SEOUL BIKE HIRE CASE STUDY

INTRODUCTION

Rental bikes have been introduced into many cities to provide quick and affordable travel for users that is environmentally friendly and doesn't contribute to road traffic.

Providing a stable supply of bikes to meet fluctuating demand is important to ensure the service remains reliable.

This report will use bike sharing demand data from Seoul to model the effects on rental bike demand.

DATA DESCRIPTION

The data set contains a count of public bike rental at peak demand hour (1800-1900) in the Seoul Bike Hiring System across 353 days. The dataset encompasses weather data, which includes Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, and Rain, along with information on the number of bikes rented during peak hours and corresponding dates, as well as details about the seasons.

Variable descriptions can be found in Table 1.

Table 1 - Variable Descriptions

Variable	Description
date	date-month-year
rented	peak rented bike count
temperature	temperature in Celsius
humidity	Relative humidity (%)
windspeed	Wind speed in m/s
visibility	visibility (multiples of 10m)
dewpoint	Dew point temperature in Celsius
SolarRadiation	Solar radiation (MJ/m2)
Rain	1: rainy, 0: no rain
Snowfall	Snow fall (cm)
seasons	Winter, Spring, Summer, Autumn
Holiday	Yes; No
wkday	Day of the week Monday – Sunday

RENTAL TRENDS BY WEEKDAY

In this analysis, the effect of weekday will be assessed using analysis of variance (ANOVA).

It stands to reason that bike usage may vary on different days, particularly between weekdays (Monday to Friday) and weekends (Saturday and Sunday) due to city working and commuting patterns. This analysis will specifically test to see if there is a statistically significant difference between weekdays and weekends.

Prior to building the ANOVA model, the data was inspected and several assumptions (required for a valid ANOVA model) were checked.

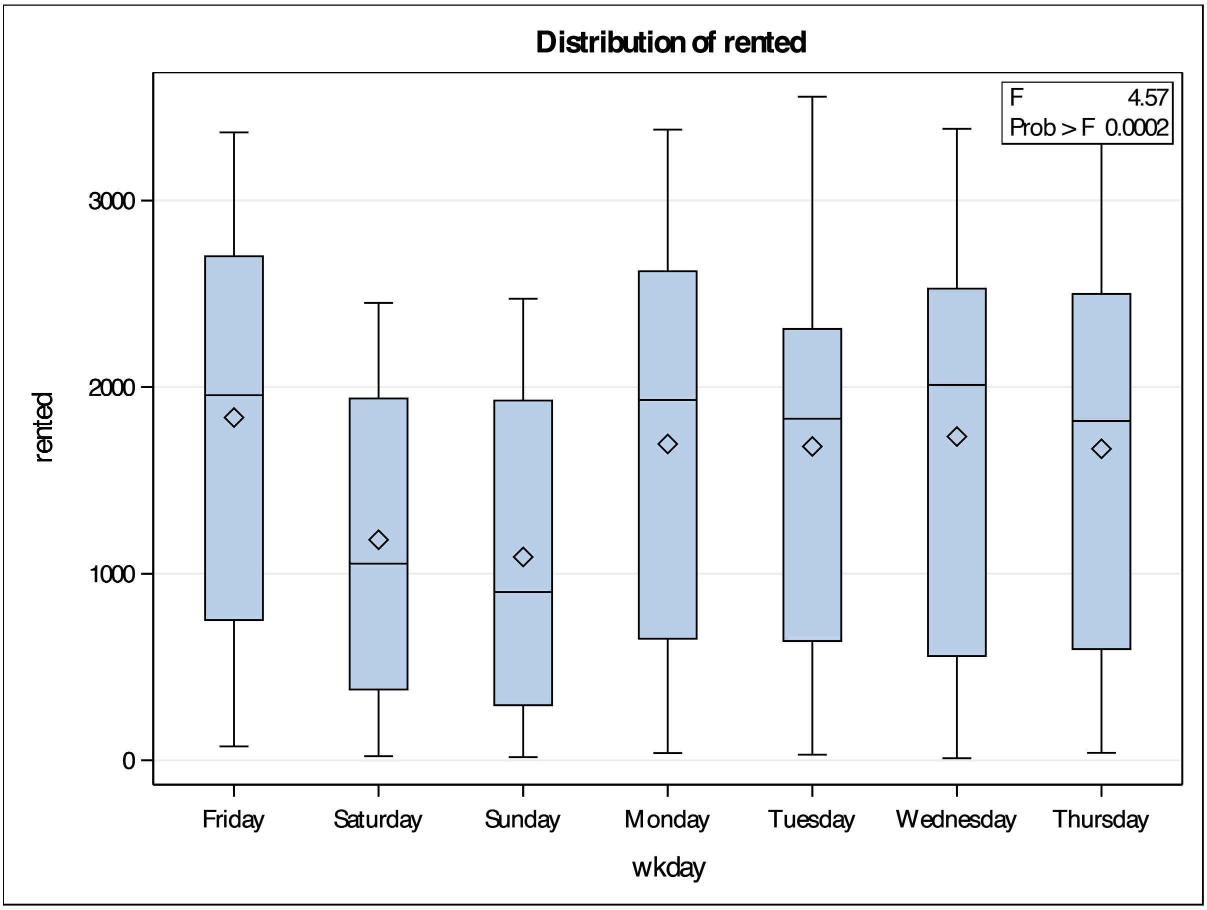
Weekday Summary Statistics

Summary statistics of the data aggregated by *wkday* are presented below.

Table 2 – Bikes *Rented* by Weekday Summary Statistics

wkday	N Obs	Mean	Std Dev	Median	Minimum	Maximum	Skewness	Kurtosis
Monday	52	1695	1064	1930	39	3380	-0.095	-1.404
Tuesday	48	1682	1022	1831	30	3556	-0.021	-1.140
Wednesday	50	1735	1076	2012	11	3384	-0.135	-1.462
Thursday	50	1669	1034	1818	40	3418	-0.057	-1.420
Friday	51	1837	994	1956	74	3365	-0.199	-1.373
Saturday	51	1182	793	1054	22	2451	0.094	-1.501
Sunday	51	1089	832	902	17	2474	0.317	-1.404

Figure 1 - Weekday Boxplots



There are observable differences between weekdays and weekends with weekends having notably lower (lower means & medians) and more consistent counts (lower standard deviations).

Relative to the average number of bikes rented, the variability is high as is the range (the difference between the minimum and maximum values). High day to day variability suggests there are other relevant factors that affect bike rentals beyond day of the week.

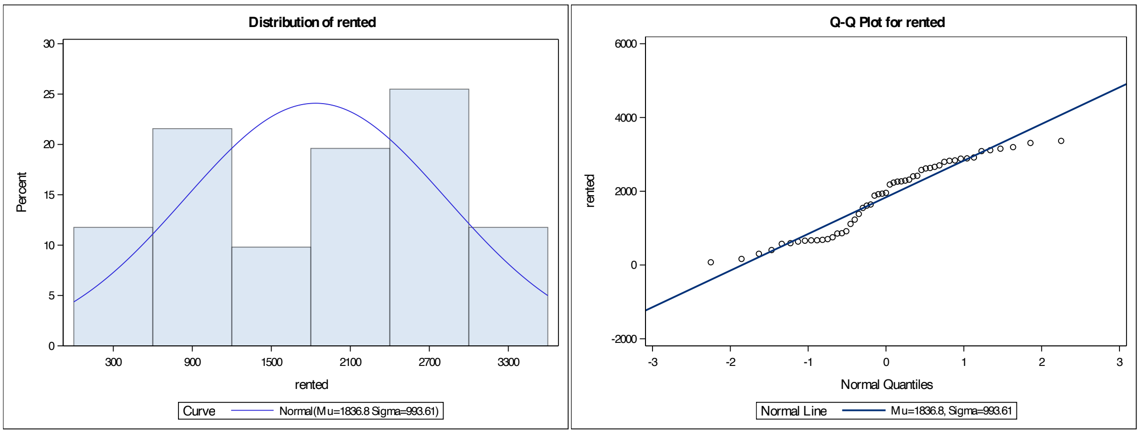
The symmetry of the distribution, skewness, is different depending on whether it is a weekday or weekend. Weekdays tend towards negative or left skewness which implies which suggests that on most weekdays, most rental counts are concentrated towards the higher end of the distribution, with fewer instances of extremely low rentals. The reverse appears true on weekends.

Weekday Distribution

For a valid ANOVA, data should be normally distributed. To assess normality, histograms and QQ plots will be inspected along with formal goodness-of-fit tests for normality.

Sample plots are presented here, all plots can be found in Appendix A

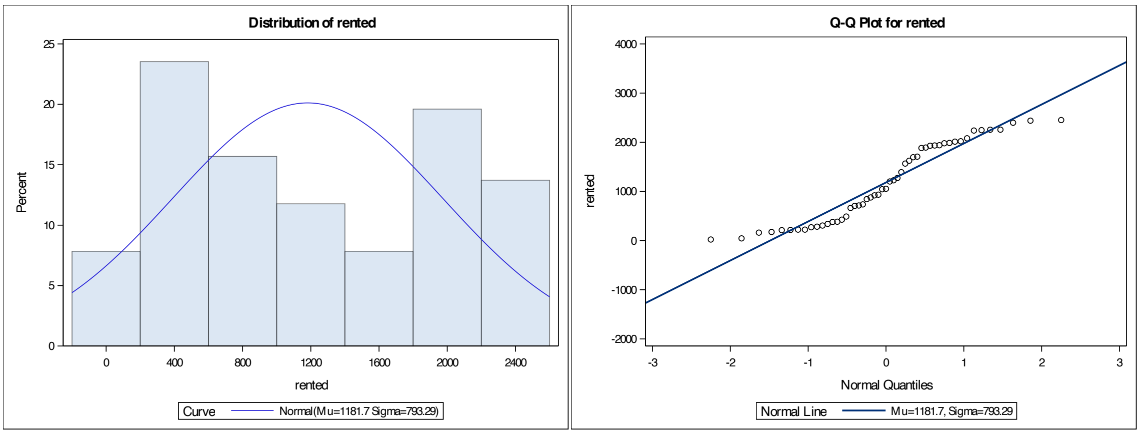
Figure 2 - Histogram and QQ plot of the Distribution of Bike Rentals on Fridays



The histogram shows evidence of two distinct populations, a bimodal distribution with peaks at approximately 900 and 2700, which likely represents a departure from a normal distribution.

Evidence of the bimodal distribution can be seen in the QQ plot also. The diagonal line represents a normal distribution, and the points deviate above and below the line around zero.

Figure 3 - Histogram and QQ plot of the Distribution of Bike Rentals on Saturdays



Similar patterns are observed on Saturdays where bimodal distributions are observed again. Similar effects are observed in most days (Appendix A).

Distributions were then tested using goodness-of-fit tests for normality (Table 3). The null hypothesis (H_0) for each test assume that the data follows a normal distribution.

Table 3 - Goodness-of-fit Tests for Normal Distribution (p = 0.05)

Test	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Kolmogorov-Smirnov	0.020	0.077	<0.010	<0.010	0.018	<0.010	<0.010
Cramer-von Mises	<0.005	0.043	<0.005	<0.005	<0.005	<0.005	<0.005
Anderson-Darling	<0.005	0.034	<0.005	<0.005	<0.005	<0.005	<0.005

Using a significance level of 0.05, all days show some degree of departure from normality.

ANOVA typically requires data to be normally distributed, so the data set as is may not be suitable. Attempts were made to transform the data (Appendix B) to see if distributions were normal on alternative scales. however no normal distributions were observed.

WEEKDAY ANOVA

A one-way ANOVA with a contrast test was performed to investigate the association between weekday and bike rentals.

Table 4 - ANOVA Summary for Bike Rentals by Weekday

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	26265699.2	4377616.5	4.57	0.0002
Error	346	331351550.7	957663.4		
Corrected Total	352	357617249.9			

R-Square	Coeff Var	Root MSE	rented Mean
0.073446	62.97246	978.6028	1554.017

The ANOVA results indicate a statistically significant relationship between weekday and bike rentals, with a corresponding p-value of 0.0002. This suggests that at least one weekday has a different effect on bike rentals compared to the others.

The fit of the model to the data is poor, however. The R-Square value, 0.073, indicates that approximately 7.3% of the variance in bike rentals can be attributed to the day they are hired meaning that other factors, not included in the model, likely have a notable role to play.

A contrast test was performed by assigning weights within the categorical variables (days) to assign each a group: 'Weekday' or 'Weekend'. The means of the groups are compared with a H_0 that there is no difference in the group means.

Table 5 - Weekdays vs Weekends Contrast Test

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Weekdays vs. Weekends	1	25061808.64	25061808.64	26.17	<.0001

The contrast "Weekdays vs. Weekends" revealed a significant p-value ($p < 0.0001$), Table 5, indicating that there are significant differences in bike rentals between weekdays and weekends.

Post-hoc analysis, pairwise comparisons of the least squares mean of each day of the week, were performed to establish which days had statistically significant differences to one another. Pair-wise comparisons were performed using Tukey-Kramer's adjustment for multiple comparisons. The H_0 for this comparison is that no difference exists and a significance value of $p = 0.05$ is used to reject the null hypothesis.

