**1- Analyse de la variance (ANOVA un facteur)**

L’analyse de variance ANOVA à 1 facteur permet de tester l’hypothèse d’égalité des
moyennes pour plus de deux échantillons. Le but sera toujours le même : vérifier l’hypothèse
nulle que les moyennes des groupes proviennent d’une même population. Pour ce faire, nous
allons utiliser l'analyse de variance univariée (ANOVA).

Les conditions du test :
Tout comme pour les autres tests d’hypothèse, il faut s’assurer de respecter certaines prémisses avant de procéder à l’analyse proprement dite :
- Les groupes sont indépendants et tirés au hasard de leur population respective
Ceci signifie qu’il n’y a ni relation entre les observations à l’intérieur d’un groupe, ni relation
entre les observations entre les groupes.
- Les valeurs des populations sont normalement distribuées
- Les variances des populations sont égales lorsque les tailles des échantillons sont
inégales
Lorsqu’on rejette l’hypothèse nulle, le test d’ANOVA ne peut pas nous dire où se situe la
ou les différences. Il faut donc effectuer d’autres tests pour savoir entre quels groupes se
trouvent cette ou ces différences. Ces tests sont appelés post-hoc ou tests a posteriori. Ils
indiquent quels groupes se distinguent.
Exemple :
On compare trois nouveaux traitements pour la glycémie pour savoir qui est le plus efficace,
pour cela, on a pris au hasard trois échantillons de patients diabétiques afin de tester nos
traitements. Les résultats des mesures sont dans le tableau suivant :

|  |  |  |
| --- | --- | --- |
| **Traitement 1** | **Traitement 2** | **Traitement 3** |
| 1,45 | 1,87 | 1,1 |
| 1,62 | 1,95 | 1 |
| 1,32 | 2,1 | 0,9 |
| 1,18 | 1,75 | 1,15 |
| 1,47 | 1,85 | 1,1 |
| 1,85 | 1,9 | 1,2 |
| 1,66 | 1,7 | 1,3 |
| 1,5 | 1,6 | 0,95 |
| 1,32 |  | 1,05 |
| 1,25 |  | 1,22 |
|  |  | 1,18 |

Question : y a-t-il une différence entre les moyennes des trois mesures au seuil de 5% ?
1- Les hypothèses à tester sont :
Hypothèse nulle : H0 : µ1 = µ2 = µ3,
Hypothèse alternative : H1 :Au moins une différence entre deux moyennes existe
(µ1, µ2et µ3 sont respectivement les moyennes des mesures pour le traitement1, 2 et 3)

**Procédure Excel :**

Aller dans « Utilitaire d’analyse », cliquez « analyse de variance : un facteur », et comparez (en précisant colonnes, ou lignes) ces résultats en faisant OK.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Analyse de variance: un facteur |  |  |  |  |
|  |  |  |  |  |  |  |
| RAPPORT DÉTAILLÉ |  |  |  |  |  |
| *Groupes* | *Nombre d'échantillons* | *Somme* | *Moyenne* | *Variance* |  |  |
| traitement 1 | 10 | 14,62 | 1,462 | 0,042351 |  |  |
| traitement 2 | 8 | 14,72 | 1,84 | 0,024229 |  |  |
| traitement 3 | 12 | 13,42 | 1,11833333 | 0,015742 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANALYSE DE VARIANCE |  |  |  |  |  |
| *Source des variations* | *Somme des carrés* | *Degré de liberté* | *Moyenne des carrés* | *F* | *Probabilité* | *Valeur critique pour F* |
| Entre Groupes | 2,52002 | 2 | 1,26001 | 46,99408 | **1,60835E-09** | 3,35413083 |
| A l'intérieur des groupes | 0,72392667 | 27 | 0,0268121 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 3,24394667 | 29 |   |   |   |   |

**Interprétation es résultats**

Le premier tableau nous donne l’analyse descriptive du tableau de chaque colonne.

Le deuxième tableau nous donne l’analyse de la variance entre les colonnes et à l’intérieur de chaque colonne et la valeur du test F et la valeur de Signification Probabilité (P-value) qui de l’ordre de **1,60835E-09** qui est largement inférieur de la valeur de référence à 0.05, alors on rejette H0 et on accepte H1, c'est-à-dire qu’il existe au moins une différence entre deux moyennes des trois mesures.

