
Split file

Use: To see if there are any trends between split categories

Comments: Whenever in split mode, whatever analysis you run will be done on both of the groups separately. Make sure to deactivate the split once you are done analyzing the differences between the groups.

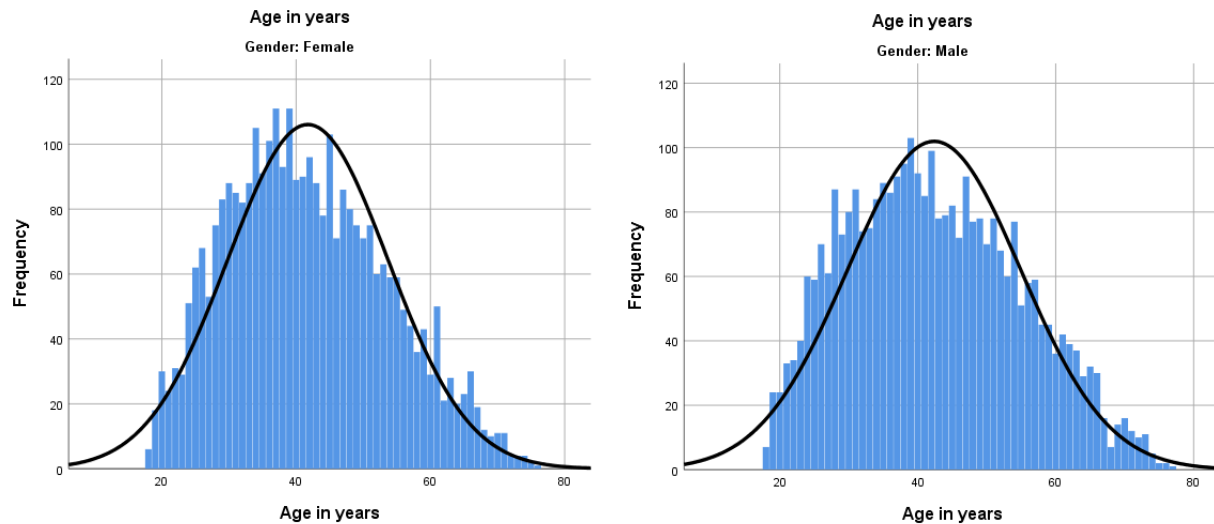
Example: Frequencies run while in split file mode

Statistics

Age in years

Gender	N		Mean	Median	Mode	Std. Deviation	Variance	Range
	Valid	Missing						
Female	3179	0	41.74	41.00	37 ^a	11.958	142.988	58
Male	3221	0	42.37	41.00	39	12.602	158.818	59

a. Multiple modes exist. The smallest value is shown

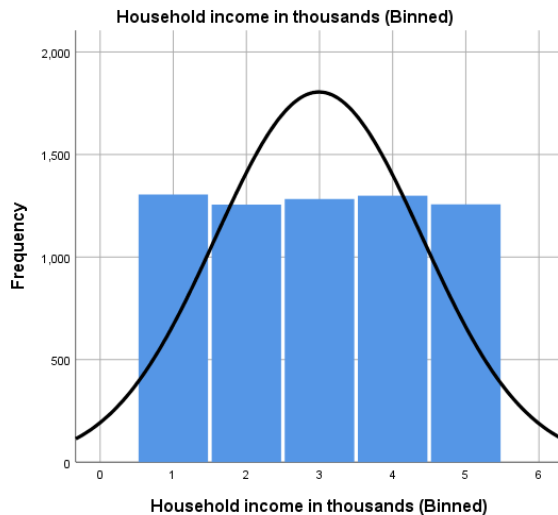
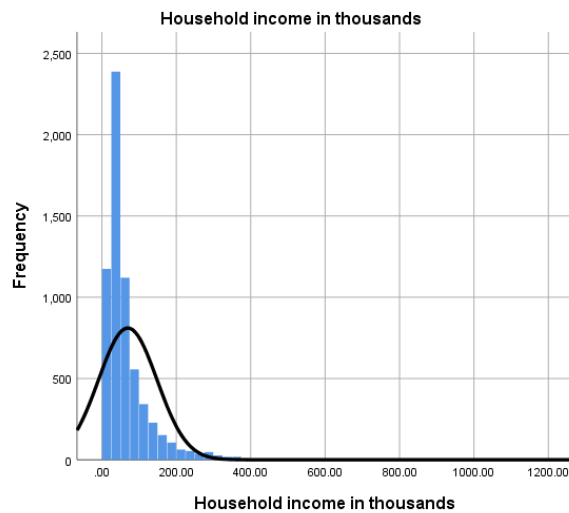


Results: In this case, there isn't a large difference between the age distributions among the two genders.

