Sorbonne Université Année 2020/2021

L3 Deuxième semestre

Statistique inférentielle
TD

Antoine Godichon-Baggioni

TD1 : Rappels de probabilités

Exercice 1 :

1. Rappeler les définitions de la convergence en loi, en probabilité, presque sûre et en moyenne quadratique .
2. Montrer que la convergence en moyenne quadratique implique la convergence en probabilité.
3. Soit a une constante et (X_n) une suite de variables aléatoires. Montrer que si (X_n) converge en loi vers a, alors (X_n) converge en probabilité vers a.
4. Soit (X_n) une suite de variables aléatoires convergeant en loi vers une constante a et soit $h : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction continue. Montrer que $h(X_n)$ converge en probabilité vers $h(a)$.
5. Etudier la convergence de la suite (X_n) dans chacun des cas suivants :
   — $X_n = 1/n$.
   — $X_n = (-1)^n$.
   — $X_n = \mathbf{1}_{A_n}$ où A_n est une suite d'évènements et $\mathbb{P}[A_n]$ converge vers 0.
   — $X_n = Z_n \mathbf{1}_{B_n}$ où Z_n converge en loi vers une variable aléatoire Z et $\mathbb{P}[B_n]$ converge vers 1.

Exercice 2 : Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées, à valeurs dans un ensemble A. Soit $D \subset A$ tel que $p = \mathbb{P}[X_1 \in D] \neq 0$.

Pour tout $n \geq 1$, on pose $S_n = \sum_{j=1}^{n} \mathbf{1}_{\{X_j \in D\}}$.

1. Calculer $\mathbb{E}[S_n]$ et $\mathbb{V}[S_n]$.
2. Montrer que la suite $\frac{S_n}{n}$ converge presque sûrement vers p.
3. Calculer l'erreur quadratique moyenne $\mathbb{E}\left[\left(\frac{S_n}{n} - p\right)^2\right]$. En déduire une majoration uniforme en p de l'erreur quadratique moyenne.
4. Démontrer que pour tout $p \in]0,1[$ et pour tout $\varepsilon > 0$,
$$\mathbb{P}\left[\left|\frac{S_n}{n} - p\right| \geq \epsilon\right] \leq \frac{1}{4n\epsilon^2}.$$
5. Enoncer le théorème de limite centrale que satisfait la variable $\dfrac{S_n}{n}$.

Exercice 3 : Soit (X_n) une suite de variables aléatoires centrées, de même variance σ^2 et satisfaisant pour tout entiers $i \neq j$
$$\mathrm{Cov}\left(X_i, X_j\right) = \sigma^2 \alpha^{|j-i|}$$
avec $\alpha \in (0,1)$. Pour tout $n \geq 1$, on pose $S_n = \sum_{i=1}^{n} X_i$.

1. Calculer $\mathbb{E}[S_n]$.

2. Montrer que
$$\mathbb{V}\left[S_n\right] = n\sigma^2 + \frac{2\alpha\sigma^2}{1-\alpha}\left((n-1) - \alpha\frac{1-\alpha^{n-1}}{1-\alpha}\right).$$
3. En déduire la convergence en moyenne quadratique de S_n/n.

Exercice 4 : On considère une suite de variables aléatoires (X_n). Dans chacun des cas suivants, donner la normalité asymptotique de $g\left(X_n\right)$.

1. $g : x \longmapsto \sqrt{x}, \theta > 0$ et
$$\sqrt{n}\left(X_n - \theta^2\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,1\right).$$
2. $g : x \longmapsto x^{-1}, \theta \neq 0$ et
$$\sqrt{n}\left(X_n - \frac{1}{\theta}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\theta^2}\right).$$
3. $g : x \longleftarrow e^x, \theta > 0$ et
$$\sqrt{n}\left(X_n - \ln(\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, (\ln\theta)^2\right)$$

Exercice 5 : Soit X une variable aléatoire suivant une loi uniforme sur $[0,1]$.

1. Donner la loi de $Z = -\log(X)$.
2. Soit $Z_1, ..., Z_n$ des variables aléatoires indépendantes et identiquement distribuées de même loi que Z. Donner la normalité asymptotique de $\overline{Z}_n$.
3. En déduire la normalité asymptotique de
$$Y_n = \frac{1}{\left(\prod_{i=1}^n X_i\right)^{1/n}}.$$

Exercice 6 (Loi exponentielle translatée). Soit Y une variable aléatoire suivant une loi exponentielle de paramètre 1, i.e de densité f définie pour tout $x \in \mathbb{R}$ par
$$f(x) = \exp\left(-x\right)\mathbf{1}_{\mathbb{R}_+^*}(x)$$
Soit θ, on considère la variable aléatoire $X = Y + \theta$ de densité
$$f_\theta(x) = \exp\left(-(x-\theta)\right)\mathbf{1}_{[\theta,+\infty[}(x).$$

1. Donner les fonctions de répartitions des variables X et Y.
2. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi que X. On considère la variable aléatoire $Z_n = \min_{i=1,\ldots,n} X_i$. Donner la fonction de répartition de Z_n.
3. En déduire que Z_n converge en probabilité vers θ.
4. Montrer que la variable $n\left(Z_n - \theta\right)$ suit une loi exponentielle de paramètre 1.

Exercice 7 : (inégalité de Hölder). L'objectif de cette exercice est de démontrer l'inégalité de Hölder "généralisée" suivante. Soient $p, q, r > 0$ tels que $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, et X, Y deux variables aléatoires admettant respectivement un moment d'ordre p et q. Alors

$$\left(\mathbb{E}\left[|XY|^r\right]\right)^{\frac{1}{r}} \leq \left(\mathbb{E}\left[|X|^p\right]\right)^{\frac{1}{p}} \left(\mathbb{E}\left[|Y|^q\right]\right)^{\frac{1}{q}}.$$

1. Montrer que pour tout $a, b \geq 0$

$$\frac{1}{r}(ab)^r \leq \frac{1}{p}a^p + \frac{1}{q}b^q.$$

2. En déduire l'inégalité de Hölder.
3. Soit (X_n) (Y_n) deux suites de variables aléatoires convergeant vers 0 respectivement à l'ordre $p > 2$ et $\frac{2p}{p-2}$, i.e

$$\mathbb{E}\left[|X_n|^p\right] \xrightarrow[n\to+\infty]{} 0 \qquad \text{et} \qquad \mathbb{E}\left[|Y_n|^{\frac{2p}{p-2}}\right] \xrightarrow[n\to+\infty]{} 0.$$

   Montrer que $X_n Y_n$ converge en moyenne quadratique vers 0.

