Quantitative research methods

Dr. Luca Longo

Artificial Intelligence and Cognitive Load Research Lab
School of Computer Science, Technological University Dublin

ADAPT, The global centre of excellence for digital content and media innovation
AIRC, The Applied Intelligence Research Center
D-Real, SFI center for research training in digitally-enhanced reality
CeADAR, EI Ireland's centre for applied Artificial Intelligence
ML-Labs, SFI center for research training in machine learning

✉ **luca.longo@tudublin.ie**
http://www.lucalongo.eu/

Definition of quantitative research - 1 of 2

‘an inquiry into a socialÂ or human problem based on testing a theoryÂ composed of variables, measured with numbers, and analyzed with statistical procedures, inÂ order to determine whether the predictive generalizations of the theory hold true.’

(Creswell, J.)

Definition of quantitative research - 2 of 2

'a formal, objective, systematic process in which numerical data are utilised to obtain information about the world'

(Burns & Grove, as cited by Cormack, 1991, p. 140).

Characteristics of quantitative research - 1 of 3

In a nutshell, quantitative research is about **quantifying** the **relationships** between **variables**.

Variable are measured and **statistical models** are built to explain what it has been **observed**.

Purposes: hypotheses testing, control, confirmation, prediction.

Characteristics of quantitative research - 2 of 3

In quantitative studies, **all aspects** of the research are **carefully designed** before data is collected.

Quantitative research is inclined to be **deductive (it tests theory)** contrarily to qualitative research (it generates theory).

The researcher tends to remain **objectively separated from the subject matter**

Characteristics of quantitative research - 3 of 3

In quantitative research, **the quantity** is the unit of analysis:

- amounts
- intensities
- values
- degrees
- frequencies

It uses **statistics** for greater **precision** and **objectivity**

Characteristics of quantitative research - key terms

- **concept**: abstract thinking to distinguish it from other elements
- **construct**: theoretical definition of a concept. It must be observable or measurable and it is linked to other concepts
- **variable**: it is present in research questions and hypotheses
- **operationalisation**: specifically how the variable is observed or measured

A variable is:

- an element that is identified in the hypothesis and/or the research question
- a property or characteristic of people/things that varies in quality or magnitude

A variable must have:

- two or more levels
- identified as independent or dependent

Quantitative research - independent variables

Independent variable:

- its manipulation/variation is the cause of change in other variables
- technically, ‘independent’ is the term reserved for experimental studies
- also refereed to as antecedent, causal, predictor experimental or treatment variable

Dependent variable:

- is the primary interest
- research questions and/or hypothesis describe, explain or predict changes in it
- it is influenced or changed by the independent variable/s
- also referred to as criteria, outcome or target variable

Quantitative research - dependent vs independent variables

- it is not possible to specify independent variables without specifying dependent variables
- the number of independent and dependent variables depends on the nature and complexity of the study
- the number and type of variables dictates which statistical test will be used

Major types of quantitative research

- descriptive research
- meta analysis research
- correlation research
- causal-comparative research
- experimental research

Quantitative descriptive research - overview

Descriptive research is to describe the **characteristics or behaviours** of a given population in a systematic and **accurate fashion** (Leary 2014).

Typically it is **not designed to test hypotheses**, rather it is aimed at **providing information** about the physical, social, behavioural, economic or psychological characteristics of some group of people (or other entities).

Quantitative descriptive research - types

Three main types of descriptive research:

- survey research
- demographic research
- epidemiological research

These types of research provide a **picture** of the extent to which **people** tend to **feel, behave and think**.

Quantitative descriptive research - surveys

Surveys are virtually used in every area of social and behavioural sciences.

Surveys are **different** from **questionnaires**.

A questionnaire is only a method for conducting a survey.

Other methods to collect data include: interviews, observational techniques

Quantitative descriptive research - survey designs

Various types of survey design exist:

- **cross-sectional**: a single group of respondents – a cross-section – of the population is surveyed
- **successive independent samples**: two or more samples of respondents answer the same questions at different points in time
- **longitudinal or panel survey**: a single group of respondents is questioned more than once

Quantitative descriptive research - demographic

Demographic research is aimed at describing and understanding patters of basic life experience of participants (birth, marriage, employment, migration, etc.).

