XXIII.—On Problems connected with Item Selection and Test Construction.

By **D. N. Lawley,** Moray House, University of Edinburgh. *Communicated by* Professor GODFREY H. THOMSON.

(MS. received December 16, 1942. Read March 1, 1943.)

1. IN constructing tests designed to measure mental ability it has been a common practice to use a fairly large number of questions or items each of which are marked 1 or 0 according as the individual tested answers rightly or wrongly. The question how these items should be selected so that the test may give maximum discrimination between the brighter and duller individuals at various levels of ability is a very important one, and a considerable amount of literature has been devoted to it. There would therefore seem to be some justification for an attempt, however slight, at a mathematical treatment of some of the problems connected with test construction. The method of approach which we shall adopt has been suggested by an article of Ferguson (1942), and is based on the elementary theory of probability.

2. We begin by making certain assumptions which, though somewhat crude, appear to be obeyed fairly well in practice. Firstly, we shall assume that all the items composing a given test are measuring the same ability x; and we may suppose that the scale in which this ability is measured is so chosen that x is normally distributed over the whole population of individuals for whom the test is designed, with zero mean and unit variance.* It is then clear that the probability of a person passing a given item will depend upon his ability as measured on this scale. Ferguson has made the hypothesis (using different notation) that the probability P is given by the relation

$$\mathrm{P}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-a}{\sigma}} e^{-\frac{u^2}{2}}du,$$

where x measures the ability of the person and a, σ are constants for the particular item.

The value of P will be exactly $\frac{1}{2}$ when $x=a$, so that, following the nomenclature used in psychophysical work, a may be termed the limen

* The ability of children will generally depend to some extent on age, but in this case we shall assume that the age-range of the population dealt with is sufficiently small for the effect of age to be neglected.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

of the item. Clearly, if α is large and positive the item will be difficult, while conversely if α is large and negative the item will be easy. The value of the constant σ determines how well the item discriminates between individuals of high and low ability: the smaller the value of σ the greater will be the increase in P for a given increase in x. Thus $1/\sigma$ may be taken as a measure of the power of discrimination of the item.

3. Let us now suppose that an item has been tried out on a sample of individuals of differing ability and that we require to estimate for that item the constants α and σ. Let us further suppose that the individuals are divided into a number of groups such that the variation in ability within each group is sufficiently small for us to be able to neglect it. Let there be m such groups and let the ith group contain N_i individuals of ability x_i (assumed known). We may denote the number of persons in the ith group who pass the item by a_i, and the number who fail by $b_i = N_i - a_i$. The probability that the sample of persons tested should answer in such a way that in each group the above numbers pass and fail is then

$$\prod_{i=1}^{m} (P_i^{a_i} Q_i^{b_i}),$$

where

$$P_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x_i - \alpha}{\sigma}} e^{-\frac{u^2}{2}} du$$

and

$$Q_i = 1 - P_i.$$

Applying the method of maximum likelihood we now find the values of α and σ for which the above expression is made a maximum. To do this we differentiate the logarithm of the expression by α and σ in turn and equate the results to zero. Differentiating with respect to α we have

$$\sum_i \left\{ \frac{a_i P_i'}{P_i} - \frac{b_i P_i'}{1 - P_i} \right\} = 0,$$

where

$$P_i' = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \alpha)^2}{2\sigma^2}}.$$

We thus derive the equation

$$\sum_i \left\{ \frac{P_i'}{P_i Q_i} (a_i - N_i P_i) \right\} = 0. \tag{1}$$

Similarly, differentiating with respect to σ gives

$$\sum_i \left\{ \frac{P_i'}{P_i Q_i} (a_i - N_i P_i)(x_i - \alpha) \right\} = 0. \tag{2}$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

The above equations are not in a very suitable form for solving directly, but if the number in each group is sufficiently large for us to be able to neglect proportionate errors of order $1/\sqrt{N_i}$ they may be replaced by simpler ones.

