BIULETYN INSTYTUTU HODOWLI I AKLIMATYZACJI ROŚLIN Nr 288 / 2020 : 23–40
E-ISSN: 2657–8913

Artykuł przeglądowy
Review paper

Historia rozwoju statystycznych metod planowania i analizy doświadczeń rolniczych na świecie oraz w Polsce

A history of the development of statistical methods for designing and analyzing agricultural experiments in the world and in Poland

Wiesław Mądry✉, Dariusz Gozdowski

Katedra Biometrii, Wydział Rolnictwa i Biologii, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
✉ e-mail: wieslaw_madry@sggw.pl

W pracy przedstawiono główne kierunki, chronologię oraz osiągnięcia w zakresie metod statystyki matematycznej w zastosowaniu do biometrii i doświadczalnictwa rolniczego, dokonywane od XVII wieku do czasów współczesnych. Uwzględniono dorobek uczonych na świecie oraz w Polsce. Podkreślono historyczne i współczesne znaczenie tych osiągnięć matematycznych i metodycznych dla rozwoju i postępu nauk empirycznych w ogóle, a zwłaszcza nauk rolniczych i biologicznych. Przedstawiono znaczenie zastosowania metod statystycznych w uznaniu empirycznych badań rolniczych, jako nauki rolnicze. Świadectwa i rozważania oraz autorskie osądy odkryć i wynalazków statystycznych na przestrzeni wieków i lat są udokumentowane i zilustrowane oryginalnymi publikacjami, a także realnymi dokonaniami pionierów statystyki, biometrii i doświadczalnictwa rolniczego oraz współczesnych uczonych w tych dziedzinach i specjalnościach.

Słowa kluczowe: metody statystyczne, metody wielowymiarowe, modele statystyczne, estymacja parametrów statystycznych, hipotezy, testowanie hipotez, układy doświadczalne, wnioskowanie statystyczne statystycy, biometrycy

The paper presents the main trends, chronology and achievements in the field of mathematical statistical methods applied to biometrics and agricultural experimentation, made from the 17th century to modern times. The achievements of scientists in the world and in Poland were taken into account. The historical and present importance of these mathematical and methodological findings for the development and progress of empirical sciences in general, and especially agricultural and biological sciences, is emphasized. The importance of using statistical methods in the recognition of empirical agricultural research as agricultural sciences is presented. Testimonies and considerations on statistical discoveries and inventions over the centuries are documented and illustrated by original publications and real activities of pioneers of statistics, biometrics and agricultural experimentation, as well as contemporary scientists in these fields.

Key words: statistical methods, multivariate methods, statistical models, estimation, hypothesis testing, experimental designs, statistical inference, statisticians, biometricians

Motto

Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house [*Henri Poincare* (*1854–1912*), *Science and Hypothesis*, *London 1905*]. The more difficult it is to acknowledge their existence, the greater the care with which we must study these phenomena [*Pierre Simon de Laplace* (*1749–1827*), *Essai philosophique sur les probabilites, Paris 1814*].

Introduction

Since ancient times, knowledge about phenomena in agriculture has been sourced from (1) observations of events in nature, intuitive linking of relationships and coexistence of events and inference about correlations, causality and forecasting (e.g. observations of weather phenomena and their effect on plants), and (2) agricultural practice of farmers, who have been experimenting with natural and agronomic factors for millennia. Farmers, by living in close contact with nature and agricultural production, improved their observation and cognitive skills. As an eternal experimenter and curator of farmlands, the farmer has made entirely new and useful discoveries. These discoveries are relevant to this day and fascinate modern people with their depth and realism, as well as the beauty of folk mastery. In Europe, the progress made by acquiring agricultural knowledge in this way increased the efficiency of crop and animal production between the 17th and 19th centuries. However,

Managing editor: H. J. Czembor.

real progress in agricultural science in Poland and other parts of Europe began after the theory about the role of minerals in plant nutrition was proven by the chemist Justus von Liebig (1803-1873) in his book *Organic chemistry in its application to agriculture and physiology* (1840), and with general advances in science and industry in the 19th century. Knowledge about agriculture mainly came from empirical research that relied on inductive and deductive reasoning. The empirical methods of research are consistent with the modern philosophy of science that has been developing since the 17th century. These methods include different principles of intersubjective establishing and confirming knowledge based on the laws of logic and probability theory. One of these principles is pathway of reproducibility and replicability in science which means that a scientific investigation is repeated in the same or similar circumstances, and new observations or experiments are made in order to verify concepts reached by inductive reasoning, which makes broad generalizations from specific observations in the first stage of investigation. Repeating of research to scientifically validate conclusions is a key requirement for the ongoing process of self-correction in the scientific method. The philosopher of science Karl Popper (1902-1994) said that "a single unreplicated empirical study is insufficient for science". A luminary of mathematical statistics, biometrics and statistical methods in experimentation, Ronald Aylmer Fisher (1890-1962) added that "a scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance" (Fisher, 1935). R. Fisher refers here both to a statistically correct and realistically replicable experiment, and a high-power statistical test (a test that ensures a low probability of a type II error) at a low level of significance (a test that ensures a low probability of type I error).

For almost a thousand years, empirical methods of research have been used increasingly often in physics, astronomy and chemistry. One example of the first methodological approaches to empirical investigation in the modern sense are studies on optics by the Arabian mathematician, physicist and astronomer Ibn al-Haythama (965–1040), a great philosopher of the Islamic Golden Age. The intellectual foundation of empirical scientific methods is a philosophical theory called empiricism. It states that knowledge comes only or primarily from sensory experience while all ideas, theories, etc. are products of this experience. Modern empiricism was created by Francis Bacon (1561–1626), a British philosopher and poet of the Renaissance and the Baroque. Two proponents of empiricism were the most brilliant and greatest polymaths in the history of the world: Leonardo da Vinci (1452–1519) and Galileo Galilei (1564–1642). Galileo (a peer of F. Bacon) was a mathematician and physicist and believed that an experiment alone was not enough for scientific knowledge. He criticized the original approach to empiricism and claimed that the mere accumulation of observations about events was not science. He argued that the role of science is to discover the repetitive coexistence of events, or the laws of nature, by means of inductive reasoning based on empiricism. Contemporary researchers, including those specialised in biology and agriculture, follow this methodology and consider empirical and theoretical studies as complementary in the process of inference about the mechanisms ruling the world.

Empirical methods in agricultural studies mainly involve comparative trials called factorial experiments. In a factorial experiment a combination of different treatments (factors) is used to stimulate specific responses in certain material in the same setting. The experiment is a scientific test of what will happen to the tested objects under specific conditions when different treatments are used. The philosophical basis of factorial experiments is the method of difference, coined in 1843 by the eminent British philosopher and follower of empiricism, John Stuart Mill (1806–1873). This method was explained in his book *A System* of *Logic*, *Ratiocinative* and *Inductive*. According to this method, if two different treatments of research material under the same circumstances result in different effects within the phenomenon under investigation, then the conclusion can be made that this effect is caused by the used factor (treatment). The concept of experimentation emerged about 150 years ago with progress in experimental studies in agriculture and applied biology, and the evolution in methodology associated with it. In general, experimentation is the process of scientific discovery of the nature of phenomena, testing hypotheses, or demonstrating selected phenomena, carried out by means of experimental methods.

Agricultural experimentation emerged and developed as the first model of experimentation in history. It is an extensive, often interdisciplinary branch of agricultural science dealing with research, development and implementation. Agricultural experimentation includes: (a) optimally designed controlled factorial experiments in a laboratory, pots and in fields, setting up on-farm experiments or observational and measurement studies, such as surveys and censuses, (b) the use of findings from these studies for inference about phenomena for research or implementation purposes. All disciplines of agricultural sciences and their subdisciplines have developed their own specific experimentation, with unique principles and methods.

Agricultural experimentation can also be used in another sense to describe an interdisciplinary, methodological and statistical speciality in two fields of science: agricultural sciences and mathematics, which is closely linked to experimental studies (design of experiments, experimental design, design and analysis of experiments). In this sense, agricultural experimentation deals with (a) the statistical methodology of designing agricultural controlled factorial experiments, on-farm experiments, and observational and measurement studies, and (b) statistical analysis of data, interpretation of results and inference. This scientific speciality is often called statistical methods of design and analysis of agricultural experiments, statistical methodology of agricultural experimentation, or the theory of agricultural experimentation (Caliński, 2012). The era of creating statistical methods for the design and analysis of agricultural experiments began at the end of the 19th century. It coincided with the emergence of a ground-breaking new discipline of statistics, i.e. mathematical statistics.

The statistical methodology of agricultural experimentation has become a separate specialised branch of biometrics, which includes the broad use of statistics in biological sciences and applied biology (primarily in the agricultural and medical sciences). This branch of biometrics is an integral element of the overall methodology of agricultural experimental studies (Oktaba, 2002; Caliński, 2012). It should be emphasized that only the application of mathematical statistics in experimentation gave the empirical experimental method scientific status.

