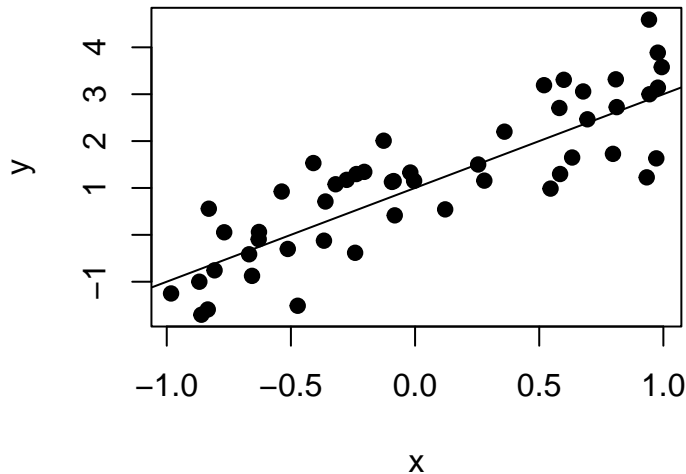


# INF511: Modern Regression I

## Lecture Materials - Introduction

In this course, we seek a deep understanding of linear modeling concepts and methods. Regression analysis allows us to quantify the statistical relationships between a dependent variable (i.e., outcome) and one or more covariates (i.e., independent variables, features). INF511 focuses on *linear* regression.



Linear regression seeks to answer the following types of questions:

- Is there a detectable linear association between my  $y$  and  $x$  variables?
  - If so, can we use this linear relationship to **predict** future observations of  $y$ ?
- Which  $x$  variables have measurable, linear effects on my independent variable,  $y$ ?
  - Put in a different way, are some  $x$  variables important for explaining patterns in my  $y$  variable, while other  $x$  variables are not measurably important?
- Does this experimental treatment have a linear effect on outcome variable,  $y$ ? (ANOVA)
  - Do these two experimental treatments have independent or interactive linear effects on the outcome variable,  $y$ ?
  - Which of my treatment options had larger linear effects on the outcome,  $y$ ?

Regression analysis is foundational to predictive modeling and machine learning techniques.

## Foundational Methods

INF511 will explore statistical and computational **methods** that allow us to estimate linear model parameters and their uncertainties, so that we can address hypotheses about the relationships between dependent and independent variables. The methods on which we will focus are:

| **Ordinary Least Squares** | **Maximum Likelihood** | **Bayesian Inference** |

These methods underpin a *huge* fraction of statistical/machine learning techniques and algorithms. If you can understand these methods and how they work for simple linear models, then you will be able to approach more complex models with confidence.

## Foundational Skills

There are some foundational skills and concepts that we will draw upon and practice throughout this course.

Computer Science	Stats/Probability	Mathematics
R programming	Probability distributions	Matrix algebra
RStudio IDE	- normal, $t$ , $F$ , $\chi^2$	Calculus
Quarto publishing	Likelihood	- derivatives
Algorithms	$p$ -values	- optimization/gradients

## Starting with the basics

Today we will review the syllabus. We will then do some practice exercises in R that will introduce some basic skills:

- R, RStudio, Quarto documents
- Simple data structures
  - objects
  - vectors/arrays
  - data frames
- Simple data visualization
  - histograms
  - scatter plots
- Brief (incomplete) example of regression analysis
  - use of `lm()` function (Least Squares method)
  - use of `summary()` function

## Up next: Probability distributions