SOUTH-WEST UNIVERSITY "NEOFIT RILSKI" - BLAGOEVGRAD

FACULTY OF NATURAL SCIENCES AND MATHEMATICS
Information Systems and Technologies

COURSEWORK

Subject: **Specialized Statistical Software**

Topic: ***Caffeine Consumption and Its Effects***

Prepared by:
Kamelija Aleksovska, 21251421002
Nevenka Lazarevska, 21251421001

Blagoevgrad, 2024

CONTENTS

I. STUDY DATA

1. **Topic:** Caffeine consumption and its effects
2. **Research target:** Students
3. **Sample:** Non-representative
4. **Sample size:** 100 people
5. **Method of data collection:** Questionnaire
6. **Research period:** 14.10.2024 – 16.10.2024

II. INTRODUCTION TO THE STUDY

The questionnaire consists of 17 questions on the topic of caffeine and caffeinated drinks. It was distributed to 100 students to gather information about their consumption habits, effects, and opinions. The questions vary in type, enabling a wide range of responses. The survey aimed to identify connections between coffee consumption, its purpose, student life, academic performance, and common side effects. The gathered data allowed us to present the topic in a more informative and engaging manner.

III. RESEARCH

a. Descriptive Statistics

The survey was presented to 100 students. Of the respondents, 49 are male, and 51 are female, as shown in the table below.

Question: **“Which is your gender?”**

Какъв е вашият пол?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Мъж	49	49,0	49,0	49,0
	Жена	51	51,0	51,0	100,0
	Total	100	100,0	100,0	

Figure 1 Table for gender

Through **Transform > Recode into Different Variables**, we coded the values for the age of the students into several categories: 18-20 years, 21-23 years, 24-27 years, 28-30 years, and 32-36 years. After that, using the tool **Analyze > Descriptive Statistics > Frequencies**, we selected **Charts > Histogram** to present the age of the students through a histogram and the corresponding frequency table.

Question: **“ What is your age group?”**

Коя е вашата възрастна група?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18-20	12	12,0	12,0	12,0
	21-23	54	54,0	54,0	66,0
	24-27	26	26,0	26,0	92,0
	28-30	6	6,0	6,0	98,0
	32-36	2	2,0	2,0	100,0
	Total	100	100,0	100,0	

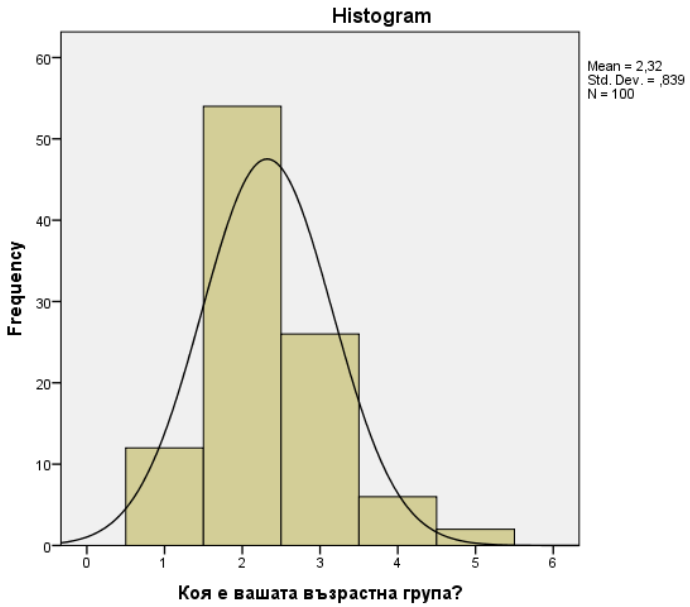

Figure 2 Table and histogram for the age of the students

Through **Data > Select Cases** we have selected the male students from the study, after which the menus **Analyze > Descriptive Statistics > Descriptives** and **Analyze > Descriptive Statistics > Frequencies** were used to present the table with the overall numerical characteristics for the selected students.

The table shows that the average age of the surveyed male students is 22.29. The exact mean is 22 and the most common age is 22 years. The standard deviation is 1.768 from the mean. The youngest surveyed student is 18 years old, while the oldest is 26 years old.

Question: **“What is your age?”**

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Каква е вашата възраст?	49	18	26	22,29	1,768	3,125
Valid N (listwise)	49					

Frequencies

Statistics

Каква е вашата възраст?

