

# Student's $z$ , $t$ , and $s$ : What if Gosset had $\mathbb{R}$ ?

James A. Hanley<sup>1</sup>    Marilyse Julien<sup>2</sup>    Erica E. M. Moodie<sup>1</sup>

<sup>1</sup>Department of Epidemiology, Biostatistics and Occupational Health,

<sup>2</sup>Department of Mathematics and Statistics, McGill University

Société statistique de Montréal  
Statistical Society of Montreal

2008.03.07

# OUTLINE

Introduction

Theory

Simulations

AfterMath

Fisher

From  $z$  to  $t$

Messages

VOLUME VI

MARCH, 1908

No. 1

---

# BIOMETRIKA.

---

THE PROBABLE ERROR OF A MEAN.

By STUDENT.

William Sealy Gosset, 1876-1937

# MR. W. S. GOSSET

Obituary, The Times

## THE INTERPRETATION OF STATISTICS

“E.S.B.” writes:-

My friend of 30 years, William Sealy Gosset, who died suddenly from a heart attack on Saturday, at the age of 61 years, was known to statisticians and economists all over the world by his pseudonym “Student,” under which he was a frequent contributor to many journals. He was one of a new generation of mathematicians who were founders of theories now generally accepted for the interpretation of industrial and other statistics.

...

The eldest son of Colonel Frederic Gosset, R.E., of Watlington, Oxon, he was born on June 13, 1876. He was a scholar of Winchester where he was in the **shooting VIII**, and went up to **Oxford** as a scholar of New College and obtained **first classes in mathematical moderations in 1897 and in natural science (chemistry) in 1899**. He was one of the early pupils of the late Professor Karl Pearson at the Galton Eugenics Laboratory, University College, London. Over 30 years ago Gosset became **chief statistician to Arthur Guinness, Son and Company, in Dublin, as was quite recently appointed head of their scientific staff**. He was much beloved by all those with whom he worked and by a select circle of professional and personal friends, who revered him as one of the most modest, gentle, and brave of men, unconventional, yet abundantly tolerant in all his thoughts and ways. Also he loved sailing and fishing, and invented an angler's self-controlled craft described in the *Field* of March 28, 1936. His widow is a sister of Miss Phillpotts, for many years Principal of Girton College, Cambridge.

March 2, 1938

Dear Mrs Gosset,

I am preparing an obituary of Mr Gosset for the Annals of Eugenics, and should be glad to publish therewith a photograph of him if you could let me have one which you would ~~judge~~ deem suitable for this purpose.

Yours sincerely,

<http://digital.library.adelaide.edu.au/coll/special//fisher/>

March 8<sup>th</sup> 1938

DAVAN HOLLOW  
DENHAM  
BUCKS

Dear Dr. Fisher

Thank you for your letter. Unfortunately my husband would never consent to be photographed during recent years. I have had some snapshots enlarged, but they are not sharp enough to be of any use for reproduction.

I have a fairly good photo of him taken about 1908, but I suppose that would not do?

My husband's brother who lives in Calcutta, sent me

a cutting from an Indian paper, giving your speech at the Congress, & it was a great pleasure to me to read in it your tribute to 'Student's' work.

Yours sincerely  
Mayjory Gosset

## “STUDENT”

Annals of Eugenics  
1939



'Student' in 1908

The untimely death of W. S. Gosset (...) has taken one of the most original minds in contemporary science.

Without being a professional mathematician, he first published, in 1908, a fundamentally new approach to the classical problem of the theory of errors, the consequences of which are only still gradually coming to be appreciated in the many fields of work to which it is applicable.

The story of this advance is as instructive as it is interesting.

---

First paragraph, *Annals of Eugenics*, **9**, pp 1-9.

# Lehmann (Breakthroughs in Statistics, Vol.II Springer-Verlag 1992)

“one of the seminal contributions to 20th century statistics”

Introduction to “Student(1908). The Probable Error of a Mean” pp 29-32.  
Volume edited by NL Johnson, N L and S Kotz.

<http://www.guinness.com/>



the beer GUINNESS in  
Darkness Reigns log in

**GUINNESS®**

**EXPLORE THE HISTORY**

Crafted since 1759, GUINNESS® has a rich history: f

Navigate your way through time by clicking on your pr  
the GUINNESS® brewery and around the world.

