Lane changes while driving occasionally cause collisions. Do we make too many lane changes?

Are Those Other Drivers Really Going Faster?

Donald A. Redelmeier and Robert J. Tibshirani

PHOTO COURTESY THOMAS KOHOUT

Introduction

Motor-vehicle travel is a mixed blessing for modern times. During the average day in the United States, for example, about 100 people step into a vehicle and do not emerge alive according to data from the *1996 Statistical Abstract of the United States*, published by the Bureau of the Census. Crashes are especially poignant if they kill healthy people who otherwise might have led long and productive lives. A 1957 *New England Journal of Medicine* study found that crashes are usually (>90%) attributed to driver error rather than failures in the vehicle or roadway. The most important factor in driver error is alcohol, contributing to about 40% of fatal collisions in 1994 according to a National Highway Traffic Safety Administration report. The other factors causing driver error are not completely known. A better understanding of such errors might allow people to benefit from motor vehicle travel at a lower personal risk.

One potential driver error is an inappropriate lane change, a vehicle maneuver that may have substantial risks for several reasons. First, it causes the individual to straddle traffic flows and be exposed to two streams of vehicles. Second, it requires the driver to make rapid judgments about sufficient spacing. Third, it increases the hazard related to other vehicles approaching along the driver's blind spot. Fourth, it disrupts the traffic pattern for following vehicles. The overall risks associated with each lane change are uncertain because the amount of normal driving spent making lane changes is not known with precision; however, rough estimates in an Ontario Ministry of Transportation report suggest about a threefold relative risk if less than 1% of normal driving involves a lane change.

We wondered whether people can accurately judge if they are in a lane that is slower than the next lane on a congested roadway. Mistaken impressions, for example, might cause a driver to incorrectly think the next lane is faster and motivate a needless lane change. Perhaps errors in judgment produce a systematic bias and create an illusion that the next lane is generally moving faster, even if all lanes have the same average speed. One basis for such error is if drivers expect that they should spend equal amounts of time passing and being overtaken. We have shown that such an expectation is mistaken when time is measured by discrete intervals. We recently popularized this finding in an exceedingly short paper published in *Nature*. This article describes the work in detail.

Methods

Individual Vehicle Perspective

Most traffic studies describe vehicle flow at a macroscopic level. Doing so makes sense because agencies tend to focus on the mobility of large populations rather than unique individuals. In addition, studying traffic at the level of an individual vehicle is problematic without elaborate video equipment or a helicopter. For our study, however, we wanted to learn about an individual person's perceptions while driving a single vehicle. To address this topic, we relied on computer simulations for testing diverse and extreme situations. Later in the article we discuss field data we also collected to test our models. Computer simulations were the core of our work because experiments on real drivers seemed to be unsafe, unethical, impractical, or too expensive.

The computer simulations tracked each vehicle each second to determine where it was and what it was doing (accelerating, decelerating, or staying at constant velocity). We assumed a vehicle would accelerate if it was traveling slower than its target speed and no other vehicle ahead was within the minimum headway distance. The vehicle would decelerate if another vehicle was ahead and within the minimum headway distance. The vehicle would maintain a constant speed if it achieved target speed and no other vehicle ahead was too close (within the minimum headway distance). Initial simulations used a minimum headway distance guaranteeing that collisions would not occur, whereas supplementary simulations applied less restrictive headways.

The baseline analysis used parameters to yield a plausible simulation yet allow subsequent sensitivity analyses. The target speed was set at 100 km/h (63 mi/h) and identical for all vehicles. The acceleration was set uniform for all vehicles and capable of going from 0 km/h to 100 km/h in 10 seconds. The deceleration was set uniform for all vehicles and capable of going from 100 km/h to 0 km/h in 5 seconds. Such performance is similar to that of a Honda Accord. The minimum headway distance (d) was a function of current velocity (v) and equal to the expression [$d = (v^2/100) + 1$]. Thus, a vehicle traveling at 100 km/hr required a minimum headway of 101 m, stayed at constant speed if the next vehicle was 110 m ahead, and decelerated if the next vehicle was 90 m ahead.

Aggregate Traffic Characteristics

Other assumptions were used to take into account roadway characteristics and traffic congestion. The most important assumption was that the number of vehicles and amount of available roadway was stable. This implied that the overall spacing of traffic, expressed as the average length of roadway available for the average vehicle, was constant and not negligible. Thus, conditions which provided little total roadway for large numbers of vehicles resulted in substantial congestion. If the average spacing was less than 100 m, then not all vehicles could maintain target speed under baseline conditions. Moreover, an average spacing that was greater than 100 m did not guarantee smooth flow unless all the vehicles were spaced fairly evenly.

A completely realistic simulation of a single lane of traffic was impossible because vehicles can differ in target speed, headway tolerance, acceleration, deceleration, starting position, current velocity, and sequence relative to others. Our model was designed with a few initial sources of randomness, then gradually made more complex. The baseline condition assumed that all vehicles had identical performance characteristics and started at zero velocity. Starting positions were initially generated using a mixture of normal distributions to provide spacing that was nonnegative with low mean and high variance. The mixture of normal distributions specifies that a random variable is drawn according to a specified probability from one of two different normal distributions.

The main advantage of the computer simulations is that they permitted construction of a second lane of traffic relatively effortlessly. Moreover, the second lane could be established with characteristics identical to the index lane, including the number of vehicles and overall spacing. By applying a different starting seed to the random generator, however, the two lanes could follow somewhat different patterns yet obtain the same average speed. Furthermore, computer simulations made it possible to prevent drivers from making lane changes. The essential contribution of a two-lane simulation was that it allowed drivers to not just determine their absolute speed on the roadway but also to determine their relative speed compared to vehicles in the next lane.

Psychological and Statistical Issues

Research in psychology suggests that people in parallel queuing processes tend to judge their speed by assessing their progress relative to those in the other queue. We defined a "passing epoch" when the index driver started behind and ended ahead of one or more drivers in the other lane after a one-second interval. We defined a "being overtaken epoch" when the index driver started ahead and ended behind one or more drivers in the other lane after a one-second interval. Other investigators use the terms "skips" and "slips" to denote these two events. All else equal, drivers prefer passing rather than being overtaken. A key issue was that people are sensitive to these events but relatively insensitive when two events occur simultaneously.

All simulations were programmed using the S-PLUS statistical language. The typical run time was about 5 minutes to create a single simulation of about 10 minutes of highway traffic. For our analyses, we typically ran about 100 simulations for both traffic lanes to obtain stable estimates for the mean number of overtaking epochs and passing epochs during each 10-minute interval for an index vehicle. Typically, this required more than a full day of computer processing time to create simulations and summary statistics. Comparisons were expressed as both the absolute number of overtaking and passing epochs as well as the relative number of overtaking and passing epochs. All comparisons were statistically significant ($p<.001$), except where noted.

