

# Méthodes statistiques pour la comparaison de plus de deux traitements I\*

Paul De Munter, professeur à l'Université de Tunis.

## PLAN

Chapitre 1. — *Comparaison simultanée de  $k > 2$  traitements. Hypothèses d'homogénéité contre alternatives générales et particulières.*

- 1.1. Formulation du problème. Hypothèses et notations.
- 1.2. Populations normales homoscedastiques.
  - 1.2.1. Homogénéité contre alternatives générales.
    - 1.2.1.1. Observations non progressives.
    - 1.2.1.2. Observations progressives.
  - 1.2.2. Homogénéité contre alternatives particulières.
- 1.3. Populations normales hétéroscedastiques.
- 1.4. Populations non normales.
  - 1.4.1. Homogénéité contre alternatives générales.
  - 1.4.2. Homogénéité contre alternatives particulières.
- 1.5. Références bibliographiques pour ce chapitre.

*Remarque* : Les autres chapitres paraîtront dans les numéros suivants de cette revue.

Ce sont les :

Chapitre 2 : Sélection du « meilleur » traitement.

Chapitre 3 : Comparaison simultanée de  $k > 2$  traitements à un témoin.

Chapitre 4 : Comparaison de  $k > 2$  traitements deux-à-deux — Classement des traitements en deux groupes — Tests de contrastes — (Populations normales homoscedastiques).

Au cours des dernières années, les méthodes d'analyse statistique d'un groupe de plus de deux moyennes traitements se sont développées et perfectionnées; des tables numériques nouvelles ont été mises à la disposition de l'expérimentateur.

Nous reprenons la question à partir du test F basé sur l'analyse de la variance dans le cas normal et passons en revue les différents problèmes pratiques qui se posent à l'expérimentateur et pour lesquels une solution existe actuellement (par exemple : comparaison des traitements deux-à-deux, sélection du « meilleur », comparaison à un témoin). Nous citons les conditions d'applicabilité des méthodes et leurs qualités; nous nous attachons particulièrement au problème de la détermination a priori de l'effectif des observations. Les méthodes sont classées, pour chaque problème pratique, suivant les conditions dans lesquelles elles sont appliquées et suivant le but qu'il faut atteindre (par exemple : cas normal ou non-normal, homogénéité des variances, effectif fixe ou progressif, test d'hypothèse ou intervalle de confiance).

Notre objectif n'est pas d'apporter une contribution originale de nature mathématique, quoique certaines méthodes ou conditions soient nouvelles (leur justification paraîtra ultérieurement). Nous voulons aider l'expérimentateur en lui permettant, une fois le problème posé, de faire l'inventaire des méthodes et de les appliquer en toute connaissance de cause.

Nous remercions Monsieur le Professeur Jean Teghem d'avoir bien voulu lire le manuscrit et d'y avoir introduit de nombreuses améliorations.

---

(\*) Cet article est le premier d'une série qui paraîtra dans les prochains numéros de cette revue.

1. — COMPARAISON SIMULTANEE DE  $k > 2$  TRAITEMENTS. HYPOTHESES D'HOMOGENEITE CONTRE ALTERNATIVES GENERALES ET PARTICULIERES.

1.1. Formulation du problème. Hypothèses et notations.

a) Alternatives générales et particulières.

Nous savons que la notion de traitement est équivalente à la notion combinée de population — fonction de distribution (De Munter, 1960). Notons  $\pi_i$  la population des réponses possibles au  $i^e$  traitement et soit  $F_i(x)$  la fonction de distribution de  $\pi_i$ , ( $i = 1, 2, \dots, k$ ). L'égalité des traitements est donc un fait équivalent à l'égalité des fonctions  $F_i$ . D'une manière générale, on est amené à tester l'hypothèse

$$H_0 : F_1(x) = F_2(x) = \dots = F_k(x),$$

pour tout  $x$ . L'hypothèse  $H_0$  exprime l'homogénéité des populations  $\pi_i$ . Lorsque  $H_0$  est vraie, il n'y a plus  $k$  populations, mais une seule, dont la fonction de distribution est la fonction  $F_1(x)$  par exemple.

Lorsqu'une méthode statistique quelconque conduit à rejeter  $H_0$ , on est en droit de penser qu'une alternative  $H$  à  $H_0$  est vraie. Or, si l'on admet qu'une alternative  $H$  exprime seulement qu'une au moins des égalités entre les  $F_i$  tombe en défaut, il y a  $2^k - 1$  alternatives  $H$ . S'il y a 5 traitements en présence, il y a 31 alternatives à l'hypothèse d'homogénéité. Nous dirons des alternatives  $H$  qu'elles sont générales.

Parmi les alternatives générales, il peut y en avoir que l'expérimentateur sait, a priori, être impossibles. Il peut en être de plus ou moins vraisemblables. Enfin, il se peut que l'expérimentateur puisse, a priori, déterminer un ensemble restreint d'alternatives seules possibles. Par exemple, seules peuvent être possibles les alternatives  $H : F_1 < F_2 < \dots < F_k$ . Nous dirons dans ce cas que les alternatives sont particulières.

Il est (intuitivement) justifié de croire qu'un test statistique de  $H_0$  contre les alternatives générales est moins efficace qu'un test de  $H_0$  contre des alternatives particulières. Un test de  $H_0$  contre  $H_g$  (alternatives générales) rejettera  $H_0$  moins facilement qu'un test de  $H_0$  contre  $H_p$  (alternatives particulières) surtout si, justement, c'est  $H_p$  qui est vraie.

Notons  $D_0$  la décision statistique de considérer  $H_0$  comme l'hypothèse vraie et soit  $D$  la décision statistique de considérer qu'une alternative  $H$  (générale ou particulière) est vraie. Prendre la décision  $D_0$  revient à accepter  $H_0$  et prendre la décision  $D$  revient à rejeter  $H_0$  en faveur d'une alternative  $H$  générale ou particulière.

b) Détermination a priori de l'effectif échantillon.

D'une manière générale, on fixe a priori la probabilité de prendre la décision  $D_0$  lorsque  $H_0$  est vraie, soit  $P\{D_0 | H_0\}$ . Nous appellerons « condition  $(C_2)$  » la nécessité pour  $P\{D_0 | H_0\}$  d'avoir une certaine valeur fixée à l'avance ; on aura généralement

$$(C_2) \equiv P\{D_0 | H_0\} = 1 - \alpha \quad (= 0,95),$$

où  $\alpha$  s'appelle « niveau de signification du test » ou « erreur de première espèce ». Mais il est tout aussi important de fixer a priori la probabilité de prendre la décision  $D$  lorsque  $H$  est vraie, soit  $P\{D | H\}$ . Nous appellerons « condition  $(C_1)$  » la nécessité pour  $P\{D | H\}$  d'avoir une valeur fixée à l'avance ; on aura généralement

$$(C_1) \equiv P\{D | H\} = 1 - \beta \quad (= 0,95 \text{ ou } 0,90),$$

où  $\beta$  s'appelle « erreur de deuxième espèce ».

Les symboles  $\alpha$  et  $\beta$  ont effectivement la signification d'une erreur. En effet, on a

$$\alpha = 1 - P\{D_0 | H_0\} = P\{D | H_0\},$$

c'est-à-dire, la probabilité de rejeter  $H_0$  alors que  $H_0$  est l'hypothèse réellement vraie. D'autre part, on a

$$\beta = 1 - P\{D | H\} = P\{D_0 | H\},$$

c'est-à-dire, la probabilité d'accepter  $H_0$  alors qu'une alternative  $H$  est vraie.

Pour exploiter la condition (C<sub>1</sub>), il faut préciser H. Par exemple, H : F<sub>1</sub>(x) = F<sub>2</sub>(x) = ... = F<sub>k-1</sub>(x) = F(x), F<sub>k</sub>(x) = F(x - c), exprimant que les k - 1 premiers traitements sont identiques et que le k<sup>e</sup> diffère, seulement en ce qui concerne sa moyenne, d'une quantité c bien précisée. Pour exploiter la condition (C<sub>1</sub>) il faut donc particulariser H. Il est évident que cette particularisation se fait dans le sens le plus plausible. Si, par exemple, on doit comparer trois engrais dont l'un diffère, par ses qualités chimiques, nettement des deux autres, les alternatives à H<sub>0</sub> sont du type cité plus haut. Dans ces conditions, on peut tirer de la condition (C<sub>1</sub>) le nombre d'observations qu'il convient de prélever dans chaque population. (C'est la raison pour laquelle nous avons affecté à la condition sur l'erreur de deuxième espèce l'indice « 1 » : on commence par fixer α et β d'où l'on tire tout d'abord l'effectif échantillon afin de pouvoir entreprendre l'essai. Ensuite, de la condition (C<sub>2</sub>) on tire la règle du test qui permet de prendre une décision). Mais il ne faut pas perdre de vue que cette manière de faire introduit une nouvelle sorte d'erreur. Il a fallu, pour calculer P{D|H} faire un choix parmi les alternatives H. Soit H' l'alternative choisie et D' la décision de considérer H' comme vraie. On a donc P{D'|H'} = 1 - β. Mais cela ne donne aucune garantie de prendre la décision D' si H''(≠ H') est vraie.

### c) Méthodes progressives.

Il est possible de fixer a priori P{D<sub>0</sub>|H<sub>0</sub>} et P{D|H} sans pour autant que cela impose un effectif échantillon fixe. Alors, la méthode est dite progressive (ou séquentielle). Cela signifie que l'effectif échantillon est aléatoire ; il dépend du comportement de la variable observée. Les observations sont prélevées successivement et à chaque prélèvement on est amené à prendre l'une des trois décisions : D<sub>0</sub>, D, faire un nouveau prélèvement. Les méthodes progressives pour tester H<sub>0</sub> contre H ne sont ni nombreuses, ni développées. Nous en décrivons une due à Johnson (1953) et améliorée par Ray (1956).

