# Reminder about Confidence Intervals

Marie Delacre[1]

[1] ULB

## Author Note

Correspondence concerning this article should be addressed to Marie Delacre, Postal address. E-mail: marie.delacre@ulb.ac.be

## Abstract

*Keywords:* keywords

Word count: X

# Reminder about Confidence Intervals

## Reference

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervales that are based on central and noncentral distributions. Educational and Psychological Measurement, 61(532).

## Method based on the use of a pivotal quantity

When computing a (supposed normal) centered variable, divided by the standard error (i.e. an independant variable closely related with the $\chi^2$ distribution), then computed quantity will follow a central $t$-distribution. This quantity is called a pivotal quantity (PQ), i.e. a quantity that is very interesting because its sampling distribution is not a function of the parameter we want to estimate (Cox & Hinkley, 1974 cited by Cumming and Finch, 2001). We can therefore use it, in order to define confidence limits for any parameter.

The method consists in four steps:

1) Compute a pivotal quantity (PQ) of the general form: (Estimator - parameter)/SE;

2) Determining the distribution of PQ;

3) Computing the confidence limits of PQ: determine a range of values, centered around 0, such as (1-alpha)% of the area under the distribution of PQ falls in this range;

4) Pivote in order to obtain the confidence interval around the parameter of interest.

As a first example, consider the case of 2 means difference, assuming normality and homoscedasticity. The pivotal quantity is defined as follows:

$$PQ = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{SE} \tag{1}$$

With $SE = \sigma_{pooled} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ and $\sigma_{pooled} = \sqrt{\frac{(n_1-1)*S_1^2+(n_2-1)*S_2^2}{n_1+n_2-2}}$

This quantity follows a $t$- distribution with $n_1 + n_2 - 2$ degrees of freedom (therefore, it depends only on $n_1$ and $n_2$, it does NOT depend on the parameter of interest, i.e. $\mu_1 - \mu_2$).

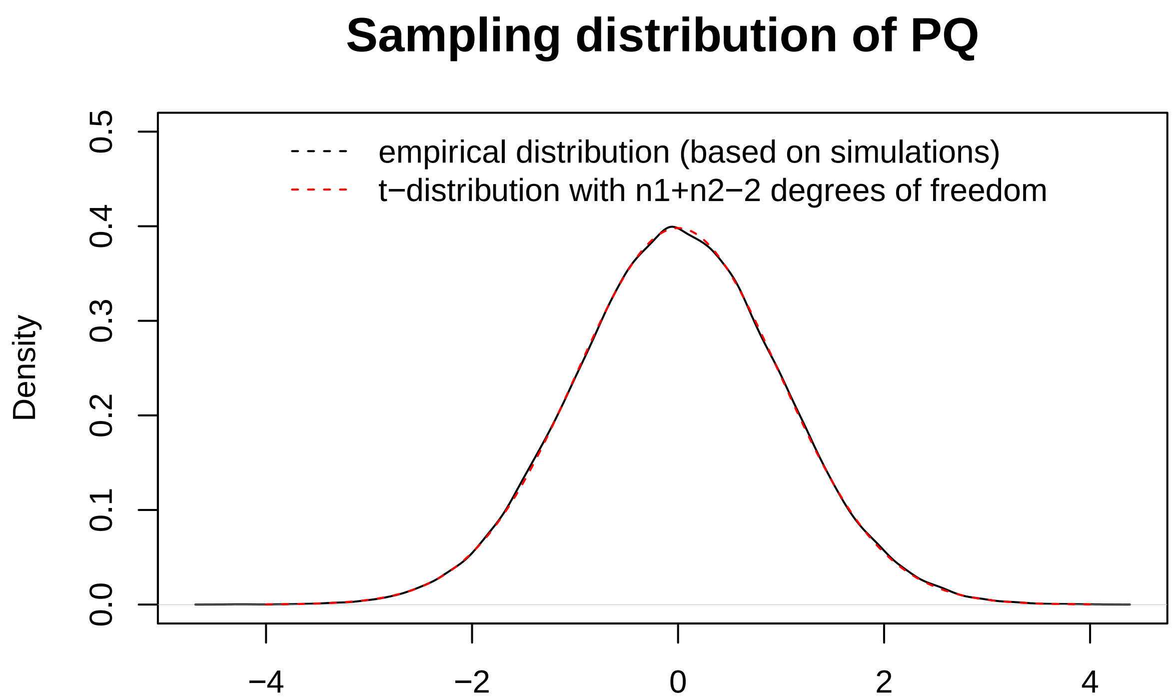


*Figure 1*. Sampling distribution of the pivotal quantity under the assumptions of normality and homoscedasticity

Because the theoretical distribution of PQ is known, one can compute the confidence limits, for any confidence level:

$$Pr[t_{n_1+n_2-2}(\frac{\alpha}{2}) < \frac{(\bar{X_1} - \bar{X_2}) - (\mu_1 - \mu_2)}{SE} < t_{n_1+n_2-2}(1 - \frac{\alpha}{2})] = 1 - \alpha \tag{2}$$

