PROPINQUITY ANALYSIS: A NEW GUIDANCE TOOL*

Marion F. Shaycoft
American Institutes for Research
Palo Alto, California

A multiple aptitude battery is a useful tool in vocational and educational guidance but there has been no clear agreement on the best way to use it. Assuming that the problem is how best to help the individual decide which of several possible occupational categories or educational programs is the one which will be best for him, three statistical techniques for developing a predictive equation for each category have been available: (1) multiple correlation other than multiple point biserial, (2) multiple discriminant analysis, and (3) multiple point biserial correlation. In the interests of brevity, the first-named of those three techniques will be referred to in this paper simply as "multiple correlation".

The term "multiple correlation technique", as used in this context, means that a separate within-group criterion of effectiveness of performance is available and that for prediction of membership in each category a separate multiple correlation coefficient and multiple regression equation are computed.

The second technique, multiple discriminant analysis, is applicable when the sole criterion available is membership in a group; in other words when there is no within-group criterion to distinguish <u>among</u> members of the group. The technique produces a set of discriminant functions, each of which is a linear composite of the original predictor variables. The discriminant functions have to be considered in combination, since they do not correspond on a one-to-one basis to the criterion categories.

---

*A very slightly abridged version of this paper was presented at the Annual Meeting of the National Council on Measurement in Education, in Minneapolis, March 6, 1970.

The third commonly used approach, multiple point biserial correlation, is really a cross between the first two. It is multiple correlation against a dichotomous criterion, and therefore, like the multiple discriminant analysis technique, it is usable even when membership-versus-nonmembership in the group is the only criterion information available. The resulting multiple regression function is exactly equivalent (except for a scaling factor) to a discriminant function that discriminates optimally between the group in question and all other groups combined.

But even the simultaneous use of all three of these techniques -- multiple correlation (against a continuous criterion), multiple discriminant analysis, and multiple point biserial R -- doesn't yield all the relevant information available. None of the three really comes to grips adequately with the kind of situation, sometimes occurring in the use of aptitude batteries, in which overqualification for group membership is just about as undesirable as underqualification is. The problem is still further aggravated when, as often happens, group membership criteria are the only kind available. Other difficulties, particularly applicable to the discriminant function approach, lie in the fact that because most of the discriminant functions are bipolar, they are hard to interpret, hard to explain, and generally obscure in meaning. This makes them peculiarly unsuitable for use in guidance. And discriminant functions have the further disadvantage that the results for one criterion category depend to an inordinate degree on what other categories happen to be included in the analysis.

In an effort to avoid all these problems and to supplement information provided by the more usual approaches, the writer has worked out a new approach, which is still in the developmental stage. This approach, called propinquity analysis, requires no criteria other than group membership, and although it superficially has some of the features of each of the other approaches it is actually quite different from any of them.*

THE NATURE AND CHARACTERISTICS OF PROPINQUITY INDEXES

Let's assume we have a scatterplot in which each individual's set of scores is represented by a point in n-dimensional space, where the standard score scale on each test constitutes a separate dimension. Let's now suppose that for purposes of developing propinquity indexes for a given occupational group the scatterplot is rescaled, weighting each dimension by a value representing, at least approximately, the relevance of the corresponding variable in identifying group members. An individual's propinquity index with respect to that group may then be defined as his geometric distance from the group centroid, in this rescaled space. A minus sign is attached to the distance, so that 0 is the maximum value of the index. A zero index indicates that the individual's scores on relevant variables coincide with the group centroid. The greater the distance, the larger the absolute value of the index. The purpose of the minus sign, therefore, is to orient the propinquity index properly, so that when it is used in a correlation matrix there won't be inconvenient negative correlations. The higher the algebraic value of the index, the closer the individual is to the centroid. (Hence the term "propinquity".)

Formula 14 (or 15)** represents the squared distance in regard to a single variable. Formula 16 gives the propinquity index, δ, with w as the weight representing the relevance of a particular variable.

Formulas 27-33 are seven different formulas giving different results, that might be used for determining the w's. All these formulas have the

*An earlier paper by Shaycoft entitled "A New Multivariate Index for Use in Educational Planning", which was presented at the American Psychological Association Convention in Washington on September 2, 1969, goes into considerably more detail on the limitations of the other three methods. Copies are available on request, while the supply lasts.

