ÉTUDE SUR LES DIFFICULTÉS DES ÉTUDIANTS EN STATISTIQUE

Mme LEBOUTET *enseigne la Statistique aux étudiants de l'Institut de Psychologie de l'Université de Paris. Le cycle des études à cet Institut est de quatre à cinq ans au minimum.*

Il comporte un enseignement obligatoire de Statistique qui comprend des travaux pratiques. Au public d'étudiants préparant des diplômes spécialisés ouvrant accès à diverses carrières, s'ajoutent des médecins, des ingénieurs, des officiers, des ecclésiastiques, etc...

Nous remercions Mme LEBOUTET *de nous avoir autorisés à reproduire cette étude destinée au* Bulletin de l'Institut de Psychologie.

Parmi les nombreuses matières inscrites à la préparation des diplômes, la Statistique semble la plus redoutée des étudiants en Psychologie de l'Université de Paris.

Dans le but de chercher les raisons de ce fait, nous avons effectué une enquête auprès des étudiants de seconde année. Les résultats nous permettent de préciser la nature des diverses difficultés rencontrées par les élèves, difficultés déjà plus ou moins dépistées au cours de plusieurs années de séances de travaux pratiques.

L'enquête a été effectuée à l'aide des moyens suivants :

I. Un *questionnaire* soumis aux étudiants consistant en une analyse dirigée de leurs difficultés. En voici le texte :

— Eprouvez-vous des difficultés en Statistique ?

— *Quelles sont vos principales difficultés ?*

1) Difficultés mathématiques ? Vous manquez de connaissances, exemples : fractions, puissances, équations.

Exemples personnels.

2) Les notions ou principes vous paraissent difficiles parce qu'ils ne correspondent à rien dans votre esprit ; exemple : corrélation, degrés de liberté, échantillonnage, probabilité.

Exemples personnels.

3) Le vocabulaire est difficile ou ambigu.

Exemples personnels.

4) Les techniques vous paraissent difficiles à appliquer : épreuves de signification, analyse factorielle, utilisation des tables.

Exemples personnels.

5) Autres causes de difficultés.

— *Les problèmes.*

Devant un problème :

1. Vous ne comprenez pas le texte. Exemples personnels précis.

2. Vous ne voyez pas comment appliquer le cours. Exemples personnels précis.

3. Le texte vous paraît abstrait et vous préféreriez travailler sur des exemples concrets. Exemples personnels précis.

4. Autres causes de difficultés vis-à-vis des problèmes.

5. Souhaits et vœux concernant les travaux pratiques.

II. Une *épreuve de Mathématiques,* comportant une trentaine de questions à résoudre, et faisant appel aux connaissances mathématiques indispensables à l'étudiant en Statistique, c'est-à-dire :

— opérations sur les fractions numériques ou littérales ;

— opérations sur les identités remarquables ;

— construction d'une droite définie par son équation ;

— résolution d'une double inégalité du premier degré ;
— résolution d'une équation du second degré ;
— opérations sur les radicaux ;
— calcul par interpolation linéaire ;
— opération sous le signe « somme ».

III. Dix questions sur le *vocabulaire statistique :* définitions à fournir sans employer de formules.

Résultats de l'enquête

I. Difficultés mathématiques : Résultats de l'épreuve de Mathématiques.

La difficulté la plus évidente avec laquelle les étudiants sont aux prises concerne le calcul mathématique. Elle constitue un très lourd handicap pour la réussite à l'examen.

Bien que les étudiants de seconde année examinés aient déjà dû se remettre depuis un an à l'utilisation des Mathématiques après le Baccalauréat, il semble que la plupart de leurs connaissances scolaires dans cette discipline aient presque complètement disparu. Le rappel s'effectue trop péniblement pour que ces notions soient disponibles au bon moment et utilisables.

Ce qui frappe le plus dans le résultat du test, n'est pas la somme des erreurs commises, mais plutôt l'absence des réponses, car la plupart des questions ne sont pas abordées.

