

Introduction

Trois points de vue sur la comparaison.

Comparer des données c'est calculer la distance qui les sépare. Si pour comparaison de deux termes est assez intuitive, il n'en va pas de même lorsqu'il s'agit de comparer plus de deux termes. Dans le cours sur l'inférence statistique vous avez vu avec le t de Student la comparaison de deux termes sur des variables numériques. L'analyse de variance est le prolongement de ce type d'analyse en généralisant l'idée de comparaison à plus de deux termes.

Comparaison sur deux termes.

Pour comparer deux termes, il est assez intuitif d'en calculer la différence par soustraction. Si on a plus de deux termes, on fait les comparaisons deux à deux et on les ordonne.

Point de vue de l'hypothèse nulle.

Une autre façon de le voir est de regarder la moyenne comme une répartition des points sur les différents individus, c'est-à-dire la note qu'il obtiendrait, si sans changer le total des points, il avait obtenu la même note. La somme des écarts à la moyenne est alors égale à la distance entre les deux individus. On peut généraliser cette idée à plus de deux individus, ce qu'on calcule alors est la variance.

Point de vue des contrastes.

Dans le prolongement de ce que nous venons de dire sur l'hypothèse nulle, on peut également attribuer un poids à chacun des individus pour calculer les différences. On calcule alors un contraste. Nous verrons dans le paragraphe consacré à cette la méthode des contrastes que la variance est une somme de contrastes, c'est-à-dire une somme de comparaisons binaires.

Point de vue vectoriel.

De ce point de vue, la comparaison revient à situer le protocole par rapport aux axes définis par les contrastes. Cet espace a autant de dimensions que le protocole a de sujets. La somme des carrés associées à une comparaison est alors le carré de la distance entre le point d'origine (moyenne du protocole) et chaque individu est situé par rapport à sa distance à cette origine.

Exercices

Le point de vue de l'hypothèse nulle.

On vous a appris dès vos premiers cours sur les méthodes statistiques qu'on pouvait voir la moyenne comme une répartition des valeurs des observations sur tous les sujets.

Devinette: Si la note de français dans une classe de collège est la même pour tous les élèves sans que cela modifie la moyenne, quelle serait cette note?

Réponse : Cette note serait la moyenne. Comparer les notes des élèves, c'est se demander si, en moyenne, les écarts à la moyenne sont supérieurs ou non à 0. Pour cela vous savez, depuis la première année, qu'on calcule la variance et l'écart-type. Si les élèves sont très proches, l'écart-type est très petit, voir nul tous les élèves ont pratiquement la même note. Si les différences inter-sujets sont importantes, cet écart-type est important.

Cette idée peut se généraliser aux groupes de sujets.

Devinette : Si les moyennes des notes de français de trois classes de collèges sont les mêmes, quelles seraient les moyennes de ces trois classes ?

Réponse : Elles seraient égales à la moyenne générale des trois classes.

Comparer des groupes, c'est se demander si, en moyenne, les écarts à la moyenne générale sont supérieurs ou non à 0. Comme pour les élèves, on va calculer la variance, en prenant comme notes pour les classes les moyennes de ces classes.

On peut également généraliser cette idée à la comparaison de tâches.

Devinette : Si les moyennes des notes à trois contrôles de français dans une classe de collège sont les mêmes, quelles seraient les moyennes à ces contrôles.

Réponse : Elle serait bien sûr égale à la moyenne générale des notes à ces trois contrôles.

Comparer des tâches, c'est se demander si, en moyenne, les écarts à la moyenne générale sont supérieurs ou non à 0. Il faut donc calculer la variance des tâches en prenant comme notes les moyennes dans chacune des tâches.

Le point de vue des contrastes.

La moyenne de deux observations a l'intéressante propriété d'être à égale distance des deux observations. Cette propriété permet de comparer ces deux observations. Reprenons notre exemple de collégiens.

Question : Lorsque vous voulez comparer deux élèves, dans une matière comme le Français, comment vous y prenez-vous ?

1^{ère} solution : vous faites la différence de deux notes. Admettons que le premier à 12 et le second 14. La comparaison des deux élèves nous montrera donc que 2 points les séparent en faveur du deuxième élève.

2^{ème} solution : vous les comparez en calculant la somme des écarts à la moyenne (en valeur absolue, sinon la somme est égale à 0). Dans notre exemple, la moyenne des deux élèves est de 13. La somme des écarts (en valeur absolue) est de 2.

Faire la différence entre deux individus ou calculer la somme des écarts à la moyenne de ces deux individus, c'est la même chose. On peut généraliser cette idée à plus de deux individus, il suffit de faire la somme de toutes les comparaisons binaires possibles. C'est ce qui sera abordé dans la méthode des contrastes. Cette idée est également généralisable à la notion de groupes et de tâches. Dans ce cas, les moyennes des groupes ou des tâches servent de notes à comparer (comme plus haut avec le point de vue de l'hypothèse nulle).

Le point de vue vectoriel.

La moyenne est également le point d'équilibre d'une distribution. Pour bien comprendre cela, imaginons que des fillettes jouent à l'élastique, puis qu'éclate une dispute concernant la possession de l'élastique. Chacune se met alors à tirer sur l'élastique.

Question : Qui va remporter la bataille, autrement dit qui tire le plus fort ?

Réponse : La question n'a de sens que si le système est en équilibre et donc ne bouge pas, sinon l'ensemble fait mouvement vers la fillette qui tire le plus fort et la question ne se pose plus. Dans un tel cas de figure, le système est en équilibre autour de la moyenne des forces exercées par chacune des fillettes. Si on représentait tout cela par des vecteurs, la fillette qui tire le plus fort est celle dont le vecteur est le plus long, autrement dit celle qui s'écarte le plus de la moyenne. Tout cela est, bien entendu, généralisable à l'échelle des groupes (plusieurs fillettes tirent dans la même direction) et à plusieurs tâches (elles tirent sur plusieurs élastiques).

De tous ces points de vue, vous devez retenir que comparer des observations, des groupes d'observations (groupes de sujets ou tâches) revient finalement à évaluer la dispersion autour de la moyenne de ces observations autrement dit à calculer la variance de ces observations.

L'inférence sur la variance : Le F de Snédécor

Rappel des principes de l'inférence.

En première année, vous avez vu comment situer l'observation d'un individu dans une distribution. On calcule, pour ce sujet, sa note z qui est l'écart à la moyenne pondérée par l'écart-type. Il s'agit dans ce cas de pondérer l'écart du sujet à la moyenne par la variation intersujet dans la distribution. On peut ensuite regarder dans la table du z la proportion de sujets plus extrêmes que le sujet qu'on cherche à situer.

En seconde année, ce principe a été étendu à l'inférence sur des échantillons. Vous avez vu qu'on pouvait situer un échantillon dans une distribution d'échantillonnage approchée en pondérant sa moyenne par l'écart-type corrigé : c'est le test du t de Student. Comme pour le z , il s'agit toujours de pondérer un écart par la variation intersujet. Dans ce cas, la moyenne de l'échantillon est comparée à la moyenne théorique (hypothèse nulle). Pour la comparaison de deux groupes d'observations (appariés ou indépendant), c'est la différence des moyennes qui est comparée à la moyenne théorique.

Relation entre t de Student et F de Snédécor

Cette année, c'est le même principe qui est étendu à la comparaison de plus de deux groupes. Une simple différence ne permettant de comparer que deux termes, pour plus de deux groupes d'observations, on retient, comme indice d'écart entre les modalités du facteur, la variance intermodalités. La variance étant un écart au carré, ce n'est plus par l'écart-type intersujets qu'il faut la pondérer, mais par la variance intersujets¹. Ce rapport entre la variance corrigée intermodalités (ou carré moyen inter) et la variance corrigée intersujets (carré moyen intersujets ou intragroupe) est ce qu'on appelle le F de Snédécor.

$$F = \frac{CM_{inter}}{CM_{intra}}$$

La conséquence directe de ce que nous venons de dire est que dans le cas d'une comparaison à un degré de liberté au numérateur (comparaison de deux groupes) le F de Snédécor est égal au carré du t de Student.

Condition de validité du test.

Les conditions de validité du test sont au nombre de trois :

- Dans le cas d'une structure d'emboîtement, on doit vérifier l'indépendance des observations. Dans le contexte expérimental, cette condition est en fait toujours respectée si les sujets sont répartis aléatoirement dans les groupes et si on a bien un emboîtement strict des sujets dans les groupes.
- Dans tous les cas, on doit vérifier la normalité des distributions. La distribution du F se distribue normalement. Pour servir de distribution d'échantillonnage approchée, cela suppose que la variable dépendante se distribue à peu près normalement ou en tout cas ne soit pas trop asymétrique.
- La dernière condition est celle de l'homogénéité des variances intragroupes ou intersujets, surtout dans le cas de groupes non équilibrés. Dans le cas où les variances seraient trop différentes, on préférera utiliser les distributions exactes (voir le cours "inférence statistique").

¹ La variance intersujets est encore appelée variance intragroupes ou variance intra dans le cas d'une structure d'emboîtement. En effet, dans ce cas, les sujets sont indissociables du groupe auquel ils appartiennent.

Notions de Plan et de protocole.

Notion de structure de protocole

Un plan est un ensemble de facteurs organisés par des relations de croisement et/ou d'emboîtement. Cette organisation des facteurs est appelée structure du protocole. Le plan est désigné par une écriture formelle exprimant les différents facteurs et le nombre de modalités de chaque facteur. Chaque facteur est désigné par une lettre et un indice indiquant le nombre de ses modalités.

Le nombre d'observations est égal au produit des indices des facteurs du plan.

Dans tous plans, il y a au moins un facteur aléatoire : le facteur sujet. On le note \underline{S} . Si les groupes sont équilibrés, c'est-à-dire contiennent le même nombre d'individus, l'indice de ce facteur est égal au nombre total de sujets divisé par le produit des indices des facteurs emboîtants.

On parle de croisement lorsque tous les individus (facteur \underline{S}) sont associés à toutes les modalités du facteur (facteur T). On note cette relation : \underline{S}^*T (lire S croix T). L'indice du facteur croisant est égal au nombre de ses modalités.

On parle de relation d'emboîtement lorsque les individus (facteur \underline{S}) ne sont associés qu'à une des modalités du facteur (facteur G). On note cette relation $\underline{S}<G>$ (lire S dans G). L'indice de ce facteur est égal au nombre de ses modalités.

Tous les plans quasi-complets qu'on peut écrire sont des combinaisons de ces deux relations.

Plan de base et plan d'analyse

Le plan de base encore appelé structure du protocole est le plan correspondant au recueil de données. Le plan d'analyse est le plan correspondant à l'intention d'analyse. Il dérive du plan de base :

- Par recodage de facteurs.
- Restriction sur les facteurs.
- Restriction sur les individus.

Notion de degrés de liberté

Les degrés de liberté correspondent au nombre de comparaisons binaires qu'on peut faire sur une partition d'un protocole.

Règles d'écriture des plans

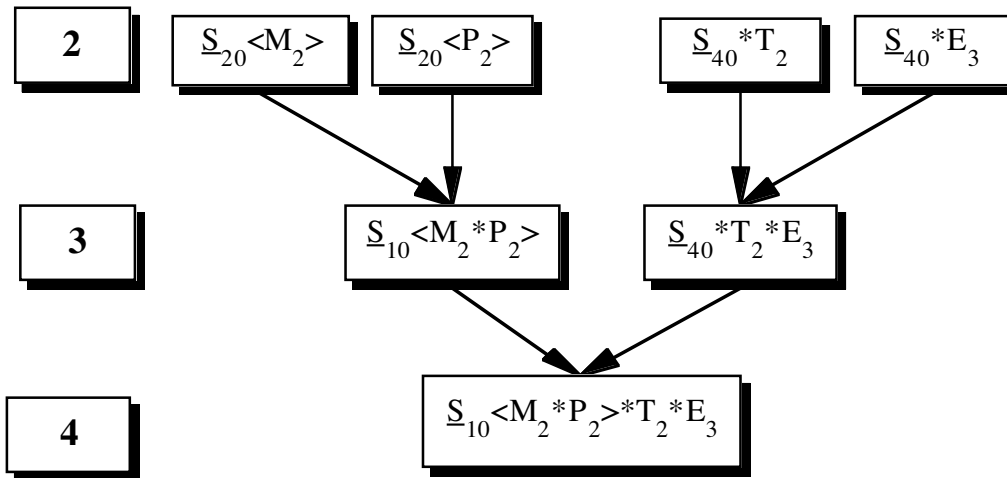
Pour écrire un plan :

- On identifie d'abord l'ensemble des facteurs et des variables dépendantes. N'entrent dans l'écriture de plan que les facteurs. Chacun de ceux-ci est noté par une lettre. Le facteur aléatoire (sujet) est souligné.
- Pour chacun des facteurs, on détermine la relation qu'ils entretiennent avec le ou les facteurs aléatoires. Pour chacun des facteurs systématiques, l'indice correspond au nombre de modalités du facteur. Pour le facteur sujet, l'indice correspond au nombre de sujet divisé par le produit des indices des facteurs emboîtants.
- On assemble ces relations dans une formule de plan en commençant par le ou les facteurs aléatoires, suivis des facteurs emboîtants (s'il y en a plusieurs, ils sont en général croisés dans les chevrons)². Enfin, on liste les facteurs croisants.
- On obtient ainsi tous les facteurs dans une écriture de plan unique.

Exemple :

² Attention, dans ce cas à l'indice de \underline{S} .

1. Dans une expérience, on a 5 facteurs S_{40} , M_2 , P_2 , T_2 , E_3 . Les facteurs systématiques entretiennent les relations suivantes avec le facteur "sujet" : $S_{20} \langle M_2 \rangle$; $S_{20} \langle P_2 \rangle$; $S_{40} * T_2$; $S_{40} * E_3$.



Plans et partitions du protocole : la décomposition

Dans l'analyse de variance, on cherche à évaluer deux types d'effets des facteurs qu'on manipule sur la variable dépendante :

L'effet simple

Il correspond à la comparaison entre les modalités d'un facteur. C'est par exemple la comparaison de deux groupes, de deux tâches.

L'effet d'interaction

L'interaction entre deux facteurs, c'est une différence de différences. C'est l'idée que la différence observée entre les modalités d'un facteur n'est pas la même en fonction des modalités du second facteur. Si on étudie par exemple deux populations d'enfants provenant de milieux favorisés et défavorisés et qu'on les soumet à deux types de pédagogies, l'effet d'interaction entre ces facteurs revient à se demander si la différence de performance entre les deux milieux est la même dans les deux pédagogies. S'il y a un effet d'interaction, vous pouvez observer sur le graphique des moyennes que les deux droites sont sécantes. L'absence d'interaction se traduit graphiquement par des parallèles.

La décomposition canonique d'un plan, c'est-à-dire l'identification des différentes sources de variation, se fait de la manière suivante :

Les sources de variation inter.

Chacun des facteurs systématiques est une source de variation, dans notre exemple, on a M, P, T, E.

On fait ensuite, s'il y a lieu la combinatoire deux à deux des différents facteurs systématiques. Dans notre exemple, on a M.P, M.T, M.E, P.T, P.E, T.E. Il s'agit des interactions, c'est-à-dire des comparaisons sur un facteur des différences entre les modalités d'un autre facteur.

On fait ensuite, s'il y a lieu, la combinaison des facteurs par trois, puis par quatre jusqu'à épuisement des différentes combinaisons. Dans notre exemple, on a M.P.E, M.P.T, M.T.E, P.T.E, M.P.T.E.

Les sources de variation intra.

Le facteur sujet est indissociable du ou des facteurs emboîtants. Le facteur de groupe pour la variation intra est la combinaison de tous les facteurs de groupes et il a pour nombre de modalités le produit des modalités des facteurs emboîtants (dans notre exemple G a $2 \times 2 = 4$ modalités). Il est éventuellement en interaction avec un des facteurs croisants, ce qui déterminera d'autres sources de variation intra. On a donc les sources de variations intra suivantes : $\underline{S}(G)$, $\underline{S}(G).T$, $\underline{S}(G).E$, $\underline{S}(G).T.E$.

Le plan comme guide de planification de l'analyse.

Un plan est une formalisation de la structure du protocole et la déclaration d'une intention d'analyse répondant à un questionnement particulier correspondant à une hypothèse. Le plan d'analyse à considérer comprend tous les facteurs intervenant dans l'hypothèse (ce qui ne correspond pas forcément à tous les facteurs du plan). L'analyse descriptive et inférentielle est guidée par la décomposition du plan.

Dans l'analyse descriptive, on commentera l'effet de chacun des facteurs systématiques (source de variation inter) pris individuellement et l'effet de chacune des interactions s'il y a lieu.

Dans l'analyse inférentielle, on associe à chaque comparaison inter un F de Snédécour (rapport entre le carré moyen inter et le carré moyen intra associé c'est-à-dire reprenant le ou les facteurs de la comparaison inter).

Exercices

Exercice 1.

Voici l'extrait d'un article relatant la méthodologie d'une étude portant sur l'évaluation d'un programme de développement du langage chez le jeune enfant.

"Cette étude a été menée selon un plan longitudinal, de la GSM³ au CP⁴. Les données ont été recueillies auprès de 592 enfants âgés de cinq ans cinq mois en moyenne (écart-type : quatre mois) au début des observations. Deux groupes ont été constitués :

Un groupe expérimental composé de 362 enfants ayant bénéficié du programme d'entraînement psychopédagogique centré sur l'oral au cours du second et du troisième trimestre de GSM et provenant de douze classes de la région de Nantes et de Poitiers ;

Un groupe contrôle constitué de 230 enfants n'ayant pas bénéficié du programme d'entraînement. Ces enfants étaient scolarisés dans la région de Bourges.

L'entraînement a consisté principalement à favoriser le développement des compétences langagières en suscitant la prise de paroles et les échanges conversationnels au sein de petits groupes de sept ou huit enfants - grands, moyens et petits parleurs - constitués en début d'année scolaire à partir d'une évaluation de leur niveau de participation verbale à la conversation scolaire en grand groupe classe. Pour chaque groupe conversationnel, un nombre minimal de sept séances d'entraînement au total a été fixé. Cependant, diverses contraintes n'ont pas permis à tous les enseignants d'atteindre cet objectif ; en outre, les absences des enfants pouvaient limiter le nombre de séances auxquelles ils participaient. Il s'en suit une variation du nombre de séances selon les classes et les enfants. Les enseignants avaient reçu pour ces entraînements un guide de travail, élaboré d'après Florin (1995, chapitre 5), qui leur avait été présenté par les équipes de circonscription. Ce guide de travail présente des orientations générales pour les activités de langage (étapes du développement langagier, diversité interindividuelle, rôle des partenaires dans les apprentissages, etc.) et des propositions pour la mise en place de groupes conversationnels et le travail sur le langage oral. Il n'y avait aucune intervention directe de l'équipe de recherche dans les classes pour les entraînements. Le suivi longitudinal du groupe expérimental comporte quatre temps d'évaluations : début de GSM (T1), fin de GSM (T2), début de CP (T3) et fin de CP (T4). Le groupe contrôle a été évalué aux temps T1 et T4 uniquement."

