Centre de Ressources
Comptabilité Finance
Lycée MARIE CURIE
Avenue du 8 mai 1945 - BP 348
38435 ECHIROLLES cedex
http://crcf.ac-grenoble.fr/

4- LES OUTILS D'AMÉLIORATION DES PERFORMANCES
Distribution d'échantillonnage et estimation

Remarque : une table de la loi normale centrée réduite est fournie à la fin de ce document

Ce que dit le programme de l'UE 8 :

4. Les outils d'amélioration des performances

Compétences attendues	Savoirs associés
- Exploiter les outils de gestion de la qualité. - Rédiger une note de synthèse sur la gestion de la qualité et identifier des solutions aux éventuels problèmes détectés.	- Contrôle statistique de la qualité, estimation ponctuelle et par intervalle de confiance d'une moyenne, d'une proportion.

Cadrage
Le calcul d'un intervalle de confiance ne sera étudié que pour une moyenne et pour une proportion. Les tests d'hypothèse ne seront pas abordés.

Explications complémentaires

Les produits ou services d'une entreprise doivent répondre à la qualité attendue par le client, respecter la règlementation et les engagements du fournisseur (par exemple sur la quantité d'un ingrédient dans un plat cuisiné).

Lorsqu'un produit est fabriqué en masse, il est souvent impossible de contrôler la conformité de la totalité de la production afin de vérifier que chaque produit respecte la qualité attendue.

Il est toutefois possible d'avoir recours à des outils statistiques nous permettant de nous assurer que la qualité de la quasi-totalité des produits est conforme aux attentes, sans avoir à contrôler individuellement chaque produit.

C'est l'objet de cet écrit qui vous propose de découvrir et de mettre en application ces outils à caractère mathématique.

Nous allons aborder deux types de problème qui relèvent de ce qu'on appelle l'induction statistique (Branche des *statistiques* permettant de tirer des conclusions sur une population à partir de l'étude d'un échantillon représentatif, aussi appelé statistique inférentielle).

On considère une population P et l'on s'intéresse à une caractéristique de cette population (diamètre d'un objet, poids d'un produit, etc.). De cette population, appelée population-mère (de taille N, un nombre entier qui peut être infini), on considère un échantillon de taille n (n étant un nombre entier).

Deux catégories de problème peuvent se poser :

- Soit on cherche, connaissant la valeur de certains paramètres dans la population-mère, à induire des renseignements sur les valeurs que peuvent prendre ces paramètres dans l'échantillon. Cette catégorie de problèmes est appelée **problèmes de distribution d'échantillonnage**. En gestion de la qualité, cela permet par exemple de vérifier qu'une machine est bien réglée, compte tenu des valeurs attendues de l'échantillon, ou encore qu'une production semble conforme aux attentes.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

- Soit au contraire, on cherche, connaissant la valeur de certains paramètres dans l'échantillon, à induire des renseignements sur les valeurs que peuvent prendre ces paramètres dans la population-mère. Cette seconde catégorie de problèmes est connue sous le nom de **problèmes d'estimation**. En gestion de la qualité, cela permet par exemple de vérifier qu'un nouveau produit permet d'avoir des performances significativement supérieures à un produit qu'il remplace. On peut par exemple étudier si l'utilisation d'une nouvelle huile moteur réduit la consommation de carburant ; ou encore qu'un nouveau modèle de pneu permet de parcourir significativement plus de kilomètres qu'un ancien modèle.

Le schéma suivant résume ces deux catégories de problème :

Distributions d'échantillonnage	Estimation
À partir de la population mère, on recherche les paramètres attendus de l'échantillon.	À partir des résultats obtenus sur un échantillon, on « estime » les paramètres de la population mère.

I. Approche de la distribution d'échantillonnage par l'exemple

Pour découvrir cette notion, nous partirons d'une application.

Application 1 – étude des échantillons d'une population

Une entreprise sous-traitante fabrique des composants de haute précision pour l'industrie aéronautique. Elle fabrique notamment des axes en acier dont la longueur attendue en millimètre est de 100.

Une très faible tolérance est admise par le client.

Même si dans la réalité, l'entreprise fabrique environ 1000 axes par jour (taille de la population-mère), nous allons dans un premier temps nous intéresser à une production de 5 unités. L'objectif étant d'observer ce qui se passe dans les échantillons de différentes tailles prélevés dans une population. Les longueurs obtenues sont les suivantes : (les axes sont notés A, B, C, D et E pour les différencier)

Population	
Axes en acier	Longueur en mm
A	99,860
B	99,880
C	100,160
D	100,140
E	99,960
Total	**500,000**
Moyenne de la population	**100,000**
Ecart-type	**0,127**

Travail à faire :

1) Recenser tous les échantillons de taille 4 et calculer la moyenne de chaque échantillon. Puis calculer la moyenne des moyennes de l'ensemble des échantillons, ainsi que l'écart-type.
2) Faire de même avec tous les échantillons de taille 3 puis de taille 2, puis de taille 1.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

3) Que constate-t-on ?

Imaginons que nous prenons au hasard un échantillon de 2 axes, et que la production totale (les cinq axes) est acceptée à condition que la moyenne des longueurs soit comprise entre 99,90 et 100,10.

4) Dans combien de cas, la production totale sera acceptée et dans combien de cas elle ne le sera pas ? Prenons-nous toujours la bonne décision sur la base d'un échantillon ?

Remarque : ce travail peut être mené à l'aide d'un tableur. La fonction à utiliser pour l'écart-type est **ECARTYPEP(plage de cellules)**. Il y a en effet d'autres fonctions écart-type pour d'autres utilisations.

Cette fonction calcule l'écart-type d'un ensemble de valeur ; le P signifie population.

Qu'avez-vous dû constater ?

- La moyenne des moyennes d'échantillon est toujours égale à la moyenne de la population (100 mm).
- Plus la taille de l'échantillon est grande plus la dispersion des moyennes est faible (se mesure à travers l'écart-type). Donc si l'on veut se rapprocher de la moyenne de la population, il faut choisir un échantillon suffisamment grand. Ce qui est bien sûr intuitif.

