

---

## Cours 11 : Homogénéité de la variance et transformations non linéaires

---

### Table des matières

Section 1. Régularité de la nature et effets linéaires.....	2
Section 2. Homogénéité des variances.....	2
Section 3. Restauration de l'homogénéité des variances.....	3
3.1. Transformation par la racine carrée.....	3
3.2. Transformation logarithmique .....	4
3.3. Transformation angulaire.....	4
Section 4. Restauration de la linéarité dans le cas d'effets courbes .....	4
Section 5. Conclusion.....	6
Exercices .....	7

### Lectures

Suggérée : Howell, chapitre 11.9.

### Objectifs

Détecter si la variance entre les groupes est homogène préalablement à la réalisation d'une analyse de variance. Si oui, effectuer des transformations dans le but de restaurer l'homogénéité des variances. Détecter si des résultats suivent une tendance précise même si les moyennes ne forment pas une ligne droite. Transformer les données pour que les résultats deviennent linéaires si une régularité est suspectée.

---

## Section 1. Régularité de la nature et effets linéaires

---

La recherche de lois scientifiques passe la découverte de régularités. La meilleure façon est de faire des mesures quand un facteur varie suivant plusieurs niveaux. Lorsqu'un graphe des résultats montre une ligne parfaitement droite, on sait qu'on vient de découvrir une régularité. Dans les faits cependant, il est rare d'obtenir des effets linéaires, surtout en psychologie. En physique, Newton a révolutionné le monde en mettant en évidence un effet linéaire de l'accélération sur la force,  $F(a) = m \times a$ . De la même façon au début du XX<sup>ième</sup> siècle, l'effet photo-électrique était très déroutant jusqu'à ce que Einstein en 1905 trouve une façon de regarder les résultats tel que l'effet photo-électrique était parfaitement linéaire. De façon générale, découvrir une relation linéaire entre un facteur et une mesure est la meilleure indication qu'il existe une régularité, surtout si l'effet se maintient lorsque l'expérience est répétée. Quand l'effet n'est pas linéaire, une façon de vérifier que l'effet est bien réel est de voir si la courbe du graphe ne peut pas être transformée de telle façon que l'effet résultant forme une ligne droite. Ce faisant, le chercheur démontre la régularité de façon frappante, et en plus, il montre quelle peut bien être la forme de la courbe avant transformation.

En psychologie, une loi importante est la loi de l'apprentissage. Cependant, il ne s'agit pas d'un effet linéaire: En début d'apprentissage, les participants apprennent très vite à faire mieux (moins d'erreur ou réponses plus rapides, selon ce que le chercheur mesure). Par contre, au fur et à mesure que les participants approchent de leur plancher, l'amélioration se fait de plus en plus lentement. L'ensemble forme une courbe du genre:

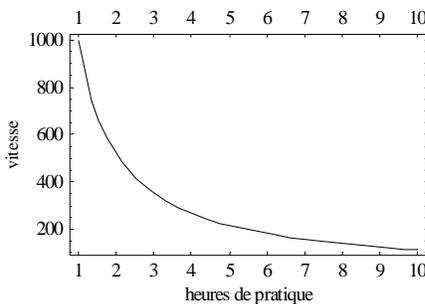


Figure 1 : Performances typiques obtenues sur 10 heures d'entraînement

Newell et Rosenbloom ont montré en 1981 la régularité de cette courbe en trouvant une transformation des axes qui rend la courbe parfaitement linéaire et ont montré sur une cinquantaine d'ensembles de données obtenus sur une grande variété de tâches très différentes les unes des autres (tel le temps pour rouler un cigare, le temps pour trouver Charlie, le nombre d'erreur de multiplication, etc.) qu'à chaque fois, la courbe devenait très linéaire. Ils ont été les premiers à détecter "la loi d'apprentissage".

---

## Section 2. Homogénéité des variances

---

Un postulat important à la base de l'ANOVA concerne l'homogénéité des variances. Plusieurs études semblent démontrer que dans le cas d'échantillons d'effectifs égaux, le test  $F$  est un test robuste, c'est à dire qu'il est peu affecté par la violation du postulat de normalité

des données brutes et la violation de l'homogénéité des variances. C'est une des raisons pour laquelle plusieurs auteurs recommandent fortement l'utilisation de groupes égaux. Toutefois, Wilcox (1987) démontre que même dans le cas d'échantillons de tailles égaux, le test  $F$  peut être affecté par une hétérogénéité sévère des variances, d'où l'importance d'une vérification préalable à l'aide de tests statistiques.

Conover et al. (1981) ont recensé et comparé 56 tests d'homogénéité des variances. Un des meilleurs est le test de Levene (1960) et sa variante, le test de Brown-Forsythe (1976).