Florin A., Guimard P., Khomsi A. (2000) Les effets d'un entraînement psychopédagogique. *Psychologie Française*. N°45-3, 219-232.

³ Grande section de maternelle.

⁴ Cours préparatoire

Sachant que les évaluations portaient à chaque temps sur les compétences langagières et scolaires évaluer à l'aide de tests appropriés donnant lieu à une note pour chacune des compétences, quelle est la structure du protocole de base (ne tenez pas compte du nombre de séances) ?

Pour répondre à la question "l'entraînement psychopédagogique permet-il le développement des compétences langagières et scolaire ?", quel plan d'analyse doit-on considérer ?

Exercice 2.

Voici l'extrait d'un article sur l'accès lexical :

"... Nous avons utilisé la tâche de dénomination de dessins avec amorçage. Les amorces (des mots présentés visuellement) pouvaient être des coordonnées sémantiques, des associés verbaux ou des mots non reliés avec les cibles dessin. De plus, parce que certaines études ont montré l'importance des paramètres temporels dans ce type de recherche (p. ex. Glaser et Dünghoff, 1984), nous avons aussi manipulé la durée de présentation de l'amorce. Ce facteur avait quatre modalités : trois visaient à évaluer les conséquences d'un traitement automatique de l'amorce (100 ms, 160 ms, 220 ms) et un dernier (600 ms) était destiné à évaluer des mécanismes, de nature différente (sur la question de l'automatisme, cf. Neely, 1991, par exemple).

Sujets. Cent vingt sujets (trente pour chaque durée de présentation) étudiants à l'université René Descartes (Paris V) ont participé à cette expérience dans le cadre de leur formation. Ils étaient tous locuteurs natifs du français et avaient une vue normale, éventuellement après correction.

Matériel. Nous avons utilisé trente dessins d'objets usuels. Ces dessins étaient issus des normes de Alario et Ferrand (1999) et de batteries de tests neuropsychologiques (Kay, Lesser et Coltheart, 1992). Chaque dessin apparaissait en noir sur blanc dans un carré de 96 x 96 pixels, au centre de l'écran de l'ordinateur. Il y avait en plus cinq dessins d'entraînement. Pour chaque dessin, trois amorces ont été utilisées : 1/ des compétiteurs sémantiquement reliés, qui n'étaient pas des associés (scie - HACHE ; cf. Wheeldon et Monsell, 1994) ; 2/ des associés qui n'étaient pas membres d'une même catégorie sémantique intuitive (bûche - HACHE ; cf. Ferrand et Alario, 1998) et 3/ des mots non-reliés (dose - HACHE).

Plan expérimental. Le facteur « durée de présentation » était manipulé entre groupes de sujets et le facteur « Relation amorce – cible » était croisé avec les sujets. Par ailleurs, les deux facteurs « durée de présentation de l'amorce » et « relation amorce – cible » étaient croisés avec les items. »

Alario F.X.(2001) Aspects sémantiques de l'accès au lexique. Psychologie française, 46-1, 17-26.

Quelle est la structure du protocole ?

Deux analyses ont été faites, l'une par sujets, l'autre par items. Quels en sont les plans d'analyse ?

Exercice 3.

Voici un extrait du compte-rendu de la procédure expérimentale utilisée dans une expérience sur la compréhension de documents techniques multimédia : 60 sujets ont participé à cette expérience.

« Les participants étaient au préalable invités à répondre à un pré-questionnaire de connaissances en mécanique. Seuls les novices étaient sélectionnés et répartis aléatoirement dans les groupes expérimentaux. Les sujets étaient placés individuellement devant l'écran d'un ordinateur sur lequel était présentée la version uniquement auditive du document ou la version multimédia. Dans ces deux conditions, l'écoute de l'explication se faisait via un casque stéréo. L'expérimentateur informait les sujets qu'ils allaient être confrontés une ou trois fois de suite, à un document portant sur le fonctionnement de pompes à air, et que leur tâche consistait à comprendre et à retenir au mieux les informations présentées en vue de répondre à un questionnaire. Ce questionnaire était présenté sous forme d'une épreuve de vérifications de phrases. Vingt phrases apparaissaient l'une après l'autre au centre de l'écran, dans un ordre aléatoire. La tâche des sujets consistait à juger de la véracité de chacune d'elles à l'aide d'un boîtier de réponses à deux boutons (vrai / faux) reliés à un ordinateur. L'appui sur l'une de ces

touches entraînait l'enregistrement de la réponse ainsi que du temps mis pour répondre depuis l'apparition de la question (temps de latence). Dix des vingt phrases étaient des paraphrases et reprenaient le contenu explicite du document en le reformulant sur le plan syntaxique. Les dix autres étaient des inférences et visaient à évaluer la compréhension de la situation décrite. »

Les hypothèses des auteurs sont les suivantes : « Nos prédictions sont que l'apprentissage est plus efficace lorsque le document est accompagné d'une illustration dynamique concurrente (i.e. réponses plus justes et plus rapides) et ce quel que soit le type de questions posé), que d'une manière générale, les présentations multiples permettent de mieux répondre aux questions posées. »

Hidrio C., Jamet E. (2002) compréhension d'un dispositif technique : apports d'une illustration dynamique et de traitements multiples. *Psychologie Française*. N°47-1, 61-67.

Décrivez la structure du protocole de base. Compte tenu des questions de recherche des auteurs, quel est le plan d'analyse ?

Nous allons maintenant présenter une situation expérimentale que nous utiliserons dans tous les exercices des prochains chapitres. La situation est la suivante :

Une expérience porte sur 84 sujets répartis par moitié dans un milieu social favorisé et dans un milieu social défavorisé. Dans chacun de ces deux groupes, on distingue deux types de pédagogie (moderne et traditionnelle). Les élèves ont passé trois épreuves notées de 0 à 10, la première fois en début d'année, et la seconde fois en fin d'année : une épreuve de combinatoire (Comb.), une épreuve de probabilité (Prob.) et une épreuve de logique (Log.). D'après Rouanet et Leroux (1993). (Voir le tableau de données en annexe 1)

Exercice 4

Décrivez les variables et la structure du protocole de base.

Exercice 5

Si on n'aborde pas la comparaison des trois épreuves, qu'elle est la structure du plan d'analyse ?

Exercice 6

Quelle est la décomposition du plan d'analyse déclaré à la question précédente ?

Exercice 7

Considérons l'hypothèse suivante : « La différence entre les deux pédagogies est plus importante en fin d'année ». A quelle source de variation renvoie cette hypothèse ?

- A : à la source de variation P
- B : à l'interaction M.P
- C : à l'interaction P.T
- D : à la source de variation $\underline{S}(G)$

Exercice 8.

A chaque source de variation « inter » correspond une hypothèse générale de différence (ou d'absence de différence H_0) et des hypothèses spécifiques précisant le sens des différences. Pour chacune des sources de variation « inter » identifiées précédemment indiquez ces hypothèses (faites un tableau pour plus de commodités).

Les corrigés

Corrigé Exercice 1.

Facteur "sujet" : S_{592} .

Facteur "groupes" : G_2 (expérimental, contrôle); $S\langle G_2 \rangle$.

Facteur "groupes conversationnels" : C_3 (grand, moyens et petits parleurs); $S\langle C_3 \rangle$.

Facteur "Temps d'évaluation" : T_2 (T_1 et T_4 , T_2 et T_3 ne concernant pas le groupe "contrôle", le plan ne serait pas complet); S^*T_2 .

Variable dépendante (VD1) : Score au test langagier, variable numérique.

Variable dépendante (VD2) : Score au test scolaire, variable numérique⁵.

La structure du protocole de base est $S\langle G_2^*C_3 \rangle^*T_2$. On notera que dans ce cas, le facteur S n'a pas d'indice parce que les groupes n'ont pas le même nombre de sujets (on dit qu'ils ne sont pas équilibrés).

La question de recherche tel qu'elle est posée peut donner lieu à plusieurs interprétations.

La première interprétation consiste à ne regarder que les sujets ayant eu un entraînement et se demander si la performance en T_4 est supérieure à T_1 . La question est alors comprise comme "lorsqu'on a un entraînement, T_4 est-il supérieur à T_1 . Cette interprétation ne permet pas l'attribution causale (que se passe-t-il s'il n'y a pas d'entraînement?). Dans ce cas, le plan d'analyse est $S_{362}^*T_2$ (on ne prend que les 362 sujets du groupe expérimental et le facteur "tâche").

Dans la seconde interprétation, "permettre" est assimilé à une condition (il n'y a que l'entraînement qui donne lieu à une augmentation de la performance). Cette interprétation permet de s'approcher d'une attribution causale (sous réserve que d'autres facteurs n'interviennent pas). Cela revient à se demander si la différence entre T_1 et T_4 est plus importante dans le groupe expérimental que dans le groupe contrôle. Le plan à considérer est alors $S\langle G_2 \rangle^*T_2$ et la question relève de l'interaction.

On pourrait peut-être en trouver d'autres. Ces exemples illustrent bien la difficulté de passer d'une problématique qui s'exprime en langage naturel à une formalisation nécessaire à l'emploi des outils statistiques et la prudence dont il faut faire preuve pour interpréter les résultats.

Corrigé Exercice 2.

Identification des variables.

Facteur "sujets" : S_{120}

Facteur "Items" : $S_{120}^*I_{30}$ (ici le facteur "Items" est considéré comme un facteur aléatoire puisque l'expérimentateur ne le contrôle pas (d'où l'intérêt de l'analyse par items).

Facteur "durée de présentation" : D_4 (100, 160, 220 et 600ms), $S_{30}\langle D_4 \rangle$ et $D_4^*I_{30}$

Facteur "Relation amorce-cible": R_3 (coordonnés, associés ou non-reliés), $S_{120}^*R_3$. et $R_3^*I_{30}$.

La variable dépendante (VD) est le temps de réponse.

La structure du protocole est $S_{30}\langle D_4 \rangle^*I_{30}^*R_3$.

$S_{30}\langle D_4 \rangle^*R_3$ pour l'analyse par sujets et $I_{30}^*D_4^*R_3$ pour l'analyse par items.

Corrigé Exercice 3.

Identification des variables et mise en relations avec le facteur « Sujet ».

Facteur « Sujet » : S_{60}

Facteur « Format » : F_2 (auditive ou multimédia); $S_{30}\langle F_2 \rangle$

Facteur « Présentation » : P_2 , une ou trois fois. $S_{30}\langle P_2 \rangle$

Facteur « Type de questions » : Q_2 , Paraphrase ou inférence. $S_{60}^*Q_2$

⁵ Ces deux compétences ne constituent pas un facteur, car il faudrait les évaluer avec la même variable dépendante (donc le même barème), ce qui n'a pas de sens.

Facteur « Questions⁶ » ; I_{20} ; $S_{60} * I_{20}$ et $I_{10} < Q_2 >$

Variable dépendante 1 (VD1) : Exactitude de la réponse : nominale dichotomique (juste ou fausse).

Variable dépendante 2 (VD2) : temps de latence, numérique.

Structure du protocole de base : $S_{15} < F_2 * P_2 > * I_{10} < Q_2 >$

Les hypothèses portent sur l'effet du « Format » (efficacité accrue avec illustration) et du facteur « Présentation » (les présentations multiples permettent de mieux répondre), ainsi que l'absence d'interaction entre le format et le type de questions (effet du format quel que soit le format de présentation). Le plan d'analyse devra retenir en plus du facteur « sujet » (qu'on ne peut pas supprimer) les facteurs F, P et Q. La VD1 et le facteur I sont recodés pour former la variable dépendante "nombre de bonnes réponses".

Le plan d'analyse devient alors : $S_{15} < F_2 * P_2 > * Q_2$

Corrigé Exercice 4.

Dans cette expérience, les individus statistiques sont les élèves. Nous avons 5 variables :

- Le milieu social (M) variable nominale dichotomique entretenant une relation d'emboîtement avec le facteur sujet (chaque sujet n'appartient qu'à un seul milieu).
- La pédagogie (P), variable nominale dichotomique entretenant une relation d'emboîtement avec le facteur sujet (chaque sujet ne voit qu'une seule pédagogie).
- Le type d'épreuve (E) variable nominale à trois modalités entretenant une relation de croisement avec le facteur sujet (chaque sujet passe les trois épreuves).
- La période de l'épreuve (T), variable ordinaire à deux modalités entretenant une relation de croisement avec le facteur sujet (Chaque sujet passe les épreuves deux fois).
- La variable performance, variable numérique à 11 modalités. C'est la variable dépendante.

Les groupes sont équilibrés. Le plan est donc $S_{21} < M_2 * P_2 > * T_2 * E_3$

Corrigé exercice 5

Dans ce cas, il faut considérer que les modalités du facteur « épreuve » correspondent à trois variables dépendantes qui seront traitées séparément. Le plan d'analyse est alors :

$$S_{21} < M_2 * P_2 > * T_2 \quad \text{-->} \quad \text{Comb, Prob, Log}$$

Nous listons après la flèche les variables dépendantes.

Note : Un plan d'analyse correspond à un point de vue sur les données. Un facteur n'est un facteur que parce qu'on considère que ses différentes valeurs (modalités) renvoient à une même dimension ou caractéristique de la situation ou des sujets. Le fait ici de ne pas s'intéresser à la comparaison entre les épreuves fait qu'elles ne sont plus considérées comme un facteur.

Corrigé Exercice 6

- Identification des sources de variation « inter ». La décomposition d'un plan commence par l'extraction des facteurs simples, dans ce cas, on a M, P, T.
- On fait ensuite la combinatoire des facteurs simples pour trouver les interactions, d'abord deux à deux. On a alors M.P, M.T, P.T (lire M point T etc.), Puis par trois (et éventuellement 4 etc.). Ici nous n'avons qu'une seule interaction triple : M.P.T (ce qui se lit M point P point T)
- Identification des sources de variation « intra »
- La variation intragroupe notée $S(G)$ (lire S dans G). Ici $G=4$ puisque nous avons deux facteurs emboîtant à deux modalités.

⁶ Ce facteur est un cas particulier puisqu'il n'entretient pas seulement de relation avec le facteur « sujet ». Ses modalités sont, en effet, emboîtées dans les modalités du facteur « type de questions ». Contrairement à l'exercice précédent, les questions ne sont pas traitées comme un facteur aléatoire puisqu'on ne fait pas d'analyse par question.

- La variation intragroupe associée à l'interaction, notée $S(G).T$ (lire S dans G point T). Ici, G est bien sûr toujours égal à 4.

Corrigé Exercice 7

- A : Non, on parle bien de la différence entre les deux pédagogies mais pour comparer la différence observée en début et en fin d'année.
- B : Non, on s'occupe bien des différence entre les pédagogies mais pas pour les comparer selon les milieux.
- C : Oui, C'est une bonne réponse.
- D. Non cette source de variation est une source « intra ». Elle concerne les différences entre les sujets à l'intérieur des groupes. La variation intersujet n'est pas contrôlée, on ne peut donc formuler d'hypothèse sur elle.

Corrigé Exercice 8

Classiquement une hypothèse se formule de façon affirmative (donc pas de conditionnel, pas d'interrogatif). Ces hypothèses sont des réponses à des questions de recherche (qui elles sont bien sûr sous forme interrogative). Les hypothèses suivantes correspondent à une décomposition canonique du plan. C'est pour les tester qu'on calcule un F de Snédécour. On peut bien sûr formuler des hypothèses qui ne correspondent pas à cette décomposition, mais nous laisserons cela de côté dans ce cours.

Sources de variation.	Hypothèse générale.	Hypothèse spécifique.
M	La performance varie selon le milieu.	Les sujets du milieu favorisé réussissent mieux les tests (ou l'inverse) ⁷
P	La performance varie selon la pédagogie.	La pédagogie moderne favorise la performance aux tests (ou l'inverse).
T	La performance varie selon la période de test.	Le test est moins bien réussi en début d'année qu'en fin d'année (ou l'inverse).
M.P	La différence entre les pédagogies n'est pas la même selon les milieux	L'écart entre les deux pédagogies est plus important pour les sujets du milieu défavorisé que pour les sujets du milieu favorisé (ou l'inverse).
M.T	La différence entre les milieux n'est pas la même en début et en fin d'année.	La différence entre les milieux est plus importante en fin d'année qu'en début d'année (ou l'inverse).
P.T	La différence entre les pédagogies n'est pas la même en début et en fin d'année.	La différence entre les pédagogies est plus importante en fin d'année qu'en début d'année (ou l'inverse).
M.P.T	L'interaction entre les milieux et les pédagogies n'est pas la même en début et en fin d'année.	L'interaction entre les milieux et les pédagogies est plus importante en début et en fin d'année (ou l'inverse).

⁷ On examine ici toutes les alternatives dans l'absolu, indépendamment du sens psychologique de ces hypothèses. Cette question du sens psychologique sera abordée plus loin.

La méthode des moyennes.

Cette méthode correspond à l'application du point de vue de l'hypothèse nulle. Elle n'est rien d'autre que l'application de la formule de définition de la variance corrigée sur le protocole dérivé par moyennage (en faisant la moyenne des observations) sur les modalités du facteur.

Cette formule de définition est :
$$Var_{corr} = \frac{\sum (x - m)^2}{n - 1}$$

Les sommes des carrés que vous allez calculer ne sont rien d'autres que les numérateurs de la variance tels que vous avez appris à les calculer en statistiques descriptives. Le nombre de degré de liberté correspondant au nombre de comparaisons qu'on peut faire sur une partition, le carré moyen n'est finalement rien d'autre que la variance corrigée associée au protocole dérivé. Concrètement, on calcule la moyenne des observations pour chacune des modalités du facteur (protocole dérivé par moyennage). On construit ensuite un tableau avec, en lignes, les modalités du facteur et en colonne les différentes étapes du calcul :

Facteur	n	m_g	m	$m_g - m$	$(m_g - m)^2$	$n(m_g - m)^2$
Modalité 1						
Modalité 2						

- La colonne n correspond au nombre d'observations pour chacune des modalités du facteur. Si les groupes sont équilibrés, c'est le nombre total d'observations divisé par le nombre de modalités du facteur, c'est-à-dire le nombre de lignes du tableau.
- Dans la colonne m_g , on met la moyenne des observations pour chacune des modalités du facteur
- La colonne m est celle de la moyenne générale, c'est-à-dire la valeur attendue.
- Dans la colonne suivante, on calcule la différence entre m_g et m.
- Puis on élève la différence au carré.
- Cette différence au carré est ensuite multipliée par n.
- Le total de la dernière colonne est la somme des carrés associée à la source de variation.

Quels que soient le plan et la partition considérés, la méthode est la même. Ce qui change c'est ce qu'on met en ligne, à savoir les modalités du facteur, le nombre d'observations et le protocole dérivé qui en découle.

Exercices.

