

## CERVEJA E ESTATÍSTICA: VIDA E OBRA DE UM MESTRE CERVEJEIRO

### BEER AND STATISTICS: LIFE AND CONTRIBUTIONS OF A MASTER BREWER

LORI VIALI\*

MÁRCIA ELISA BERLIKOWSKY\*\*

#### RESUMO

A estatística foi desenvolvida basicamente por não estatísticos. Profissionais de todas as áreas contribuíram para o seu desenvolvimento e crescimento. William Sealey Gosset, mais conhecido sob o pseudônimo de Student, foi um Químico e mestre cervejeiro da Guinness da Irlanda. Ele foi um dos pioneiros na aplicação do método científico na fabricação de cerveja e isto incluía o tratamento de dados de variadas fontes sendo que a maioria era formada por pequenos conjuntos. Na época não existia uma teoria para a tomada de decisões com base em pequenas amostras e esta foi a principal contribuição de Gosset. Além disso, pode-se dizer que ele foi um dos pioneiros na introdução do controle de qualidade na indústria, pois efetuou muitos testes para estimar a duração da cerveja sob diversas condições de armazenagem, fabricação e transporte. O procedimento desenvolvido por Student, baseado na distribuição por ele derivada, é hoje um dos principais testes utilizados para comparar pequenas amostras tanto dependentes quanto independentes.

**Palavras-Chave:** História da Estatística, Distribuição t, Student, Contribuições de William S. Gosset, Teste t de Student.

#### ABSTRACT

*The statistic was primarily developed by non-statisticians. Professionals from all areas contributed to its development and growth. William Sealey Gosset better known under the pseudonym Student was a chemist and Brewster of the Irish Guinness brewery. He was a pioneer in the application of the scientific method in brewing and this included the treatment of many data from various sources especially for small ensembles. At the time there was no theory for making decisions based on small samples and this was the main contribution of Gosset. Moreover, it can be said that he was a pioneer in the introduction of quality control in the industry, since he made many tests to estimate the duration of beer under various storage conditions, manufacturing and transportation. The procedure developed by Student, based in his derived distribution, is today one of the main tests used to compare small samples both dependents and independents.*

**Keywords:** History of Statistics, t Distribution, Student, Contributions of William S. Gosset, Student t test.

---

\* Doutor em Engenharia de Produção. Professor Titular da Faculdade de Matemática da PUCRS e professor permanente do PPGEDUCEM (Pós-Graduação em Educação e Ciências e Matemática) da PUCRS. Professor Associado do Instituto de Matemática e Estatística da Universidade Federal do Rio Grande do Sul. E-mail: viali@pucrs.br - viali@mat.ufrgs.br.

\*\* Mestre em Educação em Ciências e Matemática pela PUCRS. Doutoranda do PPEDUCEM. E-mail: aulas.marcia@gmail.com

## INTRODUÇÃO

William Sealy Gosset nasceu, em 13 de junho de 1876, na cidade de Cantuária (*Canterbury*), condado de Kent, Inglaterra. Ele era o mais velho, dos cinco filhos, de Agnes Sealy Vidal (1851-1938) e do Coronel Frederick Gosset (1846-1942) dos Engenheiros Reais. O pai de Gosset foi o último descendente de uma família de Hughenot (por parte de pai) que tiveram que abandonar a França, para fugir de perseguições, após a revogação do Édito de Nantes. Por esta mesma razão outro matemático francês Abraham de Moivre (1667-1754), descobridor da distribuição normal, teve que imigrar para a Inglaterra em 1685 (VIALI, 2014, p. 103).



William Sealy Gosset (1876-1938)\*

Seus estudos iniciaram na escola primária Cullercoat em North Shields, seguido pela escola Gore Court em Sittingbourne condado de Kent. O ensino médio foi realizado no colégio Winchester, uma escola pública em Winchester condado de Hampshire, que foi fundada em conjunto com o *New College* em Oxford, onde Gosset fez sua graduação. Um de seus colegas, em Hampshire, foi Godfrey Harold Hardy (1887 - 1947), um matemático inglês de destaque que foi professor de matemática na Universidade de Cambridge. Após se formar em Winchester Gosset tentou entrar para a academia militar dos Engenheiros Reais a exemplo de seu pai. Contudo ele foi rejeitado por ter problemas de visão. Ele se graduou em Matemática (1897) e Química (1899) pela Universidade de Oxford. Durante sua graduação foi aluno de George Biddell Airy (1801 - 1892) que contribuiu para o desenvolvimento da Astronomia e da Matemática (ZILIAK, 2009).

Em 16 de janeiro de 1906 ele casou com a técnica e capitã do time inglês de hockey feminino Marjory Surtees Phillpotts (1879 -1965). Marjory era a filha mais nova de James L. Philpotts, que foi por muitos anos diretor da Escola Secundária Bedford. Eles tiveram um filho e duas filhas (McMULLEN, 2013).

## GOSSET O QUÍMICO

Em 1899, então com 23 anos, ele foi trabalhar na cervejaria de Arthur Guinness e Cia Ltda. de Dublin, Irlanda. Ele foi um do grupo de jovens cientistas contratados pela empresa entre os anos de

1880 e 1890. Ele fez parte de uma tendência da indústria da época de contratar engenheiros cientificamente treinados. A Guinness contratou principalmente formados em química e física (TANKARD, 2001). Ele permaneceu nesse emprego por toda a vida e chegou ao topo da indústria cervejeira como cervejeiro mestre da Guinness em uma nova cervejaria em Londres. De 1899 a 1906 Gosset foi um cervejeiro aprendiz na cervejaria experimental próxima a principal. Em 1904 ele começou a enfrentar o problema de fazer inferências de pequenas amostras de malte e lúpulo, dois dos principais elementos da fabricação de cerveja (PEARSON, 1939).

