SPSS: Rank Tests

Male	Height (in)	*69*	*70*	*65*	*72*	*76*	*70*	*70*	*66*	*68*	*73*		
	Weight (lb)	192	148	140	190	248	197	170	137	160	185		

Female	Height (in)	*65*	*61*	*67*	*65*	*70*	*62*	*63*	*60*	*66*	*66*	*65*	*64*
	Weight (lb)	110	105	136	135	187	125	147	118	128	175	147	120

Here is a data set. When entering it, make it look like this:

	height	weight	gender
1	69	192	1
2	70	148	1
3	65	140	1
4	72	190	1
5	78	248	1
6	70	197	1
7	70	170	1
8	66	137	1
9	68	160	1
10	75	185	1
11	66	110	2
12	61	105	2

One-sample non-parametric tests

In general, the one-sample non-parametric tests work just like the t-tests. However, there are a couple of small differences, in particular, the fact that the hypothesis number has to be entered as data.

So, let's say we want to repeat the one-sample test.
This two-sided hypothesis was:

$$H_o : \mu_F = 63 \text{ inches} \quad \& \quad H_A : \mu_F \neq 63 \text{ inches}$$

Since the non-parametric tests look at median, we need to rewrite these as:

$$H_o : \theta_F = 63 \text{ inches} \quad \& \quad H_A : \theta_F \neq 63 \text{ inches}$$

Also, this test is only about the women, so we want to filter the cases to just be women.
Data: Select Cases: If is the way to do that, with gender = 2. You know you did it right if the diagonal lines are over the case number.

The weird thing

With t-tests, you could enter the H_0 value you wanted to in a special pop-up box. SPSS doesn't let you do that on one-sample Non-parametric tests. Instead, we "cheat" to go around this problem. We make a dummy variable where we put the null hypothesis. Then, we treat the data like it's a related sample (matched pairs). If you think about it, a matched-pairs test works much like a one-sample test on the differences between the data.

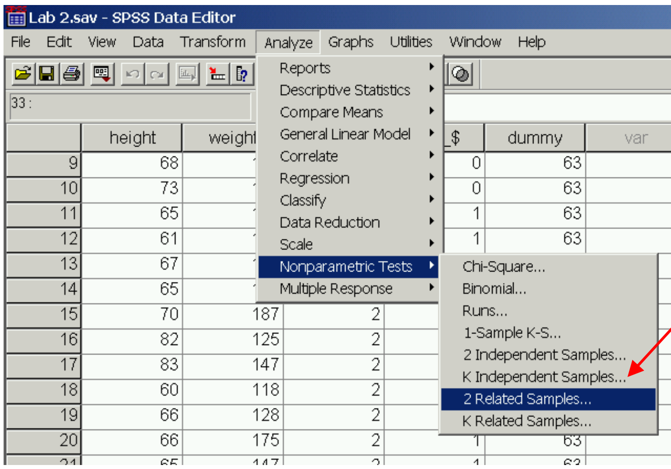
	height	weight	gender	filter_$	dummy
9	68	160	1	0	63
10	73	185	1	0	63
11	65	110	2	1	63
12	61	105	2	1	63
13	67	136	2	1	63
14	65	135	2	1	63
15	70	187	2	1	63
16	82	125	2	1	63
17	83	147	2	1	63
18	60	118	2	1	63
19	66	128	2	1	63
20	66	175	2	1	63
21	65	147	2	1	63
22	64	120	2	1	63
23					

Yep. It's really just a big column of 63s.

SPSS: Rank Tests

Simple Non-parametric Tests

Notice that there is a whole menu within Analyze for Non-parametric Statistics. Most everything we do in this part of the semester can be found here. Both of these tests can be found in **Analyze: Non-parametric: 2-Related Samples**.

