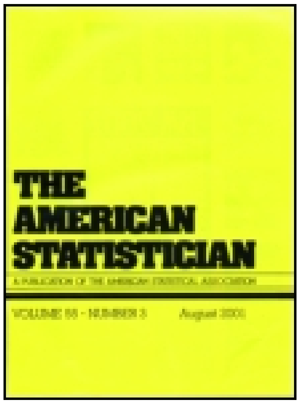
This article was downloaded by: [98.207.180.40]
On: 14 May 2013, At: 07:32
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The American Statistician
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/utas20

Twenty-Five Analogies for Explaining Statistical Concepts

Roberto Behar [a] , Pere Grima [b] & Lluís Marco-Almagro [b]
[a] Statistics at the School of Statistics Universidad del Valle, Cali, Colombia
[b] Statistics at the Universitat Politècnica de Catalunya—BarcelonaTech, Barcelona, Spain
Accepted author version posted online: 11 Dec 2012.Published online: 20 Feb 2013.

To cite this article: Roberto Behar , Pere Grima & Lluís Marco-Almagro (2013): Twenty-Five Analogies for Explaining Statistical Concepts, The American Statistician, 67:1, 44-48

To link to this article: http://dx.doi.org/10.1080/00031305.2012.752408

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://amstat.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Twenty-Five Analogies for Explaining Statistical Concepts

Roberto BEHAR, Pere GRIMA, and Lluís MARCO-ALMAGRO

The use of analogies is a resource that can be used for transmitting concepts and making classes more enjoyable. This article presents 25 analogies that we use in our introductory statistical courses for introducing concepts and clarifying possible doubts. We have found that these analogies draw students' attention and reinforce the ideas that we want to transmit.

KEY WORDS: Concepts through examples; Statistical education; Students' motivation; Teaching statistics; Undergraduate courses.

Downloaded by [98.207.180.40] at 07:32 14 May 2013

1. INTRODUCTION

The use of well-told analogies, anecdotes, and jokes is a resource that teachers use to reinforce the transmission of ideas and concepts. What is more, when used correctly, they bring an informal tone to the class which becomes more enjoyable and sparks student interest.

Recognizing the importance of analogies in the teaching and learning process is not new. Donnelly and McDaniel (1993, 2000) highlighted the possibility of using analogies for teaching in general and the excellent article of Martin (2003) documents their use in teaching statistics. There are articles with compilations of analogies (Chanter 1983; Brewer 1989) and articles dealing with one specific analogy (Feinberg 1971) about Type I and Type II errors and the judicial process. Gelman and Nolan (2002) proposed many excellent ideas for teaching statistics; Cleary (2005) published a review of that book in this very journal (with an experiment included, although the deadline for participating has surely passed).

The analogies that we present here are those that we use in our introductory courses. We think they can fit especially well in what Meng (2009) calls *happy courses*: "introductory courses that truly inspire students to *happily* learn statistics as a way of scientific thinking for whatever they do." We use some of the analogies consistently for reinforcing concepts that we are introducing: each of us has his favorites and those always appear when appropriate. Some others are used or not depending on how the lecture evolves. Sometimes a group of students have particular difficulties in understanding an idea, and an analogy can be useful to clarify it. Other times we use a suitable analogy as the answer of a question posed by a student.

Our list of analogies includes some well-known examples (like the courtroom analogy for illustrating the idea of hypothesis testing), but most of them are generally unknown; they are either our own creation or we have heard them.

Roberto Behar is Professor of Statistics at the School of Statistics Universidad del Valle, Cali, Colombia (E-mail: *roberto.behar@correounivalle.edu.co*). Pere Grima (E-mail: *pere.grima@upc.edu*) is Professor of Statistics and Lluís Marco-Almagro (E-mail: *lluis.marco@upc.edu*) is Associate Professor of Statistics at the Universitat Politècnica de Catalunya—BarcelonaTech, Barcelona, Spain. The authors thank Luis A. Escobar, Xavier Tort-Martorell, the Editor, Associate Editor, and two anonymous referees for helpful comments that resulted in substantial improvements to the article.

