STAT 830
Some Statistical History

Richard Lockhart

Simon Fraser University

STAT 830 — Fall 2020

What I assume you already know

- None of this; it is motivational, not technical

What I want you to learn

- About Eugenics, Statistics, Regression.

Some pieces of history of statistics: names

- Sir Francis Galton (1822–1911), cousin of Charles Darwin, Eugenecist.
- Sir Karl Pearson (1857–1936), Founding Editor of *Biometrika*, Founding Editor of *Annals of Eugenics*.
- Sir Ronald Aylmer Fisher (1890–1962), Geneticist, Mathematical Statistician, Eugenecist. Inventor of Likelihood, ANOVA.
- Jerzy (Spława-)Neyman (1894–1981), Egon Sharpe Pearson (1895–1980). Inventors of Frequency Theory paradigm.

Journals

- *Journal of the Royal Statistical Society*
- *Journal of the American Statistical Association*
- *Biometrika*
- *Annals of Mathematical Statistics*

The Statistical Society of London

- Founded 1834.
- Became Royal Statistical Society.
- Published *Journal of the Statistical Society of London*
- Began May 1838.
- First issue has Editorial defining statistics and giving "History of Statistics".
- Later *J Royal Statistical Society*

J Stat Soc London Volume 1, Number 1

JOURNAL

OF THE

STATISTICAL SOCIETY OF LONDON.

MAY, 1838.

INTRODUCTION.

THE Council of the Statistical Society of London is of opinion that the time has arrived when the Fellows of the Society, and the public, will hail with satisfaction the appearance of a Journal devoted to the collection and comparison of Facts which illustrate the condition of mankind, and tend to develop the principles by which the progress of society is determined.

It is within the last few years only that the Science of Statistics has been at all actively pursued in this country; and it may not, even now, be unnecessary to explain to general readers its objects, and to define its province. The word Statistics is of German origin, and is derived from the word *staat*, signifying the same as our English word *state*, or a body of men existing in a social union. Statistics, therefore, may be said, in the words of the Prospectus of this Society, to be the ascertaining and bringing together of those "facts which are calculated to illustrate the condition and prospects of society;" and the object of Statistical Science is to consider the results which they produce, with the view to determine those principles upon which the well-being of society depends.

The Science of Statistics differs from Political Economy, because, although it has the same end in view, it does not discuss causes, nor reason upon probable effects; it seeks only to collect, arrange, and compare, that class of facts which alone can form the basis of correct conclusions with respect to social and political government.

These are the objects to which, in prosecution of the ends of this Society, the Journal will be devoted; and the Council looks forward with confidence to the time when, through the exertions of its own Members and of corresponding Societies throughout the country, the Journal will become an important instrument for developing and

First Article: Pauper Children

On the Establishment of County or District Schools, for the Training of the Pauper Children Maintained in Union Workhouses. Part I

Author(s): James Phillips Kay

Source: *Journal of the Statistical Society of London*, Vol. 1, No. 1 (May, 1838), pp. 14-27

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2337851

One Table

The children above 2 years of age are divided, according to their civil condition, into the following classes :—

Bastards		543
Orphans		382
Children	deserted by father	279
,,	deserted by father and mother	54
,,	of men undergoing punishment for crime	171
,,	of persons dependent on parochial aid on account of mental or bodily infirmity	116
,,	of able-bodied widows resident in the Union Workhouse	144
,,	of able-bodied widowers resident in the Union Workhouse	36
,,	belonging to large families of able-bodied labourers, admitted into the Workhouse as relief to their parents	122
,,	of able-bodied parents resident in Workhouses	59
	Total	1906

The American Statistical Association

- Founded 1839 in Boston.
- First publications in 1888.
- First article in 1888 "Statistics of Water Power Employed in Manufacturing in the United States".
- Became *Journal of the American Statistical Association* in 1922.
- Table of contents; then *A "Statistical Paradox"*.

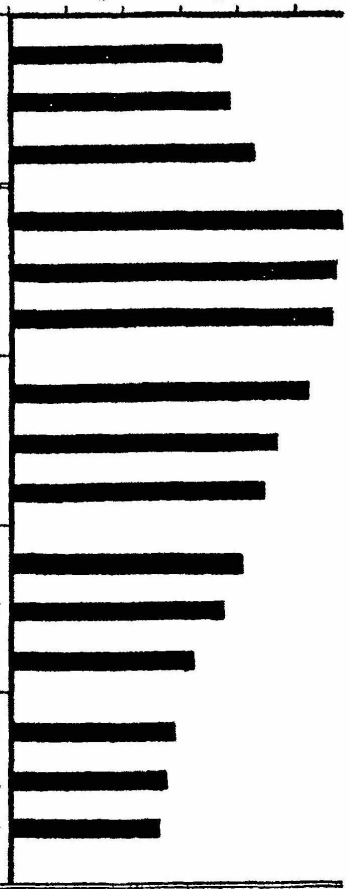
Table of Contents Extract

"Statistical Paradox"

