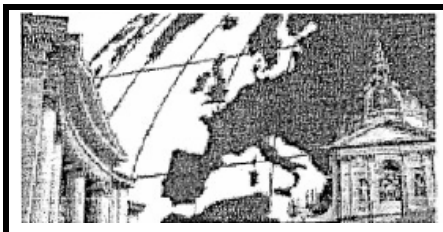

PANTHÉON - SORBONNE – UNIVERSITÉ PARIS 1

**UFR 06 – EM Sorbonne**
**L3 Finance et L3 Gestion**

# *STATISTIQUES*

***Cours de Stéphane Robin***

**Brochure de Travaux Dirigés**

# Remarques préliminaires

Nous invitons les étudiants à lire attentivement les remarques suivantes. Elles visent à assurer le bon fonctionnement des travaux dirigés tout au long de ce premier semestre universitaire.

Au niveau de la notation, la décomposition de la note finale s'effectue de la façon suivante :

**Examen : 50%**
**Moyenne de TD : 50%**

La moyenne de TD est basée sur une interrogation écrite (éventuellement deux) pondérée par une note de participation. **Toute absence non justifiée à une interrogation écrite se traduire par un zéro comptabilisé dans la moyenne de TD.**

Par ailleurs, au-delà de 4 absences non justifiées au TD, l'étudiant sera considéré comme défaillant.

Les tables statistiques (distribuées lors du 1$^{er}$ TD) et les calculatrices (non programmables) doivent être systématiquement apportées en travaux dirigés par l'étudiant.

Les calculatrices programmables (graphiques ou alphanumériques) sont strictement interdites lors des interrogations écrites en TD, et lors du partiel.

# Sommaire

TD n° 1. Rappels de statistiques, lectures et interprétations de tables (une séance)

TD n° 2. Lois de probabilités usuelles (une séance)

TD n° 3. Echantillons et échantillonnages (une séance)

TD n° 4. Estimation ponctuelle (une séance)

TD n° 5. Estimation par intervalle de confiance (deux séances)

TD n° 6. Théorie des tests (cinq séances dont une consacrée à des révisions pendant la « semaine pédagogique »)

Aux séances indiquées ci-dessus viendra s'ajouter une séance consacrée à une interrogation écrite. Celle-ci viendra s'intercaler soit entre le TD n°5 et le TD n°6, soit entre les premières séances consacrée au TD n° 6.

# Rappels des fondamentaux

***Avant de commencer à réfléchir sur les exercices proposés dans ce polycopié, fermez le cours et assurez-vous de pouvoir répondre aux questions suivantes. Si vous n'y arrivez pas, reprenez vos notes de cours et révisez vos fondamentaux avant d'entamer les exercices.***

1. Qu'est-ce qu'une population? une variable aléatoire? un échantillon? une mesure d'inférence?

2. Proposez une définition pour les termes suivants:

- Données qualitatives
- Données quantitatives
- V.A. discrète
- V.A. continue
- Moments d'une distribution
- Quantiles
- Quartiles
- Fonction de répartition
- Fonction de densité

# TD n°1 : Rappels de statistiques, lectures et interprétations de tables

**EXERCICE 1.**

On rappelle que le loto est un jeu de hasard qui se pratique de la manière suivante: chaque joueur sélectionne sur une grille 6 nombres entre 1 et 49, l'ordre de sélection n'étant pas pris en compte et chaque nombre ne pouvant être choisi plus d'une fois.

Le tirage consiste à extraire d'une urne, sans remise, 6 boules parmi 49, chacune étant associée à un nombre.

1. Montrer que la probabilité d'avoir les "6 bons numéros" en ne jouant qu'une seule grille est de $7.10^{-8}$.

2. En supposant qu'il y a deux tirages par semaine, au bout de combien de temps une personne jouant la même grille à chaque tirage aura-t-elle une probabilité de 0,5 d'avoir obtenu au moins une fois les "6 bons numéros" ?

**EXERCICE 2.**

On considère une population de 100 skieurs dans laquelle on ne trouve que des bons et des mauvais skieurs : il y a 40 bons skieurs et 60 mauvais skieurs. La probabilité qu'un bon skieur ait un accident grave un jour donné est 0,01. La probabilité qu'un mauvais skieur ait un accident un jour donné est 0,02. On suppose que les accidents des skieurs sont mutuellement indépendants (il n'y a pas de collision entre skieurs).

