Ranking and Selection Procedure for Gini Index

Saeed Baghban kohne rooz¹, Mohammad Moradi² and Hossein Jabbari Khamnei³

¹,²Department of Statistics, Faculty of Sciences, Razi University, Kermanshah, Iran
³Department of Statistics, Faculty of mathematical sciences, University of Tabriz, Tabriz, Iran

Keywords: Ranking and Selection, Gini Index, Probability of Correct Selection, Selection Procedures, Indifference and preference zone.

Abstract: Gini index, which is derived from the Lorenz curve of income inequality and shows income inequality in different populations, can be applied to ranking and selection populations. Many procedures are available for ordering and ranking income distributions where the ordering is not linear. However, the researchers often are not interested in ordering the populations but selecting the best (or worst) of available populations indicating a lower (or higher) level of disparities in incomes within the population. Madhuri S. Mulekar (2005) discussed the estimation of overlap of income distributions and selection in terms of Gini Measure of income inequality. In this paper, we simulate populations ranking and selection based on Gini index of income inequality for case that the variances are equal but known in income distributions and for case that the variances are unequal but known in income distributions.

1. Introduction

There are many different numerical measures, such as Gini coefficient (Soltow, 1971), coefficient of variation (Braun, 1988), Theil index (Theil, 1967), Atkinson ratio (Atkinson, 1970), and Nelson ratio (Nelson, 1984) among others, used to express the degree of inequality or variability in income among the members of a given population. However, Lorenz curve-based measures seem to be the most popular ones used in practice. Of those, Gini coefficient the most widely used is Lorenz curve-based measure of income inequality. Besides income or wealth inequality, Gini index is also used to study inequality in health (Illsey and Le Grand, 1987), inter-individual inequality in age at death (Anand, et al., 2001), and life tables (Shkolnikov, Andreev, and Begun, 2003). Income inequality measures are both simple as well as complex. When comparing different populations the measured differences in coefficients are often small. Although some methods of estimating variances of some inequality measures (Gastwirth, 1974; and Sandstrom, et al., 1988) have been already developed, very commonly these measures are used in applications, for comparative purposes and trend analysis, without any consideration to the sampling variation. In other words, comparisons are carried out based on simply point estimates of the coefficients. Gini index is no exception to this. For example, Braun (1988) ranked U.S. states in terms of several family income inequality measures and Moore (1996) ranked income distributions in 26 countries using geometric mean and Atkinson index related generalized welfare measures, without estimating standard errors. Karoly (1990) compared inequality in individual wage and salary income for years 1967 to 1986 using ten different inequality measures including Gini index and estimated variances of eight out of ten measures. Again, the variance of Gini coefficient was not estimated due to expenses involved in computation. Sometimes researchers are not interested in just ranking the populations by income inequalities, but they are more interested in selecting (or identifying) from a group of populations, the population with least (or most) inequality. Here we simulate procedure method for selecting a population based on Gini coefficient and obtain optimal sample size needed to make such a selection.
2. Gini index and calculation method

Suppose \( f(x) \) is the density of income \( X \), and \( F(x) = \int_0^x f(y) \, dy \) is the cumulative distribution function (CDF) of \( X \) for the individuals or families.

For mean income of the population \( \mu = \int_0^\infty x f(x) \, dx \), the Gini index is defined as

\[
G = 1 - 2 \int_0^1 \Phi(F(x)) \, dF(x) \tag{2.1}
\]

Here \( \Phi(x) \) is the share of the total income received by the population with income less than or equal to \( x \) and it is defined as

\[
\Phi(x) = \frac{1}{\mu} \int_0^x y f(y) \, dy \tag{2.2}
\]

it is known as Lorenz curve. Gini index is a function of Lorenz curve. Consider the empirical Lorenz curve characterized by a set of ordinates \( \{\Phi_l = 1, 2, \ldots, L\} \) corresponding to abscissas \( \{p_l = 1, 2, \ldots, L\} \). An income quantile \( \xi_p \) on a Lorenz curve corresponding to abscissas \( p \), \((0<p<1)\), is given by \( F(\xi_p) = p \). Then corresponding to a set of \( L \) abscissas \( p_1<p_2<\ldots<p_L \), there is a set of \( L \) population income quantiles \( \xi_{p_1} < \xi_{p_2} < \ldots < \xi_{p_L} \) and a set of \( L \) population Lorenz curve ordinates \( \Phi(\xi_{p_1}) < \Phi(\xi_{p_2}) < \ldots < \Phi(\xi_{p_L}) \), where

\[
\Phi(\xi_{p_l}) = p_l \frac{Y_l}{\mu} \tag{2.3}
\]

