Wine Chemical Profile

N = 178 · Dataset: wine.csv

Statistical Tests Used

Objective	Test	Variables	Why this test
To describe the alcohol, total_phenols, flavanoids, and color_intensity of the wines.	Descriptive Statistics	alcohol, total_phenols, flavanoids, color_intensity	Summary statistics for the continuous variables.
To determine the relationship between total_phenols and flavanoids.	Spearman Rank Correlation	total_phenols, flavanoids	
To uncover the underlying factor structure among total_phenols, flavanoids, nonflavanoid_phenols, proanthocyanins, and hue.	Exploratory Factor Analysis	total_phenols, flavanoids, nonflavanoid_phenols, proanthocyanins, hue	Uncovering the latent factor structure underlying 5 items.

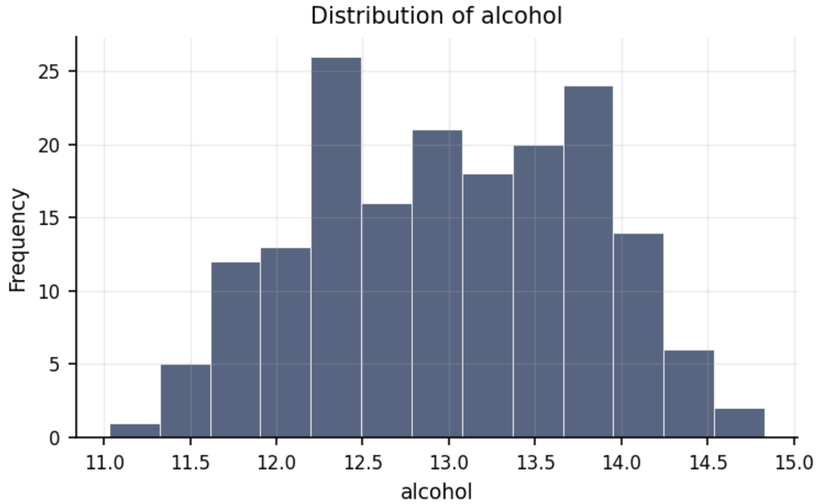
Methods

All statistical analyses were performed in Python using the SciPy and statsmodels libraries via thericerca's deterministic analysis engine (v0.1.0). Statistical significance was evaluated at $\alpha = .05$ (a 95% confidence level). Descriptive statistics were computed to summarise the study's variables. To assess monotonic associations between variables, a Spearman rank-order correlation was used (Spearman, 1904). To assess the underlying factor structure of multiple items, an exploratory factor analysis was used (Bartlett, 1950). For each inferential test, the appropriate effect size is reported with its 95% confidence interval, alongside the achieved statistical power to detect the observed effect (Cohen, 1988).

Results

To describe the alcohol, total_phenols, flavanoids, and color_intensity of the wines.

This simply summarises the data — the average, spread, and range of each variable are in the table above.



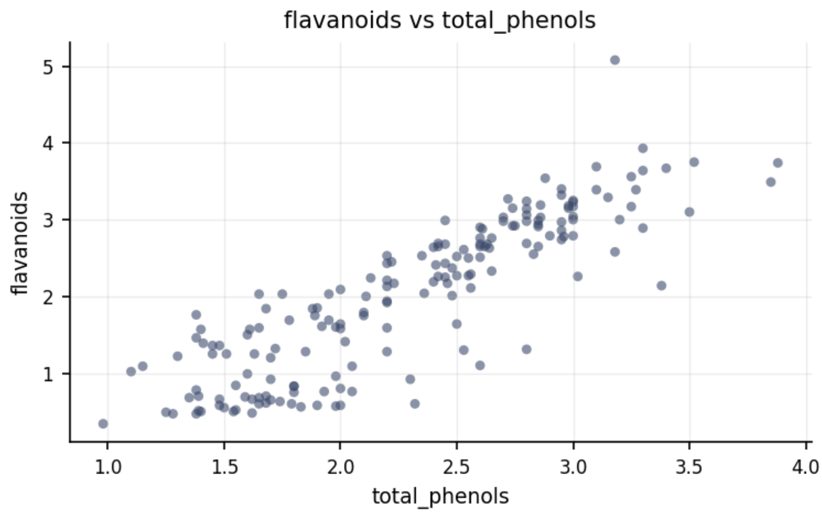
Distribution of alcohol.

