the naïve model.
- **Prediction Interval:** Prediction interval is an interval for a random variable that is yet to be observed. This sort of an interval is for an individual future observation and not for the true population mean.

## Homework

Navigate to the [Week Two Homework](#) page to view and download the homework for this week.

## Quiz

After you've gone through the materials for this week please be sure to visit the [quizzes area](#) to complete this week's quiz.



[Header image](#) is used and altered with permission from [Kevin Dooley](#) according to its [Creative Commons Attribution 2.0 Generic License](#).

---

Created Mon 2 Feb 2015 12:13 PM PST

Last Modified Tue 31 Mar 2015 8:58 AM PDT