Ces constats nous incitent à nous poser différentes question :

- jusqu'à quel point la moyenne d'un échantillon unique s'approche de la moyenne de la population et ainsi tirer des conclusions sur celle-ci, du type elle est conforme à la qualité attendue ou elle ne l'est pas ?
- connaissant la moyenne de la population, si on prélève un échantillon, comment peut-on affirmer avec un risque limité qu'il appartient à la population ? (si une machine est déréglée, nous pouvons peut-être le voir à travers un échantillon ?).

La théorie va nous apporter des éléments de réponse. Il nous faut donc désormais formaliser.

II. La distribution d'échantillonnage – Formalisation et applications

A. La notion d'échantillon aléatoire

Hypothèses et variables :

- On considère une population P et l'on s'intéresse à une caractéristique de cette population (diamètre d'un objet, poids d'un produit, etc.).
- Cette caractéristique peut être quantifiée. On désigne par x le nombre associé à cette caractéristique.
- Si on prélève au hasard un élément (ou individus) de la population P, le nombre x peut être considéré comme la réalisation d'une variable aléatoire X, appelée variable parente, de loi L (L est la loi de probabilité suivie par X, qui peut être par exemple la loi normale).
  - Quel que soit l'individu choisi dans P, x sera la réalisation d'une variable aléatoire de loi L.
- De cette population, appelée population-mère (de taille N, un nombre entier qui peut être infini), on prélève n individus dans la population. On dit alors que l'on procède à un sondage aléatoire en prélevant un échantillon de taille n (n étant un nombre entier).
  - On dispose donc de n valeurs qu'on appelle un n-échantillon empirique ou aléatoire : $(x_1,x_2,x_3,\ldots,x_n)$.
  - Chaque x_i est la réalisation d'une variable aléatoire X_i. On peut donc considérer $(x_1,x_2,x_3,\ldots,x_n)$ comme la réalisation de $(X_1, X_2, X_3, \ldots, X_n)$.

Le n uplet $(x_1,x_2,x_3,\ldots,x_n)$ s'appelle un n échantillon aléatoire. (Un n échantillon est un échantillon de taille n).

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Il existe différents types d'échantillon, nous étudierons ces deux types :

La constitution de l'échantillon peut s'effectuer **avec remise** (l'individu choisi peut l'être plusieurs fois) ; l'échantillonnage est dit **non exhaustif**.

L'échantillon peut être obtenu **sans remise** (l'individu choisi ne peut l'être qu'une seule fois); l'échantillonnage est dit **exhaustif.**

Nous traiterons essentiellement des échantillons constitués avec remise ou considérés comme tels. En effet, lorsque la population est très grande, les échantillons peuvent être considérés comme exhaustifs, puisque la probabilité de retirer le même individu après la remise est quasiment nulle.

Application 2 – exemple

Nous reprenons les données de l'application 1

Travail à faire :

1) Quelle est la caractéristique étudiée de la population ?
2) Quelles sont les valeurs de cette caractéristique ?
3) Les échantillons étudiés dans l'application 1 sont-ils avec ou sans remise ?
4) Donner des exemples d'échantillons aléatoires de taille 2.

B. Définition de la distribution d'échantillonnage des moyennes

Pour donner la définition d'une distribution d'échantillonnage des moyennes, il nous faut d'abord définir la variable aléatoire qui prend pour valeurs les différentes moyennes des échantillons de taille n. Il est d'usage de noter Tn cette variable aléatoire, nous avons :

$Tn = (X_1 + X_2 + X_3 + \ldots + X_n) / n$ (Tn est parfois noté $\overline{X}_n$)

Tn est ce qu'on appelle une statistique construite sur l'échantillon de taille n (X_1, X_2, X_3, …, X_n). Tn prend différentes valeurs en fonction des valeurs prises par les variables aléatoires X_1, X_2, X_3, …, X_n. Nous pouvons écrire que :

Tn :	{ensemble des échantillons de taille n}	⇨	R	(ensemble des nombres réels)
	Ei	⇨	E(Ei)	(moyenne de l'échantillon Ei)

Application 3 – exemple

Nous reprenons les données de l'application 1

Travail à faire :

1) Quelles sont les différentes valeurs prises par la variable aléatoire T_3, c'est-à-dire les différentes moyennes de chaque échantillon de taille 3 ?
2) Quelle est la loi de probabilité suivie par T_3 ? (c'est-à-dire les différentes probabilités élémentaires des valeurs prises par T_3).

Les valeurs prises par Tn sont aléatoires, elles dépendent de l'échantillon, aussi, on appelle distribution d'échantillonnage la distribution de probabilité de la variable aléatoire Tn construite sur l'échantillon (X_1, X_2, X_3, …, X_n). Tn suit une certaine loi de probabilité, qui dépend de la loi suivie par X (variable parente définie précédemment), cette loi peut être quelconque ou une loi connue comme la loi normale.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Les paramètres de la distribution d'échantillonnage des moyennes :

- **Cas des échantillons non exhaustifs (tirage avec remise)**

Espérance mathématique de Tn :	**E(Tn) = E(X)**
Elle se traduit par la moyenne des moyennes des échantillons et est égale à la moyenne de la population comme vous l'avez constaté dans l'application 1. Cette propriété est également vraie pour les tirages sans remise.	
L'écart-type de Tn :	**$\sigma(Tn) = \sigma(X) / \sqrt{n}$**

Traduction : si on considère l'ensemble des échantillons de taille n, la moyenne des moyennes des échantillons est la moyenne de la population. Mais cela ne veut pas dire que si l'on considère un seul échantillon, la moyenne sera celle de la population, mais on pourra par exemple dire qu'avec un risque d'erreur de tant de pourcent, l'échantillon appartient bien à la population (voir applications ci-après).

Signification de l'écart-type de Tn : Il signifie simplement que plus l'échantillon est grand plus l'écart à la moyenne est faible (plus l'échantillon est grand, plus la moyenne de l'échantillon se rapproche de la moyenne de la population, ce qui est bien sûr intuitif).

Fondamental :

- **La loi suivie par Tn dépend de la loi suivie par X.**
- **Si X suit une loi normale (N) de moyenne m (E(X)) et d'écart-type σ (σ(X)) alors Tn suit une loi normale N (m ; $\frac{\sigma}{\sqrt{n}}$).**
- **ou si la loi de X n'est pas connue, mais que n est suffisamment grand (en pratique n >= 30), alors Tn suit également une loi normale N (m ; $\frac{\sigma}{\sqrt{n}}$). Ce résultat est une application du théorème central limite.**

Application 4

La SVR, Société Vendéenne de Roulements, fabrique différents types de roulement à bille pour l'industrie. Lorsque la machine est bien réglée, le roulement à bille référence R005 produit, a un diamètre moyen de 50 millimètres avec un écart-type de 1 millimètre.

