

Aula 1

Manoel Galdino

2023-05-05

Fundamentals of Regreession

Revisão de estatística e probabilidade

A média de um conjunto de valores é dada pela soma dos valores, dividido pelo número de observações. Matematicamente: $média = \sum_{i=1}^n x_i/n$ para observações $x_1, x_2, x_3, \dots, x_n$

Em geral, as observações são uma amostra, e falamos de média amostral, \bar{x} . Ou seja:

$$\bar{x} = \sum_{i=1}^n x_i/n$$

Exercício 1. Vamos calcular, no R, a média das seguintes amostras: a) $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ b) $\{5, 5, 5, 5, 5, 5, 5\}$ c) $\{1, 3, 5, 7, 9, 11\}$ d) $\{-5, -4, -3, -2, -1, 1, 2, 3, 4, 5\}$

```
x <- c(1,2,3,4,5,6,7,8,9,10)
(media_x <- sum(x)/length(x))
```

```
## [1] 5.5
```

```
x <- c(5,5,5,5,5,5,5)
(media_x <- sum(x)/length(x))
```

```
## [1] 5
```

```
x <- c(1,3,5,7,9,11)
(media_x <- sum(x)/length(x))
```

```
## [1] 6
```

```
x <- c(-5,-4,-3,-2,-1,1,2,3,4,5)
(media_x <- sum(x)/length(x))
```

```
## [1] 0
```

```
# ou podemos simplesmente usar mean(x)
mean(x)
```

```
## [1] 0
```

A esperança de uma variável aleatória discreta X , cuja probabilidade de massa de $x \in X$ é dada por $p(x)$, é definida por: $\sum(x * p(x))$

A esperança de uma v.a. contínua X , cuja densidade é $f(x)$, é definida por: $\int f(x) * x dx$.

Similarmente, a espereança de uma função $h(X)$ é dada por $\int f(x) * h(x) dx$ e analogamente para o caso discreto.

A variância de uma variável aleatória X é dada por:

Definição 1. $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

A Covariância de duas v.a. X e Y é definida como: $Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X]) * (Y - \mathbb{E}[Y])]$.

Notem que $Cov(X, Y) = Var(X)$.

A covariância é positiva quando ambos X e Y tendem a ter valores acima (ou abaixo) de sua média simultaneamente, enquanto ela é negativa quando uma v.a. tende a ter valores acima da sua média e a outra abaixo.

Algebra com Esperança, Variância e Covariância

Sejam a e b constantes.

1. Linearidade da Esperança $\mathbb{E}[aX + bY] = \mathbb{E}[aX] + \mathbb{E}[bY] = a * \mathbb{E}[X] + b * \mathbb{E}[Y]$

Exercício: verifique, com exemplos, que isso é verdade.

2. Identidade da variância $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$ A prova será demonstrada mais adiante.

3. Identidade da Covariância

$Cov(X, Y) = \mathbb{E}[X * Y] - \mathbb{E}[X] * \mathbb{E}[Y] = \mathbb{E}[(X - \mathbb{E}[X]) * (Y - \mathbb{E}[Y])]$ Exercício para o leitor. Prove que isso é verdade.

4. Covariância é simétrica

$$Cov(X, Y) = Cov(Y, X)$$

5. Variância não é linear $Var(a * X + b) = a^2 * Var(x)$

6. Covariância não é linear

$Cov(a * X + b, Y) = a * Cov(X, Y)$ # Prova da identidade da variância Vamos mostrar que $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}^2[X]$

1. Começamos expandindo o quadrado da esperança: $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(X - \mathbb{E}[X]) * (X - \mathbb{E}[X])]$.
2. Aplicando a regra do quadrado, temos: $\mathbb{E}[(X - \mathbb{E}[X]) * (X - \mathbb{E}[X])] = \mathbb{E}[(X^2 - 2 * \mathbb{E}[X] * X + \mathbb{E}[X]^2)]$
3. Pela propriedade da esperança, sabemos que, sejam A e B duas v.a. independentes, então $\mathbb{E}[A + B] = \mathbb{E}[A] + \mathbb{E}[B]$. Então:

