

## 4. Variabilité due à l'erreur

Soumis par Éric Raufaste  
Dernière mise à jour : 28-12-2008

Objectif. Introduire la notion d'erreur sous différentes formes (erreur de mesure, fonction d'erreur, erreur statistique).

Pré-requis. Leçon : Dimensions de la variabilité en psychologie.

Résumé. La notion d'erreur dans les mesures vient de la physique où elle était initialement conçue comme l'écart entre une valeur vraie et une valeur observée. La notion de fonction d'erreur a été introduite au XIX<sup>e</sup> siècle par Laplace pour rendre compte des petits écarts constatés entre les prédictions des meilleurs modèles mathématiques et les observations de la réalité empirique que les modèles mathématiques étaient censés prédire.

### 1. Notion d'erreur de mesure

Initialement, les physiciens ont cherché à mesurer diverses grandeurs, comme par exemple la taille d'un objet, la durée d'un moment, le périmètre de la terre, etc. Or, ils se sont rapidement aperçus que si l'on répète plusieurs fois la mesure d'un même objet, on a tendance à trouver des résultats légèrement différents. Pour expliquer ces résultats, une idée naturelle consiste à considérer qu'il existe une valeur vraie de la mesure, mais que le processus de mesure étant entaché d'une multitude d'erreurs, plus ou moins grosses, la valeur observée est différente de la valeur vraie. On appelle alors erreur l'écart entre la valeur observée et la valeur vraie :

valeur observée = valeur vraie + erreur

À partir de là, le problème devient de séparer ce qui est la valeur vraie, celle que l'on recherche, de l'erreur. Séparer le bon grain de l'ivraie en quelque sorte.

Une première méthode pour cela, consiste à répéter les mesures de l'objet un grand nombre de fois. On peut alors espérer qu'en faisant la moyenne de toutes ces mesures, le résultat obtenu fournira une estimation à peu près correcte de la valeur vraie. Mais nous verrons plus loin que l'idée de multiplier le nombre de mesures de l'objet et d'en faire la moyenne ne met pas le chercheur à l'abri de différents biais.

À

### 2. Notion de fonction d'erreur

Le concept d'une fonction d'erreur a été introduit par le physicien français Pierre-Simon de Laplace.

Ce dernier était tenant d'une position déterministe très ferme, au point qu'il avait imaginé la notion d'un démon qui serait capable, ayant connaissance de toutes les positions et vitesses initiales de toutes les particules de l'univers, de pouvoir prédire tout l'état futur de l'univers. Laplace s'est notamment illustré pour avoir appliqué la mécanique newtonnienne au fonctionnement du système solaire avec un succès qui impressionna ses contemporains et établit le modèle d'un monde céleste régi avec la précision d'une horloge. Après Laplace, et jusqu'à une période relativement récente, on considérera que l'univers est entièrement prévisible.

Bien sûr, Laplace constate l'existence de petits désaccords entre les modèles mathématiques qu'il a déduits de la théorie newtonnienne et les données d'observations de la mécanique céleste. Pour traiter cette question, il introduit alors la notion de fonction d'erreur qui nous occupe ici. Cette dernière est un terme qui regroupe une multitude de choses qui, pour faire court, se résument à l'ensemble des sources de variations qui ne font pas partie du modèle mathématique du phénomène étudié lui-même.

Par exemple, si le chercheur essaie de prédire la position future d'une planète d'après les observations actuellement disponibles, il dispose d'un modèle mathématique du déplacement des planètes. En appliquant les équations de son modèle aux données d'observations, il va ainsi produire une position prédite. Même en supposant le modèle juste, si l'on regarde la position réelle de la planète dans le ciel au moment attendu, il y aura généralement de petites différences. Par exemple à cause de petites variations dans l'atmosphère terrestre qui faussent les images, ou encore d'erreurs de notations de l'observateur. L'écart entre la valeur prédite et la valeur observée entrera dans la fonction d'erreur.

Ainsi Laplace a proposé une décomposition d'une variable observée en une somme de deux autres variables, mais différente de la décomposition «valeur vraie + erreur» : il s'agit maintenant d'une décomposition valeur prédite d'après le modèle + fonction d'erreur ou résidu.

Quelle est la nature de cette fonction d'erreur ? En fait la fonction d'erreur combine une multitude de facteurs d'origines très différentes. Ainsi le résidu peut contenir une part liée à l'imprécision des instruments de mesure, une part liée à l'erreur humaine, etc.

Cette décomposition trouve aussi à s'appliquer aux mesures psychologiques mais pour certains types de données psychologiques, ce passage ne va pas de soi. Ces problèmes sont expliqués plus en détail dans la grande leçon consacrée à la psychométrie.

Pour le moment, il suffit de retenir que la décomposition de la valeur observée en valeur vraie + erreur n'est pas la seule possible. Techniquement la valeur vraie est souvent inconnue, voire même inconnaisable. La valeur prédite au contraire peut souvent être donnée a priori. Il est donc avantageux d'opérer une décomposition du type valeur observée = valeur prédite + fonction d'erreur.

À

### 3. Notion d'erreur statistique

Lorsque l'on essaie d'évaluer une valeur à partir d'un ensemble de mesures indépendantes, et qu'on fait la moyenne des observations obtenues, ce qu'on obtient n'est pas la valeur vraie de la mesure cherchée mais seulement une estimation de cette valeur.

Tout d'abord il peut y avoir des erreurs systématiques. Par exemple, si l'on mesure la largeur d'une planche avec une règle où les graduations sont trop rapprochées, on aura beau répéter les mesures un nombre infini de fois, le résultat moyen sera toujours plus grand que la largeur réelle.

Pour simplifier, imaginons que nous travaillons avec des instruments ne présentant pas de biais systématique. Il est alors possible de démontrer mathématiquement que si l'on répète un nombre infini de fois l'opération de mesure, et si l'on fait la moyenne de tous les valeurs obtenues, alors la valeur moyenne tendra vers la valeur réelle.

Seulement dans la vie réelle il est tout simplement impossible de répéter l'opération de mesure un nombre infini de fois. La moyenne d'un échantillon des valeurs tendra donc vers la moyenne réelle mais elle ne l'atteindra jamais. L'écart entre la vraie valeur et la valeur estimée à partir de l'échantillon constitue l'erreur statistique.

Il faut remarquer ici que l'erreur statistique est donc la différence entre une valeur hypothétique, qui n'existe qu'idéalement, et la moyenne des valeurs observées.