tarsal spines (in Soulé (1967) it was erroneously stated that such correlations exist). From this we might speculate that the developmental " accidents " which result in asymmetry are very localised events during the morphogenesis of the legs. Furthermore, the absence of the neighborhood effect suggests the restraint of using uncorrelated characters may be unnecessary.

It can be seen in table 2 that the most asymmetrical populations occur at the highest altitudes. If confirmed by future studies, this observation will be useful in testing the hypothesis (Soulé, 1967) that instances of strong directional selection in nature could be detected by relatively high PAP's. Another opportunity this system offers is the separation of genetic vs. environmental contributions to fluctuating asymmetry. A rearing program designed to isolate these components will begin this year.

Acknowledgments.—This research was supported by NSF Grant GB-123 to Paul R. Ehrlich, Department of Biological Sciences, Stanford University. Dan L. Lindsley and Paul R. Ehrlich kindly read the manuscript. We are grateful to Janine Cuzin-Roudy for technical assistance.

5. References

DOBZHANSKY, TH. 1950. Origin of heterosis through natural selection. *Genetics, 35*, 288-302.

EHRLICH, P. R., MASON, L. G., AND EMMEL, T. C. 1967. The population biology of the butterfly, *Euphydryas editha*. V. Character clusters and asymmetry. *Evolution, 21*, 85-91.

MATHER, K. 1943. Polygenic inheritance and natural selection. *Biol. Rev., 18*, 32-64.

MATHER, K. 1953. Genetical control of stability in development. *Heredity, 7*, 297-336.

REEVE, E. C. R. 1960. Some genetic tests on asymmetry of sternopleural chaeta number in *Drosophila*. *Genet. Res., 1*, 151-172.

SOULÉ, M. 1967. Phenetics of natural populations. II. Asymmetry and evolution in a lizard. *Amer. Nat., 101*, 141-160.

THODAY, J. M. 1953. Components of fitness. *Symp. Soc. Exp. Biol., 7*, 96-113.

THODAY, J. M. 1956. Balance, heterozygosity and developmental stability. *Cold Spring Harb. Symp. Quant. Biol., 21*, 318-326.

THODAY, J. M. 1958. Homeostasis in a selection experiment. *Heredity, 12*, 401-415.

VAN VALEN, L. 1962. A study of fluctuating asymmetry. *Evolution, 16*, 125-142.

THE SIGNIFICANCE LEVEL IN MULTIPLE TESTS MADE SIMULTANEOUSLY

D. W. COOPER

Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706

Paper No. 1218 from the Laboratory of Genetics, University of Wisconsin; supported in part by U.S. Public Health Service Research Grant E-3204 from the Institute of Allergy and Infectious Diseases, and by Grant COO-1200-33 from the U.S. Atomic Energy Commission.

THE problem of how a certain proportion of " significant " results are to be interpreted when a number of χ^2 tests have been performed on different samples from similar populations often arises in genetics. It is usually understood that when the null hypothesis is correct one result in 20 is expected to fall outside the 95 per cent. level confidence limits. Doubtless, with this in mind, many working at the 5 per cent. level tend to regard one significant result in 20 and results near that figure as chance departures. The opposite is sometimes encountered; anything below 5 per cent. is treated as significant even though the hypothesis has been simultaneously tested on other samples without a significant departure being found. Both practices fail to take

into account the *extent* of the departure. The purpose of this note is to draw attention to a more exact approach which does so. The approach is that used by Neimann-Sorensen and Robertson (1961).

I shall illustrate it with reference to a 5 per cent. significance level. The basic idea is to select a modified level of significance $\alpha_{0\cdot05}$, such that 95 per cent. of all sets of n simultaneous χ^2 tests from a random pool will not contain any result for which the associated probability falls below $\alpha_{0\cdot05}$. This makes the treatment of n tests equivalent to that for a single test. If we do this, the significance for each individual test within the set of n simultaneous tests is $\alpha_{0\cdot05}$ where

$$(1-\alpha_{0\cdot05})^n = 0\cdot95$$

I am indebted to Professor J. F. Crow for pointing out that this leads to a very simple rule; if we expand the binomial and neglect the square and higher terms

$$1-n\alpha_{0\cdot05} = 0\cdot95$$

$$\alpha_{0\cdot05} = \frac{0\cdot05}{n}$$

i.e. in order to obtain the modified significance level, $\alpha_{0\cdot05}$, one divides 0·05 by the number of simultaneous tests, n. In table 1 some examples are given showing that the error involved in this approximation can be reasonably ignored. For the level $\alpha_{0\cdot01}$ (where $(1-\alpha_{0\cdot01})n = 0\cdot99$), the error is even smaller.

TABLE 1

The relationship between the actual values of the modified significance level, $\alpha_{0\cdot05}$, and the approximate values given by dividing 0·05 by n, *the number of tests.*

n	$\alpha_{0\cdot05}\times 10^4$	$\frac{0\cdot05}{n}\times 10^4$
1	500	500
2	253	250
3	169	167
4	127	125
5	102	100
6	85	83
7	73	71
8	64	63
9	57	56
10	51	50
20	26	25
50	10	10
100	5	5

The manner in which this simple rule is to be used should be obvious. When n tests are made, and some of these have an associated probability less than 5 per cent., the probability associated with each such test is compared to $\alpha_{0\cdot05} = \left(\frac{0\cdot05}{n}\right)$. If any of these probabilities fall below $\alpha_{0\cdot05}$, the result is to be regarded as significant; if they do not so fall, they may be

regarded as chance departures. As an example we may take the data of Cooper and Rendel (1958, their table 3). In this table, data are presented upon the agreement of the phenotypic frequencies of the FV and Z cattle blood group systems with the Hardy-Weinberg expectations. There are sixteen samples of the FV system, two of which have χ_1^2 on or below the 5 per cent. level ($\chi_1^2 = 3{\cdot}86$, $P = 0{\cdot}05$ and $\chi_1^2 = 5{\cdot}47$, $P = 0{\cdot}02$). There are eleven for the Z system, one of which is below 5 per cent. ($\chi_1^2 = 15{\cdot}10$, $P = 0{\cdot}001$). For $n = 16$, $\alpha_{0{\cdot}05} = 0{\cdot}0031$ and so it is clearly not correct to regard either of the FV departures as significant. But for $n = 11$, $\alpha_{0{\cdot}05} = 0{\cdot}0045$ and so the single Z departure is significant.

