VIETNAM NATIONAL UNIVERSITY
UNIVERSITY OF INFORMATION TECHNOLOGY
INFORMATION SYSTEMS FACULTY

REPORT LAB2

SUBJECT: DATA ANALYSIS IN BUSINESS

Lecturer: Assoc. Prof. Nguyen Dinh Thuan

Instructor: TA. Nguyen Minh Nhut

Class: IS403.O22.HTCL

Group 3:

21521049 – Ho Quang Lam

21521586 – Le Thi Le Truc

21521938 – Nguyen Thanh Dat

Ho Chi Minh City, March 2024

ACKNOWLEDGEMENT

First of all, we would like to express our deepest gratitude and appreciation to our lecturers, Mr. Nguyen Dinh Thuan and Mr. Nguyen Minh Nhut, for their teaching and sharing of extensive knowledge as well as practical examples during the lectures. They have guided us in completing our Lab 01 report by providing valuable feedback, suggestions, and assistance with exercises and revisions.

The Data Analysis in Business course is an interesting and highly practical subject. However, due to our limited expertise and initial unfamiliarity with real-world applications, we acknowledge that our Lab 01 report may contain some shortcomings and inaccuracies despite our best efforts. We sincerely hope to receive further guidance and feedback from Mr. Nguyen Dinh Thuan and Mr. Nguyen Minh Nhut to improve our knowledge and equip ourselves for future projects as well as for our academic and professional endeavors.

Once again, we would like to extend our heartfelt and sincere gratitude to our lecturers and peers.

Ho Chi Minh City, March 2024

Group of student performers

Ho Quang Lam

Le Thi Le Truc

Nguyen Thanh Dat

LECTURER'S COMMENTS

TABLE OF CONTENTS

WORK DISTRIBUTION

Members / Works	Le Thi Le Truc (Leader)	Ho Quang Lam	Nguyen Thanh Dat
Problem statement	✓	✓	✓
Build the report template	✓		
Do all question 1	✓		
Do all exercise with Excel	✓		
Do all exercise with Python		✓	
Do all exercise with R			✓
Summarize and edit reports	✓	✓	✓
Completion	**100%**	**100%**	**100%**

CHAPTER 1. EXPLANATION AND ILLUSTRATIVE EXAMPLE OF LEVENE AND TUKEY TEST

a) What is Levene Test for Equality of Variances? Explanation and example.
b) What are post hoc comparison tests used for in ANOVA? Explanation and example

1.1. LEVENE'S TEST

1.1.1. EXPLANATION

Levene's test is the inferential statistical method used in SPSS to evaluate the consistency of variance for two or more groups of data

Identify hypothetical devices:

- H0: Variances between groups are equal
- H1: There is 1 variance among the 3 variances that is different from the remaining 2 variances

With significance 0.05 we have:

- If Sig (or p-value) < 0.05 (or F > Fcrit): Reject H0, means are wrong between groups different.
- If Sig (or p-value) ≥ 0.05 (or F < Fcrit): Accept H0, the mean between equal group

Levene's test formula:

$$W = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2},$$

In there:

N: Sample size

k: Number of groups in the sample

Ni : Size of group i in the sample

$Zij = |Yij - \bar{Y}i.|$;

Yij is the jth value of group i,

$\bar{Y}i.$ is the average of group i. In addition, $\bar{Y}i.$ can also be the median value of group i.

$\bar{Z}i.$: Group average of Zij $\bar{Z}..$: Overall average of Zij

If $W > F(1 - \alpha, k-1, n-k)$ then we reject H0

If we accept the hypothesis H0 of Levene's test → We can **test ANOVA**

1.1.2. EXAMPLE

Problem: Suppose we want to evaluate the effectiveness of three learning methods A, B and C (traditional learning methods, online learning methods, new learning methods) for improving scores on a math test. We collect data on students' scores (each method takes 5 different random students) after participating in each learning method. The question is determined see if there are any differences between groups

