Chapter 12 / Capítulo 12

Evidence-based biomedicine: methodology for research, standardization, and scientific procedural aspects (English Edition)
ISBN: 978-9915-704-43-2
DOI: 10.62486/978-9915-704-43-2.Ch12
Pages: 63-71

©2025 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

SPSS

In this chapter, we will first provide a brief introduction to the essential fundamentals required to perform statistical analysis. These concepts include creating databases, defining random variables, and transforming and reorganizing variables. Next, we will provide the necessary instructions to perform the analyses discussed in the previous topic.

12.1. Basic concepts of SPSS

SPSS (Statistical Package for Social Science) is widely recognized as one of the most popular statistical programs due to its user-friendly interface. The SPSS interface consists of a series of cells, where each row corresponds to an individual or patient and each column represents a variable we have measured and plan to analyze later.

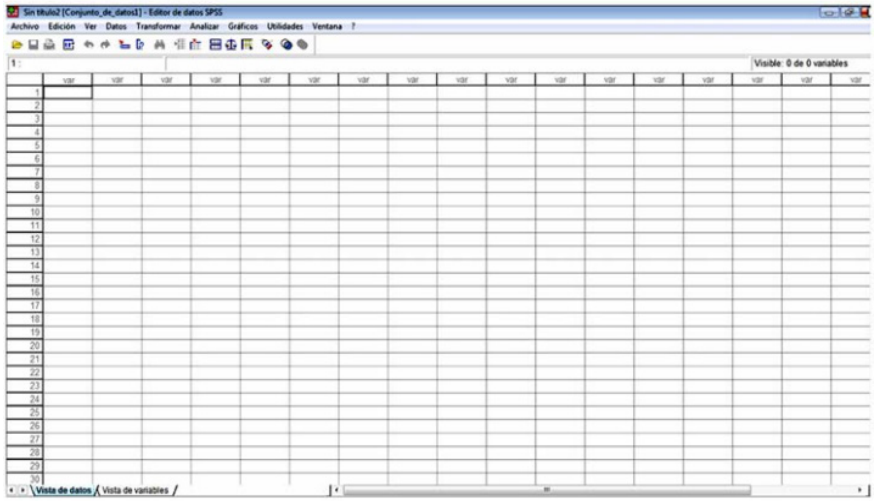

Figure 12.1. Preliminary window

12.2. Data reading

To open a data file in SPSS, we can follow these steps: In the main SPSS menu, select “File,” then “Open,” and then “Data.” We have the option of opening a file directly in .sav format, which is the native data format for SPSS, or in .xls format, which is an Excel format.

12.3. Data editor

In the main SPSS window, at the bottom left, there are two tabs: “Data View” and “Variable View.” The “Data View” shows the database we are working with, while the “Variable View” allows us to modify variable names and types used in our analysis.

Within the “Variable View” tab, you can edit the names of the variables you want to appear in the results. It is advisable to avoid using spaces and special characters in variable names.

In the “Type” box, we find a box with three ellipses. Clicking them opens a window that lets

us select the type of variable we are working with and define the cell length and the number of decimal places we want to use. By default, SPSS usually sets the length to 8 and the number of decimal places to 2.

Figure 12.2. Variable type

In the "Label" box, we can enter a description that helps us identify the variable more clearly. This description is useful for our own recognition of the variables, since the results will show the name we have assigned to each variable.

In the "Values" section, we again find a dialog box that, when selected, opens a window of value labels. These labels are useful when we want to use descriptions instead of numbers to identify the values of a variable. For example, in the case of the variable "Gender," instead of using the numerical values 1 and 2, we can use labels that clearly indicate that 1 represents men and 2 represents women.

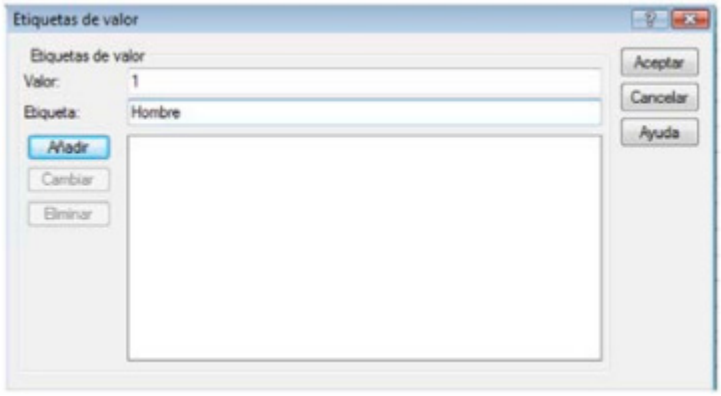

Figure 12.3. Value labels

12.4. Data transformation

Sometimes it is necessary to create a new variable from existing database variables. To do this, we can use the "Transform" function and select the "Calculate variable" option. In the "Destination variable" field, we will specify the name of the new variable we want to generate, and in the "Numeric expression" box, we will enter the formula that will be used to perform the calculation.