Um dos primeiros problemas que ele enfrentou para produzir uma cerveja perfeita foi a de adicionar a quantidade exata de fermento à contínua fermentação da cevada. Se a quantidade ficasse abaixo da ideal ele teria uma fermentação incompleta e se colocasse demais teria uma cerveja muito amarga. A temperatura ambiente era outra variável que interferia no processo. Para tal ele tinha que contar as colônias de levedura e isto era feito por intermédio de um hemacitômetro. O desafio era estimar a quantidade de colônias em tonéis inteiros com base em pequenas amostras do conteúdo. Para tal ele não podia utilizar seu conhecimento químico e teve que desenvolver habilidades estatísticas. Ele notou que a contagem de colônias de leveduras não aderiria a nenhuma das curvas assimétricas desenvolvidas por Karl Pearson (1857-1936), mas aderiam à distribuição de Poisson. Em novembro de 1904, ele apresentou um relatório à diretoria da Guinness denominado de “Aplicação da ‘Lei do Erro’ ao trabalho da cervejaria” (RAJU, 2013).

Em 1904 Gosset iniciou a utilização da análise de regressão, uma técnica que ele aprendeu lendo o “Método dos Mínimos Quadrados” de Mansfield Merriman (1848 - 1925), de 1884, e a “Teoria Algébrica e Numérica dos Erros de Observação” do astrônomo real Sir George Biddel Airy (1801-1892) de 1861. Em 1908 ele utilizou a regressão para resolver a questão de determinar “a quantidade de lúpulo” versus “a duração da cerveja”. Munido apenas de uma calculadora mecânica ele estimou parábolas da forma  $L = A + BH^2$ , onde  $L$  = tempo de vida (em dias) da cerveja,  $A$  = tempo de vida (em dias) da cerveja sem lúpulo,  $H$  = libras de lúpulo e  $B$  uma constante (parâmetro) que depende do lúpulo e de outras condições (ZILIAK, 2008). Ele concluiu que a cerveja sem lúpulo poderia durar entre 12,2 e 16,7 dias e a com lúpulo poderia ultrapassar um mês. Esta conclusão foi obtida após numerosas repetições do experimento sob as mesmas condições (ZILIAK, 2008).

A cervejaria irlandesa foi fundada, em 1759, por Arthur Guinness (1725 - 1803) e está localizada no lado sul do rio Liffey e no lado oeste da cidade próxima a passagem ou porta St. James, que já não existe mais. A Guinness é a maior cervejaria da Irlanda e foi por um bom tempo a maior do planeta e sua cerveja preta amarga é popular em todo o mundo. A empresa ficou mundialmente conhecida pela publicação do livro dos Recordes Mundiais (*The Guinness Book of Records*) em 1955.

Além de contratar formados de Cambridge e Oxford com os melhores resultados em química a cervejaria inaugurou em 1900 o seu laboratório de pesquisa chefiado pelo mais notável químico cervejeiro da época, Horace Tabberer Brown (1848 - 1925) (BOX, 1987). Após estabelecer o laboratório de pesquisa, o próximo passo da empresa foi aprimorar a qualidade da sua matéria prima: a cevada e o lúpulo. Ela iniciou um programa de cultivo de cevada executando experimentos, em conjunto com os fazendeiros, com diferentes variedades e fertilizantes. Em 1901 uma casa de malte foi construída na porta de St. James, antiga entrada ocidental da cidade, que tinha um tamanho apropriado para obter o malte de cada lote experimental de cevada separadamente. Em 1903 uma cervejaria experimental foi aberta no mesmo lugar e assim todo o processo desde o plantio até a cerveja final poderia ser controlado por uma série de observações (BOX, 1987).

## O ANO COM PEARSON

Logo no início da carreira, Gosset percebeu a necessidade de realizar análises dos vários processos da empresa que envolvia desde a produção de cevada até a fermentação da cerveja, uma vez que todos afetavam a qualidade do produto final, a cerveja. Na época a teoria da estimação de grandes amostras já estava bem estabelecida, mas Gosset notou que pouco ou nada existia sobre estimação com pequenas amostras que eram típicas do seu trabalho. Assim ele teve que desenvolver a sua própria teoria. Gosset já tinha um método aproximado de lidar com a variabilidade das pequenas amostras, mas ainda não estava satisfeito, pois ele pretendia descobrir a matemática exata por trás da inferência com pequenas amostras. Ele falou para a empresa que pretendia consultar algum “matemático” sobre o tema. A empresa o enviou, então, para o laboratório de Karl Pearson na Universidade de Londres. Pearson era uma das personalidades científicas do seu tempo e teve posteriormente o crédito de haver estabelecido o campo da estatística. Desta forma Gosset passou o ano acadêmico de 1906-1907 no laboratório de Pearson. A decisão da empresa de enviar seu funcionário para Londres foi tomada após um encontro entre Gosset e Pearson. O encontro ocorreu em julho de 1905 e tanto Gosset quanto a empresa acharam que ele foi muito produtivo, pois durante uma conversa de uma hora e meia Pearson explicou para Gosset todos os métodos estatísticos então em uso (BOLAND, 1984).



Karl Pearson (1857-1936)\*

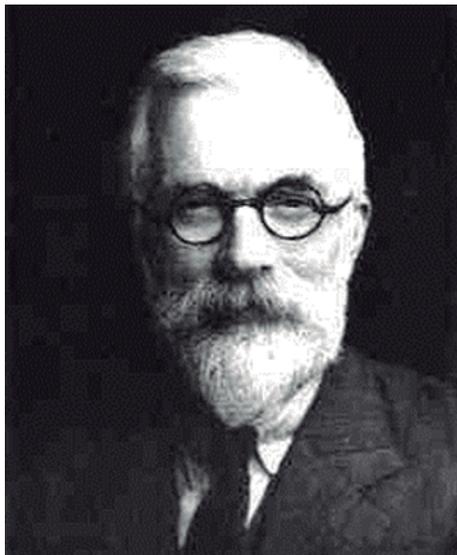
Durante sua estada Pearson ficou bem impressionado com Gosset e tentou convencê-lo a aceitar um cargo acadêmico. Contudo, nesta época, ele já estava casado e tinha um filho. Seu salário na Guinness era de £800 por ano, enquanto que o salário médio de um professor, na época, estava em torno de £600 anuais, de modo que ele recusou a oferta (BELLHOUSE, 2005).

