Johannah Cushing

Statistics II: MANOVA and GLM

May 8, 2017

SUMMARY

In this analysis each respondent is classified based on the levels (groups) of the treatment (independent variable). MANOVA is often employed to control the overall Type I error rate while assessing differences on each dependent variable collectively and individually.

HBAT aims to strengthen its customer relationship strategy and as such is looking at differences due to its distribution channel system. (x5) x5 is composed of two channels: direct through HBAT's salesforce or indirect through a broker. Three purchase outcomes have been identified as focal issues in evaluating the impacts of the two distribution systems and will be treated as the analysis variables. These are: x19, satisfaction; x20; Likelihood of recommending HBAT; and x21, likelihood of future purchase. The purpose of the analysis is to identify if differences exist between these two systems across all or a subset of the purchase outcomes.

Table 5 "Descriptive Statistics of Purchase Outcome Measures" shows group profiles on each of the purchase outcomes (x19, x20, and x21 across the two groups- direct vs. indirect distribution systems). For all three purchase outcomes, the direct distribution channel has higher mean scores and seems to be a better selling channel. Also, x21 has higher mean scores than x20 which has higher mean scores than x19.

There is a concern for adequate sample sizes across the entire MANOVA analyses and thus 200 surveys are used instead of just the initial 100. With the larger dataset, 108 firms used the indirect broker system and 92 use the direct system. With these group sizes there will be adequate statistical power at the 80%.

With MANOVA the most important assumptions are the independence of observations, homoscedasticity across groups, and normality. Outliers can influence group means and should be considered.

A random sampling plan was employed to help ensure the independence of respondents. Levene's test assesses the univariate homogeneity of variance across the two groups. For all three variables it is found to be nonsignificant (i.e. greater than .05). The Box's M test for equality of the covariance matrices shows a nonsignificant value of .607, indicating that there is no significant difference between the two groups on the three dependent variables collectively. Therefore the assumption of homoscedasticity is met for each individual variable separately as well as the three variables collectively.

Bartlett's test for sphericity is used to determine whether the dependent measures are significantly correlated. It examines the correlations among all dependent variables and assesses

whether, collectively, there is significant intercorrelation. Here we find there is a significant degree of intercorrelation (significance=.000).

By looking at boxplots (distribution of x19, distribution of x20, distribution of x21) we can address outliers. There are only a few, if any extreme points across the groups. No observations had an extreme value on all three dependent measures, and no value was so extreme as to indicate that it should be excluded. Thus all 200 observations are retained.

In step 4 we assess whether the two groups have statistically significant differences for the three purchase outcomes variables, collectively as well as individually. We specify the maximum allowable Type I error rate, accepting that 5 of 100 (.05) times we might conclude that the type of distribution channel impacts the purchase outcome variables when in fact it does not. Then we can proceed to test for differences between the two groups, perform univariate tests on each purchase outcome, and then assess power levels.

Common multivariate tests include: Pillai's criterion, Wilk's lambda, Hotelling's T^2, and Roy's greatest characteristic root. Each test shows that the set of purchase outcomes have a highly significant difference (.000) between the two types of distribution channel. This is in alignment with what is seen in table 5 and the boxplots of the distributions of the individual variables. (This data is in table 7.7; not reproduced below as not able to do so in SAS).

Table 7 also shows results for the univariate tests for each individual purchase outcome. All of the individual tests are highly significant (.000), indicating that each variable follows the pattern of higher purchase outcomes for those served by the direct distribution system than served indirectly through a broker (7.688, 7.498, and 8.051 vs. 6.325, 6.488, and 7.336 for x19, x20, and x21, respectively). The power for these tests was all 1.0, indicating that the sample sizes and effect size were sufficient to ensure that significant differences would be detected if they existed beyond the differences due to sampling error.

The Scheffe test and Tukey test are post-hocs tests to find out which pairs of means are significant. They suggest that the differences may not be significant. For further analysis we may wish to switch from x5 to x1.

