```
*** ================================================================================ ***.
***   The data set contains part of the data for a study of oral condition of cancer patients
***   conducted at the Mid-Michigan Medical Center.  The oral conditions of the patients were
***   measured and recorded at the initial stage, at the end of the second week, at the end of
***   the fourth week, and at the end of the sixth week.  The variables age, initial weight
***   and initial cancer stage of the patients were recorded. (n = 25 patients w/neck cancer)
***        Patients were divided into two groups at random:
***           One group received a placebo and the other group received aloe juice treatment.
================================================================================ ***.
```

```
*** -------------------------------------------------------------------------- ***.
*** Do some basic prep work
*** -------------------------------------------------------------------------- ***.
```

```
VARIABLE LABELS
ID      "Identification number"
TRT     "Treatment Group"
AGE     "Incoming Age"
WEIGHIN "Incoming Weight in pounds"
STAGE   "Stage of Cancer".

VALUE LABELS
TRT
   0 "control"
   1 "aloe treatment"
   999 "missing"/
AGE WEIGHIN STAGE
TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6
   999 "missing".

MISSING VALUES
TRT AGE WEIGHIN STAGE
TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6
   (999).

RECODE
TRT AGE WEIGHIN STAGE
TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6
   (SYSMIS = 999).
EXECUTE.

FORMATS
ID TRT AGE WEIGHIN STAGE
TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6 (F3).
```

← *recodes into the **SAME** variable*

Prep Work Steps

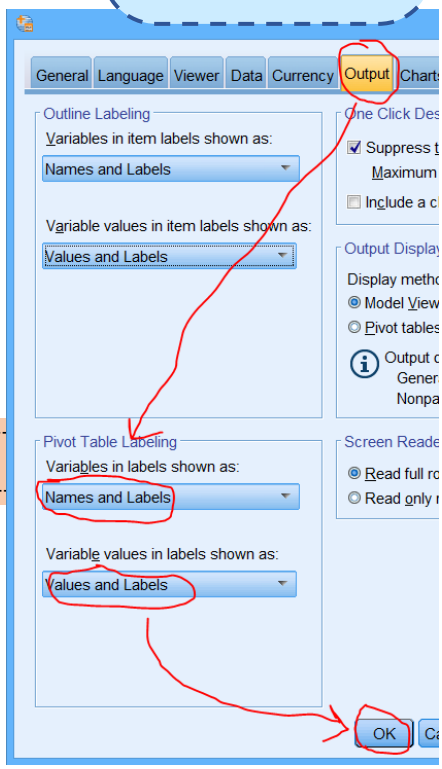
1. Declaring:
   - Variable labels
   - Value labels
   - Missing Values
2. 'plugging' blank values with missing codes
3. Format *(optional)*

From the menu bar

"Edit" > "Options"

Tab for "Output"

Change drop down selections




```
*** -------------------------------------------------------------------------- ***.
*** Compute some new variables.
*** -------------------------------------------------------------------------- ***.

NUMERIC weight_oz total_add total_sum total_avg total_mean (F2).

VARIABLE LABELS
weight_oz "Incoming Weight in ounces"
total_add  "TOTAL: ADD using the PLUS symbol"
total_sum  "TOTAL: ADD using the SUM command"
total_avg  "TOTAL: AVG using the PLUS & DIVIDE symbols"
total_mean "TOTAL: AVG using the MEAN command".

COMPUTE weight_oz  = 16 * weighin.

COMPUTE total_add  = TOTALCIN + TOTALCW2 + TOTALCW4 + TOTALCW6.
COMPUTE total_sum  = sum(TOTALCIN, TOTALCW2, TOTALCW4, TOTALCW6).
COMPUTE total_avg  = total_add / 4.
COMPUTE total_mean = mean(TOTALCIN, TOTALCW2, TOTALCW4, TOTALCW6).
EXECUTE.
```

Same math, BUT handle missing values very differently

```
*** ================================================================================ ***.
*** ================================================================================ ***.
*** CHAPTER 2: VISUALIZE YOUR DATA
*** ================================================================================ ***.
*** ================================================================================ ***.
```

```
*** -------------------------------------------------------------------------------- ***.
*** Creating FREQUENCY distribution tables.
*** -------------------------------------------------------------------------------- ***.
```

```
* One variable or you can do several variables at once

FREQUENCIES TRT.

FREQUENCIES AGE STAGE total_add.
```

TRT Treatment Group

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0 control	14	56.0	56.0	56.0
	1 aloe treatment	11	44.0	44.0	100.0
	Total	25	100.0	100.0	

```
* you include a split file to make two tables,
   by another variable.

SORT CASES by TRT.
SPLIT FILE by TRT.
FREQUENCIES STAGE.
SPLIT FILE off.
SORT CASES by ID.
```

STAGE Stage of Cancer

TRT Treatment Group			Frequency	Percent	Valid Percent	Cumulative Percent
0 control	Valid	1	8	57.1	57.1	57.1
		2	3	21.4	21.4	78.6
		3	1	7.1	7.1	85.7
		4	2	14.3	14.3	100.0
		Total	14	100.0	100.0	
1 aloe treatment	Valid	0	1	9.1	9.1	9.1
		1	4	36.4	36.4	45.5
		2	3	27.3	27.3	72.7
		4	3	27.3	27.3	100.0
		Total	11	100.0	100.0	

```
* Here is a nice way to do a
   simple crosstabulation of
   two categorical variables.

CROSSTABS STAGE BY TRT.
```

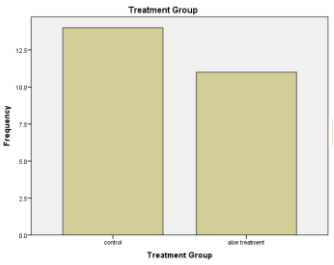
STAGE Stage of Cancer * TRT Treatment Group Crosstabulation

Count

		TRT Treatment Group		
		0 control	1 aloe treatment	Total
STAGE Stage of Cancer	0	0	1	1
	1	8	4	12
	2	3	3	6
	3	1	0	1
	4	2	3	5
Total		14	11	25

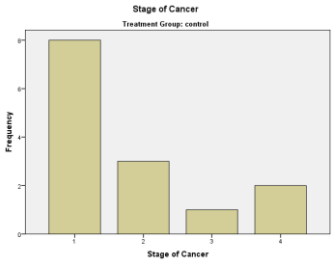
```
*** -------------------------------------------------------------------------------- ***.
*** Use the FREQUENCIES command to make BAR CHART's.
*** -------------------------------------------------------------------------------- ***.
```

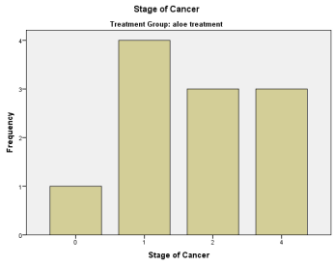
```
* You can just add the HISTOGRAM option after a slash..

FREQUENCIES TRT /BARCHART.
```



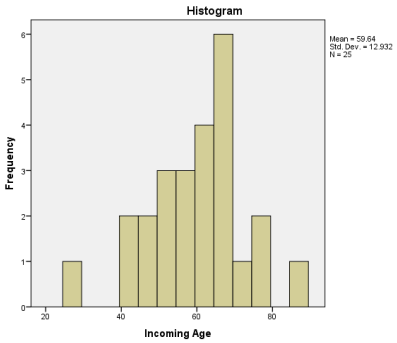

```
* or you can include the option to not bother with the table.

FREQUENCIES TRT STAGE
  /BARCHART
  /FORMAT notable.
```

```
* wrapping in the sort/split produces two separate histograms.

SORT CASES by TRT.
SPLIT FILE by TRT.
FREQUENCIES STAGE
   /BARCHART
   /FORMAT NOTABLE.
SPLIT FILE off.
SORT CASES by ID.
```

