A statistician's choice of the pen name 'Student'—and others' misspellings of his surname

This note shares 20th-century and 21st-century instances where 'Student's' real surname was misspelled by his immediate boss, his employer, *Significance* magazine, a state-run genealogy website, and—most recently—the National Archives of Ireland. It also provides a photo of the commemorative plaque whose wording drew some critical scrutiny, along with pictorial evidence that led one author to suggest why Gosset chose the pen name 'Student.'

Contents

1 1901 Census of Ireland 3

2 1907: first use of his pen name, and why *this particular* pen name? 5

3 Census of Ireland 1911; IBS-ISA Conference 2008 6

4 The birth records of his children 12

5 2013 Significance Magazine's *Timeline of Statistics* 14

6 1926 Census of the Irish Free State 15

7 BTW: Which of these titles would you vote for? 18

Students in statistics classes who wonder why there is a probability distribution named after **them** can find this answer in Wikipedia.

History [edit]

In statistics, the *t* distribution was first derived as a posterior distribution in 1876 by Helmert[18][19][20] and Lüroth.[21][22][23] As such, Student's t-distribution is an example of Stigler's Law of Eponymy. The *t* distribution also appeared in a more general form as Pearson type IV distribution in Karl Pearson's 1895 paper.[24]

In the English-language literature, the distribution takes its name from William Sealy Gosset's 1908 paper in *Biometrika* under the pseudonym "Student" during his work at the Guinness Brewery in Dublin, Ireland.[25] One version of the origin of the pseudonym is that Gosset's employer preferred staff to use pen names when publishing scientific papers instead of their real name, so he used the name "Student" to hide his identity. Another version is that Guinness did not want their competitors to know that they were using the *t* test to determine the quality of raw material.[26][27]



Statistician William Sealy Gosset, known as "Student"

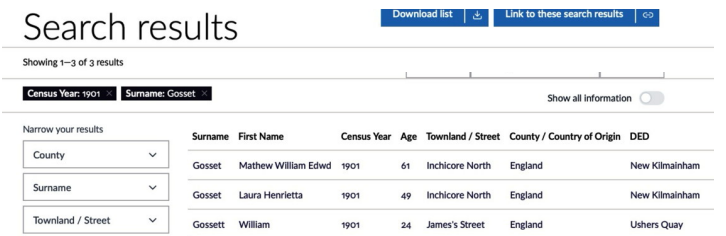
Gosset worked at Guinness and was interested in the problems of small samples – for example, the chemical properties of barley where sample sizes might be as few as 3. Gosset's paper refers to the distribution as the "frequency distribution of standard deviations of samples drawn from a normal population". It became well known through the work of Ronald Fisher, who called the distribution "Student's distribution" and represented the test value with the letter *t*.[7][28]

This note shares 20th century and 21st century instances where Student's real surname was misspelled by his immediate boss, his employer, Significance Magazine, a state-run genealogy website and—most recently—the National Archives of Ireland. It also provides a photo of the commemorative plaque whose wording drew some critical scrutiny, along with pictorial evidence that led one author to suggest why he chose the pen-name '*Student*.'

With a name that must have been quite rare in Ireland, we should be able to find him easily in the now-online databases containing the household returns in the 1901 and 1911 Censuses of Ireland, and the 1926 Census of the Irish Free State.

1 1901 Census of Ireland

Searching just on ‘Gosset’ and using the (default) approximate match option, we find just three entries, all English-born, two named Gosse**t** and one Gosse**tt**.



Search results

Download list | Link to these search results

Showing 1–3 of 3 results

Census Year: 1901 × Surname: Gosset ×

Show all information

Narrow your results

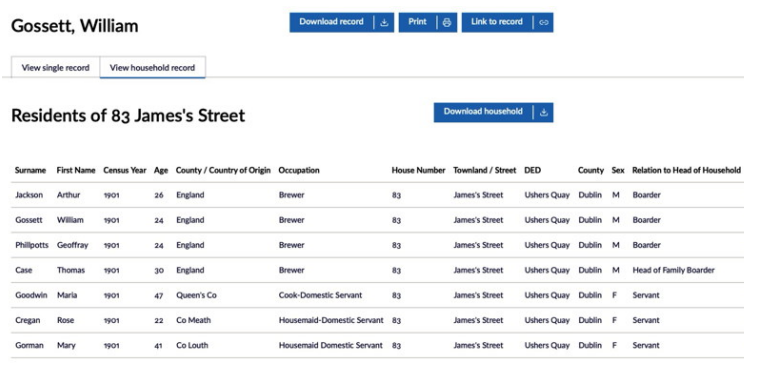
County

Surname

Townland / Street

Surname	First Name	Census Year	Age	Townland / Street	County / Country of Origin	DED
Gosset	Mathew William Edwd	1901	61	Inchicore North	England	New Kilmainham
Gosset	Laura Henrietta	1901	49	Inchicore North	England	New Kilmainham
Gossett	William	1901	24	James's Street	England	Ushers Quay

The one male named Gosse**t** turned out to be a military man. Although the other man’s surname is Gosse**tt**, the James’s Street address is promising. And indeed we find him—



