Experimental Designs in agRank Package

Shuxiao Chen

1 General Setup

Say we have n farmers as observers and m varieties to measure. Since each farmer can measure 3 varieties, $\frac{m}{3}$ farmers will be enough to compare all the varieties once. Thus, we first allocate m varieties to $\frac{m}{3}$ farmers so that each variety appears exactly once, and we call this allocation one *replica*. One replica is demostrated by a $\frac{m}{3} \times 3$ matrix, where the entries in each row are the labels of items ranked by one farmer. We then generate $ceiling(\frac{n}{m/3})$ replicas and aggregate all the replicas. We choose the first n rows of the aggregated replicas and the resulting $n \times m$ matrix represents the experimental design.

We use random allocation, alpha design, along with 15 clustering algorithms to cluster the varieties so that each farmer can choose accordingly. We first introduce two metrics that are used in the clustering algorithms. The *competence measure* of variety i in a cluster is calculated by summing up the correlation coefficients (the entry of K, the additive relationship matrix) between i and all the other varieties in this cluster. The *similarity measure* between variety i and a group of varieties is calculated by taking the mean of the correlation coefficients between i and the other group of varieties.

* The function names here are different with the function names in the paper. In the paper, only KM4, KM6, GD3, GD6 and GD9 are included, and they are namely as KM1, KM2, GD1, GD2 and GD3 respectively.

2 Random allocation

We assign the label 1 to m randomly to the $\frac{m}{3} \times 3$ matrix.

3 Alpha design

Here we use the R package Agricolae to generate (0, 1) alpha designs with treatment number = m, block size = 3 and replications = 3, so that each farmer represents a block and each replication is one replica.

4 Three clusters

Cluster the varieties into 3 clusters, and each farmer chooses one variety from each cluster.

4.1 KM1

Cluster m varieties into 3 clusters by repetitive k-means clustering, so that varieties within each cluster are most similar to each other. Then each farmer chooses one from each cluster. The algorithm to generate one replica is demostrated as follows:

Algorithm 1 Cluster varieties into 3 clusters by repetitive k-means clustering

procedure KM1(M) ▷ M is the marker matrix
    $\boldsymbol{v}$ = varieties not assigned to farmers
    while $\boldsymbol{v}$ not empty **do**
        Cluster $\boldsymbol{v}$ into 3 clusters by k-means clustering according to M
        x = the smallest size of the 3 clusters
        Randomly choose x varieties from each cluster and let them be the "chosen clusters"
        Farmers who haven't get varieties each chooses one variety from each "chosen cluster"
        Remove the varieties already chosen from $\boldsymbol{v}$

4.2 KM2

Cluster m varieties into 3 clusters by repetitive k-means clustering, assisted by the additive relationship matrix K. In each loop, instead of choosing x varieties from each cluster, we choose top x varieties in each cluster (with the largest competence measure) and let the resulting clusters be the "chosen clusters". The algorithm to generate one replica is demostrated as follows:

Algorithm 2 Cluster varieties into 3 clusters by repetitive k-means clustering, assisted by K

procedure KM2(M, K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    while $\boldsymbol{v}$ not empty **do**
        Cluster $\boldsymbol{v}$ into 3 clusters by k-means clustering according to M
        x = the smallest size of the 3 clusters
        Choose top x varieties with the largest competence measure and let them be the "chosen clusters"
        Each farmer chooses one variety from each "chosen cluster"
        Remove the varieties already chosen from $\boldsymbol{v}$

4.3 KM3

First, cluster m varieties into 3 clusters by k-means clustering and choose top x varieties with the largest competence measure (initialization step). We then let **each cluster** take turns to choose one variety with the largest similarity measure until they all reach the size of $\frac{m}{3}$. The algorithm to generate one replica is demostrated as follows:

Algorithm 3 Cluster varieties into 3 clusters using k-means to initialize and assisted by K (cluster-oriented)

procedure KM3(M, K)
    Cluster m varieties into 3 clusters by k-means clustering according to M
    x = the smallest size of the 3 clusters
    Choose top x varieties with the largest competence measure and let them be the "chosen clusters"
    while Sizes of all chosen clusters haven't reached $\frac{m}{3}$ **do**
        Each chosen cluster chooses one variety with the largest similarity measure (unless that cluster has reached the size of $\frac{m}{3}$)

4.4 KM4

First, cluster m varieties into 3 clusters by k-means clustering and choose top x varieties with the largest competence measure (initialization step). We then let **each remaining variety** take turns to choose to enter its favorate cluster (with the largest similarity measure) until all the clusters reach the size of $\frac{m}{3}$. The algorithm to generate one replica is demostrated as follows:

Algorithm 4 Cluster varieties into 3 clusters using k-means to initialize and assisted by K (variety-oriented)

procedure KM4(M, K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    Cluster $\boldsymbol{v}$ into 3 clusters by k-means clustering according to M
    x = the smallest size of the 3 clusters
    Choose top x varieties with the largest competence measure and let them be the "chosen clusters"
    Delete varieties in "chosen clusters" from $\boldsymbol{v}$
    for v in $\boldsymbol{v}$ **do**
        v chooses its favorate "chosen cluster" (with the largest similarity measure) to enter (unless that cluster has reached the size of $\frac{m}{3}$)

4.5 GD1

A greedy algorithm to cluster the varieties into 3 clusters according to the additive relationship matrix K.

Algorithm 5 Cluster varieties into 3 clusters using a greedy algorithm

procedure GD1(K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    for $i = 1, 2, 3$ **do**
        v = a randomly chosen variety from $\boldsymbol{v}$
        Let v enter cluster i and delete v from $\boldsymbol{v}$
        Compute similarity measure between v and all the varieties in $\boldsymbol{v}$
        Choose $(m/3-1)$ varieties from $\boldsymbol{v}$ with largest similarity measure and let them enter cluster i
        Delete those $(m/3-1)$ varieties from $\boldsymbol{v}$

4.6 GD2

A greedy algorithm to cluster the varieties into 3 clusters according to the additive relationship matrix K.

Algorithm 6 Cluster varieties into 3 clusters using a greedy algorithm (cluster-oriented)

procedure GD2(K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    Randomly choose a variety from $\boldsymbol{v}$, let it enter cluster 1 and delete it from $\boldsymbol{v}$
    for $i = 2, 3$ **do** ▷ Initialize 3 clusters
        for v in $\boldsymbol{v}$ **do**
            Compute similarity measure between v and all the varieties in cluster 1 to cluster $(i-1)$
        Choose the variety with the smallest similarity measure in $\boldsymbol{v}$, let it enter cluster i and delete it from $\boldsymbol{v}$
    while The size of clusters hasn't reached $\frac{m}{3}$ **do**
        Each cluster takes turns to choose the variety with the largest similarity measure (unless that cluster has reached the size of $\frac{m}{3}$)
        Delete the chosen variety from $\boldsymbol{v}$

4.7 GD3

A greedy algorithm to cluster the varieties into 3 clusters according to the additive relationship matrix K.

Algorithm 7 Cluster varieties into 3 clusters using a greedy algorithm (variety-oriented)

procedure GD3(K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    Randomly choose a variety from $\boldsymbol{v}$, let it enter cluster 1 and delete it from $\boldsymbol{v}$
    for $i = 2, 3$ **do** ▷ Initialize 3 clusters
        for v in $\boldsymbol{v}$ **do**
            Compute the similarity measure between v and all the varieties in cluster 1 to cluster $(i-1)$
        Choose the variety with the smallest similarity measure in $\boldsymbol{v}$, let it enter cluster i and delete it from $\boldsymbol{v}$
    for v in $\boldsymbol{v}$ **do**
        v chooses its favorate cluster (with the largest similarity measure) to enter (unless that cluster has reached the size of $\frac{m}{3}$)

5 $\frac{m}{3}$ clusters, similar within cluster

Cluster varieties into $m/3$ clusters, where each cluster is 3 most similar varieties. And each farmer represents one cluster.

