NUTRITION™
Nutrition xx (2009) xxx
www.nutritionjrnl.com

Statistics column

Plot your data

Seth Roberts, Ph.D.*

Department of Psychology, University of California, Berkeley, California, USA; and Tsinghua University, Beijing, China

Lesson 3: Plot your data

Writers of statistics textbooks tend to copy other textbooks rather than draw on experience. This leaves a serious gap: Techniques that are highly useful in practice are not taught. With this series of columns I am trying to fill that gap. The first column [1] pointed out that doing something (imperfect) is better than doing nothing. The second [2] was about the value of transforming your data.

A third neglected lesson about data analysis is *plot your data*. More precisely, *make all reasonable graphs of your data.* Make a histogram of every measurement (to see its distribution), plot every measurement against its date, and plot every measurement against every other measurement. This is a good way to generate ideas.

To read almost any statistics textbook, even the best (e.g., Box et al. [3]), you'd think science was all about testing ideas. It isn't. Where do the tested ideas come from? Idea *generation*, which these books ignore, is just as important as idea testing. One of the best ways to generate new ideas worth testing, I have found, is to make many graphs of my data.

It is like searching for buried treasure. New ideas worth testing are very valuable but hard to find. Only a tiny fraction (1%?) of the graphs I've made led to new ideas but some of those ideas had a big effect.

My graphs generated new ideas in two ways. 1) *Causality*. The graph suggested a cause–effect relation I hadn't thought of. 2) *Simplicity*. Something turned out to be simpler than expected. Here are examples.

Causality

Weight and sleep duration

Hoping to sleep better, I measured my sleep duration [4]. During a routine analysis of the data, I plotted sleep duration versus date. The graph showed that my sleep duration had sharply decreased several months earlier, which I hadn't noticed. The sleep change occurred at exactly at the same time I'd lost weight by changing my diet. The dietary change was to eat less-processed food, food closer to its natural state—to eat oranges instead of orange juice, for example. The upper panel of Figure 1 shows the decrease in sleep duration; the lower panel of Figure 1 shows the weight loss. Before seeing these data, I'd never suspected that weight controls sleep duration, nor had anyone else, as far as I know. I later found other evidence for this [4]. In a circuitous way, the graph of sleep duration led to several more new ideas; the first was that breakfast caused early awakening [4].

Reward expectancy and bar-press duration

Learning research is often done with rats in Skinner boxes, experimental chambers with a bar that a rat can press and a pellet dispenser. The rats press the bar to get food pellets. Figure 2 shows Skinner-box results from rats trained with a time-discrimination task called the *peak procedure* [5]. Usually the box was dark and quiet. Now and then a light or white noise went on. This marked the start of a trial. On most trials, the rat's first bar press more than 40 s later was rewarded, and the signal went off. Earlier bar presses had no effect. For example, suppose a rat presses the bar 5, 22, 28, 36, and 37 s after the start of the trial. Nothing happens. Then it presses the bar 42 s after the start of trial. *This* bar press causes the pellet dispenser to dispense a food pellet and turns off the light, ending the trial. On other trials, however, the signal lasted much longer than 40 s and no food was given. The upper panel of Figure 2 shows bar-pressing rate as a function of time since the start of the signal. Bar-pressing rate varied with time; it reached a maximum around the time that reward was most likely. These results were what I expected. In contrast, I was shocked by a parallel graph that showed bar-press *duration* (how long the rats held the bar down) as a function of time since the start of the signal (lower panel of Fig. 2). The first time I saw it I thought I'd made a mistake because it looked so different from the rate function. Eventually its shape made

* Corresponding author. Tel.: +510-418-7753; fax: +267-222-4105.
E-mail address: twoutopias@gmail.com (S. Roberts).

