DOCUMENT RESUME

ED 201 670 TM 810 265

AUTHOR Takane, Yoshio
TITLE Maximum Likelihood Additivity Analysis.
PUB DATE May 80
NOTE 19p.: Paper presented at the Annual Meeting of the Psychometric Society (Iowa City, IA, May 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Collection; Elementary Education; *Factor Analysis; *Mathematical Models; *Maximum Likelihood Statistics; Multidimensional Scaling
IDENTIFIERS *Likelihood Function Estimation

ABSTRACT

A maximum likelihood estimation procedure is developed for the simple and the weighted additive models. The data are assumed to be taken by either one of the following methods: (1) categorical ratings--the subject is asked to rate a set of stimuli with respect to an attribute of the stimuli on rating scales with a relatively few observation categories; (2) pair comparisons--the subject is asked to judge which one of two stimuli presented at a time dominates the other in some respect; or (3) directional rankings--the subject is asked to rank order stimuli in a specific direction (i.e., from the smallest to the largest or the other way around). Although the procedure mainly focuses on these three experimental methods for data collection, it is by no means restricted to them. Practical uses of the procedure are reported with an emphasis on various advantages of the procedure as a statistical method. (Author/RL)

Maximum Likelihood Additivity Analysis*

Yoshio Takane

McGill University

ABSTRACT

A maximum likelihood estimation procedure is developed for the simple and the weighted additive models. The data are assumed to be taken by either one of the following methods: categorical ratings, pair comparisons or directional rankings. Practical uses of the procedure are reported with an emphasis on various advantages of the procedure as a statistical method.

* Paper presented to the Psychometric Society meeting at the University of Iowa, May 1980.

1. Introduction

In this paper we discuss a maximum likelihood estimation procedure for additivity analysis from a variety of nonmetric data. Existing methods for nonmetric additivity analysis, notably MONANOVA [Kruskal, 1965], ADDALS [de Leeuw, Young & Takane, 1976] and WADDALS [Takane, Young & de Leeuw, 1980], are all based on the least squares principle, and are primarily descriptive in nature. The maximum likelihood method developed in this paper, on the other hand, allows various kinds of statistical inferences including tests of hypotheses about the model.

2. The Model

For illustrative convenience we only discuss the two-factor case. The method is readily generalizable to higher order designs.

Let α_i and β_j denote additive effects of the i^{th} level of Factor A and the j^{th} level of Factor B, respectively. We state the simple additive model as:

(1) $$y_{ij} = \alpha_i + \beta_j ,$$

where y_{ij} is the predicted model value for the combination of the i^{th} level of Factor A and the j^{th} level of Factor B. We also consider the weighted additive model which is stated as

(2) $$y_{ijk} = w_{k\alpha}\, \alpha_i + w_{k\beta}\, \beta_j ,$$

where y_{ijk} is the subject-specific (subjeck k) model value for the (i, j) combination, and $w_{k\alpha}$ and $w_{k\beta}$ are the weights attached to Factor A and Factor B, respectively, by subject k. This model accounts for individual differences in additivity by differential weightings of additive factors.

(The nature of this model as well as its relationship to the simple additive model (1) is fully described in Takane, Young & de Leeuw, [1980].) In order to eliminate scale indeterminacies in additive effects and individual differences weights, we may require that

$$(3) \qquad \sum_i(\alpha_i - \sum_m \alpha_m/n_\alpha)^2 = n_\alpha \quad \text{and} \quad \sum_j(\beta_j - \sum_m \beta_m/n_\beta)^2 = n_\beta$$

where n_α and n_β are the numbers of levels in Factor A and Factor B, respectively.

3. The Data

We assume that the data are collected by one of the following methods: categorical ratings, pair comparisons or directional rankings. In the categorical rating method the subject is asked to rate a set of stimuli with respect to an attribute of the stimuli on rating scales with a relatively few observation categories. In the pair comparison method, on the other hand, the subject is asked to judge which one of two stimuli presented at a time dominates the other in some respect. Finally, in the directional ranking method the subject is asked to rank order stimuli in a specific direction (i.e. from the smallest to the largest or the other way round). Although our procedure mainly focuses on the above three experimental methods for data collection, it is by no means restricted to the above three. In fact it has been shown [Takane & Carroll, in preparation] that treatments of conditionalities, missing data and tied observations in the directional ranking method allow a still wider range of data collection methods to be handled within the framework of the directional ranking method.