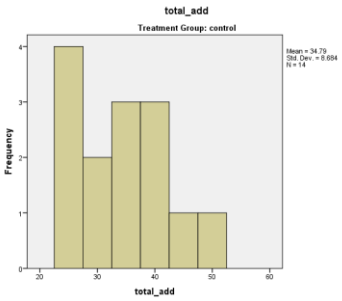


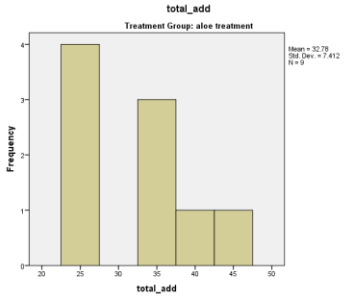


```
*** ------------------------------------------------------------------------------ ***.
*** Use the FREQUENCIES command to make HISTOGRAM's.
*** ------------------------------------------------------------------------------ ***.

* You can just add the HISTOGRAM option after a slash.

FREQUENCIES AGE /HISTOGRAM.
```




```
* or you can include the option to not bother with the table.

FREQUENCIES total_add
  /HISTOGRAM
  /FORMAT NOTABLE.

* wrapping in the sort/split produces two separate histograms
* not that they have different ranges on the x & y axis.
* This makes them hard to compare to each other.

SORT CASES by TRT.
SPLIT FILE by TRT.
FREQUENCIES total_add
  /HISTOGRAM
  /FORMAT NOTABLE.
SPLIT FILE off.
SORT CASES by ID.
```





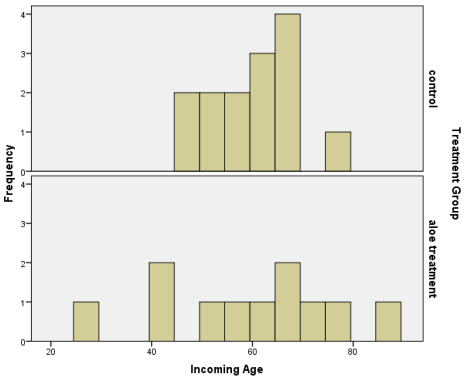

```
*** ------------------------------------------------------------------------------ ***.
*** You can also use the EXAMINE command to make HISTOGRAM's.
*** ------------------------------------------------------------------------------ ***.

* a simple histogram
   that includes a descriptive table by default.

EXAMINE AGE /PLOT HISTOGRAM.

* You can include an option to not bother with the table.

EXAMINE AGE
  /PLOT HISTOGRAM
  /STATISTICS NONE.
```

Descriptives

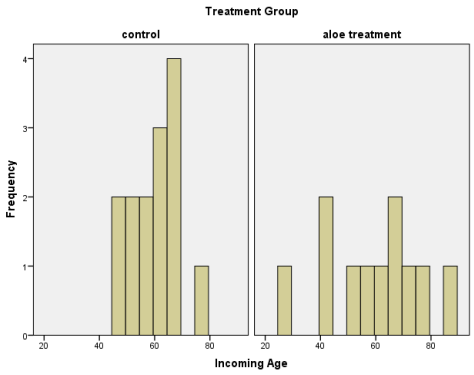
			Statistic	Std. Error
Incoming Age	Mean		59.64	2.586
	95% Confidence Interval for Mean	Lower Bound	54.30	
		Upper Bound	64.98	
	5% Trimmed Mean		59.92	
	Median		60.00	
	Variance		167.240	
	Std. Deviation		12.932	
	Minimum		27	
	Maximum		86	
	Range		59	
	Interquartile Range		16	
	Skewness		-.348	.464
	Kurtosis		.584	.902

```
* This produces a histogram of age separately for each level of TRT.
* Like using sort/split/freq, they have different ranges on the x & y axis.
* This makes them hard to compare to each other.

EXAMINE AGE BY TRT
  /PLOT HISTOGRAM
  /COMPARE GROUPS
  /STATISTICS NONE.


* If you are using the COMPARE BY option, you can also include the TOTAL option.
* This gives you results for all subjects first, followed by the groups separately.

EXAMINE AGE BY TRT
  /PLOT HISTOGRAM
  /COMPARE GROUPS
  /STATISTICS NONE
  /TOTAL.
```

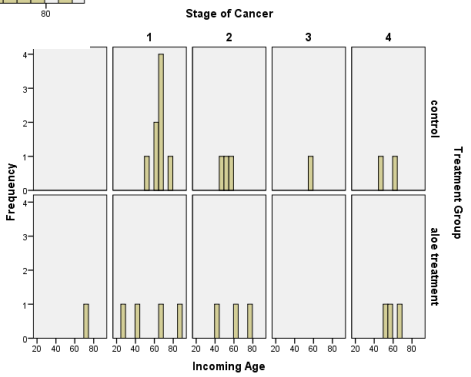
```
*** ------------------------------------------------------------------------------ ***.
*** You can also use the GRAPH command to make HISTOGRAM's.
*** ------------------------------------------------------------------------------ ***.

* Here is the basic, all default options.

GRAPH /HISTOGRAM AGE.

* a nice option this command has is to employ PANELs.
* notice that the stacked plots have the exact same ranges on the axes so you CAN COMPARE.

GRAPH
  /HISTOGRAM AGE
  /PANEL ROWVAR = TRT.
```




```
GRAPH
  /HISTOGRAM AGE
  /PANEL COLVAR = TRT.
```

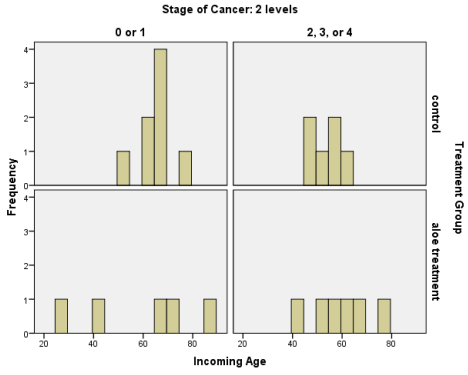



```
* You can do both rows & columns.

GRAPH
  /HISTOGRAM AGE
  /PANEL COLVAR = TRT COLVAR = STAGE.
```




```
*** ------------------------------------------------------------------------------ ***.
*** Create a RECODED version of a variable with RECODE-INTO
*** ------------------------------------------------------------------------------ ***.

* You can get fancy:
* Use NUMERIC to create a new variable that collapses the stage into two levels.
* Use LABELS to clearly tell yourself and others what is going ion.
* Use RECODE to convert values of the old variable INTO the new variable.
* You can check the number of subjects in each combination with CROSSTAB.
* Then do a grid of histograms with GRAPH.

NUMERIC          stage_2 (F3).
VARIABLE LABELS  stage_2 "Stage of Cancer: 2 levels".
VALUE LABELS     stage_2 0 "0 or 1" 1 "2, 3, or 4".

RECODE stage (0, 1 = 0)(2, 3, 4 = 1) INTO stage_2.
EXECUTE.

CROSSTABS stage BY stage_2.

CROSSTABS TRT BY stage_2.

GRAPH
  /HISTOGRAM AGE
  /PANEL ROWVAR = TRT COLVAR = stage_2.
```

← *recodes into a **DIFFERENT** variable*



```
*** ------------------------------------------------------------------------------ ***.
*** Use FREQUENCIES to find QUARTILE's, DECILE's, & PERCENTILE's
*** ------------------------------------------------------------------------------ ***.
```

```
* Quartiles split the data into four equal portions.

FREQUENCIES AGE  /NTILES(4).

* or you can include the option to not bother with the table.

FREQUENCIES AGE WEIGHIN total_add total_sum
  /FORMAT NOTABLE
  /NTILES(10) .
```

Incoming Age

N	Valid	25
	Missing	0
Percentiles	25	51.50
	50	60.00
	75	67.50

```
* To get quartiles and deciles, you must do 2 NTILE option lines.

FREQUENCIES AGE
  /FORMAT NOTABLE
  /NTILES(4)
  /NTILES(10).

* The PERCENTILE option allows you to specify a list...separated by commas.

FREQUENCIES AGE WEIGHIN
  /FORMAT NOTABLE
  /PERCENTILES(5, 50, 95).
```

Incoming Age

N	Valid	25
	Missing	0
Percentiles	10	43.20
	20	47.00
	25	51.50
	30	53.60
	40	57.20
	50	60.00
	60	63.40
	70	67.00
	75	67.50
	80	68.80
	90	77.00

		Incoming Age	Incoming Weight in pounds
N	Valid	25	25
	Missing	0	0
Percentiles	5	31.50	127.75
	50	60.00	172.80
	95	83.30	251.14

```
*** ------------------------------------------------------------------------------ ***.
*** Use EXAMINE command to get STEM-LEAF plots
*** ------------------------------------------------------------------------------ ***.
```

```
* a simple histogram that includes a descriptive table by default.

EXAMINE AGE /PLOT STEMLEAF.

* You can include an option to not bother with the table.

EXAMINE AGE WEIGHIN
  /PLOT STEMLEAF
  /STATISTICS NONE.

* This produces a plot of age separately for each level of TRT.

EXAMINE AGE BY TRT
  /PLOT STEMLEAF
  /COMPARE GROUPS
  /STATISTICS NONE.
```

