

On attribue à l'histoire de cette science à la date de commencement de 1746, bien que l'interprétation du **القديم الحجري** terme « statistique » a changé au cours du temps. ☞ Aux temps plus anciens (Paléolithique ans ~ av. J.-C; Chinois: 2238 av. J.-C; Egyptiens: 1700 av. J.-C,...), cette science ne consistait qu'à la collection d'informations des États, d'où l'étymologie du nom, de l'Allemand ACHENWALL, qui aurait créé le mot Statistik, dérivé de l'italien statista ("Homme d'État"). ☞ Plus tard, cette définition est étendue à tout type d'information collectée et, encore plus tard, les sciences statistiques incluent l'analyse et l'interprétation de ces données. ☞ En termes modernes, les statistiques incluent les ensembles de données, telles celles de la comptabilité nationale et les registres de températures, ainsi que le travail d'analyse, lequel requiert les méthodes de l'inférence statistique.

Dénombrement/Comptage Méthodes/Règles Dans le domaine de la santé, l'utilisation des statistiques est assez tardive, puisqu'il faut attendre le XVIIIème siècle pour voir apparaître les premières tables de mortalité. Le développement de la méthode statistique est quant à lui relativement récent, et ne débute véritablement que dans le courant du XVIIIème siècle. ☞ Le développement des mathématiques et des probabilités (PASCAL, FERMAT, LAPLACE, BAYES,...) permet de dégager des règles pour le traitement des données. ☞ Parallèlement, les domaines ayant recours aux méthodes statistiques se multiplient : l'agronomie, puis de la biologie, pour gagner peu à peu l'économie, l'industrie, et le domaine de la santé (PEARSON et FISHER). ☞ Au XXème siècle, le principal déterminant du développement de la statistique a été sans conteste l'apparition et la diffusion de l'outil informatique, ce qui a permis le développement de nouveaux outils statistiques qu'il était peu envisageable d'utiliser « à la main », mais a également entraîné une certaine démocratisation... ☞ De nos jours, les statistiques ont gagné tous les champs de la société....

BIO ☞ Grec : bios SCIENCE DU VIVANT Les formations mathématiques reçues en S1 et S2 sont largement suffisantes pour aborder le cours de stat. Prérequis pédagogiques

Objectifs du module : M234 « Biostatistique » ☞ Initier les étudiants à l'utilisation des statistiques en biologie. ☞ Apprendra à raisonner avant d'appliquer le modèle statistique adapté à sa problématique. ☞ Familiariser l'étudiant aux outils statistiques en insistant davantage sur le mode de raisonnement statistique que sur les aspects mathématiques. ☞ Le cours doit apprendre à l'étudiant à choisir une démarche statistique rationnelle en tenant compte des risques associés à une décision.

Compétences à acquérir Il doit donner aux étudiants les compétences suffisantes en statistique pour la suite de leurs cursus à l'université, ainsi que pour leur parcours extérieur et être capable de présenter des données brutes sous forme synthétiques (Tableaux; graphiques ou paramètres statistiques) et savoir comment peut-on généraliser les paramètres calculés au niveau d'un ou plusieurs échantillons sur la population statistique.

CNPN Partie I Statistique Descriptive Partie II Probabilité & Distributions Théoriques Partie III Théorie statistique de l'estimation Partie IV Statistique Inférentielle

Partie I Statistique Descriptive

1. Typologie des variables
2. Analyse univariée (1 dimension) : Indicateurs (position, dispersion & forme), Mesures de concentration (Degré et Indice de de concentration)
3. Analyse bivariée (2 dimensions) : Corrélation entre deux variables quantitatives, Analyse deux variables quantitatives par régression linéaire
4. Représentation graphique

☞ Analyse déductive des données Représentent l'ensemble des méthodes et techniques permettant de présenter, de décrire et de résumer des données nombreuses et variées ☞ Une variable = une quantité ou qualité susceptible de fluctuer ou de varier. Dans certains cas,

ces variables utilisées ne peuvent pas être mesurées de manière « objective » ☞ On mobilise plutôt des variables latentes ou construites. Exemple : image de marque, motivation... ☞ L'ensemble sur lequel porte l'étude s'appelle Population. ☞ On effectue généralement des mesures sur les individus (= unités statistiques) qui composent cette population + 3ème type de variables : Temporelles : Dates, Heures...