0899-9007/09/$ – see front matter
doi:10.1016/j.nut.2008.12.005

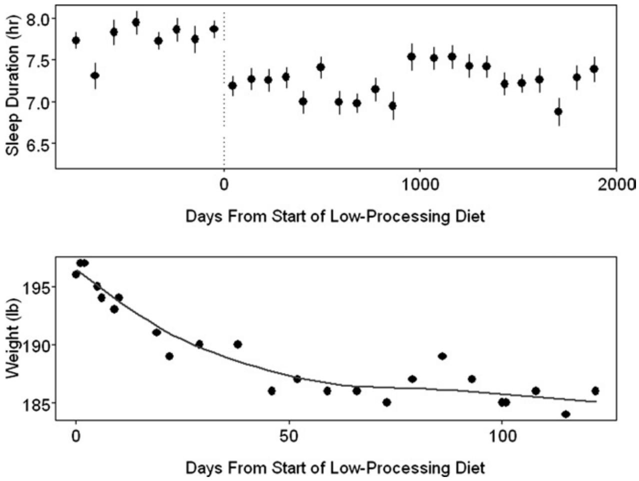

Fig. 1. Sleep duration over 5 y (upper panel) and weight loss over 4 mo (lower panel). Each point in the upper panel is a 10% trimmed mean over 84 d. The error bars are jackknife-derived SEs.

sense. The changes in mean bar-press duration reflected changes in the variability of the form of bar presses. The distribution of bar-press duration was very skew (asymmetrical); as the distribution became wider (i.e., more variable), its mean increased. When rats expected food less, the manner in which they pressed the bar became more variable. Eventually this graph (lower panel of Fig. 2) led to a new idea about what controls the variability of behavior and a new way to study it [5]. My colleagues and I have so far published two studies based on this idea [5,6].

Simplicity

Multiplicative relation

Another time-discrimination experiment of mine used the same equipment as the experiment shown in Figure 2. It

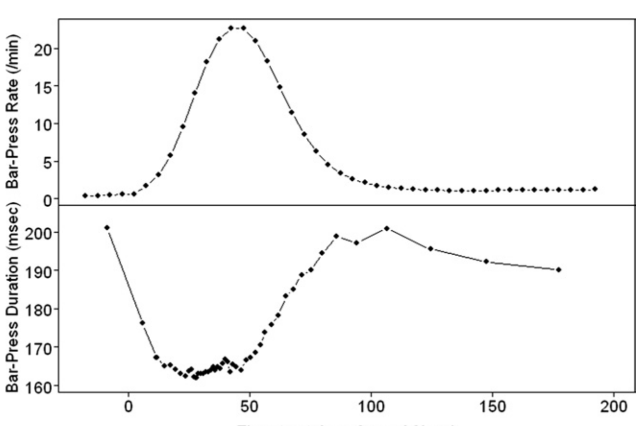

Fig. 2. Bar-press rate (upper panel) and duration (lower panel) as a function of time since the start of the signal. Points in the duration function are unequally spaced along the time axis so that each point will represent roughly the same number of bar presses.

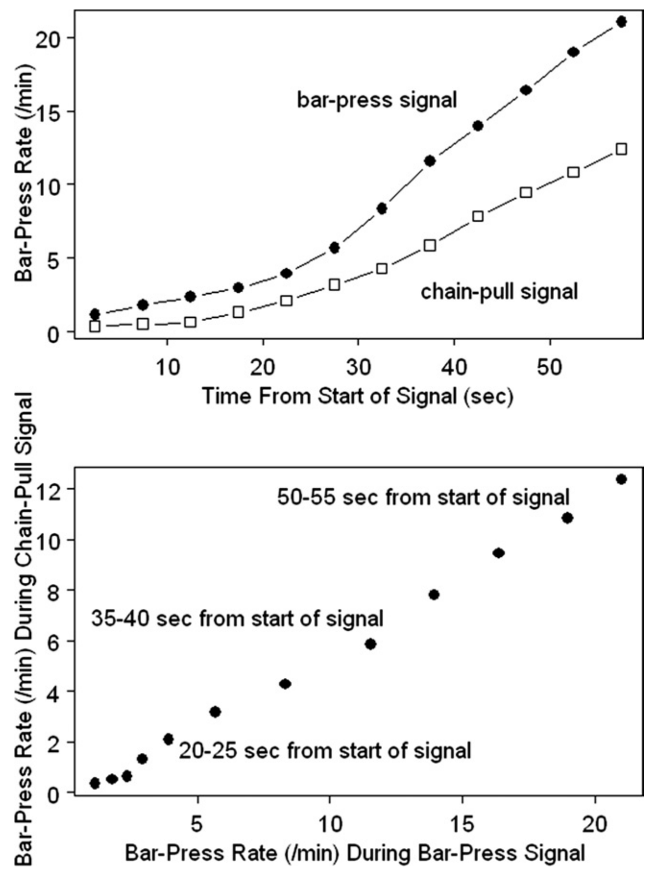

Fig. 3. (Upper panel) Bar-press rate as a function of time during two signals. During the bar-press signal, bar presses were rewarded. During the chain-pull signal, bar presses were not rewarded. (Lower panel) Bar-press rate during signal as a function of bar-press rate during the other signal. Each point is a mean over six rats.

