

Université de Franche-comté
Année Universitaire ... / ...

Licence de psychologie
3^e Année

STATISTIQUES
Recueil et Traitement des Données
JYUCP64 - Travaux dirigés

TD1

DEBUTER AVEC SPSS

Objectifs

- Ouvrir SPSS
- Saisir, ouvrir, importer/exporter et sauvegarder un fichier de données
- Transformer des données, rechercher des données
- Comprendre comment sont codées et organisées les variables sous SPSS
- Se familiariser avec les fenêtres et les menus du logiciel SPSS

Exercice préliminaire : création d'un dossier de travail et d'un fichier de données sous Excel

- A) **Sur le bureau**, créer un dossier de travail. Ouvrir un fichier Excel et constituer 4 colonnes de données factices, comme dans le tableau suivant :

a	b	c	d
1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	7
5	6	7	8
6	7	8	9
7	8	9	10
8	9	10	11
9	10	11	12
10	11	12	13

- B) Enregistrer ces données au format .xls (format Excel¹), sous le nom Test1TD1.

Exercice 1 : Ouvrir SPSS

- Dans la **fenêtre d'accueil**, tester l'option "ouverture d'un fichier existant de données" afin d'importer les données de Test1TD1.xls (cocher : "Lire le nom des variables à partir de la 1^{ère} ligne de données").

Exercice 2 : Entrer des données dans SPSS

- A) Exercice de copier/couper/coller **dans SPSS**.

1	2	1
1	3	2
1	4	3
1	5	4
1	6	5
1	7	6
1	8	7
1	9	8
1	10	9
1	11	10
2	4	3
2	5	4
2	6	5
2	7	6
2	8	7
2	9	8
2	10	9
2	11	10
2	12	11
2	13	12

On supposera que ces 40 scores factices ont été recueillis sur deux échantillons indépendants de 10 hommes et de 10 femmes ayant subis chacun deux traitements expérimentaux (à un intervalle d'une semaine). Par exemple, les données pourraient refléter le nombre de mots rappelés librement parmi une liste de 20 mots, 10 minutes après la dictée (la première colonne de scores est intitulée *semaine1*) et une semaine après (la seconde colonne de scores est intitulée *semaine2*) pour 10

¹ Le format Excel est lisible par SPSS, ainsi que d'autres formats tels que .csv (comma-separated values, utilisant les virgules pour séparer les colonnes de chiffres) ou .txt (11 texte avec colonnes séparées par des tabulations).

hommes codés 1 et 10 femmes codées 2 dans la première colonne. Réorganiser les colonnes comme montré ci-dessus (on fait en sorte que les femmes mémorisent en moyenne plus de mots que les hommes et que la mémorisation des mots se dégrade en moyenne avec le temps).

Règle d'or dans SPSS: un seul sujet par ligne. Cette règle nécessite d'utiliser une variable catégorielle nommée *genre* pour distinguer les hommes des femmes dans la première colonne (codant 1 pour les hommes et 2 pour les femmes). Supprimer les colonnes ou les lignes devenues inutiles en cliquant leur tête (partie grisée), puis presser supprimer (aucun point de doit apparaître dans les cellules vides; les cellules vides doivent être parfaitement blanches, sans quoi le point représente une donnée manquante pour une variable). Le tableau comporte au final 3 colonnes représentant respectivement les variables *genre*, *semaine1*, et *semaine2*. Les 20 lignes représentent les 20 sujets.

- B) Afin d'améliorer la lisibilité des données, 1) Recoder la variable de genre en mesure nominale et spécifier que les entrées sont de type numérique dans la fenêtre d'Affichage des variables (cliquer sur les cellules d'une ligne pour afficher les options) 2) En utilisant des étiquettes de valeurs dans la cellule "Valeurs", associer aux valeurs 1 et 2 servant à coder le *genre* les étiquettes "Homme" et "Femme". Retourner dans la fenêtre des données et cocher "Étiquettes de valeur" dans le menu "Affichage" afin de voir apparaître ces étiquettes dans le tableau 3) Vérifier que les entrées des autres colonnes sont au format numérique. 4) Pour chacune des variables, si nécessaire, ajuster le nombre de décimales et la largeur des colonnes, puis 5) Donner une étiquette et un nom différents aux variables (e.g., *Semaine 1*, pour l'étiquette et *semaine1* pour le nom de la colonne). Le nom sert uniquement de repère dans le tableur et ne doit pas dépasser 8 lettres, alors que l'étiquette est plus libre et sera utilisée dans les tableaux d'analyses et les figures.

- C) Faire une analyse de la moyenne du nombre de mots mémorisés par semaine et par genre, en sélectionnant *Analyse/Rapports/Récapitulatif_des_observations/*. Décocher "Afficher les observations" et choisir *Moyennes* dans le sous menu *Statistiques*. Penser *i)* au raccourci ctrl+A, pour sélectionner toutes les variables (*A = "All"*) *ii)* au maintien du bouton gauche de la souris pour sélectionner un ensemble de variables *iii)* les touches majuscule et de contrôle, pour sélectionner un sous-ensembles de variables. Noter les effets marginaux et les effets principaux des variables.

Exercice 3 : Transformer et rechercher des données dans SPSS

- A) On suppose que lorsque les données ont été entrées, un mot qui servait d'exemple ou d'échauffement a été comptabilisé à tort. Retirez 1, à toute les valeurs en utilisant dans le menu la fonction *Transformer/Calculer*. Deux nouvelles variables sont définies, ce qui n'est pas utile dans le cas de corrections d'erreurs. Faites en sorte de ne conserver que 2 colonnes correspondant aux valeurs corrigées.

- B) Noter que la variable indépendante "SEMAINE" n'est représentée dans SPSS qu'indirectement par ses modalités (*semaine1* et *semaine2*). Remarquer que l'ensemble des 40 scores représente la variable dépendante "Nombre de mots mémorisés". Imaginons qu'on mesure également le nombre de mots recouverts au cours des 10 premières secondes. Construire artificiellement le nombre de mots recouverts au cours des 10 premières secondes en calculant systématiquement 75% des mots recouverts. Calculer cette nouvelle variable en utilisant la fonction *Transformer/Calculer* et la fonction RND, signifiant "to round", arrondir. Par exemple, 11 sera transformé en 8 car $11 \times .75 = 8.25$ et $RND(8.25) = 8$. Comment représenter et identifier dans le tableau à la fois le nombre de mots mémorisés et le nombre de mots mémorisés en moins de 10 secondes, en Semaine 1 et en Semaine 2 ? Réorganiser le nom des variables et les étiquettes afin de clarifier les futures analyses.

- C) **Repérage de données absurdes** : Remplacer quelques valeurs en leur ajoutant une centaine. Ces valeurs pourraient correspondre à des données entrées de façon incorrecte (fautes de frappes ou 'coquilles'). Imaginer un tableau avec une centaine de valeurs, dont la relecture serait proscrite. En utilisant le *.../Récapitulatif_des_observations/*, demandez le calcul de maximum et minimum pour repérer ces valeurs absurdes pour chacune des variables. Afin de rechercher l'emplacement des valeurs absurdes, choisir la fonction "SI..." dans *Transformer/Calculer* et composer de nouvelles variables en choisissant d'attribuer aux nouvelles variables la valeur 1 si le score des hommes et des femmes est supérieur à 100. Ceci permet le repérage de données absurdes sur un large ensemble de

données (ce n'est pas le cas ici, mais ça pourrait l'être si nous avions un tableau d'un millier de lignes).

