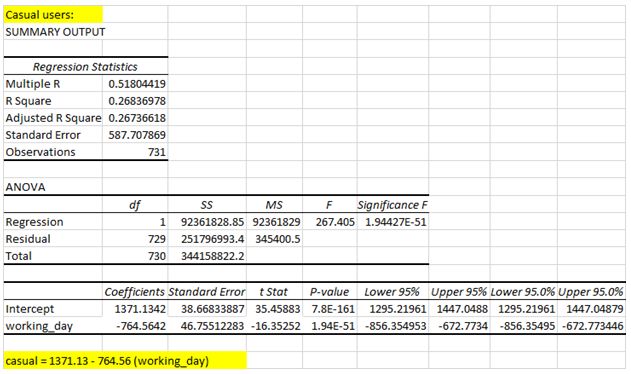
**Project Final Report**

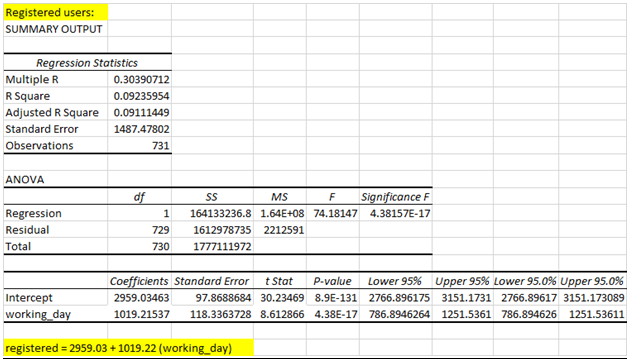
Bike-Sharing in Washington DC

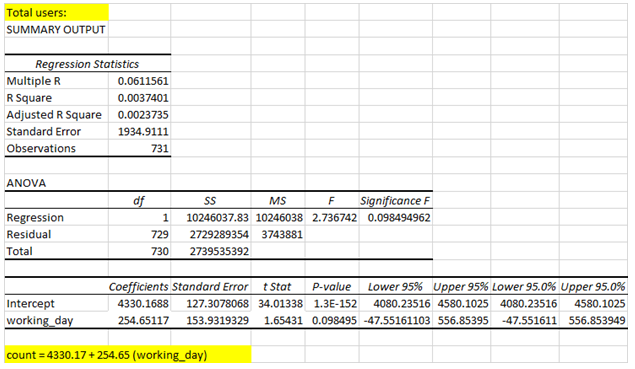
# Problem 1

* Problem Statement
  1. **Is the bike sharing system demand higher for working days or holidays?**
  2. This question uses the binary independent variable working\_day to observe if consumers rent the bikes as a mode of transportation to work or mainly for amusement. The dependent continuous variables here are casual and registered.
  3. The significance of this question is to figure out if registered customers’ consistent motive to rent the bikes is for work or amusement purposes, and similarly for casual users. The importance of knowing the reason for the high demand is to infer the purpose of bike rentals, which helps Capital Share System understand the usage of customers bike rentals and the most effective way of making their product or service available.
* Background
  1. Motivate, an experienced bike share team, conducted a survey on customers from three of their largest systems; Citi Bike, Divvy, and Capital Bikeshare, to find out the purpose individuals use bike-sharing (“Motivate,” 2016). The survey showed that 65% of Capital Bike Share members, 50% of Citi Bike members and 42% of Divvy members use bike share to commute to work (“Motivate,” 2016). The top second reason people use bike sharing is for personal and social appointment (“Motivate,” 2016). This is like the question asked in this problem. However, we used linear- linear regression method to analyze the question rather than a survey.
* Empirical Results and Discussion
  1. Regression Models

The technique used for these three regression models is linear- linear.







* 1. Dataset

Dataset 2 attached. The variables we used from the data set include registered, casual, count and working\_day.

