

Exemple 1 sur l'analyse de la variance (ANOVA)

L'analyse de la variance à un facteur Dans cet exemple, on fait une étude sur l'appréciation sensorielle de la texture de trois viandes. Seul le caractère fibreux est considéré, il est noté sur une échelle de 15 points. 15 dégustateurs différents ont noté les viandes représentées selon le tableau suivant :

A	B	C
3	10	13
5	8	11
6	5	7
3	7	11
3	5	8

Pour faire l'ANOVA à un facteur, il faut appliquer la formule suivante :

$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ tel que :

y_{ij} : est la variable à expliquer (la texture fibreuse de la viande),

α_i : est l'effet du ieme niveau du facteur (caractère fibreux de la viande i),

μ : est l'effet moyen général (caractère fibreux potentiel),

ϵ_{ij} : est la variable aléatoire résiduelle (due à l'ensemble des autres causes qui déterminent la note fibreuse),

Les ϵ_{ij} sont indépendants,

ϵ_{ij} suit une loi $N(0, \sigma^2)$.

La décomposition de l'élément y_{ij} est comme suit :

Données globales (Y) = moyenne générale (M) + écart inter-colonnes (B) + écart intra-colonnes (W) c'est à dire :

$$\begin{pmatrix} 3 & 10 & 13 \\ 5 & 8 & 11 \\ 6 & 5 & 7 \\ 3 & 7 & 11 \\ 3 & 5 & 8 \end{pmatrix} = \begin{pmatrix} 7 & 7 & 7 \\ 7 & 7 & 7 \\ 7 & 7 & 7 \\ 7 & 7 & 7 \\ 7 & 7 & 7 \end{pmatrix} + \begin{pmatrix} -3 & 0 & 3 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \end{pmatrix} + \begin{pmatrix} -1 & 3 & 3 \\ 1 & 1 & 1 \\ 2 & -2 & -3 \\ -1 & 0 & 1 \\ -1 & -2 & -2 \end{pmatrix}$$

La matrice Y des données globales est :

$$Y = \begin{pmatrix} 3 & 10 & 13 \\ 5 & 8 & 11 \\ 6 & 5 & 7 \\ 3 & 7 & 11 \\ 3 & 5 & 8 \end{pmatrix}$$

On calcule la moyenne générale des trois séries notée $\bar{\bar{Y}}$ comme suit :

$$\bar{\bar{Y}} = \frac{3 + 5 + 6 + 3 + 3 + 10 + 8 + 5 + 7 + 5 + 13 + 11 + 7 + 11 + 8}{15} = 7.$$

A partir de cette moyenne générale, on génère la matrice M tel que :

$$M = \begin{pmatrix} 7 & 7 & 7 \\ 7 & 7 & 7 \\ 7 & 7 & 7 \\ 7 & 7 & 7 \\ 7 & 7 & 7 \end{pmatrix}$$

On peut aussi calculer $\bar{\bar{Y}}$ à partir des moyennes de chaque série comme suit :

Soient $\bar{Y}_1$, $\bar{Y}_2$ et $\bar{Y}_3$ respectivement les moyennes des séries A , B et C tel que :

$$\begin{aligned} \bar{Y}_1 &= \frac{3+5+6+3+3}{5} = 4. \\ \bar{Y}_2 &= \frac{10+8+5+7+5}{5} = 7. \\ \bar{Y}_3 &= \frac{13+11+7+11+8}{5} = 10. \end{aligned}$$

Donc :

$$\bar{\bar{Y}} = \frac{\bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3}{3} = \frac{4+7+10}{3} = 7.$$

La matrice B est qui est l'écart inter-colonnes est calculée comme suit :

La première colonne c'est la différence entre $\bar{Y}_1$ et $\bar{\bar{Y}}$ c'est à dire $\bar{Y}_1 - \bar{\bar{Y}} = 4 - 7 = -3$.

La deuxième colonne c'est la différence entre $\bar{Y}_2$ et $\bar{\bar{Y}}$ c'est à dire $\bar{Y}_2 - \bar{\bar{Y}} = 7 - 7 = 0$.

