AN INDIVIDUAL-CENTERED APPROACH TO ITEM ANALYSIS WITH TWO CATEGORIES OF ANSWERS

by

G. Rasch.

1954

UNESCO RENDER

1. Introduction.

In the publication [1] a new approach to test psychology was attempted.

Traditionally the properties of a test are defined in terms of variations within some specified population. In practice such populations may be selected in various reasonable ways, and accordingly the properties referred to, e.g. the reliability coefficient, are not specific to the test itself, they may vary, even considerably, with the population chosen.

Similarly the evaluation of a subject usually is, by way of a standardization, linked up with some population and is therefore not specific to the subject per se.

The new approach aimed at developing probabilistic models, in the application of which the populations could be ruled out. It was a discovery of some mathematical significance that such models can at all be constructed, and it seemed remarkable that data collected in test psychological routine work could be fairly well expressed in terms of such models.

In the paper [2] an attempt was made to build up a general framework within which the models of [1] appeared to be special cases. And some general properties of this general framework were recognized. But only recently it has become quite clear, that the model (4.6) of [2] is in fact the complete answer to the requirement that statements about the parameters of a discrete probabilistic model and about the adequacy of such a model should be objective in a sense to be fully specified.

At present, at least, the theory leading to this result is rather involved and it is not going to be a main topic for this paper. However, it is intended that the following discussion of one of the models in [1], viz. the model for item analysis in case of only 2 possible answers, should demonstrate the nature of the type of objectivity we are aiming at, thus pointing to the more general problem to be treated elsewhere.

2. Data.

The situation to be considered is as follows:

A large number of subjects (1094 recruits of the Danish Army) were exposed to an intelligence test BBP, consisting of four subtests, two of which we shall deal with: N. (numerical sequences) and F. (geometrical shapes to be decomposed into some or all out of five given parts).

The time allowance for solving N. was 15 minutes, such chosen that - according to separate experimental evidence - almost none could be expected to have obtained an appreciable larger number of correct answers, even with unlimited time. Therefore items not reached were counted as not solved, as were items skipped. Thus each answer was recorded as correct (+) or non correct (÷). For F. the answers were recorded likewise. The time limit will be discussed in section 10.

In both subtests the items were largely arranged in order of increasing difficulty with a view to determining, if possible, the ability of each subject as a threshold.

3. Model.

The model to be suggested is based upon three assumptions.

1) To each situation of a subject (no. ν) having to solve an item (no. i) corresponds a probability of a correct answer which we shall write in the form

$$(3.1) \qquad p\{+|\nu,i\} = \frac{\lambda_{\nu i}}{1+\lambda_{\nu i}} \quad , \quad \lambda_{\nu i} \geq 0 \quad .$$

2) The situational parameter $\lambda_{\nu i}$ is the product of two factors,

$$(3.2) \qquad \lambda_{\nu i} = \xi_{\nu}\varepsilon_{i} \quad ,$$

ξ_{ν} pertaining to the subject, ε_{i} to the item.

3) All answers are, given the parameters, stochastically independent.

Each of these assumptions may call for some comments.

re 1) For description of observations two apparently antagonistic types of models are available, deterministic models (such as the Law of Gravitation) and stochastic models (such as Mendels Laws of Heredity). However, the choice of one type or the other does not imply that the phenomena observed were causally determined or that they did occur by chance.

Even if it be believed that certain phenomena can be "explained causally" (whatever such a phrase may mean) a stochastic model may at least temporarily be preferable (as in Thermodynamics).

Accordingly, by adopting a probabilistic model for describing responses to an intelligence test we have taken no sides in a possible argument about responses being ultimately explainable or not in causal terms.

re 2) In e.g. psychophysical threshold experiments a subject is usually exposed to the same stimulus a large number of times. On the assumption that the repetitions do not affect the judgements of the subject this procedure gives the opportunity of estimating each λ_{vi} separately and hence studying directly how the situational parameter varies with subject and with strength of stimulus. And then we may or may not find the multiplicative rule laid down in 2).

For the intelligence tests we shall deal with experience has shown that on one repetition the results are usually somewhat improved. A large number of repetitions have not been tried, mainly because the questions are such that it seems almost certain that several of them will easily be recognized after a certain number of repetitions. Therefore the possibilities for a direct approach would seem remote.

To compensate we take recourse to an assumption that at any rate seems rather bold, possibly even artificial, namely that λ_{vi} can be factorized into a subject parameter and an item parameter.

However, combined with the two other assumptions it produces a model that turns out to have rather remarkable properties, some of which even lead to a very careful examination of how well the model represents the data (cf. section 6).

Provided the two kinds of parameters can be operationally defined they also have a clear meaning, to be derived from the probabilities obtained by inserting (3.2) into (3.1):

$$(3.3) \qquad p\{+|v,i\} = \frac{\xi_v \varepsilon_i}{1+\xi_v \varepsilon_i}, \quad p\{-|v,i\} = \frac{1}{1+\xi_v \varepsilon_i} .$$

In fact, if the same person is given items with ε_i's approaching 0 then his probability of giving a correct answer approaches 0 while his probability of giving an incorrect answer tends to unity. And that is true for every person - provided the model holds. Similarly, when ε_i gets large the probability of + tends to 1 and the other one to 0. Thus with increasing ε_i the items become easier, so colloquially we may call ε_i "the degree of easiness" - and its reciprocal $\delta_i = 1/\varepsilon_i$ "the degree of difficulty" - of item i.

On the other hand, giving the same item to persons with ξ_v's approaching 0 we get probabilities of correct answers tending to 0 and if ξ_v increases indefinitely the probability tends to 1. And this holds for every item. Thus we may colloquially mention ξ_v as "the ability of subject v" - with respect to the kind of items in question.

In the definition of ξ_v and ε_i there is an inherent indeterminacy. In fact, if ξ_v, $v = 1,\ldots, n$ and ε_i, $i = 1,\ldots,k$ is a set of solutions to the equations

$$(3.4) \qquad \xi_v \varepsilon_i = \lambda_{vi} ,$$

then, if ξ'_ν, ε'_i is another set of solutions, the relation

$$\xi'_\nu \varepsilon'_i = \xi_\nu \varepsilon_i \tag{3.5}$$

must hold for any combination of ν and i.

Thus

$$\frac{\xi'_\nu}{\xi_\nu} = 1 : \frac{\varepsilon'_i}{\varepsilon_i} \tag{3.6}$$

must be a constant, α say, and accordingly the general solution is

$$\xi'_\nu = \alpha\xi_\nu, \quad \varepsilon'_i = \frac{1}{\alpha}\varepsilon_i, \quad \alpha>0 \text{ arbitrary} . \tag{3.7}$$

The indeterminacy may be removed by the choice of one of the items, say i = o, as the standard item having "a unit of easiness", in multiples of which the degrees of easinesses of the other items are expressed.

