Racial Isolation Decomposition

Don McLaughlin
October 8, 2002
Revised August 20, 2003

Racial isolation, *R*, for a group (*G)* (e.g., blacks) versus others, for units *A* (e.g., schools) within a larger unit *B* (e.g., a state), should be computed as follows.

Definition:

$$R_{G:AB} = 1 - \frac{\dfrac{\sum_{A\,in\,B} N_A p_{AG}(1-p_{AG})}{N_B p_{BG}}}{\dfrac{\sum_{AinB} N_A(1-p_{AG})}{N_B}} = 1 - \frac{\sum_{A\,in\,B} N_A p_{AG}(1-p_{AG})}{N_B p_{BG}(1-p_{BG})}.$$

where N_A, N_B are the numbers of students in unit *A*, in unit *B*; and p_{AG} and p_{BG} are the proportions of group members in units *A* and *B*.

That is, R is 1 minus the exposure of members of group *G* (e.g., blacks) to others divided by the exposure of all to others. R is zero if p_{AG} is the same for all *A* (complete desegregation), and R is one if p_{AG} is one or zero for all *A* (complete segregation).

Theorem 1:
$$R_{G:AB} = \frac{\sigma^2_{p_G,A/B}}{p_{BG}(1-p_{BG})}$$

That is, racial isolation is the ratio of the variance among school (unit *A*) percentages of group *G* enrollment to the overall state (unit *B*) variance in group *G* enrollment.

Theorem 2:
$$\sigma^2_{p_G,S/T} = \sigma^2_{p_G,L/T} + \sum_L \left(N_L \big/ N_T \right) \sigma^2_{p_G,S/L}$$

where *S* refers to school, *L* to LEA, and *T* to state.

This leads to:
$$R_{G:ST} = R_{G:LT} + \sum_L \left(\frac{N_L p_{LG}(1-p_{LG})}{N_T p_{TG}(1-p_{TG})} \right) R_{G:SL}.$$

That is, racial isolation between schools in a state can be partitioned into (1) racial isolation between districts in the state, plus (2) a weighted average of racial isolation within districts. The "weight" for the within-district isolation in district *L* is given by the coefficient of $R_{G:SL}$. This equation can be reversed to obtain an estimate of the average $R_{G:SL}$

Theorem 3:
$$\overline{R}_{G:SL} = (R_{G:ST} - R_{G:LT})/(1 - R_{G:LT})$$

In other words, using this equation is equivalent to weighting each $R_{G:SL}$ by $N_L\, p_{LG}\,(1-p_{LG})$ in computing the average.

Proof of theorem 1:

$$R_{G:AB} = 1 - \frac{\sum_{A\,in\,B} N_A\, p_{AG}(1-p_{AG})}{N_B\, p_{BG}(1-p_{BG})}$$, and if we let $v_{AG} = p_{AG} - p_{BG}$,

$$R_{G:AB} = 1 - \frac{\sum_{A\,in\,B} N_A (v_{AG}^2 - v_{AG}(1-2p_{BG}))}{N_B\, p_{BG}(1-p_{BG})} = \frac{\sum_{A\,in\,B} N_A v_{AG}^2}{N_B\, p_{BG}(1-p_{BG})} = \frac{\sigma^2_{p_G, A/B}}{p_{BG}(1-p_{BG})}$$

Proof of Theorem 2:

$$\sigma^2_{p_G,S/T} = \sum_{LS} N_{LS}(p_{LSG} - p_{TG})^2 \Big/ N_T = \sum_{LS} N_{LS}(p_{LSG} - p_{LG} + p_{LG} - p_{TG})^2 \Big/ N_T$$

$$= \left(\sum_{LS} N_{LS}(p_{LSG} - p_{LG})^2 + \sum_{L} N_L (p_{LG} - p_{TG})^2 \right) \Big/ N_T$$

$$= \sum_L (N_L / N_T) \left(\sum_S N_{LS}(p_{LSG} - p_{LG})^2 \Big/ N_L \right) + \sum_L N_L (p_{LG} - p_{TG})^2 \Big/ N_T$$

$$= \sum_L (N_L / N_T)\, \sigma^2_{p_G,S/L} + \sigma^2_{p_G,L/T}$$

Proof of Theorem 3:

We can rewrite $R_{G:ST} = R_{G:LT} + \sum_L \left(\frac{N_L\, p_{LG}(1-p_{LG})}{N_T\, p_{TG}(1-p_{TG})} \right) R_{G:SL}$ as

$$R_{G:ST} = R_{G:LT} + \frac{\sum_L (N_L\, p_{LG}(1-p_{LG})) R_{G:SL}}{\sum_L (N_L\, p_{LG}(1-p_{LG}))} \frac{\sum_L (N_L\, p_{LG}(1-p_{LG}))}{N_T\, p_{TG}(1-p_{TG})}$$

$$= R_{G:LT} + \frac{\sum_L (N_L\, p_{LG}(1-p_{LG})) R_{G:SL}}{\sum_L (N_L\, p_{LG}(1-p_{LG}))} (1 - R_{G:LT}) = R_{G:LT} + \overline{R}_{G:SL}(1 - R_{G:LT})$$

Rearranging terms, we have $\overline{R}_{G:SL} = (R_{G:ST} - R_{G:LT})/(1 - R_{G:LT})$.

