

## EXERCICES

### Statistique descriptive univariée : révisions

#### Exercice 1

Pour 12 patients hospitalisés pour un état dépressif majeur, on a évalué d'une part la dimension clinique de leur dépression à l'aide d'un score à un inventaire psychiatrique standardisé, noté score PSE (mesuré en points entiers), et d'autre part l'efficacité d'un traitement antidépresseur administré à ces patients, après 28 jours de traitement. On a obtenu les résultats suivants :

sujet n°	1	2	3	4	5	6	7	8	9	10	11	12
efficacité	oui	oui	oui	oui	non	oui	non	oui	oui	oui	oui	non
score PSE	7	6	6	6	5	5	5	5	7	6	6	4

Source : d'après Marchot et al, *Psychologie Médicale* 1989, 21, 905-909

- 1) On considère la variable efficacité du traitement :
  - a. Définir précisément la variable et préciser son type.
  - b. Donner les distributions d'effectifs et de proportions de la variable efficacité du traitement.
  - c. Quel est le mode de la variable efficacité du traitement ?
  - d. Représenter graphiquement la distribution de la variable efficacité du traitement.
- 2) On considère le score PSE :
  - a. Donner l'échelle descriptive associée à cette variable et préciser son type.
  - b. Donner la distribution d'effectifs du score PSE.
  - c. Quel est le mode du score PSE ?
  - d. Représenter graphiquement la distribution du score PSE.
  - e. Calculer la proportion de patients ayant un score PSE au plus égal à 5, au plus égal à 6.
  - f. Calculer la proportion de patients ayant un score PSE au moins égal à 6.
  - g. Calculer la moyenne, la variance et l'écart-type du score PSE.

#### Exercice 2

Pour étudier le niveau de mémoire des sujets d'une population on leur a demandé de mémoriser des mots de manière à se les rappeler par la suite. Pour chaque sujet, on a noté le nombre de mots correctement mémorisés et obtenu les résultats suivants :

nombre de mots mémorisés	10	11	12	13	14
nombre de sujets	12	19	0	13	6

- 1) Calculer la taille de la population étudiée.
- 2) Donner l'échelle descriptive associée à la variable nombre de mots mémorisés, et préciser son type (justifiez votre réponse).
- 3) Donner la distribution de proportions du nombre de mots mémorisés.
- 4) Quel est le mode du nombre de mots mémorisés ?
- 5) Représenter graphiquement la distribution du nombre de mots mémorisés.
- 6) Calculer la proportion de sujets qui ont mémorisé au plus 11 mots, au moins 12 mots.
- 7) Calculer le nombre moyen de mots mémorisés, la variance et l'écart-type du nombre de mots mémorisés.

#### Exercice 3

Une étude portant sur 1 264 adolescents habitant l'état de Caroline du Sud a été menée pour étudier les effets post-traumatiques de l'ouragan Hugo un an après ce cataclysme. Parmi d'autres variables on a relevé pour chaque adolescent son sexe, son âge et s'il revivait ou non l'événement (rêves, souvenirs ...).

Source : C.Z. Garrison, M.W. Weinrich, S.B. Hardin et al, *American Journal of Epidemiology* 1993, 138, 522-530

- 1) Il y a 664 filles :
  - a. donner les distributions d'effectifs et de proportions du sexe, après avoir défini précisément la variable étudiée.

- b. représenter graphiquement la distribution du sexe et préciser son mode.
- 2) Cent soixante et une filles et 82 garçons revivent l'événement :
- donner les distributions d'effectifs et de proportions de la variable "revivre l'événement"; quel est son mode ?
  - donner les distributions d'effectifs et de proportions de la variable "revivre l'événement" dans la sous-population des filles.
  - donner les distributions d'effectifs et de proportions de la variable "revivre l'événement" parmi les garçons. Comparer à celle obtenue chez les filles.
- 3) La distribution de l'âge des adolescents est la suivante :

âge (en années)	]13 ; 14]	]14 ; 15]	]15 ; 16]	]16 ; 17]	]17 ; 18]
effectif	49	479	529	169	38

- définir la variable âge.
- donner la distribution de proportions de l'âge.
- calculer puis représenter graphiquement la densité de proportions de l'âge.
- quel est l'intervalle modal de l'âge ?
- représenter graphiquement puis calculer la proportion d'adolescents de moins de 15 ans.
- représenter graphiquement puis calculer la proportion d'adolescents de plus de 15 ans.
- représenter graphiquement puis calculer la proportion d'adolescents âgés de 14 à 17 ans.
- donner la distribution des proportions cumulées de l'âge.
- représenter graphiquement la fonction de répartition de l'âge.
- en utilisant la fonction de répartition, calculer la proportion d'adolescents de moins de 15 ans, de plus de 15 ans, âgés de 14 à 17 ans.
- représenter graphiquement les trois quartiles de l'âge.
- en identifiant chaque intervalle d'âge à son milieu, calculer les moyenne, variance et écart-type de l'âge.

#### **Exercice 4**

On relève pour les sujets d'une population de 60 fumeurs, la durée pendant laquelle ils ont fumé, exprimée en années. Les résultats obtenus sont les suivants :

dizaine	unités
0	355569
1	0001123444555677888999
2	00122222344567889
3	1224568
4	02557
5	129

- Quel est le nom de ce diagramme ?
- Préciser le nom et le type de la variable étudiée.
- Calculer la moyenne, la variance et l'écart-type de la variable étudiée.
- On découpe l'échelle des durées ]0 ; 60] en 6 intervalles de longueur 10 ans : ]0 ; 10], ]10 ; 20], ... et ]50 ; 60].
  - Donner les distributions d'effectifs et de proportions de la variable étudiée pour ce découpage.
  - Représenter graphiquement la densité de proportion de la variable étudiée pour ce découpage (préciser la nature des axes représentés et l'échelle choisie pour chacun d'eux).
  - Donner l'intervalle modal de la variable étudiée pour ce découpage.
- Représenter graphiquement sur l'histogramme, puis calculer la proportion de sujets ayant fumé pendant moins de 20 ans.
  - Calculer la proportion de sujets ayant fumé pendant plus de 20 ans.
  - Représenter graphiquement sur l'histogramme, puis calculer la proportion de sujets ayant fumé pendant plus de 30 ans; plus de 50 ans.
  - Représenter graphiquement sur l'histogramme, puis calculer la proportion de sujets ayant fumé pendant plus de 20 ans mais moins de 50 ans.
- Représenter graphiquement la fonction de répartition de la variable étudiée pour ce découpage (préciser la nature des axes représentés et l'échelle choisie pour chacun d'eux).
  - Déterminer la médiane, le troisième quartile et le quantile d'ordre 95% de la variable étudiée pour ce découpage.
  - Déterminer graphiquement le premier quartile de la variable étudiée pour ce découpage.

- 7) En revenant au diagramme initial :
  - a. déterminer les trois quartiles de la variable étudiée et les comparer à ceux trouvés précédemment.
  - b. déterminer les 1<sup>er</sup> et 9<sup>ème</sup> déciles ; en déduire l'intervalle de variation à 80% de la durée.
  - c. déterminer les quantiles d'ordre 5% et 95% ; en déduire l'intervalle de variation au risque 10% de la durée.
- 8) En identifiant chaque intervalle du découpage à son milieu :
  - a. calculer la durée moyenne pendant laquelle les sujets de la population ont fumé.
  - b. calculer l'écart-type de la durée.
  - c. comparer les résultats avec ceux obtenus à la question 3.
- 9) En supposant que tous les sujets de la population seront vivants et continueront de fumer :
  - a. quelle sera la durée moyenne pendant laquelle ils auront fumé, 5 ans plus tard ?
  - b. quel sera l'écart-type de la durée 5 ans plus tard ?
  - c. quelle sera la durée médiane 5 ans plus tard ?
- 10) On rajoute à cette population de 60 sujets, 30 sujets qui fument depuis 60 ans :
  - a. quelle est la durée moyenne associée au découpage dans la population totale des 90 sujets ?
  - b. quelle est la durée médiane associée au découpage dans la population totale des 90 sujets ?

### **Exercice 5**

Pour étudier le développement mental d'enfants âgés de 18 mois, on relève pour chaque enfant de la population son score obtenu sur l'échelle de Bayley (Bayley Scales of Infant Development).