Demography is the statistical study of any dynamic population – including human beings – that changed over space and time.

Quantitative descriptive research - epidemiological

Epidemiological research is aimed at studying the occurrence of disease and death in different groups of people.

It is mainly carried out by health care practitioners, psychologists and in general behavioural scientists.

Quantitative descriptive research - describing/presenting data - criteria

Descriptions of data need to be useful and should meet a set of criteria:

- accuracy
- conciseness
- understability

Unfortunately, *the most accurate descriptions are often the least useful* because they overload the reader with information.

Quantitative descriptive research - describing/presenting data - methods

Strictly speaking, the most accurate are **raw tables of data**.

Methods of summarising and describing sets of data include:

- numerical
- graphical
- mixed methods

Quantitative descriptive research - describing/presenting data - techniques

Frequency distributions:

- simple frequencies
- grouped frequencies
- frequency histograms
- polygons

Measures of central tendency:

- mean
- median
- mode

Measures of variability

- range
- variance
- standard deviation
- the normal curve

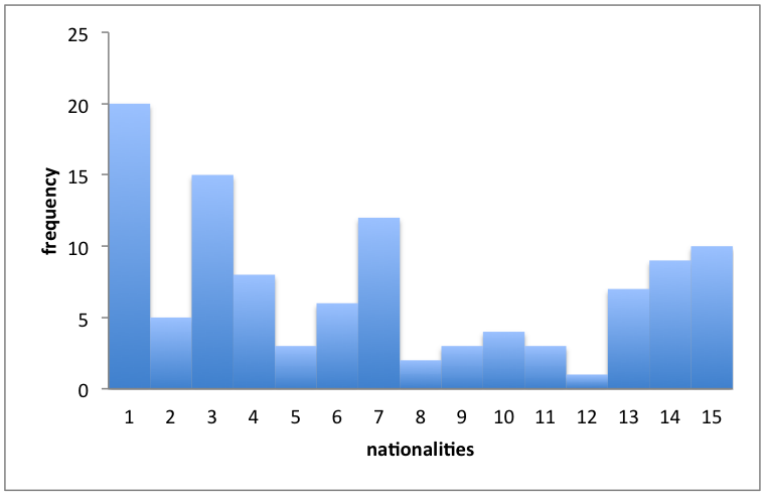
The Z-score

Quantitative descriptive research - describing/presenting data - frequency distributions - histograms and polygons

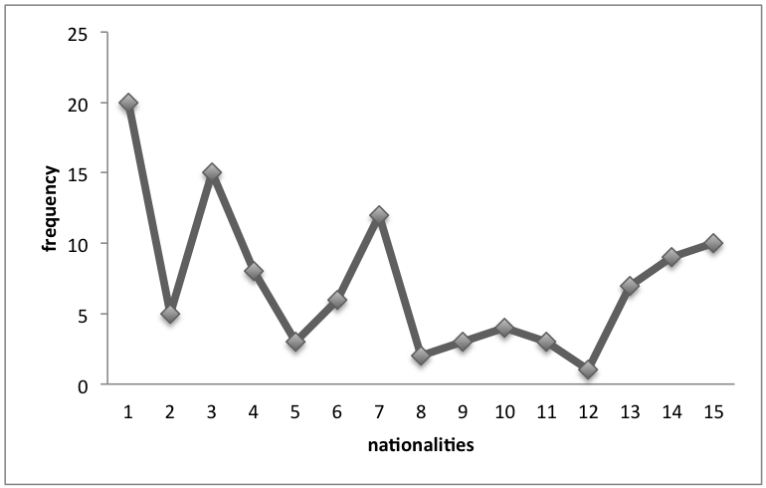
The number of the nationalities of your friends (eg. 5 people have friends of 2 different nationalities)

