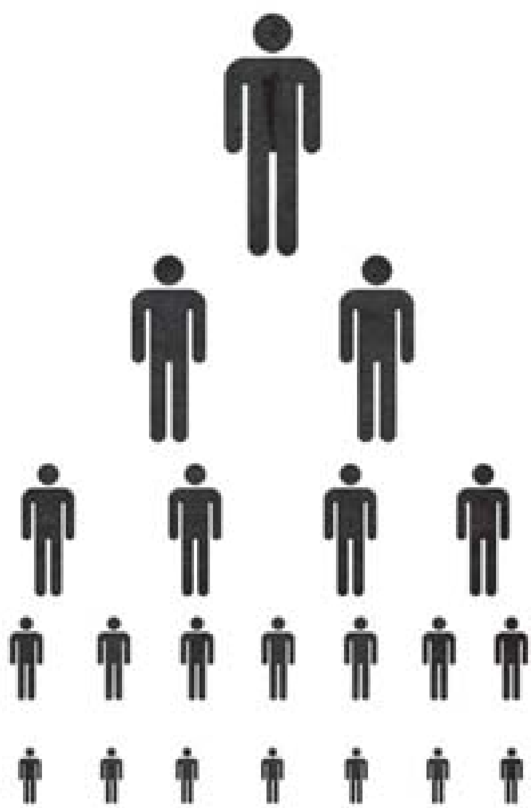
8
aula

COMBINAÇÃO LINEAR DE VARIÁVEIS ALEATÓRIAS E DISTRIBUIÇÃO DAS MÉDIAS

META
apresentar ao aluno a aplicação da combinação linear à distribuição das médias obtidas a partir de um número elevado de amostras.

OBJETIVOS
Ao final desta aula, o aluno deverá:
definir combinação linear de variáveis aleatórias;
calcular os parâmetros populacionais para as variáveis resultantes;
aplicar o método para diferentes amostras relativamente pequenas, retiradas aleatoriamente de uma mesma população;
comparar o resultado obtido com aquele de uma amostragem única relativamente numerosa.



PRÉ-REQUISITOS
O aluno deverá ter conhecimento básicos sobre expressões para média, variância e características gerais da distribuição normal.

INTRODUÇÃO

Imagine que você é um investidor e às vezes aplica em moeda estrangeira. Em um determinado momento, você tem reservas em dólares e em euros e gostaria de saber se tem um jeito de calcular qual seu total já convertido diretamente em reais, usando uma única equação matemática. Bom, neste caso, você sabe que vai ter que multiplicar cada valor nas duas moedas estrangeiras pelos respectivos fatores de conversão e depois disso somar as quantias. Esta operação que enunciamos é chamada de combinação linear e vamos ver aqui aplicações deste procedimento às distribuições estatísticas que temos estudado nesta disciplina.

O que chamamos de combinação linear é uma operação matemática em que duas variáveis x_1 e x_2 são somadas ou subtraídas após serem multiplicadas por coeficientes ou constantes específicas (a_1 e a_2), dando origem a uma nova variável y. Vamos ver uma expressão geral para ficar mais claro:

$$y = a_1x_1 + a_2x_2$$

Isto significa que não será necessariamente uma simples soma ou subtração $x_1 + x_2$ ou $x_1 - x_2$, e sim que as variáveis devem ser multiplicadas antes pelos coeficientes a_1 e a_2.

Vamos considerar que x_1 e x_2 são variáveis aleatórias independentes, com parâmetros populacionais (m_1, $ó_1^2$) e (m_2, $ó_2^2$). Após N observações, o valor médio de y pode ser calculado por:

$$\bar{y} = (1/N)\,\Sigma y = (1/N)\,\Sigma(a_1 x_1 + a_2 x_2)$$

colocando a_1 e a_2 em evidência:

$$\bar{y} = a_1[(1/N)\,\Sigma x_1] + a_2[(1/N)\,\Sigma x_2]$$

finalmente, percebendo que os termos entre colchetes são os valores médios de x_1 e x_2:

$$\bar{y} = a_1\bar{x}_1 + a_2\bar{x}_2$$

ou seja, a média da combinação linear é a combinação linear das médias! Este resultado nos é particularmente importante, pois adiante vamos comparar o efeito de se medir uma amostra relativamente numerosa de uma população, para se obter a média e a variância, com o procedimento análogo de se medir muitas amostras relativamente pequenas, tirar a média e a variância de cada uma e depois fazer a distribuição destas médias.