Visual Binning

Use: Helps minimize the effects of outliers if the data isn't normally distributed

Comments: Visual binning creates a separate variable where the data is recoded into more or less even groups. I graphed both the original data and the binned data to show the effect. This data is simply grouped into quintiles.



Optimal binning

Use: Helps show the differences between proposed subgroups

Comments: In this run, we looked at the groups that resulted when we compared household income to level of education. SPSS chose to split the subgroups at \$41,000.

		Household income in thousands							
		End Point		Number of Cases by Level of Level of education					
Bin		Lower	Upper	Did not complete high school	High school degree	Some college	College degree	Post-undergraduate degree	Total
1	a		41.00	716	927	628	524	111	2906
2		41.00	a	674	1009	732	831	248	3494
Total				1390	1936	1360	1355	359	6400

Each bin is computed as Lower <= Household income in thousands < Upper.

a. Unbounded

Multiple response frequencies

Use: Combining variables when the respondents could choose multiple variables

\$Tech_mr Frequencies

		Responses		Percent of Cases
		N	Percent	
Number of Tech Services or Appliances ^a	Wireless service	2547	6.1%	39.9%
	Multiple lines	2691	6.4%	42.1%
	Voice mail	2755	6.6%	43.1%
	Paging service	1581	3.8%	24.7%
	Internet	1636	3.9%	25.6%
	Caller ID	3267	7.8%	51.1%
	Call waiting	3247	7.8%	50.8%
	Owens TV	6337	15.2%	99.2%
	Owens VCR	6145	14.7%	96.2%
	Owens stereo/CD player	6206	14.9%	97.1%
	Owens PDA	1307	3.1%	20.5%
	Owens computer	2811	6.7%	44.0%
	Owens fax machine	1202	2.9%	18.8%
Total		41732	100.0%	653.0%

a. Dichotomy group tabulated at value 1.

Findings: Almost everyone owns a TV (99.2%). Only about a fifth of respondents own a PDA.

Multiple Response Crosstabs

Use: When you want to compare the various responses to another variable

\$Tech_mr*inccat Crosstabulation

			Income category in thousands				
			Under \$25	\$25 - \$49	\$50 - \$74	\$75+	Total
Number of Tech Services or Appliances ^a	Wireless service	Count	352	908	476	811	2547
	Multiple lines	Count	400	881	480	930	2691
	Voice mail	Count	422	907	499	927	2755
	Paging service	Count	248	464	276	593	1581
	Internet	Count	246	641	293	456	1636
	Caller ID	Count	489	1205	569	1004	3267
	Call waiting	Count	518	1146	591	992	3247
	Owens TV	Count	1116	2384	1120	1717	6337
	Owens VCR	Count	997	2313	1118	1717	6145
	Owens stereo/CD player	Count	1030	2339	1119	1718	6206
	Owens PDA	Count	191	455	231	430	1307
	Owens computer	Count	476	1049	519	767	2811
	Owens fax machine	Count	160	432	218	392	1202
	Total	Count	1166	2387	1120	1718	6391

Percentages and totals are based on respondents.

a. Dichotomy group tabulated at value 1.

Findings: Adding percentages to these can be useful to compare across rows and columns.

Nonparametric tests

Use: Determines if groups are statistically different from one another

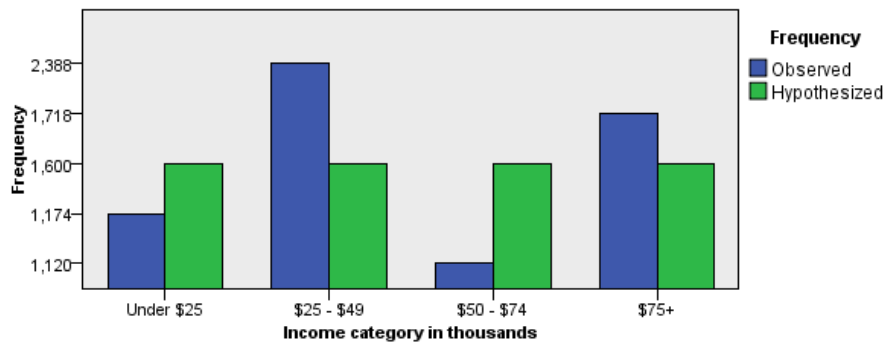
Comment: To view more details and graphs, you can double click on the yellow when in SPSS output mode.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories of Income category in thousands occur with equal probabilities.	One-Sample Chi-Square Test	.000	Reject the null hypothesis.
2	The categories of Primary vehicle price category occur with equal probabilities.	One-Sample Chi-Square Test	.000	Reject the null hypothesis.
3	The categories of Years with current employer occur with equal probabilities.	One-Sample Chi-Square Test	.000	Reject the null hypothesis.
4	The categories defined by Gender = Female and Male occur with probabilities 0.5 and 0.5.	One-Sample Binomial Test	.608	Retain the null hypothesis.

