

Danger Statisticiens !

Cette question a été posée sur un forum de mathématiques :

J'aimerais savoir quels sont les outils statistiques qui faudrait utiliser dans l'exemple suivant. On a réalisé des relevés concernant une espèce, chaque année, pendant 12 ans (voir données ci-après). On aimerait savoir, si les différences d'effectifs obtenues sont dues au simple hasard d'échantillonnage ou bien s'il y a de réelles différences. Comment, doit-on procéder ? Voici le tableau.

Année	Transect 1
2004	12
2005	7
2006	10
2007	10
2008	10
2009	8
2010	7
2011	23
2012	10
2013	19
2014	37
2015	26
somme	179
moyenne	14,91666667

Et l'information complémentaire suivante :

Concernant le protocole, il s'agit de relevés réalisés à l'identique chaque année (comptage bien spécifique le long d'un transect bien défini). Le but étant d'étudier une population de papillon et voir les évolutions de ces effectifs. L'idée générale est de savoir si les différences observées, en fonction des années, montrent une évolution de la population quelle soit positive ou négative. Ou alors, si ces différences ne sont pas significatives et uniquement dues à un delta lié à l'échantillonnage.

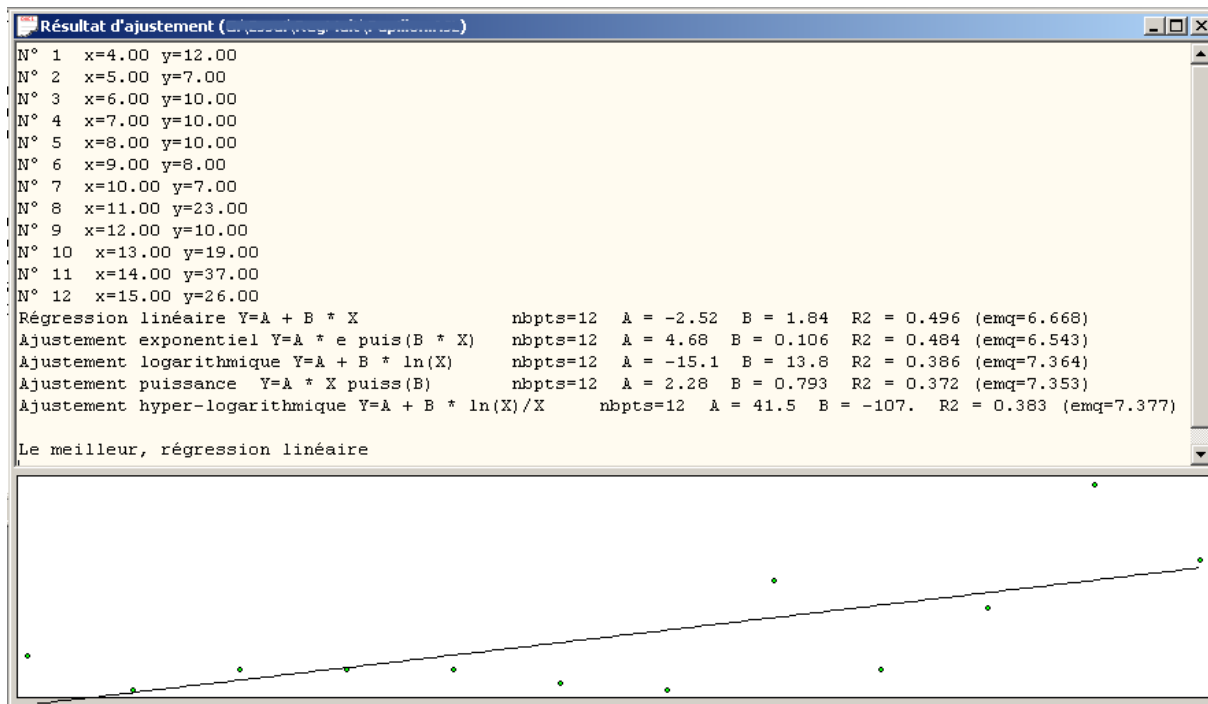
La question posée est très claire : " peut-on déduite de ces valeurs une évolution ou les différences observées ne sont due qu'au hasard ?"