**SELECT A TIME PERIOD**

<b>1700 &gt;</b>	<b>1750 &gt;</b>
<b>1950 &gt;</b>	<b>2000 &gt;</b>

**AT THE BREWERY...**

**1722**  
Richard Guinness becomes Land Steward to Rev Arthur Price, Archbishop of Cashel.

**1725**  
Arthur Guinness is born in Celbridge, County Kildare, Ireland.

**1893:**

T. B. Case becomes the **first university science graduate** to be appointed at the GUINNESS brewery.

It heralds the **beginning of 'scientific brewing'** at St. James's Gate.

<http://www.guinness.com/>

**1936**

**The first GUINNESS®  
brewery overseas is built  
at Park Royal, London.  
William Sealy Gossett,  
the father of modern  
statistics, is appointed  
Head Brewer.**

## XXIVth International Biometric Conference (Website)

“As always the IBC will be a great opportunity for scientific and social interchange - a place to present and learn of new work in biometry, and occasion to meet old and new friends, and the chance to visit a new country experiencing traditional Irish hospitality and the wonderful city of Dublin. **What better time and place to celebrate the centenary of Student's famous 1908 Biometrika paper on the t-distribution - W.S. Gossett (Student) worked in the Guinness Brewery in Dublin!** ”

Tom Louis, Organising President,  
Jean-Louis Foulley, Chair IPC,  
John Hinde, Chair LOC,  
Andrew Mead, President-elect,  
David Balding, BIR President

<http://www.cpregistrations.com/ibc/2008/default.asp?page=home>

## Lead up to 1908 article from appreciation by Egon S Pearson, 1939

1899 Hired as a staff scientist by Guinness (Dublin)

1904 “The Application of the ‘Law of Error’ to the work of the Brewery”    Airy: *Theory of Errors*

1905 Met with Karl Pearson: one of three specific problems:  
I find out the P.E. (Probable Error) of a certain laboratory analysis from  $n$  analyses of the same sample. This gives me a value of the P.E. which itself has a P.E. of  $\text{P.E.}/\sqrt{2n}$ . I now have another sample analysed and wish to assign limits within which it is a given probability that the truth must lie. e.g. if  $n$  were infinite, I could say “it is 10 : 1 that the truth lies within 2.6 of the result of the analysis,” *As however  $n$  is finite and in some cases not very large, it is clear that I must enlarge my limits, but I do not know by how much* [italics ours].

'06-'07 At Karl Pearson's Biometric Laboratory in London.

1907 Paper on sampling error involved in **counting** yeast cells.

1908 Papers on P.E. of **mean** and of **correlation coefficient**.

## D.R. Cox on Gosset's statistical papers...

“They have, as has often been remarked, an astonishing freshness and modernity, stemming perhaps from his conciseness and his ability to obtain statistically meaningful results with simple mathematics.”

D.R. Cox. *Biometrika: The First 100 Years*.

*Biometrika*, 2001, **88**, pp 3-11.

## E.S. Pearson on Gosset's 'P.E. of Mean' paper...

“It is a paper to which I think all research students in statistics might well be directed, particularly before they attempt to put together their own first paper.

The actual derivation of the distributions of  $s^2$  and  $z$ , or of  $t = z\sqrt{n-1}$  in to-day's terminology, has long since been made simpler and more precise; this analytical treatment need not be examined carefully, **but there is something in the arrangement and execution of the paper which will always repay study.**”

Pearson, E.S. 'Student' as a Statistician.

*Biometrika*, 1939, **30**, pp 210–250.

# PROBABLE ERROR

## **Webster's Revised Unabridged Dictionary (1913)**

Probable error (of an observation, or of the mean of a number), that within which, taken positively and negatively, there is an even (50%) chance that the real error shall lie. Thus, if 3(sec) is the probable error in a given case, the chances that the real error is greater than 3(sec) are equal to the chances that it is less.

[http://dict.die.net/probable\\_error/](http://dict.die.net/probable_error/)

Earliest Known Uses of Some of the Words of Probability & Statistics

<http://www.leidenuniv.nl/fsw/verduin/stathist/1stword.htm>

## Gosset's introduction to his paper

The “usual method of determining the probability that the mean of the population  $[\mu]$  lies within a given distance of the mean of the sample  $[\bar{x}]$ , is to assume a normal distribution about the mean of the sample with a standard deviation equal to  $s/\sqrt{n}$ , where  $s$  is the standard deviation of the sample, and to use the tables of the [Normal] probability integral,” i.e., to assume

$$\mu \sim N(\bar{x}, s/\sqrt{n}).$$

But, with smaller  $n$ , the value of  $s$  “becomes itself subject to increasing error.”

