

DISTRIBUIÇÃO DE STUDENT

META

Apresentar a distribuição de Student

OBJETIVOS

Ao final desta aula, o aluno deverá:

definir a variável aleatória t ;
comparar a distribuição da variável t com a distribuição normal;

usar a tabela que relaciona as áreas de cauda da curva de distribuição com o número de graus de liberdade;
calcular intervalos de confiança para a média.

PRÉ-REQUISITOS

O aluno, deverá conhecer o cálculo de intervalos de confiança a partir da distribuição normal padronizada.



Quando dizemos que saber qual o tipo de distribuição seguida por uma variável em uma dada população, por exemplo, a distribuição normal, possibilita-nos fazer inferências sobre nosso problema, um dos parâmetros

INTRODUÇÃO

que estamos interessados em obter é o intervalo de confiança da média populacional a partir de um certo número de observações (amostra). Vimos que para isto utilizamos a variável aleatória z , relativa à distribuição normal padronizada, e que precisávamos também do desvio padrão populacional (ou seja, de toda a população, o qual é impossível mensurar). Acabamos usando em seu lugar o desvio padrão amostral, obtido para uma amostra grande, como uma boa aproximação. Para distribuições rigorosamente normais, vamos aprender agora que podemos calcular intervalos de confiança usando o próprio desvio padrão amostral e um outro tipo de variável aleatória, a variável t , proposta por Student.



(Fonte: <http://www.jornalcidade.uol.com.br>)

Vamo-nos lembrar de uma atividade que fizemos algumas aulas atrás, em que calculávamos intervalos de confiança para a média a partir de apenas uma observação, usando a distribuição normal padronizada e obtivemos um valor de $1,57 < m < 1,71$ com 95 % de confiança, a partir de uma única observação de 1,64 m. Não se pode negar que seja um intervalo largo.

VARIÁVEL NORMAL PADRONIZADA

Nesta aula, vamos conhecer uma alternativa para obtermos intervalos mais estreitos mesmo com poucas observações. Começamos revendo a expressão utilizada para a variável normal padronizada:

$$z = \frac{x - \mu}{\sigma}$$

Se no lugar do valor de cada observação da variável x usarmos uma média amostral, obtida para um número qualquer de observações, e no lugar do desvio padrão populacional ó usássemos o desvio padrão amostral s , estaremos definindo uma nova variável aleatória. E foi exatamente isto o que Student (pseudônimo do químico W. S. Gosset) fez e chamou a variável de t . Modernamente, divide-se também por $(N)^{1/2}$.

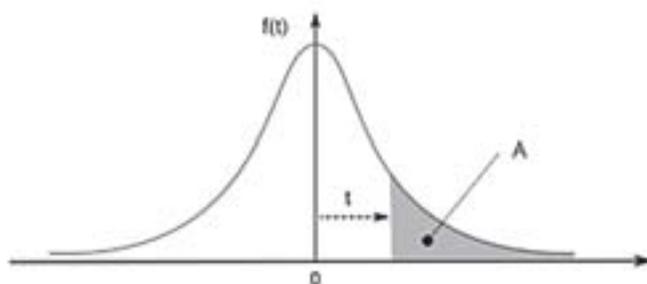
A forma moderna da distribuição de Student, portanto, fica:

$$t_{N-1} \approx \frac{\bar{x} - \mu}{s / \sqrt{N}}$$

Ela representa a distância entre a média amostral e a média populacional, dividida pelo produto entre o desvio padrão amostral e a raiz do número de observações na amostra.

CARACTERÍSTICAS

- a) A variável t se distribui de acordo com uma curva, cuja forma varia com o tamanho da amostra. Ou seja, com os graus de liberdade do desvio padrão, $N-1$. Deste modo, o que chamamos de graus de liberdade é o número de observações feitas na amostra menos 1.
- b) A distribuição de Student também é simétrica em torno da média e, quando $N-1$ cresce, a curva tende para a distribuição normal.

