Wine and Obesity Data Analysis

Adrian T. and Olivia K.
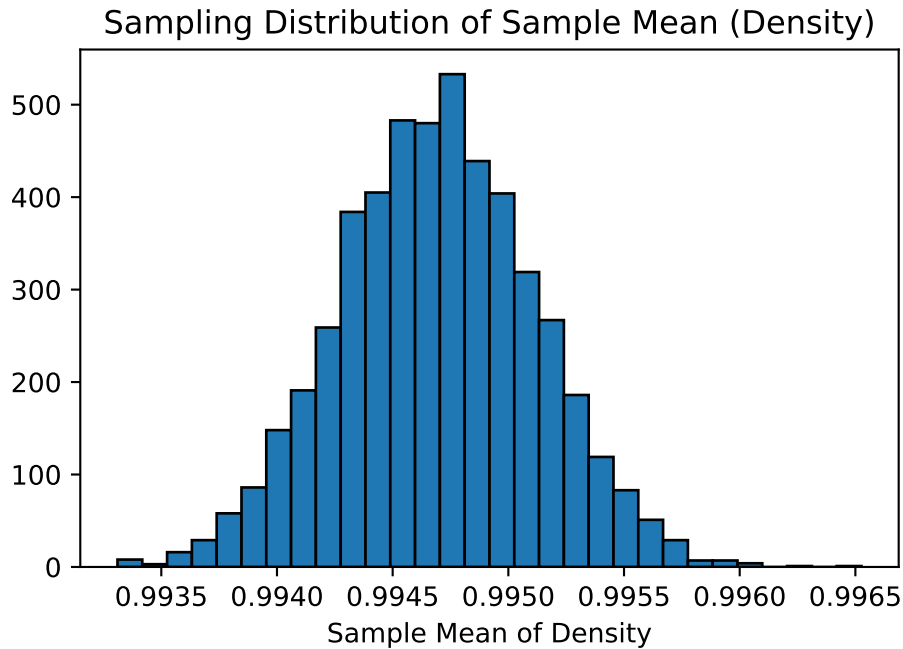
2025-12-09

1. Dataset Overview

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 bootstrap sampling distribution of the **sample mean of the `density` variable**.

Sampling Distribution of Sample Mean (Density)