On assimile la variable score à une variable de type quantitatif continu. Les scores, regroupés par intervalles, sont les suivants :

intervalle	]60 ; 70]	]70 ; 80]	]80 ; 90]	]90 ; 100]	]100 ; 110]
effectif	5	10	9	31	51

  

intervalle	]110 ; 120]	]120 ; 130]	]130 ; 140]	]140 ; 150]	]150 ; 160]
effectif	59	31	21	9	2

Source : d'après Florey C. du V., Leech A. M. et Blackhall A., *International Journal of Epidemiology* 1995, 24, 3, S21-S26

- 1) Préciser la population étudiée et sa taille.
- 2) On considère le découpage de l'échelle des scores ]60 ; 160] formé par les 10 intervalles du tableau ci-dessus.
  - a. Calculer la distribution de proportions du score associée à ce découpage.
  - b. Tracer l'histogramme du score associé à ce découpage (préciser la nature des axes représentés et l'échelle choisie pour chacun d'eux).
  - c. Quel est l'intervalle modal du score associé à ce découpage ?
  - d. Représenter graphiquement sur l'histogramme puis calculer la proportion d'enfants ayant un score inférieur à 100, inférieur à 120, compris entre 100 et 120, supérieur à 140.
- 3) a. Calculer la distribution des proportions cumulées du score associée à ce découpage.
  - b. Représenter graphiquement la fonction de répartition du score associée à ce découpage (préciser la nature des axes représentés et l'échelle choisie pour chacun d'eux).
  - c. En utilisant la fonction de répartition, calculer la proportion d'enfants ayant un score inférieur à 100, inférieur à 120, compris entre 100 et 120, supérieur à 140.
- 4) a. Déterminer graphiquement (sur le graphique obtenu en 3.b) les trois quartiles de la distribution du score.
  - b. Déduire, en utilisant le résultat obtenu en 3.c, le quantile d'ordre 95% du score.
  - c. Déterminer graphiquement l'intervalle de variation à 90% du score, l'intervalle de variation au risque 20% du score.
- 5) Calculer le score moyen des enfants de la population, la variance et l'écart-type du score, en identifiant chaque intervalle du découpage à son milieu.
- 6) On rajoute 16 enfants âgés de 18 mois et ayant tous obtenu un score inférieur à 60. Le score moyen de ces 16 enfants est de 30 :
  - a. calculer le score moyen pour la population totale.
  - b. quel est le score médian pour la population totale ?

## Lois normales (gaussiennes)

### Exercice 1

Z désigne la variable normale centrée et réduite.

- 1) Quelle est la proportion de valeurs de Z inférieures à 1,07 ?
- 2) Calculer la proportion de valeurs de Z supérieures à 1,07.
- 3) Calculer la proportion de valeurs de Z inférieures à -1,07.
- 4) Calculer la proportion de valeurs de Z supérieures à -2,58.
- 5) Calculer la probabilité pour que Z soit comprise entre 1,07 et 2,58.
- 6) Calculer la proportion de valeurs de Z comprises entre -2,76 et 1,42.
- 7) Calculer la probabilité pour que Z soit comprise entre -2,58 et -1,75.
- 8) Calculer la probabilité pour que Z soit inférieure à -1,84 ou supérieure à -0,06.
- 9) Trouver la médiane et les quartiles de Z.
- 10) Quels sont les 1<sup>er</sup>, 2<sup>ème</sup>, 8<sup>ème</sup> et 9<sup>ème</sup> déciles de Z.
- 11) Quels sont les 95<sup>ème</sup>, 99<sup>ème</sup>, 5<sup>ème</sup> et 1<sup>er</sup> percentiles de Z.
- 12) Donner les 8<sup>ème</sup>, 41<sup>ème</sup> et 77<sup>ème</sup> percentiles de Z.
- 13) Quels sont les quantiles d'ordre 97,5%, 99,5%, 2,5% et 0,5% de Z.
- 14) Déterminer l'intervalle de milieu 0 dans lequel se trouvent 90% des valeurs de Z ; comment se nomme cet intervalle ?
- 15) Quel est l'intervalle de variation de Z au niveau 80% ?
- 16) Quel est l'intervalle de variation de Z au risque 5% ?
- 17) Quel est l'intervalle de variation de Z à 99% ?
- 18) Quel est l'intervalle de variation de Z au risque 11% ?
- 19) On découpe la droite des nombres réels en 6 intervalles, choisis de la façon suivante :
  - la probabilité pour que Z soit inférieure à **E** vaut 15%
  - la probabilité pour que Z soit comprise entre **E** et **D** vaut 15%
  - la probabilité pour que Z soit comprise entre **D** et **C** vaut 20%
  - la probabilité pour que Z soit comprise entre **C** et **B** vaut 20%
  - la probabilité pour que Z soit comprise entre **B** et **A** vaut 15%
  - la probabilité pour que Z soit supérieure à **A** vaut 15%Calculer A, B, C, D et E (il est conseillé de s'aider d'un graphique).

### Exercice 2

On suppose que la distribution du quotient intellectuel QI dans une population d'enfants âgés de 7 ans suit un modèle normal de moyenne 100 et d'écart-type 10.

- 1) Un premier enfant de 7 ans a un QI de 106, un second a un QI de 85 :
  - a. quelle est la proportion d'enfants de la population qui ont un QI inférieur au QI du premier enfant ?
  - b. quelle est la proportion d'enfants de la population qui ont un QI inférieur au QI du second enfant ?
  - c. quelle est la proportion d'enfants de la population qui ont un QI compris entre le QI du second enfant et le QI du premier enfant ?
- 2) Quel est le QI au dessous duquel se situent 95% des enfants de 7 ans ?
- 3) Quel est le QI au dessous duquel se situent 5% des enfants de 7 ans ?
- 4) Quel est l'intervalle de variation à 95% du QI des enfants de 7 ans ?

### Exercice 3

La distribution de l'âge des français du recensement de 1999 est considérée comme normale de moyenne 39 ans et d'écart-type 23 ans :

- 1) Quelle est la proportion de français ayant moins de 80 ans ?
- 2) Quelle est la proportion de français ayant plus de 80 ans ?
- 3) Quelle est la proportion de français ayant moins de 20 ans ?
- 4) Quelle est la proportion de français ayant entre 20 et 60 ans ?
- 5) Quel est l'âge en dessus duquel se trouvent 5% des français ?
- 6) Quel est l'âge en dessous duquel se trouvent 10% des français ?
- 7) Dans quel intervalle se trouvent les âges de 90% des français, de 95% des français ?
- 8) Si l'on avait voulu que 20% des français soient à la retraite en 1999, quel aurait dû être l'âge de la retraite en 1999 ?

#### **Exercice 4**

Un score obtenu au questionnaire d'auto-évaluation des jeunes d'Achenbach permet de mesurer le nombre total de problèmes comportementaux signalés par l'enfant, pondérés selon la gravité du problème. Un score élevé correspond à un grand nombre de problème comportementaux. Dans une population d'enfants, la distribution de ce score est supposée normale de moyenne 50 et d'écart-type 10.

- 1) Quelle est la proportion d'enfants ayant un score inférieur à 74 dans la population ?
- 2) Quelle est la proportion d'enfants ayant un score inférieur à 34 dans la population ?
- 3) Quelle est la proportion d'enfants ayant un score compris entre 20 et 74 dans la population ?
- 4) Quel serait le score limite pour le diagnostic si l'on voulait identifier les enfants dont le score se situe dans les 2% supérieurs de la population ?
- 5) Quelles sont les bornes de l'intervalle, de milieu 50, dans lequel se situent les scores de 90% des enfants de la population ?
- 6) On désire classer les enfants en 5 classes de scores, déterminées pour contenir des proportions identiques d'enfants. On note A la borne supérieure de la première classe (contenant les scores les plus faibles), B celle de la seconde classe, C celle de la troisième et D celle de la quatrième. Représenter graphiquement les bornes sur la distribution du score puis déterminer les valeurs A, B, C et D.

#### **Exercice 5**

Le test de Fagerström mesure le degré de dépendance tabagique. Le score obtenu à ce test indique le degré de dépendance :

- score  $\leq 3$  : fumeur peu ou pas dépendant à la nicotine,
- $3 < \text{score} \leq 6$  : fumeur dépendant,
- $6 < \text{score} \leq 9$  : fumeur fortement dépendant,
- score  $> 9$  : fumeur très fortement dépendant.

La distribution du score obtenu à ce questionnaire dans une population de fumeurs, est supposée normale de moyenne 5 et d'écart-type 4,5.

