

Statistiques décisionnelles : TD 11

Exercice 1. Des enseignants de L2 maths cherchent à déterminer si le groupe de TD a une influence sur les résultats à l'examen. Pour chacun des quatre groupes de TD, on a calculé les quantités suivantes

Groupe	Effectif	Moyenne du groupe	Sommes des carrés des écarts
Groupe 1	25	13	342
Groupe 2	27	12.4	277
Groupe 3	17	9.8	204
Groupe 4	25	13	217

Conclusion au niveau 5% ?

S'il n'y avait eu que 2 groupes de TD, qu'auriez-vous fait ? C'est l'occasion de se rendre compte qu'une anova à un facteur pour 2 groupes et un test de student sont équivalents.

Exercice 2. Des professeurs d'EPS se posent la question suivante : est-ce que la couleur de cheveux a un impact sur le temps de course au 100m ? Ils décident donc de mesurer les temps de course de tous les élèves d'une même année et ils modélisent les scores obtenus comme des variables gaussiennes. Les groupes étant comparables en poids et tailles, ils estiment qu'il est raisonnable de modéliser ces temps de course par des variables gaussiennes, toutes de même variance. Comme les données sont longues à écrire, ils ont calculé les quantités suivantes :

Couleur de cheveux	Effectif	Moyenne	variance empirique
Blond	30	16.75	7.01
Brun	27	17.41	12
Châtain	25	16.9	11
Roux	35	17.15	4.73

Conclusion au niveau 5% ?

Exercice 3 (Exercice ouvert). De retour dans votre lycée peu après la fin de votre master, vous présentez votre formation à vos anciens professeurs. Intrigués, il vous demandent de leur faire une démonstration en vous donnant accès à toutes les dossiers et à toutes les notes du lycée pour que vous fassiez une démonstration de ce que vous savez faire¹.

Que faites-vous ? Pour chaque question, vous devrez le formuler sous forme mathématique et choisir le test le plus pertinent. Vous devez trouver des questions faisant appel à un minimum de cinq tests différents.

Exercice 4 (Exercice ouvert). Votre entreprise vous a chargé de déterminer si, parmi ses employés, les hommes gagnent plus que les femmes à travail égal. Pour cela, le PDG vous a donné l'accès à l'intégralité des données RH de l'entreprise. Vous disposez, pour chaque employé d'énormément de données, comme par exemple sa catégorie (disons ici commercial, ingénieur, technicien, assistant...), son ancienneté de l'entreprise, son salaire...

1. Bien entendu, c'est totalement fictif. Les dossiers étudiants sont absolument confidentiels.

Détaillez plusieurs stratégies de réponse, du plus simple au plus élaboré.

Dans un second temps, le PDG souhaite aussi savoir si les hommes et les femmes sont "bien" répartis sur les postes à responsabilité. Comment vous-y prenez-vous ?

Bonus : En général, le gender pay gap est calculé comme la différence des salaires médian ou moyen entre les hommes et les femmes, divisé par le salaire des hommes. Quels sont les avantages et inconvénients d'une telle définition ?