Sometimes can use external value of  $s$ ; but, more often ...

forced to “judge of the uncertainty of the results from a small sample, which itself affords the only indication of the variability.”

Inferential methods for such small-scale experiments had “hitherto been outside the range of statistical enquiry.”

*Although it is well known that the method of using the normal curve is only trustworthy when the sample is “large,” no one has yet told us very clearly where the limit between “large” and “small” samples is to be drawn. The aim of the present paper is to determine the point at which we may use the tables of the (Normal) probability integral in judging of the significance of the mean of a series of experiments, and to furnish alternative tables for use when the number of experiments is too few.*

## Sampling distributions studied

$$\bar{x} = \frac{\sum x}{n}; \quad s^2 = \frac{\sum (x - \bar{x})^2}{n}.$$

“when you only have quite small numbers, I think the formula with the divisor of  $n - 1$  *we used to use* is better”

... Gosset letter to Dublin colleague, May 1907

Doesn't matter, “because only naughty brewers take  $n$  so small that the difference is not of the order of the probable error!”

... Karl Pearson to Gosset, 1912

$$z = (\bar{x} - \mu)/s$$

# Three steps to the distribution of $z$

## Section I

- Derived first 4 moments of  $s^2$ .
- Found they matched those from curve of Pearson's type III.
- "it is probable that that curve found represents the theoretical distribution of  $s^2$ ." Thus, "although we have no actual proof, we shall assume it to do so in what follows."
- Found the p.d.f. of  $s$  by usual 'change of variable' method; "since the frequency of  $s$  is equal to that of  $s^2$ , all that we must do is compress the base line (axis) suitably"

Hanley2006: emphasize *change of scale* instead of *change of variable*.

## Section II

- "No kind of correlation" between  $\bar{x}$  and  $s$
- His proof is incomplete: he argues that since positive and negative values of  $\bar{x} - \mu$  are equally likely, there cannot be correlation between the absolute value of  $\bar{x} - \mu$  and  $s$ .

## Section III

- Derives the pdf of  $z$ :
  - joint distribution of  $\{\bar{x}, s\}$
  - transforms to that of  $\{z, s\}$ ,
  - integrates over  $s$  to obtain  $pdf(z) \propto (1 + z^2)^{-n/2}$ .

## Sections IV and V

- Moments, shapes and tail areas of  $s$  and  $z$  distributions
- Reason why he only tabled his curve for values of  $n \leq 10$

## Section VI: “Practical test of foregoing equations.”

[ pdf's of  $s$  and  $z$  “are compared with some actual distributions” ]

Before I had succeeded in solving my problem analytically, I had endeavoured to do so empirically.

The material used was a correlation table containing the height and left middle finger measurements of 3000 criminals, from a paper by **W. R. Macdonell** (*Biometrika*, Vol. I, p. 219).

The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random.

As each card was drawn its numbers were written down in a book, which thus contains the measurements of 3000 criminals in a random order.

## continued ...

Finally, each consecutive set of 4 was taken as a sample – 750 in all – and the mean, standard deviation, and correlation of each sample determined.

The difference between the mean of each sample and the mean of the population was then divided by the standard deviation of the sample, giving us the  $z$  of Section III.

This provides us with two sets of 750 standard deviations and two sets of 750  $z$ 's on which to test the theoretical results arrived at.

## Macdonell's data

Obtained from Central Metric Office, New Scotland Yard, reported as a  $42(1 \text{ mm bins}) \times 22(1'' \text{ bins})$  frequency table

Heights measured to nearest  $1/8$  of  $1''$  by Scotland Yard staff

Mean  $\pm$  P.E. reported as  $65.5355'' \pm 0.0313''$

- $0.0313 = 0.6745 \times sd / \sqrt{3000}$ ;
- $sd = 2.5410 \pm 0.0221$
- $0.0221 = 0.6745 \times sd / \sqrt{2n}$ .

