at marriage, years married, or years of education, reported it had taken more time to work out adjustment in sex relations than in any other area. They agreed in listing the rest of the areas in the following order: spending the family income, social activities and recreation, in-law relationships, religion in the home, and associating with mutual friends.

There was a very close relationship between the length of time required to adjust in marriage and the happiness of the marriage.

If couples failed to work out adjustment in two or more areas, they classified their marriage as average or unhappy.

The study confirms the findings of others that age at marriage, education, income, and health are associated with happiness in marriage.

A similar study representing a cross-section of the population would doubtless produce different findings. This is a study of successful marriages among the parents of college students. It was felt that in building marriage courses for college students, there is need for information on successful marriage with attention to varying lengths of time required to achieve success in the different areas. It would be of value to conduct similar studies among the parents of college students in other parts of the country. It would also be desirable to get the same type of information from people representing different social, economic, residential, nationality, and racial backgrounds.

# THE $\frac{P_1 \, P_2}{D}$ HYPOTHESIS: ON THE INTERCITY MOVEMENT OF PERSONS

George Kingsley Zipf
*Harvard University*

In the present paper we shall show with supporting data that the number of persons that move between any two communities in the United States whose respective populations are $P_1$ and $P_2$ and which are separated by the shortest transportation distance, $D$, will be proportionate to the ratio, $P_1 \cdot P_2/D$, subject to the effect of modifying factors.

The data in support of the above proposition are the highway, railway and airway data for an arbitrary set of cities during intervals of measurement in 1933-34. Before presenting the data, however, we shall give a brief theoretical discussion of the proposition itself with illustrations from other kinds of observations with which the above data are intimately connected.

## I. THEORETICAL DISCUSSION

In 1940 the author published the observation that the following equation of the generalized harmonic series described the recent distribution of communities in India, Germany, and certain other countries including the United States (for communities of 2,500 or more inhabitants), when the communities are arranged in the order of decreasing size, with $A$ representing the population of the largest community, and with the denominators referring to the ranks of the communities thus arranged:[1]

$$A \ Sn = \frac{A}{1^p} + \frac{A}{2^p} + \frac{A}{3^p} + \ldots . + \frac{A}{n^p}$$

In Figure 1 are presented the United States urban data for 1930 and 1940, as indicated, to which an ideal line, $A$, with a negative slope of 1 (i.e. $p = 1$) has been added to aid the reader's eye. The linearity of the data is apparent.

In 1941 the author presented a fuller

[1] Zipf, G. K. "The generalized harmonic series as a fundamental principle of social organization." *Psychological Record,* 4 (1940), 43.

treatment of the topic[2] and included a theoretical discussion, in the form of a lemma (ibid. pp. 91-135), of why the proportions of the generalized harmonic series will emerge in the communities of a national social system under the postulate of reducing to a minimum the sum of all products of masses moved, when multiplied by their re-

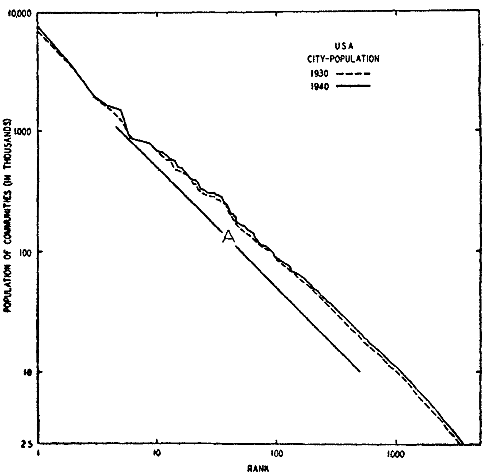


FIGURE 1. The Rank-Frequency Distribution of Communities in the United States in 1930 and 1940 (with the ideal line, *A*, with a negative slope of 1).

spective work-distances, *D*. This lemma set forth the reasons for the emergence of communities at all, and also for their number, relative sizes, and locations under the above work-minimum.

According to this lemma, the number, sizes, and locations of communities represent equilibria between *first* the economy of the population's living immediately at the source of its raw materials on the one hand (e.g. at the farm or at the mine pit), and, *second*, the economy of the population's living together in one big city where all the manufacturing is done. The *first* economy of living at the immediate source of raw materials saves the work of transporting the raw materials to the production-centers; the *second* economy of living in one big city where all production is done saves the work of transporting the goods to the consumers. Since the population cannot live both in a lot of communities scattered over the terrain and at the same time in a single big city, it is obvious that the above two economies are in conflict. This conflict, if our lemma be correct, will govern the *n*-number of different communities and their respective *P*-population-sizes, for the following reasons.

