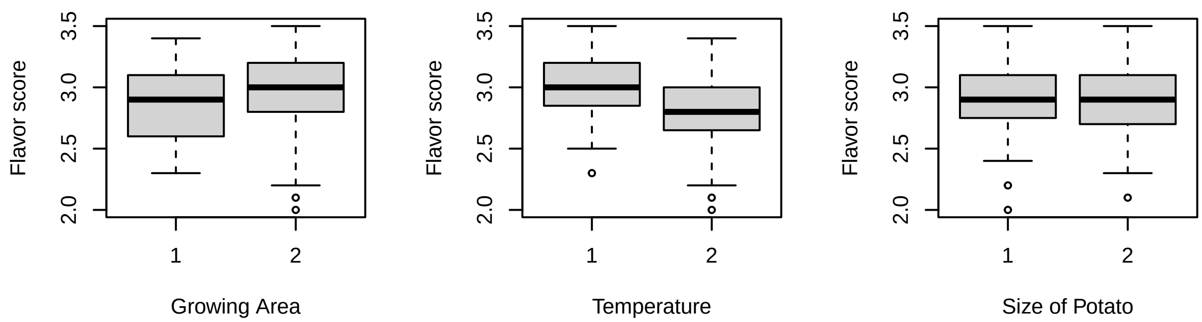
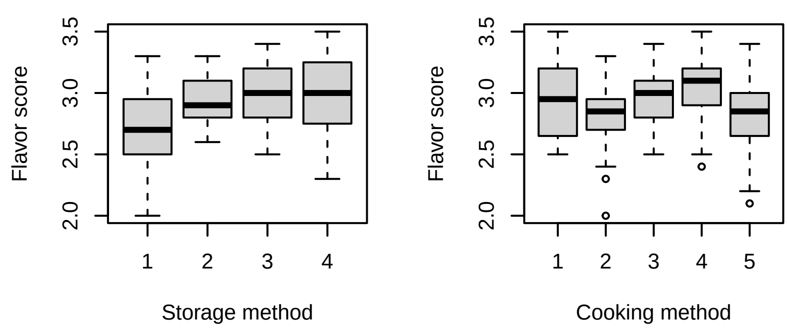
ANOVA

1 Introduction

sdfsdfsdfsdfs

2 Exploratory Data Analysis





asdadsad

3 Model Fitting

3.1 Definition of Terms

In this section we define the terms used in the model fitting as shown in later stages.

Term	Explanation
area_grow	Growing area of potato
temp	Two week holding temperature
size_potato	Size of potato
storage_period	Storage period of potato
cooking_method	Cooking method of potato
flavour	Flavour score of the cooked potato

Table 1: Term explanation of the Model

3.2 Least Squares and the Full Linear Model

Least Squares (LS) is a parameter estimation method in regression analysis that minimizes the sum of the squared residuals. The LS method finds the coefficients that minimize the sum of the squared residuals, which is the difference between the observed values and the predicted values.

In this study, we aim to investigate the relationship between the flavor score and the growing area, two-week holding temperature, size of potato, storage period, and cooking method. The full linear model is as follows:

$$\begin{aligned}\text{flavour} = {} & \beta_0 + \beta_1\text{area\_grow} + \beta_2\text{temp} + \beta_3\text{size\_potato} + \beta_4\text{storage\_period} \\ & + \beta_5\text{cooking\_method} + \beta_6\text{area\_grow} \times \text{temp} + \beta_7\text{area\_grow} \times \text{size\_potato} \\ & + \beta_8\text{area\_grow} \times \text{storage\_period} + \beta_9\text{area\_grow} \times \text{cooking\_method} \\ & + \beta_{10}\text{temp} \times \text{size\_potato} + \beta_{11}\text{temp} \times \text{storage\_period} + \beta_{12}\text{temp} \\ & \times \text{cooking\_method} + \beta_{13}\text{size\_potato} \times \text{storage\_period} + \beta_{14}\text{size\_potato} \\ & \times \text{cooking\_method} + \beta_{15}\text{storage\_period} \times \text{cooking\_method} + \epsilon\end{aligned} \tag{1}$$

where β_0 is the intercept, β_1 to β_{15} are the coefficients of the main effects and interactions, and ϵ is the error term. The full model is fitted using the least squares method to estimate the coefficients of the model.

