Letters to the Editor

NELSEN, R. B. (1998), "CORRELATION, REGRESSION LINES, AND MOMENTS OF INERTIA," *THE AMERICAN STATISTICIAN*, 52, 343–345

Having taught introductory statistics courses on a regular basis, I applaud Professor Nelsen's effort in giving the correlation coefficient ρ a direct "physical" interpretation, and I think the use of moment of inertia with respect to lines in a plane should register well with students who have some basic engineering or physics background. However, using this approach, I believe there is a more explicit "physical" interpretation of ρ than is offered by the article (e.g., as the ratio of the difference and sum of the moments of inertia about the two standard deviation lines).

Since the correlation coefficient is almost always introduced after the introduction of the mean μ (i.e., center of gravity) and the variance σ^2 (i.e., moment of inertia about μ), it should not be much trouble to introduce ρ by placing the center of gravity at the center of the plane (i.e., $(\mu_x, \mu_y) = (0, 0)$) and assuming the two marginal moments of inertia are equal (i.e., $\sigma_x^2 = \sigma_y^2 \equiv \sigma^2$). Under this setting, it follows directly from (2) of the article (or (II) here) that the relative change of the moment of inertia about the line $y = mx$ (i.e., $I(m)$) with respect to that about $y = 0$ (or $x = 0$) is

$$R(m) \equiv \frac{I(m) - \sigma^2}{\sigma^2} = -\frac{2m}{1 + m^2}\rho, \tag{I}$$

which achieves extremes at $m = \pm 1$, and thus

$$\max_m R(m) = |\rho|, \quad \text{and} \quad \min_m R(m) = -|\rho|.$$

In other words, the magnitude of ρ gives the maximum relative change in the magnitude of the moment of inertia about any line through the center of gravity compared to that about the axes. The sign of ρ tells us whether the the maximum of the moment of inertia occurs at $y = x$ (if $\rho < 0$) or at $y = -x$ (if $\rho > 0$), and equivalently about the occurrence of the minimum of the moment of inertia.

This interpretation should be quite intuitive for students with basic knowledge of physics, especially by considering the cloud of mass rotating in the 3-D fashion (with the third dimension z perpendicular to the plane) around the pole given by $y = mx$ and $z = 0$. Incidentally, one can test a student's understanding of the least-squares criterion by asking him or her why the minimum and maximum do not occur respectively at the regression line (i.e., $m = \rho$) and at the line perpendicular to it (i.e., $m = -1/\rho$).

Identity (I) also provides a nice physical interpretation of being uncorrelated: if the moments of inertia are the same about the two original axes, then the two coordinates of the mass distribution are uncorrelated if and only if the moment of inertia about the axes is rotation invariant.

For cases where $\sigma_x^2 \neq \sigma_y^2$, a complication arises because of the unbalanced inertial forces about the two axes (and thus about any line $y = mx$). This can be dealt with by considering the moment of inertia about $y = mx$ when $\rho = 0$. In other words, if we rewrite $I(m)$ of (2) as a function of both m and ρ; that is,

$$I(m, \rho) = \frac{\sigma_y^2 - 2m\rho\sigma_x\sigma_y + m^2\sigma_x^2}{1 + m^2}, \tag{II}$$

then, as a generalization of (I), we have

$$R(m, \rho) \equiv \frac{I(m, \rho) - I(m, 0)}{I(m, 0)} = -\frac{2m\sigma_x\sigma_y}{\sigma_y^2 + m^2\sigma_x^2}\rho.$$

Consequently,

$$\max_m R(m, \rho) = |\rho|, \quad \text{and} \quad \min_m R(m, \rho) = -|\rho|, \tag{III}$$

and the two extremes occur at $m = \pm\sigma_y/\sigma_x$, the standard deviation lines around which Professor Nelsen gave his interpretations.

Although the complication of $\sigma_x^2 \neq \sigma_y^2$ is unnecessary for introducing ρ since it is invariant to scale changes, the result in (III) can be helpful for someone who is more familiar with physics or mechanics than statistics to understand the role of ρ in a general mass distribution on a plane. Namely, it tells us that introducing a correlation coefficient ρ to a mass distribution in a plane is the same as modeling the extreme relative change in the moment of inertia about any line by $|\rho|$ percentage, with the sign determined by the direction at which the maximum (or minimum) occurs.

It is worth noting that the same results, and thus same interpretation, hold for the covariance $\sigma_{x,y} = \sigma_x\sigma_y\rho$, if we replace the relative change $R(m, \rho)$ by the *absolute* change $A(m, \rho) \equiv I(m, \rho) - I(m, 0)$.

Xiao-Li MENG
Department of Statistics
The University of Chicago
Chicago, IL 60637
meng@galton.uchicago.edu

SPECIAL SECTION—STATISTICS ON STATISTICIANS (1998), *THE AMERICAN STATISTICIAN*, 52, 289–314

The special section, Statistics on Statisticians (in the November 1998 issue of *The American Statistician*), discussed some information that was certainly of high interest to statisticians. We specialize in the interpretation and presentation of information, and therefore relish any information pertaining to us.

