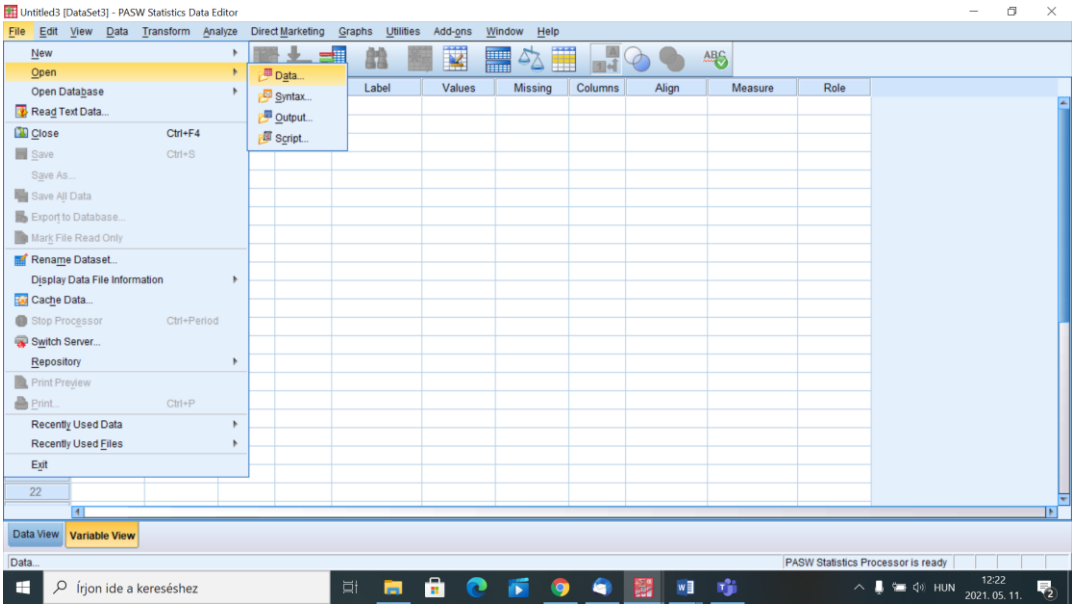
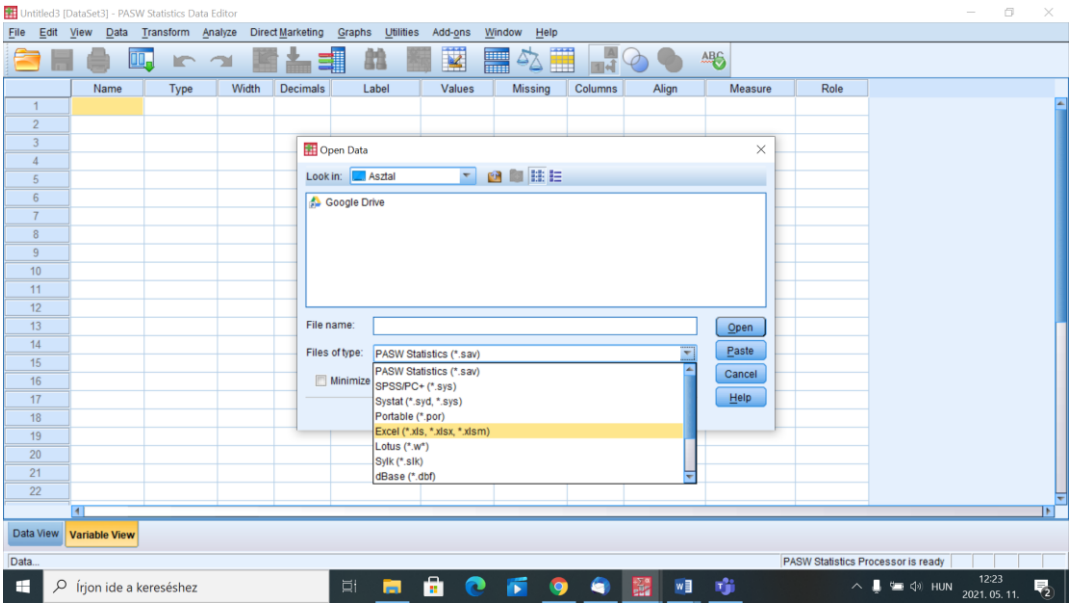
ENTER DATA INTO SPSS – SPSS GUIDELINE