- D) **Repérage de données manquantes** : Afin de trouver les données manquantes, faire une analyse des effectifs par variable (faisable par différentes fonctions...). L'exercice n'a d'intérêt que si on imagine que le tableau de données n'était pas visualisable rapidement en raison de sa grandeur.

Exercice 4 : Travailler sur un sous-ensemble de données

- A) Utiliser *Données/Sélectionner des observations* et la fonction "Selon une condition logique...SI" afin de travailler sur un sous-ensemble de données. Sélectionner par exemple les hommes.

- B) Afin d'éliminer certains résultats des analyses, sans les effacer pour autant, créer une colonne de 1 avec quelques zéros répartis au hasard et nommer la variable "filtre". Les zéros indiquent les sujets qui ne seront pas pris en compte dans les analyses SPSS subséquentes. Utiliser à nouveau *Données/Sélectionner des observations* et utiliser votre nouvelle variable filtre afin de calculer les nouvelles moyennes sur l'échantillon tronqué. L'abandon de la variable filtre permet de travailler à nouveau sur l'échantillon original. De cette façon, les scores du tableau original ne sont jamais effacés.

Exercice 5 : Tableaux de contingence

Le tableau de contingence suivant montre le lien existant entre le genre et l'approbation de programmes télévisés à caractère violents.

Genre	Approbation	
	Non	Oui
Femme	30	20
Homme	10	40

Entrer ce tableau dans SPSS en utilisant deux variables catégorielles et une colonne d'effectifs. Utiliser la fonction *Données/Pondérer les observations* afin de déclarer la colonne d'effectifs comme contenant effectivement des effectifs. Sans cela, le nombre 30 représente une mesure unique (par exemple, 30 centimètres), au lieu de 30 mesures séparées (par exemple, 30 femmes répondant par la négative). Une fois la pondération faite, SPSS ne renvoie aucun message particulier, mais sans cette opération, le calcul d'un χ^2 n'est pas possible. Vous pouvez passer à la suite sans calculer le χ^2 .

Sauvegarde

- Sauvegarder en utilisant un nom pertinent, court, sans accents, et numéroté (e.g., TD1_Ex5). Le format .sav étant spécifique à SPSS, une sauvegarde sous un format plus traditionnel (de type Excel) est également nécessaire dans le cas d'une expiration de la licence SPSS ou de la migration vers un logiciel concurrent.

Nota bene. Pour ne pas perdre le bénéfice du travail effectué dans un groupe et éviter des problèmes dus aux absences, il vaut mieux que chacun des étudiants sauvegarde les résultats sous clé USB, plutôt que sur la portion de disque nominative (Z) correspondant au compte de l'étudiant identifié par la machine. Le disque commun (Y) n'offre aucune garantie, car il est publique (n'importe qui peut effacer les fichiers d'un autre groupe). Conseil : faire une sauvegarde multiple (disque Z + disque Y + clé).

TD2

STATISTIQUES DESCRIPTIVES avec SPSS

Objectifs

- Explorer, décrire et résumer les données avec SPSS
- Copier les résultats figurant dans des tableaux ou figures dans un fichier Word
- Utiliser les normes de l'APA pour les figures et les tableaux

Remarques préliminaires

Les résultats d'analyses sous SPSS sont créés dans une fenêtre différente de celle des données. Le fichier de résultat peut être sauvegardé indépendamment des données. Les résultats peuvent être édités et exportés dans un éditeur de texte. Dans une publication, les résultats peuvent être présentés sous forme de tableaux ou de figures, accompagnés de légendes, et d'un texte s'y rapportant.

Exercice 1. Données

- A) Entrer le tableau suivant dans SPSS et sauvegarder le fichier sous le nom TD2. Préalablement, dans la colonne *Valeurs* de la fenêtre *Affichage de variables* et pour les deux variables catégorielles, ajouter les correspondances suivantes : 1 = "Groupe O", 2 = "Groupe A", 3 = "Groupe B", 4 = "Groupe AB", et 1 = "Masculin", 2 = "Féminin". En sélectionnant *Affichage/Etiquettes de valeurs* dans la fenêtre des données, il ne reste plus qu'à entrer ces chiffres dans la fenêtre de données pour voir apparaître la chaîne de caractère correspondante. B) Vérifier la présence de données manquantes en calculant les effectifs.

groupeS	sexe	taille	poids	groupeS	sexe	taille	poids
Groupe O	Masculin	178	75	Groupe O	Féminin	163	60
Groupe O	Masculin	196	100	Groupe O	Féminin	142	51
Groupe A	Masculin	145	60	Groupe A	Féminin	150	55
Groupe O	Masculin	170	71	Groupe O	Féminin	165	64
Groupe B	Masculin	180	80	Groupe A	Féminin	160	53
Groupe O	Masculin	175	69	Groupe O	Féminin	175	50
Groupe AB	Masculin	185	78	Groupe O	Féminin	182	72
Groupe A	Masculin	190	90	Groupe B	Féminin	169	65
Groupe O	Masculin	183	70	Groupe O	Féminin	162	62
Groupe B	Masculin	182	85	Groupe B	Féminin	182	80
Groupe A	Masculin	170	72	Groupe O	Féminin	165	67
Groupe O	Masculin	160	77	Groupe A	Féminin	171	50
Groupe O	Masculin	170	95	Groupe O	Féminin	146	55
Groupe B	Masculin	172	68	Groupe AB	Féminin	151	48
Groupe B	Masculin	190	120	Groupe O	Féminin	164	59
Groupe O	Masculin	180	75	Groupe B	Féminin	176	71

Fichier Word

- Ouvrir un fichier Word dans lequel vous donnerez un compte rendu détaillé des résultats qui seront obtenus dans les exercices suivants. Les tableaux et les figures seront numérotées et insérées dans un texte clair les décrivant tour à tour (e.g., Le Tableau 1 indique les effectifs par groupe sanguin tout sexe confondu. On remarque que le Groupe O est le plus présent, suivi des groupes ..., ou encore, La Figure 1 montre la répartition des effectifs selon les groupes sanguins tout sexe confondu...). Dans un compte rendu, pour une analyse donnée, éviter d'utiliser à la fois une figure ET un tableau. Choisir le format le plus adéquat.

Exercice 2. Tableaux et Figures résumés

- A) Afin de générer des tableaux au format de l'APA, sélectionner *Edition/Options.../Tableaux*, et sélectionner le format *Academic #1* ou *Academic VGA*. Pour satisfaire pleinement au style APA, aucune ligne verticale ne doit constituer un tableau. Votre choix sera effectif au moment où vous demanderez à SPSS d'analyser vos données.