1. En utilisant la formule des probabilités totales (après avoir choisi une bonne partition de la population), calculer la probabilité qu'un skieur pris au hasard dans la population de 100 skieurs ait un accident grave un jour donné.

2. Déterminez la loi du nombre d'accidents graves par jour.

3. Combien y a-t-il d'accidents graves en moyenne par jour? Quel est l'écart–type?

4. Quelle est la probabilité qu'il y ait au moins trois accidents graves un jour donné parmi les 100 skieurs ?

# TD n°2 : Lois de probabilités usuelles

**EXERCICE 1**

Une usine a 100 machines identiques qui fonctionnent simultanément. Une étude de sûreté a montré que la probabilité qu'une d'entre elle tombe en panne un jour donné est de 0,04. Trouvez la probabilité qu'il y ait plus de deux machines qui tombent en panne le même jour. Utilisez l'approximation de la loi binomiale par une loi de Poisson. Si l'approximation n'était pas utilisée, quels seraient les calculs à effectuer pour trouver la probabilité exacte?

**EXERCICE 2**

Un grand nombre d'étudiants passent un examen de statistiques noté sur 100. Les notes sont distribuées selon une loi normale de moyenne 70 et la probabilité qu'un étudiant particulier, choisi au hasard reçoive une note inférieure à 85 est de 0,9332. Quatre étudiants sont choisis au hasard. Quelle est la probabilité qu'au moins une des quatre notes dépasse 80 à cet examen?

**EXERCICE 3**

Un investisseur dispose de 1000 $ qui peuvent être placés dans deux actifs financiers. Les taux de rendement par $ de ces actifs sont désignés par les variables aléatoires X et Y. On suppose que ces variables aléatoires ont la même moyenne $\mu$ et la même variance $\sigma^2$. On désigne par C la covariance entre X et Y. L'investisseur choisit de placer $\alpha$ $ dans le premier actif et le reste dans le second. En admettant que le montant $\alpha$ caractérise un portefeuille particulier, on veut comparer les performances de tous les portefeuilles possibles.

1. Calculez l'espérance et la variance du taux de rendement du portefeuille, noté R, dans le cas où $C = 0$. Trouvez alors la valeur $\alpha$ qui rend la variance minimale. Commentez.

2. Calculez la variance de R dans le cas où $C \neq 0$. Montrez que $\alpha = 500$ est moins risqué que $\alpha = 0$ à condition que C soit inférieure à $\sigma^2$. Trouvez la valeur de C pour laquelle la diversification est la plus avantageuse.

# TD n°3 : Echantillons et échantillonnage

**EXERCICE 1**

Le taux de dividendes distribués par les entreprises dont les actions sont cotées à la Bourse de Paris (entreprises dites « du CAC 40 ») obéit à une loi normale d'écart–type 2,6 %. On tire un échantillon aléatoire de 16 entreprises du CAC 40 pour estimer le taux de dividendes moyen des actions à la Bourse de Paris.

1. Quelle est la probabilité que le taux moyen de l'échantillon excède le taux de la population de plus de 0,8 % ?
2. Quelle est la probabilité qu'il soit en dessous de celui de la population de plus de 1 % ?
3. Quelle est la probabilité qu'il diffère du taux moyen de la population de plus de 1,2 % ?
4. Un second échantillon de 40 entreprises est tiré indépendamment du premier. Sans calcul, établissez si les probabilités, calculées en 1), 2) et 3) sont plus élevées, plus faibles ou identiques à celles correspondant au second échantillon.

**EXERCICE 2**

Le prix de vente des appartements dans le XVème arrondissement de Paris suit une loi normale d'écart–type 3000 euros / $m^2$. On tire un échantillon aléatoire de 36 appartements afin d'estimer le prix moyen de vente. $\overline{X}$ désigne alors la moyenne de l'échantillon.

1. Quelle est la probabilité que l'intervalle [ $\overline{X}$ - 1500; $\overline{X}$ + 1500] contienne le prix moyen de vente de la population?
2. On définit le degré de confiance d'une estimation en calculant la probabilité que $\mu$ (la vraie valeur) appartienne à un intervalle de confiance de la forme:

$$\mu = \overline{X} \pm a,$$

où a est la marge ou « erreur d'échantillonnage ». Quelle est cette marge ici ?

