

(6) Correlation Analysis

Dr. Wan Nor Arifin

Biostatistics and Research Methodology Unit
Universiti Sains Malaysia
wnarifin@usm.my / wnarifin.github.io



Last update: Jul 17, 2023

Outlines

- Introduction
- Pearson's correlation
- Spearman's rank correlation
- Kendall's rank correlation

Expected outcomes

- Familiarize with correlation analyses between two variables
- Understand and able to interpret the results of the correlation analyses

Introduction

Correlation Analysis

- Statistical Analyses for correlations:
 - Two numerical variables: Pearson's correlation
 - Two numerical/ordinal variables: Spearman's rank correlation
 - Two ordinal variables: Kendall's rank correlation

Pearson's correlation

Pearson's correlation

- Purpose: Determine the strength of linear association between TWO numerical variables
- Assumptions:
 1. Numerical variables
 2. Normally distributed data
- Correlation coefficient, r

Pearson's correlation

Research objective:

To determine the strength of linear association between cholesterol level and age

Research question:

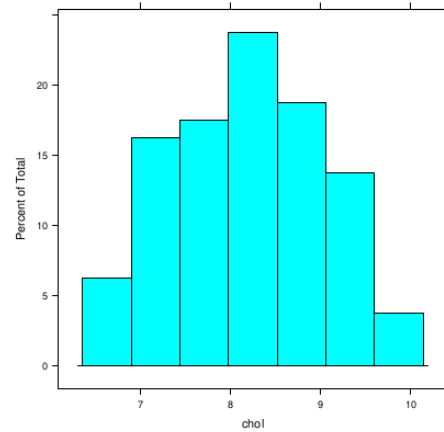
What is the strength of relationship between cholesterol level and age?

Example

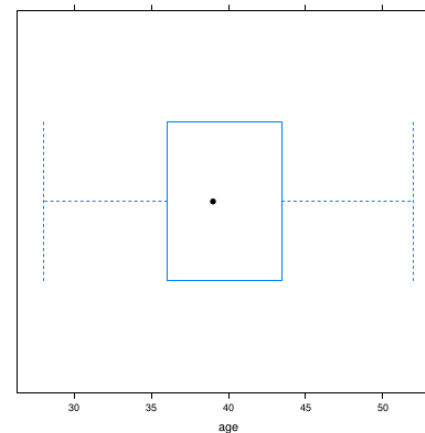
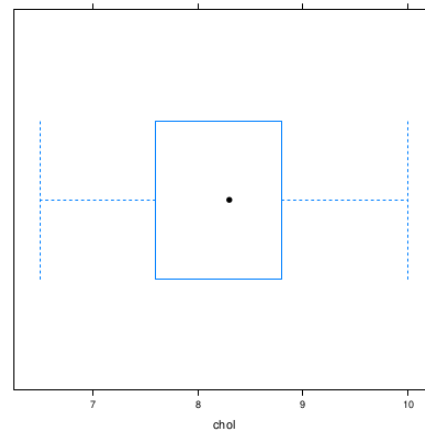
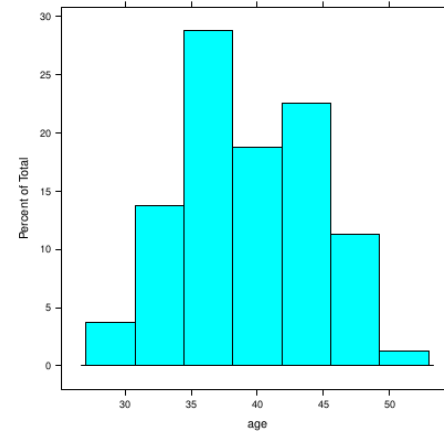
- Sample size: 80
- Variables:
 - cholesterol level in mmol/L
 - Age in years

