**Lab 3 Probability Distributions**

The North Valley Real Estate group has provided you with data on homes sole in the area last year, including the agent who sold the home, the selling price, square footage, number of bedrooms, number of bathrooms, if the property has a pool or garage, the township where the home is located, type and length of mortgage, the buyer’s FICO score, and if the loan is currently in default. You have been asked to evaluate the distributions of the selling prices and days on the market for homes in this area as well as the FICO scores of buyers in the area to provide a guideline for new real estate agents in this group as they set their goals for this coming year.

**Preparation Tasks**

1. Prepare a standardized variable for days on the market.
2. Obtain the summary Statistics for the original variable and the standardized variable including the following items.
   1. Mean
   2. Standard deviation
   3. Minimum
   4. Maximum
   5. Number of observations
   6. Median
   7. Skewness
   8. Kurtosis
   9. Histogram
3. Standardize the variable days on the market and filter your data to remove the observations that are not within ±1.96 standard deviations of the mean. Do not include the standardized variable in your new data set.
4. Obtain the summary statistics for days on the market from the filtered data set.
5. Standardize your original variable in your filtered data sets.
6. Complete a Distribution Analysis for the standardized variables.
   1. extreme values (specify n = 10)

**Analysis Tasks**

1. Provide the table including the summary statistics for days on the market and the standardized value for days on the market.
2. What do you notice when you compare the skewness and kurtosis of the original variables to the skewness and kurtosis of the standardized variables? Why do you think this happens?
3. Provide a histogram of the days on the market for homes in the last year.
4. Provide the table including the summary statistics for days on the market with the observations outside of ±1.96 standard deviations of the mean removed.
5. What happened to the skewness once the extreme values were removed?
6. Provide the histogram of the number of days homes were on the market in the last year with the observations outside of ±1.96 standard deviations of the mean removed.
7. In a normal distribution, 95% of observations ±1.96 standard deviations of the mean. For this data set, what percentage of the original observations were retained after the observations outside of this range were removed?
8. After you standardized the filtered data, how many observations fell outside of the range of ±1.96 standard deviations of the mean? What percentage of the filtered sample fell within that range? Provide the Extreme Values table.
9. The North Valley Real Estate group provides a bonus for its representatives if a property is on the market for 14 days or less.
   1. Using the original data set to two decimal places, what is the probability that one property would qualify an agent for this bonus?
   2. Using the data set with the outliers removed, what is the probability that one property would qualify an agent for this bonus?
10. When an agent lists a property, the home needs to sell by the 15th day of the month for the commission to be on the next paycheck issues on the 1st of the following month.
    1. What is the probability that a home listed on July 1st that did not sell in time to be included on the August 1st paycheck is sold by August 15th (there are 31 days in July), in time to be counted on the paycheck issued September 1 using the original data set?
    2. What is the probability that a home listed on July 1st that did not sell in time to be included on the August 1st paycheck is sold by August 15th (there are 31 days in July), in time to be counted on the paycheck issued September 1, using the data set with the outliers removed?
11. In reviewing the mean, skew, percent of observations within ±1.96 standard deviations of the mean, and results from your probability calculations for your original data set and the data set with the outliers removed, comment on the use of the Empirical rule to identify outliers? (Do any outliers need to be removed? Should a different *z* value be selected?)