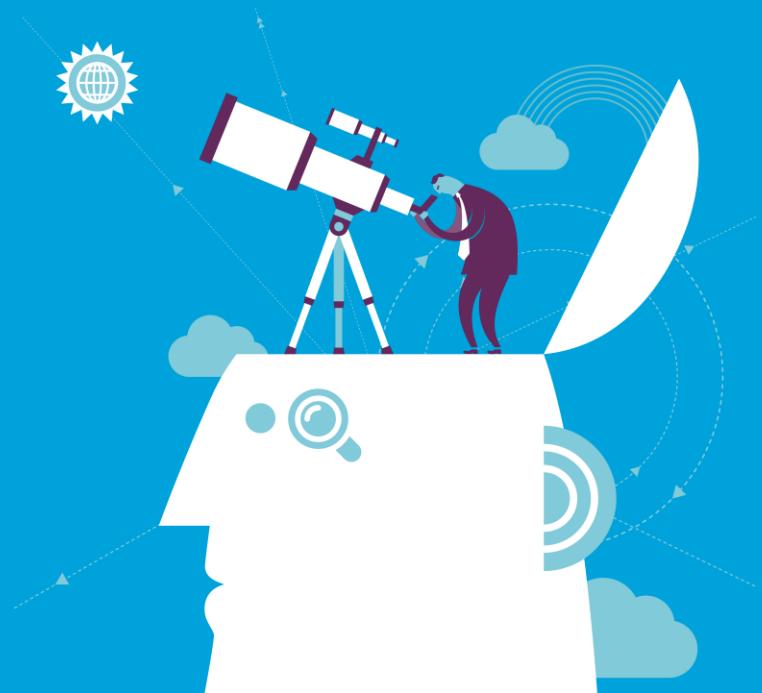


Emperical Methods for Marketing Research and Analytics Using

Prof. Dr. Martin Wetzels
Maastricht University



About Me



- ▶ **Name:** **Prof. Dr. Martin Wetzels**
Professor in Marketing and Supply Chain Research

- ▶ **Address:** **Maastricht University**
School of Business and Economics
Department of Marketing and SCM
P.O. Box 616, 6200 MD Maastricht
Tongersestraat 53, F1.05
T: +31 43 388 3250/3839
F: +31 43 388 4918
E: m.wetzels@maastrichtuniversity.nl
W: <https://www.maastrichtuniversity.nl/m.wetzels>



Using for Basic Analysis

**Prof. Dr. Martin Wetzels
Maastricht University**

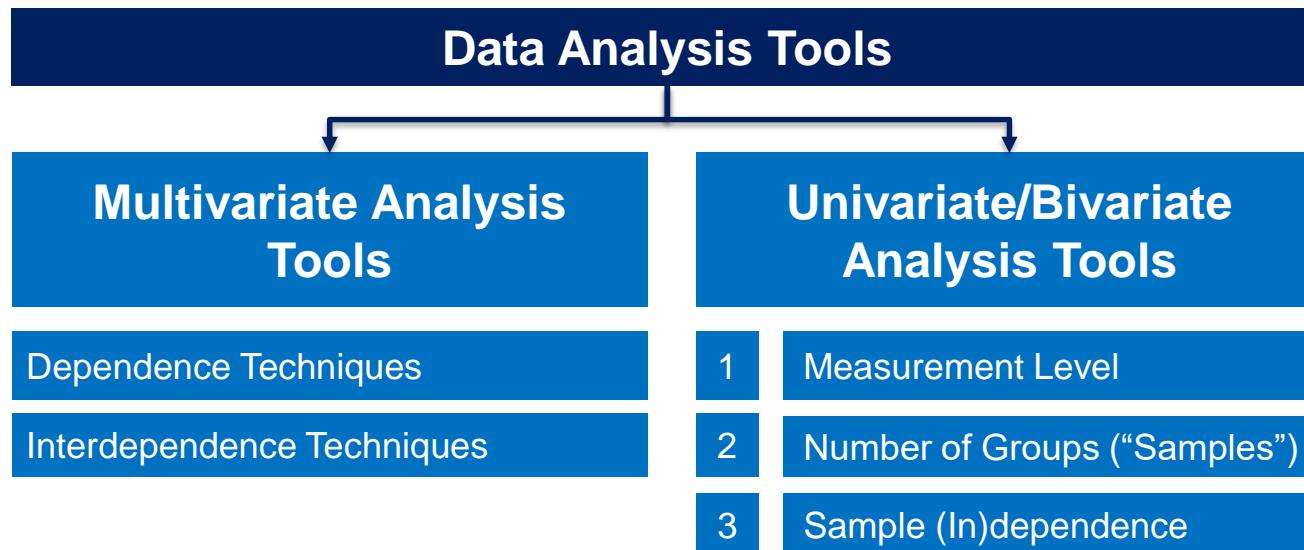


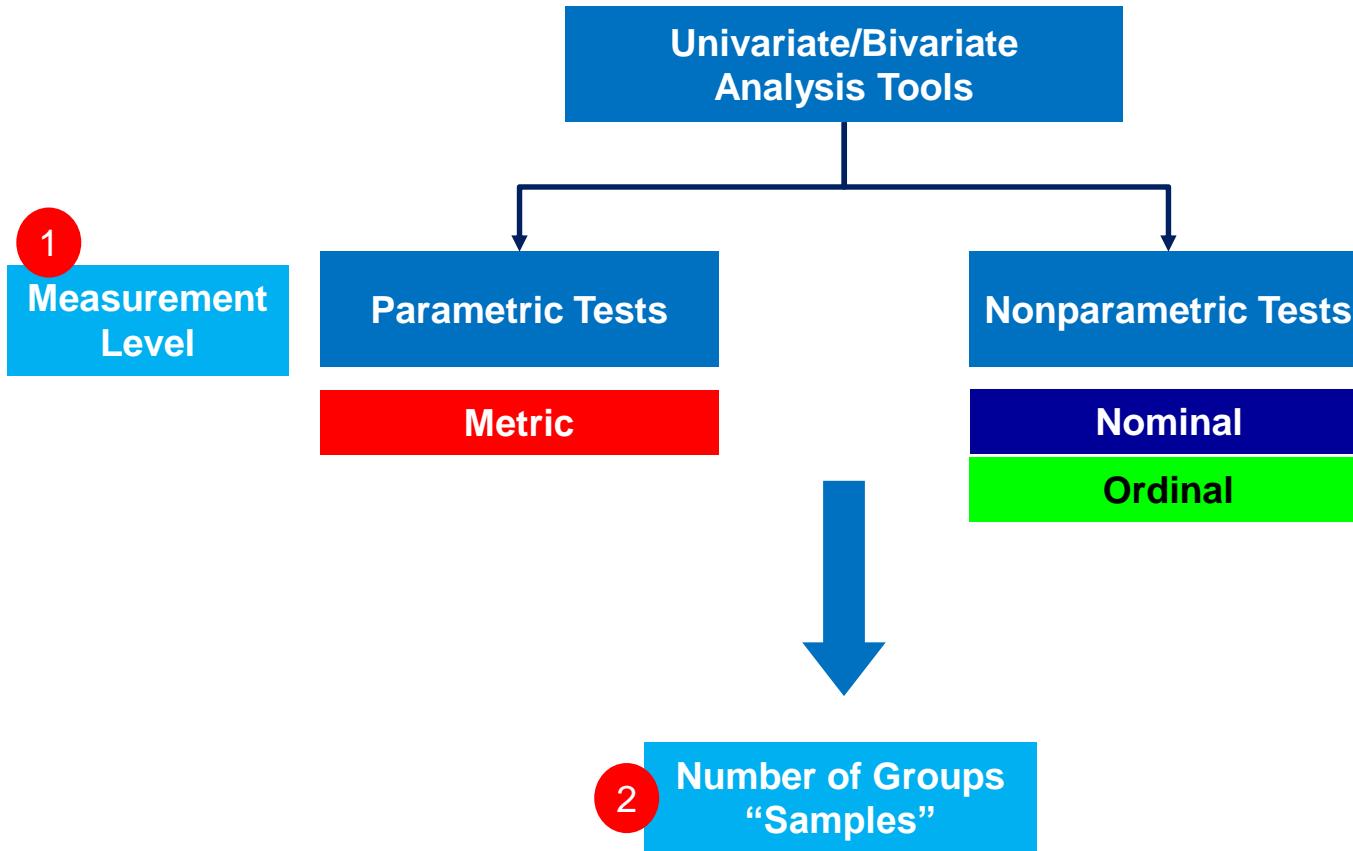
Course Outline

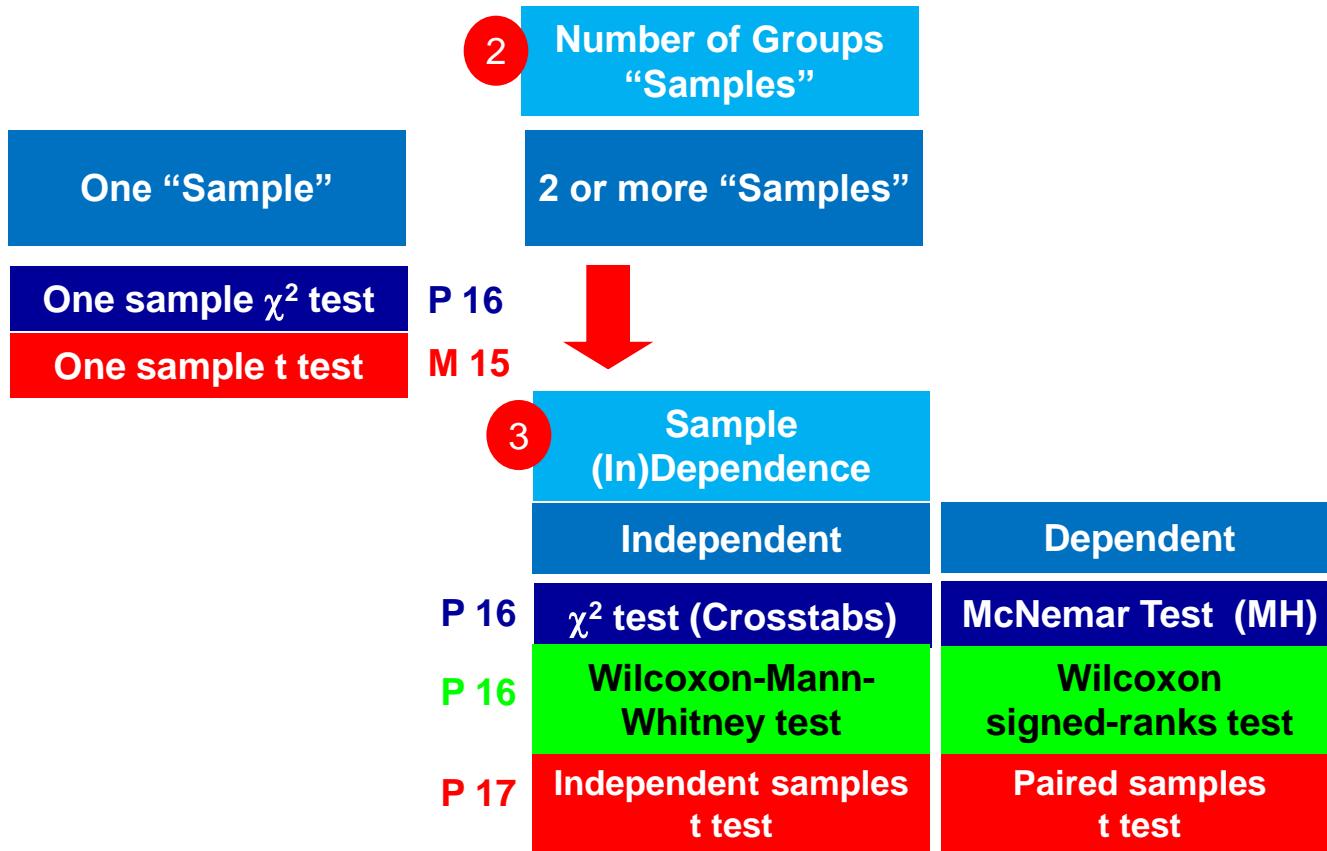
Session:	TOPIC:
DAY 1	
Session 1	INTRODUCING MULTIVARIATE ANALYSIS AND R
Session 2	USING R FOR BASIC ANALYSIS
DAY 2	
Session 3	USING R FOR AN(C)OVA
Session 4	USING R FOR REGRESSION ANALYSIS
DAY 3	
Session 5	USING R FOR SCALING AND FACTOR ANALYSIS
Session 6	USING R FOR SEM and PLS PATH MODELING

