JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MARIE-LOUISE DUFRÉNOY

Quelques aspects statistiques du concours général de français aux États-Unis

Journal de la société statistique de Paris, tome 90 (1949), p. 459-464

<http://www.numdam.org/item?id=JSFS_1949__90__459_0>

© Société de statistique de Paris, 1949, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Quelques aspects statistiques du concours général de français aux États-Unis.

L'Association américaine des Professeurs de français (A. A. T. F.) organise annuellement un concours auquel peuvent participer les élèves des écoles secondaires publiques ou privées (High Schools) au cours de leur première, deuxième, troisième ou quatrième année d'étude de la langue française.

Le Concours, auquel 36.000 élèves de 1075 écoles secondaires ont pris part en 1949, est organisé selon le principe de la sélection par échelons. On peut comparer la structure du système à celle d'une pyramide : au sommet se trouve le Comité national chargé de choisir les lauréats nationaux. Les copies jugées par le Comité national lui sont transmises par huit « Chairmen » régionaux représentant les huit divisions établies arbitrairement sur l'ensemble du territoire des États-Unis pour les besoins du Concours. Chaque région comprend un certain nombre de divisions administratives plus restreintes appelées « Chapitres » ayant chacune à sa tête un « Chapter Chairman » choisi à l'avance pour administrer le concours dans le domaine de sa juridiction.

Chaque « Chapter Chairman » reçoit, à une date fixée, les inscriptions de toutes les écoles secondaires de son ressort qui désirent participer au concours.

Le graphique I indique le classement des quarante-quatre « Chapitres » des États-Unis par ordre d'importance numérique quant au nombre d'inscriptions dans les différentes catégories I, II, III, IV, correspondant aux première, deuxième, troisième et quatrième années d'étude de la langue française.

Pour chaque catégorie I, II, III, IV, une série de questions est préparée par un comité désigné à cet effet. Chaque « Chapter Chairman » reçoit du « National Chairman » les feuilles d'examens qu'il transmet sous pli cacheté au directeur de chacune des écoles inscrites pour le concours; au jour fixé, chaque élève reçoit une feuille d'examen et répond aux questions. Le professeur de chaque groupe corrige les feuilles selon les instructions reçues et indique sur chaque copie le « pourcentage » de points obtenu; il envoie pour chaque catégorie la feuille ayant reçu le plus haut pourcentage au « Chapter Chairman ». Celui-ci reclasse les copies de chaque catégorie et transmet celle qui a obtenu le plus haut pourcentage au « Chairman régional », qui, à son tour, transmet, pour chaque catégorie, la copie ayant reçu le plus haut pourcentage, au Comité national.

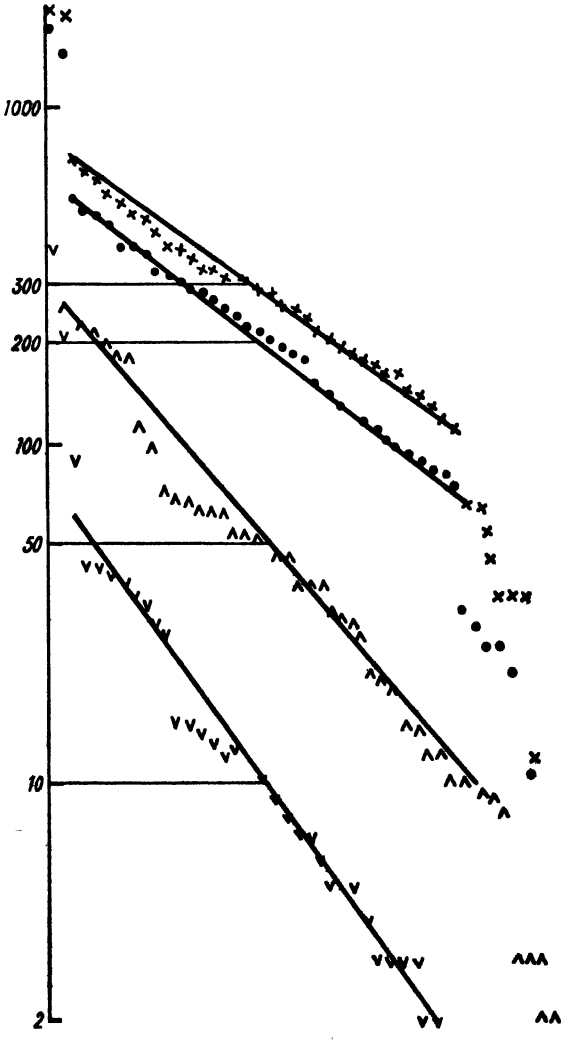
Nous choisirons comme exemple, pour l'étude statistique qui va suivre, le Chapitre de Californie, qui se classe cinquième dans l'ensemble des États-Unis (graphique 1) avec 1.500 concurrents, appartenant à 32 écoles et pour lequel les inscriptions dans les catégories I, II, III et IV se répartissent comme il est indiqué dans le graphique 2.

