

INTERACTIVE_WIKI

Juha PURANEN,
University of Helsinki,
Finland

Contact Email: jussip2@luukku.com

ABSTRACT

Statistical Wikipedia is an excellent source for anyone who is searching information on statistical terms or methods. However it is more like a book than educational material. For a student who wants to learn statistics it can be hard to understand how a statistical method works, when to use it and how to interpret the results. If Wikipedia is made for the student it can be hard to read for an statistician who already knows a lot. Is it possible to increase readability of the pages and add guidance to the pages making them more informative for students?

INTRODUCTION

At the University of Helsinki we had a small statistical encyclopedia with interactive educational material. We noticed, that it was used, not only by students, but also by scientists who needed information on the statistical methods they used in research. Because my retirement was near, we wanted to make it public and hoped that other teachers would add their material and examples to it .

The first step was to find a wiki-software with similar features as in our web-pages. After several experiments our choice was DokuWiki. DokuWiki is a flexible Wiki-software where users can develop new extensions for their needs. We used PHP and with some programming it was possible to get many of the features I had used on my web-pages.

METHODS

Methods which we used in our material to help students to learn statistics were:

Hierarchical text structure

I have found many web-pages difficult to read. The pages are too long, only a small part of the page is visible at one time. For to repeat something can be hard because it is not possible to mark important points in the text.

When a student reads a textbook first time, he/she underlines those lines he/she thinks to be important. The next time she/he reads only underlined lines. This is not possible with the net without printing first the page. To increase the readability of the pages we tried to do something similar. Our text is hierarchical. On the first page of every article we put only the essential information (the underlined lines). All the examples, derivation of the formulas, etc. we put on the sub-pages. The other method we used to help the students read the text was to hide part of the page temporally. It was possible to make this part visible or hide again easily.

Online help:

In the class we have the teacher, who helps the student.

- How are text and elements in the picture connected?
- How to interpret a term in the formula?
- How to derive the formula?
- How are the calculations made?
- How to interpret the results?

On the net is difficult to get help.

When constructing instructive web-pages we must understand what kind of problems the student can have and find a method to help the student. So far we have used very simple methods like

- Pop-up windows. To explain the substance of a single term, pop-up windows can be used.
- Links between text and picture: When the student clicks a word in the text, the

corresponding element in the picture is highlighted.

- Help to the picture can be added by the means of an image-map and when the student moves the mouse over a problematic point in the picture or clicks it, a new picture with help text is changed.

Example

Here is the partial derivative of the log-likelihood function for the normal distribution with respect to the variance.

$$\begin{aligned} & \frac{\partial}{\partial \sigma^2} l(X_1, \dots, X_n | \mu, \sigma^2) \\ &= \frac{\partial}{\partial \sigma^2} \left\{ -n \cdot \ln(\sqrt{2 \cdot \pi}) - \frac{n}{2} \cdot \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_1^n (X_i - \mu)^2 \right\} \\ &= \frac{\partial}{\partial \sigma^2} \left\{ -n \cdot \ln(\sqrt{2 \cdot \pi}) \right\} - \frac{n}{2} \cdot \frac{\partial}{\partial \sigma^2} \left\{ \ln(\sigma^2) \right\} - \frac{\partial}{\partial \sigma^2} \left\{ \frac{1}{2\sigma^2} \right\} \cdot \sum_1^n (X_i - \mu)^2 \end{aligned}$$

Tässä on huomattava, että derivointi tapahtuu varianssin σ^2 :n suhteen eikä keskihajonnan σ suhteen.

