Bike Sharing Demand

Ramaa Nathan – Fall 2018

1 Executive Summary

Over the years, the demand for bike sharing has increased worldwide. Several environmental, logistic, and personal reasons influence the demand for bike rentals. In this project, an observational study has been conducted to model the various factors that influence bike rental. The data used in this study is from a public dataset and was collected during the years 2011 and 2012 by the Capital Bikeshare program with additional labeling provided by Fanee-T (Fanaee-T, 2013) and hosted on the UC Irving Machine Learning Repository website. The original dataset has 17,389 observations with 16 different variables. A smaller subset of the relevant data is used in this study with 4,630 observations and five variables. There are three categorical variables or predictors corresponding to the type of weather (3 levels - Clear, Cloudy, and Light Rain/Snow), time of day (5 levels), and type of day (2 levels – working and non-working). There is one quantitative variable or predictor corresponding to the actual feeling temperature that accounts for the temperature measured in Celsius, windspeed, and humidity. The response variable is the number of bikes rented in a specific hour. This is an observational study and so there is no randomization here. The treatment design can be considered to be a 5x3x2 factorial model. In this study, three different models were attempted – an ANOVA model with only the three categorical variables as factors, one ANCOVA model using the actual feel temperature as the covariate and assuming no interaction between the covariate and the other factors, and second ANCOVA model using the actual feel temperature as the covariate and interactions between the covariate and two of the factors. The third model was found to show significant group effects where the time of day and type of day strongly influence the number of bikes rented after the effects of temperature, humidity, and windspeed have been accounted for. These results would be very helpful to the companies renting out the bikes in understanding the pattern of bikes rented and in the predicting of number of bikes that would be rented on a given day and time based on the predicted weather conditions.

2 Introduction

With extreme urban traffic that has prompted congestion pricing in London, Singapore, Milan, Stockholm and consideration in New York City, bike sharing has become a convenient and viable way to access metropolitan areas. Capital Bikeshare is the current bikeshare service for the District of Columbia and surrounding metropolitan areas, now with over 4,000 bikes and 500 stations across Washington DC, Virginia, and Maryland. People using the bikes can become registered users or borrow a bike for a one time use. The Capital Bikeshare website (capitalbikeshare.com) suggests scenic rides around Washington DC, presumably for casual or one time users, as well as mentioning updates such as the commuter corral service for when bike docks in prime downtown locations are filled by 8am, presumably for registered and regular users. In fact, in the 2016 survey of Capital Bikeshare users (for which there were over 6,000 respondents), 65% of Bikeshare members surveyed used the service to get to work. Bike-sharing rental process is highly correlated to the environmental and seasonal settings. For instance, weather conditions, precipitation, day of week, season, hour of the day, etc. can affect the rental behaviors.

In this observational study, the goal is to study the influence of different factors - weather (Sunny and Clear, Cloudy, Light Rain/Snow), type of day (working or non-working), time of day (1: 6am-9am, 2: 9am-12pm, 3: 12pm-4pm, 4: 4pm-7pm; 5: 7pm-11pm) on the number of bikes rented by the hour. The real feeling temperature that accounts for the actual temperature, humidity, and windspeed will be considered as the covariate. Accordingly, two separate studies have been done here – ANOVA (with only the three categorical variables) and ANCOVA (with using atemp as the co-variate).

For the first factor in the multi-factor ANOVA , the null hypothesis being tested is that time of day does not affect the number of the bikes rented and the corresponding alternate hypothesis is that the number of bikes rented depends on the time of day.

$$Ho: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$Ha: At\ least\ one\ of\ the\ \mu_i is\ different$$

For the second factor, the null hypothesis is that weather does not affect the number of the bikes rented and the corresponding alternate hypothesis is that the number of bikes rented depends on the weather.

$$Ho: \mu_{Clear} = \mu_{Cloudy} = \mu_{Light\ Rain\ or\ Snow}$$

$$Ha: At\ least\ one\ of\ the\ \mu_i\ is\ different$$

For the third factor, the null hypothesis is that the number of bikes rented is independent of the day being a working day or a non-working day and the corresponding alternate hypothesis is that number of bikes rented is different for working day versus a non-working day.

$$Ho: \mu_{working} = \mu_{non-working}$$

$$Ha: \mu_{working} \neq \mu_{non-working}$$

Finally for the interaction, the null hypothesis is that there is no interaction between any of the factors and the corresponding null hypothesis is that there is an interaction and the various combination of the factors affects the number of bike rentals.

$$Ho: There\ is\ no\ interaction; Ha: An\ interaction\ exists$$

3 Methods

The data used in this study has been collected from the Capital Bikeshare program who publish quarterly data about bike usage in their system. However, there may be many factors affecting how many people are renting a bike in a particular hour that are not included in their published data sets. Faneee-T (Fanaee-T, 2013) from the University of Porto took the time to compile hourly weather conditions and whether a date was a federal holiday for the 2011 and 2012 years of data from the Capital Bikeshare program. This data set is hosted on the UC Irving Machine Learning Repository and contains 17,389 observations of 16 different variables.

For this observational study, the data corresponding to the last 10 days of each month in 2011 and 2012 has been considered for a total of nearly 6,500 observations. In prior studies (Christiana Kemp, Summer 2018), only a subset of the variables was found to be significant in predicting the number of bikes rented by the hour. A similar subset is considered here and it includes the qualitative variables corresponding to the hour expressed in military time as values 1 to 24, weather specified as one of clear, cloudy, and light rain/snow, type of day (working or non-working) and the quantitative variable corresponding to the actual feeling temperature that takes into account the actual temperature, humidity, and windspeed. The count of total rental bikes is the response variable. Further, as the dataset contains the number of bikes rented for each hour of the day and so would add up to 24 levels of the factor hour, a new variable "timeofday" has been created that has five levels – 6am-9am; 9am-12pm; 12pm-4pm, 4pm-7pm; 7pm-11pm. The overnight data from 11pm-6am has been filtered out. The final dataset used here has 4,630 records with 3 categorical factors (weather, timeofday, workingday) and one quantitative variable (atemp).

An extensive exploratory data analysis is conducted (details are in the Appendix) and it is found that each of the factors of weather, working day, time of day, and the feeling temperature has a strong influence on the distribution of the number of the bikes rented by the hour. An ANOVA is first conducted to study the effects of the individual and combined categorical factors on the number of bikes rented. The study is next extended to include the actual feeling temperature (atemp) as a covariate in an ANCOVA model. The treatment design or the data model considered for this study is a 5x3x2 factorial fixed effects model – 5 time groups in the day, 3 types of weather conditions, and 2 types of days. The bikes are the observational units and the three-factor model can be expressed as

$$count = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk} \; where \; \alpha_i, \beta_j, \gamma_k are \; the \; treatment \; or \; group \; effects.$$

4 Results

A multi-factor ANOVA model is applied to the data. As the initial data exploration suggested strong interaction between the factors, the model included all the 2-way and 3-way interactions.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
workingday	1	1607838	1607838	80.57	0.000
weather	2	5977325	2988663	149.77	0.000
timeofday	4	6158750	1539687	77.16	0.000
workingday*weather	2	89049	44524	2.23	0.108
workingday*timeofday	4	7464582	1866146	93.52	0.000
weather*timeofday	8	821617	102702	5.15	0.000
workingday*weather*timeofday	8	507822	63478	3.18	0.001
Error	4600	91791090	19955		
Total	4629	149645553			

Figure 1: ANOVA of the 3-factor model including 2-way and 3-way interactions

As can be seen from the ANOVA table in Figure 1, the p-value of the three way interaction is less than 0.01. Using a significance level of 0.05, and with p-values less than 0.01 for all factors and interactions except for the interaction between workingday and weather, all the factors and interactions are found to be significant. As the interaction of the three way interaction (workingday *weather * timeofday) is significant, the individual factors or pair-wise interactions are not considered. There is sufficient statistical evidence to conclude that the number of bikes rented is influenced by the combination all the three factors.

A residual analysis (Figure 10) is done to check the fit of the ANOVA model. From the plot of the residuals vs the fitted value, it can be seen that the variance of the residuals is not constant and exhibits a fanning effect with the variance increasing as the count or number of bikes rented during an hour increases. The histogram indicates a normal distribution but the normal probability plot with a high Anderson Darling value and very low p-value does not indicate a normal distribution. But, this non-normality condition can be ignored to some extent as the

ANOVA model is very robust against the normality condition and due to the large number of samples. To address the problem of non-constant variances, the response variable is transformed by taking its natural log and then re-fit with an ANOVA model. The resulting ANOVA table shown in Figure 11 shows that transforming the response variable does not change the ANOVA results. Using a significance value of 0.05 and with a p-value of 0.01, the interaction of the three factors is still significant. A check of the residual plots (Figure 12) shows that the residuals are independent and have a constant variance with no outliers. The normality of the residuals has not changed much but that can again be ignored as the ANOVA is very robust towards the normality condition. Figure 13 shows the distribution of residuals for each treatment group. Within each group, the residuals have mean of 0 and mostly normally distributed even though a few of the groups exhibit a few long tails. So, the log transformed model seems to be a good fit and is selected.

