

Date: May 17, 2024

To: Perry Cox, M.D., Chief of Medicine

Subject: Analysis of Variances for Admissions and Births

---

Sacred Heart has provided the data set taken from *American Hospital Association Guide to the Health-Care Field*, published in Chicago, Illinois. The company has been provided an opportunity to expand and change the type of ownership. Before moving forward with testing the data, the company wants to make sure that the sample size and data are viable. This report will provide the Analysis of Variances (ANOVA) for admissions and births by geographical region and/or control. The data will help the company understand how different groups respond by comparing the samples to see whether there is a significant or no significant differences in the data and therefore approve the dataset for further analytical testing.

Contents of the report:

- Define Data terms
- ANOVA for Admissions
  - One- Factor ANOVA with:
    - Geographical Region
    - Control
  - Two-Factor ANOVA with Geographical Region and Control
- ANOVA for Births
  - One- Factor ANOVA with:
    - Geographical Region
    - Control
  - Two-Factor ANOVA with Geographical Region and Control
- Conclusion

### **DEFINE DATA TERMS:**

**Hospital:** The numeric value given to a hospital on the report to make running the report easier

**Geographic Region:** The variable is coded from 1 to 7, and the numbers represent the following regions-

- 1 = South
- 2 = Northeast
- 3 = Midwest
- 4 = Southwest
- 5 = Rocky Mountain
- 6 = California
- 7 = Northwest

**Control:** A type of ownership. Four categories of control are included in the database-

- 1 = government, nonfederal

2 = nongovernment, not-for-profit

3 = for-profit

4 = federal government

**Service:** The type of hospital. The two types of hospitals used in this database are-

1 = general medical

2 = psychiatric

**Beds:** Number of beds in the hospital

**Admissions:** The number of people admitted to the hospital

**Census:** Average number of patients per day in a hospital over a given period of time; admitted patients and outpatients are counted separately

**Outpat. (Outpatient) Visits:** Any visit made during the person's reference period to a hospital outpatient department

**Births:** Number of births in the hospital

**Tot. Exp. (Total Expenditure):** All costs and expenses reasonably and properly incurred in units of \$1,000.

**Payroll Exp. (Payroll Expense):** The amount of salaries and wages paid to employees in exchange for services rendered by them to a business in units of \$1,000

**Personnel:** Number of people that work in the hospital

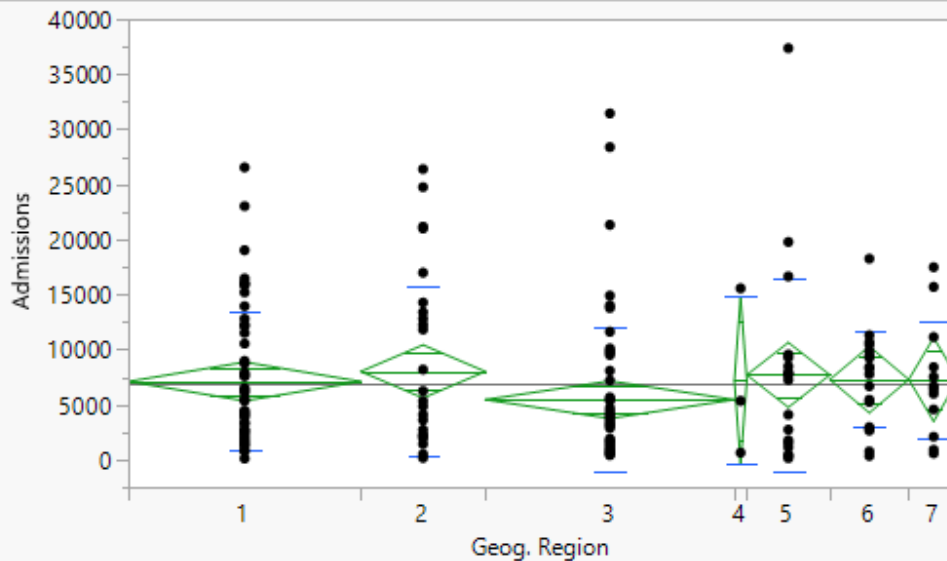
### **ANOVA for Admission:**

#### **One-Factor ANOVA for Admissions with Geographical Region**

Sacred Heart wants to know if there is a significant difference in the population means between Admissions and Geographical Region. After running the test, the data shows that the null hypothesis can't be rejected due to the statistical value, which is not small (as shown below).

## Fit Group

### Oneway Analysis of Admissions By Geog. Region



### Oneway Anova

#### Summary of Fit

Rsquare	0.020792
Adj Rsquare	-0.00965
Root Mean Square Error	6678.128
Mean of Response	6831.835
Observations (or Sum Wgts)	200

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Geog. Region	6	182761795	30460299	0.6830	0.6636
Error	193	8607296329	44597390		
C. Total	199	8790058124			