To begin, if we assume that a given large terrain has a fixed total population, then the average *P*-population-size of the *n*-different communities will be inversely proportionate to *n*. As the *n*-number of different communities increases, their average *P*-sizes decrease, and *vice versa*.

Obviously the (*first*) economy of living at the immediate source of raw materials will act in the direction of making a large *n*-number of different communities of small *P*-sizes, if the terrain is reasonably homogeneous in its distribution of raw materials per unit of area in terms of cost in man-hours in procuring them. Because of its diversifying effect in terms of the number of different communities, this (*first*) localizing economy may be called the Force of Diversification.

By the same token, the opposite (*second*) economy which places a premium upon living together in one big city may be called the Force of Unification[3] since it acts in the direction of reducing the *n*-diversity of different communities to 1, while increasing the *P*-size of that 1 community to 100%.

The actual *n*-number of different communities and their respective *P*-sizes will depend upon the comparative magnitudes of these two quasi Forces, as we pointed out at the time in our above-mentioned lemma to which we here only refer. If we assume that in a given national social economy the Forces of Diversification and of Unification

[2] Zipf, G. K. *National Unity and Disunity*, Bloomington, Ind.: Principia Press, 1941.

[3] The Force of Unification comes ever more into operation as the goods of production become ever more diversified in respect of their components and hence ever less likely to be produced from raw materials that are found together in one location. In short this Force is associated with industrialization and trade. (See below.)

are of equal magnitude at a given time, then, as far as the $n$-number and $P$-sizes of the resulting communities are concerned, we may expect to find the relationship of an equilateral hyperbola because of the nature of the factors involved. This would seem to be the case, under the assumption of a fixed total population, because one community can grow in $P$-size only at the expense of the $n$-number, or the $P$-sizes (or both) of the other communities.

If the relationship is indeed that of an equilateral hyperbola, then we should find that the $n$ number of different cities, when ranked, $r$, in the order of their decreasing $P$-sizes, would follow the equation, $r \cdot P = n$, where $r$ is the ordinal rank of the community, and where $r = 1$ represents the rank of the largest community, and where $r = n$ represents that of the smallest community, and where $r$ takes on only positive integral values from 1 through $n$.

The data of Figure 1, plotted rather to the equation, $\log r + a \log P = \log C$, is not an inordinately bad fit of our equation.

Naturally as the magnitude of either the Force of Unification or the Force of Diversification becomes greater, then the relationship between $r$ and $P$ will change accordingly (the two being logarithmically related); thus for example, as a largely rural country becomes ever more extensively industrialized, as was the case in Germany from 1870 onwards, then the slope of the resulting rank-population curve will increase accordingly, as was the case with Germany.[4]

According to the above lemma which we are here summarizing only in barest outline, the number and sizes and locations of communities in a given social economy represent equilibria in the minimizing of work in transporting raw materials through industrial processes to consumers. This lemma, however, applies only to those social systems that for the most part produce what they consume and consume what they produce, and only under the assumption that the system is minimizing its total work in the entire movement of all materials and persons.

Moreover it applies only to those cases where all members of the population get an approximately equal share of the national income in the sense that the average real income per person is about the same in any community regardless of its size, and where also an approximately equal percentage of persons in each community is gainfully employed.

Of course under the conditions of the above equal average income and of an equal proportion of gainfully employed, it follows that any community, $P$, will contribute to the total $C$ production of the system an amount in value that is proportionate to $P / C$; moreover it will receive from the system as a reward an amount in value that is proporionate to $P / C$. Or, if one will, a community, $P$, will *put-into-the-system-and-take-out-of-the-system* an amount that is proportionate to $P / C$ during an interval of measurement. Naturally we are speaking only in terms of the monetary value of goods (including services) and in no way imply that the precise goods and services that are put into the system by a $P$-population are the ones taken out as their reward (cf. lemma, *op. cit.*). On the contrary each $P$-community is receiving goods and services from, and sending goods and services to, the rest of the population in the flow of goods and services through the economy.

