1961

On the Order of a Bias

William (Bill) H. Williams
CUNY Hunter College

On the Order of a Bias
Author(s): W. H. Williams
Source: *The American Mathematical Monthly*, Vol. 68, No. 2 (Feb., 1961), pp. 137-138
Published by: Mathematical Association of America
Stable URL: http://www.jstor.org/stable/2312477
Accessed: 03-12-2015 19:30 UTC

ON THE ORDER OF A BIAS

W. H. WILLIAMS,* McMaster University

A quantity which often arises in statistics is u/v where $u=(\bar{x}-\mu_x)^2$, $v=\sum_{i=1}^{n}(x_i-\bar{x})^2$, and $\bar{x}$ is the mean of a sample $(x_1, \cdots, x_n)$ drawn from a population with mean μ_x. The form arises, for example, in evaluating the variance of the linear regression estimator of a population mean, see Cochran [1]. It has been shown ([3], [4]) that u and v are independently distributed if and only if the x distribution is Gaussian. Much literature has been devoted to the effects of non-normality on u/v, most of which is directed at the distribution properties and analysis of Type I and Type II errors. In general $E(u/v) = E(1/v)Eu + \mathrm{Cov}(u, 1/v)$, where E denotes mathematical expectation and Cov denotes the covariance of two random variables. If the higher moments are finite, it can be shown by an interesting straightforward method that the covariance term is of lower order in n than $E(1/v)Eu$.

If u/v is written as $U(1+\delta_u)(1+\delta_v)^{-1}/V$ and $1/v$ as $(1+\delta_v)^{-1}/V$ where $Eu = U$, $Ev = V \neq 0$, $\delta_u = (u-U)/U$ and $\delta_v = (v-V)/V$ then $\mathrm{Cov}(u, 1/v) = (U/V)E\{\delta_u(1+\delta_v)^{-1}\} \doteq -\mathrm{Cov}(u, v)/V^2$. The approximation involves dropping the terms $UE\{\delta_u\delta_v^{2+i}\}/V$, $i=0, 1, \cdots$. The terms inside the expectation cannot be greater than $O(n^{-1})$ by the methods of Fisher [2]. Further, $U=\sigma_x^2/n$ and $V=(n-1)\sigma_x^2$ so that the neglected terms are no more than $O(n^{-3})$.

Next, let $d_i = x_i - \mu_x$ so that $Ed_i = 0$ and $Ed_i^2 = \sigma_x^2$, then

$$\mathrm{Cov}(u, v) = \frac{1}{n^3} E\left\{\left[(n-1)\sum_{i=1}^{n} d_i^2 - \sum_{i\neq j}^{n} d_i d_j\right]\left[\sum_{i=1}^{n} d_i^2 + \sum_{i\neq j}^{n} d_i d_j\right]\right\} - \frac{n-1}{n}\sigma_x^4.$$

which will reduce to

$$\frac{n-1}{n^2}\{Ed_i^4 - 3\sigma_x^4\} \tag{1}$$

* Now at Bell Telephone Laboratories, Inc.

or alternatively, since by definition the variance of d_i^2 is $\mathrm{Var}(d_i^2) = Ed_i^4 - \sigma_x^4$,

$$\frac{n-1}{n^2}\left\{\mathrm{Var}(d_i^2) - 2\sigma_x^4\right\}. \tag{2}$$

Hence, $\mathrm{Cov}(u, v)$ is $O(n^{-1})$ and $\mathrm{Cov}(u, 1/v)$ is $O(n^{-3})$. Furthermore, Eu is $O(n^{-1})$ and $E(1/v)$ is $O(n^{-1})$ so that $EuE(1/v)$ is $O(n^{-2})$. Notice the interesting agreement of (1) and (2) with the known fact that $\mathrm{Cov}(u, v) = 0$ for normal populations. The straightforward approach used to derive these expressions can also be used to verify the stated orders of the terms neglected in the approximation.

References

1. W. G. Cochran, Sampling Techniques, New York, 1953.
2. R. A. Fisher, Moments and product moments of sampling distributions, Proc. London Math. Soc., vol. 30, 1928, pp. 199–238.
3. R. C. Geary, Distribution of students ratio for non-normal samples, J. Royal Statist. Soc., vol. 3, 1936, pp. 178–184 (Supplement).
4. E. Lukács, A characterization of the normal distribution, Ann. Math. Statist., vol. 13, 1942, pp. 91–93.