

Examen de Statistiques

Janvier 2005

Bareme:

3 Stat descriptive

3 Prob

5 Premier prob estimation stat

6 Deux prob estimation stat

3 Trois prob estimation stat

Tot = 20

1 Statistique descriptive

Répondre aux questions suivantes pour les données représentés dans la Fig. 1.

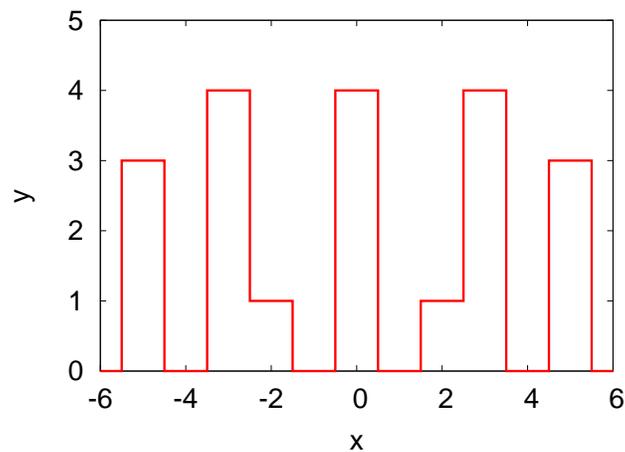


Figure 1: Description graphique des données à étudier en **1 Statistique descriptive**.

1. 0.5 Peut ce dessin représenter un diagramme en fréquence ? Pour quoi ?
2. 1 Donner la définition et la valeur des moments impairs.
3. 0.5 Est la médiane égale ou différente de la moyenne ?
4. 1 Y a-t-il une raison simple pour justifier (deviner) le résultat de deux questions précédentes sans faire les calculs ?

2 Probabilités

1. Soient A et B deux événements indépendants avec probabilités $P(A)$ et $P(B)$.
 - (a) 0.5 Donner une expression de la probabilité de l'événement $A \wedge B$ avec \wedge le "et logique".
 - (b) 0.5 Donner une expression de la probabilité de l'événement $A \vee B$ avec \vee le "ou logique".
 - (c) 1 Montrer que les événements \bar{A} et B , où \bar{A} est l'événement complémentaire de A, sont eux aussi indépendants.
2. 1 Déterminer si la fonction

$$g(x) = \begin{cases} 0, & \text{si } x \leq 0, \\ x, & \text{si } 0 \leq x \leq 2, \\ 4 - x, & \text{si } 2 \leq x \leq 3, \\ 1, & \text{si } x \geq 3, \end{cases} \quad (1)$$

peut décrire la probabilité qu'une variable aléatoire continue y soit plus grande que x , $y \geq x$. Autrement dit, peut $g(x)$ être égalée à la distribution cumulée $P(y \geq x)$? Expliquer.

3 Estimation statistique

1. On veut estimer la moyenne d'un observable x sur une population de taille N . On travaille avec des échantillons de taille n ($n < N$).

- (a) 0.5 Les moyennes de deux échantillons différents seront-elles, typiquement, identiques ou différentes ?
- (b) 0.5 Quel est le meilleur estimateur de la moyenne de la population ?
- (c) 1.5 Appelons \bar{x} la moyenne de chaque échantillon de taille n .
 - i. Est-elle une variable aléatoire ?
 - ii. Si oui, avec quelle distribution de probabilité ?
 - iii. Caractériser la valeur d'expectation et l'écart-type de la distribution de probabilité que vous avez proposé.
- (d) 0.5 Quel est le meilleur estimateur de l'écart-type de la population ?
- (e) 2 Si on prend un échantillon de taille $n = 30$ et on trouve qu'il a une moyenne $\bar{x} = 1$ et un écart-type $\sigma = 0.2$, Déterminer l'intervalle de variation de la moyenne de la population avec un niveau de confiance du 98%.

2. En arrivant au laboratoire, un chercheur vérifie que ses quatre manips soient bien réglées. Après 200 jours il trouve que le nombre de manips bien réglées (et prêtes à être utilisées) présentés dans la table suivante

Nombre de manips bien réglées	0	1	2	3	4
Nombre de jours	x_0	x_1	x_2	x_3	x_4

(x_0 est le nombre de jours où il a trouvé aucune manip réglée, x_1 est le nombre de jours où il a trouvé une manip réglée, ..., x_4 est le nombre de jours où il a trouvé toutes les manips bien réglées.) Le chercheur croit pouvoir décrire ses observations, avec une distribution binomiale de paramètre p .

- (a) 1 Expliquer comment calculer les nombres de jours avec 0, 1, 2, 3 ou 4 manips bien réglées, qu'on appellera y_i avec $i = 0, 1, 2, 3, 4$, prédits par la loi binomiale.
- (b) 0.5 Quelle quantité peut-on utiliser pour tester l'hypothèse du chercheur ? Donner la définition.
- (c) 1 Cette quantité, est-elle aléatoire ou déterministe ? Pour quoi ?

- (d) 1.5 Si la réponse à la question précédente est “aléatoire” quelle est sa distribution de probabilité ? Caractériser cette loi de probabilité.
- (e) 2 Si le chercheur trouve 5.25 pour le calcul explicite de la quantité proposée en (b), et il veut être sûr de sa proposition au 95%, doit-il rejeter son hypothèse ? Expliquer.
3. Nous voulons décrire les données présentés en Fig. 2 avec une fonction $y(x)$. La figure montre avec points carrés rouges cinq couples expérimentales (x_i, y_i) , avec $i = 1, \dots, 5$. La courbe rouge est une régression linéaire, $y = ax + b$, la courbe verte est une description quadratique, $y = ax^2 + bx + c$, tandis que la courbe bleue pointillée est un polynôme d'ordre quatre, $y = ax^4 + bx^3 + cx^2 + dx + e$, passant par les cinq points (x_i, y_i) .

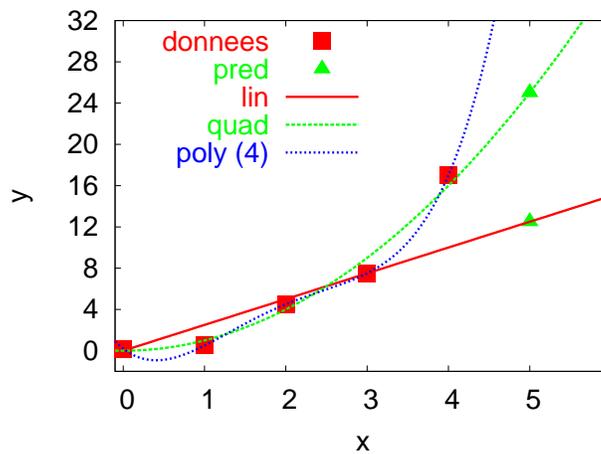


Figure 2: Les données à étudier en **3 Estimation statistique**, question 3.

- (a) 0.5 Proposer un critère pour déterminer la meilleure valeur des paramètres a et b utilisés pour la description linéaire des données.
- (b) 0.5 Proposer un critère pour déterminer la meilleure valeur des paramètres a , b et c utilisés pour la description quadratique des données.

- (c) 1 À votre avis, quelle description est la plus pertinente dans ce cas, linéaire, quadratique ou polynomiale ? Expliquer.
- (d) 1 Donner une prédiction pour la valeur “à futur” $y(x = 5)$ [choisir parmi les triangles verts correspondants aux descriptions linéaires et quadratique et la valeur prédite par la loi polynomiale, $y(x = 5) = 52.7$].