**All formulas in this paper are in the Appendix, which also contains a section defining all the notation used.

desired characteristic of giving a weight of 0 for irrelevant variables and a positive weight for relevant variables. There are no negative weights. The indicator of relevance for a variable is a function of the ratio of the standard deviation within the group to the standard deviation of the total sample. Thus there is an assumption that one way a variable's relevance may be manifested is by a somewhat restricted range, and that the more relevant the variable, the more restricted the range. This phenomenon may be seriously affected when some of the variables have skewed distributions. For such variables the ratio of group σ to total σ may be erratic, resulting in peculiar weights. Therefore it is desirable that all variables used in a propinquity analysis have normal distributions. If there is no reason to believe that the distributions are at least approximately normal they should be normalized.

Multiple regression weights (against the dichotomous criterion of group membership), as suggested by formula 39, may constitute a more effective approach than any of the seven formulas 27-33. Regression weights would of course have to be cross-validated; and small negative weights should probably be changed to 0 first. (Presumably there wouldn't be any large negative weights.) The chief disadvantage of regression weights lies in the very practical consideration that even with a high-powered computer, the computation of the betas for these squared terms is an extraordinarily complex, time-consuming, and expensive operation when the number of cases, number of variables, and number of groups are all comparatively large (as is true in the case of the Project TALENT data, to which this technique is being applied).

Because the variables used in computing the propinquity index are raw scores rather than principal components, discriminant functions, or some other kind of uncorrelated variables, propinquity indexes may lack some of the mathematical precision and invariance of statistics computed in a geometric space where orthogonal axes correspond to uncorrelated variables. But this seems a small price to pay for the twin advantages of interpretability and ready explainability.

PROPINQUILES

Now we come to a problem in connection with propinquity indexes. When the number of variables entering into the index is greater than one, the distribution is not normal and its basic shape varies with the number of variables involved. Propinquity indexes based on three variables have an entirely different distribution from those based on two, and so forth.

And even when propinquity indexes for a set of groups are all based on the same battery they still may not be directly comparable, because though all indexes technically are based on the same number of variables, some variables may have zero weights for certain groups so that the number of dimensions is in effect different for different groups. Since propinquity indexes for different categories therefore are not generally directly comparable, they have to be converted to some uniform scale in order to be compared - and percentiles serve this purpose effectively and directly. This calls for coinage of a new word. The percentiles are called "propinquiles".

PROBLEMS IN DETERMINING WEIGHTS

The nature of the change in the shape of the distribution of the propinquity index as the number of variables increases causes some of the difficulties in determining a wholly suitable a priori formula for the weights. In a small-scale tryout of formulas 27-33, for eight career groups, results suggested formula 27 was slightly better than any of the others; for six of the eight groups, it produced δ^2 values that had at least slightly higher point biserial correlations with the group membership criterion than did any of the other formulas. In addition to this rather slight empirical evidence there are some theoretical arguments in support of formula 27 -- for instance the fact that it bears a rather close relationship to the square of a type of correlation ratio. Nevertheless even this "best" a priori formula gives somewhat disappointing results, leading to the suspicion that it may not be the best weighting procedure after all.

Possibly multiple regression weights should be used, rather than any *a priori* formula, *despite* the computational complexities that this would introduce into the data analysis.

A possible alternative is the so-called city block model for defining distance. This alternative kind of propinquity index is represented by δ', and its value is given by formula 21. With it go new formulas w_1'-w_5' for *a priori* weights (Formulas 34-38). This alternative system, δ',hasn't been tried yet, and it is probably worth trying, but the chances seem slight that it will prove to have any advantages over δ, while δ has at least one major advantage over δ' -- the fact that it is invariant under rotation of axes.

This brings us to the empirical data summarized in Table 1 (in the Appendix). It is important to bear in mind that this is *preliminary* data. The study is still in progress.

EMPIRICAL DATA

Table 1 is based on about 14000 twelfth-grade boys tested in Project TALENT in 1960 for whom scores on all 109 predictor variables used in the study are available and for whom follow-up data obtained five years after the class graduated from high school are also available. These follow-up data provided the information about the long-range career plans of the TALENT sample. Ten of these career-plan groups have been selected for inclusion in Table 1.*

The total group has been divided into two parallel samples, A and B, so that propinquity indexes could be calculated for the individuals in one sample on the basis of statistics (means and standard deviations) determined from the other, in order to avoid capitalizing on chance.

