Goods Freshness Analysis

Hatim Hussaini

2024-10-12

Exploring Freshness of Goods

Is the Data Design Balanced?

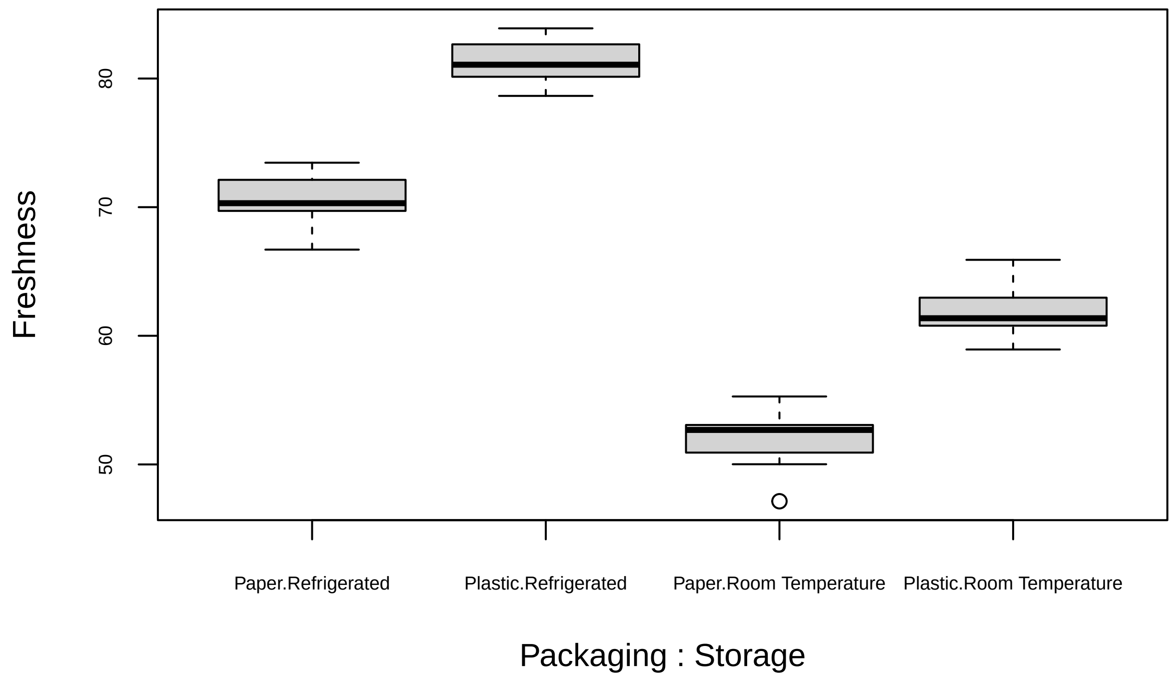
```
table(goods[1:2])
```

```
##          Storage
## Packaging Refrigerated Room Temperature
##   Paper             14               17
##   Plastic           16               18
```

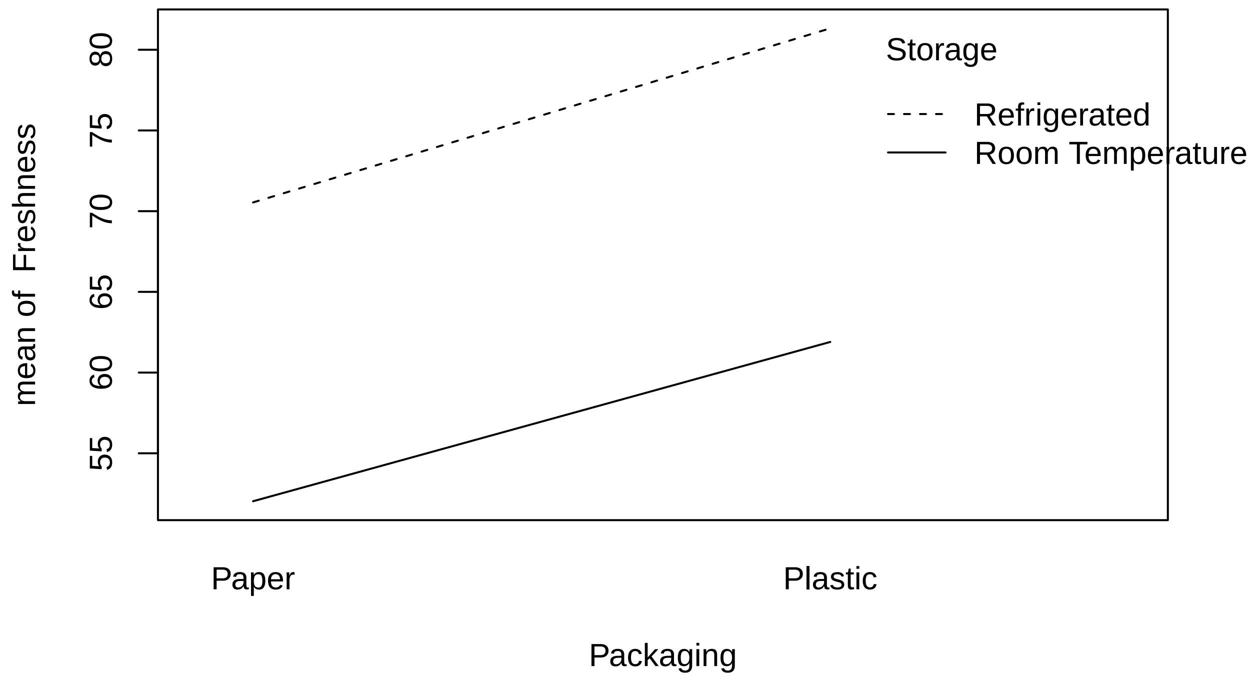
The dataset is unbalanced as categorical variables have unequal observations.

