

- Harris Z (1951) *Structural Linguistics*. Chicago, IL: University of Chicago Press.
- Lauritzen SL (1996) *Graphical Models*. Oxford, UK: Clarendon Press.
- Shannon CE (1948) A mathematical theory of communication. *The Bell System Technical Journal* 27(3–4): 379–423, 623–656.
- Shannon CE (1951) Prediction and entropy of printed English. *The Bell System Technical Journal* 30: 50–64.
- Stolz W (1965) A probabilistic procedure for grouping words into phrases. *Language and Speech* 8: 219–325.

Further Reading

- Charniak E (1993) *Statistical Language Learning*. Cambridge, MA: MIT Press.
- Jelinek F (1997) *Statistical Methods for Speech Recognition*. Cambridge, MA: MIT Press.
- Jurafsky D and Martin JH (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech*. Upper Saddle River, NJ: Prentice-Hall.
- Manning CD and Schütze H (1999) *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Statistical Methods: Overview

Introductory article

John K Kruschke, Indiana University, Bloomington, Indiana, USA

CONTENTS

The purpose of inferential statistics
Population models with discrete categories
Population models with continuous variables

Resampling methods
Bayesian approaches
Conclusion

Statistical methods specify how confident we can be to confirm or deny a claim.

THE PURPOSE OF INFERENCE STATISTICS

Our knowledge about the world is based on a finite set of observations of it. We cannot know with certainty about the true underlying state of the world; instead, we must infer what might be true, or what is probably false, given the limited data we have. The purpose of inferential statistics is to give precision to our confidence in claims about the world.

Examples of the Need for Inferential Statistics

Suppose that someone claims that Monica is a better tennis player than Conchita. How can the claim be tested? If we could somehow directly ascertain the absolute expertise of both players, by applying some kind of ‘expertisometer’ to their foreheads, then we would have proof about the claim, either for or against. But we have no such

direct method. Instead, we can only let the players compete in a finite number of games, and observe which player wins more games.

Suppose the competitors play a single game, and Monica wins. Does this outcome prove that Monica is better? No, of course not. There are numerous random influences in any given game, such as wind gusts, noises from the audience, small bumps on the court surface, and so on. To try to equalize the influence of these random factors on both players, we have them play several games on different courts.

Suppose that, after seven games, Monica has won four. Is Monica definitely the better player? A supporter of Conchita could argue that Monica just got lucky and happened to win four out of seven, even though Conchita is actually just as good a player. But if this were the case, just how lucky did Monica get? What is the probability that Monica could win four games out of seven, despite being no better a player? If the probability is fairly high, then Conchita’s supporters can retain their hope. But if the probability is extremely low, then they may have to admit that Monica is the better player.

The role of statistical inference is to determine the probability of observable events (e.g., the probability of winning four games out of seven) from hypothesized underlying states of the world (e.g., Monica and Conchita having the same level of expertise). If the hypothesized state of the world cannot easily generate the observed events, then the hypothesis is probably wrong.

Most claims about the world are of the same status as claims about tennis expertise. Consider claims that might be made in cognitive science, such as ‘my dog knows the command “stay”’. The same claim might be made instead of a robot, or of a human participant in a psycholinguistics experiment. How do we test such claims? We can only say the word ‘stay’ several times and observe what the listener does, whether the listener is a robot, a dog, or a human. From these observations, we infer whether it is tenable to claim that the listener knows the command.

The Logic of Hypothesis Testing: Population and Sampling Distributions

Suppose we make a claim about the world. This claim amounts to a specific description of the underlying population from which our observations are sampled. Given this model of the population, we determine the probability of getting any particular sample of observations. That is, we determine the distribution of samples that we would get from the hypothesized population. From this sampling distribution, we can ascertain the kind of sample we should expect to obtain, and we can determine the specific probabilities of getting samples that deviate from the expected sample. (Examples of this process are given below.) If our actual set of observations deviates wildly from the expected sample, then the hypothesized population is probably not a good description of the actual world. Standard statistical inference is just the mathematical formalization of this logic for different types of populations.