Results

Virtual Vehicles

The speed of individual drivers showed long intervals of slow velocity and short bursts of high velocity (Fig. 1). The overall pattern was nonconstant and nonperiodic with most time spent at slow velocity. As expected, the correlation between a vehicle's current speed (compared to the road) and current spacing (compared to vehicle in front) was strong and positive (the correlation was .34, with a 95% confidence interval .26 to .42). The mean velocity was about 18 km/h, which is slower than the velocity of about 31 km/h theoretically possible if all vehicles achieved uniform spacing (and also slower than theoretically possible if vehicles tailgated in an unsafe manner). The variation in speed and spacing showed no trend toward abating as the simulation duration increased.

The relative speed of one index driver compared to one other driver in the next lane also followed a complex pattern (Fig. 2). Substantial variation in relative speed and relative position was observed even over a short time interval. The curves did not represent conventional random walks because extreme differences were unlikely and there were frequent crossings of the zero point. In addition, positive serial correlation was found so that the average narrowing epoch (where the separation between the two drivers decreased) was about three times more likely to be followed by another narrowing epoch than by a widening epoch. More complicated descriptive statistics were possible by evaluating two (or more) comparator drivers in the next lane.

Summary Findings

Summary statistics were based on comparing the index driver to all the other drivers traveling in the next lane of traffic. Much of the time nothing happened and relative positions remained unchanged. Typically, during a 10-minute interval about 76 of the 600 one-second epochs (13%) were characterized by an event. Epochs in which the index car was overtaken were more frequent than epochs in which the index car was passing (43 vs. 33, ratio = 130%). Of course, the average number of cars overtaking the index car during the average overtaking epoch was proportionately smaller than the average number of cars passed by the index car during the average passing epoch. Hence, the total number of overtakes equalled the total number of passes over the full interval (each about 46).

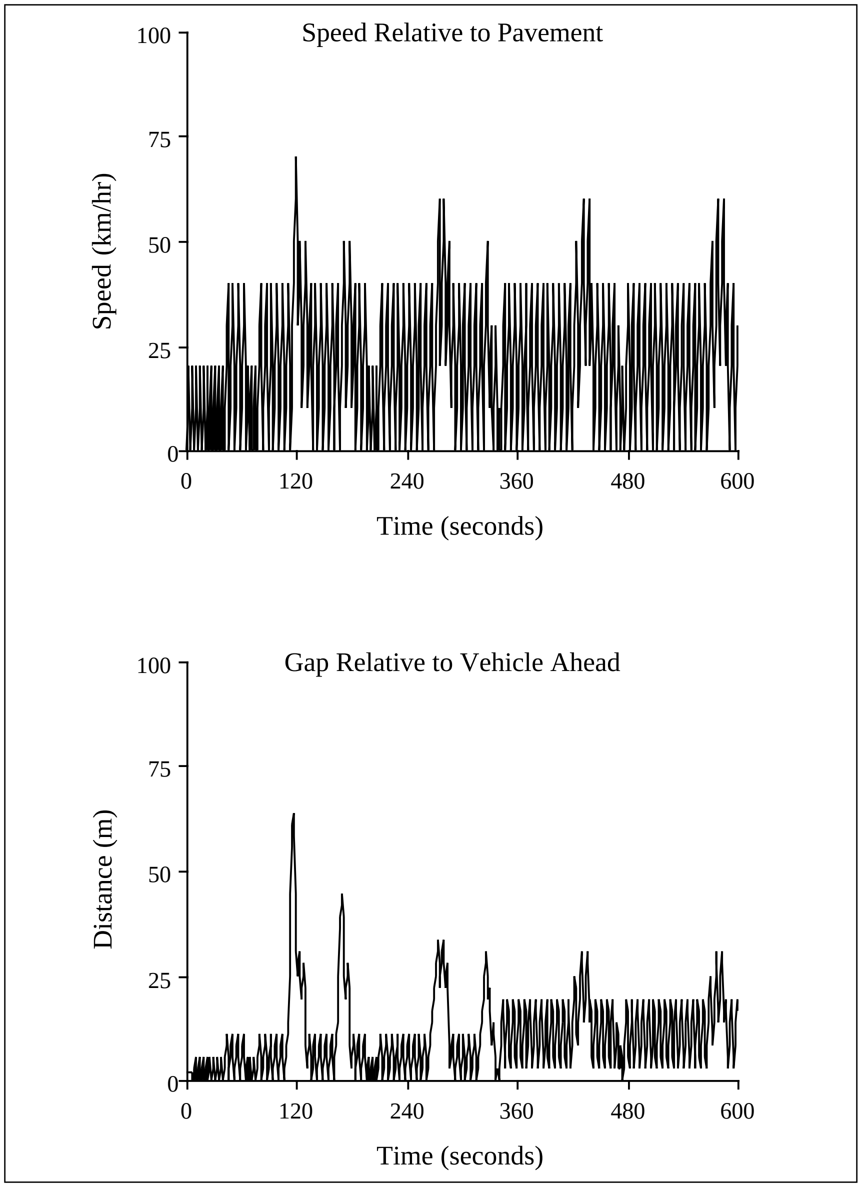


Figure 1. Speed and position of one vehicle in a single lane of traffic. Top panel shows speed of index vehicle relative to pavement. Bottom panel shows distance of index vehicle from vehicle immediately ahead. *X* axis shows 10 minute time interval as single-second epochs. *Y* axis shows speed and distance, respectively, at each time epoch. Note, for example, that a large gap appears for the index vehicle slightly before the 120-second mark, which allows a large increase in speed for the index vehicle slightly after the 120-second mark. In this simulation the vehicle was programmed to accelerate at exactly 10 m/s/s and decelerate at exactly 20 m/s/s.

