Quantitative Research
November 2019

Quantitative Research Methodology Design Based on SPSS

Eve Thullen, Claremont Graduate University

Case 1: Multi-Regression

Research Questions and Hypotheses

A researcher tried to help a private university in Southern California to find out their undergraduate student's financial abilities to pay the school and predict the amount of money the new students be able to pay in future.

Hypotheses: suppose student's financial abilities can be evaluated through their backgrounds such as their family income amount, number of children, parent's education level, etc.

1. Population and Samples

The population is undergraduate students at this private university in Southern California. Thus, the samples will be the students randomly drawn from all the departments of the school. We will collect N = 500 samples.

2. Data Collection

Data collection will come from enrollment records and survey, which will be provided by the universities. The survey questions will be the supplement with the student enrollment information. The following information about the student will be collected: gender, age, states, zip code, ethnics, number of children, totally family income (per year), average income (year per person), education level of farther, education level of mother, and the budget/available money for school (per year).

We will collect 500 records.

IRB approval will be requested to data collection. Identifying information such as student name, ID, address, etc. will be removed prior to receipt of data.

3. Data Analysis

Data will be entered into a statics program, such as SPSS, R, etc. In this case, we will use SPSS.

In SPSS, to analyze the data, we will use regression-> linear,

- Dependent Variables will be: the budget/available money for school;
  Testable IVs will be all the Independent variables.
- a significance level of 0.05 will be set.

- Step-Wise regression method will be set. SPSS will automatically throw out those independent variables whose sig t greater than 0.05 or no contribute to R.
- Collinearity Diagnostic will be checked.

4. Results Analysis

1) Entered Variables: explained the variables that can be used in this model to predict the dependent variable best.
2) Model Summary: R and R Square, Adjusted R Square, answer the question and How successful were the independent variables at predicting the dependent variable.
   - R is called "multiple R" and is the multiple correlation coefficient; It varies between 0 and +1. It is the correlation between the dependent variable and all the independent variables
   - R^2 is the multiple coefficient of determination, also ranges from 0 to +1. It tells what % of dependent variable (DV) can be attributed to the independent variables (IVs) and How successful were the independent variables at predicting the dependent variable. For example, if the R = 0.8, R Square is 0.64, means 64% of the total variation in DV can be explained and predicted by entered IV in this model.
   - Adjust R Square, which adjusted the sample R Square to population R Square.
3) ANOVA: SS_M, SS_R, SS_T and F, Sig., answer the question How confident were the independent variables at predicting the dependent variable.

From this chart, we can get the numbers about SSM, SSR, SST to tell how reliable this model is. F test determines the significance between the dependent variable and all the independent variables. The F can be calculated from SSM and SSR. If sig. ≤ 0.05, then the independent variables are collectively significant predictors of the dependent variable, means the model is reliable, the confidential level is above 95%.

4) Coefficients: Unstandardized B, Standardized Coefficients Beta, t and sig. , answer the question how good the particular independent variable is.
   - Unstandardized B tell us the model Regression Equations
     Multivariate unstandardized equation (1 dependent, more than 1 independent, and a constant)
     $Y = B_1x_1 + B_2x_2 + B_3x_3 + \ldots + \text{Constant}$
   - Standardized Coefficients Beta, tell us how good a predictor the particular independent variable is. Ranges from +1 to –1; Larger the absolute value, the stronger it is as a predictor; The sign only tells the direction (direct or inverse)
   - Sig. t value determined if variable is significant, means the particular independent variable is significant or not. If sig t ≤ 0.05, then the variable is a significant predictor.
   - If we checked Collinearity Diagnostic, the SPSS will display the Collinearity Tolerance and VIF value, which will get rid of the IVs which have high correlations with each other. For example, if two IVs correlation r = 0.8, the Collinearity tolerance will be 1- R Square = 1- 0.64 = 0.36. Higher Correlation r between two IVs will cause low Collinearity Tolerance.

Excluded Variables: the variables that not entered in the model, which either has bad sig t > 0.05 or reduce the R Square.

Case 2: Two-Way ANOVA

Question: A research wants to examine the influence of both gender and smoking status (smoker vs Non-smoker) on depression.