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

One-Sample Chi-Square Test

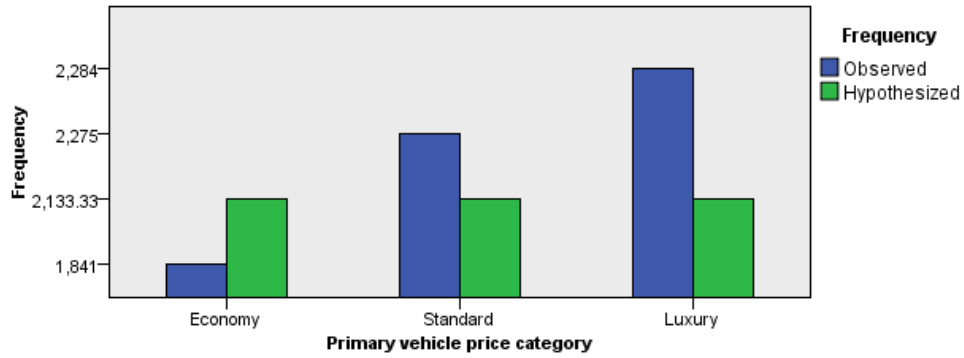


Total N	6,400
Test Statistic	654.215
Degrees of Freedom	3
Asymptotic Sig. (2-sided test)	.000

Findings: The significance for household income was 0.000, meaning that we can reject the null that the groups are statistically equal. This becomes very evident when we view the chart: the green is the expected distribution while the blue is the observed distribution. Clearly they are very different.

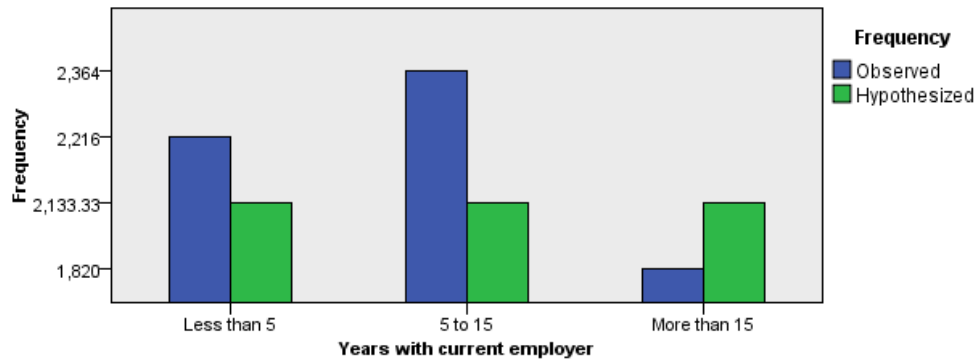
Data used is fictional and comes with the SPSS software. It's called "demo".

One-Sample Chi-Square Test



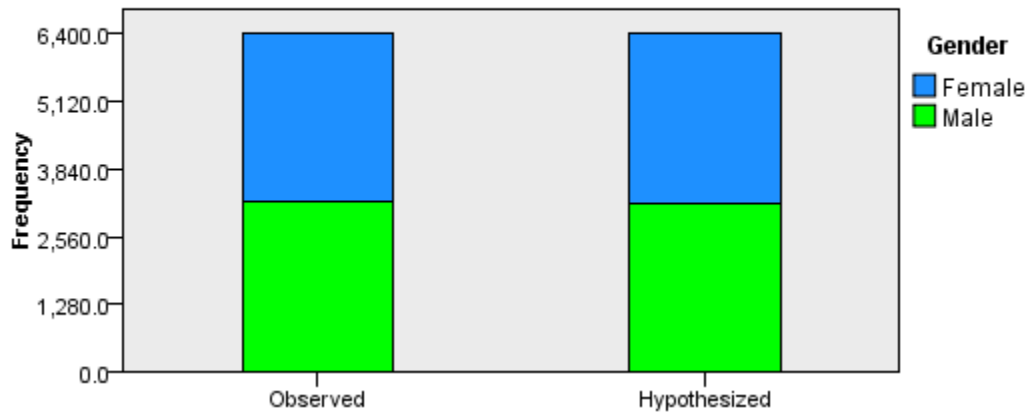
Total N	6,400
Test Statistic	60.107
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

One-Sample Chi-Square Test



Total N	6,400
Test Statistic	74.165
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

