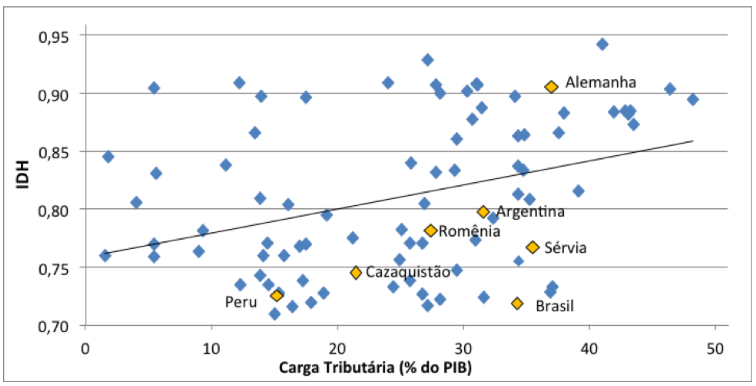
Plano Aula 07 e 08

Medidas de Associação

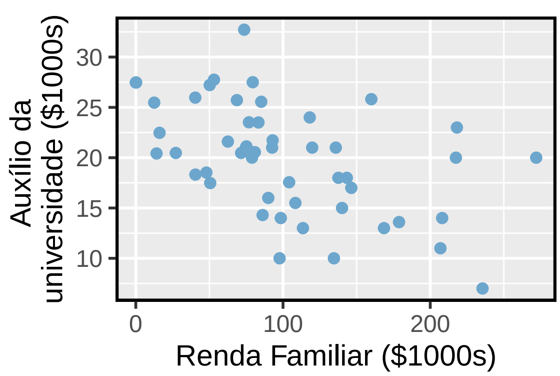
Agora, nosso interesse será analisar o relacionamento entre **duas variáveis numéricas** de interesse.

- **Exemplo 1**: O Índice de Desenvolvimento Humano (IDH) em países pode estar associado à carga tributária?



Artigo de 2012: https://carodinheiro.blogfolha.uol.com.br/2012/12/14/pagamento-de-impostos-no-brasil-e-um-investimento-sem-retorno/

- **Exemplo 2**: O valor do auxílio estudantil oferecido por uma universidade pode estar relacionado com a renda familiar dos estudantes?



Associação entre Variáveis Quantitativas (Bussab e Morettin - seção 4.5)

Para duas variáveis quantitativas também podemos estar interessados em verificar se existe associação (relação) entre elas.

Gráfico de Dispersão

- **... cont. exemplo 1**:

Como resumir a informação do gráfico acima em um só número?

Coeficiente de correlação (linear) (de Pearson)

Relembrando sobre covariância em probabilidade:
Definição (**covariância**): Sejam X e Y duas v.a. então $Cov(X,Y) = E\{[X - E(X)]\,[Y - E(Y)]\}$.

Definição (**covariância amostral**): Dados n pares de valores observados $(x_1, y_1), \ldots, (x_n, y_n)$ (de duas v.a. X e Y), chamaremos de covariância amostral antre X e Y a expressão

$$cov(X,Y) = \frac{s_{xy}}{n},$$

em que $s_{xy} = \sum_i (x_i - \overline{x})(y_i - \overline{y}) = \sum_i x_i y_i - [(\sum_i x_i)(\sum_i y_i)]/n$.

Assim, “padronizamos” a covariância para obtemos o coeficiente $corr(X,Y) \approx \frac{cov(X,Y)}{s_x \cdot s_y}$, em que s_x e s_y são os desvios padrões de X e Y respectivamente, então $-1 \leq corr(X,Y) \leq 1$.

Definição (**coeficiente de correlação**): Dados n pares de valores observados $(x_1, y_1), \ldots, (x_n, y_n)$ (de duas v.a. X e Y), chamaremos de covariância amostral entre X e Y a expressão

$$r = corr(X,Y) = \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}},$$

em que

- $s_{xx} = \sum_i (x_i - \overline{x})^2 = \sum_i x_i^2 - (\sum_i x_i)^2/n$,
- $s_{yy} = \sum_i (y_i - \overline{y})^2 = \sum_i y_i^2 - (\sum_i y_i)^2/n$ e
- $s_{xy} = \sum_i (x_i - \overline{x})(y_i - \overline{y}) = \sum_i x_i y_i - [(\sum_i x_i)(\sum_i y_i)]/n$;

ou

$$r = corr(X,Y) = \frac{cov(X,Y)}{\sqrt{\frac{s_{xx}}{n} \frac{s_{yy}}{n}}}.$$

Em Estatística Geral 2 veremos como usar os valores de uma amostra observda para testar se existe correlação (associação)?

Correlação espúria

Causalidade e correlação

Ler slides e ver vídeos da semana 4.

Continuar lista de exercícios 1-2.