Let p_i denote the observed proportion, a_i/N_i, of individuals in the ith group who pass the item, and let y_i be defined by the relation

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-\frac{u^2}{2}} du.$$

Then since p_i differs from its expected value P_i by only a small quantity (the standard error of p_i being $\sqrt{P_iQ_i/N_i}$) we have

$$p_i - P_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-\frac{u^2}{2}} du - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x_i - \alpha}{\sigma}} e^{-\frac{u^2}{2}} du$$

$$\doteqdot P_i' \left(y_i - \frac{x_i - \alpha}{\sigma} \right).$$

Hence

$$\frac{P_i'}{P_iQ_i}(a_i - N_iP_i) = \frac{N_iP_i'}{P_iQ_i}(p_i - P_i)$$

$$\doteqdot \frac{N_iP_i'^2}{P_iQ_i}\left(y_i - \frac{x_i - \alpha}{\sigma} \right)$$

$$\doteqdot \frac{N_ip_i'^2}{p_iq_i}\left(y_i - \frac{x_i - \alpha}{\sigma} \right),$$

where

$$p_i' = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_i^2}{2}}.$$

Equations (1) and (2) may now be replaced by the equations

$$\sum_i \left\{ w_i \left(y_i - \frac{x_i - \alpha}{\sigma} \right) \right\} = 0, \tag{3}$$

$$\sum_i \left\{ w_i \left(y_i - \frac{x_i - \alpha}{\sigma} \right)(x_i - \alpha) \right\} = 0, \tag{4}$$

where

$$w_i = \frac{N_ip_i'^2}{p_iq_i}.$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

Solving these, we find estimates of α and σ given by

$$1/\check{\sigma} = \frac{\sum_i (w_i x_i y_i) - \bar{x}\sum_i (w_i y_i)}{\sum_i (w_i x_i^2) - \bar{x}\sum_i (w_i x_i)},$$

$$\check{\alpha} = \bar{x} - \check{\sigma}\bar{y},$$

where

$$\bar{x} = \sum_i (w_i x_i)/\sum_i (w_i), \qquad \bar{y} = \sum_i (w_i y_i)/\sum_i (w_i).$$

It will be seen that the process of obtaining the above estimates of α and σ is exactly equivalent to that of finding the linear regression of y on x by fitting a weighted regression line through the points (x_i, y_i). The chief approximation which we have made consists in replacing the true weights $N_i P_i'^2/P_i Q_i$ by the approximate weights $N_i p_i'^2/p_i q_i$. However, even in cases where the numbers N_i are only moderately large the loss of efficiency in estimation is not serious.

The sampling variance of $1/\check{\sigma}$ is easily found to be

$$1/\sum_i \{w_i(x_i - \bar{x})^2\},$$

while that of $\check{\alpha}/\check{\sigma}$ is

$$1/\sum_i (w_i) + \bar{x}^2/\sum_i \{w_i(x_i - \bar{x})^2\}.$$

In the foregoing discussion it has been assumed that the measures of ability x_i have been previously determined, but in many cases this is not so, and we are then able to assess the individuals only by their performances on the test which is being tried out. Such cases have been dealt with by Ferguson (*loc. cit.*), who has given examples of the process of estimating the constants α and σ for various items. The estimates which he obtains by the "constant process" are identical with those which we have derived above.

4. Now suppose that we have a test containing n items. Then if $P_r = P_r(x)$ is the probability of an individual of ability x passing the rth item, the score obtained by this individual on the whole test will have an expected value of $\sum_{r=1}^{n} P_r$, while its standard error $\sigma_{E(x)}$ will be given by

$$\sigma^2_{E(x)} = \sum_{r=1}^{n} P_r(1 - P_r).$$

Let us make the assumption that the items all have the same power of discrimination, $1/\sigma_0$. There is of course no reason in practice why this should necessarily be so; but it appears that in many tests which

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

have been carefully constructed the items do not vary very greatly in their powers of discrimination, so that the above assumption does at least hold approximately. If, also, the number of items is fairly large we may regard the distribution of the values of α as being continuous. We may suppose that the number of items having values of α within the range α to $\alpha + d\alpha$ is

$$n\psi(\alpha)d\alpha,$$

where

$$\int_{-\infty}^{\infty} \psi(\alpha)d\alpha = 1.$$

The true or expected score, $F(x)$, of an individual of ability x will then be given by

$$F(x) = n\int_{-\infty}^{\infty} P\left(\frac{x-\alpha}{\sigma_0}\right)\psi(\alpha)d\alpha,$$

where

$$P\left(\frac{x-\alpha}{\sigma_0}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\alpha}{\sigma_0}} e^{-\frac{u^2}{2}}du$$

$$= \frac{1}{\sigma_0\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(v-\alpha)^2}{2\sigma_0^2}}dv.$$

Hence

$$F(x) = n\int_{-\infty}^{x} g(v)dv,$$

where

$$g(v) = \frac{1}{\sigma_0\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(v-\alpha)^2}{2\sigma_0^2}}\psi(\alpha)d\alpha.$$

