

### Introduction to STATISTICA





Australia: StatSoft Pacific Pty Ltd. Brazil: StatSoft Brazil Ltda. Czech Republic: StatSoft Czech Rep. s.r.o. Israel: StatSoft Israel Ltd. France: StatSoft France

Germany: StatSoft GmbH Hungary: StatSoft Hungary Ltd. Italy: StatSoft Italia srl

Japan: StatSoft Japan Inc. Korea: StatSoft Korea Netherlands: StatSoft Benelux BV Poland: StatSoft Polska Sp. z o. o.

Portugal: StatSoft Iberica Ltda. Russia: StatSoft Russia Singapore: StatSoft Singapore S. Africa: StatSoft S. Africa (Pty) Ltd.

Spain: StatSoft Espana Sweden: StatSoft Scandinavia AB Taiwan: StatSoft Taiwan UK: StatSoft Ltd.

Copyright StatSoft, Inc., 1984-2005. StatSoft, StatSoft logo, STAT/STICA, SEWSS, SEDAS, Data Miner, SEPATH and GTrees are trademarks of StatSoft, Inc.



### Outline

STRATETICS/A

After today's training session you will:

- Know the three basic output channels of STATISTICA and how to change the setting to your liking
- Be able to manipulate STATISTICA spreadsheets (e.g. edit variables and cases)
- Perform basic data management operations(recoding data, using spreadsheet formulas, setting case selection conditions, etc.)
- Import data from external data sources such as excel files, databases (as well as create random samples directly from the database)
- Review basic descriptive statistics such as the mean, median, standard deviation, and compute correlations.

#### A StatSoft

# **Structure of STATISTICA 10**

Sta	tistics			
48	Resume Ctrl+R			
¥	ByGroup Analysis			
Zh	Basic Statistics/Tables			
12	Multiple Regression			
1,0	ANOVA			
2	Nonparametrics			
29	Distribution Fitting			
~?	Advanced Linear/Nonlinear Models	۶	GLM	General Linear Models
¥	Multivariate Exploratory Techniques	۶	85	Generalized Linear/Nonlinear Mo
	Industrial Statistics & Six Sigma	۲	192	General Regression Models
N <sup>2</sup> li	Power Analysis		PLS	General Partial Least Squares M
16	Neural Networks		1	NIPALS Algorithm (PCA/PLS)
5	Da <u>t</u> a-Mining	۲	8.	Variance Components
2	QC Data Mining & Root Cause Analysis	۲	7	Sur <u>v</u> ival Analysis
n\$r	Text & Document Mining, Web Crawling	۲	Δ	Nonlinear Estimation
	Statistics of Block Data	,	2	Fixed Nonlinear Regression
-			Œ	Log-Linear Analysis of Frequenc
23	STATISTICA Visual Basic		4	Time Series/Forecasting
301	Probability Calculator		22	Structural Equation Modeling

Use the Statistics menu to access various analyses

STRATTISTUCA

Multiple analyses can be open simultaneously

All general purpose facilities are available in every module

Tables

#### A StatSoft'

#### STRATTISTUCA **3 Basic Channels for Output** Workbook1\* - Test of SS Whole Model vs. SS Residual (Adstudy.sta) \_ 🗆 × 🛅 Workbook1' Test of SS Whole Model vs. SS Residual (Adstudv.sta) 🗄 🖂 Basic Statistid Dependnt Multiple Multiple Adjusted df MS SS 🗄 🔄 Descriptiv Variable $\mathbb{R}^2$ Model Model Residual Model Desc GENDER 0.319861 0.102311 -0.099669 1.2605 9 0.14005 -11.0595🖻 🛅 2D Scatterpld Workbooks - 🗆 × 🛱 GAM Results.str . . . . . . . 1 . . . 5 . · · A · · 骨 🗄 💮 Nonpa Histogram of respo **Results of GAM Analysis** 🖻 🗠 🔂 Nd Fit summary (Adstu Responses vs. pre Reports 🔄 Power. Summary statistics edicted values - 🗆 🗵 1.08 Histoaram for 🖻 Genera Normal probal 🛅 GL - 🗆 × Half-normal pr Predicted and CORRECT1: No. correct solutions to first p... \_ 🗆 × □ Stand-alone Windows 20 Spline informa 18 No, correct solutions to first problem Spline informa 30 K-S d=.15196. p≥..20: Lilliefors p<.01 16 Spline informa - 🗆 × Spline informa 25 12 Spline informa - 🗆 🗵 12 20 10 Spline informa Data: Frequency table: MEASURE03 (Adstudy.sta) \_ 🗆 × Spline line and 10 15 Ca Observational Frequency table: MEASURE03 (Adstudy.sta) 0 10 Count Cumulative Percent Cumulative Category Percent Count 2 n. 5 4 8.00000 8.0000 8 8.00000 16.0000 11 19 22.00000 38.0000 5 6 25 12.00000 50.0000 67 3 28 6.00000 56.0000 3 31 6.00000 62.0000 8 7 38 14.00000 76.0000 5 43 10.00000 86.0000 4 47 8.00000 94.0000 50 6.00000 100.0000 ⊬





### **Data Files**

Title Bar

GENDER

MALE

MALE

MALE

MALE

FEMALE

FEMALE

🗰 Data: Adstudy.sta (25v by 50c)

Info Box

August 10th

R. Rafuse

T. Leiker

E. Bizot

K. French

K. Harrell

E. Van Landuvt

Info Box
Title Bar
Gile Heade
Variables

**Cases** 

FEMALE W. Noren COKE W. Willden MALE PEPSI S. Kohut FEMALE PEPSI MALE B. Madden PEPSI FEMALE PEPSI M. Bowling J. Willcoxson MALE COKE MALE J. Landrum PEPSI M. Taylor MALE COKE N.S. Madden FEMALE PEPSI K. Ridgway FEMALE PEPSI L. Cunha MALE COKE F. Wind FEMALE PEPSI 

File Header

ADVERT

Advertising Effectiveness Study.

PEPSI

COKE

COKE

PEPSI

PEPSI

COKE

Variables

÷.

MEASURE01 MEASURE02 MEASURE03 MEASURE04 MEASURE05 MEASURE06 MEA

Π

- 🗆 ×

Cases



# Variable Operations (The Vars button)

Access to the most common data management operations

Var	s <del>*</del>	
	<u>A</u> dd	
	<u>M</u> ove	
	<u>С</u> ору	
	<u>D</u> elete	
	Specs	
	All Specs	
Ś	<u>T</u> ext Labels	
	Create Subset/Random	Sampling
	⊻erify Data	•
	<u>R</u> ank	
x=?	Recalc <u>u</u> late	Shift+F9
	R <u>e</u> code	
	Replace Missing Data	
	Shi <u>f</u> t (Lag)	
	Standardize	
	Date Operations	Ctrl+Shift+O



# Variable Dialog

Each variable has a set of properties or specifications associated with it. Click on a variable and select **Specs...** from the **Vars** toolbar button menu to display the **Variable** specification dialog.

Variable 1	? ×							
Α	$\blacksquare  \blacksquare  \blacksquare  \blacksquare  \overset{\mathbf{V}}{=}  $							
Name: GENDER	Type: Double  OK							
Measurement Type: Auto	Length: 8 🖆 Cancel							
🗖 Excluded 🗖 Label 🗖 Case State	<u>M</u> D code: -9999 ► < <u>&gt;</u>							
Display format General Number Date Time Scientific Currency Percentage Fraction Custom	All Specs         Text Labels         Values/Stats         Properties         [Bundles]							
Long name (label or formula with <u>Eunctions</u> ):								
Gender of the subjects (May 15, 1	(996).							
Labels: use any text. Formulas: use variab Examples: (a) = mean(v1:v3, sqrt(v7), AG	le names or v1, v2,, v0 is case #. E) (b) = v1+v2; comment (after;)							

STRATESTRICA





# Variable Types

STATISTICA Spreadsheet data files support the four basic data types listed below:

- Double is the default format for storing numeric values in STATISTICA. Each numeric value can have a unique text label attached.
- Integer is the data type to select for whole number values.
- Byte is the data type for integers between and including 0 through 255.
- Text is optimized for storing sequences of any characters of long length.