- B) En utilisant la fonction *Analyse/Statistiques descriptives/Fréquences*, produire le tableau et la figure de la distribution des effectifs pour le groupe sanguin (choisir "Diagrammes en bâtons" et non l'histogramme dans les options, les diagrammes étant plus adaptés aux variables catégorielles). Editer le tableau en double cliquant dessus : changer par exemple le nombre de décimales dans les cellules en cliquant droit sur une sélection, effacer des colonnes ou des informations peu utiles. Citer les effectifs

cumulés dans le texte Word (A, B et AB combinés, par opposition à l'effectif du groupe O). Editer dans SPSS la figure en double-cliquant dessus, afin d'effacer le titre (qui sera ajouté dans Word), remplacer la dénomination des axes si nécessaire. Remplacer le mot "Fréquences" par "Effectifs" dans la figure. Noter que l'ordre de présentation des barres d'effectifs dépend de l'ordre des étiquettes. Une présentation non arbitraire ordonnée, dans le but de clarifier la présentation, imposerait de changer la correspondance entre étiquettes et valeurs. Une solution plus simple consiste à changer l'ordre des barres dans l'éditeur de figures par la fonction "*Séries/Affichées*". Choisir un ordre de manière à ce que la distribution des effectifs soit ascendante. Comparer *copier/coller* et *Copier_Objet/Coller*, pour exporter la figure dans Word.

- C) Pour obtenir un tableau de contingence entre sexe et groupe sanguin, choisir la fonction *Analyse/Statistiques_descriptives/Tableaux_croisés*. Décrire ce tableau sans l'analyser statistiquement. Les étudiants en avance peuvent aller à la recherche de la fonction de calcul du chi-carré d'indépendance. Cette fois, changer l'ordre des colonnes ou des lignes est plus difficile, car il correspond à l'ordre des étiquettes. Pour obtenir un ordre satisfaisant dans une table, il faut recoder les étiquettes (pas nécessaire pour cet exercice). Double cliquer sur le tableau pour l'éditer et changer les doubles barres du tableau en simples barres dans les propriétés du tableau, afin qu'il corresponde au style de l'APA (cliquer droit sur le tableau pour obtenir ses propriétés ou aller dans le menu "Format", mais seulement après avoir double-cliqué sur le tableau!). *Optionnel: Il est possible de sauvegarder ce format et l'utiliser par défaut en paramétrant les options générales des tableaux dans la fenêtre des données*. Exporter ce tableau dans Word. Pour obtenir une distribution graphique similaire, utiliser *Graphes/Bâtons* et choisir l'option *Juxtaposé*. Définir ensuite les groupes de bâtons par groupe sanguin et choisir le sexe sur l'axe des abscisses.

- D) Afin d'obtenir des indices de tendance centrale et de dispersion pour la variable taille et le poids, tout sexe confondu, utiliser la commande la plus simple pour obtenir des statistiques descriptives (*Analyse/Statistiques_descriptives/Caractéristiques*). Cocher sauvegarder les scores standardisés pour obtenir les scores z . Ensuite, calculez vous-même les scores z de la taille en utilisant la fonction *Transformer/Calculer* afin de vérifier si les scores z donnés par SPSS sont corrects (noter la moyenne et l'écart-type sur un bout de papier, puis appliquez la formule $z = (x - \text{moy}(x))/s$). Ne rien exporter sous Word.

- E) Pour obtenir un histogramme simple de la taille, utiliser *Graphes/Histogramme...*, et choisir "Afficher la courbe Gaussienne" la courbe normale de référence. Insérer cet histogramme dans Word.

- F) Afin de savoir s'il existe un lien entre la taille et le groupe sanguin, choisir dans Diagrammes la fonction *Analyse/Statistiques_descriptives/Explorer*, cocher "Les deux", choisir "Histogrammes" et conserver l'option pour "Boîtes à moustaches". Observer la distribution de la taille en fonction du groupe sanguin. Noter que les histogrammes informent sur la forme des distributions et que les diagrammes en forme de boîtes donnent une bonne représentation des données asymétriques et des résultats extrêmes (en se fondant sur les rangs des données et la distance interquartile). Il est conseillé dans le texte Word d'insérer uniquement le tableau indiquant les moyennes et les écart-types, en fonction des groupes sanguins. Insérer les diagrammes en forme de boîtes dans Word.

- G) **Remarque sur les recoupements:** Certains calculs sont disponibles dans plusieurs menus dans SPSS. *i)* Retracer un histogramme pour la taille et calculer la médiane et les quartiles en utilisant le menu *Analyse/Statistiques_descriptives/Fréquences* (Cocher "avec courbe gaussienne"). *ii)* Recalculer les différences entre moyennes de tailles en fonction du groupe sanguin, en choisissant la fonction *Analyse/Comparer_les_moyennes*.

TD3

CORRÉLATIONS ET RÉGRESSION SIMPLE

Objectifs

- Décrire des distributions bivariées
- Calculer un coefficient de corrélation adapté et en comprendre les propriétés
- Comprendre la régression simple
- Résumer et importer les résultats dans Word

Exercice 1. Diagrammes de dispersion, corrélations Bravais-Pearson, Rho de Spearman

- A) Entrer le tableau suivant dans SPSS et sauvegarder le fichier sous le nom TD3. Le tableau indique les mesures de coordination œil-main et la compétence au tennis pour 10 sujets.

<i>Sujet</i>	<i>Coor</i>	<i>Comp</i>	<i>Sujet</i>	<i>Coor</i>	<i>Comp</i>
1	4	4	6	4	2
2	4	5	7	7	5
3	5	6	8	8	6
4	2	2	9	9	9
5	10	6	10	5	3

Légende: *Coor*, Coordination œil-main; *Comp*, compétence au tennis.

- B) Etablir un **diagramme de dispersion** de ces données et décrivez le nuage de points (*Graphes/Diagramme de dispersion...*). Ici, la VD est la compétence au Tennis (en fonction des hypothèses, notez que cela pourrait être le contraire). Editer le graphe si besoin.
- C) Pourquoi un examen visuel est nécessaire avant un calcul de corrélation?
- D) Calculer la corrélation Bravais-Pearson avec SPSS et reporter le résultat dans Word sous le format suivant: "On observe une corrélation entre ... et ... ($r = \text{valeur}$; $n = \text{valeur}$; $p < \text{valeur}$). Quel est le sens et l'amplitude de la corrélation?"
- E) Etude d'un **effet de cohorte** sur la corrélation: Attribuer aux sujets 3, 5, 7, 8, 9, le sexe masculin et aux autres sujets le sexe féminin. Tracer à nouveau le diagramme de corrélation en catégorisant les points par sexe en utilisant l'option *Définir les marques par*. Calculez les corrélations par groupe en sélectionnant les données dans le tableau (*Données/Sélectionnez les Observations*). Est-il possible d'obtenir une corrélation élevée par effet de cohorte alors que les corrélations sont faibles dans les groupes?
- F) Etablir une **corrélations de point bisérielle** afin de savoir s'il existe une corrélation entre le sexe et la compétence au tennis. Etablissez un diagramme de dispersion pour ce cas.
- G) Créer une nouvelle variable chaîne de caractères comprenant une modalité par sujet, de A à J. Ajouter un nouveau sujet, nommé K, ayant un score éloigné de coordination égal à 14 et une compétence de 15. Recalculer la corrélation Bravais-Pearson entre *Comp* et *Coor* et conclure quant à la présence de **résultats extrêmes** dans les données. Retracer le diagramme de dispersion en utilisant la nouvelle variable dans le champ de *Etiqueter les observations par* et en cochant dans les options *Afficher les diagrammes avec les étiquettes d'observation*.
- H) L'objectif est de choisir un coefficient de corrélation sur des données ordinales, afin d'éviter l'effet de scores extrêmes sur le coefficient de corrélation de Bravais-Pearson. Afin d'obtenir des rangs sur les 11 sujets, recoder les variables à la main, puis vérifier le calcul manuel en utilisant une fonction SPSS destinée à l'attribution de rangs aux données (de type "Ordonner les observations"). Vérifiez l'équivalence entre le Rhô de Spearman calculé directement sur les données non rangées et la corrélation Bravais-Pearson sur les rangs. Le Rhô de Spearman (ρ) doit être reporté sous le format suivant dans Word: $\rho = \text{ou } r_s = \text{valeur}$; $n = \text{valeur}$; $p < \text{valeur}$.
- I) **Transformer les données par une relation linéaire** en multipliant par 10 et en ajoutant 1 à la la coordination ($x' = 10x + 1$). Calculer les scores standardisés pour la compétence. Quel sera le résultat de la corrélation Bravais-Pearson entre ces deux nouvelles variable (justifier, puis vérifier votre prédiction en calculant la corrélation)?