The expression for $F(x)$ is simplified considerably if it be assumed that the values of α are distributed normally about a mean $\bar{\alpha}$ with a variance of σ_1^2. Again, there is no reason why items should necessarily be selected to make this so. In many cases, however, the distribution of α will be approximately normal, and we shall therefore suppose that the above assumption is satisfied. We shall then have

$$\psi(\alpha) = \frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(\alpha-\bar{\alpha})^2}{2\sigma_1^2}},$$

and

$$g(v) = \frac{1}{2\pi\sigma_0\sigma_1}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left\{\frac{(v-\alpha)^2}{\sigma_0^2}+\frac{(\alpha-\bar{\alpha})^2}{\sigma_1^2}\right\}}d\alpha$$

$$= \frac{1}{\sqrt{2\pi(\sigma_0^2+\sigma_1^2)}}e^{-\frac{(v-\bar{\alpha})^2}{2(\sigma_0^2+\sigma_1^2)}}.$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

Hence, putting *

$$\sigma^2 = \sigma_0^2 + \sigma_1^2,$$

we find that the expected score of an individual of ability x is given by

$$\begin{aligned} \mathrm{F}(x) &= \frac{n}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(v-\bar{a})^2}{2\sigma^2}}\,dv \\ &= \frac{n}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\bar{a}}{\sigma}} e^{-\frac{u^2}{2}}\,du. \end{aligned} \tag{5}$$

The observed score of the individual, as distinct from his expected score, is of course subject to random error, the error variance being given by

$$\sigma^2_{\mathrm{E}(x)} = \frac{n}{\sigma_1\sqrt{2\pi}}\int_{-\infty}^{\infty}\left\{\mathrm{P}\left(\frac{x-\alpha}{\sigma_0}\right)\right\}\left\{1-\mathrm{P}\left(\frac{x-\alpha}{\sigma_0}\right)\right\}e^{-\frac{(\alpha-\bar{a})^2}{2\sigma_1^2}}\,d\alpha.$$

Now

$$\left\{\mathrm{P}\left(\frac{x-\alpha}{\sigma_0}\right)\right\}^2 = \frac{1}{2\pi\sigma_0^2}\int_{-\infty}^{x}\int_{-\infty}^{x} e^{-\frac{1}{2\sigma_0^2}\{(v-\alpha)^2+(w-\alpha)^2\}}\,dvdw;$$

so that

$$\frac{n}{\sigma_1\sqrt{2\pi}}\int_{-\infty}^{\infty}\left\{\mathrm{P}\left(\frac{x-\alpha}{\sigma_0}\right)\right\}^2 e^{-\frac{(\alpha-\bar{a})^2}{2\sigma_1^2}}\,d\alpha = n\int_{-\infty}^{x}\int_{-\infty}^{x} g(v,\,w)\,dvdw,$$

where

$$\begin{aligned} g(v,\,w) &= \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_1\sigma_0^2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left\{\frac{(v-\alpha)^2}{\sigma_0^2}+\frac{(w-\alpha)^2}{\sigma_0^2}+\frac{(\alpha-\bar{a})^2}{\sigma_1^2}\right\}}\,d\alpha \\ &= \frac{1}{2\pi\sigma_0\sqrt{\sigma_0^2+2\sigma_1^2}}\,e^{-\frac{(\sigma_0^2+\sigma_1^2)}{2\sigma_0^2(\sigma_0^2+2\sigma_1^2)}\left\{(v-\bar{a})^2+(w-\bar{a})^2-\frac{2\sigma_1^2(v-\bar{a})(w-\bar{a})}{(\sigma_0^2+\sigma_1^2)}\right\}}. \end{aligned}$$

Hence

$$\begin{aligned} \sigma^2_{\mathrm{E}(x)} &= \frac{n}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(v-\bar{a})^2}{2\sigma^2}}\,dv \\ &\quad - \frac{n}{2\pi\sigma^2\sqrt{1-\rho_1^2}}\int_{-\infty}^{x}\int_{-\infty}^{x} e^{-\frac{1}{2\sigma^2(1-\rho_1^2)}\{(v-\bar{a})^2-2\rho_1(v-\bar{a})(w-\bar{a})+(w-\bar{a})^2\}}\,dvdw \\ &= \frac{n}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\bar{a}}{\sigma}} e^{-\frac{u^2}{2}}\,du - \frac{n}{2\pi\sqrt{1-\rho_1^2}}\int_{-\infty}^{\frac{x-\bar{a}}{\sigma}}\int_{-\infty}^{\frac{x-\bar{a}}{\sigma}} e^{-\frac{1}{2(1-\rho_1^2)}\{t^2-2\rho_1 tu+u^2\}}\,dtdu, \end{aligned} \tag{6}$$

where

$$\rho_1 = \frac{\sigma_1^2}{\sigma_0^2+\sigma_1^2} = \frac{\sigma_1^2}{\sigma^2}.$$

