Gambling as a teaching aid in the introductory physics laboratory

L B Horodynski-Matsushigue†, P R Pascholati, V R Vanin, J F Dias‡, M-L Yoneama§, P T D Siqueira, M Amaku and J L M Duarte

Instituto de Física da Universidade de São Paulo, Caixa Postal 66318, CEP 05315-970, São Paulo, SP, Brazil

Abstract. Dice throwing is used to illustrate relevant concepts of the statistical theory of uncertainties, in particular the meaning of a limiting distribution, the standard deviation, and the standard deviation of the mean. It is an important part in a sequence of especially programmed laboratory activities, developed for freshmen, at the Institute of Physics of the University of São Paulo. It is shown how this activity is employed within a constructive teaching approach, which aims at a growing understanding of the measuring processes and of the fundamentals of correct statistical handling of experimental data.

1. Introduction

There is by now a consensus in the physics teaching community that the main concepts of the statistical theory of uncertainties (formerly known as statistical theory of errors) are far from easy to grasp [1]. On the other hand, there is an increasing awareness, also from outside the community, about the necessity of a clear understanding of the informational content of experimental data [2]. The initial training, with the aim of acquiring a critical view of measuring procedures and the resulting data, is advantageously done in the more easily controlled situation of experimental physics, where repetitions of measurements are mostly possible. In fact, in the biological sciences, and even more so in the social sciences, there are usually many more parameters of importance, whose influence on the outcomes may frequently only be estimated by a mind trained to look for regularities and irregularities. In the last few years, a reformulation of the didactic approach to the statistical theory of uncertainties was undertaken at the Institute of Physics of the University of São Paulo [3]. Two aspects are now viewed as being of prime importance: *first*, give the students the opportunity to *feel* the influence of randomness on the practical outcome of experiments prior to telling them the *rules* which are derived from it; *second*, base the whole synthesis of what the students have *seen* on a comparative discussion (guided by the teachers) of the real results of the whole class or even the totality of classes

---

† E-mail: lighia@if.usp.br
‡ Permanent address: Instituto de Física – UFRGS, CP 15051, 91501-970, Porto Alegre, RS, Brazil.
§ Permanent address: Centro de Ciências Exatas – UNISINOS – São Leopoldo, RS, Brazil.

337

(10) of the freshmen majoring in Physics. In this way a constructive approach [4] is pursued and the students are led to look at their data in a critical and comparative way.

Dice throwing, as the second activity in class, in a sequence of specially programmed laboratory works, has shown a variety of didactically interesting features which will be presented in the following sections. However, in order to take advantage of all the potentialities of such an activity, it is essential that the students have already been introduced to the statistical fluctuations of measured values. This is worked out in São Paulo in two ways: (i) in the inaugural lecture, which is in fact their first contact with the University, students are asked to read their watches simultaneously; these data are collected, histogrammed and presented to the students, shortly afterwards; (ii) in the first scheduled laboratory activity, the timing of the period of a pendulum by pairs of students also results in histograms, which the students are instructed to analyse qualitatively in a comparative way.

2. Dice throwing and the statistical expectation

Perhaps the most important issue in teaching the statistical treatment of data is to make it clear to the beginner what is meant by a statistical expectation. Dice throwing is familiar to almost everyone and can be used as a counterpart of a real experiment if conveniently exploited. At the Institute of Physics of the University of São Paulo special dice with two faces marked with a *spot* are in use, but for most purposes commercial dice, if of the same production lot, should be adequate. For almost all students it is immediately clear that the long-term expectation, or probability, of having an ace (a spot showing up) is

$$p = \tfrac{2}{6} = \tfrac{1}{3} = 0.33\overline{3}. \tag{1}$$

The experimental assessment of this expectation is, therefore, the main problem presented to the class at the very beginning of the activity.

To allow for the comparative appreciation of all data, the throwing of 10 dice at a time and the counting of the number y ($y = 0, 1, 2, \ldots, 10$) of marked faces showing up (aces or successes) is, after some initial trials (and the ensuing discussion with the class), put as a common methodology for all students. It is agreed upon to always look for indications of significant differences in the outcomes, aiming at testing the fundamental hypothesis, which is to take all dice as equal. Now, the mean value expected for y in the long term turns out to be $10p = 3.3\overline{3}$.

