INTERPRETATION STATISTIQUE DES DONNEES RESULTANT DU CLASSEMENT PAR ORDRE DE PREFERENCE OU DE MERITE DE n VINS PAR m DEGUSTATEURS

M. RIVES
Station de Recherches de Viticulture
Institut National de la Recherche Agronomique
Centre de Recherches de Bordeaux
33140 Pont-de-la-Maye (France)

La dégustation est un art qui exige des dons psychophysiologiques de discernement, une grande mémoire et de l'entraînement.

Des spécialistes en ont traité (PEYNAUD, 1971) et il n'est pas dans nos intentions de parler ici d'une discipline qui n'est pas la nôtre.

La dégustation fournit des appréciations fréquemment très exactes, quelques fois même précises, mais qui restent le plus souvent descriptives. Or les besoins de l'expérimentation exigent souvent que l'on puisse analyser biométriquement les données fournies par les dégustateurs. C'est le cas quand il faut établir un classement relatif entre des vins, par exemple pour faire une sélection sur la qualité entre différents clones, pour juger de différentes fumures, etc...

L'analyse est nécessaire car, si l'on sait que le jugement humain est inconstant et imprécis, comme toute donnée biologique, il faut pouvoir apprécier, par un test statistique, si le classement a un sens, c'est-à-dire si, compte tenu de l'imprécision des juges, on peut penser que les différences entre les vins que leurs résultats suggèrent sont réelles ou non.

On peut alors chercher à obtenir des dégustateurs des notes, par exemple de 0 à 20, soit directement, soit par le biais d'une échelle de préférence (« échelle hédonique » de PERYAM et PILGRIM 1957, cité par RIVES 1959) et l'on peut procéder par une notation globale ou par une méthode analytique, comme celle de OUGH et BAKER (1961).

Beaucoup de dégustateurs répugnent à cette contrainte. La docimologie, ou science des examens, constate que les notations chiffrées sont des instruments de mesure très imparfaits, non seulement parce que chaque juge n'a pas toujours la même doctrine, note haut ou bas, étalé ou ramassé, mais parce que surtout la justification d'un système de notes chiffrées demande des hypothèses dont on ne sait si elles sont réellement plausibles.

Il faut d'abord supposer que les objets que l'on juge se situent, par le ou les caractères jugés, dans un **continuum,** c'est-à-dire que le passage d'une note à la suivante peut se faire de manière continue (au sens de l'analyse mathématique).

Prenons un exemple : soit une boisson agréable si elle est sucrée déplaisante si elle est amère, et des échantillons allant du sucré pur à l'amer pur, en devenant de moins en moins sucrés et de plus en plus amers. Est-il possible de penser qu'il y a continuité du sucré à l'amer, ce que suppose une échelle de notes décroissantes ? Dans certains cas précis, on est sûr qu'il n'y a pas continuité : ce sont ceux où il existe un seuil de sensibilité net et tranché. L'existence du **continuum** n'est pas cependant la difficulté majeure. Même en l'admettant, en effet, il reste nécessaire d'admettre, de surcroît, que l'échelle est linéaire : c'est-à-dire que 11 représente 110 p. 100 de la satisfaction que représente 10, que 14 représente une satisfaction double de celle que décrit 7, et ainsi de suite, ce qui est plus que douteux.

La meilleure méthode est bien entendu celle des comparaisons par paires, ou celle des comparaisons triangulaires, mais on recule souvent devant le nombre de dégustations qu'elles nécessitent.

Aussi essaie-t-on souvent d'obtenir des juges non pas une notation chiffrée, mais un classement relatif des vins les uns par rapport aux autres. Ce classement peut comprendre des ex-aequos ; il a l'avantage de ne faire aucune hypothèse sur la linéarité des relations d'une place à une autre.

En examinant le classement des mêmes objets, par exemple des vins par plusieurs juges, on s'aperçoit le plus souvent qu'ils ne sont pas identiques (S'ils l'étaient, il y aurait **concordance absolue**). On se pose alors plusieurs questions : Les vins méritent-ils d'être distingués les uns des autres ; sont-ils différents ? S'ils sont différents, les juges sont-ils capables de les classer par ordre de mérite ou de préférence ?

