

TRAITEMENT STATISTIQUE D'UN PROBLEME DIFFERENTIEL D'ORIENTATION SCOLAIRE

par Ed. ROBAYE,
Université libre de Bruxelles

Dans le courant de l'année 1959, 2.181 élèves, terminant leurs études secondaires inférieures (4^e), ont été examinés à l'aide d'une batterie de tests psychologiques ⁽¹⁾. On dispose également de tous les résultats scolaires de ces élèves.

Au cours de l'année 1960, on a relevé les résultats scolaires de ceux de ces élèves qui étaient passés dans la classe suivante de l'enseignement secondaire (3^e).

Le Ministère de l'Education nationale et de la Culture a demandé au laboratoire de calcul de l'U.L.B. (Prof. Paul Gillis) d'étudier ces documents. Il s'agissait d'une part de fournir un certain nombre de paramètres statistiques à propos des variables envisagées, et d'autre part de proposer une utilisation des résultats à des fins d'orientation.

Rappelons qu'il existe des classes de 4^e anciennes, 4^e modernes, 4^e techniques, 4^e familiales, et des classes de 3^e latin-grec, 3^e latin-sciences, 3^e latin-mathématiques, 3^e économiques, 3^e scientifiques.

I. Description de l'échantillon.

Nous classerons les informations fournies au sujet de chaque élève en deux classes :

1) Informations catégorielles.

- a) l'âge (14, 15, 16 ou 17 ans; âge normal : 15 ans);

(1) — 3 épreuves verbales : BV 8 (vocabulaire)
BV 16 (analogies de phrases)
BV 50 (analogies de phrases plus faciles que le BV 16)
— 2 épreuves non verbales : BINV 53 (données graphiques)
B 9 (données en terminologie scientifique)
— 1 questionnaire de contrôle des intérêts professionnels (R. Derivière)
— toutes les épreuves ont été appliquées dans la même semaine (fin du 2^e trimestre), du 23 au 27 mars 1959, par les conseillers directeurs de tous les centres psychomédico-sociaux et conseillers de régime français (au nombre de 12), dans toutes les écoles de leur ressort (57 classes), au niveau de la fin du secondaire inférieur.

- b) la section (anciennes, modernes, latin-grec, etc.);
 - c) le sexe;
 - d) la résidence (urbaine ou rurale);
 - e) le niveau socio-culturel des parents (primaire, secondaire inférieur, secondaire supérieur, universitaire);
 - f) la classe d'intérêts professionnels définie à partir des réponses au questionnaire (S = scientifique, 0 = indifférent, L = littéraire).
- 2) *Informations quantitatives* (8 variables).
- a) les résultats scolaires de fin d'année;
 - M_4 M_3 : résultats sur 100 en mathématiques, en 4^e et en 3^e;
 - F_4 F_3 : résultats sur 100 en français, en 4^e et en 3^e;
 - T_4 T_3 : résultats sur 100 Total de l'année, en 4^e et en 3^e.
 - b) les résultats aux tests
 - V : total des trois tests verbaux;
 - NV : total des deux tests non verbaux.

La répartition des 2.181 élèves de 4^e suivant les catégories retenues est la suivante :

Répartition par âge :

| | |
|----------|-------|
| 14 ans : | 381 |
| 15 ans : | 1.007 |
| 16 ans : | 597 |
| 17 ans : | 196 |

Répartition d'après les sections d'étude :

| | |
|----------------------------|-------|
| Anciennes : | 713 |
| Modernes : | 1.015 |
| Techniques et Familiales : | 453 |

Répartition par sexe :

| | |
|-----------|-------|
| Garçons : | 1.221 |
| Filles : | 960 |

Répartition par résidence :

| | |
|-----------|-------|
| Urbains : | 885 |
| Ruraux : | 1.296 |

Répartition par niveau socio-culturel des parents :

| | |
|------------------------------------|-----|
| Parents universitaires : | 272 |
| Parents « secondaire supérieur » : | 358 |
| Parents « secondaire inférieur » : | 766 |
| Parents « primaire » : | 785 |

Les résultats au questionnaire « Derivière » se présentant sous une forme qualitative, comme les caractéristiques citées ci-dessus, nous donnons dès à présent la répartition qui y correspond.

Intérêt 0 : 944

Intérêt L : 513

Intérêt S : 724

Le tableau I donne, pour le groupe total et pour chacun des sous-groupes, la moyenne \bar{x} et l'écart-type s de la distribution des cinq épreuves relatives à la 4^e.

