ON THE STATISTICS OF SENSITIVITY DATA

By Benjamin Epstein and C. West Churchman

Frankford Arsenal

1. Introduction. "Sensitivity data" is a general term for that type of experimental data for which the measurement at any point in the scale destroys the sample; as a consequence, new samples are required for each determination. Examples of such data occur in biology in dosage-mortality determinations, in psychophysics in questions concerning sensitivity responses, and, more recently, in the theory of solid explosives, in questions concerning the sensitivity of explosive or detonative mixtures.

Methods of analyzing such data have been discussed by Bliss[1] and Spearman[2], and others. The present paper is a generalization of Spearman's result; it is the feeling of the authors that Spearman's method, if properly founded in mathematical theory, is preferable to Bliss', for it does not necessitate the assumption of some type of distribution prior to analysis, and hence resembles the standard treatment of independent observations made on the same object.

Throughout the following discussion, we let x_i be the magnitude of a certain "stimulus" (be it dosage, physical stimulus, or strength of blow) and p_i the corresponding fraction of objects unaffected by the stimulus. Bliss' method consisted in assuming that the p_i represented the cumulative distribution of some known function (in his case, the normal function), and hence the p_i could be transformed into a variable t_i linearly dependent on the x_i. The difficulty of this treatment, in addition to the distribution assumption, lies in the fact that the t_i do not have equal standard errors, and the straight line fit is very cumbersome.

Instead, Spearman makes the much simpler assumption that if p_i is unaffected at x_i, and p_{i+1} at x_{i+1}, then $p_i - p_{i+1}$ is an estimate of the fraction that is just affected (i.e., the fraction of those that have "critical" responses) at about $\frac{1}{2}(x_i + x_{i+1})$. If the x_i are evenly spaced, as we shall assume them to be throughout, and $p_1 = 1.0$ and $p_n = 0$, then any set of sensitivity data may be transformed into a set of data on critical responses classified into classes whose midpoints are evenly spaced. Without loss of generality, we shall assume the x_i's to be integers and the intervals to be unity. The data on critical responses can then be treated in the normal way, and $\bar{X}$ and all the measures of dispersion calculated in the usual fashion. In order to justify such procedures, however, it is necessary to show how the sampling errors of $\bar{X}$ and the higher moments can be estimated.

[1] C. I. Bliss, "The calculation of the dosage mortality curve," *Annals of Applied Biology*, Vol. 22, pp. 134–167.

[2] C. Spearman, "The method of 'right and wrong cases' (constant stimuli) without Gauss' formulae," *British Jour. of Psych.*, Vol. 2, 1908, pp. 227–242.

2. The moments and their errors. By definition,

$$\bar{X} = \sum_{i=1}^{n} (p_i - p_{i+1})(x_i + x_{i+1})/2 = \sum_{i=1}^{n} (p_i - p_{i+1})(x_i + \tfrac{1}{2}). \tag{1}$$

If we let x_1 represent the stimulus for which none of the samples can be affected, then

$$\bar{X} = x_1 + .5 + \sum_{i=2}^{n-1} p_i , \tag{2}$$

as Spearman has shown (3). Since x_1 is constant, and the p_i are all independent (non-correlated), it follows that (N_i being the number of objects in the ith sample)

$$\sigma_{\bar{X}}^2 = \sigma_{\Sigma p_i}^2 = \sigma_{p_1}^2 + \sigma_{p_2}^2 + \cdots + \sigma_{p_n}^2 = \sum \sigma_{p_i}^2 = \sum_{i=2}^{n-1} \frac{p_i q_i}{N_i} \tag{3}$$

(since $\sigma_{p_1}^2 = \sigma_{p_n}^2 = 0$).

Again by definition, the qth moment about the origin is

$$\mu_q' = \sum_{i=1}^{n} (p_i - p_{i+1})(x_i + \tfrac{1}{2})^q. \tag{4}$$

As before $x_1 + .5$ can be taken as the origin ($x_1 + .5 = 0$), in which case we have

$$\mu_q' = (p_1 - p_2)\cdot 0^q + (p_2 - p_3)\cdot 1^q + (p_3 - p_4)\cdot 2^q + \cdots + (p_{n-1} - p_n)(n-1)^q. \tag{5}$$

If we let $b_{q,i}$ represent the ith first difference of the consecutive qth powers of the positive integers (including 0), then

$$\mu_q' = \sum_{i=2}^{n-1} b_{q,i} p_i , \tag{6}$$

by expansion of (5). Hereafter all Σ will be taken from $i = 2$ to $i = n - 1$.

