T.D.1. Variables aléatoires

Exercice 1

Un sac contient quatre jetons noirs et quatre jetons blancs numérotés. On tire deux jetons du sac, simultanément. Soit X la variable aléatoire qui, à chaque tirage, associe le nombre de jetons noirs tirés. Déterminer la loi de probabilité de X, son espérance mathématique et la valeur approchée arrondie à 10^{-2} de son écart type.

Exercice 2

Une usine d'horlogerie fabrique une série de montres. La fabrication comporte deux phases. La première phase fait apparaître un défaut *a* dans 2 % des cas; la seconde phase, un défaut *b* dans 10 % des cas. Au cours de la fabrication, on prélève au hasard successivement cinq montres. On considère que le nombre de montres fabriquées est assez grand pour que l'on puisse supposer que les tirages sont indépendants.

1° Une montre est tirée au hasard. On définit les événements suivants :

A : « La montre tirée présente le défaut *a* »; B : « La montre tirée présente le défaut *b* ».

On suppose que les événements A et B sont indépendants.

Montrer que la probabilité qu'une montre choisie au hasard dans la production ne présente aucun des deux défauts *a* et *b* est 0,882.

2° Soit X la variable aléatoire qui associe à chaque prélèvement de cinq montres le nombre de montres sans aucun des deux défauts *a* et *b*.

a) Expliquer pourquoi X suit une loi binomiale. Donner les paramètres de cette loi.

b) Déterminer la probabilité de l'événement « Quatre montres au moins n'ont aucun défaut », On donnera la valeur de cette probabilité.

Exercice 3

Dans une entreprise de vente par correspondance un étude statistique a montré qu'il y avait 3% de bons de commande comportant au moins une erreur. On constitue au hasard un échantillon de 100 bons de commande parmi ceux traités un jour donné.

Le nombre de bons de commande traités dans cette journée est assez important pour qu'on puisse assimiler ce prélèvement à un tirage avec remise de 100 bons de commande. On désigne par X la variable aléatoire qui associe à tout échantillon de 100 bons le nombre de bons erronés. On admet que X suit la loi de Poisson.

Déterminer, la probabilité de chacun des événements suivants:

a) E1 : « il y a exactement 2 bons erronés parmi les 100 » ;

b) E2 : « il y a moins de 2 bons erronés parmi les 100 » ;

c) E3 : « il y a au moins 2 bons erronés parmi les 100».

Exercice 4

Une usine produit des bobines de fil pour l'industrie textile. On désigne par X la variable aléatoire qui, à toute bobine tirée au hasard de la production d'une journée associe la longueur, exprimée en mètres, du fil de cette bobine. On admet que X suit la loi normale de moyenne 50 et d'écart type 0,2. On prélève au hasard une bobine dans la production d'une journée. Tous les résultats approchés seront arrondis à 10^{-2}.

1° Calculer la probabilité de chacun des événements suivants :

a) la longueur du fil de la bobine est inférieure à 50,19 m ;

b) la longueur du fil de la bobine est supérieure à 50,16 m ;

c) la longueur du fil de la bobine est comprise entre 49,81 m et 50,19 m.

2° Un réglage de la machine permet de modifier l'écart type sans changer la moyenne. Quel doit être le nouvel écart type σ' pour que la probabilité que la longueur du fil de la bobine comprise entre 49,81 m et 50,19 m soit égale à 0,75 ?

T.D.2. Echantillonage

Exercice 1

On considère une population de 4 objets de masses respectives en kilos : 48, 50, 53, 54.
On désigne par X la variable aléatoire qui, à chaque objet prélevé au hasard, associe sa masse et par $\overline{X}$ la variable aléatoire qui, à tout échantillon de 2 objets, prélevé au hasard de manière non exhaustive, associe la moyenne des masses des 2 objets de cet échantillon.
1° Calculer E(X) et σ(X).
2° En faisant figurer leurs masses sous forme de couples, donner l'ensemble de tous les échantillons et les valeurs prises par la variable $\overline{X}$.
3° Calculer $E(\overline{X})$ et $\sigma(\overline{X})$, vérifier que $E(\overline{X}) = E(X)$ et que $\sigma(\overline{X}) = \dfrac{\sigma(X)}{\sqrt{n}}$.

Exercice 2

Après la correction d'une épreuve d'examen comportant un grand nombre de candidats, on constate que les notes ont pour moyenne 12 et pour écart type 3.
Soit $\overline{X}$ la variable aléatoire qui, à tout échantillon aléatoire et non exhaustif de taille 100, associe la moyenne des notes de cet échantillon.
On suppose que $\overline{X}$ suit la loi normale N(12 ; 0,3).
On se propose de prélever un échantillon aléatoire non exhaustif de 100 notes.
1° Quelle est la probabilité d'avoir la moyenne d'un tel échantillon supérieure à 12,5?
2° Quelle est la probabilité d'avoir la moyenne d'un tel échantillon comprise entre 12.5 et 12,9?

Exercice 3

Une machine fabrique des pièces en grande série. À chaque pièce tirée au hasard, on associe sa longueur exprimée en millimètres ; un définit ainsi une variable aléatoire X.
On suppose que X suit la loi normale N(m, σ), où m = 28,20 et σ = 0,027.
On admet que la variable aléatoire M qui, à tout échantillon aléatoire non exhaustif de taille n, associe la moyenne des longueurs des n pièces de l'échantillon, suit la loi normale $N\left(\mathrm{m}, \dfrac{\sigma}{\sqrt{n}}\right)$
Déterminer la valeur minimale de *n* pour que :
$P(28{,}195 \leq M \leq 28{,}205) \geq 0{,}95$.

Exercice 4

Dans une population, on constate qu'il naît 52 % de garçons et 48 % de filles.
On suppose que la variable aléatoire F qui, à tout échantillon de taille n = 400 prélevé au hasard et avec remise dans la population, associe le pourcentage de garçons dans cet échantillon, suit la loi normale $N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$, où $p = 0{,}52$.
On se propose de prélever un échantillon aléatoire non exhaustif de 400 nouveau-nés.
1° Déterminer, à 10^{-2} près, la probabilité d'avoir, dans un tel échantillon, un pourcentage de garçon compris entre 50 % et 54 %.
2° Déterminer, à 10^{-2} près, la probabilité d'avoir, dans un tel échantillon, un pourcentage de filles inférieur à 45 %.

