ANTHROPOLOGIE • XXXIX/2–3 • pp. 187–189 • 2001

SANTIAGO L. IPIÑA, ANA I. DURAND

SEXUAL DIMORPHISM, *MI* INDEX, STUDENT'S *t* AND SNEDECOR'S *F*

ABSTRACT: An index of sexual dimorphism for populations distributed as mixture models of two normal components has been proposed by the authors. Our proposal is a function of the five parameters characterizing the mixture model, that is, two mathematical expectations, two variances and one mixing proportion. That our index depends on a proportion is not always perceived correctly, so in this article we show that sample sizes, or proportions as lineal functions of sample sizes, influence on such common sample statistics as Student's t or Snedecor's F and the decisions that are made through them.

KEY WORDS: MI Index – Proportions – Sample Statistics

The authors of these comments have proposed an index (*MI*) of sexual dimorphism in populations that are distributed according to a probabilistic model denominated mixture model of two normal components (see Ipiña, Durand 2000). Our proposal is defined as a function of the five parameters that characterize the normal mixture, and it measures the overlap area between the two functions that represent the contribution of each sex in the population. Such parameters are two mathematical expectations, two variances and the proportions that correspond to each one of the sexes that make up the population. One proportion is the complement of the other so, only one of them has to be taken into account.

Since the population we are considering is a mixture of two normal populations, a mixture that brings about a new population, hence with probabilistic properties that are not present in each one of the two normal populations, it seems reasonable, even compelling, that the mixing proportion is one of the parameters that characterize the mixture.

A consequence is that such a parameter, alone, is able to discriminate two mixtures, i.e. two normal mixtures are different when they have the same means and variances but different mixing proportions. It is obvious, on the other hand, that this implies the fact that both mathematical expectations as well as variances behave in the same way, so that, for example, two normal mixtures are different when they have the same values for all parameters except the mean of one of the sexes.

However, that a mixing proportion exhibits such a peculiarity in discriminating mixtures, as mathematical expectations and variances do, is not always perceived correctly. Thus, when we analyze two normal mixtures that only differ in their mixing proportions, it may be assumed that any measure dealing with sexual dimorphism should be the same in both mixtures. Of course this is the same as saying that sexual dimorphism is related to means and variances and not to mixing proportions.

A problem arises when sexual dimorphism is analyzed from a probabilistic standpoint, i.e. when random variables, distribution functions, sample functions or hypothesis tests are involved. Such a problem is that proportions play an important role in the definition of some common statistics and the decisions that are made through them.

In effect, the estimation of a mixed proportion is a direct consequence of the sex sample sizes. Let us analyze accordingly whether or not sample sizes influence the decisions that are made in problems where hypothesis tests are concerned.

Let us suppose we have two random samples, selected from normal populations, and we feel that one of these samples has been extracted from a population with a mathematical expectation greater than the one of the other population involved. Thus, we are interested in testing:

$$H_0 : \mu_1 - \mu_2 = 0\,,$$
$$H_1 : \mu_1 - \mu_2 > 0\,.$$

Values of the sample mean, variance and size are 16, 4 and 2, respectively, in the first random sample, whereas such values are 11, 9 and 3 in the second random sample. Since the statistic we are dealing with is a quotient whose distribution function is a Student's *t* with degrees of freedom depending upon the fact that variance parameters are considered equal or not, it is mandatory to first construct and solve the test:

$$H_0 : \sigma_1^2 = \sigma_2^2\,,$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2\,.$$

As is well-known, this test is carried out by means of the quotient of sample variances statistic, distributed according to a Snedecor's *F* distribution with one and two degrees of freedom. Critical values for such a test are 0.0125 and 38.51 so that, with significance level α = 0.05, we accept the null hypothesis and conclude that our problem is homocedastic.

Consequently, we have to use the statistic distributed according to a Student's *t* distribution with three degrees of freedom. The rejection region of the test concerning mathematical expectations has, in this way, 2.353 as a lower bound so that we accept the null hypothesis since the sample value for the implied statistic is 2.0225. This means we have two samples randomly selected from one normal population.

