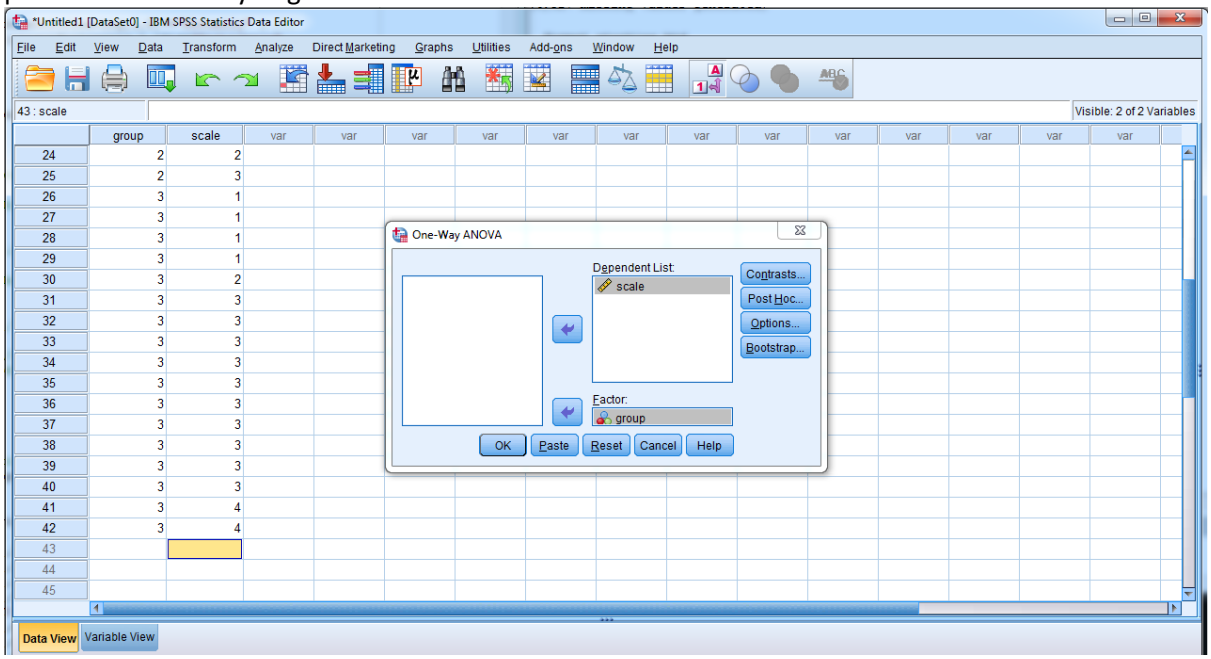
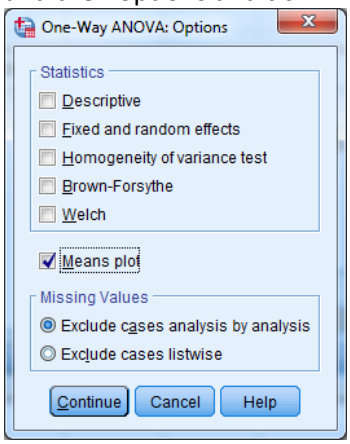
Getting means (& 95% CI) plots from SPSS via Excel

prepared grudgingly by Robert Grant (robertgrantstats.co.uk) who dislikes both packages

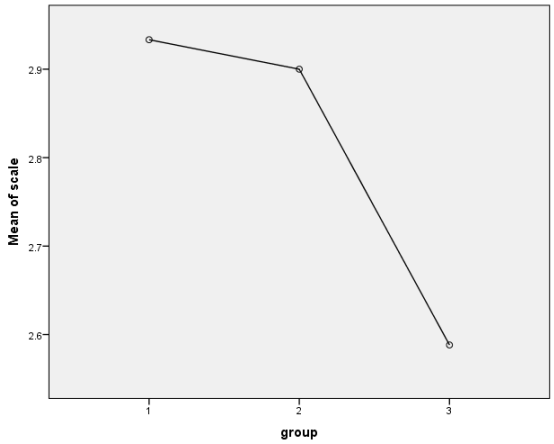
When you have a continuous outcome variable and are comparing it across various groups, it's often informative to plot the mean and its confidence interval in each group. However, this is not directly possible in SPSS. If you go to ANOVA:



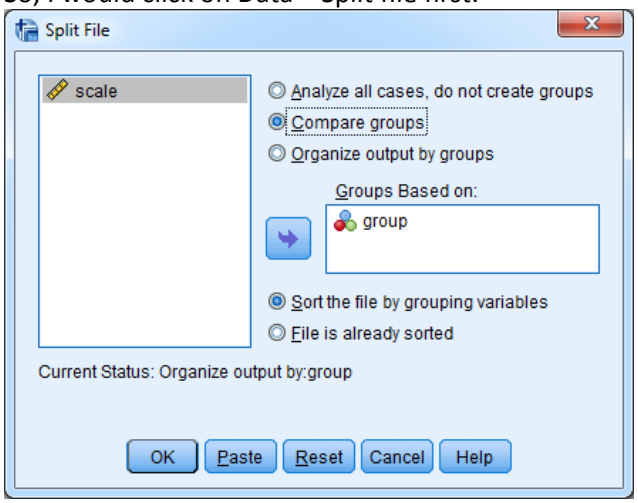
and then options and tick "means plot", unfortunately the result is not very useful



because it doesn't show confidence intervals: how uncertain are those means? Especially when some groups are smaller than others

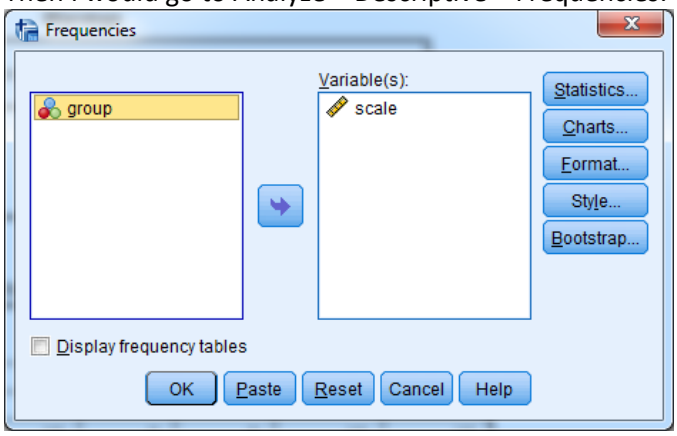


So, I would click on Data – Split file first:

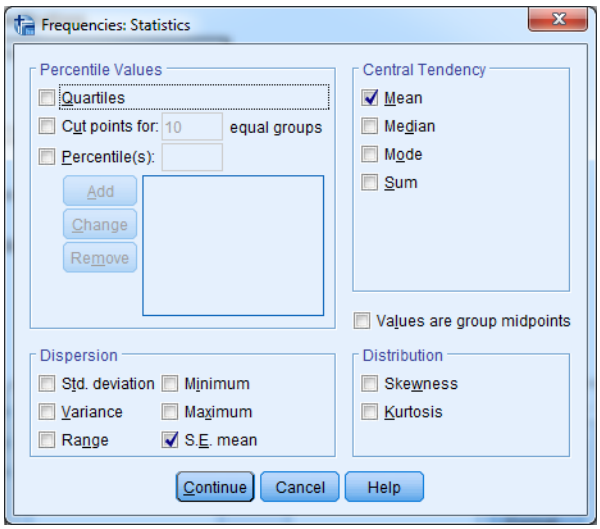


then whatever you do, it will repeat for each of the categories in the predictor variable. Don't forget to turn off split file when you're finished!

