Explicando quantis

Mário O. de Menezes

18 de março de 2019

Quantis

Em Estatística e Probabilidade, *quantis* são pontos de corte que dividem a faixa de uma distribuição de probabilidade em intervalos contínuos com probabilidades iguais, ou que divide as observações em uma amostra da mesma maneira. Há um quantil a menos do que o número de grupos criados.

Assim, quartis são os três pontos de corte que dividirão um conjunto de dados em quatro grupos de mesmo tamanho. Quantis comuns tem nomes especiais: por exemplo, *quartil*, *decil* (criando 10 grupos), etc. Os grupos criados são chamados de metades, terços, quartos, etc., apesar de algumas vezes o termo quantil ser utilizado para os grupos criados, ao invés de se utilizar os pontos de corte.

Algumas dificuldades

Esta definição de quantis e percentis não é completamente satisfatória. Por exemplo, considere a seguinte sequência de seis valores:

$$2.13.12.70.84.33.5$$2.13.12.70.84.33.5

Qual é o quartil inferior destes valores?

Não há nenhum valor que tenha 25% destes números abaixo dele e 75% acima!

Para superarmos esta dificuldade, utilizamos uma definição de percentil/quartil que tem o mesmo espírito que a que fizemos inicialmente, mas que (necessariamente) a aplica de forma aproximada quando necessário.

Definindo Quantis

Nós definimos os quantis para o conjunto de valores:

$$2.13.12.70.84.33.5$$2.13.12.70.84.33.5

da seguinte forma.

Primeiro, ordenamos os valores:

$$0.82.12.73.13.54.3$$0.82.12.73.13.54.3

Associamos os valores ordenados com frações da amostra espaçadas igualmente de zero a um.

Fração da Amostra	0	0.2	0.4	0.6	0.8	1.0
Quantil	0.8	2.1	2.7	3.1	3.5	4.3

A ordem dos quantis de

$$0.82.12.73.13.54.3$$0.82.12.73.13.54.3

pode ser obtida por interpolação linear entre os valores da tabela.

A mediana corresponde a fração da amostra de 0.5. Este valor está na metade entre 0.4 e 0.6. Então, a mediana deve ser $0.5\times2.7+0.5\times3.1=2.9$0.5×2.7+0.5×3.1=2.9

O quartil inferior corresponde a uma fração da amostra de 0.25. Este valor está um quarto do caminho entre 0.2 e 0.4. O quartil inferior pode ser calculado então como $0.75\times2.1+0.25\times2.7=2.25$

Caso geral

Dado um conjunto de valores $x_1,x_2,...,x_n$ nós podemos definir o quantis para qualquer fração p como segue:
Ordene os valores:

$$x_{(1)}\leq x_{(2)}\leq\cdots\leq x_{(n)}$$

Os valores $x_{(1)},....,x_{(n)}$ são chamados de *estatística de ordem* da amostra original.
Tomamos a estatística de ordem como sendo os quantis que correspondem às frações:

$$p_i=\frac{i-1}{n-1},(i=1,...,n)$$

A função quantil

Em geral, para definir o quantil que corresponde à fração p, nós utilizamos interpolação linear entre os dois p_i mais próximos.
Se p está a uma fração f da distância entre p a p_{i+1}, nós definimos p-ésimo quantil como sendo:

$$Q(p)=(1-f)Q(p_i)+fQ(p_{i+1})$$

Como casos especiais, definimos a mediana e os quartis como:

Quantil	**Função Quantil**
Mediana	$Q(0.5)$
Quartil inferior	$Q(0.25)$
Quartil superior	$Q(0.75)$

A função Q definida desta forma é chamada a *Função Quantil*

Exemplo 1

Usando esta fórmula para os dados acima, temos:

Fração da Amostra	**0**	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**
Quantil	0.8	2.1	2.7	3.1	3.5	4.3
p_i	p_1	p_2	p_3	p_4	p_5	p_6

Queremos calcular a Mediana, isto é, $Q(0.5)$.
Então utilizaremos:

- $i=3$
- $p_3=0.4$ e $Q(p_3)=2.7$;
- $p_4=0.6$ e $Q(p_4)=3.1$

Para a mediana $Q(0.5)$, ela está a uma fração $f=0.5$ de p_3 e p_4

Assim, o cálculo fica:

$$Q(0.5)=(1-0.5)Q(p_3)+0.5Q(p_4)=0.5\times2.7+0.5\times3.1=2.9$$

Ou seja, estamos fazendo uma interpolação linear entre os dois quantis conhecidos para encontrarmos o desconhecido.

Exemplo 2

Este faremos tudo no **R**; os passos são os seguintes:

- criamos uma sequência de 16 números aleatórios de 1 a 100;
- em seguida, obtenho a sequencia ordenada;
- depois geramos os índices dos elementos da sequência;
- e por último, gero os quantis desta sequência.

```
set.seed(123)
n <- 16
sequencia <- sample.int(100, size = n)
sequencia.ord <- sort(sequencia)
i = seq(1,n)
pi = (i - 1)/(n - 1)
kable(data.frame(Indice = i, Seq = sequencia, Seq.ord = sequencia.ord,  Quantis = round(pi*100,2)))
```

Indice	Seq	Seq.ord	Quantis
1	31	9	0.00
2	79	14	6.67
3	51	25	13.33
4	14	26	20.00
5	67	31	26.67
6	42	42	33.33
7	50	43	40.00
8	43	50	46.67
9	97	51	53.33

Indice	Seq	Seq.ord	Quantis
10	25	57	60.00
11	90	67	66.67
12	69	69	73.33
13	57	72	80.00
14	9	79	86.67
15	72	90	93.33
16	26	97	100.00

Queremos calcular os quartis: 25%, 50% e 75%. Então vamos pegar os $p_i$pis que estão imediatamente antes e depois.
Se queremos os quartis, queremos dividir a distribuição em 4 partes iguais; assim, cada parte representará 25% do todo.

Primeiro quartil (25%)

O primeiro quartil portanto estará na posição 1/4 do número de elementos que temos, arredondado para o inteiro mais próximo.

```
i = round(n * (1/4)); i # qual quantil da amostra
## [1] 4
c(pi[i], pi[i + 1]) #  quantis da amostra
## [1] 0.2000000 0.2666667
c(sequencia.ord[i], sequencia.ord[i+1]) # elementos correspondentes aos quantis
## [1] 26 31
q = 0.25  # quartil desejado
f = (q - pi[i])/(pi[i+1] - pi[i]);f
## [1] 0.75
# utilizando a formula de Q(p)
Q25 = (1 - f) * sequencia.ord[i] + f * sequencia.ord[i + 1]; Q25
## [1] 29.75
```

Para o segundo quartil (50%)

Para a mediana, isto é, o segundo quartil 50%, temos:

```
# se o comprimento da sequencia for par
if (all.equal(n%%2, 0)) {
```

```r
  # n is even
  i = n/2
} else {
  i = round(n/2,0)
```

```r
}
c(pi[i], pi[i + 1])
```

```
## [1] 0.4666667 0.5333333
```

```r
c(sequencia.ord[i], sequencia.ord[i+1])
```

```
## [1] 50 51
```

```r
q = 0.5
f = (q - pi[i])/(pi[i+1] - pi[i]);f
```

```
## [1] 0.5
```

```r
Q50 = (1 - f) * sequencia.ord[i] + f * sequencia.ord[i + 1]; Q50
```

```
## [1] 50.5
```

Para conferir, fazendo pelo **R**

```r
quantile(sequencia)
```

```
##    0%   25%   50%   75%  100% 
##  9.00 29.75 50.50 69.75 97.00
```