- 1) Quelle est la proportion de fumeurs ayant un score inférieur à 6 dans la population ?
- 2) Quelle est la proportion de fumeurs dépendant à la nicotine dans la population ?
- 3) Quelle est la proportion de fumeurs très fortement dépendant à la nicotine dans la population ?
- 4) Quel est le score au dessus duquel se situent 10% des fumeurs de la population ?
- 5) Quel est l'intervalle de variation au risque 20% du score des fumeurs de la population ?
- 6) On désire faire 4 classes de scores, différentes de celles données dans l'énoncé, définies de manière à ce qu'elles contiennent des proportions identiques de fumeurs. On note A la borne supérieure de la première classe, B celle de la seconde et C celle de la troisième classe : Représenter graphiquement la distribution du score et les bornes, puis déterminer les valeurs de A, B et C.

#### **Exercice 6**

Les scores à un inventaire d'estime de soi sont supposés suivre un modèle normal de moyenne 32 et d'écart-type 14 dans la population étudiée.

- 1) Quelle est la proportion de personnes ayant score inférieur à 50 ?
- 2) Quelle est la proportion de personnes ayant score supérieur à 20 ?
- 3) Quel est le score au dessous duquel se situent 90% des personnes ?
- 4) Quel est le score au dessous duquel se situent 10% des personnes ?
- 5) Dans quel intervalle de milieu 32 se situent les scores de 80% des personnes ?
- 6) On désire regrouper les scores en cinq classes :
  - les 10% des scores les plus élevés sont affectés à la classe 5,
  - les 20% des scores suivants sont affectés à la classe 4,
  - les 40% des scores suivants sont affectés à la classe 3,
  - les 20% des scores suivants sont affectés à la classe 2,
  - les 10% des scores les plus faibles sont affectés à la classe 1.

Représenter graphiquement la distribution des scores et utiliser les résultats des questions 3 et 4, pour trouver les quatre valeurs permettant de définir les bornes de ces cinq classes.

#### **Exercice 7**

Le test de Boëhm-R est destiné à apprécier chez l'enfant la maîtrise des concepts de base, essentiels pour la compréhension des consignes verbales et fondamentaux pour la réussite scolaire. On suppose que le résultat à ce test suit une distribution normale de moyenne 60 et d'écart-type 17 dans la population des enfants de grandes sections d'une école maternelle.

- 1) Quelle est la proportion d'enfants ayant un résultat inférieur à 70 ?

- 2) Quelle est la proportion d'enfants ayant un résultat inférieur à 30 ?
- 3) Si l'on décidait d'accepter au cours préparatoire (CP) tous les enfants ayant un résultat supérieur à 40, quelle serait la proportion d'enfants admis en CP ?
- 4) En revanche, si l'on décidait d'accepter 99% des enfants de grande section en CP, quel serait le résultat minimum requis pour passer au CP ?
- 5) Quel est l'intervalle de variation au risque 1% du score des enfants de la population ?
- 6) On désire regrouper les résultats en trois catégories (concepts non maîtrisés, concepts en cours d'acquisition, concepts maîtrisés) contenant chacune un tiers des enfants : calculer les deux valeurs permettant de définir les bornes de ces trois catégories.

### **Exercice 8**

On étudie la symptomatologie dépressive des personnes d'une population à l'aide du score obtenu sur l'échelle "Center for Epidemiological Depression" noté CES-D. On suppose que la distribution de ce score dans la population suit un modèle normal de moyenne 12 et d'écart-type 4. Plus le score est élevé, plus la symptomatologie dépressive est importante.

- 1) Si l'on considère qu'une personne a un risque élevé de dépression lorsque son score est supérieur à 17 :
  - a. quelle est la proportion de personnes n'ayant pas un risque élevé de dépression dans la population étudiée ?
  - b. quelle est la proportion de personnes ayant un risque élevé de dépression dans la population étudiée ?
- 2) Quelle est la proportion de personnes ayant un score inférieur à 10 ?
- 3) Si l'on fait l'hypothèse que 10% des personnes de la population ont un risque élevé de dépression, à partir de quel score considère-t-on que le risque de dépression d'une personne est élevé ?
- 4) Quel est le score au dessous duquel se situent 95% des personnes ?
- 5) Quel est l'intervalle de variation du score CES-D au risque 5% ?

### **Exercice 9**

La distribution des scores à un inventaire d'évaluation de l'humeur est supposée normale de moyenne 8 et d'écart-type 2 dans la population étudiée.

- 1) Quelle est la proportion de sujets ayant un score inférieur à 10 ?
- 2) Quelle est la proportion de sujets ayant un score inférieur à 3 ?
- 3) Quelle est la proportion de sujets ayant un score compris entre 3 et 12 ?
- 4) Calculer les bornes A, B, C et D permettant de créer 5 classes de scores de la manière suivante (il est conseillé de s'aider d'un graphique) :
  - la proportion de scores inférieurs à B vaut 10%
  - la proportion de scores compris entre B et D vaut 20%
  - la proportion de scores compris entre C et A vaut 20%
  - la proportion de scores supérieurs à A vaut 10%
- 5) Quel est l'intervalle de variation à 95% du score des sujets de la population ?

### **Exercice 10**

On évalue le comportement neuro-cognitif des enfants âgés de 16 à 30 mois à l'aide du score obtenu sur l'échelle "Bayley Scales of Infant Development". Plus le score est faible, plus le retard de développement est important. On suppose que dans une population d'enfants, la distribution de ce score suit un modèle normal de moyenne 100 et d'écart-type 14.

- 1) Un premier enfant a obtenu un score de 118 : quelle est la proportion d'enfants de la population ayant un score inférieur à celui de cet enfant ?
- 2) Un second enfant a obtenu un score de 62 : quelle est la proportion d'enfants de la population ayant un score supérieur à celui de cet enfant ?
- 3) Quelle est la proportion d'enfants de la population ayant un score compris entre ceux de ces deux enfants ?
- 4) Un troisième enfant a obtenu un score de 159 : quelle est la proportion d'enfants de la population ayant un score supérieur à celui de cet enfant ?
- 5) Si l'on supposait que 0,1% des enfants ont un développement plus lent que la normale, au dessous de quel score considérerait-on qu'un enfant a un développement retardé ?
- 6) Quel est l'intervalle de variation du score au risque 2% des enfants de la population ?
- 7) On désire classer les enfants en 5 classes de scores, déterminées pour contenir des proportions identiques d'enfants. On note A la borne supérieure de la première classe (contenant les scores les plus faibles), B celle de la seconde classe, C celle de la troisième et D celle de la quatrième. Représenter graphiquement les bornes sur la distribution du score puis déterminer les valeurs A, B, C et D.

## Estimation ponctuelle

### Exercice 1

Vingt enfants fréquentant la maternelle ont été soumis à des tests de Pensée Créative de Torrance, pour mesurer leur créativité. Les scores obtenus sont les suivants :

15 34 23 18 22 21 25 28 14 17 20 16 19 24 29 21 26 30 13 11

- 1) Définir la population étudiée.
- 2) Définir la variable étudiée, préciser son type ainsi que le(s) paramètre(s) correspondant(s).
- 3) Donner une estimation ponctuelle du score moyen aux tests  $\mu$  pour la population des enfants fréquentant l'école maternelle.
- 4) Donner une estimation ponctuelle de la variance du score  $\sigma^2$  et de l'écart-type du score  $\sigma$  pour la population.

### Exercice 2

On mesure pour 15 individus âgés de 20 à 30 ans, le temps nécessaire à chacun d'eux pour reproduire 16 modèles. Les résultats observés (en secondes) sont les suivants :

190 505 550 468 366 374 451 239 356 528 450 420 269 430 418

- 1) Définir la population étudiée.
- 2) Définir la variable étudiée, préciser son type ainsi que le(s) paramètre(s) correspondant(s).
- 3) Donner une estimation ponctuelle du temps moyen  $\mu$  pour la population (vérifier que  $\Sigma x = 6\ 014$ ).
- 4) Donner une estimation ponctuelle de la variance du temps  $\sigma^2$  et de son écart-type  $\sigma$  pour la population (vérifier que :  $\Sigma x^2 = 2\ 563\ 408$ ).

### Exercice 3

Parmi 262 patients lombalgiques on dénombre 136 femmes et 99 patients présentant un trouble psychologique.

- 1) Définir la population étudiée.
- 2) Estimer la proportion de femmes parmi les patients lombalgiques. Au préalable définir la variable étudiée en précisant son type ainsi que le(s) paramètre(s) correspondant(s).
- 3) Estimer la fréquence des troubles psychologiques chez les patients lombalgiques. Définir au préalable la variable étudiée en précisant son type ainsi que le(s) paramètre(s) correspondant(s).

### Exercice 4

On a administré un traitement antidépresseur à 1 197 patients souffrant de dépression résistante. Après 28 jours de traitement, on a observé une amélioration de l'état clinique pour 1 113 patients.