Overview over Data Analysis Tools for Marketing Research

Malhotra (2010); Pallant (2016)







Hypothesis Testing Procedure

Malhotra (2010); Pallant (2016)

1 Research Problem → H_0 : “No Effect”
 H_1 : “Effect”

2 ▪ Choose Appropriate Test Statistic
▪ Specify significance level ($\alpha \rightarrow 0.05$)

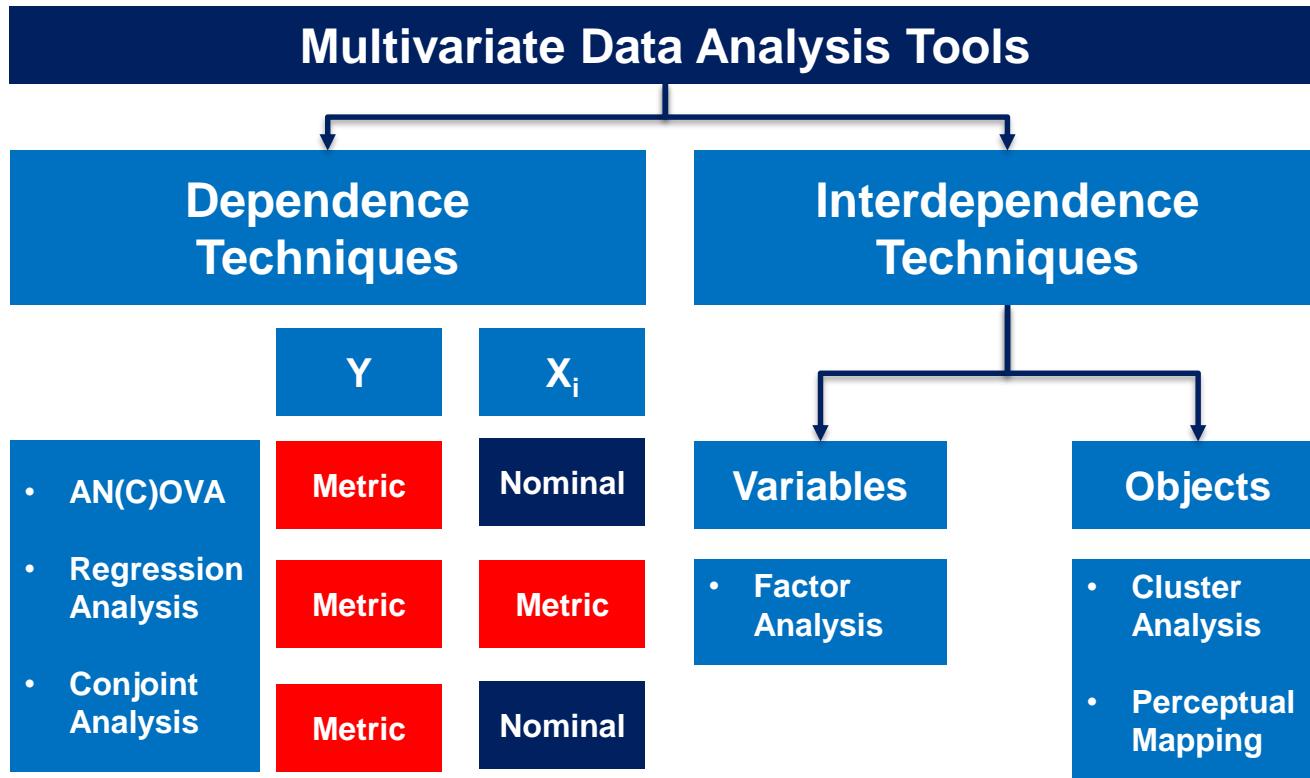
3 Collect Data

4 Calculate Test Statistic and Probability Value (p value)

5 If $p \leq \alpha$, **reject H_0**
If $p > \alpha$, **do not reject H_0**

Data Analysis Tools

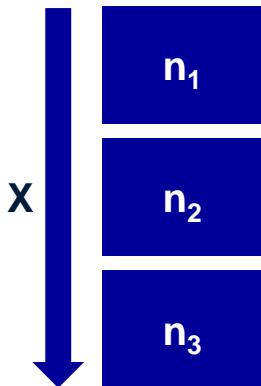
Hair et al. (2018); Malhotra (2010)



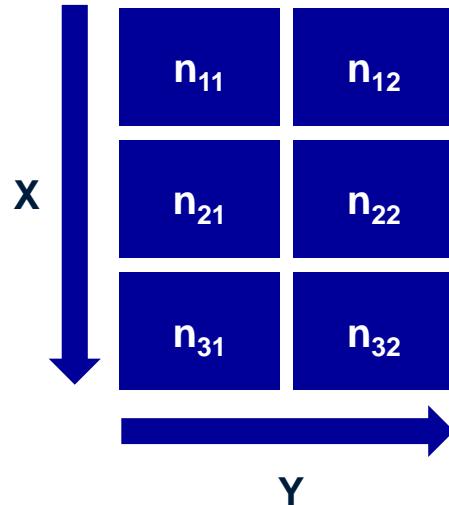
Refresher: Comparing Frequencies: χ^2 Tests

Malhotra (2010); Pallant (2016)

One-Sample χ^2 Test



“Two-Sample” χ^2 Test
for Cross-Tabulations



χ^2 Tests

Malhotra (2010); Pallant (2016)

- ▶ One-Sample χ^2 Test
 - ▶ Assumptions
 - ▶ Independent (Random) Sample
 - ▶ Nominal Data with two or more categories
 - ▶ Theory for Expected Distribution
 - ▶ Hypotheses
 - H_0 : Expected Frequencies (E) = Observed Frequencies (O)
 - H_1 : Expected Frequencies (E) \neq Observed Frequencies (O)

χ^2 Tests

Malhotra (2010); Pallant (2016)

► One-Sample χ^2 Test

- Limiting conditions one-sample chi-square test
 - $df = 1$ (or $k = 2$)
Each Expected Frequency (E) should be ≥ 5
 - $df > 1$ (or $k > 2$)
No more than 20 percent of Expected Frequencies (E) < 5
No Expected Frequencies (E) < 1

SPSS evaluates these limiting conditions



χ^2 Tests

Malhotra (2010); Pallant (2016)

- ▶ “Two-sample” χ^2 Test for Cross-Tabulations
 - ▶ Assumptions
 - ▶ Independent (Random) Samples
 - ▶ Nominal Data with two or more categories (for X and Y)
 - ▶ Theory for Expected Distribution (*Independence!*)
 - ▶ Hypotheses
 - H_0 : X and Y are independent
 - H_1 : X and Y are not independent

χ^2 Tests

Malhotra (2010); Pallant (2016)

- ▶ “Two-sample” χ^2 Test for Cross-Tabulations
 - ▶ Limiting Conditions
 - ▶ $df = 1$ (2*2 cross-table)
 - $N \geq 20$ → Fischer Exact Test
 - $20 < N \leq 40$ $\rightarrow E_{ij} \geq 5$ → χ^2 Test
 - $\rightarrow E_{ij} < 5$ → Fischer Exact Test
 - $N > 40$ → χ^2 Test with continuity correction
 - ▶ $df > 1$ ($r*k$ cross-table)
 - $E_{ij} < 5$ → Fewer than 20% of the Cells
 - $E_{ij} < 1$ → No Cell

SPSS evaluates these limiting conditions and presents the appropriate test

SPSS Data File: Internet Usage Data



Malhotra (2010, Table 15.1)

Table 15.1 Input.sav [DataSet1] - IBM SPSS Statistics Data Editor

	number	sex	familiar	iusage	ilatitude	tlatitude	ishopping	ibanking	iusagegr	var									
1	1	1	7	14	7	6	1	1	2.00										
2	2	2	2	2	3	3	2	2	1.00										
3	3	2	3	3	4	3	1	2	1.00										
4	4	2	3	3	7	5	1	2	1.00										
5	5	1	7	13	7	7	1	1	2.00										
6	6	2	4	6	5	4	1	2	2.00										
7	7	2	2	2	4	5	2	2	1.00										
8	8	2	3	6	5	4	2	2	2.00										
9	9	2	3	6	6	4	1	2	2.00										
10	10	1	9	15	7	6	1	2	2.00										
11	11	2	4	3	4	3	2	2	1.00										
12	12	2	5	4	6	4	2	2	1.00										
13	13	1	6	9	6	5	2	1	2.00										
14	14	1	6	8	3	2	2	2	2.00										
15	15	1	6	5	5	4	1	2	1.00										
16	16	2	4	3	4	3	2	2	1.00										
17	17	1	6	9	5	3	1	1	2.00										
18	18	1	4	4	5	4	1	2	1.00										
19	19	1	7	14	6	6	1	1	2.00										
20	20	2	6	6	6	4	2	2	2.00										
21	21	1	6	9	4	2	2	2	2.00										
22	22	1	5	5	5	4	2	1	1.00										
23	23	2	3	2	4	2	2	2	1.00										
24	24	1	7	15	6	6	1	1	2.00										
25	25	2	6	6	5	3	1	2	2.00										
26	26	1	6	13	6	6	1	1	2.00										
27	27	2	5	4	5	5	1	1	1.00										
28	28	2	4	2	2	2	2	2	1.00										

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

One-Sample χ^2 Test

$$\sum \frac{\alpha}{n}$$

The screenshot shows the IBM SPSS Statistics Data Editor interface. The menu bar is visible at the top, with 'Analyze' being the active tab. Under the 'Analyze' tab, the 'Nonparametric Tests' option is highlighted. A red arrow originates from the text 'One-Sample χ^2 Test' and points to the 'Chi-square...' option in the dropdown menu under 'Legacy Dialogs'. The main data grid on the right displays a dataset with 27 rows and 5 columns, labeled 'number', 'sex', 'ishopping', 'ibanking', and 'usagegr'. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode ON'.



χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

$$\sum \frac{\alpha}{n}$$

EXACT TEST
Limiting Conditions!!!

