Dénombrement et échantillonnage

Diarra Fall (*MAPMO, Université d'Orléans*)

Centre Galois,
20/06/2014, Orléans.

(diarra.fall@univ-orleans.fr)

Motivations générales

Dans de nombreuses situations, on cherche à dénombrer le nombre d'individus, d'animaux, de végétaux ou d'objets se trouvant dans une zone géographique donnée.

- Décompte exact des individus souvent très difficile, voire impossible (grande étendue, population trop importante).
- Méthodes d'échantillonnage $\rightarrow$ approximation de la taille de la population en n'observant que :
  - certains individus
  - certaines zones géographiques.

Objectifs de cet atelier :

- se familiariser avec ces méthodes au travers de quelques exemples,
- présenter les résultats mathématiques sur lesquels elles reposent.

Motivations générales

Dans de nombreuses situations, on cherche à dénombrer le nombre d'individus, d'animaux, de végétaux ou d'objets se trouvant dans une zone géographique donnée.

- Décompte exact des individus souvent très difficile, voire impossible (grande étendue, population trop importante).
- Méthodes d'échantillonnage $\rightarrow$ approximation de la taille de la population en n'observant que :
  - certains individus
  - certaines zones géographiques.

Objectifs de cet atelier :

- se familiariser avec ces méthodes au travers de quelques exemples,
- présenter les résultats mathématiques sur lesquels elles reposent.

Motivations générales

Dans de nombreuses situations, on cherche à dénombrer le nombre d'individus, d'animaux, de végétaux ou d'objets se trouvant dans une zone géographique donnée.

- Décompte exact des individus souvent très difficile, voire impossible (grande étendue, population trop importante).
- Méthodes d'échantillonnage $\rightarrow$ approximation de la taille de la population en n'observant que :
  - certains individus
  - certaines zones géographiques.

Objectifs de cet atelier :

- se familiariser avec ces méthodes au travers de quelques exemples,
- présenter les résultats mathématiques sur lesquels elles reposent.

Motivations générales

Dans de nombreuses situations, on cherche à dénombrer le nombre d'individus, d'animaux, de végétaux ou d'objets se trouvant dans une zone géographique donnée.

- Décompte exact des individus souvent très difficile, voire impossible (grande étendue, population trop importante).
- Méthodes d'échantillonnage $\rightarrow$ approximation de la taille de la population en n'observant que :
  - certains individus
  - certaines zones géographiques.

Objectifs de cet atelier :

- se familiariser avec ces méthodes au travers de quelques exemples,
- présenter les résultats mathématiques sur lesquels elles reposent.

Plan de l'atelier

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Observations répétées de phénomènes aléatoires

Chaque groupe de participants lance 10 fois 5 dés et note les résultats.
Calculer le score moyen et le nombre de 5 ou 6 observés.

- Comparer les résultats obtenus par chaque groupe. D'où proviennent ces différences ?
- Lorsqu'on lance un dé :
  - quelle est la probabilité que l'on tombe sur une certaine face ?
  - Quelle est la probabilité de tomber sur 5 ou 6 ?
  - Quel est le score moyen auquel on peut s'attendre ?
- Quelle est la proportion globale de 5 ou 6 observés ?
- Quelle est la moyenne globale des scores obtenus ?
- Que remarquez-vous ?

Loi des grands nombres

Ce phénomène, appelé *loi des grands nombres*, est à la base de nombreuses méthodes statistiques.

Théorème

Soient n observations aléatoires et indépendantes $(X_1, \ldots, X_n)$ *d'un phénomène* X *de valeur moyenne m.*
Sous des hypothèses générales, on montre que

$$\frac{(X_1 + \cdots + X_n)}{n} \underset{n \to +\infty}{\to} m. \qquad (\mathbb{P}, p.s.)$$

En d'autres termes, si n est assez grand :

- la valeur moyenne de X peut être approchée par la moyenne de n observations indépendantes.
- La probabilité d'un événement peut être approchée par la proportion de fois où il a été observé au cours de n observations indépendantes.

Loi des grands nombres

Ce phénomène, appelé *loi des grands nombres*, est à la base de nombreuses méthodes statistiques.

Théorème

Soient n observations aléatoires et indépendantes $(X_1, \ldots, X_n)$ d'un phénomène X de valeur moyenne m.
Sous des hypothèses générales, on montre que

$$\frac{(X_1 + \cdots + X_n)}{n} \underset{n \to +\infty}{\to} m. \qquad (\mathbb{P}, p.s.)$$

En d'autres termes, si n est assez grand :

- la valeur moyenne de X peut être approchée par la moyenne de n observations indépendantes.
- La probabilité d'un événement peut être approchée par la proportion de fois où il a été observé au cours de n observations indépendantes.

Loi des grands nombres

Ce phénomène, appelé *loi des grands nombres*, est à la base de nombreuses méthodes statistiques.

Théorème

Soient n observations aléatoires et indépendantes $(X_1, \dots, X_n)$ d'un phénomène X de valeur moyenne m.
Sous des hypothèses générales, on montre que

$$\frac{(X_1 + \cdots + X_n)}{n} \underset{n \to +\infty}{\to} m. \qquad (\mathbb{P}, p.s.)$$

En d'autres termes, si n est assez grand :

- la valeur moyenne de X peut être approchée par la moyenne de n observations indépendantes.
- La probabilité d'un événement peut être approchée par la proportion de fois où il a été observé au cours de n observations indépendantes.

Limite centrale et intervalles de confiance

Notation : $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ (moyenne empirique).

Construites à partir d'observations de phénomènes aléatoires, nos approximations présentent une certaine *variabilité.*

Variance : $\sigma^2_{n-1} = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right)$.

Cette variabilité diminue lorsque le nombre d'observations n croît (cf illustrations).

On définit $T = (\overline{X} - m)/\sqrt{\frac{\sigma^2_{n-1}}{n}}$.

- Pour n assez grand, les valeurs de T se répartissent à peu près suivant une loi $\mathcal{N}(0,1)$ (*théo. de la limite centrale*).
- Intervalle de confiance pour m (avec un risque de 5%) :
- $P(\overline{X} - \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96 \leq m \leq \overline{X} + \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96) \approx 0.95$.

→ illustrations exo1A et exo1B

Limite centrale et intervalles de confiance

Notation : $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ (moyenne empirique).

Construites à partir d'observations de phénomènes aléatoires, nos approximations présentent une certaine *variabilité.*

Variance : $\sigma^2_{n-1} = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right)$.

Cette variabilité diminue lorsque le nombre d'observations n croît (cf illustrations).

On définit $T = (\overline{X} - m)/\sqrt{\frac{\sigma^2_{n-1}}{n}}$.