5.1 KM5

First, cluster m varieties into $m/3$ clusters by k-means clustering. For clusters with more than 3 varieties, choose the most competent 3 varieties and let them stay in the clusters; for clusters with exactly 3 varieties, that's fine; and we want to assign the surplus varieties into clusters with less than 3 varieties. We then let **clusters with less than 3 varieties** take turns to choose one variety with the largest similarity measure, until they all reach the size of 3. The algorithm to generate one replica is demostrated as follows:

Algorithm 8 Cluster varieties into $m/3$ clusters using k-means to initialize and assisted by K (cluster-oriented)

1: **procedure** KM5(M, K)
2: Cluster m varieties into $m/3$ clusters by k-means clustering according to M
3: **for** $i = 1, \ldots, m/3$ **do**
4: **if** size of cluster $i \geq 3$ **then**
5: Choose 3 varieties with the largest competence measure to stay in cluster i
6: **while** Size of all clusters haven't reached 3 **do**
7: Each cluster chooses the variety with the largest similarity measure (unless that cluster has reached the size of 3)

5.2 KM6

First, cluster m varieties into $m/3$ clusters by k-means clustering. For clusters with more than 3 varieties, choose the most competent 3 varieties and let them stay in the clusters; for clusters with exactly 3 varieties, that's fine; and we want to assign the surplus varieties into clusters with less than 3 varieties. We then let **varieties not assigned yet** take turns to choose the cluster with the largest similarity measure to enter, until all the clusters reach the size of 3. The algorithm to generate one replica is demostrated as follows:

Algorithm 9 Cluster varieties into 3 clusters using k-means to initialize and assisted by K (variety-oriented)

procedure KM6(M, K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    Cluster $\boldsymbol{v}$ into $m/3$ clusters by k-means clustering according to M
    for $i = 1, \ldots, m/3$ **do**
        if size of cluster $i \geq 3$ **then**
            Choose 3 varieties with the largest competence measure to stay in cluster i
    Delete varieties already in clusters from $\boldsymbol{v}$
    for v in $\boldsymbol{v}$ **do**
        v chooses the cluster (with the largest similarity measure) to enter (unless that cluster has reached the size of 3)

5.3 GD4

A greedy algorithm to cluster the varieties into $m/3$ clusters according to the additive relationship matrix K.

Algorithm 10 Cluster varieties into $m/3$ clusters using a greedy algorithm

procedure GD4(A)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    for $i = 1, \ldots, m/3$ **do**
        v = a randomly chosen variety from $\boldsymbol{v}$
        Let v enter cluster i and delete v from $\boldsymbol{v}$
        Compute the similarity measure between v and all the varieties in $\boldsymbol{v}$
        Choose 2 varieties from $\boldsymbol{v}$ with largest similarity measure and let them enter cluster i
        Delete those 2 varieties from $\boldsymbol{v}$

5.4 GD5

A greedy algorithm to cluster the varieties into $m/3$ clusters according to the additive relationship matrix K.

Algorithm 11 Cluster varieties into $m/3$ clusters using a greedy algorithm (cluster-oriented)

procedure GD5(K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    Randomly choose a variety from $\boldsymbol{v}$, let it enter cluster 1 and delete it from $\boldsymbol{v}$
    for $i = 2, \ldots, m/3$ **do** ▷ Initialize $m/3$ clusters
        for v in $\boldsymbol{v}$ **do**
            Compute the similarity measure between v and all the varieties in cluster 1 to cluster $(i-1)$
        Choose the variety with the smallest similarity measure in $\boldsymbol{v}$, let it enter cluster i and delete it from $\boldsymbol{v}$
    while Size of all clusters haven't reached 3 **do**
        Each cluster takes turns to choose the variety with the largest similarity measure in $\boldsymbol{v}$ (unless that cluster has reached the size of 3)
        Delete the chosen variety from $\boldsymbol{v}$

5.5 GD6

A greedy algorithm to cluster the varieties into $m/3$ clusters according to the additive relationship matrix K.