Figure 4 - LS Means Using Tukey-Kramer Adjustment for Multiple Comparisons

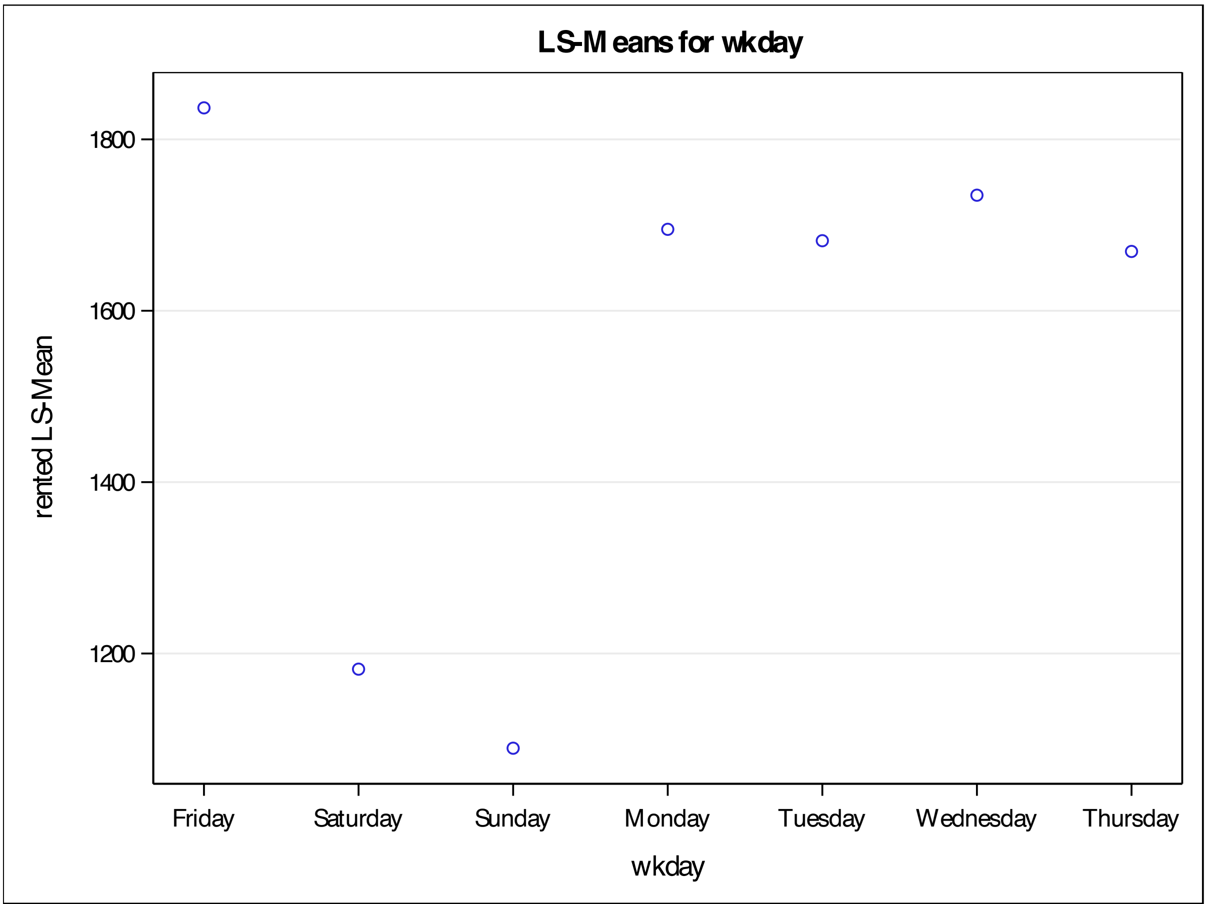


Table 6 - Pairwise Comparisons of Least Squares Means for Weekday (wkday) on Bike Rentals

Least Squares Means for effect wkday
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: rented

i/j	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
Friday		**0.0141**	**0.0026**	0.9903	0.9860	0.9985	0.9780
Saturday	0.0141		0.9991	0.1114	0.1482	0.0706	0.1614
Sunday	0.0026	0.9991		0.0301	0.0442	0.0175	0.0485
Monday	0.9903	0.1114	0.0301		1.0000	1.0000	1.0000
Tuesday	0.9860	0.1482	0.0442	1.0000		1.0000	1.0000
Wednesday	0.9985	0.0706	0.0175	1.0000	1.0000		0.9999
Thursday	0.9780	0.1614	0.0485	1.0000	1.0000	0.9999	

= $p < 0.005$

Post hoc analysis shows that the difference between weekends and weekdays is being driven by Sunday; pair-wise comparisons show that the difference between Sunday and all weekdays is statistically significant whereas Saturday varies to the rest of the week by a far lesser degree. The difference between Saturday and Friday is statistically significant, however Saturday does not appear different to the other days of the week.

WEEKDAY ANOVA DIAGNOSTICS

As a parametric test, ANOVA assumes that the data adheres to certain key assumptions: normality, homogeneity of variances, and independence of observations.

Earlier observations indicated deviations from normality. Further assessment of normality was conducted using residuals from the ANOVA. Homogeneity of variances, referred to as homoscedasticity, was evaluated through the analysis of residuals in conjunction with Levene's Test, where the null hypothesis (H_0) assumes equal variances (p = 0.05).

Figure 5 - One-Way ANOVA for Rental by Weekday Diagnostics

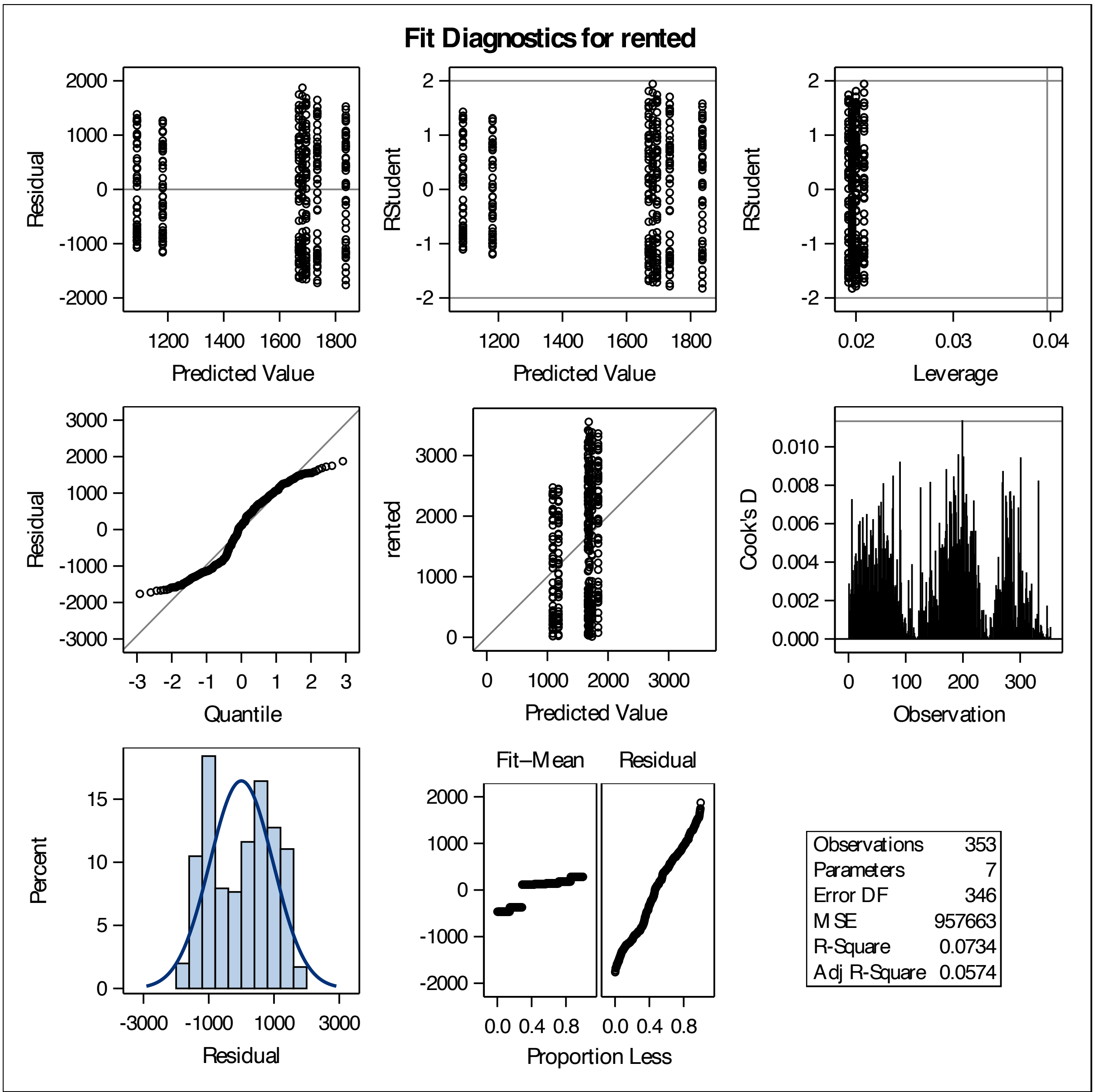


Table 7 - Levene's Test for Homogeneity of *rented* Variance in the One-Way ANOVA Model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
wkday	6	1.317E13	2.195E12	3.67	0.0015
Error	346	2.067E14	5.974E11		

The residual QQ plot and histogram both reveal the same non-normal, bimodal distribution, supporting earlier conclusions regarding the normality of the data.

The predicted values vs residual plot shows evidence of varying distributions as the scatter around zero showed inconsistency across the range of predicted values. The p-value for the Levene's test is <0.05 and the assumption of equal variances is rejected.

No outliers or high leverage points (RStudent) were identified, or influential observations (Cook's D) were noted. The model was not unduly affected by outlier results.

As each data point represents a single point in time (day), it is reasonable to assume that the data points are independent of one another.

The ANOVA model does not satisfy two of its key assumptions, normal distribution and homogeneity of variance. The potential consequences of this are that the results the parameter estimates may be biased which may in turn lead to inaccurate or unreliable conclusions when using the model.

As this ANOVA model is not suitable for analysis and efforts to improve the normality using transformations were unsuccessful, the model will be extended to include additional variables with the aim for accounting for more of the variance in bike rental numbers.

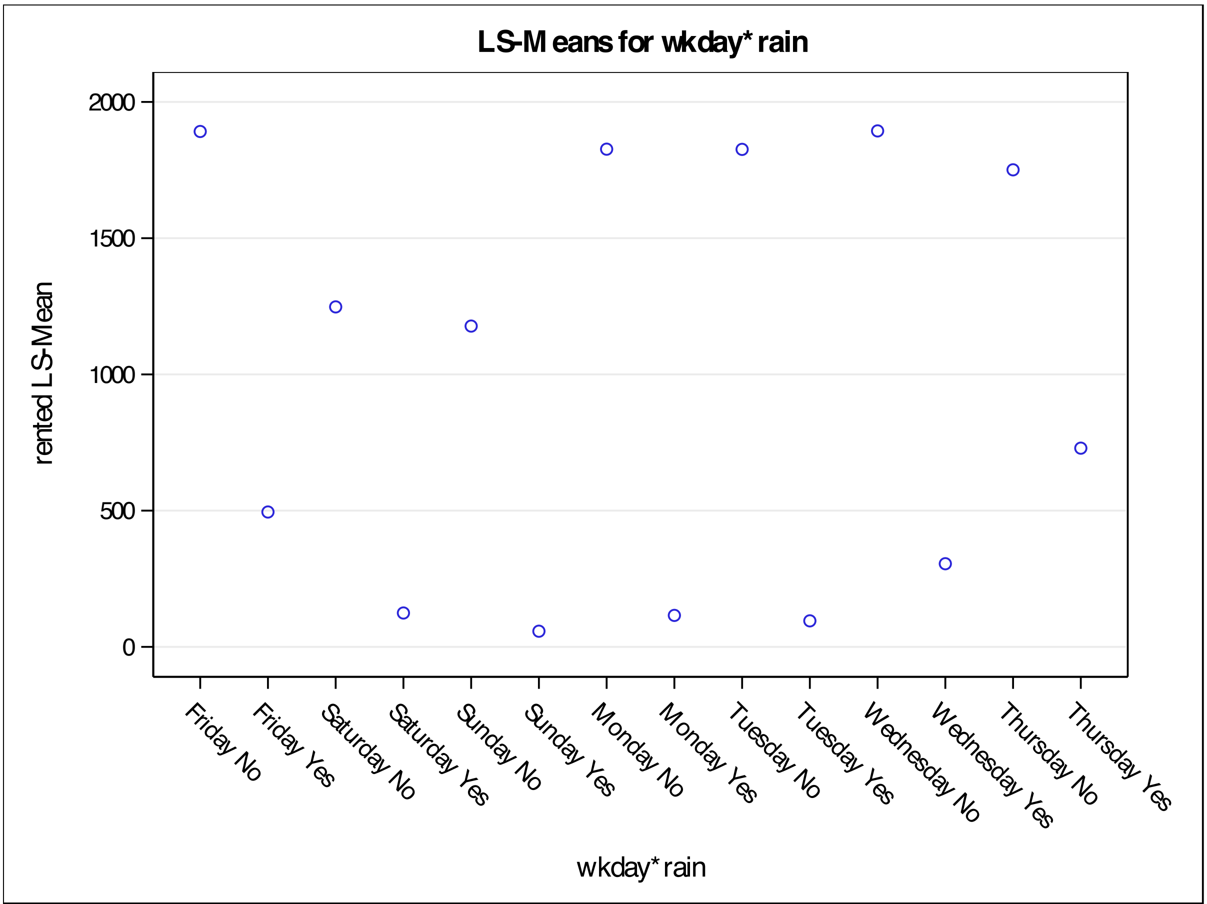
WEEKDAY FACTORIAL ANOVA

A factorial (Two-Way) ANOVA, an extension of a one-way ANOVA, was used to investigate whether there is evidence of an interaction between *wkday* and *rain*.

Table 8 - Bikes Rented by Weekday and Rain Summary Statistics

wkday	*rain*	N Obs	Mean	Std Dev	Minimum	Maximum	Median	Skewness	Kurtosis
Monday	No	48	1827	999.1	218	3380	2005	-0.188	-1.310
	Yes	4	116	56.1	39	170	127	-1.026	1.155
Tuesday	No	44	1826	941.4	351	3556	1980	-0.048	-1.050
	Yes	4	96	51.5	30	156	98	-0.290	1.517
Wednesday	No	45	1894	1009.4	297	3384	2162	-0.294	-1.327
	Yes	5	305	368.5	11	925	204	1.635	2.778
Thursday	No	46	1751	1018.7	75	3418	1884	-0.153	-1.384
	Yes	4	729	782.7	40	1850	514	1.462	2.591
Friday	No	49	1892	971.0	165	3365	2183	-0.251	-1.349
	Yes	2	495	595.4	74	916	495	.	.
Saturday	No	48	1248	769.8	162	2451	1214	0.020	-1.498
	Yes	3	124	157.7	22	306	45	1.691	.
Sunday	No	47	1177	807.0	150	2474	1212	0.227	-1.452
	Yes	4	58	51.3	17	128	43	1.181	0.428

Figure 6 - Least Squares Means for Weekday*Rain



There is a clear pattern for notably lower average numbers of bike rentals on rainy days, however there are very few observations for rainy days, meaning drawing robust conclusions about the effect of rain will be difficult.