- Pour n assez grand, les valeurs de T se répartissent à peu près suivant une loi $\mathcal{N}(0,1)$ (*théo. de la limite centrale*).
- Intervalle de confiance pour m (avec un risque de 5%) :
- $P(\overline{X} - \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96 \leq m \leq \overline{X} + \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96) \approx 0.95$.

$\rightarrow$ illustrations exo1A et exo1B

Limite centrale et intervalles de confiance

Notation : $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ (moyenne empirique).

Construites à partir d'observations de phénomènes aléatoires, nos approximations présentent une certaine *variabilité.*

Variance : $\sigma_{n-1}^2 = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right)$.

Cette variabilité diminue lorsque le nombre d'observations n croît (cf illustrations).

On définit $T = (\overline{X} - m)/\sqrt{\frac{\sigma_{n-1}^2}{n}}$.

- Pour n assez grand, les valeurs de T se répartissent à peu près suivant une loi $\mathcal{N}(0,1)$ (*théo. de la limite centrale*).
- Intervalle de confiance pour m (avec un risque de 5%) :
- $P(\overline{X} - \sqrt{\frac{\sigma_{n-1}^2}{n}} \times 1.96 \leq m \leq \overline{X} + \sqrt{\frac{\sigma_{n-1}^2}{n}} \times 1.96) \approx 0.95$.

→ illustrations exo1A et exo1B

Limite centrale et intervalles de confiance

Notation : $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ (moyenne empirique).

Construites à partir d'observations de phénomènes aléatoires, nos approximations présentent une certaine *variabilité.*

Variance : $\sigma^2_{n-1} = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right)$.

Cette variabilité diminue lorsque le nombre d'observations n croît (cf illustrations).

On définit $T = (\overline{X} - m)/\sqrt{\frac{\sigma^2_{n-1}}{n}}$.

- Pour n assez grand, les valeurs de T se répartissent à peu près suivant une loi $\mathcal{N}(0,1)$ (*théo. de la limite centrale*).
- Intervalle de confiance pour m (avec un risque de 5%) :
- $P(\overline{X} - \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96 \leq m \leq \overline{X} + \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96) \approx 0.95$.

$\rightarrow$ illustrations exo1A et exo1B

Limite centrale et intervalles de confiance

Notation : $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ (moyenne empirique).

Construites à partir d'observations de phénomènes aléatoires, nos approximations présentent une certaine *variabilité.*

Variance : $\sigma^2_{n-1} = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right)$.

Cette variabilité diminue lorsque le nombre d'observations n croît (cf illustrations).

On définit $T = (\overline{X} - m)/\sqrt{\frac{\sigma^2_{n-1}}{n}}$.

- Pour n assez grand, les valeurs de T se répartissent à peu près suivant une loi $\mathcal{N}(0, 1)$ (*théo. de la limite centrale*).
- Intervalle de confiance pour m (avec un risque de 5%) :
- $P(\overline{X} - \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96 \leq m \leq \overline{X} + \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96) \approx 0.95$.

$\rightarrow$ illustrations exo1A et exo1B

Limite centrale et intervalles de confiance

Notation : $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ (moyenne empirique).

Construites à partir d'observations de phénomènes aléatoires, nos approximations présentent une certaine *variabilité.*

Variance : $\sigma^2_{n-1} = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right)$.

Cette variabilité diminue lorsque le nombre d'observations n croît (cf illustrations).

On définit $T = (\overline{X} - m)/\sqrt{\frac{\sigma^2_{n-1}}{n}}$.

- Pour n assez grand, les valeurs de T se répartissent à peu près suivant une loi $\mathcal{N}(0,1)$ (*théo. de la limite centrale*).
- Intervalle de confiance pour m (avec un risque de 5%) :
- $P(\overline{X} - \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96 \leq m \leq \overline{X} + \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96) \approx 0.95$.

$\rightarrow$ illustrations exo1A et exo1B

Limite centrale et intervalles de confiance

Notation : $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ (moyenne empirique).

Construites à partir d'observations de phénomènes aléatoires, nos approximations présentent une certaine *variabilité.*

Variance : $\sigma^2_{n-1} = \frac{1}{n-1}\left((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\right)$.

Cette variabilité diminue lorsque le nombre d'observations n croît (cf illustrations).

On définit $T = (\overline{X} - m)/\sqrt{\frac{\sigma^2_{n-1}}{n}}$.

- Pour n assez grand, les valeurs de T se répartissent à peu près suivant une loi $\mathcal{N}(0, 1)$ (*théo. de la limite centrale*).
- Intervalle de confiance pour m (avec un risque de 5%) :
- $P(\overline{X} - \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96 \leq m \leq \overline{X} + \sqrt{\frac{\sigma^2_{n-1}}{n}} \times 1.96) \approx 0.95$.

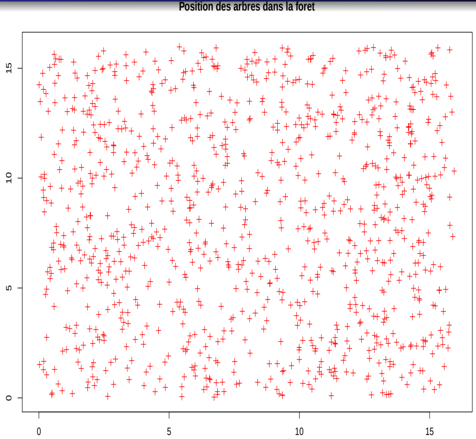
$\rightarrow$ illustrations exo1A et exo1B

Echantillonnage aléatoire

Problème concret

On cherche à évaluer le nombre N d'érables se trouvant dans une forêt s'étendant sur une grande superficie.

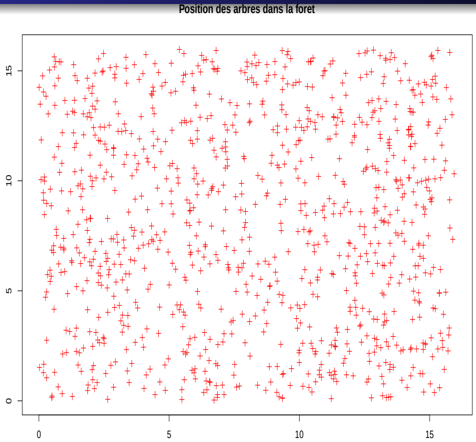
Comment faire ?



Pensez-vous que cela soit envisagable de dénombrer

- tous les érables de cette forêt ?
- les érables se trouvant dans différentes zones d'assez petite taille ?

Comment faire ?



Pensez-vous que cela soit envisagable de dénombrer

- tous les érables de cette forêt ?
- les érables se trouvant dans différentes zones d'assez petite taille ?

Comment faire ?

On suppose que l'on peut découper la forêt en 256 zones, soit 16×16, plus petites et de même forme.