### **Text Labels**

Text labels can aid in the interpretation of their respective numeric values.

Text Labels Editor	[GENDER]	? ×									
A Arial	🔽 10 🔽 B Z 👖 📥 - 🐰 🗈 🛱	ы									
Sort Now + Renumber + Copy From + Apply To											
Text Label	Numeric Description	OK									
MALE	1	Cancel									
FEMALE	2										
		<u>D</u> elete Row									
		Class All									
2 complete Text Lab	els out of 2 rows										



#### STRATISTICA

### **Variables Specifications Editor**

All variables can be reviewed or edited in the *Variable Specifications Editor* 

	Variable Specifications Editor									
€	3.	A Arial		•	10 💌	BI	u 🔺 🕺 🖻 🛍	<u>V</u> ars ▼		
		Name	Туре		MD code	Length	Long Name (label or formula)	Measurement Type	Excli	
	1	GENDER	Double	•	-9999		Gender of the subjects	Auto 💌	Γ	
	2	ADVERT	Double	•	-9999		Ad shown to the subje	Auto 💌	Γ	
	3	MEASURE01	Double	•	-9999			Auto 💌	Γ	
	- 4	MEASURE02	Double	•	-9999			Auto 💌	Γ	
	- 5	MEASURE03	Double	•	-9999			Auto 💌	Г	
	6	MEASURE04	Double	•	-9999			Auto 💌	Г	
	- 7	MEASURE05	Double	•	-9999			Auto 💌	[]	
	8	MEASURE06	Double	•	-9999			Auto 💌	Ε	
	9	MEASURE07	Double	•	-9999			Auto 💌	E	
	10	MEASURE08	Double	•	-9999			Auto 💌	Ε	
	11	MEASURE09	Double	•	-9999			Auto 💌	Ε	
	12	MEASURE10	Double	•	-9999			Auto 💌	Ε	
	13	MEASURE11	Double	•	-9999			Auto 💌	Γ	
	•		Daulata		0000			0	▶	
	Output to Spreadsheet     OK Cancel									





### Case Operations (The Cases button)

Used to perform operations on selected groups of cases in the data file



11



### **Case Names**

Case names can be used as long, unique identifiers for the observations in the spreadsheet. They are also used by default as labels for many graphs.

Use the Case Names Manager to adjust case name length and width and to transfer case names to or from a variable



STLATUSTUCA





# Example 1: Creating a hypothetical data file



Let's begin by creating a hypothetical data file. We will enter information about 18 people. The spreadsheet will contain the gender, eye color, hair color, height, weight, and age of each person.

### M StatSoft



# Example Steps

	Create New Document           Image: Workbook         Image: In-place Database Interface         Image: Browser Window           Image: Spreadsheet         Image: Browser Window         Image: Browser Window <t< th=""></t<>									
Va	iable Specification	ons Editor	▼ 10 ▼	<b>B</b> <i>I</i> ]	u 🔺 🐰 🖻 f	<u>a</u> Vars -	? ×			
	Name	Туре	MD code	Length	Long Name (label or formula)	Measurement	Exclude			
1	Gender	Text 💌		8		Auto 🔻				
2	Eye Color	Text 💌		8		Auto 💌				
3	Hair Color	Double 💌	-9999			Auto 💌				
4	Height	Double 💌	-9999			Auto 💌				
5	Weight	Double 💌	-9999			Auto 💌				
6	Age	Double 💌	-9999			Auto 💌				
•							►			
	OK Cancel									

1. Create a new data file with 6 variables and 18 cases.

STRATE STRICA

- 2. Save the spreadsheet.
- 3. Give the 6 variables these names *Gender*, *Eye Color*, *Hair Color*, *Height*, *Weight*, and *Age*.
- 4. Change the variable Type.

#### A StatSoft



# Example Steps

Data:	Informatio	n.sta* (6v b	y 18c)			
🔠 Da	ata: Inform	ation.sta* (l	6v by 18c)			_ 🗆
		_	_		_	
		2	3	4	5	6
	Gender	Eye Color	Hair Color	Height	VVeight	Age
1	Female	Blue				
2	Female	Brown				
3	Female	Green				
4	Female	Blue				
5	Female	Brown				
6	Female	Green				
- 7	Female	Blue				
8	Female	Brown				
9	Female	Green				
10	Male	Blue				
11	Male	Brown				
12	Male	Green				
13	Male	Blue				
14	Male	Brown				
15	Male	Green				
16	Male	Blue				
17	Male	Brown	_			
18	Male	Green				
I.			·			

#### 5. Enter the data.

- 6. Enter the values for *Eye Color*.
- 7. Color the cells under Gender.

SILA TRISILICA

- 8. Color the text in the cells under *Eye Color*.
- 9. Save your changes.





### **Block Stats**

Statistics for each row or column in a selected block can be computed and added to the spreadsheet by selecting the desired *Statistics of Block Data* from the shortcut menu.

🏢 Data: Adstudy.sta* (25v by 50c)												
	Advertisir	ig Effectiver	ness Study									
	1	2	3		4	5	6	6 7				
	GENDER	ADVERT	MEASUR	E01	MEASURE02	MEASURE03	MEASURE04	MEASURE05	MEASURE06	M		
R. Rafuse	MALE			9	1	6	8	1	2			
T. Leiker	MAL 💷	<u>Statistics of B</u>	ock Data 🔸		Block <u>C</u> olumns	<u>M</u> eans	8	0	0			
E. Bizot	FEN 🔝	<u>G</u> raphs of Blo	ck Data 🔹 🕨		Block <u>R</u> ows	<ul> <li>Medians</li> </ul>	9	8	8			
K. French	MAL 🕵	Graphs of <u>I</u> np	ut Data 🔷 🕨	7	9	<u>S</u> D's	5	9	9			
E. Van Landuyt	MAL v	C.4	CHLV	7	1	⊻alid N's	2	8	9			
K. Harrell	FEN 😤	շ		6	0	Sums	8	3	1			
W. Noren	FEM 🖼	<u>L</u> opy	Ltri+L	7	4	Min's	2	5	7			
W. Willden	MAL	Copy <u>w</u> ith Hea	aders	9	9	Maria	6	6	8			
S. Kohut	FEM 🖪	<u>P</u> aste	Ctrl+V	7	8	M <u>a s</u> 05%/-	3	6	9			
B. Madden	MAL	Paste Sp <u>e</u> cial		6	6	<u>2</u> 5%'s	8	3	6			
M. Bowling	FEN	Fill/Chandlardia	n Plaak 🔹	4	6	<u>7</u> 5%'s	5	6	8			
J. Willcoxson	MAL 💾	nii/otariualuiz ot	е <u>р</u> іоск •	7	3	All	7	0	6			
J. Landrum	MAL U	Clear -	•	6	2		1	8	1			
M. Taylor	MAL	<u>F</u> ormat	•	7	2	4	8	1	2			
N.S. Madden	FEM	Marking Cells	,	6	2	7	5	7	2			
K. Ridgway	FEM			3	2	5	4	4	4			
L. Cunha	MALE	COKE		2	9	9	3	1	4			
F. Wind	FEMALE	PEPSI		1	0	7	5	2	4			
				0	c	1	2	1	1			



#### STRATISTICA

### Recode

The Recode dialog is used to define new values of the selected variable depending on the specific conditions that you define.

Recode Values of Variable 4: MEA	SURE02	? ×
Category 1 Include If: MEASURE01<0	New Value 1 • value •1	OK Cancel
Category 2 Include If: MEASURE01>0	New Value 2 value 1 MD code	<u>_</u> [lear all ☐ Open <u>S</u> ave As
Category 3	New Value 3 • value MD code	Other If no conditions are met, set values to:
Category <u>4</u>	New Value 4 Value MD code	© <u>M</u> D code © value ⊙ <u>u</u> nchanged





STRATETISTETCA

Sorting can be performed on both text and numeric values, as well as on case names, in either descending or ascending order.

