

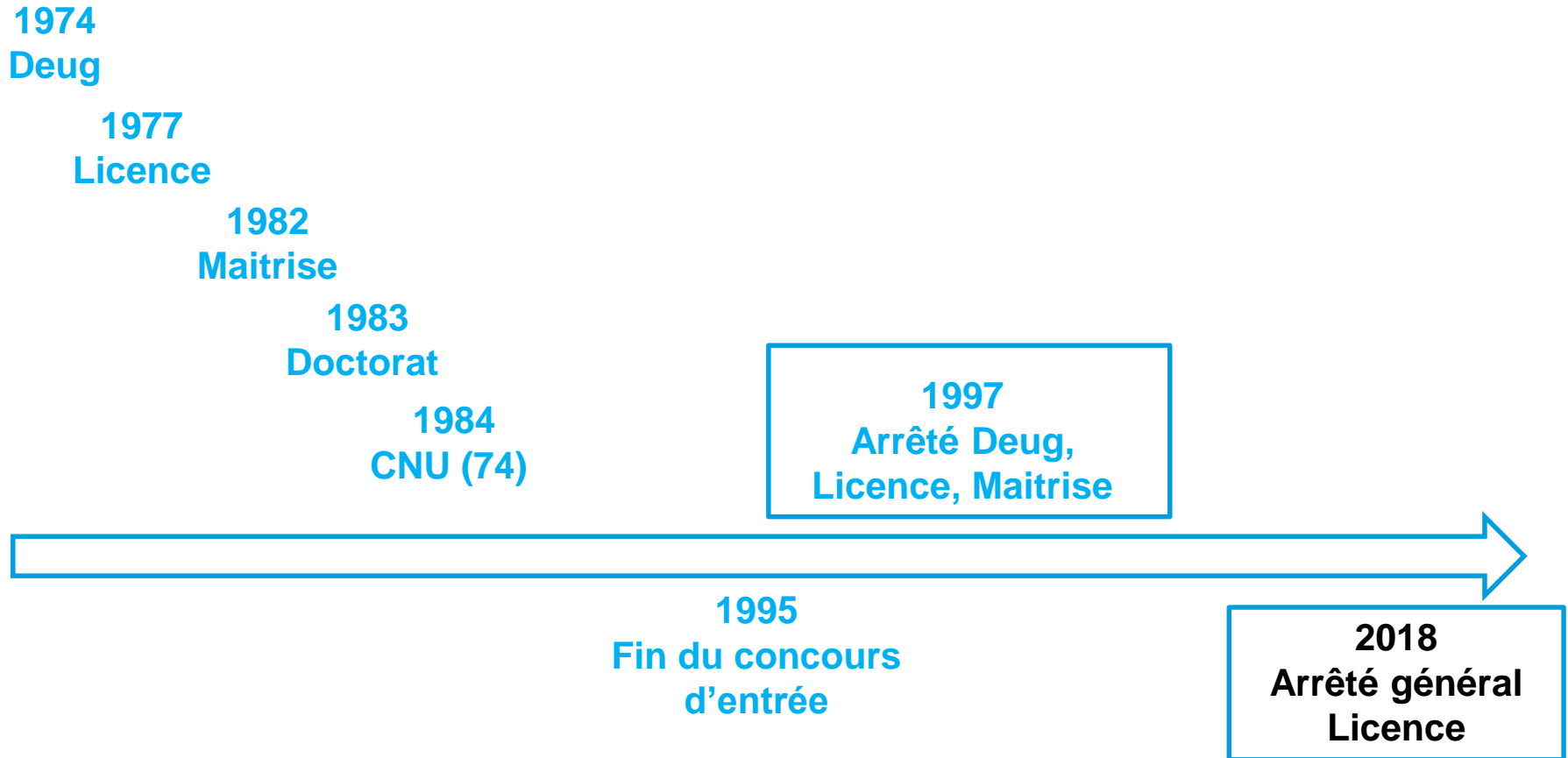
Léo Gerville-Réache

Univ. Bordeaux - Faculté des STAPS

Laboratoire IMS UMR 5218

**LA STATISTIQUE EN STAPS :
SON ENSEIGNEMENT PAR PROCURATION**

Introduction : Temporalité de la filière STAPS



Introduction : En 1997...

Arrêté du 23 mai 1997 relatif au diplôme d'études universitaires générales, à la licence et à la maîtrise STAPS

Les formations universitaires de STAPS sont conçues et organisées pour :

- *apporter aux étudiants une formation dont les contenus sont en prise directe avec l'évolution des connaissances et des formes de pratiques propres aux activités physiques et sportives (APS) et intègrent les dimensions scientifiques et techniques, artistiques et culturelles, les questions de santé et les enjeux éthiques de ces pratiques ;*
- *[...]*

Dans les deux cycles, la formation comporte la pratique d'au moins une langue vivante étrangère ainsi que de l'informatique.