Exercice 2. Régression simple

- A) L'objectif est de construire une **équation de régression**. Télécharger le fichier depuis <http://fabien.mathy.free.fr/>, rubrique *information/COURS/L3-MYUC7634*. Prendre garde de ne pas laisser SPSS considérer que la première ligne est le nom des variables (option à décocher). Nommer les colonnes de variables comme indiqué ci-dessous (S, Y, X) et attribuer aux étiquettes les valeurs suivantes: *Sujet*, *exFinal*, et *exSelect*, pour représenter respectivement les variables *Sujet*, *Note à l'examen final* et *Note à l'examen de sélection d'entrée*.

S	Y	X	S	Y	X	S	Y	X	S	Y	X
1	38	44	10	81	53	19	105	43	28	142	56
2	49	40	11	86	47	20	106	55	29	145	60
3	61	43	12	91	45	21	107	48	30	150	55
4	65	42	13	94	41	22	112	49	31	152	54
5	69	44	14	95	39	23	114	46	32	164	58
6	73	46	15	98	40	24	114	41	33	169	62
7	74	34	16	100	37	25	117	49	34	195	49
8	76	37	17	100	48	26	125	63			
9	78	41	18	103	48	27	140	52			

- B) Tracer le diagramme de dispersion. Ensuite, accéder à la fonction calculant les régressions linéaires simples à partir du menu *Analyse*. Choisir correctement la VD et la VI et transférer les noms de variable dans les zones appropriées. Cocher la case *Caractéristiques* dans l'option *Statistiques* et demandez une analyse des résidus *en cochant Diagnostique des observations*, avec la valeur 3 proposée par défaut. Qu'est ce qu'un **résidu**?
- C) Observer la sortie de la régression simple. Analyser notamment le tableau de diagnostic des observations. Si une donnée n'est pas fiable car elle conduit à un résidu standardisé à plus de 3 écart-types, relancer la procédure de régression sans le sujet concerné (l'élimination d'un sujet a été vu dans les TDs précédents; utiliser la procédure "*Si Sujet différent de 34*"). Quel est le pourcentage de variance expliquée par la note d'examen de sélection? Calculer la variance **résiduelle** $S^2_{\text{res}} = \frac{\sum (Y - \hat{Y})^2}{N - 2}$, à l'aide de la formule simplifiée $S^2_{\text{res}} = S_Y^2(1 - r^2) \frac{N - 1}{N - 2}$. Vérifier que la valeur trouvée correspond à celle donnée par SPSS.
- D) Indiquez la valeur *F* de l'analyse de variance pour la régression. Est-elle significative? Comment formule-t-on l'existence d'une relation linéaire significative dans un compte rendu? Tentez de recalculer *F* et R^2 à partir des CM et des SC du tableau ANOVA.
- E) Ecrivez l'équation de régression. Quelle serait la note à l'examen finale prédite par une note de sélection de 60? Le nuage de point confirme-t-il une relation linéaire? Apposer la droite de régression sur le nuage de points en choisissant *Graphiques/Options* dans l'éditeur de graphes, puis en cochant *Courbe d'ajustement*.

Subsidiaire

Si le temps le permet...

- A partir de l'exercice 1, établir une distribution univariée/partielle de la coordination dans un diagramme en bâton. Quel est le mode, quelle est la médiane et quelle est la forme de la distribution?
- A partir de l'exercice 1, reprendre le tableau de départ, et utiliser les fonction d'Excel afin de calculer la covariance entre les variables, les écarts-types et la corrélation. Vérifier si la covariance est divisée par *N* ou *N-1* dans Excel.

TD4

TESTS STATISTIQUES PARAMETRIQUES (OU NON) SUR 2 ECHANTILLONS PAIRES OU INDEPENDANTS

Objectifs

- Comprendre l'organisation des mesures indépendantes et paires dans la saisie des données sous SPSS.
- Comparer 2 échantillons, appariés ou indépendants.
- Comprendre les tests paramétriques z et t de Student.
- Décrire les distributions en utilisant une métrique ordinale (boîte à moustaches).
- Comprendre les tests non paramétriques de Wilcoxon et de Mann-Whitney.

Exercice 1. Données paires

1) Dans une expérimentation, on demande à 10 participants de regarder un point central sur un écran et de presser sur une touche lorsqu'un mot apparaît dans son champ périphérique. Chaque participant doit ensuite taper le mot qu'il vient d'identifier afin de contrôler si le mot a été effectivement reconnu. L'hypothèse expérimentale est que les mots représentés dans le champ visuel droit seront plus facilement reconnus que dans le champ gauche, parce que les mots présentés dans le champ droit sont traités par l'hémisphère gauche, majoritairement responsable du traitement de l'information verbale. Chaque cellule du tableau suivant représente le temps de réponse médian (en millisecondes) pour 40 mots pour un champ visuel donné. Mis à part le nom trop long des têtes de colonnes, le tableau suivant est au format SPSS. Entrer ce tableau dans SPSS.

Champ visuel gauche	Champ visuel droit
323	304
512	493
502	491
385	365
453	426
343	320
543	523
440	442
682	580
590	564

2) S'agit-il de mesures répétées ou indépendantes? Formaliser le plan d'expérience (à l'aide des symboles * et $\langle \rangle$). Le tableau précédent est-il le tableau de données brutes originel? A quoi devrait ressembler le tableau de données originel (combien de lignes, combien de colonnes, combien d'observations)?

3) a) Calculer manuellement (entendu, toute autre méthode qu'utiliser SPSS) un **t de Student** afin de savoir s'il existe une différence de moyenne significative entre les deux conditions expérimentales, avec la formule indiquée ci-dessous. La maîtrise des fonctions mathématiques (e.g., "=" somme(sélection de cellules)") et l'utilisation de calculs algébriques simples de type "=" cellule_1-cellule_2" associée au copier-coller sous Excel vous permettra d'assister vos calculs. Créer une variable d correspondant aux différences de scores et une variable $(d - \bar{d})^2$. b) Quelle est l'**erreur standard** des différences de moyennes pour ces 10 sujets? c) Quels sont les trois facteurs pouvant infléchir un test t ? Voici la formule de référence pour ce problème :

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

4) Vérifier le calcul manuel du t de student avec une fonction appropriée sous SPSS (*Analyse/Comparer les moyennes/etc.*). Le résultat est-il significatif? Décrire les résultats sous Word en écrivant le résultat du test au format APA².

