Famous Female Statisticians

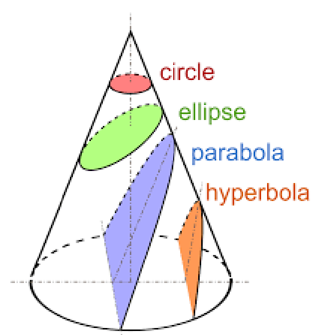
Darleen Ziege

1. Topic Selection and Introduction

I wanted to focus on another statistical application of mathematics but I was really inspired by Camilla Harris' talk about famous women in STEM and the discussion that followed so I decided to focus my attention on highlighting statistical concepts and forms of analysis specifically curated by famous women that were not traditionally taught in public school in my own personal experience. This paper is the result of that curiosity, which is a blend of something between a history, math, and women's studies application of mathematics.

2. Hypatia of Alexandria

Hypatia of Alexandria lived from around 355 A.D. to 415 A.D. and is therefore the earliest female mathematician of whose life and work reasonably detailed knowledge exists. In her time, she was the world's leading mathematician and astronomer, arguably the only known woman in history for whom such a claim can be made. Hypatia is well known for her work as a philosopher and teacher around 400 A.D. While historians have little to no evidence of her original mathematical work, most of her known contributions came in the form of written commentary on the works of others. Most significantly, Hypatia is credited with significant commentary on Apollonius' conics. The conics at this point mostly pertained to the words commonly associated with sections of cones, namely parabolas, ellipses, and hyperbolas. Another significant contribution Hypatia is credited with is preserving mathematical research documentation from the earliest centuries to allow these works to be studied for many centuries since her time.



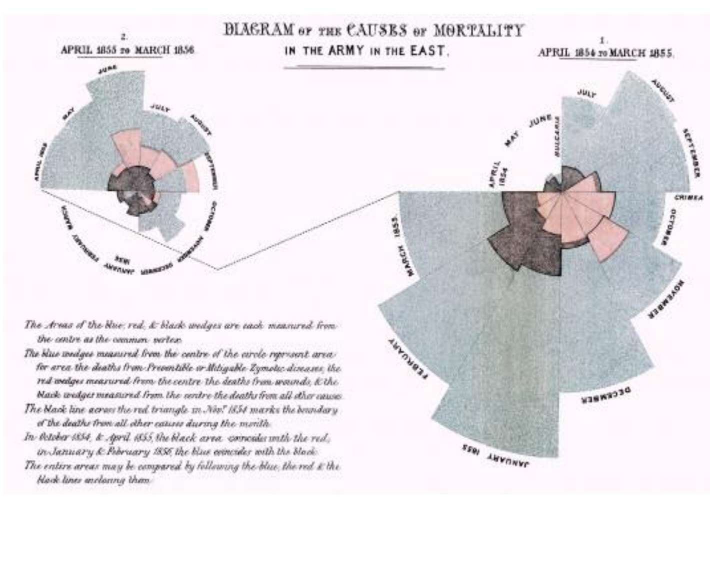
3. Florence Nightingale

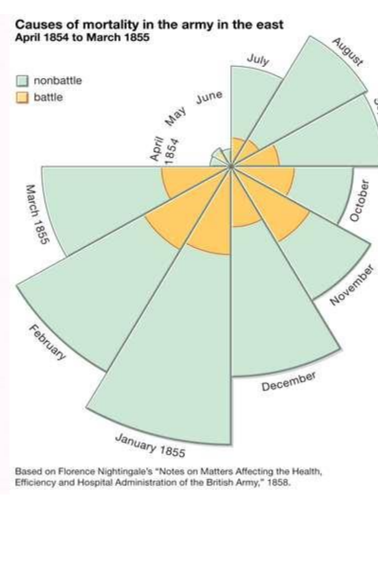
In 1820, Florence Nightingale was born as a wealthy woman in Victorian England at a time when it was frowned upon for women to receive formal educations or pursue professional careers outside of the home. Luckily, her father believed in equal education for women so he personally taught her various subjects including languages, philosophy, history, writing, and mathematics. When the Crimean war began, this allowed her to become a nurse and aid British soldiers by giving care during night rounds in hospitals. Throughout these rounds, she became known as the “Lady with the Lamp” and kept significant data records of sanitation and death rates among soldiers.