3.3 Selecting the Effective Linear Model

We apply a full ANOVA model with all the interactions and main effects to find out what are the effective terms. By fitting the full model, we obtain the following ANOVA table:

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
area_grow	1	0.529	0.5290	16.289	9.93e-05 ***
temp	1	1.089	1.0890	33.533	6.40e-08 ***
size_potato	1	0.000	0.0002	0.008	0.9302
storage_period	3	2.024	0.6747	20.777	8.60e-11 ***
cooking_method	4	1.344	0.3359	10.342	3.57e-07 ***
area_grow:temp	1	0.020	0.0202	0.624	0.4314
area_grow:size_potato	1	0.049	0.0490	1.509	0.2219
area_grow:storage_period	3	0.287	0.0955	2.941	0.0362 *
area_grow:cooking_method	4	0.166	0.0415	1.278	0.2829
temp:size_potato	1	0.016	0.0160	0.493	0.4842
temp:storage_period	3	0.937	0.3122	9.612	1.05e-05 ***
temp:cooking_method	4	0.123	0.0309	0.951	0.4376
size_potato:storage_period	3	0.943	0.3144	9.682	9.72e-06 ***
size_potato:cooking_method	4	0.076	0.0190	0.585	0.6741
storage_period:cooking_method	12	1.447	0.1206	3.713	9.90e-05 ***
Residuals	113	3.670	0.0325		

Table 2: ANOVA Results on the Full Model

By observing the ANOVA result and filtering out the insignificant ($p\_value >= 0.05$) terms, we obtain the effective linear model as follows:

$$\begin{aligned}\text{flavour} = \beta_0 &+ \beta_1 \text{area\_grow} + \beta_2 \text{temp} + \beta_3 \text{storage\_period} + \beta_4 \text{cooking\_method} \\ &+ \beta_5(\text{area\_grow} \times \text{storage\_period}) + \beta_6(\text{temp} \times \text{storage\_period}) \\ &+ \beta_7(\text{size\_potato} \times \text{storage\_period}) + \beta_8(\text{storage\_period} \times \text{cooking\_method}) + \epsilon\end{aligned} \tag{2}$$

3.4 Fitting the Effective Model

We apply the least squares method to fit the effective model and estimate the coefficients of the model. Below we show the estimated effective model.

$$
\begin{aligned}
\hat{\text{flavour}} = {} & 2.97 + 0.06 \times \text{area\_grow2} - 0.42 \times \text{temp2} - 0.058 \times \text{storage\_period2} - 0.20 \\
& \times \text{storage\_period3} + 0.21 \times \text{storage\_period4} - 0.24 \times \text{cooking\_method2} - 0.013 \\
& \times \text{cooking\_method3} + 0 \times \text{cooking\_method4} - 0.21 \times \text{cooking\_method5} \\
& - 0.05 \times \text{area\_grow2:storage\_period2} + 0.11 \times \text{area\_grow2:storage\_period3} \\
& + 0.17 \times \text{area\_grow2:storage\_period4} + 0.29 \times \text{temp2:storage\_period2} \\
& + 0.33 \times \text{temp2:storage\_period3} + 0.41 \times \text{temp2:storage\_period4} + 0.04 \\
& \times \text{storage\_period1:size\_potato2} - 0.01 \times \text{storage\_period2:size\_potato2} + 0.21 \\
& \times \text{storage\_period3:size\_potato2} - 0.23 \times \text{storage\_period4:size\_potato2} + 0.31 \\
& \times \text{storage\_period2:cooking\_method2} + 0.20 \times \text{storage\_period3:cooking\_method2} \\
& - 0.088 \times \text{storage\_period4:cooking\_method2} + 0.21 \\
& \times \text{storage\_period2:cooking\_method3} + 0.14 \times \text{storage\_period3:cooking\_method3} \\
& - 0.18 \times \text{storage\_period4:cooking\_method3} + 0.18 \\
& \times \text{storage\_period2:cooking\_method4} + 0.14 \times \text{storage\_period3:cooking\_method4} \\
& + 0.088 \times \text{storage\_period4:cooking\_method4} \\
& + 0.21 \times \text{storage\_period2:cooking\_method5} + 0.39 \\
& \times \text{storage\_period3:cooking\_method5} - 0.28 \times \text{storage\_period4:cooking\_method5}
\end{aligned}
\tag{3}
$$

In this equation, "variableLevel" represents the level of the variable, for example, area_grow2 represents the second level of the area_grow variable. The coefficients of the model are estimated using the least squares method.