**EXERCICE 3**

Une entreprise envisage l'émission d'une obligation convertible. Les décideurs pensent que les termes de l'offre vont attirer 20 % de leurs actionnaires. On suppose que cette valeur est correcte. On tire un échantillon aléatoire de 130 actionnaires.

1. Calculez la probabilité que la proportion d'actionnaires intéressés par l'offre dans l'échantillon soit supérieure à 0,15.
2. Calculez la probabilité qu'elle soit comprise entre 0,18 et 0,22.
3. L'échantillon extrait de la population des actionnaires est maintenant de taille 500. Sans faire de calculs, évaluer si les probabilités calculées dans les questions 1 et 2 sont plus élevées, plus faibles ou identiques à celles correspondant à l'échantillon de 500.

**EXERCICE 4**

Un candidat à une élection a l'intention de lancer sa campagne si son électorat initial dépasse 30 % des votants. Un échantillon de 300 votants est étudié, et on décide de lancer la campagne si la proportion dans l'échantillon qui soutient le candidat dépasse 28%.

1. Quelle est la probabilité que la décision d'entrer en campagne soit prise si en fait l'électorat qui le soutient initialement est de 20 % ?
2. Quelle est la probabilité de décider de ne pas lancer la campagne si le soutien initial est de 40 %?

# TD n°4 : Estimations ponctuelles

**EXERCICE 1:**
Afin d'apprécier l'opportunité de construire un centre sportif dans une grande ville, un organisme d'études chargé par la municipalité d'effectuer une étude de marché, interroge $n = 900$ habitants.
1. Pour estimer la proportion $p$ des habitants de la ville souhaitant la création de ce centre, on considère $Fn$, la proportion dans l'échantillon des habitants souhaitant sa construction. Justifiez ce choix et étudiez les propriétés de l'estimateur retenu (biais, convergence, efficacité).
2. Dans cet échantillon, on observe que 360 personnes souhaitent la création du centre. Donner une estimation ponctuelle de $p$.

**EXERCICE 2:**
Deux économistes estiment $\mu$, la dépense moyenne des ménages français pour l'alimentation. Ils obtiennent deux estimations sans biais et statistiquement indépendantes U et V. Le second économiste est toutefois moins soigneux que le premier, et son estimation V a un écart-type trois fois plus grand que celui de U. Les deux économistes décident de constituer un nouvel estimateur combinant leurs deux mesures afin d'obtenir une estimation globale digne d'être publiée. Ils hésitent entre trois estimateurs :

$$W_1 = \frac{1}{2}U + \frac{1}{2}V$$

$$W_2 = \frac{3}{4}U + \frac{1}{4}V$$

$$W_3 = U + 0.V = U$$

Aidez ces économistes à choisir un estimateur en déterminant:
1. Parmi ces trois estimateurs, lesquels sont sans biais.
2. Quel est (s'il existe) l'estimateur le plus efficace.

**EXERCICE 3:** ***Estimation simultanée de l'espérance mathématique et de la variance d'une loi normale***
Le coût d'un certain type de sinistres est une V.A. $X$ distribuée suivant une loi normale d'espérance mathématique $m$ et de variance $\sigma^2$ inconnues.
1. Quels sont les estimateurs du maximum de vraisemblance de $m$ et de $\sigma^2$ pour un échantillon de taille $n$ de $X$? Sont-ils sans biais?
2. Application numérique pour $n = 50$ :

$$\sum_{i=1}^{n} x_i = 6000$$

$$\sum_{i=1}^{n} x_i^2 = 842500$$

3. Quelles sont les estimations sans biais correspondantes de $m$ et de $\sigma^2$ ?

# TD n°5 : Estimations par intervalles de confiance

**EXERCICE 1 :** ***Estimation par IC d'une différence de moyenne***

Pour un échantillon aléatoire de 10 clients ayant un crédit voiture depuis un an chez un concessionnaire, on a les montants suivants de remboursement, en euros, à la fin d'un mois:
930; 1629; 449; 1580; 1293; 1098; 1559; 1190; 1891; 456.
Pour un échantillon aléatoire de 8 nouveaux clients ayant un crédit ouvert, la charge de remboursement à la fin du même mois s'élève à :
468; 730; 946; 406; 1095; 699; 591; 798.
Trouvez un intervalle de confiance à 90 % pour la différence entre les montants des deux populations. Précisez les hypothèses retenues.