#### Means for Oneway Anova

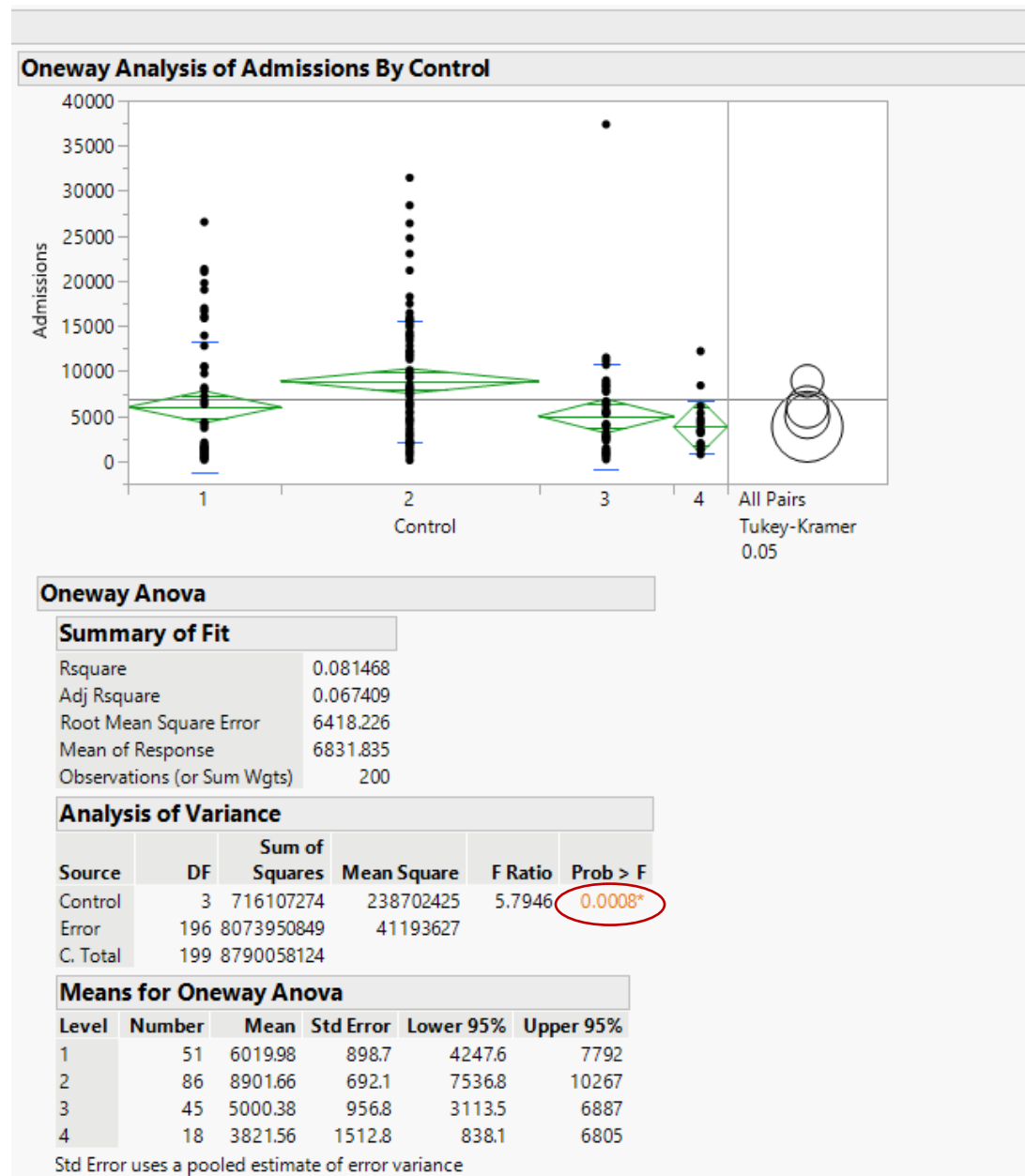
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	56	7103.11	892.4	5343	8863
2	30	8012.90	1219.3	5608	10418
3	60	5447.32	862.1	3747	7148
4	3	7182.00	3855.6	-423	14787
5	20	7699.05	1493.3	4754	10644
6	19	7283.11	1532.1	4261	10305
7	12	7288.42	1927.8	3486	11091

Std Error uses a pooled estimate of error variance

## One-Factor ANOVA for Admissions with Control

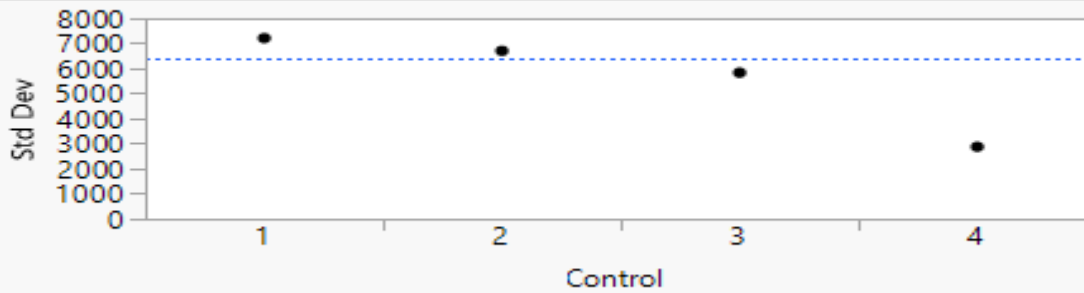
Sacred Heart wants to know if there is a significant difference in the population means for Admissions and Control. After running the test, the data shows that the null hypothesis can be

rejected due to the statistical value (as shown below), which is small. Because the null hypothesis was rejected, the data needs further analyzation by considering the whether the variances are equal or not.



To consider the equality of the variances, the data needs to be run and analyzed (as shown below). Based on the data, the statistical values are small and the variances are assumed to be unequal, which means that the ANOVA output can't be trusted and the null hypothesis continues to be rejected.

### Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
1	51	7218.155	5986.877	5177.765
2	86	6709.855	5193.220	5068.988
3	45	5840.377	3441.609	3251.533
4	18	2881.369	2062.568	1975.000

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.9443	3	196	0.4203
Brown-Forsythe	2.9934	3	196	0.0320*
Levene	5.9438	3	196	0.0007*
Bartlett	5.3465	3	.	0.0011*

Due to the variances proving to be unequal, the next step is to check the Welch's Test (as shown below). Based on the Welch's Test there is a significant difference in the population meant of control (government (nonfederal), nongovernment (not-for-profit), for-profit, federal government).

### Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
9.0729	3	83.305	<.0001*

The last step is to compare the means by using the Tukey-Kramer Honest Significant Difference, the most desired of the methods. To identify mean differences, the Connecting Letters Report needs to be examined. Based on the Connecting Letters Report (as shown below); For-Profit and Federal Government ownership have means that are similar, but their means are significantly different from the Nongovernment (not-for-profit) ownership. The Government (nonfederal) mean falls somewhere in between.