Now, if during a given interval of measurement a community, $P_1$, has a share of the flow of goods, by value, that is equal to $P_1 / C$, whereas community $P_2$ during the same time has a share equal to $P_2 / C$, then the interchange of goods, by value, between $P_1$ and $P_2$ would be proportionate to $(P_1 / C)$ $(P_2 / C)$, or $(P_1 \cdot P_2) / C^2$, *provided* that we ignore the factor of the easiest intervening transportation distance, $D$.

If we remember, however, that the number, sizes, and locations of communities depend theoretically upon the minimizing of the work of transporting mass over distance —indeed they are equilibria between the opposing Forces of Unification and Diversi-

[4] For German population data 1875-1939 cf. Zipf, *National Unity, op. cit.*, p. 140f.

fication—then the interchange of goods between communities $P_1$ and $P_2$ will tend to be inversely proportionate to their intervening easiest transportation distance, $D$. With the addition of the factor, $D$, the interchange, in value, will be directly proportionate to $\frac{P_1 P_2}{D}$ for any two cities in the economy.[5]

Obviously the factors, $P_1$, $P_2$ and $D$ are empirically ascertainable for the communities of a given national social system, like the United States, and the problem of assessing the actual rate of flow between them for some classes of goods is not imponderable, even though information on the actual monetary value of interchanged goods during a given period is unavailable to the present writer.

But if we assume, for example, that the

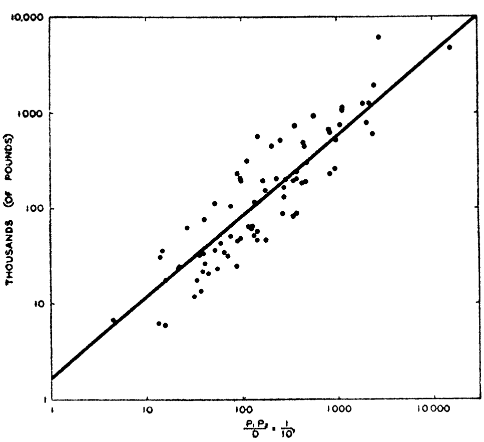


FIGURE 2. The movement of railway express (less carload lots) between 13 arbitrarily selected cities in the U.S.A. during May, 1939.

value of Railway Express per thousand pounds is fairly constant, then we can test the validity of our hypothesis by plotting the weight of Railway Express that is interchanged between a set of pairs of communities, and their respective values of $P_1 \cdot P_2 / D$. This is done in Figure 2, for the interchange of Railway Express in thousands of pounds (in less than carload lots) during the month of May, 1939, between the following 13 cities—or 78 pairs of cities: 1. Boston, 2. Buffalo, 3. Chicago, 4. Cleveland, 5. Detroit, 6. Los Angeles, 7. Milwaukee, 8. New York, 9. Philadelphia, 10. Pittsburgh, 11. St. Louis, 12. San Francisco, 13. Washington, D.C.[6]

According to our theoretical expectations the values (dots) of our data in Figure 2 should be rectilinearly distributed with a positive slope of 1. In Figure 2 the line drawn was fitted by least squares whose slope is $.85 \pm .31$, or in equation form, $\log y = .2157 + .8472 \log x$. With a *P. E.* of .2, the value of .85 may be viewed as a non-significant variation from 1.00 in a set of data of four variables in which $P$ varies from 500,000 to 700,000, and in which $D$ varies from 100 to 3,000, and where the weight of Railway Express varies from 5,000 pounds to nearly 5 million pounds. In the initial publication of these data, reasons for the deviation of .15 from our expected slope are given.

Our theory is obviously confirmed by the above data, and so too our assumptions which, being explicit, we shall not repeat here. It is interesting to note, however, the close degree to which conditions in the United States in the fourth decade approximated those that we have anticipated theoretically.

At this point it is only natural for the reader to inquire into the intercity movement of materials by other means of transportation, such as by freight, parcel post and the like which theoretically should also follow our "$P_1 \cdot P_2 / D$ hypothesis," subject to quali-

[5] Or, in equation form the $Y$-value of goods interchanged between any two cities, $P_1$ and $P_2$, separated by distance $D$ will be $Y = \frac{P_1 \cdot P_2}{D}$.