**EXERCICE 2 :** ***Intervalle de confiance unilatéral pour la variance d'une variable normale (espérance mathématique connue)***

Une machine découpe des pièces de diamètre D. On suppose que D est une variable aléatoire qui suit une loi normale d'espérance $m = 2$ et de variance $\sigma^2$.
A l'aide d'un échantillon de 30 pièces, déterminez un intervalle de confiance unilatéral pour $\sigma^2$, de la forme $\sigma^2 > A$.
On a observé:

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(d_i - m)^2 = 0{,}13$$

Quel est l'intervalle correspondant?

**EXERCICE 3 :**

1. Dans une ville d'environ 50 000 habitants, la proportion $p$ des fumeurs a été évaluée à 54%. Dans ces conditions, calculez la probabilité que, dans un échantillon aléatoire de 100 personnes, le nombre des fumeurs soit strictement supérieur à 65.
2. Par ailleurs, on a observé que la consommation journalière d'un fumeur (en nombre de cigarettes) avait pour moyenne $m = 16$ et pour écart-type $\sigma = 4{,}2$.
   2.1. Indiquez, en la justifiant, une loi approchée de la consommation moyenne, dans un échantillon aléatoire de 80 fumeurs.
   2.2. En déduire une approximation de la probabilité que, dans cet échantillon, la consommation moyenne soit strictement inférieure à 15.
3. A la suite d'une campagne publicitaire contre le tabagisme, des responsables du Ministère de la Santé, désirant en mesurer l'efficacité, décident de réaliser une enquête sur la proportion des fumeurs dans la ville ainsi que sur la consommation de tabac. Les paramètres $p$, $m$, et $\sigma$ ne sont donc plus supposés être connus. Dans un échantillon aléatoire de 200 personnes, on a dénombré 90 fumeurs.
   3.1. Proposez un estimateur de $p$ et indiquez, sans démonstration, ses propriétés.
   3.2. Construire un intervalle de confiance pour $p$, au seuil de risque 1%. Cet intervalle permet-il de conclure à l'efficacité de la campagne?
4. Dans un échantillon de 100 fumeurs, on a constaté une consommation moyenne de 14 cigarettes par jour. En supposant que $\sigma = 4{,}2$, construire un intervalle de confiance pour $m$ au seuil de risque 1 %. Concluez.

# TD n°6 : Théorie des tests

***L'ensemble des exercices suivants utilise l'approche classique de Neyman et Pearson***

## PARTIE I : EXERCICES SIMPLES D'APPLICATION DU COURS

### EXERCICE 1:

Une grande banque nationale cherche à déterminer si sa nouvelle campagne publicitaire va permettre d'accroître les ventes d'un nouveau produit d'épargne destiné aux particuliers. La banque espère que cette campagne permettra en moyenne de vendre ce produit à 50 clients supplémentaires par semaine. Dans la population de toutes les agences de cette banque réparties sur le territoire national, le nombre de clients supplémentaires par semaine est supposé suivre une loi normale de moyenne $\mu$ et d'écart–type $\sigma$ inconnus.
Sur un échantillon aléatoire simple de 20 agences suivies pendant une semaine, on observe un nombre moyen de clients supplémentaires égal à 41,3, avec un écart-type de 12,2. Testez, au seuil de 5% et au seuil de l %, l'hypothèse nulle selon laquelle le nombre moyen de clients supplémentaires dans la population des agences est égal à 50 par semaine, contre l'hypothèse alternative que ce nombre est inférieur à 50.

### EXERCICE 2 :

Dans une population donnée, on veut comparer la proportion $p$ de naissances masculines à celles des naissances féminines. Pour ce faire, on se dote d'un échantillon de 900 naissances, dans lequel on observe 477 garçons. En précisant bien, dans chaque cas, la règle de décision et la statistique utilisées, tester au seuil de risque $\alpha = 5\%$ :

1. l'hypothèse $H_0$: "$p = 0{,}5$" contre l'hypothèse $H_1$: "$p > 0{,}5$".
2. l'hypothèse $H_0$: "$p = 0{,}5$" contre l'hypothèse $H_1$: "$p \neq 0{,}5$".

Que peut-on conclure?