nationalities	frequency
1	20
2	5
3	15
4	8
5	3
6	6
7	12
8	2
9	3
10	4
11	3
12	1
13	7
14	9
15	10

frequency histogram



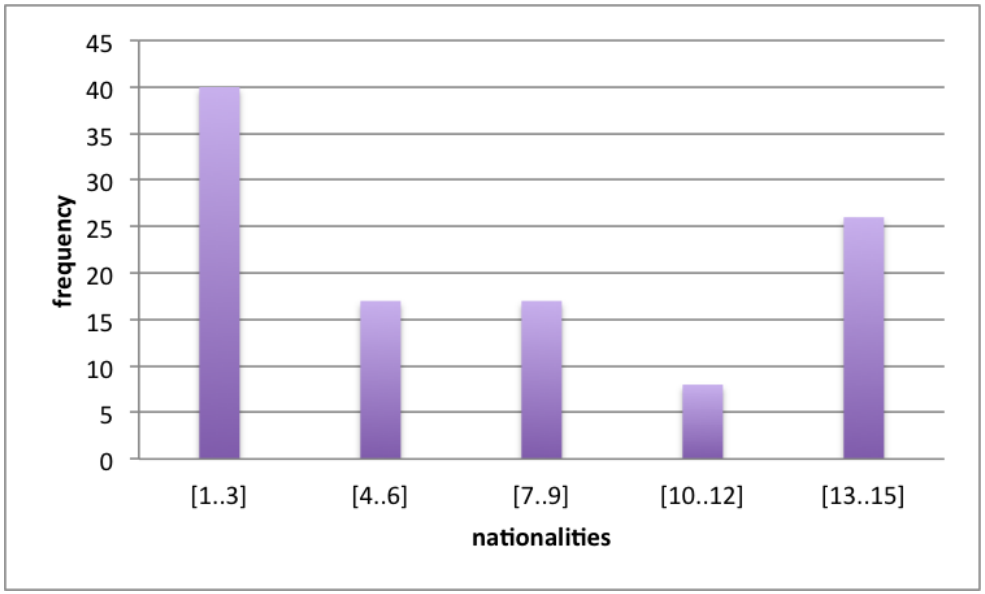
frequency polygons



Quantitative descriptive research - describing/presenting data - frequency distributions - grouped

class	frequency	relative frequency
[1..3]	51	0.37
[4..6]	6	0.16
[7..9]	17	0.16
[10..12]	7	0.07
[13..15]	13	0.24

grouped frequency histogram

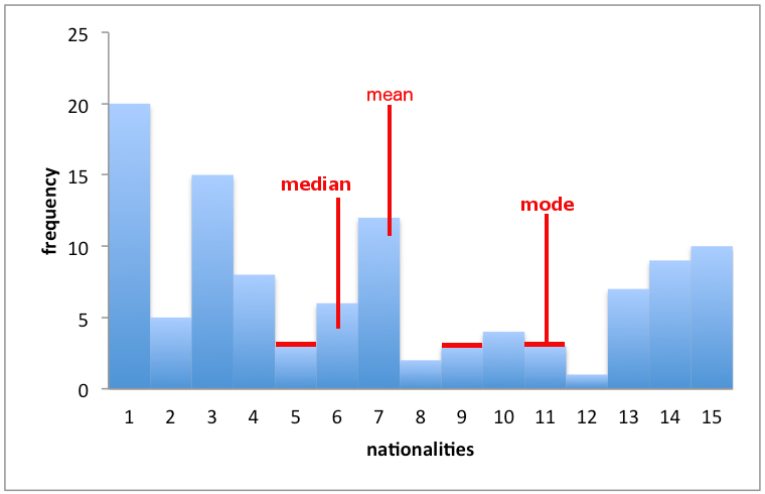


Quantitative descriptive research - describing/presenting data - measures of central tendency