L'entreprise produit chaque heure 1000 roulements à bille. Pour vérifier que la production est conforme aux attentes (et donc que la machine est bien réglée et ne nécessite pas un arrêt pour maintenance), le responsable qualité prélève chaque heure un échantillon de 50 roulements et mesure le diamètre de chacun d'eux. Compte tenu de la taille de la population (1000 roulements), le tirage peut être considéré comme un tirage avec remise (échantillon non exhaustif).

On appelle X la variable aléatoire qui correspond au diamètre d'un roulement à bille prélevé au hasard. Nous avons donc E(X) = 50 et σ(X) = 1.

Nous faisons l'hypothèse que X suit une loi normale : X → N(50 ; 1).

On note T_{50} = (X1 + X2 + X3 + …+ X50) / 50 la variable aléatoire (VA) qui a un échantillon de taille 50 associe sa moyenne. Les VA Xi sont indépendantes entre-elles.

Travail à faire :

1) Préciser la loi de probabilité suivie par la VA T_{50}.
2) L'hypothèse de normalité de la population (X suit une loi normale) était-elle nécessaire pour définir la loi suivie par T_{50} ?

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

3) Quelle est la probabilité a priori que la moyenne d'un échantillon de taille 50 qui sera prélevé au hasard sur une production de 1000 roulements soit comprise entre 49,7256 et 50,2744 millimètres ? Quelle signification pouvons-nous donner à l'intervalle [49,7256 ; 50,2744] ?
4) Après prélèvement d'un échantillon aléatoire de taille 50, nous trouvons une moyenne de 50,5 millimètres. Quels commentaires pouvons-nous faire ?

➢ **Cas des échantillons exhaustifs (tirage sans remise)**

Espérance mathématique de Tn : **E(Tn) = E(X)**

Ce qui se traduit par la moyenne des moyennes des échantillons qui est elle-même égale à la moyenne de la population.

Variance de Tn : **V(Tn) = [(N-n) / (N-1)] x V(X) / n**

L'écart-type de Tn est obtenu en prenant la racine de sa variance.

(N-n) / (N-1) est appelé facteur d'exhaustivité. N étant la taille de la population et n la taille de l'échantillon.

Important : *quand la population est infinie, ou quand n est petit par rapport à la taille de la population, le facteur d'exhaustivité est très proche de 1. On estime alors que V(Tn) = V(X) / n. en d'autres termes, les tirages sans remise sont équivalents à des tirages avec remise. On applique donc le cas précédent. C'est par exemple le cas lorsqu'on prélève un échantillon d'une production en quantité importante.*

Nous nous ramènerons donc dans la grande majorité des applications au cas des échantillons non exhaustifs (tirage avec remise)

C. Utilisation principale : élaboration d'un intervalle de confiance

La loi de distribution d'échantillonnage nous permet de dire avec quelle probabilité la moyenne d'un échantillon de taille n se trouve a priori dans un intervalle [a ; b]. On dit a priori car une fois l'échantillon prélevé, la moyenne se trouve ou ne se trouve pas dans l'intervalle, parler de probabilité serait alors un non-sens. ***Nous rappelons qu'ici les paramètres de la population sont connus (ce qui ne sera pas le cas dans les problèmes d'estimation***.

En pratique, après avoir trouvé la loi suivie par Tn ainsi que ses paramètres, il est d'usage de bâtir un intervalle de confiance centré autour de la moyenne, avec un risque d'erreur exprimé en % et noté α, dans lequel doit se trouver la moyenne de l'échantillon pour affirmer que l'échantillon est conforme au résultat attendu (machine bien réglée, production ayant la qualité requise…). La différence 1-α est appelée coefficient de confiance. Il y a toujours un risque de se tromper dans sa conclusion, le coefficient α mesure ce risque.

Intervalle centré autour de la moyenne en pratique : Tn suit une loi normale $N\left(m ; \frac{\sigma}{\sqrt{n}}\right)$

On cherche t tel que $P\left(m - t * \frac{\sigma}{\sqrt{n}} < Tn < m + t * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

α est le risque d'erreur, (1 - α) le coefficient de confiance.

Après transformation, nous obtenons :

$\pi(t) = P(T < t) = (2 - \alpha) / 2 = 1 - \alpha / 2$ avec T qui suit la loi normale centrée réduite N(0 ; 1).

Il suffit alors de trouver t dans la table de la fonction de répartition de la loi normale centrée réduite, de remplacer t par sa valeur dans l'intervalle pour déterminer l'intervalle : $\left[m - t * \frac{\sigma}{\sqrt{n}} ; m + t * \frac{\sigma}{\sqrt{n}}\right]$.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Application 5

Nous reprenons les données de l'application 4

Travail à faire :

1) Bâtir un intervalle centré autour de la moyenne de T_{50} avec un coefficient de confiance de 95 % (risque d'erreur de 5 %). Quelle signification donner à cet intervalle ?
2) Même question avec un risque d'erreur de 1 %. Comparer les deux intervalles. Conclure sur l'amplitude de l'intervalle en fonction du risque retenu.

Remarque : ne pas confondre cet intervalle de confiance, pour lequel les paramètres de la population sont connus avec l'intervalle de confiance permettant d'estimer un paramètre non connu de la population. Dans le cas ci-dessus, il s'agit de voir si un échantillon respecte les caractéristiques de la population.

Application 6

Une grande entreprise fabrique des plats cuisinés. L'emballage doit spécifier notamment la teneur en glucides. Cette entreprise possède un laboratoire permettant de faire un calcul précis de celle-ci.

La teneur est exprimée en gramme pour 100 grammes de produit.

Pour vérifier que la teneur en glucides de la production est respectée et ainsi garantir la qualité nutritive du produit, vous prélevez régulièrement des échantillons et contrôlez notamment la teneur en glucides.

La teneur en glucides du plat cuisiné de référence P187 est en moyenne de 24 g pour 100 g, avec un écart-type de 0,5 g.

L'entreprise produit environ chaque jour 5 000 plats de référence P187. La qualité est jugée conforme aux attentes quant à la teneur en glucides, si le plat contient entre 23,5 et 24,5 g de glucides pour 100 g.