All 109 variables were normalized with a mean of 0 and standard deviation of 1, on the basis of Sample A distributions. The same normalization conversion table was then applied to the Sample B cases, to make the Sample B distributions approximately normal. Propinquity

*The choice of these particular 10 career plan groups was based primarily on their size (they are all large enough to provide comparatively stable data) and the variety of fields and levels they represent.

indexes were calculated for every Sample B case, using weights, means, and standard deviations calculated from Sample A data. Formula 27 was used for the weights.

Columns 1-3 show the number of cases in each group. Column 4 shows the proportion, and column 5 shows the factor by which a point biserial correlation is multiplied to convert it to a biserial. Column 6 contains the conventional multiple point biserial R, with group membership as the dichotomous criterion and the 109 normalized variables as the predictors. These point biserials are necessarily quite low because of the very low ceiling on point biserial correlations when the dichotomous split is far out on a tail of the hypothesized normal distribution. The conversion to multiple biserial correlations, shown in column 7, makes all ten values directly comparable, since it eliminates the effect of different numbers of cases in the different groups.

Interpretation of the biserial correlations of course requires one to conceptualize that underlying each dichotomy is a unidimensional normal continuum; for instance that there is a normally distributed continuum of "CPA-ness" -- tendency to become a CPA -- and a threshhold above which everybody becomes a CPA and below which nobody does. It is important to bear in mind that this hypothesized threshhold applies only to the hypothesized normal variable underlying the dichotomized criterion -- and not to any composite of the predictive variables (except in the extremely improbable event that the composite has perfect validity).

Column 8 contains the point biserial correlation of the propinquity index, δ, with the group membership criterion, and column 9 contains the corresponding biserial. These correlations are substantial; they are lower than the corresponding multiples in columns 6 and 7 but that is to be expected since the correlations in columns 6 and 7 are based on 109 variables combined optimally and are not cross-validation values -- so there may be a little capitalization on chance there, while there is no capitalization on chance in the case of the propinquity indexes.

The critical question is whether adding the propinquity index as a 110th predictor would increase the column 6 multiple point biserial correlations significantly. The data for a definite answer on that question are not yet available, but preliminary indications are that though the amount of increase produced in the R by the addition of the propinquity index to the group of predictive variables would be relatively small (assuming that formula 27 is used for the weights) it would be significant. To the extent that that δ adds new predictive variance, not measured by the 109 predictive variables used for the column 6 multiples, R will be raised by the addition of δ as a predictor. And there is evidence that much of the propinquity index variance is unique. For instance the variables that are weighted the highest in computing δ are not necessarily the ones on which group members score especially high or especially low. They may be variables on which the average score of a group member is very close to the overall mean and the point biserial of that variable with the group membership criterion is probably correspondingly low.

The last eight columns of the table provide some evidence on this point. The highest and second highest weights for the computation of each of the 10 propinquity indexes are shown in columns 13 and 17. The standard score means and point biserials for many of these variables are quite close to 0, as can be seen in columns 15, 16, 19, and 20.

And as has already been indicated, the empirical data provide some suggestion that using multiple regression weights or some other modification of the formulas might make a substantial improvement in the results. One indication of this is in columns 10-12. Some of the dichotomous criteria turn out to have higher correlations with the δ computed for another group than with the appropriate δ. This certainly strongly suggests that better weights could be found.

CONCLUSIONS

Propinquity analysis is not in any sense of the word a replacement for the multiple correlation approach, nor for the multiple point biserial R approach. Rather, it should serve as an adjunct to both these procedures. It may be used in two ways in helping the individual develop his educational and vocational plans. Converted to a propinquile, it can constitute one of many separate items of information used in arriving at important decisions. Or in many circumstances the propinquity index may function better as one of the predictors in a multiple regression equation to predict a dichotomous group membership criterion.*

In the vocational planning or vocational guidance situation propinquity analysis is likely to be most useful for middle-level occupations, and perhaps even, to a certain extent, for lower-level occupations, rather than for the more demanding occupations, where underqualification is an overwhelmingly more potent deterrent than overqualification is, and the higher one's qualifications are the better.

