2.7 A dataset has 1000 records and 50 variables with 5% missing values, spread randomly throughout the records and variables. An analyst decides to remove records with missing values. About how many records would you expect to be removed?

Solution:

Number of Records (n) = 1000  
Number of Variables (v) = 50  
Percent of Missing Values (mv) = 0.05

The probability of no missing values = (1 - 0.05)50 ≈ 0.0769. This means that only about 7.69% of the records are expected to have no missing values. Therefore, the percentage of records you would expect to be removed is Percentage of records removed=100%−7.69%=92.31%. The analyst would expect to remove approximately 92.31% of the records, making the expected number of records the analyst would remove **923.**

2.8 Normalize the data in Table 2.17, showing calculations.

A table with numbers and a few people

Description automatically generated with medium confidence

Step 1: Load the data from the table to a data frame.

A screen shot of a computer

Description automatically generated

Step 2: Checking the mean and standard deviation of the two columns

A screen shot of a computer code

Description automatically generated

Normalization of a measurement is obtained by subtracting the average from each measurement and dividing the difference by the standard deviation.

A screenshot of a computer

Description automatically generated