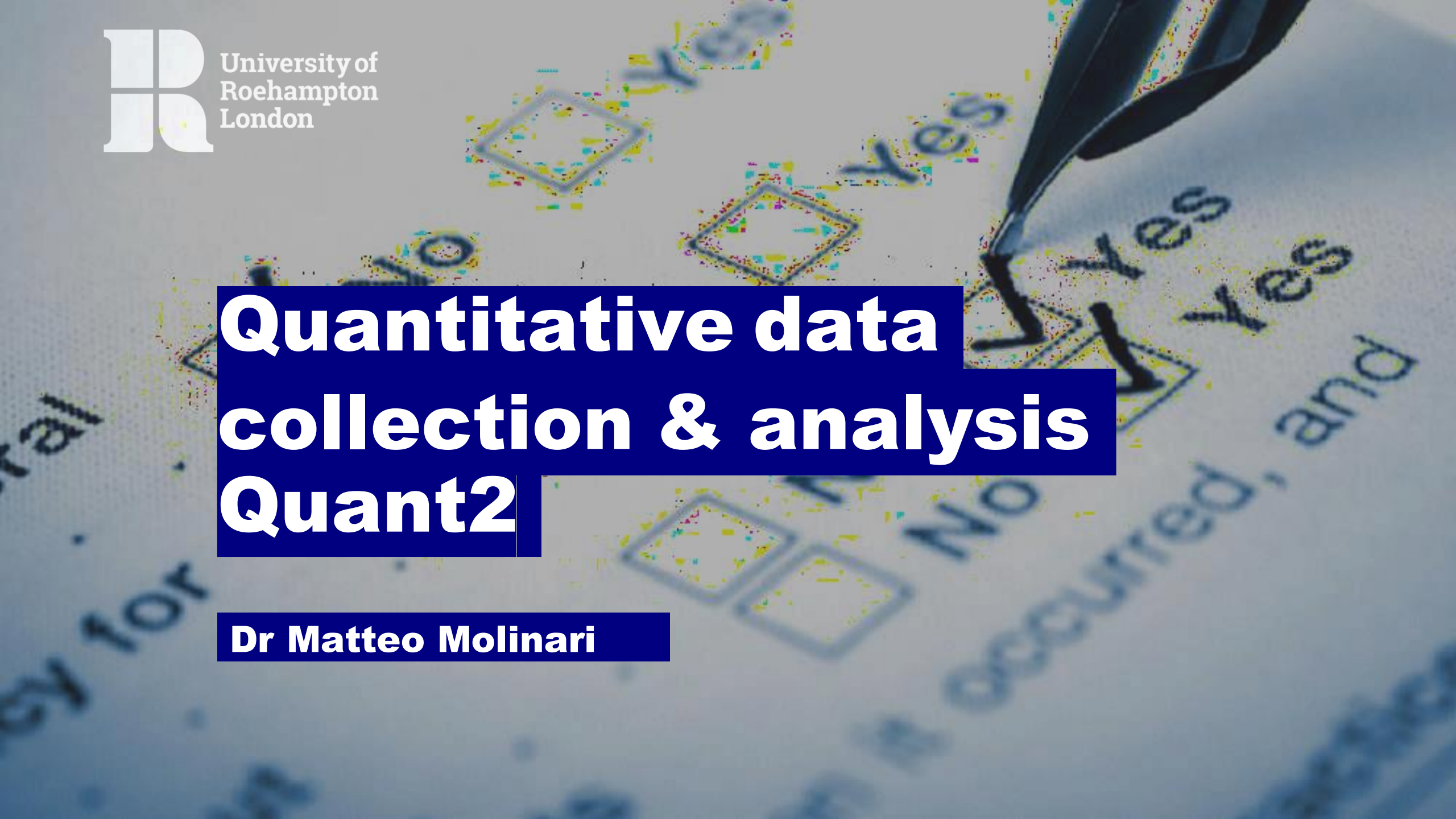


University of Roehampton London

Quantitative data collection & analysis Quant2

Dr Matteo Molinari

University of Roehampton London

Agenda

- **To understand rigour and validity of adequate analysis.**
- **To follow high standards in data collection.**