### EXERCICE 3 :

Un investisseur envisage l'acquisition d'un portefeuille d'actions dont le rendement mensuel suit une loi normale $N(m, \sigma)$, d'espérance $m$ et de variance $\sigma^2$ inconnues. L'investisseur décide d'acquérir ce portefeuille si son rendement mensuel moyen est au moins de 200 euros. Il observe ce portefeuille pendant 5 mois et trouve les rendements suivants:

| **Mois 1** | **Mois 2** | **Mois 3** | **Mois 4** | **Mois 5** |
|---|---|---|---|---|
| 150 euros | 280 euros | 170 euros | 210 euros | 180 euros |

1. Tester, au seuil de 5%, l'hypothèse $H_0$ : $m = 200$ contre l'hypothèse $H_1$ : $m < 200$ et indiquez quelle décision l'investisseur va prendre en suivant la règle qu'il s'est fixé.
2. Tester, au seuil de 5%, l'hypothèse $H_0$ : $\sigma^2 = 2500$ contre l'hypothèse $H_1$ : $\sigma^2 \neq 2500$.

**EXERCICE 4**

Lors de la fabrication d'un biocarburant, la température dans une cuve de fermentation peut être considérée comme une VA X qui évolue selon une loi normale $N(\mu, \sigma)$ de paramètres $\mu$ et $\sigma$ inconnus. On voudrait connaître la température moyenne dans la cuve pendant le processus de fabrication. On procède à 16 relevés pendant ce processus, et on obtient dans cet échantillon une température moyenne "X barre" de 120 avec un écart-type empirique S = 40.

1. Tester, au seuil de 5% puis au seuil de 1%, l'hypothèse $H_0$: $\mu = 100$ contre $H_1$: $\mu \neq 100$.
2. Tester, au seuil de 5%, l'hypothèse $H_0$: $\mu = 150$ contre $H_1$: $\mu > 150$.
3. Tester, au seuil de 5%, l'hypothèse $H_0$: $\sigma^2 = 1600$ contre $H_1$: $\sigma^2 \neq 1600$.

**EXERCICE 5**

Un échantillon aléatoire de 1250 diplômes d'université délivrés en 2008 a donné la répartition suivante:

| | **Licence** | **Master** | **Doctorat** |
|---|---|---|---|
| Genre : masculin | 501 | 162 | 27 |
| Genre : féminin | 409 | 143 | 8 |

1. Que peut-on tester ici? Exprimez 1'hypothèse nulle de façon littéraire.
2. Calculez la distance du Khi-deux et le seuil de signification empirique correspondant.

Quelle conclusion peut-on tirer de ce test?

**PARTIE II : EXERCICES DE REFLEXION**

**EXERCICE 1 (prolongé sur la page suivante):**

Dans la procédure d'exécution du plan, on sait qu'il a été envisagé un système de surveillance appelé « clignotants ». Ceux-ci doivent permettre, en particulier, de déceler suffisamment tôt si l'économie se dirige vers un état inflationniste ou déflationniste. On utilisera ici comme clignotants l'indice mensuel des prix à la consommation, exprimé en pourcentages.

Dans le cadre d'une prévision à court terme, entre 12 et 18 mois, on suppose que l'on peut représenter l'accroissement mensuel $\Delta P$ de cet indice par:

$$\Delta P = a + \varepsilon$$

où a représente le tendance et $\varepsilon$ un ensemble de facteurs accidentels ou d'erreurs de mesure que l'on peut assimiler à une variable aléatoire normale d'espérance mathématique nulle et d'écart–type $\sigma_\varepsilon = 0{,}20$, indépendant de a dans les limites envisagées.

On considère deux valeurs possibles pour a (en pourcentages)

a = 0,3 par mois en période de déflation

a = 0,5 par mois en période d'inflation

On observe les accroissements successifs de l'indice des prix sur 3 mois.

1. On définit un seuil critique A pour la moyenne de ces accroissements, au-dessus duquel on décidera des mesures de stabilisation des prix.