Generalization to *K* groups, *G*$_1$, ..., *G*$_k$.

The total isolation among a set of subpopulations is one minus the sum of exposure rates for each pair of groups, for all schools in a district, divided by the sum of exposure rates for the district.

Definition: $$R_{T:AB} = 1 - \frac{\sum_{A\,in\,B} N_A \sum_{1 \le i < j \le K} p_{AG_i} p_{AG_j}}{N_B \sum_{1 \le i < j \le K} p_{BG_i} p_{BG_j}}$$

where the first subscript T refers to "total" over all groups.

Lemma 1. $\sum_{i=1}^{K} p_{G_i}(1 - p_{G_i}) = 2 \sum_{1 \le i < j \le K} p_{G_i} p_{G_j}$ for any unit *A* or *B*.

Lemma 1 allows us to write total isolation in terms either of a weighted sum of isolation between all pairs of groups or of a weighted sum of isolation of each group from the collection of all others.

Using Lemma 1, $$R_{T:AB} = 1 - \frac{\sum_{A\,in\,B} N_A \sum_{i=1}^{K} p_{AG_i}(1 - p_{AG_i})}{N_B \sum_{i=1}^{K} p_{BG_i}(1 - p_{BG_i})}$$

We can additively partition the total multi-group isolation $R_{T:AB}$ into a weighted average of the isolation $R_{GiGj:AB}$ for the individual pairs of groups if we define $R_{GiGj:AB}$ for each pair of groups G$_i$ and G$_j$ included in the total, such that isolation between two groups is measured taking into account the other groups (in a school with equal numbers of each of *K* groups, the exposure of any group to any other would be 1/*K*). In that case,

$$R_{T:AB} = \sum_{1 \le i < j \le K} p_{BG_i} p_{BG_j} R_{G_iG_j:AB} \Big/ \sum_{1 \le i < j \le K} p_{BG_i} p_{BG_j} \text{ where } R_{G_iG_j:AB} = 1 - \frac{\sum_{A\,in\,B} N_A p_{AG_i} p_{AG_j}}{N_B p_{BG\,i} p_{BG_j}};$$

or equivalently, we can define the total isolation in terms of the weighted average of the isolation of each group from "all others."

$$R_{T:AB} = \sum_{i=1}^{K} p_{BG_i}(1 - p_{BG_i}) R_{G_i:AB} \Big/ \sum_{i=1}^{K} p_{BG_i}(1 - p_{BG_i}).$$

These results allow us to partition total multi-group isolation between schools in a state into total multi-group isolation between districts in the state and multi-group isolation between schools in each district:

Theorem 4:

$$R_{T:ST} = R_{T:LT} + \sum_L \left(\frac{N_L \sum_{1 \le i < j \le K} p_{LG_i} p_{LG_j}}{N_T \sum_{1 \le i < j \le K} p_{TG_i} p_{TG_j}} \right) R_{T:SL} = R_{T:LT} + \sum_L \left(\frac{N_L \sum_{i=1}^{K} p_{LG_i} (1 - p_{LG_i})}{N_T \sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})} \right) R_{T:SL}$$

Proof of theorem 4:

By Lemma 1, the two expressions for $R_{T:ST}$ are equivalent.

Insert the definitions for $R_{T:LT}$ and $R_{T:SL}$ in the equation:

$$R_{T:ST} = R_{T:LT} + \sum_L \left(\frac{N_L \sum_{i=1}^{K} p_{LG_i} (1 - p_{LG_i})}{N_T \sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})} \right) R_{T:SL}$$

$$R_{T:ST} = 1 - \sum_L \left(\frac{N_L \sum_{i=1}^{K} p_{LG_i} (1 - p_{LG_i})}{N_T \sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})} \right) + \sum_L \left(\frac{N_L \sum_{i=1}^{K} p_{LG_i} (1 - p_{LG_i})}{N_T \sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})} \right) \left[1 - \sum_S \left(\frac{N_S \sum_{i=1}^{K} p_{SG_i} (1 - p_{SG_i})}{N_L \sum_{i=1}^{K} p_{LG_i} (1 - p_{LG_i})} \right) \right]$$

Canceling the equal terms yields:

$$R_{T:ST} = 1 - \sum_{S \text{ in } T} \left(\frac{N_S \sum_{i=1}^{K} p_{SG_i} (1 - p_{SG_i})}{N_T \sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})} \right)$$, which is true by definition.

Note that for two groups, this is identical to the previous result.

Putting this together, we have a decomposition of total isolation by level and subgroup:

$$R_{T:ST} = \frac{\sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})\, R_{G_i:LT}}{\sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})} + \frac{\sum_{i=1}^{K} \sum_L (N_L / N_T)\, p_{LG_i} (1 - p_{LG_i}) R_{G_i:SL}}{\sum_{i=1}^{K} p_{TG_i} (1 - p_{TG_i})} \; .$$

Next step is general formula for n levels (e.g., schools within districts within states within regions within the country).

The data are n_{gclsr}, the enrollment of group g in campus c in district l in state s in region r.