### d) Homogénéité relative.

Supposons que l'on soit amené à comparer trois engrais dont l'application a pour seul but d'accroître le rendement. Et l'on suppose que ce but est atteint et seulement celui-là. En d'autres termes, l'application de ces engrais ne perturbe pas, par exemple, la dispersion des réponses. Dans ces conditions, la moyenne de la population des réponses à l'un des engrais en est la caractéristique essentielle.

Mais il se peut que les populations des réponses possibles aux différents engrais aient des fonctions de distribution de formes différentes. Cela n'intéresse pas l'expérimentateur. Ce qui l'intéresse essentiellement, c'est le comportement respectif des moyennes de ces populations. Dans ce cas, l'homogénéité des traitements est équivalente à une homogénéité relative des fonctions de distribution, relative aux valeurs moyennes seulement. L'hypothèse généralement testée est H<sub>0</sub> : μ<sub>1</sub> = μ<sub>2</sub> = ... = μ<sub>k</sub>, où μ<sub>i</sub> est la moyenne population du i<sup>e</sup> traitement.

### e) Hypothèse de normalité.

Dans ce qui précède, rien n'a été dit ni supposé quant à la forme des fonctions de distribution F<sub>i</sub>(x) des populations π<sub>i</sub>.

Lorsqu'il est concevable (i) que la variable aléatoire observée puisse prendre n'importe quelle valeur réelle, (ii) que la variable aléatoire observée résulte de l'addition d'un très grand nombre d'impulsions élémentaires représentatives de facteurs aléatoires plus ou moins indépendants, il est justifié de supposer que la fonction de distribution de cette variable est normale ou approximativement normale. Rappelons que si X est une variable normale,

$$P\{X \leq x\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt,$$

où μ est la moyenne et σ l'écart-type de X ; on note X est N(μ,σ).

Considérons les  $k$  traitements et les populations  $\pi_i$  des réponses possibles. S'il est concevable d'admettre que les réponses possibles satisfont aux conditions (i) et (ii), il est justifié de supposer que la fonction de distribution de  $\pi_i$  est normale de moyenne  $\mu_i$  et d'écart-type  $\sigma_i$ . Si les populations ont toutes la même variance ( $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$ ), nous dirons qu'elles sont homoscédastiques ; sinon elles sont hétéroscédastiques.

Nous envisagerons tout d'abord le cas de l'homoscédasticité. Il est fort important d'en bien comprendre la portée sur la nature des traitements. Exemple : Une usine de produits pharmaceutiques met au point trois fortifiants A, B et C pour enfants. Elle désire ne mettre sur le marché que le « meilleur ». Les traitements sont appliqués à trois groupes d'enfants (20 enfants par groupe, par exemple) pendant un mois. Ces enfants ont été choisis au hasard dans une population dont tous les éléments satisfont à certains critères portant sur l'âge, le poids initial, le standing social, l'état de santé, etc. De telle sorte, qu'on a éliminé les facteurs principaux d'hétérogénéité systématique pouvant affecter les réponses aux traitements. Dès l'application des traitements, la population unique des enfants se scinde en trois, respectivement caractéristique de chaque traitement. Supposer que ces populations sont normales implique une hypothèse sur les traitements : les réponses se répartissent symétriquement autour de la moyenne ; il y a autant de petits accroissements de poids que de grands. Supposer en outre que les fonctions de distribution sont homoscédastiques implique que les réponses sont également dispersées. Or, le meilleur traitement est également celui qui agit le plus indépendamment des individus, c'est-à-dire, celui qui a la dispersion la plus faible. Donc faire l'hypothèse de l'homoscédasticité, c'est faire a priori une hypothèse sur la qualité des traitements (cet exemple est repris au chapitre 2, § 1).

Admettons qu'il soit permis de faire l'hypothèse que les fonctions de distribution des populations des réponses possibles aux différents traitements soient normales homoscédastiques (ce qui se note «  $\pi_i$  est  $N(\mu_i, \sigma)$  »,  $\mu_i$  étant la moyenne et  $\sigma$  l'écart-type commun). Les populations ne diffèrent que par les valeurs moyennes  $\mu_i$ . L'hypothèse statistique  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$  est donc équivalente à l'hypothèse d'homogénéité.

Dans le cas de l'hétéroscédasticité, les fonctions de distribution des populations des réponses possibles sont  $N(\mu_i, \sigma_i)$  et l'hypothèse  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$  n'implique qu'une homogénéité relative des fonctions de distribution.

Les alternatives à  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$  sont soit générales et impliquent qu'un signe égal au moins tombe en défaut, soit particulières et sont, par exemple,  $H_p : \mu_1 \geq \mu_2 \geq \dots \geq \mu_k$  où un signe égal au moins n'existe pas.

Il est souvent possible, par une transformation adéquate des observations, de rendre homoscédastiques des fonctions de distribution présentant une certaine hétérogénéité de variance (De Munter, 1958).

## f) Non-normalité.

Lorsqu'il n'est pas permis de supposer que les fonctions de distribution des populations sont normales, il peut également être possible, par une transformation adéquate des données, de les normaliser (De Munter, 1958). Sinon, on applique des méthodes valables pour n'importe quelles fonctions de distribution. Dans ce cas, on teste l'hypothèse d'homogénéité  $H_0 : F_1 = F_2 = \dots = F_k$  contre  $H_1$  ou  $H_p$ .

### 1.2. Populations normales homoscédastiques.

#### 1.2.1. Hypothèse d'homogénéité contre alternatives générales.

##### 1.2.1.1. Observations non-progressives.

Soit  $\pi_i$  la population des réponses possibles au  $i^{\circ}$  traitement, ( $i = 1, 2, \dots, k$ ). La fonction de distribution de  $\pi_i$  est  $N(\mu_i, \sigma)$ . Soit  $Y_{ij}$  la réponse de la  $j^{\circ}$  unité expérimentale ayant subi le  $i^{\circ}$  traitement. On suppose que  $Y_{ij}$  est la somme d'une moyenne générale  $\mu$ , d'une impulsion  $\tau_i$  due essentiellement au  $i^{\circ}$  traitement et d'une variable aléatoire  $X_{ij}$  dont la fonction de distribution est  $N(0, \sigma)$ . On a donc

$$Y_{ij} = \mu + \tau_i + X_{ij},$$

où  $X_{ij}$  résulte de l'influence de tous les facteurs aléatoires de perturbation autres que le  $i^{\circ}$  traitement. On appelle donc  $X_{ij}$  l'erreur affectant la réponse de la  $j^{\circ}$  unité expérimentale.

tale ayant subi le  $i^{\circ}$  traitement. Les erreurs  $X_{ij}$  sont des variables normales homoscédastiques indépendantes. On peut généralement faire cette hypothèse d'indépendance grâce à une répartition aléatoire des traitements parmi les unités expérimentales : les facteurs non-contrôlés agissent alors sans tendance privilégiée. Le terme  $\mu$  a la signification d'une moyenne générale perçue que, d'une part les  $\tau_i$  sont tels que  $\sum \tau_i = 0$ , et d'autre part  $X_{ij}$  a pour moyenne zéro. Il en résulte que la moyenne du  $i^{\circ}$  traitement est

$$\mu_i = \mu + \tau_i$$

et que  $\sum \mu_i/k = \mu$ ; par conséquent, la différence entre les moyennes des traitements (1) et (2), par exemple, est égale à  $\tau_1 - \tau_2$ .

L'égalité des moyennes traitements est réalisée lorsque l'hypothèse  $H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0$  est vraie. L'hypothèse  $H_0$  exprime donc l'homogénéité des populations normales. Une alternative  $H_g$  est générale si elle exprime seulement que, pour deux valeurs de  $i$  au moins,  $\tau_i \neq 0$ . Tout écart entre  $H_0$  et  $H_g$  peut donc se mesurer par le coefficient

$$\lambda = \sqrt{\frac{\sum \tau_j^2}{k \sigma^2}}$$

Si par exemple,  $\tau_1 = \tau_2 = \dots = \tau_{k-2} = 0$ ,  $\tau_{k-1}$  et  $\tau_k \neq 0$ , on a

$$\lambda = \sqrt{\frac{\tau_{k-1}^2 + \tau_k^2}{k \sigma^2}}$$

L'estimation de  $\lambda$ , a priori, est un élément important de l'analyse statistique. Dans un essai variétal, par exemple, l'expérimentateur doit comparer trois variétés A, B et C de blé du point de vue du rendement en quantité de grain. Si rien n'a jamais été expérimenté sur le sujet, si l'expérimentateur n'a absolument aucune information a priori, il doit organiser un essai d'orientation qui fournira des estimations de  $\sigma$  et des moyennes  $\mu_i$ . Mais généralement, l'expérimentateur possède une certaine information et par conséquent une estimation de  $\sigma$ . Supposons, pour simplifier l'exposé, que l'expérimentateur sache à l'avance que deux variétés, A et B par exemple, sont presque identiques, tandis que la troisième se distingue des deux autres. Une certaine différence minimum entre C et (A, B) est importante (du point de vue économique, par exemple). En d'autres termes, si le rendement moyen de C dépasse les rendements moyens de A et B d'une quantité  $\Delta \geq \Delta^*$ , C présente un avantage économique certain.

On teste donc  $H_0 : \tau_A = \tau_B = \tau_C$  contre  $H_g : \tau_A = \tau_B = \tau_C = \tau + \Delta^*$ , et il convient de prendre, avec une probabilité élevée, la décision D (accepter  $H_g$ ) si  $H_g$  est vraie. Pour satisfaire à cette condition (C), (voir § 1.1.b), il importe de calculer  $\lambda$ . Sachant qu'il faut que  $\tau_A + \tau_B + \tau_C = 0$ , d'où l'on a  $\Delta^* = -3\tau$ , il vient

$$\lambda = \frac{\sqrt{2}}{3} \frac{\Delta^*}{\sigma}$$

On constate que la détermination de  $\lambda$  nécessite une information d'ordre statistique ( $\sigma$  ou une estimation de  $\sigma$ ) et une information inhérente au problème essentiel que l'on veut résoudre (une valeur de  $\Delta^*$  conditionnée par des impératifs économiques, techniques, etc.).