Statistical methods for the design and analysis of agricultural experiments began to develop rapidly, following the quantitative and qualitative progress in experimental studies in agricultural science and practice, and the theory of mathematical statistics. Beyond any doubt, the revolutionary development of the theory of mathematical statistics as a branch of applied mathematics was stimulated between the turn of the 20th century and the 1950s, mainly by the growing need for statistical methodology, necessary to conduct more and more scientifically advanced experimental studies in agricultural sciences (agricultural chemistry, plant breeding and production) and in biological sciences. From the 1950s to modern days further progress in the statistical methodology of experimentation has been strongly linked to advances in agronomy and biology (physiology, genetics), medicine, ecology, environmental sciences and other life sciences. Progress in the theory of statistical methods was also stimulated by advances in agricultural experimentation and methodology, as well as qualitative and quantitative aspects of studies (Oktaba, 2002; Caliński, 2012).

Statistical methods for the design and analysis of agricultural experiments, in line with the theory of mathematical statistics, focus on two issues. The first issue is adequate experimental design. The second issue concerns the statistical analysis of data with optimal inference about the event in terms of random variables in a general population. Therefore, statistical methods enable the most complete inductive inference possible about regularities in the investigated events based on data obtained from an experiment. Conclusions obtained in this process are reproducible in further identical or similar experiments, and this is assumed with a high probability. Therefore, conclusions from well-designed and interpreted experiments broaden scientific knowledge in accordance with the philosophy of science and the use of empirical research.

Statistical inference from experimental data is carried out on the basis of an appropriate statistical model of these data. This model formally describes the cause-effect relationships and/or interrelationships between different indicators (traits, characteristics, variables) of the investigated phenomenon. Statistical inference involves a) the most precise evaluation (estimation) of the parameters of the statistical model that characterize the regularities of the investigated phenomenon, and b) testing hypotheses about these parameters.

In further chapters we present and contemplate the achievements of prominent Polish and foreign scientists, whose talent, vision and investigative passion have shaped over the centuries the theory of mathematical statistics, biometrics and experimentation. These achievements show the power of mathematics and its role in bringing us closer to understanding the material principles about reality, necessary for the sustainable use of global resources. R. Fisher once said that "the rise of biometry in this 20th century, like that of geometry in the 3rd century before Christ, seems to mark out one of the great ages or critical periods in the advance of the human understanding". Importantly, the great mathematicians specialising in statistics and biometrics had interests in nature (physics, chemistry, biology, agriculture) and mathematics, and received a thorough education in these fields. The aim of this review is to present their achievements. Some parts of this paper were included in the Memorial Book prepared for the 100th anniversary of Agricultural Experimentation at the Warsaw University of Life Sciences.

Developments in statistics and its applications in experimentation before 1920

What we call statistics today has a rather complicated history (Oktaba, 2002; Ostasiewicz, 2012). In this chapter we present part of this history, from the modern era when statistics emerged as the art of collecting and using data from population and economic surveys, as well as probability theory, until the end of the 19th century and the first two decades of the 20th century (including the early work of R. Fisher in Rothamsted in 1919), when the theoretical foundations of mathematical statistics and its first applications in biology were created, giving a background for biometrics and experimental methods.

Although the well-documented history of statistics is not very long, its roots go back to ancient states, mainly Egypt, China, Babylon, Greece, Persia, India and Rome, where methods that are now referred to as statistical were used for censuses and administrative surveys of goods. Currently, this type of activity, i.e. economic statistics, in Poland is mainly carried out by Statistics Poland (Główny Urząd Statystyczny, GUS), which has been operating for over 100 years (since 1918). This institution, which significantly contributes to the economy and statistics, organizes, coordinates and executes national agricultural censuses and other data collection projects. The operation of Statistics Poland has been highly inspiring for the development of statistical methods, especially sampling techniques (Neyman, 1934; Kozak, 2004 a, b), but also the estimation and forecasting the outcomes of economic processes.

Modern statistics emerged with the development of modern philosophy, and the great contributor to this process was the French scientist René Descartes (1596-1650). Francis Bacon, as one of the fathers of empiricism, also had a great influence on the development of statistics. In his time, the development of statistics began to be inspired and stimulated by the recognition of empiricism as a methodological doctrine of scientific investigation.

An important event for the rise of statistics as a scientific discipline was the 1662 publication of a book by John Graunt (1620–1674), *Natural and Political Observations Made Upon the Bills of Mortality*, in which he compiled data from the bills of mortality for London. Graunt was an outstanding precursor of a trend in economic and demographic studies called political arithmetic. These studies were based on a detailed quantitative analysis of data from censuses and public records. Graunt was the first to discover that a careful analysis of numerous observational data reveals the regularities governing events in a population. The term *Statistik*, derived from the Latin word *status* and meaning 'state', first appeared in German in 1749. It was used in the work by Gottfried Achenwall (1719–1772) in the sense of knowledge about the state. For a long time, until 1850, statistics, as the art of data analysis, developed in Western Europe, mainly for the purposes of state science. Until almost the end of the 19th century, state science was an important driver of progress in statistics. From at least the middle of the 19th century statistics started to draw the interest of naturalists, biologists, farmers and doctors of medicine.

The first records evidencing interest in probability theory and statistics in Poland are *Dyskursy* by the eminent scientist Jan Śniadecki (1756-1830), as well as his manuscript of 1790 *Rachunek zdarzeń i przypadków losu* (Ostasiewicz, 2012). Progress in statistics in Poland and its application in the economy and administration of Poland's territory under the partitions in the times of the Duchy of Warsaw and the early times of the Kingdom of Poland was strongly promoted and supported by Stanisław Staszic (1755-1826), who received a comprehensive education in France (he studied in *Collège* de France and the Institute of Life Sciences in Paris), and was a prominent scientist and naturalist, as well as a public, scientific, economic, educational and social activist. In the early 19th century, Staszic published *Statystyka Polski*, where he tried to show Napoleon the historical and economic picture of the state where the Duchy of Warsaw was being established. Staszic was one of the most important advocates supporting the establishment of the Agronomic Institute in Marymont near Warsaw in 1816, whose tradition is continued now by the Warsaw University of Life Sciences.

Statistics was officially recognized as a science in the early 19th century when the *British Association* for the *Advancement of Science* created a statistical section, and the Royal Statistical Society was established in 1834. In the second half of the 19th century, awareness about the role of statistics in Western European countries was high. International statistical conferences were organized and works by European mathematicians on probability calculus were known. Monographs were also published. Ostasiewicz (2012) reports that the first books on statistics in Poland were published in 1868–1870 by Witold Załęski (1836–1908) and Zdzisław Korzybski (1834–1896). Załęski regarded the statistical method as one of the approaches to scientific investigation. He believed that in the real world events are driven by deterministic and random causes. Therefore, the main role of statistics is to eliminate random effects in the analysis of empirical data and to discover the statistical law, which reflects the law of the real world. These were very enlightened views on the role of statistics in the empirical investigation of events. Today, statistics plays the same role in empirical research in all disciplines of science. By about 1900 statistics had already become a stand-alone and mature scientific discipline, and was then on the verge of a breakthrough era of the emergence and development of the theory of mathematical statistics, which transformed old statistics into a section of applied mathematics, ensuring the scientific character of empirical research. Mathematical statistics is founded on probability theory, usually created by empirical inspiration to describe and analyse the regularities ruling events in the general population. The events are random or deterministic-random by nature, i.e. they are not fully controlled by humans. Probability theory was created in the second half of the 17th century by two French mathematicians, Blaise Pascal (1623–1662) and Pierre Fermat (1601–1665), who provided mathematical explanations for regularities observed in games of chance. Further progress in probability theory was achieved with efforts from other mathematicians: Jacob Bernoulli (1655–1705), who was the first to justify and mathematically define the law of large numbers, Abraham Moivre (1667–1754), who introduced the concept of probability distribution and generalized Jacob Bernoulli's law of large numbers (in a work from 1733), Daniel Bernoulli (1700–1782), who significantly developed the theory of probability and errors, Thomas Bayes (1702–1761), and Pierre Simon Laplace (1749–1827). Bayes was a mathematician most famous for formulating Bayes' theorem, which was published posthumously. This theorem concerns the conditional probability of two random events. Based on this theorem, the theory of Bayesian inference (Bayesian statistics) was developed. Beyes' theory of probability has multiple applications today, including in statistical methods based on linear and nonlinear models for classified experimental data (da Silva et al., 2019). Laplace created the classic definition of the probability of a random event and the paradigm of classical determinism. These theories were presented in two monographs: *Théorie analytique des probabilités* (1812), and *Essai philosophique sur les probabilités* (1814). The aforementioned paradigm recognizes the existence of objective laws of nature that completely determine real events, so man, having knowledge about all the laws governing events and being able to analyse them, could predict and partially control the course and outcomes of events. Pierre Simon Laplace and Carl Friedrich Gauss (1777–1855) proposed a formula describing the probability density function of a normally distributed random variable, modelled on the theory of errors. Gauss was the first to separate systematic and random components in the mathematical description of events in the population, before the regression concepts were proposed by statistical mathematicians between the end of the 19th century and 1920. He also made a significant contribution, together with Adrien-Marie Legendre (1752–1833), to the development of the least squares method, which was an important criterion in the statistical estimation of parameters of random variables in the general population. These two great mathematicians, Laplace and Gauss, made an early contribution to the development of mathematical statistics before its theories were elaborated.

