CORRELATIONS OF GRADES IN FRESHMAN ENGINEERING SUBJECTS AT THE UNIVERSITY OF ARIZONA

by

Margaret Nee Hillman

A Thesis

submitted to the faculty of the

Department of Mathematics

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in the Graduate College, University of Arizona

1949

Approved: R. F. Graesser, Director of Thesis

April 19, 1949. Date

CONCENTRATIONS OF CRACKS IN TURBULENT ACCUMULATING BOUNDARIES

AT THE UNIVERSITY OF ARIZONA

by

Lawrence M. Hillman

A Thesis

submitted to the faculty of the

Department of Information

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In the Graduate College, University of Arizona

1969

Approved: _____ _____
Director of Thesis Date

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The purpose of this thesis was to study the relationships among the grades in mathematics, chemistry, English, and engineering drawing, and the score in the psychological examination taken by freshmen entering the University of Arizona in September, 1947. The Psychological Examination, the 1946 College Edition, by the American Council on Education was given to each freshman upon entrance to the University. Then only the beginning students in the College of Mines and the College of Engineering during the fall semester of 1947-1948 were considered in this study.

To obtain the data, the writer examined the University of Arizona's student records in the Registrar's office. Beginning engineering and mining students taking certain courses were used. The four freshman courses considered were: mathematics, chemistry, English, and engineering drawing. The grades in these subjects and the score on the psychological test were correlated in this study.

The freshman mathematics courses were selected from these: Mathematics 22, Special College Algebra, 5 units; Mathematics 20, College Algebra, 3 units; Mathematics X, Review of High School Algebra, 3 subcollegiate units;

Mathematics 10, Solid Geometry, 3 units; and Mathematics 24, Plane Trigonometry, 2 units. If a student took more than one of these courses, the weighted average of his grades was used. The weights were the numbers of units in these courses.

The English grade was selected from English X, Subcollegiate English, 3 subcollegiate units; or English 1a, Freshman Composition, 3 units. The English placement examination given during registration week determined which of these courses the student took.

The chemistry course was either Chemistry 1a, Introductory General Chemistry, 4 units; or Chemistry 2a, General Chemistry, 4 units. The chemistry course the student took depended upon whether he had entrance credit in high school chemistry or not. Chemistry 1a is for students who have not had high school chemistry; while Chemistry 2a is for students who have entrance credit in high school chemistry.

Likewise the student took one of two engineering drawing courses, depending upon his high school background. Civil Engineering 3, Engineering Drawing, 3 units, is for students who have no entrance credit in mechanical drawing. While students with high school credit in mechanical drawing take Civil Engineering 31, Advanced Engineering Drawing, carrying 3 units.

The writer selected ninety freshmen who qualified for the correlation study in this thesis. The number of such students was limited because some students were exempt from

the first semester of college English after taking a special English examination which, if passed, entitles such students to credit for English 1a without taking it. This special English examination is given to those students who are in the top tenth of those currently taking the English placement examination. Still other students had withdrawn from one or more of the four subjects considered, thus disqualifying them as material for study. And again some freshmen did not follow the prescribed course completely. Therefore, the students considered numbered ninety.

In this correlation study the methodology of simple, partial, and multiple correlation is used. Therefore, all zero order correlations and correlation ratios were obtained for all possible pairs of subjects. There were ten such pairs of subjects possible. For each pair of subjects correlated, the equations of the regression lines were found and graphed. These regression lines were plotted on scatter-diagrams. Also the probable error of each zero order correlation was found. Following the computation of the zero order correlations, multiple and partial correlation coefficients were found, together with the regression equation of each subject on the remaining subjects.

In the second chapter, a summary of the theory of correlation is given with the definitions of the terms used. The third chapter contains an explanation of the computations and the procedure. Then the fourth chapter contains the

results and conclusions. Finally, a bibliography of references is given.

CHAPTER II

SUMMARY OF THE THEORY OF CORRELATION FOR SEVERAL VARIABLES

For the reader not familiar with the theory of correlation, we commence with the definitions of certain terms--along with an explanation of the theory of correlation.

To begin, we need to know the meaning of the arithmetic mean, since it is used in order to calculate the correlation coefficient r. The arithmetic mean of n numbers is simply the sum of the numbers divided by n.

Then we note the term standard deviation, denoted by σ. Standard deviation is defined thus: The standard deviation is the square root of the arithmetic mean of the squares of deviations of values of the variable from their arithmetic mean.

After defining the arithmetic mean and standard deviation, we need to know what correlation means. Correlation measures the relationship that exists among two or more quantities, that is, it measures the degree of dependence. According to Reitz:[1]

> The most important measure of the degree of correlation is the so-called Pearsonian

---

1. H.L. Reitz (ed.), Handbook of Mathematical Statistics (Chicago: Houghton Mifflin Company, 1924), p. 121.

> coefficient of correlation universally represented by the letter r. It is often called the product-moment coefficient.

Consider N pairs of variates

$$(x_i, y_i),\ i = 1, 2, 3, \ldots\ldots, N.$$

Let the value of y_i, when measured from the arithmetic mean with the standard deviation as a unit, be denoted by y_i^1, and in a similar manner define x_i^1. Then in terms of x_i^1 and y_i^1, the correlation coefficient is given by the formula,

$$r = \frac{1}{N} \sum_{i=1}^{i=N} x_i^1 y_i^1 .$$

This says that the

> correlation coefficient of two sets of variates, expressed with their respective standard deviations as units, may be defined as the arithmetic mean of the products of deviations of corresponding values from their respective arithmetic means.[1]

If each pair of variates is used for the rectangular coordinates of a point or dot and these points are plotted on a plane, the resulting diagram is called a scatter-diagram. If the means of the rows and the means of the columns are calculated from a correlation table such as that shown in Table I, then the scatter-diagram is also used to plot the means of the columns with small crosses and the means of the rows with small circles. These crosses for

---

1. H.L. Reitz, Mathematical Statistics (LaSalle, Illinois: The Open Court Publishing Company, 1927), p. 83.

column means and circles for row means fall upon or close to the regression lines according to the degree of correlation.

With respect to the regression line, Reitz says:

> If we mark the mean of each column of the 'scatter-diagram' by a cross, the 'line of regression of Y on X' is the straight line $Y = mX + b$ which fits 'best' this system of crosses.
>
> The line of regression of Y on X has the property that the sum of the squares of the distances (measured from it parallel to the Y-axis) of all the dots is less than from any other straight line. Moreover, the values of Y computed from the regression line are more highly correlated with the corresponding observed Y's than when calculated from any other linear function of X.
>
> For this reason, the equation of the line of regression of Y on X may be regarded as that linear relation which on the whole gives the 'best' estimate of Y corresponding to an assigned X in so far as such a prediction can be made by means of a linear function.[1]

The word array is used to designate either a row or a column of the correlation table.

The term linear regression indicates that the array means lie exactly on their respective regression lines, or do not deviate from these lines more than could be expected due to the fluctuation of sampling.

The standard error of estimate is the root-mean-square

---

1. H.L. Reitz (ed.), Handbook of Mathematical Statistics, p. 126.

error of estimate, designated by s_y, when y is estimated from x by use of the regression line of y on x. The symbol s_x has a similar meaning for x.

Once more we quote Reitz, who says:

> The mean square error in estimating the Y's, by taking the means of arrays of Y's, may be defined as the mean of the squares of the standard deviation of these arrays, the square of each standard deviation of an array being weighted with the number in the array. If the means of these arrays fall exactly on the line of regression of Y's on X's, then it can be proved that
>
> $$s_y^2 = \sigma_y^2(1-r^2),$$
>
> where s_y is the mean square error in the estimate of Y. The value of s_y^2 may also be defined as the mean of the squares of the deviations of the dots of the scatter-diagram from the line of regression of Y on X, when distances are measured parallel to the Y-axis.[1]

Since the probable error of the coefficient of correlation was calculated, a definition of the probable error is in order. The probable error is the value that any given error will as likely fall under as exceed.

If the regression of y on x is not linear, the correlation coefficient as a measure of correlation may be misleading. Hence, we need what is called the correlation ratio,

---

1. H.L. Reitz (ed.), Handbook of Mathematical Statistics, p. 128.

η_{yx}, to take the place of the correlation coefficient. The correlation ratio of y on x measures the clustering of the dots of the scatter-diagram about the means of arrays of the y's. If there is linear regression, $\eta_{yx}^2 = r^2$, approximately. The formula used to define the correlation ratio is

$$\eta_{yx}^2 = 1 - \frac{s_y'^2}{\sigma_y^2},$$

where $s_y'^2$ is the mean square vertical deviation of the dots from the means of the columns.

When we compare two subjects, we have simple or zero order correlation, but when the correlation is among three or more variables, our correlation is called multiple correlation. Reitz says:

> The basic problem of multiple correlation is to estimate the value of a variable that corresponds to assigned values of two or more other variables.[1]

According to Reitz:

> Assume that we may estimate X_1 from the assigned values of X_2, X_3, ..., X_n by means of the linear function.
>
> (1) $X_1 = b_{12.34...n} X_2 + b_{13.24...n} X_2 +$
>
> $\ldots + b_{1n...(n-1)} X_n + c$
>
> in which the b's and the c are constants, so

1. H.L. Reitz (ed.), Handbook of Mathematical Statistics, p. 139.

> chosen that the X_1's computed from (1) are the 'best' estimates of the observed X_1's which can be made by means of a linear function of the assigned values of X_2, X_3, ... X_n. The ordinary product-moment coefficient of correlation between the X_1 as thus estimated and the observed X_1 is called the 'multiple' correlation coefficient.[1]

Suppose we start with N sets of corresponding values of n variables $x_1, x_2, \ldots, x_n$. Assume that we separate the values of x_1 into classes by selecting class intervals $dx_2, dx_3, \ldots, dx_n$ of the (n-1) remaining variables. Now, when we limit the x_2's to an assigned interval dx_2, x_3's to an assigned interval dx_3, etc., the set of corresponding x's is called an array of x_1's.

We quote Reitz:

> The locus of the means of such arrays of x_1's in the theoretical distribution, as dx_2, $dx_3, \ldots dx_n$ approach zero, is called the regression surface of x_1 on the remaining variables.[2]

Referring back to equation (1) and omitting the subscripts after the decimal point, it can be shown that the linear regression surface of X_1 on the other subjects takes the form of

$$X_1 = b_{12}X_2 + b_{13}X_3 + \ldots + b_{1n}X_n + c,$$

where the X's are measured from their respective means, and the parameters $b_{12}, b_{13}, b_{14}, \ldots, b_{1n}$ can be expressed in

---

1. H.L. Reitz (ed.), Handbook of Mathematical Statistics, p. 139.
2. H.L. Reitz, Mathematical Statistics, p. 93.

terms of the standard deviations of the variables $X_1, X_2, \ldots, X_n$ and the simple correlation coefficients between these variables.

We adopt a least-squares criterion to determine the parameters $b_{12}, b_{13}, \ldots, b_{1n}$ so that

$$U = \sum(X_1 - b_{12}X_2 - b_{13}X_3 - \ldots - b_{1n}X_n - c)^2$$

shall be a minimum. Then this gives for the linear regression surface of X_1 on $X_2, X_3, \ldots, X_n$:

$$X_1 = -\sigma_1 \sum_{q=2}^{q=n} \frac{R_{1q}}{R_{11}} \frac{X_q}{\sigma_q},$$

where R_{pq} is the cofactor of the element in the pth row and qth column of the determinant:

$$R = \begin{vmatrix} 1 & r_{12} & r_{13} & \ldots\ldots & r_{1n} \\ r_{21} & 1 & r_{23} & \ldots\ldots & r_{2n} \\ r_{31} & r_{32} & 1 & \ldots\ldots & r_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{n1} & r_{n2} & r_{n3} & \ldots\ldots & 1 \end{vmatrix}.$$

By the cofactor of the element in the pth row and qth column, we mean $(-1)^{p+q}$ times the minor of that element.

We want the meaning of the standard error of estimate for n variables. This standard error $\sigma_{1.23\ldots n}$ of

estimating X_1 from assigned values of $X_2, X_3, \ldots, X_n$ is defined by

$$\sigma^2_{1.23\ldots n} = \frac{1}{N}\sum(\text{observed } X_1 - \text{computed } X_1)^2$$

By partial correlation we mean the degree of correlation between two subjects when the remaining subjects are held constant or have assigned values. For example, the partial correlation between mathematics and chemistry measures the degree of correlation between these two subjects, independent of the effect of the other subjects in the correlation study; that is, the influence of the other subjects is removed and only chemistry and mathematics are compared.

If $r_{12.345\ldots n}$ is the correlation between x_1 and x_2 after the effect of $x_3, x_4, \ldots, x_n$ has been eliminated, then $r_{12.34\ldots n}$ is defined by

$$r_{12.34\ldots n} = \frac{\sum x_{1.34\ldots n}\, x_{2.34\ldots n}}{N\sigma_{1.34\ldots n}\,\sigma_{2.34\ldots n}},$$

where the summation is extended to N pairs of residuals. That is, the partial correlation coefficient $r_{12.34\ldots n}$ is the simple coefficient of residuals

$$x_{1.34\ldots n} = x_1 - b_{13}x_3 - b_{14}x_4 - \ldots - b_{1n}x_n$$

and

$$x_{2.34\ldots n} = x_2 - b_{23}x_3 - b_{24}x_4 - \ldots - b_{2n}x_n.$$

With this we conclude our summary of the theory of correlation.