- Quel est, en fonction de N, le nombre moyen m d'érables par zone ?
- Comment pensez-vous que l'on puisse donner une approximation de m en observant seulement un petit nombre n de ces zones ?
- Comment choisir les zones à étudier ? Combien faut-il en prendre ?

Choix déterministe peut présenter des défauts (zones sélectionnées peuvent être non représentatives).
$\Rightarrow$ échantillonnage aléatoire : choisir aléatoirement (et avec remise) les zones à étudier.

Comment faire ?

On suppose que l'on peut découper la forêt en 256 zones, soit 16×16, plus petites et de même forme.

- Quel est, en fonction de N, le nombre moyen m d'érables par zone ?
- Comment pensez-vous que l'on puisse donner une approximation de m en observant seulement un petit nombre n de ces zones ?
- Comment choisir les zones à étudier ? Combien faut-il en prendre ?

Choix déterministe peut présenter des défauts (zones sélectionnées peuvent être non représentatives).
$\Rightarrow$ échantillonnage aléatoire : choisir aléatoirement (et avec remise) les zones à étudier.

Comment faire ?

On suppose que l'on peut découper la forêt en 256 zones, soit 16×16, plus petites et de même forme.

- Quel est, en fonction de N, le nombre moyen m d'érables par zone ?
- Comment pensez-vous que l'on puisse donner une approximation de m en observant seulement un petit nombre n de ces zones ?
- Comment choisir les zones à étudier ? Combien faut-il en prendre ?

Choix déterministe peut présenter des défauts (zones sélectionnées peuvent être non représentatives).
$\Rightarrow$ échantillonnage aléatoire : choisir aléatoirement (et avec remise) les zones à étudier.

Comment faire ?

On suppose que l'on peut découper la forêt en 256 zones, soit 16×16, plus petites et de même forme.

- Quel est, en fonction de N, le nombre moyen m d'érables par zone ?
- Comment pensez-vous que l'on puisse donner une approximation de m en observant seulement un petit nombre n de ces zones ?
- Comment choisir les zones à étudier ? Combien faut-il en prendre ?

Choix déterministe peut présenter des défauts (zones sélectionnées peuvent être non représentatives).

$\Rightarrow$ échantillonnage aléatoire : choisir aléatoirement (et avec remise) les zones à étudier.

Comment faire ?

On suppose que l'on peut découper la forêt en 256 zones, soit 16×16, plus petites et de même forme.

- Quel est, en fonction de N, le nombre moyen m d'érables par zone ?
- Comment pensez-vous que l'on puisse donner une approximation de m en observant seulement un petit nombre n de ces zones ?
- Comment choisir les zones à étudier ? Combien faut-il en prendre ?

Choix déterministe peut présenter des défauts (zones sélectionnées peuvent être non représentatives).
$\Rightarrow$ échantillonnage aléatoire : choisir aléatoirement (et avec remise) les zones à étudier.

Mise en pratique :

Exemples de choix aléatoires ($n = 40$) :

x	6	9	3	16	10	10	6	9	9	2	11	5	4	3	9	11
y	8	13	13	11	14	15	10	7	4	7	6	1	10	4	1	11
x	15	9	2	2	12	14	5	6	15	14	11	15	14	5	10	11
y	9	16	10	4	13	14	9	12	13	9	1	12	4	16	6	16
x	8	12	8	6	13	5	6	3								
y	10	8	15	1	1	7	10	4								

- Compter le nombre d'érables se trouvant dans les zones sélectionnées par cet échantillonnage aléatoire
  $\rightarrow$ illustration (exo2).
- Donner une approximation de m à partir des données recueillies.
- En déduire une approximation de N.

On peut aller plus loin :

- Calculer la variance de nos observations.
- Donner la valeur de l'intervalle de confiance associé à m.
- En déduire un intervalle de confiance pour N.

Mise en pratique :

Exemples de choix aléatoires ($n = 40$) :

x	6	9	3	16	10	10	6	9	9	2	11	5	4	3	9	11
y	8	13	13	11	14	15	10	7	4	7	6	1	10	4	1	11
x	15	9	2	2	12	14	5	6	15	14	11	15	14	5	10	11
y	9	16	10	4	13	14	9	12	13	9	1	12	4	16	6	16
x	8	12	8	6	13	5	6	3								
y	10	8	15	1	1	7	10	4								

- Compter le nombre d'érables se trouvant dans les zones sélectionnées par cet échantillonnage aléatoire $\rightarrow$ illustration (exo2).
- Donner une approximation de m à partir des données recueillies.
- En déduire une approximation de N.

On peut aller plus loin :

- Calculer la variance de nos observations.
- Donner la valeur de l'intervalle de confiance associé à m.
- En déduire un intervalle de confiance pour N.

Mise en pratique :

Exemples de choix aléatoires ($n = 40$) :

x	6	9	3	16	10	10	6	9	9	2	11	5	4	3	9	11
y	8	13	13	11	14	15	10	7	4	7	6	1	10	4	1	11
x	15	9	2	2	12	14	5	6	15	14	11	15	14	5	10	11
y	9	16	10	4	13	14	9	12	13	9	1	12	4	16	6	16
x	8	12	8	6	13	5	6	3								
y	10	8	15	1	1	7	10	4								

- Compter le nombre d'érables se trouvant dans les zones sélectionnées par cet échantillonnage aléatoire
$\rightarrow$ illustration (exo2).
- Donner une approximation de m à partir des données recueillies.
- En déduire une approximation de N.

On peut aller plus loin :

- Calculer la variance de nos observations.
- Donner la valeur de l'intervalle de confiance associé à m.
- En déduire un intervalle de confiance pour N.

Mise en pratique :

Exemples de choix aléatoires ($n = 40$) :

x	6	9	3	16	10	10	6	9	9	2	11	5	4	3	9	11
y	8	13	13	11	14	15	10	7	4	7	6	1	10	4	1	11
x	15	9	2	2	12	14	5	6	15	14	11	15	14	5	10	11
y	9	16	10	4	13	14	9	12	13	9	1	12	4	16	6	16
x	8	12	8	6	13	5	6	3								
y	10	8	15	1	1	7	10	4								

- Compter le nombre d'érables se trouvant dans les zones sélectionnées par cet échantillonnage aléatoire
$\rightarrow$ illustration (exo2).
- Donner une approximation de m à partir des données recueillies.
- En déduire une approximation de N.

On peut aller plus loin :

- Calculer la variance de nos observations.
- Donner la valeur de l'intervalle de confiance associé à m.
- En déduire un intervalle de confiance pour N.