Após ter passado um ano com Pearson, Gosset publicou o seu, agora famoso, artigo de 1908, que ele denominou *The Probable Error of a Mean* (Sobre o Erro Provável de uma Média), onde ele

observou que “s” (desvio padrão amostral) era um estimador errático de  $\sigma$  quando  $n$  era pequeno. Assim medidas usuais da precisão das estimativas eram inválidas para pequenas amostras. Seu artigo apresentou a distribuição de uma estatística, agora conhecida como t de Student, e introduziu a estimação com pequenas amostras por intermédio de uma família de distribuições t. A empresa concordou com a publicação do artigo com a condição de que ele utilizasse um pseudônimo (ele utilizou *Student*) e que o nome da empresa não fosse revelado e que os dados da empresa não aparecessem na publicação (ZABELL, 2008).

## O RELACIONAMENTO COM FISHER

Não é fácil mensurar a contribuição, do artigo de 1908, para a estatística, uma vez que ele é fundamental para a inferência como a conhecemos hoje, não apenas do ponto de vista da estimação, mas, também, sob a ótica dos testes de hipóteses e da análise de variância. Contudo, o trabalho de Gosset só se tornou realmente conhecido com o lançamento do livro de Fisher *Statistical Methods for Research Workers* (Métodos Estatísticos para Pesquisadores), em 1925, que foi o pioneiro na divulgação do método de *Student* para uma ampla audiência (ZILIAK, 2012).



Sir Ronald Aylmer Fisher (1890-1962)\*

Sir Ronald Fisher (1890 - 1962) que nutria grande admiração por Gosset e compartilhava com ele grande interesse pela experimentação agrícola denominou Gosset como o “Faraday da estatística”. O próprio Fisher escreveu, em 1937, por ocasião da morte de Student que “a morte tinha levado uma das mentes mais originais da ciência contemporânea”. Ele acrescentou, ainda, que embora Gosset não fosse um matemático profissional, ele tinha publicado, em 1908, uma “fundamentalmente nova abordagem ao problema clássico da teoria dos erros” (FISHER, 1939).

O relacionamento científico entre Gosset e Fisher teve início por intermédio do astrônomo Frederick John Marrian Stratton (1881 - 1960) que era o orientador de Fisher na universidade de Cambridge e sugeriu que ele enviasse uma cópia de seu primeiro artigo para Student. Fisher era então um matemático ambicioso, recém-formado, e estava à procura de um problema onde pudesse colocar a prova

seu talento (ZABELL, 2008). Em 1912 Fisher enviou a Gosset uma derivação completa e rigorosa da distribuição  $t$ , explorando sua facilidade com a geometria  $n$ -dimensional. Gosset enviou a prova para Pearson sugerindo que o resultado da prova de Fisher poderia merecer uma publicação na revista *Biometrika*. Pearson, contudo, criou dificuldades dizendo que não conseguiu entender a demonstração de Fisher (PEARSON, 1990).

## GOSSET O ESTATÍSTICO

De todos os 21 artigos de Gosset que foram publicados, 19 o foram sob o pseudônimo de “Student”. O propósito era proteger as melhorias da empresa, no controle de qualidade, dos competidores. Por muitos anos um ar de romantismo envolveu a publicação dos artigos de Student e somente algumas poucas pessoas conheciam sua real identidade, mesmo, depois de um tempo, após sua morte. O primeiro artigo publicado por Gosset foi denominado de *On the Error of Counting with a Haemocytometer* (Sobre o Erro de Contagem com um Hemocitómetro) e apareceu na revista *Biométrica* em fevereiro de 1907. Nele *Student* ilustrou a utilização prática do uso da distribuição de Poisson na contagem do número de células de leveduras em um quadrado de um hemocitómetro (BOLAND, 1984).

O artigo sobre a distribuição  $t$  foi publicado na revista *Biometrika*, em março de 1908, como um modelo que pudesse ser utilizado com pequenas amostras, onde Gosset percebeu que se elas fossem retiradas de uma população unimodal e simétrica, as mesmas poderiam ser caracterizadas pelo seu tamanho amostral.

Se uma população (variável)  $X$  de média  $\mu$  e variabilidade (desvio padrão)  $\sigma$  for padronizada por meio da transformação  $Z = (x - \mu)/\sigma$ , então a variável  $Z$  terá média “zero” e desvio padrão unitário. Essa mudança, de fato, coloca a variável centrada em torno da origem e reduz a variabilidade a um. Ela, no entanto, não altera a forma da variável, isto é, se a variável tiver certa assimetria e determinada curtose (achatamento) a mudança manterá esses valores. Assim se  $X$  for normal ela continuará tendo um comportamento normal. Gosset já conhecia isso; o que ele de fato precisava descobrir era qual o comportamento da variável  $Z = (X - \mu)/\sigma$  quando o valor de  $\sigma$  fosse desconhecido e estimado por meio de uma amostra. Também já era conhecido na época que para grandes amostras a variável  $Z$  continuava praticamente normal padrão, isto é, com média igual a zero e variabilidade igual a um (LEHMANN, 1999).

O problema de Gosset era saber qual seria o comportamento dessa variável quando as amostras fossem pequenas. O que ele percebeu é que esse resultado mantinha a simetria em torno de zero, mas não a variabilidade, ou seja, que a variabilidade dependia agora do tamanho da amostra utilizada, quanto menor, mais variável eram os resultados. Assim o modelo  $t$  de *Student* pode ser caracterizado por um único parâmetro o tamanho amostral.

Na época da publicação, a importância do resultado de Gosset não foi totalmente reconhecido. O foco entre os estatísticos contemporâneos era na teoria das grandes amostras e a ênfase de Gosset em pequenas amostras, resultantes do seu trabalho na cervejaria, o deixou à parte. De fato, foi somente após Fisher ter generalizado a distribuição  $t$  que ela teve seu uso difundido tanto na cervejaria quanto fora dela (FIENBERG e LAZAR, 2001, p. 313).

O modelo  $t$ , nem sempre, está associado com pequenas amostras. Apesar de ter-se originado, dessa forma, desenvolveu-se e ganhou vida própria. Em geral ele é especificado com um parâmetro genérico  $\nu$  que é denominado de “graus de liberdade” e quando necessário estabelecer a relação com valores amostrais coloca-se  $\nu = n - 1$ , isto é, o parâmetro do modelo probabilístico  $t$  é igual ao número de elementos amostrais subtraído de uma unidade. Tankard, 1984 destaca que em comparação com Galton, Pearson ou Fisher, Gosset escreveu poucos artigos, embora estes sejam tão influentes quanto a grande quantidade publicados pelos demais, particularmente o artigo sobre a distribuição  $z$  (como  $t$  foi originalmente denominada) de 1908. Este artigo, com pequenas modificações, levou ao desenvolvimento da análise de variância e revolucionou o modo de pensar da época levando a descoberta ou a criação a uma série de outros testes.

