ON THE STATISTICAL TREATMENT OF THE RELATION BETWEEN SEA-LEVEL CHARACTERISTICS AND HIGH-ALTITUDE ACCLIMATIZATION *

* Supplement to: Keys, A., Matthews, B.H.C., Forbes, W.H. and McFarland, R.A. (1938). Individual variations in ability to acclimatize to high altitude. *Proceedings of the Royal Society of London*, B, 126: 1-24.

Seven of the sea-level characteristics of the members of the Expedition (figs. 1–7) appear to show appreciable correlations with the subsequent acclimatization of these men at high altitude. The combination of these seven variables in a formula so as to give the highest possible total correlation with acclimatization can be made by the method of least squares. The formula has the form:

$$\text{Acclim.} = a + b(\text{pulse}) + c(pCO_2) + d(O_2\,\text{cap.})\ \text{etc.},$$

and the coefficients obtained by the method of least squares are given in the third column of Table A.

TABLE A

Coefficient of pulse rate	Values used for fig. 9	Least square values for 7 factors	4 factor formula	5 % fiducial limits
Pulse rate	−0·16	−0·7507	—	
pCO_2	−1·52	−2·534	—	
O_2 capacity	−3·10	−3·591	—	
Weight	−1·25	−0·9221	−1·628	(−0·720 to −2·532)
pO_2	−3·16	−3·590	−2·148	(−0·727 to −3·569)
Alkali reserve	+7·21	+6·472	+8·887	(5·171 to 12·603)
Age	−2·27	−2·207	−1·818	(−0·588 to −3·048)

The formula is calculated from, and tested by, only nine individuals. With the coefficients of seven independent variables, and the absolute term, the formula contains eight adjustable constants. Had one more sea-level observation been included, however worthless this one might be, the observed estimates would have been given by the formula with perfect accuracy. Since, actually, only seven variables have been used, there remains just one degree of freedom for departure from expectation. Such departures usually furnish the basis for judging the precision of a formula; in this case the basis for such judgement is exceptionally narrow.

The sum of the squares of the residues obtained by the application of these coefficients to the several observables is 13·174. For any coefficient to differ significantly from zero, the increase in the residual sum of squares due to dropping the corresponding variable out of the formula* would have to be about 160 (i.e. $12{\cdot}706^2$) times as great as this last residual, or 2127. From the least square solution it appears that the increments due to omitting each of the seven variables severally are as shown in the second column below:

Table B. Increment of residual sum of squares due to omitting each factor separately

	All factors	Omitting pulse rate	Omitting pulse rate and pCO_2	Omitting pulse rate, pCO_2 and oxygen capacity
Pulse rate	15·64	—	—	—
pCO_2	27·05	11·40	—	—
O_2	70·72	64·05	65·51	—
Weight	58·36	311·85	339·16	655·77
pO_2	190·03	218·24	523·54	464·93
Alkali reserve	210·72	352·23	355·15	1147·35
Age	531·71	516·29	509·40	444·65
Mean square residual	13·174	14·410	13·406	26·43
Needed for significance	2127	266·8	135·7	203·7

Hence it would appear that, as judged by the one degree of freedom of residual error, none of the seven coefficients differs significantly from zero at the 5 % level of significance.

This does not mean that the formula is worthless, but that all the individual coefficients might be varied largely and, provided the other coefficients were suitably adjusted, the predictive value would not appreciably be impaired. The formula in seven variables is, therefore, to a large extent arbitrary, and one or more of the variates used must certainly be redundant.

This may be judged in another way. The high ratio, 12·706, which for 5 % significance a coefficient must bear to its standard error, based on only one degree of freedom, is due to the fact that it is not improbable that this one component should happen to be very small in the sample of persons examined, compared with its value for other samples. Consequently, to bring the probability to a convincing level, the contribution of the individual variables must be very large compared with the observed residue, if these are our sole sources of information on the magnitude of the discrepancies. In this case, however, the manner in which ten independent judgements were combined to form the estimate of acclimatization provides a much more reliable estimate of the precision of this judgement, and also provides a

* For technical details see *Statistical Methods*, especially Section 29·1 and Table IV, showing e.g. 12·706 as the 5% value of t for one degree of freedom.

lower limit below which its deviations from an acclimatization formula, derived from sea-level observations, cannot possibly fall.

The total score assigned at each altitude to each subject is made up of 10 independent parts. Its variance, therefore, is equal to 10 times the variance of these parts plus 90 times the average covariance of two different parts. The method of scoring in which three subjects are given 5 marks less and three 5 marks more than the average, ensures that the sum of squares of deviations from the mean shall be $6 \times 25 = 150$. The observed sum of squares of deviations at any altitude, less 1500, may therefore be equated to 90 times the sum of products of the individual estimates. But the average covariance of different independent estimates must be the true variance of the quantity estimated; and, subtracting the sum of squares of this quantity from the sum of squares of the aggregate estimate of it, we obtain the sum of squares of the aggregate errors.