Dans l'exposé de la méthode statistique pour tester  $H_0$  contre  $H_g$ , nous nous limitons tout d'abord au cas où il y a le même nombre  $n$  d'observations prélevées dans chaque population. Il y a donc  $kn = n'$  observations au total. Soit également  $Y_i$ . La somme des observations relatives à la population  $\pi_i$  et soit  $\bar{Y}_i = Y_i./n$  la moyenne. Soit enfin  $Y..$  la somme totale de toutes les observations et  $\bar{Y}.. = Y../n'$  la moyenne générale. Notons  $C = Y^2../n'$  un terme correctif qui interviendra dans la suite.

La variation totale de l'essai est

$$S_T = \sum_{ij} (Y_{ij} - Y..) ^2 = \sum_{ij} Y_{ij}^2 - C$$

La variation (totale) due aux traitements est

$$S_t = n \sum_i (\bar{Y}_i - \bar{Y}_{..})^2 = \frac{1}{n} \sum_i Y_i^2 - C$$

tandis que la variation (totale) due aux erreurs est

$$S_e = \sum_{ij} (Y_{ij} - \bar{Y}_{i.})^2 = S_T - S_t = \sum_{ij} Y_{ij}^2 - \frac{1}{n} \sum_i Y_i^2$$

Les variations moyennes correspondantes sont

$$V_t^2 = S_t / (k - 1) \quad \text{et} \quad V_e^2 = S_e / (n^2 - k),$$

dont le rapport  $F = V_t^2 / V_e^2$  est la variable de Fisher-Snedecor avec  $(k - 1)$  et  $(n^2 - k)$  degrés de liberté.

On prend la décision  $D_0$  si la valeur observée  $f$  de la variable aléatoire  $F$  est telle que (\*)

$$f < f_{\alpha};$$

on prend la décision  $D$  si  $f > f_{\alpha}$ . Le nombre  $f_{\alpha}$  est tiré des tables de la distribution de Fisher-Snedecor; il est tel que la condition  $(C_2)$  soit satisfaite (voir plus loin et § 1.1.b).

L'effectif  $n$  commun à tous les échantillons est déterminé a priori de telle sorte que la probabilité de prendre la décision  $D$  alors que  $H_0$  est vraie soit fixée à l'avance, à savoir,

$$P \{ D | H_0 \} = 1 - \beta, \quad (C_1)$$

où  $H_0$  est spécifiée par une valeur de  $\lambda$ . Nous reproduisons aux tableaux ci-contre les valeurs de  $n$ , correspondant à différents  $\lambda$ , telles que la condition  $(C_1)$  soit satisfaite lorsque  $\alpha = 0,05$  et  $\beta = 0,10$  (table 1.1),  $\beta = 0,05$  (table 1.2).

$k \backslash \lambda$	1	0,9	0,8	0,7	0,6	0,5	0,4	0,3
3	6	7	8	10	12	18	27	47
4	5	6	7	9	11	15	23	40
5	5	5	6	7	10	11	21	35

Table 1.1. Valeurs de  $n$  correspondant à différentes valeurs de  $\lambda$  et de  $k$ , telles que  $P \{ D_0 | H_0 \} = 0,95$  et  $P \{ D | H_0 \} = 0,90$ .

$k \backslash \lambda$	1	0,9	0,8	0,7	0,6	0,5	0,4	0,3
3	7	8	9	11	15	21	34	57
4	6	7	8	10	13	17	27	48
5	6	6	7	9	12	16	24	43

Table 1.2. Valeurs de  $n$  correspondant à différentes valeurs de  $\lambda$  et de  $k$ , telles que  $P \{ D_0 | H_0 \} = 0,95$  et  $P \{ D | H_0 \} = 0,95$ .

(\*) Les majuscules sont des variables aléatoires dont les réalisations sont représentées par les minuscules correspondantes.

Les valeurs de  $n$  fournies par les tables 1.1. et 1.2 ont été tirées des abaques de Feldt et Mahmoud (1958).

La valeur  $f_{\alpha}$ , fournie par les tables de la distribution de Fisher-Snedecor avec  $(k-1)$  et  $(n-k)$  degrés de liberté, est telle que

$$P\{D_0 | H_0\} = 0,95. \quad (C_2)$$

Etant donné une valeur observée  $f$  de  $F$ , il est souvent utile de déterminer  $P\{F > f\}$ . Il est évident qu'on prend la décision  $D$  si  $P\{F > f\} < \alpha$  (généralement 0,05). Les tables de la distribution de Fisher-Snedecor fournissent les  $f_{\alpha}$  tel que  $P\{F > f_{\alpha}\} = \alpha$  mais ne fournissent pas la possibilité de déterminer  $P\{F > f\}$  quel que soit  $f$ . A cette fin, il faut disposer des tables de la fonction bêta incomplète (Pearson, 1956). Ces tables fournissent les valeurs de  $L_p(p, q)$  pour différentes valeurs de  $x$ ,  $p$  et  $q$ . Supposons que la variable  $F$  ait une distribution de Fisher-Snedecor avec  $f_1$  et  $f_2$  degrés de liberté. On a

$$P\{F > f\} = L_p\left(\frac{f_2}{2}, \frac{f_1}{2}\right)$$

où  $x = f_2/(f_2 + f_1)$ .

Le premier exemple numérique est relatif à des données artificielles. Soient trois traitements A, B et C dont les réponses possibles constituent des populations normales homoscédastiques. Afin que l'homoscédasticité ne soit pas mise en doute, les variances échantillons  $s_i^2$  sont toutes égales. Les données sont rassemblées à la table 1.3. Les effectifs échantillons ( $n = 6$ ) ont été fixés sans considération de la condition (C). Il n'empêche qu'il convient de calculer  $P\{D | H_0\}$ . Les moyennes populations sont  $\mu_1 = \mu + \tau_1$ ,  $\mu_2 = \mu + \tau_2$  et  $\mu_3 = \mu + \tau_3$  où  $\mu$  est estimée par  $\bar{y}_{..} = 12$ . Posons  $P\{D_0 | H_0\} = 0,95$ ; quelle est, dans ces conditions,  $P\{D | H_0\}$  si  $H_0: \tau_1 = -1, \tau_2 = -1$  et  $\tau_3 = 2$ ? La valeur de  $\lambda$  qui correspond à  $H_0$  est  $\sqrt{2}/\sigma^2$ , où l'on peut remplacer  $\sigma^2$  par la variance échantillon,  $s_i^2 = 2$ ; donc  $\lambda = 1$ . Par conséquent, en vertu de la table 1.1., on a  $P\{D | H_0\} = 0,90$ .

	$y_{ij}$						$n$	$y_{i.}$	$\bar{y}_{i.}$	$s_i^2$
A	10	10	12	8	9	11	6	60	10	2
B	11	9	13	11	10	12	6	66	11	2
C	15	15	13	17	14	16	6	90	15	2
							18	216	12	

Table 1.3. Données du premier exemple.

Le terme correctif <sup>(°)</sup> a pour valeur  $c = (216)^2/18 = 2592$ . La variation totale a pour valeur  $s_T^2 = (10)^2 + (10)^2 + \dots + (14)^2 + (16)^2 - 2592 = 114$ . La variation (totale) due aux traitements a pour valeur  $s_t^2 = (90)^2 - 2592 = 84$ . Par conséquent, la variation (totale) due aux erreurs vaut  $s_e^2 = s_T^2 - s_t^2 = 30$ . Les variations moyennes correspondantes sont  $v_t = s_t/2 = 42$  et  $v_e^2 = s_e/15 = 2$ . On remarque qu'effectivement, la variation moyenne résiduelle (estimation correcte de  $\sigma^2$ ) est égale à la variance (commune) correcte échantillon. La réalisation de  $F$  vaut  $f = v_t^2/v_e^2 = 21$ . La distribution de la variable  $F$  avec 2 et 15 degrés de liberté fournit la valeur  $f_{\alpha} = 3,68$  pour  $\alpha = 0,05$ , c'est-

(°) La minuscule  $c$  est la réalisation du terme correctif  $C = Y_{..}^2/n'$ ; de même pour  $s_T^2, s_t^2, \dots$ , réalisations de  $S_T^2, S_t^2, \dots$

à-dire, pour que  $P\{D_0|H_0\} = 0,95$ . Par conséquent, on est amené à prendre la décision  $D$  : rejeter  $H_0$  en faveur de  $H_1$ . Les tables de distribution de  $F$  ne permettent pas de calculer  $P\{F > 21\}$ . Mais on a, d'après les tables de la fonction bêta incomplète (Pearson, 1956),  $P\{F > 21\} = I_x(15/2, 1)$ , où  $x = 15/57 = 0,263157$  et par conséquent,  $P\{F > 21\} = 45 \cdot 10^{-6}$ .

Considérons un second exemple numérique étroitement lié au précédent : les données relatives au traitement C sont diminuées d'une quantité fixe égale à 2,84. Cette fois

$y_3 = 72,96$  (au lieu de 90) et  $\bar{y}_3 = 12,16$  (au lieu de 15) ; mais il est évident que  $s_3^2$  garde la même valeur 2. On calcule successivement  $c = 2199,17$ ,  $s_T = 44,02$ ,  $s_t = 14,02$ ,  $s_e = 30$ ,  $v_1^2 = 7,01$ ,  $v_e^2 = 2$  et finalement  $f = 3,5$ . Cette fois, la valeur observée pour  $F$  ( $f = 3,5$ ) est inférieure au seuil 5 % (3,68). On prend donc la décision  $D_0$ .

Certains auteurs ont mis en évidence (voir, par exemple, David & Johnson, 1951 et Hack, 1958) que le test- $F$  est « robuste » contre la non-normalité des populations étudiées. Autrement dit,  $P\{D_0|H_0\}$  n'est pas éloigné de 0,95 (lorsque  $f_x$  correspond à  $\alpha = 0,05$ ) dans certains cas de non-normalité même très accentuée ; d'autre part,  $P\{D|H_1\}$  n'est également pas fort influencé par la non-normalité. On a constaté également (voir, par exemple, Horsnell, 1953 et Box, 1953) que l'hétéroscélasticité des populations n'a d'importance pratique que dans le cas où les effectifs échantillons sont différents (ce problème est repris au § 1.3.).