Nous sentons bien que si les vins ne sont pas différents, par exemple si on a servi **n** fois le même vin pour tendre un piège aux juges, les classements, s'ils en font, seront l'œuvre du hasard. Mais les vins peuvent être différents et les classements des juges ne pas concorder : dans ce cas, on peut faire deux hypothèses : ou bien les juges sont mauvais et de peu de discernement et classent, en fait, à peu près au hasard, donc le « classement moyen » qu'on peut imaginer de faire n'a aucune valeur, ou bien les juges sont compétents, sensibles et ont un jugement reproductible, mais ils n'ont pas la même conception de la valeur relative des vins : dans ce cas, sans porter de jugement sur les juges, on ne peut que constater qu'on a un mauvais outil de travail car il ne permet pas d'atteindre le but poursuivi. Notons qu'un seul essai ne permet pas de distinguer entre les deux dernières hypothèses : seule l'appréciation de la concordance d'un juge avec lui-même, dans différents classements de la même série, à plusieurs reprises, permettrait de trancher.

Nous supposerons donc que nous aurons probablement choisi un jury de dégustateurs possédant le discernement nécessaire et ayant les mêmes préférences, ou capables d'établir un classement relatif de valeur par référence à une estimation connue, par exemple les préférences d'un marché.

Alors la concordance entre leurs classements pour une même série de vins établira que ces vins sont différents les uns des autres.

On se pose alors une troisième et dernière question : quel est le classement relatif réel des vins ?

On démontre (KENDALL, 1962) que, dans ce cas, la meilleure estimation possible du classement relatif réel des vins est celle obtenue en utilisant les totaux de leurs rangs dans les classements des différents juges.

Notons enfin qu'à mesure que le nombre des juges augmente, la concordance de leur classement est un argument de plus en plus fort, à lui seul, pour penser que leur système de préférences représente bien un système général commun au groupe social d'où ils sont originaires.

Notre propos est ici de montrer comment on peut mesurer la concordance entre les classements de **n** vins par **m** juges, au moyen d'une méthode dûe à KENDALL (1962) qui peut s'appliquer au jugement subjectif, par classement de n'importe quel type d'objets, et qui a été proposée par AMERINE et **al.** (1959) pour le vin. Cette méthode est fondée sur le calcul du **coefficient de concordance W.**

Les données sont disposées sous forme d'un tableau, dont les lignes correspondent aux dégustateurs-juges, et les colonnes aux vins (tableau II).

Les données, soient y_{ij}, sont les rangs ou numéros de classement attribuables aux vins d'après les classements des juges.

En cas d'ex-aequos, on applique la règle suivante : le rang y_{ij} affecté à chaque vin d'un groupe d'ex-aequos est la moyenne des rangs occupés par le groupe. Par exemple, si après deux vins classés isolément premier et second, on trouve un groupe de trois vins qui seraient classés « troisièmes ex-aequos » suivant l'usage traditionnel, leur note commune sera :

$$(3 + 4 + 5)/3 = 4 \text{ (et non 3)}$$

De même s'il y en a quatre :

$$(3 + 4 + 5 + 6)/4 = 4{,}5$$

Quand il y a des fractions, comme 4,5, il est commode de multiplier l'ensemble du tableau par 2, pour n'avoir que des entiers.

Le total de chaque ligne $Y_{i.}$ est égal à $\frac{1}{2} n (n + 1)$ c'est la somme des n premiers entiers, (dans le cas où on a éliminé des notes fractionnaires en doublant tout) $Y_{i.} = n (n + 1)$.

On calcule les totaux $Y_{.J}$ par vin, le total général Y.. puis le terme correctif :

$$TC = Y^2../nm$$

la somme des carrés de toutes les données :

$$\Sigma y^2_{ij}$$

la somme des carrés des totaux de colonnes divisés par le nombre de juges

$$\Sigma (Y^2_{.j}/m)$$

enfin les sommes de carrés d'écarts :

$$SCy = \Sigma y^2_{ij} - TC$$

$$SCv = \Sigma (Y^2_{.j}/m) - TC$$

On a alors :

$$W = \frac{SCv - 1/m}{SCy - 2/m}$$

En fait, dès que m dépasse 5, on peut négliger les « corrections de continuité » 1/m et 2/m.

