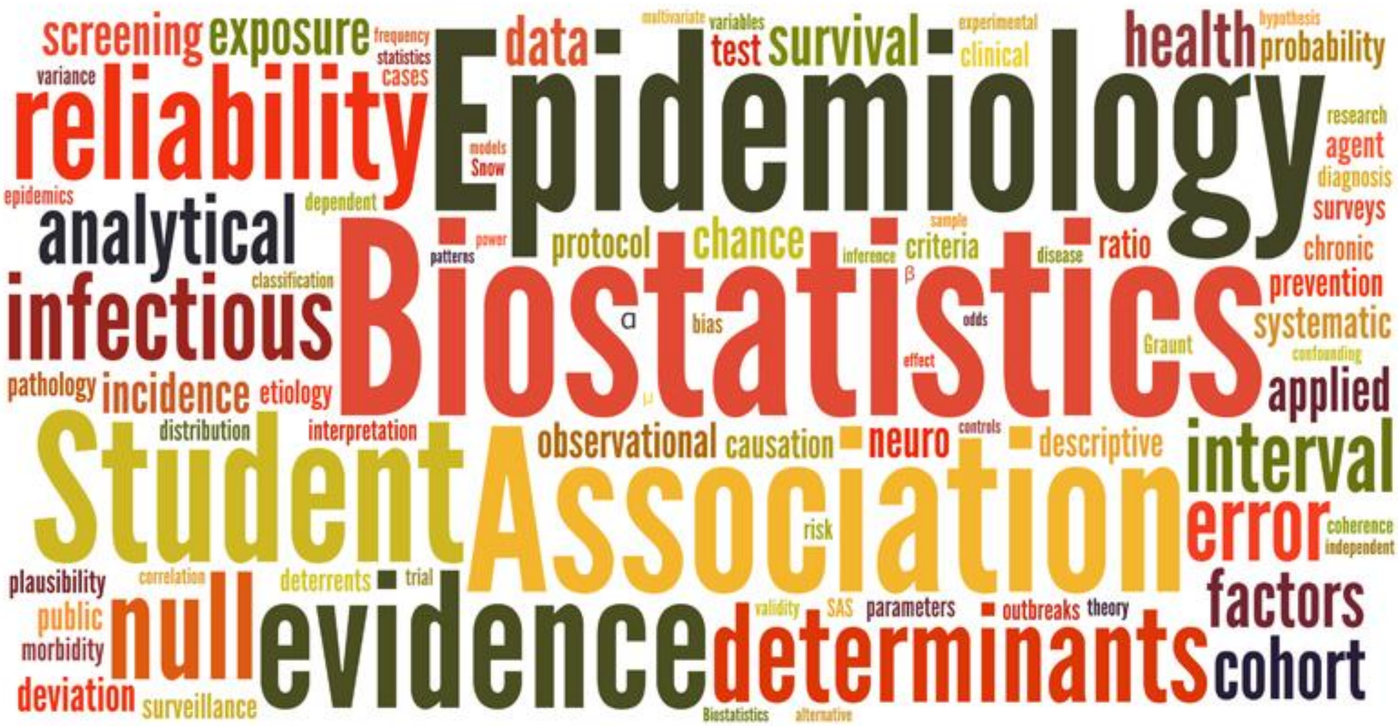


☑ Sheet (Notes) ☐ Slide ☐ Handout

Number: 11

Done By: Dana Rida

Subject: Understanding data and Introduction to SPSS

Doctor: Mohammad Nassar

Today's lecture is about how we present our data in a useful way.

this topic is not mentioned in the book, so study it from the slides and notes SPSS (a software, we don't need to know how to work on it, just know how it works): most commonly used nowadays especially in medicine , we enter the data then the SPSS will utilize it and use calculations to get what you want, and present the data in different ways, but any software used in a wrong way will give us wrong results.

We have mainly 3 spread sheets : data view sheet, variable view sheet and outcome view sheet. Usually in the SPSS we begin with the variable view sheet. The most important thing you apply in the variable sheet it automatically goes to the data view sheet to introduce your subjects.

Example

A research that compares between 3 different medications(A,B,C) to lower blood pressure and improving heart rate compared to placebo group, in each group there were 7 patients randomly selected from (A,B,C) and 7 selected from the placebo group , so when entering the data we need to know 2 things; name of the variable and the values of this variable.

Value	Label
1.00	drug A
2.00	drug B
3.00	drug C
4.00	placebo

The outcome can be anything you choose to study (heart rate, blood pressure etc,,) in the SPSS it doesn't use the word interval or ratio instead it uses the word scale or older versions "continuous level"

In the example I want to measure heart rate(hr) , so it is on the continuous level(highest level of measurement). From this level we can convert it to all types of levels, to nominal by choosing a cut-off point, or ordinal. Gender was coded 1 for males and 2 for the females which means it is nominal dichotomous.

