3. Confidence Intervals for Difference of Two Correlation Coefficients Based on Two Independent Samples

In the ith population, i=1, 2, let $\left(X^{(i)}, Y^{(i)}\right)$ be a bivariate random variable from the ith population. Let $\left(X_j^{(i)}, Y_j^{(i)}\right)$, $j = 1, ..., n_i$, be the sample of size n_i from the ith population i =1, 2, and $r^{(i)}$ the Pearson correlation coefficient, i.e.,

$$r^{(i)} = \frac{\sum_{j=1}^{n_i}\left(X_j^{(i)} - \overline{X}^{(i)}\right)\left(Y_j^{(i)} - \overline{Y}^{(i)}\right)}{\sqrt{\sum_{j=1}^{n_i}\left(X_j^{(i)} - \overline{X}^{(i)}\right)^2 \sum_{j=1}^{n_i}\left(Y_j^{(i)} - \overline{Y}^{(i)}\right)^2}}$$

$$= \frac{\sum_{j=1}^{n_i}\left(X_j^{(i)} - \overline{X}^{(i)}\right)\left(Y_j^{(i)} - \overline{Y}^{(i)}\right)}{n_i S_X^{(i)} S_Y^{(i)}},$$

where $\overline{X}^{(i)}$ and $S_X^{(i)}$ are the sample mean and standard derivation of X in the ith sample; $\overline{Y}^{(i)}$ and $S_Y^{(i)}$ are similarly defined.

Our proposed confidence intervals I, II, and III are described in the following.

3.1. Confidence Interval I

. By Slutsky Theorem and Central Limit Theorem, we have

$$\frac{\sum_{j=1}^{n_i} M_j^{(i)}}{n_i S_X^{(i)} S_Y^{(i)}} \to N\left(\rho^{(i)}, \frac{\sigma_M^{2(i)}}{n_i \sigma_X^{2(i)} \sigma_Y^{2(i)}}\right) \qquad \text{as } n \to \infty.$$

where

$M_j^{(i)} = \left(X_j^{(i)} - \overline{X}^{(i)}\right)\left(Y_j^{(i)} - \overline{Y}^{(i)}\right),$

$\sigma_M^{2(i)} = E\left(\left(X_j^{(i)} - \mu_X^{(i)}\right)^2 \left(Y_j^{(i)} - \mu_Y^{(i)}\right)^2\right) + \rho^{2(i)} \sigma_X^{2(i)} \sigma_Y^{2(i)},$

$\mu_X^{(i)} = E\left(X^{(i)}\right),$

$\mu_Y^{(i)} = E\left(Y^{(i)}\right),$

$\sigma_X^{2(i)}$ and $\sigma_Y^{2(i)}$ are variances of X and Y in the ith sample.

So we have

$$r^{(1)} - r^{(2)} \to N\left(\rho^{(1)} - \rho^{(2)}, \frac{\sigma_M^{2(1)}}{n_1 \sigma_X^{2(1)} \sigma_Y^{2(1)}} + \frac{\sigma_M^{2(2)}}{n_2 \sigma_X^{2(2)} \sigma_Y^{2(2)}}\right)$$

$$\Pr\left(\frac{r^{(1)} - r^{(2)} - \left(\rho^{(1)} - \rho^{(2)}\right)}{\zeta} \le Z\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha,$$

where $Z\left(1 - \frac{\alpha}{2}\right)$ is the $100(1 - \frac{\alpha}{2})\%$ quantile of the standard normal distribution, and

$$\zeta = \sqrt{\frac{\sigma_M^{2(1)}}{n_1 \sigma_X^{2(1)} \sigma_Y^{2(1)}} + \frac{\sigma_M^{2(2)}}{n_2 \sigma_X^{2(2)} \sigma_Y^{2(2)}}}.$$