N	Valid	49
	Missing	0
Mean		22,29
Median		22,00
Mode		22
Std. Deviation		1,768
Variance		3,125
Minimum		18
Maximum		26

Figure 3 Tables for the Age of Male Students

Through **Data > Select Cases** we have selected the female students from the study, after which the menu **Analyze > Descriptive Statistics > Frequencies** was used to present the table with the overall numerical characteristics for the selected students. The table shows that the average age of the surveyed female students is 23.49. The exact mean is 22, but the most common age is 21 years. The standard deviation is 3.563 from the mean. The youngest surveyed student is 18 years old, while the oldest is 32 years old.

Question: **“What is your age?”**

Frequencies

Statistics

Каква е вашата възраст?

N	Valid	51
	Missing	0
Mean		23,49
Median		22,00
Mode		21
Std. Deviation		3,563
Minimum		18
Maximum		32
Sum		1198

Figure 4 Table for the Age of Female Students

Through **Analyze > Descriptive Statistics > Frequencies** we have shown a Pie Chart for the motivation for drinking coffee and coffee beverages among students. The largest portion, i.e., 46, responded that they consume these beverages as a result of habit, while the next most common reason is fatigue, which was reported by 24 students.

Question: **“What is your motivation for drinking coffee?”**

Коя е вашата мотивация за пиене на кафе?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Умора	24	24,0	24,0	24,0
	Социализиране	14	14,0	14,0	38,0
	Учене	6	6,0	6,0	44,0
	Навик	46	46,0	46,0	90,0
	Друго	10	10,0	10,0	100,0
	Total	100	100,0	100,0	

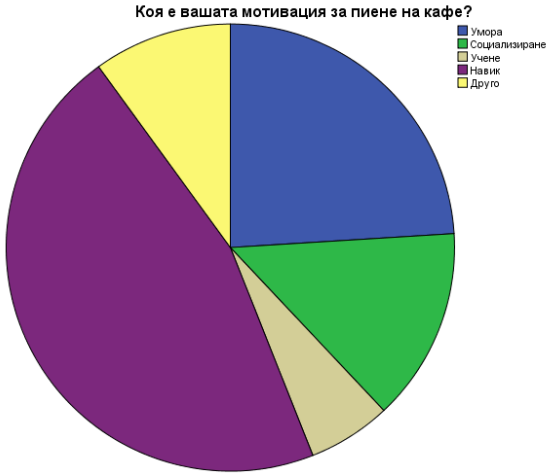

Figure 4 Table of Motivations for Coffee Consumption

Through **Analyze > Descriptive Statistics > Frequencies**, we have shown the frequency table for the opinion of students on whether there is a difference in their academic performance after coffee consumption, where approximately half of the respondents answered positively, i.e., 48, and 52 negatively.

Question: **“Do you think that there is a difference in your academic performance after drinking coffee?”**

Frequencies

Statistics

Дали смятате че има разлика във вашето академично представяне след пиене кафе?

N	Valid	100
	Missing	0

Дали смятате че има разлика във вашето академично представяне след пиене кафе?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Да	48	48,0	48,0	48,0
	Не	52	52,0	52,0	100,0
	Total	100	100,0	100,0	

Figure 5 Table of Academic Performance

Through **Analyze > Descriptive Statistics > Frequencies**, we have shown the frequency table for the side effects of consumption. With the highest frequency, out of 40 students, the response "Sometimes" is encountered, 26 have answered "Rarely," 24 "No, never," and the fewest students, 10, have answered "Yes, often." The option for Bar Chart has been used for the visual representation of the results.

Question: **“ Have you ever had side effects from coffee consumption?”**

Frequencies

Statistics

Дали сте имали някакви странични ефекти от консумация на кафе?

N	Valid	100
	Missing	0

Дали сте имали някакви странични ефекти от консумация на кафе?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Да, често	10	10,0	10,0	10,0
	Понякога	40	40,0	40,0	50,0
	Рядко	26	26,0	26,0	76,0
	Не,никога	24	24,0	24,0	100,0
	Total	100	100,0	100,0	

Figure 6 Table and for frequency of side effects

b. Crosstabs

With the help of the tool **Analyze > Descriptive Statistics > Crosstabs**, a cross table is obtained in which the statistics of male and female students from different faculties are presented, showing 9 faculties in which the surveyed students are studying.