Normality

Cholesterol

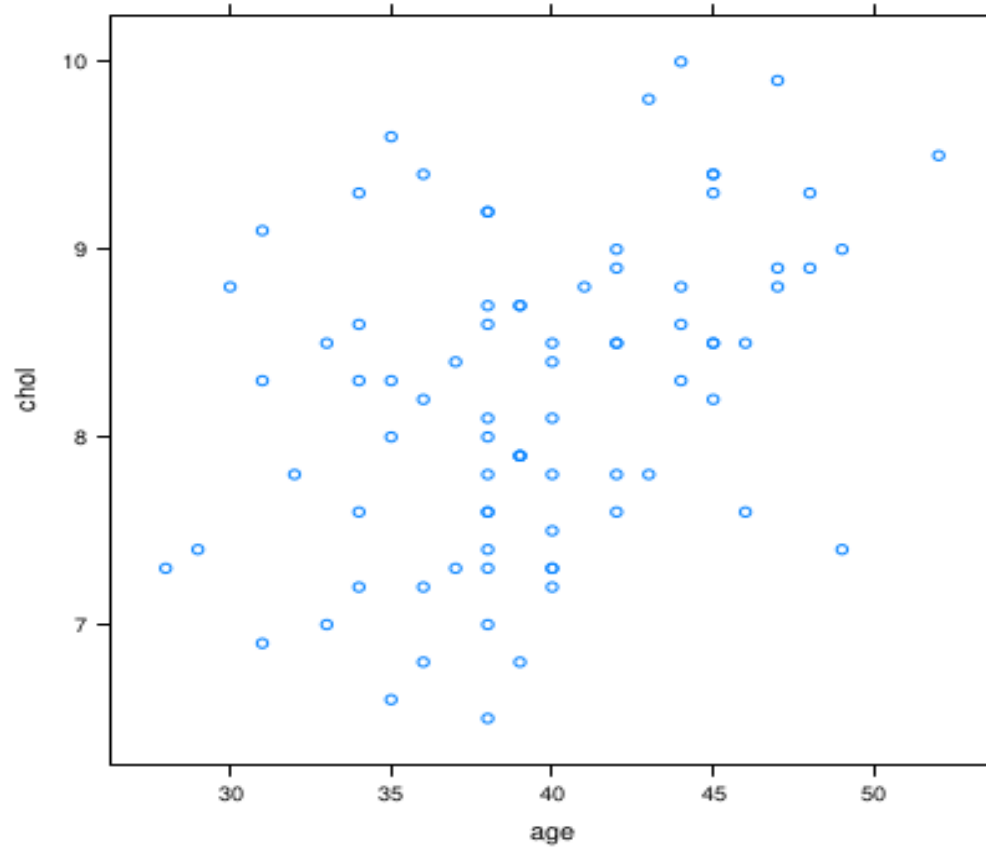


Age



Scatter Plot

Cholesterol vs Age



Results

Pearson's product-moment correlation

```
data: chol and age
t = 3.7119, df = 78, p-value = 0.0003841
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1833492 0.5595401
sample estimates:
      cor
0.3874574
```

P-value

r coefficient = 0.39

→ Significant medium correlation

Strength of correlation	$ r $ coefficient
None or very small	0 – < 0.1
Small	0.1 – < 0.3
Medium	0.3 – < 0.5
Large	0.5 – 1

Spearman's rank correlation

Spearman's rank correlation

- Purpose: Determine the strength of linear association between TWO numerical (or at least ordinal) variables
- Non-parametric version of Pearson's correlation
- Non-normally distributed numerical data and ranked data

Spearman's rank correlation

- The procedure turns numerical \rightarrow ranks
- e.g. Age = 32, 28, 41 \rightarrow 2, 1, 3
- Assumption:
 1. Numerical / ordinal variables
- Correlation coefficient, *rho* (ρ)

Spearman's rank correlation

Research objective:

To determine the strength of linear association between cholesterol level and age

Research question:

What is the strength of relationship between cholesterol level and age?

Example

- Sample size: 80
- Variables:
 - cholesterol level in mmol/L
 - Age in years
- Assuming these variables are not normally distributed

Results

Spearman's rank correlation rho

data: chol and age

S = 53143, p-value = 0.0005641

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.3771289

P-value

rho coefficient = 0.38

→ Significant medium correlation

Strength of correlation	<i>r</i> coefficient
None or very small	0 – < 0.1
Small	0.1 – < 0.3
Medium	0.3 – < 0.5
Large	0.5 – 1

Kendall's rank correlation

Kendall's rank correlation

- Purpose: Determine the strength of linear association between TWO ordinal variables
- Assumption:
 1. Ordinal variables
- Correlation coefficient, tau (τ)

Kendall's rank correlation

Research objective:

To determine the strength of ordinal association between education level and parity

Research question:

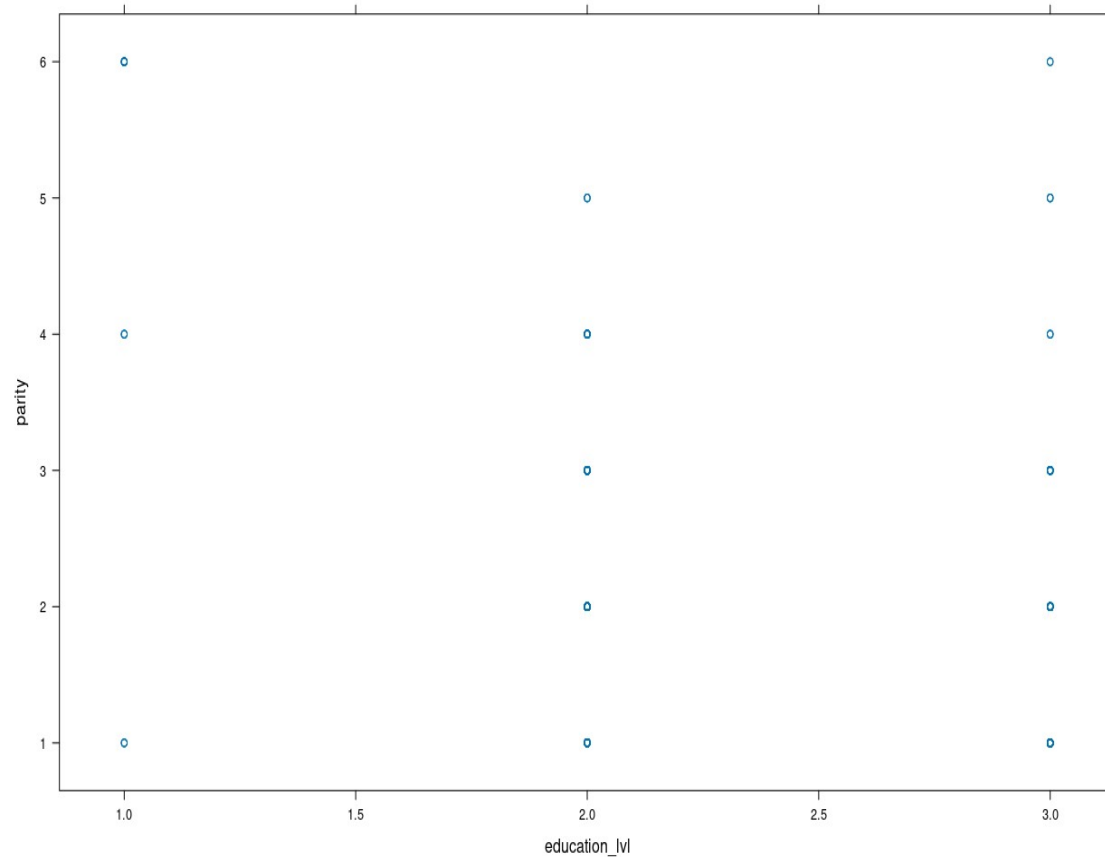
What is the strength of relationship between education level and parity?

Example

- Sample size: 248
- Variables:
 - Education years: 0-5 years = 0, 6-11 years = 1, 12+ years = 2
 - Parity

Scatter Plot

Parity vs Education Level



Results

Kendall's rank correlation tau

```
data: parity and education_lvl
z = -3.3294, p-value = 0.0008703
alternative hypothesis: true tau is not equal to 0
sample estimates:
  tau
-0.191128
```

P-value

tau coefficient = -0.19

→ Significant, small negative correlation

Strength of correlation	<i>r</i> coefficient
None or very small	0 – < 0.1
Small	0.1 – < 0.3
Medium	0.3 – < 0.5
Large	0.5 – 1

Quiz

- Describe the purpose of checking association by Pearson's correlation
- Describe the purpose of checking association by Spearman's rank correlation
- Describe the purpose of checking association by Kendall's rank correlation

Quiz

Table 6. Correlation between perceived stress and child's and caregiver's characteristics

Variable	Pearson's correlation coefficient (<i>r</i>)	<i>P</i> -value
Child characteristics		
Age	0.057	0.431
Time since diagnosis	-0.057	0.456
Care dependency	0.037 ^a	0.615
Caregiver characteristics		
Age	-0.126	0.082
Number of disabled children	0.073 ^a	0.320
Number of children	-0.219	0.002 ^b
Monthly household income	0.015 ^a	0.842
Financial support received	-0.089 ^a	0.255

^a Spearman's correlation coefficient (*r_s*)

^b Significant at 0.05 level (2-tailed)

Isa, S. N. I., Ishak, I., Ab Rahman, A., Saat, N. Z. M., Din, N. C., Lubis, S. H., & Ismail, M. F. M. (2017). Perceived stress and coping styles among Malay caregivers of children with learning disabilities in Kelantan. *The Malaysian journal of medical sciences: MJMS*, 24(1), 81.