Research Design:

- DV is depression as measured by levels numbers.
- Two independent Variables: Gender (measured by male and female), and Smoker Status (measured by Smoke and Non Smoker)
- 120 students were randomly assigned to one of the four groups based on the gender and smoker status: female and smoker, female and Non-Smoker; make smoker, male and Non-Smoker
- Resulted in a total of 30 participants in each of four groups.
- At the end of each group take the survey, the scores indicated their depression level.
- Score on the depression lever were analyzed via factorial two-way ANOVA.

Result Analysis:

- Check the *Tests of Between-Subjects Effects Table*
- Find the interaction effect Gender*Smoking Status Level first.

Case 1: If Interaction effect (Gender*Smoking Status) is statistically significant (sig.<0.05). (When you have a statistically significant interaction, reporting the main effects can be misleading. Therefore, you will need to report **the simple main effects**.) 2) For simple main effects, comparing female*smoker, female*Non-Smoker, male*smoker, male*Non-Smoker, find out which one is significance with others, we do the Post-Hoc Tests, Scheffe Test, gets *Multiple Comparison table*, get the ones below 0.05 is significant different with others.

For example, female*Smoker is significant with male*Non-Smoker, (sig.<0.05),

> we can conclusion that, 1) an interaction occurs, Gender*Smoker Status effects the depression Level. 2)Female*Smoker is significant different with Male*No-Smoker on depression level. A recommendation that based on this study would be female should avoid Smoker to reduce depression level.

Case 2: If you do not have a statistically significant interaction, Then, we will check the main effects significant level. For example, if main affect Gender is significant (sig. <0.05). We might interpret the Tukey post hoc test results for the different levels of gender, which can be found in the Multiple Comparisons

> We can conclusion that there is NO interaction occurs Gender*Smoker Status effects the depression Level. The Gender has a significant effect on depression level.

Notes: if the main effects group more than two groups, then we will need do run post-hoc testing to find out which group has significant effects on DV.

<u>Case 1: check interaction Effects</u>, significance <0.5 then need report simple main effects; the Post-Hoc Tests, Scheffe Test

Tests of Between-Subjects Effects

Dependent Variable: charges

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.220E+11[a]	3	4.067E+10	732.582	.000
Intercept	3.463E+11	1	3.463E+11	6237.077	.000
sex	151971572	1	151971572	2.737	.098
smoker	1.172E+11	1	1.172E+11	2110.798	.000
sex * smoker	492339741	1	492339741	8.868	.003
Error	7.406E+10	1334	55517658.8		
Total	4.317E+11	1338			
Corrected Total	1.961E+11	1337			

a. R Squared = .622 (Adjusted R Squared = .621)

Multiple Comparisons

Dependent Variable: SCORE final exam score

Scheffe

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 homework - 8 wks	2 homework - 16 wks	3.93	2.38	.452	-3.22	11.08
	3 no homework - 8 wks	15.19*	2.38	.000	8.05	22.34
	4 no homework - 16 wks	1.31	2.38	.959	-5.84	8.46
2 homework - 16 wks	1 homework - 8 wks	-3.93	2.38	.452	-11.08	3.22
	3 no homework - 8 wks	11.27*	2.38	.001	4.12	18.42
	4 no homework - 16 wks	-2.62	2.38	.752	-9.76	4.53
3 no homework - 8 wks	1 homework - 8 wks	-15.19*	2.38	.000	-22.34	-8.05
	2 homework - 16 wks	-11.27*	2.38	.001	-18.42	-4.12
	4 no homework - 16 wks	-13.88*	2.38	.000	-21.03	-6.73
4 no homework - 16 wks	1 homework - 8 wks	-1.31	2.38	.959	-8.46	5.84
	2 homework - 16 wks	2.62	2.38	.752	-4.53	9.76
	3 no homework - 8 wks	13.88*	2.38	.000	6.73	21.03

*. The mean difference is significant at the .05 level.