CHAPTER III

PROCEDURE

This chapter explains the procedure in more detail than was given in the Introduction. In our calculations, we follow the methods and arrangements given by Reitz.[1]

The first step toward the solution of our problem was to tabulate the grades of the ninety students in Table XIX. From this list of grades in mathematics, chemistry, English, and engineering drawing, and the scores on the psychological examination, double entry frequency tables or correlation tables, Tables I to X, were compiled. In these correlation tables, the subjects to be correlated were grouped in all possible pairs -- chemistry with mathematics, English with engineering drawing, etc. Since there were five subjects, there were ten possible pairs of subjects, and hence ten correlation tables.

Each correlation table contains several columns and several rows, each of which is a frequency distribution. In each correlation table, the totals of the frequencies in the rows were recorded in the column marked f_y, and the totals

---

1. H.L. Reitz (ed.), Handbook of Mathematical Statistics, Chapters VIII and IX.

of the frequencies in the columns were recorded in the row marked f_x. The column f_y and the row f_x each totaled ninety, the number of students in this study.

Arbitrary origins were chosen near the means, and the class intervals were used as units of measurement. Thus the column Y and the row X were set up. The next two columns and the next two rows in the correlation table were used to obtain the first and second moments about the arbitrary origins. These were used to calculate the deviations of the arithmetic means $\bar{X}$ and $\bar{Y}$ from the arbitrary origins, and to calculate the standard deviations σ_x and σ_y. The formula employed to calculate $\bar{Y}$ was

$$\bar{Y} = \frac{\sum f_y Y}{f_y},$$

and the standard deviation of y was found by the formula

$$\sigma_y^2 = \frac{\sum f_y Y^2}{f_y} - \bar{Y}^2.$$

The values for $\bar{X}$ and σ_x were found in a similar manner.

The entries in the column headed T were found by taking the sum of the products of each cell frequency in a row by the corresponding value of X. Likewise the entries in row S were calculated.

The entries in the column TY were found by multiplying the corresponding numbers in the column T and the column Y. The total of this column $(\sum TY)$ was the product moment

about the arbitrary origin, and was used in finding the correlation coefficient r by the formula,

$$r = \frac{\frac{1}{N}\sum TY - \overline{XY}}{\sigma_x \sigma_y} .$$

The column T^2 was computed in order to obtain $\sum \frac{T^2}{f_y}$ from which the correlation ratio η_{xy} was computed using

$$\eta_{xy}^2 = \frac{\frac{1}{N}\sum \frac{T^2}{f_y} - \overline{X}^2}{\sigma_x^2} = \frac{\sigma_{m_x}^2}{\sigma_x^2}$$

where $\sigma_{m_x}^2$ represents the numerator of the preceding fraction.

The column $\frac{T}{f_y}$ (or the column T times the class interval divided by f_y) was calculated in order to determine the row means. Similar calculating was done for the rows in the correlation table.

Next the root-mean-square error of estimate of y, or the standard error of y was found by using the formula,

$$s_y = \sigma_y \sqrt{1-r^2} ,$$

where $(1-r^2)^{\frac{1}{2}}$ is called the coefficient of alienation, and σ_y is the standard deviation of y.

The value of $\overline{Y}$ was multiplied by its class interval and added to its arbitrary origin to give the mean M_y of the Y's.

A similar procedure gave the mean M_x of the X's. The values of M_x and M_y entered into the regression equations:

$$Y - M_y = r \frac{\sigma_y}{\sigma_x} (X - M_x)$$

and

$$X - M_x = r \frac{\sigma_x}{\sigma_y} (Y - M_y).$$

These equations are termed, respectively, the regression of y on x, and x on y. The quantities $r \frac{\sigma_y}{\sigma_x}$ and $r \frac{\sigma_x}{\sigma_y}$ are called the regression coefficients. The preceding regression equations were changed to the form

$$Ax + By + C = 0$$

and plotted on the scatter-diagrams. The regression of y on x and the small crosses representing the corresponding column means were plotted on the scatter-diagrams, in each case, with red ink. While the regression of x on y and the small circles representing the corresponding row means were plotted, in each case, with green ink. Each pair of variates in a correlation table could have been plotted as a dot on the scatter-diagram, but this procedure was not followed.

We needed the probable error of the correlation coefficient. To obtain this, the formula

$$PE_r = 0.67449 \frac{1-r^2}{\sqrt{N}}$$

was used. The letter N represents the number of students in this study, and r is the correlation coefficient.

The correlation ratio, η_{xy}, was used with the correlation coefficient, r, to test for linearity of regression of x on y. For this purpose, the formula

$$N(\eta_{xy}^2 - r^2) < 11.37$$

was used, where N is the number of students in the study. (When the left hand member of the inequality is less than 11.37, there is linear regression.) Similarly, the linearity of regression of y on x was tested.

In order to obtain independent checks on all of the foregoing calculations, duplicate correlation tables, having different arbitrary origins, were used. All of the calculations, together with the formulas used, are set forth in Table I. In Tables II to X only the results are recorded.

When the variable x_2 is omitted, all pairs of variables increase and decrease together. Hence, their zero order correlations are positive as appear in the corresponding correlation tables. However, grades decrease numerically as proficiency in scholarship increases, while the opposite is true with respect to the psychological score, a numerically high score indicating proficiency. Hence, the true signs of the zero order correlations between the psychological score and the grades in the academic subjects are opposite to those appearing in the correlation tables.

Therefore, the corrected signs were used both in the determinant R, mentioned in the next paragraph, and also in Table XVIII.

Having completed the simple correlations as indicated above, the multiple and partial correlations among the five variables were found next. Toward this end, we first evaluated the determinant:

$$R = \begin{vmatrix} 1 & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & 1 & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & 1 & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & 1 & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & 1 \end{vmatrix},$$

where r_{pq} represents the zero order correlation between x_p and x_q, the value of r_{pp} being one.

After the evaluation of R, the multiple alienation coefficients, or k's, were found thus:

$$k_{1.2345} = \left(\frac{R}{R_{11}}\right)^{\frac{1}{2}}, \quad k_{2.1345} = \left(\frac{R}{R_{11}}\right)^{\frac{1}{2}}, \text{ etc.},$$

where the R_{pq} is the cofactor of the element in the pth row and qth column. The alienation coefficients were used to find the multiple correlation coefficients, such as $r_{1.23...n}$ by use of the relation

$$r_{1.23...n} = (1 - k_{1.23...n})^{\frac{1}{2}},$$

since

$$r^2_{1.23\ldots n} + k^2_{1.23\ldots n} = 1 .$$

And thus the multiple correlations were obtained from the simple correlations.

Next the regression of X_1 on the other four variables was calculated; viz.,

$$(1) \quad X_1 = \sum_{q=2}^{q=5} \beta_{1q} \frac{\sigma_1}{\sigma_q} X_q ,$$

where $\beta_{1q} = \frac{R_{1q}}{R_{11}}$. Similarly, the regression of each variable on the remaining ones was obtained.

The standard error of order four, $\sigma_{1.2345}$, made in estimating X_1 by the preceding equation is given by

$$\sigma^2_{1.2345} = \sigma_1^2 \frac{R}{R_{11}} .$$

This standard error was computed for each of the five variables.

Then the partial correlation, $r_{12.345}$, between X_1 and X_2, when the remaining variables are held fixed, is given by

$$r_{12.345} = \frac{- R_{12}}{(R_{11}R_{22})^{\frac{1}{2}}} .$$

This statistic was computed for each possible pair of variables.

Each calculation in multiple and partial correlation was repeated as a check, unless an independent check was possible. In the latter case the independent check was used.

The results are recorded in Tables XI to XVIII.

CHAPTER IV

RESULTS AND CONCLUSIONS

For the convenience of the reader, we repeat the meaning of the variables:

x_1 = grade in mathematics,

x_2 = score in psychological examination,

x_3 = grade in chemistry,

x_4 = grade in English composition,

x_5 = grade in engineering drawing.

Referring to Table XVII, it is observed that all zero order correlations are positive and each is, at least, six times its probable error except r_{25}, which is only about 1.2 times its probable error. Hence, all zero order correlations are significant except r_{25}; that is, with the exception of r_{25}, no one of them could have been obtained from a population in which the true correlation coefficient was zero. In the case r_{25}, the true correlation coefficient might have been zero so that r_{25} is not significant. This seems a reasonable result since the grade in engineering drawing would probably depend largely on manuel dexterity, and the psychological score on intellectual capacity.

An examination of Tables I to X shows that, for each pair of variables, both regressions are linear.

An interesting result to be observed in column (1) of Table XVIII is the high correlation of English composition with all other subjects since the r's with one subscript four come at the top of the column when the r's are arranged in order of magnitude. (One exception is r_{13}.) This result is also implied by the fact that $r_{4.1235}$ is the largest of the multiple correlations. It might be expected when we consider the importance of clear exposition in the writing of examinations and tests. That r_{13} is second in magnitude is perhaps accounted for by the fact that mathematics and chemistry are the only sciences in our study.

The r's with subscripts two are the smallest except r_{24}; also $r_{2.1345}$ is next to the smallest of the multiple correlations indicating relatively little relationship between the psychological score and proficiency in the other subjects except English. Perhaps ability to read and understand English improves the psychological score. Also higher correlation between the psychological score and chemistry, and mathematics might have been expected.

The relationships of English composition and of the psychological score with all other subjects, as shown in column (3) of Table XVIII, have just been mentioned. The lowest relationship of any subject with all others is that of engineering drawing. This seems reasonable again on account of the apparent greater dependence of engineering

drawing on manual skill.

Since all simple correlations are positive, we would expect all partial correlations to be smaller than the simple ones. This is seen to be true by comparing columns (1) and (2). The greatest reduction between simple and partial correlation is noted for the following pairs of subjects:

mathematics and English,

chemistry and engineering drawing,

chemistry and the psychological examination,

the psychological examination and engineering drawing.

This makes the net correlations between these pairs of subjects the least when the effects of all other subjects are removed. In fact, the net correlation between the last pair is actually negative, a result hardly to be anticipated, and perhaps rather difficult to explain. A low positive correlation would have seemed more reasonable.

The regression equations of each variable on the remaining four are presented in Table XV. The standard errors that would be made in using these regression equations for estimation purposes are shown in Table XVI. The regression equation is of use in two ways. First, by means of it, we may combine the variables in such a manner as to give the best possible estimate of one of the variables when the others have fixed values. The accuracy of this estimate depends upon the size of the standard error of order four.

Second, the regression equation gives the weights to be assigned to each variable in order that the values obtained from the variables shall be the best possible estimates of the variable desired. The multiple correlation coefficient measures the usefulness of the regression equation in this estimate.

To illustrate our first use of the regression equation, we shall select, at random from Table XIX, the values of x_2, x_3, x_4, and x_5 for one or two students, and substitute them in

$$(1) \quad X_1 = 0.00770\,X_2 + 0.50065\,X_3 + 0.05697\,X_4 + 0.22620\,X_5 - 0.35497.$$

The results give the estimated grade in mathematics, which can be compared with the actual grade in mathematics. This estimated grade should not differ from the actual grade by more than three standard errors. Using student No. 6 in Table XIX, equation (1) becomes

$$X_1 = 0.00770(99) + 0.50065(2) + 0.05697(3) + 0.22620(4) - 0.35497.$$

Therefore, $X_1 = 2.4843$, the estimated grade in mathematics. We note the actual grade in mathematics was 2.

We might consider another example. If we use student No. 21 (Table XIX), equation (1) becomes:

$$X_1 = 0.00770(167) + 0.50065(2) + 0.05697(2) + 0.22620(3) - 0.35497.$$

Therefore, $X_1 = 2.71477$, and we note the actual grade in mathematics was 1.6.

Since the standard error $\sigma_{1.2345}$ is 0.87371, we see the difference, in each example, between the estimated grade and the actual grade is less than three times the standard error. Similar examples could be worked out for other students and also used to estimate grades in other subjects.

To illustrate the second use of the regression equation, let us consider the relative weights to be given to

x_2 = psychological score,

x_3 = grade in chemistry,

x_4 = grade in English,

x_5 = grade in engineering drawing,

when we want to estimate x_1, the grade in mathematics. These weights are the coefficients in the regression equation. The most accurate way of evaluating the weight of each subject is to transform the following equation (see equation (1) of Chapter III):

$$x_1 = \beta_{12}\frac{\sigma_1}{\sigma_2}x_2 + \beta_{13}\frac{\sigma_1}{\sigma_3}x_3 + \beta_{13}\frac{\sigma_1}{\sigma_4}x_4 + \beta_{14}\frac{\sigma_1}{\sigma_5}x_5$$

into a form involving z's, where

$$z_i = \frac{x_i}{\sigma_i}, \quad (i = 1, 2, 3, 4, 5), \quad \text{thus:}$$

$$z_1 = \beta_{12}z_2 + \beta_{13}z_3 + \beta_{14}z_4 + \beta_{15}z_5, \quad \text{or}$$

$$z_1 = 0.13098\ z_2 + 0.52687\ z_3 + 0.04750\ z_4 + 0.17402\ z_5.$$

This is more accurate because the effect of the differences in variation of the several variables is eliminated.