Gossett, William

Download record | Print | Link to record

View single record | View household record

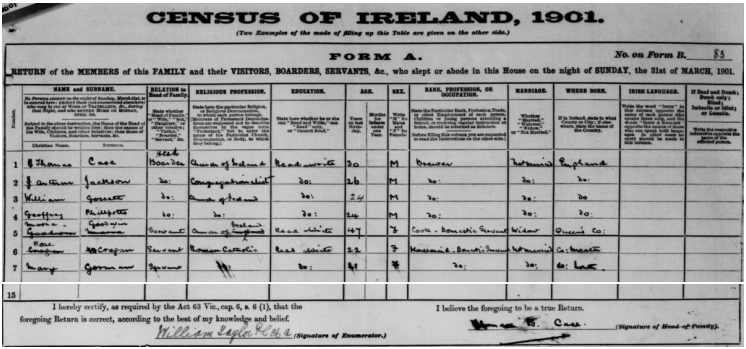
Residents of 83 James's Street

Download household

Surname	First Name	Census Year	Age	County / Country of Origin	Occupation	House Number	Townland / Street	DED	County	Sex	Relation to Head of Household
Jackson	Arthur	1901	26	England	Brewer	83	James's Street	Ushers Quay	Dublin	M	Boarder
Gossett	William	1901	24	England	Brewer	83	James's Street	Ushers Quay	Dublin	M	Boarder
Phillpotts	Geoffray	1901	24	England	Brewer	83	James's Street	Ushers Quay	Dublin	M	Boarder
Case	Thomas	1901	30	England	Brewer	83	James's Street	Ushers Quay	Dublin	M	Head of Family Boarder
Goodwin	Maria	1901	47	Queen's Co	Cook-Domestic Servant	83	James's Street	Ushers Quay	Dublin	F	Servant
Cregan	Rose	1901	22	Co Meath	Housemaid-Domestic Servant	83	James's Street	Ushers Quay	Dublin	F	Servant
Gorman	Mary	1901	41	Co Louth	Housemaid Domestic Servant	83	James's Street	Ushers Quay	Dublin	F	Servant

living at the Guinness house for unmarried brewers at St. James’ Gate. Living with him were three others: ‘Head of Family Boarder’ Thomas Case, the first university science graduate hired by Guinness in 1893; Geofffrey Phillpotts, who would come Gosset’s brother-in-law five years later; and a fourth named Jackson.[1].

Even his house-mate and immediate boss didn't get his name right, spelling it as Gosse**tt** rather than Gosse**t**.



CENSUS OF IRELAND, 1901.

(Two Examples of the mode of filling up this Table are given on the other side.)

FORM A. No. on Form B. 55

RETURN of the MEMBERS of this FAMILY and their VISITORS, BOARDERS, SERVANTS, &c., who slept or abode in this House on the night of SUNDAY, the 31st of MARCH, 1901.

Number	NAME and SURNAME. Christian Name.	Surname.	RELATION to Head of Family.	RELIGIOUS PROFESSION.	EDUCATION.	AGE.	SEX.	RANK, PROFESSION, OR OCCUPATION.	MARRIAGE.	WHERE BORN.	IRISH LANGUAGE.	If Deaf and Dumb; Deaf only; Blind; Imbecile or Idiot; or Lunatic.
1	Thomas	Case	Head	Church of Ireland	Read, write	30	M	Brewer	Unmarried	England		
2	Jonathan	Jackson	Do:	Congregationalist	Do:	28	M	Do:	Do:	Do:		
3	William	Gosset	Do:	Church of Ireland	Do:	21	M	Do:	Do:	Do		
4	Geoffrey	Pilkington	Do:	Do:	Do:	24	M	Do:	Do:	Do.		
5	Bridget [illegible]	Goodwin	Servant	Church of Ireland	Read, write	47	F	Cook - Domestic Servant	Widow	County Co:		
6	Mary Coogan	McCrogan	Servant	Roman Catholic	Read, write	23	F	Housemaid - Domestic Servant	Unmarried	Co: Meath		
7	Mary	Gorman	Servant	Do:	Do:	41	F	Do:	Do:	Do: Louth.		
15												

I hereby certify, as required by the Act 63 Vic., cap. 6, s. 6 (1), that the foregoing Return is correct, according to the best of my knowledge and belief.
William Taylor, R.M. (Signature of Enumerator.)

I believe the foregoing to be a true Return.
Thomas Case. (Signature of Head of Family).

This is understandable, as the double-t ending might look more 'English' to many people. It is only when we consult Burke's *The Landed Gentry* that we learn why it has just one t:

> **The Gossets**
> • Of "Norman Extraction"
> • Coat of arms: D'àgur, à un annulet d'or, et trois Goussés de fèves feuillées et tigées, et rangées, en pairle de même; au chef d'argent, chargé d'une aiglette de sable.
> • 1555: Adopted Protestant faith → name removed from roll of nobles.
> • 1685: Revocation of Edict of Nantes → **Jean Gosse<u>t</u>, a Huguenot**, moved to Island of Jersey.
> • Some of Jean Gosset's family settled in England.

[Another way I remember if it's one t or two t's is to recall that his 1908 paper only dealt with the <u>one</u>-sample problem.[2]]

2 1907: first use of his pen name, and why *this particular* pen name?

This is his first use of the pen name—in this first (but less well known) extra-mural publication of his, in pages 351-360 of Vol. 5 (No. 3) of Biometrika in 1907.

ON THE ERROR OF COUNTING WITH A HAEMACYTOMETER.

By STUDENT.

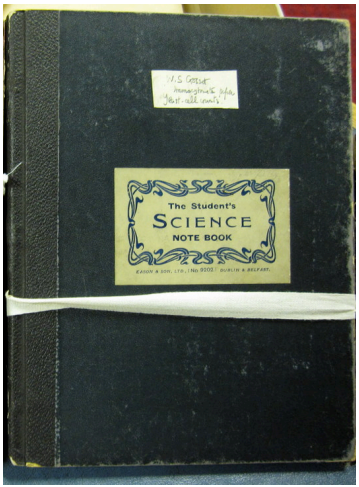
WHEN counting yeast cells or blood corpuscles with a hæmacytometer there are two main sources of error: (1) the drop taken may not be representative of the bulk of the liquid; (2) the distribution of the cells or corpuscles over the area which is examined is never absolutely uniform, so that there is an "error of random sampling."

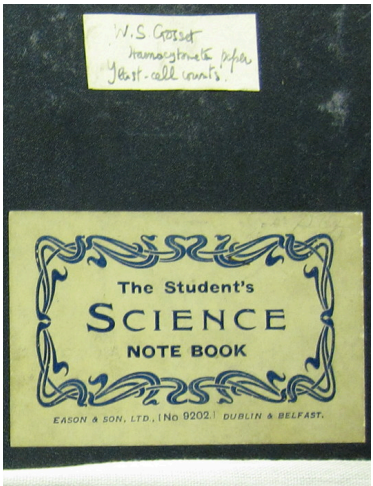
With the first source of error we are concerned only to this extent; that when the probable error of random sampling is known we can tell whether the various drops taken show significant differences. What follows is concerned with the distribution of particles throughout a liquid, as shewn by spreading it in a thin layer over a measured surface and counting the particles per unit area.