La troisième colonne c'est la différence entre $\bar{Y}_3$ et $\bar{\bar{Y}}$ c'est à dire $\bar{Y}_3 - \bar{\bar{Y}} = 10 - 7 = 3$.

$$\text{La matrice B} = \begin{pmatrix} -3 & 0 & 3 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \\ -3 & 0 & 3 \end{pmatrix}$$

La matrice W est calculée en faisant la différence entre les éléments de la colonne une la moyenne adéquate, la même chose pour les autres colonnes. En faisant ces calculs on trouve la matrice W qui est comme suit :

$$W = \begin{pmatrix} -1 & 3 & 3 \\ 1 & 1 & 1 \\ 2 & -2 & -3 \\ -1 & 0 & 1 \\ -1 & -2 & -2 \end{pmatrix}.$$

La somme des carrés totale notée **SCT** est calculée comme suit :

$$\text{SCT} = (3-7)^2 + (5-7)^2 + (6-7)^2 + (3-7)^2 + (3-7)^2 + (10-7)^2 + (8-7)^2 + (5-7)^2 + (7-7)^2 + (5-7)^2 + (13-7)^2 + (11-7)^2 + (7-7)^2 + (11-7)^2 + (8-7)^2 = 140.$$

Si on reprend le tableau vu précédemment, on remarque qu'on a 3 groupes de données on peut dire qu'on a m groupes de données et chaque groupe de données contient 5 élément, donc on peut dire que chaque groupe ou classe contient n données. En tout on m × n données. Si on connaît la moyenne générale des données qui est $\bar{\bar{Y}}$ alors les m × n données sur le tableau ne sont pas toutes utiles ou indépendantes. C'est à dire si on connaît les m × n-1 données, on peut utiliser la moyenne pour trouver la donnée manquante. On dit qu'on a m × n-1 degrés de liberté notée **DDL**. Dans notre cas le **DDL** = 3×5-1 = 14.

Pour calculer la variance des données représentées dans le tableau précédent on utilise la

formule suivante :

$$\text{La variance} = \frac{SCT}{DDL} = \frac{140}{14}.$$

La question qu'on peut poser maintenant, est ce que cette variance est due à la variation au sein du groupe (à l'intérieur du groupe) ou entre les groupes ?

Pour cela on va décomposer la somme des carrés totale **SCT** en somme des carrés qui proviennent des variations à l'intérieur des classes et en somme des carrés qui proviennent des variations entre les classes.

On commence par calculer la somme des carrés intra-classes c'est à dire on calcule les écarts par rapport à la moyenne dans chaque classe cette somme est notée par SC_{intra} .

$$SC_{intra} = (3-4)^2 + (5-4)^2 + (6-4)^2 + (3-4)^2 + (3-4)^2 + (10-7)^2 + (8-7)^2 + (5-7)^2 + (7-7)^2 + (5-7)^2 + (13-10)^2 + (11-10)^2 + (7-10)^2 + (11-10)^2 + (8-10)^2 \\ = 2 + 2 + 2 = 50.$$

$$\Rightarrow SC_{intra} = 50.$$

Maintenant si on prend par exemple la première classe A, on remarque qu'on a en réalité 4 données indépendantes car si on connaît la moyenne de la classe qui est $\bar{Y}_1$ et les 4 autres éléments de la colonne, alors on pourra calculer l'autre élément. On général, si on a une classe qui contient n données et si on connaît la moyenne, alors il y a n-1 données indépendantes. Cette valeur est notée par $DDL = m \times (n - 1)$ sachant que m représente le nombre de classes dans lesquelles on a n-1 données indépendantes.

Conclusion : le degré de liberté à l'intérieur de la $SC_{intra} = m \times (n - 1) = 12$.

maintenant, on va calculer la part de la variation totale qui est due à la variation entre les classes. Cette somme est la somme des carrés inter-classes notée par : SC_{inter} .

$$SC_{inter} = (4-7)^2 + (4-7)^2 + (4-7)^2 + (4-7)^2 + (4-7)^2 + (7-7)^2 + (7-7)^2 + (7-7)^2 + (7-7)^2 + (7-7)^2 + (10-7)^2 + (10-7)^2 + (10-7)^2 + (10-7)^2 + (10-7)^2 = 90.$$

On peut se poser la même question c'est à dire : L'intervention des variables indépendante elle est de combien si on calcule la SC_{inter} (dans ce cas on cherche le degré de liberté) ? Dans ce cas, on se pose la question : si on connaît $\bar{Y}$ on a besoin de combien de moyenne de classe indépendantes ? La réponse est m-1 c'est à dire si je connais deux moyennes et la moyenne générale, alors je peux calculer la moyenne restante. Donc le DDL pour la $SC_{inter} = m-1 = 2$.