Case 2: If no interaction effects, check the main effects; if the main effects more than three group, than run one-way ANOVA and post-hoc test

Tests of Between-Subjects Effects

Dependent Variable: charges

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.254E+9[a]	7	322000313	2.210	.031
Intercept	2.329E+11	1	2.329E+11	1598.508	.000
sex	584157098	1	584157098	4.009	.045
region_N	1.253E+9	3	417791322	2.867	.035
sex * region_N	340173346	3	113391115	.778	.506
Error	1.938E+11	1330	145729488		
Total	4.317E+11	1338			
Corrected Total	1.961E+11	1337			

a. R Squared = .011 (Adjusted R Squared = .006)

Multiple Comparisons

Dependent Variable: charges

Tukey HSD

(I) region_N	(J) region_N	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	988.809142	947.724009	.724	-1449.0045	3426.62276
	3	-1329.0269	922.029342	.474	-3700.7466	1042.69276
	4	1059.44714	947.724009	.679	-1378.3665	3497.26076
2	1	-988.80914	947.724009	.724	-3426.6228	1449.00448
	3	-2317.8361	921.278548	.058	-4687.6245	51.9523662
	4	70.6379967	946.993585	1.000	-2365.2968	2506.57276
3	1	1329.02692	922.029342	.474	-1042.6928	3700.74661
	2	2317.83606	921.278548	.058	-51.952366	4687.62449
	4	2388.4741*	921.278548	.047	18.6856305	4758.26249
4	1	-1059.4471	947.724009	.679	-3497.2608	1378.36648
	2	-70.637997	946.993585	1.000	-2506.5728	2365.29677
	3	-2388.474*	921.278548	.047	-4758.2625	-18.685631

Based on observed means.
The error term is Mean Square(Error) = 145729488.251.

*. The mean difference is significant at the

Case 3: One Way MANOVA

Question: Online travel booking tool: automated online-portal, Phone Hotline, Call-back

User acceptance – ease of use, perceived usefulness, effort to use

Research Design:

- DV is User Acceptance, measured by ease of use, perceived usefulness, effort to use (3 DVs)
- Independent Variables is Online Travel Booking Tools, which divided to 3 groups: Online-portal, Phone hotline, Call-back
- 30 people random assigned to three groups based on Online Travel Booking Tools.
- Resulted in a total 10 participants in each Booking group
- One Way MANOVA used

Result Analysis:

- Check the *Multivariate Tests table*
- Look at the second Effect, labelled "Booking Tools "and the Wilks' Lambda row:

Case 1: If Wilks' Lambda value has sig. value <0.5, then

> We can conclude that there are statistically significant differences in User Acceptance Variables

2) Then, to find out which DV is significant, we check the *Test of Between-subject Effects table.* For example, if Ease of Use significance level <0.5,
>We can conclude that online Travel Booking Tools has a statistically significant effect on the User Acceptance of Ease of Use
3) Next, to find out which Online Booking Tools is significant on Ease of Use. We run Post-Hoc test to get Multiple Comparison Table to examine. For example, automated online-portal is significant different with Call back (sig. <0.5),
>We can conclude that there is significant difference in User Acceptance of Ease of Use to use Online-portal and Call Back Tools.

Case 2: If Wilks' Lambda value has no sig. value <0.5, then these three online booking tools has no Significant impact on user acceptance.

Multivariate Tests table, Check Wilk's Lambda: significance level <0.5

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.965	18227.436[b]	2.000	1333.000	.000
	Wilks' Lambda	.035	18227.436[b]	2.000	1333.000	.000
	Hotelling's Trace	27.348	18227.436[b]	2.000	1333.000	.000
	Roy's Largest Root	27.348	18227.436[b]	2.000	1333.000	.000
region_N	Pillai's Trace	.084	19.570	6.000	2668.000	.000
	Wilks' Lambda	.916	19.949[b]	6.000	2666.000	.000
	Hotelling's Trace	.092	20.329	6.000	2664.000	.000
	Roy's Largest Root	.089	39.523[c]	3.000	1334.000	.000

a. Design: Intercept + region_N
b. Exact statistic
c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	bmi	4055.881[a]	3	1351.960	39.495	.000
	charges	1.301E+9[b]	3	433586560	2.970	.031
Intercept	bmi	1248266.99	1	1248266.99	36465.849	.000
	charges	2.335E+11	1	2.335E+11	1599.233	.000
region_N	bmi	4055.881	3	1351.960	39.495	.000
	charges	1.301E+9	3	433586560	2.970	.031
Error	bmi	45664.320	1334	34.231		
	charges	1.948E+11	1334	146007093		
Total	bmi	1307766.55	1338			
	charges	4.317E+11	1338			
Corrected Total	bmi	49720.200	1337			
	charges	1.961E+11	1337			

a. R Squared = .082 (Adjusted R Squared = .080)
b. R Squared = .007 (Adjusted R Squared = .004)