Variables			Variables	Direction	Sort By
casename	Add Ver(e) N		1-Gender	Ascending	Text
1-Gender					
2-Eye Color	- Dissetion				
3-Hair Color					
4-Height (in)	O Ascending				
5-Weight (lb)					
6-Age (yr)	- Sort Bu-				
7-Wellness 1	C Numeric				
8-Wellness 2	C Text				
9-Test Item 1					
10-Test Item 2					
11-Test Item 3	< Berrove Var(s)				
12-Test Total (Avg)					
	< < Remove All				
		-			



# **Example 2: Working with** Variables



### **1.Open the** InfoTwo.sta data file. Notice that this data file is similar to the one you created earlier. Data has been entered for the variables *Height*, Weight, and Age.

#### M StatSoft



# Example Steps

×

? ×



STRATTISTRICA

A Arial		10 💌	<b>B Z U</b>	<u>A</u> -	∦ ⊑	a 😩 🖂				
			5bû i o							
- Data:	InfoTwo.sta* (6	v by 18c)								Ě
<u>t</u>	1	2	3		4	5	6			<u>~</u>
	Gender	Eye Cold	or Hair Col	or	Height	vveight	Age			
	Data: Infoli	wo.sta- (7v	Dy 18CJ							
		1 Gender	2 Eye Color	Hair	3 Color	4 Height	5 Weight	6 Age	7 SUM var	4-6
	1	female	blue	brow	n	69	26	1 32		362
	2	female	brown	blac	Add Var	iables				
	3	female	green	red	How m	anır 1	<u> </u>	leo 0 in "Attor" f	iald to incart	Π
	4	female	blue	fold	<u>11</u> 000 III	any. In		efore first varial	ole.	
	5	female	brown	brov	<u>A</u> fter:	SUM var	4-6 t	)ouble-click on it o select variable	or press F2 from list.	Car
	7	female	green blug	rod	Mana a	Dunamic 9	Sum			
	8	female	hrown	hlor	<u>in</u> ame.	Jo yn dinio (	- Califi	<u></u>		
4 con	9	female	areen	brov	<u>M</u> D cod	te: -9999		L <u>e</u> ngth: 8	1	
	10	male	blue	blac	_ Displa	ay format —				
	11	male	brown	red	Gen	eral				
	12	male	green	blor	Num	ber				
	13	male	blue	brov	Date					
	14	male	brown	blac	Scie	ntific				
	15	male	green	red	Perc	ency entage				
	16	male	blue	blor	Frac	tion				
	17	male	brown	brov	Cust	om				
	18	male	green	blac						
					Long n	ame (label or	formula with	n <u>F</u> unctions	):	
-				_	=sum	(v4:v6)				

- 2. Enter values for the *Hair Color* variable.
- 3. Find the sum of the *Height*, *Weight*, and *Age* variables for each case.
- 4. Create a dynamic variable that automatically updates as the data change.

### M StatSoft'



# Example Steps

nclude If: Dynamic Sum	▼ <= 300		New Value 1 Value ( MD code	A		-		OK Cancel	
ntegory <u>2</u> nclude If: Dynamic Sum ntegory <u>3</u>	Sort 0 Varia Case 1-Ge 2-Eye 3-Ha	ptions bbles name nder e Color ir Color		Add Var(s)		Variables 1-Gender 9-Code Iove Variab	Dir As As Dies	rection cending	? × Sort By Text Numeric
Data: InfoT	wo.sta* (9	ight / by 18c)		C Descen	ding	From variabl	e: juode		
	1 Gender	2 Code	3 Eye Color	4 Hair Color	5 Height	6 Weight	7 Age	8 SUM var 4-6	9 Dynamic Sum
1	female	A	brown	black	66	154	78	298	298
2	2 female	A	green	red	67	198	32	297	207
								201	231
3	8 female	A	blue	blonde	70	144	35	249	249
3	8 female 1 female	A A	blue brown	blonde brown	70 65	144 185	35 50	249	237 249 300
3 2 6	6 female 1 female 5 female	A A A	blue brown green	blonde brown black	70 65 70	144 185 165	35 50 58	249 300 293	249 300 293
3 4 6 6	female female female female	A A A A	blue brown green green	blonde brown black brown	70 65 70 73	144 185 165 161	35 50 58 45	249 300 293 279	249 249 300 293 279
3 2 6 8 7	6 female 1 female 5 female 6 female 7 female	A A A B	blue brown green green blue	blonde brown black brown brown	70 65 70 73 69	144 185 165 161 270	35 50 58 45 32	249 300 293 279 362	237 249 300 293 279 371
3 4 6 7 7 8	6 female female female female female female female	A A A B B	blue brown green green blue blue	blonde brown black brown brown red	70 65 70 73 69 63	144 185 165 161 270 212	35 50 58 45 32 33	249 249 300 293 279 362 308	249 300 293 279 371 308
3 2 6 7 7 8 8	6 female 6 female 6 female 7 female 7 female 8 female 9 female	A A A B B B B	blue brown green green blue blue brown	blonde brown black brown brown red blonde	70 65 70 73 69 63 72	144 185 165 161 270 212 187	35 50 58 45 32 33 61	249 249 300 293 279 362 308 320	249 300 293 279 371 308 320
3 4 6 7 7 8 9 9 9	female female female female female female female male	A A A B B B A	blue brown green green blue blue brown blue	blonde brown black brown red blonde black	70 65 70 73 69 63 72 69	144 185 165 161 270 212 187 133	35 50 58 45 32 33 61 66	249 249 300 293 279 362 308 320 268	249 300 293 279 371 308 320 268
3 4 6 7 8 8 9 10 11	female female female female female female female male male	A A A B B B A A	blue brown green blue blue brown blue green	blonde brown black brown red blonde black blonde	70 65 70 73 69 63 72 69 69 66	144 185 165 161 270 212 187 133 179	35 50 58 45 32 33 61 66 39	249 249 300 293 279 362 308 320 268 284	237 249 300 293 279 371 308 320 268 284 284
3 4 5 7 8 9 10 11 11 12	female female female female female female female male male male male	A A A B B B A A A	blue brown green blue blue brown blue green blue blue	blonde brown black brown red blonde black blonde brown	70 65 70 73 69 63 72 69 66 70	144 185 165 161 270 212 187 133 179 180	35 50 58 45 32 33 61 66 39 45	249 249 300 293 279 362 308 320 268 284 284 295	237 249 300 293 279 371 308 320 268 284 284
23 24 25 27 26 27 28 29 20 20 20 20 20 20 20 20 20 20 20 20 20	3 female 4 female 5 female 6 female 7 female 8 female 9 female 1 male 1 male 2 male 2 male 3 male	A A A B B B A A A A	blue brown green blue blue brown blue green blue blue brown	blonde brown black brown red blonde black blonde brown black	70 65 70 73 69 63 72 69 66 70 71	144 185 165 161 270 212 187 133 179 180 174	35 35 50 58 45 32 33 61 66 39 45 43	249 300 293 279 362 308 320 268 284 284 295 288	239 249 300 293 279 371 308 320 268 284 284 295 288
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	3 female 4 female 5 female 6 female 7 female 9 female 1 male 1 male 2 male 3 male 4 male 4 male	A A A B B B A A A A A A	blue brown green blue blue brown blue green blue blue blue blue	blonde brown black brown red blonde black blonde blonde black blonde	70 65 70 73 69 63 72 69 66 70 71 65	144 185 165 161 270 212 187 133 179 180 174 175	35 50 58 45 32 33 61 66 39 45 43 43	249 300 293 362 362 308 320 288 284 285 288 284 285 288	237 249 300 293 279 371 308 320 268 284 284 295 288 295
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	6 female 6 female 6 female 7 female 8 female 9 female 9 female 1 male 2 male 2 male 4 male 4 male 4 male	A A A B B B A A A A A A A A	blue brown green blue blue brown blue green blue brown blue brown	blonde brown black brown red blonde black blonde black blonde black blonde black	70 65 70 73 69 63 72 69 66 70 70 71 65 61	144 185 165 161 270 212 187 133 179 180 174 175 215	32 35 50 58 45 32 33 61 66 39 45 43 45 43 57 20	249 300 293 279 362 308 320 268 284 295 288 297 298	237 249 300 293 279 371 308 320 268 284 295 288 295 288 297 296
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	a         female           b         female           c         male           c         male           male         male           male         male           male         male           male         male           male         male           male         male	A A A A B B B B A A A A A A A A A A A	blue brown green green blue brown blue green blue brown blue brown	blonde brown black brown red blonde black blonde black blonde blonde blonde blonde	70 65 70 73 69 63 72 69 66 70 71 65 61 62	144 185 165 270 212 187 133 179 180 174 175 215 190	355 500 588 455 322 333 61 669 399 455 433 57 200 360	249 300 293 302 308 320 268 284 295 288 295 288 297 296 288	237 249 300 293 279 371 308 320 268 288 284 295 288 297 296 288
3 4 5 6 7 7 8 8 5 5 5 10 11 11 12 13 13 14 11 12 13 14 11 17 17	6         female           6         female           6         female           7         female           8         female           9         female           9         female           1         female           2         male           3         male           4         male           5         male           6         male           7         male	A A A B B B B A A A A A A A A A B	blue brown green blue blue brown blue green blue brown blue brown green brown	blonde brown black brown red blonde blonde blonde brown black blonde brown black red	70 65 70 73 69 63 72 69 66 70 71 65 61 1 62 74	144 185 165 161 270 212 187 133 179 180 174 175 215 190 202	352 365 50 58 45 32 33 61 66 39 45 43 57 20 36 36 29	249 300 293 279 362 308 320 268 284 284 295 288 297 296 288 297 296 288 305	249 249 300 293 279 371 308 320 268 288 288 295 288 297 296 288 297 296 288 305