This result is obtained from equation 1 of Beach and Davidson (1983), where \( Y_l(l = 1, 2, \ldots, L) \) is the conditional mean of incomes less than or equal to \( \xi_{p_l} \), i.e.

\[
Y_l = E(X | X \leq \xi_{p_l}) \tag{2.4}
\]

Denote the conditional variance of \( x \) given quantile \( \xi_{p_l}(l = 1, 2, \ldots, L) \) by

\[
\lambda^2 = E[(X - Y_l)^2 | X \leq \xi_{p_l}] \tag{2.5}
\]

Let \( x_1, x_2, \ldots, x_n \) be a random sample of size \( n \) from the given population. Denote the ordered observations by \( x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} \). Then the \( p \)th income quantile denoted by \( \xi_p \) can be estimated as \( r \)th order statistic \( x_{(r)} \), where \( r = [np] \) gives the greatest integer less than or equal to \( np \). The sample estimates of Lorenz curve ordinates can be computed as (Beach and Davidson, 1983, Eq 3)

\[
\Phi(\xi_{p_l}) = p_l \frac{\hat{Y}_l}{\hat{\mu}} \tag{2.6}
\]

where \( \hat{\mu} \) is the sample mean and for \( r_l = [np_l] \) gives the greatest integer less than or equal to \( np \).

\[
\hat{Y}_l = \frac{1}{r_l} \sum_{j=1}^{r_l} x_{(j)} \tag{2.7}
\]

Then the Gini coefficient can be estimated from the sampled values as

\[
\hat{G} = 2 \sum_{l=1}^{n} \left[ p_l - \Phi(\xi_{p_l}) \right] \tag{2.8}
\]

Here is also described as mapping out of the line of equality.
2.1. Theorem: Under the conditions of strict monotonicity, and twice differentiability of the CDF, \( F(x) \), of income distribution, the sample estimator of Gini coefficient, \( \hat{G} \), is asymptotically normal, i.e. as \( n \to \infty \), \( \sqrt{n}(\hat{G} - G) \to N(0, \sigma_G^2) \), where

\[
\sigma_G^2 = \left( \frac{2}{L+1} \right)^2 \sum_{l=1}^{L} \sum_{l'=1}^{L} y_{ll'}
\]

where for \( l \leq l' = 1, 2, \ldots, L \)

\[
y_{ll'} = \frac{\omega_{ll'}}{\mu^2} + \left( \frac{p_l y_l}{\mu^2} \right) \left( \frac{p_{l'} y_{l'}}{\mu^2} \right) \sigma_G^2 = \left( \frac{p_l y_l}{\mu^2} \right) \omega_{l'1} + \left( \frac{p_{l'} y_{l'}}{\mu^2} \right) \omega_{l'1}
\]

And

\[
\omega_{ll'} = p_l [\lambda_l^2 + (1 - p_l)(\xi_l' - \gamma_l)(\xi_l' - \gamma_l') + (\xi_l - \gamma_l)(\gamma_l' - \gamma_l')]
\]

Note that \( [\omega_{ll'}] \) gives the covariance matrix of Lorenz curve ordinates (Beach and Davidson, 1983, and Gail and Gastwirth, 1978), \( \xi_{l+1} \) is the maximum income in the population, and is the variance of incomes in the population.