Investigating Features of the Data

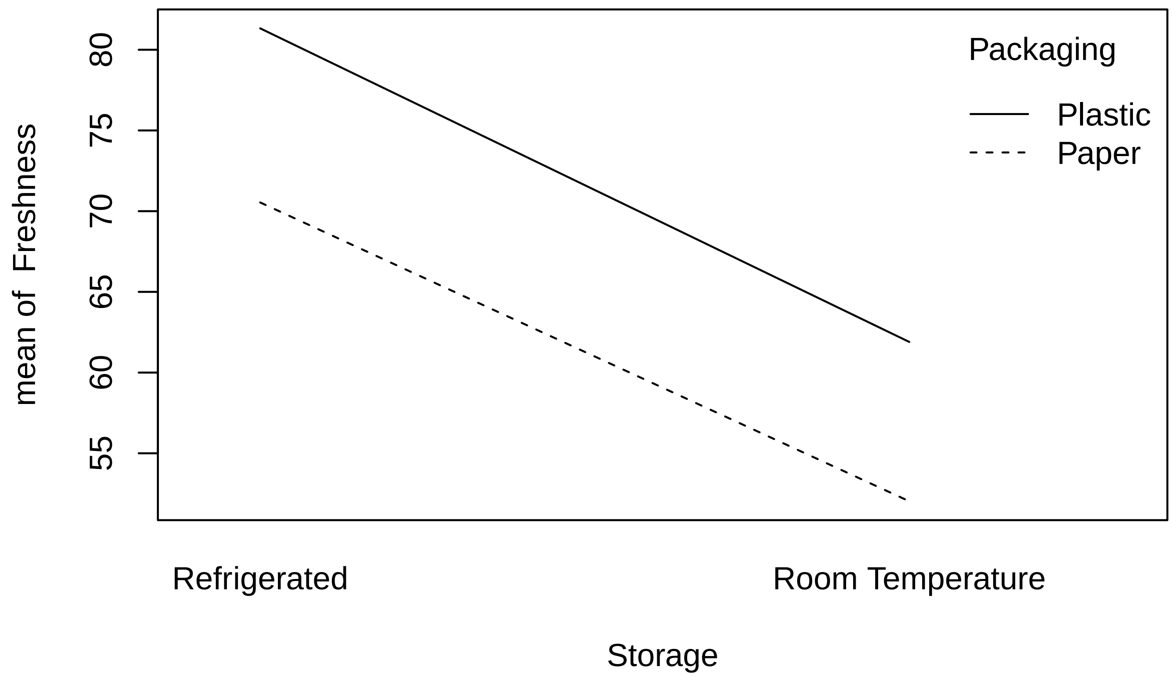
```
boxplot(Freshness ~ Packaging + Storage, data = goods, cex.axis = 0.6)
```




```r
with(goods, interaction.plot(Packaging, Storage, Freshness))
```




```r
with(goods, interaction.plot(Storage, Packaging, Freshness))
```



Insights: - Boxplots suggest equal variances, validated by similar whisker lengths. - Interaction plots show parallel lines, indicating no significant interaction between storage and packaging.

Defining the Full Mathematical Interaction Model

$$Y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \epsilon$$

Where:

- Y_{ij}: The response variable, **Freshness**.
- β_0: The intercept.
- $\beta_1, \beta_2, \beta_3$ are the coefficients for the columns X_1, X_2, X_3 respectively.
  - X_1 is the Packaging type of goods
  - X_2 is the Storage type of goods.
  - X_3 is the interaction between types of Storage and Packing of goods.
- $\epsilon_i \sim N(0, \sigma^2)$ is the random variation.

