Mann–Whitney–Wilcoxon rank sum test to compare three samples in Chapter 5. Within the framework of FRM, we can readily generalize this classic test to a general setting involving any number of groups without running into the notational complications found with multivariate U-statistics in Chapter 4, as illustrated in the next example.

Example 6 (***K-sample Mann–Whitney–Wilcoxon rank sum test***). Let y_{ki} denote some continuous response of interest from K i.i.d. samples ($1 \leq i \leq n_k$, $1 \leq k \leq K$). Let $L = \binom{K}{2}$ and $q = 1$. Consider the following FRM:

$$f(y_{ki}, y_{hj}) = I_{\{y_{ki} \leq y_{hj}\}}, \quad h(\boldsymbol{\theta}) = \theta_{kh}$$
$$\boldsymbol{\theta} = \left(\theta_{12}, \ldots, \theta_{1K}, \theta_{23}, \ldots, \theta_{2K}, \ldots, \theta_{(K-1)K}\right)^{\top}$$
$$1 \leq i \leq n_k, \quad 1 \leq j \leq n_h, \quad 1 \leq k < h \leq K$$

It follows that

$$E\left[f(y_{ki}, y_{hj})\right] = h(\boldsymbol{\theta}) = \theta_{kh}, \quad 1 \leq i \leq n_k,\ 1 \leq j \leq n_h,\ 1 \leq k < h \leq K$$

If y_{kj} have the same distribution across the K samples, then $\theta_{kh} = \frac{1}{2}$ for all $1 \leq k < h \leq K$ and vice versa. Thus, we can test the null hypothesis, $H_0 : \theta_{kh} = \frac{1}{2}$ for all $1 \leq k < h \leq K$, to determine whether there is any location shift in CDF across the K independent samples.

In Chapter 5, we developed a U-statistics based approach for simultaneously comparing the mean and variance between two groups. By combining the classic mean-based ANOVA with the generalized ANOVA for variance discussed earlier, we can readily generalize this approach by developing an FRM to jointly compare the mean and variance of a response across multiple samples, as illustrated below.

Example 7 (***Model for K-sample mean and variance***). Consider K i.i.d. samples and let y_{ki} denote the response of interest from the ith subject within the kth group with mean μ_k and variance σ_k^2 ($1 \leq i \leq n_k$, $1 \leq k \leq K$). Let

$$\mathbf{f}(y_{ki}, y_{kj}) = \begin{pmatrix} y_{ki} \\ \frac{1}{2}(y_{ki} - y_{kj})^2 \end{pmatrix}, \quad \mathbf{h}(\boldsymbol{\theta}) = \begin{pmatrix} \mu_k \\ \sigma_k^2 \end{pmatrix} = \boldsymbol{\theta}_k \tag{6.15}$$
$$\boldsymbol{\theta}_k = \left(\mu_k, \sigma_k^2\right)^{\top}, \quad \boldsymbol{\theta} = \left(\boldsymbol{\theta}_1^{\top}, \ldots, \boldsymbol{\theta}_K^{\top}\right)^{\top}, \quad (i,j) \in C_2^{n_k}, \quad 1 \leq k \leq K$$

Then, $E\left[\mathbf{f}(y_{ki}, y_{kj})\right] = \mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{\theta}_k$ and the parameter vector $\boldsymbol{\theta}_k$ contains both the mean and variance of the kth sample. If $H_0 : \boldsymbol{\theta}_k = \boldsymbol{\theta}$ for all