MKTG402(3) Labs

Ahmad Daryanto

February 13, 2024

Table of contents

Preface

This is a website for the workshops of MKTG402(3) - Introduction to Quantitative Research Methods for students at the Advanced Marketing Management (AMM) program at Lancaster University.

Instructor: **Dr. Ahmad Daryanto**

Document Version

The pdf version of this document is available to download from Moodle.

SPSS Version

I am using SPSS version 28. You can use any SPSS versions (e.g., ver 26, 27, and 29), where the statistical procedures used in the labs are the same but the output formats are slightly different from those of version 28. Therefore, I recommend you to use version 28, so you can check whether your outputs are the same with those shown in this document.

Textbook

I do not use a specific textbook. Any marketing research or quantitative research textbook can be used for your self-learning, in particular, I recommend the following text:

- Malhotra, N.K, Birks, D., and Wills, P. (2012) Marketing Research: An applied orientation, 4^{th} edition, London: Prentice-Hall, Pearson. Other books on marketing research by Malhotra are also useful.
- Sarstedt, Marko and Erik Mooi (2014). A Concise Guide to Market Research. Springer. As the title says, this book is concise focusing statistical tests typically conducted in market research. The organization of the chapters in the book match closely with what we cover in this module. You can read the e-version of the first edition via our library website.

- Feinberg, Kinnear, and Taylor. (2013). Modern Marketing Research: Concept, Methods, and Cases, 2nd edition, Cengage. Written by excellent researchers in Marketing, I also like this book. It has plenty of mini cases that help readers to see the application of concepts in practices.

Overview

Workshop 1/ Week 13: Getting started with SPSS

Tasks: In this workshop, we will label variables, create new ones, and code scales.

Workshop 2/ Week 14: Descriptive statistics, graphs and charts in SPSS

Tasks: In this workshop, we will use SPSS to produce descriptive statistics, graphs and charts for different types of variables.

Workshop 3/ Week 15: Crosstabs and two independent samples t-test

Tasks: This is the first of the bivariate analysis workshops. Here we will look at the analysis of 2 sets of nominal data in a cross tabulation and use a Chi-square test to determine whether a significant difference between the groups has occurred. In the second part of the workshop, we will look at the analysis of mean scores, i.e., comparing mean scores of two groups with two independent samples T-test.

Workshop 4/ Week 16: ANOVA and Experimentation

Tasks: Here we will use SPSS to compare the mean scores of two or more nominal groups. Using one-way ANOVA, we will determine if significant differences exists between two groups or more. Then we will look two-way ANOVA, specifically a 2x3 ANOVA between-subject design.

Workshop 5/ Week 17: Regression

Tasks: Using correlation we will determine the nature of the association between two sets of interval or ratio data. Then using linear regressions, we will develop a model to examine the influence of a set of independent variables on one dependent variable.

1 Getting started with SPSS (Week 13)

Data: Dell.sav

- Data is available on Moodle
- On Moodle, go to a section `Workshop Materials`, and open a folder `Workshop 1, 2, 3 - Data and Articles`. Download Dell.sav and **Dell Questionnaire** .

1.1 Learning objectives

The aim of this first lab exercise is to get you familiar with the SPSS program.

Learning objectives:

At the end of this lab, we hope that you will be able to:

- Explore a data set and produce descriptive statistics in table format
- Transfer tables into Word

In this lab we will use a data set from real customers that is available to download from Moodle. The data set contains information from 372 customers collected from a survey of purchasers of Dell PCs and notebooks. With this survey, Dell wants to understand their customers' primary usage for internet and their customers' satisfaction with their purchases. Customers also give information about their demographics such as age and gender. Your primary job today is to describe the characteristics of the sample.

1.2 Questionnaire and Data

Before opening that data and starting to analyse it, it is very important that you understand what was measured, how it was measured and as a result what level of measurement has been used. As we go on in the course it is very important to understand the distinctions between nominal, ordinal, interval, and ratio data.

Task

Have a look at the questionnaire and identify which questions use a nominal, interval, and ratio scale.

1.3 Working with SPSS

Open the data file Dell.sav, which you must download from Moodle. You should save the data to your own drive and then double-click the file name from windows explorer.

There are two different views of the data which can be seen by clicking the bottom left tab. The data view shows the imputed data for each respondent (each row represents a respondent and each column represents a variable or questionnaire item). The variable view shows the detail of what has been measured. The key columns are the Name, Labels and Values columns which give the shortened variable name (e.g. see variable `q2_job` in the third row), the full details of what has been measured as a label (e.g. please indicate which of these you have ever done on the Internet: Looked for a job) and the Values represent what the imputed numbers represented in the questionnaire (e.g. 0 = "don't know", 1 = "looked for a job", 2= don't look for a job).

1.4 Creating Tables

I want you to try and use SPSS on your own (I will come around and answer questions that you have). You can explore the data set as much as you wish but here are some ideas of tables which might be interesting to explore.

1.5 Frequency Tables

This table allows you to explore only one variable in each table. Use the commands **Analyze→Descriptive statistics→Frequencies**. If you click on the `Charts` tab you can also get SPSS to generate bar, pie, and histograms for the variables selected.

Note that you can get many frequency tables in the one command by selecting multiple variables from the list of those available. To select more than one variable at a time keep the `Ctrl` button down. The frequency tables will be displayed in the output window.

Task

Why not create a frequency table for the variables named `q2_job`, `q2_trip`?

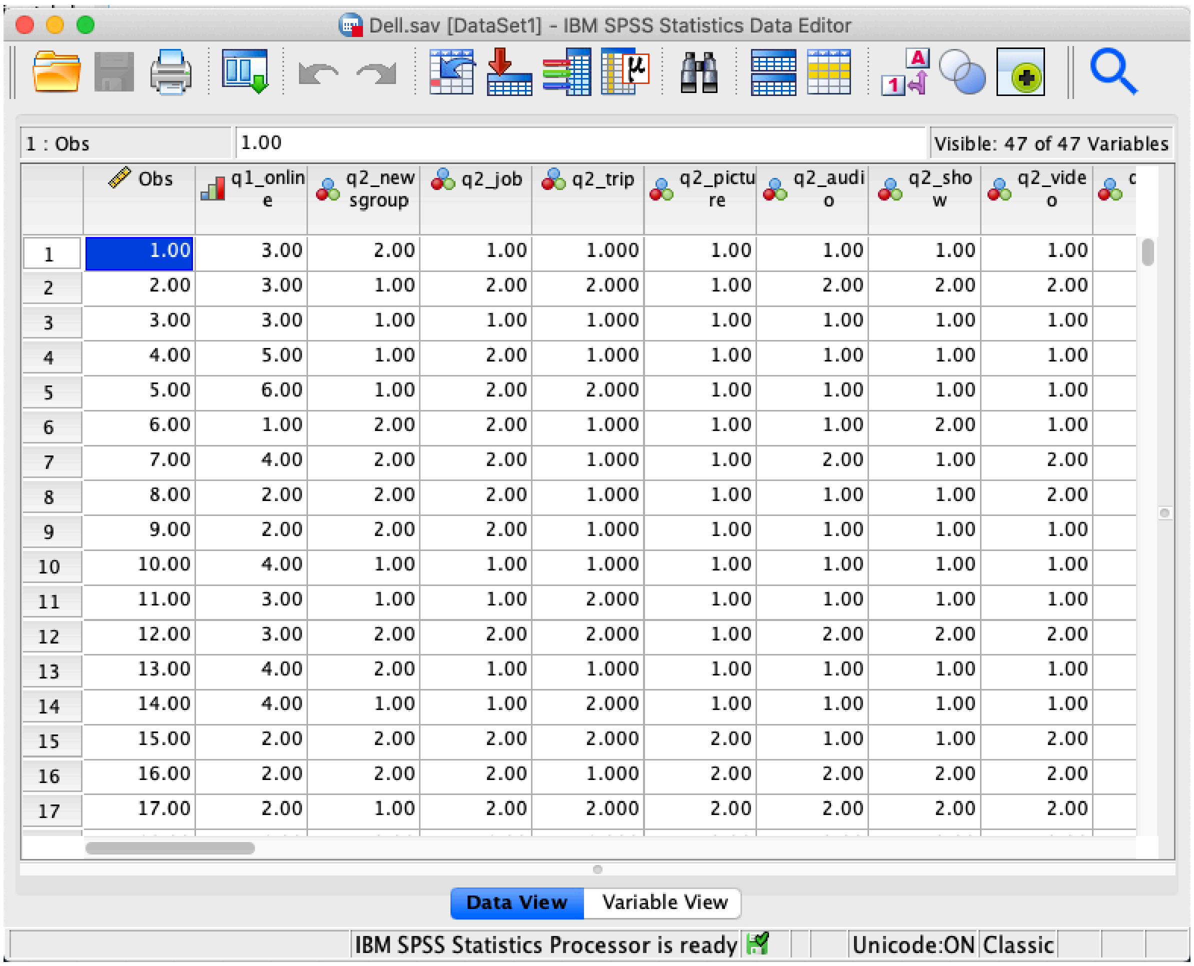


Figure 1.1: Data View: Now click variable view at the bottom of the window.