To understand the different group effects in the ANOVA model, the group means need to be compared. This 5x3x2 factorial model results in 30 different treatment groups. As can be seen in Figure 15, the ANOVA model is not balanced between these 30 groups and so it is important to compute the least square groups means which are shown in Figure 16. The Tukey mean comparison method is used for making the pairwise comparisons of the different treatment groups. The generated labels are shown in Figure 17 and the corresponding interval plot of the LS Means differences is shown below in Figure 2. The following observations can be made from the comparison of the differences of the group means:

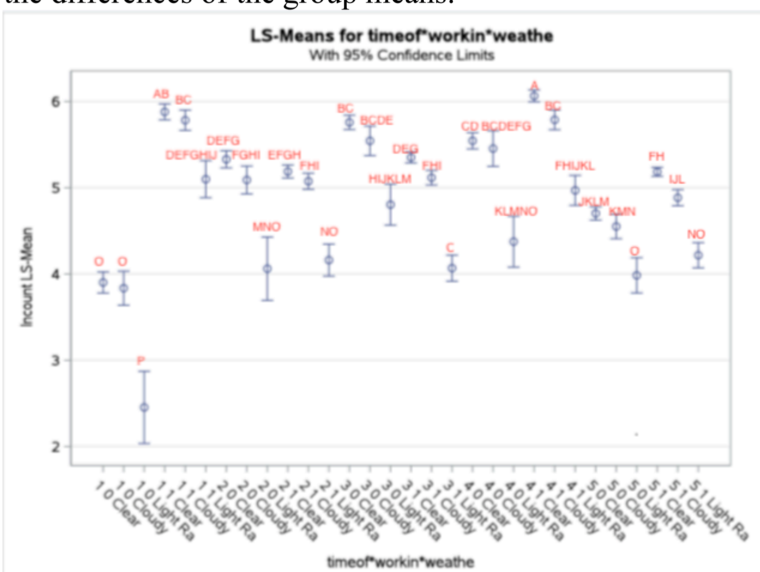

Figure 2: LS Means Interval Plot for the 3-factor ANOVA model

- On a clear working day, the average number of bikes rented between 6am and 9am and between 4pm and 7pm is significantly higher than the mean number of bikes rented during any other time. This was also observed from the plots in Figure 6 and Figure 7.
- On a non-working day when there is no rain or snow, the average number of bikes rented is high between the times of 12pm and 7pm.
- In the presence of light rain or snow, the average number of bikes rented in an hour is very low at all times of the day and on both working and non-working days except during the hours between 6am and 9am when it is the lowest on a non-working day and much higher on a working day.

- For all combinations of the type of day and time of day, the average number of bikes rented is significantly dependent on the weather. It is highest when the weather is clear and sunny. Lower when it is cloudy and lowest in the presence of light rain or snow.
- In general, the lowest number of bikes are rented between the hours of 6am and 9am on a non-working day.

A plot of the residual versus the actual feel temperature (Figure 14) shows a trend with the temperature suggesting that the temperature could be included in the model. A distinct relationship between the number of bikes rented by the hour and the actual feel temperature was also observed (Figure 8) during the exploratory data analysis. Based on these, an ANCOVA model is considered with the actual feel temperature (atemp) as the covariate.

To use the temperature variable as the covariate in the ANCOVA model, each of the 30 treatment groups is first checked for a significant regression relationship with the covariate. Using a significance value of 0.05, almost all the 30 different treatment groups (except for two) exhibit a significant linear relationship between the covariate and the response (Table 1). The treatment groups are next checked for equal regression slopes. To do this, an interaction between the covariate and each factor and between the covariate and each of the 2-way and 3-way interactions of the factor groups is checked. As seen in Figure 18, the p-values corresponding to the interactions of atemp with the individual factors and with the interaction groups are found to be significantly greater than 0.05 suggesting that the interactions are not significant and that in general the slopes can be considered to be equal. For two of the interaction groups, the atemp*timeofday and atemp*workingday*timeofday, there seems to be some interaction. This is also observed in Figure 25. Two different ANCOVA models are considered – one with equal slopes and the other without equal slopes.

The analysis of variance of the equal slopes model shows (Figure 19) that the three-way interaction term is significant and so the effects of the individual factors and the two-way interactions can be ignored. The residuals plots (Figure 20 and Figure 21) show a good fit without requiring any additional transformations. The Tukey comparisons are again used to compare the group means. The resulting interval plot (Figure 22) and the Tukey label chart (Figure 23) are similar to with a few differences when compared to the ones obtained with the 3-factor ANOVA model earlier. Some of the distinct differences in the group effects can be summarized as follows:
With the temperature, windspeed, and humidity accounted for,

- On a clear working day, the average number of bikes rented between 6am and 9am and between 4pm and 7pm are significantly higher than the mean number of bikes rented during any other time.
- There is significant difference in the mean number of bikes rented on non-working days between the 12pm-4pm and 4pm-7pm with few bikes being rented at the later hours.
- There is no significant difference in the mean number of bikes rented on working days during the mid-day and late night hours.
- In the presence of light snow or rain, there is no significant difference in the mean number of bikes rented anytime on non-working days and during the mid-day and late night on working days.

The resulting regression model is presented in Figure 24. On comparing the anova results for the ANOVA model (Figure 1) with the anova results for the ANCOVA model with equal slopes (Figure 18), it can be observed the mean square error has reduced from 19,955 to 12,169 (39%) indicating that by accounting for the temperature, the within-group variances have reduced considerably emphasizing the larger difference between the various interaction or treatment groups.

Finally, an ANCOVA model with unequal slopes is considered. As observed in Figure 25, there is a strong interaction between timeofday and workingday with no or very small interactions between weather and the other two factors. Based on this, the interactions between the covariate and the two factors of timeofday and workingday, and the two-way and 3-way interactions with these factors are included in the model (atemp, atemp*workingday, atemp*timeofday, atemp*workingday*timeofday). The resulting ANOVA table (Figure 26) shows the same reduction in the mean square error as was seen for the ANCOVA model with equal slopes. Using a significance value of 0.05, all the factors included in this model are found to be significant. The residual plots in Figure 27 and the residuals vs the atemp plot in Figure 28 suggest a good fit of the model. The interval plot of the LSMeans (Figure 3) and the corresponding table of mean comparisons with the Tukey labels (Figure 30) show a significantly smaller set but with distinct group differences. From this interval plot, it can be observed that after the effects of temperature, humidity, and windspeed have been accounted for, the mean number of bikes rented on a non-working day are distinctly different during each time period indicating that the time of day has a big effect. On a working day, there is no significant statistical difference between the mean number of bikes rented during the peak hours of 6am to 9am and 4pm to 7pm and also no significant statistical difference between the midday rentals and the late night rentals. These observations strongly match the observations made during the exploratory data analysis.

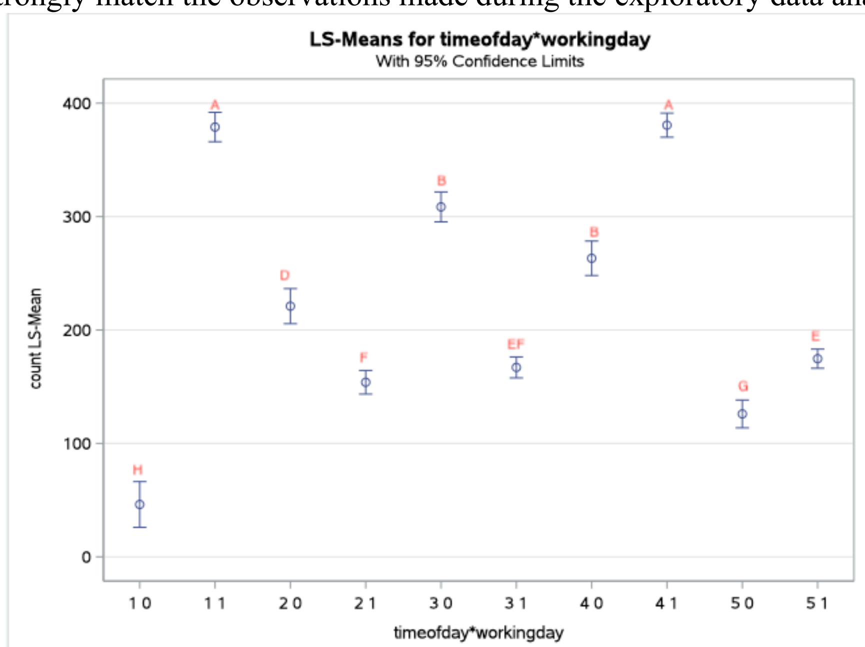

Figure 3: LS Means interval plot for ANCOVA model with unequal slopes

The regression equations generated for this model are shown in Figure 31. It can be seen that the slopes are the same for the same type of day and time period irrespective of the weather.

5 Conclusion

A number of environmental, logistic, and personal factors influence people to rent bikes. Using the data collected over a two year period by the Capital Bikeshare program and additional labeling of the data with the weather conditions and type of day, it is possible to understand the pattern of bike rentals. One of the uses of modeling this data would be to help in building a prediction model for predicting the number of bikes that would be rented on a given day based on the predicted weather and type of day. In this project, an analysis of variance study was conducted using three different models, a multifactor ANOVA, a multifactor ANCOVA assuming no interaction between the factors and the covariate, and a multifactor ANCOVA considering interaction between the covariate and the time and type of day factors. It was found that the third model using ANCOVA and only a few interactions produced the simplest model and showed very distinct group effects of the time and type of day after the effects of temperature, windspeed, and humidity have been accounted for. This model would be the recommended model for the purposes of further analysis or predictions.

References

(n.d.).

Christiana Kemp, H. M. (Summer 2018). *Bike Sharing Demand.* STAT897 Final Report .

Fanaee-T, H. a. (2013). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*(doi:10.1007/s13748-013-0040-3), 1-15.

