



## Expertise judiciaire et statistiques

---

Cet article est paru dans le numéro 139 de la **Revue Experts** en octobre 2018. Il est reproduit intégralement avec l'aimable autorisation de la revue que nous remercions. Cet article pouvant être destiné à un public beaucoup plus large, il est publié comme livre blanc Ariteam.

L'accroissement de l'enregistrement informatique des données peut amener l'expert à vouloir les exploiter. Comment faire lorsque l'on n'a pas manipulé ces notions depuis longtemps? Comment restituer les résultats dans le rapport? Quel logiciel?

**Olivier Dardare** Ingénieur, IAE, black belt six sigma certifié, Schneider electric

[olivier.dardare@ariteam.fr](mailto:olivier.dardare@ariteam.fr)

**Bertrand Vassor** Ingénieur, Expert près la Cour d'appel et Chambéry

[bertrand.vassor@ariteam.fr](mailto:bertrand.vassor@ariteam.fr)

Ariteam – 2<sup>e</sup> édition – 2020-02

Ce document est mis à disposition selon les termes de la licence Creative Commons "Attribution - Pas d'utilisation commerciale - Pas de modification 3.0 non transposé".



Il peut arriver au cours d'une expertise que l'expert envisage de réaliser une étude statistique des données dont il dispose, et que les études statistiques soient pour lui très occasionnelles voire exceptionnelles. Ce pourrait être le cas par exemple d'une expertise se tenant en milieu industriel et ayant pour but de vérifier les performances (objet du litige) d'un procédé de fabrication au cours du temps.

Pour constater la performance du procédé, l'expert a le choix, suivant les cas, d'analyser soit les données de fabrication saisies par un opérateur de production manuellement sur une feuille, soit les données issues d'enregistrements systématiques des capteurs de l'usine à des fréquences de quelques dizaines de ms (milliseconde), si cette dernière en dispose.

Dans le premier cas à partir de quelques feuilles de production l'expert donnera un avis. Mais cet avis ne sera que peu représentatif. L'analyse de l'ensemble des données de production sur plusieurs mois, voire années est bien plus riche et fournit incontestablement des résultats plus robustes.

La montée en puissance de l'enregistrement systématique des données dans tous les domaines, sans même parler des *big-data*, nécessite un emploi de plus en plus fréquent des statistiques. Les statistiques ont pour but de donner une lecture intelligible de ces grands nombres de données représentatives du phénomène observé, lecture qu'il conviendra de « traduire » pour la rendre compréhensible par les parties et les magistrats.

Dès lors comment procéder? Les questions qui se posent à l'expert sont :

1. comment récupérer les données enregistrées et les mettre en forme pour les traiter?
2. comment traiter ces données?
3. quelles analyses réaliser en fonction des compétences de l'expert? Est-ce suffisant?
4. comment restituer les résultats suivant la compétence des parties en statistiques?
5. avec quel logiciel travailler?

L'objectif de cet article est de répondre à ces questions. Il ne prend pas en compte les éventuelles questions de procédures expertales telles que l'intégrité des informations transmises par les parties ...

## 1 Récupération et mise en formes des données

Une analyse sommaire au début de l'expertise permettra à l'expert de comprendre de quelles données il peut disposer : des relevés manuels systématiques des opéra-

tions de production, des analyses des prélèvements des matières premières, sur les produits intermédiaires et finis, des données de processus de fabrication stockées dans un outil de supervision ... Ces données peuvent être présentées en tableau où l'on a en colonne les différents résultats d'analyses et en ligne les lots ou bien les dates et heures d'analyses.

Une des premières questions est de savoir comment l'expert va pouvoir rassembler, lier ces données pour les faire parler.

Cette première étape, longue et fastidieuse, représente un temps souvent supérieur au temps des analyses elles-mêmes !

Dans notre exemple, le procédé de fabrication est doté de capteurs qui transmettent les valeurs de ceux-ci à une supervision et ce en continue toutes les 50 à 70 ms. que le processus soit en production ou non. Les résultats d'analyse des matières premières, des produits intermédiaires et finis sont disponibles dans un tableur comme Excel ou Libre Office.