POPULATION MODELS WITH DISCRETE CATEGORIES

In many situations we are concerned with discrete states of the world. Familiar examples include a tossed coin coming up heads or tails (two discrete states), or the hair color of a randomly selected person being black, brown, blond or red (four discrete states). This type of situation can be contrasted with situations in which the states of the world are ordered, or ranked (e.g., grades in an

examination) or interval-scaled (e.g., response times). Situations with scaled (continuous) variables are discussed later in this article.

The Binomial Sampling Distribution

Suppose we want to test somebody’s claim that a certain coin is fair; i.e., that it comes up heads 50% of the time. This is a situation in which the world has two states: heads or tails. If the coin is fair, then when we toss it a number of times, it should come up heads in about half of those tosses. For example, if we toss the coin eight times, it should come up heads about four times. Suppose that it actually comes up heads six times out of eight tosses. Is this deviation from the expected four heads large enough for us to reject the hypothesis that the coin is fair? What if the coin comes up heads seven times out of eight, or eight times out of eight? By how much can the sample deviate from what we expect before we reject the claim that the coin is fair?

To answer this question, we must calculate the probability that a fair coin would generate so many heads in eight tosses. Figure 1 shows these probabilities. The population distribution at the left assumes a fair coin; i.e., a coin for which the probability of heads is 50%. The graph shows two bars, one labeled ‘H’ for heads and the other labeled ‘T’ for tails. The height of each bar is 0.5, indicating that we are supposing that the probability of getting a head is 0.5 and the probability of getting a tail is also 0.5. Imagine repeatedly tossing the coin eight times, as exemplified in the middle of Figure 1. Usually a fair coin will come up heads about four times out of eight, but occasionally we would expect more extreme outcomes. The exact probability of each possible outcome can be mathematically derived, and the result, called a *binomial distribution*, is plotted in the graph on the right side of Figure 1. We see that the probability of getting four heads out of eight tosses is about 27%, the probability of getting five heads is about 22%, the probability of getting six heads is about 11%, and so on. Thus, a hypothesis about something we cannot directly observe (i.e., the fairness of the coin) makes specific predictions about the probabilities of what we can observe (i.e., getting so many heads out of eight tosses).

Suppose we now actually toss the suspected coin eight times, and the result is that it comes up heads seven times. From the binomial sampling distribution, we see that the probability of getting seven heads is about 3%. The probability of getting eight heads is less than 1%. Thus, the probability of getting a result this deviant or more from the expected outcome is less than 4%.

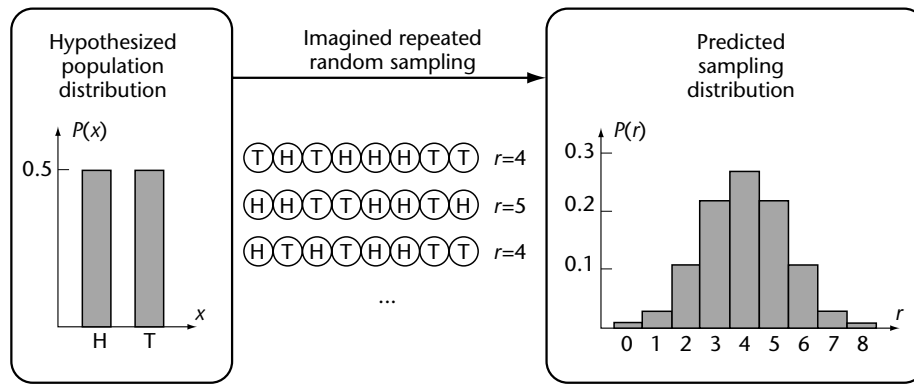


Figure 1. Illustration of a binomial sampling distribution generated by a fair coin tossed eight times. The predicted sampling distribution on the right shows the probability of getting r heads, for values of r between 0 and 8.

Therefore, because the result is so unlikely if the coin is fair, we reject the hypothesis that the coin is fair. The coin still might be fair, but it is rather unlikely given the result of seven heads out of eight flips. Note also that this is a 'one-tailed' test. A two-tailed test (in which we would have to add the probabilities of getting no heads or one head) is generally more appropriate in this and other similar situations.