---

* Henceforth the symbol σ, without subscript, has a different meaning from that which it had in previous sections.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

When $x=\bar{a}$, $\sigma^2_{E(x)}$ is readily found to be

$$\frac{n}{2\pi}\cos^{-1}\rho_1=\frac{n}{2\pi}\cos^{-1}\left(\frac{\sigma_1^2}{\sigma_0^2+\sigma_1^2}\right).$$

When $x\neq\bar{a}$, the expression for $\sigma^2_{E(x)}$ can most easily be evaluated by using Table XXIX of Pearson's (1930) *Tables for Statisticians and Biometricians*. If (using Pearson's notation) we enter this table with a value of $\frac{1}{2}(1-\alpha)$ given by

$$\tfrac{1}{2}(1-\alpha)=\tau_0=\frac{1}{\sqrt{2\pi}}\int_{\frac{x-\bar{a}}{\sigma}}^{\infty}e^{-\frac{u^2}{2}}du,$$

then $\sigma^2_{E(x)}$ is given by

$$\sigma^2_{E(x)}=n(\tau_0-\tau_0^2-\tau_1^2\rho_1-\tau_2^2\rho_1^2-\tau_3^2\rho_1^3-\ldots).$$

From the above equation it is clear that the smaller the value of σ_1^2, and hence of ρ_1, the greater the error variance $\sigma^2_{E(x)}$. Since $\rho_1=\sigma_1^2/\sigma^2$ cannot be negative, $\sigma^2_{E(x)}$ will not, however, exceed the value $n(\tau_0-\tau_0^2)$, which is attained when σ_1^2 and ρ_1 are zero; this will happen if the items selected are all of equal difficulty.

For a certain Moray House intelligence test in which n, the number of items, was 100, the value of ρ_1 was found (by a process to be described later) to be ·259. The table below gives the expected values and the error variances, as calculated by the above formulæ, of the scores in this test corresponding to various values of $\frac{x-\bar{a}}{\sigma}$.

$\frac{x-\bar{a}}{\sigma}$.	Expected Value of Score, F(x).	Error Variance of Score, $\sigma^2_{E(x)}$.
0	50·0	20·8
0·1	54·0	20·7
0·2	57·9	20·4
0·3	61·8	19·8
0·4	65·5	19·0
0·5	69·1	18·0
0·6	72·6	16·9
0·7	75·8	15·6
0·8	78·8	14·3
0·9	81·6	13·0
1·0	84·1	11·6

For the same test it was found that $\bar{a}$, the mean value of the limina of

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

the items, was ·045, while σ was 1·30. Hence, for example, in the case of a person whose ability is given by $x=0{\cdot}825$ we shall have

$$\frac{x-\bar{a}}{\sigma}=\frac{0{\cdot}78}{1{\cdot}30}=0{\cdot}60;$$

so that his score will have an expected value of 72·6 with a standard error, $\sigma_{\mathrm{E}(x)}$, of $\sqrt{16{\cdot}9}=4{\cdot}1$.

5. Since x is normally distributed with zero mean and unit standard deviation the proportion $\phi(x)$ of the population with ability less than x is given by

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}du.$$

Hence a proportion $d\phi$ of the population have abilities within the range x to $x+dx$ and have expected test scores between $\mathrm{F}=\mathrm{F}(x)$ and $\mathrm{F}+d\mathrm{F}$, where

$$d\phi=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx,$$

$$d\mathrm{F}=\frac{n}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\bar{a})^2}{2\sigma^2}}dx.$$

Thus

$$d\phi=\frac{\sigma}{n}e^{\left\{\frac{(x-\bar{a})^2}{2\sigma^2}-\frac{x^2}{2}\right\}}d\mathrm{F}.$$

If $\bar{a}=0$ and $\sigma=1$, then

$$d\phi=\frac{1}{n}d\mathrm{F},$$

and the distribution of the expected score, $\mathrm{F}(x)$, is rectangular. In practice, however, the value of σ_0, and therefore that of σ, is generally somewhat greater than unity, and the distribution then has a mode at the point where

$$x=-\frac{\bar{a}}{\sigma^2-1}.$$

For two individuals with abilities x and x' the expected value of the difference between their test scores will be $\mathrm{F}(x)\sim\mathrm{F}(x')$. To decide whether such an observed difference would be considered significant we should divide it by the standard error of the difference between the two test scores, *i.e.* by $\sqrt{\sigma^2_{\mathrm{E}(x)}+\sigma^2_{\mathrm{E}(x')}}$, thus obtaining the ratio

$$\lambda=\frac{\mathrm{F}(x)\sim\mathrm{F}(x')}{\sqrt{\sigma^2_{\mathrm{E}(x)}+\sigma^2_{\mathrm{E}(x')}}}.$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

