

Wine and Obesity Data Analysis

Adrian T. and Olivia K.

2025-12-09

1. Dataset Overview

1.1 Wine Quality Dataset

- **Collection Year:** 2009
- **Description:** Physiochemical properties for red and white Portuguese “Vinho Verde” wines.
- **Size:** 6497 rows, 12 columns
- **Variables:** fixed_acidity, volatile_acidity, citric_acid, residual_sugar, chlorides, free_sulfur_dioxide, total_sulfur_dioxide, density, ph, sulphates, alcohol, quality, color
- **Study Type:** Observational; each wine sample measured independently.

Sampling Distribution of Density

The following histogram shows the sampling distribution of the **sample mean of the density variable**.

1.2 Obesity Dataset

- **Collection Year:** 2019
- **Description:** Estimation of obesity levels in individuals from Mexico, Peru, and Colombia, based on eating habits and physical condition.

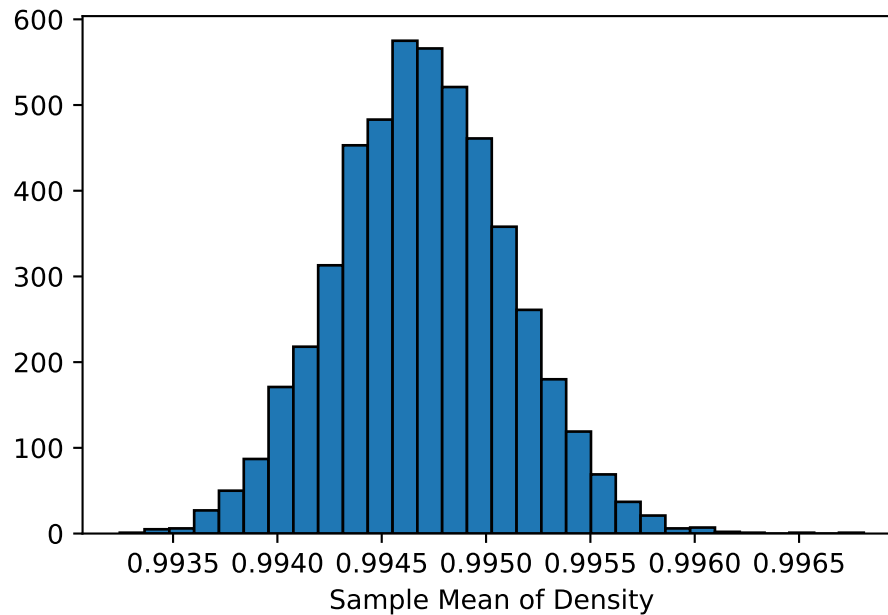


Figure 1: Sampling Distribution of Sample Mean (Density)

- **Size:** 2111 rows, 16 columns
- **Variables:** Gender, Age, Height, Weight, Family_history_with_overweight, FAVC, FCVC, NCP, CAEC, SMOKE, CH20, SCC, FAF, TUE, CALC, MTRANS, NObeyesdad
- **Study Type:** Observational; each individual measured independently.

2. One-Sample T-Test: Wine Alcohol Content

Research Question: Is the average alcohol content of red and white wine greater than 10.5%?

Hypotheses:

- (H_0 : = 10.5)
- (H_a : > 10.5)

Results:

- Sample mean: 10.4918
- t-statistic: -0.5541

- Critical t-value ($\alpha = 0.05$): 1.6451
- One-tailed p-value: 0.7102
- 95% Confidence Interval: (10.4628, 10.5208)

Conclusion: Fail to reject (H_0). There is insufficient evidence that the mean alcohol content exceeds 10.5%.

