

Research Methods I

Correlation

Stacey Coleman



Recap: Study Design



There are often several potential ways to explore a psychological research question



Aim to weigh up the benefits and limitations of each possible approach



Must make sure the design chosen will enable you to answer your research question, and reject or accept your hypotheses

Recap: Experimental Designs

When we are looking for **differences** between two or more groups, we will normally use an **experimental design**

Important in exploring
causal relationship
between variables

Can identify whether
one type of variable
affects another

Cause and effect!

Problems with the 'science' of psychology

01

Objectivity

Social nature of what we study = subjective

02

Experimenter bias

Are we influenced by our own beliefs?

03

Demand characteristics

Humans are problem solvers

Problems with the 'science' of psychology

04

Representativeness

Can we generalise? Is the sample representative?

05

Artificiality

Do people behave in a lab in the same way as in the real world?

06

Validity

Ecological validity – similar to real life?

Internal validity – scientific control in study?

The background features a light blue watercolor wash in the upper center, transitioning into a light green wash at the bottom. Scattered throughout are small, dark blue dots of varying sizes. A thin, dark blue line curves from the top right, passing through the green area and ending in a series of loops. Another thin, dark blue line curves from the top left, passing through the blue area and ending in a series of loops.

Non- experimental Designs

Non-experimental designs

- Some research questions do not suit an experimental design
- We can still use quantitative methods (numbers), but we can employ a non-experimental method to answer our research questions
 - For example, correlational design
- Non-experimental methods investigate variables that exist among people irrespective of any researcher intervention
- In these studies, no IV is manipulated
 - Instead, researchers measure variables as they naturally occur

Correlation

Non-experimental

Explores relationships!

No IV or DV

*Measurement of co-occurring
(already existing) variables*

Data

Requires data at interval level

Correlation explores a **relationship** between two co-occurring variables

Correlation

Variables are **measured**, not manipulated

Correlation analysis allows for the investigation of a **relationship** between variables: a shared variance

Two elements to a correlation value (r):

Strength

Direction

Important points re: correlation



Cause cannot be inferred from the existence of a strong correlation between two variables

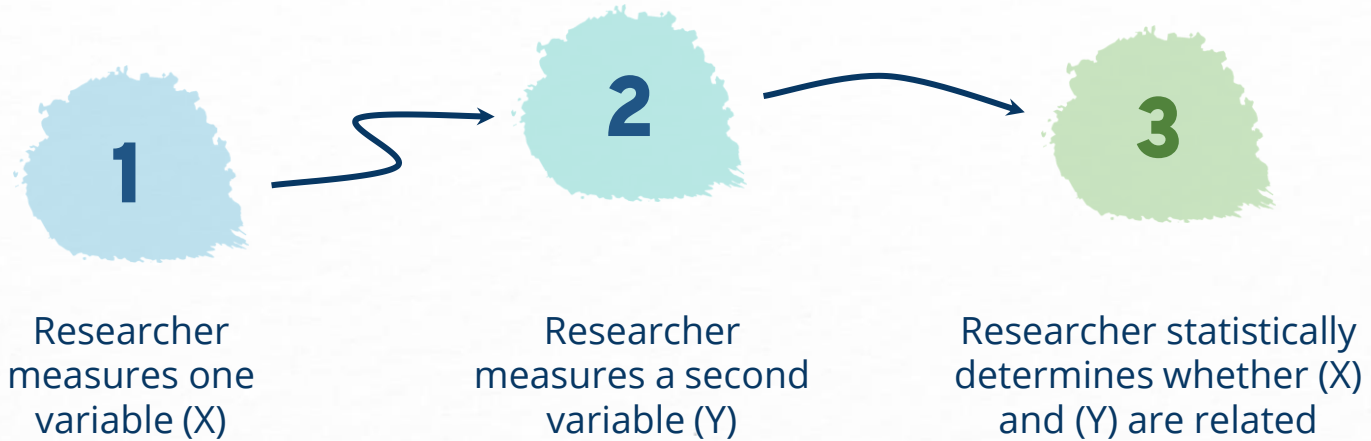
We cannot use terminology that refers to something **causing** something else!

Strength is a measure of correlation, but **significance** assesses how unlikely such a correlation was to occur under the null hypothesis

This assessment depends on the size of N

Correlation

There are three components to correlation research:



Example

Is there a relationship between levels of impulsivity and gambling behaviour?