Le mode de stockage des données de la supervision dépend de l'éditeur de cette dernière. Il dépend aussi du volume d'information à stocker. On trouve un stockage en base de données ou en fichiers. Que ce soit en bases de données ou en fichiers, généralement on a, en ligne dans notre exemple, le temps chronologique comme fil conducteur. En colonne, on trouvera le capteur et la valeur mesurée.

En fichiers ou en base de données on trouvera la présentation suivante assez souvent.

dates	heures	capteur	valeur
2018/03/05	22:00:00.075	Débitmètre 1	0.000000
2018/03/05	22:00:00.075	Débitmètre 2	0.714049
2018/03/05	22:00:00.091	Débitmètre 4	53.656685
2018/03/05	22:00:00.091	Température 1	10.005787
2018/03/05	22:00:00.091	Température 2	72.413193
2018/03/05	22:00:00.091	Débitmètre 2	0.722056
2018/03/05	22:00:00.157	Débitmètre 1	0.000000

Table 1 – Exemple

L'exemple 1 met en avant les difficultés suivantes :

- les intervalles de temps ne sont pas constants,
- il n'y a pas de colonnes par capteurs,
- des données sur les capteurs sont parfois manquantes.

Par ailleurs, les données stockées dans une base de données sont parfois compressées pour gagner de la place. Dans ce cas l'intervention du prestataire ayant mis en

place le système d'information est nécessaire pour restituer les données sous un format compréhensible généralement par fichiers au format *csv*<sup>1</sup>.

Avec un historique de 18 mois par exemple cela représente par capteur un nombre considérable de données :

$$20 \times 3600 \times 24 \times 365 \times 1.5 = 946\,080\,000 \text{ données}$$

20 étant le nombre de mesures par secondes, 24, le nombre d'heures par jour, 365, le nombre de jours par ans et 1.5 est équivalent à 18 mois.

L'emploi d'un tableur s'avère dès lors exclu. Le nombre de capteurs pouvant être important, le volume de données devient vite gigantesque.

## 2 Traitement initial et exploration des données

### 2.1 Réduction du nombre de données à analyser

Au moins dans un premier temps, l'expert qui souhaite disposer d'information sur plusieurs mois, peut probablement se contenter d'une valeur toutes les 10 mn par exemple. Toujours sur 18 mois, le nombre de lignes se trouve réduit à :

$$5 \times 24 \times 365 \times 1.5 = 65\,700$$

5 étant le nombre de données par heures.

Le nombre de données devient compatible avec une exploitation avec un tableur. Mais se pose la question de la détermination de la valeur toutes les 10 mn. Est-ce la valeur la plus proche de ces 10 mn ? Est-ce la moyenne, la moyenne pondérée avec le temps entre les mesures, mais comment l'obtenir ?

La moyenne pondérée par le temps s'obtient avec l'équation suivante :

$$\frac{\frac{v1 + v2}{2} \times (T2 - T1) + \frac{v2 + v3}{2} \times (T3 - T2)}{T3 - T1}$$

différente de la moyenne statistique d'un tableur ou d'une base de données. Dans l'exemple de la fig. 1 la moyenne sera beaucoup plus proche de M2 que de M1. Dans le cas d'une moyenne non pondérée nous aurions eu le milieu entre M1 et M2.

1. Comma-separated values : format de fichier informatique texte, ouvert et présentant les données de manière tabulaire. Les colonnes sont séparées par un signe tel que « , » ou « ; » par exemple.

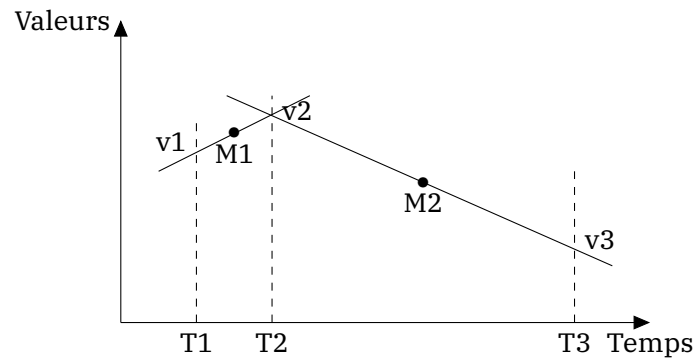


Figure 1 – Moyenne pondérée par le temps

## 2.2 Valeurs manquantes

Dès à présent se pose la question des données manquantes. Le tableau 2.2 présente ce cas.

heure	Capteur 1	Capteur 2
22:10:00	810	810
22:20:00	NA	0
22:30:00	320	320
22:40:00	523	523
moyenne	551	413,25

Table 2 – Impact des valeurs manquantes

Le résultat de la moyenne montre un résultat bien différent suivant le cas d'un capteur indiquant NA<sup>2</sup> ou 0 par défaut de mesure.