Incoming Age Stem-and-Leaf Plot

```
 Frequency    Stem &  Leaf

     1.00 Extremes    (=<27)
     2.00        4 .  24
     2.00        4 .  66
     3.00        5 .  124
     3.00        5 .  669
     4.00        6 .  0011
     6.00        6 .  577789
     1.00        7 .  3
     2.00        7 .  77
      .00        8 .
     1.00        8 .  6

Stem width:   10
Each leaf:       1 case(s)
```

```
* You can ask for multiple things at once.

EXAMINE AGE BY TRT
  /PLOT HISTOGRAM STEMLEAF
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /TOTAL.
```

Incoming Age Stem-and-Leaf Plot for
TRT= control

```
 Frequency    Stem &  Leaf

     2.00        4 .  66
     4.00        5 .  1269
     7.00        6 .  0115779
     1.00        7 .  7

Stem width:   10
Each leaf:       1 case(s)
```

Incoming Age Stem-and-Leaf Plot for
TRT= aloe treatment

```
 Frequency    Stem &  Leaf

     1.00        2 .  7
      .00        3 .
     2.00        4 .  24
     2.00        5 .  46
     3.00        6 .  078
     2.00        7 .  37
     1.00        8 .  6

Stem width:   10
Each leaf:       1 case(s)
```

```
*** ========================================================================== ***.
*** ========================================================================== ***.
*** CHAPTER 3: CENTER & SPREAD
*** ========================================================================== ***.
*** ========================================================================== ***.

*** DESCRIPTIVES /STATISTICS
*** defaults: DEFAULT (MIN, MAX, MEAN, STDDEV)
***   others: RANGE VARIANCE SKEWNESS KURTOSIS SUM

*** MEANS: MEAN & STDDEV, but BY groups

*** FREQUENCIES /STATISTICS
*** defaults: MEAN, STDDEV, MIN, MAX
***   others: SEMEAN VARIANCE, SKEWNESS, SESKEW, RANGE, MODE, KURTOSIS, SEKURT, MEDIAN, SUM

*** EXAMINE /STATISTICS
***  defaults: DESCRIPTIVES (Mean, 95% CI for mean, 5% trimmed mean, median, variance,
                             Stddev, min, max, skewness, kurtosis, range, IQR)
***    others: EXTREME, NONE, ALL


*** ------------------------------------------------------------------------- ***.
*** Use the DESCRIPTIVES command to get the SUMMARY STATISTICS (single variables).
*** ------------------------------------------------------------------------- ***.

* By default this command gives: N, min, max, mean, & STDDEV.

DESCRIPTIVES AGE.

* If you list multiple variables, it handles them one-at-a-time and makes a nice table.
* Does it make sense to calculate the mean of a categorical variable?

DESCRIPTIVES AGE TRT
   /STATISTICS DEFAULT.

* Using the option STATISTICS, you can ask for..
```

	N	Minimum	Maximum	Mean	Std. Deviation
Incoming Age	25	27	86	59.64	12.932
Treatment Group	25	0	1	.44	.507
Valid N (listwise)	25				

```
DESCRIPTIVES WEIGHIN
   /STATISTICS MIN MAX RANGE MEAN STDDEV VARIANCE SKEWNESS KURTOSIS SUM.

* ... or you can ALL.

DESCRIPTIVES
   AGE WEIGHIN
   TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6
   total_add total_sum
   total_avg total_mean
   /STATISTICS all.
```

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Incoming Age	25	59	27	86	1491	59.64	2.586	12.932	167.240	-.348	.464	.584	.902
Incoming Weight in pounds	25	137	124	261	4457	178.28	6.395	31.977	1022.498	.826	.464	.697	.902
TOTALCIN	25	8	4	12	163	6.52	.306	1.531	2.343	2.042	.464	6.347	.902
TOTALCW2	25	12	4	16	207	8.28	.508	2.542	6.460	1.148	.464	2.128	.902
TOTALCW4	25	11	6	17	259	10.36	.695	3.475	12.073	.552	.464	-.743	.902
TOTALCW6	23	16	3	19	218	9.48	.727	3.489	12.170	.882	.481	1.393	.935
TOTAL: ADD using the PLUS symbol	23	27	25	52	782	34.00	1.688	8.096	65.545	.525	.481	-.687	.935
TOTAL: ADD using the SUM command	25	27	25	52	847	33.88	1.559	7.796	60.777	.579	.464	-.484	.902
TOTAL: AVG using the PLUS & DIVIDE symbols	23	6.75	6.25	13.00	195.50	8.5000	.42203	2.02400	4.097	.525	.481	-.687	.935
TOTAL: AVG using the MEAN command	25	6.75	6.25	13.00	217.17	8.6867	.41136	2.05681	4.230	.341	.464	-.982	.902
Valid N (listwise)	23												

```
*** ------------------------------------------------------------------------------ ***.
*** Use the MEANS command to get the SUMMARY STATISTICS (a pair of variables).
*** ------------------------------------------------------------------------------ ***.

* This command calculates the SUMMARY STAT's for each level of a categorical variable.
* By default, this command gives: MEAN & STDDEV.

MEANS AGE BY TRT.
```

AGE Incoming Age

TRT Treatment Group	Mean	N	Std. Deviation
0 control	59.79	14	8.980
1 aloe treatment	59.45	11	17.218
Total	59.64	25	12.932

```
*** ------------------------------------------------------------------------------ ***.
*** Use the FREQUENCIES command to get the SUMMARY STATISTICS.
*** ------------------------------------------------------------------------------ ***.

* Easiest way - all default options.

FREQUENCIES AGE /STATISTICS all.


* Again, you can include the option to not
  bother with the table &
  do more than 1 variable.

FREQUENCIES AGE WEIGHIN
            total_add total_sum
   /FORMAT NOTABLE
   /STATISTICS all.


* Instead of requesting ALL the statistics,
   you can list the ones you want.

FREQUENCIES AGE WEIGHIN
   /FORMAT NOTABLE
   /STATISTICS MINIMUM MAXIMUM MEAN MEDIAN STDDEV RANGE SKEWNESS KURTOSIS.
```

		AGE Incoming Age	WEIGHIN Incoming Weight in pounds	total_add TOTAL: ADD using the PLUS symbol	total_sum TOTAL: ADD using the SUM command
N	Valid	25	25	23	25
	Missing	0	0	2	0
Mean		59.64	178.28	34.00	33.88
Std. Error of Mean		2.586	6.395	1.688	1.559
Median		60.00	172.80	34.00	34.00
Mode		67	164[a]	25	25
Std. Deviation		12.932	31.977	8.096	7.796
Variance		167.240	1022.498	65.545	60.777
Skewness		-.348	.826	.525	.579
Std. Error of Skewness		.464	.464	.481	.464
Kurtosis		.584	.697	-.687	-.484
Std. Error of Kurtosis		.902	.902	.935	.902
Range		59	137	27	27
Minimum		27	124	25	25
Maximum		86	261	52	52
Sum		1491	4457	782	847

```
* You can wrap it in the sort/split analyze by subgroups.
* FIRST:  you have to SORT by the variable you are going to split on.
* SECOND: make sure you use a 'temporary.' command so its not permanent.
* THIRD:  make sure you turn the split off at the end.
* FOURTH: its nice to go back to the original sorting.

SORT CASES by TRT.
SPLIT FILE by TRT.
FREQUENCIES AGE WEIGHIN
            \total_add total_sum
   /FORMAT NOTABLE
   /STATISTICS MINIMUM MAXIMUM
               MEAN MEDIAN STDDEV
   /NTILES(4).
SPLIT FILE off.
SORT CASES by ID.
```

TRT Treatment Group			AGE Incoming Age	WEIGHIN Incoming Weight in pounds	total_add TOTAL: ADD using the PLUS symbol	total_sum TOTAL: ADD using the SUM command
0 control	N	Valid	14	14	14	14
		Missing	0	0	0	0
	Mean		59.79	167.51	34.79	34.79
	Median		60.50	165.70	34.50	34.50
	Std. Deviation		8.980	23.005	8.684	8.684
	Minimum		46	124	25	25
	Maximum		77	213	52	52
	Percentiles	25	51.75	155.00	26.50	26.50
		50	60.50	165.70	34.50	34.50
		75	67.00	186.00	42.00	42.00
1 aloe treatment	N	Valid	11	11	9	11
		Missing	0	0	2	0
	Mean		59.45	191.99	32.78	32.73
	Median		60.00	181.50	33.00	33.00
	Std. Deviation		17.218	37.373	7.412	6.724
	Minimum		27	140	25	25
	Maximum		86	261	44	44
	Percentiles	25	44.00	164.00	25.50	26.00
		50	60.00	181.50	33.00	33.00
		75	73.00	226.00	39.50	37.00

```
*** ------------------------------------------------------------------------------ ***.
*** Use the EXAMINE command to get the SUMMARY STATISTICS (MORE details...including IQR !).
*** ------------------------------------------------------------------------------ ***.
```

```
* This gives a more extensive list of
   descriptives, but its all-or-none.

EXAMINE AGE
   /PLOT NONE
   /STATISTICS
     DESCRIPTIVES.
```