3. Bootstrap Approach: Wine pH

Bootstrap Distribution

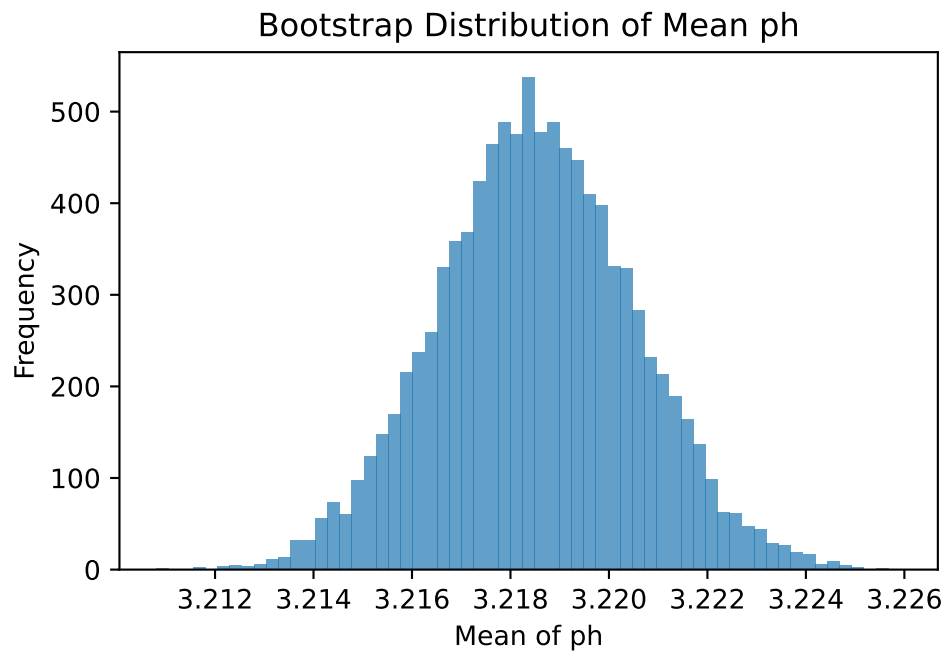


Figure 2: Bootstrap Distribution of Wine pH

QQ Plot for Bootstrap

95% Bootstrap Confidence Interval

(3.214540480221641, 3.2225303986455285)

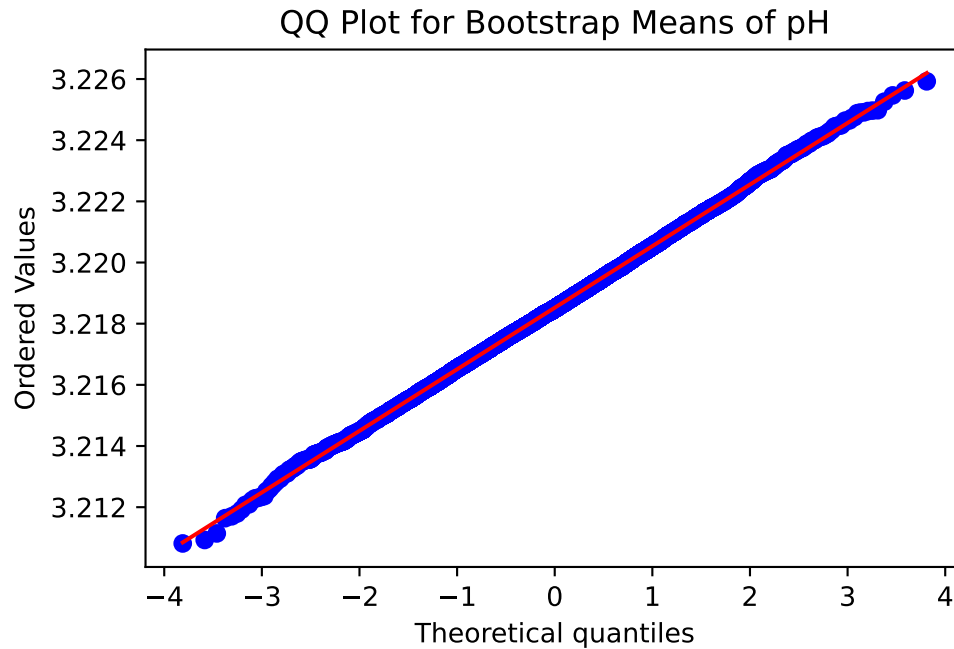


Figure 3: QQ Plot of Bootstrap Means (pH)

Conclusion: We are 95% confident that the true population mean pH lies between 3.215 and 3.222.