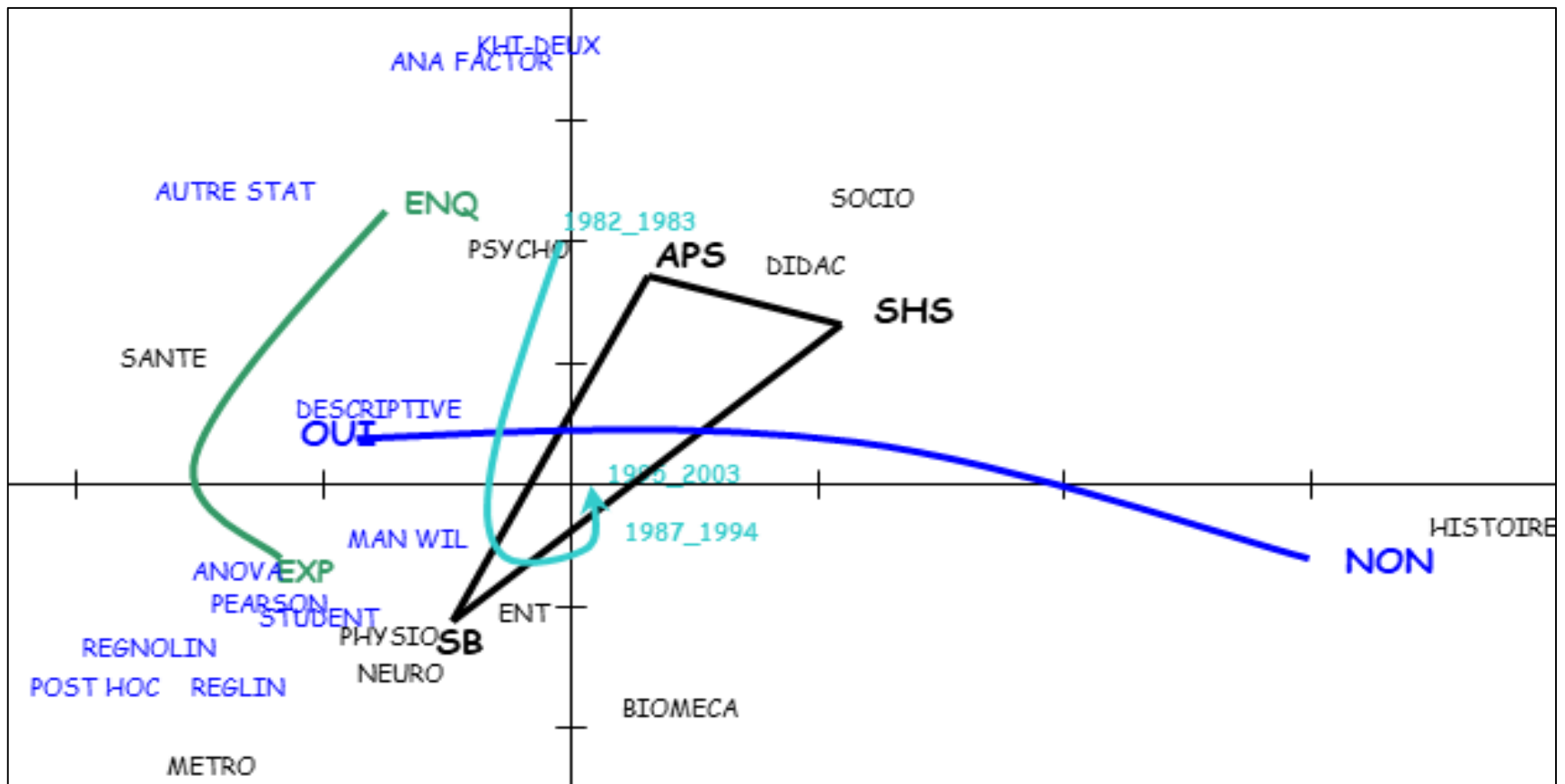
Aucune allusion à la statistique ou au traitement des données...

Introduction : Aujourd'hui...

Arrêté du 30 juillet 2018 relatif au diplôme national de licence (toutes mentions confondues) : *Tout au long du parcours personnalisé de formation, l'étudiant doit acquérir un ensemble de connaissances et compétences comprenant notamment: [...] **des compétences numériques et de traitement de l'information et des données.***

Fiches RNCP des mentions STAPS : « *Identifier, sélectionner et analyser avec esprit critique diverses ressources spécialisées pour documenter un sujet et **synthétiser ces données en vue de leur exploitation** » ...*

Introduction : La statistique en STAPS (jusqu'aux années 2000)



L'étude rassemble un échantillon de 233 articles de la revue « **Science et Motricité** » depuis son origine (en 1982) jusqu'en 2003.

Source : Sébastien Orazio (2008) : L'argument statistique dans les démarches scientifiques des activités physiques et sportives (thèse).

L'enseignement par procuration (CultureStat en STAPS)

Qu'est-ce que la culture statistique, comment l'enseigner, comment la certifier ?