RESULTS

Table 5- Descriptive Statistics of Purchase Outcome Measures

*Here x5 0 is indirect through broker, and x5 1 is direct to customer

X5	N Obs	Variable	Mean	Std Dev	N
0	108	X19	6.325	1.033	108
		X20	6.488	0.986	108
		X21	7.336	0.880	108
1	92	X19	7.688	1.049	92
		X20	7.498	0.930	92
		X21	8.051	0.745	92

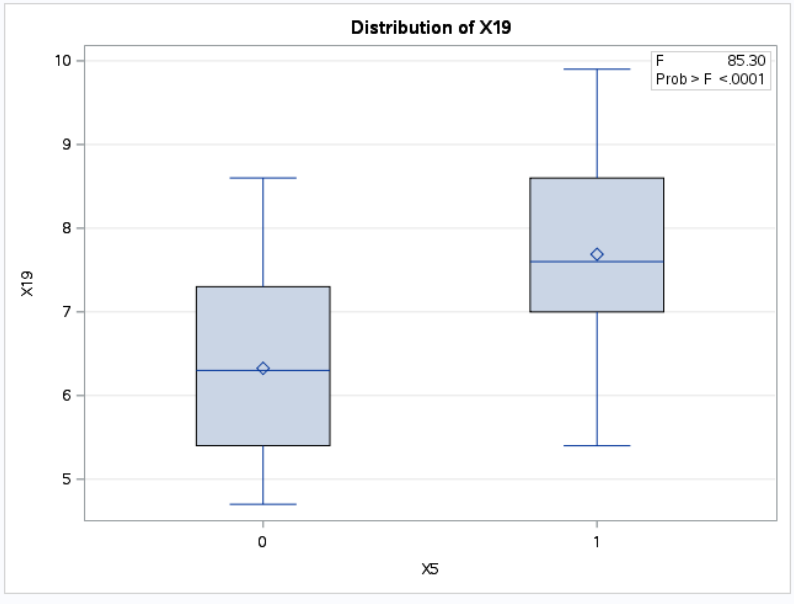
Table 6- Multivariate and Univariate Measures for Testing Homoscedasticity of x5

Dependent Variable: X19

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	92.2998522	92.2998522	85.30	<.0001
Error	198	214.2393478	1.0820169		
Corrected Total	199	306.5392000			

R-Square	Coeff Var	Root MSE	X19 Mean
0.301103	14.96261	1.040200	6.952000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
X5	1	92.29985217	92.29985217	85.30	<.0001

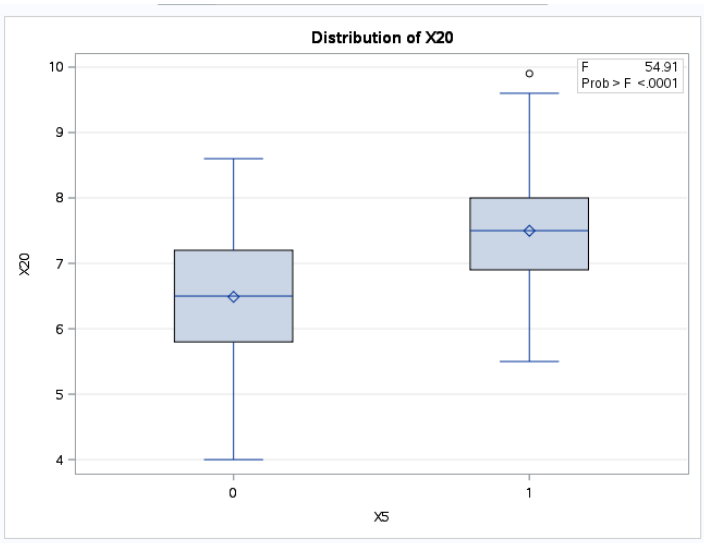


Dependent Variable: X20

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	50.6648329	50.6648329	54.91	<.0001
Error	198	182.6939171	0.9226966		
Corrected Total	199	233.3587500			

R-Square	Coeff Var	Root MSE	X20 Mean
0.217111	13.81619	0.960571	6.952500

Source	DF	Anova SS	Mean Square	F Value	Pr > F
X5	1	50.66483293	50.66483293	54.91	<.0001