Appendix

Data Source: https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset

Data: Variables included in the original dataset
There are 17,389 observations of 16 different variables with the following labels from the original data set:
instant: record index
dteday : date
season : factor with 1:spring, 2:summer, 3:fall, 4:winter
yr : year, with 0 for 2011, 1 for 2012
mnth : month
hr : hour
holiday : whether the day is a holiday (1) or not (0)
weekday : day of the week (0 to 6)
workingday : if day is neither weekend nor holiday, 1, otherwise 0.
weathersit : 1: Clear, Few clouds, Partly cloudy, Partly cloudy
2: Mist+Cloudy, Mist+Broken clouds, Mist+Few clouds, Mist
3: Light Snow, Light Rain+Thunderstorm+Scattered clouds, Light Rain+Scattered clouds
4: Heavy Rain+Ice Pellets+Thunderstorm+Mist, Snow+Fog
temp : normalized temperature in Celsius. The values are derived $\frac{T-T_{min}}{T_{max}-T_{min}}, T_{min} = -8,$
$T_{max} = 39$
atemp: normalized feeling temperature in Celsius. The values are derived via $\frac{T-T_{min}}{T_{max}-T_{min}},$
$T_{min} = -16, T_{min} = 50$
hum: normalized humidity. The values are divided by 100 (max)
windspeed: normalized wind speed. The values are divided by 67 (max)
casual: count of casual users
registered: count of registered users
cnt: count of total rental bikes including both casual and registered

Exploratory Data Analysis
The distribution of the number of bikes rented by the hour is explored in different ways – from a histogram for checking the frequencies of the number of bikes rented in an hour to the effects of the different factors on the distribution.

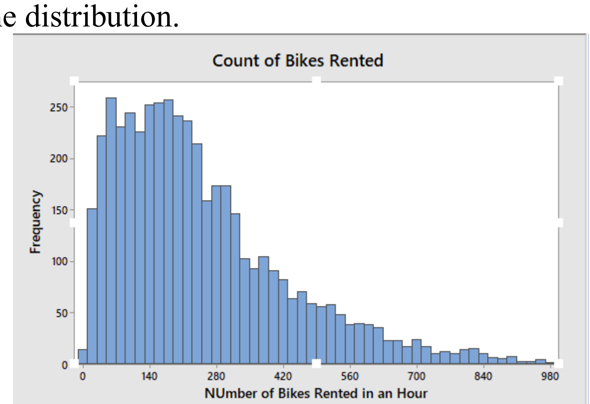

Figure 4: Histogram of number of bikes rented in an hour.

A histogram (Figure 4) is first plotted to check the frequency of the number of bikes rented by the hour. The distribution is skewed right showing that an average of 200 bikes are rented during the hour.

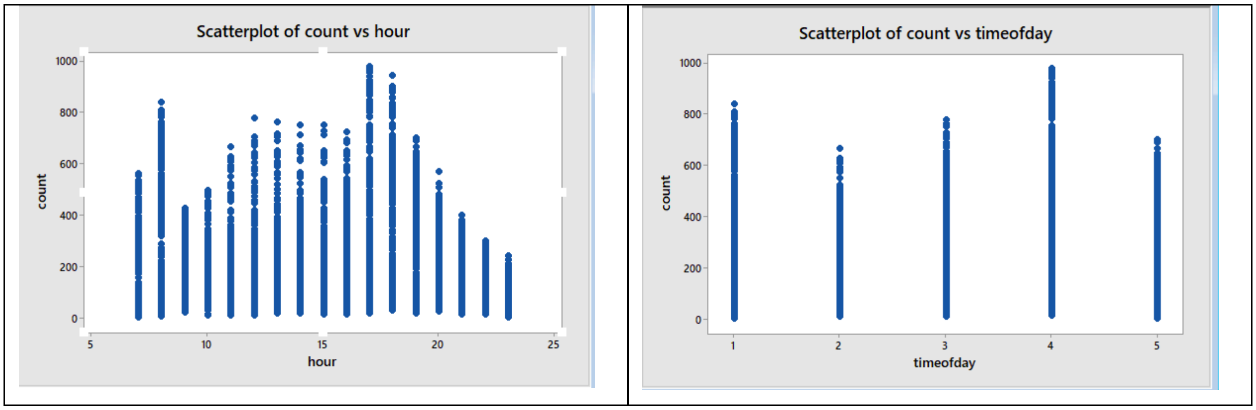

Figure 5: Distribution of bike rentals by the hour. The left panel shows the distribution by each distinct hour. The right panel shows the distribution by time of day, where the hours are grouped into 5 sections.

The overall distribution of the number of bikes rented during each hour appears unimodal and symmetric with a center around 3 pm with the hours of 7 am, 8am, 5pm, and 6 pm as outliers from the distribution.

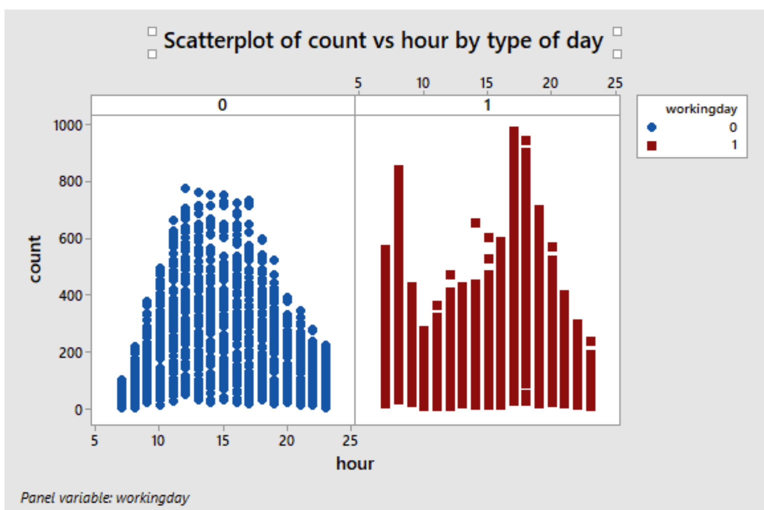

Figure 6: Distribution of bike rentals by the hour. The left panel shows the distribution of rentals on a non-working day (holidays and weekends). The right panel shows the distribution on a working day.

As seen in Figure 6, there is a strong influence of the type of day (working or non-working) on the number of bikes during the day. The number of bikes rented peaks at 8am, 5pm and 6pm on working days. This could correspond to the bikes rented for commuting purposes at the peak hours. The distribution of bike rentals on non-working days shows a distinct unimodal and symmetric pattern with higher usage between the hours of 10am and 4pm, indicating rentals for leisurely strolls on non-working days.

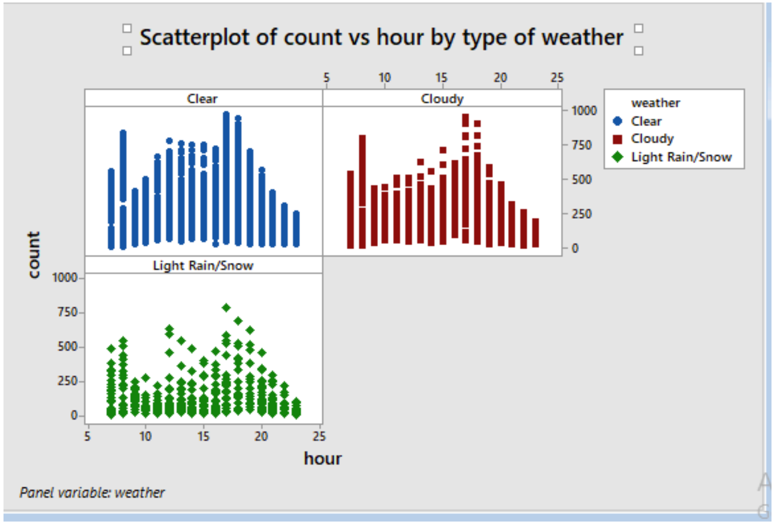

Figure 7 Distribution of bike rentals by the hour based on weather.

As seen in Figure 7, the distribution of number of bikes rented is also influenced by weather, but less so. On clear or cloudy days, the pattern is very similar except that a smaller number of bikes are rented in the middle of the day on cloudy days compared to the days on when it is clear. But the number of bikes rented at peak hours of 8am, 5pm, and 6pm seem to be the same as seen on working days. This suggests that weather is not a very strong influencer at peak hours on working days. But, as seen in the bottom panel, the number of bikes rental is substantially less when the weather is bad.

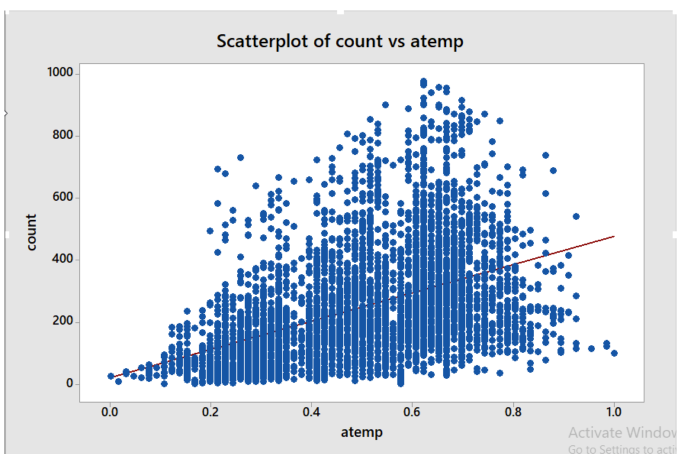

Figure 8 Distribution of number of bikes rented by the actual feeling temperature (actual temperature, humidity, and windspeed). The temperature values are normalized.

As seen in Figure 8, the number of bikes rented clearly increases with the increase in the actual feeling temperature.

ANOVA

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
workingday	1	1607838	1607838	80.57	0.000
weather	2	5977325	2988663	149.77	0.000
timeofday	4	6158750	1539687	77.16	0.000
workingday*weather	2	89049	44524	2.23	0.108
workingday*timeofday	4	7464582	1866146	93.52	0.000
weather*timeofday	8	821617	102702	5.15	0.000
workingday*weather*timeofday	8	507822	63478	3.18	$\boxed{0.001}$
Error	4600	91791090	19955		
Total	4629	149645553			

Figure 9 ANOVA of the 3-factor model including 2-way and 3-way interactions

Using a significance level of 0.05, and with p-values less than 0.01 for all factors and interactions except for the interaction between workingday and weather, all the factors and interactions are found to be significant. As the interaction of all the three factors (workingday *weather * timeofday) is significant, the individual factors or pair-wise interactions are not considered. There is sufficient statistical evidence to conclude that the number of bikes rented is influenced by the interaction all the three factors.

