

Textes et Règlementation en vigueur en matière de statistique

L'ONG GBENONKPO de la place a gagné un financement pour exécuter un projet d'appui aux producteurs au Nord-Bénin. Le bailleur a exigé à l'ONG la réalisation d'une étude diagnostique avant le démarrage des activités, car selon lui, la réalisation de l'étude diagnostique est le moyen le plus efficace pour évaluer l'impact des bénéficiaires. L'ONG est appelée à recruter un expert en statistique et suivi évaluation pour la conduite des travaux de cette étude. A cet effet, elle vous a fait appel pour la réalisation de cette étude qui concerne les données statistiques publique en république du Bénin.

A partir des connaissances que vous avez dans le domaine de statistique, associez aux connaissances reçues au cours de règlementation et texte en matière de statistique et planification, répondez aux questions.

Questions

1. Qu'appelle-t-on règlementation, loi, code, charte ?
2. Expliquer les différentes phases de la qualité et crédibilité des données.
3. Quels est le nom du document a obtenu avant le démarrage des collectes des données et l'autorité qui délivre ?
4. Quelles sont les principales constituantes de la loi statistique n°2022-07 du 27 juin 2022 ?
5. Quels sont les enjeux de règlementation et texte en matière de statistique ?
6. Quels sont les avantages de la loi de cadre et planification.

Correction

1. Qu'appelle-t-on règlementation, loi, code, charte ?

- **Réglementation** : Ensemble de règles imposées par une autorité pour organiser et encadrer un domaine spécifique.

- **Loi** : Texte voté par le pouvoir législatif définissant des obligations et des droits applicables à tous.
- **Code** : Recueil systématique de lois ou règlements relatifs à un domaine particulier (ex. Code du travail).
- **Charte** : Document définissant des principes et engagements souvent de nature éthique ou organisationnelle.

2. Expliquer les différentes phases de la qualité et crédibilité des données.

- **Phase de collecte** : Garantir la précision, la cohérence et la complétude des données dès leur acquisition.
- **Phase de traitement** : Vérification, nettoyage et structuration pour assurer la fiabilité.
- **Phase d'analyse** : Validation et interprétation pour garantir la pertinence et l'exactitude.
- **Phase de diffusion** : Transparence et traçabilité pour assurer l'accessibilité et l'utilité des données.

3. Quel est le nom du document à obtenir avant le démarrage des collectes des données et l'autorité qui le délivre ?

Le document requis est le **VISA Statistique**, généralement délivrée par l'**Institut National de la Statistique**.

4. Quelles sont les principales constituantes de la loi statistique n°2022-07 du 27 juin 2022 ?

Les principales constituantes incluent :

- Le cadre juridique de la production et diffusion des statistiques.
- Les principes de confidentialité, d'impartialité et d'accès aux données.

- Les obligations des producteurs et utilisateurs de données.
- Les sanctions en cas de non-respect des dispositions légales.
- La gouvernance du système statistique national.

5. Quels sont les enjeux de réglementation et texte en matière de statistique ?

- Assurer la fiabilité et la qualité des données produites.
- Garantir la confidentialité et la protection des informations collectées.
- Favoriser une meilleure prise de décision basée sur des données crédibles.
- Harmoniser les pratiques statistiques au niveau national et international.
- Renforcer la transparence et la confiance du public.

6. Quels sont les avantages de la loi de cadre et planification ?

- Fixer une vision et des orientations claires pour le développement statistique.
- Assurer la cohérence et l'efficacité des actions à long terme.
- Faciliter la mobilisation des ressources et des partenaires.
- Encadrer les responsabilités des acteurs impliqués.
- Promouvoir une meilleure coordination et synergie entre les différentes institutions.

Technique et méthode d'Echantillonnage et rééchantillonnage

Partie A

Répondez par « vrai » ou par << faux >> aux affirmations qui suivent, et donnez la meilleure réponse lorsque vous répondez par « faux ».