Exercice 1 à 3

Réalisez une analyse de variance complète en utilisant cette seule méthode des moyennes sur l'exemple de référence pour les trois VD : logique, combinatoire, et probabilité. Nous rappelons ci-dessous la situation expérimentale :

Une expérience porte sur 84 sujets répartis par moitié dans un milieu social favorisé et dans un milieu social défavorisé. Dans chacun de ces deux groupes, on distingue deux types de pédagogie (moderne et traditionnelle). Les élèves ont passé trois épreuves notées de 0 à 10, la première fois en début d'année, et la seconde fois en fin d'année : une épreuve de combinatoire (Comb.), une épreuve de probabilité (Prob.) et une épreuve de logique (Log.). D'après Rouanet et Leroux (1993)

Vous considérez la structure du protocole de base la suivante :

$$\underline{S}_{21} < M_2 * P_2 > * T_2 \rightarrow \text{log, comb, prob.}$$

La décomposition de ce plan est :

Sources de variation inter : M; P; T; M.P; M.T; P.T; M.P.T

Sources de variation intra : $\underline{S}(G)$; $S(G).T$.

Corrigés.

Dans les exercices suivants, nous allons calculer toutes les sommes des carrés à l'aide de cette seule méthode des moyennes. La démarche sera exposée dans son détail pour la variable « logique ». Vous aurez à appliquer la même méthode pour les deux autres variables.

Corrigé de l'exercice 1 : Analyse de variance sur la variable "logique".

La première étape consiste à calculer les moyennes pour chacun des groupes et chacune des tâches. Le tableau des moyennes est le suivant.

<i>Tableau des moyennes</i>	<i>Moderne Début</i>	<i>Classique Début</i>	<i>Moderne Fin</i>	<i>Classique Fin</i>	<i>Moyennes</i>
Favorisé	4,262	3,948	5,710	4,976	4,724
Défavorisé	3,481	2,648	4,581	3,619	3,582
Moyennes	3,871	3,298	5,145	4,298	4,153

Calcul de la somme des carrés associée à M.

<i>Effet de M</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
Favorisé	84	4,724	4,153	0,571	0,326	27,371
Défavorisé	84	3,582	4,153	-0,571	0,326	27,371
$SC_M =$						54,743

Le facteur « milieu » comporte deux modalités que nous retrouvons en ligne dans le tableau. Les moyennes des groupes sont les moyennes marginales en ligne. On notera que le nombre d'observations est égal au nombre total d'observations divisés par le nombre de modalités du facteurs. Le nombre de degrés de liberté est égal au nombre de ligne moins un. La moyenne théorique est la moyenne générale.

Calcul de la somme des carrés associée à P.

<i>Effet de P</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
Moderne	84	4,508	4,153	0,355	0,126	10,607
Classique	84	3,798	4,153	-0,355	0,126	10,607
$SC_P =$						21,215

Calcul de la somme des carrés associée à T.

<i>Effet de T</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
Début	84	3,585	4,153	-0,568	0,323	27,144
Fin	84	4,721	4,153	0,568	0,323	27,144
$SC_T =$						54,287

Calcul de la somme des carrés associée à M.P.

<i>Effet de M.P</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
fav/mod & defav/class	84	4,060	4,153	-0,093	0,009	0,734
fav/class & defav/mod	84	4,246	4,153	0,093	0,009	0,734
SC _{M.P} =						1,467

Les modalités de l'interaction sont obtenues en croisant les deux facteurs. Ici pour la première ligne, on a additionné les moyennes du groupe « milieu favorisé , pédagogie moderne » et du groupe « milieu défavorisé, pédagogie classique » pour les deux tâches (début et fin d'année). On divise par quatre (le nombre de moyennes additionnées) pour obtenir la moyenne des moyennes (à condition que les groupes soient équilibrés). Le calcul est donc le suivant :

$$((4,262+5,710)+(2,648+3,619))/4= 4,060$$

On procède de la même façon pour l'autre modalité de l'interaction.

Calcul de la somme des carrés associée à M.T.

<i>Effet de M.T</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
fav/T1 & defav/T2	84	4,102	4,153	-0,051	0,003	0,215
fav/T2 & defav/T1	84	4,204	4,153	0,051	0,003	0,215
SC _{M.T} =						0,430

Le principe de calcul de la moyenne des groupes est le même que précédemment. Attention à ne pas oublier de prendre en compte les deux modalités de P.

Calcul de la somme des carrés associée à P.T.

<i>Effet de P.T</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
fav/T1 & defav/T2	84	4,085	4,153	-0,068	0,005	0,394
fav/T2 & defav/T1	84	4,221	4,153	0,068	0,005	0,394
SC _{P.T} =						0,787

Calcul de la somme des carrés associée à M.P.T.

Le calcul direct de la somme des carrés associés à la triple interaction serait trop fastidieux. On a donc recours à une méthode plus simple qui consiste à retrancher de la somme des carrés associée au croisement des trois facteurs (ne confondez pas avec l'interaction) les autres sommes des carrés inter. Cette somme des carrés représente en effet le total des variations inter du plan.

Les modalités de ce facteur complexe sont obtenues en faisant la combinatoire des modalités de chacun des facteurs élémentaires. Le nombre de lignes du tableau est égal au produit du nombre des modalités des facteurs élémentaires. Ici chacun de nos facteurs ont deux modalités, on aura donc $2*2*2=8$ lignes dans notre tableau.

Encore une fois le nombre d'observations est égal au nombre total des observations divisé par le nombre de ligne du tableau, soit $168/8= 21$.

<i>Effet de M*P*T</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
fav/mod/T1	21	4,262	4,153	0,109	0,012	0,249
fav/mod/T2	21	5,710	4,153	1,557	2,423	50,880
fav/class/T1	21	3,948	4,153	-0,205	0,042	0,886
fav/class/T2	21	4,976	4,153	0,823	0,678	14,231
défav/mod/T1	21	3,481	4,153	-0,672	0,452	9,484
défav/mod/T2	21	4,581	4,153	0,428	0,183	3,846
défav/class/T1	21	2,648	4,153	-1,505	2,266	47,588
défav/class/T2	21	3,619	4,153	-0,534	0,285	5,987
$SC_{M*P*T} =$						133,151

Ensuite la procédure est la même qu précédemment. Une fois qu'on a la somme des carrés associée au croisement de nos trois facteurs, on peut calculer la somme des carrés associée à la triple interaction en retranchant toutes les autres sources de variation inter

$$SC_{M,P,T} = SC_{M*P*T} - (SC_M + SC_P + SC_T + SC_{M,P} + SC_{M,T} + SC_{P,T})$$

$$SC_{M,P,T} = 133,151 - (54,743 + 21,215 + 54,287 + 1,467 + 0,430 + 0,787) = 0,221$$

Calcul de la somme des carrés associée à $\underline{S(G)}$

La méthode est toujours la même, mais cette fois, c'est la variance des sujets à l'intérieur des groupes qu'on cherche. Les modalités du facteur sont donc les sujets. La somme des carrés intragroupes est la somme des sommes des carrés de chacun des groupes. Les groupes à considérer résultent du croisement des facteurs emboîtants, dans notre exemple : Le milieu et la pédagogie. Chacun de ces deux facteurs ayant deux modalités, on a $2*2 = 4$ groupes qui se compose comme suit :

- Groupe 1 = Milieu favorisé ; Pédagogie moderne
- Groupe 2 = Milieu favorisé ; Pédagogie classique
- Groupe 3 = Milieu défavorisé ; Pédagogie moderne
- Groupe 4 = Milieu défavorisé ; Pédagogie classique

Les moyennes sont calculées selon le même principe que précédemment (on additionne, puis on divise par le nombre). Pour exemple, la moyenne du groupe g1 a été calculée de la façon suivante : $m = (4,262 + 5,710) / 2 = 4,986$

On a dans cet exemple 2 observations par sujet (une en début et un en fin d'année) correspondant aux deux modalités du facteur croisant. La moyenne du sujet est la moyenne de ces deux observations.

<i>Effet de S(g1)</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
s1	2	2,100	4,986	-2,886	8,327	16,655
s2	2	3,250	4,986	-1,736	3,013	6,025
s3	2	4,250	4,986	-0,736	0,541	1,083
s4	2	1,500	4,986	-3,486	12,150	24,300
s5	2	3,300	4,986	-1,686	2,842	5,683
s6	2	4,250	4,986	-0,736	0,541	1,083
s7	2	4,200	4,986	-0,786	0,617	1,235
s8	2	4,300	4,986	-0,686	0,470	0,940
s9	2	4,700	4,986	-0,286	0,082	0,163
s10	2	4,150	4,986	-0,836	0,698	1,397
s11	2	5,000	4,986	0,014	0,000	0,000
s12	2	5,650	4,986	0,664	0,441	0,883
s13	2	3,650	4,986	-1,336	1,784	3,568
s14	2	5,900	4,986	0,914	0,836	1,672
s15	2	6,800	4,986	1,814	3,292	6,583
s16	2	7,800	4,986	2,814	7,920	15,840
s17	2	5,650	4,986	0,664	0,441	0,883
s18	2	5,200	4,986	0,214	0,046	0,092
s19	2	7,600	4,986	2,614	6,834	13,669
s20	2	7,250	4,986	2,264	5,127	10,254
s21	2	8,200	4,986	3,214	10,332	20,663
$SC_{S(g1)}=1$						132,671

<i>Effet de S(g2)</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
s1	2	8,250	4,462	3,788	14,350	28,699
s2	2	2,750	4,462	-1,712	2,931	5,861
s3	2	1,850	4,462	-2,612	6,822	13,644
s4	2	2,300	4,462	-2,162	4,674	9,348
s5	2	2,600	4,462	-1,862	3,467	6,933
s6	2	4,950	4,462	0,488	0,238	0,476
s7	2	3,100	4,462	-1,362	1,855	3,710
s8	2	3,950	4,462	-0,512	0,262	0,524
s9	2	3,950	4,462	-0,512	0,262	0,524
s10	2	4,450	4,462	-0,012	0,000	0,000
s11	2	4,750	4,462	0,288	0,083	0,166
s12	2	5,650	4,462	1,188	1,412	2,823
s13	2	3,100	4,462	-1,362	1,855	3,710
s14	2	3,950	4,462	-0,512	0,262	0,524
s15	2	6,000	4,462	1,538	2,366	4,731
s16	2	2,750	4,462	-1,712	2,931	5,861
s17	2	4,850	4,462	0,388	0,151	0,301
s18	2	6,300	4,462	1,838	3,379	6,757
s19	2	6,550	4,462	2,088	4,360	8,720
s20	2	5,300	4,462	0,838	0,702	1,405
s21	2	6,350	4,462	1,888	3,565	7,130
$SC_{S(g2)}=$						111,849

<i>Effet de S(g3)</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
s1	2	7,750	4,031	3,719	13,831	27,663
s2	2	0,750	4,031	-3,281	10,765	21,529
s3	2	2,800	4,031	-1,231	1,515	3,030
s4	2	1,700	4,031	-2,331	5,433	10,867
s5	2	2,100	4,031	-1,931	3,729	7,457
s6	2	2,450	4,031	-1,581	2,499	4,999
s7	2	3,000	4,031	-1,031	1,063	2,126
s8	2	4,150	4,031	0,119	0,014	0,028
s9	2	6,600	4,031	2,569	6,600	13,200
s10	2	3,250	4,031	-0,781	0,610	1,220
s11	2	2,750	4,031	-1,281	1,641	3,282
s12	2	3,400	4,031	-0,631	0,398	0,796
s13	2	4,600	4,031	0,569	0,324	0,648
s14	2	3,850	4,031	-0,181	0,033	0,065
s15	2	4,950	4,031	0,919	0,845	1,689
s16	2	5,750	4,031	1,719	2,955	5,910
s17	2	2,800	4,031	-1,231	1,515	3,030
s18	2	4,800	4,031	0,769	0,591	1,183
s19	2	5,850	4,031	1,819	3,309	6,618
s20	2	6,300	4,031	2,269	5,149	10,297
s21	2	5,050	4,031	1,019	1,038	2,077
$SC_{S(g3)} =$						127,715

<i>Effet de S(g4)</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
s1	2	1,450	3,133	-1,683	2,834	5,667
s2	2	2,300	3,133	-0,833	0,694	1,389
s3	2	1,550	3,133	-1,583	2,507	5,014
s4	2	2,250	3,133	-0,883	0,780	1,561
s5	2	3,300	3,133	0,167	0,028	0,056
s6	2	1,300	3,133	-1,833	3,361	6,722
s7	2	2,000	3,133	-1,133	1,284	2,569
s8	2	3,000	3,133	-0,133	0,018	0,036
s9	2	2,850	3,133	-0,283	0,080	0,161
s10	2	2,800	3,133	-0,333	0,111	0,222
s11	2	2,600	3,133	-0,533	0,284	0,569
s12	2	3,050	3,133	-0,083	0,007	0,014
s13	2	3,750	3,133	0,617	0,380	0,761
s14	2	5,400	3,133	2,267	5,138	10,276
s15	2	2,650	3,133	-0,483	0,234	0,467
s16	2	3,100	3,133	-0,033	0,001	0,002
s17	2	4,200	3,133	1,067	1,138	2,276
s18	2	3,850	3,133	0,717	0,514	1,027
s19	2	2,700	3,133	-0,433	0,188	0,376
s20	2	6,300	3,133	3,167	10,028	20,056
s21	2	5,400	3,133	2,267	5,138	10,276
$SC_{S(g4)} =$						69,493

La somme des carrés intragroupes est la somme des sommes des carrés de chaque groupe. Ainsi, on a :

$$SC_{S(G)} = 132,671 + 111,849 + 127,715 + 69,493 = 441,729$$

Calcul de la somme des carrés intragroupes associée à $\underline{S}(G).T$.

Cette somme des carrés est la variance des différences entre les observations sur les tâches pour chacun des groupes. Concrètement la procédure est très similaire à la précédente. En ligne, les individus sont toujours les sujets. Le nombre d'observations est le même. La moyenne intrasujet est la différence entre les tâches divisée par le nombre de tâches. Par exemple, pour le sujet 1, on a les observations suivantes :

1^{ère} épreuve de logique : 0,8 et 2^{ème} épreuve de logique : 3,4. On aura donc pour ce sujet :

$$(0,8 - 3,4)/2 = -1,3$$

Pour le reste, la procédure est la même que précédemment.

Effet de $\underline{S}(g1).T$	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
s1	2	-1,3	-0,724	-0,576	0,332	0,664
s2	2	-0,45	-0,724	0,274	0,075	0,150
s3	2	-0,65	-0,724	0,074	0,005	0,011
s4	2	-0,2	-0,724	0,524	0,274	0,549
s5	2	-1,3	-0,724	-0,576	0,332	0,664
s6	2	-0,45	-0,724	0,274	0,075	0,150
s7	2	-1,3	-0,724	-0,576	0,332	0,664
s8	2	0,5	-0,724	1,224	1,498	2,995
s9	2	0,1	-0,724	0,824	0,679	1,357
s10	2	-1,85	-0,724	-1,126	1,268	2,537
s11	2	-1	-0,724	-0,276	0,076	0,153
s12	2	-2,45	-0,724	-1,726	2,980	5,959
s13	2	-1,15	-0,724	-0,426	0,182	0,363
s14	2	-0,9	-0,724	-0,176	0,031	0,062
s15	2	-0,7	-0,724	0,024	0,001	0,001
s16	2	-0,2	-0,724	0,524	0,274	0,549
s17	2	0,45	-0,724	1,174	1,378	2,756
s18	2	-1,4	-0,724	-0,676	0,457	0,914
s19	2	0,4	-0,724	1,124	1,263	2,526
s20	2	-1,15	-0,724	-0,426	0,182	0,363
s21	2	-0,2	-0,724	0,524	0,274	0,549
					$SC_{S(g1).T=1}$	23,936

<i>Effet de $S(g2).T$</i>	n_q	m_q	m	m_q-m	$(m_q-m)^2$	$n_q(m_q-m)^2$
s1	2	-0,150	-0,514	0,364	0,133	0,265
s2	2	-1,050	-0,514	-0,536	0,287	0,574
s3	2	-0,050	-0,514	0,464	0,216	0,431
s4	2	-0,500	-0,514	0,014	0,000	0,000
s5	2	-0,900	-0,514	-0,386	0,149	0,298
s6	2	-0,550	-0,514	-0,036	0,001	0,003
s7	2	0,700	-0,514	1,214	1,474	2,949
s8	2	-0,150	-0,514	0,364	0,133	0,265
s9	2	-1,050	-0,514	-0,536	0,287	0,574
s10	2	0,150	-0,514	0,664	0,441	0,883
s11	2	-0,850	-0,514	-0,336	0,113	0,225
s12	2	-0,950	-0,514	-0,436	0,190	0,380
s13	2	0,200	-0,514	0,714	0,510	1,020
s14	2	-0,850	-0,514	-0,336	0,113	0,225
s15	2	0,500	-0,514	1,014	1,029	2,058
s16	2	-0,950	-0,514	-0,436	0,190	0,380
s17	2	-1,350	-0,514	-0,836	0,698	1,397
s18	2	0,400	-0,514	0,914	0,836	1,672
s19	2	-1,550	-0,514	-1,036	1,073	2,145
s20	2	-0,900	-0,514	-0,386	0,149	0,298
s21	2	-0,950	-0,514	-0,436	0,190	0,380
$SC_{S(g2).T} =$						16,421

<i>Effet de $S(g3).T$</i>	n_q	m_q	m	m_q-m	$(m_q-m)^2$	$n_q(m_q-m)^2$
s1	2	-0,750	-0,550	-0,200	0,040	0,080
s2	2	0,750	-0,550	1,300	1,690	3,380
s3	2	-0,400	-0,550	0,150	0,023	0,045
s4	2	-0,800	-0,550	-0,250	0,063	0,125
s5	2	-1,200	-0,550	-0,650	0,423	0,845
s6	2	0,550	-0,550	1,100	1,210	2,420
s7	2	-0,200	-0,550	0,350	0,123	0,245
s8	2	0,450	-0,550	1,000	1,000	2,000
s9	2	-1,100	-0,550	-0,550	0,303	0,605
s10	2	0,050	-0,550	0,600	0,360	0,720
s11	2	-1,250	-0,550	-0,700	0,490	0,980
s12	2	-0,200	-0,550	0,350	0,123	0,245
s13	2	-1,300	-0,550	-0,750	0,563	1,125
s14	2	-1,250	-0,550	-0,700	0,490	0,980
s15	2	0,050	-0,550	0,600	0,360	0,720
s16	2	-1,550	-0,550	-1,000	1,000	2,000
s17	2	-0,600	-0,550	-0,050	0,002	0,005
s18	2	-0,300	-0,550	0,250	0,063	0,125
s19	2	-0,550	-0,550	0,000	0,000	0,000
s20	2	-0,900	-0,550	-0,350	0,123	0,245
s21	2	-1,050	-0,550	-0,500	0,250	0,500
$SC_{S(g3).T} =$						17,390

<i>Effet de S(g4).T</i>	n_g	m_g	m	$m_g - m$	$(m_g - m)^2$	$n_g(m_g - m)^2$
s1	2	0,650	-0,486	1,136	1,290	2,580
s2	2	-0,700	-0,486	-0,214	0,046	0,092
s3	2	-0,450	-0,486	0,036	0,001	0,003
s4	2	-0,450	-0,486	0,036	0,001	0,003
s5	2	0,200	-0,486	0,686	0,470	0,940
s6	2	-0,100	-0,486	0,386	0,149	0,298
s7	2	0,200	-0,486	0,686	0,470	0,940
s8	2	-0,100	-0,486	0,386	0,149	0,298
s9	2	-0,850	-0,486	-0,364	0,133	0,265
s10	2	-0,800	-0,486	-0,314	0,099	0,198
s11	2	0,700	-0,486	1,186	1,406	2,812
s12	2	0,050	-0,486	0,536	0,287	0,574
s13	2	-0,650	-0,486	-0,164	0,027	0,054
s14	2	-0,600	-0,486	-0,114	0,013	0,026
s15	2	-0,850	-0,486	-0,364	0,133	0,265
s16	2	-1,200	-0,486	-0,714	0,510	1,020
s17	2	-0,900	-0,486	-0,414	0,172	0,343
s18	2	-2,150	-0,486	-1,664	2,770	5,540
s19	2	-0,700	-0,486	-0,214	0,046	0,092
s20	2	-0,100	-0,486	0,386	0,149	0,298
s21	2	-1,400	-0,486	-0,914	0,836	1,672
$SC_{S(g4).T} =$						18,311

Comme précédemment, on fera la somme des sommes des carrés de chaque groupe. On obtient ainsi :

$$SC_{S(G).T} = 23,936 + 16,421 + 17,390 + 18,311 = 76,059$$

Synthèse des calculs.

