Accid. Anal. & Prev. Vol. 17, No. 2, pp. 87–100, 1985
Printed in the U.S.A.

0001–4575/85 $3.00 + .00
Pergamon Press Ltd.

GRAPHICAL TOOL FOR COMPARING RATIOS PROPOSED IN THE LITERATURE ON TRAFFIC ACCIDENTS†

ANITA PARUNAK and ANDREW L. BAUGHMAN
The University of Michigan, Transportation Research Institute,
Ann Arbor, MI 48109, U.S.A.

(*Received* 20 *August* 1981; *in revised form* 31 *January* 1984)

Abstract—Investigators of accidents use accident rates and ratios to measure risk of being involved in a traffic accident. This paper uses two examples to present a graphical technique, that summmarizes scatters of points by ellipses, to help describe the accident involvement of drivers. The first appendix tells how the technique works and the second appendix summarizes other types of ratios used in the literature. The paper discusses how the technique may tie together ratios in the literature on accidents and ageing of drivers.

INTRODUCTION

Analysts use different criteria to measure risk of being involved in a traffic accident. Each investigation derives an accident rate or ratio and compares the rate among drivers of several subgroups. The analysts may arrive at different answers to the same research question, depending on which criterion is used. We present a graphical technique, summarizing scatters of points by ellipses, to help to unify results of various ratios and to show clearly outlying points and discrepancies in the data. See Stoto [1980] and Tukey and Parunak [1984].

Two examples are used to present the graphical technique for displaying ratios. The first example shows fatal accidents by state and age, with emphasis on the ageing driver. The second example is of fatal accidents to drivers of combination vehicles by year and state. Appendix A gives a step-by-step explanation of how to construct the ellipses. Appendix B is a review of the literature on ageing and accidents and summarizes the ratios used. Computer programs are also available in Fortran (Stoto [1980] and Baughman and Parunak [1983]) and, for a microcomputer implementation, in Pascal [O'Day, 1983] and in Basic [Baughman, 1984].

EXAMPLES TO PRESENT GRAPHICAL TECHNIQUE

Example 1. *Fatal accidents by age and state.*

Our first example illustrates the technique with data from Finesilver [1969]. Panels A through C of Table 1 give the total number of accidents and number of accidents fatal only to drivers for 26 states in 1967, for age groups 25–34, 35–44, and 65 and over. We limit ourselves to these three age groups because they highlight a progression that continues through other age groups, without unnecessarily lengthening the example. The third column in each panel gives the population in the states based on the 1970 Census. The ratios in the last two columns are y = (number of fatal accidents)/(total number of accidents) and x = (total number of accidents)/(population in age group).

These long columns of ratios are relatively silent about differences in accidents among age groups. The columns are even more silent about differences in the results among several ratios and, at the same time, among age groups. To help detect and explain these differences, we refer to a method of constructing isoquants and ellipses (see Stoto [1980]; or Tukey and Parunak [1986]). The isoquants and ellipses provide at a glance comparisons among age groups and among ratios.

Figure 1, Panels A through C, plots y against x for the three age groups. At the same time, they display a third ratio, z = (number of fatal accidents)/(population in age group), the product of x and y. A set of hyperbolic contour lines, or isoquants, references the ratio z. Each contour

---

†This research was supported in part by a grant from the Hazel I. Stoll estate to the Transportation Research Institute, The University of Michigan, and in part by private funds from H. Van Dyke Parunak.

Table 1. Data for drivers in traffic accidents by state in 26 states, 1967

A for ages 25-34

State	# Acc.	# Fatal	# Pop.	y = # Fatal / # Acc	x = # Acc / # Pop.
Alaska	2,483	11	49,299	.0044	.0504
Arizona	1,960	58	216,843	.0296	.0090
Colorado	21,701	133	291,204	.0061	.0745
Delaware	3,803	44	70,440	.0116	.0540
Dist.of Columbia	10,920	50	112,525	.0046	.0970
Illinois	123,149	645	1,383,864	.0052	.0890
Indiana	54,226	395	639,471	.0073	.0848
Iowa	17,518	173	314,768	.0099	.0557
Kentucky	22,748	236	380,990	.0104	.0597
Maine	7,288	64	109,710	.0088	.0664
Maryland	29,511	238	524,430	.0081	.0563
Michigan	95,353	576	1,082,363	.0060	.0881
Minnesota	33,219	219	456,285	.0066	.0728
Missouri	9,384	259	450,324	.0276	.0208
Montana	4,497	69	79,879	.0153	.0563
New Jersey	61,525	327	866,639	.0053	.0710
North Carolina	36,609	464	643,906	.0127	.0569
North Dakota	1,809	41	65,452	.0227	.0276
Ohio	60,572	692	1,291,822	.0114	.0469
Oklahoma	17,155	212	292,468	.0124	.0587
Oregon	21,801	143	254,577	.0066	.0856
South Carolina	19,653	215	320,245	.0109	.0614
South Dakota	2,989	46	67,072	.0154	.0446
Utah	8,790	72	129,961	.0082	.0676
Virginia	38,171	301	615,447	.0079	.0620
Wisconsin	29,527	249	506,447	.0084	.0583

B. for ages 35-44

State	# Acc.	# Fatal	# Pop.	y = # Fatal / # Acc	x = # Acc / # Pop.
Alaska	2,036	10	38,021	.0049	.0535
Arizona	1,531	61	195,323	.0398	.0078
Colorado	17,047	95	255,624	.0056	.0667
Delaware	3,111	20	65,532	.0064	.0475
Dist.of Columbia	7,638	29	84,974	.0038	.0899
Illinois	103,474	502	1,168,932	.0049	.0885
Indiana	45,564	336	582,399	.0074	.0782
Iowa	15,451	130	296,896	.0084	.0521
Kentucky	16,668	173	350,340	.0104	.0476
Maine	6,299	77	109,027	.0122	.0578
Maryland	24,965	175	471,887	.0070	.0529
Michigan	85,471	416	1,002,322	.0049	.0853
Minnesota	26,575	212	395,573	.0080	.0672
Missouri	5,064	150	509,814	.0296	.0099
Montana	3,714	58	74,998	.0156	.0495
New Jersey	59,079	245	879,421	.0041	.0672
North Carolina	27,146	318	589,087	.0117	.0461
North Dakota	1,493	25	64,198	.0167	.0233
Ohio	53,240	530	1,222,941	.0100	.0435