TABLEAU I

| | Français | | Math. | | Total | | V | | NV | |
|------------------|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|
| | \bar{x} | s |
| Groupe Total | 64 | 10 | 61 | 15 | 67 | 9 | 183 | 36 | 83 | 19 |
| 14 ans | 66 | 10 | 63 | 14 | 68 | 9 | 191 | 35 | 89 | 17 |
| 15 ans | 65 | 10 | 61 | 15 | 67 | 9 | 185 | 36 | 85 | 19 |
| 16 ans | 63 | 10 | 60 | 14 | 66 | 8 | 179 | 35 | 80 | 18 |
| 17 ans | 63 | 10 | 59 | 16 | 66 | 10 | 173 | 35 | 75 | 19 |
| Anciennes | 68 | 10 | 63 | 15 | 69 | 9 | 204 | 31 | 92 | 17 |
| Modernes | 64 | 10 | 59 | 15 | 65 | 9 | 182 | 31 | 83 | 16 |
| Tech. et Fam. | 62 | 11 | 63 | 14 | 69 | 8 | 154 | 30 | 69 | 17 |
| Garçons | 63 | 10 | 59 | 15 | 65 | 9 | 182 | 35 | 85 | 19 |
| Filles | 66 | 10 | 63 | 15 | 69 | 8 | 186 | 36 | 81 | 18 |
| Urbains | 64 | 10 | 61 | 14 | 66 | 9 | 191 | 35 | 86 | 18 |
| Ruraux | 65 | 11 | 61 | 15 | 67 | 9 | 178 | 36 | 81 | 19 |
| Niveau primaire | 64 | 11 | 60 | 14 | 67 | 9 | 175 | 36 | 80 | 18 |
| Niveau sec. inf. | 65 | 10 | 61 | 15 | 67 | 8 | 184 | 34 | 83 | 18 |
| Niveau sec. sup. | 66 | 10 | 63 | 15 | 67 | 9 | 193 | 34 | 86 | 18 |
| Niveau univ. | 66 | 11 | 61 | 16 | 67 | 10 | 193 | 35 | 87 | 19 |
| Intérêt L | 67 | 10 | 59 | 14 | 68 | 9 | 192 | 37 | 83 | 18 |
| Intérêt O | 64 | 10 | 59 | 15 | 66 | 9 | 182 | 34 | 82 | 18 |
| Intérêt S | 63 | 10 | 64 | 14 | 67 | 8 | 179 | 36 | 84 | 19 |

A l'aide d'analyses de variance à un facteur de classification, nous avons testé les différences entre moyennes au niveau de chacune des classifications qualitatives, et cela pour chaque épreuve. On a pu constater que la plupart des tests conduisent à rejeter l'hypothèse de l'homogénéité des moyennes : sur 30 tests, 3 seulement sont non significatifs, 3 sont significatifs au niveau de probabilité $P = .05$ et les 24 autres sont significatifs au niveau de probabilité $P = .01$.

Ceci provient évidemment du grand nombre d'observations sur lequel sont basés les calculs de moyennes, ce qui nous permet de considérer comme significatives des différences même petites.

Pour l'utilisateur, le problème se posera alors de décider quelles sont les différences qu'il pourra considérer comme sans intérêt d'un point de vue pratique, bien que ces différences soient significatives.

Il nous semble qu'un critère de décision possible serait le rapport entre la différence maximum des moyennes au sein des sous-groupes d'une classification qualitative et l'écart-type intragroupe.

Si l'on considère par exemple les résultats scolaires totaux, on voit que les sous-groupes d'âge ont un écart-type de l'ordre de 8,8; les moyennes s'étalent de 65.88 à 68.38; leur différence, 2.50, représente donc les 28 centièmes de l'écart-type. Suivant que cette fraction est supérieure ou inférieure à une valeur fixée *a priori*, on pourrait conserver ou abandonner la classification en quatre groupes d'âge.

C'est dans cette perspective qu'il faut lire les figures 1 à 6. Chacune d'entre elles donne pour les différentes épreuves, et pour un critère de classification, les moyennes de ces épreuves rapportées à la moyenne générale et à l'écart-type intra-groupe ⁽²⁾.

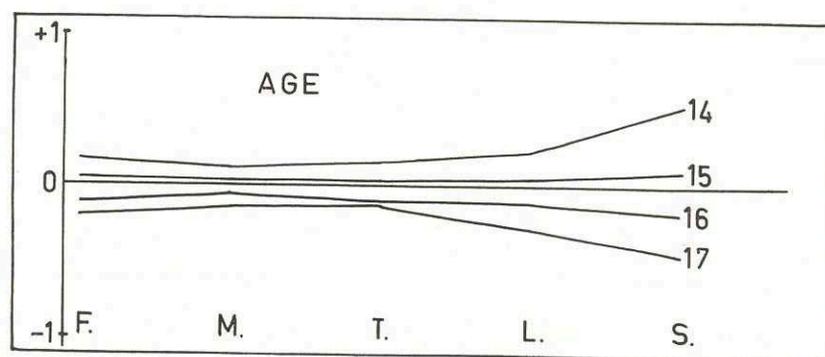


Fig. 1.

⁽²⁾ Les épreuves verbales étant normalement destinées à dépister les aptitudes littéraires sont désignées par L; les épreuves non verbales sont désignées par S (scientifiques).