RELATION OF PROPRIETORS' SALARIES TO TOTAL WAGES AND SALARIES

FOR STORES CLASSIFIED BY SIZE, 1919, 1918, AND 1914

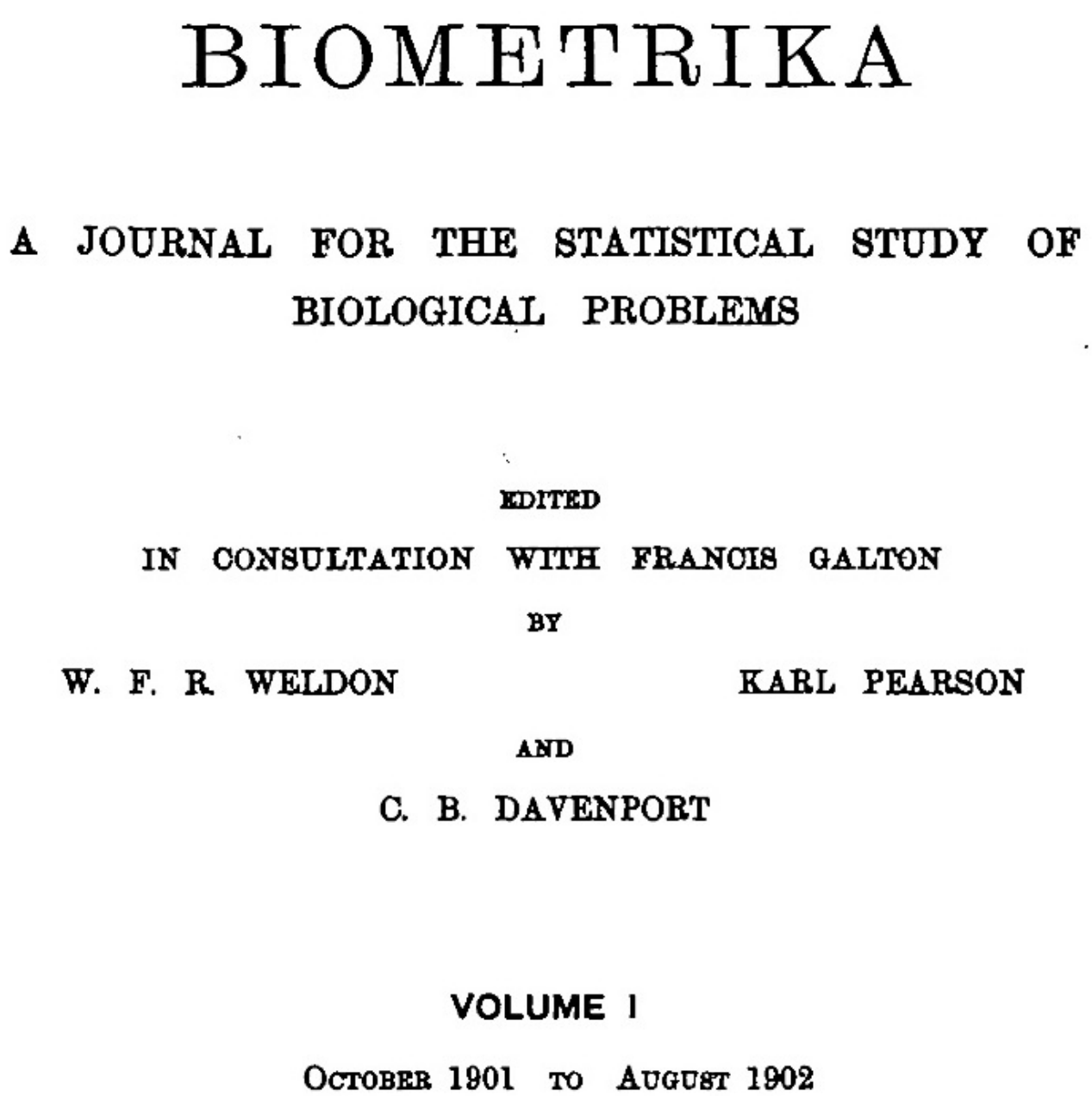
Classified Total Net Sales (in 000's)	Years	Number of Stores	Total Wages and Salaries	Proprietors' Salaries	Amount of Proprietors' Salaries per $100 of Total Wages & Salaries	
					Actual	Graphic
Total (Average)	1919	262	$3,872,634	$1,444,834	$37.31	
	1918	236	2,674,712	1,032,451	38.60	
	1914	136	957,031	408,680	42.70	
Under $40 (Group 4)	1919	37	140,794	81,671	58.01	
	1918	76	283,330	163,519	57.71	
	1914	71	254,384	145,486	57.19	
$40 to $80 (Group 3)	1919	106	674,302	353,829	52.47	
	1918	90	655,131	307,943	47.00	
	1914	45	349,922	155,814	44.53	
$80 to $180 (Group 2)	1919	81	1,166,794	472,467	40.49	
	1918	53	871,706	324,019	37.17	
	1914	17	256,560	82,180	32.03	
$180 & over (Group 1)	1919	38	1,890,744	536,867	28.39	
	1918	17	864,545	236,970	27.41	
	1914	3	96,165	25,200	26.20	



Biometrika

- Founded 1901.
- First Editor Karl Pearson.
- Dedicated to measurement in support of study of natural selection
- Mixture of papers on techniques and on data.

Biometrika Cover



BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

EDITED

IN CONSULTATION WITH FRANCIS GALTON

BY

W. F. R. WELDON KARL PEARSON

AND

C. B. DAVENPORT

VOLUME I

OCTOBER 1901 TO AUGUST 1902

Biometrika Contents

Annals of Mathematical Statistics

- First volume in 1930.
- Taken over circa 1935 by new Institute of Mathematical Statistics.
- Dedicated to mathematical papers, unlike the other journals.
- Now all take some mathematics.
- But most publish only 'theorems' not *theorems*.

Francis Galton and Regression

- First cousin of Charles Darwin.
- Deeply committed to Darwin's ideas.
- Interested in *measurement* of evolution, natural selection, and genetics, especially in humans.
- Inventor of Regression as a term and as a statistical technique.

