Research in Yield Forecasting

By

H. F. Huddleston, Principal Research Statistician,
Research Division
W. W. Wilson, Mathematical Statistician,
Research and Development Branch, Research Division

Introduction

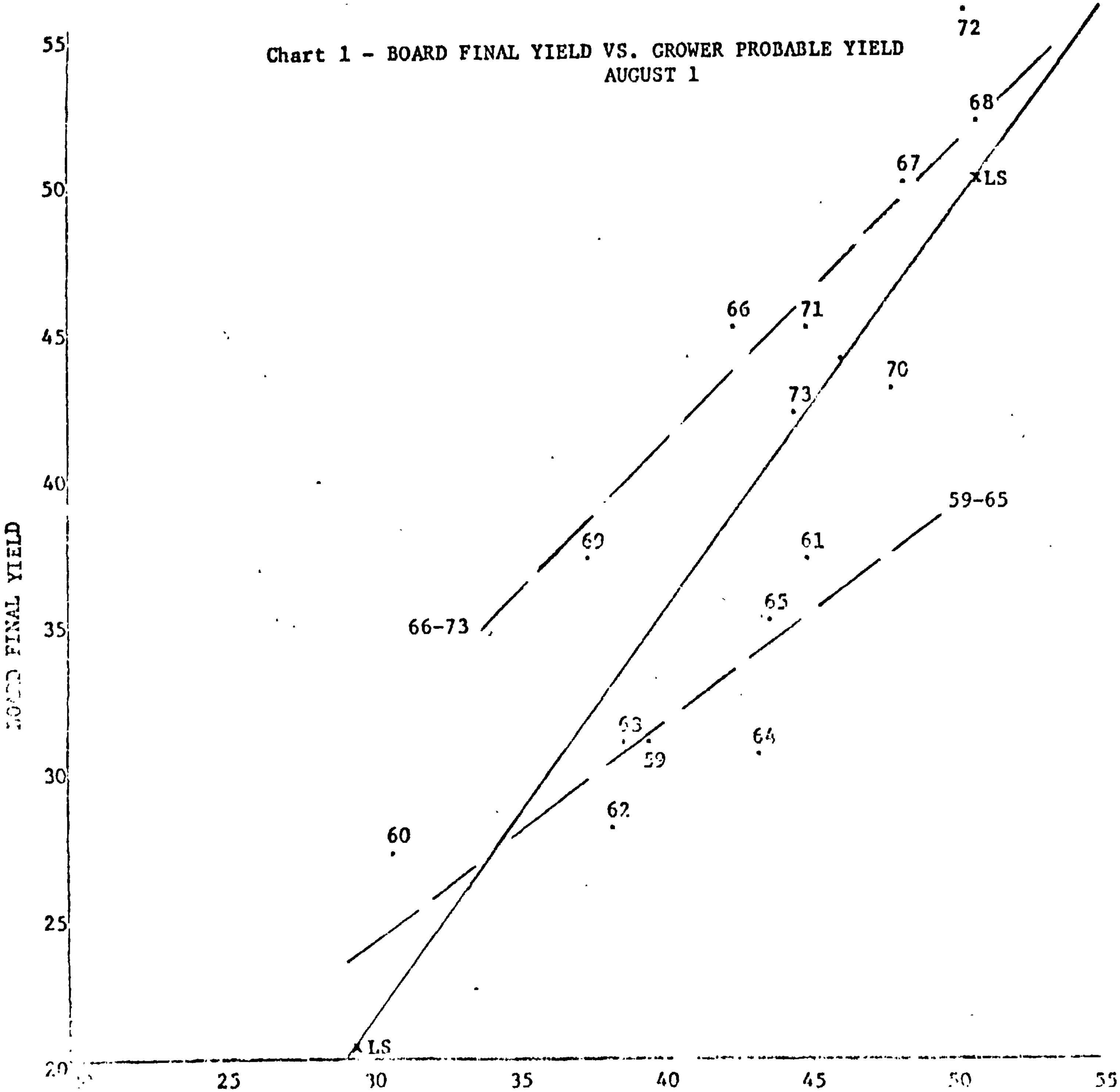
During the past three years, the major emphasis has been given to developing yield models in which the parameters are derived from the current year for use prior to harvest. These models are referred to as "within year models" and are considered more desirable than between year models if each year is different than the preceding years. These models do not require a historical series of 3-5 years of similar information before yield forecasts can be made. This last consideration has been considered quite important when starting work on a new crop or developing a system for a country without a crop forecasting system. This type of model has been considered for yield forecasts based on both grower subjective yield forecasts and objective yield methods. In the future, there will be opportunities to introduce new grower yield forecasts models based on probability samples as well as consideration of new crops of considerable economic importance for inclusion in objective yield programs.

