[Commentary by **John F Hall**] [Last updated: 12 August 2017]

John MacInnes
An Introduction to Secondary Data Analysis with IBM SPSS Statistics
(Sage, Dec. 2017)

5.1 Chapter 5 video tutorials (direct link to companion website)
[**NB:** All video tutorials for chapter 5 are on the same web page and cannot (yet) be disaggregated]

Video 5.1.7: Creating and editing a Histogram of **[depress]** (7'05")

Because **[cldgng]** is not available for Albanian respondents, JM gave them an **imputed** depression score by multiplying their score on the other seven items by a factor of **8 ÷ 7** and produces a frequency table for depression.

jmdepress

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	7.00	2	.0	.0	.0
	8.00	2449	4.3	4.4	4.4
	9.00	2977	5.2	5.3	9.8
	9.14	102	.2	.2	9.9
	10.00	4899	8.6	8.8	18.7
	10.29	114	.2	.2	18.9
	11.00	5460	9.6	9.8	28.7
	11.43	204	.4	.4	29.1
	12.00	5610	9.9	10.1	39.2
	12.57	256	.5	.5	39.7
	13.00	5167	9.1	9.3	48.9
	13.71	163	.3	.3	49.2
	14.00	4968	8.7	8.9	58.2
	14.86	221	.4	.4	58.5
	15.00	4453	7.8	8.0	66.5
	16.00	3941	6.9	7.1	73.6
	17.00	3090	5.4	5.6	79.2
	17.14	169	.3	.3	79.5
	18.00	2610	4.6	4.7	84.2
	18.29	148	.3	.3	84.4
	19.00	1749	3.1	3.1	87.6
	19.43	123	.2	.2	87.8
	20.00	1426	2.5	2.6	90.4
	20.57	106	.2	.2	90.5
	21.00	1130	2.0	2.0	92.6
	21.71	80	.1	.1	92.7
	22.00	1053	1.9	1.9	94.6
	22.86	56	.1	.1	94.7
	23.00	705	1.2	1.3	96.0
	24.00	755	1.3	1.4	97.3
	25.00	400	.7	.7	98.1
	25.14	20	.0	.0	98.1
	26.00	319	.6	.6	98.7
	26.29	17	.0	.0	98.7
	27.00	196	.3	.4	99.0
	27.43	18	.0	.0	99.1
	28.00	108	.2	.2	99.3
	28.57	18	.0	.0	99.3
	29.00	154	.3	.3	99.6
	29.71	5	.0	.0	99.6
	30.00	92	.2	.2	99.8
	30.86	2	.0	.0	99.8
	31.00	48	.1	.1	99.8
	32.00	87	.2	.2	100.0
	Total	55671	98.0	100.0	
Missing	System	1164	2.0		
Total		56835	100.0		

The table has an uneven distribution of frequency counts because of the fractional scores created, so JM says, " . it's not too easy to interpret and there are far too many numbers." He suggests one way round this is to use graphics and says a **histogram** would give a better idea of the distribution.

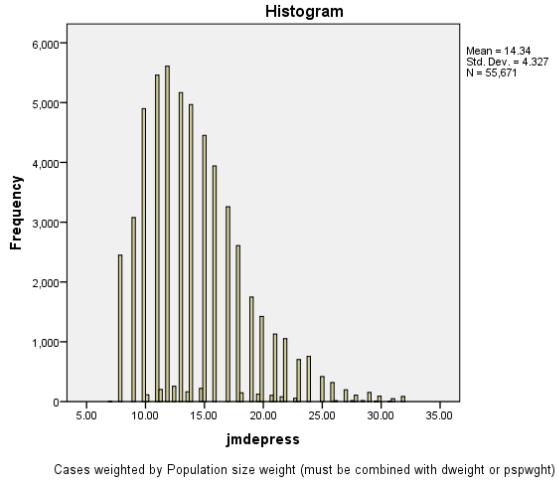
```
freq depress
        /histogram.
```

