



Setting Up Projects




Tutorial

*Time: 20 minutes
Skill level: Medium
Editions: Professional*

Learning objective(s)

- ⇒ Move questions and variables.
- ⇒ Setting different types of questions.
- ⇒ Copy variables.
- ⇒ Change the types of results you obtain by changing a question.
- ⇒ "Recode" the values in a question.

Activities

1. Open `Tutorial 7.Q` which will be in `c:\Program Files\Q\Examples` (unless Q has been installed in a different location on your computer).
2. Click on the **Variables and Questions** tab. Each row refers to a variable in the SPSS data file (in the same order as the SPSS data file).
3. Press the **Show Statistics** button () to see summary statistics – **Min**, **Mean**, **Max** and **n** – of all the variables in the data file.
4. Press  again for these to disappear.
5. The first five variables are all for administrative purposes – there is no need to use them in typical analysis. Select all of these variables and move them to the bottom of the screen by clicking  at the top of the screen (the adjacent arrow buttons are all used to move variables up or down, and to the very top or very bottom).
6. Rename the gender and age variables as `Gender` and `Age` respectively (do this by typing in the cells in the **Question** column).
7. Select the variables named `attr1_a`, `attr1_b` and `attr1_c` and create a new **Multiple Response – Categorical** question called `Attribute 1`.
8. Click on the **Tables** tab and view the summary of `Attribute 1`. You can now

see that `Attribute 1` is a bi-polar adjective scale, comparing the three brands according to whether they are exciting or boring.

9. Click on the **Variables and Questions** tab and change the name of the question to `Exciting - Boring`. This rather roundabout way of working out what to call the variable is due to the set up of the SPSS data file (i.e., the meaning of the question is not discernable from the SPSS Variable Label).
10. Click on the **Values** button for `attr2_a`. You can see that this question is measuring on a scale of Good Value to Poor Value. Click **OK**.
11. Select the variables named `attr2_a`, `attr2_b` and `attr2_c` and create a new **Multiple Response – Categorical** question called `Good Value - Poor Value`.
12. Set up the third and fourth attribute in the same way. Your screen should now look like Figure 1.

