Course: Data Analysis (task № 2)

Students' Name and Surname ____Rashid Ali_____

Please answer the questions below.

1. Open **General Social Survey.sav** file. Create a contingency table between two variables **happy** and **occcat80**.

(a) How many people who are working in "Farming, Forest, and Fishing" are Pretty Happy?
_____25_____

(b) What is the percentage of people feeling themselves Not Too Happy out of those who are working in "Precision Production, Craft, and Repair"?
_____7.4074%_____
(c) What is the percentage of people working in "Managerial and Professional Specialty" out of those who feel themselves Very Happy?

_____29.7%_____

2. Open **General Social Survey.sav** file. Select **two different pairs** of categorical variables to run Chi-square statistical test. Formulate hypotheses. Interpret the results of analysis. Make conclusions.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26.964a	10	.003
Likelihood Ratio	26.854	10	.003
Linear-by-Linear Association	11.779	1	.001
N of Valid Cases	1407		

a. 1 cells (5.6%) have expected count less than 5. The minimum expected count is 3.81.

H0= There is a no relationship between happiness (General Happiness) and occupation (Occupational Category) of people.

H1= There is a significant relationship between happiness (General Happiness) and occupation (Occupational Category) of people.

So, there is only 0.3 % chance of making mistake while rejecting H0.We reject H0 and accept H1.
So, there is a statically significant relationship between happiness (General Happiness) and occupation (Occupational Category).

3. Open **party.sav** file. Create a frequency table between a multiple-choice question *"What prevents your participation in political party work"* which is represented in the database by the following variables: mit1, mit2, mit3, mit4, mit5, and a variable *sex*. Answer the question below:

(a) How many males *don't want to take part* in a political party's work? _____10_____

(b) How many people think that they have *too little political experience*? _____29_____

(c) What is the % of females from those who *think that functions are already distributed*? _____11.5%_____

(d) What is the % of males from all males who don't know about the opportunities to participate in a political party's work? _____12.1%_____

(e) What is the % of answers "I have too little political experience" given by females from the whole sum of answers given by females? _____23.6%_____

4. Open **world_data.sav** file. Analyze **lit_male** variable.

- Run the Kolmogorov-Smirnov test to conclude whether the distribution of all the values if **lit_male** variable is significantly different from the normal. Formulate hypothesis. Make conclusions.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Males who read (%)	.169	85	.000	.880	85	.000

a. Lilliefors Significance Correction

H0: The distribution of lit_males (males who read) is normal.

H1: The distribution of lit_males (males who read) is not normal.

Sig=0,000<.05

So, we reject H0 and accept H1.It's not normal distribution.

5. Correlation
What sort of correlation would you expect to see from the following variables?
(a) Fuel bills and temperature;
_____It must be very Strong _____
(b) ice-cream sales and temperature;
_____It can be week or strong depends on people _____
(c) Number of counselling sessions and gender.
_____It can be week_____

6. Independent and dependent variables
Identify the independent and dependent variables in the following research
Questions:
(a) Does alcohol affect a person's ability to calculate mathematical problems?
_____ Alcohol (independent) and person's Ability (dependent) _____
(b) Is acupuncture better than physiotherapy in treating back pain?
___Acupuncture and Physiotherapy (both independent), Pain (dependent) _____

7. Do the bivariate correlation analysis with the use of data from **General Social Survey.sav**
file. Calculate Pearson's, Kendall's and Spearman's correlation coefficients. Evaluate the
significance of the coefficients. Indicate the coefficient of determination for every coefficient.
Interpret the results.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Respondent's Sex	.383	1510	.000	.627	1510	.000
Highest Year of School Completed	.160	1510	.000	.964	1510	.000

a. Lilliefors Significance Correction

As sig=0, 00<.05 therefore, these variables are not normally distributed and we can calculate

Pearson correlation for these variables.

We can use nonparametric correlations.

Nonparametric Correlations

Correlations

			Respondent's Sex	Highest Year of School Completed
Kendall's tau_b	Respondent's Sex	Correlation Coefficient	1.000	-.089**
		Sig. (2-tailed)	.	.000
		N	1517	1510
	Highest Year of School Completed	Correlation Coefficient	-.089**	1.000
		Sig. (2-tailed)	.000	.
		N	1510	1510

**. Correlation is significant at the 0.01 level (2-tailed).

Sig=0.000074<0.05

So, there is not statically significant relationship between Sex (Respondent's Sex) and Edu (Highest Year of School Completed).