Le test de Levene consiste à utiliser une analyse de variance simple sur des valeurs  $Z_{ki}$  qui sont la distance de chaque donnée brute  $X_{ki}$  à la moyenne de son groupe  $\bar{X}_i$ . Comme on soustrait la moyenne, toutes les valeurs  $Z_{ki}$  ne devraient refléter que la variance autour de zéro. S'il existe encore un effet du traitement, cela signifie que la variance n'est pas également répartie entre les conditions.

Formellement

$$Z_{ki} = |X_{ki} - \bar{X}_i|$$

$$H_0: \sigma^2_1 = \sigma^2_2 = \dots = \sigma^2_p$$

Le rejet de l'hypothèse nulle (un ratio  $F$  significativement supérieur à 1.00) implique la présence d'hétérogénéité dans les variances et la nécessité de recourir à une transformation non linéaire des données dans le but de ramener les scores extrêmes plus près de la moyenne.

Une variante du test de Levene est le test de Brown-Forsythe. Elle consiste à soustraire la médiane dans chaque condition plutôt que la moyenne.

Malheureusement, ces deux tests ne sont pas réalisés automatiquement par le logiciel SPSS. Il faut donc soit écrire un programme qui va calculer les  $Z_{ki}$  avant de faire une ANOVA usuelle, ou alors utiliser les autres tests incorporés dans le logiciel SPSS.

Ces tests alternatifs sont le test de Cochran et le test de Bartlett.

---

### Section 3. Restauration de l'homogénéité des variances

---

N'importe quelle transformation non linéaire peut être réalisée sur les données brutes. Cependant, il faut savoir quel est votre objectif avant de procéder.

Si l'hypothèse d'homogénéité des variances est rejetée, il est nécessaire d'effectuer l'une des transformations suivantes qui aura pour effet de normaliser les distributions, soit de rendre les variances plus comparables entre les conditions. Notons qu'une transformation est très facile à réaliser avec la plupart des logiciels d'analyses statistiques.

#### 3.1. Transformation par la racine carrée

---

Cette transformation implique de changer les  $X_{ki}$  ainsi :

$$X'_{ki} = \sqrt{X_{ki} + 0.5}$$

Elle est souvent utile lorsque la variable dépendante est une fréquence, par exemple, le nombre d'erreurs, le nombre de réponses, etc.

### 3.2. Transformation logarithmique

Cette transformation est parfois utilisée lorsque la variable dépendante est mesurée en unité de temps, par exemple, secondes, minutes, etc. Cette transformation a pour effet de stabiliser la variance et de normaliser les distributions qui ont une asymétrie positive. Elle s'obtient avec la transformation suivante :

$$X'_{ki} = \text{Log}(X_{ki} + 1.0)$$

### 3.3. Transformation angulaire

Dans une expérience, il arrive souvent que la variable dépendante soit exprimée en pourcentage, ce qui se produit par exemple lorsqu'on mesure le nombre de bonnes réponses sur un maximum possible. Dans ce cas, si les variances sont hétérogènes, il faut procéder à une transformation angulaire de la forme :

$$X'_{ki} = 2\text{ArcSin}\sqrt{X_{ki}}$$

Dans ce cas,  $X'$  devient une mesure exprimée en radians. La variable  $X$  doit être incluse entre 0 et 1, sinon la transformation donnera des résultats aberrants.

## Section 4. Restauration de la linéarité dans le cas d'effets courbes

Dans un grand nombre d'expérience, on utilise des niveaux qui s'accroissent continûment, suivant une échelle de type II. Dans ces cas, la relation devient très facile à analyser (comme on le verra aussi au cours 12), et la pente de la droite a une signification directe. Par exemple, supposons une expérience de recherche visuelle. Dans cette expérience, on montre à un sujet une scène contenant 1, 2, 3, ... objets, et le sujet doit décider aussi rapidement que possible si un objet précis (la cible) est présent. L'exemple par excellence se trouve dans les livres « Où est Charlie ». Le graphe suivant illustre le type de données que l'on peut obtenir dans un plan à 4 mesures répétées sur le nombre d'objets (1 à 4).