5) Remplacer le score du sujet 9, ayant obtenu un score médian de 682 millisecondes, par un score de 1682 millisecondes. Un tel écart n'est pas très crédible pour 40 présentations par cellule, mais le serait si l'expérimentation avait consisté à présenter un seul mot par sujet et par champ visuel (Il suffirait que le sujet manque d'attention un court moment pour faire augmenter le score d'une seconde). Représentez les distributions de scores dans des boîtes à moustaches, en utilisant une fonction SPSS. a) Calculer à nouveau le t de Student avec SPSS. Quel impact le changement de score a-t-il sur le t de Student ? b) Remplacez le t de Student par un test non paramétrique approprié (*Analyse/Tests non paramétriques*). Obtient-on grâce à cette alternative une différence significative? Vérifier le test non paramétrique manuellement. Le format de citation d'un score z est : $z = \text{valeur}, p < \text{valeur}$.

➔ **Test de Wilcoxon** pour données paires (T ou W observé est la somme des rangs, pour les rangs correspondant au signe des différences le moins fréquent):

$$z_T = \frac{T_{obs} - \mu_T}{\sigma_T}, \text{ avec } \mu_T = \frac{N(N+1)}{4} \text{ et } \sigma_T = \sqrt{\frac{N(N+1)(2N+1)}{24}}.$$

Exercice 2. Données indépendantes

- 1) Imaginez une variante de l'expérimentation indiquée dans l'exercice 1, dans laquelle on procède à des mesures indépendantes. On considérera que les observations obtenues sont strictement identiques à celles données dans l'exercice 1. Recomposer le tableau de données pour SPSS (ne pas oublier de remplacer 1682 par 682) et formaliser le plan d'expérience.
- 2) Lancer Excel, et coller les deux échantillons de résultat dans Excel. Choisir *Options/Analyse de données*. Si cette option n'est pas disponible, il faut l'ajouter (appeler l'enseignant). Ensuite, lancer le **Test F d'égalité des variances** afin de vérifier si les variances des échantillons sont égales. Compléter les deux champs du haut en indiquant les cellules des deux échantillons. Le F observé doit être supérieur à 1. Si ce n'est pas le cas, les échantillons doivent être inversés dans les champs. Le test F de différences de variances est-il significatif ? Le format d'écriture du résultat est $F(ddl1, ddl2) = \text{valeur}, p < \text{valeur}$. En fonction de la réponse, choisir le test du t de Student approprié et calculer sa valeur manuellement. Ensuite, vérifier le calcul du t de Student avec SPSS.

➔ t de Student avec calcul de variance combinée, considérant que les variances sont égales :

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_C^2}{n_A} + \frac{S_C^2}{n_B}}}, \text{ avec } S_C^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A - 1) + (n_B - 1)} \text{ et } ddl = (n_A - 1) + (n_B - 1)$$

➔ t de Student avec formule de Welch, considérant que les variances sont différentes :

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

$$* \quad ddl = \frac{\left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)^2}{\frac{\left(\frac{S_A^2}{n_A}\right)^2}{n_A + 1} + \frac{\left(\frac{S_B^2}{n_B}\right)^2}{n_B + 1}} - 2$$

² Exemple : Les résultats indiquent que ..., $t(ddl) = \dots, p < \dots$ **Ou** : Les résultats indiquent que ..., $t(ddl) = \dots, NS$; les hypothèses étant testées de manière bilatérale par défaut, on ne le précise pas; il est également possible de donner la valeur exacte de p , ceci étant valable pour tous les tests)

3) Remplacez le t de Student par un test non paramétrique approprié. (*Analyse/Tests non paramétriques*). Vérifier le calcul manuellement. Afin d'accélérer le calcul manuel, ranger les données sous SPSS, puis calculer la somme T des rangs par groupe. Obtient-on une différence significative de moyennes grâce à cette alternative ?

➔ **Test de Mann-Whitney** pour données indépendantes (T_{--} ou W_{--} est la somme des rangs pour le groupe dont la somme des rangs est la plus faible; certains manuels utilisent U , le nombre de fois que les rangs dans l'autre groupe sont plus petits). Il est conseillé d'utiliser T .

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \text{ et } \sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$z_T = \frac{T_{obs} - \mu_T}{\sigma_T}$$

OU

$$U_{\text{observé}} = T_{\text{max}} - T_{\text{observé}}$$

$$T_{\text{max}} = T_{2\text{max}} = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} \text{ ou } T_{1\text{max}} = n_1 n_2 + \frac{n_1(n_1 + 1)}{2},$$

avec n_1 l'effectif de l'échantillon sur lequel est calculé et n_2 représentant l'effectif de l'autre échantillon.

$$\mu_U = \frac{n_1 n_2}{2} \text{ et } \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$z_U = \frac{U_{obs} - \mu_U}{\sigma_U}$$

TD5

ANOVA: Analyse de variance pour plans indépendants simples

Objectifs

- Calculer une ANOVA à la main, présenter les résultats d'une ANOVA
- Effectuer une ANOVA sur SPSS
- Effectuer des comparaisons multiples a priori ou a posteriori entre moyennes ou groupes de moyennes.

Exercice 1. ANOVA simple

- 1) Entrer les données suivantes dans SPSS, en tenant compte de l'organisation du plan expérimental décrit ci-dessous:

On étudie l'influence des conditions de mémorisation sur la reconnaissance des visages. On présente à 4 groupes de 10 sujets ($N = 40$) une série de 20 photos de visages. Lors de l'examen des visages, chaque sujet doit appliquer l'une des trois consignes suivantes: inférer les traits de personnalité correspondant au visage, décrire l'aspect physique, déterminer le sexe. Les sujets du groupe contrôle ne reçoivent aucune consigne particulière autre que simplement observer les visages. Cette phase d'apprentissage est suivie d'un test de reconnaissance dans laquelle on demande aux sujets de reconnaître les 20 visages parmi 40.

L'hypothèse globale est qu'il y a un effet des consignes sur le nombre de photos mémorisées, parce que le type d'encodage et la profondeur de l'encodage varie d'une consigne à l'autre.

Traits-Pers	Traits-Phys	Sexe	Contrôle
16	14	10	6
14	12	8	9
11	9	9	7
13	10	7	5
16	11	11	4
12	9	10	8
13	13	9	10
12	14	6	11
14	11	13	7
15	10	7	9

- 2) Créer une figure en boîtes à moustaches pour ces données.
- 3) Calculer l'ANOVA manuellement, remplir la table d'ANOVA, chercher la valeur critique, conclure.

Aide 1. Formules de calcul pour l'ANOVA simple.

$$SC_{Total} = \sum x^2 - \frac{(\sum x)^2}{N}$$

$$SC_{Traitements} = \frac{\sum T_j^2}{n} - \frac{(\sum x)^2}{N} \quad (n = \text{effectif par groupe})$$

$$SC_{Erreur} = SC_{Total} - SC_{Traitements}$$

$$F = \frac{CM_{Traitements}}{CM_{Erreur}}$$

Aide 2. Table d'ANOVA.