Subsections 2.1 to 2.3, which follow, are organized so as to highlight, in a sequence of growing difficulty and through the use of the students' results, statistical concepts which are cornerstones for the later application of the statistical theory of uncertainties to measurement results. Although the present work is heavily based upon the outcomes of the gambling activities of the 1996 freshmen, some results of previous years are also included.

2.1. The meaning of a limiting distribution

It is surely not feasible to come close to the limiting distribution with real data due, among other factors, to the impossibility of maintaining the same measuring conditions for long periods. So, the existence of a limiting distribution is ultimately an act of faith, which enables one to work, in practice, with its consequences. If the hypothesis of *equal dice* can be made for a large collection of them, it is easy to collect up to some 10^4 events of simultaneous throws of 10 dice, if one has a group of 100–200 people. Therefore, one is able to study what happens to the mean, $\overline{y}$, which is the best approach to the long-term expectation of $10p$ when the sample size, N, is increased. It is also convenient to compare the outcomes of several statistically equivalent samples, which is a means of putting the fluctuations inherent to such processes into perspective. In practice, an option was made for having the gamblers collecting the data of a final $N = 500$ sample in smaller samples; to begin with, five times $N = 10$ throws of 10 dice, and, in sequence, a further nine collections of $N = 50$ repetitions of throws. Students

are instructed to plot graphs of their cumulatively increasing samples and also to compare them with the other samples of the same size already obtained. Thus, initially they compare their $N_y \times y$ (N_y being the absolute frequency of each of the y possible successes) graphs for samples with $N = 10$ and $N = 50$ and, at the end, the several $N = 500$ samples obtained in each class are jointly discussed on the blackboard. This procedure induces students to note that something like a limiting distribution is to be expected and is successively better defined as N increases. The final samples of the several classes ($N = 4 \times 10^3$) and the total sample of the classes ($N \approx 5 \times 10^4$) are afterwards made available to the students and their absolute frequencies are compared with the results of the binomial probability distribution, $P_n(y)$ for $n = 10$ trials (dice) and probability $p = 0.33\overline{3}$:

$$N_y = N \, P_{10}(y) = N \, \frac{10!}{y!(10 - y)!} p^y (1 - p)^{10-y}. \tag{2}$$

Figures 1(c) and 1(d) respectively show the N_y versus y graphs obtained by eight different classes for $N = 4000$ and the global result, $N = 46\,500$, of all classes, the latter being compared with the theoretical expectation (equation (2)). Figure 1 also displays examples of comparisons of eight samples of $N = 50$ (figure 1(a)), collected by one team of students, and of $N = 500$ (figure 1(b)), each of which is the final result of one team of the class. It can be seen that even these $N = 500$ samples are still rather variable and that only the global sample, obtained by the totality of the 1996 freshmen, seems distinctly equivalent to the binomial expectation. Statistically allowed fluctuations in each channel (indicated as approximately 68% limits, forming a belt in figures 1(a) and 1(b)), must obviously be taken into account. These results contain one of the most important messages to be got across: do not expect your histograms of data collected in actual physics experiments to look *identical* to the expected probability function, even for *very large* samples of $N \approx 50$, in spite of a possible limiting distribution of, for instance, Gaussian shape.

To complete the next step, students are further asked to calculate means, $\overline{y}$, for their samples with different N and to compare $p_{ex} = \overline{y}/10$ with the probability of obtaining an ace in a single throw, p (equation (1)). Table 1 shows results of p_{ex} for the data reproduced in figure 1 and also for eight of the $N = 10$ samples. The almost correct number of significant figures for $\overline{y}$ emerges automatically for the sample sizes chosen and is used in the first place to indicate the increasing information content of p_{ex}, when calculated from increasing samples. The technique always employed in the freshman laboratory is to compare, on the blackboard, the results of the various teams in each class, as they are being obtained: the teacher, now and then, calls for a pause in the activity and promotes a group discussion. This procedure, besides uncovering mistakes, offers in this activity opportunity to convince the students that the values *normally* differ from each other. In particular, they realize the convenience of taking care of significant figures and, especially, not to neglect zeros at the right of the number.

