Social Sciences Intro to Statistics

Week 7.1 Introduction to Bivariate Regression

Week 4: Learning goal - Demonstrate estimation and prediction of bivariate regression analysis in R.

Introduction

Load packages:

```
library(tidyverse)
library(ggplot2)
best_netflix <- read_csv("https://raw.githubusercontent.com/bcl96/Social-Sciences-Stats/ma
```

Resources used to create this lecture:

Bivariate Regression

"Bivariate regression" refers to regression models with two variables, a Y variable ("dependent variable" or "outcome") and a single X variable ("independent variable").

"Multivariate regression" refers to regression models with a Y variable and two or more X variables

This lecture – which we will teach over several weeks – teaches the fundamental concepts of bivariate regression. All of these concepts will be similar when we move on to multivariate regression.

Scatterplot

Relationships between two continuous variables

Postive relationship, negative relationship, and no relationships

Relationship between X and Y is positive

- when X is "high", Y tend to be "high"
- when X is "low", Y tends to be "low"
- e.g., number of hours (X) studying and GPA (Y)
- e.g., cost of attendance (X) and student debt (Y)

Relationship between X and Y is negative

- when X is "high", Y tend to be "low"
- when X is "low", Y tends to be "high"
- e.g., number of school absences and GPA

No relationship between X and Y

- knowing the value of X gives you does not tell you much about the value of Y
- e.g., amount of ice cream consumed and GPA (defined as "research" or "master's" universities by the Carnegie Classification that are

we will use the data frame `df_socialwork` (which combines debt/earnings data from College Scorecard and tuition/cost of attendance data from IPEDS) to run regression models of the relationship between measures of tuition/COA (X variable) and debt/earnings (Y variable) for MA programs in social work

Ways to investigate this relationship between X and Y:

- Graphically: scatterplots
- Numerically: covariance (less used), correlation

Scatterplots

Scatterplots will plot individual observations on an X and Y axis

Draw scatterplot of X (`coa_grad_res`) and Y (`debt_all_stgp_eval_mean`)

```
df_socialwork %>%  ggplot(aes(x=coa_grad_res, y=debt_all_stgp_eval_mean)) + geom_point()
```

Create scatterplot with "prediction" line

```
df_socialwork %>%  ggplot(aes(x=coa_grad_res, y=debt_all_stgp_eval_mean)) + geom_point() +
```

Covariance

Correlation