University of Roehampton London

Outline

- Descriptive Data Analysis
- Hypothesis Testing
- Regression Analysis

University of Roehampton London

Scale of measurement

Discrete	Nominal	Gender, Religious
	Ordinal	Education
Continuous	Interval	Temperature, IQ
	Ratio	Income, age

University of Roehampton London

General rule- Type of analysis

- DV and IV are Nominal scale-Cross tab
- DV and IV are Interval –Regression , Pearson correlation
- IV (2 categorical) DV (Interval)- T test
- IV (3 categorical) DV (Interval)-ANOVA

University of Roehampton London

Descriptive Analysis with One Variable

- We will look at the following statistics:

➢ Frequency Distributions (ordinal and nominal variables)

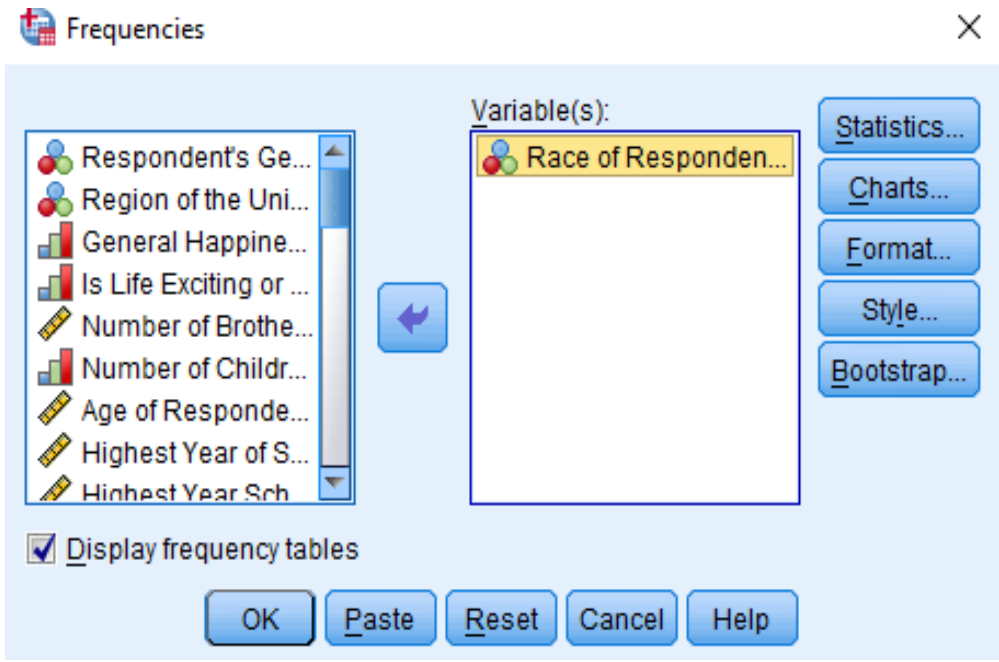
- SPSS provides a range of output e.g. cell count, percentages (%) total, as well information on missing cases

➢ Descriptive statistics (scale variables)

- measures of central tendency and dispersion e.g. min, max, mean, and standard deviation.