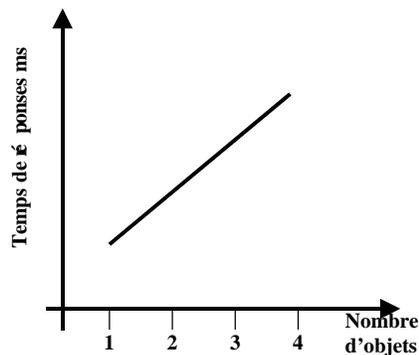
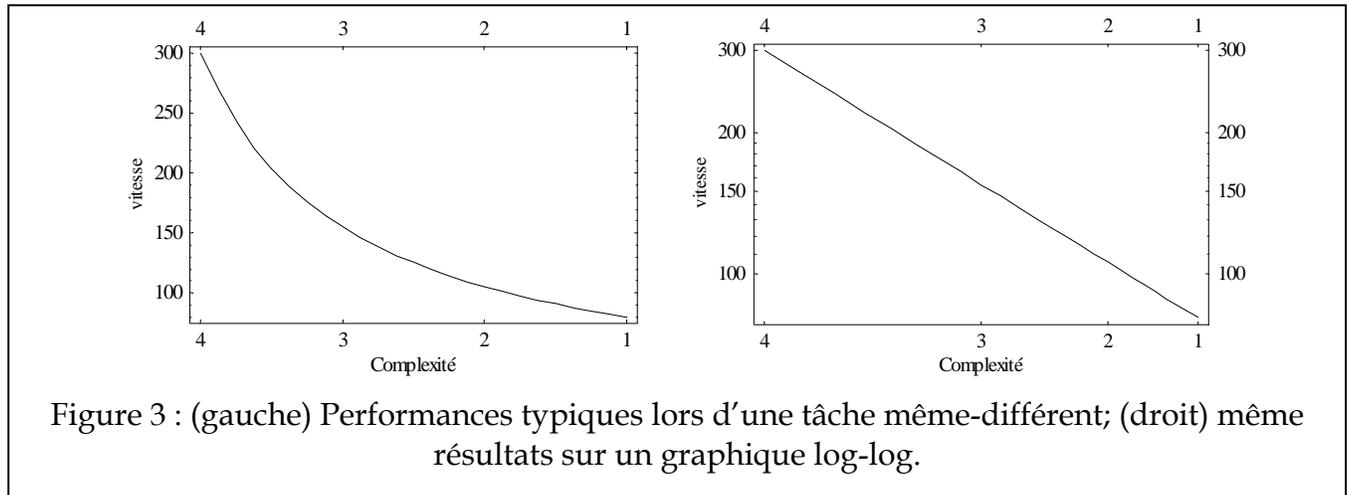


Figure 2 : Performances typiques lors de la recherche d'un objet cible parmi plusieurs objets distrayants

Dans ce cas, la pente de la fonction nous indique le temps qu'il faut pour traiter chaque objet supplémentaire. Il s'agit donc d'une indication du mouvement de l'attention.

À l'occasion, on obtient des effets qui, sans être linéaires, semblent dotés d'une régularité très précise. Par exemple, on peut étudier le temps qu'il faut pour décider si deux objets successifs sont identiques ou différents. On trouve que le temps pour dire « identique » s'accroît avec la complexité des objets, comme on le voit sur le panneau de gauche de la Figure 3. Cependant, l'accroissement n'est pas linéaire, mais forme plutôt une courbe concave. Dans ces cas, pour voir si la relation est vraiment stable plutôt que simplement illusoire, on peut faire un graphe de type log-log



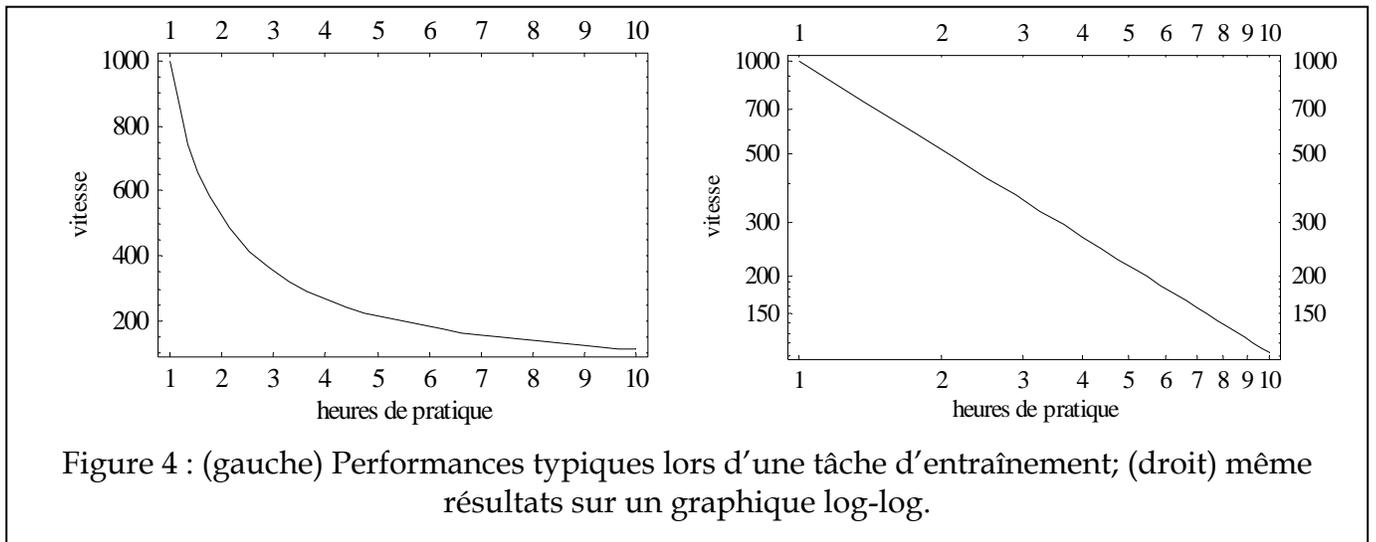
Comme on le voit sur le panneau de droite de la Figure 3, avec un graphe log-log, les données sont parfaitement linéaires, ce qui suggère que la courbe empirique obtenue à gauche n'est pas farfelue ou résultante du hasard.