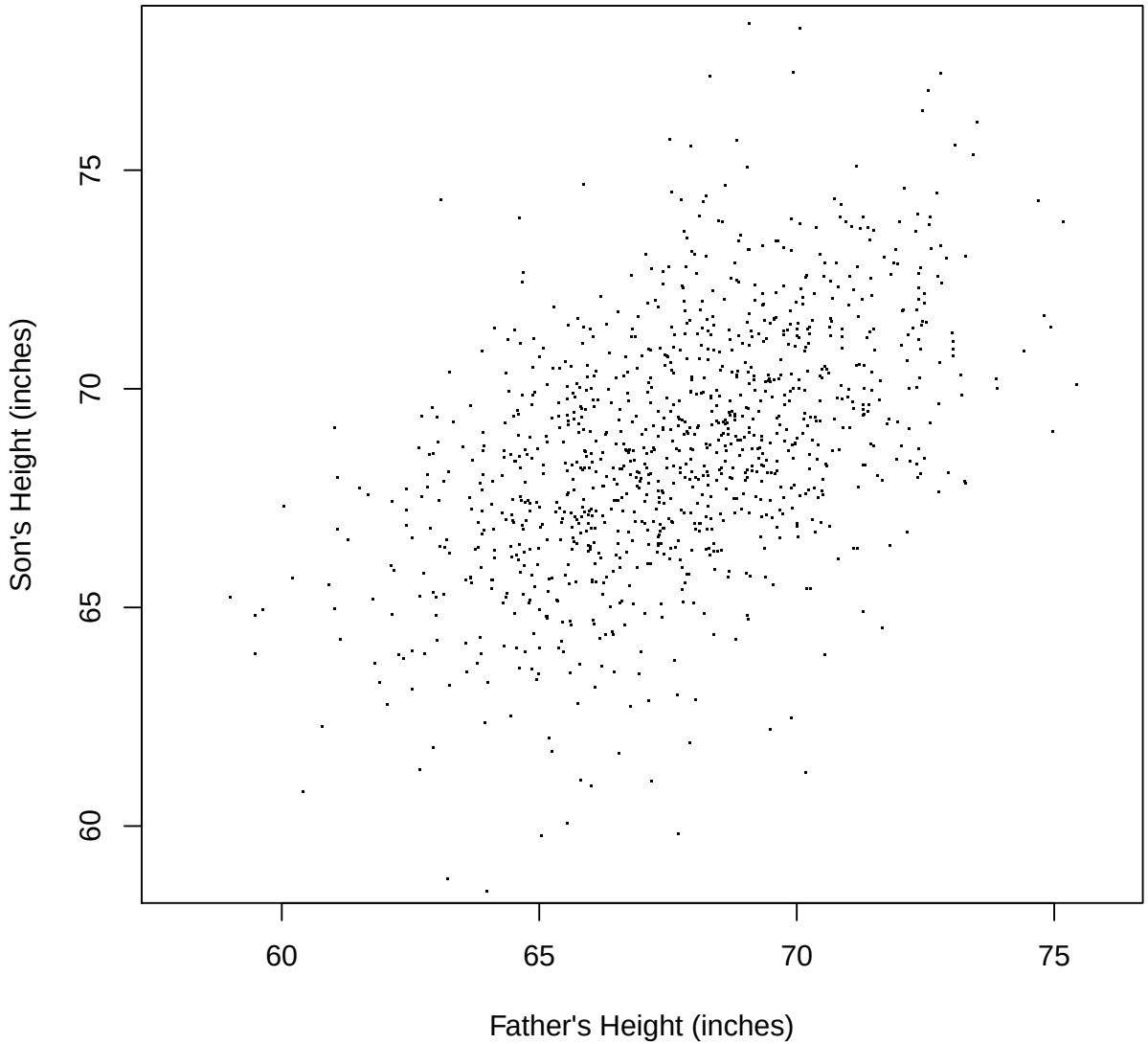
Modern notation and interpretation

DEF: If X, Y random the *regression* of Y on X is

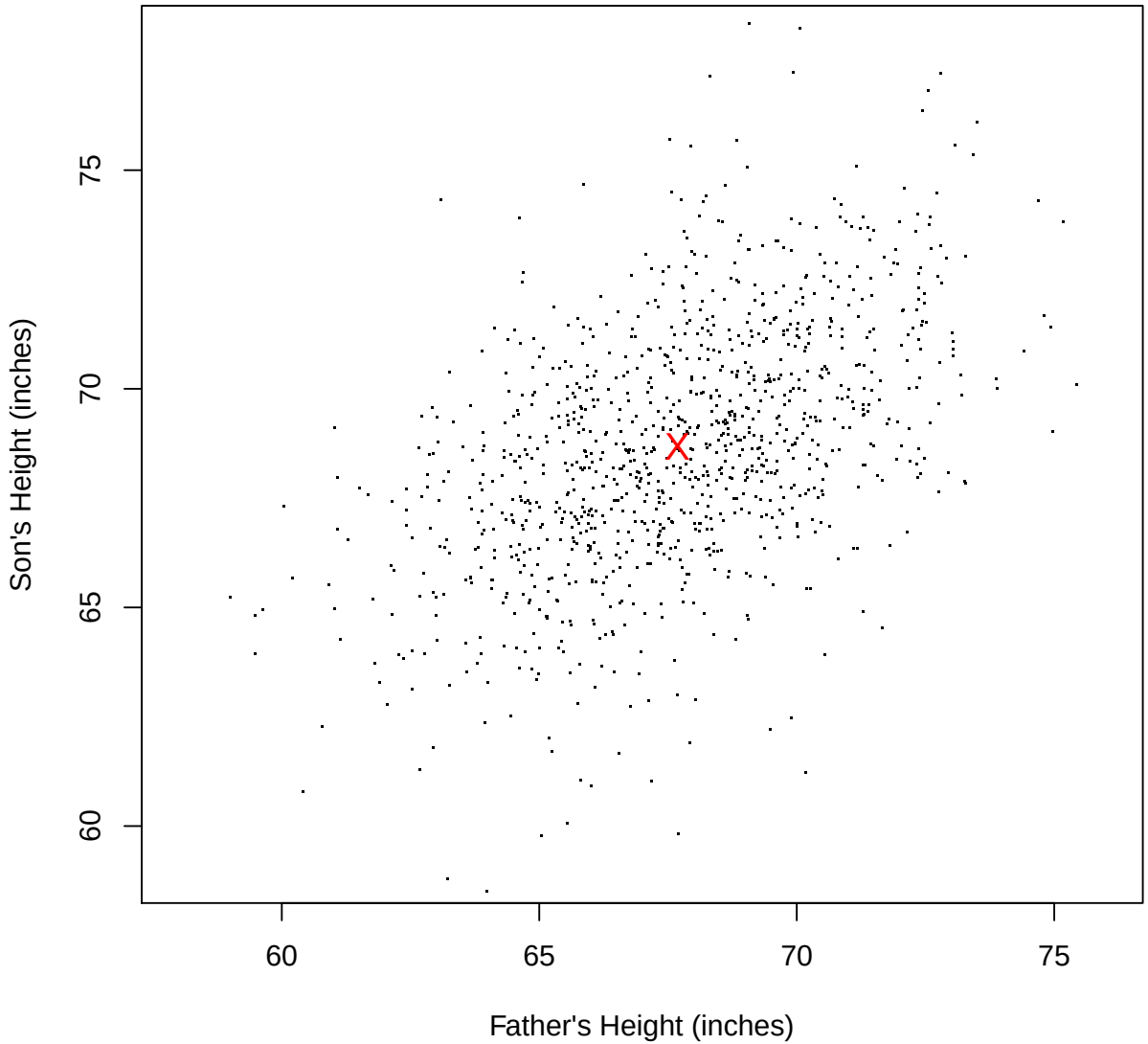
$$\mathrm{E}(Y|X = x).$$

- Example: X, Y are heights of Father, (adult) Son.
- 19th century data.
- Sons about an inch taller than fathers on average. (69 vs 68)
- What about tall fathers and their sons?