- 1) Définir la population étudiée.
- 2) Définir la variable étudiée, préciser son type ainsi que le(s) paramètre(s) correspondant(s).
- 3) Estimer ponctuellement la proportion d'amélioration de l'état clinique.

### Exercice 5

Une enquête concernant la santé et portant sur 3 279 adolescents français âgés de 12 à 20 ans, a dénombré 689 adolescents ayant pris un médicament psychotrope au cours des 12 mois précédant l'enquête. Parmi les 1 728 filles, 475 ont pris un médicament psychotrope.

Donner une estimation ponctuelle de la fréquence de consommation de médicaments psychotropes chez les adolescents français. Même question chez les filles, puis chez les garçons.

Pour chaque question, définir au préalable la population et la variable étudiées, préciser le type de la variable ainsi que le(s) paramètre(s) correspondant(s).

### **Exercice 6**

Les scores obtenus par 216 personnes à un inventaire d'évaluation de l'humeur sont les suivants :

score	1	2	3	4	5	6	7	8	9	10	11	12	13
effectif	2	2	5	3	14	37	32	39	35	16	16	13	2

- 1) Définir la population étudiée.
- 2) Définir la variable étudiée, préciser son type ainsi que le(s) paramètre(s) correspondant(s).
- 3) Estimer le score moyen dans la population.
- 4) Estimer l'écart-type du score dans la population.

### **Exercice 7**

Les résultats à un test de dextérité manuelle de 16 handicapés mentaux choisis au hasard sont les suivants :

72 73 73 74 74 75 76 66 66 67 65 73 71 70 71 70

- 1) Définir la population étudiée.
- 2) Définir la variable étudiée, préciser son type ainsi que le(s) paramètre(s) correspondant(s).
- 3) Donner une estimation ponctuelle du résultat moyen des handicapés mentaux dans cette population.
- 4) Donner une estimation ponctuelle de l'écart-type du résultat des handicapés mentaux dans cette population.

### **Exercice 8**

Une enquête norvégienne concernant la santé psychologique des femmes entre 20 et 49 ans réalisée sur 3 103 femmes a dénombré 571 femmes dépressives. Parmi les 702 femmes au foyer, 160 sont dépressives, alors que 411 le sont parmi les femmes en activité professionnelle.

- 1) Donner une estimation ponctuelle de la fréquence de la dépression chez les femmes norvégiennes de 20 à 49 ans.
- 2) Même question pour les femmes au foyer, puis pour celles qui ont une activité professionnelle.

### **Exercice 9**

La santé de 70 nouveau-nés prématurés (nés avant 30 semaines de gestation) a été mesurée par le score d'Apgar à 5 minutes. Les scores obtenus sont tels que :

$$\sum x_i = 567 \quad \sum x_i^2 = 4\,817$$

- 1) Donner une estimation ponctuelle du score d'Apgar moyen pour la population des nouveau-nés prématurés.
- 2) Donner une estimation ponctuelle de la variance du score d'Apgar et de l'écart-type de ce score pour la population des nouveau-nés prématurés.

### **Exercice 10**

Le test de Reynell mesure les capacités d'expression et de compréhension verbale des enfants de 6 mois à 6 ans en tenant compte de leur âge. Le score obtenu est tel qu'un enfant "normal" d'un âge donné a un score de 0.

Le score de Reynell a été mesuré chez 77 enfants atteints d'otite moyenne avec écoulement (OME) bilatérale chronique pour savoir si la perte d'audition temporaire occasionnée par l'OME chronique pouvait retarder à court ou à long terme l'acquisition du langage chez les jeunes enfants.

Les scores obtenus sont tels que :

$$\sum x_i = -27 \quad \sum x_i^2 = 86$$

- 1) Donner une estimation ponctuelle du score de Reynell moyen dans la population des enfants atteints d'OME chronique.
- 2) Donner une estimation ponctuelle de la variance et de l'écart-type du score de Reynell dans la population des enfants atteints d'OME chronique.

### **Exercice 11**

Les antidépresseurs tricycliques (TCA) sont associés à une augmentation du risque de chute et donc de fracture de la hanche chez les personnes âgées. Sur 5 838 personnes de plus de 66 ans consommateurs de TCA on a dénombré 1 493 fractures de la hanche.

Estimer ponctuellement la fréquence des fractures de hanche dans la population des consommateurs de TCA de plus de 66 ans.



### **Exercice 12**

Parmi 55 patients dépressifs âgés de 18 à 65 ans, 36 sont guéris après 16 semaines de psychothérapie. Parmi les réponses proposées, indiquer la(es) réponse(s) exacte(s) :

1) Définir la population étudiée :

- a.  $\mathcal{P} = \{\text{patients dépressifs âgés de 18 à 65 ans}\}$
- b.  $\mathcal{P} = \{\text{patients dépressifs âgés de 18 à 65 ans guéris après 16 semaines de psychothérapie}\}$
- c.  $\mathcal{P} = \{55 \text{ patients dépressifs âgés de 18 à 65 ans guéris après 16 semaines de psychothérapie}\}$
- d.  $\mathcal{P} = \{36 \text{ patients dépressifs âgés de 18 à 65 ans guéris après 16 semaines de psychothérapie}\}$
- e.  $\mathcal{P} = \{\text{patients âgés de 18 à 65 ans}\}$
- f.  $\mathcal{P} = \{\text{patients dépressifs}\}$
- g.  $\mathcal{P} = \{\text{patients dépressifs âgés de 18 à 65 ans, traités par psychothérapie pendant 16 semaines}\}$

2) Définir la variable étudiée :

- a.  $X = \text{patient dépressif}$
- b.  $X = \text{patient dépressif guéri}$
- c.  $X = \text{nombre de patients dépressifs}$
- d.  $X = \text{nombre de patients guéris}$
- e.  $X = \text{nombre de patients dépressifs guéris après 16 semaines de psychothérapie}$
- f.  $X = \text{guérison chez les patients dépressifs}$
- g.  $X = \text{proportion de guérison chez les patients dépressifs}$
- h.  $X = \text{guérison chez les patients dépressifs âgés de 18 à 65 ans, traités par psychothérapie pendant 16 semaines}$
- i.  $X = \text{proportion de guérison chez les patients dépressifs âgés de 18 à 65 ans, traités par psychothérapie pendant 16 semaines}$
- j.  $X = \text{guérison}$

3) Définir le type de la variable étudiée :

- a.  $X$  qualitative à deux modalités : oui, non
- b.  $X$  qualitative de moyenne  $\mu$  et d'écart-type  $\sigma$
- c.  $X$  quantitative de moyenne  $\mu$  et d'écart-type  $\sigma$

4) Définir le(es) paramètre(s) étudié(s) :

- a. deux paramètres :  $\mu$  nombre moyen de guérison et  $\sigma$  écart-type du nombre de guérison dans  $\mathcal{P}$
- b. un paramètre :  $\mu$  nombre moyen de guérison dans  $\mathcal{P}$
- c. un paramètre :  $p$  proportion de guérison dans  $\mathcal{P}$

5) Donner une estimation ponctuelle de la proportion de guérison dans la population étudiée :

- a. la proportion de guérison  $p$  est égale à 65,5%
- b. la proportion de guérison  $p$  est estimée à 0,655
- c. la proportion de guérison  $f$  est estimée à 0,655
- d. la proportion observée de guérison  $f$  est égale à 0,655
- e. la proportion de guérison est estimée par la fréquence observée égale à 65,5%

### **Exercice 13**

La prise de poids moyenne pour 38 personnes de plus de 18 ans ayant arrêté de fumer est de 1,5 kg avec un écart-type  $s$  de 2,96 kg.