[6] Originally published, Zipf, G. K., "The $P_1 P_2/D$ Hypothesis: The Case of Railway Express," *Journal of Psychology*, 22 (1946), 3-8. The data I owe to the kindness of Mr. L. O. Head, President of the Railway Express Agency. The values of $D$ for all data in the present paper are the official military (shortest railway) distances of the War Department. The values throughout of $P_1 P_2/D$ have, for ease of representation, been divided by 10 million (i.e. multiplied by $1/10^7$).

fying factors in some particular cases. Before discussing the possibility of other kinds of data it is perhaps wise to point out at once that our theory calls for the movement of *all* goods and services by *all* means of transportation; the theory will not necessarily hold for each kind of transportation since we know that for some commodities one type of transportation is cheaper than another. The fact that our theory holds so well for Railway Express suggests that the service of Railway Express is of equal value to persons, regardless of the size of the cities or of their locations; the same may well be true of parcel post for which, unfortunately, data are lacking. Nevertheless in the case of mining communities or agricultural centers we may suspect that they ship out great values of bulky materials by railway freight while receiving payment in terms of less bulky materials that are not all sent by freight. To repeat, our theory calls for *all* shipments by *all* means; hence we may expect a certain amount of variation in the data for one particular means of shipment. Information on the intercity shipment of freight is lacking. So too is lacking information on the intercity movement of money, such as by checks. Theoretically the amount in dollars interchanged by checks of all kinds between given cities—i.e. drawn in one city and cashed in another and the reverse—will follow our "$P_1 \cdot P_2 / D$ hypothesis." Although there may be tendencies for surpluses to build up in favored cities, and for capital to move as such at times in some specific directions with only a slow return flow, these variant amounts need not be necessarily large in comparison with the large amounts of the "normal flow." Data are lacking.

On the other hand, we have observed that the circulation of newspapers follows this hypothesis as well as the amount of news about a city $P_2$ at a distance, $D$, as reported in a paper in city $P_1$; the same applies also to the amount of intercity telephone calls, although in this latter instance a modifying constant is present.[7]

We mention these further considerations only to suggest the extent of the "$P_1 \cdot P_2 / D$ hypothesis" which refers to all movement, including persons.

## 2. THE MOVEMENT OF PERSONS

In turning to the question of the intercity movement of persons we have data on passenger traffic by highways, railways and airways for intervals of measurement in 1933 and 1934. For highways (i.e. busses), the data refer to December, 1933, and July, 1944. For railways they refer to one month in each quarter in 1933 (or 4 months in all). For airways they refer to all of 1933.[8]

The data are based on the number of tickets sold during the periods in question (and not the number of tickets collected by the conductors, as we should prefer). They include single-trip and round-trip tickets without information as to when the return-trip took place.

The cities arbitrarily selected for the present study were: 1. Akron, 2. Baltimore, 3. Boston (and suburbs), 4. Buffalo (and Niagara Falls), 5. Charlotte, N.C., 6. Chicago (and suburbs), 7. Cleveland (and suburbs), 8. Denver, 9. Detroit (and suburbs), 10. Flint, 11. Grand Rapids, 12. Houston, 13. Jacksonville, Fla., 14. Los Angeles (and suburbs), 15. Memphis, 16. Miami, 17. Milwaukee (and West Allis), 18. Minneapolis (and St. Paul), 19. Newark (and suburbs), 20. New Orleans, 21. New York (and suburbs), 22. Norfolk (and Newport News and Portsmouth), 23. Philadelphia (and suburbs), 24. Pittsburgh (and McKeesport),

[7] Reported *ibid.* and presented with supporting empiric data and fitted curves in Zipf, G. K., "Some Determinants of the Circulation of Information," *American Journal of Psychology,* LIX (1946), 401-421. The present $P_1 \cdot P_2/D$ relationship (a two dimensional "gravitation") like the three dimensional gravitation of physics, though theoretically rectilinear, doubly logarithmically, can be modified in respect of slope and rectilinearity by other factors.

[8] These data are contained in *Appendix I* of the *Passenger Traffic Report* prepared by the Section of Transportation Service under the office of the Federal Coordinator of Transportation. I am grateful to the American Association of Railroads for calling my attention to this valuable report, and to Mr. John R. Turney for giving me one of the copies.