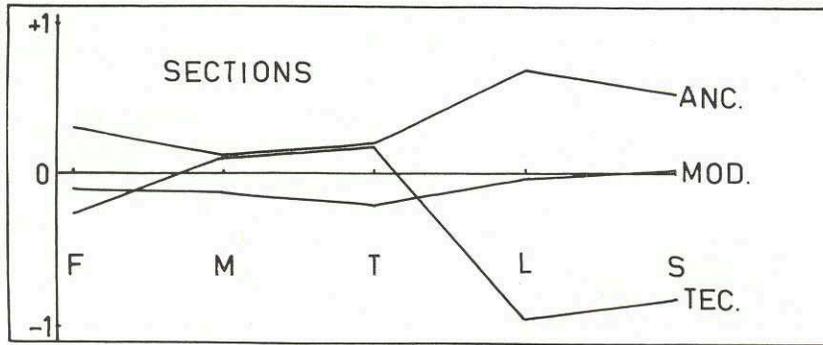


Fig. 2.

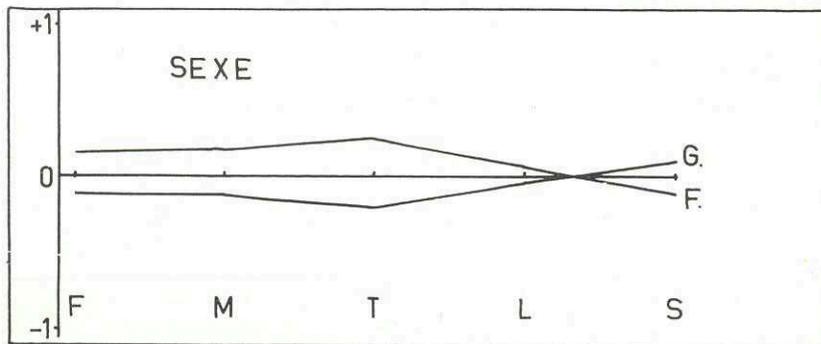


Fig. 3.

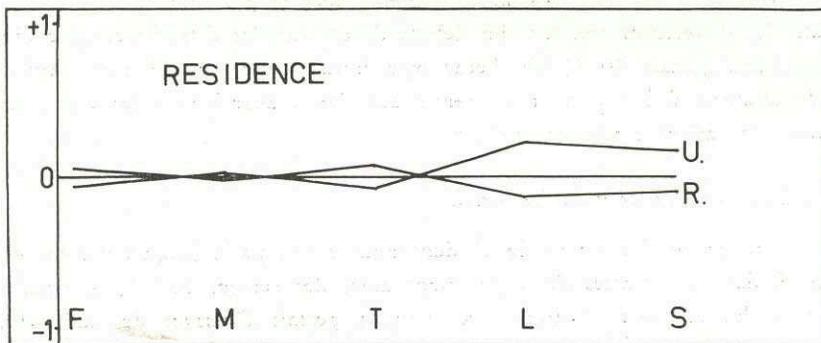


Fig. 4.

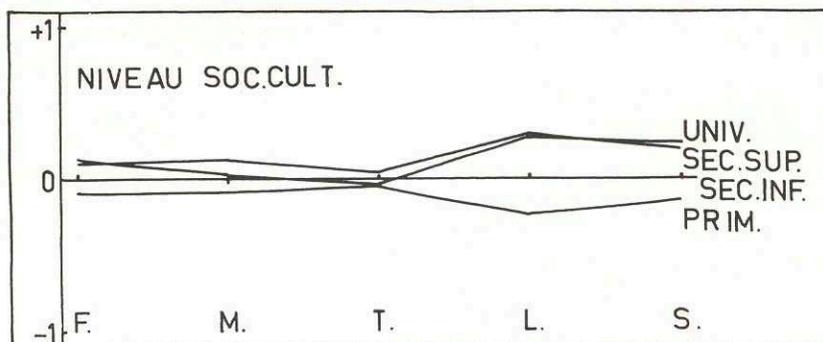


Fig. 5.

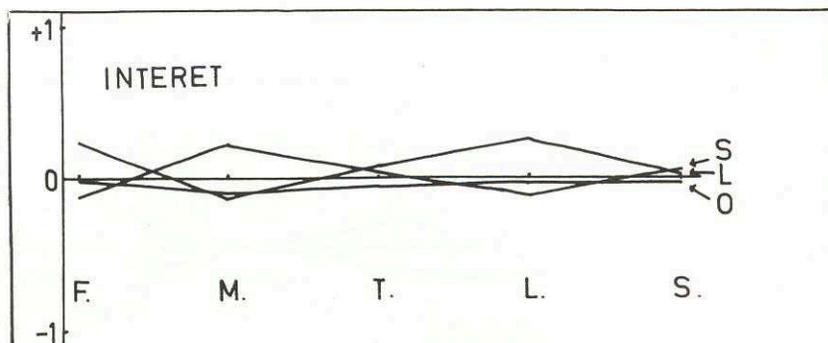


Fig. 6.

