



Le système d'information d'une entreprise :

Hiérarchisation des paramètres d'intérêt

1. Présentation du besoin

Le système d'information d'une entreprise recense toujours un grand nombre de variables, de nature très différente. S'il s'agit d'un process industriel, ce sont les mesures faites tout au long du process ; s'il s'agit d'une compagnie d'assurance, c'est l'évolution des contrats, etc. Il y a toujours une ou plusieurs "variables d'intérêt" : celles qui intéressent au premier chef le responsable de l'entreprise. Selon les cas, ce sera la qualité de la production, le chiffre d'affaires, etc.

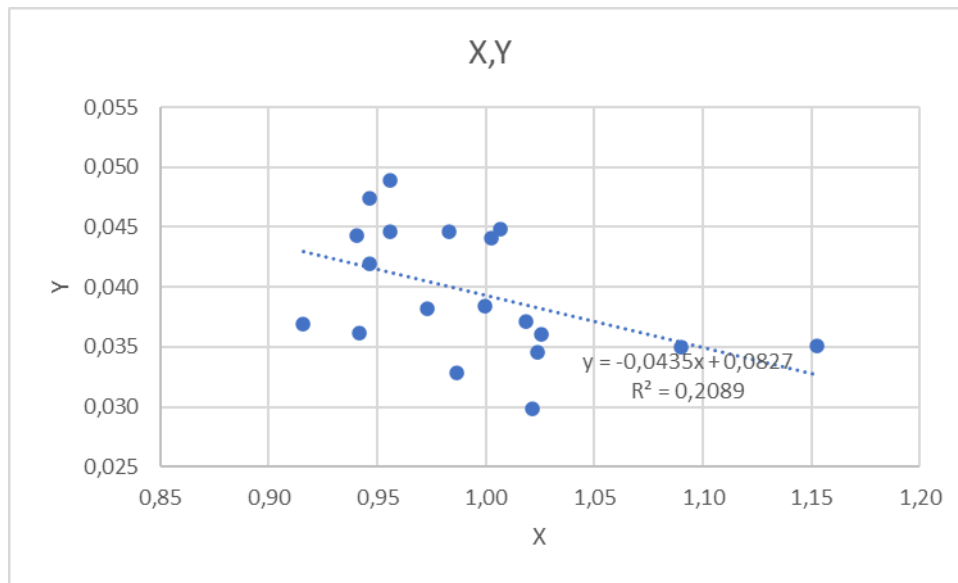
L'entreprise se demande toujours : parmi tous ces paramètres que nous enregistrons, quels sont ceux qui ont la plus forte influence sur la variable d'intérêt ? C'est évidemment ceux-là qu'il faut surveiller en premier, sur lesquels il faut tenter d'agir.

Les données se présentent en général sous la forme d'un tableau Excel ; chaque colonne représente un paramètre et chaque ligne une mesure. La méthode la plus couramment employée consiste à calculer le coefficient de corrélation entre la variable d'intérêt et chacun des paramètres ; on fera la classification au prorata de ce coefficient de corrélation.

Cette méthode est très simple à mettre en œuvre, car Excel calcule immédiatement les coefficients de corrélation. Mais elle est très insatisfaisante et, pis, elle est dangereuse.

Il faut se rappeler en effet que le coefficient de corrélation ne détecte que les corrélations linéaires : si le coefficient de corrélation, noté $\rho(X, Y)$, est très proche de 1, cela signifie que $Y \approx aX + b$. Mais, dans l'immense majorité des cas, le coefficient n'est pas proche de 1, et il est faux que si $\rho(X_1, Y) = 0.54$ tandis que $\rho(X_2, Y) = 0.32$, X_1 soit plus influent sur Y que ne l'est X_2 .

