Data Analysis

Week 2

Transformations

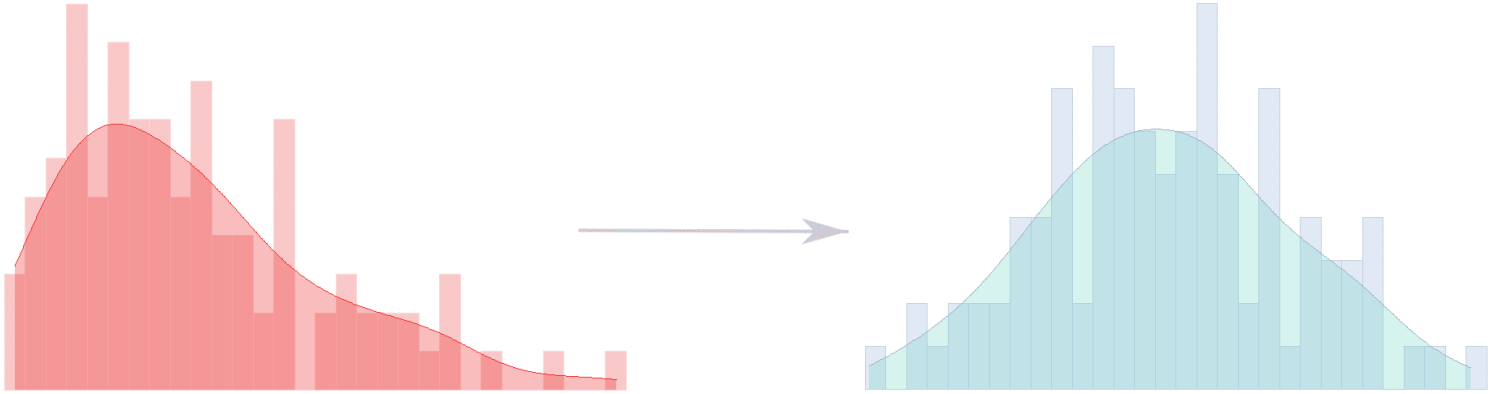
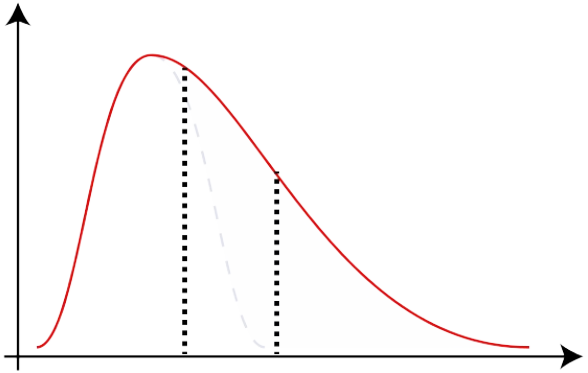


Table of content

Transformations

POSITIVELY SKEWED DATA

- Moderately: SQRT (variable)
- Strongly: LG10 (variable)
- Extremely: 1 / (variable)



NEGATIVELY SKEWED DATA

- Moderately: SQRT ((max+1) – variable)
- Strongly: LG10 ((max+1) – variable)
- Extremely: 1 / ((max+1) – variable)

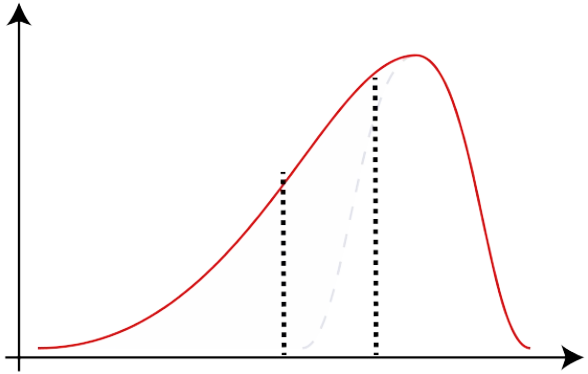


Table of content

Transformations

BOX – COX Transformation (Box & Cox, 1964)

STEP 1:

- Transform ⇢ Rank cases (enter variable) ⇢ Rank type ⇢ Check “Fractional rank”

STEP 2:

- Transform ⇢ Compute variable
- Function group ⇢ Inverse DF
- Function and Special Variables ⇢ Idf.Normal
- In Target variable ⇢ name of your transformed variable
- In numeric expression (fractional rank variable, mean of variable, standard deviation of variable)

Table of content

Reliability

- What is reliability? (week 1)
  - **Reliability:** whether an instrument can be interpreted consistently across different situations
- Cronbach's alpha (α) is the most commonly used measure of internal consistency (a measure of reliability)
- Answers the question:
  - "How much the items on a scale are measuring the same underlying dimension?"

