Assignment 4

Emil Hafeez (eh2928)

10/28/2020

```
knee_df =
  read_csv(
      "./data/Knee.csv") #load in df
```

```
## Parsed with column specification:
## cols(
##   Below = col_double(),
##   Average = col_double(),
##   Above = col_double()
## )
```

Problem 1

In the context of ANOVA models, prove the partitioning of the total variability (sum of squares).

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{\bar{y}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{y}_l\right)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{y}_i-\bar{\bar{y}}\right)^2 \tag{1}$$

We assume that the samples are drawn independently from the underlying populations, that the distributions of the error terms are normal, and that the variances of the k populations are equal (and thus that the variance of the outcome does not depend on the sample).

Let n be the total number of observations

Let $\bar{y}$ be the mean from the ith group

Let $\bar{\bar{y}}$ be the grand mean, such that $\bar{\bar{y}} = \dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{n}$