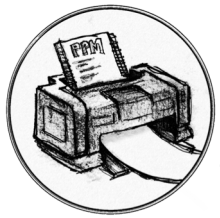
Descriptives and Frequencies

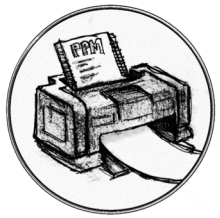
These are early analyses used to describe your entire sample. "Descriptives" gives you means and standard deviations of continuous data. "Frequencies" gives you percentages of categorical data. You'll use these values to create your Table 1.



Example of a Table 1:

Table 1. *Independent Samples T-Test. Demographic data, cardiometabolic parameters, and medical history.*

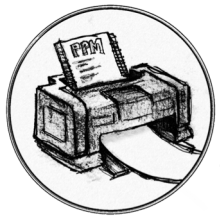
	Total Population			
Diabetes *(% yes)*	14.6%	**Non- Diabetic**	**Diabetic**	***P***
Age *(years)*	52.5 ± 22.5	49.9 ± 22.6	66.9 ± 15.2	<0.001
Sex (% male)	55.5%	56.4%	50.0%	0.034
Hemoglobin *(g/dL)*	13.5 ± 1.9	13.7 ± 1.9	12.7 ± 2.0	<0.001
Oximetry *(%)*	97.0 ± 2.9	97.2 ± 2.8	96.0 ± 3.5	<0.001
INR	1.1 ± 0.6	1.1 ± 0.5	1.3 ± 0.8	<0.001
Systolic BP *(mmHg)*	139.6 ± 24.0	138.2 ± 23.1	147.2 ± 27.3	<0.001
MAP *(mmHg)*	99.1 ± 20.7	100.6 ± 15.3	103 ± 17.0	0.015
Pulse Pressure *(mmHg)*	56.8 ± 20.5	56.4 ± 18.1	66.1 ± 22.3	<0.001



Another example of a Table 1:

Table 1. Demographic and anthropometric data, cardiometabolic parameters, and medical history.

Variable	Total Sample	Non-completers	Completers	*P*			
Sex (% female)	61.2%	65.6%	59.0%	0.666	**Males**	**Females**	***P***
HbA1c *(%)*	7.0 ± 1.1	7.2 ± 1.4	6.9 ± 0.9	0.234	7.3 ± 1.4	6.8 ± 1.0	0.117
Age *(years)*	68.3 ± 10.7	69.4 ± 12.4	67.4 ± 9.3	0.449	69.9 ± 9.8	67.2 ± 11.1	.0311
Race (% white)	66.7%	66.7%	66.7%	1.000	78.9%	60.0%	0.145
Hypertension *(% yes)*	66.7%	77.4%	59.0%	0.104	69.2%	65.0%	0.727
Hyperlipidemia *(% yes)*	53.8%	54.8%	55.3%	0.790	57.7%	51.3%	0.618
Heart Attack *(% yes)*	18.2%	16.1%	20.5%	0.562	30.8%	10.0%	0.053
Body Fat *(%)*	39.3 ± 6.9	40.9 ± 6.1	38.2 ± 7.3	0.136	33.5 ± 6.0	43.0 ± 4.5	**<0.001**
Obesity *(% yes)*	59.7%	62.5%	59.0%	0.888	65.4%	56.1%	0.458
BMI *(kg/m$^{2)}$*	32.3 ± 6.7	33.3 ± 7.7	31.6 ± 5.9	0.312	32.5 ± 5.7	32.2 ± 7.4	0.859
SBP *(mmHg)*	128.3 ± 11.9	127.3 ± 10.8	129.0 ± 12.7	0.561	127.9 ± 14.1	128.5 ± 10.4	0.857
DBP *(mmHg)*	75.0 ± 8.4	75.1 ± 9.0	78.4 ±13.9	0.917	74.6 ± 8.8	75.2 ± 8.2	0.777
HR *(bpm)*	77.0 ± 14.3	78.4 ± 13.9	78.0 ± 14.7	0.915	76.6 ±17.4	79.1 ± 12.1	0.517
Timed up-and-go *(s)*	7.3 ± 2.7	8.3 ± 3.2	7.3 ± 2.2	0.533	8.9 ± 3.8	7.5 ± 1.5	0.079
Back Scratch *(in)*	-11.9 ± 6.3	-13.3 ± 6.8	-10.9 ± 5.7	0.124	-15.4 ± 6.2	-9.7 ± 5.2	**<0.001**
Grip Strength *(kg)*	59.5 ± 21.3	57.6 ± 19.9	61.0 ± 22.4	0.520	75.8 ± 19.6	49.2 ± 15.0	**<0.001**
QoL Score	59.8 ± 17.6	59.7 ± 17.3	59.8 ± 18.0	0.972	59.9 ± 16.0	59.7 ± 18.6	0.965



Last example of a Table 1:

Table 1: *Patient baseline characteristics*

	Total	Rapid-Acting Insulin	Not Using	Sig.
N	593	26	567	
Sex (% male)	27.2%	42.3%	26.5%	$P = 0.076$
Age (years)	80.0 ± 9.1	78.2 ± 6.9	80.0 ± 9.2	$P = 0.210$
% patients age ≥ 80 years	52.5%	34.6%	52.7%	$P = 0.071$
Self-reported poor balance	2.9%	0.0%	3.0%	$P = 0.370$
Lightheadedness	13.5%	7.7%	13.8%	$P = 0.376$
Dementia	12.0%	23.1%	11.5%	$P = 0.074$
Cognitive struggles	24.6%	30.8%	24.3%	$P = 0.457$
Visual impairment	7.6%	7.7%	7.6%	$P = 0.984$
Head trauma	47.4%	42.3%	47.6%	$P = 0.596$
Concussion	2.2%	0.0%	2.3%	$P = 0.435$
Cervical injury	4.2%	3.8%	4.2%	$P = 0.924$
Back/spine injury	6.7%	11.5%	6.5%	$P = 0.319$
Hip/pelvis injury	7.6%	7.7%	7.6%	$P = 0.984$
Knee injury	9.4%	11.5%	9.3%	$P = 0.709$
Multiple previous falls	44.2%	73.1%	26.9%	**$P = 0.002$**
Number of previous falls	1.9 ± 1.3	2.6 ± 1.4	1.9 ± 1.3	**$P = 0.005$**
Return visit for fall	33.7%	53.8%	32.8%	**$P = 0.026$**
Number of return visits	0.5 ± 0.9	0.9 ± 1.0	0.5 ± 0.9	$P = 0.074$
Total Number of Falls	2.4 ± 2.1	3.5 ± 2.1	2.4 ± 2.1	**$P = 0.011$**

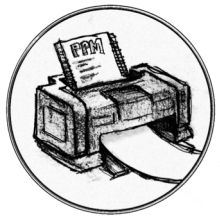
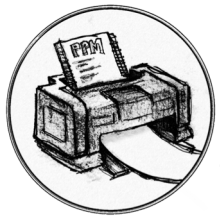


Table 1 is a characterization of your sample.

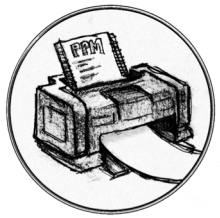
You'll describe the whole sample (descriptives and frequencies) and subsamples (t-tests and chi-square analyses). There is no p-value associated with descriptives or frequencies since no comparison is being made. It's just the averages in the sample. Whole sample. Not the *population* (i.e., everyone), but the *sample* (the group of people you analyzed, who were sampled *from* the population). Whenever there's a p-value, that means you're making inferences about the larger population (i.e., "inferential statistics"). Reporting basic characterizations of your sample (descriptives and frequencies) isn't inferring anything about the larger population. It's just describing the sample you tested.

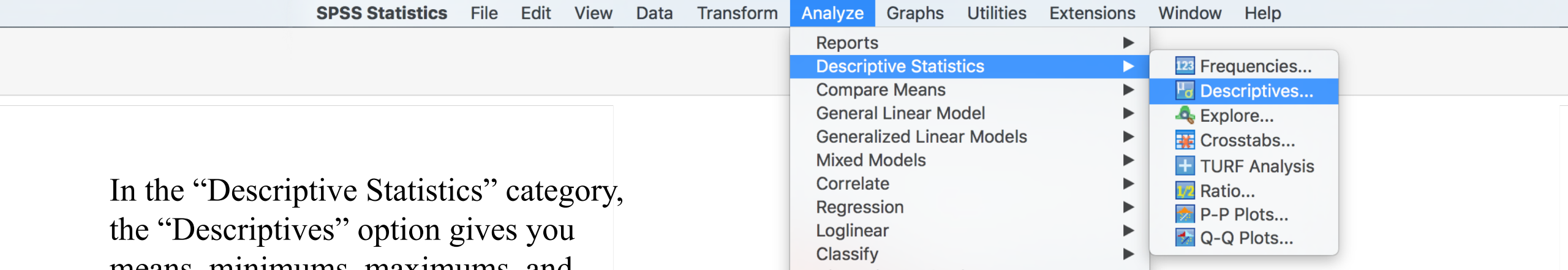


SPSS Statistics File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

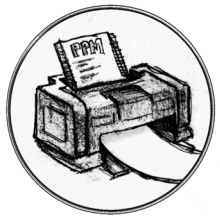
See the menu bar?

All analyses start in the Analyze tab.