Mise en pratique :

Exemples de choix aléatoires ($n = 40$) :

x	6	9	3	16	10	10	6	9	9	2	11	5	4	3	9	11
y	8	13	13	11	14	15	10	7	4	7	6	1	10	4	1	11

x	15	9	2	2	12	14	5	6	15	14	11	15	14	5	10	11
y	9	16	10	4	13	14	9	12	13	9	1	12	4	16	6	16

x	8	12	8	6	13	5	6	3								
y	10	8	15	1	1	7	10	4								

- Compter le nombre d'érables se trouvant dans les zones sélectionnées par cet échantillonnage aléatoire
  $\rightarrow$ illustration (exo2).
- Donner une approximation de m à partir des données recueillies.
- En déduire une approximation de N.

On peut aller plus loin :

- Calculer la variance de nos observations.
- Donner la valeur de l'intervalle de confiance associé à m.
- En déduire un intervalle de confiance pour N.

Mise en pratique :

Exemples de choix aléatoires ($n = 40$) :

x	6	9	3	16	10	10	6	9	9	2	11	5	4	3	9	11
y	8	13	13	11	14	15	10	7	4	7	6	1	10	4	1	11
x	15	9	2	2	12	14	5	6	15	14	11	15	14	5	10	11
y	9	16	10	4	13	14	9	12	13	9	1	12	4	16	6	16
x	8	12	8	6	13	5	6	3								
y	10	8	15	1	1	7	10	4								

- Compter le nombre d'érables se trouvant dans les zones sélectionnées par cet échantillonnage aléatoire
$\rightarrow$ illustration (exo2).
- Donner une approximation de m à partir des données recueillies.
- En déduire une approximation de N.

On peut aller plus loin :

- Calculer la variance de nos observations.
- Donner la valeur de l'intervalle de confiance associé à m.
- En déduire un intervalle de confiance pour N.

Méthode de capture re-capture

Un autre problème concret

On souhaite donner un ordre de grandeur du nombre de poissons se trouvant dans une grande étendue d'eau.
Décompte exact impossible même dans zones petites (profondeur, eau trouble etc.)
Capture des poissons et relâche immédiate de toute prise $\Rightarrow$ on ne peut observer qu'un poisson à la fois.

Expérimentation et mise en pratique :
-Vous disposez d'un récipient contenant un nombre N inconnu d'objets identiques.
-Vous ne pouvez sortir du récipient qu'un objet à la fois et devez mélanger avant chaque tirage.
-Vous avez à votre disposition un feutre indélébile.

- Comment pensez-vous que l'on puisse avoir une idée du nombre total d'objets présents dans la boîte ?
- A quoi peut servir le feutre ?

Un autre problème concret

On souhaite donner un ordre de grandeur du nombre de poissons se trouvant dans une grande étendue d'eau.
Décompte exact impossible même dans zones petites (profondeur, eau trouble etc.)
Capture des poissons et relâche immédiate de toute prise $\Rightarrow$ on ne peut observer qu'un poisson à la fois.

Expérimentation et mise en pratique :
-Vous disposez d'un récipient contenant un nombre N inconnu d'objets identiques.
-Vous ne pouvez sortir du récipient qu'un objet à la fois et devez mélanger avant chaque tirage.
-Vous avez à votre disposition un feutre indélébile.

- Comment pensez-vous que l'on puisse avoir une idée du nombre total d'objets présents dans la boîte ?
- A quoi peut servir le feutre ?

Un autre problème concret

On souhaite donner un ordre de grandeur du nombre de poissons se trouvant dans une grande étendue d'eau.
Décompte exact impossible même dans zones petites (profondeur, eau trouble etc.)
Capture des poissons et relâche immédiate de toute prise $\Rightarrow$ on ne peut observer qu'un poisson à la fois.

Expérimentation et mise en pratique :
-Vous disposez d'un récipient contenant un nombre N inconnu d'objets identiques.
-Vous ne pouvez sortir du récipient qu'un objet à la fois et devez mélanger avant chaque tirage.
-Vous avez à votre disposition un feutre indélébile.

- Comment pensez-vous que l'on puisse avoir une idée du nombre total d'objets présents dans la boîte ?
- A quoi peut servir le feutre ?

Un autre problème concret

On souhaite donner un ordre de grandeur du nombre de poissons se trouvant dans une grande étendue d'eau.
Décompte exact impossible même dans zones petites (profondeur, eau trouble etc.)
Capture des poissons et relâche immédiate de toute prise $\Rightarrow$ on ne peut observer qu'un poisson à la fois.

Expérimentation et mise en pratique :
-Vous disposez d'un récipient contenant un nombre N inconnu d'objets identiques.
-Vous ne pouvez sortir du récipient qu'un objet à la fois et devez mélanger avant chaque tirage.
-Vous avez à votre disposition un feutre indélébile.

- Comment pensez-vous que l'on puisse avoir une idée du nombre total d'objets présents dans la boîte ?
- A quoi peut servir le feutre ?

Comment faire ?

Si un nombre connu N_M d'objets portent un signe distinctif.

- Quelle serait la probabilité P de choisir au hasard un objet portant un signe distinctif, en fonction de N ?
- Supposons que l'on fasse n tirages avec remise dans le récipient, en notant si l'objet porte une marque ou non. Comment peut-on donner une approximation p_n de P ?
- Déduire de p_n une approximation de N.

Revenons à notre problème concret (objets identiques).

- Proposer une méthode permettant de se ramener à l'étude d'un récipient contenant 20 objets marqués. [CAPTURE]
- Mélanger les objets pour modéliser la répartition uniforme des individus marqués dans la zone d'étude. [DELAI D'ATTENTE]
- Effectuer ensuite $n = 45$ tirages avec remise. En déduire une approximation de N. [RECAPTURE] $\rightarrow$ illustration bouteille de lait et exo3

Comment faire ?

Si un nombre connu N_M d'objets portent un signe distinctif.

- Quelle serait la probabilité P de choisir au hasard un objet portant un signe distinctif, en fonction de N ?
- Supposons que l'on fasse n tirages avec remise dans le récipient, en notant si l'objet porte une marque ou non. Comment peut-on donner une approximation p_n de P ?
- Déduire de p_n une approximation de N.

Revenons à notre problème concret (objets identiques).

- Proposer une méthode permettant de se ramener à l'étude d'un récipient contenant 20 objets marqués. [CAPTURE]
- Mélanger les objets pour modéliser la répartition uniforme des individus marqués dans la zone d'étude. [DELAI D'ATTENTE]
- Effectuer ensuite $n = 45$ tirages avec remise. En déduire une approximation de N. [RECAPTURE] $\rightarrow$ illustration bouteille de lait et exo3

Comment faire ?

Si un nombre connu N_M d'objets portent un signe distinctif.

- Quelle serait la probabilité P de choisir au hasard un objet portant un signe distinctif, en fonction de N ?
- Supposons que l'on fasse n tirages avec remise dans le récipient, en notant si l'objet porte une marque ou non. Comment peut-on donner une approximation p_n de P ?
- Déduire de p_n une approximation de N.

Revenons à notre problème concret (objets identiques).

