ADSC2030 Project Report

Ivan Leonychev T00727314

2024-04-04

Contents

Data overview and research objective

The dataset comprises simulated data on fantasy potion efficiency (potency). The variables in the dataset as follows:

- Ingredients_Number (numerical) - number of ingredients used in potion creation.
- Magic_Powder (numerical) - magic powder used in potion creation (in grams).
- Brewing_Time (numerical) - duration of brewing (in minutes).
- Potion_Duration (numerical) - duration of potion effect after drinking (in minutes).
- Potion_Type (categorical) - type of potion effect: healing, invisibility', strength, poison, fire resistance.
- Brewer_Proficiency (categorical) - level of brewer's proficiency: novice, apprentice', regular, expert, legendary (presented as 1 - 5, where 1 - novice and 5 - legendary).
- Potency (response variable) - an effective of a potion, measured on a scale from 0 to 100 which determine the effect of the potion, for example, a fire resistance potion with potency level 56 grants 56% resistance to fire damage to everyone to drinks it. Calculated on formula (see appendix, "data simulation" chunk for formula coefficients.).

Research objective: determine, which variables have the most effect on potion potency.

Main hypothesis for the research are as follows:

- There is no significant correlation between potion type and brewing time variables.
- Number of ingredients varible has little effect on potion potency.

```
##   Ingredients_Number Magic_Powder Brewing_Time Potion_Duration     Potion_Type
## 1                  1         0.82            8               6 fire resistance
## 2                  5         5.14           36               2          poison
## 3                  1         3.90            4              52          poison
## 4                  9         9.06           22               3         healing
## 5                 10         4.47           18              54        strength
## 6                  4         8.36           45              21        strength
##   Brewer_Proficiency   Potency
## 1                  2 12.116599
## 2                  2 78.115461
## 3                  3  4.268803
## 4                  5 77.079905
## 5                  5 44.439387
## 6                  3 96.470618
```

Analysis

Linear model

```
##
## Call:
## lm(formula = Potency ~ 1 + Ingredients_Number + Magic_Powder +
##     Brewing_Time + Potion_Duration + Potion_Type + Brewer_Proficiency +
##     Potion_Type:Brewing_Time, data = potion_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1997 -0.3261  0.0000  0.4374  1.1202
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         -9.929657   1.301934  -7.627 0.000265 ***
## Ingredients_Number                   0.797225   0.105873   7.530 0.000284 ***
## Magic_Powder                         2.728786   0.097382  28.021 1.37e-07 ***
## Brewing_Time                         1.711227   0.023765  72.005 4.83e-10 ***
## Potion_Duration                     -0.285170   0.015301 -18.637 1.54e-06 ***
## Potion_Typehealing                  -1.270651   1.988653  -0.639 0.546457
## Potion_Typeinvisibility              0.213714   2.253968   0.095 0.927548
## Potion_Typepoison                   -1.291241   1.090872  -1.184 0.281306
## Potion_Typestrength                  1.698928   2.442493   0.696 0.512723
## Brewer_Proficiency                   3.690534   0.241711  15.268 4.98e-06 ***
## Brewing_Time:Potion_Typehealing      0.032693   0.060198   0.543 0.606637
## Brewing_Time:Potion_Typeinvisibility -0.004528   0.060633  -0.075 0.942892
## Brewing_Time:Potion_Typepoison       0.049698   0.039877   1.246 0.259114
## Brewing_Time:Potion_Typestrength    -0.075299   0.071422  -1.054 0.332349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.059 on 6 degrees of freedom
## Multiple R-squared:  0.9997, Adjusted R-squared:  0.9989
## F-statistic:  1349 on 13 and 6 DF,  p-value: 2.759e-09
```

The linear regression model shows a strong fit to the data with an adjusted R-squared of 0.99, indicating that 99% of the variability in the response variable is explained by the predictors. Overall, model is statistically significant with p-value < 0.05. Factors such as ingredients, magic powder, brewing time, and brewer proficiency have substantial effects on potion potency and positive correlation, while potion duration and potion type show minimal influence and have negative correlation, though potion duration is statisticaly significant.

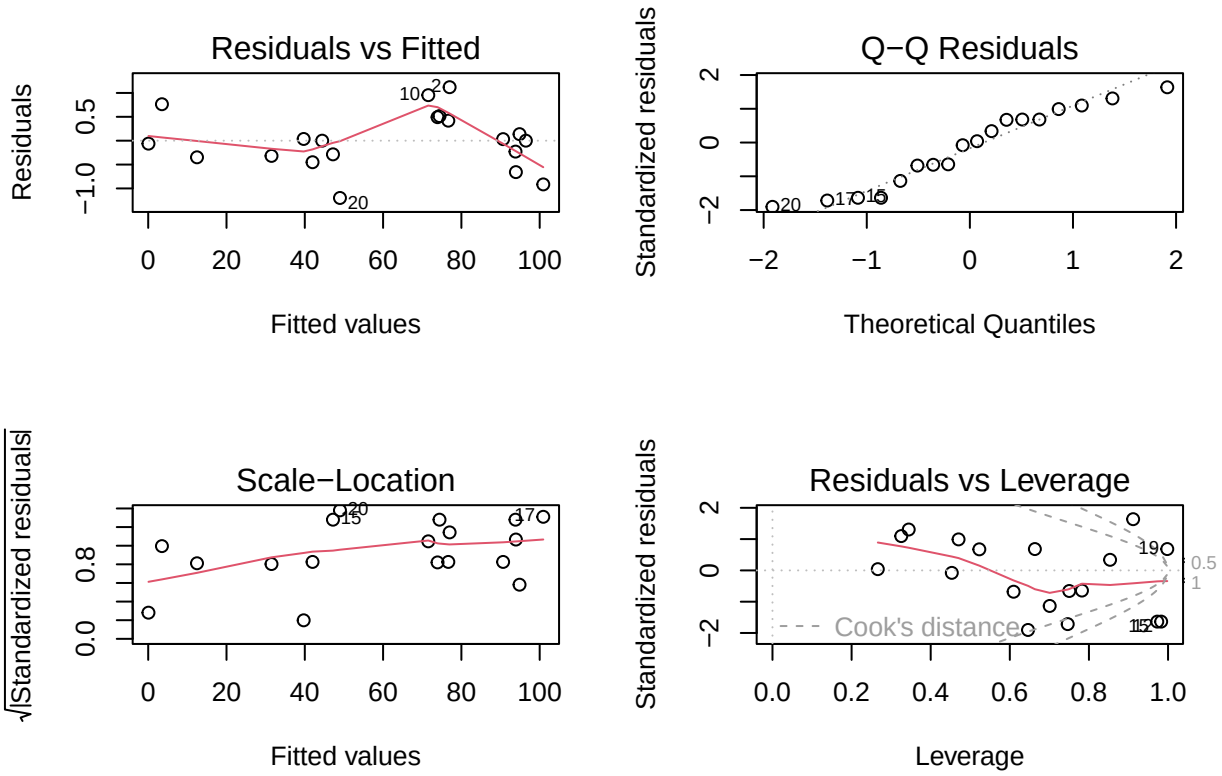
Analysis of variance (ANOVA) and factors