Upon her return to mainland England, Nightingale established the first scientifically based, secular nursing school and began a long career as a social reformer. She met with Queen Victoria and Prince Albert in 1856 to discuss British military reform surrounding the efficacy of nursing and medical staffs, as well as improved sanitation reform. Florence utilized new statistical visuals known as Coxcomb Charts to present her analyses of the data she collected at the Crimean War Barrack Hospital. For her work as a nurse, mathematician, and reformer, Nightingale was the first woman ever awarded the Order of Merit, became the first female member of the Royal Statistical Society, and became the namesake of the Nightingale Pledge.

Coxcomb Charts, also known as Polar Area Diagrams, Nightingale Rose Diagrams, or Circular Histograms are visuals like pie charts in that they indicate frequency by relative area, but different in use of fixed angles and variable radii. Each category in a Coxcomb Chart is represented by a section of the disc and has the same angle while the area of the section represents the value of the corresponding category. These are useful for demonstrating patterns rather than exact numbers since Coxcomb Charts make subtle differences difficult to see. They are also a useful form of data visualization for trend analysis that we utilize in many different fields, such as healthcare, finance, and business analytics. Below I have attached the original Nightingale Rose Diagrams, and for ease of understanding I have provided a transcription of the text in the bottom left of the original as well as a later adaptation of her research that may be easier to read:

“The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex. The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes. The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month. In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black. The entire areas may be compared by following the blue, the red, & the black lines enclosing them.”





4. Enid Charles

Enid Charles was a British socialist, feminist and statistician who lived from 1894 to 1972. She pioneered the studies of demography and population statistics as we know them today. Charles studied fertility rates and nuptiality in Canada, the United Kingdom, and Australia. She specifically projected sharp declines in the UK population if the fertility rate trends of the time continued, which was a reason she spoke out against the increasingly popular eugenics movement. She spent much of her life working for the World Health Organization as a health and population statistics consultant. Despite her well-informed statistical inferences, her working projections failed to predict the baby and marriage booms that followed the Second World War in the 1950's. Despite these failed projections, her research still laid the groundwork for the study of demography and population statistics we utilize today, such as the U.S. census that occurs every ten years. Below are images of some of her original works, utilizing graphical representations and plots that are still commonly utilized in the study of demography.

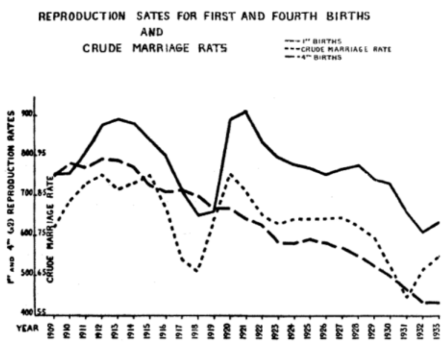


FIG. 5. Reproduction rates for first and fourth births and the crude marriage rate, 1909 to 1933.

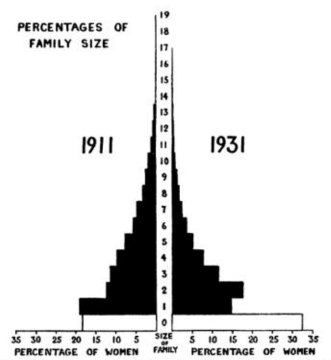


FIG. 2. Percentages of women having families of different sizes at current fertility rates, 1911 and 1931 (five-yearly averages).

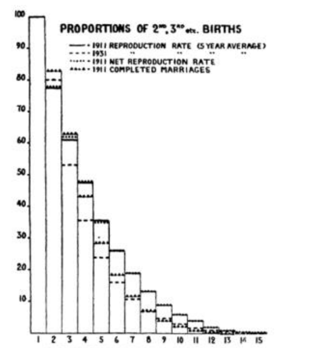


FIG. 3. Proportions of 2nd, 3rd, births to 1st in completed marriages of 1911 and according to 1911 and 1931 reproduction rates.

