

VOCABULAIRE-INDEX¹

FPStat 2, niveau 2, module 1

Introduction à la décision statistique

coefficient (d'un intervalle) de confiance (*confidence probability*) **T2.4** : proportion des intervalles de confiance (construits suivant une méthode donnée) qui contiennent le paramètre cherché parmi tous les intervalles qu'on pourrait obtenir à partir de données de même nature.

degrés de liberté (d'une statistique) (*degrees of freedom*) : paramètre des lois du χ^2 , de Student et de Fisher ou nombre de termes linéairement indépendants dans le calcul de sommes de carrés basées sur des observations indépendantes.

densité (fonction) (*density function*) **T1.16** : fonction numérique qui décrit la distribution d'une variable aléatoire continue.

écart-type ou écart quadratique moyen ou déviation standard (*standard deviation, SD*) **T1.21** : racine carrée de la variance.

échantillon (*sample*) : suite finie de variables aléatoires issues d'une même distribution et généralement supposées indépendantes.

erreur standard ou erreur type (*standard error, SE, SEM*) : écart-type d'un estimateur d'un paramètre.

estimateur (*estimator, estimate*) **T 1.31** : formule qui permet de calculer une valeur approchée du paramètre d'une population à partir d'un échantillon. Exemple : la médiane d'un échantillon est un estimateur de la médiane de la population.

estimation (*estimation*) **T 1.33** : valeur approchée du paramètre obtenue à partir d'un échantillon.

Hypothèse statistique (*statistical hypothesis*) : concerne la distribution d'une ou plusieurs variables aléatoires. L'hypothèse peut porter sur des paramètres simples (paramètres de position), sur des propriétés plus compliquées (indépendance de plusieurs variables) voire sur toute une distribution (ajustement).

Hypothèse nulle (*null hypothesis*) **T3.9** : notée habituellement H_0 . Lorsqu'on fait un test d'hypothèse, on appelle hypothèse nulle l'hypothèse sous laquelle la loi de la statistique de test est connue.

1. pour chaque terme on donne sa traduction en anglais, une définition et/ou le numéro du transparent où celui-ci est défini

Hypothèse alternative (*alternative hypothesis*) **T3.9** : notée habituellement H_1 . Hypothèse de test sous laquelle la loi de la statistique de test n'est pas connue. Par exemple «il y a une différence entre les populations» ou «la machine est dérégulée». Cette alternative peut avoir différentes formes selon la nature du problème, selon ce qu'on sait *a priori*. Par exemple elle peut être de la forme : “les deux populations d'origine ont des distributions de poids de poulets différentes”, ou de la forme : “ les poulets de type 1 sont plus gros que les poulets de type 2”. Cette deuxième alternative est adaptée si on a cherché à améliorer le régime des poulets 1.

inférence statistique (*statistical inference*) : étude des conclusions que l'on peut tirer à partir de l'échantillon d'une population et du degré d'exactitude de ces conclusions.

intervalle de confiance (*confidence interval*) **T2.5** : région calculée à partir de mesures d'une expérience qui contient le paramètre qu'on cherche à approcher dans la majorité des cas. En fait, il y a des erreurs possibles, c'est à dire des cas où l'intervalle ne contient pas le paramètre (voir coefficient de confiance).

mesures de position (*measure of location*) : aussi appelés valeurs centrales, elles servent à caractériser l'ordre de grandeur des individus d'une population. Ce sont principalement la moyenne arithmétique, la moyenne géométrique, la moyenne harmonique, la moyenne quadratique, la médiane, la médiale, le mode.

mesures de dispersion (*measure of dispersion*) : permettent de chiffrer la variabilité des valeurs autour d'un paramètre de position. Ce sont principalement la variance, l'écart-type, le coefficient de variation, l'écart moyen absolu, l'écart interquartile, l'amplitude.

niveau descriptif (niveau de signification réel, probabilité critique) (*p-value*) **T3.27** : valeur du risque de première espèce pour laquelle la statistique de test observée devient significative. Si le niveau descriptif est supérieur au risque de première espèce fixé *a priori* (par ex. 5%), on ne rejette pas l'hypothèse nulle.

pivot **T2.9** : variable aléatoire dépendant à la fois des observations et d'un paramètre à estimer et dont la distribution ne dépend pas du paramètre à estimer. Ce type de variable est utilisé dans la construction d'intervalles de confiance.

population (*population*) **T1.2** : ensemble d'individus ou d'unités auxquels on s'intéresse.

puissance d'un test (*power*) **T3.13** : probabilité de rejeter l'hypothèse nulle alors qu'elle est fautive (c'est-à-dire probabilité de voir sur les échantillons une différence qui existe). Elle est notée $1 - \beta$ ($\beta = \text{beta}$).

quantiles ou fractiles d'ordre k (*quantiles, fractiles*) **T1.23** : sont les $k - 1$ valeurs qui divisent une série statistique ou une distribution de fréquences en k classes de même effectif.

répartition (fonction) (*cumulative probability*) **T1.17** : fonction numérique associée à une distribution qui à chaque valeur associe la probabilité d'obtenir une observation plus petite ou égale à cette valeur.

risque de première espèce (niveau de signification) (*error risk of first kind, significance level*) **T3.12** : c'est la probabilité de choisir l'hypothèse alternative alors que l'hypothèse nulle est vraie. Par exemple, c'est la probabilité que la différence entre deux échantillons soit significative quand les deux échantillons proviennent d'une même population. C'est donc la probabilité de conclure à une différence qui n'existe pas. On le note généralement α (alpha) et on le choisit *a priori*.

risque d'erreur de deuxième espèce (*error risk of second kind*) **T3.12** : probabilité de choisir l'hypothèse nulle alors que c'est l'hypothèse alternative qui est vérifiée. Par exemple, c'est la probabilité de ne pas voir sur les échantillons une différence qui existe entre les populations. Il est habituellement noté β (beta).

Statistique : ensemble des méthodes scientifiques à partir desquelles on recueille, organise, résume, présente et analyse des données, et qui permettent d'en tirer des conclusions et de prendre des décisions judicieuses.

statistique de test : variable calculée à partir des observations permettant de définir une règle de décision.

test d'hypothèses (test de signification) (*test of hypothesis, significance test*) : procédure qui a pour but de vérifier, à partir des données observées dans un ou plusieurs échantillons, la validité de certaines hypothèses relatives à une ou plusieurs populations.

test bilatéral (*two-sided test, double-tailed test*) **T3.6**

test unilatéral (*one-sided test, single-tailed test*) **T3.6**

valeur critique (*critical value*) **T3.9** : c'est une valeur qui sépare en deux ensembles des valeurs de la statistique de test choisie. Si la valeur observée est d'un côté de cette valeur critique, on rejette l'hypothèse nulle comme peu vraisemblable, sinon on ne la rejette pas. Cette valeur critique se lit en général dans une table et dépend du niveau choisi et de l'alternative choisie. Dans certains tests il y a deux valeurs critiques qui définissent un intervalle et on rejette l'hypothèse nulle quand la valeur calculée sur l'expérience tombe à l'extérieur de cet intervalle.

variable aléatoire (*random variable, variate*) **T1.7** : variable associée à une expérience ou à un groupe d'expériences aléatoires servant à caractériser le résultat de cette expérience ou groupe d'expériences.