As to the *origin* of that pen name, Stephen Ziliak offers *this* possibility

> Why he chose to be known as "Student" is not known for sure but here is a clue: besides being a humble man, there is stamped on the cover of his early notebook a manufacturer's imprint, "The Student's Science Notebook, Eason and Son, Ltd., Dublin and Belfast."

Here is a 2009 photograph I took of the cover of that notebook, which he used for the 1907 paper.





Ziliak adds that [the Guinness historians] Dennison & MacDonagh (1998, p. 90, footnote 9) suggest the "Student" name originates with the managing director La Touche, but the Guinness historians show no evidence for it, and apparently they were not aware of "Student's" eponymous notebook, located in London."

Student was officially 'unmasked' in print in the Journal of the American Statistical Association in 1930[3] and in the 1938 preface[4] to the posthumous publication of his last paper.

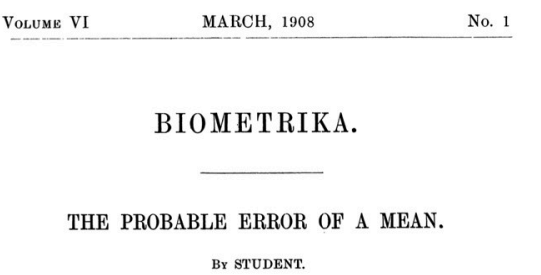
3 Census of Ireland 1911; IBS-ISA Conferen 2008

A search of the 1911 census just on 'Gosset' identifies 11 persons in four households, all with the same Gosse**t** spelling, six of

them in this suburban household.

Surname	First Name	Census Year	Age	Townland / Street	County / Country of Origin	DED
Gosset	William Sealy	1911	34	Woodpark, Part of	England	Stillorgan
Gosset	Marjory Surtees	1911	31	Woodpark, Part of	England	Stillorgan
Gosset	Isaac Henry	1911	4	Woodpark, Part of	England	Stillorgan
Gosset	Marion Bertha	1911	2	Woodpark, Part of	Co Dublin	Stillorgan
Gosset	Ruth Helen	1911	0	Woodpark, Part of	Co Dublin	Stillorgan
Gosset	Agnes Sealy	1911	59	Woodpark, Part of	England	Stillorgan

To mark the 100th anniversary of his groundbreaking paper,



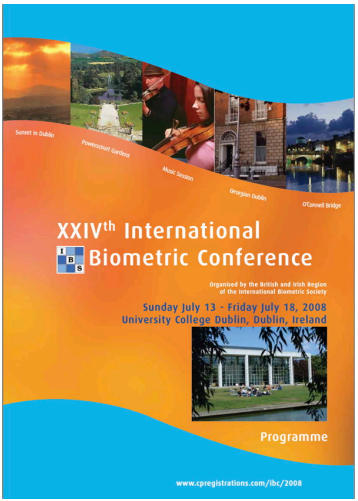
VOLUME VI MARCH, 1908 No. 1

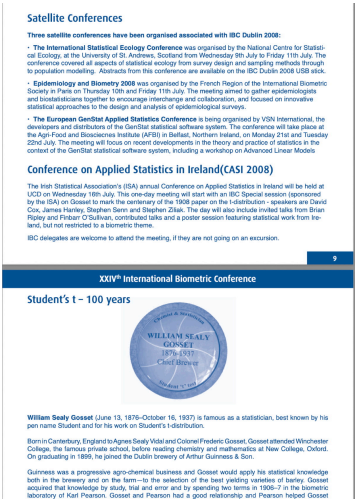
BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

the biennial conference of the International Biometric Society (IBS) was held in Dublin in 2008, in conjunction with the annual conference of the Irish Statistical Association.





Satellite Conferences

Three satellite conferences have been organised associated with IBC Dublin 2008:

- **The International Statistical Ecology Conference** was organised by the National Centre for Statistical Ecology, at the University of St. Andrews, Scotland from Wednesday 9th July to Friday 11th July. The conference covered all aspects of statistical ecology from survey design and sampling methods through to population modelling. Abstracts from this conference are available on the IBC Dublin 2008 USB stick.
- **Epidemiology and Biometry 2008** was organised by the French Region of the International Biometric Society in Paris on Thursday 10th and Friday 11th July. The meeting aimed to gather epidemiologists and biostatisticians together to encourage interchange and collaboration, and focused on innovative statistical approaches to the design and analysis of epidemiological surveys.
- **The European GenStat Applied Statistics Conference** is being organised by VSN International, the developers and distributors of the GenStat statistical software system. The conference will take place at the Agri-Food and Biosciences Institute (AFBI) in Belfast, Northern Ireland, on Monday 21st and Tuesday 22nd July. The meeting will focus on recent developments in the theory and practice of statistics in the context of the GenStat statistical software system, including a workshop on Advanced Linear Models

Conference on Applied Statistics in Ireland(CASI 2008)

The Irish Statistical Association's (ISA) annual Conference on Applied Statistics in Ireland will be held at UCD on Wednesday 16th July. This one-day meeting will start with an IBC Special session (sponsored by the ISA) on Gosset to mark the centenary of the 1908 paper on the t-distribution - speakers are David Cox, James Hanley, Stephen Senn and Stephen Ziliak. The day will also include invited talks from Brian Ripley and Finbarr O'Sullivan, contributed talks and a poster session featuring statistical work from Ireland, but not restricted to a biometric theme.

IBC delegates are welcome to attend the meeting, if they are not going on an excursion.

Student's t – 100 years

William Sealy Gosset (June 13, 1876–October 16, 1937) is famous as a statistician, best known by his pen name Student and for his work on Student's t-distribution.

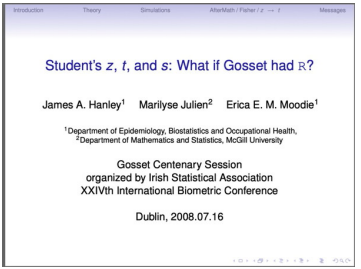
Born in Canterbury, England to Agnes Sealy Vidal and Colonel Frederic Gosset, Gosset attended Winchester College, the famous private school, before reading chemistry and mathematics at New College, Oxford. On graduating in 1899, he joined the Dublin brewery of Arthur Guinness & Son.