Measure of (X)

Measure of (Y)

Example

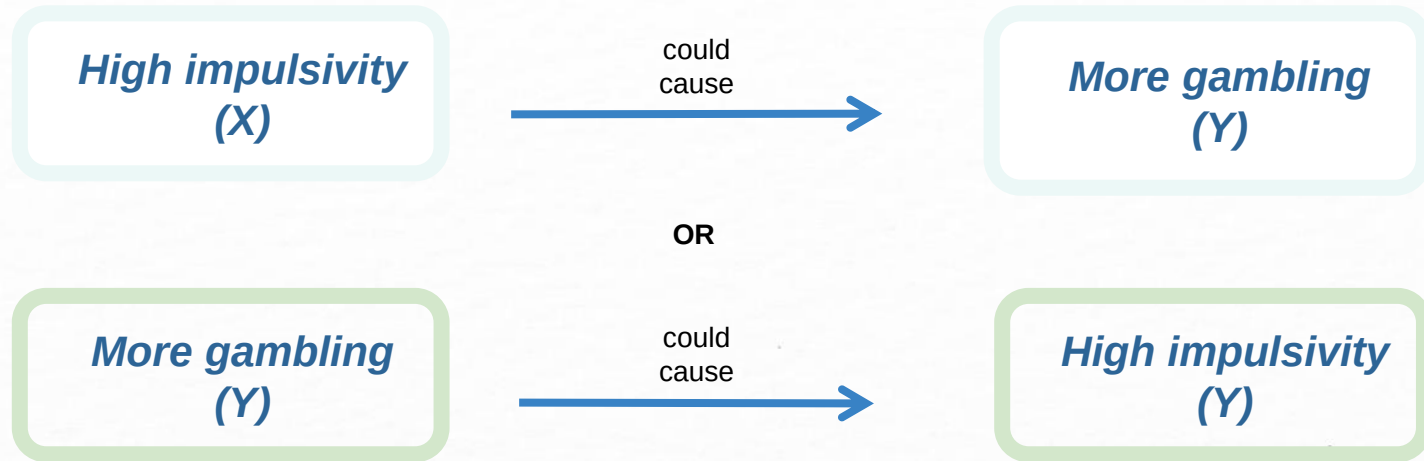
Variables are **measured**, not manipulated

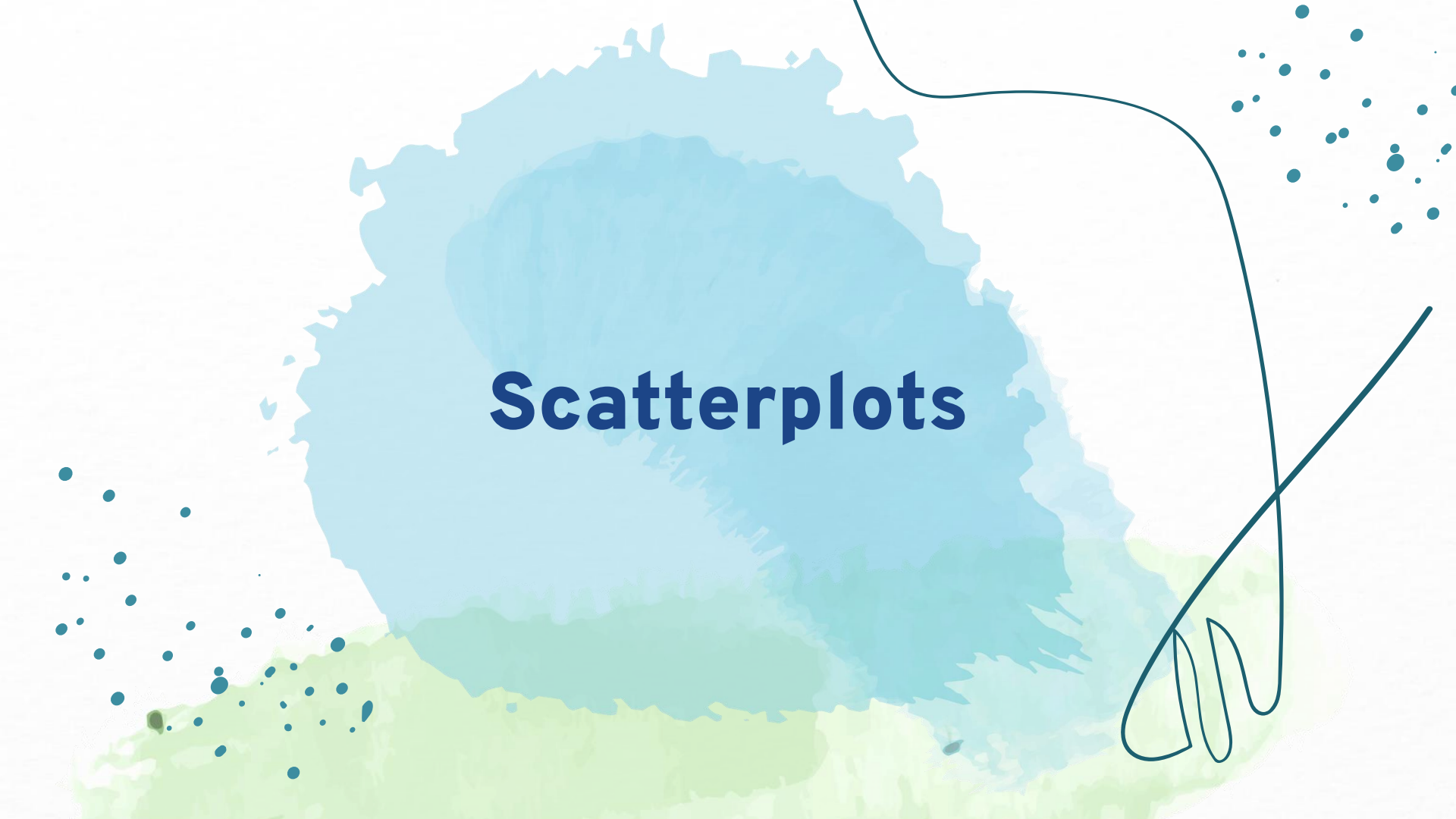
If a **significant relationship** emerges, we can say that impulsivity and level of gambling are **related**