Between the second half of the 17th century and almost the end of the 19th century probability theory and statistics attracted growing interest among outstanding mathematicians and physicists, scholars and administrators, mainly dealing with state science, political arithmetic and natural sciences. However, these two disciplines developed in considerable isolation: the first was strongly mathematical by nature, although it was inspired by empiricism, while the second was empirical and applicable. The two-centuries-long efforts of political arithmeticians, naturalists, biologists and mathematicians, who were developing and more eagerly using probability calculus for inference from data in 1875-1900, brought their research aims more closely and led to interdisciplinary cooperation. This resulted in the creation of a new section of statistics, mathematical statistics. The precursors of the theory of mathematical statistics and its first applications in biological and agricultural sciences were three British scientists: Francis Galton

(1822–1911), Karl Pearson (1857–1936) and William Sealy Gosset (1876–1937). However, the greatest contribution to the development of the theory of mathematical statistics and its applications was made by two British scientists: Ronald Fisher (1890–1962) and Egon Sharpe Pearson (1895–1980), and a Pole, Jerzy Spława-Neyman (1894–1981), who published his works after 1925 as Jerzy Neyman (Statystycy Polscy, 2012).

The first ideas, concepts and methods of mathematical statistics, like earlier statistics, emerged out of inspiration and practical needs in biological, natural and technical research, both theoretical and empirical. These categories concerned the general population of events (units) in a large population as well as a random and representative sample, the theory of point estimation of statistical parameters for a random variable in a population based on large and small samples, probability distributions and error frequencies and their variance, the statistical concept of correlation and simple correlation coefficient, and a simple linear regression. Galton, Pearson and Gosset made great contributions to define and mathematically explain many of these original, fundamental concepts and techniques of mathematical statistics and their applications in the theory of evolution, population genetics of categorical and quantitative traits, eugenics, medicine, psychology, as well as in agriculture and technology. Galton was a pioneer in the use of the theory of normal distribution to fit frequency histograms based on large samples, and worked on these problems in 1870 and 1880. At the turn of the 20th century, before the re-discovery of Mendelian laws around 1900, Galton, on the basis of contemporary biological knowledge, and later his successor Pearson, using the theory of probability and mathematical statistics, created the first school of biometrics and school of biometrics in genetics and evolution.

K. Pearson in 1893–1904 made significant progress in statistical techniques to be used in biometrics, especially the method of correlation and simple regression and multiple linear and nonlinear regression, as well as principal components analysis. The results of his pioneering work were published in 1901 (Pearson, 1901). Let us emphasize that the first two methods were originally developed in the 1880s in a graphic and computational version by Galton while he was solving various problems in biology: anthropology, quantitative genetics and eugenics. They were widely promoted by both scientists, so they quickly found applications in biological, medical, anthropological, psychological and other studies, immediately after their publication in 1901. K. Pearson was not only the father of the parameter estimation theory by developing a method of moments, but also the theory of statistical hypothesis testing, using the chi-squared distribution and the p value well -known in statistics. This approach by K. Pearson was proposed in 1900 and oriented the thinking of later creators of the theory of hypothesis testing, Fisher, Neyman and E. Pearson, but was not adopted by them in its original form (Johnson and Kotz, 1997). In 1901 K. Pearson co-founded *Biometrika*, a journal that is published to this day and is still regarded by statisticians and biometricians as highly prestigious. Until about 1900, estimation theory (without interval estimation), based on a large sample of data, was already well developed, with the greatest contribution from Galton and K. Pearson. However, Gosset regarded it as insufficient for his scientific and implementation works. As a qualified chemist, in 1899 he took the prestigious position of data scientist at the Guinness Brewery. At his place of employment, Gosset could only work on small data sets (small samples). Thus, he elaborated his own original theory of parameter estimation (for the population mean and a simple coefficient of correlation) based on a small sample using probability distributions for the estimation error. In this respect, Gosset is best known for his pioneering discovery, published in 1908 under the pseudonym Student, which was the development of the theory of the t distribution (Student 1908). The t distribution is any member of a family of continuous probability distributions that arise when estimating the mean of a normally-distributed population in situations where the sample size is small and the population's standard deviation is unknown. In 1925, the t distribution was defined by Fisher as Student's distribution, and later Student's t-distribution, and the basic statistical test was called Student's t-test (Fisher, 1925).

Student's t-distribution became the foundation for many later statistical theories and procedures, such as the confidence intervals for the mean of one and two populations coined by Neyman (Neyman, 1937), or regression analysis and multiple comparisons tests for means in experimentation (Miller, 1981). Gosset was a friend of K. Pearson (he worked as a junior researcher at Pearson's laboratory at University College, London in 1906/1907), and of Fisher and E. Pearson. Gosset also made pioneering contributions to the theory of experiment. Shortly after he was hired by the Guinness brewery in 1899, he designed and supervised field experiments with malting barley cultivars to improve beer production. Fisher had great appreciation for Gosset for his theory of small sample inference, extraordinary research intuition, his practical attitude, and his simple approach to developing mathematical statistics (Johnson and Kotz, 1997).

In the early years of the 20th century, the greatest Polish anthropologist and biometrician of his times, Jan Czekanowski (1882-1965), applied and promoted methods of correlation and regression among German anthropologists, gaining their full acceptance and recognition. Czekanowski, as a student at the University of Zurich (1902-1906), wrote an article about biometrics, and in 1913 published the first textbook on biometrics in Polish, *Zarys metod statystycznych w zastosowaniach do antropologii/Introduction to statistical methods applied in anthropology* This book was published only two years after the world's first textbook on mathematical statistics, *An introduction to the theory of statistics* by G. Yule. Czekanowski's textbook, presenting, for example, inference based on the correlation coefficient as well as simple and multiple regression, played an outstanding role in the popularization of biometrics among Polish scientists before the First World War and in the interwar period (Caliński, 2012; Statystycy Polscy, 2012). The statistician Jan Czekanowski and geographer Eugeniusz Romer (1871-1954) played a prominent role as experts in geography, demographics and anthropology, and were members of the Polish delegation to the Paris Peace Conference, held in 1919-1920 and formally ending the First World War. Czekanowski and Romer worked in Paris between the end of December 1918 until the end of October 1919 (Romer, 1989). Their enlightened scientific arguments concerning the rights of the First Republic of Poland to recover territories that belonged to Poland before partitions in 1772 had a significant effect on peace negotiations and helped convince the victorious allies about the shape of borders for independent Poland. Another Pole, Edmund Załęski (1863-1932), played a pioneering and huge role in the application of mathematical statistics in agricultural experimentation between the end of the 19th century and the 1920s. A chemist by education, with interests and practical achievements in agronomy, plant breeding, experimentation and statistics, he was an outstanding figure in Polish agricultural science, a precursor of the theory and applications of statistical methods in experimentation and plant breeding known in Poland and abroad. Załęski, from 1888, worked in plant breeding and seed production, mainly sugar beet and wheat. From 1893 he ran his own breeding company (E. Załęski i S-ka), and later worked on other breeding projects. During that time he improved his knowledge of mathematics and statistics, which he first acquired at the Technical University in Riga, and confronted it creatively and persistently with experimental practice to advance plant breeding and assessment of cultivars. He also improved methods of the design and analysis of experiments. Much before Gosset and Fisher, Załęski arrived at impressive methodological and statistical solutions, and reached very high precision in agricultural experiments (not yet known in Western Europe then), especially with respect to plant breeding. Methods proposed by Załęski were used in 1898 by Aleksander Janasz and Władysław Mayzel for their breeding experiments. In 1907 Załęski published, in 5 languages at the same time, his findings in *Instrukcya do urządzania doświadczeń porównawczych z różnemi odmianami buraków cukrowych/Manual for designing comparative experiments with different cultivars of sugar beet.* This book is considered the first systematic presentation of the methodology of agricultural experiments using probability theory and early statistical methods. By publishing his book in 1907 Załęski was at least a year ahead of the first German and English books on this subject.

In his book, E. Załęski also reported the results of his almost 20-years' work as an experimenter and presented the principles for using multiple replicates in factorial experiments (unknown at that time in the West), as well as the concept and the use of the standard method (Statystycy Polscy, 2012). In this method, a standard, i.e. the same experimental treatment (mainly a cultivar in a breeding trial), is sown every the same number of the tested treatments across all the experimental field to improve accuracy of observations of the investigated plants by effective avoiding effects of systematic differences in soil fertility. It consists in recreating, by means of linear interpolation, the hypothetical values of a given trait for tested objects in unreplicated trials that are located between systematically sown standards. Statistical analyses available back then were carried out for adjusted data, and after 1920 the analysis of variance was used. Edward Kostecki, a close co-worker of E. Załęski, acknowledged the book by his mentor published in 1907 and its role at that time. In 1933 Kostecki wrote for Gazeta Rolnicza (http:// dlibra.umcs.lublin.pl/dlibra/ plain-content?id = 16355): "the methodological principles set out in this work, more importantly, these principles immediately put into practice by Załęski and other plant breeders, were actually far ahead of the official regulations which were in force at that time in the West, but in many aspects the West has not reached the accuracy that can be achieved with the methods proposed by E. Załęski". In 1908, the International Congress on Agriculture in Vienna followed a proposal made by E. Załęski and accepted a resolution on the use of the least squares method in agricultural experimentation for the purpose of critical inference. Further impressive contributions made after 1920 by E. Załęski and his students in the theory and practice of experimentation are presented in chapter 3.

