Correlation and Chi^2

Md. Abdullah Al Mamun
2581657

Overview: Correlation

- What is correlation
- Types of Correlation
- Analysis of Correlation Coefficient and its possible values and interpretation
- Methods of studying Correlation coefficient (Scatter plot, Pearson, Spearman)
- Exercise: Hypothesis testing

Overview: Chi²

- Introduction to Chi²
- Characteristics/assumptions for Chi²
- Limitations of Chi²
- Applications of Chi² test
- Exercise of Chi² : Hypothesis Test

What is Correlation?

- Correlation means the **direction** and **strength** of **linear relationship** (Chen and Popovich, 2002).
- A **quantitative measure**
- The analysis or measurement of Correlation is called **Correlation Coefficient (r or ρ)**

Types of Correlation

Direction	Variable	Linearity
-Positive -Negative -No -Perfect	-Simple - Multiple - Partial	-Linear - Non-linear

Possible Values of correlation coefficient (r)

- $-1 \leq r \leq 1$
- r = 1 : Perfect Positive Linear Relationship
- r = -1 : Perfect Negative Linear Relationship
- r = 0 : No Linear Relationship

(Ganti, 2019)

Interpreting correlation coefficient (r)

- r > 0.7 or r < -0.7 : **Strong Correlation**
- If value of r is between 0.3 and 0.7 or -0.3 and -0.7: **Moderate Correlation**
- If value of r is between 0 and 0.3 or 0 to -0.3 : **Weak Correlation**

https://www.slideshare.net/Sudhakar8086/correlation-1?from_action=save

Facts about Correlation

- Correlations provide evidence of **association**, not causation
- It is strongly affected by outliers (calculated based on mean and SD)

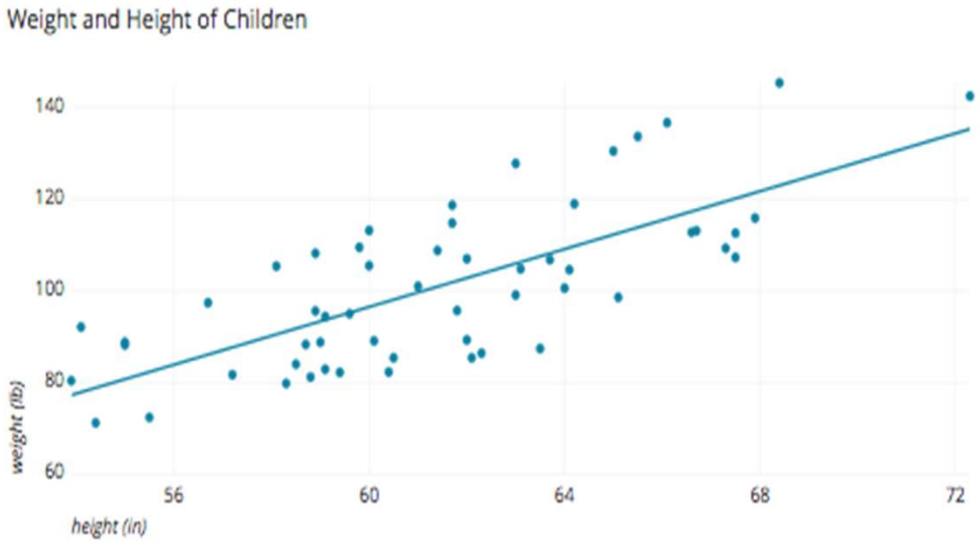
Methods of studying "r"

- Scatter Plot/Diagram Method
- Karl Pearson Coefficient Correlation Method
- Spearman's Rank Correlation Method

https://www.slideshare.net/Sudhakar8086/correlation-1?from_action=save

Scatter Diagram

- Graphical presentation of relationship between two quantitative variables
- Two variables of same individual
- Data of two variables of an individual appears as a point on the graph



(Mandrila and Balentyne, Undated) http://www.liberaldictionary.com/scatter-plot/

Pearson Coefficient Correlation Method

- Both variables are matric
- Normally distributed
- Linear relationships
- Not applicable for curve relationships between variables, no matter how strong the relationship is.

https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/

Spearman's Rank Correlation Method

- No assumptions about the distribution of the data
- When the data are measured on Scale or at least in Ordinal form
- Also called as Nonparametric Correlation

Hypothesis Test for Correlation

- $H_0 : \rho = 0$
- $H_1 : \rho \neq 0$; or $\rho > 0$; or $\rho < 0$
- Exercise: Correlation between experience and sales volume

https://tinyurl.com/uw3nwl9

How to interpret the exercise result?

- **Comment on "r"**: The value of correlation coefficient, r=0.886, which implies that there is a strong positive linear association between the variables, years of sales experience and annual sales volume.
- **Comment on Significance**: Since p-value is 0.001 (which is less than 0.01), we may reject the null hypothesis at 1% level of significance and conclude that ρ (rho) is not equal to 0.

More Exercise (Spearman's Rank)

- Exercise: Correlation between students' marks in mathematics and statistics

https://tinyurl.com/veajq6c

How do you interpret the result?

Overview: Chi²

- Introduction to Chi²
- Characteristics/assumptions for Chi²
- Limitations of Chi²
- Applications of Chi² test
- Exercise of Chi² : Hypothesis Test

Introduction to Chi²

- Chi² (χ^2) is a nonparametric test not based on distribution (normal) of any variable
- Chi² : Difference between what is observed and what is expected according to a assumed hypothesis
- Based on frequencies (not on parameters)
- Purpose is to test hypothesis, not for estimations

https://www.slideshare.net/parth241989/chi-square-test-16093013

Assumptions for Chi²

- Applicable when variables are categorical (ordinal or nominal), with at least two categories for each
- When sample size is small, lowest expected frequency should be at least 5.
- Independence of observations
- Random sampling

Limitations for Chi^2

- Can only be used when the measurements are categorical
- Only used when parametric test is not exits
- Only test hypothesis, no estimation of parameters

Application of Chi² Test

- **Test of independence of attributes**

- To see association of two attributes
- For example: new medicine to control fever
- If χ^2 (calculated)> χ^2 (tabulated) with certain level of significance for given df, then Ho is rejected, otherwise accepted.
- Ho rejected means, there is association between two attributes (new medicine is effective to control fever)

https://www.slideshare.net/parth241989/chi-square-test-16093013

Exercise of Chi² : Hypothesis Test

- Examine the relationship between education and perception of life.

https://tinyurl.com/yx4dppg2

- Can you reject the null hypothesis that education and perception of life are independent?

Table of Chi-square statistics

Exercise of Chi² : Hypothesis Test (cont…)

From the same data set

- Can you reject the null hypothesis that amount of internet use and perception of life are independent? Explain

References

- Chen, P. Y., & Popovich, P. M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks, CA: Sage Publications
- Ganti, A. (2019, November 14). Correlation Coefficient. Retrieved from https://www.investopedia.com/terms/c/correlationcoefficient.asp
- Mandrila, d. and Balentyne, P. (Undated). Scatterplots and Correlation. Retrieved from https://www.westga.edu/academics/research/vrc/assets/docs/scatterplots_and_correlation_notes.pdf
- https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/
- https://www.slideshare.net/Sudhakar8086/correlation-1?from_action=save
- https://www.slideshare.net/parth241989/chi-square-test-16093013
- http://www.liberaldictionary.com/scatter-plot/