	METHOD A	METHOD B	METHOD C
1	8	7.8	8.5
2	7.5	8.5	9
3	8.5	8	8.8
4	9	8.8	8.4
5	8.2	9.2	9.2

Determine the average value of 3 groups

MEAN	8.24	8.46	8.78

Calculate the deviation within each group

ABS(xij-meanxi)	0.24	0.66	0.28
	0.74	0.04	0.22
	0.26	0.46	0.02
	0.76	0.34	0.38
	0.04	0.74	0.42

Calculating ANOVA we get the results as shown

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	5	2.04	0.408	0.10492		
Column 2	5	2.24	0.448	0.07712		
Column 3	5	1.32	0.264	0.02488		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.093653	2	0.046827	0.67891	0.5256226	3.885294
Within Groups	0.82768	12	0.068973			
Total	0.921333	14				

We can see that: Because F < Fcrit (or p-value = 0.5256 > α = 0.05), we accept hypothesis H0.

⇨ So there is no difference in the variance of the 3 methods

1.2. TUKEY'S TEST

1.2.1. EXPLANATION

Tukey's test (also known as comparing each pair of overall averages with each other) with the assumption that two independent random samples are taken in pairs (3 or more populations) with analysis normal distribution and different methods.

Apply the **Tukey's Test** if and only if the hypothesis H0 in the ANOVA test is dropped (i.e. there is a difference between the population means).

The problem is posed next:

- Overall averages vary

- The population has a larger or smaller average.

  Use the Tukey test to compare each population pair with each other.

  Suppose we need to test the difference of three overall average ratings.

  Call it the average of 3 corresponding populations.

  We will have the following steps to perform the Tukey test:

  Step 1. Determine the hypothesis:

  TH1:

  H0: µ1= µ2

  H1: µ1 ≠ µ2

  TH2:

  H0: µ2= µ3

  H1: µ2≠ µ3

  TH3:

  H0: µ1= µ3

  H1: µ1≠ µ3

With the population k, we determine the number of hypotheses (average number of pairs needed compare) equals C_n^2

µ1, µ2, µ3 are the average values of the groups

Step 2: Calculate Tukey value:

$$T = q_{\alpha(k,n-k)}\sqrt{\frac{MSW}{n_{min}}}$$

In which:

k: group number

n: total number of elements in all samples

$q_{\propto(k,n-k)}$: value looked up from Tukey analysis table, with mean alpha, level of k and n – k

MSW: wrong method within group (determined from ANOVA step)

n_i: number of surveys of 1 group in that total. In case each group has quantity. If there are different ones, choose the smallest one

Step 3: Calculate the test value D:

D is the absolute value of the difference between the two mean values of each group

$D_{12} = |\ \overline{x1} - \overline{x2}\ |$

$D_{23} = |\ \overline{x2} - \overline{x3}|$

$D_{13} = |\ \overline{x1} - \overline{x3}\ |$

Step 4: Testing rule: If Di ≥ T => Reject hypothesis H0

1.2.2. EXAMPLE

Problem: Suppose we want to evaluate the effectiveness of three learning methods A, B and C (traditional learning methods, online learning methods, new learning methods) for improving scores on a math test. We collect data on students' scores (each method takes 5 different random students) after participating in each learning method

	METHOD A	METHOD B	METHOD C
1	8	7.8	8.5
2	7.5	8.5	9
3	8.5	8	8.8
4	9	8.8	8.4
5	8.2	9.2	9.2

Step 1. Determine the hypothesis:

TH1:

H0: μ1= μ2

H1: μ1 ≠ μ2

TH2:

H0: μ2= μ3

H1: μ2≠ μ3

TH3:

H0: µ1= µ3

H1: µ1≠ µ3

In there:

µ1: Average value of group 1

µ2: Average value of group 2

µ3 Average value of group 3

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
METHOD A	5	41.2	8.24	0.313
METHOD B	5	42.3	8.46	0.328
METHOD C	5	43.9	8.78	0.112

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.737333	2	0.368667	1.468792	0.268785	3.885294
Within Groups	3.012	12	0.251			
Total	3.749333	14				

Step 2: From the anova analysis in the levene test example, we can infer:

MSW = 0.251

n = 15

k = 3

$\bar{x}_1 = 8.24$

$\overline{x_2} = 8.46$

$\overline{x_3} = 8.78$

df = 12

α = 0.05

Looking up the Tukey distribution table, we get Q-statistic = 3.773

Because in method A, method B, and method C there are samples of 5, 5, 5 respectively, we find nmin = min{5, 5, 5} = 5

Apply the following formula to calculate the Tukey value:

$$T = 3.773 \times \sqrt{\frac{0.251}{5}} = 0.845$$

Step 3: Calculate the test value D:

D_{12} = | $\overline{x1}$ - $\overline{x2}$ | = 0.22

D_{23} = | $\overline{x2} - \overline{x3}$| = 0.32

D_{13} = | $\overline{x1}$ - $\overline{x3}$ | = 0.54

Step 4: Because

$D_{12} < T =>$ Accept hypothesis H0 (μ1 = μ2)

$D_{23} < T =>$ Accept hypothesis H0 (μ2 = μ3)

$D_{13} < T =>$ Accept hypothesis H0 (μ1 ≠ μ2)

⇨ The results of the three tests using the three learning methods are not too different

CHAPTER 2. ENERGY DRINK SURVEY

Using MS Excel, R language and Python language to perform Chi Square test on the independence of two categorical variables with the data file: *Energy Drink Survey*

Using a 5% significance level test (α=5%), determine if gender and brand preference for energy drinks can be considered independent variables.

Hypotheses:

H0: Gender and brand preference are not dependent on each other.

H1: Gender and brand preference are interdependent.

2.1. ANALYZING BY USING EXCEL

Result of Pivot Table

Count of Respondent	Column Labels			
Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
Female	9	6	22	37
Male	25	17	21	63
Grand Total	**34**	**23**	**43**	**100**

Expected frequency table

EXPECTED VALUE				
Row Labels	**Brand 1**	**Brand 2**	**Brand 3**	**Grand Total**
Female	12.58	8.51	15.91	37
Male	21.42	14.49	27.09	63
Grand Total	34	23	43	100

Test statistics calculate χ^2, p-value and df

X^2				
Row Labels	**Brand 1**	**Brand 2**	**Brand 3**	**Grand Total**
Female	1.018791733	0.7403173	2.3311188	4.0902278
Male	0.598338002	0.4347895	1.3690698	2.40219728
Grand Total	1.617129735	1.1751068	3.7001886	$\boxed{6.49242508}$
df	$\boxed{2}$			
Chi-Square value	$\boxed{5.991464547}$			

2.2. ANALYZING BY USING R

Using **chisq.test(table_name)** to compute Chi – Squared value, df, and p-value

```
> chisq.test(drink_result)

	Pearson's Chi-squared test

data:  drink_result
X-squared = 6.4924, df = 2, p-value = 0.03892
```

Using **qchisq(, df, lower.tail = FALSE)** to calculate the critical value.

```
> qchisq(0.05, 2, lower.tail = FALSE)
[1] 5.991465
```

2.3. ANALYZING BY USING PYTHON

```
c, p, dof, exp = stats.chi2_contingency(chisqt)
```

```
p
```

```
0.038921342064441915
```

```
c
```

```
6.4924250792329055
```

```
dof
```

```
2
```

```
exp
```

```
array([[12.58,  8.51, 15.91],
       [21.42, 14.49, 27.09]])
```

Explanation of values:

c: The Chi-square Test

p: p-value of the test

dof: Degrees of Freedom

expected: Expected of the test

2.4. CONCLUSION

Based on the values:

- The chi-square (χ^2) value is 6.492425079.
- The degree of freedom (df) is 2.
- The p-value is 0.038921342.
- The chi-square critical value is 5.991464547.

With a Chi-square (χ^2) value of 6.492425079 and 2 degrees of freedom (df), and considering a significance level of 0.05, we can reject the null hypothesis (H0) of independence. There is sufficient evidence to conclude that there is a significant association between the categorical variables examined.

With a p-value is 0.03892142 < 0.05 we can reject the null hypothesis(H0).