In any event, further exploration of computing procedures is necessary. It seems clear both on logical grounds and on the basis of preliminary data that the components entering into the computation of a propinquity index provide unique predictive variance -- variance not provided by the conventional scores -- and that this is not just error variance; it is specific reliable variance, and when appropriately weighted it should contribute significantly to the validity of prediction. But better weighting procedures (and perhaps other computational modifications as well) must be uncovered, in order to mine efficiently this rich and hitherto unexploited source of valid variance.

---

*This of course couldn't be proposed if the propinquity index were merely some sort of linear function of the $\underline{n}$ initial predictors, but it isn't, since squared terms are also involved. The propinquity index thus is not linearly dependent on the $\underline{n}$ variables which enter into it.

SELECTED REFERENCES ON PROJECT TALENT

Flanagan, J.C., Dailey, J.T., Shaycoft, Marion F., Gorham, W. A., Orr, D. B., & Goldberg, I. Design for a Study of American Youth, 248 pp. Boston: Houghton Mifflin, 1962.

Flanagan, J.C., Dailey, J.T., Shaycoft, Marion F., Orr, D.B., & Goldberg, I. Studies of the American High School. (Cooperative Research Project No. 226) Project TALENT Office, 1962. 375 pp.

Flanagan, J.C., Davis, F.B., Dailey, J.T., Shaycoft, Marion F., Orr, D.B., Goldberg, I., & Neyman, C.A., Jr. The American High School Student. (Cooperative Research Project No. 635) Project TALENT Office, 1964. 738 pp.

Flanagan, J.C., Cooley, W.W., Lohnes, P.R., Schoenfeldt, L.F., Holdeman, R.W., Combs, Janet, & Becker, Susan. Project TALENT One-Year Follow-Up Studies. (Cooperative Research Project No. 2333) Project TALENT Office, 1966. 348 pp.

Shaycoft, Marion F., Dailey, J.T., Orr, D.B., Neyman, C.A., Jr., & Sherman, S.E. Studies of a Complete Age Group--Age 15. (Cooperative Research Project No. 566) Project TALENT Office, 1963. 288 pp.

Shaycoft, Marion F. The High School Years: Growth in Cognitive Skills. (Cooperative Research Project No. 3051) Project TALENT Office, 1967. 376 pp.

APPENDIX

I. NOTATION

n = no. of variables (= no. of dimensions)

g = no. of groups

N_j = no. of cases in group j

N = total no. of cases

$$N = \sum_{j=1}^{g} N_j \qquad (1)$$

p_j = proportion of cases in group j

$$p_j = N_j/N \qquad (2)$$

$$q_j = 1 - p_j \qquad (3)$$

y_j = normal ordinate corresponding to point of split between p_j and q_j

X_{ik} = raw score (or normalized score if the variables have been normalized) of individual k on variable i

$i = 1,2,3,\ldots n$
$k = 1,2,3,\ldots N_j$

$\overline{X}_1$ = mean of variable X_i for total group

$$\overline{X}_i = \frac{\sum_{k=1}^{N} X_{ik}}{N} \qquad (4)$$

s_{x_i} = sample standard deviation of variable X_i for total group

$$s_{x_i} = \sqrt{\frac{\sum_{k=1}^{N} (X_{ik} - \overline{X}_i)^2}{N}} \qquad (5)$$

σ_{x_i} = corresponding estimate of population standard deviation

$$\sigma_{x_i} = s_{x_i} \sqrt{\frac{N}{N-1}} \qquad (6)$$

z_{ik} = standard score of individual k on variable i

$$z_{ik} = \frac{X_{ik} - \overline{X}_i}{\sigma_{x_i}} \qquad (7)$$

$$\overline{z}_i = 0 \qquad (8)$$

Note that σ_x not s_x is in the denominator of z_{ik} in this algebraic development.

$$\therefore \quad s_{z_i} = \sqrt{\frac{N-1}{N}} \qquad (9)$$

$$\sigma_{z_i} = 1 \qquad (10)$$

$\overline{X}_{ij}$ = means of variable X_i for group j

$$\overline{X}_{ij} = \frac{\sum_{k=1}^{N_j} X_{ik}}{N_j} \qquad (11)$$

$i = 1,2,3,\ldots n$
$j = 1,2,3,\ldots g$

$s_{x_{ij}}$ = sample standard deviation of variable i for group j

$$s_{x_{ij}} = \sqrt{\frac{\sum_{k=1}^{N_j} (X_{ik} - \overline{X}_{ij})^2}{N_j}} \qquad (12)$$

$\sigma_{x_{ij}}$ = estimate of population standard deviation of variable i for group j

$$\sigma_{x_{ij}} = s_{x_{ij}} \sqrt{\frac{N_j}{N_j - 1}} \qquad (13)$$

U_{ijk} = propinquity component indicating how much like group j individual k is, with respect to variable i.