(variable view) notice the measures it shows each type.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
interven	Numeric	9	2		{1.00, drug ...	None	8	Right	Nominal	Input
hr	Numeric	8	0		None	None	8	Right	Scale	Input
wt	Numeric	8	2		None	None	8	Right	Scale	Input
age	Numeric	8	2		None	None	8	Right	Scale	Input
gender	Numeric	8	2		{1.00, Male}...	None	8	Right	Nominal	Input
one	Numeric	8	2		None	None	8	Right	Scale	Input
two	Numeric	8	2		None	None	8	Right	Scale	Input
three	Numeric	8	2		None	None	8	Right	Scale	Input
four	Numeric	8	2		None	None	8	Right	Scale	Input

Once you completed the entering of the variables, they will be viewed in the data sheet as columns. The columns are variables and the *rows* the subjects and you start entering data.

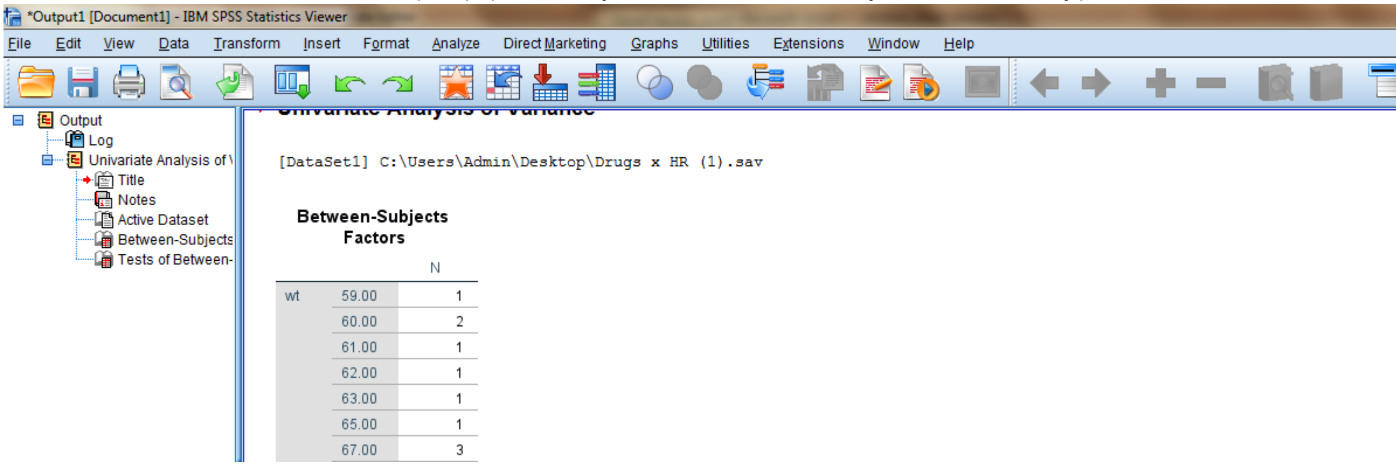
hr	wt	age	gender	one	two	three	four
70	62.00	65.00	1.00	1.00	.00	.00	.00
48	100.00	58.00	2.00	1.00	.00	.00	.00
92	60.00	71.00	1.00	1.00	.00	.00	.00
84	82.00	45.00	1.00	1.00	.00	.00	.00
66	88.00	60.00	1.00	1.00	.00	.00	.00
54	59.00	69.00	1.00	1.00	.00	.00	.00
90	88.00	90.00	1.00	1.00	.00	.00	.00

So in variable sheet I set the variables, and their values, then they will be converted to columns and rows to insert my data.

after that we begin with univarate analysis.

Univariate analysis

You check you variables one by one ,Manually check every data you introduce, for example you have entered the data for heart rates but then you got a heart rate=1200 so the researcher wrote it 1200 instead of 120 by mistake, to discover these kinds of mistakes we do univariate analysis. We go to the analysis part in the SPSS, we introduce the heart rate and the weight for example , it will open a new spread sheet named output, as we can see we will have the heart rate in frequencies, which shows us the minimal heart rate (59) and the maximum heart rate (67) (if I only took for first 7 patients only)


[DataSet1] C:\Users\Admin\Desktop\Drugs x HR (1).sav

Between-Subjects Factors

		N
wt	59.00	1
	60.00	2
	61.00	1
	62.00	1
	63.00	1
	65.00	1
	67.00	3

but while inspecting your data you got heart rate 3 or 1200 that wouldn't make sense, you go back to your data and correct it and do the analysis again. The point of this univariate analysis is to avoid bigger problems later on, these mistakes will continue with you to the inferential statistics and will be reflected on the decisions you take especially if those variables were on the continuous level ; because the means are effected by extremes (e.g 30 students all got marks above 90 but only 5 got marks below 30 this means that those 5 will effect on the 30 students and pull the mean down with their marks, vice versa) so if my data had any mistake it will have a great effect.

Presenting data

The most important way is Tables. All tables should be presenting my data.

Example

From this table we can know a lot on information like how many patients developed hematoma and received anti-coagulant? 6 patients. Another information , how many patients got hematoma and didn't receive anti-coagulant ? 0 . how many patients didn't develop hematoma and received anti-coagulant? 31. As simple as that. So just from 4 cells we got many conclusions. But this table doesn't give me information about other things ,like the weight.