Remarque : $SCT = SC_{intra} + SC_{inter}$, la même chose avec les degrés de liberté c'est à dire $DDL_{SCT} = DDL_{SC_{intra}} + DDL_{SC_{inter}}$.

On peut aussi écrire :

$$\sum_i \sum_j (y_{ij} - \bar{Y})^2 = \sum_i n_i (\bar{Y}_i - \bar{Y})^2 + \sum_i \sum_j (y_{ij} - \bar{Y}_i)^2$$

Les écart totaux = écart inter-niveaux + écart intra-niveau.

Interprétation des statistiques obtenus On souhaite tester les hypothèses :

H_0 : " il n'a y a pas d'effet produit " c'est à dire les moyennes pour les différents produits sont égales. Contre l'hypothèse H_1 : " il y a un effet produit " c'est à dire au moins de moyennes sont différentes.

Il s'agit donc de comparer la variabilité inter-niveaux à la variabilité intra-niveau du fac-

teur.

Selon les résultats vus précédemment, on peut créer le tableau suivant :

Source de variation	SC	DDL
Inter-niveaux	SC_{inter}	$m-1 = 3-1 = 2$
Intra-niveaux	SC_{intra}	$m \times (n-1) = 12$
Totale	SCT	$(m \times n) - 1 = 14$

On définit le carré moyen inter-groupes :

$$CM_{inter} = \frac{SC_{inter}}{m-1} = \frac{90}{2} = 45$$

Et le carré moyen intra-groupes qui est le $CM_{intra} = \frac{SC_{intra}}{m \times (n-1)} = \frac{50}{12} = 4,16$.

Pour tester H_0 contre H_1 , on suppose que H_0 est vraie, puis on va calculer une statistique F qui va suivre une loi de Fisher, cette statistique est une caution de deux variables qui suivent une loi de Khi2.

$$F = \frac{\frac{SC_{inter}}{m-1}}{\frac{SC_{intra}}{m \times (n-1)}} = \frac{CM_{inter}}{CM_{intra}}, \text{ alors si } \frac{SC_{inter}}{m-1} \gg \frac{SC_{intra}}{m \times (n-1)}, \text{ on peut dire que}$$

la variation due à des variations entre les classes est beaucoup plus importante que la variation due à des variations l'intérieur des classes. Si le F est grand, alors on a une très faible probabilité que l'hypothèse nulle H_0 soit vraie. Maintenant si F est petit c'est à dire $\frac{SC_{inter}}{m-1} \ll \frac{SC_{intra}}{m \times (n-1)}$ et dans ce cas on a une très grande probabilité que l'hypothèse nulle H_0 soit vraie.

Si on reprend l'exemple, $F = \frac{45}{4,16} = 10,81$.

A partir de quelle valeur observée de F rejette-t-on H_0 ? Si les résidus du modèle de l'analyse de la variance suivent une loi normale et si H_0 est vraie, on sait que F est l'observation d'une variable qui suit la loi Fisher ayant $m-1$ DDL au numérateur et $m \times (n-1)$ DDL au dénominateur noté $F_{(m-1, m \times (n-1))}$.

H_0 est rejetée si $F_{obs} > F_{(m-1, m \times (n-1))}$ pour un niveau de signification donné par α c'est à dire si p-value $< \alpha$ sur la table de Fisher.

Si on prend l'exemple de la viande, le tableau d'analyse de la variance est comme suit :

Source de variation	SC	DDL	CM	F	p-value
Type de viande	90	2	45		
Résiduelle	50	12	4,16	10,81	$p < 0,01$
Totale	140	14			

La p-value = $P(F_{\alpha, v_1, v_2} > F_{observe})$

Sur la table de Fisher :

Pour $\alpha = 5 \%$, $F_{\alpha, 2, 12} = 3,88$.

Pour $\alpha = 1 \%$, $F_{\alpha, 2, 12} = 6,93$.