	Name	Label	Variable Type	Values	Question	Tags	Question Type	Status	Data
1	gndr	Gndr. Gender	Ordered Categorical	...	Gndr. Gender	F W H	Single Response - Categorical	OK	
2	age	Age. How old are you?	Ordered Categorical	...	Age. How old are you?	F W H	Single Response - Categorical	OK	
3	buy_a	Sloppy Joes	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
4	buy_b	Giuseppe Verdi's	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
5	buy_c	Al Romano's	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
6	buy_d	Michaels Meatballs	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
7	buy_e	Puccini's	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
8	buy_f	Zeferelli's	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
9	buy_g	Big'n'Chunky	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
10	buy_h	Fasta Pasta	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
11	buy_i	Venti Mumma's	Ordered Categorical	...	Pasta Sauce Buying - monthly or more often	F W H	Multiple Response - Binary	OK	
12	attr1_a	Giuseppe Verdi's	Ordered Categorical	...	Exciting - Boring	F W H	Multiple Response - Categori...	OK	
13	attr1_b	Michaels Meatballs	Ordered Categorical	...	Exciting - Boring	F W H	Multiple Response - Categori...	OK	
14	attr1_c	Zeferelli's	Ordered Categorical	...	Exciting - Boring	F W H	Multiple Response - Categori...	OK	
15	attr2_a	Giuseppe Verdi's	Ordered Categorical	...	Good Value - Poor Value	F W H	Multiple Response - Categori...	OK	
16	attr2_b	Michaels Meatballs	Ordered Categorical	...	Good Value - Poor Value	F W H	Multiple Response - Categori...	OK	
17	attr2_c	Zeferelli's	Ordered Categorical	...	Good Value - Poor Value	F W H	Multiple Response - Categori...	OK	
18	attr3_a	Giuseppe Verdi's	Ordered Categorical	...	Reliable - Unreliable	F W H	Multiple Response - Categori...	OK	
19	attr3_b	Michaels Meatballs	Ordered Categorical	...	Reliable - Unreliable	F W H	Multiple Response - Categori...	OK	
20	attr3_c	Zeferelli's	Ordered Categorical	...	Reliable - Unreliable	F W H	Multiple Response - Categori...	OK	
21	attr4_a	Giuseppe Verdi's	Ordered Categorical	...	Premium - Budget	F W H	Multiple Response - Categori...	OK	
22	attr4_b	Michaels Meatballs	Ordered Categorical	...	Premium - Budget	F W H	Multiple Response - Categori...	OK	
23	attr4_c	Zeferelli's	Ordered Categorical	...	Premium - Budget	F W H	Multiple Response - Categori...	OK	
24	pair_0_a	Zef. [Jar-A] Which do you pref...	Ordered Categorical	...	Zef. [Jar-A] Which do you prefer?	F W H	Single Response - Categorical	OK	
25	pair_1_a	Zef. [Jar-B] Which do you pref...	Ordered Categorical	...	Zef. [Jar-B] Which do you prefer?	F W H	Single Response - Categorical	OK	
26	pair_2_a	Zef. [Jar-C] Which do you pre...	Ordered Categorical	...	Zef. [Jar-C] Which do you prefer?	F W H	Single Response - Categorical	OK	
27	pair_3_a	Zef. [Jar-D] Which do you pref...	Ordered Categorical	...	Zef. [Jar-D] Which do you prefer?	F W H	Single Response - Categorical	OK	
28	pair_4_a	Zef. [Jar-E] Which do you pref...	Ordered Categorical	...	Zef. [Jar-E] Which do you prefer?	F W H	Single Response - Categorical	OK	
29	pair_0_b	GV. [Jar-A] Which do you pref...	Ordered Categorical	...	GV. [Jar-A] Which do you prefer?	F W H	Single Response - Categorical	OK	
30	pair_1_b	GV. [Jar-B] Which do you pref...	Ordered Categorical	...	GV. [Jar-B] Which do you prefer?	F W H	Single Response - Categorical	OK	
31	pair_2_b	GV. [Jar-C] Which do you pre...	Ordered Categorical	...	GV. [Jar-C] Which do you prefer?	F W H	Single Response - Categorical	OK	
32	pair_3_b	GV. [Jar-D] Which do you pre...	Ordered Categorical	...	GV. [Jar-D] Which do you prefer?	F W H	Single Response - Categorical	OK	

Figure 1. **Variables and Questions** tab after setting up questions

13. Click on the **Tables** tab and take the time to view tables of all the questions you have just set up. If you have experience using any other market research software, you will hopefully be impressed!

14. In the blue drop-down, select **Zef. [Jar-A] Which do you prefer?** Your screen will now be showing you the result from a graded-paired comparison question. To understand what it means, you need a bit of background about the project. This project was investigating a series of five new packaging designs. The table you are looking at is showing the degrees of preference between a key competitor's packaging and a pack called "Jar-A" which was being tested in the research. The key competitor's pack was shown on the left of the screen, Jar A on the right. As you can see if you look at the table, a score of 5 indicated that the respondent was indifferent towards the two packs.

15. Click on the **Variables and Questions** tab, select variables `pair_0_a` to

`pair_4_a`, which show the relative preferences between five packs and the competitor's pack. Set a new question with name `Comparison to competitor` of type **Multiple Response Categorical**, and click **OK**.

16. View the table for **Comparison to competitor**. You will probably note that it is a bit hard to see precisely what the pattern is. For this reason, although the graded paired-comparisons are technically categorical variables, they are most commonly analyzed by researchers as if they are numeric. Change the question type to **Multiple Response – Numeric** and review the table again. Once you have added an extra decimal place, the numbers become more readily interpretable – you can easily see that the best of the packs is pack D (as it is a 9-point scale, the midpoint, indicating equivalence to the competitor's pack, is 5.0).

The "SUM" row may give a result that might seem surprising (i.e., 21.2). This is because the SUM is literally the total of the numbers above. It is not a very meaningful number in the current example, but this is because you have "tricked" Q by pretending that the variables are numeric, when in fact they are categorical. In situations where you have genuinely numeric data (e.g., number of flights on each of a series of airlines), then the SUM becomes meaningful (e.g., total number of flights). If you want to know what the average is (which would be more common on such a chart), select "Expected" from the **Statistics – Cells** option.

