**Dataset Description**

The NorthValleyRealtor.xlsx data 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 5th 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**

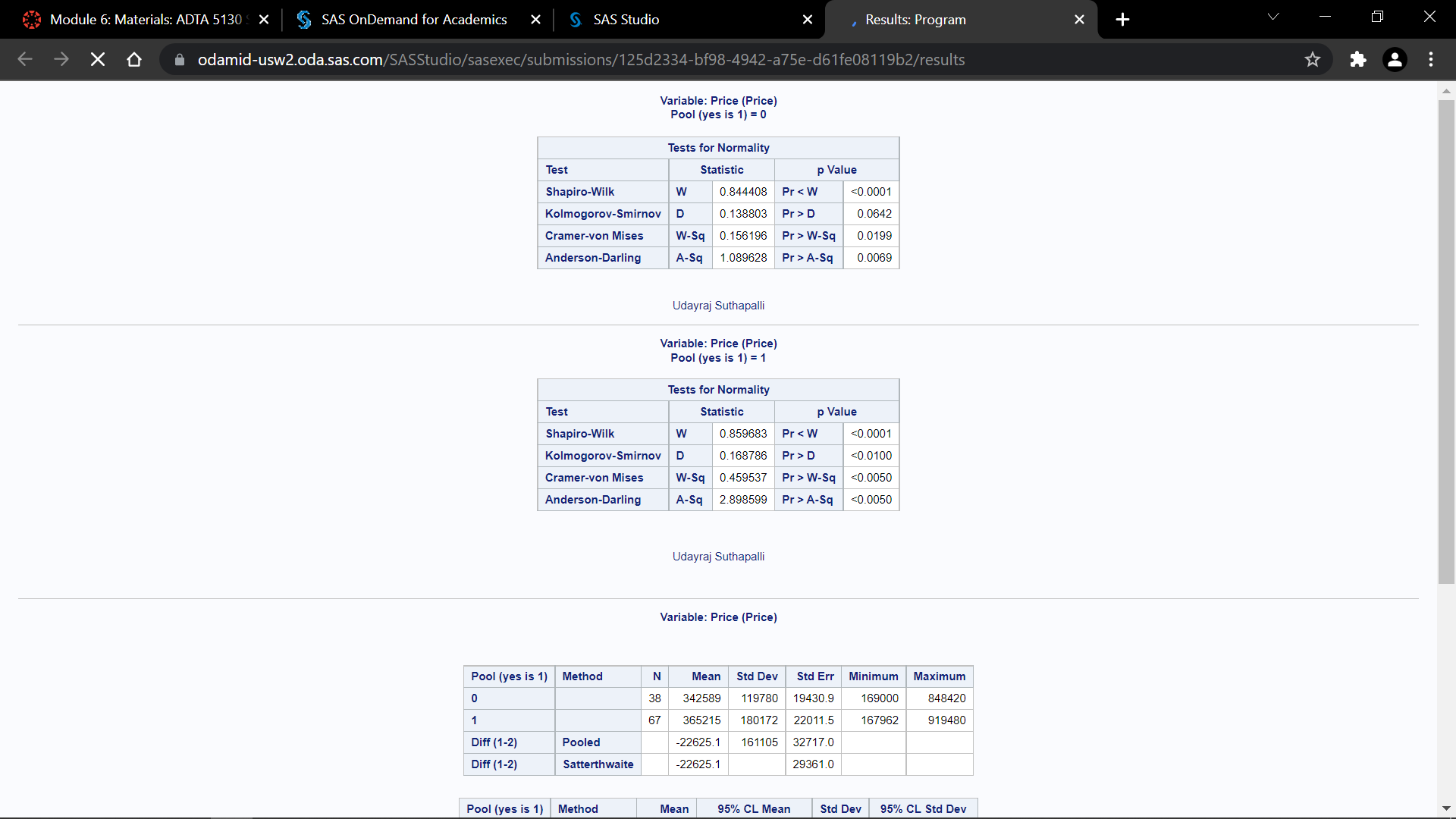
We will be using the data NorthValleyRealtor.xlsx to answer the question “Does the price of homes with a pool differ from those without a pool on average?”