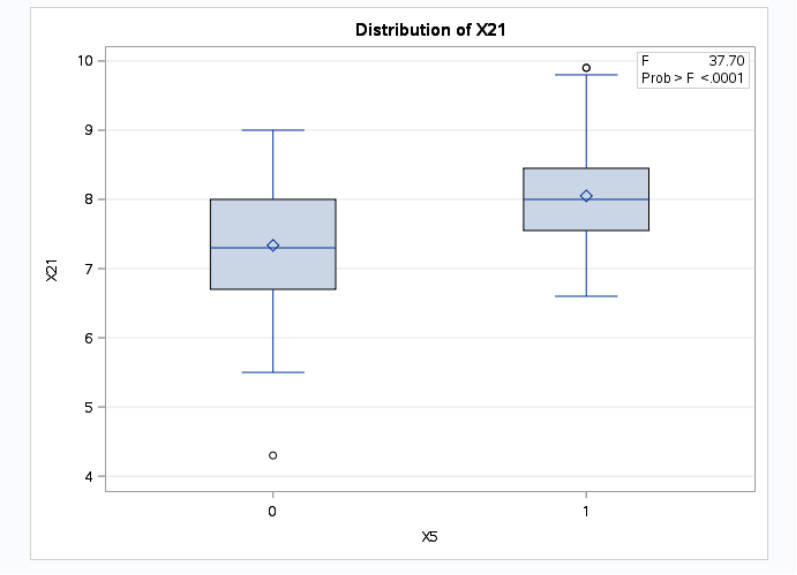


Dependent Variable: X21

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	25.3959420	25.3959420	37.70	<.0001
Error	198	133.3790580	0.6736316		
Corrected Total	199	158.7750000			

R-Square	Coeff Var	Root MSE	X21 Mean
0.159949	10.70777	0.820751	7.665000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
X5	1	25.39594203	25.39594203	37.70	<.0001


Distribution of X21
F 37.70
Prob > F <.0001
10
9
8
7
6
5
4
X21
0
1
X5

The ANOVA Procedure

Levene's Test for Homogeneity of X19 Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
X5	1	0.0482	0.0482	0.04	0.8390
Error	198	230.5	1.1640		

Bartlett's Test for Homogeneity of X19 Variance

Source	DF	Chi-Square	Pr > ChiSq
X5	1	0.0230	0.8794

Levene's Test for Homogeneity of X20 Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
X5	1	0.5739	0.5739	0.41	0.5217
Error	198	275.9	1.3932		

Bartlett's Test for Homogeneity of X20 Variance

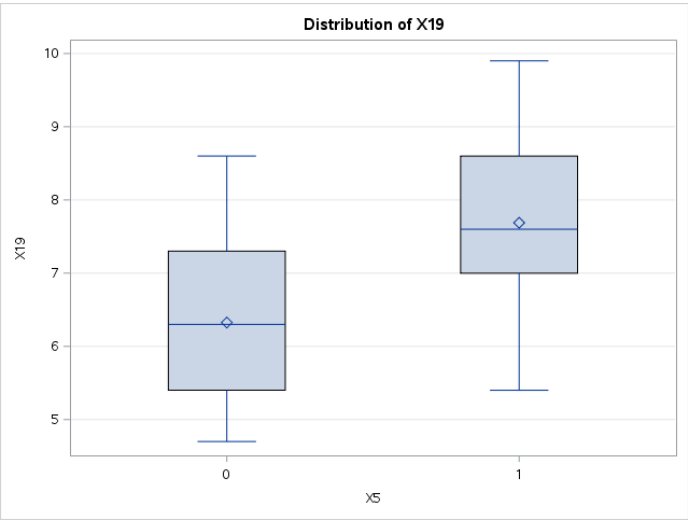
Source	DF	Chi-Square	Pr > ChiSq
X5	1	0.3321	0.5644

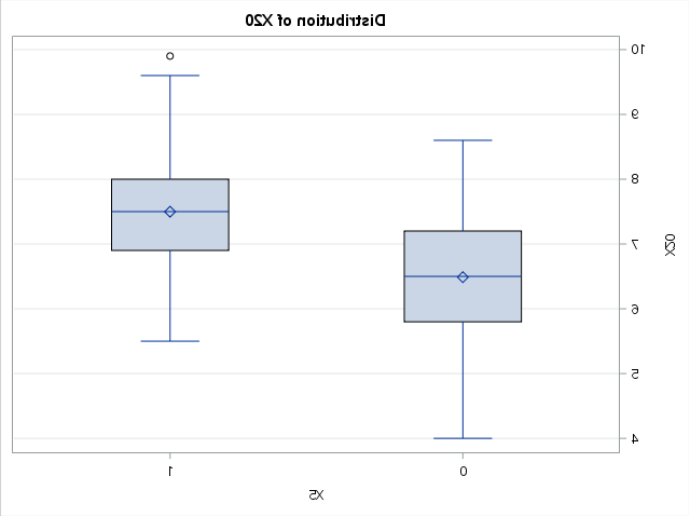
Levene's Test for Homogeneity of X21 Variance
ANOVA of Squared Deviations from Group Means