Algorithm 12 Cluster varieties into 3 clusters using a greedy algorithm (variety-oriented)

procedure GD6(K)
    $\boldsymbol{v}$ = varieties not assigned to farmers
    Randomly choose a variety from $\boldsymbol{v}$, let it enter cluster 1 and delete it from $\boldsymbol{v}$
    for $i = 2, \ldots, m/3$ **do** ▷ Initialize $m/3$ clusters
        for v in $\boldsymbol{v}$ **do**
            Compute the similarity measure between v and all the varieties in cluster 1 to cluster $(i-1)$
        Choose the variety with the smallest similarity measure in $\boldsymbol{v}$, let it enter cluster i and delete it from $\boldsymbol{v}$
    for v in $\boldsymbol{v}$ **do**
        v chooses the cluster with the largest similarity measure to enter (unless that cluster has reached the size of 3)

6 m clusters, dissimilar within cluster

Cluster varieties into $m/3$ clusters, where each cluster is 3 most **dissimilar** varieties. And each farmer represents one cluster.

6.1 GD7

A greedy algorithm to cluster the varieties into $m/3$ clusters according to the additive relationship matrix K.

Algorithm 13 Cluster varieties into $m/3$ clusters using a greedy algorithm

1: **procedure** GD7(K)
2: $\boldsymbol{v}$ = varieties not assigned to farmers
3: **for** $i = 1, \ldots, m/3$ **do**
4: v = a randomly chosen variety from $\boldsymbol{v}$
5: Let v enter cluster i and delete v from $\boldsymbol{v}$
6: Compute the similarity measure between v and all the varieties in $\boldsymbol{v}$
7: Choose 2 varieties from $\boldsymbol{v}$ with the **smallest** similarity measure and let them enter cluster i
8: Delete those 2 varieties from $\boldsymbol{v}$

6.2 GD8

A greedy algorithm to cluster the varieties into $m/3$ clusters according to the additive relationship matrix K.

Algorithm 14 Cluster varieties into $m/3$ clusters using a greedy algorithm (cluster-oriented)

1: **procedure** GD8(K)
2: $\boldsymbol{v}$ = varieties not assigned to farmers
3: Randomly choose a variety from $\boldsymbol{v}$, let it enter cluster 1 and delete it from $\boldsymbol{v}$
4: **for** $i = 2, \ldots, m/3$ **do** ▷ Initialize $nm/3$ clusters
5: **for** v in $\boldsymbol{v}$ **do**
6: Compute the similarity measure between v and all the varieties in cluster 1 to cluster $(i-1)$
7: Choose the the variety with the largest similarity measure in $\boldsymbol{v}$, let it enter cluster i and delete it from $\boldsymbol{v}$
8: **while** Size of all clusters haven't reached 3 **do**
9: Each cluster takes turns to the variety with smallest similarity measure in $\boldsymbol{v}$ (unless that cluster has reached the size of 3)
10: Delete the chosen variety from $\boldsymbol{v}$

6.3 GD9

A greedy algorithm to cluster the varieties into $m/3$ clusters according to the additive relationship matrix K.

Algorithm 15 Cluster varieties into 3 clusters using a greedy algorithm (variety-oriented)

1: **procedure** GD9(A)
2: $\boldsymbol{v}$ = varieties not assigned to farmers
3: Randomly choose a variety from $\boldsymbol{v}$, let it enter cluster 1 and delete it from $\boldsymbol{v}$
4: **for** $i = 2, \ldots, m/3$ **do** ▷ Initialize $m/3$ clusters
5: **for** v in $\boldsymbol{v}$ **do**
6: Compute the similarity measure between v and all the varieties in cluster 1 to cluster $(i-1)$
7: Choose the variety with the largest similarity measure in $\boldsymbol{v}$, let it enter cluster i and delete it from $\boldsymbol{v}$
8: **for** v in $\boldsymbol{v}$ **do**
9: v chooses its the cluster with smallest similarity measure to enter (unless that cluster has reached the size of 3)