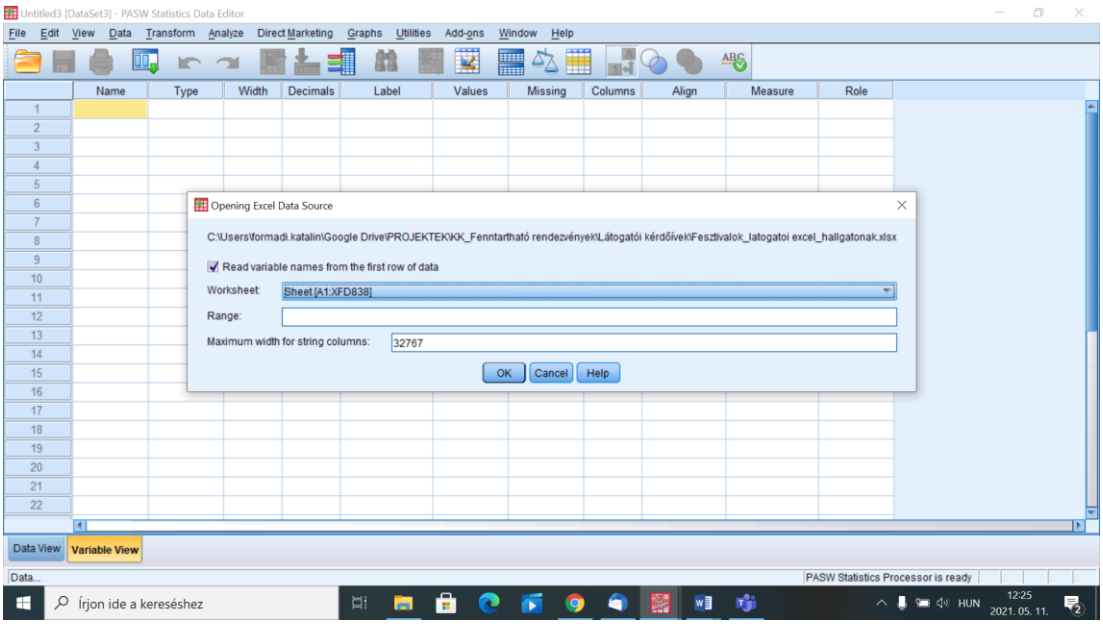
How to open the data from other sources (excel) into SPSS data?

- The excel sheet must be closed to open in SPSS.
- First open SPSS
- File → Open → Data





- Now the type of file is SPSS Statistics (.sav) →change the file types to excel (xls) or ALL TYPES than you can choose other options (excel, dBase etc.)



- Read variables names in the work sheet data – leave the sheet 1 and say ok

FREQUENCY AND DESCRIPTIVE ANALYSIS

SPSS for Beginners 2: Frequency Counts and Descriptive Statistics (10 min)

https://www.youtube.com/watch?v=4CWeHF3Mn00

ANALYZE → DESCRIPTIVES → FREQUENCIES

BAR CHARTS % , SHOW NORMAL CURVES

STATISTICS: STD. DEVIATION, VARIANCE, MIN, MAX, S.E. MEAN

ANALYZE → DESCRIPTIVES → DESCRIPTIVES

use histogram

STATISTICS: mean, min., max, St. Deviation

TYPE OF DATA	**STATISTICS**	**GRAPHS**
NOMINAL	FREQUENCY	BAR CHARTS
SCALE	MEAN, ST. DEVIATION	HISTOGRAM

ANALYZE → DESCRIPTIVES → EXPLORE

weights and heights DEPENDENT – CLICK ON STATISTICS
displays: click on both

statistics: CONFIDENTIAL 95%
click confidence interval for mean: 95%, click on also outliers, percentiles and continue

plots: choose also histograms +stem and leaf stay
click on normality plots with tests

SPLIT BY GENDER

ANALYZE → DESCRIPTIVES → EXPLORE

HEIGHTS AND WEIGHTS STAY

FACTOR LIST – MOVE GENDER TO FACTOR LIST

DESCRIPTIVES: BASIC MEANS

FREQUENCY: FLEXIBILITY TO CHOOSE

EXPLORE: VERY DETAILED SPLIT ANALYSIS – GIVING YOU ALL THE DATA INFORMATION

HOW TO STANDARIZE THE VALUES?

