FORMATION EN STATISTIQUE EN DUT STID ET TRANSPOSITION DIDACTIQUE

Étude de la transformation de Fisher dans l'analyse du coefficient de corrélation : d'une pratique fondée sur l'argument d'autorité et la lecture d'abaques à une argumentation fondée sur la simulation.

Jean-Claude Oriol* & Jean-Claude Régnier**
*CERRAL IUT LUMIERE & EA 3727 Université Lyon2
**Enseignant-Chercheur EA3727 Université Lyon2

RÉSUMÉ.
Cette communication se propose d'étudier la formation en statistique en DUT Stid et en Licence Professionnelle Chargé d'études statistiques et de comprendre la transposition didactique en acte. Notre propos concernera précisément l'analyse de la transformation de Fisher dans l'analyse du coefficient de corrélation, rencontrée en particulier par des étudiants de deuxième année de DUT STID et par les étudiants de licence professionnelle Chargé d'études statistiques. La particularité du travail présenté est contenue dans le fait que ce sont les étudiants qui construisent l'outil de simulation, à partir de consignes données par l'enseignant. Nous présenterons la séquence didactique mise en place, les principaux obstacles rencontrés et la mesure des résultats obtenus dans les apprentissages par les étudiants.

ABSTRACT
The didactic functioning of transposition in learning a statistical concept is studied in this paper. We more precisely deal with the Fisher Transformation in the study about the value of ρ, the correlation coefficient between variables x and y. of the underlying population, can be tested using the Fisher transformation of a sample's correlation coefficient r. This notion is studied in particular by DUT second year students and by third year students in "licence professionnelle".

MOTS-CLÉS : didactique, statistique, apprentissage, simulation, intervalles de confiance, transformation de Fisher.

KEYWORDS: didactic, statistic, knowlege, simulation, confidence intervals, Fisher transformation.

Introduction

En nous interrogeant sur la question de la formation citoyenne en statistique, nous sommes confrontés à des questions relatives aux concepts à enseigner et aux stratégies pédagogiques fondées sur la transposition et une ingénierie didactique. Nous nous intéressons ici non pas à la démonstration « mathématique » de l'amélioration de la convergence grâce à la transformation de Fisher appliquée au coefficient de corrélation, mais aux explications données aux étudiants dans le programme universitaire du DUT STID et à leur évolution depuis les années 70. Nous constatons ainsi le passage d'arguments d'autorité couplés à l'outil « abaque », à une argumentation plus souple fondée en partie sur la simulation.

Nous décelons dans cette modification des pratiques didactiques de l'enseignement de la statistique au niveau bac +2, d'une part une évolution des pratiques sociales et d'autre part l'émergence d'une place et d'un rôle plus importants attribués aux outils informatiques dans la pédagogie universitaire de la statistique. Il s'agit de tenter de déterminer l'intérêt d'une telle démarche mais aussi d'en relever les risques de genèse d'obstacles didactiques qui pourront perturber la formation en statistique.

Une remarque liminaire

Si l'on est attentif aux enseignements concernant le coefficient de corrélation il appert qu'une analyse d'un échantillon d'ouvrage fait ressortir une part non négligeable d'entre eux qui passent sous silence la question de la confiance à accorder à la réalisation du coefficient empirique obtenue sur un échantillon aléatoire extrait de population étudiée. Comment pouvons-nous interpréter ce fait ? peur de « rentrer » dans une question trop complexe, éloignement des justifications satisfaisantes, recommandations faites par les programmes ou leurs commentaires.

Petit retour sur l'état du problème pédagogique

Nous sommes en face d'un problème d'enseignement d'une notion que nous pouvons décrire de la façon suivante :
Il est dit que nous étudions deux variables quantitatives X et Y gaussiennes sur une population donnée de laquelle nous tirons un échantillon aléatoire de taille n. Sur cet échantillon nous calculons la réalisation du coefficient empirique de corrélation linéaire R entre X et Y. La valeur r obtenue est une estimation du coefficient ρ de corrélation linéaire entre les variables X et Y sur la population.
Il s'agit alors de tester les hypothèses : H0 : $\rho = 0$ contre H1 : $\rho \neq 0$
Sous H0, on connaît une réponse à ce problème : la variable aléatoire R coefficient empirique est telle la variable aléatoire $\frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$ suit la distribution de probabilité de la variable de un Student ayant $\upsilon = n - 2$ degrés de liberté.
Mais cette apparente simplicité est souvent mise à mal car, en particulier, la distribution de la variable R est complexe dès que ρ s'éloigne de 0.
C'est alors que l'on peut réaliser une transformation dite *transformation de Fisher* en construisant une nouvelle variable Z définie par : $Z = \frac{1}{2}\ln(\frac{1+R}{1-R})$ ou encore la fonction ArgthR (argument de tangente hyperbolique de R)
Il est dit alors que si n >30, taille de l'échantillon, Z suit approximativement une loi de Laplace-Gauss de paramètres LG(μ, σ) avec

