MP2

Yaretsy Castro, Ruth Kendall

2024-05-10

Introduction

The purpose of our project was to investigate water bottle preferences based on water bottle in which we specifically sought to derive these results based on whether a particular water bottle brand had a cap or the lack thereof. Our research question was "does water bottle brand and design influence perceived notions of water bottle preference based on affordability?" We hypothesized that water bottles pictured with the cap were more likely to be seen as less affordable and in this case ranked lower than those water bottles without a cap.

Method

Our experimental units were the participants who contributed to our study. We measured response variables using a 2 way factorial design which enabled us to examine the interaction effect between design and affordability. Specifically, the water bottle design was the explanatory variable because this factor changed as we displayed different water bottle designs between Essentia and Poland Spring brands. Whereas in the first mp we looked at simply different water brands (all with the same cap design), we now added another factor which is cap type.

To collect our data, we created a survey in which SDS 290 students were able to fill out. THere were four levels in total. Two Essentia bottles and two poland spring bottles, both with different cap design, flip and twist cap.

We hypothesize that water bottle brand and design have a significant different impact on affordability score.

Explore the Data

Lets look at our dataset. We will be conducting a two-way ANOVA. Affordability is our response variable and cap type & water type are our explanatory variable.

When we look at our cell counts, we see that our experiment is almost balanced except for flip top, contains 6 observations instead of 7.

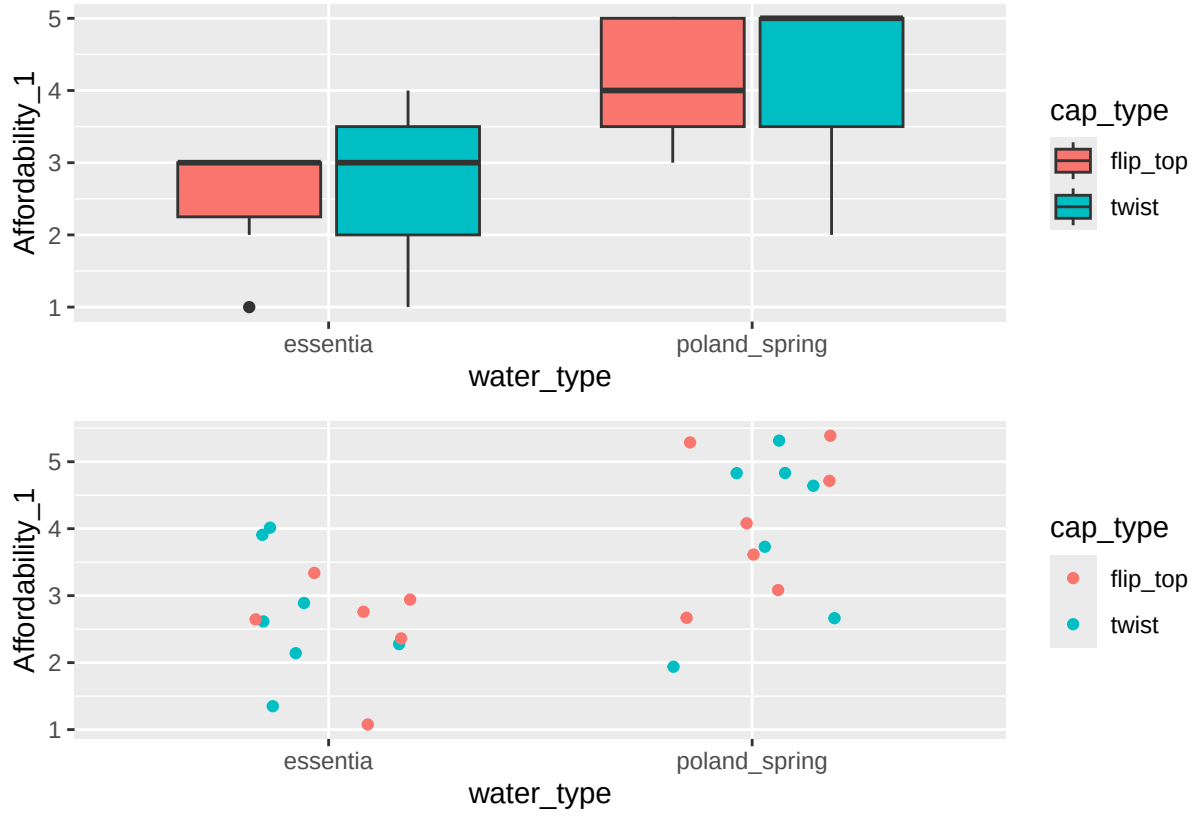
```
##          ResponseId Affordability_1    water_type cap_type
## 1 R_3L0Yi7JmdNvapgY               4 poland_spring flip_top
## 2 R_3nBCmskLmnZUTgG               2      essentia    twist
## 3 R_2rr5TduBDxcjwFu               3      essentia flip_top
## 4 R_77n1RvMSWeWn27v               4 poland_spring    twist
## 5 R_1bTxDJFnGe6jCVs               3      essentia flip_top
## 6 R_7uwzYPAIRhG7ov7               5 poland_spring flip_top
```

```
## # A tibble: 4 x 3
## # Groups:   cap_type, water_type [4]
##   cap_type water_type        n
##   <chr>    <chr>         <int>
## 1 flip_top essentia          6
## 2 flip_top poland_spring     7
## 3 twist    essentia          7
## 4 twist    poland_spring     7
```

