

Community medicine *Biostatistics activity*

Activity
Statistics: Basics – Epidemiology & Biostatistics - Video 2-6
Learn how to code your data in SPSS - Video 7-9
Descriptive Statistics Examples and SPSS analysis - Slides 79
Running and Interpreting Chi-Square on SPSS - Video 10-13
Running and Interpreting an Independent-Sample t-test on SPSS - Video 14-16
Running and Interpreting Dependent (Paired) Sample t-test on SPSS - Video 17-18
Biostatistics Activity MCQ 19

This file includes all the lectures and slides. You can rely on it completely, The links of lectures have been placed at the end of each lecture if you want to watch it, Don't worry, this file is very simple, studying it might take less than two hours.



Epidemiology and Statistics

Epidemiology and statistics are closely related, but they are not the same. Statistics serves as a tool used by epidemiologists to analyze data and derive conclusions.

Data vs. Information

• <u>Data</u> : are raw numbers or observations that, by themselves, may lack meaning. For example:

53, 61, 62: Without context, these numbers don't provide meaningful information.

• <u>Information</u>: When data are associated with context, they transform into information. For instance, if these numbers represent the ages of three individuals in 2015, they now have meaning: they are the ages of people.

What is a Variable?

• A variable is a mathematical value that can change within a specific situation or scope of a problem.

In mathematics, a variable represents a value that changes within a problem.

In research, a variable can represent characteristics like age, gender, or location.

In computers, a variable might store a number or string of text.

In epidemiology, we often categorize variables as exposures (independent variables) and outcomes (dependent variables). For example:

If we're examining the relationship between smoking rates (exposure) and cancer rates (outcome), we analyze how changes in smoking rates influence cancer rates.



In Epidemiology

Types of Variables Continuous Variables Categorical Varia

-Accept fractions

-Examples include age, height, and temperature.

Dichotomization: This refers to converting a continuous variable into a dichotomous one (i.e., having only two categories). For example:

- Age: 23, 17, 14, 35, 68, 15
- Dichotomized Age Group: <18 vs. ≥18

Categorical Variables (Discrete)

- Doesn't Accept fractions

-Examples include gender, employment status, and age group.

Dichotomous Variables: A subtype of categorical variables with only two possible values.

Example: Male/Female, Employed/ Unemployed.





We can also create categorical variables with more levels.





Population also called "Reference Population") -The topic was mentioned and explained in the main lectures (not the activities).

Introduction to the Null Hypothesis:

Definition: The null hypothesis is a statement that there is no relationship between the variables being tested.

- **Purpose**: it's crucial for performing statistical tests. It helps us determine whether to reject or fail to reject the hypothesis.
- **Important Note**: We never "accept" the null hypothesis. Instead, we either reject it or fail to reject it.

0	
(\mathbf{r})	Why do we care?
9	Statistical tests allow us to either "reject" or "fail to reject" the null hypothesis.
	$\Pi_0: \mu_1 = \mu_2$

the test group is no different from the average number of subjects in the placebo group.

Example: Suppose you're comparing the mean responses of two groups, such as a treatment group and a placebo group in a randomized control trial.

- The null hypothesis states that the mean measurement in the treatment group is equal to the mean measurement in the placebo group $(\mu_1 = \mu_2)$.
- This indicates no difference or effect between the groups.

P- Valu

The p-value is the value under this normal curve that tells you the likelihood of the null hypothesis being true. It essentially measures how likely it is that you made an error in rejecting the null hypothesis.



For example, if we are testing whether the average heights of two groups of children are different, and perform a t-test to produce a p-value

of 0.02, setting α = 0.05, we can conclude that null hypothesis is

rejected and that the two groups do indeed have different average heights.



If the p-value from the test is 0.02, which is less than 0.05, this indicates that you can reject the null hypothesis.

• Error Likelihood: The p-value of 0.02 suggests that there is only a 2% chance that you mistakenly rejected the null hypothesis. Since this is a small probability, it indicates a reliable result.

"If the p is low, the null (hypothesis) must go."

confidence Intervals

- Preferred over p-values.
- Provide a range where the true value is likely to fall.
- Example: If the average age is 21, a 95% confidence interval might be (20.5, 21.5).
- This means there's a 95% chance the actual average is within that range.