2. ANALOGIES

2.1 Statistics is Much More Than Percentages, Sports Averages, and Election Polls: Iceberg

Statistics is often confused with those aspects that are most spoken of, such as those that appear in the media. We can compare this vision of statistics with the vision of an iceberg: you see only a small part. Besides percentages or election polls, statistics plays a fundamental role in many fields of knowledge, such as quality control, the development of new medicines, marketing research, sociological studies, economic indicators, and so on. But these are parts that remain hidden to most people.

2.2 The Role of the Values We Use (Mean, Standard Deviation, . . .) in the Numerical Synthesis of Data: Police Sketch

How does one describe a face to make a police sketch? The untrained person will surely provide information that is vague, confusing, and of little help in drafting of the face being described. The police are, however, trained to focus on key elements and they know the language for describing them. It is similar to the numerical synthesis of data. We choose the measures that best describe the information in the overall dataset, and we should be able to understand and correctly interpret its values to form a reliable idea about the information contained within the data.

2.3 The Average is Insufficient for Describing the Data: The Height of Martians

If we know that Martians have an average height of 50 inches, are they taller or shorter than us Earthlings? They are not

 © 2013 American Statistical Association DOI: 10.1080/00031305.2012.752408

necessarily shorter; it could be that some are only a few inches tall and the majority is over 80 inches.

2.4 Using Only the Average Can Lead to Serious Confusion: Jokes About Averages

(a) If we are about to cross a river and we are told that the average depth is 3 feet, that does not mean we can relax. It is possible that a large section is 1.5-feet deep while another part is 10-feet deep, where we could drown.
(b) If one man eats a chicken and another man none, the average chicken that each man has eaten is one half. When we talk about a ranking of countries using per capita income, we are talking about the half chicken that each person eats but it remains unclear how many starve.
(c) If you enter the kitchen and put your head in the oven and your feet in the refrigerator, your body will be at the ideal average temperature.
(d) A statistician goes hunting with two mathematicians. They spot a duck. The first mathematician levels his rifle, fires, and misses to the right. The second mathematician levels his rifle, fires, and misses to the left. The statistician turns to his friends and says "looks to me like we got him, boys."

2.5 Properties of the Arithmetic Average: Fulcrum That Balances

When playing with a young child on a teeter-totter, it is not balanced if both the adult and the child sit at the edge of it. The adult weighs a lot more than the child, so the adult must sit closer to the fulcrum to balance the teeter-totter. We can see a dotplot as a teeter-tooter with each point having the same weight. The fulcrum where the dotplot (teeter-totter) is balanced is the average of the data.

2.6 Interpreting Standard Deviation in the Context of Normal Distribution: A Basketball Game Between Earthlings and Martians

If the heights of Martians follow a normal distribution with a mean of 55 inches and a standard deviation of 8 inches and they play a game against the Earthling team, who will win? The Earthling basketball players will be around 80 inches tall, which is to say that they are 4 standard deviations from the average (supposing that our heights are normally distributed with mean 68 inches and a standard deviation of 3 inches). If the Martians also choose from among the tallest of their species (those being 4 standard deviations from the average), their players will be around 87 inches tall, and thanks to this height advantage, they will probably beat us (although their average height is less).

2.7 That a Value Should be Considered an Anomaly Does Not Depend Only on Its Magnitude: The Velocity of Vehicles on a Highway

To study the effect of one type of speed limit sign on the velocity of cars traveling on a section of highway, a hidden radar is installed to measure the speed of each vehicle. The average car travels at about 50 mph. One vehicle passes at 10 mph and another at 90 mph. Although both are at equal extremes (they are at the same distance from the average value), the one that travels at 10 mph should not be considered in the study because it is an agricultural vehicle that cannot go faster and the signal was not intended for it. On the other hand, the car traveling at 90 mph paid no attention to the sign and therefore should not be eliminated because this information is relevant to the study.

