Stat 6021: Multicollinearity When Predictors Are Uncorrelated and When Predictors Are Perfectly Correlated

Read this after Section 3 in Guided Notes.

1 Perfectly Uncorrelated Predictors

Consider the following vectors: $\boldsymbol{x_1} = (4, 4, 4, 4, 6, 6, 6, 6)$, $\boldsymbol{x_2} = (2, 2, 3, 3, 2, 2, 3, 3)$, and $\boldsymbol{y} = (42, 39, 48, 51, 49, 53, 61, 60)$. The sample correlation coefficient between x_1 and x_2 is $r = 0$. We consider regressing y on x_1 and x_2 versus regressing y on x_2.

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x1         1 231.125 231.125  65.567 0.0004657 ***
x2         1 171.125 171.125  48.546 0.0009366 ***
Residuals  5  17.625   3.525
```

Next, consider regressing y on x_2.

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
x2         1 171.12 171.125  4.1276 0.08846 .
Residuals  6 248.75  41.458
```

Notice that $SS_R(\beta_2|\beta_1) = SS_R(\beta_2)$. This informs us the predictors are perfectly uncorrelated with each other; x_2 provides completely uncorrelated information from x_1. Next, we compare regressing y on x_2 and x_1 versus regressing y on x_1.

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x2         1 171.125 171.125  48.546 0.0009366 ***
x1         1 231.125 231.125  65.567 0.0004657 ***
Residuals  5  17.625   3.525

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value  Pr(>F)
x1         1 231.12 231.125   7.347 0.03508 *
Residuals  6 188.75  31.458
```

Notice that $SS_R(\beta_1|\beta_2) = SS_R(\beta_1)$.

2 Perfectly Correlated Predictors

Consider the following example. We have predictors $\boldsymbol{x_1} = (2, 8, 6, 10)$ and $\boldsymbol{x_2} = (6, 9, 8, 10)$, with $\boldsymbol{y} = (23, 83, 63, 103)$. The following two fitted regression equations will give perfect fits:

$$\hat{y} = -87 + x_1 + 18x_2$$

and

$$\hat{y} = -7 + 9x_1 + 2x_2$$

There exist infinitely many equations that will fit the data. $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are perfectly correlated, by $\boldsymbol{x_2} = 5 + 0.5\boldsymbol{x_1}$. If you tried to find the estimated regression coefficients, you would find no solution since $\boldsymbol{X'X}^{-1}$ does not exist when predictors are perfectly correlated.