Exercice 5

Un candidat A a obtenu 55 % des suffrages exprimés à une élection. Soit F la variable aléatoire qui, à tout échantillon de taille n = 100 prélevé au hasard et avec remise dans l'ensemble des suffrages exprimés, associe le pourcentage de voix obtenu par le candidat A dans cet échantillon. On suppose que F suit la loi normale $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ où $p = 0{,}55$.

1°Quelle est la probabilité, à 10^{-3} prés, d'avoir, dans un échantillon aléatoire non exhaustif de taille n = 100 prélevé parmi les suffrages exprimés, strictement moins de 50 % de voix pour le candidat A ?
2° Reprendre la question précédente avec n = 2000. Comparer avec le résultat de la question précédente et donner une conclusion.

Exercice 6

Une machine produit des rondelles métalliques en grande série. On suppose que la variable aléatoire qui associe à chaque rondelle prélevée au hasard dans la production son diamètre extérieur, exprimé en mm, suit la loi normale de moyenne m= 22 et d'écart type σ = 0.05.
On considère des échantillons de 100 pièces prélevés au hasard que l'on assimile à des échantillons non exhaustifs.
On définit une variable aléatoire $\overline{X}$ en associant à chaque échantillon de ce type la moyenne des diamètres des 100 rondelles.
a) Quelle est la loi suivie par $\overline{X}$? Quels sont les paramètres de cette loi ?
b) Calculer, avec la précision permise par la table du formulaire, la probabilité que la moyenne d'un tel échantillon soit supérieure à 22,01 mm.

Exercice 7

Une machine est chargée de conditionner des paquets de farine. On désigne par M la variable aléatoire qui à chaque paquet prélevé au hasard dans la production, associe sa masse exprimée en grammes ; M suit une loi normale d'écart type constant, σ = 30, et dont la moyenne m peut être modifiée.
Un paquet est refusé si sa masse est inférieure à 955 grammes.
Afin de diminuer le nombre de paquets refusés on décide de modifier le réglage de la machine.
1° Quelle doit être la valeur de m pour que la probabilité d'accepter un paquet soit égale à 0,99 ?
2° La machine est réglée de telle sorte que m = 1025. Soit $\overline{X}$ la variable aléatoire qui, à tout échantillon de 40 paquets, associe la moyenne des masses des 40 paquets. On assimile ces échantillons de 40 paquets à des échantillons aléatoires prélevés avec remise.
a) Quelle est la loi suivie par la variable aléatoire $\overline{X}$?
b) Déterminer un intervalle centré en m tel que la probabilité que $\overline{X}$ appartienne à cet intervalle soit 0,95.

T.D.3. Estimation

Exercice 1

Une entreprise utilise des camions pour transporter sa production. Elle dispose de 100 camions. Elle repère sur un échantillon de 30 jours choisis au hasard le nombre de camions en panne. Voici les résultats :

5,5,6,4,6,6,8,3,5,5,5,4,3,6,5,6,4,7,6,6,5,4,3,6,5,4,5,4,5,5.

1° Calculer la moyenne $\bar{x}$ et l'écart type s du nombre de camions en panne chaque jour pour l'échantillon étudié.

2° À partir des résultats obtenus pour cet échantillon, proposer une estimation ponctuelle de la moyenne m et de l'écart type σ du nombre de camions en panne chaque jour pour la population correspondant aux jours ouvrables de l'année.

3° On suppose que la variable aléatoire $\overline{X}$, qui à tout échantillon de taille 30 prélevé au hasard et avec remise associe la moyenne du nombre de camions en panne chaque jour, suit la loi normale $N\left(\mathrm{m}, \frac{\sigma}{\sqrt{n}}\right)$.

On prend pour valeur de σ l'estimation ponctuelle obtenue au 2°. Déterminer un intervalle de confiance de la moyenne m de la population avec le niveau de confiance 95 %.

Exercice 2

Les ampoules électriques d'un certain modèle ont une durée de vie exprimée en heures, dont la distribution est normale d'écart type σ = 200 heures. Avec cette hypothèse, on se propose d'estimer la moyenne m de la durée de vie des ampoules de la production à partir d'un échantillon de 36 ampoules dont la moyenne des durées de vie est égale à 3 000 heures.
On assimile cet échantillon à un échantillon prélevé au hasard et avec remise parmi la production.
On rappelle que la variable aléatoire $\overline{X}$ qui, à tout échantillon d'effectif n = 36 ampoules du modèle considéré, prélevé au hasard et avec remise, associe la moyenne de leurs durées de vie, suit la lui normale $N\left(\mathrm{m}, \frac{\sigma}{\sqrt{n}}\right)$.

1° Déterminer une estimation de m par un intervalle de confiance avec le niveau de confiance 95 %.

2° Même question avec le niveau de confiance 90%, puis avec le niveau de confiance 99 %.

3° Qu'observe-t-on sur les intervalles de confiance de la moyenne m de la population obtenus partir d'un même échantillon lorsque le niveau de confiance varie ?

4° Peut-on situer exactement la position de la moyenne m par rapport à l'intervalle obtenu à la question 1° ?

Exercice 3

Dans une entreprise fabriquant des accessoires pour micro-ordinateurs, on a mesuré le temps d'accès (en millisecondes) d'un échantillon de 15 lecteurs CD-Rom. Ces mesures ont été obtenues à l'aide utilitaire comprenant un logiciel d'évaluation des caractéristiques matérielles (taux de transfert, temps d'accès,...). On a calculé un temps moyen d'accès de 298,4 ms avec un écart-type de 22,9 ms.
En supposant que le temps d'accès est distribué selon une loi normale, estimer par intervalle de confiance le temps moyen d'accès des CD-Rom de ce type avec les niveaux de confiance de 90%, de 95% et de 99%.

