Reminder about Confidence Intervals

2   Reminder about Confidence Intervals

How to compute a confidence interval around a point estimator. As

illustration, we will explain how to compute a confidence interval around Cohen's d (the

explanation would be very similar for all other estimators). Under the assumption of iid

normal distribution of residuals with equal population variances across groups, in order to

test the null hypothesis that $\mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$, we can compute the following quantity:

$$t_{Student} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{SE}$$

with $SE = S_{pooled} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, $S_{pooled} = \sqrt{\frac{(n_1-1)\times S_1^2 + (n_2-1)\times S_2^2}{n_1+n_2-2}}$ and $S_j =$ the standard

deviation of the jth sample ($j = 1, 2$). Under the null hypothesis, this quantity will follow a

central t-distribution with $n_1 + n_2 - 2$ degrees of freedom. However, when the null

hypothesis is false, the distribution of this quantity is not centered and a noncentral

t-distribution arises (Cumming & Finch, 2001), as illustrated in Figure 1. Noncentral

t-distributions are described by two parameters: degrees of freedom (df) and noncentrality

parameter (that we will call Δ, Cumming & Finch, 2001), the last being a function of the

population effect size (δ) and sample sizes (n_1 and n_2):

$$\Delta = \frac{\delta}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

3   with $\delta = \frac{(\mu_1-\mu_2)-(\mu_1-\mu_2)_0}{\sigma_{pooled}}$, $\sigma_{pooled} = \sqrt{\frac{(n_1-1)\times\sigma_1^2+(n_2-1)\times\sigma_2^2}{n_1+n_2-2}}$ and $\sigma_j =$ the standard deviation of

4   the jth population ($j = 1, 2$). Considering the link between Δ and δ, it is possible to

5   compute confidence limits for Δ and multiply them by $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ in order to have confidence

6   limits for δ. In other words, we first need to determine the noncentrality parameters of the

7   t-distributions for which $t_{Student}$ corresponds respectively to the quantiles $\left(1 - \frac{\alpha}{2}\right)$ and $\frac{\alpha}{2}$:

$$P[t_{df,\Delta_L} \geq t_{Student}] = \frac{\alpha}{2}$$

8

$$P[t_{df,\Delta_U} \leq t_{Student}] = \frac{\alpha}{2}$$

9 with $df = n_1 + n_2 - 2$. Second, we multiply Δ_L and Δ_U by $\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ in order to define δ_L

10 and δ_U:

$$\delta_L = \Delta_L \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

11

$$\delta_U = \Delta_U \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

12   **Reference**

13 Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of

14   confidence intervals that are based on central and noncentral distributions. *Educational*

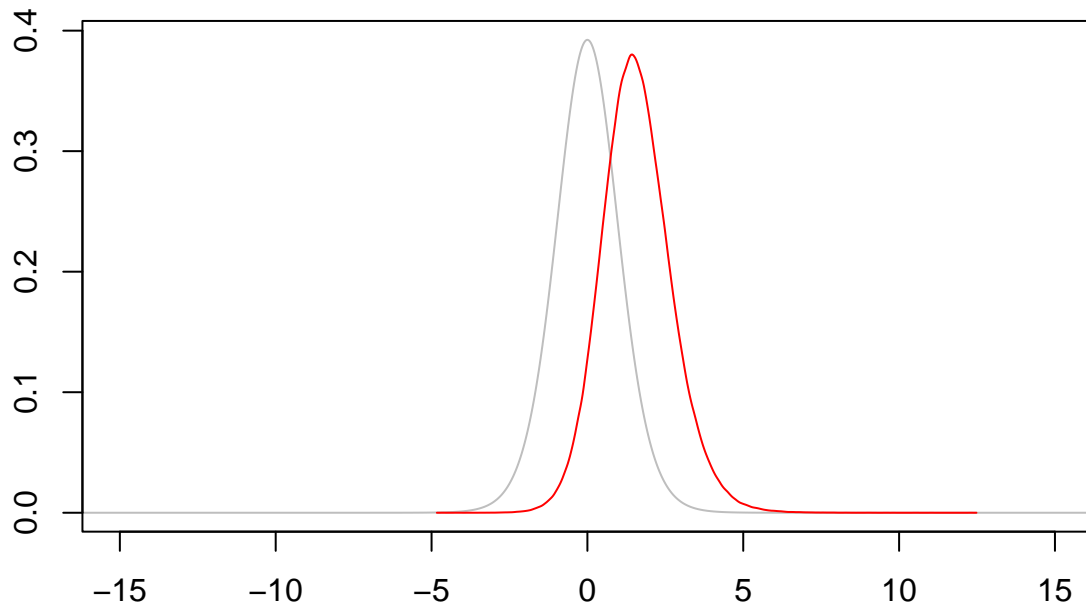15   *and Psychological Measurement, 61*(4), 532–574.

Figure 1. Sampling distribution of $t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{SE}$ when the null hypothesis is true (in grey) and when the null hypothesis is false, with $(\mu_1 - \mu_2) - (\mu_1 - \mu_2)_0 = 4$ and $SE = 5$ (in red).