50ème Journées de statistique. Gerville-Réache L., Jutand M-A. (2018).



La « capsule pédagogique »

- Une vidéo de quelques minutes,
- Une retranscription de la bande son,
- Des références, des prérequis,
- Des éléments d'indexation
- Un quiz,

Le sujet de la capsule est un concept, une méthode, un élément historique, une question d'actualité, un élément technique,...

CBEstd Communauté Bordelaise des Enseignants de statistique
<https://pedagogtec.u-bordeaux.fr/CBEstd/>

Historique Informatique Mathématique Philosophique Sociétale

Rechercher

Options d'affichage

11 résultats

▶ Ce qui pourrait se passer en 2156... Sociétale	▶ Construire un barème en sport Sociétale	▶ CultureStat, c'est quoi? (CBEstd) Historique
▶ Hasard et Incertitude Philosophique	▶ L'approche MBI pour le suivi d'un athlète Sociétale	▶ Le coefficient de variation Mathématique
▶ Le paradoxe de Simpson Philosophique	▶ Le principe d'indifférence Philosophique	▶ Qu'est-ce qu'un événement rare ? Sociétale

L'enseignement par procuration (une capsule...)

L'approche MBI pour le suivi d'un athlète

Leo Gerville-Reache



Informations

Date : 20 Septembre 2020

Status : Brouillon

Auteur : Leo Gerville-Reache

Domaine : Sociétale

Mots clés :

Transcription : [Transcription](#)

CEI/Est Communauté boréale des enseignants de statistique

Dans cette capsule je vais vous parler de la méthode MBI pour « Magnitude Based Inference » et de son application au suivi d'un athlète...

Pour vérifier sa compréhension...

Faire le quiz

L'enseignement par procuration (le livre, l'ouvrage...)

L'originalité de cet ouvrage est d'insister sur la liaison unissant les procédures statistiques avec les techniques de terrain.

Les notions théoriques ne sont pas brutalement assénées, mais présentées à l'aide d'exemples que le lecteur peut « faire fonctionner » lui-même pour se familiariser avec leur usage.

Public cible :

Cet ouvrage est destiné à ceux qui abordent les statistiques au cours de leur formation ou de leur pratique professionnelle.



Pierre Parlebas et Bernard Cyffers (1992)

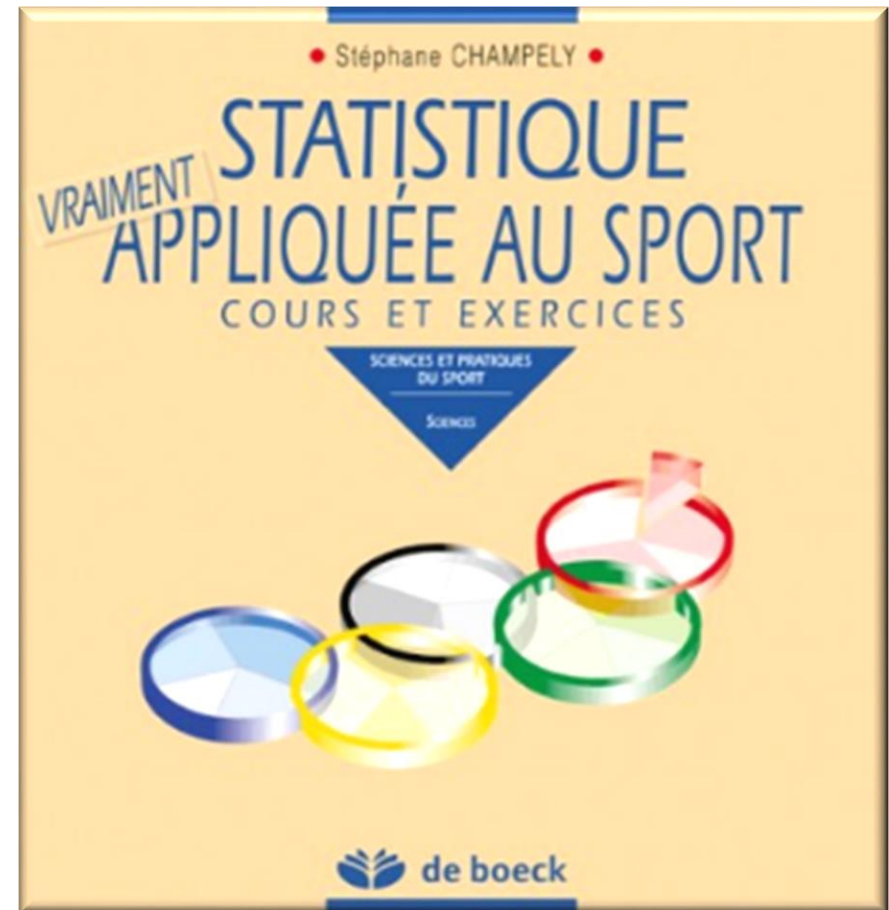
L'enseignement par procuration (le livre, l'ouvrage...)