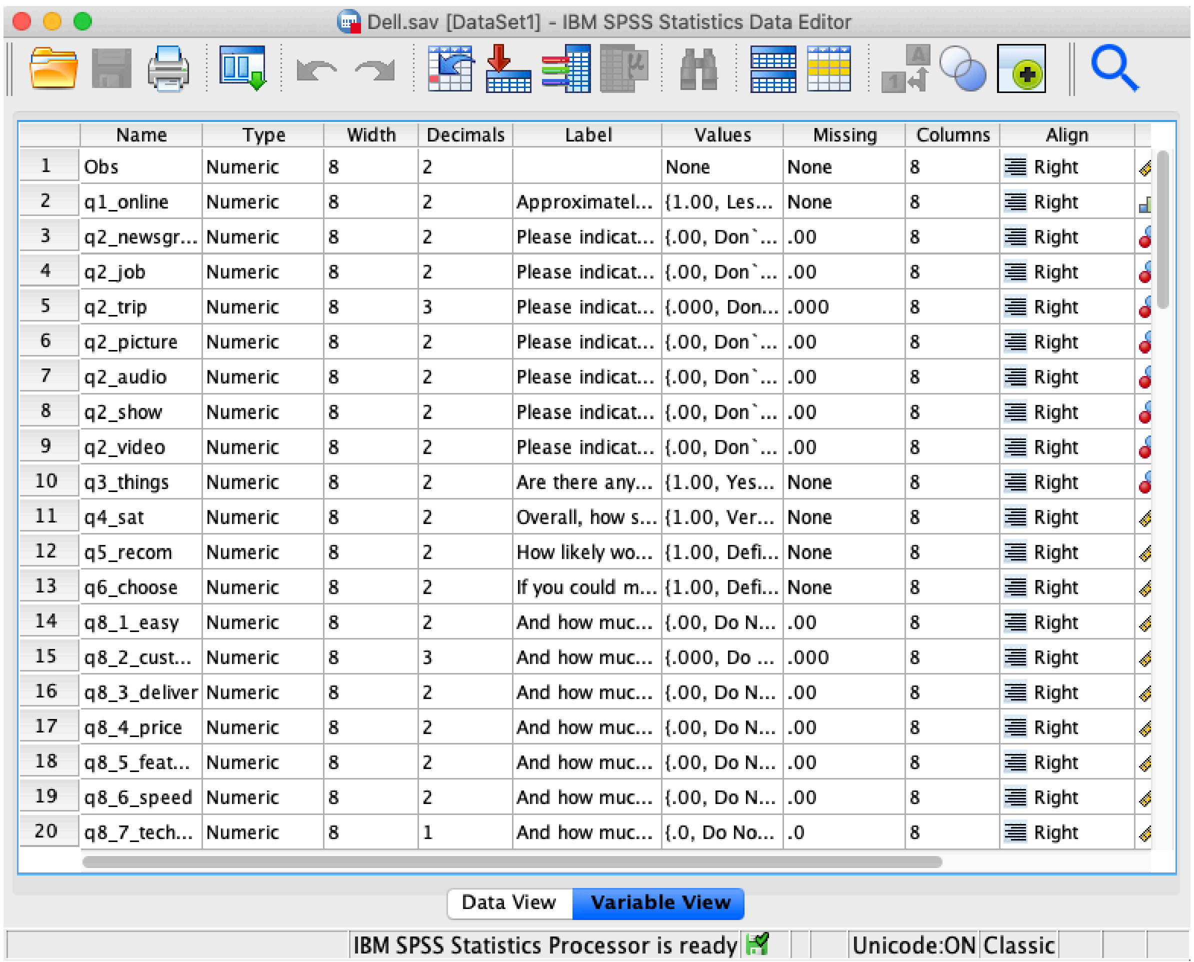


Figure 1.2: Variable View

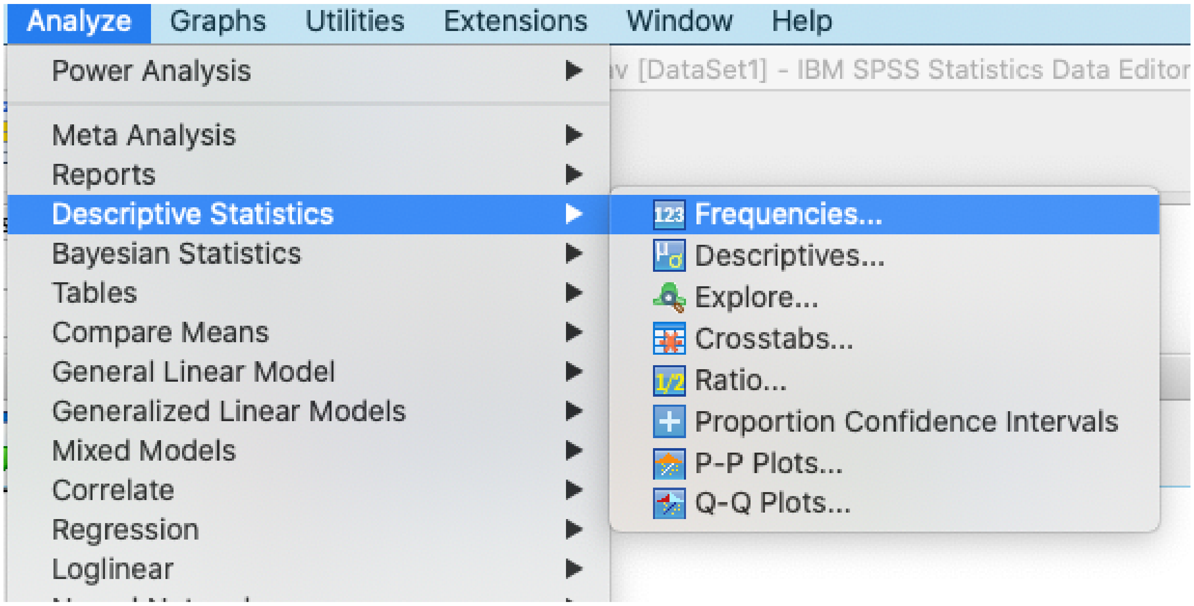


Figure 1.3: Data view with menu options displayed

i Task

Now, I want you to do more analysis to be able to answer the following questions:
1. What can you say about the sample regarding the distribution of the number of hours that customers spend online? (Q1)
2. What are the proportions of male and female in the sample? (`Q14_gender`)
3. What is the percentage of Dell customers in the sample who are willing to joint Dell loyalty program? (`Q15_loyal`)

1.6 Getting Help in SPSS

If you want to find out more about interpreting the frequency table, why not try using the `Help` tab? Select again **Analyze**→ **Descriptive statistics** → **Frequencies**, and click on the `Help` tab at the bottom. You will find that a new window opens up and provides useful information. SPSS also has a range of support features which can be found using the `Help` menu. Spend some time looking through the options of particular interest maybe the Statistics Coach, Tutorials and Case Studies – you can click this link Getting help - IBM documentation, which directly take you to the online help page of SPSS ver. 28.

Statistics

		Please indicate which of these you have ever done on the Internet: Looked for a job	Please indicate which of these you have ever done on the Internet: Booked trips
N	Valid	372	372
	Missing	0	0

Frequency Table

Please indicate which of these you have ever done on the Internet: Looked for a job

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Looked for a job	161	43.3	43.3	43.3
	Did not look for a job	211	56.7	56.7	100.0
	Total	372	100.0	100.0	

Please indicate which of these you have ever done on the Internet: Booked trips

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Planned or booked trips	239	64.2	64.2	64.2
	Did not plan or booked trips	133	35.8	35.8	100.0
	Total	372	100.0	100.0	

Figure 1.4: Outputs

1.7 Cross Tabulations

What are the proportions of male vs. female that are willing to join Dell loyalty program? To answer this question, you need to do cross tabulations. Use the commands **Analyze→Descriptive statistics→Crosstabs**.

Crosstabs are used to explore the relationship between two categorical variables (e.g., nominal by nominal, nominal vs. ordinal). Use the cells tab to select row or column percentages in order to be able to compare across the groups. The type of percentage will depend on what basis you want to compare i.e. whether the rows or the columns should add to 100%. Note that you can also produce graphs (by checking the `Display clustered bar charts` option).

> **Task**
>
> What about exploring the relationship between gender and level of education?

1.8 Copying SPSS output into word

For reports or as a record of what analysis you have done you may want to put your charts and tables into an MS Word document. This is straight forward, all you need to do is position the cursors on the SPSS output (chart or table) and right click, select `copy` and paste the table or graph into your MS Word document.

1.9 Codebook

If you use a paper-and-pencil survey, you need to enter the data manually. Before you enter the data into an SPSS data view window, I recommend that you create a codebook in advance. A codebookt contains names and descriptions of variables and coding for response answers including missing values, which is built directly from your questionnaire. A codebook is very useful and handy to place on your desk when you are working on your data where you quickly check the variables that you use, and communicate the SPSS outputs to your team members.

I also recommend that you also create a codebook when you use an online survey (e.g., https://lancasteruni.eu.qualtrics.com. Online surveys like qualtrics.com enable you to import data into an SPSS format (.sav). However, you may need to clean this data before you can use it (you may not need to do so if you want a quick report). For instance, the column label in the SPSS variable view window will be written by the software automatically. When you have a codebook, you can leave this column as blank and adjust the variable names as the way they are written in your codebook.

Example of a codebook

Variable	Construct	Statement	Response	Source
q1_online	Amount of hours spent online	Approximately how many hours per week do you spend online? (This would be the total from all the locations you might use)	1= Less than 1 hour 2=1-5 hours 3=6-10 hours 4=11-20 hours 5=21-40 hours 6=41 hours more 99=Missing values	Self-developed
...	...	...	...	
q4_sat	Satisfaction Overall	How satisfied are you with your DELL computer system?	1= Very dissatisfied 2 = Somewhat dissatisfied 3 = Somewhat satisfied 4 = Very satisfied 99 = Missing values	Adapted from Oliver(1993)

1.10 Lancaster Qualtrics

Lancaster University has a subscription to an online survey Qualtrics. You can use your university email to log onto Qualtrics.

Task

Click this link https://lancasteruni.eu.qualtrics.com, and log onto Qualtrics.

Important

Find more information about Qualtrics (e.g., training, how to create and edit a survey, etc) at the ISS page here

2 Descriptive Statistics, Graphs and Charts in SPSS (Week 14)

Data: Dell.sav

Data is available on 'Workshop Materials' folder on Moodle.

2.1 Learning objectives

The aim of this second lab exercise is to get you familiar with the SPSS program in terms of further descriptive statistics and basic computations. Through completing this lab you will be able to:

- Explore a data set and produce descriptive statistics in graph format.
- Produce descriptive statistic output.
- Compute a new variable based on the data provided.

Last time we focused on producing tables, in this lab we are now focusing on producing charts and instead of using the `Analyze` menu we will be using the `Graphs` menu.

2.2 Charts

You can use the charts from the commands already undertaken in lab 1, however if you want to explore other charts or customize a chart you should use the specifically developed chart tool. If you use SPSS ver 28 and below, click **Graphs→Legacy dialogs** and then select the type of chart that you want. In SPSS ver 29, SPSS gets rid of **Legacy dialogs** – you will see list of graphs you can use after clicking **Graphs**.

To start off with, try a pie chart for gender (`q14_gender`), a bar chart for level of education (`q11_grade`) and a box plot for satisfaction (`q4_sat`).

