By: Rajat Chandna, Jonathan Marin, Rene Alvarenga, Samira Zarandioon

**-Two Way ANOVA Analysis -how House Sale Prices Depend on Kitchen Quality and neighbourhood**

Contents

[The Motivation and Design. 2](#_Toc516854206)

[The DataSet. 2](#_Toc516854207)

[The List of Files Accomplying Analysis. 3](#_Toc516854208)

[The Questions of Interest. 3](#_Toc516854209)

[The Assumption for the analysis. 3](#_Toc516854210)

[The Analysis. 4](#_Toc516854211)

[The Statistical Conclusion. 11](#_Toc516854212)

[The Scope of inference. 12](#_Toc516854213)

[The APPENDIX. 12](#_Toc516854214)

[**SAS Code.** 12](#_Toc516854215)

# The Motivation and Design.

Different people weigh different features quite differently while deciding upon a house purchase and what price they should pay taking into account presence or absence of their highly desired features. Some people want houses in quiet neighborhoods, while others want near busy downtowns. Some weigh higher high build qualities more and yet another look for greater Living areas, while deciding on fair price to pay for the houses.

Motivated by my wife’s interest in cooking and kitchen quality, I would like access the influence on kitchen quality on sale prices of houses in different neighborhoods in the city of AMES. **Since it is common knowledge that house sale prices differ from neighborhood to another, we could have chosen neighborhood as a block but since this analysis is in observational setting, we would also like to examine influence of neighborhood on house sale prices. Hence, we are treating both Kitchen Quality and Neighborhood are our independent variables and House Sale Price as dependent variable**.

The dataset for the analysis is “CleanedDataForANOVA.csv”, which is a dataset obtained after initial cleaning on Kaggle House Sale Prices data set.

# The DataSet.

The description and levels of independent variables is as:

KitchenQual: Kitchen quality

Ex Excellent

Gd Good

TA Typical/Average

Fa Fair

Neighborhood: Physical locations within Ames city limits

Blmngtn Bloomington Heights

Blueste Bluestem

BrDale Briardale

BrkSide Brookside

ClearCr Clear Creek

CollgCr College Creek

Crawfor Crawford

Edwards Edwards

Gilbert Gilbert

IDOTRR Iowa DOT and Rail Road

MeadowV Meadow Village

Mitchel Mitchell

Names North Ames

NoRidge Northridge

NPkVill Northpark Villa

NridgHt Northridge Heights

NWAmes Northwest Ames

OldTown Old Town

SWISU South & West of Iowa State University

Sawyer Sawyer

SawyerW Sawyer West

Somerst Somerset

StoneBr Stone Brook

Timber Timberland

Veenker Veenker

Independent Variable:

Sale Price: Sale Price of the House

# The List of Files Accomplying Analysis.

All the below files have been placed in out Project GitHub Repo at

<https://github.com/JLMarin64/AppliedStatsProject1>

**CleanedDataForANOVA.csv** – The cleaned input data set.

**TwoWay\_ANOVA\_Analysis.sas** – File containing SAS code for the analysis.

**ALL\_MultipleComparisons\_Results.xlsx** – Output file containing results of all multiple comparisons.

**Statistically\_Significant\_MultipleComparisons\_Results.xlsx** - Output file containing results of all statistically significant multiple comparisons.

# The Questions of Interest.

The main goal of the analysis is to:

1. Determine the influence of 4 different levels of kitchen quality on house sale price. Is the difference between prices paid for houses with better quality kitchens differ significantly than prices paid for houses with not so great quality kitchens, after taking into account of the neighborhoods?
2. Determine the influence of Neighborhood on the house sale prices in city AMES?
3. Determine whether the influence of 4 different levels of kitchen quality on house sale price vary from neighborhood to neighborhood?

# The Assumption for the analysis.

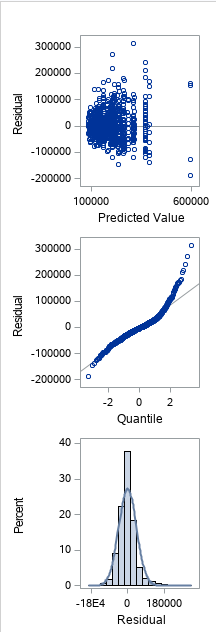
Two-way ANOVA analysis requires observations within different groups and across different groups be independent of one another. But house sale prices within same neighborhood are rarely independent of one another and suffer from spatial correlation. For this analysis, we will assume that house sale prices are independent of each other and proceed with analysis.