Lorsque les effectifs échantillons sont différents, les tables 1.1 et 1.2 ne peuvent être utilisées et par conséquent, il n'est pas possible d'exploiter la condition (C). L'analyse statistique n'est modifiée qu'en ce qui concerne la définition de  $S_t$  ; on a

$$S_t = \sum_i n_i (\bar{Y}_i - \bar{Y}_{..})^2 = \sum \frac{Y_i^2}{n_i} - C$$

où  $n_i$  est l'effectif de l'échantillon prélevé dans la population  $\pi_i$ .

#### 1.2.1.2. Observations progressives.

Les hypothèses de travail et les notations sont celles introduites au § 1.2.1.1. Rappelons ici que la réponse de la  $j^e$  unité expérimentale au  $i^e$  traitement est  $Y_{ij} = \mu + \tau_i + X_{ij}$ , où  $\sum \tau_i = 0$  et où  $X_{ij}$  est une variable  $N(0, \sigma)$ . L'hypothèse d'homogénéité est  $H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0$ , les alternatives générales étant spécifiées par les valeurs du paramètre  $\lambda = \sqrt{\sum \tau_i^2 / k \sigma^2}$ .  $D_0$  est la décision de considérer  $H_0$  vrai ;  $D$  est la décision de considérer  $H_1$  vrai. Dans le cas de l'expérimentation non-progressive, on fixe l'effectif commun  $n$  de chaque échantillon a priori de telle sorte que si  $P\{D_0|H_0\} = 0,95$  on ait soit  $P\{D|H_1\} = 0,90$ , soit  $P\{D|H_1\} = 0,95$ .

L'expérimentation progressive consiste en l'observation successive des unités expérimentales. Dans le cas du test que nous exposons ici, le démarrage se fait en observant simultanément 3 unités si  $k = 3,5$  ou 7 et 4 unités si  $k = 4,6$  ou 8, dans chaque population. Après cette première étape, on prend soit la décision  $D_0$ , soit la décision  $D$ , soit la décision  $D_0$  de continuer l'expérimentation. Dans ce dernier cas, on fait deux observations simultanées dans chaque population ; c'est la deuxième étape qui se solde par  $D_0$ ,  $D$  ou  $D_0$ , etc. (Le fait de devoir faire deux observations à chaque étape est imposé seulement par la nature des tables numériques existantes ; rien n'empêche d'interpoler (graphiquement) dans les tables et de ne faire qu'une seule observation à chaque étape).

L'effectif commun échantillon, à chaque étape, est une variable aléatoire que nous notons  $N$  ; l'effectif total correspondant est  $N^* = kN$ .

A chaque étape, on calcule  $S_T^{(N)}$  et  $S_t^{(N)}$ , les variations totale et due aux traitements, correspondant à  $N$ . Le calcul de ces variations est en tout point analogue à celui exposé dans le cas où  $n$  est fixe (§1.2.1.1.). Ensuite, on calcule

$$G^{(N)} = S_t^{(N)} / (S_T^{(N)} - S_t^{(N)})$$

Soit  $g^{(n)}$  la réalisation de  $G^{(N)}$  sur l'essai, au moment où  $N = n$ . Remarquons que

$$g^{(n)} = \frac{k-1}{n^* - k} f,$$

$\lambda$	$n$	$k = 4$		$k = 6$		$n$	$k = 3$		$k = 5$		$k = 7$			
		$10^3 a$	$10^3 b$	$10^3 a$	$10^3 b$		$10^3 a$	$10^3 b$	$10^3 a$	$10^3 b$	$10^3 a$	$10^3 b$		
0,707	4	65	1825	168	1272	3	37	1319	72	4176	184	2407		
	6	110	770	183	646	5	72	696	142	927	211	784		
	8	126	521	183	464	7	89	497	155	568	207	512		
	10	131	411	180	377	9	99	400	158	432	199	401		
	12	134	350	176	327	11	105	344	159	360				
	14	136	310			13	108	306	159	317				
	16	138	382			15			157	287				
	1	4	266	1498	398	1237	3	124	4760	331	2469	469	1925	
		6	284	898	373	763	5	195	1174	340	973	420	888	
		8	287	637	354	601	7	221	762	331	687	387	645	
		10	286	540	340	521	9	234	605	322	565	365	552	
		12	284	482	330	473	11	240	522	314	498	349	494	
		14			322	340	13	244	471	308	456	338	456	
		16	280	424			15	247	436	303	428	328	430	
		1,414	4	616	1696	765	1535	3	458	3361	747	2445		
			6	596	1128	685	1081	5	497	1403	677	1272		
8			583	935	645	913	7	507	1040	637	994			
10			574	837	620	825	9	510	891	613	869			
12			567	777	602	770	11	512	809	596	799			
14			557	708	578	706	13	512	758	584	753			
16							15	511	724	575	720			

Table 1.4. — Nombres  $a$  et  $b$  tels que  $P\{D_0 | H_0\} = 0,95$  et  $P\{D | H_b\} = 0,95$ .

où  $f$  est la réalisation correspondante de  $F$ . On prend la décision

- $D_0$  si  $g^{(n)} \leq a$  ;
- $D$  si  $g^{(n)} \geq b$  ;
- $D_c$  si  $a < g^{(n)} < b$  ;

les nombres  $a$  et  $b$  faisant l'objet de la table 1.4. (Cette table est un abrégé des résultats numériques de Ray, 1956 ; la méthode elle-même vient de Johnson, 1953). Les seuils  $a$  et  $b$  sont tels que  $P\{D_c|H_0\} = 0,95$  et  $P\{D|H_0\} = 0,95$ .

L'effectif commun  $N$  des échantillons étant une variable aléatoire, on peut, théoriquement tout au moins, en déterminer la valeur moyenne. Des résultats numériques de Ray (1956), il résulte que la valeur moyenne de  $N$  est inférieure à l'effectif fixe correspondant, d'une quantité égale, grosso modo, à 35 % de cet effectif fixe. Par exemple, dans le cas où il y a trois traitements ( $k = 3$ ) et où  $P\{D_c|H_0\} = 0,95$  et  $P\{D|\lambda = 1\} = 0,95$ , il faut fixer a priori le nombre d'observations à 7 (voir table 1.2 du § 1.2.1.1.) ; dans les mêmes conditions, la valeur moyenne de  $N$  vaut approximativement 4.

Appliquons la méthode au premier exemple du § 1.2.1.1, en supposant que les observations (table 1.3) sont dans l'ordre de leur prélèvement successif. La première étape est le calcul de  $g^{(3)}$  pour les trois premières données du tableau 1.3 que nous avons reproduites ci-contre. (On démarre avec 3 observations, parce que  $k = 3$  est impair).

A	10	10	12	3	32
B	11	9	13	3	33
C	15	15	13	3	43
				9	108

$$\text{Il vient successivement } c^{(3)} = \frac{(108)^2}{9} = 1296 ;$$

$$s_T^{(3)} = (10)^2 + (10)^2 + \dots + (13)^2 - 1296 = 38 ;$$

$$s_t^{(3)} = \frac{1}{3} [(32)^2 + (33)^2 + (43)^2] - 1296 = 24,67 ;$$

et finalement,  $g^{(3)} = s_t^{(3)} / (s_T^{(3)} - s_t^{(3)}) = 1,850$ . Or, pour  $k = 3$  et  $n = 3$  (et  $\lambda = 1$  comme dans l'exemple original),  $a = 0,124$  et  $b = 4,760$ . On prend donc la décision  $D_c$  et deux nouvelles observations dans chaque population. Mais le prix des observations étant très élevé, il y a intérêt à ne prendre qu'une seule observation dans chaque population. Il faut alors déterminer les seuils  $a$  et  $b$  par interpolation graphique ; nous verrons d'ailleurs, que seule l'interpolation pour  $b$  est indispensable. La deuxième étape est donc le calcul de  $g^{(4)}$  pour les données ci-contre.

A	10	10	12	8	4	40
B	11	9	13	11	4	44
C	15	15	13	17	4	60
					12	144

Il vient successivement  $c^{(4)} = \frac{(144)^2}{12} = 1728$  ;

$$s_T^{(4)} = s_T^{(3)} + (8)^2 + (11)^2 + (17)^2 - 1728 = 80 ;$$

$$s_i^{(4)} = \frac{1}{4} [(40)^2 + (44)^2 + (60)^2] - 1728 = 56 ;$$

et finalement,  $g^{(4)} = s_i^{(4)}/S_T^{(4)} - S_i^{(4)} = 2,333$ . Il est évident qu'il suffit d'interpoler pour  $b$  ; on obtient  $b \cong 1.800$ . Et par conséquent, on prend la décision D. On arrive au même résultat qu'au § 1.2.1.1, mais avec 4 observations seulement, au lieu de 7. (En fait, il n'y en avait que 6, parce que nous avons  $P\{D|H_0\} = 0,90$  au lieu de 0,95).

### 1.2.2. Hypothèse d'homogénéité contre alternatives particulières.

L'énoncé du problème est identique à celui qui fait l'objet du § 1.2.1 à cela près que cette fois, l'objectif est de tester  $H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0$  contre les alternatives particulières  $H_p : \tau_1 \geq \tau_2 \geq \dots \geq \tau_k$  où au moins une égalité tombe en défaut.

Soient  $\bar{y}_i$ , ( $i = 1, 2, \dots, k$ ), les moyennes calculées des différents échantillons, la moyenne  $\bar{y}_i$  correspondant à  $\tau_i$ . On ordonne les moyennes  $\bar{y}_i$  par valeurs décroissantes (c'est-à-dire, dans l'ordre des  $\tau_i$  correspondants lorsque  $H_p$  est vrai). Chaque fois que deux moyennes consécutives  $\bar{y}_i, \bar{y}_{i+1}$  sont dans l'ordre inverse à celui de  $H_p$  (c'est-à-dire  $\bar{y}_i < \bar{y}_{i+1}$ ) les deux échantillons correspondants sont réunis en un seul échantillon. L'opération de réunion des échantillons contigus est réalisée jusqu'au moment où les moyennes échantillons sont dans l'ordre de  $H_p$ . Considérons un exemple (Bartholomew, 1959, I) dont les données sont reproduites au tableau ci-contre.