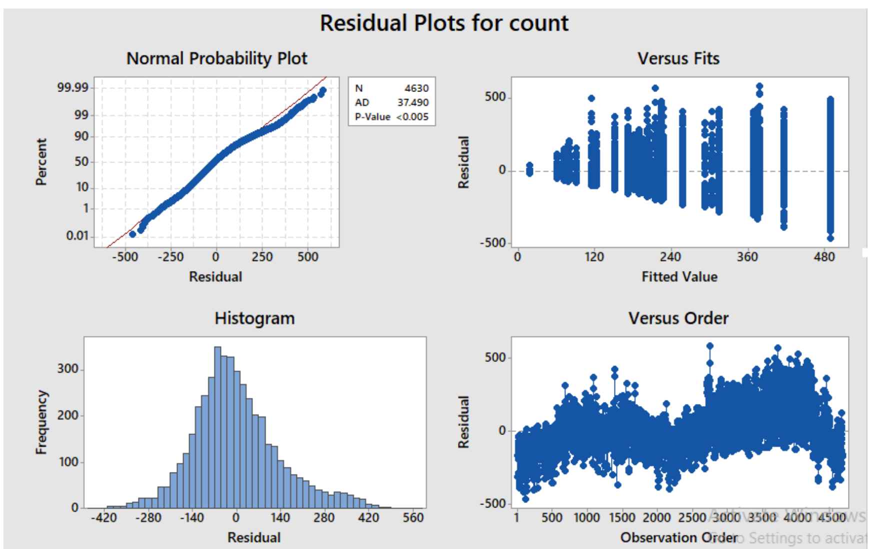

Figure 10 Residual Plots for the ANOVA model

A residual analysis is done to check the fit of the ANOVA model. From the plot of the residuals vs the fitted value, it can be seen that the variance of the residuals is not constant and exhibits a fanning effect with the variance increasing as the count or number of bikes rented during an hour increases. The histogram indicates a normal distribution but the normal probability plot with a high Anderson Darling value and very low p-value does not indicate a normal distribution. But, this non-normality condition can be ignored to some extent as the ANOVA model is very robust against the normality condition. To address the problem of non-constant variances, the response variable is transformed by taking its natural log and then re-fit with an ANOVA model.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
workingday	1	104.19	104.186	228.65	0.000
weather	2	281.36	140.682	308.75	0.000
timeofday	4	191.86	47.964	105.27	0.000
workingday*weather	2	0.97	0.487	1.07	0.344
workingday*timeofday	4	284.33	71.082	156.00	0.000
weather*timeofday	8	7.47	0.934	2.05	0.037
workingday*weather*timeofday	8	9.06	1.132	2.49	$\boxed{0.011}$
Error	4600	2095.99	0.456		
Total	4629	3507.82			

Figure 11: ANOVA model of the 3 factors with 2-way and 3-way interactions but with a natural log transformation applied to the response(count) variable.

The resulting ANOVA model shows that transforming the response variable does not change the ANOVA results. Using a significance value of 0.05 and with a p-value of 0.01, the interaction of the three factors is still significant. But there is a change in the residual plots.

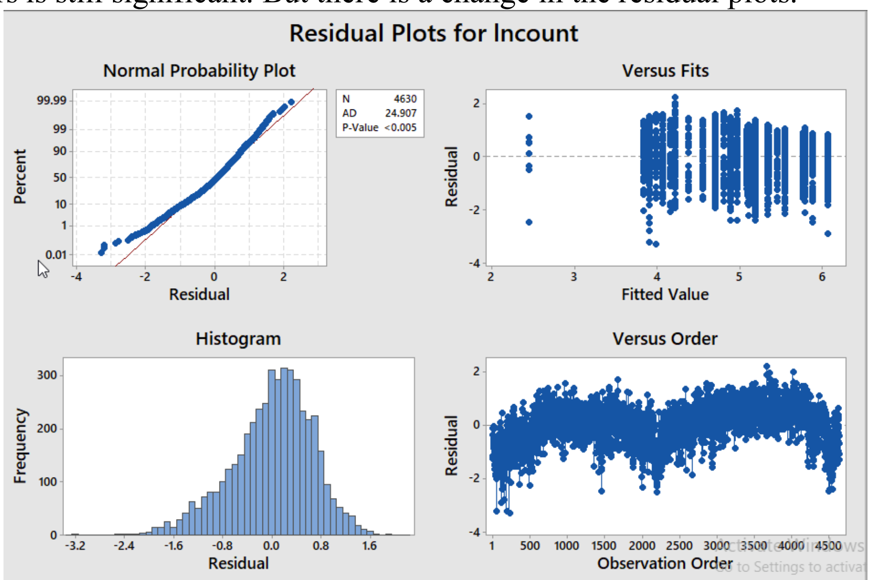

Figure 12: Residual Plots of the ANOVA model after log transformation of the response variable.

From the residuals vs fitted values plot, it can be observed that the residuals are independent and have a constant variance with no outliers. The normality of the residuals has not changed much but that can be ignored as the ANOVA is very robust towards the normality condition.

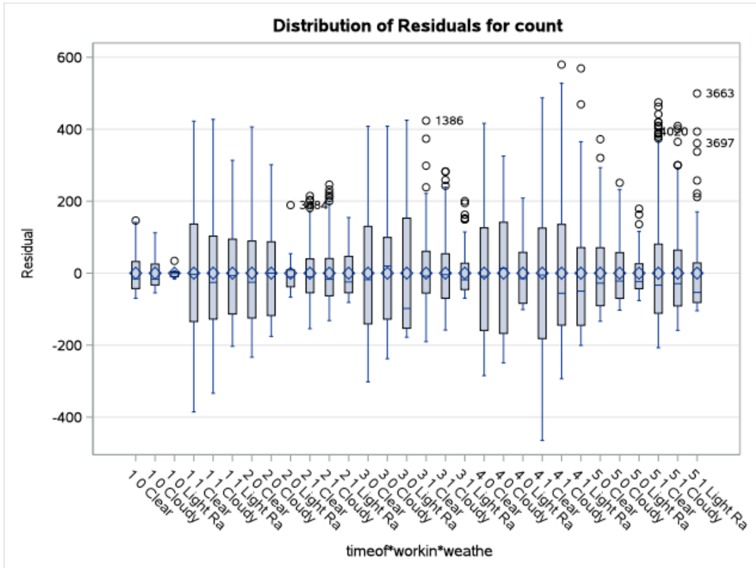

Figure 13: Distribution of residuals by each treatment group

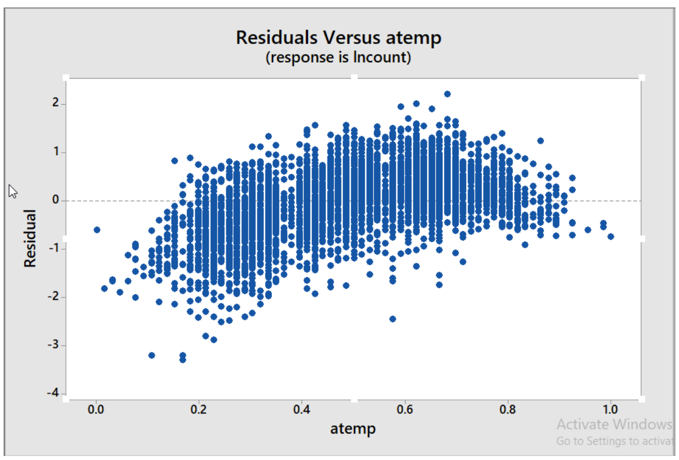

Figure 14 Residuals versus the actual feel temperature

The plot of the Residuals versus the actual feel temperature (atemp) shows a distinct increasing trend, though possibly quadratic, indicating that the atemp variable can be considered as a covariate.

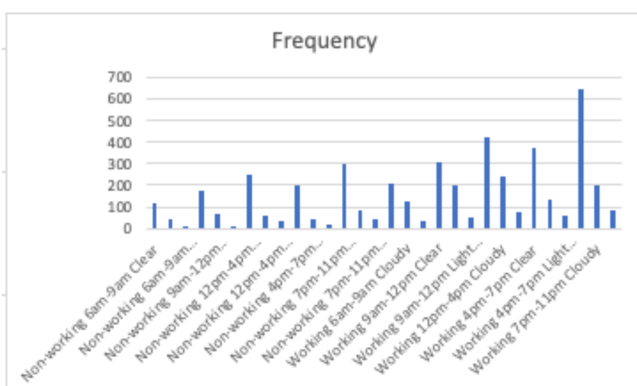

Figure 15: Group sizes for the different 3-way interaction categories

Mean Comparisons

From the ANOVA model, it was determined that the interaction of all the three categorical factors of time of day, weather, and working day is significant. In this factorial model, each of the 3-way interaction combinations is considered as a treatment. Each of these treatments (or combinations) are compared with each other using the Tukey means comparison to understand the effects of the combined factors.

