6

aula

TEOREMA DO LIMITE CENTRAL E INTERVALO DE CONFIANÇA

META

Explorar o teorema do limite central e o intervalo de confiança.

OBJETIVOS

Ao final desta aula, o aluno deverá:

aplicar o teorema do limite central e suas conseqüências para a ocorrência de distribuição normal;

compreender por que a suposição de que muitas distribuições seguem o padrão normal em geral é válida;

calcular intervalos de confiança da média, utilizando para isto a tabela de variáveis da distribuição normal padronizada.

PRÉ-REQUISITOS

O aluno deverá conhecer características da distribuição normal e uso no cálculo de probabilidades de ocorrência de valores.

(Fonte: http:// portal.mec.gov.br).

INTRODUÇÃO

Há algumas aulas, começamos a estudar a distribuição normal e vimos que ela recebe este nome por que consegue descrever com sucesso o comportamento de um número realmente expressivo de sistemas. Agora chegou a hora de pensarmos um pouco nisso, pois, será que é possível tantas coisas diferentes terem comportamento semelhante? Na realidade, vamos ver que isso é uma conseqüência do teorema do limite central, um dos teoremas mais importantes da estatística. Ele nos diz que se a flutuação total de uma certa variável aleatória depender da soma das flutuações de muitas variáveis independentes, todas com peso mais ou menos igual, então a distribuição da variável original irá tender à normalidade. Vamos desenvolver esta idéia a seguir.

(Fonte: http://www.clickchips.com.br).

DISTRIBUIÇÃO NORMAL

Inicialmente, caro aluno ou querida aluna, podemos provocar o seu raciocínio sugerindo a seguinte indagação: Por que funciona em tantos casos?

1. Muitas técnicas estatísticas comumente usadas são robustas ou insensíveis a desvio da normalidade, ou seja, não há testes realmente eficientes para se avaliar desvios da normalidade! Mas isso não é um problema sério? Não, porque há o teorema que explicitamos abaixo.
2. Efeito do teorema do limite central, o qual mostra que distribuição de flutuações como resultado de muitas variáveis independentes e de mesma importância, faz com que a sua distribuição tenderá para a normalidade.

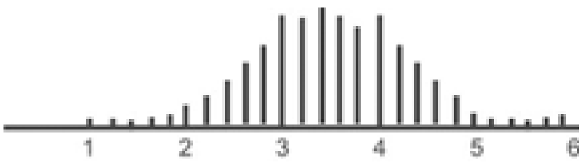
Podemos entender isso com mais clareza através do exemplo clássico dos dados. Se jogarmos N vezes um único dado e este for um dado "honesto", ou seja, não viciado e cuja probabilidade de ocorrência dos 6 lados seja exatamente a mesma, todos os lados sairão praticamente o mesmo número de vezes. Observaremos a seguinte distribuição:

nenhuma das N observações é mais importante que as outras.

(a) Um dado

Por outro lado, podemos jogar N vezes cinco dados ao em vez de um. Depois, atribuindo valores de 1 a 6 aos lados, podemos tirar a média dos valores que saíram em cada jogada de cinco dados. Neste caso, como o valor que saiu para cada dado é independente dos demais e cada um deles contribui com o mesmo peso para a média, pois são igualmente prováveis, a distribuição de médias obedece a uma distribuição normal, pois atende ao teorema do valor central. Neste caso:

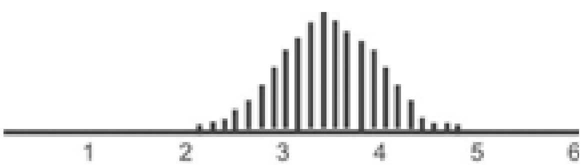
cada dado contribui com o mesmo peso para o resultado final:



(b) Cinco dados

Observando agora o que acontece com 10 dados, o resultado é semelhante, porém a observação de uma distribuição normal é bem mais pronunciada.

O valor para um certo dado não afeta os valores observados nos outros.



(c) Dez dados

Portanto, apesar de o número de vezes que cada lado sai ser praticamente igual, quando calculamos a média entre todos os valores que saíram, acabamos caindo em uma distribuição normal. O valor da média, neste caso, ficou em 3,5, que é o resultado da soma de todos os lados (1+2+3+4+5+6) dividido pelo número de lados. Médias mais baixas serão resultantes de jogadas em que predominaram lados com valores baixos, e vice-versa para médias mais altas.

A distribuição normal nada mais é do que um modelo, que supomos amplamente válido, e usamos para fazer inferências sobre parâmetros populacionais. Isto significa que podemos usar estimativas (de média e de variância, por exemplo) feitas a partir de uma amostra de um certo tamanho e usar os dados obtidos para inferir quais seriam os parâmetros para a população inteira.

Neste contexto, quanto mais o tamanho da nossa amostra se aproxima do tamanho da população, mais segura é a estimativa feita. Uma medida direta da precisão de uma estimativa é o chamado intervalo de confiança da média, e quanto menor o intervalo de confiança mais precisa é a estimativa. Vamos ver a seguir como calcular este parâmetro.

COMO CALCULAR UM INTERVALO DE CONFIANÇA PARA A MÉDIA?