$F_{obs} = 10,81 > F_{\alpha, 2, 12}$ pour les deux α , donc on dit que H_0 est rejetée et dans ce cas il existe un effet type de viande significatif concernant le caractère fibreux.

La table de Fisher pour $\alpha = 5 \%$

Table de la loi de Fisher-Snedecor, $\alpha = 5\%$										
num	den 1	2	3	4	5	6	7	8	9	10
1	161.4476	18.5128	10.1280	7.7086	6.6079	5.9874	5.5914	5.3177	5.1174	4.9646
2	199.5000	19.0000	9.5521	6.9443	5.7861	5.1433	4.7374	4.4590	4.2565	4.1028
3	215.7073	19.1643	9.2766	6.5914	5.4095	4.7571	4.3468	4.0662	3.8625	3.7083
4	224.5832	19.2468	9.1172	6.3882	5.1922	4.5337	4.1203	3.8379	3.6331	3.4780
5	230.1619	19.2964	9.0135	6.2561	5.0503	4.3874	3.9715	3.6875	3.4817	3.3258
6	233.9860	19.3295	8.9406	6.1631	4.9503	4.2839	3.8660	3.5806	3.3738	3.2172
7	236.7684	19.3532	8.8867	6.0942	4.8759	4.2067	3.7870	3.5005	3.2927	3.1355
8	238.8827	19.3710	8.8452	6.0410	4.8183	4.1468	3.7257	3.4381	3.2296	3.0717
9	240.5433	19.3848	8.8123	5.9988	4.7725	4.0990	3.6767	3.3881	3.1789	3.0204
10	241.8817	19.3959	8.7855	5.9644	4.7351	4.0600	3.6365	3.3472	3.1373	2.9782
11	242.9835	19.4050	8.7633	5.9358	4.7040	4.0274	3.6030	3.3130	3.1025	2.9430
12	243.9060	19.4125	8.7446	5.9117	4.6777	3.9999	3.5747	3.2839	3.0729	2.9130
13	244.6898	19.4189	8.7287	5.8911	4.6552	3.9764	3.5503	3.2590	3.0475	2.8872
14	245.3640	19.4244	8.7149	5.8733	4.6358	3.9559	3.5292	3.2374	3.0255	2.8647
15	245.9499	19.4291	8.7029	5.8578	4.6188	3.9381	3.5107	3.2184	3.0061	2.8450
16	246.4639	19.4333	8.6923	5.8441	4.6038	3.9223	3.4944	3.2016	2.9890	2.8276
17	246.9184	19.4370	8.6829	5.8320	4.5904	3.9083	3.4799	3.1867	2.9737	2.8120
18	247.3232	19.4402	8.6745	5.8211	4.5785	3.8957	3.4669	3.1733	2.9600	2.7980
19	247.6861	19.4431	8.6670	5.8114	4.5678	3.8844	3.4551	3.1613	2.9477	2.7854
20	248.0131	19.4458	8.6602	5.8025	4.5581	3.8742	3.4445	3.1503	2.9365	2.7740
21	248.3094	19.4481	8.6540	5.7945	4.5493	3.8649	3.4349	3.1404	2.9263	2.7636
22	248.5791	19.4503	8.6484	5.7872	4.5413	3.8564	3.4260	3.1313	2.9169	2.7541
23	248.8256	19.4523	8.6432	5.7805	4.5339	3.8486	3.4179	3.1229	2.9084	2.7453
24	249.0518	19.4541	8.6385	5.7744	4.5272	3.8415	3.4105	3.1152	2.9005	2.7372
25	249.2601	19.4558	8.6341	5.7687	4.5209	3.8348	3.4036	3.1081	2.8932	2.7298
26	249.4525	19.4573	8.6301	5.7635	4.5151	3.8287	3.3972	3.1015	2.8864	2.7229
27	249.6309	19.4587	8.6263	5.7586	4.5097	3.8230	3.3913	3.0954	2.8801	2.7164
28	249.7966	19.4600	8.6229	5.7541	4.5047	3.8177	3.3858	3.0897	2.8743	2.7104
29	249.9510	19.4613	8.6196	5.7498	4.5001	3.8128	3.3806	3.0844	2.8688	2.7048
30	250.0951	19.4624	8.6166	5.7459	4.4957	3.8082	3.3758	3.0794	2.8637	2.6996