Le contrôle est journalier et porte sur des échantillons de 100 produits.

Les variables aléatoires représentant la teneur en glucides des plats prélevés au hasard sont indépendantes entre-elles.

Travail à faire :

1) Préciser la loi de probabilité suivie par la distribution d'échantillonnage de la moyenne de la teneur en glucides du plat cuisiné de référence P187, pour des échantillons de taille 100 (qu'on pourra noter T_{100}). Vous préciserez les paramètres de cette loi (espérance mathématique et écart-type).
2) Donner un intervalle de confiance centré autour de la moyenne de la variable aléatoire T_{100}, en retenant un coefficient de confiance de 99 % (soit un risque d'erreur de 1 %).
3) Même question avec un risque d'erreur de 5 %.
4) Interpréter les résultats précédents. Sur quelle base l'entreprise peut-elle s'engager auprès de ses clients quant à la teneur en glucide ?

D. Définition de la distribution d'échantillonnage des proportions

Dans le cas de la distribution d'échantillonnage des moyennes, nous nous sommes intéressés à la valeur numérique prise par un paramètre (diamètre, poids, par exemple). Dans ce qui suit, nous allons nous intéresser à une proportion (ou fréquence).

Nous appellerons p la proportion (ou fréquence) des individus possédant le caractère c dans une population-mère. La proportion de ceux qui ne possèdent pas le caractère c dans la population est donc égale à q = (1 – p).

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Exemple : Une entreprise fabrique des instruments pour gauchers (par exemple des ciseaux pour enfants). Pour étudier le marché potentiel, elle recherche la proportion des enfants gauchers dans la population. On appelle c la caractéristique « être gaucher », et p la proportion (fréquence) d'enfants gauchers dans la population. Si l'on considère un enfant pris au hasard, soit il possède la caractéristique c avec une probabilité p, soit il ne la possède pas avec une probabilité (1-p).

Cet exemple peut être formalisé de la façon suivante : nous prenons un enfant au hasard dans la population. X est la variable aléatoire qui prend la valeur 1 avec la probabilité p si l'enfant est gaucher et 0 avec la probabilité q = (1 – p) s'il ne l'est pas. Nous savons que X suit une loi de Bernoulli de paramètre p.

Dans une classe de 30 enfants (soit un échantillon de taille 30), on demande à chaque enfant s'il est gaucher ou non et on note le résultat qui sera 1 ou 0. La proportion (ou fréquence) dans l'échantillon s'obtient en faisant le rapport nombre de gauchers divisé par 30. S'il y a par exemple 6 gauchers, la proportion est 6/30 soit 0,2.

Généralisation :

p est la proportion (ou fréquence) des individus possédant le caractère c dans une population-mère.

q = (1 – p) est la proportion de ceux qui ne possèdent pas le caractère c (c désigne une caractéristique que l'on étudie) dans la population.

On prélève au hasard un individu dans la population mère. X est la variable aléatoire qui prend la valeur 1 si l'individu possède le caractère c et 0 dans le cas contraire.

X suit une loi de Bernoulli et nous avons donc : $E(X) = p$ et $\sigma(X) = \sqrt{p\,q} = \sqrt{p\,.\,(1-p)}$

Si on prélève un échantillon de taille n et qu'on calcule la proportion (ou fréquence) des individus possédant le caractère c dans celui-ci, celle-ci est le résultat de la variable aléatoire suivante (et notée Fn) :

$$Fn = (X1 + X2 + X3 + \ldots + Xn) / n$$

À noter que chaque variable X prend la valeur 1 ou 0.

Définition de la distribution d'échantillonnage des proportions :

Nous nous intéressons à la proportion d'individus qui possèdent le caractère c dans un échantillon de taille n. On appelle distribution d'échantillonnage des proportions la distribution de probabilité de la variable aléatoire Fn construite sur l'échantillon (X1, X2, X3,…, Xn) :

Fn : {ensemble des échantillons de taille n} ➔ R (ensemble des nombres réels)

Ei ➔ E(Ei) (proportion dans l'échantillon Ei)

Paramètres de Fn :

Espérance mathématique de Fn : $E(Fn) = E(X) = p$

Ce qui se traduit par la moyenne des moyennes des échantillons qui est elle-même égale à la moyenne de la population.

L'écart-type de Fn : $\sigma(Fn) = \sigma(X) / \sqrt{n} = \sqrt{\frac{pq}{n}}$

Remarque : Fn peut également être noté Pn (ces deux notations sont couramment utilisées par les auteurs). P peut également être noté f.

La loi suivie par Fn

Si la taille de l'échantillon est faible (en pratique n <30), **Fn suit une loi binomiale** (car chaque variable Xi suit une loi de Bernoulli par hypothèse).

Si la taille de l'échantillon est suffisamment grande (en pratique n>=30), comme la loi binomiale **converge vers la loi normale**, on admettra que Fn suit une loi normale : **$N\left(p, \sqrt{\frac{pq}{n}}\right)$.**

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Application 7

Une entreprise fabrique des ustensiles en inox (plus de 15 000 par jour). Elle est très attentive à la qualité de ses produits et a mis en place à cet effet une politique « qualité ». Les nouvelles machines installées ont permis de réduire sensiblement la proportion moyenne d'ustensiles défectueux. Lorsque les machines sont bien réglées, la proportion moyenne p d'ustensiles défectueux dans la production est de 0,5 % (soit 5 pour 1000).

Afin de garantir ses clients d'un très faible niveau de produits défectueux livrés, l'entreprise s'assure régulièrement que les machines sont bien réglées. Pour cela il est prélevé chaque jour un échantillon de 100 ustensiles, et après contrôle la proportion de défectueux dans l'échantillon est calculée.

Travail à faire :

1) Préciser la loi de probabilité suivie par la distribution d'échantillonnage de la proportion (ou fréquence) d'ustensiles défectueux, pour des échantillons de taille 100 (qu'on pourra noter F_{100}). Vous préciserez les paramètres de cette loi (espérance mathématique et écart-type).
2) Quelle est la probabilité que dans un échantillon de taille 100, la proportion d'ustensiles défectueux soit supérieure à 1 % ?
3) Répondre au deux questions précédentes pour des échantillons de taille 1000.