5. Create another new variable that will contain codes.

STRATTISTICA

- 6. Sort the data file.
- 7. Rearrange the order of the variables.
- 8. Rank variables.





### **Importing Data**





s. xls	OK
1 🖨 to 150 🖨	Cancel
1 🗣 to 500 🖨	
	1 1 + to 150 + 1 1 + to 500 +

### □Use the Clipboard

### □ File – Open



# **STATISTICA** Query

### **STATISTICA** provides access to most databases (including many large system databases such as Oracle, Sybase, etc.) via **STATISTICA** Query



STRATUSTRICA



# **Exporting Data**

To export, select **Save As** from the **File** menu to display the **Save As** dialog.

A wide variety of files is available in the **Save as type** drop-down list (STATISTICA spreadsheet, Excel, SPSS (Data and Portable files), SAS (Data and Transport files), JMP, Minitab, dBASE, Text, HTML, Lotus Worksheets, Quattro Pro/DOS, and PDF).



STRATUSIUGA



#### STATISTICA

### **Random Subsets**

STATISTICA includes a facility that allows the user to randomly sample any dataset in order to create a subset dataset for analysis.



#### A StatSoft

### **Case Selection Conditions**

If you only want to analyze a specific subset of your data, you can use case selection conditions.

Cases can be selected in the spreadsheet or for a specific graph or analysis.

Spreadsheet C	Case Selection Conditions	<u>र</u>
Selections D	isplay   Subset/Random Sampling	1
C <u>A</u> ll ⊙ Spec E)	Analysis/Graph Case Selection Conditions         O       Use current Spreadsheet selection conditions         O       Use selection conditions for this Analysis/Graph only	<u> </u>
	Enable Selection Conditions     Include cases     Clear All	ОК
or case	© <u>A</u> ∥	Cancel
Exclude (	Specific, selected by:     By Expression:	<u>(</u> ☐) <u>O</u> pen
byex		<u>ave As</u>
or case		
	Exclude cases (from the set of cases defined in the "Include cases' section)	
By case num By expressio	By expregsion:	
[ <del>]]</del>	or case number: 7, 10:12	
	By case number:     Enter case numbers and/or ranges. Example: 1, 3, 5-12       By expression:     Use the same operators, functions, and syntax as in the spreadsheet formulas:       Use variable names or v1, v2     v0 is the case number (v0×4 means cases 1- Examples: (a) v1=0 OR age>18 (b) gender="MALE AND v4↔(v5+v6)	3)

STLATETSTETCA





### **Case States**

Assign case states to cases in order to customize the appearance of points in graphical displays

Assign unique point markers, or specify cases as *Excluded*, *Hidden*, *Labeled* or *Marked*.







### **Options**

Options		? ×				
Reports       Graphs 1       Graphs 2       Spreadsheets       Import       Data Miner       In-Place Database         General       Analyses/Graphs       Output Manager       Custom Lists       Configuration Manager       Macros (         Startup options: <ul> <li>Maximize STATISTICA window</li> <li>Open most recently used datafile</li> <li>Create new spreadsheet</li> <li>Do not open or create a document</li> <li>Display the welcome dialog</li> <li>Use metric measurements</li> <li>Display warning when opening STATISTICA with event Macros enabled</li> <li>Display warning when opening a document with event Macros</li> <li>Start another Analysis of the same type without asking if that is what you intend</li> <li>Computation of percentiles:</li> <li>Empirical Distribution Function w/Average</li> <li>File location</li> <li>Remember directories when opening or saving files</li> </ul>	Interface Bro SVB) Programs	Workbooks				
Default location:       C:\Documents and Settings\Kelly Ridgway\My Docume       Browse       Heset         Auto Save       Save recovery info every       10       minutes.       Disable         Recovery path:       C:\Documents and Settings\Kelly Ridgway\Application       Browse       Reset         Offer to suspend the auto-saving feature if the file size is larger than       5       megabytes         When a document exceeds the threshold size:       •       •       Do not save         WebSTATISTICA URL:       •       •       •       •						
	ОК	Cancel				

28



# **Descriptive Statistics**

The Basic Statistics and Tables module gives you more depth and control over the output than statistics of block data.

Select *Basic Statistics/Tables* from the *Statistics* menu.



STRATRISTRICA



# **Specifying an Analysis**

- 1. Use Characteristics.sta.
- 2. Select Descriptive Statistics
- 3. Click *Variables* and select the variables for the analysis



STRATRISTRICA



# **Reviewing Results**

#### Click the Summary button to produce the results spreadsheet with the default selection of statistics

5. The *Normality* tab of the *Descriptive Statistics* dialog contains many of the most common tools for checking normality assumptions

Workbook1* - Des	criptive Statist	ics (Chara	acteristics.	sta)			_ 🗆 ×
Workbook1*		Descripti	ve Statistic	s (Characte	eristics.sta)		
	Variable	Valid N	Mean	Minimum	Maximum	Std.Dev.	
Descripti	Wellness 1	100	49.37074	19.55478	79.60917	11.84380	
	Wellness 2	100	60.75626	35.02135	87.36024	11.85713	
	Test Item 1	100	9.67000	5.00000	14.00000	3.04861	
	Test Item 2	100	19.27000	15.00000	24.00000	3.08746	
	Test Item 3	100	14.23000	10.00000	19.00000	2.83860	
							▼
							Þ
	Descriptive	Statistics (C	haracteristic	s.sta)			

