1 A1-10, ANOVA Potato

1.1 Introduction

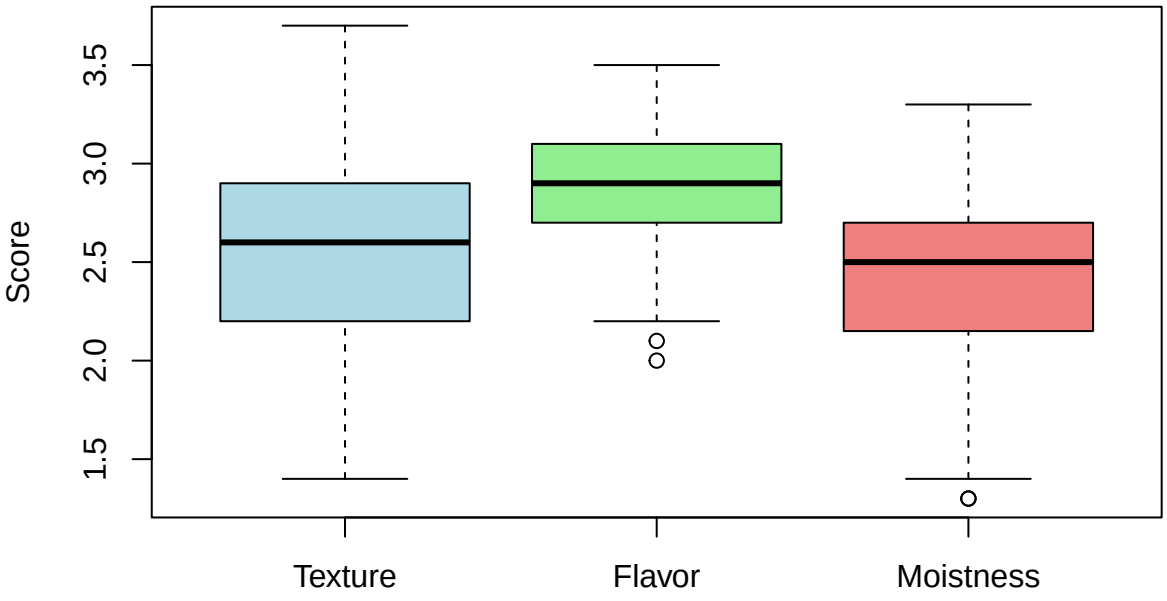
This study investigates the cooking quality of Oregon-grown Russet potatoes. Specifically, it examines how different growing areas, storage conditions, and cooking methods affect the flavor of the potatoes. The `Flavor score` is modeled as a function of:

- **Growing Area**: Southern Oregon vs. Central Oregon
- **Two-week Holding Temperature**: 75°F vs. 40°F
- **Size**: Large vs. Medium
- **Storage Period**: 0, 2, 4, and 6 months
- **Cooking Method**: Boiling, Steaming, Mashing, Baking at 350°F, Baking at 450°F

1.2 Exploratory Data Analysis (EDA)

Table 1: Summary Statistics (Categorical Variables)

Variable	Southern Oregon	Central Oregon	75°F	40°F	Large	Medium	0 months	2 months	4 months	6 months	Boil	Steam	Mash	Bake@350°F	Bake@450°F
Area	80	80													
Temp			80	80											
Size					80	80									
Storage							40	40	40	40					
Cooking											32	32	32	32	32



