

Sujet A – Evaluation 1 – TPSTA401 – 2024

Frédérique Leblanc

Exercice 1 (8pts)

Choisir comme répertoire de travail avec Rstudio (menu Session/Choose Working Directory/ ...) un répertoire de votre “home” contenant le fichier `diamantsPurs.csv`. Charger les données et les affecter à un objet nommé `diam` avec l’instruction suivante :

```
diam<-read.table("diamantsPurs.csv",header=T,sep=" ")
```

question 1 (2pts)

Extraire de ce data frame l’échantillon des prix pour les diamants pour lesquels la variable `cut` prend la modalité `Premium` et l’affecter à `x` puis affecter à `y` l’échantillon des prix des diamants pour lesquels la variable `cut` prend la modalité `VeryGood` (tronçon d’instructions suivant à compléter) :

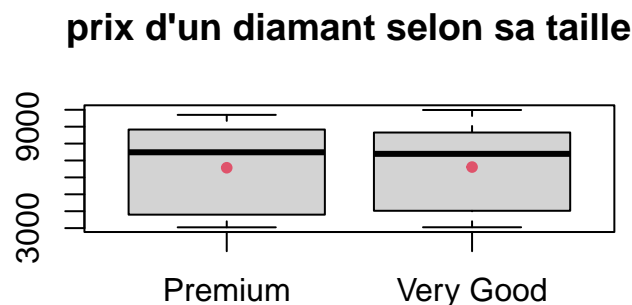
```
#extraction de l'échantillon x  
#extraction de l'échantillon y
```

question 2 (3pts)

Donner les tailles, moyennes empiriques et écart-types empiriques corrigés (noté s' en cours) de chaque échantillon (tronçon des instructions de calculs **et** texte de réponse à la question).

question 3 (3pts)

Pour comparer ces deux échantillons on peut représenter leurs boxplots sur un même graphique auquel sont ajoutées ici les positions des moyennes (points rouges) :

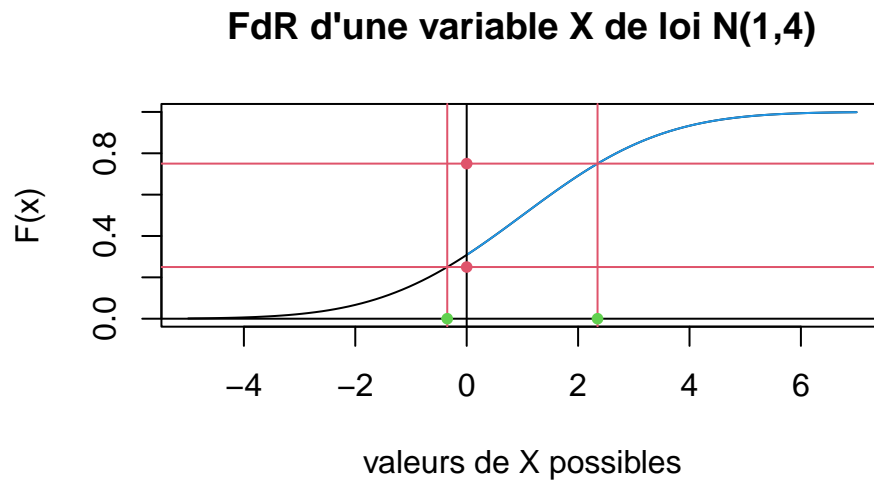


Interpréter ce graphique (3 commentaires attendus).

Exercice 2 (12 pts)

question 1 (10 pts)

Reproduire le graphique suivant où est représentée la fonction de répartition d'une variable normale de moyenne $\mu = 1$ et variance $\sigma^2 = 4$ sur l'intervalle $[-5, 7]$. Les points rouges (option `pch=20`) sont les points d'abscisse 0 et ordonnées 0.25 et 0.75 et les points verts les points d'ordonnée 0 et abscisses les antécédants par F de 0.25 et 0.75 (la couleur bleue utilisée est obtenue avec `col=4`) :



question 2 (2 pts)

Dans le graphique suivant représentant la densité d'une variable de loi $\mathcal{N}(1, 4)$ que vaut la surface délimitée par le trait vert (les deux points verts placés sur l'axe des abscisses sont les mêmes que ceux placés sur le précédent graphique) ?