If the finding is non-significant, we simply state that they are **not related**

Example

Remember that correlation does not explore causality, so we cannot say that one **causes** the other



The background features a large, irregular watercolor shape in shades of blue and green. To the left and top right, there are clusters of small, dark blue dots. A thin, dark blue line curves across the right side of the image, ending in a small, stylized loop.

Scatterplots

Scatterplots

An essential part of any preliminary analysis when conducting correlational research.

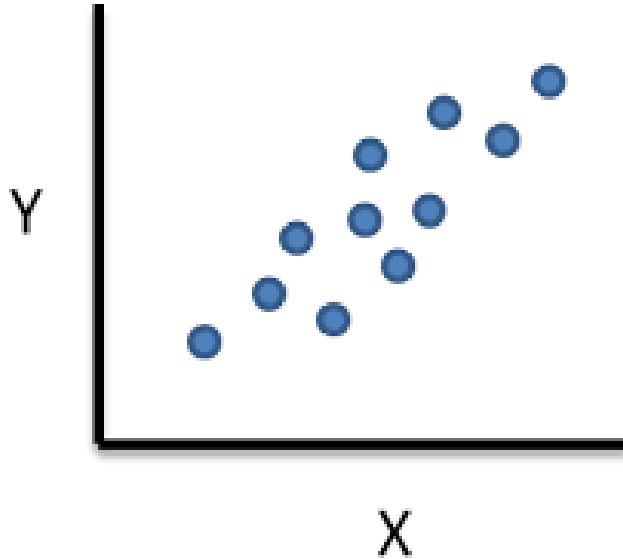
Scatterplot = graph that plots **each participants** score on one variable against their score on the other variable.

The Scatterplot allows you to see whether:

- There is a relationship between 2 variables
- What that relationship is
- What the spread of the data is like (are there outliers)

Types of Correlation / Relationship

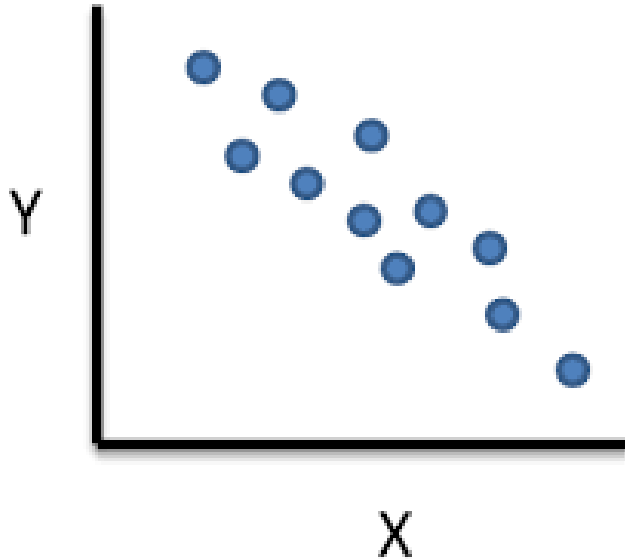
Positive Correlation



Higher scores on one variable are associated with higher scores on a second variable