Parmi les réponses proposées cocher les bonnes réponses :

1) Définir la population étudiée :

- a.  $\mathcal{P} = \{38 \text{ personnes de plus de 18 ans}\}$
- b.  $\mathcal{P} = \{38 \text{ personnes de plus de 18 ans ayant arrêté de fumer}\}$
- c.  $\mathcal{P} = \{\text{personnes de plus de 18 ans}\}$
- d.  $\mathcal{P} = \{\text{personnes de plus de 18 ans ayant arrêté de fumer}\}$
- e.  $\mathcal{P} = \{\text{personnes ayant arrêté de fumer}\}$

2) Définir la variable étudiée :

- a.  $X$  = nombre de personnes de plus de 18 ans ayant arrêté de fumer
- b.  $X$  = prise de poids moyenne pour 38 personnes de plus de 18 ans ayant arrêté de fumer
- c.  $X$  = prise de poids moyenne des personnes de plus de 18 ans ayant arrêté de fumer
- d.  $X$  = prise de poids des personnes de plus de 18 ans ayant arrêté de fumer
- e.  $X$  = personne de plus de 18 ans ayant arrêté de fumer
- f.  $X$  = prise de poids

3) Définir le type de la variable étudiée :

- a.  $X$  qualitative à deux modalités : oui, non
- b.  $X$  qualitative de moyenne  $\mu$  et d'écart-type  $\sigma$
- c.  $X$  quantitative de moyenne  $\mu$  et d'écart-type  $\sigma$
- d.  $X$  quantitative de moyenne  $\bar{x}$  et d'écart-type  $s$

4) Définir le(es) paramètre(s) étudié(s) :

- a. un paramètre :  $\mu$  prise de poids moyenne dans  $\mathcal{P}$
- b. deux paramètres :  $\mu$  prise de poids moyenne et  $\sigma$  écart-type de la prise de poids dans  $\mathcal{P}$
- c. deux paramètres :  $\mu$  prise de poids moyenne et  $\sigma$  écart-type de la prise de poids moyenne dans  $\mathcal{P}$
- d. un paramètre :  $p$  proportion de prise de poids dans  $\mathcal{P}$

5) Donner une estimation ponctuelle de la prise de poids moyenne dans la population étudiée :

- a. la prise de poids moyenne  $\bar{x}$  est estimée à 1,5 kg
- b. la prise de poids moyenne  $\mu$  est égale à 1,5 kg
- c. la prise de poids moyenne  $\mu$  est estimée à 0,04 kg
- d. la prise de poids moyenne  $\mu$  est estimée à 1,5 kg
- e. la prise de poids moyenne observée  $\bar{x}$  est égale à 1,5 kg

6) Donner une estimation ponctuelle de l'écart-type de la prise de poids dans la population étudiée :

- a. l'écart-type de la prise de poids  $\sigma$  est égal à 3 kg
- b. l'écart-type de la prise de poids moyenne  $\sigma$  est estimé à 2,96 kg
- c. l'écart-type de la prise de poids  $\sigma$  est estimé à 3 kg
- d. l'écart-type de la prise de poids  $\sigma$  est estimé à 3,04 kg
- e. l'écart-type observé de la prise de poids  $s$  est égal à 2,96 kg
- f. l'écart-type observé sans biais de la prise de poids  $s$  est égal à 3 kg
- g. l'écart-type observé de la prise de poids moyenne  $s^*$  est estimé à 3 kg

## Distributions d'échantillonnage

### Intervalles de variation

#### Exercice 1

Les normes des tests de Pensée Créative de Torrance indiquent que les élèves du secondaire obtiennent un résultat moyen de fluidité (aptitude à produire un grand nombre d'idées) de 20 avec un écart-type de 6,5. On va prélever au hasard trois échantillons d'élèves du secondaire comprenant respectivement 20, 50 et 100 élèves auxquels on administrera le test.

- 1) Peut-on prévoir (prédire) le résultat moyen obtenu sur chaque échantillon ?
- 2) Peut-on prévoir la variance du résultat moyen obtenu sur chaque échantillon, et son écart-type ?
- 3) Peut-on prévoir la variance du résultat observée sur chaque échantillon, et l'écart-type correspondant ?

#### Exercice 2

La distribution de l'âge des français au recensement de 1999 est considérée comme normale de moyenne 39 ans et d'écart-type 23 ans.

On prélève un échantillon de l'âge de taille 25 parmi les français.

- 1) Caractériser la distribution de la moyenne empirique de l'âge sur les échantillons de taille 25, en précisant sa forme, sa moyenne, sa variance et son écart-type.
- 2) Calculer la probabilité d'observer sur un échantillon de taille 25, un âge moyen inférieur à 20 ans.
- 3) Calculer la probabilité d'observer sur un échantillon de taille 25, un âge moyen compris entre 20 et 60 ans.
- 4) Donner l'intervalle de variation à 90% de l'âge moyen sur les échantillons de taille 25, puis à 95%.
- 5) On observe sur l'échantillon, un âge moyen de 35 ans : cette observation est-elle surprenante ?
- 6) Quelle aurait dû être la taille de l'échantillon pour que la demi-longueur de l'intervalle de variation à 95% de l'âge moyen n'excède pas 2 ans.

#### Exercice 3

Le résultat à un test portant sur la richesse et la précision du vocabulaire chez les enfants de 12 ans a pour moyenne 60 et pour écart-type 10. On prélève un échantillon de taille  $n$  du résultat au test issu de la population étudiée.

- 1) Indiquer l'effet de la taille de l'échantillon sur les caractéristiques de la distribution de la moyenne empirique du résultat  $\bar{X}_n$  :

taille de l'échantillon	distribution de la moyenne empirique		
	moyenne	écart-type	forme
1			
4			
8			
16			
32			
64			
100			

- 2) Evaluer la probabilité d'observer un résultat moyen supérieur à 63 sur un échantillon de taille  $n = 16$ .
- 3) Même question pour un échantillon de taille  $n = 32$ .
- 4) Comment varie cette probabilité avec la taille de l'échantillon ?

#### Exercice 4

On suppose que la distribution du quotient intellectuel QI dans une population d'enfants âgés de 7 ans suit un modèle normal de moyenne 100 et d'écart-type 10. On fait passer le test de QI à 16 enfants de 7 ans choisis au hasard et on observe un QI moyen de 106 avec un écart-type de 13.

- 1) Compléter le tableau suivant :

	population	échantillon
taille	$N =$	$n =$
moyenne	$\mu =$	$\bar{x} =$
variance	$\sigma^2 =$	$s^2 =$

- 2) Caractériser la distribution de la moyenne empirique du QI sur les échantillons de taille 16, en précisant sa forme, sa moyenne, sa variance et son écart-type.
- 3) Calculer la probabilité d'observer sur un échantillon de taille 16, un QI moyen supérieur à celui observé.
- 4) Quel serait le QI moyen limite si l'on voulait identifier les échantillons de taille 16 dont le score moyen se situe dans les 2,5% supérieurs des échantillons ?
- 5) Donner l'intervalle de variation à 95% du QI moyen sur les échantillons de taille 16.
- 6) Mêmes questions (1 à 4) pour un échantillon de taille 100 sur lequel on a observé un score moyen de 106.
- 7) Quelle aurait dû être la taille de l'échantillon pour que la demi-longueur de l'intervalle de variation à 95% du QI moyen n'excède pas 1 point ?

### **Exercice 5**

Le Child Behavior Checklist (CBCL) est un questionnaire parental permettant d'évaluer les problèmes émotionnels et comportementaux des enfants. Plus le score obtenu à ce questionnaire est élevé plus les problèmes sont importants. Dans une population d'enfants âgés de 4 à 16 ans, le score obtenu à ce questionnaire a pour moyenne 20 et pour écart-type 14.

Sur un échantillon de 36 enfants âgés de 4 à 16 ans tirés au sort dans la population, on observe un score moyen de 20,5.

- 1) Compléter le tableau suivant :

	population	échantillon
taille		
moyenne		
variance		

- 2) Caractériser la distribution de la moyenne empirique du score pour les échantillons de taille 36 en précisant sa forme, sa moyenne et son écart-type.
- 3) Calculer la probabilité d'observer sur un échantillon de taille 36, un score moyen inférieur à celui observé.
- 4) Calculer la probabilité d'observer sur un échantillon de taille 36, un score moyen supérieur à 25.
- 5) A quelle valeur seraient inférieurs 99% des scores moyens pour les échantillons de taille 36 ?
- 6) Calculer l'intervalle de variation à 98% du score moyen pour les échantillons de taille 36.
- 7) Quelle aurait dû être la taille de l'échantillon pour que la demi-longueur de l'intervalle de variation à 98% du score moyen n'excède pas 4 points ?

### **Exercice 6**

La distribution des notes obtenues par les sujets d'une population au test Stroop mesurant la résistance au stress a pour moyenne 50 et pour écart-type 25.

On tire au hasard dans la population, un échantillon de taille 30 sur lequel on observe un score moyen de 41.

- 1) Caractériser la distribution de la moyenne empirique du score sur les échantillons de taille 30, en précisant sa forme, sa moyenne, sa variance et son écart-type.
- 2) Calculer la probabilité d'observer sur un échantillon de taille 30, un score moyen inférieur à celui observé.
- 3) Calculer la probabilité d'observer sur un échantillon de taille 30, un score moyen supérieur à celui observé.
- 4) Quel serait le score moyen limite si l'on voulait identifier les échantillons de taille 30 dont le score moyen se situe dans les 5% supérieurs des échantillons ?
- 5) Donner l'intervalle de variation à 90% du score moyen sur les échantillons de taille 30.

