

# Correlação e Regressão Linear

Bioestatística em R

---

André M Ribeiro-dos-Santos

23 de 03, 2017

- Avaliar a associação entre medidas quantitativas.
- Reconhecer diferentes tipos de correlação.
- Ilustrar a relação entre medidas quantitativas.
- Reconhecer quando aplicar Pearson e Spearman.
- Conhecer principais transformações e quando aplicá-las.
- Modelar medidas através de uma regressão linear.
- Ilustrar a regressão e resíduos.
- Comparar diferentes regressões.

# Correlação

---

Em um estudo sobre diabetes, os pesquisadores observaram uma grande variação da sensibilidade à insulina entre os pacientes. Como trabalhos anteriores relacionaram essa variação com composição lipídica do tecido muscular. Foi medido a sensibilidade à insulina e composição de ácidos graxos de 10 pacientes.

### A variação da sensibilidade à insulina está relacionada a composição de ácidos graxos?

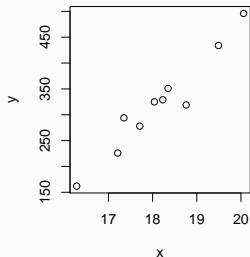
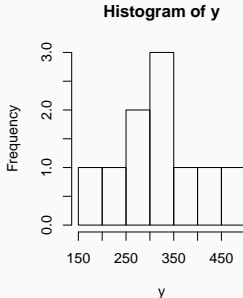
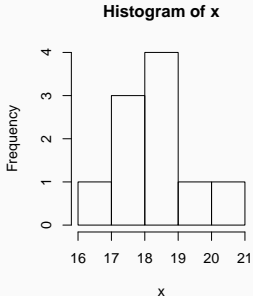
Insulin sensitivity	Fatty Acids (%)	Insulin sensitivity	Fatty Acids (%)
278	17.71	319	18.76
496	20.06	329	18.23
294	17.35	162	16.28
434	19.49	325	18.04
226	17.21	351	18.35

- As medidas em questão são categóricas ou quantitativas?
- Qual o tamanho da amostra?
- Qual a hipótese sendo avaliada?
- Qual a distribuição das medidas?

- As medidas em questão são categóricas ou quantitativas? **Ambas são quantitativas**
- Qual o tamanho da amostra? **10 pacientes**
- Qual a hipótese sendo avaliada? **As medidas são relacionadas.**

- Qual a distribuição das medidas? E como se relacionam?

```
> y <- c(278, 496, 294, 434, 226, 319, 329, 162, 325, 351)
> x <- c(17.71, 20.06, 17.35, 19.49, 17.21, 18.76,
+       18.23, 16.28, 18.04, 18.35)
>
> par(mfrow=c(1,3))
> hist(x); hist(y)
> plot(x, y)
```



Quando desejamos avaliar se a variação de uma medida afeta outra medida quantitativa, avaliamos a correlação linear das medidas com o coeficiente de correlação de Pearson ( $r$ ).

$$r = \frac{cov_{xy}}{s_x * s_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \hat{x}) * \sum *y - \hat{y})}$$

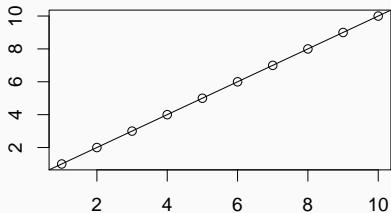
Uma propriedade interessante deste coeficiente é que  $r^2$  corresponde ao percentual da variabilidade em  $y$  explicada por  $x$  (ou vice-versa).



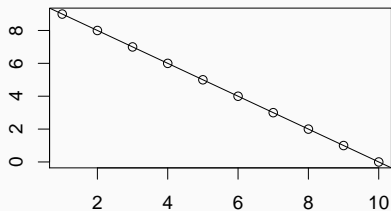
## Valores do coeficiente

O *coeficiente de correlação* ( $r$ ) assume valores entre -1 e 1, indicando uma correlação inversa em valores negativos, direta para valores positivos e zero quando não há correlação.