These baseline conditions indicated that the next lane might mistakenly appear faster than the driver's current lane, even though both lanes had identical average speeds. Results from individual simulations varied, but in 70% of cases the imbalance indicated fewer passing epochs than being overtaken epochs. For all simulations, the asymmetry was similar during the first and the second five minutes of simulation. The exact number of vehicles in each lane (varying between 100 or 1,000) made no difference provided that the average traffic density remained unchanged. One of the strongest determinants was the overall degree of congestion, where decreasing congestion

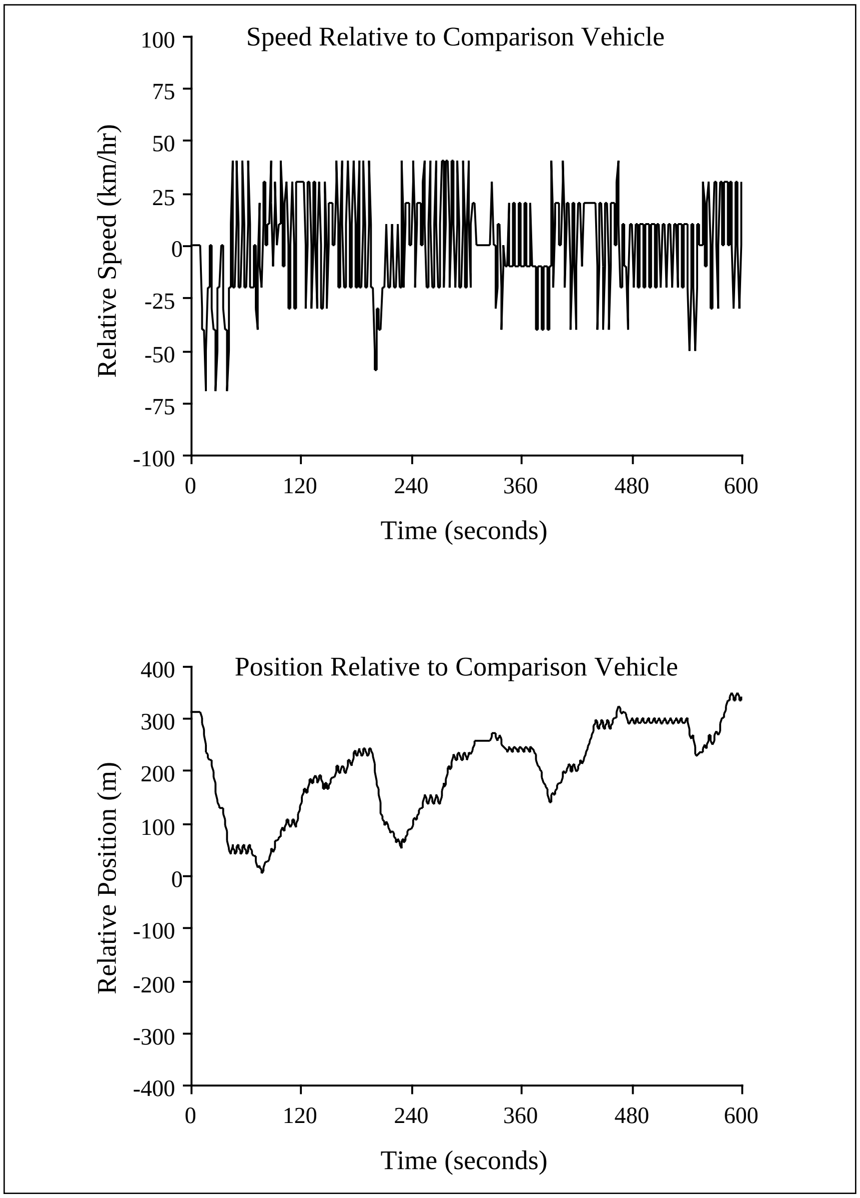


Figure 2. Speed and position of one vehicle compared to a vehicle in next lane. Top panel shows speed of index vehicle relative to one vehicle in the next lane. Bottom panel shows distance of index vehicle relative to the other vehicle. *X* axis shows 10 minute time interval as single-second epochs. *Y* axis shows speed and distance, respectively, at each time epoch. Note, for example, that during the first 30 seconds the index vehicle loses much of its advantage over the other vehicle as shown by a negative relative speed and a declining relative position. In this simulation both vehicles were programmed to accelerate at exactly 10 m/s/s and decelerate at exactly 20 m/s/s.

tended to attenuate the asymmetry (Fig. 3).

Testing Other Factors

A more subtle characteristic of roadway congestion was the initial source of randomness. Our baseline condition assumed that the starting gaps between vehicles followed a mixed normal distribution: 90% of the gaps followed a normal distribution with mean 2 and standard deviation 0.1 and 10% followed a normal distribution with mean 100 and standard deviation 5. This created a bimodal distribution of initial gaps and a skewed distribution of subsequent gaps (Fig. 4). Replacing the mixed normal random generator with a Bernoulli random generator (90% of the gaps are of length 2 and 10% are of length 100) yielded a somewhat greater subsequent apparent difference in speed (54 vs. 36, ratio = 150%). Replacing the mixed normal random generator with a Poisson random number generator having the same mean (11.8) yielded a smaller subsequent apparent difference in speed (81 vs. 70, ratio = 116%).

Altering the vehicle characteristics could accentuate or attenuate the apparent difference in speed. A particularly high target speed had virtually no effect, however, because almost no vehicles attained their target speed under baseline conditions. Vehicles with especially fast acceleration, that could go from 0 km/h to 100 km/h in 5 seconds rather than 10 seconds, generated a larger apparent difference in speed (41 vs. 22, ratio = 186%). Vehicles with especially fast deceleration, that could go from 100 km/h to 0 km/h in 2.5 seconds rather than 5 seconds, generated a smaller apparent difference in speed (68 vs. 58, ratio = 117%). Simulations in which only the index vehicle had special attributes yielded results similar to the baseline findings.

Several characteristics of the driver could change but not reverse the apparent difference in speed. Doubling the minimum headway distance for the index driver reduced the apparent difference in speed (51 vs. 46, ratio = 112%). Cutting the minimum headway distance by half — to model more tailgating — increased the apparent difference in speed (33 vs. 16, ratio = 208%). Allowing the index vehicle to maintain cruise control at the same average speed as the other lane had negligible effect on the apparent difference in speed (48 vs. 37, ratio = 130%). Allowing the index driver to make fewer glances at the next lane so that comparisons occurred every two seconds rather than every one second reduced the apparent difference in speed (16 vs. 14, ratio = 114%).