Descriptives

			Statistic	Std. Error
AGE Incoming Age	Mean		59.64	2.586
	95% Confidence Interval for Mean	Lower Bound	54.30	
		Upper Bound	64.98	
	5% Trimmed Mean		59.92	
	Median		60.00	
	Variance		167.240	
	Std. Deviation		12.932	
	Minimum		27	
	Maximum		86	
	Range		59	
	Interquartile Range		16	
	Skewness		-.348	.464
	Kurtosis		.584	.902

```
* The EXTREME option lists out the 5 highest
  and 5 lowest values, including the case #.

EXAMINE AGE
   /PLOT NONE
   /STATISTICS EXTREME.
```

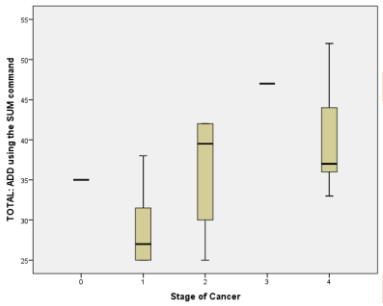
Extreme Values

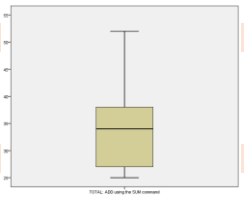
			Case Number	Value
AGE Incoming Age	Highest	1	18	86
		2	3	77
		3	16	77
		4	21	73
		5	9	69
	Lowest	1	12	27
		2	8	42
		3	10	44
		4	19	46
		5	2	46

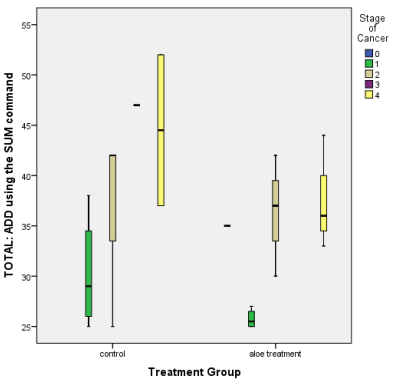
```
* This does almost the same thing as wrapping
  FREQUENCIES in SORT-SPLIT, ...
* ... except this one gives you more...
      like the Range & IQR "Inter Quartile Range!

EXAMINE AGE BY TRT
   /PLOT NONE
   /STATISTICS DESCRIPTIVES.


EXAMINE AGE BY TRT
   /PLOT NONE
   /STATISTICS all.
```

	TRT Treatment Group			Statistic	Std. Error
AGE Incoming Age	0 control	Mean		59.79	2.400
		95% Confidence Interval for Mean	Lower Bound	54.60	
			Upper Bound	64.97	
		5% Trimmed Mean		59.60	
		Median		60.50	
		Variance		80.643	
		Std. Deviation		8.980	
		Minimum		46	
		Maximum		77	
		Range		31	
		Interquartile Range		15	
		Skewness		.042	.597
		Kurtosis		-.392	1.154
	1 aloe treatment	Mean		59.45	5.192
		95% Confidence Interval for Mean	Lower Bound	47.89	
			Upper Bound	71.02	
		5% Trimmed Mean		59.78	
		Median		60.00	
		Variance		296.473	
		Std. Deviation		17.218	
		Minimum		27	
		Maximum		86	
		Range		59	
		Interquartile Range		29	
		Skewness		-.377	.661
		Kurtosis		-.216	1.279

```
*** ------------------------------------------------------
*** Create a BOXPLOT for a continuous variable.
*** ------------------------------------------------------------------------------ ***.

* Create a single boxplot for all subjects.

EXAMINE total_sum /PLOT BOXPLOT /STATISTICS NONE.

* ... separate by one grouping variable...

EXAMINE total_sum BY TRT
  /PLOT BOXPLOT
  /STATISTICS NONE
  /NOTOTAL.

* You can ask for the statistics with the boxplot.

EXAMINE total_sum BY STAGE
  /PLOT BOXPLOT
  /STATISTICS DESCRIPTIVES
  /NOTOTAL.

* ... or separate by two grouping variables.

EXAMINE total_sum BY TRT*STAGE
  /PLOT BOXPLOT
  /STATISTICS NONE
  /NOTOTAL.
```







```
*** ================================================================================ ***.
*** ================================================================================ ***.
*** CHAPTER 4:STANDARD & NORMAL
*** ================================================================================ ***.
*** ================================================================================ ***.
```

```
* Find the grand mean of age (59.64) and the standard deviation (12.932).

FREQUENCIES AGE
   /FORMAT NOTABLE
   /STATISTICS MEAN STDDEV.
```

AGE Incoming Age

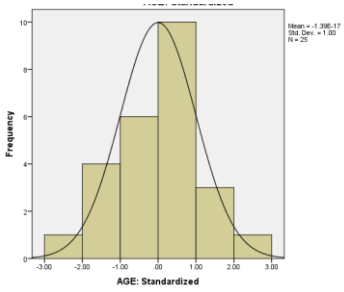
N	Valid	25
	Missing	0
Mean		59.64
Std. Deviation		12.932

```
*** -------------------------------------------------------------------------- ***.
*** LINEAR transformation.
*** -------------------------------------------------------------------------- ***.
```

```
* Calculate the z-scores
  by subtracting the M and dividing by the SD.
* Addition, Subtraction, Multiplication,
  and Division are linear transformations.

NUMERIC         AGE_z (F5.2).
VARIABLE LABELS AGE_z "AGE: Standardized".
COMPUTE         AGE_z = (AGE - 59.64) / 12.932.
EXECUTE.
```

		AGE Incoming Age	AGE_z AGE: Standardized
N	Valid	25	25
	Missing	0	0
Mean		59.64	.0000
Median		60.00	.0278
Std. Deviation		12.932	1.00001
Skewness		-.348	-.348
Kurtosis		.584	.584
Range		59	4.56
Minimum		27	-2.52
Maximum		86	2.04

```
* What happened to the skewness & kurtosis?

FREQUENCIES AGE AGE_z
   /FORMAT NOTABLE
   /STATISTICS MINIMUM MAXIMUM
               MEAN MEDIAN STDDEV
               RANGE SKEWNESS KURTOSIS
  /HISTOGRAM NORMAL.
```

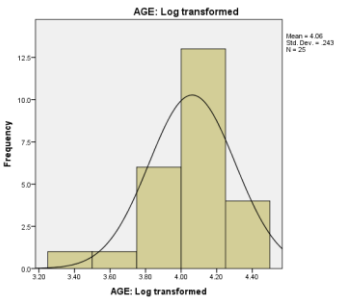


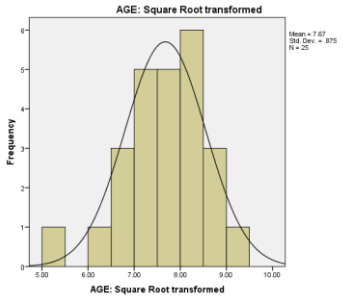

```
*** -------------------------------------------
*** NON-LINEAR transformations.
*** -------------------------------------------
```

```
* Try two non-linear transformations,
  like log and square root.

NUMERIC AGE_log AGE_sqrt (F5.2).

VARIABLE LABELS
AGE_log  "AGE: Log transformed"
AGE_sqrt "AGE: Square Root transformed".

COMPUTE AGE_log  = LN(AGE).
COMPUTE AGE_sqrt = SQRT(AGE).
EXECUTE.
```