- **mean**: average $= \sum_1^n X_i / n$
- **mediam**: middle score of ranked distribution
- **mode**: most frequent score

mean	7.20
mediam	6
mode	3

frequency histogram

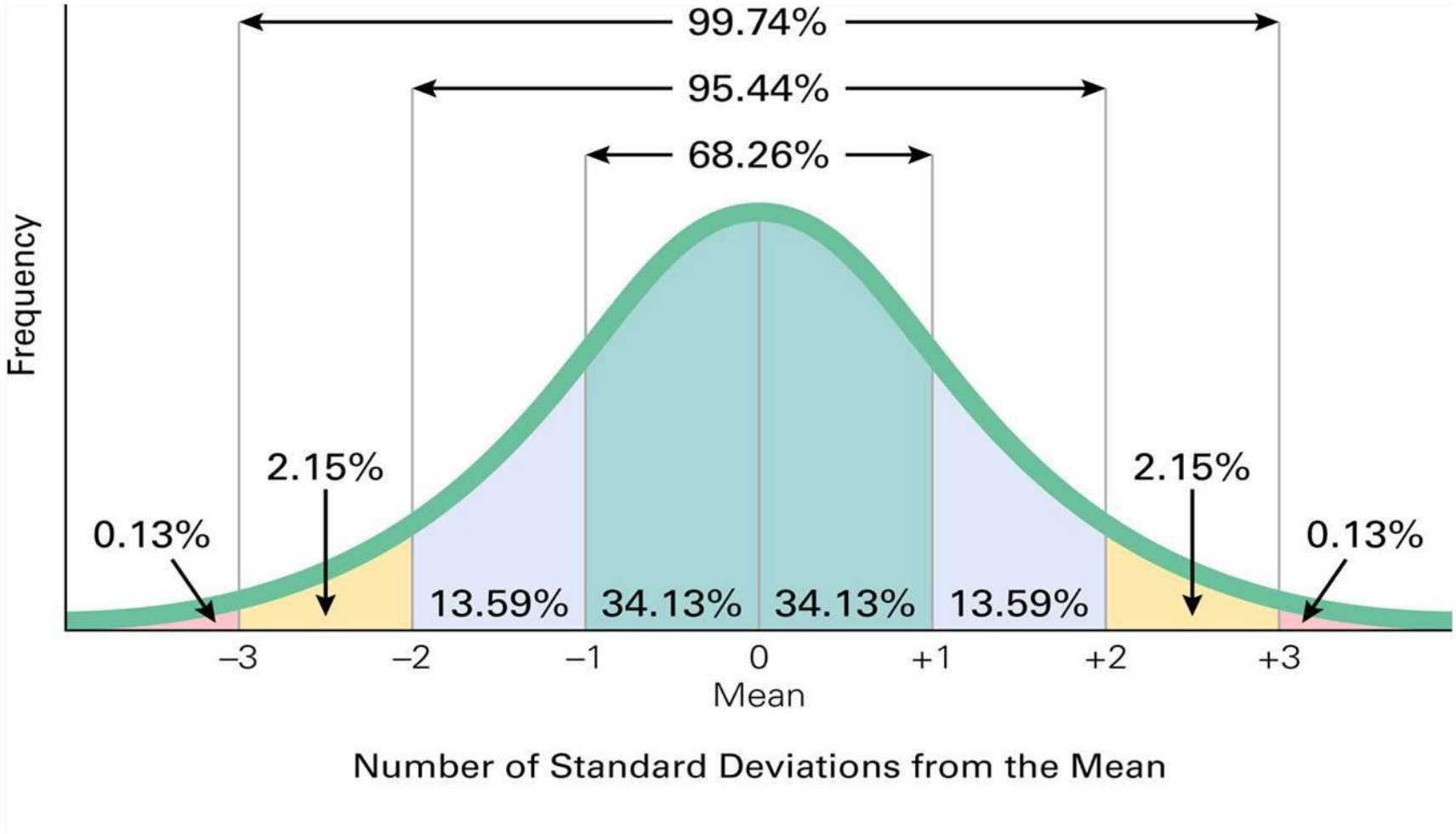


Quantitative descriptive research - describing/presenting data - measures of variability

The extent to which the scores in a distribution vary.