Test Name	Purpose	Example
T-Test	Compare the means of two groups.	Comparing the average height of children in two different schools.
Chi-Square Test	Determine if two categorical variables are associated.	Testing if smoking status (smoker/non-smoker) is associated with gender (male/ female).
ANOVA	Compare the means of three or more groups.	Comparing the average weight among three diet groups (low- fat, keto, Mediterranean).
Correlation	Assess the relationship between two continuous variables.	Checking if age and maximum heart rate are correlated.
Regression	Determines the influence of one or more variables on an outcome.	Modeling how smoking, age, and exercise influence heart disease risk.

statistical tests

Link to the lecture

The Confidence Level is complementary to α :

• Confidence Level = $1 - \lambda$

• For $\alpha = 0.05$, the confidence level is 1 - 0.05 = 0.95, or 95%.

Descriptive Statistics example

Exercise 1

Exercise 2

Computer Sales (n = 12 salespeople) Original Data: 3, 10, 2, 5, 9, 8, 7, 12, 10, 0, 4, 6 Compute the mean, median, mode, quartiles.

First order the data: 0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10, 12 $\Sigma X_i = 76$ $\overline{X} = 76 / 12 = 6.33$ computers sold Median = 6.5 computers Mode = 10 computers $Q_1 = 3.5$ computers, $Q_3 = 9.5$ computers

Data (n=16):

1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 5, 5, 6, 7, 8, 10

Compute the mean, median, mode, quartiles. Answer.

Inter-Quartile Range (IQR)

1 1 2 2 2 2 3 3 4 4 5 5 6 7 8 10

Mean = 65/16 = 4.06Median = 3.5 Mode = 2 $Q_1 = 2$ Q_2 = Median = 3.5 $Q_3 = 5.5$

Exercise 3

Data – number of Answer. First order absences (n=13) : the data: • $IQR = Q_3 - Q_1$ 0, 0, 0, 1, 1, 2, 2, 3, 3, • Example (n = 15): 0, 5, 3, 2, 1, 2, 4, 3, 1, 0, 0, 4, 5, 6, 12 6,12 0, 0, 2, 3, 4, 7, 9, 12, 17, 18, 20, 22, 45, 56, 98 Mean = 39/13 = 3.0 absences Compute the mean, Median = 2 absences • $Q_1 = 3$, $Q_3 = 22$ median, mode, Mode = 0 absences quartiles. $Q_1 = .5$ absences IQR = 22 - 3 = 19 (Range = 98) $Q_3 = 4.5$ absences

Running Descriptive Statistics in SPSS Statistics

1. STEP: Load Data into SPSS

• Commence by launching SPSS and loading your dataset, which should encompass the variables of interest – a categorical independent variable (if any) and the continuous dependent variable. If your data is not already in SPSS format, you can import it by navigating to File > Open > Data and selecting your data file.

2. STEP: Access the Analyze Menu

• In the top menu, locate and click on "Analyze." Within the

"Analyze" menu, navigate to "Descriptive Statistics" and choose "Descriptives.

"Analyze > Descriptive Statistics > Descriptives

ta •E	xample	Dataset.sav [Dat	aSet1] - IBM SPS	S Statistics	Data Editor					
Eile	Edit	View Data	Transform	Analyze	Graphs	Utilities	Extensions	Window	<u>H</u> elp	
2	•			Rep Des	orts criptive Stat	istics	P	Ereque	ancies	
		Name	Туре	Baye	sian Statis	tics	•	Deser	ntivos	Values
	1	pid	Numeric	Tabl	es			Desui	puves	None
1	2	age	Numeric	Com	nare Mean	e	,	A Explor	e	None
	3	gender	Numeric	Con	aral Linear	Model		Eross	tabs	{1, Male}
	4	ethnicity	Numeric	Con	aralizad Lin	aar Madala		TURF	Analysis	{1, White}
	5	education	Numeric	Gen	d Madala	ear models		Ratio		{1, High sch
	6	marital	Numeric	Miže	a models			P-P PI	ots	{1, Cohabiti
	7	employment	Numeric	Con	elate		,	0-0 P	ots	{1, Employe
	8	finc	Numeric	Regi	ression			2010		None
	9	pedu	Numeric	Logi	inear		,	ation (con	tinuous years)	None
	10	residentialar.	Numeric	Neu	al Net <u>w</u> ork	S	,	a		{0, Metropol
	11	asthma	Numeric	Clas	sify		•	diagnosed	by doctor	{0, No}
	12	diabetes	Numeric	Dime	ension Red	luction	,	diagnose	d by doctor	{0, No}
	13	highbp	Numeric	Sc <u>a</u> l	e		,	od pressur	e diagnosed by doctor	{0, No}
-				Alere.	and the second second	Tanks				

3. STEP: Specify Variables

• Upon selecting "Descriptives," a dialog box will appear. Transfer the continuous variable you wish to analyze into the "Variable(s)" box.