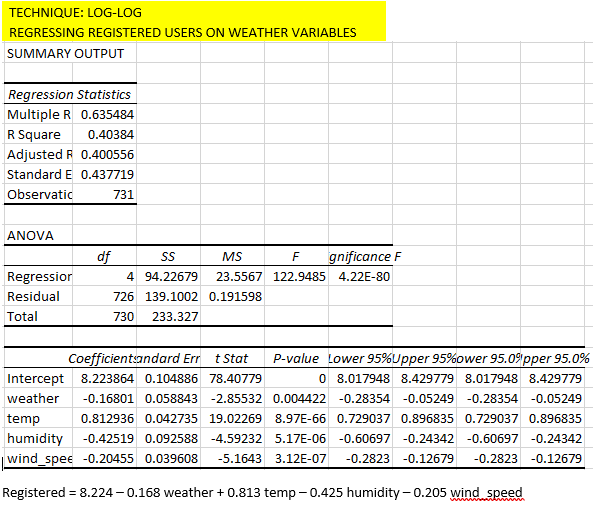
* 1. Results and Interpretation
* From the 1st regression model, we see that the working\_day coefficient has a significant negative value (since t-stat = -16.35 which is less than the t-critical value of -1.96 for the 95% confidence interval).
* Hence, we can say that the number of casual users is more on holidays than on working days.
* This makes sense since casual users would most probably have other established means of transport for commuting to work and hence, would make use of these bikes more on holidays.
* From the 2nd regression model, the working\_day coefficient has a significant positive value and hence, we can say that the number of registered users using the bikes is more during working days than holidays.
* The reason behind this is that people who registered would have mostly done so to use these bikes as a means of transport for work and hence, would prefer not to use them during their holidays as well.
* On comparing both these models, we see that the working\_day coefficient has a higher absolute value for registered users and therefore, we can conclude that whether it is a working day or not impacts registered users more than it does casual users.
* This is because most registered users almost exclusively prefer to use the bikes for work while casual users might use them for work as well, though a larger number of them do prefer to use them during the holidays.
* From the last regression model, we see that working\_day has a positive coefficient that isn’t significant. Hence, we can say that the total number of bikes rented is almost the same irrespective of whether it is a working day or not.
* This conclusion is acceptable since the increase in casual users during the holidays is nearly cancelled out by the decrease in registered users and vice-versa during working days if there is not a significant difference in the number of registered and casual users.
* A limitation of these models is that we do not consider other variables such as weather which could affect these variables and hence, we might have omitted variable bias here.
* Conclusion
* Customers who find bike rentals more essential and use it to commute to work are more likely to be registered users, while customers who find bike rentals nonessential and use it for leisure activities are more likely to be casual users.
* The results obtained imply that the number of registered users using the bikes would be more on working days than holidays and it would be the other way around for casual users. This result coincides with the survey result from Motivate which shows that individuals use bikeshare mostly to commute to work.
* Thus, in a city with less registered users, more bikes should be made available during non-working days. Additionally, if the registered user is greater, then more bikes should be made available during working days.