Reliability

SPSS: Analyze ⇢ Scale ⇢ Reliability Analysis

- Recommended values of Cronbach's alpha are > .7 (DeVillis, 2003; Kline, 2005)

- "Corrected Item-Total Correlation" values < .3 are usually cause for concern because it is an indication that this particular item might not be measuring the same construct

Correlation

When I was 8 years old, my parents bought me a guitar for Christmas. Even then, I'd desperately wanted to play the guitar for years. I could not contain my excitement at getting this gift (had it been an electric guitar I think I would actually have exploded with excitement). The guitar came with a 'learn to play' book, and after some time trying to play what was on page 1 of this book, I readied myself to unleash a riff of universe-crushing power on the world (well, 'Skip to my Lou'). But, I couldn't do it. I burst into tears and ran upstairs to hide. My dad sat with me and said something like 'Don't worry, Andy, everything is hard to begin with, but the more you practise the easier it gets.' With his comforting words, my dad was inadvertently teaching me about the relationship, or correlation, between two variables. These two variables could be related in three ways: (1) positively related, meaning that the more I practised my guitar, the better a guitar player I would become (i.e., my dad was telling me the truth); (2) not related at all, meaning that as I practised the guitar my playing ability would remain completely constant (i.e., my dad had fathered a cretin); or (3) negatively related, which would mean that the more I practised the guitar the worse a guitar player I would become (i.e., my dad had fathered an indescribably strange child).

Correlation

- A statistical technique that is used to measure and describe a relationship between two variables
- Two scores for each individual X and Y
- Scores can be presented graphically in a scatterplot and it is a common practice to check them[1]

Correlation

Would you describe the following relationships as positive or negative?

- Age of an automobile and its resale value
- Month of birth and level of anxiety
- Attitudes toward math and math achievement
- Health and length of life
- Height and life expectancy

Why?

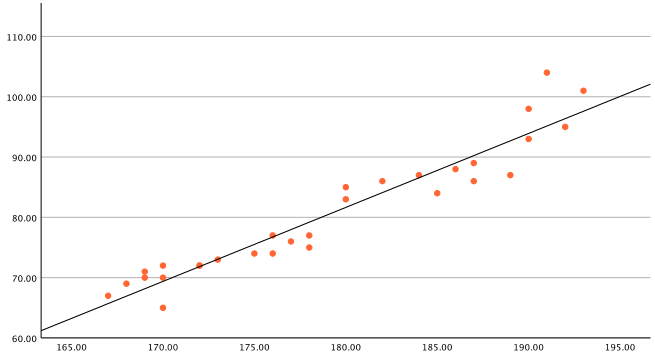
Characteristic of a relationship

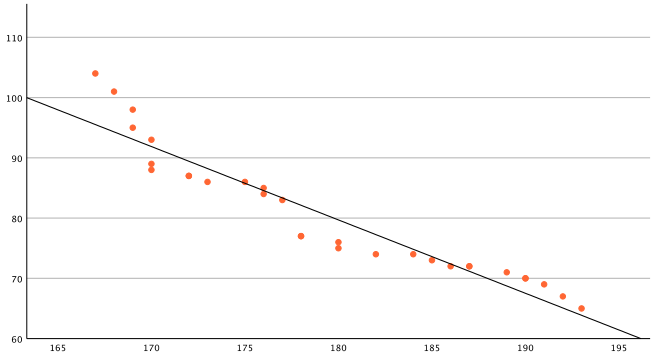
Characteristic:

- The direction of the relationship
- The form of the relationship
- The degree of the relationship