1.3 Mathematical Model

The general form of our ANOVA model is:

$$
\begin{aligned}
\hat{Y} = \beta_0 \\
&+ \beta_1 \cdot \text{Area}_{\text{Central Oregon}} \\
&+ \beta_2 \cdot \text{Temp}_{40^\circ F} \\
&+ \beta_3 \cdot \text{Size}_{\text{Medium}} \\
&+ \beta_4 \cdot \text{Storage}_{\text{2 months}} \\
&+ \beta_5 \cdot \text{Storage}_{\text{4 months}} \\
&+ \beta_6 \cdot \text{Storage}_{\text{6 months}} \\
&+ \beta_7 \cdot \text{Cooking}_{\text{Steam}} \\
&+ \beta_8 \cdot \text{Cooking}_{\text{Mash}} \\
&+ \beta_9 \cdot \text{Cooking}_{\text{Bake@350}^\circ F} \\
&+ \beta_{10} \cdot \text{Cooking}_{\text{Bake@450}^\circ F} \\
&+ \epsilon
\end{aligned}
$$

where:
- $\hat{Y}$ is the predicted **Flavor Score**
- β_i are the coefficients estimated from data
- $XXXX_i$ are indicator variables (ANOVA encoding), where $X_i = 1$ if the corresponding factor is present and $X_i = 0$ otherwise. The first level of each categorical variable is assumed present by default and thus not listed.
- $\epsilon \sim N(0, \sigma^2)$ represents residual errors.

1.4 Model Assumptions

For the ANOVA model to be valid, the following assumptions must hold:
1. **Independence**: Observations are independent.
2. **Normality**: Residuals follow a normal distribution.
3. **Homoscedasticity**: Residuals have constant variance.
4. **No Multicollinearity**: Predictors are not highly correlated.

1.5 Model Fitting

Table 2: ANOVA Results for Flavor Score

Term	DF	Sum Sq	Mean Sq	F Value	P Value
Area	1	0.53	0.53	10.0	0.00
Temp	1	1.10	1.10	21.0	0.00
Size	1	0.00	0.00	0.0	0.94
Storage	3	2.00	0.67	13.0	0.00
Cooking	4	1.30	0.34	6.5	0.00
Residuals	150	7.70	0.05	NA	NA

1.6 Model Selection

We refine our model using stepwise selection based on Akaike's Information Criterion (AIC). We see that the Size has a very large P value and thus can be removed from the model without compromising it's predictive power.

Table 3: Stepwise Model Summary

term	df	sumsq	meansq	statistic	p.value
Area	1	0.53	0.53	10.26	0
Temp	1	1.09	1.09	21.12	0
Storage	3	2.02	0.67	13.09	0
Cooking	4	1.34	0.34	6.51	0
Residuals	150	7.73	0.05	NA	NA

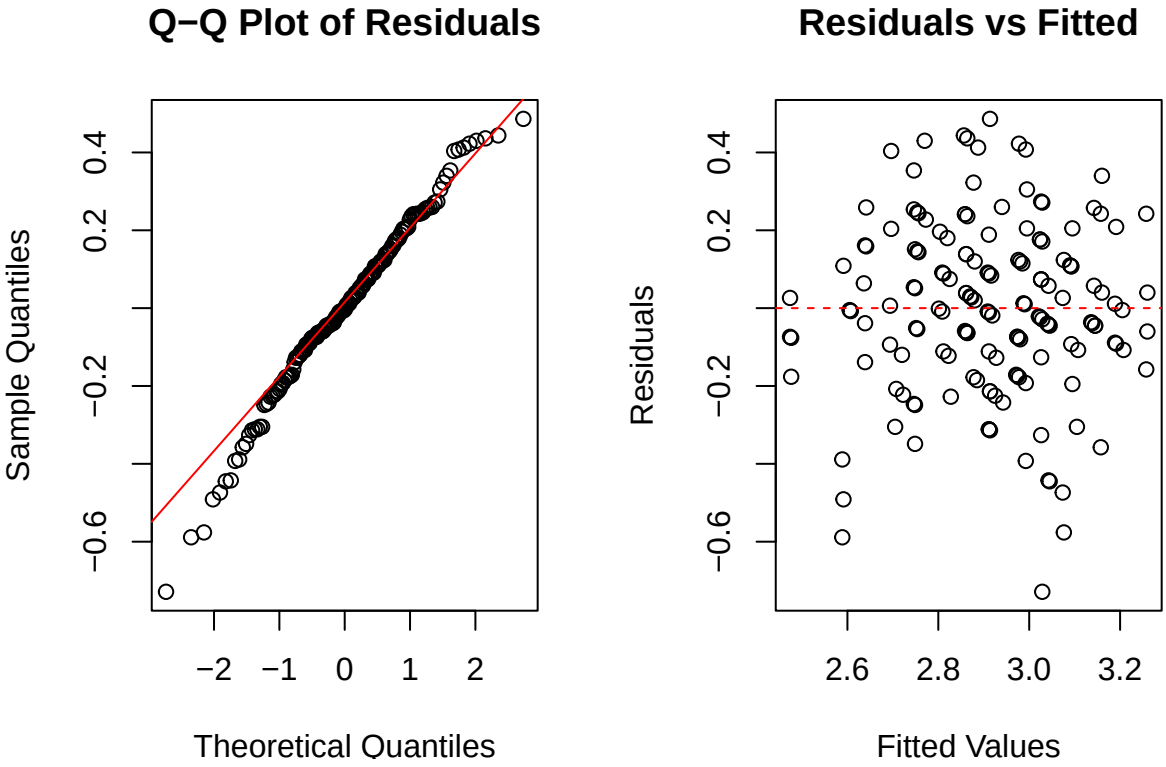
1.7 Estimated Model with Numeric Coefficients