Evidently

$$\sigma_{\mu_q'}^2 = \sum_{i=2}^{n-1} b_{q,i}^2 \left(\frac{p_i q_i}{N_i}\right), \tag{7}$$

or

$$\sigma_{\mu_q'}^2 = \sum_{i=2}^{n-1} b_{q,i}^2 \sigma_{p_i}^2 . \tag{8}$$

We are interested now in the standard error of the qth moment about the sample mean. To obtain this, compute first the correlation between the qth and rth moments about the origin.

If $\delta\mu_q'$ is taken to be the sample error in μ_q' due to deviations δp_i from the true values, then we have

$$\delta\mu_q' = \sum_{i=2}^{n-1} b_{q,i}\,\delta p_i$$

$$\delta\mu_r' = \sum_{i=2}^{n-1} b_{r,i}\,\delta p_i\,.$$

Hence

$$\delta\mu_q'\,\delta\mu_r' = \sum b_{q,i}\,b_{r,i}(\delta p_i)^2 + \sum_{i\neq j}(b_{q,i}\,b_{r,j} + b_{q,j}\,b_{r,i})\delta p_i\,\delta p_j\,.$$

Summing for all samples:

$$(9)\qquad \begin{aligned}\sigma_{\mu_q'}\sigma_{\mu_r'}r_{\mu_q'\mu_r'} &= \sum b_{q,i}\,b_{r,i}\,\sigma_{p_i}^2 + \sum_{i\neq j}(b_{q,i}\,b_{r,j} + b_{q,j}\,b_{r,i})(\sigma_{p_i}\sigma_{p_j}r_{p_ip_j})\\ &= \sum b_{q,i}\,b_{r,i}\,\sigma_{p_i}^2\,.\end{aligned}$$

Since evidently $r_{p_ip_j}$ vanishes for all $i \neq j$ (the p_i being completely independent in the statistical sense).

In particular, when $\mu_r' = \mu_1' = \bar{X}$, we have

$$(10)\qquad \sigma_{\mu_q'}\sigma_{\bar{X}}r_{\mu_q',\bar{X}} = \Sigma b_{1,i}\sigma_{p_i}^2\,.$$

By definition, the qth moment about the mean will be

$$(11)\qquad \mu_q = \Sigma(p_i - p_{i+1})(x_i + \tfrac{1}{2} - \bar{X})^q = \Sigma p_i'(x_i + \tfrac{1}{2} - \bar{X})^q$$

where $p_i' = p_i - p_{i+1}$.

For computational purposes, this may be written as

$$(12)\qquad \mu_q = \mu_q' - q\bar{X}\mu_{q-1}' + \frac{q(q-1)}{2}\bar{X}^2\mu_{q-2}' + \cdots + {}_qC_r\bar{X}^r\mu_{q-r+1} + \cdots + \bar{X}^q$$

where $\bar{X} = \Sigma p_i = \mu_1'$, if $x_1 + \frac{1}{2}$ is the origin.

To obtain $\sigma_{\mu_q}^2$, where $\bar{X}$ is estimated from the sample, we may follow the usual procedures, arguing that

$$(13)\qquad \delta\mu_q = \Sigma\{(x_i + \tfrac{1}{2})^q\delta p_i'\} - q\delta\bar{X}\Sigma(x_i + \tfrac{1}{2})^{q-1}p_i' + T$$

where T contains terms involving $\bar{X}$ and higher powers of $\bar{X}$.

From (13) we obtain

$$(14)\qquad \sigma_{\mu_q}^2 = \sigma_{\mu_q'}^2 + q^2\mu_{q-1}'^2\sigma_{\bar{X}}^2 - 2q\mu_{q-1}'\sigma_{\bar{X}}\sigma_{\mu_q'}r_{\bar{X}\mu_q'} + U$$

where U involves $\bar{X}$ and higher powers. From (3), (8), (10) and (14) we have

$$(15)\qquad \begin{aligned}\sigma_{\mu_q}^2 &= \Sigma b_{q,i}^2\sigma_{p_i}^2 + q^2\mu_{q-1}'^2\sigma_{p_i}^2 - 2q\mu_{q-1}'\Sigma b_{q,i}\sigma_{p_i}^2 + U\\ &= \Sigma(b_{q,i} - q\mu_{q-1}')^2\sigma_{p_i}^2 + U.\end{aligned}$$

We now shift the origin to $\bar{X}$. All the terms in U vanish, μ_{q-1}' becomes μ_{q-1}, and the $b_{q,i}$ values go into $\beta_{q,i}$, where

$$\beta_{q,i} = (i - \bar{X})^q - (i - 1 - \bar{X})^q.$$

That is, (15) becomes

$$\sigma_{\mu_q}^2 = \Sigma(\beta_{q,i} - q\mu_{q-1})^2\sigma_{p_i}^2 . \tag{16}$$

It is of interest to give an alternative proof of the relation (16) possessing the desirable property of being very short and simple and at the same time yielding an expression for the $\beta_{q,i}$ in terms of $b_{r,i}(1 \leq r \leq q)$ and powers of $\bar{X}$.