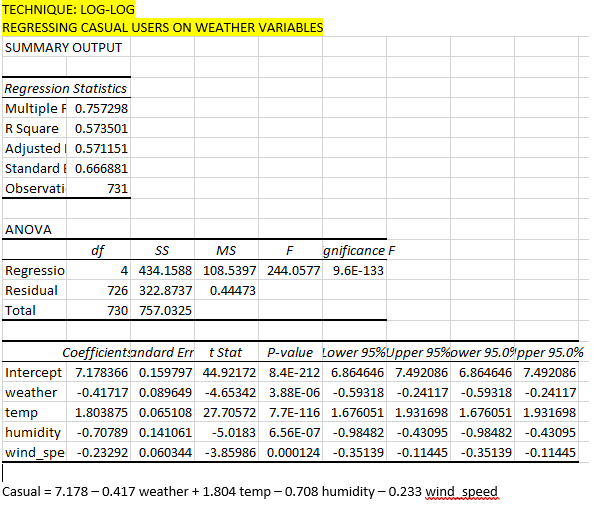
# Problem 2

* Problem Statement
  1. **How does the weather, temperature, humidity, and wind speed affect the number of bikes rented by casual customers and registered customers?**
  2. This question uses the independent discrete variable weather and the independent continuous variables temp, humidity and wind\_speedto observe if outside conditions and forecasts have a big impact on the number of bikes being rented, which are represented by the dependent continuous variables registered and casual.
  3. The significance of seeing if the weather and temperature affect the number of bikes rented helps us to figure out the disadvantages of bike transportation, predict future patterns in rental uses, and anticipate the number of bikes being utilized based on other/outside conditions.
* Background
  1. This research conducted in Washington DC bikeshare system analyzes the effect of weather on bike rentals. Variables included all trips made hourly including the weather, temperature and humidity levels (Gebhart & Noland, 2014). The authors used the dummy variable method for independent variables; weather, temperature and humidity to statistically analyze the data on the dependent variables; number of trips and average trip duration (Gebhart & Noland, 2014). The results showed that bike rentals and duration of trip are lower during cold temperatures, rainy weather and high humidity. This coincides with the question posed in problem 2 shows the role weather plays in bike sharing.
* Empirical Results and Discussion
  1. Regression Models

|  |  |
| --- | --- |
|  |  |
|  |  |

|  |  |
| --- | --- |
|  |  |





* 1. Dataset

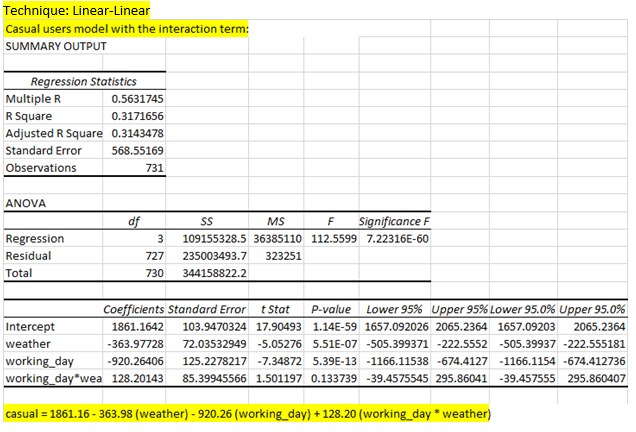
Dataset 2 attached. The variables we used from the data set include registered, casual, count, weather, temp, humidity and wind\_speed.

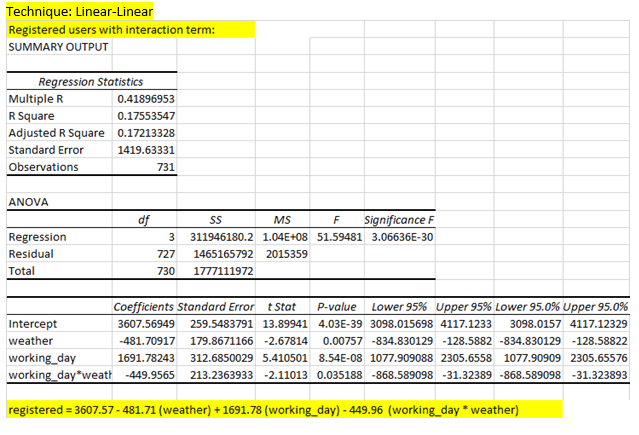
* 1. Results and Interpretation
* On comparing the four models above, we find that the log-log regression model has the highest values of adjusted R-squared for the casual and registered users. Also, the standard errors are lower in the log-log regression technique compared to others. Hence, for this problem, the log-log model is the best to use for analysis as adjusted R-squared is a good measure of fit.
* We find that weather, humidity and wind\_speed all significantly affect the number of casual as well as registered users negatively (since they have t-values less than -1.96) while temp affects them significantly positively (since it has t-values greater than 1.96).
* Temperature has the highest effect on both these users when keeping all other factors a constant, since ln(temp) coefficient has absolute values of 1.8 and 0.8 for casual users and registered users respectively, which are the highest among all the independent variables’ coefficients’ absolute values in each case.
* Thus, an increase in temp will increase the number of casual and registered users to a greater extent than the decrease caused in these numbers of users due to similar increases in wind\_speed, humidity or weather.
* Conclusion
  1. These results show that temperature has a higher impact than the other variables on the number of casual and registered users who rent bikes. If temperature doesn’t change appreciably, we consider the other factors. For casual users, apart from temperature, humidity is the more impactful, followed by weather then wind\_speed while for registered users, the more impactful is humidity, followed by wind\_speed then weather.
  2. The limitation of this model is that the unexplained differences in the influential levels between casual and registered users might be due to omitted variable bias present here.
  3. According to the results, we can recommend that the Capital Sharing System increase production of bikes at points where the temperature is higher and decrease production when the temperature is lower. Therefore, Spring and summer seasons should have more bikes available than fall and winter. These results is similar to the results in the research paper, however their results showed spring and fall were the best result.