Make sure that you use appropriate titles for your charts and label the axes accordingly. Try exploring counts and percentages as well as the different chart types.

⚠ Attention

When each dialogue box appears you need to select the correct summary. For the bar chart you want to select a simple bar chart with “summaries for groups of cases”, you also want “summaries for groups of cases” for the pie chart but for the box plot you should select “summaries for separate variables”.

If you want to explore putting data labels on your charts and for instance changing the color, in the output widow double click on the chart and it will open up in an editor view as the following screen shot shows

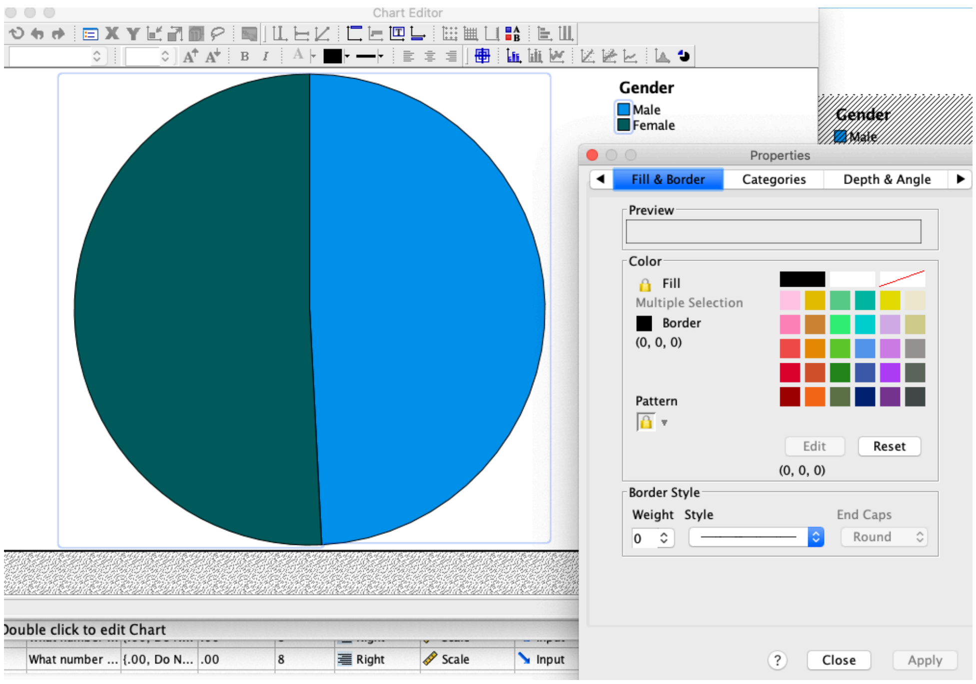


Figure 2.1: Pie chart

ℹ Task

Which type of chart is best suited to the following types of data: nominal, ordinal, interval or ratio?

2.3 Descriptive Statistics

Let’s now try using the Descriptives function using the commands **Analyze→Descriptive statistics→Descriptives**. Here you can calculate mean, standard deviations and other measures of central tendency and dispersion

> **Task**
>
> Why not try and explore the satisfaction of DELL customers (`q4_sat`) and also the amount that customers have spent on internet in the 12 month (`q16_spent`)?

2.4 Computing New Variables

It is also important to be able to perform basic manipulations of the data. The most important manipulation is the creation of new variables. Referring back to the week 2 lecture you will remember that a lot of marketing research data is collected through the use of multi-item scales. The data from the multi-item scale is then averaged into **a single composite measure**. Most of the DELL questionnaire is developed from scales used in marketing literature. You should practice creating composite variables for concepts such as Market Maven, Opinion Leadership, and Innovativeness. You can identify which items capture which concept from the variable name which is best seen from the variable view which you can click to from the bottom left hand corner of the screen.

2.5 Reliability Analysis - Computing Cronbach's alpha

Before creating a new composite variable it is important to check that the items exhibit internal consistency reliability (also discussed in the second lecture). You can do this by selecting the menu options **Analyze→ Scale→Reliability Analysis**. You then select each of the items used to measure the construct.

Conduct a reliability analysis for Opinion Leadership (Variable names: `q10_op1`, `Q10_op2`, `Q10_op3`). If the alpha value given in the output is 0.7 or above then you can create a new composite measure by averaging the items together, knowing that all the items get well as a group.

If the alpha is below 0.7 it may be because some items in the scale are reversed coded, or do not get well with the rest. In order to find out which scale items need to be reversed or removed, when the alpha dialog box is open click the `statistics` tab and then check `Scale if item deleted`.

The alpha value for Opinion Leadership is 0.927 and therefore the three items can be averaged. To do this use the menu commands **Transform→Compute**. The new variable will appear at the end of the dataset so please look for it in the data view. Note that you will be able to name the new variable whatever you like but I would suggest calling it `Opinion_avg`.

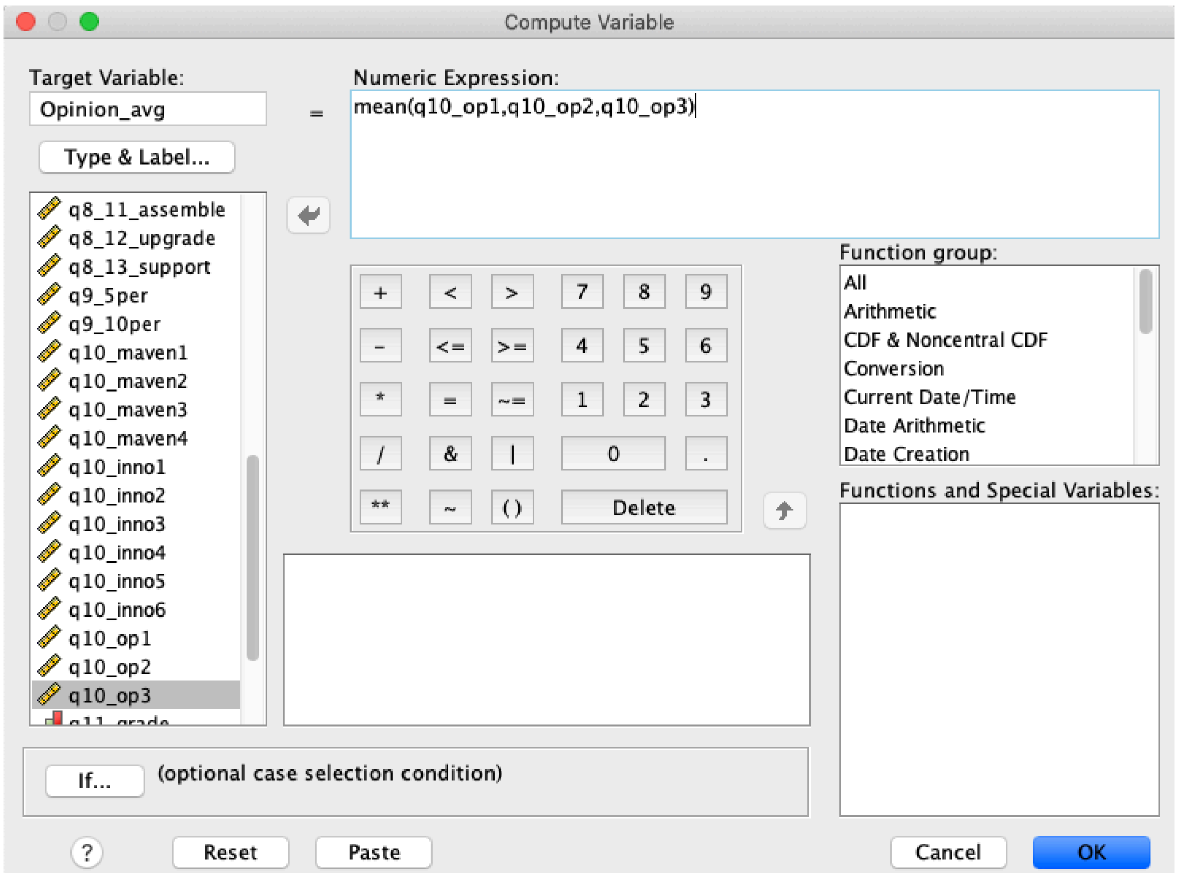


Figure 2.2: Computing mean

Task

Why not try and explore the new variable you just created? (e.g., report mean, standard deviation, inspect the shape of its distribution, create a boxplot)

2.6 Video

Boxplot

Reliability analysis

2.7 Readings

Feick, L. F., & Price, L. L. (1987). The market maven: A diffuser of marketplace information. *Journal of Marketing*, 51(1), 83-97.

Goldsmith, R. E., Flynn, L. R., & Goldsmith, E. B. (2003). Innovative consumers and market mavens. *Journal of Marketing Theory and Practice*, 11(4), 54-65.

3 Crosstabs and Two Independent Samples T-test (Week 15)

Data: Dell.sav

- Data is available on Moodle

3.1 Learning objectives

The aim of this third lab is to help you to use SPSS to examine group differences based on demographic factors. In this lab we will cover:

- Chi-square test
- The two independent samples t test

3.2 Chi-square Test

The chi-square test is useful for determining if differences exist between two categorical variables. This test can be used to substantiate perceived associations when calculating crosstabs such as those that you did in lab 1. Let's say that you want to explore if there is an association between gender (`Q14_gender`) and number of hours spent on internet (`Q1_online`). A chi-square test would be useful to assess this.

As in lab 1, use the menu options **Analyze→Descriptive statistics→Crosstabs**. I prefer to insert `q14_gender` into column and `q1_online` into row window–this way, you will see the distribution of gender within each category of `q1_online`. Select the statistics tab and click on the Chi-square option. Note that you can also click on `Phi and Cramer's V to` get an indication of the strength of the association.

Using the cells tab you should also select row or column percents to help you to see the pattern of association (in this example, I prefer to select row percentages)

In the output window you will see that three (or four with the addition Phi and Cramer's V) tables are produced. You can ignore the `Case Processing Summary`.

The second table is the crosstab table that you produced in lab 1 and the third table `Chi-Square Tests` gives you the new results. You will see that you have a table and graph like the one below for the gender by online crosstabs.

The p-value associated with the crosstabs is given by the following table

The most important column within the table is the "Asymp. Sig. (2-sided)". This is the p-value column and the result above indicates that there is an evidence of an association between the rows and columns (because the p-value is smaller than 0.05), in this case between gender of the consumer and number of hours online. It is important to look at the proportion of consumers in each of the six groups.