## Means Comparisons

### Comparisons for all pairs using Tukey-Kramer HSD

#### Confidence Quantile

q*	Alpha
2.59122	0.05

#### HSD Threshold Matrix

Abs(Dif)-HSD

	2	1	3	4
2	-2536.2	-57.6	841.4	769.4
1	-57.6	-3293.4	-2381.8	-2361.1
3	841.4	-2381.8	-3506.1	-3459.4
4	769.4	-2361.1	-3459.4	-5543.7

Positive values show pairs of means that are significantly different.

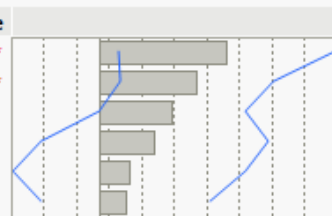
#### Connecting Letters Report

Level		Mean
2	A	8901.6628
1	A B	6019.9804
3	B	5000.3778
4	B	3821.5556

Levels not connected by same letter are significantly different.

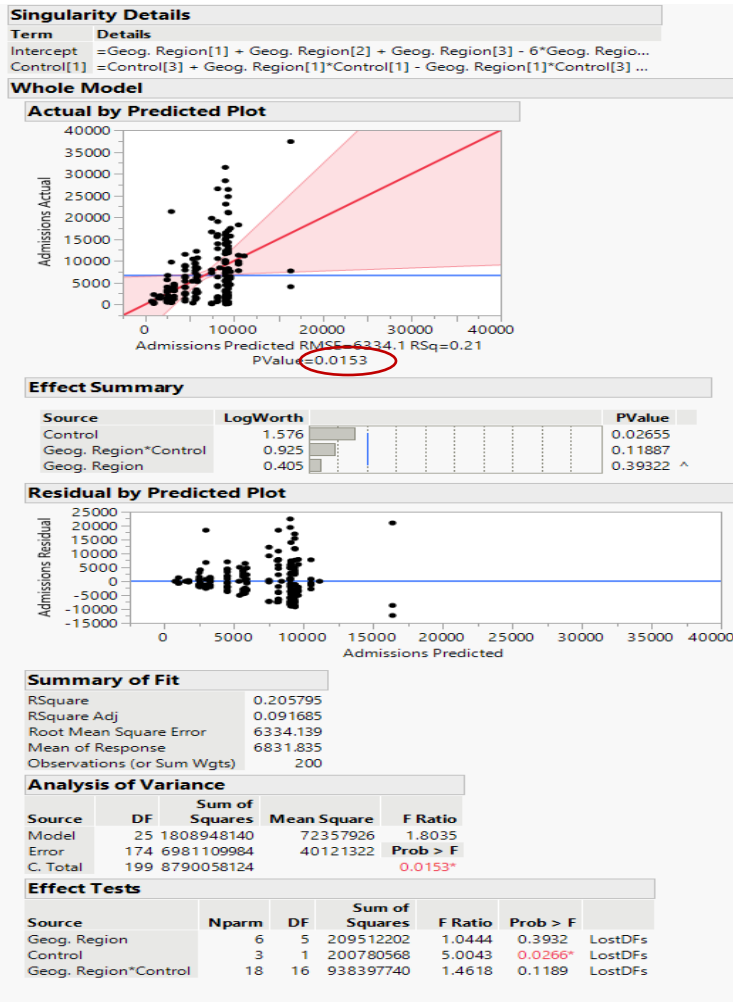
#### Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
2	4	5080.107	1663.590	769.37	9390.840	0.0136*
2	3	3901.285	1180.851	841.44	6961.135	0.0062*
2	1	2881.682	1134.334	-57.63	5820.996	0.0569
1	4	2198.425	1759.617	-2361.14	6757.987	0.5962
3	4	1178.822	1789.958	-3459.36	5817.003	0.9124
1	3	1019.603	1312.681	-2381.85	4421.054	0.8649



## Two-Factor ANOVA for Admissions with Geographical Region and Control

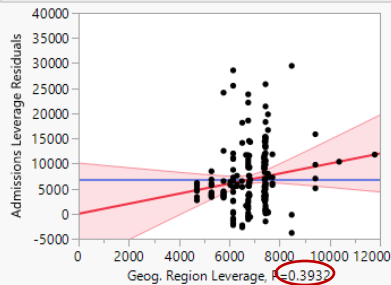
Sacred Heart would like a comparison of the mean differences for Geographical Region and Control to see if there is an interaction between Geographical Region and Control and Admissions. Based on the small statistical value, it shows that there are significant differences in the means and therefore we can initially reject the null hypothesis. The means can be compared individually by Geographical Region, Control, and Geographical Region and Control.



For Geographical region, the result is similar to the one-factor ANOVA that was done previously. The statistical value is not small, therefore the null hypothesis can't be rejected. This means that there is no significant differences in the means.

## Geog. Region

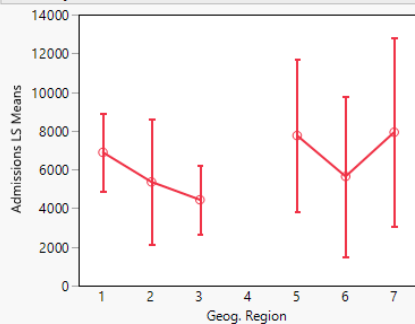
### Leverage Plot



### Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	6889.6099	1030.1532	7103.11
2	5363.4125	1648.1953	8012.90
3	4440.6447	893.5531	5447.32
4	0.0000 NonEstimable	.	7182.00
5	7759.2421	1995.0664	7699.05
6	5645.1746	2097.1934	7283.11
7	7930.6042	2461.7050	7288.42

