

Week 3 CS - SPSS

Saturday, August 27, 2022 3:03 PM

Question 1.

For Chi square, I will use employment and savings status variables. We can see that the relationship is significant.

employment * savings_status Crosstabulation							
Count		savings_status					
		<100	≥1000	100≤X<500	500≤X<1000	no known savings	Total
employment	<1	120	7	17	5	23	172
	≥7	133	14	22	20	64	253
	1≤X<4	210	18	33	26	52	339
	4≤X<7	100	9	24	9	32	174
	unemployed	40	0	7	3	12	62
Total		603	48	103	63	183	1000

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	28.662 ^a	16	.026
Likelihood Ratio	31.749	16	.011
N of Valid Cases	1000		

a. 2 cells (8.0%) have expected count less than 5. The minimum expected count is 2.98.

For Pearson, I will use age and credit amount variables. We can see that the relationship is significant.

Correlations

		age	credit_amount
age	Pearson Correlation	1	.033
	Sig. (2-tailed)		.301
	N	1000	1000
credit_amount	Pearson Correlation	.033	1
	Sig. (2-tailed)	.301	
	N	1000	1000

I will use age variable for One Sample T test. We can see that mean age of the sample is significantly different than overall population age. Positive mean indicates that the sample age is bigger than the population age.

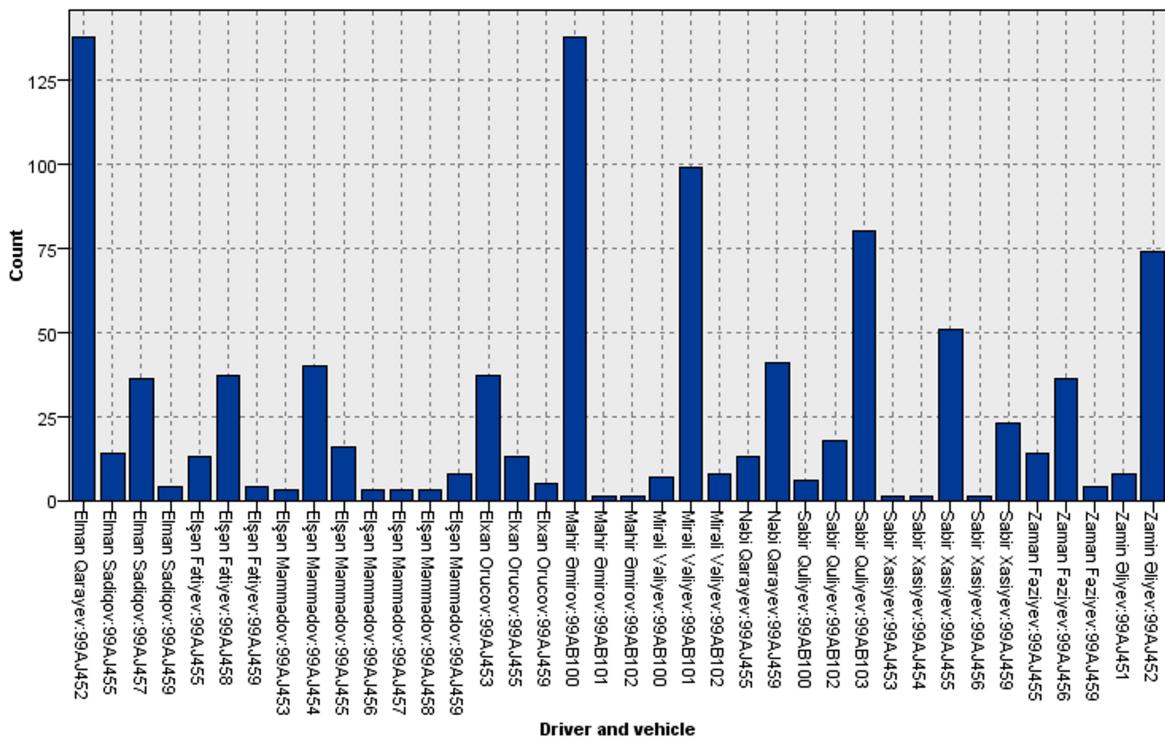
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
age	1000	35.55	11.375	.360