3. Selection procedure

Consider independent populations, \( \pi_1, \pi_2, \ldots, \pi_k \). Let \( f_i(x) \) be a density function of income in population \( i(i=1,2,\ldots,k) \). Suppose Gini coefficient is the true income inequality existing in population \( \pi_i(i=1,2,\ldots,k) \). Of course the true value of \( G_i \) in the population is unknown, and so is the correspondence between \( G_i \) and \( \pi_i(i=1,2,\ldots,k) \). Here the problem considered is that of selecting the population corresponding to the largest income inequality, i.e. the \( \max\{G_i,i=1,2,\ldots,k\} \). The ordered values of inequality measures are denoted by \( G_{[1]} \leq G_{[2]} \leq \cdots \leq G_{[k]} \). Selection of the population corresponding to \( G_{[k]} \) or any population with income inequality measure equal to is considered as the correct selection (CS). The problem of selecting the population with least inequality, i.e. the most integration of incomes can be treated similarly by selecting population corresponding to \( G_{[1]} = \min\{G_i,i=1,2,\ldots,k\} \). Indicate the estimated Gini coefficient for the population with coefficient \( G_{[i]} \) by \( \hat{G}_{[i]} \). Let us denote the parameter space, \( \Omega \), as

\[
\Omega = \{ G \in \Omega : G_{[i]} < G_i < 1 \quad \forall \ i = 1, 2, \ldots, k \}
\]

Partition the parameter space using predetermined constant \( d^*(0 < d^* < 1) \) into the preference zone (PZ) and its complement the indifference zone (IZ), such that

\[
\Omega_{PZ} = \{ G \in \Omega : G_{[k]} - G_{[i]} \geq d^* \quad \forall \ i = 1, 2, \ldots, k \}
\]

Although we would like to select a unique population associated with \( G_{[k]} \), with specified difference or distance \( d^* \) any population in IZ will be accepted as a correct decision. Therefore, if more than one population belongs to IZ, then we’ll use some randomization scheme to select a population from IZ as the best one. If a non-randomized scheme is used then it should be uncorrelated with the risk involved.
4. Estimation of n

4.1. In case of variance equality

Select the population corresponding to the \( \max(\bar{G}(1), \bar{G}(2), \ldots, \bar{G}(k)) \), i.e. \( \bar{G}_{[k]} \) as the best population. If there are tied populations then select one of the contenders for the first place by giving equal opportunity to each of the contenders.

For the predetermined risk we are interested in obtaining an optimal sample size needed to control this risk. Note that the population with largest inequality \( \{i.e., \bar{G}_{[k]}\} \) does not necessarily produce a sample with the largest estimate of Gini coefficient \( \{i.e., \bar{G}_{[k]}\} \). Using the complement of the risk, also referred to as the probability of correct selection, the above problem reduces to that of determining optimal sample size needed for

\[
P(CS) \geq p^* \quad \forall \: G \in \Omega
\]

where \( G \) is the Gini coefficient of the population. The relation between constant \( \tau \) and distance \( d^* \) is given by

\[
H(y) = \frac{\sqrt{n}d^*}{\sigma_G}.
\]

Here \( H(y) \) is the of the CDF standard normal variate, \( y = \frac{\sqrt{n}(\bar{G}_{[k]} - \bar{G}_{[i]})}{\sigma_G} \). The relation between constant \( \tau \) and distance \( d^* \) is given by

\[
\tau = \frac{\sqrt{n}d^*}{\sigma_G}
\]
The values of \( \tau \) are listed by Bechhofer (1954) and Gibbons, Olkin, and Sobel (1977, Table A.1). Here \( \tau \) can be interpreted as the standardized difference between Gini coefficients, expressed in the scale of \( \frac{\sigma_G^2}{n} \) instead of the original scale. Equating \( \tau \) with \( \frac{d^*}{\sqrt{n}/\sigma_G} \) and solving for \( n \), we get

\[
n = \left( \frac{\tau \sigma_G}{d^*} \right)^2, \tag{4.1.4}
\]

4.2. In case of unequal variances but known:

In this situation first calculate \( n_0 \), Where \( n_0 \) is given by the following equation

\[
n_0 = \left( \frac{\tau \sigma_0}{d^*} \right)^2 \tag{4.2.1}
\]

Because this method is based on the equality of variances, here consider \( \sigma_0 = 1 \). where

\[
n_i = n_0 \sigma_i^2 \quad \forall i = 1, 2, ..., k \tag{4.2.2}
\]

5. Simulation study:

**First case**: we consider five hypothetical normal populations with parameters \((\mu_1=0.85, \sigma), ~ (\mu_2=1.20, \sigma), ~ (\mu_3=1.60, \sigma), ~ (\mu_4=2.50, \sigma)\) as the density-income population. In this case the average of populations is different. Here a population with largest Gini index is defined as the best population. In this case, we consider different values for \( \sigma \) such as \((0.45, 0.52, 0.80, 0.75)\) and we consider different values for \( d^* \) such as \((0.01, 0.03, 0.05, 0.07, 0.10, 0.15, 0.20)\) and we compare the results obtained for any \( \sigma \) and \( d^* \). The result of this case is shown in the columns of tables below.