It is helpful to start with a look at grower yield appraisals (or probable yields) which are used for most crops. The impetus for this effort came from the recognition that relatively few crops were in the objective yield program and the technical assistance work for foreign governments wishing to start current statistical programs.

Part I - Grower Subjective Yield Forecasts - H. F. Huddleston

The most common approach used by SRS is the charting of grower probable yields against the Board Final Yields. This approach is based on the relationships over years being the same for a period of 5-10 years and is normally considered usable after yields have been collected for 3-5 years. In most cases, these charts are based on voluntary reports returned by mail. Consequently, the reported probable yields may not be representative and/or the grower may not be able to forecast his crop accurately. In either case, the probable yields require adjustment or correction for various kinds of biases. Frequently, there appears to be different relationships indicated for different periods of years. The dashed lines in Chart 1 indicate approximately the nature of two different regressions and the solid line the least squares regression line over both periods. This chart illustrates a common problem associated with between year regression lines. Neither the representativeness of the sample nor the

Chart 1 - BOARD FINAL YIELD VS. GROWER PROBABLE YIELD
AUGUST 1



ability of the growers to forecast their yields are measured or known. Consequently, a somewhat different approach is needed in order to overcome these shortcomings.

Several new approaches will be discussed which should provide answers to some of these problems, but obviously require evaluation as to their utility for SRS. The first method is referred to as the Grower Graded Yield Appraisal. The method seeks to determine the following: (1) What does the grower expect the yield for a specific planting of a crop will be? (2) How does the grower rate (or evaluate) the expected yield of this planting of the crop based on five descriptive categories? The acreages (or areas) planted are then summarized by the five categories and the weighed average expected yield (or expected production) is derived based on the acreages reported.

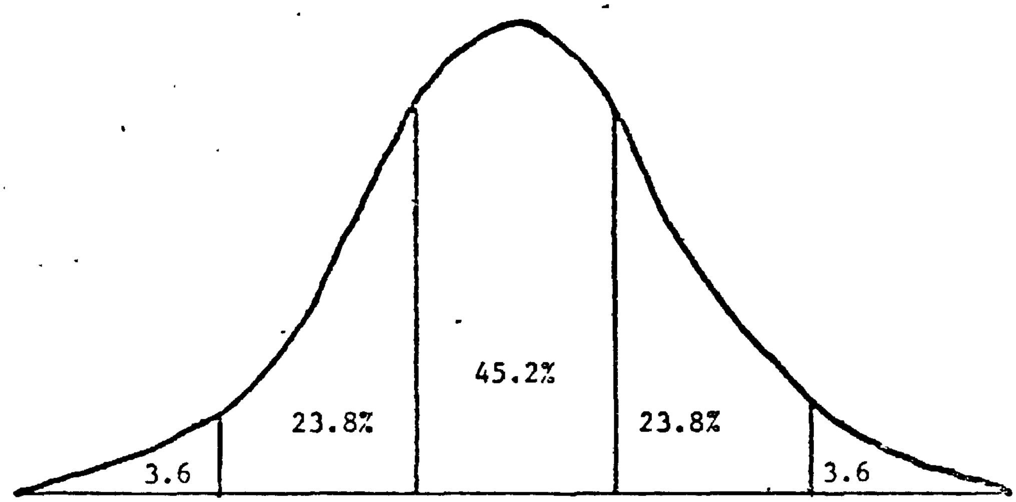
The descriptive ratings provided by the growers are assumed to be distributed normally according to the grading system suggested by some teachers whan a large number of students are to be graded. Thus the name, Grower Graded Yield Appraisal is given to the method since the grower "grades" his own yield appraisal. This grading scheme and its relation to the normal distribution is illustrated by Chart 2.