Ce problème est important. En imaginant que le capteur ne transmette pas ses valeurs durant quelques minutes, la valeur enregistrée est 0, la moyenne est donc 0, alors qu'en réalité ce sont des valeurs manquantes. L'expert devra manuellement ou presque traiter au cas par cas.

## 2.3 Exploration des données?

Après ce travail indispensable de nettoyage des données, comment l'expert va-t-il ensuite analyser les données? Avec quels objectifs? On peut supposer qu'il voudrait établir ou écarter des liens entre les variables, observer les distributions de la population, étudier des ressemblances, regarder dans notre cas les évolutions au cours du temps de ces capteurs.

2. NA : not available : donnée non disponible

La démarche d'exploration peut s'effectuer de la manière suivante :

**description des variables :** c'est la partie constatation des données. Les variables sont décrites au moins dans un premier temps par moyenne, médiane, écart-type, quartile. Les représentations avec des histogrammes, des boîtes à moustaches, des diagrammes de fréquence permettent déjà de voir des points intéressants. Dans notre cas, l'une des parties affirmait que la durée de cycle de production était inférieure à 12h et que le débit du procédé inférieur à 1 000l/h, pour un objectif à 1 400. Il s'avère que les choses étaient beaucoup plus nuancées et que 25% des cycles étaient compris entre 12h et 24h augmentant d'autant la quantité produite en un jour comme le montre la fig. 2.

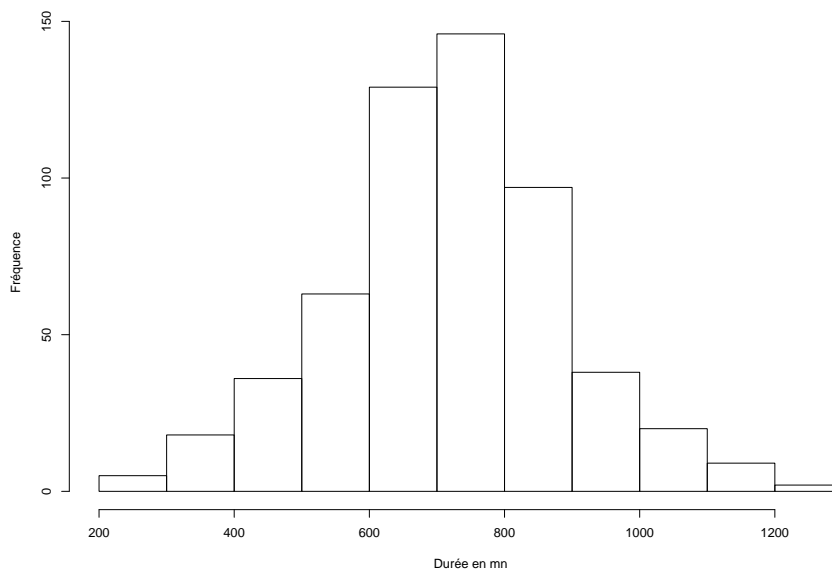


Figure 2 – Histogramme des durées de production

Dans notre cas le produit fabriqué doit être traité à une température de 70 °C. Le simple examen de la valeur minimale du capteur de température indiquant 0.5 °C a permis de détecter des problèmes n'ayant jamais fait l'objet de relevé par les opérateurs de production.

**recherche des corrélations :** la recherche des corrélations entre les variables met en avant des liens de causalité ou de relation à une même cause. Certains logiciels permettent de graphiquement voir ces corrélations ce qui simplifie considérablement le travail.

