Ahmad Mohammad – Omar AbdelKader

Supervisor: Dr. Hussein Nassar

Subject: Statistical Inference

Objective: ANOVA statistical testing

What is ANOVA and how we use it ?

ANOVA stands for Analysis of Variance, it tells you if there are any statistical difference between the means of three or more independent groups.

We use ANOVA when we want to test a particular hypothesis, it helps us to understand how the different groups respond with a null hypothesis where the mean of the different group are equal, and if there as statistically significant results (which indicates that the means are not equal). (1)

The question now, what is the statistical significant result?

This result is indicated by the F-value where the F-value is conducted form the ANOVA table and is compared to F-critical value. This process is don by the 6-steps procedure, where in the first step we state the null and the alternate hypothesis, second step choose the level of significance, third step determine the test-statistic which is ANOVA test as we said (F-distribution). Forth step, formulate the decision rule, we reject the null hypothesis if F (exported from ANOVA table) is greater than F-critical (extracted from F-distribution table). Fifth step, calculate F-value (in fact it is extracted from ANOVA table test) and make the decision rule if we want to reject or not reject H_0 (null hypothesis). Last step, is interpreting the result, if the means are equal or not.

For better understanding to how ANOVA work, we will conduct three ANOVA tests on three datasets that we extracted from an online survey filled up by 12 individuals aged between 18 and 29 years old (6 males and 6 females), in which they rate four film categories(action, romance, comedy, horror).

It is worth noting that there is three ANOVA tests: One-way ANOVA, Two-way ANOVA with replication and Two-way ANOVA without replication. In this paper we're going to observe the difference between these three tests and taking into consideration the blocking variable. For this purpose, three datasets extracted from an online survey will be studied, each dataset will be performed by specific ANOVA test deppending on the variables we have.

1-In the first dataset we are going to work on **One-way ANOVA test.**

An online survey about rating films was filled up by 12 individuals and the results are shown below.

Rating the four films categories from 1 to 10 (from worst to best)

Action	Romance	Comedy	Horror
10	3	7	5
8	7	10	3
5	8	10	3
8	7	10	1
8	7	10	7
8	2	1	6
7	8	8	5
9	10	10	1
9	4	6	10
9	8	9	2
5	9	10	7
8	8	8	3

Is there a statistically significant difference in the mean of the four categories ?

To answer this question we will perform the 6-steps procedure.

Step1 :

H_0: Null hypothesis: The means of the four categories are equal ($\mu_{action}= \mu_{romance}= \mu_{comedy}= \mu_{horror}$)

H_1: Alternate hypothesis: At least one mean is not equal.

Step2 :

Level of significance (α=0.05)

Step3 :

Select the appropriate test statistic, which is F-test statistic for ANOVA.

Step4 :

Set up the decision rule. The appropriate critical value can be found in a table of probabilities for the F distribution. In order to determine the critical value of F we need degrees of freedom, df_1=k-1 and df_2=N-k, where k=4 and N=48 then df_1=3 and df_2=44. Reject H_0 if F>F-critical (F>2.81)

Step5 :

Compute the test statistic using ANOVA table as it is shown below.

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	106.2292	3	35.40972	6.156185	0.001376	2.816466
Within Groups	253.0833	44	5.751894			
Total	359.3125	47				

F=6.15 > F-critical =2.81. So reject H_0.

Step6 :

Interepretation of the results : As we reject H_0 , so the means of the four categories are not equal. We have statistically significant evidence at α=0.05 to show that there is a difference in the means of the four categories.

Analysis of “confidence interval for the difference in treatment means”

Action		*Romance*		*Comedy*		*Horror*	
Mean	7.833333333	Mean	6.75	Mean	8.25	Mean	4.416666667
Standard Error	0.440958552	Standard Error	0.708444733	Standard Error	0.76993703	Standard Error	0.792563288
Median	8	Median	7.5	Median	9.5	Median	4
Mode	8	Mode	8	Mode	10	Mode	3
Standard Deviation	1.527525232	Standard Deviation	2.454124543	Standard Deviation	2.667140109	Standard Deviation	2.745519766

Here we will make the analysis of confidence interval for the difference in treatment means since the null hypothesis is rejected, so we can identify the pairs of means that differ with the confidence interval, knowing that the level of significance is 0.005 (two tailed) and df=11 so that t-value=2.201.

-First pair(Action and Romance) : *confidence interval for the difference in treatment means=(-1.07 , 3.23)*

Zero is included in this interval, so it appears to have **no difference** in the means of the Action and Romance.

-Second pair(Action and Comedy) : *confidence interval for the difference in treatment means=(-2.57 , 1.73)*

Zero is also included in this interval, so there is **no difference** in the means of the Action and Comedy.

-Third pair(Action and Horror) : *confidence interval for the difference in treatment means=(1.26 , 5.57)*

Zero is not captured in this interval (interval is above zero), so it appears to have a **difference** in the means of the Action and Horror.

We conclude that the mean of Action, Comedy and Romance are almost equal to each other with a 0.05 siginficance level unless the mean of these three categories are not equal to that of Horror. This proves that the alternate hypothesis is accepted (H_1 : at least one treatment mean is different).

2-In the second study we wil take into consideration the age of each individual as a second treatment variable, thus the **Two-way ANOVA test** with a **blocking variable** will be performed for better studying (age is the blocking variable).

The following table represents the result we got.

Age	Action	Romance	Comedy	Horror
20	10	3	7	5
21	8	7	10	3
19	5	8	10	3
19	8	7	10	1
25	8	7	10	7
19	8	2	1	6
19	7	8	8	5
20	9	10	10	1
23	9	4	6	10
18	9	8	9	2
24	5	9	10	7
29	8	8	8	3

Randomly we choose 4 individuals of different ages (18,19,20 and 21) to find that if the means of the ages are equal or not (considering the four ages as blocking variables) so we obtain a new table as shown below.