2.8 We Can Be More Specific in Saying That Something Depends on Chance: Types of Songs

People who know nothing about Latin music cannot distinguish between a rumba, a salsa, or a bolero: for them, these are just "songs." However, the connoisseur knows how to distinguish the different styles: when told that the next song being played is a bolero, she or he already knows its rhythm, melody, and theme. Similarly, to many people, all random variables look alike. However, a person with knowledge of statistics can assign each variable to a family of variability (normal, binomial, Poisson, etc.) and thus anticipate a lot of its properties.

2.9 All Models are Theoretical: There Are No Perfect Spheres in the Universe

It appears that the most common geometric form in the universe is the sphere. But how many mathematically perfect spheres are there in the universe? The answer is none. Neither the Earth, nor the Sun, nor a billiard ball is a perfect sphere. So, if there are no true spheres, what good are the formulas for ascertaining the area or volume of a sphere? So it is with statistical models in general and, in particular, with a normal distribution. Although one of the most commonplace examples is height distribution, if we were to have at our disposal the height of every adult on the planet, the histogram profile would not correspond to a Gaussian bell curve, not even if the data were stratified by gender, race, or any other characteristic. But the normal distribution model still provides approximate results that are good enough for practical purposes.

2.10 Adding up Random *k* Values With the Same Probability Distribution is Not the Same as Multiplying One of Them by *k*: The Weight of a Dozen Eggs

The weight of a dozen eggs (adding up the values of 12 random variables with the same distribution) presents some variability. The weight of an egg chosen by chance and multiplied by 12 (a random variable value multiplied by 12) presents more variability since the chosen egg may be large and it will be as if all the eggs are large, and if the random egg is small, it will be as if all the eggs are small. These are extreme situations that will not occur if 12 eggs are chosen at random.

2.11 Bias and Precision of an Estimator: Impacts on a Target

The center of a target corresponds to the value of an estimated parameter and the impacts are the values provided by the estimator. If the estimator is unbiased, all the impacts will be around the center, and the lower the variability is, the more

Downloaded by [98.207.180.40] at 07:32 14 May 2013

grouped the impacts appear. This image also underscores the fact that an unbiased estimator is not always the best option: it could be better if the impacts are not around the center of the target but have very small variability than if they are around the center but a lot more dispersed.

2.12 It is Not About Demonstrating the Null Hypothesis: Trials

In a trial, the null hypothesis is innocence. The objective is not to demonstrate that the accused is innocent but to see if the evidence (the data) contradicts this hypothesis. If there is no evidence, the accused cannot be declared guilty, but this does not mean innocence has been proven.

2.13 The 5% *p*-Value as a Boundary Between the Usual and Unusual is an Arbitrary Value: The Fingers of the Hand

The 5% *p*-value has been consolidated in many environments as a boundary for whether or not to reject the null hypothesis with its sole merit of being a round number. If each of our hands had six fingers, or four, these would perhaps be the boundary values between the usual and unusual.

2.14 Taking all Decisions With the Same Probability of Error is Not Reasonable: Forgetting an Umbrella or Driving on the Left Over a Blind Hill

If you leave the house one morning and shortly afterward find out that the probability of rain is 10%, you may decide not to return home for an umbrella. The probability of an error in this decision is 10%, but nobody would accuse you of being foolish. But if you were driving down an infrequently used road and came upon a blind hill with a pothole in your lane, would you drive on the left to avoid it? Few cars use this road and there is a low probability that another oncoming car will be in the left lane as you pass, but you most likely would not drive on the left because the consequences of an error would be extremely grave. It is not sensible to unify the probability of error when making decisions. In some cases, 10% is reasonable, but in others, not even 1 in 1000 is acceptable.

