Associações entre Variáveis Quantitativas - Parte 2

Aula	39
Preparada	☑
Revisada	☑
Lecionada	☑

Afinal, como medimos a “força” da associação?

Voltando ao início… Lembra-se da Variância?

Havíamos partido do desvio médio para saber o quanto erramos pela média:

$$DM = \frac{1}{n}\sum_{i=i}^{n}(x_i - \overline{x})$$

Mas ele era sempre igual a zero e precisamos “dar um jeito” nisso…

Uma das alternativas era elevar os termos ao quadrado. Surge a Variância:

$$Var(x) = \frac{1}{n-1}\sum_{i=i}^{n}(x_i - \overline{x})^2$$

Atenção: No caso da variância ser amostral o denominador da fração é n-1, caso a variância populacional, o denominador da fração é n.

Vamos então adotar uma abstração da variância…

Uma “variância conjunta” de *x* e *y*… Surge a Covariância:

$$Covariância(x,y) = \frac{1}{n-1}\sum_{i=i}^{n}[(x_i - \overline{x}) \times (y_i - \overline{y})]$$

Repare que a Covariância pode assumir qualquer valor real…

…mas ela depende das unidades adotadas, isso não é muito prático.

Karl Pearson ao resgate…

Karl Pearson é um estatístico britânico que, entre vários outros avanços, se dedicou a buscar uma maneira mais padronizada de medir a força da associação. Foi ele, em 1896, quem apresentou o conceito de Correlação:

$$Correlação(x,y) = \frac{Covariância(x,y)}{DesvioPadrão(x) \times DesvioPadrão(y)}$$

Organizando e simplificando as fórmulas…

Diretamente

$$Correlação(x,y) = r_{(x,y)} = \frac{\frac{1}{n-1}\sum_{i=i}^{n}[(x_i-\overline{x})\times(y_i-\overline{y})]}{\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(x_i-\overline{x})^2}\times\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(y_i-\overline{y})^2}}$$

- **Fatorando a fração $\frac{1}{n-1}$, temos que**

$$Correlação(x,y) = r_{(x,y)} = \frac{\frac{1}{n-1}}{\left(\sqrt{\frac{1}{n-1}}\right)^2}\times\frac{\sum_{i=i}^{n}[(x_i-\overline{x})\times(y_i-\overline{y})]}{\sqrt{\sum_{i=i}^{n}(x_i-\overline{x})^2}\times\sqrt{\sum_{i=i}^{n}(y_i-\overline{y})^2}}$$

- **Finalmente…**

$$Correlação(x,y) = \frac{\sum_{i=i}^{n}[(x_i-\overline{x})\times(y_i-\overline{y})]}{\sqrt{\sum_{i=i}^{n}(x_i-\overline{x})^2}\times\sqrt{\sum_{i=i}^{n}(y_i-\overline{y})^2}}$$

Em função do Z-score

$$Correlação(x,y) = r_{(x,y)} = \frac{\frac{1}{n-1}\sum_{i=i}^{n}[(x_i-\overline{x})\times(y_i-\overline{y})]}{\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(x_i-\overline{x})^2}\times\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(y_i-\overline{y})^2}}$$

- **Isolando a fração $\frac{1}{n-1}$ no numerador, temos que**

$$Correlação(x,y) = r_{(x,y)} = \frac{1}{n-1}\times\frac{\sum_{i=i}^{n}[(x_i-\overline{x})\times(y_i-\overline{y})]}{\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(x_i-\overline{x})^2}\times\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(y_i-\overline{y})^2}}$$

- **Lembrando que…**

$$z_{(x)} = \frac{(x_i-\overline{x})}{s_x} = \frac{(x_i-\overline{x})}{\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(x_i-\overline{x})^2}} \qquad z_{(y)} = \frac{(y_i-\overline{y})}{s_y} = \frac{(y_i-\overline{y})}{\sqrt{\frac{1}{n-1}\sum_{i=i}^{n}(y_i-\overline{y})^2}}$$

- **Finalmente**

$$Correlação(x,y) = r_{(x,y)} = \frac{1}{n-1}\sum_{i=i}^{n}\left[z_{(x)}\times z_{(y)}\right]$$

Assim, em resumo, a Correlação…

…entre *x* e *y* é denotada por $\rho_{(x,y)}$ se for na população e por $r_{(x,y)}$ se for em uma amostra.

Correlação (apenas em função das médias):

$$Correlação(x,y) = \rho_{(x,y)} = r_{(x,y)} = \frac{\sum_{i=i}^{n}[(x_i-\overline{x})\times(y_i-\overline{y})]}{\sqrt{\sum_{i=i}^{n}(x_i-\overline{x})^2}\times\sqrt{\sum_{i=i}^{n}(y_i-\overline{y})^2}}$$

Correlação (em função dos Z-scores):

$$Correlação_{Amostral}(x,y) = r_{(x,y)} = \frac{1}{n-1}\sum_{i=i}^{n}\left[z_{(x)}\times z_{(y)}\right]$$

$$Correlação_{Populacional}(x,y) = \rho_{(x,y)} = \frac{1}{n}\sum_{i=i}^{n}\left[z_{(x)} \times z_{(y)}\right]$$

Observação: Na grande maioria das vezes, vamos usar a correlação amostral…