For days with no rain, a similar pattern is observed with the effect of weekday vs weekend. This pattern appears less distinct on rainy days, however.

Table 9 – Two-Way ANOVA Summary for Bike Rentals by Weekday and Rain

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	75185328.8	5783486.8	6.94	<.0001
Error	339	282431921.1	833132.5		
Corrected Total	352	357617249.9			

R-Square	Coeff Var	Root MSE	rented Mean
0.210240	58.73558	912.7609	1554.017

As with the one-way ANOVA model, the model is again significant with a p value of <0.0001. With an R-Square of 0.210 the model now accounts for more variability (21.0%, compared to 7.3%), however a substantial amount of unexplained variability remains.

Table 10 - Type III Sum of Squares from the Two-Way ANOVA Model

Source	DF	Type III SS	Mean Square	F Value	Pr > F
*wkday*rain*	13	75185328.80	5783486.83	6.94	<.0001

The interaction term, *wkday* and *rain*, is statistically significant (p<0.0001), meaning that the relationship between these two variables is unlikely to be due to random chance. In practice, the model is suggesting that the day of the week and whether it's raining or not affects how many bikes are rented.

Table 11 - Pairwise Comparisons of Least Squares Means for Weekday (*wkday*) and Rain (rain) on Bike Rentals

Least Squares Means for effect wkday*rain

Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: rented

i/j	Fri No	Fri Yes	Sat No	Sat Yes	Sun No	Sun Yes	Mon No	Mon Yes	Tue No	Tue Yes	Wed No	Wed Yes	Thu No	Thu Yes
Fri No	.	0.6864	0.0373	0.072	0.011	0.0098	1	0.0152	1	0.0131	1	0.0174	1	0.4473
Fri Yes		.	0.9972	1	0.999	1	0.7532	1	0.7562	1	0.6867	1	0.8227	1
Sat No			.	0.7225	1	0.4071	0.1088	0.4943	0.1314	0.4637	0.0454	0.6316	0.3003	0.9982
Sat Yes				.	0.8045	1	0.1011	1	0.1038	1	0.0729	1	0.1468	0.9998
Sun No					.	0.5151	0.038	0.6053	0.0486	0.5743	0.0141	0.747	0.1329	0.9996
Sun Yes						.	0.0161	1	0.017	1	0.0101	1	0.0282	0.9989
Mon No							.	0.0245	1	0.0212	1	0.0293	1	0.5486
Mon Yes								.	0.0256	1	0.0156	1	0.0417	0.9996
Tue No									.	0.0223	1	0.031	1	0.5557
Tue Yes										.	0.0135	1	0.0365	0.9994
Wed No											.	0.0181	1	0.4499
Wed Yes												.	0.0523	1
Thu No													.	0.6679
Thu Yes														.

Post-hoc analysis using pair-wise comparisons of the LS means of the day with and without rainfall showed that rain made a statistically significant difference to bike rentals on Monday (p = 0.0245), Tuesday (p = 0.0223) and Wednesday (p = 0.0181). All other days, the effect of rain on bike rental was not statistically significant.

Comparison of the different days with rain shows no statistically significant difference between any days. Put simply: when it rains, the numbers of bikes hired are statistically similar, regardless of what day it is.

WEEKDAY FACTORIAL ANOVA DIAGNOSTICS

The two-way ANOVA model residuals were investigated to test the same key assumptions as the one-way ANOVA model.

Figure 7 - Two-Way ANOVA Model Diagnostics

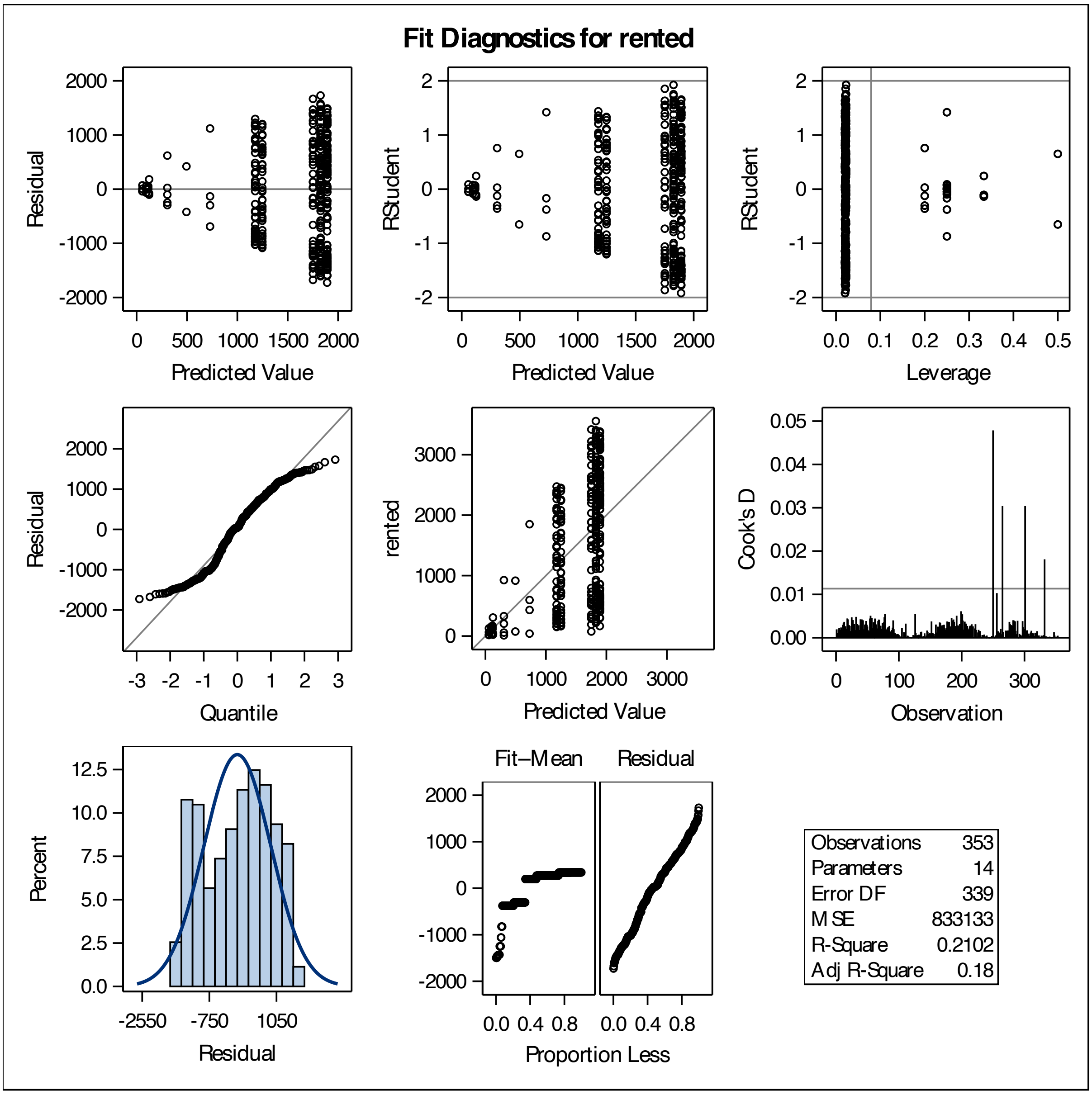


Table 12 - Levene's Test for Homogeneity of *rented* Variance in the Two-Way ANOVA Model

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
wkday*rain	12	2.221E13	1.851E12	3.64	<.0001
Error	338	1.72E14	5.088E11		

The QQ plot and histogram continue to show a departure from normality using an ANOVA model.

Pronounced heteroscedasticity is evident in the Predicted vs Residual plot, with a notable cone shape in the residuals observed. The lack of homoscedasticity is confirmed the Levene's Test ($p < 0.0001$).

In addition, this model now has several influential and high leverage points. The proportion of these outliers is low and is therefore not the primary concern with this model given the observations around normality and heteroscedasticity.

While the two-way model provides insights into the impact of rain, it does have limitations. The low number of observations for rainy days, the lack of normality in the data and the presence of heteroscedasticity are concerns that can affect the reliability of the model. While the two-way ANOVA model does account for more variability than the one-way model, a substantial amount of unexplained variability remains in both models.

As both ANOVA models had issues with normality of data, which transformation was unable to remedy, the effect of weekday will be investigated using alternative methods. Non-parametric tests, which do not require normally distributed data, were employed to analyse the effect of weekday on rentals.

NON-PARAMETRIC TESTING OF WEEKDAY

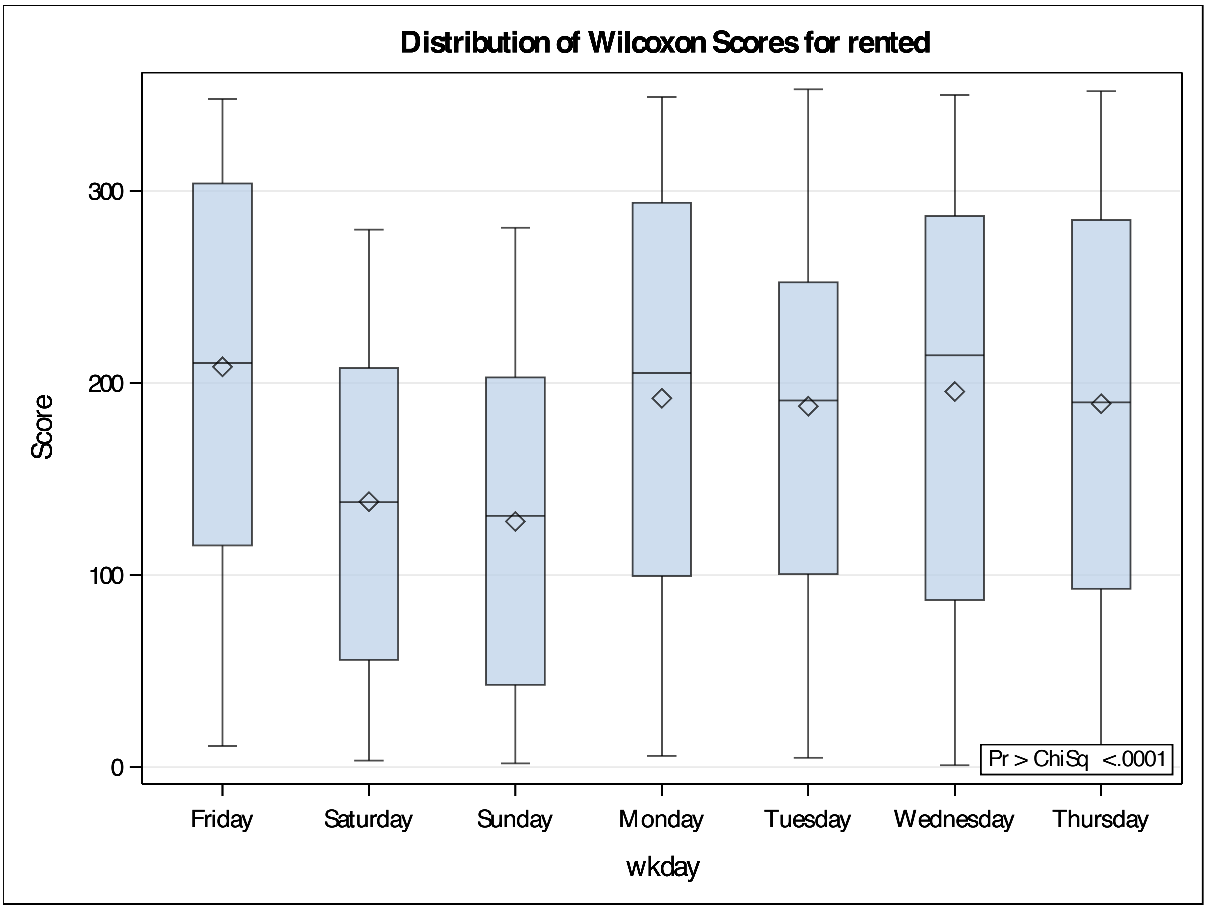
The Kruskal-Wallis Test, a non-parametric statistical test, was conducted to explore the influence of the day of the week (*wkday*) on the number of bike rentals. This test works by ranking the data within each group (*day*) calculating and comparing the sums of the ranks between each group. The null hypothesis for this test is that there is no difference between the ranks.

Table 13 - Kruskal-Wallis Test for Rental by Weekday

Wilcoxon Scores (Rank Sums) for Variable rented Classified by Variable wkday					
wkday	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Friday	51	10637	9027	674	209
Saturday	51	7053	9027	674	138
Sunday	51	6528	9027	674	128
Monday	52	9992	9204	680	192
Tuesday	48	9024	8496	657	188
Wednesday	50	9783	8850	669	196
Thursday	50	9466	8850	669	189
Average scores were used for ties.					

Kruskal-Wallis Test		
Chi-Square	DF	Pr > Chi Sq
28.0744	6	<.0001

Figure 8 - Distribution of Wilcoxon Scores for *Rented*



With a p value <0.0001, the null hypothesis is rejected, suggesting that are statistically significant differences in the number of bikes rented on different days of the week.

