Ervin Sarmiento

Probability and Applied Statistics

Project 1

This report will ***examine*** material covered in Probability and Statistics based on chapters 1 to 4. A relative frequency Histogram will be constructed based on a homework problem. Based on the lessons presented in each chapter, the following Java programs: SetOperations, Monte Carlo Simulation, and Birthday, were constructed to apply the material learned in class into a program. A statistics library of formulas from chapters **1-4** will be created using the Java language by utilizing Eclipse IDE. The following formulas: mean, median, and mode, variance, standard deviation, permutations and combinations, geometric probability distribution, hypergeometric probability distribution, Poisson probability distribution, continuous random, and probability density function, were used to create the statistics library.

**Graphical user interface, application, table

Description automatically generated** *Figure 1.1 - “Homework” Pivot Table*

A histogram is a graph that shows the frequency distributions from a set of measurements (Wackerly et al., 2008). The “graph is constructed by subdividing the axis measurement into intervals equal width” (Wackerly et al., 2008). On page 6 of the textbook, exercise 1.2, a histogram was constructed based on the set of measurements presented in the exercise. The “homework” histogram graph displays a graph of average wind speed (miles per hour) for 45 selected U.S. cities. A pivot table was created to help summarize the data before constructing the histogram (figure 1.1). The results of the histogram shown that 40 percent of the 45 selected cities recorded average wind speeds between 7.1-9.1 miles per hour and 9.1-11.1 miles per hour. Only 20 percent for the selected cities recorded wind speeds between 11.1-13.1 miles per hour.

*Text

Description automatically generated*

*Figure 2.1 - SetOperations Results*

The Set program will return the union and intersect of two array lists and the complement of an array list. This program uses a String data type, and the array lists will contain days of the week (“Monday”, “Tuesday”, “Wednesday”, “Thursday”, “Friday”, “Saturday”, and “Sunday”). The union of the two array lists will return all elements that are in at least one of the sets (Wackerly et al., 2008). The intersect will return the set of all elements in both array lists (Wackerly et al., 2008). The complement will return the set that includes all the elements in the universal set, which are not present in the given set (Wackerly et al., 2008).

Text

Description automatically generated

*Figure 2.2 – Monte Carlo Results*

The Monte Carlo simulation program simulates the probability of picking the right/winning door if given three different doors. It highlights the win percentage if switching doors is more advantageous compared to just staying with your original option of the door. The program determines that if you change doors after a wrong choice is picked, the chance of getting the right winning pick becomes higher.

Text

Description automatically generated

*Figure 2.3 – Birthday Results*

The Birthday program will determine the probability of any 2 people sharing a birthday in the class. The class size and number of runs are based on user input. Based on the results of the program, the probability of any 2 people in the class of 19 people on a run of 10,000 is .3852 or 38%.

Text

Description automatically generated

*Figure 3.1 – Mean, Median, and Mode Results*

Mean is the average a data set, which is obtained by getting the sum of the data set and dividing it by the number of values in the set. Median is the middle value of an ordered data set and the mode is the number that occurs most often in the data set.

The data set used in the program is 10, 5, 6, 2, 20, 20

The mean of the data set is 10

The median of the data set is 8

The mode of the data set is 20

Text

Description automatically generated

*Figure 3.2 – Variance and Standard Deviation*

The variance “is the sum of the square of the differences between the data set and their mean, divided by n-1” (Wackerly et al., 2008). The sample size is represented by n. The standard deviation of a data set “is the positive square root of the variance” (Wackerly et al., 2008).

The data set used in the program is 10, 40, 25, 17, 22, 15, 7, 19

The variance of the data set is 91.234375

The standard deviation of the data set is 9.551668702378658

Text

Description automatically generated

*Figure 3.3 – Permutation and Combination*

Permutation and combination can be used to find the number of ways in which a set of objects can be ordered or selected. Permutation is “an ordered arrangement of *r* distinct objects” (Wackerly et al., 2008). Combination is “the number of combinations of *n* objects taken *r* at a time is the number of subsets, each of size *r*, that can be formed from the *n* objects” (Wackerly et al., 2008).

The value of n is 10 and r is 2.

The permutation is 90

The combination is 45

Geometric probability distribution “is associated with an experiment that shares some of the characteristics of a binomial experiment” and a “random variable *Y* is the number of the trials on which the first success occurs” (Wackerly et al., 2008). Hypergeometric probability distribution is a discrete probability of an experimental setting (Wackerly et al., 2008). Poisson probability distribution is widely used to model the distribution of count data (Wackerly et al., 2008).

Continuous random is a random variable that has only continuous values (Wackerly et al., 2008).

Probability density function “is a theoretical model for the frequency distribution of a population of measurements” (Wackerly et al., 2008).

Work Cited

Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008). Mathematical statistics with applications (7th ed.). Thomson Brooks/Cole.