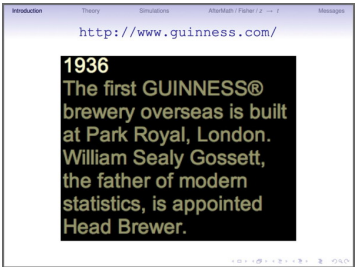
Guinness was a progressive agro-chemical business and Gosset would apply his statistical knowledge both in the brewery and on the farm—to the selection of the best yielding varieties of barley. Gosset acquired that knowledge by study, trial and error and by spending two terms in 1906–7 in the biometric laboratory of Karl Pearson. Gosset and Pearson had a good relationship and Pearson helped Gosset

In the Special Session,

08:45 Theatre P	**Session 10c: Special Session: Gosset Centenary** Chair: John Hinde
	Session Organisers: Irish Statistical Association (John Hinde)
08:45	**Student's z, t, and s: what if Gosset had R?** James Hanley, McGill University, Canada
09:15	**The Trial at Kalamazoo** Stephen Senn, University of Glasgow, United Kingdom
09:45	**GUINNESSOMETRICS: WILLIAM SEALY GOSSET AND THE ECONOMIC FOUNDATION OF 'STUDENT'S' T** Steve Ziliak, Roosevelt University, United States
10:15	**Student and Statistical Inference** David Cox, Nuffield College, United Kingdom

I showed this slide from the Guinness website on 19 July 2008, **claiming him** as **the** father of modern statistics, **but**—like his boss a century earlier—**spelling his surname wrong**! We pointed this out to the Guinness people we met at the event at the Storehouse event (see below).





They did get the spelling fixed. Here is what the website had about him for a while.

December 2008



2017

GUINNESS STOREHOUSE

Archive Fact Sheet: Guinness History Timeline

1759 Arthur Guinness signed a 9,000 year lease on a disused brewery at St. James's Gate, Dublin for an annual rent of £45.

1769 The first export shipment of six and a half barrels of GUINNESS® left Dublin on a sailing vessel bound for England.

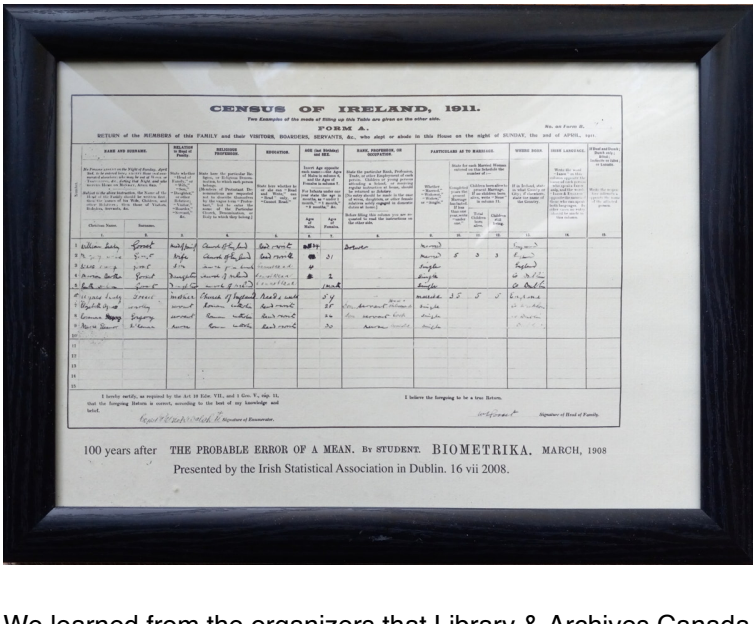
1929 First ever advertisement for GUINNESS® with the slogan "Guinness is Good For You" published in the British national press. This was soon followed by advertisements featuring the cartoon characters created by John Gilroy. His famous series of posters of the distraught zoo keeper and his mischievous animals carried the line 'My Goodness, My Guinness'.

1936 First overseas brewery built at Park Royal, London. William Sealy Gosset, the father of modern statistics, appointed Head Brewer.

1955 First Guinness Book of Records was published.

But, in 2026, when I search The Story of Guinness on their current website, I can't find Gosset's name there (maybe I could if I signed up for a tour).

As an appreciation, each of the four speakers at that 2008 IBC session was given a plaque. It included a photo of the household return from the Gosset household in that 1911 census of Ireland—the last, as it would transpire, to be held under British rule.



CENSUS OF IRELAND, 1911.

Two Examples of the mode of filling up this Table are given on the other side.

No. on Form H.

FORM A.

RETURN of the NUMBERS of this FAMILY and their VISITORS, BOARDERS, SERVANTS, &c., who slept or abode in this House on the night of SUNDAY, the 2nd of APRIL, 1911.

	NAME AND SURNAME.	RELATION to Head of Family.	RELIGIOUS PROFESSION.	EDUCATION.	AGE last Birthday and SEX	RANK, PROFESSION, OR OCCUPATION.	PARTICULARS AS TO MARRIAGE.	WHERE BORN	IRISH LANGUAGE.	If Deaf and Dumb, Blind, Lunatic, Imbecile or Idiot
	Christian Name.	Surname.			Age last Birthday. Males. Females.					
1	William Lesly	Gosset	Church of Ireland	Can read & write	38	Solicitor	Married	England		
2	[illegible]	[illegible]	Church of Ireland	Can read & write	31		Married	England		
3	[illegible]	[illegible]	Church of Ireland	Can read & write	4		Single	England		
4	[illegible]	[illegible]	Church of Ireland	Can read & write	2		Single	England		
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		Single	[illegible]		
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		Married	[illegible]		
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		Single	[illegible]		
8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		Single	[illegible]		
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		Single	[illegible]		
10										
11										
12										
13										
14										
15										

I hereby certify, as required by the Act 10 Edw. VII., and 1 Geo. V., cap. 15, that the foregoing Return is correct, according to the best of my knowledge and belief.

[illegible] Signature of Enumerator.

I believe the foregoing to be a true Return.

[illegible] Signature of Head of Family.

