TD4 : Tests statistiques

Exercice 1 : Comparaison de populations corrélables

Pour comparer les rendements de deux variétés de blé A et B, on a ensemencé respectivement 10 couples des deux parcelles voisines, l'une en variété A, l'autre en variété B.

Les 10 couples sont répartis dans différentes localités. On a obtenu les résultats suivants (en tonnes) :

Couple	1	2	3	4	5	6	7	8	9	10
A	45	32	56	49	45	38	47	51	42	38
B	47	34	52	51	48	44	45	56	46	44
C										

On donne :
— les moyennes $m_A = 44.3$ et $m_B = 46.7$
— les écarts-types de $\sigma_A = 6.69$ et $\sigma_B = 5.60$
— la covariance $cov(A, B) = 33.29$.

On supposera que les données sont gaussiennes.

1. Donner l'effectif total n_{tot} et le comparer à 30.
2. Calculer le coefficient de corrélation linéaire $R = \frac{cov(A,B)}{\sigma_A \times \sigma_B}$.
3. En déduire que les données de la récolte A et celles de la récolte B sont corrélées.
4. On pose $C = A - B$. Compléter la troisième ligne du tableau.
5. Calculer la moyenne empirique (en tonnes) m_C.
6. Calculer l'écart-type empirique (en tonnes) σ_C :

   $\sigma_C = \sqrt{\frac{1}{n_{tot}-1} \sum_{i=1}^{n_{tot}} (C_i - m_C)^2}$

7. Calculer $Z = \frac{m_C \times \sqrt{n_{tot}}}{\sigma_C}$.

   D'après la question 1, n_{tot}<30. On admet alors que Z suit une loi de Student à $n_{tot} - 1$ degrés de liberté.

   On admet les probabilités suivantes : $P(Z > 2.262) = P(Z < -2.262) = 2.5\%$

8. Effectuer un test au risque 5% pour savoir si les deux variétés produisent autant.

   H_0 :

   H_1 :

Exercice 2 : Comparaison de populations non corrélables

Deux chaînes de production produisent des transistors : une ancienne et une nouvelle, installée il y a 10 jours.

On note m_1 (respectivement m_2) la production moyenne de transistors par jour de l'ancienne (respectivement la nouvelle) ligne.

On note σ_1 (respectivement σ_2) l'écart-type de la production de transistors par jour de l'ancienne (respectivement la nouvelle) ligne.

On admettra que les écarts-type σ_1 et σ_2 sont **inconnus** mais **égaux**.

On cherche à **savoir si la nouvelle ligne produit significativement plus que l'ancienne**, c'est-à-dire si $m_2 > m_1$, en étudiant deux échantillons.

Pour tester la nouvelle ligne, on a relevé sur les $n_1 = 30$ derniers jours la production de l'ancienne ligne. On a trouvé une moyenne de $\hat{m}_1 = 2680$ transistors produits par jour avec un écart-type de $\hat{\sigma}_1 = 101.5$.

Pour la nouvelle ligne, on a pu tester seulement sur $n_2 = 10$ jours et on a trouvé $\hat{m}_2 = 2800$ transistors produits par jour avec un écart-type $\hat{\sigma}_2 = 87.4$.

1. Calculer l'estimateur de l'écart-type $\hat{\sigma} = \sqrt{\frac{n_1 \times \hat{\sigma}_1^2 + n_2 \times \hat{\sigma}_2^2}{n_1 + n_2 - 2}}$
2. Calculer $Z = \frac{\hat{m}_2 - \hat{m}_1}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
3. Calculer $n_1 + n_2$ et comparer avec 30.

   On admet alors que Z suit une loi normale centrée réduite.

   On admet la probabilité suivante : $P(Z > 1.285) = 10\%$.

4. Effectuer un test au risque 10% pour savoir si la nouvelle ligne de production produit significativement plus que l'ancienne.

   H_0 :

   H_1 :

Exercice 3 : Test du χ^2, d'indépendance et corrélations

Vous réalisez un sondage pour savoir si une campagne de publicité a été efficace. Pour cela, vous collectez les données suivantes :

Pub vue \ Produit acheté	Oui	Non
Oui	87	112
Non	47	254

1. Quel est l'effectif total des données ? Que pouvez-vous dire des effectifs de chaque classe ?

2. Représenter les données dans un tableau de fréquences (ou de pourcentages).

3. Calculer les corrélations.

4. A partir du tableau des fréquences (ou pourcentages), calculer les effectifs théoriques n_{theo} quand l'achat du produit est indépendant du fait d'avoir vu la pub.

5. On souhaite réaliser un test du χ^2 pour déterminer au risque 5% si la pub a une influence sur l'achat du produit.

   H_0 :

   H_1 :

6. On rappelle que pour une classe c donnée $\chi^2_c = \frac{(n_c - n_{c,theo})^2}{n_{c,theo}}$.

   Calculer le χ^2_c pour chaque classe.

7. Additionner tous les χ^2_c obtenus.

   La loi suivie par le χ^2_{tot} total ainsi obtenu est une loi du χ^2 à 1 degré de liberté.

   On admet la probabilité suivante : $P(\chi^2 > 3.841) = 5\%$.

8. Déterminer au risque 5% si la pub a une influence sur l'achat du produit.

9. Dans le cas où la pub a une influence, déterminer si l'influence est bonne (i.e. la pub augmente les ventes) ou néfaste (i.e. la pub diminue les ventes).