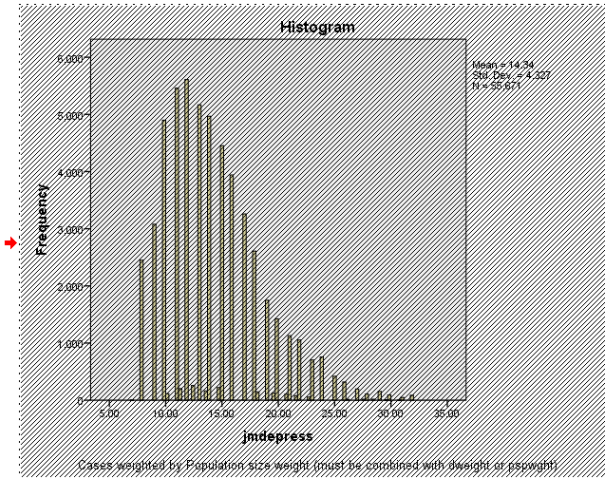
[**NB:** He has already produced the frequency table: this command produces another frequency table. If a hard copy of the output is printed, this would be a waste of paper. It is better to suppress the frequency table and display only the histogram:

```
frequencies  depress /format notable  /histogram.
```

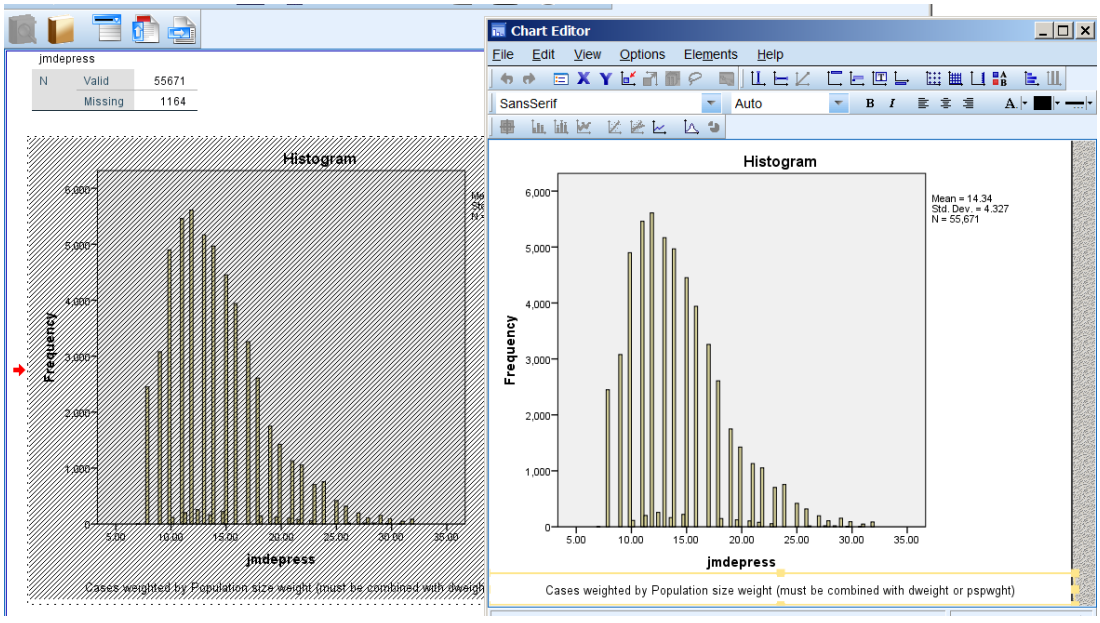
. . suppresses the frequency table and yields a chart with high columns for integer scores interleaved with small ones for the fractional scores.

Proceeds to change the width of the **bins** by double clicking on the chart :