Table 14 - Pairwise Two-Sided Multiple Comparisons of *rented* by *wkday*

Dwass, Steel, Critchlow-Fligner Method			
Variable: rented			
wkday	Wilcoxon Z	DSCF Value	Pr > DSCF
Friday vs. Saturday	3.4334	4.8555	0.0107
Friday vs. Sunday	3.9019	5.5181	0.0018
Friday vs. Monday	0.7454	1.0541	0.9897
Friday vs. Tuesday	0.9627	1.3615	0.9619
Friday vs. Wednesday	0.7166	1.0134	0.9916
Friday vs. Thursday	1.0698	1.5129	0.9370
Saturday vs. Sunday	0.6358	0.8992	0.9956
Saturday vs. Monday	-2.6780	3.7873	0.1038
Saturday vs. Tuesday	-2.4890	3.5200	0.1631
Saturday vs. Wednesday	-2.8630	4.0489	0.0636
Saturday vs. Thursday	-2.5336	3.5830	0.1473
Sunday vs. Monday	-3.0903	4.3703	0.0328
Sunday vs. Tuesday	-2.8776	4.0695	0.0611
Sunday vs. Wednesday	-3.3215	4.6973	0.0156
Sunday vs. Thursday	-3.0736	4.3467	0.0345
Monday vs. Tuesday	0.2070	0.2927	1.0000
Monday vs. Wednesday	-0.1841	0.2603	1.0000
Monday vs. Thursday	0.1573	0.2225	1.0000
Tuesday vs. Wednesday	-0.4406	0.6231	0.9994
Tuesday vs. Thursday	-0.0640	0.0905	1.0000
Wednesday vs. Thursday	0.2620	0.3705	1.0000

Pairwise comparison shows that Sunday is significantly different to Monday, Wednesday, Thursday and Friday ($p < 0.05$). Friday is also significantly different to Saturday.

WEEKDAY ANALYSIS CONCLUSION

Both parametric (ANOVA) and non-parametric (Kruskal-Wallis) tests were employed to examine the relationship between weekdays and bike rentals. The results from these tests consistently demonstrated that day of the week has a significant impact on the number of bikes rented, with Friday and Sunday being consistently notable in their differences.

Both ANOVA models were statistically significant but the proportion of variance they explained was low, suggesting other factors not included in the model contribute significantly to bike rental variations. The two-way ANOVA demonstrated the impact rain can have on bike rentals. Both ANOVA models had limitations relating to the underlying assumptions of the data, normality and heteroscedasticity, meaning the results should be interpreted with caution.

In summary, this section's analysis highlights that the day of the week plays a crucial role in bike rental patterns in Seoul. The non-parametric Kruskal-Wallis Test provides a robust and distribution-agnostic approach to establish these differences, while the parametric ANOVA models have their limitations, they offer a more nuanced understanding, including the impact of rain on specific weekdays.

RENTAL TRENDS BY WEEKDAY AND TEMPERATURE

The previous models identified weekday as a significant contributor to bike rental variability, however, weekday only explained a small proportion of the variability. The two-way ANOVA took this a step further by including *rain* in the model which showed that rain too was a relevant factor and improved the amount of variability the model could explain.

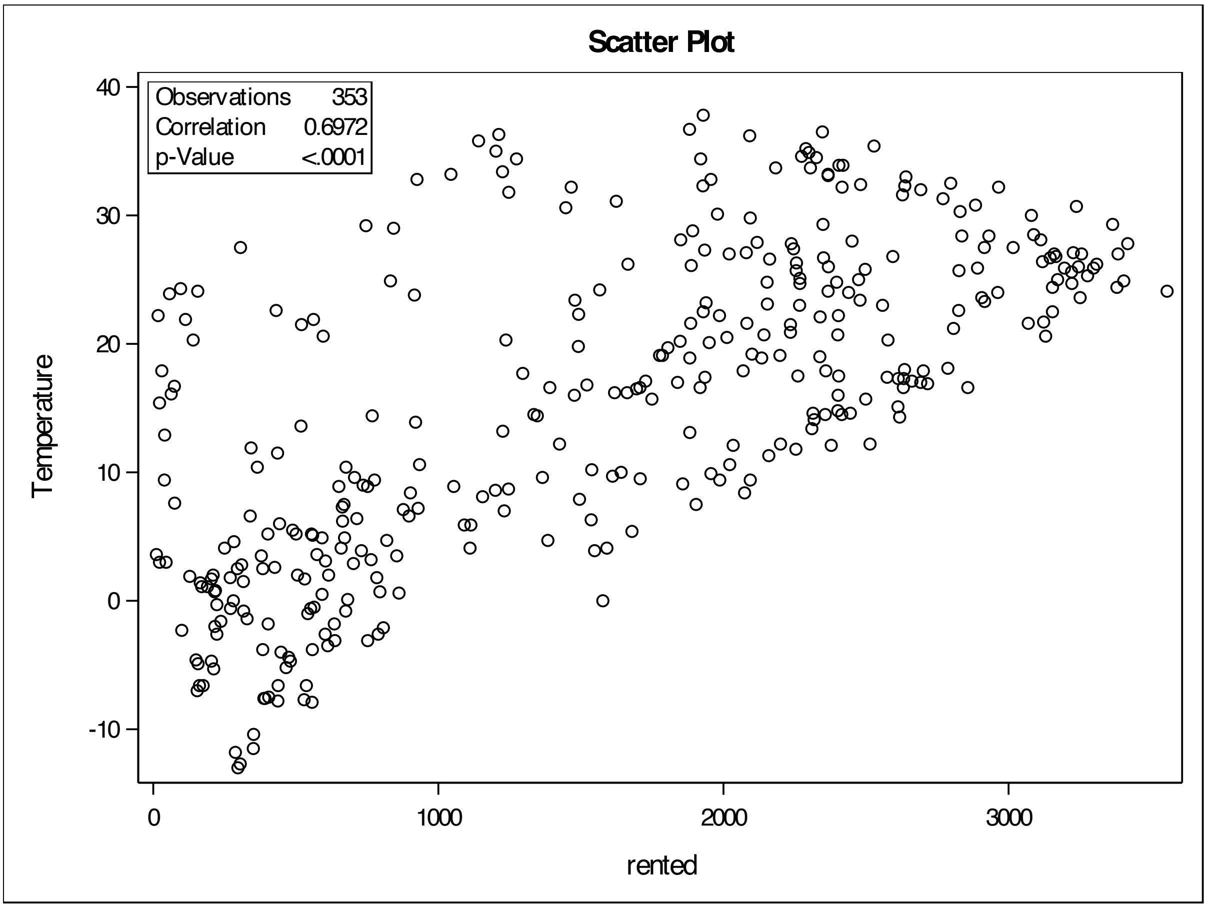
In this analysis, a covariate, *temperature*, will be introduced into an ANOVA model to create an ANCOVA model.

As *rain* demonstrated in the previous model, environmental factors can feasibly affect an individual's decision to hire a bike making investigation of temperature worthwhile. For its inclusion as a covariate in an ANCOVA model, *temperature* should have a linear relationship to *rented* and should be independent of the treatment effect (*wkday*).

WEEKDAY AND TEMPERATURE CORRELATION ANALYSIS

To measure the strength and direction the linear relationship between *temperature* and *rented*, a Pearson's Correlation Analysis was performed.

Table 15 - Pearson's Correlation Coefficient Scatter Plot



The correlation between *rented* and *temperature* was 0.697, indicating a strong positive linear relationship. The p-value of <0.0001 indicates that the correlation is statistically significant and unlikely to be due to chance.

TEMPERATURE AND WEEKDAY INDEPENDENCE

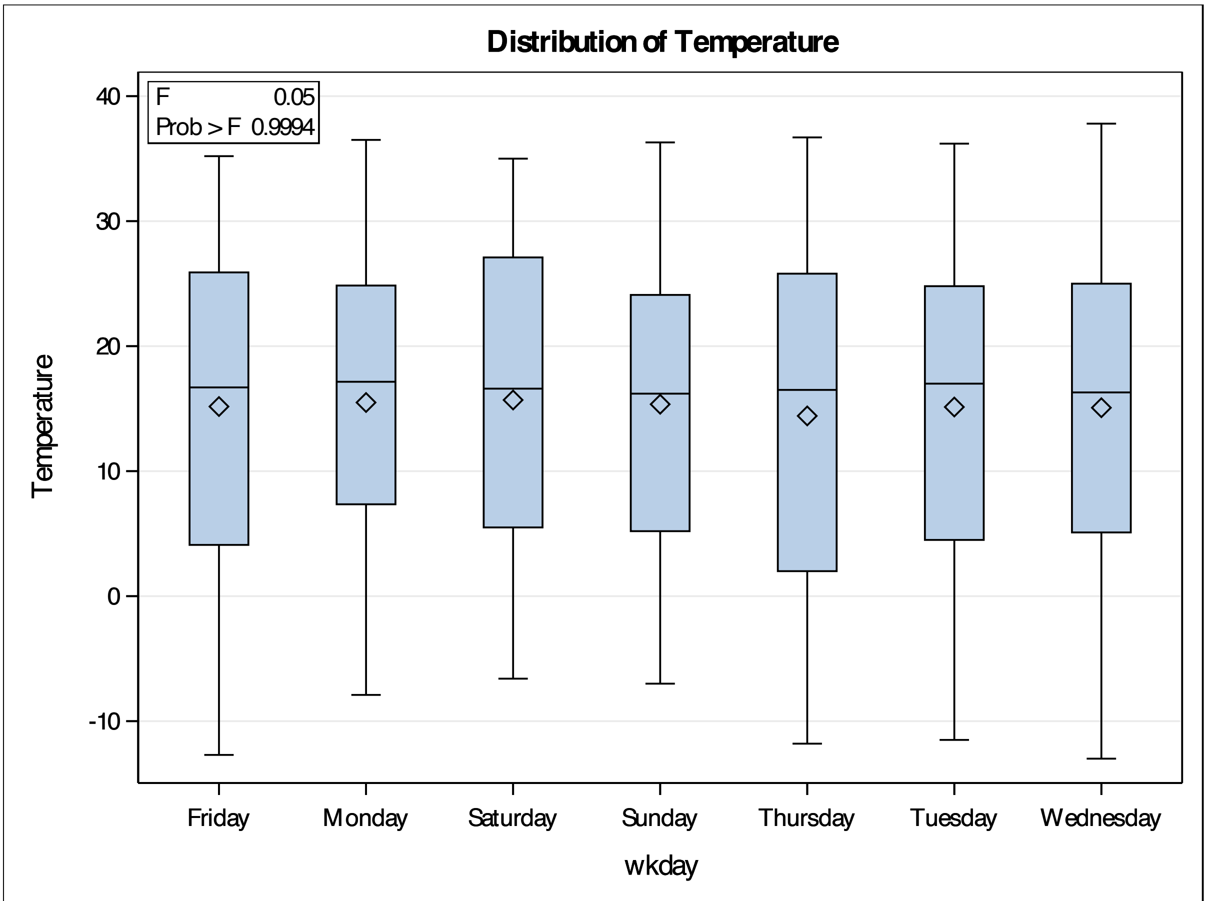
In ANCOVA, independence of the covariate (*temperature*) and treatment effect (*wkday*) refers to the assumption that these do not interact. In this instance, the assumption is that day of the week has no effect on the temperature for that day.

Independence of the covariate and the treatment effect were assessed using a one-way ANOVA. If the two variables are independent of one another then no statistically significant difference will be identified.

Table 16 - ANOVA Summary for Weekday and Temperature

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	49.46262	8.24377	0.05	0.9994
Error	346	53262.32605	153.93736		
Corrected Total	352	53311.78867			

Figure 9 - Distribution of temperature on Different Days of the Week



The ANOVA model identified no statistically significant differences within the groups (p = 0.9994). Analysis of the distributions in Figure 9 highlight that the average temperature and the variation in temperature are broadly similar across all days of the week.

As *wkday* appears to have no impact on *temperature*, it is safe to assume these variables are independent, satisfying a key assumption for ANCOVA.

EQUALITY OF SLOPES

In ANCOVA, equality of slopes refers to the assumption that the relationship between the covariate (*temperature*) and the dependent variable (*rented*) is the same across different levels or groups of the independent variable (*wkday*).

The ANCOVA model involves fitting multiple linear regressions, each corresponding to a different day of the week (*wkday*), to assess the relationship between *temperature* and *rented* on each day. If there is equality of slopes then the slope of each regression (angle of the line) will be similar, which will in turn suggest that the effect of the covariate is the same across each day.

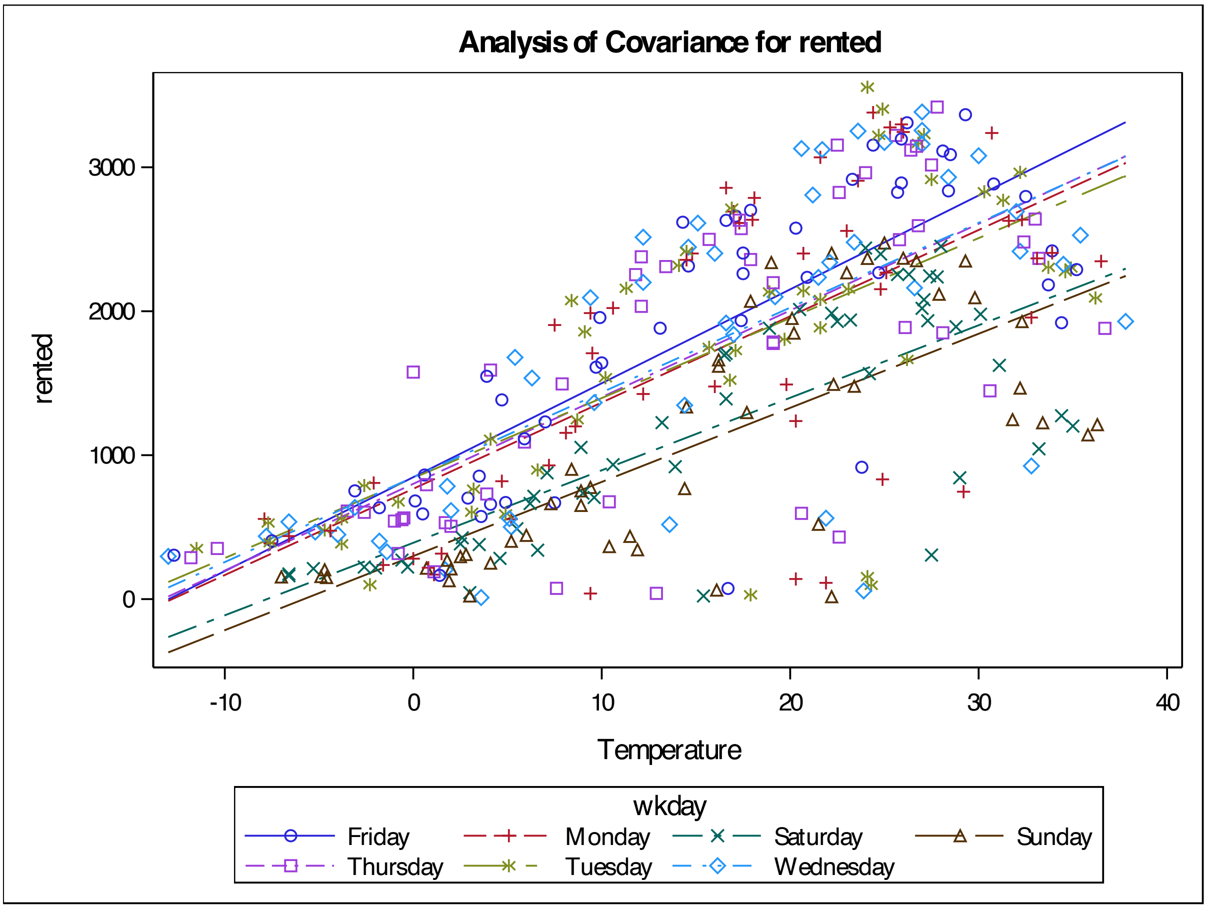
Equality of slopes is tested by including an interaction term, *temperature***wkday*, in an ANCOVA model.