		AGE Incoming Age	AGE_z AGE: Standardized	AGE_log AGE: Log transformed	AGE_sqrt AGE: Square Root transformed
N	Valid	25	25	25	25
	Missing	0	0	0	0
Mean		59.64	.0000	4.0624	7.6750
Median		60.00	.0278	4.0943	7.7460
Std. Deviation		12.932	1.00001	.24288	.87473
Skewness		-.348	-.348	-1.264	-.777
Kurtosis		.584	.584	2.901	1.421
Range		59	4.56	1.16	4.08
Minimum		27	-2.52	3.30	5.20
Maximum		86	2.04	4.45	9.27

```
* What happened to the skewness & kurtosis?

FREQUENCIES AGE AGE_z AGE_log AGE_sqrt
   /FORMAT NOTABLE
   /STATISTICS MINIMUM MAXIMUM
               MEAN MEDIAN STDDEV
               RANGE SKEWNESS KURTOSIS
  /HISTOGRAM NORMAL.
```





```
*** ------------------------------------------------------------------------------ ***.
*** Selecting a SUBSET of your dataset for a SINGLE STEP.
*** ------------------------------------------------------------------------------ ***.
```

```
* Find Mean/Median/Mode for age, but ONLY among
  participants with STAGE 1 (note the n).

* Note: you MUST 'run' ALL THREE lines at the same time,
       Or non-selected cases will be removed from your dataset.

TEMPORARY.
SELECT IF stage = 1.
FREQUENCIES age /STATISTICS MEAN MEDIAN MODE.
```

AGE Incoming Age

N	Valid	12
	Missing	0
Mean		61.67
Median		66.00
Mode		67

```
* Find Mean/Median/Mode for age, but ONLY among
  participants with STAGE 1 or stage 2 (note the n).

* Note: typing the 2 letters 'or' is the same as using the '|' symbol.

TEMPORARY.
SELECT IF stage = 1 or stage = 2.
FREQUENCIES age /STATISTICS MEAN MEDIAN MODE.

TEMPORARY.
SELECT IF stage = 1 | stage = 2.
FREQUENCIES age /STATISTICS MEAN MEDIAN MODE.
```

AGE Incoming Age

N	Valid	18
	Missing	0
Mean		59.89
Median		61.00
Mode		67

```
* Find Mean/Median/Mode of age, but ONLY among
  participants in the TREATMENT group w/STAGE 1 (note the n).

* Note: typing the 3 letters 'and' is the same as using the '&' symbol.

TEMPORARY.
SELECT IF trt = 1 and stage = 1.
FREQUENCIES age /STATISTICS MEAN MEDIAN MODE.

TEMPORARY.
SELECT IF trt = 1 & stage = 1.
FREQUENCIES age /STATISTICS MEAN MEDIAN MODE.
```

AGE Incoming Age

N	Valid	4
	Missing	0
Mean		55.50
Median		54.50
Mode		27[a]

a. Multiple modes exist. The smallest value is shown

```
* Get really fancy by combining SELECT IF and SPLIT FILE together.

* Find Mean/Median/Mode of age, for each treatment/control group,
  but ONLY among participants with at least stage 2.

* Note: Keep the TEMPORARY and SELECT IF lines together and
       'Run' the entire code chunk at the same time.

SORT CASES BY trt.
SPLIT FILE BY trt.
TEMPORARY.
SELECT IF stage >= 1.
FREQUENCIES age /STATISTICS MEAN MEDIAN MODE.
SPLIT FILE off.
SORT CASES BY id.
```

AGE Incoming Age

0 control	N	Valid	14
		Missing	0
	Mean		59.79
	Median		60.50
	Mode		46[a]
1 aloe treatment	N	Valid	10
		Missing	0
	Mean		58.10
	Median		58.00
	Mode		27[a]

a. Multiple modes exist. The smallest value is shown

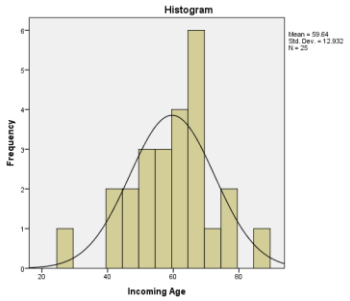
```
*** ================================================================================ ***.
*** ================================================================================ ***.
*** CHAPTER 5: TESTING ASSUMPTION OF NORMALITY
*** ================================================================================ ***.
*** ================================================================================ ***.
```

```
* You can get an idea shape from skewness & kurtosis, but a histogram is better.

* NOTE: the key word 'NORMAL' after the optional '/HISTOGRAM' will draw a bell curve on top.
```

```
FREQUENCIES AGE
   /FORMAT NOTABLE
   /HISTOGRAM NORMAL
   /STATISTICS MEAN MEDIAN
               SKEWNESS SESKEW
               KURTOSIS SEKURT.
```

AGE Incoming Age

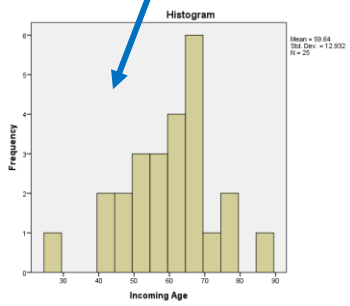
N	Valid	25
	Missing	0
Mean		59.64
Median		60.00
Skewness		-.348
Std. Error of Skewness		.464
Kurtosis		.584
Std. Error of Kurtosis		.902

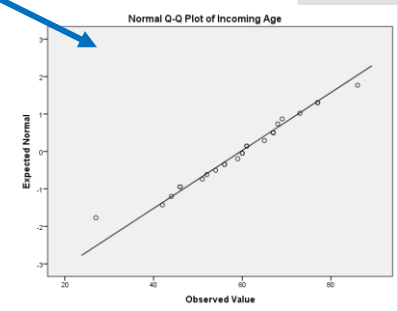



```
* The NPPLOT creates a QQplot that is even better.
```

```
* Beware, the formal tests for normality tend to be:
*   too sensative for large datasets and
*   too unreliable for small datasets.
```

```
EXAMINE AGE
  /PLOT HISTOGRAM NPPLOT.
```





Descriptives

			Statistic	Std. Error
AGE Incoming Age	Mean		59.64	2.586
	95% Confidence Interval for Mean	Lower Bound	54.30	
		Upper Bound	64.98	
	5% Trimmed Mean		59.92	
	Median		60.00	
	Variance		167.240	
	Std. Deviation		12.932	
	Minimum		27	
	Maximum		86	
	Range		59	
	Interquartile Range		16	
	Skewness		-.348	.464
	Kurtosis		.584	.902

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
AGE Incoming Age	.080	25	.200*	.983	25	.940

```
* The default CI for the mean is 95%,
  but you can use "/CINTERVAL 99" for 99% confidence interval
```

```
EXAMINE AGE
  /CINTERVAL 99.
```

Descriptives

			Statistic	Std. Error
AGE Incoming Age	Mean		59.64	2.586
	99% Confidence Interval for Mean	Lower Bound	52.41	
		Upper Bound	66.87	
	5% Trimmed Mean		59.92	
	Median		60.00	
	Variance		167.240	
	Std. Deviation		12.932	
	Minimum		27	
	Maximum		86	
	Range		59	
	Interquartile Range		16	
	Skewness		-.348	.464
	Kurtosis		.584	.902

```
*** ================================================================================== ***.
*** ================================================================================== ***.
*** CHAPTER 6: SINGLE SAMPLE t-TEST
*** ================================================================================== ***.
*** ================================================================================== ***.
```

```
* The t-test come with a confidence interval for THE DIFFERENCE IN MEANS (default is 95% CI).
```

```
T-TEST
  /TESTVAL 50
  /VARIABLES AGE.
```

One-Sample Test

	Test Value = 50					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
AGE Incoming Age	3.727	24	.001	9.640	4.30	14.98

```
* You can change the confidence level to with the CRITERIA option.
```

```
T-TEST
  /TESTVAL 50
  /VARIABLES AGE
  /CRITERIA CI(.99).
```

One-Sample Test

	Test Value = 50					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference Lower	Upper
AGE Incoming Age	3.727	24	.001	9.640	2.41	16.87

```
* To get a Confidence Interval around the mean, use a TEST VALUE of ZERO.
* Ignore the significance (p-value) as it is nearly always significant.
```

```
T-TEST
  /TESTVAL 0
  /VARIABLES AGE WEIGHIN
       total_add total_sum
  /CRITERIA CI(.90).
```

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference Lower	Upper
AGE Incoming Age	23.059	24	.000	59.640	55.21	64.07
WEIGHIN Incoming Weight in pounds	27.877	[illegible]	.000	178.280	167.34	189.22
total_add TOTAL: ADD using the PLUS symbol	20.141	22	.000	34.000	31.10	36.90
total_sum TOTAL: ADD using the SUM command	21.729	24	.000	33.880	31.21	36.55