On calcule alors :

$$F = (m - 1)\ W/(1 - W) = \frac{(m - 1)\ SCv}{SCy - SCv}$$

avec pour degrés de liberté :

$$dl\ numérateur = (n - 1) - 2/m$$

$$dl\ dénominateur = (m - 1)\ (n - 1 - 2/m)$$

Il y a lieu ici de tenir compte de -2/m, surtout pour le dénominateur. On arrive donc à des nombres de degrés de liberté fractionnaires (ceux-ci ne figurant pas dans les tables, on prendra les entiers les plus proches).

F est d'autant plus grand que la **concordance** entre les juges est meilleure, donc que le classement des vins est plus évident.

Nous prendrons deux exemples : dans le premier onze vins ont été classés par six juges.

TABLEAU I

Les numéros en marge représentent les juges (lignes) et l'ordre de classement (colonnes) ; les nombres dans le corps du tableau représentent les numéros des vins. Les groupes d'ex-aequos sont soulignés.

Ordre	1	2	3	4	5	6	7	8	9	10	11
Juges											
1	4	8	5	6	7	1	2	3	9	11	10
2	4	5	6	8	9	7	2	3	1	11	10
3	8	9	6	2	3	7	1	4	5	11	10
4	8	9	3	1	4	5	7	2	6	11	10
5	8	9	1	4	6	7	3	5	2	11	10
6	8	4	9	7	1	6	2	5	11	3	10

TABLEAU II

Les notes obtenues en résolvant les ex-aequos et en doublant sont réarrangées : les colonnes correspondent maintenant aux vins, les lignes correspondent toujours aux dégustateurs et les nombres du tableau II sont les rangs y_{ij} multipliés par 2.

	1	2	3	4	5	6	7	8	9	10	11	$Y_{i.}$
1	14	14	14	2	8	8	8	4	20	20	20	132
2	20	14	14	6	6	6	11	6	6	20	20	132
3	16	10	10	16	16	4	10	4	4	21	21	132
4	11	18	4	11	11	18	11	4	4	22	18	132
5	7	17	17	7	7	7	7	7	7	22	17	132
6	14	14	21	5	14	14	5	5	5	21	14	132
$Y_{.j}$	82	87	80	47	72	57	55	30	46	126	110	792
R	8	9	7	3	6	5	4	1	2	11		

On a :

$$TC = \overline{792}^2/66 = 8504$$

$$\Sigma y^2 = \overline{14}^2 + \overline{14}^2 + \ldots\ldots + \overline{21}^2 + \overline{14}^2 = 11850$$

$$\Sigma\,(Y^2/m) = (\overline{82}^2 + \overline{87}^2 + \ldots + \overline{126}^2 + \overline{110}^2)/6 = 10892$$

$$SCy = 11850 - 9504 = 2346$$

$$SCv = 10892 - 9504 = 1888$$

$$W = \frac{1388 - 1/6}{2346 - 2/6} \simeq \frac{1388}{2346} = 0{,}5916$$

$$F = \frac{5 \times 0{,}5916}{1 - 0{,}5916} = \frac{2{,}9580}{0{,}4084} = 7{,}244$$

$$\text{ou} \quad F = \frac{5 \times 1388}{2346 - 1388} = \frac{6940}{958} = 7{,}244$$

avec

$$dl\ n = 10 - 2/6 = 9{,}668 \sim 10$$

$$dl\ d = 5 \times 9{,}667 = 48{,}333 \sim 48$$

avec 10 et 48 degrés de liberté, cette valeur de F a une probabilité d'environ $8 \cdot 10^{-7}$; elle est donc très significative : il y a une bonne concordance entre les juges.

Les vins sont donc très probablement différents, et le classement R, déduit des totaux Y_j est la meilleure estimation qu'on puisse faire de la valeur relative de ces vins.

Notre exemple fournit l'occasion de montrer qu'en dehors de l'analyse purement biométrique des données, il est bon de regarder celles-ci avec les yeux du bon sens.

En effet, il est apparu par ailleurs que l'un des vins, le numéro 10, avait un défaut : c'est ce qui explique qu'il ait été mis en dernière position par tous les juges. Quant au vin 11, il était unique en son genre et servait de témoin : c'est un cépage pur, alors que tous les autres vins sont des mélanges. Le vin 11 a été classé avant-dernier par 5 juges sur 8, antépénultième par le sixième juge.