L'examen de ces diagrammes permet de constater que, dans l'ensemble, les différences sont plus importantes au niveau des tests qu'au niveau des résultats scolaires. D'autre part, les différences les plus importantes sont observées pour les tests NV aux différents âges et pour les tests V et NV dans les différentes sections. En dehors de ces cas, les différences observées n'excèdent jamais 50 % de l'écart-type intra-groupe, ce qui nous semble suffisamment faible pour nous sentir autorisé à abandonner les classifications qualitatives envisagées au départ.

II. Essais d'orientation différentielle.

Une partie des élèves de 4^e dont nous avons parlé jusqu'à présent est passée dans des classes de 3^e, relevant aussi des centres P.M.S., et nous a fourni des résultats scolaires. Ceci nous a permis d'essayer des méthodes de prédiction des résultats en 3^e.

Voici tout d'abord la répartition des effectifs suivant les différentes possibilités qui se présentent aux élèves. (La colonne « résidu » correspond aux élèves pour lesquels nous n'avons pas eu les résultats de 3^e).

TABLEAU II

| | 4 ^e Anciennes | 4 ^e Modernes | 4 ^e Techn. (1) | 4 ^e Techn. (2) | Total général |
|----------------------|--------------------------|-------------------------|------------------------------|------------------------------|------------------|
| 3 ^e L. G. | 169 | 9 | — | 2 | |
| 3 ^e L. M. | 134 | — | — | — | |
| 3 ^e L. S. | 55 | — | — | — | |
| 3 ^e Eco. | 3 | 164 | — | — | |
| 3 ^e Sc. | 22 | 144 | — | — | |
| 3 ^e Tech. | — | — | 79 | — | |
| Résidu | 330 | 698 | 177 | 201 | |
| Total | 713 | 1 015 | 250 | 203 | 2 181 |

(1) garçons uniquement

(2) filles uniquement (4^e familiales).

Comme on peut le constater, certains effectifs sont fort faibles et ne nous permettraient par d'établir des étalonnages ou des pronostics valables. Nous les ferons par conséquent passer au résidu.

Nous éliminerons d'autre part les élèves de 4^e familiale qui ne permettent aucun pronostic.

TABLEAU III

| | 4 ^e Anciennes | 4 ^e Modernes | 4 ^e Techniques |
|----------------------|--------------------------|-------------------------|---------------------------|
| 3 ^e L. G. | A = 169 | | |
| 3 ^e L. M. | B = 134 | | |
| 3 ^e L. S. | C = 55 | | |
| 3 ^e Eco. | | E = 164 | |
| 3 ^e Sc. | | F = 144 | |
| 3 ^e Tech. | | | H = 73 |
| Résidu | D = 355 | G = 707 | I = 177 |
| Total | $\alpha = 713$ | $\beta = 1015$ | $\gamma = 250$ |

Il nous reste alors le tableau III dans lequel figurent les effectifs et la lettre sous laquelle les groupes seront parfois désignés.

Exemple : le groupe A est constitué par 169 élèves qui sont passés de 4^e ancienne en 3^e latin-grec.

Nous avons vu précédemment que les variables qualitatives telles que le sexe, la résidence, etc. n'influençaient guère les moyennes. Nous avons fait l'hypothèse qu'il en était de même au sein des 3^{es}. Cette hypothèse a l'avantage d'éviter la multiplication des sous-groupes au détriment de leur effectif.

Nous fournissons dans le tableau IV les compositions des différents groupes ainsi que du tiers supérieur des groupes de 3^e au point de vue du

TABLEAU IV
*Contrôle des compositions
du point de vue des variables qualitatives (en pourcents)*