Durante o ano que passou no laboratório de Pearson, Gosset também tentou obter a distribuição do coeficiente de correlação pelo mesmo método que havia utilizado para chegar a distribuição das médias para pequenas amostras. Contudo, ele conseguiu apenas uma solução para o caso particular de o coeficiente ser igual a zero. Segundo Box, 1987 foi quase um milagre que ele conseguiu chegar tão longe, em virtude de sua matemática precisar ter a ajuda de palpites inspirados.

Em 1907 Gosset retornou a Dublin como chefe da cervejaria experimental, posição que ele ocupou até o ano de 1935. Ele se ofereceu como voluntário para lutar na primeira grande guerra, mas foi recusado em função do seu problema de visão. No início de 1920 ele se tornou o chefe do Departamento de Estatística que ele criou na Guinness. Sua primeira tarefa foi a de estimar o efeito de uma campanha publicitária sobre a venda de cerveja na Escócia. Em setembro de 1935 ele foi promovido a cervejeiro mestre da nova fábrica, da empresa, em Park Royal, Londres (que hoje está fechada) e, em setembro de 1937, foi nomeado cervejeiro chefe de toda a Guinness. Com a empresa fabricando até 100 milhões de galões de cerveja anualmente, Gosset introduziu o lado quantitativo da elaboração científica de cerveja e com ele um conjunto de recursos experimentais e estatísticos (ZILIAK, 2009).

A estatística inglesa Florence Nightingale David (1909 - 1993) tinha grande admiração por Gosset, em uma entrevista, em 1989, declarou a seu respeito:

Gosset foi um homem extraordinário. Eu acho que ele foi, de fato, uma grande influência para a estatística. [...] uma pessoa extraordinária. Ele fazia perguntas, não existem dúvidas sobre isso. Ele fazia as perguntas e Fisher e Pearson as colocavam em linguagem estatística e então Neyman passava para a linguagem matemática. Mas eu acho que muito disto veio de Gosset. Eu tinha um grande respeito por ele, ele foi um grande homem (LAIRD, 1989, p. 238-239).

Lehmann a propósito da declaração de F. N. David coloca que:

dez anos depois quando reli esta entrevista eu fiquei intrigado pela sua declaração sobre a influência de Student e eu a investiguei em um artigo, de 1999, com base na correspondência dele com Fisher e Egon Pearson. Foi fascinante seguir o desenvolvimento gradual da teoria dos testes normais de Fisher e a teoria de Neyman-Pearson e o impacto de Student foi, sem dúvida, seminal (LEHMANN, 2008, p. 119).

## A DISTRIBUIÇÃO T

Gosset partiu das seguintes hipóteses: seja  $X_1, X_2, \dots, X_n$  uma sequência de variáveis aleatórias independentes e identicamente distribuídas tal que  $\mu = E(X_i)$  e  $s^2 = \text{Var}(X_i)$  e que cada  $X_i$  tenha uma distribuição normal, isto é,  $X_i \sim N(\mu, \sigma)$ . Sejam  $\bar{x}$  e  $S^2$  respectivamente a média e a variância amostral de  $X_i$ . Então a estatística:  $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$  tem uma distribuição t com  $v = n - 1$  graus de liberdade, isto é, ela

tem uma função densidade de probabilidade dada por:

$$t_v(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{v}{2}\right) \left(1 + \frac{x^2}{v}\right)^{\frac{v+1}{2}}} \quad v = 1, 2, 3, \dots \quad x \in \mathfrak{R}$$

onde o único parâmetro  $v$  (graus de liberdade) é igual a  $n - 1$  e  $\Gamma$  é a função gama, que é uma extensão da função fatorial. Para qualquer valor inteiro e positivo de  $v$  a distribuição t assume uma forma muito parecida com a curva normal-padrão ( $Z$ ), sendo que a aproximação será tanto melhor, quanto maior for o valor de  $v$ . Para chegar ao resultado acima Gosset teve que mostrar que:

- $\bar{x} \sim N(\mu, \sigma^2/n)$
- $(n - 1)S^2/\sigma^2 \sim \chi_{n-1}^2$
- $\bar{x}$  e  $S^2$  eram independentes
- $t \sim t_{n-1}$

O primeiro destes resultados já era bem conhecido na época de William e poderia ser encontrado nos livros textos sobre a teoria dos erros como de Airy de 1861 e Czuber de 1891. O segundo e o terceiro tópico do artigo de Gosset foi utilizado para derivar as três últimas das propriedades acima (ZABEL, 2008). Na época ele trabalhou com a quantidade

$$z = (\bar{X} - \mu) / \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$$

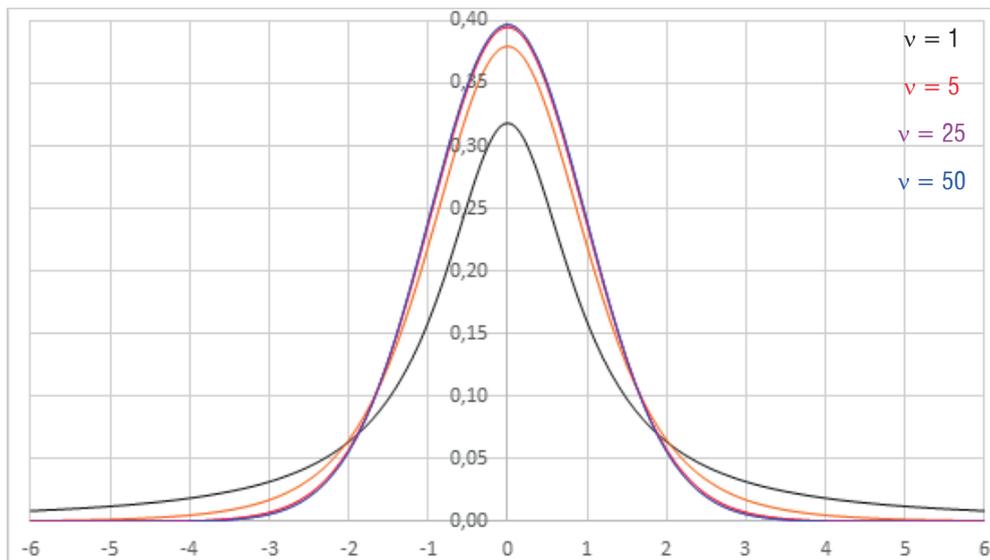
$t = z\sqrt{n-1}$  foi introduzida por Fisher em 1925 e posteriormente adotada por Gosset (Eisenhart, 1979). Na época que *Student* escreveu o seu artigo a distribuição de  $S^2$  já era conhecida, mas não na Inglaterra. Friedrich Robert Helmert (1843 - 1917) um geodesta alemão, que foi professor da universidade de Berlin, já a tinha derivado, em 1872 (HALD, 1998). Gosset poderia ter simplesmente citado o artigo de Elmert, mas ele não conhecia este resultado e assim conjecturou corretamente que  $(n - 1)S^2/\sigma^2 \sim \chi_{n-1}^2$  e escreveu “em consequência é provável que a curva encontrada represente a distribuição teórica de  $S^2$ , assim, embora, nós não tenhamos, de fato, uma prova, iremos assumir que isto é de fato assim!”

