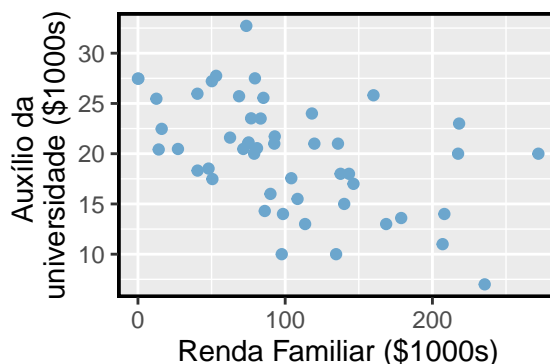


Plano Aula 27 e 28

Análise de Correlação e Regressão

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

- **Exemplo 1:** 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?

Coefficiente 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 $(x_1, y_1), \dots, (x_n, y_n)$ (de duas v.a. X e Y), chamaremos de covariância amostral entre X e Y a expressão

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

.

Assim, “padronizamos” a covariância para obtemos o coeficiente $corr(X, Y) = \frac{cov(X, Y)}{s_x \cdot s_y}$, $-1 \leq corr(X, Y) \leq 1$.

Definição (**coeficiente de correlação**): Dados n pares de valores observados $(x_1, y_1), \dots, (x_n, y_n)$ (de duas

v.a. X e Y), chamaremos de covariância amostral entre X e Y a expressão

$$\text{corr}(X, Y) = \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \bar{x})}{s_x} \cdot \frac{(y_i - \bar{y})}{s_y}.$$

Como usar os valores de cada observação para testar se existe correlação (associação)?

Teste para o Coeficiente de Correlação (Bussab e Morettin - seção 14.5)

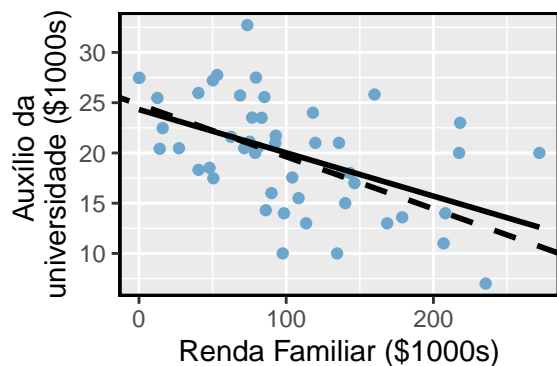
- Hipóteses?
- Quais as suposições necessárias???

Análise de Regressão

Agora, nosso interesse será em estimar uma relação **linear** entre **duas variáveis numéricas** de interesse.

Regressão Linear Simples (Bussab e Morettin - capítulo 16)

- **cont. Exemplo 1:** O valor do auxílio estudantil oferecido por uma universidade pode estar relacionado com a renda familiar dos estudantes?



Estimação dos parâmetros (Bussab e Morettin - seção 16.2)

Relembrando sobre esperança condicional em probabilidade...

Modelo populacional

sejam X e Y duas v.a. queremos estimar a esperança condicional de Y em função de (dado que) $X = x$,

$$E(Y|X = x) = \alpha + \beta \cdot x,$$

ou seja, queremos estimar os parâmetros α e β .

- Para uma amostra de tamanho n podemos escrever que cada observação (x_i, y_i) , para $i = 1, \dots, n$, segue o modelo

$$y_i = \alpha + \beta \cdot x_i + e_i.$$

- chamamos e_i de erro amostral e assumimos que:
 - $E(e_i) = 0$,
 - $\text{Var}(e_i) = \sigma^2$, para todo $i, j = 1, \dots, n$,
 - $\text{Cov}(e_i, e_j) = 0$ para $i \neq j$.
- Assim $E(y_i) = \alpha + \beta \cdot x_i$, $\text{Var}(y_i) = \sigma^2$ e $\text{Cov}(y_i, y_j) = 0$.
 - o parâmetro σ^2 também precisa ser estimado.

Método dos Mínimos Quadrados (Ordinários)

Para estimar α e β podemos pensar em minimizar os erros e_i , ou

$$SQ(\alpha, \beta) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (\alpha + \beta \cdot x_i)]^2$$

- Derivando $SQ(\alpha, \beta)$ em relação a α e β , igualando a zero e resolvendo o sistema de equações temos
 - $\hat{\alpha} = ???$ e $\hat{\beta} = ???$.
- Reta estimada (modelo ajustado): $\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i$
 - Interpretação de $\hat{\alpha}$ e $\hat{\beta}$;
 - **Prever**, para um dado valor $X = x$, quanto esperamos observar o valor de Y ?

Coefficiente de determinação R^2 (Bussab e Morettin - seção 16.3)

Intervalos de Confiança e Testes de hipóteses (Bussab e Morettin - seção 16.4)

Para α , β e σ^2 . (suposições???)

Se adicionalmente assumimos $e_i \sim Normal$, então $y_i \sim Normal(\alpha + \beta \cdot x, \sigma^2)$.

- Também $\hat{\alpha} \sim Normal$ e $\hat{\beta} \sim Normal$.

Previsão e predição (Bussab e Morettin - seção 16.4.4)

- para o valor esperado $E(Y_i|x_i) = y_i$;
- para uma futura (nova) observação y_f .

Correlação espúria

Causalidade e correlação

Ler slides e ver vídeos da semana 14.

Fazer lista de exercícios 3-4.

Fazer o Quiz da semana 14 - VALE NOTA!!!