Exercice 4

Dans une école de 300 élèves, on choisit 17 élèves au hasard (tirage sans remise). Leurs résultats au test D 48 sont les suivants ; ils sont exprimés sous forme de QI :109, 97, 92, 122, 103, 98,107, 138, 111, 95, 101,102,104, 95, 115, 107, 103. Etablir un intervalle de confiance à 95% pour le QI moyen des élèves de l'école.

Exercice 5

La directrice des ressources humaines de l'entreprise Giscom a effectué un sondage auprès de 100 employés prélevés au hasard et sans remise parmi les 500 employés de l'entreprise dans la catégorie « main-d'œuvre spécialisée » pour connaître leur préférence concernant une modification importante de la semaine de travail (4 jours de 10 heures au lieu de 5 jours de 8 heures). Sur les 100 employés interrogés, 58 étaient en faveur du nouvel horaire de travail.

a) Quelle est l'estimation ponctuelle de la proportion p pour l'ensemble des employés de l'entreprise de cette catégorie en faveur de ce nouvel horaire de travail?
b) Quel est taux de sondage ? Est-ce nécessaire de faire intervenir le facteur de correction dans le calcul de l'écart-type de la proportion?
c) Calculer, avec un niveau de confiance de 99%, l'intervalle de confiance associé à l'estimation de p.
d) Quelle est la marge d'erreur associée à l'estimation de p pour ce sondage?

Exercice 6

Sur une portion de route où la vitesse des véhicules est limitée à 90 km/h, on effectue un contrôle des vitesses avec un instrument de mesure de grande précision.
On mesure la vitesse (en km/h) d'un véhicule sur vingt et on obtient les résultats suivants pour un échantillon de 100 véhicules que l'on assimile à un échantillon obtenu par prélèvement aléatoire avec remise :

Vitesse (en km/h)	Effectif
[75, 80[	5
[80, 85[	10
[85, 90[	20
[90, 95[	36
[95, 100[	15
[100, 105[	8
[105, 110[	6

1° En supposant que les valeurs observées sont celles du centre de la classe, calculer à 10^{-2} prés, la moyenne $\overline{x}$ et l'écart type s des vitesses pour cet échantillon en admettant que, dans chaque classe, tous les éléments sont au centre.
2° À partir des résultats obtenus pour cet échantillon, proposer une estimation ponctuelle de la moyenne m et de l'écart type σ des vitesses des 2000 véhicules de la population observée.
3° On suppose que la variable aléatoire $\overline{X}$ qui, à tout échantillon de taille n = 100 obtenu comme précédemment, associe la moyenne des vitesses de l'échantillon suit une loi normale. Déterminer un intervalle de confiance de la vitesse moyenne m de la population avec le niveau de confiance 99 %.
4° Quelle doit être la taille minimale n de l'échantillon pour connaître, avec le niveau de confiance 95 %, la vitesse moyenne de la population à 0.5 km/h près?

T.D.4. Tests paramétriques

Exercice 1

Sur une portion d'autoroute où la vitesse des automobiles est limitée à 130 km/h, on effectue un contrôle des vitesses sur un échantillon de 72 automobiles que l'on assimile à un échantillon obtenu par prélèvement aléatoire et avec remise dans la population constituée de toutes les voitures circulant à cet endroit dans l'intervalle de temps considéré. On obtient les résultats suivants :

Vitesse(en km/h)	Effectif
[90,100[	1
[100, 110[	3
[110, 120[	6
[120, 130[	16
[130, 140[	24
[140, 150[	13
[150, 160[	9

On note m et σ la moyenne et l'écart type des vitesses de l'ensemble de la population considérée. On suppose que la variable aléatoire $\overline{X}$ qui, à tout échantillon de taille n = 72 prélevé comme celui-ci, associe sa vitesse moyenne, suit la loi normale $N\left(\mathrm{m}, \frac{\sigma}{\sqrt{n}}\right)$.

A. Test bilatéral

1° Construire un test bilatéral permettant de décider si, au seuil de 5%, la vitesse moyenne de toutes les automobiles circulant à cet endroit et à ce moment est égale à 130 km/h.
Utiliser ce test avec l'échantillon de l'énoncé.

2° Même question au seuil de 1 %.

B. Test unilatéral

1° Construire un test unilatéral permettant de décider si, au seuil de 5 %, la vitesse moyenne des mêmes automobiles est strictement supérieure à 130 km/h.
Utiliser ce test avec l'échantillon de l'énoncé.

2° Même question au seuil de 1 %.

Exercice 2

Une grande surface vend du matériel photographique. On note X la variable aléatoire qui, à chaque ticket de caisse prélevé au hasard dans le stock de tickets d'un mois associe son montant en euros. On note m et σ la moyenne et l'écart type des montants de la population de tickets.
À la suite d'une campagne promotionnelle d'un concurrent, le responsable de la grande surface redoute que le montant moyen des tickets (donc des ventes) soit modifié.
Afin de contrôler que la moyenne m des achats reste égale à 550 euros, il se propose de construire un test d'hypothèse unilatéral. Le seuil de signification du test est fixé à 0,05.
On désigne par $\overline{X}$ la variable aléatoire qui, à chaque échantillon aléatoire de 50 tickets de caisse, associe la moyenne des montants en euros de ces tickets (le stock de tickets est assez important pour qu'on puisse assimiler des prélèvements à des tirages avec remise).

1) Donner les hypothèses H_0 et H_1.

2) Déterminer la région critique et formuler la règle de décision du test.

3) Appliquer le test sachant que la moyenne des montants d'un échantillon aléatoire de 50 tickets prélevés après la campagne du concurrent est $\bar{x} = 530$ et $s = 195$.

Exercice 3

1) Un échantillon provenant d'une population normale se caractérise ainsi : $n = 17$, $\bar{x} = 10$, s = 10. La moyenne théorique serait-elle égale à 13 (risque 5 %) ?

2) Même question si $n = 170$.