12.5. Recoding

When we need to group or recode values in a variable, we use the "Recode" function in SPSS. This can be done in two ways: "Recode in the same variables": This option allows us to modify the values of an existing variable without creating a new one. To do this, select the variable you want to change, then click the "Recode in the same variables" option. Next, in the "Old and New Values" section, specify the original values you want to recode and the new values you want to assign. For example, if you have a 'weight' variable and want to recode it into intervals, select the "Range" option and define the desired intervals. For example, we could recode people with a weight between 50 and 55 into a single interval and assign them a new value, while leaving the other values unchanged. Then, we click on "Add" and follow the process.

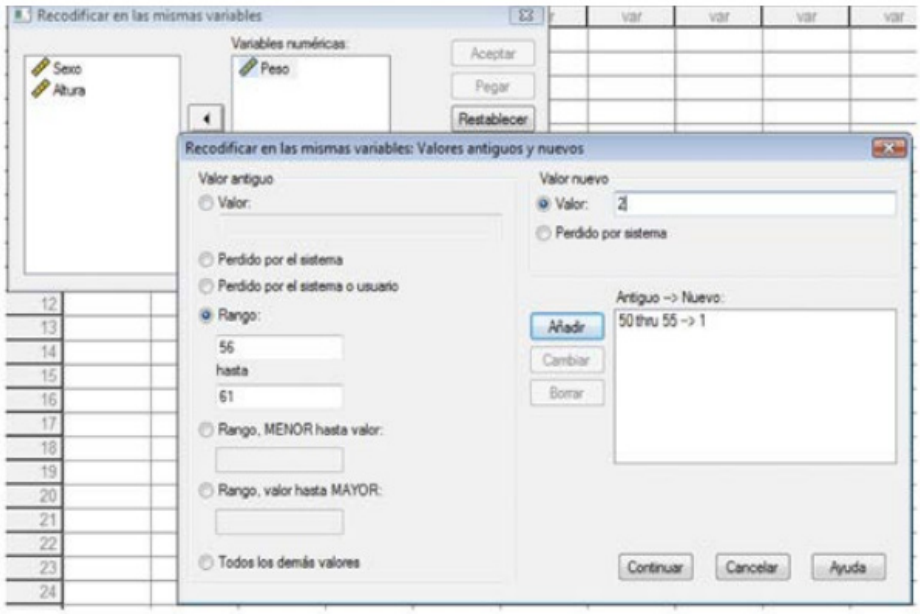

Figure 12.4. Recoding in the same variables

When we need to change the values of a variable and want the resulting variable to be different from the original, we use the "Recode into different variables" option in SPSS. In this case, we are creating a new variable, so we are asked to provide a name for the result variable. It is essential that, after specifying the new variable's name, we click the "Change" button. Then, the same recoding window used to define the old and new values will open, allowing us to perform the recoding as explained above.

12.6. Data filtering

In some situations, we may want to focus only on a specific part of our database. To do this, we can use the "Select Cases" function in SPSS. To do so, we select "Data," then "Select Cases," and then "If meets condition." For example, if we want to analyze the variables only for people over 15 years of age, we can set the condition "Age>15." If, in addition to being over 15 years of age, we want them to be female, we can combine both conditions using the "&" operator and express it as "Age>15 & Sex=2," as shown in the figure below.

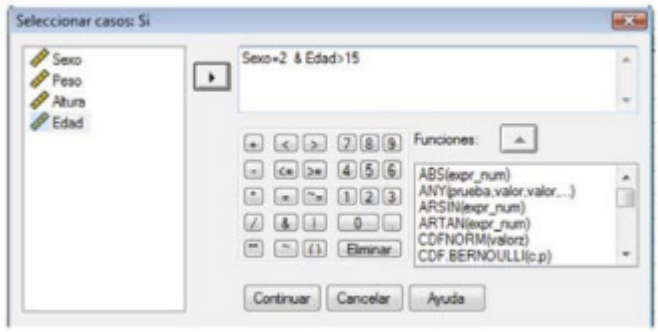

Figure 12.5. Case selection

12.7. Descriptive statistics in SPSS

To perform a descriptive statistics analysis in SPSS, we can follow these steps: Go to the "Analyze" option, then select "Descriptive Statistics" and then choose "Frequencies." In this section, select the variable you want to work with and then click on "Statistics."