5. Florence Nightingale David

Florence Nightingale David (or F.N. David) was an English statistician named after Florence Nightingale who lived from 1909 to 1993. F.N. David studied mathematics with the intention of becoming an actuary but was rejected because actuarial firms only accepted men at the time. Instead, she began working for Karl Pearson; a renowned mathematician credited with establishing the field of mathematical statistics. While working for Pearson, David computed solutions to complex integrals and developed the distribution for Pearson's Correlation Coefficient, generally denoted as rho (ρ). The correlation coefficient between x and y is computed as

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

where cov(X, Y) is the covariance between random variable X and random variable Y, and sigma (σ) represents their standard deviations. Since covariance and standard deviations can be expanded into an expression of expected values, rho expands to

$$\rho_{X,Y} = \frac{\mathbb{E}[\,X\,Y\,] - \mathbb{E}[\,X\,]\,\mathbb{E}[\,Y\,]}{\sqrt{\mathbb{E}[\,X^2\,] - (\mathbb{E}[\,X\,])^2}\,\sqrt{\mathbb{E}[\,Y^2\,] - (\mathbb{E}[\,Y\,])^2}}.$$

where the expected values expression is equal to

$$E(X) = \sum_{x_i} x_i P(X = x_i),\ \ When\ X\ is\ a\ discrete\ random\ variable$$

$$E(g(X)) = \sum_{x_i} g(x_i) P(X = x_i),\ \ (g\ is\ an\ arbitrary\ function)$$

$$E(X) = \int_a^b x f_X(x)dx \quad (a \le X \le b)\ ,When\ X\ is\ a\ continuous\ random\ variable$$

.

When solving this expression for x and y, the Pearson Correlation yields the distribution of R values that are utilized in general studies of probability. This was an incredibly tedious process at the time, which is what makes F.N. David's work so valuable. I've included the distributions below to highlight the extent of the work she contributed; each value is the outcome of a series of complex integrals solved anew.

df\ α	0.2	0.1	0.05	0.02	0.01	0.001
1	0.951057	0.987688	0.996917	0.999507	0.999877	0.999999
2	0.800000	0.900000	0.950000	0.980000	0.990000	0.999000
3	0.687049	0.805384	0.878339	0.934333	0.958735	0.991139
4	0.608400	0.729299	0.811401	0.882194	0.917200	0.974068
5	0.550863	0.669439	0.754492	0.832874	0.874526	0.950883
6	0.506727	0.621489	0.706734	0.788720	0.834342	0.924904
7	0.471589	0.582206	0.666384	0.749776	0.797681	0.898260
8	0.442796	0.549357	0.631897	0.715459	0.764592	0.872115
9	0.418662	0.521404	0.602069	0.685095	0.734786	0.847047
10	0.398062	0.497265	0.575983	0.658070	0.707888	0.823305
11	0.380216	0.476156	0.552943	0.633863	0.683528	0.800962
12	0.364562	0.457500	0.532413	0.612047	0.661376	0.779998
13	0.350688	0.440861	0.513977	0.592270	0.641145	0.760351
14	0.338282	0.425902	0.497309	0.574245	0.622591	0.741934
15	0.327101	0.412360	0.482146	0.557737	0.605506	0.724657
16	0.316958	0.400027	0.468277	0.542548	0.589714	0.708429
17	0.307702	0.388733	0.455531	0.528517	0.575067	0.693163
18	0.299210	0.378341	0.443763	0.515505	0.561435	0.678781
19	0.291384	0.368737	0.432858	0.503397	0.548711	0.665208
20	0.284140	0.359827	0.422714	0.492094	0.536800	0.652378
21	0.277411	0.351531	0.413247	0.481512	0.525620	0.640230
22	0.271137	0.343783	0.404386	0.471579	0.515101	0.628710
23	0.265270	0.336524	0.396070	0.462231	0.505182	0.617768
24	0.259768	0.329705	0.388244	0.453413	0.495808	0.607360
25	0.254594	0.323283	0.380863	0.445078	0.486932	0.597446
26	0.249717	0.317223	0.373886	0.437184	0.478511	0.587988
27	0.245110	0.311490	0.367278	0.429693	0.470509	0.578956
28	0.240749	0.306057	0.361007	0.422572	0.462892	0.570317
29	0.236612	0.300898	0.355046	0.415792	0.455631	0.562047
30	0.232681	0.295991	0.349370	0.409327	0.448699	0.554119