Boxplot and Dotplot

Next we look at our boxplot. We notice that there is lower skewness for most of our levels except for Poland Spring & Flip Top, which has a high skewness. We do see that the affordability score is higher for Poland Spring water, suggesting that poland spring may be statistically significantly higher than Essentia.

We then look at the dotplots the distributions for both Essentia and Poland Spring waters have similar distributions. We also notice, just like the boxplots, that the distribution of Poland Spring is higher than Essentia.



Descriptive Statistics and Equal variance

When we look at the descriptive statistics, we notice that the mean for essentia water are almost the same. When we look at Poland Spring, the means, 4.143, is exactly the same. When we look at standard deviation, we notice that it is higher when the cap type is twist.

Now that we are looking at standard deviation, lets check if it passes the condition of same standard deviation. When we divide the maximum standard deviation by the lowest standard deviation, we get 1.452186. Since it is less than 2, we say that this passes the condition.

```
## `summarise()` has grouped output by 'cap_type'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 4 x 5
## # Groups:   cap_type [2]
##   cap_type water_type        n  mean    sd
##   <chr>    <chr>         <int> <dbl> <dbl>
## 1 flip_top essentia          6  2.5  0.837
## 2 flip_top poland_spring     7  4.14 0.900
## 3 twist    essentia          7  2.71 1.11
## 4 twist    poland_spring     7  4.14 1.21
```

```
## # A tibble: 1 x 1
##   `max(sd)/min(sd)`
##               <dbl>
## 1              1.45
```

Formal Anlysis

Next, we do a formal ANOVA analysis without the interaction.We notice that water type is statistically significant with f-statistic of 15.3 and a p-value of 0.0007, which is less than our benchmark 0.05. When we look at water type, we see that cap type is not statistically significant because the f-statistic is less than 1, 0.0126, and the p-value is more than 0.05, 0.911.