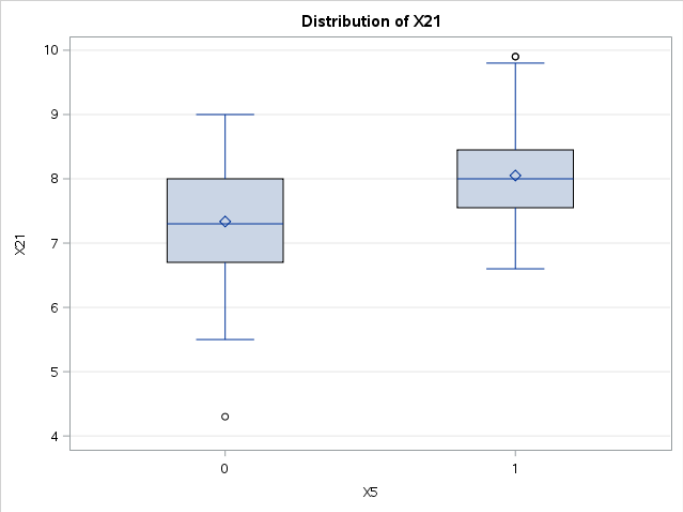
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
X5	1	2.3759	2.3759	2.42	0.1216
Error	198	194.6	0.9830		

Bartlett's Test for Homogeneity of X21 Variance

Source	DF	Chi-Square	Pr > ChiSq
X5	1	2.6886	0.1011







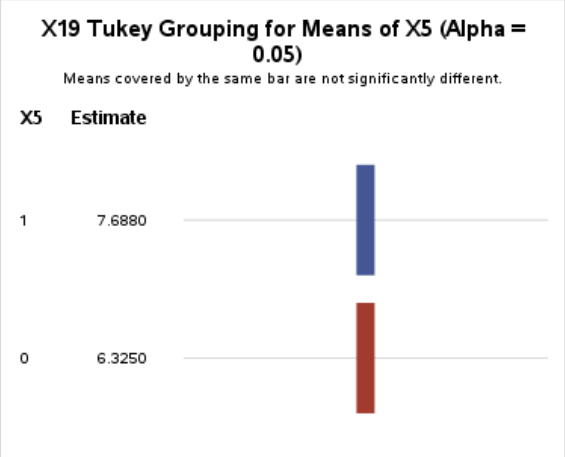
Level of X5	N	X19		X20		X21	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
0	108	6.32500000	1.03283703	6.48796296	0.98585487	7.33611111	0.88015062
1	92	7.68804348	1.04879233	7.49782609	0.92996257	8.05108696	0.74487178

Tukey's Studentized Range (HSD) Test for X19

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	198
Error Mean Square	1.082017
Critical Value of Studentized Range	2.78885
Minimum Significant Difference	0.291
Harmonic Mean of Cell Sizes	99.36

Note: Cell sizes are not equal.

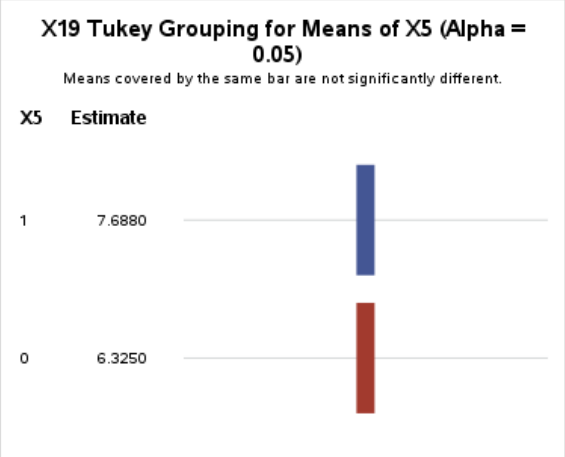


Tukey's Studentized Range (HSD) Test for X19

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	198
Error Mean Square	1.082017
Critical Value of Studentized Range	2.78885
Minimum Significant Difference	0.291
Harmonic Mean of Cell Sizes	99.36

Note: Cell sizes are not equal.



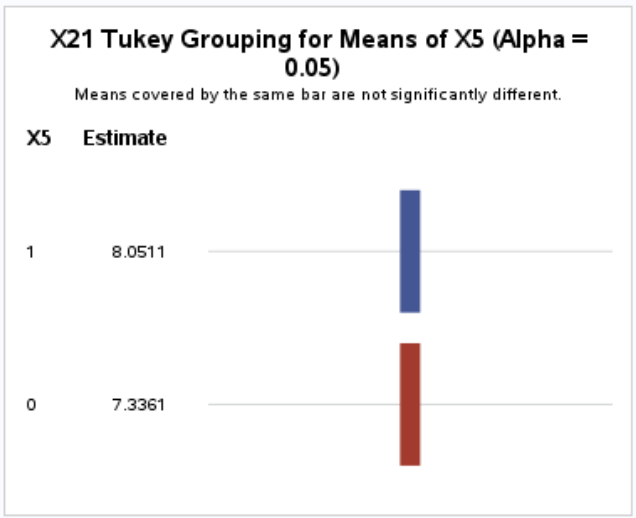
The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for X21

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	198
Error Mean Square	0.673632
Critical Value of Studentized Range	2.78885
Minimum Significant Difference	0.2296
Harmonic Mean of Cell Sizes	99.36

Note: Cell sizes are not equal.

