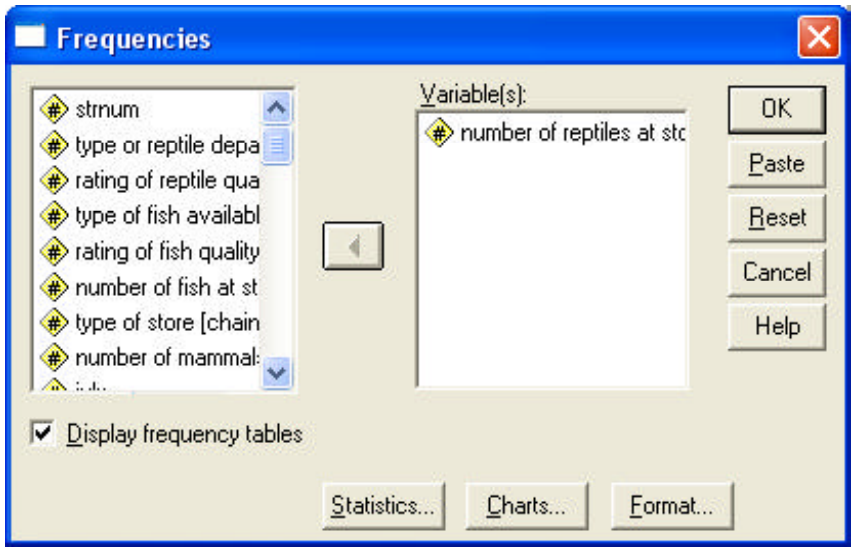
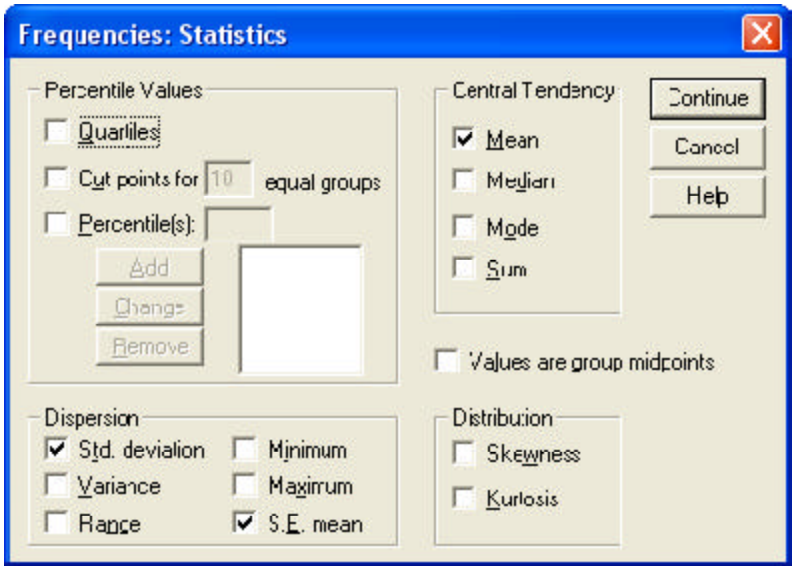
Frequency Distributions, Mean and Standard Deviation -- Univariate Statistics

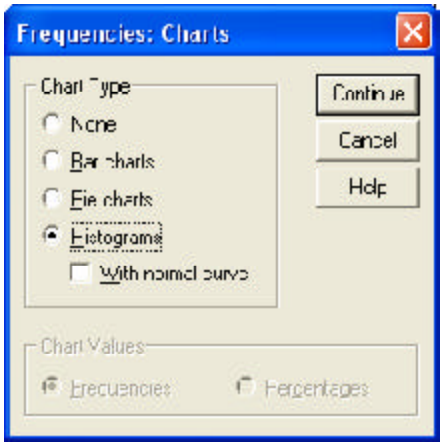
Application: To obtain a graph and a summary of the distribution of scores.