Types of Correlation / Relationship

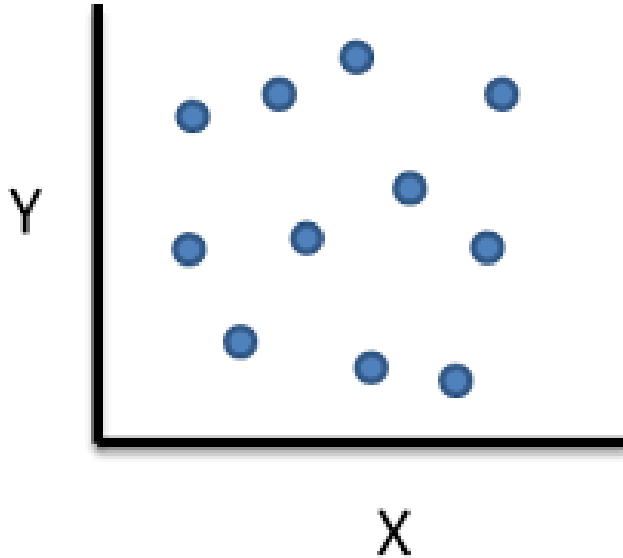
Negative Correlation



Higher scores on one variable are associated with lower scores on a second variable

Types of Correlation / Relationship

No Correlation



The two sets of scores are completely unrelated

How do we know if two variables are related or not?

We carry out a statistical analysis (using SPSS)...

Pearson's Product Moment Correlation (Pearson's correlation coefficient or Pearson's r)

This analysis provides us with two vital numbers to interpret:

1. A value for the **strength** of the relationship/correlation between the two variables
2. A value for the **significance** (probability) that these two variables are, in fact, related

Measurement of Correlation

The strength of the relationship between two variables is the degree to which one variables tends to vary with the other

- If one increases, does the other increase/decrease, too?
- Strength is expressed on a scale ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation)

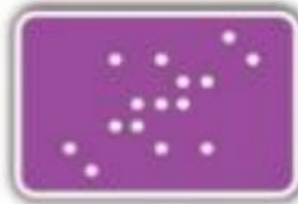
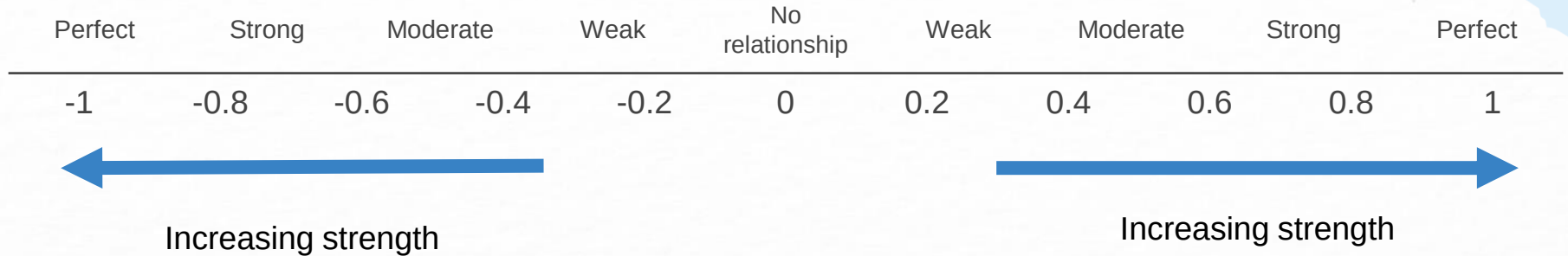
The numerical value indicates the strength