By inspection of the preceding equation, we see the grade in chemistry influences most the grade in mathematics, since the coefficient of z_3 is larger than the coefficients of the other z's. Further inspection shows engineering drawing to be next in importance. The psychological score and English rank third and fourth, respectively. That is, the weight of chemistry has about three times the weight of engineering drawing, four times the weight of the psychological score, and almost 11.1 times the weight of English.

In a similar manner, the equations

$$z_2 = 0.17226\ z_1 + 0.017631\ z_3 + 0.58727\ z_4 - 0.28830\ z_5,$$
$$z_3 = 0.69273\ z_1 + 0.010263\ z_2 + 0.44218\ z_4 + 0.07236\ z_5,$$
$$z_4 = 0.03606\ z_1 + 0.33904\ z_2 + 0.43858\ z_3 + 0.23558\ z_5,$$
$$z_5 = 0.23844\ z_1 - 0.30043\ z_2 + 0.12954\ z_3 + 0.42552\ z_5$$

are obtained. Each variable in the right members is weighted by a coefficient which indicates its importance in estimating the variable in the left member of the equation.

With this we conclude our study of the correlation of freshman engineering grades of ninety freshmen entering the University of Arizona in September, 1947.

Table I

Mathematics

Chemistry

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row mean
5			2	5	5	12	3	36	108	27	81	729	60.75	2.25	4.25
4		3.5	1.5	7	4	16	2	32	64	27.5	55	756.25	47.26563	1.71875	3.71875
3	1	7	10	6	1	25	1	25	25	24	24	576	23.04	0.96	2.96
2	2	9	2	3		16	0	0	0	6	0	36	2.25	0.375	2.375
1	12	6	2	1		21	-1	-21	21	-8	8	64	3.04762	-0.38095	1.61905
F_x	15	25.5	17.5	22	10	90		72	218	76.5	168		136.35324		
X	-1	0	1	2	3										
$F_x X$	-15	0	17.5	44	30	76.5									
$F_x X^2$	15	0	17.5	88	90	210.5									
S	-11	8	17	34	24	72									
SX	11	0	17	68	72	168									
S^2	121	64	289	1156	576										
S^2/F_x	8.06667	2.50980	16.51429	52.54545	57.6	137.23621									
S/F_x	-.73333	.31373	.97143	1.54545	2.4										
Column Mean	1.26667	2.31373	2.97143	3.54545	4.4										

$\bar{Y} = \frac{\sum f_y Y}{f_y} = 0.8$ $\quad \bar{Y}^2 = 0.64$ $\quad \bar{X} = \frac{\sum F_x X}{F_x} = 0.85$

$\bar{X}^2 = 0.7225$ $\quad \sigma_y^2 = \frac{\sum f_y Y^2}{f_y} - \bar{Y}^2 = 1.78222$

$\sigma_y = 1.33500$ $\quad \sigma_x^2 = \frac{\sum F_x X^2}{F_x} - \bar{X}^2 = 1.61639$

$\sigma_x = 1.27137$ $\quad r = \frac{\frac{1}{N}\sum TY - \bar{X}\bar{Y}}{\sigma_x \sigma_y} = 0.69916$

$r^2 = 0.48882$ $\quad \sqrt{1-r^2} = 0.71497$ $\quad 1-r^2 = 0.51118$

$S_x = \sigma_x \sqrt{1-r^2} = 0.90899$ $\quad S_y = \sigma_y \sqrt{1-r^2} = 0.95448$

$M_x = A_x + \bar{X} = 2.85$ $\quad M_y = A_y + \bar{Y} = 2.8$

28

Table I Continued

$$Y - M_y = r\frac{\sigma_y}{\sigma_x}(X - M_x) \qquad 0.73415X - Y + 0.70768 = 0$$

$$X - M_x = r\frac{\sigma_x}{\sigma_y}(Y - M_y) \qquad X - 0.66584Y - 0.98566 = 0$$

$$P.E._r = 0.67449\,\frac{1-r^2}{\sqrt{N}} = 0.03634 \qquad \therefore r = 0.69916 \pm 0.03634$$

$$\sigma_{m_x}^2 = \frac{1}{N}\sum\frac{T^2}{f_y} - \bar{X}^2 = 0.79254 \qquad \eta_{xy} = \frac{\sigma_{m_x}^2}{\sigma_x^2} = 0.49031$$

$$\eta_{xy} = 0.70022 \qquad N(\eta_{xy}^2 - r^2) = 1.34272 < 11.37$$

$\therefore$ There is Linear Regression

$$\sigma_{m_y}^2 = \frac{1}{N}\sum\frac{S^2}{f_x} - \bar{Y}^2 = 0.88518 \qquad \eta_{yx}^2 = \frac{\sigma_{m_y}^2}{\sigma_y^2} = 0.44667$$

$$\eta_{yx} = 0.70475 \qquad N(\eta_{yx}^2 - r^2) = 0.70661 < 11.37$$

$\therefore$ There is Linear Regression

Diagram I

Means of Columns Marked +

Means of Rows Marked ○

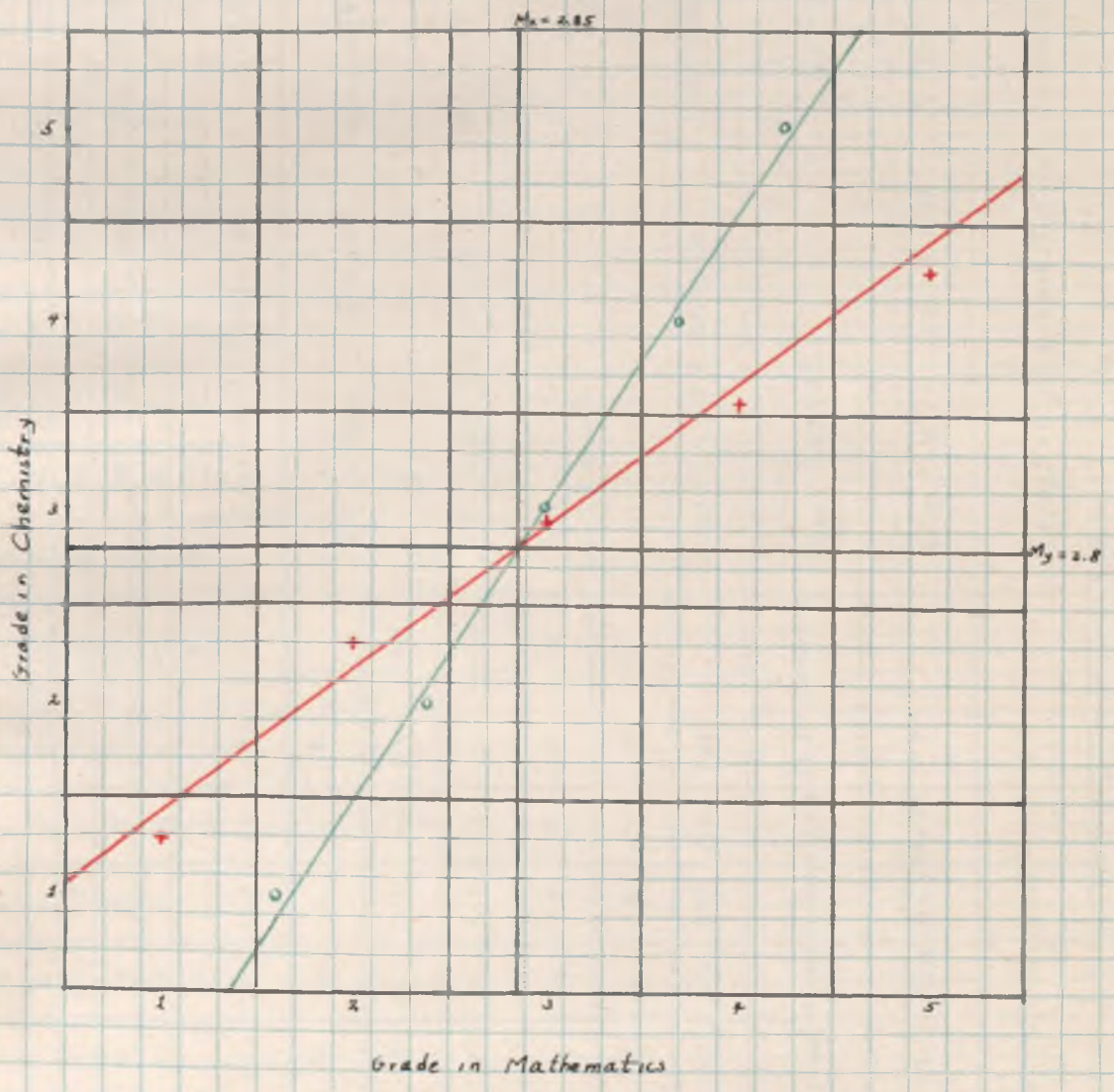

Table II

Mathematics

English

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row Mean
5		2	2	5	6	15	3	45	135	30	90	900	60	2	4
4	1	2.5	3.5	8	3	18	2	36	72	27.5	55	756.25	42.01389	1.527778	3.527778
3	6	9	11	7	1	34	1	34	34	22	22	484	14.23529	0.64706	2.64706
2	8	10	1	2		21	0	0	0	-3	0	9	0.42857	-0.14285	1.85714
1		2				2	-1	-2	2	0	0	0	0	0	2
f_x	15	25.5	17.5	22	10	90		113	243	74.5	167		116.67775		
X	-1	0	1	2	3										
$f_x X$	-15	0	17.5	44	30	76.5									
$f_x X^2$	15	0	17.5	88	90	210.5									
S	8	18	24	38	25	113									
SX	-8	0	24	76	75	167									
S^2	64	324	576	1444	625										
S^2/f_x	4.26667	12.70588	32.91429	65.63636	62.5	178.02320									
S/f_x	0.53333	2.70588	1.37143	1.72727	2.5										
Column Mean	2.53333	2.70588	3.37143	3.72727	4.5										

$\bar{Y} = 1.25556$ $\quad \bar{X} = 0.85$ $\quad \sigma_y = 1.05999$ $\quad \sigma_x = 1.27137$

$r = 0.58497 \pm 0.04677$ $\quad M_x = 2.85$ $\quad M_y = 3.25556$

$S_x = 1.03115$ $\quad S_y = 0.85971$

$0.48771\,X - Y + 1.86557 = 0$ $\quad X - 0.70163\,Y - 0.56581 = 0$

$\eta_{xy} = 0.59587$ $\quad \eta_{yx} = 0.59757$

$N(\eta_{xy}^2 - r^2) = 1.15837 < 11.37$ $\quad \therefore$ There is Linear Regression

$N(\eta_{yx}^2 - r^2) = 1.17259 < 11.37$ $\quad \therefore$ There is Linear Regression

