L'INDUCTION STATISTIQUE

L'induction statistique a pour but de faire émerger des propriétés d'un ensemble de variables. Elle s'appuie sur les résultats de la statistique mathématique, qui applique des calculs mathématiques rigoureux utilisant la théorie des probabilités. Sans les statistiques mathématiques, un calcul sur des données (par exemple une moyenne), n'est qu'un indicateur. Ce sont les statistiques mathématiques qui lui donnent le statut d'estimateur dont on maîtrise le biais et l'incertitude.

I. La procédure que suit l'induction statistique

La procédure stipule de passer par quatre étapes :

a. Poser les hypothèses de travail

Les tests statistiques sont des tests d'hypothèses parce que leur application peut conduire au rejet de l'hypothèse nulle (H_0). L'hypothèse nulle stipule toujours qu'il n'existe pas de différence à part les différences produites par le hasard de l'échantillonnage.

H_0 : $F = F_0$, où F fait référence à la distribution (cumulative) d'une population, et F_0 est une distribution précisée préalablement par le chercheur.

Les hypothèses peuvent être très variables, mais H_0 postule toujours une absence de différence.

b. Choisir le seuil de confiance de la décision

Il existe toujours un risque d'erreur dont il est important de tenir compte.

Pour déterminer si la différence est significative (probablement pas due au hasard), il faut d'abord choisir un seuil qui définit le risque d'erreur qu'on est prêt à accepter. Ce seuil de signification définit la région de rejet de l'hypothèse nulle. Par convention, on utilise en général un seuil de 10%, 5% ou 1%.

L'hypothèse nulle sera rejetée si la différence observée a une probabilité très faible de se produire par pur hasard (inférieur au seuil).

Si l'hypothèse nulle est rejetée, l'hypothèse alternative (H_1) est acceptée, ce qui veut dire qu'il est permis de penser que les deux échantillons proviennent de populations différentes. Les deux populations sont alors significativement différentes pour au moins un de leur paramètre.

Comme il est impossible d'être toujours parfaitement confiant dans notre décision, nous choisissons à priori un seuil de confiance (α) (ex : 5%) tel que notre conclusion serait correcte à 95% du temps.

c. Chercher le test adéquat en fonction des hypothèses et du seuil

Il est nécessaire de connaître suffisamment la population pour pouvoir émettre des postulats. Est-elle distribuée normalement? Ou est-elle une suite qui n'obéit pas à la loi normale (suite d'essai de Bernoulli)?

Par souci de généralité, on préfère des postulats très généraux, nous postulons donc, grâce à l'approximation normale de la distribution binomiale, que les mesures sont normalement distribuées. Parfois, des approximations sont nécessaires. Bien identifier les postulats peut parfois être difficile, mais si on utilise un test alors que le postulat à ce test n'est pas satisfait, on peut tirer des conclusions qui ne seront pas fondées.

Un test statistique est une recette qu'on applique, il se formule toujours suivant une règle du genre : « Rejeter H0 si $k > s(\alpha)$ ». Dans cette formulation, k est une statistique (spécifiée par le test). La façon d'obtenir la valeur critique est aussi spécifiée par le test, et dépend uniquement du seuil choisi à priori et du type de test.

Un test correct garanti que la probabilité de rejeter H0 alors que H0 est vrai (dû à des fluctuations de l'échantillon) tout en supposant que l'échantillon est représentatif, choisi sans biais, et que des contrôles sont en place. Ces derniers facteurs (facteurs externes à la méthode statistiques) ne peuvent pas être quantifiés par le test statistique et sont donc de la responsabilité du chercheur (ce sont des facteurs liés à la méthode expérimentale).

Le test quantifie la probabilité d'obtenir une certaine différence par pur hasard. Si cette probabilité est faible (l'événement est improbable), nous concluons que la différence est significative.

Explication : si nous pouvions répéter l'expérience un grand nombre de fois. Il est certain que d'un échantillon à l'autre, la statistique (k) variera légèrement. En fait, si l'on fait effectivement ces nombreuses réplications, on pourrait trouver une valeur **z** tel que ($z > k$) dans 95% des situations. Dans ces cas, étant peu probable que ($k > z$) par pur hasard ; nous pensons alors, qu'elle est due à la manipulation expérimentale.

d. Appliquer le test et conclure

Le chercheur calcule à partir de son échantillon la bonne statistique k. Il trouve aussi (souvent à partir de tables) la valeur critique $s(\alpha)$. La conclusion découle de la comparaison entre la statistique empirique et la valeur critique.

II. principe des tests : hypothèse nulle

Les statistiques inférentielles permettent d'assigner une probabilité à l'obtention d'un résultat pour une hypothèse donnée. Si cette probabilité est trop faible, on rejette l'hypothèse.