4 Model Assessment

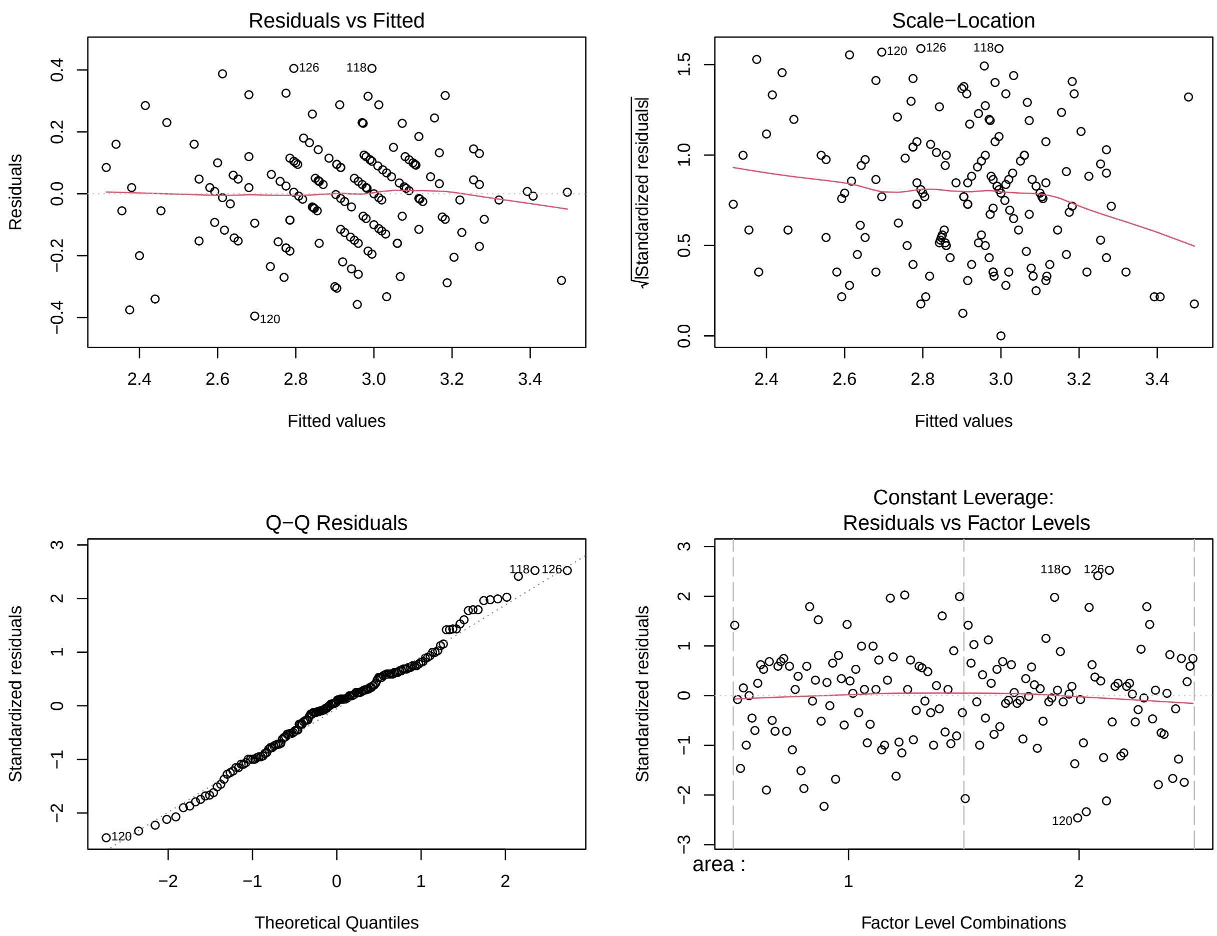
4.1 Model Assumptions

The assumptions of the model are as follows:

1. The errors have mean 0: $E(\epsilon_i) = 0$.
2. Errors are homoscedastic (constant variance): $\text{Var}(\epsilon_i) = \sigma^2$, for all i.
3. Errors are uncorrelated: $\text{Cov}(\epsilon_i, \epsilon_j) = 0$, for all $i \neq j$.
4. Errors are normally distributed: $\epsilon_i \sim N(0, \sigma^2)$.

4.2 Graphical Assessment

We assess the model assumptions by examining the residuals of the model. We plot the residuals against the predicted values to check for homoscedasticity and against the fitted values to check for normality.



5 Conclusions

In this study, we investigated the relationship between the flavor score of cooked potatoes and the growing area, two-week holding temperature, size of potato, storage period, and cooking method. We applied the ANOVA model to fit the full linear model, to find out the effective terms, and to fit the effective model with least square.

We find out that many of the interactions are not significant, and the effective model is presented in Equation 2, to obtain the relation between the flavour score and the effective terms, least square method is applied to estimate the coefficients of the model, resulting in Equatio 3. Finally, we assess the model assumptions by examining the residuals of the model as seen in Section 4.2.