**Work-Life Flourishing Project**

**Contexte**:

Cet ensemble de données est basé sur une enquête de la vie réelle menée en France. L’objectif est de comprendre Prospérer à la fois dans le travail et en dehors du travail. L’ensemble de données utilisé ici n’est qu’un sous-ensemble des données brutes complètes. Il comprend certaines variables sociodémographiques en plus des éléments psychométriques utilisés pour mesurer l’épanouissement. Certains éléments du questionnaire ont été supprimés pour faciliter l’analyse.

Chaque instance de l’ensemble de données est un individu humain.

NBENFANT : Représente le nombre d’enfants pour les répondants.

SEX : Représente le sexe du répondant.

TAILLE : Représente la taille de l’entreprise du répondant.

ANCIENNETE : Représente le mandat du répondant dans l’entreprise.

AGE : Âge du répondant.

SITUFAM : La situation personnelle (familiale) du répondant.

NIVEAUETUDE : Le niveau de scolarité du répondant.

**Sujet 1 :** Les éléments MHC sont utilisés pour mesurer la santé mentale positive (PMH) selon (Keyes, 2002). Les articles de la MHCB mesurent les PMH dans la vie professionnelle, tandis que les articles de la MHCC mesurent les HPM dans la vie privée. Les items 1 à 3 du MHC mesurent le bien-être émotionnel (EWB), les items (4 à 8) mesurent le bien-être social (SWB) et les items 9 à 14 mesurent le bien-être psychologique (PWB).

Info : la littérature, SWB et PWB se réfère également au fonctionnement positif.

**Sujet 2 :** La prochaine série d’items mesure les émotions : les items NE mesurent les émotions négatives tandis que les items PE mesurent les émotions positives.

**Thème 3 :** Les éléments Force mesurent différentes facettes des forces de caractère.

**Thème 4 :** Les éléments de la FQ mesurent la fréquence d’utilisation des forces de caractère.

Le codage de chacune de ces variables est fourni dans les différentes feuilles Excel du fichier du questionnaire.

**Concernant les mesures en psychométrie**

En psychologie et en sciences sociales, certains chercheurs considèrent les variables ordinales comme des variables quantitatives parce qu’elles ont un ordre naturel et peuvent être comparées de façon significative à l’aide d’opérations mathématiques comme la moyenne, la médiane et l’écart-type. Toutefois, la distinction entre les variables ordinales et quantitatives n’est pas toujours claire et peut varier selon la question de recherche, la conception et les méthodes d’analyse des données.

Voici quelques références qui traitent de cette question :

* "The Handbook of Research Methods in Social and Personality Psychology" edited by Harry T. Reis and Charles M. Judd
* "Statistics for the Social Sciences" by Agresti and Finlay
* "Quantitative Social Science: An Introduction" by Kosuke Imai, Gary King, and Elizabeth A. Stuart

Ces références donnent un aperçu complet de la question et des différentes perspectives sur le traitement des variables ordinales comme variables quantitatives dans la recherche en sciences sociales et comportementales.

***En marketing comme en comportement organisationnel – ou en règle générale en psychométrie – les chercheurs et les praticiens considèrent que les scores ordinaux supérieurs à 5 éléments sont quantitatifs. Dans ce projet, il est recommandé d’appliquer la même logique.***

**PART 1 – Data Cleaning a Preparation**

Lorsque les gens n’ont pas beaucoup d’expérience dans la réalisation d’enquêtes, ils commettent des erreurs. C’est le cas dans cette enquête. Voici les problèmes :

- Toutes les non-réponses ont été codées à 0.

- Mais certaines autres modalités ont été codées zéro, comme SEX : Female

- Parfois, le code « Ne sait pas » est 7.

De plus, certaines personnes ont commencé à répondre aux premiers sujets, mais se sont arrêtées en chemin et n’ont pas terminé les autres. Cela générera d’autres valeurs manquantes.