### Least Squares Means Plot



## LSMeans Differences Tukey HSD

$\alpha = 0.050$   $Q = 2.98315$

LSMean[i]	LSMean[j]						
	1	2	3	4	5	6	7
Mean[i] - Mean[j]							
Std Err Dif							
Lower CL Dif							
Upper CL Dif							
1		0	1526.2	2448.97	6889.61	-869.63	1244.44
		0	1943.65	1363.69	17514.1	2245.33	2336.54
		0	-4272	-1619.1	-45358	-7567.8	-5725.8
		0	7324.39	6517.06	59136.8	5828.52	8214.7
2			-1526.2	0	922.768	5363.41	-2395.8
			1943.65	0	1874.83	17660.3	2587.82
			-7324.4	0	-4670.1	-47320	-10116
			4271.99	0	6515.67	58046.8	5324.04
3					-2449	-922.77	0
					1363.69	1874.83	0
					-6517.1	-6515.7	0
					1619.13	4670.13	0
4						-6889.6	-5363.4
						17514.1	17660.3
						-59137	-58047
						45357.6	47320
5							869.632
							2245.33
							-5828.5
							7567.79
6							
							2336.54
							-8214.7
							5725.83
7							
							1040.99
							2668.56
							-6919.7
							9001.71

Level	Least Sq Mean
7	A 7930.6042
5	A 7759.2421
1	A 6889.6099
6	A 5645.1746
2	A 5363.4125
3	A 4440.6447
4	A 0.0000

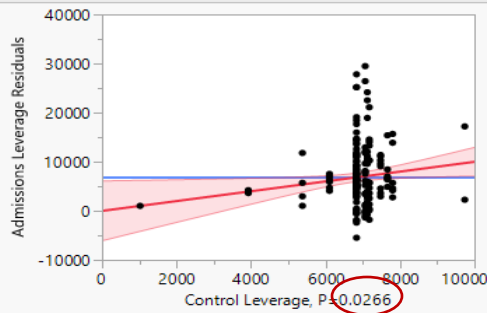
Levels not connected by same letter are significantly different.

For Control, the result is similar to the one-factor ANOVA that was done previously. The statistical value is small coupled with the Connecting Letters Report, which shows 2 different letters (A and B), therefore the null hypothesis can be rejected. Based on the information below, there is a significant difference among the means.



## Control

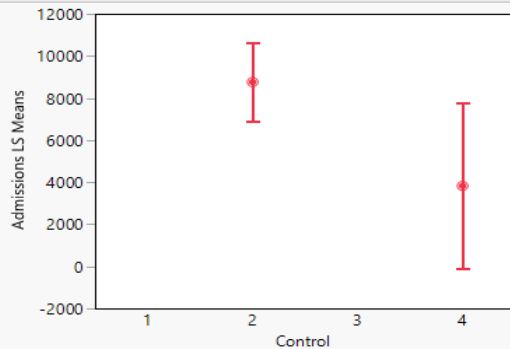
### Leverage Plot



### Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	0.0000	NonEstimable	6019.98
2	8763.3077	935.7752	8901.66
3	0.0000	NonEstimable	5000.38
4	3827.9464	1997.9144	3821.56

### Least Squares Means Plot



## LSMeans Differences Tukey HSD

$\alpha = 0.050$   $Q = 2.59404$

		LSMean[j]			
LSMean[i]	Mean[i]-Mean[j]	1	2	3	4
	Std Err Dif				
	Lower CL Dif				
	Upper CL Dif				
		0	-8763.3	0	-3827.9
1		0	8557.88	7882.77	8549.91
		0	-30963	-20448	-26007
		0	13436.2	20448.3	18350.9
2		8763.31	0	8763.31	4935.36
		8557.88	0	3192.11	2206.2
		-13436	0	482.841	-787.63
		30962.8	0	17043.8	10658.4
3		0	-8763.3	0	-3827.9
		7882.77	3192.11	0	3663.23
		-20448	-17044	0	-13331
		20448.3	-482.84	0	5674.64
4		3827.95	-4935.4	3827.95	0
		8549.91	2206.2	3663.23	0
		-18351	-10658	-5674.6	0
		26006.8	787.632	13330.5	0

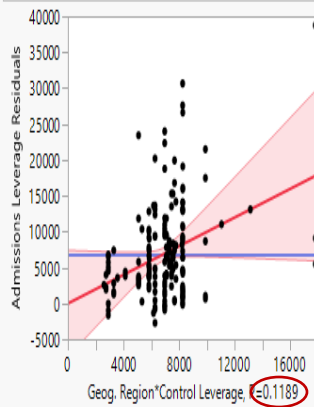
Level		Least Sq Mean
2	A	8763.3077
4	A B	3827.9464
1	A B	0.0000
3	B	0.0000

Levels not connected by same letter are significantly different.

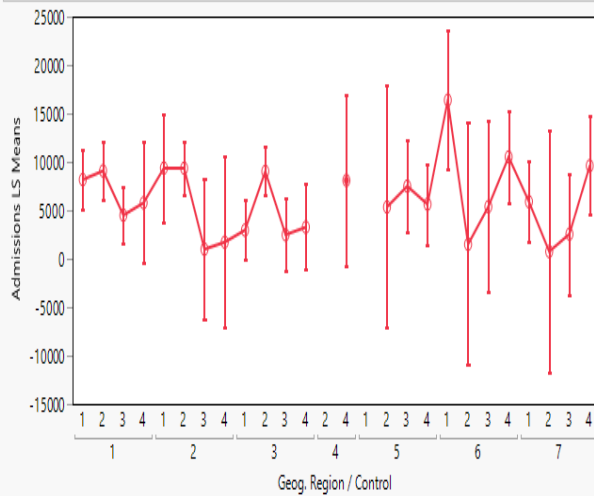
The comparison of Geographical Region and Control shows a not small statistical value coupled with the Connecting Letters Report, which only shows letter A, confirms that there is no significant difference among the means.