<i>Source de variation</i>	<i>ddl</i>	<i>SC</i>	<i>CM</i>	<i>F</i>	<i>p</i>
S<M*P>*T	167	650,939	3,898		
M	1	54,743	54,743	9,914	
P	1	21,215	21,215	3,842	
T	1	54,287	54,287	57,100	
M.P	1	1,467	1,467	0,266	
M.T	1	0,430	0,430	0,452	
P.T	1	0,787	0,787	0,828	
M.P.T	1	0,221	0,221	0,233	
S(G)	80	441,729	5,522		
S(G).T	80	76,059	0,951		

Le tableau ci-dessus synthétise les résultats de l'analyse de variance. Le carré moyen est la somme des carrés divisée par le nombre de degré de liberté. Concrètement, il représente la variance corrigée du protocole dérivé associé à une source de variation. Le F de Snédécour est la statistique sur laquelle nous allons faire l'inférence. On le calcule en faisant le rapport entre le carré moyen inter et le carré moyen intra (voir le formulaire). Nous avons laissé vide la colonne p où doit être inscrite la proportion d'échantillons pour lesquels le F est supérieur au F calculé. Nous reprendrons cela dans la section « interprétation des résultats ».

On remarquera que la somme des sommes des carrés des sources de variation inter et intra est égale à la somme des carrés du plan total (S<M*P>*T). Concrètement, cela signifie

qu'une des sommes des carrés peut-être calculée par différence. C'est généralement ce qu'on fait pour la dernière somme des carrés ($S(G).T$) qui lorsqu'on a plus de deux modalités au facteur T devient difficile à calculer à la main.

**Corrigé de l'exercice 2 : Analyse de variance sur la variable
« combinatoire ».**

<i>Source de variation</i>	<i>ddl</i>	<i>SC</i>	<i>CM</i>	<i>F</i>	<i>p</i>
<i>S<M*P>*T</i>	167	711,779	4,262		
M	1	5,075	5,075	1,040	
P	1	31,547	31,547	6,466	
T	1	138,249	138,249	80,729	
M.P	1	0,461	0,461	0,094	
M.T	1	7,209	7,209	4,209	
P.T	1	1,929	1,929	1,126	
M.P.T	1	0,004	0,004	0,002	
S(G)	80	390,306	4,879		
S(G).T	80	137,000	1,713		

Corrigé de l'exercice 3: Analyse de variance sur la variable « probabilités ».

<i>Source de variation</i>	<i>ddl</i>	<i>SC</i>	<i>CM</i>	<i>F</i>	<i>p</i>
<i>S<M*P>*T</i>	167	589,073	3,527		
M	1	8,550	8,550	1,707	
P	1	8,641	8,641	1,725	
T	1	16,657	16,657	9,932	
M.P	1	13,430	13,430	2,681	
M.T	1	4,307	4,307	2,568	
P.T	1	2,453	2,453	1,463	
M.P.T	1	0,050	0,050	0,030	
S(G)	80	400,817	5,010		
S(G).T	80	134,168	1,677		

La méthode des totaux

Comme la précédente, cette méthode relève de l'application du point de vue de l'hypothèse nulle. Elle n'est rien d'autre que l'application de la formule de calcul de la variance corrigée sur le protocole dérivé par sommage (en faisant la somme des observations) sur les modalités du facteur.

La formule de calcul de la variance est :
$$Var_{corr} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

Comme à chaque fois, les sommes des carrés qu'on calcule correspondent au numérateur de la variance corrigée. Concrètement, on calcule d'abord la somme des observations pour chacune des modalités du facteur (protocole dérivé par sommage). On construit ensuite un tableau avec en ligne les modalités du facteur plus une ligne pour le total, et en colonne les différences étapes du calcul.

Facteur	T	n	T ² /n
Modalité 1			a
Modalité 2			b
Total			c

Dans la colonne T, on mettra le total des observations pour la modalités et dans la colonne n, le nombre d'observations pour le facteur. C'est le nombre total d'observations divisé par le nombre de modalités du facteur, c'est-à-dire le nombre de lignes du tableau.

Dans la dernière colonne, on élève le total au carré, puis on divise ce total par le nombre d'observations.

La somme des carrés associée à la source de variation est la somme des résultats de la dernière colonne pour chacune des modalités du facteur moins le résultat de la dernière colonne pour le total, soit (a+b)-c.

La méthode vaut pour tous les plans et toutes les sources de variation qu'on peut considérer. Dans les exercices suivants, nous allons calculer à l'aide de cette seule méthode les sommes des carrés associées aux différentes sources de variation du plan qui nous sert d'exemple de référence. Comme précédemment, nous détaillerons seulement les calculs sur la variable « logique ». Vous aurez ensuite à transposer l'exemple sur les deux autres variables.

Exercices

Réalisez une analyse de variance complète en utilisant cette seule méthode des totaux sur l'exemple de référence pour les trois VD : logique, combinatoire, et probabilité. Nous rappelons ci-dessous la situation expérimentale :

Une expérience porte sur 84 sujets répartis par moitié dans un milieu social favorisé et dans un milieu social défavorisé. Dans chacun de ces deux groupes, on distingue deux types de pédagogie (moderne et traditionnelle). Les élèves ont passé trois épreuves notées de 0 à 10, la première fois en début d'année, et la seconde fois en fin d'année : une épreuve de combinatoire (Comb.), une épreuve de probabilité (Prob.) et une épreuve de logique (Log.). D'après Rouanet et Leroux (1993)

Vous considérez la structure du protocole de base la suivante :

$$\underline{S}_{21} < M_2 * P_2 > * T_2 \rightarrow \text{log, comb, prob.}$$

La décomposition de ce plan est :

Sources de variation inter : M; P; T; M.P; M.T; P.T; M.P.T

Sources de variation intra: S(G); S(G).T

Corrigés.

Dans les exercices suivants, nous allons calculer, à l'aide de cette seule méthode des totaux, les sommes des carrés associées aux différentes sources de variation du plan qui nous sert d'exemple de référence. Comme précédemment, nous détaillerons seulement les calculs sur la variable « logique ». Vous aurez ensuite à transposer l'exemple sur les deux autres variables.

Analyse de variance sur la VD "logique".

La première étape consiste à calculer les moyennes pour chacun des groupes et chacune des tâches. Le tableau des totaux est le suivant.

<i>Tableau des totaux</i>	<i>Moderne Début</i>	<i>Classique Début</i>	<i>Moderne Fin</i>	<i>Classique Fin</i>	<i>Total</i>
Favorisé	89,500	82,900	119,900	104,500	396,800
Défavorisé	73,100	55,600	96,200	76,000	300,900
Total	162,600	138,500	216,100	180,500	697,700

Calcul de la somme des carrés associée à M.

<i>Effet de M</i>	<i>T_g</i>	<i>n_g</i>	<i>T²/n</i>
Favorisé	396,800	84	1874,408
Défavorisé	300,900	84	1077,867
Total	697,700	168	2897,531
		SC _M	54,743

Comme avec la méthode des moyennes, n_g est égal au nombre total des observations divisé par le nombre de modalités du facteur, soit le nombre de lignes. Le nombre de degrés de liberté est égal au nombre de ligne moins 1 (on exclut de ce nombre de lignes la ligne « total »). La somme des carrés est égale à la somme des sommes des carrés pour les modalités (en blanc) moins la somme des carrés pour le total en gris soit :

$$SC_M = (1874,408 + 1077,867) - 2897,531 = 54,743.$$

Calcul de la somme des carrés associée à P.

<i>Effet de P</i>	<i>T_g</i>	<i>n_g</i>	<i>T²/n</i>
Moderne	378,700	84	1707,306
Classique	319,000	84	1211,440
Total	697,700	168	2897,531
		SC _P	21,215

Calcul de la somme des carrés associée à T.

<i>Effet de T</i>	<i>T_g</i>	<i>n_g</i>	<i>T²/n</i>
Début	301,100	84	1079,300
Fin	396,600	84	1872,519
Total	697,700	168	2897,531
		SC _T	54,287

Calcul de la somme des carrés associée à M.P.

Les modalités de l'interaction sont déterminées comme précédemment dans la méthode des moyennes.

<i>Effet de M.P</i>	T_g	n_g	T^2/n
M1P1+M2P2	341,000	84	1384,298
M1P2+M2P1	356,700	84	1514,701
Total	697,700	168	2897,531
		$SC_{M,P}$	1,467

Calcul de la somme des carrés associée à M.T.

<i>Effet de M.T</i>	T_g	n_g	T^2/n
M1T1+M2T2	344,600	84	1413,680
M1T2+M2T1	353,100	84	1484,281
Total	697,700	168	2897,531
		$SC_{M,T}$	0,430

Calcul de la somme des carrés associée à P.T.

<i>Effet de P.T</i>	T_g	n_g	T^2/n
P1T1+P2T2	343,100	84	1401,400
P1T2+P2T1	354,600	84	1496,919
Total	697,700	168	2897,531
		$SC_{P,T}$	0,787

Calcul de la somme des carrés associée à M.P.T.

<i>Effet de M*P*T</i>	T_g	n_g	T^2/n
M1P1T1	89,500	21	381,440
M1P1T2	82,900	21	327,258
M1P2T1	119,900	21	684,572
M1P2T2	104,500	21	520,012
M2P1T1	73,100	21	254,458
M2P1T2	55,600	21	147,208
M2P2T1	96,200	21	440,688
M2P2T2	76,000	21	275,048
Total	697,700	168	2897,531
		SC_{M*P*T}	133,151

Comme avec la méthode des moyennes, la somme des carrés associée à la triple interaction sera calculée par différence entre la somme des carrés associée au croisement des facteurs et les sommes des carrés associées aux autres sources de variation inter.

$$SC_{M,P,T} = SC_{M*P*T} - (SC_M + SC_P + SC_T + SC_{M,P} + SC_{M,T} + SC_{P,T})$$

$$SC_{M,P,T} = 133,151 - (54,743 + 21,215 + 54,287 + 1,467 + 0,430 + 0,787) = 0,221$$

Calcul de la somme des carrés associée à S(G).

Comme précédemment la somme des carrés intragroupes est calculée en faisant la somme des sommes des carrés de chaque groupe.

Effet de $S(g1)$	T_g	n_g	T^2/n
S1	4,2	2	8,820
S2	6,5	2	21,125
S3	8,5	2	36,125
S4	3	2	4,500
S5	6,6	2	21,780
S6	8,5	2	36,125
S7	8,4	2	35,280
S8	8,6	2	36,980
S9	9,4	2	44,180
S10	8,3	2	34,445
S11	10	2	50,000
S12	11,3	2	63,845
S13	7,3	2	26,645
S14	11,8	2	69,620
S15	13,6	2	92,480
S16	15,6	2	121,680
S17	11,3	2	63,845
S18	10,4	2	54,080
S19	15,2	2	115,520
S20	14,5	2	105,125
S21	16,4	2	134,480
Total	209,4	42	1044,009
		$SC_{S(g1)}$	132,671

Effet de $S(g2)$	T_g	n_g	T^2/n
S1	16,5	2	136,125
S2	5,5	2	15,125
S3	3,7	2	6,845
S4	4,6	2	10,580
S5	5,2	2	13,520
S6	9,9	2	49,005
S7	6,2	2	19,220
S8	7,9	2	31,205
S9	7,9	2	31,205
S10	8,9	2	39,605
S11	9,5	2	45,125
S12	11,3	2	63,845
S13	6,2	2	19,220
S14	7,9	2	31,205
S15	12	2	72,000
S16	5,5	2	15,125
S17	9,7	2	47,045
S18	12,6	2	79,380
S19	13,1	2	85,805
S20	10,6	2	56,180
S21	12,7	2	80,645
Total	187,4	42	836,161
		$SC_{S(g2)}$	111,849

<i>Effet de S(g3)</i>	T_g	n_g	T^2/n
S1	15,5	2	120,125
S2	1,5	2	1,125
S3	5,6	2	15,680
S4	3,4	2	5,780
S5	4,2	2	8,820
S6	4,9	2	12,005
S7	6	2	18,000
S8	8,3	2	34,445
S9	13,2	2	87,120
S10	6,5	2	21,125
S11	5,5	2	15,125
S12	6,8	2	23,120
S13	9,2	2	42,320
S14	7,7	2	29,645
S15	9,9	2	49,005
S16	11,5	2	66,125
S17	5,6	2	15,680
S18	9,6	2	46,080
S19	11,7	2	68,445
S20	12,6	2	79,380
S21	10,1	2	51,005
Total	169,3	42	682,440
		$SC_{S(g3)}$	127,715

<i>Effet de S(g4)</i>	T_g	n_g	T^2/n
S1	2,9	2	4,205
S2	4,6	2	10,580
S3	3,1	2	4,805
S4	4,5	2	10,125
S5	6,6	2	21,780
S6	2,6	2	3,380
S7	4	2	8,000
S8	6	2	18,000
S9	5,7	2	16,245
S10	5,6	2	15,680
S11	5,2	2	13,520
S12	6,1	2	18,605
S13	7,5	2	28,125
S14	10,8	2	58,320
S15	5,3	2	14,045
S16	6,2	2	19,220
S17	8,4	2	35,280
S18	7,7	2	29,645
S19	5,4	2	14,580
S20	12,6	2	79,380
S21	10,8	2	58,320
Total	131,6	42	412,347
		$SC_{S(g4)}$	69,493

Comme précédemment, la somme des carrés intragroupes est la somme des sommes des carrés de chaque groupe. Ainsi, on a :

$$SC_{S(G)} = 132,671 + 111,849 + 127,715 + 69,493 = 441,729$$

Calcul de la somme des carrés associée à $S(G).T$.

Ici, le total pour chacun des sujets est obtenu en faisant la somme des observations du sujet pour les deux tâches. Le reste reprend la procédure exposée pour la somme des carrés intragroupes (calcul de la somme des carrés pour chaque groupe et addition)

<i>Effet de S(g1).T</i>	T_g	n_g	T^2/n	<i>Effet de S(g2).T</i>	T_g	n_g	T^2/n
S1	-2,6	2	3,380	S1	-0,3	2	0,045
S2	-0,9	2	0,405	S2	-2,1	2	2,205
S3	-1,3	2	0,845	S3	-0,1	2	0,005
S4	-0,4	2	0,080	S4	-1	2	0,500
S5	-2,6	2	3,380	S5	-1,8	2	1,620
S6	-0,9	2	0,405	S6	-1,1	2	0,605
S7	-2,6	2	3,380	S7	1,4	2	0,980
S8	1	2	0,500	S8	-0,3	2	0,045
S9	0,2	2	0,020	S9	-2,1	2	2,205
S10	-3,7	2	6,845	S10	0,3	2	0,045
S11	-2	2	2,000	S11	-1,7	2	1,445
S12	-4,9	2	12,005	S12	-1,9	2	1,805
S13	-2,3	2	2,645	S13	0,4	2	0,080
S14	-1,8	2	1,620	S14	-1,7	2	1,445
S15	-1,4	2	0,980	S15	1	2	0,500
S16	-0,4	2	0,080	S16	-1,9	2	1,805
S17	0,9	2	0,405	S17	-2,7	2	3,645
S18	-2,8	2	3,920	S18	0,8	2	0,320
S19	0,8	2	0,320	S19	-3,1	2	4,805
S20	-2,3	2	2,645	S20	-1,8	2	1,620
S21	-0,4	2	0,080	S21	-1,9	2	1,805
Total	-30,4	42	22,004	Total	-21,6	42	11,109
		$SC_{S(g1)}$	23,936			$SC_{S(g1)}$	16,421

<i>Effet de S(g3).T</i>	T_a	n_a	T^2/n	<i>Effet de S(g4).T</i>	T_a	n_a	T^2/n
S1	-1,5	2	1,125	S1	1,3	2	0,845
S2	1,5	2	1,125	S2	-1,4	2	0,980
S3	-0,8	2	0,320	S3	-0,9	2	0,405
S4	-1,6	2	1,280	S4	-0,9	2	0,405
S5	-2,4	2	2,880	S5	0,4	2	0,080
S6	1,1	2	0,605	S6	-0,2	2	0,020
S7	-0,4	2	0,080	S7	0,4	2	0,080
S8	0,9	2	0,405	S8	-0,2	2	0,020
S9	-2,2	2	2,420	S9	-1,7	2	1,445
S10	0,1	2	0,005	S10	-1,6	2	1,280
S11	-2,5	2	3,125	S11	1,4	2	0,980
S12	-0,4	2	0,080	S12	0,1	2	0,005
S13	-2,6	2	3,380	S13	-1,3	2	0,845
S14	-2,5	2	3,125	S14	-1,2	2	0,720
S15	0,1	2	0,005	S15	-1,7	2	1,445
S16	-3,1	2	4,805	S16	-2,4	2	2,880
S17	-1,2	2	0,720	S17	-1,8	2	1,620
S18	-0,6	2	0,180	S18	-4,3	2	9,245
S19	-1,1	2	0,605	S19	-1,4	2	0,980
S20	-1,8	2	1,620	S20	-0,2	2	0,020
S21	-2,1	2	2,205	S21	-2,8	2	3,920
Total	-23,1	42	12,705	Total	-20,4	42	9,909
		$SC_{S(g1)}$	17,390			$SC_{S(g1)}$	18,311

Comme précédemment, on fera la somme des sommes des carrés de chaque groupe.
On obtient ainsi :

$$SC_{S(g),T} = 23,936 + 16,421 + 17,390 + 18,311 = 76,059$$

La méthode des contrastes.