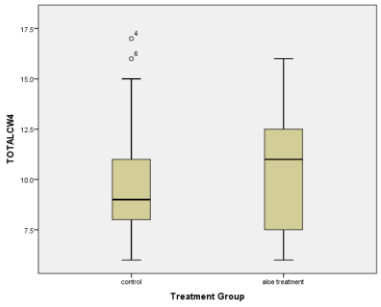
```
* ------------------------------------------------------------------------------------ *
* SPSS seems to have removed the bootstraping option.
* I think it is an add-on utility now which costs MEGA $$$.
* ------------------------------------------------------------------------------------ *
```

```
*** ================================================================================ ***.
*** ================================================================================ ***.
*** CHAPTER 7: 2 INDEPENTENT SAMPLE MEANS t-TEST
*** ================================================================================ ***.
*** ================================================================================ ***.
```

```
* GOAL: Is there a difference between the treatment & placebo?
```

```
*** -------------------------------------------------------------------------------- ***.
*** Check for normality - 4 wk in each group.
*** -------------------------------------------------------------------------------- ***.
```

```
EXAMINE TOTALCW4 BY TRT
  /PLOT HISTOGRAM BOXPLOT NPPLOT
  /STATISTICS DESCRIPTIVES
  /NOTOTAL.
```

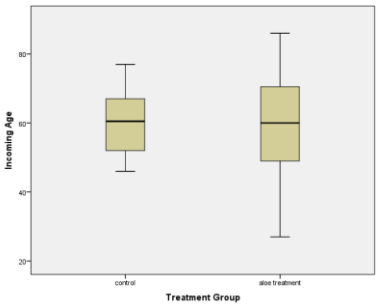


		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	TRT Treatment Group	Statistic	df	Sig.	Statistic	df	Sig.
TOTALCW4	0 control	.196	14	.149	.883	14	.064
	1 aloe treatment	.140	11	.200*	.929	11	.401

```
*** -------------------------------------------------------------------------------- ***.
*** H0: age in placebo group     =     age in aleo treatment group.
*** H1: age in placebo group   not =   age in aleo treatment group.
*** -------------------------------------------------------------------------------- ***.
```

```
* Look at side-by-side boxplots to get a feel for the data (= spreads? and = centers?).
* Use the Histograms and QQplots to judge normaity more than skewness & kurtosis
```

```
EXAMINE AGE BY TRT
  /PLOT HISTOGRAM BOXPLOT NPPLOT
  /STATISTICS DESCRIPTIVES
  /NOTOTAL.
```



Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	TRT Treatment Group	Statistic	df	Sig.	Statistic	df	Sig.
AGE Incoming Age	0 control	.108	14	.200*	.967	14	.828
	1 aloe treatment	.124	11	.200*	.983	11	.979

```
* Step 1) Leven's Test for Equality of Variance
*          If the SIG (p-value) is < .05 --> variances are different --> look at bottom row
*          If the SIG (p-value) is > .05 --> variances are nearly =  --> look at top row

* Step 2) t-Test for Equality of Means (default is 95% CI or alpha = .05)
```

```
T-TEST GROUPS TRT (0 1)
  /VARIABLES AGE.
```

Group Statistics

	TRT Treatment Group	N	Mean	Std. Deviation	Std. Error Mean
AGE Incoming Age	0 control	14	59.79	8.980	2.400
	1 aloe treatment	11	59.45	17.218	5.192

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AGE Incoming Age	Equal variances assumed	4.566	.043	.062	23	.951	.331	5.322	-10.678	11.341
	Equal variances not assumed			.058	14.231	.955	.331	5.719	-11.917	12.580

```
*** ------------------------------------------------------------------------------ ***.
*** H0: Total in placebo group     =     Total in aleo treatment group.
*** H1: Total in placebo group   not =   age in aleo treatment group.
*** ------------------------------------------------------------------------------ ***.
```

```
* You can compare several variables at once.
```

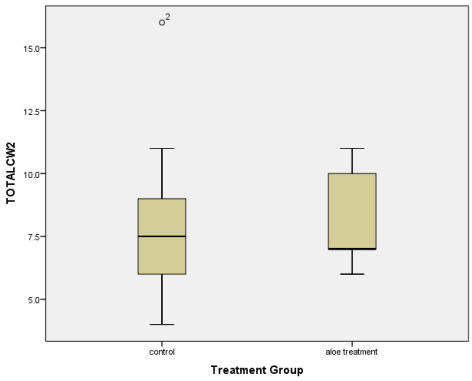
```
EXAMINE TOTALCW2 TOTALCW4 TOTALCW6 BY TRT
  /STATISTICS DESCRIPTIVES
  /NOTOTAL.
```

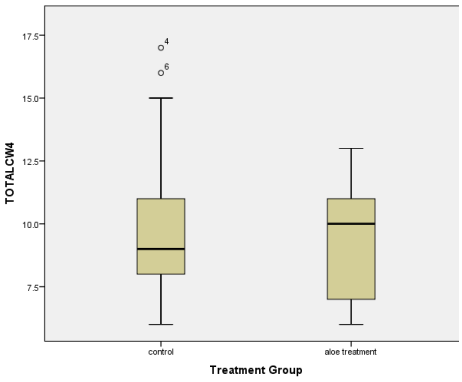
Tests of Normality

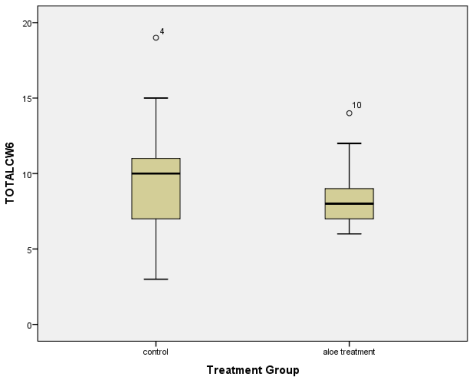
	TRT Treatment Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
TOTALCW2	0 control	.174	14	.200*	.884	14	.065
	1 aloe treatment	.293	9	.025	.822	9	.037
TOTALCW4	0 control	.196	14	.149	.883	14	.064
	1 aloe treatment	.177	9	.200*	.926	9	.446
TOTALCW6	0 control	.179	14	.200*	.960	14	.728
	1 aloe treatment	.244	9	.130	.861	9	.099

```
* The boxplots are my favorite.
```

```
EXAMINE TOTALCW2 TOTALCW4 TOTALCW6 BY TRT
  /PLOT  BOXPLOT
  /STATISTICS none
  /NOTOTAL.
```








```
* You can do several t-tests at once, too.
```

```
T-TEST GROUPS TRT (0 1)
  /VARIABLES TOTALCW2 TOTALCW4 TOTALCW6.
```

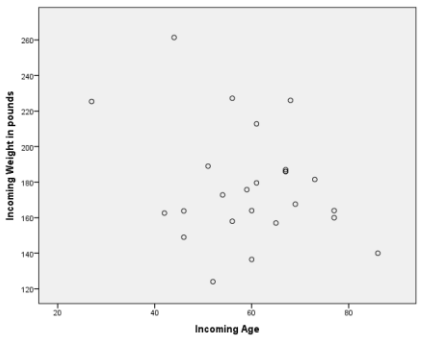
Group Statistics

	TRT Treatment Group	N	Mean	Std. Deviation	Std. Error Mean
TOTALCW2	0 control	14	8.14	3.009	.804
	1 aloe treatment	11	8.45	1.916	.578
TOTALCW4	0 control	14	10.14	3.592	.960
	1 aloe treatment	11	10.64	3.472	1.047
TOTALCW6	0 control	14	9.93	3.970	1.061
	1 aloe treatment	9	8.78	2.635	.878

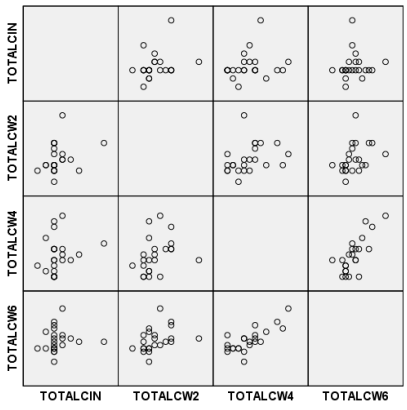
Independent Samples Test

		Levene's Test for Equality of Variances F	Sig.	t-test for Equality of Means t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
TOTALCW2	Equal variances assumed	.657	.426	-.299	23	.768	-.312	1.044	-2.472	1.848
	Equal variances not assumed			-.315	22.197	.756	-.312	.990	-2.364	1.741
TOTALCW4	Equal variances assumed	.000	.995	-.346	23	.733	-.494	1.426	-3.444	2.457
	Equal variances not assumed			-.347	21.950	.732	-.494	1.420	-3.440	2.452
TOTALCW6	Equal variances assumed	.668	.423	.765	21	.453	1.151	1.505	-1.978	4.280
	Equal variances not assumed			.835	20.941	.413	1.151	1.378	-1.714	4.016