Now suppose that $x-x'$ is small Then

$$\sqrt{\sigma^2_{E(x)}+\sigma^2_{E(x')}} \doteq \sigma_{E(x)}\sqrt{2}.$$

Also, if $\Delta\phi$ is the proportion of individuals with abilities between x and x',

$$F(x) \sim F(x') = \Delta F \doteq \frac{n}{\sigma} e^{\left\{\frac{x^2}{2} - \frac{(x-\bar{a})^2}{2\sigma^2}\right\}} \Delta\phi.$$

Hence, for a given value of $\Delta\phi$,

$$\lambda \propto \frac{n}{\sigma} \times \frac{1}{\sigma_{E(x)}} e^{\left\{\frac{x^2}{2} - \frac{(x-\bar{a})^2}{2\sigma^2}\right\}}.$$

The above expression is a measure of how well the test as a whole discriminates at any point. When $x=\bar{a}$,

$$\sigma^2_{E(x)} = n \cos^{-1}\left(\frac{\sigma_1^2}{\sigma^2}\right),$$

and the expression takes the value

$$\frac{\sqrt{n}e^{\frac{\bar{a}^2}{2}}}{\sqrt{\left\{(\sigma_0^2+\sigma_1^2)\cos^{-1}\left(\frac{\sigma_1^2}{\sigma_0^2+\sigma_1^2}\right)\right\}}};$$

which for given values of σ_0 and $\bar{a}$ is a maximum when $\sigma_1=0$. This illustrates the obvious fact that in order to obtain maximum discrimination at a given point, $x=\bar{a}$, the items in a test should be selected so that the values of a are all as near as possible to $\bar{a}$.

In general it is clear that the discrimination will be greatest for large values of x, either positive or negative. It appears, therefore, that the order in which individuals are placed by their test performance is more reliable at either end of the scale than in the middle.

6. The mean score of the whole population on the test is obtained by finding the mean value of $F(x)$, and is given by

$$\begin{aligned}
\xi_T &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(x) e^{-\frac{x^2}{2}} dx \\
&= \frac{n}{2\pi\sigma} \int_{-\infty}^{\infty} \int_{-\infty}^{-\bar{a}} e^{-\frac{1}{2}\left\{\frac{(v+x)^2}{\sigma^2} + x^2\right\}} dx dv \\
&= \frac{n}{\sqrt{2\pi(\sigma^2+1)}} \int_{-\infty}^{-\bar{a}} e^{-\frac{v^2}{2(\sigma^2+1)}} dv \\
&= \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\bar{a}}{\sqrt{\sigma^2+1}}} e^{-\frac{u^2}{2}} du. \qquad (7)
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

In the same way we may derive the variance of $F(x)$. This is found to be

$$\sigma_F^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\{F(x)\}^2 e^{-\frac{x^2}{2}}dx - \xi_T^2$$

$$= \frac{n^2}{2\pi\sigma\sqrt{\sigma^2+2}}\int_{-\infty}^{-\tilde{a}}\int_{-\infty}^{-\tilde{a}} e^{-\frac{(v^2-2\rho vw+w^2)}{2(\sigma^2+1)(1-\rho^2)}}dvdw - \xi_T^2$$

$$= \frac{n^2}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{-\frac{\tilde{a}}{\sqrt{\sigma^2+1}}}\int_{-\infty}^{-\frac{\tilde{a}}{\sqrt{\sigma^2+1}}} e^{-\frac{1}{2(1-\rho^2)}\{t^2-2\rho tu+u^2\}}dtdu - \xi_T^2, \qquad (8)$$

where

$$\rho = \frac{1}{\sigma^2+1} = \frac{1}{\sigma_0^2+\sigma_1^2+1}.$$

When $\tilde{a}$ is small the simplest way of evaluating the above expression is to expand in powers of $\tilde{a}$. We then have

$$\sigma_F^2 = \frac{n^2}{2\pi}\left\{\sin^{-1}\rho - \frac{\tilde{a}^2}{(\sigma^2+1)}\left(1-\sqrt{\frac{1-\rho}{1+\rho}}\right)+O(\tilde{a}^4)\right\}. \qquad (9)$$

