STAT 32950: Homework 0

Robert Winter

Table of Contents

1 Exercise 1: Typing Math Formula

For this question, type out your answers, including the mathematical formula. Consider a dataset containing n pairs of real numbers, (x_i, y_i), $i = 1, ..., n$. Suppose the data are n independent observations of a pair of random variables (X, Y). Type the formula of the Pearson sample correlation coefficient of (X, Y) in terms of the x_i's and y_i's.

The Pearson sample correlation coefficient of (X, Y) is

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i \ \text{ and } \ \bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i.$$

2 Exercise 2: Producing R Plot

Produce data and 2-dimensional scatter plot with least square line by using the following R commands. Replace ??? in the last command by an appropriate plot title.

```
x = runif(30); y = x^3 + rnorm(30)/3
plot(x,y); abline(lm(y~x))
title(main = "Linear Regression of Y on X")
```

Linear Regression of Y on X