### **Exercice 7**

Le test de Fagerström mesure le degré de dépendance tabagique. Le score obtenu à ce test indique le degré de dépendance. Dans une population de fumeurs, le score obtenu à ce questionnaire a pour moyenne 5 et pour écart-type 4,5 (on ne suppose plus sa distribution normale).

On dispose d'un échantillon de 45 fumeurs tirés au sort dans cette population.

- 1) Caractériser la distribution de la moyenne empirique du score sur les échantillons de taille 45, en précisant sa forme, sa moyenne, sa variance et son écart-type.
- 2) Calculer la probabilité d'observer sur cet échantillon, un score moyen inférieur à 6.
- 3) Calculer la probabilité d'observer sur cet échantillon, un score moyen inférieur à 3.
- 4) Calculer la probabilité d'observer sur cet échantillon, un score moyen compris entre 3 et 6.
- 5) Calculer la probabilité d'observer sur cet échantillon, un score moyen supérieur à 9.

- 6) En considérant les scores moyens de tous les échantillons de taille 45, à quelle valeur 90% de ces scores moyens seront-ils inférieurs ?
- 7) Donner l'intervalle de variation à 80% du score moyen sur les échantillons de taille 45.

### Exercice 8

Selon une étude sur le comportement du consommateur 25% d'entre eux sont influencés par la marque de commerce lors de l'achat d'un produit. Un échantillon de 100 consommateurs choisis au hasard sont interrogés.

1) Compléter le tableau suivant :

	population	échantillon
taille	$N =$	$n =$
proportion	$p =$	$f =$

- 2) Caractériser la distribution de la fréquence empirique de l'influence par la marque de commerce sur les échantillons de taille 100, en précisant sa forme, sa moyenne, sa variance et son écart-type.
- 3) Calculer la probabilité pour qu'au moins 35 consommateurs se déclarent influencés par la marque de commerce.
- 4) Calculer la probabilité pour que moins de 20 consommateurs se déclarent influencés par la marque de commerce.
- 5) Donner l'intervalle de variation à 90% de la proportion des consommateurs influencés par la marque de commerce sur un échantillon de taille 100.
- 6) Sur un échantillon de 100 consommateurs, on observe que 31 d'entre eux sont influencés par la marque de commerce : que peut-on en déduire ?

### Exercice 9

La proportion de lancers d'une pièce de monnaie équilibrée qui tombent sur le côté face est de 50%. On procède à une série de  $n$  lancers de la pièce.

1) Indiquer l'effet de la taille de l'échantillon sur les caractéristiques de la distribution de la fréquence empirique du résultat  $F_n$  :

taille de l'échantillon $n$	distribution de la fréquence empirique		
	moyenne	écart-type	forme
5			
10			
30			
50			
100			

- 2) Evaluer la probabilité d'observer plus de 18 faces sur un échantillon de 30 lancers.
- 3) Evaluer la probabilité d'observer plus de 60% de faces sur un échantillon de 50 lancers.
- 4) Comment varie cette probabilité avec la taille de l'échantillon ?
- 5) Combien de faces s'attend-on "raisonnablement" à observer sur un échantillon de 100 lancers ?
- 6) On observe 35 faces sur un échantillon de 100 lancers : que peut-on en déduire ?
- 7) Quelle aurait dû être la taille de l'échantillon pour que la demi-longueur de l'intervalle de variation à 95% de la proportion de face n'excède pas 5%.

### Exercice 10

Dans une université 45% des étudiants s'adonnent à au moins une activité physique par semaine. Sur un échantillon de 400 étudiants tirés au sort dans cette université on dénombre 205 étudiants pratiquant au moins une activité physique par semaine.

1) Compléter le tableau suivant :

	population	échantillon
taille		
proportion		

- 2) Caractériser la distribution de la fréquence empirique de la pratique d'au moins une activité physique par semaine sur les échantillons de taille 400, en précisant sa forme, sa moyenne, sa variance et son écart-type.
- 3) Le résultat observé sur l'échantillon est-il surprenant ?
- 4) Donner l'intervalle de variation à 95% de la proportion des étudiants pratiquant au moins une activité physique par semaine, pour un échantillon de taille 400.
- 5) Quelle aurait dû être la taille de l'échantillon pour que la demi-longueur de l'intervalle de variation à 95% de la proportion des étudiants pratiquant au moins une activité physique par semaine n'excède pas 2%.

### **Exercice 11**

Dix pour cent des français consomment des antidépresseurs. On interroge un échantillon de 30 français puis un autre échantillon de 80 français.

- 1) Quelle est la distribution de la proportion empirique de consommateurs d'antidépresseurs sur les échantillons de français de taille 30 (forme, moyenne et écart-type).
- 2) Caractériser la distribution de la proportion empirique de consommateurs d'antidépresseurs sur les échantillons de français de taille 80, en précisant sa forme, sa moyenne et son écart-type.
- 3) Calculer la probabilité d'observer sur un échantillon de français de taille 80, plus de 12 consommateurs d'antidépresseurs.
- 4) Calculer la probabilité d'observer sur un échantillon de français de taille 80, moins de 2 consommateurs d'antidépresseurs.
- 5) Donner l'intervalle de variation au niveau 95% de la proportion empirique de consommateurs d'antidépresseurs sur les échantillons de français de taille 80.
- 6) Sur les 80 personnes interrogées, dix consomment des antidépresseurs : que peut-on en déduire ?

## Estimation ponctuelle et estimation par intervalle

### Exercice 1

On sait que les résultats à un test de QI se distribuent normalement avec un écart-type  $\sigma=13$  dans une population d'étudiants. Sur un échantillon de taille 30, on observe  $\bar{x} = 111$ .

- 1) Estimer la moyenne  $\mu$  des résultats avec un intervalle de confiance à 95%, à 90%, et à 99%.
- 2) Mêmes questions pour un résultat moyen identique observé sur 50 étudiants, puis sur 100 étudiants.
- 3) Quelle doit être la taille minimum de l'échantillon pour que la demi-longueur de l'intervalle de confiance à 95% n'excède pas 1 ?

### Exercice 2

Trente enfants fréquentant la maternelle ont été soumis à des tests de Pensée Créative de Torrance. Les scores obtenus sont tels que :

$$\sum x_i = 639 \quad \sum x_i^2 = 14\,691$$

- 1) Donner une estimation ponctuelle du score moyen aux tests  $\mu$  pour la population des enfants fréquentant l'école maternelle.
- 2) Donner une estimation ponctuelle de la variance du score  $\sigma^2$  et de l'écart-type du score  $\sigma$  pour la population.
- 3) Estimer par intervalle de confiance au niveau 95%, le score moyen  $\mu$ .

### Exercice 3

On mesure pour 60 individus âgés de 20 à 30 ans, le temps nécessaire à chacun d'eux pour reproduire 16 modèles. Les résultats observés (en secondes) sont tels que :

$$\sum x_i = 24\,056 \quad \sum x_i^2 = 10\,253\,632$$

- 1) Donner une estimation ponctuelle du temps moyen  $\mu$  pour la population.
- 2) Donner une estimation ponctuelle de la variance du temps  $\sigma^2$  et de son écart-type  $\sigma$  pour la population.
- 3) Estimer dans la population le temps moyen  $\mu$  par intervalle de confiance aux niveaux 90%, 95% et 99%.

### Exercice 4

On s'intéresse au temps de mémorisation d'un texte (mesuré en minutes) par les étudiants d'une promotion. Un échantillon de 37 étudiants fournit les valeurs  $\bar{x} = 25$  et  $s = 5$ .

Pour un risque  $\alpha=5\%$ , construire un intervalle de confiance du temps moyen de mémorisation  $\mu$  de la promotion.

### Exercice 5

On observe sur 100 sujets, le temps nécessaire pour parcourir un labyrinthe. Le temps moyen est de 8,76 minutes avec un écart-type  $s^*$  de 2,3 minutes.

- 1) Donner une estimation ponctuelle du temps moyen de parcours du labyrinthe.
- 2) Estimer le temps moyen par intervalle de confiance au niveau 90%.
- 3) En supposant que  $\sigma=2,3$ , quelle devrait être la taille de l'échantillon pour que la marge d'erreur dans l'estimation du temps moyen au niveau 90% ne soit pas supérieure à 0,3 minute ?

