Joint Confidence Regions for Rankings based on Correlated Estimates

Matala, Shaine Rosewel

University of the Philippines

November 24, 2025

Background of the Study I

Klein et al. (2020) illustrates the problem of ranking the US states in terms of travel time to work. In this context, θ_k is the true mean travel time to work for state k, and we wish to find a joint confidence statement about $r_1, r_2, \ldots, r_K$.

Rank r_k

A mathematical definition of r_k is as follows:

$$r_k = \sum_{j=1}^{K} I(\theta_j \leq \theta_k) = 1 + \sum_{j:j \neq k} I(\theta_j \leq \theta_k), \qquad \text{for } k = 1, 2, \ldots, K. \tag{1}$$

Background of the Study II

Estimated rank

Let $\hat{r}_k$ denote the estimated rank of the kth observation.

$$\hat{r}_k = 1 + \sum_{j:j \neq k} I(\hat{\theta}_j \leq \hat{\theta}_k), \qquad \text{for } k = 1, 2, \ldots, K. \tag{2}$$

Motivation

Objective

This research aims to do the following:

- Develop a procedure to construct joint confidence intervals for the ranks and the ranked parameters when the estimates to be ranked may be correlated.
- Evaluate the performance of the proposed approaches under various parameter settings.
- Apply the proposed approaches to a real-life example.

Definitions and Assumptions

- Define $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K)'$ and assume that $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)'$ is unknown and $\boldsymbol{\Sigma}$ is a known $K \times K$ positive definite matrix. The diagonal elements of $\boldsymbol{\Sigma}$ are $\sigma_1^2, \ldots, \sigma_K^2$.
- Note that in the literature on inferences on the ranks, it is customary to assume that the variances are known.

Procedure I

1. Derive simultaneous confidence intervals for $\theta_1, \theta_2, \ldots, \theta_K$ of the form

$$\mathfrak{R}_1 = [\hat{\theta}_1 \pm t \times \sigma_1] \times [\hat{\theta}_2 \pm t \times \sigma_2] \times \cdots \times [\hat{\theta}_K \pm t \times \sigma_K]. \tag{3}$$

2. Once the confidence intervals in (3) have been obtained, we can then use the result of Klein et al. (2020) in (**??**) to get the lower and upper bounds on the ranks $r_k, k = 1, 2, \ldots, K$. That is, we also get a joint confidence region for $r_1, r_2, ..., r_K$.

Proposed methodology to compute the joint confidence region for the unordered parameters: Algorithm 1

Let the data be represented by $\hat{\boldsymbol{\theta}} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_K\right)'$ and suppose that $\boldsymbol{\Sigma}$ is known

1: **for** $b = 1, 2, \ldots, B$ **do**

2: Generate $\hat{\boldsymbol{\theta}}_b^* \sim N_K\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right)$ and write $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*\right)'$

3: Compute

$$t_b^* = \max_{1 \leq k \leq K} \left| \frac{\hat{\theta}_{bk}^* - \hat{\theta}_k}{\sigma_k} \right|$$

4: **end for**

5: Compute the $(1-\alpha)$-sample quantile of $t_1^*, t_2^*, \ldots, t_B^*$, call this $\hat{t}$.

6: The joint confidence region for $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)'$ is given by

$$\mathfrak{R}_1 = \left[\hat{\theta}_1 \pm \hat{t} \times \sigma_1\right] \times \left[\hat{\theta}_2 \pm \hat{t} \times \sigma_2\right] \times \cdots \times \left[\hat{\theta}_K \pm \hat{t} \times \sigma_K\right].$$

Algorithm 1: Quantile Calculation

We want the joint confidence region in (3) to satisfy the following probability condition:

$$P\left(\hat{\theta}_k - t \cdot \sigma_k \leq \theta_k \leq \hat{\theta}_k + t \cdot \sigma_k,\ \forall\, k = 1, 2, \ldots, K\right) = 1 - \alpha. \quad (4)$$

