

ANALYSE DE LA VARIANCE (ANOVA)

Problème

Les données :

Un organisme socio-économique étudie le budget familial hebdomadaire alloué aux repas. Parmi les facteurs susceptibles d'influer sur ce budget, cette étude s'intéresse au facteur "taille de la famille" et au facteur "catégorie socio-professionnelle : CSP". L'enquête réalisée par cet organisme permet de classer les familles en 15 catégories obtenues par croisement de la CSP et la taille de la famille.

Taille de la famille :

- Catégorie A : Familles sans enfant.
- Catégorie B : Familles ayant 1 ou 2 enfants.
- Catégorie C : Familles ayant 3 ou 4 enfants.
- Catégorie D : Familles ayant 5 ou 6 enfants.
- Catégorie E : Familles ayant plus de 6 enfants.

CSP :

- CSP 1
- CSP 2
- CSP 3

Pour chaque famille on calcule le budget par personne et on obtient le fichier de données (disponible sur Moodle) : DonneesBudget.

Ce fichier comporte le nom des variables en première ligne. Le séparateur décimal est la virgule, les champs sont séparés par des tabulations.

Travail attendu :

Pour ce travail, vous serez encore une fois encouragés à prendre en main de manière très autonome l'étude de ce jeu de données. Il sera attendu trois principaux types d'analyses,

à savoir des :

- Statistiques descriptives univariées,
- Statistiques descriptives bivariées,
- ANOVAs à un ou deux facteurs.

Le choix des variables et modalités à étudier en priorité est *libre* et devra faire l'objet d'une **problématisation bien claire**. Encore plus que pour le travail sur la régression linéaire multiple, l'accent devra être mis sur la **concision** et la **clarté** dans les questions traitées, ainsi que dans les réponses apportées. L'objectif doit devenir de plus en plus de bien s'approprier les analyses pour étudier en profondeur **des points précis**, plutôt que d'enchaîner les graphiques et les modèles, sans recul critique ou objectif préalable. N'hésitez à proposer des méthodes originales et interprétations de votre propre initiative, par exemple en imaginant quelles pourraient être les CSP qui sont ici seulement désignées par un numéro.

Quelques règles de rédaction d'un rapport :

Le rapport est un document qui doit être de qualité avec une présentation réfléchie et structurée. Il a pour but de présenter le cheminement de votre étude. Ce n'est pas une liste exhaustive des travaux effectués mais une vitrine du travail effectué. Voici quelques règles simples à respecter.

Organisation du rapport

- Le rapport commence toujours par une introduction générale du sujet d'intérêt.
- Ensuite vous présenterez la base de données : contexte et but de l'étude, présentation des différentes variables, problématique.
- Viennent ensuite les résultats pertinents de statistiques descriptives. Les statistiques descriptives peuvent être résumées sous forme de tableaux et de figures et doivent être commentées (valeurs pertinentes pour l'étude, valeurs atypiques, valeurs manquantes, dissymétrie des distributions, ...).
- L'étude des relations entre les variables doit être présentée avec soin. Pour chaque relation étudiée, présentez précisément les variables étudiées, l'outil statistique utilisé, le résultat de la comparaison ou du test, la conclusion que l'on peut en tirer. Terminez par un commentaire replaçant le résultat obtenu dans le contexte de l'étude.
- N'oubliez pas d'interpréter **TOUS** vos résultats.
- Vous conclurez votre rapport en donnant deux conclusions à cette étude :
 1. La première s'adressant à un spécialiste de la statistique.
 2. La seconde étant destinée à un non spécialiste de la statistique.
- Si vous insérez des annexes veillez à ce qu'elles soient en nombre limité.