

IST 5535 Lab 2: Collinearity

Instruction: Finish the following steps in R Markdown. Submit your R Markdown and HTML report to assignment on Canvas.

Tasks:

1. Perform the following commands in R:

```
set.seed(1)
x1<- runif(100)
x2 <- 0.5* x1 + rnorm(100) /10
y <- 2 + 2* x1 + 0.3* x2 + rnorm(100)
```

The last line corresponds to creating a linear model in which y is a function of $x1$ and $x2$. Write out the form of the linear model. What are the regression coefficients?

2. What is the correlation between $x1$ and $x2$? Create a scatterplot displaying the relationship between the variables.
3. Using this data, fit a least squares regression to predict y using $x1$ and $x2$. Describe the results obtained. What are β_0 , β_1 , and β_2 ? 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 $H_0 : \beta_2 = 0$?
4. Now fit a least squares regression to predict y using only $x1$. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?
5. Now fit a least squares regression to predict y using only $x2$. Comment on your results. Can you reject the null hypothesis $H_0 : \beta_1 = 0$?
6. Do the results obtained in (3)–(5) contradict each other? Explain your answer.
7. Now suppose we obtain one additional observation, which was unfortunately mismeasured.

```
x1 <- c(x1 , 0.1)
x2 <- c(x2 , 0.8)
y <- c(y ,6)
```

Re-fit the linear models from (3) to (5) using this new data. What effect does this new observation have on the each of the models? In each model, is this observation an outlier? A high-leverage point? Both? Explain your answers.