

TO: Hospital Managers  
FROM: Charlie Evert  
DATE: June 7, 2021  
RE: Hospital Admissions and Births are Better Analyzed Via Region than Control

There are differences in hospitals for births and admissions. These differences may be observed amongst particular groups more than others. The following appendices will uncover the differences and similarities for groupings of hospitals. These groupings are by type of ownership (Control) and by geographic region. The major findings of this data are:

- Admissions by Geographic Region and Births by Geographic Region have means that are statistically the same. Thus, each region contributes to the overall model without outliers. Admissions by Control and Births by Control have statistically different means, and should be segmented into several models for analysis. Thus, data would be better analyzed (in terms of leading to reliable insights) by Geographic Region than by Control.
- All 4 models have very poor correlations, and predictive analytics derived from these models should not be seen as very reliable.

The following details the organization of the appendix:

- A. Admissions by Geographic Region
  - B. Admissions by Control
  - C. Births by Control
  - D. Births by Geographic Region
- 
1. Descriptive Statistics
  2. Box Plot
  3. Test for Unequal Variance
  4. Test for Goodness of Fit
  5. Analysis of Variance
  6. Connecting Letters Report

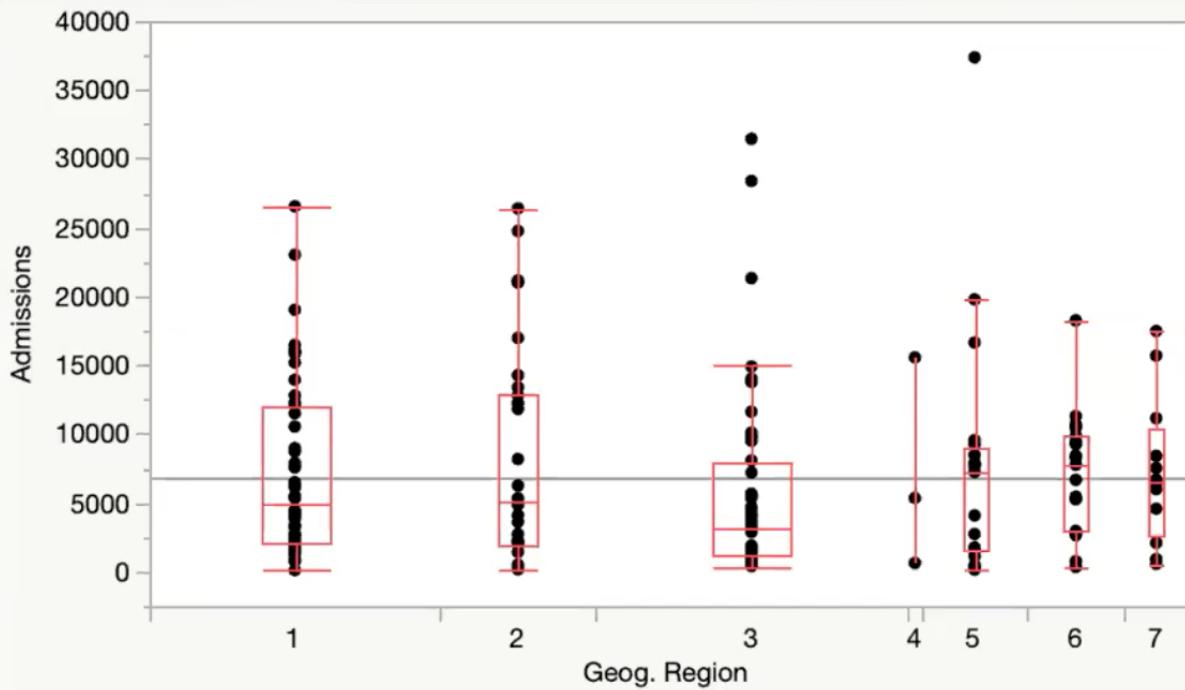
A1

	N	Mean	Variance	Std Dev
Geog. Region	Admissions	Admissions	Admissions	Admissions
1	56	7103	39043166	6248
2	30	8013	59624887	7722
3	60	5447	42528807	6521
4	3	7182	58229251	7631
5	20	7699	76628788	8754
6	19	7283	18622621	4315
7	12	7288	28544384	5343

The means seem to be fairly similar, with an outlier for Region 3. Standard deviations vary more widely, with region 5 and 6 appearing to be much different from the others. Region 4 has very few observations, so data collected for this region should be viewed as less indicative of the overall descriptive statistics.