*Cet ouvrage donne une multitude d'informations précieuses sur la **statistique descriptive** vraiment appliquée à son champ disciplinaire (psychologie du sport, marketing sportif, etc.).*

[...] les développements mathématiques ont été limités à leur plus simple expression.

Public cible :

*Conçu comme un manuel, ce livre s'adresse tout particulièrement **aux professeurs et étudiants de 1^{er} et 2^e cycle en STAPS** ainsi qu'à tout intervenant dans le domaine du sport.*



Stéphane Champely (2003)

L'enseignement par procuration (le livre, l'ouvrage...)

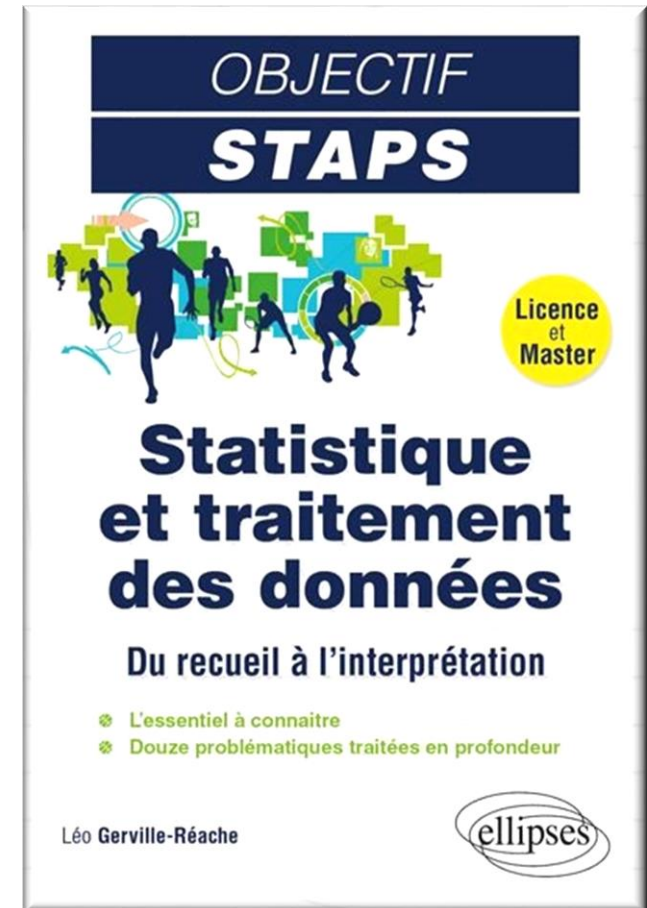
*Cet ouvrage propose une approche **généraliste et spécifique** de la statistique appliquée aux sciences du sport.*

*Ce qu'il faut connaître de la donnée à la comparaison, de l'enquête à l'expérimentation en passant par **la modélisation** est développé sans formalisme mathématique inutile.*

En faisant le choix de nous plonger dans 12 problématiques « très STAPS », l'auteur s'écarte volontairement des manuels classiques pour nous faire découvrir les principaux enjeux de la statistique... en STAPS.

Public cible :

*Écrit pour les **étudiants de licences, masters et doctorat STAPS**, il s'adresse également aux enseignants et enseignants-chercheurs en sciences du sport, aux data analysts traitant de données sportives, [...]*

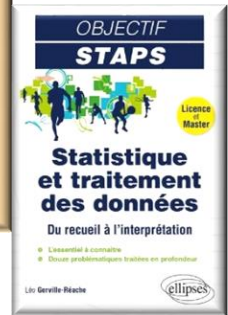
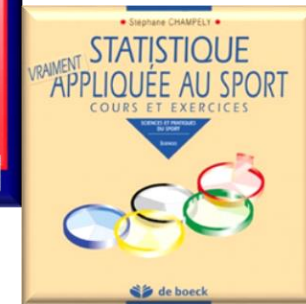


Léo Gerville-Réache (2022)

Les grands chapitres...

Pierre Parlebas et Bernard Cyffers (1992)

- Préface - Avant-propos - Lexique de rappel (20p)
- Observation expérimentale et méthode statistique (40p)
- Méthodologie expérimentale - Concepts de base (15p)
- Rappel de notions mathématiques (30p)
- La statistique descriptive (90p)
- La loi normale (35p)
- Étalonnage-Normalisation (40p)
- Estimations - Intervalles de confiance (40p)
- Tests d'hypothèses (40p)
- Tables – Bibliographie - (20p)



Stéphane Champely (2003)

- Avant-propos et Introduction (5p)
- Produire de bonnes données (50p)
- Organiser des données (50p)
- Comparer, relier ou croiser des données (55p)
- Généraliser les résultats obtenus (60p)
- Tables statistiques (15p)
- Bibliographie (10p)

Léo Gerville-Réache (2022)

Introduction (10p)

Partie 1 : Les fondamentaux (70p)

