January 22nd, 1936

Dear Dr Anderson,

Thanks for sending me the Iris data, which very prettily illustrate the point I had in mind in respect of determining the compound measurement which shall discriminate differences between species, or, for that matter, other things, such as effects of genes, or environmental ~~manifestation~~ modification, better than any of the simple measurements of which they are composed. In respect of your two samples, the compound measurement, I find, is approximately sepal length + 6 (sepal width) - 7 (petal length) - 10 (petal width). More accurately, the coefficients I have determined are: 1, 5.9037, - 7.1299, -10.1036.

For Iris setosa the mean value of this compound is 12.334 mm. Its variance 7.3092 mm^2 + standard deviation 2.7036. For Iris versicolor the mean value is - 21.481 mm. The variance is 14.8448 mm^2 and the standard deviation 3.8529 mm. The difference between the means is thus nearly twelve times one standard deviation and nearly nine times the other. Indeed it exceeds seven times the standard ~~deviation~~)

deviation between a pair of plants, one of each species.

Naturally, of course, you had the species well separated already for the values of petal length and petal width only, though not so widely separated in relation to the introspecific variances as the compound makes possible. You will notice, in respect of sepal length, that, although setosa has actually a smaller mean than versicolor, it appears in the formula with a positive coefficient, though the smallest coefficient of the four. One may say that, though the sepals in setosa are actually shorter than they are in versicolor, yet that they are longer than would be expected of a versicolor plant having in the other three measurements the mean values found in setosa. This is a useful warning against thinking that one could build up a discriminating compound, using coefficients merely based on the mean differences and their variances without taking into account the whole system of co-variances between the measurements of each species.

The formula is intended to be used as a designation of an individual plant, e.g. some of the suspected hybrids may show themselves to have compound measurements obviously intermediate between the ranges of the pure species. Again it might be that,

in some of your series, the average compound measurement is different for different colours of the sepal. I have not worked out the compounds for your 100 individual plants, though with setosa I think the highest score is 18.84 for a plant marked BB light, while the smallest compound is a plant marked BV. The only one marked V very dark has also a small value. Any such effect, if it were confirmed by parallelism between the two species, would, I think, be interesting.

I have not yet gone into the effect of sampling on such a formula as I have derived from your data. This is an interesting problem in itself, which I mean to look into more fully.

I expect to be some time in the States this year, having fixed, so far, to be three weeks at Harvard during the tercentenary conference and celebrations, about five weeks in June and July in Ames, and about a month in California in September to October. I hope I may have an opportunity, some time from June to October, of talking over with you further problems you have in mind.

Yours sincerely,