L_{ijk} is an alternative to U_{ijk}

w_{ij} = weight representing the relevance of component U_i as an indicator of membership of individual in group j.

w'_{ij} = weight representing the relevance of component L_i as an indicator of membership of individual in group j.

δ_{jk} = propinquity index for individual k indicating how much like group j he is, assuming the U_{ijk} formula has been used for defining likeness.

δ'_{jk} = propinquity index for individual k indicating how much like group j he is, assuming the L_{ijk} formula has been used for defining likeness.

P_{jk} = propinquile for individual k corresponding to his propinquity index for group j. The propinquile is a percentile based on the individuals in group j. A propinquile of 100 corresponds to a propinquity index of 0.

II. FORMULAS

$$U_{ijk} = \left(\frac{z_{ik} - \bar{z}_{ij}}{\sigma_{z_{ij}}} \right)^2 \qquad (14)$$

$$= \left(\frac{X_{ik} - \bar{X}_{ij}}{\sigma_{x_{ij}}} \right)^2 \qquad (15)$$

$$\delta_{jk} = -\sqrt{\sum_{i=1}^{n} w_{ij} \, U_{ijk}} \qquad (16)$$

$$\sigma_{z_{ij}} = \frac{\sigma_{x_{ij}}}{\sigma_{x_i}} \qquad (17)$$

$$= \frac{\sigma_{x_{ij}}}{s_{x_i}} \sqrt{\frac{N-1}{N}} \qquad (18)$$

$$L_{ijk} = \sqrt{U_{ijk}} \qquad (19)$$

$$= \left| \frac{X_{ik} - \bar{X}_{ij}}{\sigma_{x_{ij}}} \right| \qquad (20)$$

$$\delta'_{jk} = -\sum_{i=1}^{n} w'_{ij} \, L_{ijk} \qquad (21)$$

Formulas for intermediate values in computing w_{ij} and w'_{ij}

$$w''_1 = 1 - \sigma^2_{z_{ij}} \quad (22)$$

$$w''_2 = \frac{1}{\sigma^2_{z_{ij}}} - 1 \quad (23)$$

$$w''_3 = 1 - \sigma_{z_{ij}} \quad (24)$$

$$w''_4 = \frac{1}{\sigma_{z_{ij}}} - 1 \quad (25)$$

$$w''_5 = \log_{10} \frac{1}{\sigma_{z_{ij}}} \quad (26)$$

Formulas for w_{ij} and w'_{ij}

$w_1 = w''_1$ with negative values changed to 0 (27)

$w_2 = w''_2$ " " " " " " (28)

$w_3 = w''_3$ " " " " " " (29)

$w_4 = w''_4$ " " " " " " (30)

$w_5 = w''_5$ " " " " " " (31)

$$w_6 = {w_3}^2 \quad (32)$$

$$w_7 = {w_4}^2 \quad (33)$$

$$w'_1 = \sqrt{w_1} \quad (34)$$

$$w'_2 = \sqrt{w_2} \quad (35)$$

$$w'_3 = w_3 \quad (36)$$

$$w'_4 = w_4 \quad (37)$$

$$w'_5 = w_5 \quad (38)$$

w_o = multiple regression beta weight when the predictors are U-variables (39)

w'_o = multiple regression beta weight when the predictors are L-variables (40)

TABLE 1: Multiple point biserial R's with group membership, point biserial r's with δ, corresponding biserial correlations, and related data, for 10 selected "purified" career groups

