

My journey to answering Bob's question on 01/27/2006:

First, I needed to save the data in a format that Stata could read in. This meant I needed to re-arrange the data, change variable names, remove comments, and save it in a text file. What I did for Bob's data was to create a column called `toxin` with 0 for controls and 1 for toxins, to put the two saline columns into a single column, and to put the two morphine columns into a single column. I had to make sure the two numbers for a subject are still on the same row. When Stata read in the data, there were three variables (`toxin`, `saline`, `morphine`) and 14 rows, with a missing value for `morphine`.

Then I graphed the data using

```
scatter morphine saline if (toxin==0) ||  
    scatter morphine saline if (toxin==1), mcol(red)  
    legend(label(1 "Control") label(2 "Toxin"))
```

Another way of graphing the data is

```
scatter morphine saline, by(toxin)
```

Apparently morphine brings down the response time for controls no matter what the baseline response time is. While for the toxin group, morphine appears to have effect if the baseline response time is long but mixed results if the baseline time is relatively low.

The baseline response times for the two groups are comparable. This can be seen by using

```
dotplot saline, over(toxin)
```

Now to capture the effect of morphine, we can use two scores: difference or ratio. They have different meanings and thus different interpretations. It is always good to determine which one to use prior to data analysis. If this is not feasible, we can try both. Then we are more on the exploratory track, which allows ourselves to see which score may capture the effect better. However, it is wrong if you try both scores and pick the better one to report without disclosing the fact that you have tried both and picked the "winner".

I tried both. I generated new variables `diff` and `ratio` using:

```
gen diff = morphine - saline  
gen ratio = morphine/saline
```

and did the following (repeat this for `ratio`):

```
sort toxin  
by toxin: summarize diff  
graph box diff, over(toxin)  
dotplot diff, over(toxin)  
ttest diff, by(toxin)  
ranksum diff, by(toxin)
```

Note that Stata is inconsistent in using the grouping variable like `toxin` in different commands. When you use the `by var:` prefix, you need to sort the data first (once for all), although in my case the data have already been ordered according to `toxin`. I did summary, boxplots, and dotplots for the two groups. I also did a t-test and a non-parametric test called rank-sum test (also called Wilcoxon's test and Mann-Whitney test). The validity of the t-test relies on the assumption that the distributions of `diff` (or `ratio`) for the two groups should be like normal distributions, which is difficult to check given the small sample sizes. Both tests yielded p-values that are not significant at level 0.05, despite the difference we can see. This is because the sample sizes are small.

More sophisticated analyses can be carried out on this data set. We will try them out later on.