

# Experience in teaching biological statistics

Grigorii Kozlov<sup>1</sup> and Mikhail Pushkarev<sup>1,\*</sup>

<sup>1</sup>Saint-Petersburg State Institute of Technology, 26 Moskovsky Prospect, 190013 St. Petersburg, Russia

**Abstract.** The work summarizes the many years of experience in teaching the discipline “Biological statistics” to students - biotechnologists of the St. Petersburg State Institute of Technology (technical university). The paper provides data on the features of teaching the course and its relationship with other disciplines of the curriculum, topics of coursework performed by students, publication activity of students, integration of educational research work of students with topics of the university and industrial partners of the university, as well as attracting students to the development of the discipline.

## 1 Introduction

Statistics, as applied to a particular specialty, is mathematical knowledge that is most frequently used by specialists in the natural sciences and humanities. A key problem in teaching this subject is the fact that many students who successfully pass the exams are subsequently unable to fully apply it to practice.

Teaching biological statistics is based on the course “Additional chapters of mathematics”, which deals with the theoretical foundations of statistical methods.

The basis for presenting material in the course is:

- emphasis on the application of methods to practice;
- division of competences into those necessary for everyone (formed within the framework of the basic course) and those necessary for a limited number of students (taught in the form of an elective or consultation).

Based on the analysis of the needs of the disciplines following the curriculum, the basic structure of the course (the competencies “to be able” required by all students without exception) is as follows:

- make the primary grouping of data;
- record the measurement results with the error that accompanies them;
- resolve the issue of the difference in values;
- assess the influence of the factor on the result of the experiment;
- assess the degree of relationship between two variables;
- build a calibration graph;
- identify typical gross errors in statistical data processing and experiment planning;
- identify the main signs of falsification of scientific data and/or imitation of statistical processing.

---

\* Corresponding author: [malepush@bk.ru](mailto:malepush@bk.ru)

Competencies that are needed for a limited number of students (as a rule, for the performance of diploma theses):

- optimization (nutrient media or conditions of the technological process of cultivation) using a full factorial experiment and fractional factorial designs;
- “elements of style” in biological statistics.

## **2 Features of carrying out various types of classes**

### **2.1 Lectures**

Reading lectures implies managing the mood of the audience, and one of the most important points is periodic relaxation, which allows relieving tension when assimilating complex material. However, “lyrical digressions” should be strictly within the framework of the subject and not only relieve tension, but also focus on the material, only in another form.

This requirement is best met by an amusing or thought-provoking fact from the history of statistics, mathematics or metrology (i.e. disciplines that are strictly related to the topic), or a biological fact that makes one think, which illustrates the material just analyzed.

Since the availability of electronic teaching materials provides, among other things, for their independent development, such “relaxation points” must be placed as slides in the presentation.

The variety of additional material is very important. Little-known portraits of great scientists in student and postgraduate age evoke positive emotions and guaranteed interest of the student audience. The use of well-known but vivid examples also enlivens the audience without distracting it from the topic of the lecture, relaxation occurs due to the change of occupation, but the subject remains the same – first we analyze the statistical series from the formal, and then from the emotional side.

Thus, information about the brain size of cetaceans and elephants (larger than that of humans), although generally known, nevertheless causes a guaranteed revival in the audience, a brief discussion about the encephalization coefficient [1] of various biological species not only “rehabilitate” *Homo sapiens*, but also a vivid example of the importance of taking into account biological specificity in data processing.

The need for validity of conclusions and the inadmissibility of the mechanical application of statistical methods are well illustrated by mathematical sophisms.

Illustration of statistical patterns with materials from dissertations of politicians and celebrities. A large number of well-known politicians in Russia and in the world hold academic degrees.

The audience is interested in the works of those authors who received an academic degree while being already famous. The use of the works of authors who came to politics from science or education after defending their dissertations is inappropriate, since they do not cause almost any revival in the audience. Emotionally, the dissertation, for example, of the vice-rector for scientific work, who became the deputy minister of education, is a priori recognized by the students as completely legitimate and does not cause any emotions.

Facts that cause sad reflections or are not convenient for everyone – this way of focusing attention comes into some conflict with the use of inclusive language. However, it serves an exclusively positive cause of maintaining the tone of thought. One of these points is illustrating the concept of statistical probability – “according to Swedish statistics for the period from 1866 to 1905 (40 years old), the proportion of newborn boys ranged from 0.512 to 0.518 (0.515 on average)” [2]. This leads to the sad conclusion that the planned

sacrifice to natural selection is  $\approx 5.8\%$  of the male part of the population from each generation (see formula (1)):

$$\frac{0,515 - 0,485}{0,515} * 100\% = 5,8252427 \dots \approx 5,8\% \quad (1)$$

Information from the history of science related to the subject also allows switching to the humanitarian perception of the topic of the lecture without being distracted from its essence. The perfect example is Pascal's triangle for calculating binomial coefficients, which has a very ancient history. Experiments on checking statistical laws (binomial, law of large numbers, etc.) carried out in different eras. Illustration of statistical patterns with topical statistical material (demographics, mortality when taking selfies).