- Descriptives		~
Descriptives unique individual identificatio gender [gender] gender [gender] device identificatio marital and cohabitation statu menjoyment status (collapsed) menjoyment status (collapsed) whether has astima diagnos whether has astima diagnos whether has anal blood pres whether has acacer diagnos whether has achizophrenia di whether has achizophrenia di whether has achizophrenia di whether has anxiety diagnos whether has achizophrenia di whether has achol problem whether has ocid diagnosed	Variable(s):	ACT Contraction of the second
Adaptacolus cumptom coupritu		No
Save standardized values as variables		No
OK Pas	te Reset Cancel Help	No

4. STEP: Define Options

• Click on the "Options" button within the "Descriptives" dialog box to access additional settings. Here, you can request various descriptive statistics such as mean, median, mode, standard deviation, and more. Adjust the settings according to your analytical requirements



5. STEP: Generate Descriptive Statistics:

Once you have specified your variables and chosen options, click the "OK" button to execute the analysis. SPSS will generate a comprehensive output, including the requested descriptive statistics for your dataset.



Link to the lecture

This lecture includes these slides . However, it only teaches you how to input information into the computer, which is something that is impossible to be tested on or come up in the exam.

Chi- square in SPSS

The Chi-square test is applied when we are working with categorical variables "Nominal Variables" (e.g., gender, disease status, education level) rather than continuous data (e.g., height, weight, or age).

- 1. Open SPSS:
- 2. Enter your data into SPSS with the following columns:

<u>Subject</u>: An identifier for each participant , <u>Gender</u>: Categorical variable (Male , Female) , <u>Age</u> ,<u>Smoker</u>: Categorical variable (Smoker ,Non-smoker) , <u>Cholesterol</u>.

1004	43													Visible: 5 of	5 Variable
	subject	gender	age	smoker	cholesterol	100	Nar -	100	NM .	100	Var .	Var	934	107	var
1	1	Mate	- 44	Smoker	2.26										
2	2	Male	20	Non-smoker	6.00										
3	3	Male	35	Non-smoker	6.15										
4	4	Male	43	Non-smoker	6.00										
5	5	Female	17	Non-smoker	7.83										
8	6	Female	22	Non-smoker	4.56										
7	7	Female	48	Smoker	7.18										
8		Male	68	Smoker	4.28										
9	. 9	Male	30	Non-smoker	3.31										
10	10	Female	44	Non-smoker	7.31										
11	11	Male	53	Smoker	7.84										
12	12	Female	40	Non-smoker	7.38										
13	13	Female	32	Non-smoker	6.59										
54	14	Female	17	Smoker	9.04										
15	15	Female	29	Non-smoker	5.03										
16	16	Female	21	Non-smoker	6.36										
17	17	Female	31	Smoker	6.66										
18	18	Male	55	Non-smoker	5.10										
19	19	Female	56	Smoker	5.18										
20	20	Male	48	Non-smoker	5.07										
21	21	Female	41	Non-smoker	4.33										
22	22	Female	35	Smoker	5.82										
				10000	1000			_			_				

3. Descriptive Statistics:

Go to Analyze \rightarrow Descriptive Statistics \rightarrow Crosstabs to check the distribution of variables like gender, smoker status, etc.