4. The Method

Maximum likelihood multidimensional scaling procedures have already been developed for the three data collection methods mentioned above [Takane, 1978; Takane, 1980; Takane & Carroll, in preparation]. The construction of likelihood functions in the current procedure is very similar to that in its MDS counterparts with the difference being that the former fits the additive model, while the latter fits the distance model. Here we only briefly discuss general strategic schemes for the construction of likelihood functions specific to the data collection methods. Details as well as modifications of the basic schemes (e.g., a provision for tied observations) may be found in the references given above.

4.1. Categorical ratings

Let

$$\lambda_{ijk} = y_{ijk} + e_{ijk}$$

where $e_{ijk} \sim N(o,\sigma_k^2)$. The probability that stimulus o_{ijk} (defined by the i^{th} level of Factor A and the j^{th} level of Factor B as perceived by subject k) is judged to be in the m^{th} category (C_m) is stated as

$$\Pr(o_{ijk} \varepsilon C_m) = \Pr(b_{k(m-1)} < \lambda_{ijk} < b_{km})$$

$$= \int_{a_{ijk(m-1)}}^{a_{ijkm}} \phi(z)\ dz \equiv p_{ijkm}$$

where b_{km} and $b_{k(m-1)}$ are upper and lower boundaries of category m for subject k, $a_{ijkm} = (b_{km} - y_{ijk})/\sigma_k$ (m=1, ..., M where M is the number of observation categories), and where ϕ is the standard normal density function. We may impose various restrictions on category boundaries [Takane, 1980].

Let Z_{ijkm} denote the observed frequency with which o_{ijk} is put category m by subject k. Then the joint probability of Z_{ijkm} (m=1, M) is given by

$$p_{ijk} = \prod_{m=1}^{M} p_{ijkm}^{Z_{ijkm}} \quad .$$

The joint likelihood of the total set of observations is, in turn, stated as

$$L = \prod_{k,i,j} p_{ijk} \; .$$

4.2. Pair comparisons

The probability that o_{ijk} is judged to be larger than o_{lmk} is given by

$$\Pr(o_{ijk} \succ o_{lmk}) = \Pr(\lambda_{ijk} > \lambda_{lmk})$$

$$= \int_{-\infty}^{a_{ijlmk}} \phi(z)\ dz \equiv p_{ijlmk}$$

where $a_{ijlmk} = (y_{ijk} - y_{lmk})/(\ 2\ _k)$. Let Z_{ijlmk} be the frequency with which o_{ijk} is judged larger than o_{lmk} out of N_{ijlmk} replications. The likelihcod of the total set of observations can then be written as

$$L = \prod p_{ijlmk}^{N_{ijlmk}} (1 - p_{ijlmk})^{N_{ijlmk} - Z_{ijlmk}}$$

[Takane, 1978].

4.3. Directional rankings

Let $o_k^{(1)} \succ o_k^{(2)} \succ \ldots. \succ o_k^{(M)}$ be the observed ranking. Let $y_k^{(m)}$ be the model value corresponding to $o_k^{(m)}$. We assume that the ranking is obtained from the largest element to the smallest by successive first choices. The probability of the m^{th} first choice,

$$p_k^{(m)} \equiv \Pr(o_k^{(m)} \succ o_k^{(m+1)}, \ldots, o_k^{(m)} \succ o_k^{(M)})$$

$$= \Pr(\lambda_k^{(m)} > \lambda_k^{(m+1)}, \ldots, \lambda_k^{(m)} > \lambda_k^{(M)})$$

is given by an appropriate integral of the multivariate normal distribution, which is approximated by

$$p_k^{(m)} = \frac{\exp(s_k y_k^{(m)})}{\sum_{j=m}^{M} \exp(s_k y_k^{(j)})},$$

where s_k is approximately $\pi/(\sqrt{3}\sigma_k)$. For the likelihood of a ranking we take a product of $p_k^{(m)}$; i.e., $p_k = \prod_{m=1}^{M} p_k^{(m)}$. Finally, the joint likelihood of multiple rankings obtained from different subjects is defined by the product of p_k over k. For treatments of ties see Takane and Carroll [in preparation].