Test of Between-subject Effects table Check which DV is significance

Post-Hoc test to get Multiple Comparison Table
Check which group of IV is significant :

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) region_N	(J) region_N	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
bmi	1	2	-.02628	.459324	1.000	-1.20779	1.15522
		3	-4.18249*	.446870	.000	-5.33196	-3.03301
		4	-1.42311*	.459324	.011	-2.60462	-.24161
	2	1	.02628	.459324	1.000	-1.15522	1.20779
		3	-4.15620*	.446507	.000	-5.30474	-3.00767
		4	-1.39683*	.458970	.013	-2.57743	-.21624
	3	1	4.18249*	.446870	.000	3.03301	5.33196
		2	4.15620*	.446507	.000	3.00767	5.30474
		4	2.75937*	.446507	.000	1.61084	3.90791
	4	1	1.42311*	.459324	.011	.24161	2.60462
		2	1.39683*	.458970	.013	.21624	2.57743
		3	-2.75937*	.446507	.000	-3.90791	-1.61084
charges	1	2	988.809142	948.626254	.725	-1451.3161	3428.93434
		3	-1329.0269	922.907126	.475	-3702.9955	1044.94167
		4	1059.44714	948.626254	.679	-1380.6781	3499.57233
	2	1	-988.80914	948.626254	.725	-3428.9343	1451.31605
		3	-2317.8361	922.155616	.058	-4689.8716	54.1994379
		4	70.6379967	947.895135	1.000	-2367.6066	2508.88255
	3	1	1329.02692	922.907126	.475	-1044.9417	3702.99551
		2	2317.83606	922.155616	.058	-54.199438	4689.87157
		4	2388.4741*	922.155616	.048	16.4385588	4760.50956
	4	1	-1059.4471	948.626254	.679	-3499.5723	1380.67805
		2	-70.637997	947.895135	1.000	-2508.8826	2367.60656
		3	-2388.474*	922.155616	.048	-4760.5096	-16.438559

Based on observed means.
The error term is Mean Square(Error) = 146007092.869.
*. The mean difference is significant at the

Case 4: One Way MANCOVA

Question: Do different income classes report a significantly different satisfaction with life when looking where they live (urban, suburban, rural) when controlling for e.g. marital status, job satisfaction, social support system

Research Design:

- DV is Life Satisfaction measured by three live areas: Urban, Suburban, and Rural

- Independent Variables are different Income Classes (measured by Low, Media, High);
- Covariate factors are: marital status, job satisfaction, and social support system
- 120 people randomly assigned to three groups based on income levels: low income group; media income group; and high income group
- Resulted in a total 40 participants in each of income group.
- One Way MANCOVA used

Results Analysis:

- Check the *Multivariate Tests table*
- Only Look at the Effect, labelled "Income Classes "and the Wilks' Lambda row:

Case 1: If Wilks' Lambda value has sig. value <0.5, then

>We can conclude that there are statistically significant differences in Life Satisfaction Variables between different Income Levels, after controlling for marital status, job satisfaction, and social support system.

2) Then, to find out which DV is significant, we check the *Test of Between-subject Effects table.* For example, if Urban significance level <0.5,

>We can conclude that different Income Levels has a statistically significant effect on Life Satisfaction in Urban area.

3) Next, to find out which Income class has significant effect in urban area, we do the Estimated Marginal Means test.

a. From *Estimate Table*, can find the adjusted mean after controlling Covariate factors.

b. Also, we get *Pairwise Comparison* table, (from *Multiple Comparison with Bonferroni* based on estimated marginal means), we find which Income level is significant with others, for example, the High Income Level is significant with Low Income. (sig.<0.5).

> We can conclude that there is significance difference in Life Satisfaction in Income Classes between High Income and Low Income that living in Urban area.