Frequencies

- Select *Analyze>Descriptive Statistics>Frequencies*
- Highlight the variable(s) you want in the left hand panel and click on the arrow to move it into the right hand panel and click on 'OK'
- Your frequency table will appear in the output window
- Other options for descriptive statistics include percentiles (more appropriate for scale variables)
- You can sort your results in the output (e.g. using categories)



University of Roehampton London

Frequency Distribution for Race

Statistics

Race of Respondent

N	Valid	1517
	Missing	0

Race of Respondent

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	White	1264	83.3	83.3	83.3
	Black	204	13.4	13.4	96.8
	Other	49	3.2	3.2	100.0
	Total	1517	100.0	100.0	

University of Roehampton London

Descriptives

- Select *Analyze>Descriptive Statistics>Descriptives*
- Highlight the variable (e.g. age) and move it across.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age of Respondent	1514	18	89	45.63	17.808
Valid N (listwise)	1514				

- To obtain the median or quartiles, you will find it as an option under *Frequencies>Statistics* or use the *Analyze>Explore* option
- Stats can be affected by missing values and thus, you may need to discard or apply corrective measures on missing cases/values

University of Roehampton London

Descriptive Analysis with Two Variables

We will now look at the following techniques:

- Cross Tabulations
- Descriptive statistics by category
- Correlations

University of Roehampton London

Cross-Tabulations

- To examine the relationship between two nominal variables, select *Analyze>Descriptive Statistics>Crosstabs*
- Select the variables you want as the row and column variables and move them into the respective boxes
  - e.g. to see how happiness varies by gender, select happy as your row variable and gender as your column variable
- Click 'OK' to see the frequencies for each category
- Percentages of the results can be obtained by clicking on the *Cells* button on the right hand side of the screen
  - e.g. click on the 'row' option under *Cells* to get row percentages
- Contingency (and other) tests can be obtained under *Statistics*

University of Roehampton London

Crosstab of Happiness and Gender

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
General Happiness * Respondent's Gender	1504	99.1%	13	0.9%	1517	100.0%

General Happiness * Respondent's Gender Crosstabulation

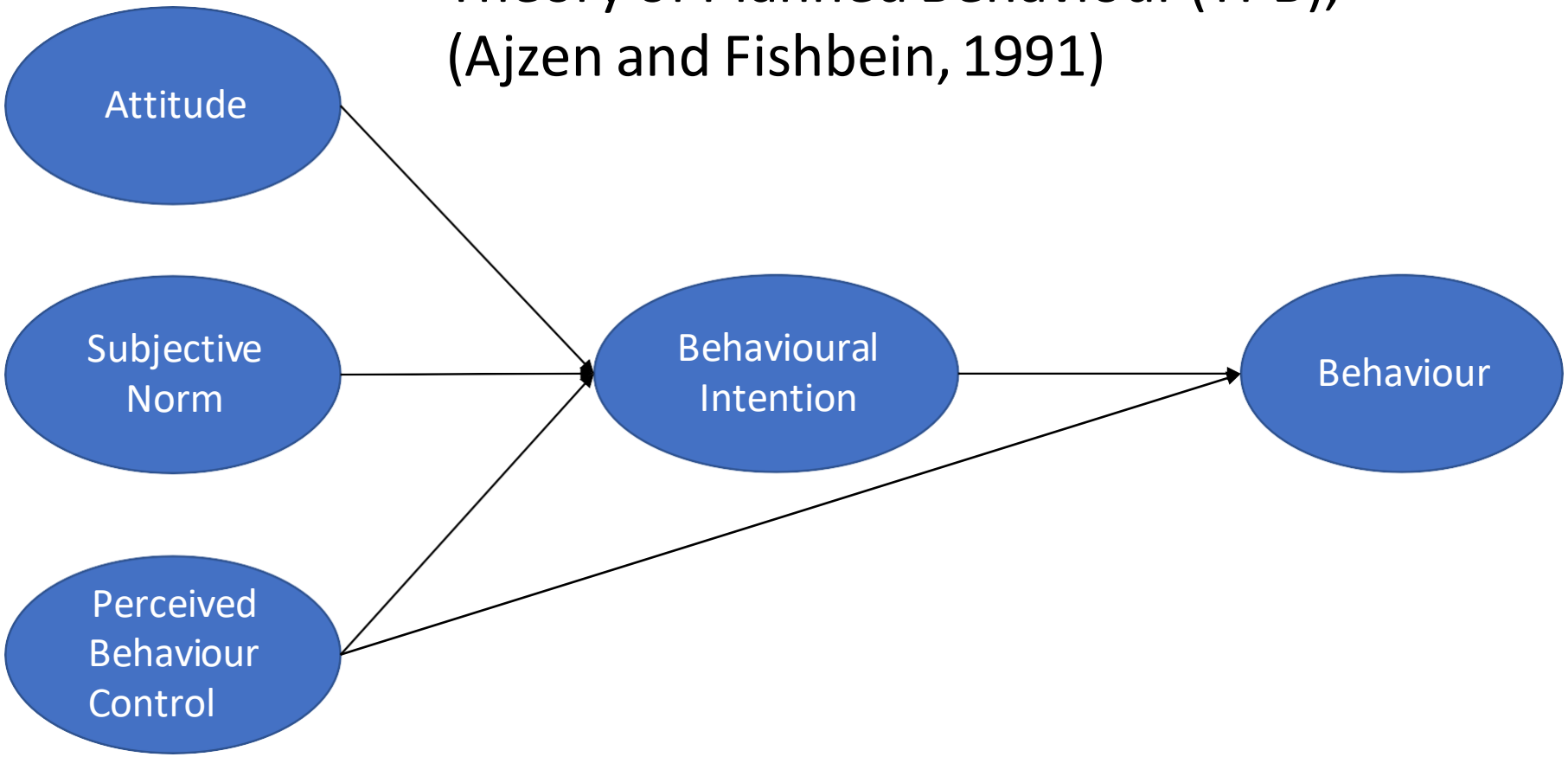
			Respondent's Gender		
			Male	Female	Total
General Happiness	Very Happy	Count	206	261	467
		% within General Happiness	44.1%	55.9%	100.0%
	Pretty Happy	Count	374	498	872
		% within General Happiness	42.9%	57.1%	100.0%
	Not Too Happy	Count	53	112	165
		% within General Happiness	32.1%	67.9%	100.0%
Total		Count	633	871	1504
		% within General Happiness	42.1%	57.9%	100.0%