Table 1. *(Continued)*

B. for ages 35-44 State	# Acc.[1]	# Fatal[1]	# Pop.[2]	$y = \frac{\#\ \text{Fatal}}{\#\ \text{Acc}}$	$x = \frac{\#\ \text{Acc}}{\#\ \text{Pop}}$
Oklahoma	14,395	170	283,055	.0118	.0509
Oregon	19,899	118	225,782	.0059	.0881
South Carolina	14,733	172	286,000	.0117	.0515
South Dakota	2,557	31	68,821	.0121	.0372
Utah	6,864	37	107,548	.0054	.0638
Virginia	30,406	223	549,962	.0073	.0553
Wisconsin	23,731	162	470,246	.0068	.0505
C. for ages 65 and over					
Alaska	140	2	6,887	.0143	.0203
Arizona	569	26	161,474	.0457	.0035
Colorado	5,348	49	187,891	.0092	.0285
Delaware	856	13	43,833	.0152	.0195
Dist.of Columbia	1,201	5	70,803	.0042	.0170
Illinois	28,068	248	1,093,654	.0088	.0257
Indiana	15,199	155	493,809	.0102	.0308
Iowa	5,076	100	350,293	.0197	.0145
Kentucky	5,282	82	337,428	.0155	.0157
Maine	2,730	39	114,592	.0143	.0238
Maryland	4,581	35	299,682	.0076	.0153
Michigan	24,657	219	752,955	.0089	.0327
Minnesota	8,936	92	408,919	.0103	.0219
Missouri	3,154	129	560,656	.0409	.0056
Montana	1,446	24	68,736	.0166	.0210
New Jersey	13,562	83	696,989	.0061	.0195
North Carolina	6,943	117	414,120	.0169	.0168
North Dakota	848	21	66,368	.0248	.0128
Ohio	15,666	288	997,694	.0184	.0157
Oklahoma	6,966	98	299,756	.0141	.0232
Oregon	8,544	67	226,799	.0078	.0377
South Carolina	3,441	62	190,960	.0180	.0180
South Dakota	1,359	21	80,484	.0155	.0169
Utah	2,466	30	77,561	.0122	.0318
Virginia	7,056	71	366,021	.0101	.0193
Wisconsin	8,931	100	472,865	.0112	.0189

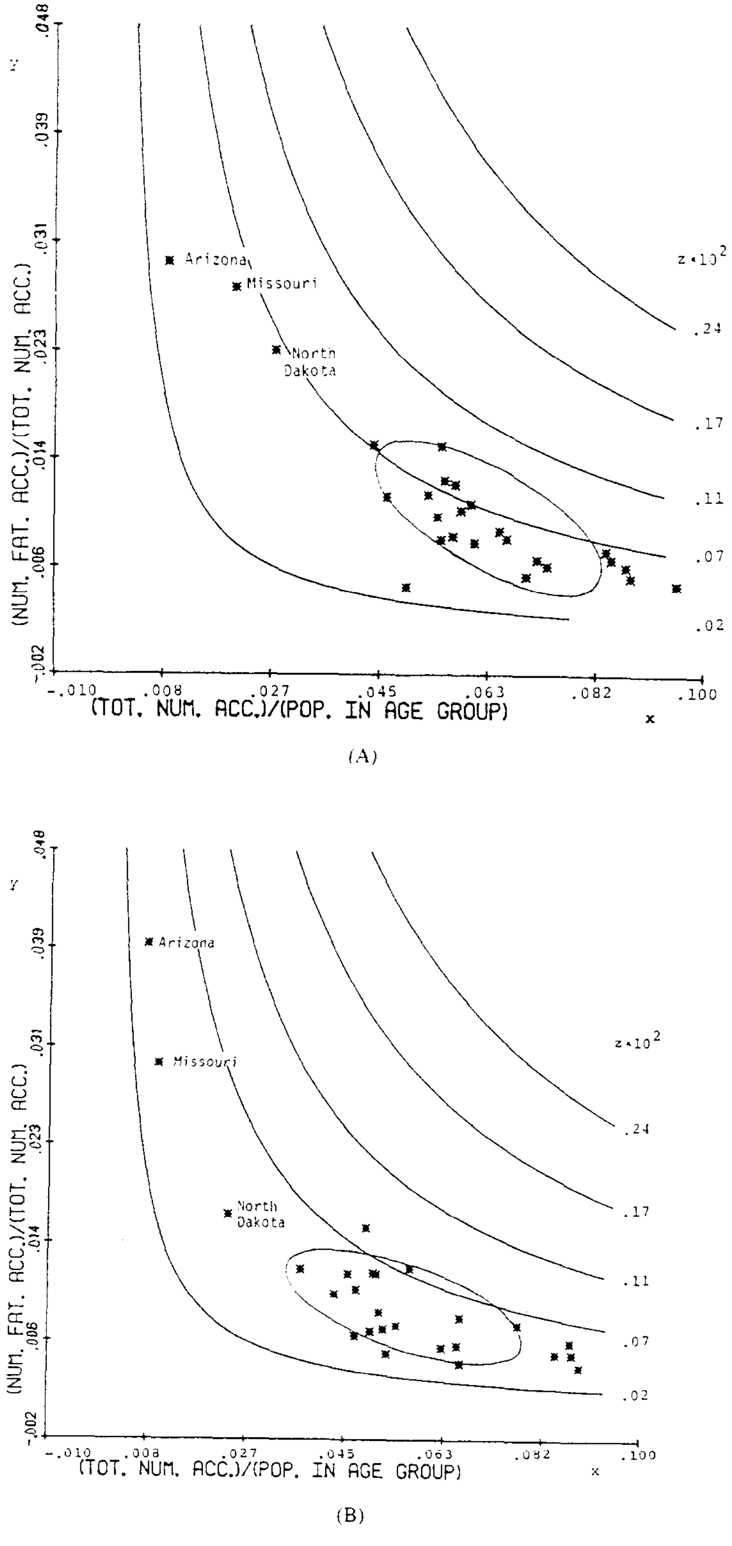
[1] Extracted from Finesilver [1969] [Note: # Acc. and # Fatal are 1966 figures for Maryland, Missouri, New Jersey, and Oregon].

[2] 1970 Census of Population: Volume 1 Characteristics of the Population, Part 1 United States Summary Section 1, Table 62.

line is the locus of points with the same value of z. The labels for the isoquants are on the right side of the graph.

We construct an ellipse around the scatter of points, to cover about three-fourths of the points in each panel. Appendix A presents the mechanics of the construction. The ellipse focuses attention on the central cluster of points.

So two steps are involved in the graphical displays. One step is the construction of isoquants. They enable the display of a third ratio, along with the first two ratios. The other step is the construction of ellipses, to outline the central cluster of points.