Examples of egregious mistakes that cause laughter from personal "collections" of peer-reviewed works and some sites. Such techniques, especially with explanations of the biological meaning of the phenomenon, contribute to the consolidation of the material not only on the formal, but also on the emotional level.

An important aspect is the credibility of the lecturer. In addition to the quality of teaching materials and preparation for classes, the teacher must show his personal contribution and/or the contribution of his students to the development of the scientific field in which he teaches. In the course of lectures, it is absolutely necessary to give 2-3 slides with links to your work in the subject area (albeit very insignificant).

## 2.2 Workshops

The first stage of studying the subject is a detailed analysis of the data processing algorithm. Therefore, tasks are solved manually, and only then, after mastering the basic methods, the use of software products begins. The tasks require a lot of routine calculation work, and here it is also necessary to arouse interest not only in the best part of the audience (as a rule, 2-4 people out of 25-30), but also among all students. The task should arouse keen interest of the entire audience. Since a person himself is most interesting for a person, the initial data for working out methods for solving problems are the impersonal biometric data of students in a group (height, weight, gender, eye color and shoe size). Based on these data, problems are solved on almost all topics of the course. The second aspect is the periodic solution of problems that describe interesting facts and dependencies – for example, the problem of the Prussian cavalymen.

A classic problem on Poisson's law is "the distribution of the number of deaths from a hoof kick in 10 Prussian cavalry corps over 20 years (1875-1894)":

**Table 2.** The distribution of the number of deaths from a hoof kick in 10 Prussian cavalry corps over 20 years (1875-1894).

m(number of cases in one cavalry corps per year) = xi	f <sub>i</sub>
0	109
1	65
2	22
3	3
4	1

The third point, the most interesting are the tasks not from classical textbooks, but the tasks developed on the basis of course research papers of students of past years. This approach shows that their own work is of great practical value and demand, which is one of the most powerful motivators.

Analysis of typical mistakes in using the criteria that we go through, including the example of critical reviews of dissertations and articles.

### 2.3 Course work

According to the degree of motivation, students can be conditionally divided into two groups: motivated and those who just need to pass the subject. Among the first group, original ideas for course works were widely used. Themes of works that have an original idea, which were subsequently published (for example [4]) and/or took part in competitions, can be grouped as follows (33.65% of the total):

1. Works based on impersonal medical, veterinary, sports data. Performed in collaboration with medical, veterinary professionals and trainers.

Medicine and veterinary medicine:

- Dynamics of decrease in visual acuity of school students.
- Dependence of the number of teeth on the sex and age of a person.
- Dependence of blood sugar level on body mass index.
- Factors causing overweight in cats.
- Relationship of various diseases with overweight.

Physical culture and sports:

- The influence of various factors on the strength indicators of athletes.
- The influence of physical activity on the indicators of vital body functions.

2. Works based on social media surveys and questionnaires related to passive experiment and observation. Completed independently.

Anthropometry for sales:

- Distribution of shoe size.
- Distribution of finger sizes.
- Study of the target audience for the sale of earrings.

Beauty:

- Distribution of the fused earlobe.
- Distribution of hair length by body type.
- Distribution of dimples on the cheeks and chin.
- Comparison of the preferred relative hair length of a woman's hairstyle by residents of Russia and Mexico.
- Dependence of eye color on solar radiation.

Study:

- Dependence of academic performance on the presence of wisdom teeth.
- Correlation of left-handedness in children and parents.
- Factors affecting student performance.
- The influence of the geomagnetic environment on the well-being of students.

Biometrics for forensics:

- Influence of body mass index on foot width.

3. An active experiment or observation was conducted.

Ecology:

- Study of the mass distribution of acorns from various parks of St. Petersburg (2 works in the center and suburbs).

Study:

- The effect of chewing gum on students' cognitive abilities.

Influence of various factors on the reading speed of people aged 18-25.

Medicine:

- Studying the nature of the distribution of the pulse rate of young men and women at rest.
- Influence of energy drinks on the somatometric indicators of different age groups.
- Effects of green tea and coffee intake on heart rate.

Physical culture and sports, entertainment:

- Endurance when playing “peepers” depending on eye color.
- The influence of physical activity on the heart rate in boys and girls 17-22 years old.

Biometrics for forensics:

- Correlation of stride length and anthropometric data.

12.5% of the work was carried out on typical topics: studying the nature of the distribution of a particular biological trait (geometric size, weight, size of shoes and gloves, hairstyle parameters), determining the correlation coefficient of variables, etc., but an original object of observation was selected that requires efforts in data collection (mainly anthropometry of the foot, hand, heart rate, nevi). These works are not sufficient for publication, but their material can be used both to develop a fund of assessment tools (typical tasks) and to illustrate the patterns studied during the lecture with original statistical material, for example, the height of an ideal man, how much an average student weighs, how many moles a person can have.