If $x_1 + .5$ is taken as the origin, then (11) may be written as (17)

$$\mu_q = \Sigma(i - \bar{X})^q p_i' . \tag{17}$$

The application of the δ-operation to both sides of (17) yields:

$$\delta\mu_q = \Sigma(i - \bar{X})^q \delta p_i' - q\Sigma(i - \bar{X})^{q-1} p_i' \delta\bar{X} \tag{18}$$

$$= \Sigma\,(\beta_{q,i} - q\mu_{q-1})\delta p_i . \tag{19}$$

Repetition of a previous argument gives the result:

$$\sigma_{\mu_q}^2 = \Sigma(\beta_{q,i} - q\mu_{q-1})^2\sigma_{p_i}^2 . \tag{20}$$

In order to derive the relation connecting the $\beta_{q,i}$ with $b_{r,i}(1 \leq r \leq q)$ we expand $\Sigma(i - \bar{X})^q \delta p_i'$ in equation (18). This expansion yields:

$$\begin{aligned} \sum(i - \bar{X})^q \delta p_i' &= \sum(i^q - {}_qC_1 i^{q-1}\bar{X} + {}_qC_2 i^{q-2}\bar{X}^2 \\ &\quad + \cdots + (-1)^{q-1}{}_qC_{q-1} i\bar{X}^{q-1} + (-1)^q\bar{X}^q)\delta p_i' \\ &= \sum (b_{q,i} - {}_qC_1 b_{q-1,i}\bar{X} + {}_qC_2 b_{q-2,i}\bar{X}^2 \\ &\quad + \cdots (-1)^{q-1} q\bar{X}^{q-1} b_{1,i})\delta p_i \end{aligned} \tag{21}$$

i.e.,

$$\begin{aligned} \beta_{q,i} = b_{q,i} - {}_qC_1 b_{q-1,i}\bar{X}\, {}_qC_2 b_{q-2,i}\bar{X}^2 \\ + \cdots + (-1)^{q-1} q\bar{X}^{q-1} b_{1,i} . \end{aligned} \tag{22}$$

The relationship (16) combined with (22) enables one to compute the standard errors of a number of useful statistics. In particular in case $q = 2$ it follows that

$$\sigma_{\mu_2}^2 = \sigma_{\sigma^2}^2 = \Sigma(b_{2,i} - 2\bar{X})^2\sigma_{p_i}^2 . \tag{23}$$

Combining (23) with the well-known result that

$$\sigma_\sigma = \sigma_{\mu_2}/2\sigma \tag{24}$$

we see that

$$\sigma_\sigma = \frac{\sqrt{\sum \{(2i - 3) - 2\bar{X}\}^2 \sigma_{p_i}^2}}{2\sqrt{\sum (2i - 3)p_i - (\sum p_i)^2}} . \tag{25}$$

Formula (25) is useful in significance tests involving the standard deviations of sensitivity data.

3. Standard errors of the moments in standard units. We now turn our attention to the derivation of the standard error of the higher moments when expressed in standard units. Before proceeding with the derivation it is convenient to find the correlation between the qth and rth moments about the mean. This result is an immediate consequence of (19), for since

$$\delta\mu_q = \Sigma\{\beta_{q,i} - q\mu_{q-1}\}\delta p_i$$

and

$$\delta\mu_r = \Sigma\{\beta_{r,i} - r\mu_{r-1}\}\delta p_i \tag{26}$$

it follows that

$$\delta\mu_q\delta\mu_r = \Sigma\{\beta_{q,i} - q\mu_{q-1}\}\{\beta_{r,i} - r\mu_{r-1}\}(\delta p_i)^2 + Z \tag{27}$$

where Z contains terms $\delta p_i \delta p_j (i \neq j)$. Hence, as before,

$$\sigma_{\mu_q}\sigma_{\mu_r}r_{\mu_q\mu_r} = \Sigma\{\beta_{q,i} - q\mu_{q-1}\}\{\beta_{r,i} - r_{\mu_{r-1}}\}\sigma_{p_i}^2 . \tag{28}$$