1. Dans un échantillon représentatif, les caractéristiques de l'échantillon correspondent à celles de la population. **Vrai**
2. Lors de l'analyse des données, il n'est pas nécessaire de comprendre la conception de l'échantillon, mais plutôt la manière dont la conception a été réellement exécutée sur le terrain. **Faux.** Il est essentiel de comprendre la conception de l'échantillon avant d'analyser les données, car cela influence la validité des résultats.
3. L'échantillonnage non probabiliste ne peut jamais produire d'échantillon représentatif. **Faux.** Un échantillonnage non probabiliste peut produire un échantillon représentatif, mais cela repose sur des hypothèses et n'est pas garanti scientifiquement.
4. Si la base de sondage et la population sont très similaires, on a peu d'erreur d'échantillonnage. **Vrai**
5. Dans un échantillonnage probabiliste, il n'est pas nécessaire que les chances que les unités d'observation soient sélectionnées dans l'échantillon soient égales. **Vrai**
6. Un échantillon sélectionné avec des chances de sélection inconnues peut être lié de manière appropriée à la population à partir de laquelle l'échantillon a été tiré. **Faux.** Sans connaître les chances de sélection, il est difficile de faire des inférences valides sur la population.
7. Dans les méthodes non aléatoires, la seule façon de mesurer la qualité des données en résultant consiste à comparer certains des résultats de l'enquête à l'information dont on dispose au sujet de la population. **Vrai**
8. La précision des estimations de l'échantillon dépend beaucoup de la taille de l'échantillon mais aussi de la taille de la population. **Faux.** La précision dépend

principalement de la taille de l'échantillon et non de celle de la population, sauf pour les populations très petites.

9. Généralement, toutes les méthodes d'échantillonnage probabilistes permettent de tirer des échantillons représentatifs de la population cible. **Vrai**

10. L'échantillonnage aléatoire simple et l'échantillonnage stratifié sont considérés comme supérieurs (à toutes les méthodes d'échantillonnage) en termes de représentativité. **Faux.** D'autres méthodes, comme l'échantillonnage en grappes ou systématique, peuvent être tout aussi efficaces selon le contexte.

Partie B

- a. Qu'entend-on par échantillon et quel est son intérêt ?
- b. Qu'est-ce qu'une base de sondage ?
- c. Citez les différentes méthodes d'échantillonnage probabilistes puis faites une brève description de l'échantillonnage systématique.
- d. Citez deux types de biais d'échantillonnage, et décrivez-les brièvement.
- e. Quel est l'inconvénient principal de l'échantillonnage stratifié.
- f. Donnez deux avantages de l'échantillonnage non probabiliste.

Partie C

1. Qu'est-ce que le rééchantillonnage ?
2. Quelle idée sous-tend les méthodes de rééchantillonnage ?
3. Citez trois types de technique de rééchantillonnage, puis décrivez brièvement les principes de ces méthodes.
4. Citez deux avantages de la méthode du bootstrap ?

Correction

Partie B

1. **Un échantillon** est un sous-ensemble représentatif d'une population, utilisé pour tirer des conclusions sur cette dernière. Il permet de réduire le coût et le temps d'une enquête tout en obtenant des résultats fiables.
2. **Une base de sondage** est la liste exhaustive des éléments de la population cible à partir de laquelle un échantillon peut être sélectionné.
3. **Méthodes d'échantillonnage probabilistes :**
 - Échantillonnage aléatoire simple
 - Échantillonnage stratifié
 - Échantillonnage en grappes
 - Échantillonnage systématique

Échantillonnage systématique : Il consiste à sélectionner les unités à des intervalles fixes à partir d'une liste ordonnée après avoir choisi un point de départ aléatoire.

