Benefits and Risks of Correlations in Decision Making Using R for Analysis

Chihiro Sato (400924769)

Fresenius University of Applied Science

Data Science and Data Analytics (WS 2025/26)

Prof. Dr. Stephan Huber

2025-12-10

Author Note

Correspondence concerning this article should be addressed to Chihiro Sato (400924769), Email: sato.chihiro@stud.hs-fresenius.de

Abstract

This handout examines the role of correlation analysis in business decision-making and research, with a particular focus on implementation using R. While correlations can help identify relationships between variables and support predictive insights, they may also lead to misleading conclusions due to various factors such as the distinction between correlation and causation. Using two practical examples and corresponding datasets, I demonstrate how correlation analysis can inform business decisions and how confounding factors can distort interpretations. Additionally, the paper introduces key R functions for computing correlations and visualizing results, aiming to provide a practical guide for analysts and decision-makers.

Keywords: correlation analysis, business decision-making, cor()

Benefits and Risks of Correlations in Decision Making Using R for Analysis

Word count: 983

1 Correlation

1.1 What Is Correlation?

Correlation is a statistical measure that indicates how strongly and in which way two or more variables are related. The correlation coefficient demonstrates the strength and direction of a relationship, ranging from -1 to +1 (see JMPStatisticalDiscovery (2015)):

- **+1:** Perfect positive correlation (as one variable increases, the other increases proportionally)
- **0:** No correlation (no linear relationship between the variables)
- **-1:** Perfect negative correlation (as one variable increases, the other decreases proportionally)

There are three major types of correlation coefficients: Pearson, Spearman, and Kendall, each with distinct characteristics. Pearson correlation is the most commonly used statistic to measure the degree of relationship between linearly related variables. If (a) the variables are continuous, (b) the relationship between the variables is approximately linear, and (c) there are no extreme outliers, valid results can be ensured using Pearson correlation. If the relationship between variables is non-linear, Spearman correlation can be used to determine the strength and direction of monotonic relationships. It can be applied to either continuous or ordinal data and is relatively robust to outliers. Kendall's correlation is a non-parametric measure suitable for exploring relationships between two ordinal variables (see Brideau (2025) and Bishara and Hittner (2012)).

Table 1 below illustrates key differences between Pearson and Spearman correlations:

1.2 Role of Correlations in decision making

When making decisions, humans operate with limited cognitive capacity, incomplete information, and constant uncertainty (Simon (1979)). In business or research contexts, correlation can help identify which factors may influence outcomes. A strong correlation indicates a significant relationship, providing valuable guidance for decision-making.

1.3 Benefits of Using Correlations

Using correlations in decision making gives you many advantages.Below is several examples:

Table 1

Comparison between Pearson and Spearman

Feature	Pearson	Spearman
Relationship	Linear	Monotonic (linear or not)
Data type	Continuous	Continuous or ordinal
Outliers	Sensitive	Robust
Normality	Assumes approx normal	No strict assumption

- **Identifying relationships:** Correlation analysis reveals how two variables are interconnected.
- **Prediction and forecasting:** Based on the relationship identified through past correlations, you can make predictions and correlations about future outcomes.
- **Strategy optimization:** Analyzing historical data allows you to understand which initiatives have been most effective and enables you to formulate strategies accordingly.
- **Resource allocation:** Understanding which variables have the greatest impact allows you to allocate resources more efficiently. past effective and efficient practices, you can focus on impactful variables.
- **Risk management:** Recognizing correlations among variables helps you to prepare for and mitigate potential risks.

see Rinehart (2023)

1.4 Risks of Using Correlations

While correlations can simplify understanding relationships between variables, they also come with certain risks.:

- **Correlation is not causation:** Just because two variables are correlated, it does not mean one causes the other; Misinterpretation can lead to wrong decisions.
- **External factors:** It is necessary to consider external influences (or seasonal trends) that might impact the observed relationship between variables.
- **Data quality & quantity:** Meaningful correlations require sufficient, high-quality data; otherwise, conclusions may be inaccurate.

- **False correlations:** Random data which actually have no relationship may accidentally look meaningful and lead you to wrong conclusions.
- **Non-linear relationships:** You may not be able to identify correlation when relationships are not linear.

see Rinehart (2023)

1.5 Benefits and Risks of Using R

While correlations offer significant advantages and some potential drawbacks in decision-making, leveraging R for correlation analysis can maximize their potential:

Benefits of using R:

- **R has a powerful statistical and data visualization capabilities:** R is designed for statistical analysis and data visualization. With advanced packages such as corrr, ggplot2, and caret, it simplifies the process of calculating, visualizing, and interpreting correlations.
- **Extensive package ecosystem:** R offers access to over 20,000 packages, ensuring that whatever data analysis task you need, there's likely an R package available to support it.
- **Transparency and reproducibility:** R maintains a clear code history, making it easy to track which analysis methods were used and allowing immediate reproduction of results for validation and collaboration.
- **Integration with predictive modeling**