Alfred Hall (1864–1942), director of Rothamsted Experimental Station in 1902-1912, promoted the use of statistical methods in many agricultural experiments carried out at this famous agricultural research institution, and recognized the great value of data that had been gathered there since 1843. Hall worked to improve the methodology of agricultural experiments in the same years as Załęski in Poland, or Gosset in England and Ireland, whom the Polish scientist valued and collaborated with. In 1909 Hall published an article about the role of experimental error in field trials (Hall, 1909). He concluded that "the magnitude of experimental error attaching to one or more field plots is a question of extreme importance in Agricultural Science, because upon its proper recognition depends the degree of confidence which may be attached to the results obtained in field work". For this reason, Hall used some basic statistical measures (means and their standard deviations), to "show that a pair of plots similarly treated may be expected to yield considerably different results, even when the soil appears to be uniform and the conditions under which the experiment is conducted are carefully designed to reduce errors in weighting and measurement". As we can see, Hall, with this reasoning, was another contributor, next to Załęski and Gosset, to the new theory of testing statistical significance of differences between mean values for treatments developed in the 1920s and 1930s both by Fisher, and jointly by Neyman and E. Pearson. A year later, Hall and his agronomists made an attempt to determine the number of replicates in field trials at Rothamsted that would ensure statistically significant differences between means obtained for treatments. The results of their study were published in 1911 in the Journal of Agricultural Science. It was a real innovation and heralded changes at Rothamsted that were to come 10 years later. Another merit of Hall and his collaborators was the implementation and promotion in 1910 of uniformity trials in Western countries. In these experiments, also known as blank trials or dummy trials, no treatment or variety differences are involved, in order to measure how crop yields vary over space. These experiments are designed to assess the spatial differences in soil fertility across the field, based on the data on yields or other agronomic variables for the smallest plots established on this field. These data also allow for the determination of the size and shape of plots and the number of plots per block, which ensures the minimum possible assessment of the error of variance on the experimental field. Therefore, uniformity trials improve the precision of experiments and the power of statistical and scientific hypothesis testing. Importantly, the concept of uniformity trials in experiments on plant breeding and cultivars was also independently created and used by E. Załęski in the early 20th century. Uniformity trials are still valid and used in field experiments, although they create a great practical challenge for researchers. Analysis of data from uniformity trials carried out by Hall and his collaborators was an important inspiration for Fisher when creating randomised block designs (Speed, 1992).

It should be emphasized that the development of the statistical methodology of agricultural experimentation in the world between the late 19th century and 1920 was under the strong influence of E. Załęski in Poland and Gosset and Hall in Great Britain (Johnson and Kotz, 1997; Statystycy Polscy, 2012). They all based their work on the fundamental achievements of F. Galton and K. Pearson in statistics and biometrics.

In 1912 Hall left his job at Rothamsted Experimental Station. The use of statistical methods in this research institution was to be resumed not earlier than after the First World War. In the autumn of 1919 the station, headed in 1912-1943 by an agricultural chemist, Edward John Russell (1872-1965), hired the mathematician, promising statistician and geneticist Ronald A. Fisher. Before taking this job, in 1912-1918, Fisher published several important articles and books about the foundations of mathematical statistics and its applications in genetics. One of them was a pioneering work fundamental for quantitative genetics (Fisher, 1918). It presents the mathematical basis of quantitative genetics and a linear model: $P = G + E$, where the *phenotypic value* (*P*) of an individual in a population is the sum of genetic (G) and environmental (E) effects. This equation was the result of the achievements in genetics in the first 15 years of its development and of mathematical reasoning. It is a mathematical model describing genetic and environmental determinants of phenotypic variability in living organisms, and is the basis of modern quantitative genetics and its applications. Rothamsted Experimental Station expected Fisher to analyse large data sets from field experiments carried out since 1843. No one before him was able to elaborate these data, and Fisher did it using modern statistical methods. Later results of work carried out at Rothamsted by Fisher, his students and successors surpassed all expectations. They are a valuable contribution to the development of the theory of mathematical statistics and its applications in agricultural experimentation. These achievements are presented in Chapter 3.

Achievements in the methodology of experimentation at Rothamsted, in the world and in Poland in 1920-1945

Significant advances in the modern theory and applications of mathematical statistics, including estimation theory, started with Gosset's research on small statistical samples, in particular his paper on Student's *t*-distribution published in 1908. Two major statisticians, Fisher and Neyman, stimulated the rapid development of these disciplines after 1920. At the end of 1919 Fisher started work at Rothamsted Experimental Station, and Neyman after arriving from Russia to Poland in 1921, was hired by the Institute of Agriculture in Bydgoszcz, and later at the Warsaw University of Life Sciences (in 1923) and the Institute of Nęcki in Warsaw (in 1928). Rothamsted Experimental Station was the first research institution in the world, where the Department of Statistics, established by Fisher in 1920, was involved in regular cooperation with experimenting scientists. A justified assumption was made that progress in statistical methodology applied in experimentation can only be stimulated by the growing needs of experimenting researchers. Such a strategy of organizing research work at this renowned institution turned out to be productive both for the station and the world, giving all of us today the best model to follow. Thus, it is clear that agricultural experimentation was the most important driver of the historical development of modern statistical theory in the 1920s and 1930s (Statystycy Polscy, 2012).

Impressive contributions made by the greatest statisticians, Fisher and Neyman, as well as the nature and significance of their accomplishments were presented by Erich L. Lehmann (1917-2009) in his extensive monograph (Lehman, 2011). Lehman was the author of acclaimed works and monographs on the history of the theory of mathematical statistics and its applications, and worked at the University of California in Berkeley, US, where Neyman had also worked since 1938. Lehmann, a student and collaborator of Neyman, documents in detail and explains the accomplishments of Fisher and Neyman and their contribution to the development of the classical theory of statistical hypothesis testing, designing experiments, observational studies or quasi-experiments. Sometimes their contributions were complementary, and sometimes they led similar research in parallel, and especially in the later stages of their work it was often strongly contrasting. This whole story and the products of the efforts and intellectual attitudes of these people with great minds and powerful spirits is extremely valuable, and they created the foundations for modern statistics and empirical disciplines which, because of statistics, are science, and not just a kind of art.

In his pioneering paper from 1922, Fisher presented the mathematical foundations of statistical inference, focusing on his own and other achievements at the time. They concerned the philosophical foundations of mathematical statistics and the estimation of population parameters (Fisher, 1922). Fisher argued that "the object of statistical methods is the reduction of data", and "a quantity of data (…) is to be replaced by relatively few quantities which shall adequately represent the whole". To accomplish this object, Fisher recalls basic statistical terms already used (by Galton and K. Pearson) or introduces his own (today well known to all statisticians or those who apply statistics). These terms include hypothetical infinite population, a random sample representing this population, law of distribution of this hypothetical population, and population parameters. These parameters sufficiently describe the probability distribution for a random variable in the population. Fisher mentioned three types of problems associated with statistical inference: (1) specification of the mathematical form of the population, i.e. distribution for a random variable, (2) the choice of methods to estimate the values of the parameters of the population, and (3) specification of probability distributions for estimates. He also defined the properties of estimates, i.e. consistency and efficiency. Fisher introduced the validity method for estimating population parameters and demonstrated that the estimates of parameters of the normal distribution in the population are more accurate than those obtained using the method of moments proposed by K. Pearson.

A breakthrough moment for the development of the theory of mathematical statistics and methods of the design and analysis of experiments in the 1920s came with the creation of mathematical foundations (theory) for statistical hypothesis testing, which are hypotheses on the values of a random variable in a population. Methods based on these theories, applied in experimental studies, allow for a reliable discrimination between the actual effects of investigated causes (factors) and those resulting from a random sampling error or any uncontrolled variation. Two theories of statistical hypothesis testing, today considered classical, although based on different principles of probabilistic logic, were developed by Fisher, and by Neyman and E. Pearson. The first theory that Fisher started elaborating in 1921 (Fisher, 1921), includes tests of significance. The second theory, created a few years later by Neyman and E. Pearson between the end of the 1920s and early 1930s (Neyman, Pearson, 1928), includes tests of statistical hypotheses. These theories in their final form were proposed a few years later (Neyman, Pearson 1933; Fisher, 1935a).