4. **Deux types de biais d'échantillonnage :**
 - **Biais de sélection :** Se produit lorsque certaines unités de la population ont une probabilité plus faible d'être incluses dans l'échantillon.
 - **Biais de non-réponse :** Se produit lorsque les participants choisis ne répondent pas, créant une distorsion dans les résultats.
5. **Inconvénient principal de l'échantillonnage stratifié :** Il peut être coûteux et complexe à mettre en œuvre, nécessitant une connaissance préalable de la population pour bien définir les strates.
6. **Deux avantages de l'échantillonnage non probabiliste :**

- Moins coûteux et plus rapide à mettre en œuvre.
 - Utile dans les études exploratoires ou lorsqu'une base de sondage n'est pas disponible.
-

Partie C

1. **Le rééchantillonnage** est une méthode statistique qui consiste à générer de nouveaux échantillons à partir d'un échantillon existant, souvent en utilisant des techniques comme le bootstrap ou le jackknife.
2. **Idée sous-jacente** : Le rééchantillonnage repose sur la simulation pour estimer la variabilité des statistiques sans avoir à faire des hypothèses strictes sur la distribution des données.
3. **Trois techniques de rééchantillonnage** :
 - **Bootstrap** : Tirages répétés avec remise pour estimer la distribution d'une statistique.
 - **Jackknife** : Exclusion séquentielle d'une observation à la fois pour évaluer la variabilité des estimations.
 - **Permutation** : Réarrangement aléatoire des valeurs observées pour tester des hypothèses statistiques.
4. **Deux avantages de la méthode du bootstrap** :
 - Ne nécessite pas d'hypothèses strictes sur la distribution des données.
 - Fournit des intervalles de confiance plus précis, même pour des échantillons de petite taille.

Master ENSPD/Epreuve d'Anglais 2023-2024

I. Reading comprehension

Text: The New Gods

1. By the time you sit down at your desk this morning, hundreds of brands will have hit your retina-toothpaste, underwear, cereal, newspaper, billboards, even the back of your ticket, the railway sign and the paper bag for your croissant. We live in a brandscape where an increasing proportion of our lives and corporate logos are now in every civic space, from the school playground and hospital to art gallery, the sponsored political debate and the marathon. And this is global work down a street in any city in the world and there will be enough brands to make you feel at home.

2. For the first time, marketing-its purpose and its method-has shifted into the centerground of intense political debate and activism. The anti-globalization movement doesn't demonstrate outside multinational corporate headquarters, it attacks brands on the streets, the more familiar the better-MackDonald's Starbuck-because it makes for great media coverage and symbolizes the movement's criticism of global capitalism.

3. Brands have rightly become a political battleground, they represent power. Firstly, they are worth a lot of money. The physical assets now represent just the tip of the iceberg of the company's worth. Far more important is the perception millions of consumers have of the company. Reputation is worth millions because, in a highly competitive market, it brings you loyal customers and good staff.

4. Seondly, brands generate more trust than any institution-government, church and politics all fall before the credibility of some brands. People have great affection and loyalty to them and will pay over odds for the logo. The argument runs that a successful brand offers consistency of quality, a point of certainty in an uncertain world. So, we use brands and we decode other people's use of brands to establish their status. We no longer identify with churches, political parties or even our local community: we construct our sense of who we are through our association with brands.

5. Brands are no longer simply about the qualities of the product or service they sell, they are promoted as a set of values, a philosophy, even an ideology. For example, the Orange brand represents a bright optimistic future, while Nike is about individual achievement. Brands 2 are being used to humanize corporations by appropriating characteristics such as courage, honesty, friendliness and fun. As they take on good causes as Aids or death penalty, brands are implicitly bidding to legitimize the corporation as a morally and socially responsible institution. The aim is that the audience will forget the downsizing, the sweatshop wages, the environmental waste, the huge pay inequalities and believe that the corporation is a good thing.

(SOURCE: MADELEINE BUNTING, THE GUARDIAN, JULY 9, 2001, SLIGHTLY ADAPTED)

QUESTIONS

A. Answer by T (true) or F (False) according to the text

1. The first paragraph states that a decreasing proportion of our working lives is mediated by brands.
2. It is also clearly implied that brands cannot be seen or found in cities of the third world.
3. In the past or before July 2001, the purpose and method of marketing had not shifted to the centerground of activism.
4. The physical assets of a company are said to be less important or valuable than its intangible or reputational assets.
5. The credibility of brands is getting higher and higher than that of politics or religion.
6. Nowadays, people hardly construct their sense of what they are identifying with politics or religion.
7. By appropriating human properties/attributes like courage and honesty, brands don't treat or present corporations as 'human beings'.