Besides the averages of the eight values of p_{ex} ($\overline{p}_{ex}$), table 1 also presents the standard deviations for each N, $s_p(N)$, which represent the fluctuation of the particular p_{ex} values tabulated about their respective mean. The kind of global overview of the results exemplified by the last columns of table 1 is discussed with the students in a special activity [3] called a *synthesis class* which is delivered after about two months of experimental activities. By the time this *synthesis class* is presented, the students have already been exposed to two further experimental activities dealing with measurements. An analysis of table 1, as done in the *synthesis class*, demonstrates that, as expected, the significant figures of the results increase rather slowly. On the other hand, after a thorough examination of the table, students clearly grasp the information contained in the significant figures, so that a value of $p_{ex} = 0.42$ may be a possible outcome of a $N = 10$ sample, but a value of $p_{ex} = 0.420$ should not be a result of $N = 500$; that for the $N = 4000$ samples of each class the first two figures are surely 0.33 but the fluctuation affects the third figure of p_{ex}, meaning that the results of classes 1 and 8 are *no better* than those of class 3 or, in particular, of class 6.
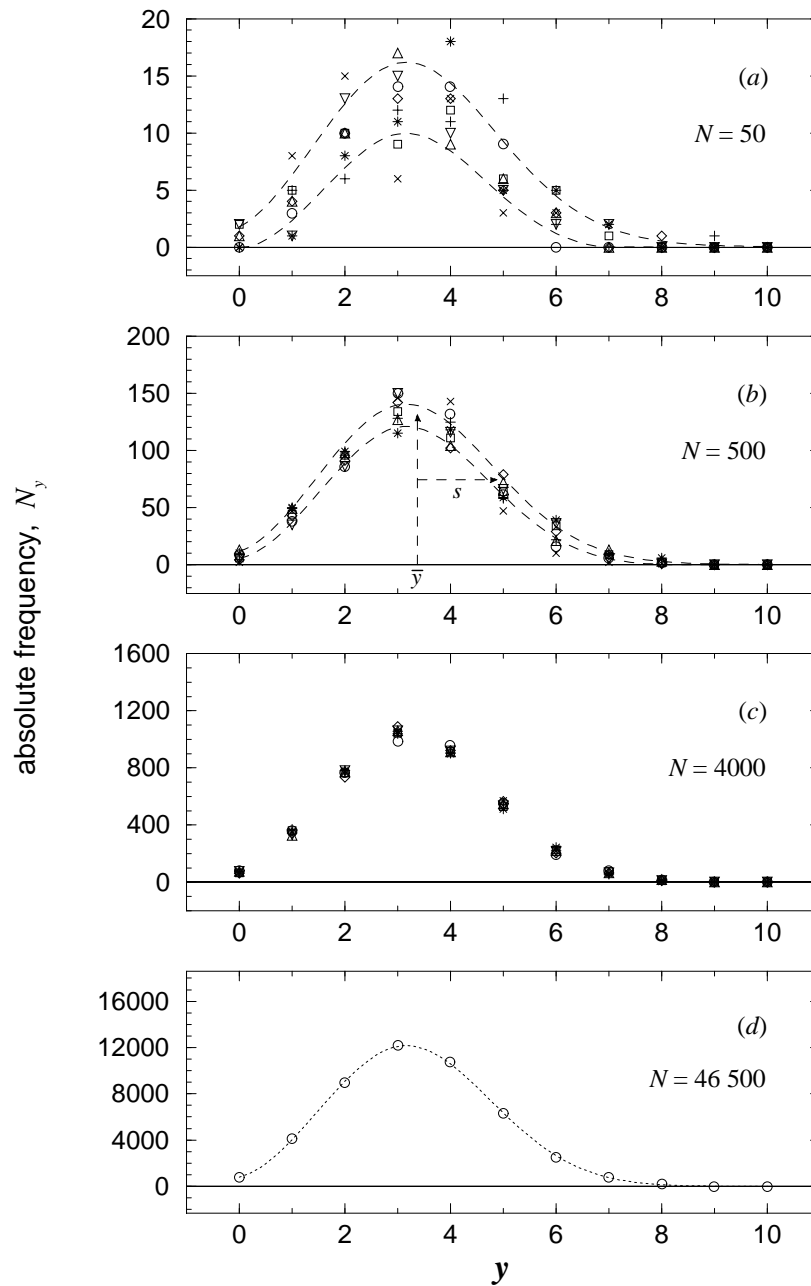
Figure 1. Absolute frequency N_y, of y sucesses, in the throwing of 10 dice with two marked faces, for eight different samples obtained by: (*a*) one team of students throwing 50 times; (*b*) several teams, each totalling 500 times; (*c*) eight different classes with $N = 4000$. (*d*) The 1996 global result, for $N = 46\,500$, compared with the binomial expectation shown as the dotted line connecting the calculated points. In the $N = 50$ and $N = 500$ cases the limits within which approximately 68% of the experimental frequencies are statistically expected to fall are shown by upper and lower dashed lines. A different symbol is used for each sample. In the samples with $N = 500$, (*b*), one typical mean value, $\overline{y}$, and standard deviation, s, are respectively indicated by vertical and horizontal dashed lines.