Um caso extremo aqui é uma amostra com o menor tamanho possível, ou seja, a partir de uma única observação. Podemos calcular o intervalo de confiança para a média através da seguinte equação:

$$x_i - z\sigma < \mu < x_i + z\sigma$$

em que:

x_i: observação única

z: variável padronizada (lembre-se de que estas variáveis estão associadas às probabilidades de ocorrência de observações aleatórias em determinadas faixas de valores)

ó: desvio padrão populacional: se não conhecemos a média populacional, cujo intervalo de confiança queremos determinar temos que saber o desvio padrão, senão é impossível estimar intervalos para ambos ao mesmo tempo. Mas, na realidade, também não sabemos seu valor, então acabamos assumindo que uma boa aproximação é o desvio padrão obtido para uma amostra relativamente grande.

Exemplo

Lembre-se dos nossos dados a respeito das alturas das pessoas. Medimos 140 pessoas e obtivemos uma média de 1,70 m e um desvio padrão amostral de ó = 0,03641. Suponha que temos para a população um valor de desvio padrão também de ó = 0,03641, só para podermos resolver o problema, considerando que nossa amostra foi bastante numerosa a ponto de termos obtido uma excelente estimativa. Fazemos agora uma única observação de x_i = 1,64 m e queremos determinar o intervalo em que se encontra a média, com 95 % de confiança. Ou seja, calcularemos o intervalo de confiança da média com 95 % (como atividade, confira que o valor de z com 95 % de confiança é 1,96). Utilizando a equação acima, obtemos:

1,64 -1,96ó < m < 1,64 +1,96ó

utilizando o valor de ó = 0,03641, obtemos:

1,57 <   < 1,71

ou seja, por estes valores podemos dizer que a média deve estar entre 1,57 m e 1,71 m. De fato, pode parecer que é uma faixa muito ampla, mas, considerando-se que esta estimativa foi realizada a partir de uma única medida, ela é relativamente bem sucedida. No caso, a média amostral obtida a partir de 140 valores foi de 1,70 m. Sendo assim, a estimativa do intervalo de confiança da média populacional a partir de uma amostra única é possível, apesar do fato de cada medida ser independente das demais, ou seja, o valor obtido com uma medida não nos permite saber nada a respeito de uma medida seguinte.

ATIVIDADES

Calcule o intervalo de confiança da média com 60 % e 99,8 % de confiança a partir de observações únicas com os seguintes valores (use o mesmo valor de desvio padrão do exemplo):

a) 1,61 m

b) 1,68 m

c) 1, 78 m

Para resolver, siga o procedimento que foi detalhado no desenvolvimento.

CONCLUSÃO

A partir do que aprendemos nesta aula, podemos concluir que sempre que um conjunto de dados está sujeito a diversas variáveis independentes e com importâncias de mesma magnitude, as médias do conjunto de dados original tendem a seguir uma distribuição normal (isto é o que chamamos de teorema do limite central). Este é o fator determinante da grande aplicabilidade da distribuição normal. Concluímos também que a partir de uma única medida escolhida aleatoriamente dentro de uma população, podemos estimar o valor da média populacional, utilizando o modelo de distribuição normal.

RESUMO

Nesta aula, aprendemos as principais razões para que a distribuição normal descreva de modo satisfatório o comportamento dos mais variados tipos de conjuntos de dados. Em primeiro lugar, isto é decorrência do Teorema do Limite Central, um dos mais importantes da estatística. Segundo este teorema, se a flutuação total de uma certa variável aleatória depender da soma das flutuações de muitas variáveis independentes, todas com peso mais ou menos igual, então a distribuição da variável original irá tender à normalidade. Além disto, muitas técnicas estatísticas comumente usadas são robustas ou insensíveis a desvio da normalidade, ou seja, não há testes realmente eficientes para se avaliar desvios da normalidade! Mas isso não é um problema sério? Não, porque justamente por isto o teorema do limite central se aplica a todos estes casos. Isto pode ser compreendido por analogia ao comportamento dos dados (de jogos). Cada lado de um dado tem igual probabilidade de ocorrer, em uma jogada, ou seja, todos têm pesos iguais. Se jogarmos o dado 6 vezes, a tendência é que cada lado saia uma vez e não teremos uma distribuição normal. Por outro lado, podemos tomar vários dados (10, por exemplo) e fazer N jogadas, jogando os 10 dados a cada jogada. Podemos somar a cada jogada os números correspondentes aos lados que saíram para cada dado e tirar a média, representando em seguida a distribuição de médias obtidas. Como cada lado tem probabilidade igual de sair, a distribuição de médias será uma distribuição normal centrada em 3,5, que é o resultado da soma de todos os lados (1+2+3+4+5+6) dividido pelo número de lados. Médias mais baixas serão resultantes de jogadas em que predominaram lados com valores baixos, e vice-versa para médias mais altas. Finalmente calculamos intervalos de confiança para a média, a partir de uma dada observação, em um dado nível de confiança.

PRÓXIMA AULA

Na próxima aula, vamos estudar como é possível descobriros que dois conjuntos de dados estão relacionados (por exemplo, um conjunto de pesos de alguma fruta e outro conjunto com os seus volumes: será que podemos supor que quanto maior o peso, maior também será o volume?). Vamos também tomar contato com outro tipo de distribuição derivada da distribuição normal, mas que nos exime de ter dados sobre a variância populacional, ou seja, podemos trabalhar apenas com dados amostrais.

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.