People's Beliefs

To contrast against our formal analysis, we surveyed students attending a safety course and asked about their intuitive beliefs concerning driving in heavy traffic (n=110). The majority (74%) believed that if the next lane has the

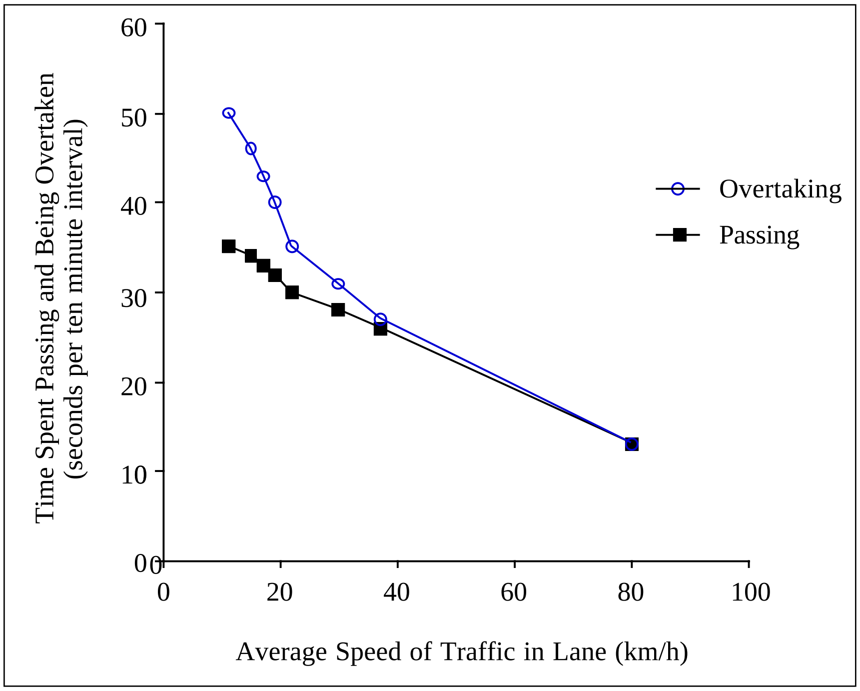


Figure 3. Relation of traffic speed to time spent passing and being overtaken. Figure shows relation of traffic speed to amount of time spent passing and being overtaken. *X* axis shows the overall average speed of a lane. *Y* axis shows the total number of single-second epochs in a 10 minute interval in which index driver passes or index driver is overtaken by one or more vehicles in next lane. Note, for example, at an average speed of 15 km/h a driver spends about 46 seconds being overtaken by vehicles in next lane and about 33 seconds passing vehicles in next lane. Lesser congestion leads to higher speeds, fewer passing and overtaking epochs, and a smaller asymmetry.

same average speed as the current lane the amount of time spent passing and being overtaken should be equal. The remainder (26%) were split almost evenly on whether the amount of time passing should exceed or should trail the amount of time spent being overtaken (10% vs. 16%). Many (46%) indicated they would consider changing lanes if the relative difference exceeded 10%. The majority (72%) indicated they would consider changing lanes if the relative difference exceeded 50%.

As another test of validity we also videotaped traffic sequences by mounting a camera in a moving vehicle and recording the side-view perspective of the next lane on a congested roadway during rush hour. A section of videotape was then selected that showed four minutes of continuous traffic with many overtaking and passing events and with an overall speed slightly slower in the next lane. The section of videotape was screened to another group of driving students (n=120) who were asked to judge whether the next lane moved faster or slower than the index lane. The majority stated that the sequence was realistic (86%), that the next lane moved faster (70%), and that they would have made a lane change if safe to do so (65%).

Discussion

We developed a computer simulation to evaluate the speed of individual vehicles on a congested roadway. We then constructed a second lane of traffic having similar characteristics and assessed the apparent speed of the second lane relative to an index vehicle in the first lane. We found that the second lane could be mistakenly perceived as going faster because a driver generally spent more time being overtaken than passing. No combination of assumptions or parameters reversed this asymmetry. A videotape sequence obtained from field observations confirmed people's mistaken impressions of speed on a congested roadway. Together, these findings suggest a roadway illusion — namely, that the next lane on a congested roadway appears to be moving faster than the driver's current lane even if both lanes have the same average speed.

The mistaken impression arises due to several factors. The basic explanation is that vehicles spread out when going quickly and pack together when going slowly (Fig. 5). A driver on a congested roadway can pass many vehicles in a brief interval, whereas it takes much more time for the driver to be overtaken by the same number of vehicles. Hence, a journey has fewer passing epochs than overtaking epochs. Every driver, therefore, should normally expect to spend more time going slower than going faster compared to others. In contrast to this principle of queuing theory, people do not formally integrate the frequency and intensity of every moment and, instead, expect that the amount of good times should balance the amount of bad times. Of course, people's mistaken judgments are sometimes fortuitously correct and the next lane turns out to be faster.

Several factors can alter the apparent asymmetry in speed. We found that greater congestion, particularly if distributed unevenly, intensified the illusion. Vehicles with powerful engines or weak brakes also accentuated the illusion. Drivers who tended to tailgate or who made frequent glances at the comparison lane were more prone to the illusion. Together, these observations suggest that the apparent faster speed of the next lane is a function of both the roadway, the vehicle, and the driver. The essential role of congestion also suggests that the illusion may be a new phenomenon given that, for example, from 1985 to 1995 in the United States the number of vehicle miles travelled increased by 32% but the amount of available roadway increased by only 1%.

We studied faulty intuitions related to queuing theory, but other aspects of human reasoning could reinforce the impression that the next lane moves faster than the driver's current lane. Differential surveillance can occur because drivers direct more attention ahead than behind; consequently, vehicles that are passed (victories) turn invisible whereas vehicles that overtake (defeats) stay conspicuous. Moreover, glances at the next