Quanto a distribuição de  $S^2$  e de  $\bar{x}$ , Fisher (1939) coloca que Gosset mostrou de uma forma trabalhosa que  $S^2$  era não correlacionada tanto com  $\bar{x}$  quanto com  $\bar{x}^2$  e acrescenta “este foi a falha mais marcante em seu argumento, pois, de fato, não era meramente a distribuição de  $S$ , descoberta

por Helmert, mas a distribuição conjunta exata de  $S$  e  $\bar{x}$  que ele precisava para desenvolver o seu teste”. Outros detalhes da demonstração podem ser vistos em *Student*, 1908.

A Figura 1 apresenta a forma da distribuição de Student para alguns valores de  $v$ .

**Figura 1** - Modelos da t de Student para alguns graus de liberdade.



Fonte: construção do autor, utilizando uma planilha.

A função gama, que é utilizada pela distribuição t, é considerada uma extensão do fatorial para o domínio dos números complexos, com exceção dos números inteiros negativos. Essa função é geralmente vista em cursos avançados de cálculo e é definida por:

$$Gama(x) = \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Ela apresenta algumas propriedades peculiares, como, por exemplo:

Se  $n$  é um inteiro, então:

$$\Gamma(n) = (n - 1)! \text{ e assim } \Gamma(1) = 0! = 1, \Gamma(2) = 1! = 1, \Gamma(3) = 2! = 2, \text{ etc.}$$

Ela é uma função recursiva, isto é:

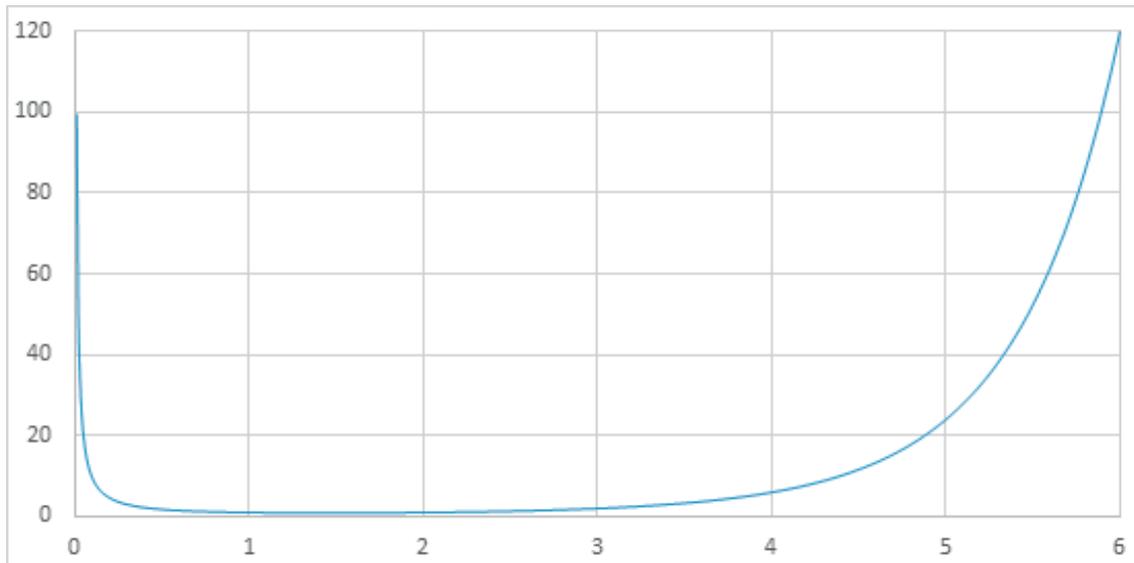
$$\Gamma(x + 1) = x \cdot \Gamma(x).$$

Para valores não inteiros, como por exemplo,  $1/2$  e  $3/2$  ela vale:

$$\Gamma(1/2) = \sqrt{\pi} \text{ e } \Gamma(3/2) = \frac{1}{2} \sqrt{\pi}.$$

Graficamente a função Gama tem um comportamento estranho, especialmente para os números negativos devido aos pontos de descontinuidade. A Figura 2 mostra o gráfico da função Gama apenas para os números positivos que são o interesse aqui, uma vez que uma fdp (função densidade de probabilidade) é sempre positiva.

**Figura 2** - Gráfico da função Gama no domínio dos números reais positivos.



Fonte: construção do autor utilizando uma planilha.

Convém observar que a função Gama não faz parte exatamente do modelo criado por Student, mas que ela é utilizada como uma constante de normalização, isto é, para que a área total sob a curva tenha o valor igual a 1, que é uma das características de uma função densidade de probabilidade. De fato, se ignoramos a constante de normalização na expressão da distribuição t ela ficará reduzida a:

$$t_v(x) = C \left( 1 + \frac{x^2}{v} \right)^{-\frac{v+1}{2}} \quad v = 1, 2, 3, \dots \text{ e } x \in \mathfrak{R}, \text{ onde a constante } C \text{ é dada por:}$$

$$C = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma\left(\frac{v}{2}\right)}$$

### Propriedades da distribuição t

A expectância (média) ou valor esperado é igual a zero para  $v > 1$  e se  $v = 1$  a distribuição t é idêntica à distribuição de Cauchy<sup>1</sup> e a expectância (valor esperado) não existe. A mediana e a moda da distribuição são ambas iguais a zero. A variância é igual a  $v/(v - 2)$  e note-se que ela tende a 1 à medida que o grau de liberdade tende ao infinito. Assim a variância existe para  $v > 2$  e é indefinida para outros valores. A assimetria é 0 para  $v > 3$  e a curtose é igual a  $6/(v - 4)$  e portanto definida apenas para  $v > 4$  e indefinida para outros valores.