**2- Analyse de la variance (ANOVA deux facteurs facteur)**

L’analyse de variance ANOVA à deux facteurs permet de tester l’hypothèse d’égalité des
moyennes pour plus de deux échantillons tout comme l’Anova à un facteur, la différence c’est
que l’Anova 2 teste cette égalité pour deux facteurs indépendants au lieu d’un.
Il faut préciser que dans l’Anova 2 on distingue deux types de test : Anova 2 sans répétition et
Anova 2 avec répétition (avec interaction).

Exemple :
On a choisis trois parcelles de terrains à superficie égale dans trois régions différentes, et on y
a testé quatre types d’Angré pour un légume donné. La récolte de ce dernier (en tonne) dans
les trois régions pour chaque Angré a été présentée dans le tableau suivant :

|  |  |  |  |
| --- | --- | --- | --- |
|   | **Tlemcen** |  Oran  | Saida |
| Angré 1  | 10 | 11 | 10 |
| Angré 2  | 14 | 12 | 14 |
| Angré 3  | 15 | 16 | 14 |
| Angré 4  | 17 | 19 | 16 |

Question : existe-il une différence de récole dut à la région ou à l’Angré ou les deux à la fois ?

1. Les hypothèses à tester sont :
* Premier facteur : facteur de la région

Hypothèse nulle : H0 : µ1 = µ2 = µ3,

Hypothèse alternative : H1 :Au moins une différence entre deux moyennes existe

(µ1, µ2et µ3 sont respectivement les moyennes des récoltes pour les trois régions)

* Deuxième facteur : facteur de l’Angré

Hypothèse nulle : H0 : µ’1 = µ’2 = µ’3= µ’4,

Hypothèse alternative : H1 : Au moins une différence entre deux moyennes existe

(µ’1, µ’2, µ’3et µ’4 sont respectivement les moyennes des récoltes pour les trois régions)

**Procédure Excel :**

Aller dans « Utilitaire d’analyse », cliquez « analyse de variance : deux facteurs sans répétition », et comparez (en précisant colonnes, ou lignes) ces résultats en faisant OK.

|  |  |
| --- | --- |
| Analyse de variance: deux facteurs sans répétition d'expérience |  |
|  |  |  |  |  |  |  |
| *RAPPORT DÉTAILLÉ* | *Nombre d'échantillons* | *Somme* | *Moyenne* | *Variance* |  |  |
| Angré 1  | 3 | 31 | 10,3333333 | 0,33333 |  |  |
| Angré 2  | 3 | 40 | 13,3333333 | 1,33333 |  |  |
| Angré 3  | 3 | 45 | 15 | 1 |  |  |
| Angré 4  | 3 | 52 | 17,3333333 | 2,33333 |  |  |
|  |  |  |  |  |  |  |
| Tlemcen | 4 | 56 | 14 | 8,66667 |  |  |
|  Oran  | 4 | 58 | 14,5 | 13,6667 |  |  |
| Saida | 4 | 54 | 13,5 | 6,33333 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANALYSE DE VARIANCE |  |  |  |  |  |
| *Source des variations* | *Somme des carrés* | *Degré de liberté* | *Moyenne des carrés* | *F* | *Probabilité* | *Valeur critique pour F* |
| Lignes | 78 | 3 | 26 | 19,5 | **0,00169824** | 4,75706266 |
| Colonnes | 2 | 2 | 1 | 0,75 | **0,512** | 5,14325285 |
| Erreur | 8 | 6 | 1,33333333 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 88 | 11 |   |   |   |   |

**Interprétation es résultats**

Le premier tableau donne l’analyse descriptive des deux facteurs ( angré et région).

Le deuxième tableau montre l’analyse de variance des lignes (angré) et l’analyse de variance des colonnes (regions) et leurs valeurs du test F ainsi que la probabilité de chaque facteur. La signification du facteur lignes (angré) = 0.00169 elle est inferieur à 0.05 donc on rejette H0 et on accepte H1 c'est-à-dire qu’il existe au moins une différence entre deux moyennes.

La signification du facteur colonne (régions) = 0.512 elle est supérieur à 0.05 donc on rejette H1 et on accepte H0 c'est-à-dire qu’il n’existe pas de différence significative entre les régions.