STRATETISTECA

Workbook1* - Freq	uency table: Height (in) (C	haracte	ristics.sta)			
Basic Statistics/T		Frequency table: Height (in) (Characteristics.sta) K-S d=.09198, p>.20; Lilliefors p<.05 Shapiro-Wilk W=.97272, p=.03582				
Frequenc		Count	Cumulative	Percent	Cumul %	% of all
	Category		Count	of Valid	of Valid	Cases
	55.00000 <x<=60.00000< th=""><th>2</th><th>2</th><th>2.00000</th><th>2.0000</th><th>2.00000</th></x<=60.00000<>	2	2	2.00000	2.0000	2.00000
	60.00000 <x<=65.00000< th=""><th>19</th><th>21</th><th>19.00000</th><th>21.0000</th><th>19.00000</th></x<=65.00000<>	19	21	19.00000	21.0000	19.00000
	65.00000 <x<=70.00000< th=""><th>60</th><th>81</th><th>60.00000</th><th>81.0000</th><th>60.00000</th></x<=70.00000<>	60	81	60.00000	81.0000	60.00000
	70.00000 <x<=75.00000< th=""><th>19</th><th>100</th><th>19.00000</th><th>100.0000</th><th>19.00000</th></x<=75.00000<>	19	100	19.00000	100.0000	19.00000
	Missing	0	100	0.00000		0.00000
	<b> </b> •					
	Frequency table: Height (ir	n) (Charac	teristics.sta)			••



#### STATISTICA

### **Example 3: Descriptive Stats**



- 1. Open the *Characteristics.sta* data file.
- 2. From the *Statistics* menu, select *Basic Statistics/Tables*, then select *Descriptive Statistics*.

#### M StatSoft



# Example Steps



3. Click the *Variables* button and highlight variables *4-8.* 

STRATRISTRI

- 4. Click the *Summary* button.
- Resume the analysis by clicking the *Descriptive Statistics* button on the *Analysis* bar, then click the *Histograms* button.

### M StatSoft



# Example Steps



📅 Normal P-Plot: Height (in) 📅 Normal P-Plot: Weight (lb) 📅 Normal P-Plot: Age (yr)

 Resume the analysis and click on the *Advanced* tab. Select only the statistics shown at left.

STRATTISTRIC

- Click the Summary: Descriptive statistics button
- 8. Resume the analysis and on the *Prob.& Scatterplots* tab, click the *Normal probability plot* button.

4





STRATTISTRICA

Correlation is a measure of the relationship between two or more variables. The measurement scales used should be at least interval scales. Correlation coefficients can range from -1.00 to +1.00. The value of -1.00 represents a perfect negative correlation while a value of +1.00 represents a perfect positive correlation. A value of 0.00 represents a lack of correlation.



#### STATISTICA

### **Example 4: Correlations**



 Continue using the Characteristics.sta data file. Select Basic Statistics/Tables from Statistics menu, then select Correlation matrices and click the OK button to display the Product-Moment and Partial Correlations dialog.


# Example Steps



 Click the One variable list button and select variables 4-11. Click OK.

SILATUSILIC

- 3. Click the *Summary* button.
- Resume analysis. Click the Options tab. Select the Display r, p-levels, and N's option button. Click Summary
- 5. Resume analysis. On the *Quick* tab, click *Scatterplot matrix for variables*. Select variables *4-6* and click *OK*.





### t-tests

- There are four different types of t-tests in the Basic Statistics and Tables module. The following examples illustrate when to use each test.
- 1. Independent, by groups
- 2. Independent, by variables
- 3. Dependent samples
- 4. Single sample



STATISTICA

## Example 5: t-test Independent, By Groups



When one variable contains codes for two groups and the second variable contains measurements or values of a dependent variable, one should use a t-test by groups to compare the group means.



# Example Steps

T-Test for Independent Samples by Groups: Characteristics.sl	ta ?_×
☑ Variables:       Dependent: Height (in)         Grouping:       Gender	
Code for Group 1: "male" Code for Group 2: "female"	
Quick Advanced Options	🔊 Options 🔻
Summary: <u>I</u> -tests	CASES S
<u>∎esē</u> <u>B</u> ox & whisker plot	□ Weighted moments
	DF =
Workbook1.stw* - T-tests; Grouping: Gender (Characteristics.sta)	
Workbook1.stw*     T-tests; Grouping: Gender (Characteristics.sta)       Gasc Statistics/T     Group 1: male       Group 1: male     Group 2: female	Ă
Mean Mean t-value df p Valid N Variable male female	Valid N Std.Dev. Std female male fer
Height (in) 67.78846 68.00000 -0.294428 98 0.769054 52	48 3.488561 3.6
T-tests; Grouping: Gender (Characteristics.sta)	

 Use Characteristics.sta. Select Statistics-Basic Statistics/Tables, then, select t-test, independent, by groups and click OK.

PAJPI (SIPI (C)A

- 2. Click *Variables*, select *Height* (*in*) as the dependent variable, and *Gender* as the grouping variable. Click *OK*.
- 3. Click Summary.





### Example 6: t-test, Independent, By Variables



When the two groups to be compared reside in separate variables, it is more appropriate to choose a t-test by variables.



## **Example Steps**

T-Test for Independent Samples by V	ariables: Character ?
Image: Second contraction of the second contraction o	Cancel
Quick Options	Deptions V
Summary: <u>1</u> -tests       Box & whisker plot	DF =
	Nonstandard data arrangement: It is assumed that each variable contains the data for one group.

Workbook1.stw - Masic Statistics/Tab - Masic Tutest for indeps		T-test for Independent Samples (CharacteristicsHe Note: Variables were treated as independent samples					
T test for Inc		Mean	Mean	t-value	df	p	Valid N
1-test for the	Group 1 vs. Group 2	Group 1	Group 2				Group 1
	Male Height vs. Female Height	67.78846	68.00000	-0.294428	98	0.769054	5,
							T
	[•]						•
	, T-test for Independent Samples (Chara	acteristicsHeig	ght.sta)				

### 1. Use *CharacteristicsHeight.sta*.

- 2. On the *Basic Statistics and Tables* dialog, select *t-test, independent, by variables* and click *OK*.
- 3. Select *Male Height* as the first variable and *Female Height* as the second variable.
- 4. Click Summary.









### Example 7: t-test, Dependent Samples



If the two groups being compared were measured twice on the same variable, then a considerable portion of the within-group variation can be attributed to the individual differences between measurements on the same subjects.



# Example Steps

<mark> </mark>	pendent Samples: Characte      tem 1	ristics.sta ? _ X Summary Cancel	
Workbook1.stw* - T-test for Der	pendent Samples (Characteristics sta)		
Workbook1.stw <sup>*</sup> Basic Statistics/T     Garage T-test for dep     T-test for dep     Variable	T-test for Dependent Samples (Charact Marked differences are significant at p       Mean     Std.Dv.       N     Diff.       S     S	c.05000 td.Dv. t df p Diff.	
Workbook1.stw* - Bo	<b>x &amp; Whisker Plot</b> Box & Whisker Ple	ot.	<u>-   0   ×</u>
T-test for dep T-test for dep Box & W	2 Test Item 1 vs. Test It	em 2	
2	o	Ē.	
1	8		
1	6		
	2		
1			
	8 Test lien 1 Test	tem 2	E .96*SE
	T-test for Dependent Samples (Characteristics.s	ta) 📅 Box & Whisker Plot	

 Use Characteristics.sta. On the Basic Statistics and Tables dialog, select t-test, dependent samples.

CASTISSIII (CA

- 2. Select *Test Item 1* as the first variable and *Test Item 2* as the second variable.
- 3. Click the *Summary* button.
- 4. Resume the analysis. Click the **Box & whisker plots** button.



STATISTICA

### **Example 8: t-test, Single Sample**



Using the single sample t-test, you can compare the mean of a particular variable to a specified value.



# Example Steps



PAD MISSINICIA

🔀 T-Test for Single Means: Characteristics.sta 👘	?_×
Variables: Weight (lb)	Summary
Quick Advanced Options	Cancel
Summary: <u>T</u> -tests	🔊 Options 🔻
Reference values	
📀 Test <u>a</u> ll means against: 🛛 200 🚔	
C Test means against different	CRSES S
	Weighted moments
	_ DF =
<u>∎o</u> ox & whisker plot	© W-1 O N-1
	<u>M</u> D deletion
	C Casewise
	Pairwise
	Pairwise

- Use Characteristics.sta. On the Basic Statistics and Tables dialog, Select t-test, single sample.
- 2. Use Weight (lb) as variable.
- 3. Select the *Test all means against* option button and enter *200* into the adjacent box.