Both theories of statistical hypothesis testing start with the same assumption that statistical inference is a procedure based on scientific induction, that is, inference in which we go "from the specific to the general or, in statistical language, from a sample to population" (Oktaba, 2002; Lehmann, 2011). In Fisher's approach, only one hypothesis is proposed, i.e. the null hypothesis, or H_0, which corresponds to the adopted research model. The test statistic T of known probability distribution is chosen if H_0 is true. A large absolute value T calculated for a sample, and thus the low probability p ($p<\alpha$, where α is the significance level, or probability of type I error) provides the investigator with evidence against H_0 and means that H_0 should be rejected. If $p>\alpha$, H_0 is not rejected and this creates a serious further research problem, since this conclusion does not mean that the H_0 hypothesis is almost certainly true. In the Neyman-Pearson approach, two hypotheses are formulated, i.e. the null hypothesis H_0 and an alternative hypothesis H_1. The procedure for testing the hypothesis is as follows: reject H_0 if $|T| \geq c$ and accept alternative H_1, or accept H_0, if $|T| < c$, where c is a predefined critical value of the test function at the predefined probability of type I error α and unknown probability of type II error β.

Currently, in applied statistics, the Neyman-Pearson theory of hypothesis testing with the significance level α is widely recognized as the standard procedure (Lehmann, 2011). However, Fisher's method of significance testing, where the *p*-value is only evidence against the null hypothesis (not the significance level α), has dominated testing practice. In the current approach to testing statistical hypotheses, both methods proposed by Fisher and Neyman-Pearson are used to create a specific, hybrid procedure suitable for its purpose (Greenland et al., 2016). It is worth noting that in each commercially available statistical software all tools designed for hypothesis testing support the calculation of the probability value *p*. Looking closely at these statistical procedures, we can see that their use by professional researchers does not have to create a threat to the validity of statistical and scientific inference. After all, researchers are not able to determine in their research the probability of type I error β and the power of the I-β test, although they can use these statistics as a source of inspiration to improve the theory and practice of experimental methods aimed at reducing the variance of estimation errors. Nowadays, statisticians and biometricians are involved in a lively debate regarding the role of Neyman's theory on confidence intervals, the Fisher and Neyman-Pearson theory on hypothesis testing, and the suitability of tools for statistical inference relying on these theories for science (Hurlbert and Lombardi, 2009; Greenland et al., 2016).

Fisher, with his creative and productive 14 years of work at Rothamsted Experimental Station in 1919-1933, dominated the history of the development of statistical methods for the design and analysis of agricultural experiments between the two World Wars (1918-1939). These were the great years of making ground-breaking discoveries and milestone achievements in agricultural experimentation. Fisher's methodical solutions of those times remain the basis of agricultural experimentation to this day (Speed, 1992). Fisher wrote two pioneering monographs on the statistical theory of experimentation, the first issued in 1925 (Fisher, 1925) and the second in 1935 (Fisher, 1935 b). Both monographs were later reprinted many times. Original terminology on statistical methods used in these monographs, and later perpetuated as classical, comes from the nature and methodology of agricultural experiments. When working at Rothamsted, in addition to developing the theory of hypothesis testing, Fisher made a great contribution to biometrics and experimentation, creating the foundations for the maximum likelihood estimation, a procedure for analysis of variance and the F test (from the name Fisher), the concepts of establishing blocks and their randomization in experimental designs, and a concept of new experimental designs: complete randomized design, randomized design, split-plot, Latin squares and factorial designs. Unlimited access to the valuable retrospective experimental data at Rothamsted helped a lot in assessing the suitability and effectiveness of these new statistical methods. To commemorate the centenary of Fisher's appointment at Rothamsted Experimental Station and the contribution of local statisticians to the development of modern statistical methods an international conference organized under the auspices of the Biometric Society was held in Rothamsted in July 2019. In 1933 Fisher ended his praiseworthy work at Rothamsted and moved to University College in London, where he focused again on statistical genetics and eugenics. Fisher's worthy successor at Rothamsted was Frank Yates (1902-1994), hired in 1931. Although they only worked together for two years until 1933, Fisher and Yates remained close associates and friends for another 29 years. In the 1930s Yates cooperated with Fisher on the concepts of new experimental designs, i.e. the Latin square, incomplete block design (very useful in biological and agricultural research), lattice squares and factorial designs. Yates also created the theory of confounding in factorial experiments and fractional replications, split-plot designs, balanced and partially balanced incomplete block designs, lattice squares and quasi-Latin squares. Without all these new experimental designs, experimentation in agriculture, life sciences and technical sciences would not be as advanced as it is now. Yates also made a serious contribution to the very broad application of its *computationally* challenging ideas, making a breakthrough in the computerization of the Rothamsted Experimental Station and setting standards for others. In 1936 Fisher and Yates published extensive s*tatistical tables, a valuable* source of information necessary for the users of statistics (Fisher, Yates, 1936).

At Rothamsted Yates cooperated closely with William Cochran (1909–1980), who worked there in 1934–1939. Within these six years Cochran and Yates developed pioneering methods for the analysis of data from long-term agricultural experiments, mainly regarding crop rotation (Cochran, 1939).

In the 1930s Harold Hotelling (1895-1973) initiated progress in multivariate statistics, strongly advanced later and today widely used in experimentation. Hotteling was a precursor of multivariate statistics and in 1931 he developed T-squared (T^2) distribution, which is a generalization of Student's *t*-distribution in a multivariate setting, related to the *F*-distribution (Hotelling, 1931*). Hotelling's T-Squared test is based on* T^2 *distribution and is used for testing hypotheses on the equality of differences in multivariate population means, assuming a normal distribution. T-squared distribution is related to the* Mahalanobis *distance* (Mahalanobis, 1930, 1936), developed by Prasant Chandr Mahalanobis (1893–1972). This measure is valued and widely used in biometrics, in addition to Euclidean distance, especially in discriminant analysis, canonical variate analysis and cluster analysis (Caliński et al., 1985). In 1933 Hotelling developed principal component analysis (PCA) (Hotelling, 1933), for which the mathematical basis was proposed in 1901 by K. Pearson. PCA is often used for exploratory data analysis in biometrics, mainly to identify common factors for many variables in the analysed population and to visualise the multidimensional similarity of objects defined by the Euclidean distance and approximated in a small number (2-3) of dimensions (Johnson and Kotz, 1997). In PCA, principal components are computed by eigendecomposition of the data covariance or correlation matrix, or singular value decomposition of the data matrix. Principal component analysis has become the basis of many classical and the latest methods for various specific applications. These include factor analysis/exploratory factor analysis (EFA), most suitable in agricultural experimentation, canonical correlation analysis (also proposed by Hotelling in 1935), k-means clustering, as well as methods based on additive-multiplicative statistical models for two-way data classification, such as the additive main effects and multiplicative interaction model (AMMI) and the genotype and genotype × environment interaction model (GGE). These two methods were developed at the end of the 20th century and are used in the analysis of data from a series of cultivar experiments for an in-depth assessment and visualisation of the genotype-environment interaction (Gauch et al., 2008).

In 1936, Fisher, after the pioneering achievements of Hotelling and Mahalanobis, published his next groundbreaking work on linear discriminant analysis (Fisher, 1936). Today, discriminant analysis is a term referring to a multivariate technique used to separate groups of observations based on variables measured on each experimental unit. Fisher's concept of discriminant analysis is used in many modern analytical methods, including data mining and gene microarray data analysis.

Polish forerunners of statistics and experimental studies (E. Załęski, J. Czekanowski, J. Neyman) also made great contributions to the development of statistical methodology for agricultural experimentation and world science in this field. Among them, the most admirable is Neyman for publishing works on the theory of experimentation in 1923-1934. He was the first before Fisher to create a probabilistic language adequate to describe randomized experiments. This allowed for more effective design of experiments and, more importantly, reliable inference. Neyman published his first findings on this subject in the Polish language (Spława-Neyman, 1923) when he worked at the Institute of Agriculture in Bydgoszcz. The idea of this work reached the broad community of statisticians much later, in 1990, when D. M. Dąbrowska and T. P. Speed translated Neyman's paper for the journal Statistical Science. The ground-breaking innovation presented in this paper was the idea of a completely randomized experimental design and the probabilistic model of data from such experiments, as well as the link between the design of randomised experiments and the probabilistic model that enabled effective statistical inference. The idea of this model is similar to that presented by Fisher in his monograph (1925) and known today as the linear analysis of variance (ANOVA model). One could therefore conclude that the two luminaries of the theory of statistics, biometrics and experimentation, Neyman and Fisher, independently created statistical models and designs of controlled factorial experiments for agriculture and other disciplines. This speculation or assumption can be made because Fisher did not read Neyman's paper in the Polish language, as well as the fact that the first work by Fisher with a preliminary version of the analysis of variance (ANOVA) for use in experimentation was published in 1923 (Fisher, Mackenzie, 1923) and the aforementioned Fisher's textbook in 1925. In his model Neyman for the first time used the term 'true yield', which is associated with the expected value of a random variable and the real effect of the factor under study. Over the years, Neyman's contribution to the methodology of agricultural experimentation was appreciated worldwide and he was recognized, next to Fisher, as a forerunner of not only the theory of mathematical statistics, but also statistical theory in experimentation. Neyman explained the dominant and initiating role of Fisher in the theory of experimentation, also referring to his personal input to creating the foundations of randomization in the design of experiments (Neyman, 1979).