Let us now derive the standard errors of the moments in standard units, i.e., of

$$\alpha_q = \mu_q/\sigma^q. \tag{29}$$

Now in general,

$$\delta\alpha_q = \frac{\sigma^q\delta\mu_q - q\sigma^{q-1}\mu_q\delta\sigma}{\sigma^{2q}} = \frac{\sigma\delta\mu_q - q\mu_q\delta\sigma}{\sigma^{q+1}} \tag{30}$$

and since

$$\delta\mu_2 = 2\sigma\delta\sigma , \quad \text{or} \quad \delta\sigma = \delta\mu_2/2\sigma \tag{31}$$

we have

$$\delta\alpha_q = \frac{2\sigma^2\delta\mu_q - q\mu_q\delta\mu_2}{2\sigma^{q+2}} \tag{32}$$

and hence

$$(\delta\alpha_q)^2 = \frac{4\sigma^4(\delta\mu_q)^2 + q^2\mu_q^2(\delta\mu_2)^2 - 4q\sigma^2\mu_q\delta\mu_q\delta\mu_2}{4\sigma^{2(q+2)}} \tag{33}$$

$$\sigma_{\alpha_q}^2 = \frac{4\mu_2^2\sigma_{\mu_q}^2 + q^2\mu_q^2\sigma_{\mu_2}^2 - 4q\mu_2\mu_q\sigma_{\mu_q}\sigma_{\mu_2}r_{\mu_q\mu_2}}{4\mu_2^{q+2}} . \tag{34}$$

In this case, it follows that

$$\sigma_{\alpha_q}^2 = \frac{4\mu_2^2\sum(\beta_{q,i} - q\mu_{q-1})^2\sigma_{p_i}^2 + q^2\mu_q^2\sum\beta_{2,i}^2\sigma_{p_i}^2 - 4q\mu_q\mu_2\sum(\beta_{q,i} - q\mu_{q-1})\beta_{2,i}\sigma_{p_i}^2}{4\mu_2^{q+2}} \tag{35}$$

or

$$\sigma_{\alpha_q}^2 = \frac{\sum(2\mu_2(\beta_{q,i} - q\mu_{q-1}) - q\mu_q\beta_{2,i})^2\sigma_{p_i}^2}{4\mu_2^{q+2}} \tag{36}$$

If the qth moment about the mean vanishes, then

$$\sigma^2_{\alpha_q} = \frac{4\mu_2^2 \sum (\beta_{q,i} - q\mu_{q-1})^2 \sigma^2_{p_i}}{4\mu_2^{q+2}} = \frac{\sigma^2_{\mu_q}}{\mu_2^q}. \tag{37}$$

It is readily seen that the standard errors of the skewness and flatness are special cases of formula (36) when $q = 3$ and $q = 4$ respectively.

4. Some minimization problems. In the analysis of sensitivity data it is most desirable to minimize $\sigma^2_{\bar{X}}$ or $\sigma^2_{\sigma^2}$ in order to increase the precision of significance tests involving $\bar{X}$ or σ respectively. Therefore, it is of interest to solve the following problem: Suppose that we have a sample of size N which is to be subdivided into n samples of size N_i to be tested at a number of fixed levels $\{x_i\}$ $i = 1, 2 \cdots n$, $\sum_{i=1}^{n} N_i = N$. Then what choice of values $\{N_i\}$ will minimize

$$\sigma^2_{\bar{X}} = \sum_{i=1}^{n} \frac{p_i q_i}{N_i}, \qquad \text{where} \quad \sum_{i=1}^{n} N_i = N?$$

In order to solve this problem most quickly we use the method of Lagrange multipliers, i.e., we minimize the expression

$$L_1(N_i, \lambda) = \sum_{i=1}^{n} \frac{p_i q_i}{N_i} + \lambda\left(\sum_{i=1}^{n} N_i - N\right). \tag{38}$$

Taking the partial derivatives with respect to N_i we obtain the n equations

$$\frac{p_i q_i}{N_i^2} = \lambda, \qquad \text{i.e.,} \quad N_i = \frac{\sqrt{p_i q_i}}{\lambda^{1/2}}, \qquad i = 1, 2 \cdots, n. \tag{39}$$

