

3.07pt

# Estatística Descritiva III

Bacharelado em Economia - FEA - Noturno

1<sup>o</sup> Semestre 2017

Profs. Vanderlei C. Bueno e Gilberto A. Paula

## Sumário

3.07pt

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Tabelas de Contingência
- 4 Diagrama de Dispersão
- 5 Correlação Linear
- 6 Reta de Regressão

## Objetivos da Aula

### Objetivos da Aula

Nesta aula apresentaremos alguns procedimentos descritivos para o estudo de associação entre duas (ou mais) variáveis qualitativas ou quantitativas. Inicialmente, discutiremos a descrição de dados qualitativos através de **Tabelas de Contingência**. Em seguida, discutiremos procedimentos para avaliar a associação entre duas variáveis quantitativas, tais como **Diagramas de Dispersão**, **Coeficiente de Correlação Linear** e a **Reta de Regressão**. Motivaremos com o mesmo exemplo ilustrativo da aula anterior.

## Sumário

3.07pt

- 1 Objetivos da Aula
- 2 Fluência da Fala**
- 3 Tabelas de Contingência
- 4 Diagrama de Dispersão
- 5 Correlação Linear
- 6 Reta de Regressão

# Fluência da Fala

## Descrição do Projeto

# Fluência da Fala

## Descrição do Projeto

- Dados do projeto Perfil Evolutivo da Fluência da Fala de Falantes do Português Brasileiro

# Fluência da Fala

## Descrição do Projeto

- Dados do projeto Perfil Evolutivo da Fluência da Fala de Falantes do Português Brasileiro
- Estudo realizado pela Faculdade de Medicina - USP e pela Faculdade de Filosofia, Letras e Ciências Humanas - USP



# Fluência da Fala

## Descrição do Projeto

- Dados do projeto Perfil Evolutivo da Fluência da Fala de Falantes do Português Brasileiro
- Estudo realizado pela Faculdade de Medicina - USP e pela Faculdade de Filosofia, Letras e Ciências Humanas - USP
- Projeto desenvolvido em 2006 referente a tese de doutorado

## Fluência da Fala

### Descrição do Projeto

- Dados do projeto **Perfil Evolutivo da Fluência da Fala de Falantes do Português Brasileiro**
- Estudo realizado pela Faculdade de Medicina - USP e pela Faculdade de Filosofia, Letras e Ciências Humanas - USP
- Projeto desenvolvido em 2006 referente a tese de doutorado
- Análise Estatística realizada pelo Centro de Estatística Aplicada (CEA) do IME-USP (**Dados CEA0P16**)

# Fluência da Fala

## Objetivo do Projeto

Avaliar o perfil da fluência da fala de acordo com a idade, gênero e grau de escolaridade.

## Fluência da Fala

### Objetivo do Projeto

Avaliar o perfil da fluência da fala de acordo com a idade, gênero e grau de escolaridade.

### Amostra

A amostra consistiu de 594 indivíduos residentes na cidade de São Paulo com idade entre 2 e 99 anos.

## Fluência da Fala

### Descrição do Experimento

Foram obtidas de cada indivíduo amostras de fala auto-expressiva. O indivíduo era apresentado a uma figura e orientado a discorrer sobre a mesma durante um tempo mínimo de 3 minutos e máximo de 6 minutos. Para crianças de 2 e 3 anos, as amostras foram obtidas com a colaboração dos pais.

# Fluência da Fala

Algumas Variáveis do Estudo

# Fluência da Fala

## Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)

# Fluência da Fala

## Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)



# Fluência da Fala

## Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)
- Grau de escolaridade (pré-escola a superior)

# Fluência da Fala

## Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)
- Grau de escolaridade (pré-escola a superior)
- Fluxo de palavras por minuto (FPM)

# Fluência da Fala

## Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)
- Grau de escolaridade (pré-escola a superior)
- Fluxo de palavras por minuto (FPM)
- Fluxo de sílabas por minuto (FSM)

## Fluência da Fala

### Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)
- Grau de escolaridade (pré-escola a superior)
- Fluxo de palavras por minuto (FPM)
- Fluxo de sílabas por minuto (FSM)
- Número de interjeições durante o discurso (INTERJ)

## Fluência da Fala

### Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)
- Grau de escolaridade (pré-escola a superior)
- Fluxo de palavras por minuto (FPM)
- Fluxo de sílabas por minuto (FSM)
- Número de interjeições durante o discurso (INTERJ)
- Número de palavras não terminadas durante o discurso (PNT)

## Fluência da Fala

### Algumas Variáveis do Estudo

- Gênero (1:feminino e 2:masculino)
- Idade (em anos)
- Grau de escolaridade (pré-escola a superior)
- Fluxo de palavras por minuto (FPM)
- Fluxo de sílabas por minuto (FSM)
- Número de interjeições durante o discurso (INTERJ)
- Número de palavras não terminadas durante o discurso (PNT)
- Número de pausas durante o discurso (PAUSA).