Table (1): Factors associated with the development of hematoma. (n = 66)

Factors	Hematoma Yes(n=6)	Hematoma No(n=60)	P value
Prophylactic anticoagulants (Heparin)			
Yes	6	31	
No	0	29	0.031*
Smoking			
Yes	0	14	0.335
No	6	46	
Medical illness			
Present	1	15	0.548
Not present	5	45	
Menstrual phase			
phase I (periovulatory)	1	26	0.38
phase II (perimenstrual)	4	28	0.42
phase III (amenorrhea)	1	6	0.45

* Significant using Fisher's exact test at α=0.05, 2-tailed

Instead the second table can give me all the info I need about the weight(total mass). And do a lot of calculations (means, median etc,)

Table (2): Variables associated with hematoma post breast reduction.

Variables	Hematoma Mean (SD) (n=6)	No Hematoma Mean (SD) (n=60)	P Value
Age (Years)	39.83 (6.8)	34.75(8.8)	0.175 [a]
Total mass (gm)	3186.67(932.4)	1829(969.6)	0.002 *a
Left breast mass (gm)	1585(433.3)	868.1(449.4)	< 0.001 *a
Right breast mass (gm)	1601.6(502)	940.3(469.4)	0.002* a
Hospital stay(day)	6.3(1.2)	4.7(1.6)	0.02 *a
Duration of operation(hr)	3.4(0.66)	2.8(0.66)	0.05 *a
Estimated blood loss(ml)	566.6(87.5)	327.5(118.7)	< 0.001 *a
Low SBP first period (mmHg)	86.1(5.49)	88.6(8.5)	0.48 [a]
Low SBP second period (mmHg)	81.3(5.3)	89.8(9.5)	0.03 *a
Mean BP first period (mmHg)	63.6(1.2)	71.2(5.8)	0.002 *b
Mean BP second period (mmHg)	62.3(2.05)	75.1(6.9)	< 0.001 *b
Peak SBP first period (mmHg)	97.1(4.02)	107.8(10)	0.012 *a
Peak SBP second period (mmHg)	90.8(3.7)	115.9(13.6)	< 0.001 *b
BMI (4 International Categories)	50.83 (mean rank)	31.7 (mean rank)	0.018 *b
Hemoglobin preoperative (gm/dl)	12.6(0.79)	12.4(1.17)	0.663 [a]
Hemoglobin postoperative (gm/dl)	9.3(1.16)	10.3(1.1)	0.03 *a

* Significant at α=0.05, 2-tailed
a t-test

Another important topic not in the book :

PX = y

if we say P 75 of the students marks is 80 is it a good or bad indicator? This means that 75% of the students got a mark less than 80 , which is not that good. If we said P90 =60 then this means 90% got marks less than 60 also a bad indicator. What if P30= 90 this means 30% got marks less than 90 and the other 70% all above 90 which is a good indicator. This shows how your data is variable. if the data was much variable there will be a lot of confounding variables in the relation between the dependent and independent variables and effect their relations , for example I'm comparing two groups; one group uses a certain drug and a placebo group, the age could be a confounding variable if the placebo group were all old people and the other group young, the results might not be related to the drug I used instead to the age difference, * There are four commonly used measures of variability: range, mean, variance and standard deviation, we need to know the relationships between these variables not only calculations..*

in the SBSS it gives you the numbers till 16 digits so it is important to round your results.

Charts

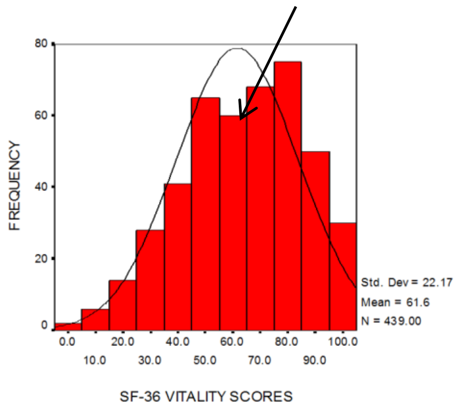
We have different types of charts; Bar chart divided to simple bar chart, clustered bar chart, stacked bar chart, horizontal bar chart, other than bar chart we have pie chart, histogram and box plots.

In histogram and box plots they can give us more than one result at the same time.

in simple bar chart I will work only on one variable, but in the clustered I will be working on more than one.

Histogram is used mostly in epidemiology, ***usually in bar chart and pie chart we describe nominal and maximum ordinal data*** , so in representing nominal data use a bar or pie chart, but when it is continuous then use a histogram, box plot or line chart (each line represent data and numbers)

The difference between bar chart and histogram ; in the level of measurements, number of bars can be up to 30 in the histogram, I can get the mean and standard deviation etc. in the histogram which is something I can't get in the bar chart, histogram gives us a normal curve we can know the skewed from it, which means I can compare my results (bars) with the normal curve (the line), as I can notice here there is a positive skewness(explained more in upcoming lectures) at the point 60(the bar) away from normality(the line above it) the 60 should be skewed to the right a little bit so it would be within the normal distribution.



Polygon; the number of bars less than histogram..