TD2 : Estimation

Exercice 1 : estimation de la moyenne et de la variance. Soit X une variable aléatoire de moyenne μ et de variance σ^2 inconnues. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi que X.

1. Rappeler l'estimateur de la moyenne. Montrer qu'il est sans biais, fortement consistant et donner son erreur quadratique moyenne ainsi que sa normalité asymptotique.

2. On souhaite maintenant estimer σ^2. On propose l'estimateur suivant :

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2.$$

   Expliquer ce choix.

3. Montrer que

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \left(\overline{X}_n - \mu\right)^2.$$

4. Soit $\tau^4 = \mathbb{E}\left[(X-\mu)^4\right]$. A l'aide du Théorème de Slutsky, donner la normalité asymptotique de $\hat{\sigma}_n^2$.

5. Calculer $\mathbb{E}\left[\hat{\sigma}_n^2\right]$. L'estimateur $\hat{\sigma}_n$ est-il sans biais ?

6. En déduire un estimateur sans biais de σ^2 et donner sa normalité asymptotique.

Exercice 2 : estimation de la covariance. Soit (X, Y) un couple de variables aléatoires réelles d'espérances respectives μ_X, μ_Y et de variances respectives σ_X^2, σ_Y^2. On s'intéresse ici à l'estimation de la covariance C de X, Y, définie par

$$C = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Soient $(X_1, Y_1), \ldots, (X_n, Y_n)$ des couples de variables aléatoires indépendants et de même loi que (X, Y). On s'intéresse à l'estimateur C_n défini par

$$C_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)\left(Y_i - \overline{Y}_n\right).$$

1. Justifier la proposition de cet estimateur.

2. Montrer que

$$\sum_{j=1}^{n}\left(X_j - \overline{X}_n\right)\left(Y_j - \overline{Y}_n\right) = \sum_{j=1}^{n}\left(X_j - \mu_X\right)\left(Y_j - \mu_Y\right) - n\left(\overline{X}_n - \mu_X\right)\left(\overline{Y}_n - \mu_Y\right).$$

3. Montrer que

$$n^2 \left(\overline{X}_n - \mu_X\right) \left(\overline{Y}_n - \mu_Y\right) = \sum_{j=1}^{n} \left(X_j - \mu_X\right) \left(Y_j - \mu_Y\right) + \sum_{i \neq j} \left(X_i - \mu_X\right) \left(Y_j - \mu_Y\right).$$

4. Calculer $\mathbb{E}\left[C_n\right]$. Que pouvez vous en déduire ?
5. Proposer un estimateur sans biais de C.
6. En posant $Z_j = \left(X_j - \mu_X\right) \left(Y_j - \mu_Y\right)$, montrer que C_n converge en probabilité vers C.
7. Montrer, soit à l'aide des inégalités de Markov et de Cauchy-Schwarz, soit à l'aide du théorème de Slutsky, que
$$\sqrt{n} \left(\overline{X}_n - \mu_X\right) \left(\overline{Y}_n - \mu_Y\right) \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$
8. Soit $\tau^4 = \mathbb{E}\left[\left(X - \mu_X\right)^2 \left(Y - \mu_Y\right)^2\right] < +\infty$. Donner la normalité asymptotique de C_n, et en déduire celle de l'estimateur sans biais.

Exercice 3 : méthode des moments. Soit $\theta \in \Theta$ où Θ est un ouvert de $\mathbb{R}$, et $\varphi : \Theta \longrightarrow \varphi(\Theta)$ un C^1-difféomorphisme. Soit $k \in \mathbb{N}^*$ tel que

$$\mathbb{E}\left[X^k\right] = \varphi(\theta).$$

De plus, on suppose que X admet un moment d'ordre $2k$.

1. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et identiquement distribuées de même loi que X. Proposer un estimateur de $\varphi(\theta)$.
2. Est-il consistant ? Asymptotiquement normal ?
3. En déduire un estimateur de θ.
4. Est-il consistant ?
5. On suppose que $\varphi'(\theta) \neq 0$. En déduire la normalité asymptotique de l'estimateur de θ.

Exercice 4 : (loi géométrique). Soit X une variable aléatoire suivant une loi géométrique de paramètre p, i.e pour tout entier $k \geq 1$, $\mathbb{P}\left[X = k\right] = (1-p)^{k-1} p$.

1. Rappeler l'espérance et la variance de X.
2. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi que X. Par la méthode des moments, donner un estimateur $\hat{p}_n$ de p.
3. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?
4. Donner l'estimateur $\hat{p}_n^{MV}$ du maximum de vraisemblance de p. Que pouvez vous en conclure ?

Exercice 5 : (loi exponentielle). Soit X une variable aléatoire suivant une loi exponentielle de paramètre $\theta > 0$. On rappelle que la densité de X est définie pour tout $x \in \mathbb{R}$ par

$$f(x) = \theta \exp\left(-x\theta\right) \mathbf{1}_{[0,+\infty[}(x).$$

1. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.
2. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi que X. A l'aide de la méthode des moments, proposer un estimateur $\hat{\theta}_n$ de θ.
3. Est-il consistant ? Fortement consistant ?
4. Donner sa normalité asymptotique.
5. Donner l'estimateur (si il existe) du maximum de vraisemblance $\hat{\theta}_n^{MV}$ de θ.
6. Que pouvez vous en conclure ?

Exercice 6 : (loi de Rayleigh). Soit θ un entier positif. On considère une variable aléatoire X suivant une loi de Rayleigh de paramètre θ, i.e de f_θ définie pour tout $x \in \mathbb{R}$ par $f_\theta(x) = \lambda x \exp\left(-\frac{x^2}{\theta}\right) \mathbf{1}_{\mathbb{R}_+}(x)$.

1. Calculer λ.
2. Calculer l'espérance et la variance de X.
3. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi que X. En déduire un estimateur $\hat{\theta}_n$ de θ.
4. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?
5. Donner l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{MV}$ de θ.
6. Donner la normalité asymptotique de cet estimateur.
7. Que pouvez-vous en conclure ?

Exercice 7 : (loi de Poisson). On considère une variable aléatoire X suivant une loi de Poisson de paramètre θ.

1. Soit $(X_1, \ldots, X_n)$ i.i.d de même loi que X. A l'aide de la méthode des moments, proposer un estimateur de θ.
2. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?
3. Donner l'estimateur du maximum de vraisembance.
4. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?