# The Analysis.

***Inputting the data and plotting the means plot.***



***Above plot depicting mean sale prices based upon levels of Kitchen quality across different neighborhoods alludes to a non-additive model as difference in mean sale prices among Kitchen Quality appear to be different across different neighborhoods. Also, spread of the error bars hint at higher spreads at higher mean values, thus towards non-constant variance. Residual plot could clarify this trend further.***



***Above residual plot reveals a funnel shaped pattern for distribution of residuals across fitted values. Also, a pattern of deviation towards left side could be seen from Normal QQ Plot. A log transformation would help in rectifying such situation.***

***Also, since houses in given data set are concentrated under 4000 sq ft of GrLiving Area, we would remove any observations that have GrLivingArea > 4000 since such houses are not representative of houses in the dataset while having high influence on sale prices.***



***After log transformation on sales prices and some outliers removal, plotting the mean log sale prices based upon levels of Kitchen quality across different neighborhoods.***



***As seen from plot above, that after log transformation on sales price, the variance is nearly equal across different levels of Kitchen Quality and Neighborhood. Also, plot alludes towards interaction between Kitchen Quality and Neighborhood and hence towards non-additive model in which average log sales prices among levels of Kitchen Quality (from Excellent, Good, Average to Fair) differ with the different Neighborhoods. Hence, running the saturated, non-additive model first.***

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **KitchenQual** | 4 | Ex Fa Gd TA |
| **Neighborhood** | 25 | Blmngtn Blueste BrDale BrkSide ClearCr CollgCr Crawfor Edwards Gilbert IDOTRR MeadowV Mitchel NAmes NPkVill NWAmes NoRidge NridgHt OldTown SWISU Sawyer SawyerW Somerst StoneBr Timber Veenker |

|  |  |
| --- | --- |
| **Number of Observations Read** | 1456 |
| **Number of Observations Used** | 1456 |

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 70 | 155.8014039 | 2.2257343 | 42.54 | **<.0001** |
| **Error** | 1385 | 72.4578965 | 0.0523162 |  |  |
| **Corrected Total** | 1455 | 228.2593005 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **LogSalePrice Mean** |
| --- | --- | --- | --- |
| 0.682563 | 1.902582 | 0.228727 | 12.02194 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **KitchenQual** | 3 | 7.89484489 | 2.63161496 | 50.30 | **<.0001** |
| **Neighborhood** | 24 | 23.12545364 | 0.96356057 | 18.42 | **<.0001** |
| **KitchenQu\*Neighborho** | 43 | 5.22502408 | 0.12151219 | 2.32 | **<.0001** |

***As we see from above tables that:***

1. ***Overall Model with Kitchen Quality and Neighborhood explaining/predicting log house sales prices is significant at alpha = 0.05, with p-value < 0.0001.***
2. ***Variables Kitchen Quality and Neighborhood combined explain about 68.2% variability in sale prices. This practical effect size is quite significant.***
3. ***As seen from Type 3 SS table, the effects of Kitchen Quality are significant even after accounting for effects of Neighborhood and interaction. Also, effects of Neighborhood are significant even after accounting for effects of Kitchen Quality and interaction. And finally, so does the effects of interaction are significant after accounting for effects of Neighborhood and Kitchen Quality.***

***Now since interaction is significant, the interpretation of the main effects would not be correct in isolation to interaction. We will perform multiple comparisons in presence of interaction once we find out interaction is important. We will first validate model assumptions.***