Nous étudierons la distribution des « pourcentages de points » dans la catégorie I, qui est non seulement la plus nombreuse mais celle qui représente la première orientation des jeunes Américains vers l'étude du français.

Le graphique 3 représente comparativement les distributions de pourcentages de points dans la population des vingt-neuf copies adressées au Chairman du Chapitre de Californie, comme étant chacune la meilleure dans la catégorie I

pour chacune des écoles de Californie participant au concours dans cette catégorie, et les distributions de pourcentages de points parmi les copies des différents élèves de catégorie I participant au concours dans chacune des trois écoles différentes.

Chacune de ces distributions s'obtient en classant les feuilles d'examen par



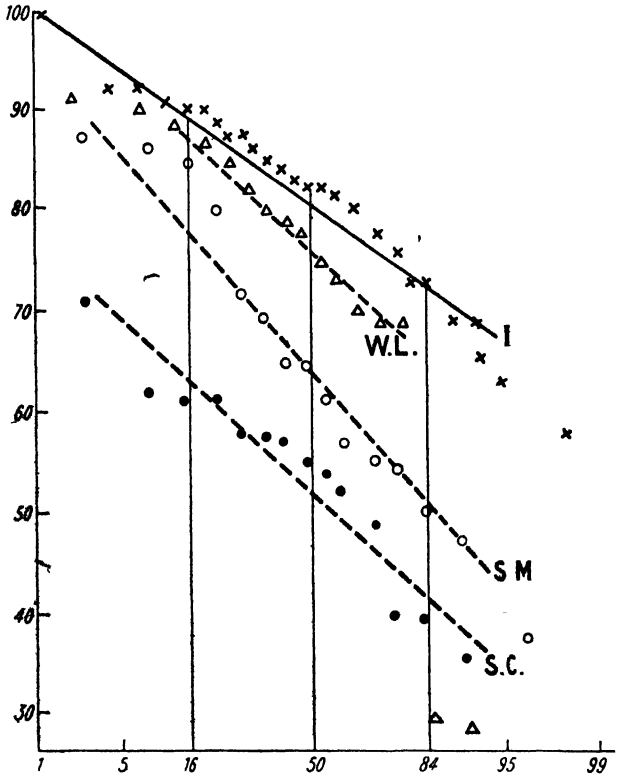
Graphique 1 — Droites de régression obtenues en portant en ordonnées, sur échelle logarithmique, les « fréquences d'élèves inscrits dans chaque catégorie » (I, II III, IV) pour chacun des 44 « chapitres » de l'Association américaine des Professeurs de français Les chapitres étant classés de gauche à droite, sur l échelle des abscisses, par importance numérique décroissante, le Chapitre de Californie se situe au 5e rang c est celui qui sera pris comme exemple pour l analyse statistique Les situations de la droite supérieure, représentant le « Français I », et des droites situées au dessous (correspondant aux Français II, III et IV) témoignent du « taux de mortalité » du Français dans les High Schools Par exemple pour 34 des Chapitres, les fréquences d'inscriptions se situent, entre 692 et 110, autour de la moyenne géométrique 300 pour le Français I, elles se situent autour des moyennes 200, 50 et 10 pour les Français II, III, IV Les différences de pentes montrent d'ailleurs que le taux de mortalité varie grandement d une école a l'autre

ordre décroissant de pourcentage de points, puis en portant successivement chaque pourcentage de points à son niveau sur l'échelle arithmétique des ordonnées et à son rang sur l'échelle de probabilité normale en abscisses.

Cette opération peut se faire rapidement de façon semi-automatique en

employant la règle à calcul multiple réalisée à cet effet sur le principe suivant (1).

Étant donné que dans une distribution normale de 100 valeurs, 15 représentent la « tête » dépassant la valeur moyenne de plus de σ et 16 représentent la « queue », manifestant vis-à-vis de la moyenne d'un déficit correspondant au moins à σ, nous pouvons aisément nous représenter combien de valeurs seront comprises dans la « tête » et la « queue » que délimitent dans toute popu-