Based on five-year follow-up of twelfth-grade boys*; (N = 14123)

n = 109
Predictor variables: 109 normalized** variables from the TALENT Battery

j	Career group code #	Career	(1) No. of cases A	(2) No. of cases B	(3) No. of cases T	(4) p	(5) $\sqrt{pq}/y$	(6) Mult. pt. bis. R	(7) Mult. bis. R	(8) r with corresp δ: Pt. bis. $r_{j\delta}$	(9) r with corresp δ: Bis $r_{j\delta}$	(10) For highest $\lvert r_{j\delta}\rvert$ among the 10 for group j: Pt. bis. $r_{j\delta}$	(11) For highest $\lvert r_{j\delta}\rvert$ among the 10 for group j: Bis. $r_{j\delta}$	(12) Group Code for δ	(13) For weights (w_{ij}) for computing appropriate δ***: Highest w_{ij}	(14) Corresp. variable (i)	(15) $\bar{z}_{ij}$	(16) Bis. r_{ji}	(17) Second highest w_{ij}	(18) Corresp. variable (i)	(19) $\bar{z}_{ij}$	(20) Bis. r_{ji}
1	748	Salesman	175	175	350	.02494	2.6734	.156	.417	.060	.162	(.060)	(.162)	(748)	.3482	R-142	-.070	-.040	.3112	P-820	-.330	-.059
2	730	Business mgmt.	256	255	511	.03634	2.3482	.167	.391	.057	.133	.072	.169	393	.2994	A-410	.190	.085	.2791	P-702	.134	.030
3	716	CPA	140	140	280	.01995	2.8949	.183	.530	.067	.193	(.067)	(.193)	(716)	.5124	R-101	.122	.046	.4168	R-311	.306	.055
4	393	Lawyer	197	196	393	.02793	2.5697	.259	.665	.112	.289	-.149	-.383	639	.5633	R-136	.473	.267	.4782	R-101	.110	.069
5	329	M.D. (NEC)	115	114	229	.01625	3.1174	.281	.876	.120	.374	-.179	-.557	899	.6794	R-136	.590	.212	.6049	R-250	1.002	.365
6	242	Elec. engineer	122	121	243	.01724	3.0504	.246	.749	.108	.331	-.123	-.376	899	.4973	R-270	.756	.336	.4779	P-701	1.177	.448
7	811	Electronic tech.	106	106	212	.01511	3.2010	.178	.570	.049	.156	(.049)	(.156)	(811)	.4287	R-234	.028	-.115	.3984	P-714	.519	.205
8	125	Electronics (NEC)	78	78	156	.01112	3.5874	.166	.594	.033	.120	.034	.123	811	.4122	R-250	-.025	-.097	.3991	R-232	-.294	-.072
9	639	Farming (NEC)	122	122	244	.01739	3.0414	.278	.846	.104	.315	-.131	-.398	329	.4639	R-104	-.748	-.286	.4573	R-250	-.630	-.316
10	899	Unskilled labor	139	138	277	.01967	2.9087	.195	.567	.112	.325	-.141	-.410	329	.5486	R-333	-.770	-.279	.4857	R-106	-.888	-.335
⋮																						
Total			7106	7017	14123	1.00000																
Foot note #'s	1	2	3	4		4	4	4,6-7		4-6				1	3,8	9	3	4,6	3,8	9	3	4,6

* The group was divided into 2 samples, A and B; sample A consisting of the 1st, 3rd, 5th,... cases in each career group, and sample B consisting of the 2nd, 4th, 6th....

** Normalization was based on the sample A distribution.

*** δ is the propinquity index to which column 8 (or 9) applies.

1. TALENT career plan code series Q-2-cp-A
2. NEC stands for "Not elsewhere classified".
3. Based on sample A
4. Based on sample B
5. $\bar{X}$ and σ values entering into the computation of w_{ij} (Formula 27) and δ_j are based on sample A; 109 U-variables based on the normalized TALENT variables are used.
6. Dichotomous group membership variable is the criterion (X_j).
7. The 109 normalized TALENT variables are the predictors.
8. Excluding w_{ij}'s for dichotomous predictor variables.
9. Code for TALENT variables represented in columns 14 and 18

R-101 Screening
R-104 Music Information
R-106 Math Information
R-136 Journalism Information
R-142 Bible Information

R-232 Capitalization (English subtest)
R-234 English Usage (English subtest)
R-250 Reading Comprehension

R-270 Mechanical Reasoning
R-311 Arith. Reasoning
R-333 Adv. h.s. math.

A-410 Arith Comp. attempts

P-701 Interest in phys. sci., engineering, math.
P-702 Interest in biol. sci. and medicine
P-714 Mechanical-Technical interest

P-820 High school grades