Exercice 4

Deux entreprises A et B fabriquent des ampoules électriques d'un même modèle.
Un échantillon de taille $n = 100$ ampoules de la première entreprise a donné une durée de vie moyenne $\bar{x}_1 = 94$ h et un écart type $s_1 = 12$ h. Un échantillon de taille $n_2 = 200$ ampoules de la seconde entreprise a donné une durée de vie moyenne $\bar{x}_2 = 90$ h et un écart type $s_2 = 16$h.

On note $\overline{X}_1$ la variable aléatoire qui, à chaque échantillon de 100 ampoules de la première entreprise, associe sa moyenne et $\overline{X}_2$ la variable aléatoire qui à chaque échantillon de 200 ampoules de la seconde entreprise associe sa moyenne.
Tous les échantillons considérés sont supposés prélevés au hasard et avec remise.
On suppose que les variables aléatoires $\overline{X}_1$, $\overline{X}_2, D = \overline{X}_1 - \overline{X}_2$ suivent les lois normales de moyennes respectives m_A, m_B, m_A-m_B inconnues et on estime l'écart type de D par $\sqrt{\dfrac{\hat{s}_1^2}{100} + \dfrac{\hat{s}_2^2}{200}}$ où $\hat{s}_1$ et $\hat{s}_2$ sont les estimations des écarts types pour les deux populations des ampoules fabriqués respectivement par A et B.

A. Test bilatéral
1° Construire un test bilatéral permettant de décider s'il y a une différence significative au seuil de 5 % entre les durées de vie des ampoules fabriquées par les entreprises A et B.
Utiliser ce test avec les échantillons de l'énoncé.
2° Reprendre la même question avec le seuil de 1 %.

B. Test unilatéral
1° Construire un test unilatéral permettant de décider si les durées de vie des ampoules fabriquées part l'entreprise A sont, au seuil de 5 %, significativement supérieures à celles de l'entreprise B.
Utiliser ce test avec les échantillons de l'énoncé.
2° Reprendre la question 1° avec le seuil de 1 %.

T.D.4. Tests paramétriques (suite)

Exercice 5

Un candidat à une élection fait effectuer un sondage dans sa circonscription comportant 185842 électeurs : sur 1068 personnes interrogées, 550 déclarent vouloir voter pour ce candidat.
On suppose que cet échantillon peut être assimilé à un échantillon prélevé au hasard dans la population des électeurs de la circonscription. Soit F la variable aléatoire, qui, à tout échantillon de taille n = 1068 prélevé au hasard dans cette population, associe le pourcentage d'électeurs de cet échantillon voulant voter pour le candidat. On suppose que F, suit a loi normale $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$, où p est le pourcentage inconnu des électeurs de la circonscription voulant voter pour le candidat.

1° Construire un test permettant d'accepter ou de rejeter, au seuil de signification de 5 %, l'hypothèse selon laquelle, au vu d'un sondage portant sur 1068 personnes interrogées, le candidat sera élu au premier tour, c'est-à-dire :
a) Énoncer une hypothèse nulle H_o et une hypothèse alternative H_1 pour ce test unilatéral ;
b) Déterminer le nombre réel positif a tel que, sous l'hypothèse H_o $P(F \leq a) = 0,95$;
c) Énoncer la règle de décision du test :
d) Utiliser ce test avec l'échantillon de l'énoncé.

2° Même question avec le seuil de signification de 1 %.

Exercice 6

Une entreprise s'interroge sur le lancement d'un produit nouveau auprès d'une population fixée. Au moment où un pourcentage inconnu p_1 de cette population est favorable à cette nouveauté, un premier sondage effectué sur un échantillon de 100 personnes donne 25 % de réponses favorables. Un mois plus tard, alors que le pourcentage inconnu de la population favorable à ce produit est devenu p_2, un second sondage effectué auprès de 400 personnes donne 33 % de réponses favorables.

On suppose que tous les échantillons intervenant ici peuvent être considérés comme prélevés au hasard et avec remise et que la variable aléatoire F_1, (resp. F_2) qui, à un tel échantillon de taille n_1 = 100 (resp. n_2 = 400), associe son pourcentage de réponses favorables, suit la loi normale $N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right)$ (resp. $N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right)$)
On suppose que F_1 et F_2 sont indépendantes.
Peut-on, au seuil de signification de 5 %, considérer que le pourcentage de réponses favorables sur l'ensemble de la population n'a pas changé d'un sondage à l'autre?

Exercice 7

La Mauricienne, compagnie d'assurances générales, veut améliorer le temps de réponse du personnel affecté au service d'indemnisation. Les clients ont accès au service à l'aide d'un numéro 800 et l'information est gérée à l'aide d'une base de données.
Selon la responsable du service, le temps moyen de réponse à l'aide de cette base de données est de 14 secondes. Un nouveau programme de gestion de base de données a été mis en application depuis quelques semaines et on aimerait vérifier s'il y a eu une diminution significative du temps moyen de réponse du personnel. Un échantillon aléatoire de 32 appels a permis d'obtenir les résultats ci-dessous concernant le temps de réponse.

Temps de réponse							
13,37	13,09	13,48	13,64	13,51	13,59	13,43	13,45
13,06	13,98	14,01	13,09	13,52	13,97	13,33	13,66
13,24	13,09	13,48	13,34	13,56	13,29	13,23	13,46
13,85	13,09	13,32	13,64	13,11	13,38	13,44	13,41

a) Préciser les hypothèses statistiques qui sont appropriées ici.
b) Est-ce que l'hypothèse de normalité du temps de réponse est requise ici pour exécuter le test d'hypothèse? Expliquer.
c) Donner la règle de décision à adopter pour un échantillon de taille n = 32 et un seuil de signification α = 0,01, selon les hypothèses statistiques précisées en a).
d) Peut-on conclure que ce nouveau programme de gestion de base de données a permis de réduire de façon significative le temps moyen de réponse? Justifier votre conclusion.