# Fluência da Fala

## Variáveis Qualitativas

---

# Fluência da Fala

## Variáveis Qualitativas

- Nominal: gênero



# Fluência da Fala

## Variáveis Qualitativas

- Nominal: gênero
- Ordinal: grau de escolaridade

# Fluência da Fala

## Variáveis Qualitativas

- Nominal: gênero
- Ordinal: grau de escolaridade

## Variáveis Quantitativas

# Fluência da Fala

## Variáveis Qualitativas

- Nominal: gênero
- Ordinal: grau de escolaridade

## Variáveis Quantitativas

- Contínua: idade, fluxo de palavras por minuto e fluxo de sílabas por minuto

## Fluência da Fala

### Variáveis Qualitativas

- Nominal: gênero
- Ordinal: grau de escolaridade

### Variáveis Quantitativas

- Contínua: idade, fluxo de palavras por minuto e fluxo de sílabas por minuto
- Discreta: número de interjeições durante o discurso, número de palavras não terminadas durante o discurso e número de pausas durante o discurso

# Sumário

3.07pt

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Tabelas de Contingência**
- 4 Diagrama de Dispersão
- 5 Correlação Linear
- 6 Reta de Regressão

# Tabelas de Contingência

## Definição

Tabelas de contingência são utilizadas para descrever a distribuição (em frequências ou em porcentagens) de duas ou mais variáveis qualitativas.

# Fluência da Fala

## Tabela de Frequências

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	52	96	44	24	20	24	260
Masculino	48	69	21	24	24	20	206
Total	100	165	65	48	44	44	466

## Fluência da Fala

### Tabela de Frequências

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	52	96	44	24	20	24	260
Masculino	48	69	21	24	24	20	206
Total	100	165	65	48	44	44	466

### Interpretações



## Fluência da Fala

### Tabela de Frequências

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	52	96	44	24	20	24	260
Masculino	48	69	21	24	24	20	206
Total	100	165	65	48	44	44	466

### Interpretações

- em virtude de alguns indivíduos não terem informado o nível de instrução foram considerados apenas 466 indivíduos da amostra

## Fluência da Fala

### Tabela de Frequências

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	52	96	44	24	20	24	260
Masculino	48	69	21	24	24	20	206
Total	100	165	65	48	44	44	466

### Interpretações

- em virtude de alguns indivíduos não terem informado o nível de instrução foram considerados apenas 466 indivíduos da amostra
- desses 466 indivíduos, por exemplo, 44 têm ensino médio completo e 165 fundamental incompleto

## Fluência da Fala

## Tabela de Porcentagens

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	11,2	20,6	9,4	5,2	4,3	5,2	55,8
Masculino	10,3	14,8	4,5	5,2	5,2	4,3	44,2
Total	21,5	35,4	13,9	10,3	9,4	9,4	100,0

## Fluência da Fala

### Tabela de Porcentagens

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	11,2	20,6	9,4	5,2	4,3	5,2	55,8
Masculino	10,3	14,8	4,5	5,2	5,2	4,3	44,2
Total	21,5	35,4	13,9	10,3	9,4	9,4	100,0

### Interpretações

## Fluência da Fala

### Tabela de Porcentagens

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	11,2	20,6	9,4	5,2	4,3	5,2	55,8
Masculino	10,3	14,8	4,5	5,2	5,2	4,3	44,2
Total	21,5	35,4	13,9	10,3	9,4	9,4	100,0

### Interpretações

- nota-se que 55,8% da amostra é do sexo feminino enquanto 44,2% é do sexo masculino

## Fluência da Fala

### Tabela de Porcentagens

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	11,2	20,6	9,4	5,2	4,3	5,2	55,8
Masculino	10,3	14,8	4,5	5,2	5,2	4,3	44,2
Total	21,5	35,4	13,9	10,3	9,4	9,4	100,0

### Interpretações

- nota-se que 55,8% da amostra é do sexo feminino enquanto 44,2% é do sexo masculino
- nota-se, por exemplo, que 20,6% da amostra é do sexo feminino e têm fundamental incompleto

## Fluência da Fala

### Porcentagens Marginais por Gênero

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	20,0	36,9	16,9	9,2	7,7	9,2	100,0
Masculino	23,3	33,5	10,2	11,7	11,7	9,7	100,0

## Fluência da Fala

### Porcentagens Marginais por Gênero

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	20,0	36,9	16,9	9,2	7,7	9,2	100,0
Masculino	23,3	33,5	10,2	11,7	11,7	9,7	100,0