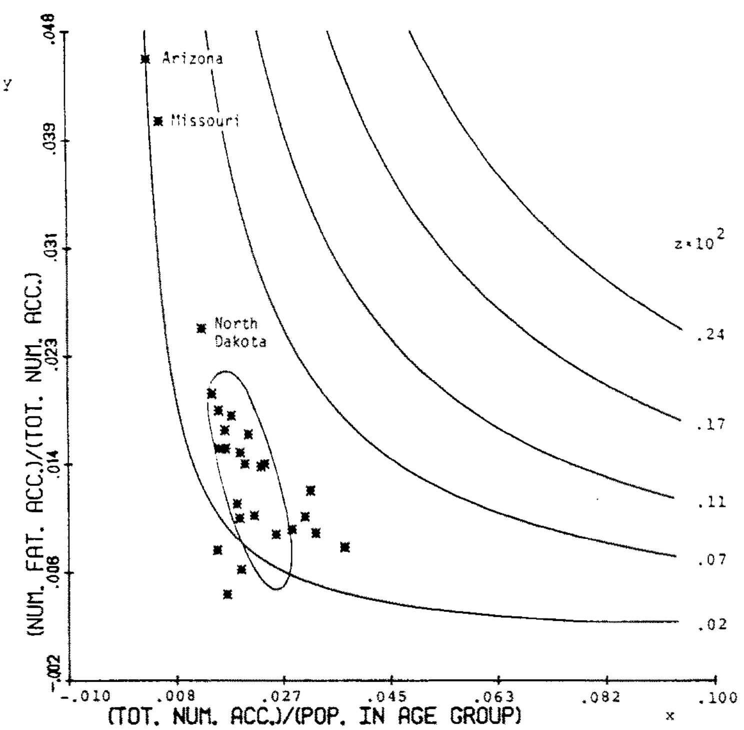


(A)

(B)

Fig. 1. Scatter plot by y = (number of fatal accidents)/(total number of accidents) against x = (total number of accidents)/(population in age group) in 26 states in 1967. The contour lines are $z = y \times x$ = (number of fatal accidents)/(population in age group). (A) For ages 25–34. (B) For ages 35–44. (C) For ages 65 and over.

In addition to rendering broad patterns in the data, the plots readily identify outliers. For example, we can identify three points as outliers in the y dimension in Panel A. Their labels are Arizona, Missouri and North Dakota. When the same three states are distinct outliers in



(C)

Fig. 1. Continued.

the other two panels, we wonder if the methods of collecting or reporting data are different there than in the other states. If we were viewing only one dimension, for example the x dimension in Panel C, we would not be able to label Arizona, Missouri and North Dakota as outliers. The added dimension shows their deviation clearly.

Now that the ellipses have delineated the central points and outliers, we let the ellipses alone show the progression of the age groups. Figure 2 displays the three ellipses together. All the points are omitted, so the progression of the ellipses without the distraction of many dots is visible.

The slopes of the ellipses for ages 25–34 and 35–44 are almost identical. The spread and level in all three dimensions is about the same for both age groups. The ellipse for the drivers 65 and over is distinct from the other two ellipses. The oldest drivers fare worst in terms of the y ratio, (number of fatal accidents)/(total number of accidents). They are decidedly better than those in the other two age groups for the x ratio. They are a little better in terms of the z ratio. Perhaps their outcomes result from their being more careful drivers but more fragile when they have an accident.

These conclusions are based on the progression of the ellipses. They may point toward a slight increase in safety on the highways, because the number of older drivers is increasing somewhat. The number of drivers over 65 years old increased from 5.4 million in 1958 to 13.7 million in 1980. Also the proportion of drivers in this age group to all drivers increased from 6.7% to 9.4% between these years.

Next, we briefly compare our ratios and results with the ratios and results from other studies in the literature. We then point out how the use of an isoquant plot with ellipses can sharpen the conclusion of previous, and future, studies.

Comparison with other studies.

We reviewed the literature on accidents and the age of the drivers to find out what ratios other studies used and how they answered the question to whether drivers over 65 are more likely than drivers in other age groups to be involved in a traffic accident.

Appendix B provides a comprehensive bibliography to this literature. It classifies the ratios and presents them in the classified format. We used both a computerized search, MEDLARSII

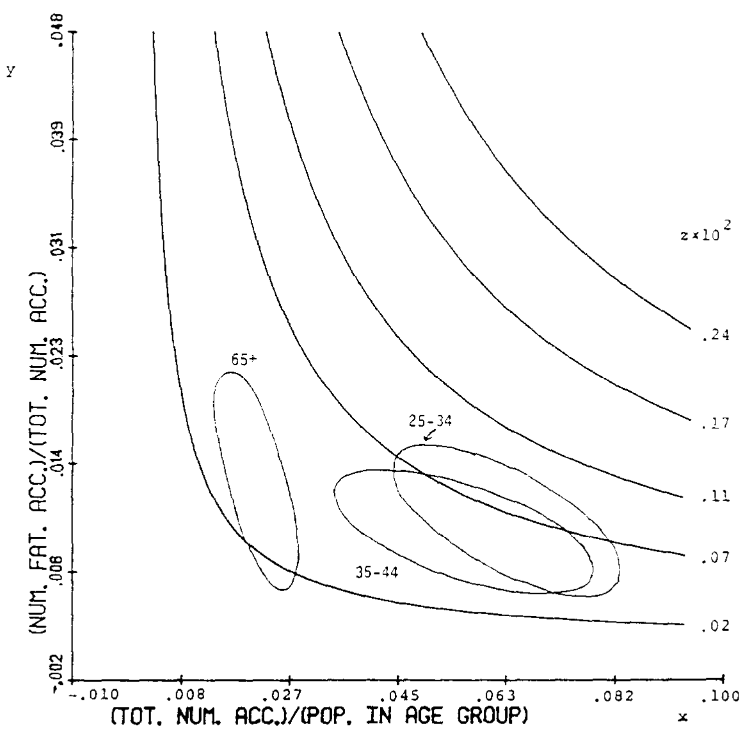


Fig. 2. Ellipses for three age groups from Fig. 1 for 26 states in 1967. y = (number of fatal accidents)/(number of accidents), x = (number of accidents)/(population in age group), z = (number of fatal accidents)/(population in age group).

data base of the National Interactive Retrieval Service, through the library of The University of Michigan Medical School, as well as more conventional library resources. We concentrated on articles about ageing and accidents that included statistics or data analysis.

Analysts use different criteria to measure risk of being involved in a traffic accident. Each investigation derives an accident rate or ratio and compares this rate among drivers of several age groups. The analysts arrive at different answers, depending upon which criterion is used. Some analysts believe drivers over 65 have fewer accidents than drivers of other age groups. Other analysts prove exactly the converse, and still others land somewhere in between.

Application of ellipses to other studies.

