A DIDACTIC PROPOSAL FOR "STATISTICAL AND NUMERICAL METHODS", AN OPTIONAL COURSE IN THE LAST YEAR OF HIGH SCHOOL IN GALICIA (SPAIN) ®

Salvador Naya, Ricardo Cao
Universidade da Coruña, Spain
Aurora Labora
IES Alfredo Brañas, Carballo, A Coruña, Spain
Matilde Ríos
CPI Cruz do Sar, Bergondo, A Coruña, Spain

Some examples from the teaching proposal, elaborated by us, for the subject "Statistical and Numerical Methods" (SNM) are presented. SNM is an optional course in the last year of High School in Galicia (Spain). More specifically, we are concerned with the introduction of some new concepts at this teaching level in the Galician education system, namely, Markov chains, statistical inference and time series.

INTRODUCTION

When Baía Publishing Company entrusted us with writing a text book on the new optional course SNM (see Cao, Labora, Naya & Ríos, 2001) we faced the real challenge of how to introduce some "difficult" content that would be taught, for the first time, to Galician students at a High School level. After thinking about the issue we decided to take a simple example as a starting point (in order to motivate the student) and then go through the different concepts of the didactic unit using the selected familiar example.

Our method emphasizes applications of the statistical techniques rather than theoretical formalism. To do this, we worked on a situation that happens to a fiction character, Don Anselmo, whose problem has a statistical nature related to the concepts that will be introduced in the rest of the unit. This character is used as some link from chapter to chapter along the book.

MARKOV CHAINS

To introduce the concept of a Markov chain we start with the following problem. Don Anselmo always walks from home to his office and also on his way back. To prevent getting wet when it rains he has two umbrellas and decides to follow some plan. Before leaving for work, early in the morning, he looks at the sky and, if it is raining, he takes one umbrella with him (if there is one available at home). In such a case he puts his umbrella in the stand at his office. When he returns home he looks at the sky again and takes one umbrella if necessary (provided that there is one availablein his office). Don Anselmo lives in Galicia, a wetland in the northwest of Spain where the probability of rain falling (when he leaves home or his office) is 1/2.

In this setup, the number of umbrellas in the same place where Don Anselmo is (at home or in his office) can be described using a Markov chain. The states of the chain are 0, 1 and 2, since these are the possible numbers of umbrellas. We denote by stage 1 the instant when Don Anselmo leaves home (the first day when the whole process starts), by stage 2 when he arrives his office that day, by stage 3 when he arrives home (which is equivalent –in terms of number of umbrellas– to the instant when he will leave the next day) and so on. The whole process is a compound experiment with an undefined number of stages. Furthermore, it is obvious that the probability that Don Anselmo has 0, 1 or 2 umbrellas with him, given the number of umbrellas in previous stages, only depends on the number of umbrellas in the last stage. More precisely, if he had 0 umbrellas in stage n (let's assume, at home) it is clear that in stage n+1 (in his office with our assumption) will be 2, with or without rain. If Don Anselmo has 1 umbrella in stage n, he may have 1, as well, in stage n+1 (provided that it does not rain, with probability 1/2) or 2 (provided that it rains, with probability 1/2). Finally if he has 2 umbrellas in stage n, he may have 0, with probability 1/2 (no rain) or 1, with the same probability (rain), in stage n+1. We have then that the number of umbrellas that Don Anselmo has with him is a homogeneous Markov chain with transition probabilities matrix:

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

In the remainder of the unit we keep on introducing new concepts using the same example. For instance, if Don Anselmo has 2 umbrellas at home (before going for work) the probability that, in three days, he will not have any umbrella when he returns home is:

$$p_{20}^{(6)} = \frac{7}{64} = 0.1094$$

To compute this we explicitly find the sixth power of the transition matrix and ignore the possibility of using the diagonalization of the transition matrix to compute the n-th power of the matrix ($P^n = H \cdot J^n \cdot H^{-1}$). The reason is that these algebraic techniques are beyond the knowledge of the student at this High School level.