17. Turn `pair_0_b` to `pair_4_b` into a **Multiple Response – Numeric** question called `Comparison to current`. View the table – again, D is the most popular pack and that all the packs have scores above the midpoint of 5.0.
18. You can also use Q to combine the two numeric questions you have created. Go back to the **Variables and Questions** tab and select `pair_0_a` to `pair_4_b`, right-click and select **Copy and Paste Variable(s) | Exact Copy**.
19. Using the new variables, set a new question called `Comparison to competitor and current` with a question type of **Two Dimensional – Numeric**. Click **Yes** on the message box. Note that the **Two Dimensional Variable Layout** dialog box will appear as shown in Figure 2, allowing you to choose whether or not the variables are ordered by rows, and the number of rows in the arrangement. The number of columns will be inferred by dividing the total number of variables by this number. Click **OK**.

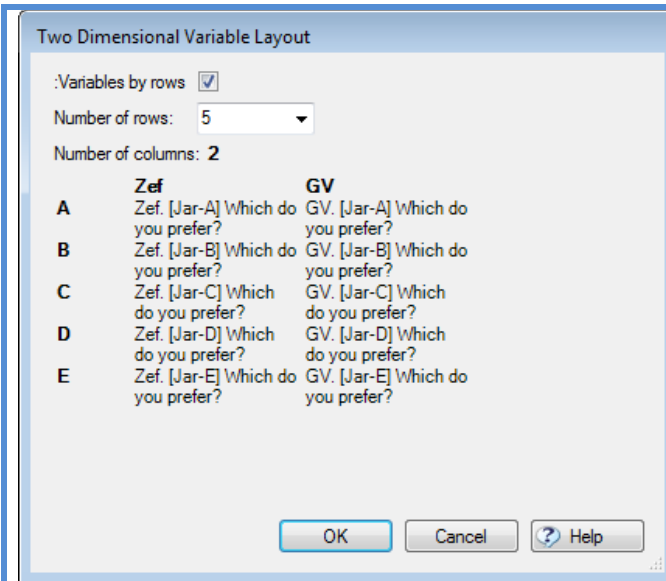


Figure 2. Two Dimensional Variable Layout dialogue box

20. Go to **Tables** and select the newly created question to view the resulting table.

The *Two Dimensional* type of question is for situations, such as this, where each variable relates to two dimensions (i.e., preference for the new pack type, A to E, and which pack it is compared against, current or competitor).

Note that now none of the numbers are in different colors. This tells us that there is no significant interaction between the rows and the columns in the table (to use the language of chi-square tests, they are independent). Stated more simply: if you look at the table, you can see that in all cases the GV number is higher than the Zef number, and the differences are roughly equivalent by rows.

In the case of the actual experiment that this data relates to, it meant that the same conclusion regarding the popularity of different packs was obtained regardless of whether the new packs were compared with the competitor's pack or the current pack. If you did have significant differences on this table, it would indicate that some of the new packs performed relatively well against the competitor, but not so well against the current packaging, and vice versa.

Importantly, a two dimensional table only tests for such interactions, as this is the usual research objective when looking at two dimensional tables (for example, the classic problem is one of whether or not there is an interaction between different brands and image statements – see *Correspondence Analysis* in the *Q Reference Manual*).

If you are interested in the more basic question of whether there is a statistical

significance between any of the averages, instead change the **Question Type** to **Multiple Response – Numeric** and select **SUMMARY** in the brown drop-down.

21. Create a **Multiple Response – Categorical** question for the fit variables (`fit_0_a` to `fit_4_a`), called `Fit`.
22. Click on the **Values** button. You will see that the values do not match up to the labels! You can type over the top of the numbers in the **Value** column, and the numbers we enter will be used in computing the average. Replace the number 1 with 0, 2 with 1, 3 with 2, ..., and 11 with 10.

Replacing one value with another is known in most statistical packages as *recoding* (e.g., in SPSS, this is an example of Recode into same Variable). Q differs from most other packages in that while it performs the recoding, it retains the original values. Similarly, you can type over the labels for different values but Q retains the original labels. Tick the **Show source values and labels** check box to see the default values.