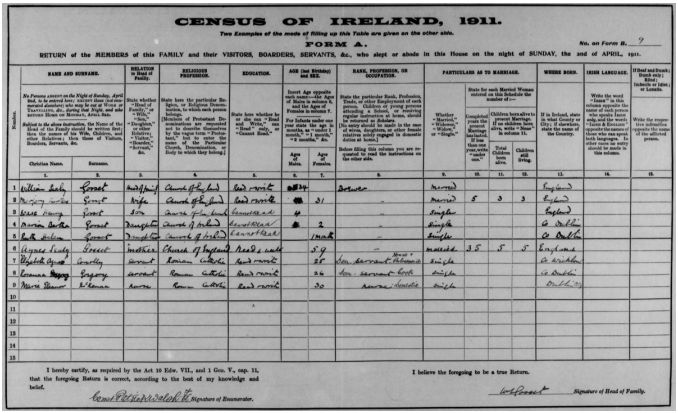
100 years after THE PROBABLE ERROR OF A MEAN. By STUDENT. BIOMETRIKA. MARCH, 1908

Presented by the Irish Statistical Association in Dublin. 16 vii 2008.

We learned from the organizers that Library & Archives Canada, under a contract with the National Archives of Ireland, had already digitised and put online the 1911 census returns for Dublin, and would have all of the counties completed soon. As I describe elsewhere, the contract was to “digitise, index and publish online both the 1901 and 1911 paper census returns”; the data releases took place between 2007 and 2010. And they sparked my interest in census data, not just for family reasons, but also for teaching and for documenting local and national history.

Online at the National Archive, in addition to the transcribed data,

you can find this slightly more readable version of the image from which Gosset's data were transcribed.



CENSUS OF IRELAND, 1911.

Two Examples of the mode of filling up this Table are given on the other side.

FORM A.

No. on Form B. 9

RETURN of the MEMBERS of this FAMILY and their VISITORS, BOARDERS, SERVANTS, &c., who slept or abode in this House on the night of SUNDAY, the 2nd of APRIL, 1911.

Houses.	NAME and SURNAME. Christian Name.	Surname.	RELATION to Head of Family.	RELIGIOUS PROFESSION.	EDUCATION.	AGE (last Birthday) and SEX. Age at 1 Male.	Age at 1 Female.	RANK, PROFESSION, or OCCUPATION.	PARTICULARS as TO MARRIAGE. Whether "Married," "Widower," "Widow," or "Single."	Completed years the present Marriage has continued.	Children born alive to present Marriage. Sons only.	Children born alive to present Marriage. Daughters only.	Total Children born alive.	Children still living.	WHERE BORN. If in Ireland, state the Townland, Street, &c., and the County. If elsewhere, state the name of the Country.	IRISH LANGUAGE. Write the word "Irish" in this column opposite the name of every person who speaks Irish only, and the words "Irish and English" opposite the names of those who can speak both languages. No other entry should be made in this column.	If Deaf and Dumb; Dumb only; Blind; Imbecile or Idiot; or Lunatic. Write the corresponding letter opposite the name of the afflicted person.
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.		14.	15.	
1	William Sealy	Gosset	Head of Family	Church of England	Read & write	44		Brewer	Married						England		
2	Marjory Sealy	Gosset	Wife	Church of England	Read & write		31		Married	5	3	3			England		
3	Isobel Emily	Gosset	Dau.	Church of England	Read & write	4			Single						England		
4	Dorothy Sealy	Gosset	Daughter	Church of Ireland	Read & write	2			Single						Co. Dublin		
5	Ruth Sealy	Gosset	Daughter	Church of Ireland	Read & write	1 month			Single						Co. Dublin		
6	Agnes Sealy	Grosset	Mother	Church of England	Read & write	69		— Head of Family	Married	35	5	5	5		England		
7	Elizabeth Byrne	Cawley	Servant	Roman Catholic	Read & write	25		Gen. Servant (Domestic)	Single						Co. Wicklow		
8	Emma Mary	Gregory	Servant	Roman Catholic	Read & write	24		Gen. Servant (Cook)	Single						Co. Dublin		
9	Mary Ellen	O'Connor	Servant	Roman Catholic	Read & write	20		Nurse Domestic	Single						Dublin City		
10																	
11																	
12																	
13																	
14																	
15																	

I hereby certify, as required by the Act 10 Edw. VII., and 1 Geo. V., cap. 11, that the foregoing Return is correct, according to the best of my knowledge and belief.

Ernest Pellissier Golah R.M. Signature of Enumerator.

I believe the foregoing to be a true Return.

W.S. Gosset Signature of Head of Family.

The Gosset Commemorative Plaque









Some statisticians at that 2008 event were surprised to see the uppercase **'T'** in the sign displayed at the entrance to the reception.

The inscription on the plaque itself led to further discussion: should the single quotes surround the statistic *t* or the name *Student*?

[The last three photographs in the panel were taken by Peter Macdonald. All of his photographs of this event, and of the IBC2008 itself, can be found here. The first one, and a few more of mine, can be found at the bottom of this page.]

Peter's notes on the photo at the bottom right read: "P1070403: The commemorative plaque. Note the errors: Gosset was Master Brewer (Guinness never had a Chief Brewer), and the quotes should have been around the possessive pseudonym "Student's" rather than the t."

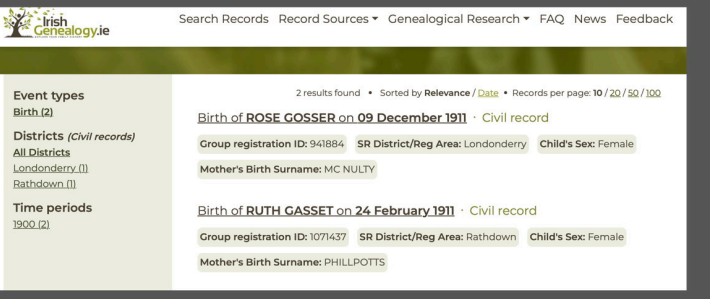
The two 'unveilers' of the plaque are Dermot Roaf and Jane (Roaf) Galbraith, grandchildren of Gosset.