Les résultats de l'épreuve ont été les suivants : 1er groupe : effectif 22, moyenne 8,69 sur 20 ; 2e groupe : effectif 41, moyenne 8,10 sur 20.

Remarquons que si, dans chaque groupe, nous éliminons les quelques étudiants français titulaires du Baccalauréat de Mathématiques Elémentaires ou de Certificats de Licence scientifiques (au nombre de 7), les moyennes précédentes, affectant les bacheliers littéraires, qui constituent la majorité, s'abaissent aux chiffres suivants : 7,5 et 7,6.

Ces résultats sont particulièrement faibles et un gros effort personnel doit être fait par les étudiants pour les améliorer.

II. Difficultés en Statistique.

Les questionnaires de difficultés en Statistique et pour le Vocabulaire ne nous fournissent pas, et ne peuvent pas nous fournir, de résultats numériques aussi immédiats, en dehors de l'inventaire des difficultés principales, du relevé des fréquences de ces réponses, du relevé des erreurs ou omissions pour le vocabulaire.

Nous énoncerons séparément les conclusions concernant chacun de ces deux questionnaires.

A) *Questionnaire des difficultés :* De nombreuses difficultés ont été signalées en dehors de celles déjà proposées par le questionnaire. Citons, par exemple, les limites de confiance, l'ajustement, la droite de régression, la méthode des moindres carrés.

Les plus fréquentes sont suscitées par les tests de signification, la notion d'échantillon, les degrés de liberté, l'analyse factorielle, l'utilisation du chi carré. Ce sont en quelque sorte les noyaux de résistance contre lesquels viennent buter les étudiants.

Il nous semble alors possible de découvrir sous la variété des réponses conscientes et disparates, le petit nombre de réponses essentielles susceptibles de les expliquer et de les unifier. Mais ces réponses ne peuvent se présenter à l'esprit des étudiants, car s'ils pouvaient appréhender l'essence même de leurs difficultés, ces difficultés s'évanouiraient ou reculeraient sur un plan d'abstraction ou d'axiomatisation plus élevé ; elles ne seraient plus alors une gêne pour la manipulation des statistiques élémentaires.

Les difficultés majeures paraissent provenir d'une incompréhension de l'esprit même de la Statistique, confondu simplement avec celui des Mathématiques, sans doute parce que ces deux sciences opèrent sur des nombres. Il semble, en effet, que, malgré leurs faiblesses en calcul, les étudiants aient assimilé quelque peu l'esprit des Mathématiques, sinon leurs techniques. Ils ont acquis de cette science une idée d'absolu qui fait que des résultats numériques ne peuvent être pour eux que vrais ou faux. Cela les conduit souvent à formuler des conclusions péremptoires comme : « Ainsi, l'hypothèse nulle est fausse et nous pouvons affirmer que... », ou des réponses qui sont énoncées sans indication du seuil de probabilité auquel l'on se place.

Mais l'esprit de la Statistique est différent de celui des Mathématiques. Du fait que les résultats proviennent toujours d'échantillons prélevés dans une population donnée et que ces échantillons présentent une très grande variabilité, les données recueillies ne peuvent jamais être étudiées, comme en Mathématiques, avec une certitude totale. Il y a toujours, en ce qui les concerne, une certaine marge d'incertitude qui limite l'effort de généralisation et d'induction. Remarquons qu'il en est de même pour toute science expérimentale.

Si l'on veut déterminer la valeur d'une statistique, il n'est jamais possible de le faire avec une rigueur mathématique. Il faut se contenter de connaître un intervalle à l'intérieur duquel peut se trouver la valeur cherchée, encore faut-il y associer l'indication de la probabilité selon laquelle l'assertion est vraie, car il y a un certain nombre de chances sur cent pour qu'elle soit fausse. On ne peut donc jamais être sûr de la vérité d'une affirmation concernant une donnée calculée sur un échantillon.