### Exercice 6

Les résultats à un test de dextérité manuelle de 32 handicapés mentaux choisis au hasard sont tels que :

$$\sum x_i = 2\,272 \quad \sum x_i^2 = 161\,664$$

Donner un intervalle de confiance à 99% du résultat moyen des handicapés mentaux.

### Exercice 7

La santé de 70 nouveau-nés prématurés (nés avant 30 semaines de gestation) a été mesurée par le score d'Apgar à 5 minutes. Les scores obtenus sont tels que :

$$\sum x_i = 567 \quad \sum x_i^2 = 4\,817$$

- 1) Donner une estimation ponctuelle du score d'Apgar moyen pour la population des nouveau-nés prématurés.



- 2) Donner une estimation ponctuelle de la variance du score d'Apgar et de l'écart-type de ce score pour la population des nouveau-nés prématurés.
- 3) Estimer par intervalle de confiance au niveau 90%, le score d'Apgar moyen pour la population des nouveau-nés prématurés.

### **Exercice 8**

Le test de Reynell mesure les capacités d'expression et de compréhension verbale des enfants de 6 mois à 6 ans en tenant compte de leur âge. Le score obtenu est tel qu'un enfant "normal" d'un âge donné a un score de 0.

Le score de Reynell a été mesuré chez 77 enfants atteints d'otite moyenne avec écoulement (OME) bilatérale chronique pour savoir si la perte d'audition temporaire occasionnée par l'OME chronique pouvait retarder à court ou à long terme l'acquisition du langage chez les jeunes enfants.

Les scores obtenus sont tels que :

$$\sum x_i = -27 \qquad \sum x_i^2 = 86$$

- 1) Donner une estimation ponctuelle du score de Reynell moyen dans la population des enfants atteints d'OME chronique.
- 2) Donner une estimation ponctuelle de la variance et de l'écart-type du score de Reynell dans la population des enfants atteints d'OME chronique.
- 3) Estimer par intervalle de confiance au niveau 95%, le score de Reynell moyen dans la population des enfants atteints d'OME chronique.
- 4) Quelle devrait être la taille de l'échantillon pour que la marge d'erreur dans l'estimation du score de Reynell moyen au niveau 95% ne soit pas supérieure à 0,1, en supposant que l'écart-type observé du score reste le même ?

### **Exercice 9**

Lors d'une élection présidentielle opposant deux candidats A et B, on a interrogé au hasard 100 électeurs : 52 ont déclaré avoir l'intention de voter pour A et les autres pour B.

- 1) Estimer ponctuellement la proportion d'intentions de vote pour le candidat A.
- 2) Estimer par intervalle de confiance à 95% la proportion d'intentions de vote pour le candidat A.
- 3) Même question pour une fréquence identique d'intentions de vote pour A observée sur 1 000 électeurs.
- 4) Quelle devrait être la taille de l'échantillon pour que la marge d'erreur dans l'estimation de cette proportion au niveau 95% n'excède pas 0,5%, en supposant que la fréquence observée d'intentions de vote pour A reste identique ?

### **Exercice 10**

Parmi 262 patients lombalgiques on dénombre 136 femmes et 99 patients présentant un trouble psychologique.

- 1) Donner une estimation ponctuelle de la proportion de femmes parmi les patients lombalgiques.
- 2) Déterminer un intervalle de confiance au niveau 95% de la proportion de femmes parmi les patients lombalgiques.
- 3) Estimer la fréquence des troubles psychologiques chez les patients lombalgiques.
- 4) Déterminer un intervalle de confiance au niveau 99% de la proportion de troubles psychologiques parmi les patients lombalgiques.

### **Exercice 11**

On a administré un traitement antidépresseur à 1 197 patients souffrant de dépression résistante. Après 28 jours de traitement, le traitement a été efficace pour 957 patients.

- 1) Estimer ponctuellement la proportion d'efficacité du traitement.
- 2) Estimer par intervalle de confiance à 90% la proportion d'efficacité du traitement.
- 3) Quelle devrait être la taille de l'échantillon pour que la demi-longueur de l'intervalle de confiance à 90% soit inférieure à 2% , en supposant que la fréquence observée d'efficacité du traitement reste la même ?

### **Exercice 12**

Les antidépresseurs tricycliques (TCA) sont associés à une augmentation du risque de chute et donc de fracture de la hanche chez les personnes âgées. Sur 5 838 personnes de plus de 66 ans consommateurs de TCA on a dénombré 1 493 fractures de la hanche.

- 1) Estimer ponctuellement la fréquence des fractures de hanche dans la population des consommateurs de TCA de plus de 66 ans.

- 2) Estimer par intervalle de confiance à 99% la fréquence des fractures de hanche dans la population des consommateurs de TCA de plus de 66 ans.

### **Exercice 13**

Une enquête concernant la santé et portant sur 3 279 adolescents français âgés de 12 à 20 ans, a dénombré 689 adolescents ayant pris un médicament psychotrope au cours des 12 mois précédant l'enquête. Parmi les 1 728 filles, 475 ont pris un médicament psychotrope, et 214 parmi les garçons.

- 1) Donner une estimation ponctuelle de la fréquence de consommation de médicaments psychotropes chez les adolescents français.
- 2) Estimer par intervalle de confiance à 95% la fréquence de consommation de médicaments psychotropes chez les adolescents français.
- 3) Mêmes questions chez les filles, puis chez les garçons.
- 4) Quelle devrait être la taille de l'échantillon pour que la précision de l'intervalle de confiance à 95% de la fréquence de consommation de médicaments psychotropes chez les adolescents français soit inférieure à 1%, en supposant que la fréquence observée de consommation de médicaments psychotropes n'est pas modifiée ?

### **Exercice 14**

Une enquête norvégienne concernant la santé psychologique des femmes entre 20 et 49 ans réalisée sur 3 103 femmes a dénombré 571 femmes dépressives. Parmi les 702 femmes au foyer, 160 sont dépressives, alors que 411 le sont parmi les femmes en activité professionnelle.

- 1) Donner une estimation ponctuelle de la fréquence de la dépression chez les femmes norvégiennes de 20 à 49 ans.
- 2) Estimer par intervalle de confiance à 95% la fréquence de la dépression chez les femmes norvégiennes de 20 à 49 ans.
- 3) Quelle devrait être la taille de l'échantillon pour que la demi-longueur de l'intervalle de confiance à 95% de la fréquence de la dépression chez les femmes norvégiennes de 20 à 49 ans soit inférieure à 1%, en supposant que la fréquence observée de dépression reste la même ?

### **Exercice 15**

Une publicité vantant les vertus d'une crème antirides annonce comme performances prouvées :

- "après un mois, une diminution de la profondeur des rides pour 93% des cas",
- "après 3 mois, une amélioration du micro relief cutané dans 84% des cas".

Il est précisé en petits caractères : "tests effectués sur 30 volontaires".

- 1) Pour combien de volontaires testées a-t-on observé une diminution de la profondeur des rides après un mois, une amélioration du micro relief cutané après 3 mois ?
- 2) Estimer par intervalle de confiance à 95% la proportion d'efficacité de la crème sur la diminution de la profondeur des rides après un mois.
- 3) Estimer par intervalle de confiance à 95% la proportion d'efficacité de la crème sur l'amélioration du micro relief cutané après 3 mois.

### **Exercice 16**

La prise de poids moyenne pour 38 personnes de plus de 18 ans ayant arrêté de fumer est de 1,5 kg avec un écart-type  $s^*$  de 3 kg. Parmi les réponses proposées indiquer celles qui sont exactes.

