TD 1
Introduction : Tests sur l'espérance d'une ou deux populations

Exercice 1

Un entreprise fabrique des cables dont la charge de rupture notée X est supposée suivre une loi $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma = 75kg$. La charge de rupture moyenne attendue est de $600kg$. L'entreprise a changé son procédé de fabrication et prétend maintenant la charge de rupture moyenne attendue est de $650kg$. Une association de consommateurs a vérifié la charge de rupture 9 cables indépendants et a trouvé une charge moyenne de rupture observée de $\overline{x} = 631kg$. On souhaite faire le test de $(H_0) : \mu = 600$ contre $(H_1) : \mu = 650$

1. Donner les deux erreurs que l'on peut commettre quand on effectue un test statistique.
2. Donner la définition du risque de première espèce.
3. Donner la définition du risque de seconde espèce.
4. Pour le test ci-dessus, donner la statistique de test, sa loi sous l'hypothèse nulle, la zone de rejet, ainsi que la règle de décision, au niveau $\alpha \in]0, 1[$.
5. Mettre en oeuvre le test aux niveaux $\alpha = 2.5\%$, $\alpha = 5\%$ et $\alpha = 10\%$. Commenter.
6. Donner la définition de la p-value (ou niveau de signification). Que peut-on dire de la p-value au vue des résultats de la question précédente ?
7. Calculer la p-value du test et conclure.
8. Donner la définition de la puissance d'un test. Calculer la puissance du test précédent, au niveaux $\alpha = 2.5\%$, $\alpha = 5\%$ et $\alpha = 10\%$. Commenter.
9. En prenant $\alpha = 5\%$, représenter (en les précisant) sur un même graphique les lois de $\overline{X}$, sous l'hypothèse (H_0) et sous l'hypothèse (H_1). On indiquera également où se lisent sur le graphique le niveau, la puissance et la p-value.

Exercice 2

Le temps de réaction moyen des souris d'un certain élevage à un test déterminé est de 19 minutes. On désire expérimenter un produit pharmaceutique sur ces souris. On administre à 8 d'entre elles une dose de ce produit et l'on observe les temps de réaction suivants (en minutes) :

$$15 \quad 14 \quad 21 \quad 12 \quad 17 \quad 12 \quad 19 \quad 18$$

On suppose les temps de réaction normalement distribués.

1. Ecrire une fonction `my.t.test` ayant comme arguments un vecteur d'observations, une valeur de référence μ_0 ainsi qu'une chaine de caractères pouvant prendre les valeurs `"two.sided"`, `"less"` ou `"greater"` et retournant la p-value du test de Student.
2. Utiliser la fonction créée pour tester l'action du produit.
3. Vérifier le résultat avec la fonction `t.test`
4. A l'aide de la fonction `power.t.test`, calculer la valeur de la puissance π pour diverses valeurs de μ_1, le paramètre sous l'hypothèse alternative. Tracer la courbe de puissance.

Exercice 3

La notice d'un sirop contre la toux indique comme valeur de référence pour la concentration de l'agent actif A dans une bouteille de sirop la quantité $m_0 = 40g/l$. Le contrôleur de la fabrication décidera d'arrêter provisoirement la production si la moyenne m inconnue est strictement inférieure à cette valeur de référence. Il souhaite ne prendre qu'un risque minime (de 0.01) en décidant d'arrêter à tort la production.
Le contrôleur de la fabrication prélève 9 bouteilles au hasard dans la production et mesure la quantité d'agent actif A. Les résultats pour les 9 dosages indépendants sont les suivants (en g/l) :

38,7 39.6 37.9 40.6 40.5 37.7 41.2 37.5 39.1

On suppose que la quantité d'agent actif A conditionné dans une bouteille de sirop suit une loi normale centrée sur la vraie valeur m.
Proposer un test de niveau 1% permettant de savoir quelle décision doit prendre le contrôleur de la fabrication.

Exercice 4

Une nouvelle technique de dosage du glucose sanguin vient d'être mise au point. Pour un même échantillon de sang, sept dosages effectués à l'aide de cette nouvelle technique ont donné les résultats suivants (en g/l) :

$$1.17 \quad 1.16 \quad 1.16 \quad 1.19 \quad 1.21 \quad 1.19 \quad 1.18$$

On admet que les sept mesures sont des variables aléatoires indépendantes et identiquement distribuées de loi $N(\mu, \sigma^2)$ où σ^2 caractérise la précision du procédé.

1. La technique utilisée jusqu'alors était caractérisée par un écart-type de 0.05 mg/l. Peut-on dire que la nouvelle technique est plus précise que l'ancienne ?
2. Déterminer un intervalle de confiance de μ de niveau 95%.

Exercice 5

On désire savoir si une certaine variété de blé a dans une région A un rendement supérieur à celui qu'elle a dans une région B. Pour cela, on dispose de résultats obtenus sur seize parcelles différentes (en quintaux par hectare) :

Région A	48.0	48.2	50.3	53.5	54.6	56.4	57.8	58.5	60.5
Région B	44.2	46.3	48.3	48.5	50.5	51.2	55.4		

On note X (resp Y) la variable aléatoire correspondant au rendement observable sur une parcelle de la région A (resp B). On admet que X et Y suivent des lois normales d'espérances μ_A et μ_B et de variances σ_A^2 et σ_B^2.