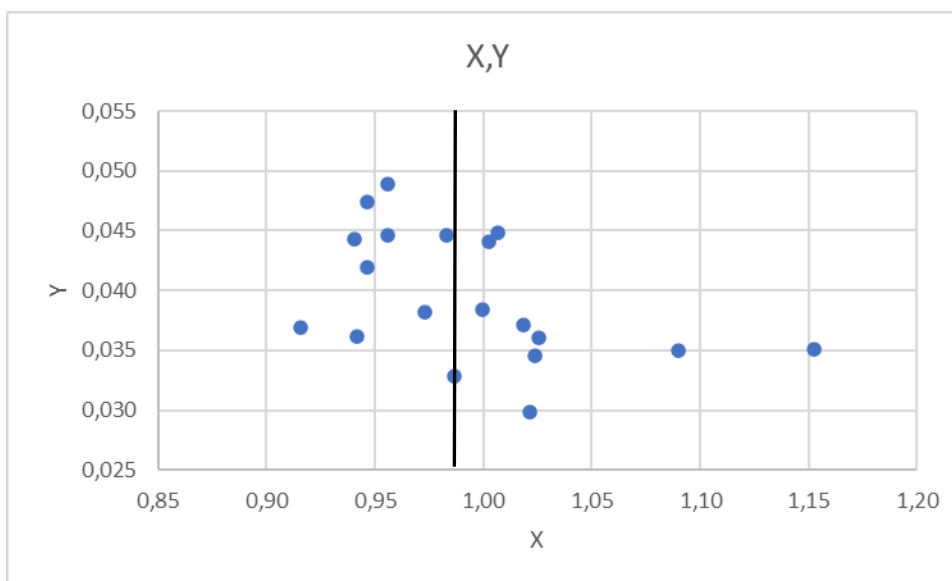
Prenons un exemple très simple, avec peu de données ; la représentation est faite sous la forme d'un nuage de points, X en abscisse et Y en ordonnée :

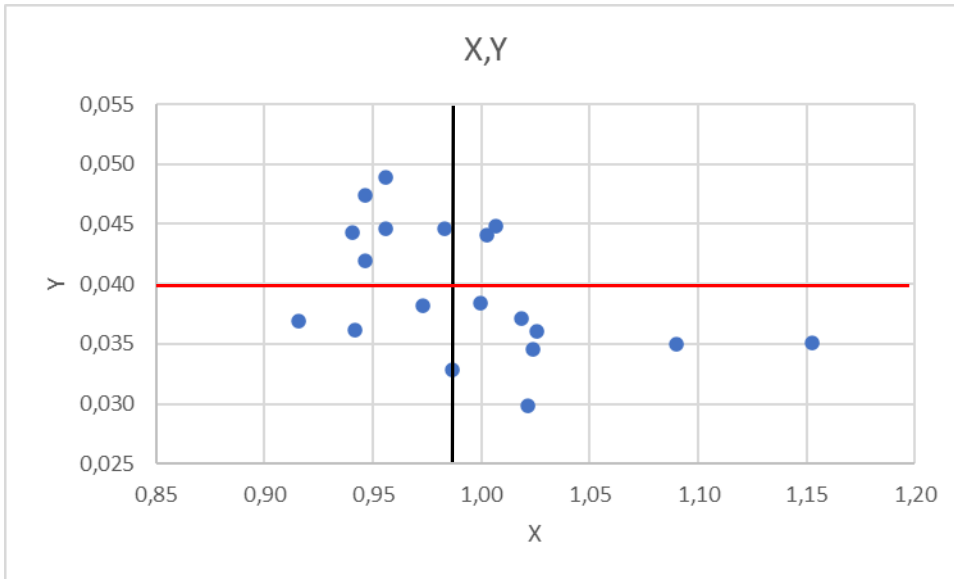


La droite de tendance est dirigée vers le bas, laissant penser que plus X est grand, plus Y est faible, mais le coefficient de détermination est très mauvais ($R^2 \approx 0.21$) et il est évident que l'approximation du nuage par une droite est très mauvaise. Si on regarde le nuage, aucune conclusion claire n'apparaît.

La méthode de hiérarchisation que la SCM a introduite, voici une dizaine d'années environ, est purement probabiliste et permet de s'affranchir de ces difficultés. Elle procède de la manière suivante :

On commence par calculer la médiane de X (il y a autant de valeurs de $X > \text{médiane}$ que de valeurs $X < \text{médiane}$) ; elle vaut 0.986 dans notre exemple et il y a 9 valeurs de X au-dessus et 9 au-dessous. C'est le segment vertical dans la figure ci-dessous.