Characteristic of a relationship

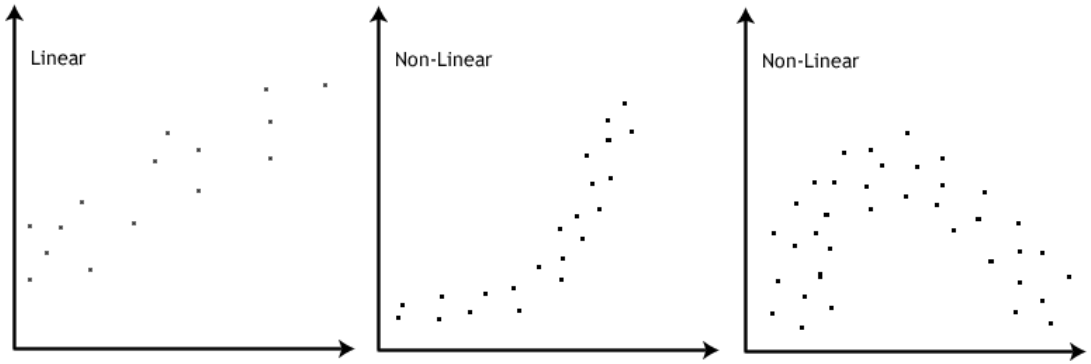
- The direction of the relationship:
  - Positive
  - Negative



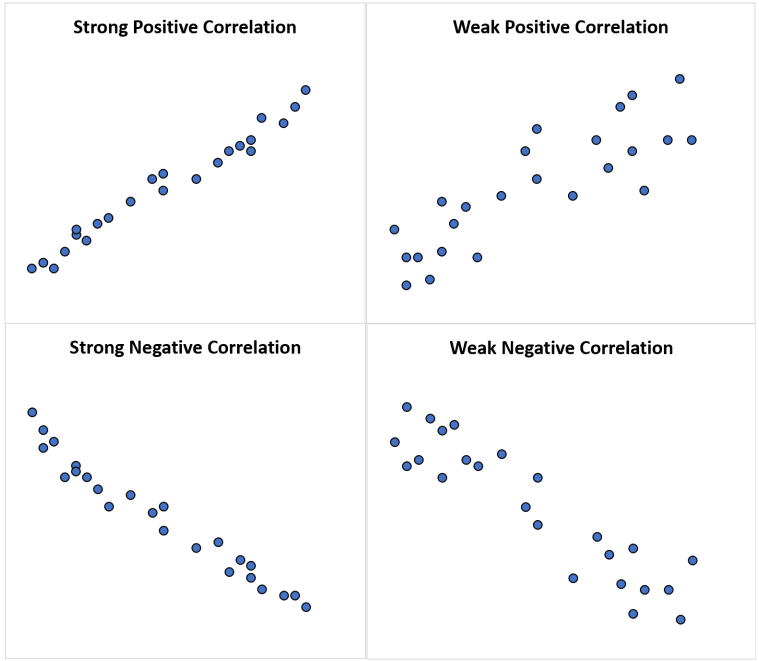


Characteristic of a relationship

- The form of the relationship:
  - Linear
  - Non-linear



Characteristic of a relationship



- The degree of the relationship:
  - Strong
  - Weak

Correlation Coefficient

- Ranges from -1 to +1
- The +/- sign before *r* indicates the direction of the relationship between two variables
- The absolute value of the *r* shows the degree or strength of the relationship between two variables
- *r* does not (and cannot) tell whether the relationship is linear or not

Correlation Coefficient

r Coefficient

- Although it is difficult to state what is strong or weak relationship between variables, Green, Salkin and Akey (2000) suggest:

*"for the behavioural sciences, correlation coefficients of .10, .30 and .50 (absolute value!) are interpreted as **small**, **medium**, and **large** coefficients, respectively"*