- La donnée (10p)
- L'outil informatique (5p)
- La description (10p)
- La comparaison (10p)
- La modélisation (10p)
- L'expérimentation (10p)

Partie 2 : Les problématiques (120p)

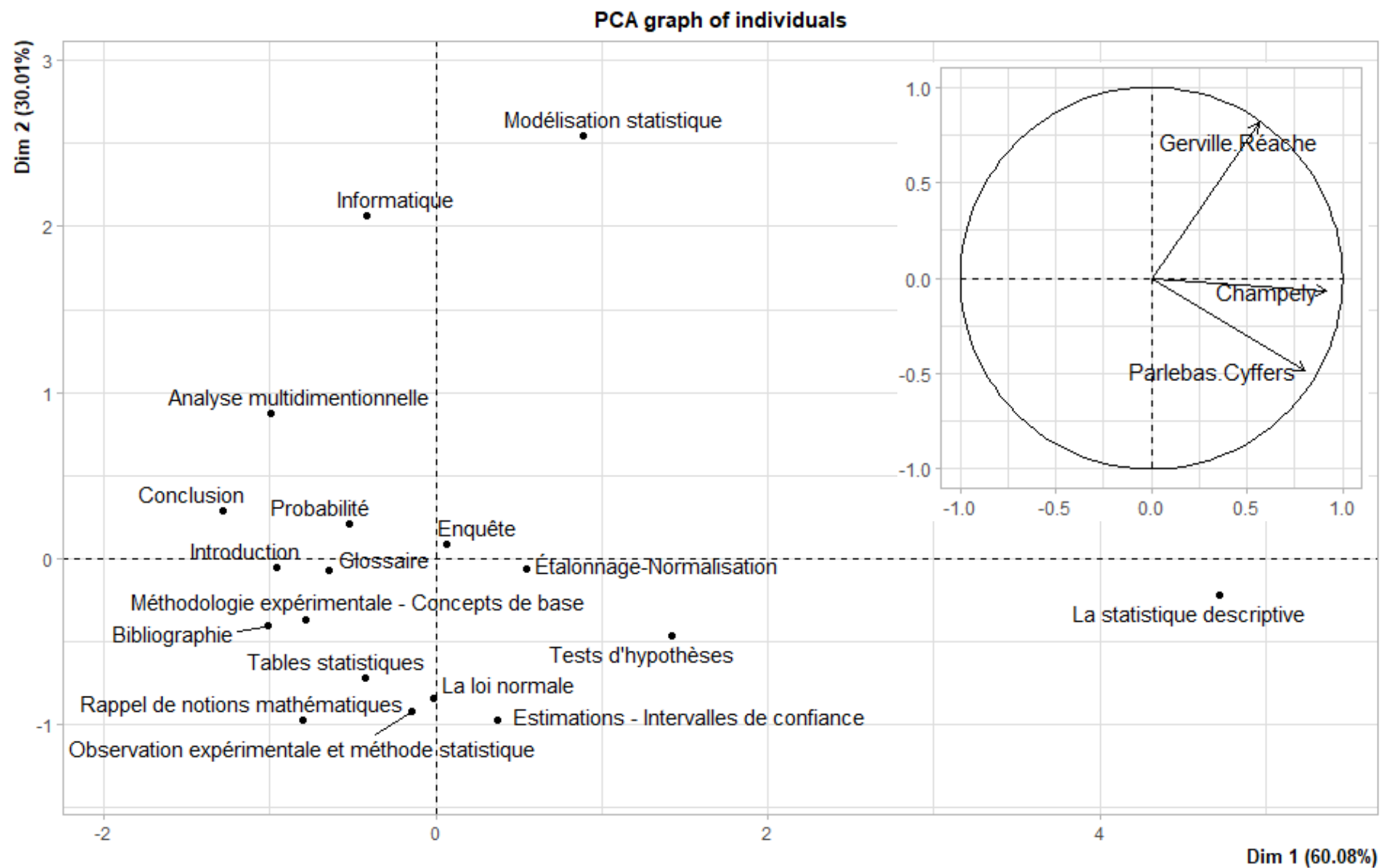
Conclusion (10p)

Glossaire (10p)

Surface rédactionnelle

Thème	Parlebas/Cyffers	Champely	Gerville-Réache
La statistique descriptive	24,3%	24,4%	15,9%
Tests d'hypothèses	10,8%	13,3%	6,0%
Estimations - Intervalles de confiance	10,8%	6,7%	2,0%
Étalonnage-Normalisation	10,8%	4,4%	7,9%
Observation expérimentale et méthode statistique	10,8%	2,2%	2,0%
La loi normale	9,4%	4,4%	2,0%
Rappel de notions mathématiques	8,1%	0,0%	0,0%
Tables statistiques	4,0%	6,7%	0,0%
Glossaire	4,0%	2,2%	4,0%
Méthodologie expérimentale - Concepts de base	4,0%	2,2%	2,0%
Introduction	1,6%	2,2%	2,8%
Enquête	0,8%	11,1%	4,0%
Bibliographie	0,5%	4,4%	0,0%
Modélisation statistique	0,0%	8,9%	19,8%
Probabilité	0,0%	6,7%	4,0%
Informatique	0,0%	0,0%	15,9%
Analyse multidimensionnelle	0,0%	0,0%	7,9%
Conclusion	0,0%	0,0%	4,0%

Surface rédactionnelle



Dim 1 (60.08%)



Le test de Student... (en 1992)

Une approche par taille d'échantillon

Décomposition des formules

Utilisation de tables statistiques

Nombreux exemples et exercices corrigés
(avec décomposition des calculs)

STATISTIQUE APPLIQUÉE AUX APS

et on compare u à u_α .