No caso da variância, temos:

$$\sigma_y^2 = [(1/N-1)]\Sigma(y_i-\bar{y})^2$$

e, finalmente:

$$\sigma_y^2 = a_1^2 x_1^2 + a_2^2 x_2^2$$

Isto considerando-se que as variáveis são independentes. Há o caso mais geral, em que a última expressão inclui também o coeficiente de correlação entre as variáveis, mas isto vai além dos nossos propósitos.

Considerando agora um número grande de variáveis aleatórias, a combinação linear fica:

$$y = a_1 x_1 + a_1 x_1 + \ldots + a_p x_p = \sum_i a_i x_i$$

A média, por sua vez, pode ser representada por:

$$\mu_y = \sum_i a_i \mu_{xi}$$

finalmente, no caso da variância:

$$\sigma_y^2 = \Sigma a_i^2 x_i^2$$

Isto pode ser aplicado a variáveis correlacionadas ou não. Um caso bastante interessante a que podemos aplicar este tratamento é o nosso problema de medição de alturas das pessoas. Imagine que em vez de medirmos 140 pessoas, tivéssemos medido 10 de cada vez e tirássemos a média e a variância. Repetindo isso 140 vezes (140 amostras de 10, perfeitamente aleatórias), teríamos 140 valores de médias e 140 variâncias. Como será que ficaria a distribuição destes valores médios resultantes? Podemos aplicar neste caso a combinação linear de variáveis aleatórias, tratando-as como independentes.

A média dos N valores pode ser considerada como um caso particular de combinação linear, em que todos os coeficientes a_i são iguais a 1/N:

$$\bar{x} = \frac{1}{N}x_1 + \frac{1}{N}x_2 + \ldots + \frac{1}{N}x_N$$

Vamos comparar esta expressão com a expressão original utilizada para calcular a média de uma amostra de N observações:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Verificamos que é a mesma expressão, somente na de baixo 1/N está em evidência! Ou seja, quando fazemos a distribuição de médias, não modificamos o valor médio: ambas as distribuições se situam em torno do mesmo valor médio.

A variância da média pode ser calculada pela expressão abaixo, considerando as variáveis todas independentes, como é o caso (lembre-se de que os coeficientes a_i são iguais a 1/N):

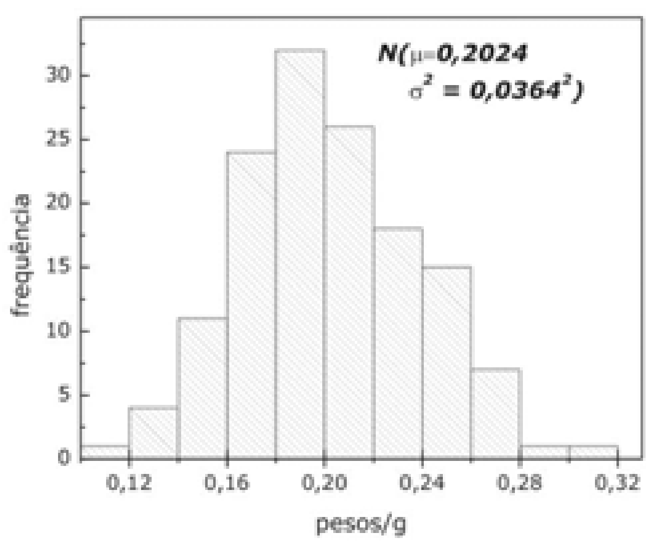
$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{N}$$

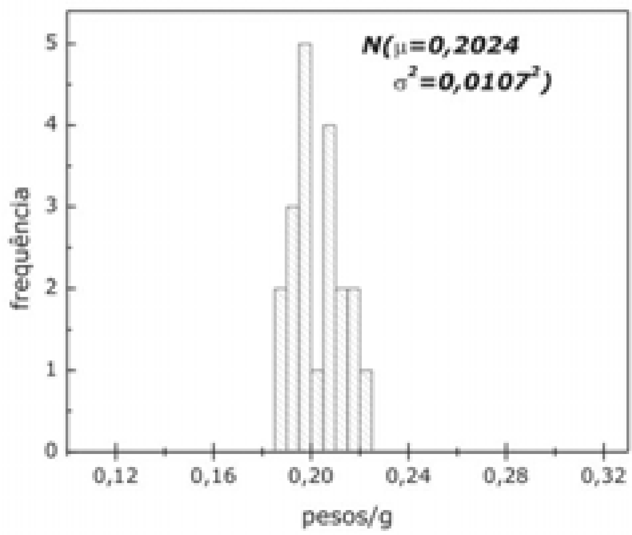
Esta expressão é interessante, pois mostra que a variância da distribuição de médias é igual à variância da amostra de N valores dividida por N, ou seja, é bem menor e será cada vez menor quanto maior for N. Isso também pode ser compreendido se considerarmos que quando fazemos cada média das 140 amostras de 10 observações, este procedimento mascara os limites inferior e superior, permitindo com que só visualizemos a média. Depois, ao fazer a média das médias, estes limites já se perderam e a nova variância só representa a faixa em que estas médias apareceram.