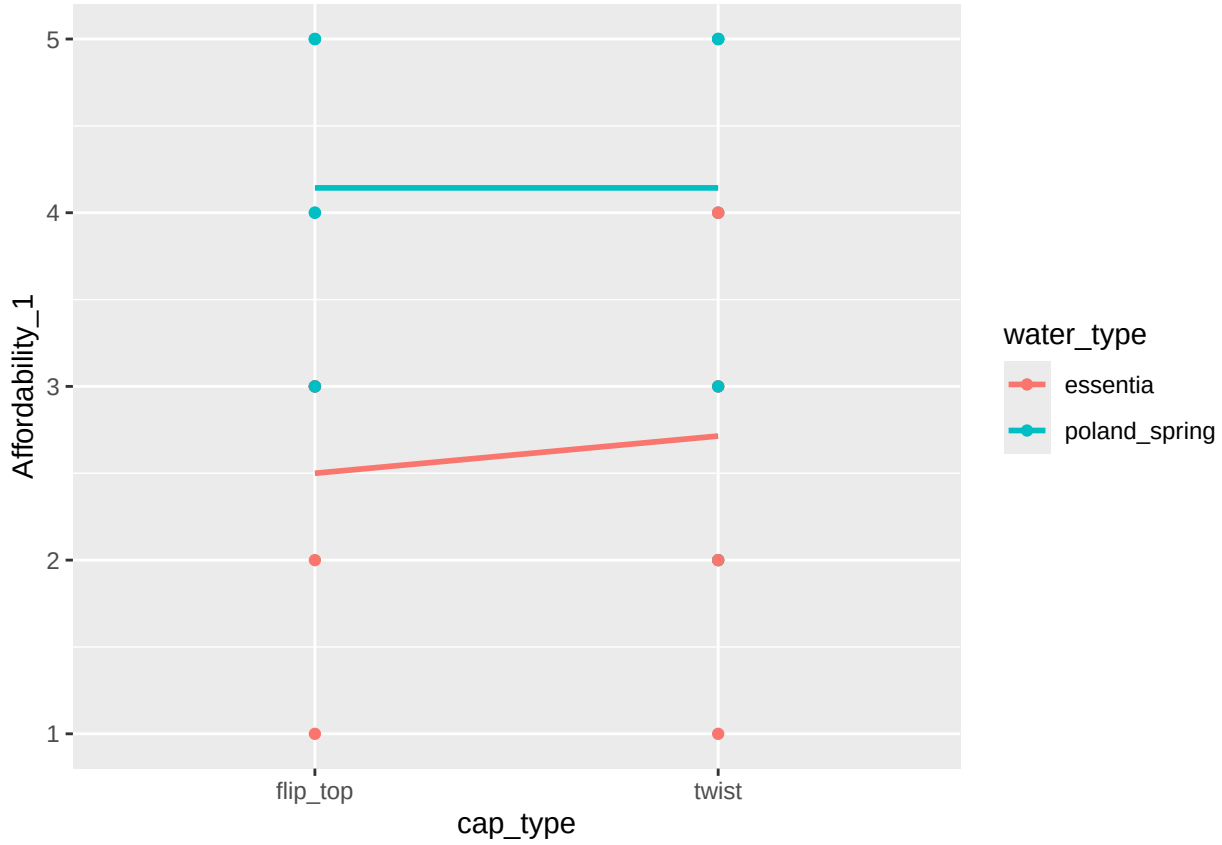
```
# Formal Analysis
mod1 <- lm(Affordability_1 ~ cap_type + water_type, data = ds) # ANOVA Model without interaction term
anova(mod1)
```

```
## Analysis of Variance Table
##
## Response: Affordability_1
##            Df Sum Sq Mean Sq F value    Pr(>F)
## cap_type    1  0.013   0.013  0.0126 0.9114030
## water_type  1 15.786  15.786 15.3257 0.0006533 ***
## Residuals  24 24.720   1.030
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interaction Next, we check if there is an interaction in between our variables, cap type and water type. From the interaction plot, we notice that there might be an interaction, because the slopes are different from each other. From the interaction plot we also see that there are main effects in water type and main effect in cap type when the water type is poland spring but not for essentia.

```
#Interaction plot
ggplot(ds, aes(x = cap_type, y = Affordability_1,
                    group = water_type,
                    color = water_type)) +
  geom_point() +
  geom_smooth(method = "lm", se = 0)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Two-Way ANOVA Since we think there may be an interaction we conduct a two-way ANOVA. From the ANAOVA table we see that the only significant different mena is water type. Water type has a a f-statistic of 14.7331 and a low p-value of 0.0008 which is less than 0.05. The other variables, cap type & cap type * water type have a f-statistic less than 1 and a p-value more than 0.05. For this reason it is best to stick to our first model without the interaction term.

