

Université de Bretagne Sud
Eléments de statistique

Travaux Pratiques R

Guillaume Franchi

Année universitaire 2024/2025

■ Quelques statistiques

Exercice 1

On a recensé, dans une entreprise, les salaires bruts annuels de tous les employés, en milliers d'euros. Les données sont disponibles dans le fichier `salaires.csv`, disponible sur la page web

https://guillaumefranchi.github.io/index_fr.html

- 1) A l'aide de la fonction `summary`, effectuer un rapide descriptif des données.
- 2) A l'aide de la fonction `boxplot`, dessiner le diagramme en boîtes représentant la distribution des salaires de l'entreprise.
- 3) A l'aide de la fonction `hist`, dessiner l'histogramme des salaires de l'entreprise. Qu'en pensez-vous ?
- 4) On suppose que la distribution des salaires suit une loi $\mathcal{N}(\mu, \sigma^2)$, où μ et σ sont inconnus. Proposer des estimations $\hat{\mu}$ et $\hat{\sigma}$ de μ et σ .
- 5) La commande suivante

```
t = seq(from=0,to=60,length.out=500),
```

crée un vecteur de taille 500, qui est en fait une subdivision régulière de l'intervalle $[0, 60]$, contenant 500 points. A l'aide de la fonction `dnorm`, évaluer la fonction de densité f de la loi $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ en chacun des points de cette subdivision.

- 6) A l'aide de la fonction `lines`, superposer la fonction f à l'histogramme dessiné précédemment. Notre estimation du modèle vous semble-t-elle pertinente ?
- 7) On rencontre une personne au hasard dans l'entreprise.
 - a) A l'aide de la fonction `pnorm`, calculer la probabilité que cette personne ait un salaire brut annuel compris entre 30 000 et 40 000 euros.
 - b) Déterminer cette même probabilité, sachant que cette personne a un salaire supérieur à 20 000 euros.

Exercice 2

Dans le fichier `composants.csv`, on a recensé la durée de vie de 1 500 composants électroniques identiques, en heures.

- 1) Reprendre les questions de l'exercice précédent et dessiner le diagramme en boîtes de cette série statistique, ainsi que son histogramme.
- 2) Peut-on supposer ici que la durée de vie d'un composant suit une loi normale ? Pourquoi ?

- 3) Proposer de modéliser cette durée de vie par une autre loi, et en préciser le(s) paramètre(s).
- 4) Calculer une estimation de ce(s) paramètre(s), puis, en s'inspirant de l'exercice précédent, superposer la fonction de densité estimée de la loi de durée de vie d'un composant à l'histogramme tracé.
- 5) Avec la fonction `pexp`, déterminer la probabilité qu'un composant ait une durée de vie supérieure à 2000 heures.
- 6) Déterminer la probabilité qu'un tel composant tienne encore 2000 heures au moins, sachant qu'il fonctionne depuis 2000 heures déjà. Que remarquez-vous ?

■ Quelques simulations

Exercice 3

On souhaite simuler $N = 1000$ simulations indépendantes d'une loi normale $\mathcal{N}(0, 1)$.

- 1) Effectuer ces simulations avec la fonction `rnorm`.
- 2) Dessiner l'histogramme de la série statistique obtenue, et superposer la fonction de densité de la loi $\mathcal{N}(0, 1)$.
- 3) Répéter ce même type de simulations pour des lois $\mathcal{N}(\mu, \sigma^2)$ pour les valeurs :
 - ▷ $\mu = 0$ et $\sigma = 0.1$;
 - ▷ $\mu = 0$ et $\sigma = 10$;
 - ▷ $\mu = 0$ et $\sigma = 100$.

Que constate-t-on ?

Exercice 4

- 1) Pour $N \in \{10, 30, 1000\}$, réaliser 1000 simulations d'une loi binomiale $\mathcal{B}(N, p)$, pour $p = 0.6$.
- 2) Tracer ensuite les histogrammes de ces réalisations. Superposer à ces histogrammes la courbe de la densité d'une loi normale soigneusement choisie.