A2

## Oneway Analysis of Admissions By Geog. Region



I hypothesize that the means will be statistically the same, since the grey line representing the overall mean intersects each box plot. Further testing will need to be done to confirm this, but the distributions of each plot appear very similar. There are a few outliers, specifically in regions 3 and 5, but most data falls within the box plots.

A3

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.7081	6	193	0.6434
Brown-Forsythe	0.7617	6	193	0.6009
Levene	1.2131	6	193	0.3011
Bartlett	1.8396	6	.	0.0872

Warning: Small sample sizes. Use Caution.

## ▼ Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
0.5835	6	22.998	0.7398

Since the probability is greater than .005, variances are statistically equal and this model should be seen as reliable. The O'Brien test is seen as reliable, and since this data passes this test, we should rely upon its findings.

# 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

This is a poor model in terms of goodness of fit. Only 2% of the variation of means are explained in the model. This is a weak R squared, and shows that the data does not correlate greatly.

**▼ Analysis of Variance**

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

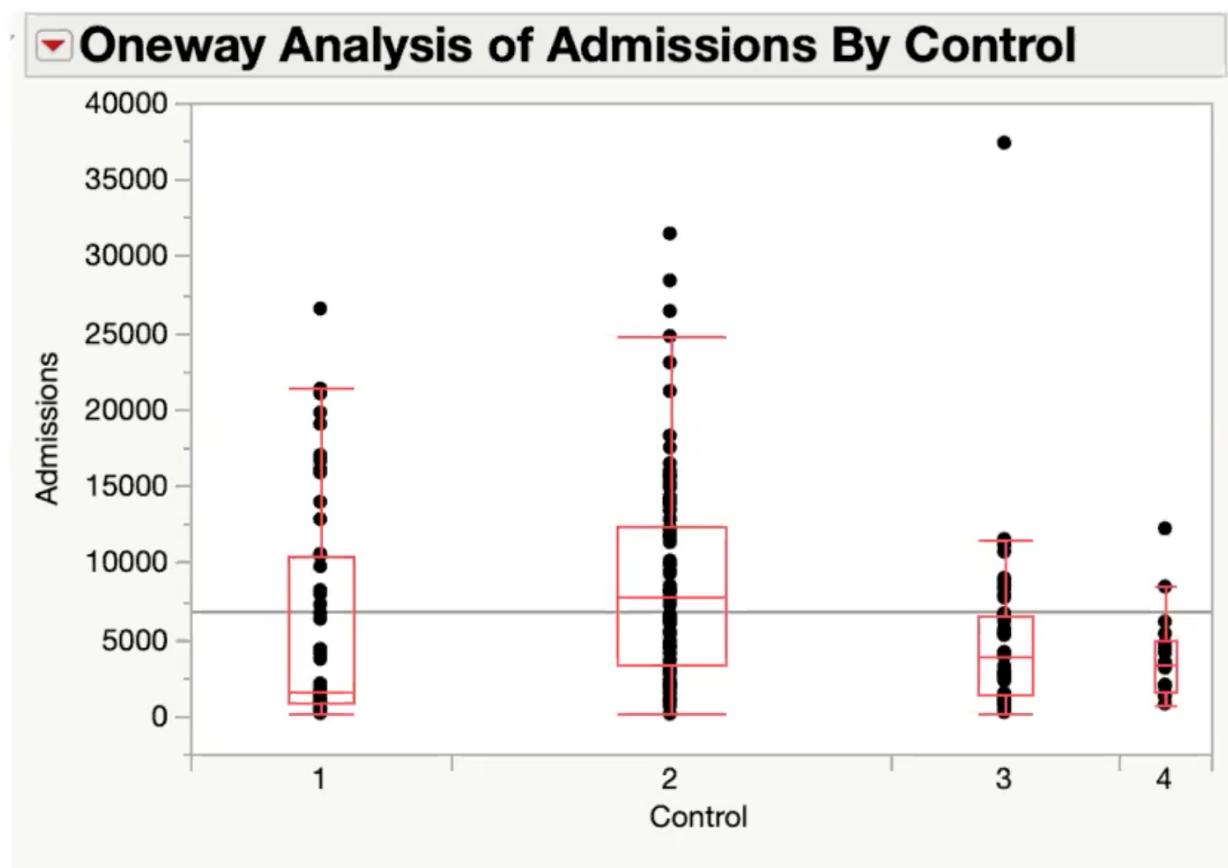
The P value is greater than .005, so this model is statistically significant. Thus, we should use the model since variances fall within acceptable range. Since there are no statistical differences in means, there is no need for further analysis.