One-sample χ^2 Test
 $H_0: M=30\%, F=70\%$

Table 15.1 Inputs.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

number sex familiar usage iatitude tlatitude ishopping ibanking iusagegr var var

1 1 1 7 14 7 6 1 1 2.00
2 2 2 2 2 3 3 2 2 1.00
3 3 2 2 2 3 3 2 2 1.00
4 4 2 2 2 2 2 2 2 1.00
5 5 1 2 2 2 2 2 2 2 2.00
6 6 2 2 2 2 2 2 2 2 2.00
7 7 2 2 2 2 2 2 2 2 2.00
8 8 2 2 2 2 2 2 2 2 2.00
9 9 2 2 2 2 2 2 2 2 2.00
10 10 1 2 2 2 2 2 2 2 2.00
11 11 2 2 2 2 2 2 2 2 2.00
12 12 2 2 2 2 2 2 2 2 2.00
13 13 1 2 2 2 2 2 2 2 2.00
14 14 1 2 2 2 2 2 2 2 2.00
15 15 1 2 2 2 2 2 2 2 2.00
16 16 2 2 2 2 2 2 2 2 2.00
17 17 1 2 2 2 2 2 2 2 2.00
18 18 1 2 2 2 2 2 2 2 2.00
19 19 1 2 2 2 2 2 2 2 2.00
20 20 2 2 2 2 2 2 2 2 2.00
21 21 1 2 2 2 2 2 2 2 2.00
22 22 1 2 2 2 2 2 2 2 2.00
23 23 2 2 2 2 2 2 2 2 2.00
24 24 1 7 15 6 6 1 1 1 2.00
25 25 2 6 6 5 3 1 2 2.00
26 26 1 6 13 6 6 1 1 2.00
27 27 2 5 4 5 5 1 1 1.00
... 20 2 4 0 2 0 0 0 0 0 0.00

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode.ON

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

One-sample χ^2 Test
 $H_0: M=30\% (=9), F=70\% (=21)$

Sex			
	Observed N	Expected N	Residual
Male	15	9.0	6.0
Female	15	21.0	-6.0
Total	30		

Sex			
	Observed N	Expected N	Residual
Male	15	15.0	.0
Female	15	15.0	.0
Total	30		

	Sex
Chi-Square	5.714 ^a
df	1
Asymp. Sig.	.017
Exact Sig.	.026
Point Probability	.018

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 9.0.

$$\chi^2 (1) = 5.714 \text{ (p=0.017)}$$

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{6^2}{9} + \frac{6^2}{21} = 5.714$$

!

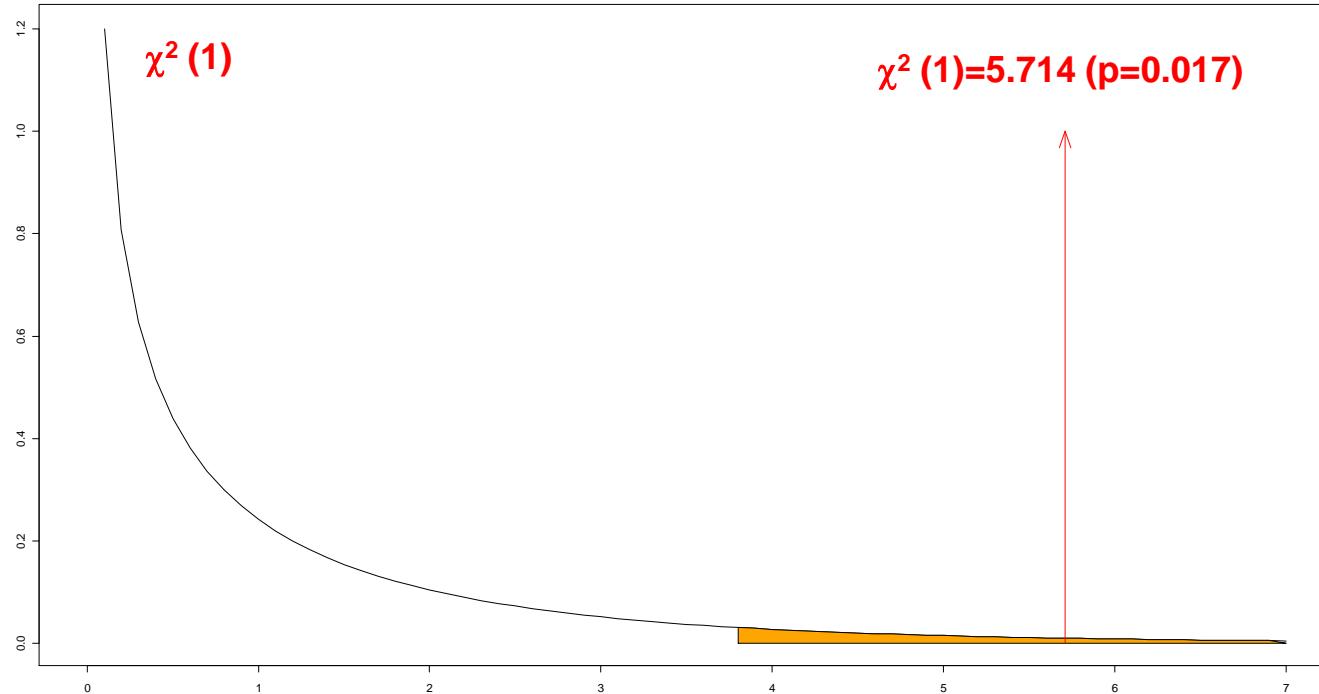
	Sex
Chi-Square	.000 ^a
df	1
Asymp. Sig.	1.000
Exact Sig.	1.000
Point Probability	.144

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 15.0.

!

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)



χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

χ^2 Test for Crosstabulations

$$\sum \frac{\alpha}{\cdot}$$

H_0 : Sex and Internet shopping are independent

The screenshot shows the IBM SPSS Statistics Data Editor interface. On the left, there is a data view window displaying a table with columns 'number' and 'sex'. The 'sex' column has values 1 and 2. On the right, there is a large empty grid labeled 'Visible: 9 of 9 Variables' with columns labeled 'var'.

The top menu bar is visible with options: File, Edit, View, Date, Transform, Analyze, Direct Marketing, Graphs, Utilities, Extensions, Window, Help. The 'Analyze' menu is currently open, showing various statistical analysis options. The 'Crosstabs...' option is highlighted with a yellow box and a red arrow pointing towards it from the top right corner of the slide.

The 'Crosstabs...' option is located under the 'Descriptive Statistics' section of the Analyze menu. Other options in this section include Frequencies..., Descriptives..., Explore..., and TURF Analysis. Below this section, other menu items like General Linear Model, Mixed Models, Correlate, Regression, Loglinear, Neural Networks, Classify, Dimension Reduction, Scale, Nonparametric Tests, Forecasting, Survival, Multiple Response, Missing Value Analysis..., Multiple Imputation, Complex Samples, Simulation..., Quality Control, ROC Curve..., RanFor Prediction, RanFor Estimation..., and Spatial and Temporal Modeling... are listed.

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

$$\sum \frac{\alpha}{\cdot}$$

The image shows two overlapping dialog boxes from SPSS:

Crosstabs dialog box (left):

- Row(s): Sex [sex]
- Column(s): Internet Shopping [isho...]
- Exact... button
- Statistics... button (highlighted by a red arrow)
- Cells... button
- Format... button
- Layer 1 of 1
- Previous and Next buttons
- Display clustered bar charts checkbox (checked)
- Suppress tables checkbox
- OK, Paste, Reset, Cancel, Help buttons

Crosstabs: Statistics dialog box (right):

- Chi-square checkbox (checked)
- Correlations checkbox
- Nominal group:
 - Contingency coefficient checkbox (checked)
 - Phi and Cramer's V checkbox (checked)
 - Lambda checkbox
 - Uncertainty coefficient checkbox
- Ordinal group:
 - Gamma checkbox
 - Somers' d checkbox
 - Kendall's tau-b checkbox
 - Kendall's tau-c checkbox
- Nominal by Interval group:
 - Eta checkbox
 - Kappa checkbox
 - Risk checkbox
 - McNemar checkbox
- Cochran's and Mantel-Haenszel statistics checkbox
- Test common odds ratio equals: 1 input field
- Continue, Cancel, Help buttons

$$\sum \frac{\alpha}{\cdot}$$

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

The image shows two overlapping SPSS dialog boxes. The left dialog is titled 'Crosstabs' and has 'Row(s): Sex [sex]' and 'Column(s): Internet Shopping [isho...]' selected. The right dialog is titled 'Crosstab Cell Display' and includes options for 'Counts' (Observed checked), 'z-test' (Compare column proportions checked, Adjust p-values (Bonferroni method) checked), 'Percentages' (Row checked), 'Residuals' (Unstandardized checked), and 'Noninteger Weights' (Round cell counts selected). A red arrow points from the 'Exact...' button in the Crosstabs dialog to the 'Exact' button in the Crosstab Cell Display dialog. Above the Crosstab Cell Display dialog, the text ' $> 2 \times 2$ ' is written in red.

**EXACT TEST
Limiting Conditions!!!**

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

Sex * Internet Shopping Crosstabulation				
Sex		Internet Shopping		Total
		Yes	No	
Male	Count	11	4	15
	% within Sex	73.3%	26.7%	100.0%
	% within Internet Shopping	64.7%	30.8%	50.0%
Female	Count	6	9	15
	% within Sex	40.0%	60.0%	100.0%
	% within Internet Shopping	35.3%	68.2%	50.0%
Total	Count	17	13	30
	% within Sex	56.7%	43.3%	100.0%
	% within Internet Shopping	100.0%	100.0%	100.0%

Crosstabulation

H_0 : Sex and Internet shopping are independent

$$\chi^2(1)=3.394 \text{ (} p=0.065 \text{)}$$

$$\chi^2_{cc}(1)=2.172 \text{ (} p=0.141 \rightarrow (2*2) \text{)}$$

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.394 ^a	1	.065		
Continuity Correction ^b	2.172	1	.141		
Likelihood Ratio	3.466	1	.063		
Fisher's Exact Test				.139	.070
Linear-by-Linear Association	3.281	1	.070		
N of Valid Cases	30				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.50.

b. Computed only for a 2x2 table

**EXACT TEST
Limiting Conditions!!!**

!

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

			ishopping		Total	
			1 Yes	2 No		
sex	1 Male	Count	22	8	30	
		% within sex	73.3%	26.7%	100.0%	
		% within ishopping	64.7%	30.8%	50.0%	
	2 Female	Count	12	18	30	
		% within sex	40.0%	60.0%	100.0%	
		% within ishopping	35.3%	69.2%	50.0%	
Total		Count	34	26	60	
		% within sex	56.7%	43.3%	100.0%	
		% within ishopping	100.0%	100.0%	100.0%	

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.787 ^a	1	.009		
Continuity Correction ^b	5.498	1	.019		
Likelihood Ratio	6.932	1	.008		
Fisher's Exact Test				.018	.009
Linear-by-Linear Association	6.674	1	.010		
N of Valid Cases	60				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 13.00.

b. Computed only for a 2x2 table

Crosstabulation Double Sample Size

H_0 : Sex and Internet shopping are independent

$$\chi^2(1)=6.787 \text{ (} p=0.009 \text{)}$$

$$\chi^2_{cc}(1)=5.498 \text{ (} p=0.019 \text{) } \rightarrow (2*2)$$

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Sex * Internet Shopping Crosstabulation

