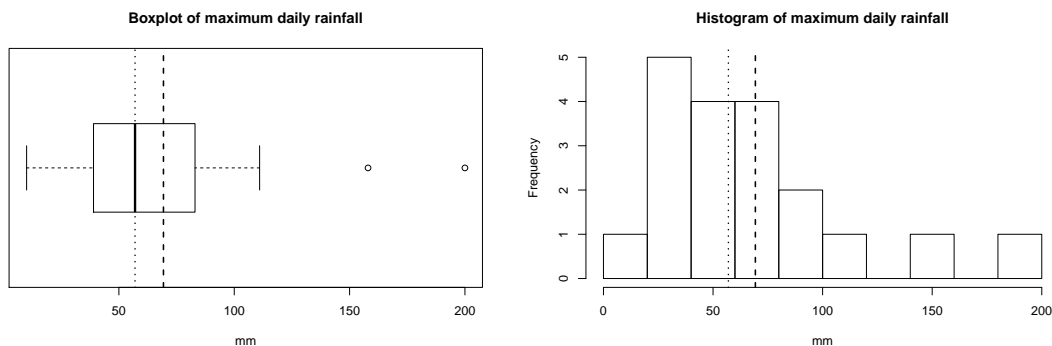


CORRIGÉ 2

Exercice 1. En traçant la moyenne (69.32 mm) et la médiane (57 mm) sur les graphiques, on voit que la moyenne est décalée vers les deux observations atypiques. Même si la tendance centrale des maxima des pluies journalières typiques est mieux représentée par la médiane, la différence entre la moyenne et la médiane met en évidence les deux années marquées par une journée de novembre particulièrement pluvieuse. En regardant l'histogramme, on peut aussi constater que les données à droite de la médiane sont plus dispersées que les données à gauche. Cela nous indique que la probabilité d'observer une quantité de pluie journalière beaucoup plus grandes que la médiane n'est pas aussi petite que l'on pourrait le penser (elle est non-négligeable).



- Exercice 2.** (a) Chaque proportion est entre 0 et 1, et la somme est bien égale à 1.
 (b) Le pourcentage des femmes divorcées ou jamais mariées est égal à $7.1\% + 35.3\% = 42.4\%$.
 (c) Le pourcentage des femmes qui ne sont pas mariées est égal à $100\% - 57.4\% = 42.6\%$.

Exercice 3. (a) On a vu dans le cours que, pour la distribution $\mathcal{N}(\mu, \sigma^2)$, le pourcentage d'observations dans $[\mu - \sigma, \mu + \sigma]$ est 68 %. Ici, avec $\mu = \bar{x} = 176.6$ et $\sigma = s_x = 7.99$, cet intervalle est $[168.61, 184.59]$ et la fréquence relative vaut 63.3 %, ce qui est relativement proche de la proportion attendue sous la distribution normale.

(b) On standardise par la transformée $x \mapsto (x - \bar{x})/s_x$. Donc la probabilité d'être inférieur ou égal à 170 sous la distribution $\mathcal{N}(\bar{x}, s^2)$ est la même que la probabilité d'être inférieur ou égal à $(170 - 176.6)/7.99 = -0.83$ sous la distribution $\mathcal{N}(0, 1)$. La probabilité cherchée est la valeur $\Phi(-0.83)$. Cette valeur n'est pas dans le tableau, mais on peut utiliser la symétrie pour obtenir $\Phi(-0.83) = 1 - \Phi(0.83) = 1 - 0.79673 = 0.20327 = 20.327\%$. La fréquence relative d'observations inférieures ou égales à 170 cm est 23.3 %, ce qui est proche de la valeur sous le modèle normal.

Exercice 4. (a) La première distribution décrit des données qui sont plus grandes qu'une certaine valeur (par exemple, si cette valeur est 0, les données sont positives), et qui sont en grande partie concentrées près de ce nombre. Le pourcentage attendu d'observations décroît rapidement en fonction de la distance par rapport à ce nombre.

La deuxième distribution décrit des données qui sont contenues entre deux nombres, et qui sont réparties entre eux de façon uniforme.

La troisième distribution décrit des données qui sont plus grandes qu'un certain nombre, et qui sont en grande partie concentrées dans une région à droite de ce nombre. Le pourcentage attendu d'observations décroît en fonction de la distance par rapport à cette région, et ce de façon plus rapide vers la gauche que vers la droite.

La quatrième distribution décrit des données qui sont dispersées de façon symétrique autour d'un certain nombre, et qui sont concentrées autour de ce nombre. Le pourcentage attendu d'observations décroît rapidement en fonction de la distance par rapport à ce nombre.

- (b) La troisième pour l'exercice 1 et la quatrième pour l'exercice 3. On pourrait dire que la première est également possible pour l'exercice 1, mais elle tolère moins les grandes valeurs que la troisième.
- (c) Oui, la quatrième.