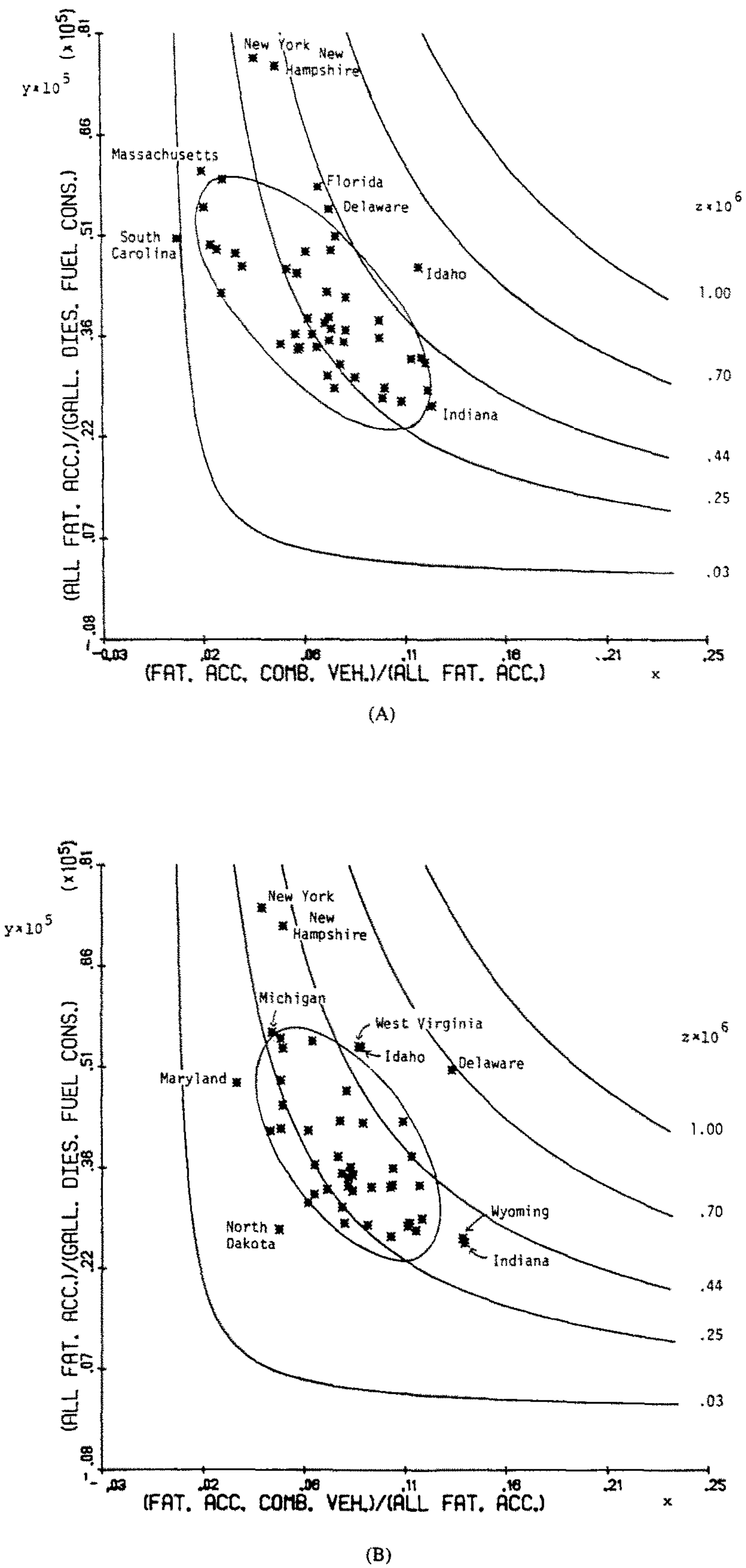
The diversity of the conclusions points to the need for further data analysis. Rather than itemize the conclusions, article by article, we have chosen to illustrate how the isoquant plot with ellipses can resolve ambiguities within studies and aid comparisons among studies. We mention two applications of the ellipses to previous studies that would have sharpened their results considerably. The data from the other articles are not sufficient to enable us to construct ellipses and thereby compare their results with ours. However, we feel that the graphical techniques we present are extremely important to be incorporated into future planning and studies.

McFarland and O'Doherty [1952] present the relation between driver age and automobile accidents using two different graphs. One is a plot of the ratio of number of drivers in accidents to the number of licensed drivers versus driver age. The second graph plots the percent of drivers involved in accidents judged "at fault" versus driver age.

The McFarland and O'Doherty graphs bring to our attention an application of the isoquant plots. We could look at the two ratios from their plots plus a third ratio, all on one graph. We would use the relation:

$$\frac{\text{\# Drivers in Acc.}}{\text{\# Licensed Drivers}} \times \frac{\text{\# Drivers at Fault}}{\text{\# Drivers in Acc.}} = \frac{\text{\# Drivers at Fault}}{\text{\# Licensed Drivers}} \quad (y \times x = z)$$

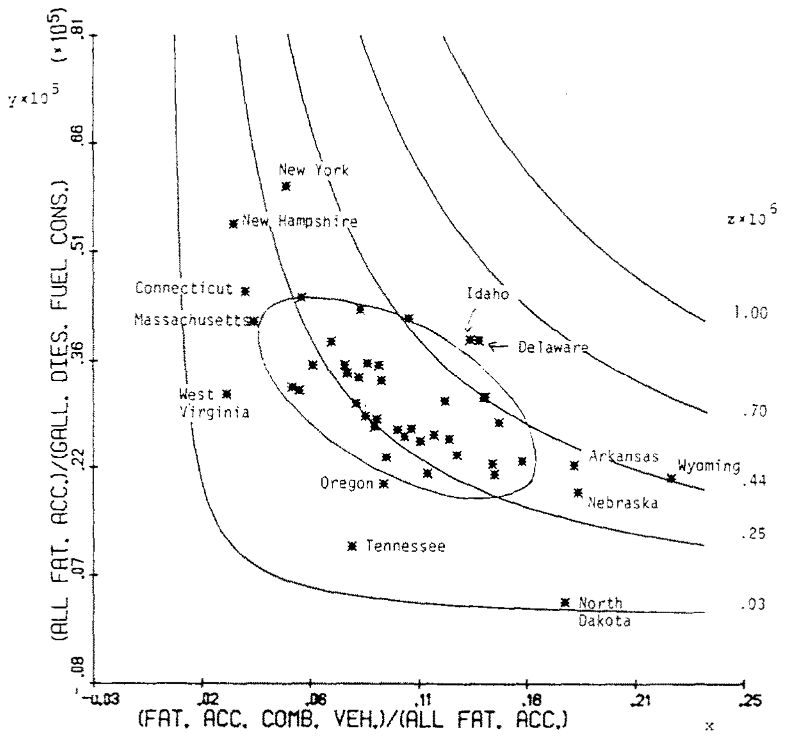
Points could represent age groups and successive plots various states, or vice versa.



(A)

(B)

Fig. 3. Scatter plot of y = (all fatal accidents)/(gallons of diesel fuel consumed) against x = (fatal accidents involving drivers of combination vehicles)/(all fatal accidents) in 48 states for all age groups combined. The contour lines are $z = y \times x$ = (fatal accidents involving drivers of combination vehicles)/(gallons of diesel fuel consumed). (A) For 1975. (B) For 1976. (C) For 1979.



(C)

Fig. 3. Continued.