| Groupe | Sexe | | Résidence | | Niveau soc.-cult. | | Ages | | Questionnaire d'intérêt | | |
|---|------|----|-----------|----|----------------------|----|------|----|----------------------------|----|----|
| | G | F | U | R | N+ | N- | A+ | A- | L | O | S |
| A | 53 | 47 | 47 | 53 | 53 | 47 | 84 | 16 | 46 | 41 | 13 |
| A sup. | 42 | 58 | 37 | 63 | 51 | 49 | 93 | 7 | 40 | 40 | 19 |
| B | 70 | 30 | 49 | 51 | 48 | 52 | 87 | 13 | 14 | 44 | 42 |
| B sup. | 67 | 33 | 42 | 58 | 51 | 49 | 93 | 7 | 14 | 42 | 44 |
| C | 62 | 38 | 76 | 24 | 65 | 35 | 84 | 16 | 24 | 40 | 36 |
| C sup. | 35 | 65 | 70 | 30 | 65 | 35 | 95 | 5 | 25 | 40 | 35 |
| | 61 | 39 | 50 | 50 | 46 | 54 | 78 | 22 | 32 | 43 | 25 |
| Total 4 ^e Anc. (avant sélection) | 61 | 39 | 51 | 49 | 49 | 51 | 81 | 19 | 31 | 43 | 26 |
| E | 55 | 45 | 36 | 64 | 23 | 77 | 67 | 33 | 31 | 49 | 20 |
| E sup. | 38 | 62 | 39 | 61 | 20 | 80 | 75 | 25 | 26 | 52 | 21 |
| F | 60 | 40 | 37 | 63 | 31 | 69 | 67 | 33 | 14 | 31 | 55 |
| F sup. | 54 | 46 | 38 | 62 | 29 | 71 | 69 | 31 | 15 | 25 | 60 |
| G | 51 | 49 | 42 | 58 | 22 | 78 | 54 | 46 | 22 | 47 | 31 |
| Total 4 ^e Mod. (avant sélection) | 53 | 47 | 41 | 59 | 23 | 77 | 58 | 42 | 22 | 45 | 33 |
| H | 100 | 0 | 33 | 67 | 3 | 97 | 52 | 48 | 8 | 42 | 49 |
| H sup. | 100 | 0 | 35 | 65 | 4 | 96 | 58 | 42 | 12 | 46 | 42 |
| I | 100 | 0 | 19 | 81 | 8 | 92 | 44 | 56 | 6 | 33 | 61 |
| Total 4 ^e Tech. (avant sélection) | 100 | 0 | 23 | 77 | 6 | 94 | 46 | 54 | 7 | 36 | 58 |

sexe, de la résidence, du niveau socio-culturel (nous avons dichotomisé ce dernier en groupant ensemble, d'une part les niveaux universitaires et secondaires supérieurs, et d'autre part les niveaux secondaires inférieurs et primaires). Au point de vue des âges, A+ désigne les élèves d'âge normal ou avancé, A— désigne les retardés d'un ou deux ans. Les groupes correspondant au questionnaire d'intérêt sont désignés respectivement par L, 0 et S.

L'examen attentif du tableau IV permet de faire quelques constatations.

Fixons par exemple notre attention sur la variable « sexe ». Si l'on compare les groupes A, B, C (3^{es}) au groupe D (4^e), on voit que les garçons sont relativement plus nombreux en 3^e latin-math. et 3^e latin-sciences et relativement moins nombreux en 3^e gréco-latine qu'en 4^e ancienne. On peut aussi constater que parmi les meilleurs élèves de 3^e (1/3 supérieur), les filles sont relativement plus nombreuses que dans le groupe total correspondant de 3^e. De même, les garçons sont relativement plus nombreux en 3^e scientifique et économique qu'en 4^e moderne, mais les filles sont relativement plus nombreuses dans les groupes supérieurs de 3^e moderne que dans les 3^{es} modernes entières.

Pour la variable « résidence », on voit que dans les sections anciennes, et à l'exception de la 3^e latin-sciences entière, les ruraux sont proportionnellement plus nombreux en 3^e qu'en 4^e, et dans les groupes supérieurs que dans les groupes totaux de 3^e. En moderne, il y proportionnellement plus de ruraux en 3^e qu'en 4^e, mais les urbains réussissent sensiblement mieux.

Au point de vue de la variable « âge », il y a, d'une manière générale, proportionnellement plus de « normaux » et d'« avancés » en 3^e qu'en 4^e, et en tête de promotion, que dans les classes totales.

Les variables « niveau socio-culturel » et « questionnaire d'intérêts » ne semblent pas fournir de conclusions très évidentes.

Prédiction des résultats de 3^e.

Nous avons tout d'abord essayé de prédire les trois résultats scolaires à partir des cinq résultats de 4^e, et cela pour chacune des sections de 3^e. Ce travail impliquait l'analyse de matrices de corrélation de dimensions allant jusqu'à 5 sur 5. Il s'agit d'en extraire 75 coefficients de régression partiels. Ce travail a été fait par l'ordinateur du laboratoire de calcul, et a malheureusement mené à des résultats inutilisables. Aucune cohérence ne se manifestait dans la grandeur des coefficients de régression, ni même dans leur signe. Nous avons pensé que ce résultat était dû à des interférences entre

variables (le résultat total, par exemple, contient le résultat en français et en mathématiques), et d'autre part certains coefficients de corrélation sont affectés d'erreurs d'échantillonnage qui risquent de faire perdre leur sens aux résultats.

Devant cet état de choses, nous nous sommes tenu le raisonnement suivant :

- 1) le prédicteur « total 4^e » étant très bien représenté par les prédicteurs « français 4^e » et « mathématiques 3^e », nous nous contentons de ces deux derniers comme prédicteurs scolaires;
- 2) les prédicteurs « tests » étant très mal corrélés avec les critères de 3^e, nous ne les utilisons pas dans ce moment-ci du travail;
- 3) la multiplicité des critères étant plutôt un handicap, nous nous limiterons à la prédiction de « total 3^e ».

