**Predict the New York City Subway Ridership**

**Overview**

This is my first project (out of five) of Udacity Data Analyst Nanodegree. In this project, I work on the NYC Subway data to draw an interesting conclusion about the dataset itself and especially to figure out if more people ride the subway when it is raining versus when it is not raining.

Data Science Skills Applied:

* Data wrangling
* Linear regression and statistical tests
* Effective visualization

**Section 0: References**

* Statistical Data Analysis in Python <https://github.com/fonnesbeck/statistical-analysis-python-tutorial>
* The Wikipedia page of Mann–Whitney U test <http://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test>
* The Wikipedia page of Skewness <http://en.wikipedia.org/wiki/Skewness>
* The Wikipedia page of Power transform <http://en.wikipedia.org/wiki/Power_transform#Box.E2.80.93Cox_transformation>
* One-tail vs. two-tail P values <http://graphpad.com/guides/prism/6/statistics/index.htm?one-tail_vs__two-tail_p_values.htm>

[Python for Data Analysis, By Wes McKinney, 2012](http://shop.oreilly.com/product/0636920023784.do)

**Section 1. Statistical Test**

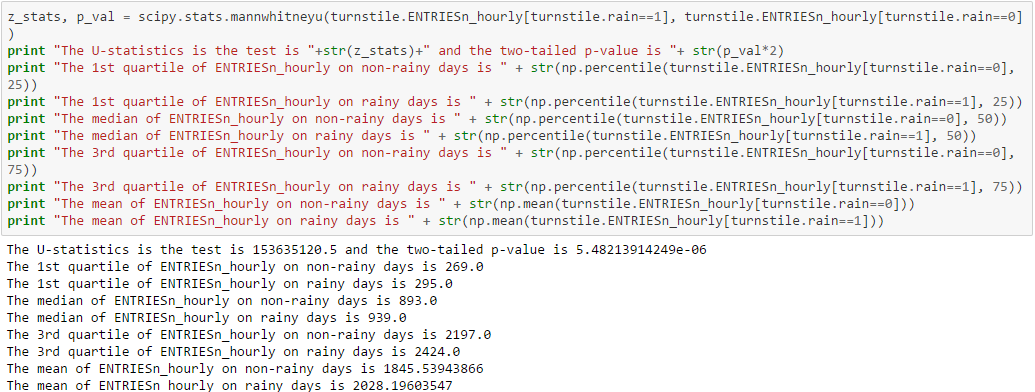
**1. Which statistical test did you use to analyze the NYC subway data? Did you use a one-tail or a two-tail P value? What is the null hypothesis? What is your p-critical value?**  
I used Mann–Whitney U test based on the observation of non-normal distribution of the data. The null hypothesis of the test is  
The distributions of ENTRIESn\_hourlyrainy and ENTRIESn\_hourlynon−rainy are the same,  
versus the alternative hypothesis:  
The distributions of ENTRIESn\_hourlyrainy and ENTRIESn\_hourlynon−rainy differ by a location shift μ and μ≠0  
Since the difference of riderships between rainy and non-rainy days can go both directions, the two-tailed P value is more appropriate here.

In inferential statistics on observational data, the **null hypothesis** refers to a general statement or default position that there is no relationship between two measured phenomena

**2. Why is this statistical test applicable to the dataset? In particular, consider the assumptions that the test is making about the distribution of ridership in the two samples**  
According to [the wikipedia page of Mann–Whitney U test](http://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test), Mann–Whitney U test is a non-parametric test that is good for testing whether a particular population tends to have larger values than the other. That is exactly what we are after.

**3. What results did you get from this statistical test? These should include the following numerical values: p-values, as well as the means for each of the two samples under test.**  
As printed below, the two-tailed p-value is 5.48213914249e-06. The mean of ENTRIESn\_hourly on non-rainy days and rainy days are 1845.53943866 and 2028.19603547 respectively.

**4. What is the significance and interpretation of these results?**  
Set the significance level α=0.05, the p-value is far less than α. So the result would be considered statistically significant and the null hypothesis would be rejected.



**Section 2. Linear Regression**

**1. What approach did you use to compute the coefficients theta and produce prediction for ENTRIESn\_hourly in your regression model:**  
**a. Gradient descent (as implemented in exercise 3.5)**  
**b. OLS using Statsmodels**  
**c. Or something different?**  
I used OLS from statsmodels.formula.api module. Also, in order to get the design matrix I imported dmatrices from pasty. (The details are shown below)

**2. What features (input variables) did you use in your model? Did you use any dummy variables as part of your features?**  
I included ENTRIESn, precipi, hour, tempi, UNIT, day\_week, pressurei, fog and conds in the model. Among those predictors, fog and conds are dummry variables. Also, three other categorical variables: hour, UNIT and day\_week. All of them would be presented as multiple dummy variables in the model even though that's not the way they how they present in the dataset.

**3.** **Why did you select these features in your model? We are looking for specific reasons that lead you to believe that the selected features will contribute to the predictive power of your model.**

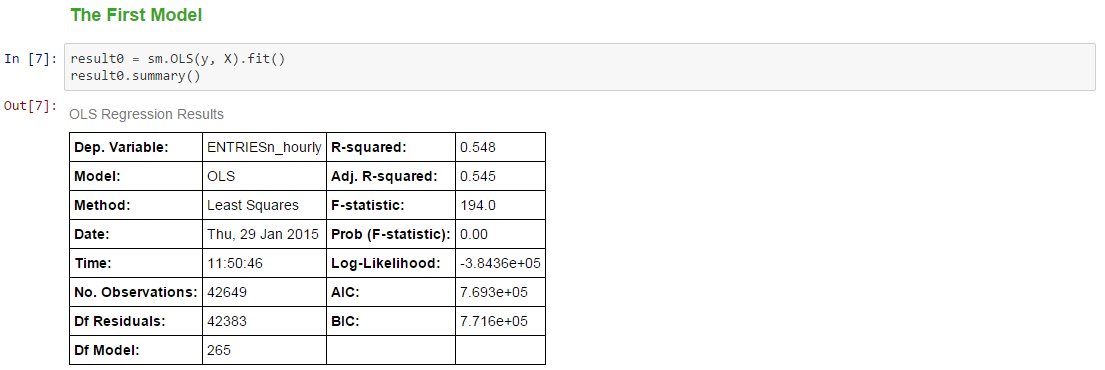
* **Your reasons might be based on intuition. For example, response for fog might be: “I decided to use fog because I thought that when it is very foggy outside people might decide to use the subway more often.**
* **Your reasons might also be based on data exploration and experimentation, for example: “I used feature X because as soon as I included it in my model, it drastically improved my R2 value.**

When I first started, I tried to include as many weather related variables as possible because I think people's choices of transportation highly depend on the condition of the weather.  
But some of the variables are highly correlated, such as pressurei and meanpressurei, precipi, meanprecipi and rain etc. To avoid the problems can be caused by multicollinearity, I decided to include only one of each group of the highly correlated variables. Since we try to predict the ENTRIESn\_hourly, I chose the hourly information of certain variable in each group rather than the daily mean value. It turned out all the coefficients of weather related variables I chose have the p-value less than 0.05, so I kept all of them.  
Also, I think different exits of one station and different stations see different volume of riders so I included UNIT in the model. And it turned out that it drastically improved the value of R2. But one thing we need to keep in mind is that the larger the number of explanatory is, the larger the value of R2. So the larger R2doesn't necessarily mean better model. But if the improvement is huge, I would include the newly-added variable.  
Since more people commute on weekdays than weekend, so I think the day\_of\_week information could also be useful to be included in the model.  
For the feature of EXITSn\_hourly, I originally included it into the model due to the improvement of R2 it contributed. But for more general use of the model, I decided to take it out. Since the information of EXITSn\_hourly won't be available when predict the ENTRIESn\_hourly accordingly.

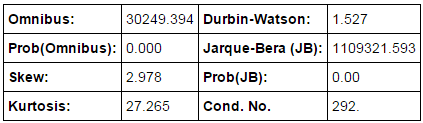
**4. What is your model’s**R2**(coefficients of determination) value?**  
R2 is 0.548 for the first model and is improved to 0.745 for the second and final model I fit.

**5. What does this**R2**value mean for the goodness of fit for your regression model?**  
**Do you think this linear model to predict ridership is appropriate for this dataset, given this**R2**value?**  
R2 is a measurement of how well the model fits the data, which ranges from 0 to 1, and 1 means the model fits the data perfectly. The first model has R2=0.548 and skewness=2.978. I don't think this simple linear regression fits the data well, so I applied Box-Cox transformation to the dependent variableENTIRES\_hourly and got the second model. In the output of the second model, the R2 was increased to 0.745 and the skewness was reduced to around −0.426. As far as I concerned, the linear regression model can be used to predict the ridership but some transformation should be added to the dependent variable or/and the explanatory variables.

