Statistical Modeling Course

Collinearity Lab

This lab focuses on the *collinearity* problem. Perform the following commands in `R` . The last line corresponds to creating a linear model in which `y` is a function of `x1` and `x2`.

```
set.seed(1)
x1 = runif(100)
x2 = 0.5*x1 + rnorm(100)/10
y = 2 + 2*x1 + 0.3*x2 + rnorm(100)
df = tibble(y, x1, x2)
```

Problem 1

What is the correlation between `x1` and `x2`? What is the variance inflation factor? How about the condition number of $X^T X$?

Problem 2

Using this data, fit a least squares regression to predict `y` using `x1` and `x2`. How do these relate to the true β_0, β_1, and β_2? Can you reject the null hypothesis $H_0 : \beta_1 = 0$? How about the null hypothesis $H_0 : \beta_2 = 0$?

Answer:

Problem 3

Now fit a least squares regression to predict `y` using only `x1` . Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

Answer:

Problem 4

Now fit a least squares regression to predict `y` using only `x2` . Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?

Answer:.

Problem 5

Do the results obtained in Problem 2 and 4 contradict each other? Explain your answer. *Answer:*