Dans cette perspective, on pourrait concevoir une orientation faite de la manière suivante : lorsqu'un élève de 4^e ancienne se présente à l'orienteur, celui-ci, à l'aide des résultats en français et en mathématiques de l'élève, peut prédire le résultat total que celui-ci obtiendra

- a) s'il passe en 3^e gréco-latine;
- b) s'il passe en 3^e latin-mathématiques;
- c) s'il passe en 3^e latin-sciences.

Le même travail peut être fait avec les élèves de 4^e moderne passant soit en 3^e économique, soit en 3^e scientifique.

Nous pouvons également, mais ici il n'y a plus lieu de faire un pronostic différentiel, prédire le résultat total en 3^e technique pour les élèves fréquentant la 4^e technique.

Ce procédé nous a amené à établir des diagrammes qui ont été obtenus de la manière suivante :

Considérons les équations permettant de prédire T_{3LG} , T_{3LM} et T_{3LS} pour les élèves de 4^e ancienne :

$$T_{3LG} = .3013 F + .2791 M + 27,8856$$

$$T_{3LM} = .6118 F + .2972 M + .4596$$

$$T_{3LS} = .3867 F + .2381 M + 19,9384$$

Considérons ensuite, dans un système d'axes (M, F), les droites d'équations :

$$T_{3LG} = T_{3LM}$$

$$T_{3LM} = T_{3LS}$$

$$T_{3LG} = T_{3LS}$$

Il est facile de montrer que ces droites, que nous appellerons séparatrices, sont concourantes, et qu'elles déterminent trois régions :

- 1) une région où $T_{3LG} > (T_{3LM}, T_{3LS})$
- 2) une région où $T_{3LM} > (T_{3LG}, T_{3LS})$
- 3) une région où $T_{3LS} > (T_{3LG}, T_{3LM})$

La seule séparatrice qui nous intéresse est celle qui sépare LG et LM, les deux autres tombant en dehors de la partie utile du diagramme. Son équation est :

$$.3105 F + .0181 M = 27,2260.$$

Dans chacune des régions il est possible de tracer des courbes de niveau à T constant. Il est donc théoriquement possible de faire l'orientation d'un élève de 4^e ancienne ayant obtenu M_0 et F_0 en cherchant le point de coordonnées correspondantes; la région dans laquelle tombe le point indiquera la section de 3^e vers laquelle il faut l'orienter pour être assuré qu'il obtiendra en 3^e un résultat total supérieur à celui qu'il obtiendrait dans toute autre section. La courbe de niveau indiquera la valeur de ce résultat (fig. 7).

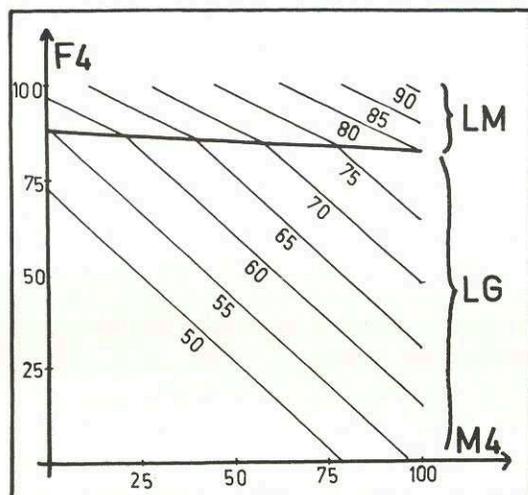


Fig 7

Le même travail effectué à propos des élèves de section moderne nous donne, pour les équations de prédiction :

$$T_{3Eco} = .3706 F + .2331 M + 21,8694$$

$$T_{3Sc} = .4715 F + .2447 M + 13,8761$$

et pour la séparatrice Eco-Sc :

$$.1009 F + .0116 M = 7,9933 \quad (\text{fig. 8})$$

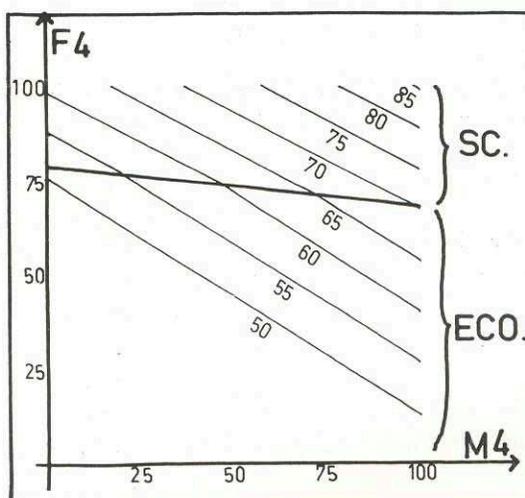


Fig 8

Au terme de ce premier essai, on a été amené à faire les remarques suivantes :

1) Le résultat « total 3^e latin-sciences » sera toujours inférieur au résultat « total 3^e gréco-latine » et « total 3^e latin-math. ». Ceci voudrait dire qu'en utilisant la méthode d'orientation suggérée ci-dessus, il ne faudrait jamais envoyer un élève en 3^e latin-sciences, quels que soient ses résultats en 4^e !