Le principe des tests statistiques est de postuler l'hypothèse nulle : on fait de sorte que les différences observées entre des valeurs observées ou entre une valeur observées et une valeur théorique sont dues aux fluctuations d'échantillonnage.

A. La formulation d'hypothèses

La formulation d'hypothèses et l'évaluation de leur validité sont une branche importante des statistiques inférentielles. Afin de pouvoir décider entre plusieurs hypothèses possibles, on met en avant une hypothèse particulière que l'on appelle l'hypothèse nulle (notée H0) ; l'hypothèse alternative est notée H1.

En général H1 est le contraire de H0 mais il arrive que l'hypothèse H1 soit plus restrictive.

H1: $a = b$

H1: $a < b$

H1: $a > b$

Le résultat d'un test est de rejeter H0 en faveur de H1 ou bien ne pas rejeter H0. On ne conclut jamais par rejeter H1 et encore moins par accepter H1.

Les tests procèdent tous de la même manière : on dispose d'une variable de décision X qui suit une loi théorique P connue lorsque l'hypothèse H0 est vraie. On détermine donc une zone de rejet de probabilité qui délimite les valeurs acceptables que peut prendre X.

Si la valeur calculée pour la variable X se trouve dans la zone de rejet, on rejettera l'hypothèse H0 (avec un risque de le faire à tort). Cette zone est en deux parties pour un test bilatéral ou en une seule pour un test unilatéral. Tout dépend de la manière dont on a formulé l'hypothèse alternative H1.

Par exemple, si on teste les effets d'un vaccin avec l'hypothèse H0: le vaccin n'a pas d'effet, on ne se contente pas d'une hypothèse H1 qui dirait le vaccin a de l'effet car on souhaitera, en plus, que cet effet ne soit pas négatif et donc on exigera que la zone soit en un seul morceau.

B. Conditions de rejet de l'hypothèse nulle

Si la probabilité de l'hypothèse nulle est trop faible, on la rejette et on accepte l'hypothèse non nulle on conclut alors que les échantillons comparés proviennent de populations différentes. On dit alors qu'il existe une différence statistiquement significative.

Par convention, on fixe en général le seuil de signification à 5 %

$p<0,05$: différences statistiquement significatives

$p<0,01$: différences statistiquement très significatives

$p<0,001$: différences statistiquement hautement significatives

Le seuil de signification est déterminé avant d'effectuer le test ; le degré de signification est déterminé par le test (c'est la probabilité de rejeter l'hypothèse nulle).

La différence est significative si le degré de signification est inférieur au seuil de signification.

III. Les risques d'erreur

Lorsque nous portons un jugement sur l'hypothèse H0, il est facile de voir que cela comporte un risque. On dispose d'une information insuffisante et la prise de décision implique un double risque d'erreur. En fait, deux types d'erreurs sont possibles : l'erreur de type I et l'erreur de type II.

1. on peut décider que H0 est fausse alors qu'elle est vraie. Cela veut dire qu'on décide que la différence est significative alors qu'elle ne l'est pas vraiment. C'est le risque de première espèce.
2. on peut décider que H0 est vraie alors qu'elle est fausse. Cela veut dire qu'on décide que la différence est non significative alors qu'elle l'est raiment. C'est le risque de deuxième espèce.

L'erreur de type I est commise lorsque l'on rejette H0 alors qu'en réalité cette hypothèse est vraie. Elle est donc commise lorsqu'on croit que la différence observée découle d'une différence significative alors qu'elle est due au hasard. Cette erreur est heureusement quantifiable en termes de probabilité, puisqu'il s'agit du seuil de signification fixé. Le chercheur décide donc lui-même du risque à prendre de commettre l'erreur.

De même, il peut arriver de ne pas rejeter l'hypothèse nulle alors que cette hypothèse est fausse. Cette erreur est commise lorsqu'on croit que la différence observée (étant faible) est due au hasard alors qu'en fait elle résulte de différences dans les populations. Il s'agit alors d'une **erreur de type II**.

Les erreurs de type I et de type II sont étroitement liées. Ainsi, moins le seuil de signification est sévère (i.e. un seuil à 10% plutôt qu'à 5%) plus il est risqué de commettre l'erreur type I. Inversement, plus le seuil est sévère (i.e. un seuil à 1% plutôt que 5%), plus on risque de commettre l'erreur de type II.

Plus la probabilité observée (pobs) est petite, plus on doute H0. Si pobs est petite, il y a deux possibilités:

- soit (a) H0 est vraie, et un événement rare s'est passé,
- soit (b) H0 est fausse.

Le choix entre ces possibilités dépend de la manière de juger l'importance des deux types d'erreurs possibles:

Erreur de Type I: H0 est vraie, mais on la rejette.