Table 17 - ANCOVA Summary for Model Assessing *temperature* wkday* Interaction

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	203478378.9	15652183.0	34.42	<.0001
Error	339	154138871.0	454686.9		
Corrected Total	352	357617249.9			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Temperature*wkday	6	1224641.5	204106.9	0.45	0.8456

Figure 10 – ANCOVA Multiple Linear Regression



The interaction term between *temperature* and *wkday* does not show statistical significance (p = 0.8456) suggesting that the influence of temperature on bike rentals does not significantly vary across different days of the week.

Observation of the regression lines shows similar slopes with no significant crossing or interaction observed.

The assumption of equality of slopes is satisfied and the interaction term does not need to be included in the model.

ONE-WAY ANCOVA RELATING RENTED AND WKDAY WITH TEMPERATURE

Table 18 - ANCOVA Summary for Analysis of Bike Rentals by Weekday and Temperature

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	202253737.4	28893391.1	64.16	<.0001
Error	345	155363512.5	450329.0		
Corrected Total	352	357617249.9			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
wkday	6	28410506.2	4735084.4	10.51	<.0001
Temperature	1	175988038.2	175988038.2	390.80	<.0001

R-Square	Coeff Var	Root MSE	rented Mean
0.565559	43.18264	671.0656	1554.017

The significant p-value (*p* = < 0.0001) for the model suggest that its component factors, *wkday* and *temperature*, have a statistically significant impact on the number of bikes rented. Analysing the Type III Sum of Squares (SS), both *wkday* and *temperature* have a significant impact on *rented*, when account for other effects.

The R-square for the model is 0.566 suggesting that the model accounts for approximately 56.6% of the variability in *rented*.

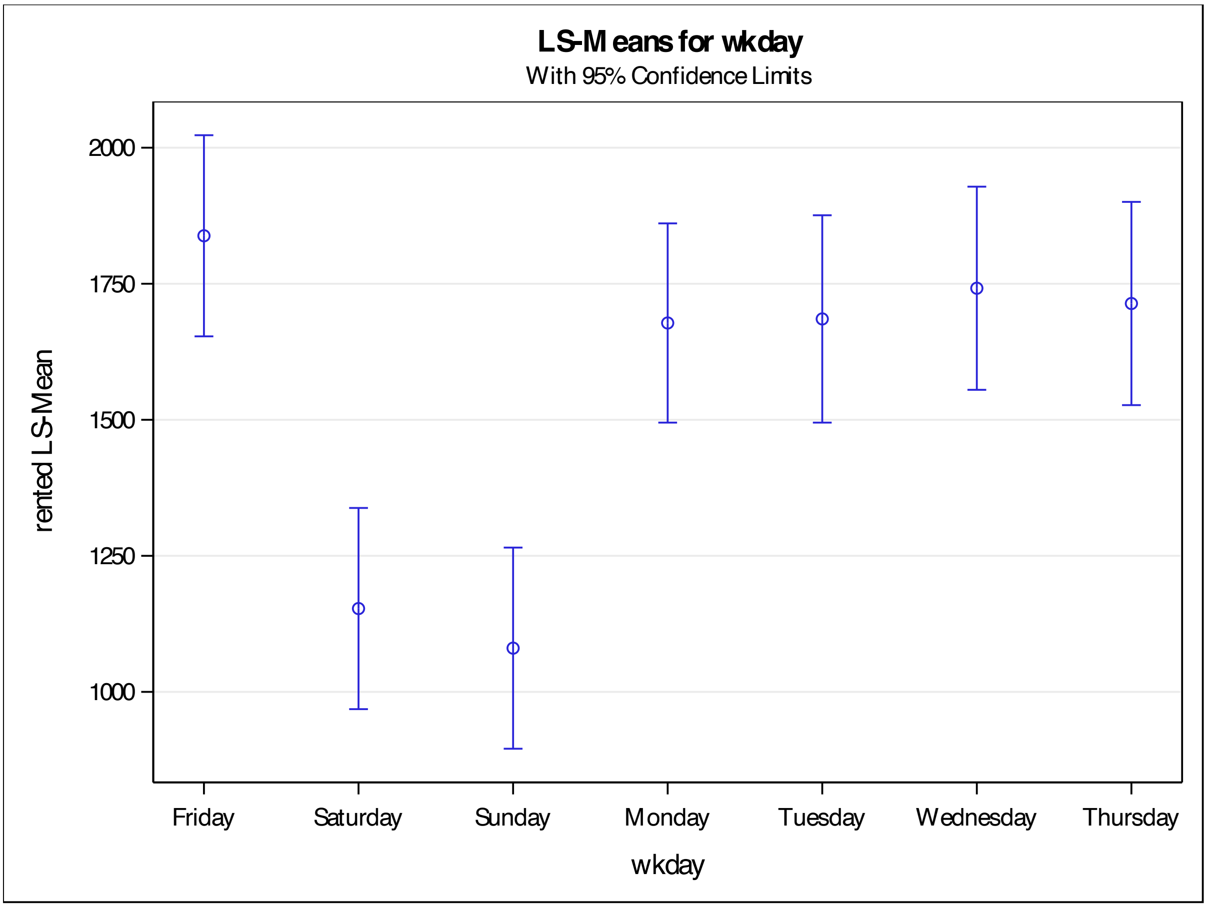
Table 19 - Pairwise Comparisons of Least Squares Means for Weekday (*wkday*) on Bike Rentals with *temperature* as a Covariate

Least Squares Means for effect wkday
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: rented

i/j	Friday	Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday
Friday		<.0001	<.0001	0.8894	0.9178	0.9912	0.9672
Saturday	<.0001		0.9981	0.0017	0.0019	0.0003	0.0007
Sunday	<.0001	0.9981		0.0002	0.0002	<.0001	<.0001
Monday	0.8894	0.0017	0.0002		1.0000	0.9991	1.0000
Tuesday	0.9178	0.0019	0.0002	1.0000		0.9996	1.0000
Wednesday	0.9912	0.0003	<.0001	0.9991	0.9996		1.0000
Thursday	0.9672	0.0007	<.0001	1.0000	1.0000	1.0000	

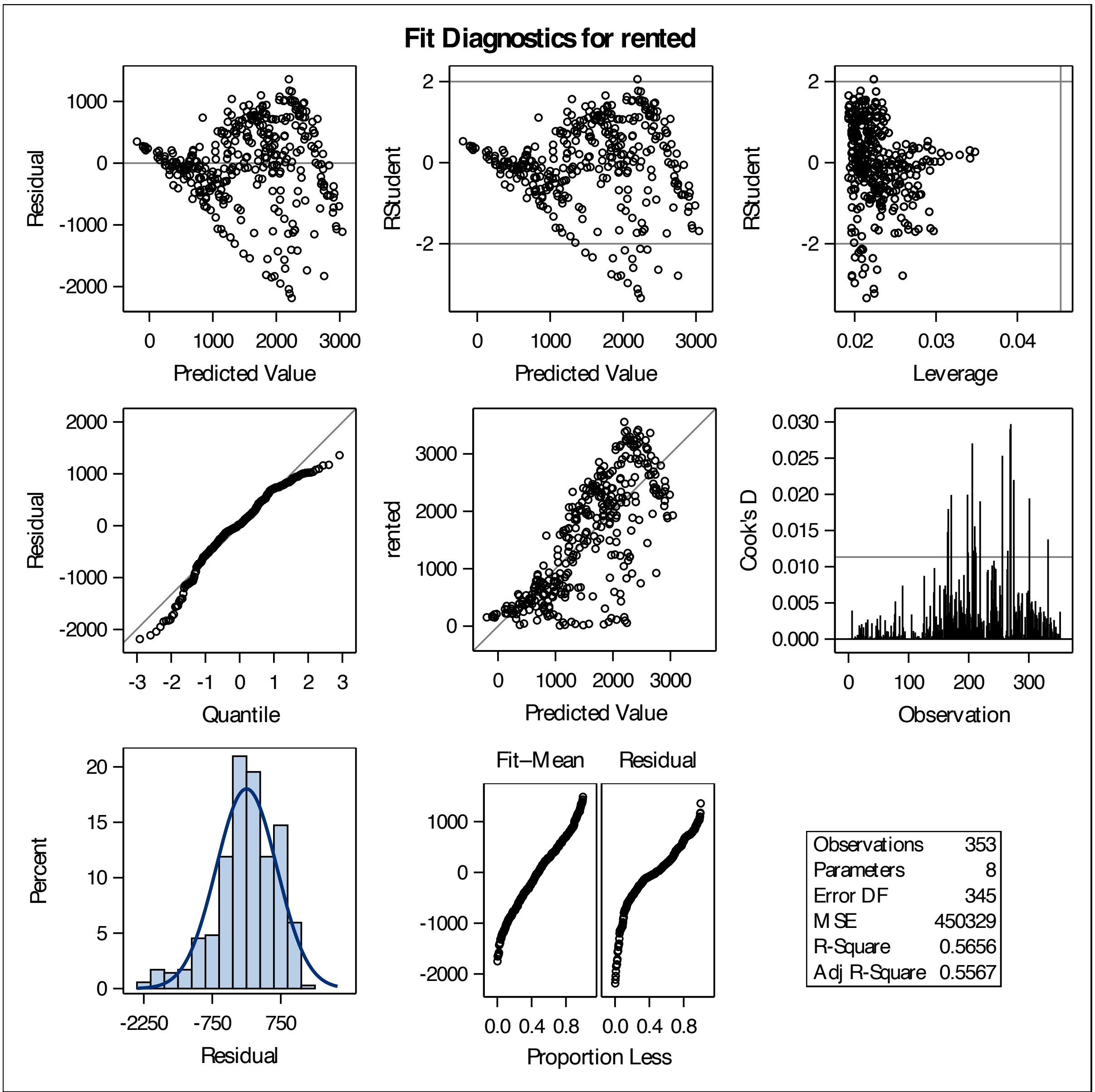
Figure 11 – Average Number of Bikes Rented on Each Day of the Week with *temperature* as a Covariate



Post-hoc analysis using pair-wise comparisons with Tukey-Kramer adjustments (Table 19) shows that Saturday and Sunday are significantly different to the weekdays, Monday to Friday. This difference can be observed in; Saturday and Sunday are notably lower than the other days of the week and there is no overlap of the 95% confidence limits.

ONE-WAY ANCOVA DIAGNOSTICS

Figure 12 - One-Way ANCOVA Model Diagnostics



ANCOVA models are somewhat resilient to deviations from normality and deviations observed in the histogram and residual QQ plot are unlikely to have a significant impact on the model.

Heteroscedasticity is observed in the predicted vs residual plots.

Outliers are observed in the RStudent plots as are influential observations in the Cook's D plot.

While the model has room for improvement, the issues highlighted are not severe and the model could be considered fit for purpose provided its limitations are considered.

ONE-WAY ANOVA AND ANCOVA COMPARISONS

Both models identify the importance of weekday on bike rentals, however the ANCOVA model takes this further and includes the role of temperature. The inclusion of *temperature* leads to a better understanding of bike rental factors, as the ANCOVA model accounts for considerably more variation in bike rentals (56.6% compared to 7.3%). The magnitude of this improvement highlights the importance of temperature in people's bike hiring patterns.

Both models had issues with heterogeneous variance, however issues with normality that were noted in the ANOVA model were less pronounced in the ANCOVA model, resulting in a lower risk of bias and inaccurate or unreliable inferences. The ANCOVA model did have an increase in the number of outliers and influential observations that may need further investigation.

In summary, the ANCOVA model is the more comprehensive model that fits the data better. Some of the underlying assumptions of ANOVA/ANCOVA are improved in the ANCOVA model but issues remain.