Exercice 8 (loi exponentielle translatée). Soit Y une variable aléatoire suivant une loi exponentielle de paramètre 1, i.e de densité f définie pour tout $x \in \mathbb{R}$ par

$$f(x) = \exp(-x) \mathbf{1}_{\mathbb{R}_+}(x)$$

Soit θ, on considère la variable aléatoire $X = Y + \theta$ de densité

$$f_\theta(x) = \exp(-(x-\theta)) \mathbf{1}_{[\theta,+\infty[}(x).$$

1. Calculer $\mathbb{E}[Y]$ et en déduire $\mathbb{E}[X]$.

2. Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi que X. En déduire un estimateur $\hat{\theta}_n$ de θ. Cet estimateur est-il consistant ? Fortement consistant ? Sans biais ?
3. Calculer $\mathbb{V}[Y]$ et en déduire $\mathbb{V}[X]$.
4. Donner la normalité asymptotique de $\hat{\theta}_n$.
5. Donner l'erreur quadratique moyenne de $\hat{\theta}_n$.
6. Donner l'estimateur $\hat{\theta}_n^{MV}$ du maximum de vraisemblance.
7. Soit $Z_n = \min_{i=1,\ldots,n} X_i$. Rappeler la loi de $n\,(Z_n - \theta)$.
8. Quel estimateur choisiriez vous ?
9. Calculer sonr erreur quadratique moyenne.

Exercice 9 : (loi uniforme dilatée). Soit $\theta > 0$. On considère une variable aléatoire X suivant une loi uniforme sur $[0, \theta]$. Soit $x_1, ..., x_n$ des réalisation des variables aléatoires indépendantes $X_1, ..., X_n$ et de même loi que X.

1. Par la méthode des moments, proposer un estimateur convergent de θ et donner sa convergence.
2. Cet estimateur est-il sans biais ?
3. Donner son erreur quadratique moyenne.
4. Donner sa normalité asymptotique.
5. Donner l'estimateur du maximum de vraisemblance de θ.
6. Calculer la fonction de répartition de X.
7. On considère maintenant $X_{(n)} = \max_{i=1,\ldots,n} X_i$. Donner sa fonction de répartition.
8. Montrer que $X_{(n)}$ converge en probabilité vers θ.
9. A l'aide du lemme de Borel-Cantelli, en déduire la forte consistance de $X_{(n)}$.
10. Donner la fonction de répartition d'une loi exponentielle de paramètre θ^{-1}.
11. Montrer que $n\left(\theta - X_{(n)}\right)$ converge en loi vers une loi exponentielle de paramètre θ^{-1}.
12. Quel estimateur de θ choisiriez vous ?

Exercice 10 : (loi normale). Soit X une variable aléatoire suivant une loi normale de moyenne μ et de variance σ^2. Soit $x_1, \ldots, x_n$ des réalisations des variables aléatoires indépendantes $X_1, \ldots, X_n$ de même loi que X.

1. Ecrire la vraisemblance
2. En déduire les estimateur du maximums de vraisemblance de μ et σ^2.
3. Commenter.

TD3 : Intervalles de confiance

Exercice 1 : Dans un centre avicole, des études antérieures ont montré que le poids d'un oeuf choisi au hasard peut être considéré comme la réalisation d'une variable aléatoire gaussienne X, d'espérance μ et de variance σ^2. On admet que les poids des oeufs sont indépendants les uns des autres. On prend un échantillon de $n = 36$ oeufs que l'on pèse. Les mesures obtenues (exprimées en g) 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

TABLE 1 – Mesure des poids des oeufs (en g).

1. La figure ci-dessous présente l'histogramme en fréquences des poids des oeufs sur lequel on a superposé une version lissée de l'histogramme. Quelles conclusions peut-on tirer de cet histogramme sur la distribution des oeufs ?

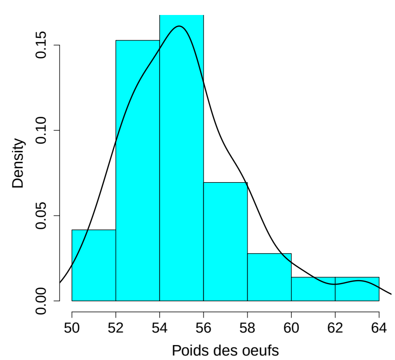

FIGURE 1 – Histogramme en fréquence des poids des oeufs.

2. Donner une estimation de la moyenne μ et de la variance σ^2. On les notera m et s^2.

   Pour vous aider, on fournit les informations suivantes : si $x_1, x_2, \ldots, x_{36}$ désignent les poids mesurés, alors
$$\sum_{j=1}^{36} x_j = 1982,99 \quad \text{et} \quad \sum_{j=1}^{36} x_j^2 = 109481,1$$

3. Construire un intervalle de confiance au niveau 95% pour le poids moyen des oeufs. Commenter le résultat obtenu. Pour s'aider, soit $Z \sim \mathcal{N}(0,1)$, T_{35} et T_{36} suivant des loi de

Student à 35 et 36 degrés de liberté. On dispose des résultats suivants :

$$\mathbb{P}\left[Z \leq 1.64\right] = 0.95, \quad \mathbb{P}\left[Z \leq 1.96\right] = 0.975, \quad \mathbb{P}\left[Z \leq 2.58\right] = 0.995$$
$$\mathbb{P}\left[T_{36} \leq 1.69\right] = 0.95 \quad \mathbb{P}\left[T_{36} \leq 2.03\right] = 0.975 \quad \mathbb{P}\left[T_{36} \leq 2.72\right] = 0.995$$
$$\mathbb{P}\left[T_{35} \leq 1.69\right] = 0.95 \quad \mathbb{P}\left[T_{35} \leq 2.03\right] = 0.975 \quad \mathbb{P}\left[T_{35} \leq 2.72\right] = 0.995$$

4. Construire un intervalle de confiance au niveau 95% pour la variance. Commenter. Pour s'aider, soit χ_{35} et χ_{36} des variables suivants des loi du Khi-deux à 35 et 36 degrés de liberté. On dispose des résultats suivants :

$$\mathbb{P}\left[\chi_{36} \leq 51\right] = 0.95 \quad \mathbb{P}\left[\chi_{36} \leq 54.4\right] = 0.975 \quad \mathbb{P}\left[\chi_{36} \leq 23.3\right] = 0.05 \quad \mathbb{P}\left[\chi_{36} \leq 21.3\right] = 0.025$$
$$\mathbb{P}\left[\chi_{35} \leq 49.8\right] = 0.95 \quad \mathbb{P}\left[\chi_{35} \leq 53.2\right] = 0.975 \quad \mathbb{P}\left[\chi_{35} \leq 22.5\right] = 0.05 \quad \mathbb{P}\left[\chi_{35} \leq 20.6\right] = 0.025$$

5. A quel niveau de confiance correspondrait un intervalle centré en m et de demi-longueur 0.76 ?

Exercice 2 : On considère une variable aléatoire X dont la loi dépend d'un paramètre inconnu $p > 0$ et telle que $\mathbb{E}\left[X^2\right] = \frac{1}{p^2}$ et $\mathbb{E}\left[X^4\right] = \frac{1}{p^6} + \frac{1}{p^4}$, et on pose $\theta = p^{-2}$.