By this choice - or an equivalent one - the whole set of ξ_ν's and ε_i's is fixed.

In particular

$$\xi_\nu = \lambda_{\nu o} \quad , \tag{3.8}$$

i.e. the parameter of a subject is a very simple function of his probability of giving a correct answer to the standard item, in fact

$$\lambda_{\nu o} = \frac{p(+|\nu,o)}{1-p(+|\nu,o)} \tag{3.9}$$

is the "betting odds" on a correct answer.

Now we may be able to find a person who has in fact his $\xi = 1$. We may refer to him as a standard subject ($\nu = o$). And then the item parameter

$$\varepsilon_i = \lambda_{oi} \tag{3.10}$$

is the same simple function of the probability that the standard person gives a correct answer to this item.

rel 3) To some psychologists this assumption at first sight appears to be rather startling, since it is well known that usually quite high correlation coefficients between responses to different items are found.

This fact is, however, a consequence of the assumption.

To exemplify we may consider five subjects with the parameters

o.o1, o.1o, 1.o, 1o, 1oo,

being given three items with the parameters

1.o, 2.o, 1o.

The probabilities of correct answer become:

ξ \ ε	1.o	2.o	1o
o.o1	o.o1	o.o2	o.1o
o.1	o.1o	o.17	o.5o
1.o	o.5o	o.67	o.89
1o	o.89	o.95	o.99
1oo	o.99o	o.995	o.999

where it should be noted that for each item the probabilities vary from small values to almost unity.

Now the theoretical correlation coefficient in a population of individuals equals the correlation coefficient of the probabilities, calculated with the parameter distribution as a weight function. It follows that even with moderate variation of ξ, from o.1 to 1o say, we should obtain quite high correlation coefficients. But, of course, if ξ is the same - or practically the same - for all individuals the correlation coefficient becomes o or close to it.

Thus, the model suggested is not at all at variance with the well known findings, but, clearly, under this model the correlations do not represent intrinsic properties of the items, being mainly governed by the variation of the person parameters.

In order now in a more positive way to elucidate what is implied by our assumption we shall turn our attention to the very basic concept, that of probability.

Probabilities are often thought of as a kind of formalization or idealization of relative frequencies. In the model suggested this point of view does not apply directly since large scale repetitions of a test are not conceivable. However, in modern axiomatics of probability theory, initiated by A.N. Kolmogoroff [1], probabilities are just real numbers between o and 1 that obey a certain set of rules. Within that framework we may very well allot probabilities to events that cannot be repeated, such as the two possible responses of a person to an item (i) in an intelligence test. With a view to the following discussion we shall temporarily use the notations $p\{(\overset{i}{+})\}$ and $p\{(\overset{i}{-})\}$ for such probabilities for a given person.

---

1) See e.g. W. Feller: An Introduction to Probability Theory and Its Applications. I. Second.ed. Wiley. New York 1957. (Introduction and chapt. I, IV, V.

Considering next his possible responses to two items, i and j, they may also be allotted probabilities: $p\{(^i_+), (^j_+)\}$ etc. Now our third assumption states inter alia that his responses to i and j should be "stochastically independent". Technically this is expressed in the following relations

$$(3.11) \qquad \begin{cases} p\{(^i_+), (^j_+)\} = p\{(^i_+)\}\, p\{(^j_+)\} \\ p\{(^i_+), (^j_-)\} = p\{(^i_+)\}\, p\{(^j_-)\} \quad \text{etc.}, \end{cases}$$

but what does it mean?

If in the first of these equations we divide by $p\{(^j_+)\}$ and in the second by $p\{(^j_-)\}$ we get

$$(3.12) \qquad p\{(^i_+)\} = \frac{p\{(^i_+),(^j_+)\}}{p\{(^j_+)\}} = \frac{p\{(^i_+),(^j_-)\}}{p\{(^j_-)\}} \ .$$

In order to realize the content of these relations we shall for a moment return to relative frequencies as a background for probabilities.

Imagine a large number, N, of subjects with the same parameter, ξ, as being exposed to the two items in question. Then for instance $Np\{(^j_+)\}$ "stands for" the number out of the N subjects which gave the right answer to item j. Similarly $Np\{(^i_+),(^j_+)\}$ "stands for" the number out of the N subjects which gave the right answer to both i and j. Thus the latter number divided by the former one:

$$(3.13) \qquad \frac{Np\{(^i_+),(^j_+)\}}{Np\{(^j_+)\}}$$

"stands for" the relative frequency of right answers to both i and j among those who gave the right answer to j. Or, simpler, the relative frequency of right answers to i, provided the answer to j was correct.

Reducing by N we just have the ratio (3.12) of the two probabilities which we shall call: <u>the conditional probability of a + answer to i, given a + answer to j</u>. The notation is

$$(3.14) \qquad p\{(^i_+)\,|\,(^j_+)\} = \frac{p\{(^i_+),(^j_+)\}}{p\{(^j_+)\}} \ .$$

Thus the relations (3.12) may be written

$$(3.15) \qquad p\{(^i_+)\,|\,(^j_+)\} = p\{(^i_+)\,|\,(^j_-)\} = p\{(^i_+)\} \ ,$$

i.e. <u>the probability of a + answer to i is independent of whether the answer to j is + or -</u>, it is just the probability of a + answer to i.

And of course the same holds for a - answer to i. This is a specification of the statement that the answers to i and j are stochastically independent.

Assumption 3, however, requires still more.

First it requires for each subject that the answers to all questions should be stochastically independent. Technically this is expressed in the equation

$$(3.16) \qquad p\left\{\left(\overset{1}{+}\right),\left(\overset{2}{+}\right),\ldots\left(\overset{k}{+}\right)\right\} = p\left\{\left(\overset{1}{+}\right)\right\} p\left\{\left(\overset{2}{+}\right)\right\}\ldots p\left\{\left(\overset{k}{+}\right)\right\}$$

and all its analogues. The content of this statement is that the probability of a certain answer to an item or of a combination of answers to a set of items is unaffected by the answers given to the other items.

In its full scope assumption 3) requires that the probabilities corresponding to any combination of subjects and items should be unaffected by all the other answers.

4. Comparison of two items.

As an introduction to the more general treatment of the model in section 5 we shall consider how two items may be compared.