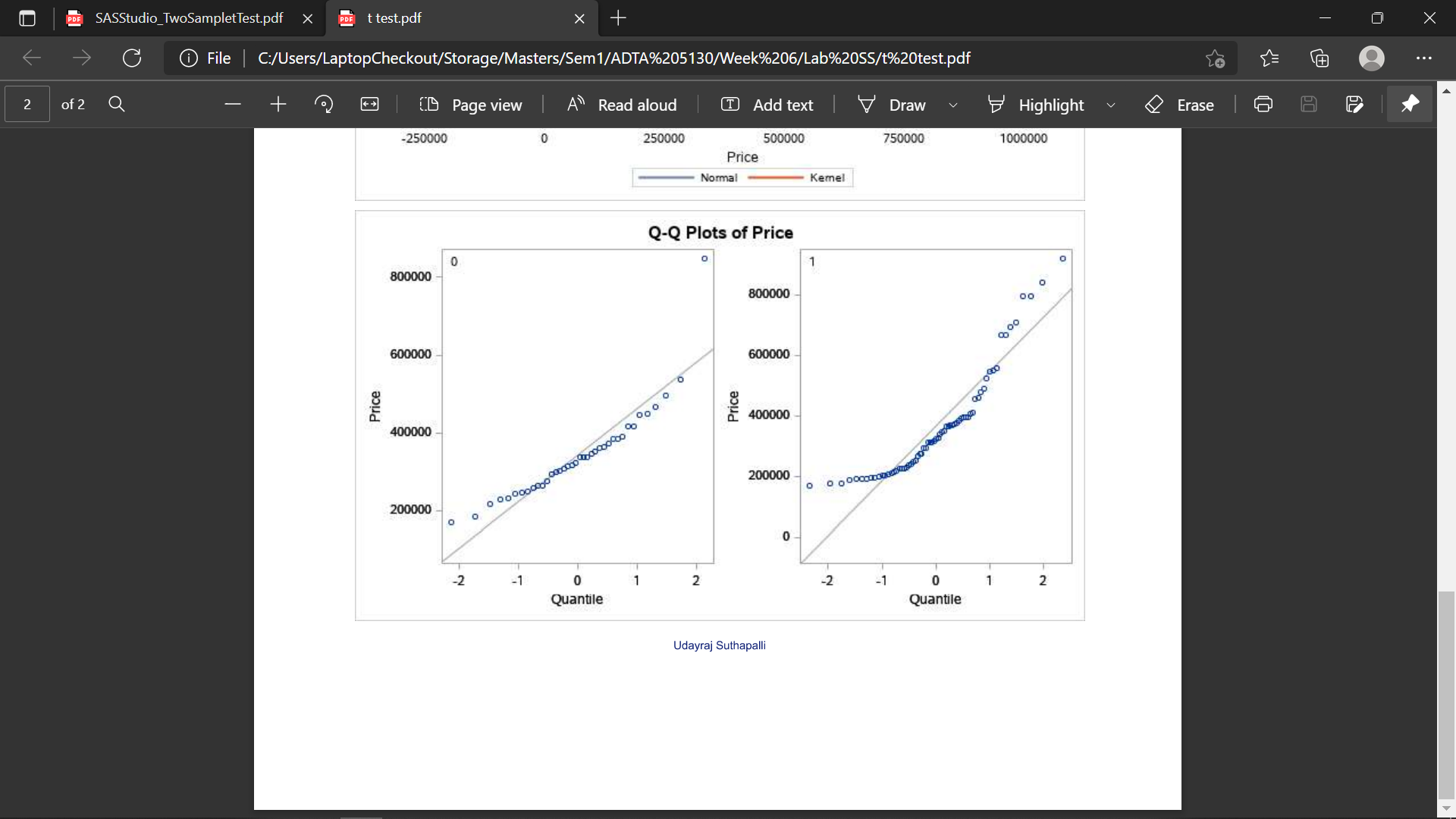
1. State the null hypothesis and alternate hypothesis for this research question.

H0: - The mean price of home with pool is equal to the mean price of home without pool.

HA: - The mean price of home with pool is not equal to the mean price of home without pool.