Ete Edit	Yew Data Transform	Brafty Dred Harteling Gr	aphis	Littles Add-ges (jedow Help									Ele	dt ylew	Data 1	ransform	Analize Dred Mariating Grad	ns Utilities Add-gns)	Window Help								
2	🖨 🛄 🖛	Regords Descriptive Statistics	:1	* * =	43 Ⅲ									. 🗁			F	Regota P Descriptive Statistics	Pressenties	43 🎟			~6					
4:age	43	Tables										Visible: 5 of	5 Variables	4:ape		43		Tabled .	ER Descentions								Visible 5 of 5	Variables
	subject gender	Compare Means		olesteral var	100	VW	100	VOF.	100	100	101	VAI			10	ubject	gender	Compare Means +	A Contract	Var	4.50	1 100	100	Var		100	UM	100
1	1 Male	Ceneral Linear Model	. [2.26										1		1	Male	General Linear Nodel	ord Extract									-
2	2 Male	Generalized Linear Models		6.00										2		2	Male	Generalized Linear Models *	III Crosstate .									
3	3 Male	Must Bodala		6.15										3		3	Male	Maned Madels .	EE Bato.									
4	4 Male	Constitu	11	6.00										4		4	Male	Consists	EP Plots									
5	5 Female	Success	1	7.83										5		5	Female	Parameters .	C-Q Plots									
6	6 Female	Rebassion		4.56										6		6	Female	Debearder	4.56									
7	7 Female	ribuen	1	7.58										7		7	Female	- Sharen	7.18									
	8 Male	Neural Network	1	4.28											-	8	Male	Neural Networks	4.28									
	9 Male	Classify		2.21											_	9	Male	Classify +	3.31									
10	10 Female	Dimension Reduction	*	7.31										10		10	Female	Ornension Reduction	7.31									
11	11 Male	Scate	•	7.84										11		11	Male	Scale +	7.84									
12	12 Female	Honparametric Testa		7.38										12	-	12	Female	Honparametric Tests	7.38									
13	13 Famala	Forecasting		6.69										13		13	Female	Forecasting +	6.59									
54	14. Female	Deniral		9.04										14	1	14	Female	Davent +	9.04									
55	16 Female	Myttple Response		6.03										15		15	Female	Myttple Response +	5.03									
16	16 Famila	Making Value Analysis.		6.36										16		16	Female	Missing Value Analysis	6.36									
17	17 Female	Multiple Imputation		6.66										17		17	Female	Multale Imputation +	6.66									
18	1R Male	Complex Samples		5.10										10		18	Male	Complex Samples +	5.10									
19	10 Famile	Dualh Control		6.18										19		19	Female	Quality Control .	5.18									
20	20 Male	2000 Octo		6.07										20		20	Male	No. Com	6.07									
31	21 Eemale	En rice corps.	_	4.33										24	-	21	Cample	The second	4.32									
21	22 Family	M Peaker		6.82													Female	16 Smither	6.02									
	22. Female	75 35568	-	3.84	-		-	-	_				-		-		reman	20 0000	1.44			-	-	-	_	_		
-			-			_	_	_	_				151	-		_	_				_	_	_	_				
Data View	Variable View													Cieta V	Yet Yet	the Vew												
Analyze									IKN SPSS	Statistics Proc	essor is ready			Descri	New Statut	ics.								ISM SPISS	1 Statistics Pro	cessor is rea	1.1.1.1.1.1.1	
							B	(H)(m	Date Darabet	n yulua (m	etgenere De	na (mina)	Adapter 2		ec.													
								- - -	Ing at	Cesarate	in the second	Citrare,	ndet.	01.0.0	10													
								141	49	Tapes		(Corres	phen.		-		_	Tube 5 of 5 youtes										
								1	T Mar	- General	and a Model	A Corner			-		-											
							- 1	3	3 Mai	Canada Salar Sa	pd Linear Binders 1	Higes.																
									4 184	· Corviate		Bern	a.															
								1	5 Fanta	(papers)		BO GPS																
								1	7 Fand	Incide	-	7.10																

4. Reset \rightarrow Select Gender for the Row(s) & Select Smoker for the Column(s) \rightarrow Click on Cells and check the box for Expected to see the expected frequencies in the table (this is important for the Chi-square test) \rightarrow continue



1909	43											Visible: 5 of	r 5 Variabi
	subject	gender	age	smoker		cholesterol var	var var var	100	100	100	-var	1.00	100
1	1	Male	44	50	Court			II I					
2	2	Male	20	Non-s		CH CLOTHERE CHE DIRDERA	and the second	100 C					
3	3	Male	35	Non-si	-	Counts	alest	d					
4	4	Male	43	Non-se	53	V Observed	Compare column proportions	BCK.					
5	5	Female	17	Non-s	10	K Expected	Advationations Displayers method						
6	6	Female	22	Non-s		Hide small counts	-						
7	7	Female	48	S		Late Ban 16		and the second s					
	8	Male	68	S				130					
9	9	Maie	30	Non-si		Percentages	Residuals	7					
10	10	Female	- 44	Non-all		Bow	E Unstandardized						
11	11	Male	53	S		Column	E Standardized						
12	12	Female	40	Non-s		E Total	E &djusted standardized						
13	13	Female	32	Non-a									
14	14	Female	17	S		Norweeger Weights	2	1					
15	15	Female	29	Non-si-	-	Round cell counts	O Round case weights						
16	16	Female	21	Non-se	100	O Truncate cell counts	O Truncate case weights						
17	17	Female	31	5	204	O No adjustment							
18	18	Male	55	Non-si	1.54	Ines	Cancel Itale						
19	19	Female	56	S		Leona .							
20	20	Male	48	Non-sug	-			-					
21	21	Female	41	Non-smi	oker	4.33							
22	22	Female	35	Sm	oker	5.82							
	0	-	_		-		the second s						14

