

Support statistique Mémoire de fin d'études

(Ce papier et d'autres se trouvent au site <http://sozio.logik.ch> sous "Statistiksupport PH Wallis")

La HEP offre aux étudiantes et aux étudiants la possibilité de contacter le support statistique dans le cadre de la rédaction du mémoire de fin d'études. Ce papier décrit les procédures à observer lors d'un tel contact. Ces règles maximisent l'efficacité pour vous et pour le support statistique (minimiser le temps investi).

Le support statistique

- vous conseille par rapport aux problèmes de statistiques et
- se charge du traitement des données que vous lui remettez (tests statistiques, statistiques descriptives).
- Le support statistique vous offre aussi la mise sur internet d'un questionnaire.
- Le support statistique est prêt à contrôler les passages concernant les statistiques dans votre travail de mémoire avant le remise du travail.

Le support statistique ne peut cependant vous conseiller par rapport au contenu (p.ex. formulation des hypothèses ou des questions du questionnaire), ce qui relève de la compétence du directeur du mémoire de fin d'études. Exception : le directeur de mémoire vous renvoie pour ces questions au support statistique.

Pour mieux comprendre les procédures décrites et pour pouvoir collaborer efficacement avec le support statistique, nous commençons par une petite répétition de quelques termes que vous avez rencontrés dans d'autres thèmes mais que vous avez peut-être oubliés.

1.1 Terminologie

Pour contacter le support statistique, la connaissance de la signification des termes suivants est indispensable : variables, valeurs d'une variable, codage d'une variable, échelles de mesure (nominal, ordinal, métrique), hypothèses (univariées, bivariées, multivariées).

- Une variable est une partition d'un ensemble d'objets A (= ensemble de base ou échantillon). Une partition de A est définie comme un ensemble de sous-ensembles non-vides de A , de sorte que chaque objet de A est attribué à exactement un sous-ensemble de la partition de A . (plus formellement : Les intersections des éléments différents de la partition sont vides et l'union de tous les éléments de la partition est identique à A)

Exemples :

- Ensemble de base A = l'ensemble des habitants du Valais. La variable "genre" constitue une partition de A . Celle-ci comprend l'ensemble des hommes et l'ensemble des femmes. Chaque habitant est une femme ou un homme et aucun n'est les deux à la fois.
- Ensemble de base A = l'ensemble des élèves de Naters. La variable "attitude envers l'enseignement du français" avec les valeurs "très content", "content", "peu content", "mécontent" constitue une partition de l'ensemble A (si les élèves peuvent s'attribuer eux-mêmes ces valeurs et qu'ils doivent se décider pour l'une de ces catégories).
- Ensemble de base A = l'ensemble des élèves des classes de 6ème primaire en Valais. Variable "taille" et les valeurs "taille en mètres". (par ex. 1.83, 1.65). Si nous mesurons les élèves d'une manière assez précise, nous obtenons un seul élève par valeur.
- Ensemble de base A = l'ensemble des parents valaisans. La définition suivante d'une variable est incorrecte : les valeurs de la variable "lit beaucoup", "ne lit que les journaux", "ne lit jamais". Il peut y avoir des parents qui ne lisent que les journaux mais qui lisent beaucoup.
- Ensemble de base A = l'ensemble des habitants de St. Maurice. La définition suivante d'une variable est incorrecte : les valeurs de la variable : "fait du ski", "fait du snowboard", "fait du patinage". Premièrement, une personne peut pratiquer les trois sports. De plus, il peut y avoir des personnes qui ne font aucun de ces sports.