B1

Control	N	Admissions		
		Mean	Variance	Std Dev
1	51	6019.9803921569	52101766.059608	7218.1553086372
2	86	8901.6627906977	45022158.437893	6709.8553216812
3	45	5000.3777777778	34110005.10404	5840.3771371411
4	18	3821.5555555556	8302285.5555556	2881.3686948316

Means appear to be very different in that there could be several groups formed from each. The same can be said for standard deviations. It seems that control 1, 3 and 4 should be grouped, but 1, 2 and 3 could be grouped as well. Control 2 and 4 should not be in a group.

B2



Overall mean is not within all boxes, thus means can be hypothesized to be statistically different. This will need further analysis to confirm.

B3

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*

## ▼ 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*

Variances are approximately equal between 4 levels of control according to the O'Brien test. They are not for every other listed test, but the O'Brien test is what we are analyzing to determine this.

## Oneway Anova

### ▼ Summary of Fit

Rsquare	0.081468
Adj Rsquare	0.067409
Root Mean Square Error	6418.226
Mean of Response	6831.835
Observations (or Sum Wgts)	200

Better than other model, still bade but explains 8% of variations in mean. Weak r squared. The R squared explains 8% of the variations from the mean, which is significantly better than the previous model. However, this is still a very low R squared value and shows a very weak correlation between the variables.

<b>Analysis of Variance</b>					
<b>Source</b>	<b>DF</b>	<b>Sum of Squares</b>	<b>Mean Square</b>	<b>F Ratio</b>	<b>Prob &gt; F</b>
Control	3	716107274	238702425	5.7946	0.0008*
Error	196	8073950849	41193627		
C. Total	199	8790058124			

This test is significant, as indicated by the star. For average admissions by control, at least one of the means is statistically different from others. Likely, as discussed before, 2 and 4 should not be grouped together.

## ▼ 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.

This test was done since the model failed its analysis of variance test.

Level 2 and 1 have means that are statistically the same.

Level 1, 3 and 4 have means that are statistically the same.

All 4 do not have means that are statistically the same, they are in the aforementioned groups though.

2 and 3 as well as 2 and 4 have statistically different means.