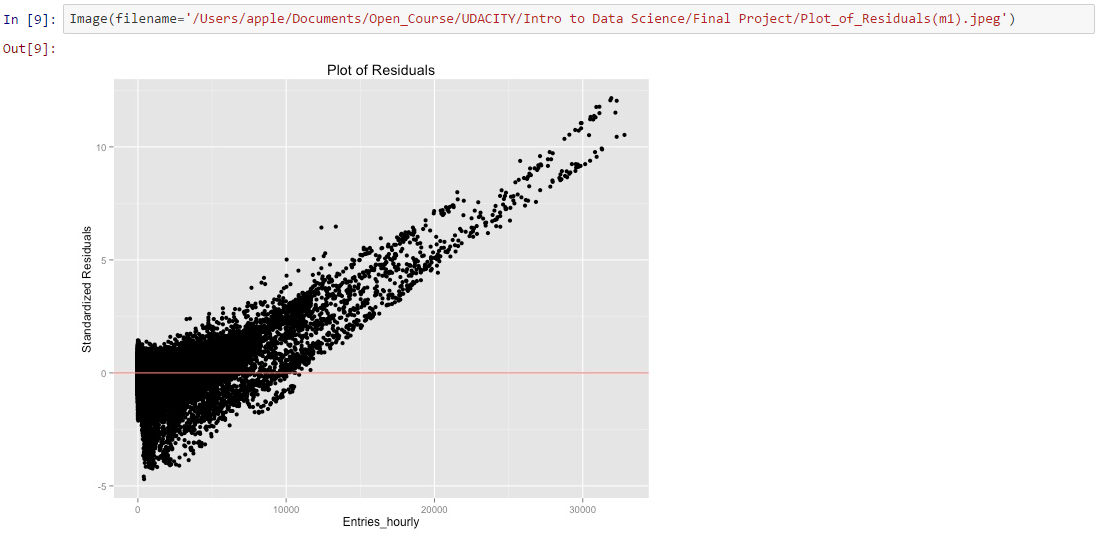




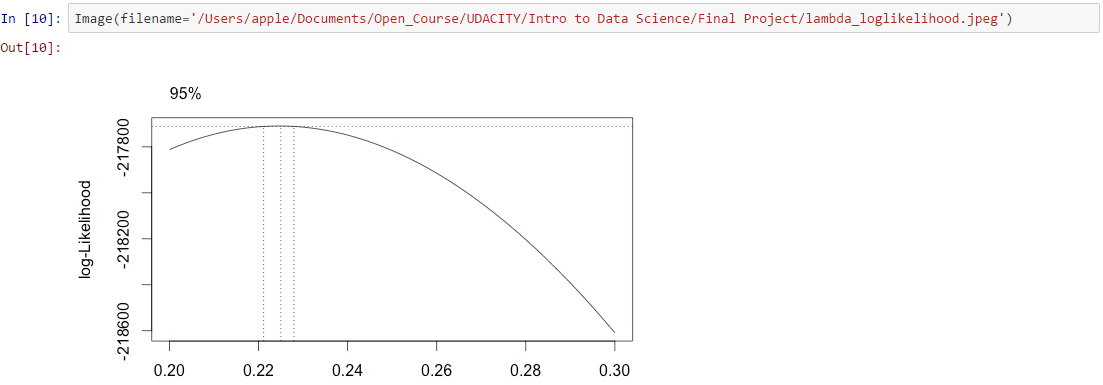
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[95.0% Conf. Int.]** |
| **Intercept** | **-200.8842** | **158.820** | **-1.265** | **0.206** | **-512.175 110.407** |
| **C(hour)[T.4]** | **-1131.9455** | **33.161** | **-34.135** | **0.000** | **-1196.942 -1066.949** |
| **C(hour)[T.8]** | **-790.7972** | **35.967** | **-21.987** | **0.000** | **-861.293 -720.301** |
| **C(hour)[T.12]** | **1762.4570** | **35.493** | **49.656** | **0.000** | **1692.890 1832.024** |
| **C(hour)[T.16]** | **1042.4766** | **34.942** | **29.835** | **0.000** | **973.990 1110.963** |
| **C(hour)[T.20]** | **1869.4146** | **33.368** | **56.024** | **0.000** | **1804.012 1934.817** |
| **UNIT[T.R004]** | **375.6526** | **215.123** | **1.746** | **0.081** | **-45.993 797.299** |
| **UNIT[T.R005]** | **367.5182** | **216.046** | **1.701** | **0.089** | **-55.937 790.974** |
| **UNIT[T.R006]** | **541.3942** | **213.667** | **2.534** | **0.011** | **122.603 960.186** |
| **UNIT[T.R007]** | **177.2457** | **216.666** | **0.818** | **0.413** | **-247.425 601.916** |
| **UNIT[T.R008]** | **173.0801** | **216.987** | **0.798** | **0.425** | **-252.220 598.380** |
| **UNIT[T.R009]** | **174.2035** | **215.121** | **0.810** | **0.418** | **-247.437 595.844** |
| **UNIT[T.R011]** | **7195.4456** | **212.850** | **33.805** | **0.000** | **6778.256 7612.635** |
| **UNIT[T.R012]** | **8547.2850** | **212.310** | **40.258** | **0.000** | **8131.153 8963.417** |
| **UNIT[T.R013]** | **2445.7259** | **212.310** | **11.520** | **0.000** | **2029.593 2861.858** |
| **UNIT[T.R016]** | **622.6446** | **212.847** | **2.925** | **0.003** | **205.460 1039.830** |
| **UNIT[T.R017]** | **4060.7205** | **212.310** | **19.126** | **0.000** | **3644.588 4476.853** |
| **UNIT[T.R018]** | **7695.2027** | **212.198** | **36.264** | **0.000** | **7279.291 8111.114** |
| **UNIT[T.R019]** | **3170.5001** | **211.994** | **14.956** | **0.000** | **2754.988 3586.013** |
| **UNIT[T.R020]** | **6236.8173** | **212.310** | **29.376** | **0.000** | **5820.685 6652.950** |
| **UNIT[T.R021]** | **4545.8963** | **212.854** | **21.357** | **0.000** | **4128.697 4963.095** |
| **UNIT[T.R022]** | **9381.2097** | **212.310** | **44.186** | **0.000** | **8965.077 9797.342** |
| **UNIT[T.R023]** | **6016.3872** | **212.310** | **28.338** | **0.000** | **5600.255 6432.520** |
| **UNIT[T.R024]** | **3130.9796** | **212.264** | **14.750** | **0.000** | **2714.939 3547.020** |
| **UNIT[T.R025]** | **5267.3926** | **211.994** | **24.847** | **0.000** | **4851.880 5682.905** |
| **UNIT[T.R027]** | **2830.8226** | **212.310** | **13.333** | **0.000** | **2414.690 3246.955** |
| **UNIT[T.R029]** | **7092.7958** | **212.310** | **33.408** | **0.000** | **6676.663 7508.928** |
| **UNIT[T.R030]** | **2962.9893** | **212.310** | **13.956** | **0.000** | **2546.857 3379.122** |
| **UNIT[T.R031]** | **4214.7581** | **212.310** | **19.852** | **0.000** | **3798.626 4630.891** |
| **UNIT[T.R032]** | **4309.7141** | **212.578** | **20.274** | **0.000** | **3893.058 4726.370** |
| **UNIT[T.R033]** | **8097.7689** | **212.310** | **38.141** | **0.000** | **7681.636 8513.901** |
| **UNIT[T.R034]** | **980.2317** | **216.981** | **4.518** | **0.000** | **554.945 1405.519** |
| **UNIT[T.R035]** | **2661.2248** | **212.846** | **12.503** | **0.000** | **2244.043 3078.406** |
| **UNIT[T.R036]** | **638.0853** | **214.460** | **2.975** | **0.003** | **217.740 1058.430** |
| **UNIT[T.R037]** | **761.4289** | **212.749** | **3.579** | **0.000** | **344.438 1178.420** |
| **UNIT[T.R038]** | **112.0226** | **215.058** | **0.521** | **0.602** | **-309.495 533.540** |
| **UNIT[T.R039]** | **623.6562** | **217.893** | **2.862** | **0.004** | **196.581 1050.731** |
| **UNIT[T.R040]** | **1163.9147** | **212.471** | **5.478** | **0.000** | **747.467 1580.362** |
| **UNIT[T.R041]** | **2962.8011** | **212.310** | **13.955** | **0.000** | **2546.669 3378.934** |
| **UNIT[T.R042]** | **446.5499** | **213.964** | **2.087** | **0.037** | **27.177 865.923** |
| **UNIT[T.R043]** | **2749.9355** | **212.310** | **12.952** | **0.000** | **2333.803 3166.068** |
| **UNIT[T.R044]** | **4541.3442** | **212.310** | **21.390** | **0.000** | **4125.212 4957.477** |
| **UNIT[T.R046]** | **8207.7958** | **212.310** | **38.659** | **0.000** | **7791.663 8623.928** |
| **UNIT[T.R049]** | **2635.6076** | **212.310** | **12.414** | **0.000** | **2219.475 3051.740** |
| **UNIT[T.R050]** | **3891.7282** | **212.856** | **18.283** | **0.000** | **3474.526 4308.930** |
| **UNIT[T.R051]** | **4997.3495** | **212.310** | **23.538** | **0.000** | **4581.217 5413.482** |
| **UNIT[T.R052]** | **1117.8135** | **215.974** | **5.176** | **0.000** | **694.500 1541.127** |
| **UNIT[T.R053]** | **3090.5975** | **212.471** | **14.546** | **0.000** | **2674.150 3507.045** |
| **UNIT[T.R054]** | **1322.4815** | **212.854** | **6.213** | **0.000** | **905.283 1739.680** |
| **UNIT[T.R055]** | **8254.3781** | **211.928** | **38.949** | **0.000** | **7838.995 8669.761** |
| **UNIT[T.R056]** | **1308.5108** | **212.850** | **6.148** | **0.000** | **891.321 1725.700** |
| **UNIT[T.R057]** | **4745.7366** | **212.310** | **22.353** | **0.000** | **4329.604 5161.869** |
| **UNIT[T.R058]** | **503.1240** | **212.577** | **2.367** | **0.018** | **86.469 919.779** |
| **UNIT[T.R059]** | **1043.5487** | **214.850** | **4.857** | **0.000** | **622.438 1464.659** |
| **UNIT[T.R060]** | **650.8217** | **213.967** | **3.042** | **0.002** | **231.442 1070.202** |
| **UNIT[T.R061]** | **445.5578** | **217.940** | **2.044** | **0.041** | **18.391 872.725** |
| **UNIT[T.R062]** | **2599.5968** | **212.310** | **12.244** | **0.000** | **2183.464 3015.729** |
| **UNIT[T.R063]** | **992.1186** | **217.617** | **4.559** | **0.000** | **565.585 1418.652** |
| **UNIT[T.R064]** | **698.9402** | **215.128** | **3.249** | **0.001** | **277.284 1120.596** |
| **UNIT[T.R065]** | **693.2507** | **216.343** | **3.204** | **0.001** | **269.214 1117.288** |
| **UNIT[T.R066]** | **88.3618** | **216.960** | **0.407** | **0.684** | **-336.885 513.608** |
| **UNIT[T.R067]** | **764.0117** | **217.562** | **3.512** | **0.000** | **337.585 1190.438** |
| **UNIT[T.R068]** | **343.5447** | **217.564** | **1.579** | **0.114** | **-82.884 769.974** |
| **UNIT[T.R069]** | **840.2987** | **215.664** | **3.896** | **0.000** | **417.594 1263.004** |
| **UNIT[T.R070]** | **1650.5753** | **212.310** | **7.774** | **0.000** | **1234.443 2066.708** |
| **UNIT[T.R080]** | **3474.9033** | **212.310** | **16.367** | **0.000** | **3058.771 3891.036** |
| **UNIT[T.R081]** | **3426.7653** | **212.851** | **16.099** | **0.000** | **3009.572 3843.958** |
| **UNIT[T.R082]** | **1365.0878** | **212.856** | **6.413** | **0.000** | **947.886 1782.289** |
| **UNIT[T.R083]** | **2989.3119** | **212.310** | **14.080** | **0.000** | **2573.179 3405.444** |
| **UNIT[T.R084]** | **9893.5861** | **212.310** | **46.600** | **0.000** | **9477.454 1.03e+04** |
| **UNIT[T.R085]** | **2473.3386** | **212.858** | **11.620** | **0.000** | **2056.133 2890.545** |
| **UNIT[T.R086]** | **2462.9033** | **212.310** | **11.600** | **0.000** | **2046.771 2879.036** |
| **UNIT[T.R087]** | **1090.3806** | **213.405** | **5.109** | **0.000** | **672.103 1508.659** |
| **UNIT[T.R089]** | **376.4327** | **212.847** | **1.769** | **0.077** | **-40.752 793.618** |
| **UNIT[T.R090]** | **330.0599** | **217.566** | **1.517** | **0.129** | **-96.373 756.493** |
| **UNIT[T.R091]** | **1035.2047** | **216.599** | **4.779** | **0.000** | **610.666 1459.743** |
| **UNIT[T.R092]** | **1917.9085** | **214.763** | **8.930** | **0.000** | **1496.969 2338.848** |
| **UNIT[T.R093]** | **1949.4127** | **215.361** | **9.052** | **0.000** | **1527.300 2371.525** |
| **UNIT[T.R094]** | **1691.2597** | **212.471** | **7.960** | **0.000** | **1274.811 2107.708** |
| **UNIT[T.R095]** | **2100.3448** | **213.309** | **9.847** | **0.000** | **1682.256 2518.434** |
| **UNIT[T.R096]** | **2281.1654** | **212.198** | **10.750** | **0.000** | **1865.253 2697.078** |
| **UNIT[T.R097]** | **2904.2405** | **212.198** | **13.686** | **0.000** | **2488.329 3320.152** |
| **UNIT[T.R098]** | **1697.4624** | **212.310** | **7.995** | **0.000** | **1281.330 2113.595** |
| **UNIT[T.R099]** | **2254.4893** | **212.310** | **10.619** | **0.000** | **1838.357 2670.622** |
| **UNIT[T.R100]** | **485.0711** | **213.030** | **2.277** | **0.023** | **67.528 902.614** |
| **UNIT[T.R101]** | **2688.8818** | **212.310** | **12.665** | **0.000** | **2272.749 3105.014** |
| **UNIT[T.R102]** | **3577.7205** | **212.310** | **16.851** | **0.000** | **3161.588 3993.853** |