Graphique 2. — Les 32 écoles participant aux Concours en Californie ayant été classées quant au nombre des élèves de la catégorie I, et ce nombre étant porté en ordonnée sur échelle logarithmique, chacune à son rang, on obtient une série de points (x) par lesquels on peut faire passer la droite de régression (trait continu) On porte alors au rang de chaque école, en abscisse, la fréquence des élèves participant au concours dans les catégories II (point noir) et III (triangle). On note que, pour certaines écoles, les ordonnées diffèrent peu d'une catégorie à l'autre, tandis que pour d'autres écoles, une diminution considérable de la fréquence, de I à II, puis à III, témoigne d'un « taux de mortalité élevé »

lation, au-dessus et au-dessous de la moyenne la valeur $\pm \sigma$; à titre d'exemple, citons pour les cas de distributions de 3, 6, 9, 15, 21, 27... valeurs, les rangs occupés par les positions 16.7 %, 50 % et 83.3 % sur l'échelle de probabilité normale :

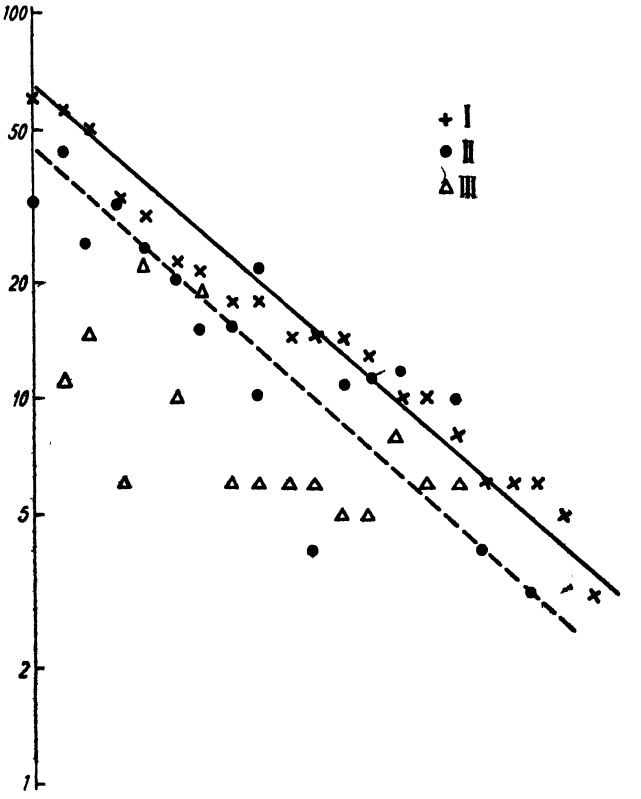
	16.7	50	83.3
3	1er	2e	3e
9	2e	5e	8e
15	3e	8e	13e
21	4e	11e	18e
27	5e	14e	23e

En d'autres termes, les trois valeurs d'une population normale définissent la

(1) Dufrénoy et F. M. Goyan, « A Graphical Calculator for Statistical Analysis », *J. Am. Pharmac. Assoc.* 36(10) : 309-314, 1947.

position de la moyenne et, approximativement celles de la moyenne $\pm \sigma$; dans des populations de 9, 15, 21, 27 valeurs, la position de la moyenne $+ \sigma$ est approximativement définie par celle de la 2e, 3e, 4e, 5e valeur.

L'échelle de probabilité normale, symétrique par rapport à la valeur de 10 %, porte vers la gauche les valeurs décroissantes de pourcentages, 40 %, 30 %, 20 %... 1 %... et vers la droite les valeurs croissantes 60 %, 70 %, 80 %...



Graphique 3 — Distribution des pourcentages de points (ordonnées) dans une population des 29 copies représentant chacune la meilleure copie de chacune des écoles participant au concours dans cette catégorie, (trait continu) et distribution des pourcentages attribués, par trois professeurs différents, chacun aux copies de tous leurs élèves participant au concours dans la catégorie 1.

99 %. Ces pourcentages sont disposés sur l'échelle de telle manière que, dans une distribution normale de cent notes attribuées chacune à une feuille d'examen et rangées de la plus élevée à la plus basse, la 50e occupera le milieu de l'intervalle entre 50 et 49 % (soir 49.5 %) et la 51e occupera la position correspondant à 50.5 %; la 49e et la 52e occuperont respectivement les positions 49.5 % et 51.5 % et ainsi de suite, de sorte que la première (la plus haute) occupera la position 0.5 % et la dernière, la plus basse, la position 99.5 %.

Si au lieu de 100 feuilles d'examen à noter nous n'en avons que 25, chacune des 25 notes se placera au milieu de l'intervalle correspondant à 4% sur l'échelle; la note se classant 13e se placera au niveau 50 %, la 12e au niveau de 46 %, la 12e à 38 %, la seconde à 6 % et la première à 2 %.