1. Proposer un estimateur $\hat{\theta}_n$ de θ. Est-il consistant ? Fortement consistant ?
2. Est-il sans biais ?
3. Donner son erreur quadratique moyenne.
4. Donner sa normalité asymptotique.
5. En déduire un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ.
6. En déduire un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour p.
7. Proposer un estimateur $\hat{p}_n$ de p et montrer sa forte consistance.
8. Donner sa normalité asymptotique.
9. En déduire un nouvel intervalle de confiance asymptotique de niveau $1 - \alpha$ pour p.

Exercice 3 : Soit $\theta > 0$ et Y une variable aléatoire suivant une loi exponentielle de paramètre θ^{-1}. Soit $X = Y + \theta$, sa densité f_θ est définie pour tout x par

$$f_\theta(x) = C_\theta \exp\left(-\frac{x-\theta}{\theta}\right) \mathbf{1}_{[\theta,+\infty[}(x).$$

Dans ce qui suit, on considère $x_1, \ldots, x_n$ qui sont des réalisations des variables aléatoires indépendantes $X_1, \ldots, X_n$ de même loi que X.

1. Que vaut C_θ ?
2. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.

3. Par la méthode des moments, en déduire un estimateur de θ.
4. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?
5. Soit $\alpha \in (0,1)$, construire un intervalle de confiance asymptotique de niveau $1-\alpha$ du paramètre θ.
6. Donner l'estimateur du maximum de vraisemblance.
7. On considère maintenant $X_{(1)} = \min_{i=1,\dots,n}(X_i)$. Donner la fonction de répartition de X. En déduire celle de $X_{(1)}$.
8. Donner la loi de $n\left(X_{(1)} - \theta\right)$.
9. En déduire la convergence en probabilité de $X_{(1)}$.
10. Montrer que l'estimateur $X_{(1)}$ est biaisé mais asymptotiquement sans biais.
11. En déduire la convergence en moyenne quadratique de $X_{(1)}$.
12. Donner un nouvel intervalle de confiance pour le paramètre θ.
13. Commenter

Exercice 4 : Soit X une variable aléatoire de densité $f_\theta(x) = \frac{2}{\theta^2} x \mathbf{1}_{[0,\theta]}(x)$, avec $\theta > 0$. On considère dans ce qui suit des réalisations $x_1, \dots, x_n$ des variables aléatoires indépendantes $X_1, \dots, X_n$ de même loi que X.

1. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.
2. Par la méthode des moments, en déduire un estimateur $\hat{\theta}_n$ de θ.
3. Est-il consistant ? Fortement consistant ? Asymptotiquement normal ?
4. Soit $\alpha \in (0,1)$, construire un intervalle de confiance asymptotique de niveau $1-\alpha$ du paramètre θ.
5. Donner l'estimateur du maximum de vraisemblance.
6. On considère maintenant $X_{(n)} = \max_{i=1,\dots,n} X_i$. Donner la fonction de répartition de X.
7. En déduire la fonction de répartition de $X_{(n)}$.
8. En déduire la densité de $X_{(n)}$.
9. L'estimateur $X_{(n)}$ est-il sans biais ?
10. Calculer le risque quadratique $\mathbb{E}\left[\left(X_{(n)} - \theta\right)^2\right]$.
11. Donner la convergence en loi de $n\left(\theta - X_{(n)}\right)$.
12. Donner un nouvel intervalle de confiance au niveau $1-\alpha$ de θ.
13. Quel intervalle de confiance choisiriez vous ?

Exercice 5 : Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes suivant une loi de Poisson de paramètre $\theta > 0$.

1. Donner $\mathbb{E}[X_1]$ et $\mathbb{V}[X_1]$. En déduire un estimateur $\hat{\theta}_n$ de θ.
2. Donner la normalité asymptotique de l'estimateur $\hat{\theta}_n$.
3. Par la méthode du plug-in, donner un intervalle de confiance asymptotique à 95% pour θ.
4. Trouver une fonction $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ telle que

$$\sqrt{n}\left(g\left(\hat{\theta}_n\right)\right) - g(\theta)) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

5. En déduire un nouvel intervalle de confiance pour θ.
6. Vérifier que les deux résultats sont équivalents, i.e si on note $[a_n, b_n]$ et $[a'_n, b'_n]$ les deux intervalles obtenus, on a

$$\frac{a_n}{a'_n} \xrightarrow[n\to+\infty]{\mathbb{P}} 1 \qquad \text{et} \qquad \frac{b_n}{b'_n} \xrightarrow[n\to+\infty]{\mathbb{P}} 1.$$

Exercice 6 : Loi de Laplace translatée. Soit Y une variable aléatoire suivant une loi de Laplace, i.e de densité f_Y définie pour tout $x \in \mathbb{R}$ par

$$f_Y(x) = \frac{1}{2}\exp\left(-|x|\right).$$

Soit $\theta > 0$ et X une variable aléatoire suivant une loi de Laplace translatée, i.e de densité f_X définie pour tout $x \in \mathbb{R}$ par

$$f_X(x) = \frac{1}{2}\exp\left(-|\theta - x|\right).$$

Dans ce qui suit on considère des variables aléatoires indépendantes $X_1, \ldots, X_n$ de même loi que X.

1. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.
2. Par la méthode des moments , donner un estimateur de θ. Est-il consistant ? Asymptotiquement normal ? Que pouvez vous en conclure ?
3. Construire un intervalle de confiance asymptotique de niveau 0.90 pour θ.
4. Que vaut l'estimateur du maximum de vraisemblance ?
5. On note $\hat{m}_n$ l'estimateur du maximum de vraisemblance, on suppose qu'il existe et est consistant. Que pouvez-vous en conclure ?

Exercice 7 : Débiaiser ou ne pas débiaiser. On considère une variable aléatoire $X \sim \mathcal{U}([0,\theta])$ avec $\theta > 0$ et on considère l'estimateur $\hat{\theta}_n = X_{(n)}$.

1. Est-il biaisé ? Calculer l'erreur quadratique moyenne.
2. Proposer un estimateur non biaisé et calculer son erreur quadratique moyenne.