- 1) Donner une estimation ponctuelle de la prise de poids moyenne dans la population étudiée :
  - a. la prise de poids moyenne dans  $\mathcal{P}$  est égale à  $\bar{x} = 1,5$  kg
  - b. la prise de poids moyenne dans  $\mathcal{P}$  est égale à  $\mu = 1,5$  kg
  - c. la prise de poids moyenne dans  $\mathcal{P}$  est estimée par la moyenne observée  $\bar{x} = 1,5$  kg
  - d. la prise de poids moyenne dans  $\mathcal{P}$  est estimée par  $F_n = 1,5$  kg
  - e. la prise de poids moyenne dans  $\mathcal{P}$  est estimée par  $\bar{X}_n = 1,5$  kg
  - f. la prise de poids moyenne dans  $\mathcal{P}$  est estimée par  $\bar{x} = 1,5$  kg
  - g. la prise de poids moyenne dans  $\mathcal{P}$  est estimée par  $p = 1,5$  kg
- 2) Vrai ou faux :
  - a. l'estimation ponctuelle de la prise de poids moyenne dans  $\mathcal{P}$  fournit une information concernant la précision de l'estimation effectuée

- b. pour obtenir une certaine confiance dans l'estimation de la prise de poids moyenne dans  $\mathcal{P}$ , on a recours à l'estimation par intervalle de confiance
  - c. dans l'estimation par intervalle de confiance de la prise de poids moyenne dans  $\mathcal{P}$ , on associe à l'intervalle de confiance une probabilité de contenir la vraie valeur de la moyenne
  - d. la probabilité pour qu'un intervalle de confiance à 95% de la prise de poids moyenne  $\mu$  dans  $\mathcal{P}$ , contienne la vraie valeur de la moyenne  $\mu$  est de 95%
  - e. plus le niveau de confiance est élevé, plus l'amplitude de l'intervalle de confiance de la prise de poids moyenne est grande
  - f. plus la taille de l'échantillon est élevée, plus l'amplitude de l'intervalle de confiance à 95% de la prise de poids moyenne est grande
  - g. plus la variance observée de la prise de poids est élevée, plus l'amplitude de l'intervalle de confiance à 95% de la prise de poids moyenne est grande
  - h. l'intervalle de confiance à 95% de la prise de poids moyenne  $\mu$  dans  $\mathcal{P}$  est centré sur la vraie valeur de la moyenne  $\mu$
  - i. sur 100 intervalles de confiance à 95% de la prise de poids moyenne  $\mu$  dans  $\mathcal{P}$  calculés à partir de 100 échantillons de taille 38 issus de  $\mathcal{P}$ , environ 95 intervalles contiennent  $\mu$
- 3) Estimer par intervalle de confiance au niveau 95% la prise de poids moyenne dans la population étudiée :
- a. la prise de poids moyenne dans  $\mathcal{P}$  se situe entre 0,5 et 2,5 kg ; la méthode utilisée comporte un risque d'erreur de 5%
  - b. au risque 5%, la prise de poids moyenne dans  $\mathcal{P}$  varie entre 0,7 et 2,3 kg dans l'échantillon observé
  - c. au risque 5%, la prise de poids moyenne dans  $\mathcal{P}$  est comprise entre 0,5 et 2,5 kg
  - d. la prise de poids moyenne dans  $\mathcal{P}$  est comprise entre 0,5 et 2,5 kg
  - e. 95% des échantillons de la variable prise de poids issus de  $\mathcal{P}$  ont une prise de poids moyenne comprise entre 0,5 kg et 2,5 kg
  - f. la marge d'erreur dans l'estimation de la prise de poids moyenne dans  $\mathcal{P}$  au niveau de confiance 95% est de  $\pm 0,8$  kg
  - g. la marge d'erreur dans l'estimation de la prise de poids moyenne dans  $\mathcal{P}$  au niveau de confiance 95% est de  $\pm 1$  kg
  - h. l'intervalle  $[0,5 - 2,5]$  a 95% de chances de contenir la prise de poids moyenne observée
  - i. l'intervalle  $[0,5 - 2,5]$  a 5% de chances de ne pas contenir la prise de poids moyenne dans  $\mathcal{P}$
  - j. une confiance de 95% est accordée au fait que l'intervalle  $[0,5 - 2,5]$  contienne la prise de poids moyenne dans  $\mathcal{P}$
  - k. 95% des personnes ont une prise de poids moyenne comprise entre 0,5 kg et 2,5 kg
  - l. la prise de poids moyenne dans  $\mathcal{P}$  est comprise entre 0,5 kg et 2,5 kg pour 95% des personnes

### **Exercice 17**

Parmi 55 patients dépressifs âgés de 18 à 65 ans, 36 sont guéris après 16 semaines de psychothérapie. Parmi les réponses proposées indiquer celles qui sont exactes.

- 1) Estimer la proportion de guérison dans la population étudiée :
- a. la proportion de guérison dans  $\mathcal{P}$  est estimée par  $F_n = 65,5\%$
  - b. la proportion de guérison dans  $\mathcal{P}$  est égale à la proportion observée de guérison  $f = 65,5\%$
  - c. la proportion de guérison dans  $\mathcal{P}$  est égale à  $p = 0,655$
  - d. la proportion de guérison dans  $\mathcal{P}$  est estimée par la fréquence observée de guérison  $f = 0,655$
  - e. la proportion de guérison dans  $\mathcal{P}$  est estimée par  $\bar{x} = 0,655$
- 2) Vrai ou faux :
- a. pour obtenir une certaine confiance dans l'estimation de la proportion de guérison dans  $\mathcal{P}$ , on a recours à l'estimation par intervalle de confiance
  - b. la probabilité pour qu'un intervalle de confiance au risque 1% de la proportion de guérison  $p$  dans  $\mathcal{P}$ , contienne la vraie valeur de la proportion  $p$  est de 99%
  - c. dans l'estimation par intervalle de confiance de la proportion de guérison dans  $\mathcal{P}$ , on associe à l'intervalle de confiance une probabilité de contenir la vraie valeur du paramètre

- d. la probabilité pour que la vraie valeur de la proportion de guérison  $p$  dans  $\mathcal{P}$ , appartienne à l'intervalle de confiance à 99% de  $p$  est de 99%
  - e. l'intervalle de confiance au risque 1% de la proportion de guérison  $p$  dans  $\mathcal{P}$  est centré sur la fréquence de guérison observée
  - f. plus le risque est élevé, plus l'amplitude de l'intervalle de confiance de la proportion de guérison est faible
  - g. plus la taille de l'échantillon est élevée, plus l'amplitude de l'intervalle de confiance à 99% de la proportion de guérison est faible
  - h. sur 100 intervalles de confiance à 99% de la proportion de guérison dans  $\mathcal{P}$  calculés à partir de 100 échantillons de taille 55 issus de  $\mathcal{P}$ , un intervalle (environ) ne contient pas la vraie proportion de guérison dans  $\mathcal{P}$
- 3) Estimer par intervalle de confiance à 99% la proportion de guérison dans la population étudiée. Parmi les réponses suivantes trouver celles qui sont justes :
- a. la proportion de guérison dans  $\mathcal{P}$  est comprise entre 0,49 et 0,82
  - b. la proportion de guérison dans  $\mathcal{P}$  se situe entre 49% et 82% ; la méthode utilisée comporte un risque d'erreur de 1%
  - c. au risque 1%, la proportion de guérison dans  $\mathcal{P}$  est comprise entre 0,49 et 0,82 dans l'échantillon observé
  - d. au risque 1%, la proportion de guérison dans  $\mathcal{P}$  se situe entre 52,9% et 78,1%
  - e. 99% des échantillons de la variable guérison issus de  $\mathcal{P}$  ont une proportion de guérison comprise entre 49% et 82%
  - f. la précision dans l'estimation de la proportion de guérison dans  $\mathcal{P}$  au niveau 99% est de  $\pm 12,6\%$
  - g. la précision dans l'estimation de la proportion de guérison dans  $\mathcal{P}$  au risque 1% est de  $\pm 16,5\%$
  - h. une confiance de 99% est accordée au fait que l'intervalle  $[0,49 - 0,82]$  contienne la proportion de guérison dans  $\mathcal{P}$
  - i. l'intervalle  $[0,529 - 0,781]$  a 1% de chances de ne pas contenir la proportion de guérison dans  $\mathcal{P}$
  - j. les conditions d'applications ne sont pas vérifiées

### **Exercice 18**

Associer à chaque définition la bonne notation.

<b>Définition</b>	<b>Notation</b>
fréquence empirique •	♦ $\sigma$
moyenne observée •	♦ $p$
moyenne empirique •	♦ $s^*$
moyenne observée sur l'échantillon •	♦ $f$
taille de l'échantillon •	♦ $F_n$
fréquence observée sur l'échantillon •	♦ $N$
écart-type observé •	♦ $\bar{X}_n$
écart-type observé sans biais •	♦ $\bar{x}$
écart-type empirique •	♦ $\mu$
moyenne dans la population •	♦ $s$
écart-type dans la population •	♦ $X$
proportion dans la population •	♦ $S_n^*$
intervalle de confiance à 95% de la moyenne •	♦ $IC_{95\%}(p)$
intervalle de confiance au risque 5% de la proportion •	♦ $I_{95\%}(F_n)$
intervalle de variation à 95% de la fréquence observée •	♦ $IC_{95\%}(\mu)$
intervalle de fluctuation au risque 5% de la moyenne observée •	♦ $I_{95\%}(\bar{X}_n)$
proportion observée •	♦ $n$
écart-type empirique sans biais •	♦ $S_n$