Par exemple, pour $\alpha = 0,05$, on accepte H_0 si $|u| < 1,96$.

b) σ_1 et σ_2 sont inconnus. On retrouve les deux cas :

• **Grands échantillons** : avec les mêmes conditions

$n > 120$: populations quelconques

$30 < n < 120$: populations normales.

On estime :

$$\sigma_1^2 \text{ par } s_1^2 = \frac{\sum_1 (\bar{x} - \bar{x}_1)^2}{n_1 - 1}$$

et :

$$\sigma_2^2 \text{ par } s_2^2 = \frac{\sum_2 (x - \bar{x}_2)^2}{n_2 - 1}$$

puis :

$$\sigma_d^2 \text{ par } s_d^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

On est ramené au cas a) précédent en formant :

$$u = \frac{d}{s_\alpha} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

• **Petits échantillons** : $n < 30$

Nous devons supposer, non seulement que les populations mères sont normales, mais également que leurs variances sont égales $\sigma_1^2 = \sigma_2^2$. Autrement dit, on suppose partir de deux populations normales identiques, puisqu'elles ont même moyenne et même écart-type.

La meilleure estimation de la variance commune est :

$$s^2 = \frac{\sum_1 (x - \bar{x}_1)^2 + \sum_2 (x - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

Cette estimation dépend de $\nu = n_1 + n_2 - 2$ degrés de liberté, puisqu'on a utilisé les deux relations :

$$\sum_1 (x - \bar{x}_1) = 0$$

Le test de Student... (en 2003)

La formule

Décomposition des calculs

Utilisation de tables statistiques

Nombreux exemples (avec décomposition des calculs) et très nombreux exercices **NON**-corrigés.

Tester des hypothèses (paramétriques) Chapitre 12

12.9 Test de Student « d'égalité de deux moyennes »

Quel est l'effet de la castration sur le développement des muscles squelettiques ? À court de volontaires humains, Doutreloux *et al.* [Doutreloux 01] ont tenté de le mesurer sur des lapins. Trente mâles composent le groupe de contrôle, tandis que trente autres ont été castrés à trois jours. Tous les animaux ont été abattus à 98 jours et on obtient comme force de résistance du muscle *longissimus dorsi* une moyenne de $40,3 \pm 0,9$ (N) pour le groupe castré et $45,3 \pm 1,2$ pour les mâles entiers. Les cinq newtons de différence observés entre ces deux groupes constituent-ils une différence statistiquement significative ?

Le modèle probabiliste le plus classique considère que les mesures des deux groupes sont produites par deux lois normales, de même écart type $\sigma^{(14)}$, mais pourvues d'espérances mathématiques propres μ_1 et μ_2 . Les hypothèses à tester sont $H_0 : \mu_1 = \mu_2$ contre $H_a : \mu_1 \neq \mu_2$.

La statistique de test sera la différence entre les deux moyennes observées $\bar{y}_1 - \bar{y}_2$ (soit ici $40,3 - 45,3 = -5,0$). On a vu dans la section 11.6 que la distribution d'échantillonnage de cette différence $\bar{y}_1 - \bar{y}_2$ suit une loi normale d'espérance $\mu_1 - \mu_2$, qui est ici *nulle par hypothèse*, et dont l'écart-type dépend de σ , écart-type commun aux deux populations.

On a également décrit dans le même temps comment estimer σ par $\hat{\sigma}$, compromis entre les deux dispersions observées. Du fait de cette approximation, on utilise une loi de Student dont le nombre de degrés de liberté est $n_1 + n_2 - 2$ pour situer la valeur normalisée.