Standardization comes into picture when features of input data set have large differences between their ranges.

ANALYZE → DESCRIPTIVES → DESCRIPTIVES

move gender, leave heights and weights

save standardized value as (click on that box)

Z-score is one of the most popular methods to standardize data, and can be done by subtracting the mean and dividing by the standard deviation for each value of each feature.

$$z = \frac{value - mean}{standard\ deviation}$$

Once the standardization is done, all the features will have a mean of zero, a standard deviation of one, and thus, the same scale.

Z new variables Z heights and Z weights

Z score: how far the raw score differs from the mean in standard deviation units

- negative (-) Z score means always below the average
- positive (+) Z score means always above the average

Other tutorial videos to watch:

Descriptive Statistics and Z scores in SPSS – SPSS for beginners

https://www.youtube.com/watch?v=99fGYHGyO5U

How to explore relationship between variables in SPSS?

1) Categorical vs. categorical variables (nominal or ordinal)

Eg. Relationship between gender and education

Education/ Gender	*Primary education*	*Secondary education*	*Higher education*
Male			
Female			

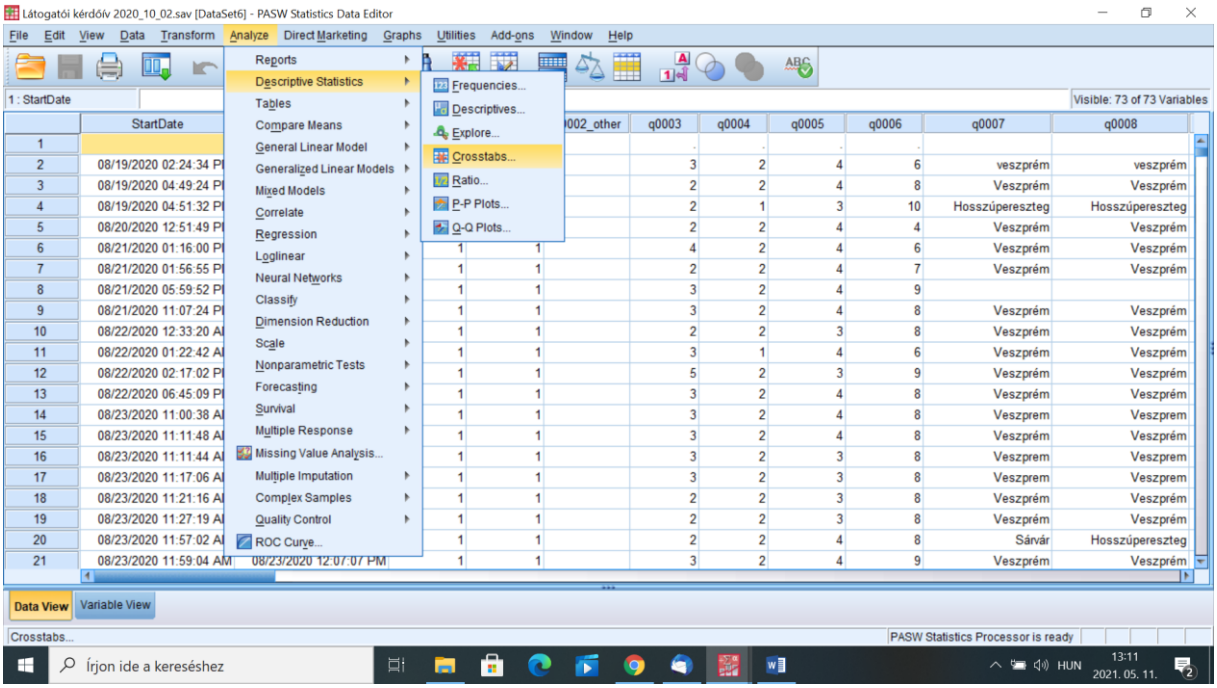
ANALYZE → *DESCRIPTIVES* → CROSSTABS

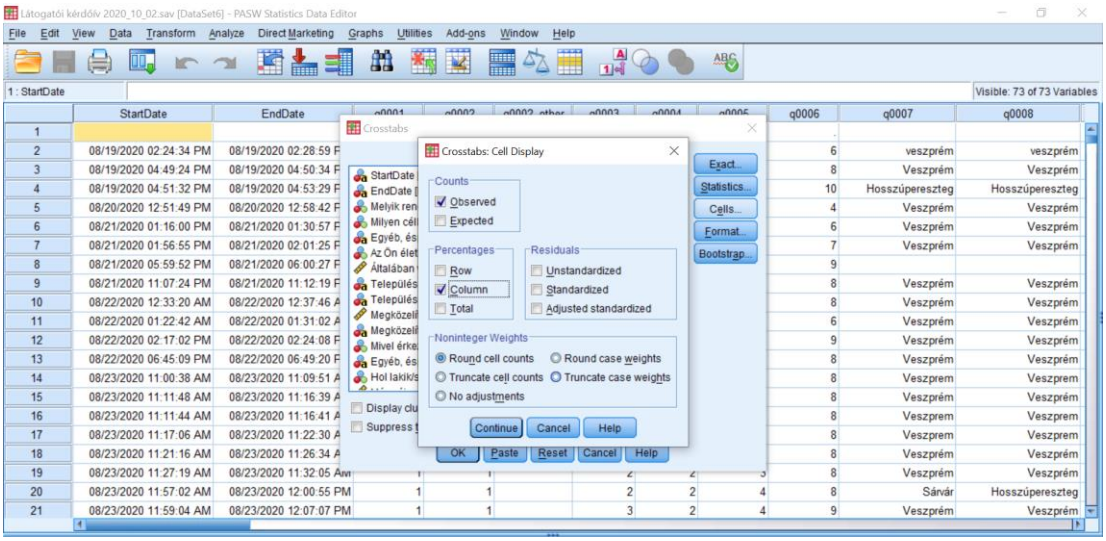
Rows: gender (categorical) - independent

Column: educational level - dependent

Click on Cells – change the counts: observed

Percentages: Column (% within gender of respondents)





CROSSTABS - **CHI-Square**

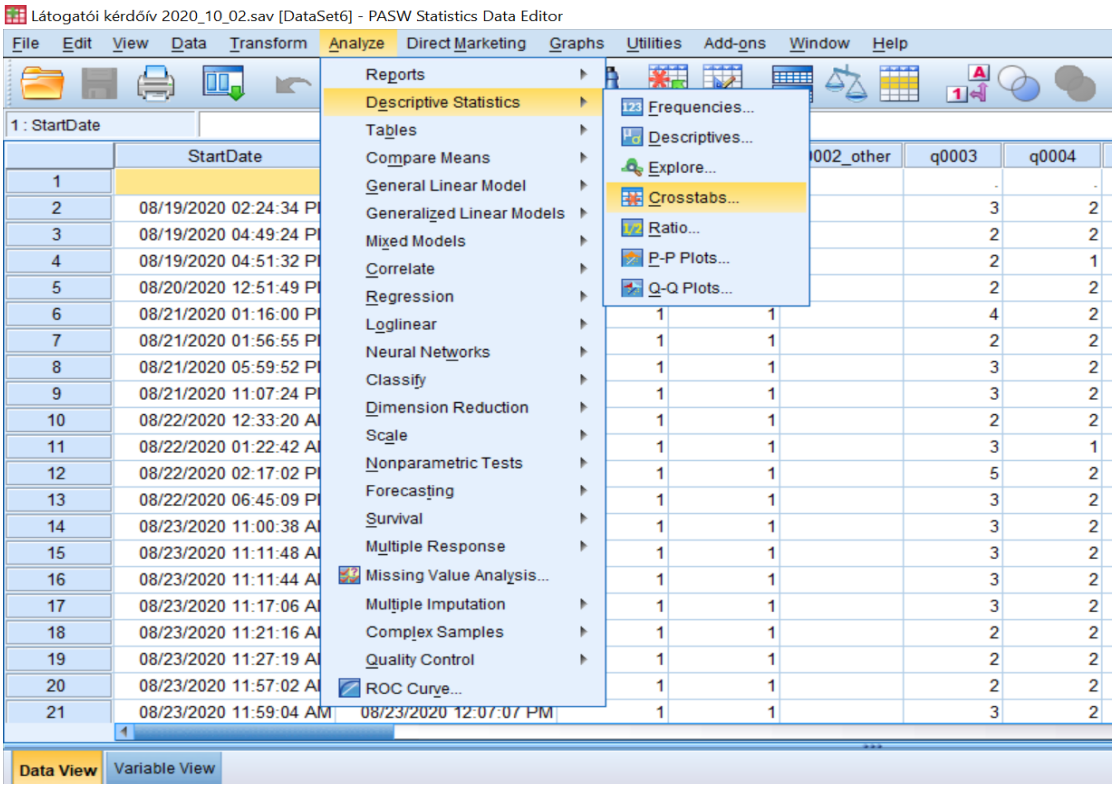
Analyze → Descriptive Statistics → Crosstabs → Statistics → √ Chi-square →