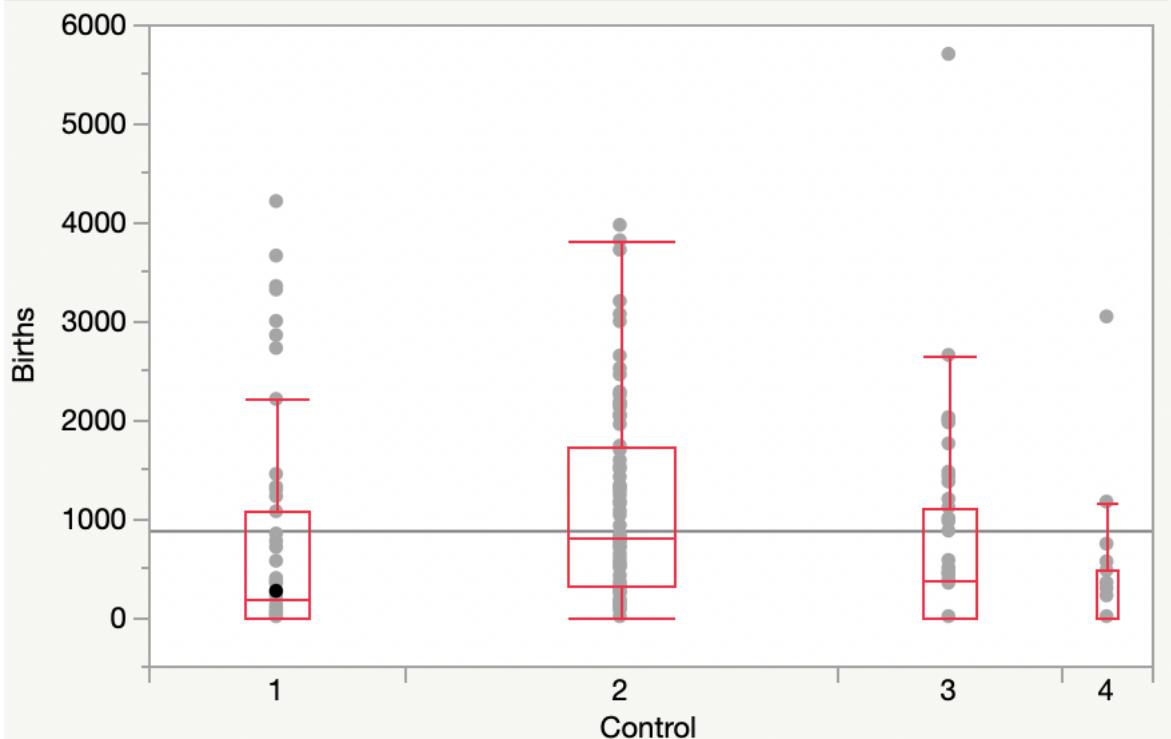
C1

Control	N	Births	
		Mean	Std Dev
1	51	765.11764705882	1142.5638126084
2	86	1136.5813953488	1033.304504577
3	45	694.377777777778	1037.6052255267
4	18	377.5	740.51057505258

Means and number of observations appear to be very different for each control. Standard deviations, however, are quite similar. This may suggest a normal distribution, in which means are approximately equal since there is a great variation for N.

C2

## Oneway Analysis of Births By Control



At least federal hospitals have mean number of births that differ from others. They will be lower than other hospitals; for whatever reasons there are less births in federal hospitals. Further analysis will be needed to determine if this is so, and why.

C3

Level	Count	Std Dev	MeanAbsDif	MeanAbsDif
			to Mean	to Median
1	51	1142.564	864.3230	736.5490
2	86	1033.305	829.2342	801.2558
3	45	1037.605	717.0815	640.2000
4	18	740.511	452.0556	377.5000

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.4053	3	196	0.7493
Brown-Forsythe	1.3449	3	196	0.2610
Levene	1.9783	3	196	0.1185
Bartlett	1.3556	3	.	0.2544

▼ **Welch's Test**

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
4.9757	3	70.014	0.0035*

Good result, high P values. Variances are statistically the same so we can continue with analysis. O'Brien test is reliable.

**▼ 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

R squared is very low. Of changes in # of births, model is only accounting for 5% of the differences in average births. This is between the R squared for Appendix A4 and B4, and follows the same insights as those two.

## ▼ 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			

0.01 is greater than .005, so at least one of means is statistically different from others. Thus, my hypothesis in part C2 was correct.

▼ **Connecting Letters Report**

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

Levels 2 and 4 are statistically different in terms of their means. Mean number of births in non-profit hospitals are significantly higher than those in a federal hospital.

All others are statistically the same.