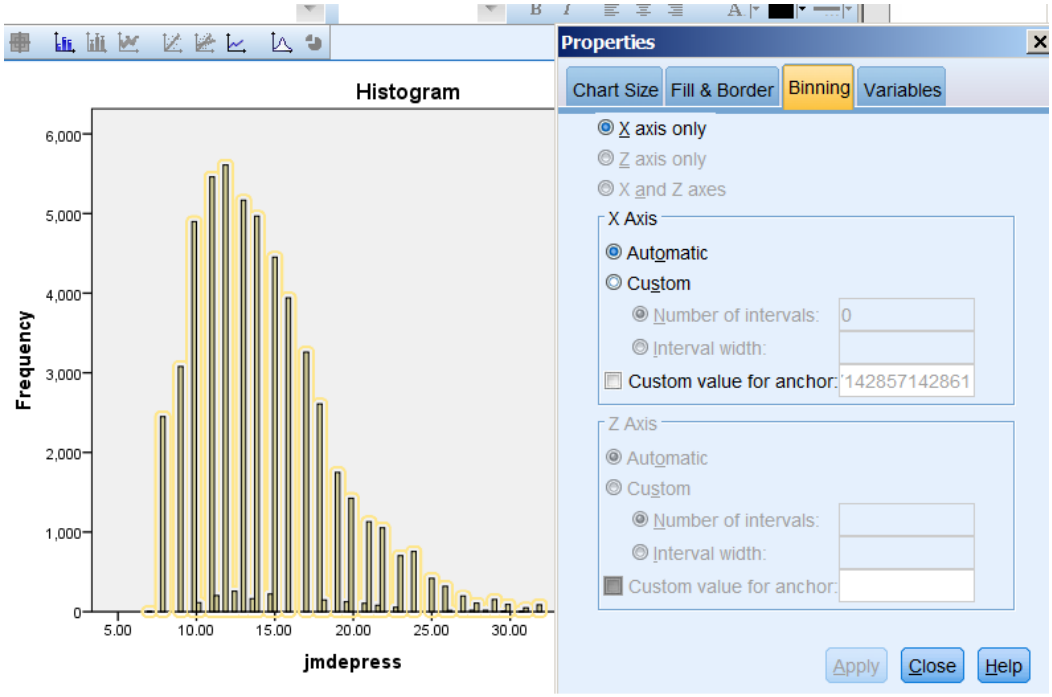




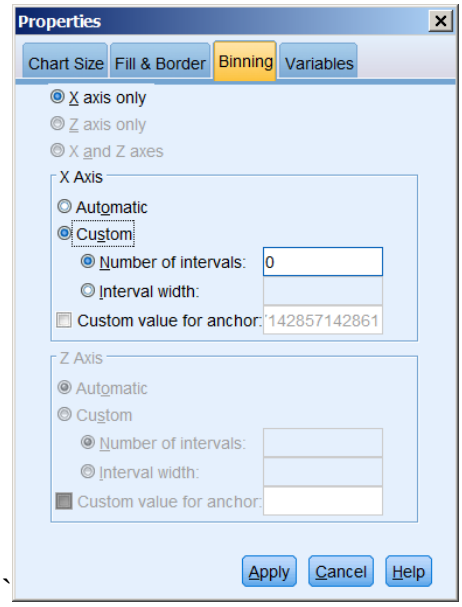
. . to bring up the Chart Editor. Double click on any bar in the Chart Editor



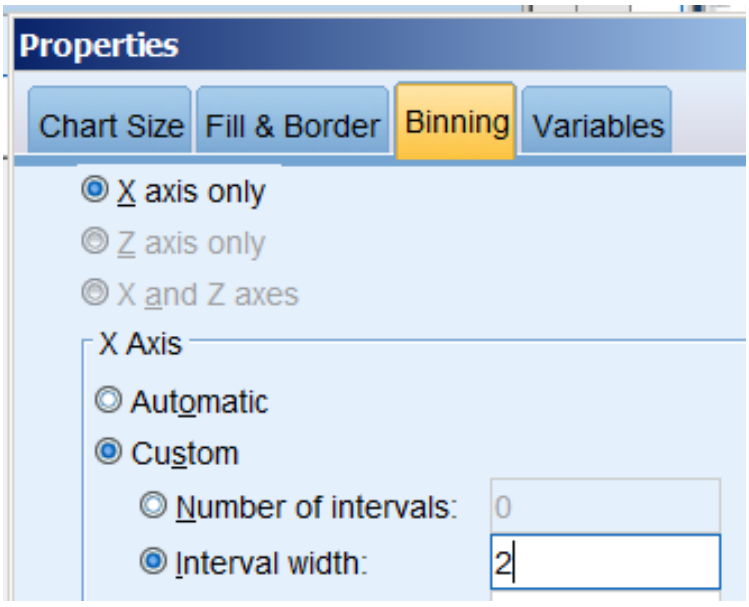
. . to open the Properties window:



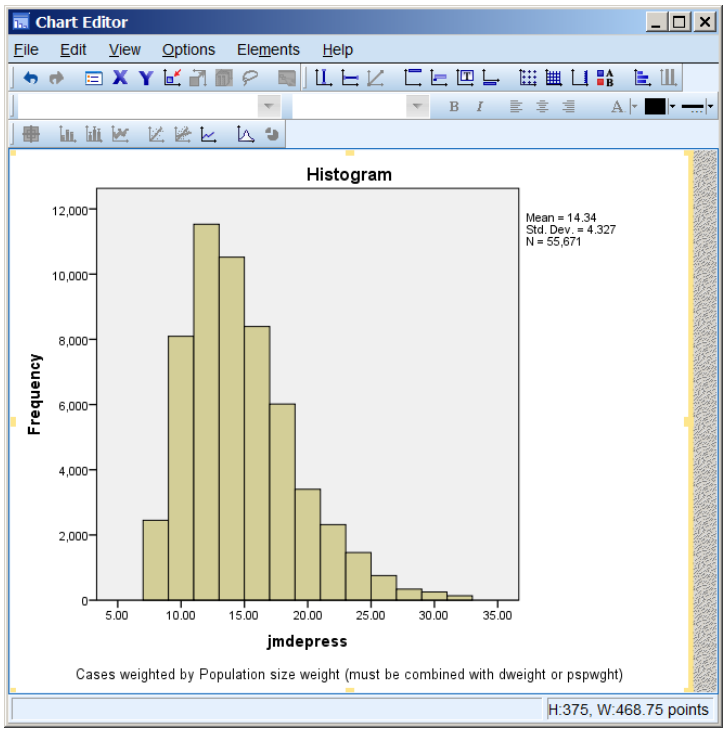
Click on the Custom button



. . and enter the number **2** in the Interval width box



. . to yield a histogram with wider **bins**, ". .which is much easier to interpret."

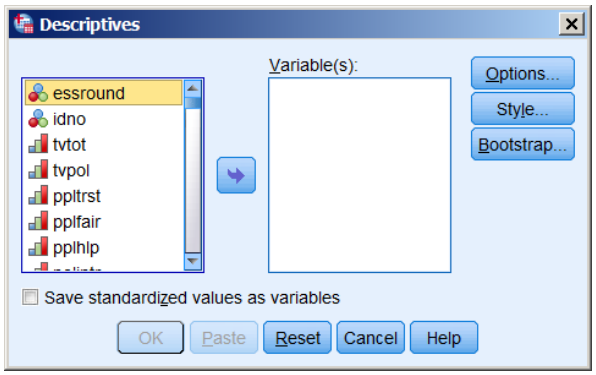


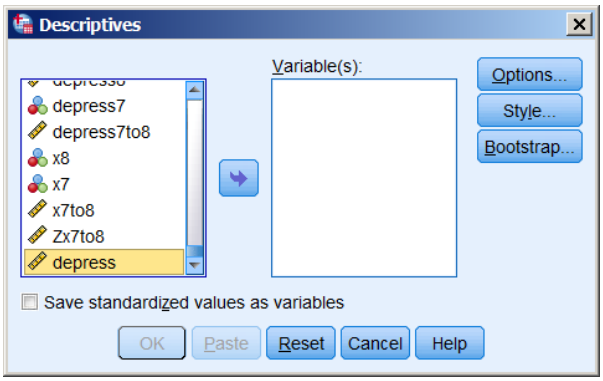
JM comments on the "tail" of people to the right, but misses the opportunity to mention **skew**.