Traçons ensuite une horizontale : ici, en rouge, $Y = 0.40$ et comptons le nombre de points au-dessus et au-dessous, dans chacun des deux cas : $X < m$ et $X > m$:

Si $X < m$, il y a 3 points au-dessous de $Y = 0.40$ et 6 au-dessus

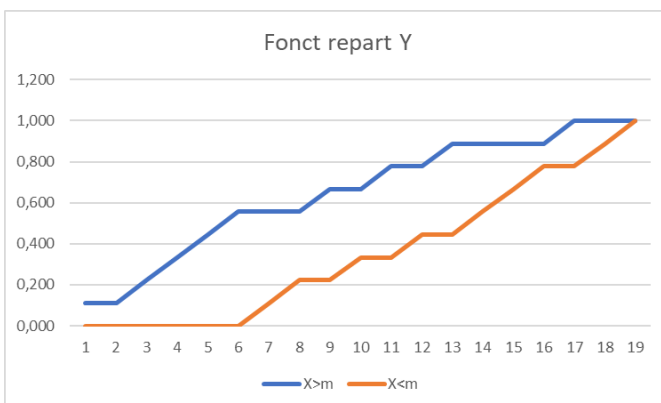
Si $X > m$, il y a 7 points au-dessous de $Y = 0.40$ et 2 au-dessus

Si $X < m$, la probabilité d'avoir $Y < 0.40$ est donc $3/9$ et la probabilité d'avoir $Y > 0.40$ est $6/9$.

Si $X > m$, la probabilité d'avoir $Y < 0.40$ est donc $7/9$ et la probabilité d'avoir $Y > 0.40$ est $2/9$.

Autrement dit, il est beaucoup plus probable d'avoir Y faible lorsque $X > m$ que lorsque $X < m$. Le résultat est similaire si on déplace l'horizontale en rouge, vers le haut ou vers le bas. On conçoit donc bien que X a une influence sur Y .

Pour rendre ceci quantitatif, on trace la fonction de répartition de Y , dans les deux cas $X < m$ et $X > m$; rappelons qu'elle est définie par $F(y) = P(Y \leq y)$; voici les graphes :



L'aire comprise entre la courbe en marron et la courbe en bleu caractérise l'influence du paramètre X sur Y ; on calcule cette aire pour chaque paramètre X et on dispose ainsi d'un

moyen très simple pour hiérarchiser les paramètres : celui pour lequel l'aire séparant les courbes est la plus forte est le plus influent.

2. Méthode de calcul

L'aire sous le graphe de la fonction de répartition peut être déterminée très simplement ; voir l'article http://www.scmsa.eu/archives/BB_fn_repartition_2020_01_18b

3. Livre

On trouvera les justificatifs théoriques dans le livre :

[NMP] Bernard Beauzamy : Nouvelles Méthodes Probabilistes pour l'évaluation des risques. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN 978-2-9521458-4-8. ISSN 1767-1175, avril 2010.

4. Fiches de compétences associées

Amélioration de la qualité d'un process industriel :

http://scmsa.eu/fiches/SCM_Qualite_Process.pdf

Méthodes robustes :

http://scmsa.eu/fiches/SCM_Methodes_robustes.pdf

Le contrôle de la qualité :

http://scmsa.eu/fiches/SCM_Controlle_Qualite.pdf

5. Contrats récents

- SNECMA Propulsion Solide, 2009-2010 : Méthodes probabilistes pour la fiabilité
- ArcelorMittal, 2011-2012 : Méthodes probabilistes pour la qualité d'un usinage
- Air Liquide, 2011 : Construction d'un "indice de proximité" entre pipelines
- Areva, 2013 : Hiérarchisation des paramètres intervenant dans un process industriel.
- DCNS, site d'Indret, 2013 : Soudage par Faisceau d'Electrons ; hiérarchisation de l'influence des paramètres par analyse probabiliste
- Solétanche-Bachy, 2015 : Hiérarchisation des paramètres influant sur la déformation d'un ouvrage d'art
- Carrefour, 2016 : Hiérarchisation des paramètres intervenant sur la vente de jouets
- Taxis G7, 2016 : Analyse probabiliste de bases de données et hiérarchisation de paramètres
- RTE, 2018 : Hiérarchisation des paramètres intervenant sur la durée de vie des équipements
- Eramet, 2018-2019 : Hiérarchisation des paramètres intervenant dans un process industriel (décarburation du manganèse)
- Johnson & Johnson, site de Sézanne, 2019 : Hiérarchisation des paramètres intervenant dans un process industriel
- Orano Mining, 2019 : Hiérarchisation des paramètres intervenant dans un process industriel (extraction de minerai)

- CEA, 2019 : Hiérarchisation de paramètres
- Arcelor Mittal Research, 2019-2020 : Amélioration d'un process industriel
- SARP Industries, site de Limay, 2019-2021 : Hiérarchisation des paramètres intervenant dans un process industriel
- Befesa-Valera, 2022 : Hiérarchisation des paramètres intervenant dans un process industriel