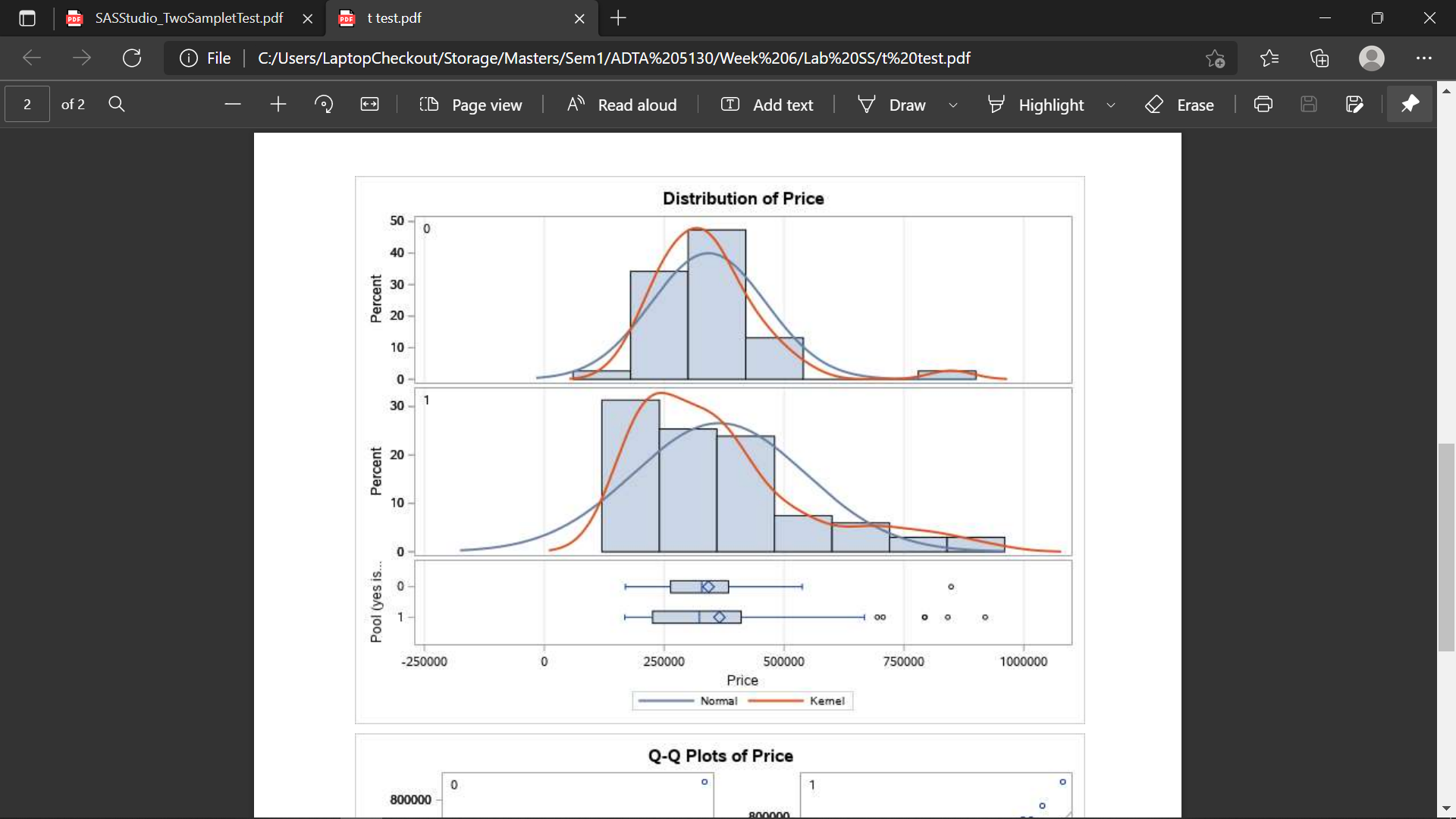
1. Perform a t test with appropriate settings and provide a screen shot of test results for normality or Q-Q plot for normality. According to either of Shapiro-Wilk test (significance level of 0.05) **or** the Q-Q plots, is the price of homes normally distributed?