La figure 3 indique le degré de corrélation en fonction de l'intensité de la couleur. Plus le bleu (corrélation positive) et le rouge (corrélation négative) est foncé, plus la corrélation est forte. Près de 40 variables sont étudiées. Les capteurs travaillant « ensemble » sont clairement identifiables.

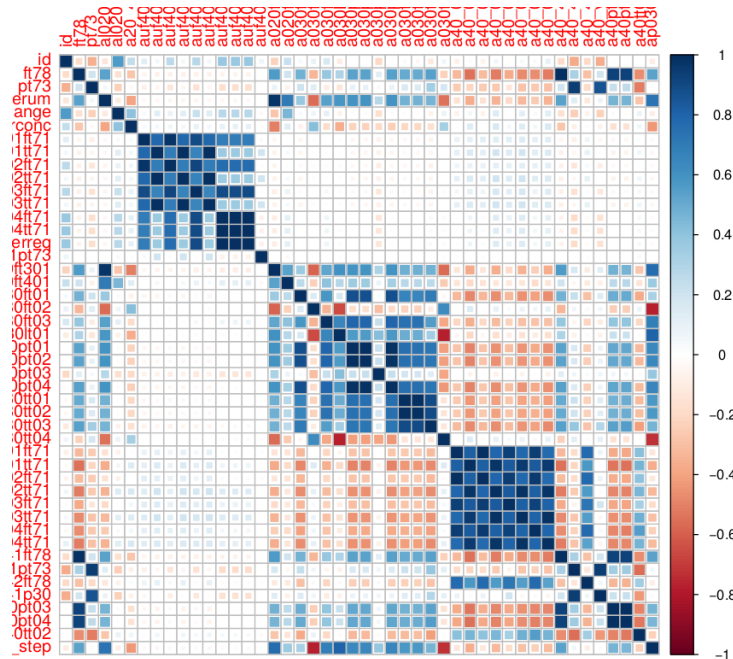


Figure 3 – Tableau des corrélations 2 à 2

### 3 Quelles analyses statistiques ?

#### 3.1 Analyses courantes

Quelles démarches d’analyse l’expert peut-il adopter ? Après avoir traité ces données pour permettre au logiciel de faire les calculs et fait une première démarche exploratoire, l’expert commencera très probablement par décrire plus finement les données avant de les interpréter.

Ensuite il s’attachera à regarder quelle loi statistique les données peuvent-elles suivre. Dans notre cas, les données étant nombreuses, on peut supposer que la loi normale<sup>3</sup> sera privilégiée, tout en se souvenant qu’on ne démontre pas la normalité d’une distribution mais la non-normalité. Les diagrammes de distribution et quantiles-quantiles permettront de comprendre dans un premier temps cette normalité ou non.

Ensuite il pourra suivant les cas, faire les comparaisons de moyenne, d’écart-type, au moyen des tests de Student, Fisher, des analyses de la variance. Peut-être sera-il amené à vouloir analyser les liens entre les variables. Dans notre cas, y a-t-il des liens entre les températures, les pressions et les débits de notre processus. Le nombre de variables étant assez importants, l’utilisation de logiciels comme R s’avère indispensable.

Dans notre cas, le calcul du préjudice passe par une modélisation des capacités de

3. appelé également loi de Gauss

production. La régression linéaire nous est apparue comme étant possible. Elle a permis de modéliser la courbe de production « moyenne » avec ses différents cycles et par voie de conséquence la taille de lot économiquement optimale, différente en l'occurrence de la taille de lot de l'exploitant. Cette courbe est associée aux courbes de prédiction et de confiance. Cette faculté d'afficher les courbes de prédiction et de confiance permet de conforter les valeurs et a également un intérêt didactique comme le montre la figure 4.