Exercice 8

Une entreprise A fabrique en grande série des pièces pour l'industrie. Pour analyser la qualité de la fabrication, on effectue un test bilatéral permettant, à la suite du prélèvement au hasard d'un échantillon de n_1 = 164 pièces dans la production, de tester au seuil de 5 %, l'hypothèse H_o selon laquelle le pourcentage de pièces défectueuses dans la production est 4%.
Soit F_1 la variable aléatoire qui, à tout échantillon de 164 pièces, associe le pourcentage de pièces défectueuses de cet échantillon.

1° Donner l'hypothèse alternative H_1.

2° Déterminer la région critique du test au seuil de 5%.

3° Énoncer la règle de décision du test.

T.D.5. Analyse de la variance

Exercice 1

Six milieux de culture d'une bactérie pathogène ont été réparés avec 6 pH différents. Pour tester l'influence du pH sur la croissance bactérienne, trois microgouttes d'une suspension de cette bactérie ont été séparément déposées sur chacun des milieux, et les colonies formées après incubation ont été dénombrées. Les résultats sont les suivants.

	pH 5,3	pH 5,8	pH 6,3	pH 7,0	pH 7,5	pH 8,1
Nombre	0	20	85	75	33	64
de	0	22	101	73	52	66
colonies	0	16	88	60	51	54

En admettant que les conditions de l'analyse de la variance sont réalisées, en déduire si l'effet du pH est significatif.

Exercice 2

Le tableau suivant représente les durées de vie de 3 échantillons de tubes de télévision de types différents de la même marque. Tester s'il y a, ou pas, différence au risque de 5%.

Echantillon 1	407	411	409		
Echantillon 2	404	406	408	405	402
Echantillon 3	410	408	406	408	

Exercice 3

Une expérience a été réalisée pour déterminer si la quantité de matière grasse absorbée par des beignets pendant leur cuisson dépend de la matière grasse utilisée. Pour chacune des quatre matières grasses, on a constitué six fournées de 24 beignets chacune. Les données du tableau suivant représentent les quantités en grammes de matière grasse absorbée par fournée.

1	2	3	4
164	178	175	155
172	191	193	166
168	197	178	149
177	182	171	164
156	185	163	170
195	177	176	168

T.D.6. Tests non paramétriques

Exercice 1

Un homme politique s'interrogeant sur ses chances éventuelles de succès aux élections présidentielles commande un sondage qui révèle que sur 2000 personnes, 19% ont l'intention de voter pour lui. Il demande alors à une agence de publicité de promouvoir son image. Un second sondage, réalisé après cette campagne publicitaire, auprès de 1000 personnes, montre que 240 d'entre elles ont l'intention de voter pour lui.
Peut-on considérer que cette campagne a été efficace ? Expliquer.

Exercice 2

Deux groupes A et B se composent chacun de 100 personnes atteintes d'une même maladie. On a administré un sérum au groupe A mais pas au groupe B que l'on appelle groupe de contrôle ; mais les deux groupes sont traités de la même façon. On a constaté que 75 malades du groupe A et 65 malades du groupe B ont guéri. Tester l'hypothèse que le sérum est une aide efficace dans la guérison de la maladie, en considérant un seuil de 5%.

Exercice 3

Une entreprise de service informatique veut ouvrir un bureau dans la région de Lanaudière. Un sondage a été commandé auprès d'une firme réputée pour connaître leur opinion concernant la situation économique de cette région auprès d'individus âgés de 18 ans et plus. La question posée était la suivante:
«*Selon vous, dans les 12 prochains mois, la situation économique va-t-elle s'améliorer, rester stable, se détériorer, ne sais pas*?»
Les résultats obtenus auprès de 250 répondants sont présentés dans le tableau croisé suivant selon la catégorie d'emploi.

	Profes-sionnels	Employés de bureau	Autres
S'améliorer	52	31	15
Reste stable	20	16	24
Se détériorer	13	14	21
Ne sais pas	8	17	19

a) Préciser l'hypothèse nulle et l'hypothèse alternative que l'on veut soumettre au test de khi-deux.
b) Déterminer la répartition théorique des répondants, sous l'hypothèse que les caractères «opinion sur la situation économique» et «catégorie d'emploi» sont indépendants.
c) Calculer l'ampleur de la disparité des deux distributions à l'aide de khi-deux.
d) En utilisant un seuil de signification de 5%, déterminer quelle hypothèse est favorisée.
e) Si les caractères «opinion sur la situation économique» et «catégorie d'emploi» s'avèrent liées, préciser le coefficient V de Cramer

Exercice 4

Soit A, B et C, trois marques de détergent à lessive. Historiquement, 30 % de la population utilise la marque A, 40 % utilise la marque B, 20 % utilise la marque C et le reste (10 %) n'utilise aucun de ces trois marques. Après avoir mené une campagne publicitaire intensive dans le but d'accroître sa part du marché, la compagnie qui fabrique la marque A a effectué un sondage auprès de 500 répondants et a obtenu les résultats suivants : 176 des répondants utilisent la marque A, 195 utilisent la marque B, 81 utilisent la marque C et 48 n'utilisent aucune des trois marques.
1) Au seuil critique $\alpha = 5$ % peut-on conclure que les parts de marché ont changé ? Justifier à l'aide du test du χ^2 d'ajustement.
2) Même question au seuil $\alpha = 1$ % .

Exercice 5

On a noté pour 25 employés le nombre de jours d'absentéisme de l'année écoulée. Les résultats sont regroupés dans le tableau :

21	23	33	32	37
40	37	29	23	29
24	32	24	46	32
17	29	26	46	27
36	38	28	33	18

Si l'on veut utiliser ces données, par exemple, pour calculer un intervalle de confiance pour le nombre moyen de jours d'absentéisme, il faut supposer que ce nombre suit une distribution normale dans la population. Vérifions cette hypothèse.

Classes (nombre de jours d'absentéisme)	Effectifs
$X \leq 25$	
$25 < X \leq 30$	
$30 < X \leq 35$	
$35 \leq X$	

a) Soit X une variable aléatoire qui associe à chaque employé le nombre de jours d'absentéisme de l'année écoulée. Compléter le dépouillement des données.

b) Peut-on considérer comme vraisemblable l'hypothèse selon laquelle les données concernant le nombre de jours d'absentéisme sont distribuées selon une loi normale? Utiliser un seuil de signification de 5 %.

c) Estimer (par intervalle de confiance) le nombre moyen de jours d'absentéisme dans la population des employés de l'entreprise (au niveau de confiance de 95%).