Experience with this approach in the Dominican Republic indicates that the growers do grade their yields in approximately this manner. That is, 40 to 50 percent of the acreage is reported by growers to have an expected yield which is "average" early in the crop season. The remaining expected yields are either one category above or below the average. This result suggests most growers merely report an average yield early in the crop season. The interpretation of the expected yield as being related to a harvested yield may be in serious error in any year that is not average or normal. Stated another way, most growers may either not be skillful forecasters of crop yields or do not wish to forecast a yield different from their average for purposes of reporting to public agencies. It may be that the most useful information comes from those growers who report a yield which is not average.

The procedure for reporting yield prospects to users for the coming harvest is as follows: (1) report the actual acreage percentages reported by the growers for the grade categories used, (2) report the average expected yield, and (3) derive from the model a within year average yield for the current year based on (1) and (2). The rationale behind this approach is to provide the growers expected yield, the descriptive appraisals, and the derived within year average yield so the data user may agree or disagree with this information as they see fit. Expected production can also be reported to the user in place of yield if this is considered preferable or if probable production was reported. If the within year derived average yield differs from the grower's last years average yield (or a five year average), the user is aware of this difference and may wish to place a somewhat different interpretation or evaluation on

Chart 2

Grower Graded Yield Appraisal Curve for a Large Number of Fields



Grade Scale	F	D	C	B	A
Possible crop descriptions corresponding to grades	Crop failure	Below average	Average	Above average	Much above average
	Very poor crop	Poor crop	Normal crop	Good crop	Very good crop
	No Harvest				Outstanding
					Excellent

Yield scale: 0 .2 .4 .6 .8 1.0

Midpoint of interval	.1	.3	.5	.7	.9

Where the range of the yield scale is 0 to 1.0 and each of the 5 grades covers one fifth of the X-axis.

$$E(X) = \sum_{1}^{5} p_i X_i = .50 \quad \text{(Average Yield)}$$

A scale from 0 to 2 could also be used in which case the E(X) would be 1.0.

crop prospects. For application to specific crops, the normal distribution may be skewed slightly by altering the tail probabilities and X scale values of the model. For example, in the Domincan Republic, coffee and rice are expected to have crop failures less frequently and outstanding crops more frequently than shown in chart 2 because of increased management inputs. Corn and beans are two crops which would be expected to have their distributions skewed in an opposite manner.

A second method is available which leads to essentially the same information. It can be referred to as the Growers Average Yield and Appraisal. For each planting of their crop early in the season, the grower is asked for the expected yield followed by a question to determine what the grower considers an average yield to be for the crop planted in the same field. The grower's expected yield (or production) and the average yield for the same acreage are reported for the data user's evaluation. The grower's within year average yield permits the user to judge whether this figure is consistent with the reported yield of the previous year or years.

An equally important phase of the yield information is to obtain similar information for the same growers after harvest. This second survey provides annual harvested acreage and crop production as well as a grower evaluation of the crop just harvested by five categories. That is the grower is asked to grade the harvested yield (or production) by the categories given. This information provides a basis for evaluating how good the growers are at forecasting their crop yields early in the season and whether they evaluate the harvested crop in a manner consistent with the model. Early in the season, there appears to be a tendency for the growers to be somewhat pessimistic and after harvest to have a rosier evaluation with regard to the past season.

