

Tout ce que vous avez toujours voulu savoir sur le TBL sans jamais oser le demander

Définitions

Avant de rentrer dans le vif du sujet, quelques définitions et explications.

TBL vient des initiales de ses concepteurs : MM. Tarasov, Bauer et Long, trois éminents statisticiens qui se sont penchés sur les méthodes statistiques permettant d'éliminer les erreurs de jugement intentionnelles ou non.

Le TBL est utilisé dans les compétitions de F3A et de F3P. Un algorithme similaire (TBLP) est utilisé en voltige grandeur.

Le but recherché est de supprimer autant que possible l'influence des « styles » des juges (un juge note plus sévèrement qu'un autre) et la « distorsion » de la notation (un juge avantage ou désavantage un ou plusieurs concurrents).

Dans la suite, nous allons parler de scores. Le score est la somme des produits des points par les coefficients des figures. Pour un vol donné et pour un pilote donné, un juge aura donc attribué un score.

Pour la bonne compréhension de ce qui suit, il est nécessaire de réviser quelques notions de statistique, mais rien de très compliqué !

Je ne reviendrai pas sur le calcul d'une moyenne, mais il faut parler de la notion d'écart-type : l'écart-type (Standard Deviation en anglais) mesure la dispersion (la variance) d'une série de nombres. Dans notre cas, ces nombres seront les scores des juges pour un tour de vol donné.

Plus le juge aura donné des scores différents, plus l'écart-type sera grand. A la limite, si le juge donne le même score à tous les pilotes, l'écart-type sera de 0. (Pour ceux que cela intéresse, la formule de calcul figure en fin de document)

Principe du calcul du TBL

Il est important de comprendre que le TBL ne s'applique qu'aux scores (et non aux notes de figures) pour un seul tour de vol et non pour l'ensemble de la compétition. On ne peut donc pas parler de « notes après TBL », mais seulement de « scores après TBL ».

Par ailleurs, le rapport à 1000 est calculé après TBL.

Etape 1

Le calcul commence par une étape de normalisation. En effet, les juges ont des styles de jugement différents. Tel juge notera « plus durement », c'est-à-dire qu'il donnera moins de points, c'est-à-dire que sa moyenne de score sera plus basse que celle de l'ensemble du collège des juges.

Tel juge donnera des scores compris entre 480 et 520, tel juge donnera des scores entre 300 et 550.

On ne peut donc pas comparer directement ces scores et il est donc nécessaire de les « normaliser ». C'est le but de l'étape 1.

A la fin de cette étape, tous les juges auront la même moyenne de score et le même écart-type.

Pour éclaircir les choses, nous allons travailler sur un exemple :

	Juge n° 1	Juge n° 2	Juge n° 3	Juge n° 4	Juge n° 5	Score	Rang
Pilote n° 1	700	610	300	725	620	591	6
Pilote n° 2	600	545	650	625	590	602	5
Pilote n° 3	760	695	750	945	740	778	2
Pilote n° 4	800	710	785	755	825	775	3
Pilote n° 5	900	815	895	965	920	899	1
Pilote n° 6	650	595	630	710	660	649	4
Moyenne	735,0	661,7	668,3	787,5	725,8		
Ecart-type	108,4	97,7	204,5	136,9	127,8		

Le tableau indique les scores attribués aux 6 pilotes par les 5 juges.