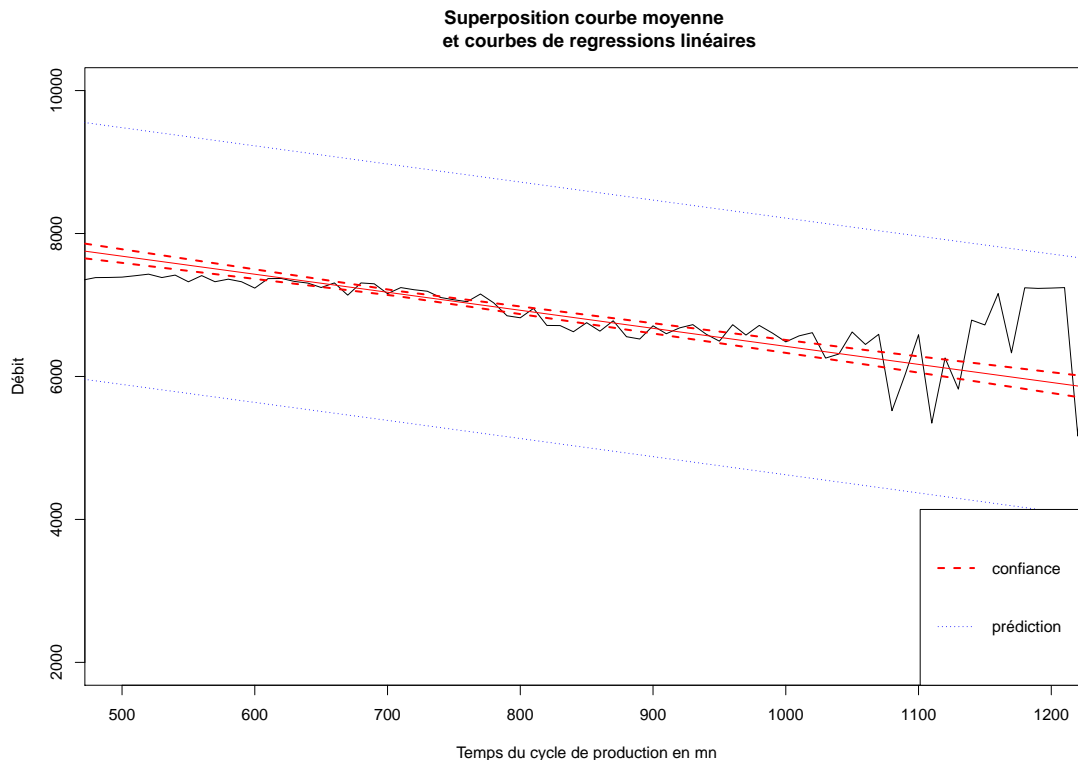


Figure 4 – Regression linéaire

### 3.2 Analyse de données multidimensionnelles

D'autres analyses, certes plus complexes, mais néanmoins abordables peuvent être mises en œuvre lorsqu'il y a un grand nombre de données et variables présentées en tableau où les individus sont en lignes et les variables en colonnes. Ce sont entre autres les analyses factorielles. Ces méthodes descriptives visent à structurer l'information portée par les nombreuses variables pour les résumer par des axes d'importance décroissante. Elles permettent d'étudier simultanément les individus pour en dégager une notion de « profil ». Il s'agit des analyses factorielles :

**ACP** : analyse en composantes principales, où les individus sont décrits par des variables quantitatives;



**AFC** : analyse factorielle de correspondance, dédiée aux tableaux de contingence <sup>4</sup> ;

**ACM** : analyse des correspondance multiples, où les individus sont décrits par des variables qualitatives ;

**AFDM** : analyse factorielle de données mixtes, où les individus sont décrits par des variables qualitatives et quantitatives ;

**AFM** : analyses factorielles multiples, où les individus sont décrits par des variables qualitatives ou quantitatives, variables structurées en groupes.

**AFMH** : analyse factorielle multiple hiérarchique, généralisation de l'AFM à des groupes hiérarchiques.

Ces méthodes peuvent être complétées par les classements ascendant hiérarchique (CAH).

## 4 Restitution

Après avoir mené ses investigations l'expert devra restituer aux parties d'une part et dans son rapport d'autre part ses conclusions.

La restitution des données aux parties peut provoquer plusieurs effets. Alors que les résultats des analyses statistiques n'étaient constituées que de notions comme moyennes, écart-type, quartile, c'est à dire des notions d'un niveau de terminale scientifique, ou des notion de régressions linéaires vues en niveau licence scientifique, nous avons pu constater un rejet complet d'une partie, bien que cette dernière prétendait avoir des ingénieurs dans ses effectifs. Peu habitué probablement aux démarches statistiques, cette partie a surtout compris que toute l'histoire de la production sur les 18 derniers mois devenait compréhensible, lisible, transparente à tous. Elle même n'avait pas songé à extraire les données pour les exploiter *a minima*, avant le début de l'expertise, en vue de tenter de comprendre où sont les problèmes : conception des équipements, dimensionnement des équipements, opérations de production mal maîtrisées ...