На кой факултет се обучавате? * Какъв е вашият пол? Crosstabulation

Count

		Какъв е вашият пол?		
		Мъж	Жена	Total
На кой факултет се обучавате?	Природо-математически факултет	7	9	16
	Правно-исторически факултет	8	10	18
	Стопански факултет	6	10	16
	Философски факултет	4	8	12
	Факултет по педагогика	6	2	8
	Факултет по изкуствата	2	0	2
	Технически факултет	6	6	12
	Факултет Обществено здраве, здравни грижи и спорт	10	6	16
Total		49	51	100

Figure 7 Crosstab for the faculties and gender of the students

With the help of the tool **Analyze > Descriptive Statistics > Crosstabs**, another crosstab is obtained that presents the faculties and possible reasons for coffee consumption among students from the respective faculties.

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
На кой факултет се обучавате? * Коя е вашата мотивация за пиене на кафе?	100	100,0%	0	0,0%	100	100,0%

На кой факултет се обучавате? * Коя е вашата мотивация за пиене на кафе? Crosstabulation

Count

		Коя е вашата мотивация за пиене на кафе?					
		Умора	Социализиране	Учене	Навик	Друго	Total
На кой факултет се обучавате?	Природо-математически факултет	4	2	0	10	0	16
	Правно-исторически факултет	6	2	2	4	4	18
	Стопански факултет	6	2	0	6	2	16
	Философски факултет	4	0	0	8	0	12
	Факултет по педагогика	2	2	0	4	0	8
	Факултет по изкуствата	0	0	0	2	0	2
	Технически факултет	2	0	0	8	2	12
	Факултет Обществено здраве, здравни грижи и спорт	0	6	4	4	2	16
Total		24	14	6	46	10	100

Figure 8 Crosstab for the faculties and the motivation of the students

c. Chi-Square Tests

With the help of the tool **Analyze > Descriptive Statistics > Crosstabs**, the relationship between two categorical characteristics is checked, and it turns out that after testing the hypotheses, we obtain the result that there is no statistically significant dependence between.

Gender and academic performance after coffee consumption since the asymptotic significance is **0.553**, i.e., it is greater than 0.05 and in this case, the null hypothesis is accepted, i.e., H0.

Какъв е вашият пол? * Дали смятате че има разлика във вашето академично представяне след пиене кафе? Crosstabulation

			Дали смятате че има разлика във вашето академично представяне след пиене кафе?		
			Да	Не	Total
Какъв е вашият пол?	Мъж	Count	25	24	49
		Expected Count	23,5	25,5	49,0
	Жена	Count	23	28	51
		Expected Count	24,5	26,5	51,0
Total		Count	48	52	100
		Expected Count	48,0	52,0	100,0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,351[a]	1	,553		
Continuity Correction[b]	,154	1	,695		
Likelihood Ratio	,351	1	,553		
Fisher's Exact Test				,689	,347
Linear-by-Linear Association	,348	1	,555		
N of Valid Cases	100				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 23,52.

b. Computed only for a 2x2 table

Therefore, since the value of **Chi Square** shows that there is no significant relationship between gender and academic performance after drinking coffee, the values of **Cramer's V** and **Phi** also show that there is no relationship between them because they are approximately 0, i.e., **0.059**.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,059	,553
	Cramer's V	,059	,553
N of Valid Cases		100	

Figure 9 Chi-Square Test

d. Cramer's V

Again, we use the same menu to show the value of Cramer's number. Compared to the previous example, for the relationship between the faculty and coffee consumption while studying, **Cramer's number** is **0.342**, which is greater than the previous example, meaning that it is a stronger relationship, but still not significant, since the number is close to 0.