Cette méthode est l'application du point de vue du même nom. Elle n'est pas souvent employée parce qu'un peu fastidieuse, mais permet de bien comprendre ce que représente la variance corrigée, c'est-à-dire une comparaison deux à deux des observations. En pratique, elle présente surtout un intérêt pour les comparaisons partielles et est indispensable pour le calcul de la somme des carrés associés à la composante linéaire. La formule appliquée dans le calcul des contrastes est la suivante :

$$SC_C = \frac{\left(\sum_{g=1}^k c_g m_g \right)^2}{\sum_{g=1}^k \frac{c_g^2}{n_g}}$$

Comme à chaque fois, les sommes des carrés qu'on calcule correspondent au numérateur de la variance corrigée. L'application de cette méthode a pour point de départ le tableau des moyennes.

Facteur	n_g	m_g	C	$C_g m_g$	$-C_g m_g$	$-C_g m_g^2$	C^2/n_g	$-c^2/n_g$	SC
Modalité 1				a			b		.a+b
Modalité 2									

- La première colonne contient le nombre d'observations de chaque groupe. La seconde est celle des moyennes de chaque groupe.
- Dans la colonne C, on va indiquer les contrastes, c'est-à-dire des coefficients. En pratique, on choisit pour la première ligne un contraste égale au nombre de lignes de la comparaison moins un, les autres lignes du tableau ayant pour contraste : -1.
- On fait ensuite le produit des contrastes de chaque ligne par la moyenne correspondante.
- Dans la première case de la 5^{ème} colonne, on additionne les produits des moyennes par les contrastes.
- Dans la 6^{ème} colonne, on calcule le rapport des carrés des contrastes par le nombre d'observations.
- Ensuite, on fait la somme de la 6^{ème} colonne dans la première case de la 7^{ème} colonne.
- La somme des carrés associée à cette comparaison est égale au rapport entre le résultat de la 5^{ème} colonne et le résultat de la 7^{ème} colonne.

Si le facteur a plus de deux modalités, le nombre de contraste à construire est égal au nombre de modalités moins un. On peut appliquer cette méthode quel que soit le plan et quelle que soit la source de variation considérée.

Exercices

Comme pour les précédentes méthodes, nous allons exemplifier l'application de cette méthode sur la variable « logique » du protocole qui nous sert d'exemple de référence. Vous aurez donc à calculer l'analyse de variance à l'aide de la seule méthode des contrastes. Nous rappelons à nouveau l'énoncé de l'exemple de référence pour vous éviter de fastidieux retours en arrière. Comme précédemment, nous détaillerons seulement les calculs sur la variable « logique ». Vous aurez ensuite à transposer l'exemple sur les deux autres variables.

Réalisez une analyse de variance complète en utilisant cette seule méthode des contrastes sur l'exemple de référence pour les trois VD : logique, combinatoire, et probabilité. Nous rappelons ci-dessous la situation expérimentale :

Une expérience porte sur 84 sujets répartis par moitié dans un milieu social favorisé et dans un milieu social défavorisé. Dans chacun de ces deux groupes, on distingue deux types de pédagogie (moderne et traditionnelle). Les élèves ont passé trois épreuves notées de 0 à 10, la première fois en début d'année, et la seconde fois en fin d'année : une épreuve de combinatoire (Comb.), une épreuve de probabilité (Prob.) et une épreuve de logique (Log.). D'après Rouanet et Leroux (1993) .

Vous considérez la structure du protocole de base la suivante :

$\underline{S}_{21} < M_2 * P_2 > * T_2 \rightarrow \text{log, comb, prob.}$

La décomposition de ce plan est :

Sources de variation inter: M; P; T; M.P; M.T; P.T; M.P.T

Sources de variation intra: $\underline{S}(G)$; $\underline{S}(G).T$

Corrigés

Analyse de variance sur la VD "logique".

Comme pour les précédentes méthodes, nous allons exemplifier l'application de cette méthode sur la variable « logique » du protocole qui nous sert d'exemple de référence.

La première étape consiste à calculer les moyennes pour chacun des groupes et chacune des tâches. Le tableau des moyennes est le suivant.

Tableau des moyennes	Moderne Début	Classique Début	Moderne Fin	Classique Fin	Moyennes
Favorisé	4,262	3,948	5,710	4,976	4,724
Défavorisé	3,481	2,648	4,581	3,619	3,582
Moyennes	3,871	3,298	5,145	4,298	4,153

Calcul de la somme des carrés associée à M.

Le facteur n'ayant que deux modalités, le premier contraste (1^{ère} ligne) sera fixé à 2-1=1. Le contraste suivant sera de -1. En réalité, on peut choisir n'importe quelle valeur pour le contraste, à condition que la somme des contrastes soit nulle. Ainsi si on avait choisi 2 pour valeur du premier contraste, il aurait fallu choisir -2 pour le second, la somme des deux étant 0. Pour des raisons pratiques évidentes (la multiplication par 1 c'est facile), il est recommandé de choisir pour premier contraste le nombre de lignes moins un, puis d'affecter un contraste de -1 à toutes les autres lignes.

<i>Effet de M</i>	n_g	m	C	$C_g m_g$	$\Sigma C_g m_g$	$\Sigma C_g m_g^2$	C^2/n_g	$\Sigma c^2/n_g$	SC
Favorisé	84	4,724	1	4,724	1,142	1,303	0,012	0,024	54,743
Défavorisé	84	3,582	-1	-3,582			0,012		
SC_M									54,743

- 5^{ème} colonne : Une fois les contrastes fixés, on multiplie les moyennes par leur contraste.
- 6^{ème} colonne : On additionne ensuite ces produits.
- 7^{ème} colonne : On élève ensuite cette somme au carré
- 8^{ème} colonne : On élève le contraste (4^{ème} colonne) au carré et on le divise par n_g .
- 9^{ème} colonne : On fait la somme des valeurs de la 8^{ème} colonne.

La somme des carrés est le rapport entre la valeur de la 7^{ème} et de la 9^{ème} colonne, soit :
 $SC_M = 1,303/0,024 = 54,747$

Calcul de la somme des carrés associée à P.

<i>Effet de P</i>	n_g	m	C	$C_g m_g$	$\Sigma C_g m_g$	$\Sigma C_g m_g^2$	C^2/n_g	$\Sigma c^2/n_g$	SC
Moderne	84	4,508	1	4,508	0,711	0,505	0,012	0,024	21,215
Classique	84	3,798	-1	-3,798			0,012		
SC_P									21,215

Calcul de la somme des carrés associée à T.

<i>Effet de T</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC	
Début	84	3,585	1	3,585	-1,137	1,293	0,012	0,024	54,287	
Fin	84	4,721	-1	-4,721			0,012			
									SC_T	54,287

Calcul de la somme des carrés associée à M.P.

<i>Effet de M.P</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC	
M1P1+M2P2	84	4,060	1	4,060	-0,187	0,035	0,012	0,024	1,467	
M1P2+M2P1	84	4,246	-1	-4,246			0,012			
									$SC_{M.P}$	1,467

Calcul de la somme des carrés associée à M.T.

<i>Effet de M.T</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC	
M1T1+M2T2	84	4,102	1	4,102	-0,101	0,010	0,012	0,024	0,430	
M1T2+M2T1	84	4,204	-1	-4,204			0,012			
									$SC_{M.T}$	0,430

Calcul de la somme des carrés associée à P.T.

<i>Effet de P.T</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC	
P1T1+P2T2	84	4,085	1	4,085	-0,137	0,019	0,012	0,024	0,787	
P1T2+P2T1	84	4,221	-1	-4,221			0,012			
									$SC_{P.T}$	0,787

Calcul de la somme des carrés associée à M.P.T.

Jusqu'à maintenant, nous avons vu l'application de la méthode des contrastes sur des comparaisons à un degré de liberté (facteur à 2 modalités). Dans le cas de la somme des carrés associée au croisement des facteurs la comparaison a : $(2*2*2)-1=7$ degrés de liberté. Il va donc falloir construire 7 comparaisons.

La première comparaison est celle de la première modalité comparée aux sept autres. Puis on aura la seconde modalité comparée aux six suivantes, puis la troisième comparée aux 5 restantes etc. Nous présentons l'ensemble des contrastes dans le tableau suivant :

<i>Effet de M*P*T</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
M1P1T1	21	4,262	7	29,833	0,871	0,759	2,333	2,667	0,285
M1P1T2	21	5,710	-1	-5,710			0,048		
M1P2T1	21	3,948	-1	-3,948			0,048		
M1P2T2	21	4,976	-1	-4,976			0,048		
M2P1T1	21	3,481	-1	-3,481			0,048		
M2P1T2	21	4,581	-1	-4,581			0,048		
M2P2T1	21	2,648	-1	-2,648			0,048		
M2P2T2	21	3,619	-1	-3,619			0,048		
M1P1T2	21	5,710	6	34,257	11,005	121,105	1,714	2,000	60,552
M1P2T1	21	3,948	-1	-3,948			0,048		
M1P2T2	21	4,976	-1	-4,976			0,048		
M2P1T1	21	3,481	-1	-3,481			0,048		
M2P1T2	21	4,581	-1	-4,581			0,048		
M2P2T1	21	2,648	-1	-2,648			0,048		
M2P2T2	21	3,619	-1	-3,619			0,048		
M1P2T1	21	3,948	5	19,738	0,433	0,188	1,190	1,429	0,131
M1P2T2	21	4,976	-1	-4,976			0,048		
M2P1T1	21	3,481	-1	-3,481			0,048		
M2P1T2	21	4,581	-1	-4,581			0,048		
M2P2T1	21	2,648	-1	-2,648			0,048		
M2P2T2	21	3,619	-1	-3,619			0,048		
M1P2T2	21	4,976	4	19,905	5,576	31,094	0,762	0,952	32,649
M2P1T1	21	3,481	-1	-3,481			0,048		
M2P1T2	21	4,581	-1	-4,581			0,048		
M2P2T1	21	2,648	-1	-2,648			0,048		
M2P2T2	21	3,619	-1	-3,619			0,048		
M2P1T1	21	3,481	3	10,443	-0,405	0,164	0,429	0,571	0,287
M2P1T2	21	4,581	-1	-4,581			0,048		
M2P2T1	21	2,648	-1	-2,648			0,048		
M2P2T2	21	3,619	-1	-3,619			0,048		
M2P1T2	21	4,581	2	9,162	2,895	8,382	0,190	0,286	29,338
M2P2T1	21	2,648	-1	-2,648			0,048		
M2P2T2	21	3,619	-1	-3,619			0,048		
M2P2T1	21	2,648	1	2,648	-0,971	0,944	0,048	0,095	9,909
M2P2T2	21	3,619	-1	-3,619			0,048		
SC_{M*P*T}									133,151

La somme des carrés associée au croisement des facteurs est égale à la somme des sommes des carrés de chaque contraste . Concrètement, on fait le total de la dernière colonne.

Comme avec les méthodes précédentes, la somme des carrés associée à la triple interaction sera calculée par différence entre la somme des carrés associée au croisement des facteurs et les sommes des carrés associées aux autres sources de variation inter.

$$SC_{M,P,T} = SC_{M*P*T} - (SC_M + SC_P + SC_T + SC_{M,P} + SC_{M,T} + SC_{P,T})$$

$$SC_{M,P,T} = 133,151 - (54,743 + 21,215 + 54,287 + 1,467 + 0,430 + 0,787) = 0,221$$

Calcul de la somme des carrés associée à $S(G)$

Le principe général reste le même qu'avec les autres méthodes : Calculer les sommes des carrés intra de chaque groupe et ensuite les additionner. Compte tenu de la longueur des calculs, nous ne détaillerons pas les calculs pour tous les groupes, mais seulement pour le premier.

- Le nombre d'observations est calculé comme pour les autres méthodes.
- La moyenne est bien la moyenne des observations pour chaque sujet.

Nous avons 21 sujets dans le groupe 1, il va donc falloir construire 21-1=20 tableaux de contrastes en reprenant le même principe que celui que nous avons énoncé pour la somme des carrés associée au croisement des facteurs :

- Le premier sujet contre les 20 suivants.
- Le second sujet contre les 19 suivants.
- Le troisième sujet contre les 18 suivants etc.

On applique ensuite la même procédure. La somme des carrés intra pour le groupe 1 est la somme des sommes des carrés associées aux 20 contrastes.

Notez, et c'est toujours comme cela, que le nombre de degrés de liberté est égal au nombre de contrastes à construire, c'est-à-dire au nombre de comparaisons qu'on peut faire. On a, dans cet exemple 20ddl par groupe soit 4*20 ddl pour la comparaison intragroupes.

<i>Effet de $S(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S1	2	2,1	20	42	-60,6	3672,36	200	210	17,487
S2	2	3,25	-1	-3,25			0,5		
S3	2	4,25	-1	-4,25			0,5		
S4	2	1,5	-1	-1,5			0,5		
S5	2	3,3	-1	-3,3			0,5		
S6	2	4,25	-1	-4,25			0,5		
S7	2	4,2	-1	-4,2			0,5		
S8	2	4,3	-1	-4,3			0,5		
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S2	2	3,25	19	61,75	-37,6	1413,76	180,5	190	7,441
S3	2	4,25	-1	-4,25			0,5		
S4	2	1,5	-1	-1,5			0,5		
S5	2	3,3	-1	-3,3			0,5		
S6	2	4,25	-1	-4,25			0,5		
S7	2	4,2	-1	-4,2			0,5		
S8	2	4,3	-1	-4,3			0,5		
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S3	2	4,25	18	76,5	-18,6	345,96	162	171	2,023
S4	2	1,5	-1	-1,5			0,5		
S5	2	3,3	-1	-3,3			0,5		
S6	2	4,25	-1	-4,25			0,5		
S7	2	4,2	-1	-4,2			0,5		
S8	2	4,3	-1	-4,3			0,5		
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S4	2	1,5	17	25,5	-68,1	4637,61	144,5	153	30,311
S5	2	3,3	-1	-3,3			0,5		
S6	2	4,25	-1	-4,25			0,5		
S7	2	4,2	-1	-4,2			0,5		
S8	2	4,3	-1	-4,3			0,5		
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S5	2	3,3	16	52,8	-37,5	1406,25	128	136	10,340
S6	2	4,25	-1	-4,25			0,5		
S7	2	4,2	-1	-4,2			0,5		
S8	2	4,3	-1	-4,3			0,5		
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S6	2	4,25	15	63,75	-22,3	497,29	112,5	120	4,144
S7	2	4,2	-1	-4,2			0,5		
S8	2	4,3	-1	-4,3			0,5		
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S7	2	4,2	14	58,8	-23,05	531,3025	98	105	5,060
S8	2	4,3	-1	-4,3			0,5		
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S8	2	4,3	13	55,9	-21,65	468,7225	84,5	91	5,151
S9	2	4,7	-1	-4,7			0,5		
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S9	2	4,7	12	56,4	-16,45	270,6025	72	78	3,469
S10	2	4,15	-1	-4,15			0,5		
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S10	2	4,15	11	45,65	-23,05	531,3025	60,5	66	8,050
S11	2	5	-1	-5			0,5		
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S11	2	5	10	50	-13,7	187,69	50	55	3,413
S12	2	5,65	-1	-5,65			0,5		
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S12	2	5,65	9	50,85	-7,2	51,84	40,5	45	1,152
S13	2	3,65	-1	-3,65			0,5		
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S13	2	3,65	8	29,2	-25,2	635,04	32	36	17,640
S14	2	5,9	-1	-5,9			0,5		
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de S(g1)</i>	n_g	m	C	$C_g m_g$	$SC_g m_g$	$SC_g m_g^2$	C^2/n_g	Sc^2/n_g	SC
S14	2	5,9	7	41,3	-7,2	51,84	24,5	28	1,851
S15	2	6,8	-1	-6,8			0,5		
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S15	2	6,8	6	40,8	-0,9	0,81	18	21	0,039
S16	2	7,8	-1	-7,8			0,5		
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S16	2	7,8	5	39	5,1	26,01	12,5	15	1,734
S17	2	5,65	-1	-5,65			0,5		
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S17	2	5,65	4	22,6	-5,65	31,9225	8	10	3,192
S18	2	5,2	-1	-5,2			0,5		
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S18	2	5,2	3	15,6	-7,45	55,5025	4,5	6	9,250
S19	2	7,6	-1	-7,6			0,5		
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S19	2	7,6	2	15,2	-0,25	0,0625	2	3	0,021
S20	2	7,25	-1	-7,25			0,5		
S21	2	8,2	-1	-8,2			0,5		

<i>Effet de $\underline{S}(g1)$</i>	<i>n_g</i>	<i>m</i>	<i>C</i>	<i>$C_g m_g$</i>	<i>$SC_g m_g$</i>	<i>$SC_g m_g^2$</i>	<i>C^2/n_g</i>	<i>Sc^2/n_g</i>	<i>SC</i>
S20	2	7,25	1	7,25	-0,95	0,9025	0,5	1	0,902
S21	2	8,2	-1	-8,2			0,5		

$$SC_{\underline{S}(g1)} = 17,487 + 7,441 + 2,023 + 30,311 + 10,340 + 4,144 + 5,060 + 5,151 + 3,469 + 8,050 + 3,413 + 1,152 + 17,640 + 1,851 + 0,039 + 1,734 + 3,192 + 9,250 + 0,021 + 0,902 = 132,671$$

Calcul de la somme des carrés associée à S(G).T

Nous ne détaillerons pas ici le calcul de cette somme des carrés à l'aide de la méthode des contrastes. La procédure est similaire à la précédente, en prenant comme moyenne pour chaque sujet, la différence moyenne entre les tâches (voir la méthode des moyennes pour ce calcul).

Les relations entre les plans.

Comprendre ces relations est un puissant moyen de contrôle permettant d'éviter bien des erreurs dans le cas de calculs à la main ou à l'aide d'un tableur. Nous avons vu que la somme des carrés du plan total est la somme des sommes des carrés de toutes les sources de variation et qu'on peut calculer par différences certaines d'entre elles à condition de connaître les autres, notamment la somme des carrés du plan total. Il en va de même pour les degrés de liberté de chacune des sources de variation, dont le total est égal au nombre de ddl du plan. Il existe également des relations du même type entre les plans.

La structure du protocole de l'exemple de référence est $S_{21} \times M_2 \times P_2 \times T_2$. Nous redonnons pour plus de facilités le tableau d'analyse de variance sur la variable « logique ».

Source de variation	ddl	SC	CM	F	p
$S \times M \times P \times T$	167	650,939	3,898		
M	1	54,743	54,743	9,914	
P	1	21,215	21,215	3,842	
T	1	54,287	54,287	57,100	
M.P	1	1,467	1,467	0,266	
M.T	1	0,430	0,430	0,452	
P.T	1	0,787	0,787	0,828	
M.P.T	1	0,221	0,221	0,233	
S(G)	80	441,729	5,522		
S(G).T	80	76,059	0,951		

Imaginez qu'on ne se soit intéressé qu'au facteur T. Notre plan d'analyse aurait été $S_{84} \times T_2$. Un tel plan se décompose en S , T et $S.T$.

