Nature du problème

Il s'agit de chercher des informations sur une population d'effectif important à partir de l'étude d'un échantillon de quelques dizaines d'unités. Ce type de situation se rencontre fréquemment dans le monde industriel, car le plus souvent il n'est pas possible d'étudier la population entière.
La méthode consiste tout d'abord à proposer un nombre comme moyenne, proportion ou écart-type de la population. C'est l'estimation ponctuelle.
Le résultat n'étant pas toujours utilisable de façon satisfaisante on utilise alors la notion d'intervalle de confiance associé à un coefficient de confiance.

A) Estimation ponctuelle.

1 Moyenne:

D'une manière générale on choisit la moyenne $\overline{x_e}$ d'un échantillon prélevé au hasard dans une population comme estimation ponctuelle de la moyenne inconnue m de cette population: $m = \overline{x_e}$

2 Proportion:

D'une manière générale on choisit la proportion f_e des éléments possédant une certaine propriété dans un échantillon prélevé au hasard dans une population comme estimation ponctuelle de la proportion inconnue p des éléments de cette population ayant cette même propriété: $p = f_e$

3 Variance , écart-type :

D'une manière générale on choisit le nombre $\frac{n}{n-1}\sigma_e^2$ où n est l'effectif et σ_e^2 la variance d'un échantillon prélevé au hasard dans une population comme estimation ponctuelle de la variance inconnue σ^2 de cette population.

$$\sigma^2 = \frac{n}{n-1}\sigma_e^2$$

De même on choisit le nombre $\sqrt{\frac{n}{n-1}}\,\sigma_e$ où n est l'effectif et σ_e l'écart-type d'un échantillon prélevé au hasard dans une population comme estimation ponctuelle de l'écart-type inconnu de cette population.

Cette valeur est appelée déviation standard et souvent notée *s*. $s = \sigma = \sqrt{\frac{n}{n-1}}\,\sigma_e$

B) Estimation par intervalle de confiance.

L'estimation ponctuelle de la moyenne d'une variable aléatoire ou de la fréquence d'un caractère sur une population ne donne pas de renseignement sur la valeur de cette estimation. On précisera donc la probabilité pour que ce paramètre se trouve dans un intervalle centré sur la valeur d'estimation.

1 Moyenne :

En général, on peut considérer que la variable aléatoire $\overline{X}$ qui à tout échantillon de taille *n* fixée, associe la moyenne de cet échantillon, suit la loi normale $N\,(\mu\ ,\ \sigma_{\overline{X}})$, avec $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$, *m* et σ étant la moyenne et l'écart-type de la population et *n* l'effectif des échantillons. (cf distribution d'échantillonnage des moyennes).

$T = \dfrac{\sqrt{n}}{\sigma}\,(\ \overline{X}\ - m\ )$ suit la loi normale centrée réduite $N\ (\ 0\ ,\ 1\ )$.

Un coefficient de confiance $1-\alpha$ choisi à l'avance (α étant le risque) permet de définir un nombre positif t tel que : $p\,(-t \le T \le t)$ soit égal à ce coefficient de confiance.

On a donc $p\,(-t \le T \le t) = 1-\alpha \Leftrightarrow p\left(-t \le \dfrac{\sqrt{n}}{\sigma}(\overline{X} - m) \le t\right) = 1-\alpha$

$$\Leftrightarrow p\left(-t\frac{\sigma}{\sqrt{n}} \le \overline{X} - m \le t\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\Leftrightarrow p\left(m - t\frac{\sigma}{\sqrt{n}} \le \overline{X} \le m + t\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

En conséquence: $p\left(\overline{x}_e - t\dfrac{\sigma}{\sqrt{n}} \le \overline{X} \le \overline{x}_e + t\dfrac{\sigma}{\sqrt{n}}\right) = 1-\alpha$

L'intervalle $\left[\ \overline{x}_e - t\dfrac{\sigma}{\sqrt{n}}\ ;\ \overline{x}_e + t\dfrac{\sigma}{\sqrt{n}}\ \right]$ est donc l'intervalle de confiance de la moyenne *m* de la population avec le coefficient de confiance $1-\alpha$ ayant pour centre la moyenne $\overline{x}_e$ de l'échantillon considéré.

Remarques:

- Pour calculer les extrémités de cet intervalle de confiance, l'écart-type σ de la population doit être connu. Dans certains cas, notamment lorsque l'effectif *n* de l'échantillon est grand, on peut prendre pour valeur de σ son estimation ponctuelle : $s = \sigma = \sqrt{\dfrac{n}{n-1}}\ \sigma_e$, σ_e étant l'écart-type de l'échantillon.