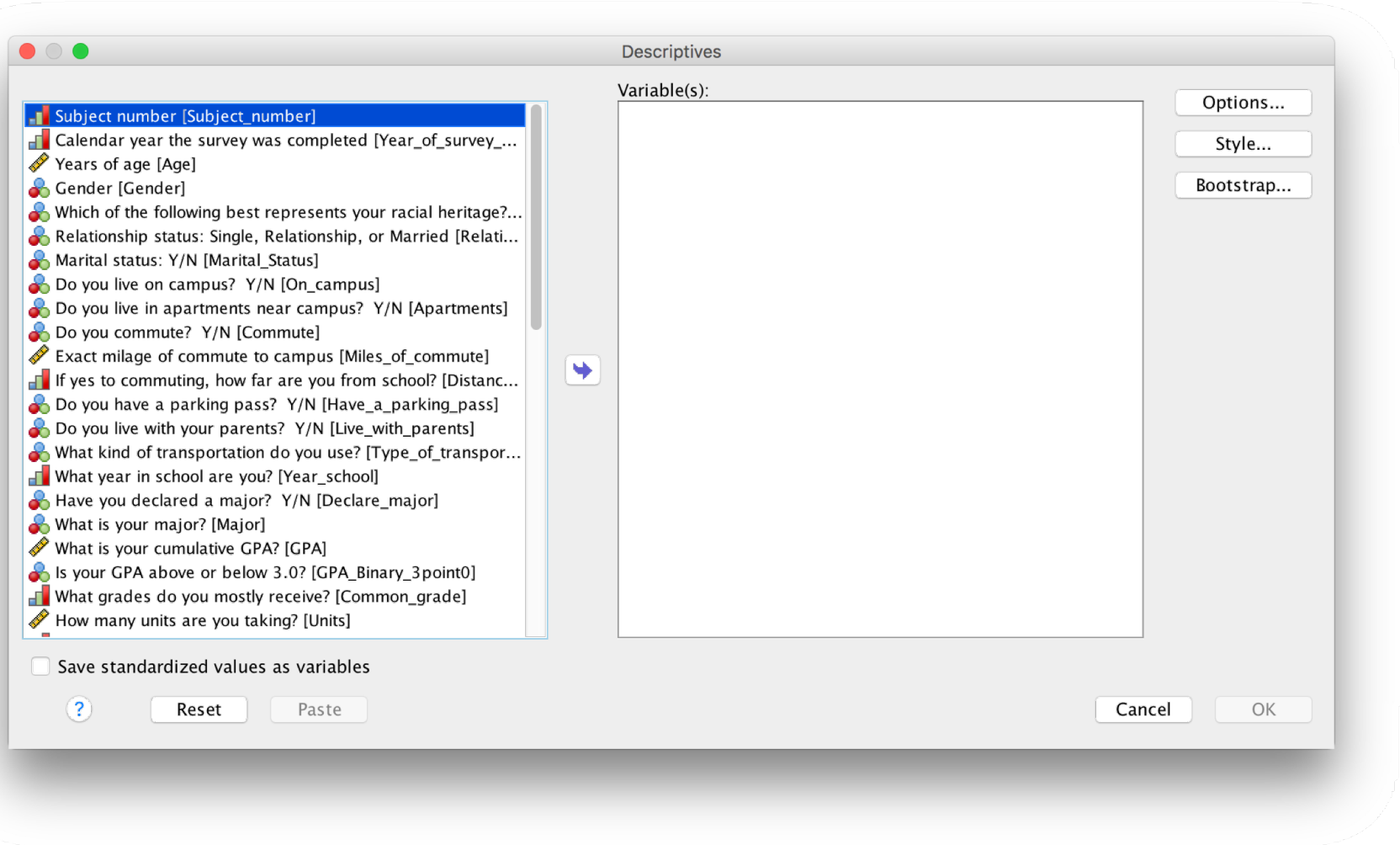


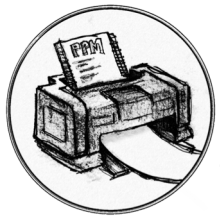


In the "Descriptive Statistics" category, the "Descriptives" option gives you means, minimums, maximums, and standard deviations.



This box will appear. Every variable in your database is in the left column. Move all of the continuous (i.e., "scale") variables of interest to the right column; don't move any categorical variables. The mean of bodyweight or age (continuous) makes sense. The mean of gender (categorical) doesn't.

