

Préliminaire

Quelques précautions à prendre en cas d'analyse statistique

Introduction

A few precautions to be taken in statistical analysis

Robert Chevrou

Direction Départementale de l'Agriculture et de la Forêt, Montpellier (France)

Les exemples cités ci-après ont été tirés de l'ouvrage de Joseph Klatzmann « Attention statistiques ! comment en déjouer les pièges » (*Cahiers Libres*, 405, éd. La Découverte, Paris, 1985, 126 p).

1. Attention aux effets mécaniques dans nos observations

p. 42 de l'ouvrage cité : « *durée moyenne de vie. La durée de vie moyenne des médecins est supérieure à celle de la population entière* » ; explication proposée : « *les médecins veillent plus et mieux au bon état de leur santé* ».

La durée de vie moyenne des académiciens est encore plus grande : bien écrire rend immortel !

Peut-être les explications proposées sont-elles pertinentes, mais les différences observées découlent aussi d'un phénomène mécanique : on devient médecin après de longues études, vers 25 à 30 ans ; les médecins déclarés ont ainsi échappé à la mortalité touchant les âges 0 à 25 ans, qui affecte la durée moyenne de vie de l'ensemble de la population.

Le phénomène est encore plus marqué pour les académiciens : on devient académicien à un âge très avancé, donc après avoir échappé à toute mortalité précoce.

Des phénomènes mécaniques peuvent toucher les données concernant les feux de forêt. Faute d'analyses approfondies par des statisticiens expérimentés, il n'est pas possible d'en dresser l'inventaire.

On peut penser que la densité locale des éclo-sions de feux, ainsi que les feux « simultanés », résultent de phénomènes mécaniques liés à la météorologie locale (sécheresse, temps chaud, et

The examples mentioned here are taken from the work by Joseph Klatzmann entitled Attention statistiques! comment en déjouer les pièges (Cahiers Libres, 405, éd. La Découverte, Paris, 1985, 126 pp.).

1. Beware of the mechanical effects in our observations

Page 42 of the book: "average length of life. The average length of life of doctors is greater than that of the population as a whole"; The explanation proposed is that "doctors take increasing better care of their own health".

The average length of life of members of the Académie Française (French Academy) is even greater. Writing well makes you immortal!

The explanations proposed may be pertinent but the differences also result from a mechanical phenomenon. One becomes a doctor at about 25 to 30 years old after long training. The doctors in the statistics thus escape deaths for the 0 to 25 year bracket that affect the average length of life of the population as a whole.

The phenomenon is even more marked for academicians. One becomes an academician at a very advanced age, thus after avoiding an early death.

Mechanical phenomena may affect data on forest fires. In the absence of in-depth analysis by experienced statisticians, it is not possible to draw up an inventory.

It can be considered that the local density of outbreaks of fire and "simultaneous" fires result from mechanical phenomena related to local weather conditions (drought, hot weather and wind), the

vent), à la concentration du public (qui sort par beau temps), à la topographie (relief qui joue sur le vent, exposition qui accentue la siccité de la végétation), et à la répartition spatiale de la végétation.

L'hypothèse de commando de pyromanes, si elle n'est pas à écarter, est en général peu vraisemblable. Elle a l'avantage d'éviter de s'engager dans des réflexions de fond et des études longues et coûteuses, mais l'inconvénient de nous faire prendre, à terme et faute d'en fournir la preuve, pour des benêts.

2. Attention aux relations farfelues, notamment l'évolution dans le temps des surfaces annuelles brûlées

L'ouvrage cité donne quelques exemples pour illustrer des relations farfelues, et la nécessité de corriger les données des variations saisonnières (p. 99).

Il faudrait, pour comparer entre elles des années successives, que les nombres de feux et les surfaces brûlées soient corrigés des variations climatiques (ou météorologiques) annuelles.

L'évolution des surfaces brûlées suit peut-être plus celle du climat que celle de l'efficacité de la prévention et de la lutte.