5. In the Crosstabs dialog box, click on the Statistics button \rightarrow Select Chi-square to run the test \rightarrow Click Continue \rightarrow ok

1:age	43													Visible: 5 of	5 Variabl
	subject	gender	age	smoker	choleste	rol var	var	var	var	var	var	var	Var.	Var	Var
1	1	Male	44	Stac	osstabs	1.00				22					
2	2	Male	20	Non-s	6	-	-		C D						
3	3	Male	35	Non-s		Crosstabs: Statis	tics	2		Exact.					
4	4	Male	43	Non-s	subject nu	Chi source)		Completions	_ I (Statistics					
5	5	Female	17	Non-s	Cholester	A Considerate				Cells					
6	6	Female	22	Non-s		Nominal		Ordinal		O'Las.					
7	7	Female	48	S		Contingency	coefficient	🔄 Gamma		Format					
8	8	Male	68	S		Phi and Cran	ner's V	Somers' d		Bootstrap					
9	9	Male	30	Non-sil		🖾 Lambda		Kendall's ta	n-p						
10	10	Female	44	Non-s		Uncertainty of	coefficient	E Kendali's ta	u- <u>c</u>						
11	11	Male	53	S		-Nominal by Inter	nat	E Kanaa							
12	12	Female	40	Non-s		E Da		E Disk							
13	13	Female	32	Non-s		E Fo		E Mehlemar							
14	14	Female	17	S				- montenae							
15	15	Female	29	Non-s	_	Cochran's an	d Mantel-Ha	enszel statistics							
16	16	Female	21	Non-s		Test common	odds rabo	equals: 1							
17	17	Female	31	S	Display du	Continu	a) Canre	Help							
18	18	Male	55	Non-s	Suppress t				_						
19	19	Female	56	S	(3)	OK Pas	e Reset	Cancel Hel	P						
20	20	Male	48	Non-sv	-	-									
21	21	Female	41	Non-smoker		4.33									
22	22	Female	35	Smoker		5.82									
	14	_			_		-	1.	_		-			-	151

The Chi-square test for independence or for association is used to determine if there is a statistically significant relationship or association between two categorical variables. However, it does not provide any information about the strength or magnitude of that relationship

File Edit View Data Transfo	orm Insert Format Analyze	Direct Ma	rketing Graph	s <u>U</u> tilities	s Add- <u>o</u> r	s <u>Window</u>	Help	
😑 🗟 🖨 🧟 🦉) 🛄 🖛 🧃 🛛		1	0		•		
* * + -								
Cutput	+ Crosstabs							-
Crosstabs Crosstabs	[DataSet2] Z:\Chris	tine 201 Ca	3july17\Sta	tistics Summary	and SPS	SS\SPSS G	iides\Hea	lthStats.sav
Gender (male or fe		T		Case	95	_		
Chi-Square Tests		1	/alid	Missi	ing	Tot	al	
	Condes (mais estemais)	N	Percent	N	Percent	N	Percent	
	* smoke (smoker or non- smoker)	176	100.0%	0	0.0%	176	100.0%	
1	Gender (male	or female)	* smoke (smoke	er or non-si	moker) Cr noke (smol smok	osstabulation ker or non- er)		1
				Smo	oker	Non-smoker	Total	_
	Gender (male or female)	Male	Count Expected Coun		46	4	1 87	
		Female	Count Expected Ooun	t .	33 39.9	5	6 89	
	Total		Count	-	79	9	7 176	
			Expected Coun	t	79.0	97.	0 176.0	
			Chi-Square	Tests				_
KI E		Value	df	Asymp. Sig	Exa	ct Sig. (2-	Exact Sig. (1	
								IBM SPSS Statistics Processor is ready

Gender (male or female) * smoke (smoker or non-smoker) Crosstabulation

			smoke (smo smo	oker or non- ker)	
			Smoker	Non-smoker	Total
Gender (male or female)	Male	Count	46	41	87
		Expected Count	39.1	47.9	87.0
	Female	Count	33	56	89
		Expected Count	39.9	49.1	89.0
Total		Count	79	97	176
		Expected Count	79.0	97.0	176.0

Look at the table showing the Observed and Expected frequencies for each combination of Gender and Smoker categories.

