EXERCISE 2 Single – Factor Experiments: CRD

Ayop, Kyla M. & Garridos, Charlene P.

2024-01-24

Contents

1 R Libraries

```r
library(tidyverse)
library(car)
library(agricolae)
```

2 Problem

Questionnaire Color. In an experiment to investigate the effect of color paper (blue, green, orange) on response rates for questionnaires distributed by the "windshield method" in supermarket parking lots, 15 representative supermarket parking lots were chosen in a metropolitan area and each color was assigned at random to five of the parking lots. The response rates (in percent) follow.

COLOR	RESPONSE RATE
BLUE	28 26 31 27 35
GREEN	34 39 25 31 29
ORANGE	31 25 27 29 28

1. Do visual inspections of the raw data.
2. Is there a difference in the response rate among the different colors? Use $\alpha = 0.05$.

- Treatment: Color (different)
- Treatment Levels: Blue, Green and Orange
- Experimental Units : Supermarket Parking Lots
- Response Variable : Response Rate per Questionnaires
- Experimental Model: Fixed Effects Model

3 Data

```r
color=rep(c('blue','green','orange'), each=5)
response =c(28,26,31,27,35, 34,39,25,31,29, 31,25,27,29,28)
color=as.factor(color)
ques=data.frame(color,response)
class(ques$color)
```

```
## [1] "factor"
```

```r
levels(ques$color)
```

```
## [1] "blue"   "green"  "orange"
```

```
ques
```

```
##     color response
## 1    blue       28
## 2    blue       26
## 3    blue       31
## 4    blue       27
## 5    blue       35
## 6   green       34
## 7   green       39
## 8   green       25
## 9   green       31
## 10  green       29
## 11 orange       31
## 12 orange       25
## 13 orange       27
## 14 orange       29
## 15 orange       28
```

4 Summary

```
desc.data <- ques %>%
group_by(color) %>%
summarize(
mean = mean(response),
standard_deviation=sd(response),
no_of_replicates=n(),
)
desc.data
```

```
## # A tibble: 3 x 4
##   color   mean standard_deviation no_of_replicates
##   <fct>  <dbl>              <dbl>            <int>
## 1 blue    29.4               3.65                5
## 2 green   31.6               5.27                5
## 3 orange  28                 2.24                5
```

With five number of replicates each, the treatment levels of blue, green, and orange are then summarize by their mean and standard deviation. The blue color has a mean of 29.4 and a standard deviation of approximately 3.65. The color of green with the highest mean of 31.6 and a standard deviation of 5.27. And lastly, with the least resulting mean of 28 and standard deviation of 2.24 is the orange paper color.

5 Visualization

```
#visual inspection using box plot
ggplot(ques, aes(x = color, y = response, fill = color)) +
  geom_boxplot(fill = c("blue", "green", "orange")) +
  stat_summary(fun = "mean", geom = "point", shape = 8, size = 2, color = "white")
```