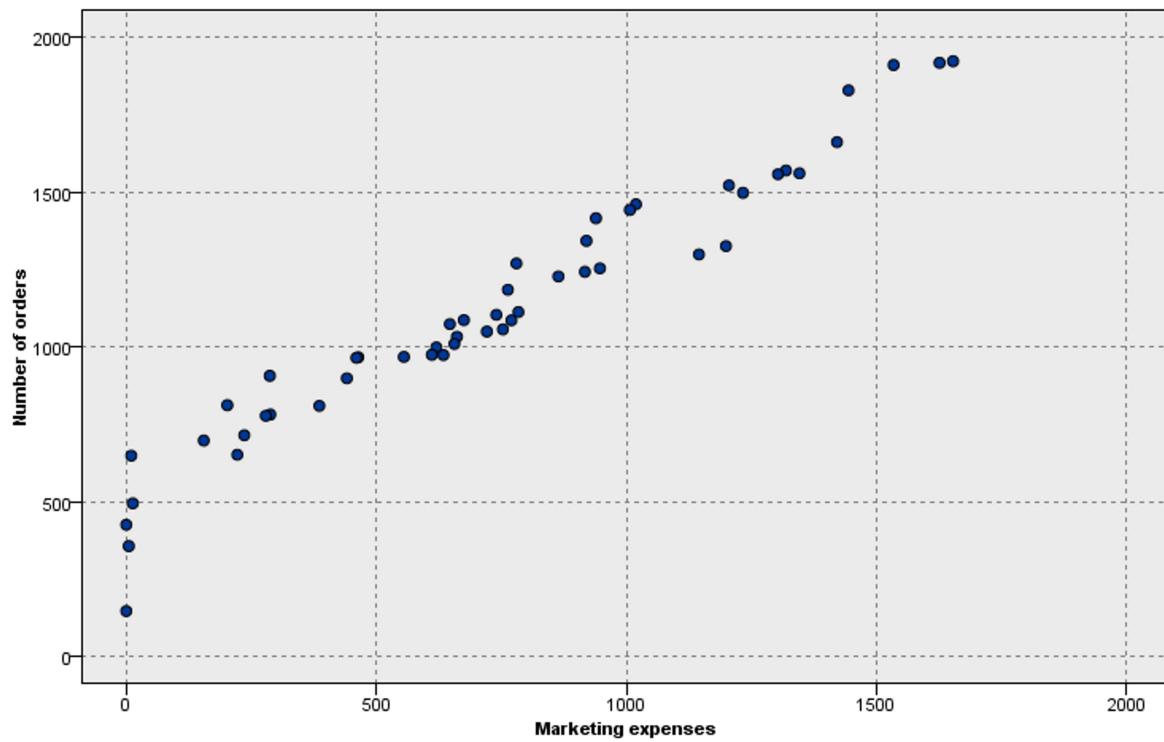
One-Sample Test

Test Value = 0							
t	df	Significance		Mean Difference	95% Confidence Interval of the Difference		
		One-Sided p	Two-Sided p		Lower	Upper	
age	98.815	999	.000	.000	35.546	34.84	36.25

Question 2.