La somme des carrés de S n'est que la somme des carrés des moyennes de chaque sujet sur chaque tâche (cf tableau suivant).

S	T1	T2	m_i	$xi-m$	$n(xi-m)^2$
1	0,8	3,4	2,1	-2,053	8,429
2	2,8	3,7	3,25	-0,903	1,631
3	3,6	4,9	4,25	0,097	0,019
4	1,3	1,7	1,5	-2,653	14,077
5	2	4,6	3,3	-0,853	1,455
6	3,8	4,7	4,25	0,097	0,019
7	2,9	5,5	4,2	0,047	0,004
8	4,8	3,8	4,3	0,147	0,043
9	4,8	4,6	4,7	0,547	0,598
10	2,3	6	4,15	-0,003	0,000
...
77	4,8	6	5,4	1,247	3,110
78	1,8	3,5	2,65	-1,503	4,518
79	1,9	4,3	3,1	-1,053	2,218
80	3,3	5,1	4,2	0,047	0,004
81	1,7	6	3,85	-0,303	0,184
82	2	3,4	2,7	-1,453	4,222
83	6,2	6,4	6,3	2,147	9,219
84	4	6,8	5,4	1,247	3,110
Moyenne	4,153			SC_S	519,154

Quant à la somme des carrés de $\underline{S.T}$, elle est la somme des carrés des différences entre les tâches pour chaque sujet⁸ (cf tableau suivant) .

<i>S</i>	<i>Log</i>	<i>Log2</i>	<i>différence</i>	<i>xi-m</i>	<i>n(xi-m)2</i>
1	0,8	3,4	-1,300	-0,732	1,070
2	2,8	3,7	-0,450	0,118	0,028
3	3,6	4,9	-0,650	-0,082	0,013
4	1,3	1,7	-0,200	0,368	0,272
5	2	4,6	-1,300	-0,732	1,070
6	3,8	4,7	-0,450	0,118	0,028
7	2,9	5,5	-1,300	-0,732	1,070
8	4,8	3,8	0,500	1,068	2,283
9	4,8	4,6	0,100	0,668	0,894
10	2,3	6	-1,850	-1,282	3,285
...
76	3,1	4,4	-0,650	-0,082	0,013
77	4,8	6	-0,600	-0,032	0,002
78	1,8	3,5	-0,850	-0,282	0,159
79	1,9	4,3	-1,200	-0,632	0,798
80	3,3	5,1	-0,900	-0,332	0,220
81	1,7	6	-2,150	-1,582	5,003
82	2	3,4	-0,700	-0,132	0,035
83	6,2	6,4	-0,100	0,468	0,439
84	4	6,8	-1,400	-0,832	1,383
Moyenne			-0,568	$SC_{S,T}$	77,498

Le tableau d'analyse de variance pour le plan d'analyse $\underline{S}_{84} * T_2$ sur notre exemple de référence est le suivant :

<i>Source de variation</i>	<i>ddl</i>	<i>SC</i>	<i>CM</i>	<i>F</i>	<i>p</i>
S*T	167	650,939	3,898		
T	1	54,287	54,287	58,141	
S	83	519,154	6,255		
S.T	83	77,498	0,934		

Nous aurions pu reconstituer ce tableau à partir de l'analyse de variance sur le plan $S_{21} < M_2 * P_2 > * T_2$.

- La somme des carrés du plan total est la même, puisqu'on travaille sur les mêmes observations.
- La somme des carrés associée à T avait déjà été calculée, c'est la même.
- Quant à $SC_{\underline{S.T}}$, elle nous est donnée dans le tableau par le total des sommes des carrés des autres sources de variation impliquant T. soit :

$$SC_{\underline{S.T}} = SC_{M.T} + SC_{P.T} + SC_{M.P.T} + SC_{\underline{S}(G).T} = 77,498$$

- La somme des carrés associée à \underline{S} se calculant par différence ou en additionnant les sources de variation n'impliquant pas T.

On remarquera que le F de Snédécour est un peu plus important que dans l'analyse sur le plan de base. Cela tient au fait qu'à la source de variation $\underline{S}(G).T$, on a ajouté les interactions.

⁸ Rappelez-vous comment on calcule la variance corrigée qui sert à calculer s dans le t de Student pour les groupes appariés.

Celles-ci étant faibles, cela n'a que peu de conséquences, mais si elles avaient été plus importantes, cela aurait pu avoir une incidence sur la significativité du F calculé.

Ceci n'est qu'un exemple et il n'est pas question de connaître toutes les relations. Il faut retenir de cela que le principe de l'analyse de variance est bien de calculer la part de la variance totale qui revient à chacun des sources de variation et que ces parts de variance entretiennent des relations additives.

Commentaires et conclusions.

Venons en maintenant à la partie, souvent la plus difficile, mais aussi la plus intéressante : le commentaire et la formulation de la conclusion

Schématiquement, on peut décomposer la construction du commentaire en quatre grandes étapes

Rappel des hypothèses. Les résultats de l'analyse de variance ne peuvent prendre leur sens qu'en relation avec les hypothèses qu'on cherche à tester et la justification de ces hypothèses. Rappelons qu'une hypothèse consiste à affirmer quelque chose sur l'importance et/ou le sens de l'écart entre les modalités d'un facteur systématique.

Analyse descriptive. Cette étape est indispensable. C'est elle qui permet de montrer à l'aide de graphique et/ou de tableaux de moyennes, les écarts observés et le sens de ceux-ci. Le commentaire descriptif des résultats s'appuie sur la décomposition du plan en sources de variation systématiques (il faut donc commenter chaque facteur).

Calcul du F et Lecture de la proportion dans la table. La lecture de cette table s'effectue en cherchant la colonne correspondant au nombre de degrés de liberté du numérateur et la ligne correspondant au nombre de degrés de liberté du dénominateur. Dans le pavé correspondant, on cherchera la valeur la plus proche et inférieure à la valeur de F calculée. La valeur de p est lue en tête de ligne. Le résultat sera déclaré significatif si la valeur de p est inférieure à .025.

Conclusions. La conclusion est d'ordre psychologique (sinon on change de métier). Il s'agit d'une mise en relation de la problématique qui a présidé à la mise en place de l'expérience et des résultats au niveau descriptif et inférentiel.

Exercices

Pour bien comprendre cela, reprenons notre exemple de référence.

Une expérience porte sur 84 sujets répartis par moitié dans un milieu social favorisé et dans un milieu social défavorisé. Dans chacun de ces deux groupes, on distingue deux types de pédagogie (moderne et traditionnelle). Les élèves ont passé trois épreuves notées de 0 à 10, la première fois en début d'année, et la seconde fois en fin d'année : une épreuve de combinatoire (Comb.), une épreuve de probabilité (Prob.) et une épreuve de logique (Log.). D'après Rouanet et Leroux (1993).

Vous considérez la structure du protocole de base la suivante :

$S_{21} < M_2 * P_2 > * T_2 \rightarrow \text{log, comb, prob.}$

La décomposition de ce plan est :

Sources de variation inter: M; P; T; M.P; M.T; P.T; M.P.T

Sources de variation intra: $\underline{S}(G)$; $\underline{S}(G).T$

Commentez les résultats de l'expérience pour les trois variables dépendantes.

Corrigés

Les hypothèses.

Nous l'avons dit plus haut, ce sont des affirmations à propos de l'effet des facteurs systématiques (ceux qu'on a manipulés). On va se servir de la décomposition en sources de variation pour lister les hypothèses. L'objectif de cette expérience est de tester la pertinence d'une pédagogie moderne par rapport à une pédagogie classique. On peut considérer que la nouvelle pédagogie sera plus intéressante que l'autre à condition que :

- les sujets réussissent mieux les épreuves avec la nouvelle pédagogie
- La nouvelle pédagogie ne creuse pas, voir atténue les différences sociales.
- L'effet d'apprentissage soit plus important avec la nouvelle pédagogie.

L'effet de M. Si on a contrasté ces deux milieux, c'est qu'on fait l'hypothèse qu'il influence la performance, ce qui somme toute est assez triviale. Cette influence devrait jouer en faveur du milieu favorisé où l'on devrait observer de meilleures performances. L'hypothèse peut donc se formuler ainsi : Les sujets du milieu favorisé ont des scores plus importants que les sujets du milieu défavorisé.

L'effet de P. Ici, il apparaît clair qu'on cherche à étudier l'intérêt d'une pédagogie moderne par rapport à une pédagogie classique. L'hypothèse sera donc : la pédagogie moderne favorise la réussite aux épreuves.

L'effet de T. Comme dans beaucoup d'expériences en psychologie, la répétition de mesure est utilisée pour mesurer l'effet d'apprentissage. On fait donc l'hypothèse de meilleures performances en T2 qu'en T1.

L'effet d'interaction M.P. L'interaction est une différence des différences. Elle peut se formuler de la façon suivante : la différence entre les deux pédagogies n'est pas la même selon le milieu. Une des conditions nécessaires pour qu'un changement pédagogique soit intéressant, c'est précisément qu'il ne creuse pas les disparités sociales ou qu'il les réduise. On doit donc s'attendre, si la pédagogie moderne est aussi intéressante qu'elle y paraît à ce qu'il n'y ai pas d'interaction entre le milieu et la pédagogie (effet non significatif) ou que l'interaction aille dans le sens d'une diminution des différences entre les milieu.

L'interaction M.T. Ici, il s'agit de s'interroger sur la variation de la différence entre les deux milieux pour les mesures en début et en fin d'année. La différence entre T1 et T2 reflétant un effet d'apprentissage, cela revient à supposer un effet d'apprentissage plus important pour un des deux milieux. Cette hypothèse peut raisonnablement être retenue. On formulera donc l'hypothèse suivante : La différence entre T1 et T2 est plus importante dans le milieu favorisé que dans le milieu défavorisé.

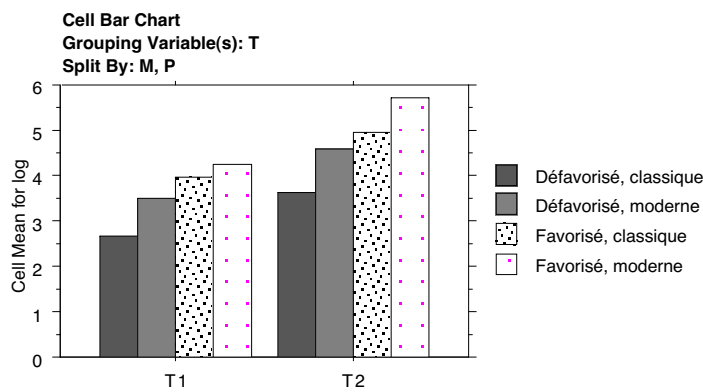
L'interaction P.T. Postuler une interaction entre ces deux facteurs, c'est postuler un effet d'apprentissage (différence entre T1 et T2) plus important dans une des deux pédagogies. On peut donc poser l'hypothèse d'une différence entre T1 et T2 plus importante dans le groupe ayant eu une pédagogie moderne que dans le groupe ayant eu une pédagogie classique.

L'interaction M.P.T : Les effets de triple interaction sont un peu compliqués à interpréter. Il s'agirait ici de voir si l'interaction entre deux des trois facteurs est la même selon les modalités du troisième facteur. Nous laisserons de côté cette source de variation dans le cadre de ce cours.

Analyse des résultats sur la variable « logique ».

Dans une analyse descriptive, on commence par donner un résumé graphique et/ou numérique des résultats. Ceux de cette expérience, pour la variable "logique", sont présentés ci-dessous sous la forme d'un tableau de moyennes.

<i>Tableau des moyennes</i>	<i>Moderne Début</i>	<i>Classique Début</i>	<i>Moderne Fin</i>	<i>Classique Fin</i>	<i>Moyennes</i>
Favorisé	4,262	3,948	5,710	4,976	4,724
Défavorisé	3,481	2,648	4,581	3,619	3,582
Moyennes	3,871	3,298	5,145	4,298	4,153



On peut également résumer ces données sous forme graphique.

Formellement nous avons dissocié en introduction de cette partie du cours l'analyse descriptive et l'analyse inférentielle pour montrer qu'elles ne se confondent pas et qu'elles sont tous les deux indispensables. En pratique, elles sont réalisées simultanément et commentées ensemble. C'est ce que nous allons faire après avoir rappeler les résultats de l'analyse de variance sur la variable "logique".

<i>Source de variation</i>	<i>ddl</i>	<i>SC</i>	<i>CM</i>	<i>F</i>	<i>p</i>
<i>S<M*P>*T</i>	167	650,939	3,898		
M	1	54,743	54,743	9,914	<.005
P	1	21,215	21,215	3,842	ns
T	1	54,287	54,287	57,100	<.0005
M.P	1	1,467	1,467	0,266	ns
M.T	1	0,430	0,430	0,452	ns
P.T	1	0,787	0,787	0,828	ns
M.P.T	1	0,221	0,221	0,233	ns
S(G)	80	441,729	5,522		
S(G).T	80	76,059	0,951		

Lecture de la table du F. On regarde la table dans la colonne correspondant au nombre de ddl de la source de variation inter et dans la ligne correspondant au nombre de ddl de la source de variation intra. Dans cet exemple précis, toutes les sources de variation inter ont un ddl et toutes les sources de variation intra ont 80 ddl. La ligne 80 n'existant pas dans la table qui est dans votre livre, on prendra la ligne immédiatement inférieure c'est-à-dire 60. Nous reproduisons ci-dessous le pavé correspondant :

ddl		1
60	.10	2,79
	.05	4,00
	.025	5,29
	.01	7,08
	.005	8,49
	.001	12,0
	.0005	13,6

On cherche ensuite la valeur la plus élevée, mais inférieure au F calculé et on lit la valeur de p. Si la valeur de p est inférieure au seuil critique, on déclare le résultat significatif au seuil correspondant. À titre d'exemple, le $F_{M[1,80]}=9,914$ (entre crochets sont indiqués les ddl du numérateur et du dénominateur. La valeur la plus élevée inférieure à ce F est 8,49. Elle correspond à la valeur de p .005. ce seuil étant inférieur à .025 (seuil critique), on peut déclarer le test significatif à un seuil $<.005$. En revanche $F_{P[1,80]}=3,842$, ce qui correspond à la valeur de p .10. Cette valeur de p étant supérieure au seuil critique, le résultat est déclaré non significatif (ns).

Commentaires : pour structurer le commentaire et éviter les répétitions, il est conseillé de s'appuyer sur la décomposition en sources de variation du plan et sur les hypothèses. Nous avons séparé plus haut les commentaires descriptifs et inférentiels pour des raisons pédagogiques. En pratique, les deux commentaires sont construits simultanément pour éviter la redondance. C'est donc de cette manière que nous allons présenter le corrigé.

L'effet de M. Conformément à l'hypothèse, on observe bien que les sujets du milieu favorisé ont des notes plus importantes au test de logique que les sujets du milieu défavorisé. Cette différence est significative ($F_{M[1,80]}=9,914$; $p<.005$). On peut donc affirmer que la performance au test de logique dépend du milieu d'origine des sujets. Les sujets du milieu favorisé ont de meilleurs résultats.

L'effet de P. On observe des différences entre les deux types de pédagogie qui vont dans le sens de l'hypothèse. Les résultats des sujets ayant reçu une pédagogie moderne sont supérieurs aux résultats des sujets ayant eu une pédagogie classique. Cette différence n'est cependant pas significative ($F_{P[1,80]}=3,842$; ns). On ne peut donc pas affirmer que la pédagogie influence la performance au test de logique.

L'effet de T. Les résultats montrent que dans tous les groupes, la performance augmentent entre le début et la fin de l'année. Cet effet est significatif ($F_{T[1,80]}=54,287$; $p<.0005$). Il y a donc bien un effet d'apprentissage entre les deux tâches.

L'effet de M.P. On n'observe pas d'interaction entre les milieux et les types de pédagogie (la différence entre milieu favorisé et milieu défavorisé n'est pas plus importante dans la pédagogie classique que dans la pédagogie moderne). La comparaison étant non significative ($F_{M,P[1,80]}=0,266$; ns), on conclura que le type de pédagogie ne change rien à la différence entre les deux milieux.

L'effet de M.T. La question de cet effet est celle de l'importance de l'effet d'apprentissage plus important dans l'un ou l'autre des deux milieux. Les résultats montrent qu'il n'y a pas d'interaction entre ces facteurs ($F_{M,T[1,80]}=0,430$; ns). L'importance de l'apprentissage est donc le même dans les deux milieux.

L'effet P.T. On interroge ici une variation possible de l'effet d'apprentissage selon le type de pédagogie. On s'attend à un effet de T plus important dans la pédagogie moderne que dans la pédagogie classique. Les résultats de l'expérience montrent qu'il n'en est rien ($F_{P,T[1,80]}=0,828$; ns). L'effet d'apprentissage ne varie pas en fonction de la pédagogie.

L'effet M.P.T sera laissé de côté dans le cadre de ce cours.

Conclusion. La problématique de cette expérience est d'évaluer la pertinence d'une pédagogie (moderne) par rapport à une autre (classique). Une pédagogie est plus intéressante qu'une autre si elle permet d'obtenir de meilleurs résultats (effet de P), si elle ne creuse pas les écarts entre les groupes de sujet (effet de M.P) ou permet de les réduire (à condition que ce ne soit pas à la baisse). Elle peut également être intéressante si elle permet un apprentissage plus important (effet de P.T). Les résultats sur l'épreuve de logique de cette expérience ne permettent de mettre en évidence aucune de ces conditions. On conclura donc que, sur cette épreuve de logique, la pédagogie moderne n'est pas plus intéressante que la pédagogie classique.

Analyse des résultats sur la variable « Combinatoire ».

Le jeu d'hypothèses sur l'effet de chacun des facteurs est le même que pour la variable « logique ». Nous présentons ci-dessous l'analyse descriptive et inférentielle de cette variable.

<i>Tableau des moyennes</i>	<i>Moderne Début</i>	<i>Classique Début</i>	<i>Moderne Fin</i>	<i>Classique Fin</i>	<i>Moyennes</i>
Favorisé	4,471	3,914	6,076	5,110	4,893
Défavorisé	3,805	3,057	6,257	5,062	4,545
Moyennes	4,138	3,486	6,167	5,086	4,719

<i>Source de variation</i>	<i>ddl</i>	<i>SC</i>	<i>CM</i>	<i>F</i>	<i>p</i>
<i>S<M*P>*T</i>	167	711,779	4,262		
M	1	5,075	5,075	1,040	ns
P	1	31,547	31,547	6,466	<.025
T	1	138,249	138,249	80,729	<.0005
M.P	1	0,461	0,461	0,094	ns
M.T	1	7,209	7,209	4,209	ns
P.T	1	1,929	1,929	1,126	ns
M.P.T	1	0,004	0,004	0,002	ns
S(G)	80	390,306	4,879		
S(G).T	80	137,000	1,713		

Commentaires :

Effet de M. D'un point de vue descriptif, on observe une légère différence des moyennes en faveur des sujets du milieu favorisé. Cependant cette différence n'est pas significative ($F_M[1,80]=1,040$; ns). On ne peut donc pas dire que la performance varie en fonction du milieu.