Si l'on veut, en outre, comparer une statistique déterminée sur un échantillon à une donnée correspondante sur un autre échantillon réel ou théorique, en établissant le caractère valable ou non d'un « écart significatif », on retrouve la même incertitude et l'on ne peut émettre de conclusion qu'en indiquant le nombre de chances d'erreur ; encore faut-il disposer de tables convenables, établies à partir de distributions que l'on a pu étudier. L'affinement des techniques statistiques ne permet en général que de cerner l'erreur ou sa probabilité, non de la déterminer. Nous sommes donc bien loin de l'absolue vérité mathématique.

Le danger est aggravé pour les étudiants du fait que, en Statistique, comme dans beaucoup de sciences, on essaie d'adapter les données de l'expérience à un modèle mathématique. Ce peut être une distribution normale, ou un schéma linéaire comme pour l'analyse de la variance ou l'analyse factorielle. La sécurité intellectuelle que donne l'utilisation des Mathématiques peut faire oublier que ce schéma risque de s'ajuster fort mal à la réalité, d'où un emploi abusif des Mathématiques.

Inversement, les notions mathématiques peuvent être insuffisamment sollicitées, en particulier le calcul des probabilités auquel le développement de la Statistique doit beaucoup ; mais qui n'est pas à la portée des étudiants, si bien que ceux-ci, ignorant l'origine des formules, des tables, ou la justification de propriétés préliminaires conditionnant le bon usage de certaines techniques (par exemple homogénéité de la variance pour l'analyse de la variance) sont excusables d'utiliser techniques, formules et tables comme des recettes, ce qui crée un malaise chez les plus consciencieux. Nous ne voyons pas de remède à cette situation.

En conclusion, nous dirons que la difficulté majeure vient de l'incompréhension de l'esprit de la Statistique, confondu avec celui des Mathématiques dont les techniques propres ne sont d'ailleurs pas suffisamment maîtrisées pour permettre d'aller au fond des problèmes purement statistiques.

B) *Le Vocabulaire statistique :* Une difficulté non négligeable réside dans le vocabulaire tout nouveau que l'étudiant doit assimiler et qui concerne des notions qui ne lui sont nullement familières. Nous nous sommes aperçus que, d'une année à l'autre, les mêmes mots, et, au-delà, les mêmes notions, soulevaient toujours les mêmes difficultés, d'où le questionnaire sur le vocabulaire statistique.

Les définitions inexactes ou manquantes concernent principalement les mots : fréquence, erreur-type, paramètre, estimation. Mais ces lacunes ne sont que des exemples parmi beaucoup d'autres relevés au cours des séances de travaux pratiques.

Nous demandons donc aux étudiants un effort particulier pour remédier à ces déficiences.

Nous leur proposerons ultérieurement, dans le *Bulletin de Psychologie*, un Lexique de la Statistique. Il sera centré sur la compréhension des éléments essentiels de Statistique, et fournira, avec la définition des termes, des renseignements sur les difficultés de leur emploi.

*
**

Nous terminerons cette étude par des indications pratiques destinées à favoriser la résolution des difficultés relevées.

1. Révision des programmes d'algèbre de Troisième et de Seconde, faite individuellement. Il est impossible de s'attarder à ces questions au cours d'une heure de travaux pratiques par quinzaine, pour chaque groupe.

2. Etude spéciale du vocabulaire statistique à l'aide des cours et du lexique à paraître.

3. Participation à des séances de travaux pratiques organisées à l'Institut Pédagogique National (ex-Musée Pédagogique) les jours suivants : samedis : 18 janvier, 1er et 15 février, 1er, 15 et 29 mars 1958, de 10 heures à 12 heures, Salle Paul-Langevin, 1er étage (pour les étudiants de seconde année).

Mme L. LEBOUTET (*janvier 1958*).

Références bibliographiques

G. DARMOIS : L'enseignement du Calcul des probabilités et des méthodes statistiques, *Bulletin de la Société Française de Pédagogie*, n° 107, 1954, et *Cahiers Pédagogiques pour l'Enseignement du Second Degré*, 2e année, n° 3, 15 novembre 1955.