weather	timeofday	workingday	Estimate	Standard Error	DF
Clear	4	1	6.0649	0.03509	4600
Clear	1	1	5.8795	0.04703	4600
Cloudy	4	1	5.7873	0.05831	4600
Cloudy	1	1	5.783	0.05966	4600
Clear	3	0	5.7581	0.04244	4600
Clear	4	0	5.5446	0.04822	4600
Cloudy	3	0	5.5439	0.08714	4600
Cloudy	4	0	5.4545	0.1054	4600
Clear	3	1	5.3515	0.0327	4600
Clear	2	0	5.3293	0.05059	4600
Clear	2	1	5.1887	0.03853	4600
Clear	5	1	5.1848	0.02646	4600
Cloudy	3	1	5.1168	0.04321	4600
Light Ra	1	1	5.0991	0.1095	4600
Cloudy	2	0	5.0901	0.08247	4600
Cloudy	2	1	5.0756	0.04773	4600
Light Ra	4	1	4.9695	0.08863	4600
Cloudy	5	1	4.8849	0.04749	4600
Light Ra	3	0	4.8041	0.1212	4600
Clear	5	0	4.7027	0.03923	4600
Cloudy	5	0	4.5509	0.07237	4600
Light Ra	4	0	4.375	0.1509	4600
Light Ra	5	1	4.2177	0.07454	4600
Light Ra	2	1	4.1613	0.09452	4600
Light Ra	3	1	4.0675	0.07693	4600
Light Ra	2	0	4.0614	0.1872	4600
Light Ra	5	0	3.9846	0.1042	4600
Clear	1	0	3.9018	0.06241	4600
Cloudy	1	0	3.8354	0.1006	4600
Light Ra	1	0	2.4531	0.2135	4600

Figure 16: Least Squares Means for each of the three-way interaction groups and sorted in descending order by the group means estimates.

workingday*weather*timeofday	N	Mean	Grouping
1 Clear 4	370	6.06489	A
1 Clear 1	206	5.87946	A B
1 Cloudy 4	134	5.78735	B C
1 Cloudy 1	128	5.78298	B C
0 Clear 3	253	5.75812	B C
0 Clear 4	196	5.54460	C D
0 Cloudy 3	60	5.54385	B C D E
0 Cloudy 4	41	5.45447	B C D E F G
1 Clear 3	426	5.35152	D E G
0 Clear 2	178	5.32935	D E F G
1 Clear 2	307	5.18867	E F G H
1 Clear 5	651	5.18484	F H
1 Cloudy 3	244	5.11677	F H I
1 Light Rain/Snow 1	38	5.09910	D E F G H I J
0 Cloudy 2	67	5.09012	F G H I
1 Cloudy 2	200	5.07564	F H I
1 Light Rain/Snow 4	58	4.96955	F H I J K L
1 Cloudy 5	202	4.88488	I J L
0 Light Rain/Snow 3	31	4.80406	H I J K L M
0 Clear 5	296	4.70271	J K L M
0 Cloudy 5	87	4.55094	K M N
0 Light Rain/Snow 4	20	4.37500	K L M N O
1 Light Rain/Snow 5	82	4.21766	N O
1 Light Rain/Snow 2	51	4.16125	N O
1 Light Rain/Snow 3	77	4.06750	O
0 Light Rain/Snow 2	13	4.06145	M N O
0 Light Rain/Snow 5	42	3.98457	O
0 Clear 1	117	3.90179	O
0 Cloudy 1	45	3.83542	O
0 Light Rain/Snow 1	10	2.45307	P

Figure 17 Tukey Mean Comparison Test Results

ANCOVA

Step 1: Are all regression slopes = 0? Checks for a linear relationship between the response and the covariate for each treatment group

Type of Day	Time of Day	Weather	Group size	p-value
Non-working	6am-9am	Clear	117	< 0.001
Non-working	6am-9am	Cloudy	45	0.0008
Non-working	6am-9am	Light Rain/Snow	10	0.9474
Non-working	9am-12pm	Clear	178	< 0.001
Non-working	9am-12pm	Cloudy	67	< 0.001
Non-working	9am-12pm	Light Rain/Snow	13	0.0988
Non-working	12pm-4pm	Clear	253	< 0.001
Non-working	12pm-4pm	Cloudy	60	< 0.001
Non-working	12pm-4pm	Light Rain/Snow	31	< 0.001
Non-working	4pm-7pm	Clear	196	< 0.001
Non-working	4pm-7pm	Cloudy	41	< 0.001
Non-working	4pm-7pm	Light Rain/Snow	20	0.0304
Non-working	7pm-11pm	Clear	296	< 0.001
Non-working	7pm-11pm	Cloudy	87	< 0.001
Non-working	7pm-11pm	Light Rain/Snow	42	0.0008
Working	6am-9am	Clear	206	< 0.001
Working	6am-9am	Cloudy	128	< 0.001
Working	6am-9am	Light Rain/Snow	38	0.0006

Working	9am-12pm	Clear	307	0.0001
Working	9am-12pm	Cloudy	200	< 0.001
Working	9am-12pm	Light Rain/Snow	51	0.0015
Working	12pm-4pm	Clear	426	< 0.001
Working	12pm-4pm	Cloudy	244	< 0.001
Working	12pm-4pm	Light Rain/Snow	77	< 0.001
Working	4pm-7pm	Clear	370	< 0.001
Working	4pm-7pm	Cloudy	134	< 0.001
Working	4pm-7pm	Light Rain/Snow	58	< 0.001
Working	7pm-11pm	Clear	651	< 0.001
Working	7pm-11pm	Cloudy	202	< 0.001
Working	7pm-11pm	Light Rain/Snow	82	< 0.001

Table 1: Checks for a linear relationship between the response variable (count) and the covariate for each three-way interactino group

Using a significance value of 0.05, only two of the 30 different treatment groups do not exhibit a significant linear relationship between the covariate and the response.

Step 2: Are the regression slopes all equal?

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
atemp	1	6385276	6385276	446.99	0.000
workingday	1	188919	188919	13.22	0.000
weather	2	467397	233698	16.36	0.000
timeofday	4	284244	71061	4.97	0.001
atemp*workingday	1	1264	1264	0.09	0.766
atemp*weather	2	89057	44528	3.12	0.044
atemp*timeofday	4	501984	125496	8.79	0.000
workingday*weather	2	128797	64398	4.51	0.011
workingday*timeofday	4	115381	28845	2.02	0.089
weather*timeofday	8	170949	21369	1.50	0.153
atemp*workingday*weather	2	89418	44709	3.13	0.044
atemp*workingday*timeofday	4	1078709	269677	18.88	0.000
atemp*weather*timeofday	8	126714	15839	1.11	0.354
workingday*weather*timeofday	8	102231	12779	0.89	0.520
atemp*workingday*weather*timeofday	8	79851	9981	0.70	0.693
Error	4570	65283112	14285		
Lack-of-Fit	940	21108906	22456	1.85	0.000
Pure Error	3630	44174207	12169		
Total	4629	149645553			

Figure 18: ANOVA table to check if the regression slopes of all the treatment groups are equal

The p-values corresponding to the interactions of atemp with the individual factors and with combinations of factors are found to be significantly greater than 0.05 suggesting that the interactions are not significant and that in general the slopes can be considered to be equal. Only in two cases, the atemp*timeofday and atemp*workingday*timeofday, there seems to be some interaction.

We initially ignore the presence of the two interactions and fit an equal slopes ANCOVA model without the interactions.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
atemp	1	21685165	21685165	1422.56	0.000
workingday	1	875408	875408	57.43	0.000
weather	2	2616466	1308233	85.82	0.000
timeofday	4	4981670	1245417	81.70	0.000
workingday*weather	2	41772	20886	1.37	0.254
workingday*timeofday	4	7280003	1820001	119.39	0.000
weather*timeofday	8	621077	77635	5.09	0.000
workingday*weather*timeofday	8	507189	63399	4.16	0.000
Error	4599	70105925	15244		
Lack-of-Fit	969	25931719	26761	2.20	0.000
Pure Error	3630	44174207	12169		
Total	4629	149645553			

Figure 19: ANOVA table for the ANCOVA model with equal slopes

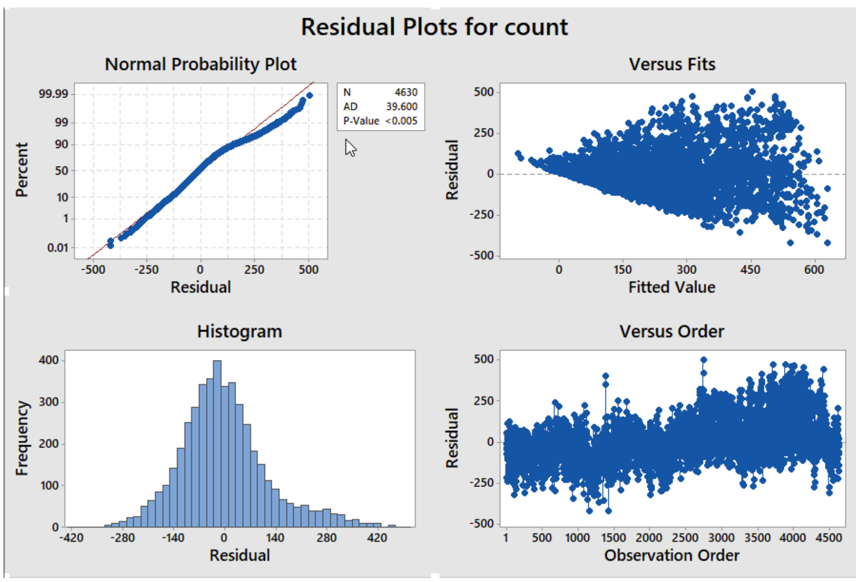

Figure 20: Residual plots of the ANCOVA model with equal slopes

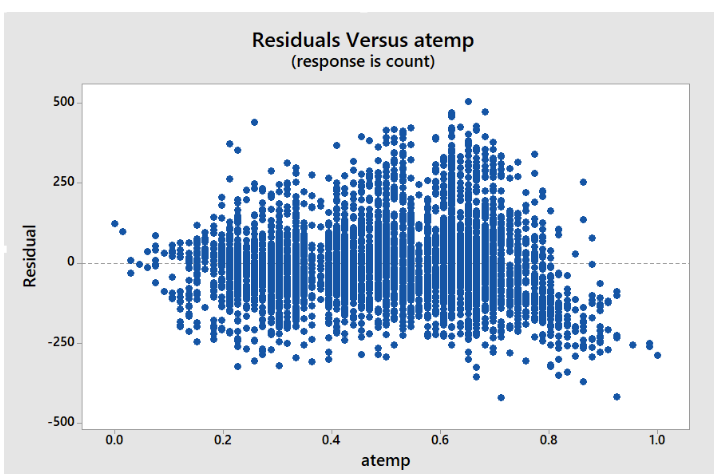

Figure 21: Plot of Residual vs atemp for the ANCOVA model with equal slopes.