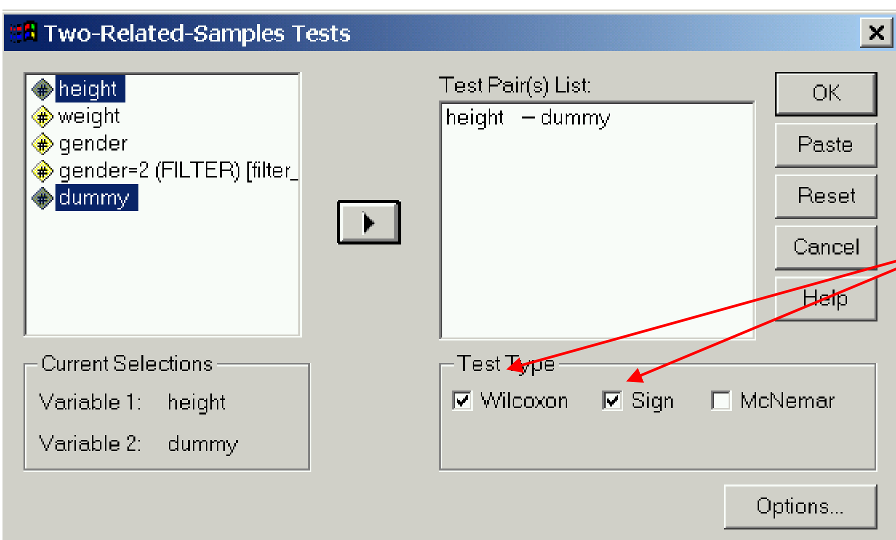


Sign Test

This test compares the number of points above and below the specified median. When the null hypothesis is true, roughly the same number of points lie above and below this hypothesized median.

Wilcoxon Signed Rank Test

This test compares the ranks of the data above and below the median. When the null hypothesis is true, roughly the same total rank of data points lie above and below this hypothesized median.



Once you get this far, running one or both test is easy. Just put your variables in the Test Pair list, and click the box(es) you want. In fact, running both tests is no harder than running just one.

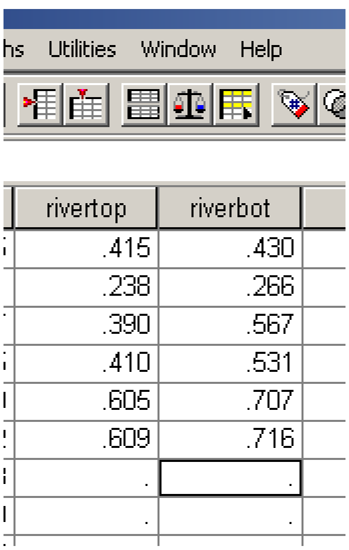
SPSS: Rank Tests

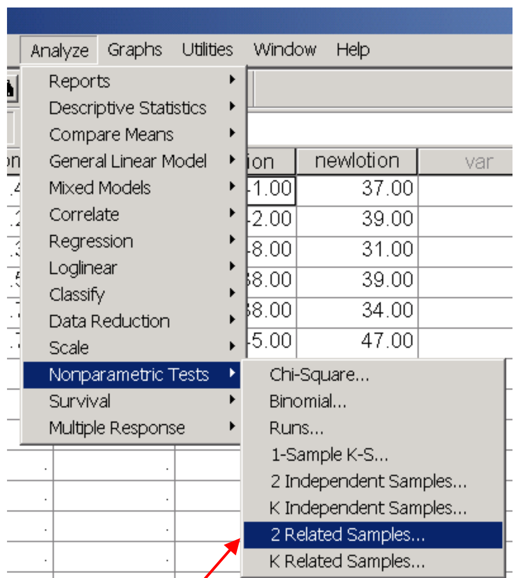
The Wilcoxon Signed-Rank Test for matched, ordinal data

Although a matched-pairs t-test is great, sometimes we aren't sure that our data has met all of the assumptions for t-tests. In that case, we should instead use the Wilcoxon or the Sign Test.
It works just like it did for one-sample, and is even easier, since we don't have to make that "dummy" variable.

Here's is data taken on sediment at the top and bottom of a river. It looks pretty numeric, but we don't really know how it is laid out. Notice that it is matched, because the top and bottom are from the same points in the river. As a hypothesis, we write it:

$H_o : \theta_{top} = \theta_{bottom}$ $\qquad$ $H_A : \theta_{top} \neq \theta_{bottom}$ (two-sided)