Because the $t$-distribution is symmetrically centered around 0, one can deduce that $t_{n_1+n_2-2}(\frac{\alpha}{2}) = -t_{n_1+n_2-2}(1 - \frac{\alpha}{2})$, and therefore:

$$Pr[-t_{n_1+n_2-2}(1-\frac{\alpha}{2}) < \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{SE} < t_{n_1+n_2-2}(1-\frac{\alpha}{2})] = 1-\alpha \tag{3}$$

In pivoting the inequation, one can deduce that:

$$Pr[-t_{n_1+n_2-2}(1-\frac{\alpha}{2}) \times SE < (\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2) < t_{n_1+n_2-2}(1-\frac{\alpha}{2}) \times SE] = 1-\alpha \tag{4}$$

$$\leftrightarrow Pr[-(\bar{X}_1-\bar{X}_2)-t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE < -(\mu_1-\mu_2) < -(\bar{X}_1-\bar{X}_2)+t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE] = 1-\alpha \tag{5}$$

$$\leftrightarrow Pr[(\bar{X}_1-\bar{X}_2)+t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE > \mu_1-\mu_2 > (\bar{X}_1-\bar{X}_2)-t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE] = 1-\alpha \tag{6}$$

$$\leftrightarrow Pr[(\bar{X}_1-\bar{X}_2)-t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE < \mu_1-\mu_2 < (\bar{X}_1-\bar{X}_2)+t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE] = 1-\alpha \tag{7}$$

As a second example, consider the case of 2 means difference, assuming normality and heteroscedasticity. The pivotal quantity is defined as follows:

$$PQ = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{SE} \tag{8}$$

With $SE = \sqrt{\frac{S_1^2}{n1}+\frac{S_2^2}{n2}}$

This quantity follows a $t$- distribution with $\frac{(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1-1}+\frac{(\frac{S_2^2}{n_2})^2}{n_2-1}}$ degrees of freedom (therefore, it depends on $n_1$ and $n_2$, $S_1$ and $S_2$, and does NOT depend on the parameter of interest, i.e. $\mu_1 - \mu_2$).

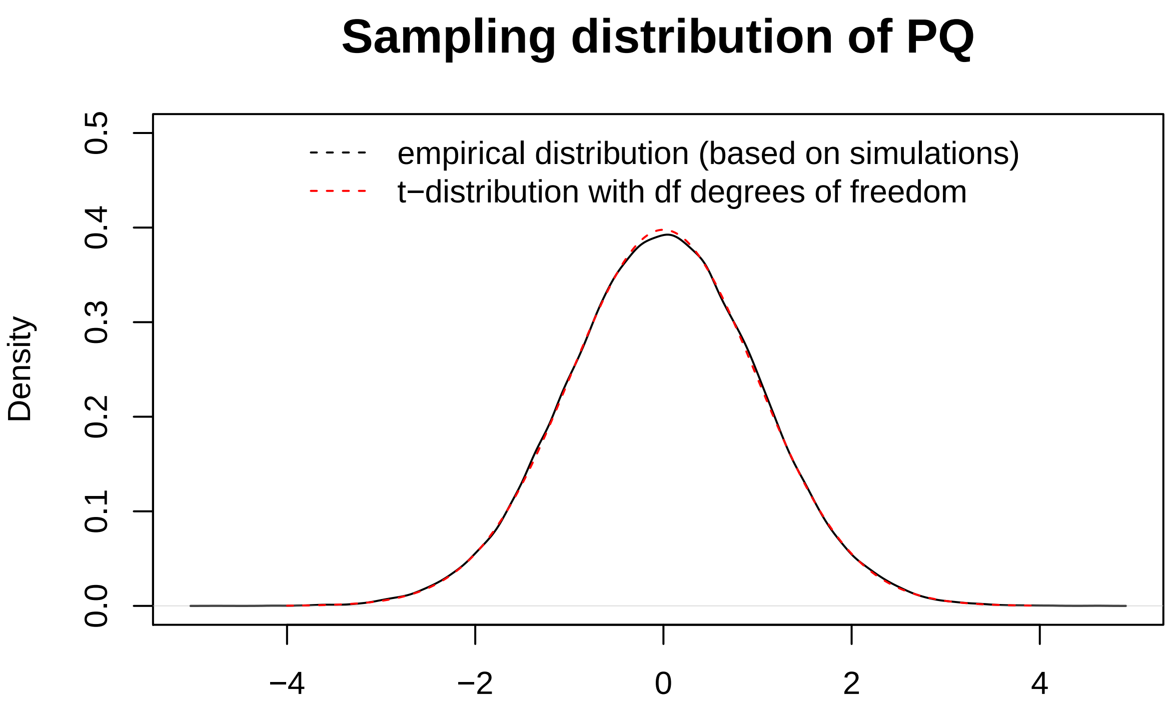


*Figure 2*. Sampling distribution of the pivotal quantity under the assumptions of normality and heteroscedasticity

Because the theoretical distribution of PQ is known, one can compute the confidence limits, for any confidence level (see the first example for more details):

$$Pr[(\bar{X_1}-\bar{X_2})-t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE < \mu_1-\mu_2 < (\bar{X_1}-\bar{X_2})+t_{n_1+n_2-2}(1-\frac{\alpha}{2})\times SE] = 1-\alpha \quad (9)$$

With SE $= \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$