According to (3.11) and (3.3) the probability of correct answers to both item i and item j is

$$(4.1) \qquad \begin{cases} p\left\{\left(\overset{i}{+}\right),\left(\overset{j}{+}\right)\middle|\,\xi\right\} = p\left\{\left(\overset{i}{+}\right)\middle|\,\xi\right\} p\left\{\left(\overset{j}{+}\right)\middle|\,\xi\right\} \\ \qquad\qquad = \dfrac{\xi^2\varepsilon_i\varepsilon_j}{(1+\xi\varepsilon_i)(1+\xi\varepsilon_j)} \end{cases}$$

for a subject with the parameter ξ. Similarly

$$(4.2) \qquad p\left\{\left(\overset{i}{+}\right),\left(\overset{j}{-}\right)\middle|\,\xi\right\} = \frac{\xi\varepsilon_i}{(1+\xi\varepsilon_i)(1+\xi\varepsilon_j)} \ ,$$

$$(4.3) \qquad p\left\{\left(\overset{i}{-}\right),\left(\overset{j}{+}\right)\middle|\,\xi\right\} = \frac{\xi\varepsilon_j}{(1+\xi\varepsilon_i)(1+\xi\varepsilon_j)} \ ,$$

$$(4.4) \qquad p\left\{\left(\overset{i}{-}\right),\left(\overset{j}{-}\right)\middle|\,\xi\right\} = \frac{1}{(1+\xi\varepsilon_i)(1+\xi\varepsilon_j)} \ .$$

With the notations

$$(4.5) \qquad a_i = \begin{cases} 1 \text{ in case of answer + to item i} \\ 0 \text{ in case of answer - to item i} \end{cases}$$

and

$$(4.6) \qquad a_{\cdot} = a_i + a_j$$

the probabilities of $a_{\cdot} = 0$ and 2 are given by (4.1) and (4.4) while the probability of $a_{\cdot} = 1$ is the sum of (4.2) and (4.3):

$$(4.7) \qquad p\{a_{\cdot} = 1 | \xi\} = \frac{\xi(\varepsilon_i+\varepsilon_j)}{(1+\xi\varepsilon_i)(1+\xi\varepsilon_j)} \ .$$

Now the conditional probability of $a_i = 1$ provided $a_{\cdot} = 1$ is - analogous to (3.14) - obtained by dividing (4.7) into (4.2). However, by that operation the common denominator and ξ in the numerators cancel and we are left with

$$(4.8) \qquad p\{a_i = 1 | a_{\cdot} = 1, \xi\} = \frac{\varepsilon_i}{\varepsilon_i+\varepsilon_j} \ ,$$

<u>irrespective of the subject parameter ξ.</u>

Considering, then, a number, n, of subjects, all of which happened to have $a_{\cdot} = 1$, the probability that c of them have $a_i = 1 (a_j = 0)$ is given by the binomial law

$$(4.9) \qquad p\{c | n\} = \binom{n}{c} \left(\frac{\varepsilon_i}{\varepsilon_i+\varepsilon_j}\right)^c \left(\frac{\varepsilon_j}{\varepsilon_i+\varepsilon_j}\right)^{n-c} \ .$$

Accordingly by

$$(4.10) \qquad \frac{\varepsilon_i}{\varepsilon_i+\varepsilon_j} \approx \frac{c}{n}$$

the ratio $(\varepsilon_i/\varepsilon_j)$ is estimated <u>independently of the person parameters</u>, the distribution of which is therefore irrelevant in the connection.

Furthermore we may get a check on the model by first stratifying the subjects according to any principle - educational level or socio-economic status or even according to the total test score - and apply (4.10) to each of the groups. For the model to hold the ratio $\varepsilon_i/\varepsilon_j$ should be the same in all of the groups and the variation of the estimates obtained should therefore concord with the binomial distributions (4.9).

The appropriate test for this constancy has a remarkable property.

Denote the local c's and n's by c_g, n_g with g = 1,...h and their totals by $c_.$ and $n_.$. Since the groups could be collected into one group of size $n_.$ to which (4.9) applies we have

$$(4.11) \qquad p\{c_.|n_.\} = \binom{n_.}{c_.}\left(\frac{\varepsilon_i}{\varepsilon_i+\varepsilon_j}\right)^{c_.}\left(\frac{\varepsilon_j}{\varepsilon_i+\varepsilon_j}\right)^{n_.-c_.} .$$

On the other hand the joint probability of the numbers $c_1,\ldots,c_h$ is - due to their stochastic independence -

$$(4.12) \qquad p\{c_1,\ldots,c_h|n_1,\ldots,n_h\} = \prod_{g=1}^{h}\binom{n_g}{c_g}\cdot\left(\frac{\varepsilon_i}{\varepsilon_i+\varepsilon_j}\right)^{c_.}\left(\frac{\varepsilon_j}{\varepsilon_i+\varepsilon_j}\right)^{n_.-c_.} .$$

In consequence the conditional probability of $c_1,\ldots,c_h$ given the total $c_.$, obtained by dividing (4.11) into (4.12), becomes independent of ε_i and ε_j:

$$(4.13) \qquad p\{c_1,\ldots,c_h|c_.,n_1,\ldots,n_h\} = \frac{\prod_{g=1}^{h}\binom{n_g}{c_g}}{\binom{n_.}{c_.}} .$$

It follows that as far as only the items i and j are concerned <u>the testing of the model</u> may be directed in such a way that it is <u>independent of all of the parameters</u>.

In the formal derivation of the fundamental relation (4.8) subjects and items may of course be interchanged.

Thus the comparison of two subjects μ and ν by means of an item with parameter ε leads to the conditional probability

$$(4.14) \qquad p\{a_\mu=1|a_.=1,\varepsilon\} = \frac{\xi_\mu}{\xi_\mu+\xi_\nu}$$

if a_μ, a_ν and $a_.$ have a meaning similar to (4.5) and (4.6). And (4.14) is <u>independent of which item was used</u>.

In principle, therefore, it should be possible to estimate the ratio ξ_μ/ξ_ν independently of the item parameters and also to test the model independently of all parameters. In practice, however, this method does not work because the number of items - in contrast to the number of subjects - usually is small, in our actual cases only 17 and 18.

5. Generalization to k items.

In generalizing the results of the preceding section we shall first consider the responses of an individual with parameter ξ to k items.

With the notation (4.5) and the adaptation

$$a_\cdot = a_1 + \ldots + a_k \tag{5.1}$$

of (4.6) we may condense (3.3) to

$$p\{a_i|\xi\} = \frac{(\xi\varepsilon_i)^{a_i}}{1+\xi\varepsilon_i} \tag{5.2}$$

and the generalization of (4.1) through (4.4) to

$$\begin{cases} p\{a_1,\ldots,a_k|\xi\} = p\{a_1|\xi\} \cdots p\{a_k|\xi\} \\ \qquad = \dfrac{\xi^{a_\cdot}\varepsilon_1^{a_1}\ldots\varepsilon_k^{a_k}}{\prod_{i=1}^{k}(1+\xi\varepsilon_i)} \end{cases} \tag{5.3}$$

In consequence of this result we shall derive the probability that $a_\cdot$ takes on a specified value r.