E. Élaboration d'un intervalle de confiance autour d'une proportion (fréquence)

Position du problème

Nous considérons une population mère dans laquelle p est la proportion (ou fréquence) des individus possédant le caractère c dans une population-mère et, q = (1 – p) la proportion de ceux qui ne possèdent pas le caractère c.

On prélève au hasard un individu dans la population mère. X est la variable aléatoire qui prend la valeur 1 avec la probabilité p si l'individu possède le caractère c (c désigne une caractéristique que l'on étudie), et 0 avec la probabilité q = (1 – p) dans le cas contraire.

Nous prélevons un échantillon de taille n et l'on calcule la proportion (ou fréquence) des individus possédant le caractère c dans celui-ci ; la proportion est le résultat de la variable aléatoire suivante (et notée Fn) : Fn = (X1 + X2 + X3 +…+ Xn) / n.

Nous cherchons à bâtir un intervalle centré autour de la proportion p, avec un risque d'erreur exprimé en % et noté α (ou ce qui est équivalent avec un coefficient de confiance 1-α), dans lequel doit a priori se trouver la proportion possédant le caractère c dans l'échantillon. Nous cherchons ce que nous pouvons appeler une assurance raisonnable. Nous ferons l'hypothèse souvent vérifiée que Fn suit une loi normale (de moyenne p et d'écart-type $\sqrt{\frac{pq}{n}}$).

Quel peut être l'intérêt d'un tel intervalle de confiance ? Nous partirons d'un exemple. Une machine bien réglée produit en moyenne p = 0,2 % de produits défectueux. Pour vérifier régulièrement le réglage de la machine, nous prélevons des échantillons de 50 produits. Nous avons au préalable bâti un intervalle de confiance au risque de 1 % de se tromper (dans la conclusion), dans lequel nous devons trouver le pourcentage de produits défectueux pour affirmer (avec une assurance raisonnable) que la machine est ou n'est pas bien réglée. Nous pouvons ainsi vérifier que la proportion trouvée dans un échantillon se trouve dans l'intervalle. Si la proportion s'y trouve, nous dirons que la machine est bien réglée, tout en prenant un risque moyen de se tromper une fois sur 100 (la machine étant en fait mal réglée). Si la proportion ne s'y trouve pas, nous procéderons alors à un réglage de la machine, et en moyenne une fois sur 100, nous le ferons inutilement (la machine étant en réalité bien réglée).

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Intervalle de confiance centré pour la proportion, avec un risque d'erreur α :

$$[p - t * \sqrt{\frac{p(1-p)}{n}} ; p + t * \sqrt{\frac{p(1-p)}{n}}]$$

p est la proportion (ou fréquence) des individus possédant le caractère c dans une population-mère.

On calcule t à partir de la loi normale (on suppose que toutes les conditions sont remplies). Nous cherchons t dans la table tel que :

$\pi(t)$ **= P(T < t) = (2 - α) / 2= 1 - α / 2 avec T qui suit la loi normale centrée réduite N(0 ; 1).**

Application 8

Reprendre les données de l'application 7 (échantillons de taille 1000).

Travail à faire :

1) Donner un intervalle de confiance centré autour de la proportion (ou fréquence) de la variable aléatoire F_{1000} (proportion ou fréquence de produits défectueux dans des échantillons de taille 1000), en retenant un coefficient de confiance de 95 % (soit un risque d'erreur de 5 %).
2) Interpréter le résultat précédent. Sur quelle base l'entreprise peut-elle s'engager auprès de ses clients quant au nombre de produits défectueux susceptibles d'être livrés malgré les contrôles qualités de l'entreprise ?

III. Les problèmes d'estimation

Estimation
À partir des résultats obtenus sur un échantillon, on cherche à estimer les paramètres de la population mère.

L'estimation est une technique qui permet à partir des valeurs d'un échantillon d'estimer la valeur d'un paramètre de la population mère d'où est issu l'échantillon. C'est une technique qui est notamment utilisée en matière de sondage (il n'est souvent pas possible d'interroger l'intégralité d'une population dans un délai raisonnable).

Exemples d'utilisation en gestion :

Connaître la durée de vie moyenne d'une ampoule produite (ou plus généralement d'un produit fabriqué) ou la durée d'utilisation optimale d'un produit (c'est une durée moyenne qui ne nécessite pas d'être calculée sur la totalité des produits, ce qui ne serait d'ailleurs pas possible puisque cette durée doit être anticipée pour être indiquée sur certains produits de consommation).

A. Approche de l'estimation par l'exemple

Contexte :

La société « Les miels de l'Aunis » exploitent depuis plus de 40 ans les produits de la ruche (miel, pollen, propolis, gelée royale, et cire). Ses produits ont acquis une excellente réputation, ils sont commercialisés en vente directe, mais aussi à travers des magasins spécialisés et en grande surface depuis peu. Un site de vente par internet a également été créé.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Pour se développer la société a lancé depuis quelques mois une nouvelle gamme de produits dans le domaine de l'Apithérapie. (L'Apithérapie est l'utilisation thérapeutique des produits de la ruche pour le bien-être et la beauté. miel, pollen, propolis, gelée royale, cire et même venin).

Le démarrage de cette nouvelle activité est difficile aussi les dirigeants désirent mesurer auprès de sa zone de chalandise (estimée à 100 000 personnes) le taux de notoriété de ses produits d'Apithérapie.

Comme il n'est pas possible d'interroger toute la population visée, il est décidé de retenir un échantillon de 1000 personnes et de leur poser la question suivante : « **connaissez-vous l'existence des produits d'Apithérapie commercialisés par notre société ?** ». Les deux réponses possibles à ce ***sondage*** sont donc oui ou non.

Résultat obtenu au sondage :

- Nombre de réponse « oui » : 150 soit 150 / 1000 = 15 %,
- Nombre de réponse « non » :850 soit 850 / 1000 = 85 %.

Exploitation des résultats :

Sur la base des résultats, les dirigeants estiment que le taux de notoriété de leur produit est de 15 %, mais ils s'interrogent sur la qualité scientifique de ce résultat.

Ils se posent les questions suivantes :

- dans quelles conditions cette estimation est-elle exploitable ?
- l'échantillon retenu était-il correct, était-il suffisant ?
- la taille de l'échantillon a-t-elle un impact sur la qualité de l'estimation et dans quelle mesure ?
- ne serait-il pas plus intéressant de bâtir un intervalle de confiance dans lequel on a de grandes chances de trouver le taux réel de notoriété dans la population totale ?