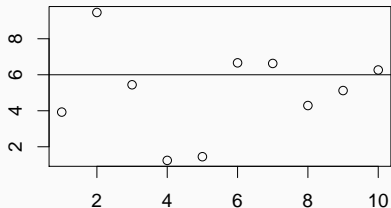
**$r = 1$**



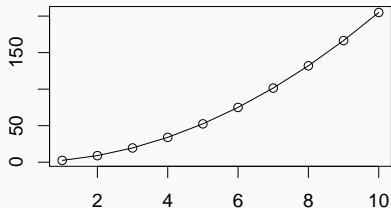
**$r = -1$**



**$r = 0$**



**non-linear**



```
> ?cor
> ## Correlation, Variance and Covariance (Matrices)
> ## Description:
> ##      'var', 'cov' and 'cor' compute the variance of 'x' and the
> ##      covariance or correlation of 'x' and 'y' if these are vectors. If
> ##      'x' and 'y' are matrices then the covariances (or correlations)
> ##      between the columns of 'x' and the columns of 'y' are computed.
> ## Usage:
> ##      var(x, y = NULL, na.rm = FALSE, use)
> ##      cov(x, y = NULL, use = "everything",
> ##          method = c("pearson", "kendall", "spearman"))
> ##      cor(x, y = NULL, use = "everything",
> ##          method = c("pearson", "kendall", "spearman"))
```

```
> cov(x, y) / sqrt(var(x) * var(y))
```

```
## [1] 0.9668936
```

```
> cor(x, y)
```

```
## [1] 0.9668936
```

```
> cor(x, y)^2
```

```
## [1] 0.9348833
```

Assumindo que  $r$  apresenta uma distribuição normal, podemos testar a significância da correlação observada.

$$H_0 : r = 0; \quad H_a : R \neq 0$$

$$t = \frac{r}{EP_r} = r \sqrt{\frac{n-2}{1-r^2}}$$

```
> ?cor.test
> ## Test for Association/Correlation Between Paired Samples
> ## Description:
> ##      Test for association between paired samples, using one of
> ##      Pearson's product moment correlation coefficient, Kendall's
> ##      tau or Spearman's rho.
> ## Usage:
> ##      cor.test(x, y,
> ##              alternative = c("two.sided", "less", "greater"),
> ##              method = c("pearson", "kendall", "spearman"),
> ##              exact = NULL, conf.level = 0.95,
> ##              continuity = FALSE, ...)
> ##      ## S3 method for class 'formula'
> ##      cor.test(formula, data, subset, na.action, ...)
```

```
> cor.test(x, y)
```

```
##
```

```
## Pearson's product-moment correlation
```

```
##
```

```
## data: x and y
```

```
## t = 10.717, df = 8, p-value = 5.05e-06
```

```
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## 0.8620949 0.9923783
```

```
## sample estimates:
```

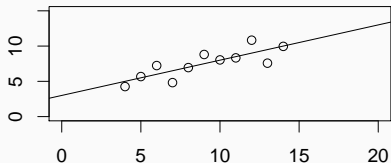
```
## cor
```

```
## 0.9668936
```

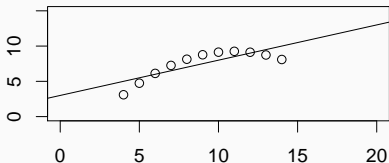
## Erros comuns

1. Correlação não implica em causa.
2. Focar no P-value, no lugar do coeficiente.
3. Assumir correlação sem plotar relação.

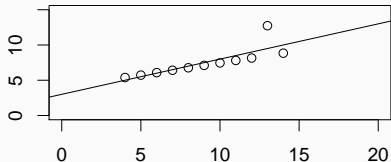
**Cor(x, y) = 0.816**



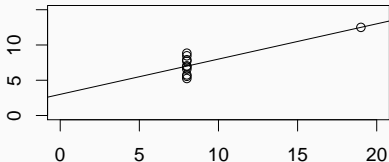
**Cor(x, y) = 0.816**



**Cor(x, y) = 0.816**



**Cor(x, y) = 0.817**







Parece que o jogo virou





## Regressão Linear

---

Imagine...