If $r = o$ every $a_i = o$, thus

$$p\{a_\cdot = o|\xi\} = \frac{1}{\gamma(\xi)} \tag{5.4}$$

where for short we write

$$\prod_{i=1}^{k}(1+\xi\varepsilon_i) = \gamma(\xi) \tag{5.5}$$

r=1 may be obtained in k different ways:

$$\begin{aligned} &a_1 = 1,\ a_2 = \ldots = a_k = o, \\ &a_1 = o,\ a_2 = 1,\ a_3 = \ldots = a_k = o \\ &- - - - - - - - - - - - \\ &a_1 = a_2 = \ldots = a_{k-1} = o,\ a_k = 1 \end{aligned} \tag{5.6}$$

with the probabilities

$$(5.7) \qquad \frac{\xi\varepsilon_1}{\gamma(\xi)}, \frac{\xi\varepsilon_2}{\gamma(\xi)}, \cdots, \frac{\xi\varepsilon_k}{\gamma(\xi)},$$

the sum of which is the probability

$$(5.8) \qquad p\{a_\cdot = 1|\xi\} = \frac{\xi(\varepsilon_1+\ldots+\varepsilon_k)}{\gamma(\xi)}.$$

$r = 2$ may be obtained in $\binom{k}{2}$ different ways, namely by taking any two of the a_i's to be 1, the rest of them being o. The probabilities of these combinations are

$$(5.9) \qquad \frac{\xi^2\varepsilon_1\varepsilon_2}{\gamma(\xi)}, \frac{\xi^2\varepsilon_1\varepsilon_3}{\gamma(\xi)}, \frac{\xi^2\varepsilon_2\varepsilon_3}{\gamma(\xi)}, \ldots, \frac{\xi^2\varepsilon_{k-1}\varepsilon_k}{\gamma(\xi)}$$

and the sum of them is

$$(5.10) \qquad p\{a_\cdot=2|\xi\} = \frac{\xi^2(\varepsilon_1\varepsilon_2+\ldots+\varepsilon_{k-1}\varepsilon_k)}{\gamma(\xi)}.$$

In general $a_\cdot = r$ may be obtained in $\binom{k}{r}$ different ways, namely by taking any r out of the k a_i's to be 1, the rest of them being o. The probabilities of these combinations being

$$(5.11) \qquad \frac{\xi^r\cdot\varepsilon_1\ldots\varepsilon_r}{\gamma(\xi)}, \frac{\xi^r\cdot\varepsilon_1\ldots\varepsilon_{r-1}\varepsilon_{r+1}}{\gamma(\xi)}, \ldots, \frac{\xi^r\cdot\varepsilon_{k-r+1}\ldots\varepsilon_k}{\gamma(\xi)}$$

the probability of $a_\cdot = r$ becomes

$$(5.12) \qquad p\{a_\cdot=r|\xi\} = \frac{\gamma_r\xi^r}{\gamma(\xi)}$$

where for short

$$(5.13) \qquad \gamma_r = \varepsilon_1\ldots\varepsilon_r+\ldots+\varepsilon_{k-r+1}\ldots\varepsilon_k.$$

In particular for $r = k$ (5.13) has only one term

$$(5.14) \qquad \gamma_k = \varepsilon_1\varepsilon_2\ldots\varepsilon_k.$$

If in (5.12) we let r pass through the values o,1,...,k all possibilities have been exhausted and therefore the probabilities must add up to unity:

$$(5.15) \qquad \sum_{r=o}^{k} p\{a_\cdot = r|\xi\} = 1.$$

Hence

$$\gamma(\xi) = \sum_{r=0}^{k} \gamma_r \xi^r \quad , \tag{5.16}$$

i.e. γ_r are the coefficients in the expansion of the product (5.5) in powers of ξ[x].

If the ε's were known the γ_r's might be computed and it would be possible from an observed $a_.$ to estimate ξ and to indicate the precision of the estimate, for instance in terms of confidence intervals. Thus $a_.$ is what is called an <u>estimator</u> for ξ. How to compute an estimate from the estimator is not our concern at present, but as an estimator $a_.$ has an important property.

On dividing (5.1o) into (5.3) we obtain the conditional probability of the a_i's, given that their sum is r. Through this operation both the common denominator and the common power of ξ cancel and so we get

$$p\{a_1,\ldots,a_k | a_.=r,\xi\} = \frac{\varepsilon_1^{a_1}\ldots\varepsilon_k^{a_k}}{\gamma_r} \tag{5.17}$$

which is <u>independent of ξ</u>, the parameter to be estimated.

In order to realize the significance of this result we may turn to an obvious, but fundamental principle of science, namely that <u>if we want to know something about a quantity</u> - e.g. a parameter of a model - <u>then we have to observe something that depends on that quantity</u> - something that changes if the said quantity varies materially.

For the purpose of estimating the parameter ξ of a person the observations $a_1,\ldots,a_k$ are at disposal. On repetition of the experiment they would - according to our theory - vary at random in concordance with the distribution (5.3) which depends on ξ. $a_.$ also is a random variable, the distribution of which (5.12) depends on ξ and therefore it may be used for the estimation. But what (5.17) tells is that <u>the constellation of o's and 1's producing $a_.$, which also varies at random, has a distribution that does not depend on ξ</u>. From the fundamental principle it then follows that once $a_.$ has been recorded any extra information about <u>which</u> of the items were answered correctly is, according to our model, <u>useless as a source of inference about ξ</u> (but not for other purposes as will presently be seen).

---

x) In the algebra they are known as "the elementary symmetric functions" of $\varepsilon_1,\ldots\varepsilon_k$.

The capital discovery that such situations exist was made by R.A. Fisher in 1922, and following his terminology we shall call $a_.$ 'a sufficient statistic - or estimator - for the parameter in question.

In the present situation, however, the sufficiency of $a_.$ needs a qualification as being relative, since it rests upon the condition that ε_i's are known. As long as such knowledge is not available the sufficiency as such is not very helpful, but the important point of (5.17) then is that it depends solely upon the ε's, not on ξ.

From (5.17) we may therefore proceed as we did from (4.8) - which actually is a special case of (5.17), viz. k=2 - considering a collection of subjects which all happened to have $a_. = r$.

Specifying by $a_{\nu i}$ the a_i of subject no. ν and denoting by $(a_{\nu i})$, given ν, the set of responses $a_{\nu 1},\ldots a_{\nu k}$, i.e.

$$(a_{\nu i}) = (a_{\nu 1},\ldots,a_{\nu k}) \tag{5.18}$$

we may rewrite (5.17) in the form

$$p\{(a_{\nu i})\,|\,a_{\nu .}=r\} = \frac{\varepsilon_1^{a_{\nu 1}}\ldots\varepsilon_k^{a_{\nu k}}}{\gamma_r}\,, \quad \nu = 1,\ldots,n\,. \tag{5.19}$$

The responses of the n persons being independent their joint probability is obtained by multiplying the n probabilities (5.19). Denoting for short the whole set of nxk responses by $((a_{\nu i}))$ - the double bracket indicating variation over both ν and i - we get

$$p\{((a_{\nu i}))\,|\,(a_{\nu .}=r)\} = \frac{\varepsilon_1^{a_{.1}}\ldots\varepsilon_k^{a_{.k}}}{\gamma_r^n} \tag{5.20}$$

where

$$a_{.i} = \sum_{\nu=1}^{n} a_{\nu i}\,. \tag{5.21}$$

(5.20) implies that - as a consequence of the model - we have to deal with the total number of correct answers to each item for the n persons in question.