In the statistics window, we can find all the measures of central tendency, dispersion, and position available for the selected variable. To calculate specific percentiles, select the "Percentiles" option and enter the desired value in the corresponding box. For example, if we want to calculate the 50th percentile, we select 'Percentiles' and type "50" in the adjacent box.

To generate graphical representations in SPSS, we can access the "Graphs" option and select "Old dialog boxes." Doing so displays all the graph options available in SPSS.

Figure 12.6. Descriptive statistics

12.8. Inferential statistics in SPSS

To calculate the confidence interval for the mean in SPSS, follow these steps: Go to the "Analyze" option, then select "Descriptive Statistics" and choose "Explore." In the 'Explore' window, select the quantitative variable of interest in the "Dependent" section. In the "Statistics" section, you can specify the desired confidence level, which is set to 95 % by default.

To perform a Chi-Square test in SPSS, as in the case of a study on the relationship between lung cancer and smoking, where the null hypothesis is that the incidence of lung cancer is the same in smokers and non-smokers, you can follow these steps: Go to "Analyze," then select "Statistics" and choose "Contingency Tables." Next, in the 'Statistics' section, select "Chi-Square." This will allow us to perform the Chi-Square analysis to evaluate the relationship between the two variables, in this case, smoking and the incidence of lung cancer.

Tabla de contingencia Cáncer * Fumador

Recuento

		Fumador		
		Si	No	Total
Cáncer	Si	11	2	13
	No	1	11	12
Total		12	13	25

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	14,547[b]	1	,000		
Corrección por continuidad[a]	11,652	1	,001		
Razón de verosimilitudes	16,571	1	,000		
Estadístico exacto de Fisher				,000	,000
Asociación lineal por lineal	13,965	1	,000		
N de casos válidos	25				

a. Calculado sólo para una tabla de 2x2.
b. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 5,76.

Figure 12.7. Contingency

As a first step, SPSS provides a contingency table that shows the relationship between the two variables. The observed values correspond to the figures in each cell of the table. In addition, SPSS allows us to calculate percentages relative to each cell.

In the following table, we can see the Chi-Square statistic (14,54) and the p-value for this contrast (0,00). Fisher's exact statistic is also shown, providing p-values for both the one-tailed and two-tailed tests (both equal to 0,00). In general, Fisher's statistic is more accurate than the Chi-Square test and can be used in all cases, unlike the Chi-Square test, which is only appropriate when the expected frequencies are greater than 5.

Since the p-value is less than 0.05 (indicating bilateral significance under Fisher's exact test), we would have reason to reject the null hypothesis (H0) and conclude that there is a difference in cancer incidence between smokers and non-smokers. It is important to note that although two significance values are provided for Fisher's test (one-tailed and two-tailed), we are performing a two-tailed test in this case.

The McNemar test is a nonparametric test used to compare two related samples. To perform this test in SPSS, we can follow these steps: In the "Statistics" section, select the "McNemar" option. When applying this test, we will obtain a contingency table and another table with the p-value (exact significance (two-tailed)). Alternatively, we can also obtain this test by selecting

"Statistics," then "Nonparametric Tests," and finally "2 Related Samples." Next, we choose the two variables we want to compare.

To perform a Student's t-test in SPSS, it is important to verify two aspects of the quantitative or dependent variable:

Normality: If the sample size is 30 or fewer, we can assess normality using the Kolmogorov-Smirnov test. This is done by selecting "Analyze," then "Nonparametric Tests," and "1-Sample Ks." If the p-value (two-tailed) is greater than 0,05, we have no evidence to reject the hypothesis that the variable follows a normal distribution. Otherwise, we could consider transforming the variable or applying a nonparametric test instead of the Student's t-test.

Homoscedasticity (only for two independent samples): When comparing two independent samples, it is necessary to verify homoscedasticity. To do this, we use Levene's test.

These steps will help us determine whether we can apply the Student's t-test appropriately or whether we need to consider non-parametric alternatives if the assumptions are not met.

To perform a Student's t-test for independent samples in SPSS, assuming that the dependent variable meets the assumption of normality, we can follow these steps:

We access the "Analyze" option, then select "Compare Means" and finally "Independent Samples t-Test."