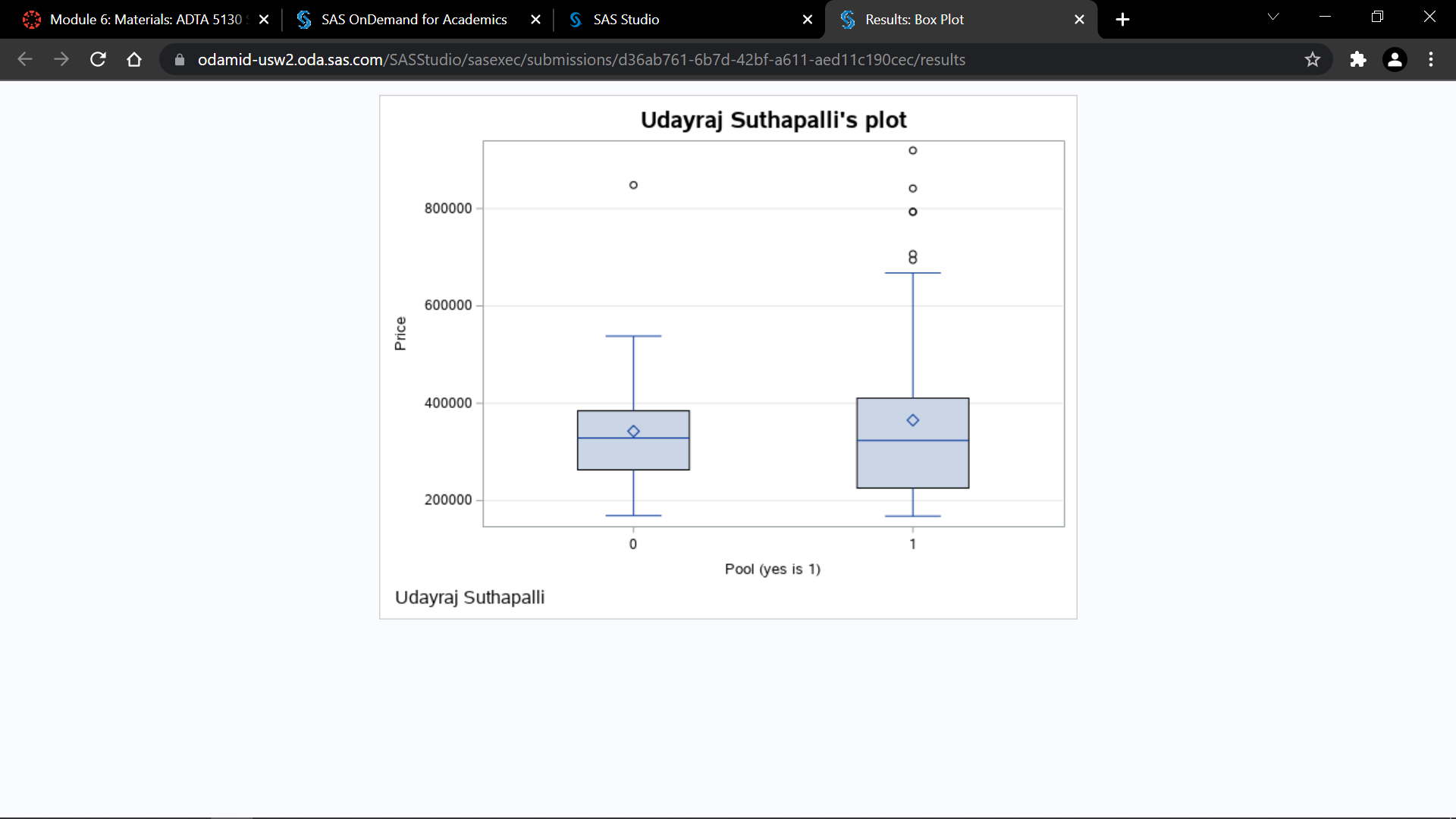




The distribution is normal as both mean and standard deviation of sample data and normally distributed data are equal.