- Une variable correspond à une colonne d'un tableau de données, tandis que les lignes correspondent aux objets d'étude. Traduit dans le langage du tableau, la définition de la variable signifie : Chaque cellule d'une colonne doit contenir une donnée et aucune cellule ne peut contenir plus d'une donnée. Dans la réalité, il y a toujours des données qui manquent. La construction des variables doit observer les conditions citées au niveau conceptuel !
- Codage d'une variable et échelles de mesure : Nous codons une variable en attribuant des nombres aux objets d'un ensemble de base ou de l'échantillon. L'action d'attribuer s'appelle "mesurer". L'attribution en tant que telle est une application au sens mathématique et s'appelle "l'échelle de mesure". L'attribution doit obéir à certaines règles pour qu'elle mérite la qualification "d'échelle de mesure".
 - a) Il faut attribuer le même nombre aux objets d'une même valeur de la variable. Si l'échelle n'observe que cette règle, nous appelons la variable "nominale". (par ex. variables "genre", "appartenance religieuse", "langue maternelle", etc.). (Bien que l'échelle de mesure attribue les nombres aux objets de l'ensemble de base, nous dirons parfois plus simplement qu'elle attribue les nombres aux valeurs de la variable).
 - b) Si l'on peut déterminer un ordre "naturel" pour les valeurs d'une variable, on peut fixer l'échelle de sorte que les nombres attribués reflètent l'ordre entre les valeurs de la variable. Si de plus (a) est rempli, nous appelons une telle variable "ordinaire". A part l'observation de ces deux règles, l'attribution reste arbitraire et la distance entre les nombres ne comporte aucune information (par ex. attitude envers l'enseignement du français et les valeurs de l'échelle de mesure 4 pour "très content", 3 pour "content", 2 pour "peu content", 1 pour "mécontent". L'ordre entre les valeurs est "être plus content que" et la relation mathématique ">" entre les nombres reflète cet ordre. On pourrait tout aussi bien définir l'échelle suivante : 1000 pour "très content", 1 pour "content", 0.2 pour "peu content", 0.1 pour "mécontent".).
 - c) Si (a) et (b) sont respectés et si la distance entre les nombres de l'échelle joue un rôle (par ex. mesure de la distance, du temps, de la vitesse, du poids), nous parlons d'une variable "métrique". Nous obtenons immédiatement les valeurs de l'échelle en appliquant les opérations de mesure. Souvent, mais pas toujours, les valeurs de la variable ne contiennent qu'un seul objet (par ex. nous mesurons la taille des élèves d'une classe au millimètre près). Si vous voulez vous informer davantage sur les échelles de mesure, voir [Variablen, Ausprägungen und Skalen](#) sur le site internet indiqué (texte en allemand).
- Les hypothèses constituent le centre d'une analyse statistique. On peut distinguer deux types d'hypothèses :
 - a) Les hypothèses strictes correspondant à des propositions générales (exemples pour des propositions générales : Tous les êtres humains sont mortels. Tous les métaux conduisent le courant). On peut infirmer une hypothèse stricte par un seul cas qui contredit la proposition : p.ex. Un seul morceau de métal qui ne conduit pas le courant falsifie l'hypothèse ci-dessus.
 - b) Les hypothèses statistiques sont des propositions qui postulent une tendance (par ex. Les fumeurs vivent moins longtemps que les non-fumeurs. Pour deux cas concrets, un fumeur spécifique peut vivre plus longtemps qu'un non-fumeur spécifique. C'est pourquoi la proposition générale "Tous les non-fumeurs vivent plus longtemps que tous les fumeurs" est fausse. Pour cet exemple, une proposition de tendance peut être formulée ainsi : en moyenne les non-fumeurs vivent plus longtemps que les fumeurs). Les propositions statistiques ne sont ainsi pas des propositions générales. C'est pourquoi on ne peut pas argumenter avec des cas isolés par rapport à des hypothèses statistiques - un cas isolé ne peut confirmer ou infirmer une telle hypothèse. Les hypothèses statistiques sont la règle dans le domaine des sciences humaines.
- On peut distinguer des hypothèses (a) univariées, (b) bivariées et (c) multivariées.
 - a) Les hypothèses univariées postulent un fait par rapport à une variable (par ex. La majorité des étudiants de la HEP est satisfait de l'échange.).

- b) Les hypothèses bivariées affirment une relation tendancielle entre deux variables (par ex. Plus il y a de travaux de groupe dans l'enseignement du français à l'école, plus les enfants sont motivés pour le français. Plus le revenu des parents d'un enfant est élevé, plus la chance de l'enfant d'entrer au collège est grande). Presque toutes les hypothèses qui surgissent dans le cadre d'un travail de diplôme relèvent de cette catégorie. On utilise deux variables pour la falsification ou la confirmation d'une hypothèse.
- c) Les hypothèses multivariées affirment une relation tendancielle entre plusieurs variables explicatives et une variable dépendante (par ex. les variables explicatives "revenu des parents", "langue maternelle", "genre", etc. et la variable dépendante ou expliquée "compétence de lecture" - étude Pisa). Les statistiques multivariées constituent des méthodes pour estimer le poids des différents facteurs qui "influencent" une certaine variable.

1.2 Procédures à observer lors du contact avec le support statistique

Dans le cadre des recherches pour votre mémoire, différentes étapes pouvant concerner le support statistique s'imposent. Vous pouvez contacter le support statistique pour chaque étape - mais seulement après une préparation intensive.

