80

80

A Statistical Test of Significance Applied to A Sociological Problem: Variation in Accident Rates From Motor Vehicles

W. Edwards Deming
Bureau of the Budget and New York University

A. D. Battey
National Safety Council

Josephine D. Cunningham
New York State Commission Against Discrimination

Reprinted from American Sociological Review, Vol. 17, No. 6, December 1952

Reprinted from AMERICAN SOCIOLOGICAL REVIEW, Vol. 17, No. 6, December 1952

A STATISTICAL TEST OF SIGNIFICANCE APPLIED TO A SOCIOLOGICAL PROBLEM: VARIATION IN ACCIDENT RATES FROM MOTOR VEHICLES

W. EDWARDS DEMING

Bureau of the Budget and New York University

A. D. BATTEY

National Safety Council

JOSEPHINE D. CUNNINGHAM

New York State Commission Against Discrimination

Headlines in the Richmond *Times-Dispatch* for October 8, 1951 told that "Big United States Traffic Toll Is a Blot on Nation; State Shares Blame." The text was based on figures just released by the National Safety Council on deaths by motor-vehicle by state of accident, 1950. The national average was 7.5 deaths per 100 million vehicle-miles, and the figure for Virginia was 9.2. The *Times-Dispatch* suggested the usual reasons for the high rate in Virginia (speeding, drunk drivers, courts not tough enough, state highway patrol undermanned, etc.), and it outlined a program calculated to reduce the rate in the future.

Neither the text nor the comment was accompanied by any inquiry into the question of whether the rate for Virginia was actually *significantly* higher than the national average. There was no recognition of the fact that if there are to be any accidents at all, then without any *identifiable* cause whatever—i.e., purely by chance—the rates will vary from one area to another, and from one year to another; that some rates must be above the average, and some below.

The purpose of this paper is to illustrate how a simple statistical tool may assist in the discovery of the existence or the absence of identifiable causes of apparent differences between the behavior of the populations in different areas or classes. The discovery of differences and the causes of differences embrace a great deal of sociological research. The application here is to the causes of variation, from one area to another, in the rates of deaths from motor vehicles. The same technique has immediate application to other types of accidents, and to studies in the differences in birthrates, morbidity, and deaths from other causes.

More details on a possible application to insurance may help to clarify these proposals. Some insurance companies that write policies against damage done by an automobile divide their policy-holders into two classes—those who had an accident during the past year, and those who did not. Suppose that the company gives a reduced rate to those who had no accident last year. Are such people entitled to a reduction? Will the two classes be significantly different? That is, will the second class during the next year or five years have fewer accidents than the first class, and so merit the reduced rate? The answer is no; the separation is meaningless and ineffective except for (a) some possible psychological incentive helpful to drivers placed in Class 2 and discouraging to those placed in Class 1; and (b) the fact that it does place in Class 1 the drivers that have an average of several accidents per year. The two classes would probably otherwise have practically the same accident rates during the coming year or five years.

What is wrong? Statistically, a year offers too small a sample of exposure for most drivers. In a year's time only one in seven vehicles is involved in an accident of a type and severity deemed reportable to an insurance company. More time, more vehicle-miles, are necessary to effect a meaningful separation. Drivers with no reportable accident over a suitable period, possibly five years, possibly eight, could lay claim with some scientific justification to reduced rates. The length of a suitable period could be calculated by the techniques described here from the records of accidents reported. The same techniques would separate out another class of driver who has more accidents and is not entitled to insurance at the rates paid by "normal" drivers. Such separations would not be clean, but they would be economic, and in line with good business.[1]

[1] Most of the psychological benefits of reward for a clean record could be retained by reducing the premium by only a mere pittance for each successive clean year. At the end of so many clean years the policy-holder would definitely fall into one class or the other. Meanwhile an acci-

Conclusions. The statistical reasoning that is explained later on, combined with the results in Table 2, leads us to the following conclusions concerning deaths from motor vehicles:

(1) Efforts that have been expended to bring high death rates into line have apparently not yet been wholly successful, because there are many rates significantly high (H in Table 2) year after year.