### Interpretações



## Fluência da Fala

### Porcentagens Marginais por Gênero

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	20,0	36,9	16,9	9,2	7,7	9,2	100,0
Masculino	23,3	33,5	10,2	11,7	11,7	9,7	100,0

### Interpretações

- no grupo feminino 9,2% têm ensino médio incompleto

## Fluência da Fala

### Porcentagens Marginais por Gênero

Gênero	Nível de Instrução						Total
	PE	FI	FC	MI	MC	Sup	
Feminino	20,0	36,9	16,9	9,2	7,7	9,2	100,0
Masculino	23,3	33,5	10,2	11,7	11,7	9,7	100,0

### Interpretações

- no grupo feminino 9,2% têm ensino médio incompleto
- no grupo masculino 10,2% têm fundamental completo

## Fluência da Fala

## Porcentagens Marginais por NI

Gênero	Nível de Instrução					
	PE	FI	FC	MI	MC	Sup
Feminino	52,0	58,2	67,7	50,0	45,5	54,5
Masculino	48,0	41,8	32,3	50,0	54,5	45,5
Total	100,0	100,0	100,0	100,0	100,0	100,0

## Fluência da Fala

### Porcentagens Marginais por NI

Gênero	Nível de Instrução					
	PE	FI	FC	MI	MC	Sup
Feminino	52,0	58,2	67,7	50,0	45,5	54,5
Masculino	48,0	41,8	32,3	50,0	54,5	45,5
Total	100,0	100,0	100,0	100,0	100,0	100,0

### Interpretações

## Fluência da Fala

### Porcentagens Marginais por NI

Gênero	Nível de Instrução					
	PE	FI	FC	MI	MC	Sup
Feminino	52,0	58,2	67,7	50,0	45,5	54,5
Masculino	48,0	41,8	32,3	50,0	54,5	45,5
Total	100,0	100,0	100,0	100,0	100,0	100,0

### Interpretações

- dentre os indivíduos com curso superior 54,5% é do sexo feminino

## Fluência da Fala

### Porcentagens Marginais por NI

Gênero	Nível de Instrução					
	PE	FI	FC	MI	MC	Sup
Feminino	52,0	58,2	67,7	50,0	45,5	54,5
Masculino	48,0	41,8	32,3	50,0	54,5	45,5
Total	100,0	100,0	100,0	100,0	100,0	100,0

### Interpretações

- dentre os indivíduos com curso superior 54,5% é do sexo feminino
- dentre os indivíduos com ensino médio incompleto metade é do sexo masculino

## Sumário

3.07pt

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Tabelas de Contingência
- 4 Diagrama de Dispersão**
- 5 Correlação Linear
- 6 Reta de Regressão

## Diagrama de Dispersão

### Definição

Os diagramas de dispersão são representações dos valores observados de duas (ou mais) variáveis quantitativas num mesmo gráfico. Um dos objetivos desse tipo de gráfico é avaliar a possível existência de alguma relação funcional entre as variáveis em estudo.



## Diagrama de Dispersão

### Definição

Os diagramas de dispersão são representações dos valores observados de duas (ou mais) variáveis quantitativas num mesmo gráfico. Um dos objetivos desse tipo de gráfico é avaliar a possível existência de alguma relação funcional entre as variáveis em estudo.

- Por exemplo, verificar se há **relação linear** ou alguma **relação não linear** entre as duas variáveis quantitativas.

## Diagrama de Dispersão

### Definição

Os diagramas de dispersão são representações dos valores observados de duas (ou mais) variáveis quantitativas num mesmo gráfico. Um dos objetivos desse tipo de gráfico é avaliar a possível existência de alguma relação funcional entre as variáveis em estudo.

- Por exemplo, verificar se há **relação linear** ou alguma **relação não linear** entre as duas variáveis quantitativas.

Quando há suspeita de uma das variáveis ser preditora de uma outra (**relação de causa-efeito**), a variável preditora é colocada no eixo-x enquanto a variável resposta é colocada no eixo-y.