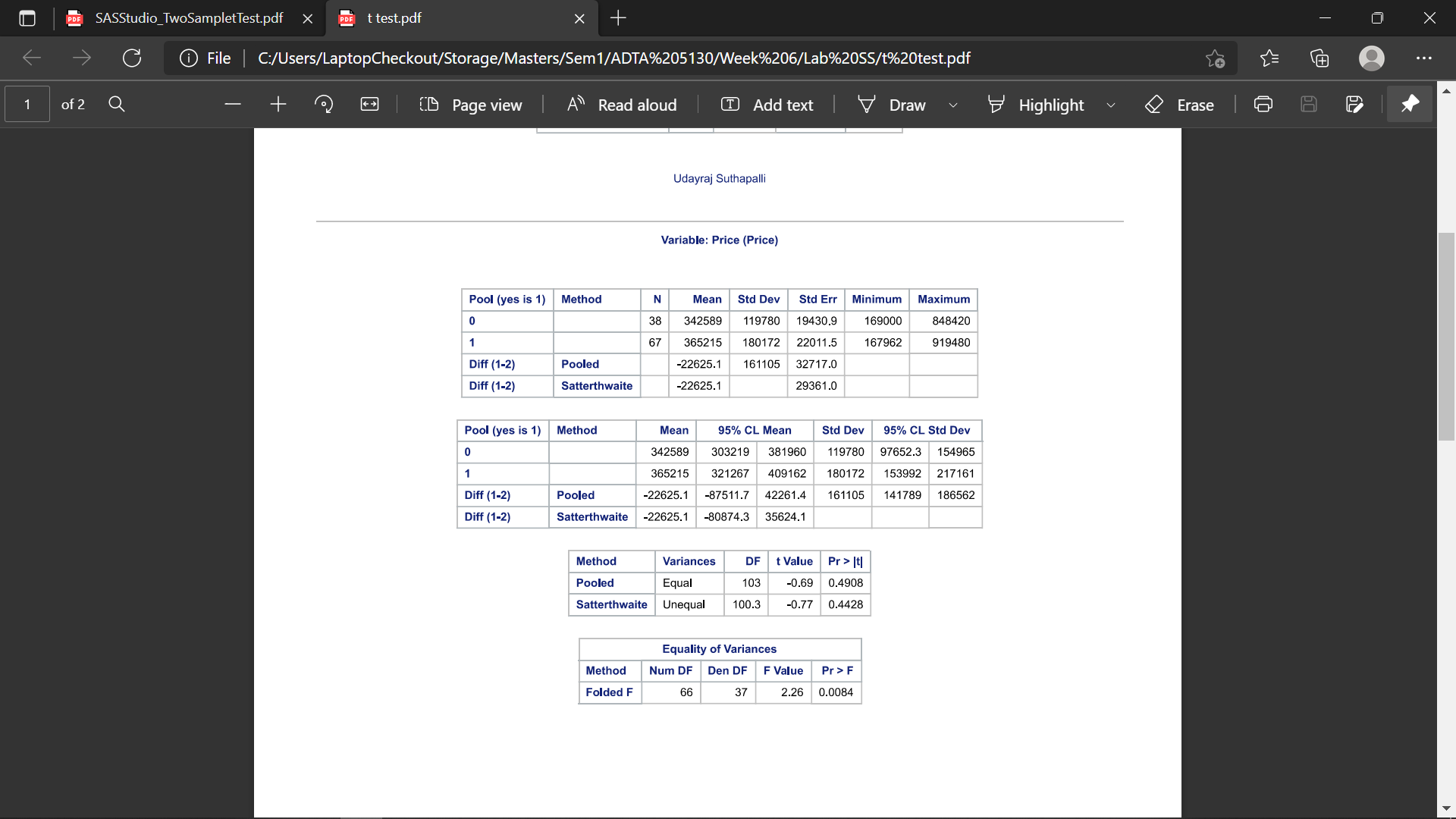
1. Provide a screenshot of the box plots of home prices. You can simply generate them using appropriate settings of the t test. Do you think the variances of home prices are equal for the two groups?





No, the variances of home prices are not equal from the above box plots. We can observe lot of difference and outliers.

1. Provide a screenshot of the test results including **t value and p-value** of both pooled and Satterthwaite t tests.



1. Which *t* test method, Satterthwaite or pooled, should you use to test the hypothesis? Why?

As we can observe a inequality between the variances of prices of homes for groups with pool and without pool. We need to consider the Satterthwaite results which is 0.4428 which is greater than the significance level 0.05. In our case it for both pooled and Satterthwaite values the p-value is greater than 0.05 which means using both the values we can reject the null hypothesis.