The Chi-squared Distribution

An analogous procedure is applied when we have a situation with more than two discrete states. For example, suppose we want to know if four hair colors – black, brown, blond and red – are equally likely to occur in the general population. We hypothesize that each of the four colors has a probability of 0.25 in the population. We then imagine repeatedly sampling from this population. Suppose we select 40 people at random in each sample. If the four colors are equally likely, then we would expect to get about 10 people of each hair color, but by chance we would often get a few more or a few less of each color. For any given sample, a number called chi-squared (χ^2) describes how much the sample deviates from what we expect. For every possible sample of 40 people, there is a corresponding value of chi-squared which measures how deviant that sample is from the expected 10 people in each category. In most cases the samples will have moderate chi-squared values, but on rare occasions the samples will have extreme chi-squared values. The hypothesis of equally likely colors thereby generates a predicted sampling distribution for the chi-squared values. This sampling distribution shows the probability of getting each possible chi-squared value from the hypothesized population.

We now get an actual sample of 40 randomly selected people, and we determine the chi-squared value of this sample. If this actual chi-squared value is very extreme in the predicted sampling distribution, we reject the hypothesis.

The chi-squared method can also be applied to the situation of just two discrete states, and therefore subsumes the binomial distribution. Nevertheless, the binomial is often used in research because of its simplicity.

POPULATION MODELS WITH CONTINUOUS VARIABLES

The methods described above dealt with discrete variables; e.g. heads versus tails or category of hair color. The same logic can be applied to continuously scaled variables such as temperature, response time, blood pressure, etc.

The t Distribution

Suppose someone claims that a certain drug decreases blood pressure. (Notice that blood pressure is measured on a scale with meaningful intervals, not just discrete categories such as yellow or green.) How do we test this claim about the drug? One approach is to randomly assign people either to a group that takes the drug or to a comparison group that does not take the drug. (Usually the comparison group will be given some innocuous replacement substance, called a placebo, in a so-called 'double blind' procedure, so that neither the subject nor the person administering the drug knows whether it is really the drug or the placebo. This equalizes any psychosomatic effects in the two groups.) Then we measure the blood pressure of everyone in the two groups, and compare the mean

blood pressures of the two groups. If the drug-treated group has lower mean blood pressure than the control group, are we certain that the drug works as claimed? Not necessarily. It could be that the drug has no real effect, but that the people in the drug-treated group happened to have lower blood pressure because of other random influences. So we need to determine just how probable it would be for the drug-treated group to have a mean that much lower, if the actual effect of the drug were zero.

First we create a model of the hypothesized population in which the drug has no overall effect. We conceive of the control group as having a distribution of blood pressures, some higher than average and some lower. We conceive of the drug-treated group as having the very same distribution of blood pressures: that is, the drug has no effect.

What shape of distribution should we assume? The usual assumption is a normal distribution – the well-known bell curve, as shown in the left panel of Figure 2. There are several motivations for using the normal distribution. One is that it has convenient mathematical properties for deriving probabilities in the sampling distribution. Another is that the central limit theorem shows that many different distributions become approximately normal when the processes that generate them are repeated many times and the results averaged. Finally, many distributions in the real world are approximately normal. The normal curve shows the relative probability of each blood pressure. Thus, the most probable blood pressure is the value under the peak of the normal curve. Blood pressures higher or lower than this average are less probable.

Having hypothesized this model of the population, we then imagine repeatedly sampling from it. Suppose we randomly select four people from each

group. The mean of each sample will be about the same as the mean of the population, but the two means will not be exactly the same because of the random sampling. This difference between the group means is not very meaningful by itself, because its value depends on the scale units in which it is measured (e.g., a temperature difference of 3 degrees has different meanings depending on whether the scale is Fahrenheit or Celsius). Therefore the difference between group means is divided by the expected dispersion of this difference across samplings, and the resulting ratio is called the t statistic. Usually the difference between groups means, and hence the value of t , will be close to zero, on the hypothesis that we are sampling from a population with identical groups. Sometimes, however, it will be significantly above or below zero. The sampling distribution for t is illustrated in the right panel of Figure 2. Thus, a hypothesis about something we cannot directly observe (i.e., the equality of the distributions for the two groups) makes specific predictions about the probabilities of what we can observe – values of the t statistic.