Exercice 6

L'entreprise Paramus fabrique des cartes d'interface SCSI pour un fabricant de dispositifs de surveillance. L'entreprise effectue un contrôle sur des échantillons de 20 cartes prélevées après chaque heure de production. Aucune non-conformité n'est tolérée. Aussitôt qu'une non-conformité est observée sur une carte, on la déclare non conforme. La répartition du nombre d'échantillons sans carte non conformes, avec 1 carte non conforme,..., 6 cartes non conformes par échantillon de 20, est présentée dans le tableau ci-contre. On a observé au total 80 échantillons de taille 20 sur une période de deux semaines.

Nombre des cartes non conformes	Nombre d'échantillons
0	13
1	21
2	19
3	12
4	9
5	4
6	2

La variable "nombre de cartes d'interface non conformes" correspond aux conditions d'application de la loi binomiale. On aimerait, à l'aide d'un test du χ^2 au seuil de signification $\alpha = 0.05$, tester l'hypothèse selon laquelle les observations (la distribution observée) se comportent d'après une loi binomiale.

Exercice 7

Les données suivantes (n = 84) représentent le nombre hebdomadaire de problèmes de systèmes pour un réseau informatique. Ces problèmes sont survenus suite à un changement de configuration du système.

Nombre hebdomadaire de problèmes de système :

3	3	0	0	4	3	5	6	0	1	0	2
0	2	3	3	0	2	3	3	2	1	0	0
2	0	2	3	0	2	1	0	1	0	2	2
0	0	0	0	0	6	0	0	0	0	2	6
3	0	0	4	3	0	0	2	3	3	3	0
3	3	0	0	0	0	1	0	1	1	0	0
4	1	1	1	3	1	0	0	6	4	3	1

Est-ce que ce relevé permet de supporter, au seuil de signification α=0,01, l'hypothèse selon laquelle le nombre hebdomadaire de problèmes de système se comporte d'après une loi de Poisson?

a) Compléter le dépouillement des données selon une distribution de fréquences.

Nombre de problèmes de système par semaine	Effectif observé
0	34
1	
2	
3	
4	
5	
6	
7 et plus	

b) Estimer le paramètre λ du loi de Poisson par $\hat{\lambda} = \bar{x}$

c) Appliquer le test d'ajustement.

T.D.7. Fiabilité

Exercice 1

On a mesuré pour 20 éléments du même type la durée de vie, en heures, avant la première défaillance.

Intervalle de temps	Nombre d'éléments défaillants dans cet intervalle	Instant t_i (en heures)	Nombre total n_i d'éléments défaillants à l'instant t_i	Estimation de $F(t_i)$ par $\frac{n_i}{20}$	Estimation de $R(t_i) = 1 - F(t_i)$
[0,500]	7	500	7	0,35	0,65
]500,1000]	4	1000	11	0,55	0,45
]1000,1500]	3	1500	14	0,70	0,30
]1500,2000]	2	2000	16	0,80	0,20
]2000,2500]	2	2500	18	0,90	0,10
]2500,3000]	1	3000	19	0,95	0,05
]3000,4000]	1	4000	20	1,00	0

a) A partir des valeurs observées sur un échantillon, représenter sur un graphique l'allure de la fonction de défaillance F(t) et la fonction de fiabilité R(t) obtenue à l'aide de la méthode des rangs bruts. Ensuite, calculer F(t_i) et R(t_i) à l'aide de la méthode des rangs moyens et représenter les graphiques correspondants. Comparer ces graphiques.

b) En utilisant l'estimation de F(t) selon la méthode des rangs moyen, calculer le taux d'avarie moyen pour chaque intervalle de temps considéré. Dans quel intervalle de temps le taux d'avarie moyen a été le plus élevé ?

c) Tracer le nuage de points M_i(t_i, R(t_i)) sur du papier semi-logarithmique; en déduire que la variable aléatoire mesurant la durée de vie des éléments suit approximativement une loi exponentielle. Déterminer à une heure près, graphiquement la MTBF. En déduire l'expression de R(t).

Exercice 2

On a relevé, durant une période de 1500 heures, la durée de vie de 24 éléments identiques, mis en service à la même heure. On a obtenu les résultats suivants :

Durée de vie (en heurs)	Nombre d'élément défaillant
[0, 100]	5
]100, 200]	4
]200, 300]	3
]300, 400]	3
]400, 500]	2
]500, 600]	1
]600,750]	2
]750, 1000]	2
[1000, 1500]	2

1° En utilisant la méthode des rangs moyens, compléter le tableau :

t_i	n_i	$F(t_i)$	$R(t_i)$

2° Tracer le nuage de points $M_i(t_i, R(t_i))$ sur du papier semi-logarithmique; en déduire que la variable aléatoire mesurant la durée de vie des éléments suit une loi exponentielle. Déterminer graphiquement la MTBF.

3° En déduire le paramètre et l'écart type de cette loi exponentielle, ainsi que l'expression de $R(t)$.

4° Déterminer graphiquement et par le calcul l'instant où la fiabilité d'un élément est 50 %. Étant donné qu'à l'instant t_o la fiabilité d'un élément est 50 %, déterminer, à cet instant, la fiabilité d'un système de deux éléments montés en parallèle (on admet que les deux éléments fonctionnent de façon indépendante).

5° On monte en série deux éléments. Montrer que la variable aléatoire mesurant la durée de vie du système obtenu suit une loi exponentielle dont on déterminera le paramètre. (On admettra encore l'indépendance du fonctionnement des deux éléments.)