One-Sample Binomial Test



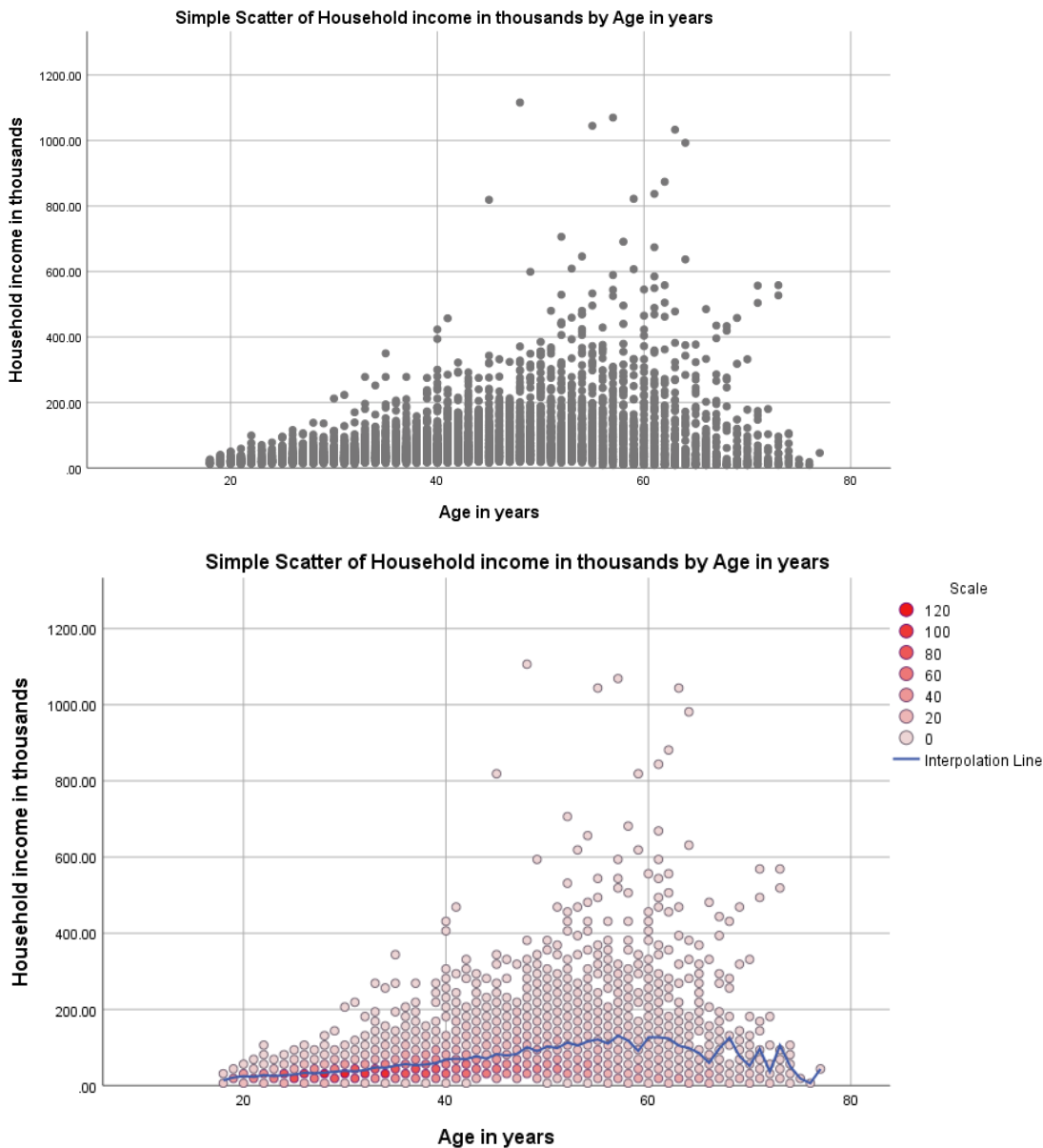
Total N	6,400
Test Statistic	3,179.000
Standard Error	40.000
Standardized Test Statistic	-.512
Asymptotic Sig. (2-sided test)	.608

Findings: Gender is the only variable with a high significance level. This means that we fail to reject the null, meaning that there is not statistical evidence suggesting that the two groups are different. As we can see from the graph, they do in fact look alike.

Simple Scatterplot

Use: Finding a trend; exploratory

Comments: When there's a lot of data, you might need to do some editing to make it readable. The first graph is the original scatterplot while the second is the edited scatterplot.



Findings: Changing the color to red somehow makes it the most readable. I tried having it at blue and then at green but the colors did not quite get this large range. Moreover, adding an interpolation line gives some idea of what the trend in the data could be. I changed that line to blue to create some contrast against the red.