$k$	1	2	3	4	5
$y_i$	110	123	126	140	96
$n_i$	4	4	4	4	4
$\bar{y}_i$	27,5	30,75	31,5	35	24
				29,5	
				30,17	
			30,31		

L'hypothèse  $H_p$  exprime que les moyennes population sont dans l'ordre  $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4 \leq \mu_5$ . Les moyennes échantillons sont dans cet ordre, sauf en ce qui concerne  $\bar{y}_4$  et  $\bar{y}_5$ . Les échantillons (5) et (4) sont d'abord réunis ; l'échantillon global a pour moyenne 29,5 inférieure à la moyenne de l'échantillon (3). On réunit donc (3) à (5 + 4) ; ce nouvel échantillon a pour moyenne 30,17 inférieure à la moyenne de l'échantillon (2). On réunit donc (2) à (5 + 4 + 3). Cette fois, la moyenne du nouvel échantillon (30,31) est supérieure à la moyenne de l'échantillon (1). Par conséquent, les 5 échantillons d'origine sont ramenés à 2.

Soit  $l < k$ , le nombre d'échantillons qui subsistent après les opérations successives de réunion. Si les moyennes originales  $\bar{y}_i$  sont dans l'ordre de  $H_p$   $l = k$ . En réalité, le nombre d'échantillons qui subsistent après les opérations de réunion est une variable aléatoire : il dépend de l'ordre des variables aléatoires  $\bar{X}_i$ . Nous devons donc noter ce

nombre  $L$ . Soit d'ailleurs  $P_k\{L = l\}$  la probabilité pour que,  $H_0$  étant vraie, les  $k$  échantillons originaux se réduisent à un groupe de  $l$  échantillons. On a évidemment  $P_k\{L = 1\} = 1/k$  et  $P_k\{L = k\} = 1/k!$  On démontre que, (Miles, 1959),

$$P_k\{L = l\} = [P_{k-1}\{L = l-1\} + (k-1)P_{k-1}\{L = l\}]/k,$$

$$P_k\{L = l\} = \frac{1}{l} \sum_{j=l-1}^{k-1} P_{k-j}\{L = 1\} P_j\{L = l-1\}.$$

Nous empruntons à Miles (1959), un extrait d'une table fournissant les valeurs de  $P_k\{L = l\}$  pour différentes valeurs de  $l$  et de  $k$  (table 1.5).

$l$	$k = 2$	3	4	5	6	7	8	9
1	5000	3333	2500	2000	1667	1429	1250	1111
2	5000	5000	4383	4167	3805	3500	3241	3019
3		1667	2500	2917	3125	3222	3257	3255
4			417	833	1181	1458	1679	1854
5				83	208	347	486	619
6					14	42	80	125
7						2	7	15
8							0,2	0,9

Table 1.5.  $10^4 \cdot P_k\{L = l\}$ .

Supposons donc que les  $k$  échantillons d'origine soient ramenés à  $l$  (une réalisation de  $L$ ). Soit  $G$  la variable définie sur le groupe  $l$  exactement de la même manière que  $F$  l'a été au § 1.2.1.1. pour le groupe  $k$ . Il est évident que la distribution conditionnelle de  $G$  (la condition étant  $L = l$ ) est une distribution de Fisher-Snedecor avec  $(l-1)$  et  $(n'-1)$  degrés de liberté ( $n'$  étant l'effectif total de l'essai). Par conséquent,

$$P\{G \geq g\} = \sum_{l=2}^k P_k\{L = l\} \cdot P\{F(l-1, n'-l) \geq g\},$$

où  $g$  est un nombre quelconque, par exemple la réalisation de  $G$  sur les échantillons. Généralement, les tables de la distribution de Fisher-Snedecor ne donnent pas  $P\{F(l-1, n'-1) \geq g\}$  pour les valeurs de  $(l-1)$ ,  $(n'-1)$  et  $g$  observées dans la pratique. Mais en vertu d'une remarque faite au § 1.2.1.1 on peut également écrire

$$P\{G \geq g\} = \sum_{l=2}^k P_k\{L = l\} \cdot I_{x_1} \left( \frac{n'-l}{2}, \frac{l-1}{2} \right)$$

où  $x_1 = (n'-l)/[n'-l + (l-1)g]$ .

Reprenons dans le même ordre les deux exemples du § 1.2.1.1. Supposons tout d'abord que l'on teste  $H_0$  contre  $H_p: \mu_C > \mu_B > \mu_A$ . Dans ce cas, comme les moyennes échantillons sont dans l'ordre de  $H_p$ ,  $l = k = 3$  et la réalisation de  $G$  est  $g = f = 21$ . On calcule facilement, grâce aux tables de la fonction bêta incomplète (Pearson, 1956) et à la table 1.5, la probabilité pour que  $G$  soit égal à ou dépasse 21, soit

$$P\{G \geq g\} = P_3\{L = 2\} \cdot I_{x_2} \left( 8, \frac{1}{2} \right) + P_3\{L = 3\} \cdot I_{x_3} \left( \frac{15}{2}, 1 \right) = 162 \cdot 10^{-6}.$$

On prend donc la décision  $D$  en faveur de  $H_p$ .

Supposons ensuite, toujours pour le même exemple, que l'on veuille tester  $H_0$  contre  $H_p: \mu_C > \mu_A > \mu_B$ . Cette fois, les moyennes observées  $\bar{y}_A$  et  $\bar{y}_B$  sont dans l'ordre

opposé à celui de  $H_p$ . On réunit donc les échantillons correspondants et l'on constitue ainsi un nouvel échantillon (AB) d'effectif 12 et de moyenne 10,5 (inférieure à  $\bar{y}_C$ ). L'analyse de la variance, pour les deux échantillons (AB) et (C) fournit les résultats suivants :

$$c = (216)^2/18 = 2592 ; s_T = 114 ; s_t = 81 ;$$

$$s_g = 33 ; v_t^2 = s_t = 81 ; v_g^2 = s_g/16 = 2,0625 ;$$

et finalement,  $g = v_t^2/v_g^2 = 39,27$ . La probabilité pour  $G$  d'être égal à ou de dépasser  $g$  est

$$\begin{aligned} P\{G \geq g\} &= P_3\{L = 2\} \cdot P\{F(1, 16) \geq g\} + P_3\{L = 3\} \cdot P\{F(2, 15) \geq g\} \\ &= \frac{1}{2} I_{x_2}\left(8, \frac{1}{2}\right) + \frac{1}{6} I_{x_3}\left(\frac{15}{2}, 1\right) \end{aligned}$$

où  $x_2 = 0,29$  et  $x_3 = 0,16$ . Donc  $P\{G \geq g\} = 6 \cdot 10^{-6}$ . On prend la décision  $D$  en faveur de  $H_p$ . Si l'on confronte ce résultat avec le précédent, il faut conclure que le traitement  $C$  se distingue des deux autres eux-mêmes indistingables.

Considérons le second exemple du § 1.2.1.1 et testons  $H_0$  contre  $H_p: \mu_C > \mu_B > \mu_A$ . Les moyennes observées  $\bar{y}_A = 10$ ,  $\bar{y}_B = 11$  et  $\bar{y}_C = 12,16$  sont dans l'ordre de  $H_p$ . La valeur calculée de  $g$  est donc celle obtenue pour  $f$ , à savoir, 3,5. Et

$$P\{G \geq 3,5\} = \frac{1}{2} I_{x_2}\left(8, \frac{1}{2}\right) + \frac{1}{6} I_{x_3}\left(\frac{15}{2}, 1\right)$$

où  $x_2 = 0,8205$  et  $x_3 = 0,6818$ . Par conséquent,  $P\{G \geq 3,5\} = 0,0493$ . On est donc amené cette fois à prendre la décision  $D$ , c'est-à-dire, à rejeter  $H_0$  en faveur de  $H_p$ . Rappelons que le test-F de  $H_0$  contre  $H_p$  avait conduit à prendre la décision  $D_0$ . On voit que le test-G de  $H_0$  contre  $H_p$  rejette  $H_0$  (en faveur de  $H_p$ ). Le caractère particulier des alternatives rend le test-G plus « exigeant » en ce qui concerne  $H_0$ ; ceci illustre une remarque faite au § 1.1.a.

### 1.3. Populations normales hétéroscédastiques. Hypothèse d'homogénéité relative contre alternatives générales.

Lorsqu'on n'a aucune raison impérative de croire que les populations étudiées sont non-normales, mais que les variances populations sont vraisemblablement inégales et que les effectifs échantillons sont différents, on peut, pour tester  $H_0$  contre  $H_p$ , utiliser la méthode suivante, due à Welch (1951). Une méthode analogue due à James (1951), a également été proposée; nous n'en parlerons pas: les deux méthodes n'ont pas été comparées systématiquement.

L'hétérogénéité des variances populations peut être testée de différentes manières. L'une d'elles est le test  $\chi^2$  dû à Bartlett (voir, par exemple, Hald, 1955, p. 290). Box (1953), a montré que, dans l'ignorance où l'on se trouve généralement quant aux fonctions de distribution des populations étudiées, l'application préalable d'un test d'homogénéité des variances peut conduire à des mécomptes plus graves encore que ceux qui peuvent advenir si l'on s'en passe. D'après Box (1953), l'application d'un tel test, préalablement à l'analyse de la variance ou la méthode de Welch, « is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port! »

Dans le cas où les fonctions de distribution des populations étudiées peuvent être supposées normales ou approximativement telles, on adopte l'attitude suivante :

— si les effectifs échantillons sont égaux ou presque, on applique le test-F exposé au § 1.2.1.1 ;

— si les effectifs échantillons sont très inégaux et si les variances populations sont différentes, on applique la méthode de Welch (1951), exposée ci-dessous.