Suppose we now actually test the drug on four randomly selected people, with four other randomly selected people getting the placebo. Suppose that we get a t value of 2.45. This value implies that the difference between the group means is fairly large compared with the expected variation of this difference. From the sampling distribution of t , it can be determined that the probability of getting a t value this extreme is less than 5%. Therefore, because this extreme value of t is so unlikely if the two groups have the same underlying population, we reject the hypothesis that the drug has no effect. Note that the drug might nevertheless have no effect, but it is rather unlikely given the result that $t = 2.45$.

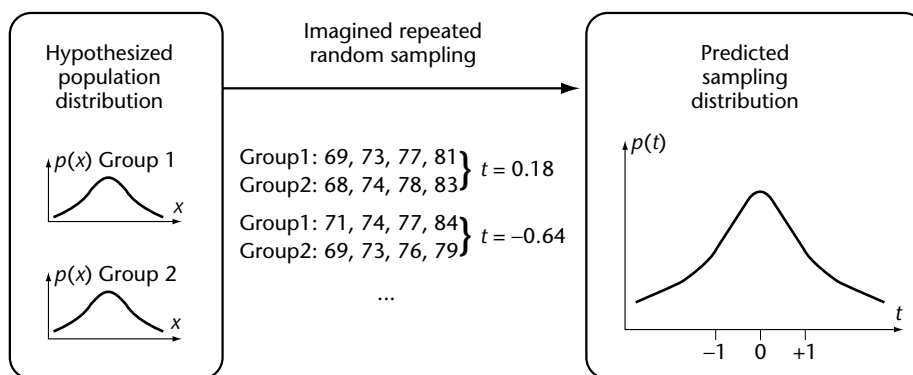


Figure 2. Illustration of the t sampling distribution generated by comparing four scores sampled from each of two identical normal distributions.

The F Distribution and Analysis of Variance

An analogous method can be applied when there are more than two groups. For example, there might be several different drugs that are claimed to be effective in reducing blood pressure. To test whether the drugged and the control groups differ in their overall mean blood pressures, we can measure the overall dispersion among the group means relative to the dispersion within the groups. This ratio is called the F statistic. The logic of using it is the same as for other statistics. First, hypothesize identical normal populations for the groups. Then imagine repeated random sampling from these populations, each time computing the value of the F statistic. This generates a predicted sampling distribution for F . Now get actual random samples of blood pressures from people who have taken the various drugs or the placebo, and compute the F statistic. If the actual F statistic deviates greatly from what would be expected from the hypothesis of equality, then we reject that hypothesis.

This procedure is called *analysis of variance* (ANOVA) because the overall variance among the individual scores is decomposed (analyzed) into (1) the variance between group means and (2) the variance within groups. The ratio of these two components of the overall variance is the F statistic.

ANOVA can be applied when there are just two groups, and so it subsumes the t statistic. Nevertheless the t statistic is often used in research because of its simplicity.

Linear Regression and the General Linear Model

Many types of claims can be expressed as mathematical hypotheses about a population. We can imagine more and more complicated hypotheses, yet apply the same logic as in the previously discussed situations.

For example, suppose that we believe that the blood pressure drug should yield a decrease in blood pressure proportional to the amount of drug administered. We might also imagine that exercise increases blood pressure, in proportion to the rate of exertion. This type of hypothesis can be expressed mathematically as a so-called *linear model*. Denote predicted blood pressure by \hat{y} , drug quantity by x_d , and exercise rate by x_r . Then a linear relationship between blood pressure, drug and exercise can be described as $\hat{y} = \beta_0 + \beta_d x_d + \beta_r x_r$, where β_0 is base-line blood pressure and β_d and β_r are numbers that reflect how big an influence each

factor (drug and exercise) has on blood pressure. If β_d is a large negative number, then predicted blood pressure drops rapidly as drug dosage increases. If β_r is a large positive number, then predicted blood pressure increases rapidly as exercise rate increases.

We can construct a series of hypothetical populations in which one or more of the factors (drug dosage or exercise rate) has no effect. For example, if both β_d and β_r are zero, then we are hypothesizing no effect of the drug and no effect of exercise on blood pressure. We can then imagine repeatedly randomly sampling from this hypothesized population, and for each sample computing the value of a statistic analogous to F . From this process we generate a sampling distribution predicted by the hypothesis. Then we can get an actual sample, and determine whether the hypothesized population could generate this actual sample with reasonable probability.