**Problem 3**

* Problem Statement
  1. **Does the reason for the bike sharing system’s high demand on a working day or nonworking day decrease or increase the weather variable’s worth/importance?**
  2. For this question, we will determine if riding the bikes mainly to commute to work will make weather an insignificant factor, compared to if riding bikes for leisure activities will cause weather to be more significant. We use the binary independent variable working\_day, discrete independent variable weather, the interaction term between the two, and the continuous dependent variables casual and registered here.
  3. Considering both casual and registered customers to determine the importance of the weather variable, decline or incline, based on each type of customer’s use/purpose.
  4. The importance of this question is to observe the influence weather has on bike sharing. This information will help Capital Bike Sharing System make predictions and anticipate bikes utilization in Washington DC.
* Background
  1. The research was conducted on Bike Share Toronto in Canada which operates throughout the year. The authors conducted a regression analysis to show the correlation between temperature and bike sharing; the result showed to be significant (El-Assi, Salah & Nurul, 2017). Other variables were included such as land use (bike lanes, paths etc.) and week days.
  2. Regression analysis was performed on trip generation, trip attraction and station-to-station trips on the natural logarithm of trip counts (the dependent variable) (El-Assi, Salah & Nurul, 2017). They compared trip generation and trip attraction on weekdays to weekends. This research coincides with the question posed in problem 3, since it analyzes weather and working day on bike rentals. The distributed lag model was used since it reflects temporary changes in weather variation and other variables (El-Assi, Salah & Nurul, 2017). These models were estimated using the Ordinary Least Square (OLS) method, analyzed with the multivariable regression model variance inflation factor, multilevel mixed effects model and Maximum Likelihood method (El-Assi, Salah & Nurul, 2017). While, we used the nonlinear regression model; linear-linear.
* Empirical Results and Discussion
  1. Regression Models

We implement the regression models with the interaction term between weather and working\_day variable to determine the impact of weather on the number of casual and registered users given that it is a working day (or not).





* 1. Dataset

Dataset 3 attached. The variables we used from the data set include weather, working\_day, working\_day\*weather

* 1. Results and Interpretation:
* From the 1st regression model, we see that the interaction term (working\_day \* weather) is insignificant in the case of casual users and hence, we can say that the weather does not drastically affect the number of casual users using bikes based on whether it is a working day or not.
* For the 2nd regression model, the interaction term is significant for registered users and hence, we can say that the effect of weather is significantly different based on whether it is a working day or not. Since the interaction term’s coefficient has a significant negative value here, we can say that the number of registered users reduces significantly on a working day if the weather is bad (since worse weather is assigned larger numeric values).
* This is explainable from the results we obtained in problem 1 which showed us that the number of registered users is affected more by whether it is a working day or not and hence, the effect of weather on a working day or holiday will be more significant for registered users as obtained here.
* Conclusion
  1. Our results showed that weather has an impactful effect on registered customers when it is a working day. In the case of bad weather, we can conclude that the number of registered users reduces significantly if it is a working day.
  2. In conclusion, we recommend that the Capital Share System should reduce the number of bikes put out on working days when the weather is bad. The bad weather days include days in which it is snowing, raining, has high wind speed, or when it is extremely cold.