Source	ddl	SC	CM	F	Sig.
Traitements/Entre
Erreur/Dans		
Total			

Aide 3. Présentation des résultats.

- La valeur critique, à chercher dans la table du F , est notée $F_{.05}(ddl_numérateur, ddl_dénominateur) = \dots$

Cette valeur critique doit être comparée au F calculé. Les résultats du présent test F de l'ANOVA doivent être reportés de la façon suivante dans un texte: L'ANOVA montre/ne montre pas un effet global de la profondeur de l'encodage sur la mémorisation des 20 photos, $F(ddl_numérateur, ddl_dénominateur) = \text{valeur observée}$; $p < .05$ ou $p < .01$ ou $p < .001$ ou $p = \text{valeur exacte}$ ou NS.

- Dans la table d'ANOVA, ajouter une ou deux étoiles au F calculé en fonction de la significativité du résultat.

- 4) Vérifier vos résultats avec SPSS. Quelle est la valeur exacte de p cette fois-ci?

Exercice 2. Comparaisons multiples a priori orthogonales

Nous souhaitons tester les contrastes a priori et orthogonaux suivants sur les données précédentes:

La **première hypothèse spécifique** est que les participants des 3 premiers groupes auront des performances plus élevées que dans le groupe contrôle, parce que le traitement des visages est plus profond dans les groupes expérimentaux, en comparaison avec le groupe contrôle dans lequel le traitement des visages est superficiel.

($g1, g2, g3$ vs $g4$)

La **seconde hypothèse spécifique** est que les participants des 2 premiers groupes auront des performances plus élevées que celle du 3^e groupe, car ils réalisent un encodage verbal profond alors que ceux du troisième groupe opèrent une classification sommaire ne nécessitant pas de traitement verbal.

($g1, g2$ vs $g3$)

La **troisième hypothèse spécifique** est qu'inférer des traits de personnalité représente un traitement plus profond que décrire l'aspect physique d'un visage. Pour cette raison, les résultats du premier groupe seront plus élevés que ceux du 2nd.

($g1$ vs $g2$)

- 1) Indiquer les coefficients des contrastes pour ces trois hypothèses. La somme des coefficients pour chaque contraste doit être égale à zéro.

	Groupe				Somme
	Traits-Pers	Traits-Phys	Sexe	Cont.	
ψ_1	= 0 ?
ψ_2	= 0 ?
ψ_3	= 0 ?

- 2) Ces contrastes ont-ils une chance d'être tous orthogonaux, étant donné leur nombre total (3)?
 3) Vérifier l'orthogonalité dans la famille de contrastes en calculant pour chacune des paires de contrastes, le produit des coefficients par traitement. Si la somme des produits est égale à zéro, les contrastes de la paire considérée sont orthogonaux.

Exemple

	Traits-Pers	Traits-Phys	Sexe	Cont.	Somme
ψ_1	Coeff1	Coeff2	Coeff3	Coeff4	
ψ_2	Coeff1'	Coeff2'	Coeff3'	Coeff4'	
Produit:	Coeff1 × Coeff1'	Coeff2 × Coeff2'	Coeff3 × Coeff3'	Coeff4 × Coeff4'	= 0 ?

- 4) Construire la matrice des comparaisons et indiquer dans cette matrice les résultats calculés précédemment. Vérifier que la famille de comparaison est orthogonale.
- 5) Calculer les contrastes et les F associés:

$$SC_{\psi} = \frac{n(\sum C_k \bar{x}_k)^2}{\sum C_k^2} \quad F_{\psi} = \frac{CM_{\psi}}{CM_{Erreur}} = \frac{\frac{SC_{\psi}}{ddl}}{CM_{Erreur}} = \frac{\frac{SC_{\psi}}{1}}{CM_{Erreur}} = \frac{SC_{\psi}}{CM_{Erreur}}$$

- 6) Etant donné que $k-1$ contrastes sont additifs, ce qui est le cas ici, remplir la table ANOVA des contrastes, puis indiquer leur significativité après avoir trouvé les valeurs critiques.

Source	SC	ddl	CM	F
Traitements
ψ_1
ψ_2
ψ_3
Erreur	
Total		

- 7) Lancer à nouveau une ANOVA sur les données en entrant les coefficients pour les contrastes dans l'option "Contrastes...". Bien préciser les coefficients égaux à zéro. Dans les résultats indiqués par SPSS, les contrastes sont testés par un test t . Vérifier l'équivalence entre les valeurs t^2 et les F calculés précédemment. Vérifier que les significativités sont également équivalentes.

Exercice 3. Autres comparaisons multiples

1) Relancer l'ANOVA. Demander un test de Dunnett visant à comparer tous les groupes expérimentaux au groupe contrôle et un test de Tukey visant à établir toutes les comparaisons par paires, progressivement, en commençant par les plus éloignées. Noter les résultats afin de vérifier la correspondance avec les résultats obtenus en cours sur les mêmes données.

Exercice 4. Taille de l'effet expérimental

Calculer le rapport de corrélation η^2 (équivalent au R^2) pour ces données et conclure quand au pourcentage de variance lié aux traitements expérimentaux.

$$\eta^2 = \frac{SC_{Traitement}}{SC_{Total}}$$

Subsidiaire: Test non paramétrique

Effectuer un test Kruskal-Wallis (équivalent non paramétrique de l'ANOVA) pour les données indépendantes. La logique de ce test est fondée sur un classement préliminaire des scores, indépendamment du groupe d'appartenance. Ce test est-il nécessaire ici?

TD6

ANOVA: Analyse de variance pour plans factoriels

Objectifs

- Calculer une ANOVA factorielle à la main, présenter les résultats d'une ANOVA
- Effectuer une ANOVA factorielle sur SPSS
- Comprendre les effets simples, les effets globaux et les effets d'interaction.

Exercice 1. ANOVA factorielle à deux facteurs

- 1) D'après l'hypothèse d'Eysenck (1974), le niveau de rétention d'une liste de mots dépend de la profondeur de leur traitement dans la phase de mémorisation. Dans le cas d'un apprentissage incident/implicite, diverses profondeurs de traitement sont possibles: par exemple, si on demande à un sujet de lire une liste de mots et de se contenter de compter le nombre de lettres dans chacun d'eux, le niveau de traitement est plus bas que si on lui demande de trouver un adjectif modifiant le mot. Dans le cas d'une mémorisation intentionnelle, on peut s'attendre à un bon niveau de rétention résultant de l'effort volontaire. Une technique de mémorisation optimale pourrait être la formation d'une image mentale représentant le mot. Pour étudier le lien entre profondeur de traitement et taux de rétention, Eysenck a envisagé cinq conditions expérimentales: apprentissage *intentionnel*, formation d'*images*, création d'*adjectifs*, création de *rimes*, *addition* des lettres. Eysenck postule également un effet de l'âge du fait que le traitement est plus laborieux chez les jeunes. Non seulement les enfants devraient retenir plus de mots, mais la différence avec les sujets plus âgés devrait s'accroître avec la complexité du traitement requise dans les consignes, puisque les enfants jeunes déploient plus d'efforts pour les traitements complexes que les sujets plus âgés. Le tableau de résultats récoltés sur 100 sujets figure ci-dessous. Préparer la structure du tableau dans SPSS: entrer le nom des variables, leurs étiquettes et les valeurs des modalités (avec une correspondance numérique-chaîne, e.g., 1= 55-65 ans et 2 = 18-30 ans). Une fois le tableau validé par l'enseignant, télécharger le contenu du tableau depuis <http://fabien.mathy.free.fr/>, rubrique *information*, L3-MYUC7634, TD6.

