Exercice 8

① Déterminer la liste des fréquences cumulées ne nécessite pas de faire appel à une quelconque fonction Python si on note que les caractères observés sont distincts

A la "main", on commencera par trier les données de l'échantillon et on fera:

x_i	14	16	25	29	38	41	...	238	325	410
n_i	1	1	1	1	1	1	...	1	1	1
$F_c(x_i)$	$1/20$	$2/20$	$3/20$	$4/20$	$5/20$	$6/20$		$18/20$	$19/20$	1

Il suffit alors de tracer les fréquences cumulées correspondant à chacun des caractères x_i.

Avec Python, on pourra faire :

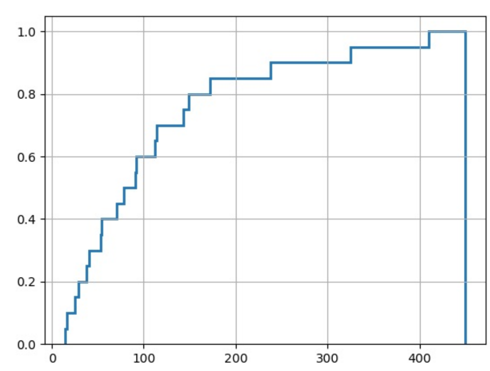
```
def frequencesCumulees(S):
    St = sorted(S)     # série triée
    n = len(St)
    Fc = [k/n for k in range(1, n+1)]
    return St, Fc
```

On donnera une représentation graphique avec la fonction "plt.stairs" pour tracer le polygône des fréquences cumulées ; Il faut juste compléter la liste "St" par une valeur maximale jusqu'à laquelle la ligne sera tracée...

Par exemple:

```
St, Fc = frequencesCumulees(S)
St += [450]
plt.stairs(Fc, St)
plt.grid()
plt.show()
```

va tracer:



avec

$F_c = [0.05, 0.1, 0.15, 0.2, 0.25, 0.3, \ldots$
$\ldots 0.9, 0.95, 1]$

② Les fonctions 'moyenne()', 'variance()' et 'mediane()' doivent pouvoir être écrites SANS recours à aucune fonction prédéfinie Python (y compris 'sum()')

→ S'entraîner à les rédiger.

On obtient $\boxed{m = 113,25 \;;\; s^2 = 10513,5 \text{ et } med = 84.5}$

③ On teste l'hypothèse que les durées de survie peuvent être modélisées par une VAR $X \hookrightarrow \mathcal{E}(\lambda)$

Sous cette hypothèse $F_c(x_i) \cong F_X(x_i) = 1 - e^{-\lambda x_i}$ (*)
(car $x_i \in S$ donc $x_i > 0$)

(*) $\Leftrightarrow e^{-\lambda x_i} \cong 1 - F_c(x_i)$

$\Leftrightarrow -\lambda x_i \cong \ln(1 - F_c(x_i)) \Rightarrow x_i \cong -\frac{1}{\lambda}\ln(1 - F_c(x_i))$

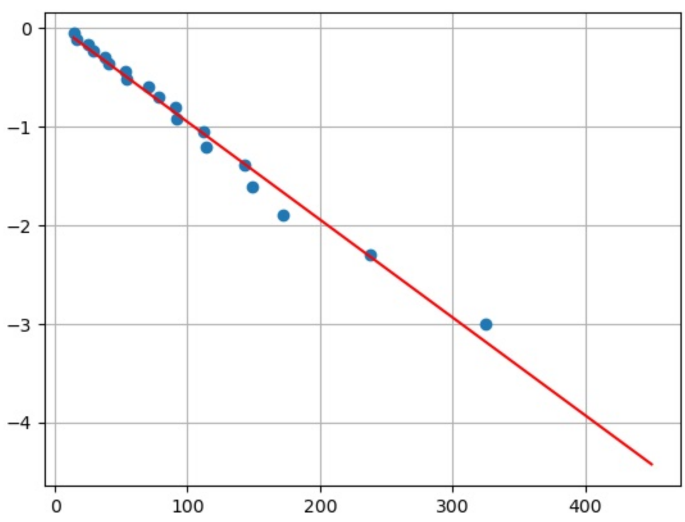
Conséquence : L'hypothèse sera jugée validée si, après avoir posé $Y_i = \ln(1 - F_c(x_i))$, on observe une relation linéaire entre $S = [x_0, \dots, x_{19}]$ et Y.

```
n = len(Fc)
Y = [np.log(1 - Fc[k]) for k in range(n-1)]
plt.plot(St[:-2], Y, 'o')
```

on enlève les deux dernières valeurs de St : 450 et 410 car $F_c(410) = 1$.

ne pas oublier que $F_c(n-1) = 1$ et donc $\ln(1 - F_c(1))$ n'est pas défini...

!) Le calcul du coefficient de corrélation donne $r_{S,Y} = -0.99$ confirme une relation affine entre S et Y de la forme : $Y = aS + b$



Pour obtenir une valeur approchée de a et b on calcule les coefficients de la droite de régression (tracée en rouge sur la figure ci-contre)

→ les fonctions permettant d'obtenir a et b sont aussi à savoir écrire !

Si on a le droit à la bibliothèque 'numpy' on pourra se contenter de faire :

$(a, b) = np.polyfit(St[:-2], Y, 1)$

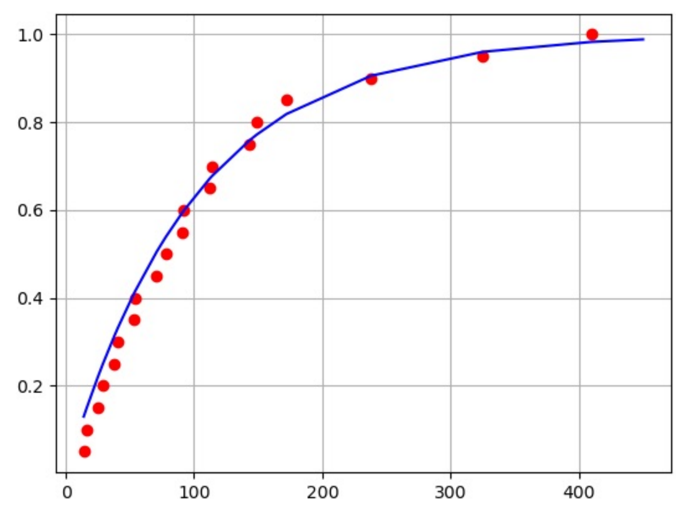
c) sous l'hypothèse d'une loi exponentielle :

$Y = -\lambda S$ et donc $\lambda = -a \simeq 0.01$

On confrontera alors les données expérimentales au modèle dans lequel $X \hookrightarrow \mathcal{E}(\lambda)$ en traçant sur une même figure :

$x \mapsto 1 - e^{-\lambda x}$ et F_C

Ce qui donne :



!) on note que :

$E(X) = \frac{1}{\lambda} = 100 \simeq 113$

et

$V(X) = \frac{1}{\lambda^2} = 10000 \simeq 10513$

qui sont les moyennes et variances expérimentales