$$\mu = E(Z) = \frac{1}{2}\ln(\frac{1+\rho}{1-\rho}) + \frac{\rho}{2(n-1)} \text{ et } \sigma^2 = V(Z) = \frac{1}{n-3}$$

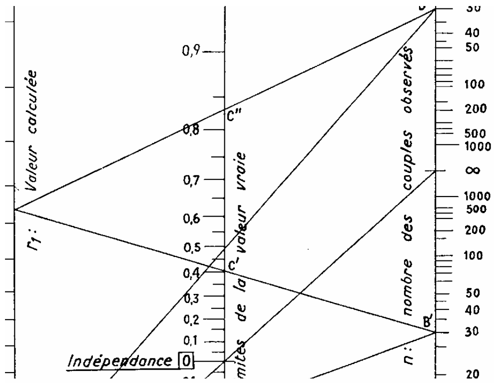
De là il ressort la possibilité de construire un intervalle de confiance à niveau 1-α par rapport à Z et donc par rapport à R moyennant la transformation réciproque :

$$R = \frac{\exp(2Z) - 1}{\exp(2Z) + 1}$$

Quelques solutions apportées dans des ouvrages de base

Dans un certain nombre d'ouvrages, c'est une solution graphique qui est fournie. Ainsi on donne un schéma sous la forme d'abaque permettant de se dispenser de ces calculs.
Signalons deux réalisations :

Tout d’abord dans (Liorzou 1979 p.175)



Ou bien encore dans (Wonnacott Wonnacott 1972 p.545)

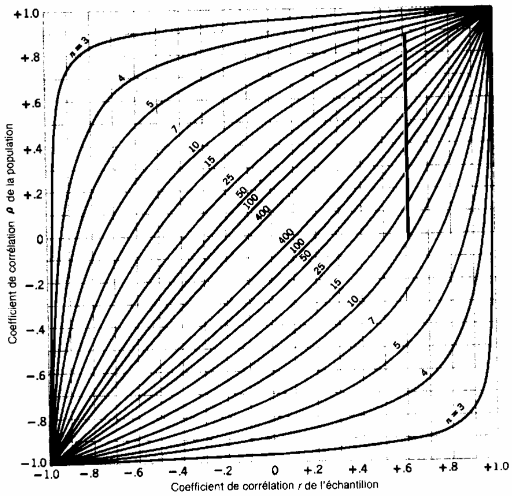
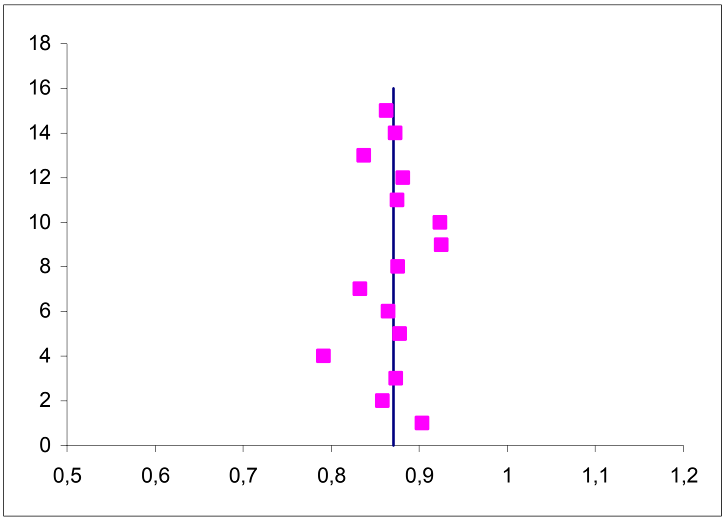


FIGURE 15·4
Bande de confiance à 95 % pour le coefficient de corrélation ρ dans le cas d'une population normale bivariée, pour différentes valeurs de la taille n de l'échantillon.