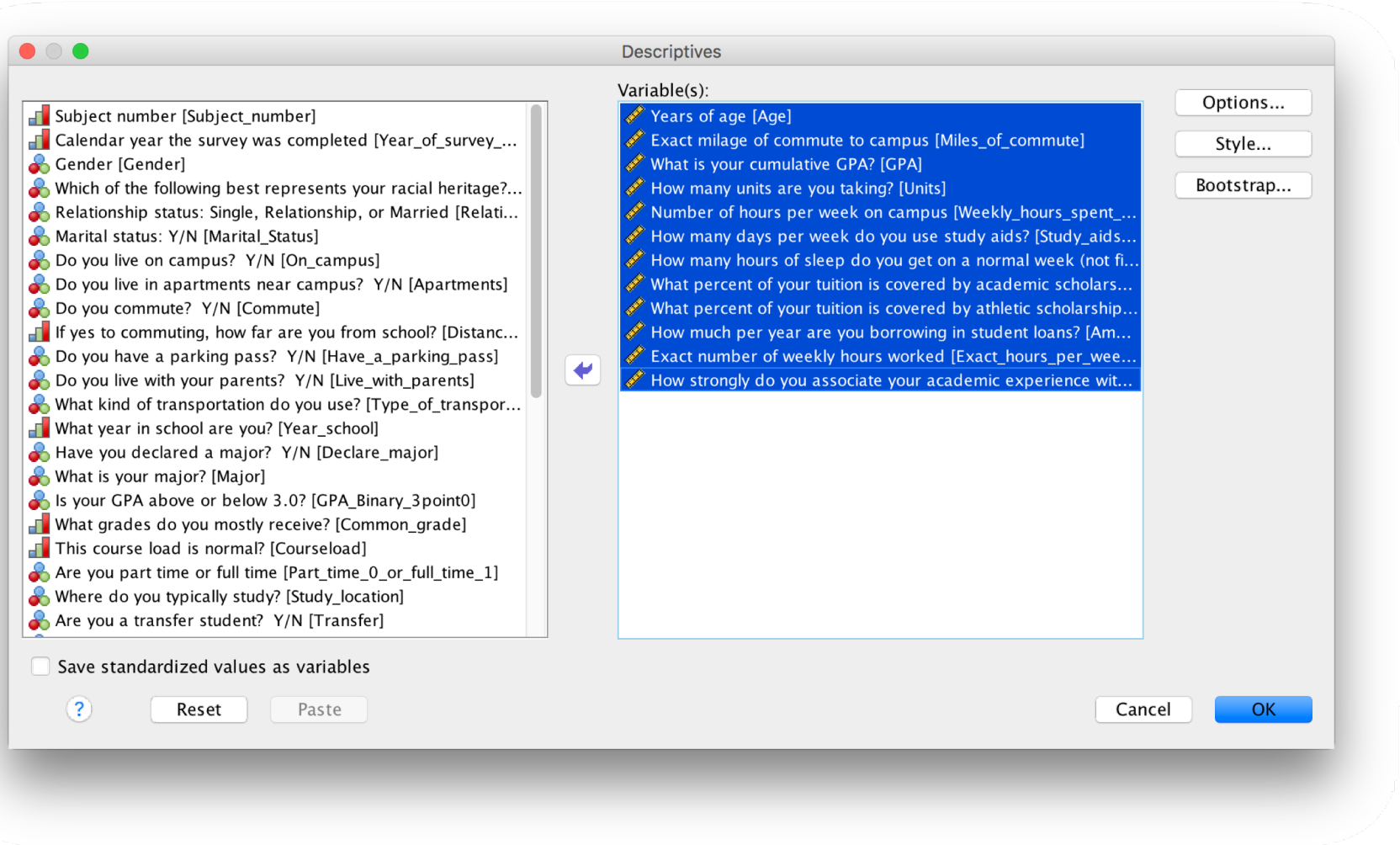




Scale variables have an icon that looks like a little ruler.

Drag over any variable you want values for.

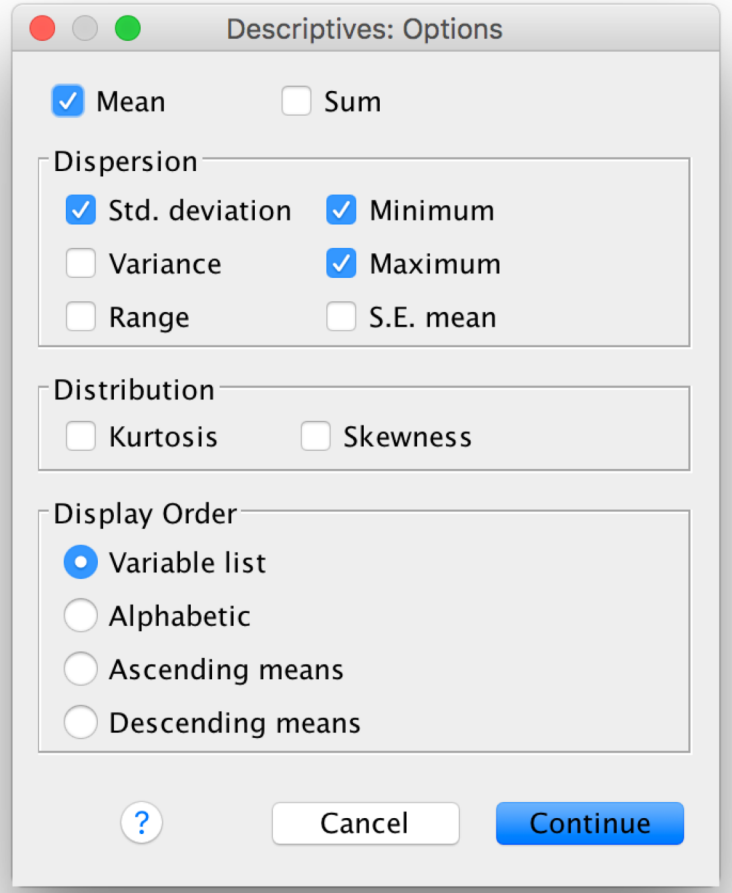
If you hit OK, it'll give you those basic values (min, max, mean, SD).



If you want additional data (variance, skewness, kurtosis), select Options. This box appears.

Select any additional outputs.

Then click Continue, then OK.