3. On considère l'estimateur $\hat{\theta}_{\alpha,n} = \alpha X_{(n)}$. Calculer l'erreur quadratique moyenne.
4. Choisir α afin de minimiser l'erreur quadratique moyenne.
5. Conclure.

Exercice 8 : J'existe ? On considère une variable aléatoire X suivant une loi binomiale $X \sim \mathcal{B}\left(p, \lambda^{-1}\right)$ avec $\lambda > 1$ et $\hat{\theta}(X)$ un estimateur de λ.

1. Calculer $\mathbb{E}\left[\hat{\theta}(X)\right]$.
2. Sachant qu'un polynôme non nul de degré p admet au plus p racines, qu'observez-vous si $\hat{\theta}(X)$ est non biaisé ?

Exercice 9 : Estimation de deux paramètres. Soit $Y \sim \mathcal{E}(\lambda)$ avec $\lambda > 0$ inconnu. Soit $X = Y + \theta$ avec $\theta > 0$ inconnu. On admettra que X a pour densité $f_{\theta,\lambda}$ définie pour tout x par

$$f_{\theta,\lambda}(x) = \lambda \exp\left(-\lambda(x-\theta)\right) \mathbf{1}_{[\theta,+\infty[}(x).$$

Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d de même loi que X.

1. Calculer $\mathbb{E}[X]$ et $\mathbb{V}[X]$.
2. Déterminer l'estimateur du maximum de vraisemblance de θ.
3. Calculer la fonction de répartition de $X_{(1)}$.
4. En déduire l'erreur quadratique moyenne de $X_{(1)}$.
5. Déduire de la question 5
$$\sqrt{n}\left(X_{(1)} - \theta\right) \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$
Que pouvez-vous en déduire ?
6. En déduire un estimateur de λ.
7. Montrer que l'estimateur de λ est consistant. Pour s'aider, on admettra que si (A_n) et (B_n) sont deux suites de variables aléatoires convergeant en probabilités vers a et b, et $g : I \times J \longrightarrow \mathbb{R}$ est une fonction continue en (a,b), alors
$$g\left(A_n, B_n\right) \xrightarrow[n\to+\infty]{\mathbb{P}} g(a,b).$$
avec I, J des intervalles ouvert de $\mathbb{R}$.
8. Montrer que
$$\sqrt{n}\left(\overline{X}_n - \hat{\theta}_n - \frac{1}{\lambda}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right).$$
9. En déduire la normalité asymptotique de l'estimateur de λ.
10. Soit $\alpha \in (0,1)$, en déduire un intervalle de confiance asymptotique de λ.

11. Montrer que
$$n\left(X_{(1)} - \theta\right) \sim \mathcal{E}(\lambda).$$
12. Pour tout $\alpha \in (0,1)$, donner le quantile $q_{\lambda,1-\alpha}$ d'ordre $1-\alpha$ de la loi exponentielle de paramètre λ.
13. Donner la convergence de
$$n\left(X_{(1)} - \theta\right) - \left(q_{\hat{\lambda}_n,1-\alpha} - q_{\lambda,1-\alpha}\right).$$
avec $q_{\hat{\lambda}_n} = \frac{-\ln(\alpha)}{\hat{\lambda}_n}$.
14. Déterminer un intervalle de confiance asymptotique de niveau $1-\alpha$ pour θ.

Exercice 10 : On considère une urne dans laquelle il y a m_1 boules rouges et m_2 boules noires. Le nombre de boules de chaque couleur est inconnu, et le nombre total de boules est bien trop conséquent pour que l'on s'amuse à les compter. L'objectif est donc de proposer différentes stratégies pour estimer la proportion $p = \frac{m_1}{m_1+m_2}$ de boules rouges.

1. On propose d'effectuer N tirages avec remise et on note $X_i = 1$ si le i-ème tirage est une boule rouge et $X_i = 0$ sinon.
   (a) Proposer un estimateur de p et donner son erreur quadratique moyenne.
   (b) Donner sa normalité asymptotique et en déduire un intervalle de confiance asymptotique de niveau $1-\alpha$ de p.
2. On propose d'effectuer N tirages sans remise et on note Y le nombre de boules rouges tirées ($N << m_1 + m_2$).
   (a) Quelle est la loi de Y ? On admettra que $\mathbb{E}[Y] = Np$ et $\mathbb{V}[Y] = Np(1-p)\frac{m_1+m_2-N}{m_1+m_2-1}$.
   (b) Proposer un nouvel estimateur de p et donner son erreur quadratique moyenne. Quel estimateur choisiriez-vous ?
   (c) Donner un intervalle de confiance de niveau au moins $1-\alpha$ de p.
3. On propose d'effectuer la même expérience n fois, mais en n'effectuant que $K = \frac{N}{n}$ (on suppose K entier) tirages sans remise et on note Y_i le nombre de boules rouges tirées à la i-ème expérience.
   (a) Proposer un nouvel estimateur de p et donner son erreur quadratique moyenne. Quel estimateur choisiriez-vous ?
   (b) Donner sa normalité asymptotique et en déduire un nouvelle intervalle de confiance.
4. On a $m_1 = 2500$, $m_2 = 7500$, $N = 1000$, $K = 10$ et on obtient des intervalles de confiance (dans l'ordre) de taille 0.054, 0.14, 0.053. Comment interpréter ces résultats ?

TD4 : Théorème de Cochran et modèle linéaire

Exercice : Le principe de la régression linéaire est de modéliser une variable y à partir de variables explicatives $\mathbf{x} = \left(x_1, \ldots, x_p\right)^T$, i.e de considérer

$$y = \beta_1 x_1 + \ldots + \beta_p x_p,$$

où $\beta = \left(x_1, \ldots, x_p\right)$ est inconnu. En pratique, on dispose d'un échantillon $\left(\mathbf{x}_1, y_1\right), \ldots, \left(\mathbf{x}_n, y_n\right)$, mais on obtient jamais réellement une droite (erreurs de mesures...). On va donc considérer le modèle linéaire

$$y = \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

avec $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$. On parle alors de modèle linéaire gaussien. On suppose maintenant que les données suivent le modèle suivant :

$$Y_i = \beta_1 x_{i,1} + \ldots + \beta_p x_{i,p} + \epsilon_i,$$

avec

- Y_i est une variable aléatoire et on observe les réalisations y_i.
- Les $\mathbf{x}_i = \left(x_{i,1}, \ldots, x_{i,p}\right)^T$ sont déterministes.
- Le paramètre $\beta = \left(\beta_1, \ldots, \beta_p\right)^T$ est inconnu et déterministe.
- Les ϵ_i sont i.i.d et $\epsilon_1 \sim \mathcal{N}\left(0, \sigma^2\right)$.