If there was a statistically significant association you would ask the following:

1. Where do you see the association between the proportions of consumers?
2. How strong is the association?

> **i Task**
>
> Have a look at the clustered bar chart, and explain the pattern of the association between `hours spend online` and `gender`. Is the association significant? (inspect the p-value of the Pearson Chi-Square)

> **i Task**
>
> Based on the Phi value, assess the strength the association between `hours spend online` and `gender`?

> **! What a Phi value tells us**
>
> Phi tells us about the strength of an association. Its value ranges from 0 to 1. Cohen (1988) provides the following guideline:
>
> - 0.1 is considered small
> - 0.3 medium
> - 0.5 large

3.2.1 Post-Hoc Test

If the association were significant, you would want to know further which group comparisons are actually significantly different. This is called post-hoc tests or multiple comparison tests.

Approximately how many hours per week do you spend online? This would be the total from all the locations you might use * Gender Crosstabulation

			Gender		
			Male	Female	Total
Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	Less than 1 hour	Count	7	13	20
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	35.0%	65.0%	100.0%
	1 to 5 hours	Count	52	87	139
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	37.4%	62.6%	100.0%
	6 to 10 hours	Count	57	48	105
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	54.3%	45.7%	100.0%
	11 to 20 hours	Count	32	23	55
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	58.2%	41.8%	100.0%
	21 to 40 hours	Count	20	14	34
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	58.8%	41.2%	100.0%
	41 hours or more	Count	15	4	19
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	78.9%	21.1%	100.0%
Total		Count	183	189	372
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	49.2%	50.8%	100.0%

Figure 3.1: Frequency Table

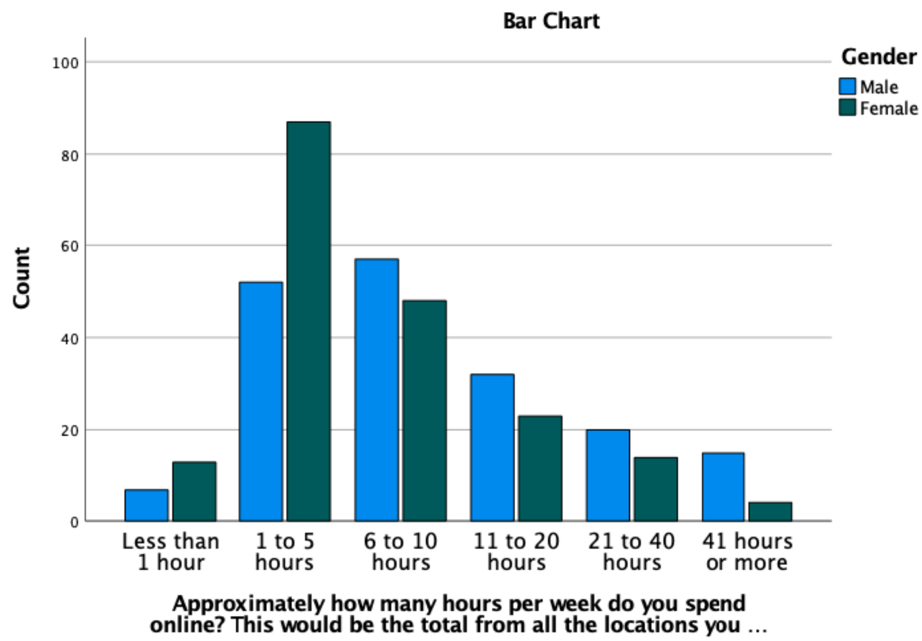


Figure 3.2: Clustered Bar Chart

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	20.193[a]	5	.001
Likelihood Ratio	20.739	5	<.001
Linear-by-Linear Association	17.521	1	<.001
N of Valid Cases	372		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 9.35.

The result is not reliable if count is less than 5

Figure 3.3: SPSS Output: Chi-Square Tests

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.233	.001
	Cramer's V	.233	.001
N of Valid Cases		372	

Figure 3.4: Effect size

In essence, you want to test all pairwise comparisons to detect where the significant occurs. For example, is the proportion of male vs female significantly different within “less than 1-hour”, within “1-5 hours” group, etc?

How?

Use the menu options **Analyze→Descriptive statistics→Crosstabs**. Insert `gender` into “Columns” and `q1_online` into “rows”. Click `Cells` tab on the right, then `Tick`Row percentages`. Under z-test, tick`Compare columns proportions`and`Adjust p-values (Bonferroni method)‘.

What does SPSS do with “Adjust p-values”? The answer is that SPSS will conduct z-tests for each comparison. Because of conducting multiple comparisons tests, the significance level at each test should not be 5% anymore. It should be divided by the number of comparisons. For example, if you have 6 pairwise comparisons, the sig. level at each test would be 0.05/6=0.008. This is to make sure that all comparisons tests were maintained at alpha=5%. If you do not adjust it and use alpha=5% in each of the pairwise tests, the probability of getting at least one significant result due to chance is high.

3.2.2 Post-Hoc Test Output

The output of the post-hoc test is given below

> **Task**
>
> Looking at the output of the post-hoc test, what do you conclude?

Approximately how many hours per week do you spend online? This would be the total from all the locations you might use * Gender Crosstabulation

			Gender: Male	Gender: Female	Total
Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	Less than 1 hour	Count	7_{a}	13_{a}	20
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	35.0%	65.0%	100.0%
	1 to 5 hours	Count	52_{a}	87_{b}	139
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	37.4%	62.6%	100.0%
	6 to 10 hours	Count	57_{a}	48_{a}	105
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	54.3%	45.7%	100.0%
	11 to 20 hours	Count	32_{a}	23_{a}	55
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	58.2%	41.8%	100.0%
	21 to 40 hours	Count	20_{a}	14_{a}	34
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	58.8%	41.2%	100.0%
	41 hours or more	Count	15_{a}	4_{b}	19
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	78.9%	21.1%	100.0%
Total		Count	183	189	372
		% within Approximately how many hours per week do you spend online? This would be the total from all the locations you might use	49.2%	50.8%	100.0%

Each subscript letter denotes a subset of Gender categories whose column proportions do not differ significantly from each other at the .05 level.

"If a pair of values is significantly different, the values have *different* subscript letters assigned to them" (IBM SPSS)

Figure 3.5: Output of the post-hoc test

3.3 Two Independent Samples T Test

This test allows you to explore differences in mean values across two groups e.g., between low vs. high-income groups; male vs. female consumers; those who received a reward vs. nothing; those who get a flu jab vs. placebo, country A vs. country B, etc.

Now, we want to assess if there are differences in the Opinion of Leadership (average scores of `q10_op1`, `q10_op2`, `q10_op3`) across gender (`q14_gender`).

As in the previous lab, you need to conduct reliability analysis to find out whether the three items of Opinion Leadership get well together, if yes, then you need to create a composite score by taking the average of the three items (I would suggest that you name the new variable as `Opinion_avg`).

In this lab, you will test following hypothesis:

H0: Opinion leadership of male consumers = Opinion leadership of female consumers.

H1: Opinion leadership of male consumers > Opinion leadership of female consumers.

> **Task**
>
> The above alternative hypothesis is directional. Formulate a non-directional alternative hypothesis instead.

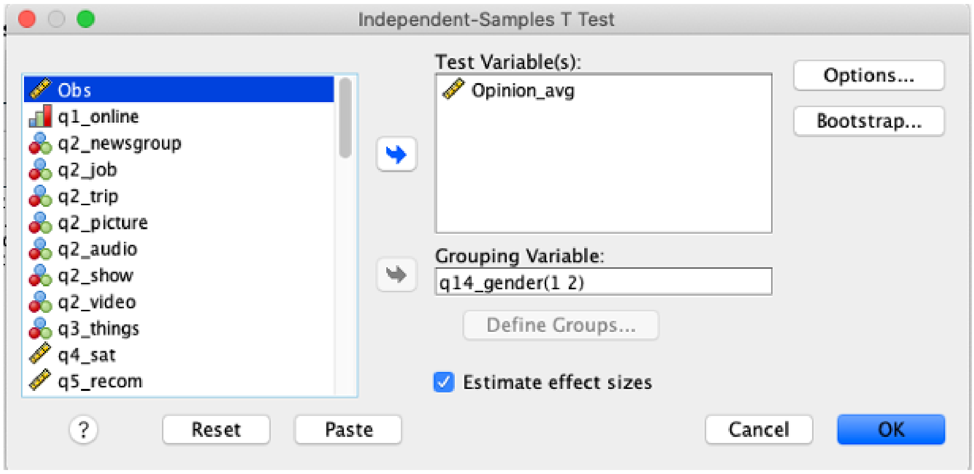
> **Warning**
>
> Note that when using a two independent samples test, the grouping variable (e.g., gender) is a qualitative variable – nominal or ordinal, and the test variable (e.g., amount money spent, scores on opinion leadership) is always a quantitative variable – interval or ratio.

To conduct a T test, use the menu options **Analyze→Compare means→Independent-Samples T test**. The test variables will be `Opinion_avg`. The grouping variable is `gender`. You need to “define groups” this means telling SPSS that gender is measured using the numbers 1 and 2 (you will see from the data view that `loyalty card status` is coded as 1 and 2). By default, ‘Estimates effect size’ was ticked’.

3.3.1 Outputs

In the output window you will have three tables, the first giving the “Group Statistics”, the second giving the results of the “Independent Samples Test” (The important column is the sig. columns which give the p-values.), and the third gives the estimate of the effect size.

First, we need to look at the descriptive statistics to get a clear picture of the mean score of female vs. male on `Opinion_avg`, and assess the difference between the means.