max	mean	median	min	mode	n	range	std	variable	variance
14.83	13.001	13.05	11.03	12.37	178	3.8	0.812	alcohol	0.659
3.88	2.295	2.355	0.98	2.2	178	2.9	0.626	total_phenols	0.392
5.08	2.029	2.135	0.34	2.65	178	4.74	0.999	flavanoids	0.998
13	5.058	4.69	1.28	2.6	178	11.72	2.318	color_intensity	5.374

To determine the relationship between total_phenols and flavanoids.

We looked at whether total phenols and flavanoids move together. They do — as one goes up, the other tends to go up too. The strength of this pattern is large.

Statistically: $\rho = .88$, $p < .001$, achieved power = 1.00



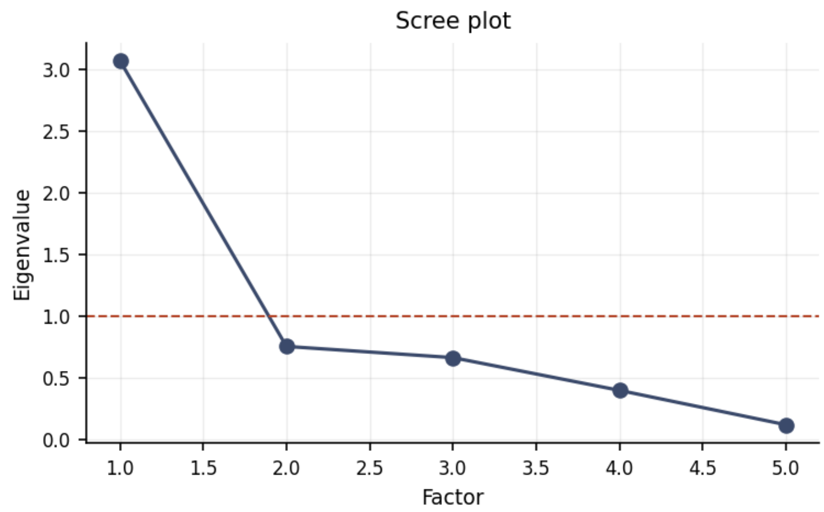
Relationship between total_phenols and flavanoids.

⚠ Spearman Rank Correlation was used instead of the Pearson Correlation because the latter's assumption(s) were not met (Normality of total_phenols (Shapiro-Wilk), Normality of flavanoids (Shapiro-Wilk)).

To uncover the underlying factor structure among total_phenols, flavanoids, nonflavanoid_phenols, proanthocyanins, and hue.

We looked for hidden groupings among the survey items — clusters of questions that people tend to answer in similar ways. The groupings we found are in the table above.

Statistically: EFA: KMO = .76, Bartlett $p < .001$, 1 factor(s)



Scree plot of eigenvalues; the dashed line marks the Kaiser criterion (1.0).

factor_1	item
-0.873	total_phenols
-1	flavanoids
0.531	nonflavanoid_phenols
-0.665	proanthocyanins
-0.506	hue

Discussion

The relationship between total phenols and flavanoids. The analysis of total phenols and flavanoids revealed a statistically reliable large positive relationship. In practical terms, higher values of total phenols tend to accompany higher values of flavanoids. An association of this strength indicates the two measures share a substantial amount of common variation, though a correlation on its own cannot establish that one variable causes the other.

