Théorie statistique de l'estimation

Lorsque que l'on peut caractériser les propriétés d'un échantillon, il est souvent très utile de remonter aux caractéristiques de la population. On utilise alors les principes de la théorie statistique de l'estimation qui s'appuient sur des méthodes *"d'inférence statistique"*.
L'inférence statistique consiste à induire les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population. Les caractéristiques de l'échantillon, une fois connues, reflètent avec une certaine marge d'erreur (la limite de confiance, le risque) celles de la population.
On pose μ_S et σ_S la moyenne et l'écart type d'une distribution d'échantilonnage d'une statistique S. Si la distribution d'échantilonnage est normale (c'est pratiquement toujours le cas pour des grands échantillons), on peut ramener cette distribution à une loi normale centrée réduite en réduisant l'écart

$$z = \frac{S - \mu_S}{\sigma_S}.$$

A partir de la fonction de répartition de la loi normale centré réduite, $\Phi(z)$, on peut alors déterminer les intervalles de confiances $[-z_c, z_c]$, des statistique S et μ_S. Ces intervalles de confiance sont souvent appelés coefficents de confiance ou valeurs critiques.

Niveau de Confiance	z_c
99.73 %	3.00
99 %	2.58
98%	2.33
96%	2.05
95.45%	2.00
95%	1.96
90%	1.645
80%	1.28
68.27%	1.00
50%	0.675

Intervalles de confiance des moyennes

Si la statistique S est la moyenne de l'échantillon $\bar{X}$, les limites de l'intervalle de confiance de la moyenne de la population sont

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}}$$

si l'échantillonnage est issu d'une population infinie, ou d'une population finie mais avec remise, ou

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$$

si l'échantillonnage est issu d'une population finie sans remise.
En prenant le cas avec remise comme exemple, on peut aussi écrire que

$$P(\bar{X} - z_c \frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{X} + z_c \frac{\sigma}{\sqrt{N}}) = \Phi(zc) - \Phi(-zc).$$

> Un échantillon aléatoire de 50 notes de mathématiques sur un total de 200 présente une moyenne de 75 et un écart type de 10.
> (a) Quelles sont les limites de l'intervalle de confiance à 95% de l'estimation de la moyenne des 200 notes ?
> (b) Quel est le degré de confiance avec lequel nous pouvons dire que la moyenne des notes est 75 ± 1?

> En mesurant des temps de réaction, un psychologue estime que l'écart type est de 0.05 s. Quelle devrait être la taille de l'échantillon de mesure pour que l'on soit sûr à
> (a) 95%
> (b) 99%
> que l'erreur sur le temps de réaction moyen ne dépasse pas 0.01 s.

Intervalles de confiance des pourcentages

Si la statistique S est un pourcentage de succès P dans un échantillon de taille N tiré d'une population binômiale dans laquelle p est le pourcentage de succés ($q = 1 - p$), les limites de l'intervalle de confiance du pourcentage dans la population sont

$$P \pm z_c \sqrt{\frac{pq}{N}}$$

si l'échantillonnage est issu d'une population infinie, ou d'une population finie mais avec remise, ou

$$P \pm z_c\sqrt{\frac{pq}{N}}\sqrt{\frac{N_p - N}{N_p - 1}}$$

si l'échantillonnage est issu d'une population finie sans remise.
<u>Note :</u> Souvent on utilise le pourcentage P de l'échantillon pour estimer p.
En prenantt le cas avec remise comme exemple, on peut aussi écrire que

$$P\left(P - z_c\sqrt{\frac{P(1-P)}{N}} \leq p \leq P + z_c\sqrt{\frac{P(1-P)}{N}}\right) = \Phi(zc) - \Phi(-zc).$$

55% des électeurs d'un échantillon aléatoire de 100 électeurs, choisi parmis tous les électeurs d'une région, est favorable à un candidat. Trouver les limites de l'intervalle de confiance à (a) 95%, (b) 99%, (c) 99.73% du pourcentage des électeurs de la région favorables à ce candidat. (2) Quelle devrait être la taille de l'échantillon d'électeurs pour être confiant à (a) 95%, (b) 99.73% que le candidat soit élu?

Sur 40 lancers d'une pièce, 24 "*face*" sortent. Trouver les limites de l'intervalle de confiance à (a) 95%, (b) 99.73% du pourcentage de "*face*" qui serait obtenu pour un nombre illimité de lancers de la pièce.

Intervalles de confiance des différences et des sommes

Si S_1 et S_2 sont deux statistiques d'échantillons présentant chacune une distribution d'échantillonnage normale, les limites de l'intervalle de conficance de la différence des paramètres correspondant sont données par

$$S_1 - S_2 \pm z_c\sigma_{S_1 - S_2} = S_1 - S_2 \pm z_c\sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}.$$

Les limites de l'intervalle de conficance de la somme des paramètres sont donnés par

$$S_1 + S_2 \pm z_c\sigma_{S_1 - S_2} = S_1 + S_2 \pm z_c\sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}.$$

Intervalles de confiance des différences de moyenne

$$\bar{X}_1 - \bar{X}_2 \pm z_c \sigma_{\bar{X}_1 - \bar{X}_2} = \bar{X}_1 - \bar{X}_2 \pm z_c \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

Un échantillon de 150 lampes de qualité A montre une durée de vie moyenne de 1400 heures avec un écart type de 120 heures. Un échantillon de 200 lampes de qualité B montre une durée de vie moyenne de 1200 heures avec un écart type de 80 heures. Donner les limites de l'intervalle de confiance de à (a) 95%, (b) 99% de la différence de la durée de vie moyenne entre les qualités A et B.

La force électromotrice moyenne de batteries produites par une entreprise est 45.1 V avec un écart type de 0.04 V. Si quatre de ces batteries sont connectées en série, trouver les limites de l'intervalle de confiance à (a) 95%, (b) 99%, (c) 99.73%, (d) 50% de la force électromotrice moyenne totale.

Intervalles de confiance des différences de pourcentages

$$P_1 - P_2 \pm z_c \sigma_{P_1 - P_2} = P_1 - P_2 \pm z_c \sqrt{\frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}}.$$

Dans un échantillon aléatoire de 400 adultes et de 600 adolescents qui ont regardé un programme de télévision donné, 100 adultes et 300 adolescents ont indiqué qu'ils avaient apprécié. Construire les limites de l'intervalle de confiance à (a) 95%, et (b) 99%, de la différence de pourcentage de l'ensemble des adultes et de l'ensemble des adolescents qui ont regardé le programme et l'ont apprécié.