Dans cet exemple, le collège de juge possède les particularités suivantes :

Juge1 : les scores sont normalement répartis sans écart significatif

Juge 2 : Les scores sont aussi normalement répartis avec un niveau relativement bas par rapport à la moyenne

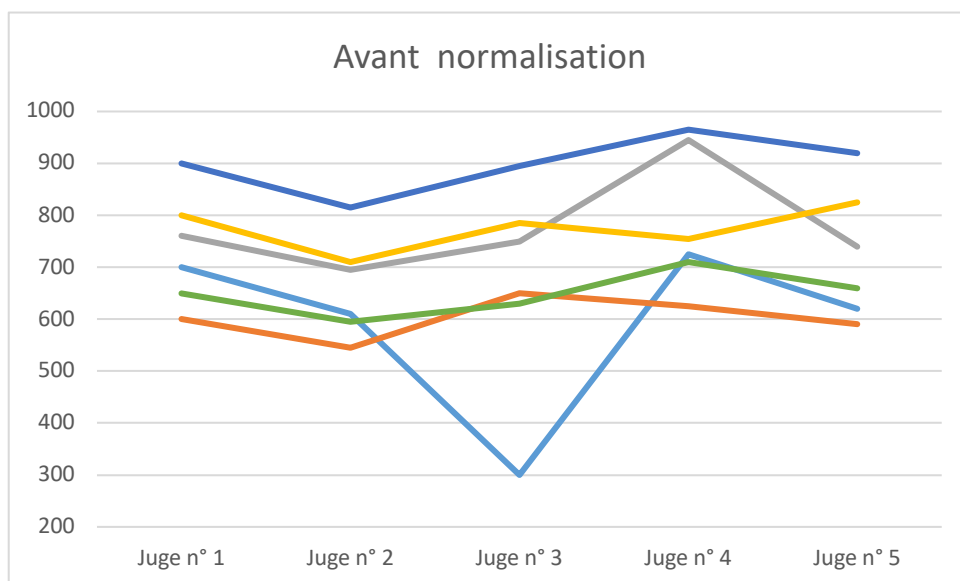
Juge 3 : La moyenne est correcte, mais il a sévèrement jugé le pilote n° 1

Juge 4 : Ce juge a été très clément avec le pilote n° 3 et très sévère avec le pilote n° 4

Juge 5 : similaire au juge 1

Les deux colonnes de droite indiquent les scores de l'ensemble du jury (la moyenne des scores) et le classement brut.

La représentation graphique de ces scores figure ci-dessous :



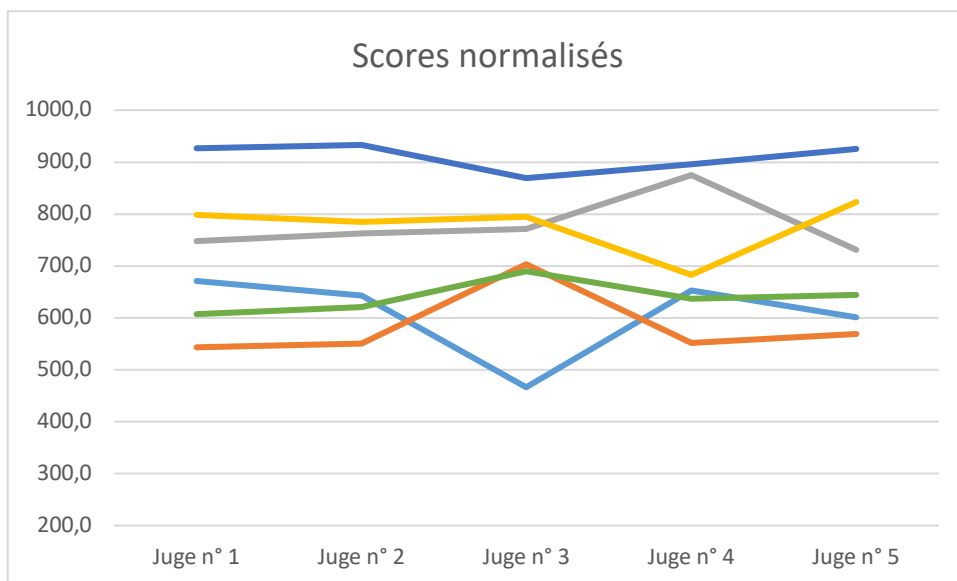
Sans préjuger de la suite, nous pouvons déjà constater que les moyennes vont de 661 à 785 et que les écarts-types vont de 97 à 204.