rewarded rats for pressing a bar during one signal (light or sound) and pulling a chain during another signal (sound or light). During both signals, only the first response (bar press or chain pull) more than 60 seconds after the start of the signal was rewarded. At that point, when the reward (a food pellet) was given, the signal (light or sound) went off and there was a period of darkness and silence before the next trial. Figure 3 shows half of the results—the bar-pressing data. The upper panel shows the time and signal discriminations that this procedure produced. The rats pressed the bar more often as the time of reward approached; and they pressed the bar more often during the bar-press signal than during the chain-pull signal. One day, just to see what would happen, I plotted bar-press rates during the bar-press signal against bar-press rates during the chain-pull signal. The lower panel of Figure 3 shows the graph I made. Each point is a different time during the trial. To my surprise, most of the points fell on a straight line. This indicated that effects of time and signal combined in a proportional way; in technical terms, the two factors, time and signal, had multiplicative effects. Saul Sternberg [7] had pointed out that additive factors with reaction time could be explained

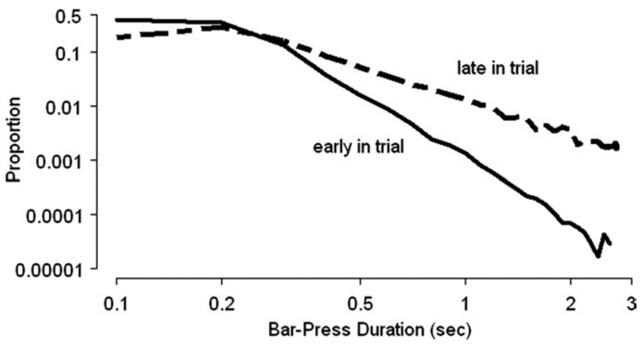

Fig. 4. Distribution of bar-press distributions. Each point is a median over 36 rats. The two functions are before (early in trial) and after (late in trial) the first bar-press more than 40 s after the start of the trial.

by assuming that the underlying mechanism could be divided into serially arranged parts. I realized that multiplicative factors with rate could be explained the same way. I started to look for them and found several other examples [8]. The theoretical conclusions that these examples suggested are supported by other lines of evidence [9].

Power-law distribution (or close to one)

I made a histogram of the bar-press durations from the experiment of Figure 2. I was surprised to see that with a log-transformation of both axes (x, value; y, relative frequency) most of the distribution was linear (Fig. 4). The two lines in Figure 4 come from different parts of the trial (before and after the first response after 40 s). The remarkable feature of the data is the near linearity. Distributions linear on a log-log scale are often called power-law distributions; the technical name is *Pareto distribution*. They occur in a wide range of situations [10]; what those situations have in common suggests why the rat's brain generates a near power-law distribution in this case [6].

Discussion

These examples reflect my research. Graphs can generate ideas in other ways. A project to ensure the quality of the joints of a high-speed train track measured the current at the time they were welded together. A graph of current versus time of day showed that the current was higher at night. This surprised the welders, who had assumed that the source of current (the power grid) was constant [11]. A medical researcher might make a histogram showing that a measurement of disease has a bimodal distribution. This might suggest that the disease has two different causes.

I have chosen extreme examples, cases where graphs led to new theories and new research. Those are cases of ideas with a big impact. But, as Saul Sternberg commented on a draft of this column, plotting *any* data is likely to give you a better mental picture of it. It will show unusual points, linearity, or bimodality. This could affect how you analyze your data (what tests you do, what data you include in those tests). I suspect a power-law distribution: a small fraction of graphs have a big effect and a large fraction have a small but non-zero effect.