Ces deux abaques sont assez efficaces mais les étudiants qui les utilisent perdent de vue les fluctuations d’échantillonnages et ont des difficultés à faire le lien avec les éléments de théorie que nous avons rappelés ci-dessus, comme nous avons pu le constater à partir des données construites par observation dans la salle de classe. C’est pour cela que nous avons été conduit à réaliser une transposition didactique de l’approche de ce concept de corrélation linéaire fondée sur une approche liée à la simulation.

Une séquence didactique d’observation des fluctuations d’échantillonnage de R

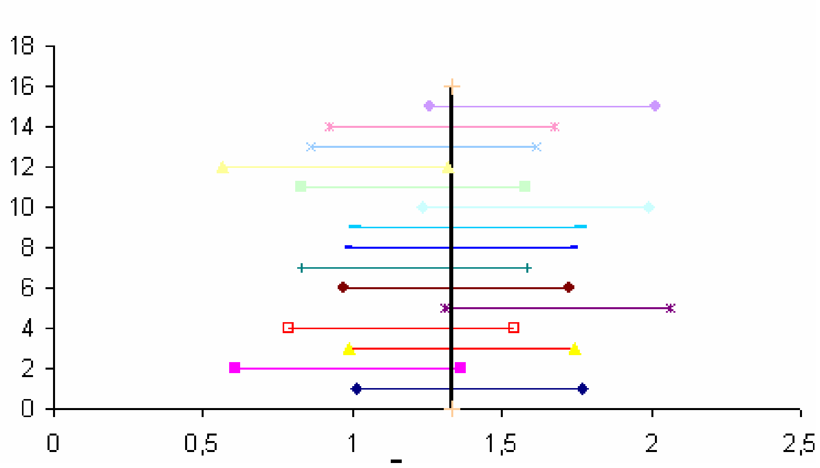
Nous avons réalisé une macro sous Excel qui fournit des réalisations du coefficient empirique R à partir d’échantillons aléatoires avec remise extrait d’une population de taille N=1000. Les deux variables étudiées X et Y connues sur la population donne donc la possibilité de connaître la valeur certaine de ρ. Le graphique ci-dessus donne les résultats d’une expérience fondées sur 15 échantillons et permet de comparer les 15 réalisations notées r_1 à r_{15} à la valeur de ρ.



L'introduction de la transformation de Fisher et de la variable Z

Sur ces quinze échantillons les étudiants ont appliqué la transformation de Fisher puis ont calculé et représenté graphiquement les intervalles de confiance. En voici un exemple :

N°échantillon	Réalisations de Z	Intervalle de confiance	
1	1,42583773	1,04864271	1,80303274
2	1,16252098	0,78532597	1,539716
3	1,51054716	1,13335214	1,88774217
4	1,40355715	1,02636214	1,78075217
5	1,45978321	1,0825882	1,83697823
6	1,38219475	1,00499973	1,75938976
7	1,26767646	0,89048145	1,64487148
8	1,374048	0,99685298	1,75124301
9	1,32848267	0,95128765	1,70567768
10	1,45518598	1,07799097	1,832381
11	1,45997334	1,08277832	1,83716835
12	1,76132777	1,38413276	2,13852279
13	1,13325308	0,75605807	1,51044809
14	1,3976812	1,02048619	1,77487621
15	1,16694815	0,78975313	1,54414316



Conclusion
L'objet de notre communication sera d'exposer comment, en quoi et à quelles conditions cette activité pédagogique constitue une transposition didactique qui permet aux étudiants une appropriation d'un savoir délicat sans perdre de vue les idées qui ont présidé à la transformation de Fisher pour étudier le coefficient de corrélation d'une population à partir de celui d'un échantillon.

Bibliographie

[1] Chevallard Yves (1991), *La transposition didactique*, Grenoble : La Pensée Sauvage
[2] Liorzou A. (1979}, *Initiation pratique à la statistique*, Paris : Eyrolles
[3] Pagès J.(2005) *Statistiques générales pour utilisateurs*, Rennes PUR
[4] Wonnacott, RJ et Wonnacott, TH (1972) *Introductory Statistics*, New York John Wiley & Sons