T. D. n^{o} 4
Échantillonnage. Estimation d'un paramètre.

Exercice 1 Estimateur de l'espérance, page 242 du livre « Initiation à la statistique avec `R` ».
Considérons le fichier `Mesures` qui est dans le package `BioStatR`. Nous voulons connaître une estimation sans biais de l'espérance μ de la variable `taille` des glycines blanches. Pour cela, suivre la démarche suivante.

a) Extraire les 54 données des glycines blanches des 252 données des quatre espèces en vous servant de la commande `subset`. Si vous ne savez pas vous servir de la commande `subset`, alors consultez l'aide en ligne. N'oubliez pas de donner un nom à votre jeu de données qui provient de l'extraction.
b) Calculer une estimation de l'espérance de la variable `taille` des glycines blanches en vous servant de la commande `mean`.
c) Conclure.

Exercice 2 Estimateurs de la variance, pages 243 et 244 du livre « Initiation à la statistique avec `R` ».
Considérons le fichier `Mesures` qui est dans le package `BioStatR`. Nous voulons connaître une estimation de la variance σ^2 de la variable `taille` des glycines blanches dont l'espérance est inconnue. Pour cela, suivre la démarche suivante.

a) Extraire les 54 données des glycines blanches des 252 données des quatre espèces en vous servant de la commande `subset`. Si vous ne savez pas vous servir de la commande `subset`, alors consultez l'aide en ligne. N'oubliez pas de donner un nom à votre jeu de données qui provient de l'extraction.
b) Calculer une estimation biaisée de la variance de la variable `taille` des glycines blanches en vous servant de la commande `var` et `length`.
c) Calculer une estimation non biaisée de la variance de la variable `taille` des glycines blanches en vous servant de la commande `var`.
d) Conclure.

Exercice 3 Estimateur d'une proportion, pages 244 et 245 du livre « Initiation à la statistique avec `R` ».
Considérons le fichier `Mesures` qui est dans le package `BioStatR`. Nous voulons connaître une estimation sans biais de la proportion des gousses de glycine blanche qui ont moins de trois graines présentes dans une gousse. Pour cela, suivre la démarche suivante.

a) Extraire les 54 données des glycines blanches des 252 données des quatre espèces en vous servant de la commande `subset`. Si vous ne savez pas vous servir de la commande `subset`, alors consultez l'aide en ligne. N'oubliez pas de donner un nom à votre jeu de données qui provient de l'extraction.

b) Calculer une estimation de la proportion des gousses de glycine blanche qui ont moins de trois graines présentes dans une gousse en vous servant de la commande `cumsum` et `table`.

c) Conclure.

Exercice 4 La masse d'un œeuf.
Dans un centre avicole, des études antérieures ont montré que la masse d'un œuf choisi au hasard peut être considérée comme la réalisation d'une variable aléatoire normale X, de moyenne μ et de variance σ^2.
Nous admettons que les masses des œufs sont indépendantes les unes des autres. Nous prenons un échantillon de $n = 36$ œufs que nous pesons. Les mesures sont données (par ordre croissant) dans le tableau suivant :

$50,34$	$52,62$	$53,79$	$54,99$	$55,82$	$57,67$
$51,41$	$53,13$	$53,89$	$55,04$	$55,91$	$57,99$
$51,51$	$53,28$	$54,63$	$55,12$	$55,95$	$58,10$
$52,07$	$53,30$	$54,76$	$55,24$	$57,05$	$59,30$
$52,22$	$53,32$	$54,78$	$55,28$	$57,18$	$60,58$
$52,38$	$53,39$	$54,93$	$55,56$	$57,31$	$63,15$

a) Tracer la boîte à moustaches et l'histogramme correspondant à cette série de mesures.

b) Calculer la moyenne empirique et l'écart-type empirique corrigé de cette série statistique.

c) Donner une estimation des paramètres μ et σ^2.

Exercice 5 Les fêtes de fin d'année et les kilos en trop. D'après un examen.
Monsieur A vient de passer les fêtes de fin d'année et souhaite se peser pour voir son poids. Pour cela, Monsieur A dispose de trois balances :

a) La première balance fournit un poids, noté X, sans biais mais avec une erreur de mesure de variance 4 kg^2.

b) La deuxième balance a été récemment déréglée, et elle fournit un résultat biaisé, noté Y : elle surévalue le vrai poids de B kg en moyenne. La variance de son résultat est de 1 kg^2.

c) La troisième balance fournit un résultat sans biais, noté Z, de variance 10 kg^2.

Monsieur A se pèse une fois sur chaque balance. Nous supposons que les trois balances donnent des résultats indépendants.
Soit P le vrai poids de Monsieur A.
Nous notons $EQM(T)$ l'erreur quadratique moyenne obtenue avec un estimateur T, et nous dirons qu'un estimateur F est meilleur qu'un estimateur G si

$$EQM(F) < EQM(G).$$

a) Quelle(s) est (sont) la (les) proposition(s) ou expression(s) exacte(s) ?
   (i) $EQM(X) = 16$
   (ii) $EQM(Y) = B^2 + 1$
   (iii) $EQM(Z) = 100$
   (iv) X est meilleur que Z pour estimer P, le vrai poids de Monsieur A.
   (v) Si $B < 1,5$ alors Y est meilleur que X.

b) Nous supposons pour la suite de l'exercice que $B = 2$.
   Nous construisons de nouveaux estimateurs :
   (i) $Y' = Y - 2$
   (ii) $T_1 = 0,5\,X + 0,5\,Y'$
   (iii) $T_2 = 0,4\,X + 0,4\,Y' + 0,2Z$.

   Si nous ordonnons les estimateurs du meilleur au moins bon, quel est le bon ordre pour ces estimateurs ?

c) Nous cherchons à construire le meilleur estimateur possible avec X et Y', c'est-à-dire celui qui aura la plus petite erreur quadratique moyenne, parmi les estimateurs de la forme :

$$T_3 = aX + (1-a)Y',$$

   avec $0 < a < 1$. Est-ce que T_3 est un estimateur sans biais de P ? Que vaut l'EQM de T_3 ?

d) Pour trouver T_3, nous allons chercher le minimum de $EQM(T_3)$ en fonction de a, en dérivant $EQM(T_3)$ par rapport à a.
   Soit $\tilde{a}$ la valeur de a telle que $EQM'(T_3) = 0$. En se pesant avec les deux premières balances, Monsieur A a observé les poids suivants :

$$x = 85kg, \quad y = 82kg.$$

   Pour ces deux valeurs, que valent $\tilde{a}$ et t_3 ?