***As seen from residual plots above, the residuals are fairly randomly distributed across fitted values. The spread appears to be small as compared to other levels at value 13(log sale price) but that’s just because there are only two values at this value. The QQ Plot and Histogram reveal fairly normally distributed residuals at different levels. So, our model assumptions are met. Next, we check whether interaction is important. For this we find statistical significant differences in median sale price across kitchen Quality for same sites. ( p-val adjusted via tukey-kramer procedure in multiple comparisons). Results are as below:***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **KitchenQual** | **Neighborhood** | **\_KitchenQual** | **\_Neighborhood** | **Adjp** | **Estimate\_NormalScale** | **UpperCI\_NormalScale** | **LowerCI\_NormalScale** |
| Gd | CollgCr | TA | CollgCr | <.0001 | 1.50141 | 1.835 | 1.22846 |
| Ex | CollgCr | TA | CollgCr | <.0001 | 2.08535 | 3.11825 | 1.39459 |
| Ex | NridgHt | Gd | NridgHt | <.0001 | 1.44229 | 1.79186 | 1.16092 |
| Gd | SawyerW | TA | SawyerW | <.0001 | 1.54717 | 2.02513 | 1.18202 |
| Ex | Edwards | Fa | Edwards | <.0001 | 2.47161 | 4.76778 | 1.28128 |
| Fa | BrkSide | Gd | BrkSide | <.0001 | 0.53059 | 0.84826 | 0.33188 |
| Gd | OldTown | TA | OldTown | 0.0004 | 1.32964 | 1.66506 | 1.06179 |
| Ex | Edwards | TA | Edwards | 0.0005 | 2.01738 | 3.53487 | 1.15134 |
| Ex | Somerst | Gd | Somerst | 0.0007 | 1.48529 | 2.04811 | 1.07713 |
| Fa | Edwards | Gd | Edwards | 0.0231 | 0.6332 | 0.97906 | 0.40951 |
| Fa | BrkSide | TA | BrkSide | 0.0294 | 0.70609 | 0.98868 | 0.50428 |
| Ex | StoneBr | Gd | StoneBr | 0.0913 | 1.46262 | 2.1748 | 0.98366 |
| Ex | NAmes | TA | NAmes | 0.1066 | 1.57765 | 2.55318 | 0.97486 |
| Gd | Edwards | TA | Edwards | 0.1161 | 1.28905 | 1.68846 | 0.98412 |
| Ex | OldTown | TA | OldTown | 0.1354 | 1.5031 | 2.33093 | 0.96928 |
| Ex | CollgCr | Gd | CollgCr | 0.2255 | 1.38893 | 2.01211 | 0.95876 |
| Ex | BrkSide | Fa | BrkSide | 0.2315 | 2.41755 | 6.56218 | 0.89064 |
| Ex | OldTown | Fa | OldTown | 0.2993 | 1.73175 | 3.27993 | 0.91434 |
| Fa | IDOTRR | Gd | IDOTRR | 0.3259 | 0.61218 | 1.08958 | 0.34395 |
| Gd | NAmes | TA | NAmes | 0.3513 | 1.14932 | 1.35552 | 0.97448 |
| Ex | Somerst | TA | Somerst | 0.3768 | 1.60122 | 2.8124 | 0.91165 |
| Gd | Mitchel | TA | Mitchel | 0.4063 | 1.32206 | 1.85275 | 0.94338 |
| Fa | OldTown | Gd | OldTown | 0.4076 | 0.65278 | 1.09338 | 0.38973 |
| Gd | IDOTRR | TA | IDOTRR | 0.5193 | 1.44829 | 2.30555 | 0.90978 |
| Gd | Crawfor | TA | Crawfor | 0.5318 | 1.2495 | 1.65451 | 0.94364 |
| Ex | Gilbert | TA | Gilbert | 0.5462 | 2.1441 | 5.63345 | 0.81605 |
| Fa | Crawfor | Gd | Crawfor | 0.5469 | 0.63042 | 1.13098 | 0.35141 |
| Ex | Crawfor | Fa | Crawfor | 0.6213 | 1.95252 | 4.6565 | 0.81871 |
| Ex | Edwards | Gd | Edwards | 0.7165 | 1.56501 | 2.85781 | 0.85704 |