Descriptives

	Respondent's Gender			Statistic	Std. Error
Highest Year of School Completed	Male	Mean		13.23	.125
		95% Confidence Interval for Mean	Lower Bound	12.99	
			Upper Bound	13.48	
		5% Trimmed Mean		13.27	
		Median		13.00	
		Variance		9.876	
		Std. Deviation		3.143	
		Minimum		3	
		Maximum		20	
		Range		17	
		Interquartile Range		4	
		Skewness		-.203	.097
		Kurtosis		.425	.194
	Female	Mean		12.63	.096
		95% Confidence Interval for Mean	Lower Bound	12.44	
			Upper Bound	12.82	
		5% Trimmed Mean		12.65	
		Median		12.00	
		Variance		8.062	
		Std. Deviation		2.839	
		Minimum		0	
		Maximum		20	
		Range		20	
		Interquartile Range		2	
		Skewness		-.199	.083
		Kurtosis		1.006	.165

University of Roehampton London

Theory of Planned Behaviour (TPB), (Ajzen and Fishbein, 1991)



Correlations

- To examine the relationship between two continuous variables, select *Analyze>Correlate>Bivariate*

- Select the variables you want and move them into the box on the right hand side by clicking on the right arrow
  - e.g. to see how an individual's attitude is associated with their purchasing intention, select att and intention
  - The default correlation coefficient is Pearson

- Click 'OK. The output will include significance levels for the correlations.