- Observed Counts: These are the actual counts in your data for each combination of gender and smoking status.
- Expected Counts: These are the counts you would expect if there was *no* association between gender and smoking status (i.e., if the two variables were independent).
 - large difference between observed and expected counts suggests that the variables (gender and smoking status) are not independent.
 - If the difference is not large, it suggests that gender and smoking status are independent

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	4.437 ^a	1	.035		
Continuity Correction ^b	3.821	1	.051		
Likelihood Ratio	4.455	1	.035		
Fisher's Exact Test				.049	.025
Linear-by-Linear Association	4.411	1	.036		
N of Valid Cases	176				
a. 0 cells (0.0%) have e b. Computed only for a	xpected count le 2x2 table	ess than 5	The minimum ex	pected count is 3	9.05.
b. compared only for a		_	Double-click to		

At the bottom of the Chi-square test table in SPSS, you'll see a line that tells you the percentage of cells that have an expected count less than 5. You want this percentage to be 0% (or very close to 0%). This means no cell has an expected count less than 5, and the Chi-square test is valid.

If this percentage is greater than 20%, it indicates that the assumption has been violated, and the Chi-square test results may not be reliable

 $\boldsymbol{0}$ cells have an expected count less than 5, which means the assumption is not violated.

The Pearson Chi-square test gives a statistically significant result (p = 0.035), which is less that alpha (0.05) so we reject the null hypothesis meaning gender and smoking status are dependent on each other.

All of the following about degree of freedom is correct except:
Advantages of large degree of freedom often depends on the type of the analysis
Degrees of freedom typically (but not always) relate the size of the sample.
A higher degree of freedom means more power to reject a false null hypothesis and find a significant result.

 Higher degrees of freedom generally mean smaller sample sizes

Link to the lecture

T-test on SPSS

You have two variables:

- Dependent variable: Frontal brain volume (measured on an interval scale).
- Grouping variable: Smoker vs. Non-smoker (categorical variable coded as 0 for non-smokers and 1 for smokers).

	frontal_brain_volume	group	tev	var	var	var	Var	var	1
17	4.00	.00						1	
18	4.20	1.00							
19	4.00	1.00							
20	2.60	1.00							
21	4.90	1.00							-
22	4.40	1.00							
23	4.40	1.00						1	
24	5.50	1.00							
25	5.10	1.00							
26	5.10	1.00							
27	3.20	1.00							
28	3.90	1.00							
29	3.20	1.00							
30	4.90	1.00							
31	4.30	1.00							
32	4.80	1.00							
33	2.40	1.00							
34	5.50	1.00							

الدر اسة كانت تهدف إلى مقارنة حجم الدماغ الأمامي (Frontal Brain) (Volume) بين المدخنين و غير المدخنين لمعرفة ما إذا كان هناك اختلاف إحصائي بين المجمو عتين.

البيانات المستخدمة:

- تم جمع البيانات من 36 مشاركًا في الدراسة.
- تم تقسيم المشاركين إلى مجموعتين: 17 شخصًا غير مدخنين و 19 شخصًا مدخنًا.
- 1. Go to Analyze > Compare Means > Independent Samples T-Test.

File Edit	ident_samples_t_test_ view Data Transform	smokers_brain_volume.sav	DataSet1] - IBM SPSS Statistics Data Editor
		Regorts Descriptive Statistics	
T, Ironia_oral		Compare Means	Visite. 2 012 Valiable
1	Trontal_brain_volume	General Linear Model	Means Var y
2	7.30	Generalized Linear Models	One-Sample T Test
3	6.50	Mixed Models	kdependent-Samples T Test
4	6.30	Correlate	PaNed-Samples T Test
5	7.00	Regression	Dne-Way ANOVA
6	5.90	Loginear	
7	5.20	Classify	•
8	5.00	Dimension Reduction	
9	4.70		
10	5,70	Nonparametric Tests	
11	5.70	Forecasting	
12	3.30	Multiple Response	
13	5.00	Missing Value Analysis	
14	4.60	Multiple Imputation	
15	4.80	Quality Control	
16	3.80	ROC Curve	
17	4.60	Amos 19	
	4	-	
Data View	Variable View		

2. Place the dependent variable (frontal brain volume) into the "Test Variables" box & Place the grouping variable (smoker vs. non-smoker) into the "Grouping Variable" box.