Part II - Objective Yield Forecasting Methods - W. W. Wilson

Objective methods of forecasting crop yields have many things in common with subjective methods. Over-the-years regression models have been developed which are similar to the subjective charts. The models differ from the charts in two important respects:

1. The variables result from objective observations.
2. More than one data point is available for each year.

The availability of more than one point (usually many) for a year allows the development of within year regression models. For example (see exhibit 1), a scatter of data points in a single year may be fitted by a regression line which expresses the dependence of average mature corn ear weight on ear length. This fit of a model relates only to that specific year.

Now, similar models may be fitted to the data for a series of three years (see exhibit 2). A problem encountered in connection with regression analysis for data grouped by years is in selecting which model best fits the data. The questions which need to be answered are:

1. Should a different model be used for each year?
2. Should models based on a common slope but unique intercepts be used? (see exhibit 3)
3. Or, can the data be combined in a single regression model which ignors the grouping by years? (see exhibit 4)

Sequential statistical tests are available which provide answers to these questions.

However, in our application of forecasting yield a single model is required. The logical model to select (exhibit 4) ignors grouping of the data by years. If either of the alternative models have been selected based on the sequential tests, then we should not be surprised if problems in fitting this model to data from a fourth year occur. Even if the true relationship (see exhibit 5) in year 4 is not any more different than in years 1 and 3, substantial departures may occur. For an average ear length L, an overestimate of B-C occurs if line 4B represents the regression based on year 4 data. An underestimate of magnitude A-B results if 4A represents the actual relationship between the variables in the fourth year. If the fourth year is more unusual than any of the years used to develop the pooled model, greater departures may result. Of course, the regression for the fourth year is not known until the growing season is over. This limits our knowledge of the severity of the departure at the time the forecast is made.

Within year objective methods of forecasting crop yields are being investigated. These methods rely on data only from the current year. As such, they have the opportunity of reflecting unique characteristics of the year for which the forecast is desired.

Within year models depend on relating a response (what we want to forecast) to a variable with a known value at maturity. Various measures of time provide a suitable variable for this purpose.

A time measurement related to the beginning of growth has been found useful in growth models (see exhibit 6). This graph shows the model for average corn grain weight per plant as dependent on time. The measure of time used is days since silk emergence. Note the uniformity of weight in the harvest or mature period. The model provides an estimate of grain weight at any given time. The forecast is dependent on how well the model represents the actual situation and on our ability to know what value of time corresponds to maturity. In this case, the time value at maturity is any value in the flat region. A growth model for estimating average weight per grape has also been developed (see exhibit 7). The time variable used is not as closely related to the beginning of growth as was that variable in the corn model.

Within year models for survival of fruit, nuts, ears and so forth may also be developed. They complement the growth models. Together average per unit weight at harvest and number of units at harvest provide an indication of biological yield. This graph (see exhibit 8), shows the dependence of a survival ratio on days after a base estimate of plants with ears per acre. A base estimate of plants with ears per acre is made at day zero, so that the ratio for day zero is 1. The forecast survival ratio in the mature period can be multiplied times the base estimate to adjust it to number of units at harvest.

Research on both growth and survival models is continuing for corn and may be applicable to other crops. For example, survival models will be investigated to forecast the portion of papayas set each week surviving to harvest some 5 or 6 months later. Because previous year data is required for developing over-the-year models, within year methods may be most useful in developing and implementing objective yield forecast procedures for new crops.