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} l(X_1, \dots, X_n | \mu, \sigma^2) &= 0 - \frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{1}{2} \cdot \frac{(-1)}{\sigma^4} \cdot \sum_1^n (X_i - \mu)^2 \\ &= -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2 \cdot \sigma^4} \sum_1^n (X_i - \mu)^2 \end{aligned}$$

The last line in the upper formula and the first line in lower formula are image-maps. If one has problems understanding how the partial derivative is made, it is possible to see how the terms are connected by moving the mouse over the lines with image-maps (picture 2).

$$\begin{aligned} & \frac{\partial}{\partial \sigma^2} l(X_1, \dots, X_n | \mu, \sigma^2) \\ &= \frac{\partial}{\partial \sigma^2} \left\{ -n \cdot \ln(\sqrt{2 \cdot \pi}) - \frac{n}{2} \cdot \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_1^n (X_i - \mu)^2 \right\} \\ &= \frac{\partial}{\partial \sigma^2} \left\{ -n \cdot \ln(\sqrt{2 \cdot \pi}) \right\} - \frac{n}{2} \cdot \frac{\partial}{\partial \sigma^2} \left\{ \ln(\sigma^2) \right\} - \frac{\partial}{\partial \sigma^2} \left\{ \frac{1}{2\sigma^2} \right\} \cdot \sum_1^n (X_i - \mu)^2 \end{aligned}$$

Tässä on huomattava, että derivointi tapahtuu varianssin σ^2 :n suhteen eikä keskihajonnan σ suhteen.

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} l(X_1, \dots, X_n | \mu, \sigma^2) &= 0 - \frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{1}{2} \cdot \frac{(-1)}{\sigma^4} \cdot \sum_1^n (X_i - \mu)^2 \\ &= -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2 \cdot \sigma^4} \sum_1^n (X_i - \mu)^2 \end{aligned}$$

Because we are just changing pictures (in this case 4 pictures), it is possible to use part of the pictures for additional help (picture 3).

Because $\frac{\partial 1}{\partial x} = \frac{-1}{x^2} \Rightarrow \frac{\partial 1}{\partial \sigma^2} = \frac{-1}{\sigma^4}$

$$= \frac{\partial}{\partial \sigma^2} \left\{ -n \cdot \ln(\sqrt{2 \cdot \pi}) \right\} - \frac{n}{2} \cdot \frac{\partial}{\partial \sigma^2} \left\{ \ln(\sigma^2) \right\} - \frac{\partial}{\partial \sigma^2} \left\{ \frac{1}{2\sigma^2} \right\} \cdot \sum_1^n (X_i - \mu)^2$$

Tässä on huomattava, että derivointi tapahtuu varsinkin σ^2 :n suhteen eikä keskihajonnan σ suhteen.

$$\frac{\partial}{\partial \sigma^2} l(X_1, \dots, X_n | \mu, \sigma^2) = 0 - \frac{n}{2} \cdot \frac{1}{\sigma^2} - \frac{1}{2} \cdot \frac{(-1)}{\sigma^4} \cdot \sum_1^n (X_i - \mu)^2$$

$$= -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{\sum_1^n (X_i - \mu)^2}{2 \cdot \sigma^4}$$

VISUALIZATIONS AND SIMULATIONS:

JavaScript

With JavaScript can easily visualize dynamic phenomena like:

- How do parameters change the form of the distribution?
- How does the system work?
- What happens in the formula?

I usually do it with a series of pictures and JavaScript which changes the pictures automatically or manually. To change pictures manually a parameter scale and image-map can be used.

