WEEK 7 ASSIGNMENT

Concepts of Statistics 2 – DATA-51200 | Spring 2 2020

Christina Morgenstern

---

1. On your own words, summarize (in less than one page) Stage 4: Estimation of the MANOVA Model and Assessing Overall Fit

After the research objective has been stated, the variables have been selected and the assumptions of MANOVA analysis have been checked, the model is being estimated by assessing differences in group means. To test for statistical significance of the model and depending on the situation, several tests can be used: Roy´s greatest characteristic root (gcr), Wilks´lambda, Pillai´s criterion and Hotelling´s T^2 are widely applied measures. Dependent on the individual research design and taking into account the assumptions underlying MANOVA one of these statistical measures is applied for significance testing. The alpha level, effect size and sample size all affect the significance level of the above-mentioned tests. By tuning these parameters to some degree, the level of power in the statistical tests can be adjusted. In most cases all four tests will deliver similar results but there can be variation in some cases. In general, maintaining power at around 0.8 for a selected alpha level is acceptable for MANOVA.
Furthermore, in MANOVA analysis the effects of multicollinearity of the dependent variables i.e. the strength and direction of the correlations on the power must also be taken into account. Mediation and moderation are additional relationships that can be considered in MANOVA with these relationships providing insights into "why" the treatment impacted outcome and "when" the treatment effect occurred, respectively. Also, a combination of the effects might be assessed in an extended framework. To test for these effects, several tests such as the Sobel test for estimating mediation effects are available and so are software tools.

References:
[1] Multivariate Data Analysis by Joseph F. Hair Jr, William C. Black, Barry J. Babin and Rolphe E. Anderson, Pearson, 8th edition, 2019

2. Since the GUI version of SAS UE does NOT support MANOVA, this part will focus on ANOVA. For the data set associated with this homework (HBAT200.sav), reproduce Tables 5, 6 and 7 and figure 6 from MANOVA and GLM chapter. You may not be able to reproduce all (similar) entries. (You may use any software and programming language you feel comfortable dealing with. Make sure to include your codes, diagrams and results)

To create the desired tables and figures, I used SAS Studio.

To answer the research question if there are any differences between HBAT´s two distribution channels (X5), direct through HBAT´s salesforce or indirect through a broker, on the purchase outcomes satisfaction (X19), likelihood of recommending HBAT (X20) and likelihood of future purchase (X21), a MANOVA analysis should be performed.

To create table 5, the summary statistics task in SAS studio was used, with analysis variables X19, X20 and X21 and the classification variable X5. In the options section, I choose to display mean, standard deviation and the number of observations. Please note, that the total of either values couldn´t be selected for display. I also choose to display the boxplots for the variables X19, X20 and X21 for the two classes of variable X5.

The following SAS code was generated:

```
ods noproctitle;
ods graphics / imagemap=on;

proc means data=WORK.ANOVA chartype mean std n vardef=df;
      var x19 x20 x21;
      class x5;
run;

proc sort data=WORK.ANOVA out=WORK.TempSorted2236;
      by x5;
run;

proc boxplot data=WORK.TempSorted2236;
      plot (x19 x20 x21)*x5 / boxstyle=schematic;
run;

proc datasets library=WORK noprint;
      delete TempSorted2236;
      run;
```

Table 1 displays the summaries of the group profiles on each purchase outcome. Within the group “direct to customer”, the mean values are higher for each of the variables. To check, whether these differences are significant, a MANOVA analysis should be applied.

X5 - Distribution System	N Obs	Variable	Label	Mean	Std Dev	N
Indirect through broker	108	x19	X19 - Satisfaction	6.3250000	1.0328370	108
		x20	X20 - Likely to Recommend	6.4879630	0.9858549	108
		x21	X21 - Likely to Purchase	7.3361111	0.8801506	108
Direct to customer	92	x19	X19 - Satisfaction	7.6880435	1.0487923	92
		x20	X20 - Likely to Recommend	7.4978261	0.9299626	92
		x21	X21 - Likely to Purchase	8.0510870	0.7448718	92

Table 1. Descriptive statistics of purchase outcome measures (X19, X20 and X21) for groups of X5 (distribution system).