- On ne peut pas savoir si la moyenne m de la population appartient ou non à l'intervalle de confiance associé au seul échantillon prélevé.
- De plus, si m (moyenne de la population) appartient à cet intervalle, on ne sait pas où m se situe dans cet intervalle centré sur $\bar{x}_e$ (moyenne de l'échantillon).
- On peut retenir : si $1-\alpha = 0,95$ alors $t = 1,96$ et si $1-\alpha = 0,99$ alors $t = 2,58$

2 Proportion :

A l'aide d'un échantillon, nous allons définir, avec un coefficient de confiance choisi à l'avance, un intervalle de confiance de la proportion p des éléments de la population possédant une certaine propriété.
Considérons le cas où la variable F qui, à tout échantillon aléatoire de taille n fixée, associe la proportion des éléments de cet échantillon possédant la propriété considérée, suit la loi normale $\mathcal{N}\left(p, \sqrt{\frac{pq}{n}}\right)$ avec $q = 1 - p$.

En conséquence:

L'intervalle $\left[f_e - t\sqrt{\frac{pq}{n}} ; f_e + t\sqrt{\frac{pq}{n}} \right]$ est l'intervalle de confiance d'une proportion p de la population avec le coefficient de confiance $1-\alpha$ ayant pour centre la proportion f_e de l'échantillon considéré.

EX 1

Une société s'approvisionne en pièces brutes qui, conformément aux conditions fixées par le fournisseur, doivent avoir une masse moyenne de 780 g.
Au moment où 500 pièces sont réceptionnées, on en prélève au hasard un échantillon de 36 pièces dont on mesure la masse.
On obtient les résultats suivants:

Masse en grammes	[745 ; 755 [	[755 ; 765 [	[765 ; 775 [	[775 ; 785 [	[785 ; 795 [	[795 ; 805]
Nombre de pièces	2	6	10	11	5	2

a) A combien peut-on estimer la moyenne et l'écart-type des masses pour la population constituée des 500 pièces à l'aide des résultats obtenus sur cet échantillon ?
b) On prélève maintenant une succession d'échantillons de même effectif $n = 36$.
Soit $\bar{X}$ la variable aléatoire qui à tout échantillon prélevé associe sa moyenne. $\bar{X}$ suit la loi normale $\mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)$.
On veut obtenir un intervalle qui contienne la moyenne de la population avec une confiance de 95 %.
Déterminer t tel que p $(-t \leq T \leq t) = 0,95$, T suivant $\mathcal{N}(0,1)$.
Déterminer un encadrement de m avec le coefficient de confiance 95 %. (On prendra comme valeur de σ la valeur estimée en a)

EX 2

Une entreprise utilise des camions pour transporter sa production. Elle dispose de 100 camions.
Elle repère sur un échantillon de 30 jours choisis au hasard le nombre de camions en panne. Les résultats sont :
5 5 6 4 6 6 8 3 5 5 5 4 3 6 5 6 4 7 6 6 5 4 3 6 5 4 5 4 5 5

1. Calculer la moyenne $\overline{x}_e$ et l'écart-type σ_e du nombre de camions en panne chaque jour pour l'échantillon considéré.
2. A partir des résultats obtenus pour cet échantillon, proposer une estimation ponctuelle de la moyenne μ et de l'écart-type *s* du nombre de camions en panne chaque jour pour la population.
3. On suppose que la variable aléatoire $\overline{X}$ qui, à tout échantillon de taille 30 prélevé au hasard et avec remise, associe la moyenne du nombre de camions en panne chaque jour suit la loi normale

$N\ (\mu\ ,\ \frac{s}{\sqrt{n}}\ )$. Déterminer un intervalle de confiance de la moyenne μ de la population avec le coefficient de confiance 95 %

4. Même question avec le coefficient de confiance 99 %.

EX 3

Un candidat à une élection fait effectuer un sondage dans sa circonscription comportant 85842 électeurs: sur 1068 personnes interrogées, 550 déclarent vouloir voter pour ce candidat.
On suppose que cet échantillon peut être assimilé à un échantillon prélevé au hasard dans la population des électeurs de la circonscription. Soit F la variable aléatoire qui à tout échantillon de taille $n = 1068$ prélevé au hasard dans cette population associe le pourcentage d'électeurs de cet échantillon voulant voter pour le candidat.

On suppose que F suit la loi normale $N\ (p\ ,\ \sqrt{\frac{pq}{n}}\ )$, où *p* est le pourcentage inconnu des électeurs de la circonscription voulant voter pour le candidat et $q = 1 - p$.
Déterminer un intervalle de confiance de *p* avec le coefficient de confiance 95 %
Au vu du résultat de ce sondage, le candidat a-t-il raison de penser que si les élections avaient eu lieu au moment où le sondage a été réalisé et si les réponses au sondage étaient sincères, il aurait été élu au premier tour ?