## Our simulations, 100 years later

- Reproduced means and sd's reported by Macdonell.
- Repeated Gosset's procedure to create 750 samples.
- Occasionally, all 4 persons from same 1" bin  $\rightarrow s = 0$ .  
Replaced  $z = \pm\infty$  by  $\pm$  largest observed  $|z|$ .
- $X^2$  goodness of fit statistic for 750  $s/\sigma$ , and 750  $z$  values.
- Repeated procedure 100 times:- 100  $X^2$  values :-  
check repeatability of Gosset's  $X^2$  statistics; cards sufficiently shuffled?
- Single set of 75,000 samples of size 4, *sampled with replacement*, and with *Scotland Yard precision* (1/8 of 1").  
How much more smooth/accurate might Gosset's empirical frequency distribution of  $s$  have been?

## RESULTS: his and ours

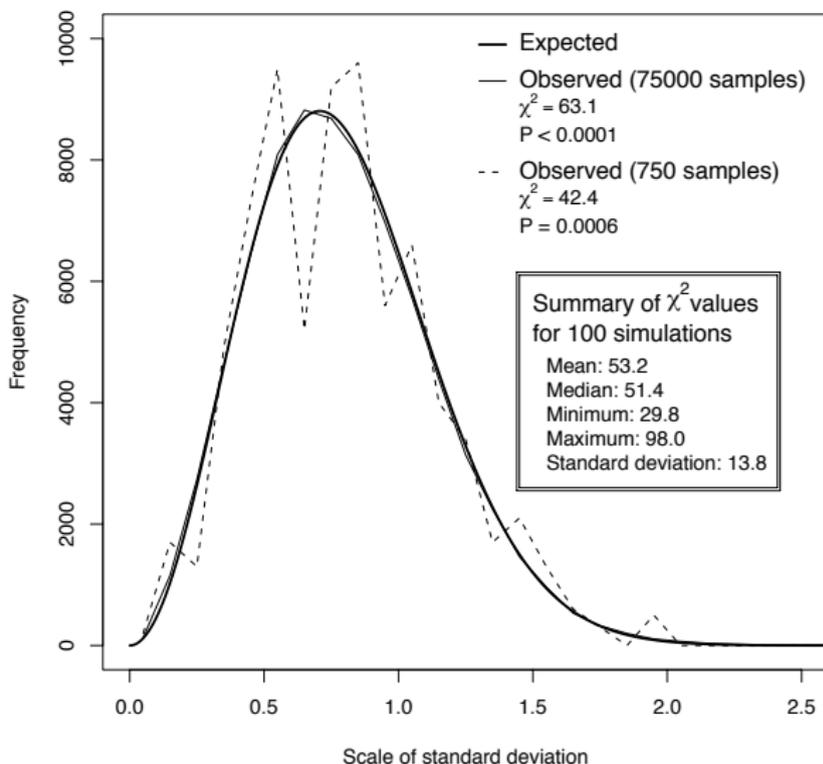
### Shuffling:

	No. samples/750 with $s = 0$							
	0	1	2	3	4	5		All
Ours:	21	41	17	16	4	1		100
Gosset's:				1				1

Gosset's double precautions – very thorough shuffling *and* drawing cards at random – appear to have worked.