- Proposer une méthode permettant de se ramener à l'étude d'un récipient contenant 20 objets marqués. [CAPTURE]
- Mélanger les objets pour modéliser la répartition uniforme des individus marqués dans la zone d'étude. [DELAI D'ATTENTE]
- Effectuer ensuite $n = 45$ tirages avec remise. En déduire une approximation de N. [RECAPTURE] $\rightarrow$ illustration bouteille de lait et exo3

Comment faire ?

Si un nombre connu N_M d'objets portent un signe distinctif.

- Quelle serait la probabilité P de choisir au hasard un objet portant un signe distinctif, en fonction de N ?
- Supposons que l'on fasse n tirages avec remise dans le récipient, en notant si l'objet porte une marque ou non. Comment peut-on donner une approximation p_n de P ?
- Déduire de p_n une approximation de N.

Revenons à notre problème concret (objets identiques).

- Proposer une méthode permettant de se ramener à l'étude d'un récipient contenant 20 objets marqués. [CAPTURE]
- Mélanger les objets pour modéliser la répartition uniforme des individus marqués dans la zone d'étude. [DELAI D'ATTENTE]
- Effectuer ensuite $n = 45$ tirages avec remise. En déduire une approximation de N. [RECAPTURE] $\rightarrow$ illustration bouteille de lait et exo3

Comment faire ?

Si un nombre connu N_M d'objets portent un signe distinctif.

- Quelle serait la probabilité P de choisir au hasard un objet portant un signe distinctif, en fonction de N ?
- Supposons que l'on fasse n tirages avec remise dans le récipient, en notant si l'objet porte une marque ou non. Comment peut-on donner une approximation p_n de P ?
- Déduire de p_n une approximation de N.

Revenons à notre problème concret (objets identiques).

- Proposer une méthode permettant de se ramener à l'étude d'un récipient contenant 20 objets marqués. [CAPTURE]
- Mélanger les objets pour modéliser la répartition uniforme des individus marqués dans la zone d'étude. [DELAI D'ATTENTE]
- Effectuer ensuite $n = 45$ tirages avec remise. En déduire une approximation de N. [RECAPTURE] $\rightarrow$ illustration bouteille de lait et exo3

Comment faire ?

Si un nombre connu N_M d'objets portent un signe distinctif.

- Quelle serait la probabilité P de choisir au hasard un objet portant un signe distinctif, en fonction de N ?
- Supposons que l'on fasse n tirages avec remise dans le récipient, en notant si l'objet porte une marque ou non. Comment peut-on donner une approximation p_n de P ?
- Déduire de p_n une approximation de N.

Revenons à notre problème concret (objets identiques).

- Proposer une méthode permettant de se ramener à l'étude d'un récipient contenant 20 objets marqués. [CAPTURE]
- Mélanger les objets pour modéliser la répartition uniforme des individus marqués dans la zone d'étude.
  [DELAI D'ATTENTE]
- Effectuer ensuite $n = 45$ tirages avec remise. En déduire une approximation de N. [RECAPTURE]
  $\rightarrow$ illustration bouteille de lait et exo3

Pour aller un peu plus loin

Lorsque le nombre d'observations et la proportion d'objets marqués sont assez grands, on peut également utiliser le théorème de la limite centrale pour montrer que :

$$P(p_n - \sqrt{\frac{p_n(1-p_n)}{n}} \times 1,96 \leq P \leq p_n + \sqrt{\frac{p_n(1-p_n)}{n}} \times 1,96) \approx 0,95.$$

Supposons que sur 1000 observations on ait une proportion de 0.199 poissons marqués.

- Donner un intervalle de confiance pour P.
- En déduire un intervalle de confiance pour N.

Pour aller un peu plus loin

Lorsque le nombre d'observations et la proportion d'objets marqués sont assez grands, on peut également utiliser le théorème de la limite centrale pour montrer que :

$$P(p_n - \sqrt{\frac{p_n(1-p_n)}{n}} \times 1,96 \leq P \leq p_n + \sqrt{\frac{p_n(1-p_n)}{n}} \times 1,96) \approx 0,95.$$

Supposons que sur 1000 observations on ait une proportion de 0.199 poissons marqués.

- Donner un intervalle de confiance pour P.
- En déduire un intervalle de confiance pour N.

Méthode du maximum de vraisemblance

Un autre exemple concret

Idée principale : choisir la valeur de N avec laquelle on a le plus de chance d'obtenir les observations qui ont été faites.

On cherche à donner une approximation du nombre de rhinocéros se trouvant dans une zone géographique assez vaste.

On suppose que les rhinocéros ont

- une probabilité de $\frac{1}{3}$ d'aller se désaltérer à l'un des rares points d'eau en fin d'après-midi.
- des comportements indépendants les uns des autres.

On propose de se rendre à plusieurs reprises autour de ce point d'eau et de compter le nombre de rhinocéros que l'on observe. On fait les observations suivantes :

Nombre de rhinocéros	0	1	2	2	1

Un autre exemple concret

Idée principale : choisir la valeur de N avec laquelle on a le plus de chance d'obtenir les observations qui ont été faites.

On cherche à donner une approximation du nombre de rhinocéros se trouvant dans une zone géographique assez vaste.

On suppose que les rhinocéros ont

- une probabilité de $\frac{1}{3}$ d'aller se désaltérer à l'un des rares points d'eau en fin d'après-midi.
- des comportements indépendants les uns des autres.

On propose de se rendre à plusieurs reprises autour de ce point d'eau et de compter le nombre de rhinocéros que l'on observe. On fait les observations suivantes :

Nombre de rhinocéros	0	1	2	2	1

Un autre exemple concret

Idée principale : choisir la valeur de N avec laquelle on a le plus de chance d'obtenir les observations qui ont été faites.

On cherche à donner une approximation du nombre de rhinocéros se trouvant dans une zone géographique assez vaste.

On suppose que les rhinocéros ont

- une probabilité de $\frac{1}{3}$ d'aller se désaltérer à l'un des rares points d'eau en fin d'après-midi.
- des comportements indépendants les uns des autres.

On propose de se rendre à plusieurs reprises autour de ce point d'eau et de compter le nombre de rhinocéros que l'on observe. On fait les observations suivantes :

Nombre de rhinocéros	0	1	2	2	1

La méthode du maximum de vraisemblance

- On note b_i le nombre de rhinocéros observés le jour i.
- On peut montrer que le nombre de rhinocéros venant s'abreuver en fin d'après-midi un jour donné peut être modélisé par une variable aléatoire B de loi $Bin(N, \frac{1}{3})$.
- On peut donc donner une expression explicite de la probabilité d'avoir observé $0, 1, 2, 2, 1$ rhinocéros en fonction de N :

$$P_N(Observations : 0, 1, 2, 2, 1) = \left(\frac{2}{3}\right)^{5N} \frac{N^4(N-1)^2}{2^8}.$$

- Est-ce que N peut valoir 0 ou 1 ?