In general, the greater the value of ρ the larger will be the variance, σ_F^2, of the true scores. Now large values of ρ correspond to small values of σ^2. Therefore, in order to spread out well the scores of the individuals on the test it is necessary to decrease the value of $\sigma^2 = \sigma_0^2 + \sigma_1^2$ as much as possible. This can clearly be done in two ways. The first of these is to decrease σ_0^2; which means increasing the discriminating power, $1/\sigma_0$, of the items. The other, and more practicable, method is to select the items so that the variance σ_1^2 of their limina is as small as possible. It has often been thought that in constructing a test one should choose items of widely varying degrees of difficulty in order to discriminate well at different levels; and this would be correct if the power of discrimination of these items were larger than it actually is. Unless, however, new types of test items are discovered having a much greater discriminatory power than those at present in use (which is unlikely) it would seem best to choose items which vary within a comparatively narrow range of difficulty.

The error variance, $\sigma_{E(x)}^2$, of a test score is different, as we have seen, at various levels of ability. The mean value of $\sigma_{E(x)}^2$, averaged over the whole population, is however given by

$$\sigma_E^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sigma_{E(x)}^2 e^{-\frac{x^2}{2}}dx.$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

Now

$$\sigma^2_{E(x)} = \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{-\bar{a}} e^{-\frac{(v+x)^2}{2\sigma^2}} dv - \frac{n}{2\pi\sigma^2(1-\rho_1^2)} \int_{-\infty}^{-\bar{a}} \int_{-\infty}^{-\bar{a}} e^{-\frac{1}{2\sigma^2(1-\rho_1^2)}\{(v+x)^2 - 2\rho_1(v+x)(w+x) + (w+x)^2\}} dv dw,$$

where, as before, $\rho_1 = \sigma_1^2/\sigma^2$.

Hence

$$\sigma^2_E = \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\bar{a}}{\sqrt{\sigma^2+1}}} e^{-\frac{u^2}{2}} du - \frac{n}{2\pi\sqrt{1-\bar{\rho}^2}} \int_{-\infty}^{-\frac{\bar{a}}{\sqrt{\sigma^2+1}}} \int_{-\infty}^{-\frac{\bar{a}}{\sqrt{\sigma^2+1}}} e^{-\frac{1}{2(1-\bar{\rho}^2)}\{t^2 - 2\bar{\rho}tu + u^2\}} dt du, \quad (10)$$

where

$$\bar{\rho} = \frac{\rho_1\sigma^2 + 1}{\sigma^2 + 1} = \frac{\sigma_1^2 + 1}{\sigma^2 + 1}.$$

Again, when $\bar{a}$ is small we may write

$$\sigma^2_E = \frac{n}{2\pi}\left\{\cos^{-1}\bar{\rho} - \frac{\bar{a}^2}{(\sigma^2+1)}\sqrt{\frac{1-\bar{\rho}}{1+\bar{\rho}}} + O(\bar{a}^4)\right\}. \quad (11)$$

The variance, σ^2_T of the observed test scores may be regarded as being the sum of two independent components; one of these components is the variance, σ^2_F, of the true test scores, while the other is due to the differences between the observed and the true scores. The latter component will thus be the mean error variance, σ^2_E. Hence

$$\sigma^2_T = \sigma^2_F + \sigma^2_E. \quad (12)$$

Let us suppose that two parallel, and as nearly as possible similar, forms of the same test are given to the same population of individuals. The reliability coefficient of the test is then defined to be the correlation between the scores on the two parallel forms. It has been shown that this correlation coefficient is equal to the proportion of the total variance which is not due to error. The reliability coefficient is therefore given by

$$\rho_T = \frac{\sigma^2_F}{\sigma^2_T} = 1 - \frac{\sigma^2_E}{\sigma^2_T}. \quad (13)$$

Now from equations (9) and (11) we see that while the error variance, σ^2_E, is proportional to n, the number of items composing the test, the variance of the true scores, σ^2_F, is proportional to n^2. When n is large, $1 - \rho_T = \sigma^2_E/\sigma^2_T$ will therefore be of order $1/n$. It is thus clear that by increasing the number of items in a test we increase its reliability coefficient, ρ_T.