Diagram II

Means of Columns Marked +

Means of Rows Marked ○

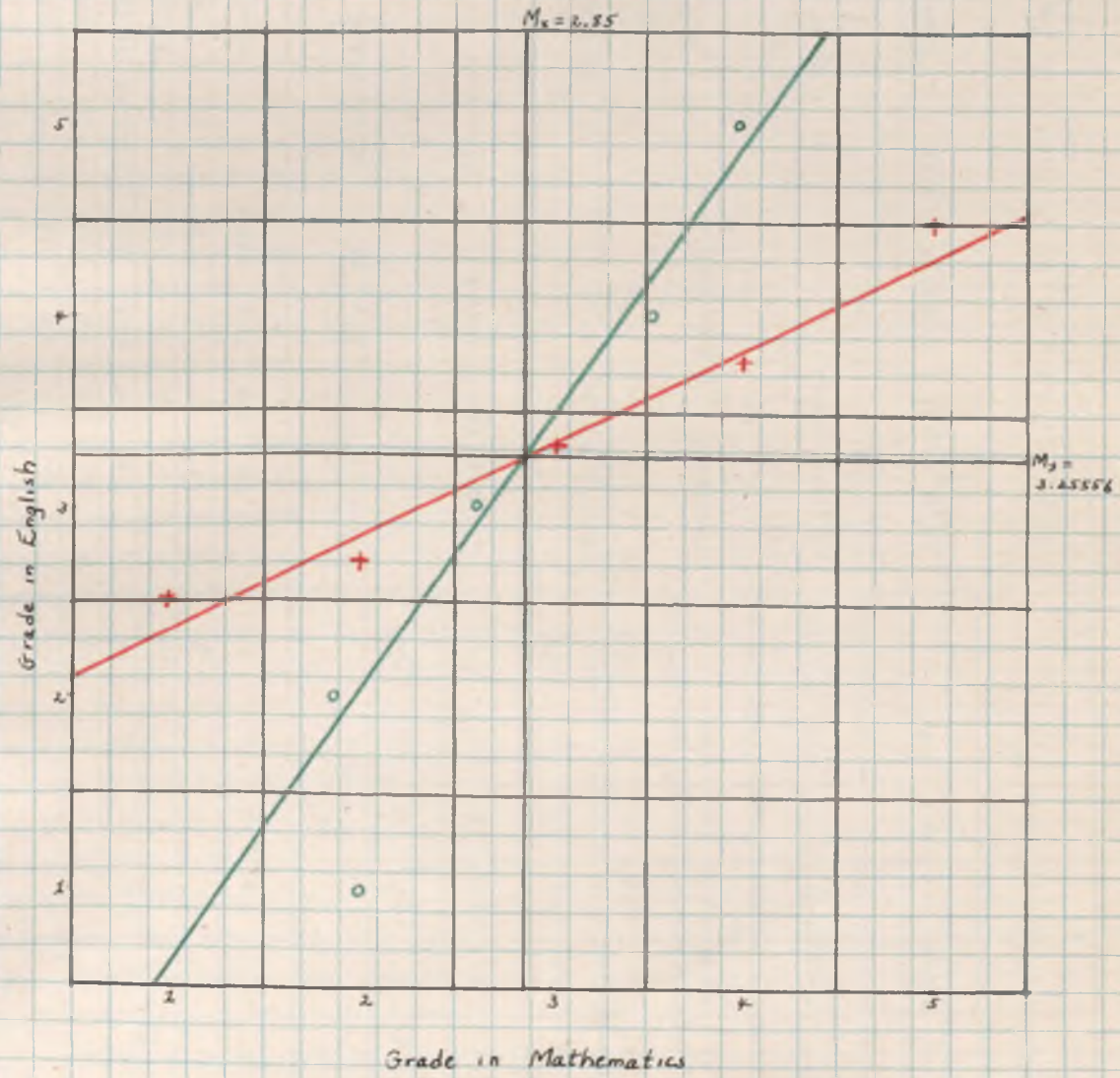

Table III

Mathematics

Engineering Drawing

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	T Y	T^2	T^2/f_y	T/f_y	Row Mean
5			2	4	4	10	3	30	90	22	66	484	48.4	2.2	4.2
4	1	5	3		4	13	2	26	52	14	28	196	15.07692	1.07692	3.07692
3	8	11.5	9.5	17	2	48	1	48	48	41.5	41.5	1722.25	35.88021	0.86458	2.86458
2	5	6	2	1		14	0	0	0	-1	0	1	0.07143	-0.07143	1.92857
1	1	3	1			5	-1	-5	5	0	0	0	0	0	2
f_x	15	25.5	17.5	22	10	90		99	195	76.5	135.5		99.42856		
X	-1	0	1	2	3										
$f_x X$	-15	0	17.5	44	30	76.5									
$f_x X^2$	15	0	17.5	88	90	210.5									
S	9	18.5	20.5	29	22	99									
SX	-9	0	20.5	58	66	135.5									
S^2	81	342.25	420.25	841	484										
S^2/f_x	5.4	13.42157	24.01429	38.2272	48.4	129.46313									
S/f_x	0.6	0.72549	1.17143	1.31818	2.2										
Column Mean	2.6	2.72549	3.17143	3.31818	4.2										

$\bar{Y} = 1.1$ $\quad \bar{X} = 0.85$ $\quad \sigma_y = 0.97809$ $\quad \sigma_x = 1.27137$

$r = 0.45882 \pm 0.06317$ $\quad M_x = 2.85$ $\quad M_y = 3.1$

$S_x = 1.12965$ $\quad S_y = 0.86906$

$0.35298\,X - Y + 2.09400 = 0$ $\quad X - 0.59640\,Y - 1.00116 = 0$

$M_{xy} = 0.48630$ $\quad M_{yx} = 0.48870$

$N(M_{xy}^2 - r^2) = 2.33657 < 11.37$ $\quad \therefore$ There is Linear Regression

$N(M_{yx}^2 - r^2) = 2.54790 < 11.37$ $\quad \therefore$ There is Linear Regression

Diagram III

Means of Columns Marked +

Means of Rows Marked ○

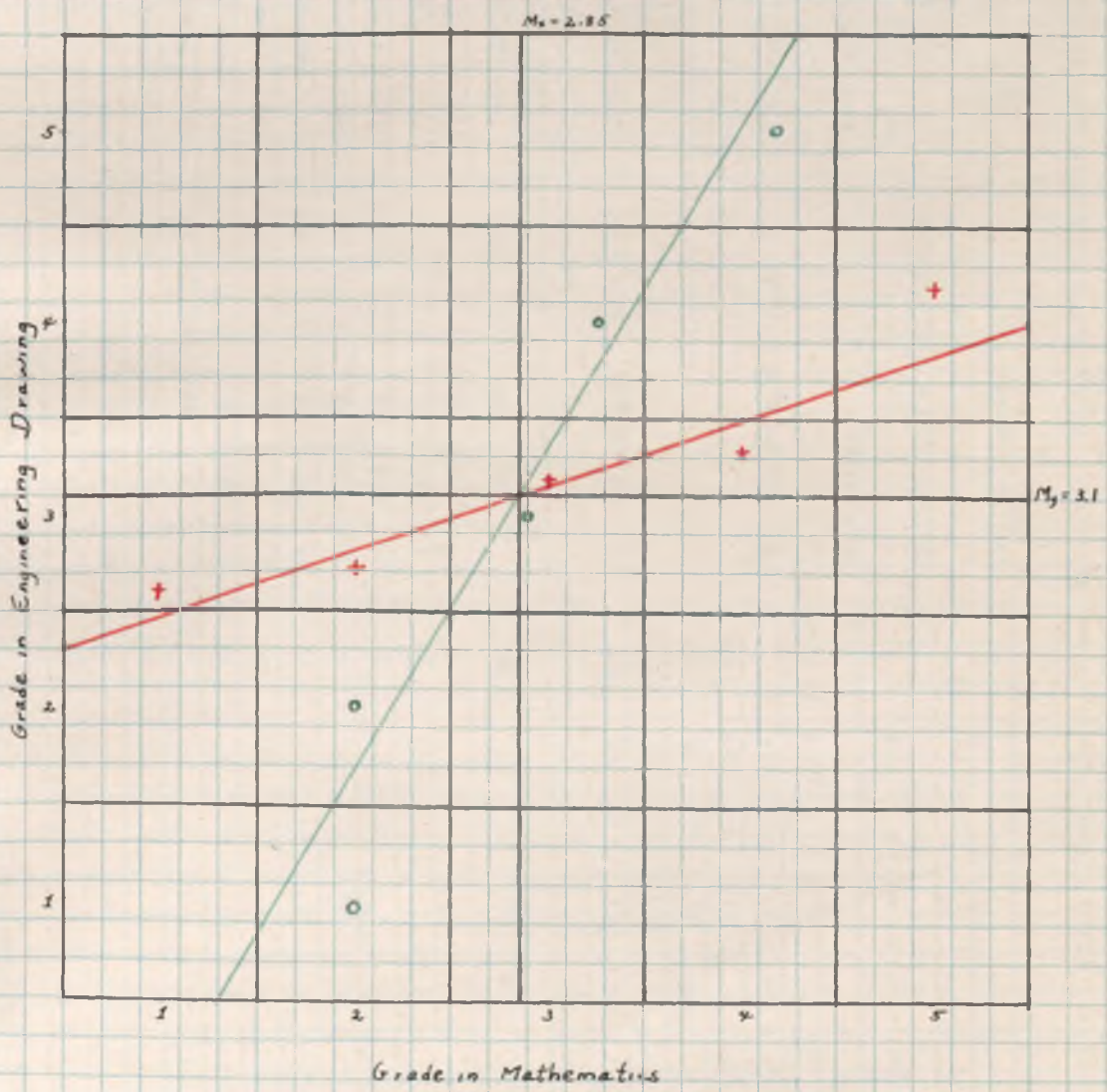

Table IV

Mathematics

Psychological Score

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row Mean
165		2.5	0.5			3	4	12	48	0.5	2	0.25	0.08333	0.16667	2.16667
154	1	4	1	1		7	3	21	63	2	6	4	0.57143	0.28571	2.28571
143	4	2				6	2	12	24	-4	-8	16	2.66667	-0.66667	1.3333
132	4	4	4	2	2	16	1	16	16	10	10	100	6.25	0.625	2.625
121	3	4	4	3	2	16	0	0	0	13	0	169	10.5625	0.8125	2.8125
110	2	2	5	8	1	18	-1	-18	18	22	-22	484	26.88889	1.22222	3.22222
99	1	4	2	5	2	14	-2	-28	56	17	-34	289	20.64286	1.21429	3.21429
88		2	1	3		6	-3	-18	54	7	-21	49	8.16667	1.16667	3.16667
77		1			2	3	-4	-12	48	6	-24	36	12	2	4
66					1	1	-5	-5	25	3	-15	9	9	3	5
f_x	15	25.5	17.5	22	10	90		-20	352	76.5	-106		96.83234		
X	-1	0	1	2	3										
$f_x X$	-15	0	17.5	44	30	76.5									
$f_x X^2$	15	0	17.5	88	90	210.5									
S	11	10	-3	-22	-16	-20									
SX	-11	0	-3	-44	-48	-106									
S^2	121	100	9	484	256										
S^2/f_x	8.06667	3.92157	0.51429	22	25.6	60.10252									
$11S/f_x$	8.06667	4.31373	-1.88571	-11	-17.6										
Column Mean	124.06667	125.31373	119.11429	110	103.4										

$\bar{Y} = -0.22222$ $\bar{X} = 0.85$ $\sigma_x = 1.27137$ $\sigma_y = 1.96513$ C.I

$\sigma_y = 21.61641$

$r = -0.39581 \pm 0.05996$ $M_x = 2.85$ $M_y = 118.55556$

$S_x = 1.16754$ $S_y = 19.85108$

$6.72968\ X + Y - 137.73514 = 0$ $X + 0.02328\ Y - 5.60991 = 0$

$M_{xy} = 0.46759$ $M_{yx} = 0.40018$

$N(M_{xy}^2 - r^2) = 5.57832 < 11.37$ ∴ There is Linear Regression $N(M_{yx}^2 - r^2) = 0.31304 < 11.37$ ∴ There is Linear Regression

Diagram IV

Means of Columns Marked +; Means of Rows Marked ○

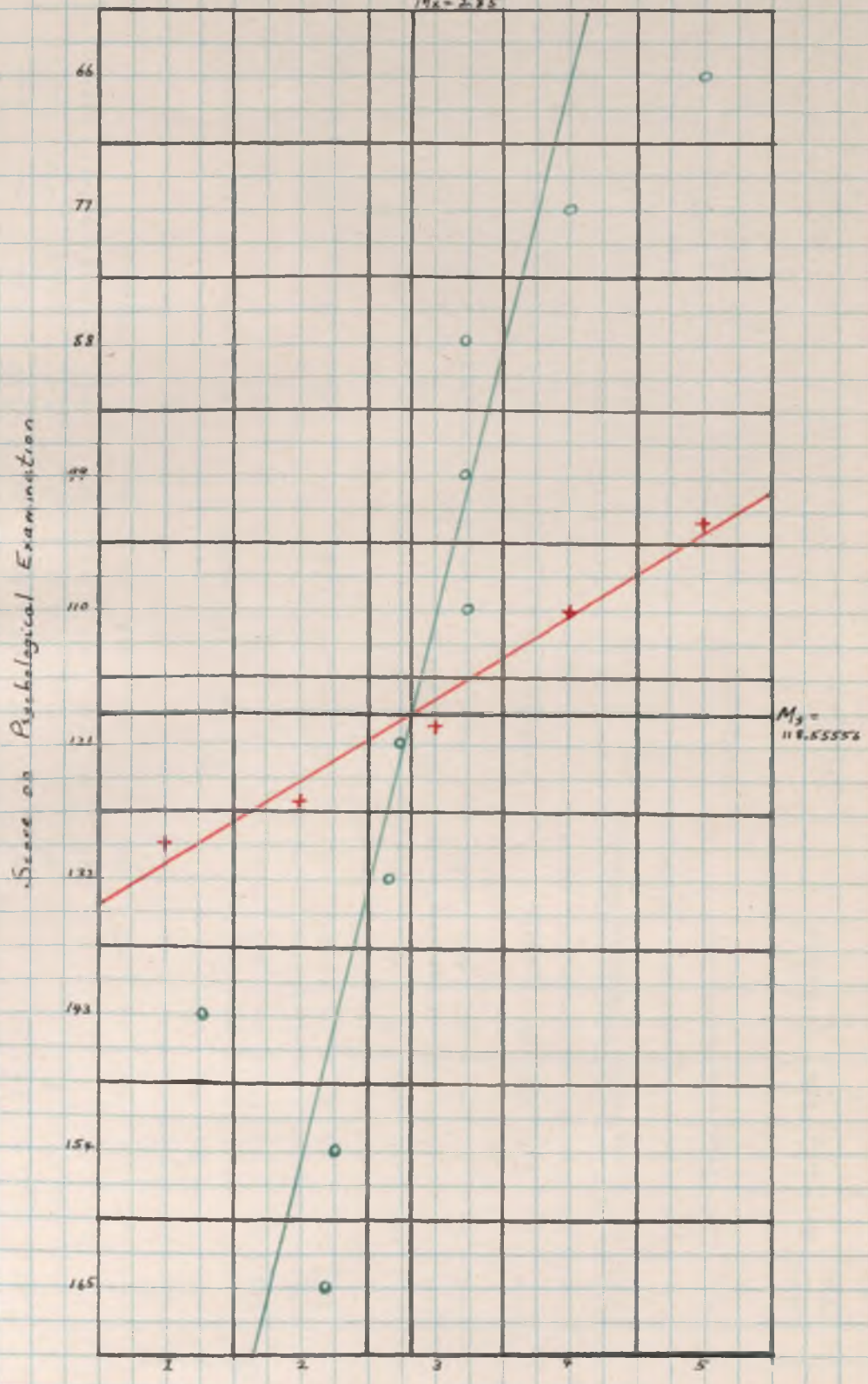

Table V

English

Chemistry

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row Mean
5			1	2	9	12	3	36	108	32	96	1024	85.33333	2.66667	4.66667
4	1		3	10	2	16	2	32	64	28	56	784	49.	1.75	3.75
3		2	14	5	4	25	1	25	25	36	36	1296	51.84	1.44	3.44
2		6	9	1		16	0	0	0	11	0	121	7.5625	0.6875	2.6875
1	1	13	7			21	-1	-21	21	6	-6	36	1.71429	0.28571	2.28571
f_x	2	21	34	18	15	90		72	218	113	182		195.45312		
X	-1	0	1	2	3										
$f_x X$	-2	0	34	36	45	113									
$f_x X^2$	2	0	34	72	135	243									
S	1	-11	16	31	35	72									
SX	-1	0	16	62	105	182									
S^2	1	121	256	961	1225										
S^2/f_x	0.5	5.76191	7.52941	53.38889	81.66667	148.84687									
S/f_x	0.5	-0.52381	0.47059	1.72222	2.33333										
Column Mean	2.5	1.47619	2.47059	3.72222	4.33333										