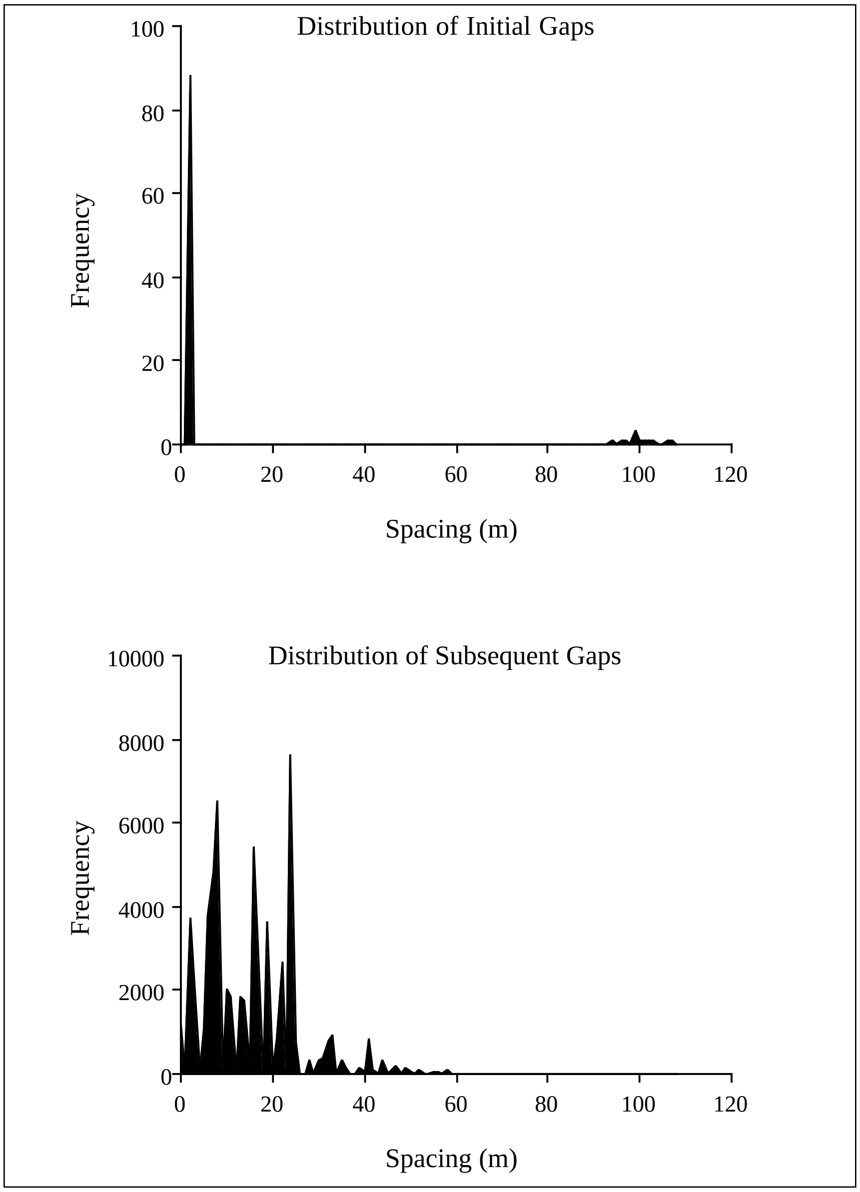


Figure 4. Characterizing initial and subsequent degrees of congestion. Top panel shows distribution of initial gaps between 100 consecutive vehicles at the first single-second epoch. Bottom panel shows distribution of subsequent gaps between 100 consecutive vehicles for the remaining 599 single second epochs. *X* axis shows spacing in meters, and *Y* axis shows corresponding frequency. Note, for example, that in this one simulation the initial bimodal distribution collapses substantially, but incompletely and unevenly, to an average spacing that is still about 15 meters between consecutive vehicles.

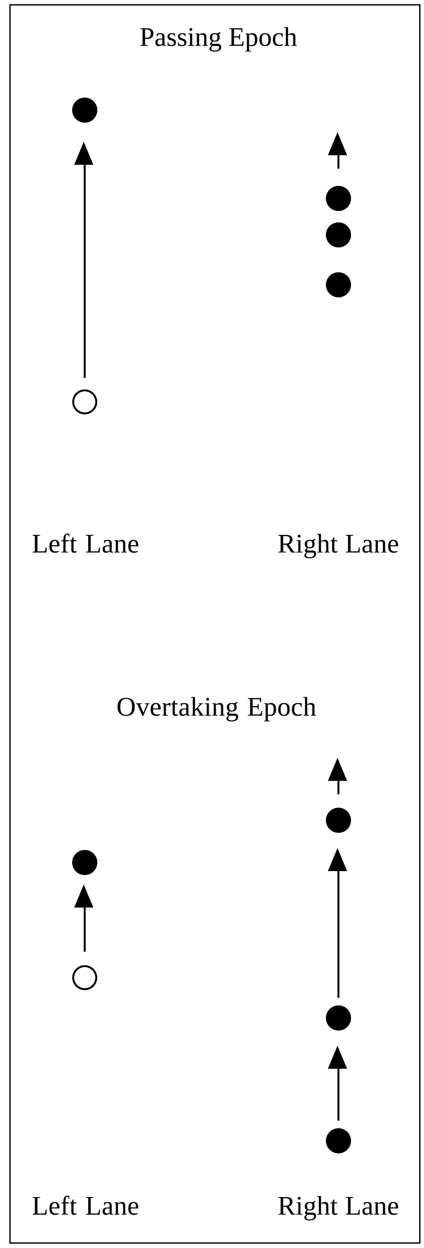


Figure 5. Asymmetry between passing and overtaking vehicles in the next lane. Top panel portrays an aerial view of a passing epoch. Bottom panel shows an aerial view of an overtaking epoch. Two lanes are shown, marked left and right, each with traffic flow in an upward direction. The index vehicle is represented by a hollow circle and all other vehicles are represented by solid circles. Note that the index vehicle can pass three vehicles in a single second if the right lane is sufficiently slow and congested; in contrast, the index vehicle is overtaken by each vehicle separately when the right lane is fast and uncrowded.

lane may be more frequent when drivers are relatively idle. Even if attention were omnidirectional and evenly paced, human psychology may make losses appear more salient than the corresponding gains, causing the joy of passing to feel less intense than the frustration of being overtaken. Finally, misconceptions about chance, like those described in the basketball "hot hand" literature, can make even small streaks seem unduly large.

The most important limitation of our analysis is the dearth of empiric data on the speed of individual drivers in congested roadways. This shortfall in background data is not surprising. Police investigations determine peak velocities for purposes of law enforcement. Traumatologists estimate impact velocities as an index for the severity of injury. Traffic engineers assess aggregate flow rates in efforts to maximize the mobility of populations. We are aware of no previous studies focusing on our line of investigation. The pioneering nature of our findings, therefore, calls for more research directly measuring the speed of individual drivers over extended intervals under natural circumstances. The data we obtained from simulations are already quite intricate, suggesting that data from field studies could be even more complicated.

An understanding of the highway illusion may encourage drivers to consider a few safety strategies. First and foremost, individuals should recognize the illusion and resist small temptations to change lanes. In addition, drivers might want to focus more on the clock and less on social comparisons when evaluating their progress in heavy traffic. Incidentally, drivers could try directing more attention toward the rear-view mirror and less toward the side-view window to enhance their subjective satisfaction. Finally, people could remember that some human factors leading to aggressive driving are not idiosyncratic pathology but normal psychology. Authorities stress that the secret for safe driving is to never be in a rush. Our findings suggest that naive attempts to rush may be misguided without a careful understanding of queuing theory.

Aftermath

After publishing our study in *Nature*, we were interviewed by many journalists and encountered five common questions. Some have wondered whether it is ever worthwhile to change lanes while driving. Clearly, a driver must occasionally avoid a fixed obstacle, prepare for an exit, avoid a forced exit, or escape from behind a truck. On occasions when the roads are congested, as well, there may be times when the next lane is moving faster and the driver can gain by making a lane change. Our main finding, however, suggests that the gains from discretionary lane changes are smaller than usually believed and moreover, that discretionary judgments are fallible when estimating speeds on congested roadways. A more conservative attitude, we hope, might encourage drivers to be selective about discretionary lane changes.