Equivalently, we require

$$P\left(\max_{k=1,2,\ldots,K} \left|\frac{\hat{\theta}_k - \theta_k}{\sigma_k}\right| \leq t\right) = 1 - \alpha. \quad (5)$$

Proposed methodology to compute a joint confidence region for the ordered parameters

1: **for** $b = 1, 2, \ldots, B$ **do**

2: Generate $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*\right)' \sim N_K\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \ldots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values

3: Compute $\hat{\sigma}_{b(k)}^*$ using:

- asymptotic variance definition
- second-level bootstrap

4: Compute

$$t_b^* = \max_{1 \le k \le K} \left| \frac{\hat{\theta}_{b(k)}^* - \hat{\theta}_k^*}{\hat{\sigma}_{b(k)}^*} \right|$$

5: **end for**

6: Compute the $(1-\alpha)$-sample quantile of $t_1^*, \ldots, t_B^*$, call this $\hat{t}$.

7: The joint confidence region of $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$ is given by

$$\mathfrak{R}_2 = \left[\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)}\right] \times \left[\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)}\right] \times \cdots \times \left[\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)}\right]$$

Algorithm 2

$$\hat{\sigma}^*_{b(k)} = \sqrt{\left[\text{kth ordered value among } \left\{\hat{\theta}^{*2}_{b1} + \sigma^2_1, \ldots, \hat{\theta}^{*2}_{bK} + \sigma^2_K\right\}\right] - \hat{\theta}^{*2}_{(k)}}$$

Algorithm 3

for $c = 1, 2, \ldots, C$ **do**

  Generate $\hat{\boldsymbol{\theta}}^{**}_{bc} = \left(\hat{\theta}^{**}_{bc1}, \hat{\theta}^{**}_{bc2}, \ldots, \hat{\theta}^{**}_{bcK}\right) \sim N_K\left(\hat{\boldsymbol{\theta}}^{*}_{b}, \boldsymbol{\Sigma}\right)$ and let $\hat{\theta}^{**}_{bc(1)}, \hat{\theta}^{**}_{bc(2)}, \ldots, \hat{\theta}^{**}_{bc(K)}$ be the corresponding ordered values of $\hat{\theta}^{**}_{bc1}, \hat{\theta}^{**}_{bc2}, \ldots, \hat{\theta}^{**}_{bcK}$

  Compute $\hat{\sigma}^{*}_{b(k)} = \dfrac{\sum_{c=1}^{C}\left(\hat{\theta}^{**}_{bc(k)} - \bar{\hat{\theta}}^{**}_{b\cdot(k)}\right)^2}{C-1}, \quad \bar{\hat{\theta}}^{**}_{b\cdot(k)} = \dfrac{1}{C}\displaystyle\sum_{c=1}^{C} \hat{\theta}^{**}_{bc(k)}$

end for

Evaluation Algorithm

For given values of $\mathbf{\Sigma}$ and $\theta_1, \theta_2, \ldots, \theta_K$ (with corresponding $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$ for rank-based methods)

1: **for** replications $= 1, 2, \ldots, 5000$ **do**
2: Generate $\hat{\boldsymbol{\theta}} \sim N_K(\boldsymbol{\theta}, \mathbf{\Sigma})$
3: Compute the confidence region $\mathfrak{R}_1$ for the unordered parameters using Algorithm **??** and the confidence region for the ordered parameters $\mathfrak{R}_2$ using Algorithms **??** and **??**.
4: For the unordered parameters, check if $(\theta_1, \theta_2, \ldots, \theta_K) \in \mathfrak{R}_1$ and compute T_1, T_2, and T_3. For the ordered parameters, check if $\left(\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}\right) \in \mathfrak{R}_2$
5: **end for**
6: Compute the proportion of times that the condition in line 4 is satisfied and the average of T_1, T_2, and T_3.