$\bar{Y} = 0.8$ $\bar{X} = 1.25556$ $\sigma_x = 1.05999$ $\sigma_y = 1.33500$

$r = 0.71923 \pm 0.03679$ $M_x = 3.25556$ $M_y = 2.8$

$S_x = 0.73645$ $S_y = 0.92752$

$0.90583X - Y - 0.14899 = 0$ $X - 0.57107Y - 1.65655 = 0$

$M_{xy} = 0.72786$ $M_{yx} = 0.75422$

$N(M_{xy}^2 - r^2) = 1.12325 < 11.37$ ∴ There is Linear Regression

$N(M_{yx}^2 - r^2) = 4.64161 < 11.37$ ∴ There is Linear Regression

Diagram V

Means of Columns Marked +

Means of Rows Marked o

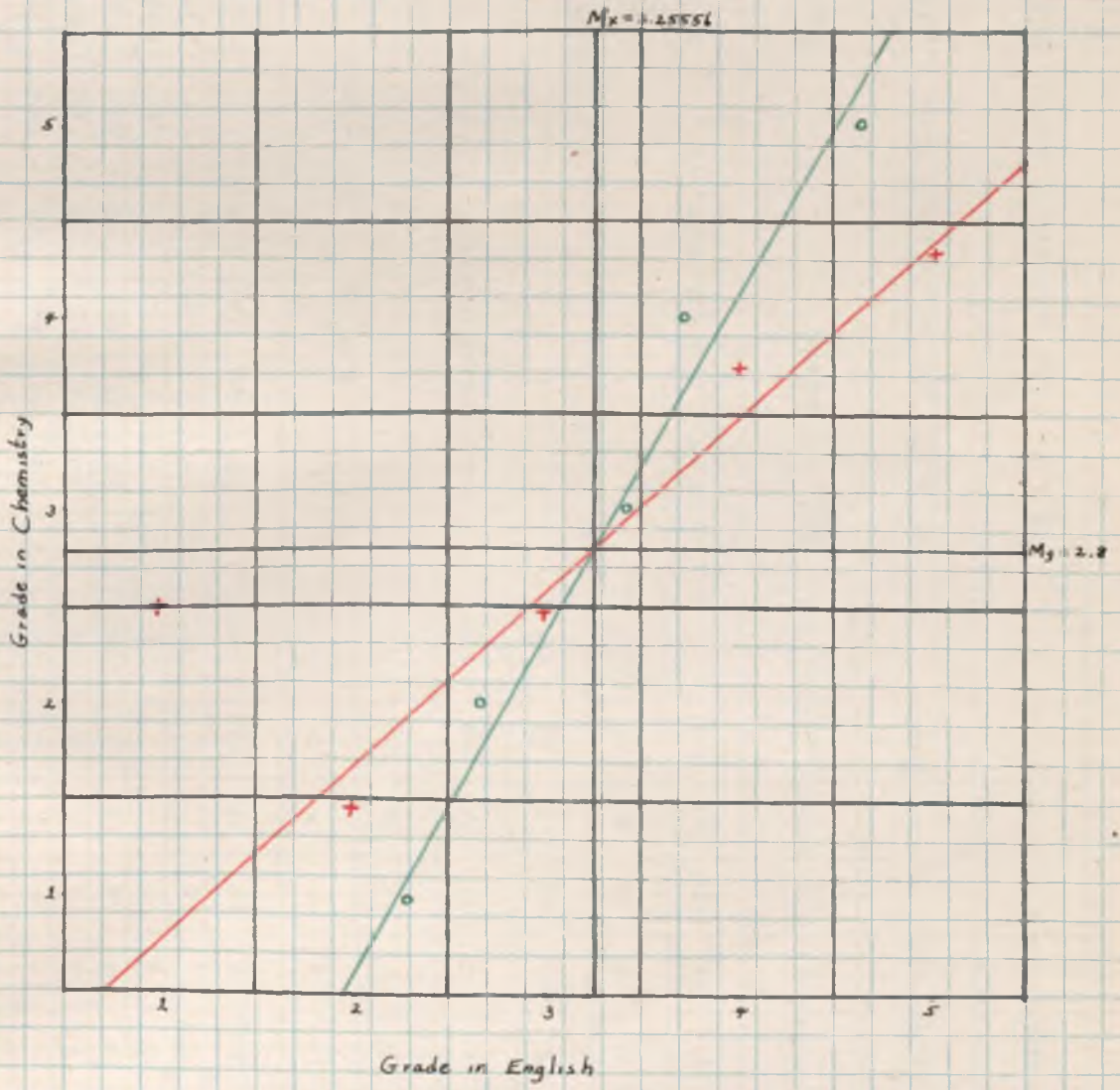

Table VII

Chemistry

Engineering Drawing

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row Mean
5		1	3	1	5	10	3	30	90	20	60	400	40	2	4
4	1	3	4	1	4	13	2	26	52	17	34	289	22.23077	1.30769	3.30769
3	10	9	12	14	3	48	1	48	48	39	39	1521	31.68750	0.81250	2.81250
2	7	2	5			14	0	0	0	-2	0	4	0.28571	-0.14286	1.85714
1	3	1	1			5	-1	-5	5	-2	2	4	0.8	-0.4	1.6
f_x	21	16	25	16	12	90		99	195	72	135		95.00398		
X	-1	0	1	2	3										
$f_x X$	-21	0	25	32	36	72									
$f_x X^2$	21	0	25	64	108	218									
S	9	17	28	19	26	99									
SX	-9	0	28	38	78	135									
S^2	81	289	784	361	676										
S^2/f_x	3.85714	18.0625	31.36	22.5625	56.33333	132.17548									
S/f_x	0.42857	1.0625	1.12	1.1875	2.16667										
Column Mean	2.42857	3.0625	3.12	3.1875	4.16667										

$\bar{Y} = 1.1$ $\bar{X} = 0.8$ $\sigma_{\bar{y}} = 0.97809$ $\sigma_{\bar{x}} = 1.33500$

$r = 0.47482 \pm 0.05507$ $M_x = 2.8$ $M_y = 3.1$

$S_x = 1.17491$ $S_y = 0.86080$

$0.34788\,X - Y + 2.12594 = 0$ $X - 0.64808\,Y - 0.79094 = 0$

$M_{xy} = 0.48290$ $M_{yx} = 0.51993$

$N(M_{xy}^2 - r^2) = 0.69626 < 11.37$ $\therefore$ There is Linear Regression

$N(M_{yx}^2 - r^2) = 4.03895 < 11.37$ $\therefore$ There is Linear Regression

37

Diagram VI

Means of Columns Marked +

Means of Rows Marked ○

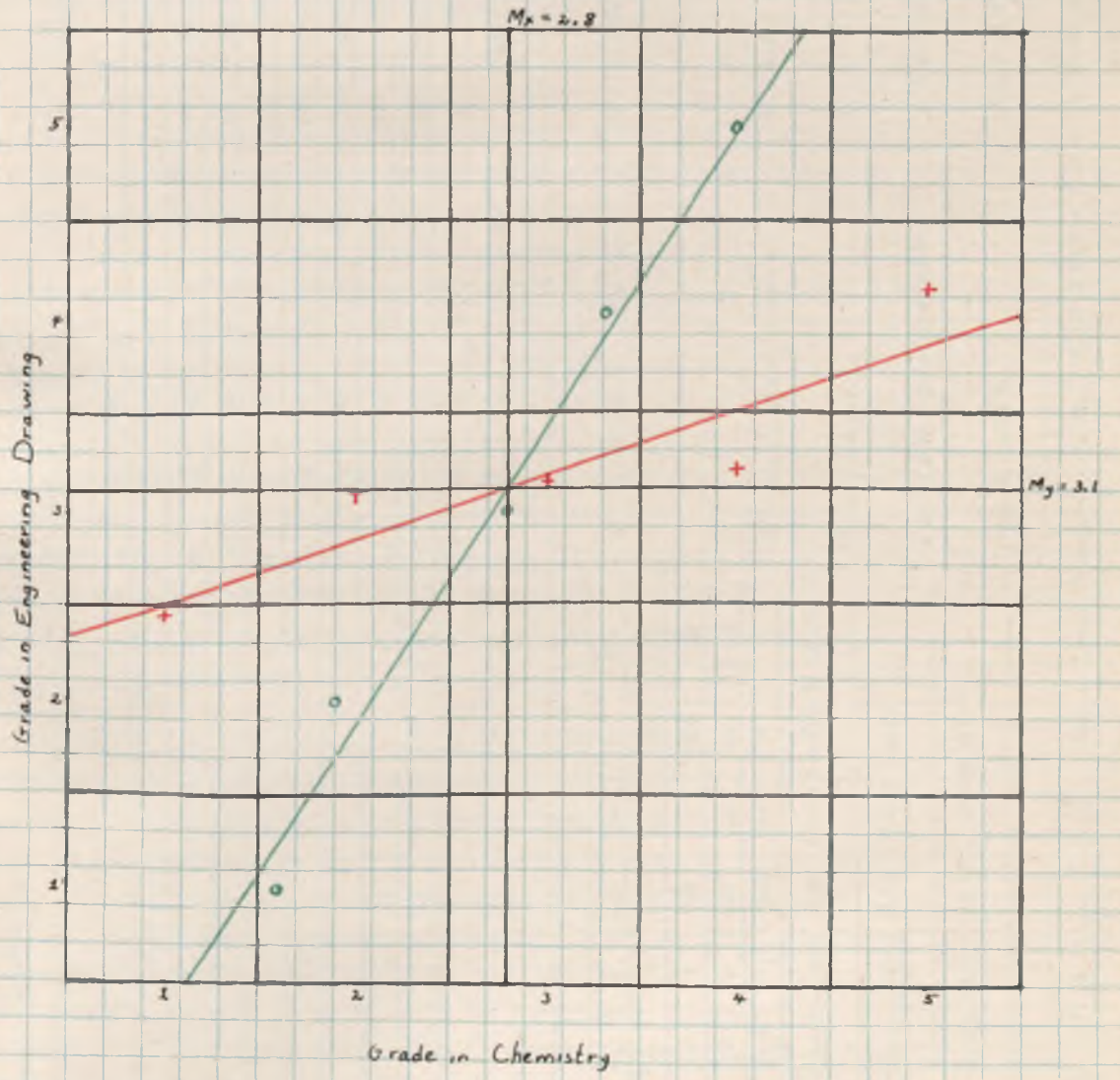

Table VII

Chemistry

Psychological Score

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row Mean
165		1	1	1		3	4	12	48	3	12	9	3	1	3
154	2	4	1			7	3	21	63	-1	-3	1	0.14286	-0.14285	1.85714
143	4	1		1		6	2	12	24	-2	-4	4	0.66667	-0.33333	1.66667
132	6	2	5	2	1	16	1	16	16	6	6	36	2.25	0.375	2.375
121	5	2	3	5	1	16	0	0	0	11	0	121	7.5625	0.6875	2.6875
110	3	5	5	1	4	18	-1	-18	18	16	-16	256	14.22222	0.88889	2.88889
99	1	1	6	4	2	14	-2	-28	56	19	-38	361	25.78571	1.35714	3.35714
88			3	2	1	6	-3	-18	54	10	-30	100	16.66667	1.66667	3.66667
77			1		2	3	-4	-12	48	7	-28	49	16.33333	2.33333	4.33333
66					1	1	-5	-5	25	2	-15	9	9	3	5
F_x	21	16	25	16	12	90		-20	352	72	-116		95.62996		
X	-1	0	1	2	3										
$F_x X$	-21	0	25	32	36	72									
$F_x X^2$	21	0	25	64	108	218									
S	15	13	-18	-7	-23	-20									
SX	-15	0	-18	-14	-69	-116									
S^2	225	169	324	49	529										
S^2/f_x	10.71429	10.5625	12.96	3.0625	44.08333	81.38262									
IIS/f_x	7.85714	8.9375	-7.92	-4.8125	-21.08333										
Column Mean	128.85714	129.9375	113.08	106.1875	79.91667										