Soit  $\pi_i$  la population de toutes les réponses possibles au  $i^{\text{e}}$  traitement ( $i = 1, 2, \dots, k$ ). La fonction de distribution de  $\pi_i$  est  $N(\mu_i, \sigma_i)$ . Soit  $Y_{ij}$  la réponse de la  $j^{\text{e}}$  unité expérimentale ayant subi le  $i^{\text{e}}$  traitement. On suppose, comme au § 1.2.1.1, que  $Y_{ij} = \mu_i + X_{ij}$  où  $X_{ij}$  est une variable  $N(0, \sigma_i)$ ;  $\mu_i (= \mu + \tau_i)$  est donc la moyenne de la population  $\pi_i$ . L'hypothèse que l'on teste,  $H_0: \mu_1 = \mu_2 = \dots = \mu_k$  est une hypothèse d'homogénéité relative (voir § 1.1.d). Les alternatives sont générales. Les variances populations étant différentes il ne peut être question d'une estimation de la forme  $\sum_{ij} (y_{ij} - \bar{y}_{ij})^2$ .

Il faut estimer chaque variance population  $\sigma_i^2$  au moyen de la variance échantillon

$$S_i^2 = \frac{1}{n_i - 1} \sum_j (Y_{ij} - \bar{Y}_i)^2$$

Posons

$$W_i = n_i / S_i^2$$

au moyen de quoi on définit une moyenne pondérée (relativement à l'hétérogénéité des variances)

$$\bar{\bar{Y}} = \frac{\sum W_i \bar{Y}_i}{\sum W_i}$$

En suite de quoi on peut définir une variation totale (pondérée) due aux traitements

$$\bar{S}_t = \sum_i W_i (\bar{Y}_i - \bar{\bar{Y}})^2$$

On définit aussi

$$S_v = \sum_i \frac{1}{n_i - 1} \left( 1 - \frac{W_i}{\sum W_i} \right)^2$$

Nous verrons plus loin comment il convient de calculer les réalisations de  $S_t$  et  $S_v$  avec le maximum d'économie. On définit ensuite une variable

$$\bar{\bar{F}} = \frac{\frac{1}{k-1} \bar{S}_t}{1 + \frac{2(k-2)}{k^2-1} S_v}$$

dont Welch (1951) a montré qu'elle avait une distribution d'échantillonnage approximativement de Fisher-Snedecor, avec

$$f_1 = k - 1$$

$$f_2 = (k^2 - 1) / 3s_v$$

degrés de liberté ( $s_v$  est la réalisation de  $S_v$  sur l'essai).

Le calcul de  $\bar{\bar{F}}$  et  $s_v$  est relativement simple si l'on utilise les formules suivantes

$$\bar{S}_t = \sum w_i \bar{y}_i + \bar{y}^2 \sum w_i - 2 \bar{y} \sum w_i \bar{y}_i$$

$$s_v = \sum \frac{1}{n_i - 1} + \frac{1}{(\sum w_i)^2} \sum \frac{w_i^2}{n_i - 1} - \frac{2}{\sum w_i} \sum \frac{w_i}{n_i - 1}$$

dont chacun des termes s'obtient aisément si l'on aménage les calculs suivant les tableaux 1.6 et 1.7.

$Trts$ $i$	$y_{ij}$	$n_i$	$y_{i.}$	$\bar{y}_{i.}$	$\bar{y}_{i.}^2$	$1/(n_i - 1) \sum y_{ij}^2 - \bar{y}_{i.}^2$	$s_i^2$	$w_i$
1								
2								
⋮								
		$n'$	$y_{..}$			$\sum \frac{1}{n_i - 1}$		$\sum w_i$

Table 1.6. Données, et calcul de  $w_i$  et  $\sum w_i$ .

$Trts$ $i$	$w_i \bar{y}_{i.}$	$w_i \bar{y}_{i.}^2$	$w_i/(n_i - 1)$	$w_i^2$	$w_i^2/(n_i - 1)$
1					
2					
⋮					
	$\sum w_i \bar{y}_{i.}$	$\sum w_i \bar{y}_{i.}^2$	$\sum \frac{w_i}{n_i - 1}$		$\sum \frac{w_i^2}{n_i - 1}$

Table 1.7. Calcul de  $\bar{s}_i$  et  $s_v$ .

L'exemple suivant est emprunté à Welch (1951). Nous l'avons quelque peu modifié afin d'illustrer le fait suivant : Les variances échantillons étant très différentes le test-F n'est pas applicable ; si malgré tout, il est appliqué, il conduit à un résultat différent de celui auquel conduit le test- $\bar{F}$ . Il est d'ailleurs vraisemblable que la probabilité de prendre la décision  $D_0$  avec le test-F est inférieure à la probabilité de prendre la décision  $D_0$  avec le test- $\bar{F}$ , lorsque les moyennes populations sont égales.

L'exemple est relatif à trois populations  $\pi_1$ ,  $\pi_2$  et  $\pi_3$  dans lesquelles on a prélevés des échantillons d'effectifs 20, 10, 10, de moyennes respectives 27,845, 24,1, 22,2 et variances ( $s_i^2$ ) 60,1, 6,3, 15,4. Les données sont rassemblées au tableau 1.8.

$Trts$ $i$	$n_i$	$\bar{y}_{i.}$	$\bar{y}_{i.}^2$	$s_i^2$	$w_i$	$\frac{1}{n_i - 1}$
1	20	27,8454	775,3663	60,1	0,333	0,0526
2	10	24,1	580,810	6,3	1,587	0,1111
3	10	22,2	492,840	15,4	0,649	0,1111
	40				2,569	0,2748

Table 1.8. Données et calcul de  $w_i$ .

$T_{rts}$ $i$	$w_i \bar{y}_i$	$w_i^2 \bar{y}_i$	$\frac{w_i}{n_i - 1}$	$w_i^2$	$\frac{w_i^2}{n_i - 1}$
1	—	—	—	0,1109	—
2	—	—	—	2,5186	—
3	—	—	—	0,4212	—
	61,9267	1499,7956	0,2659		0,3324

Table 1.9. Calcul de  $\bar{s}_t$  et  $s_{T^2}$ .

Le test-F est tout d'abord réalisé. On obtient successivement  $c = 26.005,31$  ;  $s_t = 238,52$  ;  $s_e = 1337,20$  ;  $v_i^2 = 119,262$  ;  $v_e^2 = 36,140$  ; et finalement  $f = 3,3$ . Or, la valeur du seuil 5 % de la distribution de Fisher-Snedecor avec  $k - 1 = 2$  et  $n' - k = 37$  degrés de liberté est 3,255. Par conséquent, on prend la décision D en faveur de  $H_g$ .

Les variances échantillons étant très différentes, les effectifs échantillons étant très différents, les populations étant supposées normales, l'on se trouve dans les conditions d'application du test-F. Les calculs de  $w_i$ ,  $\bar{s}_t$  et  $s_e$  sont résumés aux tableaux 1.8 et 1.9 et complétés par

$$\bar{v} = 24,1054 ; \bar{y}^2 = 581,0689 ; (\Sigma w_i)^2 = 6,5998 ;$$

et finalement  $\bar{s}_t/2 = 3,513$  ; d'autre part,  $1 + \frac{6}{8} s_e = 1,029$ . Donc,  $\bar{f} = 3,35$ . Or, la valeur du seuil 5 % de la distribution de Fisher-Snedecor avec  $k - 1 = 2$  et  $(k^2 - 1)/3 s_e = 22,56$  degrés de liberté, est 3,43. Par conséquent, on prend la décision  $D_0$ .

#### 1.4. Populations non-normales.

##### 1.4.1. Hypothèse d'homogénéité contre alternatives générales.

Soient  $\pi_i$ , ( $i = 1, 2, \dots, k$ ),  $k$  populations dont les fonctions de distribution sont les fonctions  $F_i(x)$  dont les dérivées sont continues. On ne sait rien de la forme des fonctions  $F_i$  et en particulier, on ne peut pas supposer que ce sont des fonctions de distribution normales.

Soit  $F(x)$  une fonction de distribution dont la forme analytique est inconnue. Supposons que la moyenne de la population dont  $F(x)$  est la fonction de distribution soit  $\mu$  et que son écart-type soit  $\sigma$ . Les fonctions de distribution  $F_i$  ont toutes la même forme analytique que  $F$ , mais différent l'une de l'autre par la valeur d'un paramètre de position. On a

$$F_i(x) = F(x - v_i) ,$$

où  $v_i$  représente l'effet du  $i^e$  traitement. La valeur moyenne de la population dont  $F_i$  est la fonction de distribution est  $\mu + v_i$ . Donc, plus  $v_i$  est grand, plus la moyenne du  $i^e$  traitement est élevée. D'autre part, on suppose que l'effet d'un traitement ne se manifeste que sur la position de la fonction de distribution. Cela signifie, par exemple, qu'un traitement n'est en aucune manière caractérisé par une valeur de dispersion. Toutes les fonctions de distribution  $F_i$  ont le même écart-type  $\sigma$ .

Si  $v_1 = v_2 = \dots = v_k$ , les populations  $\pi_i$  sont identiques ; les traitements ne se distinguent pas l'un de l'autre. Par conséquent, l'égalité des  $v_i$  implique l'homogénéité des populations et constitue l'hypothèse nulle  $H_0$  qu'il faut tester. Nous désignons par  $D_0$  la décision statistique de considérer  $H_0$  comme l'hypothèse vraie. Les alternatives à  $H_0$

expriment qu'au moins un  $\nu_i$  se distingue des autres. Il y a donc  $(2^k - 1)$  alternatives. Parmi elles, considérons  $H_k: \nu_1 = \nu_2 = \dots = \nu_{k-1} = 0 = \nu_k - \Delta$  où  $\Delta$  est un nombre positif. Nous supposons que

$$\Delta = (\mu + 1)/\sigma \sqrt{m},$$

où  $m \geq 1$ . L'hypothèse  $H_k$  exprime que les moyennes des traitements sont telles que

$$\mu_1 = \mu_2 = \dots = \mu_{k-1} = \mu,$$

$$\mu_k = \mu + \Delta,$$

$m$  étant un paramètre variable auquel l'expérimentateur peut donner n'importe quelle valeur supérieure à un. Nous désignons par  $D_k$  la décision statistique de considérer  $H_k$  comme l'hypothèse vraie.