A few skeptics have questioned whether a research study can change people's actions and lead to safer driving. We recognize that knowledge does not always control behavior, yet have some reason for optimism. For example, in the United States seatbelts were initially received with great reluctance (17% adherence in 1982) but subsequently followed with growing popularity (69% adherence in 1988). Because inappropriate lane changes cause approximately 300,000 collisions each year in the United States (National Highway Traffic Safety Administration 1995; Ontario Ministry of Transportation 1998), a small improvement in behavior could still yield substantial benefits. In addition, a decreased frequency of lane changes might offer indirect benefits by creating an environment of greater lane discipline and less incentive for otherwise cautious drivers to tailgate as a way of preempting a lane incursion by others.

Some people have correctly noted that passing and being overtaken are really instantaneous events that are equal for both vehicles. We agree with this strict definition but emphasize that people tend to consider the events in more extended intervals. Some journeys seem to entail a lot of time spent passing or being overtaken, contrary to the theory that each event is instantaneous. The roadway illusion occurs if people do not follow strict definitions and tend to view experiences as time intervals, such as single-second epochs. Indeed, some people go further and err by driving as if it were a competitive sport (evaluating progress relative to others) rather than as a logical problem-solving exercise (evaluating progress by arrival time). Hence, the misperception stems from human psychology and not from logical theory.

Some journalists have wondered whether our findings might apply to slow speed queues such as those found in grocery stores and banks. Our research focused on high-speed travel because of the serious consequences of a motor-vehicle collision. For indoor settings, beliefs about being in the wrong lane may primarily stem from biases of multiple comparisons and ignoring other lanes that are moving slower than you. One more paradox is that consumers are happy to spend an hour navigating around the grocery store gathering items but become upset by having to wait five minutes for checkout. A better perspective might be to evaluate the total time spent door-to-door and place no special emphasis on the final few minutes. Apparently, people's feelings do not always match the reality of a situation.

We have also been quizzed on whether our results are likely to apply to congested roadways outside of Canada. We only obtained videotapes from the Trans-Canada highway, and we do not know the extent of roadway congestion elsewhere. The illusion would not apply to roadways that maintain organized laminar flow, where faster cars are streamlined to more central lanes. Instead, the illusion occurs on congested roadways because flow is disorganized, the laminar structure disintegrates, and drivers become impatient. Our main observation relates to people's misunderstandings of the laws of probability, and the laws of probability are enforced in all countries. Wherever the roads are congested, the best way to arrive five minutes sooner is to start five minutes earlier.

References and Further Reading

Bureau of the Census (1997), *Statistical Abstract of the United States 1996* (116th ed.), Washington, DC: Government Printing Office.

McFarlane, R.A., and Moore, R.C. (1957), "Human Factors in Highway Safety: A Review and Evaluation," *New England Journal of Medicine*, 256, 792-799.

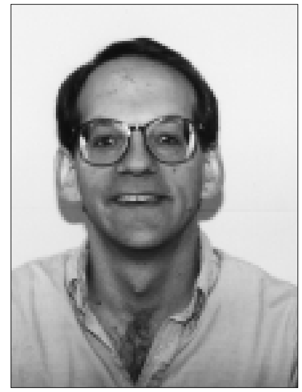
National Highway Traffic Safety Administration (1995), *Traffic Safety Facts 1994*., Washington, DC: U. S. Department of Transportation.

Ontario Ministry of Transportation (1998), *Ontario Road Safety Annual Report 1996*, Downsview, ON: Safety Research Office, Ministry of Transportation.

Redelmeier, D.A., and Tibshirani, R.J. (1999), "Why Cars in the Next Lane Seem to Go Faster," *Nature*, 401, 35.

AP Statistics

W. Robert Stephenson



Hal Stern

■ 1. Question

How important is randomization?

■ 2. Activity

Across the midwest farmers are constantly looking for that competitive edge that will increase profits. Responding to this demand, seed companies have developed, through cross breeding, hybrid varieties of corn with higher and higher yields. More recently, through genetic engineering, there are now corn varieties that are resistant to the affects of herbicide residue and others that can combat pests like the European corn borer. Once a variety of corn is developed, the true test of its values comes in field trials. A field trial is a designed experiment used to compare varieties of corn (or soybeans, or wheat, etc.) in terms of average yield (or some other measure of quality). Sir Ronald Fisher developed many of the methods of applied statistics while analyzing agricultural field experiments at Rothamsted in England. The following activity simulates an agricultural field experiment, or field trial, conducted to compare two varieties of corn, A and B.

■ Class Activity Introduction

Researchers at a large seed company are planning a field trial to compare two hybrid varieties of corn. The response of interest is the yield, in bushels per acre. The better variety will be the one with the highest yields but the researchers recognize that variation in soil composition, fertility and drainage will have effects on the growth of plants and thus yield. There is a filed with 36 plots available for the experiment. On 18 plots variety A will be planted and on the other 18 plots variety B will be planted. The researchers wish to see if the two varieties have equal yields, on average, or if the two varieties differ significantly. If the two varieties really do differ, the researchers would like their experiment and the subsequent statistical analysis to detect this true difference. The ability of a statistical procedure to detect a true difference is called the power of the procedure. The researchers must decide how to assign the varieties to the plots.

Convenience Assignment

It is easiest to plant one variety on 18 plots on one side of the field and the other variety on the 18 plots on the other side. Modern machinery cam plant up to 18 rows at a time, so planting in this way can be done in one or two passes through the field. Below is a picture of such an assignment and the yields, in bushels per acre, for each plot.

A 130	A 149	A 139	B 155	B 137	B 145
A 149	A 133	A 152	B 131	B 147	B 136
A 141	A 156	A 137	B 146	B 132	B 148
A 150	A 142	A 155	B 136	B 152	B 133
A 139	A 155	A 139	B 147	B 137	B 153
A 155	A 138	A 150	B 137	B 145	B 136

Summary	n	mean	std. dev
A	18	144.9	8.29
B	18	141.8	7.65

Based on this assignment, by convenience, does one variety appear to have a larger mean yield? Is there a significant difference in mean yields between the two corn varieties?

W. Robert Stephenson and Hal Stern are University Professors of Statistics at Iowa State University, Ames, IA.

Systematic Assignment

Many people think that an alternating sequence is a random, or at least an unbiased, sequence. For example, when assigning participants to treatment and control, taking every other participant (alternating) for the treatment group appears random. However, if participants are lined up alternating between male and female then all the males will be in one group and all the females in the other. Gender and group would be completely confounded. That is the effects of treatment and control are inseparable from gender effects. In a field, an alternating pattern would be like a checkerboard. Below is a picture of such an alternating pattern and the yields, in bushels per acre, for each plot.