| Ex | NAmes | Fa | NAmes | 0.735 | 1.71085 | 3.54015 | 0.82681 |
| Gd | BrkSide | TA | BrkSide | 0.7548 | 1.33077 | 1.96568 | 0.90094 |
| Ex | Timber | TA | Timber | 0.7872 | 1.45755 | 2.45528 | 0.86525 |
| Gd | Veenker | TA | Veenker | 0.8094 | 1.57891 | 2.99045 | 0.83364 |
| Ex | Gilbert | Gd | Gilbert | 0.9231 | 1.89923 | 4.97441 | 0.72513 |
| Ex | NAmes | Gd | NAmes | 0.964 | 1.37269 | 2.26031 | 0.83363 |
| Ex | Crawfor | TA | Crawfor | 0.9809 | 1.53803 | 3.10309 | 0.76231 |
| Ex | NoRidge | TA | NoRidge | 0.9863 | 1.7757 | 4.60109 | 0.6853 |
| Ex | Timber | Gd | Timber | 0.989 | 1.32358 | 2.1173 | 0.82741 |
| Gd | NWAmes | TA | NWAmes | 0.9973 | 1.14326 | 1.45172 | 0.90034 |
| Gd | Gilbert | TA | Gilbert | 0.9976 | 1.12893 | 1.40309 | 0.90835 |
| Ex | BrkSide | TA | BrkSide | 0.9979 | 1.70701 | 4.47574 | 0.65104 |
| Fa | Edwards | TA | Edwards | 0.999 | 0.81622 | 1.18964 | 0.56002 |
| Ex | NoRidge | Gd | NoRidge | 0.9998 | 1.42539 | 2.84808 | 0.71336 |
| Ex | Veenker | TA | Veenker | 1 | 1.48233 | 3.38094 | 0.6499 |
| Ex | ClearCr | Gd | ClearCr | 1 | 0.74287 | 1.53718 | 0.35901 |
| Fa | Crawfor | TA | Crawfor | 1 | 0.78772 | 1.41316 | 0.43908 |
| Gd | ClearCr | TA | ClearCr | 1 | 1.16342 | 1.69201 | 0.79996 |
| Fa | NAmes | Gd | NAmes | 1 | 0.80234 | 1.41797 | 0.45399 |
| Fa | SWISU | Gd | SWISU | 1 | 0.81134 | 1.41684 | 0.46461 |
| Ex | Blmngtn | Gd | Blmngtn | 1 | 1.3996 | 3.73441 | 0.52455 |
| Gd | NoRidge | TA | NoRidge | 1 | 1.24577 | 2.48919 | 0.62347 |
| Ex | BrkSide | Gd | BrkSide | 1 | 1.28272 | 3.54955 | 0.46354 |
| Ex | ClearCr | TA | ClearCr | 1 | 0.86428 | 1.7751 | 0.42081 |
| Ex | Crawfor | Gd | Crawfor | 1 | 1.23091 | 2.48346 | 0.61009 |
| Ex | MeadowV | Gd | MeadowV | 1 | 1.06977 | 4.11202 | 0.27831 |
| Ex | MeadowV | TA | MeadowV | 1 | 0.94604 | 2.52907 | 0.35388 |
| Ex | NWAmes | Gd | NWAmes | 1 | 1.05675 | 2.12957 | 0.52438 |
| Ex | NWAmes | TA | NWAmes | 1 | 1.20813 | 2.40245 | 0.60754 |
| Ex | OldTown | Gd | OldTown | 1 | 1.13046 | 1.80836 | 0.70668 |
| Ex | Veenker | Gd | Veenker | 1 | 0.93883 | 2.08227 | 0.42329 |
| Fa | IDOTRR | TA | IDOTRR | 1 | 0.88661 | 1.36461 | 0.57605 |
| Fa | NAmes | TA | NAmes | 1 | 0.92214 | 1.60525 | 0.52973 |
| Fa | OldTown | TA | OldTown | 1 | 0.86797 | 1.41349 | 0.53298 |
| Fa | SWISU | TA | SWISU | 1 | 0.90398 | 1.49195 | 0.54773 |
| Fa | Sawyer | Gd | Sawyer | 1 | 0.9339 | 2.5245 | 0.34548 |
| Fa | Sawyer | TA | Sawyer | 1 | 0.96054 | 2.508 | 0.36788 |
| Gd | BrDale | TA | BrDale | 1 | 1.14945 | 3.07285 | 0.42997 |
| Gd | MeadowV | TA | MeadowV | 1 | 0.88434 | 2.36413 | 0.3308 |
| Gd | NPkVill | TA | NPkVill | 1 | 1.09998 | 3.01966 | 0.40069 |
| Gd | SWISU | TA | SWISU | 1 | 1.11418 | 1.74102 | 0.71303 |
| Gd | Sawyer | TA | Sawyer | 1 | 1.02853 | 1.40442 | 0.75324 |
| Gd | Somerst | TA | Somerst | 1 | 1.07806 | 1.75813 | 0.66104 |
| Gd | Timber | TA | Timber | 1 | 1.10121 | 1.57941 | 0.7678 |

