Autocorrelação

Prof. Renzo Flores-Ortiz

Correlação (populacional)

- A **correlação populacional** mede a **intensidade** e a **direção** da **associação linear** entre duas variáveis aleatórias **na população**.
- O coeficiente de correlação populacional de Pearson, denotado por ρ (rho), é definido como:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

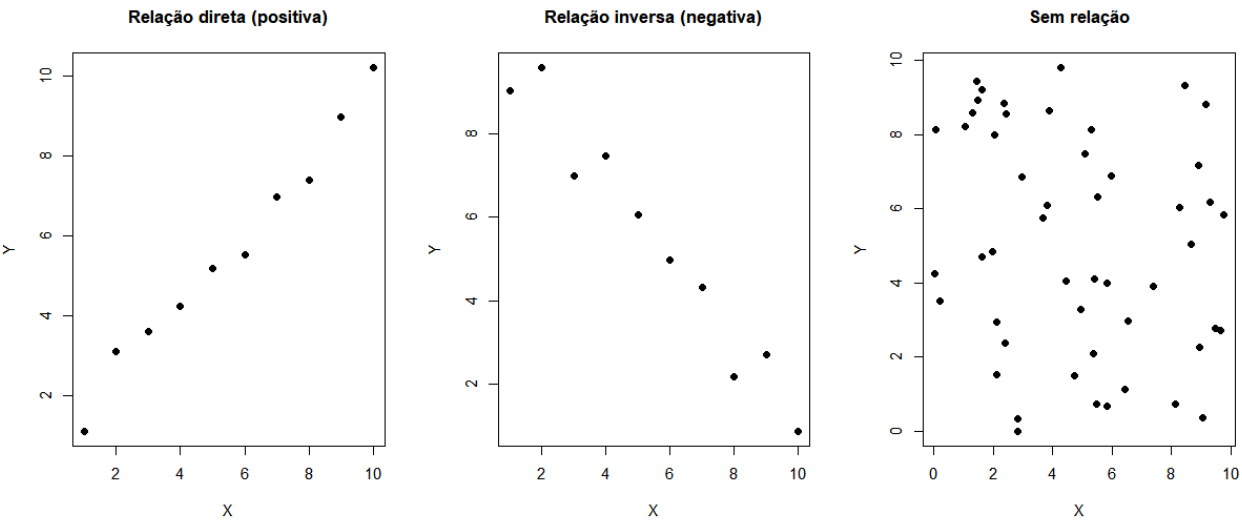
Onde

- $Cov(X,Y)$ é a covariância populacional entre X e Y.
- σ_X e σ_Y são os desvios padrão populacionais de X e Y.

O valor de ρ sempre está entre -1 e 1:

- ρ = 1: correlação perfeita positiva (associação linear crescente)
- ρ = −1: correlação perfeita negativa (associação linear decrescente)
- ρ = 0: ausência de associação linear

Tipos de relações entre variáveis



Correlação amostral

- A **correlação amostral** é o **estimador** de ρ, baseado em uma **amostra** extraída da população.
- Esse estimador é conhecido como **coeficiente de correlação amostral de Pearson**, denotado por r:

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Onde:

- (x_i, y_i) são as observações amostrais,
- $\bar{x}$ e $\bar{y}$ são as médias amostrais de x e y,
- n é o tamanho da amostra.

Resumo comparativo

Conceito	**Correlação (ρ)**	**Correlação amostral (r)**
Contexto	População (teórico)	Amostra (observado)
Tipo de medida	Parâmetro	Estatística (estimador)
Fórmula	Usa variâncias e covariância populacionais	Usa variâncias e covariância amostrais
Objetivo	Medir associação linear na população	Estimar ρ a partir de dados amostrais

Aplicações da Correlação em Séries Temporais

- A **correlação** mede o grau de associação linear entre duas variáveis. Em **séries temporais**, ela é usada para investigar **dependências ao longo do tempo**, seja **dentro da própria série** ou **entre séries diferentes**.
- A seguir estão listadas essas e outras aplicações:

1. **Autocorrelação**: correlação de uma série com ela mesma.
2. Correlação entre duas séries temporais diferentes.
3. Seleção de variáveis e defasagens para modelagem preditiva.
4. Diagnóstico de modelos (autocorrelação dos resíduos).

Autocorrelação (populacional)

- A **autocorrelação** mede a **intensidade** e a **direção** da **associação linear entre os valores da mesma variável** de um **processo estocástico** em dois **instantes distintos**, separados por uma **defasagem** (ou *lag*) h.

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{Cov(X_t, X_{t+h})}{Var(X_t)}$$

- $\rho(h)$ é a **autocorrelação populacional** no **lag** h.
- $\gamma(h) = Cov(X_t, X_{t+h})$ é a **função de autocovariância** no lag h.
- $\gamma(0) = Var(X_t) = \sigma^2$ é a **variância populacional** do processo estocástico.
- **Supõe-se que o processo estocástico seja estacionário** (média e variância constantes, e covariância dependendo apenas da defasagem).
- Tecnicamente, a autocorrelação é o **coeficiente de correlação de Pearson** entre X_t e X_{t+h}.
- Também pode ser interpretada como uma **padronização da autocovariância**.