На кой факултет се обучавате? * Дали докато учите пиете кафе? Crosstabulation

			Дали докато учите пиете кафе?			
			Да	Не	Понякога	Total
На кой факултет се обучавате?	Природо-математически факултет	Count	6	4	6	16
		Expected Count	5,8	3,2	7,0	16,0
	Правно-исторически факултет	Count	0	6	12	18
		Expected Count	6,5	3,6	7,9	18,0
	Стопански факултет	Count	8	2	6	16
		Expected Count	5,8	3,2	7,0	16,0
	Философски факултет	Count	6	0	6	12
		Expected Count	4,3	2,4	5,3	12,0
	Факултет по педагогика	Count	2	2	4	8
		Expected Count	2,9	1,6	3,5	8,0
	Факултет по изкуствата	Count	0	0	2	2
		Expected Count	,7	,4	,9	2,0
	Технически факултет	Count	6	4	2	12
		Expected Count	4,3	2,4	5,3	12,0
	Факултет Обществено здраве, здравни грижи и спорт	Count	8	2	6	16
		Expected Count	5,8	3,2	7,0	16,0
Total		Count	36	20	44	100
		Expected Count	36,0	20,0	44,0	100,0

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,484	,054
	Cramer's V	,342	,054
N of Valid Cases		100	

Figure 10 Chi-Square Test and Cramer V

e. Paired Samples T Test

The **Paired Samples Correlation** shows a strong positive relationship between the two measurements (**r = 0.884**, p < 0.001), meaning that students' sleep durations before and after coffee are highly related.

However, the **Paired Samples T-Test** result shows **p = 1.000**, which is much greater than **0.05**. This indicates that there is no statistically significant difference in the mean hours of sleep before and after caffeine consumption. In other words, students generally maintained the same average sleep duration regardless of coffee intake.

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Колко часа спите без консумация на кофеин?	6,70	100	1,685	,168
	Колко часа спите след консумация на кофеин?	6,70	100	1,920	,192

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Колко часа спите без консумация на кофеин? & Колко часа спите след консумация на кофеин?	100	,884	,000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Колко часа спите без консумация на кофеин? - Колко часа спите след консумация на кофеин?	,000	,899	,090	-,178	,178	,000	99	1,000

Figure 11 Paired Samples T-Test

f. Independent-Samples T-Test

Using the **Analyze → Compare Means → Independent-Samples T-Test** tool, an independent samples t-test was conducted. The analysis examined the criterion **age between male and female students**. Since the Significance (p-value) is 0.000, which is less than 0.05, we conclude that there is a statistically significant difference between the variances (**Levene's Test**).

In this case, the value for Sig. (2-tailed) is **0.035**, which is also less than 0.05, indicating that the difference is statistically significant. Therefore, we can conclude that there is a significant statistical difference in terms of age between the two genders.

T-Test

[DataSet1] C:\Users\User\Desktop\Kamelija\Coffee_Drinking_Statistics.sav

Group Statistics

	Какъв е вашият пол?	N	Mean	Std. Deviation	Std. Error Mean
Каква е вашата възраст?	Мъж	49	22,29	1,768	,253
	Жена	51	23,49	3,563	,499

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Каква е вашата възраст?	Equal variances assumed	21,733	,000	-2,128	98	,036	-1,204	,566	-2,328	-,081
	Equal variances not assumed			-2,154	73,853	,035	-1,204	,559	-2,319	-,090

Figure 12 Independent-Samples T-Test

Using the same test, an analysis was performed for the criterion **amount of caffeine consumed by two independent groups of students**. The Significance (p-value) is **0.539**, which is greater than 0.05, indicating that there is no statistically significant difference between the variances. The Sig. (2-tailed) value is **0.737**, which is also greater than 0.05. Therefore, the conclusion is that there is **no statistically significant difference** between the two student groups regarding caffeine consumption.

T-Test

[DataSet1] C:\Users\User\Desktop\Kamelija\Coffee_Drinking_Statistics.sav

Group Statistics

	Дали смятате че има разлика във вашето академично представяне следлнение кафе?	N	Mean	Std. Deviation	Std. Error Mean
Колко чаши най-често консумирате през деня?	Да	48	2,33	,953	,138
	Не	52	2,27	,952	,132

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Колко чаши най-често консумирате през деня?	Equal variances assumed	,380	,539	,336	98	,737	,064	,191	-,314	,442
	Equal variances not assumed			,336	97,352	,737	,064	,191	-,314	,442

Figure 13 Independent-Samples T-Test

g. ANOVA

With the tool **Analyze > Compare Means > One-Way ANOVA**, the relationship between the **number of cups of coffee consumed** (quantitative variable) and the **perceived influence on concentration** (categorical factor) was analyzed.

