

TD 1 : Statistiques descriptives

Exercice 1. Quartiles approchés et histogrammes

Table 1 – Distribution du nombre de cigarettes fumées par jour pour les 484 mères fumeuses.

Nb de cig.	Nb. de fumeurs (%)
0–5	16
5–10	25
10–15	14
15–20	4
20–30	32
30–40	5
40–60	4
Total	100

- a) A partir du tableau 1 trouver les quartiles approchés de la distribution du nombre de cigarettes fumées par jour des mères fumeuses pendant la grossesse.
- b) Combiner les 4 dernières classes de ce tableau et faire un histogramme à partir de ce nouveau tableau. Comment l’histogramme a-t-il évolué par rapport à celui présent dans le cours? Expliquez.

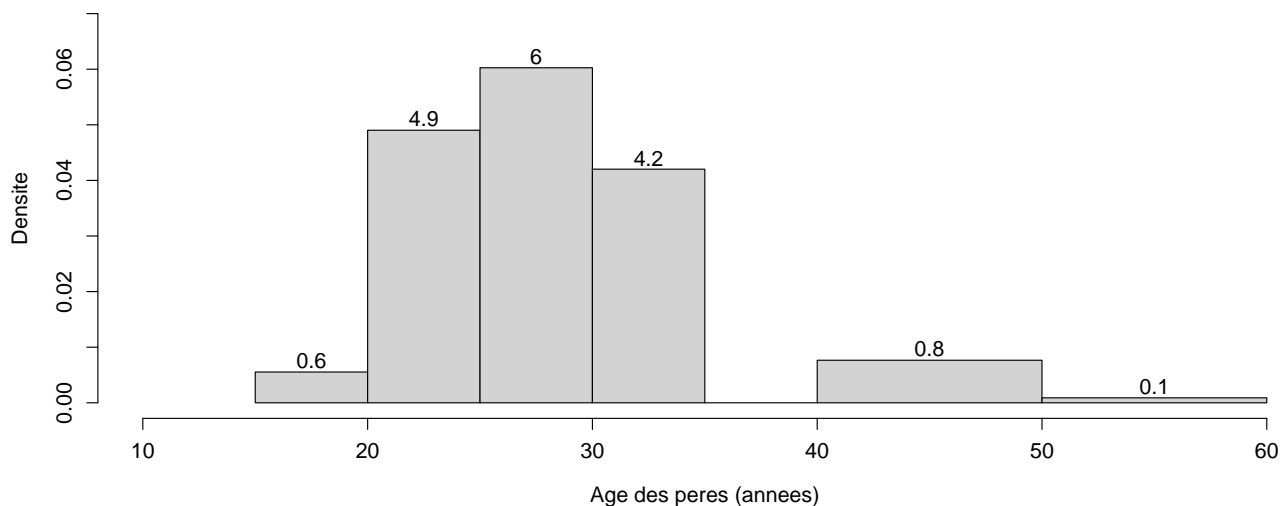


Figure 1 – Histogramme de l’âge des pères dans l’étude CHDS. Les nombres indiquent la hauteur (en %) de chaque rectangle. Le rectangle pour la classe [35, 40) est manquant.

- c) Regardez l’histogramme de la figure 1. On a oublié de faire le graphique pour la classe [35, 40). Comblez cette lacune.



Exercice 2. La loi Normale

- a) Utilisez l’approximation normale pour estimer la proportion de mères mesurant entre 61.5 et 64.5 pouces sachant que la taille moyenne est 64 pouces et l’écart-type 2.5 pouces.

- b) Supposons que l'on dispose de 100 observations issues d'une loi normale centrée réduite. Quelle proportion des observations peut-on espérer voir en dehors des « moustaches » du boxplot ?



Exercice 3. QQ-plot

- a) Ci-dessous vous trouverez les quantiles associés aux probabilités 0.05, 0.10, ..., 0.95 de la durée de la grossesse (en jours) pour les mères de l'étude CHDS. Tracez ces quantiles en fonction de ceux d'une loi $U(0, 1)$. Décrivez la forme de la distribution de la durée de la grossesse par rapport à la loi uniforme.

252, 262, 267, 270, 274, 276, 277, 278, 280, 281, 283, 284, 286, 288, 290, 292, 296, 302

- b) Supposez que les quantiles y_p d'une $N(\mu, \sigma^2)$ sont tracés en fonctions des quantiles z_p d'une $N(0, 1)$. Montrez que la pente et l'ordonnée à l'origine de la droite des points sont σ et μ respectivement.



Exercice 4. Moyenne empirique

Montrez que \bar{x} est la valeur qui minimise (par rapport à c) la fonction suivante

$$f(c) = \sum_{i=1}^n (x_i - c)^2$$



Exercice 5. Moyennes et écart-types

- a) Dans une étude du Missouri, le poids moyen à la naissance des bébés issus de mères fumeuses est 3180g et l'écart-type de 500g. Quel est le poids moyen et l'écart-type en onces sachant qu'il y a 0.035 onces dans 1g.
- b) Soient x_1, \dots, x_n quelques observations. Pour des raisons de commodités, Bob a changé les unités menant à de nouvelles observations

$$y_i = ax_i + b, \quad i = 1, \dots, n.$$

Exprimez la moyenne et l'écart-type des y_i en fonction de ceux des x_i .

