NON-PARAMETRIC TEST
ANOVA

Aditya Joshi

GRAPHIC ERA DEEMED TO BE UNIVERSITY 05-05-2020

A **non parametric test** (sometimes called a distribution free test) does not assume anything about the underlying distribution (for example, that the data comes from a normal distribution). That's compared to parametric test, which makes assumptions about a population's parameters (for example, the mean or standard deviation); When the word "non parametric" is used in stats, it doesn't quite mean that you know nothing about the population. It usually means that you know the population data does not have a normal distribution.

Analysis of Variance (ANOVA)

Variance

The average of the **squared** differences from the Mean.

The heights of dogs (at the shoulders) are: 600mm, 470mm, 170mm, 430mm and 300mm.

Find out the Mean, the Variance, and the Standard Deviation.

Your first step is to find the Mean:

$$\begin{aligned} \text{Mean} &= 600 + 470 + 170 + 430 + 300/5 \\ &= 1970/5 \\ &= 394 \end{aligned}$$

To calculate the Variance, take each difference, square it, and then average the result:

$$\begin{aligned} &\text{Variance} \\ \sigma^2 &= 206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2 /5 \\ &= 42436 + 5776 + 50176 + 1296 + 8836/5 \\ &= 108520/5 \\ &= 21704 \end{aligned}$$

So the Variance is 21,704

And the Standard Deviation is just the square root of Variance, so:

Standard Deviation

$\sigma = \sqrt{21704}$

$= 147.32...$

$= 147$ *(to the nearest mm)*

Degrees of freedom

the number of independent values or quantities which can be assigned to a statistical distribution.

Lets say the beverages the people drink effects the reaction time set up an experiment with three groups of people

1st group Water to drink

2nd group is sugary fruit Juice

3rd group is coffee

Now test everyone reaction time and know the difference in reaction time in groups

Null hypothesis says Ho: the mean reaction time of all three group is same

If there are two groups you can find the difference between them with t-Test

If you have three groups of more you need to use different approach which is called analysis of variance ANOVA

When you do the experiment, the score is same the total variation of all the scores is made up of two parts

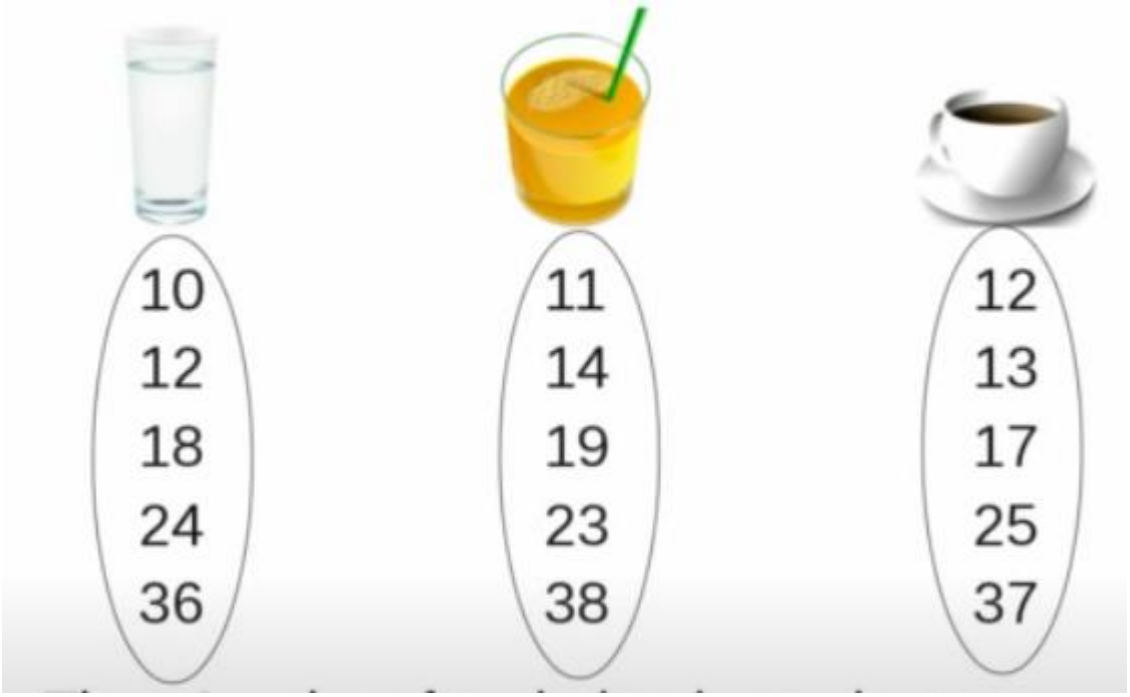
1. the variation with in each group because the people in each have different reaction time