1.1. Formalisez le problème de test ainsi posé:

- Paramètres d'intérêt, lois de probabilités des variables aléatoires utilisées
- Hypothèses
- Décisions
- Règle de décision: $C_1$

**(suite du premier exercice de la partie II du TD):**

1.2. On fixe le seuil critique A = 0,35. Quels sont les deux risques associés à la règle de décision $C_1$?

1.3. On choisit comme hypothèse de base $H_0$ celle où a = 0,5

1.3.1. Quelle est l'interprétation pratique de ce choix?

1.3.2. Quels sont les risques de 1ère et 2ème espèce correspondants?

1.4. En conservant a = 0,5 comme hypothèse de base $H_0$, on fixe maintenant le risque de 1ère espèce $\alpha$ à 5%

1.4.1. Quel doit être le seuil critique correspondant?

1.4.2. Quelle sera la valeur du risque de 2ème espèce $\beta$ ?

2. On définit maintenant un seuil critique B pour le maximum des observations, au-dessus duquel on décidera des mesures de stabilisation des prix:

2.1. Quelle est la nouvelle règle de décision adoptée: $C_2$ ?

2.2. On fixe le seuil critique B = 0,50 Quels sont les deux risques associés à la règle de décision $C_2$?

2.3. On choisit comme hypothèse de base $H_0$, celle où a = 0,5 et on fixe le risque de 1ère espèce $\alpha$ à 5%

2.3.1.Quel doit être le seul critique B correspondant ?

2.3.2.Quelle sera la valeur du risque de 2ème espèce $\beta_2$ ?

3. Comparez $C_1$ et $C_2$, en prenant comme hypothèse de base a = 0,5.

**EXERCICE 2:**

Une société a expérimenté que l'année passée, le montant de ses commandes (en milliers d'euros) était distribuée selon une loi normale $N(250, 75)$. On s'intéresse aux commandes de cette année, en supposant, pour simplifier les calculs, que l'écart-type est demeuré inchangé. Sur un échantillon de $n = 225$ commandes, on observe un montant moyen de 260 K€.

1. Tester, au seuil de 5 %, l'hypothèse $H_0$ : "$\mu = 250$" contre l'hypothèse $H_1$ : "$\mu \neq 250$", où $\mu$ désigne la moyenne des commandes cette année.
2. Toujours pour un échantillon de $n = 225$ commandes, mais avant tout sondage, on décide d'accepter $H_0$ : "$\mu = 250$" si la moyenne empirique est inférieure à 261,50. Précisez alors $H_1$, et calculez le seuil du test correspondant à cette règle de décision.
3. Combien faudrait-il examiner de commandes pour que $\mu$ appartienne à un intervalle de confiance bilatéral d'amplitude 10 avec un niveau de confiance de 10 %?

**EXERCICE 3:**

On désire comparer les résultats d'un nouveau traitement découvert récemment et d'un médicament existant. Le taux de guérison avec le traitement traditionnel est de 60 %. Sur 100 individus, 72 ont été guéris en utilisant le nouveau médicament.

1. En supposant que les traitements sont équivalents, déterminez la loi de $X$, taux de guérison observé sur un échantillon de 100 personnes. Quelle est la probabilité d'avoir plus de 60 guérisons dans un échantillon de cette taille?
2. Le Ministère de la Santé doit décider l'autorisation de ce nouveau traitement. Etant donné que le traitement est très onéreux, le ministère ne souhaite pas le rembourser s'il n'est pas plus efficace que le traitement traditionnel. Le ministère choisit le processus de décision suivant: si le taux de guérison dans l'échantillon est supérieur à $\theta$, un seuil fixé, le nouveau traitement sera remboursé. Dans le cas contraire, il ne le sera pas. Ecrire les hypothèses testées et les risques de 1ère espèce $\alpha$ et de 2nde espèce $\beta$. Compte tenu de ses préférences, que doit faire le Ministère pour déterminer $\theta$ ?
3. Le ministère a choisi $\alpha = 0,05$. Que va-t-il décider?

4. Une association de malades conteste le processus de décision du ministère. Ils craignent qu'à l'avenir, les malades ne puissent accéder à des traitement efficaces pour des raisons financières. Ils affirment qu'il y a 89 % de chances que le Ministère refuse de rembourser un traitement qui aurait un taux de guérison de 0,62. Ont-ils raison, et pourquoi?
5. En choisissant d'autres valeurs du taux de guérison, tracez la courbe d'efficacité de ce test. Que constatez-vous? Que feriez-vous alors?

