

# Exercice de statistiques descriptives

MACS3 – Institut Galilée

2021-2022

## Contexte

Le fabricant d'un système mécanique propose une garantie associée à son produit. Pendant une période d'exploitation il a fait évoluer son matériel et a distribué trois types de versions de ce produit. Pendant cette même période il a observé une série de pannes qui ont conduit à des retours en atelier pour réparation ou échange. Ces interventions sont à sa charge tant que dure la garantie, il cherche donc à optimiser son appareil en identifiant les éléments les plus impactant.

Nous disposons de deux tableaux de données récupérés pendant cet intervalle de temps. Ils sont stockés dans deux feuilles du fichier « data\_exercice.xlsx ».

Le premier tableau « Facteurs » (1) donne une série de pièces qui sont passées en atelier référencées par leur identifiant et le type de pannes (codées de A à I).

Tableau 1. Liste de retours en atelier.

	Version	D1	D2	Panne	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
Id																			
348606	v3	13665.09	11751.56	G	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
348762	v2	4888.38	3944.48	A	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
348855	v1	10273.96	13592.97	A	1.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
348906	v2	2149.26	3965.84	C	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
348946	v3	1.13	6.23	D	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
394339	v1	2398.99	2279.29	F	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
394414	v1	16258.44	15915.87	I	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
394490	v1	2141.35	1755.97	A	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
394590	v1	622.63	3266.30	A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
394609	v1	4052.18	4509.63	A	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0

461 rows x 19 columns

Associé à chaque événement, on dispose de la version de la pièce, de deux délais D1 et D2 correspondant à deux durées de fonctionnement (il y a deux modes de fonctionnement de l'appareil) et 15 diagnostics (X1 à X15). Chaque diagnostic correspond à un composant de l'appareil, il vaut 0 si aucun problème n'a été détecté sur ce composant et 1 si on repère un endommagement.

Le second tableau (2) « Cout » donne des coût en euros associés à des réparations.

Comme tous les ateliers ne fournissent pas le détail de leurs coûts il n'y a pas nécessairement un coût obtenu pour chaque réparation et on peut aussi avoir des informations manquantes.

Tableau 2.  
Impact financier.

	Panne	Cout
Id		
349483	A	15.84
350021	D	19.14
350212	B	9.90
351351	F	4.40
351351	F	28.16
...	...	...
388579	C	55.00
388579	C	4.40
388839	I	55.00
388839	I	55.00
394414	I	55.00

109 rows x 2 columns

## Questions

Le problème consiste à aider le fabricant à identifier les composants sur lesquels il devrait le plus s'interroger.

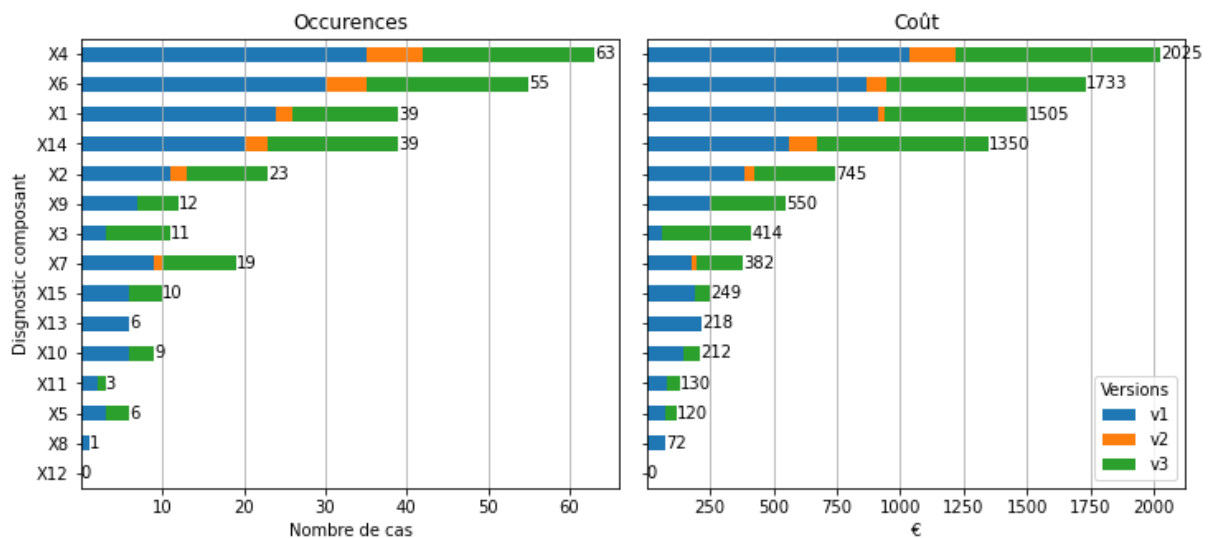
Le premier tableau peut permettre de trouver des relations entre les composants de l'appareil et les pannes observées. Il y a peut-être des relations intéressantes à observer.

A l'aide du second tableau on devrait pouvoir se faire une idée du coût de chaque panne, et en croisant les deux tableaux il devrait être possible d'identifier les composants les plus importants. On peut aussi confronter les nombres d'occurrences de dégradation des composants avec leurs impacts respectifs et regarder ce qu'il en est des différentes versions de l'appareil.

*A vous de jouer.*

*J'attends des modèles, des affichages, des statistiques... et des explications.*

*Par exemple vous pouvez essayer de reproduire le graphe suivant et en donner une explication.*



*Bien entendu j'attends aussi les codes !*