WIKIPÉDIA

Quantil

Origem: Wikipédia, a enciclopédia livre.

Quantis são pontos estabelecidos em intervalos regulares a partir da função distribuição acumulada (FDA), de uma variável aleatória. Os quantis dividem os dados ordenados em *q* subconjuntos de dados de dimensão essencialmente igual. Dessa forma dão origem a *q*-Quantis; os quantis são estabelecidos a partir de pontos de corte que determinam as fronteiras entre os subconjuntos consecutivos. Visto de outra forma, o *k*-ésimo *q*-quantil é o valor *x* tal que a probabilidade de um evento da variável aleatória será inferior *x* é no máximo *k*/*q* e a probabilidade da variável aleatória ser superior ou igual a *x* é pelo menos *(q-k)*/*q*. Há *q - 1* quantis, sendo *k* um inteiro satisfazendo 0 <*k* <*q*.

Índice

Quantis específicos

Alguns quantis têm nomes especiais:

- Os 100-quantis são chamados percentis → P
- Os 12-quantis são chamados duo-deciles → Dd
- Os 10-quantis são chamados decis → D
- Os 5-quantis são chamados quintis → QU
- Os 4-quantis são chamados quartis → Q
- OS 3-quantis são chamados tercis → T

De um modo mais geral, pode-se considerar a função quantil para qualquer distribuição. Esta é definida por variáveis reais, entre zero e um, e, matematicamente, é a inversa da função distribuição acumulada.

Quantis de uma população

Para uma população de valores discretos ou para uma densidade populacional contínua o *k*-ésimo *q*-quantil é o valor onde a função distribuição acumulada cruza *k*/*q*. Isto é, *x* é o *k*-ésimo *q*-quantil de uma variável *X* se

$\Pr[X < x] \leq k/q$ (ou, de forma equivalente, $\Pr[X > x] \geq 1 - k/q$)

e

$\Pr[X \leq x] \geq k/q$ (ou, de forma equivalente, $\Pr[X \geq x] \leq 1 - k/q$).

Para uma população finita de *N* valores indexados de 1 ,..., a *N* de menor para maior, o *k*-ésimo *q*-quantil desta população pode ser computado através do valor de $I_p = N\frac{k}{q}$. Se I_p não for um inteiro, em seguida, arredonda-se para o próximo inteiro para obter o índice apropriado; o valor correspondente é o *k*-ésimo *q*-quantil. Por outro lado, se I_p é um número inteiro, então qualquer valor correspondente a esse índice até o valor correspondente ao próximo pode ser tomado como o quantil, de maneira convencional (embora arbitrária) se considera a média dos dois valores (ver #estimativa dos Quantis).

Se, em vez de usar inteiros *k* e *q*, o "*p*-quantil" é baseado em um número real *p* com 0 <*p*<1, então, *p* substitui *k*/*q* nas fórmulas acima. Alguns programas como algumas folhas de cálculo consideram o mínimo e o máximo como o 0º e 100º percentil, respectivamente; no entanto, essa terminologia é uma extensão além das tradicionais definições estatísticas .

Quantis de uma amostra

A abordagem é diferente para uma amostra finita selecionada aleatoriamente a partir da população. o *k*-ésimo *q*-quantil de uma amostra pode ser estimado através do valor de $I_s = (N+1)\frac{k}{q}$. Se I_s for um inteiro, então é o índice do valor a ser considrado o *k*-ésimo *q*-quantil da amostra. Por outro lado, se I_s não é um inteiro, mas está entre 1 e *N*, então, normalmente é usada uma média (ponderada) dos valores observados para os índices inteiros adjacentes.

Quando I_s é menor que 1 ou maior que *N* o *k*-ésimo *q*-quantil da amostra não é normalmente definido.

Se, em vez de usar inteiros *k* e *q*, o "*p*-quantil" é baseado em um número real *p* com 0 <*p*<1 e, então, *p* substitui *k*/*q* nas fórmulas acima.

Esta abordagem de estimativa está intimamente relacionado com o resultado de estatísticas de ordem. Especificamente, o I_s-ésimo menor de *N* valores escolhidos independentemente da distribuição uniforme entre [0,1] é uma variável aleatória com média $p = I_s/(N+1)$.

Exemplos

Considere uma **população** de 10 dados {3, 6, 7, 8, 8, 10, 13, 15, 16, 20}.

- O primeiro quartil é determinado por 10*(1/4) = 2.5, que pode ser arredondado a 3, isto é, 3 é o indice na população (classificada de menor a maior valor), no qual aproximadamente 1/4 dos valores são menores do que esse terceiro valor, que, neste caso, é 7.
- O segundo quartil (igual à mediana) é determinado por 10*(2/4) = 5, um valor inteiro, mas como o número de observações (10) é um número par, então a média do quinto e sexto valor será considerado; isto é, (8+10)/2 = 9, embora qualquer valor entre 8 e 10 poderia ser considerado como a mediana. Se o número de observações é impar, o valor da mediana (ou segundo quartil) corresponde ao índice (número_de_observações + 1)/2.
Então, no caso deste exemplo, se acrescentarmos o valor 9, fazendo 11 observações, então (11 + 1)/2 = 6. O que significaria que o sexto valor, ou seja 9, seria o segundo quartil,

onde a metade dos valores seriam maiores que este valor (maiores que 9, correspondente ao índice 6 em 11), é a outra metade seriam menores que dito índice.
- O terceiro quartil para a população original (sem o 9) é determinado por 10*(3/4) = 7.5, que pode ser arredondado a 8, índice que corresponde ao número 15.

A motivação para este método é que o primeiro quartil deve dividir os dados entre o quarto inferior e os tres quartos superiores. Idealmente, isso significaria 2.5 da amostra estão abaixo do primeiro quartil e 7.5 são superiores, neste caso significa que um terço da amostra de dados está "dividida em duas", tornando a terceira parte da amostra com o primeiro e segundo quartos.

Agora considere uma **amostra** dos mesmos 10 valores {3, 6, 7, 8, 8, 10, 13, 15, 16, 20}, que são retirados aleatoriamente de alguma população desconhecida.

- O primeiro quartil pode ser estimado através do (10+1)*(1/4) = 2.75, que se encontra entre 2 e 3, porém está mais próximo a este último. Uma estimativa para o primeiro quartil é a média ponderada do segundo e terceiro menores valores, que são de 6 e 7, neste caso. Concretamente, a estimativa é de $0,25(6)+0,75(7)=6,75$.
- O segundo quartil pode ser estimado através de (10+1)* (2/4) = 5.5, que se encontra entre 5 e 6. Uma estimativa é, assim, a média do quinto e sexto menores valores, $0,5(8)+0.5(10)=9$.
- O terceiro quartil pode ser estimado através de (10+1)*(3/4) = 8,25, que se encontra entre 8 e 9. Uma estimativa é, assim, a média ponderada do oitavo e nono menores valores, $0,75(15)+0,25(16)=15,25$.

Se tivesse tido também um valor 9 entre os valores 8 e 10, fazendo um total de 11 valores, os quartis teriam indices de $(N+1)\frac{k}{q}$ ou 3, 6 e 9, respectivamente. Assim, as estimativas quartis seriam os valores 7, 9 e 15, respectivamente. Note-se que estes valores particionam os restantes oito valores ordenados em quatro grupos de igual tamanho {3,6}, {8,8}, {10,13} e {16,20}.

Ver Também

- Estatística descritiva
- Quartil
- Percentil
- Q-Q plot
- Função quantil

Referências

- R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, 1980.