1. Il faut savoir ce qu'on veut connaître (par écrit). Il faut formuler par écrit ses hypothèses. Les hypothèses ne sont pas des questions mais des propositions. La recherche sert à falsifier ou à confirmer ces propositions (on ne peut falsifier une question!!!!).
2. Si l'on rédige un questionnaire ou si l'on récolte des données d'une autre manière (par observation, par recherche dans des archives, par la lecture d'un texte), il faut savoir par quelles variables on veut tester quelles hypothèses (fixer par écrit). Ce n'est qu'ensuite qu'on peut récolter des données. Concernant les variables, il faut aussi déterminer à l'avance quelles sont leurs valeurs (par écrit).
3. Il faut récolter uniquement les données dont on a besoin pour l'examen des hypothèses (ou si vous avez dans le cadre de votre travail un besoin spécifique de connaissance dans le domaine des statistiques descriptives). Faire des questionnaires les plus courts possibles.
4. Lors de la définition des variables, il est très important de vérifier si une partition de l'ensemble étudié est définie par la variable : tous les objets doivent être attribués à une valeur de la variable et à une seule valeur.
Pour les questionnaires, cela implique que les questions à plusieurs réponses sont à éviter, sauf si l'on compte le nombre des réponses, ce qui crée une nouvelle variable. Si l'on ne compte pas les réponses, on crée pour chaque sous-question une variable dichotome (= ayant deux valeurs). Est-ce qu'on compte les personnes qui n'ont pas coché la réponse en question comme des personnes ayant dit "non" à la sous-question? Si l'on pose des questions à plusieurs réponses il faut savoir ce qu'on va faire avec les réponses avant de lancer le questionnaire!!!!
5. Il faut tester le questionnaire auprès de plusieurs personnes (de la même catégorie que les personnes interrogées plus tard, mais aussi auprès d'autres personnes) afin de détecter des polysémies, des questions peu claires, etc.
6. Les tests statistiques supposent que les données sont le résultat de variables aléatoires indépendantes qui suivent une distribution identique. Ces conditions sont rarement satisfaites pour des données scolaires. Parfois on essaye de garantir le respect de ses conditions par le tirage aléatoire d'un échantillon : on fixe un ensemble de base et on garantit par la procédure de tirage p.ex. la même probabilité d'être membre de l'échantillon. Une procédure pour réaliser cela est de trouver une liste des objets de l'ensemble de base (p.ex. une liste des enseignants du Valais) dans un tableau Excel, d'attribuer des nombres aléatoires (fonction Excel : "alea()") et ordonner la liste selon la ligne avec les nombres aléatoires. On prend ensuite les n premiers objets de la liste (n = taille de l'échantillon). Il y a d'autres procédures pour tirer des échantillons aléatoires. En cas de problèmes, contactez le support statistique.

Il faut cependant relever le fait, que de telles procédures sont en général trop lourdes pour des travaux de mémoire. Pour tester des méthodes didactiques on les applique à des classes entières. Il faudrait alors tirer un échantillon aléatoire de classes plutôt que d'élèves - une quantité assez grande de classes et leur soumissions à des procédures didactiques différentes n'est guère possible avec les moyens à disposition dans le cadre d'un travail de mémoire. Le but des tests statistiques dans ce contexte serait alors de familiariser l'étudiant-e avec certaines procédures (en avouant le "péché", celui-ci devient moins grave).