A variável t é simétrica em relação ao valor  $x = 0$ , isto é,  $t(x) = t(-x)$ . Ela é duplamente assintótica, ou seja a medida que x cresce ou decresce indefinidamente t(x) tende a zero. A função é côncava e possui dois pontos de inflexão, isto é, pontos onde a função muda a concavidade em  $x = \pm\sqrt{v(v+1)}$ . Esses pontos convergem para  $x = \pm 1$  à medida que o grau de liberdade v tende ao infinito.

<sup>1</sup> Em homenagem ao barão e matemático francês Louis-Augustin Cauchy (1789 - 1857).

Certos valores de  $\nu$  dão origem a casos especiais. Para  $\nu = 1$  a função densidade de probabilidade, como mencionado acima, é a distribuição do Cauchy na sua forma padrão, isto é:

$$t_1(x) = \frac{1}{\pi(1+x^2)} \text{ e a função de distribuição é dada por: } T_1(x) = \frac{1}{2} + \frac{1}{\pi} \arctg(x).$$

Para  $\nu = 2$  a fdp e a função de distribuição são dadas por:

$$t_2(x) = \frac{1}{(2+x^2)^{3/2}} \text{ e } T_2(x) = \frac{1}{2} \left[ 1 + \frac{x}{\sqrt{2+x^2}} \right].$$

Para  $\nu = 3$  a fdp e a função de distribuição são dadas por:

$$t_3(x) = \frac{6\sqrt{3}}{\pi(3+x^2)^2} \text{ e } T_3(x) = \frac{1}{2} + \frac{1}{\pi} \arctg\left(\frac{x}{\sqrt{x}}\right) + \frac{\sqrt{3}x}{\pi(x^2+3)}.$$

Para  $\nu = 4$  a fdp e a função de distribuição são dadas por:

$$t_4(x) = \frac{3}{8} \left( \frac{4}{4+x^2} \right)^{5/2} \text{ e } T_4(x) = \frac{1}{2} + 1 + \frac{x(x^2+6)}{2(x^2+4)^{3/2}}.$$

Pelos quatro exemplos acima é possível notar um padrão para a função de distribuição. Pode-se ver que quando  $n$  é par o resultado é sempre algébrico e quando  $n$  é ímpar a função acumulada é uma mistura de funções algébricas e trigonométricas (SHAW, 2006). Pode-se notar, entretanto que uma generalização dessas expressões não é simples e nem tampouco seria muito produtivo de ser feito.

Se  $\nu$  for infinito a função densidade de probabilidade será dada por:

$$t_\infty(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \text{ que é a função densidade de probabilidade da distribuição normal padrão.}$$

Neste caso, não existe, de forma analítica, a função acumulada da variável.

### Relação com outras distribuições

A distribuição  $t$  (de Student) está relacionada com outras distribuições.

- A medida que o grau de liberdade tende ao infinito a distribuição  $t$  tende para a normal padrão, isto é, com média zero e desvio padrão igual a um;
- Quando o número de graus de liberdade é igual a um a distribuição  $t$  é igual a distribuição de Cauchy com parâmetros zero e um;
- Se uma variável aleatória  $X$  tem uma distribuição  $t$  então a variável  $X^2$  tem uma distribuição de Snedecor<sup>2</sup> (Fisher) com graus de liberdade um no numerador e  $\nu$  no denominador (DODGE, 2010).

### GOSSET PIONEIRO DO CONTROLE DE QUALIDADE

O trabalho de Gosset na Guinness envolveu principalmente o controle de qualidade. Ele entrou para a cervejaria praticamente na época em que a empresa iniciou o controle da matéria prima. O trabalho de Gosset envolvia mais do que simplesmente registrar dados de controle da matéria prima

<sup>2</sup>George Waddel Snedecor (1881-1974). Matemático e físico americano.

que consistia principalmente na cevada que origina o malte, o lúpulo que é o responsável pelo sabor e a preservação da cerveja e o fermento (leveduras) que age na fermentação do malte produzindo álcool e dióxido de carbono e da água que é a base do todo o processo.

Antes da Guinness iniciar o seu processo de contratação de jovens cientistas de Oxford e Cambridge, nos anos de 1890, a fabricação de cerveja era um processo misterioso que não tinha mudado praticamente em centenas de anos. Gosset e seus colegas aplicaram as realizações da revolução científica ao processo de fabricação da cerveja iniciando pelo controle da matéria prima com novos métodos. Eles utilizaram instrumentos científicos para registrar a temperatura e outros aspectos do processo de fabricação e introduziram medidas cuidadosas em todo o processo, medidas estas que requeriam algum tipo de tratamento estatístico (TANKARD, 1984).

Um dos primeiros problemas que William enfrentou foi como os vários fatores afetavam o tempo que a cerveja Guinness permanecia potável, uma vez que a cerveja, não leva conservantes e não é pasteurizada. Isto era avaliado pela sua acidez. Um dos fatores envolvidos era a condição do malte, mas outro fator complicador que também afetava a acidez e que deveria ser levado em conta na análise era a temperatura no processo de fabricação e durante a armazenagem. Ao mesmo tempo em que a empresa estava implantando o controle no processo de fabricação ele, também, estava em curso no campo e envolvia uma série de fatores que incluíam desde a variedade de cevada até o tipo e a quantidade de fertilizantes utilizados. Esses experimentos, da mesma forma, levavam a algum tipo de tratamento estatístico. Tanto os experimentos de campo quanto os do processo eram limitados quanto ao número de repetições que poderiam ser realizadas. Na cervejaria um experimento poderia levar um dia inteiro e isto significava que era possível de oito a doze repetições. Isto requeria que os estatísticos da empresa aplicassem conceitos estatísticos a amostras muito menores do que as que eram utilizadas por outros estatísticos. Por exemplo, tanto Pearson quanto Weldon<sup>3</sup> nos seus trabalhos biométricos estavam acostumados a trabalhar de centenas ou milhares (TANKARD, 1984).