# Geog. Region\*Control

Leverage Plot



Least Squares Means Plot



Level		Least Sq Mean
5,3	A	16386.333
7,3	A	11142.000
6,2	A	10524.143
7,2	A	9610.667
2,1	A	9381.600
2,2	A	9352.050
1,2	A	9083.176
3,2	A	9030.840
7,4	A	8408.000
1,1	A	8180.235
4,2	A	8099.500
5,1	A	7504.857
6,3	A	5907.556
1,4	A	5786.750
5,2	A	5642.778
6,1	A	5393.000
4,4	A	5347.000
1,3	A	4508.278
3,4	A	3272.875
3,1	A	2967.500
7,1	A	2561.750
3,3	A	2491.364
2,4	A	1722.000
5,4	A	1503.000
2,3	A	998.000
6,4	A	756.000
4,1	A	0.000
4,3	A	0.000

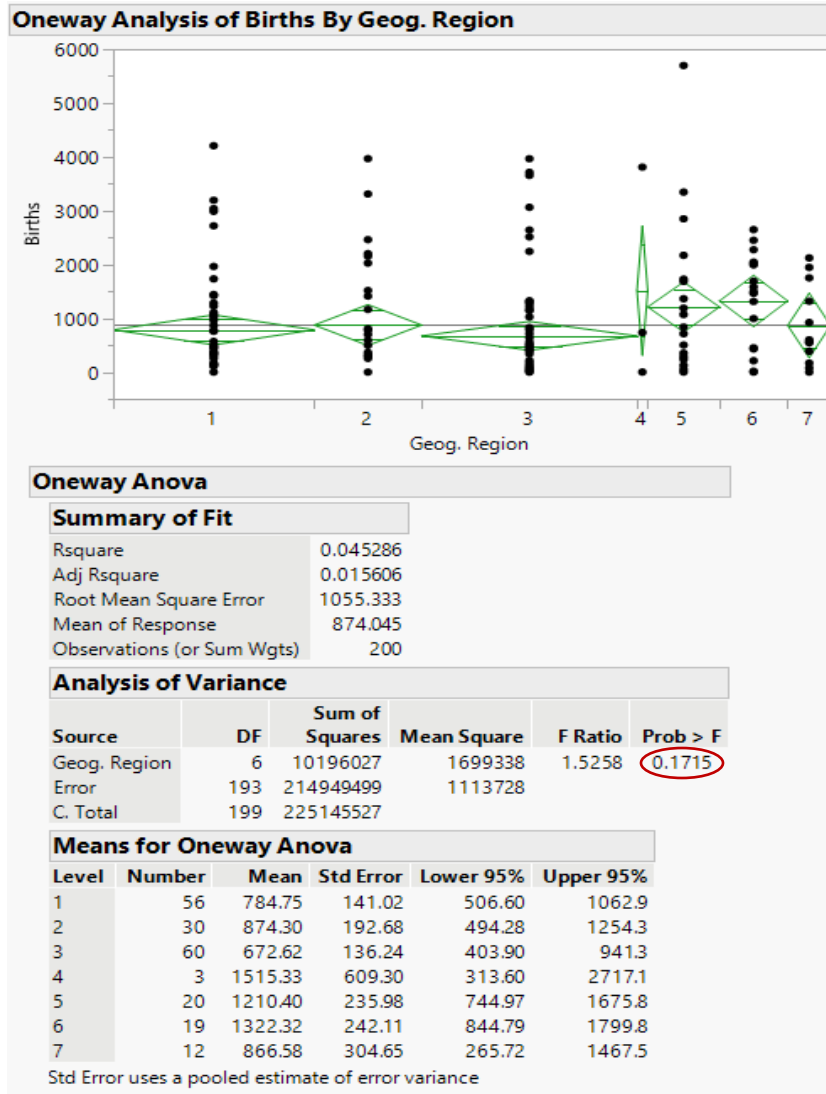
**LSMeans Differences Tukey HSD**  
 $\alpha = 0.050$   $Q = 3.77749$

[illegible]

### ANOVA for Births:

## One-Factor ANOVA for Births with Geographical Region

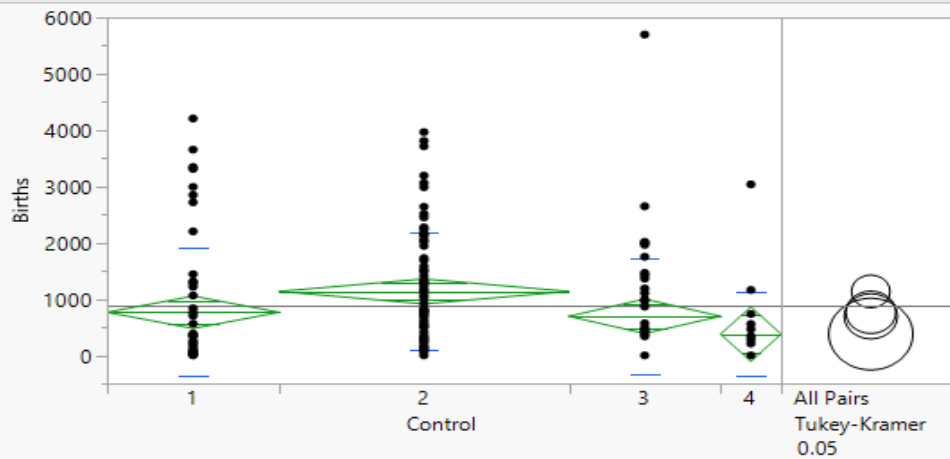
Sacred Heart wants to know if there is a significant difference in the population means for Births and Geographical Region. After running the test, the data shows that the null hypothesis can't be rejected due to the statistical value, which is not small (as shown below).



## One-Factor ANOVA for Births with Control

Sacred Heart wants to know if there is a significant difference in the population means for Admissions and Control. After running the test, the data shows that the null hypothesis can be rejected due to the statistical value (as shown below), which is small. Because the null hypothesis was rejected, the data needs further analyzation by considering the whether the variances are equal or not.