Age	Action	Romance	Comedy	Horror
18	9	8	9	2
19	5	8	10	3
20	10	3	7	5
21	8	7	10	3

Is there a statistically significant difference in the mean of the blocking variable (age of the 4 individuals) ?

Also to answer this question we'll need the 6-steps approach stating the null and the alternate hypothesis regarding the means of age of the individuals without denying the means of the four categories.

Step1 :

For the 4 categories:

H_0: Null hypothesis: The means of the four categories are equal ($\mu_{action}= \mu_{romance}= \mu_{comedy}= \mu_{horror}$)

H_1: Alternate hypothesis: The means are not equal.

For the blocking variable (age):

H_0: Null hypothesis: The means of the age of all individuals are equal ($\mu_{18}= \mu_{19}= \mu_{20}= \mu_{21}$)

H_1: Alternate hypothesis: The means are not equal.

Step2 :

Level of significance (α=0.05)

Step3 :

Select the appropriate test statistic, which is F-test statistic for ANOVA.

Step4 :

Set up the decision rule. The appropriate critical value can be found in a table of probabilities for the F distribution. In order to determine the critical value of F we need degrees of freedom. For blocks (k-1=4-1=3), for treatments(4-1=3), for the error ((k-1)(b-1)=9).

If F>F-critical of the treatments(columns) reject H_0 (null hypothesis of the four categories)

If F>F-critical of the blocks(rows) reject H_0 (null hypothesis of the blocking variable)

Step5 :

Compute the test statistic using ANOVA table as it is shown below.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.6875	3	0.5625	0.126365	0.942099	3.862548
Columns	75.6875	3	25.22917	5.667707	0.018479	3.862548
Error	40.0625	9	4.451389			
Total	117.4375	15				

For rows (age : blocking variable) F=0.126 < F=3.86 ; so don't reject H_0.

For columns (categories) F=5.66 > F=3.86 ; so reject H_0.

Step6 :

Interepretation of the results : Observing the Rows in the ANOVA table we conclude that the means of the ages are equal. However, at least one film rate is different from others through observing the columns in the ANOVA table.

3-Now we are going to study the interaction of the gender and the film rating, where we divide the 12 individuals into two equal parts in which we got 6 males and 6 females, so we use the **Two-way ANOVA test with replication.**

For this purpose we got this table.

gender	action	romance	comedy	horror
male	10	3	7	5
	8	7	10	3
	8	7	10	1
	8	7	10	7
	9	8	9	2
	8	8	8	3
female	5	8	10	3
	8	2	1	6
	7	8	8	5
	9	10	10	1
	9	4	6	10
	5	9	10	7

Is there a satistical difference between the means of the four categories(action, romance, comedy, horror) ? Is there a statistical difference between the genders(male and female) ? And is there an interaction between rating the four categories and the gender of the individuals ?

The 6-step approach will be also done to answer these three questions.

Step1 :

For the 4 categories:

H_0: Null hypothesis: The means of the four categories are equal ($\mu_{action}= \mu_{romance}= \mu_{comedy}= \mu_{horror}$)

H_1: Alternate hypothesis: The means are not equal.

For the genders:

H_0: Null hypothesis: The means of the two genders are equal.

H_1: Alternate hypothesis: The means are not equal.

Interaction:

H_0: Null hypothesis: There is an interaction between the rating of categories and the gender of the individuals.

H_1: Alternate hypothesis: There is no interaction.

Step2 :

Level of significance (α=0.05)

Step3 :

Select the appropriate test statistic, which is F-test statistic for ANOVA.

Step4 :

Set up the decision rule. The appropriate critical value can be found in a table of probabilities for the F distribution. In order to determine the critical value of F we need degrees of freedom. For the gender (2-1=1), for categories (4-1=3), for the inetraction ((2-1)(4-1)=3).

If F>F-critical of the categories(columns) reject H_0 (null hypothesis of the four categories).

If F>F-critical of the gender(sample) reject H_0 (null hypothesis of the blocking variable).

If F>F-critical of the ineraction reject H_0 .

Step5 :

Compute the test statistic using ANOVA table as it is shown below.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.520833	1	0.520833	0.090253	0.765412	4.084746
Columns	106.2292	3	35.40972	6.135981	0.001559	2.838745
Interaction	21.72917	3	7.243056	1.255114	0.302721	2.838745
Within	230.8333	40	5.770833			
Total	359.3125	47				

For sample : F=0.09 < F-critical=4.08 ; do not reject H_0 (null hypothesis of the gender)

For columns : F=6.13 > F-critical=2.83 ; reject H_0 (null hypothesis of the categories)

For interaction : F=1.25 < F-critical=2.83 ; do not reject H_0 (null hypothesis of the interaction)

Step6 :

Interepretation of the results : Considering the gender, we accept the null hypothesis but we'll reject the null hypothesis of the categories as we conclude that the mean of the genders are equal but that of the four categories aren't equal. Also we don't reject the null hypothesis of the interaction, thus we conclude that there is an interaction between the rating of the four film categories and the gender.

Finally, from the first study we conclude that the means of the four films categories, but as we study the confidence interval for the difference in treatment means we deduce that the mean of Action, Romance, Comedy are almost equal to each other but they are not equal to that of Horror. In the second study we observe that the means of the ages (18,19,20 and 21) are equal to each other and through the third study we conclude that there is an intercation between the gender and rating the film categories.

References:

(1) https://www.qualtrics.com/uk/experience-management/research/anova/?rid=ip&prevsite=en&newsite=uk&geo=LB&geomatch=uk