4.4. Numerical method

The log likelihood may be optimized by various numerical methods. The current MAXADD, a FORTRAN program to perform the analysis described in this paper, uses Fisher's scoring algorithm. From a limited experience of the author the convergence is very quick and smooth.

5. Some Empirical Results

We present some empirical results obtained by the method proposed in this paper. The data we analyze pertain to developmental change in the structure of weights attached to height and width of rectangles in largeness judgments. Kempler [1971] constructed a set of 100 rectangles by factorially combining 10 height levels and 10 width levels each ranging from 10 inches to 14.5 inches in half-inch intervals. Four groups of children (1^{st}, 3^{rd}, 5^{th} and 7^{th} graders) judged each of the 100 stimuli as to whether it looked "large" or "small" (two-category rating judgments).

We previously analyzed the same set of data using a different procedure called WADDALS [Takane, Young & de Leeuw, 1980]. The first figure shows the change in weights attached to height and width of rectangles as a function of age groups (as revealed by the WADDALS procedure). We can see a fairly consistent tendency that the weight attached to height decreases with age. For this analysis we aggregated the data by age groups by counting the frequency with which each rectangle is judged as "large". This frequency was used as an ordinal measure of the perceived largeness of rectangles. The primary focus of this analysis, however, was on the group differences disregarding the individual differences within the groups. This may not be justifiable, so we have performed MAXADD analyses of individual data.

The first table summarizes the results of separate MAXADD analyses of Kempler's data by grade. The first column represents the weighted additive model (WAM) with individual differences in dispersion (σ_k), the second column the simple additive model (SAM) with σ_k and the last column SAM without individual differences in dispersion (σ). In all analyses category boundaries were allowed to vary over individuals. Three figures are reported in each cell of the table. The top one is the log likelihood multiplied by -2, the middle one is the effective number of parameters in the fitted model (d.f.), and the bottom the value of the AIC statistic [Akaike, 1974], which is defined by

AIC = -2 x log likelihood + 2 x d.f. of the model.

The best fitting model is the one with a minimum AIC value. We see that the weighted additive model is the best in all age groups according to this criterion. There seems to be substantial individual differences in the weight structure within the age groups. We can also observe that the differences in the AIC values between WAM and SAM tend to diminish as we go from the 1st graders to the 7th graders.

This point may be more clearly seen in the next four figures which display estimated individual weights for each age group. One tendency is evident; the plots of weight estimates tend to converge in the middle as the age goes up. For example, in grade 1 there are quite a few children who put a disproportionately large weight on height (including those two who totally ignore the width dimension), while those extreme subjects decrease in number and also in its degree until the majority of subjects put approximately equal weights on both height and width of rectangles. So the group differences we found previously with the WADDALS analysis of Kempler's data seem to be largely due to the difference in the constitution of the groups, which are heterogeneous in themselves, but which tend to get more homogeneous with age.

We have not done a joint analysis of all data in all age groups. The reason is simply that there are too many subjects. (For WAM with individual dispersions and category boundaries we need to estimate up to 300 parameters.) Consequently we had to resample portions of the available data set to examine the goodness of fit of the model in joint analyses. (The algorithm is being revised so that MAXADD can accommodate a problem of this size, however.) Two sets of data were subsampled, each consisting of data from 24 subjects (6 in each age group). A summary of the results are shown in the next table.

(The entries in this table are analogous to those in the previous table.) For both sets of data the joint analysis with WAM with individual dispersions and boundaries outperforms the others, indicating that the additive effects may be assumed common to all age groups, though the weights are different not only for different age groups but also for individuals within the groups.