Table 1. Experimental probability, p_{exi} ($i = 1, 2, \ldots, 8$) for a dice to show an ace, obtained from eight samples of $N = 10, 50, 500, 4000$, and one sample of 46 500 throwings of 10 dice with two marked faces compared with the value expected for $N \to \infty$, p. Also indicated in the two last rows are the corresponding *mean values* $\overline{p}_{ex}$ and a measure of the dispersion of the eight values, $s_p(N)$.

			N			
	10	50	500	4000	46 500	∞ (expected)
p_{ex1}	0.42	0.332	0.329	0.3331	0.332 884	$0.333\,33\overline{3}$
p_{ex2}	0.28	0.334	0.337	0.3342		
p_{ex3}	0.35	0.332	0.340	0.3348		
p_{ex4}	0.28	0.318	0.329	0.3344		
p_{ex5}	0.26	0.326	0.342	0.3324		
p_{ex6}	0.37	0.366	0.331	0.3295		
p_{ex7}	0.36	0.310	0.315	0.3318		
p_{ex8}	0.33	0.382	0.338	0.3331		
$\overline{p}_{ex}$	0.331	0.338	0.333	0.3329		
$s_p(N)$	0.055	0.024	0.009	0.0017		

It is also important to stress that, without further information, it is impossible to know if the global result for p_{ex}, shown in the fifth row, is compatible with the expected value.

2.2. The width of a distribution as estimated through the standard deviation

Besides the concept connected with the proper existence of a statistical distribution of data, another difficult concept [1] refers to the meaning attached to its width. Most students see the width of the distribution of measurements as a direct indication for the uncertainty in the *mean value* and thus anticipate that it should decrease as the number N of events in a sample is increased. It is, therefore, of utmost didactical interest to convince the class, at the very beginning of the experimental activities, that the width of the data distribution reflects the relevant characteristics of the whole measurement process in each experimental situation and, ideally, only fluctuates in the vicinity of a constant value. Therefore, making the students calculate standard deviations for samples of increasing N, and showing the near constancy of the results, helps them to clarify the concept, especially if the calculated values are displayed on the respective graphs. On each histogram of their previous experimental results they had already been prompted to indicate estimates for *a best value* for their measurements and for *a width* of the experimental distribution, meant as an indication of the dispersion of their values, so as to encompass the *majority* of the events. Afterwards, they had been instructed to calculate the arithmetic mean $\overline{x}$ and the standard deviation estimate

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N - 1}} \tag{3}$$

and to compare these two quantities with their *previously guessed values*. No *formalization* of a concept for the standard deviation was attempted on that occasion. The values of $\overline{x}$ and $\pm s$ were simply represented on the histogram by vertical and horizontal arrows, respectively, and it was stressed that the value $2s$ gives a result very similar to the value estimated for the *width*, while $\overline{x}$ is certainly not far from the previously guessed *best value*.