| **UNIT[T.R103]** | **1313.6490** | **215.973** | **6.082** | **0.000** | **890.337 1736.961** |
| **UNIT[T.R104]** | **1277.5786** | **212.471** | **6.013** | **0.000** | **861.131 1694.027** |
| **UNIT[T.R105]** | **3227.4194** | **212.310** | **15.201** | **0.000** | **2811.287 3643.552** |
| **UNIT[T.R106]** | **984.2341** | **217.562** | **4.524** | **0.000** | **557.809 1410.660** |
| **UNIT[T.R107]** | **391.3213** | **217.893** | **1.796** | **0.073** | **-35.754 818.396** |
| **UNIT[T.R108]** | **5116.5861** | **212.310** | **24.100** | **0.000** | **4700.454 5532.718** |
| **UNIT[T.R111]** | **3113.7097** | **212.310** | **14.666** | **0.000** | **2697.577 3529.842** |
| **UNIT[T.R112]** | **1622.8519** | **212.198** | **7.648** | **0.000** | **1206.939 2038.765** |
| **UNIT[T.R114]** | **845.7473** | **212.265** | **3.984** | **0.000** | **429.703 1261.791** |
| **UNIT[T.R115]** | **1217.9894** | **211.994** | **5.745** | **0.000** | **802.477 1633.502** |
| **UNIT[T.R116]** | **3092.8549** | **212.310** | **14.568** | **0.000** | **2676.723 3508.987** |
| **UNIT[T.R117]** | **800.0034** | **217.239** | **3.683** | **0.000** | **374.211 1225.796** |
| **UNIT[T.R119]** | **1787.1329** | **213.876** | **8.356** | **0.000** | **1367.931 2206.334** |
| **UNIT[T.R120]** | **1444.2900** | **215.358** | **6.706** | **0.000** | **1022.183 1866.397** |
| **UNIT[T.R121]** | **1349.1933** | **215.134** | **6.271** | **0.000** | **927.527 1770.860** |
| **UNIT[T.R122]** | **2516.4348** | **213.313** | **11.797** | **0.000** | **2098.337 2934.533** |
| **UNIT[T.R123]** | **1516.0927** | **213.967** | **7.086** | **0.000** | **1096.713 1935.472** |
| **UNIT[T.R124]** | **526.2671** | **216.654** | **2.429** | **0.015** | **101.620 950.914** |
| **UNIT[T.R126]** | **1754.7635** | **212.310** | **8.265** | **0.000** | **1338.631 2170.896** |
| **UNIT[T.R127]** | **4695.4732** | **212.310** | **22.116** | **0.000** | **4279.341 5111.606** |
| **UNIT[T.R137]** | **2400.2867** | **211.928** | **11.326** | **0.000** | **1984.904 2815.670** |
| **UNIT[T.R139]** | **2429.2339** | **212.578** | **11.427** | **0.000** | **2012.577 2845.891** |
| **UNIT[T.R163]** | **3231.2097** | **212.310** | **15.219** | **0.000** | **2815.077 3647.342** |
| **UNIT[T.R172]** | **1797.2474** | **212.310** | **8.465** | **0.000** | **1381.115 2213.380** |
| **UNIT[T.R179]** | **6678.0216** | **212.310** | **31.454** | **0.000** | **6261.889 7094.154** |
| **UNIT[T.R181]** | **1655.5474** | **214.260** | **7.727** | **0.000** | **1235.593 2075.502** |
| **UNIT[T.R183]** | **681.1718** | **217.894** | **3.126** | **0.002** | **254.096 1108.248** |
| **UNIT[T.R184]** | **871.8202** | **216.972** | **4.018** | **0.000** | **446.550 1297.090** |
| **UNIT[T.R186]** | **975.9048** | **213.693** | **4.567** | **0.000** | **557.061 1394.748** |
| **UNIT[T.R188]** | **2231.4168** | **212.578** | **10.497** | **0.000** | **1814.761 2648.073** |
| **UNIT[T.R189]** | **1290.0428** | **214.843** | **6.005** | **0.000** | **868.946 1711.140** |
| **UNIT[T.R194]** | **1893.9229** | **214.249** | **8.840** | **0.000** | **1473.991 2313.855** |
| **UNIT[T.R196]** | **1228.9639** | **213.131** | **5.766** | **0.000** | **811.223 1646.705** |
| **UNIT[T.R198]** | **2013.4380** | **212.855** | **9.459** | **0.000** | **1596.239 2430.637** |
| **UNIT[T.R199]** | **616.8865** | **214.256** | **2.879** | **0.004** | **196.940 1036.833** |
| **UNIT[T.R200]** | **1015.4243** | **213.094** | **4.765** | **0.000** | **597.756 1433.092** |
| **UNIT[T.R202]** | **2188.6368** | **212.474** | **10.301** | **0.000** | **1772.184 2605.089** |
| **UNIT[T.R203]** | **1661.4764** | **215.435** | **7.712** | **0.000** | **1239.219 2083.734** |
| **UNIT[T.R204]** | **1358.7474** | **212.310** | **6.400** | **0.000** | **942.615 1774.880** |
| **UNIT[T.R205]** | **1464.6334** | **213.026** | **6.875** | **0.000** | **1047.098 1882.169** |
| **UNIT[T.R207]** | **1907.0942** | **212.586** | **8.971** | **0.000** | **1490.422 2323.766** |
| **UNIT[T.R208]** | **2468.7554** | **213.026** | **11.589** | **0.000** | **2051.220 2886.291** |
| **UNIT[T.R209]** | **671.4965** | **217.613** | **3.086** | **0.002** | **244.971 1098.022** |
| **UNIT[T.R210]** | **432.3620** | **214.846** | **2.012** | **0.044** | **11.260 853.464** |
| **UNIT[T.R211]** | **2316.1775** | **212.310** | **10.909** | **0.000** | **1900.045 2732.310** |
| **UNIT[T.R212]** | **1596.9424** | **212.577** | **7.512** | **0.000** | **1180.288 2013.597** |
| **UNIT[T.R213]** | **1052.2149** | **214.552** | **4.904** | **0.000** | **631.690 1472.740** |
| **UNIT[T.R214]** | **543.0353** | **217.944** | **2.492** | **0.013** | **115.861 970.209** |
| **UNIT[T.R215]** | **1503.3707** | **212.858** | **7.063** | **0.000** | **1086.166 1920.576** |
| **UNIT[T.R216]** | **675.6479** | **212.851** | **3.174** | **0.002** | **258.456 1092.840** |
| **UNIT[T.R217]** | **880.8674** | **217.593** | **4.048** | **0.000** | **454.380 1307.355** |
| **UNIT[T.R218]** | **1946.4051** | **212.199** | **9.173** | **0.000** | **1530.492 2362.319** |
| **UNIT[T.R219]** | **1228.6045** | **212.473** | **5.782** | **0.000** | **812.153 1645.056** |
| **UNIT[T.R220]** | **1390.9033** | **212.310** | **6.551** | **0.000** | **974.771 1807.036** |
| **UNIT[T.R221]** | **1290.2269** | **217.939** | **5.920** | **0.000** | **863.063 1717.391** |
| **UNIT[T.R223]** | **2085.1114** | **212.198** | **9.826** | **0.000** | **1669.198 2501.024** |
| **UNIT[T.R224]** | **629.2467** | **215.128** | **2.925** | **0.003** | **207.591 1050.903** |
| **UNIT[T.R225]** | **482.2256** | **213.694** | **2.257** | **0.024** | **63.381 901.070** |
| **UNIT[T.R226]** | **658.7450** | **216.599** | **3.041** | **0.002** | **234.207 1083.283** |
| **UNIT[T.R227]** | **1010.1667** | **212.310** | **4.758** | **0.000** | **594.034 1426.299** |
| **UNIT[T.R228]** | **1020.0371** | **217.298** | **4.694** | **0.000** | **594.128 1445.946** |
| **UNIT[T.R229]** | **458.0839** | **216.652** | **2.114** | **0.034** | **33.442 882.725** |
| **UNIT[T.R230]** | **414.6042** | **215.436** | **1.924** | **0.054** | **-7.654 836.863** |
| **UNIT[T.R231]** | **874.8830** | **213.689** | **4.094** | **0.000** | **456.049 1293.717** |
| **UNIT[T.R232]** | **912.5408** | **216.345** | **4.218** | **0.000** | **488.500 1336.582** |
| **UNIT[T.R233]** | **979.1672** | **218.595** | **4.479** | **0.000** | **550.717 1407.617** |
| **UNIT[T.R234]** | **187.5643** | **217.614** | **0.862** | **0.389** | **-238.963 614.092** |
| **UNIT[T.R235]** | **2481.2267** | **212.577** | **11.672** | **0.000** | **2064.572 2897.882** |
| **UNIT[T.R236]** | **1482.1863** | **213.024** | **6.958** | **0.000** | **1064.654 1899.718** |
| **UNIT[T.R237]** | **591.0371** | **216.917** | **2.725** | **0.006** | **165.874 1016.200** |
| **UNIT[T.R238]** | **2069.0831** | **212.199** | **9.751** | **0.000** | **1653.169 2484.997** |
| **UNIT[T.R239]** | **807.3280** | **212.310** | **3.803** | **0.000** | **391.196 1223.460** |
| **UNIT[T.R240]** | **2566.6070** | **213.132** | **12.042** | **0.000** | **2148.864 2984.350** |
| **UNIT[T.R242]** | **449.6953** | **214.540** | **2.096** | **0.036** | **29.193 870.198** |
| **UNIT[T.R243]** | **1322.4704** | **215.664** | **6.132** | **0.000** | **899.764 1745.177** |
| **UNIT[T.R244]** | **1563.2459** | **215.673** | **7.248** | **0.000** | **1140.523 1985.969** |
| **UNIT[T.R246]** | **557.9131** | **217.318** | **2.567** | **0.010** | **131.966 983.860** |
| **UNIT[T.R247]** | **16.7267** | **219.264** | **0.076** | **0.939** | **-413.034 446.488** |
| **UNIT[T.R248]** | **3011.5636** | **212.578** | **14.167** | **0.000** | **2594.906 3428.221** |
| **UNIT[T.R249]** | **1263.2812** | **213.962** | **5.904** | **0.000** | **843.912 1682.650** |
| **UNIT[T.R250]** | **873.1428** | **214.551** | **4.070** | **0.000** | **452.618 1293.668** |
| **UNIT[T.R251]** | **1182.9546** | **213.122** | **5.551** | **0.000** | **765.232 1600.677** |
| **UNIT[T.R252]** | **865.9115** | **212.856** | **4.068** | **0.000** | **448.710 1283.113** |
| **UNIT[T.R253]** | **695.9316** | **215.985** | **3.222** | **0.001** | **272.597 1119.267** |
| **UNIT[T.R254]** | **2555.3395** | **212.200** | **12.042** | **0.000** | **2139.423 2971.256** |
| **UNIT[T.R255]** | **732.9026** | **213.091** | **3.439** | **0.001** | **315.241 1150.565** |
| **UNIT[T.R256]** | **970.6468** | **212.860** | **4.560** | **0.000** | **553.436 1387.857** |
| **UNIT[T.R257]** | **1794.5001** | **212.310** | **8.452** | **0.000** | **1378.368 2210.632** |
| **UNIT[T.R258]** | **1430.7036** | **213.131** | **6.713** | **0.000** | **1012.962 1848.445** |
| **UNIT[T.R259]** | **831.6052** | **213.982** | **3.886** | **0.000** | **412.197 1251.014** |