EXHIBIT 1

AVERAGE MATURE EAR WEIGHT
VS
AVERAGE EAR LENGTH

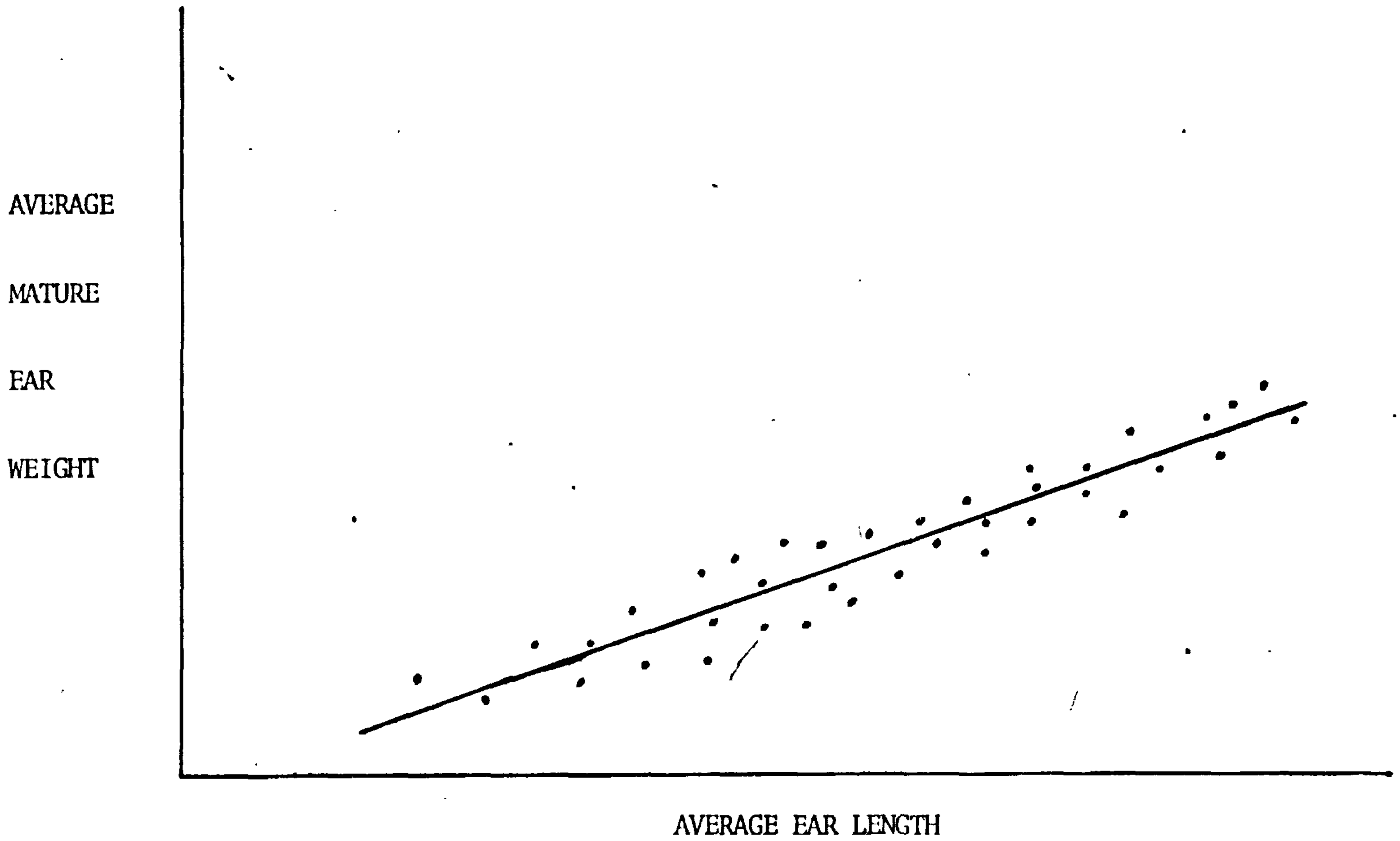


EXHIBIT 2