References

Akaike, H. A new look at the statistical model identification. IEEE transactions on automatic control, 1974, 19, 716-723.

Kempler, B. Stimulus correlates of area judgments: A psychophysical developmental study. Developmental Psychology, 1971, 4, 158-163.

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformations of data. Journal of the Royal Statistical Society, Series B, 1965, 27, 251-265.

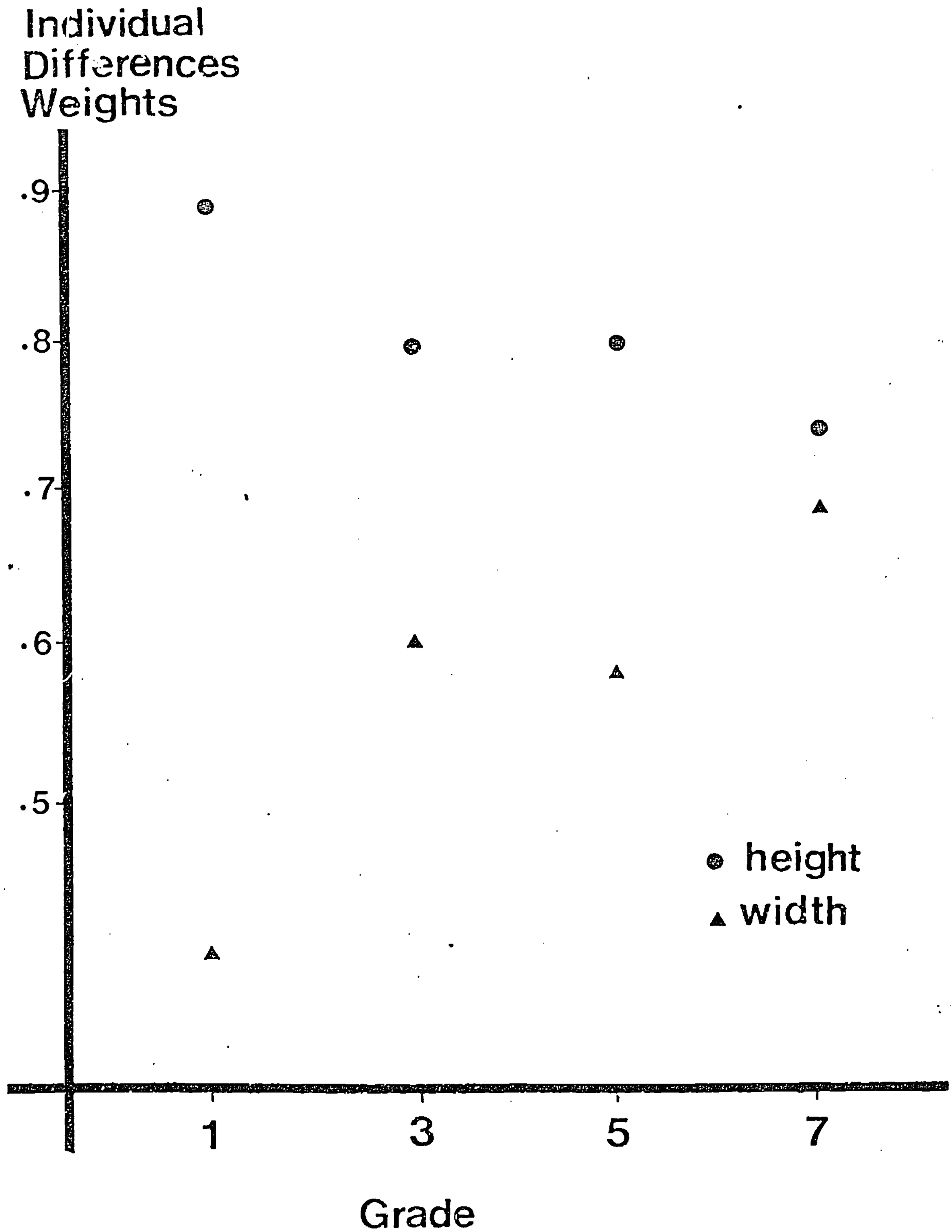
de Leeuw, J., Young, F.W., & Takane, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. Psychometrika, 1976, 41, 471-503.