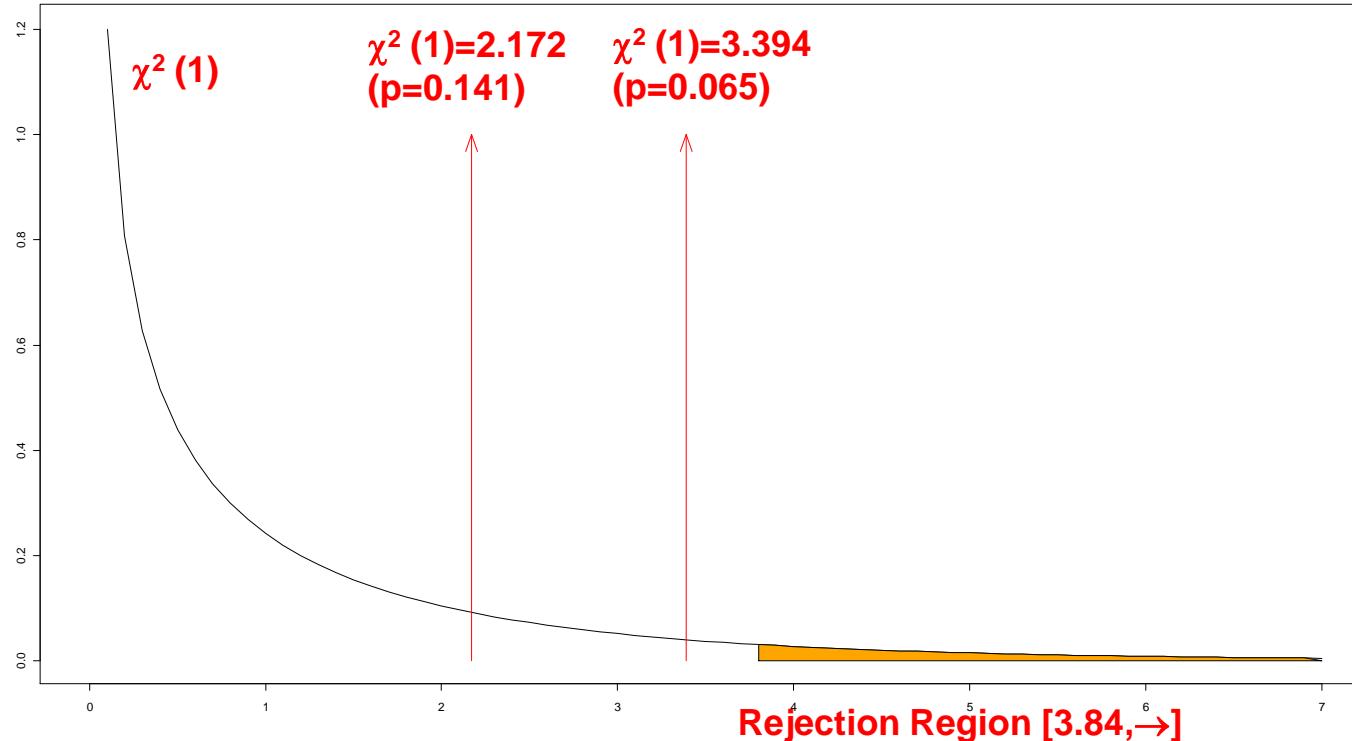
			Internet Shopping		Total	
Sex	Male	Count	Yes	No		
		Expected Count	8.5	6.5	15.0	
		Residual	2.5	-2.5		
	Female	Count	6	9	15	
		Expected Count	8.5	6.5	15.0	
		Residual	-2.5	2.5		
Total		Count	17	13	30	
		Expected Count	17.0	13.0	30.0	

$$E_{11} = \frac{(15*17)}{30} = 8.5$$

$$\chi^2 = \frac{2.5^2}{8.5} + \frac{-2.5^2}{6.5} + \frac{-2.5^2}{8.5} + \frac{2.5^2}{6.5} = 3.3937$$

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)



χ^2 Tests

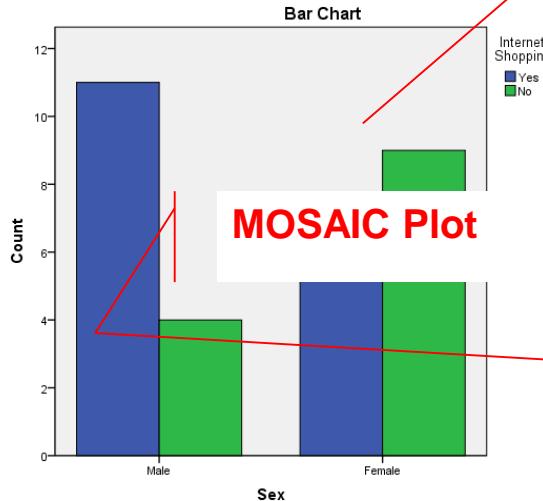
Internet Usage Data (Malhotra, 2010, Table 15.1)

$$\sum \frac{\alpha}{\beta}$$

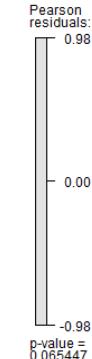
Symmetric Measures		Value	Approx. Sig.
Nominal by Nominal	Phi	.336	.065
	Cramer's V	.336	.065
	Contingency Coefficient	.319	.065
N of Valid Cases		30	

EFFECT SIZE!

Clustered Bar Chart



MOSAIC Plot



χ^2 Tests

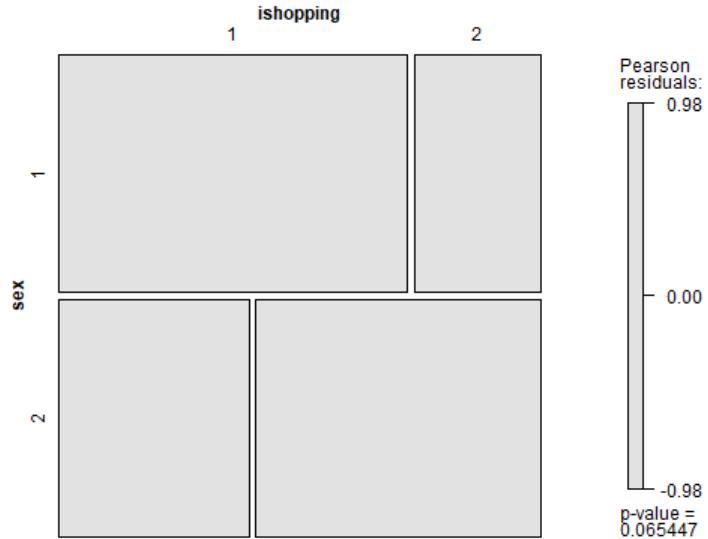
Internet Usage Data (Malhotra, 2010, Table 15.1)

$$\sum \frac{\alpha}{\cdot}$$

```
1  * Encoding: UTF-8.  
2  * Mosaic plot.  
3  
4  * Open SPSS Data File.  
5  
6  BEGIN PROGRAM R.  
7  
8  # Mosaic plot|  
9  
10 install.packages("vcd", repos="http://cran.rstudio.com")  
11 library(vcd)  
12  
13 DATA<-spssdata.GetDataFromSPSS()  
14  
15 mosaic(~sex + ishopping, data=DATA, shade=TRUE, legend=TRUE)  
16  
17 END PROGRAM.  
18
```

χ^2 Tests

Internet Usage Data (Malhotra, 2010, Table 15.1)



χ^2 Tests

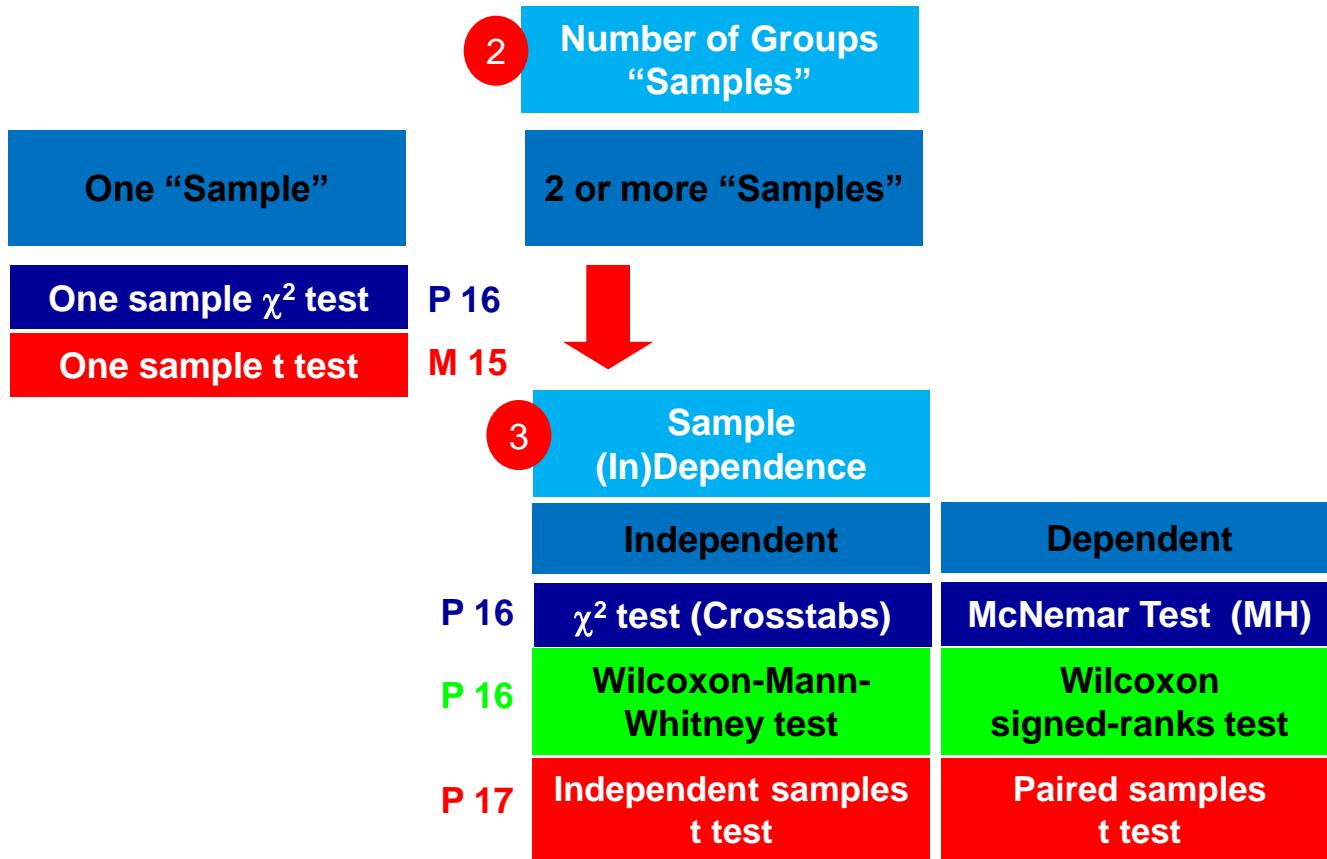
Internet Usage Data (Malhotra, 2010, Table 15.1)

$$\sum \alpha_i$$

The screenshot shows the IBM SPSS Statistics Data Editor interface. The menu bar is visible at the top, and the 'Graphs' tab is selected. A context menu is open over the data table, with the 'Mosaic and Association Plots' option highlighted. A red arrow points from the text 'Measure type: Nominal' to the 'Nominal' entry in the 'Measure' column of the data table.