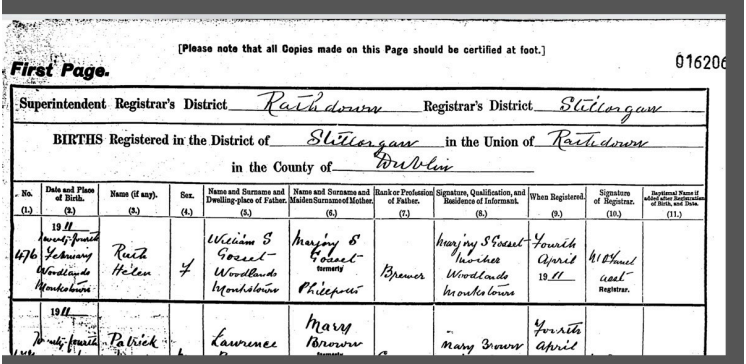
4 The birth records of his children

Dermot told us he was pretty sure he was only one at the reception who had met Gosset in person. He was just a baby when his mother brought him into the hospital so that his grandfather, who had suffered a heart attack, could meet him. His mother was Ruth Helen Gosset, born just a little over a month before the 1911 Census. Thanks to the default search setting—which does not insist on 'exact name matches only'—at this https://www.irishgenealogy.ie website, one can find the handwritten, as well as this **transcribed**, record of her birth registration.

The 1911 return indicated that his oldest child, a son, was born in England. The reader can check the transcription of the birth registration of his older daughter, Marion (?Marian) Bertha, also born in Woodlands, Monkstown, Dublin.







[Please note that all Copies made on this Page should be certified at foot.]

First Page.

016206

Superintendent Registrar's District Rathdown Registrar's District Stillorgan

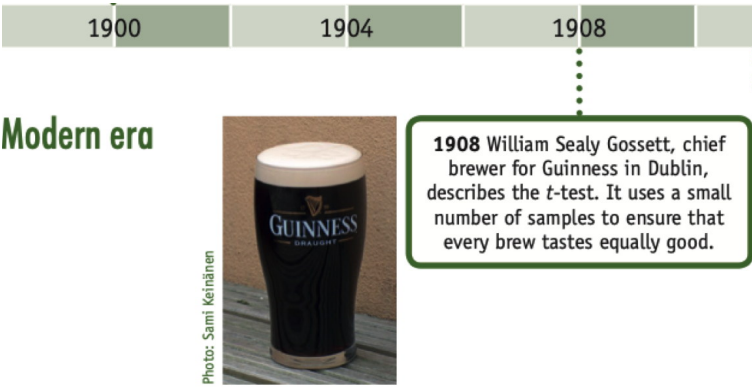
BIRTHS Registered in the District of Stillorgan in the Union of Rathdown

in the County of Dublin

No. (1.)	Date and Place of Birth. (2.)	Name (if any). (3.)	Sex. (4.)	Name and Surname and Dwelling-place of Father. (5.)	Name and Surname and Maiden Surname of Mother. (6.)	Rank or Profession of Father. (7.)	Signature, Qualification, and Residence of Informant. (8.)	When Registered. (9.)	Signature of Registrar. (10.)	Baptismal Name if added after Registration of Birth, and Date. (11.)
1476	1911 Twenty-fourth February Woodlands Monkstown	Ruth Helen	F	William S Gosset Woodlands Monkstown	Margery S Gosset formerly Chicester	Brewer	Margery S Gosset Mother Woodlands Monkstown	Fourth April 1911	H O'Neill Asst Registrar.	
	1911 Twenty-fourth	Patrick		Laurence	Mary Brown formerly		Mary Brown	Fourth April		

5 2013 Significance Magazine's *Timeline of Statistics*

Significance Magazine marked its 10th birthday with "a fold-out timeline of more or less everything that is important in the history of statistics." [5] It introduced the 1900's with Student's paper.



Those searching for it today will find this version along with a note that "This article was published online on 18 December 2013. Subsequently, a typo was identified in the timeline, and the article was corrected on 1 December 2014."

1908 William Sealy Gosset, chief brewer for Guinness in Dublin, describes the *t*-test. It uses a small number of samples to ensure that every brew tastes equally good.

6 1926 Census of the Irish Free State

It is not clear that the Library & Archives Canada people had been warned that a few forms in the 1901 and 1911 census of Ireland might have been filled using the old Gaelic script. Some time after the contract ended, a Kerry historian discovered that the software had not properly interpreted **na Uíbheácha a Dó** rcríofa ra rean-ċló Ġaelaċ in his ancestor's name. It led to the newspaper headline "Irish names translated as 'gobbledygook' in census"—and a story in which the person heading the project at the National Archives conveniently blamed the Canadian contractors, who "made a hash of their attempts to deal with a **language none of them knew**."

The people at the National Archives are a little humbler in 2026, having realized that **their own in-house transcription** of the 1926 census returns also "made a hash of some of their attempts to deal with a **language they all know**." The issue isn't so much the language; it's the handwriting!

In an essay earlier this year, I described some of their 'mutations' of **ages** of persons on the District Enumeration District (DED) where I grew up. But, until this week[6] I had not appreciated just how consequential some mutations can be, especially for those who rely on censuses for genealogical research.

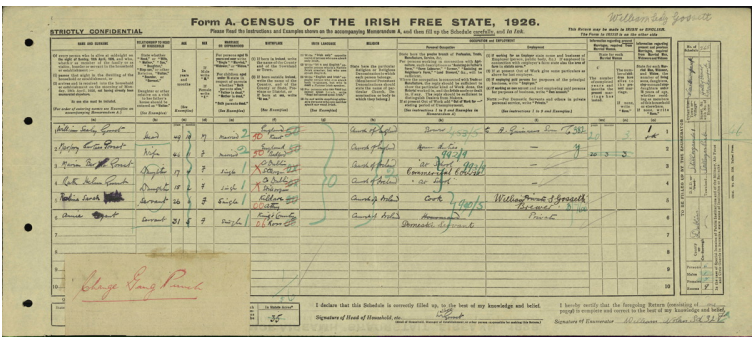
Imagine my surprise—and frustration—when **my search, again just on the surname, did not find a single "Gosset" (or "Gossett" !) recorded in the entire country**. Had he gone on an unrecorded sabbatical? Were the entire family on a late-season ski-ing holiday in Europe? Or on an overseas trip to a women's field hockey tournament?

Why not use other search fields? The *occupation* field isn't transcribed/put online yet. I didn't think of looking for the address he lived at in 1911. There must be a lot of men named **William** and aged about 50 living somewhere in Dublin, but how many would have the names **William Sealy**? The answer: 1!