Un système de correction devra être établi pour le nombre d'éclousions, et un autre pour les surfaces brûlées. Il suffit que la végétation ne soit pas humide pour que le feu prenne ; il faut qu'elle soit très sèche, et que, de surcroît, le vent soit fort, pour que le feu puisse échapper aux pompiers.

Il faut, ou bien s'abstenir d'analyser l'évolution au cours du temps des nombres de feux et des surfaces brûlées, ou bien construire des systèmes de correction des variations climatiques.

3. Comparons ce qui est comparable

p. 33 de l'ouvrage cité : « *dans les années cinquante, l'URSS construisait 1 logement par minute* » (sous-entendu qu'on n'en faisait pas autant à l'Ouest).

Mais, avec un effort équivalent, le Luxembourg ne pourrait pas construire 1 logement par minute pour une population inférieure à 0,2% de celle de l'URSS ! et où aurait-on mis ces logements, et qui les aurait utilisés ? Pour des comparaisons raisonnables et raisonnables entre pays, on calcule généralement le nombre de logements construits pour 1000 habitants.

concentration of the public (who go out in fine weather), topography (the relief affects wind and exposition affects the dryness of plants) and the spatial distribution of plants.

Although the hypothesis of a gang of pyromaniacs cannot be rejected, it is generally fairly unlikely. It has the advantage of avoiding in-depth reflection and long and expensive studies but, after a while and when there is no proof, has the disadvantage of showing that we are stupid.

2. Beware of unlikely connections and especially the trend in time of the area burned annually

The book mentioned illustrates improbable connection with a few examples and shows the need to correct seasonal variation data (p. 99).

Comparison of two successive years would require that the number of fires and the burned areas should be weighted by the annual climatic (or meteorological) variations.

The trend in burned area may perhaps be governed more by the climate than by the effectiveness of fire prevention and fire-fighting.

A weighting system should be established for the number of outbreaks of fire and another for the burned areas. It is sufficient for vegetation not be wet for a fire to break out. It must be very dry and there must also be strong wind for a fire to be beyond control by fire-fighters.

It is necessary either to avoid analysing the evolution of the number of outbreaks of fire and the areas burned in time or to draw up correction systems for climatic variations.

3. Only comparable things should be compared

It is stated on p. 33 of the book that "the USSR built one dwelling per minute in the 1950s" (it being inferred that the same did not happen in the West).

However, with the same effort, Luxembourg could not build a dwelling a minute with a population of less than 0.2% of that of the USSR! Where would these dwelling have been put and who would have used them? The number of dwellings built per thousand people is generally used in reasoned and reasonable comparisons of countries.

Les comparaisons des nombres de feux et des surfaces brûlées entre régions, départements, ou pays, doivent être réalisées sur des bases similaires, notamment en se rapportant toujours à une même unité de référence, par exemple :

- nombre de feux pour 1000 ha de territoire ;
- nombre de feux pour 1000 ha de forêt ou de zone « naturelle » ;
- nombre de feux pour 1000 habitants ;
- nombre de feux pour 1000 unités de bétail ;
- nombre de feux pour 1000 points cumulés d'indice de risque météo ;
- idem pour les surfaces brûlées.

Et il serait bon que les données soient corrigées des variations climatiques.

Les données totales concernant des territoires (pays, régions, etc.) de superficies trop différentes tendent à biaiser les comparaisons, notamment dans leurs représentations cartographiques. On devrait s'astreindre à ne présenter que des données ramenées à la même unité de référence.

Comparisons of the numbers of fires and of burned areas between regions, administrative departments or countries should be performed on similar bases and should in particular refer to the same unit of reference, e.g.:

- the number of fires per 1000 ha of land;*
- the number of fires per 1000 ha of forest or "natural" area;*
- the number of fires per 1000 inhabitants;*
- the number of fires per 1000 head of cattle;*
- the number of fires per 1000 points of cumulated meteorological risk;*
- the same should be applied to burned areas.*

It would also be preferable to apply a climatic variation weighting to the data.

The overall data for zones (countries, regions, etc.) and areas that are too different tend to bias comparisons, especially when these are shown cartographically. Care should be taken to show only data that have been keyed to the same unit of reference.