25. St. Louis (and East St. Louis), 26. San Diego, 27. San Francisco (and suburbs), 28. South Bend, 29. Washington, D.C.

These 29 arbitrarily selected cities (or about 400 pairs of cities) are widely scattered over the United States and vary sufficiently in size—roughly from 100,000 to 7,500,000—so that our sample is fair. Cities much smaller than the above would not have shown any passengers in some cases during so short a period of measurement. When no passengers travelled between a given pair of cities—notably often in the case of airways for which, incidently, Newark and New York are combined—then, obviously the pair of cities is not represented on the graph.

On the other hand the above entities include in some cases suburban populations whose sizes are impossible to disclose. Hence it was decided to use as values of $P$ the populations of the cities *exclusive* of those mentioned in the parentheses above. This will introduce a certain amount of variation in our data which we mention here at the very start; the variation will be neither favorable nor unfavorable to our hypothesis. Yet because of the ambiguity of our $P$-values, no lines were calculated.

While we are still on the subject of our data let us remember that they refer to a very depressed period in our history when the national economy was giving way, in part, to a sectionalism (i.e. the Force of Diversification) that might well influence long $D$-travel adversely. Let us also remember that the then differences in fares between bus and railway will affect the resulting passenger and rate distributions, particularly since the difference in fares becomes ever more pronounced with long distances. Let us also remember the possible influence of the preferential fares of round-trip tickets. Let us remember that with railways there are different rates for different classes of travel (e.g. Pullman, coach, etc.) which are lumped together in the data. Let us remember that private automobile traffic (and commutation traffic) is not included, and so too the movement of persons by foot with an occasional "thumbed" ride.

Yet with all these factors in mind which, if excessive, can modify both the slope and even the rectilinearity of the distributions, let us note from our below data how basic the "$P_1 \cdot P_2 / D$ hypothesis" is.

## *A. The Case of Highway Traffic*

In Figure 3 are presented logarithmically on the ordinate the number of passengers moving by bus between the above-mentioned

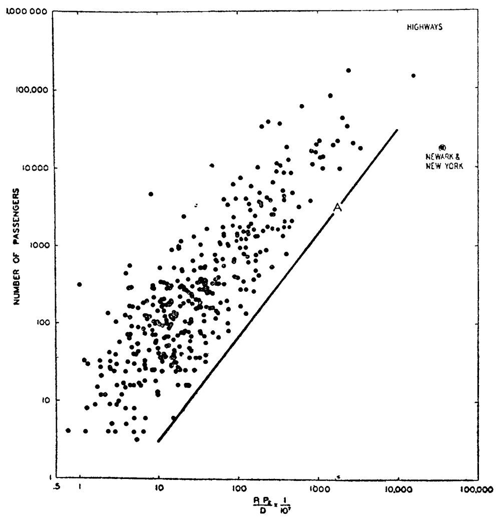


FIGURE 3. The number of passengers travelling by highway carriers between 29 arbitrary cities during December, 1933, and July, 1934 (the ideal line, $A$, has a slope of 1.25).

pairs of cities whose respective values of $P_1 \cdot P_2 / D$ are plotted logarithmically on the abscissa. The line, $A$, with an arbitrary slope of 1.25, which has been added to aid the reader's eye, represents about the upper limit of slope of the distribution which, in view of the amount of variation, is not inconsistent with a slope of 1 (may the reader cover up $A$ and draw a line of his own). In any event the rectilinearity of the data—*extending over 5 logarithm cycles*—is unmistakable. The position of the point for Newark-New York is understandable when we remember the excellent competitive rail service between these two cities.

At this point it should be borne in mind that if fares are in proportion to distance, then the aggregate fares for travel between cities will also be proportionate to $P_1 \cdot P_2 / D$, since the fewer persons who travel longer distances will pay proportionately larger fares for their tickets and the reverse.

In Figure 4 are presented the aggregate fares paid by the above bus-passengers. The line through the points was drawn by eye with a positive slope of 1, so that the reader may decide for himself whether this theoretical value is a fair value for the data. The points for Miami, Florida, are encircled to show the effect of seasonal variation (the measurement included December, 1933); the points for San Diego, Los Angeles and Jacksonville which are likewise winter resorts diverge similarly but are not indicated. The points for Chicago-South Bend and Newark-New York are indicated to suggest the presence not only of serious competition for short hauls but also, in comparison with Figure 3, the presence of cut-rates for those short hauls.

In view of the nature of the data of the above two figures—even in spite of the presence of other factors—it is not rash to suggest that our theoretical hypothesis has been confirmed.