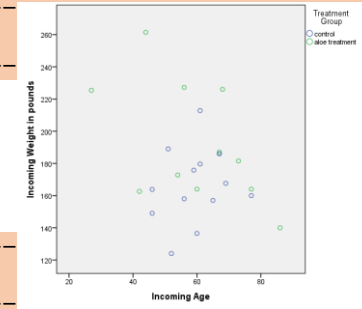
```
*** ============================================================
*** ============================================================
*** CHAPTER 9: 2 CORRELATION
*** ============================================================
*** ============================================================

*** ============================================================
*** Creating SCATTERPLOTS: x-variable "WITH" y-variable.
*** ============================================================

GRAPH /SCATTERPLOT(BIVAR) AGE WITH WEIGHIN.
```



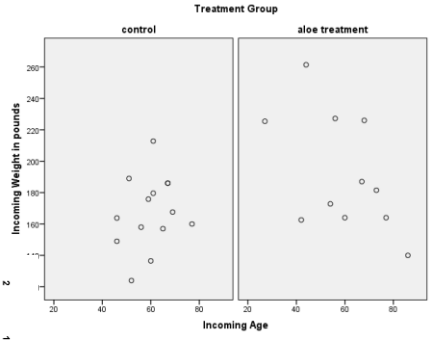

```
GRAPH /SCATTERPLOT(BIVAR) TOTALCIN WITH TOTALCW2.
GRAPH /SCATTERPLOT(BIVAR) TOTALCW2 WITH TOTALCW4.
GRAPH /SCATTERPLOT(BIVAR) TOTALCW4 WITH TOTALCW6.
```



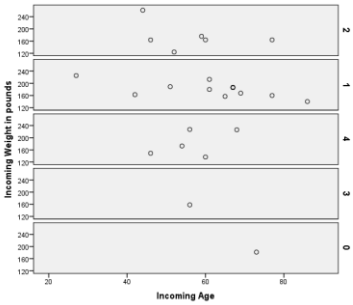

```
*** ------------------------------------------------------------
*** Scatterplot MATRIX between more than 2 variables.
*** ------------------------------------------------------------

GRAPH /SCATTERPLOT(MATRIX)
            TOTALCIN TOTALCW2
            TOTALCW4 TOTALCW6.

*** ------------------------------------------------------------
*** Color "BY" a categorical variable.
*** ------------------------------------------------------------

GRAPH /SCATTERPLOT(BIVAR) AGE WITH WEIGHIN BY TRT.
GRAPH /SCATTERPLOT(BIVAR) AGE WITH WEIGHIN BY STAGE.
```



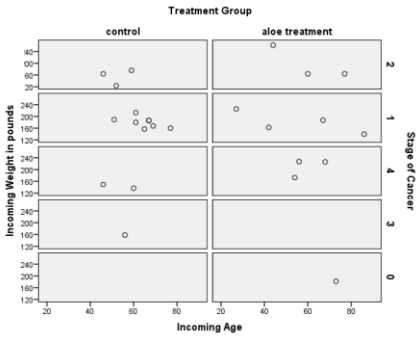

```
*** ------------------------------------------------------------
* use "SELECT IF" to restrict to only cases in stage 0, 1 or 2.
*** ------------------------------------------------------------

TEMPORARY.
SELECT IF STAGE <= 2.
GRAPH /SCATTERPLOT(BIVAR) AGE WITH WEIGHIN.
```




```
*** ------------------------------------------------------------
*** separate treatment/control into panels (columns).
*** ------------------------------------------------------------

GRAPH
  /SCATTERPLOT(BIVAR) AGE WITH WEIGHIN
  /PANEL COLVAR=TRT.
```




```
* separate stage into panels (rows).

GRAPH
  /SCATTERPLOT(BIVAR) AGE WITH WEIGHIN
  /PANEL ROWVAR=STAGE.

* separate treatment/control into panels (columns)
   AND stage into panels (rows).

GRAPH
  /SCATTERPLOT(BIVAR) AGE WITH WEIGHIN
  /PANEL COLVAR=TRT ROWVAR=STAGE.
```



```
*** ========================================
*** Pearson's Correlation Coefficient, "r".
*** ========================================

*** Only 2 variables only.

CORRELATIONS AGE WEIGHIN.
```

Correlations

		AGE Incoming Age	WEIGHIN Incoming Weight in pounds
AGE Incoming Age	Pearson Correlation	1	-.288
	Sig. (2-tailed)		.163
	N	25	25
WEIGHIN Incoming Weight in pounds	Pearson Correlation	-.288	1
	Sig. (2-tailed)	.163	
	N	25	25

```
*** ----------------------------------------
*** More than two variables = "correlation matrix".
*** ----------------------------------------

* Default: pair-wise deletion
  (cells can all have different # valid).

CORRELATIONS
TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6.
```

		TOTALCIN	TOTALCW2	TOTALCW4	TOTALCW6
TOTALCIN	Pearson Correlation	1	.314	.222	.098
	Sig. (2-tailed)		.126	.287	.657
	N	25	25	25	23
TOTALCW2	Pearson Correlation	.314	1	.337	.378
	Sig. (2-tailed)	.126		.099	.075
	N	25	25	25	23
TOTALCW4	Pearson Correlation	.222	.337	1	.763
	Sig. (2-tailed)	.287	.099		.000
	N	25	25	25	23
TOTALCW6	Pearson Correlation	.098	.378	.763	1
	Sig. (2-tailed)	.657	.075	.000	
	N	23	23	23	23

```
* Option: list-wise deletion
  (all cells have the same N, given under table).

CORRELATIONS
TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6
  /MISSING LISTWISE.
```

		TOTALCIN	TOTALCW2	TOTALCW4	TOTALCW6
TOTALCIN	Pearson Correlation	1	.282	.206	.098
	Sig. (2-tailed)		.192	.346	.657
TOTALCW2	Pearson Correlation	.282	1	.314	.378
	Sig. (2-tailed)	.192		.145	.075
TOTALCW4	Pearson Correlation	.206	.314	1	.763
	Sig. (2-tailed)	.346	.145		.000
TOTALCW6	Pearson Correlation	.098	.378	.763	1
	Sig. (2-tailed)	.657	.075	.000	

a. Listwise N=23

```
*** ----------------------------------------
*** Use the "with" option to create a smaller matrix.
*** ----------------------------------------

CORRELATIONS
      TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6
      WITH AGE WEIGHIN.
```

		AGE Incoming Age	WEIGHIN Incoming Weight in pounds
TOTALCIN	Pearson Correlation	.256	.170
	Sig. (2-tailed)	.217	.418
	N	25	25
TOTALCW2	Pearson Correlation	-.106	.274
	Sig. (2-tailed)	.615	.185
	N	25	25
TOTALCW4	Pearson Correlation	.162	-.095
	Sig. (2-tailed)	.438	.651
	N	25	25
TOTALCW6	Pearson Correlation	.030	-.078
	Sig. (2-tailed)	.891	.725
	N	23	23

```
*** ----------------------------------------
*** use "SELECT IF" to restrict to only cases in stage 3
*** ----------------------------------------

TEMPORARY.
SELECT IF STAGE = 3 or STAGE = 4.
CORRELATIONS TOTALCIN TOTALCW2 TOTALCW4 TOTALCW6.
```

```
* use "SELECT IF" to restrict to only cases in stage 0, 1 or 2.

TEMPORARY.
SELECT IF STAGE <= 2.
CORRELATIONS AGE WEIGHIN.
```

```
*** ----------------------------------
*** “SPLIT FILE" calculate on subgroups.
*** ----------------------------------

SORT CASES by TRT.
SPLIT FILE by TRT.
CORRELATIONS AGE WEIGHIN.
SPLIT FILE off.
SORT CASES by ID.
```