**Second case**: we consider five hypothetical normal populations with parameters \((\mu, \sigma_1=0.45), ~ (\mu, \sigma_2=0.52), ~ (\mu, \sigma_3=0.80), ~ (\mu, \sigma_4=0.75)\) as the density-income population. In this case the Standard deviation of populations is different. Here a population with largest Gini index is defined as the best population. In this case, we consider different values for \( \mu \) such as \((0.85, 1.20, 1.60, 2.50)\) and we consider different values for \( d^* \) such as \((0.01, 0.03, 0.05, 0.07, 0.10, 0.15, 0.20)\) and we compare the results obtained for any \( \mu \) and \( d^* \). Selection procedure is such as first case. The result of this case is shown in the rows of tables below.

**Third case**: we consider five hypothetical normal populations with parameters \((\mu_1=0.85, \sigma_1=0.45), ~ (\mu_2=1.20, \sigma_2=0.52), ~ (\mu_3=1.60, \sigma_3=0.80), ~ (\mu_4=2.50, \sigma_4=0.75)\) as the density-income population. In this case the average and Standard deviation of populations are different. Here a population with largest Gini index is defined as the best population. In this case, we consider different values for \( d^* \) such as \((0.01, 0.03, 0.05, 0.07, 0.10, 0.15, 0.20)\) and we compare the results obtained for any \( \mu, \sigma \) and \( d^* \). Selection procedure is such as first case. The result of this case is shown in the diagonals of tables below.

To do these cases, first we estimate \( \sigma_G^2 \) for each cell in the tables below, where \( \sigma_G^2 \) is obtained from equation (2.1.1). Then \( n \) calculated using equation (4.1.4) or (4.2.1), and then the Gini coefficient for each population, using equation (2.8) calculated and then obtain any population that would satisfy in the condition \( \bar{G}_{i0} - \bar{G}_{i0} \leq d^*, \forall i = 1, 2, ..., A \).

If only one population satisfies the above condition, it is the best population and if more than one population satisfies in the above condition, we’ll use some randomization scheme to select a population as the best population. In this article consider \( p^* \) equal 0.90.
6. Result:

|   | $\mu$ | 0.45  | 0.52  | 0.80  | 0.75  | Best population |
|---|-------|-------|-------|-------|-------|-----------------|
| $d^* = 0.01$ | $\sigma_b$ | $2.0664$ | $2.2921$ | $3.1909$ | $2.7818$ | 3 |
| 0.85 | $n$ | 217186 | 240909 | 335362 | 292373 | |
| & $\sigma_b$ | 0.2973 | 0.3456 | 0.5303 | 0.4983 | |
| 1.20 | $n$ | 179489 | 175015 | 212974 | 195597 | |
| & $\sigma_b$ | 0.2112 | 0.2445 | 0.3762 | 0.3519 | |
| 1.60 | $n$ | 136425 | 137426 | 162291 | 159923 | |
| & $\sigma_b$ | 0.1586 | 0.1823 | 0.2817 | 0.2645 | |
| 2.50 | $n$ | 90047  | 95252  | 120010 | 113477 | |
| & $\sigma_b$ | 0.1013 | 0.1172 | 0.1805 | 0.1692 | |
| Best population | 1 | 1 | 1 | 1 | 1 |
| $d^* = 0.03$ | 0.45 | 0.52 | 0.80 | 0.75 | Best population |
|--------------|------|------|------|------|----------------|
| $\mu$ | $\sigma^2 = 2.1990$ | $\sigma^2 = 2.5932$ | $\sigma^2 = 3.5276$ | $\sigma^2 = 2.9138$ | 3 |
| 0.85 | $n = 25680$ | $n = 30283$ | $n = 41195$ | $n = 34028$ | 3 |
| $\bar{G}_{41} = 0.2978$ | $\bar{G}_{41} = 0.3446$ | $\bar{G}_{41} = 0.5238$ | $\bar{G}_{41} = 0.4987$ | 3 |
| $\sigma^2 = 1.5939$ | $\sigma^2 = 1.5338$ | $\sigma^2 = 2.2917$ | $\sigma^2 = 1.9124$ | 3 |
| 1.20 | $n = 18614$ | $n = 17912$ | $n = 26763$ | $n = 22334$ | 3 |
| $\bar{G}_{41} = 0.2124$ | $\bar{G}_{41} = 0.2428$ | $\bar{G}_{41} = 0.3735$ | $\bar{G}_{41} = 0.3563$ | 3 |
| $\sigma^2 = 1.3083$ | $\sigma^2 = 1.3711$ | $\sigma^2 = 1.5119$ | $\sigma^2 = 1.3730$ | 3 |
| 1.60 | $n = 15279$ | $n = 16013$ | $n = 17657$ | $n = 16034$ | 3 |
| $\bar{G}_{41} = 0.1575$ | $\bar{G}_{41} = 0.1840$ | $\bar{G}_{41} = 0.2801$ | $\bar{G}_{41} = 0.2663$ | 3 |
| $\sigma^2 = 0.8846$ | $\sigma^2 = 0.9027$ | $\sigma^2 = 1.0515$ | $\sigma^2 = 1.0772$ | 3 |
| 2.50 | $n = 10331$ | $n = 10542$ | $n = 12281$ | 3 |
| $\bar{G}_{41} = 0.1011$ | $\bar{G}_{41} = 0.1013$ | $\bar{G}_{41} = 0.1811$ | 3 |
| d' = 0.05 | 0.45 | 0.52 | 0.80 | 0.75 | Best population |
|-----------|------|------|------|------|----------------|
|           | \( \mu \) | \( \sigma \) |       |       | 3              |
| 0.85      |       |       |       |       | 3              |
|           | \( \sigma_1^2 = 2.1625 \) | \( \sigma_2^2 = 2.2350 \) | \( \sigma_3^2 = 3.3455 \) | \( \sigma_4^2 = 3.1503 \) | 3 |
|           | \( n = 9092 \) | \( n = 9396 \) | \( n = 14065 \) | \( n = 13245 \) | 3 |
| 1.20      |       |       |       |       | 4              |
|           | \( \sigma_1^2 = 1.5936 \) | \( \sigma_2^2 = 1.6217 \) | \( \sigma_3^2 = 2.1695 \) | \( \sigma_4^2 = 1.9213 \) | 4 |
|           | \( n = 6700 \) | \( n = 9121 \) | \( n = 8078 \) | \( n = 8078 \) | 4 |
| 1.60      |       |       |       |       | 3              |
|           | \( \sigma_1^2 = 1.2760 \) | \( \sigma_2^2 = 1.3370 \) | \( \sigma_3^2 = 1.5868 \) | \( \sigma_4^2 = 1.4173 \) | 3 |