However, the derivation of their joint distribution we shall take as a special case of the further generalization in the following section.

6. Separation of parameters.

Let us finally consider the responses of n individuals with the parameters $\xi_1, \ldots, \xi_n$ to k items with the parameters $\varepsilon_1, \ldots, \varepsilon_k$. With the notation $a_{\nu i}$ introduced in the last section the model (5.2) now takes the form

$$p\{a_{\nu i} | \xi_\nu, \varepsilon_i\} = \frac{(\xi_\nu \varepsilon_i)^{a_{\nu i}}}{1+\xi_\nu \varepsilon_i} \quad , \tag{6.1}$$

and on the assumption of stochastic independence of all of the responses $a_{\nu i}$, $\nu = 1, \ldots n$, $i = 1, \ldots k$, the joint probability of the whole set $((a_{\nu i}))$ of them becomes

$$\begin{aligned} p\{((a_{\nu i}))|(\xi_\nu),(\varepsilon_i)\} &= \prod_{\nu=1}^{n} \prod_{i=1}^{k} p\{a_{\nu i}|\xi_\nu,\varepsilon_i\} \\ &= \frac{\prod_{\nu=1}^{n} \prod_{i=1}^{k} (\xi_\nu \varepsilon_i)^{a_{\nu i}}}{\prod_{\nu=1}^{n} \prod_{i=1}^{k} (1+\xi_\nu \varepsilon_i)} \quad . \end{aligned} \tag{6.2}$$

As regards the numerator we notice that the parameter ξ_ν occurs in k places, each time raised to a power $a_{\nu i}$ which altogether makes $\xi_\nu^{a_{\nu\cdot}}$ and that the parameter ε_i occurs in n places, each time raised to a power $a_{\nu i}$, adding up to a total power of $a_{\cdot i}$. If furthermore the denominator is denoted by

$$\gamma((\xi_\nu),(\varepsilon_i)) = \prod_{\nu=1}^{n} \prod_{i=1}^{k} (1+\xi_\nu \varepsilon_i) \tag{6.3}$$

we may simplify (6.2) to

$$p\{((a_{\nu i}))|(\xi_\nu),(\varepsilon_i)\} = \frac{\prod_{\nu=1}^{n} \xi_\nu^{a_{\nu\cdot}} \cdot \prod_{i=1}^{k} \varepsilon_i^{a_{\cdot i}}}{\gamma((\xi_\nu),(\varepsilon_i))} \quad . \tag{6.4}$$

This formula is the generalization of (5.3) to n persons, but in consequence of (6.4) we now have to derive the probability that $a_{1\cdot}, \ldots a_{n\cdot}$ and $a_{\cdot 1}, \ldots a_{\cdot k}$ take on two specified sets of values $r_1, \ldots, r_n$ and $s_1, \ldots s_k$.

In analogy to section 5. - cf. in particular the logical chain of (5.11) through (5.13)-we should find all possible ways of building up zero - one - matrices $((a_{\nu i}))$, that have the same row totals $a_{\nu\cdot} = r_\nu$, $\nu = 1, \ldots n$ and column totals $a_{\cdot i} = s_i$, $i=1, \ldots k$, state the probability of each realization and add up all such probabilities to a total joint probability of the two sets of totals considered. However, this procedure is greatly simplified by the fact that

all the probabilities to be added are equal, viz. - according to (6.4) -

$$(6.5) \qquad \frac{\prod_{\nu=1}^{n} \xi_\nu^{r_\nu} \cdot \prod_{i=1}^{k} \varepsilon_i^{s_i}}{\gamma((\xi_\nu),(\varepsilon_i))} \ .$$

Thus we just have to count the number of different ways in which it is algebraically possible to build up a zero - one - matrix with the row totals of $r_\nu, \nu=1,\ldots,n$ and the column totals of s_i, $i=1,\ldots,k$.

Determining this number is a combinatorial problem and it appears to be rather difficult, but at present we need nothing more than a notation. For this number we shall write

$$(6.6) \qquad \begin{bmatrix} r_1,\ldots,r_n \\ s_1,\ldots,s_k \end{bmatrix} = \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix}$$

and then we have

$$(6.7) \qquad p\left\{(a_{\nu\cdot}=r_\nu),(a_{\cdot i}=s_i)\,\middle|\,(\xi_\nu),(\varepsilon_i)\right\} = \begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix} \frac{\prod_{\nu=1}^{n} \xi_\nu^{r_\nu} \prod_{i=1}^{k} \varepsilon_i^{s_i}}{\gamma((\xi_\nu),(\varepsilon_i))} \ .$$

This joint probability distribution of the row totals $a_{\nu\cdot}$ and the column totals $a_{\cdot i}$ contains just as many parameters as observables, and the latter would therefore seem suitable for estimation purposes. How true this is becomes clear when we divide (6.7) into (6.4)(or (6.5)) to obtain the probability of the whole set of observations, on the condition that the totals of rows and columns are given. In fact, all parametric terms cancel so we are left with a conditional probability

$$(6.8) \qquad p\left\{((a_{\nu i}))\,\middle|\,(a_{\nu\cdot}=r_\nu),(a_{\cdot i}=s_i)\right\} = \frac{1}{\begin{bmatrix} (r_\nu) \\ (s_i) \end{bmatrix}}$$

that is independent of all of the parameters.

Therefore, once the totals have been recorded any further statement as regards <u>which of the items</u> were answered correctly by <u>which persons</u> is, according to our model, <u>useless as a source of information about the parameters</u>. (Which other use may be made of the $a_{\nu i}$'s will emerge at a later stage of our discussion). Thus the row totals and the column totals are <u>not only suitable</u> for estimating the parameters, they simply <u>imply every possible statement about the parameters that can be made on the basis of the observations</u> $((a_{\nu i}))$. Accordingly we shall, in continuation of the terminology introduced in sect. 5, characterize the row totals $a_{\nu\cdot}$, $\nu=1,\ldots n$ and the column totals $a_{\cdot i}$, $i=1,\ldots k$ as a <u>set of sufficient estimators for the parameters</u> $\xi_1,\ldots\xi_n$ and $\varepsilon_1,\ldots\varepsilon_k$.