*B. The Case of Railway Traffic*

In Figure 5 are presented the data for railway passengers of all classes between the above cities plotted as indicated, with the line, *A*, added with a positive slope of 1.00 to indicate the theoretically expected slope.

The existence of a correlation between our factors is unmistakable. Equally unmistakable is its divergence from the expected slope (the slope of the data being much nearer 1.50). This divergence is significant for us (as well as for the Federal Coordinator of Transportation who apparently undertook this study in order to find out what was wrong with the railroads in the 1930's). After all, there was no such marked divergence from the slope of 1.00 with the highway data of Figures 3 and 4.

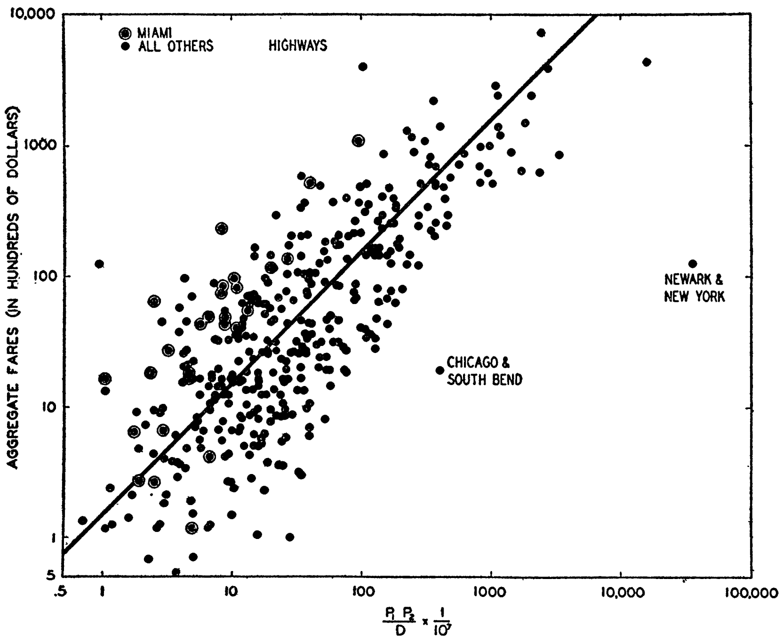


FIGURE 4. The aggregate fares (in hundreds of dollars) paid by the highway passengers reported in Figure 3. The ideal line has a slope of 1.

This divergence of Figure 5 means that as $P_1 \cdot P_2 / D$ increases, the number of passengers increases by approximately the 1.5 power thereof. In short there is a systematic premium upon small $D$ distances, since the same 29 cities are used throughout, with $D$'s of varying sizes.

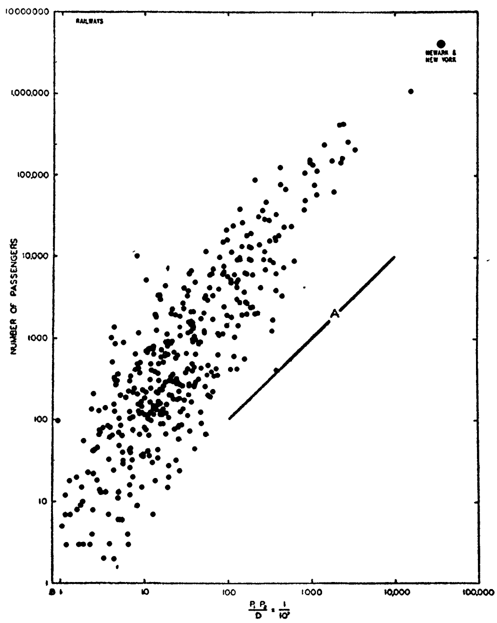


FIGURE 5. The number of passengers travelling by railways between 29 arbitrary cities during one month in each quarter of 1933 (the ideal line, $A$, has a slope of 1.00).

This systematic premium upon small $D$-distances may mean several things. In the first place it may mean, in view of the Pareto curve which presumably applies to the distribution of incomes within cities, that railway fares, in comparison with bus fares, were so high that a logarithmically decreasing number of persons out of those that travelled *could* afford to take trips of increasing $D$-distance.[9] In other words, as $D$ increased, the resulting absolute increase in cost prohibited an ever increasing percentage of persons from travelling by rail. Perhaps that is the reason why the bus-travel approximated our theoretical value more closely than did the rail-travel.