AVERAGE MATURE EAR WEIGHT
VS
AVERAGE EAR LENGTH

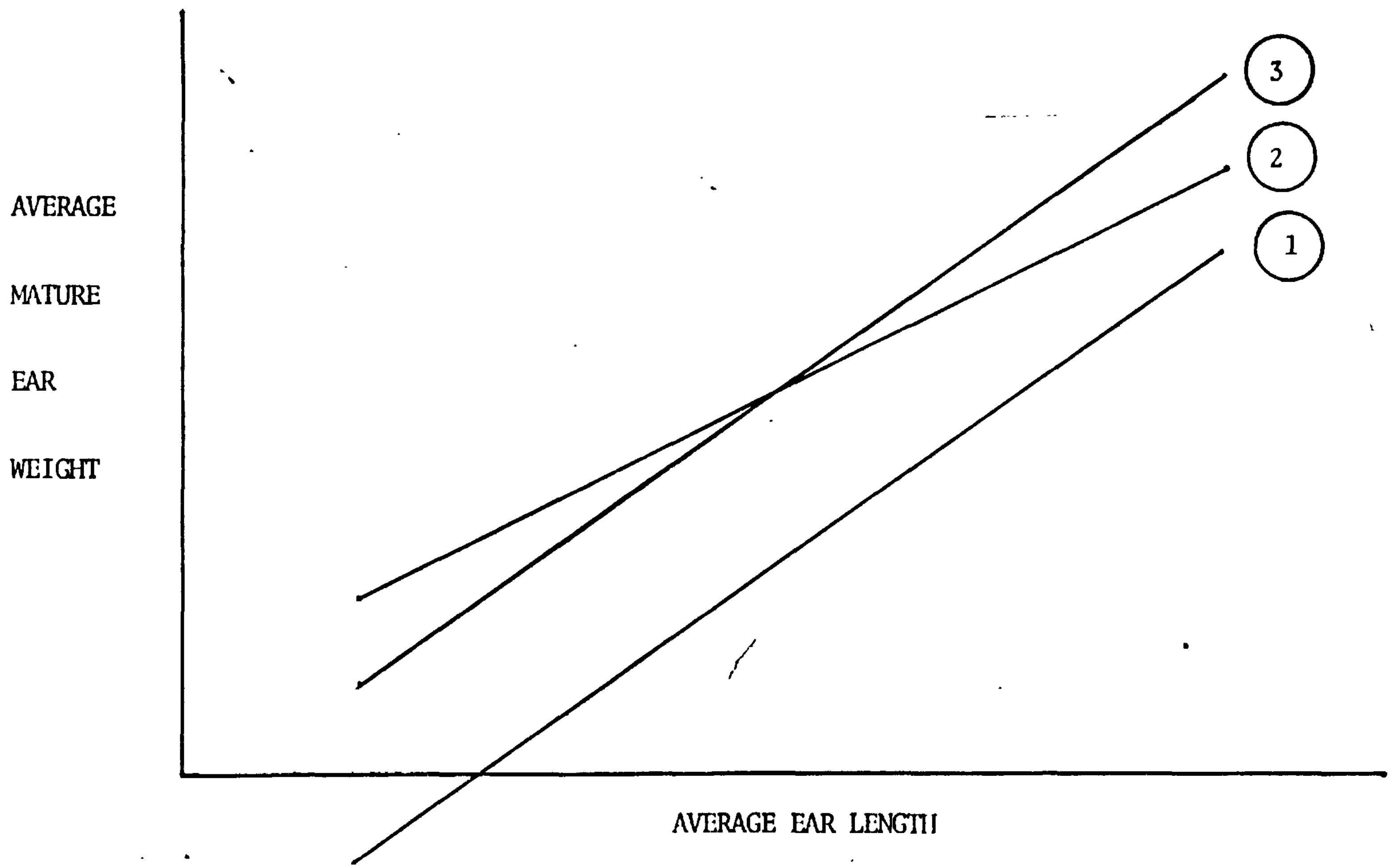


EXHIBIT 3

AVERAGE MATURE EAR WEIGHT