df\ α	0.2	0.1	0.05	0.02	0.01	0.001
35	0.215598	0.274611	0.324573	0.380976	0.418211	0.518898
40	0.201796	0.257278	0.304396	0.357787	0.393174	0.489570
45	0.190345	0.242859	0.287563	0.338367	0.372142	0.464673
50	0.180644	0.230620	0.273243	0.321796	0.354153	0.443201
60	0.164997	0.210832	0.250035	0.294846	0.324818	0.407865
70	0.152818	0.195394	0.231883	0.273695	0.301734	0.379799
80	0.142990	0.182916	0.217185	0.256525	0.282958	0.356816
90	0.134844	0.172558	0.204968	0.242227	0.267298	0.337549
100	0.127947	0.163782	0.194604	0.230079	0.253979	0.321095
125	0.114477	0.146617	0.174308	0.206245	0.227807	0.288602
150	0.104525	0.133919	0.159273	0.188552	0.208349	0.264316
175	0.096787	0.124036	0.147558	0.174749	0.193153	0.245280
200	0.090546	0.116060	0.138098	0.163592	0.180860	0.229840
250	0.081000	0.103852	0.123607	0.146483	0.161994	0.206079
300	0.073951	0.094831	0.112891	0.133819	0.148019	0.188431
350	0.068470	0.087814	0.104552	0.123957	0.137131	0.174657
400	0.064052	0.082155	0.097824	0.115997	0.128339	0.163520
450	0.060391	0.077466	0.092248	0.109397	0.121046	0.154273
500	0.057294	0.073497	0.087528	0.103808	0.114870	0.146436
600	0.052305	0.067103	0.079920	0.094798	0.104911	0.133787
700	0.048427	0.062132	0.074004	0.087789	0.097161	0.123935
800	0.045301	0.058123	0.069234	0.082135	0.090909	0.115981
900	0.042711	0.054802	0.065281	0.077450	0.085727	0.109385
1000	0.040520	0.051993	0.061935	0.073484	0.081340	0.103800
1500	0.033086	0.042458	0.050582	0.060022	0.066445	0.084822
2000	0.028654	0.036772	0.043811	0.051990	0.057557	0.073488
3000	0.023397	0.030027	0.035775	0.042457	0.047006	0.060027
4000	0.020262	0.026005	0.030984	0.036773	0.040713	0.051996
5000	0.018123	0.023260	0.027714	0.032892	0.036417	0.046512

Additional contributions by David include statistical models in 1939 to predict possible consequences of bombs exploding in high density populations such as the city of London including numbers of living and dead, reactions to fires and damages to buildings and utilities; models which were updated as World War II progressed. She also pursued extending the Markoff Theorem on Least Squares, a truncated version of the Poisson distribution, and contagious distributions in plant populations, just to name a few. Her discoveries and contributions are still utilized to this day in probability theory, modeling and regression, and biology.

6. Conclusions

While there are many contributions in statistics by women, these are a few of the most iconic female statisticians that I chose to focus on. It is remarkable to note that many of these women became well known for their accomplishments despite the sexism and various adversities they faced throughout their careers and general lives. What's important to note is that even though this list is by no means holistic or even comprehensive, women of color in statistics have had an even more strenuous path to recognition for their contributions in the field. The women I've focused on – except for Hypatia due to the early time period – are all from western, European cultures that despite the struggles of women's recognition, have still received more pathways to professional success than their eastern and non-European counterparts. While I personally did not choose to focus on any such intersectional representation in this presentation (frankly because such women are much harder to research), I find it important to note in the concluding portion of this paper that it is a necessary disparity to analyze and learn from other such women as well.

7. References

Charles, Enid. *The Changing Structure of the Family in Australia.* Economica, 1937.

Greenwald, Sarah & Mendez, Edith P. *Hypatia, the First Known Woman Mathematician,* Accessed 2020.

Homecoming and Legacy. Britannica, Accessed 2020.

Kruse, Megan. *Florence N. David (1909-1993)*. AmStat News, 2016.

Nightingale, Florence. *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army*, 1858

Pearson Correlation Coefficient. Wikipedia, Accessed 2020.

[Prob & Stat] 3. Expected Value, Variance, and Standard Deviation, 2019.

Riddle, Larry. *Polar-Area Diagram.* Agnes Scott College, 2019.

Wargon, Sylvia. *Legacy of Enid Charles, 1894-1972*, Canadian Studies in Population, 2005.