In this context, calculating s for the dispersion of the y successes around their mean value $\overline{y}$ and indicating it on the graphs, in a way similar to that already employed, reinforced the concept that the standard deviation encompasses a region around the mean where *most* of

the events lie. It was stressed that this *most* may be quantitatively different for each kind of probability function, in particular for dice throwing and their former measurement results. A typical standard deviation is indicated in figure 1(*b*).

All values of *s* obtained by the students are collected and also discussed during the aforementioned *synthesis class*. It is verified that even for samples of $N = 10$, the standard deviation should be represented with two significant figures if the first digit is a 1 or 2. It is also seen that *s* remains constant within its increasing significant figures as N is increased, but that fluctuations as large as 11% are *usual*, even for $N = 50$.

2.3. The standard deviation of the mean, s_m

As already stated, students intuitively expect the result of a measurement to be better determined as the sample size N increases. Unfortunately, most students associate this gain in precision for the final result (that is, the smaller uncertainty of the *mean value*) directly with a decrease of the standard deviation estimate for the dispersion, s. If repeated samples of size N are taken, their *means*, $\overline{y}_N$, tend to obey a limiting distribution (of Gaussian form, in the central limit theorem situation) with width

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \tag{4}$$

the small sample estimator being

$$s_m = s_m(N) = \frac{s}{\sqrt{N}}. \tag{5}$$

Since the standard deviation, s, is approximately constant, the *width of the distribution of the means* is expected to decrease by a factor $1/\sqrt{N}$ with increasing sample size. To assure students of this fact and to show them that s_m is to be taken as the *statistical uncertainty* of the mean, their sample means $\overline{y}_N$ were collected, histogrammed and presented to them in the *synthesis class*. Figure 2 compares the histograms which resulted for 100 means of samples with $N = 10$ and $N = 50$, and for 93 means (the totality of the student teams) for $N = 500$. The information contained in figure 2 must be thoroughly discussed, but it is usually possible to convince the freshmen that the widths of the *distributions of the means* decrease with N in the expected proportion. Therefore, there is a *statistical expectation* for each *individual mean* to lie nearer the final expected value, represented by the mean of the means $\overline{\overline{y}}_N$, as N increases and, hopefully, this message is also grasped by the students. The caption of figure 2 presents, for each N, the values of the experimental standard deviations $s_{\overline{y}}(N)$ calculated for the approximately 100 means shown in the figure, as an experimental measure of their fluctuation around $\overline{\overline{y}}_N$, in comparison with the expected value, $\sigma_m(N)$, obtained from equation (4). Observing these values, good accordance may be noted, if allowance is given to the statistical variation of $\sim 7\%$ expected for the standard deviation for samples of about 100 events [5]. In a future paper, the authors intend to show how the data collected in the throwing of dice may also be used to ascertain the statistical accordance of the fluctuations observed, both for the N_y frequencies and for the standard deviations, s, with their respective expectations.