0 means there is no correlation

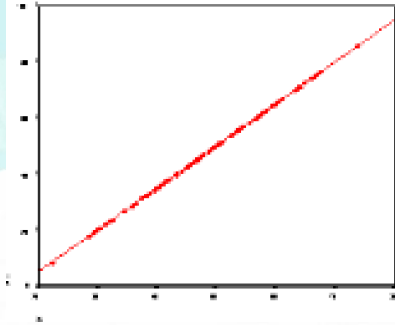
The sign (+/-) indicates direction

- This final number is referred to as a ***correlation coefficient*** and is represented as r

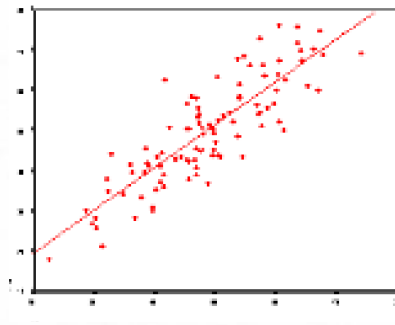
Measurement of Correlation



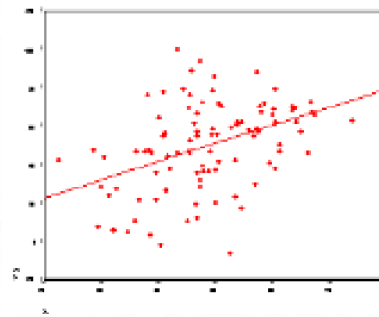
$r = 1.00$



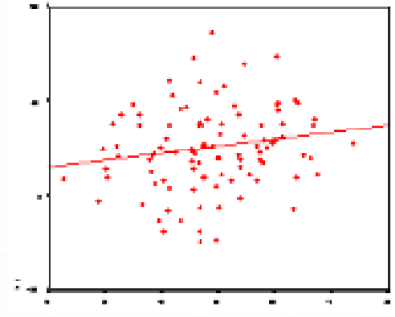
$r = .85$



$r = .42$



$r = .17$



Strength of the relationship is indicated by its numerical value

Expresses the strength of the linear relationship between the variables.

The line running through the points is called a regression line and represents the "best fit".

The background features a central, large, irregular watercolor splash in shades of light green and teal. Surrounding this central splash are smaller, lighter blue watercolor splashes in the corners. Scattered throughout the background are numerous small, dark blue dots, some of which are grouped together in clusters. There are also some faint, dark blue line drawings, including a teardrop shape on the left and some curved lines on the right.

Example studies

Example 1

After running a correlation analysis on data from a study measuring levels of impulsivity and gambling behaviours, the following values were obtained:

$$r = .74$$

$$p = .001$$

1. What is the **DIRECTION** of the relationship?
2. What is the **STRENGTH** of the relationship?
3. Are the two variables **statistically related**?

Example 2

After running a correlation analysis on data from a study measuring temperature and the sale of hot drinks, the following values were obtained:

$$r = - .52$$

$$p = .042$$

1. What is the **DIRECTION** of the relationship?
2. What is the **STRENGTH** of the relationship?
3. Are the two variables **statistically related**?



SPSS Procedure

SPSS Procedure

1

- Analyze

2

- Correlate

3

- Bivariate

Reporting Correlation Findings

- When presenting the results of a correlation analysis you need to remember to state:
- the **Pearson's statistic (r)**
- the **degrees of freedom**
- and the **probability/significance**

Df value is **not** generated by SPSS, but it needs to be reported to adhere to APA standards.

Df for Pearson's correlation:

$N - 2$

SPSS Output

Correlations

		Impulsivity	Gambling
Impulsivity	Pearson Correlation	1	.820**
	Sig. (2-tailed)		.004
	N	10	10