La méthode du maximum de vraisemblance

- On note b_i le nombre de rhinocéros observés le jour i.
- On peut montrer que le nombre de rhinocéros venant s'abreuver en fin d'après-midi un jour donné peut être modélisé par une variable aléatoire B de loi $Bin(N, \frac{1}{3})$.
- On peut donc donner une expression explicite de la probabilité d'avoir observé $0, 1, 2, 2, 1$ rhinocéros en fonction de N :

$$P_N(Observations : 0, 1, 2, 2, 1) = \left(\frac{2}{3}\right)^{5N} \frac{N^4(N-1)^2}{2^8}.$$

- Est-ce que N peut valoir 0 ou 1 ?

La méthode du maximum de vraisemblance

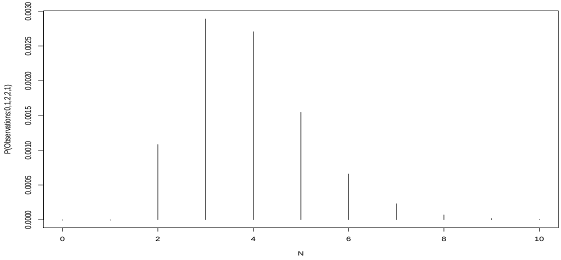
- On note b_i le nombre de rhinocéros observés le jour i.
- On peut montrer que le nombre de rhinocéros venant s'abreuver en fin d'après-midi un jour donné peut être modélisé par une variable aléatoire B de loi $Bin(N, \frac{1}{3})$.
- On peut donc donner une expression explicite de la probabilité d'avoir observé $0, 1, 2, 2, 1$ rhinocéros en fonction de N :

$$P_N(Observations : 0, 1, 2, 2, 1) = \left(\frac{2}{3}\right)^{5N} \frac{N^4(N-1)^2}{2^8}.$$

- Est-ce que N peut valoir 0 ou 1 ?

La méthode du maximum de vraisemblance

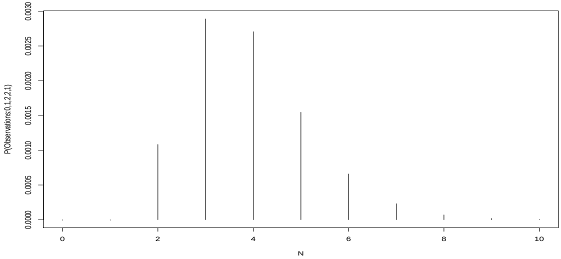
- A partir du graphique ci-dessous, pour quelle valeur de N la probabilité est la plus forte que se produise ce que l'on a observé ?



- En déduire la valeur de N la plus vraisemblable.
- $\rightarrow$ illustration exo4

La méthode du maximum de vraisemblance

- A partir du graphique ci-dessous, pour quelle valeur de N la probabilité est la plus forte que se produise ce que l'on a observé ?



- En déduire la valeur de N la plus vraisemblable.
- $\rightarrow$ illustration exo4

Un autre exemple :

Imaginons par exemple qu'un nouveau participant arrive en fin de séance et qu'il ne sache pas combien de dés vous avez lancé au début.

S'il dispose uniquement du nombre de 5 ou 6 observés pour chaque lancé de 5 dés, il peut utiliser la méthode du maximum de vraisemblance pour déterminer à partir des observations la valeur la plus vraisemblable de dés que vous avez jetés.

En effet, ici aussi le nombre de 5 ou 6 obtenus à chacun de vos lancés peut être modélisé par une variable de loi $Bin(N, \frac{1}{3})$, où N représente le nombre inconnu de dés que vous avez lancés.

$\rightarrow$ illustration : résultats dans groupes (et exo4B)

Conclusion

Conclusion et commentaires

- Méthodes statistiques permettent de donner une approximation du nombre d'individus mais aussi de donner un ordre de grandeur de la précision de notre estimation.
- On observe une certaine variabilité de nos résultats, qui diminue cependant lorsque le nombre d'observations augmente.
- Il est important de prendre en compte la qualité des données (taille de l'échantillon, représentativité) lorsque l'on cherche à interpréter les résultats donnés par une méthode d'estimation (par exemple un sondage).
- Des méthodes plus fines et précises sont généralement utilisées.

Conclusion et commentaires

- Méthodes statistiques permettent de donner une approximation du nombre d'individus mais aussi de donner un ordre de grandeur de la précision de notre estimation.
- On observe une certaine variabilité de nos résultats, qui diminue cependant lorsque le nombre d'observations augmente.
- Il est important de prendre en compte la qualité des données (taille de l'échantillon, représentativité) lorsque l'on cherche à interpréter les résultats donnés par une méthode d'estimation (par exemple un sondage).
- Des méthodes plus fines et précises sont généralement utilisées.

Conclusion et commentaires

- Méthodes statistiques permettent de donner une approximation du nombre d'individus mais aussi de donner un ordre de grandeur de la précision de notre estimation.
- On observe une certaine variabilité de nos résultats, qui diminue cependant lorsque le nombre d'observations augmente.
- Il est important de prendre en compte la qualité des données (taille de l'échantillon, représentativité) lorsque l'on cherche à interpréter les résultats donnés par une méthode d'estimation (par exemple un sondage).
- Des méthodes plus fines et précises sont généralement utilisées.

Conclusion et commentaires

- Méthodes statistiques permettent de donner une approximation du nombre d'individus mais aussi de donner un ordre de grandeur de la précision de notre estimation.
- On observe une certaine variabilité de nos résultats, qui diminue cependant lorsque le nombre d'observations augmente.
- Il est important de prendre en compte la qualité des données (taille de l'échantillon, représentativité) lorsque l'on cherche à interpréter les résultats donnés par une méthode d'estimation (par exemple un sondage).
- Des méthodes plus fines et précises sont généralement utilisées.

Tests médicaux

Les données

- On souhaite étudier l'efficacité d'un test médical permettant de dépister une maladie présente chez 0.5% de la population.
- Le fabricant de ce test indique qu'il est positif pour 99% des malades et négatif pour 95% des personnes saines.

- On dispose de données (simulées) concernant un échantillon de 1.000.000 d'individus.

- Pour chacun de ces individus on observe une variable "Santé" donnant l'état de santé de l'individu et une variable "Test" donnant le résultat du test de dépistage.