Now the frame is set for returning to a quantitative appreciation of the information contained in table 1. It may be observed that the standard deviations of the means (note $p_{ex} = \overline{y}/10$), presented there, are of the correct magnitude, even if only eight pieces of data were considered, that is $s_p(N) \simeq s_{\overline{y}}(N)/10$. It is, by now, also possible to decide that the final result of the 1996 gambling activity should be accompanied by $s_m(N)$, calculated through equation (5), giving $p_{ex} = 0.332\,884 \pm 0.000\,694$. In the future paper, it will be made clear that three significant figures should be presented for the uncertainty of this final result. Finally, very good accordance with the expected value of p (equation (1)) can, therefore, be claimed.
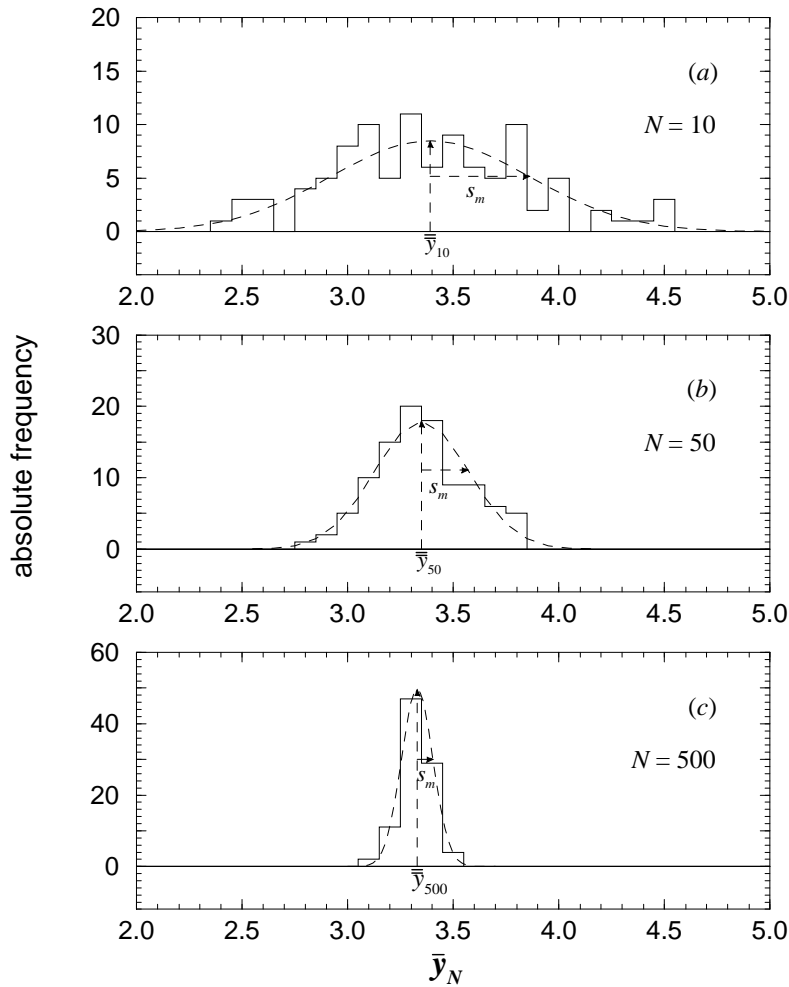
Figure 2. Absolute frequency histograms of approximately 100 experimental results for the means, $\bar{y}_N$, of the number of successes, y, in 10 simultaneous throwings of marked dice ($p = \frac{1}{3}$), for samples of sizes (*a*) $N = 10$, (*b*) 50 and (*c*) 500 compared with their Gaussian representation (dashed curves). The means of the means, $\bar{\bar{y}}_N$, and a measure of their dispersion, $s_\mathrm{m} = s_{\bar{y}}(N)$, are respectively indicated by vertical and horizontal dashed arrows (see text). For comparison, the theoretically expected values (equation (4)) are given in brackets: $s_{\bar{y}}(10) = 0.47\ (0.47)$, $s_{\bar{y}}(50) = 0.22\ (0.21)$ and $s_{\bar{y}}(500) = 0.075\ (0.067)$.

In retrospect, in the *synthesis class*, with help especially of figures 1 and 2, much emphasis has to be put on the fact that the standard deviation *s* represents a measure of the half-width of the original distribution and, as such, $\bar{x} \pm s$ is the interval where *one* new try (with 10 dice or one new measurement) should lie, with some associated *probability*. On the other hand, the standard deviation of the mean, s_m, represents the half-width of distribution of the *means* of several samples, each containing *N repetitions*. Thus $\bar{\bar{x}} \pm s_\mathrm{m}$ (where $\bar{\bar{x}}$ is the mean of the means), represents the interval where a new *mean* of *N* repetitions is to be found with about 68% probability, since this distribution is approximately Gaussian.