Applying Slutsky Theorem again, in large sample

$$\Pr\left(\frac{r^{(1)} - r^{(2)} - \left(\rho^{(1)} - \rho^{(2)}\right)}{\widehat{\zeta}} \le Z\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha \tag{3.1}$$

where

$$\widehat{\zeta} = \sqrt{\frac{S_M^{2(1)}}{n_1 S_X^{2(1)} S_Y^{2(1)}} + \frac{S_M^{2(2)}}{n_2 S_X^{2(2)} S_Y^{2(2)}}},$$

$$S_M^{2(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(X_j^{(i)} - \mu_X^{(i)}\right)^2 \left(Y_j^{(i)} - \mu_Y^{(i)}\right)^2 + r^{2(i)} S_X^{2(i)} S_Y^{2(i)}, \ i = 1, 2.$$

According (3.1), the proposed confidence interval I for $\left(\rho^{(1)} - \rho^{(2)}\right)$ is given by

$$\left[r^{(1)} - r^{(2)} - Z\left(1 - \frac{\alpha}{2}\right)\widehat{\zeta}, r^{(1)} - r^{(2)} + Z\left(1 - \frac{\alpha}{2}\right)\widehat{\zeta}\right] \cap [-2, 2]$$

Note that this confidence interval does not require distributional assumptions on (X, Y). We only need the sample size to be sufficiently large to ensure the Central Limit Theorem holds, which is the theoretical basis of this confidence interval.

3.2. Confidence Interval II

. The confidence interval II we proposed is based on Theorem 2. Assume that the observations from each of two populations are bivariate normal and hence $\kappa = 0$.

Let $d = r^{(1)} - r^{(2)}$. According to Theorem 2, the asymptotic distribution of d is

$$N(\rho^{(1)} - \rho^{(2)}, \frac{\left(1-\rho^{2(1)}\right)^2}{n_1 - 1} + \frac{\left(1-\rho^{2(2)}\right)^2}{n_2 - 1}),$$

where the two samples are independent.

So

$$\frac{d - \left(\rho^{(1)} - \rho^{(2)}\right)}{\sqrt{\frac{\left(1-\rho^{2(1)}\right)^2}{n_1-1} + \frac{\left(1-\rho^{2(2)}\right)^2}{n_2-1}}} \to N(0,1) \ \ as\ n \to \infty,$$

and by Slutsky Theorem,

$$\frac{d - \left(\rho^{(1)} - \rho^{(2)}\right)}{\sqrt{\frac{\left(1-r^{2(1)}\right)^2}{n_1-1} + \frac{\left(1-r^{2(2)}\right)^2}{n_2-1}}} \to N(0,1) \ \ as\ n \to \infty,$$

Hence,

$$\Pr\left(\left|\frac{d - \left(\rho^{(1)} - \rho^{(2)}\right)}{\sqrt{\frac{\left(1-\rho^{2(1)}\right)^2}{n_1-1} + \frac{\left(1-\rho^{2(2)}\right)^2}{n_2-1}}}\right| \leq Z\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$$

and we obtain the $100\,(1-\alpha)\,\%$ confidence interval of $\left(\rho^{(1)} - \rho^{(2)}\right)$ as

$$\left[d - Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\left(1-r^{2(1)}\right)^2}{n_1-1} + \frac{\left(1-r^{2(2)}\right)^2}{n_2-1}}, d + Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{\left(1-r^{2(1)}\right)^2}{n_1-1} + \frac{\left(1-r^{2(2)}\right)^2}{n_2-1}}\right] \cap [-2, 2]$$

Note that, unlike confidence interval I, here we need to assume normality for the observations.

3.3. Confidence Interval III

. Assume $\left(X^{(i)}, Y^{(i)}\right)$, $i = 1$, 2, are bivariate normal. From Theorem 3,

$$z^{(i)} \to N\left(\xi^{(i)}, \frac{1}{n^{(i)}-1}\right) \qquad \text{as } n_i \ \to \infty,$$

where

$$z^{(i)} = \frac{1}{2}\log\frac{1+r^{(i)}}{1-r^{(i)}},$$

$$\xi^{(i)} = \frac{1}{2}\log\frac{1+\rho^{(i)}}{1-\rho^{(i)}},\ \ i = 1,\ 2.$$

so

$$z^{(1)} - z^{(2)} \to N\left(\xi^{(1)} - \xi^{(2)}, \frac{1}{n_1-1} + \frac{1}{n_2-1}\right) \qquad \text{as } n \to \infty$$

and

$$\Pr\left(\left|\frac{\left(z^{(1)}-z^{(2)}\right)-\left(\xi^{(1)}-\xi^{(2)}\right)}{\sqrt{\frac{1}{n_1-1}+\frac{1}{n_2-1}}}\right| \leq Z\left(1-\frac{\alpha}{2}\right)\right)$$

