Dataset Description

The data NorthValleyRealtor.xlsx contain information on homes sold by the North Valley Real Estate group within the last year. Within this file you will find the following fields:

- Record - Property identification number
- Agent – Name of the real estate agent assigned to the property
- Price – Market price in US dollars
- Size – Livable square feet of the property
- Bedrooms – Number of bedrooms
- Baths – Number of baths, which takes numbers 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5
- Pool – Does this home have a pool? (1 = yes, 0 = no)
- Garage – Does the home have a detached garage? (1 = yes, 0 = no)
- Days – Number of days the property was on the market
- Township – Area where the property is located
- Mortgage type – Fixed or adjustable. The fixed mortgage is a 30 year, fixed interest rate loan. The adjustable rate loan begins with an introductory rate of 3% for the first five years, then the interest rate is based on the current interest rates plus 1% (i.e., the interest rate AND the payment is likely to change each year after the 5^{th} year.).
- Years – The number of years that the mortgage loan has been paid
- FICO – the credit score of the mortgage loan holder. The highest score is 850; an average score is 680; a low score is below 680. The score reflects a person's ability to pay their debts.
- Default – Is the mortgage loan in default? (1 = yes, 0 = no)

t test

We will be using the data NorthValleyRealtor.xlsx to answer the question "Does the price of homes with a pool differ from those without a pool on average?"

1. State the null hypothesis and alternate hypothesis for this research question. (5pts)
   H_0: The mean prices of homes with and without a pool are equal.
   H_A: The mean price of homes with a pool is different from those without a pool.

2. Perform a t test with appropriate settings and provide a screen shot of test results for normality or Q-Q plot for normality. (5pts) According to either of Shapiro-Wilk test (significance level of 0.05) **or** the Q-Q plots, is the price of homes normally distributed? (5pts)

Variable: Price (Price)
Pool (yes is 1) = 0

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.844408	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.138803	Pr > D	0.0642
Cramer-von Mises	W-Sq	0.156196	Pr > W-Sq	0.0199
Anderson-Darling	A-Sq	1.089628	Pr > A-Sq	0.0069

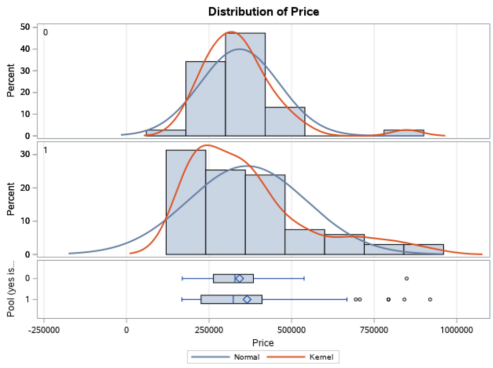
Variable: Price (Price)
Pool (yes is 1) = 1

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	W	0.859683	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.168786	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.459537	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.898599	Pr > A-Sq	<0.0050

OR

The *p* value is less than 0.05, therefore we can conclude that there is sufficient evidence to reject the null hypothesis that the variables are normally distributed.
Or the Q-Q plots show deviation from normal reference line for both pool groups.

3. Provide a screenshot of the box plots of home prices. You can simply generate them using appropriate settings of the t test. (5pts) Do you think the variances of home prices are equal for the two groups? (5pts)



The boxplots show that the two pool groups show different variances of home prices.

4. Provide a screenshot of the test results including **t value and p-value** of both pooled and Satterthwaite t tests. (5pts)

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	103	-0.69	0.4908
Satterthwaite	Unequal	100.3	-0.77	0.4428

5. Which *t* test method, Satterthwaite or pooled, should you use to test the hypothesis? Why? (5pts)
   Satterthwaite due to unequal variances.

6. Complete the following sentence for the interpretation of the test. (2.5pt*2=5pts)

   We can conclude with a *p* value of __0.4428__, that the prices of homes with and without a pool are __equal__ (equal/unequal).

ANOVA

We will be using the data NorthValleyRealtor.xlsx to answer the question "Does the presence of detached garage and presence of pool affect the home price?"

1. Perform an ANOVA with interaction between Garage and Pool, and Price as the dependent variable. Take a screenshot of the ANOVA results (either Type 3 Tests of Fixed Effects or Type 3 Analysis of Variance is good). (5pts)

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Garage (Yes is 1)	1	101	15.25	0.0002
Pool (yes is 1)	1	101	0.00	0.9771
Garage (Y*Pool (yes	1	101	1.09	0.2990

Or

Type 3 Analysis of Variance								
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Garage (Yes is 1)	1	339636707616	339636707616	Var(Residual) + Q(Garage (Yes is 1),Garage (Y*Pool (yes)	MS(Residual)	101	15.25	0.0002
Pool (yes is 1)	1	18379112	18379112	Var(Residual) + Q(Pool (yes is 1),Garage (Y*Pool (yes)	MS(Residual)	101	0.00	0.9771
Garage (Y*Pool (yes	1	24271167592	24271167592	Var(Residual) + Q(Garage (Y*Pool (yes)	MS(Residual)	101	1.09	0.2990
Residual	101	2.2496279E12	22273543716	Var(Residual)	.	.	.	.

2. According to the results, which one of the independent variables and/or their interaction significantly affect the home prices at significance level of 0.05? (5pts)
   The Garage is the only significant factor because of its p-value less than 0.05. Note we don't need further pairwise t tests because garage only have two levels.

   5pts for attempting for the lab