Independent: educational level, gender, location

Dependent: participation in an event, shopping behaviour

There is significance between two variables if $p < 0,05$

The level of education influence the participation in a marching event (if participate or not)



a5. Mi a legmagasabb iskolai végzettsége? * a6. Részt vett-e a felvonuláson? Crosstabulation

% within a5. Mi a legmagasabb iskolai végzettsége?

		a6. Részt vett-e a felvonuláson?		
		Igen	Nem	Total
a5. Mi a legmagasabb iskolai végzettsége?	Alapfokú	62.0%	38.0%	100.0%
	Középfokú	8.6%	91.4%	100.0%
	Fels?fokú	6.6%	93.4%	100.0%
Total		19.5%	80.5%	100.0%

Chi-Square Tests

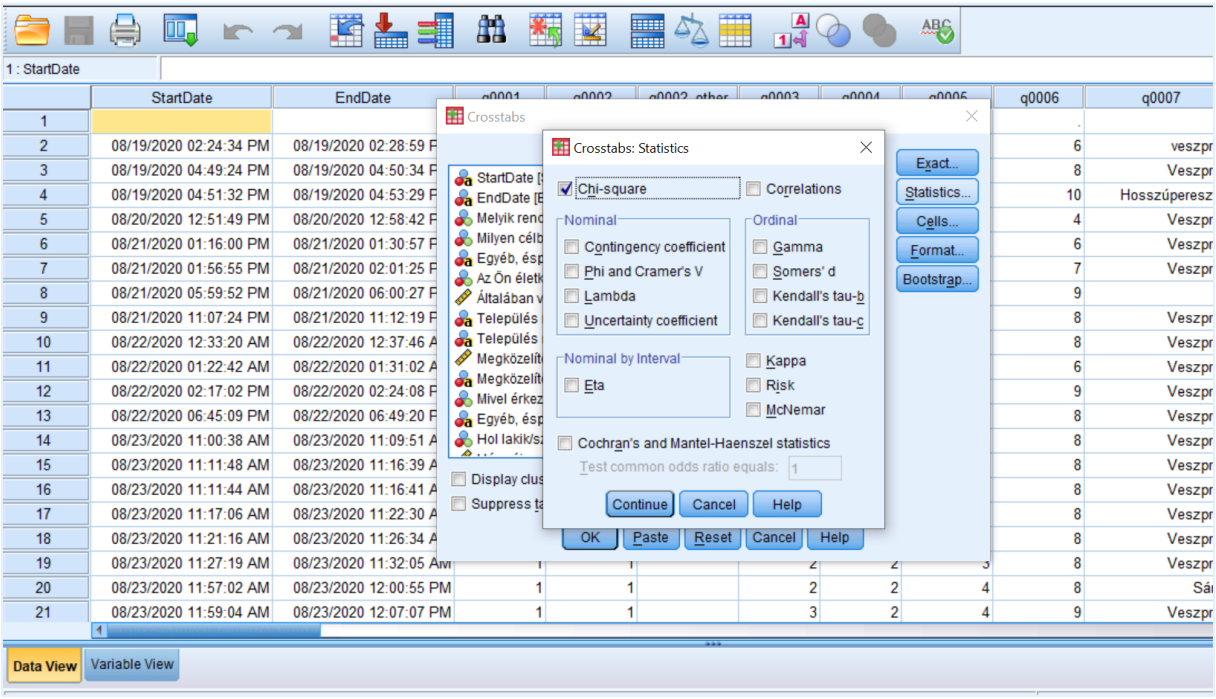
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	74.179[a]	2	.000
Likelihood Ratio	63.884	2	.000
Linear-by-Linear Association	54.427	1	.000
N of Valid Cases	226		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 9.73.

Those who has primary education (alapfokú), shows higher interest in participating in a march than the higher level of education (felsőfokú).