DÉFINITION 12.13
Pour tester dans deux populations normales, d'écart-types identiques σ , l'hypothèse que leurs espérances mathématiques sont égales ($H_0 : \mu_1 = \mu_2$), on utilise le *test de Student pour échantillons indépendants*. On calcule pour les moyennes observées \bar{y}_1 et \bar{y}_2 des deux échantillons de tailles n_1 et n_2 le score normalisé suivant qui servira de statistique de test :

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\hat{\sigma} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

où $\hat{\sigma} = \sqrt{\frac{(n_1 - 1) \times (\hat{\sigma}_1)^2 + (n_2 - 1) \times (\hat{\sigma}_2)^2}{n_1 + n_2 - 2}}$ ($\hat{\sigma}_1$ et $\hat{\sigma}_2$ étant les écarts-types observés). La probabilité critique de l'hypothèse nulle

- contre l'hypothèse $H_{a(1)} : \mu_1 > \mu_2$ est égale à la probabilité que la loi de Student à $n_1 + n_2 - 2$ degrés de liberté soit plus élevée que t .
- contre l'hypothèse $H_{a(2)} : \mu_1 < \mu_2$ est égale à la probabilité que la même loi soit plus petite que t .
- contre l'hypothèse $H_{a(3)} : \mu_1 \neq \mu_2$ est égale à la probabilité que cette loi soit plus éloignée de zéro que t .

Pour notre exemple, nous obtenons

$$\hat{\sigma} = \sqrt{\frac{(30 - 1) \times (0,9)^2 + (30 - 1) \times (1,2)^2}{30 + 30 - 2}} = 1,06$$

d'où

$$t = \frac{40,3 - 45,3}{1,06 \sqrt{\left(\frac{1}{30} + \frac{1}{30}\right)}} = -18,3.$$

Il faut déterminer la probabilité qu'une loi de Student à 58 degrés de liberté soit inférieure à $-18,3$ ou supérieure à $18,3$. Elle est quasiment nulle. On en déduit que l'hypothèse nulle n'est pas cohérente avec les données et qu'il existe un écart statistiquement significatif entre les deux groupes.

12.10 Exercices

EXERCICE 211
Les affirmations suivantes vous paraissent-elles vraies ou fausses ?

- Le test « d'égalité entre deux moyennes » utilise comme hypothèse nulle $\mu_1 = \mu_2$.
- Dans les tests appariés, les deux échantillons ont forcément la même taille.
- On utilise des échantillons appariés pour obtenir une comparaison plus précise qu'avec des échantillons indépendants.

EXERCICE 212
Dans le cadre du test Z « d'égalité de deux proportions », dire dans les cas (a, b, c) suivants : (1) Quelle est la différence entre les deux proportions observées ? (2) Quelle est la proportion p , compro-

(14) Une hypothèse forte dont il convient de vérifier la compatibilité avec les données. Il existe d'ailleurs des tests de cette hypothèse, le plus connu, celui de Fisher, étant loin d'être le meilleur... on lui préférera le plus confidentiel test de Levene [Brown 74]. Lorsque l'écart-type des deux groupes est définitivement différent, on peut utiliser une transformation (cf. section 7.11), un test adapté comme celui de Welch [Welch 38]], des tests dits robustes [Staudte 90] inspirés des statistiques vues en section 6.12, des tests non paramétriques (chapitre 13) ou des procédures basées sur des simulations informatiques (test de randomisation, bootstrap...)



Le test de Student... (en 2022)

La formule

L'application avec R/Excel, utilisation de la p-value (pas de tables statistiques)

Une lecture des « propriétés » de la formule

Peu d'exemples et pas d'exercice.
12 problématiques qui mettent en œuvre les différents concepts

Ce qu'il faut bien comprendre c'est que vous pouvez, à partir de vos données, calculer des moyennes sur deux populations distinctes et observer une différence. Mais cette différence n'est pas nécessairement significative, statistiquement significative. Encore faut-il que l'écart entre ces moyennes, relativement à la variabilité des données et à leur nombre soit suffisant...

Regardons la statistique du test de Student pour deux échantillons indépendants. On suppose que l'on a deux échantillons de données X_1, \dots, X_n et Y_1, \dots, Y_n dont les distributions suivent des lois normales (ou bien sont de tailles suffisantes, en pratique, supérieures à quelques dizaines).

Ce test de Student teste l'hypothèse H_0 : « Égalité des deux moyennes ». C'est fondamental lors de l'utilisation d'un test de respecter ses conditions d'utilisation et son hypothèse H_0 (l'hypothèse que le test étudie, pas celle que vous avez en tête). L'hypothèse H_1 est ici: « Différence des deux moyennes ».

Imaginez par exemple des résultats de 60 mètres pour des filles de 4^e d'une part et de garçons de 4^e d'autre part (ce sont deux échantillons qu'il est raisonnable de supposer indépendants et suivant des lois normales). Voici les deux premiers calculs que fera tout logiciel:

$$S^2 = \frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2} \quad \text{et} \quad t = \frac{|\bar{X}_n - \bar{Y}_n|}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$$

Clairement, un décryptage s'impose! S^2 est la moyenne pondérée des variances des échantillons de X et de Y . Cette moyenne est alors utilisée pour relativiser l'écart des moyennes dans la statistique de Student (t). Bien sûr, comme pour l'intervalle de confiance, la taille des échantillons est également là pour relativiser tout cela. En bref, la statistique « t » sera d'autant plus grande que la différence entre les moyennes sera grande, d'autant plus grande que les écarts-types des X et Y seront petits, d'autant plus grande que les échantillons seront de grandes tailles (toutes choses égales par ailleurs).