A l'aide d'un coefficient, nous allons donc « normaliser » les scores. Les formules mathématiques sont explicitées à la fin de ce document.

Le résultat va être le suivant :

Score normalisé							
	Juge n° 1	Juge n° 2	Juge n° 3	Juge n° 4	Juge n° 5	Moyenne	Rang
Pilote n° 1	671,0	642,4	466,4	652,5	601,1	606,7	5
Pilote n° 2	543,3	550,3	703,3	551,3	568,6	583,3	6
Pilote n° 3	747,6	762,9	770,9	874,9	731,0	777,5	2
Pilote n° 4	798,7	784,2	794,6	682,8	823,0	776,7	3
Pilote n° 5	926,4	933,0	869,1	895,2	925,9	909,9	1
Pilote n° 6	607,1	621,2	689,7	637,3	644,4	639,9	4
Moyenne	715,7	715,7	715,7	715,7	715,7		
Ecart-type	138,4	138,4	138,4	138,4	138,4		

Les scores sont « normalisés », c'est-à-dire que les moyennes sont identiques, les écarts-types aussi. L'étape 1 est terminée, le graphique ressemble à ceci :



Etape 2

A partir de ces scores normalisés, on va ensuite calculer un intervalle de confiance à 90 %. Sans entrer dans les détails statistiques, cet intervalle détermine la valeur minimum et la valeur maximum entre lesquelles le « vrai score » se trouve avec une probabilité de 90 %.

Muni de cet intervalle de confiance, nous allons pouvoir « élaguer » le tableau des scores, c'est-à-dire que l'on va éliminer les scores qui se trouvent en dehors de cet intervalle de confiance. Le tableau des scores est donc maintenant celui-ci :

	Juge n° 1	Juge n° 2	Juge n° 3	Juge n° 4	Juge n° 5
Pilote n° 1	671,0	642,4	(466,4)	652,5	601,1
Pilote n° 2	543,3	550,3	(703,3)	551,3	568,6
Pilote n° 3	747,6	762,9	770,9	(874,9)	731,0

Pilote n° 4	798,7	784,2	794,6	(682,8)	823,0
Pilote n° 5	926,4	933,0	869,1	895,2	925,9
Pilote n° 6	607,1	621,2	689,7	637,3	644,4

Nous avons donc éliminé deux scores pour le juge 3 et deux scores pour le juge 4.

On peut donc recalculer le classement : les nouveaux scores sont les moyennes des scores restants (on ne tient plus compte des scores qui ont été éliminés).

	Juge n° 1	Juge n° 2	Juge n° 3	Juge n° 4	Juge n° 5	Score	Rang	Avant TBL
Pilote n° 1	671,0	642,4	(466,4)	652,5	601,1	641,7	4	6
Pilote n° 2	543,3	550,3	(703,3)	551,3	568,6	553,4	6	5
Pilote n° 3	747,6	762,9	770,9	(874,9)	731,0	753,1	3	2
Pilote n° 4	798,7	784,2	794,6	(682,8)	823,0	800,1	2	3
Pilote n° 5	926,4	933,0	869,1	895,2	925,9	909,9	1	1
Pilote n° 6	607,1	621,2	689,7	637,3	644,4	639,9	5	4

Le classement a été beaucoup modifié. Il s'agit cependant d'un exemple pour lequel les valeurs ont été choisies dans un but pédagogique.

Il faut noter que ce processus est itératif, c'est-à-dire que l'on va le répéter jusqu'à ce qu'il ne reste plus de scores à éliminer.

Accessoirement, ces calculs permettent aussi d'estimer la « performance » des juges, c'est-à-dire le pourcentage de leurs scores qui ont été acceptés, l'écart entre leur moyenne et la moyenne générale et l'écart entre leur écart-type et l'écart-type général.

Conclusion

Qu'est-il arrivé ?