Of note, Gosset is also considered, next to Fisher and Neyman, a forerunner of statistical methodology for agricultural experimentation. He ongoingly improved statistical methods for the analysis of data from these experiments, and from the 1920s he productively co-operated with Fisher, who called Gosset "the Faraday of statistics". The results of this co-operation were fruitful for both scientists. For example, a debate among statisticians on the importance of systematic and randomized experimental designs that continued in the 1930s, with a large contribution by Fisher, was initiated by Gosset in 1931. Gosset's methodological ideas on field experiments, also inspired by discussions with Fisher, were published in 1923-1937 (Speed, 1992; Johnson and Kotz, 1997).

E. Załęski, who began his pioneering research in agricultural experimentation at the end of the 19th century, presented his enormous and innovative achievements in the textbook *Metodyka doświadczeń rolniczych/Methodology of agricultural experiments* (1927), just about two years after the publication of two important works in 1925 in the United Kingdom and Germany, i.e. a monograph on *Statistical methods for research workers* by Fisher and *Der Feldversuch* by Theodor Roemer (1883–1951). In his original work, which made no reference to the seminal monograph by Fisher published in 1925, Załęski introduced the use of the first procedures for statistical data analysis and inference in agricultural experimentation. These procedures were developed in the late 19th and early 20th centuries by representatives of the British early school of biometry and experimentation, i.e. Galton, K. Pearson and Gosset. Załęski discussed the application of the theory of probability distribution for experimental errors and estimation parameters, and the methods of correlation and linear regression. He also presented logical and nature-related evidence for designing replicated experiments in pots and on field, and serial replicated experiments (called by him collective experiments) and long-term studies, as well as statistical methods for data analysis. With reference to the design of field experiments he emphasized the nature and role of fluctuating (or perfectly random) variability and systematic differences in experimental conditions. Załęski presented the theory and practice of standard methods that were used by him in field studies from at least the beginning of the 20th century, and which were also appreciated by other great Polish experimentalists recognized worldwide (e.g. Stefan Barbacki and Józef Przyborowski) and used for a long time until the 1970s (Statystycy Polscy, 2012).

The standard method, although simple in terms of design and data analysis, and in those days modern and very helpful for a reliable assessment of plant breeding material, was gradually replaced by alternative methods, first by the incomplete block design proposed by Yates and, as of the 1940s and further in the 1980s, by methods of data covariance analysis accounting for the nearest neighbourhood on the experimental field. Another valuable paper by E. Załęski, *Regjonalizacja*, *czyli dobór roślin uprawnych/Regionalization of agricultural crops*, published in 1929, is a fresh attempt to assess the results of serial and long-term replicated experiments on cultivars, which were later addressed in papers by Polish biometricians, including Jerzy Neyman, Stefan Barbacki, Regina Elandt, Tadeusz Caliński and others.

E. Załęski and his school at the Jagiellonian University also had great practical merits for agricultural experimentation in Poland and the recommendations on cultivars and fertilization of plants. The Department of Plant Breeding and Experimentation at the Faculty of Agriculture, Jagiellonian University in Krakow, headed by E. Załęski in 1919-1932, and in 1933-1939 by Załęski's student and a worthy successor Józef Przyborowski (1895-1939), cooperated with the Seed Section of the very active Małopolska Agricultural Association, but most importantly with the Ministry of Agriculture and Agricultural Reforms. The Committee for Cooperation in Experimentation at this ministry, led by Przyborowski, coordinated experimental studies on important crops. The experiments were established across Poland according to the latest standardized rules for statistical design of unreplicated and replicated trials and for data analysis, developed and published by Przyborowski (e.g. *Zasady organizacji i wykonywania doświadczeń odmianowych ze zbożami i ziemniakami/Rules for the organization and performance of experiments on cereal and potato cultivars,* Kraków, 1925). Recommendations on statistical methods for the design and analysis of experiments, as well as their results with the interpretation of national data were published regularly by Przyborowski and Wileński in the 1930s in short scientific and implementation monographs (e.g. *Metoda przeprowadzania doświadczeń z zastosowaniem poletek/Experimental method with standard plots,* Kraków, 1937; *Analiza zmienności wyników doświadczeń wielokrotnych/Analysis of data variability for replicated experiments*, Kraków, 1938) and in the Polish journal *Przegląd doświadczalnictwa rolniczego*. These publications are available, for example, in the Library of the Faculty of Economics and Agriculture, University of Agriculture in Kraków, which continues the tradition and research of the Faculty of Agriculture at the Jagiellonian University in Krakow. In this way Przyborowski and Wileński promoted with a strong commitment the conceptual and practical aspects of modern statistics in experimentation, mainly for the breeding and assessment of cultivars. The significant achievements of outstanding pre-war Polish scientists dealing with the theory and practice of experimentation, E. Załęski, Przyborowski, Wileński and Barbacki, created a brave and enlightened prototype for the post-war evaluation of cultivars and today's, very modern on a European scale, post-registration variety testing system, but also inspired pre-registration and registration experiments. Stefan Barbacki (1902-1979), a student of E. Załęski at the Jagiellonian University in Krakow, had great merits in the 1930s and after World War II for experimentation in Poland. Barbacki worked for 20 years (1926-1945) at the State Research Institute of Rural Husbandry in Puławy. In 1935 Barbacki published an excellent textbook, *Ogólna metodyka doświadczeń polowych w zarysie/Foundations of the general methodology of field experiments.* This textbook was published in the same year as Fisher's *The design of experiments* and Yates' *Complex experiments*. In his monograph, Barbacki introduced in Poland his own as well as Fisher's and Yates' methods for the design of unreplicated and replicated experiments, as well as statistical data analysis. Today, it can be said with full confidence that the textbook by E. Załęski in 1927 formed the basis of modern experimental methods, while the textbook by Barbacki in 1935 introduced the latest theoretical and methodological achievements of those years into agricultural experimentation in Poland. In 1939, Barbacki published another monograph *Analiza zmienności w zagadnieniach doświadczalnictwa rolniczego/Analysis of variation in agricultural experiments*, but all its printed copies were destroyed in September 1939. Here is a passage from proofing columns of this monograph that miraculously survived: "Statistical methods help us in experimental studies but they cannot create anything new. If a research problem in an experiment is wrongly formulated, statistical methods will not change it. They might provide an exact answer, but not to the problem we actually want to explain" (Caliński, 2012). Barbacki was the only Pole to co-author a publication with Fisher (Barbacki, Fisher, 1936).

Progress in the statistical methodology of experimentation after World War II

An important achievement in the theory of experimentation was the development of multiple comparisons for means using concurrent statistical procedures (significance tests and confidence intervals) used after the analysis of variance. These procedures were based on the work by D. Newman (Newman, 1939), in which he started from Gosset's ideas to formulate and illustrate the first multiple range test. General rules for multiple comparisons of means were defined in their present form in the years 1947-1955 by three key researchers, i.e. D.B. Duncan, Henry Scheffe (1907–1977) and John Tukey (1915–2000). There is still no full agreement between statisticians as to which procedures are the best. There are three categories of multiple comparison tests: (1) analysis of contrasts (Scheffe test), (2) procedures based on the studentized distribution range for the grouping of means (LSD, Newman-Keuls, Tukey and Duncan method), and (3) inference based on confidence intervals (Scheffe, Benferroni and Dunnet test), (Miller, 1981).

Another landmark event was the introduction by Scheffe in 1956 of a linear mixed model for Fisher's analysis of variance (Scheffe, 1959). This idea, though previously considered by Yates in the 1940s in the theory of incomplete block designs, in the version proposed by Scheffe became a milestone for the progress of the modern theory of design and analysis of simple and complex experiments with fixed factors and random factors, in observations repeated for the same units and in methods for complete and incomplete series of agricultural experiments.

The theory of incomplete block designs was developed creatively and effectively in Poland and abroad by Tadeusz Caliński (1928-) and his co-workers at the Poznań School of Biometry, and foreign co-workers. An intra-block analysis for block designs was proposed. Researchers specified different new block designs, orthogonal and non-orthogonal designs, resolvable block designs and affine μ-resolvable designs (widely used in cultivar trials), balanced incomplete block (BIB) designs and partly balanced incomplete block (PBIB) designs. There was also a rapid development of multivariate methods and their applications, such as multivariate analysis of variance (MANOVA) and canonical analysis. A key contribution to this was made by Calyampudi Radhakrishna Rao (1920-), a PhD student of Fisher in 1948, Caliński (a PhD student of Barbacki in 1961, and John Gower (1930-2019), a co-worker of Yates at Rothamsted.

We should also emphasize the great merits of many Polish statisticians dealing with experimentation, and farmers in the development of the methodology of agricultural experimentation in Poland and abroad after World War II. Great contributions to the scientific, educational and organizational aspects of progress in statistical methodology of agricultural experimentation during this period were made by Stefan Barbacki, Zygmunt Nawrocki, Regina Elandt, Wiktor Oktaba and Tadeusz Caliński (in chronological order of professional activity) (Statystycy Polscy, 2012).