And

2. the variation between the groups because the drinks you give in the three groups are different

Example 1.

Look at this set of scores



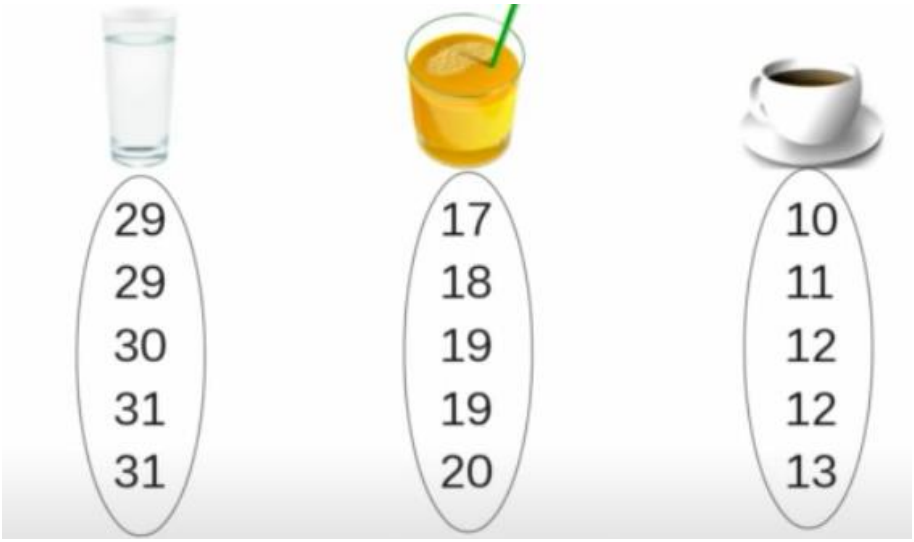
there is lot if variation in each group it is sorted to look easier some people are faster and some are slower

but each group looks pretty much the same

conclusion: it's the people that make the difference not the drink

then you will accept null hypothesis that type of drink does not affect reaction time

Example 2. now look at the different set of numbers



There is not much variation in this group

But the groups are very different from one another

In this case you will reject the null hypothesis

Conclusion: its's the drink that makes the difference, not the people

Here is the idea behind ANOVA

Figure out how much of the total variance comes from:

The variance between the groups

The variance with in the groups

Calculate the ratio

$$F = \frac{\text{between group}}{\text{within groups}}$$

Larger the ratio, the more likely it is that the groups have different means(reject Ho)

In previous example it is obvious what the variance was

Now look at these numbers

The variance between and within groups is not so obvious

The result of an ANOVA shows

$F(2,12)=4.27$, $p=.04$

Reject the null hypothesis

What is 2,12 is doing here it is the degrees of freedom between groups and within groups