For an article by Siebrecht *et al.* [1959] we note another potential use of the isoquant plots. Their index is:

$$\text{Accident-Violation Index} = (\text{Accident Ratio} - 1/2\ \text{Violation Ratio}) \times 10$$

where

$$\text{Accident Ratio} = \frac{\text{Number of Recorded Accidents}}{\text{Annual Mileage} \times \text{Years of Driving}} \times 10^{-5}$$

$$\text{Violation Ratio} = \frac{\text{Number of Recorded Violations}}{\text{Annual Mileage} \times \text{Years of Driving}} \times 10^{-5}$$

The isoquant plots can present both the accident and violation ratios, as well as a third ratio (number of recorded accidents)/(number of recorded violations). The necessary equation is

$$\frac{\text{Number of Violations}}{\text{Annual Mileage} \times \text{Years of Driving}} \times \frac{\text{Number of Accidents}}{\text{Number of Violations}}$$
$$= \frac{\text{Number of Accidents}}{\text{Annual Mileage} \times \text{Years of Driving}} \quad (y \times x = z)$$

From the graph we can then ascertain the mathematical relation among the ratios. We can at least confirm to ourselves that the data support the linear combination: $z - (1/2)y$ that the authors employ.

Example 2. Fatal accidents to drivers of combination vehicles.

The comments on the McFarland and O'Doherty and the Seibrecht *et al.* articles show that

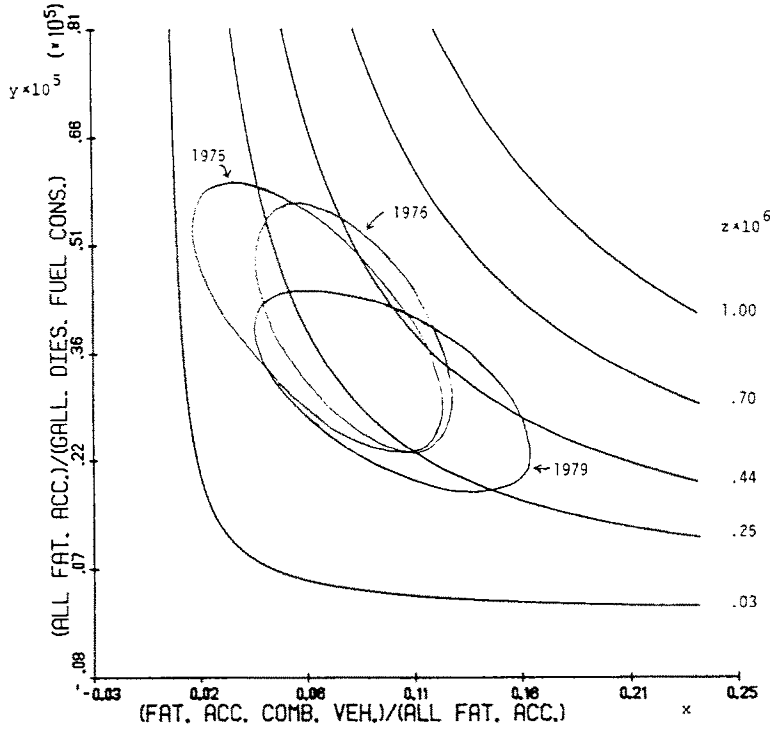


Fig. 4. Ellipses for three years from Fig. 3 for all age groups combined in 48 states. y = (all fatal accidents)/(gallons of diesel fuel consumed), x = (fatal accidents involving drivers of combination vehicles)/(all fatal accidents), z = (fatal accidents involving drivers of combination vehicles)/(gallons of diesel fuel consumed).

only may formulate numerous combinations of ratios for accidents. The model is

$$y = b/c, \quad x = a/b, \quad z = a/c \qquad y \times x = z.$$

where b and c are variables representing measures of driving exposure such as driving distance or number of licensed drivers.

By exploring different formulations, analysts do not have to rely on only one accident ratio to make conclusions. Our second example emphasizes this idea. O'Day, *et al.* [1980] present data for one accident ratio, labelled x = (fatal accidents involving drivers of combination vehicles)/(all fatal accidents). This index gives two of the three parameters in the above model, a and b, needed to apply the technique. They suggest another measure of driving exposure, gallons of diesel fuel consumed, for each state and year from another data source. With the third parameter, c, we formulate two new accident ratios, y = (all fatal accidents)/(gallons of diesel fuel consumed) and z = (fatal accidents involving drivers of combination vehicles)/(gallons of diesel fuel consumed). We plot the x and y ratios and view the z ratio in Panel A of Fig. 3. We have three ratios to look at simultaneously, and thus more information to characterize the accident history of combination vehicles.

Panels A through C of Fig. 3 show three of five ellipses for the accident experience of combination vehicles (tractor-trailers) from 1975–79 in 48 states. In the first example an ellipse was for an age group. Here, each ellipse represents a year. Again, we picture only three years (1975, 1976, and 1979), because they show a trend that continues through other years, without lengthening the example.

As in our first example, the ellipses tell about the behavior of the three accident ratios, x, y, and z, and they point out differences among states and peculiar trends that may otherwise go unnoticed. In Panels A through C of Fig. 3, states falling outside the ellipses are identified. New York and New Hampshire are the only states that fall outside the ellipse for all five years.

These states consistently exhibit high values for y. Indiana and Wyoming fall outside four of the five ellipses (data not shown). They consistently exhibit large values for x. A few more states fall outside the ellipses for more than one year. These outliers are in roughly the same position relative to the other points in each panel.

We summarize changes in these accident ratios over the five-year period with the following observations based on Fig. 4:

(1) The ratio of number of fatal accidents involving drivers of combination vehicles to total number of fatal accidents (x ratio) increases;

(2) The y ratio, total number of fatal accidents to gallons of diesel fuel consumed, decreases; and

(3) The z ratio, number of fatal accidents involving drivers of combination vehicles to gallons of diesel fuel consumed, remains about the same.

Summary of graphical tool.

There are two parts to the graphical tool. One part is an isoquant plot, based on ratios, and the other part is an ellipse. Appendix A gives a step-by-step explanation of how the ellipses are constructed.

Isoquants allow the display of, not just one ratio, but three ratios on a single graph. This helps the eye to make comparisons. The isoquants lay the foundation for drawing ellipses to summarize the data.

One ellipse allows a summary of three ratios for a single year or a single age group. The tilt and spread of the ellipse describe the ratios for the central group of points. The ellipses label points outside as outliers. Several ellipses on the same graph show how the ratios change over time or across age groups.

REFERENCES

Accident Facts. National Safety Council.

Allgaier E., *Age, Exposure, and Accidents.* Traffic Engineering and Safety Department, American Automobile Association Research Report, Washington, D.C., 1964.

Allgaier E., Accident involvement of senior drivers, *Traffic Digest Rev.* **13**(3), 17–19, 1965.

American Medical Association and American Association of Motor Vehicle Administrators., *Proc. National Conf. Aging Driver*, Washington, D.C., 2–4 May, 1974.