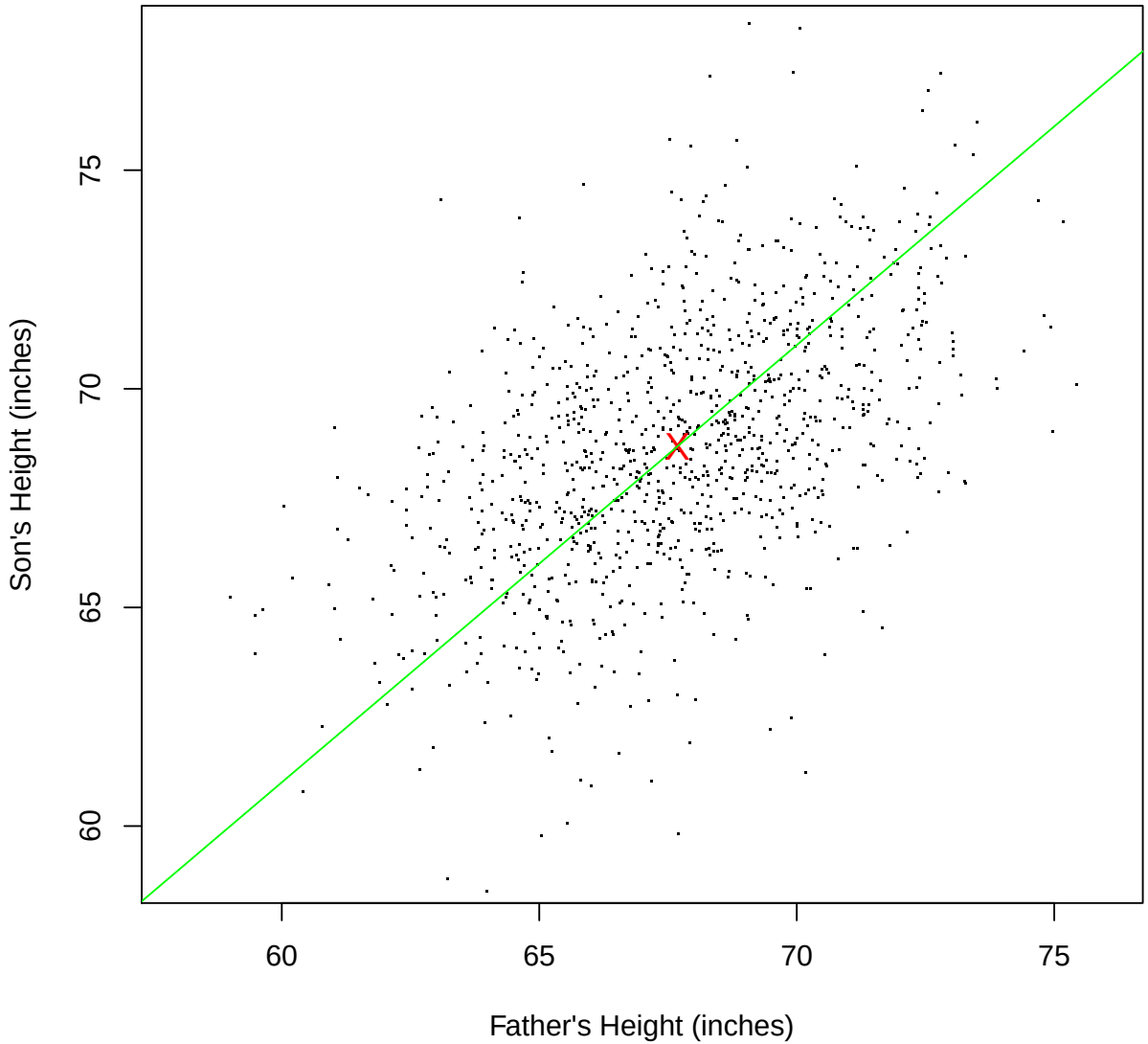
Son versus Father



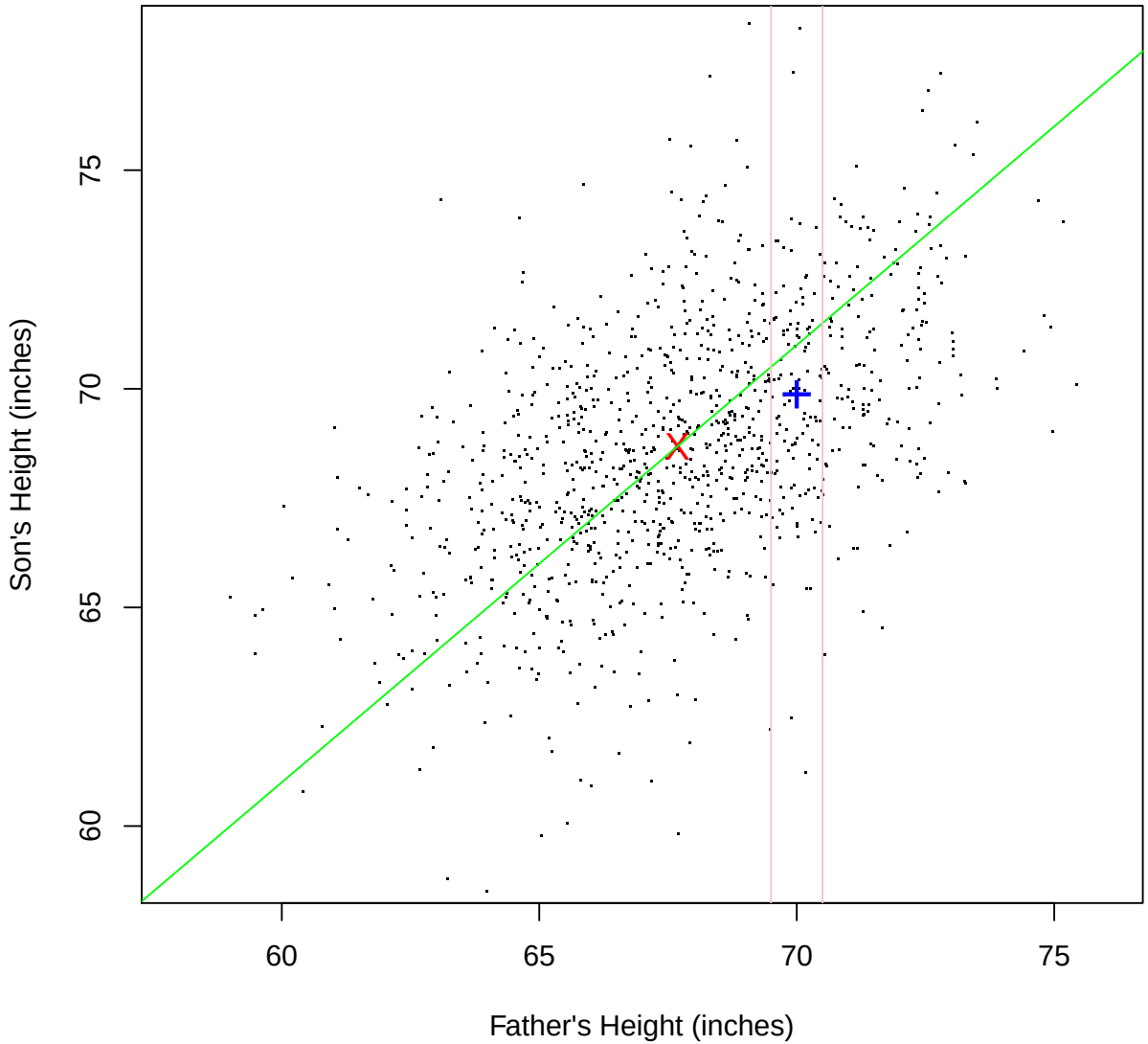
Add $(\bar{x}, \bar{y})$



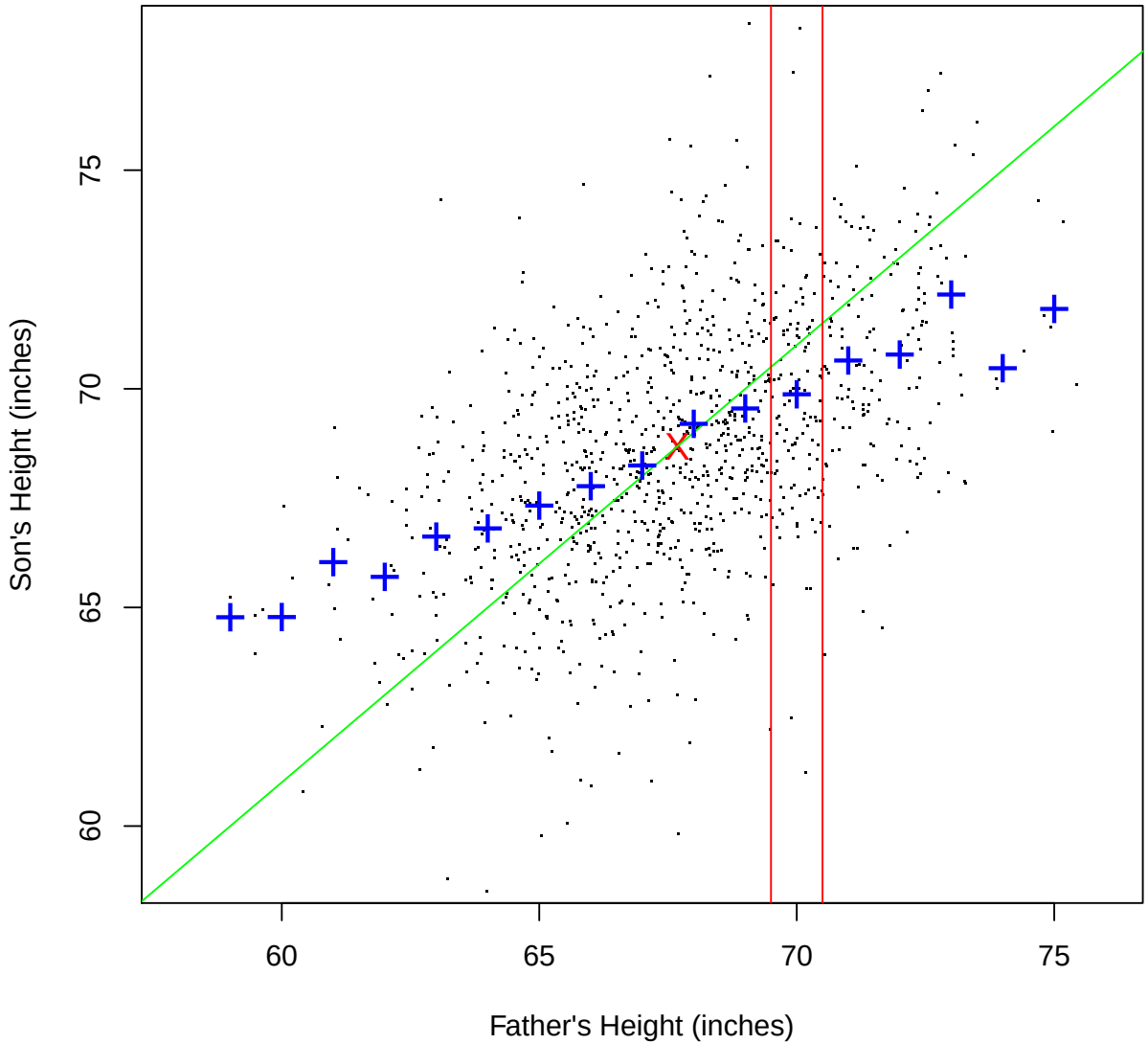
Add line $y = x + 1$


Son's Height (inches)
60
65
70
75
Father's Height (inches)
60
65
70
75

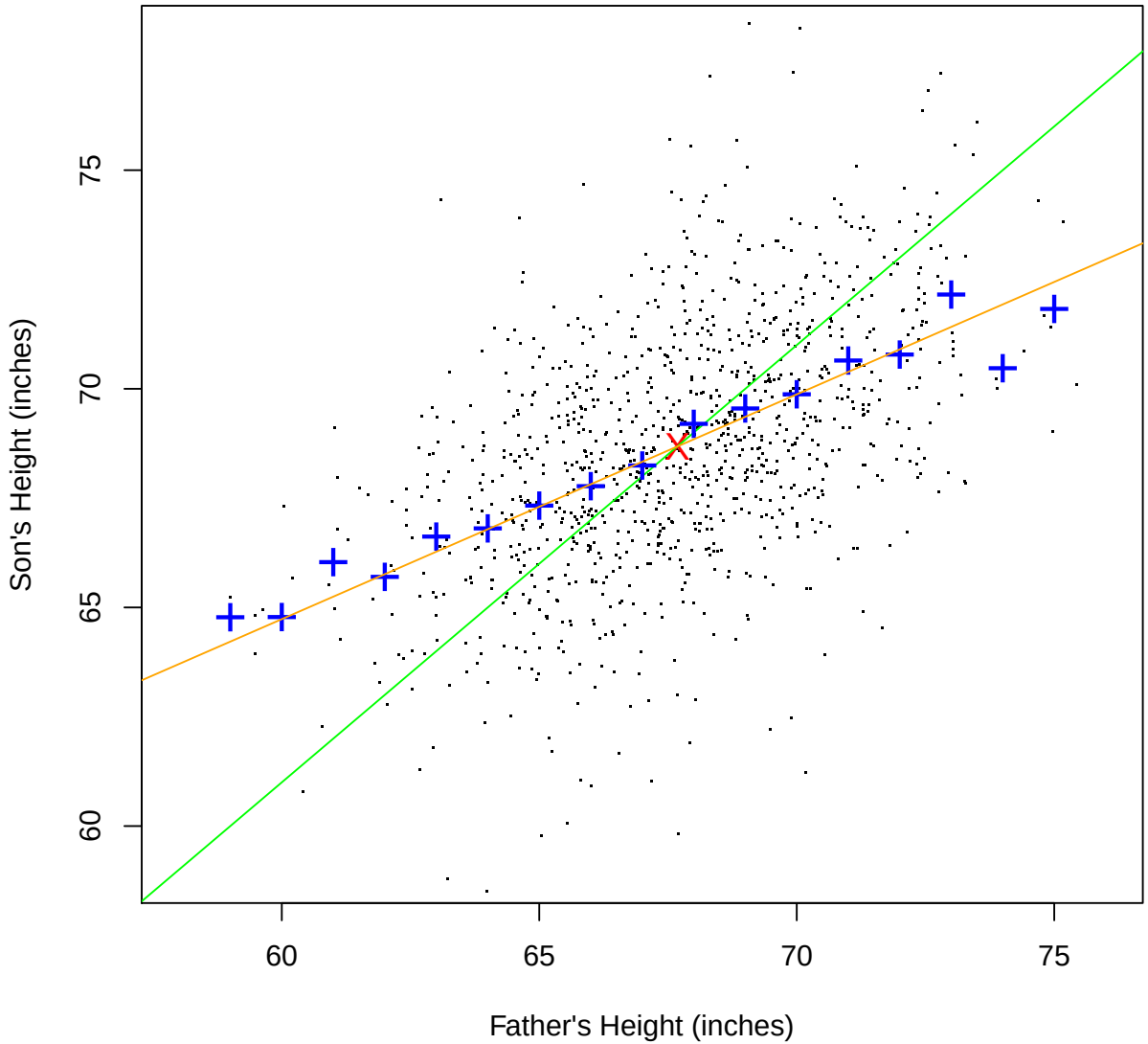
Fathers who are 70 inches tall



Average height of son for $x = 59 : 75$



Add regression line



Regression to the mean

- Galton observed sons are closer to average than fathers in each height group.
- Called this *regression* to the mean; sons are more "mediocre" than fathers.
- Turned this into mathematics about the normal (Gaussian) distribution.
- But the name is a misinterpretation: sons just as variable as fathers

Karl Pearson

- Founder of *Biometrika*
- *Eugenecist*
- Professor of Eugenics, U C London, 1911-1933.
- Founder of *Annals of Eugenics*, 1925.
- That became *Annals of Human Genetics*, 1945.

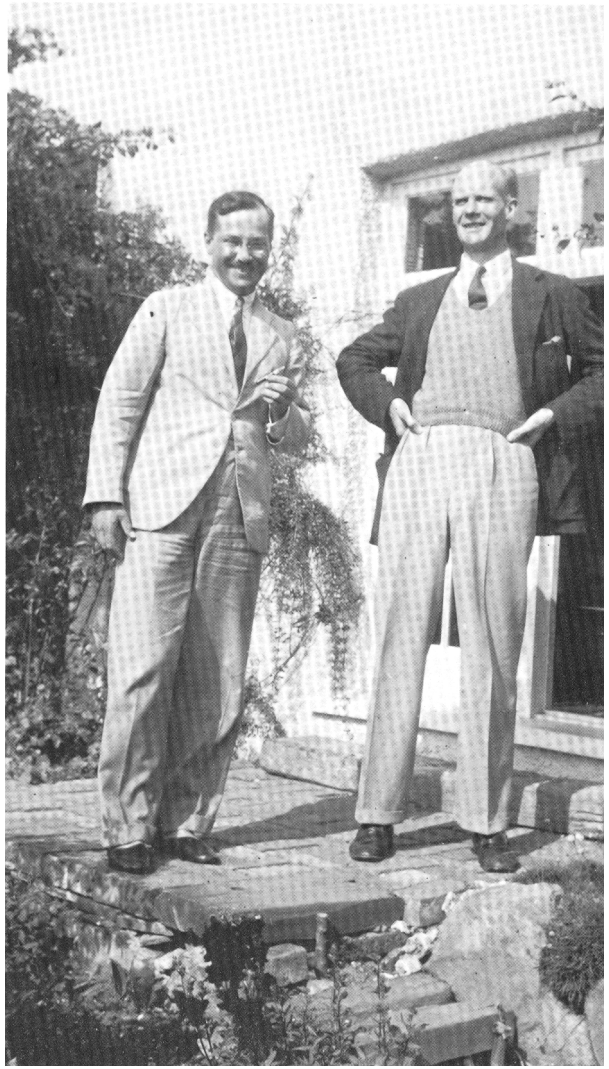