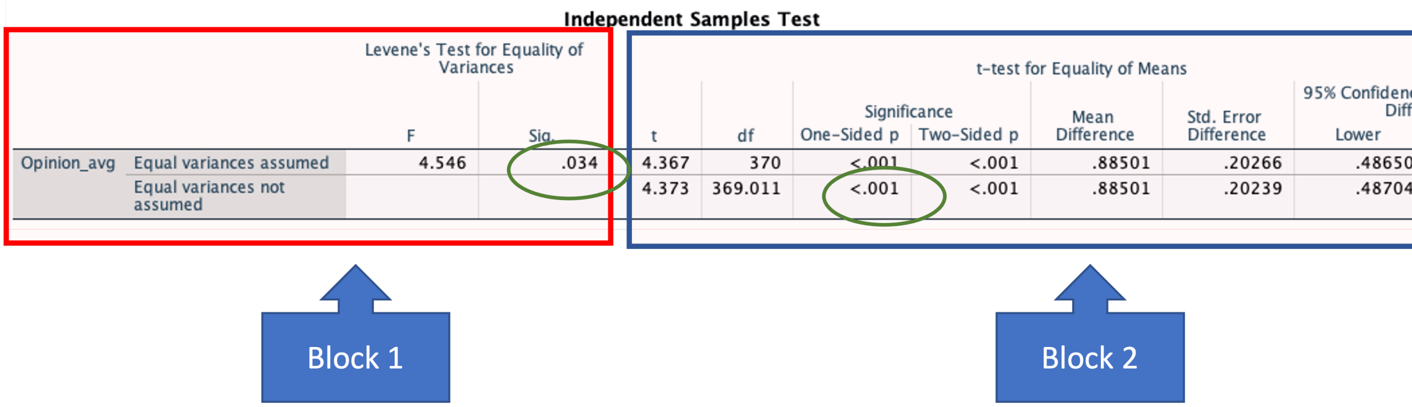
Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Opinion_avg	Male	183	4.6557	1.86900	.13816
	Female	189	3.7707	2.03320	.14789

The std. deviation is helpful in understanding the Levene's test results reported in the second table.

> **Task**
>
> Look at the descriptive statistic above, say something about the mean scores and their standard deviations.

Now, let us discuss the second table.



Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
						Significance				95% Confiden… Diff…
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower
Opinion_avg	Equal variances assumed	4.546	.034	4.367	370	<.001	<.001	.88501	.20266	.48650
	Equal variances not assumed			4.373	369.011	<.001	<.001	.88501	.20239	.48704

The second column in the table gives the significance value for Levene's test (**Block 1**). This tells you if you can assume that the variance for the groups is equal, which cannot be assumed for this example and therefore when interpreting the result of the T-tests you should look at the p-value for "equal variances not assumed" (bottom row of **Block 2**). The p-value associated with the one-sided test was circled because our alternative hypothesis was directional (the males' scores is higher than the females' scores).

> **Task**
>
> The one-sided p-value in the second block is significant (circled in green), what do you conclude?

3.3.2 Is the Difference Practically Meaningful?

If the mean difference is significant. We should ask if the different is meaningful or not. To answer this question, we need to assess the effect size table.

What is effect size? Imagine that if you took a paracetamol when you had a migraine and the pill reduces the pain slightly but it is not really helping you much – so the effect of the pill is small.

i Guidelines on the effect size of mean difference (Cohen 1988)

d = 0.2 is considered small
0.5 medium
0.8 large
For example, if the difference between two groups' means < 0.2 standard deviations, despite the difference being significant, it is not practically meaningful.

Let us look at the effect size output below:

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
Opinion_avg	Cohen's d	1.95416	.453	.247	.658
	Hedges' correction	1.95813	.452	.246	.657
	Glass's delta	2.03320	.435	.227	.643

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

i Task

Look at the Cohen's d in the above table, what does it tell you?

3.3.3 Interpretation of the Findings

Looking at the descriptive statistics and statistical results, we can conclude that there are statistically significant differences between male and female consumers on their ratings on opinion leadership. Specifically, male consumers have a higher rating of opinion leadership compared to that of female consumers ($M_{Male} = 4.66$ vs. $M_{Female} = 3.77$, Cohen's d = 0.45). Furthermore in terms of the variability in the opinion leadership across the two groups (male and female consumers), the Levene's test show that there are differences in the spread, that is, the variability in the opinion of leadership is not the same for both groups ($SD_{Male} = 1.87$ vs. $SD_{Female} = 2.03$).

i Task

Why not investigating if there are differences in the satisfaction level between those who are willing to participate in Dell loyalty program and those who are not.

3.4 Video

One-sample T-test

Two independent samples T-test

Crosstabs

3.5 Readings

Feick, L. F., & Price, L. L. (1987). The market maven: A diffuser of marketplace information. *Journal of Marketing*, 51(1), 83-97.

Goldsmith, R. E., Flynn, L. R., & Goldsmith, E. B. (2003). Innovative consumers and market mavens. *Journal of Marketing Theory and Practice*, 11(4), 54-65.

4 ANOVA and Experimentation (Week 16)

Data: MBGshort.sav

Data is available on 'Workshop Materials' folder on Moodle.

4.1 Learning objectives

The aim of this lab is to help you to use SPSS to analyze data from an experimentation. Specifically, we want to examine mean differences across three experimental groups (one-way ANOVA design) and analyze data of a more complex experiment (i.e., two-way ANOVA in the form of a 3x2 between-subject design).

Learning objectives: At the end of this lab, we hope that you will be able to

- Analyze data from a one-way ANOVA (Analysis of Variance) experiment.
- Produce and interpret basic SPSS outputs from a two-way ANOVA experiment.

In this lab, we are going to look at how one-way ANOVA can be used to extend the two independent samples T test to look for means differences across three or more groups and secondly to show how data resulted from manipulation of more than one factor can be analyzed by extending one-way ANOVA method. For instance, to look at how types of return condition facilitated by money-back guarantees (MBG for short) (No MBG, 15 days, 30 days) and types of product (search vs. experience product) influence perceived product quality. We focus our attention to between-subject ANOVA designs where each respondent is randomly assigned to only one of the experimental conditions.

4.2 Mean Differences Across Three or More Groups

One-way ANOVA can also be used to explore differences in a variable across three or more groups. This is more useful than the two independent sample T test simply because it can be used for more groups and can tell us where the location of the differences. For instance you might find significant mean differences on perceived product quality across three groups (e.g., No MBG, 15 days, and 30 days) and to be more specific, the 15 days and 30 days group do not differ. But perceived product quality of participants in the No MBG group is different than those in the 15 days and 30 days group.

In this part, we are going to use the data collected by a former Advanced Marketing Management student at Lancaster University for his dissertation about the effect of MBG on perceived durability of a product. He created an experiment by devising three different scenarios where each scenarios contains information about each of MBG conditions (No MBG, 15 days, 30 days). Respondents were randomly assigned to read one of three different questionnaires. In each questionnaire, he put an image of a product (he chose a laptop) and information about the MBG condition. Other information across different questionnaires was kept similar (e.g., product specification, laptop price, etc).

Figure 4.1: An example of stimulus used in the 30 days MBG condition. This image was created by William, a former AMM student for his MSc dissertation

Let us explore if the difference exist in the perceived product durability across the three groups MBG. We will use quality5 (i.e., 'the product of this offer would be likely to be durable') as the dependent variable.

To conduct an ANOVA, use the menu options **Analyze**→Compare means→One-way ANOVA. Click `Options` then tick `Descriptive`. You can also tick **Options**→means plot to display a graph that shows the means of the three groups Click `Continue` then `OK`. The dependent variable is `quality5` (i.e., 'the product of this offer would be likely to be durable'). The factor is `tMBG`.

In the output window the SPSS produces the Descriptive table and ANOVA table.

Descriptives

The product of this offer would be likely to be durable

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
No MBG	143	2.85	1.439	.120	2.62	3.09	0	6
15 days	140	3.22	1.341	.113	3.00	3.45	0	6
30 days	131	3.34	1.335	.117	3.11	3.57	0	6
Total	414	3.13	1.386	.068	3.00	3.27	0	6

ANOVA

The product of this offer would be likely to be durable

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	18.099	2	9.050	4.796	.009
Within Groups	775.594	411	1.887		
Total	793.693	413			

The ANOVA table reveals that perceived product durability differ across the experimental groups because the sig-value is less than 5%. The descriptive table gives an indication of why the differences exist. No MBG group has a lower mean compared the 15-days and 30-days group. The 15-days and 30-days groups are similar. However, you need to do additional test to get more details about the differences. Let's try again putting some important options.

4.3 Homogeneity of Variance and Post-Hoc Tests

4.3.1 Homogeneity of Variance Test

Repeat the analysis above and click the `Options` tab. Check on `Options` and tick `Homogeneity of Variance test`. To determine which of the three groups differ, you can do what is called a Post Hoct test - click `Post Hoc`. There are many tests available. Among many options, check `Scheffe` for situation where Equal Variances Assumed; check `Games-Howell` under Equal Variances Not Assumed.

Homogeneity variances test determines whether you would use Scheffe or Games-Howell test. You use this rule:

> **Tip**
>
> If sig-value < 0.05, use Games and Howell, otherwise use Scheffe

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
The product of this offer would be likely to be durable	Based on Mean	.099	2	411	.906
	Based on Median	.109	2	411	.897
	Based on Median and with adjusted df	.109	2	409.066	.897
	Based on trimmed mean	.080	2	411	.923

The sig-value of the homogeneity of variance test is 0.906, therefore you focus on the outputs of the Scheffe test in the next table. Let us interpret the results of the Scheffe test.

4.3.2 Post Hoc Test

In the SPSS outputs, you have the following table

Multiple Comparisons

Dependent Variable: The product of this offer would be likely to be durable

	(I) type of MBG	(J) type of MBG	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	No MBG	15 days	-.368	.163	.080	-.77	.03
		30 days	-.490*	.166	.013	-.90	-.08
	15 days	No MBG	.368	.163	.080	-.03	.77
		30 days	-.122	.167	.766	-.53	.29
	30 days	No MBG	.490*	.166	.013	.08	.90
		15 days	.122	.167	.766	-.29	.53
Games-Howell	No MBG	15 days	-.368	.165	.068	-.76	.02
		30 days	-.490*	.168	.010	-.89	-.10
	15 days	No MBG	.368	.165	.068	-.02	.76
		30 days	-.122	.163	.733	-.51	.26
	30 days	No MBG	.490*	.168	.010	.10	.89
		15 days	.122	.163	.733	-.26	.51

*. The mean difference is significant at the 0.05 level.

Figure 4.2: Post Hoc Tests

You can see from the above table that No MBG is not statistically different from 15 days (p = 0.08) and 15 days is not statistically different from the 30 days (p = 0.766). But no MBG has a statistical difference with the 30 days (p = 0.013) (the p-value between No MBG and 15 days are close to significant, if sample size is large, it would be likely to be significant). The next output from the Scheffe test below clarifies the difference.