Obesity Dataset

- **Collection Year:** 2019

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

- **Size:** 2111 rows, 16 columns

- **Variables:** `Gender, Age, Height, Weight, Family_history_with_overweight, FAVC, FCVC, NCP, CAEC, SMOKE, CH2O, 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%?
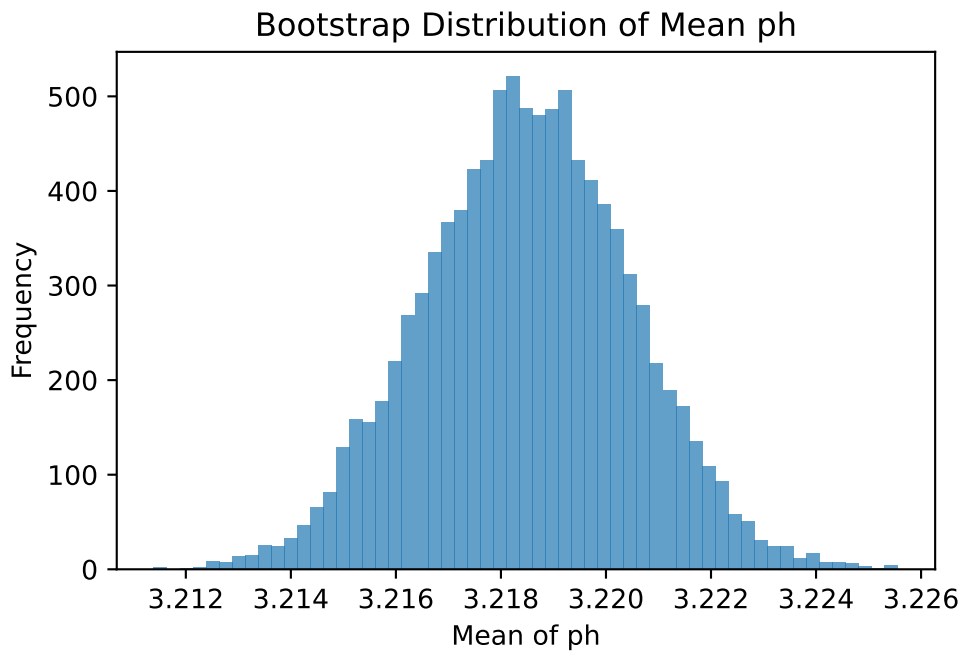
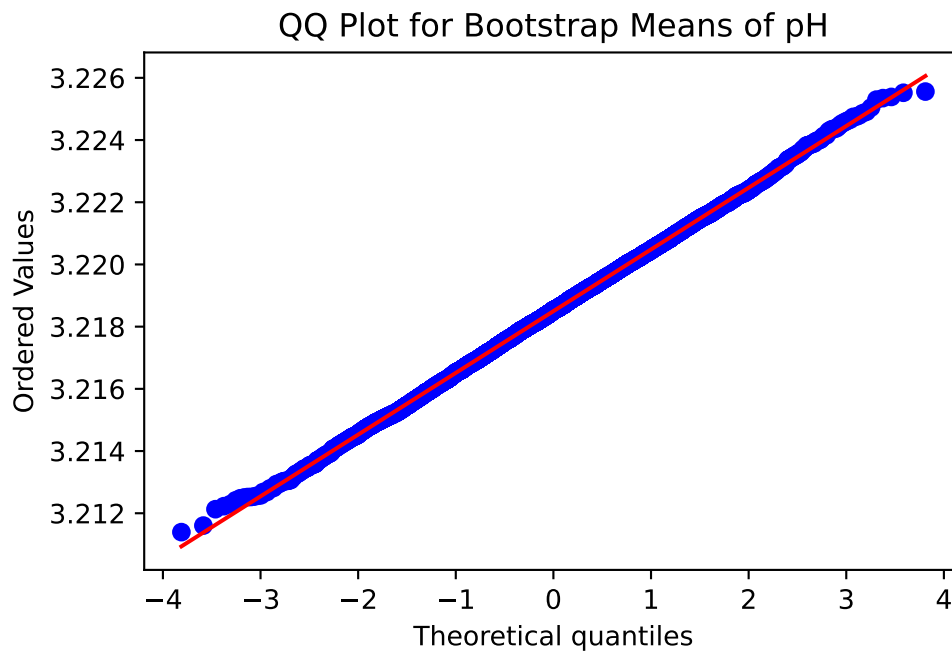**Hypotheses:**
- (H_0: = 10.5)
- (H_a: > 10.5)

Results:
- Sample mean: 10.4918
- t-statistic: -0.5541
- Critical t-value (= 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

Bootstrap Distribution of Mean ph

QQ Plot for Bootstrap

QQ Plot for Bootstrap Means of pH

95% Bootstrap Confidence Interval

(3.2146389872248733, 3.2223041788517777)