Cette partie n'a jamais accepté que l'on pouvait s'affranchir de faire des tests *in situ* pour constater que l'équipement ne fonctionnait pas correctement. Un test d'un jour permet effectivement de voir, de toucher le problème éventuellement, mais disposer de 18 mois d'historique a une valeur probante bien supérieure.

L'emploi de techniques plus élaborées comme l'analyse de la variance, ou les analyses factorielles nécessite, lors d'une réunion de restitution par exemple, d'expliquer *a minima* les techniques, la raison du choix, les résultats obtenus et les conclusions que l'on peut tirer. L'utilisation de graphiques, qui suivant les logiciels peuvent être très élaborés parfois, sera un plus. Nous avons constaté que cette

---

4. de manière simplifiée, tableau croisé dynamique

restitution permettait de factueliser les discours des partis et également pour elles de mieux comprendre le processus en cause.

Comme dans tout rapport, il conviendra d'expliquer en termes compréhensibles par les magistrats, les conseils, la démarche et les conclusions auxquelles l'expert arrive.

L'interprétation des résultats statistiques pouvant être soumis à de fortes ambiguïtés, il convient d'être très rigoureux dans la méthode pour éliminer au maximum les incertitudes. Il est préférable d'employer une formule comme « il est statistiquement significatif ... » plutôt que « l'intervalle de confiance à 95% ... » ce qui laisse la possibilité à l'une des parties de mettre en avant que rien n'est certain! Mais il souhaitable également de regarder la qualité des résultats et leur fiabilité avec la *p-value*<sup>5</sup>. Les logiciels fournissent en général cette valeur. Cette *p-value* permet de quantifier la fiabilité des calculs. Habituellement, si elle est inférieure à 5%, le résultat du test est déclaré comme statistiquement significatif. Bien sûr ce seuil abaissé, ou une *p-value* faible, renforce la significativité du test.

## 5 Quels logiciels?

L'évolution des outils informatiques est telle qu'il est devenu inopportun de faire les calculs de statistique manuellement. Les questions qui se posent sont déterminantes pour le choix du logiciel : tableur ou un logiciel spécifique? L'expert a besoin avant tout d'un logiciel fiable, c'est à dire dont il est certain qu'il donne la bonne réponse à un calcul. Si le calcul d'une moyenne est aisément vérifiable, qu'en est-il d'un résultat d'un écart-type dans un cas simple ou d'un test statistique de Fisher? Comment l'algorithme est-il mis en œuvre et vérifié? Comment les éventuelles anomalies de calcul (*bug*) sont-elles corrigées? Comment les données manquantes sont-elles prises en compte?

Dispose-t-il des fonctionnalités qui seront utilisées dans les analyses? Et fournit-il des commentaires pertinents sur les limites des méthodes employées? Il serait dommage en effet après avoir réalisé la partie descriptive des données (moyenne, écart-type, quartiles, médiane) d'être confronté à l'impossibilité lors d'une régression linéaire de disposer des intervalles de confiance, ou de vouloir faire une ACP et de ne pas disposer de cette fonctionnalité.

Ensuite, les calculs pourront-ils être vérifiables, contrôlables? Autrement dit, sont-ils reproductibles par les parties, et si oui, facilement? Une autre manière de répondre à cette question est de savoir si les parties peuvent comprendre le raisonnement suivi, les données calculées à partir d'un rapport du logiciel, d'un code logiciel compréhensible avec un minimum d'apprentissage?

---

5. valeur p en français

Enfin est-il aisément manipulable pour permettre à l'expert, qui n'aurait pas pratiqué d'études statistiques depuis quelque temps, un apprentissage rapide ? Permet-il de faire des analyses à tâtons, facilement reproductible, extensible ?

Bien évidemment, il ne faut pas négliger l'aspect économique, quel est le coût d'acquisition de licence si le logiciel est propriétaire ou l'effort de prise en main s'il est *Open source* ?