Opening Editorial: some quotes

- *The time seems fully ripe for the issue of a journal which shall devote its pages wholly to the scientific treatment of racial problems in man. . . .*
- *By whatever manner we approach heredity and selection in man, we still meet the dominating fact that probability lies at the basis of our knowledge. . . .*
- *It is not a quarter of a century since the Council of the Royal Society issued a solemn warning that mathematics must not be mixed with biology, thus proscribing biometry from their publications. . . .*
- *Wiser for our text-book writers to adopt a home-made term—Racial Hygiene—like the Germans, than purloin a word which its inventor attached to a very different range of ideas and methods. . . .*

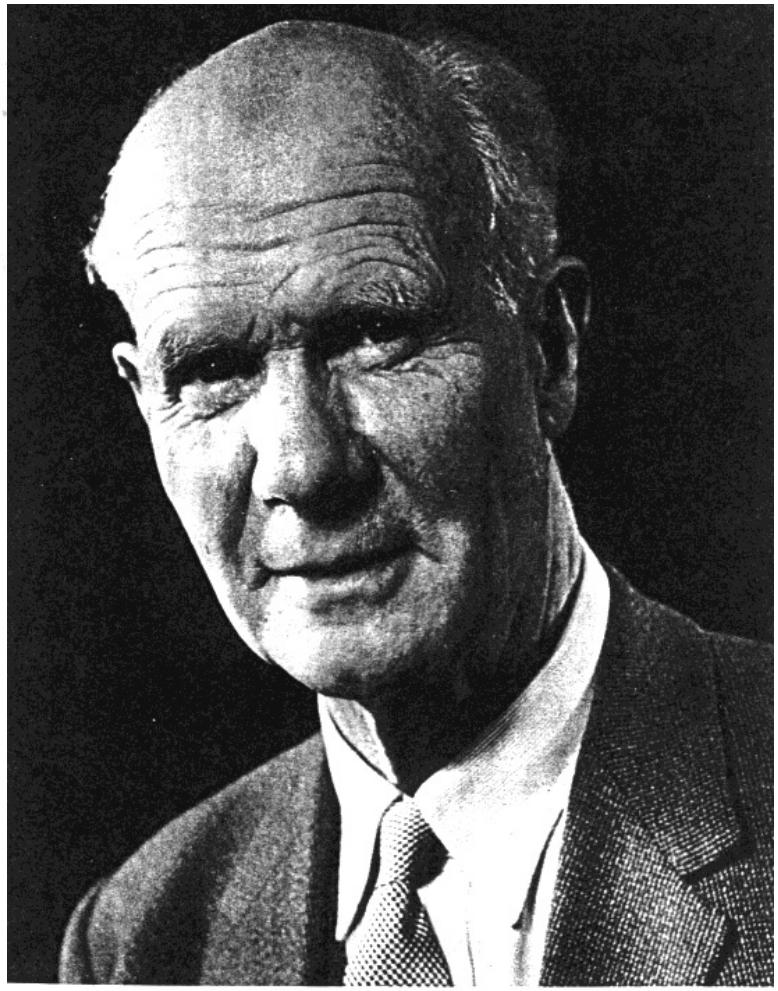
More quotes

- *Eugenics stands where Astronomy stood before the age of Newton and Laplace, much has been observed, little codified and all awaits its mathematical analysis. . . .*
- *Let us bear in mind the words of Galton written almost in the last years of his life, words not of despair, but of wise caution: When the desired fullness of information shall have been acquired, then and not till then, will be the fit moment to proclaim a "Jehad" or Holy War against customs and prejudices that impair the physical and moral qualities of our race. That has been the spirit in which the Laboratory he founded has been conducted, and that will be essentially our guide in the control of this journal.*

Neyman Pearson Statistics



The Pearsons



EGON SHARPE PEARSON

Berkeley department founded 1954

Jerzy Neyman

- Neyman and Pearson developed their ideas at University College London.
- Most important ideas: evaluate frequency behaviour of procedures: hypothesis tests and confidence intervals.
- Worst case analysis.
- Neyman came to London in 1925 to work with *Karl* Pearson.
- Instead worked with *Egon* Pearson.
- *Neyman from Life*, by Constance Reid.
- Scary classes for students.