$\bar{Y} = -0.22222$ $\bar{X} = 0.64$ $\sigma_x = 1.33500$ $\sigma_y = 1.96513$

$\sigma_y = 21.61641$

$r = -0.42353 \pm 0.05834$ $M_x = 2.8$ $M_y = 118.55556$

$S_x = 1.20935$ $S_y = 19.58190$

$6.85786X + Y - 1327.5755 = 0$ $X + 0.02614Y - 5.90102 = 0$

$M_{xy} = 0.48692$ $M_{yx} = 0.47050$

$N(M_{xy}^2 - r^2) = 5.19440 < 11.37$ $\therefore$ There is Linear Regression $N(M_{yx}^2 - r^2) = 3.77914 < 11.37$ $\therefore$ There is Linear Regression

Diagram VII

Means of Columns Marked + ; Means of Rows Marked o

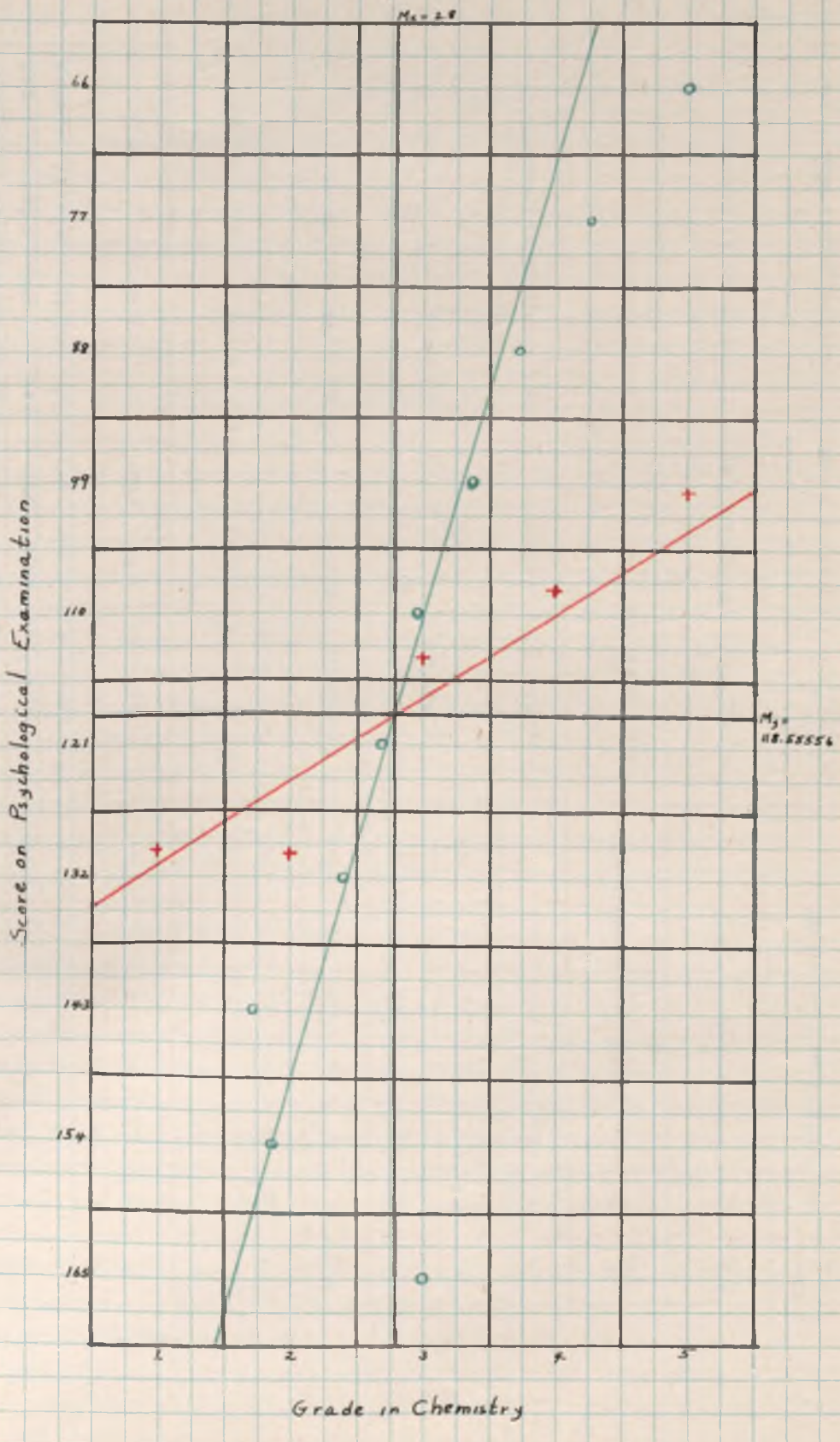

Table VIII

English

Engineering Drawing

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row Mean
5			2	2	6	10	3	30	90	24	72	576	57.6	2.4	4.4
4		2	6	1	4	13	2	26	52	20	40	400	30.76923	1.53846	3.53846
3	1	10	19	13	5	48	1	48	48	59	59	3481	72.52083	1.22917	3.22917
2		7	5	2		14	0	0	0	9	0	81	5.78571	0.64286	2.64286
1	1	2	2			5	-1	-5	5	1	-1	1	0.2	0.2	2.2
f_x	2	21	34	18	15	90		99	195	113	170		166.87578		
X	-1	0	1	2	3										
$f_x X$	-2	0	34	36	45	113									
$f_x X^2$	2	0	34	72	135	243									
S	0	12	35	21	31	99									
SX	0	0	35	42	93	170									
S^2	0	144	1225	441	961										
S^2/f_x	0	6.85714	36.02941	24.5	64.06667	131.45322									
S/f_x	0	0.57143	1.02941	1.16667	2.06667										
Column Mean	2	2.57143	3.02941	3.16667	4.06667										

$\bar{Y} = 1.1$ $\bar{X} = 1.25556$ $\sigma_y = 0.97809$ $\sigma_x = 1.05999$

$r = 0.48977 \pm 0.05404$ $M_x = 3.25556$ $M_y = 3.1$

$S_x = 0.92416$ $S_y = 0.85275$

$0.45193X - Y + 1.62872 = 0$ $X - 0.53078Y - 1.61014 = 0$

$m_{xy} = 0.497197$ $m_{yx} = 0.51180$

$N(m_{xy}^2 - r^2) = 0.65979 < 11.37$ ∴ There is Linear Regression

$N(m_{yx}^2 - r^2) = 1.98617 < 11.37$ ∴ There is Linear Regression

Diagram VIII

Means of Columns Marked +

Means of Rows Marked o

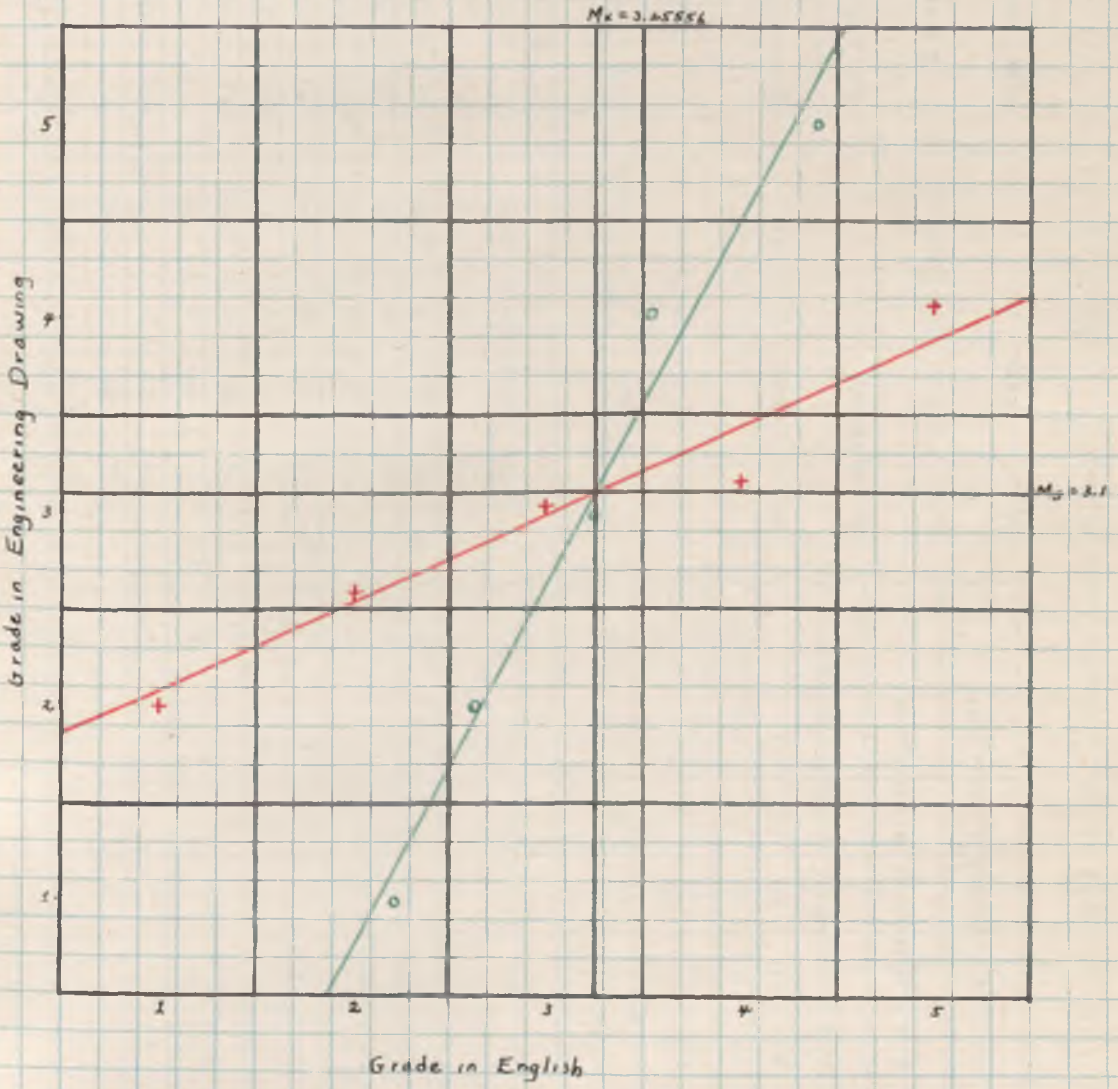

Table IX

English

Psychological Score

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row mean
165		1	1	1		3	4	12	48	3	12	9	3	1	3
154	1	2	4			7	3	21	63	3	9	9	1.28571	0.42857	2.42857
143	1	4	1			6	2	12	24	0	0	0	0	0	2
132		5	9	2		16	1	16	16	13	13	169	10.5625	0.8125	2.8125
121		3	8	4	1	16	0	0	0	19	0	361	22.5625	1.1875	3.1875
110		5	5	3	5	18	-1	-18	18	26	-26	676	37.55556	1.444444	3.444444
99		1	5	6	2	14	-2	-28	56	23	-46	529	37.78571	1.64286	3.64286
88			1	1	4	6	-3	-18	54	15	-45	225	37.5	2.5	4.5
77				1	2	3	-4	-12	48	8	-32	64	21.33333	2.66667	4.66667
66					1	1	-5	-5	25	3	-15	9	9	3	5
f_x	2	21	34	18	15	90		-20	352	113	-130		180.58532		
X	-1	0	1	2	3										
$f_x X$	-2	0	34	36	45	113									
$f_x X^2$	2	0	34	72	135	243									
S	5	16	9	-16	-34	-20									
SX	-5	0	9	-32	-102	-130									
S^2	25	256	81	256	1156										
S^2/f_x	12.5	12.19048	2.38235	14.22222	77.06667	118.36172									
$\Pi S/f_x$	27.5	8.38095	2.9876	-7.77778	-29.93333										
Column Mean	148.5	139.38095	23.9 1176	111.22222	96.06667										

$\bar{Y} = -0.22222$ $\bar{X} = 1.25556$ $\sigma_x = 1.05999$ $\sigma_y = 1.96513$ C.I.

