

Medicine effect with analysis of variance ANOVA in R

Caio di Felice Cunha

Analyzing the Financial Market

A pharmaceutical company tested three formulations of a pain relief drug for headache (specifically migraine) sufferers. For the experiment, 27 volunteers were selected and 9 were randomly assigned to one of the three drug formulations.

Subjects were instructed to take the medication during their next migraine episode and to report their pain on a scale of 1 to 10 (10 being the most severe pain). We'll create a mass of test data simulating these results during the hands-on classes that follow.

We need to compare whether all three drugs generate the same effect or not. I.e:

- If the AVERAGE of each group is the same or very similar, it means that all three drugs are effective.
- If the AVERAGE is different, it means that out of three, only one or two drugs are effective.

Based on the answer we want to obtain, let's define the hypotheses for the statistical test:

- H0 (Null Hypothesis)= The AVERAGE of the groups is the same.
- Ha (Alternative Hypothesis) = The MEAN of all groups is not the same.

ANOVA will be used as a statistical test in this scenario. Our goal is to reject or not H0.

Stage 1 - Create the data

```
# Data provided:
# Medicine A = 4 5 4 3 2 4 3 4 4
# Medicine B = 6 8 4 5 4 6 5 8 6
# Medicine C = 6 7 6 6 7 5 6 5 5

# List with the answers of the volunteers
volunteer_note <- c(4,5,4,3,2,4,3,4,4,6,8,4,5,4,6,5,8,6,6,7,6,6,7,5,6,5,5)
volunteer_note
```

```
## [1] 4 5 4 3 2 4 3 4 4 6 8 4 5 4 6 5 8 6 6 7 6 6 7 5 6 5 5
```

```
# List with the total tests of the 3 drugs
medicine <- c(rep("A",9), rep("B",9), rep("C",9))
medicine
```

```
## [1] "A" "A" "A" "A" "A" "A" "A" "A" "A" "B" "B" "B" "B" "B" "B" "B" "B" "B" "C"
## [20] "C" "C" "C" "C" "C" "C" "C" "C" "C"
```

```
# Now we create a dataframe
df <- data.frame(volunteer_note, medicine)
```

Stage 2 - Applying the test

```
# We apply ANOVA
test_anova <- aov(volunteer_note ~ medicine, data = df)
test_anova
```

```
## Call:
## aov(formula = volunteer_note ~ medicine, data = df)
##
## Terms:
##              medicine Residuals
## Sum of Squares  28.22222  28.44444
## Deg. of Freedom      2      24
##
## Residual standard error: 1.088662
## Estimated effects may be unbalanced
```

```
# test summary
summary(test_anova)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## medicine      2  28.22  14.111    11.91 0.000256 ***
## Residuals    24  28.44   1.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Stage 3 - Conclusion

P-value < 0.05, that is, $0.000256 < 0.05$, so we reject H_0 . The averages of the groups are not the same and consequently the drugs do not have the same effect

Disclaimer:

Disclaimer: a good part of this project was largely done in the Data Science Academy, Big Data Analytics with R and Microsoft Azure Machine Learning course (part of the Data Scientist training)

End