

A COMMON ERROR in STATISTICS.

By FRANCIS GALTON, F.R.S.

I DESIRE to direct attention to the fact that a common method of calculating mean values in certain classes of observations is vitiated by an error that, curiously enough, has not attracted attention, though it is large and of a character that no multiplication of observations can eliminate. It is a fundamental oversight of somewhat serious importance, that detracts from the professed accuracy of many published tables of results.

To fix the ideas, I will explain myself by taking a definite case, leaving it to be understood that the criticism has much wider application. I will suppose then, that we are dealing with a mass of observations referring to the stature of a population, and that we adopt the common method of including all measurements noted as of, say, "65 inches and less than 66 inches," in a column headed 65—; those of "66 inches and less than 67 inches," in a column headed 66—, and so on; then of counting the entries in the columns, and referring the means of them severally to 65·5, 66·5, &c., inches. Let us imagine that the identical observations have been separately recorded on two forms, *a* and *b*, but that in *a*, the entries are to the nearest inch; in *b*, to the nearest tenth of an inch. In the *a* cases, all measurements of 65·51 inches would have been entered as 66; consequently the column headed 66—, would include all actual measurements lying between 65·5 and 66·5, whose mean value would be 66·00 inches. In the *b* cases the column headed 66—, would include all actual measurements lying between 65·95 and 66·95, whose mean value would be 66·45 inches. Hence the results given by the same observations, however numerous they may be, and whatever their intrinsic value, will differ by nearly half an inch, simply according to the degree of minuteness with which they were written down, whether to the nearest inch or the nearest tenth. The conclusion is, that the degree of minuteness to which measurements are recorded, is an element of calculation that must be ascertained and allowed for, whenever the above method of obtaining mean values is employed.