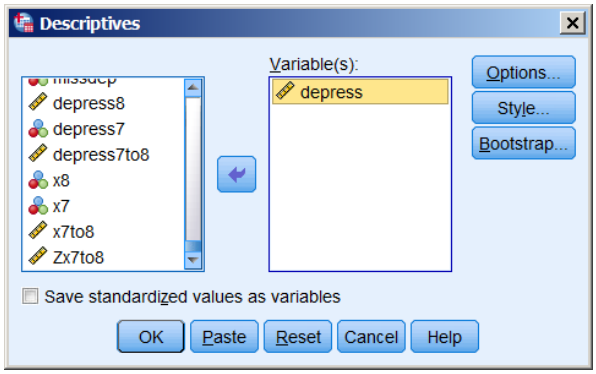
Doesn't point out that, for **Ordinal** variables, because there is no fixed interval between the points, a **barchart** has spaces between the bars, but for **Scale** variables, which have a fixed interval between the points, the bars in a **histogram** are actually **touching**.

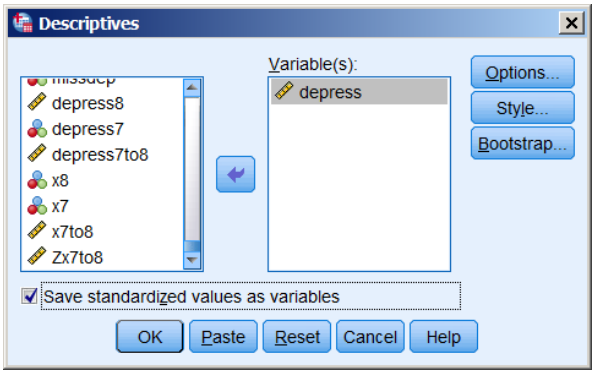
A **mean** of 15.4 with a **standard deviation** of 4.56 on a scale ranging from 8 to 32 is not particularly intuitive, so he demonstrates that the depression score can be **standardised** to have a mean of **0** and a standard deviation of **1**.