As (6.7) contains both sets of parameters a direct utilization of this formula would apparently lead to a simultaneous estimation of both sets. However, in view of previous results, cf. the comments following (5.17), it would seem appropriate to ask whether it is possible - also in this general case - to estimate the item parameters independently of the person parameters and, if so, vice versa as well.

In order to approach this problem we shall first derive the distribution of the row totals, appearing as exponents of the ξ's, irrespective of the values of the column totals, by summing (6.7) over all possible combinations of $s_1,\ldots s_k$. During this summation the denominator $\gamma((\xi_\nu),(\varepsilon_i))$ keeps constant and the same holds for the terms $\xi_\nu^{r_\nu}$, $\nu=1,\ldots,n$ in the numerator. Thus, on introducing the notation

$$\gamma_{(r_\nu)}((\varepsilon_i)) = \sum_{(s_i)} \begin{bmatrix}(r_\nu)\\(s_i)\end{bmatrix} \varepsilon_1^{s_1}\ldots\varepsilon_k^{s_k} \tag{6.9}$$

we obtain

$$p\{(a_{\nu\cdot}=r_\nu)\,|\,(\xi_\nu),(\varepsilon_i)\} = \frac{\gamma_{(r_\nu)}((\varepsilon_i))\cdot\prod_{\nu=1}^{n}\xi_\nu^{r_\nu}}{\gamma((\xi_\nu),(\varepsilon_i))} \tag{6.10}$$

from which it is seen that the ξ_ν's might be estimated from the row totals if the ε_i's - and therefore also the polynomials (6.9) - were known.

Similarly we may sum (6.7) over all possible combinations of $r_1,\ldots r_n$, keeping $s_1,\ldots s_k$ fixed. Substituting in (6.9) $\xi_1,\ldots\xi_n$ for $\varepsilon_1,\ldots\varepsilon_k$ and in consequence interchanging the r's and the s's we get

$$\gamma_{(s_i)}((\xi_\nu)) = \sum_{(r_\nu)} \begin{bmatrix}(s_i)\\(r_\nu)\end{bmatrix} \xi_1^{r_1}\ldots\xi_n^{r_n} \tag{6.11}$$

where by the way

$$\begin{bmatrix}(s_i)\\(r_\nu)\end{bmatrix} = \begin{bmatrix}(r_\nu)\\(s_i)\end{bmatrix} . \tag{6.12}$$

With this notation the summation yields on analogy to (6.10):

$$p\{(a_{\cdot i}=s_i)\,|\,(\xi_\nu),(\varepsilon_i)\} = \frac{\gamma_{(s_i)}((\xi_\nu))\cdot\prod_{i=1}^{k}\varepsilon_i^{s_i}}{\gamma((\xi_\nu),(\varepsilon_i))} , \tag{6.13}$$

and accordingly the ε_i's might be estimated from the column totals provided the ξ_ν's were known.

Thus, we might estimate the ξ's if the ε's were known, and the ε's if the ξ's were known! And both estimations would even have been relatively sufficient. In fact, on dividing (6.1o) into (6.7) to obtain the conditional probability of $(a_{.i})$ for given $(a_{\nu.})$ we get

$$(6.14) \qquad p\{(a_{.i}=s_i)|(a_{\nu.}=r_\nu),(\xi_\nu),(\varepsilon_i)\} = \left[\begin{matrix}(r_\nu)\\(s_i)\end{matrix}\right] \frac{\prod_{i=1}^{k} \varepsilon_i^{s_i}}{\gamma_{(r_\nu)}((\varepsilon_i))}$$

which is independent of the parameters ξ_ν to be estimated. And similarly the division of (6.13) into (6.1o) gives

$$(6.15) \qquad p\{(a_{\nu.}=r_\nu)|(a_{.i}=s_i),(\xi_\nu),(\varepsilon_i)\} = \left[\begin{matrix}(r_\nu)\\(s_i)\end{matrix}\right] \frac{\prod_{\nu=1}^{n} \xi_\nu^{r_\nu}}{\gamma_{(s_i)}((\xi_\nu))}$$

which is independent of the ε's.

But, of course, as long as neither set of parameters is known these possibilities are of no avail.

It is one of the characteristic features of the model under consideration that this vicious circle may be broken, the instrument being a reinterpretation of the formulae (6.14) and (6.15).

In fact, as (6.14) depends on the ε's, but not on the ξ's, this formula gives the opportunity of estimating the ε's without dealing with the ξ's. Thus the objections to both (6.7) and (6.13) have been eliminated. The unknown ξ's in these expressions have been replaced by observable quantities, viz. the individual totals $a_{\nu.}$.

Similarly in (6.15) the ε's of (6.7) and (6.1o) have been replaced by the item totals $a_{.i}$, in consequence of which we may estimate the ξ's without knowing or simultaneously estimating the ε's.

Thus the estimation of the two sets of parameters may be separated from each other.

In this connection we may return to (6.8), noticing that this formula is a consequence of the model structure - (3.3) and the stochastic independence - irrespective of the values of the parameters of which the right hand term is independent.

Therefore, if from a given matrix $((a_{\nu i}))$ we construct a quantity which would be useful for disclosing a particular type of departure

from the model, then its sampling distribution as conditioned by the marginals $(a_{\nu.})$ and $(a_{.i})$ will be independent of all of the parameters.

Thus the testing of the model may be separated from dealing with the parameters.

The question of how to perform such testing in practice and also that of turning the observed row and column totals into adequate estimates of the ξ's and the ε's we shall not enter upon on this occasion.

In [1], chapter VI these questions were dealt with by simple methods which were taken to be acceptable approximations. In case of the numerical sequences the observations passed the test for the model satisfactorily, but the model failed completely in the case of the geometrical shapes. In the latter subtest the time allowance for some technical reasons had been cut down below the optimal limit, but a reanalysis of the data - to be reported elsewhere - has shown that when allowance is made for the working speed for each subject, then the data fit the model just as well as for the numerical sequences.

However, from a theoretical point of view the method used was unsatisfactory (cf. [1], chapter X, in particular pp. 181-182). By now we are in the process of working out better methods, and therefore we shall, for the time being, leave the documentation of the applicability of the model at a reference to the earlier work.

7. Generalization of the model.

As a possible generalization to the case of more responses than two the following model may be suggested.

Consider a number of subjects (ν=1,...,n) being exposed to the same set of stimuli (i=1,...,k). In each case is recorded one response out of a finite set of possible responses

$$x^{(1)},\ldots,x^{(\mu)},\ldots,x^{(m)} \quad , \tag{7.1}$$

the set being the same in all cases.