### Oneway Analysis of Births By Control



### Oneway Anova

#### Summary of Fit

Rsquare	0.055179
Adj Rsquare	0.040718
Root Mean Square Error	1041.786
Mean of Response	874.045
Observations (or Sum Wgts)	200

#### Analysis of Variance

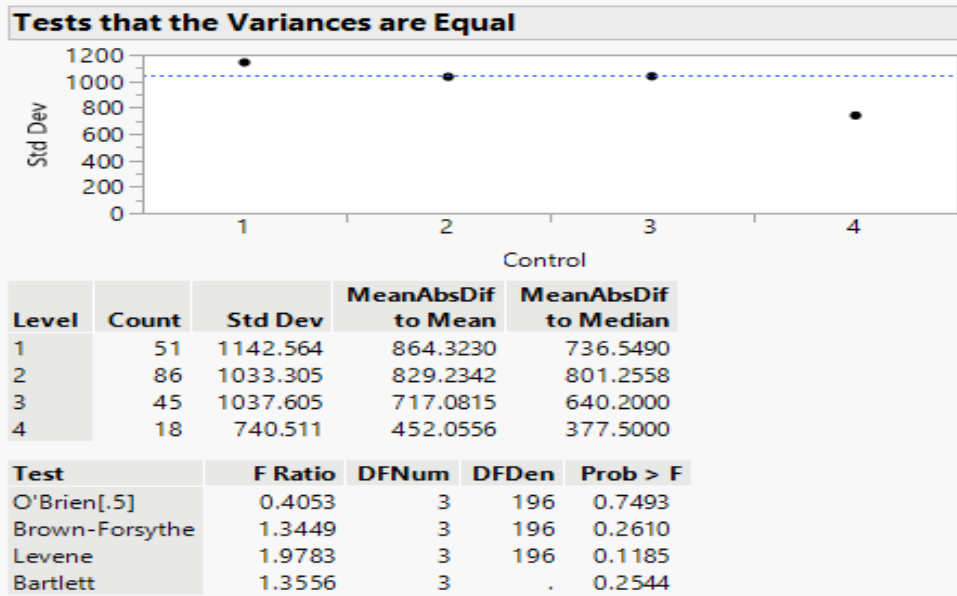
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Control	3	12423343	4141114	3.8156	0.0109*
Error	196	212722183	1085317		
C. Total	199	225145527			

#### Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	51	765.12	145.88	477.4	1052.8
2	86	1136.58	112.34	915.0	1358.1
3	45	694.38	155.30	388.1	1000.7
4	18	377.50	245.55	-106.8	861.8

Std Error uses a pooled estimate of error variance

To consider the equality of the variances, the data needs to be run and analyzed (as shown below). Based on the data, the statistical values are not small and the null hypothesis can't be rejected. In this case, the Welch's Test can be skipped.



The last step is to compare the means by using the Tukey-Kramer Honest Significant Difference. To identify mean differences, the Connecting Letters Report needs to be examined. Based on the Connecting Letters Report (as shown below); Government (nonfederal) and For Profit ownership have means that are similar and fall in the middle. The means are significantly different between the Nongovernment (not-for-profit) ownership and the Federal Government.

## Means Comparisons

### Comparisons for all pairs using Tukey-Kramer HSD

#### Confidence Quantile

q*	Alpha
2.59122	0.05

#### HSD Threshold Matrix

Abs(Dif)-HSD

	2	1	3	4
2	-411.67	-105.64	-54.46	59.38
1	-105.64	-534.58	-481.37	-352.48
3	-54.46	-481.37	-569.10	-435.98
4	59.38	-352.48	-435.98	-899.83

Positive values show pairs of means that are significantly different.

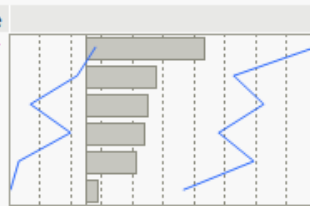
#### Connecting Letters Report

Level		Mean
2	A	1136.5814
1	A B	765.1176
3	A B	694.3778
4	B	377.5000

Levels not connected by same letter are significantly different.

#### Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
2	4	759.0814	270.0285	59.377	1458.786	0.0276*
2	3	442.2036	191.6719	-54.461	938.868	0.0999
1	4	387.6176	285.6154	-352.476	1127.711	0.5278
2	1	371.4637	184.1214	-105.636	848.564	0.1851
3	4	316.8778	290.5401	-435.977	1069.732	0.6957
1	3	70.7399	213.0702	-481.373	622.852	0.9873



## Two-Factor ANOVA for Births with Geographical Region and Control

Sacred Heart would like a comparison of the mean differences for Geographical Region and Control to see if there is an interaction between Geographical Region and Control and Births. Based on the small statistical value, it shows that there are significant differences in the means and therefore we can initially reject the null hypothesis. The means can be compared individually by Geographical Region, Control, and Geographical Region and Control.

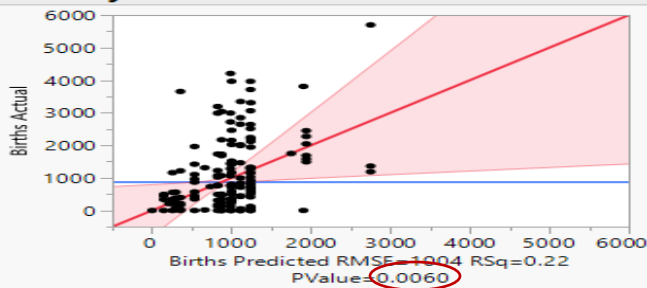
### Singularity Details

#### Term Details

Intercept =Geog. Region[1] + Geog. Region[2] + Geog. Region[3] - 6\*Geog. Region[1]  
Control[1] =Control[3] + Geog. Region[1]\*Control[1] - Geog. Region[1]\*Control[3] ...