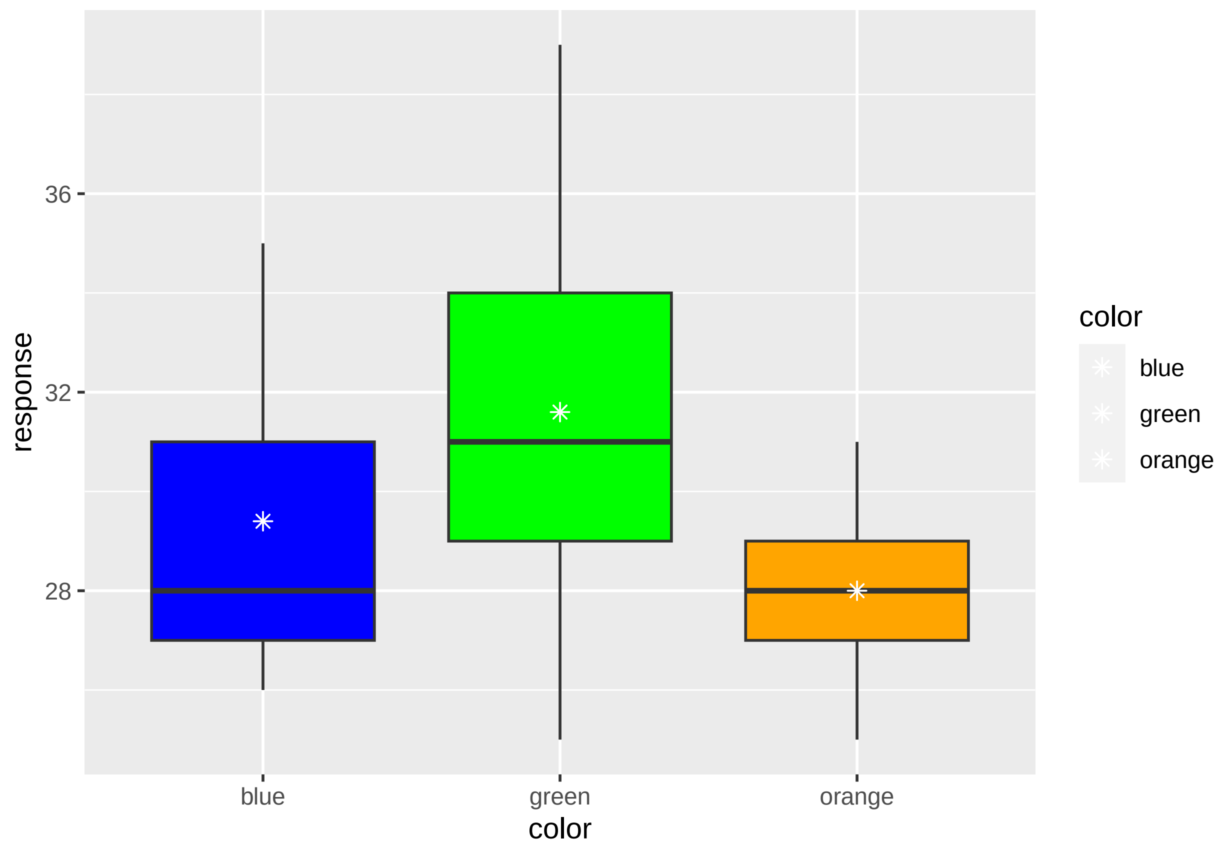

Figure 1: Boxplot Representation of the Effect of Paper Color on Questionnaires Response Rates in Supermarket Parking Lots

The figure above shows a visual summary of the variability of the data. It is to visually inspect the spread of the data. The colors blue and green are skewed to the right, whereas the color orange is symmetrically skewed. Subjectively, the difference between orange and the other colors is noticeable.

6 Hypothesis and Test Statistic

H_O: There are no differences in the mean responses rate percentage among the different color papers.

$$[\mu_1 = \mu_2 = \mu_3]$$

H_a: At least one color paper gave a different response rate percentage.

$$[\mu_i \neq \mu_j, i \neq j]$$

Test statistic: One-way analysis of Variance at $\alpha = 0.05$.

$$F_c = \frac{MST_r}{MSE}$$

7 Data Analysis

```
#ANOVA
anova=aov(response~color, data=ques)
summary(anova)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## color        2  32.93   16.47   1.072  0.373
## Residuals   12 184.40   15.37
```

The results of the test shows the independent variable color with the model error residuals, with the degrees of freedom 2 and 12, respectively.

Moreover, the total variability in the data is measured by adding up sum of squares due to treatment which is equal to 32.93 that is between treatment and the sum of squares due to error equal to 184.40 which is within treatments, hence, the total variability is equal to 217.33.

The mean square of treatment is equal to 16.47 is the measure of variability between groups and this is how much of the overall variability in the data which is greater than the mean square of error equal to 15.37 that measures the unexplained variability within groups that cannot be accounted for by the treatments.

The F value is equal to 1.072 is the test statistic from the F test, since the F value is small,more likely the variation associated with the independent variable is purely from to chance. Along with Pr(>F) equal to 0.373 which is the p-value of the F statistic.

8 Alternative of ANOVA

```
#Another way to run ANOVA
model <- lm(response~color, data=ques)
anova(model)
```

```
## Analysis of Variance Table
##
## Response: response
##           Df  Sum Sq Mean Sq F value Pr(>F)
## color      2  32.933  16.467  1.0716 0.3731
## Residuals 12 184.400  15.367
```

```
summary(model)
```

```
##
## Call:
## lm(formula = response ~ color, data = ques)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -6.6   -2.5   -0.6    2.0    7.4
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   29.400      1.753  16.770 1.08e-09 ***
## colorgreen     2.200      2.479   0.887    0.392
## colororange   -1.400      2.479  -0.565    0.583
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.92 on 12 degrees of freedom
## Multiple R-squared:  0.1515, Adjusted R-squared:  0.01012
## F-statistic: 1.072 on 2 and 12 DF,  p-value: 0.3731
```

9 Interpretation and Conclusion

Decision Rule: Reject H_o at $\alpha = 0.05$ if $F_c > F_t = F_{2,12}(0.05) = 3.89$. Otherwise fail to reject H_o.

Decision: Since $F_c = 1.07 < F_t = 3.89$, then we fail to reject H_o.

Interpretation: Based on the results, the estimated response rate of the intercept under the coefficients is 29.400 and labeled '***' which means it is highly significant. Moreover, the p-value of the intercept is 1.08e-09, which is less than the significant level of 0.05. Hence, it indicates strong evidence against the null hypothesis that there are no differences in the mean response rate percentage among the different color papers.

On the contrary, the other colors have a lower estimated response rate and a p-value greater than the significant level of 0.05. This implies that neither of the colors has a significant effect on the response rate.

Even though the intercept is significant, neither of the different colors has a statistically significant effect on the response rate. In addition, the p-value is 0.3731, which is greater than the significant level of 0.05. Hence, there is not enough evidence to reject the null hypothesis and accept the alternative hypothesis.

Conclusion: We fail to reject the null hypothesis at 5% level of significance. There is not sufficient evidence to support the claim that at least one color paper gave a different response rate percentage.