(2) The reason may be found in diffused effort in trying to pin the reasons on to too many causes. One of the main aims of this paper is to suggest that *the identifiable and removable causes of significantly high rates can be discovered most effectively by looking for causes other than common causes* that run almost nation-wide.

(3) Studies of conditions in areas that have significantly low rates (those marked L in Table 2) might identify some of the causes of low rates. The lessons so learned might then be applied in areas marked H.

(4) In areas whose rates are not significantly high or low, efforts should be concentrated on common causes and not on attempts to discover causes identifiable with the particular area.

(5) Efforts to decrease the national average rates, urban and rural, should be concentrated principally on the removal of *common causes,* which will be causes not specifically identified with areas that are significantly high or low.

The Necessity for Classifying Rates as Significant or Not Significant. In the interest of effectiveness of effort, explanations of the variations of accident rates from one area to another, and consequent attempts to institute reforms for safety, should hinge upon the discovery of which rates represent departures *significantly different* from the average, and why. When a rate is significantly above or significantly below the average, an identifiable cause (explanation) other than chance variation should be presumed and sought. But when a rate is not significantly different from the average, it will not be profitable to look for an identifiable cause for the difference between (e.g.) Virginia and the national average—such differences occur too readily by chance, and it is better to study areas where there is indication of an identifiable cause. Moreover, much harm is done when, subsequent to a "drive" for safety, next year's rate goes below the average, not because of the drive, or in spite of it, but purely by chance variation. The decrease in the rate would then be wrongly attributed to the drive, instead of to coincidental chance variations—up one year, down the next. When a rate is not significantly high or low, efforts to find why it is merely above or below the national average will be misguided, misspent, misleading, and ineffective. Medical men know well the wrong done by coincidence of improvement and quack medicine.

The Mathematical Model for Chance Variations of Rates. Let black and white sand, thoroughly mixed, fall from a height on to a checkerboard. Suppose that there are 420,000 million grains of sand, 30,411 of which are black. Let 100 million grains of sand on the checkerboard constitute one unit; then there are, on the average, 7.24 black grains in every one unit of sand.

In 1949, vehicles travelled 420,000 million miles in the United States.[2] We assume that a grain of sand is a vehicle-mile, that a unit of sand is 100 million vehicle-miles, and that a black grain is a fatality. Then the annual exposure within any state is composed of some number of units of sand. A small state will contain a few units of sand, while a large state may contain many units. Not all the squares of the checkerboard will contain the same number of grains of sand, nor the same number of units, nor the same proportion of black grains. All this is well known from experience.

The Mathematical Definition of Significance. The experiment with the black and white sand will serve as the probability model. We shall test whether this mechanism is the right one. Significant differences will indicate departures from the model provided by the sand. With the aid of theoretical statistics we may compare the number of deaths in any state with the expected number of grains of black sand. First, the "expected" or mathematical average number of grains of black sand found in n units of sand (1 unit$=10^8$ grains of sand) that fall on the checkerboard is just 7.24 n. This proportion

dent would restore the rates of Class 1 and the chain would start again. In fact, the whole system might be placed under a continuous inspection plan, whereby a driver will stay in one class, or shift, depending on his record. Such a plan would take account of changes in eyesight and reflexes. New machinery for accounting now eliminates the excessive cost and time required for processing the records. The statistical techniques required have been developed by Harold F. Dodge and Miss Mary N. Torrey, "Continuous Sampling Inspection Plans," *Bell Telephone System, Technical Publications Monograph 1834* (1951).

[2] The U. S. totals used in this paper for 1948 and 1949 exclude the State of Rhode Island, which state failed to supply certain information. The exception of Rhode Island for the years 1948 and 1949 is to be understood in the text throughout. The U. S. total for all years excludes also the District of Columbia, which is entirely urban.

agrees with the original mixture, in which out of every 10^8 grains of sand, 7.24 were black.

Second, we do not expect to find exactly the same number of grains of black sand in every n units of sand on the checkerboard. Instead, in repeated experiments we observe variations above and below the average.

Third, experience shows that these variations will be described accurately enough by a distribution known as the Poisson series, by which, if m is the theoretical average or "expected" number of grains of black sand, the standard deviation of the observed number in repeated experiments will be the square root of m.