The product of this offer would be likely to be durable

	type of MBG	N	Subset for alpha = 0.05 1	2
Scheffe[a,b]	No MBG	143	2.85	
	15 days	140	3.22	3.22
	30 days	131		3.34
	Sig.		.085	.762

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 137.808.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Means of the same column are not significantly different. Mean of no-MBG group ≠ mean of 30 days group because they are not in the same column.

P-value (or sig.value) of no-MBG group vs. 15-days group

Figure 4.3: Homogenous Subsets

4.4 Two-way ANOVA Experiment

The student thought that the results might not be the same if other product is used. He collected new data but use cloth (i.e, T-shirt) as the focal product. In the data, he created a new variable `tGood` and coded it as 0 for laptop and 1 for cloth.

Laptop is an example of search goods where consumers can fully search for information about the attributes of the product prior to purchase. In contrast, cloth is an example of experience goods where consumers can only acquire limited information without their direct experiences – to know whether or not a cloth fits you well then you have to wear it!

> **Note**
>
> The experiment above can be described in shorthand notation as a 3 (type of MBG: No MBG, 15 days, 30 days) X 2 (type of product: search good vs. experience good) between-subject ANOVA design or 3 X 2 Between-Subjects ANOVA, for short. X is read as "by". Between-subject is another term for randomization, which means that participants were randomly assigned to experimental groups. Because there are two variables being manipulated (type of MBG, type product), in general the design is referred to as a **two-way ANOVA** design.

> **Note**
>
> Because there are two variables being manipulated (type of MBG, type of product), the design is classified as a **two-way ANOVA** design.

To analyze data from this experiment, use the menu options **Analyze→General Linear Model →Univariate**. Assign `quality5` as the dependent variable and `tMBG` and `tGood` as the independent variables. It would also be useful to plot the relationship between the variables. So click on the plot tab and specify `tMBG` as the horizontal axis and `tGood` as the separate lines. Also in order to see the means for each group you need to click on the `Options` tab and check `Descriptive statistics`. The following tables provide descriptive statistics for each experimental group.

Between-Subjects Factors

		Value Label	N
type of MBG	0	No MBG	143
	1	15 days	140
	2	30 days	131
type of good	0	search good	215
	1	experience good	199

Figure 4.4: Cell frequency: This gives the number of respondents within each experimental condition

The most important table is the following table.

Tests of Between-Subjects Effects

Dependent Variable: The product of this offer would be likely to be durable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	40.266[a]	5	8.053	4.361	<.001
Intercept	4064.099	1	4064.099	2200.813	<.001
tMBG	19.262	2	9.631	5.216	.006
tGood	.700	1	.700	.379	.539
tMBG * tGood	21.719	2	10.860	5.881	.003
Error	753.427	408	1.847		
Total	4857.000	414			
Corrected Total	793.693	413			

a. R Squared = .051 (Adjusted R Squared = .039)

The previous table tests us if there are differences due to type of MBG condition and type of product under evaluations. It also tells us if there is an interaction between the two factors. We should interpret the sig-value as before, with a sig-value less than 0.05 indicating that there are statistically significant mean differences across levels of factors.

The table shows us that `tMBG` is significant (sig-value=0.006<0.05). This means that respondents have different perception regarding the durability of the products across the three levels of MBG regardless of product types. `tGood` is not significant (sig-value=0.539), which means

Descriptive Statistics

Dependent Variable: The product of this offer would be likely to be dura

type of MBG	type of good	Mean	Std. Deviation	N
No MBG	search good	3.00	1.405	76
	experience good	2.69	1.469	67
	Total	2.85	1.439	143
15 days	search good	3.31	1.380	72
	experience good	3.13	1.303	68
	Total	3.22	1.341	140
30 days	search good	2.99	1.261	67
	experience good	3.72	1.315	64
	Total	3.34	1.335	131
Total	search good	3.10	1.355	215
	experience good	3.17	1.422	199
	Total	3.13	1.386	414

Figure 4.5: This table provides the mean for each group

that respondents have similar perceptions about the product durability regardless of types of MBG offered. You can say, these situations reflect the presence of the main effect of `tMBG` but not `tGood`.

The most interesting findings from the above table is the sig-value related to the expression tMBG*tGood, which reflects the interaction effect between `tMBG` and `tGood`. The interaction effect means that the effect of `tMBG` is different across levels of `tGood`. The meaning will become clearer (hopefully!) when you see the plots of the means below.

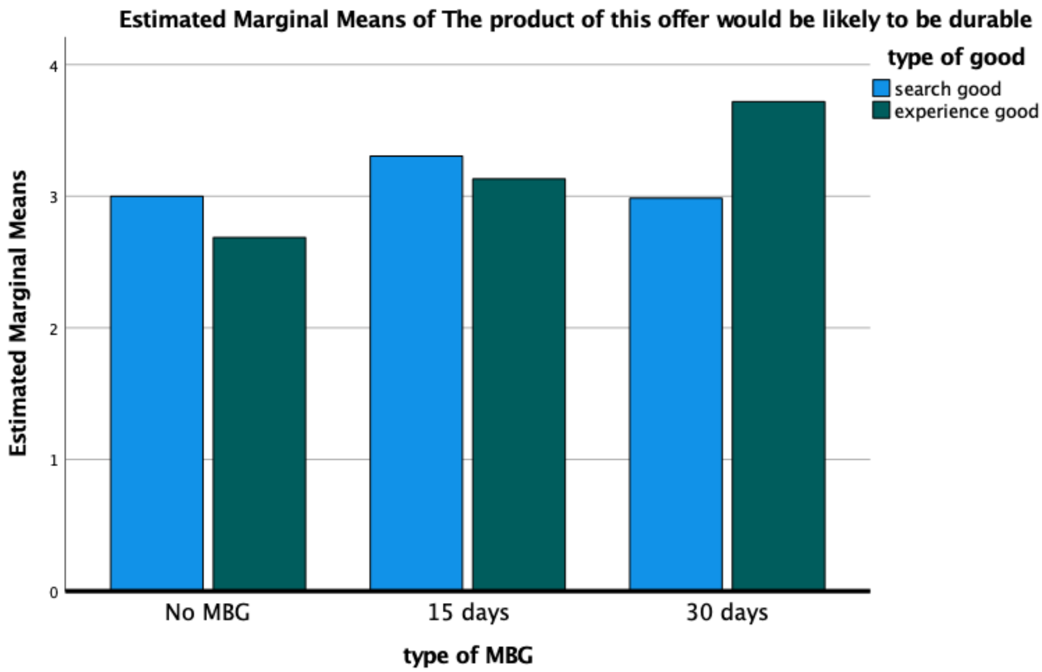


Figure 4.6: Bar chart showing means

As you can see from the figure above, in the No MBG and 15 days, the perceived durability of search good is higher than that of the experience good. But when the MBG is 30days, the perceived durability of search good is lower than that of the experience good.

Line graph is also often used to plot the interaction.

> **Task**
>
> What do you think about the potential managerial implications of the above findings?

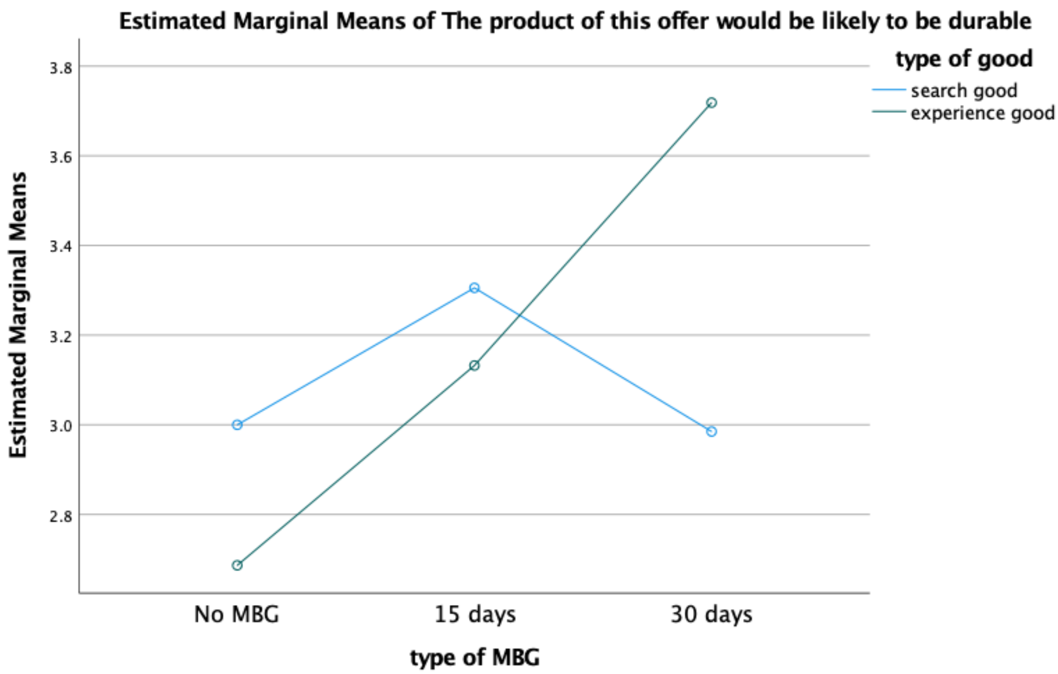


Figure 4.7: Line graph showing an interaction effect exists

4.5 Learning Experimentation from Published Research

You can learn experimentation from published research. You can even replicate results of published studies as more authors nowadays made their data publicly available as a way to increase research transparency. Open Science Forum https://osf.io facilitates such initiatives. You can click this link or this link as examples of such documentation.

4.6 Video

Lecture Week 15 on Experimentation

2023 2024

One-way ANOVA

Analyzing data from two-way ANOVA

5 Regression (Week 17)

Data: RobotGLP.sav

Data is available on 'Workshop Materials' folder on Moodle.

5.1 Learning objectives

The aim of this lab is to help you to use SPSS to conduct regression analysis, which is useful in explaining the relationship between a set of *independent* variables and a *dependent variable*.

