T. D. n° 2
Analyse de variance à un facteur : quelques cas atypiques

Les deux premiers exercices sont issus du livre d'exercices de Pierre Dagnelie intitulé Statistique théorique et appliquée*, Tome 2, éditions De Boeck.*
Le troisième exercice est issu du livre d'exercices de François Husson et de Jérôme Pagès intitulé Statistiques générales pour utilisateurs*, éditions PUR.*

Exercice 1. Quinzaine de veaux

Quinze veaux ont été répartis au hasard en trois lots, alimentés chacun de façon différente. Les gains de poids, observés durant la même période et exprimés en kilogrammes, sont présentés ci-dessous, une donnée étant manquante.

Alimentation 1	42,1	37,7	45,1	43,2	
Alimentation 2	45,2	54,2	38,1	48,3	55,1
Alimentation 3	48,3	44,1	56,9	42,2	54,0

Doit-on considérer que les différences de moyennes entre alimentations sont significatives ? Dans l'affirmative, estimer ces différences de moyennes, et déterminez-en les limites de confiance.

Exercice 2. Rendements fouragers

On s'intéresse à l'ensembe des prairies d'une région donnée et on souhaite identifier l'importance, absolue ou relative, de la variabilité de la production fourragère, d'une part, d'une prairie à l'autre, et d'autre part, d'un endroit à l'autre, à l'intérieur des différentes prairies. Dans ce but, on a tout d'abord choisi au hasard trois prairies, dans l'ensemble du territoire, puis au sein de chacune de ces trois prairies, cinq petites parcelles, de deux mètres carrés. Dans l'optique d'un échantillonage à deux degrés, les trois prairies constituent trois unités du premier degré, et les quinze petites parcelles quinze unités du deuxième degré.

Dans chacune des petites parcelles, on a mesuré les rendements en matière sèche à une date donnée. Les valeurs observées, exprimées en tonne par hectare, figurent dans le tableau ci-dessous.

	Prairie 1	Prairie 2	Prairie 3
Parcelle 1	2,06	1,59	1,92
Parcelle 2	2,99	2,63	1,85
Parcelle 3	1,98	1,98	2,14
Parcelle 4	2,95	2,25	1,33
Parcelle 5	2,70	2,09	1,83

Les rendements sont-ils homogènes ?

Exercice 3. De l'analyse de la variance à la comparaison multiple de moyennes

On s'intéresse à l'efficacité de cinq insecticides pour lutter contre des chenilles arpenteuses (chenilles de lépidoptères de la famille des *Geometridae*). Pour cela, on dispose de vingt parcelles de cinquante choux. On applique chacun des cinq insecticides sur sur quatre parcelles. L'efficacité des insecticides est évaluée en fonction du nombre de chenilles comptabilisées dans chaque parcelle : les valeurs observées, exprimées en dizaine de chenilles, figurent dans le tableau ci-dessous. On pourra considérer ces valeurs comme issues de réalisations d'une variable aléatoire quantitative continue.

Insecticide 1	Insecticide 2	Insecticide 3	Insecticide 4	Insecticide 5
2	2	4	4	10
4	4	6	7	13
6	5	8	9	14
7	11	11	14	18

1. Proposer un modèle puis un test (hypothèses, statistique du test, loi de la statistique du test sous $\mathcal{H}_0$) permettant de conclure à l'existence d'un effet traitement. Au vu du tableau ci-dessus, conclure (au seuil 5 %).

On veut maintenant voir s'il existe une différence entre les traitements 4 et 5. Soit D la différence entre les moyennes des traitements 4 et 5. On réalise d'abord le test usuel de comparaison de deux moyennes (avec variances inconnues mais égales à σ^2).

2. Donner une estimation de la variance commune aux deux traitements et en déduire une estimation de la variance de la différence de moyenne $D = \overline{Y_5} - \overline{Y_4}$ (notée σ_D^2) ? Peut-on considérer que les traitements 4 et 5 ont la même efficacité (hypothèses, statistique du test, loi de la statistique du test sous $\mathcal{H}_0$, décision au seuil 5 %).

Dans la question précédente, la variance commune a été estimée à partir des seuls traitements 4 et 5. Or, dans l'analyse de la variance à un facteur, on a supposé les cinq variances des traitements égales. Ceci suggère d'estimer la variance commune σ^2 à partir de l'ensemble des traitements, même pour comparer les moyennes de seulement deux traitements. L'objectif des question suivantes est de construire le test d'égalité des efficacités des traitements 4 et 5 avec cette nouvelle estimation.

3. Donner une estimation de cette variance commune fondée sur les vingt parcelles. En déduire une estimation de σ_D (écart-type de la différence entre la moyenne $\overline{X_4}$ et la moyenne $\overline{X_5}$). Quel est l'intérêt de cette nouvelle estimation ?
4. Quel est le nombre de degrés de liberté ν associé à la variance commune calculée à la question précédente ? En déduire la loi de $D/\hat{\sigma}_D$ sous $\mathcal{H}_0$. Quelle décision prenez-vous au seuil 5 % ?

Pour réaliser de manière systématique le test d'égalité des moyennes de deux populations, on peut procéder comme suit :

- On détermine un seuil, appelé $PPDS$ (Plus Petite Différence Significative), calculé de la manière suivante : $PPDS = t_\nu(1-\alpha/2)\hat{\sigma}_D$ où $t_\nu(1-\alpha/2)$ est le quantile d'ordre $1-\alpha/2$ de la loi de Student à ν degrés de liberté.

- Si la valeur absolue de la différence des moyennes des échantillons est inférieure à la $PPDS$, alors on considère que les moyennes des deux populations sont égales (on les considère différentes sinon).

5. Calculer la $PPDS$ dans cet exercice. Quels sont les traitements aussi efficaces que le traitement 4 (au seuil 5 %) ?