Curva de Student (Fonte: <http://www.mspc.eng.br>)

A importância prática da distribuição de Student é que podemos usar uma tabela de áreas de cauda, relacionadas às probabilidades de que observações aleatórias estejam nestes intervalos, da mesma forma que a distribuição normal padronizada. Para usar a tabela, apresentada a seguir, é importante verificarmos que, na primeira coluna temos o número de graus de liberdade, $n = N-1$ e, na primeira linha, temos alguns valores de áreas de cauda, lembrando que a distribuição t também é simétrica em torno da média. Portanto, para um nível de confiança de 95 % ou 0,95, precisamos procurar a metade dos 0,05 restantes, ou seja, 0,025. Se quisermos achar o valor de t para o nível de 95 % de confiança e 9 graus de liberdade, buscamos a intersecção entre $n = 9$ e 0,025, obtendo $t = 2,262$.

**Tabela parcial de pontos de probabilidade da distribuição
t com n=N-1 graus de liberdade.**

n	0,25	0,1	0,05	0,025	0,01	0,0083	0,005
1	1	3,078	6,314	12,706	31,821	38,343	63,656
2	0,816	1,886	2,92	4,303	6,965	7,664	9,925
3	0,765	1,638	2,353	3,182	4,541	4,864	5,841
4	0,741	1,533	2,132	2,776	3,747	3,966	4,604
5	0,727	1,476	2,015	2,571	3,365	3,538	4,032
6	0,718	1,44	1,943	2,447	3,143	3,291	3,707
7	0,711	1,415	1,895	2,365	2,998	3,13	3,499
8	0,706	1,397	1,86	2,306	2,896	3,018	3,355
9	0,703	1,383	1,833	2,262	2,821	2,936	3,25
10	0,7	1,372	1,812	2,228	2,764	2,872	3,169
11	0,697	1,363	1,796	2,201	2,718	2,822	3,106
12	0,695	1,356	1,782	2,179	2,681	2,782	3,055
13	0,694	1,35	1,771	2,16	2,65	2,748	3,012
14	0,692	1,345	1,761	2,145	2,624	2,72	2,977
15	0,691	1,341	1,753	2,131	2,602	2,696	2,947

Vamos ver agora como calcular intervalos de confiança para a média populacional a partir da distribuição de Student.

Como:

$$\mu = \bar{x} \pm t_{N-1} \frac{s}{\sqrt{N}}$$

Pode-se realizar este cálculo usando o valor de t obtido na tabela, para um certo intervalo de confiança e grau de liberdade, além do valor da média amostral, usando-se a expressão abaixo:

$$\bar{x} - t_{N-1} \frac{s}{\sqrt{N}} < \mu < \bar{x} + t_{N-1} \frac{s}{\sqrt{N}}$$

Por exemplo:

para 95 % de confiança $t_{N-1} = 2,447$ com 6 graus de liberdade

$$\bar{x} - 2,447 \frac{s}{\sqrt{N}} < \mu < \bar{x} + 2,447 \frac{s}{\sqrt{N}}$$



ATIVIDADES

Aplice aos valores de pesos de azeitonas usados há duas aulas:

0,9504; 2,1384; 1,436; 1,8952; 1,4608; 1,488; 1,636, com média amostral igual a 1,5721 e desvio padrão 0,3772.

Concluimos, com base no conteúdo estudado nesta aula, que a distribuição de Student, similar à distribuição normal, nos permite calcular intervalos de confiança,

CONCLUSÃO

porém, sem a necessidade de conhecermos o desvio padrão populacional, o qual demandaria que realizássemos medidas em toda a população, o que é impossível. No caso da distribuição normal, a suposição de que o desvio padrão obtido para uma amostra relativamente grande seria uma boa aproximação para o valor populacional. Contudo, a distribuição de Student nos propicia uma alternativa a esta aproximação, por utilizar o próprio desvio padrão amostral.