The ANOVA results show:

Between-group variance = 5.441
Within-group variance = 14.599
Significance (Sig.) = 0.000 (< 0.05)

This means there is a **statistically significant difference in the number of cups of coffee consumed among groups with different opinions on coffee's effect on concentration**. However, **post-hoc tests were not performed**, so the specific group differences are not identified.

The nonparametric **Kruskal-Wallis test** confirms this result with **p = 0.000**, supporting the conclusion of a significant difference between the groups.

Oneway

Descriptives

Колко чаши най-често консумирате през деня?

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Много положително	14	2,71	,469	,125	2,44	2,98	2	3
Положително	62	2,06	,439	,056	1,95	2,18	1	3
Отрицателно	12	2,00	,000	,000	2,00	2,00	2	2
Няма ефект	12	2,00	,000	,000	2,00	2,00	2	2
Total	100	2,14	,450	,045	2,05	2,23	1	3

ANOVA

Колко чаши най-често консумирате през деня?

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	5,441	3	1,814	11,926	,000
Within Groups	14,599	96	,152		
Total	20,040	99			

Figure 14 ANOVA

Nonparametric Tests

```
[DataSet1] C:\Users\User\Desktop\Kamelija\Coffee_Drinking_Statistics.sav
```

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Колко чаши най-често консумирате през деня? is the same across categories of Как смятате че кофеинът влияе на фокуса?.	Independent-Samples Kruskal-Wallis Test	,000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Figure 15 Nonparametric Tests

h. Nonparametric Tests (Independent Samples)

With the tool **Analyze > Nonparametric Tests > Independent Samples**, a study of categorical and quantitative characteristics was conducted. In the test conducted, a study of a categorical characteristic (**the gender of the students**) is placed in the group field and a quantitative characteristic (**the age at which students drink coffee**).

For this purpose, **the Mann Whitney U Test** (used for comparing two independent groups) was used. From the study, it is found that the value of the significant label

Sig. is **0.901**, which is greater than 0.05, and the final decision is to accept the null hypothesis that there is no significant statistical difference between the groups.

Nonparametric Tests

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of От коя възраст пиете кафе? is the same across categories of Какъв е вашият пол?.	Independent-Samples Mann-Whitney U Test	,901	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

Figure 16 Nonparametric Tests

i. Regression Analysis

With the tool **Analyze-> Regression -> Linear**, a regression analysis of the data was performed where the dependent variable refers to the hours of sleep after coffee consumption (quantitative characteristic), and the independent variable is the number of cups of coffee consumed during the day (quantitative characteristic). The ANOVA table shows the overall significance of the model. The value of **Sig.= 0.546** is greater than 0.05, which means that the regression model is not statistically significant. In the coefficients table,

The coefficient for the predictor **"How many cups of coffee do you consume per day?"** is **0.124**, which means that with each additional cup of coffee, the hours of sleep slightly increase (on average by 0.124 hours).

However, this relationship is statistically insignificant because the value of **Sig.= 0.546**. Thus, the presented graph shows us that there is no dependency between the two characteristics, i.e., if one increases or decreases, there will be no changes in the other. From the graph, we can conclude that the linear dependency is very weak and there is no clear trend in the data.

Model Summary and Parameter Estimates

Dependent Variable: Колко часа спите след консумация на кофеин?

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	,004	,366	1	98	,546	6,416	,124

The independent variable is Колко чаши най-често консумирате през деня?.

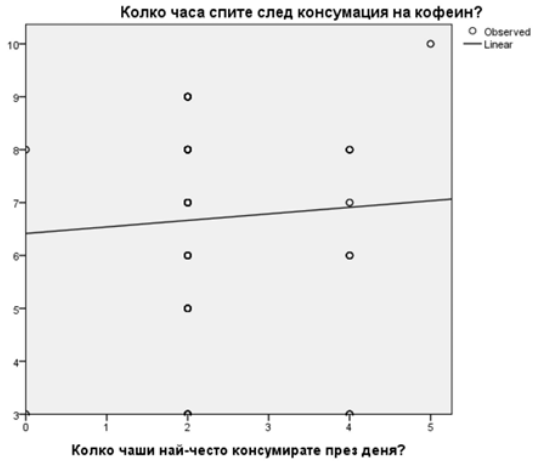

Figure 17 Regression Analysis

With the tool **Analyze -> Correlate > Bivariate**, an analysis of the same data was made and the value of Sig. = 0.546.