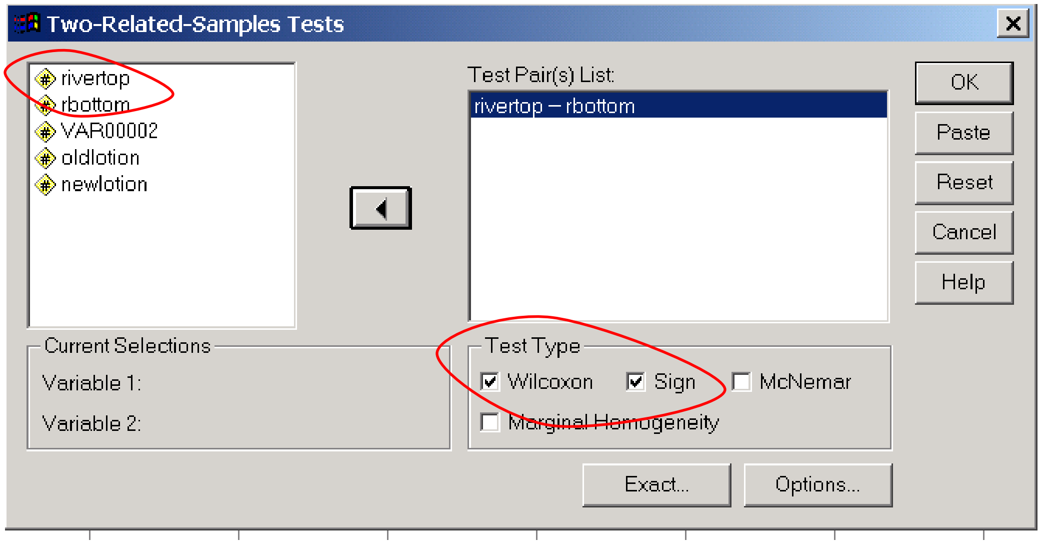




So, we need to click on
Analyze: Non-Parametric: 2-related Samples

This will bring up the data window. Click on the two variables representing our paired data and then use the arrow key to select them. Be sure to click both variables before clicking the arrow.

SPSS: Rank Tests



Since it's free, go ahead and click both boxes (Wilcoxon and Sign Tests).

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
rbottom - rivertop	Negative Ranks	1[a]	3.00	3.00
	Positive Ranks	5[b]	3.60	18.00
	Ties	0[c]		
	Total	6		

a. rbottom < rivertop
b. rbottom > rivertop
c. rbottom = rivertop

Test Statistics[b]

	rbottom - rivertop
Z	-1.572[a]
Asymp. Sig. (2-tailed)	.116

a. Based on negative ranks.
b. Wilcoxon Signed Ranks Test

Sign Test

Frequencies

		N
rbottom - rivertop	Negative Differences[a]	1
	Positive Differences[b]	5
	Ties[c]	0
	Total	6

a. rbottom < rivertop
b. rbottom > rivertop
c. rbottom = rivertop

Test Statistics[b]

	rbottom - rivertop
Exact Sig. (2-tailed)	.219[a]

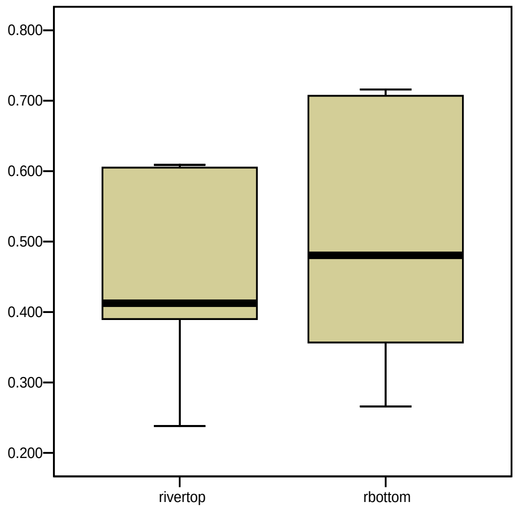
a. Binomial distribution used.
b. Sign Test

