by
Lloyd S. Nelson

Evaluating Overlapping Confidence Intervals

RECENTLY an informal poll showed that far too many people incorrectly believe that two means can be said to be significantly different at exactly the 5 percent significance level if the lower 95 percent confidence limit of the larger mean equals the upper 95 percent confidence limit of the smaller mean. The situation is illustrated in Figure 1A, where $\bar{Y}_1$ and $\bar{Y}_2$ represent means having 95 percent confidence intervals that just do not overlap. In fact the significance of the difference between such means can lie anywhere from 0.05 to 0.0056.

Twenty years ago Barr (1969) explained the exact relationship between a significance test on the difference between two means (a one-interval test) and the overlap of confidence intervals around each mean (a two-interval test). It would appear useful to review this. The more widespread use of confidence intervals in recent years has made the subject treated here more important. In particular it was brought to light in connection with box plots by McGill, Tukey, and Larsen (1978).

Example

Consider the following data and analyses. We shall assume normality of errors and a common, known standard deviation.

$$\bar{X}_1 = 20.602 \quad \bar{X}_2 = 24.369$$

$$n_1 = 16 \qquad n_2 = 25$$

$$\sigma = 6$$

$$H_0: \quad \mu_1 = \mu_2$$

$$H_1: \quad \mu_1 \neq \mu_2 \text{ (2-sided test)}$$

$$\alpha = 0.05$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$= \frac{20.602 - 24.369}{6} \sqrt{\frac{(16)(25)}{16 + 25}} = -1.961.$$

Do not reject H_0 if $-1.96 \leq Z \leq 1.96$. Conclusion: Reject H_0 because $Z = -1.961$ is less than $z_{0.025} = -1.960$. A confidence interval on the true difference $(\mu_1 - \mu_2)$ is

$$(\bar{X}_1 - \bar{X}_2) + z_{0.025}\sigma \sqrt{\frac{n_1 + n_2}{n_1 n_2}} < \mu_1 - \mu_2$$

$$< (\bar{X}_1 - \bar{X}_2) + z_{0.975}\sigma \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

$$= -3.767 + (-1.96)(6)\sqrt{0.1025}$$

$$< \mu_1 - \mu_2 < -3.767 + (1.96)(6)\sqrt{0.1025}$$

$$= -7.532 < \mu_1 - \mu_2 < -0.002.$$

Notice that this 95 percent confidence interval barely excludes zero as an admissible value for $\mu_1 - \mu_2$. This is consistent with the above hypothesis test that barely rejected $\mu_1 - \mu_2 = 0$ at the 5 percent significance level.

Let us now look at 95 percent confidence intervals for μ_1 and μ_2, separately. These are

$$\bar{X}_1 + z_{0.025}\sigma/\sqrt{n_1} < \mu_1 < \bar{X}_1 + z_{0.975}\sigma/\sqrt{n_1}$$

$$17.662 < \mu_1 < 23.542 \qquad (1)$$

and

$$\bar{X}_2 + z_{0.025}\sigma/\sqrt{n_2} < \mu_2 < \bar{X}_2 + z_{0.975}\sigma/\sqrt{n_2}$$

$$22.017 < \mu_2 < 26.721. \qquad (2)$$

The plot in Figure 1B is of these two intervals. About 32 percent of the shorter interval overlaps about 26 percent of the longer interval. This hardly appears to be in accord with the fact that the difference is significantly different from zero at the 5 percent level. The difficulty lies in the fact that the two-interval test was not carried out correctly.

Extention of Example

It was shown by Barr (1969) that the length of the confidence intervals for the two-interval test must be constructed with the multiplier

$$z' = \frac{\sqrt{n_1 + n_2}}{\sqrt{n_1} + \sqrt{n_2}} z_{0.975}$$

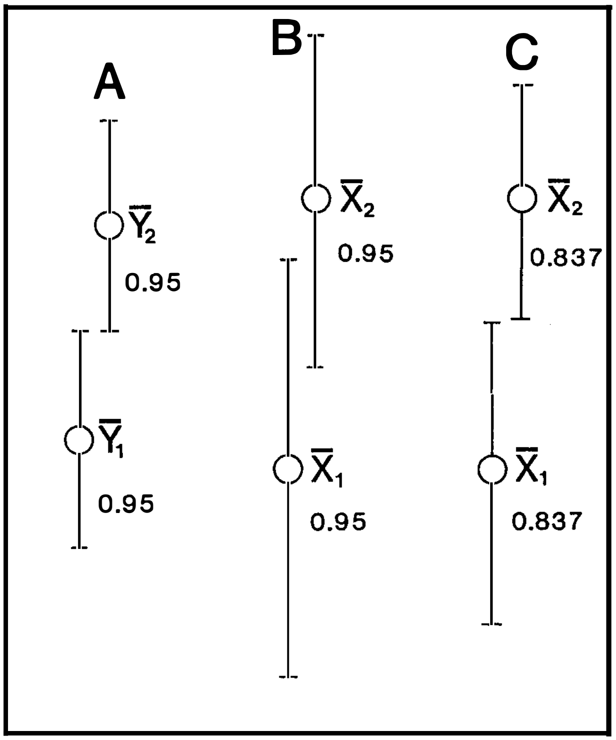

FIGURE 1. Each Confidence Interval is Labeled with Its Confidence Coefficient. (A) General Representation Showing the Lower 95 Percent Confidence Limit of the Larger Mean Equal to the Upper 95 Percent Confidence Limit of the Smaller Mean. (B) 95 Percent Confidence Intervals for the Means in the Example. (C) 83.7 Percent Confidence Intervals for the Means in the Example.

if significance at the 5 percent level is to be declared when the intervals are just nonoverlapping.

Substituting $-z'$ for $z_{0.025}$ and z' for $z_{0.975}$ in equations (1) and (2) gives the intervals

$$18.51 < \mu_1 < 22.69$$

$$22.70 < \mu_2 < 26.04$$

which now enables significance at exactly the 5 percent level to be judged by their just not overlapping. The fact that these intervals barely fail to overlap agrees with the previous rejection of $\mu_1 - \mu_2 = 0$ at slightly greater than 0.05 significance. Figure 1C shows this situation. Looking up $z' = 1.394$ in a normal table shows that these limits have a confidence coefficient of 0.837. We can easily find the exact significance associated with a hypothesized difference of zero when the 95 percent confidence intervals are just not overlapping. It ranges from 0.0056 for means with equal standard errors up to 0.05 for means with vastly different standard errors. Thus the most common case of means with nearly equal standard errors can have its significance greatly understated.

When a common standard deviation (as is assumed in the example given here) is estimated from the data, then Student's t can be used in place of z. Also the procedure can be generalized to the situation in which the two samples have different known standard deviations.

References

BARR, D. R. (1969). "Using Confidence Intervals to Test Hypotheses." *Journal of Quality Technology* 1, pp. 256–258.

MCGILL, R.; TUKEY, J. W.; and LARSEN, W. A. (1978). "Variations of Box Plots." *The American Statistician* 32, pp. 12–16.

Key Words: *Confidence Intervals, Significance Tests.*