CHAPTER 3. INSURANCE SURVEY

Using MS Excel, R language and Python language to perform ANOVA with data file (including Levene, ANOVA, Tukey Test): *Insurance survey*

3.1. ANALYZING BY USING EXCEL

3.1.1. LEVENE’S TEST

Identify hypothetical devices:

H0: Satisfaction and education level are not dependent on each other.

H1: Satisfaction and educational level are dependent on each other

We calculate Mean of each group by average() function

	College graduate	Graduate degree	Some college
	5	3	4
	3	4	1
	5	5	4
	3	5	2
	3	5	3
	3	4	4
	3	5	4
	4	5	
	2		
MEAN	3.444444444	4.5	3.142857143

Determine the absolute difference between each score and the corresponding category mean

ABS(MEAN-X)	1.555555556	1.5	0.857142857
	0.444444444	0.5	2.142857143
	1.555555556	0.5	0.857142857
	0.444444444	0.5	1.142857143
	0.444444444	0.5	0.142857143
	0.444444444	0.5	0.857142857
	0.444444444	0.5	0.857142857
	0.555555556	0.5	
	1.444444444		

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	9	7.333333	0.814815	0.280864
Column 2	8	5	0.625	0.125
Column 3	7	6.857143	0.979592	0.356657

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.472744	2	0.236372	0.943358	0.405206	3.4668
Within Groups	5.261855	21	0.250565			
Total	5.734599	23				

We can see that: Because F < F crit (or p-value = 0.40521 > α = 0.05), we accept hypothesis H0 => ANOVA'S TEST

⇨ **Conclusion:** So Satisfaction and Education level do not depend on each other.

3.1.2. ANOVA'S TEST

Determine your hypothesis:

H0: μ1 = μ2 = μ3

H1: There is at least one mean value that is different from the remaining mean values

In there:

μ1: Average value of College graduates

μ2: Average value of Graduate degree

μ3: Average value of Some college

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
College graduate	9	31	3.444444	1.027778		
Graduate degree	8	36	4.5	0.571429		
Some college	7	22	3.142857	1.47619		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	7.878968	2	3.939484	3.924652	0.035635	3.4668
Within Groups	21.07937	21	1.003779			
Total	28.95833	23				

Because F > Fcrit -> Reject hypothesis H0

⇨ **Conclusion:** So there is at least one average value that is different from the remaining values

3.1.3. TUKEY'S TEST

Since we reject H0, we will perform an in-depth ANOVA test to confirm

Specify which group's average is different from which group's average, larger or smaller.

Determine the hypothesis:

Case 1:

H0: $\mu1 = \mu2$

H1: $\mu1 \neq \mu2$

Case 2:

H0: $\mu2 = \mu3$

H1: $\mu2 \neq \mu3$

Case 3:

H0: $\mu1 = \mu3$

H1: $\mu1 \neq \mu3$

Step 1: Calculate Q-statistics

We have: k = 3, df = 21

Looking up the Tukey distribution table, we get Q-statistic = 3.565

Step 2: Calculate the comparison standard T

$$T = 3.565 * \sqrt{\frac{1.00377928949358}{7}} = 1.34998$$

Step 3: Calculate the difference between the 2 groups

MEAN (C - G)	1.05555556
MEAN (G - S)	1.35714286
MEAN (C - S)	0.3015873

Step 4: Compare the difference between the two pairs with T and draw conclusions

College graduate vs Graduate degree < T => $\mu1 = \mu2$

College graduate vs Some college < T => $\mu1 = \mu3$

Graduate degree vs Some college > T => $\mu2 \neq \mu3$

⇨ **Conclusion:** There are two pairs of groups: College Graduate & Graduate

Degree, College Graduate & Some College the mean value is the same, while Graduate Degree & Some College have the mean value different

3.2. ANALYZING BY USING R

3.2.1. LEVENE'S TEST

Identify hypothetical devices:

H0: Satisfaction and education level are not dependent on each other.

H1: Satisfaction and educational level are dependent on each other

```
> leveneTest(Satisfaction., Education, center="mean")
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  2  0.9434 0.4052
      21
```

Because p-value = 0.9434 > α = 0.05, we accept H0.

Therefore, there were no differences in methods between the three groups.