- Other coefficients can be added (e.g. Kendall's Tau and Spearman)

- Be careful on how you interpret correlation results: a high correlation does not necessarily imply a causation

Correlation Between Individual's attitude and intention

Correlations

		Attitude	Intention
Attitude	Pearson Correlation	1	.419**
	Sig. (2-tailed)		.000
	N	1510	1232
Intention	Pearson Correlation	.419**	1
	Sig. (2-tailed)	.000	
	N	1232	1233

**. Correlation is significant at the 0.01 level (2-tailed).

University of Roehampton London

Tests of Statistical Significance

- **Chi-Square Tests**
  - Establishes how confident we are that there exists a relationship between two variables in a population

- **Comparing means (t-tests)**
  - Allows you to test whether a variable's mean is equal to a particular value or is different to that of another variable in a population

- **Tests of Correlation**
  - Provides information about likelihood that variables are statistically associated in a population

University of Roehampton London

Statistical significance when a relationship is found

- How confident can we be that the findings from a sample can be generalized to the population as a whole?
- How risky is it to make this inference?
- Hypothesis testing is usually applied to probability (or representative) samples

What is a Hypothesis?

An informed speculation about a relationship between some variables based on theoretical underpinnings but has not been proven yet. There are 2 types of hypotheses:

- **Null hypothesis** (H_0): represents the status quo that is adopted until it is proven false i.e. effect is absent, there is no association, difference or relationship:

"There is no relationship between students' attendance and their academic performance"

- **Alternative hypothesis** (H_a): *converse* of null hypothesis which represents the theory that we will adopt when there is corroborative evidence i.e. effect is present, there is an association, difference or relationship:

"Class attendance has a positive effect on students' academic performance"

  - **One-tailed** vs **Two-tailed** hypothesis test
  - **Type I** & **Type II** errors

University of Roehampton London

What is a Hypothesis (cont.)?

- Single-tailed test: directional $(X > Y \text{ or } X < Y)$
  Predicts the nature of the difference:

  "Female managers earn less than their male counterparts"

  "An increase in the number of promotional emails sent to customers will increase customer complaints"

  - Can also be a default when one of the sides is impossible

- Two-tailed test: non-directional $(X \neq Y)$
  Does not predict the nature of the difference. You are interested in values above or below the hypothesised value (no specific direction)

  "There is a difference in the wages earned between employees working in London and the South East"

  - May be you know that both outcomes are possible or you are not sure

University of Roehampton London

Steps for Hypothesis Testing

1. **Set up** a null hypothesis (H_0) as well as an alternative hypothesis (H_1);