|           | \( n = 5365 \) | \( n = 5622 \) | \( n = 6671 \) | \( n = 5959 \) | 3 |
| 2.50      |       |       |       |       | 3              |
|           | \( \sigma_1^2 = 0.8621 \) | \( \sigma_2^2 = 0.8891 \) | \( \sigma_3^2 = 1.1036 \) | \( \sigma_4^2 = 1.0443 \) | 3 |
|           | \( n = 3625 \) | \( n = 3738 \) | \( n = 4640 \) | \( n = 4391 \) | 3 |
| Best population | 1 | 1 | 1 | 1 | 1 |
| $d^* = 0.07$ | 0.45 | 0.52 | 0.80 | 0.75 | Best population |
|-------------|------|------|------|------|-----------------|
| $\mu$       | $\sigma$ | $\sigma^2$ | $\sigma^2$ | $\sigma^2$ | 3               |
| 0.85        | $n = 4655$ | $G_{11} = 0.2992$ | $G_{12} = 0.3448$ | $G_{13} = 0.5250$ | $G_{14} = 0.4867$ | 3               |
|             | $\sigma^2 = 2.1698$ | $\sigma^2 = 2.4621$ | $\sigma^2 = 3.4778$ | $\sigma^2 = 2.9356$ |                 |
| 1.20        | $n = 3456$ | $G_{21} = 0.2109$ | $G_{22} = 0.2456$ | $G_{23} = 0.3782$ | $G_{24} = 0.3508$ | 4               |
|             | $\sigma^2 = 1.6111$ | $\sigma^2 = 1.6788$ | $\sigma^2 = 2.1171$ | $\sigma^2 = 2.1611$ |                 |
| 1.60        | $n = 2804$ | $G_{31} = 0.1593$ | $G_{32} = 0.1844$ | $G_{33} = 0.2788$ | $G_{34} = 0.2672$ | 4               |
|             | $\sigma^2 = 1.3071$ | $\sigma^2 = 1.3329$ | $\sigma^2 = 1.6006$ | $\sigma^2 = 1.5842$ |                 |
| 2.50        | $n = 1870$ | $G_{41} = 0.1032$ | $G_{42} = 0.1152$ | $G_{43} = 0.1763$ | $G_{44} = 0.1722$ | 3               |
|             | $\sigma^2 = 0.8717$ | $\sigma^2 = 0.8815$ | $\sigma^2 = 1.0577$ | $\sigma^2 = 1.1007$ |                 |
| Best population | 1 | 1 | 1 | 1 | 3 |
| $d^* = 0.10$ | 0.45 | 0.52 | 0.80 | 0.75 | Best population |
|-------------|------|------|------|------|-----------------|
| $\mu$       |      |      |      |      |                 |
|             | $\sigma_2^2 = 2.2062$ | $\sigma_2^2 = 2.2657$ | $\sigma_2^2 = 2.7915$ | $\sigma_2^2 = 2.6999$ | 3 |
| 0.85        | $n = 2319$ | $n = 2382$ | $n = 2934$ | $n = 2838$ |                 |
|             | $G_{21} = 0.2934$ | $G_{22} = 0.3377$ | $G_{23} = 0.5577$ | $G_{24} = 0.5217$ |                 |
|             | $\sigma_2^2 = 1.6379$ | $\sigma_2^2 = 1.6664$ | $\sigma_2^2 = 2.0923$ | $\sigma_2^2 = 1.9507$ | 4 |
| 1.00        | $n = 1722$ | $n = 1772$ | $n = 2200$ | $n = 2051$ |                 |
|             | $G_{21} = 0.2092$ | $G_{22} = 0.3737$ | $G_{23} = 0.3691$ | $G_{24} = 0.2569$ | 4 |
| 1.20        | $n = 1257$ | $n = 1372$ | $n = 1554$ | $n = 1257$ |                 |
|             | $G_{21} = 0.1593$ | $G_{22} = 0.1822$ | $G_{23} = 0.2815$ | $G_{24} = 0.2569$ | 4 |
| 1.60        | $n = 919$ | $n = 959$ | $n = 1167$ | $n = 1088$ |                 |
|             | $G_{21} = 0.1017$ | $G_{22} = 0.1166$ | $G_{23} = 0.1848$ | $G_{24} = 0.1737$ | 3 |
| 2.50        | 1 | 2 | 1 | 1 | 3 |