	Name	Type	Width	Decimals	Label
1	number	Numeric	11	0	Response
2	sex	Numeric	11	0	Sex
3	familiar	Numeric	11	0	Familiar
4	iusage	Numeric	11	0	Internet Usage
5	iattitude	Numeric	11	0	Attitude
6	tattitude	Numeric	11	0	Attitude toward ... [1, Very Unf... None]
7	ishopping	Numeric	11	0	Internet Shopping [1, Yes]... None
8	ibanking	Numeric	11	0	Internet Banking [1, Yes]... None
9	iusagegr	Numeric	8	2	Internet Usage ... [1.00, Light ... None]
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					

Measure type:
Nominal



Chi-Square Goodness of Fit



```
1
2
3 # Chi Square Goodness of Fit Test
4
5 attach(Table15.1)
6
7 table(sex)
8
9 ftable(sex)
10
11 prop.table(table(sex))
12
13
14 chisq.test(table(sex), p = c(0.30,0.70))
15
16 chisq.test(table(sex), p = c(0.30,0.70), simulate.p.value=TRUE)
17
18
19
```



Chi-Square Goodness of Fit

R Console

```
> table(sex)
sex
 1 2
15 15
>
> ftable(sex)
sex 1 2
      15 15
>
> prop.table(table(sex))
sex
 1 2
0.5 0.5
>
>
> chisq.test(table(sex), p = c(0.30,0.70))

Chi-squared test for given probabilities

data: table(sex)
X-squared = 5.7143, df = 1, p-value = 0.01683

>
> chisq.test(table(sex), p = c(0.30,0.70), simulate.p.value=TRUE)

Chi-squared test for given probabilities with simulated p-value (based on
2000 replicates)

data: table(sex)
X-squared = 5.7143, df = NA, p-value = 0.02099
```

Chi-Square Test for Independence (Crosstabs)



```
20
21 # Chi Square for Crosstabulation
22
23 chisq.test(table(sex,ishopping))
24
25 chisq.test(table(sex,ishopping), correct=FALSE)
26
27 chisq.test(table(sex,ishopping), simulate.p.value=TRUE)
28
29 # Package vcd
30
31 library(vcd)
32
33 fisher.test(table(sex,ishopping))
34
35 assocstats(table(sex,ishopping))
36
37
38
```

T

Chi-Square Test for Independence (Crosstabs)



Console

```
> chisq.test(table(sex,ishopping))

Pearson's Chi-squared test with Yates' continuity correction

data: table(sex, ishopping)
X-squared = 2.1719, df = 1, p-value = 0.1405

>
> chisq.test(table(sex,ishopping), correct=FALSE)

Pearson's Chi-squared test

data: table(sex, ishopping)
X-squared = 3.3937, df = 1, p-value = 0.06545
```

I

Chi-Square Test for Independence (Crosstabs)



```
> fisher.test(table(sex,ishopping))

Fisher's Exact Test for Count Data

data: table(sex, ishopping)
p-value = 0.1394
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.7101042 25.9039064
sample estimates:
odds ratio
 3.919659

>
> assocstats(table(sex,ishopping))
          X^2 df P(> X^2)
Likelihood Ratio 3.4661  1 0.062639
Pearson        3.3937  1 0.065447

Phi-Coefficient : 0.336
Contingency Coeff.: 0.319
Cramer's V       : 0.336
```

Graphs

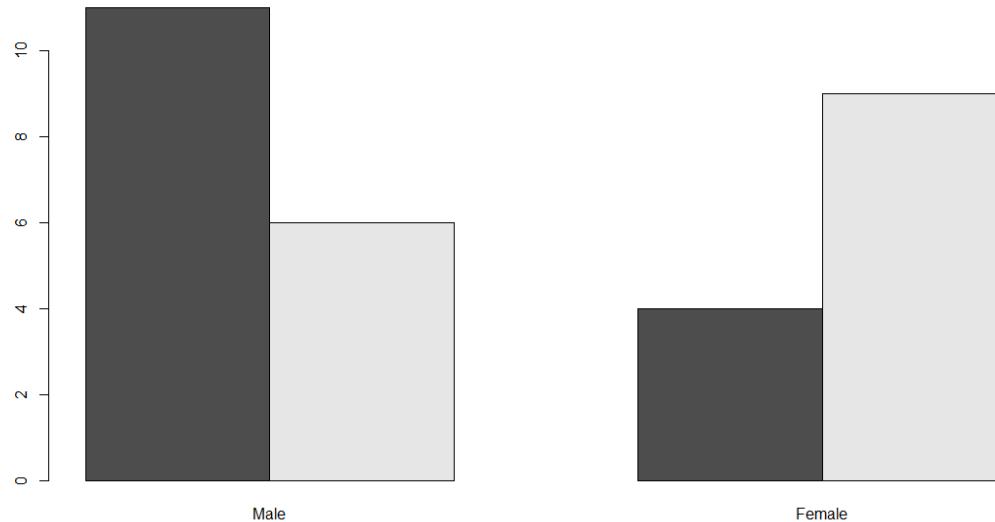


```
39
40 # Plots
41
42 barplot(table(sex,ishopping), names.arg=c("Male", "Female"), beside=TRUE)
43
44 mosaicplot(table(sex,ishopping),color=c("red", "blue"))
45
46
```

Graphs



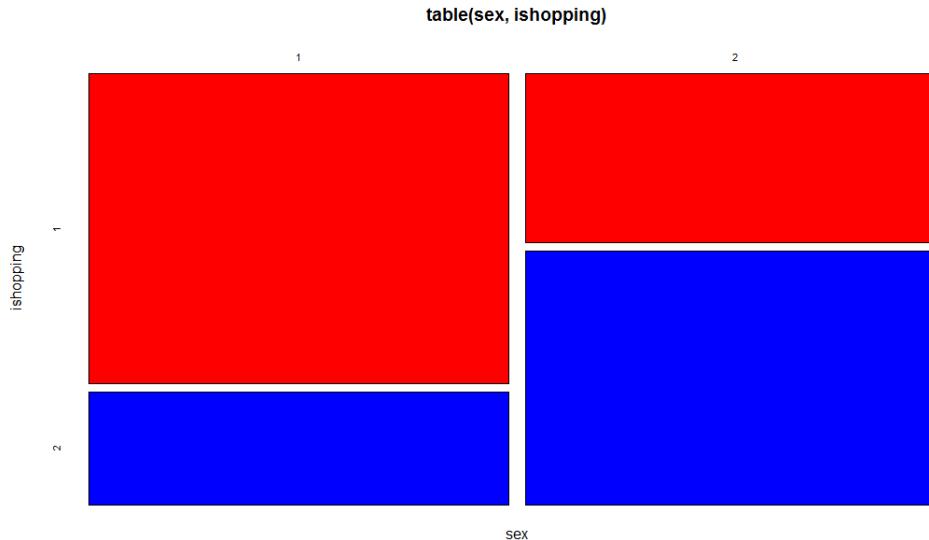
 **Graphics**

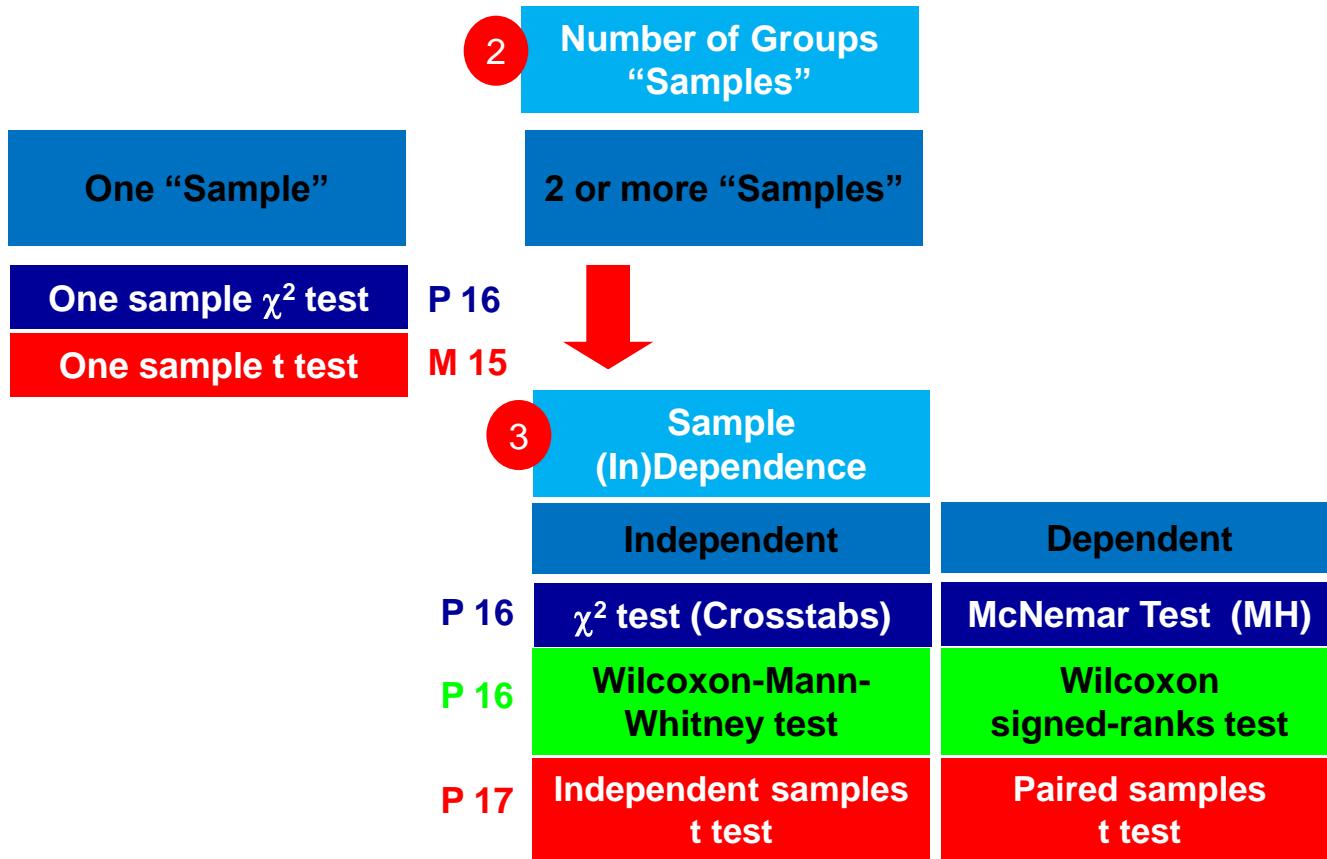


Graphs



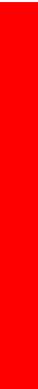
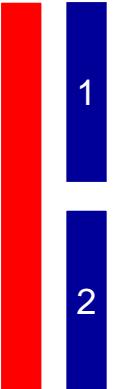
R Graphics





T-Tests

Malhotra (2010); Pallant (2016)

TEST	One-Sample	“Two”-Sample Independent-Samples	“Two”-Sample Paired-Samples
		 X_{1i} X_{2i}	
H_0 (one-sided)	$M \geq 3$	$M_{X1} \geq M_{X2}$	$M_{Y1} \geq M_{Y2}$
H_0 (two-sided)	$M = 3$	$M_{X1} = M_{X2}$	$M_{Y1} = M_{Y2}$

SPSS conducts two-sided tests by default!!!

T Tests

Pallant (2016), Ch. 17



- ▶ Assumptions
 - ▶ Random Sample(s)
 - ▶ Normal distribution (Central Limit Theorem, n>30)
 - ▶ Unknown Population Variance (Equal for independent-samples t test; *homogeneity assumption*)

- At least 3 scale points
- Symmetric distribution

T Tests

$$\sum \alpha$$

Internet Usage Data (Malhotra, 2010, Table 15.1)

The screenshot shows two windows of the IBM SPSS Statistics Data Editor. The top window is titled "Table 15.1 Input.ERROR.sav [DataSet1] - IBM SPSS Statistics Data Editor" and displays a list of variables: number, sex, familiar, iusage, iatitude, tatitude, ishopping, ibanking, iusagegr, and var. The bottom window is titled "Table 15.1 Input.sav [DataSet1] - IBM SPSS Statistics Data Editor" and shows the variable properties for each column.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	number	Numeric	11	0	Respondent Nu...	None	None	8	Right	Scale	Input
2	sex	Numeric	11	0	Sex	{1, Male}...	None	8	Right	Scale	Input
3	familiar	Numeric	11	0	Familiarity	{1, Very Unf...	9	8	Right	Scale	Input
4	iusage	Numeric	11	0	Internet Usage ...	None	None	8	Right	Scale	Input
5	iatitude	Numeric	11	0	Attitude toward ...	{1, Very Unf...	None	8	Right	Scale	Input
6	tatitude	Numeric	11	0	Attitude toward ...	{1, Very Unf...	None	8	Right	Scale	Input
7	ishopping	Numeric	11	0	Internet Shopping	{1, Yes}...	None	8	Right	Scale	Input
8	ibanking	Numeric	11	0	Internet Banking	{1, Yes}...	None	8	Right	Scale	Input
9	iusagegr	Numeric	8	2	Internet Usage ...	{1.00, Light ...	None	8	Right	Scale	Input
10											
19	19	1	7	14	6	6	1	1	2.00		
20	20	2	6	6	6	4	2	2	2.00		
21	21	1	6	9	4	2	2	2	2.00		
22	22	1	5	5	5	4	2	1	1.00		
23	23	2	3	2	4	2	2	2	1.00		
24	24	1	7	15	6	6	1	1	2.00		
25	25	3	6	6	5	3	1	2	2.00		



T-Tests in SPSS

$$\sum \alpha$$

Table 15.1 Input.sav [DataSet3] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Reports Descriptive Statistics Tables Compare Means General Linear Model Generalized Linear Models Mixed Models Correlate Regression Loglinear Neural Networks Classify Dimension Reduction Scale Nonparametric Tests Forecasting Survival Multiple Response Missing Value Analysis... Multiple Imputation Complex Samples Quality Control ROC Curve...