Effet de P. On observe sur cette variable une performance plus importante pour les sujets ayant reçu une pédagogie moderne que pour les sujets ayant reçu une pédagogie classique. Cet effet est significatif ($F_P[1,80]=6,466$; $p<.025$). On peut donc dire que la pédagogie moderne favorise la performance à cette épreuve de combinatoire.

Effet de T. Comme avec l'épreuve de logique, on retrouve un effet d'apprentissage dans cette épreuve, les performances en fin d'année étant bien meilleures qu'en début ($F_T[1,80]=80,729$; $p<.005$)

Effet de M.P. Dans cette épreuve, comme dans la précédente, l'effet de la pédagogie est le même dans les deux milieux ($F_{M.P}[1,80]=0,094$; ns)

Effet de M.T. L'effet d'apprentissage reste le même dans les deux milieux ($F_{M.T}[1,80]=4,209$; ns).

Effet de P.T. L'effet d'apprentissage est le même pour les deux pédagogies ($F_{P.T}[1,80]=1,126$; ns).

Conclusion. Des trois conditions énumérées dans les hypothèses pour considérer que la pédagogie moderne est préférable à la pédagogie classique, deux sont réunies. La pédagogie moderne permet d'obtenir de meilleurs résultats sur l'épreuve de combinatoire (effet

de P) et ne génère pas plus de différence sociale (pas d'effet M.P). On peut donc dire qu'en ce qui concerne l'épreuve de combinatoire, la pédagogie moderne est préférable.

Analyse des résultats sur la variable "Probabilités".

Comme précédemment, le jeu d'hypothèses sur l'effet de chacun des facteurs est le même que pour la variable « logique ». Nous présentons ci-dessous l'analyse descriptive et inférentielle de cette variable.

<i>Tableau des moyennes</i>	<i>Moderne Début</i>	<i>Classique Début</i>	<i>Moderne Fin</i>	<i>Classique Fin</i>	<i>Moyennes</i>
Favorisé	5,681	6,000	6,838	6,743	6,315
Défavorisé	6,081	5,338	6,667	5,371	5,864
Moyennes	5,881	5,669	6,752	6,057	6,090

<i>Source de variation</i>	<i>ddl</i>	<i>SC</i>	<i>CM</i>	<i>F</i>	<i>p</i>
<i>S<M*P>*T</i>	<i>167</i>	<i>589,073</i>	<i>3,527</i>		
M	1	8,550	8,550	1,707	ns
P	1	8,641	8,641	1,725	ns
T	1	16,657	16,657	9,932	<.005
M.P	1	13,430	13,430	2,681	ns
M.T	1	4,307	4,307	2,568	ns
P.T	1	2,453	2,453	1,463	ns
M.P.T	1	0,050	0,050	0,030	ns
S(G)	80	400,817	5,010		
S(G).T	80	134,168	1,677		

Commentaires.

Effet de M. Comme précédemment, on observe une tendance chez les sujets de milieu favorisé à avoir de meilleurs résultats sur cette variable. Cette différence n'est cependant pas significative ($F_{M[1,80]}=1,707$; ns). On ne peut donc pas dire que le milieu influence la performance.

Effet de P. Descriptivement, les résultats des sujets ayant eu une pédagogie moderne sont supérieurs à ceux qui ont reçu une pédagogie classique. Cependant cette différence n'est pas significative ($F_{P[1,80]}=1,725$; ns). Les deux types de pédagogie conduisent donc à des performances similaires dans cette épreuve.

Effet de T. On retrouve, comme pour les autres épreuves un effet de la répétition de mesures ($F_{T[1,80]}=9,932$; $p<.005$) qui va dans le sens d'une meilleure performance entre en T2 qu'en T1. Il y a donc bien un effet d'apprentissage entre les deux tests.

Nous passerons rapidement sur les effets d'interaction qui sont tous non significatifs, comme dans les autres épreuves (les commentaires seront donc les mêmes).

Conclusion : dans cette troisième épreuve, on ne peut pas dire que la pédagogie moderne apporte quelque chose de plus que la pédagogie classique. Elle ne favorise pas la performance des sujets (absence d'effet de P et de P.T) quel que soit le milieu (absence d'interaction M.P)

L'inférence fiduciaire.

Principes et Définition.

Dans l'inférence statistique, on teste l'hypothèse d'une différence des moyennes dans la population parente égale à 0 ($\delta=0$). Dans l'analyse fiduciaire, on cherche :

- Soit la probabilité que δ soit supérieur à une valeur donnée, considérée comme notable.
- Soit la probabilité que δ soit inférieur à une valeur donnée, considérée comme négligeable.

Il n'y a pas de règles strictes pour déterminer les valeurs notables et négligeables d'un point de vue descriptif. Cela dépend du type de données traitées. Pour cela, au niveau inductif, on a recours à la distribution de δ étant donnée la valeur de d_{obs} qu'on appelle la distribution fiduciaire Δ

Méthodologie.

Au niveau descriptif, on observe une différence positive ou négative, qu'on peut juger négligeable ou notable, intuitivement ou à l'aide de critères tel que ceux de Rouanet (1993) qui consiste à prendre la différence des moyennes rapportée à l'écart-type des différences individuelles en considérant les deux règles suivantes :

- Si $|d|/s > 0,6 \Rightarrow$ différence notable.
- Si $|d|/s < 0,4 \Rightarrow$ différence négligeable.

Au niveau de l'inférence, on aura pu se prononcer sur la significativité de la différence.

On a donc 4 cas :

Au niveau descriptif, la différence est

	Au niveau descriptif, la différence est	
Au niveau inférentiel la différence est	Notable	Négligeable
Significative	A	B
Non significative	C	D

Dans les cas C et D, Il n'y a pas d'effet notable possible au niveau inductif. On peut observer un effet inductivement négligeable dans les 4 cas.

Deux règles sont à respecter :

- On ne cherche à faire une conclusion sur un effet notable que si l'effet est notable descriptivement.
- Le test de signification est une étape obligatoire dans la recherche d'un effet notable.

Méthodologie de l'inférence fiduciaire

Les explications suivantes ne concernent que les comparaisons à un ddl au numérateur sur une variable numérique (comparaison de deux moyennes). Le point de départ de l'inférence fiduciaire est le résultat d'un test de signification (t ou F).

Test de la notabilité d'une différence.

- Calculer le dénominateur du t de Student⁹ : $e=d_{\text{obs}}/t$
- Calculer s en appliquant la formule selon le cas :

⁹ Si vous partez du F, il suffit d'en calculer la racine carrée pour obtenir t.

- Groupes indépendants : $s = e / \sqrt{1/n + 1/n}$
- Groupes appariés : $s = e \sqrt{n}$
- Juger de la notabilité au niveau descriptif en vérifiant que : $|d_{\text{obs}}|/s > 0,6$.
- Choisir une garantie γ . Lire la table de l'inférence fiduciaire à la ligne correspondant aux ddl du dénominateur du t de Student puis appliquer la formule qui y est indiquée.
- Si le résultat est supérieur à 0,6s, alors l'effet peut être déclaré inductivement notable avec une garantie γ . ce qui ne veut pas dire qu'il soit négligeable.

Exemple.

- Dans l'exemple du cours, $d_{\text{obs}}=3,81$, $F_{[1,75]}=71,01$, $s= 1,61$ et $t_{\text{obs}}=8,43$. Au niveau descriptif, l'effet est notable car $3,81/1,61= 2,36$ et $2,36 > 0,6$.
- On a donc : $e=3,81/8,43 = 0,452$
- On a 75 ddl au dénominateur du t, on doit donc lire la table à la ligne 80 de la première colonne : $d_{\text{obs}}-2,37$ e
- On a choisi la partie gauche de la table parce que la différence observée est positive : 3,81. Si la différence avait été négative, nous aurions regardé la partie droite.
- On a choisi la garantie γ la plus importante permise par la table, soit $\gamma=0,99$.
- On calcule ensuite la quantité $d_{\text{obs}}-2,37$ e = 2,74. Cette différence peut être considérée comme notable puisque $1,61*0,6 < 2,74$. Si cela n'avait pas été le cas, nous n'aurions pas pu conclure à la notabilité au niveau inductif de d_{obs} pour ce seuil de garantie
- Dès lors que d_{obs} est supérieure à la quantité ainsi calculée, alors on peut la déclarer notable avec une garantie de 0,99 (la garantie γ qu'on a choisie). Dans le cas contraire, on ne peut pas la déclarer notable, ce qui ne veut pas dire qu'elle est négligeable. Dans notre exemple, on a bien d_{obs} supérieure à la quantité calculée. L'effet peut donc être déclaré notable au niveau inductif.

Test de la négligeabilité d'un effet.

- Juger de la négligeabilité de l'effet au niveau descriptif en vérifiant que : $|d_{\text{obs}}|/s < 0,4$ ¹⁰
- Choisir la table correspondant à la garantie γ souhaitée.
- Chercher dans la table la valeur correspondant au nombre de ddl du dénominateur de t_{obs} (en ligne) et à la valeur absolue de t_{obs} (en colonne)¹¹..
- Calculer ensuite $|d| = \text{val} * e$ où val est la valeur lue dans la table
- Si $|d| < 0,4s$, alors l'effet peut être déclaré inductivement négligeable avec une garantie de γ . Dans le cas contraire, on conclura qu'il n'est pas négligeable, ce qui ne veut pas dire qu'il soit notable.

Exemple

Dans cet exemple, la valeur absolue de t est de 1,12 et le dénominateur du F a 30 degrés de liberté (32 sujets-2 groupes indépendants). On a donc $d_{\text{obs}}=0,625$; $e= 0,558$ et $s = 1,58$.

- Négligeabilité de la différence : $0,625/1,58 = 0,395 < 0,4$
- La garantie γ est fixée à .90.
- Dans la table, on cherche en tête de colonne la valeur 1,12. Elle n'existe pas, mais peut être encadrée par 1,10 et 1,20. Dans ces deux colonnes, on cherche la ligne correspondant à 30 ddl. On y lit les valeurs suivantes : 2,41 et 2,51. On applique ensuite une règle de trois : $(1,1-1,12)/(1,1-1,2)=0,2$. Il faut donc ajouter à la valeur trouvée pour 1,10 (soit 2,41) les deux dixièmes de l'écart entre 2,41 et 2,51 soit : $0,2(2,51-2,41)=0,02$. Pour une valeur absolue de $t_{\text{obs}}=1,12$, une table plus précise nous donnerait : $2,41+0,02=2,41+0,1*0,2=2,43$
- sachant que $e= d_{\text{obs}}/t$ soit $e=0,625/1,12=0,558$

¹⁰ La valeur de s sera calculée comme dans le test de notabilité.

¹¹ Si on ne la trouve pas, on cherche les deux valeurs l'encadrant et l'on applique une règle de 3 pour trouver la valeur correspondant à notre t_{obs} .

- $|d| = 2,43 \cdot 0,558 = 1,36$
- On calcule ensuite le rapport de $|d|/s = 1,36/1,58 = 0,86$. On peut conclure que l'effet n'est pas négligeable puisqu'il représente 86 % de s .

Exercices.

Exercice 1

L'analyse de variance sur la variable logique de l'exemple de référence montre un effet significatif du facteur « milieu » ($F_M[1,80]=9,914$; $p<.005$) en faveur du milieu favorisé ($m_{m1}=4,724$; $m_{m2}=3,582$). Cet effet est-il notable pour une garantie de .95 ? Pour une garantie de .90 ?

Exercice 2.

Dans le même exemple de référence, l'analyse de variance met en évidence un effet du facteur « tâches » ($F_T[1,80] = 57,1$; $p<.0005$) en faveur de la tâche en fin d'année ($m_{t1}=3,585$; $m_{t2}=4,721$). Cet effet est-il notable pour une garantie de .99 ?

Exercice 3.

Toujours dans l'exemple de référence, l'analyse de variance de la variable « logique » montre que le groupe de sujets ayant eu une pédagogie moderne ($m_{p1}=4,508$) réussit mieux l'épreuve que le groupe de sujets ayant eu une pédagogie classique ($m_{p2}=3,798$). Cependant, l'effet de P n'est pas significatif ($F_P[1,80]=3,842$; ns). Est-il pour autant négligeable pour garantie de .95.

Corrigés :**Corrigé de l'exercice 1**

Les données de départ sont $F_M[1,80]=9,914$; $p<.005$ et $m_{m1}=4,724$; $m_{m2}=3,582$. On peut donc calculer :

$$d_{obs} = 4,724 - 3,582 = 1,142$$

$$t_{obs} = \sqrt{F} = \sqrt{9,914} = 3,149$$

$$e = d_{obs}/t_{obs} = 1,142/3,149 = 0,363$$

Les groupes sont indépendants, on a donc :

$$s = e \sqrt{1/n_1 + 1/n_2} = 0,363 * \sqrt{1/42 + 1/42} = 1,662$$

Test de la notabilité avec $\gamma=.95$

- Notabilité descriptive : $1,142/1,662=0,687$.
- Lecture de la table des différences positives, colonne $\gamma=.95$, ligne 80: $d_{obs}-1,66 e$
- Application de la formule : $1,142-1,66*0,363 = 0,54$
- Conclusion : $0,54 < 0,6s$ donc l'effet de M n'est pas notable avec une garantie de .95

Test de la notabilité avec $\gamma=.90$

- Notabilité descriptive : $1,142/1,662=0,687$ (comme précédemment).
- Lecture de la table des différences positives, colonne $\gamma=.90$, ligne 80: $d_{obs}-1,29 e$
- Application de la formule : $1,142-1,29*0,363 = 0,67$
- Conclusion : $0,67 < 0,6s$ donc l'effet de M n'est pas notable avec une garantie de .90

Corrigé de l'exercice 2.

Les données de départ sont $F_T[1,80]=57,00$; $p<.0005$ et $m_{m1}=3,585$; $m_{m2}=4,721$. On peut donc calculer :

$$d_{obs} = 3,585 - 4,721 = -1,136$$

$$t_{obs} = \sqrt{F} = \sqrt{57,100} = 7,556$$

$$e = d_{obs}/t_{obs} = -1,136/7,556 = -0,150$$

Les groupes sont appariés, on a donc :

$$s = e \sqrt{n} = 0,150 \sqrt{84} = 1,378$$

- Notabilité descriptive : $1,136/1,378=0,824$.
- Lecture de la table des différences négatives colonne $\gamma=.99$, ligne 80: $d_{obs}+2,237 e$
- Application de la formule : $-1,136+2,37*0,363 = -0,78$.
- Conclusion : $0,78 < 0,6s$ donc l'effet de T n'est pas notable avec une garantie de .99.

Corrigé de l'exercice 3

On cherche à tester la négligeabilité de la différence. Les données du problème sont : $m_{p1}=4,508$; $m_{p2}=3,798$; $F_P[1,80]=3,842$; ns. On peut donc calculer :

$$d_{obs} = 4,508 - 3,798 = 0,710$$

$$t = \sqrt{F} = \sqrt{3,842} = 1,960$$

$$e = d_{obs}/t = 0,710/1,960 = 0,362$$

$$s = e \sqrt{1/n_1 + 1/n_2} = 0,362 * \sqrt{1/42 + 1/42} = 1,660$$

- Négligeabilité descriptive : $0,710/1,660=0,428$. On est dans la zone d'incertitude (ni négligeable, ni notable descriptivement).
- Lecture de la table avec $\gamma=.95$, ligne 60, colonne 1,90 et colonne 2,00. On trouve les valeurs 3,57 et 3,67. On peut donc calculer la valeur correspondant à $t=1,96$: $((1,90-1,96)/(1,90-2)*(3,67-3,57))+3,57=3,63$
- Application de la formule : $3,63 * e = 3,63*0,362=1,31$
- Conclusion : $|d|/s = 1,31/1,660 = 0,79$. On ne peut donc pas conclure que la différence est négligeable.

Foire aux questions.

Voici une petite sélection de vos questions sur le cours.

Que signifient les notions d'effet et de sens de l'effet ?

La notion d'effet concerne l'existence d'une différence (est-elle significative ou non). Le sens de l'effet porte sur la direction des différences de moyennes (dans quel ordre peuvent être rangées les moyennes).

Comment fait-on pour lire la table du F de Snédecor ?

Pour lire la table du F de Snédecor, vous devez connaître le nombre de degrés de liberté (ddl) au numérateur et au dénominateur. Vous cherchez la colonne correspondante au nombre de ddl du numérateur. Puis dans cette colonne, vous cherchez la ligne correspondant au nombre de ddl du dénominateur. Vous trouvez à l'intersection les différentes valeurs de correspondant aux différentes proportions d'échantillons pour lesquels la valeur de F est dépassée (seuil). Si la valeur observée est supérieure à la valeur de la table, vous pouvez conclure que le F est significatif au seuil correspondant.

Pourquoi doit-on faire une comparaison inter et intragroupes ?

Le F de Snédecor est un rapport entre une variation systématique (comparaison inter groupes) et une variation aléatoire (comparaison intra groupe). Cet indice est construit sur le même principe que le t de Student. Ce dernier est en effet un rapport entre une différence (variation intergroupes) et un estimateur de l'écart-type de la population parente, soit donc une variation intragroupes.

Dans quelle partie du cours trouve-t-on les explications concernant le calcul des différents ddl ?

Vous trouverez le détail des calculs des ddl est synthétisé à la page 628 et dans le formulaire du présent document. Les différents plans peuvent en effet être considérés comme des sous plan de $S \times G \times T$ Ainsi :

$S \times G$ se décompose en G et S(G) que vous retrouverez en tant que tel à la page 628.

$S \times T$ se décompose en T et S.T. Pour ce dernier, il faut considérer qu'on n'a qu'un seul groupe. Le calcul des ddl sera le même que pour S(G).T avec $K=1$, soit $(J-1) \times (L-1)$.

$S \times A \times B$ est un cas particulier de $S \times G$ où G est le croisement de deux facteurs. Il se décompose en A, B et S(G). Pour A et B le nombre de ddl se calcule comme pour G. Pour S(G), il faut considérer que le nombre de groupe est égal au produit des nombres de modalités pour A et B.

Pourquoi dit-on que le t de Student est une comparaison à un ddl alors qu'on ne regarde jamais la table pour un ddl?

Dans le t de Student, la comparaison est toujours à un degré de liberté. Quand on dit cela, on parle des ddl du numérateur. Ce qu'on cherche dans la table c'est la ligne correspondant aux ddl du dénominateur (en général $n-1$). Tout comme le F, le t est un rapport entre une variation intergroupe (la différence des moyennes) et une variation intragroupes ou intersujets (l'écart type). Le fait que la variation intergroupe est toujours un seul ddl dans le t de Student conduit à ne tenir compte que des ddl du dénominateur dans la lecture de la table.

Si F est significatif, est-il nécessaire de déterminer le sens de l'effet en calculant la différence des moyennes ?