***As we see from the table above that while in some neighborhoods such as CollgCr, there is statistically significant (p-value < .0001 from multiple pooled t test adjusted via tukey-kramer adj) multiplicative increase by factor for 2.09 times in median sales price (109% increase in the median sales price with 95% CI of 39% to 212% ) between Excellent Quality Kitchen and Average Quality Kitchen, such a statistically significant difference is not present between Excellent Quality Kitchen and Average Quality Kitchen in neighborhood of NW Ames( p-val = approx. 1 from multiple pooled t test adjusted via tukey-kramer adj). Such differences in multiplicative factor of median sales price among levels of Kitchen Quality is differing with neighborhood. Also, even after adjusting p-values for many comparisons (2485 simultaneous comparisons) and being more conservative on statistically significant findings, interaction effects are observed in some pairs, hence interactions are important in case at hand. Since interactions are both significant and important, interpretation of main effects would not be justified. Hence, we would now treat each combination of KitchenQual\*Neighborhood as an independent treatment and run multiple comparisons. The results below found 863 pairs out of 2485 as statistically significant. The estimate and 95% CI for multiplicative factor for the ratio of median sale prices are also added into the table.***

***Note: Final Multiple Comparisons table for statistically significant results and detailed table containing all the comparisons are added with the final deliverables.***

# The Statistical Conclusion.

***There is sufficient evidence at alpha=0.05 to suggest that median sales prices among different levels of kitchen quality vary with neighborhood in which house is located (p-value for interaction term < 0.0001 from F-Test). That is to say influence of kitchen quality on sales prices vary with levels of neighborhood. Features Kitchen Quality and Neighborhood combined explain about 68.2% variability(R-Sq) in sale prices. This practical effect size is quite significant. Comparisons between log sales price corresponding to different levels of kitchen quality variable for same sites further supports the hypothesis that interaction is important and significant (p-vals were adjusted by tukey-kramer adjustment). Since interaction was significant, finding main effects would not be justified. Hence, each combination of KitchenQual\*Neighborhood was treated as an individual mean or treatment and multiple comparisons were carried out by adjusting p-val by tukey-kramer adjustment. 863 pairs out of all 2485 comparisons are found to be statistically significant. The results in each row of table is to be interpreted as***

***ratio of (median sale price for kitchen Quality Q1 at neighborhood N1 / median sale price for kitchen Quality Q2 at neighborhood N2 ) = Estimate\_NormalScale***

# The Scope of inference.

***Since this was an observational study, any causal conclusions as kitchen Quality and neighborhood together cause the change in median sales price would be speculative as an observational study can uncover associations and not causations. As the selection method for the houses in the study was not specified, generalizing the results to the whole population of houses in the area would be based on an assumption that these samples are as representative as random samples from such a population are. Results of the study could be biased if it was found that a high degree of spatial correlation is present in sale prices within or among neighborhood. Hence, results of this study should be carefully accessed.***

# The APPENDIX.

## **SAS Code.**

FILENAME REFFILE '/RajatChandna/folders/myfolders/CleanedDataForANOVA.csv';

PROC IMPORT DATAFILE=REFFILE

DBMS=CSV

OUT=WORK.KaggData replace;

GETNAMES=YES;

RUN;

proc means data=KaggData n mean max min range std fw=8;

class KitchenQual Neighborhood;

var SalePrice;

output out=meansout mean=mean std=std;

title 'Summary of Sales Prices';

run;

/\* Remove the spurious obervations \*/

data summarystats;

set meansout;

if \_TYPE\_=0 then delete;

if \_TYPE\_=1 then delete;

if \_TYPE\_=2 then delete;

run;

/\* Prepare the data to be plotted, calc upper and lower ends of standard dev bars \*/

data plottingdata;

set summarystats;

lower=mean - std;

upper=mean + std;

run;

/\* Sort the data by Neighborhood \*/

proc sort data=plottingdata;

by Neighborhood;

run;

proc print data=plottingdata;

run;

/\* Plotting the data \*/

proc sgplot data=plottingdata;

scatter x=Neighborhood y=mean / group=KitchenQual yerrorlower=lower

yerrorupper=upper

markerattrs=(symbol=CircleFilled) ;

series x=Neighborhood y=mean / group=KitchenQual ;

title1 'Plot Means with Standard Deviations Bars from Calculated Data';

label mean='Average Sales Price';

run;

/\* Above Plot shows that spread is higher at some higher values of mean, hence

Non Constant variance could be present( points at site No Ridge). Residual plot would clarify further \*/

proc glm data=KaggData plots=(DIAGNOSTICS RESIDUALS);

class KitchenQual Neighborhood;

model SalePrice = KitchenQual Neighborhood KitchenQual\*Neighborhood;

run;

/\* Residual Plots show funnel like pattern and also skewed residuals.