Correlations

Correlations

		Колко чаши най-често консумирате през деня?	Колко часа спите след консумация на кофеин?
Колко чаши най-често консумирате през деня?	Pearson Correlation	1	,061
	Sig. (2-tailed)		,546
	N	100	100
Колко часа спите след консумация на кофеин?	Pearson Correlation	,061	1
	Sig. (2-tailed)	,546	
	N	100	100

Figure 18 Correlations

IV. SUMMARY

Gender of respondents: 49% men and 51% women.

Age: Students are categorized into groups: 18-20, 21-23, 24-27, 28-30, 32-36 years.

Average age by gender:

Men: Average age 22.29, most common 22, range 18-26, standard deviation 1.768.
Women: Average age 23.49, most common 21, range 18-32, standard deviation 3.563.

A statistically significant difference is observed in the age characteristics of men and women (**Independent-Samples T-Test**). Women have a higher average age, and the difference is significant (**0.035** < 0.05).

Motivation for coffee consumption: 46% indicate habit, 24% fatigue, 14% socializing, 6% studying, and 10% for some other reason.

Academic performance after coffee consumption: 48% believe in improvement, 52% do not notice a difference. The **Chi-Square test** showed that there is no statistically significant relationship between the gender of the student and the perceived influence of coffee on their performance **(Sig. = 0.553** > 0.05).

Side effects from coffee: Most students report side effects only sometimes (40%), 26% rarely, 24% never, and 10% often, which indicates that these effects are not frequent and do not pose a serious problem for the respondents.

Coffee Consumption and Faculty (Crosstabs): There is no statistically significant relationship between the faculty of the students and their coffee consumption (**Cramer's V = 0.342** i.e. Cramer's V ≈ 0).

Correlation Between Hours of Sleep Before and After Coffee Consumption (Paired-Samples T-Test): A result of strong positive correlation **0.884** is obtained, with significance **0.000**, indicating that the relationship is statistically significant. However, the value of **P=1.000** in the **Paired Samples Test** table is greater than 0.05, which means that there is no significant statistical difference between the parameters (variances).

Correlation Between Academic Performance and Number of Cups of Coffee (Independent Samples T-Test): The result for **Sig.=0.539** indicates that there is no significant statistical difference between the groups and therefore **Sig.=0.737** again points to the acceptance of the null hypothesis of absence of a statistically significant difference.

Number of Cups of Coffee and Concentration (ANOVA): The number of cups of coffee that students consume shows a significant statistical dependence with the perceived influence on concentration (**Sig.=0.000** < 0.05).

Differences Regarding the Age of Starting to Drink Coffee (Nonparametric Test): There is no statistically significant difference between men and women regarding the age at which they start drinking coffee **(Sig.=0.901** > 0.05).

Amount of Coffee and Duration of Sleep (Regression Analysis): The regression model shows that there is no statistically significant relationship between the number of cups of coffee consumed during the day and the number of hours of sleep **(Sig.=0.546** > 0.05).

V. CONCLUSION

The study summarizes key aspects of student behavior and perceptions regarding coffee consumption, providing the following evidence and conclusions:

1. **Coffee is a fundamental part of students' daily lives:**

   The main motivation for consumption is habit (46%), followed by the need to overcome fatigue (24%). This highlights the social and functional nature of coffee drinking in the student population.

2. **Side effects and impact on concentration:**

   Side effects occur rarely or moderately often, suggesting that coffee does not pose a significant health problem for most students.
   According to the results from ANOVA, the number of cups of coffee is statistically significantly related to improvement in concentration, which may explain why students perceive coffee as a study aid.

3. **Gender differences:**

   Age analysis shows statistically significant differences between men and women, with women having a higher average age. However, there are no statistically significant differences between genders regarding the age at which they start drinking coffee.

4. **Academic performance and gender:**

   The lack of a statistically significant relationship between gender and perceived academic performance after coffee consumption (Chi-Square Test) shows that this factor is not related to differences in gender identity.

5. **Amount of Coffee Drinks and Sleep:**

   According to the regression analysis, it can be concluded that coffee consumption does not have a significant impact on the duration of sleep.
   There is a strong correlation between hours of sleep without and after coffee consumption, and there is no statistically significant difference between the variances.