It turns out that this time round his surname fell victim to a bigger and more serious *transcription* error. In retrospect, it looks like the National Archives people oversold their process/prowess.[7]

THE NAI were soon to find out from the public that the 'modern computer programming and intelligent robotic process automation' and even the 'human in the loop' cannot ensure that it gets *every* case exactly right. Gosset's handwritten return is an interesting example.



His handwritten lowercase 's' loops in a way that easily fools automated scanners into reading an 'r'. Indeed, it would fool many humans as well—unless a reader recalled the name from a statistics instructor who was a big Gosset-at-Guinness fan!

One wonders if one of the NAI data-cleaning processes involved making a frequency distribution and looking at all of the 'singleton' (unique) surnames.

Some people may have wondered why the NAI did not anticipate

the challenge of using OCR on handwriting, and why they did not use volunteer human transcribers, as was done by `https://duchas.ie` with the material from the Schools Folklore Collection. Presumably, such a *meitheal* of volunteer transcribers, as were used in the Schools project, was not feasible, especially given the scale, privacy and deadline constraints they were under.

Had the NAI had the luxury of a *meitheal* of volunteer transcribers, the extra information in the employment column, where a census clerk used darker ink (and even crossed out the second 's' in the clearly written 'Gossett') would have saved Gosset from this latest indignity.

It's a 'data' error that—until it is corrected—continues to make William Sealy Gosset the anonymous statistician his Guinness bosses wanted him to be.[8]

Worth reading

- Joan Fisher Box. 'Guinness, Gosset, Fisher, and Small Samples'. Statistical Science, Vol. 2, No. 1 (Feb., 1987), pp. 45-52. Stable URL: https://www.jstor.org/stable/2245613

  As for staying under the radar, here is a link to a mysterious photograph with the caption "RAF with 'Student' " which I (and obviously others) have seen in Fisher's archives in Adelaide.[9] In that same collection, one finds this touching letter from Marjory Gosset to Fisher dated Nov. 8th, 1937, thanking him for his letter of sympathy and stating how highly she valued his tributes to her husband "both as a friend & as 'Student'."

- Stephen T. Ziliak 'Guinnessometrics: The Economic Foundation of "Student's" t' . Journal of Economic Perspectives—Volume 22, Number 4—Fall 2008—Pages 199–216 Retrospectives.
  It also provides the **location of the notebook**: Folder 282, "Student's Haemocytometer Paper on Yeast-cell Counts," 1905–1907, in the Pearson Papers, UCL.

`james.hanley@mcgill.ca`
`https://jhanley.biostat.mcgill.ca`

2026.11.04

7 BTW: Which of these titles would you vote for?

(all from Google AI)
Option 1: The Meme-Culture Angles (Highly Relatable)

These play on common social media formats about being ignored or ghosted.

• Ghosted by the Algorithm: The Century-Long Erasing of William Sealy Gosset

• William Sealy Gosset Can't Catch a Break (A modern, universal equivalent to Rodney Dangerfield's catchphrase)

• The NAI Said "Human in the Loop" and Still Called the Father of Modern Statistics "Gorret"

• Who is Student? Not Even His Boss Knew
Option 2: The Data & Tech Snark Angles

These use modern data science terminology to poke fun at the institutional errors.

• Garbage In, "Gorret" Out: How Automated Scanners Made a Hash of Gosset

• William Sealy Gosset: Overfitted, Mistranscribed, and Unrecognized

• An Elegant Pen Name vs. Century-Old Typos: Guess Who Wins?

• The "One-Sample" Statistician With Way Too Many Misspelled Names
Option 3: The Short & Punchy "Clickbait" Style

Perfect if you want something brief that immediately makes a reader want to know what happened.

• The Unmasking and Un-Spelling of "Student" • Will the Real W.S. Gosset Please Stand Up?

• Chief Brewer, Master Statistician, Anonymous "Gorret"
Option 4: A Literal Rodney Dangerfield Translation

If you still want the soul of the Rodney Dangerfield line but translated for modern UK/Irish ears:

• No Respect at All: The Non-Stop Misspelling of W.S. Gosset

Do any of these styles fit the vibe you want to present to the ISA and Significance, or should we lean closer into a Guinness-them pun?

Endnotes

1. Joan Fisher Box's article, which begins with a rarely seen 1899 photo of Gosset, has a list of the early chemistry-trained hires from Oxford and Cambridge
2. In a 1908+100 presentation and at a session to mark his birthday I describe how, in 1912, Major Greenwood showed one of the first extra-mural users of his table an ad-hoc way apply it in a two-sample problem. It was only in 1925 that Fisher showed that it was a special case of a much wider family where the numerator involves a Gaussian random variable, and the denominator the square root of an independent chi-squared random variable.
3. "American students of statistics have long speculated as to the identity of "Student," whose papers of 1908 and 1912 in Biometrika inspired R. A. Fisher, and who was the first to take the fundamental step of examining the distribution in random samples of the ratio of mean to standard error, thus opening the way for escape from the haze of hypothetical standard errors and inverse probability which have obfuscated the theory of statistics. I have heard guesses in this country identifying "Student" with Egon S. Pearson and with the Prince of Wales.
   He is now so well known in Great Britain that no confidence is violated in revealing that he is W. S. Gosset, a research chemist employed by a large Dublin brewery. This concern years ago adopted a rule forbidding its chemists to publish their findings. Gosset pleaded that his mathematical and philosophical conclusions were of no possible practical use to competing brewers, and finally was allowed to publish them, but under a pseudonym, to avoid difficulties with the rest of the staff. Following his example "Sophister" and Mathetes," younger chemists employed by the same brewery, have published contributions to the theory of statistics. This same firm has made large grants to the Rothamsted station for research on barley, which incidentally has stimulated the work in statistics." [H. Hotelling. British statistics and statisticians today. Journal of the American Statistical Association Volume 25, 1930 - Issue 170, pp 186-190]
4. "With very deep regret the Editorial Committee has to report the death on 16 October 1937, of Mr W. S. Gosset, whose scientific contributions under

the pseudonym of “Student” are well known to all statisticians. It is hoped to include some account of his life and work in the next issue of the Journal. Mr Gosset had been working at the following paper during the past summer, and a fortnight before his death had discussed the draft, which is printed below, with Dr J. Neyman and Prof. E. S. Pearson. It was then agreed that certain points in sections 2 and 3 needed clarification and Mr Gosset proposed to undertake this work himself; unfortunately this final revision was never completed. Dr Neyman and Prof. Pearson have therefore added in a separate Note (pp. 380-88 below) some comments, for which they take full responsibility, regarding the points on which they know Mr Gosset had intended to enlarge.-ED.”
“Student” Comparison between balanced and random arrangements of field plots. Biometrika Vol. 29, No. 3/4 (Feb., 1938), pp. 363-378