TRT Treatment Group			AGE Incoming Age	WEIGHIN Incoming Weight in pounds
0 control	AGE Incoming Age	Pearson Correlation	1	.235
		Sig. (2-tailed)		.418
		N	14	14
	WEIGHIN Incoming Weight in pounds	Pearson Correlation	.235	1
		Sig. (2-tailed)	.418	
		N	14	14
1 aloe treatment	AGE Incoming Age	Pearson Correlation	1	-.534
		Sig. (2-tailed)		.091
		N	11	11
	WEIGHIN Incoming Weight in pounds	Pearson Correlation	-.534	1
		Sig. (2-tailed)	.091	
		N	11	11

```
*** ================================================================================ ***.
*** ================================================================================ ***.
*** CHAPTER 10: SIMPLE LINEAR REGRESSION
*** ================================================================================ ***.
*** ================================================================================ ***.
```

```
* simplest syntax: must tell is Y and what is X
   (leave everything we can to default settings).

REGRESSION
  /DEPENDENT WEIGHIN
  /METHOD=ENTER AGE.
```

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	220.690	30.108		7.330	.000
	AGE Incoming Age	-.711	.494	-.288	-1.440	.163

a. Dependent Variable: WEIGHIN Incoming Weight in pounds

```
* include asking for the 95% CI for the slope.

REGRESSION
  /STATISTICS COEFF CI(95) R
  /DEPENDENT WEIGHIN
  /METHOD=ENTER AGE.
```

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	220.690	30.108		7.330	.000	158.408	282.972
	AGE Incoming Age	-.711	.494	-.288	-1.440	.163	-1.733	.310

```
* you can ask to include descriptives

REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /DEPENDENT WEIGHIN
  /METHOD=ENTER AGE.
```

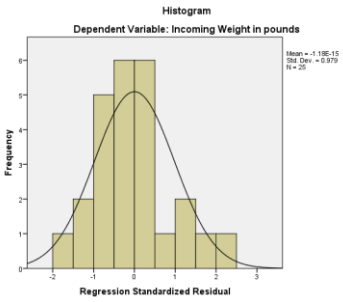
Descriptive Statistics

	Mean	Std. Deviation	N
WEIGHIN Incoming Weight in pounds	178.28	31.977	25
AGE Incoming Age	59.64	12.932	25

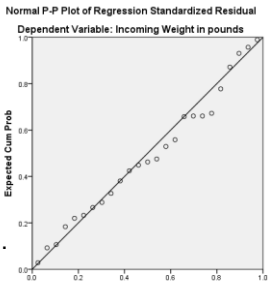
Correlations

		WEIGHIN Incoming Weight in pounds	AGE Incoming Age
Pearson Correlation	WEIGHIN Incoming Weight in pounds	1.000	-.288
	AGE Incoming Age	-.288	1.000
Sig. (1-tailed)	WEIGHIN Incoming Weight in pounds	.	.082
	AGE Incoming Age	.082	.
N	WEIGHIN Incoming Weight in pounds	25	25
	AGE Incoming Age	25	25

```
* you can ask for residual plots to test the assumptions, too.

REGRESSION
  /DEPENDENT WEIGHIN
  /METHOD=ENTER AGE
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```




```
* most of the optional commands used here (from point-click).

REGRESSION
  /DESCRIPTIVES MEAN STDDEV CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /DEPENDENT WEIGHIN
  /METHOD=ENTER AGE
  /RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```




```
* you can SAVE variables, too.
*   ...optional: specify the name of the new variable in parenthese.

* PRED.     Unstandardized predicted values.
* RESID.    Unstandardized residuals.
* MCIN.     Lower (LMCIN) and upper (UMCIN) bounds for the confidence interval of the MEAN Y
* ICIN.     Lower and upper bounds for the prediction interval for a SINGLE observation

*  The default confidence interval is 95%, can be reset with the CIN subcommand of CRITERIA.

REGRESSION
  /DEPENDENT WEIGHIN
  /METHOD=ENTER AGE
  /SAVE PRED RESID MCIN ICIN.

REGRESSION
  /CRITERIA=CIN(90)
  /DEPENDENT WEIGHIN
  /METHOD=ENTER AGE
  /SAVE PRED(weight_pred) RESID(weightin_resid) MCIN(weight_90).
```

PRE_1	RES_1	LMCI_1	UMCI_1	LICI_1	UICI_1
183.71279	-59.71279	168.59868	198.82691	117.25473	250.17086
165.93532	-5.93532	143.98083	187.88982	97.59619	234.27446
178.02400	-41.52400	165.07546	190.97255	112.02474	244.02327
177.31291	2.28709	164.29524	190.33057	111.30005	243.32577
178.73510	-2.93510	165.77528	191.69492	112.73363	244.73658
171.62412	-4.02412	155.53227	187.71596	104.93689	238.31134
173.04631	12.95369	158.07788	188.01474	106.62123	239.47140
180.86840	-22.86840	167.40159	194.33521	114.76550	246.97130
177.31291	35.48709	164.29524	190.33057	111.30005	243.32577

```
*** ================================================================================ ***.
*** ================================================================================ ***.
*** CHAPTER 11: PAIRED T-TEST FOR MATCHED PAIRS DATA (vs Independent groups)
*** ================================================================================ ***.
*** ================================================================================ ***.
```

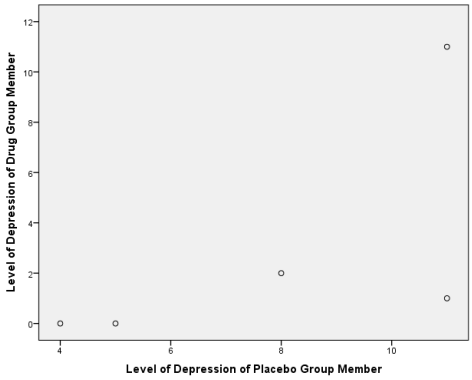
**** NEW DATA SET (not the same cancer dataset as the pages before this) ****

* Enter data as **matched pairs**.

```
DATA LIST LIST
    /pair drug placebo.
BEGIN DATA
    1   11   11
    2    1   11
    3    0    5
    4    2    8
    5    0    4
END DATA.
```

pair	drug	placebo
1	11	11
2	1	11
3	0	5
4	2	8
5	0	4

```
FORMATS pair drug placebo (F2).

VARIABLE LABELS
pair    "Identification of Matched Pair"
drug    "Level of Depress: Drug Group"
placebo "Level of Depress: Placebo Group".

GRAPH /SCATTERPLOT(BIVAR) placebo WITH drug.
```




```
T-TEST PAIRS placebo WITH drug (PAIRED).
```

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	placebo Level of Depression of Placebo Group Member - drug Level of Depression of Drug Group Member	5.000	3.606	1.612	.523	9.477	3.101	4	.036

```
COMPUTE diff = placebo - drug.
EXECUTE.

T-TEST TESTVAL 0
  /VARIABLE diff.
```

One-Sample Test

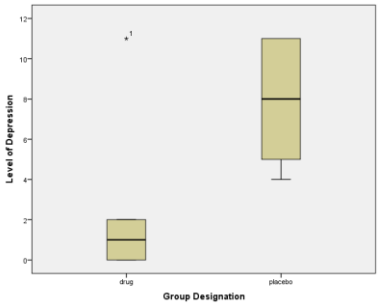
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
diff	3.101	4	.036	5.000	.52	9.48

* Enter **SAME** data as **independent groups**.

```
DATA LIST LIST
    /ID pair grp dep.
BEGIN DATA
   1   1   1   11
   2   1   2   11
   3   2   1    1
   4   2   2   11
   5   3   1    0
   6   3   2    5
   7   4   1    2
   8   4   2    8
   9   5   1    0
  10   5   2    4
END DATA.
```

ID	pair	grp	dep
1	1	drug	11
2	1	placebo	11
3	2	drug	1
4	2	placebo	11
5	3	drug	0
6	3	placebo	5
7	4	drug	2
8	4	placebo	8
9	5	drug	0
10	5	placebo	4

```
FORMATS ID pair grp dep (F2).

VARIABLE LABELS
ID    "Identification of Individual"
pair  "Identification of Matched Pair"
grp   "Group Designation"
dep   "Level of Depression".

VALUE LABELS grp 1 "drug" 2 "placebo".

EXAMINE dep BY grp
  /PLOT BOXPLOT
  /STATISTICS NONE
  /NOTOTAL.
```



* This ignores correlation!

```
T-TEST GROUPS=grp(1 2)
  /VARIABLES=dep.
```

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
dep Level of Depression	Equal variances assumed	.200	.667	-1.964	8	.085	-5.000	2.546	-10.870	.870
	Equal variances not assumed			-1.964	7.173	.089	-5.000	2.546	-10.990	.990