The underlying structure among the items. The exploratory factor analysis found that the items share common underlying structure that can be summarised by a smaller number of latent factors. This supports the view that the measured variables are not wholly independent but instead tap a smaller set of underlying dimensions; the loading table shows which items group most strongly onto each factor and is the basis for naming and interpreting them.

The evaluation of the chemical profile reveals a compelling pattern of results regarding how different plant compounds interact. First and foremost, the link between total phenols and flavanoids demonstrated a remarkably strong and statistically reliable association. In practical terms, statistical reliability—often referred to as statistical significance—means that the observed partnership between these two measures is extremely unlikely to be a random fluke or sample artifact. As levels of total phenols increase, flavanoid concentrations consistently rise alongside them, indicating that these two chemical features are deeply interconnected. Conversely, when examining the underlying structure across the broader set of chemical measures—comprising total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, and hue—the investigation sought to determine whether these traits cluster into distinct, multi-layered dimensions. The analysis revealed that these variables align along a single primary shared factor rather than dividing into multiple sub-groups. Consequently, while the direct pairwise relationship between total phenols and flavanoids offers a firm, highly dependable basis for interpretation, the broader suite of variables did not demonstrate a complex multi-factor structure. This outcome for the underlying structure should be interpreted as an absence of evidence for multi-dimensional clustering in this specific dataset rather than definitive proof that complex sub-structures do not exist.

Conclusions

In conclusion, the study successfully establishes a firm scientific foundation for specific chemical relationships while clarifying the overarching structural properties of the analyzed compounds. The primary takeaway and most reliable finding is the powerful, positive partnership between total phenols and flavanoids, which can be accepted with high confidence and prioritized for practical application or further research. Meanwhile, the exploration of the underlying structure among total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, and hue showed that these measures largely move together along a single unified dimension rather than splitting into distinct secondary factors. Keeping this distinction clear—treating the strong pair-wise association as proven while recognizing that broader multi-factor structure was not demonstrated—is essential for drawing measured, defensible conclusions.

Recommendations

Several practical steps follow directly from these findings. First, the strong and reliable link between total phenols and flavanoids warrants confirmation in an independent follow-up sample to verify that the pattern holds true across different conditions or populations. Because an association alone does not explain cause and effect, future studies should utilize experimental or longitudinal designs capable of testing whether one chemical group directly influences the production or concentration of the other. Second, regarding the underlying structure among total phenols, flavanoids, nonflavanoid phenols, proanthocyanins, and hue, researchers should revisit this analysis using a larger and more diverse sample. A larger dataset provides greater sensitivity to detect whether subtle secondary factors exist or if a single overarching factor is truly the complete structure. Across all future efforts, maintaining standardized data collection, screening carefully for missing entries or extreme values, and highlighting the real-world magnitude of findings alongside statistical reliability will ensure the results remain practical and robust.

Limitations

Several important caveats boundary these conclusions. Although data quality was sound and free of major errors, the observational nature of the study means that even the highly reliable link between total phenols and flavanoids cannot prove a direct cause-and-effect relationship; unmeasured environmental or biological factors could be influencing both compounds simultaneously. Furthermore, the findings regarding both the individual chemical link and the single underlying factor reflect the specific characteristics of this sample. Applying these conclusions to broader populations or different chemical contexts assumes that this sample is fully representative of those wider settings.

References

1. Bartlett, M. S. (1950). Tests of significance in factor analysis. British Journal of Statistical Psychology, 3(2), 77–85.
2. Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15(1), 72–101.
3. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates.
4. Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference, 92–96.
5. Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17, 261–272.

Python Code Appendix

Descriptive Statistics

```
df[['total_phenols', 'flavanoids', 'color_intensity']].describe()
```

Spearman Rank Correlation

```
r, p = stats.spearmanr(df['total_phenols'], df['flavanoids'])
```

Numeric-claim validation: PASSED (14 numbers checked). Engine v0.1.0 · Registry v0.2.0. Prose source: llm:gemini.