$$r = .82$$

$$p = .004 \text{ (or report as } p < .05)$$

$$df = N - 2$$

$$df = 8$$

Reporting Correlation Findings

$$r(df) = \underline{\hspace{1cm}}, p </> .05$$

OR

$$r(df) = \underline{\hspace{1cm}}, p = \underline{\hspace{1cm}}$$

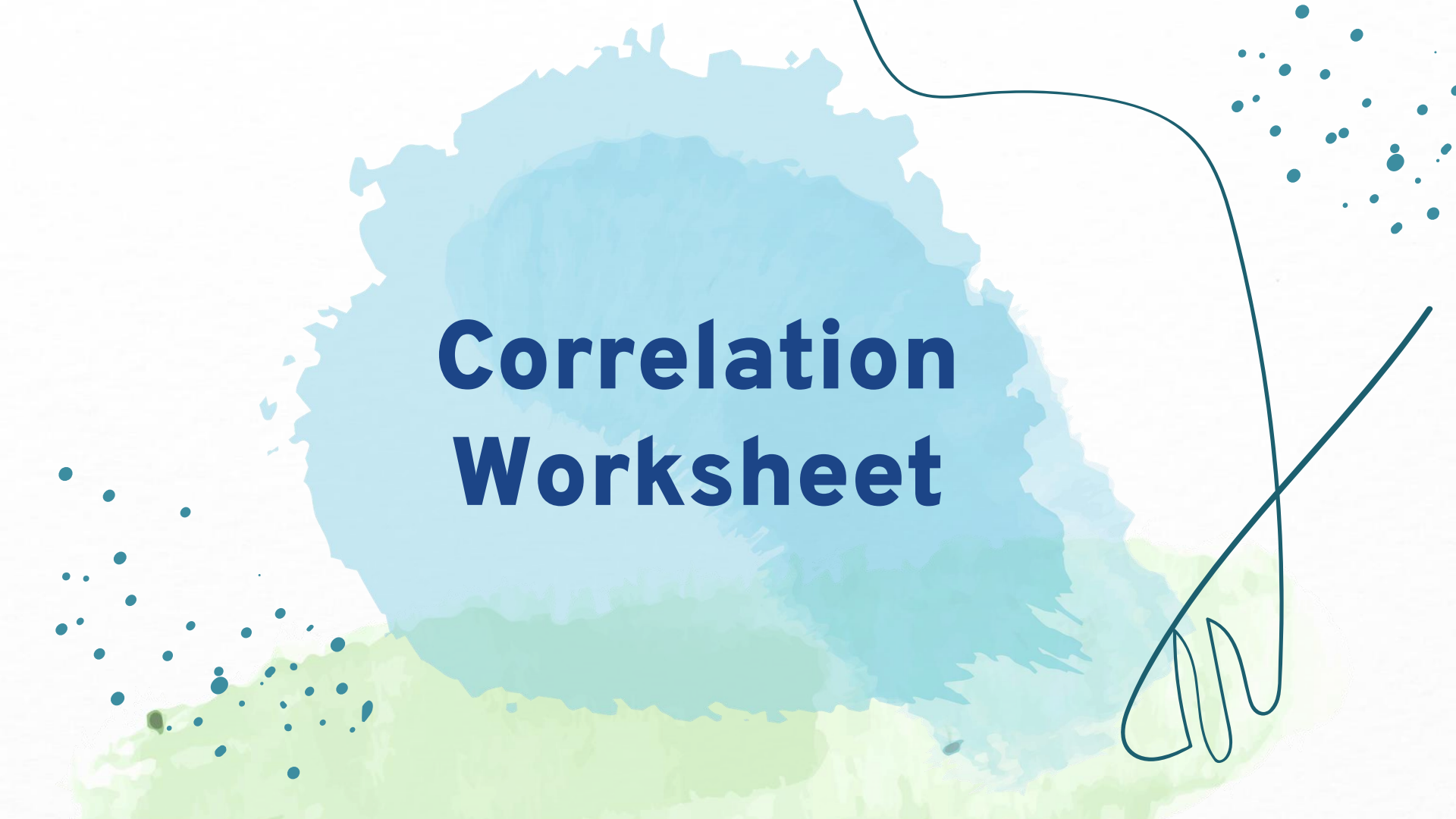
$$\begin{aligned} r &= .82 \\ df &= 8 \\ p &= .004 \end{aligned}$$

$$r(8) = .82, p = .004$$

Reporting Correlation Findings: Results Section

The relationship between impulsivity and gambling behaviour was investigated using Pearson product-moment correlation coefficient.

There was a strong, positive correlation between the two variables, $r(8) = .82, p = .004$, with high levels of impulsivity associated with greater engagement in gambling.