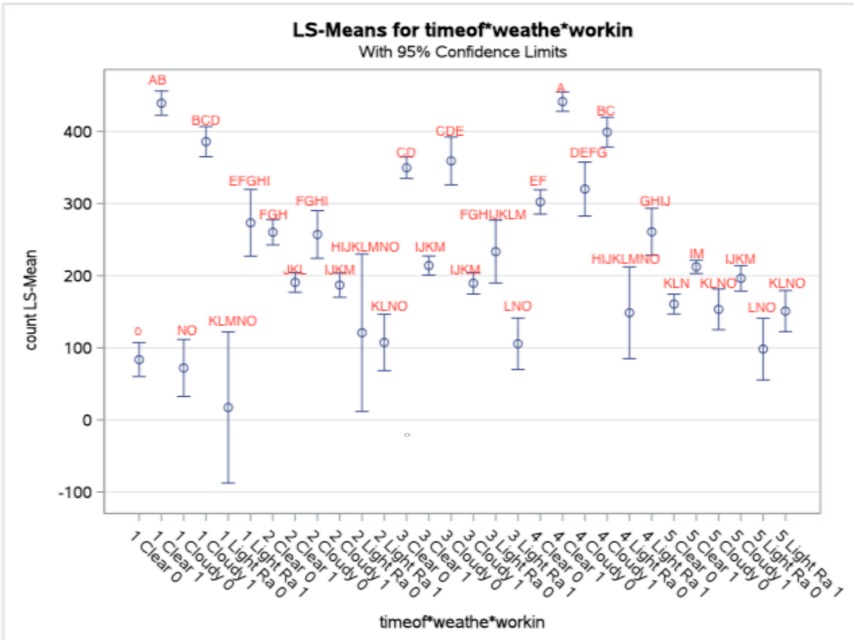

Figure 22: LSMeans Interval plot for an ANCOVA model with equal slopes

workingday*weather*timeofday	N	Mean	Grouping														
1 Clear 4	370	455.825	A														
1 Clear 1	206	433.229	A	B													
1 Cloudy 4	134	390.446		B	C												
1 Cloudy 1	128	383.886		B	C	D											
0 Clear 3	253	354.904			C	D											
0 Cloudy 3	60	332.900			C	D	E										
0 Cloudy 4	41	306.577				D	E	F	G								
0 Clear 4	196	306.084					E	F									
0 Clear 2	178	259.721						F	G	H							
1 Light Rain/Snow 1	38	258.264					E	F	G	H	I						
0 Cloudy 2	67	241.793						F	G	H	I						
1 Light Rain/Snow 4	58	236.016							G	H	I	J					
0 Light Rain/Snow 3	31	219.045						F	G	H	I	J	K	L	M		
1 Clear 5	651	213.461									I				M		
1 Cloudy 2	200	194.146									I	J	K		M		
1 Cloudy 5	202	193.965									I	J	K		M		
1 Cloudy 3	244	189.889									I	J	K		M		
1 Clear 3	426	187.016									I	J	K		M		
1 Clear 2	307	177.857										J	K	L			
0 Clear 5	296	160.810											K	L		N	
0 Light Rain/Snow 4	20	154.907								H	I	J	K	L	M	N	O
0 Cloudy 5	87	153.921											K	L		N	O
0 Light Rain/Snow 2	13	153.511								H	I	J	K	L	M	N	O
1 Light Rain/Snow 5	82	143.245											K	L		N	O
1 Light Rain/Snow 2	51	127.724											K	L		N	O
1 Light Rain/Snow 3	77	126.325												L		N	O
0 Light Rain/Snow 5	42	111.757												L		N	O
0 Clear 1	117	103.264															O
0 Cloudy 1	45	93.890														N	O
0 Light Rain/Snow 1	10	74.285											K	L	M	N	O

Figure 23: Tukey comparisons for the ANCOVA model with equal slopes

Regression Equations for ANCOVA model with equal slopes:

workingday	weather	timeofday			
0	Clear	1	count	=	-95.9 + 401.7 atemp
0	Clear	2	count	=	60.5 + 401.7 atemp
0	Clear	3	count	=	155.70 + 401.7 atemp
0	Clear	4	count	=	106.9 + 401.7 atemp
0	Clear	5	count	=	-38.39 + 401.7 atemp
0	Cloudy	1	count	=	-105.3 + 401.7 atemp
0	Cloudy	2	count	=	42.6 + 401.7 atemp
0	Cloudy	3	count	=	133.7 + 401.7 atemp
0	Cloudy	4	count	=	107.4 + 401.7 atemp
0	Cloudy	5	count	=	-45.3 + 401.7 atemp
0	Light Rain/Snow	1	count	=	-124.9 + 401.7 atemp
0	Light Rain/Snow	2	count	=	-45.7 + 401.7 atemp
0	Light Rain/Snow	3	count	=	19.8 + 401.7 atemp
0	Light Rain/Snow	4	count	=	-44.3 + 401.7 atemp
0	Light Rain/Snow	5	count	=	-87.4 + 401.7 atemp
1	Clear	1	count	=	234.03 + 401.7 atemp
1	Clear	2	count	=	-21.35 + 401.7 atemp
1	Clear	3	count	=	-12.19 + 401.7 atemp
1	Clear	4	count	=	256.62 + 401.7 atemp
1	Clear	5	count	=	14.26 + 401.7 atemp
1	Cloudy	1	count	=	184.7 + 401.7 atemp
1	Cloudy	2	count	=	-5.1 + 401.7 atemp
1	Cloudy	3	count	=	-9.31 + 401.7 atemp
1	Cloudy	4	count	=	191.2 + 401.7 atemp
1	Cloudy	5	count	=	-5.24 + 401.7 atemp
1	Light Rain/Snow	1	count	=	59.1 + 401.7 atemp
1	Light Rain/Snow	2	count	=	-71.5 + 401.7 atemp
1	Light Rain/Snow	3	count	=	-72.9 + 401.7 atemp
1	Light Rain/Snow	4	count	=	36.8 + 401.7 atemp
1	Light Rain/Snow	5	count	=	-56.0 + 401.7 atemp

Figure 24: Regression equations for the ANCOVA model with equal slopes

ANCOVA model with unequal slopes

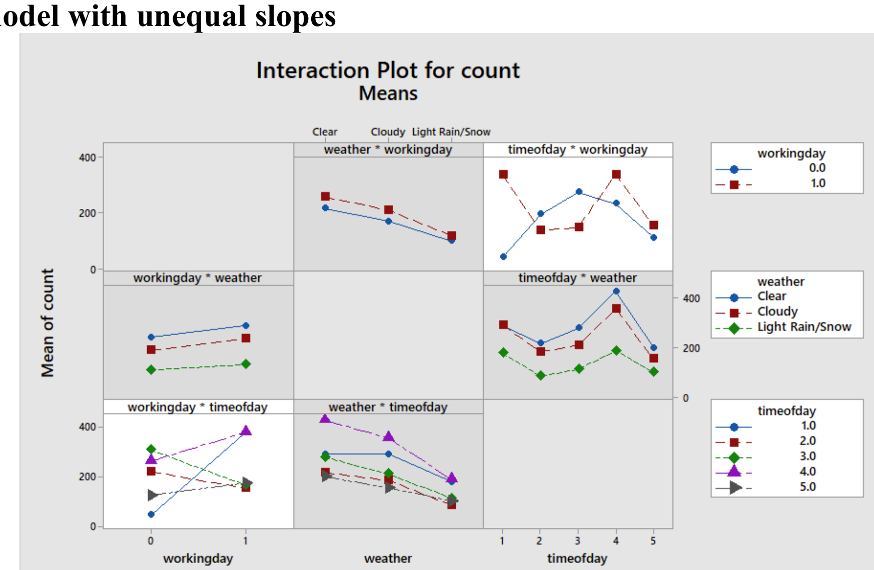

Figure 25: Interaction plot for count means

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
atemp	1	19199639	19199639	1329.31	0.000
workingday	1	806107	806107	55.81	0.000
weather	2	4051126	2025563	140.24	0.000
timeofday	4	878979	219745	15.21	0.000
atemp*workingday	1	79751	79751	5.52	0.019
atemp*timeofday	4	1565675	391419	27.10	0.000
workingday*timeofday	4	224455	56114	3.89	0.004
atemp*workingday*timeofday	4	2476919	619230	42.87	0.000
Error	4608	66554543	14443		
Lack-of-Fit	978	22380336	22884	1.88	0.000
Pure Error	3630	44174207	12169		
Total	4629	149645553			