Exercice 3

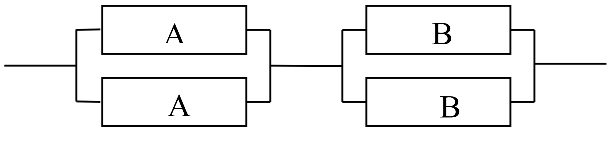
Soit T_1 et T_2 les variables aléatoires qui associent à des composants respectivement du type A et du type B tirés au hasard leur durée de vie exprimée en heures.
On admet que T_1 et T_2 suivent les lois exponentielles de paramètres respectifs :

$$\lambda_1 = 0{,}0012 \text{ et } \lambda_2 = 0{,}0007.$$

1° Quelle est, à 10^{-4} prés, la probabilité qu'un composant du type A soit encore en état de marche au bout de 1000 heures? Même question pour un type B.

2° Déterminer, à 1 jour prés, le nombre de jours au bout desquels 70 % des composants du type A ont eu leur première défaillance.

3° Deux composants du type A et deux composants de type B sont associés comme suit:



Déterminer la fiabilité d'un tel système au bout de 1000 heures.(Les défaillances des composants de type. A ou B sont indépendantes les unes des autres).

Exercice 4

Un technicien supérieur en maintenance a été chargé d'étudier plus particulièrement le cas de la pièce JB008. Son historique lui permet de connaître les durées de vie des pièces de ce type déjà utilisées. Elles sont consignées dans le tableau suivant.

N° d'ordre	Durée de vie (jours)	N° d'ordre	Durée de vie (jours)
1	32	11	460
2	45	12	139
3	120	13	58
4	210	14	323
5	70	15	160
6	21	16	379
7	85	17	183
8	10	18	241
9	277	19	102
10	599		

1° On note R(t) la probabilité de survie du matériel à la date t.
En utilisant la méthode des rangs moyens, compléter le tableau :

t_i	n_i : nombre total d'éléments défaillants à l'instant t_i	F(t_i)	R(t_i)

En utilisant une feuille de papier semi-logarithmique, justifier l'approche de R(t) par une loi exponentielle.
Déterminer, à une unité prés, graphiquement, la MTBF d'une pièce JB008. Montrer que l'on peut prendre pour valeur approchée du paramètre λ de la loi exponentielle la valeur 0,005.

2° Déterminer, à une unité prés, par le calcul à quel instant t_o la fiabilité de JB 008 est égale à 80 %. Comparer avec le résultat obtenu graphiquement.

3° On envisage de placer deux pièces JB 008 en parallèle, c'est-à-dire de telle sorte que le système fonctionne tant que l'une des deux pièces est en état de fonctionnement.
Étant donné qu'à l'instant t_o la fiabilité d'une pièce JB 008 est de 80%, déterminer, à cet instant, celle du système ainsi formé. On admet que les deux pièces fonctionnent de façon indépendante.

4° Quelle aurait été la fiabilité à l'instant t_o si on avait placé les deux pièces en série, c'est-à-dire de telle sorte que le système ainsi formé soit défaillant dés que l'une des deux pièces casse ? (On admettra encore l'indépendance du fonctionnement des deux pièces.)

T.D.8. Régression et corrélation

Exercice 1

Un chimiste veut étudier la perte de poids (en grammes) d'un certain mélange en fonction de la durée d'exposition à l'air (en heures).
Les données sont présentées dans le tableau ci-contre.

a) Dans cette étude, quelle est la variable dépendante ? La variable explicative ?
b) Déterminer la droite de régression et le coefficient de corrélation.
c) D'après la droite de régression, estimer la perte de poids pour des durées d'exposition de 3, 4, 5 et 7 heures.

Perte de poids	Durée d'exposition
2	3
2,5	4
3,1	5
3,6	6
1,8	3
2,4	4
3	5
3,4	6

Exercice 2

L'indice des dépenses de santé aux Etats-Unis entre1976 et 1984, sur la base d'un indice fixé à 100 en 1967 est donné par le tableau.

a) Représenter ces données dans le plan.
b) Donner la droite des moindres carrés correspondant à ces données.
c) Donner le coefficient de corrélation.
d) Donner une estimation de l'indice pour 1985 et comparer l'estimation à la valeur réelle qui est 396,6.

Année	Indice
1976	184,7
1977	202,4
1978	219,4
1979	239,7
1980	265,9
1981	294,5
1982	328,7
1983	357,3
1984	378

Exercice 3

Afin d'étudier les effets d'une station de lagunage sur un cours d'eau dans lequel elle se déverse, des variables physico chimiques et biologiques ont été mesurées. Des mesures de ont été effectuées à 12 dates de l'année. Les résultats concentration de l'eau en chlorophylle *a* (exprimés en $\mu g \cdot l^{-1}$) et la demande chimique en oxygène (DCO exprimée en mg d'oxygène par litre) sont les suivants :

Chl.*a* :	9,03	4,90	14,92	18,54	9,12	9,92	11,81	19,72	2,84	4,16	14,61	19,24
DCO :	43	32	43	42	26	26	27	38	29	28	37	41

a) Observe-t-on une corrélation entre les deux variables ?
b) Si cette corrélation est significative, quelle est la relation entre les deux paramètres dans l'optique de la prévision d'une des variables à partir de l'autre ?
c) Quelle proportion de variance d'une des variables est expliquée par sa corrélation avec l'autre ?
d) A l'aide de droites de régression calculer la valeur de la DCO pour une teneur en chlorophylle de 15 $\mu g \cdot l^{-1}$, et la valeur de la teneur en chlorophylle pour une DCO de 37 $mg \cdot l^{-1}$.

Exercice 4

On mesure le poids frais et le poids sec de 20 prélèvements de plancton. Les résultats sont les suivants (exprimés en g par 10 m^3 d'eau de mer).

Poids

frais	20,4	28,4	48,7	28,8	32,9	100,2	32,2	27,8	27	36,7	20,4	24,3	24,3	18	31,7	25,7	41,2	53	61	61,2
sec	3,6	3,4	5,6	4,1	3,3	9,3	3,7	3,2	2,9	4,5	2,6	6,5	3,1	2,6	4,4	2,8	4,6	6	7,2	6,3

Calculer le coefficient de corrélation linéaire entre le poids frais et le poids sec. Est-il significatif et à quel seuil ? Repérer deux *outsiders* parmi les couples de valeurs ; éliminer ces valeurs et reprendre la question.