A 130	B 137	A 139	B 155	A 149	B 145
B 137	A 133	B 140	A 143	B 147	A 148
A 141	B 144	A 137	B 146	A 144	B 148
B 138	A 142	B 143	A 148	B 152	A 145
A 139	B 143	A 139	B 147	A 149	B 153
B 143	A 138	B 138	A 149	B 145	A 148

Summary	n	mean	std. dev
A	18	142.3	5.75
B	18	144.5	5.37

Based on this assignment, alternating, does one variety appear to have a larger mean yield? Is there a significant difference in mean yields between the two corn varieties?

Discuss the results from the analysis of the convenience assignment data and those from the analysis of the alternating assignment data. Some may find it a bit disturbing that B appears better for one assignment while A appears better for the other. Of course, this could be due to chance variation. It could also be due to a poor assignment of treatments. For example, the right side/left side assignment is vulnerable to bias due to soil fertility, or drainage that is different from one side of the field to the other. The checkerboard assignment is also susceptible to fertility, drainage or other gradients.

Random Assignment

What if chance is used to assign varieties to plots? How, physically, would you randomly assign varieties to plots? Come up with a randomization scheme to assign variety A to 18 plots and variety B to the remaining 18 plots. Record your assignments in the table below.

Randomized Assignment

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Once you have completed your random assignment, ask your instructor for “The Truth” — this sheet gives the yield for each plot using either variety. “The Truth” was used to fill in the yields for the plots in the convenience and alternating patterns you looked at earlier. In general, “The Truth” is not available since it requires knowing what would happen to the same plot of land using each of the treatments.

Write down the yields for your random assignment — if you have an A in the row 1, column 1 plot then you would put down 130 whereas if you have a B in the row 1, column 1 plot you would put down 118 for the yield. Repeat for all squares. This gives you 18 A yields and 18 B yields. Based on this assignment, at random, did you find a significant difference in mean yield between the two corn varieties?

Share and discuss your results. Examine “The Truth” more closely. Which variety appears to have the larger yield? By how much?

■ 3. Suggested Solution

Convenience Assignment

Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is 1.17 with an associated two sided P-value of 0.25. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. If you are using the conservative degrees of freedom, min(n_1-1, n_2-1) = 17, the P-value would be 0.26. Although variety A has a slightly larger mean yield, there is not a statistically significant difference between the sample mean yields for the two varieties.

THE TRUTH

A = 130 B = 118	A = 149 B = 137	A = 139 B = 127	A = 167 B = 155	A = 149 B = 137	A = 157 B = 145
A = 149 B = 137	A = 133 B = 121	A = 152 B = 140	A = 143 B = 131	A = 159 B = 147	A = 148 B = 136
A = 141 B = 129	A = 156 B = 144	A = 137 B = 125	A = 158 B = 146	A = 144 B = 132	A = 160 B = 148
A = 150 B = 138	A = 142 B = 130	A = 155 B = 143	A = 148 B = 136	A = 164 B = 152	A = 145 B = 133
A = 139 B = 127	A = 155 B = 143	A = 139 B = 127	A = 159 B = 147	A = 149 B = 137	A = 165 B = 153
A = 155 B = 143	A = 138 B = 126	A = 150 B = 138	A = 149 B = 137	A = 167 B = 145	A = 148 B = 136

Alternating Assignment

Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is -1.20 with an associated two sided P-value of 0.24. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. If you are using the conservative degrees of freedom, min(n_1-1, n_2-1) = 17, the P-value would be 0.25. Although variety B has a slightly larger mean yield, there is not a statistically significant difference between the sample mean yields for the two varieties.

Random Assignment

How one randomly assigns varieties to plots is a good class discussion question. Some students might suggest flipping a coin for each plot; heads = A and tails = B. This is random but will not assure 18 plots with variety A and 18 with variety B.

One way to randomly assign the varieties to the plots is to use a die.

—Roll the die, this will give the row number for the plot

—Roll the die again, this will give the column number for the plot

—Assign variety A to the plot with the (row,column) numbers from above

—Repeat the steps above until 18 plots have variety A

—Fill in the remaining 18 plots with variety B

Another way to randomly assign the varieties to plots is to use the TI-83 to generate a random assignment. Essentially what we want to do is to select 18 of the 36 plots at random to receive variety A. The remaining 18 plots (they are chosen at random by default) will receive variety B. To do this first label the plots sequentially from 1 to 36. Then using the TI-83 calculator;

—Put the numbers 1, 2, 3,..., 36 in L1.

—Generate 36 uniform random digits in L2. Math → PRB → 1:rand → ENTER rand (36) → STO → L2

—Arrange L2 in ascending order while carrying the entries from L1 along. 2nd → LIST → OPS → 1:SortA(→ ENTER SortA(L2, L1) → ENTER

—Read off the first 18 numbers in list L1. These plot numbers will receive variety A. The remaining plot numbers will receive variety B.

Example Randomization with yields

A 130	A 149	A 139	A 167	B 137	A 157
A 149	B 121	B 140	B 131	A 159	B 136
A 141	A 156	A 137	A 158	A 144	A 160
A 150	B 130	B 143	B 136	B 152	B 133
B 127	A 155	B 127	A 159	A 149	B 153
B 143	B 126	B 138	A 149	B 145	B 136

Summary	n	mean	std. Dev
A	18	150.4	9.49
B	18	136.3	8.74

Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is 4.64 with an associated two sided P-value that is virtually zero. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. Even using the conservative degrees of freedom, min(n_1-1, n_2-1) = 17, the P-value is virtually zero. Varieties have different mean yields and that difference is statistically significant.

Closer examination of "THE TRUTH" reveals that variety A has a yield that is 12 higher than variety B on every plot. The true difference in yield between variety A and variety B is 12 bushels per acre.

■ 4. Discussion

Assignment by convenience or using an alternating pattern failed to uncover the true difference between the two varieties. “THE TRUTH” was set up in such a way that the convenience pattern and alternating pattern would mislead the experimenter. If you look closely at “THE TRUTH” you will see that there are alternating high/low yield gradients running diagonally across the field. By planting one variety on one side of the field, or in the alternating pattern, the superiority of variety A is hidden by these diagonal yield gradients. In real fields the truth is not known but non-random assignment of varieties to plots can mislead the experimenter in much the same way. The hidden patterns in real fields can confound the effects of the varieties.

Randomization, the random assignment of varieties to plots tends to take hidden patterns (or lurking variables) and spread their effects evenly across the treatment groups. This allows us to see the underlying truth most of the time. This disclaimer, “most of the time,” is important. Even with randomization, we are not guaranteed to find a statistically significant difference even when a real difference does exist. In fact, the chance that a test of hypothesis can detect a difference when one exists is called the power of the test. By looking at the results of tests based on many random assignments, this activity can be used to simulate the power of the two sample t-test to detect a difference in mean yield of 12 bushels per acre. When this randomization activity was done by 40 AP statistics teachers at a short course, all but one of the teachers obtained a t-test statistic that was statistically significant. That is, the simulated power was 39 out of 40 or 97.5%.

