L'échantillonnage

Monsieur Gauss

I Rappel : la loi normale

Le fait que des valeurs soient distribuées selon une loi normale est fréquent dans les processus de fabrication industriels.
En effet, on cherche à atteindre une certaine valeur moyenne $\bar{x}$: vis de longueur 4 cm, pots d'encre de 1 kg, plaques offset d'épaisseur 0,2 mm…
Mais, en raison de petites imperfections dans la chaîne de fabrication, les articles obtenus s'écartent plus ou moins de la valeur moyenne.

Exercice pour voir où vous en êtes…

1) X suit $N(0 ; 1)$. Calculez $p(X < 1{,}57)$; $p(0{,}34 < X < 2{,}12)$; $p(-1{,}22 < X < 1{,}22)$

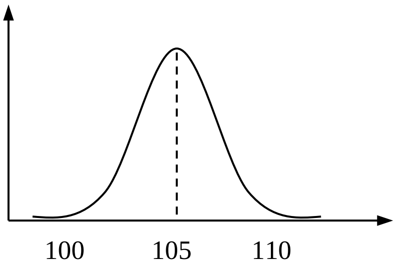
2) Y suit $N(25 ; 3{,}6)$. Calculez $p(Y < 27{,}9)$; $p(23{,}2 < Y < 26{,}8)$; $p(23{,}5 < Y < 26{,}5)$

II Echantillons

La situation classique est celle où l'on cherche à vérifier si la production d'un produit répond aux critères fixés par le client au départ.

Exemple : Fabrication de lampes de 100 W

On espère à la sortie de la chaîne de fabrication une répartition du genre :



o.k. ?

Pour vérifier la qualité de la production, on prélève des **échantillons** de n lampes (avec $n \geq 30$ car….) pour lesquels on mesure la puissance.

On calcule alors la puissance moyenne des lampes de l'échantillon : m

Question : La valeur de m sera-t-elle égale à 105 W ??

III Lois d'échantillonnages

1) Loi d'échantillonnage de la moyenne

On considère ici une variable aléatoire X caractérisant un article.
On suppose que X suit une loi normale de moyenne $\bar{x}$ et d'écart-type σ.

Lorsque l'on mesure la moyenne sur des échantillons nommés E1, E2, E3… on obtient les valeurs notées : m1, m2, m3…

n° de l'échantillon	1	2	3	…
moyenne calculée	m1	m2	m3	…

Ainsi, on peut définir une nouvelle variable aléatoire notée M.

M représente la moyenne observée sur un échantillon. C'est bien une variable aléatoire puisque l'on ne peut deviner *a priori* la valeur de la moyenne pour un échantillon donné. La preuve en a été apportée !

Théorème

$$\left\{ \begin{array}{l} \text{M est une variable aléatoire qui suit la loi normale } N\left(\bar{x} \; ; \; \frac{\sigma}{\sqrt{n}} \right) \\ \bar{x} = \text{moyenne sur l'ensemble de la population} \\ n = \text{nombre d'éléments dans chaque échantillon} \\ \sigma = \text{écart-type de la population} \end{array} \right.$$

Pour ce théorème, on considère que les échantillons sont constitués d'éléments tirés *au hasard* et, de plus, que le tirage est *non-exhaustif* , ou avec remise…
Là où on pioche, il en reste assez pour piocher encore….

Application

La tension des piles rectangulaires de petit format est indiquée à 9 Volt.
On suppose dans cet exemple que la fabrication de ces piles suit une loi normale de moyenne 9,5 V et d'écart-type égal à 0,33 V.

Ainsi, si l'on considère des lots de n = 30 piles prélevées au hasard, on peut calculer la tension moyenne notée T. D'après le théorème précédent, on sait que T est une variable aléatoire qui suit la loi normale $N\left(9,5 \; ; \; \frac{0,33}{\sqrt{30}}\right)$ ce qui donne $N(9,5 \; ; \; 0,06)$.