1. Vérifier que le modèle peut s'écrire comme

$$Y = X\beta + \epsilon,$$

avec $Y = \left(Y_1, \ldots, Y_n\right)^T$, $\epsilon = \left(\epsilon_1, \ldots, \epsilon_n\right)^T$ et

$$X = \begin{pmatrix} x_{1,1} & \ldots & x_{1,p} \\ \vdots & \vdots & \vdots \\ x_{n,1} & \vdots & x_{n,p} \end{pmatrix}$$

2. Donner la loi de ϵ et en déduire la loi de Y. Quelle est la loi de Y_i ?

3. On considère à partir de maintenant que $\operatorname{rang}(X) = p$, et on note $D = \operatorname{Im}(X)$. On s'intéresse à l'estimateur des moindres carrés défini par

$$\hat{\beta} = \arg\min_{h \in \mathbb{R}^p} \|Y - Xh\|^2$$

Montrer que la matrice $X^T X$ est symétrique et définie positive. On rappellera qu'une matrice symétrique M est définie positive si pour tout $h \in \mathbb{R}^p \backslash \{0\}$,

$$h^T M h > 0$$

4. On note $G(h) = \|Y - Xh\|^2$. Calculer le gradient et la Hessienne de G et en déduire $\hat{\beta}$.
5. Soit $\mathbb{R}^d = E \oplus_{\perp} F$. Soit P une matrice $p \times p$. On rappelle que P est le projecteur orthogonal sur E (parallèlement à F) si
   — P est symétrique.
   — $P^2 = P$.
   — Pour tout $h \in E$, $P(h) = h$.
   — Pour tout $h \in F$, $P(h) = 0$.
   Montrer que $P_D = X \left(X^T X\right)^{-1} X^T$ est le projecteur orthogonal sur D parallèlement à $D^{\perp}$.
6. Que pouvez vous en déduire sur $X\hat{\beta}$?
7. Donner la loi de $X\hat{\beta}$ et en déduire celle de $\hat{\beta}$.
8. On suppose σ^2 connu. Soit $x_0 \in \mathbb{R}^p \backslash \{0\}$, donner un intervalle de confiance de niveau au moins $1 - \alpha$ de $x_0^T \beta$. Que se passe-t-il si σ^2 est inconnu ?
9. On suppose maintenant que σ^2 est inconnu et on considère l'estimateur

$$\hat{\sigma}^2 = \frac{1}{n-p} \left\|Y - X\hat{\beta}\right\|^2$$

   (a) Expliquer ce choix d'estimateur.
   (b) Exprimer $\hat{\sigma}^2$ à l'aide de projections.
   (c) Enoncer le théorème de Cochran dans ce cas.
   (d) En déduire un intervalle de confiance pour $x_0^T \beta$.

TD5 : Tests

Exercice 1 : Dans les années 70, les athlètes féminines de RDA étaient réputées pour leur forte corpulence et soupçonnées par le comité éthique olympique de dopages via la prise de substances hormonales virilisantes (dites androgènes). Des mesures ont été effectuées sur la quantité d'androgènes par litre de sang chez 9 athlètes, et on obtient les résultats suivants :

3.22 3.07 3.17 2.91 3.40 3.58 3.23 3.11 3.62

On veut tester l'hypothèse nulle "les athlètes de RDA ne sont pas dopées", sachant que chez une femme "lambda", le quantité moyenne d'androgènes est de 3.1

1. Quel test faut-il effectuer ?
2. Quels sont les hypothèses à vérifier ?
3. Creer un vecteur data comprenant toutes les données.
4. Rentrer la commande suivante et commenter :

   ```
   hist(data)
   ```

5. Faire le test au risque de 5%. Pour cela on pourra s'aider, i.e rentrer

   ```
   help(t.test)
   ```

6. Pourquoi peut-on remettre en question le protocole expérimental et donc la conclusion du test ?

Exercice 2 : Soient $X_1, X_2, \ldots, X_p$ des variables aléatoires indépendantes et de même loi $\mathcal{N}(\mu_1, \sigma_1^2)$ et soient $Y_1, Y_2, \ldots, Y_q$ des variables aléatoires indépendantes et de même loi $\mathcal{N}(\mu_2, \sigma_2^2)$. On suppose que les deux échantillons sont indépendants et de même variance, c'est à dire que $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

On s'intéresse à l'estimation de la différence $\mu_1 - \mu_2$.

1. Proposer un estimateur de $\mu_1 - \mu_2$. On le notera D.
2. Etablir la loi de cet estimateur.
3. Proposer un estimateur de σ^2. On le notera S^2 et démontrer que $\frac{(p+q-2)S^2}{\sigma^2}$ suit une loi du Khi-deux à $(p+q-2)$ ddl et S^2 indépendant de $\overline{X}_p$ et $\overline{Y}_q$.
4. Etablir alors que
$$\frac{D - (\mu_1 - \mu_2)}{S\sqrt{\dfrac{1}{p} + \dfrac{1}{q}}} \sim T_{p+q-2}$$
5. Construire un intervalle de confiance pour la différence $(\mu_1 - \mu_2)$ au niveau de confiance $(1-\alpha)$ avec $\alpha \in]0, 1[$.

6. Soient $x_1, \ldots, x_{15}$ des réalisations de $X_1, \ldots, X_{15}$ et $y_1, \ldots, y_9$ des réalisations de $Y_1, \ldots, Y_9$. Tester au risque de 1% l'hypothèse nulle H0 : "$\mu_1 = \mu_2$" contre l'hypothèse alternative H1 : "$\mu_1 \neq \mu_2$" sachant que :

$$\sum_{i=1}^{15} x_i = 16.2 \qquad \sum_{i=1}^{15} x_i^2 = 28.7$$
$$\sum_{i=1}^{9} y_i = 8.9 \qquad \sum_{i=1}^{9} y_i^2 = 31.5.$$

   On pourra également s'aider de la commande suivante :

```
help(qt)
```

Exercice 3 : Lors d'une petite expérimentation sur des souris atteintes d'une maladie mortelle, on a tiré au sort parmi 16 souris, 7 qui reçoivent un nouveau traitement alors que les 9 autres sont des contrôles qui reçoivent un placebo. Leurs durées de survie sont mesurées en jours et donnent les résultats suivants :

Survie (en jours)	
Groupe 1 (Placebo)	Groupe 2 (Traitement)
$n_1 = 9$ mesures 52, 10, 40, 104, 50, 27, 146, 31, 46	$n_2 = 7$ mesures 94, 38, 23, 197, 99, 16, 141
$\sum_{j=1}^{n_1} x_{j,1} = 506$	$\sum_{j=1}^{n_2} x_{j,2} = 608$
$\sum_{j=1}^{n_1} x_{j,1}^2 = 42842$	$\sum_{j=1}^{n_2} x_{j,2}^2 = 79556$

On supposera que les données du groupe 1 sont des réalisations indépendantes d'une variable aléatoire X_1 de loi normale $\mathcal{N}(\mu_1, \sigma_1^2)$ et que les données du groupe 2 sont des réalisations indépendantes d'une variable aléatoire X_2 de loi normale $\mathcal{N}(\mu_2, \sigma_2^2)$.