One-sample t test

Independent-samples t test

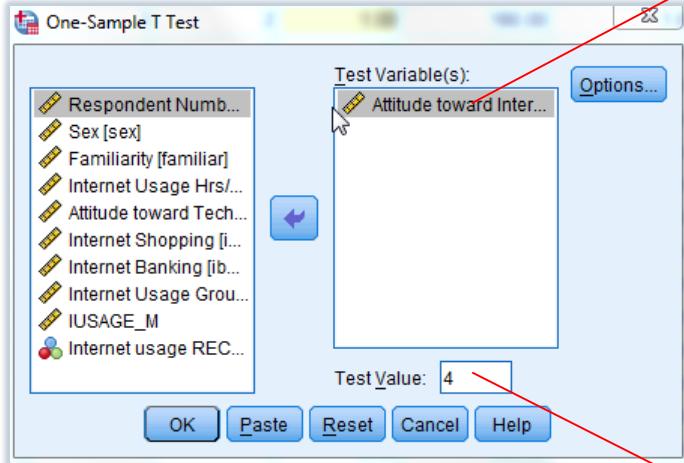
Paired-samples t test

	number	sex
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	
8	8	
9	9	
10	10	
11	11	
12	12	
13	13	
14	14	
15	15	
16	16	
17	17	
18	18	
19	19	
20	20	
21	21	

	banking	z	p
1	2.00		
2	1.00		
3	2.00		
4	1.00		
5	2.00		
6	2.00		
7	2.00		
8	2.00		
9	2.00		
10	2.00		
11	2.00		
12	2.00		
13	2.00		
14	2.00		
15	2.00		
16	2.00		
17	2.00		
18	2.00		
19	2.00		
20	2.00		
21	2.00		

One-Sample T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)



Test Variable:
iattitude

Test Value = 4
(two-sided test!!!)

One-Sample T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Attitude toward Internet	30	5.17	1.234	.225

$$\begin{aligned}H_0: M &= 4 \\H_1: M &\neq 4\end{aligned}$$

$$\begin{aligned}t(29) &= 5.178 \quad (p < 0.001) \\ES &= \eta^2 = t^2/(t^2+df) = 0.480\end{aligned}$$

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
Attitude toward Internet	5.178	29	.000	1.167	.71	1.63

Sig. = p value!

Effect Size (ES)

REMINDER

- ▶ Effect Size (ES)

	(Partial) η^2 (Cohen, 1988)	r (Cohen, 1988)
Small	0.01	0.10
Medium	0.06	0.30
Large	0.138	0.50

One-Sample T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Attitude toward Internet	30	5.17	1.234	.225

$$H_0: M=5$$
$$H_1: M \neq 5$$

$$t(29)=0.740 \quad (p_2=0.465)$$
$$ES = \eta^2 = t^2/(t^2+df) = 0.019!$$

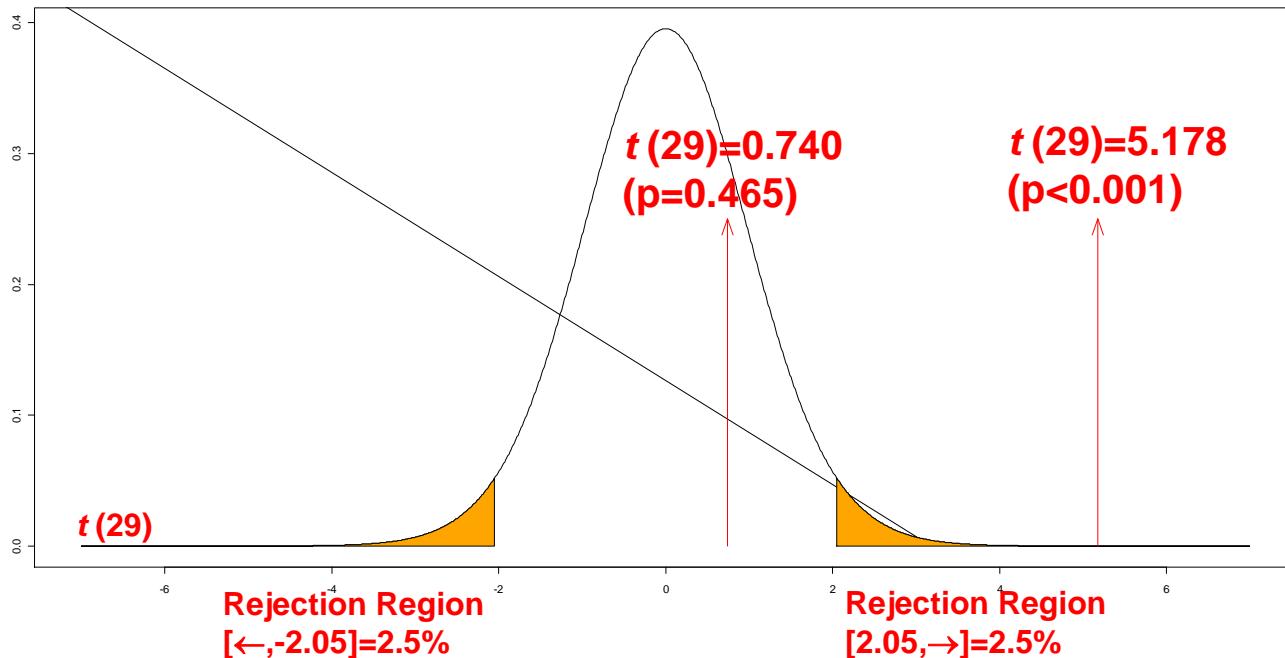
One-Sample Test

	Test Value = 5					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
Attitude toward Internet	.740	29	.465	.167	-.29	.63

Sig. = p value!

One-Sample T-Test

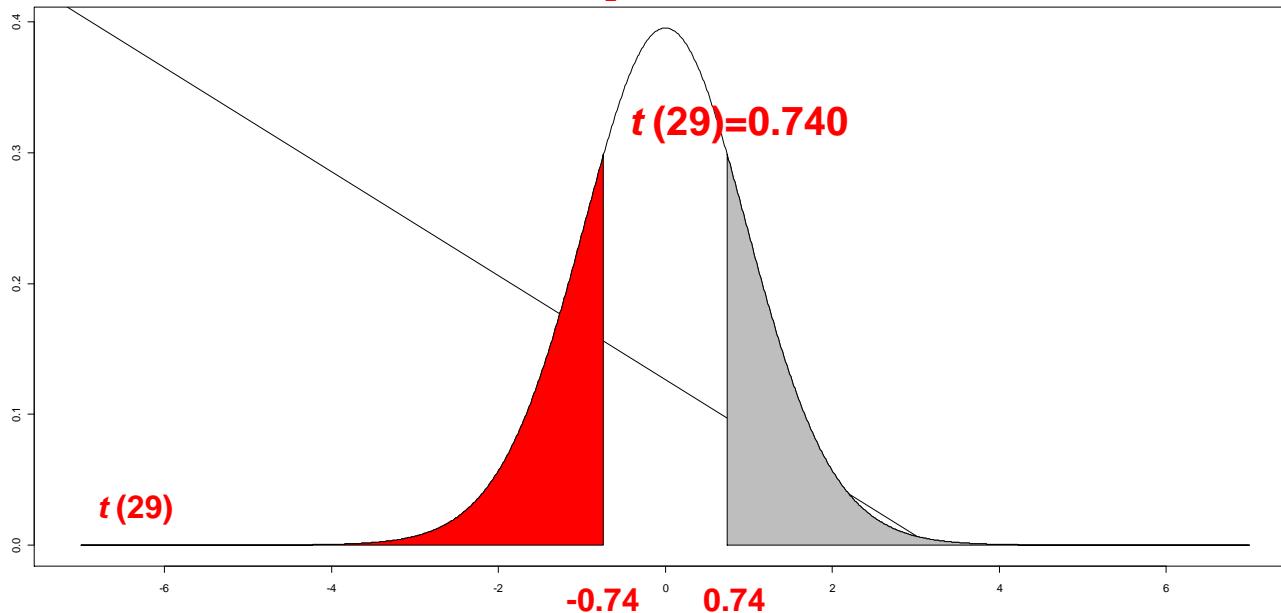
Internet Usage Data (Malhotra, 2010, Table 15.1)



One-Sample T-Test

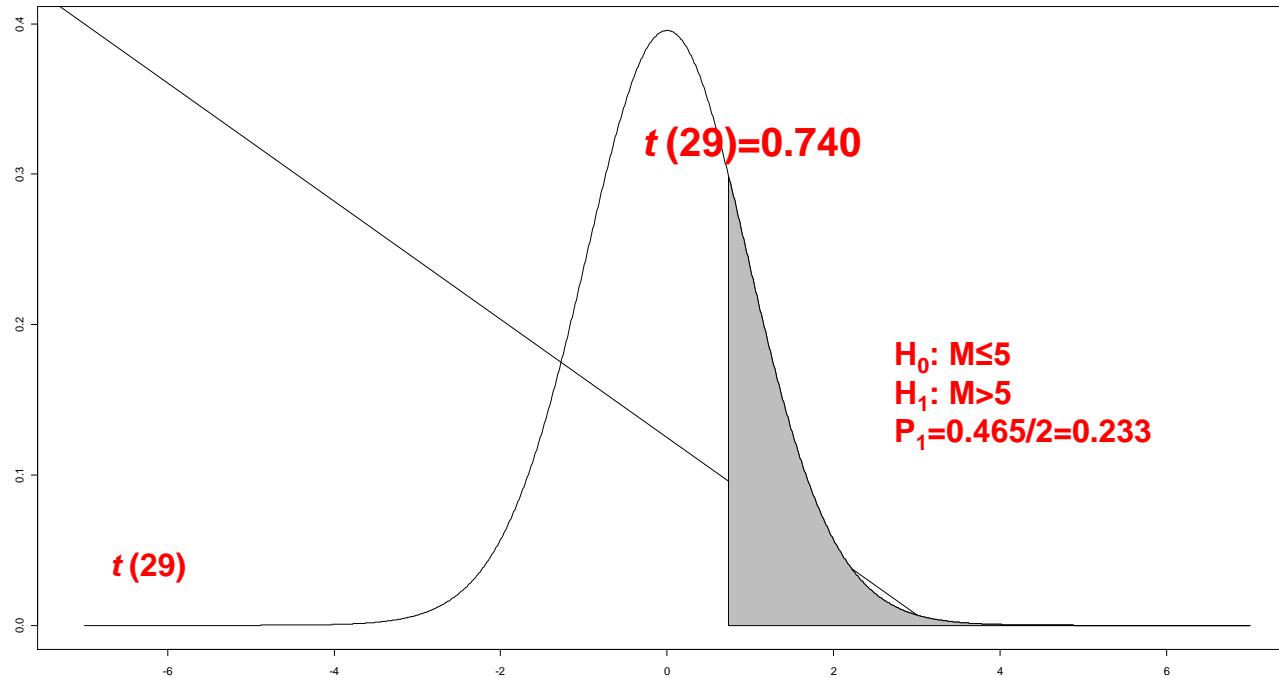
Internet Usage Data (Malhotra, 2010, Table 15.1)