The estimated model with numeric values is:

$$
\begin{aligned}
\hat{Y} &= 2.8 \\
&+ 0.11 \cdot AreaCentralOregon \\
&- 0.17 \cdot Temp40°F \\
&+ 0.0025 \cdot SizeMedium \\
&+ 0.22 \cdot Storage2months \\
&+ 0.27 \cdot Storage4months \\
&+ 0.27 \cdot Storage6months \\
&- 0.13 \cdot CookingSteam \\
&+ 0.031 \cdot CookingMash \\
&+ 0.1 \cdot CookingBake@350°F \\
&- 0.13 \cdot CookingBake@450°F \\
&+ \epsilon, \quad \epsilon \sim \mathcal{N}(0, 0.23^2)
\end{aligned}
$$

1.8 Model Diagnostics

To assess whether the ANOVA assumptions hold, we examine residual diagnostics.



1.9 Conclusion

This study analyzed the factors influencing the **flavor score** of Oregon-grown Russet potatoes using an **ANOVA model**. The key variables considered included **growing area, storage conditions, size and cooking methods**.

1.9.1 Main Findings

1. **Significant Factors**
    - **Holding temperature (75°F vs. 40°F)** had the **largest effect** on flavor, with lower temperatures reducing flavor scores.
    - **Storage duration** influenced flavor, with **2- to 6-month storage improving flavor** compared to fresh potatoes.
    - **Cooking method** significantly affected flavor, with **boiling and high-temperature baking (450°F) resulting in lower scores**, while **baking at 350°F** yielded better results.
    - **Growing area (Southern vs. Central Oregon)** also had a **notable impact**.
2. **Insignificant Factor**
    - **Potato size (Large vs. Medium)** had a very high p-value ($p = 0.94$), indicating **no significant effect on flavor**. It was **removed from the final model** to improve simplicity.

1.9.2 Model Fit & Assumption Verification

- The **residual analysis** (Q-Q plot and residuals vs. fitted values) indicated that the model **reasonably meets normality and homoscedasticity assumptions**.
- The estimated residual standard deviation was **0.232**, confirming a **good model fit**.

1.9.3 Interpretation & Implications

- **Storage at moderate temperatures and baking at 350°F** maximizes potato flavor.
- **Cold storage (40°F) negatively affects flavor**, which may influence **storage and distribution practices**.
- **Cooking methods play a crucial role**; steaming and high-temperature baking should be avoided for better flavor retention.