L'étape 1 a permis de résoudre les écarts des juges 2 et 3 (scores trop bas) et du juge 4 (scores trop hauts). Cette étape a aussi permis d'identifier un problème de style de jugement pour le juge 3.

En examinant le tableau de l'étape 2, vous pouvez voir :

- que les juges 3 et 4 qui n'étaient pas "corrects" pour deux pilotes ont vu leurs scores éliminés.
- Que le score du juge 3 qui avait surévalué le pilote 2 a lui aussi été éliminé

Note : si un pilote pour n'importe quelle raison ne peut pas effectuer son vol ou bien si son modèle ne satisfait pas au test de bruit, le vol est noté zéro. Ce score ne doit alors pas être pris en compte dans le calcul TBL.

Formules

Pour ceux qui veulent aller au bout des choses ou reproduire le fonctionnement, voici l'intégralité des formules utilisées :

Etape 1

N_j : nombre de juges ayant donné au moins un score différent de 0. Permet de ne compter que les juges ayant réellement officié.

N_p : nombre de pilotes ayant reçu au moins un score différent de 0. Cf remarque ci-dessus.

tabScore(i, j) : tableau des scores bruts. i représente le pilote, j représente le juge. tabScore(25, 3) représente le score du pilote 25 pour le juge 3.

Moy(j) : tableau des moyennes des scores. Moy(2) contient la moyenne des scores du juge 2 pour tous les pilotes.

MoyGen : moyenne générale des scores pour tous les pilotes et tous les juges.

SD(j) : tableau des écarts-types (Standard Deviation) des juges. SD(3) contient l'écart type des scores du juge 3 pour tous les pilotes.

$$SD(j) = \sqrt{\frac{1}{Np} \sum_{i=1}^{Np} [\{tabScore(i, j) - Moy(j)\}^2]}$$

SDGen : écart-type pour tous les juges et tous les pilotes.

$$SDGen = \sqrt{\frac{1}{Np * Nj} \sum_{i=1}^{Np} \sum_{j=1}^{Nj} [\{tabScore(i, j) - MoyGen\}^2]}$$

Kj(j) : Tableau des coefficients utilisés pour normaliser les scores. Un coefficient pour chaque juge

$$Kj(j) = SDGen / SD(j)$$

tabScore2(i, j) : tableau des scores **normalisés**. i représente le pilote, j représente le juge.

tabScore2(25, 3) représente le score du pilote 25 pour le juge 3.

$$tabScore2(i, j) = Kj(j) * (tabScore(i, j) - Moy(j) + MoyGen)$$

Etape 2

Cette étape va être répétée jusqu'à ce qu'il n'y ait plus de score à éliminer.

tabMoyPilote(i) : tableau de la moyenne des scores des pilotes. tabMoyPilote(12) représente la moyenne des scores de tous les juges pour le pilote n° 12. Attention, on ne tient pas compte des scores déjà éliminés !

tabSDPilote(i) : tableau des écarts-types des scores des pilotes. tabSDPilote(12) représente l'écart-type des scores de tous les juges pour le pilote n° 12. Même remarque pour les scores éliminés.

$$tabSDPilote(i) = \sqrt{\frac{1}{Nj} \sum_{j=1}^{Nj} [\{tabScore2(i, j) - tabMoyPilote(i)\}^2]}$$

Calcul de l'intervalle de confiance à 90% pour chaque pilote :

tabMinScore(i) : tableau contenant la limite basse de l'intervalle de confiance pour le pilote i

$$tabMinScore(i) = tabMoyPilote(i) - tabSDPilote(i) * 1.645$$

tabMaxScore(i) : tableau contenant la limite haute de l'intervalle de confiance pour le pilote i

$$tabMaxScore(i) = tabMoyPilote(i) + tabSDPilote(i) * 1.645$$

Elimination des scores en dehors de l'intervalle de confiance : si tabScore2(i) < tabMinScore(i) ou tabScore2(i) > tabMaxScore(i), alors tabScore2(i) = 0.