Il est, bien entendu, important de décrire l'effet mis en évidence par l'analyse de variance. Dire qu'un facteur a un effet n'a aucun intérêt d'un point de vue psychologique si on ne dit pas quel est cet effet et notamment dans quel sens il va.

Peut-on indiquer le nombre de ddl au numérateur et au dénominateur du t de Student comme on le fait parfois pour le F ?

Non, l'indication des ddl du numérateur et du dénominateur du F a pour but d'indiquer ce qu'on lit dans la table. Pour le t, cette indication n'est pas nécessaire, le numérateur a toujours 1 ddl et le dénominateur est de n-1.

Doit-on réviser le point de vue vectoriel et algébrique pour l'examen ?

Point de vue algébrique et point de vue vectoriel font partie du cours que vous devez connaître, même si les sujets d'examen consistent le plus souvent en une mise en œuvre des méthodes d'analyse.

Comment formuler les conclusions pour qu'elles soient les plus explicites possibles ?

La conclusion dans une analyse statistique comporte toujours deux aspects :

- La conclusion statistique. Elle est, dans un premier temps, descriptive. Cette conclusion porte sur le sens de l'effet, c'est-à-dire quel est le sens des différences, dans quel groupe ou condition obtient-on une meilleure performance etc. On passe ensuite au niveau inférentiel, c'est-à-dire celui de la généralisabilité de l'effet observé (mon résultat est-il significatif ou non). Selon le point de vue adopté, on formulera la conclusion en termes d'homogénéité ou de typicité (approche combinatoire) ou bien en termes de conservation ou de rejet de l'hypothèse nulle (approche fréquentiste). Le troisième niveau est celui de l'inférence inductive. Il porte sur l'importance de l'effet, c'est-à-dire de la négligeabilité ou de la notabilité de l'effet observé.
- La conclusion psychologique. Cette conclusion relève de l'interprétation des résultats. Il faut donc mettre en relation les résultats des analyses aux différents niveaux et la problématique de l'expérience, c'est-à-dire les hypothèses. Concrètement cela revient à se demander en quoi ce qu'on a observé nous renseigne sur le fonctionnement psychologique. Par exemple si on compare l'effet de deux types d'apprentissage sur la réussite à un examen et qu'on observe dans le premier que la performance est meilleure que dans le second (conclusion descriptive). Que par ailleurs au niveau inférentiel les deux groupes ne sont pas homogènes (conclusion inférentielle du point de vue combinatoire) et que l'effet est notable (conclusion inductive), on pourra conclure que la première forme d'apprentissage est plus efficace que la seconde (ou plus conforme aux conditions de développement si les conditions d'apprentissage sont liées à des hypothèses sur le développement).

Bibliographie et liens

Plus de cours ?

Howell D.C. (1998) Méthodes statistiques en sciences humaines, Bruxelles : De Boeck Université, Collection Méthodes en sciences humaines.

Rouanet H., Le Roux B., Bert M.C. (1987) Statistique en sciences humaines, procédures naturelles, Paris, Dunod.

Plus d'entraînement ?

Gueguen N. (2001) Statistique pour psychologue : Cours et exercices, Paris, Dunod, collection " Psycho sup ".

Rouanet H., Le Roux B. (1995) Statistiques en sciences humaines : Exercices et solutions, Paris, Dunod, collection " Psycho sup ".

Et des liens...

Encore des cours (en français) !

Intitiation aux méthodes statistiques :

<http://ibm2.cicrp.jussieu.fr/grasland/STAT98/STAT98.htm>

SEL : <http://www.inrialpes.fr/sel/telecharger.html>

Statnet : <http://www.agro-montpellier.fr/cnam-lr/statnet/>

Mas11 :

http://nte-serveur.univ-lyon1.fr/nte/immediato/math2002/Mass11/cours/cours_table.htm

Fiches de statistique : http://hpa.free.fr/Fiches_de_Stat.htm

Encore des cours (en anglais)!

HyperStat Online : <http://davidmlane.com/hyperstat/index.html>

The Statistics Homepage : <http://www.statsoft.com/textbook/stathome.html>

StatNotes: <http://www2.chass.ncsu.edu/garson/pa765/statnote.htm>

Introductory Statistics: <http://www.psychstat.smsu.edu/introbook/sbk00.htm>

Et si l'ordinateur vous aidait ?

StatView 5 Demos & Downloads : <http://www.statview.com/product/demo.shtml>

Formulaire.

Statistiques	Formules
Comparaison de groupe appariés	$t_{obs} = \frac{(d - \delta_0)}{s / \sqrt{n}}$
Comparaison de groupes indépendants	$t_{obs} = \frac{m \ominus m''}{s \sqrt{\frac{1}{n \ominus} + \frac{1}{n''}}}$
Point de vue de l'hypothèse nulle	$SC_G = \sum_{g=1}^k n_g (m^g - m)^2 = \sum_{g=1}^k \frac{T_g^2}{n_g} - \frac{T^2}{n}$
Point de vue des contrastes	$SC_C = \frac{\left(\sum_{g=1}^k c_g m^g \right)^2}{\sum_{g=1}^k \frac{c_g^2}{n_g}}$
F de Snédécór	$\frac{CM_{inter}}{CM_{intra}}$

Sources de variation	Degrés de liberté	Somme des carrés	CM	F
S _{J<G>}	(J*K)-1	$SC_{S<G>*T} = \sum (x - m)^2 = \sum x^2 - \frac{(\sum x)^2}{n}$		
G	K-1	$SC_G = \sum_{g=1}^k n_g (m^g - m)^2 = \sum_{g=1}^k \frac{T_g^2}{n_g} - \frac{T^2}{n}$	$\frac{SC_G}{ddl_G}$	$\frac{CM_G}{CM_{S(G)}}$
S(G)	(J-1)K	$SC_{S(G)} = \sum_{g=1}^k \sum_{s=1}^{n_g} (x^{s<g>} - m^g)^2 = \sum_{j=1}^j \frac{T_j^2}{n_j} - \frac{T^2}{n_g}$	$\frac{SC_{S(G)}}{ddl_{S(G)}}$	

Sources de variation	Degrés de liberté	Somme des carrés	CM	F
S _{J*T_L}	(J*L)-1	$SC_{S<G>*T} = \sum (x - m)^2 = \sum x^2 - \frac{(\sum x)^2}{n}$		
T	L-1	$SC_T = \sum_{t=1}^k n_t (m^t - m)^2 = \sum_{t=1}^k \frac{T_t^2}{n_t} - \frac{T^2}{n}$	$\frac{SC_T}{ddl_T}$	$\frac{CM_T}{CM_{S,T}}$
S.T	(J-1)(L-1)	$SC_{S,T} = SC_{S*T} - SC_T$	$\frac{SC_{S,T}}{ddl_{S,T}}$	

Sources de variation	Degrés de liberté	Somme des carrés	CM	F
$\underline{S}_J < A_K * B_L >$	$(J * K - 1 * L) - 1$	$SC_{\underline{S} < A * B >} = \sum (x - m)^2 = \sum_{i=1}^n x^2 - \frac{(\sum x)^2}{n}$		
A	K-1	$SC_A = \sum_{a=1}^k n_a (m^a - m)^2 = \sum_{a=1}^k \frac{T_a^2}{n_a} - \frac{T^2}{n}$	$\frac{SC_A}{ddl_A}$	$\frac{CM_A}{CM_{\underline{S}(G)}}$
B	L-1	$SC_B = \sum_{b=1}^l n_b (m^b - m)^2 = \sum_{b=1}^l \frac{T_b^2}{n_b} - \frac{T^2}{n}$	$\frac{SC_T}{ddl_T}$	$\frac{CM_T}{CM_{\underline{S}(G)}}$
A.B	$(K-1)(L-1)$	$SC_{A.B} = \sum_{ab=1}^l n_{ab} (m^{ab} - m)^2 = \sum_{ab=1}^l \frac{T_{ab}^2}{n_{ab}} - \frac{T^2}{n}$	$\frac{SC_{A.B}}{ddl_{A.B}}$	$\frac{CM_{A.B}}{CM_{\underline{S}(G)}}$
S(G)	$(J-1)(K * L)$	$SC_{\underline{S}(G)} = \sum_{g=1}^k \sum_{s=1}^{n_g} (x^{s < g >} - m^g)^2 = \sum_{j=1}^j \frac{T_j^2}{n_j} - \frac{T_g^2}{n_g}$	$\frac{SC_{\underline{S}(G)}}{ddl_{\underline{S}(G)}}$	

Sources de variation	Degrés de liberté	Somme des carrés	CM	F
$S_J < G_K > * T_L$	$(J * L) - 1$	$SC_{\underline{S} < G > * T} = \sum (x - m)^2 = \sum x^2 - \frac{(\sum x)^2}{n}$		
G	K-1	$SC_G = \sum_{g=1}^k n_g (m^g - m)^2 = \sum_{g=1}^k \frac{T_g^2}{n_g} - \frac{T^2}{n}$	$\frac{SC_G}{ddl_G}$	$\frac{CM_G}{CM_{\underline{S}(G)}}$
T	L-1	$SC_T = \sum_{t=1}^l n_t (m^t - m)^2 = \sum_{t=1}^l \frac{T_t^2}{n_t} - \frac{T^2}{n}$	$\frac{SC_T}{ddl_T}$	$\frac{CM_T}{CM_{\underline{S}(G), T}}$
G.T	$(K-1)(L-1)$	$SC_{G.T} = SC_{G * T} - SC_G - SC_T$ $= \left(\sum \frac{x_{gt}^2}{n_{gt}} - \frac{T^2}{n} \right) - SC_G - SC_T$	$\frac{SC_{G.T}}{ddl_{G.T}}$	$\frac{CM_{G.T}}{CM_{\underline{S}(G), T}}$
S(G)	$(J-1)K$	$SC_{\underline{S}(G)} = \sum_{g=1}^k \sum_{s=1}^{n_g} (x^{s < g >} - m^g)^2 = \sum_{j=1}^j \frac{T_j^2}{n_j} - \frac{T_g^2}{n_g}$	$\frac{SC_{\underline{S}(G)}}{ddl_{\underline{S}(G)}}$	
S(G).T	$(J-1)K * (L-1)$	$SG_{\underline{S}(G), T} = SC_{\underline{S} < G > * T} - SC_G - SC_T - SC_{G.T} - SC_{\underline{S}(G)}$	$\frac{SC_{\underline{S}(G), T}}{ddl_{\underline{S}(G), T}}$	

Inférence Fiduciaire

Calcul de e: cas de groupes appariés	$e = s / \sqrt{n} = d_{obs} / t_{obs} = d_{obs} / \sqrt{F_T}$
Calcul de e: cas de groupes indépendants	$e = s \sqrt{1/n^{\text{C}} + 1/n^{\text{n}}} = d_{obs} / t_{obs} = d_{obs} / \sqrt{F_G}$
Recherche d'un effet notable	$Val_{seuil} = d_{obs} - Val_{table} * e$
Recherche d'un effet négligeable	$ d = val_{table} * e$

Annexe 1: Données de l'exemple de référence

Une expérience porte sur 84 sujets répartis par moitié dans un milieu social favorisé et dans un milieu social défavorisé. Dans chacun de ces deux groupes, on distingue deux types de pédagogie (moderne et traditionnelle). Les élèves ont passé trois épreuves notées de 0 à 10, la première fois en début d'année, et la seconde fois en fin d'année : une épreuve de combinatoire (Comb.), une épreuve de probabilité (Prob.) et une épreuve de logique (Log.). D'après Rouanet et Leroux (1993).

<i>S</i>	<i>G</i>	<i>M</i>	<i>P</i>	<i>Log</i>		<i>Comb</i>		<i>Prob</i>	
				<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>	<i>T1</i>	<i>T2</i>
1G1	Favorisé	moderne		0,8	3,4	2	3,8	2,9	5,8
2G1	Favorisé	moderne		2,8	3,7	1,9	1,7	4,6	6,1
3G1	Favorisé	moderne		3,6	4,9	1,5	5	5,7	4,9
4G1	Favorisé	moderne		1,3	1,7	4,9	5,5	3,8	7,5
5G1	Favorisé	moderne		2	4,6	2,7	6,7	2,1	0,3
6G1	Favorisé	moderne		3,8	4,7	4	6	5,2	7
7G1	Favorisé	moderne		2,9	5,5	1,4	3,9	3,3	8,1
8G1	Favorisé	moderne		4,8	3,8	4,6	7,8	5,3	7,9
9G1	Favorisé	moderne		4,8	4,6	3,4	6,2	6,6	9,6
10G1	Favorisé	moderne		2,3	6	2,4	8,9	5,8	5,3
11G1	Favorisé	moderne		4	6	4,9	5,8	5,5	6,5
12G1	Favorisé	moderne		3,2	8,1	2,3	6,4	5,1	5,8
13G1	Favorisé	moderne		2,5	4,8	6,8	6,3	7,6	9
14G1	Favorisé	moderne		5	6,8	6,6	4,9	6,5	3,7
15G1	Favorisé	moderne		6,1	7,5	6,6	5,7	5,7	6,6
16G1	Favorisé	moderne		7,6	8	6	6,8	6,4	8,3
17G1	Favorisé	moderne		6,1	5,2	6,9	6,3	7,1	8,3
18G1	Favorisé	moderne		3,8	6,6	9,3	9,1	7,7	7,4
19G1	Favorisé	moderne		8	7,2	6,2	7,2	6,8	9,6
20G1	Favorisé	moderne		6,1	8,4	2,5	6,6	6,6	7,5
21G1	Favorisé	moderne		8	8,4	7	7	9	8,4
22G1	Favorisé	classique		8,1	8,4	4,7	5,1	8,2	9,7
23G2	Favorisé	classique		1,7	3,8	2,9	2,9	4	4,2
24G2	Favorisé	classique		1,8	1,9	1,7	4,9	3	3,7
25G2	Favorisé	classique		1,8	2,8	3,8	4,1	7,8	5,5
26G2	Favorisé	classique		1,7	3,5	4,4	4,7	5,1	6,1
27G2	Favorisé	classique		4,4	5,5	2,5	3,1	5,6	7
28G2	Favorisé	classique		3,8	2,4	5,4	7,1	6,6	8,6
29G2	Favorisé	classique		3,8	4,1	2,4	6,2	4,6	5,3
30G2	Favorisé	classique		2,9	5	4,3	7,5	6,9	4,8
31G2	Favorisé	classique		4,6	4,3	5,1	5,2	6,5	9,7
32G2	Favorisé	classique		3,9	5,6	3,4	7	5,5	8,2
33G2	Favorisé	classique		4,7	6,6	3,2	2,6	6,5	8,2
34G2	Favorisé	classique		3,3	2,9	2,9	5	6	7,3
35G2	Favorisé	classique		3,1	4,8	4	3,9	5,3	6,7
36G2	Favorisé	classique		6,5	5,5	4,1	4,9	5,3	9,3
37G2	Favorisé	classique		1,8	3,7	3,8	5,1	4,3	4,2
38G2	Favorisé	classique		3,5	6,2	4,1	5,1	7,4	1,8

39G2	Favorisé	classique	6,7	5,9	5,1	8,1	7,1	8
40G2	Favorisé	classique	5	8,1	6,7	5,6	7,1	7,3
41G2	Favorisé	classique	4,4	6,2	3	4,7	7,1	7,7
42G2	Favorisé	classique	5,4	7,3	4,7	4,5	6,1	8,3
43G2	Défavorisé	moderne	7	8,5	5,3	9,9	7,1	6,6
44G3	Défavorisé	moderne	1,5	0	3,8	3,4	5,1	7,7
45G3	Défavorisé	moderne	2,4	3,2	1,6	2,6	5,2	0,8
46G3	Défavorisé	moderne	0,9	2,5	1,9	4,9	4,9	5,9
47G3	Défavorisé	moderne	0,9	3,3	4,3	6,6	4,5	6,2
48G3	Défavorisé	moderne	3	1,9	3,1	6,4	6,2	6,3
49G3	Défavorisé	moderne	2,8	3,2	3,5	1,8	5,7	3,4
50G3	Défavorisé	moderne	4,6	3,7	1,1	3,8	5,4	6,8
51G3	Défavorisé	moderne	5,5	7,7	4,1	6,6	6,6	7,4
52G3	Défavorisé	moderne	3,3	3,2	2,6	5,5	7,1	6,5
53G3	Défavorisé	moderne	1,5	4	4,3	6,6	5,6	5,4
54G3	Défavorisé	moderne	3,2	3,6	6	9,5	6,4	8,9
55G3	Défavorisé	moderne	3,3	5,9	2,5	9	5,6	7,9
56G3	Défavorisé	moderne	2,6	5,1	4,3	9,7	5,6	6
57G3	Défavorisé	moderne	5	4,9	3,4	9,2	6,6	8,1
58G3	Défavorisé	moderne	4,2	7,3	2,4	4,7	5,4	6,5
59G3	Défavorisé	moderne	2,2	3,4	3,5	4,7	8,3	8,2
60G3	Défavorisé	moderne	4,5	5,1	8,2	8	8,1	9,1
61G3	Défavorisé	moderne	5,3	6,4	2,5	4	4,9	7,9
62G3	Défavorisé	moderne	5,4	7,2	7,2	9,3	8,1	9
63G3	Défavorisé	moderne	4	6,1	4,3	5,2	5,3	5,4
64G4	Défavorisé	classique	2,1	0,8	0,9	3	3,3	0,5
65G4	Défavorisé	classique	1,6	3	2,4	2,5	3,6	1,8
66G4	Défavorisé	classique	1,1	2	3	3,3	6,5	4,2
67G4	Défavorisé	classique	1,8	2,7	3,6	5,9	4,4	4,3
68G4	Défavorisé	classique	3,5	3,1	1,1	4,9	4,1	3
69G4	Défavorisé	classique	1,2	1,4	2,6	3,1	2,8	1,8
70G4	Défavorisé	classique	2,2	1,8	4,4	4	3	6,8
71G4	Défavorisé	classique	2,9	3,1	3,7	5	4,1	3,9
72G4	Défavorisé	classique	2	3,7	3,7	6,9	6,4	5,4
73G4	Défavorisé	classique	2	3,6	2,9	3,8	5	7,2
74G4	Défavorisé	classique	3,3	1,9	2,6	8,7	6,9	7,8
75G4	Défavorisé	classique	3,1	3	3,5	5,8	5,1	5,9
76G4	Défavorisé	classique	3,1	4,4	2	2,8	4,7	4
77G4	Défavorisé	classique	4,8	6	2,4	6,8	6,4	6,3
78G4	Défavorisé	classique	1,8	3,5	2,9	3,1	5,8	5,3
79G4	Défavorisé	classique	1,9	4,3	4	5,1	9,3	7,9
80G4	Défavorisé	classique	3,3	5,1	1,6	4	5,8	7,2
81G4	Défavorisé	classique	1,7	6	4,6	7,6	6,4	8,1
82G4	Défavorisé	classique	2	3,4	1,4	6,2	4,8	6,7
83G4	Défavorisé	classique	6,2	6,4	6,7	8,8	8,6	7,6
84G4	Défavorisé	classique	4	6,8	4,2	5	5,1	7,1