Unlike the 1970 U.S. draft lottery for military service in Vietnam

# Distribution of $s/\sigma$

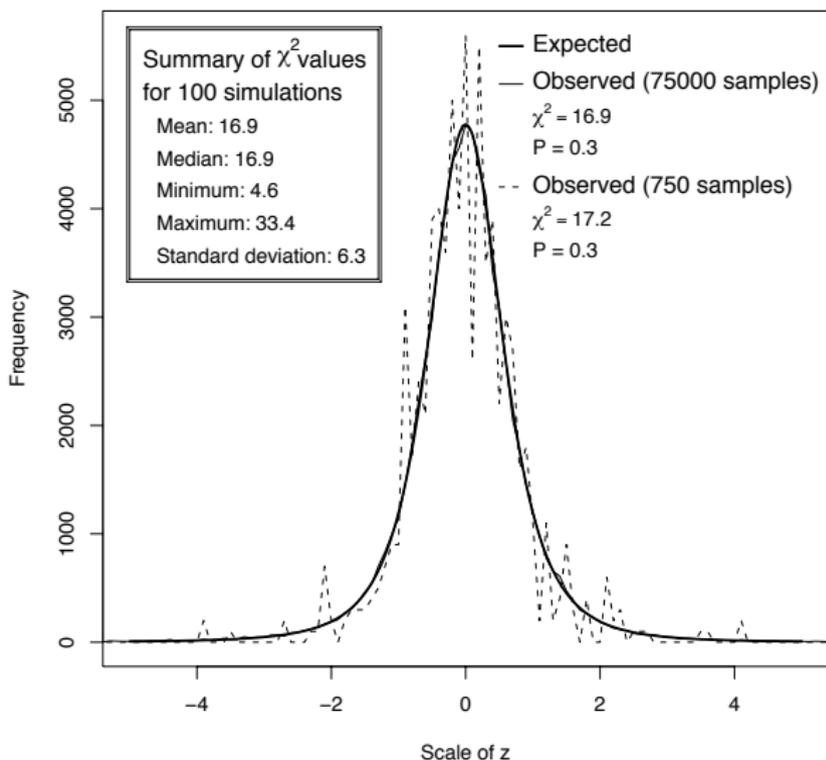


Dotted line: Sample statistics obtained from one set of 750 random samples generated by Gosset's procedure.

Inset: distr'n of 100  $\chi^2$  statistics (18 intervals).

Thin solid line: distr'n of statistics obtained from 75,000 samples of size 4 sampled with replacement from 3000 heights recorded to nearest 1/8".

# Distribution of z



Dotted line: Sample statistics obtained from one set of 750 random samples generated by Gosset's procedure.

Inset: distribution of 100  $\chi^2$  statistics (15 intervals).

Thin solid line: distr'n of statistics obtained from 75,000 samples of size 4 sampled with replacement from 3000 heights recorded to nearest 1/8".

## If Gosset had $R$ :

“Agreement between observed and expected frequencies of the 750  $s/\sigma$ 's was not good”. He attributed this to coarse scale of  $s$ .

His  $X^2$  statistic, summed across 18 bins, was 48.1 – just below median (51) in our series, in which values ranged from 30 to 98.

Distribution of our 75,000  $s/\sigma$  values also shows pattern of large deviations similar to those in table on p. 15 of his paper.

Scotland Yard precision and today's computing power would have left Gosset in no doubt that the distribution of  $s$  which he “assumed” was correct was in fact correct.

Grouping had not had so much effect on distr'n of  $z$ 's: “close correspondence between the theory and the actual result.”

## Remaining sections; and first dozen years

VII Tabulated the  $z$  distribution for  $n \leq 10$ .

VIII Explained its use.

IX 3 fully worked e.g.'s, all of "paired- $t$ " type,  $n = 10, 6, 2$ .  
 4. e.g.  $n = 11$  experiments, with  $\bar{d} = 33.7$  and  $s = 63.1$ .  
 Uses approx'n

$$\Delta \sim N(\bar{d}, s/\sqrt{n-3})$$

0.934 probability "that kiln-dried barley seed gives a higher barley yield than non-kiln-dried seed."

Approx'n remarkably accurate: extended  $z$  table of 1917, and  $p_t$  function in  $\mathbb{R}$ , both yield exact probability of 0.939.

>1908 "the  $z$ -test was used in brewery at once, but I think very little elsewhere for probably a dozen years." (E.S. Pearson)

1922 Gosset sent Fisher a copy of his new tables, "you are the only man that's ever likely to use them!"

1912 'Extra-mural' use by epidemiologist Janet Lane-Claypon.

## Fisher and Gosset

**1912 Fisher (age 22) to Gosset:** rigorous & elegant derivation of z distribution, one that ultimately led Fisher to realize the far wider applicability of variants of Student's z.

### Gosset to Karl Pearson

*Would you mind looking at it for me; I don't feel at home in more than three dimensions even if I could understand it otherwise (...) It seemed to me that if it's all right perhaps you might like to put the proof in a note. It's so nice and mathematical that it might appeal to some people.*

### Pearson to Gosset

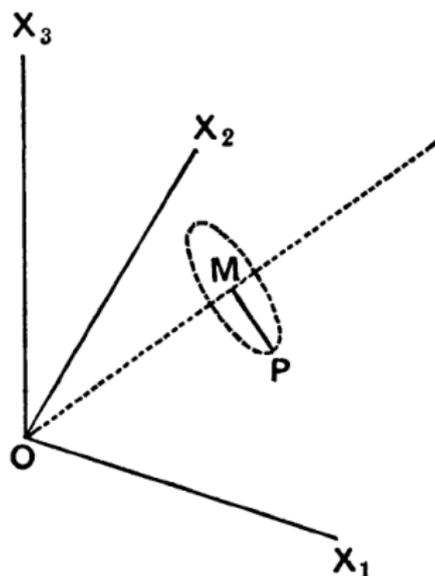
*I do not follow Mr. Fisher's proof and it is not the kind of proof which appeals to me.*