Fourth, we may choose 2 standard deviations (2-sigma) as the dividing line between significant and nonsignificant variations. It is better to err occasionally on the side of looking for causes of real differences when such causes do not actually exist, than to fail too often to look for causes of differences when the causes really do exist. Hence, we use 2-sigma limits, and not the more usual 3-sigma limits.

With this model (mechanism) we need only: set n for some particular state equal to the number of units of vehicle-miles in that state (1 unit $=10^8$ vehicle-miles); calculate the expected number of deaths $m=7.24\ n$; calculate the square root of m; inquire whether the recorded number of deaths D for that state falls inside or outside the 2-sigma interval. If D is below the lower limit the rate is significantly low. If D is above the upper limit the rate is significantly high. If D falls inside the interval the rate is not significantly above or below the national average.

Procedure of Calculation. For an illustration of these steps of calculation we may use the figures for Virginia for the year 1949. In that year there were 807 motor-vehicle traffic deaths, and 8,845,000,000 vehicle-miles of travel, with 9.1 deaths per 100 million vehicle-miles.

$n=88.45$ units. (This is the number of vehicle-miles for Virginia in 1949, expressed in units of 10^8.)

$m=7.24n=7.24\times 88.45=640$ expected number of deaths.

$\sqrt{m}=\sqrt{640}=25$ deaths, standard deviation.

$m+2\sqrt{m}=640+50=690.$

$m-2\sqrt{m}=640-50=590.$

The recorded number D of deaths was 807. This number falls well above the upper limit 690; consequently, the rate for Virginia was significantly high in 1949.

Separate Analysis for Urban and Rural Areas. The validity of this statement depends in part on how adequately the total of vehicle-miles measures the chance of death in a motor-vehicle accident. It is known, with fair accuracy, that nationally the mileage death rate for cities and towns is only half as high as the rate for rural areas. A highly urbanized state might therefore have a death rate that is significantly low, statewide, even though its urban rate is no lower than the rates in cities and towns of other states, and even though its rural rate is as high as the rural rates recorded elsewhere.

It seems desirable, then, to separate the urban areas from the rural areas. Unfortunately, this cannot be carried out with mileage rates. Some of the mileage estimates for urban areas probably contain sizable errors, although the extent cannot be measured. The estimates of rural mileage, while probably not free of error, are satisfactory for the present purpose.

An alternative measure of the urban exposure to motor-vehicle accidents is population. As more than half the urban accident deaths are of pedestrians, the use of the ratio of deaths to number of inhabitants in the urban areas seems justifiable. However, in interpreting the results (Table 2), it must be borne in mind that one factor in a significantly high rate may be extra heavy use of motor vehicles in the cities and towns of that state.

Application of the statistical method to the urban and rural death rates is the same as described above for state-wide rates. The national death rate for 1949 for accidents in cities and towns with more than 2,500 inhabitants was 9.41 per 100,000 people. In Virginia for 1949 the estimated population in places with more than 2,500 residents was 1,258,000. Now $9.41\times 1{,}258{,}000/100{,}000=118$, the number of expected deaths for urban Virginia. Moreover, the square root of 118 is 11, which is the standard deviation of the observed number in repeated experiments. The 2-sigma range is therefore 118 plus or minus 2×11, or 96 to 140. As the recorded total number of deaths for urban Virginia was 106, urban Virginia was well within this range, and was therefore not significantly different from the national urban average.