2.15 Using Reference Distributions: Screening X Rays

When a doctor looks at a presumably healthy patient's X ray to see if there is something abnormal, what he or she does is mentally compare it with the X ray of a healthy person, taking into account that not all healthy people have exactly the same X ray. The same is true when comparing a test statistic with its reference distribution in hypothesis testing. The reference distribution is the set of X rays of healthy people that the doctor has in mind, and the test statistic is the patient's X ray. If the patient's X ray looks normal in its reference distribution, the doctor will say that everything is fine (he cannot think the opposite with the available information). Otherwise, the doctor will say that there is something strange, something not normal for a healthy person. There will also be cases where the doctor will doubt (it could be normal but is not frequent) and will ask for additional tests. Luckily, when the reference distribution is a known probability distribution and the test statistic is a value, it is possible to quantify the degree of compatibility through the *p*-value. For doctors, the quantification of their possible doubts is not so easy.

2.16 Effect of Sample Size on the Comparison of Treatments: Magnification of Binoculars

In the distance, we see two animals but we do not know if they are two dogs or a dog and a cat. If we use binoculars with very low magnification (low sample size), we are not able to ascertain whether or not they are both dogs. With an increase in magnification, we can distinguish what they are. If we use a telescope that provides a much greater magnification, we may see in great detail the whiskers of both animals, which are certain to be different although both are dogs. With enough magnification (sample size), we will always see differences even though the animals are the same type. A significant difference (we are sure that the whiskers are different) may be irrelevant to what matters.

2.17 The Concept of Confidence Level (1): A Person Who Tells the Truth 95% of the Time

An exact 95% confidence interval is calculated such that it includes the true value of the estimated parameter 95% of the time. We do not know, however, if the interval we have is one of those that are correct or not. It is like a person who tells the truth 95% of the time, but we do not know whether a particular statement is true or not.

2.18 The Concept of Confidence Level (2): Number of Computers That Estimate Correctly

Students are asked in a computer lab class to generate random numbers from a normal population. From the sample obtained, each is asked to compute a 95% confidence interval for the mean of the population. We then ask students to raise their hand if the 95% confidence interval they just calculated does not include the true value of the mean. Almost no hands are raised. We discuss with the students the fact that, with a 95% confidence level, the true value is captured in 95% of cases. We repeat the procedure with other levels of confidence and corroborate that when using, for example, a 50% confidence level, only around 50% of students get the mean in the interval.

2.19 The Sample Size Versus the Size of the Population: A Spoon for Tasting the Soup

At home, we use a small pot to make soup when we are only two people. To taste whether or not there is enough salt, we use a teaspoon. Some weekends, there are up to 12 members of the family present, and then we use a pot that is six times larger. Should the spoon also be six times larger? No, the size of the spoon (sample) does not increase proportionally to the size of the pot (population).

Downloaded by [98.207.180.40] at 07:32 14 May 2013

2.20 The Importance of Sample Representation: Stirring the Pot

Something essential when tasting soup is to stir it well to ensure that the content of the spoon is representative of the content of the pot. If it is not stirred and the sample is removed from the top part, it may be that salt has accumulated in this area and therefore the taste will be more salty, when in fact the whole is lacking in salt. The mistake of not stirring is not corrected simply by using a larger spoon, because the error will be the same: the problem of a sample not being random is not resolved by making it larger.

2.21 Increasing the Sample Size Does Not Always Affect the Usefulness of the Estimate: The Depth of a Lake

If you cannot swim but you have to cross a river, surely, you want the river to be as shallow as possible (better 2 feet than 4 feet, just in case . . .). But once you have enough data to assert that the river is too deep to cross (say, about 8 feet), collecting more data to get the exact depth is pointless.

2.22 Sample Size and Population Variability: Analyzing a Drop of Blood

It only takes one drop of blood to find out what blood group you belong to because all drops are the same type (no variability). It does not matter if it is from a 7-pound baby or her father who weighs more than 200 pounds, just a drop is enough to determine your blood group. The greater the variability in the population, the larger the sample size necessary for estimating the value of the parameter.

2.23 The Difference Between Correlation and Cause and Effect: The Number of Firefighters and the Damage Caused by Fire

There may be a correlation between two variables without them having a causal relationship. It may be that there is a third, hidden variable that is related to the two, for example, the number of firefighters and fire damage (related variable: size of the fire), the size of shoes worn by children and mathematical skills (child's age), milk consumption, and the rate of deaths from cancer (development of the country).