I found it rather peculiar that the most recent data presented by many of the authors dated from 1995. I recognize that the authors were not themselves entirely at fault, as they submitted their manuscripts to *The American Statistician* in 1996. Certainly the world has changed quite a lot since then. Iglewicz provided the results of an interesting survey on biopharmaceutical statisticians, but did not state the time frame during which he completed the survey. The type of information that he presented would be particularly valuable to statisticians contemplating a job change. However, the older the information, the less value it has.

Shettle and Gaddy presented quite a lot of statistics, some of which seemed rather disjointed at times. For example, they stated that "understanding the nature of the labor market for statisticians ... requires us to take a look at the demographics of this group." The authors proceeded to provide fascinating demographic data on classes of people receiving degrees in statistics, but then did not tie the data back to understanding the nature of the labor market for statisticians. How does knowing that 17% of the people with statistical degrees were Asian, or that 81% of the people with doctoral degrees in statistics were male, help us understand the nature of the labor market any better? Without providing more inferences, the authors seemed almost as if they were presenting a lot of statistics purely for the sake of presenting statistics.

The article by DeMets, Woolson, Brooks, and Qu summarized *Amstat News* advertisements from 1990 to 1994. These data are already five years out of date, and probably of little relevance. Now, many job hunters rely on the World Wide Web as a search tool. The United States Office of Personnel Management, for example, lists many federal vacancies. One advantage of this list (if salary information is of interest) is that there is no mystery about the salary range.

Demets et al. suggest asking *Amstat News* advertisers to submit confidentially a salary range—which, as the authors state, is often omitted for strategic reasons—for the purpose of tracking salary trends. However, I suspect many employers would then deliberately volunteer salary information which was distorted for the purpose of bringing down the overall salary estimates.

I think that it would be valuable for the American Statistical Association to poll its members annually to gather information on statisticians, as Iglewicz suggests. However, merely collecting and reporting statistics on nominal salaries would be quite misleading. For one thing, a nominal salary of \$50,000 in San Francisco is very different from a nominal salary of \$50,000 in Grand Rapids. For another, many people may have a high nominal salary but few paid holidays, vacation days, or health benefits. Many employers no longer provide pensions or other retirement benefits. Different people obviously value benefits such as paid holidays and pensions (i.e., deferred salaries) differently. Therefore, it will be important to collect and report basic statistics beyond nominal salaries.

I agree with Shettle and Gaddy that it will also be very important to ask how happy statisticians are in their jobs, and whether they find their jobs to be sufficiently challenging and satisfying. A common conception in American culture is that people with higher nominal salaries are necessarily happier than people with lower nominal salaries. It would be extremely fascinating to measure the extent to which this relationship holds among members of the American Statistical Association.

It would also be quite worthwhile for either *The American Statistician* or *Amstat News* to review periodically job search tools available on the Internet.

Willard C. LOSINGER
2424 Montomorency Street
Fort Collins, CO 80526

REFERENCES

DeMets, D. L., Woolson, R., Brooks, C., and Qu, R. (1998), "Where the Jobs Are: a Study of *Amstat News* Advertisements," *The American Statistician*, 52, 303–307.

Iglewicz, B. (1998), "Selected Information on the Statistics Profession," *The American Statistician*, 52, 289–294.

Shettle, C., and Gaddy, C. (1998), "The Labor Market for Statisticians and Other Scientists," *The American Statistician*, 52, 295–302.

PARR, W. C., AND SMITH, M. A. (1998), "DEVELOPING CASE-BASED BUSINESS STATISTICS COURSES," *THE AMERICAN STATISTICIAN*, 52, 330–337.

CORRECTIONS AND SUPPLEMENTS

We alert readers of our recent article in *The American Statistician* to an additional resource for case-based teaching of business statistics courses. The books by Foster, Stine, and Waterman, *Basic Business Statistics: A Casebook* and *Business Analysis Using Regression: A Casebook*, rely on cases to lead students through fundamental statistical concepts. A detailed review of these books may be found in *The American Statistician*, Vol. 52, No. 3, pg. 281.

WAND, M.P. (1997), "DATA-BASED CHOICE OF HISTOGRAM BIN WIDTH," *THE AMERICAN STATISTICIAN*, 51, 59–64

CORRECTION

The following corrections have been pointed out by Alan Polansky, Northern Illinois University. On page 61 the expression for g_{22} should be

$$g_{22} = [-2L^{(4)}(0)/\{\mu_2(L)\widehat{\psi}_6^{\mathrm{NS}} n\}]^{1/7},$$

and $g_{22}^{[4]}$ should be

$$g_{22}^{[4]} = [-24L_{[4]}^{(6)}(0)/\{\mu_4(L_{[4]})\widehat{\psi}_{10}^{\mathrm{NS}} n\}]^{1/11}.$$

On page 64 the expression for $M(h)$ should be

$$M(h) = n^{-1}h^{-1} - \frac{1}{12}h^2\psi_2 + \frac{1}{30}h^4\psi_4 + o(n^{-1}h^{-1} + h^4).$$

Subsequently the expressions for $M'(h)$ and $M''(h)$ also need to have their second terms multiplied by $\frac{1}{2}$.

M. P. WAND
Department of Biostatistics
Harvard School of Public Health
Boston, MA 02115