Partial Correlation Coefficients

Example. In a study, X_1 = weekly amount of coffee/tea sold by a refreshment stand at a summer resort, and X_2 = weekly number of visitors to the resort. If X_2 is large, so should X_1 be, right? Actually no! With a certain resort, $r_{12} = -0.3$. Why? Consider X_3 = average weekly temperature at the resort. Both X_1 and X_2 are related to X_3. If temperature is high, there will be more visitors, but they will prefer cold drinks to coffee/tea. If temperature is low, there will be fewer visitors, but they will prefer coffee/tea. Say, $r_{13} = -0.7$, $r_{23} = .8$. It is then more meaningful to investigate the relationship between X_1 and X_2 conditional on X_3 (i.e., when X_3 is kept fixed) to eliminate the effect of X_3.

Partial correlation coefficient between X_1 and X_2 when X_3 is fixed is

$$r_{12.3} = Corr(X_1|X_3, X_2|X_3) = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}.$$

Suppose $\mathbf{X} \sim N_m(\mu, \Sigma)$ and partition $\mathbf{X}$, μ and Σ as:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix},$$

where $\mathbf{X}_1$ is k-dimensional. Then $\mathbf{X}_1|\mathbf{X}_2 \sim N_k(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2), \Sigma_{11.2})$, where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}' = ((\sigma_{ij.k+1,\ldots,m}))$. Note that $\sigma_{ij.k+1,\ldots,m}$ = partial covariance between X_i and X_j conditional on $\mathbf{X}_2 = (X_{k+1}, \ldots, X_m)'$. Therefore the partial correlation coefficient between X_i and X_j given $\mathbf{X}_2$ is

$$\rho_{ij.k+1,\ldots,m} = \frac{\sigma_{ij.k+1,\ldots,m}}{\sqrt{\sigma_{ii.k+1,\ldots,m}}\sqrt{\sigma_{jj.k+1,\ldots,m}}}.$$

Recall the notation, ρ for the population and r for a sample. From the expression for $\Sigma_{11.2}$ note that $\sigma_{ij.l} = \sigma_{ij} - \sigma_{il}\sigma_{jl}/\sigma_{ll}$. Thus,

$$\begin{aligned} \rho_{ij.l} &= \frac{\sigma_{ij.l}}{\sqrt{\sigma_{ii.l}}\sqrt{\sigma_{jj.l}}} = \frac{\sigma_{ij} - \frac{\sigma_{il}\sigma_{jl}}{\sigma_{ll}}}{\sqrt{\left(\sigma_{ii} - \frac{\sigma_{il}^2}{\sigma_{ll}}\right)\left(\sigma_{jj} - \frac{\sigma_{jl}^2}{\sigma_{ll}}\right)}} \\ &= \frac{\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} - \frac{\sigma_{il}\sigma_{jl}}{\sigma_{ll}\sqrt{\sigma_{ii}\sigma_{jj}}}}{\sqrt{\left(1 - \frac{\sigma_{il}^2}{\sigma_{ii}\sigma_{ll}}\right)\left(1 - \frac{\sigma_{jl}^2}{\sigma_{jj}\sigma_{ll}}\right)}} = \frac{\rho_{ij} - \rho_{il}\rho_{jl}}{\sqrt{(1-\rho_{il}^2)(1-\rho_{jl}^2)}}. \end{aligned}$$

Simultaneous confidence sets

When we have a scalar parameter, such as the mean μ of X, we can construct a confidence interval for it using a sample of observations:
$\bar{X} \pm \frac{s}{\sqrt{n}} t_{n-1}(1-\alpha/2)$. What about the vector β of regression coefficients? We know that if $Y = X\beta + \epsilon$, where $\epsilon \sim N_n(0, \sigma^2 I_n)$, then $(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta) \sim \sigma^2\chi_r^2$ independent of RSS $= Y'(I-P)Y \sim \sigma^2\chi^2_{n-r}$, so that

$$\frac{(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)/r}{Y'(I-P)Y/(n-r)} \sim F_{r,n-r},$$

and hence

$$P\left((\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta) \le \frac{r}{n-r}\ \ Y'(I-P)Y\ \ F_{r,n-r}(1-\alpha)\right) = 1-\alpha.$$

Therefore,

$$C = \left\{\beta : (\beta-\hat{\beta})'X'X(\beta-\hat{\beta}) \le \frac{r}{n-r}\ \ \text{RSS}\ \ F_{r,n-r}(1-\alpha)\right\}$$

is a $100(1-\alpha)\%$ confidence set for β. This is an ellipsoid, and if p is not small (1 or 2), a set which is difficult to appreciate.

Suppose we are only interested in $a'\beta$ for some fixed a. Then
$a'\hat{\beta} \pm t_{n-r}(1-\alpha/2)\sqrt{\text{RSS}\ /(n-r)}\sqrt{a'(X'X)^{-}a}$ is a $100(1-\alpha)\%$ confidence interval for $a'\beta$. Let us see if we can extend this when we are interested in deriving a simutaneous confidence set of coefficient $1-\alpha$ for $a_1'\beta, a_2'\beta, \ldots, a_k'\beta$.