3. How to conclude that a systematic difference with expectations exists

Table 2 shows some examples (arbitrarily taken from table 1) of the experimentally verified probability of having an ace in a single dice throw, as obtained in 1996, for several sample sizes. The value of the experimental uncertainty, as given by $s_m(N)$ calculated through equation (5), is given in parenthesis with the due significant figures. Next, table 3 compares the final results of each year, starting in 1991 when this activity began, in fact with a slightly different character and also differently manufactured dice. What can be concluded about the repeated gambling activity and about the dice in use?

Table 2 shows that most of the partial results of 1996, as well as the final result, are in very good agreement with the theoretical expectation $p = 0.333\overline{3}$ within the 68% confidence interval ($1s_m$). Clear exceptions are the $N = 10$ result $p_{ex1} = 0.42 \pm 0.06$ and a $N = 500$ one, $p_{ex7} = 0.3150 \pm 0.0056$, the former being encompassed by the 95% confidence interval ($2s_m$), while the latter is just outside this interval, but well inside the $3s_m$ one, which is about what is statistically expected. Students are thereby alerted to be cautious with their rejection impetus, in particular if small samples are examined. Table 3, on the other hand, demonstrates accordance within $1s_m$ since 1992 with the expectation $p = 0.333\overline{3}$ for the same dice, still in use, leading to the conclusion that there has been fair play for some years. However, the value obtained in 1991 with the first set of differently manufactured dice is clearly well outside the $3s_m$ interval, being considerably lower than expected. It must be stressed to the students that one was able to reach this conclusion only after having gathered the $N = 500$ results of *many* teams and, in particular, the $N \sim 4000$ results of the 10 classes which then gambled ($p_{ex} \pm 0.002$: 0.321, 0.324, 0.326, 0.331, 0.327, 0.329, 0.331, 0.328, 0.330 and 0.333). Almost all are below the expected value of p by several s_m, indicating that something unusual happened in 1991.

Table 2. Some examples of experimentally verified values (1996) for the probability p_{ex} of having an ace showing up with dice having two marked faces, obtained with samples of increasing size N. The associated uncertainty given by s_m (equation (5)) is given in parenthesis.

N	p_{ex1}	p_{ex4}	p_{ex7}
10	0.42(6)	0.28(7)	0.36(4)
50	0.332(17)	0.318(20)	0.310(23)
500	0.329 4(61)	0.328 6(70)	0.315 0(56)
4 000	0.333 13(239)	0.334 38(236)	0.331 83(234)
46 500		0.332 884(694)	

Table 3. Final experimental results for the probability of having an ace showing up, p_{ex}, and for the width of the associated distribution measured through the standard deviation, $s_p = s/10$, obtained from the global samples of indicated size (N), with the same dice being thrown in lots of 10 in the successive years. In 1991 differently manufactured dice were in use. Uncertainties of the final values p_{ex}, as given by s_m, are presented in parenthesis, truncated to one figure for ease of comparison.

Year	N	p_{ex}	s_p
1991		0.3280(7)	
1992	43 042	0.3328(7)	0.1492
1993	44 324	0.3334(7)	0.1488
1994	34 446	0.3332(8)	0.1488
1995	35 400	0.3329(8)	0.1492
1996	46 500	0.3329(7)	0.1487

4. Discussion and conclusion

It has been shown in this work that the activity of throwing dice contains a wealth of statistical information of didactical interest. The concepts of (i) limiting distribution; (ii) standard deviation and (iii) standard deviation of the mean and its change by a factor $1/\sqrt{N}$ as the sample size N is increased, were all demonstrated with the aid of the practical results obtained by the students themselves. Gambling is a very popular activity in Brazil and probably in most other countries. So, if correctly exploited, it is possible to maintain the interest of the class in the activity, even during the boring 500 repetitions. For instance, this can be done by prompting them to see if the somewhat expected *one event* with 10 aces will show up in *their class*. Moreover, during the activity the students are prompted to explore their partial results, obtained for $N = 10$ and $N = 50$, by drawing comparative graphs and getting a feeling for the variability of statistical events. The discussion of the results for p_{ex}, the experimental probability of having an ace, as a function of increasing sample size N, should be done at a growing level of sophistication based on statistical grounds and can be used to *calibrate* students' expectations to reality. In fact, it is this *calibration* of what to expect qualitatively for the small (usually $N \leqslant 10$, as is well known) samples of actual measurements, that is one of the most relevant results of the proposed activity. To succeed in his or her task of promoting a statistical understanding of the data taking process, the teacher must be always alert on this issue, coming back to the discussion, whenever pertinent, using whatever 'peg' presents itself in future activities.