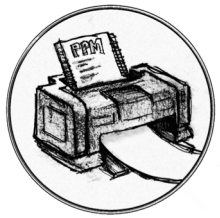


The basic outputs look like this:

If you selected additional data options, those will be reported, too.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Years of age	104	17	25	19.69	1.495
Exact milage of commute to campus	110	.0	55.0	4.150	9.6564
What is your cumulative GPA?	49	2.05	4.00	3.2663	.47827
How many units are you taking?	110	3	20	15.67	2.427
Number of hours per week on campus	61	0	168	74.56	66.147
How many days per week do you use study aids?	49	.0	7.0	1.714	2.3805
How many hours of sleep do you get on a normal week (not finals week or when you have midterms)	110	3	10	6.57	1.078
What percent of your tuition is covered by academic scholarships?	48	0	1	.35	.337
What percent of your tuition is covered by athletic scholarships?	49	0	1	.02	.143
How much per year are you borrowing in student loans?	47	0	60000	9808.52	13540.815
Exact number of weekly hours worked	110	0	40	6.59	7.893
How strongly do you associate your academic experience with financial concerns?	49	2.0	10.0	6.531	2.2370
Valid N (listwise)	0				

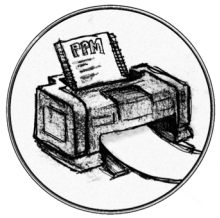


This is where the means and standard deviations go in a Table 1 (in the column that describes your entire sample).

Report data as mean plus or minus the standard deviation.

Table 1: *Subject baseline characteristics*

	Total	Men	Women	Sig.
N	45			
Age (years)	20.2 ± 0.7			
BMI (kg/m^2)	26.5 ± 2.8			
GPA	2.9 ± 0.7			
Nightly sleep (hours)	7.5 ± 1.4			
Employed (%)				
Weekly work (hours)	4.0 ± 5.2			
Academic Scholarship (%)				
Athletic Scholarship (%)				



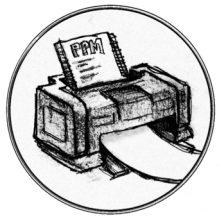
Done with Descriptives.

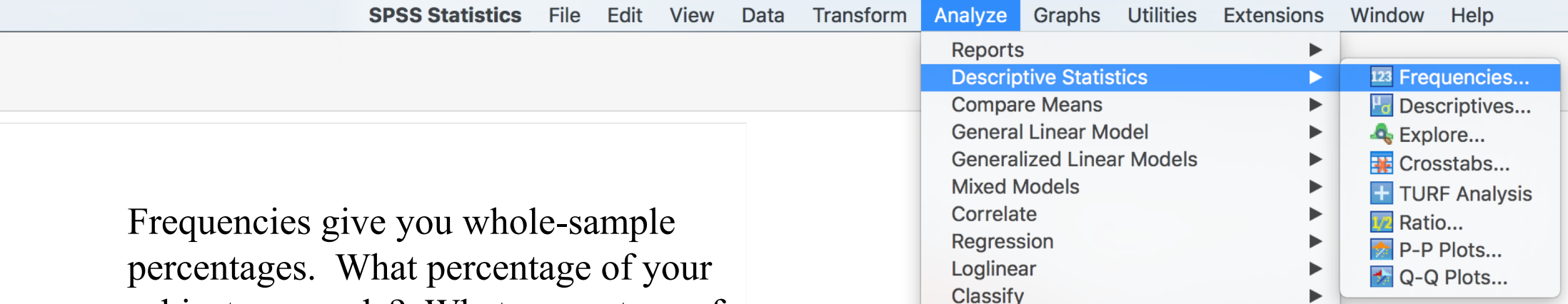
Next: Frequencies.

That's how you get your percentages. And this is where you put them in your Table 1:

Table 1: *Subject baseline characteristics*

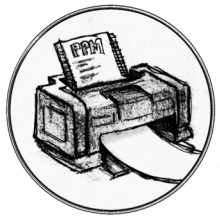
	Total	Men	Women	Sig.
N	45			
Age (years)	20.2 ± 0.7			
BMI (kg/m²)	26.5 ± 2.8			
GPA	2.9 ± 0.7			
Nightly sleep (hours)	7.5 ± 1.4			
Employed (%)				
Weekly work (hours)	4.0 ± 5.2			
Academic Scholarship (%)				
Athletic Scholarship (%)				