Utiliser un graphe de type log-log revient à transformer les données brutes  $X_{ki}$  et les niveaux  $a_i$ . La relation est simplement  $a'_i = \text{Log}(a_i)$  et  $X'_{ki} = \text{Log}(X_{ki})$ . Ce genre de transformation est souvent utilisé en ingénierie, et commence à l'être en psychologie, notamment en perception, où il semble exister un grand nombre de lois très stables, mais non linéaires.

Un autre exemple concerne la loi d'apprentissage. Il semble bien établi que les performances s'améliorent avec la pratique suivant une équation du type:

$$\text{Perf}(N) = a + b N^{-c}$$

où  $\text{Perf}$  mesure la performance après  $N$  heures d'entraînement, et  $a$ ,  $b$ ,  $c$  sont les paramètres du sujet. Souvent,  $\text{Perf}$  est le temps pris pour compléter une tâche, temps qui se réduit avec la pratique. La courbe ressemble à celle du panneau de gauche de la Figure 4 qui devient linéaire sur un graphe de type log-log, comme on le voit sur le panneau de droite de la Figure 4.



---

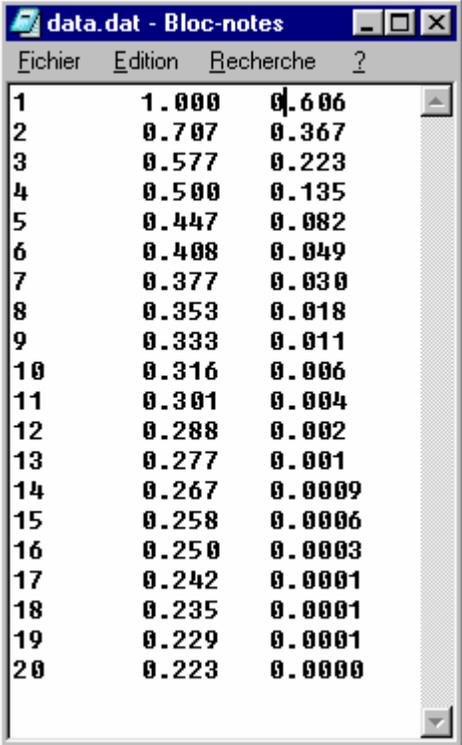
## Section 5. Conclusion

---

*Exercices*

1. Soit les données de la Figure 5 (fictives) représentant des taux d'erreur d'un participant dans deux tâches différentes en fonction du nombre d'heure d'entraînement (première colonne):

Log(session)? Si oui, quel type de graphe avez-vous fait?



Session	Tâche 1 (Colonne 2)	Tâche 2 (Colonne 3)
1	1.000	0.606
2	0.707	0.367
3	0.577	0.223
4	0.500	0.135
5	0.447	0.082
6	0.408	0.049
7	0.377	0.030
8	0.353	0.018
9	0.333	0.011
10	0.316	0.006
11	0.301	0.004
12	0.288	0.002
13	0.277	0.001
14	0.267	0.0009
15	0.258	0.0006
16	0.250	0.0003
17	0.242	0.0001
18	0.235	0.0001
19	0.229	0.0001
20	0.223	0.0000

Figure 5 : Données de la question 1

- Est-ce que les effets sont linéaires dans la tâche 1 (colonne 2)? dans la tâche 2 (colonne 3). Pour répondre à la question, faites un graphe des moyennes.
- Calculer dans deux autres colonnes le  $\log(1/n)$  du taux d'erreur dans la tâche 1, dans la tâche 2. Est-ce que l'effet devient linéaire en fonction de session? Si oui, quel type de graphe avez-vous fait?
- Calculer dans une dernière session le  $\log$  du nombre d'heure d'entraînement. Est-ce que l'effet devient linéaire en fonction de