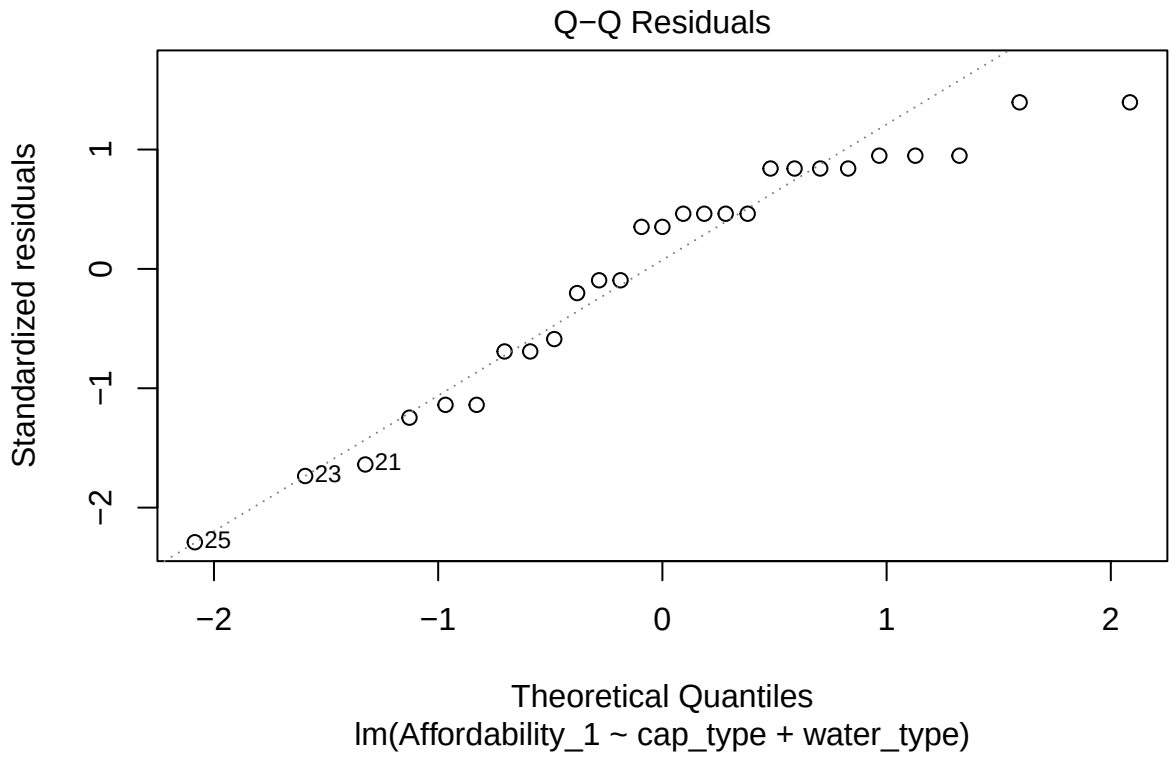
```
#2-way ANOVA
mod2 <- lm(Affordability_1 ~ cap_type * water_type, data = ds)
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: Affordability_1
##                     Df  Sum Sq Mean Sq F value    Pr(>F)
## cap_type             1  0.0130  0.0130  0.0122 0.9131651
## water_type           1 15.7855 15.7855 14.7331 0.0008397 ***
## cap_type:water_type  1  0.0771  0.0771  0.0720 0.7908387
## Residuals           23 24.6429  1.0714
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual Analysis We next look at our residuals to check if it passes our conditions of normality and zero mean. When we look at the qq-plot, we see that the our data is mostly normal therefore it passes our

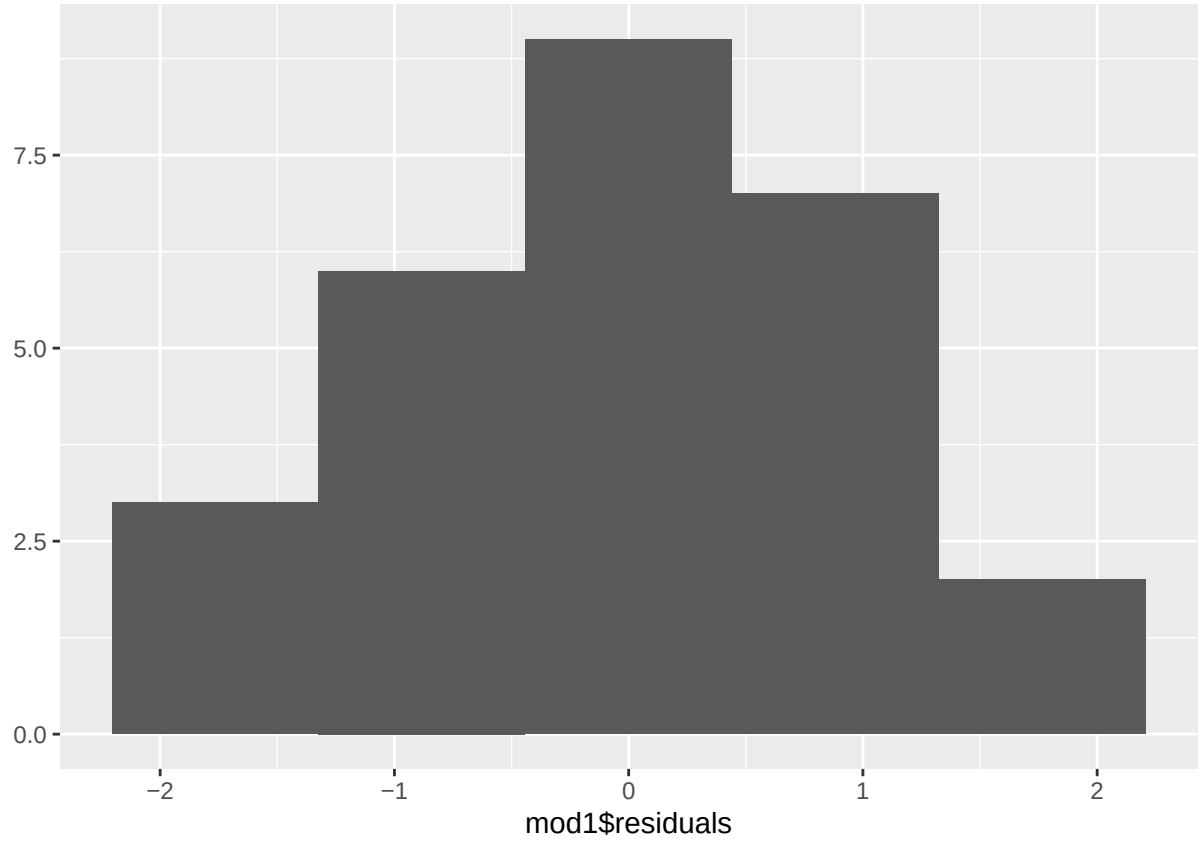
condition. When we look at our histogram we notice that zero is in the centere and therefore also passes the zero mean condition.

```
#Residual analysis
plot(mod1, which = 2)
```




```
qplot(x = mod1$residuals, bins = 5)
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



LSD

Since our model passed our condition, we look at LSD. From our ANOVA table we saw that water type is tatistically different from one another. As we see from below, poland spring and essentia are indeed statistically as the mean values are different by 1.527472 affordability score.

```
library(agricolae)
LSD.test(mod1, "water_type", p.adj = "none")$groups
```

```
##                Affordability_1 groups
## poland_spring         4.142857      a
## essentia              2.615385      b
```

```
4.142857 - 2.615385
```

```
## [1] 1.527472
```

Conclusion

Based on our analysis there is no interaction between our two explanatory variable, cap type and water type. Due to this we decided to stick to our non-interactive model. When we conducted our analysis, we noticed that only water type is statistically different from one another by 1.527472 afordability score.