#### M StatSoft'



# Example Steps





4. Click the Summary button.

SILAILISILICA

 Resume the analysis. Click the *Box & whisker plot* button. Select *Mean/SE/1.96\*SE* in the *Box-whisker Type* dialog box. Click *OK*.

STATISTICA

### **Breakdown and One-way ANOVA**

- Typically used as an EDA Technique, breakdowns answer the question, "Are the groups different regarding the dependent variable?"
- Examine group means with one-way ANOVA.
- Investigate variation with homogeneity of variance tests
- Conduct a variety of Post-hoc tests

#### STRATISTICA

### Example 9: Breakdowns and One-way ANOVA



 Continue using the Characteristics.sta data file.
 Select Breakdown & Oneway ANOVA from the Basic Statistics and Tables (Startup Panel), and click OK.

#### M StatSoft



## Example Steps

Statistics by Groups (F	}reakdo <mark>wn)</mark> : Ch	aracteristics.sta	1	?_×				
Individual tables Lists of ta	ibles			OK				
<u>G</u> rouping variables:	Not selected			Cancel				
Dependent variables: none								
Output Tables	Statistics		s					
Summary table of mea	ns   🔽 <u>N</u> is   🔽 Std. de	evs 🗖	S <u>u</u> ms – – Variances – Г	- Weighted				
Within-group correlation	ns 🛛 🗖 Min & r	ma <u>x</u>	Median &	DF =				
Select up to 6 lists of grouping variabl	es:				? ×			
1-Gender         1-Gender           2-Eye Color         3-Hair Color           3-Hair Color         3-Hair Color           4-Height (in)         4-Height (in)           5-Weight (lb)         5-Weight (lb)           6-Age (yr)         7-Wellness 1           8-Wellness 2         8-Wellness 2           9-Test Item 1         9-Test Item 1           10-Test Item 2         10-Test Item 3           12-Test Total (Avg)         12-Test Total (Avg)	1-Gender 2-Eye Color 8-Haight (in) 5-Weight (ib) 6-Age (vr) 7-Weilness 1 8-Weilness 2 9-Test Item 3 10-Test Item 3 12-Test Total (Avg)	1-Gender 2-Eye Color 3-Hair Color 4-Height (in) 5-Weight (ib) 6-Age (yr) 7-Wellness 1 8-Wellness 2 9-Test Item 1 10-Test Item 2 11-Test Item 3 12-Test Total (Avg)	1-Gender 2-Eye Color 3-Hair Color 4-Height (in) 5-Weight (ib) 6-Age (yr) 7-Weilness 1 8-Weilness 2 9-Test Item 3 10-Test Item 3 12-Test Total (Avg)	1-Gender 2-Eye Color 3-Hai Color 4-Height (in) 5-Weight (lb) 6-Age (yr) 7-Wellness 1 8-Wellness 1 8-Wellness 2 9-Test Item 1 10-Test Item 3 11-Test Item 3 12-Test Total (Avg)	OK Cancel			
Spread Zoom Spread Zoom List1: List2:	Spread Zoom List3:	Spread Zoom List4:	Spread Zoom List5:	Spread Zoom List6:				
1 2	3							

2. On the *Statistics by Groups* (*Breakdown*) dialog, select *List of tables* tab, and click *Grouping variables*. For *List1* highlight *Gender*, for *List2* highlight *Eye Color*, and for *List3* highlight *Hair Color*.

MAND MISSIPIC

3. Click **OK**. Click **Dependent variables**. Highlight the *Height (in)* and *Weight (lb)* variables and click **OK**.



# Example Steps



CASPI (SSPI) (C)

Breakdown resi	Brea	akdown Tabl N=10	e of Descript )0 (No missir	ive Statistics Ig data in dep	(Characteris ). var. list)	tics.sta)	_
Breakdown	Gender	Eye Color	Hair Color	Height (in) Means	Height (in) N	Height (in) Sum	He S
	male	blue	brown	66.92857	14	937.000	
	male	blue	red	67.28571	7	471.000	:
	male	blue	black	70.75000	4	283.000	
	male	blue	blonde		0		
	male	green	brown	69.66667	3	209.000	
	male	green	red	68.00000	5	340.000	1
	male	green	black	67.00000	5	335.000	
	male	green	blonde		0		
	male	brown	brown	67.33333	3	202.000	
	male	brown	red	67.00000	2	134.000	-
			1				$\mathbf{F}$
		ndividual table: <u>V</u> ariab Dependent: I Grouping: I <u> <u>Codes</u> for</u>	s Lists of table les Height (in)-Weig Hair Color r grouping varial	is ht (lb) ples none			

4. On the List of tables tab, select Sums, Variances, Std. devs, Min & max and N in the Statistics section. Under Output tables, select Summary table of means. Click OK.

5. Resume analysis. Specify the *Individual tables* tab as shown. Click *OK*.

ise

#### StatSoft<sup>®</sup>



## **Example Steps**

> 12.875 0.720260 0.542287 96 1322.240 0.646114 0.587280



STRATTISTUCA

Statistics by Groups -	- Results: Cha	racteristics.sl	ta	li i	? _ ×				
DEPENDENT: 2 varial GROUPING: 1-Hair (	bles: Heigh Col(4): bro	t (in) Weig wn red blac	ht (lb) k blonde		En ±			6	. Clio dia
Quick Descriptives AN	IOVA & tests C	iorrelations   Pos	st-hoc	Su (Characteri	mmary stics.sta)			<mark>ิส</mark>	. Clic
Workbook1	1							7	
🗓 🚊 🖂 Basic Statistic		Breakdown	Table of Des	criptive Statis	tics (Charac	teristics.sta)	A.	ill o	Dro
🕞 🚊 🔄 Breakdow		N	l=100 (No m	issing data in	i dep. var. lis	t)			. 116
Break	Hair Color	Height (in)	Height (in)	Height (in)	Height (in)	Height (in)	Height (in)		
Break	ļ	Means	N	Sum	Std.Dev.	Variance	Minimum		kov
	brown	68.04348	46	3130.000	3.189960	10.17585	58.00000		ney
	red	67.13636	22	1477.000	3.907203	15.26623	57.00000		
	black	68.44444	27	1848.000	3.826359	14.64103	61.00000		_
	blonde	66.80000	5	334.000	4.381780	19.20000	61.00000		Δn
	All Grps	67.89000	100	6789.000	3.572892	12.76556	57.00000		
							▼		
Lan d	Breakdow	n Table 🔠 🖤	orkbook1* - #	Analysis of Va	riance (Char	acteristics.sta	1]		
		🔁 🗠	orkbook1*			Ana	lveie of Variand	o (Chara	ctarietice eta)
		⊡··∰	Basic Statistic			Mark	od offorte aro e	e (Chara ignifican	totn < 05001
		Ē	🔄 Breakdow	1	22	df M		df	MS
			Break	Variable	Effect	Effect Effe	o oo oot Error	Error	Error
			Break	Hoight (in)	27.819		2731 1236 C	96	12,875,0.7
			- Analy:	Woight (lh)	27.013	3 954	2731 1230.0	90	12:070 0:73
					2302.333	5 054.	5170 120900.0	30	1322.240 0.04
				<u> </u>					
		•	E F	Analysis (	of Variance (Cha	aracteristics.sta)			

- 6. Click **OK** to display results dialog.
- 7. Click the **Summary** button
- 8. Press CTRL+R on your keyboard, then click the

#### Analysis of Variance button





# Example Steps



 Resume the analysis. Click the *Categorized box & whisker plot* button. Select both dependent variables and click *OK*.

