A Trilogy of "Item" Biographies of Our Favorite Statisticians
Bruce Ratner, Ph.D.

I encountered a chance-moment of finding a series of three statistical items. Although, each is in one sense complete, the items have a close mutual relation to form a trilogy of "item" biographies of our favorite statisticians. I present: a clever, self-taught statistician; the statistician, who invented the second most used statistic; and perhaps, the most passionate statistician.

Einstein: A Clever, Self-taught Statistician

In December 1900 a 21-year-old recent college graduate student – Albert Einstein – submitted his first paper for publication. The paper used modeling (least-squares method, which had been well developed at that time) and data analysis in support of a couple of scientific propositions. The paper shows that Einstein was a clever, self-taught statistician handily using statistical tools and data in support of two scientific propositions, which can be summarized by Einstein in his conclusion: "To each atom corresponds a molecular attraction field that is independent of the temperature and of the way in which the atom is chemically bound to other atoms." Inexpectant, the paper shows that Einstein was an "average" data analyst as it indicates Einstein's tendency to make trivial arithmetic mistakes and some clumsiness in data recording. The objective of this fascinating article is to help provide a better appreciation of Einstein as an active user of statistical arguments in this and other if his important publications. (This abstract draws from Iglewicw, B., (2007) "Einstein's First Published Paper," *The American Statistician*, Vol. 61, No. 4, 339-342.)
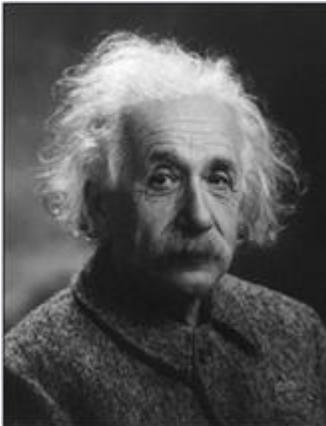
**Karl Pearson: Everybody Knows His Correlation Coefficient,
but Not How "Close" the Binomial Distribution is to a Normal Distribution**

Karl Pearson was born 150 years ago on March 27, 1857; he died April 27, 1936. He made important contributions to statistics; he is usually remembered for two path-breaking achievements: his "product-moment" estimate of the correlation coefficient (dating from 1896), and the chi-square test (introduced in 1900). But, on the 150th anniversary of his birth, there is a small striking discovery he made in 1895 that is virtually unknown today, yet well worth knowing.

Everybody knows that the binomial distribution is "like" a normal distribution if the number of independent trials (n) is large and the probability of success (p) on a single trial is not too near 0 or 1. Everybody knows this because Abraham De Moivre proved it to be true in 1733. Accordingly, every student knows that for most practical purposes, if you want to calculate the probability that a binomial count X will fall between two limits, P[a *le* X *le* b], where *le* means "less than or equals," you would be foolish to do other than use a normal approximation.

Because we are always reminded that this is an approximation, there is a nagging doubt as to how close the binomial and normal distributions really are. Pearson discovered something quite remarkable: There is a case where the agreement is much, much closer than anyone would have expected. In fact, if one particular definition of "agreement" is adopted, and if p = ½, the "agreement" is actually **exact** for all n (even for n =1), providing one minor fudge factor is allowed (replacing n+1 instead of n). Thus, Pearson discovered in 1895 that the normal and symmetric binomial distributions are more similar than De Moivre realized. I suspect most modern statisticians are equally unaware of this surprising identity, but now you know. (This abstract draws from Stigler, S., (2008) "Remembering Karl Pearson After 150 Years," *STATS*, Issue 49, 3-4.)

Karl Pearson in his office, 1934

**Florence Nightingale: You Know Her as the Pioneer of Modern Nursing,
But as a Passionate Statistician!**

Florence Nightingale (May 12, 1820 – August, 13 1910), who came to be known as "The Lady with the Lamp," was the pioneer of modern nursing, and a noted statistician. Florence Nightingale had exhibited a gift for mathematics from an early age and excelled in the subject under the tutorship of her father. She had a special interest in statistics, a field in which her father was an expert. Nightingale – a self-education statistician – made extensive use of statistical analysis in the compilation, analysis and presentation of statistics on medical care and public health.

Inspired by what she took as a "Christian divine calling," Nightingale had a strong desire to have a career in medicine. In 1851, Florence's father gave her permission to train as a nurse, despite her mother's disapproval because nursing was a career with a poor reputation, filled mostly by poorer women.

Nightingale was a pioneer in the visual presentation of information. In 1857, she invented the *pie chart*, which helped her disprove the medical assumptions of her day. Using fatality counts from the Crimean War, Nightingale develops a progressive series of statistical diagrams that reveal startling information: most soldiers did not die of their wounds, as reported, but in army hospitals, from diseases related to poor hygiene. When further data showed army death rates twice that of the civilian population, Nightingale traces the cause to overcrowded, disease-ridden barracks. She used in her report to the royal commission to force the British army to maintain nursing and medical care to soldiers in the field.

In her later life, Nightingale made a comprehensive statistical study of sanitation in Indian rural life and was the leading figure in the introduction of improved medical care and public health service in India.

In 1859, Nightingale was elected the first female member of the Royal Statistical Society and she later became an honorary member of the American Statistical Association.
(This abstract draws from the movie synopsis found from this link:
http://ffh.films.com/id/9104/The_Pasionate_statistican_Florence_Nightingale.htm)