1. Creer un vecteur placebo et un vecteur traitement.
2. Calculer la moyenne des durées de survie des souris des groupe 1 et 2 On les notera m_1 et m_2 respectivement. Commenter les résultats obtenus. En particulier, que peut-on dire sur l'effet du traitement sur la durée de survie ? On pourra s'aider de la commande suivante

```
help(mean)
```

3. Construire des intervalles de confiance au niveau de confiance 95% pour les moyennes réelles μ_1 et μ_2 et calculer leur réalisation. Commenter les résultats obtenus. En particulier, que peut-on dire sur l'effet du traitement sur la durée de survie ? On pourra s'aider de la commande suivante :

```
help(qt)
```

4. On suppose que $\sigma_1^2 = \sigma_2^2$. Tester au risque de 1% l'hypothèse nulle H0 : "$\mu_1 = \mu_2$" contre l'hypothèse alternative H1 : "$\mu_1 \neq \mu_2$". On pourra s'aider de la commande suivante :

```
help(t.test)
```

5. Quelles conclusions peut-on tirer de cette expérience ? Pour argumenter, on pourra s'aider de la commande suivante :

```
boxplot(placebo,traitement,names=c("placebo","traitement"))
```

6. Tester au risque de 1% l'égalité des variances. Pour cela, on pourra s'aider de la commande

```
help(var.test)
```

.

Exercice 4 : On veut tester la précision d'une balance en effectuant une série de 15 mesures du poids d'un kilo de riz. On obtient les mesures suivantes :

Poids (en g)
$n = 15$ mesures 996.17, 994.45, 998.78, 997.2, 1007.01, 998.45, 1003.93, 995.23, 997.01, 999.36, 997.64, 993.81, 1004.33, 991.38, 1000.97
$\sum_{j=1}^{n} x_j = 14975.72$
$\sum_{j=1}^{n} x_j^2 = 14951732$

On supposera que les données sont des réalisations indépendantes d'une variable aléatoire X de loi normale $\mathcal{N}(\mu, \sigma^2)$.

1. Créer un vecteur poids contenant les mesures. Rentrer la commande suivante et commenter :

```
plot(hist(poids))
```

2. Calculer le poids mesuré moyen que l'on notera m. On pourra s'aider de la commande suivante :

```
help(mean)
```

3. Donner une estimation de σ^2. On la notera s^2. On pourra s'aider de la commande suivante :

```
help(var)
```

4. Construire un intervalle de confiance à 90% pour la moyenne. On pourra s'aider de la commande suivante :

```
help(qt)
```

5. Construire un intervalle de confiance à 95% pour la variance. On pourra s'aider de la commande suivante :

```
help(qchisq)
```

6. Tester au risque de 1% la précision de la balance. On pourra s'aider de la commande suivante

```
help(t.test)
```

.

Exercice 5 : On s'intéresse au salaire journalier des employés d'une entreprise. On obtient les salaires suivants :

Salaires (en euro)
$n = 16$ mesures 41, 40, 50, 45, 41, 41, 40,43, 45, 52, 40, 48, 50, 40, 47, 46
$\sum_{j=1}^{n} x_j = 709$
$\sum_{j=1}^{n} x_j^2 = 31675$

On supposera que les données sont des réalisations indépendantes d'une variable aléatoire X de loi normale $\mathcal{N}(\mu, \sigma^2)$.

1. Calculer le salaire moyen mesuré que l'on notera m. On pourra s'aider de la commande suivante

```
help(mean)
```

2. Donner une estimation de σ^2. On la notera s^2. On pourra s'aider de la commande suivante

```
help(var)
```

3. L'entreprise prétend payer en moyenne ses salariés plus de 47 euros par jour. Au risque de 5%, pouvez vous confirmer cette affirmation ? Au risque de 1% ? On pourra s'aider de la commande suivante

```
help(t.test)
```

4. L'entreprise prétend également avoir très peu de différences de salaires au sein de l'entreprise, i.e avoir une variance des salaires $\sigma_0^2 = 5$. Dit-elle vrai ?On pourra s'aider de la commande suivante

```
help(qchisq)
```

Exercice 6 : On souhaite comparer les longueurs des mâchoires inférieures de 10 chacals mâles et 10 chacals femelles. On a les mesures suivantes :

Longueur (en mm)	
Groupe 1 (Mâles)	Groupe 2 (Femelles)
$n_1 = 10$ mesures 120, 107, 110, 116, 114, 111, 113, 117, 114, 112	$n_2 = 10$ mesures 110, 111, 107, 108, 110, 105, 107, 106, 111, 111
$\sum_{j=1}^{n_1} x_{j,1} = 1134$	$\sum_{j=1}^{n_2} x_{j,2} = 1086$
$\sum_{j=1}^{n_1} x_{j,1}^2 = 128720$	$\sum_{j=1}^{n_2} x_{j,2}^2 = 117986$

On supposera que les données du groupe 1 sont des réalisations indépendantes d'une variable aléatoire X_1 de loi normale $\mathcal{N}(\mu_1, \sigma^2)$ et que les données du groupe 2 sont des réalisations indépendantes d'une variable aléatoire X_2 de loi normale $\mathcal{N}(\mu_2, \sigma^2)$.