## 5.1 Fiabilité

Les logiciels de type tableur sont-ils fiables, autrement dit, donnent-ils la bonne réponse pour la question posée ? La littérature spécialisée en statistique ne recommande pas l'usage de tableurs de type Excel, et ceci pour deux raisons principales : La première concerne l'absence de protection contre les erreurs de calcul d'un certain nombre de fonctions statistiques de base. Un exemple d'erreur grossière de calcul est indiqué dans la littérature<sup>6</sup>. La seconde raison réside dans le type même de fonctions implémentées : elles ne sont pas toujours identiques aux procédures standard communément utilisées par la communauté scientifique, et ne sont pas suffisamment documentées pour permettre une utilisation performante adaptée à un usage de travail réel.

RExcel fait figure d'exception, dans la mesure où Excel se comporte simplement comme une interface de calcul au logiciel R, lui-même plus éprouvé et documenté qu'Excel.

Excel, ou ses équivalents, est à privilégier pour deux activités : la gestion d'une base de données de taille modérée, et un moyen d'organiser le calcul (préparation et vérification des feuilles de données, tris et filtrages, nommage des colonnes, etc... ). Le calcul à proprement parler devrait plutôt être confié à un logiciel de calcul de type R.

## 5.2 Fonctionnalités

Les outils statistiques sont capables de traiter aussi bien de faibles échantillons que de lourds tableaux, d'effectuer des comparaisons entre différentes configurations et d'effectuer un très vaste ensemble d'analyses statistiques. Les traitements actuels à disposition, d'accès libre ou payant, sont en mesure aujourd'hui de donner des résultats également très utiles sur ce qu'on appelle des données de survie<sup>7</sup>, et l'incertitude sur ces résultats.

---

6. par exemple *Statistical Analysis and Data Display*, R.Heiberger et B. Holland, Springer, 2015, pages 869–871, au sujet d'Excel 2013 pour Windows, et Excel 2011 pour Macintosh

7. données représentées par un temps, par exemple de fonctionnement, ainsi qu'un état à la fin de ce temps : état fonctionnel, ou état défaillant; lorsque l'état est fonctionnel, la donnée est dite censurée.

En ce qui concerne le logiciel libre R, les fonctions intégrées offertes sont si compactes qu'un traitement puissant ne requiert que quelques lignes de code judicieusement écrites, souvent quasiment offertes dans la littérature ou accessible par internet. Les traitements peuvent aussi être « non paramétriques » (statistique de Kaplan-Meier par exemple), autrement dit ne requérant aucune hypothèse sur la forme de la distribution (Weibull par exemple pour une des plus connues), ce qui évite des hypothèses sur les distributions; des tests de comparaisons existent (Log-Rank test par exemple) pour conduire à une conclusion. Il faut retenir de tout cela que les traitements sont d'une grande puissance, qu'ils demandent bien évidemment un minimum de connaissance dans leur mise en œuvre et interprétation, et qu'ils délivrent des informations précieuses sur des problèmes assimilables à des données de survie.

### 5.3 Reproductibilité

La démarche d'analyse est-elle facilement reproductible par les parties? À partir des données transmises, les parties ont-elles la possibilité de reconduire facilement la démarche, de la critiquer? L'utilisation d'un tableur ne facilite pas ce point. L'expert s'il fourni son fichier de travail embarquera avec celui-ci l'ensemble des données. Ensuite, même avec un rapport détaillé, il est bien difficile de reprendre toutes les formules. Il faut déjà savoir dans quelles cellules du tableur elles sont situées ce qui, parmi plusieurs onglets, n'est pas une sinécure.

L'utilisation d'un logiciel fonctionnant avec des scripts est un indéniable avantage. Qu'est-ce qu'un logiciel avec un langage par script? Un script est composé de commandes qui s'exécutent les unes après les autres. Le langage script, comme celui de R<sup>8</sup> par exemple, permet à un utilisateur non informaticien d'effectuer des commandes. Dans l'exemple qui suit, *str* correspond à structure et donne la structure des données, *summary* restitue le résumé des données étudiées (moyenne, écart-type, modalités ...), *barplot* affiche un diagramme à barre, *boxplot* affiche les données sous forme de boîte à moustaches.

```
1 donnees <- read.csv2("/home/monFichier.csv")
2 str(donnees, max.level=NA)
3 summary(donnees)
4 barplot(donnees)
5 boxplot(donnees$duree)
```

Les fichiers scripts sont des fichiers de petits volumes de type texte. Ils peuvent être envoyés aux parties indépendamment des données volumineuses.