STRATITISTRIC,

The **ANOVA & tests** tab and the **Post-hoc** tab are also available if you would like to further test for equal variances or to see which levels differ from each other (via post-hocs).



## **Nonparametric Methods**

STRATTISTUCIA

- Used when the researcher knows nothing about the parameters of the variable of interest in the population or in cases where the usual parametric assumptions do not hold
- Do not rely on the estimation of parameters

#### STRATISTICA

### Example 10: Descriptives & Percentiles



#### **Ordinal descriptive statistics**

computes a wide variety of measures of location (mean, median, mode, etc.) and dispersion (variance, average deviation, quartile range, etc.) to provide a more "complete picture" of your data

For this example, use *Characteristics.sta*.



Example Steps

Monparam	etric Statist	ics: Char	acteristi	cs.sta		?_×		
Quick				1		OK)		
🔜 🛗 🛄 🖓 🖸 🕯	escriptive S	tatistics:	Charac	teristics.	sta		?.	-   ×
	uick   <mark>⊇ ⊻</mark> ariable Compute perc <u>F</u> irst:	es: Te entile bour 25.	st Item 1-7 ndaries	Fest Item 3	}		Su <u>m</u> ma Cancel Options	ıy •
	<u>S</u> econd:	75.	<b>▲</b>			SELECT CRSES		<u>w</u>
Workbook1.stw*	- Descriptive	Statistics (	Character	istics.sta)			_ 🗆 ×	cs
📉 Workbook Listw*		D	escriptive	Statistics	(Character	istics.sta)		ics
Nonparam	Variable	Mean	Valid N	Median	Mode	Frequency of Mode	Min	c
	Test Item 1	9.67000	100	10.00000	14.00000	15	5.	
	Test Item 2	19.27000	100	19.00000	15.00000 multiple	15	15.	
	rest item 5	14.23000	100	14.00000	multiple		10.	
							•	
<u>∢</u>	Descriptive	Statistics (Ch	aracteristics	sta)				1

 From the Statistics menu, select Nonparametrics to display the Nonparametric Statistics (Startup Panel).

CAJNISINI(C)

 Choose Ordinal descriptive statistics (median, mode, ...) and click OK. On the Descriptive Statistics dialog, click Variables and select Test Item 1, Test Item 2, and Test Item 3. Click OK.

3. Click the *Summary* button.









4. Press CTRL+R and click the Box & whisker plot for all variables button. On the Box-Whisker Type dialog, select the Median/Quart.
/Range option button and click OK.

CASPICSSICIC

 Resume the analysis. Enter 10 and 90 into the *First* and *Second* boxes, respectively, under *Compute percentile boundaries*. Click *Summary*.



# Example 11: Comparing 2 Independent Samples (Groups)



Nonparametric alternatives for the t-test for independent samples available in *STATISTICA* are the Wald-Wolfowitz runs test, the Mann-Whitney U test, and the Kolmogorov-Smirnov two sample test.

## For this example, use *Characteristics.sta*.





# Example Steps

Comparing Two	Groups: Ch	naracteristi	cs.sta	? _ )	< l	
	]			-W U <u>t</u> est	]	
Dependent: Test It Grouping: Gende	em 1 er		Ca	ancel	L	
Codes for: Group <u>1</u> :	"male"	Group	<u>2</u> : ["female	'		
Quick			🔊 🛛	Iptions 🔻		
Wald-\	Volfowit <u>z</u> runs	test	SELECT CRSES S	6 ⊻		
Kolmogorov-	Smirnov two-s	ample test	Double-cli	ck on the		
Workbook1* - Man	in-Whitney U T	est (Charact	eristics.sta)			_
Workbook1* ⊡ ··· ⊕ ·· ⊕ ·· ⊕ ·· ⊕ ·· ⊕ ·· ⊕ ·· ⊕ ·		Man	n-Whitney U By va	Test (Char riable Gen	acteristics.« der	sta) oo
		Rank Sum	Rank Sum	U	Z	p-level
	variable	male	female			
	Test Item 1	2612.500	2437.500	1234.500	-0.093141	0.92579
	1•1			1		
	Mann-White	ney U Test (Cha	racteristics.sta)			

 From the Statistics menu, select Nonparametrics, then choose Comparing two independent samples (groups) and click OK.

STRATE STRICA

- 2. Click the Variables button
  and highlight Test Item 1 in
  the Dependent variable list
  box and Gender in the Indep.
  (grouping) variable box.
  Click OK.
- 3. Click the *M-W U test* button.





## Example Steps



4. Now, look at a box plot to visualize the results. To do this, resume the analysis. Click the *Box & Whisker plot by group* button on the *Comparing Two Groups* dialog.

SILAINISILI

The plot confirms our conclusion that there is little difference between males and females in respect to *Test Item 1*.

**60** 



# Example 12: Comparing Two Dependent Samples



When the data are nonparametric, use the Sign test or Wilcoxon's matched pairs test to compare two dependent samples.

For this example, use *Characteristics.sta*.



# Example Steps



STRATE STRICA

🎑 Comparing two vari	ables: Characteristics.s	ta  ?	_ ×		
🕵 Variables	I	🖪 🛛 Sign	test		
List 1: Test Item 1		Cance	el		
List 2: Test Item 2-Test	Item 3				
Quick	<u> </u>	Option	ns 🔻		
	I SEI	ECT _ L	<b>a</b> 1		
Sign te:	st <u>CR</u>	SES S			
Workbook1* - Sign	n Test (Characteristics.sta)				_ [
Workbook1*		Sign Marked te	Test (Cha sts are sig	racteristics Inificant at	sta) p <.05000
Sign Tes	Pair of Variables	No. of Non-ties	Percent v < V	Z	p-level
Sign Te:	Pair of Variables Test Item 1 & Test Item 2	No. of Non-ties 100	Percent v < V 100.0000	Z 9.900000	p-level
Sign Tet	Pair of Variables Test Item 1 & Test Item 2 Test Item 1 & Test Item 3	No. of Non-ties 100 92	Percent v < ∨ 100.0000 86.9565	Z 9.900000 6.985233	p-level 0.000000 0.000000
Sign Te:	Pair of Variables Test Item 1 & Test Item 2 Test Item 1 & Test Item 3	No. of Non-ties 100 92	Percent v < ∨ 100.0000 86.9565	Z 9.900000 6.985233	p-level
Sign Tes	Pair of Variables Test Item 1 & Test Item 2 Test Item 1 & Test Item 3	No. of Non-ties 100 92	Percent v < V 100.0000 86.9565	Z 9.900000 6.985233	p-level 0.000000 0.000000

- 1. Select *Nonparametrics* from the *Statistics* menu.
- 2. Choose **Comparing two** dependent samples (variables) and click **OK**.
- Select the variable Test Item 1 in the First variable list box and select both Test Item 2 and Test Item 3 in the Second variable list box.
- 4. Click the Sign test button.





## Example Steps



SILATUSILI



5. Resume the analysis and create a box plot to graphically visualize the results. Click the **Box &** whisker plots for all variables button. Select all three variables and click **OK**. Select the *Median/Quart./ Range* option button on the Box-Whisker Type dialog, and click **OK** to produce the graph.





## Example 13: Comparing Multiple Independent Samples



The nonparametric equivalents to the one-way ANOVA method are the Kruskal-Wallis analysis of ranks and the Median test.

For this example, use *Characteristics.sta*.