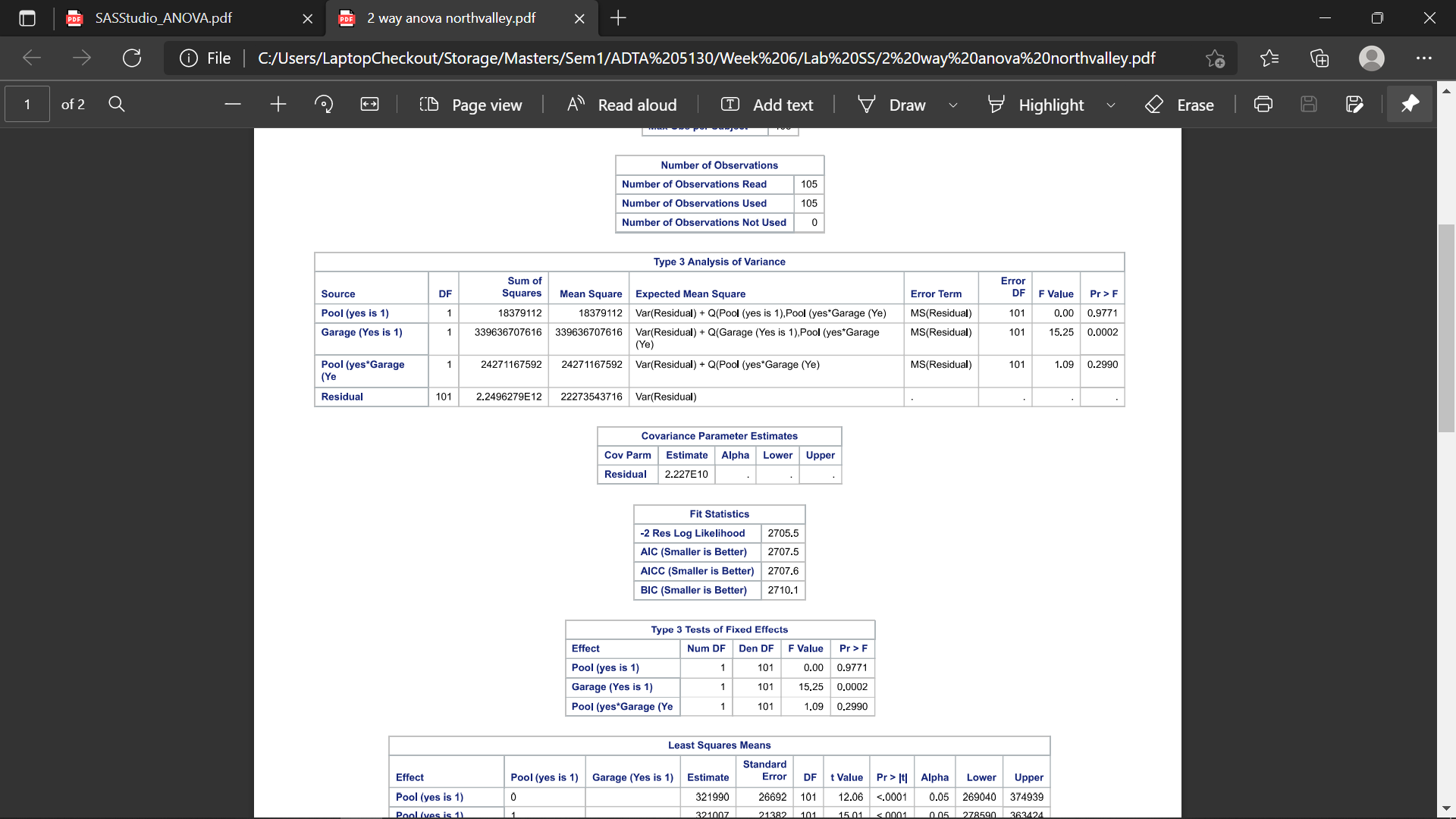
1. Complete the following sentence for the interpretation of the test.

We can conclude with a *p* value of \_0.4428\_\_\_, that the prices of homes with and without a pool are \_Unequal\_\_\_ (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).



1. 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?

From the observed P-values for pool, garage and combination of pool and garage which are 0.9771,0.0002 and 0.2990 respectively the most significant variable in determining the house price is only pool and not the interacted variable of pool and garage.