FIGURE 1 – table de Fisher pour $\alpha = 5\%$

num	den 11	12	13	14	15	16	17	18	19	20
1	4.8443	4.7472	4.6672	4.6001	4.5431	4.4940	4.4513	4.4139	4.3807	4.3512
2	3.9823	3.8853	3.8056	3.7389	3.6823	3.6337	3.5915	3.5546	3.5219	3.4928
3	3.5874	3.4903	3.4105	3.3439	3.2874	3.2389	3.1968	3.1599	3.1274	3.0984
4	3.3567	3.2592	3.1791	3.1122	3.0556	3.0069	2.9647	2.9277	2.8951	2.8661
5	3.2039	3.1059	3.0254	2.9582	2.9013	2.8524	2.8100	2.7729	2.7401	2.7109
6	3.0946	2.9961	2.9153	2.8477	2.7905	2.7413	2.6987	2.6613	2.6283	2.5990
7	3.0123	2.9134	2.8321	2.7642	2.7066	2.6572	2.6143	2.5767	2.5435	2.5140
8	2.9480	2.8486	2.7669	2.6987	2.6408	2.5911	2.5480	2.5102	2.4768	2.4471
9	2.8962	2.7964	2.7144	2.6458	2.5876	2.5377	2.4943	2.4563	2.4227	2.3928
10	2.8536	2.7534	2.6710	2.6022	2.5437	2.4935	2.4499	2.4117	2.3779	2.3479
11	2.8179	2.7173	2.6347	2.5655	2.5068	2.4564	2.4126	2.3742	2.3402	2.3100
12	2.7876	2.6866	2.6037	2.5342	2.4753	2.4247	2.3807	2.3421	2.3080	2.2776
13	2.7614	2.6602	2.5769	2.5073	2.4481	2.3973	2.3531	2.3143	2.2800	2.2495
14	2.7386	2.6371	2.5536	2.4837	2.4244	2.3733	2.3290	2.2900	2.2556	2.2250
15	2.7186	2.6169	2.5331	2.4630	2.4034	2.3522	2.3077	2.2686	2.2341	2.2033
16	2.7009	2.5989	2.5149	2.4446	2.3849	2.3335	2.2888	2.2496	2.2149	2.1840
17	2.6851	2.5828	2.4987	2.4282	2.3683	2.3167	2.2719	2.2325	2.1977	2.1667
18	2.6709	2.5684	2.4841	2.4134	2.3533	2.3016	2.2567	2.2172	2.1823	2.1511
19	2.6581	2.5554	2.4709	2.4000	2.3398	2.2880	2.2429	2.2033	2.1683	2.1370
20	2.6464	2.5436	2.4589	2.3879	2.3275	2.2756	2.2304	2.1906	2.1555	2.1242
21	2.6358	2.5328	2.4479	2.3768	2.3163	2.2642	2.2189	2.1791	2.1438	2.1124
22	2.6261	2.5229	2.4379	2.3667	2.3060	2.2538	2.2084	2.1685	2.1331	2.1016
23	2.6172	2.5139	2.4287	2.3573	2.2966	2.2443	2.1987	2.1587	2.1233	2.0917
24	2.6090	2.5055	2.4202	2.3487	2.2878	2.2354	2.1898	2.1497	2.1141	2.0825
25	2.6014	2.4977	2.4123	2.3407	2.2797	2.2272	2.1815	2.1413	2.1057	2.0739
26	2.5943	2.4905	2.4050	2.3333	2.2722	2.2196	2.1738	2.1335	2.0978	2.0660
27	2.5877	2.4838	2.3982	2.3264	2.2652	2.2125	2.1666	2.1262	2.0905	2.0586
28	2.5816	2.4776	2.3918	2.3199	2.2587	2.2059	2.1599	2.1195	2.0836	2.0517
29	2.5759	2.4718	2.3859	2.3139	2.2525	2.1997	2.1536	2.1131	2.0772	2.0452
30	2.5705	2.4663	2.3803	2.3082	2.2468	2.1938	2.1477	2.1071	2.0712	2.0391

FIGURE 2 – Suite table de Fisher pour $\alpha = 5\%$