Calculate degrees of freedom as follows

F(b,w)

b is the degrees of freedom for variance between groups

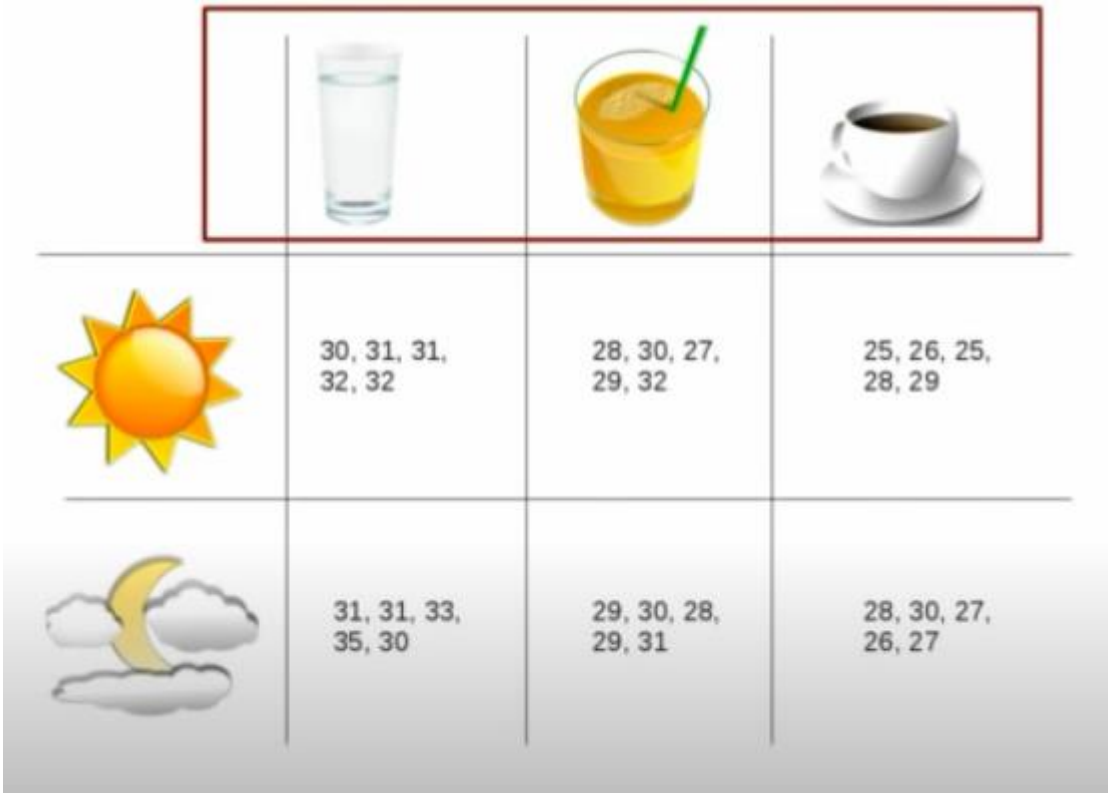
w is the degrees of freedom for variance within groups

b=number of groups-1

w=total number of observations – number of groups

it also works when you have multiple variables

if you test three groups for the reaction time in the morning and test another three groups in the evening



ANOVA tells you if there is a significant effect for the type of drink or type of day makes the difference or some interaction for eg coffee might be more effective in the morning then in the evening

Recap:

If variation is between group then there is probably significant effect

If within groups there is not probably significant effect

Implementation in R


```
Group1 <- c(2,3,7,2,6)
Group2 <- c(10,8,7,5,10)
Group3 <- c(10,13,14,13,15)
Combined_Groups <- data.frame(cbind(Group1, Group2, Group3))
Stacked_Groups <- stack(Combined_Groups)
```

```
Anova_Results <- aov(values ~ ind, data = Stacked_Groups)
summary(Anova_Results)
```

```
> summary(Anova)
            Df Sum Sq Mean Sq F value   Pr(>F)
ind          2  203.3   101.7   22.59 8.54e-05 ***
Residuals   12   54.0     4.5
                                      0.0000845
```

$F(2,12) = 22.59,\ p<.05$