| **UNIT[T.R260]** | **992.0162** | **221.714** | **4.474** | **0.000** | **557.453 1426.579** |
| **UNIT[T.R261]** | **1512.0778** | **214.461** | **7.051** | **0.000** | **1091.729 1932.426** |
| **UNIT[T.R262]** | **321.8208** | **219.273** | **1.468** | **0.142** | **-107.958 751.599** |
| **UNIT[T.R263]** | **33.6215** | **214.565** | **0.157** | **0.875** | **-386.930 454.173** |
| **UNIT[T.R264]** | **380.7876** | **212.588** | **1.791** | **0.073** | **-35.889 797.464** |
| **UNIT[T.R265]** | **748.1496** | **216.650** | **3.453** | **0.001** | **323.511 1172.788** |
| **UNIT[T.R266]** | **867.5987** | **212.198** | **4.089** | **0.000** | **451.686 1283.512** |
| **UNIT[T.R269]** | **786.5132** | **212.853** | **3.695** | **0.000** | **369.317 1203.710** |
| **UNIT[T.R270]** | **366.6593** | **217.307** | **1.687** | **0.092** | **-59.267 792.585** |
| **UNIT[T.R271]** | **249.9518** | **216.985** | **1.152** | **0.249** | **-175.344 675.247** |
| **UNIT[T.R273]** | **1250.1405** | **217.641** | **5.744** | **0.000** | **823.561 1676.720** |
| **UNIT[T.R274]** | **843.0347** | **216.357** | **3.897** | **0.000** | **418.971 1267.098** |
| **UNIT[T.R275]** | **847.4550** | **214.463** | **3.952** | **0.000** | **427.104 1267.806** |
| **UNIT[T.R276]** | **1317.0377** | **212.310** | **6.203** | **0.000** | **900.905 1733.170** |
| **UNIT[T.R277]** | **293.2596** | **220.298** | **1.331** | **0.183** | **-138.529 725.048** |
| **UNIT[T.R278]** | **253.2720** | **216.662** | **1.169** | **0.242** | **-171.389 677.933** |
| **UNIT[T.R279]** | **711.5016** | **213.876** | **3.327** | **0.001** | **292.300 1130.704** |
| **UNIT[T.R280]** | **521.2291** | **219.616** | **2.373** | **0.018** | **90.778 951.680** |
| **UNIT[T.R281]** | **1182.0846** | **213.971** | **5.525** | **0.000** | **762.698 1601.472** |
| **UNIT[T.R282]** | **1486.6409** | **212.854** | **6.984** | **0.000** | **1069.442 1903.840** |
| **UNIT[T.R284]** | **674.3082** | **212.856** | **3.168** | **0.002** | **257.107 1091.510** |
| **UNIT[T.R285]** | **517.1633** | **217.647** | **2.376** | **0.017** | **90.570 943.756** |
| **UNIT[T.R287]** | **410.7664** | **218.269** | **1.882** | **0.060** | **-17.046 838.579** |
| **UNIT[T.R291]** | **1710.2689** | **212.310** | **8.056** | **0.000** | **1294.136 2126.401** |
| **UNIT[T.R294]** | **830.8731** | **214.845** | **3.867** | **0.000** | **409.773 1251.973** |
| **UNIT[T.R295]** | **538.9171** | **228.760** | **2.356** | **0.018** | **90.542 987.292** |
| **UNIT[T.R300]** | **2119.3334** | **212.310** | **9.982** | **0.000** | **1703.201 2535.466** |
| **UNIT[T.R303]** | **1127.6864** | **213.968** | **5.270** | **0.000** | **708.305 1547.067** |
| **UNIT[T.R304]** | **984.9062** | **212.856** | **4.627** | **0.000** | **567.704 1402.108** |
| **UNIT[T.R307]** | **244.4578** | **217.892** | **1.122** | **0.262** | **-182.615 671.531** |
| **UNIT[T.R308]** | **719.0975** | **215.663** | **3.334** | **0.001** | **296.393 1141.802** |
| **UNIT[T.R309]** | **755.3874** | **215.361** | **3.508** | **0.000** | **333.276 1177.498** |
| **UNIT[T.R310]** | **1205.3853** | **216.599** | **5.565** | **0.000** | **780.847 1629.923** |
| **UNIT[T.R311]** | **275.9163** | **215.435** | **1.281** | **0.200** | **-146.341 698.174** |
| **UNIT[T.R312]** | **225.7917** | **213.126** | **1.059** | **0.289** | **-191.940 643.524** |
| **UNIT[T.R313]** | **-55.7180** | **218.221** | **-0.255** | **0.798** | **-483.435 371.999** |
| **UNIT[T.R318]** | **433.4214** | **213.404** | **2.031** | **0.042** | **15.145 851.698** |
| **UNIT[T.R319]** | **1235.7679** | **214.251** | **5.768** | **0.000** | **815.832 1655.704** |
| **UNIT[T.R321]** | **981.9302** | **212.310** | **4.625** | **0.000** | **565.798 1398.063** |
| **UNIT[T.R322]** | **1663.1797** | **214.256** | **7.763** | **0.000** | **1243.234 2083.126** |
| **UNIT[T.R323]** | **1136.3434** | **216.332** | **5.253** | **0.000** | **712.329 1560.358** |
| **UNIT[T.R325]** | **263.9566** | **215.056** | **1.227** | **0.220** | **-157.558 685.471** |
| **UNIT[T.R330]** | **867.0443** | **214.846** | **4.036** | **0.000** | **445.942 1288.146** |
| **UNIT[T.R335]** | **272.0412** | **218.640** | **1.244** | **0.213** | **-156.497 700.579** |
| **UNIT[T.R336]** | **-81.5989** | **218.608** | **-0.373** | **0.709** | **-510.074 346.876** |
| **UNIT[T.R337]** | **-28.3027** | **217.313** | **-0.130** | **0.896** | **-454.241 397.636** |
| **UNIT[T.R338]** | **-148.2565** | **215.429** | **-0.688** | **0.491** | **-570.503 273.990** |
| **UNIT[T.R341]** | **445.2449** | **212.820** | **2.092** | **0.036** | **28.114 862.375** |
| **UNIT[T.R344]** | **318.9827** | **218.964** | **1.457** | **0.145** | **-110.191 748.156** |
| **UNIT[T.R345]** | **332.6241** | **215.444** | **1.544** | **0.123** | **-89.650 754.898** |
| **UNIT[T.R346]** | **1117.3419** | **215.748** | **5.179** | **0.000** | **694.471 1540.213** |
| **UNIT[T.R348]** | **33.5851** | **215.356** | **0.156** | **0.876** | **-388.517 455.688** |
| **UNIT[T.R354]** | **84.1236** | **217.644** | **0.387** | **0.699** | **-342.462 510.710** |
| **UNIT[T.R356]** | **1022.2773** | **214.835** | **4.758** | **0.000** | **601.197 1443.358** |
| **UNIT[T.R358]** | **114.2985** | **217.639** | **0.525** | **0.599** | **-312.279 540.876** |
| **UNIT[T.R370]** | **339.7682** | **215.120** | **1.579** | **0.114** | **-81.871 761.408** |
| **UNIT[T.R371]** | **555.1715** | **216.340** | **2.566** | **0.010** | **131.141 979.202** |
| **UNIT[T.R372]** | **541.3490** | **217.282** | **2.491** | **0.013** | **115.472 967.226** |
| **UNIT[T.R373]** | **486.8968** | **217.645** | **2.237** | **0.025** | **60.309 913.485** |
| **UNIT[T.R382]** | **816.3804** | **214.531** | **3.805** | **0.000** | **395.896 1236.865** |
| **UNIT[T.R424]** | **187.6872** | **218.594** | **0.859** | **0.391** | **-240.762 616.137** |
| **UNIT[T.R429]** | **928.7969** | **212.747** | **4.366** | **0.000** | **511.808 1345.786** |
| **UNIT[T.R453]** | **1626.2537** | **219.625** | **7.405** | **0.000** | **1195.784 2056.723** |
| **UNIT[T.R454]** | **-55.2780** | **217.931** | **-0.254** | **0.800** | **-482.427 371.871** |
| **UNIT[T.R455]** | **-113.1488** | **217.643** | **-0.520** | **0.603** | **-539.733 313.435** |
| **UNIT[T.R456]** | **77.1825** | **214.853** | **0.359** | **0.719** | **-343.934 498.299** |
| **UNIT[T.R459]** | **-36.5929** | **262.117** | **-0.140** | **0.889** | **-550.348 477.162** |
| **UNIT[T.R464]** | **-225.5016** | **216.987** | **-1.039** | **0.299** | **-650.801 199.798** |
| **C(day\_week)[T.1]** | **355.3942** | **35.346** | **10.055** | **0.000** | **286.115 424.674** |
| **C(day\_week)[T.2]** | **379.9457** | **36.782** | **10.330** | **0.000** | **307.852 452.039** |
| **C(day\_week)[T.3]** | **359.4636** | **37.832** | **9.501** | **0.000** | **285.312 433.616** |
| **C(day\_week)[T.4]** | **313.6387** | **37.461** | **8.372** | **0.000** | **240.215 387.062** |
| **C(day\_week)[T.5]** | **-552.2097** | **37.491** | **-14.729** | **0.000** | **-625.693 -478.726** |
| **C(day\_week)[T.6]** | **-881.5418** | **34.782** | **-25.345** | **0.000** | **-949.716 -813.368** |
| **C(fog)[T.1]** | **-852.4608** | **181.680** | **-4.692** | **0.000** | **-1208.557 -496.365** |
| **C(conds)[T.Fog]** | **1629.9286** | **347.046** | **4.697** | **0.000** | **949.711 2310.146** |
| **C(conds)[T.Haze]** | **142.7717** | **60.740** | **2.351** | **0.019** | **23.721 261.823** |
| **C(conds)[T.Heavy Rain]** | **972.6288** | **226.466** | **4.295** | **0.000** | **528.751 1416.506** |
| **C(conds)[T.Light Drizzle]** | **-575.3013** | **117.056** | **-4.915** | **0.000** | **-804.734 -345.869** |
| **C(conds)[T.Light Rain]** | **-12.2641** | **56.665** | **-0.216** | **0.829** | **-123.329 98.800** |
| **C(conds)[T.Mist]** | **1625.7737** | **409.538** | **3.970** | **0.000** | **823.071 2428.477** |
| **C(conds)[T.Mostly Cloudy]** | **-327.7625** | **34.326** | **-9.549** | **0.000** | **-395.042 -260.483** |
| **C(conds)[T.Overcast]** | **-195.8482** | **29.399** | **-6.662** | **0.000** | **-253.470 -138.226** |
| **C(conds)[T.Partly Cloudy]** | **-33.1191** | **47.879** | **-0.692** | **0.489** | **-126.963 60.725** |
| **C(conds)[T.Rain]** | **291.0285** | **128.296** | **2.268** | **0.023** | **39.566 542.491** |
| **C(conds)[T.Scattered Clouds]** | **-44.3321** | **42.191** | **-1.051** | **0.293** | **-127.028 38.364** |
| **precipi** | **-103.0038** | **19.596** | **-5.256** | **0.000** | **-141.412 -64.596** |
| **tempi** | **-166.5737** | **12.788** | **-13.026** | **0.000** | **-191.637 -141.510** |
| **pressurei** | **-56.8553** | **10.571** | **-5.378** | **0.000** | **-77.576 -36.135** |