Les données

- Nombre de malades=4.971 ; Nombre de saines=995.029
  Nombre de tests += 54.549 ; Nombre de tests -= 945.451
  Nombre de de tests + chez malades=4.920 ;
  Nombre de tests - chez malades=51
  Nombre de tests + chez saines=49.629 ;
  Nombre de tests - chez saines=945.400
- Quelle est la proportion de malades ?
  nombre de malades/1.000.000=0,004971.
- Quelle est la proportion de tests positifs ?
  nombre de tests +/1.000.000=0,054549
- Proportion de tests positifs chez les malades ?
  nb. tests + chez malades/ nb. malades= 0,98974049
- Proportion de tests positifs chez les saines ?
  nb. tests + chez saines/ nb. saines=0,04987694

Les données

- Nombre de malades=4.971 ; Nombre de saines=995.029
  Nombre de tests += 54.549 ; Nombre de tests -= 945.451
  Nombre de de tests + chez malades=4.920 ;
  Nombre de tests - chez malades=51
  Nombre de tests + chez saines=49.629 ;
  Nombre de tests - chez saines=945.400
- Quelle est la proportion de malades ?
  nombre de malades/1.000.000=0,004971.
- Quelle est la proportion de tests positifs ?
  nombre de tests +/1.000.000=0,054549
- Proportion de tests positifs chez les malades ?
  nb. tests + chez malades/ nb. malades= 0,98974049
- Proportion de tests positifs chez les saines ?
  nb. tests + chez saines/ nb. saines=0,04987694

Les données

- Nombre de malades=4.971 ; Nombre de saines=995.029
  Nombre de tests += 54.549 ; Nombre de tests -= 945.451
  Nombre de de tests + chez malades=4.920 ;
  Nombre de tests - chez malades=51
  Nombre de tests + chez saines=49.629 ;
  Nombre de tests - chez saines=945.400
- Quelle est la proportion de malades ?
  nombre de malades/1.000.000=0,004971.
- Quelle est la proportion de tests positifs ?
  nombre de tests +/1.000.000=0,054549
- Proportion de tests positifs chez les malades ?
  nb. tests + chez malades/ nb. malades= 0,98974049
- Proportion de tests positifs chez les saines ?
  nb. tests + chez saines/ nb. saines=0,04987694

Les données

- Nombre de malades=4.971 ; Nombre de saines=995.029
  Nombre de tests += 54.549 ; Nombre de tests -= 945.451
  Nombre de de tests + chez malades=4.920 ;
  Nombre de tests - chez malades=51
  Nombre de tests + chez saines=49.629 ;
  Nombre de tests - chez saines=945.400
- Quelle est la proportion de malades ?
  nombre de malades/1.000.000=0,004971.
- Quelle est la proportion de tests positifs ?
  nombre de tests +/1.000.000=0,054549
- Proportion de tests positifs chez les malades ?
  nb. tests + chez malades/ nb. malades= 0,98974049
- Proportion de tests positifs chez les saines ?
  nb. tests + chez saines/ nb. saines=0,04987694

Les données

- Nombre de malades=4.971 ; Nombre de saines=995.029
  Nombre de tests += 54.549 ; Nombre de tests -= 945.451
  Nombre de de tests + chez malades=4.920 ;
  Nombre de tests - chez malades=51
  Nombre de tests + chez saines=49.629 ;
  Nombre de tests - chez saines=945.400
- Quelle est la proportion de malades ?
  nombre de malades/1.000.000=0,004971.
- Quelle est la proportion de tests positifs ?
  nombre de tests +/1.000.000=0,054549
- Proportion de tests positifs chez les malades ?
  nb. tests + chez malades/ nb. malades= 0,98974049
- Proportion de tests positifs chez les saines ?
  nb. tests + chez saines/ nb. saines=0,04987694

Les données

- Nombre de malades=4.971 ; Nombre de saines=995.029
  Nombre de tests += 54.549 ; Nombre de tests -= 945.451
  Nombre de de tests + chez malades=4.920 ;
  Nombre de tests - chez malades=51
  Nombre de tests + chez saines=49.629 ;
  Nombre de tests - chez saines=945.400
- Quelle est la proportion de malades ?
  nombre de malades/1.000.000=0,004971.
- Quelle est la proportion de tests positifs ?
  nombre de tests +/1.000.000=0,054549
- Proportion de tests positifs chez les malades ?
  nb. tests + chez malades/ nb. malades= 0,98974049
- Proportion de tests positifs chez les saines ?
  nb. tests + chez saines/ nb. saines=0,04987694

Les données

- Nombre de malades=4.971 ; Nombre de saines=995.029
  Nombre de tests += 54.549 ; Nombre de tests -= 945.451
  Nombre de de tests + chez malades=4.920 ;
  Nombre de tests - chez malades=51
  Nombre de tests + chez saines=49.629 ;
  Nombre de tests - chez saines=945.400
- Quelle est la proportion de malades ?
  nombre de malades/1.000.000=0,004971.
- Quelle est la proportion de tests positifs ?
  nombre de tests +/1.000.000=0,054549
- Proportion de tests positifs chez les malades ?
  nb. tests + chez malades/ nb. malades= 0,98974049
- Proportion de tests positifs chez les saines ?
  nb. tests + chez saines/ nb. saines=0,04987694

Les données

- Ces résultats vous semblent-ils conformes à ce qui est annonce par le fabriquant ?
- Un patient arrive à l'hôpital : le test est positif. Quel est le risque qu'il soit effectivement malade ?
  On regarde la proportion de malades parmi les individus pour lesquels le test est positif :
  4920/(4920+49629)=4920/54549=0,09019
- Commenter ce résultat.

Les données

- Ces résultats vous semblent-ils conformes à ce qui est annonce par le fabriquant ?
- Un patient arrive à l'hôpital : le test est positif. Quel est le risque qu'il soit effectivement malade ?
  On regarde la proportion de malades parmi les individus pour lesquels le test est positif :
  4920/(4920+49629)=4920/54549=0,09019
- Commenter ce résultat.

Les données

- Ces résultats vous semblent-ils conformes à ce qui est annonce par le fabriquant ?
- Un patient arrive à l'hôpital : le test est positif. Quel est le risque qu'il soit effectivement malade ?
  On regarde la proportion de malades parmi les individus pour lesquels le test est positif :
  4920/(4920+49629)=4920/54549=0,09019
- Commenter ce résultat.

Sondages

Les données

Une enquête est menée auprès d'un échantillon de n électeurs potentiels. On leur demande s'ils sont plutôt favorables au candidat 1 ou au candidat 2. On obtient 53% d'intentions de vote pour le candidat numéro 1 (on note $p_n = 0.53$).

Peut-on affirmer que le candidat 1 rassemble plus d'intentions de vote dans l'ensemble de la population ?

Quelle information nous manque t-il ?

On suppose que $n = 100$, l'intervalle de confiance pour p est $[0.4321784; 0.6278216]$. Peut-on conclure quelque chose ?

Quelle taille d'échantillon est nécessaire pour avoir une précision de 0.03 ?