In 1945, Stefan Barbacki moved from Puławy to the Faculty of Agriculture and Forestry at the University of Poznań, where he co-created and headed the Department of Experimental Agriculture and Biometry, renamed in 1951 the independent Higher School of Agriculture. In Poznań's scientific and agricultural community dealing with plant breeding and evaluation of varieties in the Wielkopolska region and across Poland Barbacki promoted and developed his pre-war achievements and concepts regarding the design of experiments and their statistical analysis, and also put into practice the ideas coined by Załęski. Barbacki's monograph *Doświadczenia kombinowane/Combined experiments,* published in 1951, is an unsurpassed example of how complex problems can be described in simple language. In 1961, Professor Barbacki co-founded the Department of Plant Genetics of the Polish Academy of Sciences, which in 1979 was transformed, with his great involvement, into the Institute of Plant Genetics of the Polish Academy of Sciences. He was also one of the initiators of the Research Centre for Cultivar Testing (COBORU), established in 1966, and later for many years was the chairman of its Scientific Council (Statystycy Polscy, 2012). Barbacki has enormous merits for Polish science, both in the pre-war and post-war times, as an outstanding scientist -experimenter dealing with statistical methods and practical aspects of research, as well as genetics and plant breeding. He was also a wonderful mentor and organizer of science. He made his impact forever because of his own accomplishments and those of his many excellent students. His students: Regina Elandt, Tadeusz Caliński, Eugeniusz Bilski and many others, are luminaries of Polish and world biometrics and agricultural experimentation.

Zygmunt Nawrocki (1910–1978) studied at the Jagiellonian University in Kraków under the supervision of J. Przyborowski. When working on practical aspects of plant breeding during and after WWII, in 1950 he was granted a doctoral degree in agricultural sciences at the Faculty of Agriculture of the University of Maria Curie -Skłodowska in Lublin, for his dissertation *O metodzie dyskryminacji populacji hodowlanych, opartej na pomiarze wielu cech osobników do nich należących/A method of discriminating breeding populations based on the measurement of many traits of their individuals*. In 1951 Nawrocki started work at the Warsaw University of Life Sciences. Nawrocki made a huge contribution to the development of post-war biometry in experimentation in Poland in the 1950s, 1960s and 1970s. After the war he revived and continued the school of statistics and biometry created before World War II by Jerzy Spława-Neyman at the Warsaw University of Life Sciences. Before 1951 Nawrocki was among the first researchers in Poland to use Yates' lattice squares in field experiments, in particular plant breeding experiments, for the assessment of a large number of varieties. In the 1960s he initiated the departure from strict models of experimental design in agriculture, especially with regards to plant breeding. Many of his concepts (e.g. the N design, an unreplicated trial for a large number of breeding treatments and replicated design for a standard cultivar) were introduced in the practice of agricultural experimentation in Poland. Nawrocki adapted and applied the theory of orthogonal projections coined by the meritorious Dutch biometrician Leo Corsten (1924-2013), to perform the analysis of variance for classified unbalanced data. It was a big step ahead in those days, when the statistical theory of mixed models and appropriate information technology were not yet advanced (Statystycy Polscy, 2012). From 1957 until his death Nawrocki was a member of the Scientific Council of the Plant Breeding and Acclimatization Institute in Radzików, and a member of the Scientific Council of the Potato Research Institute in Bonin. A doctoral student and co-worker of Zygmunt Nawrocki was Zbigniew Laudański (1942–2017). He made a significant contribution to the further development of Nawrocki's methods (in theoretical terms by using the theory of orthogonal projections and information technology) in plant breeding and agricultural sciences.

Regina Elandt (1918–2011) formed her scientific interests in Barbacki's school in Poznań, where in 1955 under his supervision she obtained a doctoral degree for a dissertation, *O pewnych testach interakcji w doświadczeniach wieloletnich i wielokrotnych. Zagadnienie rejonizacji./On certain interaction tests in long-term and replicated experiments. The problem of regionalization.* She presented her numerous scientific accomplishments regarding theoretical and practical aspects of mathematical statistics in agricultural experimentation, plant breeding and genetics in an excellent monograph (Elandt, 1964). It is a great work by Elandt, known and appreciated by her contemporaries, even today, for the timeless modernity and accessible presentation of achievements in biometrics and its applications. Other achievements of Elandt, who had great merits for experimentation during her work in Poland (1946-1964), concern the problems of biometrics in human genetics, epidemiology and survival analysis (Statystycy Polscy, 2012).

Widely known, meaningful and original scientific achievements by Wiktor Oktaba (1920-2009) include a diverse range of topics and address the following problems: regression, univariate and multivariate analysis of variance, variance components estimation, estimation and verification of hypotheses for univariate models of asymptotical constants, theory of experimental designs, multivariate Zyskind-Martin models, theory of models with missing observations, matrix algebra, and history of statistics. Professor Oktaba created the Lublin School of Statistics and Experimentation. He published many valuable academic textbooks (Statystycy Polscy, 2012). He also initiated the organization of annual conferences under the name Colloquium in Biometry, which have been held for many years up to the present time as the *International Biometrical Colloquium*. Another outstanding and well-deserved statistician specializing in experimentation at the Lublin School of Statistics was Tadeusz Przybysz (1929–2007). His most important scientific accomplishments include the development of the methodology of incomplete block design and crop rotation experiments. He introduced a generalized Yates' method, a method for the analysis and comparison of crop rotation, accounting for the test plant. Przybysz formulated several models for crop rotation experiments, such as hierarchical, a combination of cross-classification with hierarchical classification, and a model combining split-plot design with incomplete block design. For individual models, Przybysz provided parameter estimators and the appropriate form of analysis of variance for hypothesis testing.

Tadeusz Caliński is among the most outstanding, after Jerzy Neyman, Polish statisticians, biometricians and experimenters recognized in Poland and worldwide. He has made dominant contribution to statistics and biometrics, including agricultural experimentation on a national and international scale. By developing Poznań's school of Barbacki, Caliński created the great Poznań School of Biometry. The most laudable achievements by Caliński and his many outstanding co-workers cover a wide range of problems. The most important of them concern: the theory of incomplete block designs, multivariate methods, simultaneous comparison statistical procedures, mixed models for the assessment of varieties in multi-environment trials and the analysis of crop genotype-environment interactions.

A significant contribution to the progress in statistical methodology for agricultural experimentation in Poland was made by Leokadia Ubysz -Borucka (1919–1989), her student Jan Trętowski (1942–1993), and Ryszard Wójcik (1937–2003). They developed and considerably enriched (together with the previously mentioned Laudański) achievements by Nawrocki, contributing to the establishment of the third Polish research and educational centre for biometrics and agricultural experimentation at the Warsaw University of Life Sciences. Their major accomplishments include multidisciplinary improvement and applications of experimental designs, univariate and multivariate methods in agronomic and breeding experiments and in the assessment of many varieties of crops based on unbalanced data from multi-environment trials. This centre works actively, mainly in the areas outlined by its founders, as well as in new areas of agricultural sciences, primarily geostatistics, precision agriculture, interpretation and use of satellite data, and mathematical modelling of events.

A significant contribution to progress in biometrics in Poland and worldwide was made by the Wrocław School of Statistics, initiated by Hugo Steinhaus (1887-1972), and later developed and headed by Julian Perkal (1913–1965). This school mainly gathered researchers from the University of Wrocław and the University of Agriculture (today the University of Life Sciences) in Wrocław. An important research area of the Wrocław school of applied mathematics was multivariate analysis, and the most important achievements concern the new method of graphic taxonomy, especially the theory and applications (algorithms) of graphs. The optimal graph model was named Wrocław's dendrite (Statystycy Polscy, 2012).

The latest trends in the development of statistical methods of experimentation

Recent years have brought considerable progress in statistical methods of experimentation that are more demanding in terms of computation. This is mainly due to the widespread use of computers and statistical software (SAS, GENSTAT, Statistica, R, ASReml, IBM SPSS and XLStat). In such circumstances, since the turn of the 21st century, mixed models and methods have been widely developed and used to analyze classified balanced data, but mostly unbalanced data. Such data are acquired mainly from series of replicated and long-term experiments on cultivars and agronomic objects for reliable and comprehensive assessment of the agronomic, ecological and economic values of new cultivars and agronomic practices, expressed in terms of their stability and adaptability for important agricultural traits (Caliński et al., 2005; Smith et al., 2005; Van Eeuwijk et al., 2016; Studnicki et al., 2017). Effective progress has also been made in the application of multivariate methods, for which foundations were already developed much earlier. These include cluster analysis and PCA with their numerous modifications for quantitative and categorized variables for the classification of genotypes (Crossa and Franco, 2004). The classical methods used in agricultural experimentation, derived from PCA, are the AMMI and GGE procedures, based on fixed or mixed additive and multiplicative models, i.e. AMMI and GGE. AMMI and GGE are used to assess, visualize, interpret and exploit the genotype-environment interactions in agriculture. These factor interactions are very important for breeding, assessment and recommendations on cultivars, based on balanced or unbalanced classified data obtained from a series of cultivar trials (Smith et al., 2005; Gauch et al., 2008; Van Eeuwijk et al., 2016; da Silva et al., 2019).