The final goal of the activity is, of course, to show the freshmen how they can draw statistically based *quantitative* conclusions. It was stressed that no valid conclusion can be drawn unless a confidence interval can be stated for the result, at a certain confidence level. In physics it is usual to take a 68% confidence level and to presume that this corresponds to a '$\pm 1\sigma$' interval, which is, strictly speaking, only valid for purely Gaussian distributions. Students were convinced that they should go through the trouble of calculating standard deviations s (if samples of size $N \geqslant 4$ are available) and that the correct statistical uncertainty of the mean values (usually taken as final results) is the standard deviation of the mean, $s_{\mathrm{m}} = s/\sqrt{N}$, which should be substituted for the vague *one sigma* concept.

Last but not least, students were reminded to be careful with respect to their tendency to reject unexpected data on *first impression* and not to forget that any statistical affirmation and conclusion is made on a certain *confidence level*. In this sense the fortunate fact of actually having an unexpected outcome in the throws of 1991 helped in demonstrating this to the students. Indeed, no single team could have convincingly argued the 1991 dice to be untrue. It needed the statistically increased information, provided by the joint examination of the about hundred $N = 500$ results, to make the affirmation within a reasonable confidence level. The results of 1991 were, *a posteriori*, interpreted as due to the way of marking the dice, with paint sprayed over two entire contiguous faces, which then adhered slightly preferentially with the rubber surface over which the dice were thrown, therefore showing an unmarked face 'up' with a somewhat increased probability.

Acknowledgments

References

[1] Journeaux R and Séré M G 1994 Traitement statistique des incertitudes en physique: problèmes scientifiques et didactiques *Eur. J. Phys.* **15** 286–92

[2] Bates B T, Zhang S, Dufek J S and Chen F C 1996 The effects of sample size and variability on the correlation coefficient *Med. Sci. Sports Exerc.* **28** 386–91

Price P N 1997 Predictions and maps of county mean indoor radon concentrations in the Mid-Atlantic States *Health Phys.* **72** 893–906

Ruth C and Joseph P M 1997 Estimation of a photon energy spectrum for a computed tomography scanner *Med. Phys.* **24** 695–702

Brown I F, Turcq B and Alechandre A 1995 Teaching concepts of accuracy, precision, opportunity cost in environmental sciences: arms, legs and significant figures *Ciência e Cultura* **47** 41–4

Raičević J J, Merkle M, Ehrhardt J and Ninković M M 1997 Loss of lifetime due to radiation exposure-averaging problems *Health Phys.* **72** 550–7

[3] Horodynski-Matsushigue L B, Pascholati P R, Vuolo J H, Yoneama M-L, Dias J F, Siqueira P T D and Amaku M 1997 Abstract published in *Programas e Resumos* of the *Simpósio Nacional de Ensino de Física – Novos Horizontes (Belo Horizonte, Brazil)* p 100 (the full contribution (in Portuguese) will appear in the proceedings of the symposium, to be published)

[4] Redish E F 1994 Implications of cognitive studies for teaching physics *Am. J. Phys.* **62** 796–803

McDermott L C 1993 Guest comment: how we teach and how students learn—a mismatch? *Am. J. Phys.* **61** 295–8

French A P 1988 Some thoughts on introductory physics courses *Am. J. Phys.* **56** 110–3

[5] Helene O A M and Vanin V R 1991 *Tratamento Estatístisco de Dados em Física Experimental* 2nd edn (São Paulo: Editora Edgard Blücher) (in Portuguese)