**EXERCICE 4**

Un indice de difficulté de compréhension de textes écrits (*fog index* = moyenne pondérée du nombre moyen de mots par phrase et du pourcentage de mots de plus de trois syllabes) est utilisé afin de comparer deux revues : une revue de vulgarisation scientifique et un magazine sportif. On considère que l'indice est distribué selon une loi normale $N(\mu_1, \sigma_1)$ dans la population des publicités parues dans la première, et selon une loi normale $N(\mu_2, \sigma_2)$ dans la population du second (les deux populations étant indépendantes).
Un échantillon de 6 publicités parues dans la revue de vulgarisation scientifique a donné les valeurs suivantes :

| | | | | | |
|---|---|---|---|---|---|
| 15,75 | 11,56 | 11,16 | 9,92 | 9,23 | 8,20 |

Un échantillon (indépendant du précédent) de 6 publicités parues dans un magazine sportif a donné les valeurs suivantes :

| | | | | | |
|---|---|---|---|---|---|
| 9,17 | 8,44 | 6,10 | 5,78 | 5,58 | 5,36 |

1. Testez au seuil de 10% l'hypothèse que les variances des indices sont égales dans les deux populations.
2. Testez ensuite, toujours au seuil de 5 %, l'hypothèse nulle que les moyennes des indices sont les mêmes dans les deux populations contre l'hypothèse alternative que la revue scientifique est plus difficile à lire que le journal sportif.

**EXERCICE 5**

Soit X le nombre des accidents qui se produit par mois, en scooter des mers, sur une plage déterminée. Nous supposons que $X$ est une variable aléatoire suivant une loi de Poisson de paramètre $\theta$. La municipalité, inquiète, décide de relever durant 5 mois le nombre d'accidents sur cette plage. Si ce total dépasse la valeur $k$, la municipalité décidera d'interdire l'usage des scooters ($\theta = 3$),sinon elle les tolèrera ($\theta = 1$).
1. On choisit $k = 9$
Définir entièrement le problème de test ainsi posé ainsi que la règle de décision adoptée. Quels sont les deux risques associés à cette règle de décision ?
2. On retient comme hypothèse de base, la situation jugée "acceptable". Interpréter ce choix. Quelle est la valeur critique $k_1$ correspondant à un risque de 1ère espèce $\alpha_1 = 0.014$ ? Quelle est la puissance de ce test?
3. On retient maintenant l'autre situation comme hypothèse de base. Interpréter ce choix. Quelle est la valeur critique $k_2$ correspondant à un risque de 1ère espèce $\alpha_2 = 0.018$ ? Quelle est la puissance de ce test?
4. On dénombre un total de 8 accidents pendant 5 mois sur cette plage. Quelle décision la municipalité prendra-t-elle dans chacun des cas 2 et 3 ?

**EXERCICE 6**

On s'intéresse au taux de réussite au permis de conduite en fonction du genre. On dispose d'un échantillon de 429440 individus sur lequel on a dénombré 221023 personnes de sexe masculin et 208417 personnes de sexe féminin.

1. Tester, au seuil de 5 %, l'hypothèse de base selon laquelle "le taux de réussite au permis de conduire des personnes de sexe masculin est de 50 %".
2. Calculer la puissance du test pour une proportion de candidats masculins de 0,505.
3. Proposer alors un test unilatéral de l'hypothèse de base, en justifiant le type de test choisi.

**EXERCICE 7**

Afin de déterminer la santé financière d'une banque on détermine son ratio de liquidité $X$. Dans le secteur des banques d'affaires, on admet que ce rapport est une variable aléatoire $X_1$ qui suit une loi normale $N(m_1,\sigma_1)$. On observe un échantillon de 10 banques de ce secteur et on trouve :

$$\overline{X}_1 = 6{,}1\% \qquad \text{et} \qquad S_1 = \sqrt{\frac{\sum_{i=1}^{10}(x_{1i} - \overline{X_1})^2}{9}} = 0{,}3\%$$

Dans le secteur des banques commerciales, on admet que ce rapport est une variable aléatoire $X_2$ qui suit une loi normale $N(m_2, \sigma_2)$. On observe un échantillon de 20 banques de ce secteur et on trouve

$$\overline{X}_2 = 6{,}6\% \qquad \text{et} \qquad S_2 = \sqrt{\frac{\sum_{i=1}^{20}(x_{2i} - \overline{X_2})^2}{19}} = 0{,}45\%$$