En bref, vous avez plusieurs types de données manquantes que vous devez encoder correctement.

**Handling missing data**

Le principe que nous allons appliquer est le suivant :

1- Lorsqu’une ligne ou une colonne a plus de 50 % de valeurs manquantes, nous supposons qu’elle apporte plus d’erreurs dans l’analyse qu’elle apporte de l’information.

a. Par conséquent, nous supprimons toute ligne et toute colonne pour lesquelles il manque plus de 50 % de données.

b. Faut-il commencer par les lignes ou les colonnes ? Réfléchissez bien !

2- Après avoir supprimé les lignes et les colonnes inappropriées, sélectionnez une méthode appropriée pour imputer les valeurs manquantes et les appliquer à vos données.

**Computing new categorical variables**

**L’épanouissement** exige une réponse de "presque tous les jours" (5) ou "tous les jours" (6) à 1 ou plus des 3 questions de bien-être émotionnel (EWB), et à 6 ou plus des 11 questions de fonctionnement positif (SWB & PWB).

**Languir** exige une réponse de "une ou deux fois" (2) ou "jamais" (1) à 1 ou plus des 3 questions de bien-être émotionnel, et à 6 ou plus des 11 questions de fonctionnement positif.

**La santé mentale modérée** désigne ceux qui ne sont ni florissants ni languissants.

1. Créer deux variables catégorielles fournissant le statut Florissant de chaque répondant : une variable pour le contexte professionnel et une autre pour le contexte privé.

**L’épanouissement émotionnel** mesure le rapport entre les émotions positives et négatives, il est aussi appelé le rapport de postivité (PsurN).

- Lorsque le ratio de positivité est supérieur à 3, le répondant est considéré comme épanouissant émotionnellement.

- Lorsque le rapport de positivité se situe entre 3 et 2, le répondant est en santé émotionnelle modérée.

- Lorsque le rapport de positivité est entre 2 et 1, le répondant languit émotionnellement.

- Lorsque le rapport de positivité est inférieur à 1, la personne est considérée comme déprimée ou dans un épisode dépressif.

1- Créer une variable catégorielle indiquant le statut affectif de chaque répondant

**PART 2 – Principal Component Analysis (PCA)**

Given the theoretical information provided here, you should consider conducting PCA on the different topics, focusing on one topic at a time (Topic 1 may be considered as comprising 2 sub-topics, each respectively associated to the work or non-work context).

We also recommend that the number of axis retained be somewhat guided by theoretical knowledge.

**PART 3 – Multiple Correspondance Analysis (MCA)**

Now apply MCA on all the categorical variables of the dataset and provide some insights on Flourishing, and if applicable on socio-demographic profiles.

**PART 4 - Clustering**

Use the aggregated factor scores as variables for clustering.

After selecting the right number of groups and assigning each individual in a cluster group, give the best possible interpretation of each group based on the previous results.

Does your result change if instead of using the aggregated factor, you used the factor score instead (the coordinate of each factor on the variables.

Does your result change if instead of the average aggregated score, you used the standardized aggregated score instead ?

**PART 5 – Multiple Linear Regression (MLR)**

Create several quantitative variables by computing the aggregated score for each topic :

* For each factor and its corresponding (questionnaire) items, compute the aggregated scores as the average of those items

At this stage the individual items are no longer used in the analyses as they are all replaced by their aggregated scores.

Build a MLR model predicting Flow-Performance & Mastery from all the other scores. Select the best, most parsimonious models and check the conditions for inference.

**PART 6 – Logistic Regression**

Now, binarize the SITUFAM variable into Couple vs Single.

Predict the binary SITUFAM variable from all the aggregated variables.

Select the most parsimonious model and its performance using all the most reliable metrics.

**Report**

Write a clear but synthetic report of you analysis, highlighting the main results, the convergence as well as the divergence amont these.

How does the result highlight the current knowledge that we might have on Flourishing ? What are the new perspectives that these results open ?