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[2 * \mathbb{E}[X] * X] + \mathbb{E}[\mathbb{E}[X]^2]$$

4. Outra propriedade da esperança é que, seja a uma constante e X uma v.a., então $\mathbb{E}[a * X] = a * \mathbb{E}[X]$.

$$Var(X) = \mathbb{E}[X^2] - 2 * \mathbb{E}[\mathbb{E}[X] * X] + \mathbb{E}[\mathbb{E}[X]^2]$$

5. Nós sabemos que $\mathbb{E}[X]$ é uma constante (é uma média da v.a.). E a média de uma constante é a própria constante. Portanto, $\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]$. E usaremos também que $\mathbb{E}[a * X] = a * \mathbb{E}[X]$ e, por fim, o fato de que uma constante ao quadrado é em si uma constante.

$$Var(X) = \mathbb{E}[X^2] - 2 * \mathbb{E}[X] * \mathbb{E}[\mathbb{E}[X]] + \mathbb{E}[X]^2$$

$$Var(X) = \mathbb{E}[X^2] - 2 * \mathbb{E}[X] * \mathbb{E}[X] + \mathbb{E}[X]^2$$

$$Var(X) = \mathbb{E}[X^2] - 2 * \mathbb{E}[X]^2 + \mathbb{E}[X]^2$$

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Como Queríamos Demonstrar (CQD).

Erro Quadrático Médio

ref <https://stats.stackexchange.com/questions/520286/how-can-i-prove-mathematically-that-the-mean-of-a-distribution-is-the-measure-th>

<https://statproofbook.github.io/P/mse-bnv.html>

Seja m um chute (ou, dito de maneira mais bonita, uma estimativa, por exemplo, a média. Ou ainda, uma previsão) de uma variável aleatória Y . Seja o erro, $e = m - Y$ a diferença entre o chute e as realizações da variável aleatória.

Definição 1.2. Chama-se Erro Quadrático Médio (EQM) de m o valor:

$$EQM(m) = \mathbb{E}[e^2] = \mathbb{E}[(m - Y)^2]$$

Podemos reescrever a equação acima do seguinte modo:

$\mathbb{E}[(m - Y)^2] = \mathbb{E}[(m + \mathbb{E}[m] - \mathbb{E}[m] - Y)^2]$. Aqui eu somei e adicionei o mesmo valor, o que não altera a equação.

Reordenando os termos, temos:

$$\mathbb{E}[(m - \mathbb{E}[m] + \mathbb{E}[m] - Y)^2]$$

Veja que se eu chamar $m - \mathbb{E}[m] = a$ e $\mathbb{E}[m] - Y = b$, tenho:

$\mathbb{E}[(a + b)^2]$. Aplicando a regra do quadrado:

$$\mathbb{E}[a^2 + 2 * a * b + b^2]$$

Substituindo a e b , temos: $\mathbb{E}[(m - \mathbb{E}[m])^2 + 2 * (m - \mathbb{E}[m]) * (\mathbb{E}[m] - Y) + (\mathbb{E}[m] - Y)^2]$

Sabemos que $\mathbb{E}[m] = m$, pois m é constante. Então, $(m - \mathbb{E}[m]) = 0$ e todo o termo multiplicado por 2 vai para zero. De forma que obtemos:

$$\mathbb{E}[(m - \mathbb{E}[m])^2 + (\mathbb{E}[m] - Y)^2]$$

Pela linearidade da esperança:

$$\mathbb{E}[(m - \mathbb{E}[m])^2] + \mathbb{E}[(\mathbb{E}[m] - Y)^2]$$

Note que $\mathbb{E}[(m - \mathbb{E}[m])^2] = Var(m)$ e $\mathbb{E}[(\mathbb{E}[m] - Y)^2] = bias(m)^2$

Portanto, podemos escrever o EQM como: $Var(m) + bias(m)s^2$