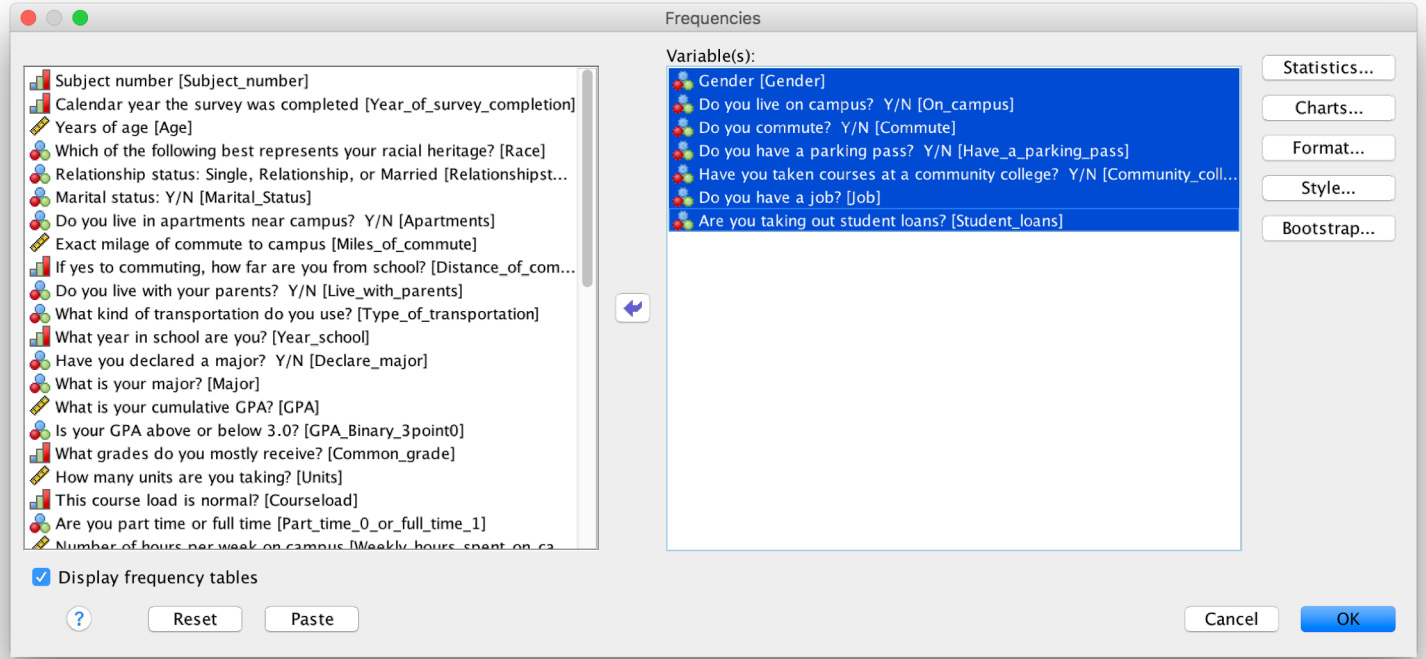
Frequencies give you whole-sample percentages. What percentage of your subjects are male? What percentage of your subjects are employed? Or some race. Or belong to some political party. What percentage of your subjects are over the age of 65? What percentage have killed a bird with a circular saw? Percentages. That's what "Frequencies" gives you.

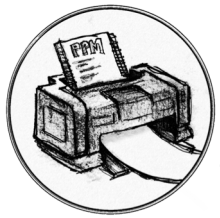
Here's where to run it in the menu:



Drag categorical variables of interest from the left box (all of your variables) to the right box (the ones you'll be getting data on).

If percentages are all you need, select OK. If additional data are needed, select the "Statistics" option.