8. Brands sometimes help people to remember the wrongdoings or negative actions of corporations.

B. Answer these questions with information from the text

1. What are referred to as 'the New Gods' and why are they so referred to?

2. Why do demonstrators prefer to attack the corporate brands rather than the corporate headquarters?

II. Match each word/phrase in colum A with its definition in columb B. Write down numbers and letters only

Columb A

1. Statistics

2. Descriptive statistics

3. Sample

4. Population

5. Monitoring

6. Evaluation

7. Input

8. Outputs.

9. Local development

10. Unit

Columb B

a. is the collection of all individuals or items under consideration in a statistical study

- b. consists of a body of methods for collecting and analyzing data
- c. is that part of the population from which information is collected
- d. consist of methods for organizing and summarizing information
- e. is a process that systematically and objectively assesses all the elements of a programme (e.g. design, implementation and results achieved) to determine its overall worth or significance
- f. is a continuous process of collecting and analysing information about a programme, and comparing actual against planned results in order to judge how well the intervention is being implemented
- g. The resources, both financial and human resources required to undertake your project..
- h. it is the set of structural transformations (demographic, economic, social, mental, political, etc.) that make economic growth and the rise in living standards possible and accompany in a given place, region or country.
- i. The immediate results of a grantee's activities-the processes, products, goods and services delivered through funded activities e.g. publications, manuals, dataset, models, workshops, stakeholder meetings etc. produced.
- j. is a standardized and agreed-upon value used to express a specific quantity.

Correction

I. Reading Comprehension

A. Answer by T (True) or F (False) according to the text

1. **F** — The first paragraph mentions that **an increasing proportion** of our lives is mediated by brands.
2. **F** — The text does not imply that brands cannot be found in third world cities. It talks about brands being globally present.

3. **T** — The text says that the purpose and method of marketing shifted into the center of political debate **for the first time**.
4. **T** — The text states that the physical assets of a company are less important than its intangible or reputational assets.
5. **T** — The text mentions that brands generate more trust than any institution, including government, church, and politics.
6. **T** — The text suggests that people now construct their sense of identity through their association with brands rather than identifying with political parties or religion.
7. **F** — The text says that brands present corporations as “human beings” by appropriating human qualities such as courage and honesty.
8. **F** — The text says that brands attempt to make the audience forget the negative aspects (like downsizing or environmental waste) and believe in the positive aspects of the corporation.

B. Answer these questions with information from the text

1. **'The New Gods'** refer to **brands**. They are referred to as such because they have taken on an almost sacred role in our lives, influencing our identity, loyalty, and perceptions more than traditional institutions like politics or religion.
 2. Demonstrators prefer to attack **corporate brands** rather than the corporate headquarters because **brands are more familiar** and make for great media coverage, symbolizing the movement's critique of global capitalism.
-

II. Match each word/phrase in column A with its definition in column B.

1. **Statistics** — **b.** consists of a body of methods for collecting and analyzing data

2. **Descriptive statistics** — **d.** consist of methods for organizing and summarizing information
3. **Sample** — **c.** is that part of the population from which information is collected
4. **Population** — **a.** is the collection of all individuals or items under consideration in a statistical study
5. **Monitoring** — **f.** is a continuous process of collecting and analyzing information about a programme, and comparing actual against planned results in order to judge how well the intervention is being implemented
6. **Evaluation** — **e.** is a process that systematically and objectively assesses all the elements of a programme (e.g. design, implementation and results achieved) to determine its overall worth or significance
7. **Input** — **g.** The resources, both financial and human resources required to undertake your project
8. **Outputs** — **i.** The immediate results of a grantee's activities-the processes, products, goods and services delivered through funded activities e.g. publications, manuals, dataset, models, workshops, stakeholder meetings, etc. produced
9. **Local development** — **h.** it is the set of structural transformations (demographic, economic, social, mental, political, etc.) that make economic growth and the rise in living standards possible and accompany in a given place, region or country
10. **Unit** — **j.** is a standardized and agreed-upon value used to express a specific quantity