- **Range:** the difference between the largest and the smallest score in a distribution
- **Variance**: sum of the squared differences between each score and the mean
- **Standard deviation:** square root of variance

Quantitative descriptive research - describing/presenting data - measures of variability - std & normal curve



Quantitative descriptive research - describing/presenting data - Z-score

The **Z-score**, or *standard score* is used to describe a particular participant's score (or a particular data point in a distribution) relative to the rest of the data.

$$z = \frac{(y_i - \overline{y})}{s}$$

with y_i the participant score, $\overline{y}$ the mean and s the standard deviation.

Z indicates how far from the mean, in terms of standard deviation, the participant's score falls.

In a nutshell, a **meta-analysis** is a synthesis of available studies about a topic aimed at producing a single summary.

It includes:

1. a systematic identification of similar studies
2. a definition of the study to keep for the meta-analysis
3. the use of structured formats to formally summarised information taken from selected studies
4. a summary of the estimate of the effect and a test of the homogeneity of the studies

Quantitative research - correlation - 1 of 2

Correlational research is aimed at investigating whether & **to what degree** a **relationship** exists between **two/more** (numerical) **variables**.

When two variables are correlated, the relationship can be used for prediction (eg. **regression**).

Notes: even in the presence of a significant relationship between two variables it does not mean that one causes the other.

Correlation does not mean causation.

Correlation implies prediction but not causation.

A correlation coefficient is a statistics that indicates the degree to which two variables are related to each other **linearly**.

The most commonly used measure of correlation is the **Pearson correlation coefficient**, indicated by the letter *r* **or** ρ.

Two random variables X and Y is defined as

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{Var}(X)\mathbf{Var}(Y)}}$$

with *cov* the covariance and *var* the standard deviation of X

Two samples x_i and y_j is defined as

$$r = \frac{\mathbf{S_{xy}}}{\sqrt{\mathbf{S_{xx}}\mathbf{S_{yy}}}}$$

with S the sum

ρ and r return values in the range $[-1..1] \in \Re$

Given two variables X and Y, r can be rewritten as:

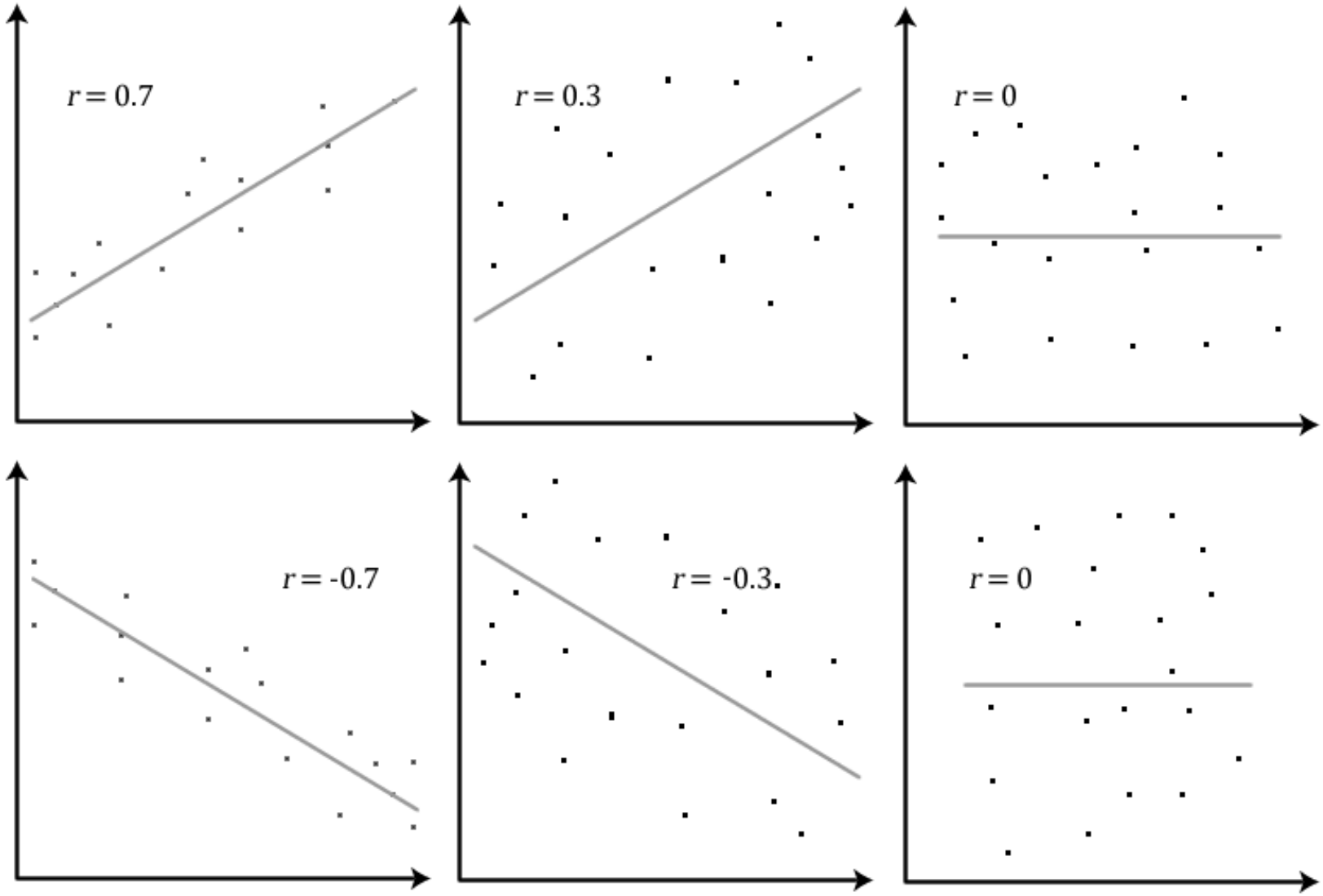
$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