Função de autocorrelação (FAC ou ACF)

- É a função que associa a cada defasagem h o correspondente coeficiente de autocorrelação $\rho(h)$.

$$FAC: h \mapsto \rho(h) = corr(X_t, X_{t+h}) \quad ou\ seja \quad FAC = \{\rho(h) | h = 0, 1, 2, \dots, h_{max}\}$$

- A FAC descreve o comportamento da dependência linear entre os **valores do processo estocástico ao longo do tempo, considerando todas as defasagens possíveis**.

- É útil na **identificação de modelos ARIMA**, ajudando a determinar o número adequado de defasagens (*lags*).

Autocorelação amostral

- A **autocorrelação amostral** no lag h é o estimador de $\rho(h)$, calculado a partir de uma amostra finita da série temporal.
- Para uma amostra $\{x_1, x_2, \dots, x_T\}$, com média amostral $\bar{x}$:

$$r(h) = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

- $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$ é a média amostral da série,
- $h \in \{0, 1, 2, \dots\}$ é o valor da defasagem (*lag*),
- x_t é o valor observado no tempo t,
- x_{t+h} é o valor observado h períodos à frente,
- Os valores de $r(h)$ estão entre −1 e 1.

Função de autocorrelação amostral (FAC amostral ou estimada)

- É a função que associa cada defasagem h ao correspondente coeficiente de autocorrelação amostral $r(h)$:

$$FAC\ amostral: h \mapsto r(h) \qquad ou \qquad FAC\ amostral = \{r(h) | h = 0, 1, \dots, h_{max}\}$$

- Essa função é empírica, calculada a partir dos dados disponíveis.
- Pode ser visualizada por meio do gráfico de autocorrelação (por ex., com a função acf() no R).
- Aplicações típicas:
  - Diagnosticar dependência temporal.
  - Avaliar a estacionariedade.
  - Identificar modelos apropriados (ex: número de defasagens (lags) em modelos AR).

Exemplo: calculo da autocorrelação amostral com defasagem $h = 1$

- Considere a série temporal de $T = 5$ observações igualmente espaçadas no tempo: $\{3, 5, 7, 6, 8\}$

- Passo 1: Calcular a média amostral: $\bar{x} = \frac{3+5+7+6+8}{5} = 5.8$

- Passo 2: Aplicar a fórmula da autocorrelação amostral

A fórmula da autocorrelação amostral com defasagem h é:

$$r(h) = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

- Para *h*=1:

$$r(1) = \frac{\sum_{t=1}^{4}(x_t - 5.8)(x_{t+1} - 5.8)}{\sum_{t=1}^{5}(x_t - 5.8)^2} \approx 0.13$$

t	x_t	x_{t+1}
1	3	5
2	5	7
3	7	6
4	6	8

- <u>Interpretação</u>: A autocorrelação com defasagem 1 apresenta um valor positivo, indicando que os valores consecutivos da série tendem a variar na mesma direção — ou seja, há uma associação linear positiva entre os valores separados por um período. Contudo, essa associação é fraca, dada a baixa magnitude da autocorrelação observada.

Exemplo: calculo da autocorrelação amostral com defasagem $h = 2$

- Considere a série temporal de $T = 5$ observações igualmente espaçadas no tempo: $\{3, 5, 7, 6, 8\}$

- Passo 1: Calcular a média amostral: $\bar{x} = \frac{3+5+7+6+8}{5} = 5.8$

- Passo 2: Aplicar a fórmula da autocorrelação amostral

A fórmula da autocorrelação amostral com defasagem h é:

$$r(h) = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

- Para *h*=1:

$$r(2) = \frac{\sum_{t=1}^{3}(x_t - 5.8)(x_{t+2} - 5.8)}{\sum_{t=1}^{5}(x_t - 5.8)^2} \approx -0.06$$

t	x_t	x_{t+2}
1	3	7
2	5	6
3	7	8

- <u>Interpretação</u>: A autocorrelação com defasagem 2 apresenta um valor negativo, indicando que os valores separados por dois períodos tendem a variar em direções opostas — ou seja, há uma associação linear negativa entre os valores separados por dois períodos. Contudo, essa associação é fraca, dada a baixa magnitude da autocorrelação observada.

Exemplo: calculo da autocorrelação amostral com defasagem $h = 0$

- Considere a série temporal de $T = 5$ observações igualmente espaçadas no tempo: $\{3, 5, 7, 6, 8\}$

t	x_t	x_{t+0}
1	3	3
2	5	5
3	7	7
4	6	6
5	8	8

- Passo 1: Calcular a média amostral: $\bar{x} = \frac{3+5+7+6+8}{5} = 5.8$

- Passo 2: Aplicar a fórmula da autocorrelação amostral

A fórmula da autocorrelação amostral com defasagem h é:

$$r(h) = \frac{\sum_{t=1}^{T-h}(x_t - \bar{x})(x_{t+h} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

- Para *h*=0:

$$r(0) = \frac{\sum_{t=1}^{5}(x_t - 5.8)(x_{t+0} - 5.8)}{\sum_{t=1}^{5}(x_t - 5.8)^2} = \frac{\sum_{t=1}^{5}(x_t - 5.8)^2}{\sum_{t=1}^{5}(x_t - 5.8)^2} = 1$$

- <u>Interpretação</u>: A autocorrelação de defasagem 0 é sempre igual a 1, pois a associação linear de cada valor consigo mesmo é perfeita e positiva.