Segundo Ziliak (2008), Gosset foi um pioneiro do controle da qualidade industrial, utilizando a análise de regressão em experimentos repetidos para determinar padrões de controle sobre inúmeras variáveis envolvidas na produção de cerveja. Isto ocorreu antes de o engenheiro americano W. Edwards Deming (1900-1993) ter introduzido a estatística e o controle de qualidade na manufatura japonesa. Ele resolveu um antigo problema relacionado às mudanças dos dados de uma série temporal com uma original diferença em um método de diferenças em 1909 e 1914.

## **GOSSET O FINAL**

Ao final do ano de 1935 a família Gosset deixou a Irlanda uma vez que ele fora promovido a cervejeiro mestre da nova planta da Guinness em Londres, que foi a primeira cervejaria da Guinness construída fora de Dublin. Cerca de dois anos depois em um sábado, 16 de outubro de 1937, em Beaconsfield, Inglaterra Gosset faleceu de um ataque cardíaco (o terceiro em um curto espaço de tempo). Ele tinha à época 61 anos e deixou seus pais, a esposa, os filhos e um neto. Ao todo William dedicou 38 anos a cervejaria Guinness e a Estatística. Ele publicou ao todo 21 artigos, praticamente um por ano, com algumas falhas. Considerando que ele não era nem acadêmico e tampouco tinha pós-graduação, isto é, de fato, surpreendente, não pela quantidade, mas sim pela qualidade e consequências de suas publicações. Suas pesquisas e seus métodos foram em grande parte para resolver

<sup>3</sup> Walter Frank Raphael Weldon (1860-1906). Zoólogo e estatístico britânico.

problemas práticos que lidavam com a produção da cerveja e se originavam devido a variações na cevada, lúpulo, malte e de outras condições experimentais (BOLAND, 2011).

## CONCLUSÃO

A estatística, diferentemente da probabilidade, é uma ciência aplicada e foi desta forma que ela se originou e evoluiu. Contudo, nem sempre ela é lecionada ou ensinada desta forma. Muitas vezes os cursos que envolvem estatística são lecionados com um enfoque puramente algorítmico e com isto a essência da disciplina é perdida. Isto tem contribuído para que boa parte dos alunos que já possuem um grande temor da matemática passem, também, a temer a estatística.

De fato, a abordagem não precisa ser desta forma. A estatística pode ser vista com um enfoque totalmente aplicado tendo sua utilidade destacada. O uso da história deve sempre acompanhar as aulas e a apresentação dos conteúdos, pois é raro que existam situações que não possam ser ilustradas por algum fato ou personagem da história. Infelizmente os professores que trabalham com a disciplina tem pouco ou nenhum conhecimento do seu desenvolvimento histórico, com isto eles estão perdendo um ótimo recurso didático e uma boa oportunidade de tornar a disciplina mais leve e atrativa.

Um dos objetivos deste artigo e outros é destacar a história da estatística e principalmente dar voz aos seus protagonistas, evidenciando, desta forma, que a disciplina não nasceu pronta e tampouco seu desenvolvimento foi desprovido de muito suor, estudo e empenho dos que se dedicaram a ela. Além disso, o seu desenvolvimento ilustra que ela foi feita principalmente por cientistas não estatísticos, pois esta profissão praticamente só começou a existir após a segunda grande guerra. O propósito é que este trabalho, entre outros sirva para divulgar, principalmente, entre os professores de estatística, a história e a evolução da mesma, de forma que eles possam lecionar não apenas tratamento de dados, mas uma disciplina viva e dinâmica e que os alunos percebam que ela foi construída por profissionais que dedicaram muitas vezes toda uma carreira a ela.

Aqui foi ilustrada a vida e obra de um profissional que, de fato, não era estatístico. Gosset era formado em Química e Matemática, mas sua principal contribuição, além do que ele exercia como mestre cervejeiro, foi na Estatística. Isso merece destaque, pois ele não era um acadêmico. Geralmente a publicação de artigos e pesquisas científicas é feita nas universidades. Com Gosset foi diferente, apesar de não em grande quantidade, de boa qualidade mesmo não estando inserido no meio acadêmico. O surpreendente é que ele anotava e divulgava o seu trabalho sem ter nenhum título de pós-graduação, uma vez que ele não possuía nem mestrado e nem doutorado. A estatística ele aprendeu com os livros disponíveis à época e já mencionados no artigo e com o ano que ele passou com Pearson em Londres.

A maior contribuição de William Gosset não foi a derivação de uma nova distribuição, mas sim a contribuição para o estudo das pequenas amostras e a derivação da distribuição exata de estatísticas como foi observado por Fisher no seu livro de 1925. Foi a partir desta ideia de Gosset que Fisher derivou a distribuição exata do coeficiente de correlação, mostrando com isso que a evolução da ciência é iterativa com alguns dando pequenos incrementos enquanto outros dando passos muito mais largos.