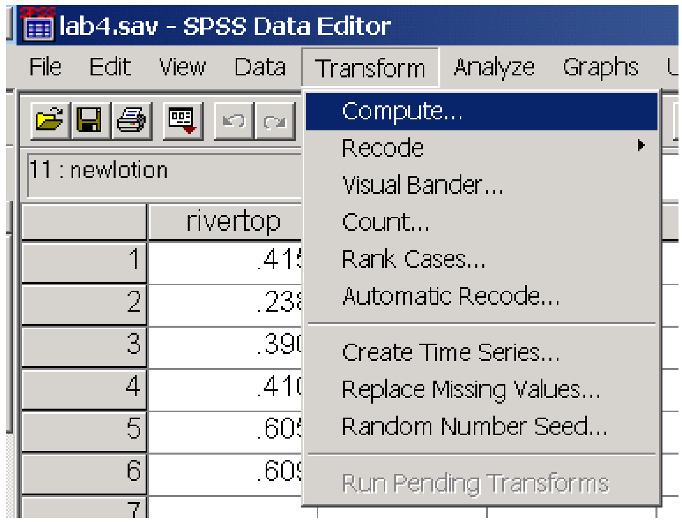
We can see that the p value for the test is no longer below our cutoff of 0.05 with either test and so we would fail to reject the null hypothesis and conclude there is significant evidence that the concentration of pollutants differs from the top and bottom of the river.

So, if the assumptions are not met, then we have saved ourselves a heap of trouble. On the other hand, these tests are "weaker" than the t-test, so, if the data is normal "enough," we've missed a chance to find a difference.

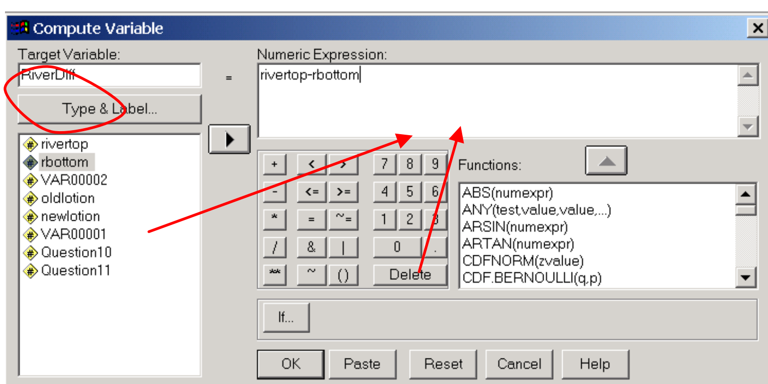
SPSS: Rank Tests

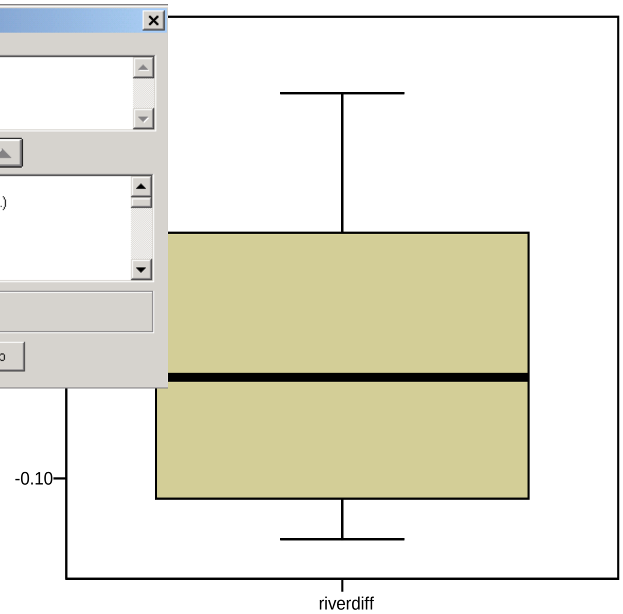
Difference Plots: Making a plot of the two data sets using **Graphs:Boxplot** (remember to click "make separate plots for separate variables") works OK, but doesn't really show the matched nature of the data (See how boring it is? Of course, making a plot of the 5 number summary of 6 numbers will never be that interesting). So, we'd rather make a plot of the difference. To find this, we can just subtract one from the other. You can do this by hand, but it seems the sort of thing a computer can do for you. The command you want is **Transform: Compute**





Now, you get the popup box below, and you can make a new column that is a function of the old columns. I named the column "RiverDiff," and made it by taking "rivertop" and subtracting "riverbottom." Click OK, and you'll see a new column in your data window. You can either type a minus sign or click on the little one in the box. You can see a zillion other functions, but this is the only one we'll probably use.





Now, make a boxplot of the new variable.

This one is much more interesting. You can see where zero is, and how much sits above and below zero.