

AN APPLICATION OF ANALYSIS SITUS
TO STATISTICS*

BY HAROLD HOTELLING

1. *Introduction.* The theoretical distribution curve of correlation coefficients obtained by sampling under any particular conditions may or may not extend to the limits $r = \pm 1$. Whether it does extend to either of these limits, and if so its order of contact with the r -axis, are determined by features of the problem which will be shown to be essentially topological. These properties of the curve are independent of any influence which some members of the sample may exert upon others, provided this does not amount to complete determination. They are, moreover, to some extent independent of the nature and existence of any correlation between the variates in the population from which samples are drawn, and indeed of the distribution in this population of the variates. Finally, they are independent of heterogeneity in the population. First it will be well to notice the relation of these ideas to *time series*.

2. *Time Series.* It is well known that the correlation coefficient and other statistical measures do not have their usual significance when computed from observations ordered in time. On account of the lack of independence of successive observations, the known sampling distributions and probable error formulas are inapplicable, and no adequate substitutes have been discovered. Economic and social statisticians and meteorologists thus labor under a serious handicap which is largely absent from biometric work.

A contribution toward the removal of this handicap has recently been made by G. Udny Yule in a presidential address to the Royal Statistical Society.† He gives the

* Presented to the Society, San Francisco Section, October 30, 1926.

† *Why do we sometimes get nonsense correlations between time series?* Journal of the Royal Statistical Society, vol. 89 (1926), pp. 1-69.