4. Analysis of Variance (Wine Dataset)

F Test

ANOVA Table

ANOVA Assumptions Check

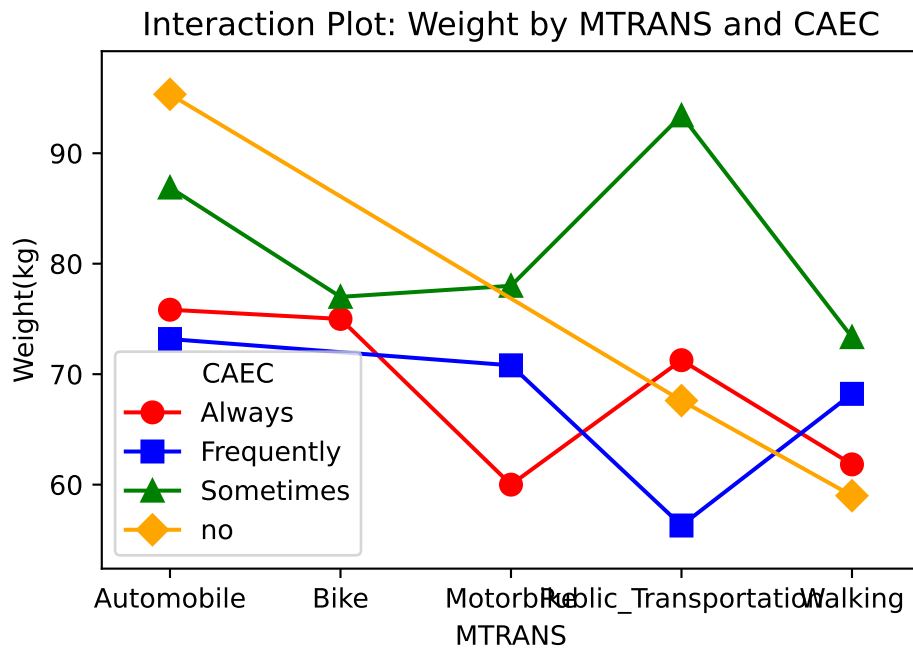
Conclusion

5. Multiple Comparisons (Obesity Dataset)

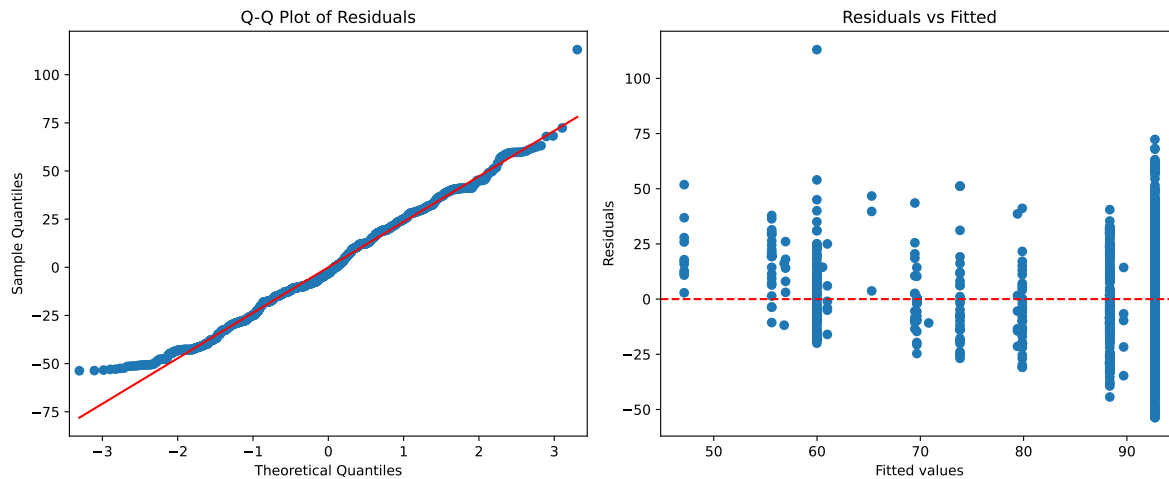
ANOVA Model and Assumptions

- Two-way ANOVA with interaction: $\text{Weight} \sim \text{MTRANS} * \text{CAEC}$.

<Figure size 3600x2400 with 0 Axes>



QQ Plot and Residual Analysis:



- QQ plot shows approximate normality.
- Residual plot shows no clear pattern.
- **Conclusion:** ANOVA assumptions satisfied (normality, independence, equal variance).

F-Tests for Factors

Factor	F-statistic	F-critical	p-value	Conclusion
MTRANS	6.85	2.376	1.78×10^{-4}	Reject H : At least one group mean differs
CAEC	149.69	2.609	1.11×10^{-14}	Reject H : At least one group mean differs

Interpretation: Both MTRANS and CAEC significantly affect Weight; differences between group means are unlikely due to chance.

References

Cortez, Paulo, et al. "Wine Quality." UCI Machine Learning Repository, 2009, <https://doi.org/10.24432/C56S3T>.

"Estimation of Obesity Levels Based On Eating Habits and Physical Condition ." UCI Machine Learning Repository, 2019, <https://doi.org/10.24432/C5H31Z>.

Code Appendix