I learned to plot my data from John Tukey's *Exploratory Data Analysis* [12]. Under its influence, I started plotting my data a lot. Then something surprising happened: A few of my graphs suggested new ideas—not just new ideas about my data, new ideas about how the world worked. Tukey's book had said nothing about this. I eventually grasped a larger point:

You should analyze your data two ways:

1. *To test ideas you already have* (mostly by doing statistical tests).
2. *To generate new ideas* (mostly by making graphs).

Statistics texts, with a few exceptions (e.g., De Veaux and Velleman [13]), teach only the first way.

It wasn't always like this. R. A. Fisher's *Statistical Methods for Research Workers* [14], published in 1925, had a chapter about graphs. Graphs are "no substitute for such critical tests as may be applied to the data," wrote Fisher, "but are valuable in suggesting such tests" (p. 27). After Fisher, interest in graphs disappeared from mainstream statistics research until *Exploratory Data Analysis* (1977). More recently, William Cleveland of Bell Laboratories has done much to promote and improve exploratory graphics [15,16], especially by inventing loess, a way to draw a line through a scatterplot [17]. I use loess often (e.g., lower panel of Fig. 1).

To almost all statisticians, however, the use of data to generate ideas is unfamiliar. They don't talk or write about it, and statistics texts don't include it. The result is that scientists are poorly taught, they analyze their data poorly, and a lot of buried treasure remains buried. I can think of three possible reasons for this unfortunate state of affairs. 1) *Separation of statistics professors from the practice of science*. Maybe they fail to understand the importance of idea generation. Most scientists realize you can't test the ideas you already have forever. Eventually you need new ones. But I doubt they say this to statistics professors. 2) *Difficulty of finding examples*. In my experience, examples of graphs suggesting ideas with big consequences are rare. And they often require subject-matter knowledge to understand. Fisher didn't give any examples of graphs "suggesting such tests." 3) *Lack of interest in idea generation among statistics professors*. Statistics professors may understand perfectly well the importance of idea generation but ignore it because it doesn't fit their research needs. Idea generation is hard to study. It is unpredictable and episodic, with a long time between episodes (my examples span 25 y). However, as Nassim Taleb has argued [18], to equate *rare* with *unimportant* is a huge mistake.

I thank Saul Sternberg and Andrew Gelman for helpful comments.

References

[1] Roberts S. Something is better than nothing. Nutrition 2007;23:911–2.
[2] Roberts S. Transform your data. Nutrition 2008;24:492–4.
[3] Box GEP, Hunter WG, Hunter JS. Statistics for experimenters: an introduction to design, data analysis, and model building. New York: Wiley; 1978.
[4] Roberts S. Self-experimentation as a source of new ideas: Ten examples about sleep, mood, health, and weight. Behav Brain Sci 2004;27:227–88.
[5] Gharib A, Derby S, Roberts S. Timing and the control of variation. J Exp Psychol Anim Behav Proc 2001;27:165–78.
[6] Gharib A, Gade C, Roberts S. Control of variation by reward probability. J Exp Psychol Anim Behav Proc 2004;30:271–82.
[7] Sternberg S. The discovery of processing stages: extensions of Donders' method. Acta Psychol 1969;30:276–315.
[8] Roberts S. Evidence for distinct serial processes in animals: the multiplicative-factors method. Anim Learn Behav 1987;15:135–73.
[9] Sternberg S. Separate modifiability, mental modules, and the use of pure and composite measures to reveal them. Acta Psychol 2001;106:147–246.
[10] Buchanan M. Ubiquity. New York: Three Rivers Press; 2001.
[11] Lu A. Unpublished work of Professor Anli Lu. Beijing: Department of Mechanical Engineering, Tsinghua University, 2008.
[12] Tukey JW. Exploratory data analysis. Reading, MA: Addison-Wesley; 1977.
[13] De Veaux RD, Velleman PF. Intro stats. Boston: Addison-Wesley; 2003.
[14] Fisher RA. Statistical methods for research workers. London: Oliver & Boyd; 1925.
[15] Cleveland WS. The elements of graphing data. New York: Chapman and Hall; 1985.
[16] Cleveland WS. Visualizing data. Summit, NJ: Hobart Press; 1993.
[17] Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74:829–36.
[18] Taleb NN. The black swan: the impact of the highly improbable. New York: Random House; 2007.