■ 5. More on Power

Let’s look at power in a little more detail. What we would like to know is of all the possible randomizations of varieties to plots how many would produce a significant difference in sample mean yields? There are over 9 billion possible randomizations so enumerating all of them is out of the question. We can tackle this problem theoretically with some simplifying assumptions. For the two sample problem, it is easiest to look at power assuming normally distributed values with a common, and known variance. For the corn yield example we might assume that the yields for variety A are normally distributed with a mean μ_A and variance σ^2=87. Additionally, let’s assume that the yields for variety B are normally distributed with a mean μ_B and variance σ^2=87. The value 87 for the population variance is obtained from the values reported in “THE TRUTH.” We need to first establish what is a statistically significant difference. To do this we can use the 68-95-99.7 (or empirical) rule. Recall that approximately 95% of normally distributed values are within 2 standard deviations of the mean. So any difference whose absolute value is greater than 2 standard deviations is statistically significant at approximately the 5% level. We have the variances for individual yields but we need the variance (to get the standard deviation) of the difference in sample mean yields.

Sample mean yields (n=18) for variety A will be normally distributed with a center at μ_A and a variance

$$\frac{\sigma^2}{n} = \frac{87}{18} = 4.833$$

Similarly, sample mean yields (n=18) for variety B will be normally distributed with a center at μ_B and variance

$$\frac{\sigma^2}{n} = \frac{87}{18} = 4.833$$

The difference in sample mean yields will be normally distributed with a center at μ_A - μ_B and a variance of

$$\frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{87}{18} + \frac{87}{18} = 9.667$$

Thus the standard deviation for the difference in two sample mean yields (n=18) is

$$\sqrt{9.667} = 3.11$$

Any absolute difference in sample mean yields larger than two standard deviations (6.22) would be considered statistically significant.

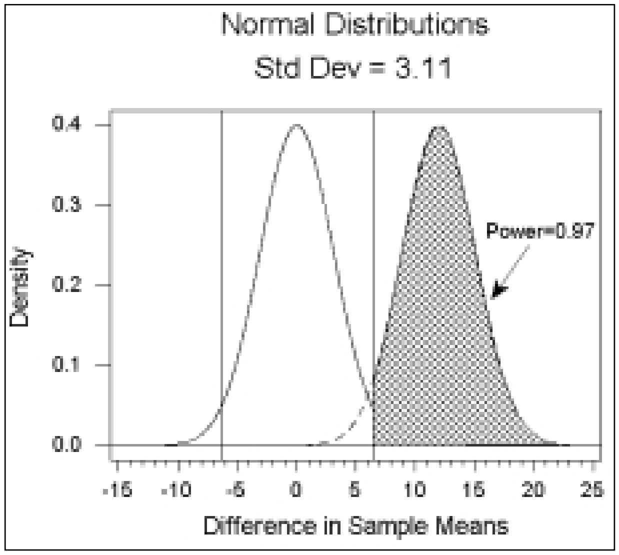
To calculate the power all we would need to do is to compute the probability of getting a difference in sample mean yields that is less than –6.22 or greater than 6.22 when we assume the true difference in means $\mu_A - \mu_B = 12$. This is just the probability that a normal random variable with mean 12 and standard deviation 3.11 takes on a value less than –6.22$ or greater than 6.22. We can obtain the standardized values

$$z_1 = \frac{-6.22 - 12}{3.11} = -5.86$$

$$z_2 = \frac{6.22 - 12}{3.11} = -1.86$$

The normal cumulative distribution function (cdf) for z_1 is zero and so contributes nothing to the power calculation. The cdf for z_2 is 0.03, so the chance of being greater than z_2 =–1.86, and thus the power, is 1 – .03 = .97. The computation of the power is illustrated in the figure below.

Power is actually a function of how big a difference you want to detect. In the calculation above, the true difference of 12 will be picked up most of the time by a two independent sample test when randomization is used to assign varieties to plots. The power will be much lower for smaller true differences. You can adjust "THE TRUTH" so that variety A beats variety B by say 6 bushels. You will find that the power as calculated above (think about moving the right hand normal curve in the figure above so that it is centered at 6 instead of 12) is less than before (around 0.50). Power is clearly a function of the size of the true difference. Procedures have more power to detect large differences than small differences. Power is also affected by sample size. We know that larger sample sizes are good because they reduce the variation in the sample mean. It is nice to know that larger sample sizes also provide more power for much the same reason. Think about how the graph above would change if the sample size, the number of plots receiving each variety increased.



Continued from page 13

Salter, L. (1988) *Mandated Science: Science and Scientists in the Making of Standards*. Dordrecht, Holland: Kluwer Academic Publishers.

Scott, M., Cutter, S.L., Menzel, C., and Ji, M. (1997) "Spatial accuracy of the EPA's environmental hazards databases and their use in environmental equity analysis". *Applied Geographic Studies*. 1, 45–61.

Sexton, K. Olden, K., and Johnson, B.L. (1993) "'Environmental justice': the central role of research in establishing a credible scientific foundation for informed decision making". *Toxicology and Industrial Health*, 9, 685–727.

Sheppard, E., Leitner, H., McMaster, R.B., and Tian, H. (1999) "GIS-based measures of environmental equity: exploring their sensitivity and significance". *Journal of Exposure Analysis and Environmental Epidemiology*. 9, 18–28.

Szwarcwald, C.L., Bastos, F.I., Viacava, F., and Andrade, C.L.T. (1999) "Income inequality and homicide rates in Rio de Janeiro, Brazil". *American Journal of Public Health*. 89, 845–850.

Talen, E., and Anselin, L. (1998) "Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds". *Environment and Planning A*. 30, 595–613.

United Church of Christ (1987) *Toxic Wastes and Race in the United States: A National Report on the Racial and Socio-Economic Characteristics of Communities with Hazardous Waste Sites*. New York: United Church of Christ.

Wagener, D.K. and Williams, D.R. (1993) "Equity in environmental health: data collection and interpretation issues". *Toxicology and Industrial Health*, 9, 775–795.

Wallace, L.A. (1995) "The exposure of the general population to benzene". *Cell Biology and Toxicology*. 5, 297–314.

Waller, L.A., Louis, T.A. and Carlin, B.P. (1997) "Bayes methods for combining disease and exposure data in assessing environmental justice". *Environmental and Ecological Statistics*. 4, 267–281.

Waller, L.A., Louis, T.A., and Carlin, B.P. (1999) "Environmental justice and statistical summaries of differences in exposure distributions". *Journal of Exposure Analysis and Environmental Epidemiology*. 9, 56–65.