Boxplots for the three purchase outcome variables X19, X20 and X21 were used to graphically visualize the distribution of the three variables across the two distribution groups (see Figure 1). This also helps to detect outliers as denoted by small circles outside the boxplots of variables X20 and X21.

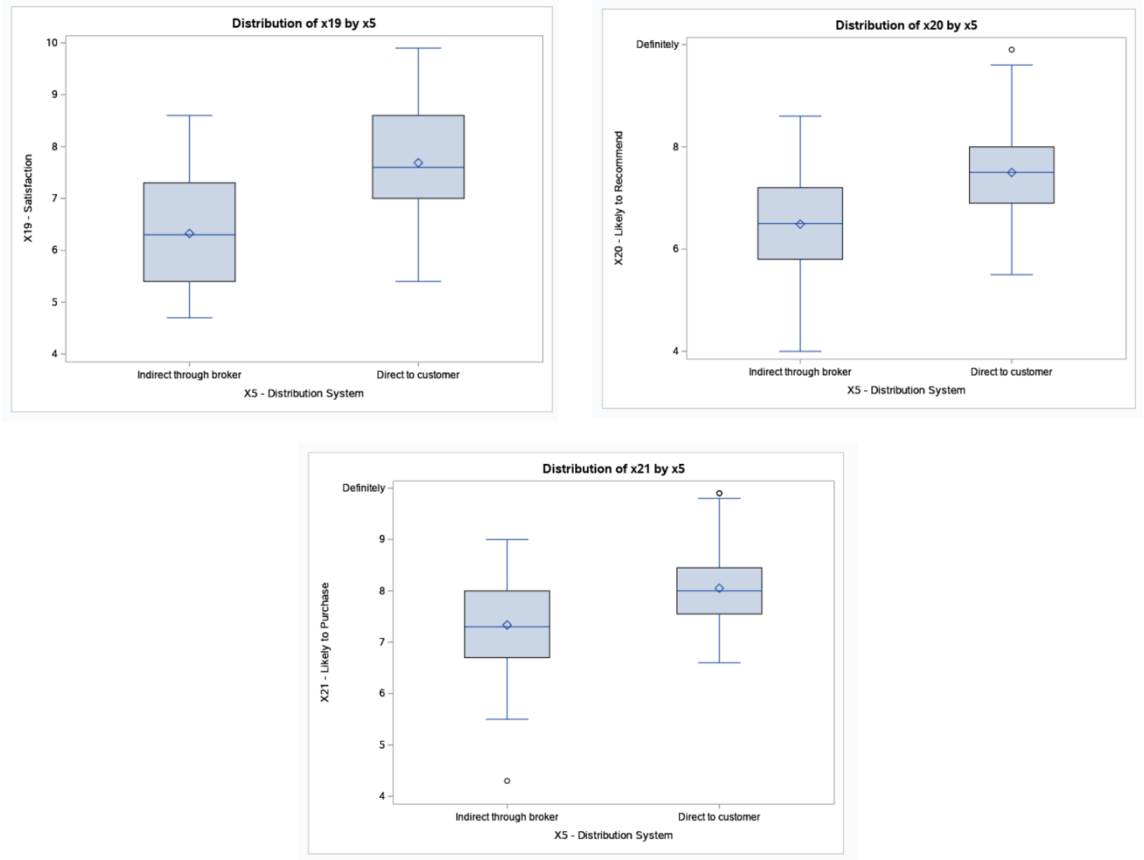

Figure 1. Boxplots of purchase outcome measures (X19, X20 and X21) for groups of X5 (distribution system).

Independence of observations, homoscedasticity across groups and normality are the most important assumptions in MANOVA.

To create table 6, univariate and multivariate tests should be performed to test for homoscedasticity of X5. One-way ANOVA was used to perform Levene´s and Bartlett´s tests.