Takane, Y. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent - theory and evaluations. Japanese Psychological Research, 1978, 20, 7-17 and 105-114.

Takane, Y. Multidimensional scucessive categories scaling: A maximum likelihood method. Paper submitted for publication, 1980.

Takane, Y., & Carroll, J. D. Maximum likelihood multidimensional scaling from directional rankings of similarities. Manuscript in preparation.

Takane, Y., Young, F. W., & de Leeuw, J. An individual differences additive model: an alternating least squares method with optimal scaling features. Psychometrika, 1980, 45, in press.

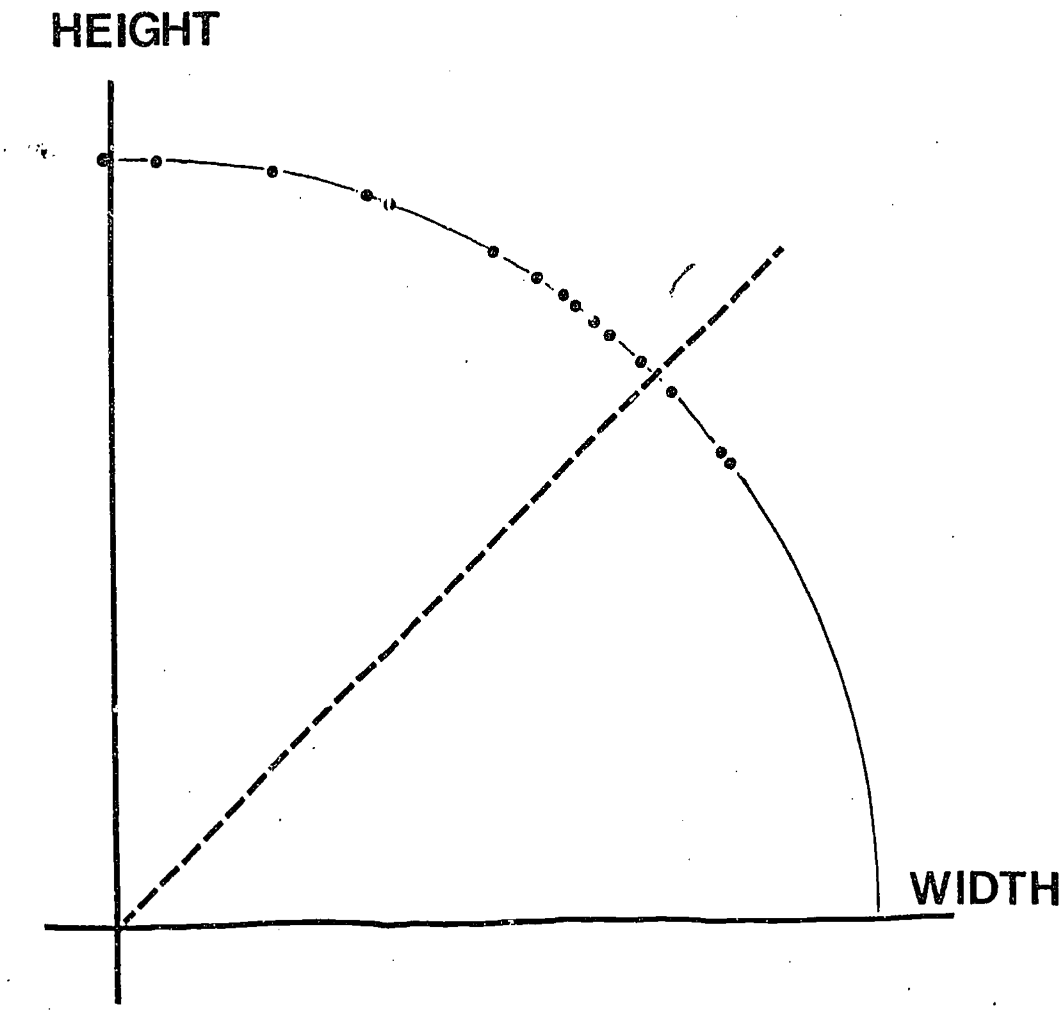


Change in weights attached to height and width of rectangles as a function of age-groups.