Nous allons tenter de répondre à toutes ces questions à travers les outils théoriques proposés par les statisticiens.

B. Pour estimer correctement : estimateur sans biais et convergent

Dans l'exemple développé précédemment, nous avons utilisé intuitivement un estimateur, c'est-à-dire une variable aléatoire dont le résultat nous donne une estimation. Pour donner un gage de qualité à l'estimation que l'on obtient, il convient de choisir un estimateur pertinent, et le meilleur d'entre eux que l'on appelle un **estimateur efficace**.

La notion d'estimateur

Nous considérons une population. Nous désirons mesurer une caractéristique de cette population (par exemple le taux de notoriété d'un produit). On prélève donc un **échantillon de taille n** noté (X_1, X_2, X_3, …, Xn). Chaque variable aléatoire X_i suit la même loi de probabilité, qu'elle soit connue ou pas).

Toute fonction de (X_1, X_2, X_3, …, Xn) est appelé estimateur (en clair à chaque échantillon prélevé, nous obtiendrons une valeur qui est fonction de ce que l'on cherche à estimer).

Important : il y a souvent confusion entre la notion d'estimateur et celle d'estimation. L'estimateur est la variable aléatoire, c'est-à-dire la formule qui nous permet de calculer l'estimation.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Quand un estimateur est-il efficace (c'est-à-dire pertinent) ?

1- L'estimateur doit être sans biais

Un estimateur sans biais est un estimateur qui lorsqu'on retient tous les échantillons d'une même taille et que l'on calcule la moyenne des moyennes de tous les échantillons, nous retrouvons exactement le paramètre de la population. (Voir le tout début du chapitre sur les distributions d'échantillonnage pour le voir à travers un exemple concret).

2- L'estimateur doit être convergent

Si l'on prend un échantillon de plus en plus grand, l'estimateur doit permettre de faire une estimation qui se rapproche de la valeur réelle de la population. Et si n est égal à la taille de la population, l'estimation doit correspondre exactement à la valeur du paramètre dans la population.

Dans ce qui suit, nous n'utiliserons que des estimateurs sans biais et convergents, encore appelés absolument corrects. Nous retiendrons pour chaque paramètre le meilleur estimateur absolument correct possible, cet estimateur est appelé estimateur efficace.

C. Estimateur et estimation d'une moyenne

Nous allons ici nous intéresser à la moyenne d'une caractéristique dans une population.

Soit (X_1, X_2, X_3, …, X_n) un échantillon de taille n. Chaque variable aléatoire Xi suit la même loi de probabilité, de moyenne m et d'écart-type σ.

L'estimateur efficace pour estimer au mieux une moyenne est : $Mn = (X_1 + X_2 + X_3 + …+ X_n) / n$

L'estimation ponctuelle de m notée m_n est $(x_1 + x_2 + …+x_n) / n$ (c'est-à-dire la moyenne de l'échantillon prélevé).

$(x_1 + x_2 + …+x_n)$ étant les valeurs du paramètre trouvées dans l'échantillon.

Remarque : Nous ferons l'hypothèse que les conditions sont réunis pour considérer que Mn suit une loi normale $\mathbf{N\left(m ; \frac{\sigma}{\sqrt{n}}\right)}$. Si σ n'est pas connu (σ étant l'écart-type dans la population), on utilise une estimation de celui-ci (voir plus loin).

Exemple

On mesure chaque jour le nombre d'œufs pondus par une population de 1 000 poules dans une journée (elles ne pondent pas tous les jours). Pour connaître le nombre moyen d'œufs pondus par jour, nous allons par exemple reproduire l'expérience pendant 50 jours. Les résultats obtenus sont les suivants :

Jour	Nb œufs	Jour	Nb œufs	Jour	Nb œufs	Jour	Nb œufs	Jour	Nb œufs
1	**728**	11	**737**	21	**748**	31	**747**	41	**748**
2	**753**	12	**765**	22	**769**	32	**760**	42	**759**
3	**789**	13	**778**	23	**767**	33	**758**	43	**766**
4	**735**	14	**752**	24	**758**	34	**750**	44	**775**
5	**791**	15	**780**	25	**775**	35	**785**	45	**784**
6	**827**	16	**812**	26	**87**	36	**795**	46	**796**
7	**832**	17	**825**	27	**815**	37	**820**	47	**802**
8	**784**	18	**799**	28	**787**	38	**792**	48	**794**
9	**801**	19	**805**	29	**808**	39	**806**	49	**788**
10	**766**	20	**785**	30	**787**	40	**785**	50	**779**

On a donc un échantillon de taille 50, le nombre d'œufs pondus chaque jour est le résultat d'une variable aléatoire X_i (X_1 pour le jour 1, X_2 pour le jour 2, etc.). La taille de la population est ici inconnue, non finie (c'est le nombre de jours…).

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Le nombre total d'œufs pondus sur les 50 jours relevés est de 38 334, et la moyenne arrondie à l'entier supérieur est de 767 œufs (38 334 / 50). **767 œufs représentent une estimation du nombre d'œufs pondus par jour par les mille poules.**

D. Estimateur et estimation d'une proportion (ou fréquence)

Nous allons ici nous intéresser à la proportion moyenne d'une caractéristique dans une population.

Soit une population et c un caractère de celle-ci. Soit p la proportion d'individus qui possèdent le caractère c.

On considère un échantillon de taille n, (X_1, X_2, X_3, …, X_n). Chaque variable aléatoire Xi suit la même loi de probabilité, une loi de Bernoulli de paramètre p.

Pourquoi une loi de Bernoulli ?

Car si l'on considère un individu de la population, soit il possède le caractère c, avec une probabilité p, soit il ne le possède pas, avec une probabilité (1-p).

La variable aléatoire Pn = ($X_1 + X_2 + X_3 + …+ X_n$) / n est un estimateur efficace de p.

L'estimation ponctuelle de p, notée p_n ou f_n est ($x_1 + x_2 + …+x_n$) / n (c'est-à-dire la proportion ou fréquence de l'échantillon prélevé).

Loi suivie par Pn (l'estimateur) :

Si les conditions d'approximation de la loi binomiale par la loi normale sont satisfaites alors Pn suit une loi normale N (p, $\sqrt{\frac{pq}{n}}$).