Quiz

Correlation between the IDAF-4C and IDAF-S Modules

Spearman's rank correlation yielded r values ranging from 0.278 to 0.566. Item Q3F (*Not knowing what the dentist is going to do*) had the highest correlation with the IDAF-4C module, while item Q3G (*The cost of dental treatment*) had the lowest correlation with the IDAF-4C module. All correlations were significant at the $P < 0.001$ level. The results are shown in [Table 3](#).

Table 3

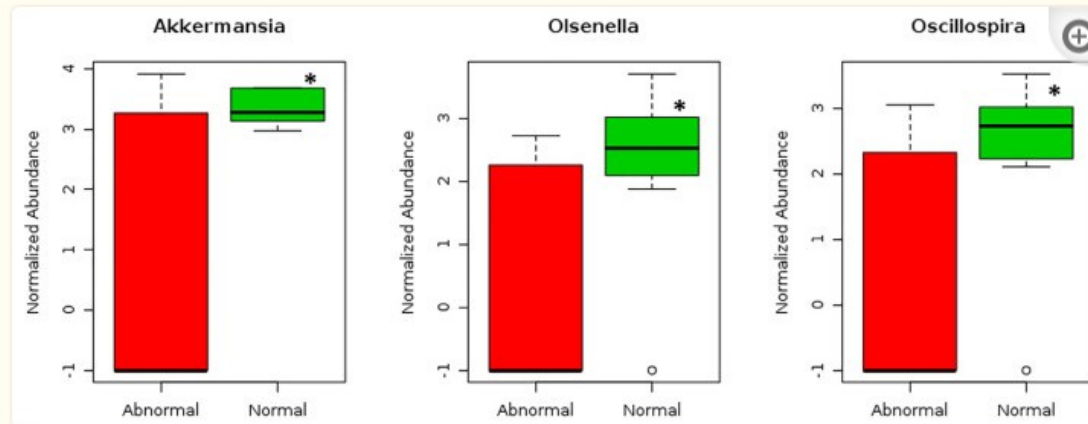
Correlation between the IDAF-4C and each item in the IDAF-S module ($n = 320$)

Item in IDAF-S Module	Spearman's rank correlation, r
Q3A	0.386
Q3B	0.437
Q3C	0.397
Q3D	0.407
Q3E	0.547
Q3F	0.561
Q3G	0.278
Q3H	0.496
Q3I	0.380
Q3J	0.393

[Open in a separate window](#)

Baharuddin, I. H., Arifin, W. N., Kueh, Y. C., & Abd Rahman, Normastura (2018). Adaptation and validation of the Malay version of the Index of Dental Anxiety and Fear (IDAF-4C+) for Malaysian secondary school children. *The Malaysian journal of medical sciences: MJMS*, 25(3), 111.

Quiz



[Figure 5](#)

The normalised abundance of *Akkermansia*, *Olsenella* and *Oscillospira* are significantly higher in the normal BMI group as compared to the abnormal BMI group during pregnancy. *Independent samples *t*-test ($P < 0.05$, FDR = 0.05)

Correlation between the identified key genera-level relative abundances and BMI was further investigated. Interestingly, a negative correlation was found between BMI and *Akkermansia* ($P < 0.05$; tau = -0.413; Kendall rank correlation testing), *Olsenella* ($P < 0.05$; tau = -0.601; Kendall rank correlation testing) and *Oscillospira* ($P < 0.05$; tau = -0.593; Kendall rank correlation testing).

Abdullah, B., Idorus, M. Y., Daud, S., Aazmi, S., Pillai, T. K., & Zain, Z. M. (2023). Gut Microbiota Composition in the First and Third Trimester of Pregnancy among Malay Women is Associated with Body Mass Index: A Pilot Study. *The Malaysian journal of medical sciences: MJMS*, 30(1), 116.

Thank You