1. Créer des vecteurs "males" et "femelles", rentrer la commande suivante et commenter :

   ```
   boxplot(males,femelles)
   ```

2. Calculer la moyenne des longueurs des mâchoires des groupe 1 et 2. On pourra s'aider de la commande suivante :

   ```
   help(mean)
   ```

   On les notera respectivement m_1 et m_2.
3. Donner une estimation de σ^2. On la notera s^2.
4. Tester au risque de 5% le fait que le sexe des individus n'a pas d'incidence sur la longueur moyenne de leur mâchoire. On pourra s'aider de la commande suivante :

   ```
   help(t.test)
   ```

5. On suppose maintenant que les variables aléatoires X_1 et X_2 sont de variance σ_1^2 et σ_2^2. Tester au risque de 5% l'égalité de ces variances. On pourra s'aider de la commande suivante :

   ```
   help(var.test)
   ```

Exercice 7 : On considère un groupe de 28 individus souffrant d'un même handicap. Les individus ont été répartis en deux groupe, suivant deux apprentissages différents. Le premier consiste en de l'imitation (les sujets doivent imiter les gestes faits), le second consiste en la guidance (les sujets sont aidés physiquement pour effectuer les gestes). Le tableau ci-dessous donne les scores obtenus par les différents individus.

Scores	
Groupe 1 (Imitation)	Groupe 2 (Guidance)
$n_1 = 15$ mesures 19, 16, 24, 13, 9, 14, 17, 10, 19, 22, 23, 5, 7, 13, 11	$n_2 = 13$ mesures 15, 18, 23, 10, 8, 11, 12, 14, 21, 15, 18, 6, 7
$\sum_{j=1}^{n_1} x_{j,1} = 222$	$\sum_{j=1}^{n_2} x_{j,2} = 178$
$\sum_{j=1}^{n_1} x_{j,1}^2 = 3766$	$\sum_{j=1}^{n_2} x_{j,2}^2 = 2778$

On supposera que les données du groupe 1 sont des réalisations indépendantes d'une variable aléatoire X_1 de loi normale $\mathcal{N}(\mu_1, \sigma_1^2)$ et que les données du groupe 2 sont des réalisations indépendantes d'une variable aléatoire X_2 de loi normale $\mathcal{N}(\mu_2, \sigma_2^2)$.

1. Créer un vecteur imitation et un vecteur guidance.Effectuer un test de Shapiro et conclure. On pourra s'aider de la commande suivante :

   ```
   help(shapiro.test)
   ```

2. Rentrer la commande suivante et commenter :

   ```
   boxplot(imiation,guidance,names=c("imitation","guidance"))
   ```

3. Tester au risque de 5% le fait que les variabilités des scores dans chacun des groupes ne sont pas différentes. On pourra s'aider de la commande suivante :

   ```
   help(var.test)
   ```

4. Tester au risque de 5% le fait que la méthode choisie n'impacte pas le score moyen. On pourra s'aider de la commande suivante

   ```
   help(t.test)
   ```

Exercice 8 : On étudie l'influence du magnésium sur la croissance d'une moisissure. On procède à deux expériences distinctes :
`Expérience 1`**:** On cultive la moisissure dans 20 boîtes, on injecte une dose de 5mg dans 10 boîtes et une dose de 10mg dans les 10 autres. La croissance moyenne dans les 10 premières boîtes est de 1.03μm avec une variance de 0.05μm^2, la croissance moyenne dans les 10 dernières boîtes est de 1.12μm avec une variance de 0.1μm^2.
`Expérience 2:` On cultive les moisissures dans 10 boîtes, puis on sépare chaque boîte en deux. Dans une partie, on injecte 5mg de magnésium, dans l'autre 10mg de magnésium. La différence moyenne est égale à 0.08μm et la variance de la différence est égale à 0.02μm^2.

1. Expliquer la différence entre les deux expériences.
2. Pour chaque expérience, après avoir rappelé le cadre théorique, mettre en oeuvre un test au risque de 5% pour l'égalité des croissances moyennes. On pourra saider de la commande suivante :

   ```
   help(qt)
   ```

3. Les deux tests mènent-ils à la même conclusion ?

Exercice 9 : Une entreprise a mis au point un nouveau traitement contre le phylloxera, puceron qui ravage les vignes. Il est testé sur une parcelle de 600 plants sur lesquels on observe les résultats suivants :

Effet	Eradication	Amélioration	Sans effet
Nombre de plants	280	210	110

Les résultats promis par l'entreprise sont de 60% d'éradication, 30% d'amélioration et 10% sans effet.

1. Tester au risque de 1% la véracité des dires de l'entreprise. On pourra s'aider de la commande suivante

   ```
   help(qchisq)
   ```

2. On traite une deuxième parcelle avec le traitement habituel. Les réulstats observé sur 400 plants sont les suivants :

Effet	Eradication	Amélioration	Sans effet
Nombre de plants	220	90	90

   (a) Proposer un test qui permet de rejeter ou non l'hypothèse "le nouveau traitement est différent de l'ancien".

   (b) Tester au risque de 5% si les traitements sont différents. On pourra s'aider de la commande suivante

   ```
   help(qchisq)
   ```

Exercice 10 : On souhaite savoir si le rhésus dépend du groupe sanguin. Pour cela, on dispose du tableau de données suivant :

	O	A	B	AB	Total
Rhésus +	370	381	62	28	
Rhésus -	70	72	12	5	
Total					

Au risque de 5%, tester si le rhésus est indépendant du groupe sanguin. On pourra s'aider de la commande suivante

```
help(qchisq)
```

Exercice 11 : Le couvert végétal du domaine vital d'un orignal (élan d'amérique) se compose de peuplements feuillus (25% de la superficie du domaine vital), de peuplements mixtes (38% de la superficie), de peuplements résineux (25.8%) et d'un marécage (11.2%). Dans ce domaine, l'orignal a été localisé à 511 reprises au cours de l'année. Sur les 511 localisations, 118 se trouvaient dans le feuillus, 201 dans les peuplements mixtes, 110 dans les résineux, et 83 dans les marécages.

1. On veut montrer que l'orignal fréquente préférentiellement certains milieux. Proposer un test statistique pour vérifier cette hypothèse.

2. Réaliser le test et proposer une interprétation biologique du résultat. On pourra s'aider de la commande suivante

   ```
   help(qchisq)
   ```

Exercice 12 : Dans une étude sur un répulsif de moustique, on compte le nombre de piqûres de chaque personne à partir d'un échantillon de 150 personnes. On obtient

Nb de piqûres	0	1	2	3	4	5	6	>6
Nb d'individus	32	54	34	21	6	2	1	0

Tester au risque de 5% que le nombre de piqûres pour une personne est une variable aléatoire suivant une loi de Poisson de paramètre 1. Pour s'aider, soit $X \sim \mathcal{P}(1)$, on a

k	0	1	2	3	4	5	6	>6
$\mathbb{P}[X = k]$	0.37	0.37	0.18	0.061	0.015	0.0031	0.00051	$8.10^{-}5$

On pourra s'aider de la commande suivante

```
help(qchisq)
```