Ander A., Einfluss von lebensalter und fahrpraxis bei verkehrsunfallen hefte unfallheik. *Proc., National Conf. Aging Driver*, **66**, 1961, Washington, D.C., 2–4 May, 1974.

Baughman, A. L. and Parunak A., *Computer Program ELLIPSE.FTN.* The University of Michigan, Transportation Research Institute, Ann Arbor, Michigan, 1983.

Baughman A. L., *Computer program ELLIPSE.BAS.* Centers for Disease Control, Atlanta, GA, 1984.

Bureau of Public Roads Study. Reported in *The Federal Role in Highway Safety*, a letter from the Secretary of Commerce, 1959.

Burg A., *The Relationship Between Vision Test Scores and Driving Record: General Findings.* Los Angeles, California: Department of Engineering, Reports Group, University of California, Report No. 67–24, 1967.

Bygren L. O., The driver's exposure to risk of accident, *Scandanavian J. Social Med.* **2**, 49–65, 1974.

Carr B. R., A statistical analysis of rural Ontario traffic accidents using induced exposure data, *Accid. Anal. & Prev.* **1**, 343–357, 1969.

Carroll P. S., The meaning of driving exposure, *HIT Lab Reports*, The University of Michigan, Transpn. Research Institute, Ann Arbor, Michigan, April 1971.

Crancer A., *Involvement of the Problem Driver in Fatal Motor Vehicle Accidents.* Washington Department of Motor Vehicles: Division of Research and Statistical Analysis, Report No. 002, 1967.

Finesilver S. G., *The Older Driver: a Statistical Evaluation of Licensing and Accident Involvement in 30 States.* University of Denver, College of Law, Denver, Colorado, 1969.

Foldvary L. A., Road accident involvement per miles travelled—IV., *Accid. Anal. & Prev.* **10**, 143–176, 1978.

Freeman J. T., Elderly drivers: growing numbers and growing problems, *Geriatrics* **27**(7), 46–56, 1972.

Hall W. K., An empirical analysis of accident data using induced exposure, *HIT Lab. Reports*, September 1970. The University of Michigan, Transpn. Research Institute, Ann Arbor, Michigan, 1970.

Kent D. P. and Novotny G. B., Aging and automotive accidents, *Geriatrics* **16**, 271–277, 1964.

Lauer A. R., Age and sex in relation to accidents, *Highway Res. Board Bull.* **60**, 25–35, 1952.

Luepker R. V. and Smith M. L., Mortality in unionized truck drivers, *J. Occupational Med.* **20**(10), 677–682, 1978.

Marsh B. W., Aging and driving, *Traffic Eng.* **31**, 11–29, 1960.

McDole T. L., Age and sex of drivers involved in fatal crashes in Michigan during 1968, *HIT Lab. Reports*, February 1971, The University of Michigan, Transpn. Research Institute, Ann Arbor, Michigan, 1971.

McFarland R. and O'Doherty, B., Work and occupational skills, In *Handbook of Aging and the Individual*, Edited by J. E. Birren, pp. 481–483. The University of Chicago Press, Chicago, Illinois, 1952.

McFarland R. A., Tune G. S. and Welford A. T., On the driving of automobiles by older people. Presented at the Annual Meeting of the Gerontological Society, Boston, Massachusetts, November 1963.

Mosteller F. and Tukey J. W., *Data Analysis and Regression.* Addison-Wesley, Reading, Massachusetts, 1977.

Munden J. M., The accident rates of car drivers by age, *Int. Road Safety and Traffic Rev.* **14**(1), 28–29, 1966.

O'Day J., Computer Program ELLIPSE21. The University of Michigan, Transportation Research Institute, Ann Arbor, Michigan, 1983.

O'Day J., Filkins L. and Kaplan R., Combination Vehicles: Five-year Accident Experience, University of Michigan, Transpn. Research Institute, Ann Arbor, Michigan, 1980.

Planek T. W., Condon M. E. and Fowler R. C., An investigation of the problems and opinions of aged drivers. National Safety Council, Report No. 5/68, Chicago, Illinois, 1968.

Planek T. W. and Fowler R. C., Traffic accident problems and exposure characteristics of the aging driver, *J. Gerontology* **26**(2), 224–230, 1971.

Planek T. W. and Overland R. B., Profile of the aging driver, *Traffic Safety* **73**(1), 9–10, 1973.

Schwenk, L. C., Age and sex of motor vehicle fatality victims, *Proc. Iowa Academy of Sci.* **63**, 615–619, 1956.

Siebrecht E. B., Schumacher C. F. and Lauer A. R., Accident characteristics of drivers at various age levels. *Traffic Safety Res. Rev.* **3**(4), 29–32, 1959.

Stoto M., Fertility change in Russia summarized: the march of the ellipses, Chap. 6 from *Human Fertility in Russia Since the 19th Century*, by A. J. Coale, B. Anderson and E. Harm, 1980.

Swanson C. O., Schwenk L. C. and Lauer A. R., Age and fatal motor vehicle accidents, *Highway Res. Bull.* **212**, 21–26, 1959.

Thorpe J. P., Calculating relative involvement rates in accidents without determining exposure, *Aust. Road Res.* **2**(1), September 1964.

Tukey, J. W., *Exploratory Data Analysis*, Addison-Wesley, Reading, Massachusetts, 1977.

Tukey, J. W. and Parunak A., Display of points (x, y) Chap. B6 in *Statistician's Guide to Exploratory Data Analysis* (to be published), 1986.

Waller J. A., Injury in aged: clinical and epidemiological implications, *New York State J. Med.* **12**, 2200–2208, 1974.

Waller P. F. and Reinfurt D. W., *The Who and When of Accident Risk: Can Driver License Programs Provide Countermeasures*? University of North Carolina, Chapel Hill, North Carolina, 1973.

Washington State Patrol, *Senior Drivers Study*, Olympia, Washington, 1969.

APPENDIX A. CONSTRUCTING AN ELLIPSE

(A) *Outline for constructing an ellipse*

We present a general outline for drawing an ellipse. We show how to determine the shape, the location, and the size of an ellipse. Outliers do not disturb the method. In the first step we describe how to construct a preliminary line. We need the preliminary line to serve as a guide for determining shape and location in the second step. In the last step we fix the overall size of the ellipse. This outline shows the construction of the ellipse for the 26 states in Panel A of Figure 1. (See also Stoto [1980].)

Step 1. *Construct the preliminary line* (*Fig*. 5)

(a) Find the fourths for points ordered along the x-axis and draw a line at each fourth. Given 26 points in all, the fourths are the $6\frac{1}{2}$th from either end.

(b) Find the fourths for points ordered along the y-axis and draw a line at each fourth.

(c) Note which interquartile distance on the graph is greater. In Fig. 5, the x distance is greater. Then for points ordered along the x axis find the median values for x and y within the upper and lower sets of points. (The fourths for the axis whose interquartile distance is greater define upper (U) and lower (L).) Draw a line through these two points: (x_L, y_L) and (x_U, y_U).