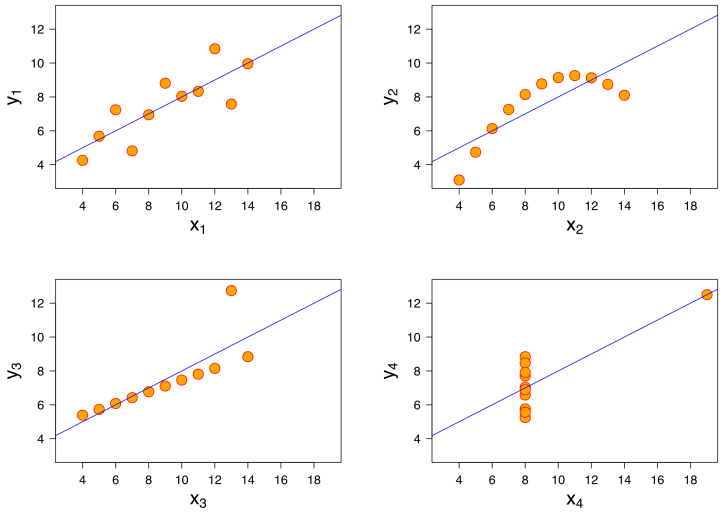
Correlation Coefficient

Size of correlation	Interpretation
0.90 – 1.00 (-0.90 to -1.00)	Very high positive (negative) correlation
0.70 – 0.90 (-0.70 to -0.90)	High positive (negative) correlation
0.50 – 0.70 (-0.50 to -0.70)	Moderate positive (negative) correlation
0.30 – 0.50 (-0.30 to -0.50)	Low positive (negative) correlation
0.00 – 0.30 (-0.00 to -0.30)	Negligible correlation

Anscombe's quartet

- Four sets of data with the same correlation of 0.816
- The four *y* variables have the same mean (7.5), variance (4.12), correlation (0.816), and regression line (y=3+0.5x)

- What is different?[1]



Pearson Correlation

Measures degree and direction of a linear relationship between two variables.

- Variables should be measured in interval or ratio scale
- The relationship should be linear

SPSS: Analyze ⇢ Correlate ⇢ Bivariate

- One tailed (directional): “The more anxious someone is about an exam, the worse their mark will be.”
- Two tailed (non-directional): “I am not sure whether exam anxiety will improve or reduce exam marks.”
- Statistically sig. positive correlation between height and weight, $r = .957$ ($p < .0005$)

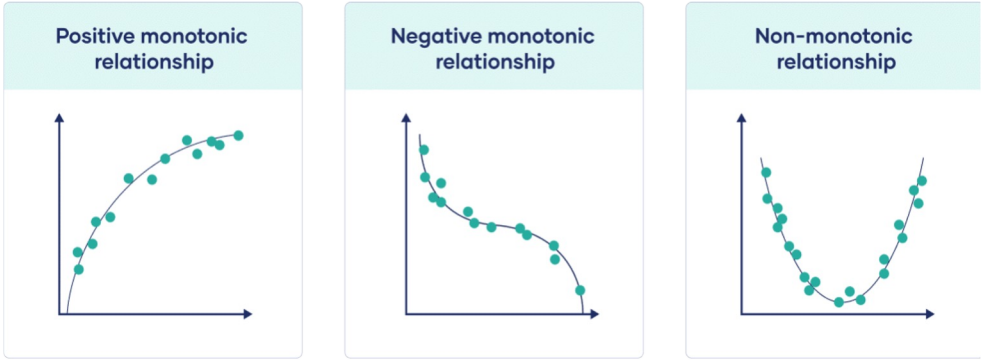
Spearman Correlation

Measures the relationship between variables and it is used when:

- Variables are measured on an ordinal scale of measurement;
- The relationship is not linear form but it is **monotonic** and the researcher wants to measure the degree of consistency of the relationship between two variables.

Spearman Correlation

- Monotonic relationship means consistently one-directional relationship (consistently increasing or decreasing) between two variables.



Spearman Correlation

Computing Spearman correlation (what the process involves):

1. List the cores from smallest to largest
2. Assign a rank (first, second, etc.) to each person
3. When two or more scores are tied, compute the mean of their ranked positions
4. Use the same Pearson formula for the ranked data, the result is called Spearman correlation.

Other correlation tests

- Point-Biserial Correlation: The relationship between one continues and one dichotomous variable
- Phi-Coefficient: The relationship between two dichotomous variables
- Chi-Square: Non-parametric test. Tests how well the obtained sample proportions fit the population proportions

Correlation

Correlation describes a relationship between two variables. It does not imply causality!

however

This should not be taken to mean that correlations cannot indicate the potential existence of causal relations.