Summing over all values of i we obtain

$$N = \sum_{i=1}^{n} \frac{\sqrt{p_i q_i}}{\lambda^{1/2}} \quad \text{or} \quad \lambda^{1/2} = \frac{N}{\sum_{i=1}^{n} \sqrt{p_i q_i}}; \tag{40}$$

i.e., the best choice of values for $\{N_i\}$ is given by

$$N_i = \frac{N\sqrt{p_i q_i}}{\sum_{i=1}^{n} \sqrt{p_i q_i}}. \tag{41}$$

The value of $\sigma^2_{\bar{X}}$ for this choice of the set $\{N_i\}$ is

$$\frac{\left(\sum_{i=1}^{n} \sqrt{p_i q_i}\right)^2}{N}. \tag{42}$$

It is obvious that this is actually a minimum. In particular, it is less than the value of $\sigma^2_{\bar{X}}$ for $N_i = N_j = N/n$ (the number of groups is n). This follows from the application of Schwartz' inequality to (42), for

$$\frac{\left(\sum_{i=1}^{n} \sqrt{p_i q_i}\right)^2}{N} \leq \frac{n \sum_{i=1}^{n} p_i q_i,}{N} \tag{43}$$

which equals the value of $\sigma_{\bar{X}}^2$ for $N_1 = N_2 = \cdots = N_n = N/n$. The equality holds if and only if $p_1 = p_2 = \cdots = p_n$.

Suppose next that we wish to minimize

$$(44)\qquad \sigma_{\sigma^2}^2 = \sum_{i=1}^{n} \frac{(b_{2,i} - 2\bar{X})^2 p_i q_i}{N_i'} = \sum_{i=1}^{n} \frac{\beta_{2,i}^2 p_i q_i}{N_i'},$$

where

$$\beta_{2,i} = b_{2,i} - 2\bar{X}.$$

We proceed as before to minimize the expression

$$(45)\qquad L_2(N_i', \lambda) = \sum_{i=1}^{n} \frac{\beta_{2,i}^2 p_i q_i}{N_i'} + \lambda\left(\sum_{i=1}^{n} N_i' - N\right).$$

Taking partial derivatives with respect to N_i' we obtain

$$(46)\qquad \frac{\beta_{2,i}^2 p_i q_i}{N_i'^2} = \lambda \quad \text{i.e.} \quad N_i' = \frac{|\beta_{2,i}| \sqrt{p_i q_i}}{\lambda^{1/2}}, \qquad i = 1, 2, \cdots, n$$

or summing over all values of i we obtain

$$(47)\qquad N = \sum_{i=1}^{n} \frac{|\beta_{2,i}| \sqrt{p_i q_i}}{\lambda^{1/2}} \quad \text{or} \quad \lambda^{1/2} = \sum_{i=1}^{n} \frac{|\beta_{2,i}| \sqrt{p_i q_i}}{N};$$

i.e., the best choice of values for $\{N_i'\}$ is given by

$$(48)\qquad N_i' = \frac{N |\beta_{2,i}| \sqrt{p_i q_i}}{\sum_{i=1}^{n} |\beta_{2,i}| \sqrt{p_i q_i}}.$$

The minimum value of $\sigma_{\sigma^2}^2$ is given by

$$(49)\qquad \frac{\left(\sum_{i=1}^{n} |\beta_{2,i}| \sqrt{p_i q_i}\right)^2}{N}.$$

In practice we desire a set $\{N_i\}$ which will make $\sigma_{\bar{X}}^2$ and $\sigma_{\sigma^2}^2$ small simultaneously. Unfortunately this is not in general possible. In fact, it may be asserted that the set $\{N_i\}$ minimizing, $\sigma_{\bar{X}}^2$ will yield a large value of $\sigma_{\sigma^2}^2$ and similarly the set $\{N_i'\}$ minimizing $\sigma_{\sigma^2}^2$ will yield a large value of $\sigma_{\bar{X}}^2$. The reason for this curious behavior lies in the fact that the only difference between the set $\{N_i\}$ and the set $\{N_i'\}$ is the set of numbers $\{|\beta_{2,i}|\} = \{|(2i - 3) - 2\bar{X}|\}$. These numbers, however, change the character of the sets $\{N_i\}$ and $\{N_i'\}$. In particular $\{N_i'\}$ takes on its largest values for both small and large values of i, whereas $\{N_i\}$ takes on small values in these regions; $\{N_i'\}$ takes on small values for those values of i which are the integral values closest to $\bar{X} + 3/2$, whereas $\{N_i\}$ takes on large values for such values of i. It is this curious juxtaposition of $\{N_i\}$ and $\{N_i'\}$ that renders it impossible to choose sets of numbers $\{N_i\}$ minimizing $\sigma_{\bar{X}}^2$ and $\sigma_{\sigma^2}^2$ simultaneously.