	Conditions de rétention				
	Addition	Rimes	Adjectifs	Images	Intentionnel
Sujets âgés	9	7	11	12	10
	8	9	13	11	19
	6	6	8	16	14
	8	6	6	11	5
	10	6	14	9	10
	4	11	11	23	11
	6	6	13	12	14
	5	3	13	10	15
	7	8	10	19	11
	7	7	11	11	11
	<u>70</u>	<u>69</u>	<u>110</u>	<u>134</u>	<u>120</u>
Sujets jeunes	8	10	14	20	21
	6	7	11	16	19
	4	8	18	16	17
	6	10	14	15	15
	7	4	13	18	22
	6	7	22	16	16
	5	10	17	20	22
	7	6	16	22	22
	9	7	12	14	18
	7	7	11	19	21
	<u>65</u>	<u>76</u>	<u>148</u>	<u>176</u>	<u>193</u>

- 2) Formaliser le plan expérimental.
- 3) Produire la figure de résultats la plus adaptée à cette étude avec les outils graphiques SPSS. Respecter le format APA (réfléchir à une solution en noir et blanc).
- 4) Calculer l'ANOVA factorielle manuellement pour ces données. Construire la table ANOVA, rechercher les valeurs critiques, décrire les effets globaux et d'interaction présents. Noter si les résultats sont significatifs à .05 ou à .01 à l'aides d'étoiles assignées aux F calculés.
- 5) Vérifier vos calculs par SPSS. Choisir une ANOVA factorielle dans le menu Modèle linéaire général/Univarié³. Dans les options, choisir "Overall" ou "Complet". Demander à nouveau une figure avec l'option "Diagrammes", puis choisir les variables avec soin et cliquer sur "ajouter" afin que votre choix soit pris en compte.
- 6) Calculer tous les effets simples ainsi que leur significativité et les inscrire dans une seconde table d'ANOVA.
- 7) Insérer tous vos résultats dans Word et effectuer un compte rendu.

Formules

$$SC_{totale} = SC_{centre} + SC_{dans}$$

(ou, $SC_{totale} = SC_{inter} + SC_{intra}$; ou encore $SC_{totale} = SC_{traitement} + SC_{erreur}$)

$$SC_{centre} = SC_{cellules} = SC_{lignes} + SC_{colonnes} + SC_{interaction}$$

$$SC_{total} = \sum x^2 - \frac{(\sum x)^2}{N}$$

$$SC_{cellules} = \frac{\sum T_{cel}^2}{n_{cel}} - \frac{(\sum x)^2}{N}$$

$$SC_{lignes} = \frac{\sum T_{lig}^2}{n_{lig}} - \frac{(\sum x)^2}{N}$$

$$SC_{colonnes} = \frac{\sum T_{col}^2}{n_{col}} - \frac{(\sum x)^2}{N}$$

$$F = CM / CM_{intra}$$

³ L'ANOVA est un cas particulier des **modèles linéaires généraux**, dans lesquels les données sont expliquées par une combinaison linéaire de variables. Le modèle est **univarié** car il ne comporte qu'une VD. Une analyse prenant en compte plusieurs VD simultanément est dite **multivariée**.

TD7

ANOVA: Analyse de variance pour mesures répétées

Objectifs

- Calculer une ANOVA pour mesures répétées à la main et en présenter les résultats.
- Effectuer une ANOVA pour mesures répétées sur SPSS.

Exercice 1. ANOVA pour mesures répétées simples

Dans une expérience sur les techniques graphiques, on demande à chacun des 10 participants experts dans l'art du dessin de reproduire un portrait en utilisant successivement un crayon, un pinceau ou un feutre (ordre contrebalancé). L'esthétisme des dessins, évalué par un groupe de 10 juges apparaît dans le tableau ci-dessous (une note plus élevée correspondant à un esthétisme plus élevé). Chaque juge évalue les trois dessins d'un seul participant.

Crayon	Pinceau	Feutre
10	12	14
18	10	16
20	15	16
12	10	12
19	20	21
25	22	20
18	16	17
22	18	18
17	14	12
23	20	18

- 1) Télécharger le tableau depuis <http://fabien.mathy.free.fr/>, rubrique *Information/Cours*, pour l'insérer dans SPSS et Excel.
- 2) Calculer manuellement l'ANOVA pour ces données et présentez les résultats dans un tableau d'ANOVA. Est-ce que le type d'instrument utilisé affecte l'évaluation qu'un dessin reçoit ? Indiquer également le pourcentage de variance expliquée par le facteur Instrument.

Formules (à lire en lignes)

$$SC_{total} = \sum x^2 - \frac{(\sum x)^2}{N} \quad \rightarrow \quad SC_{traitements} = \frac{\sum T_a^2}{n} - \frac{(\sum x)^2}{N}$$

$$SC_{sujets} = \frac{\sum T_s^2}{a} - \frac{(\sum x)^2}{N} \quad \quad \quad SC_{erreur} = SC_{total} - SC_{traitements} - SC_{sujets}$$

- Test F : CM_{erreur} au dénominateur;

- Force de l'effet expérimental : $\eta^2_{partiel} = \frac{SC_{Traitements}}{SC_{Traitements} + SC_{Erreur}}$

- 3) Afin de lancer une ANOVA intra-sujets dans SPSS, choisir "Modèle linéaire général", puis "Mesures répétées". Nommer le facteur dans la zone "Nom du critère intra-sujet". Dans la zone "Nombre de niveaux", entrer le nombre de conditions expérimentales pour le facteur. Ensuite, cliquer sur *ajouter*, puis, *Définir*. Dans la fenêtre suivante, transférer les variables dans la zone "Variables intra-Sujets" afin de faire disparaître les points d'interrogation. Demander un diagramme des données et, en option, demander un tableau de statistiques descriptives et une estimation de la taille de l'effet. Dans les données volumineuses fournies par SPSS, retrouver les résultats calculés à la main.

Remarque: Pour la lecture de la significativité du test F , tenir compte du résultat du test de

sphéricité. Si le test de sphéricité n'est pas significatif, il indique que les covariances entre les conditions ne sont pas hétérogènes (cela signifiant qu'il existe des corrélations similaires entre les paires de conditions, prouvant à indiquer que les différences individuelles sont constantes entre conditions). Si le test de sphéricité⁵ n'est pas significatif, lire la ligne "Sphéricité supposée" dans la table d'ANOVA.

- 4) En imaginant qu'on souhaite calculer toutes les comparaisons par paires, calculer le taux d'erreur par comparaison par la procédure de Bonferroni, à partir d'un taux d'erreur par famille de .05. Lancer ensuite tous les *t* de Student possibles et trouver les paires significatives.