2) Pour un résultat donné en 4^e mathématiques, on peut constater que ce sont les élèves qui ont le meilleur résultat en français qui obtiennent le meilleur résultat en 3^e latin-math. Ceci voudrait dire par exemple qu'un élève médiocre en mathématiques et fort en français réussira mieux en 3^e latin-math, qu'en 3^e gréco-latine ! Une constatation identique peut être faite au niveau de la section moderne.

Nous pensons que ces résultats, apparemment paradoxaux, peuvent s'expliquer par le niveau moyen actuel plutôt élevé en latin-math, et plutôt bas en latin-sciences.

Analyse factorielle d'Hotelling.

Devant la situation décrite dans les paragraphes précédents, et en particulier devant les remarques qui terminent le dernier paragraphe, nous avons

décidé l'abandon du critère; il nous a semblé qu'il avait été utilisé au maximum et qu'il ne fallait pas exagérer son importance à cause de circonstances contingentes telles que le niveau actuel moyen des sections, l'attitude différentielle des professeurs vis-à-vis des sections, facteurs qui sont difficiles sinon impossibles à évaluer à l'aide des données dont nous disposons.

Nous nous sommes proposé alors de faire une analyse factorielle d'Hotelling sur les quatre variables de 4^e (français, mathématiques, tests littéraires, tests scientifiques) et cela, d'une part pour la section ancienne, et d'autre part pour la section moderne.

Rappelons en quelques mots en quoi consiste la méthode.

On se propose de remplacer les quatre variables ci-dessus, que nous appellerons « variables initiales » et qui sont corrélées entre elles, par quatre variables indépendantes, dites « variables d'Hotelling », et qui sont des combinaisons linéaires des variables initiales. On peut évaluer la « fraction de la variance totale » expliquée par chacune des variables d'Hotelling. D'autre part, pour des questions de facilité d'utilisation, il est possible de calculer les coefficients de telle manière que l'on ait, pour les variables d'Hotelling, la moyenne et l'écart-type que l'on désire. Nous avons pensé commode de choisir la moyenne 50 et l'écart-type 10, qui sont les valeurs correspondant au classique « T-score ».

Avant de donner les équations qui permettront de calculer les variables d'Hotelling à partir des variables initiales, faisons quelques remarques :

1) La première variable d'Hotelling est, comme on pouvait s'y attendre, une variable d'aptitude générale se présentant sous la forme d'une combinaison linéaire additive des variables initiales. Cette variable explique 49 % de la variance totale dans le groupe « anciennes » et 47 % dans le groupe « modernes ».

2) La deuxième variable est une variable d'attitude « examens » contre « tests ». Dans l'équation, le français et les mathématiques apparaissent avec un même signe, opposé au signe commun des tests littéraires et scientifiques. Cette variable explique 22 % de la variance dans le groupe « anciennes » et 24 % dans le groupe « modernes ».

3) La troisième variable d'Hotelling mesure une aptitude différentielle envers « activités scientifiques » et « activités littéraires »; le français et le test littéraire apparaissent dans l'équation avec le signe —; les mathématiques

et le test scientifique avec le signe +. Cette variable explique 18 % de la variance dans le groupe « anciennes » et 18 % également dans le groupe « modernes ».

4) La dernière variable d'Hotelling explique respectivement 10 et 11 % de la variance totale, mais n'est pas susceptible d'une interprétation pédagogique. Elle peut être considérée comme due aux erreurs d'échantillonnage.

5) La similitude des grandeurs trouvées dans les deux analyses d'Hotelling, tant du point de vue des fractions de variance que des grandeurs et des signes des coefficients, est un résultat encourageant qui permet de croire que les erreurs d'échantillonnage n'ont pas trop influencé les analyses.

6) Les signes des variables ont été arrangés de telle sorte que les valeurs élevées correspondent :

- pour la première variable, à une aptitude générale élevée,
- pour la deuxième variable, à une attitude orientée vers les examens plutôt que vers les tests,
- pour la troisième variable, à une attitude orientée plutôt vers les activités scientifiques que vers les activités littéraires.

Equations pour le groupe « anciennes ».

$$I = .40 F + .27 M + .10 L + .18 S - 30,87$$

$$II = .48 F + .30 M - .18 L - .36 S + 67,79$$

$$III = -.28 F + .22 M - .26 L + .43 S + 68,14$$

Equations pour le groupe « modernes ».

$$I = .38 F + .25 M + .11 L + .22 S - 27,46$$

$$II = .47 F + .34 M - .17 L - .33 S + 58,35$$

$$III = -.52 F + .33 M - .21 L + .43 S + 65,91$$

Essai de validation des variables d'Hotelling.