A second possible explanation is that of preferential fares for shorter (strip) tickets and for round-trip tickets—a consideration which in turn brings up the whole question of railway fares.

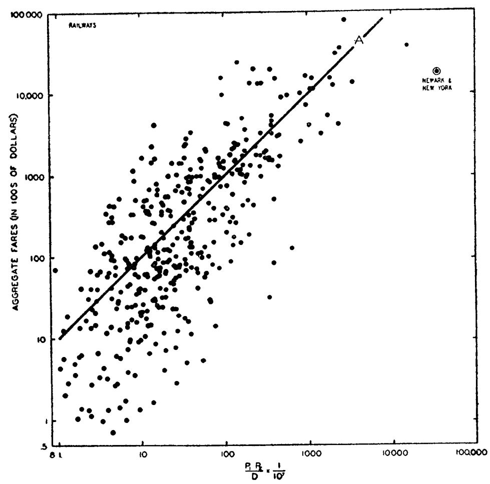


FIGURE 6. The aggregate fares (in hundreds of dollars) paid by the railway passengers reported in Figure 5. The ideal line has a slope of 1.

In Figure 6 are presented the aggregate fares for the passengers of Figure 5. The arbitrary line drawn with a positive slope of 1.00 is added to aid the reader's eye. If the reader covers the points from .5 through 10 on the ordinate, the line is not a bad fit. It certainly fits these points more closely than a line of slope 1.00 would fit the passenger data of Figure 5. Therefore we may suspect (along with everybody else) that in some way or other the railway fares of 1933 were out of line with bus fares, particularly in reference to the longer hauls. In this connection we point out that the passenger

[9] The slope of the Pareto income curve is sufficient to account for the divergence of slope of Figure 5. The same considerations should also apply to bus travel, although the lower income brackets are much, much more populous (at least two-thirds of the consumer units getting less than $2,000 during the years in question—and this sum is at the lower limit of the Pareto curve, cf. Zipf, G. K., *National Unity, op. cit.* p. 303f. for data and references).

traffic between Newark and New York as indicated in Figure 5 is quite in line, whereas the fares paid in Figure 6 between the same cities are well below the line; hence in this case, at least, there was a preferential fare, if we may assume the validity of our data.

This question of the general validity of our data for the specific question we are asking is serious. For example, as above mentioned, the various classes of railway passenger travel are combined so that we have no way of knowing which are Pullman passengers and which are coach passengers; which are extra fare passengers and which are taking advantage of excursion rates. Not only can these differences in rate-class result in an increased variability in the rate-distribution, as seems to be the case when we compare the rate-distribution of Figure 6 with the passenger-distribution of Figure 5. It can also induce a curvature in the line of a general downward concavity of a type which the reader may even choose to see already present in Figure 5; for if long distance through-travel carries with it extra fares, then, as $D$ increases on the whole *from right to left,* the values on the ordinate will fall ever more below the expected line, since the gross cost of distant travel will make an ever larger share of persons consider the economy of some alternate form of transportation.

The pooling of single-trip and round-trip fares in the aggregate revenues reported makes the underlying data difficult to manipulate. Some of the fares reported were doubtless for return-trips that had not yet occurred; as long as the data refer to tickets sold (and not to tickets collected by the conductor) an interpolation is risky. Although studies now in process show that the percentage of round-trip tickets tends to increase with $P_1 \cdot P_2 / D$, nevertheless the increase will scarcely be sufficient to explain the marked deviation of the slope of Figure 5 from the theoretically expected slope.

And so, as we look at the distributions of Figures 5 and 6 in comparison with those of Figures 3 and 4 (and even with Figure 1 and 2), and as we reflect upon the implications of our theoretical "$P_1 \cdot P_2 / D$ hypothesis," we can only conclude that in 1933 it was not as economical to travel per mile by rail over variable $D$-lengths as it was by bus. Hence as the railroads today strive ever more to give a better comparative service at a lower comparative cost for long and short hauls, their passenger-distributions as well as their rate-distributions should ever more approximate our theoretical expectations. And by the terms, *comparative* service and *comparative* cost, we mean not only in comparison with bus travel, but also in comparison with private automobile travel and with airway travel.