Épreuve : Statistique inférentielle

NB: Dans toute l'épreuve et partout où c'est utile, z_q désigne le quantile d'ordre q de la loi normale centrée et réduite et $t_{k,q}$ est le quantile d'ordre q d'une loi de student à k degré de liberté. On donne $z_{0.05} = -1.64$ et $z_{0.025} = -1.96$

EXERCICE 1

L'achat de céréales dans les marchés du Bénin comporte une expérience de mesures de quantité grâce à un bol dont la contenance pèse normalement 1 kg. Mais une mesure de ce bol pèse rarement 1 kg à cause des différents aléas qui entourent cette mesure. Soit $XN(m, \sigma)$ (loi normale d'espérance m et d'écart-type σ) la variable aléatoire désignant le poids d'un bol de céréale mesuré par une vendeuse prise au hasard. Il vous est demandé de choisir la réponse correcte parmi les propositions.

1/ Un client achète $n = 4$ mesures de céréales; on note \bar{x} , la moyenne empirique et s l'écart-type empirique de ces 4 mesures. Parmi les phrases ci-dessous, la plus correcte est:

- (a), donnera 1 kg et s donnera 0.2
- (b), ne donnera pas 1 kg et s ne donnera pas 0.2
- (c) \bar{x} peut donner 1 kg et s peut donner 0.2
- (d) \bar{x} peut donner 1 kg, mais il est peu probable que s donne 0.2

2/ 25 clients viennent acheter chacun $n = 4$ kg de céréale; on note \bar{x} la moyenne empirique des moyennes de chaque client et s l'écart-type de cette distribution de moyennes. Parmi les phrases ci-dessous, la plus correcte est:

- (a) donnera 1 kg et s donnera 0.2
- (b) ne donnera pas 1 kg et s ne donnera pas 0.2
- (c) peut donner 1 kg mais s ne peut pas donner 0.2

(d) peut donner 1 kg, et il est fort probable que s donne 0.2

3/ Un client achète $n = 4$ mesures de céréales et souhaite calculer la probabilité (arrondie au centième près) que le poids moyen de ces mesures soit supérieur à 1.2 kg.

(a) il peut utiliser l'écart-type empirique s_1 (non corrigé) dans ce contexte.

(b) il doit utiliser l'écart-type corrigé s_2 dans ce contexte.

(c) qu'il utilise s , ou s_2 , il aura les mêmes résultats

(d) Aucune des réponses précédentes n'est correcte

4/ Le nombre minimal n de mesures que doit acheter un client pour que la moyenne empirique se situe dans un rayon de 0.15 kg autour de m dans au moins 95% des cas est:

(a) $n = 35$

(b) $n = 54$

(c) $n = 23$

(d) tel qu'aucune des réponses précédentes n'est correcte

EXERCICE 2 ($2.5 * 4 = 10$ points)

Pour vérifier la qualité de la formation donnée aux étudiants à l'ENSPD, un ingénieur en statistique s'intéresse aux nombres de fois qu'un étudiant issu de cette école candidate à une offre d'emploi jusqu'à son premier recrutement. Il décrit ce nombre par une variable aléatoire X suivant un loi géométrique de paramètre p tel que

$$\mathbb{P}\{X = l\} = p * (1 - p)^{l-1}, l \in \mathbb{N} \text{ et } p \in]0, 1[.$$

Pour chacune des questions suivantes, choisir la ou les bonnes réponse(s). Lorsque vous faites plus de choix qu'il en faut, chaque choix incorrecte enlève les des points alloués à la bonne réponse de la question.