Log Transformation could help in rectifying such situation \*/

/\* Performing Log Transformation and Model Assumption Validation \*/

data kaggdata;

set kaggdata;

LogSalePrice = Log(SalePrice); /\* In \*/

run;

/\* To check, for outliers based upon high GrLiving Area as they might not

be representative of all other houses in the area\*/

proc sgscatter data=Work.Kaggdata;

plot SalePrice\*GrLivArea;

run;

/\* Houses with Gr Living Area > 4000 are not representative of other houses in data set

hence removing there outlier values \*/

data kaggdata;

set kaggdata;

if GrLivArea < 4000;

run;

proc means data=KaggData n mean max min range std fw=8;

class KitchenQual Neighborhood;

var LogSalePrice;

output out=meansoutlog mean=mean std=std;

title 'Summary of Log Sales Prices';

run;

/\* Remove the spurious obervations \*/

data summarystatsafterlogT;

set meansoutlog;

if \_TYPE\_=0 then delete;

if \_TYPE\_=1 then delete;

if \_TYPE\_=2 then delete;

run;

/\* Prepare the data to be plotted, calc upper and lower ends of standard dev bars \*/

data plottingdataafterlogT;

set summarystatsafterlogT;

lower=mean - std;

upper=mean + std;

run;

/\* Sort the data by Neighborhood \*/

proc sort data=plottingdataafterlogT;

by Neighborhood;

run;

proc sgplot data=plottingdataafterlogT;

scatter x=Neighborhood y=mean / group=KitchenQual yerrorlower=lower

yerrorupper=upper

markerattrs=(symbol=CircleFilled) ;

series x=Neighborhood y=mean / group=KitchenQual ;

title1 'Plot Means with Standard Deviations Bars from Calculated Data';

label mean='Average Log Sales Price';

run;

/\* Running the Model, Using proc GLM for obtaining Type 3 SS table and R sq(effect size)

and then using Proc Mixed to obtain formatted Multiple Comparisons table, if necessary\*/

proc glm data=KaggData;

class KitchenQual Neighborhood;

model LogSalePrice = KitchenQual Neighborhood KitchenQual\*Neighborhood;

run;

/\* Storing Comparison Table to a dataset so that only statistically

significant differences can be extracted \*/

ods output diffs=ComparisonData;

proc mixed data=KaggData plots=RESIDUALPANEL;

class KitchenQual Neighborhood;

model LogSalePrice = KitchenQual Neighborhood KitchenQual\*Neighborhood;

lsmeans KitchenQual\*Neighborhood/pdiff diff cl adjust=tukey;

run;

ods output on;

ods exclude none;

/\* Find Same site Diff in Kitchen Qual on Sale Price \*/

/\* This further bolsters that interaction is important as we see for some sites that

differences in Kitchen Qual on Price are not significant but for few sites, diff are highly significant \*/

data CompareSameSite\_KitchenQual;

set ComparisonData;

if Neighborhood = \_Neighborhood;

run;

/\* To convert log back to normal scale. Interpretation would be Media Y/Median X = e ^ estimate\*/

data CompareSameSite\_KitchenQual;

set CompareSameSite\_KitchenQual;

Estimate\_NormalScale = exp(Estimate);

UpperCI\_NormalScale = exp(AdjUpper);

LowerCI\_NormalScale = exp(AdjLower);

run;

proc sort data=CompareSameSite\_KitchenQual;

by AdjP;

run;

proc print data=CompareSameSite\_KitchenQual;

var KitchenQual Neighborhood \_KitchenQual \_Neighborhood Adjp Estimate\_NormalScale UpperCI\_NormalScale LowerCI\_NormalScale;

run;

/\* To find all ab kind off means that are statistically significant \*/

/\* 863 off 2485 pairs are statistically significant, interaction is important one and complex in nature as

nearly 25% of pairs differ \*/

data StatisticallySigDiffs;

set ComparisonData;

Estimate\_NormalScale = exp(Estimate);

UpperCI\_NormalScale = exp(AdjUpper);

LowerCI\_NormalScale = exp(AdjLower);

if AdjP <= 0.05;

run;

proc sort data=StatisticallySigDiffs;

by AdjP descending Estimate\_NormalScale;

run;

proc print data=StatisticallySigDiffs;

var KitchenQual Neighborhood \_KitchenQual \_Neighborhood Adjp Estimate\_NormalScale UpperCI\_NormalScale LowerCI\_NormalScale;

run;