SUMMARY

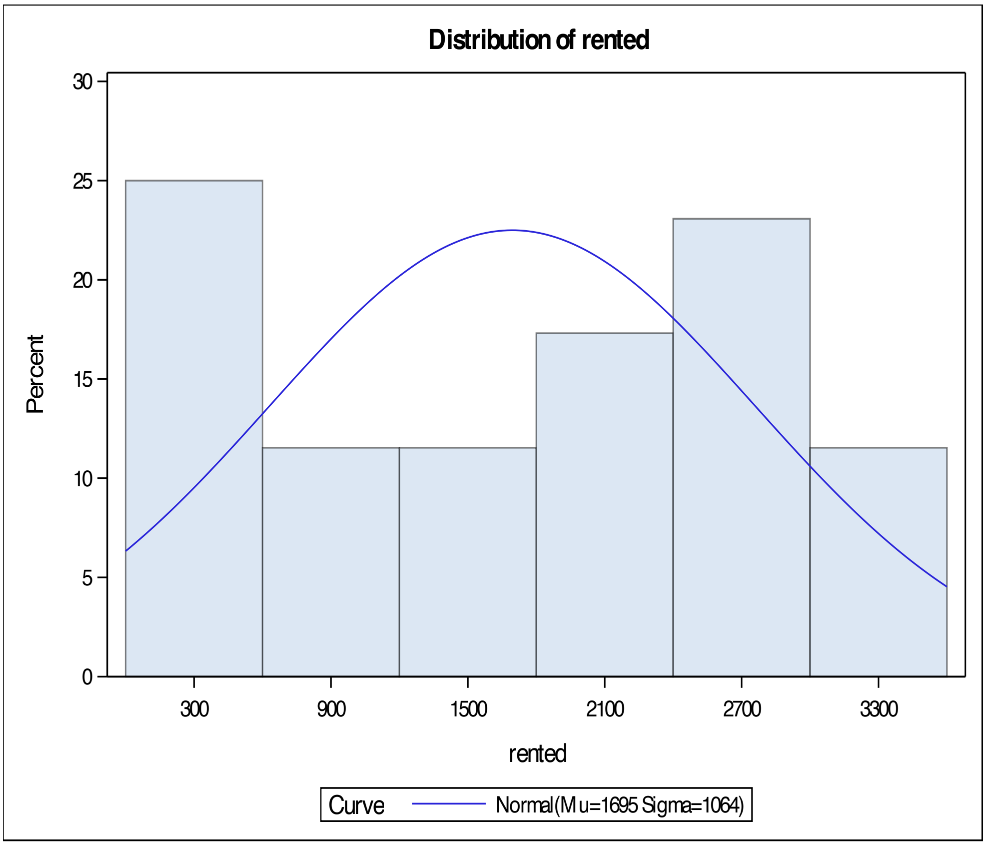
The models presented here highlight the significant role that several variables played on bike hiring patterns in Seoul.

All models agreed that day of the week plays a significant role, with demand for bikes being significantly lower at the weekend. Based on these findings, cities wanting to maintain a stable supply of bikes should ensure maximum availability through the week and focus on stock replenishment or maintenance at the weekend.

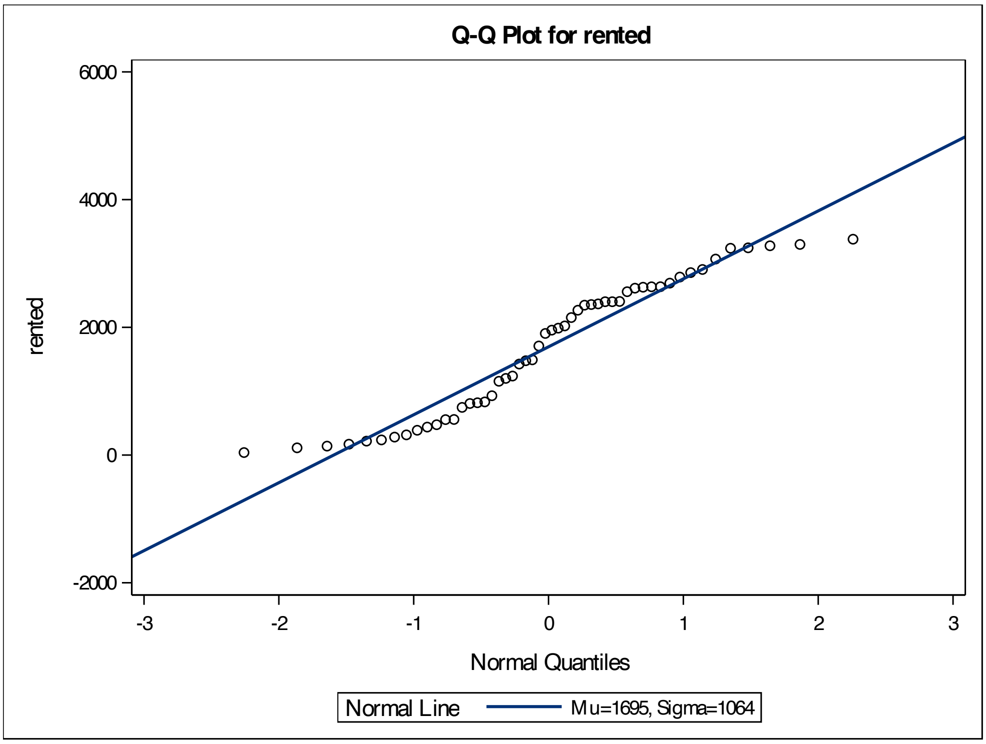
The variables *rain* and *temperature* were also explored. The models found some statistically significant effects of rain on bike hiring patterns, with lower demand on rainy days, but this was not consistent or conclusive. The effect of *temperature* was statistically significant; however, the relationship was more complex. Notably there was a strong positive trend for an increase in the number of bikes rented as the temperature increased, however, there were several outliers at higher temperatures with disproportionately lower rental counts. Cities wanting to maintain a stable supply of bikes can therefore use weather forecasting, particularly for temperature and rainfall, to estimate demand in advance.

APPENDIX A

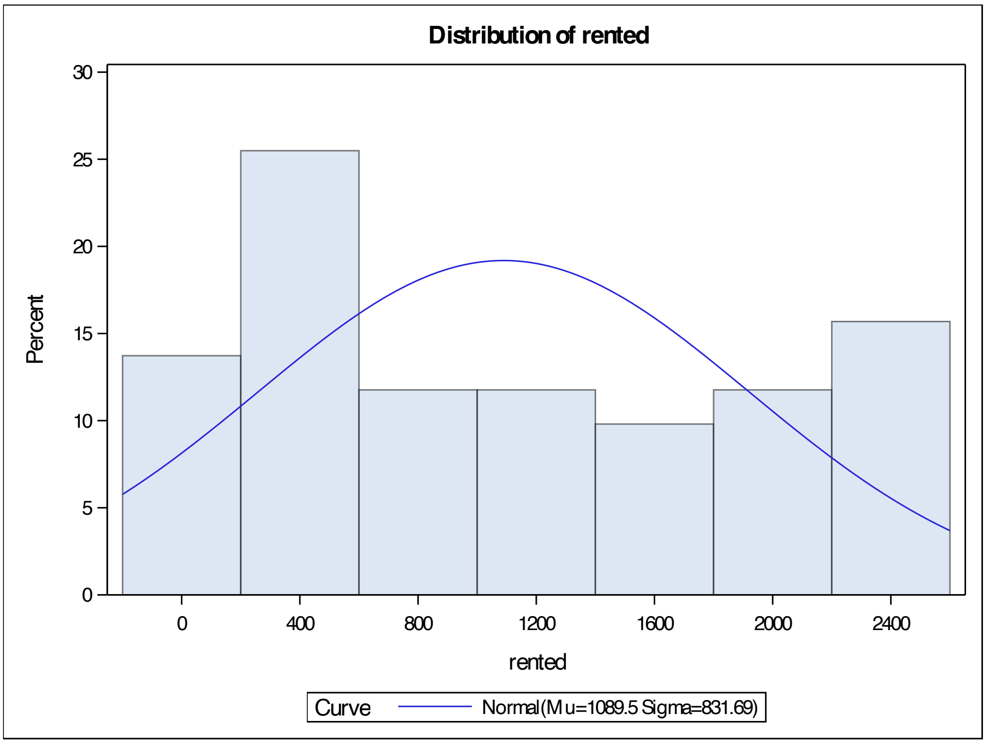
Histogram for the Distribution of *rented* on Mondays



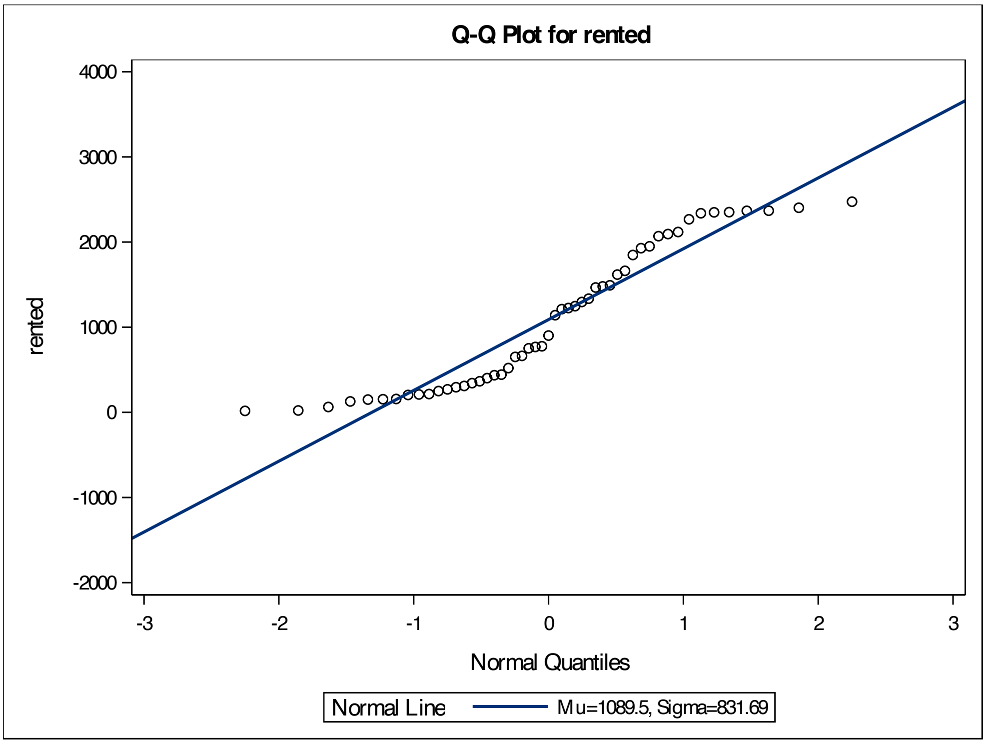
QQ Plot for the Distribution of *rented* on Mondays



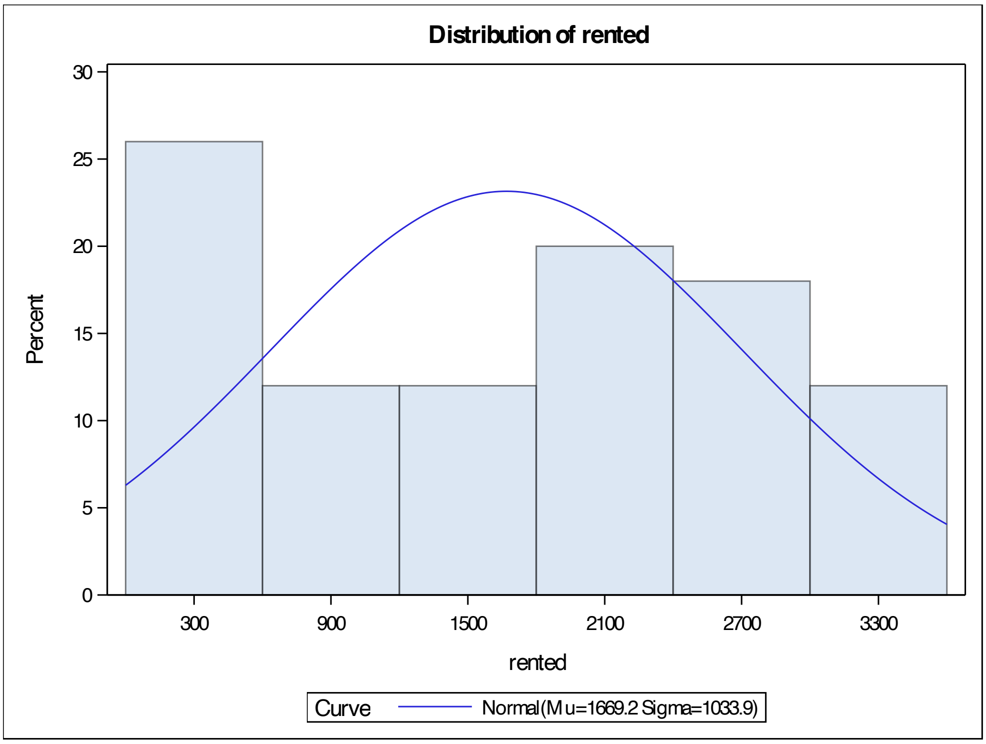
Histogram for the Distribution of *rented* on Sundays



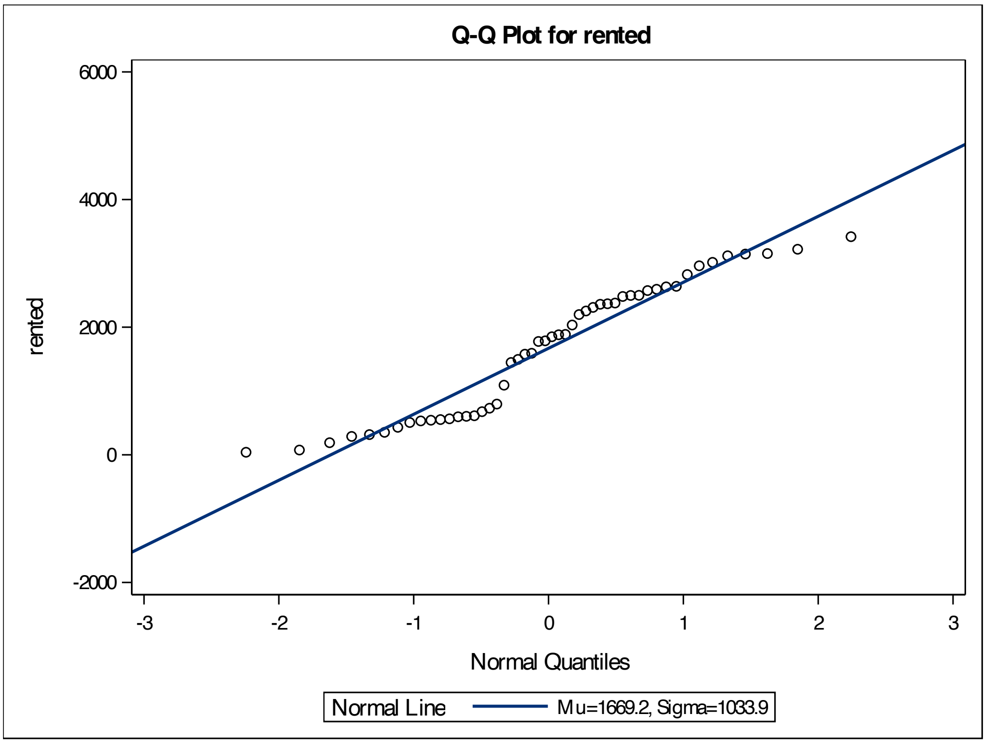
QQ Plot for the Distribution of *rented* on Sundays