#### M StatSoft



Example Steps



UNIISSINIC

Cuick	llis ANOVA and Medi	an Test:	Characte	eristics.st	a <b>? _ &gt;</b> Summary	<b>≤</b> ]
🖳 🛛 🖳	is				lancel	
Dependent v	ariables: Test Item 1				Dotions 🔻	1
Grouping vari	able: Hair Color x1* - Kruskal-Wallis ANOV	A by Ranks	; Test Item	1 (Charac	teristics.sta)	
Workbook1       Image: Strate Str	* ametrics (I skal-Wallis Kruskal-W Median T Depend.: Co Test Item 1 brown red	skal-Wallis ependent (g skal-Wallis de Valid N 1 46 2 22	ANOVA b grouping) va test: H ( 3 <b>Sum of</b> <u>Ranks</u> 2239.500 1275.500	y Ranks; 1 ariable: Hai }, N= 100) =	Test Item 1 ( r Color =2.104098 p	Characteri ) =.5511
Workbook1* - M	edian Test, Overall Media	n = 10.000	0; Test Iter	n 1 (Charao	cteristics.sta	
Workbook		t, Overall I Indepe Chi-S	Median = 1 indent (grou quare = 1.9	0.0000; Te uping) varia 308146, df:	st Item 1 (C ble: Hair Co = 3, p = .59*	haract 🗾 Ior 17
Media	Dependent: Test Item 1	brown	red	black	blonde	Tota
	<= Median: observed expected	28.00000	10.00000	15.00000	2.000000	55.00
	obsexp. > Median: observed	2.70000	-2.10000	12.00000	3.000000	45.00
	expected obsexp.	-2.70000	2.10000	-0.15000	0.750000	100.00
		40.00000	22.00000	27.00000	5.000000	

 Select Nonparametrics from the Statistics menu.

N/A

- 2. Choose *Comparing multiple indep. samples (groups)*, then click the *Variables* button and select *Test Item 1* as the *Dependent variable list*, and *Hair Color* as the *Indep. (grouping) variabl*e. Click *OK*.
- Click Summary. The Median test results are displayed next in the workbook



## Example Steps



 Now, look at a box plot to graphically visualize the results. This graph is available via the **Box &** whisker button on the Kruskal-Wallis ANOVA and Median Test dialog.

SILAINISILI

Notice that each analysis in this workbook was a new analysis. If you placed all of the documents in the same workbook, your workbook could look similar to this.



## **Example 14: Frequency Tables**



Frequency or one-way tables represent the simplest method for analyzing categorical (nominal) data. They are often used as an exploratory procedure to review how different categories of values are distributed in the sample.

SIN MANNISSIN NI (C)A

For this example, use Sports.sta.

#### M StatSoft



# Example Steps



CADDIS DDI

Frequency Tables: Sport	ts.sta	<u> ? _   ×</u>					
Variables: Football		Summary					
Quick Advanced Options	Descr. Normality	Cancel					
Workbook1.stw* - I	Frequency table: Footba	all: "Watching footba	all" (Sports.sta)	_ 🗆 ×			
Basic Statistics/T	Category	Frequ Coun	Frequency table: Football: "Watch Count Cumulative Percent Cur Count P				
	Always: Always intere	sted 3	9 39 39.00000 6 55 16.00000				
<u>u</u>	Frequency Tables: Sports.sta						
		otball Marathon	Summary	] =			
	Quick Advanced Op	tions Descr. Normality	y Cancel				
	Descrip	otive statistics	🔁 Options 🔻	╻╞╴┚			
	Box & whisker p						
	Normal pro	obability plots ( <u>2)</u>					
	Half-normal (	probability plots ( <u>3</u> )					
	Detrended norm	nal probability plots ( <u>4</u> )					
	💯 3D histograms, b	ivariate distributions ( <u>5</u> )		1			
			<u>MD</u> deletion	1			
			C Casewise Pairwise				
				-			

- Choose Frequency tables from the Basic Statistics and Tables (Startup Panel). Select Football as the Variable, and click OK.
- 2. Click the Summary button
- 3. Press CTRL+R and select Football and Marathon.
- Return to the *Frequency Tables* dialog and select the *Descr.* tab.



# Example Steps



5. Click the **3D histograms**, **bivariate distributions** button. Select *Football* for the first variable list and *Marathon* for the second variable list.

CALICIES ICIC

Click OK.

From this graph, you can see that regarding the two sports selected, the largest group of men said that they never watch marathons on television and always watch football.



# Example 15: Crosstabulation Tables



Crosstabulation allows us to examine frequencies of observations that belong to specific categories on more than one variable. By examining these frequencies, we can identify relationships between variables.

For this example, use Sports.sta.

70



## Example Steps



STRATE STRICA

Quick	sstabulation Tables Ro Advanced Options Summary: Review <u>s</u> umma Categorized histogra	esults: Sports. ary tables	sta			? _ ×       Summary       Cancel       Options ▼       npute Max.		
6	Workbook1.stw <sup>*</sup> - S Workbook1.stw <sup>*</sup> ⊡ - → Basic Statistics/T ⊡ - → Crosstabulatic	Summary Frequency Table (Sports.sta)						
	Summary	Football	Baseball Always	Baseball Usually	Baseball Sometimes	Baseball Never	Row Totals	
		Always	24	8	5	2	39	
		Usually	2	5	7	2	16	
		Sometimes	2	3	19	2	26	
		Never	0	1	6	12	19	
		All Grps	28	17	37	18	100	
								7
								×
		📰 Summary F	requency Tab	ole (Sports.sta	)			

- 1. Choose *Tables and banners* from the *Startup Panel* and click *OK*.
- 2. On the **Crosstabulation** tab, click the **Specify tables** (select variables), and select Football in List1 and Baseball in List2. Click OK. Click OK on the **Crosstabulation** Tables dialog.
- 3. On the **Crosstabultaion Tables Results** dialog, click **Summary**.



# Example Steps



 On the Options tab, select the Pearson & M-L Chisquare check box. On the Advanced tab, click Detailed two-way tables.

SILA ILISILI

 Press CTRL+R, then click 3D histograms on the Advanced tab.
#### A StatSoft



# Example Steps



PAJPI (SIPI (C)

Workbook1* - Summary Frequency Table (Sports.sta)						
workbook1* ⊡ ∰ Basic Statistics/T ⊡ ∰ Crosstabulati	Summary Frequency Table (Sports.sta) Marked cells have counts > 10 (Marginal summaries are not marked)					
2-Way Si Statistics Statistics Summary	Football	Baseball	Basketball Always	Basketball Usually	Basketball Sometimes	Basket Neve
	<u>Always</u>	Always	11	8	2	
	Always	Usually	2	2	3	
	Always	Sometimes	2	0	0	
	Always	Never	0	0	0	
	lotal		15	10	5	
	Usually	Always	0	0	1	
	Usually	Usually	0	1	4	
	Usually	Sometimes	1	0	5	
	<u>Usually</u>	Never	0	0	1	
	Total		1	1	11	
	Sometimes	Always	0	1	1	
	Sometimes	Usually	0	0	2	
	Sometimes	Sometimes	2	2	12	
	Sometimes	Never	0	0	1	
	Total		2	3	16	
🔹 💽 👘 Bivariate Distribution: Football x Baseball 🏢 Summary Frequency Table (Sports.sta)						

- Press CTRL+R, then cancel to return to *Crosstabulation Tables* dialog. Specify new variables: *Football* on *List1*, *Baseball* on *List2*, and *Basketball* on *List3*.
- Click OK to run the analysis. Click Summary. In this survey, 12 men said they watch baseball sometimes, football sometimes, and basketball sometimes.



### **Example 16: ByGroup Analysis**



Use **ByGroup** analysis to do the wide variety of analyses and graphs in *STATISTICA* while taking a grouping variable into consideration.

STRATEISTEICA

For this example, use *Characteristics.sta*.

#### M StatSoft



## Example Steps



 Select Statistics - ByGroup Analysis. In the ByGroup Statistics Browser select Basic Statistics and Tables in the left pane and Correlation Matrices in the right pane.
Click OK.

SIPAIRISIPICA

 On the Quick tab, click Variables and select Height(in), Weight(lb), and Age(yr) in the first list; click OK. Click By Variables and select Gender. Click OK.

#### A StatSoft



## Example Steps



3. Select the **General** tab and select **All results** for the **Detail of computed results reported**.

STRATISICIA

- 4. Click **OK** to display the results.
- When you compare the results for *Males* and *Females*, you can see that only the correlation between *Weight* and *Age* for *Males* was significant.