Il ne faut pas espérer que les variables d'Hotelling soient plus valides que les pronostics faits à l'aide de la régression multiple, puisque, par définition, ces derniers représentent un optimum dans le sens des moindres carrés. Nous avons pensé toutefois intéressant de mettre en relation les variables d'Hotelling et le critère en essayant de contourner les défauts du critère auquel il a été fait allusion précédemment.

Nous avons procédé de la façon suivante :

1) Au niveau du critère (résultats total de 3^e), nous avons partagé les observations en trois groupes : les « meilleurs », les « moyens », les « moins bons », et ceci pour chacune des cinq sections de 3^e. En fait, pour chaque section, nous avons pris le tiers supérieur, le tiers central et le tiers inférieur de chaque groupe.

TABLEAU V

| | Variable I | | Variable II | | Variable III | |
|-------------------------|------------|----------|-------------|----------|--------------|----------|
| | Moy. | Ec. type | Moy. | Ec. type | Moy. | Ec. type |
| 3 ^e LG int. | 47,1 | 7,6 | 49,4 | 10,3 | 47,0 | 9,1 |
| LG moy. | 52,0 | 6,5 | 51,2 | 9,8 | 47,8 | 10,4 |
| LG sup. | 59,5 | 9,6 | 56,2 | 9,6 | 46,7 | 8,5 |
| 3 ^e LM inf. | 50,8 | 6,7 | 46,7 | 6,6 | 54,3 | 9,6 |
| LM moy. | 55,2 | 5,9 | 50,0 | 9,9 | 53,3 | 9,0 |
| LM sup. | 64,1 | 6,1 | 53,1 | 7,8 | 53,5 | 9,8 |
| 3 ^e LS inf. | 44,8 | 7,2 | 49,3 | 11,1 | 53,6 | 7,1 |
| LS moy. | 47,6 | 6,0 | 49,5 | 8,1 | 50,4 | 8,8 |
| LS sup. | 57,3 | 6,5 | 53,5 | 7,1 | 53,4 | 9,1 |
| 3 ^e Ec. inf. | 49,8 | 7,5 | 49,0 | 8,8 | 48,1 | 8,9 |
| Ec. moy. | 51,1 | 7,0 | 50,0 | 9,7 | 48,5 | 8,6 |
| Ec. sup. | 56,5 | 8,2 | 56,3 | 8,3 | 49,9 | 7,9 |
| 3 ^e Sc. inf. | 51,8 | 9,5 | 48,8 | 9,9 | 53,2 | 7,2 |
| Sc. moy. | 56,6 | 6,8 | 52,8 | 8,7 | 52,5 | 9,1 |
| Sc. sup. | 62,0 | 8,2 | 53,6 | 8,3 | 50,5 | 9,2 |

2) Pour chacun de ces groupes, et dans chacun des sous-groupes, nous avons calculé la moyenne et l'écart-type de chacune des variables d'Hotelling. Nous donnons ces valeurs dans le tableau V. Pour simplifier davantage, nous pouvons convenir de caractériser les élèves, non plus par ses trois valeurs des variables d'Hotelling, mais bien par un signe « plus » ou « moins » suivant que la valeur de la variable d'Hotelling est supérieure ou inférieure à la moyenne 50. Nous pouvons ainsi observer des « types » d'élèves du genre : (+++), (++-), (+-+), etc.

On peut alors constater que les meilleurs élèves de gréco-latine et d'économique ont une configuration $(++-)$, c'est-à-dire qu'en moyenne ces élèves sont de bons élèves, orientés vers les examens, orientés vers les activités littéraires. D'autre part, les meilleurs élèves de 3^e latin-math., latin-sciences et scientifiques ont une configuration moyenne $(+++)$. Dans l'ensemble, il s'agit donc d'élèves doués, orientés vers les examens, orientés vers les activités scientifiques.

Les élèves « moyens » des différentes sections ont une configuration semblable à celle des bons élèves, mais d'une façon moins marquée. Une exception cependant : les élèves moyens de 3^e latin-sciences ont une configuration moyenne $(-++)$. Toutefois, si l'on mélange le groupe supérieur et le groupe moyen de cette section, la configuration moyenne est malgré tout $(+++)$.

Les mauvais élèves de chaque section se caractérisent essentiellement par une première variable d'Hotelling inférieure à 50 (sauf toutefois en section scientifique, où elle est à peine supérieure à 50); la deuxième variable d'Hotelling est systématiquement inférieure à 50 pour ces mauvais élèves.

En résumé, la cohérence de ces résultats est encourageante et permet de penser que l'utilisation judicieuse des variables d'Hotelling peut effectivement aider l'orienteur dans sa tâche.