Dans la catégorie I du *National French Contest*, 29 feuilles d'examen ont été envoyées par 29 écoles comme représentant chacune la meilleure copie pour le groupe I de l'école. Les 29 feuilles étant classées par ordre décroissant de note

(exprimée en pourcentage des 100 points correspondant à l'absence de toute erreur) la 15e feuille devrait, dans l'hypothèse de distribution normale, se situer au rang 50 % sur l'échelle de probabilité, et la 14e au rang 46.5 % (correspondant à 50-(100/29), c'est-à-dire 50-3.45); la 16e devrait se placer au rang 53.45; de façon générale, les « pourcentages de points » doivent se distribuer de telle façon que le plus élevé se situe sur l'échelle de probabilité au rang 1.725 (correspondant à 3.45/2) et le 29e au rang 98.275.

Si la distribution est normale, on doit pouvoir faire passer par les 29 points une droite de régression qui intercepte la verticale élevée du milieu (50 %) de l'axe des abscisses à un niveau correspondant à la moyenne arithmétique de la distribution; les verticales élevées de 16 % et 84 % interceptent la droite de régression aux niveaux qui définissent au-dessus et au-dessous de la moyenne la valeur de $\pm \sigma$ (standard deviation).

Le graphique 1 montre qu'en fait aucune des distributions n'est normale; une droite passant par la série des points représentant les meilleures réponses, laisse très au-dessous d'elle un certain nombre de points représentant les plus mauvaises réponses. Dans la plupart des cas, on peut négliger l'effet de ces mauvaises réponses et considérer l'ensemble de la distribution comme normale.

De façon générale cependant et surtout pour les groupes d'élèves les mieux préparés, où la probabilité de réponses sans fautes ou presque sans fautes est assez grande, on obtient une meilleure rectilinéarité de la ligne de régression en portant en ordonnées non plus les « pourcentages de points obtenus » mais le logarithme du nombre des fautes, c'est-à-dire le logarithme de la différence entre 100 % et le pourcentage de points obtenu (si ce pourcentage est 100 % la différence devient zéro et ne trouve pas de place sur l'échelle logarithmique; si dans une distribution, la meilleure copie a 100 % et la seconde 98 % on renonce à localiser la première et on inscrit la seconde à son rang sur l'échelle des abscisses et à son niveau sur l'échelle logarithmique des ordonnées).

L'étude statistique des distributions de pourcentages de points attribués à des feuilles d'examen du *French National Contest* montre que ces distributions peuvent être considérées comme normales si l'on néglige la queue de la distribution. On peut alors aisément déterminer graphiquement la moyenne arithmétique et la déviation standard de chaque distribution.

Cependant, les distributions sont log-normales plutôt que normales; l'accord avec une distribution normale est d'autant moins bon et l'accord avec une distribution log-normale d'autant meilleur que le groupe d'élèves étudiés comprend davantage d'élèves bien préparés, faisant peu ou pas de fautes en même temps que quelques élèves peu aptes à s'assimiler la matière enseignée.

A la fin de leur dernière année de « High School » les élèves (seniors) généralement âgés de 18 ans, peuvent chercher immédiatement à gagner leur vie, ou s'efforcer de se faire admettre dans un collège pour y obtenir le diplôme de « Bachelor »; la proportion des « seniors » espérant pouvoir poursuivre leurs études varie de moins de 10 % dans certaines « Public High Schools » à 100 % dans certains établissements privés.

Le pourcentage des « seniors » étudiant le français n'est que de 3 % dans certaines « Public High Schools » et peut atteindre 50 % dans certaines écoles privées; mais, à de rares exceptions près, tous les « seniors » étudiant le fran-

çais se destinent au Collège, et ont atteint le niveau de scolarité, A ou B, qui doit leur en ouvrir les portes.

Il est généralement admis que pour obtenir ce que l'on est convenu d'appeler « du crédit » pour le français en entrant au Collège, c'est-à-dire pour ne pas être tenus de s'inscrire dans le cours destiné aux débutants et numéroté (1), les élèves de « High School » doivent avoir étudié la langue pendant deux ans au minimum.

Chaque élève de la catégorie I représente un étudiant potentiel de français. Chaque inscription dans une classe de français représente un choix fait par l'élève, non seulement entre plusieurs langues, mais encore entre plusieurs disciplines. Ce choix est le signe de l'éveil d'un intérêt qui peut se développer, se maintenir pendant un certain temps ou disparaître. Le développement de cet intérêt ne dépend pas seulement de l'aptitude de l'élève à s'assimiler la matière enseignée, mais surtout de la façon dont cette matière est présentée. De là un taux de mortalité très variable d'une école à l'autre. Tel est l'un des enseignements du concours où nous avons été appelé à assumer les responsabilités de « Chairman » pour le Chapitre de Californie.

Marie-Louise DUFRÉNOY.