Analyzing the Effect of Packaging and Storage on Freshness

```
int.goods <- lm(Freshness ~ Storage * Packaging, data = goods)
anova(int.goods)
```

```
## Analysis of Variance Table
##
## Response: Freshness
##                   Df Sum Sq Mean Sq   F value Pr(>F)
## Storage            1 5947.5  5947.5 1983.8043 <2e-16 ***
## Packaging          1 1716.8  1716.8  572.6484 <2e-16 ***
## Storage:Packaging  1    3.4     3.4    1.1295 0.2921
## Residuals         61  182.9     3.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction p-value is insignificant in the linear model with a p-value greater than 0.05 at 0.29. We can go ahead and remove the interaction term and build a model without it.

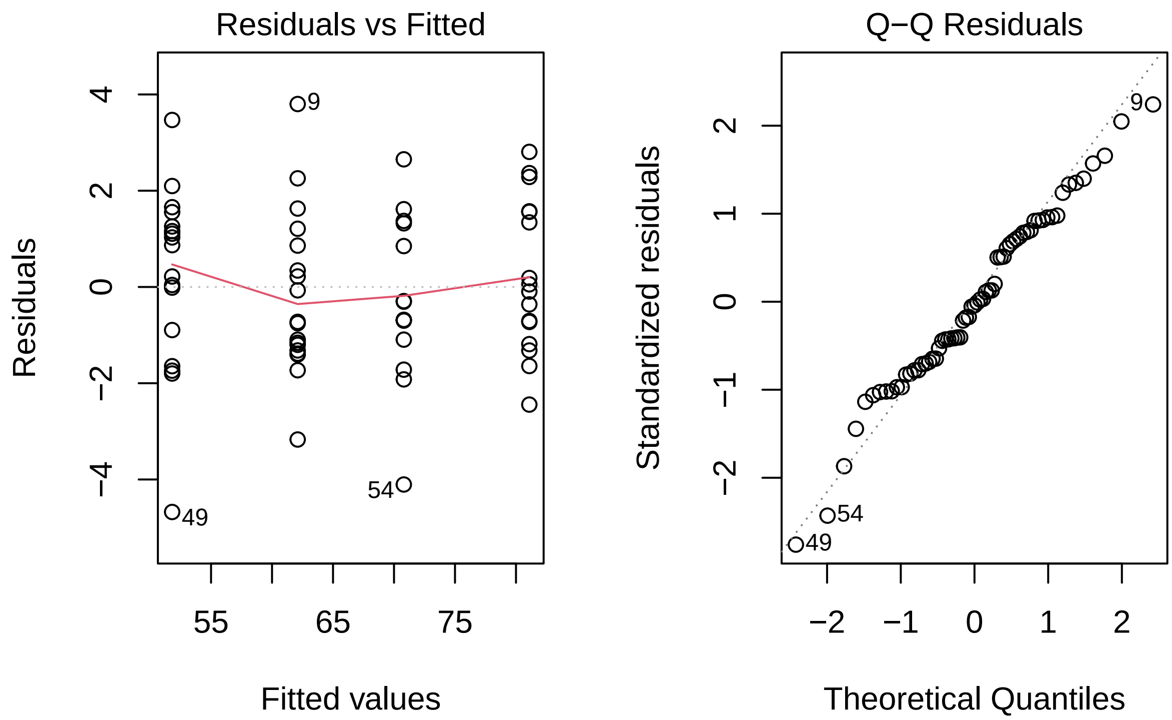
```r
final.goods <- lm(Freshness ~ Storage + Packaging, data = goods)
anova(final.goods)
```

```
## Analysis of Variance Table
##
## Response: Freshness
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Storage    1 5947.5  5947.5 1979.67 < 2.2e-16 ***
## Packaging  1 1716.8  1716.8  571.45 < 2.2e-16 ***
## Residuals 62  186.3     3.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The new linear regression model looks appropriate with both Storage and Packaging playing an important role in predicting the freshness score, with p-values well below 0.05.

Validating the Final Model

```r
par(mfrow = c(1:2))
plot(final.goods, which = 1:2)
```



- The QQ residuals plot closely follows the dotted line at a roughly 45° angle, and the points are near the dotted line, suggesting that the model is normal. However, there are some deviations, but they are not too serious to raise concern.
- The Residuals vs. Fitted plot tests for linearity and constant variance assumptions of the model. The points in the plot are scattered randomly around the red line with no particular trend, validating these assumptions as well.

Conclusion: The final model effectively predicts freshness based on storage and packaging, aligning with statistical assumptions.