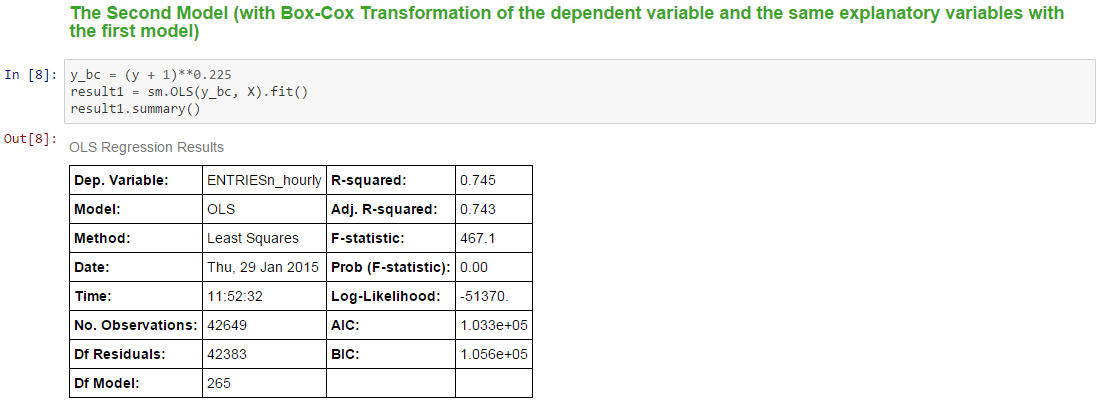
I was not impressed by with the linear regression models I got after adding/removing all the possible explanatory variables. Then I normalized all the numerical variables since the magnitude for different variables vary a lot. But 0.548 was the highest R2 value I got among all the reasonable models. So I used R(which I'm more familiar with) and examined the plot of the dependent variable ENTRIES\_hourly versus residuals and found a linear trend there.