7. Lors de la saisie des données, on code les valeurs des variables, en fixant des nombres pour les valeurs (cela n'est pas nécessaire pour les variables métriques puisque la mesure est donnée immédiatement par la réponse). Pour les données ordinales, le codage doit respecter l'ordre des valeurs de la variable (par ex. "meilleur" correspond à la relation ">" entre les nombres de l'échelle). Pour les variables nominales, le codage est arbitraire. Pour les données métriques, vous obtenez immédiatement le codage (sauf si vous classifiez les données).
8. Pour la remise des données auprès du support statistique, il faut saisir les données d'abord dans un seul document excel (un seul tableau excel), si vous voulez comparer statistiquement les résultats de deux ensembles de base (p.ex. enseignants et élèves). Si vous saisissez les données de deux ensembles de bases, il faut que ces deux ensembles aient une relation - si vous voulez les comparer. Il faut se décider pour un de ces ensembles et exprimer l'autre par le premier (par ex. Vous avez des données de plusieurs classes et les opinions des différents enseignants. On peut attribuer l'opinion des enseignants aux élèves (variable : a un enseignant ayant telle et telle idée).
Une variable correspond à une colonne et une ligne correspond à un objet d'étude. Chaque cellule (= intersection entre ligne et colonne) doit être remplie par une seule donnée (ne pas insérer des colonnes ou des lignes vides dans un tableau excel). Voir exemple : [Beispiel Datentabelle](#) sur le site internet indiqué ci-dessus. Il faut procéder à une saisie des données la plus détaillée possible (on peut toujours recoder après la saisie, mais c'est embêtant de devoir reconsulter les questionnaires parce qu'on a trop réduit l'information en saisissant les données).
9. Si l'on saisit deux fois les mêmes variables par rapport à un ensemble d'étude, il faut toujours respecter la règle : une ligne correspond à un objet d'étude (p.ex. on teste les capacités de lecture. On fait une expérience et on refait le même test auprès des mêmes personnes. Il faut avoir une variable "nombre des points avant l'expérience" et une autre variable "nombre des points après l'expérience").
10. Il faut remettre les données auprès du support statistique sous forme de document "SPSS" (logiciel de statistiques installé dans les salles d'informatique ; exception : si vous avez peu de variables, le support statistique fait ce travail sur la base d'un tableau Excel correctement construit pour vous). Il est cependant recommandé de saisir les données d'abord à l'aide d'excel et de copier ensuite les données dans SPSS. SPSS est un logiciel avec différents éditeurs. Nous nous intéressons seulement à l'éditeur des données et l'éditeur des variables. L'un des deux apparaît après avoir ouvert le programme (l'autre peut être atteint en pressant sur "affichage de données" ou "affichage de variables" en bas à gauche de la fenêtre). Pour copier les données, on les copie dans excel (Ctrl + c), puis on passe à SPSS (éditeur de données) et on presse Ctrl + v (ou "édition", "coller"). Ensuite vous devez étiqueter correctement les variables et les valeurs des variables. Pour cela il faut ouvrir l'éditeur des variables. Pour étiqueter les variables et les valeurs des variables, il faudrait choisir des expressions courtes - ces expressions apparaîtront dans les documents que le support statistique vous envoie. Concrètement, vous nommez les variables sous "étiquette". Sous "valeurs" vous indiquez les étiquettes des valeurs (cliquer deux fois à droite dans la cellule ; ces étiquettes apparaissent dans les tableaux et graphiques que le support statistique vous envoie. Faire attention à l'orthographe ! Expressions courtes - sinon une partie des expressions n'est pas imprimée). S'il faut donner les mêmes étiquettes aux valeurs de plusieurs variables, on peut les copier (Ctrl+c) et les coller (Ctrl + v). (Variante : on peut importer

un tableau excel (fichier, ouvrir, données, choisir format *.xls et chercher le tableau excel, etc. Cette variante est surtout utile, si l'on a entré les noms des variables avec les noms des valeurs dans la première ligne du tableau excel).

Attention : Ne pas transformer ni la première ni la dernière colonne en SPSS sous affichage de variables. Ne travailler que sur les colonnes "étiquette" et "valeurs".

11. Lors de la remise des données auprès du support statistique il faut transmettre :

- (a) le tableau SPSS (comprenant les étiquettes des variables et des valeurs; si votre tableau contient peu de variables, un tableau Excel suffit)
- (b) une liste des hypothèses et l'indication par quelles variables vous voulez tester quelles hypothèses (entre parenthèses après chaque hypothèse).
- (c) Si l'on a récolté les données à l'aide d'un questionnaire, il faut joindre le questionnaire.
- (d) Une description de la procédure du prélèvement des données (ensemble de base, taille de l'ensemble de base, taille de l'échantillon, procédure du prélèvement de l'échantillon).

Si les données sont transmises correctement, le support statistique vous envoie les résultats de l'analyse statistique (html : statistique descriptive univariée, tests des hypothèses; vous pouvez copier les tableaux et les graphiques dans le document pour les insérer dans votre texte). adresse e-mail : Paul.Ruppen@hepvs.ch; Indiquer lors de la remise des données le numéro de téléphone et les heures où l'on peut vous atteindre. En principe la coopération avec le support statistique se fait par e-mail, si nécessaire par téléphone, en cas de blocage aussi par rendez-vous sur place. (Version : 20 septembre 2021)