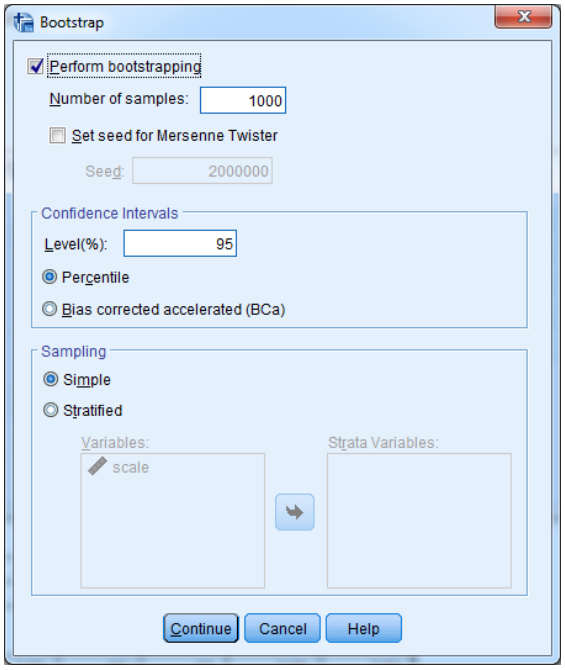
Then I would go to Analyze – Descriptive – Frequencies:



and click on statistics

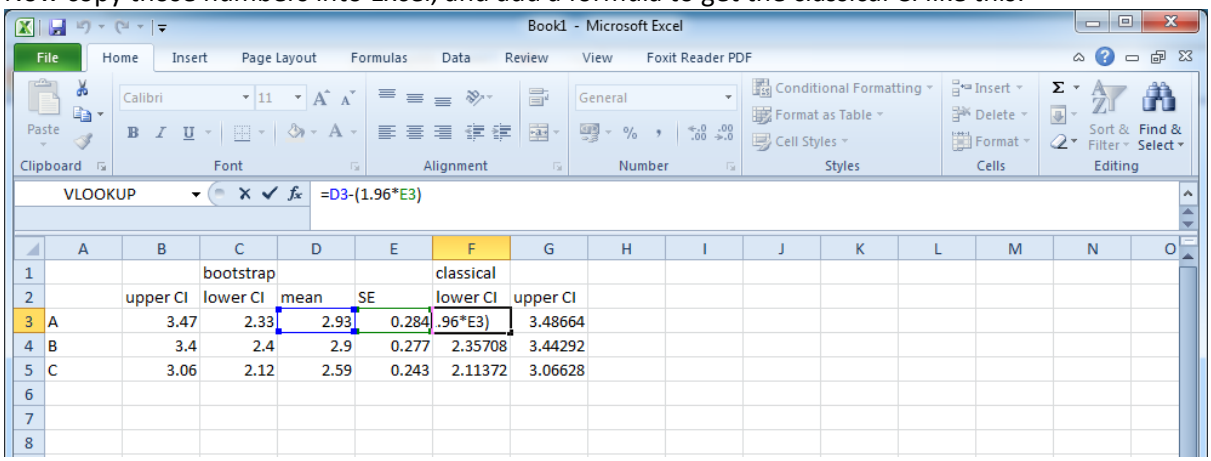


then 'continue', and then click on 'Bootstrap' (do you have a bootstrap button here? If not, either hand over the $$$ to get it from IBM, or just ignore the bootstrap stuff that follows (or get better stats software!))



Tick 'Perform bootstrapping' and then Continue. Don't forget to untick 'Display frequency tables'! Then when you click OK, you will get the mean, its standard error (from which you can calculate a classical approximate 95% CI) and the bootstrap CI (which helps when you have non-normal data with a ceiling effect).