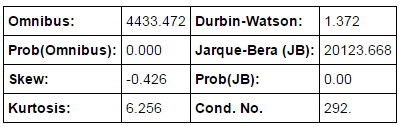
Then I decided to apply Box-Cox transformation to the dependent variable, hoping to fix the problem of non-normal distributed residuals from the previous model. Again, I used R to find the optimal λ value for the transformation.  
Box-Cox transformation is a common type of Power transforms, which is a useful data transformation technique used to stabilize variance, make the data more normal distribution-like, improve the validity of measures of association such as the Pearson correlation between variables and for other data stabilization procedures.



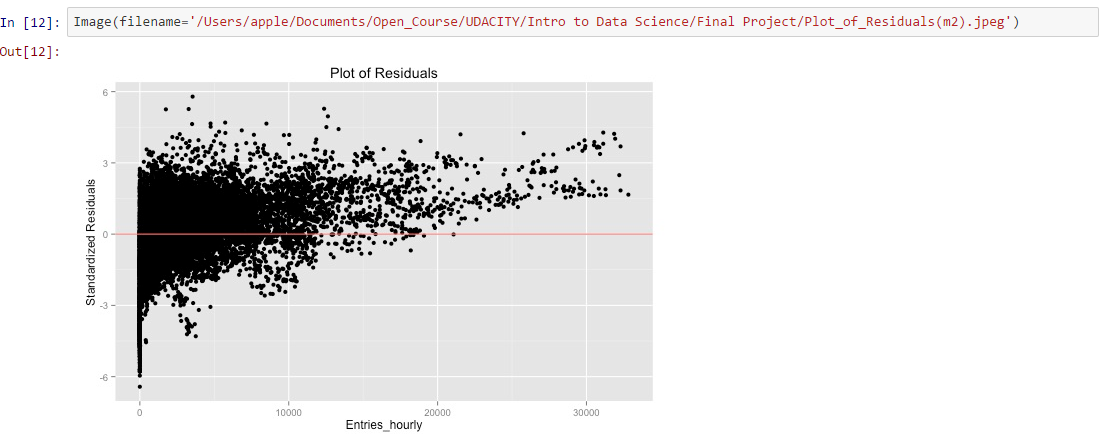
From the log-likelihood graph above, we can see the log-likelihood was maximized when λ=0.225. So I decided to fit another linear regression model with the all the same explanatory variables but Box-Cox transformed dependent variable of λ=0.225.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **coef** | **std err** | **t** | **P>|t|** | **[95.0% Conf. Int.]** |  |
| **Intercept** | **2.8643** | **0.065** | **44.357** | **0** | **2.738 2.991** |
| **C(hour)[T.4]** | **-1.4487** | **0.013** | **-107.445** | **0** | **-2.897** |
| **C(hour)[T.8]** | **-0.7079** | **0.015** | **-48.406** | **0** | **-1.416** |
| **C(hour)[T.12]** | **1.1735** | **0.014** | **81.321** | **0** | **1.145 1.202** |
| **C(hour)[T.16]** | **0.8245** | **0.014** | **58.039** | **0** | **0.797 0.852** |
| **C(hour)[T.20]** | **0.9429** | **0.014** | **69.5** | **0** | **0.916 0.969** |
| **UNIT[T.R004]** | **0.6665** | **0.087** | **7.62** | **0** | **0.495 0.838** |
| **UNIT[T.R005]** | **0.6292** | **0.088** | **7.163** | **0** | **0.457 0.801** |
| **UNIT[T.R006]** | **0.8978** | **0.087** | **10.335** | **0** | **0.728 1.068** |
| **UNIT[T.R007]** | **0.3665** | **0.088** | **4.16** | **0** | **0.194 0.539** |
| **UNIT[T.R008]** | **0.3839** | **0.088** | **4.352** | **0** | **0.211 0.557** |
| **UNIT[T.R009]** | **0.3947** | **0.087** | **4.513** | **0** | **0.223 0.566** |
| **UNIT[T.R011]** | **3.8899** | **0.087** | **44.949** | **0** | **3.720 4.060** |
| **UNIT[T.R012]** | **4.2805** | **0.086** | **49.587** | **0** | **4.111 4.450** |
| **UNIT[T.R013]** | **2.5159** | **0.086** | **29.145** | **0** | **2.347 2.685** |
| **UNIT[T.R016]** | **0.9832** | **0.087** | **11.362** | **0** | **0.814 1.153** |
| **UNIT[T.R017]** | **2.8683** | **0.086** | **33.227** | **0** | **2.699 3.037** |
| **UNIT[T.R018]** | **4.1246** | **0.086** | **47.807** | **0** | **3.956 4.294** |
| **UNIT[T.R019]** | **2.7133** | **0.086** | **31.479** | **0** | **2.544 2.882** |
| **UNIT[T.R020]** | **3.2992** | **0.086** | **38.219** | **0** | **3.130 3.468** |
| **UNIT[T.R021]** | **3.0193** | **0.087** | **34.888** | **0** | **2.850 3.189** |
| **UNIT[T.R022]** | **4.2529** | **0.086** | **49.268** | **0** | **4.084 4.422** |
| **UNIT[T.R023]** | **3.6233** | **0.086** | **41.975** | **0** | **3.454 3.793** |
| **UNIT[T.R024]** | **2.878** | **0.086** | **33.348** | **0** | **2.709 3.047** |
| **UNIT[T.R025]** | **3.4709** | **0.086** | **40.268** | **0** | **3.302 3.640** |
| **UNIT[T.R027]** | **2.23** | **0.086** | **25.833** | **0** | **2.061 2.399** |
| **UNIT[T.R029]** | **3.8007** | **0.086** | **44.029** | **0** | **3.631 3.970** |
| **UNIT[T.R030]** | **2.5214** | **0.086** | **29.21** | **0** | **2.352 2.691** |
| **UNIT[T.R031]** | **3.1874** | **0.086** | **36.925** | **0** | **3.018 3.357** |
| **UNIT[T.R032]** | **3.0722** | **0.086** | **35.545** | **0** | **2.903 3.242** |
| **UNIT[T.R033]** | **4.2607** | **0.086** | **49.358** | **0** | **4.091 4.430** |
| **UNIT[T.R034]** | **1.6725** | **0.088** | **18.958** | **0** | **1.500 1.845** |
| **UNIT[T.R035]** | **2.5875** | **0.087** | **29.9** | **0** | **2.418 2.757** |
| **UNIT[T.R036]** | **0.9736** | **0.087** | **11.166** | **0** | **0.803 1.145** |
| **UNIT[T.R037]** | **1.0833** | **0.087** | **12.523** | **0** | **0.914 1.253** |
| **UNIT[T.R038]** | **0.0868** | **0.087** | **0.992** | **0.321** | **-0.085 0.258** |
| **UNIT[T.R039]** | **0.9535** | **0.089** | **10.763** | **0** | **0.780 1.127** |
| **UNIT[T.R040]** | **1.2248** | **0.086** | **14.178** | **0** | **1.055 1.394** |
| **UNIT[T.R041]** | **2.355** | **0.086** | **27.282** | **0** | **2.186 2.524** |
| **UNIT[T.R042]** | **0.81** | **0.087** | **9.311** | **0** | **0.639 0.980** |
| **UNIT[T.R043]** | **2.1319** | **0.086** | **24.697** | **0** | **1.963 2.301** |
| **UNIT[T.R044]** | **3.0691** | **0.086** | **35.554** | **0** | **2.900 3.238** |
| **UNIT[T.R046]** | **4.0909** | **0.086** | **47.391** | **0** | **3.922 4.260** |
| **UNIT[T.R049]** | **2.4097** | **0.086** | **27.915** | **0** | **2.240 2.579** |
| **UNIT[T.R050]** | **2.8953** | **0.087** | **33.454** | **0** | **2.726 3.065** |
| **UNIT[T.R051]** | **3.3777** | **0.086** | **39.13** | **0** | **3.209 3.547** |
| **UNIT[T.R052]** | **1.8213** | **0.088** | **20.741** | **0** | **1.649 1.993** |
| **UNIT[T.R053]** | **2.7266** | **0.086** | **31.562** | **0** | **2.557 2.896** |
| **UNIT[T.R054]** | **1.5565** | **0.087** | **17.985** | **0** | **1.387 1.726** |
| **UNIT[T.R055]** | **4.2552** | **0.086** | **49.383** | **0** | **4.086 4.424** |
| **UNIT[T.R056]** | **1.8209** | **0.087** | **21.041** | **0** | **1.651 1.991** |
| **UNIT[T.R057]** | **3.4509** | **0.086** | **39.977** | **0** | **3.282 3.620** |
| **UNIT[T.R058]** | **0.911** | **0.086** | **10.54** | **0** | **0.742 1.080** |
| **UNIT[T.R059]** | **1.5171** | **0.087** | **17.367** | **0** | **1.346 1.688** |
| **UNIT[T.R060]** | **1.1619** | **0.087** | **13.356** | **0** | **0.991 1.332** |
| **UNIT[T.R061]** | **1.1196** | **0.089** | **12.635** | **0** | **0.946 1.293** |
| **UNIT[T.R062]** | **2.5699** | **0.086** | **29.771** | **0** | **2.401 2.739** |
| **UNIT[T.R063]** | **1.5756** | **0.088** | **17.807** | **0** | **1.402 1.749** |
| **UNIT[T.R064]** | **1.28** | **0.087** | **14.634** | **0** | **1.109 1.451** |
| **UNIT[T.R065]** | **1.332** | **0.088** | **15.143** | **0** | **1.160 1.504** |
| **UNIT[T.R066]** | **0.4899** | **0.088** | **5.554** | **0** | **0.317 0.663** |
| **UNIT[T.R067]** | **1.4071** | **0.088** | **15.908** | **0** | **1.234 1.581** |
| **UNIT[T.R068]** | **0.8554** | **0.088** | **9.67** | **0** | **0.682 1.029** |
| **UNIT[T.R069]** | **1.5057** | **0.088** | **17.171** | **0** | **1.334 1.678** |
| **UNIT[T.R070]** | **2.0673** | **0.086** | **23.948** | **0** | **1.898 2.236** |
| **UNIT[T.R080]** | **2.884** | **0.086** | **33.41** | **0** | **2.715 3.053** |
| **UNIT[T.R081]** | **2.9041** | **0.087** | **33.557** | **0** | **2.734 3.074** |
| **UNIT[T.R082]** | **1.7201** | **0.087** | **19.876** | **0** | **1.551 1.890** |
| **UNIT[T.R083]** | **2.6038** | **0.086** | **30.164** | **0** | **2.435 2.773** |
| **UNIT[T.R084]** | **4.5064** | **0.086** | **52.204** | **0** | **4.337 4.676** |
| **UNIT[T.R085]** | **2.5173** | **0.087** | **29.087** | **0** | **2.348 2.687** |
| **UNIT[T.R086]** | **2.2421** | **0.086** | **25.973** | **0** | **2.073 2.411** |
| **UNIT[T.R087]** | **1.4066** | **0.087** | **16.211** | **0** | **1.236 1.577** |
| **UNIT[T.R089]** | **0.6854** | **0.087** | **7.919** | **0** | **0.516 0.855** |
| **UNIT[T.R090]** | **0.7872** | **0.088** | **8.899** | **0** | **0.614 0.961** |
| **UNIT[T.R091]** | **1.6607** | **0.088** | **18.857** | **0** | **1.488 1.833** |
| **UNIT[T.R092]** | **2.1248** | **0.087** | **24.334** | **0** | **1.954 2.296** |
| **UNIT[T.R093]** | **2.1844** | **0.088** | **24.946** | **0** | **2.013 2.356** |
| **UNIT[T.R094]** | **2.0499** | **0.086** | **23.73** | **0** | **1.881 2.219** |
| **UNIT[T.R095]** | **2.2795** | **0.087** | **26.284** | **0** | **2.110 2.450** |
| **UNIT[T.R096]** | **2.4706** | **0.086** | **28.636** | **0** | **2.301 2.640** |
| **UNIT[T.R097]** | **2.5697** | **0.086** | **29.785** | **0** | **2.401 2.739** |
| **UNIT[T.R098]** | **2.0866** | **0.086** | **24.172** | **0** | **1.917 2.256** |
| **UNIT[T.R099]** | **2.3592** | **0.086** | **27.331** | **0** | **2.190 2.528** |
| **UNIT[T.R100]** | **0.8628** | **0.087** | **9.961** | **0** | **0.693 1.033** |
| **UNIT[T.R101]** | **2.6842** | **0.086** | **31.095** | **0** | **2.515 2.853** |
| **UNIT[T.R102]** | **3.047** | **0.086** | **35.298** | **0** | **2.878 3.216** |
| **UNIT[T.R103]** | **1.9517** | **0.088** | **22.226** | **0** | **1.780 2.124** |
| **UNIT[T.R104]** | **1.8127** | **0.086** | **20.983** | **0** | **1.643 1.982** |
| **UNIT[T.R105]** | **2.9619** | **0.086** | **34.312** | **0** | **2.793 3.131** |
| **UNIT[T.R106]** | **1.5952** | **0.088** | **18.034** | **0** | **1.422 1.769** |
| **UNIT[T.R107]** | **0.9518** | **0.089** | **10.743** | **0** | **0.778 1.125** |
| **UNIT[T.R108]** | **3.3052** | **0.086** | **38.289** | **0** | **3.136 3.474** |
| **UNIT[T.R111]** | **2.7868** | **0.086** | **32.284** | **0** | **2.618 2.956** |
| **UNIT[T.R112]** | **2.1232** | **0.086** | **24.609** | **0** | **1.954 2.292** |
| **UNIT[T.R114]** | **1.4011** | **0.086** | **16.234** | **0** | **1.232 1.570** |
| **UNIT[T.R115]** | **1.7713** | **0.086** | **20.55** | **0** | **1.602 1.940** |
| **UNIT[T.R116]** | **2.6886** | **0.086** | **31.147** | **0** | **2.519 2.858** |
| **UNIT[T.R117]** | **1.0922** | **0.088** | **12.366** | **0** | **0.919 1.265** |
| **UNIT[T.R119]** | **2.1573** | **0.087** | **24.808** | **0** | **1.987 2.328** |
| **UNIT[T.R120]** | **1.9383** | **0.088** | **22.137** | **0** | **1.767 2.110** |
| **UNIT[T.R121]** | **1.9018** | **0.087** | **21.742** | **0** | **1.730 2.073** |
| **UNIT[T.R122]** | **2.5448** | **0.087** | **29.341** | **0** | **2.375 2.715** |
| **UNIT[T.R123]** | **1.9751** | **0.087** | **22.704** | **0** | **1.805 2.146** |
| **UNIT[T.R124]** | **1.1461** | **0.088** | **13.011** | **0** | **0.973 1.319** |
| **UNIT[T.R126]** | **2.1271** | **0.086** | **24.641** | **0** | **1.958 2.296** |
| **UNIT[T.R127]** | **3.2099** | **0.086** | **37.185** | **0** | **3.041 3.379** |
| **UNIT[T.R137]** | **2.5455** | **0.086** | **29.542** | **0** | **2.377 2.714** |
| **UNIT[T.R139]** | **2.3628** | **0.086** | **27.337** | **0** | **2.193 2.532** |
| **UNIT[T.R163]** | **2.8403** | **0.086** | **32.904** | **0** | **2.671 3.009** |
| **UNIT[T.R172]** | **2.0407** | **0.086** | **23.641** | **0** | **1.872 2.210** |
| **UNIT[T.R179]** | **3.7245** | **0.086** | **43.146** | **0** | **3.555 3.894** |
| **UNIT[T.R181]** | **2.1039** | **0.087** | **24.151** | **0** | **1.933 2.275** |
| **UNIT[T.R183]** | **1.3255** | **0.089** | **14.962** | **0** | **1.152 1.499** |
| **UNIT[T.R184]** | **1.4384** | **0.088** | **16.305** | **0** | **1.266 1.611** |
| **UNIT[T.R186]** | **1.401** | **0.087** | **16.124** | **0** | **1.231 1.571** |
| **UNIT[T.R188]** | **2.4316** | **0.086** | **28.133** | **0** | **2.262 2.601** |
| **UNIT[T.R189]** | **1.9629** | **0.087** | **22.471** | **0** | **1.792 2.134** |
| **UNIT[T.R194]** | **2.253** | **0.087** | **25.864** | **0** | **2.082 2.424** |
| **UNIT[T.R196]** | **1.7425** | **0.087** | **20.109** | **0** | **1.573 1.912** |
| **UNIT[T.R198]** | **2.285** | **0.087** | **26.403** | **0** | **2.115 2.455** |
| **UNIT[T.R199]** | **1.2758** | **0.087** | **14.646** | **0** | **1.105 1.447** |
| **UNIT[T.R200]** | **1.4418** | **0.087** | **16.641** | **0** | **1.272 1.612** |
| **UNIT[T.R202]** | **2.3682** | **0.086** | **27.413** | **0** | **2.199 2.537** |
| **UNIT[T.R203]** | **1.1368** | **0.088** | **12.978** | **0** | **0.965 1.308** |
| **UNIT[T.R204]** | **1.659** | **0.086** | **19.218** | **0** | **1.490 1.828** |
| **UNIT[T.R205]** | **1.9848** | **0.087** | **22.916** | **0** | **1.815 2.155** |
| **UNIT[T.R207]** | **2.2567** | **0.086** | **26.109** | **0** | **2.087 2.426** |