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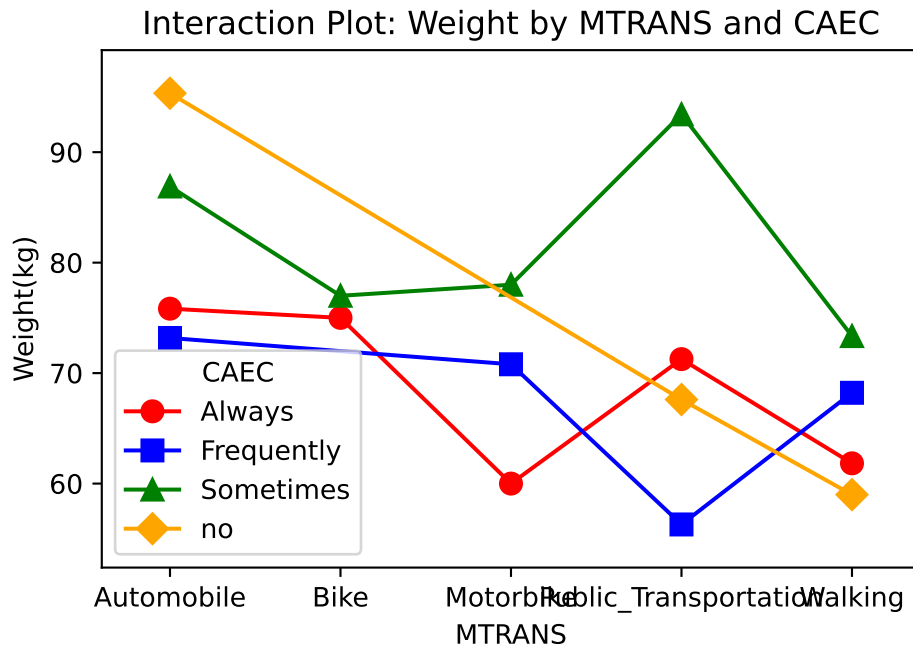
---

4. Analysis of Variance (Wine Dataset)
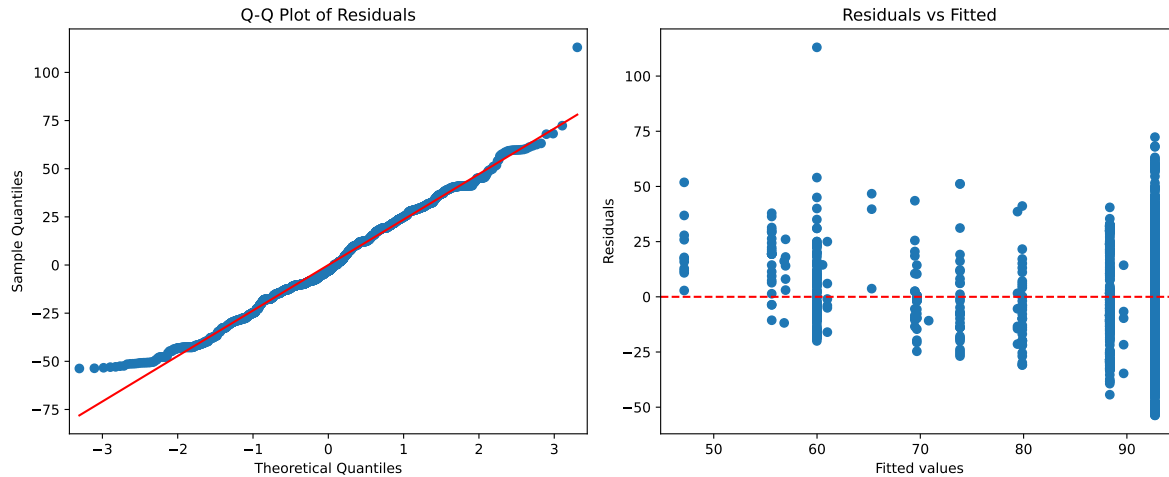
---

5. Multiple Comparisons (Obesity Dataset)

ANOVA Model and Assumptions

- Two-way ANOVA with interaction: `Weight ~ MTRANS * CAEC`.

`<Figure size 3600x2400 with 0 Axes>`

Interaction Plot: Weight by MTRANS and CAEC

- **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	Reject H : At least one group mean differs
CAEC	149.69	2.609	1.11×10^{1}	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

- UCI Wine Quality Dataset

- UCI Obesity Dataset

- Stat319 GitHub Repository