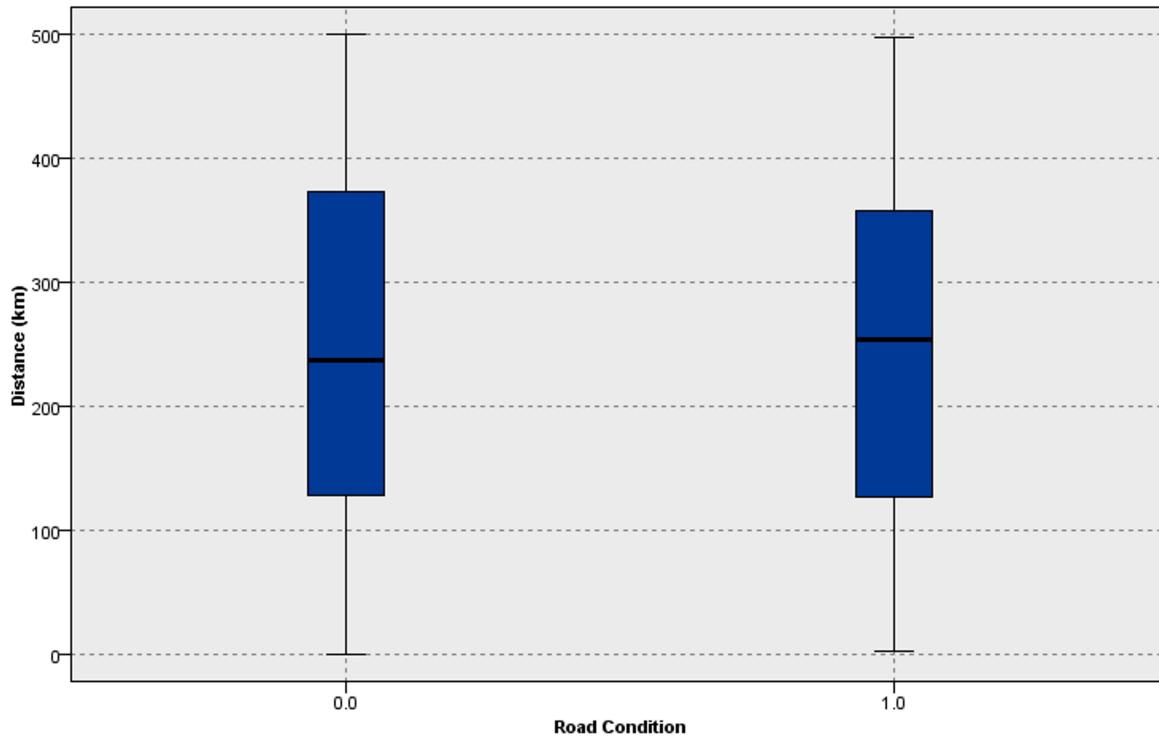


Question 3.