1

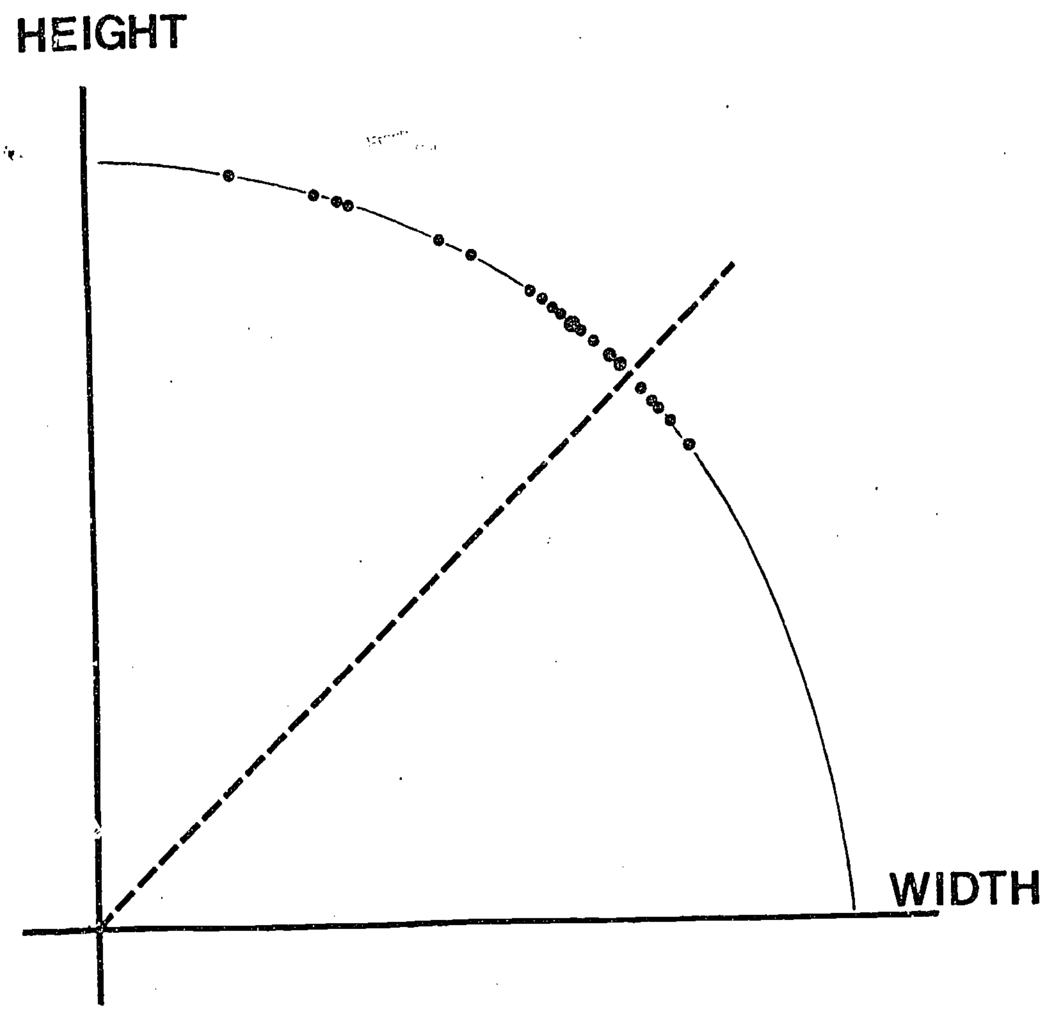
INDIVIDUAL DIFFERENCES WEIGHTS ATTACHED TO HEIGHT AND WIDTH

GRADE 1



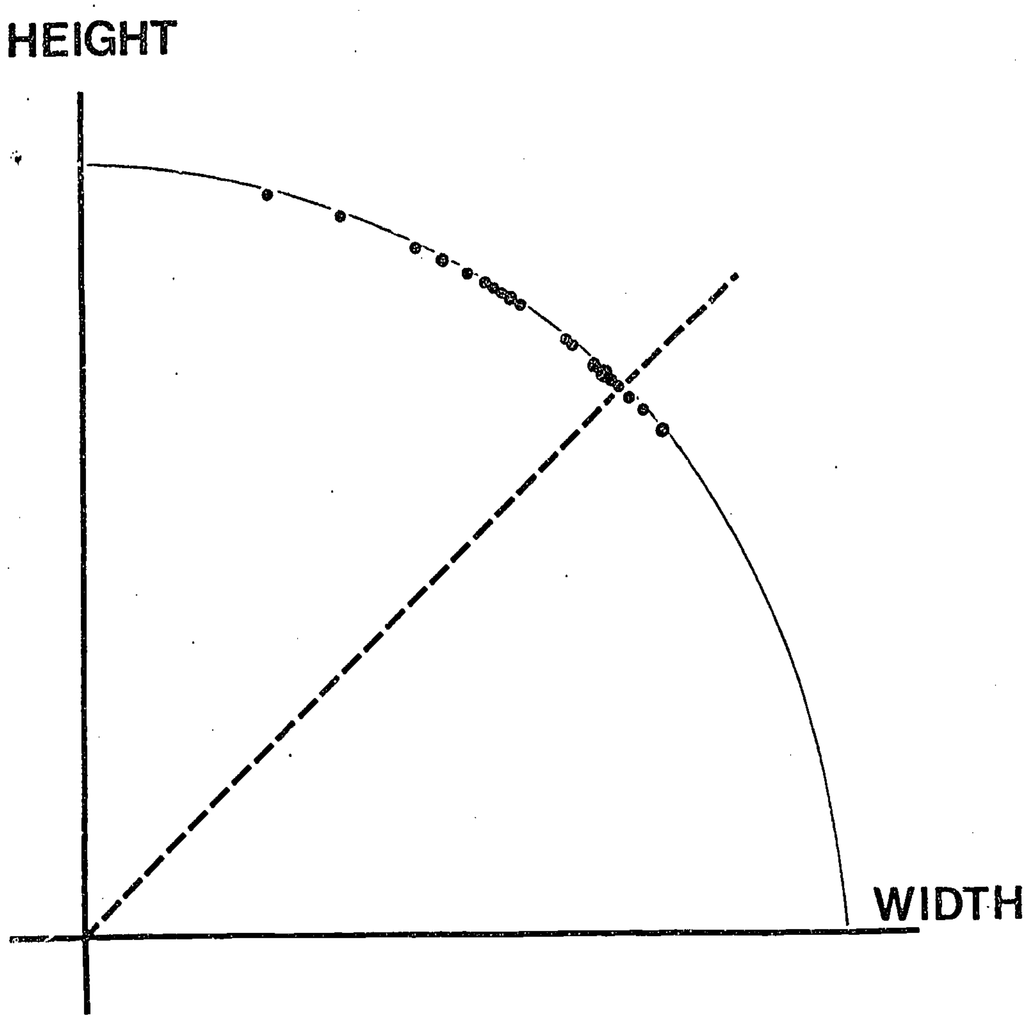
INDIVIDUAL DIFFERENCES WEIGHTS ATTACHED TO HEIGHT AND WIDTH

GRADE 3



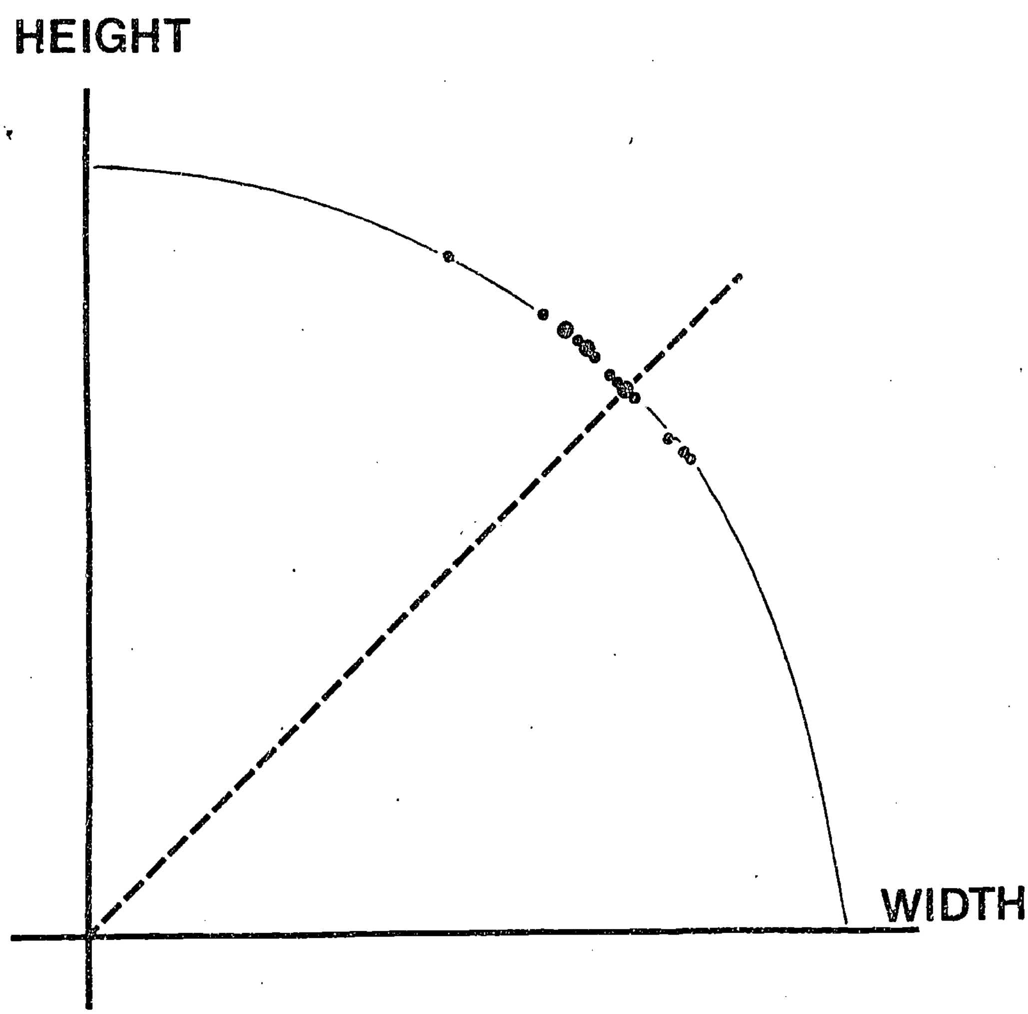
INDIVIDUAL DIFFERENCES WEIGHTS ATTACHED TO HEIGHT AND WIDTH

GRADE 5



INDIVIDUAL DIFFERENCES WEIGHTS ATTACHED TO HEIGHT AND WIDTH

GRADE 7



COMPARISONS OF SEPARATE ANALYSIS BY GRADE AND JOINT ANALYSIS

	Separate analysis (by grade)			Joint analysis		
	WAM σ_K	SAM σ_K	SAM σ	WAM σ_K	SAM σ_K	SAM σ
Set 1	2992.4	3130.9	3185.5	3050.0	3272.6	3371.2
	136	120	96	88	66	42
	3264.4	3370.9	3377.5	3226.0*	3404.6	3455.2
Set 2	2562.4	2742.1	2802.5	2634.5	2868.1	2930.5
	136	120	96	88	66	42
	2834.4	2982.1	2994.5	2810.5*	3000.1	3014.5

Legend:

-2 x log likelihood (+ const.)
D.F. of the model
AIC (+ const.)

* Minimum AIC solution