| **UNIT[T.R208]** | **2.5003** | **0.087** | **28.868** | **0** | **2.331 2.670** |
| **UNIT[T.R209]** | **1.2845** | **0.088** | **14.518** | **0** | **1.111 1.458** |
| **UNIT[T.R210]** | **0.9475** | **0.087** | **10.847** | **0** | **0.776 1.119** |
| **UNIT[T.R211]** | **2.4097** | **0.086** | **27.915** | **0** | **2.241 2.579** |
| **UNIT[T.R212]** | **2.0122** | **0.086** | **23.282** | **0** | **1.843 2.182** |
| **UNIT[T.R213]** | **1.5396** | **0.087** | **17.65** | **0** | **1.369 1.711** |
| **UNIT[T.R214]** | **1.0354** | **0.089** | **11.685** | **0** | **0.862 1.209** |
| **UNIT[T.R215]** | **1.9216** | **0.087** | **22.204** | **0** | **1.752 2.091** |
| **UNIT[T.R216]** | **1.2106** | **0.087** | **13.988** | **0** | **1.041 1.380** |
| **UNIT[T.R217]** | **1.4498** | **0.088** | **16.387** | **0** | **1.276 1.623** |
| **UNIT[T.R218]** | **2.2048** | **0.086** | **25.555** | **0** | **2.036 2.374** |
| **UNIT[T.R219]** | **1.6323** | **0.086** | **18.895** | **0** | **1.463 1.802** |
| **UNIT[T.R220]** | **1.7053** | **0.086** | **19.756** | **0** | **1.536 1.875** |
| **UNIT[T.R221]** | **1.8658** | **0.089** | **21.056** | **0** | **1.692 2.039** |
| **UNIT[T.R223]** | **2.4156** | **0.086** | **27.998** | **0** | **2.246 2.585** |
| **UNIT[T.R224]** | **1.215** | **0.087** | **13.891** | **0** | **1.044 1.386** |
| **UNIT[T.R225]** | **0.8349** | **0.087** | **9.61** | **0** | **0.665 1.005** |
| **UNIT[T.R226]** | **1.1868** | **0.088** | **13.477** | **0** | **1.014 1.359** |
| **UNIT[T.R227]** | **1.4202** | **0.086** | **16.452** | **0** | **1.251 1.589** |
| **UNIT[T.R228]** | **1.5647** | **0.088** | **17.71** | **0** | **1.391 1.738** |
| **UNIT[T.R229]** | **0.8798** | **0.088** | **9.987** | **0** | **0.707 1.052** |
| **UNIT[T.R230]** | **0.815** | **0.088** | **9.305** | **0** | **0.643 0.987** |
| **UNIT[T.R231]** | **1.4934** | **0.087** | **17.189** | **0** | **1.323 1.664** |
| **UNIT[T.R232]** | **1.4973** | **0.088** | **17.022** | **0** | **1.325 1.670** |
| **UNIT[T.R233]** | **1.5335** | **0.089** | **17.254** | **0** | **1.359 1.708** |
| **UNIT[T.R234]** | **0.6198** | **0.088** | **7.005** | **0** | **0.446 0.793** |
| **UNIT[T.R235]** | **2.6565** | **0.086** | **30.735** | **0** | **2.487 2.826** |
| **UNIT[T.R236]** | **1.7992** | **0.087** | **20.773** | **0** | **1.629 1.969** |
| **UNIT[T.R237]** | **1.2304** | **0.088** | **13.951** | **0** | **1.058 1.403** |
| **UNIT[T.R238]** | **2.3919** | **0.086** | **27.724** | **0** | **2.223 2.561** |
| **UNIT[T.R239]** | **1.4299** | **0.086** | **16.565** | **0** | **1.261 1.599** |
| **UNIT[T.R240]** | **2.5146** | **0.087** | **29.018** | **0** | **2.345 2.684** |
| **UNIT[T.R242]** | **0.9963** | **0.087** | **11.422** | **0** | **0.825 1.167** |
| **UNIT[T.R243]** | **1.8817** | **0.088** | **21.46** | **0** | **1.710 2.054** |
| **UNIT[T.R244]** | **1.8669** | **0.088** | **21.29** | **0** | **1.695 2.039** |
| **UNIT[T.R246]** | **1.0911** | **0.088** | **12.349** | **0** | **0.918 1.264** |
| **UNIT[T.R247]** | **0.3239** | **0.089** | **3.633** | **0** | **0.149 0.499** |
| **UNIT[T.R248]** | **2.8801** | **0.086** | **33.323** | **0** | **2.711 3.050** |
| **UNIT[T.R249]** | **1.6038** | **0.087** | **18.436** | **0** | **1.433 1.774** |
| **UNIT[T.R250]** | **1.3145** | **0.087** | **15.069** | **0** | **1.144 1.486** |
| **UNIT[T.R251]** | **1.5661** | **0.087** | **18.074** | **0** | **1.396 1.736** |
| **UNIT[T.R252]** | **1.4206** | **0.087** | **16.415** | **0** | **1.251 1.590** |
| **UNIT[T.R253]** | **1.1984** | **0.088** | **13.647** | **0** | **1.026 1.371** |
| **UNIT[T.R254]** | **2.4674** | **0.086** | **28.598** | **0** | **2.298 2.636** |
| **UNIT[T.R255]** | **1.1622** | **0.087** | **13.414** | **0** | **0.992 1.332** |
| **UNIT[T.R256]** | **1.5781** | **0.087** | **18.234** | **0** | **1.408 1.748** |
| **UNIT[T.R257]** | **2.1316** | **0.086** | **24.694** | **0** | **1.962 2.301** |
| **UNIT[T.R258]** | **1.9778** | **0.087** | **22.824** | **0** | **1.808 2.148** |
| **UNIT[T.R259]** | **1.3011** | **0.087** | **14.955** | **0** | **1.131 1.472** |
| **UNIT[T.R260]** | **1.249** | **0.09** | **13.855** | **0** | **1.072 1.426** |
| **UNIT[T.R261]** | **1.9762** | **0.087** | **22.663** | **0** | **1.805 2.147** |
| **UNIT[T.R262]** | **0.8266** | **0.089** | **9.272** | **0** | **0.652 1.001** |
| **UNIT[T.R263]** | **0.0449** | **0.087** | **0.514** | **0.607** | **-0.126 0.216** |
| **UNIT[T.R264]** | **0.8271** | **0.086** | **9.569** | **0** | **0.658 0.996** |
| **UNIT[T.R265]** | **1.292** | **0.088** | **14.668** | **0** | **1.119 1.465** |
| **UNIT[T.R266]** | **1.3361** | **0.086** | **15.487** | **0** | **1.167 1.505** |
| **UNIT[T.R269]** | **1.365** | **0.087** | **15.773** | **0** | **1.195 1.535** |
| **UNIT[T.R270]** | **0.8689** | **0.088** | **9.834** | **0** | **0.696 1.042** |
| **UNIT[T.R271]** | **0.6723** | **0.088** | **7.62** | **0** | **0.499 0.845** |
| **UNIT[T.R273]** | **1.7934** | **0.088** | **20.266** | **0** | **1.620 1.967** |
| **UNIT[T.R274]** | **1.0458** | **0.088** | **11.888** | **0** | **0.873 1.218** |
| **UNIT[T.R275]** | **1.533** | **0.087** | **17.581** | **0** | **1.362 1.704** |
| **UNIT[T.R276]** | **1.6483** | **0.086** | **19.095** | **0** | **1.479 1.818** |
| **UNIT[T.R277]** | **0.863** | **0.09** | **9.635** | **0** | **0.687 1.039** |
| **UNIT[T.R278]** | **0.726** | **0.088** | **8.242** | **0** | **0.553 0.899** |
| **UNIT[T.R279]** | **1.2215** | **0.087** | **14.047** | **0** | **1.051 1.392** |
| **UNIT[T.R280]** | **1.0729** | **0.089** | **12.016** | **0** | **0.898 1.248** |
| **UNIT[T.R281]** | **1.6365** | **0.087** | **18.811** | **0** | **1.466 1.807** |
| **UNIT[T.R282]** | **1.7346** | **0.087** | **20.044** | **0** | **1.565 1.904** |
| **UNIT[T.R284]** | **1.237** | **0.087** | **14.294** | **0** | **1.067 1.407** |
| **UNIT[T.R285]** | **0.9049** | **0.088** | **10.226** | **0** | **0.731 1.078** |
| **UNIT[T.R287]** | **1.0365** | **0.089** | **11.68** | **0** | **0.863 1.210** |
| **UNIT[T.R291]** | **1.7456** | **0.086** | **20.222** | **0** | **1.576 1.915** |
| **UNIT[T.R294]** | **1.3618** | **0.087** | **15.59** | **0** | **1.191 1.533** |
| **UNIT[T.R295]** | **0.913** | **0.093** | **9.816** | **0** | **0.731 1.095** |
| **UNIT[T.R300]** | **2.4184** | **0.086** | **28.016** | **0** | **2.249 2.588** |
| **UNIT[T.R303]** | **1.5371** | **0.087** | **17.668** | **0** | **1.367 1.708** |
| **UNIT[T.R304]** | **1.1126** | **0.087** | **12.856** | **0** | **0.943 1.282** |
| **UNIT[T.R307]** | **0.7862** | **0.089** | **8.875** | **0** | **0.613 0.960** |
| **UNIT[T.R308]** | **1.3247** | **0.088** | **15.107** | **0** | **1.153 1.497** |
| **UNIT[T.R309]** | **1.3713** | **0.088** | **15.661** | **0** | **1.200 1.543** |
| **UNIT[T.R310]** | **1.6956** | **0.088** | **19.254** | **0** | **1.523 1.868** |
| **UNIT[T.R311]** | **0.8762** | **0.088** | **10.003** | **0** | **0.705 1.048** |
| **UNIT[T.R312]** | **0.5652** | **0.087** | **6.523** | **0** | **0.395 0.735** |
| **UNIT[T.R313]** | **0.0306** | **0.089** | **0.344** | **0.731** | **-0.143 0.204** |
| **UNIT[T.R318]** | **0.9319** | **0.087** | **10.74** | **0** | **0.762 1.102** |
| **UNIT[T.R319]** | **1.5842** | **0.087** | **18.186** | **0** | **1.413 1.755** |
| **UNIT[T.R321]** | **1.4301** | **0.086** | **16.567** | **0** | **1.261 1.599** |
| **UNIT[T.R322]** | **2.0048** | **0.087** | **23.013** | **0** | **1.834 2.176** |
| **UNIT[T.R323]** | **1.8004** | **0.088** | **20.469** | **0** | **1.628 1.973** |
| **UNIT[T.R325]** | **0.682** | **0.087** | **7.799** | **0** | **0.511 0.853** |
| **UNIT[T.R330]** | **1.5494** | **0.087** | **17.737** | **0** | **1.378 1.721** |
| **UNIT[T.R335]** | **0.5539** | **0.089** | **6.231** | **0** | **0.380 0.728** |
| **UNIT[T.R336]** | **-0.3389** | **0.089** | **-3.813** | **0** | **-0.678** |
| **UNIT[T.R337]** | **-0.1285** | **0.088** | **-1.454** | **0.146** | **-0.302 0.045** |
| **UNIT[T.R338]** | **-0.9876** | **0.088** | **-11.276** | **0** | **-1.975** |
| **UNIT[T.R341]** | **0.8017** | **0.087** | **9.265** | **0** | **0.632 0.971** |
| **UNIT[T.R344]** | **0.7721** | **0.089** | **8.673** | **0** | **0.598 0.947** |
| **UNIT[T.R345]** | **0.7985** | **0.088** | **9.116** | **0** | **0.627 0.970** |
| **UNIT[T.R346]** | **1.5864** | **0.088** | **18.085** | **0** | **1.414 1.758** |
| **UNIT[T.R348]** | **0.1766** | **0.088** | **2.016** | **0.044** | **0.005 0.348** |
| **UNIT[T.R354]** | **0.1727** | **0.088** | **1.952** | **0.051** | **-0.001 0.346** |
| **UNIT[T.R356]** | **1.5591** | **0.087** | **17.849** | **0** | **1.388 1.730** |
| **UNIT[T.R358]** | **0.2289** | **0.088** | **2.587** | **0.01** | **0.055 0.402** |
| **UNIT[T.R370]** | **0.5983** | **0.087** | **6.84** | **0** | **0.427 0.770** |
| **UNIT[T.R371]** | **0.8501** | **0.088** | **9.665** | **0** | **0.678 1.023** |
| **UNIT[T.R372]** | **1.0273** | **0.088** | **11.629** | **0** | **0.854 1.200** |
| **UNIT[T.R373]** | **0.9213** | **0.088** | **10.412** | **0** | **0.748 1.095** |
| **UNIT[T.R382]** | **1.3786** | **0.087** | **15.805** | **0** | **1.208 1.550** |
| **UNIT[T.R424]** | **0.5528** | **0.089** | **6.22** | **0** | **0.379 0.727** |
| **UNIT[T.R429]** | **1.5088** | **0.086** | **17.443** | **0** | **1.339 1.678** |
| **UNIT[T.R453]** | **1.9068** | **0.089** | **21.353** | **0** | **1.732 2.082** |
| **UNIT[T.R454]** | **0.0423** | **0.089** | **0.477** | **0.633** | **-0.131 0.216** |
| **UNIT[T.R455]** | **-0.1154** | **0.088** | **-1.304** | **0.192** | **-0.289 0.058** |
| **UNIT[T.R456]** | **0.217** | **0.087** | **2.484** | **0.013** | **0.046 0.388** |
| **UNIT[T.R459]** | **-1.5938** | **0.107** | **-14.955** | **0** | **-3.188** |
| **UNIT[T.R464]** | **-1.9718** | **0.088** | **-22.35** | **0** | **-3.944** |
| **C(day\_week)[T.1]** | **0.2267** | **0.014** | **15.777** | **0** | **0.199 0.255** |
| **C(day\_week)[T.2]** | **0.2273** | **0.015** | **15.202** | **0** | **0.198 0.257** |
| **C(day\_week)[T.3]** | **0.2146** | **0.015** | **13.951** | **0** | **0.184 0.245** |
| **C(day\_week)[T.4]** | **0.2204** | **0.015** | **14.47** | **0** | **0.191 0.250** |
| **C(day\_week)[T.5]** | **-0.3168** | **0.015** | **-20.785** | **0** | **-0.634** |
| **C(day\_week)[T.6]** | **-0.5468** | **0.014** | **-38.668** | **0** | **-1.094** |
| **C(fog)[T.1]** | **-0.6684** | **0.074** | **-9.049** | **0** | **-1.337** |
| **C(conds)[T.Fog]** | **1.2918** | **0.141** | **9.155** | **0** | **1.015 1.568** |
| **C(conds)[T.Haze]** | **0.004** | **0.025** | **0.163** | **0.871** | **-0.044 0.052** |
| **C(conds)[T.Heavy Rain]** | **0.4852** | **0.092** | **5.269** | **0** | **0.305 0.666** |
| **C(conds)[T.Light Drizzle]** | **-0.3719** | **0.048** | **-7.813** | **0** | **-0.744** |
| **C(conds)[T.Light Rain]** | **-0.0133** | **0.023** | **-0.578** | **0.563** | **-0.058 0.032** |
| **C(conds)[T.Mist]** | **1.0553** | **0.167** | **6.338** | **0** | **0.729 1.382** |
| **C(conds)[T.Mostly Cloudy]** | **-0.185** | **0.014** | **-13.258** | **0** | **-0.37** |
| **C(conds)[T.Overcast]** | **-0.1477** | **0.012** | **-12.356** | **0** | **-0.295** |
| **C(conds)[T.Partly Cloudy]** | **-0.0429** | **0.019** | **-2.201** | **0.028** | **-0.086** |
| **C(conds)[T.Rain]** | **0.2535** | **0.052** | **4.859** | **0** | **0.151 0.356** |
| **C(conds)[T.Scattered Clouds]** | **-0.079** | **0.017** | **-4.605** | **0** | **-0.158** |
| **precipi** | **-0.0546** | **0.008** | **-6.856** | **0** | **-0.109** |
| **tempi** | **-0.0894** | **0.005** | **-17.191** | **0** | **-0.179** |
| **pressurei** | **-0.0305** | **0.004** | **-7.088** | **0** | **-0.061** |