Thus if $x_1, \ldots x_{10}$ are the aggregate scores given to 10 subjects at any altitude,

$$S(x) = 500,$$

$$S(x-50)^2 = A,$$

then $\dfrac{A-1500}{90}$ is the estimated mean sum of products of two individual assessments.

Hence
$$\frac{10}{9}(A-1500)$$

is an estimate of the true sum of squares of the total quantity assessed. Subtracting from A we find

$$\frac{15{,}000-A}{9}$$

as the sum of squares of the errors of the total assessments.

For the four altitudes used we have

Altitude km.	A	$\dfrac{15{,}000-A}{9}$
2·81	7009	887·$\dot{8}$
3·66	8700	700
4·71	8498	722·$\dot{4}$
5·34, 6·14	9311	632·1
		2942·4

The total assessments at the four altitudes were averaged to give the acclimatization figure, so we may divide the total by 16 to obtain the sum of squares of deviations due to errors of estimation. Since there are 9 subjects this should be further divided by 9, giving 20·434 as a minimal estimate of the variance by which a perfect formula ought to deviate from the observations.

No contribution of less than four times this amount (80) could reasonably be accepted as significant, so that the contributions tested severally of pulse rate, pCO_2, weight and oxygen capacity indicate that at least one of these variables should be rejected.

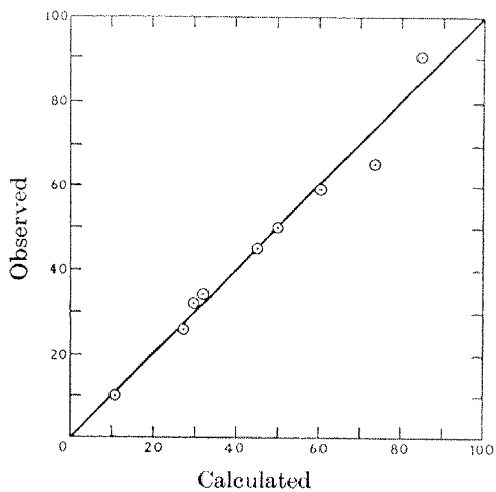

FIG. 24. Prediction formula based on only four observables at sea-level. Compare fig. 9, and the more accurate seven-factor predictions of Table C.

TABLE C. EXPECTATIONS FROM LEAST SQUARES FORMULAE WITH SEVEN AND FOUR VARIABLES LEAST SQUARES FORMULA

	7 factors			4 factors	
	Obs.	Exp.	Diff.	Exp.	Diff.
B	59	61·088	−2·088	59·527	−0·527
C	32	32·011	−0·011	29·675	+2·325
D	34	34·096	−0·096	31·863	+2·137
E	46	44·997	+1·003	45·632	+0·368
F	65	65·127	−0·127	72·688	−7·688
H	10	8·475	+1·525	10·761	−0·761
K	90	89·167	+0·833	84·221	+5·779
Mc	26	26·916	−0·916	27·496	−1·496
T	50	51·121	−1·121	50·136	−0·136

The effect of omitting pulse rate is shown in the third column of Table B. The mean square residue is only slightly raised, from 13·171 to 14·408. Since this is based now on two degrees of freedom, the amount required for significance is much diminished, being only 266·8 in place of 2127. The three variables, weight, alkali reserve, and age now give significant regressions. The contributions of oxygen capacity and of pCO_2 are, however, diminished to so small a value as to show that they also ought probably to be omitted.

The fourth column shows the effect of leaving out both pulse rate and pCO_2. In addition to the three others pO_2 now certainly makes a significant contribution. The mean square residue is, however, still lower than the value 20·434, which from the method of estimation appears to be a minimum value for the real errors of prediction. Seeing that the omission of oxygen capacity only raises the mean square to 26·43, the advantage of retaining this variable is very doubtful. The remaining four variables, used jointly, do, however, give a prediction with a standard error of about 5 on the scale used. Fig. 24 shows the observed acclimatization compared with the predicted values of Table C.

For this formula I have added, in Table A, the coefficients and their 5 % fiducial limits, i.e. the limits within which the true value will lie in 95 cases out of 100.

From Table C it appears that the largest discrepancies from the four factor formula are between the two climbers rated highest. This suggests that with more assessors, or a more careful method of estimation in which each assessor places all the subjects in order of merit, F and K would be placed more nearly equal.