Exemplo

Suponha que pegamos 1 kg de milho de pipoca e pesamos um único de 140 grãos. Depois, pesamos 140 amostras contendo 7 grãos cada uma. As figuras abaixo apresentam, respectivamente, a distribuição de valores obtidos para a amostra de 140 grãos e para as 140 amostras de 7 grãos.

Note que o valor médio do peso é m = 0,2024 g em ambos os casos, mas que a variância é menor para a distribuição de médias.





CONCLUSÃO

Como conclusão desta aula, podemos destacar que, com base no método de combinação linear de variáveis aleatórias, pode-se obter a distribuição de médias amostrais e comparar os resultados em termos de parâmetros populacionais com aquele obtido a partir de uma única amostra numerosa. Todavia, como ambas abordagens se referem à mesma população (pacote de 1 kg de milho), a distribuição obtida a partir de uma única amostra numerosa e mais representativa da população. Isto porque na distribuição de médias as caudas em ambos os lados da curva (próximos aos limites inferior e superior da distribuição) tendem a não aparecer, pois entram no cálculo das médias.

RESUMO

Nesta aula, introduzimos o conceito de combinação linear, como uma operação matemática em que duas variáveis x_1 e x_2 são somadas ou subtraídas após serem multiplicadas por coeficientes ou constantes específicas (a_1 e a_2), dando origem a uma nova variável y. Em seguida, verificamos quais as expressões para o cálculo do valor médio desta nova variável aleatória e da variância. A importância disto para nossos estudos é que um dos resultados mais interessantes que observamos foi que a média da combinação linear é a combinação linear das médias. Com isso, pudemos aplicar estes resultados para nossas distribuições de variáveis aleatórias independentes (apesar do fato de que mesmo variáveis correlacionadas podem ser sujeitas à combinação linear). Finalmente, comparamos o efeito de se medir uma amostra relativamente numerosa de uma população, para se obter a média e a variância, com o procedimento análogo de se medir muitas amostras relativamente pequenas, tirar a média e a variância de cada uma e depois fazer a distribuição destas médias. Para isto, vamos tratar essas médias como variáveis independentes e aplicar a combinação linear de variáveis aleatórias. Como já foi mencionado que a média das combinações lineares é igual à combinação linear das médias, quando fazemos a média dos valores médios das N populações "pequenas", o valor médio observado é o mesmo observado para uma única amostra numerosa, mas a variância é bem menor. Se observarmos a expressão para a variância das combinações lineares, vemos que é a variância da amostra de N valores dividida por N. Isso também pode ser compreendido se considerarmos que quando fazemos cada média das N amostras de poucas observações, este procedimento mascara os limites inferior e superior, e permite-nos visualizar somente a média. Depois, ao fazer a média das médias, esses limites já se perderam e a nova variância só representa a faixa em que estas médias apareceram.

COMBINAÇÃO LINEAR

As combinações lineares são úteis em várias situações da química, sendo usadas, por exemplo, na representação de orbitais moleculares como combinações lineares de orbitais atômicos. Este método é conhecido com o método LCAO (Combinação Linear de Orbitais Atômicos), que parte da hipótese de que uma ligação química entre dois átomos (A e B) resulta da sobreposição de orbitais atômicos F_A e F_B, dando origem a um orbital molecular chamado de y e que é ocupado por dois elétrons, um de cada átomo. Representa-se a combinação linear da seguinte forma:

$$y = F_A + F_B$$

PRÓXIMA AULA

Na próxima aula, veremos que ao assumirmos explicitamente que estamos diante de uma distribuição normal, o que não fizemos até agora, os parâmetros amostrais seguem distribuições específicas, possibilitando-nos obter intervalos de confiança.

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.