The following SAS code was generated when using X19 as dependent variable and X5 as class variable

```
Title;
ods noproctitle;
ods graphics / imagemap=on;

proc glm data=WORK.ANOVA;
       class x5;
       model x19=x5;
       means x5 / hovtest=bartlett welch plots=none;
       lsmeans x5 / adjust=bon pdiff alpha=.10;
       run;
quit;
```

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		

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		

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		

Table 2. Results of Levene´s test for outcome measures (X19, X20 and X21) for groups of X5 (distribution system).

The results from the Levene´s tests show that all three variables are nonsignificant i.e. > 0.05 (see values in table 2 Pr > F cells).

Bartlett's Test for Homogeneity of x19 Variance			
Source	DF	Chi-Square	Pr > ChiSq
x5	1	0.0230	0.8794

Bartlett's Test for Homogeneity of x21 Variance			
Source	DF	Chi-Square	Pr > ChiSq
x5	1	2.6886	0.1011

Bartlett's Test for Homogeneity of x20 Variance			
Source	DF	Chi-Square	Pr > ChiSq
x5	1	0.3321	0.5644

Table 3. Results of Bartlett´s test for outcome measures (X19, X20 and X21) for groups of X5 (distribution system).

In the Bartlett´s test the dependent variables were assessed separately and not as in the given table 6 for correlation among all dependent variables.

For the Box´s Test of Equality of Covariance Matrices (top part of table 6), the following SAS code was used:

```
proc discrim data=WORK.ANOVA method=normal pool=test wcov;
      class x5;
      var x19 x20 x21;
run;
```

Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
4.520531	6	0.6066

Table 4. Multivariate test of homoscedasticity for dependent variables X19, X20 and X21 with variable X5.

The result of the Box´s M test for equality of the covariance matrices shows with 0.607 a nonsignificant value, suggesting no significant difference between the two groups on the three independent variables.

To determine group differences in purchase outcome measures X19, X20 and X21 across groups of X5, different multivariate tests were performed using SAS Studio Canonical Correlation task. The results from this analysis are creating the upper part of table 7.

The following SAS code was generated:

```
ods noproctitle;
ods graphics / imagemap=on;

proc cancorr data=WORK.ANOVA ncan=1 redundancy;
                                    /*** The VAR statement defines Variable set 1 ***/
                                    var x5;

                                    /*** The WITH statement defines Variable set 2 ***/
                                    with x19 x20 x21;
run;
```

Multivariate Statistics and Exact F Statistics					
S=1 M=0.5 N=97					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.69314489	28.92	3	196	<.0001
Pillai's Trace	0.30685511	28.92	3	196	<.0001
Hotelling-Lawley Trace	0.44269981	28.92	3	196	<.0001
Roy's Greatest Root	0.44269981	28.92	3	196	<.0001

Table 5. Multivariate tests for group differences in purchase outcome measures (X19, X20 and X21) across groups of X5 (distribution system).

For the univariate tests, one-way ANOVA was performed X5 and each of the dependent variables X19, X20 and X21 separately:

The following SAS code was generated using the dependent variable X19 and X5 as class variable.

```
Title;
ods noproctitle;
ods graphics / imagemap=on;

proc glm data=WORK.ANOVA plots=none;
                                    class x5;
                                    model x19=x5;
                                    means x5 / hovtest=levene plots=none;
                                    lsmeans   x5   /   adjust=tukey   pdiff   alpha=.05
plots=none;
                                    run;
quit;
```

To test for the other two variables, X20 and X21, the code was amended, and the model run again with each of the two variables separately.

Dependent Variable: x19 X19 - Satisfaction					
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			

Table 6. One-way ANOVA using X5 as the independent variable and X19 as the dependent variable.

Dependent Variable: x20 X20 - Likely to Recommend

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			

Table 7. One-way ANOVA using X5 as the independent variable and X20 as the dependent variable.

Dependent Variable: x21 X21 - Likely to Purchase

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			

Table 8. One-way ANOVA using X5 as the independent variable and X21 as the dependent variable.