Exercice 2. ANOVA pour mesures répétées complexes

Une expérience est conçue pour étudier l'influence de la solidité des formes dans leur détection. L'expérience manipule la forme comme facteur secondaire (cercle, carré ou triangle) et la solidité comme facteur primaire (pleine vs simple contour). La VD est le nombre d'erreurs commises en détectant les formes. Les sujets sont confrontés à toutes les combinaisons de traitements. Les résultats sont présentés dans le tableau ci-dessous (Les données sont également présentes dans le fichier téléchargé dans l'exercice 1). Les six colonnes indiquent dans l'ordre les résultats pour dix sujets, pour les conditions *cercplei*, *cerccont*, *carrplei*, *carrcont*, *triplei*, *triacont*.

4	2	2	8	7	5
3	6	2	6	8	9
2	10	2	5	5	3
1	8	5	5	2	9
4	6	4	5	5	10
3	6	4	6	9	12
7	12	2	6	4	8
6	10	9	5	0	10
4	5	7	6	8	12
2	12	12	8	10	12

- 1) Quel est le plan d'expérience?
- 2) Calculer manuellement l'ANOVA et compléter le tableau d'ANOVA indiqué ci-dessous, puis conclure au regard des hypothèses de départ.

Aide

<i>Source</i>	<i>ddl</i>	<i>SC</i>	<i>CM = SC/ddl</i>	<i>F</i>
Inter-Sujets	9	109.33		
Forme (A)	2	46.03		
Erreur(S × A)	18			
Solidité (B)	1	117.6		
Erreur(S×B)	9			
Interaction (A × B)	2	23.7		
Erreur(S × A × B)	18			
Total	59	606.33		

⁵ Le problème de sphéricité est hors programme.

Formules (à lire en lignes)

$$SC_{total} = \sum x^2 - \frac{(\sum x)^2}{N} \xrightarrow{\text{sens de lecture}} SC_{traitements} = \frac{\sum T_{Cellule}^2}{n_{Cellule}} - \frac{(\sum x)^2}{N}$$

$$SC_A = \frac{\sum T_a^2}{n_a} - \frac{(\sum x)^2}{N} \qquad SC_B = \frac{\sum T_b^2}{n_b} - \frac{(\sum x)^2}{N}$$

$$SC_{sujets} = \frac{\sum T_s^2}{a \times b} - \frac{(\sum x)^2}{N} \qquad SC_{erreur} = SC_{total} - SC_{traitements} - SC_{sujets}$$

$$SC_{A \times B} = SC_{traitements} - SC_A - SC_B \qquad SC_{sujets \times A \times B} = SC_{erreur} - SC_{sujets \times A} - SC_{sujets \times B}$$

- Pour calculer l'interaction $SC_{sujets \times A}$, construire le tableau $sujets \times A$ afin de calculer les totaux pour chacune des cellules, puis, calculer :

$$SC_{cellules(sujets \times A)} = \frac{\sum T_{cellule}^2}{b} - \frac{(\sum x)^2}{N}$$

Ensuite, considérer que: $SC_{cellules(sujets \times A)} = SC_{sujets} + SC_A + SC_{sujets \times A}$ et en déduire $SC_{sujets \times A}$.

- Procéder de même pour $SC_{sujets \times B}$.

- Ensuite, considérer que: $SC_{erreur} = SC_{sujets \times A} + SC_{sujets \times B} + SC_{sujets \times A \times B}$ et en déduire l'interaction $SC_{sujets \times A \times B}$.

- Tests F : Chacun des facteurs et leur interaction ayant une interaction propre avec le facteur sujet, le calcul des F pour chacun des effets comporte un terme d'erreur spécifique.

$$F = \frac{CM_A}{CM_{sujets \times A}} \qquad F = \frac{CM_B}{CM_{sujets \times B}} \qquad F = \frac{CM_{A \times B}}{CM_{sujets \times A \times B}}$$

- 3) Saisir les données dans SPSS. Comme pour l'exercice précédent, lancer une ANOVA pour mesures répétées (définir 2 facteurs, *Forme* et *Solidité*, et choisir les options identiques à l'exercice 1), puis vérifier vos calculs.

TD8

ANOVA: Analyse de variance pour plans mixtes

Objectifs

- Calculer manuellement une ANOVA pour plans mixtes et vérifier les calculs avec SPSS
- Présenter les résultats de l'ANOVA.

Exercice 1. ANOVA pour un plan mixte (un facteur fixe intra et un facteur inter)

Une expérience est conçue pour explorer l'hypothèse selon laquelle les étudiants ingénieurs ont un sens plus développé des formes géométriques que les étudiants de psychologie. Trois formes (triangle, carré, et rectangle) sont présentées dans des conditions sous optimales à des étudiants psychologues ou ingénieurs. Chacune des trois formes est présentée à chaque sujet 50 fois. Les sujets ont 3 secondes pour choisir la forme perçue sur un clavier à trois touches. On mesure le nombre de formes correctement identifiées. Les résultats de l'expérience sont présentés dans le Tableau ci-dessous. Les cellules sont téléchargeables sur <http://fabien.mathy.free.fr/>, rubrique *Information/Cours*.

Psychologie			Ingénierie		
Triangle	Carré	Rectangle	Triangle	Carré	Rectangle
2	12	7	13	3	35
8	10	9	21	4	30
4	15	3	26	10	35
6	9	7	22	8	30
9	13	8	20	9	28
7	14	8	19	8	27

- 1) Explorer rapidement les résultats à l'aide de boîtes à moustaches (Sélectionner *Juxtaposé* et *Récapitulatif pour variables distinctes*).
- 2) Quel est le plan d'expérience ?
- 3) Calculer manuellement l'ANOVA pour ces données et présenter les résultats dans un tableau d'ANOVA (les formules sont au verso). Est-ce que le fait d'être étudiant en psychologie ou en ingénierie affecte la reconnaissance de formes en 2D dans des conditions sous-optimales ? Les formes sont-elles traitées de manière équivalente ? Y a-t-il une interaction entre le type de forme traitée et les études suivies, dans la reconnaissance de formes en conditions sous optimales ? Observe-t-on des inversions de tendances dans les performances entre les groupes ?
- 4) Vérifier vos calculs par SPSS. La procédure est identique à celle du TD précédent.

Formules de calcul et méthode pas à pas:

$$SC_{total} = \sum x^2 - \frac{(\sum x)^2}{N}$$

$$SC_{Sujets} = \frac{\sum T_{Sujets}^2}{b} - \frac{(\sum x)^2}{N}$$

$$SC_A = \frac{\sum T_a^2}{n \times b} - \frac{(\sum x)^2}{N}$$

$$SC_{Sujets/Groupe} = SC_{Sujets} - SC_A$$

$$SC_{Intra-Sujets} = SC_{Total} - SC_{Sujets}$$

$$SC_B = \frac{\sum T_b^2}{n \times a} - \frac{(\sum x)^2}{N}$$

$$SC_{Cellules} = \frac{\sum T_{a \times b}^2}{n} - \frac{(\sum x)^2}{N}$$

$$SC_{A \times B} = SC_{Cellules} - SC_A - SC_B$$

$$SC_{(Sujets/Groupe) \times B} = SC_{Intra-Sujets} - SC_B - SC_{A \times B}$$

$$F = \frac{CM_A}{CM_{Sujets/Groupe}}$$

$$F = \frac{CM_B}{CM_{(Sujets/Groupe) \times B}}$$

$$F = \frac{CM_{A \times B}}{CM_{(Sujets/Groupe) \times B}}$$