<u>Remarque</u> : Pn est souvent noté Fn pour fréquence (les termes proportion et fréquence sont ici synonymes).

Exemple

Nous reprenons l'exemple précédent sur les poules pondeuses, et nous nous intéressons cette fois-ci à la fréquence de ponte d'une poule par jour.

Soit un jour donné, une poule pond avec une probabilité p et ne pond pas avec une probabilité (1-p). La ponte ou non d'une poule peut donc être formalisée par la variable aléatoire X suivante (qui suit une loi de Bernoulli) :

X = 1 avec la probabilité p (fréquence de ponte)

X = 0 avec la probabilité (1-p)

Si l'on reprend les résultats précédents obtenus au cours des 50 jours, les poules étant au nombre de 1 000, en divisant chaque jour le nombre d'œufs pondus par 1 000, nous obtenons la proportion de poules qui ont pondu le jour considéré. (*Voir tableau page suivante*).

Calculons p_n = (somme des fréquences des 50 jours) / 50 = 38,334 / 50 = 0,76668 soit 76,668 %.

0,76668 ou 76,668 % représente une estimation de la proportion des poules qui pondent par jour. Autrement dit si l'on prend une poule au hasard un jour donné, la probabilité qu'elle ponde ce jour-là est de 0,76668.

<u>Remarque</u> : si l'on prend une autre exploitation, la proportion estimée pourra être différente car les conditions peuvent être différentes (nourriture, cadre, catégorie de poule, etc.) Cela peut permettre d'étudier l'influence de facteurs sur la ponte.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Jour i	p_i	Jour i	p_i	Jour i	p_i	Jour i	p_i	Jour i	p_i
1	**0,728**	11	**0,737**	21	**0,748**	31	**0,747**	41	**0,748**
2	**0,753**	12	**0,765**	22	**0,769**	32	**0,760**	42	**0,759**
3	**0,789**	13	**0,778**	23	**0,767**	33	**0,758**	43	**0,766**
4	**0,735**	14	**0,752**	24	**0,758**	34	**0,750**	44	**0,775**
5	**0,791**	15	**0,780**	25	**0,775**	35	**0,785**	45	**0,784**
6	**0,827**	16	**0,812**	26	**0,087**	36	**0,795**	46	**0,796**
7	**0,832**	17	**0,825**	27	**0,815**	37	**0,820**	47	**0,802**
8	**0,784**	18	**0,799**	28	**0,787**	38	**0,792**	48	**0,794**
9	**0,801**	19	**0,805**	29	**0,808**	39	**0,806**	49	**0,788**
10	**0,766**	20	**0,785**	30	**0,787**	40	**0,785**	50	**0,779**

Interprétation : par exemple le 1er jour, 72,8 % des poules ont pondus.

E. Estimation par intervalle de confiance

La valeur d'une estimation (dite estimation ponctuelle) ne donne qu'un résultat imparfait et dépend notamment de la taille de l'échantillon. Par ailleurs le risque n'est pas négligeable d'avoir une estimation ponctuelle qui ne soit pas le reflet de la réalité, du fait notamment des erreurs d'échantillonnage (voir première partie sur la distribution des échantillons).

Pour limiter les risques et pour ne pas se contenter d'une valeur ponctuelle, il est souvent préférable de bâtir un intervalle de confiance permettant d'encadrer une valeur.

https://www.insee.fr/fr/metadonnees/definition/c1386

Les estimations que fournit une enquête par sondage s'écartent légèrement des résultats qu'aurait donnés une interrogation exhaustive. Si le sondage est aléatoire, la notion d'intervalle de confiance permet de donner une idée de cet écart. Lorsqu'un intervalle de confiance à 95 % est fourni pour une grandeur, cela signifie que cet intervalle a 95 % de chances de contenir la valeur qu'aurait donnée une interrogation exhaustive.

L'intervalle de confiance ne prend en compte que le fait que les résultats proviennent d'une enquête par sondage aléatoire, et non les autres sources d'erreurs : réponses inexactes ou mal interprétées, biais des non-réponses...

Nous donnons simplement dans ce qui suit le résultat, sans formaliser. Il conviendra de développer et de comprendre l'intérêt à travers les applications.

On cherche un intervalle de confiance avec un risque d'erreur α (risque de se tromper dans notre conclusion) dans lequel la moyenne ou la proportion de la population devrait se trouver. Le coefficient de confiance est (1- α).

Intervalle de confiance centré pour la moyenne, avec un risque d'erreur α (exprimé en %) :

$$\left[m_n - t * \frac{\sigma}{\sqrt{n}} \; ; \; m_n + t * \frac{\sigma}{\sqrt{n}}\right]$$

Avec m_n, la moyenne estimée (moyenne de l'échantillon) et σ, l'écart-type de la population. **Si σ n'est pas connu, on utilise son estimation notée généralement s_n.**

une estimation (dite ponctuelle) s_n de σ est donnée par la racine carrée de [n/(n-1)] * $(\sigma_n)^2$

σ_n étant l'écart-type de l'échantillon : $sn = \sigma n\sqrt{n/(n-1)}$

(à noter que $\frac{\sigma}{\sqrt{n}}$ **= sn /** $\sqrt{n-1}$**)**

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

On calcule t à partir de la loi normale (on suppose que toutes les conditions sont remplies). Nous cherchons t dans la table tel que :

$\pi(t)$ = P(T < t) = (2 - α) / 2= 1 - α / 2 avec T qui suit la loi normale centrée réduite N(0 ; 1).

Intervalle de confiance centré pour la proportion, avec un risque d'erreur α (exprimé en %) :

$$\left[f - t * \sqrt{\frac{f(1-f)}{n}} ; f + t * \sqrt{\frac{f(1-f)}{n}}\right]$$

Avec f, la proportion ou fréquence estimée (proportion de l'échantillon).

On calcule t à partir de la loi normale (on suppose que toutes les conditions sont remplies). Nous cherchons t dans la table tel que :

$\pi(t)$ = P(T < t) = (2 - α) / 2= 1 - α / 2 avec T qui suit la loi normale centrée réduite N(0 ; 1).

Exemple

Nous reprenons l'exemple précédent sur les poules pondeuses.