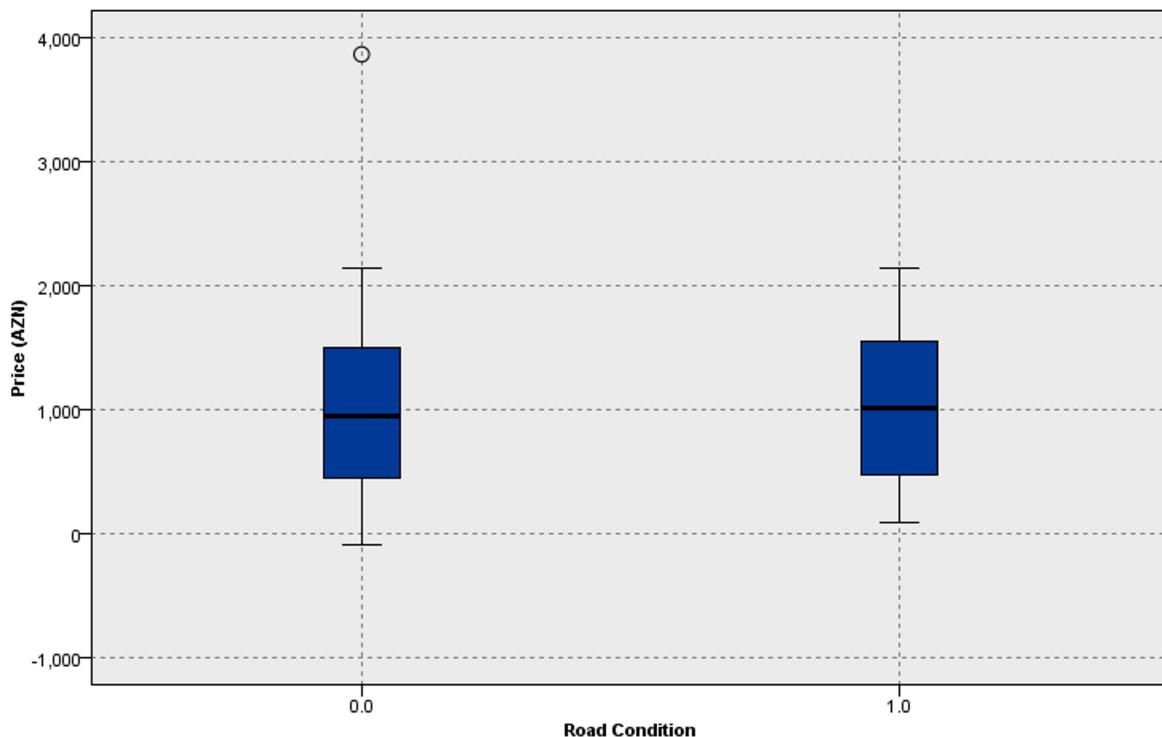


Question 4.

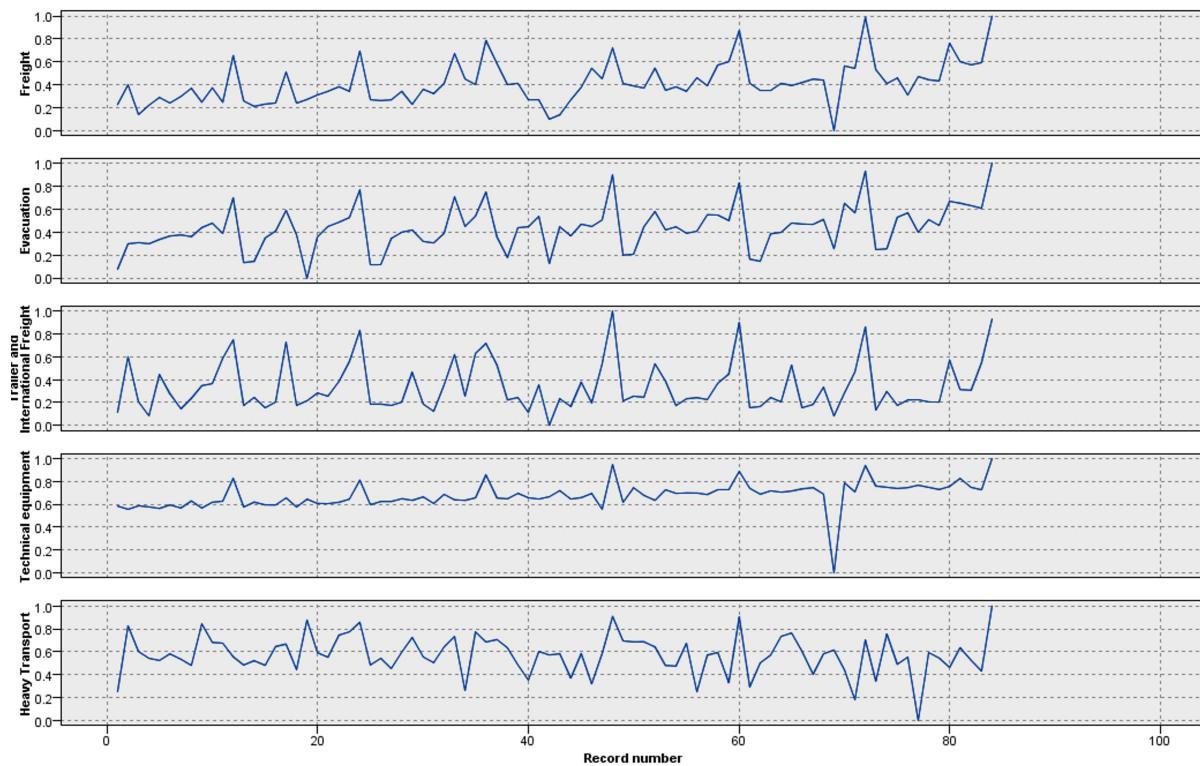
Distance and road condition box plot:



Price and road condition box plot:



Question 5.