Example:

X = $\{1, 3, 4, 4\}$ Y=$\{2, 5, 5, 8\}$

$r = 0.866$ (positive linear correlation)

Quantitative research - correlation - Pearson correlation - plots



Is the value of *r* statistically significant?

Statistical significance exists when a correlation coefficient calculated on a sample has a very low probability of being zero in the population.

It is affected by:

1. sample size
2. magnitude of correlation
3. degree of cautiousness allowed (eg. $p < 0.05$)

Quantitative research - correlation - Pearson correlation - statistical significance of *r* - p-value

The $\mathbf{p-value}$ is a number between 0 and 1 that represents the probability that data would have arisen if the null hypothesis was true.

Typically researchers will consider a correlation to be significantly different from zero if there is less than 5% chance that a correlation as large as the one they obtained could have come from a population with a true correlation of zero.

Quantitative research - correlation - Pearson correlation - testing statistical significance of *r*

The calculation of the $p - value$ is based on a number of assumptions that are beyond the scope of this presentation. However, these can be found in standard statistical tables [1]:

N	0.1	0.05	0.01	0.001
4	0.900	0.950	0.990	0.999
5	0.805	0.878	0.959	0.991
6	0.729	0.811	0.917	0.974
7	0.669	0.754	0.875	0.951
8	0.621	0.707	0.834	0.925
9	0.582	0.666	0.798	0.898
10	0.549	0.632	0.765	0.872
11	0.521	0.602	0.735	0.847
12	0.497	0.576	0.708	0.823
13	0.476	0.553	0.684	0.801
14	0.458	0.532	0.661	0.780
15	0.441	0.514	0.641	0.760
20	0.378	0.444	0.561	0.679

N	0.1	0.05	0.01	0.001
30	0.306	0.361	0.463	0.570
40	0.264	0.312	0.403	0.501
50	0.235	0.279	0.361	0.451
60	0.214	0.254	0.330	0.414
70	0.198	0.235	0.306	0.385
80	0.185	0.220	0.286	0.361
90	0.174	0.207	0.270	0.341
100	0.165	0.197	0.256	0.324
200	0.117	0.139	0.182	0.231
300	0.095	0.113	0.149	0.189
400	0.082	0.098	0.129	0.164
500	0.074	0.088	0.115	0.147
1000	0.052	0.062	0.081	0.104

[1] column 1 is the number of participants, the other columns are Two-Tailed Probabilities

Quantitative research - correlation - Pearson correlation - hypothesis testing example

Research question: Is the relationship between height and intelligence (QI) of students significant?