STATISTICAL INFERENCE

We introduce the main concepts of statistical inference (more specifically the sampling distribution of the sample mean) using a trivial population of the four boys and girls living in the same flats building as Don Anselmo. Their names and heights (in centimetres) are given in the first two columns of Table 1:

Table 1.
Trivial Population of the Four Boys and Girls

Name	Height (cm)	Sample	Mean height($\overline{X}$)
Xoan	180	(Xoan, Uxía)	172
Uxía	164	(Xoan, Bieito)	170
Bieito	160	(Xoan, Iria)	179
Iria	178	(Uxía, Bieito)	162
		(Uxía, Iria)	171
		(Bieito, Iria)	169

The random variable X= "height of a boy/girl neighbour to Don Anselmo" has the following probability mass:

$p_X(160) = 1/4,\ p_X(164) = 1/4,\ p_X(178) = 1/4,\ p_X(180) = 1/4.$

The population mean and variance of this small population are:

$$\mu = \frac{180+164+160+178}{4} = 170.5$$

$$\sigma^2 = \frac{180^2 + 164^2 + 160^2 + 178^2}{4} - 170.5^2 = 29145 - 29070.25 = 74.75$$

Let's assume that Don Anselmo ignores these values and decides to estimate the mean height by using a sample of size 2. Under a random sampling without replacing the six possible samples of size 2 are: (Xoan, Uxía), (Xoan, Bieito), (Xoan, Iria), (Uxía, Bieito), (Uxía, Iria) and (Bieito, Iria). Hence, the sample mean (as a random variable) has a discrete distribution with the following probability mass:

$$p_{\overline{X}}(162) = 1/6,\ p_{\overline{X}}(169) = 1/6,\ p_{\overline{X}}(170) = 1/6,$$
$$p_{\overline{X}}(171) = 1/6\ p_{\overline{X}}(172) = 1/6,\ p_{\overline{X}}(179) = 1/6.$$

We compute the expectation of the sample mean:

$$E\left(\overline{X}\right)=\frac{172+170+179+162+171+169}{6}=\frac{1023}{6}=170.5$$

to obtain the bias: $E\left(\overline{X}\right)-\mu=170.5-170.5=0$.This means that $\overline{X}$ is an unbiased estimator of μ.By similar calculations we obtain $Var\left(\overline{X}\right)=MSE\left(\overline{X}\right)=24.197$.

TIME SERIES ANALYSIS

Finally in the didactic unit devoted to time series analysis we start with a weekly series of the number of attendants (per day) to the English school where our character goes every day. The manager of the school records the number of students per day during a period of four weeks to study the time evolution of these numbers. The values are presented in Table 2.

Table 2

Daily Variations in the Number of Students

Monday	Tuesday	Wednesday	Thursday	Friday	Weekly
97	85	101	103	118	504
98	85	99	102	113	497
96	84	100	103	118	501
99	86	102	104	119	510

One can observe in the previous table that the day of the week makes a real difference in the number of attendants to the English school. We check it by computing the seasonal indices. This is done by the method of daily average ratios with respect to the tendency. We obtain the following table:

Table 3

Seasonal indices and intermediate calculations

	Monday	Tuesday	Wednesday	Thursday	Friday	Average
x_{1j}	97	85	101	103	118	100.8
x_{2j}	98	85	99	102	113	99.4
x_{3j}	96	84	100	103	118	100.2
x_{4j}	99	86	102	104	119	102
$\overline{x}_{\bullet j}$	97.5	85	100.5	103	117	
$\overline{x}'_{\bullet j}$	97.5	84.912	100.324	102.736	116.648	$\overline{x}'_{\bullet\bullet}=100.424$
s_{ij}	-2.924	-15.512	-0.1	2.312	16.224	

Similarly, we may compute some other autocorrelation coefficients and make a plot with the correlogram, as well as the tendency component, the seasonal indices, etc.

REFERENCES

Cao, R., Labora, A., Naya, S., & Ríos, M. (2001). *Métodos Estatísticos e Numéricos*. Baía Edicións.