For purpose of the research the studying of the the next model was selected: Potency ~ 1 + Ingredients_Number + Magic_Powder + Brewing_Time + Potion_Duration + Potion_Type + Brewer_Proficiency + Potion_Type:Brewing_Time.

```
##                           Df Sum Sq Mean Sq   F value   Pr(>F)
## Ingredients_Number         1   2425    2425  2162.809 6.62e-09 ***
## Magic_Powder               1     52      52    46.231 0.000496 ***
## Brewing_Time               1  15948   15948 14224.024 2.34e-11 ***
## Potion_Duration            1    773     773   689.506 2.01e-07 ***
## Potion_Type                4    128      32    28.518 0.000480 ***
## Brewer_Proficiency         1    334     334   298.069 2.42e-06 ***
## Brewing_Time:Potion_Type   4      4       1     0.891 0.523165
## Residuals                  6      7       1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA results suggest that Ingredients_Number, Brewing_Time, Potion_Duration, Potion_Type, and Brewer_Proficiency are significant predictors of the response variable. However, the interaction between Brewing_Time and Potion_Type does not significantly contribute to the variability of the response.

Checking model assumption



Does not appear to pass based on visuals.

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.1550068, Df = 1, p = 0.6938
```

Appears to pass.

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.98219, p-value = 0.9591
```

Appears to pass.

Conlusion

Variables with the highest correlation coefficient with potion potency are Magic_Powder, Brewer_Proficiency and Brewing_Time.

Main hypothesis for the both proved correct:

- There is no significant correlation between potion type and brewing time variables and it is not statistically significant.
- Number of ingredients coefficient $<$ than 1, so it doesn't have much effect on potion potency.

Appendix

```
knitr::opts_chunk$set(echo = TRUE)
library(dplyr)
library(tidyr)
library(sjPlot)
library(daewr)
library(lme4)
library(knitr)
library(gmodels)
library(car)
library(MASS)
# Set seed for reproducibility
set.seed(42)

# Define the number of samples
num_samples <- 20

# Define the possible values for categorical variables
potion_types <- c('healing', 'invisibility', 'strength', 'poison', 'fire resistance')
brewer_proficiency_levels <- c('novice', 'apprentice', 'regular', 'expert', 'legendary')

# Generate synthetic data for the dataset
ingredients_number <- sample(1:10, num_samples, replace = TRUE)
magic_powder <- round(runif(num_samples, min = 0, max = 10), 2) # Represented in grams with two digits
brewing_time <- sample(1:60, num_samples, replace = TRUE) # Minutes
potion_duration <- sample(1:60, num_samples, replace = TRUE) # Minutes
potion_type <- sample(potion_types, num_samples, replace = TRUE)
brewer_proficiency <- sample(1:5, num_samples, replace = TRUE) # Assuming 1 corresponds to novice and 5

# Simulate effectiveness of potions
effectiveness <- 0.1 * ingredients_number + 0.3 * magic_powder + 0.2 * brewing_time -
  0.3 * sqrt(potion_duration) + 0.4 * brewer_proficiency

# Normalize effectiveness to range [0, 100]
effectiveness <- (effectiveness - min(effectiveness)) / (max(effectiveness) - min(effectiveness)) * 100

# Create the dataset
potion_data <- data.frame(
  Ingredients_Number = ingredients_number,
  Magic_Powder = magic_powder,
  Brewing_Time = brewing_time,
  Potion_Duration = potion_duration,
  Potion_Type = potion_type,
  Brewer_Proficiency = brewer_proficiency,
  Potency = effectiveness
)

head(potion_data)
model <- lm(Potency ~ 1 + Ingredients_Number + Magic_Powder +
               Brewing_Time + Potion_Duration + Potion_Type +
               Brewer_Proficiency + Potion_Type:Brewing_Time,
             data = potion_data)
```

```r
summary(model)
anova <- aov(Potency ~ 1 + Ingredients_Number + Magic_Powder +
               Brewing_Time + Potion_Duration + Potion_Type +
               Brewer_Proficiency + Potion_Type:Brewing_Time,
             data = potion_data)

summary(anova)
par(mfrow = c(2, 2))

plot(anova)
# Constant Variance of Residuals #

ncvTest(model)
# Normality #

shapiro.test(model$residuals)
```