The remaining 53.85% of the work was carried out on a typical topic – studying the nature of the distribution of one or another biological trait (geometric size, weight) using a typical object that does not cause interest and does not require efforts during measurements (various cereals, acorns, needles, nuts, etc.). These works are insufficient for publication and are used to develop a fund of assessment tools (typical tasks).

To increase the motivation of students to the level necessary for mastering the competencies to the fullest extent, their work must be integrated into a large research project (for example, modernization of the collection of tasks for the course for the next year of study, ecological mapping of a district or city, etc.), which is part of the subject of the department. However, in this case, about 5% of students perform work with an original idea. As a rule, these are high-level athletes (members of the Russian national team, the work is carried out together with a coach or a sports doctor and is devoted to the training process), and those students who started scientific work on the first year (they process their own scientific data, or anonymized medical data, the work is carried out jointly with the scientific advisor).

## 2.4 Attestation

Attestation of students takes place in the form of testing. This format of assessing knowledge, in addition to the well-known advantages and disadvantages, has a property that is very important for the taught subject – it is quantitative data that can be processed by statistical methods. In the course of the study, the test results of previous years are analyzed, including the analysis of the coincidence of erroneous answers. This procedure arouses a keen interest among students and is of great pedagogical importance, since it demonstrates the uselessness of attempts to cheat – after the attestation, statistical processing and analysis of test results are carried out jointly with the students.

## 3 Teaching in a pandemic, remote access of students to classes

Teaching in a pandemic, remote access of students to classes. The widespread use of remote access technologies and distance learning during the pandemic has opened up wide opportunities for improving the quality of the educational process - after the restrictive

measures in the training process were canceled, an elective position “Zoom-commissioner” was introduced in the group - a student from a group carrying out a centralized broadcast of a lecture or workshop for ill students, students absent for other valid reasons, as well as students on internships and double-degree students. His responsibilities include broadcasting to the lecturer questions asked by remote listeners (as a rule, the number of remote students does not exceed 5 people). He also makes a video recording of the lecture for those who cannot participate in the Zoom-conference (for example, double-degree students with overlaps in the schedule of our university and partner university). Video recording of lectures followed by individual viewing has a serious drawback, which is why it is not used as basic electronic educational and methodological materials (list of topics, tasks for independent work, checklist, lecture presentations, text of lectures, a selection of original articles, standards and reviews on the subject of term papers and electronic versions of textbooks) - in most cases, pre-sounding of the video is required. In addition to the low (non-studio) quality of the sound recording, the unprocessed version of the lecture also contains, in addition to the presentation of the material, the lecturer's questions and remarks necessary to control the audience's understanding of the material, concentration of the audience's attention, etc. These remarks, questions and comments are pertinent and useful only in an interactive form of communication, when the reaction of the audience is clear, and when watching without the opportunity to ask the lecturer a question or answer it, they hinder the assimilation of the material. Thus, when preparing video lectures, one should not neglect the generally accepted methods from other areas of knowledge, namely cinematography, which clearly require a separate sound recording for a movie, including an educational one.

One of the negative aspects of the self-isolation regime is the impossibility of conducting laboratory and research work by students, and only their demonstration by the teacher is possible. For the teaching of biological statistics, this circumstance caused difficulties only partially, since a significant part of the coursework was carried out on the basis of the results of questionnaires on social networks. However, for work related to experiment or observation, self-isolation was a serious obstacle. The way out of this situation in the conditions of departure of a significant part of the students to their places of residence in the regions of Russia and foreign countries was the implementation of coursework on bioindication using local plant indicators. In fact, the work was carried out in a forced geographic observation network. At present, this successful experience has been extended to remotely studying foreign students and students studying for double degree.

## **Contribution of the authors**

Kozlov G.V. - materials (experience of teaching bachelor students, advising the students of the last year and graduate students), discussion, writing and editing the paper

Pushkarev M.A. - materials (experience in advising the students of the last year and graduate students), discussion, writing and editing the paper.

The work was performed within the State task of the Ministry of Science and Higher Education of the Russian Federation (785.00. X6019).

## **References**

1. J. Harry, *Evolution of the brain and intelligence* (Academic Press, 1973)
2. G.F. Lakin, *Biometrics. Textbook for biol. spec. of universities*, 4th ed., rev. and add. (M.: Higher school, 1990)

3. P.V. Terentyev, N.S. Rostova, Workshop on biometrics. Tutorial. (L.: Publishing house of Leningrad University, 1977)
4. G. Kozlov, M. Pushkarev, A. Kozlov, E. Perepelitsa, Bioindication for the Search of Microorganisms-Destructors. In: Murgul V., Pukhkal V. (eds) *International Scientific Conference Energy Management of Municipal Facilities and Sustainable Energy Technologies EMMFT 2019. EMMFT 2019, Advances in Intelligent Systems and Computing*, Springer, Cham. **1259**, 678-684 (2021) [https://doi.org/10.1007/978-3-030-57453-6\\_64](https://doi.org/10.1007/978-3-030-57453-6_64)