1. Testez, au seuil $\alpha = 5\%$, $H_0$: "$\sigma_1 = \sigma_2$" contre $H_1$: "$\sigma_1 < \sigma_2$". Que concluez-vous?
2. Répétez ce test en supposant $m_1$ et $m_2$ connues et égales respectivement à 6,1 % et 6,6 %. Votre conclusion change-t-elle?
3. On suppose maintenant que les variances $\sigma^2_1$ et $\sigma^2_2$ sont inconnues mais égales. Testez l'hypothèse $H_0$: "$m_1 = m_2$" contre $H_1$: "$m_1 \neq m_2$" aux seuils de $1^o/_{oo}$, 1% et 5%. Que concluez-vous?

**EXERCICE 8**

Un horticulteur cherche à créer une nouvelle variété de roses "Eternité". Il souhaite que sa rose, une fois coupée, puisse avoir une durée de vie moyenne de 15 jours. Etant donnés les coûts élevés de conception du produit engagés, pour que sa rose soit rentable, il ne peut pas se permettre de supporter de nouveaux coûts de développement. Afin d'obtenir une aide à la décision à propos de la commercialisation potentielle de sa rose, il fait appel à un statisticien. L'horticulteur dispose d'un plant expérimental de 150 roses, toutes étiquetées et répertoriées. Le statisticien coupe 30 roses au hasard et enregistre leur durée de vie ($d_i$), reportée dans le tableau page suivante.

1. Décrivez la série ainsi obtenue en en précisant les principales caractéristiques.
2. Soit $p_0$ la proportion de roses dont la durée de vie est au moins de 15 jours. Soit $p_1$ la proportion de roses dont la durée de vie est au plus de 12 jours.
   2.1. Préciser la population étudiée et l'échantillon utilisé.
   2.2. Donner un estimateur et une estimation pour chacun des paramètres $p_0$ et $p_1$.
   2.3. Proposer un intervalle de confiance bilatéral pour chacun des paramètres, en supposant que l'horticulteur tolère un risque de 10 %. Commenter les résultats ainsi obtenus. Donner une "fourchette" pour le nombre de roses ayant une durée de vie d'au moins 15 jours.

| **No. de la rose** | $d_i$ | **No. de la rose** | $d_i$ |
|---|---|---|---|
| 1 | 18 | 16 | 18 |
| 2 | 14 | 17 | 16 |
| 3 | 13 | 18 | 14 |
| 4 | 16 | 19 | 17 |
| 5 | 15 | 20 | 14 |
| 6 | 14 | 21 | 14 |
| 7 | 15 | 22 | 16 |
| 8 | 18 | 23 | 17 |
| 9 | 14 | 24 | 16 |
| 10 | 13 | 25 | 11 |
| 11 | 15 | 26 | 16 |
| 12 | 17 | 27 | 13 |
| 13 | 14 | 28 | 11 |
| 14 | 17 | 29 | 13 |
| 15 | 14 | 30 | 17 |

3. Soit D une variable aléatoire représentant la durée de vie des roses, dont les $d_i$ sont les réalisations. On note $E(D) = m$ et $Var(D) = \sigma^2$

3.1. Calculer les estimations usuelles pour chacun de ces deux paramètres, en précisant bien les formules utilisées.

3.2.Peut-on considérer, avec un risque de 1ère espèce de 5%, que $D$ suit une loi normale?

4. Sachant que l'horticulteur tolère un risque de 1ère espèce de 10%, faire un test statistique pour déterminer si l'on peut commercialiser la variété de roses "Eternité". Que peut-on conclure des résultats de ce test? Ces résultats changeraient-ils pour un seuil de 5%?

5. L'horticulteur ne peut attendre les résultats du statisticien et décide de commercialiser sa rose. Il s'intéresse désormais au rendement potentiel de cette nouvelle variété. Le rendement d'une variété dite classique est une variable aléatoire X qui suit une loi normale d'espérance $m = 10$ et d'écart–type $\sigma = 2$. Dans l'échantillon considéré, on trouve un rendement moyen égal à 15 avec une variance de 3,5. Déterminer, à l'aide d'un test statistique au seuil de 10%, si la nouvelle variété "Eternité" (dont le rendement suit également une loi normale) est plus rentable, en moyenne, qu'une variété de rose classique.