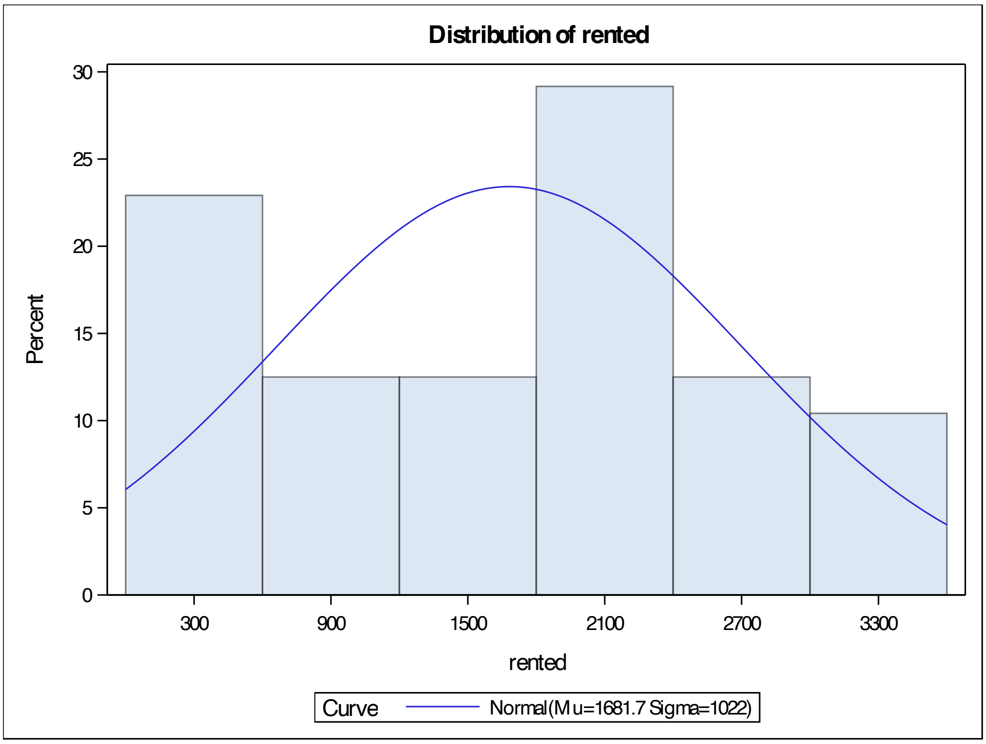
Histogram for the Distribution of *rented* on Thursdays



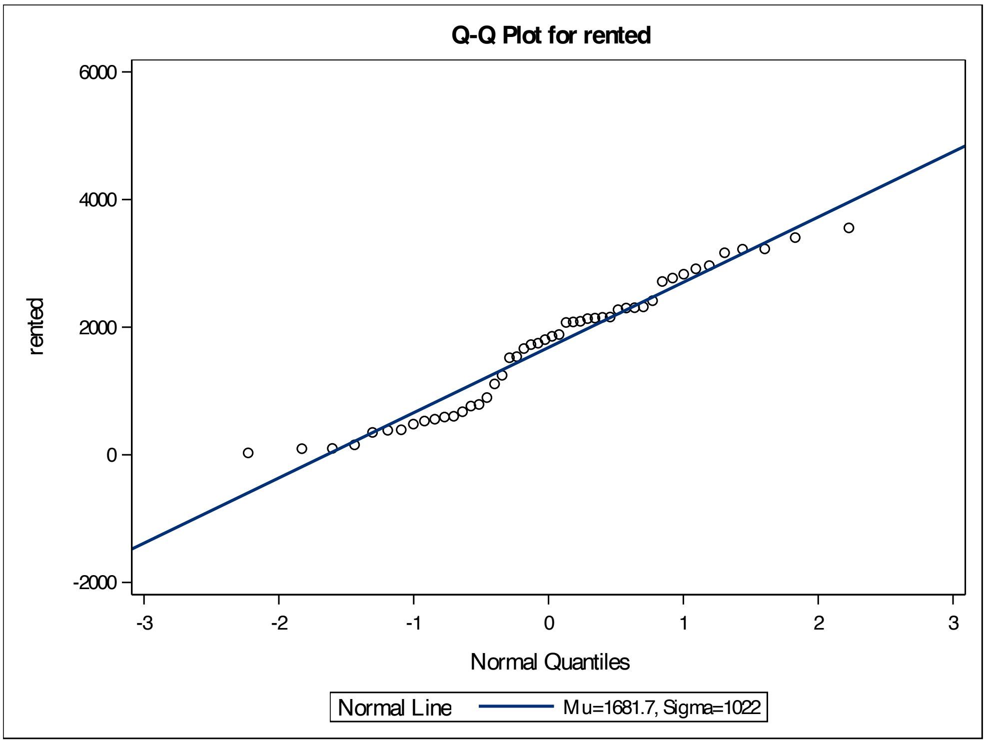
QQ Plot for the Distribution of *rented* on Thursdays



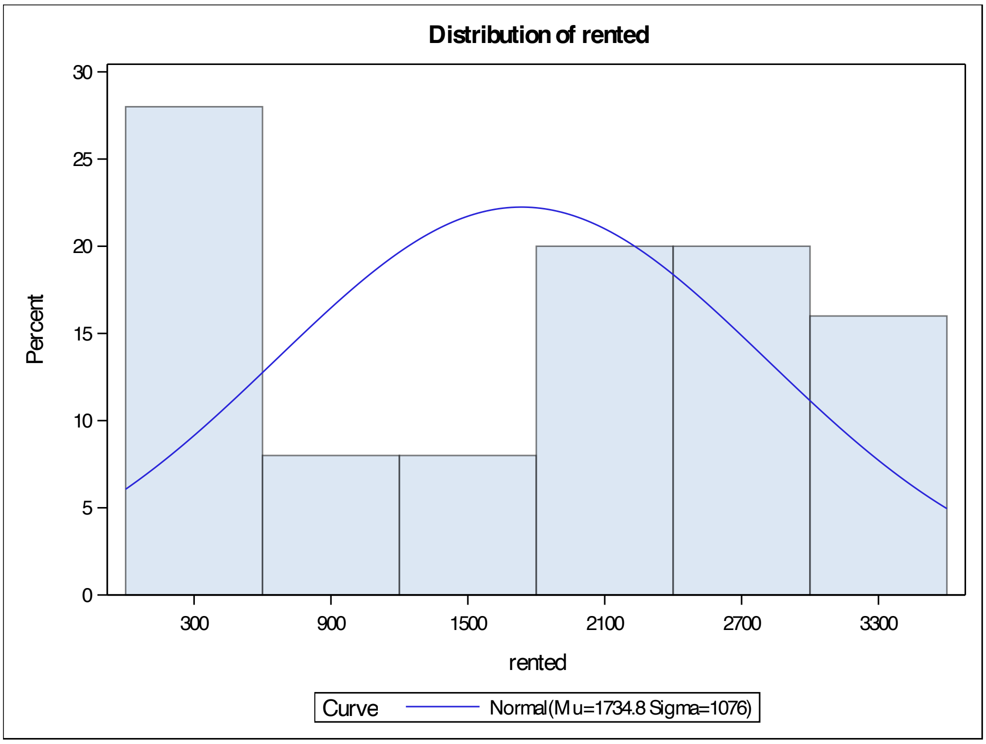
Histogram for the Distribution of *rented* on Tuesdays



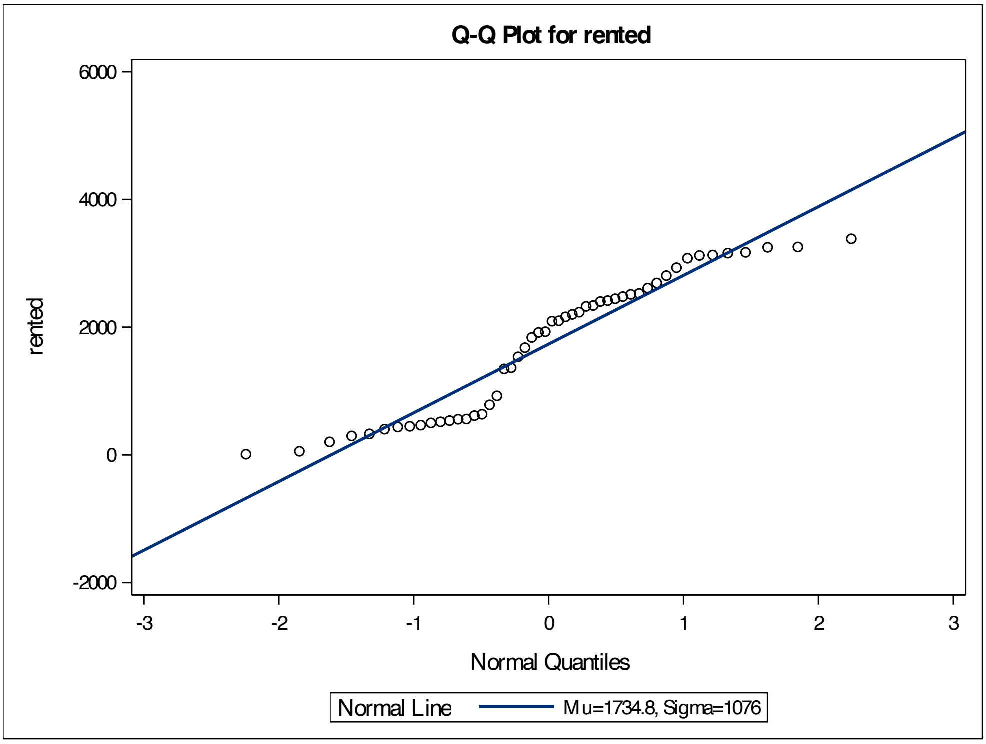
QQ Plot for the Distribution of *rented* on Tuesdays



