Some Examples Demonstrating the Need for Using Two-Group Tests

Kevin E. Thorpe
Department of Public Health Sciences
University of Toronto

November 1, 2005

Introduction

A question that occasionally arises in introductory statistics courses concerns the necessity for performing two-group tests. Often, people wonder why it is not acceptable to compute 95% confidence intervals on the outcome separately for the two groups and use these as the basis for determining whether or not the two groups are statistically significantly different. For example, one might be tempted to conclude that two groups are "different" if one confidence interval includes "zero" and the other does not. Alternatively, one might want to conclude that if there is overlap between the confidence intervals, the two groups are not "different."

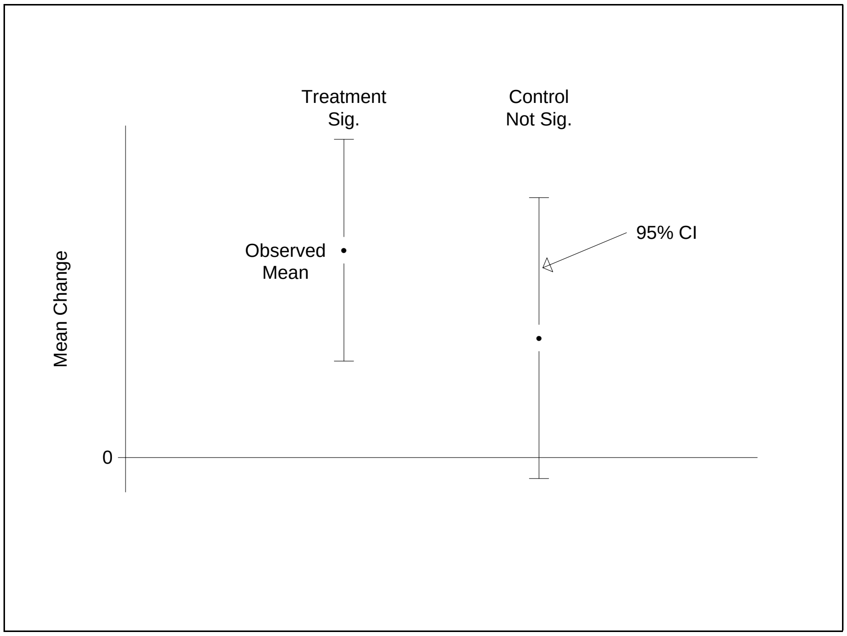
The fallacies in the above reasoning will be demonstrated by means of a couple of examples.

Example 1

Consider the following sample data from two populations $G1$ and $G2$.

i	$G1$	$G2$
1	1.412	0.041
2	2.338	−1.306
3	3.546	−0.068
4	0.119	−0.070
5	−0.162	2.255
6	0.496	2.462
7	1.768	0.970
8	2.913	2.959
9	1.545	2.459
10	1.942	−0.521
$\bar{x}$	1.592	0.918
s	1.190	1.507

The following plot shows the 95% confidence intervals resulting from the given sample data.



As you can see, a test of the hypothesis $H_0 : \mu_{G1} = 0$ versus $H_1 : \mu_{G1} \neq 0$ would result in $p < 0.05$ while testing $H_0 : \mu_{G2} = 0$ versus $H_1 : \mu_{G2} \neq 0$ would give $p > 0.05$. Based on this you may want to conclude that $\mu_{G1} \neq \mu_{G2}$.

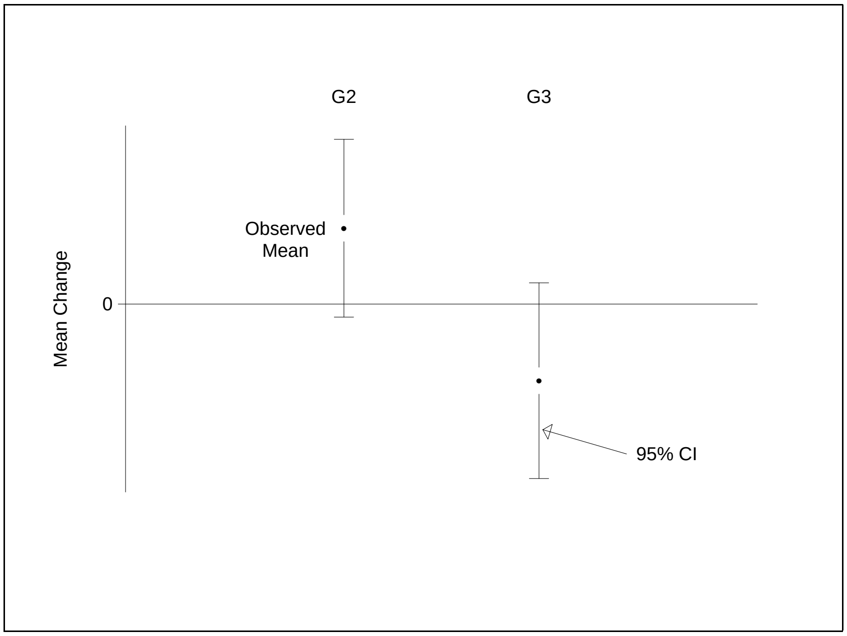
However, performing a formal t-test of $H_0 : \mu_{G1} = \mu_{G2}$ versus $H_1 : \mu_{G1} \neq \mu_{G2}$ gives $p = 0.28$. In other words, you would not be able to conclude that the means in populations $G1$ and $G2$ were statistically significantly different.

Example 2

In this example, we will compare the sample from $G2$ to a sample from a population denoted by $G3$. The resulting data are as shown.

i	$G2$	$G3$
1	0.041	−1.787
2	−1.306	2.357
3	−0.068	0.944
4	−0.070	−1.718
5	2.255	−2.069
6	2.462	−1.927
7	0.970	−1.585
8	2.959	−3.187
9	2.459	0.001
10	−0.521	−0.332
$\bar{x}$	0.918	−0.930
s	1.507	1.657

Consider the plot of the 95% confidence intervals generated from these data.



In this plot, there is some overlap of the confidence intervals and both confidence intervals also include 0. A formal t-test of $H_0 : \mu_{G2} = \mu_{G3}$ versus $H_1 : \mu_{G2} \neq \mu_{G3}$ gives $p = 0.018$. This would be considered strong evidence of a difference in the means between the two populations.