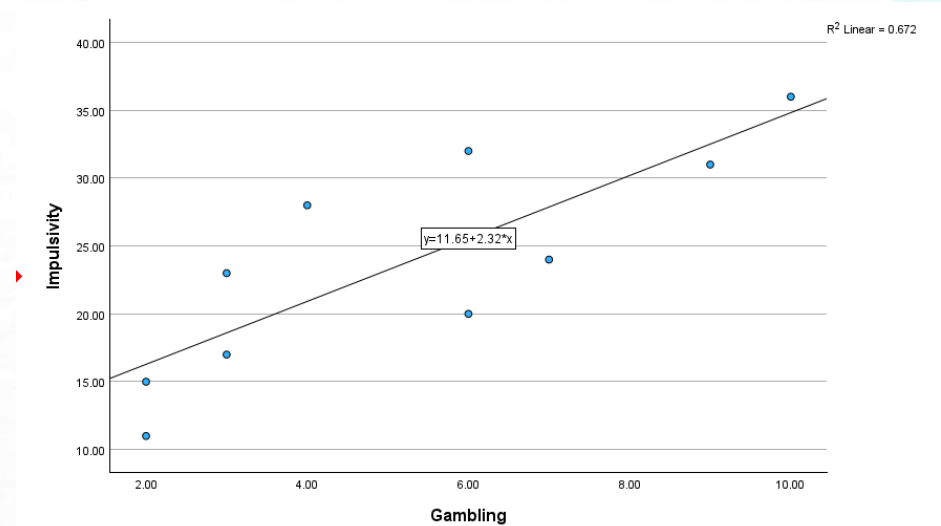
The background features a light blue watercolor wash in the upper center, transitioning into a light green wash at the bottom. Scattered throughout are small, dark blue dots. A thin, dark blue line curves from the top right, looping around the right side of the central text area, and ending in a series of three loops at the bottom right.

Correlation Worksheet

Reporting Correlation Findings: Results Section

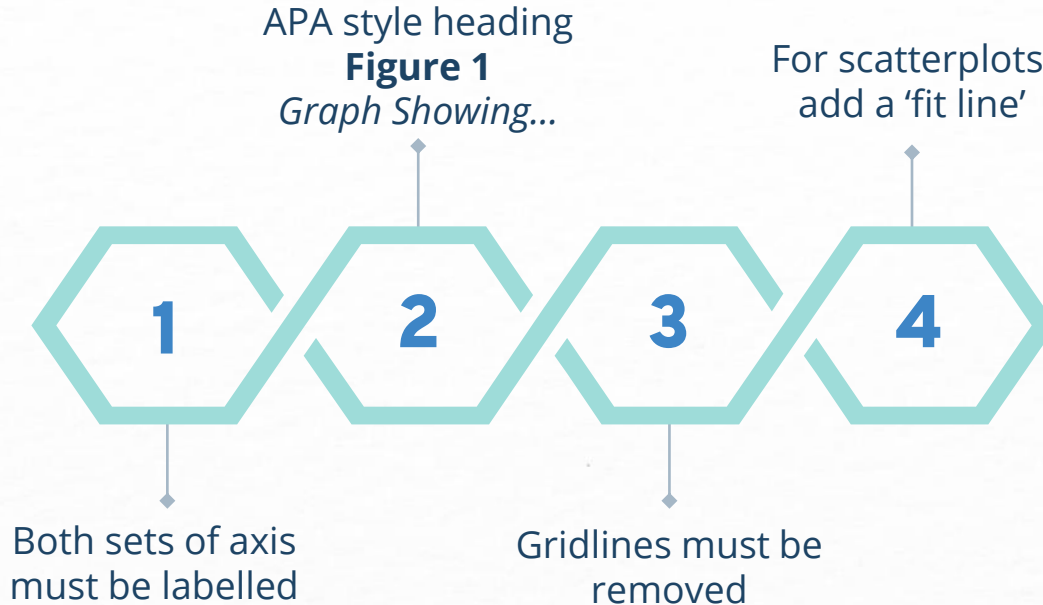
If a significant relationship is observed, it is useful to also provide a scatterplot as a visual representation of the correlation

Although SPSS can provide a scatterplot – you must create your own if you intend to include a chart(figure) of the data in the results section of a report



Scatterplots

There are specific APA conventions for labelling figures that are added to research reports

