Examination for UE3.3 « Refresher courses » of master BME-BIN: Statistics Applied to Biology

Isabelle Rivals, Équipe de Statistique Appliquée, ESPCI Paris Tech

Duration: 40 minutes. Pocket calculator and lecture notes are authorized.

1 Fight against doping

We are concerned with the detection of the anabolic steroid nandrolone in the professional sports world. The prevalence of its use in the environment is 1%, the sensitivity of the test is 99% but it also generates 5% of false positives, see the following citation[1]:

> Dans le cas d'un produit dopant anabolisant comme la nandrolone ou de toute substance qui réunit les mêmes conditions supposées à savoir une prévalence de consommation de 1% et une méthode de détection sensible à 99% et spécifique à 95%, la probabilité d'une consommation dans un délai compatible avec le moment de prélèvement avant la détection en laboratoire n'est que de 83% soit 17% d'erreurs judiciaires ou disciplinaires prévisibles. En vertu de la loi des grands nombres, il y aura assurément 17 déclarations erronées de culpabilité sur 100 échantillons examinés si on ne dispose d'aucun autre indice. Et, pour ne pas arranger les choses, plus la prévalence de la consommation est faible plus le nombre de déclarations de culpabilité erronées augmente.

a) Compute the probability that a sports professional has used nandrolone given that he was positive to the test.

b) The conclusion of the above citation is that there will be 17% of judiciary errors. What do you think about this conclusion?

2 Cell size statistics

a) A biologist measures the sizes $\{y^k\}$ of $n = 6$ cells (arbitrary unit):

7.98 2.03 1.49 2.45 0.60 1.22

Provide the expressions of the estimates of the mathematical expectation and of the variance of the cell size, as well as their values.

b) Provide a 95% confidence for the mathematical expectation of the size. What are the underlying assumptions?

c) What is wrong with this confidence interval?

d) As a matter of fact, the sizes are log normally distributed, i.e. their logarithm is normally distributed. The natural logarithms of the above values, i.e. the $\{z^k = ln(y^k)\}$ are:

2.0769 0.7080 0.3988 0.8961 -0.5108 0.1989

where $\bar{z} = \frac{1}{n}\sum_{k=1}^{n} z^k = 0.6280$, and $\sum_{k=1}^{n}(z^k - \bar{z})^2 = 3.7113$. Provide a 95% confidence interval for the mathematical expectation of the natural logarithm of the cell size.

e) Deduce a confidence interval for the typical cell size.

[1] Comment valider une analyse, Dr. Yves Jacomet, 7eColloque National de la Fondation Sport Santé (16-17 mars 2007, Paris).