**Best population**

1 | 2 | 1 | 1 | 3
| $d^* = 0.15$ | $\sigma$ | $\mu$ | 0.45 | 0.52 | 0.80 | 0.75 | Best population |
|-------------|---------|-------|------|------|------|------|-----------------|
|             |         |       |      |      |      |      |                 |
| 0.85        | $\hat{\sigma}^2 = 2.2550$ | $\hat{n} = 1054$ | $\bar{C}_{11} = 0.3057$ | $\hat{\sigma}^2 = 2.1603$ | $\bar{C}_{12} = 0.3439$ | $\bar{C}_{13} = 0.5264$ | $\hat{\sigma}^2 = 3.1348$ | $\bar{C}_{14} = 0.4960$ | 4 |
| 1.20        | $\hat{\sigma}^2 = 1.6325$ | $\bar{C}_{22} = 0.2378$ | $\hat{\sigma}^2 = 1.7086$ | $\bar{C}_{21} = 0.3942$ | $\bar{C}_{23} = 0.3565$ |                 |
| 1.60        | $\hat{\sigma}^2 = 1.2945$ | $\bar{C}_{24} = 0.2696$ | $\hat{\sigma}^2 = 1.3007$ | $\bar{C}_{33} = 0.1816$ | $\bar{C}_{32} = 0.2732$ |                 |
| 2.50        | $\hat{\sigma}^2 = 0.8727$ | $\bar{C}_{44} = 0.1008$ | $\hat{\sigma}^2 = 0.9066$ | $\bar{C}_{41} = 0.1201$ | $\bar{C}_{42} = 0.1738$ |                 |
|              | $\bar{C}_{11} = 0.517$ | $\hat{\sigma}^2 = 1.1062$ | $\bar{C}_{12} = 0.1703$ | $\bar{C}_{13} = 0.1703$ |                 |
|              | $\bar{C}_{22} = 0.424$ | $\hat{\sigma}^2 = 1.0437$ | $\bar{C}_{23} = 0.488$ | $\bar{C}_{24} = 0.488$ |                 |
|              | $\bar{C}_{33} = 0.517$ | $\hat{\sigma}^2 = 1.1062$ | $\bar{C}_{32} = 0.1703$ | $\bar{C}_{34} = 0.1703$ |                 |
|              | $\bar{C}_{44} = 0.424$ | $\hat{\sigma}^2 = 1.0437$ | $\bar{C}_{41} = 0.488$ | $\bar{C}_{42} = 0.488$ |                 |
|              | $\bar{C}_{31} = 0.488$ | $\hat{\sigma}^2 = 1.1062$ | $\bar{C}_{34} = 0.1703$ | $\bar{C}_{32} = 0.1703$ |                 |
|              | $\bar{C}_{43} = 0.488$ | $\hat{\sigma}^2 = 1.0437$ | $\bar{C}_{41} = 0.488$ | $\bar{C}_{42} = 0.488$ |                 |
|              | $\bar{C}_{42} = 0.488$ | $\hat{\sigma}^2 = 1.1062$ | $\bar{C}_{41} = 0.488$ | $\bar{C}_{42} = 0.488$ |                 |

Best population: 4
Conclusion:

We consider five hypothetical normal populations with different parameters as the density-income populations. The average and Standard deviation of populations are different. Here a population with largest Gini index is defined as the best population. We consider different values for $d^*$ and we compare the results obtained for any $\mu, \sigma$ and $d^*$. The results are shown in the rows, columns and diagonals of above tables.

| $\sigma \mu$ | 0.45 | 0.52 | 0.80 | 0.75 | Best population |
|-------------|------|------|------|------|----------------|
| $d^* = 0.20$ |      |      |      |      |                |
| 0.85        | $\sigma^2 = 2.3093$ | $\sigma^2 = 2.3338$ | $\sigma^2 = 3.0700$ | $\sigma^2 = 2.9528$ | 4               |
|             | $n = 607$ | $n = 614$ | $n = 807$ | $n = 776$ |                |
|             | $G_{11} = 0.3034$ | $G_{12} = 0.3713$ | $G_{13} = 0.4738$ | $G_{14} = 0.5021$ |                |
| 1.20        | $\sigma^2 = 1.6476$ | $\sigma^2 = 1.6183$ | $\sigma^2 = 2.0212$ | $\sigma^2 = 2.0064$ | 4               |
|             | $n = 433$ | $n = 614$ | $n = 532$ | $n = 528$ |                |
|             | $G_{21} = 0.1956$ | $G_{22} = 0.2268$ | $G_{23} = 0.3972$ | $G_{24} = 0.3526$ |                |
| 1.60        | $\sigma^2 = 1.2783$ | $\sigma^2 = 1.3162$ | $\sigma^2 = 1.5623$ | $\sigma^2 = 1.5189$ | 3               |
|             | $n = 336$ | $n = 346$ | $n = 411$ | $n = 400$ |                |
|             | $G_{31} = 0.1508$ | $G_{32} = 0.1928$ | $G_{33} = 0.2665$ | $G_{34} = 0.2617$ |                |
| 2.50        | $\sigma^2 = 0.8626$ | $\sigma^2 = 0.9075$ | $\sigma^2 = 1.0498$ | $\sigma^2 = 1.0509$ | 2               |
|             | $n = 227$ | $n = 239$ | $n = 276$ | $n = 277$ |                |
|             | $G_{41} = 0.1036$ | $G_{42} = 0.1240$ | $G_{43} = 0.1773$ | $G_{44} = 0.1651$ |                |
| Best        | 2     | 3     | 1     | 1     | 3               |
| population  |      |      |      |      |                |
References:

[1] Akinson, A.B. (1970). On the measurement of inequality. Journal of Economic Theory, 2, 244-263.