Intervalle de confiance à 95 % autour de la moyenne (risque d'erreur 5 %)

- Nous cherchons un intervalle centré autour de la moyenne d'œufs pondus par jour par les 1 000 poules. (une estimation de la moyenne est de 767 œufs par jour)
    - Nous cherchons dans un premier temps t tel que $\pi(t)$ = P(T < t) = 1 – (5% / 2) = 1 – (0.05 / 2) = 0,975.
    - Dans la table de la loi normale centrée réduite, nous trouvons t = 1,96.
    - Il nous faut ensuite calculer une estimation de l'écart-type de la population. Pour cela nous calculons l'écart-type dans l'échantillon, nous obtenons : σ_n = 100,17.
    - Nous multiplions ensuite cet écart-type par la racine carrée de n/n-1 = $\sqrt{(\frac{50}{49})}$. Une estimation de l'écart-type est donc : s_n = 100,17 * $\sqrt{(\frac{50}{49})}$ = 101,19.

      <u>À noter que</u> Racine carrée de 50 / 49 est très proche de 1, donc quand n est grand (n est la taille de l'échantillon, ici 50 pour 50 jours), nous pouvons raisonnablement retenir l'écart-type de l'échantillon sans multiplier par le facteur racine carrée de n/n-1 (retenir 100,17 ou 101,19 ne change pas fondamentalement l'intervalle obtenu).
- Donc l'intervalle cherché est [767-1,96*101,19/$\sqrt{50}$; 767-1,96*101,19/$\sqrt{50}$] = [738,95 ; 795,05] arrondi à **[738 ; 795]**. Cela signifie que le nombre estimé d'œufs pondus par jour est compris entre 738 et 795, le risque d'erreur de se tromper étant de 5 %. À noter que pour réduire l'intervalle, il faut augmenter le risque de se tromper.
  <u>Attention</u> ; ***ne pas oublier de diviser l'écart-type estimé par la racine carrée de n***.

L'intervalle de confiance permet d'apporter une information supplémentaire par rapport à une estimation ponctuelle. En matière de gestion il est préférable de savoir que certains jours le nombre d'œufs pourra ne pas dépasser 738 pour éviter l'insatisfaction de clients, ou encore pour dimensionner son exploitation en fonction de la demande (nombre de poules nécessaires pour produire au minimum tant d'œufs).

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Intervalle de confiance à 95 % autour de la fréquence (risque d'erreur 5 %)

- Nous cherchons un intervalle centré autour de la fréquence moyenne de ponte d'une poule. La fréquence estimée est de f = 0,76668 arrondie à 0,767
    - Nous cherchons dans un premier temps t tel que $\pi(t)$ = P(T < t) = 1 – (5% / 2) = 1 – (0.05 / 2) = 0,975.
    - Dans la table de la loi normale centrée réduite, nous trouvons t = 1,96.
    - Nous remplaçons ensuite f et t par leur valeur dans l'intervalle

$$\left[f - t * \sqrt{\frac{f(1-f)}{n}} ; f + t * \sqrt{\frac{f(1-f)}{n}}\right]$$

    - Nous obtenons :

$[0{,}767 - 1{,}96 * \sqrt{0.767 * (1 - 0.767)/50}$; $0{,}767 + 1{,}96 * \sqrt{0.767 * (1 - 0.767)/50}]$

Soit **[0,760 ; 0.774].** Cela signifie que la proportion moyenne estimée de poules qui pondent chaque jour est comprise entre 0,760 (76 %) et 0,774 (77,4 %). Le risque de se tromper est de 5 %.

Application 9 – Estimation d'une moyenne – Estimation par intervalle de confiance pour la moyenne

Une entreprise (dont le marché est mondial) fabrique un produit qui permet de nettoyer les injecteurs des véhicules diesels. Ce produit permet en particulier de réduire la consommation de gazole.

Elle vient de mettre au point un nouveau produit et affirme que celui-ci permet de diminuer significativement la consommation de gazole et bien au-delà des produits actuels du marché.

L'entreprise fait appel à un bureau d'étude chargé de tester le nouveau produit afin de démontrer l'efficacité du nouveau produit.

Ce bureau d'étude effectue des essais de véhicules sur circuit et obtient les résultats suivants :

Nombre de véhicules testés : 50 (véhicule identiques)

Consommation moyenne et écart-type par véhicule :

Condition de test **Échantillon de 50 voitures**	**Moyenne m_n de l'échantillon** **L / 100 km**	**Écart-type de l'échantillon** **L / 100 km**
Sans utilisation de produit nettoyant pour injecteurs	6,75	0,12
Avec utilisation d'un produit concurrent	6,40	0,10
Avec utilisation du nouveau nettoyant mis au point	6,08	0,08

Travail à faire :

Peut-on en conclure que le nouveau produit est plus efficace ?

Aide – Approche proposée

- Donner une estimation de la moyenne pour chaque cas ;
- Donner une estimation par intervalle de confiance autour de la moyenne en retenant un risque de 5 % puis de 1 % (pour le cas de l'utilisation du nouveau nettoyant mis au point) ;
- Commenter.

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Application 10 – Estimation par intervalle de confiance pour la fréquence (proportion) – D'après un sujet de DECF

La Société VTT-EVASION est une petite entreprise spécialisée dans le montage et la commercialisation de VTT (vélos tout terrain). Ce secteur d'activité est en plein développement. Plusieurs études sont menées afin d'améliorer les performances de l'entreprise.

Dans une production de VTT sortis des ateliers, on choisit un échantillon aléatoire, assimilé à un échantillon non exhaustif, de 350 VTT dont le montage est soigneusement contrôlé. Sur les 350 VTT contrôlés, on en a compté 21 présentant au moins un défaut de montage.

Travail à faire

Déterminer l'intervalle de confiance à 92 % de la proportion de VTT présentant au moins un défaut de montage.

Remarque : il s'agit bien d'un problème d'estimation car nous partons d'un échantillon (les 350 VTT) pour estimer la proportion de la population (ensemble de la production).

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr

Table de la loi normale centrée réduite

Fonction de répartition de la loi normale centrée réduite : $P(T \leq t) = \pi(t)$

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

TABLE POUR LES GRANDES VALEURS DE t

t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
$\pi(t)$	0,99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,99984	0,99992	0,99997	0,99999

(cc) BY-NC-SA Réseau CRCF - Ministère de l'Éducation nationale - http://crcf.ac-grenoble.fr