A tiny little bit of history linked to statistics and mathematics

What has the 'Student t-test' to do with beer and why the 'Scottish Book' is an amazing story about Polish mathematicians?

In an internet blog of the journal 'Plos' in 2016, a distributor lamented that students are bored by statistics and reluctant to appreciate lectures about mathematics as very interesting topics. The author in 'Plos' suggested to change this negative attitude, in resorting to some interesting features in the history of statistical methods and mathematic theorems. The examples given were quite exciting and the readers of this 'Journal Club' might enjoy some of those as well.

Student t-test

The 'Student t-test' goes back to a publication of William Sealy Gosset (1876 – 1937) in Biometrika with the title 'The Probable Error of the Mean'. Gosset was born at Canterbury in England as a descendent of a Huguenot family. Those familiar with European history will remember that the Huguenots were migrating from France to countries near and far, because King Louis XIV (le Roil Soleil – the Sun King - living in Versailles palace under more than lousy hygienic conditions) evoked the Edict of Nantes from his grandfather Henry IV. King Henry granted the Calvinist Protestants 1598 certain rights in the otherwise predominantly catholic country. To abolish these rights triggered the Huguenots to leave France migrating, among other countries, into England. The Huguenots proved to be of great benefit for the countries they moved in because they were highly elaborate and very skilled workers. Gosset's descent might explain his intelligence.

Gosset studied mathematics and natural science and graduated from Oxford 1899. He was employed from the Guinness brewery in England and worked with this company throughout his life time. His keen interest in the process of brewing beer, from the quality of barley up to fermentation of yeast, finally made him to be the chief brewer of the London plant of the company. The Guinness Company supported his further study in statistics at the University College at London, where Pearson (remember the Pearson correlation coefficient r!) was teaching. At that time, it was possible to use statistical methods based on large sample sizes. Gosset however only had small samples to evaluate and realized that there is no reliable method for statistics with a small sample size. He discussed a solution to that problem with Pearson, but it seems that the latter did not realize the significance of it. So Gosset decided to publish his ideas how to statistically treat small sample sizes. His employer was not very fond to allow him to publish his work. It was feared that competitors of the company might get an insight in to the secrets of production details of the Guinness s brewery. Therefore, Gossett published the paper under the pseudonym 'Student'. Nowadays Gosset is considered as a very important contributor to the development of statistical methods. He published numerous articles all of those, with one exception, under his pseudonym. Even after his death it was not generally known who was behind the now named 'Student-t-test' familiar to any student in the field of science linked to statistical methods. Up to now the t-test of Student is one of the most used procedures in the statistical toolbox. But what is the t-test? Suppose you want to compare the mean outcomes in two groups, say the mean BMI. Then it is natural to take the difference between the two means. But what is its distribution? This will depend on the

scale of the measurements taken. However, if we devide by the standard deviation of the mean difference then a certain distribution arises, the *t-distribution of Student.* This is always the case, whatever outcome is considered. The only underlying assumption is the normality of the outcome data. Part of the mathematical beauty of the underlying result was the surprising fact that sample mean and sample variance are independent which makes the the t-distribution so unique. Given the complexity of the proof it can be fairly assumed that Gosset had plenty of free time in his job brewing Guinness. Although the application areas are quite different, Albert Einstein seemed to have a similar freedom in developing the foundations of the relativity theory while working at the patent office in Bern (Sitzerland). Not only a comedian might wonder which lessons can be drawn from this for the productivity of science?

Scottish book

To raise the spirit of students of mathematics a number of appealing anecdotes are available such as the story around the 'Scottish book'. Among those experts of mathematics, the 'Scottish book' is famous as a collection of mathematical problems being solved from those fascinated by the challenge. It is safe to suspect that only a very tiny minority of the general public will find enjoyment in solving a mathematical problem while sitting in an old fashion coffee shop, similar to the ones around the Habsburg Monarchy in the 19th Century. In fact, such a minority existed in the Polish city Lwów. A group of mathematicians regularly met in a restaurant called Scottish Café, close to the university in the years before the Second World War. Someone of this group bought a folder, which was kept at a secret spot by the head waiter. Those who wanted to suggest how a particular mathematical problem could be solved did note that down in the folder. He could expect to get a price, ranging from a bottle of brandy up to a living goose. It is said that even some entries from Russian mathematicians found its way into the book.

Why even some Russian entries found its way into the book, one again must have some historical knowledge about recent European history. In this case about the Second World War. The war was started by Hitler ordering the German Army to invade Poland. The Polish army strongly resisted the takeover of the City of Lwów by the German troops. At the end the Germans didn't conquer the city but the Polish army had to give in to the Russian Red Army appearing on the scene more or less at the same time as the Germans. Finally, Hitler ordered the German troops to withdraw in favor of the Russians. All this is described in the Battle of Lwów. At the end the former Polish city Lwów turned in to what nowadays is Lviv in the Ukraine. The Ukraine formerly was in cooperated into the Soviet Union. How this could have happened is another twist in history. Some of the territory the Russians conquered in Poland remained finally within the Soviet Union, due to the insistence of Stalin at the Potsdam Conference at the End of the Second World War in Europe.

To compensate the loss for Poland the former German province of Silesia became Polish. That might explain, why a number of the Polish group of mathematicians, who contributed to the Scottish Book and made it through the war became famous mathematicians finally staying in Wroclaw, which was the former capital city of Silesia known as Breslau. Well, also this story starting with a nice anecdote related to mathematics enables a short detour into history which might

raise the all-round education of some interested students looking beyond statistics and mathematics.

The Scottish Book in its original form contained 193 mathematical problems to which Stanislaw Mazur alone contributed 43 problems. Many problems (and their solutions which were often prized, for example, with a living goose) lay the basis of modern functional analysis which is the basis of the existing stochastic science including the current art of data science machinery. The last problem was contributed by Hugo Steinhaus in 1941 which raised the question of the likely count distribution of the counts of matches in a match box. This problem has more practical impact than it might seem at first glance. It has close connection to rejection sampling in quality control. Suppose a match box should have a target number of matches, but there is random variation due to production. At which count should the box be rejected by the inspector? If the distribution of the count of matches is known, this number could be determined. This brings me to an anecdote that the famous statistician C.R. Rao told at dinner in his home. C.R. was (and supposingly is) a very friendly person with great hospitality and often invited visitors and friends over for dinner at his home. During one of these dinner parties the conversation came to the statistical methods used in quality control. He then tells the following. An American company orders a number of items from a Japanese company and, under specifications, allows them to have at maximum 5 products out of 1000 to be faulty. After some time the package is delivered in America and when the package is opened there are 1000 items, 995 wrapped in white paper, 5 in brown paper. The customer wonders about the 5 brown items and the Japanese company explains that these are the 5 faulty items that had been ordered. The alert reader will have noticed that C.R. was probably hinting to the Japanese Taguchi methods which substantially improved quality control. In any case, it was said it was a very nice evening.

Prof. Dankmar Böhning from the Southampton Statistical Science Research Institute, University Southampthon, UK significantly contributed and revised this entry into the Journal Club. The original manuscript first appeared in the last of the three editions of the ‘Khon Kaen Public Health Forum’ in 2016. The English language of the Forum was edited by Peter Bradshow. Frank P. Schelp is responsible for the content of the manuscript and points of views expressed might not reflect the stance and policy of the Faculty of Public Health, Khon Kaen University, Thailand