1. Montrer que, au niveau $\alpha = 5\%$, les variances σ_A^2 et σ_B^2 ne sont pas significativement différentes. Pour la suite, on notera σ^2 la valeur commune de ces deux variances.
2. On peut alors disposer de trois estimateurs sans biais de σ^2. Le premier est associé à l'ensemble des 9 rendements provenant de la région A. Le deuxième est associé à l'ensemble des 7 rendements provenant de la région B. Le troisième est associé à l'ensemble des 16 rendements. A partir de leur écart quadratique moyen relativement à σ^2, choisir le meilleur de ces 3 estimateurs. En déduire une estimation de σ^2.
3. Montrer que, au niveau $\alpha = 5\%$, on peut accepter l'hypothèse $\mu_A > \mu_B$.

Exercice 6

On veut étudier l'effet secondaire d'un certain médicament sur le taux de cholestérol. On administre le médicament à 10 personnes et on relève pour chacune de ces personnes le taux de cholestérol. On obtient pour cet échantillon un taux moyen de 2.4 g/l avec une variance empirique (non corrigée) de 12 $(g/l)^2$. Sur un autre échantillon de 8 personnes n'ayant pas été traitées avec ce médicament, on relève un taux moyen de 2 g/l avec une variance empirique (non corrigée) de 9 $(g/l)^2$. On suppose que le taux de cholestérol d'un individu est une variable aléatoire de loi gaussienne d'espérance μ_A et de variance σ_A^2 pour une personne ayant pris le médicament et μ_B et σ_B^2 pour une personne n'ayant pas pris le médicament.

1. Peut-on conclure à partir des données observées à une différence significative entre les variances σ_A^2 et σ_B^2 au niveau $\alpha = 5\%$?
2. Comparer μ_A et μ_B à l'aide d'un test au niveau $\alpha = 5\%$.

Exercice 7

Un laboratoire de recherche étudie, sur une nouvelle espèce de vers (C. marginalus), les gènes pouvant être impliqués dans la mort programmée. La distribution de la durée de vie de ces vers est inconnue mais des études antérieures ont montré que l'espérance de vie de ces vers est $\mu = 250$ heures et que l'écart-type de leur durée de vie est $\sigma = 24$ heures. Après modification d'un gène supposé intervenir dans la mort programmée, on a relevé sur un échantillon de 80 vers une durée de vie moyenne de 256 heures.

1. Peut-on conclure que la modification de ce gène induit une augmentation significative (au niveau 5%) de l'espérance de vie des vers ?
2. On suppose que la modification du gène induit une réelle augmentation de l'espérance de vie de 7 heures. Quelle est la puissance du test mis en oeuvre ?
3. Combien de vers faudrait-il inclure dans l'étude pour avoir une puissance de 90% ?

Exercice 8

On envisage d'ajouter un adjuvant au traitement usuel d'un certain type de rhumatisme. Sans adjuvant, la durée séparant deux crises de récurrence rhumatismale peut être modélisée par une variable aléatoire suivant une distribution normale d'espérance $\mu = 560$ (exprimée en jours). On administre le traitement avec adjuvant à 10 sujets. Les durées de récurrence observées sont les suivantes :

$$646, 573, 485, 752, 742, 636, 607, 665, 506, 575.$$

Au niveau $\alpha = 5\%$, l'adjuvant modifie-t-il significativement la durée moyenne de récurrence ?

Exercice 9

Un producteur de lait souhaite comparer le rendement moyen des vaches normandes et hollandaises de son unité de production. Pour ce faire, il a relevé la production de lait (exprimée en kg) de 10 vaches prises au hasard dans chaque groupe. On suppose que la production dans chaque groupe suit une distribution normale.

Normandes	552	464	423	506	497	544	486	531	496	501
Hollandaises	487	489	470	482	494	500	504	567	482	526

Conclure au vu de ces données.

Exercice 10

Un laboratoire pharmaceutique produit des tubes de pommade dont les poids suivent une distribution normale. On dispose de deux échantillons issus de 2 sites de production différents. Les poids sont donnés dans le tableau suivant :

										$\sum x_i$	$\sum x_i^2$
Echantillon 1	56.4	57.5	55.8	54.3	58.9	56.9	54.8	54.2	58.1	506.9	28572.45
Echantillon 2	54.6	58.2	60.3	59.5	61.1	58.7	59.8	57.5		469.7	27605.93

1. Les variances des 2 échantillons sont-elles significativement différentes ?
2. Le poids des tubes est-il significativement différent d'un site de production à l'autre ?

Exercice 11

On fait une numération globulaire à un groupe de 20 personnes à deux périodes différentes de l'année. Pour chaque sujet, on note les résultats des deux numérations (à multiplier par 10^5) :

Sujet	1	2	3	4	5	6	7	8	9	10
Janvier	46	38	42	47	48	40	40	43	42	49
Septembre	48	47	44	45	51	44	47	48	47	57

On suppose que les sujets sont mutuellement indépendants et suivent une loi gaussienne. Tester au niveau 0.05 l'hypothèse selon laquele les résultats de la numération sont les mêmes aux deux périodes.

Exercice 12

La quantité de bactéries par cm^3 de lait provenant de 8 vaches différentes est estimée juste après la traite et 24h plus tard. La distribution des résultats obtenus est supposée normale. Au niveau $\alpha = 5\%$, existe-t-il un accroissement significatif du nombre de bactéries par cm^3 de lait au cours du temps ?

Vache	1	2	3	4	5	6	7	8
Juste après la traite	12000	13000	21500	17000	15000	22000	11000	21000
24h après la traite	14000	20000	31000	28000	26000	30000	16000	29000