

TYPE DE CENSURE (F8)

(26 / 09 / 2020, © Monfort, Dicostat2005, 2005-2020)

Un **échantillon aléatoire** peut être censuré de plusieurs manières (cf **censure**).

(i) Ainsi, on considère un **espace probabilisé** (Ω, \mathcal{F}, P) , une **vars** $\xi : \Omega \mapsto \mathbf{R}$, un **échantillon iid** $X = (X_1, \dots, X_N)$ constitué de **copies** indépendantes et équidistribuées selon la **loi** P^ξ de ξ , et l'on note $X^{(\cdot)}$ l'échantillon ordonné associé à X (cf **statistique d'ordre**). Alors :

(a) étant donné un couple $(a, b) \in \mathbf{R}^2$ (connu ou non, aléatoire ou non), on appelle **censure de type I**, ou **censure de première catégorie**, ou encore **censure de type B**, de X la **situation statistique** dans laquelle on ne peut observer, ou n'utiliser, (eg pour l'**inférence statistique**) que le sous-échantillon $(X^{(\alpha)}, \dots, X^{(\beta)})$ tq :

$$(1) \quad \begin{aligned} &1 \leq \alpha \leq \beta \leq N, \\ &X^{(\alpha-1)} < a \leq X^{(\alpha)} \leq \dots \leq X^{(\beta)} \leq b < X^{(\beta+1)}, \end{aligned}$$

avec la convention selon laquelle $X^{(\alpha-1)} = -\infty$ si $\alpha = 1$, et $X^{(\beta+1)} = +\infty$ si $\beta = N$;

(b) étant donné une **proportion** $p \in]0, 1[$ (connue ou non), on appelle **censure de type II**, ou **censure de seconde catégorie**, ou encore **censure de type A**, de X la situation dans laquelle on n'observe ou n'utilise que le sous-échantillon $(X^{(\alpha)}, \dots, X^{(\beta)})$ tq :

$$(2) \quad \begin{aligned} &1 \leq \alpha \leq \beta \leq N, \\ &(N - \beta + \alpha - 1) / N = p, \end{aligned}$$

ie une proportion p des coordonnées de X parmi les **valeurs extrêmes** est « éliminée » (ie non utilisée, ou non observée).

Dans ce cas, on parle aussi d'**équilibre (bilatéral)** de l'échantillon X , ainsi que de toute **caractéristique** empirique pouvant s'en déduire (**moment**, **médiane**, etc).

Les deux types de censure précédents sont les plus classiques. Ces **censures** ont été supposées **bilatérales**, mais il est possible de définir des **censures** de type I ou de type II **unilatérales** (soit à droite, soit à gauche), en fonction du **problème statistique** considéré.

(ii) On distingue, par ailleurs, entre les censures de type aléatoire et les censures de type non aléatoire. Dans le premier cas, les « valeurs » censurantes (eg a et b) sont elles-mêmes des variables aléatoires, qui peuvent aussi varier en fonction des observations.