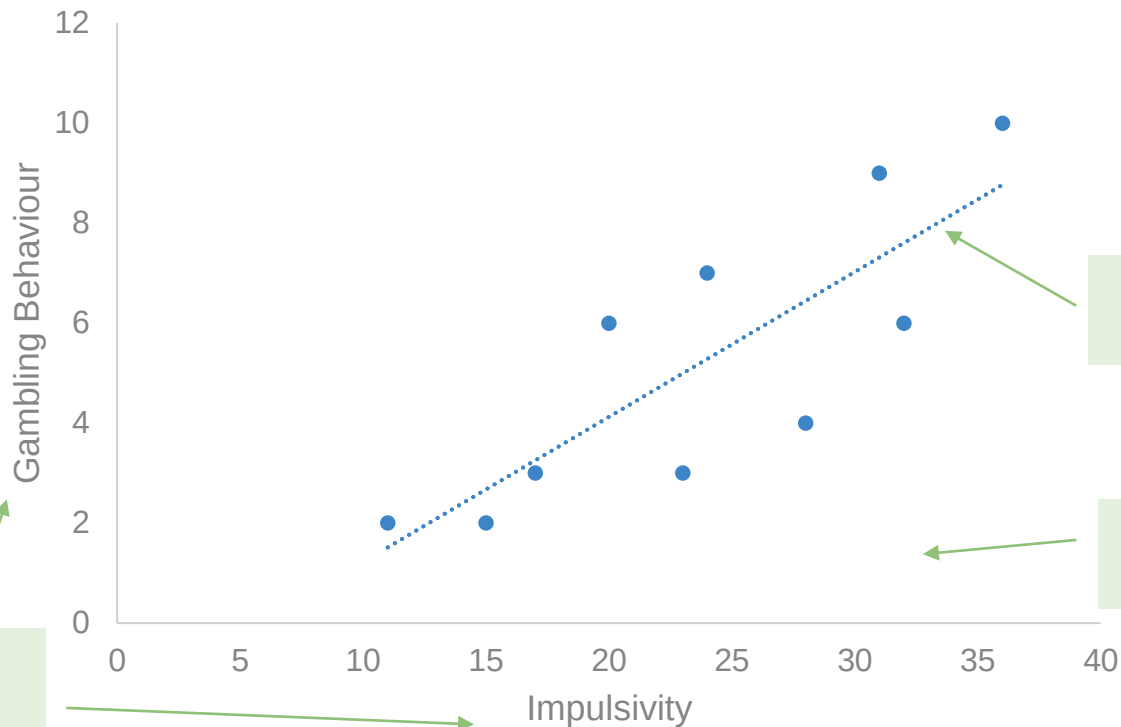


USE HOUSE
RULES
DOCUMENT!

Figure 1

Scatterplot Demonstrating the Relationship Between Levels of Impulsivity and Gambling Behaviour

APA style heading
Figure 1
Graph Showing...

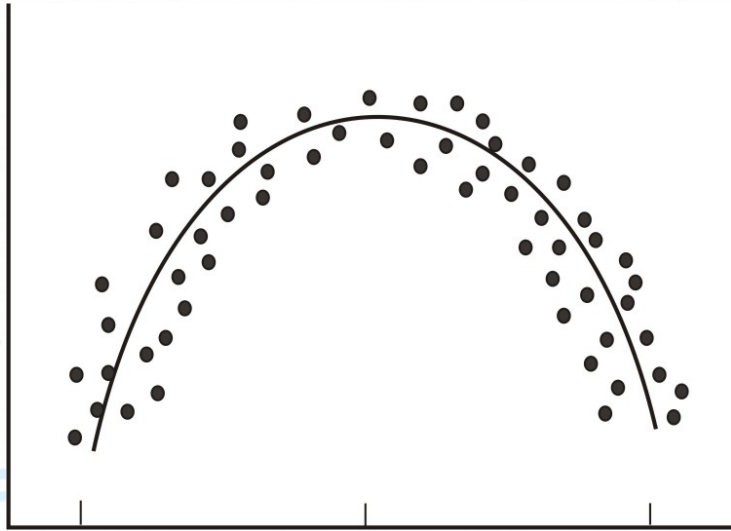


Both sets of axis
must be labelled

For scatterplots,
add a 'fit line'

Gridlines must be
removed

Curvilinear Relationships



- Pearson's r is a good measure only for linear relationships, in which the points are best approximated by a straight line.
- It is **not** a good measure for nonlinear relationships, in which the points are better approximated by a curved line.
- It is important to make a scatterplot and confirm that a relationship is approximately linear **before** using Pearson's r

Spearman's Rank (Spearman's Rho)

- A Pearson's correlation cannot be used if:
 - Data are not parametric, or
 - When variables are not capable of quantitative measurement but can be arranged in serial order, or
 - When you do not get a linear relationship
- Instead, the non-parametric Spearman Rank correlation would be used (Spearman's rho).



**Any
Questions?**