In each case a probability is allotted to each response category $x^{(\mu)}$:

$$p\{x^{(\mu)} | \nu,i\} = \frac{\lambda_{\nu i}^{(\mu)}}{\gamma_{\nu i}} \tag{7.2}$$

where $\lambda_{\nu i}^{(1)},\ldots,\lambda_{\nu i}^{(m)}$ are positive numbers adding up to $\gamma_{\nu i}$. With the clause that the responses, given all the parameters $\lambda_{\nu i}^{(\mu)}$, are stochastically independent these requirements generalize the assumptions 1) and 3) of sect. 3.

In an attempt at generalizing the remaining assumption (3.2) we shall simply assume that a multiplicative rule holds for each category, i.e.

$$\lambda_{\nu i}^{(\mu)} = \xi_{\nu}^{(\mu)} \varepsilon_i^{(\mu)} \tag{7.3}$$

for all ν and i.

In order to see (7.3) as generalizing (3.2) we may consider the case m = 2 when

$$\gamma_{\nu i} = \xi_{\nu}^{(1)} \varepsilon_i^{(1)} + \xi_{\nu}^{(2)} \varepsilon_i^{(2)} \tag{7.4}$$

and therefore

$$\begin{cases} p\{x^{(1)} | \nu, i\} = \dfrac{\xi_{\nu}^{(1)} \varepsilon_i^{(1)}}{\xi_{\nu}^{(1)} \varepsilon_i^{(1)} + \xi_{\nu}^{(2)} \varepsilon_i^{(2)}} \\ p\{x^{(2)} | \nu, i\} = \dfrac{\xi_{\nu}^{(2)} \varepsilon_i^{(2)}}{\xi_{\nu}^{(1)} \varepsilon_i^{(1)} + \xi_{\nu}^{(2)} \varepsilon_i^{(2)}} \end{cases} . \tag{7.5}$$

On dividing through by $\xi_{\nu}^{(2)} \varepsilon_i^{(2)}$ and introducing the notations

$$\xi_{\nu} = \frac{\xi_{\nu}^{(1)}}{\xi_{\nu}^{(2)}} , \quad \varepsilon_i = \frac{\varepsilon_i^{(1)}}{\varepsilon_i^{(2)}} \tag{7.6}$$

we obtain

$$\begin{cases} p\{x^{(1)} | \nu, i\} = \dfrac{\xi_{\nu} \varepsilon_i}{\xi_{\nu} \varepsilon_i + 1} \\ p\{x^{(2)} | \nu, i\} = \dfrac{1}{\xi_{\nu} \varepsilon_i + 1} \end{cases} , \tag{7.7}$$

which is equivalent to (3.2).

With suitable notations the formal theory of the model for m>2 runs perfectly parallel to the developments of the sections 4 through 6, but as the algebra required is somewhat more advanced we shall leave it aside as being for the present purpose less material than the conclusion which is quite analogous to the main result of sect. 6:

From an analogue of (6.14) it follows that it is possible to estimate and otherwise appraise the stimulus parameters $\varepsilon_i^{(\mu)}$, i=1,...k, μ=1,...m without implying the subject parameters $\xi_\nu^{(\mu)}$, ν=1,...n,μ=1,...m.

Similarly it follows from an analogue of (6.15) that it is possible to estimate and otherwise appraise the subject parameters $\xi_\nu^{(\mu)}$ without implying the stimulus parameters $\varepsilon_i^{(\mu)}$.

From an analogue of (6.8) it follows that it is possible to direct the testing of the model structure - as given by (7.2), (7.3) and the independence - in such a way that the test becomes independent of all of the parameters.

The first two statements we may call the mutual separability of the parameters while the third statement may be termed the separability of the model structure from the parameters. (cf. [1], chapter X, sect. 5 and [2]).

8. Specific objectivity.

The formula (6.15) or its analogue may of course be applied to any subgroup of the total collection of subjects having been exposed to the k stimuli. Thus the parameters of the subjects in the subgroup may be evaluated without regard to the parameters of the other subjects.

In particular the parameters of any two subjects may be compared on their own virtues alone, quite irrespective of the group or population to which - for some reason - they may be referred. Thus, as indicated in the introduction, the new approach, when applicable, does rule out populations from the comparison of individuals.

Similarly the formula (6.14) or its analogue may be applied to any subset of the k stimuli, and accordingly their parameters may be evaluated without regard to the parameters of the other stimuli. In particular the parameters of any two stimuli may be compared separately.

With these additional consequences the principle of separability leads to a singular objectivity in statements about both parameters and model structure.

In fact, the comparison of any two subjects may be carried out in such a way that no other parameters are involved than those of the two subjects - neither the parameter of any other subject nor any of the stimulus parameters.

Similarly, any two stimuli may be compared independently of all other parameters than just those of the two stimuli - the parameters of all other stimuli as well as the parameters of the subjects having been replaced by observable numbers.

It is suggested that comparisons carried out under such circumstances will be designated as "specifically objective". And the same term would seem appropriate for statements about the model structure which are independent of all of the parameters specified in the model, the unknown values of them being, in fact, irrelevant for the structure of the model.

Of course, specific objectivity is no guarantee against the subjectivity of the statistician when he choses his fiducial limits or when he judges about which kind of deviations from the model he will look for. Neither does it save the statistician from the risk of being offered data marred by the subjective attitude of the psychologist during his observations.

Altogether, when introducing the concept of specific objectivity I am not entering upon a general philosophical debate on the meaning and the use of objectivity at large. At present the term is strictly limited to observational situations that can be covered by the stimulus - subject-response scheme, to be described in term of a probabilistic model which specifies parameters for stimuli and for subjects. And the independence of unwarranted parameters, entering into the characterization of specific objectivity, pertains only to such parameters as are specified in the model.

What has been demonstrated in details in the case of two response categories and indicated for a finite number of categories then is that the specific objectivity in all three directions can be attained in so far as the type of model defined by (7.1), (7.2) and the independence holds.

Recently it has been shown that - but for unimportant mathematical restrictions - the inverse statement is also true:

If a set of observational situations can at all be described by a probabilistic model (7.1), including the independence, then the multiplicative rule (7.2) is also necessary for obtaining specific objectivity as regards both sets of parameters and the model structure as well.

In particular, if only two responses are available then the observations must conform to the simple model (7.6) (or equivalently (3.3)) if it be possible to maintain specific objectivity in statements about subjects, stimuli and model.

9. Reduction of parameters.

The general model allots m parameters to a subject and equally many to a stimulus, and even if, by an argument similar to (7.4) through (7.7), the number may be reduced to m-1 it easily becomes unduly large.

As a case in point we may think of a number of texts, each of 2oo words, to be used in testing for proficiency in reading aloud. With number of errors as response categories each child as well as each text should, according to the model (7.2) and (7.3), be characterized

by no less than 2oo parameters ! However, in [1], chapter II data of this type were well described by a Poisson model with only one parameter per subject and per text. The number of errors being denoted by a, the model states that

$$p\{a|v,i\} = e^{-\lambda_{vi}} \cdot \frac{\lambda_{vi}^{a}}{a!} \tag{9.1}$$

with

$$\lambda_{vi} = \xi_v \varepsilon_i \ , \tag{9.2}$$

and in chapter VIII separability theorems leading up to specific objectivity have been proved.