1. Typologie des variables = Discontinues = série groupée ☞ Comptable (valeurs isolées) = Série non groupée ☞ Mesurable sur une échelle Fréquence, Répartition Proportionnelle (% , %o...) Indicateurs de position, de dispersion, de forme... Partie I Statistique Descriptive 1. Typologie des variables 2. Analyse univariée (1 dimension) : Indicateurs (position, dispersion & forme), Mesures de concentration (Degré et Indice de de concentration) 3. Analyse bivariée (2 dimensions) : Corrélation entre deux variables quantitatives, Analyse deux variables quantitatives par régression linéaire 4. Représentation graphique ☞

Analyse déductive des données Représentent l'ensemble des méthodes et techniques permettant de présenter, de décrire et de résumer des données nombreuses et variées 3 Indicateurs caractérisent la distribution d'une série statistique quantitative: Les paramètres de position : Moyenne « arithmétique »,

Fractile (Médiane, quartile, déciles, centiles,...), Min, Max, Mode Les paramètres de dispersion : Variance, Écart-type, Coefficient de Variance, Étendue (Amplitude), Ecart Interquartiles, Interquartile relatif, Intervalle de Kelley, Interdécile relatif Les paramètres de forme : Aplatissement (Kurtosis), Asymétrie (Skewness) 2.1. Indicateurs 2. Analyse univariée Il existe d'autres types de moyennes : ●

Moyenne géométrique d'une série de valeurs positives est la racine nième du produit des n valeurs. Elle est toujours inférieure ou égale à la moyenne arithmétique. ● Moyenne harmonique d'une série de valeurs positives est égale à l'inverse de la moyenne des inverses. ● Moyenne quadratique est la racine carrée de la moyenne arithmétique des carrés. ☞ Moyenne arithmétique Cette moyenne arithmétique est la plus ancienne méthode employée pour caractériser un ensemble de données et indiquer une tendance centrale. Elle représente le centre de gravité de la distribution. Soit $x_1, x_2, \dots, x_i, \dots, x_n$ Si les données sont organisées en classes (variable continue) de centre c_i , on remplacera x_i par c_i

Inconvénient : ce paramètre est très sensible aux valeurs extrêmes de la série Paramètres de position ☞

Moyenne arithmétique pondérée Il faut tenir compte des coefficients Matière Coefficient note notes coefficientées Physique 4 12 $4 \times 12 = 48$ Chimie 4 8 $4 \times 8 = 32$ Philosophie 1 5 $1 \times 5 = 5$ Histoire/Géographi

e 1 14 $1 \times 14 = 14$ Maths 10 9,5 $10 \times 9,5 = 95$ Total des coefficients : $4 + 4 + 1 + 1 + 10 = 20$ Total des notes coefficients : 194 Moyenne pondérée : $= 9,7$ soit la note est de $9,7 / 20$ ☞ Fractile La médiane est la valeur de la série qui partage la distribution en 2 sous-ensembles d'égal effectif. Le quartile est la

valeur de la série qui partage la distribution en 4 sous-ensembles d'égal effectif. Le centile (ou percentile) est la valeur de la série qui partage la distribution en 100 sous ensembles d'égal effectif. Min=Q0=Do=Co Max=Q4=D10=C100 Q1=D2.5=C25 Q3=D7.5=P75 Me=Q2=D5=C50 Mode ???? ☞ Mode = la valeur dominante Correspond à la valeur la plus fréquente. x_i correspondant au n_i (ou f_i) maximum. Il peut y avoir un ou plusieurs modes. 10, 9, 12, 11, 10, 8, 14, 11, 9, 16, 5, 12, 10, 11, 10, 13 Moyenne = 10,7 ; Médiane = 10,5 ; Mode = 10