Soit  $X_{i1}, X_{i2}, \dots, X_{in_i}$  l'échantillon aléatoire prélevé dans  $\pi_i$ .

Suivant une notation, reprise au § 3.3, soit  $\langle X_1 - X_2 - \dots - X_k \rangle$  la suite ordonnée par valeurs croissantes de toutes les variables observées.

Chaque variable  $X_{ij}$ , ( $j = 1, 2, \dots, n_i$ ), a dans la suite ordonnée une certaine place, un certain rang  $R'_{ij}$ . Soit  $R_i = \sum_j R'_{ij}$  la somme des rangs des variables de l'échantillon prélevé dans  $\pi_i$ . Posons également  $n' = \sum_i n_i$ , l'effectif total de l'expérience.

Considérons la variable aléatoire

$n_1$	$n_2$	$n_3$	$h_z$	$10^3 P \{ H > h_z \}$	$n_1$	$n_2$	$n_3$	$h$	$10^3 P \{ H > h_z \}$
2	2	2	457	67	5	2	1	500	48
3	2	2	471	48	5	2	2	504	56
3	3	1	514	43	5	3	1	487	52
3	3	2	514	61	5	3	2	525	49
3	3	3	560	50	5	3	3	534	50
4	2	1	482	57	5	4	1	486	56
4	2	2	512	52	5	4	2	527	50
4	3	1	521	50	5	4	3	563	50
4	3	2	540	52	5	4	4	562	50
4	3	3	573	50	5	5	1	491	53
4	4	1	496	48	5	5	2	525	51
4	4	2	524	52	5	5	3	563	51
4	4	3	558	50	5	5	4	564	50
4	4	4	569	49	5	5	5	566	51

Table 1.10. — Valeurs de  $h_z$  telles que  $P \{ H < h_z \} \cong 0,95$ .

$$H = \frac{12}{n'(n' + 1)} \sum_1^k \frac{R_i^2}{n_i} - 3(n' + 1).$$

La variable  $H$  a été proposée par Wallis et Kruskal, (1952), en remplacement de la variable  $F$  de Fisher-Snedecor utilisable seulement dans le cas de distributions normales homoscédastiques. Les propriétés du test de  $H_0$  basé sur  $H$  ont été étudiées par Andrews, (1954). La distribution d'échantillonnage, lors  $H_0$  est vrai, a été ajustée récemment par Wallace, (1959).

On prend la décision  $D_0$  si

$$H < h_z$$

où  $h_\alpha$  est un nombre tel que  $P\{D_0 | H_0\} = 1 - \alpha$ . Lorsque  $k = 3$  et  $n_i \leq 5$ , les valeurs de  $h_\alpha$  telles que  $P\{D_0 | H_0\}$  soit voisin de 0,95 sont fournies à la table 1.10, empruntée à Wallis et Kruskal (1952). Lorsque  $k = 3$  et  $n_i > 5$ , ou bien, lorsque  $k > 3$ , on a

$$P\{D_0 | H_0\} = P\{\chi_{k-1}^2 < h_\alpha\}$$

et dans ces conditions,  $h_\alpha$  est le seuil  $\alpha$  % de la distribution  $\chi_{k-1}^2$  (distribution  $\chi^2$  avec  $(k-1)$  degrés de liberté). Par exemple, si  $\alpha = 0,05$  et  $k = 7$ ,  $h_\alpha = 12,6$ .

Supposons que l'on ait des raisons de craindre que la population  $\pi_k$  se distingue seule des autres. Dans ce cas, l'alternative à  $H_0$  dont il convient de tenir compte est  $H_k$ . Le test doit être tel que la probabilité de prendre la décision  $D_k$  si  $H_k$  est vraie, soit contrôlable et fixée à un niveau élevé. Nous avons considéré ce problème dans le cas où tous les effectifs échantillons sont égaux à  $n$ . Il convient de fixer  $n$  à l'avance de telle sorte que  $P\{D_k | H_k\} \geq 0,90$ . On démontre qu'il faut que

$$n > \frac{\lambda k}{\Delta^2 (k-1)},$$

où  $\lambda$  dépend de  $k$  et est fourni à la table 1.11.

$k$	3	4	5	6	7
$\lambda$	13	14	15	16	17

Table 1.11. — Valeurs de  $\lambda$  correspondant à différentes valeurs de  $k$  et telles que  $P\{D_k | H_k\} \geq 0,90$ , lorsque  $\alpha = 0,095$ .

Dans l'exemple qui suit, le calcul de  $H$  n'est pas gêné par la présence d'observations égales dans la suite ordonnée  $\langle X_1 - X_2 - \dots - X_k \rangle$ . Si cette circonstance survient, on remplace le rang de chaque observation double (ou triple, ou quadruple, etc.) par la moyenne de leurs rangs.

	1 <sup>er</sup> trait.		2 <sup>e</sup> trait.		3 <sup>e</sup> trait.		4 <sup>e</sup> trait.		
	obs.	rangs	obs.	rangs	obs.	rangs	obs.	rangs	
	-2	1	-1,95	2	-1,9	3	1	13	
	-0,8	4	-0,7	5	-0,5	6	1,5	14	
	0	7	0,1	8	0,2	9	2,2	18	
	0,3	10	0,4	11	0,5	12	2,5	19	
	2,0	15	2,1	16	2,15	17	3	20	
$R_i$		37		42		47		84	210
$R_i^2$		1369		1764		2209		7056	
$R_i^2/n$		274		353		442		1411	2480

Supposons qu'un expérimentateur doive comparer quatre traitements. Il ne peut supposer que les populations des réponses sont normales ; mais l'expérience lui enseigne que les populations ne diffèrent que par la valeur d'un paramètre central, la moyenne ou la médiane. L'un des traitements, le 4<sup>e</sup> par exemple, est suspecté d'avoir une valeur contrale supérieure à celle des trois autres ; il convient que  $P\{D_4 | \Delta = 2\} > 0,90$ . Pour satisfaire à cette condition il faut prendre ( $\mu$  étant voisin de zéro et  $\sigma$  voisin de 1)

$$\frac{\lambda k}{4(k-1)} = \frac{14.4}{4.3} \cong 5$$

observations dans chaque population. Si les quatre traitements sont identiques, il convient que  $P\{D_0 | H_0\} = 0,95$ . Par conséquent, on prend la décision  $D_0$  si  $h < 7,81$ . Les données sont reproduites au tableau ci-contre.

La somme des rangs (210) est indiquée au tableau comme vérification. Il faut en

$$\Sigma R_i = \frac{n^3(n+1)}{2} \left( = \frac{20 \times 21}{2} \right)$$

On indique aussi  $\Sigma R_i^2/n = 2480$  et on calcule ensuite la valeur observée de  $H$ , soit

$$h = \frac{12.2480}{20.21} - 3.21 \cong 8$$

On est donc amené à ne pas prendre la décision  $D_0$ . Et comme le 4<sup>e</sup> échantillon présente l'anomalie à laquelle on s'attendait de sa part, on prend la décision  $D_4$  ; on a  $P\{D_4 | H_4\} > 0,90$ .

#### 1.4.2. Hypothèse d'homogénéité contre alternatives particulières.

Soient  $\pi_i$ , ( $i = 1, 2, \dots, k$ ),  $k$  populations dont les fonctions de distribution  $F_i(x)$  inconnues sont non-normales, mais admettent des dérivées continues. On désire tester l'hypothèse d'homogénéité  $H_0 : F_1 = F_2 = \dots = F_k$  contre les alternatives particulières  $H_p : F_1 > F_2 > \dots > F_k$ . Le test que nous exposons est une généralisation, proposée par Jonckheere, (1954), du test basé sur le coefficient des rangs de Kendall.

Il importe de bien comprendre le sens des alternatives  $H_p$ . Considérons deux populations seulement  $\pi_1$  et  $\pi_2$  dont les fonctions de distribution sont  $F_1$  et  $F_2$ . Soit également  $X_1$  la variable aléatoire dont les valeurs possibles constituent la population  $\pi_1$  et soit  $X_2$  la variable aléatoire dont les valeurs possibles constituent la population  $\pi_2$ . Lorsque  $F_1 = F_2$ , la probabilité pour que  $X_1$  soit supérieure (ou inférieure) à  $X_2$  vaut exactement  $\frac{1}{2}$ . Par contre, si  $F_1 > F_2$ ,  $P\{X_1 > X_2\} < \frac{1}{2}$  et si  $F_1 < F_2$ ,  $P\{X_1 > X_2\} > \frac{1}{2}$ . Si  $F_1 > F_2$ , on a aussi que  $v_1 < v_2$ ,  $v_1$  et  $v_2$  étant respectivement les médianes des populations  $\pi_1$  et  $\pi_2$  ; si  $F_1 < F_2$ ,  $v_1 > v_2$ .

Généralisant ce qui précède, appelons  $X_i$  la variable aléatoire dont les valeurs possibles constituent la population  $\pi_i$ . Et soit  $N_{ij}$  la variable aléatoire représentant le nombre de fois que  $X_i < X_j$ . La variable aléatoire sur laquelle le test de  $H_0$  est basé est

$$S = 2 \sum_{i < j} N_{ij} - \sum_{i < j} n_i n_j,$$

où  $n_i$  et  $n_j$  sont respectivement les effectifs des échantillons prélevés dans  $\pi_i$  et  $\pi_j$ . Considérons un exemple simple qui permet d'illustrer le calcul d'une réalisation  $s$  de  $S$ . Nous avons reproduit au tableau ci-contre les observations prélevées dans trois populations  $\pi_1$ ,  $\pi_2$  et  $\pi_3$  ; nous les avons ordonnées par valeurs croissantes.