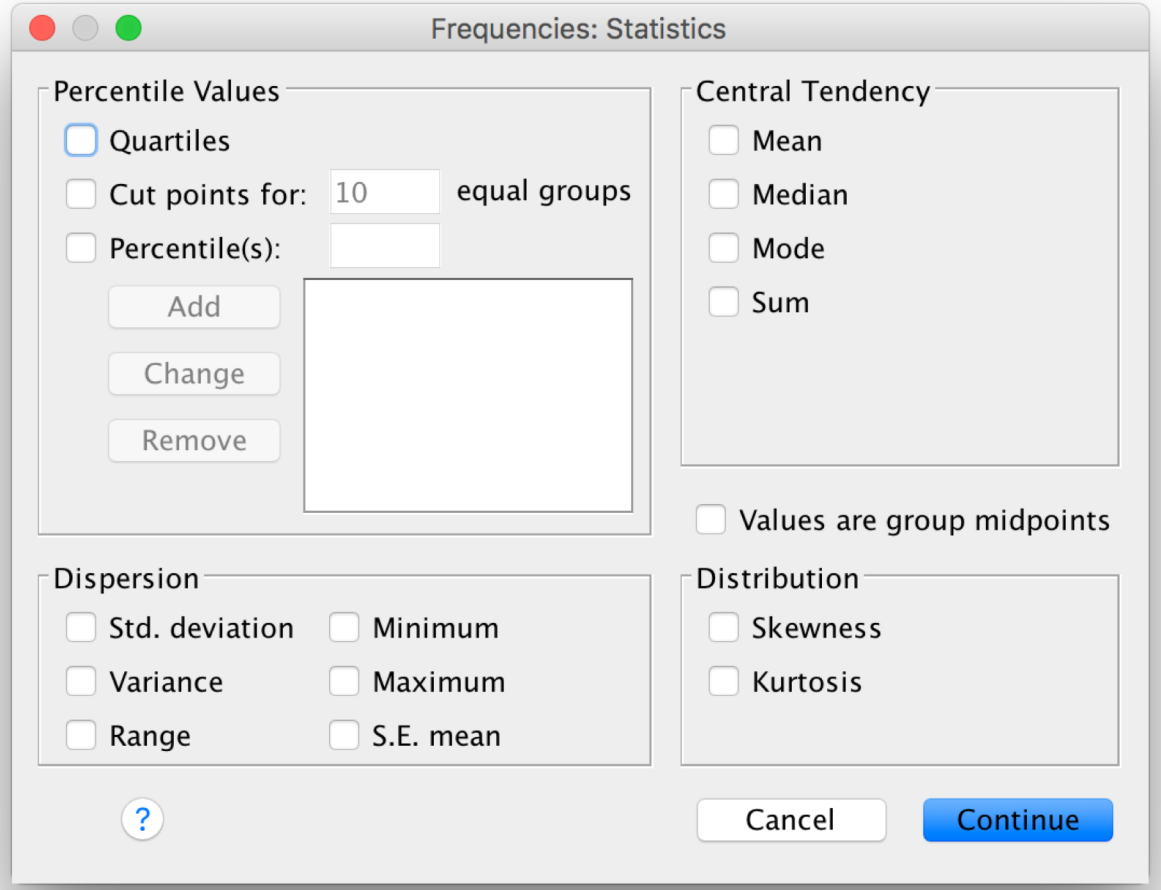


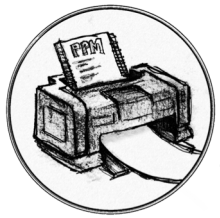


PacificPaperMill.com

In the Statistics box, there are other measurements of central tendency (median and mode) that can be reported.

If you want those, select them, click Continue, then click OK.

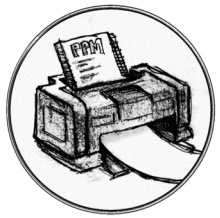




Outputs for basic frequencies:

"Percent" and "Valid Percent" are the same number if there are no missing values.

If there are missing values (e.g., a subject didn't have their sex or ethnicity coded), then "Valid Percent" will be a higher number, as it eliminates that subject from the calculation.



PacificPaperMill.com

Frequencies

Statistics

		Gender	Do you live on campus? Y/N	Do you commute? Y/N	Do you have a parking pass? Y/N	Have you taken courses at a community college? Y/N	Do you have a job?	Are you taking out student loans?
N	Valid	110	110	110	61	49	110	49
	Missing	0	0	0	49	61	0	61

Frequency Table

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	29	26.4	26.4	26.4
	Female	81	73.6	73.6	100.0
	Total	110	100.0	100.0	

Do you live on campus? Y/N

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	47	42.7	42.7	42.7
	Yes	63	57.3	57.3	100.0
	Total	110	100.0	100.0	

Do you commute? Y/N

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	72	65.5	65.5	65.5
	Yes	38	34.5	34.5	100.0
	Total	110	100.0	100.0	

Do you have a parking pass? Y/N

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No Parking Pass	24	21.8	39.3	39.3
	Have a parking pass	37	33.6	60.7	100.0
	Total	61	55.5	100.0	
	System	49	44.5		
		110	100.0		

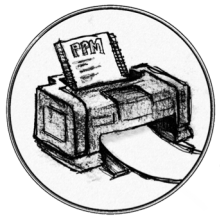
It doesn't have to be dichotomous variables. Frequencies can be generated for any number of categories (e.g., political party, year in school, astrological sign).

Or they can be generated for continuous data, although that output will be huge (but that's how you get medians and modes).

Frequency Table

Astro_Sign

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unknown	8	11.8	24.2	24.2
	Aries	3	4.4	9.1	33.3
	Taurus	2	2.9	6.1	39.4
	Gemini	1	1.5	3.0	42.4
	Cancer	3	4.4	9.1	51.5
	Leo	4	5.9	12.1	63.6
	Virgo	4	5.9	12.1	75.8
	Libra	2	2.9	6.1	81.8
	Scorpio	2	2.9	6.1	87.9
	Sagit	4	5.9	12.1	100.0
	Total	33	48.5	100.0	
Missing	System	35	51.5		
Total		68	100.0		

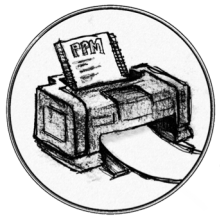


When you have your relevant percentages, add them to Table 1. That completes the left column (the column that describes the entire sample… not the whole *population*, but the whole sample).

The rest of the table is for comparison statistics, as in: are subsamples different from each other?

Table 1: *Subject baseline characteristics*

	Total	Men	Women	Sig.
N	45			
Age (years)	20.2 ± 0.7			
BMI (kg/m²)	26.5 ± 2.8			
GPA	2.9 ± 0.7			
Nightly sleep (hours)	7.5 ± 1.4			
Employed (%)	22.5%			
Weekly work (hours)	4.0 ± 5.2			
Academic Scholarship (%)	12.9%			
Athletic Scholarship (%)	11.5%			

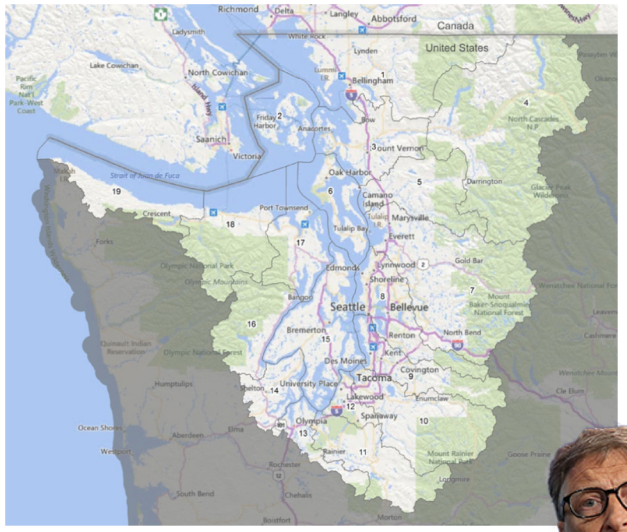


Although percentages and means are typically what get reported to characterize central tendency, median (middle value) and mode (most common value) are occasionally appropriate measurements.