At the end of this lab, we hope that you will be able to

- Understand the meaning of independent and dependent variables
- Select appropriate independent variables to explain a dependent variable
- Produce and interpret basic SPSS outputs for multiple regression
- Understand the meaning of multicollinearity, how to detect and remedy it

> **!** Important
>
> FYI, many students of the previous cohorts used regression when they wrote their MSc dissertations.

5.2 Why Regression

In marketing research, we often need to determine the impact of a set of marketing variables on one variable – a factor we want to understand or predict. Furthermore, among those variables we may want to find out which variables matter most, and which variables are not so important that we can ignore. Regression analysis can help us findings the answers to these questions.

In this workshop, we want to investigate factors that influence consumers' decision to join a green loyalty program – This LP encourages behaviours from hotel guests that are good for environment (e.g., reuse towels).

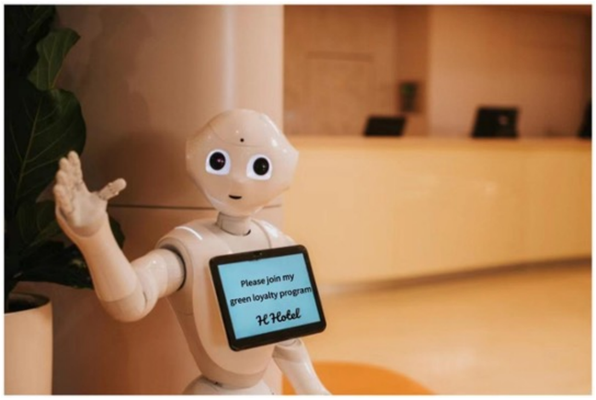


Figure 5.1: Hotel receptionist (Image courtesy of Jingxi)

In this lab, we consider the following case. A hotel manager who recently launch a green loyalty program (GLP) wants to know about factors that affect consumers' intention to join the hotel GLP. The hotel manager decides to develop a survey and asks the hotel's guests to fill in an offline survey. The survey form was handed in to hotel guests by a frontline service robot (see the image above). The survey contains items to measure the following constructs:

- Intention to join the green loyalty program
- Anticipated guilt if not joining the green loyalty program
- Perceived attractiveness of hotel receptionist
- Age
- Gender

`Anticipated guilt`, `Perceived Attractiveness`, `Age`, `Gender` are called the independent variables and `Intention to join the green loyalty program` is the dependent variable. The notion of the dependent variable comes from our prediction that its value depends on the values of the independent variables. The relationship between `Intention to join the green LP` the and the four independent variables can be written as:

Intention to join the green LP ~ Anticipated Guilt + Perceived Attractiveness + Age + Gender + error.

You can read the above expression as a consumer' intention to join the green loyalty program is influenced by the consumer's anticipated of guilt, perceived attractiveness, age, and gender, and some unknown factors represented by an error term. The relationship can be represented by a mathematical expression as below. The names of the variables are shortened.

$$Intent = \beta_0 + \beta_1 * Guilt + \beta_2 * Attract + \beta_3 * Age + \beta_4 * Gender. \tag{5.1}$$

where β_0, β_1, β_2, β_3, and β_4 are parameters that capture the impact of each of the independent variables on `intention`.

We use regression procedure in SPSS to find out the estimates for all parameters using sample data (β_0 is just a constant so it is not of our interest).

> **!** Important
>
> Variable `Intent`, `Guilt`, and `Attract` were average scores of a multi-item scale where each item in the scale were measured using a Likert scale ranging from 1 = strongly disagree to 7=strongly agree. `Age` is a continuous variable, and `Gender` is a 0,1 variable.

If you want to examine whether `Guilt` is a significant factor, then you want to test $H0$: $\beta_1 = 0$ against $H1$: $\beta_1 \neq 0$. SPSS will report the p-value associated with $H0$. If p-value is less than 0.05, you reject $H0$ otherwise retain it. If you reject $H0$, you can conclude that `Guilt` is a significant factor that influences `intent`.

If you hypothesize that `Guilt` is a significant factor and has a positive impact on satisfaction, then you want to test $H0$: $\beta_1 = 0$ against $H1$: $\beta_1 > 0$. SPSS will report the p-value associated with $H0$. You have to divide the p-value by 2. If the p-value/2 is less than 0.05, you reject $H0$ otherwise retain it. If you reject $H0$, you can conclude that `Guilt` has a significant positive influence on `Intent`.

5.3 Conducting Multiple Regression with SPSS

Open RobotGLP.sav – Thanks to Jingxi for allowing us to use a subset of her data to test the above model.

To conduct a regression analysis in SPSS, click the following: **Analyze→Regression→Linear.**

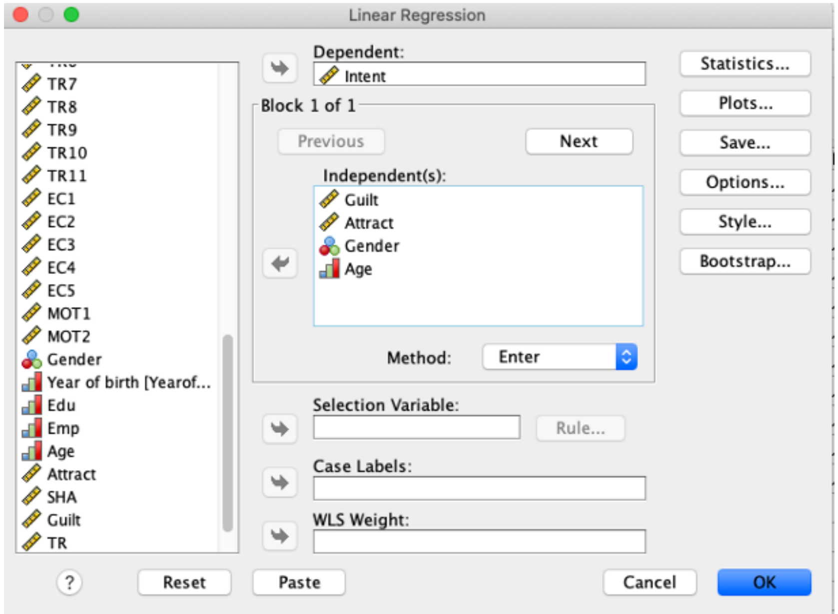
Enter `Intent` into the dependent variable box.

Enter `Guilt`,`Attract`, `Age,`Gender‘ into the independent variable(s) box.

Click `OK`

SPSS produces four tables. The first one does not offer much information. Therefore, we focus our attention to the next three tables.

1. ANOVA table: Is the model meaningful?



ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	66.437	4	16.609	18.250	<.001[b]
	Residual	135.607	149	.910		
	Total	202.044	153			

a. Dependent Variable: Intent

b. Predictors: (Constant), Age, Guilt, Attract, Gender

This table tells whether or not the model is meaningful. If the sig value is less than 0.05, then the model is meaningful. If p-value is greater than 0.05, then model should be dismissed and don't interpret other outputs. In this example, the sig. value is less than 0.05; therefore we have a meaningful model. We can proceed with the next output.

2. Model Summary table: How good is the model?

Having known the model is meaningful, how do we know whether it is good enough? The Model Summary table provides the answer to this question.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.573[a]	.329	.311	.95400

a. Predictors: (Constant), Age, Guilt, Attract, Gender

The model summary table tells you how well the independent variables explain variation in the dependent variable. The adjusted R^2 is 0.311 indicating that about 31% of the variation in the intention score is explained by the four independent variables (Guilt, Attract, Age, and Gender). Theoretically, the maximum possible value for adjusted R-square is 100% indicating a perfect model!

Adjusted R^2 is R^2 that is adjusted for the number of independent variables in the model. The more independent variable you have in the model, the larger the R^2 will be. The adjusted R^2 prevents the inflation. You can use the following convention to qualify the impact of the set of the independent variables on the dependent variable[1]:

R^2 0.02 small effect

0.13 medium

0.26 large

> **Task**
>
> Knowing the adjusted R^2, indicate the strength of the impact of Guilt, Attract, Age, and Gender.

> **Warning**
>
> Maximizing R^2 value should not be your main goal in regression. Do not be tempted to select variables with the aim of increasing R^2 value. R^2 value in social sciences are typically in the range of 0.1 to 0.5. Selecting variables should be motivated by theories.

3. Coefficients

[1] Ellis, Paul D. 2011. The Essential Guide to Effect Sizes, p.41

t = B / Std. error

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.909	.583		1.560	.121
	Guilt	.251	.059	.294	4.263	<.001
	Attract	.385	.070	.384	5.478	<.001
	Gender	.037	.173	.016	.213	.831
	Age	.038	.014	.203	2.767	.006

a. Dependent Variable: Intent

A t-value is used by SPSS to compute a p-value.

If p-value <= 0.05, the associated variable is significant (null hypothesis is rejected).

If p-value > 0.05, the associated variable is **NOT** significant (null hypothesis is not rejected).

For 1-sided test, if you used the older version of SPSS eg ver 26 or below, you need to divide the p-value by half.

This table tells us which of the independent variables significantly explain or predict the dependent variable. In this case three variables significantly explain `Intent`. These are `Guilt`, `Attract`, and `Age`, whereas `Gender` is not significant. Furthermore, the standardized beta values tell us which variable has the strongest impact on `Intent`. In this case `Attract` is the strongest predictor followed by `Guilt` and `Age`.

> **Task**
>
> If you want to make a prediction, use the unstandardized coefficients (B). If someone assigns a rating of three on `Guilt`, `Attract` can you predict her intention level? (round-off your answer)

5.4 Multicollinearity Problems

High correlation among independent variables is problematic in multiple regression because it is hard for us to determine the individual contribution of each of the independent variables in the model. For example, if `Attract` is highly correlated with `Guilt`, then we cannot conclude that `Attract` have the strongest influence on `Intent` because `Guilt` also gives contribution to the magnitude of Attract-Intent relationship (it looks stronger than it should be). In other words, the coefficients associated with the regression estimates are biased – not the same as their true population values.

The situation where an independent variable is highly correlated with another variable is referred to as multicollinearity problem. To diagnose whether multicollinearity exists, you can check it in two ways.

1. Inspect the correlation coefficient among the independent variables. To do correlation analysis in SPSS, follow this step: **Analyze→Correlation→Bivariate**. If two variables is highly correlated with the correlation coefficient is larger than 0.9, your model may suffer from a multicollinearity problem. In practice, you should be cautious, if the correlation is 0.7.