Case 2: If Wilks' Lambda value has no sig. value <0.5, then the Income Classes has no Significant impact on Life Satisfaction on the controlling of marital status, job satisfaction, and social support system, etc.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.722	1727.300[b]	2.000	1332.000	.000
	Wilks' Lambda	.278	1727.300[b]	2.000	1332.000	.000
	Hotelling's Trace	2.594	1727.300[b]	2.000	1332.000	.000
	Roy's Largest Root	2.594	1727.300[b]	2.000	1332.000	.000
sex_N	Pillai's Trace	.004	2.852[b]	2.000	1332.000	.058
	Wilks' Lambda	.996	2.852[b]	2.000	1332.000	.058
	Hotelling's Trace	.004	2.852[b]	2.000	1332.000	.058
	Roy's Largest Root	.004	2.852[b]	2.000	1332.000	.058
region_N	Pillai's Trace	.084	19.489	6.000	2666.000	.000
	Wilks' Lambda	.916	19.866[b]	6.000	2664.000	.000
	Hotelling's Trace	.091	20.243	6.000	2662.000	.000
	Roy's Largest Root	.089	39.363[c]	3.000	1333.000	.000

a. Design: Intercept + sex_N + region_N

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	bmi	4141.888[a]	4	1035.472	30.284	.000
	charges	1.914E+9[b]	4	478457212	3.285	.011
Intercept	bmi	118206.437	1	118206.437	3457.109	.000
	charges	1.664E+10	1	1.664E+10	114.267	.000
sex_N	bmi	86.007	1	86.007	2.515	.113
	charges	613069167	1	613069167	4.209	.040
region_N	bmi	4034.975	3	1344.992	39.336	.000
	charges	1.270E+9	3	423412889	2.907	.034
Error	bmi	45578.313	1333	34.192		
	charges	1.942E+11	1333	145656709		
Total	bmi	1307766.55	1338			
	charges	4.317E+11	1338			
Corrected Total	bmi	49720.200	1337			
	charges	1.961E+11	1337			

a. R Squared = .083 (Adjusted R Squared = .081)

b. R Squared = .010 (Adjusted R Squared = .007)

Estimated Margin Means:

Estimates

Dependent Variable	region_N	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
bmi	1	29.175[a]	.325	28.537	29.812
	2	29.205[a]	.324	28.568	29.841
	3	33.349[a]	.307	32.748	33.950
	4	30.598[a]	.324	29.962	31.235
charges	1	13409.289[a]	670.492	12093.955	14724.624
	2	12430.909[a]	669.490	11117.541	13744.278
	3	14716.455[a]	632.646	13475.365	15957.546
	4	12351.938[a]	669.463	11038.623	13665.254

a. Covariates appearing in the model are evaluated at the following values: sex_N = 1.51.

Pairwise Comparisons

Dependent Variable	(I) region_N	(J) region_N	Mean Difference (I–J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
bmi	1	2	-.030	.459	1.000	-1.243	1.183
		3	-4.174*	.447	.000	-5.354	-2.994
		4	-1.424*	.459	.012	-2.637	-.211
	2	1	.030	.459	1.000	-1.183	1.243
		3	-4.144*	.446	.000	-5.323	-2.965
		4	-1.394*	.459	.015	-2.606	-.182
	3	1	4.174*	.447	.000	2.994	5.354
		2	4.144*	.446	.000	2.965	5.323
		4	2.750*	.446	.000	1.571	3.930
	4	1	1.424*	.459	.012	.211	2.637
		2	1.394*	.459	.015	.182	2.606
		3	-2.750*	.446	.000	-3.930	-1.571
charges	1	2	978.380	947.501	1.000	-1525.108	3481.869
		3	-1307.166	921.861	.939	-3742.908	1128.576
		4	1057.351	947.488	1.000	-1446.103	3560.805
	2	1	-978.380	947.501	1.000	-3481.869	1525.108
		3	-2285.546	921.183	.079	-4719.497	148.405
		4	78.971	946.766	1.000	-2422.575	2580.517
	3	1	1307.166	921.861	.939	-1128.576	3742.908
		2	2285.546	921.183	.079	-148.405	4719.497
		4	2364.517	921.122	.062	-69.274	4798.308
	4	1	-1057.351	947.488	1.000	-3560.805	1446.103
		2	-78.971	946.766	1.000	-2580.517	2422.575
		3	-2364.517	921.122	.062	-4798.308	69.274

Based on estimated marginal means

*. The mean difference is significant at the

b. Adjustment for multiple comparisons: Bonferroni.