5. It began at 450 BC, when “Hippias of Elis uses the average value of the length of a king’s reign (the mean) to work out the date of the first Olympic Games, some 300 years before his time”) —and ended with two 2012 events: “Nate Silver, statistician, successfully predicts the result in all 50 states in the US Presidential election. He becomes a media star”, and “The Large Hadron Collider confirms existence of a Higgs boson-like particle with probability of five standard deviations – around one chance in 3.5 million that all they are seeing is coincidence.”

6. In a *statistical* manuscript I was submitting, I wanted to show an example of low intra-class correlations, and so I used the **ages** of the persons in the 200 or so households in Bere Island in 1926.

   > I chose it to show the large intra-household variation, and thus the low ICC. The five persons (rows 26-30) in one of those households (rows 26-30) were my great-grandfather (age 76), my grandfather (46), my father (17) and my aunts (15 and 12). The ICC based on the originally-released transcriptions of the handwritten household returns was even lower, as the AI-based OCR software transcribed the 12 as 2!

   However, the journal insists that my manuscript not identify me. And so, to make it anonymous, I wrote

   > To preserve *my* anonymity, this household is ***NOT*** that of ***MY*** family, but (as a placeholder, for now) of a person in an urban household who, when publishing statistical papers, also sought to preserve *their* anonymity!

   and proceeded to find the link to the return from the Gosset household in 1926. I planned to insert it instead, just so the reviewers would see what data were available.

7. A few days before the data were released, one of its people told RTE readers that the **digitization** process

employed cutting edge technology. The National Archives acquiring two powerful high-speed scanners to facilitate digitisation of the census forms after their conservation, and to capture high-quality colour images, scanned to 400 DPI (dots per inch).

The process resulted in the creation of approximately 1.5 million images, which were digitised in a number of digital formats, that are stored and backed up to ensure their long-term preservation, and that there is no risk of loss or corruption.

And, as for the **transcription**, she went on to say

Finally, the data contained on the scanned census forms was transcribed - in other words extracting the text from the forms in order to make the documents, and the information they contained, searchable when converted into an online database. Transcription presented its own challenges. Millions of individual data points, such as names, ages, townlands and streets, all had to be transformed into a structured, searchable database.

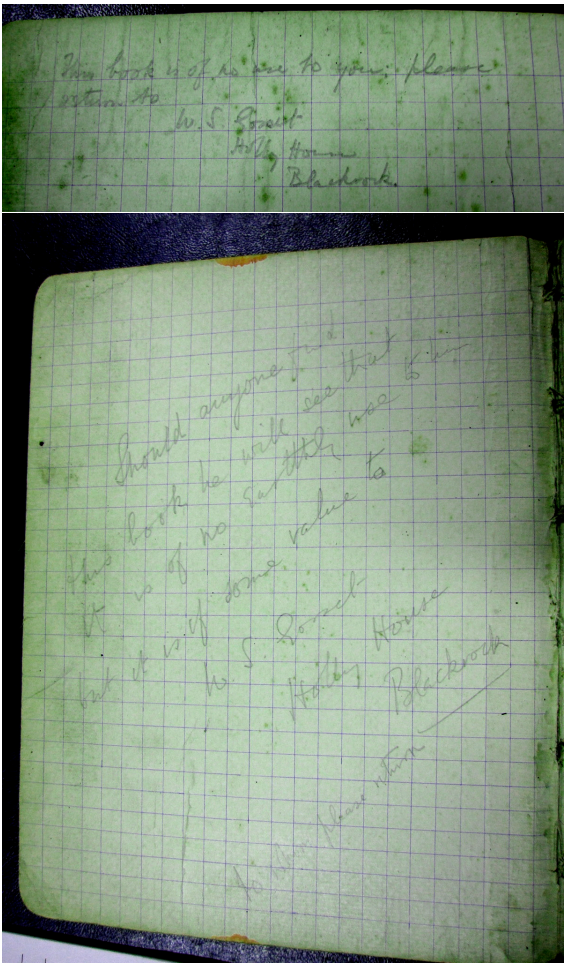
To manually transcribe the data in the 1926 Census would take one person an estimated 23 years. To make the project achievable within a realistic timeframe, we relied on modern computer programming and intelligent robotic process automation. However, automation was only one part of the story.

Our automated processes were initially trained using patterns identified in the 1911 Census and then refined through repeated physical checks. But because the 1926 returns introduced new complexities, particularly the inclusion of forms completed in Irish, automation alone was never going to be sufficient. Irish-language entries had not appeared in earlier censuses, and many were written in seanchló, a traditional Gaelic typeface with unique characters and a visual structure entirely different from English script.

These features made it particularly difficult to train automated reading models to a fully reliable standard, so the Irish-language forms were instead transcribed by fluent Irish speakers on the National Archives team.

Because of these challenges, every single piece of transcribed data from every page of the census was manually checked, twice, by National Archives staff. This "human in the loop" stage was essential to ensure that names were spelled correctly, ages were accurate, and the automated outputs matched the original records.

8. Thanks to the Irish Statistical Association, primary school 'students' in Blackrock will know both his pen name and his real name. And if people ever find his notebooks



lying around, they will know where to return them to.

9. We can tighten the (left-censoring of the) date to 1937, and—although I am not sure that it really is 'RAF' on the left—confidently take it that the one on the right is the **ever-elusive** William Sealy Gosset. But can we tighten further and specify **where** and **when**? Here is a poorer quality, but not cropped, photo that I took in 2016 in a hurried visit to that Fisher Collection in Adelaide.



This exchange between Marjory Gosset and Fisher suggests that this is a rare photograph indeed.