---

8. R est un logiciel de statistique basé sur un langage de script

## 5.4 La démarche d'analyse à tâtons, ajout de données

La démarche exploratoire des données se fait bien souvent à tâtons. L'emploi d'un logiciel qui facilite cette démarche est un plus. En effet, ce n'est pas le même travail que de chercher dans une cellule d'un tableur une formule à modifier que d'aller modifier directement telle ou telle ligne de commande dans un script, là où le texte est déjà visible.

L'exploration des données peut se faire aussi par ajout de nouvelles données dans le temps, si l'expertise dure quelques mois. Un logiciel qui intégrera ce point sans modification de code sera préféré.

## 5.5 Quels logiciels? Quel apprentissage?

Dans le cas d'un expert effectuant ponctuellement des analyses statistiques, le choix est assez simple. Il sera guidé par le niveau d'analyse requise (moyenne, écart-type...) ou des analyses plus complexes. Le coût d'acquisition et le temps d'apprentissage peuvent être un critère. Compte-tenu de l'accroissement des demandes en analyses statistiques, le temps d'apprentissage peut être vu comme un investissement, un développement d'une nouvelles compétence.

Si les données ne sont pas trop nombreuses, les attendus simples, alors un tableur sera suffisant. Si au contraire, il est nécessaire d'aller plus loin, un logiciel comme R peut être employé. R est un logiciel *open-source*, c'est à dire fruit de la coopération de très nombreux spécialistes mondiaux, d'accès gratuit, dédié uniquement aux statistiques et disposant d'une communauté très active. Très employé dans le milieu de la recherche, il est intégré aussi aux bases de données Oracle et SQL de Microsoft. Il dispose d'un langage de script de haut niveau permettant à l'utilisateur de s'affranchir de nombreuses lignes de code. Il possède, si on le souhaite, des interfaces graphiques très ergonomiques via RStudio ou RCommander.

L'apprentissage pour une remise à niveau des statistiques peut se faire par un mooc<sup>9</sup>. Il en est de même pour travailler avec R<sup>10</sup>.

## 6 Ariteam

Ariteam est une tribu de 7 consultants aguerris à la quête de l'efficience aux confins des galaxies Projet & Process, ayant pour objectif d'aider les entreprises à maximiser leur organisation en optimisant les processus et en digitalisant les outils.

---

9. mooc : massive open online course, autrement dit, formation à distance

10. Nous citons tout particulièrement deux formations de la plateforme France Université Numérique : <https://www.fun-mooc.fr/> : *Introduction à la statistique avec R* par Université de Paris Sud et *Analyse des données multidimensionnelles* par AgroCampus Ouest

Ariteam est hébergé par Aritem Consulting.

Les domaines de compétence ou certifications sont :

1. Management de projets :

- planification et coûténance (IPMA),
- Management of Portfolio,
- HERMES 5 Foundation,
- Prince2 practitioner et accréditation formateur,
- PMP,
- Scrumaster par Scrum Alliance,
- Expert auprès de l'AFNOR pour la standardisation des bonnes pratiques.

2. Management et processus :

- MBA - IAE Grenoble,
- Lean Six Sigma black belt,
- Pilotage par les processus : BPMN 2.

3. Domaine IT :

- certification ITIL v3,
- SAP Certified - Associate Business foundation & Integration ERP 6.0,
- Microsoft certified technology specialist : Office project and Project Server 2007,
- BeoTic, CA PPM, Sciforma : administrateur.

4. Statistiques :

- Introduction à la statistique avec R,
- analyse des données multidimensionnelles et *data mining*,
- analyse textuelle (*text mining*).

## Table des figures

1	Moyenne pondérée par le temps . . . . .	4
2	Histogramme des durées de production . . . . .	5
3	Tableau des corrélations 2 à 2 . . . . .	6
4	Regression linéaire . . . . .	7