Vigorous debate

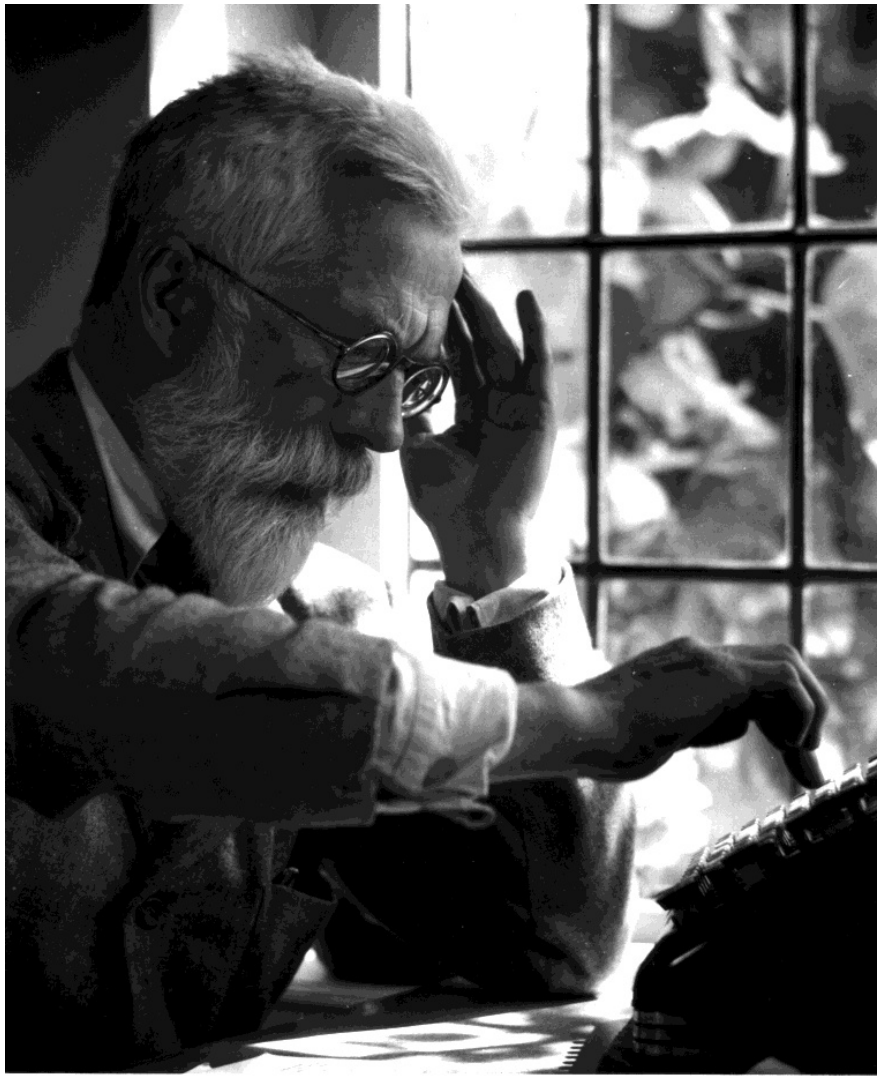
- "Vigorous" and "Vicious" are not so different.
- Last major player of period: Sir Ronald Aylmer Fisher.
- Only one paper in *Biometrika*: distribution of Pearson's r.
- Long battle with Karl Pearson. (Mathematics becomes personal.)
- Fisher battled others regularly.
- Eugenics: Fisher edited *Annals of Eugenics* after KP died.
- Second holder of *Professor of Eugenics* chair.

Fisher on Karl Pearson

- KP hated being wrong.
- He was wrong about degrees of freedom in chi-squared and Fisher said so.
- Fisher (*Annals of Eugenics*, 1937):

> *His example on this point is valuable; whereas he was a clumsy mathematician. Had it not been for his arrogant temper, his taste for numerical example might well have saved him from serious theoretical mistakes.*

Fisher and Neyman



R.A. Fisher, P.C. Mahalanobis, and Jerzy Neyman

Karl Pearson (On the difference between 0 correlation and independence):

> *That may be true in Poland, Mr Neyman, but it is not true here!*

Neyman (1938,*Biometrika*) on Fisher, 1937:

> *Readers of statistical journals can hardly fail to have noticed what appears to be a regular campaign carried on by Prof. R. A. Fisher to discredit the work of the late Karl Pearson. Owing to the tone and form of these depreciations, one feels reluctant to reply, but it seems to be useful to point out just one instance illustrating the methods used by Prof. Fisher.*

A handful of quotations 37

Fisher (*JASA*, 1943):

> *It is not my purpose to make Dr. Berkson seem ridiculous, nor, of course, to prevent him from providing innocent amusement. Had he looked up Hersh's original paper he would have been spared a blunder, ...*

Berkson (*Biometrics*, 1954):

> *I consented to comment on the remarks of Sir Ronald Fisher only with considerable reluctance. The passages of his article that have to do with my work are so far out of the bounds of reasonableness or relevancy that on first reading them I could only believe that he had been misinformed regarding my statement's ...*

On Bartlett

Fisher (*Biometrics*,1954):

> *It is a great pity that Cochran in this paper does not clearly point out that such adjustments have no useful function, at least finally, if it is intended to perform a correct analysis. The subsequent papers (5, 6) by Bartlett (1947) and Anscombe (1948)) show no such consciousness of the situation as they would have obtained had Cochran expressed himself more definitely.*
>
> *It is unfortunate that Bartlett did not restate his own views on this topic without making misleading allusions to mine.*

Fisher (*JRSS*-B, 1957):

> *If Professor Neyman were in the habit of learning from others he might profit from . . .*

Neyman (1951):

> *In particular, three major concepts were introduced by Fisher and consistently propagandized by him in a number of publications. These are mathematical likelihood as a measure of the confidence in a hypothesis, sufficient statistics, and fiducial probability. Unfortunately, in conceptual mathematical statistics Fisher was much less successful than in manipulatory, and of the three above concepts only one, that of a sufficient statistic, continues to be of substantial interest. The other two proved to be either futile or self-contradictory and have been more or less generally abandoned.*