Risks of using R:

- **Correlation ≠ causation:** While R can effectively identify correlations between variables, it does not establish causality. Misinterpreting correlation as causation can lead to flawed decisions.
- **Complexity for non-experts:** R requires a solid understanding of statistics and programming. Misuse of functions or misinterpretation of outputs can result in incorrect conclusions.
- **False sense of precision**

see Githiora (2025)

Table 2 shows functions useful for visualization of correlations in R

Table 2

Useful functions for visualization

Package	Function	Use
graphics	Pairs plot	Scatterplot matrix
`GGally`	GGally	Scatterplot matrix
corrplot	corrplot	Heatmap
ggraph	ggraph	Network graph of correlations
graphics	barplot	Bar plot of correlation with target variable
ggcorrplot	ggcorrplot	Correlogram, a polished alternative to corrplot
plotly, heatmaply	plotly, heatmaply	Interactive visualization

2 Example Case

2.1 Example 1: Benefit

We will examine the benefits and risks of using correlations in decision-making through example cases. For illustrating the benefits, we will use the “marketing” dataset.

```
data("marketing")
head(marketing)
```

```
  youtube facebook newspaper sales
1  276.12    45.36     83.04 26.52
2   53.40    47.16     54.12 12.48
3   20.64    55.08     83.16 11.16
4  181.80    49.56     70.20 22.20
5  216.96    12.96     70.08 15.48
6   10.44    58.68     90.00  8.64
```

Below are example functions for calculating correlations based on the Pearson linear relationship.

```r
cor(marketing$youtube, marketing$sales)
```

```
[1] 0.7822244
```

```r
cor.test(marketing$youtube, marketing$sales)
```

```
    Pearson's product-moment correlation


data:  marketing$youtube and marketing$sales
t = 17.668, df = 198, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7218201 0.8308014
sample estimates:
      cor
0.7822244
```

```r
M <- cor(marketing, use = "pairwise.complete.obs", method = "pearson")
round(M, 3)
```

```
          youtube facebook newspaper sales
youtube     1.000    0.055     0.057 0.782
facebook    0.055    1.000     0.354 0.576
newspaper   0.057    0.354     1.000 0.228
sales       0.782    0.576     0.228 1.000
```

For Spearman correlation coefficient, you can simply select Spearman as below:

```r
M_s <- cor(marketing, use = "pairwise.complete.obs", method = "spearman")
round(M_s, 3)
```

```
          youtube facebook newspaper sales
youtube     1.000    0.056     0.051 0.801
facebook    0.056    1.000     0.317 0.554
newspaper   0.051    0.317     1.000 0.195
sales       0.801    0.554     0.195 1.000
```

2.2 Example 2: Risk

We will demonstrate an example where an external factor influences two variables, creating the appearance of a correlation between them.

```
data("airquality")
aq <- na.omit(airquality[, c("Ozone", "Solar.R", "Temp")])
```

To remove the impact of an external factor (i.e., temperature), the following function can be used:

```
library(ppcor)
pcor.test(x = aq$Ozone, y = aq$Solar.R, z = aq["Temp"])
```

```
   estimate    p.value statistic   n gp  Method
1 0.2089543 0.02847063  2.220534 111  1 pearson
```

```
pc <- pcor.test(x = aq$Ozone, y = aq$Solar.R, z = aq["Temp"])
```

3 References

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of pearson, spearman, transformation, and resampling approaches. *Psychological Methods*, *17*(3), 399.

Brideau, L. (2025). *Correlation: Pearson, spearman, and kendalls tau*. https://library.virginia.edu/data/articles/correlation-pearson-spearman-and-kendalls-tau

Githiora, Y. (2025). *Advantages and disadvantages of using r for data analysis*. https://africaresearchub.com/advantages-and-disadvantages-of-using-r-for-data-analysis/

JMPStatisticalDiscovery. (2015). *Correlation*. https://www.jmp.com/en/statistics-knowledge-portal/what-is-correlation#:~:text=What%20is%20correlation?,two%20measures:%20elevation%20and%20temperature

Rinehart, R. (2023). *How to use correlation in business decision making*. https://www.servcorp.co.uk/en/blog/business-networking/how-to-use-correlation-in-business-decision-making/

Simon, H. A. (1979). Rational decision making in business organizations. *The American Economic Review*, *69*(4), 493–513.

Appendix

Affidavit

I hereby affirm that this submitted paper was authored unaided and solely by me. Additionally, no other sources than those in the reference list were used. Parts of this paper, including tables and figures, that have been taken either verbatim or analogously from other works have in each case been properly cited with regard to their origin and authorship. This paper either in parts or in its entirety, be it in the same or similar form, has not been submitted to any other examination board and has not been published. I acknowledge that the university may use plagiarism detection software to check my thesis. I agree to cooperate with any investigation of suspected plagiarism and to provide any additional information or evidence requested by the university.