[2] Anand, S. (1983). Inequality and Poverty in Malaysia: Measurement and Decomposition. Oxford University Press, New York.

[3] Anand, S., Diderichsen, F., Evans, T., Shkolnikov, V., and Wirth, M. (2001). Measuring disparities in health: methods and indicators. In: Challenging Inequities in Health: From Ethics to Action. Ed. By T. Evans, M. Whitehead, F. Diderichsen, A. Bhuiya. Oxford University Press, New York, 49-67.

[4] Barrett, G.F. and Pendakur, K. (1995). The Asymptotic Distribution of the Generalized Gini Indices of Inequality, The Canadian Journal of Economics, 28, 1042-1055.

[5] Beach, C.M. and Davidson, R. (1983). Distribution-free statistical inference with Lorenz curves and income shares. The Review of Economic Studies, 50, 723-735.

[6] Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances, Annals of Mathe. Statist., 25, 16-39.

[7] Bechhofer, R.E., Dunnett, C.W., and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common known variance. Biometrika, 41, 170-176.

[8] Bhattacharya, N. and Mahalanobis, B. (1967). Regional disparities in household consumption in India. Journal of American Statistical Association, 62, 143-161.

[9] Braun, D. (1988). Multiple measurements of U.S. income inequality. The review of Economics and Statistics. 70, 398-405.

[10] Chakravarty, S.R. (1988). Extended Gini indices of inequality, International Economic Review, 29, 147-156.

[11] Dalton, H. (1920). The measurement of the inequality of incomes, Economic Journal, 30,348-361.

[12] Formby, J.P., Seaks, T.G., and Smith, W.J. (1989). On the measurement and trend of inequality: A reconsideration. The American Economic Review, 79, 256-264.

[13] Gail, M.H., and Gastwirth, J.L. (1978). A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve, Journal of the American Statistical Association, 73, 787-793.

[14] Gastwirth, J.L. (1974). Large sample theory of some measures of income inequality. Econometrica, 42, 191-196.

[15] Gibbons, J.D., Olkin, I., and Sobel, M. (1977). Selecting and Ordering Populations. Wiley Publications, New York.

[16] Hayre, L.S. and Gittins, J.C. (1981). Sequential selection of the larger of two normal means, J. Amer. Statist. Assoc. 76, 696-700.

[17] Illsey, R. and Le Grand, J. (1987). The measurement of inequality in health, In: Health and Economics, Ed. By A. Williams, Macmillan, London, 13-36.

[18] Karoly, L.A. (1990). Changes in the distribution of individual earnings in the United States. The Review of Economics and Statistics, 74, 107-115.

[19] Madhuri S. Mulekar. (2005). Selection using Gini Coefficient, Presented at the Gini-Lorenz Conference, Sienna, Italy.

[20] Moore, R.E. (1996). Ranking income distributions using the geometric mean and a related general measure, Southern Economic Journal, 63, 69-75.

[21] Nelson, J.I. (1984). Income inequality: The American States, Social Science Quarterly, 5, 854-860.

[22] Rao, C.R. (1965). Linear Statistical Inference and Its Applications, John Wiley and Sons, New York, pp 355.

[23] Sandstrom, A., Wretman, J.H., and Walden, B. (1988). Variance estimators of the Gini sampling. Journal of Business and Economic Statistics, 6, 113-119.
[24] Shkolnikov, V.M., Andreev, E.E., and Begun, A.Z. (2003). Gini Coefficient as a life table function: Computation from discrete data decomposition of differences and empirical examples. Demographic Research, 80, 305-385.

[25] Soltow, L. (1971). Patterns of wealthholding in Wisconsin since 1850. Wisconsin University Press, Madison.

[26] Stuart, A. and Ord, J.K. (1994). Kendall’s Advanced Theory of Statistics: Volume 1, Distribution Theory, Sixth Ed. Edward Arnould, Great Britain.

[27] Theil, H. (1967). Economics and Information Theory, North-Holland Publishing Co., Amsterdam.

[28] Zitikis, R. (2003). Asymptotic estimation of the E-Gini index, Economic Theory, 19, 587-601.