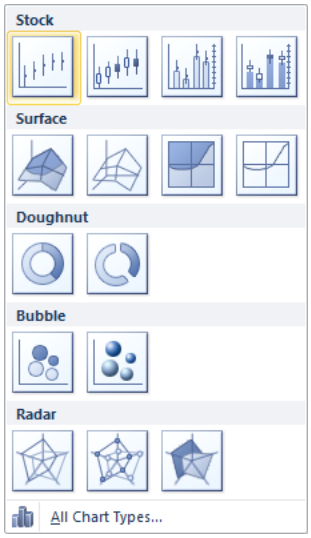
Now copy those numbers into Excel, and add a formula to get the classical CI like this:



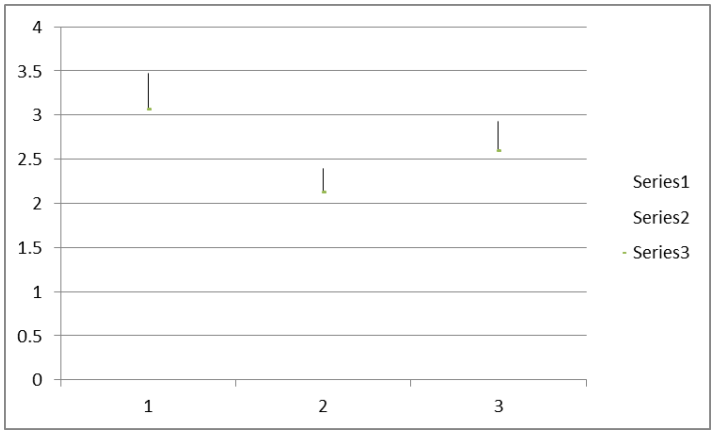
	A	B	C	D	E	F	G
1			bootstrap			classical	
2		upper CI	lower CI	mean	SE	lower CI	upper CI
3	A	3.47	2.33	2.93	0.284	.96*E3)	3.48664
4	B	3.4	2.4	2.9	0.277	2.35708	3.44292
5	C	3.06	2.12	2.59	0.243	2.11372	3.06628
6							
7							
8							

the lower CI is mean – (1.96*SE) and the upper CI is mean + (1.96*SE). Note the odd way of writing upper CI, then lower CI, then mean. That's how Excel will want it for the chart. If the classical and bootstrap options are notably different, you may find that the bootstrap avoids the upper or lower limit of the outcome (floor/celiling effect), which is good. But if you have very few data or some suspicious outlier, it might be best to ignore the bootstrap. Otherwise, the choice doesn't matter much. If you're worried about these CIs, talk to a statistician (and I mean statistician here, not any old computer science graduate friend with a hip job title like Data Samurai).

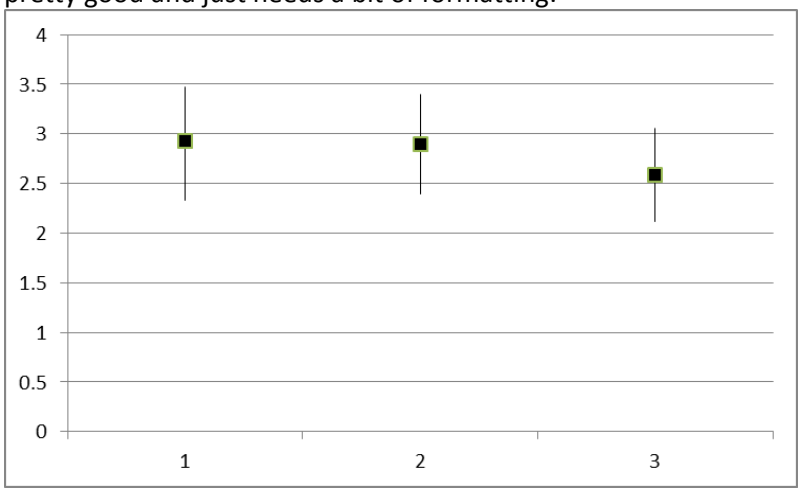
Now, highlight the numbers in upper, lower and mean, go to Insert and choose other chart types, where you should see the "high-low-close" option



This will read the data in columns but we need to switch it to rows:



Right click on the chart, go to Select data, and then click 'Switch Row/Column'. Now it should look pretty good and just needs a bit of formatting:



Mission accomplished; you now have a plot of the mean and its 95% CI in each group.