Similarly, the 1949 national death rate for accidents in rural areas and towns under 2,500 inhabitants was 10.37 per 10^8 rural vehicle-miles.[3] Virginia's rural vehicle mileage was

[3] One point should be noted concerning the computation of the rural expectancy of death. The U. S. rural rate was obtained by dividing deaths in rural areas *and* in towns of less than 2,500 population by the vehicle mileage given for rural areas only. The total of deaths recorded for the U. S. rural thus includes deaths from a greater area than is included in the rural mileage. The reason for this inconsistency is that the best available series of death records classifies towns of less than 2,500 population as rural areas, while

TABLE 1. SAMPLE CALCULATIONS FOR A FEW STATES, FOR 1949

(Source of the data for D, the National Office of Vital Statistics and state traffic authorities; for n, estimates of the National Safety Council based on data from the Bureau of Public Roads)

Area	D Number of Deaths	n*	m**	$\sqrt{m}$	$m \pm 2\sqrt{m}$	Significant?
U. S.***						
Total	30,411	4,200	30,411	..	....	....
Urban	8,046	855	8,046	..	....	....
Rural	22,365	2,156	22,365	..	....	....
Alabama						
Statewide	674	61.1	442	21	400–484	High
Urban	113	11.6	109	10½	88–130	No
Rural	561	32.4	336	18½	299–373	High
Utah						
Statewide	172	22.8	165	13	139–191	No
Urban	51	4.0	38	6	26– 50	High
Rural	121	12.4	129	11½	106–152	No
Washington						
Statewide	444	76.0	550	23½	503–597	Low
Urban	135	12.3	116	11	94–138	No
Rural	309	36.4	377	19½	338–416	Low

* For the U. S. total, and for rural and statewide areas, n denotes vehicle-miles in units of 10^8. For urban areas, n denotes population in units of 10^5.

** The national death rates used in computing m, the "expected" number of state deaths are: 7.24 for the U. S. total, 9.41 for U. S. urban, and 10.37 for U. S. rural.

*** Excluding Rhode Island which state failed to report sufficient information, and the District of Columbia which is an entirely urban area.

estimated at approximately 64.7×10^8, so the expected number of deaths was $10.37 \times 64.7 \times 10^8/10^8$, or 671, and the standard deviation is here the square root of 671 or 26. The 2-sigma range is 671 plus or minus 2×26, or 619 to 723. With a recorded death total of 701, the rural experience—like the urban—is within the range, and is not significantly different from the national rural average.

Table 1 summarizes the computations for 1949 for three other states as additional illustrations of the method.

We had already noted that if the 1949 Virginia experience is considered in total, the deaths appear to be significantly more numerous than would be expected on the basis of the national experience. This conclusion is changed by separate consideration of the urban and rural death records, for we have just seen that the state's urban and rural experience did not differ significantly from that of the nation's urban and rural averages.

Conversely, Ohio, the statewide experience in 1947, 1948 and 1949 did not differ significantly from the respective national averages for urban and rural areas combined, yet Ohio's urban rate in all three years showed deaths above the national urban average.

Interpretation of the Table of Significance, Table 2. This table shows whether the death rates—statewide, urban, and rural—were significantly above or below the national average, for the years 1946, 1947, 1948, and 1949. It shows that practically all states retain their

---

the only estimates available on mileage classifies them as urban places. The rate obtained is too high to represent the rural experience alone, but as its only purpose is to estimate the number of deaths in rural areas and towns of less than 2,500 population for each state, the question of its accuracy as a rural rate does not arise. Here we only need estimates that will be useful in the problem of determining whether a state's rural area is significantly different from the national rural average.

One possible defect is that in towns of less than 2,500 inhabitants, the ratio of population to rural mileage is not the same in all states. It seems reasonable to say that in states where the ratio of population to mileage is greatest, the recorded deaths should exceed the expected deaths. However, this is not indicated by the data. In Iowa and Minnesota this ratio was twice the national average; moreover, in these states the populations in towns with less than 2,500 inhabitants are absolutely as well as relatively large (463,000 in Iowa and 402,000 in Minnesota according to the 1950 Census); yet as Table 2 shows, in the rural parts of these states the recorded deaths were significantly low in all four years covered by this study. Their experience thus indicates that the method that we have used here for obtaining the expected death totals does not of itself always raise the recorded deaths above expectancy.

