Use of the gravity-potential model in studies of spatial interactance has resulted in the accumulation of a large body of literature, which indicates that the model is helpful in explaining the spatial distribution of many phenomena involving human activity over space. Its use was suggested by Carey (1858-59), and was first applied to a study of student enrollment by Stewart (1941).

The general formula for the gravity-potential model is derived by analogy from the gravity model of Newtonian physics. In its simplest form, it is hypothesized that

\[ I_{ij} = k \frac{P_i P_j}{D_{ij}} \]

where:
- \( I_{ij} \) is the number of units of the form of interactance of the phenomenon being studied between \( i \) and \( j \)
- \( P_i \) is the population of area \( i \)
- \( P_j \) is the population of area \( j \)
- \( D_{ij} \) is the distance between area \( i \) and area \( j \)
- \( k \) is a constant of proportionality.

Stated verbally, the number of units of the dependent variable originating in area \( j \) is directly proportional to the population of that area multiplied by the population of area \( i \) and inversely proportional to the distance between area \( i \) and area \( j \).

Empirically, this relationship has been found to hold for a wide variety of phenomena, including consumer shopping behavior, migration, and airplane, rail, and road trips. Since the rationale for use of the model is based on probability reasoning, it applies to aggregates rather than to individuals. It is hypothesized that there is a certain probability \( p \) that any individual in a given population will engage in the particular type of behavior defined by the dependent variable. Thus, for a large number of people, the proportion defined by \( p \) may be expected to engage in this behavior. Stated otherwise, a given population \( P_j \) may be expected to generate \( p P_j \) units of this behavior; thus, the number of units of the behavior generated by that population will be directly proportional to the population. Since, in addition, the behavior is one which involves overcoming some aspect of the time-distance-cost relationship, it is further hypothesized that the greater the distance, time, cost, or some combination of these factors, the smaller the probability that the individual will engage in that behavior. That is, a person is more likely to move or to travel a short distance than a longer one.
Correlation-regression analysis is used to obtain the predicted values of interactance $l_{ij}$. The coefficient of correlation, $r$, is used to test the correspondence between these and the observed values for each unit area; its square, the coefficient of determination, $r^2$, represents the proportion of the spatial variation in the observed distribution of the phenomenon which is explained by the model. Since the value of $r$ in many of the early studies was not as high as had been hoped, many approaches to improving it were used. Measures for the $P$ values other than the total population of the area, which were believed to reflect better the actual population which might generate the particular activity were tried. Likewise, different measures for distance employing some combination of time-distance-cost were used. Some investigators found that the use of an exponent for the $D$ value leads to more accurate prediction, although at present no theoretical basis for selecting an exponent exists, and it must be derived empirically. The $P$ and $D$ values may also be weighted in some manner, as by income. Another approach involves the addition of independent variables to the regression equation, so that other factors thought to be relevant may be taken into account.

Stewart (1941) used the basic equation, except for modifying the $P$ value. Since he was investigating student enrollments at Princeton and Harvard, he used the native white male population, from which the majority of the students were drawn. Distances for each state were estimated by drawing concentric circles with centers at the college being studied. He found that the number of undergraduates from each ring was roughly proportional to the source population and inversely proportional to its distance from the college. Since he did not compute coefficients of correlation, his results cannot be compared precisely with those of other studies.

In his study of student enrollment at Bowling Green University in Ohio, McConnell (1965) experimented with a number of different modifications, retaining ones which seemed to him most justifiable and which yielded the best predictions. He tried two different population measures: total population size, and population weighted by per capita income. He also applied an exponent to the $D$ value, substituting $D^2$ in two formulations. In other trials, the number of intervening opportunities, as measured by the number of other colleges closer to the county of residence than Bowling Green, was substituted for the $D$ value.

None of the six combinations tried yielded values which were significantly higher statistically than those of the original model. For this formulation, the value of $r$ was $0.905$, $r^2 = 0.819$. The highest value of $r$, $0.919$, $r^2 = 0.844$, was obtained by weighting population with income and using the square of the distance, but this was found not to differ significantly from the value of $r$ in the original formulation.

Examination of the residuals from regression showed that enrollments for counties located nearer other large public and private universities than to Bowling Green and those farthest away from Bowling Green tended to be over-predicted. Counties located nearer Bowling Green than other large universities,