RESUMO



Calculamos, anteriormente, o intervalo de confiança da média populacional a partir de um certo número de observações (amostra), utilizando a variável aleatória z , relativa à distribuição normal padronizada pois precisávamos também do desvio padrão populacional (ou seja, de toda a população, que é impossível mensurar). Como não tínhamos tal parâmetro, usamos em seu lugar o desvio padrão amostral, obtido para uma amostra grande, como uma boa aproximação. Para distribuições rigorosamente normais, aprendemos, neste aula, que podemos calcular intervalos de confiança usando o próprio desvio padrão amostral e um outro tipo de variável aleatória, a variável t , proposta por Student. Para a determinação da nova variável, a variável t , partimos da expressão para a variável normal padronizada e , no lugar do valor de cada observação da variável x , usamos uma média amostral, obtida para um número qualquer de observações e , no lugar do desvio padrão populacional σ , usamos o desvio padrão amostral s . Assim, a forma da distribuição de Student representa a distância entre a média amostral e a média populacional, dividida pelo produto entre o desvio padrão amostral e a raiz do número de observações na amostra. A importância prática da distribuição de Student consiste em podermos usar uma tabela de áreas de cauda, relacionadas às probabilidades de que observações aleatórias estejam nestes intervalos, da mesma forma que a distribuição normal padronizada. Na primeira coluna desta tabela, temos o número de graus de liberdade, $n = N-1$. Na primeira linha, temos alguns valores de áreas de cauda, lembrando que a distribuição t também é simétrica em torno da média.

A DESCOBERTA DE W. S. GOSSET

Leia a seguir trecho extraído da página:

<http://www.pucrs.br/famat/statweb/historia/daestatistica/biografias/Gosset.htm> (08/02/2008)

...Em 1900, o Laboratório de Pesquisa Guinness foi inaugurado e era chefiado por Horace Brown, um renomado químico cervejeiro. Brown estava tentando obter a matéria prima para a cerveja ao menor custo possível, mas com a melhor qualidade possível. Para isto, fatores como variedades de cevada e lúpulo, condições de cultivo e maturação. Após alguns anos de pesquisa, Gosset teve a oportunidade de trabalhar como estatístico. Ele foi capaz de obter dados de diferentes exemplos de fabricação para tentar encontrar o que seria o melhor. Como os novos cervejeiros trabalhavam em conjunto, era natural para eles levarem os dados para que Gosset resolvesse os problemas numéricos. Ele escrevia artigos em suas horas de folga que publicava sob o pseudônimo de “Student”. Estes artigos eram sobre a distribuição da média e sobre o coeficiente de correlação. Ele nunca trabalhou de forma isolada. Ele conduziu experimentos com Hunter e Bennett de Ballinacurra, Buffin de Cambridge e Beaven de Warminster. Estes experimentos consistiam em testar variedades de sementes. Ele trabalhou com Neyman e com R. A. Fisher que não tinha boas relações com Pearson que não era o caso de Gosset que se dava bem com ambos. Ele descobriu a forma da distribuição “t” através de uma combinação de trabalho empírico e matemático com números aleatórios em uma das primeiras aplicações do método de Monte-Carlo...



PRÓXIMA AULA

Na aula 10, conheceremos a distribuição F e um exemplo prático de aplicação em Química.

REFERÊNCIAS

BARROS NETO, B.; SCARMINIO, I. E.; BRUNS, R. E. **Planejamento e otimização de experimentos**. Campinas: Editora da Unicamp, 1995.

BOX, G. E. P.; HUNTER, W. G.; HUNTER, J. S. **Statistics for experimenters. An introduction to design, data analysis and model building**. New York: Wiley, 1978.

BUSSAB, W. O.; MORETIN, P. A. **Estatística básica**, São Paulo: Atual, 1985.