Si on regarde alors le lot des neuf autres vins (tableau I), on constate que les classements sont d'une concordance assez peu évidente.

Ne pourrait-il pas se faire que la signification très importante trouvée plus haut résulte de l'opposition d'une masse indifférenciée de neuf vins à peu près semblables et de deux vins très différenciés et faciles à distinguer du lot précédent ?

Pour le savoir, nous refaisons le calcul en omettant les deux dernières colonnes du tableau II.

On obtient alors :

$SCy = 1425{,}26$
$SCv = 504{,}56$
$W = 0{,}35401$
$F = 2{,}740$ avec 8 et 38 degrés de liberté

ce qui correspond à une probabilité de 0,0172. La concordance est donc encore significative au sens statistique. Mais elle l'est beaucoup moins.

Cela rappelle opportunément que l'outil statistique n'est qu'un outil, qui ne donne de bons résultats que s'il est bien employé, c'est-à-dire si l'on soumet en même temps les données et leur analyse à la critique de la logique.

TABLEAU III

Classement → / Dégustateurs ↓	1	2	3	4	5	6	7
1	1	2	4	7	3	5	6
2	5	6	7	1	2	3	4
3	7	4	3	1	5	6	2
4	1	7	4	5	3	6	2
5	7	5	2	3	4	1	6
6	1	2	7	6	3	4	5

TABLEAU IV

Vins → / Dégustateurs ↓	1	2	3	4	5	6	7	
1	2	4	12	6	12	12	8	56
2	11	11	11	11	2	4	6	56
3	7	14	7	4	11	11	2	56
4	2	14	10	7	7	12	4	56
5	12	6	8	10	3	14	3	56
6	2	4	11	11	11	11	6	56
$Y_{.j}$	36	53	59	49	46	64	29	336
R = classement	2	5	6	4	3	7	1	

Notre deuxième exemple n'illustre pas mieux le détail des calculs, mais il a l'avantage de présenter un cas où le même jury que dans le premier cas a été mis en présence de vins qui ne différaient que très peu les uns des autres.

Les données et les résultats figurent aux tableaux III, IV et V.

TABLEAU V

Somme des carrés totale	616
Somme des carrés " vins "	152
W	0,2466
F	1,6367
avec 5 et 28 degrés de liberté, ce qui correspond à une probabilité de 0,183 : le F n'est pas significatif.	

Le résultat est confirmé par celui de la dégustation des mêmes vins par un autre jury composé de dégustateurs également compétents.

Cet exemple illustre le fait que la signification du F, c'est-à-dire l'obtention d'une valeur calculée grande dont la probabilité est faible, doit être considérée comme un signe de concordance des juges, et comme un signe que les vins sont différents.

Dans la mesure où les vins sont différents, si les juges concordent pour les classer, ce sont de bons juges (au moins du point de vue expérimental qui nous préoccupe et qui est d'obtenir un classement).

Dans la mesure où le F n'est pas significatif, on peut penser que les juges sont incompétents ou inconstants ce qui revient au même du point de vue expérimental. Mais si nous savons par une expérience antérieure que nos juges sont bons, comme c'est le cas pour notre deuxième exemple, nous pouvons penser devant un F non significatif que leur incapacité à s'accorder sur un classement vient de ce qu'il n'y a pas de classement possible parce que les vins sont trop semblables.

C'est ce que confirme la concordance des résultats des deux jurys.

Manuscrit reçu le 18 Mai 1972.

REFERENCES BIBLIOGRAPHIQUES

AMERINE M.A., ROESSLER E.B. et FILIPELLO F., 1959. Modern sensory methods of evaluating wines. **Hilgardia., 28,** 477-561.

KENDALL M.G., 1962. **Rank correlation methods.** Ch. Griffin., Londres, 3e éd. 199 p.

OUGH C.S. et BAKER G.A., 1961. Small panel sensory evaluations of wines by scoring. **Hilgardia., 30,** 587-619

PERYAM D.R. et PILGRIM F.J., 1957. Hedonic scale method of measuring food preferences. **I.F.T. Symp.** 9.14.

PEYNAUD E. 1971. **Connaissance et travail du vin.** Dunod, Paris.

RIVES M., 1959. Note sur la méthode de l'échelle hédonique pour mesurer les préférences gustatives. **Ann. Technol. Agric.,** 8, 101-105.