$\rightarrow$ illustration EXOSONDAGE

Les données

Une enquête est menée auprès d'un échantillon de n électeurs potentiels. On leur demande s'ils sont plutôt favorables au candidat 1 ou au candidat 2. On obtient 53% d'intentions de vote pour le candidat numéro 1 (on note $p_n = 0.53$).

Peut-on affirmer que le candidat 1 rassemble plus d'intentions de vote dans l'ensemble de la population ?

Quelle information nous manque t-il ?

On suppose que $n = 100$, l'intervalle de confiance pour p est $[0.4321784; 0.6278216]$. Peut-on conclure quelque chose ?

Quelle taille d'échantillon est nécessaire pour avoir une précision de 0.03 ?

$\rightarrow$ illustration EXOSONDAGE

Les données

Une enquête est menée auprès d'un échantillon de n électeurs potentiels. On leur demande s'ils sont plutôt favorables au candidat 1 ou au candidat 2. On obtient 53% d'intentions de vote pour le candidat numéro 1 (on note $p_n = 0.53$).

Peut-on affirmer que le candidat 1 rassemble plus d'intentions de vote dans l'ensemble de la population ?

Quelle information nous manque t-il ?

On suppose que $n = 100$, l'intervalle de confiance pour p est $[0.4321784; 0.6278216]$. Peut-on conclure quelque chose ?

Quelle taille d'échantillon est nécessaire pour avoir une précision de 0.03 ?

$\rightarrow$ illustration EXOSONDAGE

Les données

Une enquête est menée auprès d'un échantillon de n électeurs potentiels. On leur demande s'ils sont plutôt favorables au candidat 1 ou au candidat 2. On obtient 53% d'intentions de vote pour le candidat numéro 1 (on note $p_n = 0.53$).

Peut-on affirmer que le candidat 1 rassemble plus d'intentions de vote dans l'ensemble de la population ?

Quelle information nous manque t-il ?

On suppose que $n = 100$, l'intervalle de confiance pour p est $[0.4321784; 0.6278216]$. Peut-on conclure quelque chose ?

Quelle taille d'échantillon est nécessaire pour avoir une précision de 0.03 ?

$\rightarrow$ illustration EXOSONDAGE

Les données

Une enquête est menée auprès d'un échantillon de n électeurs potentiels. On leur demande s'ils sont plutôt favorables au candidat 1 ou au candidat 2. On obtient 53% d'intentions de vote pour le candidat numéro 1 (on note $p_n = 0.53$).

Peut-on affirmer que le candidat 1 rassemble plus d'intentions de vote dans l'ensemble de la population ?

Quelle information nous manque t-il ?

On suppose que $n = 100$, l'intervalle de confiance pour p est $[0.4321784; 0.6278216]$. Peut-on conclure quelque chose ?

Quelle taille d'échantillon est nécessaire pour avoir une précision de 0.03 ?

$\rightarrow$ illustration EXOSONDAGE

Sondages sur données sensibles

Méthode utilisée

On souhaite évaluer la proportion P de personnes faisant des téléchargements illégaux dans une population.
On ne peut pas leur poser directement la question car ils risquent de ne pas nous dire la vérité par peur d'éventuels problèmes judiciaires.
Un statisticien propose la démarche suivante : demander aux individus de lancer un dé dans un isoloir puis de répondre seulement à :

- Q1 : "avez-vous obtenu 1 ou 2 ?", s'ils téléchargent illégalement.
- Q2 : "avez-vous obtenu 3,4,5 ou 6 ?", s'ils ne téléchargent pas illégalement.

Seule la personne interrogée connait la question à laquelle elle répond.

Méthode utilisée

On souhaite évaluer la proportion P de personnes faisant des téléchargements illégaux dans une population.
On ne peut pas leur poser directement la question car ils risquent de ne pas nous dire la vérité par peur d'éventuels problèmes judiciaires.
Un statisticien propose la démarche suivante : demander aux individus de lancer un dé dans un isoloir puis de répondre seulement à :

- Q1 : "avez-vous obtenu 1 ou 2 ?", s'ils téléchargent illégalement.
- Q2 : "avez-vous obtenu 3,4,5 ou 6 ?", s'ils ne téléchargent pas illégalement.

Seule la personne interrogée connait la question à laquelle elle répond.

Comment marche-t-elle ?

- Quelle est la probabilité qu'une personne qui télécharge illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne qui ne télécharge pas illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne interrogée réponde OUI ?

$$p_{OUI} = p_{PIRATES}\frac{1}{3} + (1 - p_{PIRATES})\frac{2}{3} = \frac{2 - p_{PIRATES}}{3}.$$

Comment marche-t-elle ?

- Quelle est la probabilité qu'une personne qui télécharge illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne qui ne télécharge pas illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne interrogée réponde OUI ?

$$p_{OUI} = p_{PIRATES}\frac{1}{3} + (1 - p_{PIRATES})\frac{2}{3} = \frac{2 - p_{PIRATES}}{3}.$$

Comment marche-t-elle ?

- Quelle est la probabilité qu'une personne qui télécharge illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne qui ne télécharge pas illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne interrogée réponde OUI ?

$$p_{OUI} = p_{PIRATES}\frac{1}{3} + (1 - p_{PIRATES})\frac{2}{3} = \frac{2 - p_{PIRATES}}{3}.$$

Comment marche-t-elle ?

- Quelle est la probabilité qu'une personne qui télécharge illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne qui ne télécharge pas illégalement réponde OUI ?
- Quelle est la probabilité qu'une personne interrogée réponde OUI ?

$$p_{OUI} = p_{PIRATES}\frac{1}{3} + (1 - p_{PIRATES})\frac{2}{3} = \frac{2 - p_{PIRATES}}{3}.$$

Comment marche-t-elle ?

- Comment peut-on donner une valeur approchée de p_{OUI} ?
- En déduire une valeur approchée de $p_{PIRATES}$.

$$p_{PIRATES} = 2 - 3p_{OUI} \Rightarrow \hat{p}_{PIRATES} = 2 - 3\hat{p}_{OUI}$$

- Application. → illustration EXOSUP

Comment marche-t-elle ?

- Comment peut-on donner une valeur approchée de p_{OUI} ?
- En déduire une valeur approchée de $p_{PIRATES}$.

$$p_{PIRATES} = 2 - 3p_{OUI} \Rightarrow \hat{p}_{PIRATES} = 2 - 3\hat{p}_{OUI}$$

- Application. → illustration EXOSUP

Comment marche-t-elle ?

- Comment peut-on donner une valeur approchée de p_{OUI} ?
- En déduire une valeur approchée de $p_{PIRATES}$.

$$p_{PIRATES} = 2 - 3p_{OUI} \Rightarrow \hat{p}_{PIRATES} = 2 - 3\hat{p}_{OUI}$$

- Application. $\rightarrow$ illustration EXOSUP