This result points to an interpretation of the basic model (7.2)-(7.3) that becomes crucial for applying it in practice, viz. that it is not necessary to assume that the m parameters for a subject or the m parameters for a stimulus are functionally independent. If in the present case we replace μ by a and write

$$\xi_v^{(a)} = \xi_v^a \ , \ \varepsilon_i^{(a)} = \frac{\varepsilon_i^a}{a!} \ , \ \gamma_{vi} = e^{\xi_v \varepsilon_i} \tag{9.3}$$

and if we allow for an infinite, but enumerable set of categories (a=0,1,2,...), the model reduces to (9.1)-(9.2).

It is a trivial implication of the general theory that each $\xi_v^{(a)}$ and each $\varepsilon_i^{(a)}$ may be estimated objectively, the remarkable point being that the specific objectivity also holds for the new parameters ξ_v and ε_i in the reduced model.

This situation raises the question of when and how the basic model may be reduced to simpler models, i.e. models where the m parameters per subject and per stimulus can be expressed in terms of considerably fewer elements for which the specific objectivity is preserved.

This problem has also been solved and the solution was presented as formula (4.6) in [2] with a demonstration that the model possesses the separability property leading up to specific objectivity. So far, however, the proof that - within certain limitations - this type of model is the only mathematical possibility has not been published.

10. Fields of application.

The problems we have been dealing with in the present paper were formulated within a narrow field, viz. psychological test theory. However, with the generalization of sects. 7 and 9 and with the discovery of specific objectivity we have reached at concepts of such generality that the original limitation is no longer justified.

Extensions into other fields of psychology, such as psychophysical threshold experiments and experiments on perception of values offer themselves, but the stimulus-subject-response framework is by no means limited to psychology.

Thus in a recent publication [3] the above mentioned Poisson model was employed in an investigation of infant mortality in Denmark in the period 1931-1960.

In each year the number of infant deaths (of all causes or of a particular cause) out of the number of children born was recorded for both boys and girls, born in or out of wedlock. In this case the years served as "subjects", the combination of sex and legitimacy of the children as the "stimulus", while the number of infant deaths out of the number of children born were the responses.

From economy we may take household budgets as an example. The families serve as "subjects", income and expenditures as classified into a few types, serve as "stimuli", while the amount earned or spent are the responses.

These examples may indicate that the framework covers a rather large field, at least within Social Sciences. Delineating the area within which the models described here apply is a huge problem, the enquiry into which has barely started.

But already the two intelligence tests mentioned in sect. 2 and discussed at the end of sect. 6 are instructive as regards the sort of difficulties we should be prepared to meet. For one of them, the numerical sequences, the analysis in [1] Chapter VI showed a perfectly satisfactory fit of the observations to the model, i.e. in this case specific objectivity may be obtained on the basis of the response pattern for each subject. For the other test, the geometrical shapes, the analysis most unambiguously showed that the separability did not hold. As a matter of fact, in the first case the analysis, in agreement with the theory, ended up with a bunch of parallel lines with unit slope, while in the second case I got a family of straight lines with all sorts of slopes, in complete disagreement with the model.

The same kind of picture had been obtained in[a/] different intelligence test which, however, was of the omnibus type, containing items that presumably called upon very different intelligence functions. In this case, therefore, the data could not be expected to allow for a description comprising only one parameter for each subject.

The items in the numerical sequences are much more uniform in that they require that the testee realizes a logical structure in a sequence of numbers. And according to the analysis the items were sufficiently uniform - although of very different levels of difficulty - to allow for a description of the data by one parameter only for each subject as well as for each item.

The items of the figure test were constructed just as uniformly as the numerical sequences and therefore it was somewhat of a surprise that they turned out quite adversely.

To this material I could add observations on two other tests,

constructed with equal care. One was a translation of the idea in Raven's Matrix Test into letter combinations, at the same time substituting the multiple choice by a construction, on the part of the testee, of the answer. For this test the results were just as beautiful as for the numerical sequences. The other one was a set of verbal analogies where the number of answers offered was practically infinite, with the effect that the multiple choice was in fact eliminated. Here the results of the testing were just as depressing as for the figure test.

This contrast, however, led to the resolution of the mystery. The difference between the two pairs of tests was not one in construction principles, but one in the administration of the tests.

For all four tests the adequate time allowance was determined by means of special experiments. On applying them to random samples of 2oo recruits it turned out that the number of correct answers formed a convenient distribution for the letter matrix test and for the numerical sequences, but verbal analogies and the figure test were too easy, so the distributions showed an undesirable accumulation of many correct answers.

This happened in 1953 when only the barest scraps of the theory had been developed, and yielding to a considerable time pressure the test constructor, consulting me on the statistical part of the problem, severely cut down the time allowances so as to move the distributions to the middle of the range. And while succeeding in that we spoiled the test, turning it into a mixture of a test for capacity and a test for speed!

More recently, however, I have had the opportunity of reanalyzing both sets of data, grouping primarily the subjects according to their working speed, as given by the number of items done, and applying to each group the technique of [1], chapter VI. The result was startling: Within each speed group I found the bunch of parallel lines with unit slope required by the theory, and their mutual distances - measuring the relative values of the logε's - were equal in overlapping items, i.e. the relative difficulties of the items were independent of the working speed. Altogether, with speed as an ancillary information specific objectivity may be attained as regards the properties the tests really aimed at measuring.

Turning the final statement upside down we get the morale of this story: <u>Observations may easily be made in such a way that specific objectivity, otherwise available, gets lost</u>.

This, for instance, easily happens when qualitative observations with, say, 5 categories of responses for convenience are grouped into 3 categories. If the basic model holds for the 5 categories it is mathematically almost impossible for the 3-categories-model also to hold. Thus the grouping, tempting as it may be, will usually tend to slur the specific objectivity.

In concluding I therefore feel the necessity of pointing out that the problem of the relation of data to models is not only one of trying to fit data to an adequately chosen model from our inventory and see if it works, a question in the opposite direction is equally relevant: <u>How to observe in such a way that specific</u>

objectivity obtains ?

Literature:

[1] G.Rasch, Probabilistic Models for Some Intelligence and Attainment Tests. The Danish Institute for Educational Research 1960.

[2] G.Rasch, On General Laws and the Meaning of Measurement in Psychology. PROCEEDINGS OF THE FOURTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, University of California Press 1960.

[3] P.C.Matthiessen, Spædbørnsdødeligheden i Danmark 1931-60. Det statistiske Departement, København 1964.