Analyze >> Descriptive Statistics >> Descriptives









Check Save standardized values box then Paste to get z-scores but it doesn't work so he types in syntax direct:

```
desc depress
  /save .
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
depress	55671	7.00	32.00	14.3428	4.32651
Valid N (listwise)	55671				

The standardised variable has its name pre-fixed with Z and is appended to the end of the file.

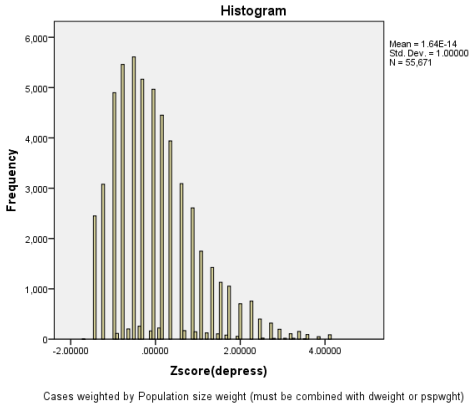
	Name	Measure	
638	depress	Scale	
639	Zdepress	Scale	Zscore(depress)
640			

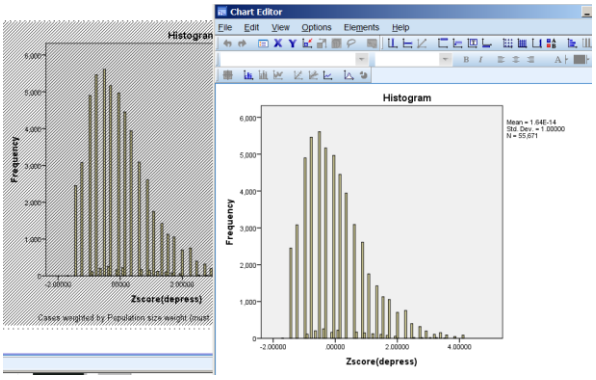
Recalls the syntax, modifies it to produce the histogram with

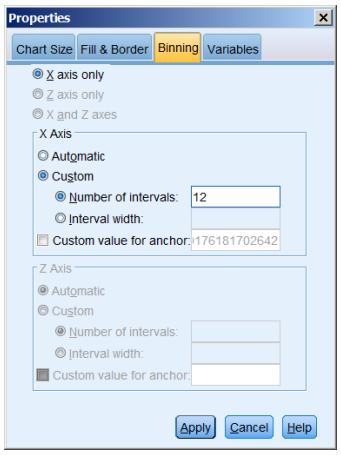
```
freq zdepress
     /his .
```

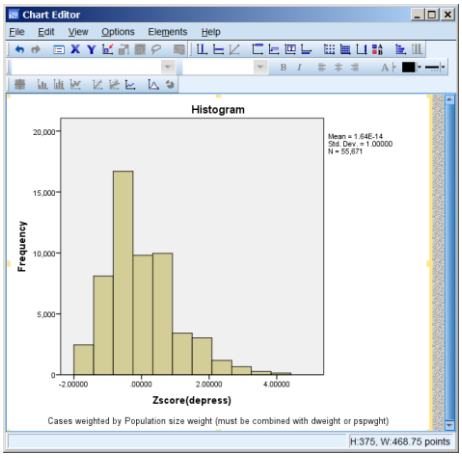
Zscore(depress)

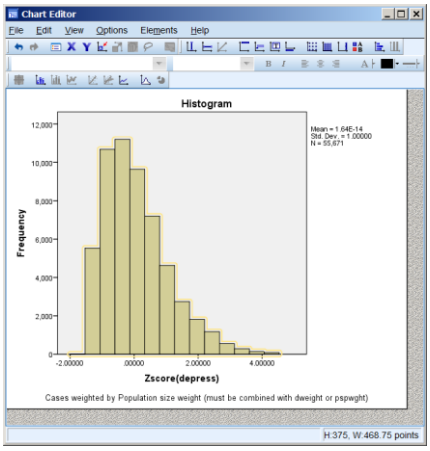
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-1.69716	2	.0	.0	.0
	-1.46602	2449	4.3	4.4	4.4
	-1.23489	2977	5.2	5.3	9.8
	-1.20187	102	.2	.2	9.9
	-1.00376	4899	8.6	8.8	18.7
	-.93772	114	.2	.2	18.9
	-.77262	5460	9.6	9.8	28.7
	-.67357	204	.4	.4	29.1
	-.54149	5610	9.9	10.1	39.2
	-.40941	256	.5	.5	39.7
	-.31036	5167	9.1	9.3	48.9
	-.14526	163	.3	.3	49.2
	-.07922	4968	8.7	8.9	58.2
	.11889	221	.4	.4	58.5
	.15191	4453	7.8	8.0	66.5
	.38304	3941	6.9	7.1	73.6
	.61418	3090	5.4	5.6	79.2
	.64720	169	.3	.3	79.5
	.84531	2610	4.6	4.7	84.2
	.91135	148	.3	.3	84.4
	1.07644	1749	3.1	3.1	87.6
	1.17550	123	.2	.2	87.8
	1.30758	1426	2.5	2.6	90.4
	1.43965	106	.2	.2	90.5
	1.53871	1130	2.0	2.0	92.6
	1.70381	80	.1	.1	92.7
	1.76984	1053	1.9	1.9	94.6
	1.96796	56	.1	.1	94.7
	2.00098	705	1.2	1.3	96.0
	2.23211	755	1.3	1.4	97.3
	2.46324	400	.7	.7	98.1
	2.49626	20	.0	.0	98.1
	2.69438	319	.6	.6	98.7
	2.76041	17	.0	.0	98.7
	2.92551	196	.3	.4	99.0
	3.02457	18	.0	.0	99.1
	3.15664	108	.2	.2	99.3
	3.28872	18	.0	.0	99.3
	3.38778	154	.3	.3	99.6
	3.55287	5	.0	.0	99.6
	3.61891	92	.2	.2	99.8
	3.81702	2	.0	.0	99.8
	3.85004	48	.1	.1	99.8
	4.08118	87	.2	.2	100.0
	Total	55671	98.0	100.0	
Missing	System	1164	2.0		
Total		56835	100.0		











Present results to non-expert audience. Rescales from 0 to 10

```
compute depress10 = (depress-8)/2.4.
freq    depress10.

means   depress10 by cntry by gndr.
```

z-scores: same histogram. Repeat chart edit to 12 bins. "Up the number of bins a bit" to 15 yields smoother distribution.

Draws no conclusions.

End of: **5.1.7:** Creating and editing a Histogram

Back to: **MacInnes (2017)**