2.24 Variable Selection in a Regression Equation: Bringing a Consultant to an Exam

Suppose that you have to take an exam that covers 100 different topics and that you do not know any of them. The rules, however, state that you can bring two classmates as consultants. Suppose also that you know which topics each of your classmates is familiar with. If you could bring only one consultant, it is easy to figure out who you would bring: it would be the one who knows the most topics (the variable most correlated with the response). Let us say this is Paul, who knows 85 topics. With two consultants, you might choose Paul first and, for the second option, it seems reasonable to choose the second most knowledgeable classmate (the second most highly correlated variable), for example, Albert, who knows 75 topics. The problem with this strategy is that it may be that the 75 subjects Albert knows are already included in the 85 that Paul knows and, therefore, Albert does not provide any knowledge beyond that of Paul. A better strategy is to select the second not by considering what he or she knows regarding the entire agenda, but by looking for the person who knows more about the topics that the first does not know (the variable that best explains the residuals of the equation for the variables previously entered). It may even happen that the best pair of consultants are not the most knowledgeable, as there may be two who complement each other perfectly in such a way that one knows 55 topics and the other knows the remaining 45, while the most knowledgeable consultant does not perfectly complement anybody.

2.25 Residuals Should Not Contain Information: A Trash Bag

Residuals are what remain after removing all the information from the data. Since they should carry no information, we consider them as "trash." It is necessary to make sure that we do not throw out any trash that has value (information) and that can be exploited to better explain the behavior of the dependent variable.

3. FINAL CONSIDERATIONS

An old and renowned professor of statistics once said that instead of preparing the formal presentation of the classes and then, on the fly, improvising anecdotes and analogies, it may be better to prepare the latter and improvise the rest of the class. His argument was based on many anecdotes and stories about his former students who mostly remembered only anecdotes, analogies, similes, or dramatizations and, along with the anecdotes, they retained in their memory the concepts that he had attempted to pass on.

Our experience is that in addition to creating a casual atmosphere in class, analogies are effective in helping the students to understand and remember the ideas we want to convey. We hope that some of those listed here may be useful to our colleagues.

[Received July 2012. Revised November 2012.]

REFERENCES

Brewer, J. K. (1989), "Analogies and Parables in the Teaching of Statistics," *Teaching Statistics*, 11, 21–23. [44]

Chanter, D. O. (1983), "Some Anecdotes and Analogies for Illustrating Statistical Ideas," *Teaching Statistics*, 5, 14–16. [44]

Cleary, R. J. (2005), Review of *Teaching Statistics: A Bag of Tricks*, *The American Statistician*, 59, 275. [44]

Donnelly, C. M., and McDaniel, M. A. (1993), "Use of Analogy in Learning Scientific Concepts," *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 975–987. [44]

Downloaded by [98.207.180.40] at 07:32 14 May 2013

——— (2000), "Analogy With Knowledgeable Learners: When Analogy Confers Benefits and Exacts Costs," *Psychonomic Bulletin and Review*, 7, 537–543. [44]

Feinberg, W. E. (1971), "Teaching the Type I and II Errors: The Judicial Process," *The American Statistician*, 25, 30–32. [44]

Gelman, A., and Nolan, D. (2002), *Teaching Statistics: A Bag of Tricks*, Cambridge, MA: Oxford University Press. [44]

Martin, M. A. (2003), "'It's Like . . . You Know': The Use of Analogies and Heuristics in Teaching Introductory Statistical Methods," *Journal of Statistics Education* [online], 11, 2. Available at *http://www.amstat.org/publications/jse/v11n2/martin.html*. [44]

Meng, X. (2009), "Desired and Feared—What Do We Do Now and Over the Next 50 Years?" *The American Statistician*, 63, 202–210. [44]

Downloaded by [98.207.180.40] at 07:32 14 May 2013