⇨ Qualified to conduct ANOVA test

3.2.2. ANOVA'S TEST

Determine your hypothesis:

H0: $\mu1 = \mu2 = \mu3$

H1: There is at least one mean value that is different from the remaining mean values

In there:

$\mu1$: Average value of College graduates

$\mu2$: Average value of Graduate degree

$\mu3$: Average value of Some college

```
> aov(Satisfaction. ~ Education, data = is)
Call:
   aov(formula = Satisfaction. ~ Education, data = is)

Terms:
                Education Residuals
Sum of Squares   7.878968 21.079365
Deg. of Freedom         2        21

Residual standard error: 1.001888
Estimated effects may be unbalanced
> rs = aov(Satisfaction. ~ Education, data = is)
> summary(rs)
            Df Sum Sq Mean Sq F value Pr(>F)
Education    2  7.879   3.939   3.925 0.0356 *
Residuals   21 21.079   1.004
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> qf(p = 0.05, 2, 21, lower.tail = FALSE)
[1] 3.4668
```

Because p-value = 0.0356 < α = 0.05

⇨ Reject hypothesis H0. So there is at least one average value that is different from the remaining values

3.2.3. TUKEY'S TEST

Because we have rejected the hypothesis H0, we use the Tukey test for analysis more deeply about the superiority and inferiority between group averages, specifically the average of any other group with small groups, which group is larger and smaller

Determine the hypothesis:

Case 1:

H0: $\mu 1 = \mu 2$

H1: $\mu 1 \neq \mu 2$

Case 2:

H0: $\mu 2 = \mu 3$

H1: $\mu 2 \neq \mu 3$

Case 3:

H0: µ1= µ3

H1: µ1≠ µ3

```
> TukeyHSD(rs)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Satisfaction. ~ Education, data = is)

$Education
                                      diff        lwr         upr     p adj
Graduate degree-College graduate  1.0555556 -0.1715336  2.28264475 0.1003252
Some college-College graduate    -0.3015873 -1.5742334  0.97105876 0.8230559
Some college-Graduate degree     -1.3571429 -2.6641246 -0.05016107 0.0409193
```

Because:

College Graduate & Graduate Degree has p-value = 0.1 > α => Accept the hypothesis H0(µ1 = µ2)

College Graduate & Some College has p-value = 0.8 > α => Accept hypothesis H0(µ1 = µ3)

Graduate Degree & Some College has p-value = 0.04 < α => Accept hypothesis H1 (µ1 ≠ µ3)

Conclude: There are two pairs of groups: College Graduate & Graduate Degree, College Graduate & Some College the mean value is the same, while Graduate Degree & Some College have the mean value different

3.3. ANALYZING BY USING PYTHON

3.3.1. LEVENE'S TEST

Identify hypothetical devices:

H0: Satisfaction and education level are not dependent on each other.

H1: Satisfaction and educational level are dependent on each other

```
[ ] from scipy.stats import levene

    stat, p = levene(*df_gr, center='mean')
    stat, p
```

```
(0.9433580072525427, 0.40520616699352924)
```

Because p-value = 0.405 > $\alpha = 0.05$, we accept H0.

Conclusion: So there is no difference in the variance of the 3 populations.

⇨ Qualified to conduct ANOVA test

3.3.2. ANOVA'S TEST

Determine your hypothesis:

H0: $\mu 1 = \mu 2 = \mu 3$

H1: There is at least one mean value that is different from the remaining mean values

In there:

$\mu 1$: Average value of College graduates

$\mu 2$: Average value of Graduate degree

$\mu 3$: Average value of Some college

```
[ ] from scipy.stats import f_oneway

    fvalue, pvalue = f_oneway(*df_gr)
    fvalue, pvalue
```

```
(3.9246517319277117, 0.03563539756488997)
```

Because p-value = 0.0356 < $\alpha = 0.05$

⇨ Reject hypothesis H0. So there is at least one mean value that is different from the mean values remaining

3.3.3. TUKEY'S TEST

Because we have rejected the hypothesis H0, we use the Tukey test for analysis more deeply about the superiority and inferiority between group averages, specifically the average of any other group with small groups, which group is larger and smaller