$\sigma_{\bar{y}} = 21.61641$

$r = -0.55949 \pm 0.04884$ $M_x = 3.25556$ $M_y = 118.55556$

$S_x = 0.87856$ $S_y = 17.91645$

$11.40973\,X + Y - 155.70056 = 0$ $X + 0.02744\,Y - 6.50818 = 0$

$M_{xy} = 0.618690$ $M_{yx} = 0.57251$

$N(M_{xy}^2 - r^2) = 6.27735 < 11.37$ ∴ There is Linear Regression $N(M_{yx}^2 - r^2) = 1.32621 < 11.37$ ∴ There is Linear Regression

Diagram IX

Means of Columns Marked + ; Means of Rows Marked ○

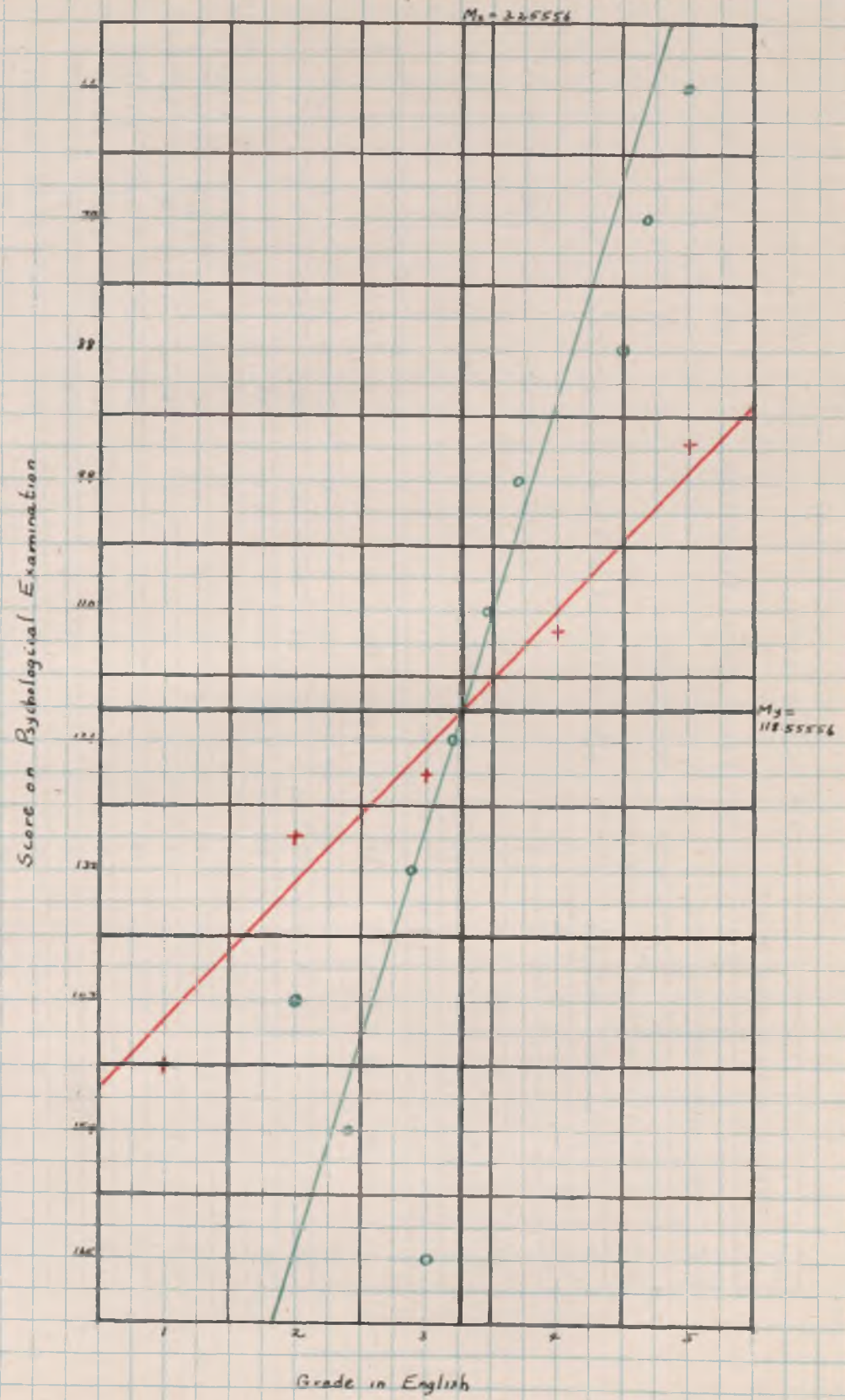

Table X

Engineering Drawing

Psychological Score

	1	2	3	4	5	f_y	Y	$f_y Y$	$f_y Y^2$	T	TY	T^2	T^2/f_y	T/f_y	Row Mean
165			2	1		3	4	12	48	4	16	16	5.33333	1.33333	3.33333
154	1		4	2		7	3	21	63	7	21	49	7	1	3
143		1	5			6	2	12	24	5	10	25	4.16667	0.83333	2.83333
132	3	1	5	4	3	16	1	16	16	19	19	361	22.56250	1.1875	3.1875
121		4	9	2	1	16	0	0	0	16	0	256	16	1	3
110		3	12	1	2	18	-1	-18	18	20	-20	400	22.22222	1.11111	3.11111
99		5	6	2	1	14	-2	-28	56	13	-26	169	12.07143	0.92857	2.92857
88	1		4		1	6	-3	-18	54	6	-18	36	6	1	3
77			1	1	1	3	-4	-12	48	6	-24	36	12	2	4
66					1	1	-5	-5	25	3	-15	9	9	3	5
f_x	5	14	48	13	10	90		-20	352	99	-37		116.35615		

X	-1	0	1	2	3	
$f_x X$	-5	0	48	26	30	99
$f_x X^2$	5	0	48	52	90	195
S	3	-10	-5	5	-13	-20
SX	-3	0	-5	10	-39	-37
S^2	9	100	25	25	169	
S^2/f_x	1.8	7.14286	0.52083	1.92308	16.9	28.28677
SS/f_x	6.6	-7.85714	-1.14583	+1.23077	-14.3	
Column Mean	127.6	113.14286	109.35417	125.23077	106.7	

$\bar{Y} = -.22222$ $\bar{X} = 1.21$ $\sigma_{\bar{x}} = 0.97809$ $\sigma_y = 1.96513$ C.I

$\sigma_{\bar{y}} = 2.1.61691$

$r = -0.08671 \pm 0.07056$ $M_x = 3.1$ $M_y = 118.55556$

$S_x = 0.99441$ $S_y = 21.53499$

$1.91638X + Y - 12.49633 = 0$ $X + .08392\ Y - 3.56515 = 0$

$\eta_{xy} = 0.29428$ $\eta_{yx} = 0.26192$

$N(\eta_{xy}^2 - r^2) = 7.11717 < 11.37$ $\therefore$ There is Linear Regression $N(\eta_{yx}^2 - r^2) = 5.49730 < 11.37$ $\therefore$ Linear Regression

Diagram X

Means of Columns Marked + ; Means of Rows Marked ○

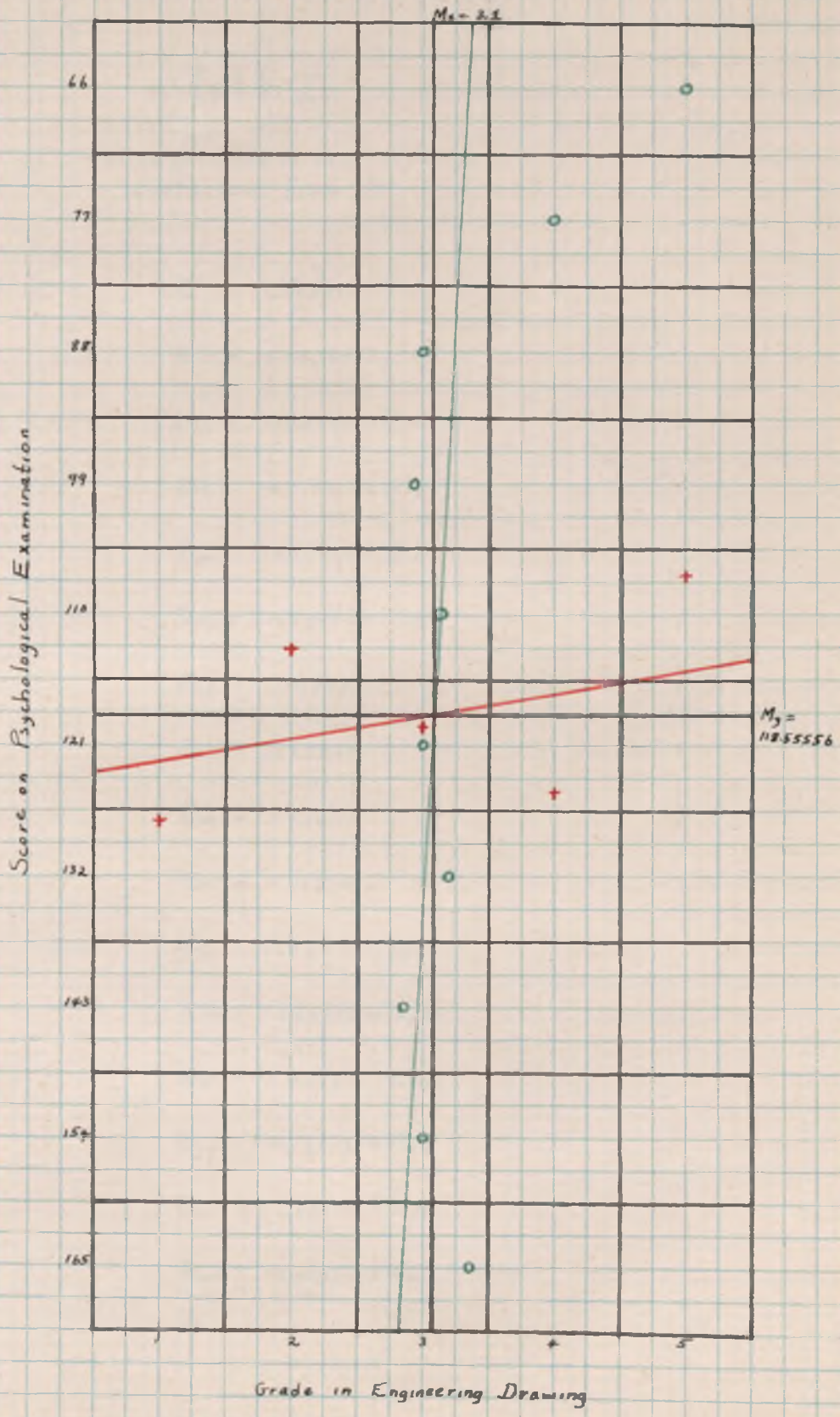

Table XI

Value of Determinant R and Cofactors from R

$R = 0.10555$

$R_{11} = 0.21350$

$R_{12} = -0.04928 = R_{21}$

$R_{13} = -0.11771 = R_{31}$

$R_{14} = -0.01042 = R_{41}$

$R_{15} = -0.03889 = R_{51}$

$R_{22} = 0.18998$

$R_{23} = 0.29203$

$R_{24} = 0.29403$

$R_{25} = 0.16312$

$R_{32} = -0.05230 = R_{23}$

$R_{34} = -0.07982 = R_{43}$

$R_{35} = 0.04901 = R_{53}$

$R_{42} = -0.12942 = R_{24}$

$R_{45} = -0.02113 = R_{54}$

$R_{44} = -0.06936 = R_{55}$

Table XII

Multiple Alienation Coefficients

$$k_{1.2345} = \sqrt{\frac{R}{R_{11}}} = 0.68722$$

$$k_{2.1345} = \sqrt{\frac{R}{R_{22}}} = 0.78802$$

$$k_{3.1245} = \sqrt{\frac{R}{R_{33}}} = 0.60121$$

$$k_{4.1235} = \sqrt{\frac{R}{R_{44}}} = 0.59875$$

$$k_{5.1234} = \sqrt{\frac{R}{R_{55}}} = 0.80441$$

Table XIII

Quotients of Standard Deviations

$\frac{\sigma_1}{\sigma_2}$	= 0.05882	$\frac{\sigma_3}{\sigma_4}$	= 1.25944
$\frac{\sigma_1}{\sigma_3}$	= 0.95234	$\frac{\sigma_3}{\sigma_5}$	= 1.36490
$\frac{\sigma_1}{\sigma_4}$	= 1.19942	$\frac{\sigma_4}{\sigma_1}$	= 0.83374
$\frac{\sigma_1}{\sigma_5}$	= 1.29985	$\frac{\sigma_4}{\sigma_2}$	= 0.04904
$\frac{\sigma_2}{\sigma_1}$	= 17.00241	$\frac{\sigma_4}{\sigma_3}$	= 0.79400
$\frac{\sigma_2}{\sigma_3}$	= 16.19208	$\frac{\sigma_4}{\sigma_5}$	= 1.08373
$\frac{\sigma_2}{\sigma_4}$	= 20.39301	$\frac{\sigma_5}{\sigma_1}$	= 0.76932
$\frac{\sigma_1}{\sigma_5}$	= 22.10057	$\frac{\sigma_5}{\sigma_2}$	= 0.04525
$\frac{\sigma_3}{\sigma_1}$	= 1.05005	$\frac{\sigma_5}{\sigma_3}$	= 0.73265
$\frac{\sigma_3}{\sigma_2}$	= 0.06176	$\frac{\sigma_5}{\sigma_4}$	= 0.92274

Table XIV

β's Used in the Regression Equations

$$\beta_{12.345} = -\frac{R_{12}}{R_{11}} = 0.18098 \qquad \beta_{31.245} = -\frac{R_{31}}{R_{33}} = 0.69273$$

$$\beta_{13.245} = -\frac{R_{13}}{R_{11}} = 0.52687 \qquad \beta_{32.145} = -\frac{R_{32}}{R_{33}} = 0.01026$$

$$\beta_{14.235} = -\frac{R_{14}}{R_{11}} = 0.04750 \qquad \beta_{34.125} = -\frac{R_{34}}{R_{33}} = 0.44218$$

$$\beta_{15.234} = -\frac{R_{15}}{R_{11}} = 0.17462 \qquad \beta_{35.124} = -\frac{R_{35}}{R_{33}} = 0.07216$$

$$\beta_{21.345} = -\frac{R_{21}}{R_{22}} = 0.17226 \qquad \beta_{41.235} = -\frac{R_{41}}{R_{44}} = 0.03606$$

$$\beta_{23.145} = -\frac{R_{23}}{R_{22}} = 0.07631 \qquad \beta_{42.135} = -\frac{R_{42}}{R_{44}} = 0.33904$$

$$\beta_{24.135} = -\frac{R_{24}}{R_{22}} = 0.58727 \qquad \beta_{43.125} = -\frac{R_{43}}{R_{44}} = 0.93858$$

$$\beta_{25.134} = -\frac{R_{25}}{R_{22}} = -0.28830 \qquad \beta_{45.123} = -\frac{R_{45}}{R_{44}} = 0.28558$$

