# Overlapping Confidence Intervals and Statistical Significance

Andrea Knezevic

## 1 Introduction

In this document, we address the following question: can we judge whether two statistics are significantly different based on whether or not their confidence intervals overlap? The short answer is: not always. If two statistics have non-overlapping confidence intervals, they are necessarily significantly different but if they have overlapping confidence intervals, it is not necessarily true that they are not significantly different.

## 2 Example

We can illustrate this with a simple example. Suppose we are interested in comparing means from two independent samples. The mean of the first sample is 9 and the mean of the second sample is 17. Let's assume that the two group means have the same standard errors, equal to 2.5. The 95 percent confidence interval for the first group mean can be calculated as: $9 \pm 1.96 \times 2.5$ where 1.96 is the critical t-value. The confidence interval for the first group mean is thus $(4.1, 13.9)$. Similarly for the second group, the confidence interval for the mean is $(12.1, 21.9)$. Notice that the two intervals overlap. However, the t-statistic for comparing two means is

$$t = \frac{17 - 9}{\sqrt{2.5^2 + 2.5^2}} = 2.26,$$

which indicates that the null hypothesis, that the means of the two groups are the same, should be rejected at the $\alpha = 0.05$ level.

To verify the above conclusion, consider the 95 percent confidence interval for the difference between the two group means:

$$(17 - 9) \pm 1.96 \times \sqrt{2.5^2 + 2.5^2},$$

which yields the interval $(1.09, 14.91)$. The interval does not contain zero, hence we reject the null hypothesis that the group means are the same.

Generally, when comparing two parameter estimates, if the 95-percent confidence intervals do not overlap, then the null hypothesis of zero difference between the parameters will be rejected at the 0.05 level. However, the converse is not true. That is, if the 95-percent confidence intervals *do* overlap, then we cannot determine whether the null hypothesis of zero difference between the parameters will be rejected at the 0.05 level. It is necessary to perform the statistical test for this null hypothesis using the appropriate test statistic, or to compute the confidence interval for the difference in the parameters.

For an explanation of why this is true when comparing two means using independent samples, see the Technical Details section at the end of this document.

# 3 Technical Details

Consider the comparison of two means based on independent samples. The bounds of each individual confidence interval are based on the magnitude of the corresponding standard errors, $SE_1$ and $SE_2$. The bounds of the confidence interval for the mean *difference* are based on the standard error of the difference, which is $\sqrt{SE_1^2 + SE_2^2}$. It is the relationship between the standard error for the difference and the individual standard errors which leads to the conclusions about overlapping confidence intervals described in this newsletter.

If we label the two sample means as $x_1$ and $x_2$, then the individual confidence intervals are

$$x_1 \pm t \times SE_1 \quad \text{and} \quad x_2 \pm t \times SE_2$$

where $t$ is an appropriate critical value. These individual confidence intervals do not overlap when $x_1 - t \times SE_1 > x_2 + t \times SE_2$ or $x_2 - t \times SE_2 > x_1 + t \times SE_1$.

The null hypothesis of zero population difference between the means is rejected when $|x_1 - x_2| > t \times \sqrt{SE_1^2 + SE_2^2}$, i.e. when the confidence interval for the difference does not include zero.

After some algebraic manipulation, we see that

- The null hypothesis of zero mean difference is rejected when $|x_1 - x_2| > t \times \sqrt{SE_1^2 + SE_2^2}$
- The individual confidence intervals do not overlap when $|x_1 - x_2| > t \times (SE_1 + SE_2)$

Now, it can be shown that $\sqrt{SE_1^2 + SE_2^2} \le SE_1 + SE_2$ is always true. This means that as $|x_1 - x_2|$ (the distance between the means) increases, it will exceed $t \times \sqrt{SE_1^2 + SE_2^2}$ before it exceeds $t \times (SE_1 + SE_2)$. When this is true, that is, when $t \times \sqrt{SE_1^2 + SE_2^2} \le |x_1 - x_2| \le t \times (SE_1 + SE_2)$, then the individual confidence intervals will overlap but the null hypothesis of zero mean difference will be rejected.

Figure 3.1 illustrates how $|x_1 - x_2|$ can be large enough so that the null hypothesis of zero mean difference is rejected, but not large enough so that the two individual confidence intervals do not overlap. When $|x_1 - x_2|$ is greater than $t \times (SE_1 + SE_2)$ and the confidence intervals do not overlap, then we know that $|x_1 - x_2|$ is also greater than $t \times \sqrt{SE_1^2 + SE_2^2}$ and hence the null hypothesis of zero mean difference is rejected.

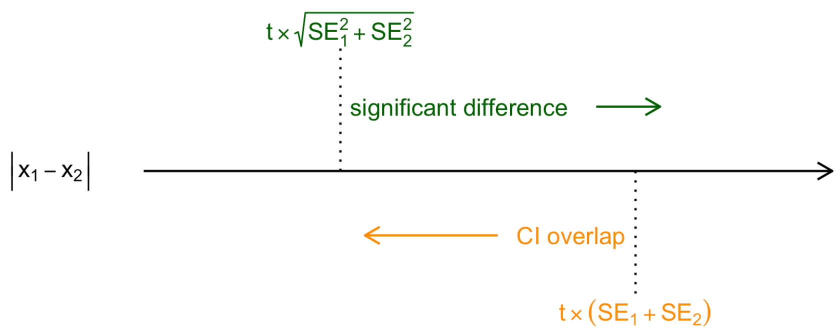


*Figure 3.1: Relationship among $|x_1 - x_2|$, the standard errors, confidence interval overlap, and rejection of the null hypothesis of zero mean difference.*

---

Created October 2008. Last updated April 2022.