Step 2. *Determine the shape and location of the ellipse* (*Fig*. 6)

(a) Find the fourths for points ordered along the preliminary line and draw a line at each fourth.

(b) Find the fourths for points ordered along the perpendicular to the preliminary line and draw a line at each fourth to complete the rectangle.

(c) Draw a point halfway between the lines at each fourth from Step 2(a) and halfway between the lines at each fourth from Step 2(b). This point will be the center of the ellipse.

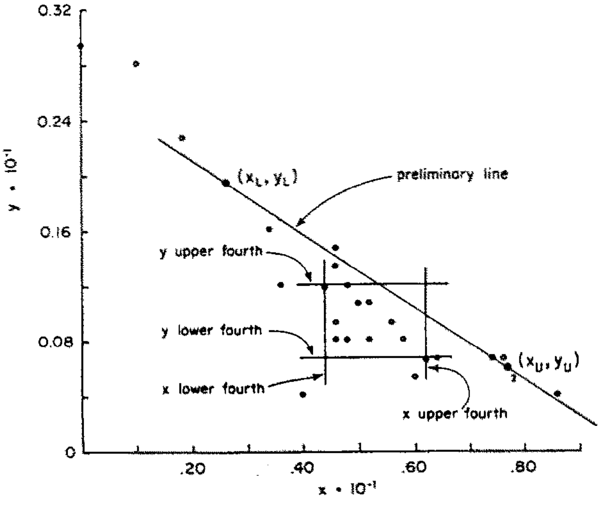


Fig. 5. Illustration of Step 1. (The small 2 near the lower end of the preliminary line indicates the position of two points.)

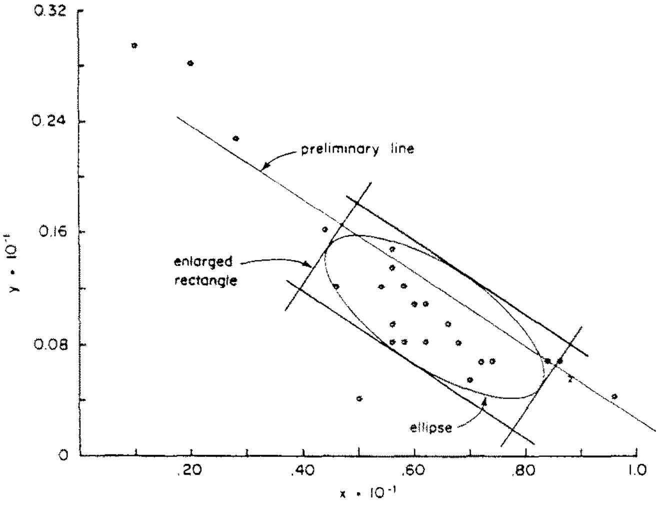


Fig. 6. Illustration of Step 2.

Step 3. Fix the size of the ellipse (Fig. 7)

(a) Measure the width and length of the rectangle in Step 2 and multiply each distance by two.

(b) Draw the new enlarged rectangle. The center of the enlarged rectangle will be the same as the center of the rectangle in Step 2.

(c) Draw the ellipse inside this rectangle.

(B) *Interpreting an ellipse*

The horizontal extent and the vertical extent of an ellipse show the variability of x and y, respectively. The direction of the major axis reveals any tendency for x and y to be related. The two variables are statistically independent if the major axis is horizontal or vertical. They are positively related if the slope is positive, and negatively related if the slope is negative. A close relation is implied if an ellipse that is tilted is also narrow.

(C) *Comments on constructing an ellipse*

1. *Finding the Fourths.* The outline for constructing an ellipse requires finding the fourths for sets of ordered points. We refer the reader to Tukey [1977, pp. 29–39] or Mosteller and Tukey [1977, pp. 43–49] for illustrations of finding the fourths and other summary values in a batch of data.

2. *Scale Factor.* The outline is useful for most sets of (x,y) points. For some applications, one can improve the ellipse by modifying the outline slightly. One modification concerns the size of the ellipse. In Step 3 we multiply by two both the width and the length of the rectangle in Step 2. We choose a scale factor of two because the resultant ellipse summarizes the batch of points, as well as being sensitive to outliers.

An extremely large value for the scale factor illustrates how scaling is important. If we choose a scale factor much greater than two (e.g. four or five), the ellipse will surround every point. In this situation we preserve the orientation of the ellipse, but we lose crucial information about outliers and the degree of relation between x and y. If we choose a scale factor much less than two (e.g. 1.5 or 1.3), again we know about the direction of the ellipse and possibly the

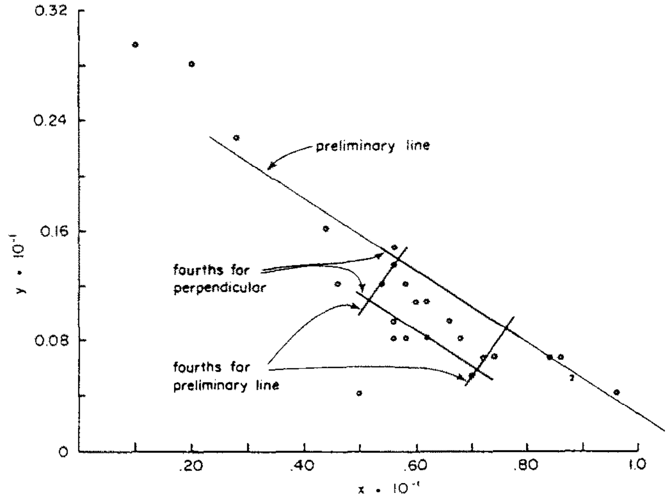


Fig. 7. Illustration of Step 3.

relation between x and y, but we must label too many points as outliers. At either extreme, we lose accuracy of interpretation about the behavior of the z ratio, for example.

Even a small change in the scale factor can significantly alter how well the ellipse surrounds the points. For instance, in our second example (Panels A–C of Fig. 3) we choose a scale factor of 2.5 for all sets of points. Employing 2.0 gives too small an ellipse, but 3.0 gives too large an ellipse. Because testing more than one scale factor is easily accomplished, even by hand, we recommend starting with a scale factor of two and finding a factor that surrounds about 75% of the points. An iterative program would be useful.

3. *Location*. A second modification of the general outline for constructing an ellipse concerns the location of the ellipse. From Step 2(c) we observe that the center of the rectangle is the center of the ellipse. In very asymmetrical sets of data, centering the ellipse in this way will not capture well the central points. For these situations, we recommend another method to locate the center of the ellipse. First, we find the intersection of the median position of the points ordered along the preliminary line with the median position of the points ordered along the perpendicular to this line. This point becomes the center of the ellipse. The major axis will have the same slope.