Erreur de Type II: H0 est fausse, mais on l'accepte.

Les deux types de risques sont antagonistes.

Si on diminue le risque de 1re espèce, on augmente le risque de 2e espèce.

Étant donné que le risque de 2e espèce n'est pas connu (à la différence du risque de 1re espèce) en absence de différence significative, on ne peut pas conclure à l'absence de différence, car on ne contrôle pas le risque d'erreur attaché à cette conclusion.

Il y a une différence souvent oubliée entre ne pas conclure qu'il existe une différence, et conclure qu'il n'existe pas de différence.

Le risque de se tromper

Le risque de conclure à tort à une différence, c'est-à-dire le risque de conclure à une différence significative alors qu'il n'y en a pas, n'est pas le risque de 1re espèce. En effet, il s'agit de la probabilité de rejeter l'hypothèse nulle et que l'hypothèse nulle soit vraie.

Donc, le risque de conclure à tort dépend du seuil de signification choisi mais également de la probabilité à priori que l'hypothèse nulle soit vraie.

IV. La signification statistique ou (niveau de p (probabilité))

La signification statistique d'un résultat est une mesure estimée du degré auquel il est "vrai" (dans le sens du "représentant de la population"). Techniquement, la valeur du niveau de p représente un index décroissant de la fiabilité d'un résultat. Le haut niveau de p, ne montre pas, que la relation observée entre les variables dans l'échantillon est un indicateur fiable de la relation entre les variables respectives dans la population.

Spécifiquement, le niveau de p représente la probabilité de l'erreur qui est impliquée en acceptant notre résultat observé comme valide, c.-à-d., en tant que "représentant de la population."

Exemple : un de niveau de p de 0.05 (c.-à-d., 1/20) indique qu'il y a une probabilité de 5% que la relation entre les variables trouvées dans notre échantillon est un "échec". En d'autres termes, si nous répétions des expériences comme la nôtre l'un après l'autre, nous pourrions compter qu'approximativement dans chaque 20 répliques de l'expérience il y aurait d'une dans laquelle la relation entre les variables en question ne serait pas justifiée.

Dans beaucoup de domaines de recherche, le de niveau de p de .05 est habituellement admis comme le niveau d'erreur limite acceptable.

Comment déterminer qu'un résultat est "vraiment" significatif ?

Il n'y a aucune manière d'éviter le caractère arbitraire dans la décision finale quant à quel niveau d'importance sera traité le « vraiment significatif ». C'est-à-dire, le choix d'un certain niveau d'importance, au delà duquel les résultats seront rejetés ou considérés en tant qu'invalide ou est arbitraire.

Typiquement, dans beaucoup de sciences, les résultats qui rapportent le .05 de p sont considérés come la limite du statistiquement significatif ; mais il faut se rappeler, que ce niveau d'importance implique toujours une probabilité élevée de l'erreur (5%).

Des résultats qui sont significatifs au niveau de .01 de p sont généralement considérés comme statistiquement significatifs, et le .005 de p ou les niveaux de .001 de p sont considérés comme "fortement" significatifs. D'une façon générale, ces classifications ne représentent rien d'autres que des conventions arbitraires qui sont officieusement basées sur l'expérience de la recherche.

❖ L'intervalle de confiance

C'est l'intervalle autour de la moyenne calculée de l'échantillon dans lequel la moyenne de la population a une probabilité donnée de se trouver.

Exemple d'une intervalle de confiance à 95 % : la valeur moyenne de la population dont est issu l'échantillon a 95 chances sur 100 de se trouver dans l'intervalle.

Le niveau de signification ou la probabilité observée (pobs)

La valeur observée de T est **tobs**.

Le niveau de signification **pobs** donne la probabilité d'observer l´événement.

Plus pobs est petite, plus on doute que H0 soit vraie ; on le calcule comme si H0 était vraie.

On utilise souvent des niveaux conventionnels, tels que : alpha = 0.05, 0.01, 0.001 qui correspondent aux événements avec des probabilités de 1/20, 1/100, 1/1000.

On dit que l'on rejette H0 à niveau 0.05 si pobs < 0.05.

La notion de p-valeur

Les logiciels de calcul statistique expriment souvent le résultat d'un test en fournissant une grandeur appelée p-valeur (en anglais p-value).

La règle veut que l'on rejette l'hypothèse nulle H0 dès que la p-valeur est inférieure au risque choisi.

La p-valeur dépend donc des données de l'échantillon testé. Exprimer le résultat du test au moyen de cette valeur permet d'avoir une réponse indépendante du risque et de conclure simplement en comparant le risque à la p-valeur, et donc de pouvoir répondre pour diverses valeurs de risques.