### Whole Model

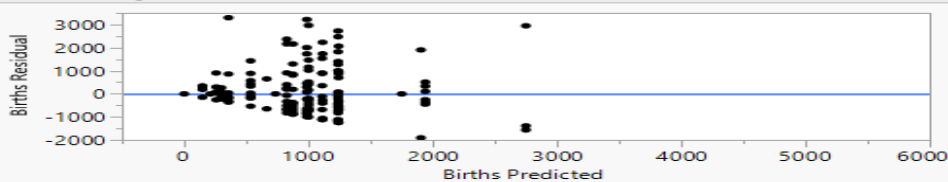
#### Actual by Predicted Plot



#### Effect Summary

Source	LogWorth	PValue
Control	2.146	0.00715
Geog. Region*Control	1.290	0.05125
Geog. Region	0.589	0.25735

#### Residual by Predicted Plot



#### Summary of Fit

RSquare	0.220957
RSquare Adj	0.109026
Root Mean Square Error	1004.009
Mean of Response	874.045
Observations (or Sum Wgts)	200

#### Analysis of Variance

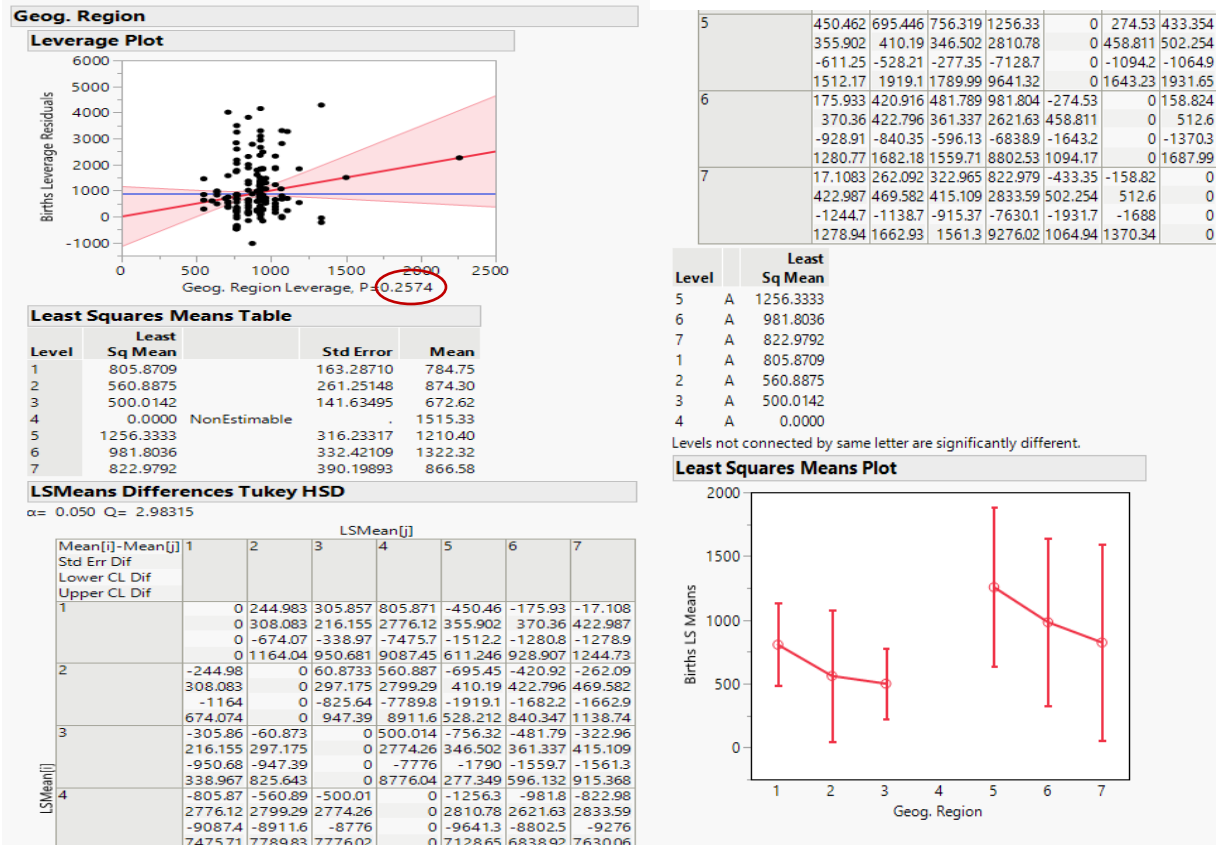
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	25	49747529	1989901	1.9740
Error	174	175397998	1008034	Prob > F
C. Total	199	225145527		0.0060*

#### Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F	LostDFs
Geog. Region	6	5	6657493	1.3209	0.2574	LostDFs
Control	3	1	7468420	7.4089	0.0072*	LostDFs
Geog. Region*Control	18	16	27344020	1.6954	0.0513	LostDFs

For Geographical region, the result is similar to the one-factor ANOVA that was done previously. The statistical value is not small, therefore the null hypothesis can't be rejected. This means that there is no significant differences in the means.

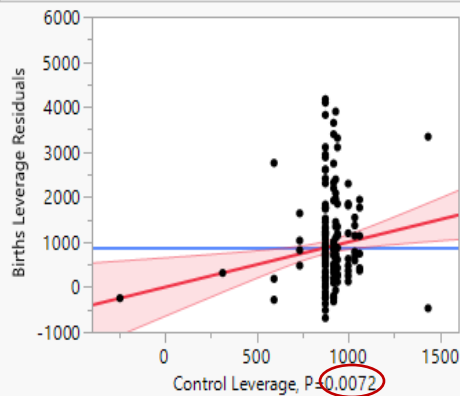




For Control, the result is similar to the one-factor ANOVA that was done previously. The statistical value is small coupled with the Connecting Letters Report, which shows 2 different letters (A and B), therefore the null hypothesis can be rejected (as shown below).

## Control

### Leverage Plot



### Least Squares Means Table

Level	Sq Mean	Least	Std Error	Mean
1	0.0000	NonEstimable	.	765.12
2	1289.7694		148.32748	1136.58
3	0.0000	NonEstimable	.	694.38
4	337.9107		316.68461	377.50