In the "Compare Variables" box, we select the dependent variable we want to analyze, and in the "Grouping Variable" box, we enter the dichotomous independent variable. For example, if we want to compare cholesterol levels between men and women, the cholesterol variable will be placed in the "Compare variables" field, and the sex variable will be entered in the "Grouping variable."

The program will provide results in a table titled "Independent samples test." This table will include information on the homogeneity of variances, evaluated using Levene's test. If the p-value is less than 0,05, we conclude that the variances are not equal. Otherwise ($p>0,05$), we assume homoscedasticity of variances.

Depending on the results of Levene's test, we will look at the first row (if we assume equal variances, $p>0.05$) or the second row (if we do not assume equal variances, $p<0,05$) in the results table.

The table will present the comparison of means between the two independent populations under study, in this case, men and women.

If we find that the p-value is less than 0,05 (e.g., 0,002), this indicates significant differences in the means of the two populations.

A relevant result is the confidence interval for the difference in means, which provides information about the direction and magnitude of the difference. A wider interval indicates lower statistical power in the study.

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Peso	Se han asumido varianzas iguales	,338	,577	4,375	8	,002	17,58333	4,01873	8,31613	26,85054
	No se han asumido varianzas iguales			4,654	7,762	,002	17,58333	3,77804	8,82444	26,34223

Figure 12.8. T-Student for independent samples

To perform a Student's t-test for paired samples in SPSS, let's assume we want to evaluate the differences in weight before and after following a diet. The procedure is like that for independent samples, but in this case, we select "paired samples." Next, we will obtain a table with statistics for the samples and the test results.

The table will provide the mean of the differences, the standard deviation, a confidence interval indicating the study's power, and the p-value. If the p-value is less than 0.05, this indicates that the mean weight difference before and after the diet is significant.

The Mann-Whitney U test is an alternative to the Student's t-test for two independent samples, especially useful when the sample size is less than 30 or when the variable is ordinal rather than quantitative. For example, it can be used to compare the effectiveness of two diets (1 and 2) for weight reduction, where the weight reduction variable is coded into three categories: 0 (no weight reduction), 1 (moderate reduction), and 2 (high weight reduction).

To perform the Mann-Whitney U test in SPSS, go to the "Analyze" option, then select "Nonparametric Tests" and finally "Two Independent Samples." When performing this test, we will obtain both one-tailed and two-tailed significance results.

If our objective is to determine whether there are differences between the two groups, attention should be paid to two-tailed significance. It is important to note that, unlike parametric tests, this test does not yield a confidence interval.

The Wilcoxon Test for related data is an alternative similar to the Student's t-test for comparing two related samples. It is particularly used when the sample size is less than 30 or when we are working with an ordinal rather than a quantitative variable.

To perform the Wilcoxon test in SPSS, go to the "Analyze" option, then select "Nonparametric Tests" and then "Two Related Samples." This test provides adequate results when these conditions are met and is useful for comparing two groups of related data, such as before and after a treatment or intervention.

Analysis of Variance (ANOVA) is used when we want to compare the means of multiple groups, such as to determine whether there are differences in hypertension levels among patients who have received five different treatments. This analysis is used when comparing more than two samples.

In SPSS, the procedure for performing an ANOVA is as follows: go to "Analyze," then select

"Nonparametric Tests" and then "One-Way ANOVA." Among the available options, we can choose the "Homogeneity Contrast" option if we assume equal variances across groups, which is a necessary assumption for applying this test.

If we cannot assume that the variances are equal, within the "One-way ANOVA" options, we can select Welch's t-test or Brown-Forsythe's test. These tests are appropriate when the variances are not homogeneous between groups.

The Kruskal-Wallis test is a nonparametric alternative to ANOVA, especially useful when the sample size is less than 30 or when the variable is ordinal rather than quantitative.

To perform the Kruskal-Wallis test in SPSS, we can access it through the "Analyze" option, then select "Nonparametric Tests" and then "Comparison of k Independent Samples." This test is appropriate when we want to compare the means of several independent samples rather than using a parametric ANOVA.

To perform Pearson and Spearman correlation analyses in SPSS, you can follow these steps: go to "Analyze," then select 'Correlations' and "Bivariate." Next, choose the variables you want to analyze. To calculate Pearson's correlation coefficient, SPSS has it preselected by default; to calculate Spearman's correlation coefficient, you must check this option.

Once the analysis is complete, SPSS will provide similar results tables for both correlation coefficients. These tables will include the coefficient values and p-values for each comparison. It is important to note that when comparing a variable with itself, the coefficient value will always be 1, and in this case, SPSS will not calculate the significance value