> **Task**
>
> Check the correlation coefficients among the predictors in the model.

2. Rerun your regression.

- Click: **Analyze→Regression→Linear**.
- Click `Statistics` options.
- Select `Collinearity diagnostic`. Click `Continue` and Click `OK`.
- Inspect the **VIF (Variance Inflation Factor** values in the output.

> **Tip**
>
> Multicollinearity exists if VIF (variance inflation factor) > 10.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	.909	.583		1.560	.121		
	Guilt	.251	.059	.294	4.263	<.001	.948	1.055
	Attract	.385	.070	.384	5.478	<.001	.915	1.093
	Gender	.037	.173	.016	.213	.831	.846	1.183
	Age	.038	.014	.203	2.767	.006	.837	1.195

a. Dependent Variable: Intent

> **Task**
>
> See the above output, does the model suffer from a multicollinearity problem?

5.5 If Multicollinearity Exists

What should you do if multicollinearity exists?

You can deal with it using one of the following techniques:

1. Omit one or more highly correlated independent variables
2. Create a composite variable e.g., by taking the average score of two variables if the two variables causes multicollinearity **averaging** or using Principle Component Analysis (beyond the scope of our workshops)
3. Use the model that suffers from multicollinearity but for prediction purposes only.
4. Collect more samples

5.6 Heteroskedasticity Problem

One of the assumptions in regression is that residuals or errors should be constant across any values of independent variables. This is referred to as the **homoskedasticity** assumption. The opposite of homoskedasticity is **heteroskedasticity** where residuals are not constant (i.e., heteroskedastic errors) (*learn how to pronounce these words, it took me a while to get used to them!*).

If heteroskedasticity exists, hypotheses tests about the regression parameters are not correct anymore. See lecture slide for more details. One of the recommended techniques to handle heteroskedasticity is to adjust the standard errors of the regression estimates (i.e., values in the `Std. error` column in the SPSS output).

> **Task**
>
> Rerun your regression with heteroskedasticity-adjusted standard errors and compare the results with the original regression you had previously conducted. Continue reading the text below.

To rerun your regression, install and use the 'HeteroskedasticityV3.spd' macro developed by yours truly[2]. The macro produces regression outputs with/without heteroscedasticity-adjusted standard errors.

5.7 Installing the HeteroskedasticityV3 Macro

SPSS in the Lab PCs or from cloud may not permit you to install the add-on macro because you do not have an admin right to do so. However, it is worth trying to install the macro, it

[2]Daryanto, A. (2020). Tutorial on heteroskedasticity using heteroskedasticityV3 SPSS macro. The Quantitative Methods for Psychology, 16(5), 8-20

may work! Alternatively, you need to install the macro in your personal PC where you have an admin right to install it.

You can download the macro from these websites:

- https://github.com/ahmaddaryanto/Heteroskedasticity
- Click the green button `Code`, and select `Download ZIP`.
- Go to the download folder on your PC or laptop.
- Locate the HeteroskedasticityV3.spd and install.
- If you have an admin right, double click the file, and follow the instruction on your screen
- If the above does not work, click **Extensions→Utilities→Install Custom Dialog**. Locate the file, and install.

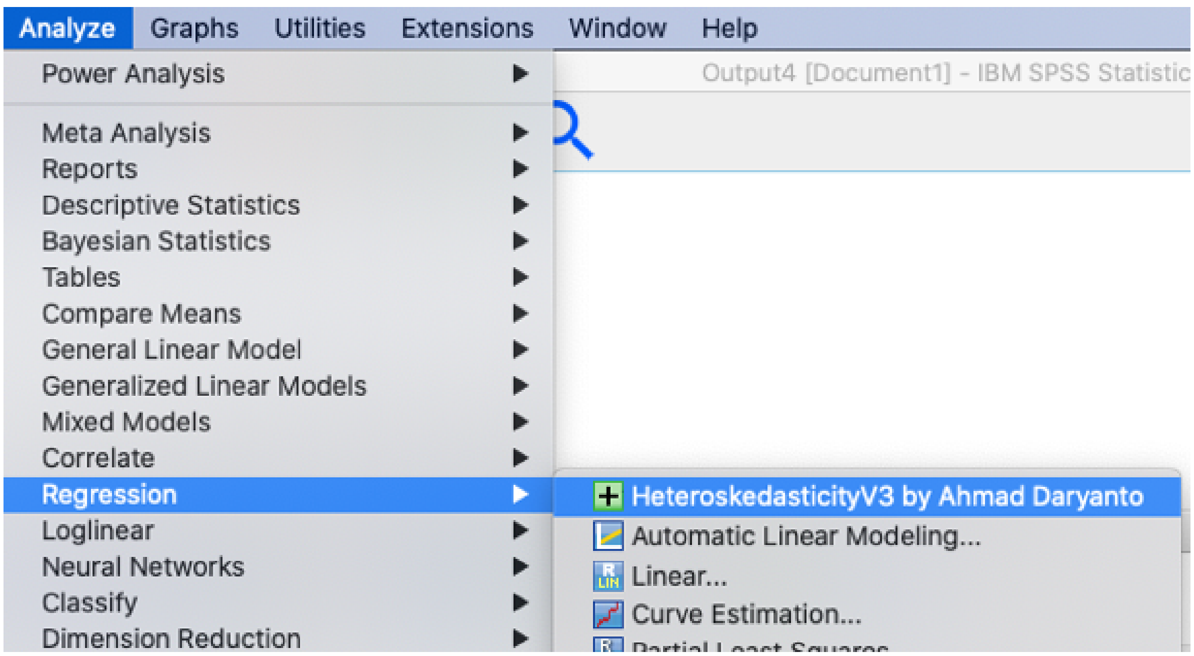


Figure 5.2: HeteroskedaticityV3 Macro in the SPSS menu after installation

If you do not manage to install the macro. Do not panic! There is another option without installation – see below:

5.8 Running the Macro without Installation

You can run the macro without having it installed on your machine. Follow these steps:

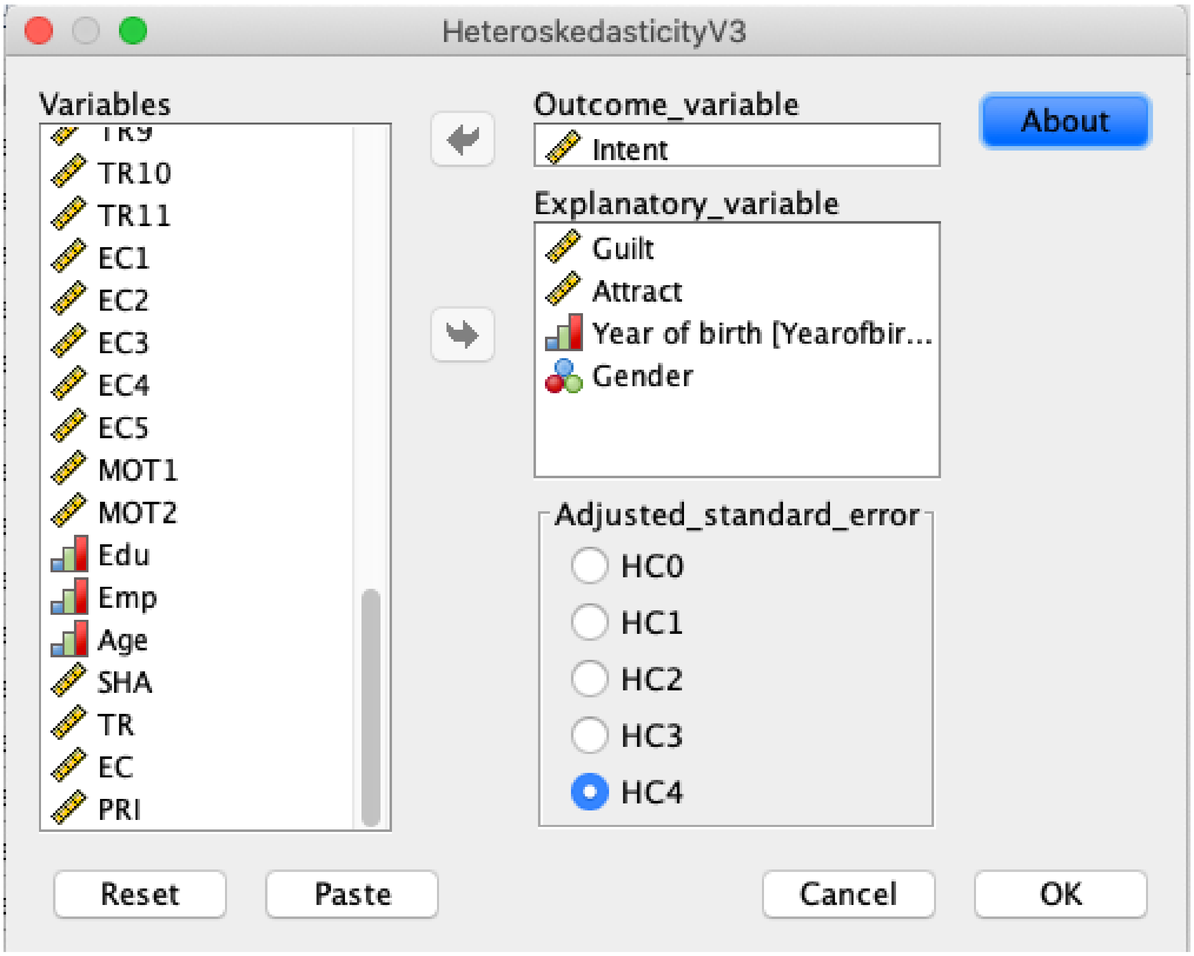


Figure 5.3: HeteroskedaticityV3 Macro

- Go back to the download folder in your PC where you have downloaded several files from this page https://github.com/ahmaddaryanto/Heteroskedasticity (see Section 5.7)
- HeteroskedasticityV3.sps was included when you download the files from the github page.
- Open the HeteroskedasticityV3.sps on your SPSS – it will be opened as a syntax file,
- Run the syntax file (highlight all the lines, run the selection button, i.e., the green button).
- Next, open this file: **Runthemacro.sps**.
- Change the DV and IVs according to your model specifications, and
- Run the file.

5.9 Video

[Lecture Week 16 on Correlation and Regression]

2023

Regression with a categorical iv

Regression diagnostics