Of course our interests in the passenger data and their corollary rate data are purely theoretical, as we attempt to demonstrate the presence of a basic principle. It may well be that we shall never have completely satisfactory data for all person-movement by all means; indeed with governmental fiat in the setting of the rates of public carriers we may never have the prerequisite condition of "free" rates, even for the several classes of public carrier traffic. The data of other countries may be more instructive.

It may be that information on actual traffic is not essential. After all, the information about the places of origin of hotel guests as reported by hotel registers will also tell a lot, should the hotels care to make the information available.

### *C. Airway Traffic*

Rather for completeness than for decisive information do we present in Figure 7 the information for airway passengers during 1933 when air-travel was scarcely even in its infancy.

Yet in spite of the marked variation in the distribution of points in Figure 7, there is an unmistakable positive correlation between the number of passengers carried, and their corresponding values of $P_1 \cdot P_2 / D$.

What the distribution will look like a few years hence after air-travel has taken its place as a customary means of travel is an interesting question to ponder. Much will

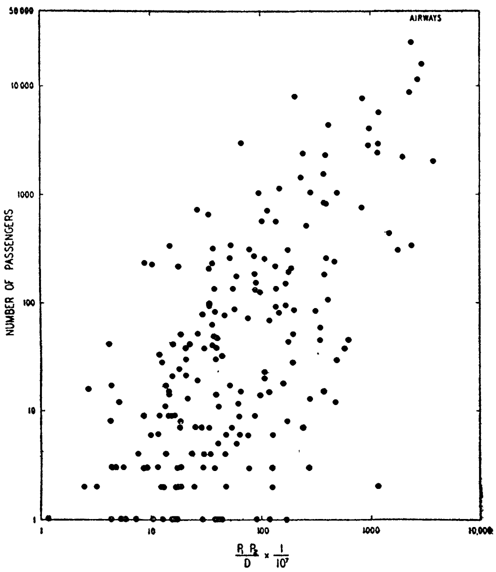


FIGURE 7. The number of passengers travelling by airway between arbitrary cities in 1933.

depend upon its rates and services in competition with those of other public carriers.

### 3. SUMMARY

In the present paper we have briefly set forth (1) the theoretical reasons for expecting that the inter-community movement of goods (by value) and of persons between any two communities, $P_1$ and $P_2$, that are separated by an easiest transportation-distance, $D$, will be directly proportionate to the product, $P_1 \times P_2$, and inversely proportionate to the distance, $D$.[10]

And we have also (2) presented data for the number of passengers and also for the amount of their aggregate fares (except for airways) for, *A*, Highway (public bus) travel, *B*, Railway-travel, and *C*, Airway-travel. The Highway data approximated our expected curve with considerable closeness. Though the Railway data revealed an unmistakable rectilinear correlation between our factors, the slope was greater than that anticipated theoretically; reasons were presented for the deviation of the slope for railway passengers during the depressed year, 1933. The Airway data also revealed a correlation even for the early year of its development, 1933, though the variation was considerable.

These data will be treated further and in greater detail in connection with other sets of data in the writer's forthcoming book, *The Principle of Least Effort.*

[10] In the light of our findings, we cannot agree as a general principle with our distinglished colleague, Dr. Samuel A. Stouffer who stated ["Intervening opportunities: a theory relating mobility to distance," *American Sociological Review,* 5 (1940), 846]: "the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities." On the other hand we yield to no one in our admiration of Dr. Stouffer's observations and of their applicability to intracity movement (to be discussed in a future publication).

# THE VOLUNTARY ASSOCIATIONS OF URBAN DWELLERS*

MIRRA KOMAROVSKY
*Barnard College*

THIS is a study of organized group affiliations of 2,223 adult residents of New York City. The study is focused on *class* differences but it also contains some data on sex, religious, and other factors in extent and pattern of participation.

* This study was done under a grant from the Columbia University Council for Research in the Social Sciences. The author wishes to express her gratitude to Professor Robert M. MacIver for his encouragement and guidance.

Studies of social participation published during the past decade or two dealt largely with rural areas and smaller communities.[1]

[1] *Social Participation Differences Among Tenure Classes* by C. Arnold Anderson, Bryce Ryan. Rural Soc., Vol. 8, No. 3, September, 1943; *Social Organizations in a Small City* by F. A. Bushee. *American Journal of Sociology,* Vol. 51, November, 1945; *Socio-Economic Circumstances about Adult Participation* by A. A. Kaplan (Ph.D.). Teachers