1/ Dans ce modèle, l'information de Fisher pour une observation est

(a) $I(p) = 1/(p^2)$

(b) $J(p) = 1/(p^2 * (1 - p))$

(c) $I(p) = 1/(p(1 - p))$

(d) $I(p) = 1/(p * (1 - p)^2)$

2/ Soit X_1, X_2, \dots, X_n un n-échantillon iid de même loi que X . L'estimateur du maximum de vraisemblance \hat{p}_n de p est donné par

(a) $\hat{p}_n = \sum_{i=1}^n X_i / n$

(b) $\hat{p}_n = 2 \sum_{i=1}^n X_i / n$

(c) $\hat{p}_n = n \sum_{i=1}^n X_i$

(d) $\hat{p}_n = 1 / (2n \sum_{i=1}^n X_i^2)$

3/ $T_n = \sqrt{n} (\tilde{p}_n - p) / \sqrt{\tilde{p}_n (1 - \tilde{p}_n)}$ est une statistique

(a) asymptotiquement pivotale pour p (b) pivotale pour p .

(c) suivant une loi normale (d) suivant une loi de Student

4/ A partir de la bonne réponse à la question 3 /, on obtient un intervalle de confiance de niveau (ou niveau asymptotique) $1 - \alpha$ pour p donné par

(a) $IC_{1 - \alpha}(p) = [\hat{p}_n - z_{1 - \alpha/2} \sqrt{\hat{p}_n (1 - \hat{p}_n)}; \hat{p}_n + z_{1 - \alpha/2} \sqrt{\hat{p}_n (1 - \hat{p}_n)}]$

(b) $IC_{1 - \alpha}(p) = [\hat{p}_n (1 - z_{1 - \alpha/2}^2 / n) / \sqrt{1 - \hat{p}_n}; \hat{p}_n (1 + z_{1 - \alpha/2}^2 / n) / \sqrt{1 + \hat{p}_n}]$

(c) $IC_{1 - \alpha}(p) = [\hat{p}_n - t_{n-1, 1 - \alpha/2} \sqrt{\hat{p}_n (1 - \hat{p}_n)}; \hat{p}_n + t_{n-1, 1 - \alpha/2} \sqrt{\hat{p}_n (1 - \hat{p}_n)}]$

(d) $IC_{1-\alpha}(p) = [\hat{p}_n (1 - t_{n,1-\alpha/2} \sqrt{1-\hat{p}_n}), \hat{p}_n (1 + t_{n,1-\alpha/2} \sqrt{1-\hat{p}_n})]$

Correction

Voici les réponses à vos questions :

EXERCICE 1

La plus correcte est :

(d) peut donner 1 kg, mais il est peu probable que s_1 donne 0.2

Explication : La moyenne empirique d'un échantillon de taille $n = 4$ peut donner 1 kg en moyenne, mais il est peu probable que l'écart-type empirique des mesures soit exactement 0.2, car il est sujet à la variation d'échantillon.

La plus correcte est :

(b) il doit utiliser l'écart-type corrigé s_2 dans ce contexte

Explication : L'écart-type corrigé (s_2) est utilisé dans le calcul de la probabilité sur un échantillon pour compenser l'estimation de la variance à partir de l'échantillon.

La plus correcte est :

(b) $n = 54$

Explication : La formule pour le nombre minimal n qui donne un rayon de 0.15 autour de la moyenne avec 95% de confiance est basée sur la formule de l'intervalle de confiance pour la moyenne, et $n = 54$ est la valeur correcte.

EXERCICE 2

La réponse correcte est :

(c) $I(p) = 1/(p(1-p))$

Explication : L'information de Fisher pour la loi géométrique est donnée par cette expression.

La réponse correcte est :

(a) $\hat{p}_n = (\sum_{i=1}^n X_i) / n$

Explication : L'estimateur du maximum de vraisemblance de p pour un échantillon iid de loi géométrique est la moyenne des X_i .

La réponse correcte est :

(a) asymptotiquement pivotale pour p

Explication : La statistique T_n est asymptotiquement pivotale pour p , car elle converge en distribution vers une loi normale.

La réponse correcte est :

$$IC_{1-\alpha}(p) = [(\hat{p}_n - z_{1-\alpha/2} \sqrt{(1-\hat{p}_n)/n}); (\hat{p}_n + z_{1-\alpha/2} \sqrt{(1+\hat{p}_n)/n})]$$

Explication : L'intervalle de confiance pour p est donné par cette formule, qui utilise le quantile de la loi normale.