$\pi_1$	$\pi_2$	$\pi_3$
1	3	5
4	6	8
9	11	13

Dans cet exemple,  $n_1 = n_2 = n_3 = 3$ . La deuxième somme intervenant au second membre de S vaut donc

$$\sum_{i < j} n_i n_j = n_1 n_2 + n_1 n_3 + n_2 n_3 = 27 .$$

De même, la première somme vaut

$$2 \sum_{i < j} n_{ij} = 2 (n_{12} + n_{13} + n_{23}) ,$$

où  $n_{12}$  est le nombre de fois qu'une observation de  $\pi_1$  est inférieure à une observation de  $\pi_2$ ,  $n_{13}$  et  $n_{23}$  étant définis de la même manière. Par conséquent,  $n_{12} = 3 + 2 + 1 = 6$ ,  $n_{13} = 3 + 3 + 1 = 7$  et  $n_{23} = 3 + 2 + 1 = 6$ . Donc  $2 \sum_{i < j} n_{ij} = 38$  et  $s = 11$ .

Lorsque  $H_0$  est vrai, la variable  $X_i$  n'a pas plus de chance de dépasser la variable  $X_j$  que de lui être inférieure. Par conséquent, la valeur moyenne de  $N_{ij}$ , lorsque  $H_0$  est vrai, est  $n_i n_j / 2$  et la valeur moyenne de S vaut zéro. Par contre, lorsque  $H_p$  est vrai,  $N_{ij}$  a tendance à prendre des valeurs supérieures à  $n_i n_j / 2$  et S a donc tendance à prendre des valeurs positives. La valeur maximum de S est  $\sum_{i < j} n_i n_j$ . Les grandes valeurs de S sont

donc significatives d'un écart par rapport à  $H_0$  en faveur de  $H_p$ .

Posons

$$\sigma^2 = \frac{1}{18} [n^2 (2n' + 3) - \sum n_i^2 (2n_i + 3)] ,$$

où  $n' = \sum n_i$ ; si tous les effectifs échantillons sont égaux à  $n$ , on a  $n' = kn$  et

$$\sigma^2 = \frac{1}{18} k(k-1)n^2 [2n(k+1) + 3] .$$

Soit  $D_0$  la décision statistique de considérer  $H_0$  comme vraie et soit  $D_p$  la décision statistique de considérer  $H_p$  comme vraie. La règle que nous donnons ci-dessous est telle que  $P\{D_0 | H_0\} \geq 0,95$ . Les nombres du tableau 1.12 permettent de l'appliquer dans le cas où tous les effectifs échantillons sont égaux ( $\leq 10$ ) et où  $k \leq 6$ . On prend la décision  $D_0$  si  $s$  est inférieur au nombre de la table 1.12 correspondant à  $n$  et à  $k$ ; on prend la décision  $D_p$  si  $s$  lui est supérieur ou égal. Pour les configurations d'effectifs qui ne sont pas prévues à la table 1.12, on prend la décision  $D_0$  si  $s < 1,64 \sigma$  et la décision  $D_p$  si  $s \geq 1,64 \sigma$ , où  $\sigma$  est la racine carrée positive de  $\sigma^2$ .

$n \backslash k$	3	4	5	6
2	10°	14°	20°	26°
3	17°	26°	34°	43
4	24°	38°	49	65
5	33°	49	69	90
6	41	64	90	118
7	51	80	113	149
8	62	97	138	173
9	74	116	164	216
10	86	136	192	253

Table 1.12. Seuils de signification 5% pour le test - S. Les nombres marqués de (°) sont empruntés à Jonckheere (1954); les autres sont obtenus grâce à l'approximation normale indiquée par cet auteur qui fait en outre remarquer que l'approximation est pratiquement très bonne même dans les cas extrêmes où les effectifs sont très différents.

Appliquons la méthode à un exemple emprunté à Jonckheere (1954) ; les données sont rassemblées au tableau 1.13. Les quatre populations

	$\pi_1$	$\pi_2$	$\pi_3$	$\pi_4$
	19	21	40	49
	20	61	99	110
	60	80	100	151
	130	129	149	160
$x_{i.}$	229	291	388	470
$\bar{x}_{i.}$	57,25	72,75	97	117,50
$s_i^2$	2037,69	1508,19	1491,5	1919,25

Table 1.13. Données de l'exemple.

$\pi_1$ ,  $\pi_2$ ,  $\pi_3$  et  $\pi_4$  ont des fonctions de distribution non-normales. Les échantillons ont tous le même effectif  $n = 4$ . Les variances échantillons sont fort différentes. Nous sommes donc dans les conditions d'application des méthodes indépendantes de l'hypothèse de normalité : soit le test - H (§ 1.4.1), soit le test - S. Le choix entre les deux tests s'établit sur la base des alternatives à l'hypothèse d'homogénéité : si l'on teste  $H_0$  contre  $H_g$ , on utilise le test - H ; si l'on teste  $H_0$  contre  $H_p$ , on utilise le test - S. A titre d'illustration, nous appliquons aux données de la table 1.13 successivement le test - F (seulement valable si l'on suppose la normalité des populations), le test - H et le test - S.

Le test - F (voir § 1.2.1.1.) conduit à la valeur observée  $f = 1,216$  ; or, le seuil 5 % de la distribution de Fisher-Snedecor avec 3 et 12 degrés de liberté est 3,49. On prend donc la décision  $D_0$  (contre la décision  $D_g$ ).

Le test - H (voir § 1.4.1) conduit à la valeur observée  $h = 3,73$  ; or, le seuil 5 % de la distribution de Fisher-Snedecor avec 3 et 12 degrés de liberté est 3,49. On prend donc la décision  $D_0$  (contre la décision  $D_g$ ).

Les réalisations de la variable aléatoire  $N_{ij}$  sont  $n_{12} = 11$ ,  $n_{13} = 12$ ,  $n_{14} = 13$ ,  $n_{23} = 11$ ,  $n_{24} = 12$ , et  $n_{34} = 12$  ; donc  $s = 46$ . Or, le seuil 5 % correspondant à  $n = 4$  et  $k = 4$  (voir table 1.12) est 38. On prend donc la décision  $D_p$  (contre la décision  $D_0$ ). On rejette l'hypothèse d'homogénéité en faveur de l'alternative  $H_p$  :  $F_1 > F_2 > F_3 > F_4$ . Autrement dit, on rejette l'hypothèse d'égalité des médianes contre l'alternative  $v_1 < v_2 < v_3 < v_4$  ; mais il est bien entendu que le test - S est relatif à l'hypothèse  $H_0$  d'homogénéité totale ; la formulation des conclusions en termes de médianes n'est qu'un aspect de la question : si l'on avait pris la décision  $D_0$ , cela aurait signifié que l'on acceptait l'hypothèse que les quatre populations sont en tout point identiques (et pas seulement de médianes égales).

Comme l'exemple 1.2.2., celui-ci illustre la même remarque faite au § 1.1.a sur la plus grande efficacité d'un test de  $H_0$  contre  $H_p$ .

## REFERENCES BIBLIOGRAPHIQUES POUR LE CHAPITRE 1

- ANDREWS, C. (1954). — *Asymptotic behavior of some rank tests for analysis of variance*. A.M.S., pp. 724-735.
- BARTHOLOMEW, D.J. (1959). — *A test of homogeneity of ordered alternatives*. Bka, I, pp. 36-48, II, pp. 328-335.

- BOX, G.E.P., (1953). — *Non-normality and tests on variances*. Bka, pp. 318-335.
- DAVID, F.N. & JOHNSON, N.L. (1951). — *The effect of non-normality on the Power function of the F-test in the analysis of variance*. Bka, pp. 43-57.
- DE MUNTER, P. (1954). — *Sur les transformations de variables aléatoires*. « Bull. Soc. Belge de St. », pp. 99-111.
- DE MUNTER, P. (1960). — *Sur la notion de traitement et le problème de la comparaison de plus de deux traitements*. « Bull. Inst. Agr. Gembloux » (centenaire).
- FELDT, L.S. & MAHMOUD, M.W. (1958). — *Power function charts for specifying numbers of observations in analysis of variance of fixed effects*. A.M.S., pp. 871-877.
- HACK, H.R.S. (1958). — *An empirical investigation into the distribution of the F-ratio in samples from two non-normal populations*. Bka, pp. 260-265.
- HALD, A. (1955). — *Statistical Theory with Engineering Applications*. J. Wiley, N.Y.
- HORSNELL, G. (1953). — *The effect of unequal group variances on the F. test for the homogeneity of group means*. Bka, pp. 128-136.
- JAMES, G.S. (1951). — *The comparison of several groups of observations when the ratios of the population variances are unknown*. Bka, pp. 324-329.
- JOHNSON, N.L. (1953). — *Some notes on the application of sequential methods in the analysis of variance*. A.M.S., pp. 614-623.
- JONCKHEERE, A.R. (1954). — *A distribution-free k-sample test against ordered alternatives*. Bka, pp. 133-145.
- KRUSKAL, W.H. & WALLIS, W.A. — *Use of ranks in one-interion variance analysis*. 1952 : J.A.S.A., pp. 583-621 ; 1953 : J.A.S.A., pp. 907-911.
- MILES, R.E. (1959). *The complete amalgamation into blocks, by weighted means, of a finite set of real numbers*. Bka, pp. 317-327.
- PEARSON, E.S. & HARTLEY, H.O. (1951). — *Charts of the power function of analysis of variance tests, derived from the non-central F-distribution*. Bka, pp. 112-130.
- PEARSON, K. (1956). — *Tables of the Incomplete Beto-Function*. University Press, Cambridge.
- RAY, W.D. (1956). — *Sequential analysis applied to certain experimental designs in the analysis of variance*. Bka, pp. 388-399.
- WALLACE, D.L. (1959). — *Simplified beta-approximations to the Kruskal-Wallis H-test*, J.A.S.A., pp. 225-230.
- WELCH, B.L. (1951). — *On the comparison of several mean values : an alternative approach*. Bka, pp. 330-336.
-