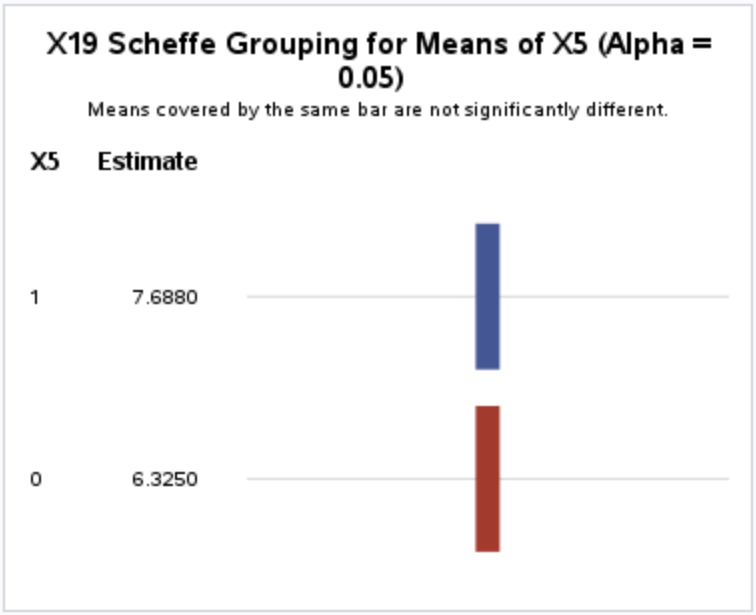


Scheffe's Test for X19

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	198
Error Mean Square	1.082017
Critical Value of F	3.88885
Minimum Significant Difference	0.291
Harmonic Mean of Cell Sizes	99.36

Note: Cell sizes are not equal.



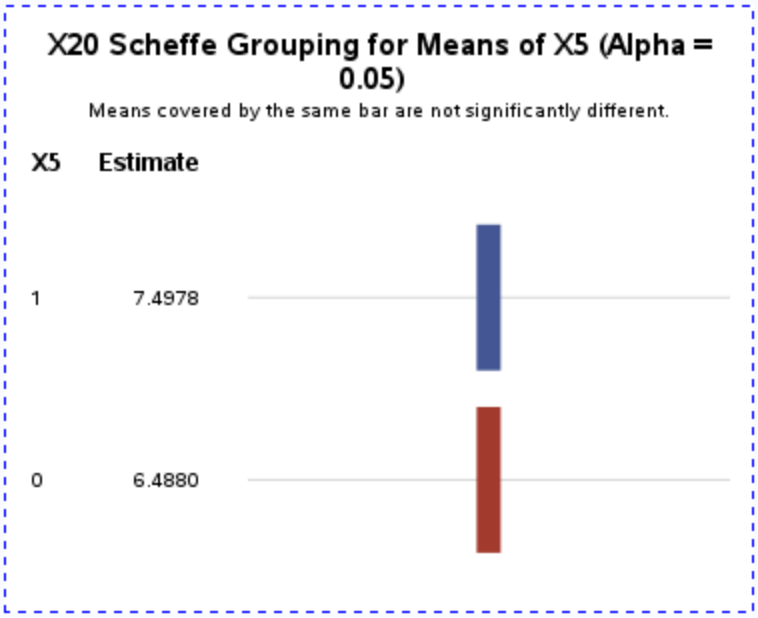
The ANOVA Procedure

Scheffe's Test for X20

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	198
Error Mean Square	0.922697
Critical Value of F	3.88885
Minimum Significant Difference	0.2688
Harmonic Mean of Cell Sizes	99.36

Note: Cell sizes are not equal.



Scheffe's Test for X21

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	198
Error Mean Square	0.673632
Critical Value of F	3.88885
Minimum Significant Difference	0.2296
Harmonic Mean of Cell Sizes	99.36

Note: Cell sizes are not equal.