Histogram for the Distribution of *rented* on Wednesdays

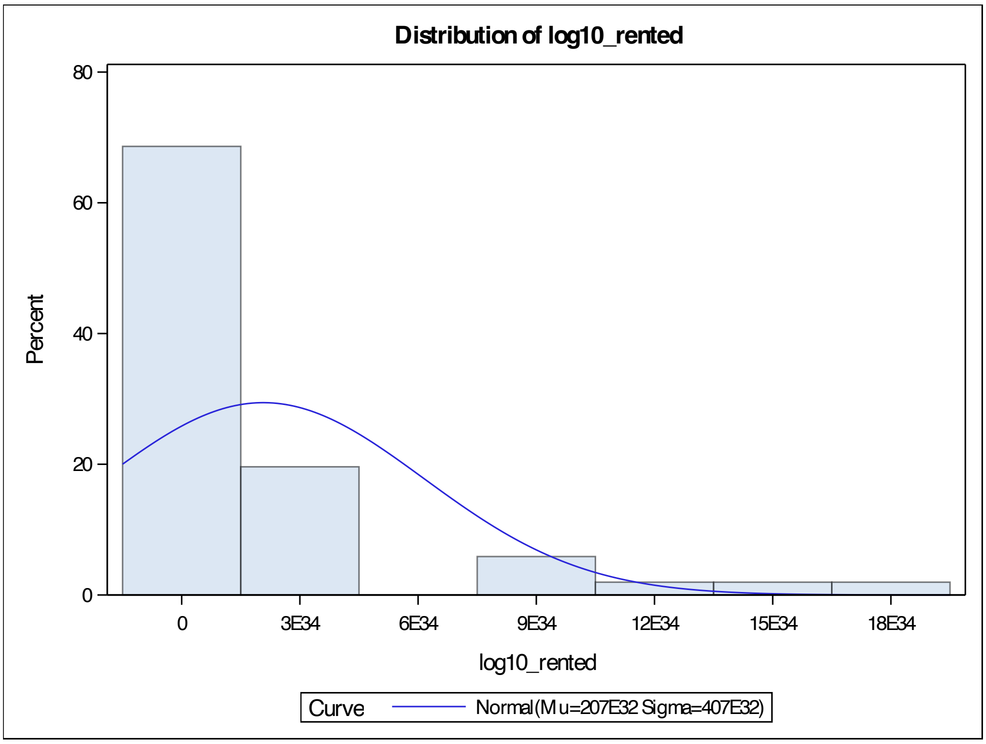


QQ Plot for the Distribution of *rented* on Wednesdays

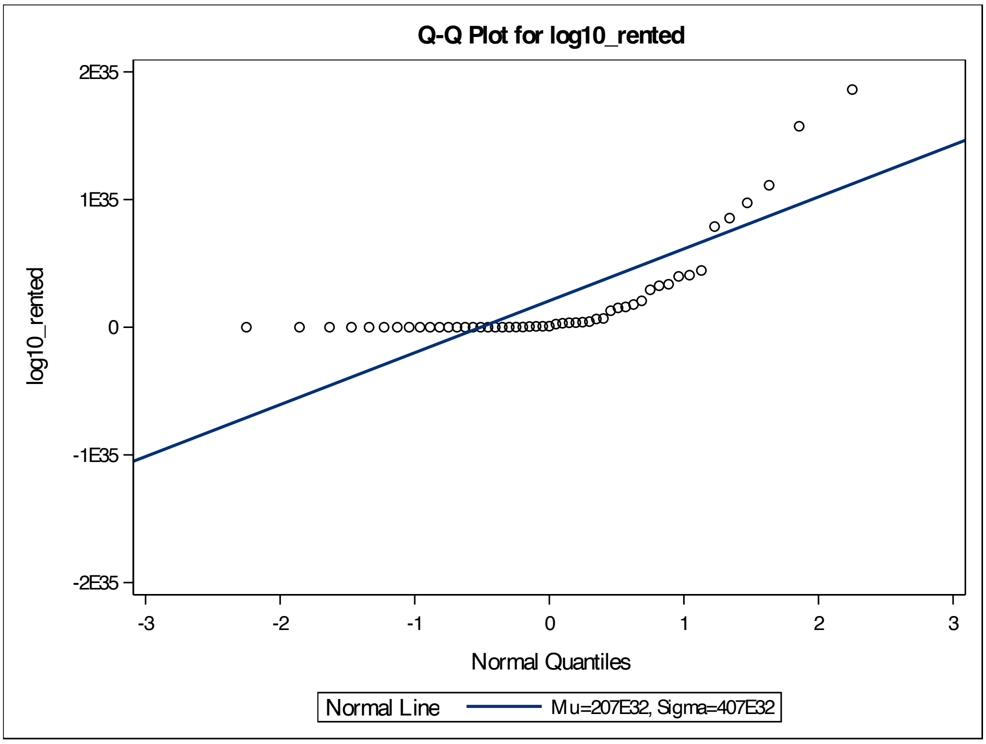


APPENDIX B

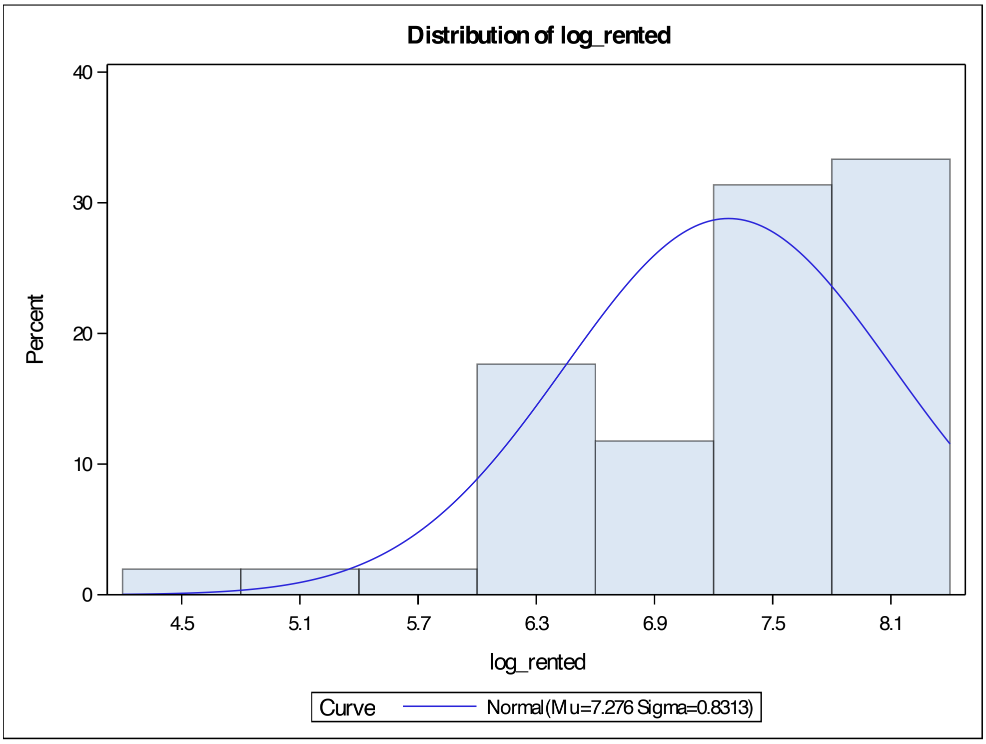
Histogram for the Distribution of *rented* on Fridays with a Log10 Transformation



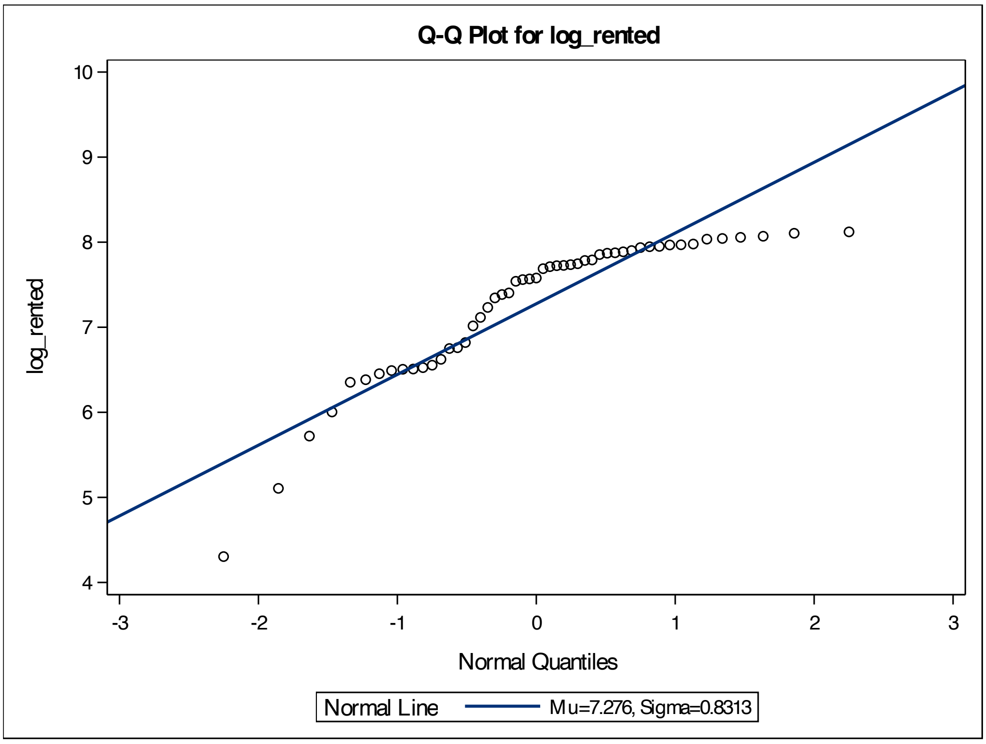
QQ Plot for the Distribution of *rented* on Fridays with a Log10 Transformation



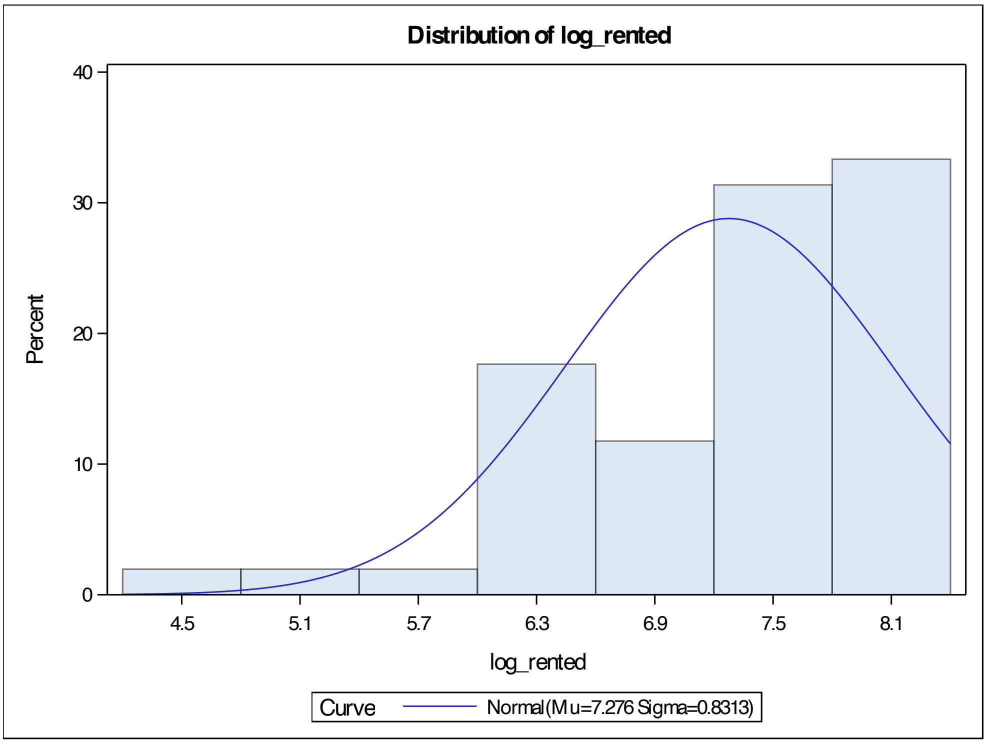
Histogram for the Distribution of *rented* on Fridays with a Log Transformation



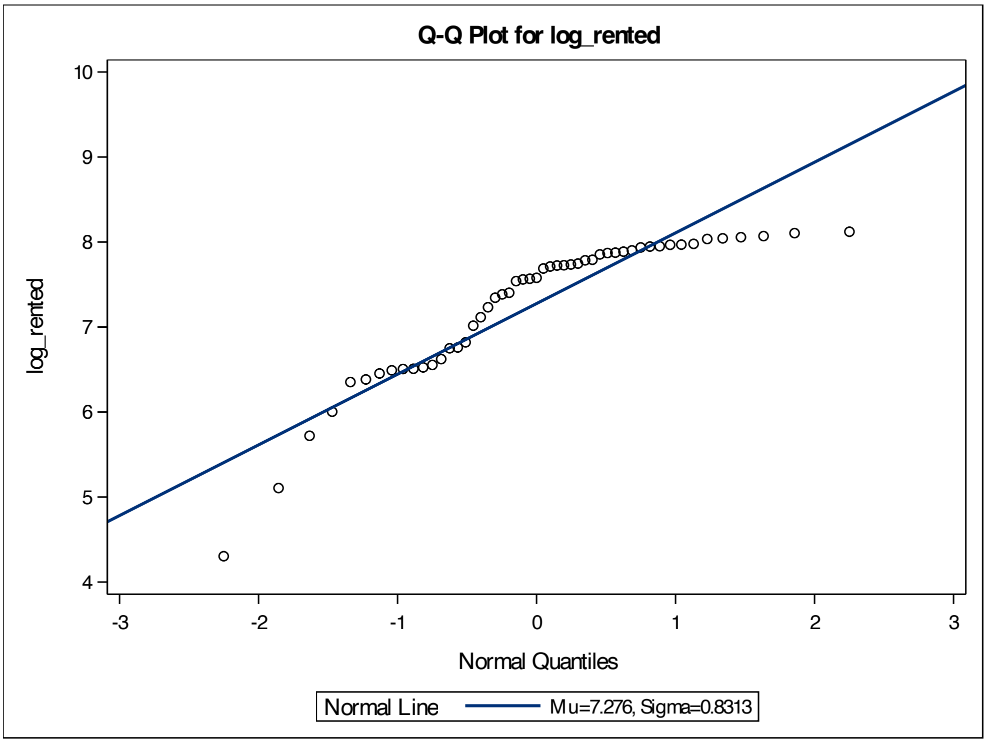
QQ Plot for the Distribution of *rented* on Fridays with a Log Transformation



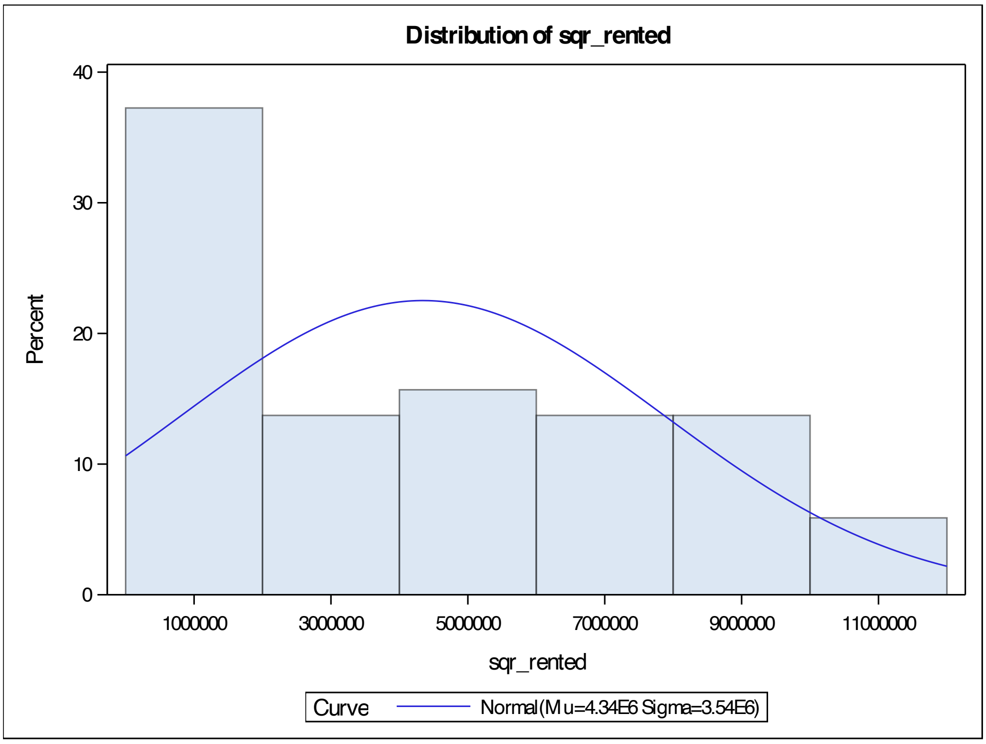
Histogram for the Distribution of *rented* on Fridays with a Log Transformation



QQ Plot for the Distribution of *rented* on Fridays with a Log Transformation



Histogram for the Distribution of *rented* on Fridays with a Square Transformation



QQ Plot for the Distribution of *rented* on Fridays with a Square Transformation