Imagine you're estimating the average number of fingers on a human hand. The *median* number of fingers is a way to estimate central tendency in the presence of extreme numbers. Someone who has no fingers (presumably no arms) would reduce the mean in a sample of 10 people, but not the median. Likewise, if someone had lots of additional fingers – for example: if Count Rugen had fifty more fingers growing out of his pelvis – the mean would be affected, but the median wouldn't. The median gives you an idea of what the average would be *without* outliers.

Although percentages and means are typically what get reported to characterize central tendency, median (middle value) and mode (most common value) are occasionally appropriate measurements.

The Puget Sound region of Washington has >150 different zip codes. What is the average household net worth of different Puget Sound zip codes?



When you compare the mean value of 98039 (Medina) to 98110 (Bainbridge Island), Medina is higher. However, if the *median* value is used in the place of the mean, the average net worth of Medina households is decreased by about $45 million. How?

Puget Sound BUSINESS JOURNAL

The Sound's wealthiest zip codes

By Jeanne Lang Jones – Staff Writer

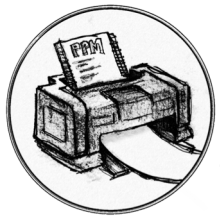
Feb 6, 2005, 21:0pm **Updated** Feb 3, 2005, 13:6pm

Although percentages and means are typically what get reported to characterize central tendency, median (middle value) and mode (most common value) are occasionally appropriate measurements.

If a parent is trying to estimate the risk she's subjecting her child to by signing him up for football, she might want to know the average number of times per game that someone is carried off of the field in a stretcher. In a 10-game season, last year's per-game stretcher data were: 0, 0, 0, 0, 0, 0, 1, 1, 2, 6. The median and mode both = 0. But "on average, 0 athletes are carried off of the field in a stretcher" is misleading since the actual number of stretcher episodes is 10 in a 10-game season.

Telling a parent 0 will communicate a very different message from "once per game on average; it could be your kid… you never know."

Select your measurement of central tendency (usually mean, occasionally medium, seldom mode) by its appropriateness to the research question.



Final, final example of a Table 1:

Table 1. Demographic and anthropometric data, cardiometabolic parameters, and medical history.

Variable	**Total Population**	**Group Comparisons**					
Sex	55.5% male						
		Men	**Women**	***P***			
Depression (%)	1.1%	0.8%	1.5%	0.117	**Depressed**	**Non-Depressed**	***P***
Hemoglobin (g/dL)	13.5 ± 1.9	14.2 ± 1.8	12.6 ± 1.7	<0.001	12.2 ± 1.9	13.5 ± 1.9	<0.001
Age (years)	52.5 ± 22.5	48.5 ± 20.4	57.5 ± 23.9	<0.001	73.5 ± 17.7	52.3 ± 22.4	<0.001
Obesity (%)	16.6%	16.7%	16.6%	0.966	8.0%	16.7%	0.244
Mean Arterial Pressure (mmHg)	99.1 ± 20.7	100.3 ± 19.0	97.5 ± 22.6	0.002	93.4 ± 31.2	99.1 ± 20.6	0.167
Pulse (bpm)	87.4 ± 16.9	87.6 ± 17.1	87.2 ± 16.7	0.507	89.7 ± 20.3	87.4 ± 16.9	0.511
Oximetry (% Saturation)	97.0 ± 2.9	97.1 ± 2.7	96.9 ± 3.2	0.038	95.5 ± 4.0	97.0 ± 2.9	0.079
Blood pH	7.3 ± 0.4	7.4 ± 0.2	7.3 ± 0.6	0.338	7.4 ± 0.1	7.3 ± 0.4	0.732
Diabetes Mellitus (%)	14.6%	13.2%	16.4%	0.032	12.0%	14.6%	0.712
Cancer (%)	0.9%	0.7%	1.1%	0.343	0.0%	0.9%	0.631
Respiratory Disease (%)	11.3%	8.5%	14.9%	<0.001	4.0%	11.4%	0.245
Bleeding Disorder (%)	10.9%	8.4%	14.1%	<0.001	24.0%	10.8%	0.035
Cirrhosis (%)	0.5%	0.6%	0.5%	0.842	0.0%	0.6%	0.710
Dialysis (%)	0.5%	0.4%	0.7%	0.334	0.0%	0.6%	0.710
Myocardial Infarction (%)	2.1%	2.5%	1.7%	0.245	0.0%	2.2%	0.458
Cerebrovascular Accident (%)	3.8%	2.7%	5.2%	0.002	16.0%	3.7%	0.001
Pregnant (%)	0.2%	0.0%	0.4%	0.025	0.0%	0.2%	0.830