Let us now suppose we have two random samples that are the same as the previous ones but with different sample sizes. Namely, the first sample has sample mean and variance of 16 and 4, respectively, but size 16, and the second sample has sample mean and variance of 11 and 9, but size 6. Given that sample sizes have changed, it is mandatory to perform the above test concerning variances again, so that critical values for a test with α = 0.05 are now 0.279 and 6.43 for a Snedecor's *F* distribution with fifteen and five degrees of freedom. The decision is again the acceptance of the null hypothesis so we are still considering a homocedastic problem.

The statistic that is involved in the test concerning mathematical expectations, is now distributed according to a Student's *t* distribution with twenty degrees of freedom and defines a critical value that is 1.72. Since the sample value for such a statistic is 4.5584, our decision now is to reject the null hypothesis and to conclude that the two random samples analyzed were not selected from a unique population.

As a result we can see that we have made two opposite decisions in spite of the fact that the only difference between the two above-mentioned experimental situations is in the sample sizes. We would like to point out that the two above examples are by no means exceptional or elaborate. The reader is kindly asked to carry out the same calculations as above, with a random sample whose mean is 16, whose variance is 6.25 and whose size is 8, and another random sample whose mean, variance and size are 14, 9 and 6, respectively. Our reader will conclude that the null hypothesis is accepted in the case where the test to solve is the same as the above-mentioned test concerning mathematical expectations. However, if the random samples to analyze are the same as the former, but with sample sizes 18 and 8, respectively, the null hypothesis is rejected.

In fact, sample sizes influence on the Student's *t* or Snedecor's *F* distribution functions in such a way that the degrees of freedom of these distributions are a consequence of these sample sizes, as is well-known. Furthermore, the above decisions involving the statistic with a Student's *t* distribution can be easily obtained by considering the statistic with a Snedecor's *F* distribution.

In effect, let 15 and 5 be the variances of two random samples extracted from normal populations. Suppose we are interested in testing whether or not such samples come from homocedastic populations and, initially, the sample sizes are 8 and 13. With significance level α = 0.05, the corresponding statistic with a Snedecor's *F* distribution of seven and twelve degrees of freedom delimits an acceptance region with minimum endpoint equal to 0.2141 and maximum endpoint equal to 3.61. As a consequence we accept the null hypothesis, concluding that the two random samples have been selected from a unique population as far as the σ^2 variance parameter is concerned.

If we now have two random samples with variances, again 15 and 5, but sizes 11 and 21, respectively, the decision to make is to reject the null hypothesis so that we are now analyzing a heterocedastic problem. In effect, the minimum and maximum endpoints of the acceptance region are now 0.29239 and 2.77.

Thus, we see that sample sizes are directly implied in decision-makings such as those mentioned above. It is clear that, from a mathematical standpoint, sample sizes influence the values obtained in statistics such as the one with a Student's *t* distribution. In effect, when two random samples selected from normal populations, are analyzed and it is felt that a homocedastic problem is concerned, the statistic definition, as a quotient of a standardized normal random variable and the square root of a chi-square random variable with $n_1 + n_2 - 2$ degrees of freedom divided by such degrees of freedom (and provided that these random variables are independent), is quite straightforward,

$$T = \frac{\overline{X}_1 - \overline{X}_2 - d_0}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}},$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

with d_0 some specified value in the null hypothesis, and for a statistic distributed according to a Student's *t* model with $n_1 + n_2 - 2$ degrees of freedom. This means we have a sample function concerning two sample means, two sample variances and two sample sizes which, in turn, implies that the value of *T* will change as some of the sample means or variances, or some of the sample sizes change.

To conclude, if there are some doubts about the reliability of the *MI* index of sexual dimorphism because it is a function involving the mixing proportions of the normal mixture, then such doubts should be extended to such common statistics as the ones with the Student's *t* or the Snedecor's *F* distributions, and to the decisions that these statistics generate. We do not think that such a state of affairs is the most reasonable one.

REFERENCES

IPIÑA S. L., DURAND A. I., 2000: A measure of sexual dimorphism in populations which are univariate normal mixtures. *Bull. Math. Biol.* 62: 925–941.

Santiago L. Ipiña
Matemática Aplicada
Facultad Biología
Universidad Complutense Madrid
28040 Madrid
Tel.: +34 91 394 5060
Fax: +34 91 394 5051
E-mail: ipina@eucmax.sim.ucm.es