TABLE 2. SIGNIFICANCE OF TRAFFIC DEATHS FROM MOTOR VEHICLES, STATEWIDE, URBAN, AND RURAL, 1946 TO 1949

(H denotes significantly high, L denotes significantly low, — not significant, in comparison with the national average)

States	1946			1947			1948			1949		
	State-wide	Urban	Rural	State-wide	Urban	Rural	State-wide	Urban	Rural	State-wide	Urban	Rural
Alabama	H	—	H	H	—	H	H	H	H	H	—	H
Arizona	H	—	—	H	H	—	H	—	H	H	—	—
Arkansas	H	—	—	H	—	—	H	—	—	H	—	—
California	H	H	H	H	H	H	—	H	H	H	H	H
Colorado	—	—	—	—	—	—	—	—	—	—	—	—
Connecticut	L	L	L	L	L	L	L	L	L	L	L	L
Delaware	—	—	L	—	—	L	—	—	L	—	—	—
Florida	H	H	H	H	H	H	H	H	H	H	—	H
Georgia	H	H	—	H	—	—	H	—	H	H	—	H
Idaho	—	—	—	—	H	—	H	H	—	—	—	L
Illinois	—	H	L	H	H	—	H	H	—	—	H	H
Indiana	—	H	—	H	H	H	H	H	H	H	H	—
Iowa	L	—	L	L	—	L	L	—	L	L	—	L
Kansas	L	L	L	L	L	L	—	L	L	—	—	—
Kentucky	H	—	H	H	—	H	H	—	H	H	H	H
Louisiana	H	—	H	H	—	H	H	L	H	H	—	H
Maine	L	L	—	L	—	L	—	L	—	L	—	L
Maryland	—	L	L	—	—	—	—	—	L	—	—	—
Massachusetts	L	L	L	L	L	L	L	L	L	L	L	L
Michigan	L	—	—	L	—	—	—	H	—	—	—	—
Minnesota	L	L	L	L	—	L	L	—	L	L	—	L
Mississippi	H	—	H	H	—	H	H	—	H	H	—	H
Missouri	L	—	H	L	L	H	L	L	H	L	L	H
Montana	—	—	—	—	L	—	—	—	—	—	—	—
Nebraska	L	L	L	—	—	—	L	—	L	L	—	L
Nevada	H	—	—	Data incomplete			H	—	—	H	—	—
New Hampshire	—	L	—	—	—	—	L	—	L	L	—	L
New Jersey	L	L	L	L	L	L	L	L	L	L	L	L
New Mexico	H	—	H	H	—	H	H	—	H	H	—	H
New York	L	L	L	L	L	L	L	L	L	L	L	L
North Carolina	H	—	H	H	L	H	H	L	H	H	—	H
North Dakota	—	L	—	L	—	—	—	—	—	—	L	L
Ohio	L	H	—	—	H	—	—	H	—	—	H	—
Oklahoma	—	L	L	—	L	L	—	—	L	—	L	L
Oregon	H	H	—	—	—	—	—	H	L	L	—	L
Pennsylvania	L	L	H	L	L	H	L	L	H	L	L	H
Rhode Island	L	L	L	L	L	L	No data			No data		
South Carolina	H	—	H	H	—	H	H	—	H	H	—	H
South Dakota	—	—	—	—	—	—	L	—	L	L	L	L
Tennessee	H	—	H	H	—	H	H	H	H	H	H	H
Texas	—	—	L	—	—	L	H	—	—	—	—	—
Utah	—	H	—	—	—	—	H	—	H	—	H	—
Vermont	—	—	L	L	—	L	—	—	—	L	L	L
Virginia	H	—	H	H	L	—	H	L	—	H	—	—
Washington	—	H	—	L	H	—	L	H	L	L	—	L
West Virginia	H	—	—	H	—	—	H	—	H	H	—	—
Wisconsin	L	L	L	—	—	—	—	—	—	—	L	—
Wyoming	H	—	—	—	—	—	H	—	—	H	—	—

respective positions through the four years. That is, a state that has a significantly high rate, either in total or for urban or rural areas, in one year retains it in other years, and a state that has a significantly low rate in one year retains it in other years. In other words, a pattern set in any one year holds in most cases through other years, before and after.