Figure 26: ANOVA table for ANCOVA model with unequal slopes

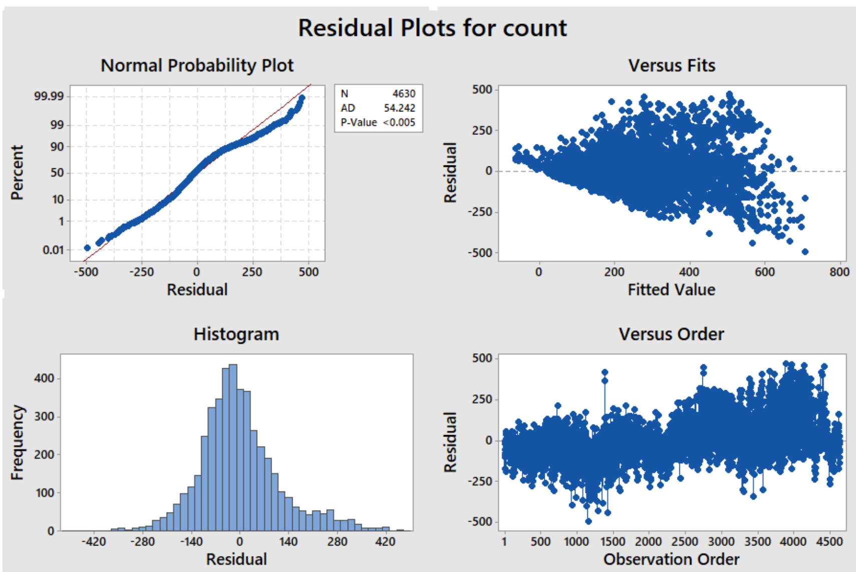

Figure 27: Residual plots for ANCOVA model with unequal slopes

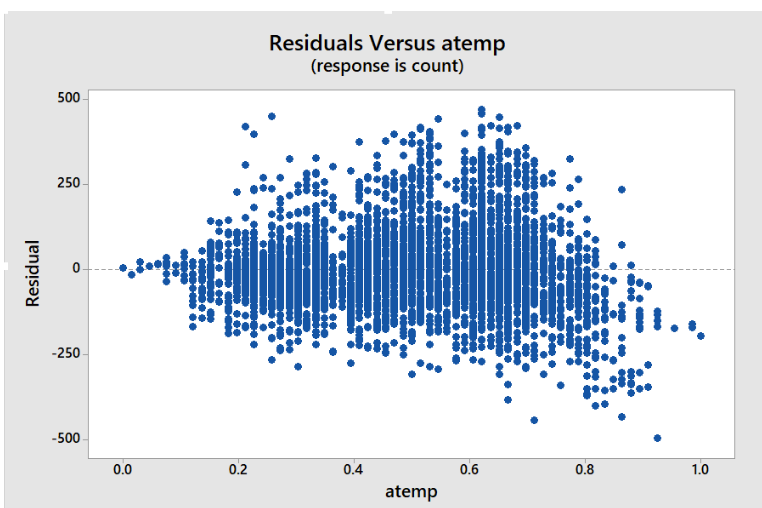

Figure 28: Residual vs atemp for ANCOVA model with unequal slopes

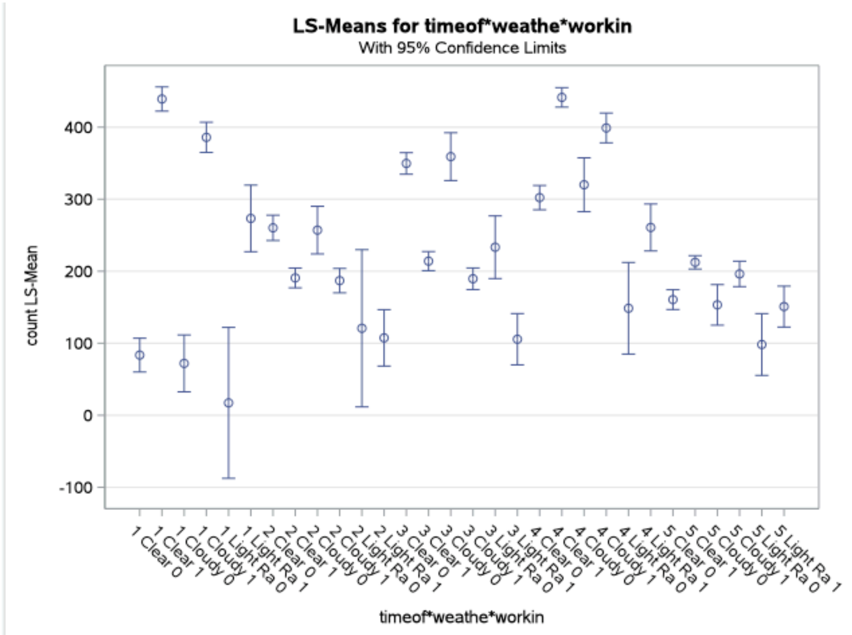

Figure 29: LS Means Interval Plot for ANCOVA model with unequal slopes

Grouping Information Using the Tukey Method and 95% Confidence

workingday*timeofday	N	Mean	Grouping							
1 4	562	380.633	A							
1 1	372	378.980	A							
0 3	344	308.557		B						
0 4	257	263.241			C					
0 2	258	221.031				D				
1 5	935	174.705					E			
1 3	747	167.018					E	F		
1 2	558	153.916						F		
0 5	425	125.989							G	
0 1	172	46.151								H

Means that do not share a letter are significantly different.

Figure 30: Tukey Comparison chart for ANCOVA model with unequal slopes

workingday	weather	timeofday			
0	Clear	1	count	=	20.7 + 135.3 atemp
0	Clear	2	count	=	25.6 + 478.2 atemp
0	Clear	3	count	=	55.0 + 595.3 atemp
0	Clear	4	count	=	8.1 + 598.5 atemp
0	Clear	5	count	=	-18.9 + 376.2 atemp
0	Cloudy	1	count	=	3.1 + 135.3 atemp
0	Cloudy	2	count	=	7.9 + 478.2 atemp
0	Cloudy	3	count	=	37.4 + 595.3 atemp
0	Cloudy	4	count	=	-9.5 + 598.5 atemp
0	Cloudy	5	count	=	-36.5 + 376.2 atemp
0	Light Rain/Snow	1	count	=	-86.6 + 135.3 atemp
0	Light Rain/Snow	2	count	=	-81.8 + 478.2 atemp
0	Light Rain/Snow	3	count	=	-52.4 + 595.3 atemp
0	Light Rain/Snow	4	count	=	-99.2 + 598.5 atemp
0	Light Rain/Snow	5	count	=	-126.2 + 376.2 atemp
1	Clear	1	count	=	160.4 + 524.8 atemp
1	Clear	2	count	=	134.9 + 122.4 atemp
1	Clear	3	count	=	122.3 + 174.2 atemp
1	Clear	4	count	=	94.6 + 660.8 atemp
1	Clear	5	count	=	-2.3 + 441.0 atemp
1	Cloudy	1	count	=	142.8 + 524.8 atemp
1	Cloudy	2	count	=	117.2 + 122.4 atemp
1	Cloudy	3	count	=	104.7 + 174.2 atemp
1	Cloudy	4	count	=	77.0 + 660.8 atemp
1	Cloudy	5	count	=	-20.0 + 441.0 atemp
1	Light Rain/Snow	1	count	=	53.1 + 524.8 atemp
1	Light Rain/Snow	2	count	=	27.5 + 122.4 atemp
1	Light Rain/Snow	3	count	=	14.9 + 174.2 atemp
1	Light Rain/Snow	4	count	=	-12.8 + 660.8 atemp
1	Light Rain/Snow	5	count	=	-109.7 + 441.0 atemp

Figure 31: Regression Equations for the ANCOVA model with unequal slopes

Regression equations for ANCOVA model with unequal slopes and including all the interaction terms:

workingday	weather	timeofday			
0	Clear	1	count	=	9.0 + 150.3 atemp
0	Clear	2	count	=	26.0 + 472.3 atemp
0	Clear	3	count	=	71.7 + 560.6 atemp
0	Clear	4	count	=	13.5 + 582.2 atemp
0	Clear	5	count	=	-34.9 + 394.2 atemp
0	Cloudy	1	count	=	4.4 + 136 atemp
0	Cloudy	2	count	=	-22.4 + 563.7 atemp
0	Cloudy	3	count	=	-13.6 + 751.7 atemp
0	Cloudy	4	count	=	-70.2 + 787 atemp
0	Cloudy	5	count	=	-42.0 + 393.8 atemp
0	Light Rain/Snow	1	count	=	16 + 2 atemp
0	Light Rain/Snow	2	count	=	8.9 + 226 atemp
0	Light Rain/Snow	3	count	=	-124.3 + 721 atemp
0	Light Rain/Snow	4	count	=	-17.1 + 334 atemp
0	Light Rain/Snow	5	count	=	-21.1 + 241 atemp
1	Clear	1	count	=	173.0 + 536.6 atemp
1	Clear	2	count	=	146.2 + 89.8 atemp
1	Clear	3	count	=	148.3 + 132.5 atemp
1	Clear	4	count	=	155.0 + 577.5 atemp
1	Clear	5	count	=	-9.3 + 447.0 atemp
1	Cloudy	1	count	=	147.2 + 481.5 atemp
1	Cloudy	2	count	=	88.3 + 199.0 atemp
1	Cloudy	3	count	=	76.6 + 227.7 atemp
1	Cloudy	4	count	=	68.5 + 666.5 atemp
1	Cloudy	5	count	=	-22.8 + 441.8 atemp
1	Light Rain/Snow	1	count	=	-7.4 + 566 atemp
1	Light Rain/Snow	2	count	=	13.5 + 189 atemp
1	Light Rain/Snow	3	count	=	-5.3 + 224 atemp
1	Light Rain/Snow	4	count	=	-162.1 + 852.9 atemp
1	Light Rain/Snow	5	count	=	-101.5 + 509.0 atemp

Regression Equations from a Multi Regression Model with all interactions:

Regression Equation

weather	workingday	timeofday			
Clear	0	1	count	=	9.0 + 150.3 atemp
Clear	0	2	count	=	26.0 + 472.3 atemp
Clear	0	3	count	=	71.7 + 560.6 atemp
Clear	0	4	count	=	13.5 + 582.2 atemp
Clear	0	5	count	=	-34.9 + 394.2 atemp
Clear	1	1	count	=	173.0 + 536.6 atemp
Clear	1	2	count	=	146.2 + 89.8 atemp
Clear	1	3	count	=	148.3 + 132.5 atemp
Clear	1	4	count	=	155.0 + 577.5 atemp
Clear	1	5	count	=	-9.3 + 447.0 atemp
Cloudy	0	1	count	=	4.4 + 136 atemp
Cloudy	0	2	count	=	-22.4 + 563.7 atemp
Cloudy	0	3	count	=	-13.6 + 751.7 atemp
Cloudy	0	4	count	=	-70.2 + 787 atemp
Cloudy	0	5	count	=	-42.0 + 393.8 atemp
Cloudy	1	1	count	=	147.2 + 481.5 atemp
Cloudy	1	2	count	=	88.3 + 199.0 atemp
Cloudy	1	3	count	=	76.6 + 227.7 atemp
Cloudy	1	4	count	=	68.5 + 666.5 atemp
Cloudy	1	5	count	=	-22.8 + 441.8 atemp
Light Rain/Snow	0	1	count	=	16 + 2 atemp
Light Rain/Snow	0	2	count	=	8.9 + 226 atemp
Light Rain/Snow	0	3	count	=	-124.3 + 721 atemp
Light Rain/Snow	0	4	count	=	-17.1 + 334 atemp
Light Rain/Snow	0	5	count	=	-21.1 + 241 atemp
Light Rain/Snow	1	1	count	=	-7.4 + 566 atemp
Light Rain/Snow	1	2	count	=	13.5 + 189 atemp
Light Rain/Snow	1	3	count	=	-5.3 + 224 atemp
Light Rain/Snow	1	4	count	=	-162.1 + 852.9 atemp
Light Rain/Snow	1	5	count	=	-101.5 + 509.0 atemp

The regression equation for an ANCOVA model with all interactions are the same as obtained with a multi regression model.

SAS Code: (Please note that some of the plots presented in this report were generated from Minitab).

```
data bikerentals;
		infile '/folders/myfolders/stat502/sascode/bikecount.csv'
				dsd delimiter=',' missover firstobs=2;
   input workingday weather $ atemp humidity windspeed count hour timeofday;
   * timeofday hour weather $ atemp humidity windspeed count;
   lncount = log(count);
run;

proc freq data=bikerentals;
tables weather;
run;

proc print data=bikerentals(obs=20);
title "Raw Data";
run;

proc freq data=bikerentals;
title "Frequency Table";
tables hour timeofday workingday weather timeofday*workingday /nocum;
*weight count;
run;

proc summary data=bikerentals;
var atemp humidity windspeed;
output out=qsum1;
run;

proc print data=qsum1;
title "Summary";
run;

ods graphics on;

/* ANOVA model using hour as the factor */
proc mixed data=bikerentals method=type3;
title "ANOVA";
class workingday hour weather;
model count = workingday hour weather hour*workingday hour*weather hour*workingday*weather;
store outhour;
run;

proc plm restore=outhour;
lsmeans workingday*hour*weather /adjust=tukey plot=meanplot cl lines;
*lsmeans workingday*hour /adjust=tukey plot=meanplot cl lines;
*lsmeans weather /adjust=tukey plot=meanplot cl lines;
ods exclude diffs diffplot;
run;

/* ANOVA model using timeof day as the factor */
title "ANOVA with interaction";
proc mixed data=bikerentals method=type3;* plots=all;
class timeofday workingday weather;
model count = workingday timeofday weather
		workingday*timeofday workingday*weather timeofday*weather
```

```
          workingday*timeofday*weather;
store outaday;
run;

proc plm restore=outaday;
lsmeans workingday*weather*timeofday /adjust=tukey plot=meanplot cl lines;
ods exclude diffs diffplot;
run;

title "ANOVA with interaction and ln values";
proc mixed data=bikerentals method=type3;* plots=all;
class timeofday workingday weather;
model lncount = workingday timeofday weather
          workingday*timeofday workingday*weather timeofday*weather
          workingday*timeofday*weather;
store outlnday;
run;

proc plm restore=outlnday;
lsmeans workingday*weather*timeofday /adjust=tukey plot=meanplot cl lines;
ods exclude diffs diffplot;
run;

title "ancova";
*test for working day and temperature;
*Step 1: Check if regression slopes are zero;
*non-working day;
proc mixed data=bikerentals;
where workingday=0 and timeofday=1 and weather="Clear";
model count=atemp;
title "01Clear";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=1 and weather="Cloudy";
model count=atemp;
title "01Cloudy";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=1 and weather="Light Ra";
model count=atemp;
title "01Rain";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=2 and weather="Clear";
model count=atemp;
title "02Clear";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=2 and weather="Cloudy";
model count=atemp;
title "02Cloudy";
run;
```

```
proc mixed data=bikerentals;
where workingday=0 and timeofday=2 and weather="Light Ra";
model count=atemp;
title "02Rain";

proc mixed data=bikerentals;
where workingday=0 and timeofday=3 and weather="Clear";
model count=atemp;
title "03Clear";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=3 and weather="Cloudy";
model count=atemp;
title "03Cloudy";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=3 and weather="Light Ra";
model count=atemp;
title "03Rain";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=4 and weather="Clear";
model count=atemp;
title "04Clear";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=4 and weather="Cloudy";
model count=atemp;
title "04Cloudy";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=4 and weather="Light Ra";
model count=atemp;
title "04Rain";

proc mixed data=bikerentals;
where workingday=0 and timeofday=5 and weather="Clear";
model count=atemp;
title "05Clear";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=5 and weather="Cloudy";
model count=atemp;
title "05Cloudy";
run;

proc mixed data=bikerentals;
where workingday=0 and timeofday=5 and weather="Light Ra";
model count=atemp;
title "05Rain";
```

```
*working day;
proc mixed data=bikerentals;
where workingday=1 and timeofday=1 and weather="Clear";
model count=atemp;
title "11Clear";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=1 and weather="Cloudy";
model count=atemp;
title "11Cloudy";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=1 and weather="Light Ra";
model count=atemp;
title "11Rain";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=2 and weather="Clear";
model count=atemp;
title "12Clear";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=2 and weather="Cloudy";
model count=atemp;
title "12Cloudy";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=2 and weather="Light Ra";
model count=atemp;
title "12Rain";

proc mixed data=bikerentals;
where workingday=1 and timeofday=3 and weather="Clear";
model count=atemp;
title "13Clear";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=3 and weather="Cloudy";
model count=atemp;
title "11Cloudy";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=3 and weather="Light Ra";
model count=atemp;
title "13Rain";
run;

proc mixed data=bikerentals;
```

```
where workingday=1 and timeofday=4 and weather="Clear";
model count=atemp;
title "14Clear";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=4 and weather="Cloudy";
model count=atemp;
title "14Cloudy";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=4 and weather="Light Ra";
model count=atemp;
title "14Rain";

proc mixed data=bikerentals;
where workingday=1 and timeofday=5 and weather="Clear";
model count=atemp;
title "15Clear";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=5 and weather="Cloudy";
model count=atemp;
title "15Cloudy";
run;

proc mixed data=bikerentals;
where workingday=1 and timeofday=5 and weather="Light Ra";
model count=atemp;
title "15Rain";
run;


title; run;
*Step 2: Check if regression slopes are equal;

proc mixed data =bikerentals method=type3;
class workingday;
model count=workingday atemp workingday*atemp;
run;

* interaction is significant implying that the slopes are not equal;
* check values at different points of temperature;
proc mixed data =bikerentals method=type3;
class workingday;
model count=workingday workingday*atemp /noint solution;
run;

title "ANOCOVA with all interactions";
proc mixed data =bikerentals method=type3 plots= residualpanel;
class timeofday weather workingday;
model count=timeofday workingday weather atemp
        timeofday*workingday timeofday*weather workingday*weather timeofday*workingday*weather
        atemp*workingday atemp*timeofday atemp*weather
```

```
            atemp*workingday*timeofday atemp*workingday*weather atemp*timeofday*weather
            atemp*workingday*timeofday*weather /noint solution;
lsmeans workingday*timeofday*weather/pdiff at atemp=0.25;
lsmeans workingday*timeofday*weather/pdiff at atemp=0.50;
lsmeans workingday*timeofday*weather/pdiff at atemp=0.75;
lsmeans workingday*timeofday*weather/pdiff at atemp=0.95;
store outinanc;
run;

proc plm restore=outinanc;
lsmeans workingday*weather*timeofday /adjust=tukey plot=meanplot cl lines;
ods exclude diffs diffplot;
run;

title "ANOCOVA without interactions";
proc mixed data =bikerentals method=type3;
class timeofday weather workingday;
model count=timeofday workingday weather atemp
            timeofday*workingday timeofday*weather workingday*weather timeofday*workingday*weather
             /noint solution;
store outanc;
run;

proc plm restore=outanc;
lsmeans workingday*weather*timeofday /adjust=tukey plot=meanplot cl lines;
ods exclude diffs diffplot;
run;

title "ANOCOVA with interactions and reduced params";
proc mixed data =bikerentals method=type3 plots= residualpanel;
class timeofday weather workingday;
model count=timeofday workingday weather atemp
            timeofday*workingday
            atemp*workingday atemp*timeofday
            atemp*workingday*timeofday  /noint solution;
lsmeans workingday*timeofday/pdiff at atemp=0.25;
lsmeans workingday*timeofday/pdiff at atemp=0.50;
lsmeans workingday*timeofday/pdiff at atemp=0.75;
lsmeans workingday*timeofday/pdiff at atemp=0.95;
store outintr;
run;

proc plm restore=outintr;
lsmeans workingday*timeofday /adjust=tukey plot=meanplot cl lines;
ods exclude diffs diffplot;
run;
```