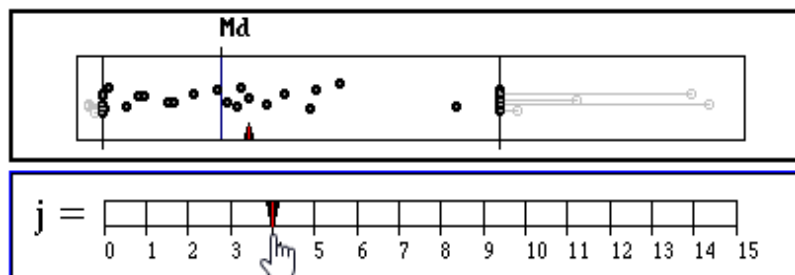
Example

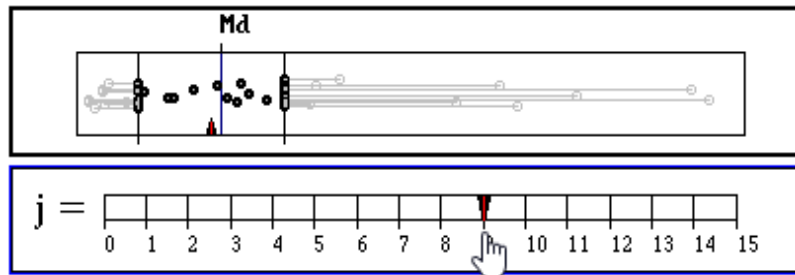
$$\bar{x}_{Wj} = \frac{1}{n} \left(j \times x_{(j+1)} + \sum_{i=j+1}^{n-j} x_{(i)} + j \times x_{(n-j)} \right)$$

Winsorized mean is a robust location parameter defined:

It is possible to visualize the content of the formula with following picture:

The picture has two parts. In the upper part is the data, Median of the data and a red arrowhead describing the mean (winsorized mean.). The lower part (j-scale) is a picture-map, and when the student moves the mouse over the it, he/she changes the winsorization parameter j. The upper picture shows the corresponding winsorized data and winsorized mean (in picture 5 j=4 and in picture 6 j=9).





Flash

With Flash applets it is possible to visualize almost everything. Only a good idea and some programming is needed. I have made several Flash-applets for classroom use and found them useful in many cases. Because it is easier to maintain one Flash-applet than many similar applets, I have put together several applets which use the same components. For example my histogram applet can:

- Generate samples from normal distributed population.
- Use real data.
- Plot histogram from the data
- Change the classification (class-limit bin-width)
- Test, if the parent distribution is normal using the χ^2 -test .
- Plot the theoretical and observed frequency histogram used in the χ^2 -test .
- Plot different types of kernel density estimators.

It is difficult to use this applet on the web because it does too many activities at the same time. The students do not know what they should learn from it. On the web one applet can visualize only one statistical feature at a time. I have solved this by a parameter-file which is read at the start and which defines those functions the applet will use in that example. The Flash-applets worked well in DocuWiki until the project ended and PHP was updated to a new better, safer version, which was not backward compatible.

ALTERNATIVE VERSIONS

When I start a new course in statistics I usually say to my students: *“When you are studying statistics, it is good to have at least two books. If there is something you don't understand in one book, the way it is explained in the other book can be easier for you.”*

It is the same on the net. Students are not the same. The way how they learn varies. It is good if we can have different approaches to the statistical methods and a lot of examples from different fields of science.

One advantage with web-based material is the possibility to add as much material as you like (and can). There are no limits as in a book. The readability of the text depends on the way how the material is organized .

SUMMARY

In this experimental project there were only 4 persons on our team, but we could show that with quite simple methods it is possible to add interactivity into Wikipedia and improve it toward a more educational tool.

We tried to get other teachers involved with our "Statistical Wiki"-project to get their articles and examples in it. It did not work. There are several reasons:

- Teachers have their own teaching and research, which takes a lot of time.
- We did not have any money to pay to the teachers.
- It is not difficult to add material to our "Statistical Wiki". However many teachers, who are not familiar with the web, believe it is impossible and don't dare to take the first step.

To make our "Statistical Wiki" better, we need a bigger team. At every university in the project we need at least one active member who searches good material, makes an agreement with the teachers for using it and adds it to the statistical Wiki. I do believe that this kind of solution on a laptop or tablet-PC can be very useful when one is studying statistics.

REFERENCE

Juha Puranen (2005) Statistical Simulations in the Web
<http://www.stat.auckland.ac.nz/~iase/publications/13/Puranen.pdf>
<https://www.dokuwiki.org/>
www.mathstat.helsinki.fi/tilastosanasto (in Finnish)