

Concepts to know

Marcelo Coca Perrillon
University of Colorado AMC
marcelo.perrillon@ucdenver.edu

Although this class is about linear and logistic regression, you need to be familiar with basic mathematical statistics, which does require calculus. We won't be taking complicated derivatives or integrals but you do need to know what they mean. Below is a list of key concepts that you should know and where to find good introductions. We will review some of these concepts during the second week of class. But since we will move fairly quickly, I suggest you study on your own before the start of the class. In my experience, not knowing the basics hinders learning applied statistics. You do not need to know how to do complicated proofs to do well this semester but you do need to master the concepts behind proofs. When reviewing the concepts below, focus on the *meaning* and the big picture.

1 Concepts

1.1 Calculus

Understand the concept of a **limit**. Calculus is built on the concept of limits. Derivatives and integrals are limits. Many proofs in mathematical statistics involve limits as well. A good place to start: <https://www.khanacademy.org/math/differential-calculus/limit-basics-dc/limits-introduction-dc/v/introduction-to-limits-hd>.

Once you understand the concept of a limit, make sure that you understand that the **derivative** of a function $y = f(x)$ is the change in y for a small change in x . "Small" means small as possible, or alternative, when the change h tends to zero:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Recall, too, that the derivative $f'(x)$ is the slope of a function at some particular point or the instantaneous rate of change (stare at the above formula for a while and recall the formula for the slope of a curve and you will see the connection). This is a good place to start remembering: <https://www.khanacademy.org/math/calculus-home/taking-derivatives-calc>.

Also, remember that we use derivatives to find **critical points** (maxima and minima of functions). See this video:

<https://www.khanacademy.org/math/calculus-home/derivative-applications-calc/critical-points-calc/v/minima-maxima-and-critical-points>

Then move on to **integrals**. You can think of integrals as antiderivatives. More important for statistics, however, is to understand that the definite integral is the

area under a curve (remember the fundamental theorem of calculus?). Start here: <http://tinyurl.com/jr99fbd>.

1.2 Statistics

We will review the basics of mathematical statistics during class, but you do need to remember what is a **probability density function** (pdf) and a **cumulative distribution function** (cdf). Remember that we can describe the distribution of a discrete or continuous variable using a mathematical expression for the pdf (or cdf). Once we have that expression, we can calculate **probabilities**, **expected values**, and **variances**.

We will work with the **normal**, **Bernoulli**, and **Binomial** distributions during the semester but we will use other distributions to simulate data (for example, Gamma and Beta). Make sure you know what are the parameters of a pdf. For example, the normal distribution has two parameters: μ and σ . The Bernoulli distribution has one, p . The Binomial two: n and p . The Gamma distribution has two: the shape, α , and scale, β .

Please go over these videos:

<https://www.khanacademy.org/math/statistics-probability/random-variables-stats-library>

2 Stata

By the end of the semester, you should know Stata very well. Stata is your friend. It is easy to learn, easy to use and the documentation is fantastic. The PDF documentation is essentially a very good stats textbook. And no, learning **R** doesn't make you a statistician. **R** is good for newer methods, but the documentation is not great and there are many ways of doing basic things, which makes learning the basics confusing.

Make sure you have access to Stata. I'll go over the basics. Each class will have tons of examples. I'll also post all the code that I used to make slides so you can practice. But it won't hurt you if you start learning on your own before the class starts. There are plenty of resources online. Start here:

<http://www.ats.ucla.edu/stat/stata/sk/default.htm>

Click on "Class notes with movies."