When you are doing anything relating to the values or value labels for a variable which is a part of a question, it is important to select all the other variables that make up that question unless you only want to make the change to a selected variable (which will be rare). When editing values by right-clicking on a table in the tables tab, all constituent variables that make up the selected question are edited at once

23. Click **OK**.
24. Change the name of the question to `Fit (out of 10)` and change the question type to **Multiple Response – Numeric**.
25. Click on the **Tables** tab and view a summary of **Fit (out of 10)**. The table should look like Figure 3 (you may need to add a decimal place). As none of the numbers are shaded, this tells us that there is no significant difference between the packs.

Average	Average
A] Giuseppe Verdi's	6.6
B] Giuseppe Verdi's	6.0
C] Giuseppe Verdi's	6.5
D] Giuseppe Verdi's	6.6
E] Giuseppe Verdi's	6.0
SUM	31.7

Figure 3. Comparison of means

26. Select **Fit (out of 10)** in the brown drop-down (i.e., you will have the same question in both the rows and columns of the table). Your screen should look like Figure 4. You have now created a *correlation* matrix, which shows the strength of the relationship between each measure of pack fit.

Correlation	A] Giuseppe Verdi	B] Giuseppe Verdi	C] Giuseppe Verdi	D] Giuseppe Verdi	E] Giuseppe Verdi	SUM
A] Giuseppe Verdi	1.0	.5	.4	.3	.4	.7
B] Giuseppe Verdi	.5	1.0	.2	.3	.2	.6
C] Giuseppe Verdi	.4	.2	1.0	.3	.4	.7
D] Giuseppe Verdi	.3	.3	.3	1.0	.2	.7
E] Giuseppe Verdi	.4	.2	.4	.2	1.0	.7
SUM	.7	.6	.7	.7	.7	1.0

Figure 4. Correlation matrix

As all the values in the table are positive, this tells us that people who considered one pack to have a good fit were, all else being equal, more likely to think that all the packs had a good fit (not a particularly interesting finding). Note that there is a diagonal line of correlations, called the *main diagonal*, of 1.0s going down the table. This is because, by definition, a variable has a positive correlation with itself of 1.0.

Significance tests of correlations have the *null hypothesis* that the correlation is 0.0 (i.e., that there is no relationship between the variables). In the example shown in Figure 4, all the cells are blue because all the correlations are greater than this benchmark value of 0.0. To see if there are significant differences between two correlations you need to use a *planned test*.

27. Select **Comparison to current** in the blue drop-down and ensure two decimal places are selected. The table now shows the correlations between two questions (see Figure 5).

Comparison to current						
Fit (out of 100)						
Correlation	A] Giuseppe Verdi	B] Giuseppe Verdi	C] Giuseppe Verdi	D] Giuseppe Verdi	E] Giuseppe Verdi	SUM
A] Which do you prefer?	.19	.02	.38	.18	.25	.30
B] Which do you prefer?	.05	.24	.24	.12	-.03	.18
C] Which do you prefer?	.14	-.01	.51	.20	.22	.32
D] Which do you prefer?	.11	.03	.24	.67	.17	.39
E] Which do you prefer?	.07	-.01	.22	.15	.46	.27
SUM	.16	.08	.45	.38	.30	.42

Figure 5. Correlations between two questions

The pattern of correlations shown in Figure 5 is marginally more interesting than the earlier correlation matrix. If you look at the top left number, 0.19, it indicates that there is a (weak) correlation between preference for pack A (rows) and fit of pack A (columns). That is, it suggests that the better a consumer considered the pack to fit with the brand, the more likely they were to consider the pack to be superior to the current pack.

If you look at the cell immediately below, it also has a correlation of greater than 0.0 (i.e., .05), but it is not significant. If you look at the rest of the table you will see that in all cases, the correlations in the *main diagonal* are equal to or higher than the other correlations in the same column. That is, as a general rule, the higher the fit score a respondent gave to a pack, the more likely they were to prefer the pack. And, in the case of Pack D, this relationship is at its strongest.