Exercice 5

On étudie l'influence d'un antibiotique sur une culture bactérienne. On répartit dans 10 tubes des volumes égaux de culture additionnés d'une quantité X d'antibiotique, et on mesure, après incubation, la densité optique D . Les résultats sont les suivants.

X:	0,2	0,2	0,4	0,4	0,6	0,6	0,8	0,8	1,0	1,0
D:	19	21	35	38	64	66	115	130	200	210

a) Un ajustement linéaire semble-t-il justifié ?
b) En posant $Z = \ln D$, déterminer une équation de régression de Z en X .
c) En déduire une expression de D en fonction de X . Donner une prévision de D pour une quantité d'antibiotique $X = 0,5$.

T.D.8. Régression et corrélation (suite)

Exercice 6

Les données suivantes ont été obtenues à partir d'une étude sur les concentrations de mercure sous phase gazeuse présentes à l'intérieur d'un édifice à bureaux. Les variables considérées ici sont la concentration de mercure (en ng/m^3) qui est la variable dépendante et la température extérieure (en °C) qui est la variable indépendante ou explicative. Les mesures ont été effectuées au troisième étage d'un édifice à bureaux de 9 étages situé à ville St-Laurent; l'étude a été réalisée sur une période de 7 jours. Les conditions externes ont été obtenues d'Environnement Canada et sont constituées de moyennes sur l'heure. Voici une partie des résultats qui ont été observés au cours des quatre premiers jours de mesure.

Obs no.	Température	Concentration de mercure	Obs no.	Température	Concentration de mercure
1	25	43,9	23	19	32,5
2	18	31,7	24	21	34,6
3	19	30,1	25	15	21,1
4	22	45,2	26	24	47,2
5	20	41,4	27	21	44,2
6	21	45,4	28	13	22,5
7	13	22,3	29	22	40,3
8	22	38,2	30	12	20,1
9	14	22,9	31	13	25
10	12	20,1	32	22	45,2
11	19	31,4	33	14	20,4
12	12	19,1	34	12	20,2
13	22	42,3	35	20	30,4
14	23	45,5	36	18	30,5
15	23	47,2	37	12	17,8
16	23	45,1	38	16	30,7
17	13	28,4	39	24	43,4
18	12	20,1	40	22	40,2
19	22	42,3	41	25	41,2
20	23	44,2	42	24	42
21	22	38,3	43	20	34,4
22	15	24,8	44	13	26,2

Source: Adapté de Casimir A. ,Poissant L. et Béron P. (1996). Vapeur mercurielle à l'intérieur d'un édifice à bureaux: une étude de cas. Vecteur Environnement, vol 29, no 6, pp 31-37.

On veut dans un premier temps déterminer s'il existe une relation entre les vapeurs de mercure et les conditions météorologiques à l'extérieur de l'édifice.
Le travail suivant est requis:
a) Tracez le diagramme de dispersion en utilisant la température extérieure comme variable explicative et la concentration de mercure comme variable dépendante.
b) Que peut-on dire du nuage de points?
c) Peut-on affirmer au seuil de signification 5% qu'il existe, pour cet édifice, une corrélation positive entre la température extérieure et la concentration de mercure?
d) Déterminez l'équation de régression empirique.
e) Tracez cette droite dans le nuage de points.
f) Indiquez également dans le nuage de points l'équation de régression et le coefficient r^2.
g) Quelle interprétation peut-on donner au coefficient de détermination?

T.D.9. Séries chronologiques

Exercice 7

L'entreprise BLH se spécialise dans la fabrication de dispositifs électroniques de sécurité. Les ventes trimestrielles des 5 dernières années sont présentées ci-dessous. Le but d'exercice est de donner les prévisions pour l'année qui suit.

	Année 1	Année 2	Année 3	Année 4	Année 5
Trimestre 1	1291	1422	1515	1738	1778
Trimestre 2	1771	1782	1942	2057	2264
Trimestre 3	2006	2112	2233	2408	2597
Trimestre 4	1603	1658	1755	1925	2111

1) Première méthode
   a) Représenter graphiquement les données.
   b) Estimer la tendance générale à l'aide d'une droite de régression.
   c) Estimer les composantes saisonniers.
   d) Donner les prévisions pour l'année qui suit à l'aide du modèle additif. Représenter sur le même graphique la série initiale et les prévisions.

2) Deuxième méthode
   a) Représenter graphiquement les données.
   b) Lisser la série à l'aide de moyennes mobiles évaluées sur une période de 1 ans. Ajuster une courbe de tendance.
   c) Estimer les composantes saisonniers.
   d) Donner les prévisions pour l'année qui suit. Représenter sur le même graphique la série initiale et les prévisions. Comparer avec le résultat précédant.

Exercice 8

L'entreprise Sécuripak se spécialise dans l'installation de systèmes de sécurité dans les secteurs résidentiels et commerciaux. Le nombre d'installations effectuées lors des quatre dernières années est présenté dans le tableau ci-dessous.

a) Représenter graphiquement cette série.
b) Lisser la série à l'aide de moyennes mobiles évaluées sur une période de 3 mois.
c) Même question avec une période de 12 mois. Comparer avec le résultat précédent. Ajuster une courbe de tendance.
d) Donner les prévisions (à l'aide du modèle multiplicatif) du nombre d'installations qui serait requis pour la 5ème année?

	2003	2004	2005	2006
Janvier	85,5	101	118,2	130,4
Février	60,1	64,9	74,3	85,3
Mars	71,6	76,9	89,2	100
Avril	80,1	98,7	94,4	112,4
Mai	120,8	126,2	132,6	138,7
Juin	110,4	110,5	119,9	122,7
Juillet	94,9	108,1	114,2	110,5
August	138,7	144,1	167,9	195,7
Septembre	129,5	142,2	158,6	164,9
Octobre	144,9	159,8	186,1	210,2
Novembre	104,6	123,3	128,4	138,5
Décembre	100,9	112,8	132,6	128,2