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

7. When a test containing n items is given to a random sample of N persons we can find for each person the number of items answered correctly, *i.e.* the score of that person. Alternatively, however, we can find for each item the number of persons who answer it correctly; this may be termed the score of the item. There is thus a certain reciprocity between persons and items. For an item with limen α we may derive the expected value $f(\alpha)$ and the error variance $\sigma^2_{e(\alpha)}$ of its score, just as in the case of persons. We then have

$$f(\alpha)=\frac{N}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\alpha}{\sqrt{\sigma_0^2+1}}}e^{-\frac{u^2}{2}}du, \tag{14}$$

$$\sigma^2_{e(\alpha)}=\frac{N}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\alpha}{\sqrt{\sigma_0^2+1}}}e^{-\frac{u^2}{2}}du-\frac{N}{2\pi\sqrt{1-\rho_1'^2}}\int_{-\infty}^{-\frac{\alpha}{\sqrt{\sigma_0^2+1}}}\int_{-\infty}^{-\frac{\alpha}{\sqrt{\sigma_0^2+1}}}e^{-\frac{1}{2(1-\rho_1'^2)}\{t^2-2\rho_1' tu+u^2\}}dtdu, \tag{15}$$

where

$$\rho_1'=\frac{1}{\sigma_0^2+1}.$$

Similarly, the mean score of the n items is

$$\xi_t=\frac{N}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\bar{\alpha}}{\sqrt{\sigma^2+1}}}e^{-\frac{u^2}{2}}du=\frac{N}{n}\xi_T; \tag{16}$$

while the variance of $f(\alpha)$ is given by

$$\sigma_f^2=\frac{N^2}{2\pi\sqrt{1-\rho'^2}}\int_{-\infty}^{-\frac{\bar{\alpha}}{\sqrt{\sigma^2+1}}}\int_{-\infty}^{-\frac{\bar{\alpha}}{\sqrt{\sigma^2+1}}}e^{-\frac{1}{2(1-\rho'^2)}\{t^2-2\rho' tu+u^2\}}dtdu-\xi_t^2, \tag{17}$$

where

$$\rho'=\frac{\sigma_1^2}{\sigma^2+1}=\frac{\sigma_1^2}{\sigma_0^2+\sigma_1^2+1}.$$

Also, the mean error variance of the item scores is

$$\sigma_e^2=\frac{N}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\bar{\alpha}}{\sqrt{\sigma^2+1}}}e^{-\frac{u^2}{2}}du-\frac{N}{2\pi\sqrt{1-\bar{\rho}^2}}\int_{-\infty}^{-\frac{\bar{\alpha}}{\sqrt{\sigma^2+1}}}\int_{-\infty}^{-\frac{\bar{\alpha}}{\sqrt{\sigma^2+1}}}e^{-\frac{1}{2(1-\bar{\rho}^2)}\{t^2-2\bar{\rho}tu+u^2\}}dtdu \tag{18}$$

$$=\frac{N}{n}\sigma_E^2,$$

where, as before,

$$\bar{\rho}=\frac{\sigma_1^2+1}{\sigma^2+1}.$$

The total variance σ_t^2 of the observed item scores is then given by

$$\sigma_t^2=\sigma_f^2+\sigma_e^2.$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

8. By means of the formulæ of the last two sections it is possible, knowing only the mean and variance of both the person and the item scores, to calculate the constants σ_0^2, σ_1^2 and $\bar{\alpha}$. If n and N are fairly large (say not less than 100) and $\bar{\alpha}$ is small, then to a first approximation

$$\sigma_T^2 = \frac{n^2}{2\pi}\sin^{-1}\rho = \frac{n^2}{2\pi}\sin^{-1}\left(\frac{1}{\sigma_0^2+\sigma_1^2+1}\right),$$

$$\sigma_t^2 = \frac{N^2}{2\pi}\sin^{-1}\rho' = \frac{N^2}{2\pi}\sin^{-1}\left(\frac{\sigma_1^2}{\sigma_0^2+\sigma_1^2+1}\right).$$

These equations are easily solved for σ_0^2 and σ_1^2 in terms of σ_T^2 and σ_t^2, thus yielding first approximations to σ_0^2, σ_1^2 and all the other constants functionally dependent on them. $\bar{\alpha}$ is then found from the equation

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\bar{\alpha}}{\sqrt{\sigma^2+1}}} e^{-\frac{u^2}{2}}du = \xi_T/n = \xi_t/N.$$

Second approximations for σ_0^2 and σ_1^2 may now be obtained by substituting the first approximation values of the constants in the right-hand sides of the equations

$$\sin^{-1}\left(\frac{1}{\sigma_0^2+\sigma_1^2+1}\right) = \frac{2\pi\sigma_T^2}{n^2} + \frac{\bar{\alpha}^2}{(\sigma^2+1)}\left\{1-\sqrt{\frac{1-\rho}{1+\rho}}\right\} - \frac{1}{n}\cos^{-1}\bar{\rho},$$

$$\sin^{-1}\left(\frac{\sigma_1^2}{\sigma_0^2+\sigma_1^2+1}\right) = \frac{2\pi\sigma_t^2}{N^2} + \frac{\bar{\alpha}^2}{(\sigma^2+1)}\left\{1-\sqrt{\frac{1-\rho'}{1+\rho'}}\right\} - \frac{1}{N}\cos^{-1}\bar{\rho}.$$