Determine the hypothesis:

Case 1:

H0: $\mu1 = \mu2$

H1: $\mu1 \neq \mu2$

Case 2:

H0: $\mu2 = \mu3$

H1: $\mu2 \neq \mu3$

Case 3:

H0: $\mu1 = \mu3$

H1: $\mu1 \neq \mu3$

```
[ ]  from statsmodels.stats.multicomp import pairwise_tukeyhsd
     tukey = pairwise_tukeyhsd(endog=df['Satisfaction* '],groups=df.Education,alpha=0.05)
     print(tukey)
```

```
          Multiple Comparison of Means - Tukey HSD, FWER=0.05
=======================================================================
     group1            group2      meandiff p-adj   lower   upper  reject
-----------------------------------------------------------------------
College graduate Graduate degree    1.0556 0.1003 -0.1715  2.2826  False
College graduate     Some college  -0.3016 0.8231 -1.5742  0.9711  False
 Graduate degree     Some college  -1.3571 0.0409 -2.6641 -0.0502   True
-----------------------------------------------------------------------
```

Because:

College Graduate & Graduate Degree has p-value = 0.1 > α => Accept the hypothesis H0($\mu1 = \mu2$)

College Graduate & Some College has p-value = 0.8 > α => Accept hypothesis H0($\mu1 = \mu3$)

Graduate Degree & Some College has p-value = 0.04 < α => Accept hypothesis H1 ($\mu1 \neq \mu3$)

Conclude: There are two pairs of groups: College Graduate & Graduate Degree, College Graduate & Some College the mean value is the same, while Graduate Degree & Some College have the mean value different

CHAPTER 4. VIETNAM NATIONAL HIGHSCHOOL EXAM SCORE 2018

4.1. ANALYZING BY USING R

4.1.1. ANOVA'S TEST

Determine your hypothesis:

H0: $\mu 1 = \mu 2 = \mu 3$

H1: There is at least one mean value that is different from the remaining mean values

In which:

$\mu 1$: Average value of Block A

$\mu 2$: Average value of Block B

μ3: Average value of Block C

```
> rs = aov(Diem~Khoi, data = exam_score)
> rs
Call:
   aov(formula = Diem ~ Khoi, data = exam_score)

Terms:
                    Khoi  Residuals
Sum of Squares    152724   12073340
Deg. of Freedom        2    1073350

Residual standard error: 3.353845
Estimated effects may be unbalanced
> summary(rs)
                 Df   Sum Sq Mean Sq F value Pr(>F)    
Khoi              2   152724   76362    6789 <2e-16 ***
Residuals   1073350 12073340      11                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Because p-value = 2.e-16 < α = 0.05

⇨ Reject hypothesis H0. Therefore, there is at least one mean value that is different from the remaining values

4.2. ANALYZING BY USING PYTHON

4.2.1. ANOVA'S TEST

Determine the hypothesis:

Case 1:

H0: μ1= μ2

H1: μ1 ≠ μ2

Case 2:

H0: μ2= μ3

H1: μ2≠ μ3

Case 3:

H0: μ1= μ3

H1: μ1≠ μ3

In which:

μ1: Average value of Block A

μ2: Average value of Block B

μ3: Average value of Block C

```
[ ] from scipy.stats import f_oneway

    fvalue, pvalue = f_oneway(*repaired_df)
    fvalue, pvalue
```

```
(6788.78521446278, 0.0)
```

The ANOVA test results showed that there were significant differences between the KhoiA, KhoiB and KhoiD groups.

The F-Statistic value (6788.785) is significantly large, meaning that there is a significant difference between groups.

The p-value (0.0) is very small, lower than the significance level of 0.05, showing that there is enough evidence to reject the hypothesis of no difference between groups.

REFERENCES

[1] N.M.Nhut, "LAB02. INFERENTIAL STATISTICS" - IS403 – Business data analysis, UIT, 2023

[2] N.Đ.Thuan, "Chapter 7: Statistical Inference" - Business data analysis, UIT, 2023

[3] https://real-statistics.com/statistics-tables/studentized-range-q-table/

[4] https://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm

LINK_CANVA