Première hypothèse, les observations montrent une évolution. Alors, on cherche un fonction qui permet de chiffrer cette évolution. Les méthodes sont bien connues : par régression linéaire, on cherche une fonction simple qui minimise le carré des écarts entre les valeurs observées et les valeurs calculées, pour chaque année.

Pour une série de données, X, Y, le logiciel calcule 5 formules de régression et pour chacune le coefficient de détermination.

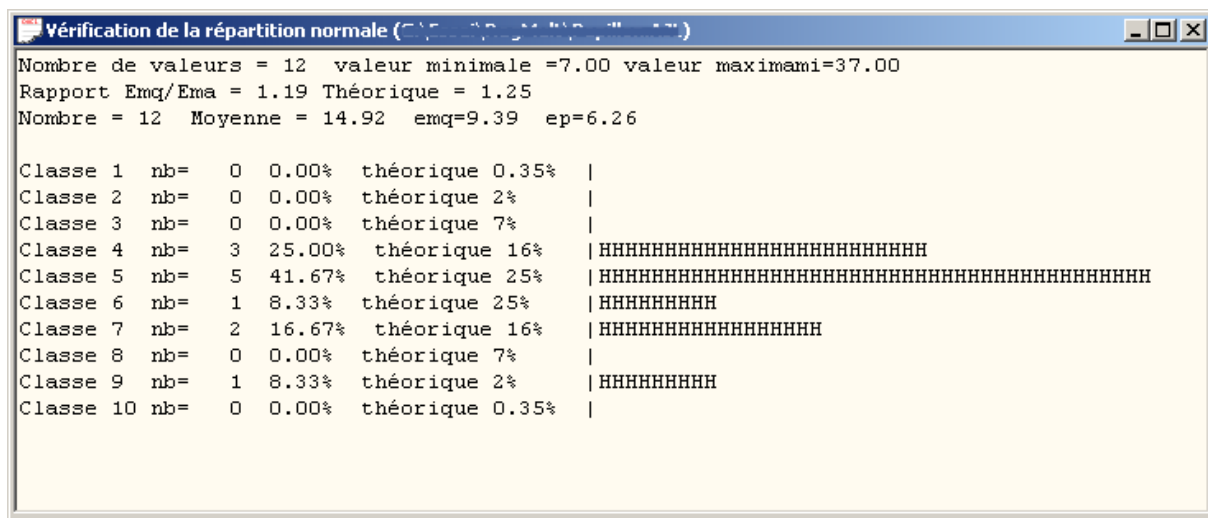
La meilleure formule est celle qui donne un coefficient de détermination le plus proche de 1. On observe généralement que les coefficients de détermination sont supérieurs à 0.8 et souvent supérieurs à 0.9.

Par contre, un coefficient de détermination inférieur à 0.5 indique qu'il n'y a pas vraiment de rapport (ou corrélation) entre les différents couples d'observation. Dans le cas présent, c'est d'autant plus vrai que les 5 fonctions testées donnent des coefficients de détermination assez voisins.



Autre approche : peut-on considérer que ces observations ont un caractère aléatoire, c'est à dire, ne dépendant que du hasard ?

Principe de la méthode : soit une liste d'observations, on cherche à savoir si chaque élément dépend de la même loi de probabilité, en l'occurrence, ne dépend que du hasard.



Il est bien évident que, étant donné le petit nombre d'observations, la répartition suivant les classes n'a pas tout à fait l'aspect que l'on observe avec un grand nombre de valeurs.

Cependant le résultat est suffisamment caractéristique pour conclure que les écarts entre ces observations ne sont dus qu'au hasard, ou pour être rigoureux, qu'on ne peut pas conclure le contraire.

Conclusion.

La question a été clairement posée, et, probablement, il ne s'agit pas d'un exercice, mais d'une question posée par un professionnel, sur un forum de mathématiques. On peut faire ce que l'on veut avec des chiffres, et en l'occurrence, les réponses qui ont été données sont plus celles attendues par un professeur pour son énoncé théorique que celles attendues d'un professionnel de la part d'un professionnel.