Null hypothesis (H_0):

'height & intelligence of students are linearly unrelated'

Alternative hypothesis (H_1):

'height & intelligence of students are linearly related'

Formally: $\mathbf{H_0}$: $p = 0$, $\mathbf{H_1}$: $p \neq 0$

Example: participants: $N = 64$, Correl. coefficient: $r = 0.15$, $\alpha = 0.05$

The minimum value of r that is significant with 64 students lies within [0.254..0.235] (roughly 0.245).

Because $r = 0.15 < 0.245$ there is strong evidence to accept the null hypothesis (thus reject the alternate hypothesis).

The strength of a correlation is assessed only by its magnitude, not whether it is statistically significant.

In behavioural studies, a **rule of thumbs** is to regard correlations as:

- **weak** in magnitude: [0-0.30]
- **moderate** in magnitude: [0.30-0.50]
- **strong** in magnitude: [0.50-1]]

Factors that might distort correlation coefficients: restricted ranges, outliers, reliability of measures

The assumptions behind Pearson Correlation are:

- **level of measurement**: each variable should be continuous
- **related pairs**: each participant/observation should have a pair of values
- **absence of outliers**: not outliers in either variable
- **normality of variables**: each variable should follow a normal distribution
- **linearity**: a 'straight line' relationship between the variable should be formed
- **homoscedasticity**: literally, same variance

If these assumptions are not met, then other techniques can be used: **Spearman** or **Kendall rank-order** correlations.

Quantitative research - correlation - limitations

Correlation research can provide only **tentative conclusions** about **cause-and-effect** relationships, demonstrating how variables are correlated.

Researchers would also like to understand **why** and **how** variables are related.

Quantitative research - causal-comparative - 1 of 2

Causal-comparative research (ex post facto) is aimed at establishing and assessing **cause-effect relationships** among the variables of the study.

The objective is to establish **how** values of the independent variable/s have a **significant effect** on the dependent variable and **why**.

Casual/comparative research usually involves **group comparisons** making up the values of the independent variable/s (eg. male vs. female).

The independent variable/s is/are **not under the control of the researcher** who has to take the values as they come (eg. male or female).

The variable under investigation is the **dependent variable**.

Several methods exist for understanding **why** and **how** variables are related

- **regression**: develop equations that describe how variables are related allowing the prediction of a variable given one (or more) variables
- **cross-lagged panel & structural equation analysis**: explore the likely direction of causality between two (or more) variables that are correlated
- **multilevel modelling**: examine relationship among variables that are measured at different levels of analysis
- **factor analysis**: identify basic dimensions that underlie sets of correlations

Quantitative research - experimental

Experimental research is like causal-comparative research: it aims to establish cause-effect relationship among the groups of participants. However, in **experimental research the researcher has control of the independent variables**.

Two types of experimental design:

- true experimental design
- quasi-experimental design

True experimental design

The researcher randomly assigns participants to the groups/conditions that form the independent variable/s of the study and then **measures the effect this group membership has on the dependent variable** (manipulation, randomization, control group).

Quasi-experimental design

It provide alternate means for examining causality when experimental control is not useful/possible. This designs should account for as many threats to validity as possible.

Suggested readings

- Creswell, J. 1994 **Research Design: Qualitative and Quantitative Approaches**. (Sage)
- Leary, 2014 **Introduction to behavioural research methods** (Pearson)
- Nicol, A. and Pexman, P. 2003 **Displaying Your Findings: A Practical Guide for Creating Figures, Posters, and Presentations**, Sixth Edition (American Psychological Association publisher)
- Marder, M.P. 2011 **Research methods for science**. (Cambridge university press)