If the hypothesis that neither factor has an effect can be rejected, then we can hypothesize that one or both of the factors does have an influence; that is, we allow one or both of β_d and β_r to be nonzero. Suppose we hypothesize that the drug has no effect, so $\beta_d = 0$, but exercise does have an effect, so $\beta_r \neq 0$. We want to test whether this hypothesis can be rejected, but what should be the exact value of β_r for our population model? We set it to whatever value best fits the data, to give the hypothesis its best chance of being confirmed. The degree to which a model fits data is usually measured as the 'sum squared error' $\sum_i (y_i - \hat{y}_i)^2$ between actual data points y_i and predicted values \hat{y}_i , where \sum_i indicates summation over all the data points. The value of β_r that minimizes this quantity is used for the population hypothesis. Then we proceed with the imagined sampling from the hypothesized population, to assess the probability that our actual data could have come from such an underlying population.

This type of procedure, using linear models, is called *multiple linear regression*. It turns out that models of group differences in ANOVA can all be expressed as particular cases of linear models. Therefore, the general linear model subsumes multiple linear regression and ANOVA. In much research, however, ANOVA is used because of its conceptual simplicity.

The Model Comparison Approach

It is often the case that complicated hypotheses about the world can be modeled as extensions of simpler models. More complicated models will

involve more free parameters, like the coefficients in the linear model described above. As parameters are added to a model, it will usually fit data better, but only at the cost of being a more complex model. We want the model to be as simple as possible while also fitting the data well. Does the inclusion of particular new parameters improve the fit of the model significantly?

To answer this question, we use the same logic as has been described above. Assume that the ‘simple’ model, the one with fewer parameters, is a correct description of the world. (The parameter values in this simple model are set to whatever best fits our present data set.) The simple model incorporates an assumption about the distribution of values in the population; typically the population is assumed to be normally distributed around means specified by a linear function. We then imagine repeatedly sampling from this ‘simple’ population. For every sample, we compute the improvement in fit that is gained by including the extra parameters. This measure of improvement in fit is called the *generalized F ratio*. Note that a given sample from the simple population will almost surely be fitted better by a model with more parameters, because the sample will deviate randomly from the simple population’s form, and this random deviation can be partially accommodated by the extra parameters.

The repeated sampling generates a sampling distribution of improvements in fit achieved by the more complex model, even when the simpler model is correct. Once we know this sampling distribution, we can see whether the improvement in fit for our actual data set lies near the expected value of this sampling distribution. If it does, then the simpler model may be adequate. On the other hand, if the actual improvement in fit deviates greatly from what we would expect from the simpler model, then we reject the simpler model.

All of the methods described in this section are special cases of this model comparison approach. The model comparison approach is conceptually powerful and helps unify and generalize the various traditional statistical tests. However, the specific methods, such as t tests, ANOVA, and multiple linear regression, continue to be popular.

RESAMPLING METHODS

All of the methods described so far use the same basic logic: hypothesize a specific form of the population distribution (i.e., start with a model of the population) and then derive the predicted sampling distribution of whatever statistic seems

useful. For example, the t statistic was sampled from normally distributed populations.

But what if we do not have a specific population distribution in mind? For example, what if a normal distribution seems inappropriate for modeling our particular population? We could just hypothesize a different, more appropriate distribution – but what if we have no theoretical commitment to any particular distribution? To answer this question, we ask another question: what is the best information we have about the underlying population? It is our actual sample. So we let the data themselves serve as the population, from which we then resample to generate a sampling distribution.

For example, suppose we are measuring scores on a statistics examination. Three students who read an encyclopedia article about statistics scored 58, 70, and 89. Three other students who did not read the article scored 50, 55, and 68. On average, the group who read the article did better, but maybe this was merely a random effect of the sampling. Maybe the two groups really come from the same underlying population, with no genuine difference between them. If we do not want to assume any specific form for this mutual underlying population, we can just let the six scores themselves be the population from which we sample. So the question becomes this: if we repeatedly sample from this population of six scores, how likely is it that a random sampling would yield a difference between groups as large as what we found in our actual sample? It turns out that there are 20 ways of sampling two groups of three from the six scores, and of these 20, there are 4 ways that yield group differences as large or larger than the one actually obtained. Because 4 out of 20 is not very low, we would probably not reject the hypothesis that the two groups come from the same underlying population.