• Define the groups by selecting "0" for non-smokers and "1" for smokers.. continue





÷.	Outp	out2 [Document2] - IBA	A SPSS Statistics	Viewer							
File	Edit	<u>View Data</u> <u>Transfo</u>	rm Insert Form	at <u>Analyze</u>	Graphs	Utilities	Add-ons	Window	Help	_	
0	🚬 H 🖯 🖉 🖉 📖 🗠 🛥 🧱 🎬 📲 ⊘ 🌑 🚑 🎲 🖻 🕯										
<		+ + -									
ıt .0g '-T€		[DataSet1] C:\Doc	cuments and S	Settings\	King of	the Mod	is\Deskt	op\01d	Stats Lect	ures and Mov:	1e
2	1			Group Statis	tics						
	-		group	N	Mean	Std. Dev	iation	Std. Error Mean	Ì		
õ		frontal_brain_volume	non-smokers smokers	17 19	5.3294 4.2947	1.0	06170 94662	.2575	50 17		
			2					Independe	nt Samples Te	st	
					Levene	s Test for E Variances	Equality of S				
							p_value				
					F		Sig.	- t	df	Sig. (2-tailed)	
		frontal_brain_volume	Equal variances assumed			174	.67	9 3.0	32 34	.004	T
			Equal variances assumed	not				3.0	72 32.321	.004	
11	1										P
Double click to edd Plunt Table RM SPSS Statistics Processor is ready H: 1 0.8 W S 80 in								82 in			

1. Group: Non-smokers (17 participants) and Smokers (19 participants).

2. Standard Deviation:

This tells you the spread or variability of the brain volume measurements within each group. A larger standard deviation means more variability in measurements.

 The standard error measures how accurately the sample mean represents the population mean. A smaller standard error indicates more precision in the sample mean.

Levene's Test for Equality of Variances:

- Purpose: Tests whether the variances of the two groups (smokers and non-smokers) are equal.
- If the significance (p-value) from Levene's test is greater than 0.05, we can assume that the variances are equal. In this case, the p-value is 0.679, so the assumption of equal variances is not violated.

			Inde	ependent Sa	mples Tes	t
		Levene's Test fo Variano	or Equality of ces	1		
		F	Sig.	t	df	Sig. (2-tailed)
frontal_brain_volume	Equal variances assumed	.174	.679	3.092	34	.00
	Equal variances not assumed			3.072	32.321	.004

T-test Results:

- T-statistic: 3.092
- Degrees of freedom (df): 34
- P-value: 0.004

Since the p-value (0.004) is less than the typical significance level (0.05), we reject the null hypothesis that there is no difference between the brain volumes of smokers and non-smokers. The

Deviation

1.06170

94662

Std. Error

Mean

25750

21717

Statistical significance: This means that the difference you observed between the groups (e.g., frontal brain volume between smokers and nonsmokers) is unlikely to be due to random variability or chance. In other words, the result is probably not due to random sampling or measurement errors.

وجود علاقة إحصائية بين التدخين وحجم الدماغ الأمامي (أي أن المدخنين وغير المدخنين يختلفون في الحجم). لكن لا يمكن الحز م أن التدخين هو السبب المباشر

كن لا يمكن الجزم ان التدخين هو السبب المباشر
لهذا الاختلاف دون إجراء مزيد من الأبحاث
ودر اسات طويلة المدي

Independent Samples Test									
or Equality of ces				t-test for Equality	of Means				
						95% Confidenc Differ	e Interval of the ence		
Sig.	1	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper		
.679	3.092	34	.004	1.03467	.33466	.35457	1.71478		
	3.072	32.321	.004	1.03467	.33685	.34880	1.72055		

Cohen's d (Effect Size):

- Cohen's d measures the size of the effect or the magnitude of the difference between the two groups (smokers vs. non-smokers).
- In this case, Cohen's d is about 1.03, which indicates a large effect size. A value above 0.8 is typically considered a large effect.

SPSS also provides confidence intervals for the mean difference between groups, but this information is rarely reported.

Link to the lecture

Paired samples T-test on SPSS

The study involves 50 participants, and their resilience scores were measured before and after a training program (Time 1 and Time 2).

The goal is to determine if there is a statistically significant difference between the resilience scores before and after the training.

1. In the top menu, click on Analyze →Compare Means → Select Paired-Samples T Test.



هذا الاختبار يساعدنا في معرفة ما إذا كانت الاختلافات بين القياسين (قبل وبعد التدريب) كبيرة بما يكفي لكي نعتبرها ذات دلالة إحصائية

2. Select the column for Resilience Time 1 and move it to Variable 1 & Select the column for Resilience Time 2 and move it to Variable $2 \rightarrow ok$



outpi	ut2 [Document]	2] - IBM !	spss Statist	ics Viewe					
Edit	View Data	Iranaform	invert P	ormut Ana	ilyzn <u>C</u> imphu	Utilities Add-ons	Window Help		
a k				-	1	X 🚣 🗐	9	1	
	+ +	-			🖸 🖻				
			Paired Sa	mples Stat	listics	,			
			Mean	N	Std. Deviation	Std. Error Mean			
	Pair1 resili	ience1	52.56	50	4.563	.645			
	resili	ience2	58.80	50	4.651	.658			
	Pair1 reali	ience1 & r	ealtience2	50	.755	.000	ar Tart		
						Paired Differen	65 1651		
	Parted University of the Difference Interval of the Difference								
				Mean	Std. Deviation	Std. Error Mean	Lower	Upper	1
1				0.040	0.000	450	7.169	£ 224	
	Pair1 realli	ience1 - ru	silience2	-8.240	3.223	.400	-7,150	-3.324	-13,689