A significantly high or low rate indicates that there is an identifiable cause other than chance variation. An identifiable cause of a high rate might be roadways too narrow (especially on mountain curves), or high speeds on flat plains. An identifiable cause, present one year, will usually continue to act until it is removed. Hence, a state that has a significantly high rate in any one year will usually continue to have a high rate until definite steps are taken to identify and remove the cause. Similarly, a state that has a significantly low rate will continue to have a low rate unless the identifiable causes change. On the other hand, in a state that is so close to the average that its difference from the average is not significant, it will be difficult or even hazardous to try to identify causes for the difference. Effort should there be concentrated on *common causes.*

For such reasons the conclusions already listed at the beginning of this paper seem to follow.

APPENDIX: SOURCES OF THE DATA

There are six series of state death totals from motor-vehicle accidents, prepared by state traffic authorities, vital statistics divisions of state health departments, and the National Office of Vital Statistics.

The records bureaus in the offices of state traffic authorities compile detailed information by place of accident on deaths that result from accidents on traffic ways involving motor vehicles. In some states a record is also kept of deaths in motor-vehicle accidents occurring off traffic ways, chiefly home driveway accidents, but these are recorded separately from the traffic deaths. These totals of traffic deaths would be the best series for use in the study if the division between urban and rural accidents were consistent from state to state. However, for most states "urban" includes all incorporated places regardless of size, plus the larger unincorporated places; but for some states "urban" means places of 1,000 or more inhabitants, for others it means places of 2,500 or more, and for a few it includes only places of 5,000 or more.

State health departments compile information on deaths from motor-vehicle accidents according to place of death, and in many states also by place of residence of deceased. If only statewide data were needed, the totals by place of death would be useful, but an urban-rural division by place of death is misleading. Many persons injured in rural accidents are taken to urban hospitals before they die. The urban record is thus loaded with deaths that are a part of the rural accident experience. The totals of deaths classified according to place of residence of deceased are even less useful for this study, for they include as rural the deaths of residents of rural areas in urban accidents, and as urban the deaths of urban residents in rural accidents.

The National Office of Vital Statistics, for the years with which this study deals, has published three series of state motor-vehicle death totals. Two of them have the same bases as the data from state health departments—place of death and place of residence of the deceased—and are therefore of limited usefulness for this particular study. The third series, however, is based on place of accident, and includes an urban-rural division that is the same for all states. This is the series finally selected for our calculations as the best available. It required some minor adjustment, however, for discrepancies in reporting.

A comparison of the statewide totals with those prepared by state traffic authorities showed that some states apparently missed, in their reports to the N.O.V.S., some of the traffic deaths. This could happen through failure to sort out all the deaths from motor vehicles, from the death certificates from all causes. Occasional differences of opinion on reportability or on the classification of a death as traffic or non-traffic may be expected, and if all the differences were small they could be disregarded. However, the totals reported by the National Office of Vital Statistics for nine states were more than five per cent below the totals reported by the state traffic authorities. For these states we used the figures from the state traffic authorities if we could make an urban-rural division at places of 2,500 or more inhabitants; otherwise we used the urban and rural totals from the National Office of Vital Statistics, both increased proportionately to equal in total the state traffic authority's total of deaths. The errors involved in this procedure are undoubtedly small compared with the 2-sigma limits that determine significant variations.

The population estimates that we used were straight-line interpolations for July 1 of each year between the Census populations of 1940 and 1950. It is unlikely that estimates for these years made by more complicated procedures would yield results materially different.

The state totals of rural vehicle mileage were derived from estimates of the consumption of gasoline and from trends of rural traffic compiled from traffic counters placed by the Bureau of Public Roads. The estimates have a firm basis in the detailed information obtained in the state highway planning surveys conducted in the late 1930's, while the trends in traffic volume (including data for passenger cars and for trucks of different load-capacities), and information on the estimated gasoline consumption, make possible reasonable approximations of rural vehicle-miles for recent years.

The mileage estimates do not have the precision of the death totals, but the errors are not big enough to produce changes in the conclusions from Table 2. Details of the computations are not given in this article, but it may be of interest to the reader that in 70 per cent of the entries in Table 2 in which an area is classified H or L, the recorded number of deaths exceeded the limits of significance by 10 per cent or more. In these cases only a sizable change in the estimate of mileage could shift a state to another classification of significance (H to not significant, etc.).