Comparing the summary table of the second model with the first model, the value of R2 was significantly improved and the skewness was reduced from 2.978to −0.426, which is far closer to zero. Skewness is a measurement of asymmetry of the distribution and of normal distribution the skewness is zero. So a good linear regression model should have the value of skewness close to zero.  
Then I examined the dependent variables versus residuals again. Although the residuals still not look like randomly distributed in the plot, the linear trend is not obvious any more (if there is any).



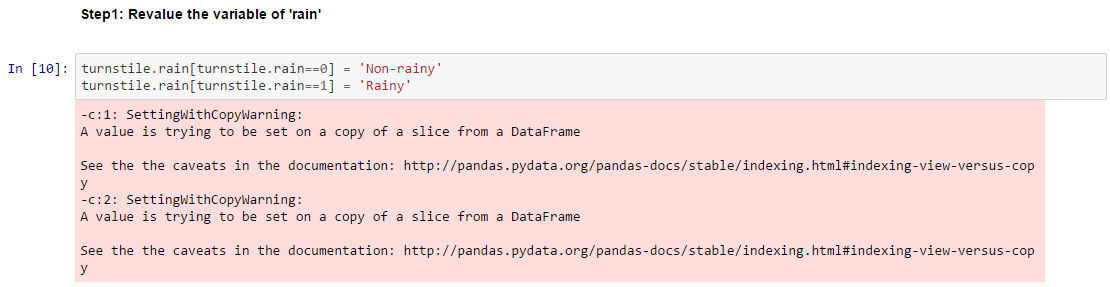
**Section 3. Visualization**

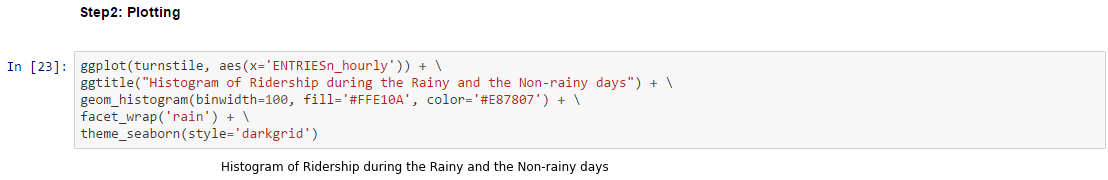
Please include two visualizations that show the relationships between two or more variables in the NYC subway data. You should feel free to implement something that we discussed in class (e.g., scatter plots, line plots, or histograms) or attempt to implement something more advanced if you'd like.  
Remember to add appropriate titles and axes labels to your plots. Also, please add a short description below each figure commenting on the key insights depicted in the figure.

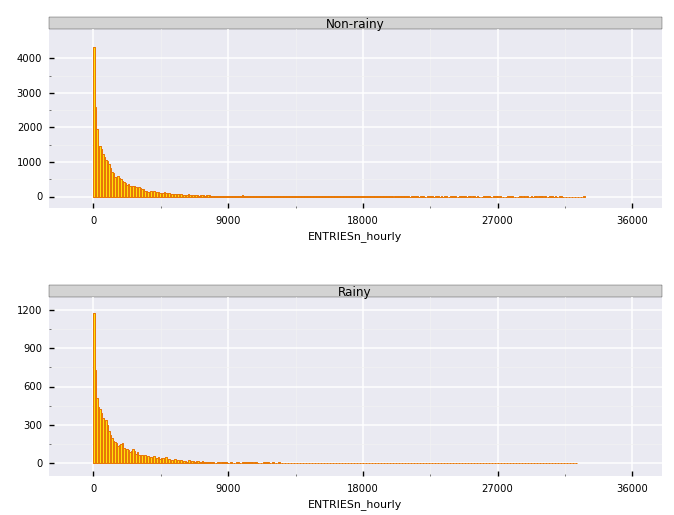
1. One visualization should contain two histograms: one of ENTRIESn\_hourly for rainy days and one of ENTRIESn\_hourly for non-rainy days. You can combine the two histograms in a single plot or you can use two different plots.  
   For the histogram, you should have intervals representing the volume of ridership (value of ENTRIESn\_hourly) on the x-axis and the frequency of occurrence on the y-axis. For example, you might have one interval (along the x-axis) with values from 0 to 1000. The height of the bar for this interval will then represent the number of records (rows in our data) that have ENTRIESn\_hourly that fall into this interval.  
   Remember to increase the number of bins in the histogram (by having larger number of bars). The default bin width is not sufficient to capture the variability in the two samples.
2. One visualization can be more freeform, some suggestions are:  
   a. Ridership by time-of-day or day-of-week  
   b. How ridership varies by subway station  
   c. Which stations have more exits or entries at different times of day

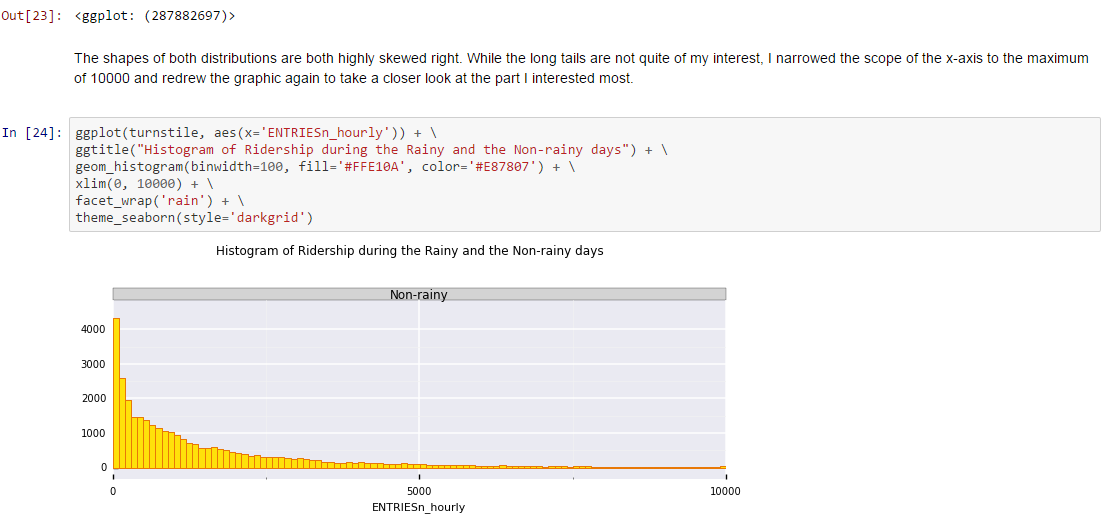
**Two histograms: one of ENTRIESn\_hourly for rainy days and one of ENTRIESn\_hourly for non-rainy days**

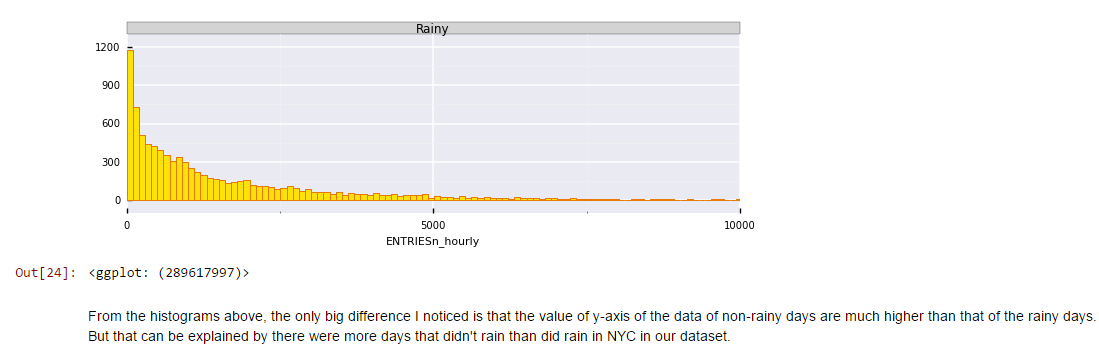
**Step1: Revalue the variable of 'rain'**







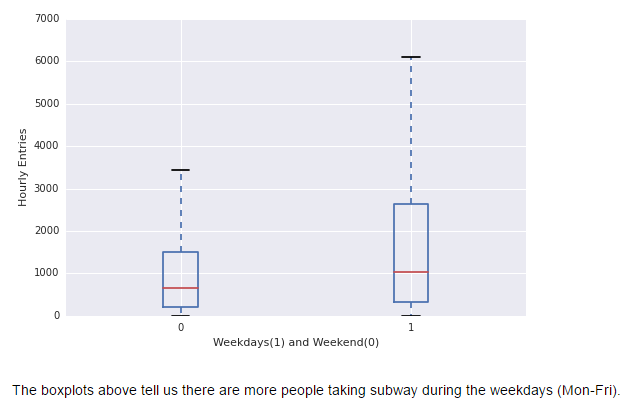


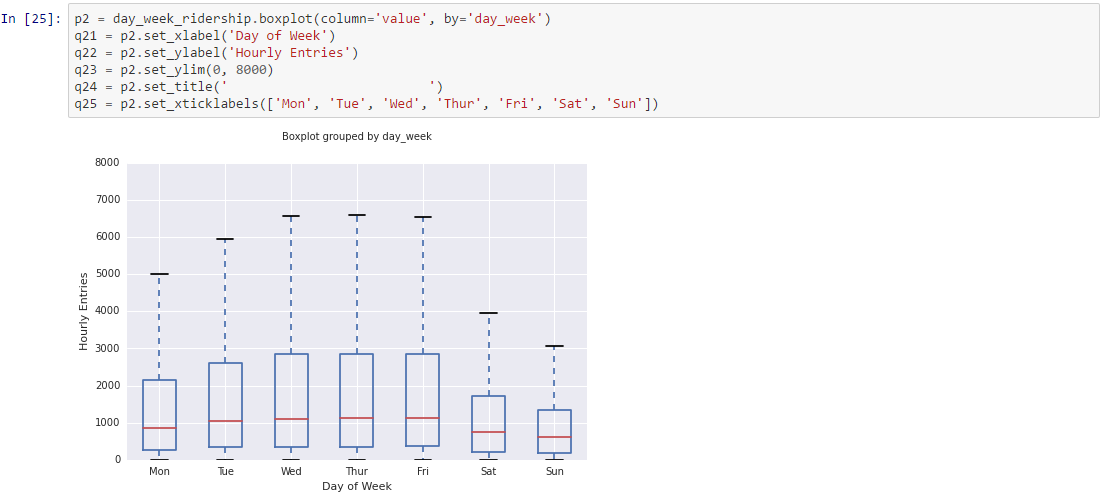




After narrowing the y-axis to (0,3000), we can see clearly all the first quartile, the median and the third quartile of the hourly entries of rainy days are higher then those of non-rainy days.







If we draw one box plot for each day of the week, we can see there is the smallest volume of riders in the subway on Sunday and the second smallest on Saturday. Surprisingly, less people take subway on Monday comparing to other weekdays.

**Section 4. Conclusion**

*Please address the following questions in detail. Your answers should be 1-2 paragraphs long.*

1. From your analysis and interpretation of the data, do more people ride the NYC subway when it is raining versus when it is not raining?
2. What analyses lead you to this conclusion?

From the result of the Mann–Whitney U test, the two-tailed p-value is 5.48213914249e-06, which indicates the probability of the distributions of ENTRIESn\_hourlyrainy and ENTRIESn\_hourlynon−rainy are the same is less than α. So there's a statistical significant difference between the riderships of rainy and non-rainy days. Furthermore, all the descriptive listed (including 1st quartile, median, mean and 3rd quartile) of rainy days are greater than those of non-rainy days and the same information also shows in the box plots. Therefore, I conclude that there are more people ride the NYC subway when it is raining versus when it is not raining.

**Section 5. Reflection**

*Please address the following questions in detail. Your answers should be 1-2 paragraphs long.*

1. Please discuss potential shortcomings of the data set and the methods of your analysis.
2. (Optional) Do you have any other insight about the dataset that you would like to share with us?

After plotting several box plots and histograms, I think for the purpose of comparing the distributions of different groups the box plots do a better job than histograms. So I will go with box plots next time I need to compare different distributions. But I do like how histograms give me more control, such as adjusting the width of the bars etc, so that I can get the information of the distribution I am looking for, but probably not for comparison purpose.

I explained how I got the two regression models in details from the regression section. Even though the second model was improved in terms of the R2, skewness and some other measurement of the model, I think there is still something missing in the second model. Maybe it is the interaction between two or more explanatory variables, or some transformation of some explanatory variables, or some other variables, which are not in the dataset, need to be added to the model. The expertise knowledge is desperately needed to improve my model.  
I would like to add another variable to indicate whether a certain day was a holiday or not. According to the previous analysis, whether a day is weekend or not makes a big difference in the volume of the ridership. So it is reasonable to analyze if a holiday usually witnesses a change of the volume of the ridership of NYC subway. For the weather in May, there may have not been many extreme weather conditions, but if we expand the timespan of the dataset I think it would be a good idea to add the weather information about the extreme weather, such as snowstorm in the winter time.