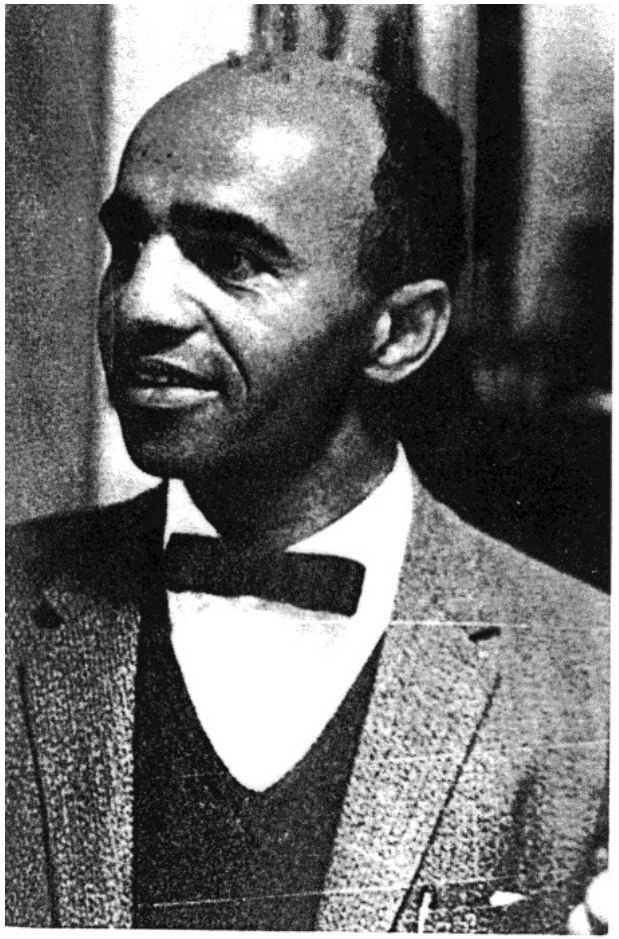
And from Pearson

> *Most readers will regret the inclusion of the Note on Paper (29,1937) which, if nothing more, shows in its last sentences, a profound ignorance of Karl Pearson's character and, indeed, of his contemporaries.*

and the footnote

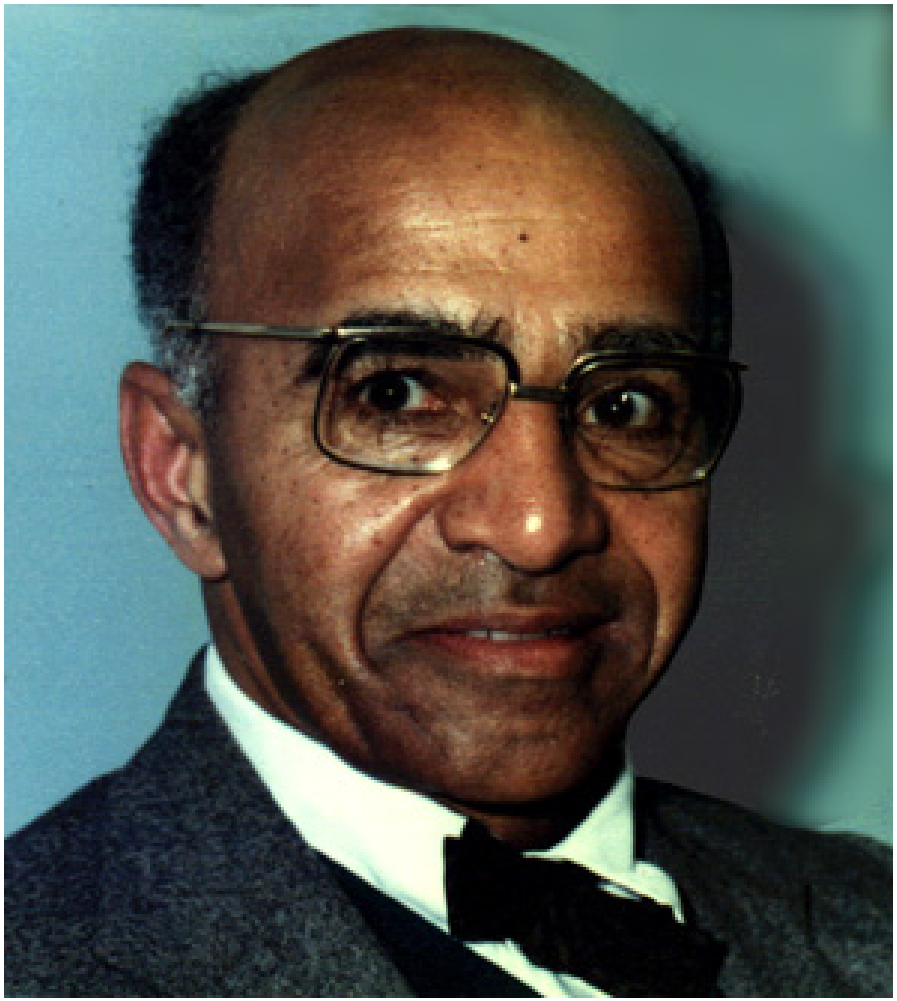
> *In view of the author's comment on the reason for the absence of the Metron papers, it should be made clear that permission to reproduce this paper from Biometrika would of course have been given had it been asked for. ED.*

David Harold Blackwell: the 50s maybe



David Blackwell

Picture: the 80s maybe



Some points about David

- First black member of National Academy of Sciences (1965); second was Wilson elected 1991. Two more in 2003!
- In Albers and Alexanderson *Mathematical People: Profiles and Interviews*

  > *I've worked in so many areas—I'm sort of a dilettante. Basically, I'm not interested in doing research and I never have been. I'm interested in understanding, which is quite a different thing.*

- Tribute is in *Notices of AMS*.
- Long written interview from 2002-2003 in *Bancroft Library Regional Oral History Project*.

Neyman picture

Neyman when I knew him

Jerzy (Jerry) Neyman and Elizabeth (Betty) Scott

Neyman Seminar Features 48

- Every Wednesday
- "Speak up, Jerry"; "I can't hear you, Betty."
- Research funding devoted to RA who bought cakes (3).
- Neyman knew numbers.

After the Talk

- Off to faculty club.
- Three toasts
- "The speaker".
- "The international intellectual community".
- "To all the ladies present and some of those absent".

Seminar taken over by Lucien Le Cam

The Coffee Room; an important place to learn

- Regulars: Neyman, Scott, Doksum, Le Cam, former student of Einstein.
- Saturday afternoons: Betty brought cake.
- Deserved by anyone working on a Saturday.
- We only lived a 10 minute walk from the cake.

Neyman summary

"Too bad!"
"Life is complicated but not uninteresting"