**1915** Proof published in Fisher's 1915 corr'l'n paper (Biometrika).

## Fisher's geometric vision

- “the form establishes itself instantly, when the distribution of the sample is viewed geometrically.”
- “exceedingly beautiful interpretation in generalised space” used to derive the distribution of  $s$
- “expanded version”

[www.epi.mcgill.ca/hanley/Student](http://www.epi.mcgill.ca/hanley/Student)



Fisher's derivation “shows that for normal distributions there is no correlation between deviations in the mean and in the standard deviation of samples, a familiar fact.” – Biometrika Editorial

## End of the chapter on $z$

- “the paper by Mr Fisher and (..) more or less complete the work on the distribution of standard-deviations outlined by “Student” in 1908.” Biometrika Editorial, 1915
- The 1917 publication of Gosset’s extended tables – from  $n = 2$  to  $n = 30$  – of his  $z$  distribution did indeed end the chapter on  $z$ .
- But a new more extensive and much more important one on  $t$  – that took until 1925 to reach publication – was being opened by Gosset and Fisher.

## From $z$ to $t$

- Fisher-Gosset collaboration:  $z \rightarrow t = z\sqrt{n-1}$ .
- Fisher: uses for 2-sample problems and regr'n coeffs.
- Student published his  $t$  tables in *Metron* in 1925
- Fisher put own version in *Stat'l Methods for Research Workers*.
- Joan Fisher: describes collaboration & dealings with K Pearson.
- **Gosset**: had to fight for access to **portable** hand operated Baby Triumphator calculator in heavy day use at brewery
- **Fisher's** motor Millionaire at Rothamsted: large electrically powered **desktop** machine.
- Eisenhart: evidence "seems to indicate that decision to shift from  $z$ - to  $t$ - **form** originated with **Fisher**, but choice of **letter "t"** to denote new form was due to "**Student.**" "
- Boland; Lehmann: add'nl accounts of G's life & work.

## Triumphator A ser 43219



<http://www.calculators.szrek.com/>

## Millionaire Ser 1200



## To students of statistics in 2008 ... (I)

Fisher1939:

- “of (Gosset’s) personal characteristics, the most obvious were a clear head, and a practice of forming independent judgements.”
- The other was the importance of his work environment: “one immense advantage that Gosset possessed was the concern with, and responsibility for, the practical interpretation of experimental data.”

Gosset stayed very close to these data.

We should too!

## To students of statistics in 2008 ... (II)

Compared with what Gosset could do, today we can run much more extensive simulations to test our new methods.

Which pseudo-random observations are more appropriate: those from perfectly behaved theoretical populations, or those from real datasets, such as Macdonell's?

In light of how Gosset included the 3 infinite z-ratios, we might re-examine how we deal with problematic results in our runs.

## To students of statistics in 2008 ... (III)

The quality of writing – and statistical writing – is declining.

Today's students – and their teachers – would do well to heed E.S. Pearson's 1939 advice regarding writing and communication.

I encourage you to read the primary work and other writings of authors such as Galton, Karl Pearson, Gosset, Fisher, E.S. Pearson, Cochran, Mosteller, David Cox, Stigler, and others, not only for interesting statistical content, but also for style.

## To students of statistics in 2008 ... (IV)

When JH was a student, very little of the historical material we have reviewed here was readily available.

Today, we are able to obtain it, review it, and follow up leads – all from our desktops – via Google, and using JSTOR and other online collections.

Statistical history need no longer be just for those who grew up in the years “B.C.”

Become Students of the History of Statistics

---

“B.C.”: Before Computers.

# FUNDING / CO-ORDINATES

Natural Sciences and Engineering Research Council of Canada

`James.Hanley@McGill.CA`

`http://www.epi.mcgill.ca/hanley`



**McGill**

**Biostatistics  
Biostatistique**

---

<http://www.mcgill.ca/epi-biostat-occh/grad/biostatistics/>