Un client explique avoir contrôlé un lot avec une moyenne T = 9,31. Il se plaint de cette valeur anormalement basse selon lui…

Calculons alors la probabilité qu'un lot de 30 piles donne une moyenne T < 9,32 pour tenter d'apprécier le caractère « anormal » de cette affaire…

$$p(\, T < 9{,}32) \;=\; p\left(\frac{T - 9{,}5}{0{,}06} < \frac{9{,}32 - 9{,}5}{0{,}06} \right) \;=\; p\left(\frac{T - 9{,}5}{0{,}06} < -3 \right)$$

$$= \; p(\, Y < -3 \,) \quad \text{où Y suit la loi normale centrée réduite } N(\, 0 \; ; \; 1 \,)$$

$$= \; 1 - \Pi(\, 3 \,) \; = \; 1 - 0{,}99865 = \; 0{,}00135$$

Remarque : La calculatrice permet de trouver rapidement le même résultat…

Ainsi, ce calcul montre une possibilité de 0,135% de chance de tomber sur un lot ayant une tension moyenne si basse… Finalement, le client semble sincère, la valeur moyenne est effectivement « anormalement » basse !

Exercice I

Une production en série d'articles est réglée pour un poids moyen de 250 g et un écart-type de 10 g.

Quelle est la loi suivie par le poids moyen mesuré sur des échantillons de 40 articles ?

Calculez la probabilité d'observer un poids moyen compris entre 247 et 253 grammes.

Calculez la probabilité d'observer un poids moyen qui soit inférieur à 247 g ou bien supérieur à 253 g.

2) Loi d'échantillonnage pour la fréquence

On considère ici une situation présentant seulement deux possibilités dont on connaît les probabilités respectives.
Le cas typique est celui de la fabrication d'articles. On mesure sur la production d'un grand nombre d'articles la proportion des articles défectueux et celle d'articles corrects.
On peut aussi travailler sur le pourcentage de clients satisfaits par un nouveau produit…
On choisit alors au hasard des échantillons d'articles et on note pour chacun la proportion, la fréquence d'articles défectueux.

<u>Cette fréquence est encore un nombre aléatoire qui suit une loi normale.</u>

Théorème

Si un caractère étudié présente seulement deux éventualités A et B, l'une « favorable » et l'autre pas, dont les proportions respectives sont notées p et q (avec q = 1 – p), la loi d'échantillonnage suivie par la fréquence des cas « favorables » est la loi normale de moyenne p et d'écart-type $\sqrt{\frac{p(1-p)}{n}}$

C'est la loi $N(p\ ;\ \sqrt{\frac{p(1-p)}{n}})$

n est la taille de l'échantillon

Cette loi est toujours rappelée dans un examen…

Etudions un exemple

Un assureur constate qu'en moyenne 78% de sa clientèle est satisfaite. Il souhaite proposer un nouveau service et choisit au hasard 30 de ses clients pour une étude de marché.
Calculons la probabilité qu'au sein de ce groupe « test » il y ait moins de 60 % de clients satisfaits (ce qui poserait trop de problèmes…).
Notons F la variable aléatoire donnant la proportion de clients satisfaits au sein d'un échantillon de 30 clients.

D'après le cours : F suit la loi $N\left(0{,}78\ ;\ \sqrt{\frac{0{,}78\,(1-0{,}78)}{30}}\right)$ soit $N(0{,}78\ ;\ 0{,}08)$

On peut alors calculer p(F < 0,6) = ? Montrez que p(F < 0,6) = 0,01

Exercice II

Un imprimeur se rend compte d'une erreur survenue pendant la reliure d'une série de livres. D'après les retours récents, il parvient à estimer que 15 % de sa production est défectueuse.

Avant d'envoyer un lot de 200 de ces livres, il préfère envisager de contrôler ce lot pour s'éviter des retours inutiles (et toujours délicats en terme d'image de marque…).

1) On note **f** la variable aléatoire donnant la proportion de livres défectueux lorsque l'on considère des échantillons de n = 200 livres. Déterminez la loi suivie par **f**.

2) Calculez p(**f** < 0,05). Justifiez qu'il est préférable de contrôler ce lot après avoir expliqué la signification de « p(**f** < 0,05) ».

3) Calculez p(0,1 < **f** < 0,2)

4) En déduire un encadrement du nombre de livres défectueux qu'il doit retirer pour être pratiquement certain que le lot devienne correct.

5) Pendant le contrôle, l'imprimeur arrive à 25 livres retirés. Peut-il alors arrêter le contrôle en se déclarant satisfait ?

Dans ce cours, on connaît parfaitement la population (sa moyenne et son écart-type). Donc, on arrive à « prédire » certaines choses sur les échantillons étudiés.
A l'inverse, si l'on regarde très attentivement un échantillon, peut-on dire des choses sur la population de départ ?
C'est le cours « Statistiques inférentielles » qui permet de répondre à cette question…