The above methods may readily be applied to actual data. When the true values of ξ_T, ξ_t, σ_T^2 and σ_t^2 are replaced by estimates obtained by the method of moments the resulting estimates of σ_0^2, σ_1^2, etc., will not in general be fully efficient; but usually only rough values of these constants are required.

The Moray House intelligence test of 100 items previously referred to was given to a random sample of 249 children of approximately eleven years of age. The mean and variance of the individual scores were found to be 48·9 and 618 respectively, while those of the item scores were 121·7 and 1655 respectively. Thus

$$N = 249, \qquad n = 100,$$
$$\xi_T = 48{\cdot}9, \qquad \xi_t = 121{\cdot}7,$$
$$\sigma_T^2 = 618, \qquad \sigma_t^2 = 1655.$$

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

Applying the process outlined above we then find that

$$\rho = \frac{1}{\sigma_0^2 + \sigma_1^2 + 1} = \cdot 370,$$

$$\rho' = \frac{\sigma_1^2}{\sigma_0^2 + \sigma_1^2 + 1} = \cdot 163.$$

Hence $\sigma_1^2 = 0 \cdot 44$, $\sigma_0^2 = 1 \cdot 26$, and $\sigma^2 = 1 \cdot 70$. It is also found that $\bar{a} = \cdot 045$. The other constants are given by

$$\rho_1 = \frac{\sigma_1^2}{\sigma_0^2 + \sigma_1^2} = \cdot 259,$$

$$\rho_1' = \frac{1}{\sigma_0^2 + 1} = \cdot 442,$$

$$\bar{\rho} = \frac{\sigma_1^2 + 1}{\sigma^2 + 1} = \cdot 533.$$

The mean error variance, σ_E^2, of the individual scores is now calculated from equation (11) to be 16·0. Hence, of the total variance of 618 the part, σ_F^2, due to real differences between individuals, is estimated to be $618 - 16 = 602$. Using equation (13) the reliability coefficient of the test is then given by

$$\rho_T = \tfrac{602}{618} = \cdot 974.$$

As a check on this figure the reliability coefficient was calculated by the so-called "split-half" method, which yielded the somewhat higher value of ·980. It is well known, however, that this method tends to give exaggerated estimates of the reliability.

It is interesting to compare the results of the last four sections with those of Walker (1931, 1936, 1940), although he has adopted a rather different method of approach to the whole problem.

Summary.

9. A formula is adopted which gives the probability of an individual of given ability passing a test item in terms of two quantities constant for that item. A method of estimating these two constants is given. Making certain assumptions concerning the items composing a test, formulæ are then derived giving the expected value and the standard error of the test score of any person. It is shown that there is a certain reciprocity between persons and items, and corresponding formulæ are given for the item scores. Finally, the results are applied to actual data

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282

obtained on a Moray House intelligence test, an estimate being made of the reliability of the test.

In conclusion I should like to thank Professor Godfrey H. Thomson for his help and valuable criticism in connection with this paper. I should also like to take this opportunity of thanking the Carnegie Trust for the Universities of Scotland for grants to cover the cost of the setting and printing of mathematical formulæ in two papers previously published in the Society's *Proceedings*.

REFERENCES TO LITERATURE.

FERGUSON, G. A., 1942. "Item Selection by the Constant Process," *Psychometrika*, vol. vii, pp. 19–29.

PEARSON, K., 1930. *Tables for Statisticians and Biometricians*, Part I, Cambridge University Press, 3rd ed., pp. 42–51.

WALKER, D. A., 1931. "Answer-pattern and Score-scatter in Tests and Examinations," *Brit. Journ. Psychol.*, vol. xxii, pp. 73–86.

——, 1936. "Answer-pattern and Score-scatter in Tests and Examinations," *Brit. Journ. Psychol.*, vol. xxvi, pp. 301–308.

——, 1940. "Answer-pattern and Score-scatter in Tests and Examinations," *Brit. Journ. Psychol.*, vol. xxx, pp. 248–260.

(*Issued separately May* 4, 1943.)

Downloaded from https://www.cambridge.org/core. Access paid by the UC Merced Library, on 10 Mar 2021 at 23:13:06, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0080454100006282