Paired Samples Statistics Table:

- This will display the means, standard deviations, and sample sizes for both Resilience Time 1 and Resilience Time 2.
- -This will give you an idea of the central tendency of each group.

Paired Samples Test Table:

- This is where you'll find the results of the t-test.
 - Mean Difference: The difference between the means of Time 1 and Time 2 (e.g., -6.24).
 - Standard Deviation of the Difference: The variability of the difference scores.
 - Standard Error Mean: The error associated with the mean difference.
 - t-value: The calculated t-statistic (e.g., -13.69).
 - Degrees of Freedom (df): The number of participants minus 1 (e.g., 49 if you have 50 participants).
 - Sig. (2-tailed): The p-value (e.g., 0.000). If this value is less than 0.05, the difference is statistically significant.

 Part of Samples Statistics
 Part of Samples Statistics

 Part of Samples Statistics
 Part of Samples Test

 Part of Samples Statistics
 Part of Samples Test

 Part of Samples Test
 Part of Samples Test

 Main
 Dist. Deviation

 Main
 Dist. Deviation

 Main
 Dist. Deviation

 Main
 Correlations

 Main
 Dist. Deviation

 Main

Paired Samples Correlations Table:

- This shows the correlation between the two sets of scores (before and after).
- You might see a correlation value of 0.755, indicating that participants' scores tend to be similar between the two time points.

كان لدى المشاركين متوسط مرونة قبل التدريب قدره 52.56، ومتوسط مرونة بعد التدريب قدره 58.80.

- الفرق بين هذين المتوسطين هو 6.24.
- القيمة p=0.000 تشير إلى أن هذا الفرق ليس صدفة، وأن التدريب بالفعل أدى إلى زيادة في مستوى المرونة

Link to the lecture

p-value (Sig.) is less than 0.05, you can reject the null hypothesis, indicating that there is a statistically significant difference between the two time points. the p-value of 0.000 indicates a very statistically significant difference

Biostatistics activity MCQ

Blood pressure of a patient is the output that depends on patient's age and comorbidity. The best way to describe patient's blood pressure reading is:	Blood pressure of a patient is the output that depends of patient's age and comorbidity. The best way to describe patient's blood pressure reading is:
O A dependant variable	An independent variable
O An independent variable	Well done
© Check	► Contin
An example of descriptive statistics is:	An example of descriptive statistics is:
O T-test and chi square test	T-test and chi square test
O Measures of validity and relaibility	Measures of validity and relaibility
 Only measures of central tendancy (Mean, Median and Mode) and dispersion (SD) 	 Only measures of central tendancy (Mean, Median and Mode) and dispersion (SD)
O ANOVA, MANOVA and ANCOVA	ANOVA, MANOVA and ANCOVA
Check	—— 1/1
	Fill in the missing words
	You have the following data (n=16):
ll in the missing words	1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 5, 5, 6, 7, 8, 10
u have the following data (n=16):	
	Compute the Mean, Median, Mode, Quartiles
1, 2, 2, 2, 2, 3, 3, 4, 4, 5, 5, 6, 7, 8, 10	(Q1, Q2, Q3)
ompute the Mean, Median, Mode, Quartiles (1, Q2, Q3)	4.06 X Mean = 4.06 3-6 X Median = 3.5 2 X Mode = 2 2 X Q1 = 2
	3.5 x Q2 = 3.5
	≝ 5.5 × Q3 = 5.5
	0/6 G



تم بحمدلله

اللهم يسر ليكل صعب، واجعل الخير في عملي ونتائجه، واجعلني من الذين يسيرون على طريق الحق والرشاد. اللهم إني أسالك علمًا نافعًا، ورزقًا طيبًا، وعملًا متقبلاً، وقلبًا سليمًا، ولسانًا ذاكرًا، ورزقًا مباركًا في الدنيا والآخرة ، اللهم افتح لي أبواب الفهم والنجاح، وبارك لي في سعيي واجعل توفيقك رفيقي في كل خطوة أخطوها.

For any feedback , click here