Resampling methods are becoming increasingly popular because of the increasing availability of computer software. Their main advantage is that they do not require specific assumptions about the shape of the population distribution. A disadvantage is that important statistical concepts (such as power and confidence interval) are difficult to quantify by these methods, because to do so would require specification of alternative hypothetical populations.

BAYESIAN APPROACHES

All the methods described above are based on determining the probability of the data given a

hypothesis. The logic is that if the data are very unlikely to have come from the hypothesized model, then the model is probably wrong. But this logic is incomplete. The probability of data D given a hypothesis H is not necessarily the same as the probability of the hypothesis given the data. In mathematical notation, usually $P(H|D) \neq P(D|H)$, where $P(A|B)$ means the probability of A given B . The methods described above provide $P(D|H)$, but we would really like to know $P(H|D)$.

For example, consider a deck of playing cards. If I select a card at random, and tell you that it is a king, then what is the probability that it is also a heart? The answer is $1/4$, because of the four kings in the deck, one is a heart. If instead I tell you that the card I selected is a heart, then what is the probability that it is also a king? The answer is $1/13$, because of the 13 hearts in the deck, one is a king. Thus $P(\text{heart}|\text{king}) \neq P(\text{king}|\text{heart})$.

The relationship between $P(H|D)$ and $P(D|H)$ is given by Bayes' theorem: $P(H|D) = P(D|H)p(H) / \sum_i P(D|H_i)P(H_i)$, where the terms H_i refer to all possible hypotheses about the situation. At first it might seem impossible to consider all possible hypotheses, and to establish the prior probability $P(H_i)$ that each is true. But in many situations this can be done in reasonable ways.

For example, suppose that we are questioning the fairness of a coin. We believe that there is a 70% chance that the coin is fair; i.e., $P(H_{.50}) = 0.70$. We believe that there is a 25% chance that the coin is biased to yield heads 35% of the time; i.e., $P(H_{.35}) = 0.25$. And we believe that there is a 5% chance that the coin is biased to yield heads 10% of the time; i.e., $P(H_{.10}) = 0.05$. Suppose we actually toss the coin four times, and it comes up heads zero times; this result is our data, denoted D . The binomial distribution indicates that $P(D|H_{.50}) = 0.0625$, $P(D|H_{.35}) = 0.1785$, and $P(D|H_{.10}) = 0.6561$. Thus, the probability that the data could be generated by a fair coin is pretty low, and we might even want to reject the hypothesis that the coin is fair. We also see that the most biased of the three hypotheses has the highest probability of generating the data.

The most biased hypothesis, however, is also the hypothesis that we believed, before flipping the coin, was very unlikely to be true. So how much should we increase our belief in this unlikely hypothesis? Using Bayes' theorem, we obtain $P(H_{.10}|D) = P(D|H_{.10})P(H_{.10}) / [P(D|H_{.10})P(H_{.10}) + P(D|H_{.35})P(H_{.35}) + P(D|H_{.50})P(H_{.50})] = 27\%$. Similarly, we find that $P(H_{0.35}|D) = 37\%$ and $P(H_{0.50}|D) = 36\%$. Thus, before tossing the coin, we thought that $P(H_{.10})$ was only 5%, but after tossing the coin, we think that the probability is 27%, more

than five times as great. Moreover, before tossing the coin, we believed that $P(H_{.50})$ was 70%, but after tossing the coin, we think that the probability is 36%, only about half as likely. Nevertheless, despite the data, the chances that the coin is severely biased are still not as high as the chances that the coin is unbiased, or only moderately biased. To revise our prior beliefs more radically, we would need more convincing data.

There are several advantages of a Bayesian approach to inference. Firstly, it forces the theorist to consider multiple hypotheses and the prior probability of each. This encourages breadth of theorizing. Secondly, it takes into account the prior probabilities of hypotheses, so that bizarre hypotheses with small prior probabilities are unlikely to be accepted even if they could generate the data, and very plausible hypotheses with high prior probabilities are not necessarily rejected even if a particular data set is not easily generated by them. Thirdly, the framework can be extended to take into account the costs or pay-offs of different kinds of correct or incorrect decisions about hypotheses. On the other hand, a disadvantage of Bayesian statistics is that it is often impossible to specify the prior probabilities of all possible hypotheses.