### LSMeans Differences Tukey HSD

$\alpha = 0.050$   $Q = 2.59404$

### LSMeans Differences Tukey HSD

$\alpha = 0.050$   $Q = 2.59404$

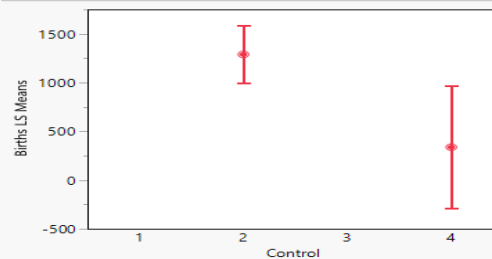
	LSMean[j]	1	2	3	4
Mean[i]-Mean[j]					
Std Err Dif					
Lower CL Dif					
Upper CL Dif					
1					
		0	-1289.8	0	-337.91
		0	1356.49	1249.48	1355.23
		0	-4808.6	-3241.2	-3853.4
		0	2229.02	3241.2	3177.6
2		1289.77		0	1289.77
		1356.49		0	505.973
		-2229		0	-22.747
		4808.56		0	2602.29
3		0	-1289.8		0
		1249.48	505.973		0
		-3241.2	-2602.3		0
		3241.2	22.7474		0
4		337.911	-951.86	337.911	
		1355.23	349.7	580.65	
		-3177.6	-1859	-1168.3	
		3853.43	-44.721	1844.14	

### Least Squares Means Table

Level	Sq Mean
2	A 1289.7694
4	B 337.9107
1	A B 0.0000
3	A B 0.0000

Levels not connected by same letter are significantly different.

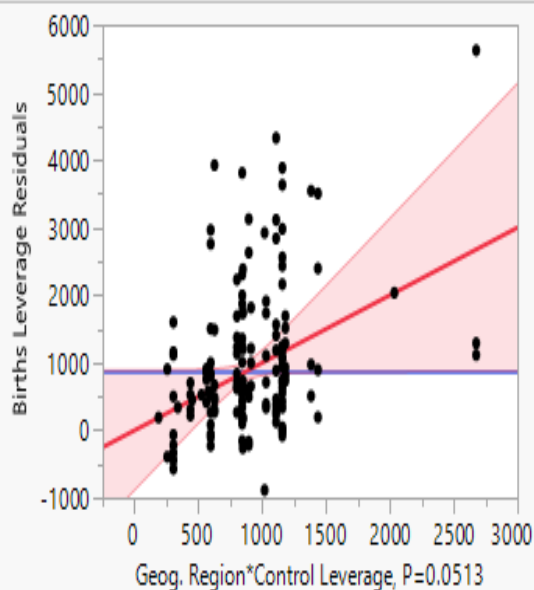
### Least Squares Means Plot



The comparison of Geographical Region and Control shows statistical value that is slightly not small, but the Connecting Letters Report, which shows a majority of a combination of A and B. This seems to have conflicting information and overall shows that there is a slight significance in the means.

## Geog. Region\*Control

### Leverage Plot

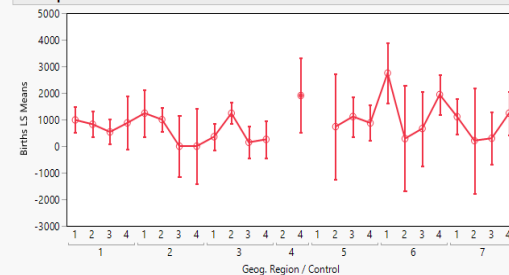


### Least Squares Means Table

Level	Sq Mean
5.3	A 2750.667
6.2	A B 1940.714
4.2	A B 1905.500
7.3	A B 1751.000
2.1	A B 1242.800
7.2	A B 1242.167
3.2	A B 1239.000
5.1	A B 1114.000
6.3	A B 1111.000
2.2	A B 1000.750
1.1	A B 987.118
1.4	A B 876.000
5.2	A B 874.667
1.2	A B 825.588
4.4	A B 735.000
6.1	A B 664.500
1.3	A B 534.778
3.1	B 356.500
7.1	A B 298.750
5.4	A B 286.000
3.4	A B 257.375
6.4	A B 211.000
3.3	B 147.182
4.1	A B 0.000
4.3	A B 0.000
2.3	A B -2.274e-13
2.4	A B -2.728e-12
7.4	A B -4.775e-12

Levels not connected by same letter are significantly different.

### Least Squares Means Plot



LSMeans Differences Tukey HSD  
n = 0.00, C = 5.774

		Mean(I) - Mean(J)	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	4.1	4.2	4.3	4.4	5.1	5.2	5.3	5.4	6.1	6.2	6.3	6.4	7.1	7.2	7.3	7.4
1.1	1.1																													
	1.2																													
1.2	1.1																													
	1.2																													
1.3	1.1																													
	1.2																													
1.4	1.1																													
	1.2																													
2.1	1.1																													
	1.2																													
2.2	1.1																													
	1.2																													
2.3	1.1																													
	1.2																													
2.4	1.1																													
	1.2																													
3.1	1.1																													
	1.2																													
3.2	1.1																													
	1.2																													
3.3	1.1																													
	1.2																													
3.4	1.1																													
	1.2																													
4.1	1.1																													
	1.2																													
4.2	1.1																													
	1.2																													
4.3	1.1																													
	1.2																													
4.4	1.1																													
	1.2																													
5.1	1.1																													
	1.2																													
5.2	1.1																													
	1.2																													
5.3	1.1																													
	1.2																													
5.4	1.1																													
	1.2																													
6.1	1.1																													
	1.2																													
6.2	1.1																													
	1.2																													
6.3	1.1																													
	1.2																													
6.4	1.1																													
	1.2																													
7.1	1.1																													
	1.2																													
7.2	1.1																													
	1.2																													
7.3	1.1																													
	1.2																													
7.4	1.1																													
	1.2																													

## CONCLUSION:

Based on the ANOVA tests conducted, there are a few things that can be taken away. Firstly, the Geographical Region variable had no significant differences between any of the population means regardless of whether it was in relation to Admissions or Births, which makes Geographical Region a good set of data to use. Secondly, the control variable consistently showed significant differences among the population means regardless of whether it was in relation to Admissions or Births, which makes it not a good set of data to use on its own. Finally when Geographical Region and Control were used in conjunction with each other, there was different effects for Admissions and Births. When the test was conducted for Admissions with Geographical region and Control were used in conjunction with each other, there was no significant differences for Births with Geographical region and Control, there was no significant differences in the population means based on the statistical value, but when you look at the Connecting Letter code it showed a significant difference among the population means.