In recent years an increasing role has been attributed to geostatistical methods, whose aim is to analyse geographic data for different spatial coverage. With regard to agricultural research, this applies, for example, to precision agriculture, including advanced interpolation of various soil properties. A new concept in this field is pedometrics, a discipline created mainly by the initiative of researchers from the Universities of Sydney (Australia) and Wageningen (Netherlands). The objective of pedometrics is to use a range of quantitative methods, especially geostatistics, for mapping properties of the soil at different spatial scales.

There is a growing amount of data available from different agricultural experiments and for this reason, like in any disciplines of science, meta-analyses of data from agricultural experiments are gaining more and more importance. Various statistical methods for aggregate data are used in meta-analyses, including fixed and mixed linear and nonlinear models for incomplete (unbalanced) data. Analysis of data from agricultural experiments increasingly often relies on machine learning, including models that employ artificial neural networks, (ANN), decision trees, support vector machine (SVM) and Bayesian networks for the prediction of variables (e.g. yield) or the incidence of specific events (e.g. plant infestation by diseases). This category also includes advanced crop simulation models, which are especially useful for predicting the growth and yield of crops in research based on the agricultural effects of climate change.

Rapid advances in molecular biology have stimulated a demand for statistical methods for analyzing data on the expression, evolution and structure of genes. One such method is quantitative trait locus (QTL) mapping, which explains relationships between phenotypes and genotypes in living organisms. Other methods used for this purpose are variance analysis, composite interval mapping (CIM), and profiling gene expression using DNA microarrays. One method for the assessment of genetic similarity is the analysis of molecular variance (AMOVA). Molecular phylogenetics or genotype classification based on molecular data (e.g. DNA sequencing) employs Markov Chain Monte Carlo (MCMC).

Acknowledgements

The authors are very grateful to Professor dr hab. Tadeusz Caliński for his effort in reading the original version of this text, and his kind attitude, comments and suggestions on many justified changes to improve the content.

The authors also extend their sincere thanks to two anonymous reviewers whose fundamental and detailed remarks helped to improve the quality of the manuscript.

References

Barbacki S., Fisher R. 1936. A test of the supposed precision of systematic arrangements. Annals of Eugenics 7: 189 — 193.

Caliński T. 2012. Rozwój i osiągnięcia w biometrii polskiej. Przegląd statystyczny, Numer specjalny I: 47 — 52.

Caliński T., Camussi A., Ottaviano E., Kaczmarek Z. 1985. Genetic distances based on quantitative traits. Genetics, 111:945 — 962.

Caliński T., Czajka S., Kaczmarek Z., Krajewski P., Pilarczyk W. 2005. Analyzing multi-environment variety trials using randomization-derived mixed models. *Biometrics*, 61:448 — 455.

Caliński T., Harabasz J. 1974. A dendrite method for cluster analysis. *Communications in Statistics* 3: 1 — 27.

Cochran W. G. 1939. Long-term agricultural experiments. J. Roy. Stat. Soc. (Suppl.) 6: 104–148.

Crossa J., Franco J. 2004. Statistical methods for classifying genotypes. Euphytica 137: 19 — 37.

da Silva C. P., de Oliveira L. A., Nuvunga J. J., Pamplona A. K. A., Balestre M. 2019. Heterogeneity of variances in the Bayesian AMMI model for multienvironment trial studies. Crop Sci. 59: 2455 — 2472.

Elandt R. 1964. Statystyka matematyczna w zastosowaniu do doświadczalnictwa rolniczego. PWN, Warszawa

Fisher R. A. *1925.* Applications of "Student's" distribution*. Metron 5: 90–104.*

Fisher R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52: 399 — 433.

Fisher R. A. 1921. Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. The J. Agri. Sci. 11: 107 — 135.

Fisher R. A. 1922. On the mathematical foundations of theoretical statistics, *Philos. Trans. Roy. Soc., Lon. Ser. A,* 222**:** 309 — 368.

Fisher R. A. 1925. Statistical methods for research workers, 1st ed. Oliver and Boyd, Edinburgh.

Fisher R. A. 1935 a. The logic of inductive inference (with discussion). J. Roy. Stat. Soc. 98: 39 — 82.

Fisher R. A. 1935 b. The design of experiments, 1st ed. Oliver and Boyd, Edinburgh.

Fisher R. A. 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7: 179 — 188.

Fisher R. A., Mackenzie W. A. 1923. Studies in crop variation. II. The manurial response of different potato varieties. *J. Agric. Sci.* 13: 311 — 320.

Fisher R. A., Yates F. *1936*. *Statistical Tables* for Biological, Agricultural and Medical Research. Oliver and Boyd, Edinburgh.

Gauch H.G., Jr., Piepho H.P., Annicchiarico P. 2008. Statistical analysis of yield trials by AMMI and GGE: Further considerations. Crop Sci. 48:866–889

Greenland S., Senn S. J., Rothman K. J., Carlin J. B., Poole C., Goodman S. N., Altma D. G. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur. J. Epidemiol. 31:337 — 350.

Hall A. D. 1909. The experimental error in field trials. J. Board Agr. 16: 365 — 370.

Hotelling H. 1931. *The generalization of Student's ratio. Ann. Math. Stat. 2: 360 — 378.*

Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol., 24: 417 — 441.

Hurlbert S. H., Lombardi C. M. 2009. Final collapse of the Neyman-Pearson decision theoretic framework and rise of the neoFisherian. Ann. Zool. Fennici 46: 311 — 349.

Johnson N. L., Kotz S. 1997. Leading personalities in statistical sciences. John Wiley & Sons, New York.

Lehman E. L. 2011. Fisher, Neyman and the creation of classical statistics. Springer, New York.

Kozak M. 2004 a. Alokacja próby między warstwy w przypadku cechy wielowymiarowej. Wiadomości statystyczne Nr 7: 13 — 21.

Kozak M. 2004b. Efektywność schematów losowania w badaniach gospodarstw rolnych. Wiadomości statystyczne Nr 9:20 — 26

Mahalanobis P. 1930. On tests and measures of group divergence I. Theoretical formulae. J. and Proc. Asiat. Soc. of Bengal, 26: 541 — 588.

Mahalanobis P. 1936. On the generalized distance in statistics. Proc. Nat. Inst. Sci. 2:49 — 55.

Miller R.G. 1981. Simultaneous statistical inference. Springer-Verlag, New York.

Newman D. 1939. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. Biometrika 31: 20 — 30.

Neyman J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. J. Roy. Stat. Soc., 97:558 — 625

Neyman J. 1937. Outline of a theory of statistical estimation based on the classical theory of probability. Philos. Trans. Roy. Soc. London, Series A. 236: 333 — 380.

Neyman J. 1979. Narodziny statystyki matematycznej. Wiadomości Matematyczne, 22: 91 — 106.

Neyman J., Pearson E. S. 1928. On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika 20A, Pt. I: 175 — 240; Pt. II: 263–294.

Neyman J., Pearson E. S. 1933. On the problem of the most efficient tests of statistical hypotheses. Phil. Trans. Roy. Soc. London. Series A, 231: 289 — 337.

Oktaba W. 2002. Historia teorii eksperymentu. Lubelskie Towarzystwo Naukowe, Lublin.

Ostasiewicz W. 2012. Rozwój myśli statystycznej w Polsce w XIX wieku. Przegląd statystyczny, Numer specjalny I: 34 — 46.

Pearson K. 1*901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2:559–572*).

Scheffe H. 1959. The analysis of variance. John Wiley & Sons, Inc., N. Y.

Smith A. B., Cullis B. R., Thompson R. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. J. Agric. Sci. 143: 449 — 462

Speed T. P. 1992. Introduction to Fisher (1926) — The arrangement of field experiments. W: Breakthroughs in statistics, Vol. II. (Ed by Kotz S., Johnson N. L.), Springer-Verlag, New York: 71 — 82.

Spława-Neyman J. 1923. Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych. Roczn. Nauk Roln. i Leśnych, 10: 1 — 51.

Statystycy Polscy. 2012. (Redakcja Adamczewski W. i in.) GUS i Polskie Towarzystwo Statystyczne, Warszawa

*Student [*W.S. Gosset*]. 1908.* The probable error of a mean. Biometrika *6: 1 — 25.*

Studnicki M., Paderewski J., Piepho H.P., Wójcik-Gront E. 2017. Prediction accuracy and consistency in cultivar ranking for factor-analytic linear mixed models for winter wheat multienvironmental trials *Crop Sci.* 57: 2506 — 2516.

Van Eeuwijk F. A., Bustos-Korts D. V., Malosetti M. 2016. What should students in plant breeding know about the statistical aspects of genotype × environment interactions? Crop Sci. 56: 2119 — 2140.

Yates F. 1935. Complex experiments. J. Roy. Stat. Soc. (Suppl.) 2: 181 — 247.

Romer E. 1989. Pamiętnik paryski (1918–1919). Wyd. Zakład Narodowy Imienia Ossolińskich, Wrocław.

Załęski E. 1927. Metodyka doświadczeń rolniczych. Wydawnictwo Rozpraw Biologicznych, Lwów.