Covariance Matrix $\mathbf{\Sigma}$

- The covariance matrix $\mathbf{\Sigma}$ need not be a diagonal matrix.
- We assume that $V(\hat{\boldsymbol{\theta}}) = \mathbf{\Sigma}$ is known and express $\mathbf{\Sigma}$ as in (6), where $\mathbf{R}$ is the population correlation matrix.

$$\mathbf{\Sigma} = \mathbf{\Delta}^{1/2}\mathbf{R}\mathbf{\Delta}^{1/2}; \quad \mathbf{\Delta} = \text{diag}\left\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2\right\}. \tag{6}$$

- The diagonal elements of $\mathbf{\Sigma}$, which are $\sigma_k^2 = V(\hat{\theta}_k)$ for $k = 1, 2, \ldots, K$, are treated as known quantities in practice.

Correlation Structures: Equicorrelation

- We assume certain correlation structures among the $\hat{\theta}$s.

Correlation Structures: Equicorrelation

- We assume certain correlation structures among the $\hat{\theta}$s.
- Equicorrelation is considered for simplicity.

Correlation Structures: Equicorrelation

- We assume certain correlation structures among the $\hat{\theta}$s.
- Equicorrelation is considered for simplicity.
- This assumes that the k variables are equally correlated, i.e., that $\rho_{jk} = \rho$ where $\rho \in [-1, 1]$ for $j \neq k \in \{1, \dots, K\}$.

$$\mathbf{R}_{\text{eq}} = (1-\rho)\,\mathbf{I}_K + \rho \mathbf{1}_K \mathbf{1}_K' = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}_{K \times K} \tag{7}$$

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys
- In a block correlation matrix $\mathbf{R}_{block}$ with G blocks, each diagonal block represents an equicorrelation structure within group g, denoted by

$$\mathbf{R}_{\mathrm{eq,g}} = (1-\rho_g)\,\mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g}\mathbf{1}'_{n_g}$$

where ρ_g is the within-block correlation and n_g is the number of variables in block g such that $\sum_{g=1}^{G} n_g = K$.

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys
- In a block correlation matrix $\mathbf{R}_{block}$ with G blocks, each diagonal block represents an equicorrelation structure within group g, denoted by

$$\mathbf{R}_{\text{eq,g}} = (1 - \rho_g)\,\mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g} \mathbf{1}'_{n_g}$$

  where ρ_g is the within-block correlation and n_g is the number of variables in block g such that $\sum_{g=1}^{G} n_g = K$.
- The off-diagonal blocks capture between-block correlations, represented by $\mathbf{C}_{g'g} = \mathbf{C}_{gg'} = \rho_{gg'} \mathbf{1}_{n_g} \mathbf{1}'_{n_g}$ where $g \neq g' \in \{1, \ldots, G\}$

Correlation Structures: Block correlation

- Useful in the context of pre-election surveys
- In a block correlation matrix $\mathbf{R}_{block}$ with G blocks, each diagonal block represents an equicorrelation structure within group g, denoted by

$$\mathbf{R}_{\mathrm{eq,g}} = (1 - \rho_g)\, \mathbf{I}_{n_g} + \rho_g \mathbf{1}_{n_g} \mathbf{1}_{n_g}'$$

  where ρ_g is the within-block correlation and n_g is the number of variables in block g such that $\sum_{g=1}^{G} n_g = K$.
- The off-diagonal blocks capture between-block correlations, represented by $\mathbf{C}_{g'g} = \mathbf{C}_{gg'} = \rho_{gg'} \mathbf{1}_{n_g} \mathbf{1}_{n_g}'$ where $g \neq g' \in \{1, \ldots, G\}$
- The full block correlation matrix can be expressed as in (8).

$$\mathbf{R}_{\mathrm{block}} = \begin{bmatrix} \mathbf{R}_{eq,1} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1G} \\ \mathbf{C}_{21} & \mathbf{R}_{eq,2} & \cdots & \mathbf{C}_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{G1} & \mathbf{C}_{G2} & \cdots & \mathbf{R}_{eq,G} \end{bmatrix}_{K \times K} \tag{8}$$

Correlation Structures: Distance-based correlation

- Useful in the mean travel time from the study of Klein et al. (2020).