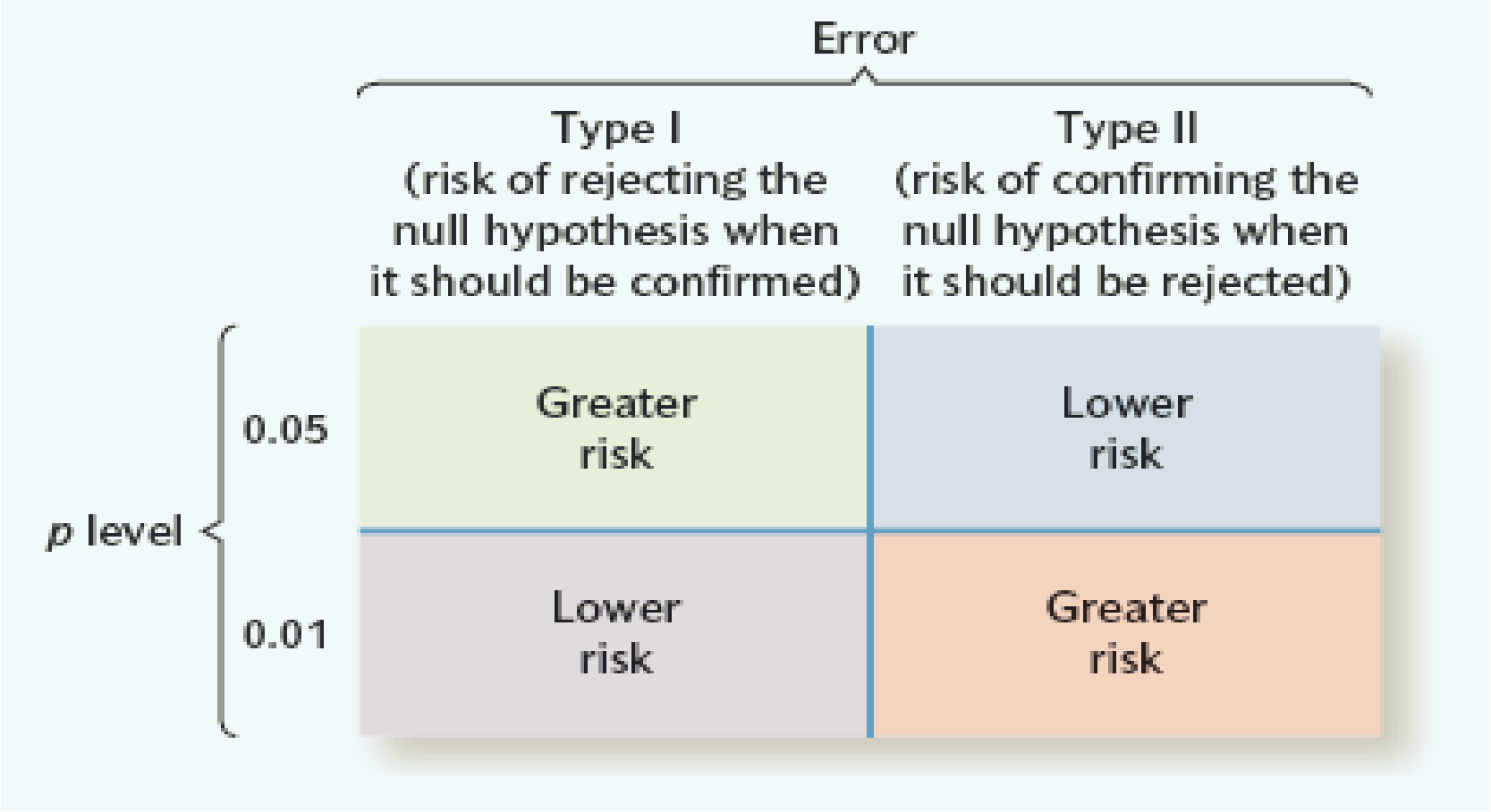
2. **Decide** on an acceptable level of statistical significance (1% or 5%);

3. **Use** an appropriate statistical test (could either be set up as one or two-tailed alternative);

4. If acceptable level is attained, **reject** null hypothesis; If not attained, **do not reject** it.

University of Roehampton London

…but you might be wrong to accept or reject the null hypothesis



Type I & Type II errors

University of Roehampton London

Null Hypothesis	Type I Error / False Positive	Type II Error / False Negative
Wolf is not present	Shepherd thinks wolf is present (shepherd cries wolf) when no wolf is actually present	Shepherd thinks wolf is NOT present (shepherd does nothing) when a wolf is actually present

University of Roehampton London

Null Hypothesis	Type I Error / False Positive	Type II Error / False Negative
Person is not guilty of the crime	Person is judged as **guilty** when the person actually **did not** commit the crime (convicting an innocent person)	Person is judged **not guilty** when they actually **did** commit the crime (letting a guilty person go free)

University of Roehampton London

Different Tests in SPSS

- One can test the extent to which two (nominal or ordinal) variables are associated and the most common test of association is the Chi-Square Test

- The null hypothesis in this test assumes that the variables are independent

- These tests can be performed by selecting *Analyze>Descriptive Statistics>Crosstabs* and then clicking on the *Statistics* button and ticking the appropriate boxes.