Table 2. Summary of ratios* by year for measures of individual and group exposure

Author [Year]	Ratios* Considered
Allgaier [1965]	$\dfrac{\text{fatal accidents}}{\text{100,000 licensed drivers arrested/years of driving}}$
	$\dfrac{\text{all accidents}}{\text{100 licensed drivers/years of driving}}$
	$\dfrac{\text{percent of drivers involved in accidents}}{\text{percent of drivers arrested}}$
Munden [1966]	$\dfrac{\text{car driver casualties}}{\text{100 million miles driven}}$, $\dfrac{\text{casualty rate for age group } i}{\text{casualty rate for all age groups}}$
	$\dfrac{\text{drivers involved in fatal and serious accidents}}{\text{100 million miles driven}}$
	$\dfrac{\text{drivers involved in all injury accidents}}{\text{100 million miles driven}}$
Burg [1967]	$\dfrac{\text{accidents}}{\text{100,000 vehicle miles}}$
Crancer [1967]	$\dfrac{\text{percent of fatal accidents}}{\text{percent of licensed drivers}}$
Finesilver [1969]	$\dfrac{\text{percent of accidents}}{\text{percent of drivers}}$
Waller and Reinfurt [1973]	$\dfrac{\text{percent of accident trips}}{\text{percent of exposure trips}}$
Bygren [1974]	$\dfrac{\text{percent of drivers involved in fatal accidents}}{\text{percent of mileage driven}}$
Folduary [1978]	$\dfrac{\text{number of accidents}}{\text{vehicle miles of travel, in million miles}}$
Luepker and Smith [1978]	$\dfrac{\text{number of fatal accidents}}{\text{population}}$
Lauer [1952]	$\dfrac{\text{accidents}}{\text{100,000 miles}}$, $\dfrac{\text{accidents}}{\text{licensees}}$
Bureau of Public Roads Study [1959]	$\dfrac{\text{drivers involved in accidents}}{\text{100 million vehicle miles of travel}}$
McFarland and O'Doherty [1959]	$\dfrac{\text{drivers in accidents}}{\text{licensed drivers}}$, $\dfrac{\text{drivers adjudged "at fault"}}{\text{drivers involved in accidents}}$
Siebrecht et al. [1959]	(Accident Index - 1/2 Violation Index) 10
	where: Accident Index = $\dfrac{\text{number of recorded accidents}}{\text{annual mileage} \cdot \text{years of driving/100,000}}$
	Violation Index = $\dfrac{\text{number of recorded violations}}{\text{annual mileage} \cdot \text{years of driving/100,000}}$
Swanson et al. [1959]	$\dfrac{\text{driver fatalities}}{\text{driver licensees}}$, $\dfrac{\text{driver fatalities}}{\text{100 million miles vehicular travel}}$
Ander [1961]	$\dfrac{\text{number of car crashes}}{\text{number of drivers}}$
McFarland et al. [1963]	$\dfrac{\text{accidents}}{\text{licensees}}$, $\dfrac{\text{drivers, held to be at fault, involved in accidents}}{\text{drivers involved in accidents}}$
Allgaier [1964]	$\dfrac{\text{fatal accidents}}{\text{100,000 drivers}}$, $\dfrac{\text{all accidents}}{\text{200 drivers}}$
	$\dfrac{\text{license withdrawals for fatal accidents}}{\text{1,000,000 drivers}}$
Kent and Novotny [1964]	$\dfrac{\text{percent of accidents}}{\text{percent of drivers}}$

*The numerators and denominators are both for specific age groups.

Table 3. Summary of ratios by year for measures of paired exposure

Author [year]	Ratios Considered
Thorpe [1964]	Relative accident likelihood = $\frac{T}{2T - S}$ where, S = proportion of ith driver-vehicle combination found in single-vehicle accidents, T = proportion of ith driver-vehicle combination found in collision accidents
Carr [1969]	$\frac{\text{Frequency of occurrence of ith category in the responsible population}}{\text{Frequency of occurrence of ith category in the non-responsible population}}$
Hall [1970]	Similar to Carr [1969].

$$\frac{y_U - y_L}{x_U - x_L}.$$

The major and minor axes will still have the same initial length, as in Step 2(b), and the axes will still be subject to a multiplicative factor in Step 3(a).

In this alternative method for centering the ellipse, the median along the preliminary line may not be halfway between the fourths along the line. The distance between the fourths, the inter-quartile distance, is the length of an axis. We can choose to center the axis on the median along the preliminary line and not to let the center remain at the point halfway between the fourths. In this way, for asymmetrical data, the ellipse encloses the most dense group of points, labelling the others as outliers.

APPENDIX B. REVIEW OF LITERATURE ON AGEING AND ACCIDENTS IN ARTICLES WITH STATISTICS AND DATA ANALYSIS

This Appendix presents the results of a literature review to see what criteria other researchers used to answer the question of older drivers' involvement in traffic accidents.

Many analysts use data about both accidents and exposure to identify high-risk drivers. They divide the number of reported accidents by the corresponding measure of exposure. We first discuss the meaning of driver exposure. Then we classify the approaches highway-safety researchers take to analyzing data for accidents on the basis of which measure of exposure they use.

Carroll [1971] proposes the following definition: "Driving exposure is the frequency of traffic events which create a risk of accident." He calls measures of exposure that the driver controls, direct measures, for example, driving distance, driving time, traffic volume on the road he chooses. We favor the more descriptive and more specific term: individual exposure. He calls those that the driver does not control, indirect measures, for example, number of registered vehicles, number of licensed drivers, gallons of gas consumed by all vehicles. We would choose the expression: group exposure. In Fig. 1, the denominator of x = (number of accidents)/(population in age group) is the population of people in a given age group and state. It is a measure of group exposure. This ratio takes into account the unequal proportions of licensed drivers that occur among states. Table 2 summarizes the ratios in the literature based on both individual and group exposure. This table of ratios is particularly interesting because of the increased possibilities for data analysis by isoquants it suggests.

The words "direct" and "indirect" have a different connotation in the statistical literature (see Mosteller and Tukey [1977, Chapter 11]), so we think it is better not to use them in this context. Both types of exposure data can serve as part of accident rates. The most commonly used measure of exposure is driving distance of one driver expressed in miles travelled.

Either individual or group driving exposure comprises one approach of highway-safety researchers to analyzing data for accidents and exposure. Another approach is to select a control population. The usual control population is drivers deemed not to be responsible in a two-vehicle accident. The road conditions for the responsible and nonresponsible driver are then quite similar. We divide the number of accidents "caused" by drivers in a particular category by the number of accidents not "caused" by the drivers in that category. The literature refers to the numbers of nonresponsible drivers involved in accidents as "induced-exposure data" (Thorpe [1964]; Car [1969]; Hall [1970]). This ratio requires the assumption that there is a responsible and a non-responsible driver in every accident. We like the mnemonic title: paired exposure.

Fewer investigators study the relation between traffic accidents and age of driver using paired-exposure data than do those using individual and group-exposure data. Table 3 summarizes these ratios.