Question 6.

For driver and vehicle transaction ID distance:

Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
Driver and vehicle		Categorical	--	--	--	--	--	38	1002
Transaction ID		Continuous	1001.000	9990.000	5556.639	2628.969	-0.058	--	1002
Distance (km)		Continuous	0.000	500.000	254.441	147.070	-0.033	--	1002

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value
Driver and ve...	Categorical	--	--	Never	Fixed		100	1002	0
Transaction ID	Continuous	0	0 None	Never	Fixed		100	1002	0
Distance (km)	Continuous	0	0 None	Never	Fixed		100	1002	0

We can see there are no outliers, and distributions are approximately symmetric.

For marketing expenses and number of orders:

Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
Marketing expenses		Continuous	0.000	1653.000	737.180	458.980	0.164	--	50
Number of orders		Continuous	147.000	1923.000	1120.180	403.105	0.024	--	50

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete
Marketing ex...	Continuous	0	0 None	Never	Fixed		10
Number of or...	Continuous	0	0 None	Never	Fixed		10

There are no outliers, and both distributions are approximately symmetric.

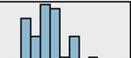
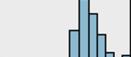
For road condition distance price:

Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
Road Conditi...		Continuous	0.000	1.000	0.234	0.424	1.266	--	231
Distance (km)		Continuous	0.000	500.000	247.346	145.619	0.047	--	231
Price (AZN)		Continuous	-89.300	3868.900	1015.627	647.972	0.493	--	230

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Comp
Road Conditi...	Continuous	0	0 None	Never	Fixed		
Distance (km)	Continuous	0	0 None	Never	Fixed		
Price (AZN)	Continuous	1	0 None	Never	Fixed		9

Price variable has 1 outlier. Road condition is highly skewed, and the rest are approximately symmetric.

For time series:

Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
Date		Continuous	2013-01-01	2019-12-01	--	--	--	--	84
Freight		Continuous	224.518	3760.966	1683.262	645.337	0.978	--	84
Evacuation		Continuous	1400.000	7924.597	4246.665	1255.526	0.408	--	84
Trailer and International Freight		Continuous	498.355	3723.157	1592.602	699.621	1.249	--	84
Technical equipment		Continuous	14.700	1426.300	970.339	164.599	-1.764	--	84
Heavy Transport		Continuous	1706.120	3566.660	2771.684	320.111	-0.359	--	84

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records
Date	Continuous	0	0 None	Never	Fixed		100	8
Freight	Continuous	2	0 None	Never	Fixed		100	8
Evacuation	Continuous	0	0 None	Never	Fixed		100	8
Trailer and In...	Continuous	1	0 None	Never	Fixed		100	8
Technical eq...	Continuous	0	1 None	Never	Fixed		100	8
Heavy Trans...	Continuous	1	0 None	Never	Fixed		100	8

We have 4 outliers here, Trailer and Internations Freight and Technical Equipment is highly skewed, Freight is moderately skewed and the rest is approximately symmetric.

Question 7.

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records
Date	Continuous	0	0 None	Never	Fixed		100	8
Freight	Continuous	0	0 None	Never	Fixed		100	8
Evacuation	Continuous	0	0 None	Never	Fixed		100	8
Trailer and In...	Continuous	0	0 None	Never	Fixed		100	8
Technical eq...	Continuous	1	0 None	Never	Fixed		100	8
Heavy Trans...	Continuous	0	0 None	Never	Fixed		100	8

After changing the settings, only 1 outlier is left and 0 extreme.