University of Roehampton London

A Basic Example of Crosstabs

Attitude

Cheated	Liked	Disliked
Yes	12	33
No	44	66

Comparing Means

- Various statistical tests can be found under *Analyze>Compare Means*

- To test whether means are different from zero (or another value)
  - ➢ Select *One-Sample T Test*

- To test whether means are different from each other
  - ➢ Select *Paired-Sample T Test*

Paired sample t-test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Male_time	11.01	974	4.118	.132
	Female_time	11.02	974	3.407	.109

H0= no difference between male and female group

Paired Samples T

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Male_time Female_time	-.007	3.115	.100	-.203	.189	-.072	973	.943

P=0.943>0.05
Null Hypothesis not rejected
No difference between two groups

University of Roehampton London

Correlation and Significance

- How confident can you be about a relationship between two variables?
- Whether a correlation coefficient is statistically significant depends on:
  - Size of the coefficient (the higher the better)
  - Size of your sample (the larger the better)
- If $p < 0.05$, one can reject the null hypothesis

Correlation and Significance in SPSS

- You can obtain correlation coefficients for more than 2 variables e.g. attitude and their SN and PCB

Correlations

		AVGatt	AVGintention	AVGsn	AVGpbc
AVGatt	Pearson Correlation	1	.213**	.147*	.081
	Sig. (2-tailed)		.003	.045	.270
	N	187	187	187	187
AVGintention	Pearson Correlation	.213**	1	.567**	.380**
	Sig. (2-tailed)	.003		.000	.000
	N	187	187	187	187
AVGsn	Pearson Correlation	.147*	.567**	1	.566**
	Sig. (2-tailed)	.045	.000		.000
	N	187	187	187	187
AVGpbc	Pearson Correlation	.081	.380**	.566**	1
	Sig. (2-tailed)	.270	.000	.000	
	N	187	187	187	187

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05level (2-tailed).

p-values are generated automatically. Stars (*) next to a value signify that there is a significant relationship between the variables

University of Roehampton London

Regression Analysis

- Similar to correlation but also gives you a coefficient attached to a constant
- OLS estimates a line of best fit with an intercept and slope
- Coefficients should be interpreted differently depending on how the variables have been entered (e.g. in levels or logs)
- Associated *p*-values allow you to test for significant relationships
- The R-squared statistic in the output shows the overall significance of the regression
  - How much the independent variables explain the variations in the dependent variables

Simple Regression Analysis in SPSS

- Select *Analyze>Regression>Linear*

- To estimate the effect of attitude on intention
  - Move intention into the Dependent variable box
  - Move attitude into the Independent variables box
  - Click OK

- A range of other options are available with the *Statistics* button

Simple Regression Output from SPSS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.213[a]	.045	.040	.70377	.045		1	185	.003

a. Predictors: (Constant), AVGatt

The R and R square value are statistically significant

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.351	1	4.351	8.784	.003[b]
	Residual	91.629	185	.495		
	Total	95.980	186			

a. Dependent Variable: AVGintention

A consumer's intention is expected to be around 2.566 if attitude is zero.

A one unit increase in consumer's intention will result in an increase of 0.261 in the consumer's attit ude.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	2.566	.355		7.237	.000
	AVGatt	.261	.088	.213	2.964	.003

a. Dependent Variable: AVGintention

The equation for predicting intention is

intention = *2.566+ 0.261**att*

Dr Matteo Molinari

Multiple Regression Analysis in SPSS

- Can add more explanatory variables into the analysis e.g. age, subjective norm, etc.
- Just move SN into the independent variable box and click OK
- Analysis will affect the coefficient on attitude as well as *R-squared*

University of Roehampton London

Multiple Regression Output from SPSS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.582[a]	.339	.332	.58716	.339	47.201	2	184	.000

a. Predictors: (Constant), AVGsn, AVGatt

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.451	.321		4.528	.000
	AVGatt	.163	.074	.132	2.185	.030
	AVGsn	.472	.052	.548	9.043	.000

a. Dependent Variable: AVGintention