VS
AVERAGE EAR LENGTH

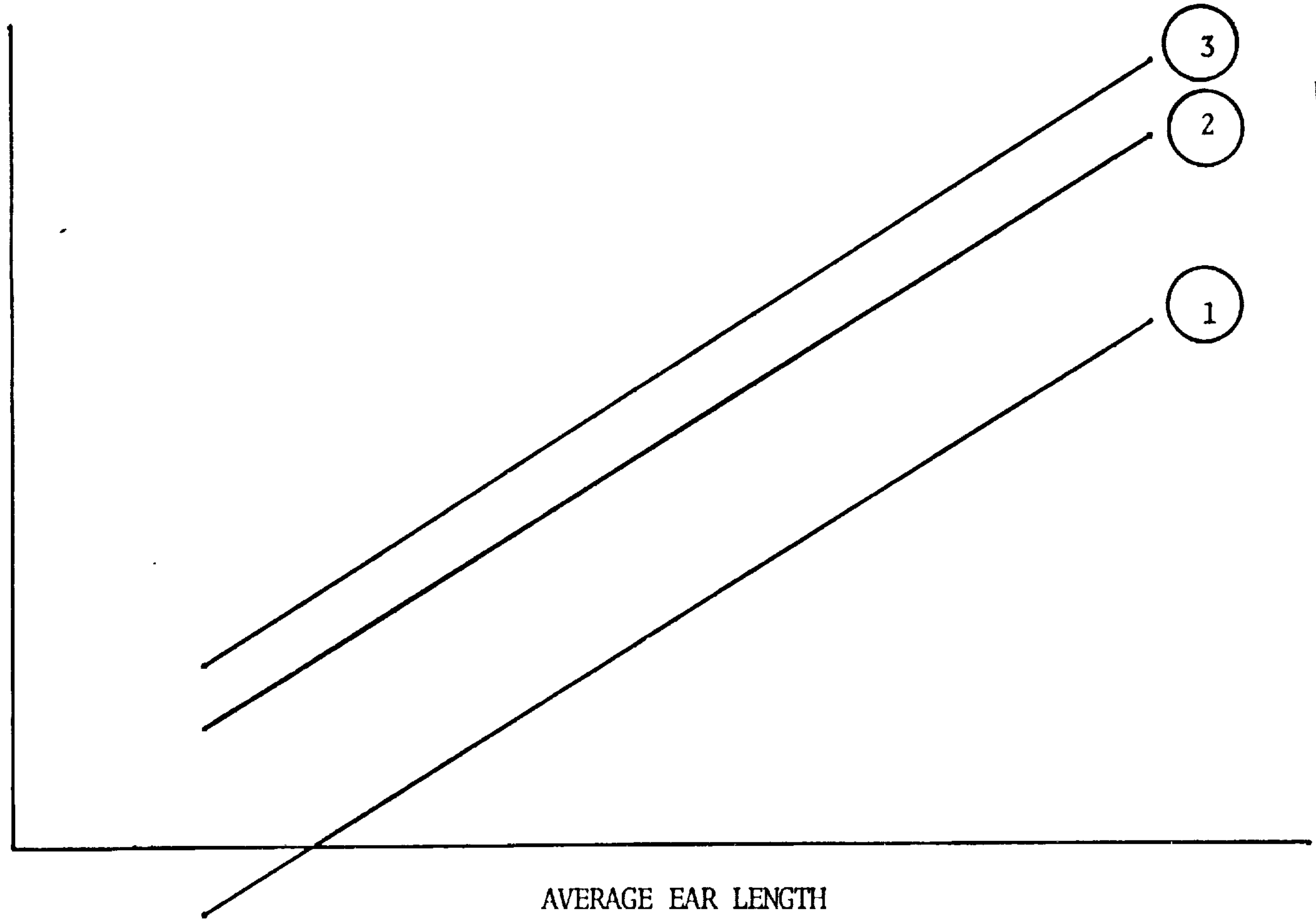


EXHIBIT 4

AVERAGE MATURE EAR WEIGHT
VS