## REFERÊNCIAS

- AIRY, George B. **On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations**. 3. ed. London (UK): Mac-Milan, 1861.
- BELLHOUSE, D. R. Probability and Statistics Ideas in the Classroom - Lessons from History. **International Statistical Institute, 55<sup>th</sup> Session**, Sydney, Australia, 2005.
- BOLAND, Philip J. A biographical glimpse of William Sealy Gosset. **The American Statistician**, v. 38, n. 3, p. 179-183, August 1984.
- \_\_\_\_\_. William Sealy Gosset - An Inspiring 'Student'. **Proceedings of 58<sup>th</sup> World Statistical Congress**, Dublin, Ireland, 2011.
- BOX, Joan Fisher. Guinness, Gosset, Fisher, and Small Samples. **Statistical Science**, v. 2, n. 1, p. 45-52, Feb. 1987.
- CZUBER, E. **Theorie der Beobachtungsfehler**. Leipzig: Teubner. 1891.
- DODGE, Yadolah. **The Concise Encyclopedia of Statistics**. New York (NY): Springer, 2010.
- EISENHART, Churchill. On the Transition from 'Student's' z to 'Student's' t. **The American Statistician**, v. 33, n. 1, p. 6-10, 1979.
- FIENBERG, Stephen, LAZAR, Nicole. William Sealy Gosset. In: (Heide, C. C., Seneta, E., eds) **Statisticians of the Centuries**. New York (NY): Springer-Verlag, 2001.
- FISHER, R. A. Student. **Annals of Eugenics**, v. 9, p. 1-9, 1939.
- \_\_\_\_\_. **Statistical Methods for Research Workers**. 13th ed. New York: Hafner, 1925.
- HALD, A. **A History of Mathematical Statistics from 1750 to 1930**. New York: Wiley. 1998.
- HELMERT, F. R. **Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate**, Leipzig: Teubner. 1872.
- LAIRD, Nan M. A Conversation with F. N. David. **Statistical Science**, v. 4, n. 3, p. 235-246, 1989.
- LEHMANN, E. L. "Student" and Small-Sample Theory. **Statistical Science**, v. 14, n. 4, p. 418-426, 1999.
- \_\_\_\_\_. **Reminiscences of a Statistician: the Company I Kept**. New York (NY): Springer, 2008.
- McMULLEN, L. "Student" as man. **Biometrika**, v. 30, n. ¾, Jan. 1939, p. 205-210.
- PEARSON, E. S. "Student" as statistician. **Biometrika**, v. 30, n. ¾, p. 210-250, 1939.
- \_\_\_\_\_. WISHART J. (eds.) "Student's" Collected Papers. **Biometrika Office**, University College, 1942.

PLACKETT, R. L.; BARNARD, G. A. **Student: A Statistical Biography of William Sealy Gosset. Based on the writings of E. S. Pearson.** Oxford (UK): Clarendon Press, 1990.

RAJU, Tonse N. K. William Sealy Gosset and William A. Silverman: Two "Students" of Science. **Pediatrics**, v. 116, n. 3, p. 732-736, September 2005.

SHAW, William T. Sampling Student's T distribution - use of the inverse cumulative distribution function. **Journal of Computational Finance**, v. 9, n. 4, p. 37-73, Summer 2006.

STUDENT. The probable error of a mean. **Biometrika**, v. 6, n. 1, p. 1-25, March 1908.

TANKARD, James W. **The Statistical Pioneers.** Cambridge (MA): Schenkman, 1984.

The Guinness Book of Records. Ireland: **Guinness Brewery**, 1955.

VIALI, Lori. Algumas considerações sobre a denominada curva normal. **Revista Vidya**. Edição Especial de Aniversário, v. 34, n. 1, jan./jun. de 2014, p. 99-116

ZABEL, S. L. On Student's 1908 Article "The Probable Error of a Mean". **Journal of the American Statistical Association**, v. 103, n. 481, p. 1-7, March 2008.

ZILIAK, Stephen T. Great Lease, Arthur Guinness - Lovely Day for a Gosset! **Journal of Wine Economics** ("Beeronomics" issue), September 2009.

\_\_\_\_\_. Guinnessometrics: The Economic Foundation of "Student's" t. **The Journal of Economic Perspectives**, v. 22, n. 4, Fall, 2008, p. 199-216.

## APÊNDICE

A lista completa das publicações de Student (William Sealey Gosset) (PEARSON e WISHART, 1942):

01. On the error of counting with a haemocytometer. **Biometrika**, v. 5, n. 351, p. 351-360, Feb. 1907.

02. The probable error of a mean. **Biometrika**, v. 6, n. 1, p. 1-25, March 1908.

03. Probable error of a correlation coefficient. **Biometrika**, v. 6, n. 2/3, p. 302-10, Sep.1908.

04. The distribution of the means of samples which are not drawn at random. **Biometrika**, v. 7, n. 1/2, p. 210-214, Oct. 1909.

05. Appendix to Mercer and Hall's paper on 'Experimental error of field trials'. **Journal of Agricultural Science**, v. 4, p. 128-131, 1911.

06. The correction to be made to the correlation ratio for grouping, **Biometrika**, v. 9, n. 1/2, p. 316-320, 1913.

07. The elimination of spurious correlation due to position in time and space. **Biometrika**, v. 10, n. 1, p. 179-180, 1914.

08. Tables for estimating the probability that the mean of a unique sample of observations lies between  $-\infty$  and any given distance of the mean of the population from which the sample is drawn. **Biometrika**, v. 11, n. 4, p. 414-417, May 1917.
09. An explanation of the deviations from Poisson's law in practice. **Biometrika**, v. 12, n. 3/4, p. 211-15, 1919.
10. An experimental determination of the probable error of Dr. Sperman's correlation coefficients, **Biometrika**, v. 13, n. 2/3, p. 263-82, 1921.
11. Note by 'Student' with regard to his paper 'On testing varieties of cereals'. **Biometrika**, v. 16, n. 3/4, p. 411, 1924.
12. New tables for testing the significance of observations. **Metron**, v. 5, n. 3, p. 105-120, 1925.
13. Mathematics and agronomy. **Journal of American Society of Agronomy**, v. 18, n. 8, p. 703-791, 1926.
14. Errors of routine analysis. **Biometrika**, v. 19, n. 1/2, p. 151-64, 1927.
15. The Lanarkshire milk experiment. **Biometrika**, v. 23, n. 3/4, p. 398-406, Dec. 1931.
16. Article on 'Yield Trials' in Bailliére's. **Encyclopedia of Scientific Agriculture**, 1931.
17. On the 'z' test. **Biometrika**, v. 23, n. 3/4, p. 407, December 1931.
18. Evolution by selection: the implications of Winter's selection experiment. **The Eugenics Review**, v. 24, n. 4, p. 293-96, January 1933.
19. A calculation of the minimum number of genes in Winter's selection experiment. **Annals of Eugenics**, v. 6, n. 1, p. 77-82, October 1934.
20. Co-operation in large-scale experiments. A discussion opened by W. S. Gosset. **Supplement to the Journal of Royal Statistical Society**, v. 3, n. 2, p. 115-36, 1936.
21. Random and balanced arrangements. **Biometrika**, v. 29, n. 3/4, p. 363-78, 1938.

(\*) Figuras de domínio público disponíveis na rede.

---

**RECEBIDO EM:** 15 set. 2016.

**CONCLUÍDO EM:** 25 nov. 2016.