Let $\delta = z^{(1)} - z^{(2)}$. The $100\,(1-\alpha)\,\%$ confidence interval for $\xi^{(1)} - \xi^{(2)}$ is given by

$$\delta - Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n_1-1}+\frac{1}{n_2-1}} \leq \left(\xi^{(1)}-\xi^{(2)}\right) \leq \delta + Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n_1-1}+\frac{1}{n_2-1}}$$

that is

$$\begin{aligned}\delta - Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n_1-1}+\frac{1}{n_2-1}} \ &\leq\ \frac{1}{2}\log\frac{1+\rho^{(1)}-\rho^{(2)}-\rho^{(1)}\rho^{(2)}}{1-\rho^{(1)}+\rho^{(2)}-\rho^{(1)}\rho^{(2)}} \\ &\leq\ \delta + Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n_1-1}+\frac{1}{n_2-1}}\end{aligned}$$

The above inequality can be simplified as

$$\exp\left(2\tau^{(1)}\right) \leq \frac{1+\rho^{(1)}-\rho^{(2)}-\rho^{(1)}\rho^{(2)}}{1-\rho^{(1)}+\rho^{(2)}-\rho^{(1)}\rho^{(2)}} \leq \exp\left(2\tau^{(2)}\right)$$

where

$$\tau^{(1)} = \delta - Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$\tau^{(2)} = \delta + Z\left(1-\frac{\alpha}{2}\right)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

Note that $1-\rho^{(1)}+\rho^{(2)}-\rho^{(1)}\rho^{(2)} = \left(1-\rho^{(1)}\right)\left(1+\rho^{(2)}\right) > 0$. So

$$\begin{aligned}\left(1-\rho^{(1)}+\rho^{(2)}-\rho^{(1)}\rho^{(2)}\right)\exp\left(\tau^{(1)}\right) &\leq 1+\rho^{(1)}-\rho^{(2)}-\rho^{(1)}\rho^{(2)} \\ &\leq \left(1-\rho^{(1)}+\rho^{(2)}-\rho^{(1)}\rho^{(2)}\right)\exp\left(\tau^{(2)}\right).\end{aligned}$$

Substituting $r^{(1)}r^{(2)}$ to $\rho^{(1)}\rho^{(2)}$, the above inequality reduces to

$$\begin{aligned}\left(1-\rho^{(1)}+\rho^{(2)}-\rho^{(1)}\rho^{(2)}\right)\exp\left(2\tau^{(1)}\right) &\leq 1+\rho^{(1)}-\rho^{(2)}-r^{(1)}r^{(2)} \\ &\leq \left(1-\rho^{(1)}+\rho^{(2)}-\rho^{(1)}\rho^{(2)}\right)\exp\left(2\tau^{(2)}\right),\end{aligned}$$

which in turn leads to

$$\left(1-r^{(1)}r^{(2)}\right)\frac{\exp\left(2\tau^{(1)}\right)-1}{\exp\left(2\tau^{(1)}\right)+1} \leq \left(\rho^{(1)}-\rho^{(2)}\right) \leq \left(1-r^{(1)}r^{(2)}\right)\frac{\exp\left(2\tau^{(2)}\right)-1}{\exp\left(2\tau^{(2)}\right)+1}$$

We thus propose the confidence interval III for $(\rho_1-\rho_2)$ given by

$$\left[\left(1-r^{(1)}r^{(2)}\right)\frac{\exp\left(2\tau^{(1)}\right)-1}{\exp\left(2\tau^{(1)}\right)+1}, \left(1-r^{(1)}r^{(2)}\right)\frac{\exp\left(2\tau^{(2)}\right)-1}{\exp\left(2\tau^{(2)}\right)+1}\right]. \tag{3.2}$$

We note that the interval (3.2) is always within the range of $[-2,2]$, the most widest range that $\rho^{(1)}-\rho^{(2)}$ can take.