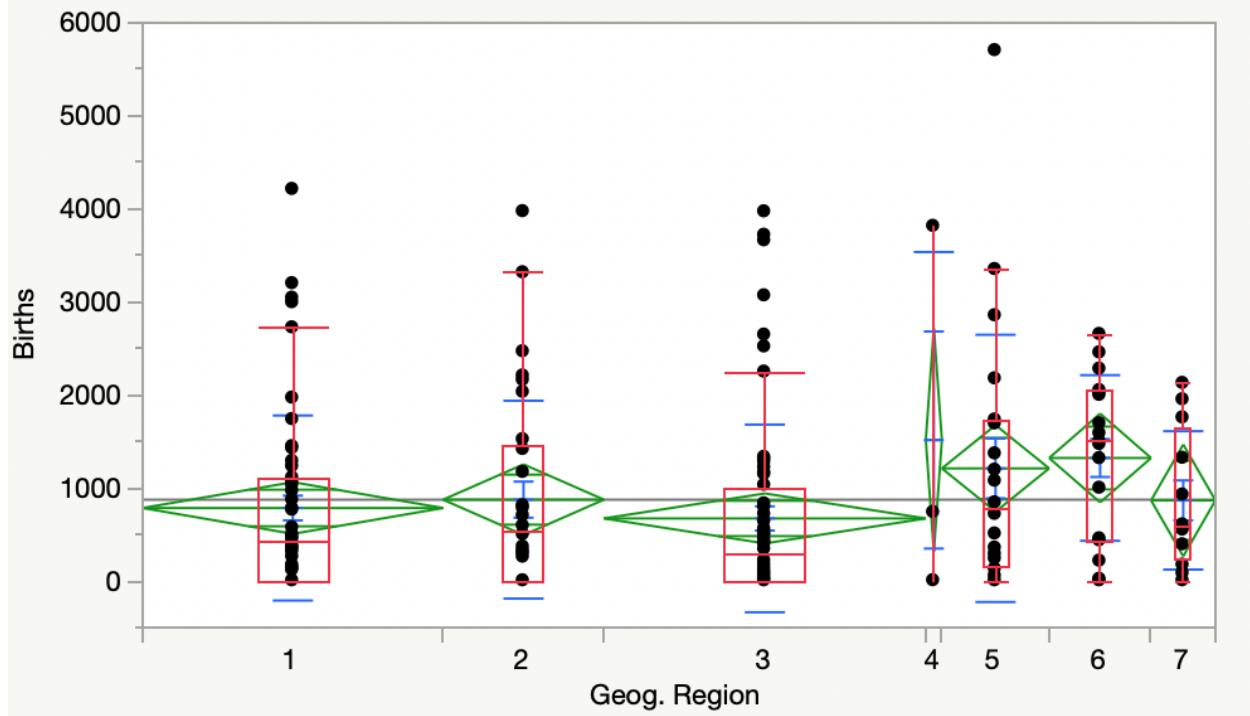
D1

Geog. Region	N	Births	
		Mean	Std Dev
1	56	784.75	994.53512667066
2	30	874.3	1063.5325376408
3	60	672.616666666667	1008.6267739898
4	3	1515.33333333333	2020.8439656078
5	20	1210.4	1436.3199467645
6	19	1322.3157894737	890.47035340454
7	12	866.58333333333	743.87932285398

N and means vary wildly for each region. Standard deviations are more similar, with some clear outliers (Region 4 is around 600 more than the next highest standard deviation. This suggests a normal distribution, and some means may be statistically different from others.

D2

## Oneway Analysis of Births By Geog. Region



It appears that the means are statistically the same, since the overall average intersects each box in the box plot. There are some clear outliers in 1, 2, 3 and especially 5.

D3

Test	F Ratio	DFNum	DDF Den	Prob > F
O'Brien[.5]	1.4276	6	193	0.2058
Brown-Forsythe	0.7084	6	193	0.6432
Levene	1.3247	6	193	0.2478
Bartlett	1.7455	6	.	0.1061

Warning: Small sample sizes. Use Caution.

### ▼ Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DDF Den	Prob > F
1.3121	6	22.777	0.2918

The probability of .2 within the O'Brien test is much greater than the .005 level of significance. Thus, variances are within an acceptable range.

D4	
▼ Summary of Fit	
Rsquare	0.045286
Adj Rsquare	0.015606
Root Mean Square Error	1055.333
Mean of Response	874.045
Observations (or Sum Wgts)	200

There is a very low R squared value that indicates the model explains only 4.5% of the variations from the mean. This is in line with C4, and still demonstrates a lackluster predictive model.

## ▼ Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Geog. Region	6	10196027	1699338	1.5258	0.1715
Error	193	214949499	1113728		
C. Total	199	225145527			

Since the P value is greater than the .005 level of significance, this model should be accepted as statistically significant with no significant variation of means. Thus, we should use the model since variances fall within acceptable range. Since there are no statistical differences in means, there is no need for further analysis.