This approach allows the hypothesis to be falsified only for those individual results which fall below $\alpha_{0{\cdot}05}$. It cannot be used to falsify the hypothesis for the entire set, though the set contains some results below $\alpha_{0{\cdot}05}$. The interpretation to be placed upon the result in the example given is that in the sample leading to $\chi_1^2 = 15{\cdot}10$, random mating for the Z system did not obtain while in the other samples it did.

In some situations the usual small table of χ^2 may not give the associated probability with sufficient accuracy in order to make the above comparison, *e.g.* if the χ^2 table gives $0{\cdot}01 > P > 0{\cdot}001$ and $\alpha_{0{\cdot}05} = 0{\cdot}0050$ ($n = 10$). If so, a more detailed table of χ^2 such as table 7 of Pearson and Hartley (1958) should be consulted. Crow's (1945) chart of χ^2 allows the associated probability to be found for χ^2 up to 50 and degrees of freedom to 20, and will also be useful in this regard.

The decision as to what constitutes a set of n simultaneous χ^2 tests will depend upon the null hypothesis chosen by the investigator. In the examples given above there were two null hypotheses, viz. that the Z phenotypic ratios did not depart from Hardy-Weinberg expectation, and that the FV phenotypic ratios did not depart from Hardy-Weinberg expectation. These hypotheses are appropriate when one is interested in testing for the action of some force which may operate on some but not other blood group systems, *e.g.* selection. But if one were interested in testing for the existence of something which would be expected to affect all blood group systems, *e.g.* population structure, it would be necessary to consider the results from all systems simultaneously.

The modification of the significance level proposed here is most useful when one wishes to avoid rejecting the null hypothesis when it is in fact true (type I error). This is particularly true for situations where the experiments cannot be easily repeated. If they can be, then it is probably desirable to do so, rather than to apply the modified significance level and thus run the risk of missing real departures from the hypothesis.

Finally, it is noted that a rule similar to that adduced here has been pointed out by Fisher (1960, and earlier editions) in connection with the analysis of variance. He writes (p. 60); " Thus in comparing the best with the worst of ten varieties, we have chosen the pair with the largest apparent difference out of 45 pairs, which might equally have been chosen. We might, therefore, require the probability of the observed difference to be as small as 1 in 900, instead of 1 in 20, before attaching statistical significance to the contrast."

Acknowledgments.—I wish to thank Professor J. F. Crow, Professor A. B. Chapman, Ben A. Taylor and Kenneth K. Kidd for helpful discussion of this problem.

References

COOPER, D. W., AND RENDEL, J. 1968. Incomplete family data, selection and population structure for bovine transferins and blood groups. *Heredity, 23,* 49-66.

CROW, J. F. 1945. A chart of the X^2 and *t* distributions. *J. Am. Stat. Assoc., 40,* 376.

FISHER, R. A. 1960. *The Design of Experiments,* 7th Ed. Hafner Publishing Co. Inc., New York.

NEIMANN-SORENSEN, A., AND ROBERTSON, A. 1961. The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agric. Scand., 11,* 163-196.

PEARSON, E. S., AND HARTLEY, H. O. 1958. *Biometrika Tables for Statisticians,* Vol. I. Cambridge University Press.

STABILISING SELECTION IN THE ABSENCE OF DOMINANCE: AN ADDITIONAL NOTE

M. J. KEARSEY and J. S. GALE

Department of Genetics, University of Birmingham

1. Introduction

In an earlier paper (Gale and Kearsey, 1968) it was shown, for a 2-locus model, that stabilising selection can lead to stable equilibria, even if the genes act in a purely additive manner in regard to their effect on the primary character under selection. The necessary condition for such equilibria to occur, given an additive system, is that the effect of gene substitutions on the primary character shall differ from one locus to the other. If this disparity between loci is small, equilibria will result only if the loci are closely linked; the greater the disparity, the looser need be the linkage. If the disparity is very large, stable equilibria will obtain even in the absence of linkage. Results strongly suggest that in situations where stable equilibria exist, they will usually be attained, irrespective of the initial gametic frequencies.

It is natural to enquire whether these results are applicable to more realistic situations, where the primary character is under the control of many loci. In view of the many parameters required to describe such situations, a systematic investigation of this problem would be a formidable undertaking and seems scarcely justified at present, since the values of such parameters are known only for a few special cases. We shall therefore confine ourselves to describing the results obtained by computer simulation of a few special situations, using a 3-locus deterministic model. It will be shown that, at least for the cases considered, results on the 3-locus model do not differ in any essential way from those obtained on the 2-locus model considered earlier.

2. Models of stabilising selection

A wide variety of models representing the change of fitness with phenotype under stabilising selection have been proposed at various times. Unfortunately there is little evidence which would enable us to decide which models are really appropriate. In the present paper, we shall consider 2 models of the same general type but with different intensities of