CONCLUSION

For statistical methods to be applied, society must provide two conditions. Firstly, people must have the liberty to doubt claims. If claims about the world are dictated by authorities, without permission to doubt the claims, then there is no room for inferential statistics. Secondly, people must have the liberty to gather data that are randomly sampled and representative of the underlying population about which the claim is made. If a claim is made that Monica is better than Conchita, but the only games you are allowed to observe have the wind against Conchita, then your observations are not a good test of the claim. Thus, without liberty, statistical methods cannot be meaningfully applied, and when statistical methods have been appropriately applied, liberty has been exercised. *Ubi dubium, ibi libertas*: where there is doubt, there is liberty. Doing a statistical test is a political act, and an expression of the culture in which we work.

Liberty should not be confused with license. Scientific research depends on the liberty to apply statistical methods to random, representative samples. But care must be taken that the methodological strictures of statistical method and experimental design do not override the ethical principles

of human beings. The methods of science are an exquisitely refined expression of human intellect combined with human awe and curiosity about the mysteries of nature. Statistical methods must be applied with humility and responsibility; they must not violate the rights of the very beings whose culture created the methods.

Further Reading

Bronowski J (1973) Knowledge or certainty. In: *The Ascent of Man*, chap. XI, pp. 352–377. Boston, MA: Little, Brown.

Huff D (1954) *How to Lie With Statistics*. New York, NY: W. W. Norton.

Judd CM and McClelland GH (1989) *Data Analysis: A Model-Comparison Approach*. San Diego, CA: Harcourt Brace Jovanovich.

Maxwell SE and Delany HD (2000) *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Mahwah, NJ: Lawrence Erlbaum.

Wickens TD (1998) Drawing conclusions from data: statistical methods for coping with uncertainty. In: Scarborough D and Sternberg S (eds) *Methods, Models and Conceptual Issues: An Invitation to Cognitive Science*, vol. IV, pp. 585–634. Cambridge, MA: MIT Press.

Stereotypes

Introductory article

Steven Fein, Williams College, Williamstown, Massachusetts, USA

William von Hippel, University of New South Wales, Sydney, New South Wales, Australia

CONTENTS

Introduction

Stereotype formation

Stereotype maintenance

Stereotype application

Stereotype change

Conclusion

Stereotypes are consensual beliefs about group characteristics that influence the perception, interpretation, and evaluation of others, sometimes blatantly but often in a manner so subtle that they are outside awareness. Because they serve basic cognitive and motivational functions, stereotypes are highly resistant to change.

INTRODUCTION

On 4 February 1999, just after midnight, 22-year-old Amadou Diallo entered his apartment building. He was spotted by members of the Street Crime Unit, an elite corps of New York City police officers who had been extraordinarily successful in reducing crime, but were often criticized for being too aggressive. In particular, in a practice known as ‘racial profiling’, the Street Crime Unit was routinely selecting African American and Hispanic men to be stopped and frisked. That winter night in 1999, Amadou Diallo was targeted to be one of them. Four white, plainclothes police officers from the unit spotted Diallo, who matched the

general description of a suspected rapist. The police thought that Diallo looked suspicious as he appeared to duck into his building to avoid them. As they approached and ordered him to freeze, he reached into his pocket and began to pull out a wallet. Thinking that the wallet was a gun, the police opened fire. Nineteen of their 41 shots hit Diallo, and he lay dead in the vestibule. Diallo was unarmed.

In the days and weeks that followed, numerous protests erupted in New York City and around the country. The controversial use of ‘racial profiling’ came under national attention and renewed attack. The four officers responsible for the shooting of Diallo were acquitted of any wrong-doing, but many felt that their stereotypes of African Americans played a critical role in their misperception of his wallet as a gun, and their decision to fire so many bullets. In contrast, various politicians, columnists, and citizens defended the police, noting how difficult it is to make life-or-death decisions in the blink of an eye. Although we’ll never know what role stereotypes played in the Diallo shooting,