Correlation Structures: Distance-based correlation

- Useful in the mean travel time from the study of Klein et al. (2020).
- Spatial dependence can be modeled using a stationary Matérn correlation function, which for two locations $\mathbf{s}_i$ and $\mathbf{s}_j$ is expressed as

$$\rho_{\text{matern}} = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa \parallel \mathbf{s}_i - \mathbf{s}_j \parallel)^{\nu} K_{\nu}(\kappa \parallel \mathbf{s}_i - \mathbf{s}_j \parallel)$$

  where $\parallel \cdot \parallel$ denotes the Euclidean distance and K_ν is the second kind of the modified Bessel function. It has a scale parameter $\kappa > 0$ and a smoothness parameter $\nu > 0$. ρ_{matern} reduces to the exponential correlation when $\nu = 0.5$ and to Gaussian correlation function when $\nu = \infty$.

Code blocks

Alert block

$$E = mc^2$$

Examples

Example blocks are automatically green in color

Blue block

- happens with level 2, 3 headings
- this is only true for ‘Madrid’ theme in R Markdown!!

This works, incremental bullets

- Bullet 1

This works, incremental bullets

- Bullet 1
- Bullet 2

This nests, but does not increment

- Bullet 1
- Bullet 2
    - subbullet 1
    - subbullet 2

This increments

- Bullet 1

This increments

- Bullet 1
- Bullet 2

This increments

- Bullet 1
- Bullet 2
  - subbullet 1

This increments

- Bullet 1
- Bullet 2
    - subbullet 1
    - subbullet 2

This increments too

- Bullet 1

This increments too

- Bullet 1
- Bullet 2

This increments too

- Bullet 1
- Bullet 2
    - subbullet 1

This increments too

- Bullet 1
- Bullet 2
    - subbullet 1
    - subbullet 2

Algo Font Size Adjustments sample

1: **for** $b = 1, 2, \ldots, B$ **do**

2: Generate $\hat{\boldsymbol{\theta}}_b^* = \left(\hat{\theta}_{b1}^*, \hat{\theta}_{b2}^*, \ldots, \hat{\theta}_{bK}^*\right)' \sim N_K\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right)$ and let $\hat{\theta}_{b(1)}^*, \hat{\theta}_{b(2)}^*, \ldots, \hat{\theta}_{b(K)}^*$ be the corresponding ordered values

3: Compute $\hat{\sigma}_{b(k)}^* = \hat{\sigma}_{b(k)}^* = \sqrt{\left[\text{kth ordered value among } \left\{\hat{\theta}_{b1}^{*2} + \sigma_1^2, \ldots, \hat{\theta}_{bK}^{*2} + \sigma_K^2\right\}\right] - \hat{\theta}_{(k)}^{*2}}$

4: Compute

$$t_b^* = \max_{1 \leq k \leq K} \left| \frac{\hat{\theta}_{b(k)}^* - \hat{\theta}_k^*}{\hat{\sigma}_{b(k)}^*} \right|$$

5: **end for**

6: Compute the $(1-\alpha)$-sample quantile of $t_1^*, \ldots, t_B^*$, call this $\hat{t}$.

7: The joint confidence region of $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(K)}$ is given by

$$\mathfrak{R}_2 = \left[\hat{\theta}_{(1)} \pm \hat{t} \times \hat{\sigma}_{(1)}\right] \times \left[\hat{\theta}_{(2)} \pm \hat{t} \times \hat{\sigma}_{(2)}\right] \times \cdots \times \left[\hat{\theta}_{(K)} \pm \hat{t} \times \hat{\sigma}_{(K)}\right]$$

where

$$\hat{\sigma}_{(k)} = \sqrt{\text{kth ordered value among } \{\hat{\theta}_1^2 + \sigma_1^2, \ \ldots, \ \hat{\theta}_K^2 + \sigma_K^2\} - \hat{\theta}_{(k)}^2}$$

Text

Today is going to be a great day

Today is. . .

A new beginning.

Your Reference

- slideshare
- themes
- incremental bullets

Klein, M., Wright, T., & Wieczorek, J. (2020). A joint confidence region for an overall ranking of populations. *Journal of the Royal Statistical Society*, 589–606.