AVERAGE EAR LENGTH

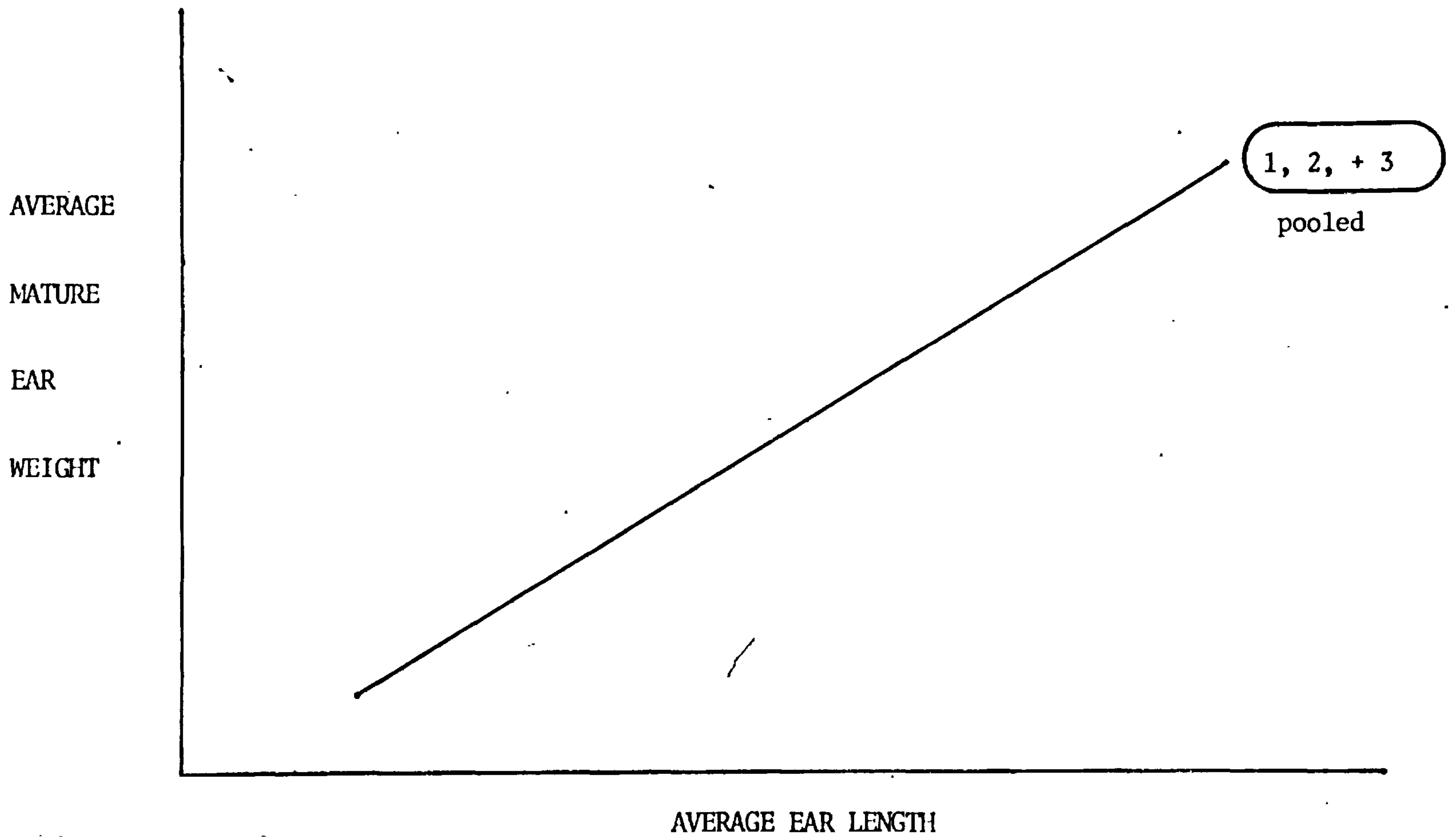


EXHIBIT 5

AVERAGE MATURE EAR WEIGHT
VS
AVERAGE EAR LENGTH