$$\begin{aligned}H_0: M &= 5 \\H_1: M &\neq 5 \\P_2 &= 0.465\end{aligned}$$



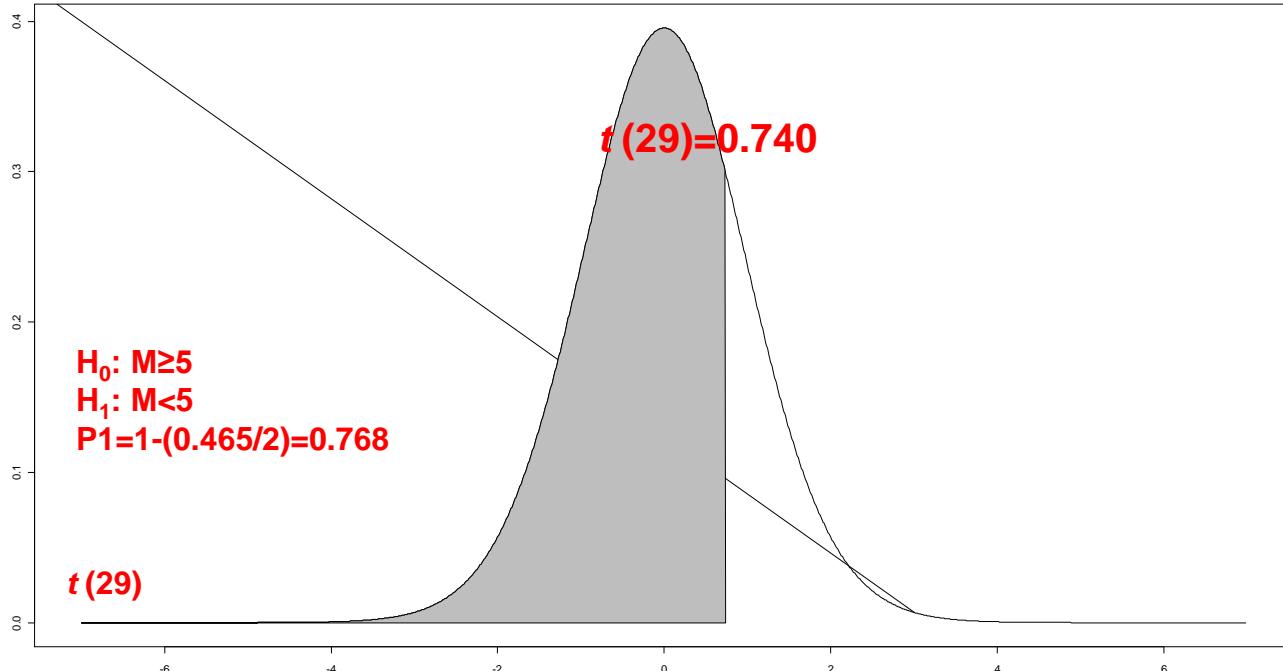
One-Sample T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)



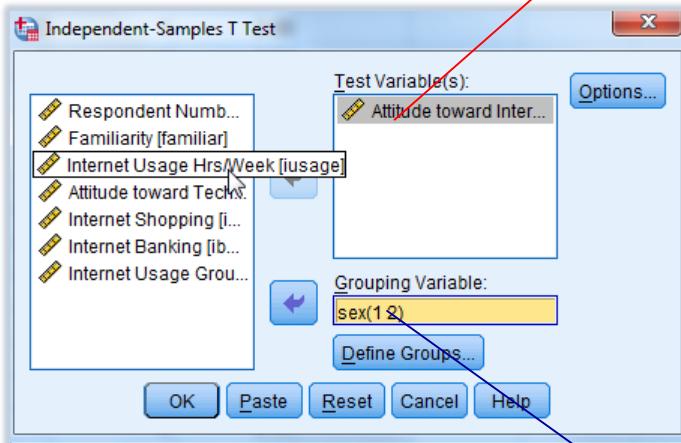
One-Sample T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)



Independent-Samples T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)



Attitude toward Internet
(iattitude)
("Dependent" Variable)

Sex (1='Male', 2='Female')
Grouping Variable
("Independent Variable")

Independent-Samples T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)

Group Statistics					
	Sex	N	Mean	Std. Deviation	Std. Error Mean
Attitude toward Internet	Male	15	5.60	1.183	.306
	Female	15	4.73	1.163	.300

$$H_0: M_{MALE} = M_{FEMALE}$$

$$H_1: M_{MALE} \neq M_{FEMALE}$$

F test: H_0 cannot be rejected
Equal variances assumed

$$t(28)=2.023 \quad (p_2=0.053)$$

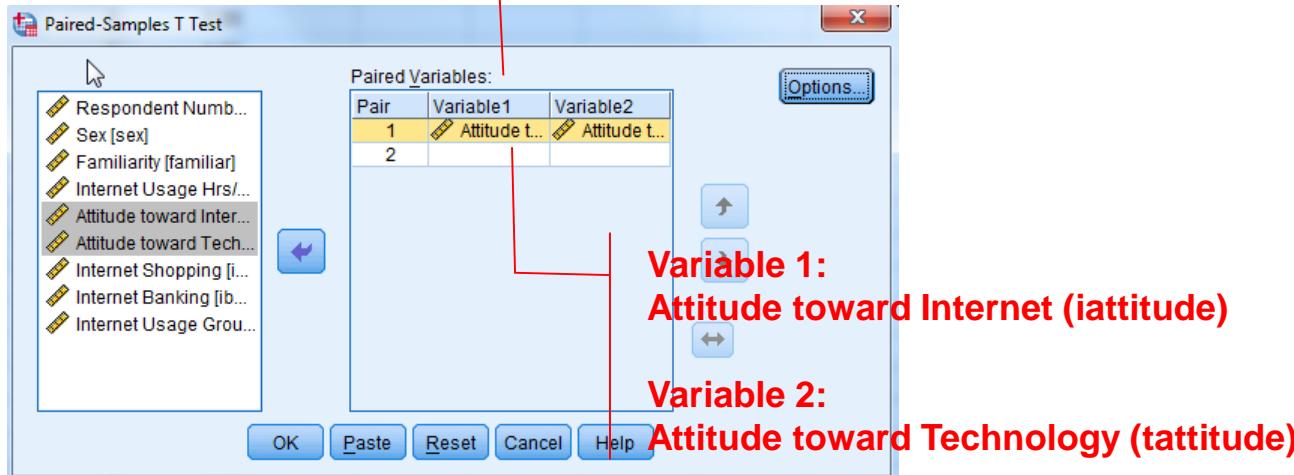
$$ES=\eta^2= t^2/(t^2+df)=0.128$$

		Levene's Test for Equality of Variances		t-test for Equality of Means							95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference				
									Lower	Upper		
Attitude toward Internet	Equal variances assumed	.001	.969	2.023	28	.053	.867	.428	-.011	1.744		
	Equal variances not assumed			2.023	27.992	.053	.867	.428	-.011	1.744		

Paired-Samples T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)

$$H_0: M_{Y1} = M_{Y2}$$
$$H_0: M_D = M_{Y1} - M_{Y2} = 0$$



Paired-Samples T-Test

Internet Usage Data (Malhotra, 2010, Table 15.1)

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Attitude toward Internet	5.17	30	1.234
	Attitude toward Technology	4.10	30	.225
			1.398	.255

$$\begin{aligned} H_0: M_D &= 0 \\ H_1: M_D &\neq 0 \end{aligned}$$

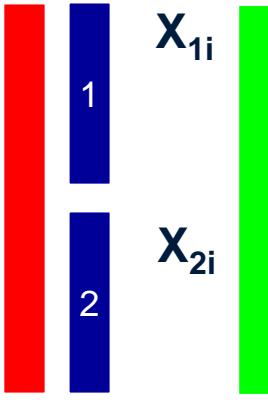
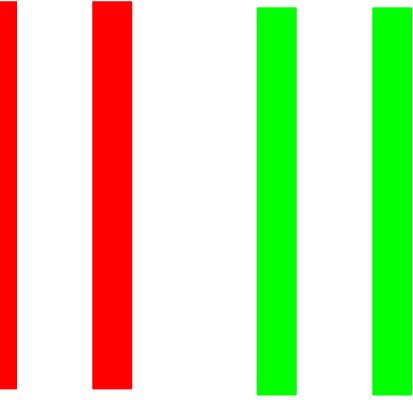
Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
Pair 1	Attitude toward Internet - Attitude toward Technology	1.067	.828	.151	.758	-1.376	7.059	.000			

$$\begin{aligned} t(29) &= 7.059 (p < 0.001) \\ ES &= \eta^2 = t^2 / (t^2 + df) = 0.632 \end{aligned}$$

Nonparametric Alternatives

Malhotra (2010)

TEST	“Two”-Sample			
	Wilcoxon-Mann-Whitney Test		Wilcoxon Signed Rank Test	
		X_{1i}		Y_{1i}
	$Y_i \quad X_i$	Y_i	$Y_{1i} \quad Y_{2i}$	$Y_{1i} \quad Y_{2i}$
H_0 (one-sided)	$\text{MEDIAN}_{X_1} \geq \text{MEDIAN}_{X_2}$		$\text{MEDIAN}_{Y_1} \geq \text{MEDIAN}_{Y_2}$	
H_0 (two-sided)	$\text{MEDIAN}_{X_1} = \text{MEDIAN}_{X_2}$		$\text{MEDIAN}_{Y_1} = \text{MEDIAN}_{Y_2}$	

SPSS conducts two-sided test by default!!!

Nonparametric Alternatives

Malhotra (2010)

$$\sum \alpha$$

Legacy Dialogs!

Wilcoxon-Mann-Whitney Test

Wilcoxon Signed Rank Test

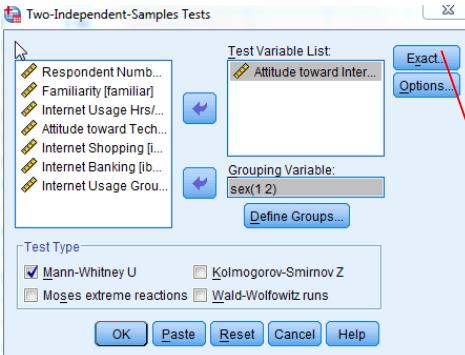
The screenshot shows the IBM SPSS Statistics Data Editor interface. The menu bar is visible with 'File', 'Edit', 'View', 'Data', 'Transform', 'Analyze', 'Graphs', 'Utilities', 'Add-ons', 'Window', and 'Help'. The 'Analyze' menu is open, showing various statistical options. The 'Nonparametric Tests' option is highlighted in yellow. A sub-menu for 'Legacy Dialogs' is also open, listing several nonparametric tests: Chi-square, Binomial, Runs, 1-Sample K-S, 2 Independent Samples, K Independent Samples, 2 Related Samples, and K Related Samples. The 'Independent Samples...' option is highlighted with a blue box. A red arrow originates from the text 'Legacy Dialogs!' and points to the 'Legacy Dialogs' option in the 'Nonparametric Tests' submenu. A green box highlights the 'Wilcoxon Signed Rank Test' option. The main data grid shows a sample dataset with columns: number, sex, latitude, tlatitude, isshopping, ibanking, iusagegr.

$$\sum \alpha$$

Nonparametric Alternatives

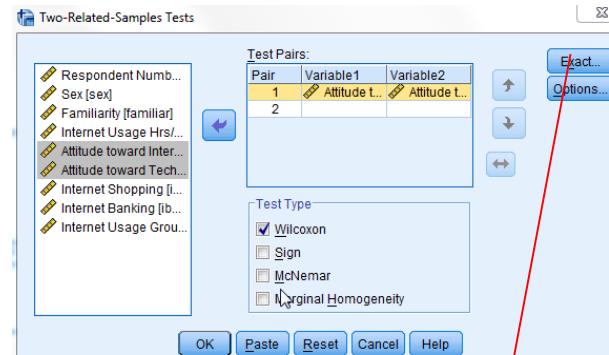
Internet Usage Data (Malhotra, 2010, Table 15.1)

Wilcoxon-Mann-Whitney Test



EXACT!