Table XIV Continued

$$\beta_{51.234} = \frac{R_{51}}{R_{55}} = 0.2488$$

$$\beta_{52.134} = \frac{R_{52}}{R_{55}} = 0.3004$$

$$\beta_{53.124} = \frac{R_{53}}{R_{55}} = 0.1295$$

$$\beta_{54.123} = \frac{R_{54}}{R_{55}} = 0.2529$$

Table XV

Regression Equations

$$X_1 = \beta_{1.2345}\frac{\sigma_1}{\sigma_2}X_2 + \beta_{1.3245}\frac{\sigma_1}{\sigma_3}X_3 + \beta_{1.4235}\frac{\sigma_1}{\sigma_4}X_4 + \beta_{1.5234}\frac{\sigma_1}{\sigma_5}X_5 + M_1 - \beta_{1.2345}\frac{\sigma_1}{\sigma_2}M_2 - \beta_{1.3245}\frac{\sigma_1}{\sigma_3}M_3 -$$

$$\beta_{1.4235}\frac{\sigma_1}{\sigma_4}M_4 - \beta_{1.5234}\frac{\sigma_1}{\sigma_5}M_5$$

OR

$$X_1 = 0.00770X_2 + 0.50065X_3 + 1.05697X_4 + 0.22620X_5 - 0.25497$$

$$X_2 = \beta_{21.345}\frac{\sigma_2}{\sigma_1}X_1 + \beta_{23.145}\frac{\sigma_2}{\sigma_3}X_3 + \beta_{24.135}\frac{\sigma_2}{\sigma_4}X_4 + \beta_{25.134}\frac{\sigma_2}{\sigma_5}X_5 + M_2 - \beta_{21.345}\frac{\sigma_2}{\sigma_1}M_1 - \beta_{23.145}\frac{\sigma_2}{\sigma_3}M_3 -$$

$$\beta_{24.135}\frac{\sigma_2}{\sigma_4}M_4 - \beta_{25.134}\frac{\sigma_2}{\sigma_5}M_5$$

OR

$$X_2 = 2.92884X_1 + 0.28548X_3 + 11.97620X_4 - 6.37159X_5 + 90.17172$$

Table XV Continued

$$X_3 = \beta_{3.124}\frac{\sigma_1}{\sigma_1}X_1 + \beta_{32.145}\frac{\sigma_2}{\sigma_1}X_1 + \beta_{34.125}\frac{\sigma_3}{\sigma_4}X_4 + \beta_{35.124}\frac{\sigma_1}{\sigma_5}X_5 + M_3 - \beta_{31.245}\frac{\sigma_3}{\sigma_1}M_1 - \beta_{32.145}\frac{\sigma_2}{\sigma_2}M_2 -$$

$$\beta_{34.125}\frac{\sigma_3}{\sigma_4}M_4 - \beta_{35.124}\frac{\sigma_3}{\sigma_5}M_5$$

OR

$$X_3 = 0.72740X_1 + 0.0006338X_2 + 0.55690X_4 + 0.09826X_5 - 1.46741$$

$$X_4 = \beta_{4.123}\frac{\sigma_4}{\sigma_1}X_1 + \beta_{42.135}\frac{\sigma_4}{\sigma_2}X_2 + \beta_{43.125}\frac{\sigma_4}{\sigma_3}X_3 + \beta_{45.123}\frac{\sigma_4}{\sigma_5}X_5 + M_4 - \beta_{41.235}\frac{\sigma_4}{\sigma_1}M_1 - \beta_{42.135}\frac{\sigma_4}{\sigma_2}M_2 -$$

$$\beta_{43.125}\frac{\sigma_4}{\sigma_3}M_3 - \beta_{45.123}\frac{\sigma_4}{\sigma_5}M_5$$

OR

$$X_4 = 0.030061X_1 + 0.016627X_2 + 0.34823X_3 + 0.25531X_5 - 0.56783$$

Table ~~XV~~ Continued

$$X_E = \left(\beta_{51.234} \frac{\sigma_5}{\sigma_1}\right) X_1 + \left(\beta_{52.134} \frac{\sigma_5}{\sigma_2}\right) X_2 + \left(\beta_{53.124} \frac{\sigma_5}{\sigma_3}\right) X_3 + \left(\beta_{54.123} \frac{\sigma_5}{\sigma_4}\right) X_4 + M_5 - \left(\beta_{51.234} \frac{\sigma_5}{\sigma_1}\right) M_1 - \left(\beta_{52.134} \frac{\sigma_5}{\sigma_2}\right) M_2 -$$

$$\left(\beta_{53.124} \frac{\sigma_5}{\sigma_3}\right) M_3 - \left(\beta_{54.123} \frac{\sigma_5}{\sigma_4}\right) M_4$$

or

$$X_E = 0.18344 X_1 - 0.01359 X_2 + 0.09491 X_3 + 0.19239 X_4 + 3.16851$$

Table XVI

Standard Errors of Order Four.

$$\sigma_{1.2345} = \sigma_1 \left(\frac{R}{R_{11}} \right)^{\frac{1}{2}} = 0.87331$$

$$\sigma_{2.1345} = \sigma_2 \left(\frac{R}{R_{22}} \right)^{\frac{1}{2}} = 1.703345$$

$$\sigma_{3.1245} = \sigma_3 \left(\frac{R}{R_{33}} \right)^{\frac{1}{2}} = 0.80260$$

$$\sigma_{4.1235} = \sigma_4 \left(\frac{R}{R_{44}} \right)^{\frac{1}{2}} = 0.68466$$

$$\sigma_{5.1234} = \sigma_5 \left(\frac{R}{R_{55}} \right)^{\frac{1}{2}} = 0.78679$$

Table XVII

Zero Order Correlations Divided by Probable Errors

$$\frac{r_{34}}{P.E._{r_{34}}} = \frac{0.71923}{0.03678} = 19.55492$$

$$\frac{r_{13}}{P.E._{r_{13}}} = \frac{0.69916}{0.03644} = 19.18661$$

$$\frac{r_{14}}{P.E._{r_{14}}} = \frac{0.58497}{0.04677} = 12.50738$$

$$\frac{r_{24}}{P.E._{r_{24}}} = \frac{0.55949}{0.04884} = 11.45557$$

$$\frac{r_{45}}{P.E._{r_{45}}} = \frac{0.48977}{0.05404} = 9.06310$$

$$\frac{r_{35}}{P.E._{r_{35}}} = \frac{0.47452}{0.05507} = 8.62212$$

$$\frac{r_{15}}{P.E._{r_{15}}} = \frac{0.45882}{0.06317} = 7.26326$$

$$\frac{r_{23}}{P.E._{r_{23}}} = \frac{0.42453}{0.05834} = 7.25968$$

$$\frac{r_{12}}{P.E._{r_{12}}} = \frac{0.39581}{0.05996} = 6.60123$$

$$\frac{r_{25}}{P.E._{r_{25}}} = \frac{0.08671}{0.07056} = 1.22888$$

Table XVIII

Simple, Multiple, and Partial Correlations

Simple Correlations (1)	Partial Correlations (2)	Multiple Correlations (3)
$r_{34} = 0.71923 \pm 0.03678$	$r_{34.125} = 0.44037$	$r_{4.1235} = 0.80094$
$r_{13} = 0.69916 \pm 0.03644$	$r_{13.245} = 0.46092$	$r_{3.1245} = 0.79909$
$r_{14} = 0.58497 \pm 0.04677$	$r_{14.235} = 0.04138$	$r_{1.2345} = 0.72645$
$r_{24} = 0.55949 \pm 0.04884$	$r_{24.135} = 0.44624$	$r_{2.1345} = 0.61566$
$r_{45} = 0.48977 \pm 0.05404$	$r_{45.123} = 0.31650$	$r_{5.1234} = 0.59408$
$r_{35} = 0.47482 \pm 0.05507$	$r_{35.124} = 0.09681$	
$r_{15} = 0.46882 \pm 0.06317$	$r_{15.234} = 0.20370$	
$r_{23} = 0.42253 \pm 0.05834$	$r_{23.145} = 0.01345$	
$r_{12} = 0.39581 \pm 0.05996$	$r_{12.345} = 0.15020$	
$r_{25} = 0.08671 \pm 0.07056$	$r_{25.134} = -0.29429$	

TABLE XIX

TABLE OF GRADES IN FRESHMAN SUBJECTS AND SCORES IN PSYCHOLOGICAL EXAMINATION

Student no.	x_1 Grade in math.	x_2 Score on psy. test	x_3 Grade in chem. 1a or 2a	x_4 Grade in Eng.	x_5 Grade in engr. draw.
(1)	3	118	5	4	4
(2)	3	120	2	3	5
(3)	3	123	3	3	3
(4)	2	99	2	3	4
(5)	1	129	1	2	4
(6)	4	99	3	3	3
(7)	4	115	3	2	3
(8)	1.4	120	3	4	3
(9)	2	84	3	5	3
(10)	1	156	1	2	3
(11)	5	134	5	3	4
(12)	1	127	2	3	3
(13)	1	138	1	2	3
(14)	5	124	4	5	4
(15)	4	83	5	5	3
(16)	4.4	112	5	5	5
(17)	1	102	1	3	2
(18)	5	126	4	4	3
(19)	2	137	1	2	2

TABLE XIX (cont.)

Student no.	x_1 Grade in math.	x_2 Score on psy. test	x_3 Grade in chem. 1a or 2a	x_4 Grade in Eng.	x_5 Grade in engr. draw.
(20)	4.4	99	4	4	3
(21)	1.6	167	2	2	3
(22)	1	132	1	2	1
(23)	3.6	110	2	3	3
(24)	4.29	108	5	5	3
(25)	2	126	3	3	3
(26)	4.4	121	4	4	3
(27)	4	103	3	3	3
(28)	2.4	135	2	3	1
(29)	2.6	119	1	3	3
(30)	2.6	105	3	5	3
(31)	4.6	106	5	5	4
(32)	5	81	5	5	4
(33)	4	132	3	4	5
(34)	4	122	4	3	3
(35)	2	73	3	4	3
(36)	2	144	4	1	3
(37)	1	144	1	2	2
(38)	1.6	89	4	5	3
(39)	3	111	3	3	3
(40)	1	125	1	2	2

TABLE XIX (cont.)

Student no.	x_1 Grade in math.	x_2 Score on psy. test	x_3 Grade in chem. la or 2a	x_4 Grade in Eng.	x_5 Grade in engr. draw.
(41)	2.4	123	2	2	2
(42)	1	132	1	3	3
(43)	3	137	4	3	3
(44)	2.5	167	4	4	3
(45)	5	103	5	5	5
(46)	3	107	3	3	2
(47)	1.4	114	1	3	2
(48)	4	154	2	3	3
(49)	3.6	84	4	4	3
(50)	2.6	85	3	3	1
(51)	2.4	134	1	2	1
(52)	1	144	1	3	3
(53)	3	101	5	5	4
(54)	2.6	136	3	3	3
(55)	4.4	106	4	4	3
(56)	3	106	1	2	3
(57)	3.4	137	3	3	5
(58)	2	149	2	3	4
(59)	3	155	3	3	3
(60)	4.6	132	4	4	5
(61)	3.6	130	1	2	3

TABLE XIX (cont.)

Student no.	x_1 Grade in math.	x_2 Score on psy. test	x_3 Grade in chem. 1a or 2a	x_4 Grade in Eng.	x_5 Grade in engr. draw.
(62)	2	96	3	2	2
(63)	4	115	2	3	3
(64)	1	145	1	2	3
(65)	5	70	5	5	5
(66)	1	123	1	3	2
(67)	3.4	96	3	4	2
(68)	1.6	148	2	2	3
(69)	5	101	4	4	3
(70)	1.6	150	2	2	4
(71)	3.6	92	3	5	5
(72)	1	113	2	2	3
(73)	5	72	5	5	5
(74)	1.8	127	3	3	4
(75)	2.8	111	2	4	3
(76)	2	104	3	3	2
(77)	3.6	101	3	4	2
(78)	2	157	1	1	1
(79)	1.8	114	2	2	2
(80)	4.4	107	5	5	5
(81)	4	105	5	4	3
(82)	2	154	2	3	3
(83)	2	108	1	2	3

TABLE XIX (cont.)

Student no.	x_1 Grade in math.	x_2 Score on psy. test	x_3 Grade in chem. 1a or 2a	x_4 Grade in Eng.	x_5 Grade in engr. draw.
(84)	3.875	101	4	4	3
(85)	2.4	100	4	4	3
(86)	3	131	3	3	4
(87)	4.2	119	4	3	3
(88)	2.4	163	3	3	4
(89)	2	116	1	2	2
(90)	2.4	123	1	3	3

BIBLIOGRAPHY

Books

Holzinger, Karl J. Statistical Methods for Students in Education. Boston: Ginn and Co., 1928.

Reitz, Henry Louis (ed.). Handbook of Mathematical Statistics. Chicago: Houghton Mifflin Co., 1924.

Reitz, Henry Louis. Mathematical Statistics. LaSalle, Illinois: The Open Court Publishing Co., 1927.

Unpublished Material

Allen, Ruth Weyer. "Correlation among Certain Grammar School Subjects." Unpublished Master's thesis, Department of Mathematics, University of Arizona, 1941. Pp. 67.