Analyze/Statistics ➔ Descriptive Statistics ➔ Frequencies

- highlight each desired variable and click the arrow button
- Click “Statistics” — indicate the univariate statistics you want computed (get mean, std dev, S.E. mean)
- Click “Charts” -- click “Histogram” for quantitative variables and “Bar Chart” for qualitative variables







'number of reptiles at store'

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4	3	25.0	25.0	25.0
	5	1	8.3	8.3	33.3
	7	1	8.3	8.3	41.7
	10	2	16.7	16.7	58.3
	12	2	16.7	16.7	75.0
	14	2	16.7	16.7	91.7
	15	1	8.3	8.3	100.0
	Total	12	100.0	100.0	

Statistics

'number of reptiles at store'

N	Valid	12
	Missing	0
Mean		9.25
Std. Error of Mean		1.23
Std. Deviation		4.27

Chart
Gives a quick picture of the data. Can be used to look for noncontinuities (gaps) and skewness (asymmetry) in the distribution.

Frequency Listing
Tells the number of occurances of each value for this variable (e.g., there were 3 stores with 4 reptiles and 2 stores with 10

If there are missing values, then you should use the "Valid Percent" column to find the % of cases in each value/category.

Quantitative data -- mean and standard deviation are meaningful.

Combining the mean and std (see below) tells us that...

about 68% of stores in this population display between 4.983 & 13.517 reptiles

about 96% of the stores display between .716 & 17.784 reptiles

'type of reptile department'

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	not separate	6	50.0	50.0	50.0
	separate	6	50.0	50.0	100.0
	Total	12	100.0	100.0	

Statistics

'type of reptile department'

N	Valid	12
	Missing	0
Mean		1.50
Std. Error of Mean		.15
Std. Deviation		.52

'type of store'

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	'chain store'	5	41.7	41.7	41.7
	privately owned	3	25.0	25.0	66.7
	coop	4	33.3	33.3	100.0
	Total	12	100.0	100.0	

Statistics

'type of store'

N	Valid	12
	Missing	0
Mean		1.92
Std. Error of Mean		.26
Std. Deviation		.90

Binary variables -- The decimal part of the mean will tell you the proportion of the scores in the distribution with the higher coded value.

The mean gives useful information because there is only one combination of two conditions of a binary variable that will yield this particular mean (i.e., only a 50%-50% combination of these two types of departments will give a mean of 1.5). Any other combination would give a different mean (e.g., 25% not separate and 75% separate would give a mean = 1.75, compared to a combination of 75% not separate and 25% separate would give a mean = 1.25)

For these data...

Delete the whole number part of the mean, leaving .5
which tell us that .5 (1/2, 50%) of the sample had separate reptile departments

Qualitative variables-- mean and standard deviation are **not meaningful,** because there are multiple combinations of proportions of the three conditions that will yield this particular mean.

For example, a different mix of these frequencies, say 5 privately owned stores, 3 chain stores and 4 coops would also give a mean of 1.917. In fact, any combination that has the same sum ((5 coded 1 = 5) + (3 coded 2 = 6) + (4 coded 3 = 12) = 23) will have the same mean (23 / 12 = 1.917).

Thinking about the Standard Deviation and the Standard Error of the Mean

The standard deviation (Std.) tells about the expected variation of individual scores in the distribution around the mean of that distribution. The standard error of the mean (SEM) tells about the expected variation in estimates of the mean of the population across multiple samplings of a specific size.

Huh? We're used to drawing a single sample and treating the mean of that sample as our estimate of the population mean. We know that estimate isn't exactly accurate, but by how much is likely to be off? If we took multiple samples from the same population and computed the mean for each, we could compute the standard deviation of those means, as an index of the expected variability in sample estimates of the population mean. The standard error of the mean gives us this information, even if we have only one sample. How does this work? Assuming the data are normally distributed, we can estimate the variability of multiple sample means around the true population mean from the standard deviation of a single sample (the more variability there is among the individuals, the more variability there is likely to be among the sample means - get it !!). That is why the formula for the standard error (SE = std / $\sqrt{N}$) is based on the standard deviation and the sample size.

Example of a Table to show univariate statistics for a mix of quantitative and qualitative variables.

Table 1
Summary of Story Type, Reptile Department Type and Number of Reptiles

Variable	Univariate Statistics		
Number of reptiles	M = 9.25	S = 4.27	N = 12
Type of Reptile Department	Separate Department		6 (50%)
	Not Separate Department		6 (50%)
Type of Store	Part of National Chair		5 (41.7%)
	Privately Owned		3 (25%)
	Member of Regional Coop		4 (33.3%)

Be sure to use the “valid %” w:hen presenting the % for each category of a qualitative variable.