Wilcoxon Signed Rank Test



EXACT!

Exact Tests ...

- ▶ Nonparametric Tests
 - ▶ Sample Size Matters ... (Mundry and Fischer, 1998; Siegel and Castellan, 1988)
 - ▶ Large samples (“asymptotic testing”)
 - ▶ Small samples (“exact testing”)
 - ▶ What “exactly” is large? (Mundry and Fischer, 1998)
 - ▶ Wilcoxon signed-ranks test $n>15$
 - ▶ Wilcoxon-Mann-Whitney test $n_1=3 \text{ or } 4 \text{ and } n_2>12 \text{ or } n_1>4 \text{ and } n_2>10$
 - ▶ Kruskal-Wallis test $k>3 \text{ and all } n>5$

Exact Tests

Mundry and Fischer (1998)

ANIMAL BEHAVIOUR, 1998, 56, 256–259
Article No. ar980756



Use of statistical programs for nonparametric tests of small samples often leads to incorrect P values: examples from
Animal Behaviour

ROGER MUNDRY & JULIA FISCHER

Institut für Verhaltensbiologie, Freie Universität Berlin

(Received 5 December 1997; initial acceptance 20 January 1998;
final acceptance 9 February 1998; MS. number SC-1124)

Nonparametric Alternatives...

Internet Usage Data (Malhotra, 2010, Table 15.1)

Wilcoxon-Mann-Whitney Test

Ranks				
	Sex	N	Mean Rank	Sum of Ranks
Attitude toward Internet	Male	15	18.57	278.50
	Female	15	12.43	186.50
	Total	30		

RANKS

Independent Samples t Test

Group Statistics					
	Sex	N	Mean	Std. Deviation	M, SD
Attitude toward Internet	Male	15	5.60	1.183	.306
	Female	15	4.73	1.163	.300

Data Rank

1	1
2	2.5
2	2.5
3	4
4	5
5	6.5
5	6.5



$$\text{Sum of Ranks} = \frac{n(n+1)}{2} = 28$$

Nonparametric Alternatives...

Internet Usage Data (Malhotra, 2010, Table 15.1)



	number	sex	familiar	iusage	iattitude	tattitude	ishopping	ibanking	iusagegr	Riattitu
1	1	1	7	14	7	6	1	1	2.00	28 000
2	5	1	7	13	7	7	1	1	2.00	28 000
3	10	1	9	15	7	6	1	2	2.00	28 000
4	13	1	6	9	6	5	2	1	2.00	22 000
5	14	1	6	8	3	2	2	2	2.00	2 000
6	15	1	6	5	5	4	1	2	1.00	14 000
7	17	1	6	9	5	3	1	1	2.00	14 000
8	18	1	4	4	5	4	1	2	1.00	14 000
9	19	1	7	14	6	6	1	1	2.00	22 000
10	21	1	6	9	4	2	2	2	2.00	6 500
11	22	1	5	5	5	4	2	1	1.00	14 000
12	24	1	7	15	6	6	1	1	2.00	22 000
13	26	1	6	13	6	6	1	1	2.00	22 000
14	29	1	4	4	5	3	1	2	1.00	14 000
15	30	1	3	3	7	5	1	2	1.00	28 000
16	2	2	2	2	3	3	2	2	1.00	2 000
17	3	2	3	4	3	1	2	1	1.00	6 500
18	4	2	3	3	7	5	1	2	1.00	28 000
19	6	2	4	6	5	4	1	2	2.00	14 000
20	7	2	2	2	4	5	2	2	1.00	6 500
21	8	2	3	6	5	4	2	2	2.00	14 000
22	9	2	3	6	6	4	1	2	2.00	22 000
23	11	2	4	3	4	3	2	2	1.00	6 500
24	12	2	5	4	6	4	2	2	1.00	22 000
25	16	2	4	3	4	3	2	2	1.00	6 500
26	20	2	6	6	6	4	2	2	2.00	22 000
27	23	2	3	2	4	2	2	2	1.00	6 500
28	25	2	6	6	5	3	1	2	2.00	14 000
29	27	2	5	4	5	5	1	1	1.00	14 000
30	28	2	4	2	3	2	2	2	1.00	2 000

Nonparametric Alternatives

Internet Usage Data (Malhotra, 2010, Table 15.1)

Wilcoxon-Mann-Whitney Test

Ranks			
	Sex	N	Mean Rank
Attitude toward Internet	Male	15	18.57
	Female	15	12.43
	Total	30	186.50

RANKS

Wilcoxon Signed Rank Test

Ranks			
	N	Mean Rank	Sum of Ranks
Attitude toward Internet -	Negative Ranks	1 ^a	7.50
Attitude toward Technology	Positive Ranks	23 ^b	12.72
	Ties	6 ^c	292.50
	Total	30	

- a. Attitude toward Internet < Attitude toward Technology
 b. Attitude toward Internet > Attitude toward Technology
 c. Attitude toward Internet = Attitude toward Technology

Test Statistics^b

	Attitude toward Internet
Mann-Whitney U	66.500
Wilcoxon W	186.500
Z	-1.960
Asymp. Sig. (2-tailed)	.050
Exact Sig. [2*(1-tailed Sig.)]	.056 ^a
Exact Sig. (2-tailed)	.052
Exact Sig. (1-tailed)	.026
Point Probability	.003

$$\text{U}=66.5 \\ \text{z}=-1.96 \quad (\text{p}_2=0.05)$$

$$\text{ES} \\ \text{r=z/}\sqrt{\text{n}} \\ \text{r}=-1.96/\sqrt{30}=-0.358$$

- a. Not corrected for ties.
 b. Grouping Variable: Sex

Test Statistics^b

	Attitude toward Internet - Attitude toward Technology
Z	-4.207 ^a
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

$$\text{z} = -4.207 \quad (\text{p}_2<0.001)$$

$$\text{ES} \\ \text{r=z/}\sqrt{\text{n}} \\ \text{r}=-4.207/\sqrt{30}=-0.768$$

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test



One Sample T Test



```
1 # T Test
2
3 attach(Table15.1)
4
5 # One-sample T Test
6
7 t.test(iattitude, mu=4)
8
9 t.test(iattitude, mu=5)
10
11
```

One Sample T Test



```
> t.test(iatitude, mu=4)

One Sample t-test

data: iatitude
t = 5.178, df = 29, p-value = 1.547e-05
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
4.705848 5.627485
sample estimates:
mean of x
5.166667

>
> t.test(iatitude, mu=5)

One Sample t-test

data: iatitude
t = 0.7397, df = 29, p-value = 0.4654
alternative hypothesis: true mean is not equal to 5
95 percent confidence interval:
4.705848 5.627485
sample estimates:
mean of x
5.166667
```

Independent Samples T Test



```
13
14 # Independent Samples T Test
15
16 t.test(iattitude ~ sex, alternative="two.sided", var.equal=FALSE)
17
18 t.test(iattitude ~ sex, alternative="less")
19
20 t.test(iattitude ~ sex, alternative="greater")
21
22 |
23
```

Independent Samples T Test



R Console

```
> t.test(iattitude ~ sex, alternative="two.sided", var.equal=FALSE)

  Welch Two Sample t-test

data: iattitude by sex
t = 2.0232, df = 27.992, p-value = 0.05269
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.01080031 1.74413364
sample estimates:
mean in group 1 mean in group 2
5.600000      4.733333
```

Independent Samples T Test



```
> t.test(iattitude ~ sex, alternative="less")

Welch Two Sample t-test

data: iattitude by sex
t = 2.0232, df = 27.992, p-value = 0.9737
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
-Inf 1.59537
sample estimates:
mean in group 1 mean in group 2
      5.600000      4.733333

>
> t.test(iattitude ~ sex, alternative="greater")

Welch Two Sample t-test

data: iattitude by sex
t = 2.0232, df = 27.992, p-value = 0.02635
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
0.1379632 Inf
sample estimates:
mean in group 1 mean in group 2
      5.600000      4.733333
```

Levene's Test



```
22  
23 # Levene's Test|  
24  
25 library(lawstat)  
26  
27 Levene.test(iattitude,sex, location="mean")  
28  
29
```

Levene's Test



R Console

```
> Levene.test(iattitude,sex, location="mean")
classical Levene's test based on the absolute deviations from the mean (
  none not applied because the location is not set to median )

data: iattitude
Test Statistic = 0.0015, p-value = 0.9695
```

Paired Samples T Test



```
27  
28 # Paired Samples T Test  
29  
30 t.test(iattitude, tattitude, paired=TRUE)  
31  
32
```

Paired Samples T Test



```
> t.test(iattitude, tattitude, paired=TRUE)

Paired t-test

data: iattitude and tattitude
t = 7.0587, df = 29, p-value = 9.16e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.7576051 1.3757282
sample estimates:
mean of the differences
 1.066667
```

Wilcoxon-Mann-Whitney Test



```
35
36 # Wilcoxon-Mann-Whitney Test
37
38 wilcox.test(iattitude ~ sex, exact=FALSE, correct=FALSE)
39
40
```

Wilcoxon-Mann-Whitney Test



R Console

```
> wilcox.test(iattitude ~ sex, exact=FALSE, correct=FALSE)

wilcoxon rank sum test

data: iattitude by sex
W = 158.5, p-value = 0.05001
alternative hypothesis: true location shift is not equal to 0
```

Wilcoxon Signed Rank Test



```
44
45 # wilcoxon Signed Rank Test
46
47 wilcox.test(iatitude,tatitude, paired=TRUE, exact=FALSE, correct=FALSE)
48
```

Wilcoxon Signed Rank Test



R Console

```
> wilcox.test(iattitude,tattitude, paired=TRUE, exact=FALSE, correct=FALSE)
wilcoxon signed rank test

data: iattitude and tattitude
V = 292.5, p-value = 2.592e-05
alternative hypothesis: true location shift is not equal to 0
```

References

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Cohen, J. (1992). "A Power Primer," *Psychological Bulletin*, 112 (1), 155-159. [condensed version].
- Everitt, B. S. (1992). *The Analysis of Contingency Tables*. CRC Press.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behavior Research Methods*, 39 (2), 175-191.
- Hair, J.F., Jr., Black, W.C., Babin, B.J, and Anderson, R.E. (2018). *Multivariate Data Analysis*.Cengage.
- Howell, D. C. (2012). *Statistical Methods for Psychology*. Cengage Learning.
- Malhotra, N. (2010). *Marketing Research: An Applied Orientation*. Upper Saddle River: Pearson/Prentice-Hall.

References

- Mundry, R. and Fischer, J. (1998). Use of Statistical Programs for Nonparametric Tests of Small Samples often Leads to Incorrect P Values: Examples from Animal Behaviour. *Animal Behaviour*, 56, 256-259.
- Pallant, J. (2016). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows*. Maidenhead: Open University Press/McGraw-Hill.
- Siegel, S. and Castellan, N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

R eferences

- Chapman, C., & Feit, E. M. (2015). *R for Marketing Research and Analytics*. New York, NY: Springer.
- Crawley, MJ (2013). *The R Book*. Chichester, UK: John Wiley and Sons.
- Everitt, BS and Hothorn, T (2006). *A Handbook of Statistical Analysis Using R*. Boca Raton, FL: Chapman and Hall/CRC.
- Field, A, Miles, J and Field, Z (2012). *Discovering Statistics Using R*. Los Angeles, CA: Sage Publications.
- Kabacoff, RI (2011). *R in Action*. Shelter Island, NY: Manning.
- Muenchen, RA (2009). *R for SAS and SPSS Users*. New York, NY; Springer Science and Business Media.



Thank you for
Your Attention!