If there is a measured significance, we continue to check the following tests to measure associations between the variables:

- 1 nominal and one ordinal OR 2 nominal variables**: Cramer's V**
- 2 ordinal variables: **Gamma**



Cramer's V

A more interpretable measure of association between variables

Measures the association in a contingency table on a scale from 0 till 1

IF 1 nominal and one ordinal OR 2 nominal variables**: Cramer's V**

The value is between 0 and 1.

- If V=0, there is independency, the variables are not associated.
- If V=1, there is a strong dependence. They are perfectly associated.
- If V < 0,25 , the association is weak
- If V > 0,75, the association is very strong
- If 0,25 < V< 0,75, the association is moderate

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.573	.000
	Cramer's V	.573	.000
N of Valid Cases		226	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

GAMMA

The measures of association for ordinal variables. In case of 2 ordinal variables' association: use **Gamma**

Measure the association between the dependent ordinal variable (how happy you are?) and independent ordinal variable measure the condition of Health (excellent, good, fair, poor)

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Gamma	.385	.042	8.561	.000
N of Valid Cases		991			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

- Gamma value ranges from -1 and +1
- Gamma of 0.00 reflects no association
- Gamma of 1.00 reflects a positive perfect relationship between variables eg. People generally happier when they are healthier
- Gamma of -1.00 reflects a negative perfect relationship between variables

2) Categorical vs. numerical variables

Eg. Relationship between gender and heights/weights/income

ANALYZE → DESCRIPTIVES → COMPARE MEANS

COMPARE MEANS - we want to analyse the differences of heights or weights or income by gender?

DEPENDENT VALUE: heights or weights, income etc. (scale)

INDEPENDENT VALUE: gender (categorical/nominal)

Data to show: MEAN, NUMBER OF CASES, ST. DEVIATION, MEDIAN

3) Scale vs. scale variables

CORRELATIONS PEARSON'S R - MEASURES THE RELATIONS BETWEEN TWO VARIABLES

ANALYZE → CORRELATE→ BIVARIATE correlations

add two numeric continuous variables (heights and weights eg.) or age and TV hours (the number of TV hours viewing per day)

eg. pairing your heights and weights

test of significance: TWO-TAILED

- PEARSON'R (r is the coefficient of the correlation)
- FLAG SIGNIFICANT TEST – CLICK OK

we got the CORRELATION MATRIX with correlation coefficient

Correlations

		AGE OF RESPONDENT	HOURS PER DAY WATCHING TV
AGE OF RESPONDENT	Pearson Correlation	1	.139**
	Sig. (2-tailed)		.000
	N	1491	966
HOURS PER DAY WATCHING TV	Pearson Correlation	.139**	1
	Sig. (2-tailed)	.000	
	N	966	973

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

The values ranges from +1 and -1

- 0: no correlations
- -1 and +1 is perfect correlation
- Positive values means an increase in 1 variable is associated with an increase in the other variables
- Negative values means an increase in 1 variable is associated with a decrease in the other variables

Very weak correlation (0)

-1 ……0……..+1

stronger stronger (closer to either +1 or -1)

Eg. Age is strongly correlated to age. Pearson's r value for comparing age to age is 1, suggesting a strong correlation. Pearson's r value of 2 different variables (the value is .139). This suggests that someone ages, they watch more TV.

Sample size only those who has both data/values/variables – if one is missing either heights or weights is not possible to measure

LARGER SAMPLE SIZES ARE GENERALLY SHOWS MORE SIGNIFICANCE

DO IT AGAIN

ANALYZE → CORRELATE→ BIVARIATE correlations

add two variables (heights and weights eg.)

if we add another variables like GENDER + THE HEIGHTS AND WEIGHTS into it we do a POINT BISERIAL CORRELATION, just click ok

to measure the statistical significance between heights, weights and gender

How to make a picture of correlation?

use the GRAPHS menu (next to ANALYZE)

GRAPHS → CHART BUILDER (or in new versions Chart Legacy) → choose from graph types, eg. scatter plot

height will be the X axes (move with the cursor)

weight will be the y axes

Other tutorial videos to watch for correlation:

SPSS Correlations – explore the relationship between our two-scale level variables: heights and weights, and measure with Pearson's correlation.

https://www.youtube.com/watch?v=6EH5DSaCF_8

EXPORT DATA /OUTPUT TABLES FROM SPSS TO WORD

File → export or from output table the export window (click on the export icon)

you have to do it from the output tables

(objects to export = word/RTF (*doc)