Vous vous demandez peut-être où est passée la p-value (p)? Je ne vais pas rentrer dans les détails mais c'est en appliquant la fonction des quantiles de Student avec le bon degré de liberté sur la valeur de t que l'on va obtenir p .

Je ne suis pas sûr que cela soit parfaitement clair pour vous. L'essentiel est de comprendre qu'à partir des données, nous sommes en mesure de calculer une première statistique t puis une seconde, la fameuse p-value.

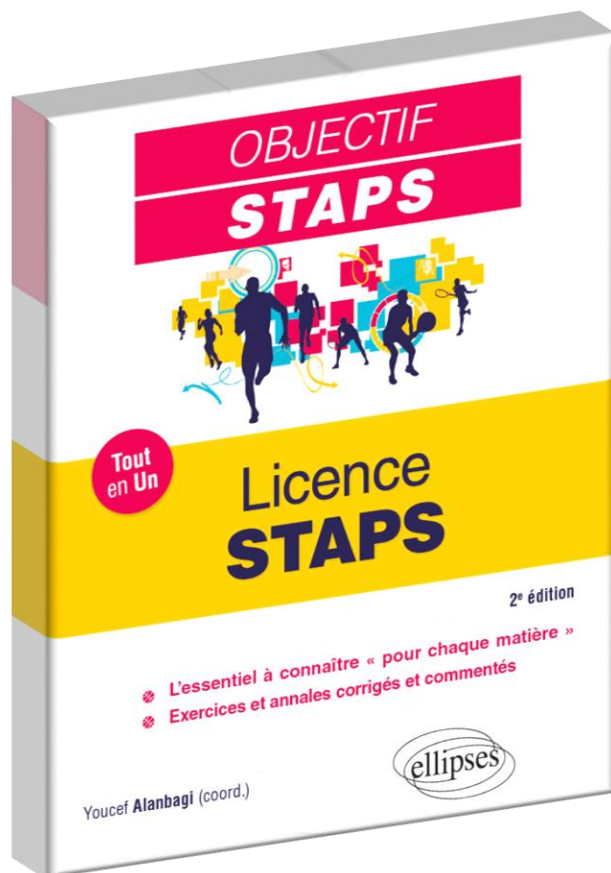
Je vous ai montré la statistique t pour vous donner un exemple de construction et ses propriétés en fonction des moyennes, des écarts-types et des tailles d'échantillons. Si vous regardez bien, ce sont les mêmes principes que pour l'intervalle de confiance.

L'approche par problématique

- Le barème (évaluation, docimologie, z-Score)*
- L'accès au sport (Inégalité, territorialité, indice, composition de z-scores)**
- Le Parcoursup STAPS (probabilité de réussite, critère d'évaluation, classement)*
- La motivation dans le sport (motivation, validation, ACP, AFCM)**
- Le patinage artistique (jury sportif, valeur aberrante, Anova)***
- La batterie de test (validité, fidélité, corrélation)*
- Le suivi de l'athlète (NHST, MBI, SWC, valeur aberrante)**
- L'entraînement et la performance (forme, fatigue, modèle de Banister)***
- Le handicap de jeu au ping (probabilité de victoire, égalité des chances)***
- Le jeu du penalty (théorie de jeux, stratégie mixte, rationalité, espérance)***
- Le pari sportif au rugby (cote, attaque/défense, modèle polytomique)***
- Une perle *

Discussion : L'enseignement de la statistique en STAPS...

TABLE DES MATIÈRES



Préface	9
Introduction	13
■ L'enseignement de l'éducation physique et sportive à l'école	17
■ Psychologie sociale	67
■ Sociologie	115
■ Socio-histoire et épistémologie des APS.....	181
■ Anglais.....	241
■ Anatomie	279
■ Biomécanique	335
■ Neurosciences comportementales	397
■ Physiologie.....	543
■ Méthodologie de l'entraînement	683
Postface	707
Liste des auteurs	713

Bibliographie (short)

[1] Champely S. (2003). Statistique vraiment appliquée au sport (cours et exercices), Editions De Boeck, Louvain.

[2] Gerville-Réache L., Jutand M-A. (2018). Qu'est-ce que la culture statistique, comment l'enseigner, comment la certifier ? 50ème Journées de statistique, Paris, 4p.

[3] Gerville-Réache L. (2022). Statistique et traitement des données, Collection Objectif STAPS, Editions Ellipses, Paris.

[4] Orazio S. (2008) : L'argument statistique dans les démarches scientifiques des activités physiques et sportives (thèse de l'Université Bordeaux Ségalen)

[5] Parlebas P., Cyffers B. (1992). Statistique appliquée aux activités physiques et sportives, Collection Etudes et Formation, INSEP-Éditions, Paris.

[6] Trabal P. (2000). Pratiquer les statistiques (initiation à l'usage des acteurs du sport et de l'éducation physique), Vuibert, collection STAPS.

Questions

Merci pour votre attention