# Diagrama de Dispersão

Exemplos

# Diagrama de Dispersão

## Exemplos

- idade e altura das crianças

# Diagrama de Dispersão

## Exemplos

- idade e altura das crianças
- tempo de prática de esportes e ritmo cardíaco

# Diagrama de Dispersão

## Exemplos

- idade e altura das crianças
- tempo de prática de esportes e ritmo cardíaco
- fluxo de sílabas e fluxo de palavras por minuto

# Diagrama de Dispersão

## Exemplos

- idade e altura das crianças
- tempo de prática de esportes e ritmo cardíaco
- fluxo de sílabas e fluxo de palavras por minuto
- tempo de estudo e nota na prova

# Diagrama de Dispersão

## Exemplos

- idade e altura das crianças
- tempo de prática de esportes e ritmo cardíaco
- fluxo de sílabas e fluxo de palavras por minuto
- tempo de estudo e nota na prova
- taxa de desemprego e taxa de criminalidade

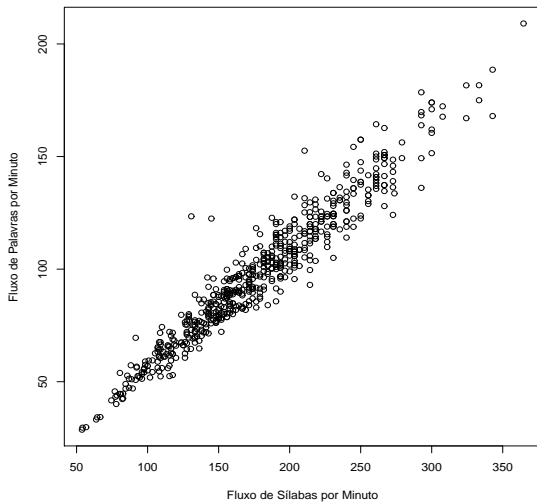


# Diagrama de Dispersão

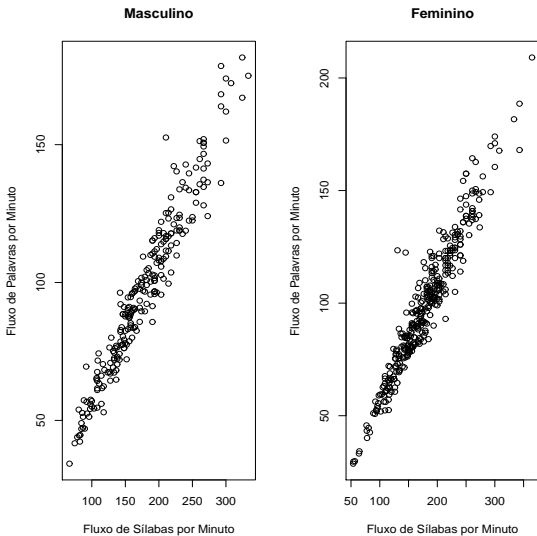
## Exemplos

- idade e altura das crianças
- tempo de prática de esportes e ritmo cardíaco
- fluxo de sílabas e fluxo de palavras por minuto
- tempo de estudo e nota na prova
- taxa de desemprego e taxa de criminalidade
- expectativa de vida e taxa de analfabetismo

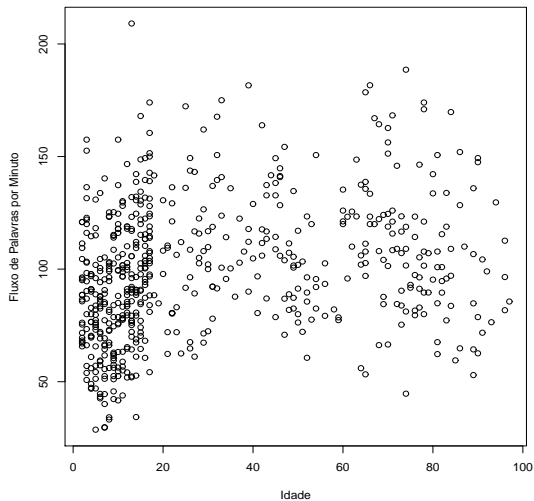
## Dispersão entre FSM e FPM



## Dispersão entre FSM e FPM segundo o Gênero



# Dispersão entre FPM e Idade



# Diagrama de Dispersão

Comentários

## Diagrama de Dispersão

### Comentários

- nota-se uma forte relação linear crescente entre FSM e FPM, ou seja, à medida que FSM aumenta também aumenta o FPM. Essa tendência é mantida quando há estratificação por gênero

## Diagrama de Dispersão

### Comentários

- nota-se uma forte relação linear crescente entre FSM e FPM, ou seja, à medida que FSM aumenta também aumenta o FPM. Essa tendência é mantida quando há estratificação por gênero
- todavia, quando a Idade do indivíduo é considerada, nota-se uma leve relação linear crescente

# Sumário

3.07pt

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Tabelas de Contingência
- 4 Diagrama de Dispersão
- 5 Correlação Linear**
- 6 Reta de Regressão



# Coefficiente de Correlação Linear

## Definição

O **coeficiente de correlação linear de Pearson** é uma medida resumo que quantifica a força da relação linear entre duas variáveis quantitativas.

## Coeficiente de Correlação Linear

### Definição

O coeficiente de correlação linear de Pearson é expresso na seguinte forma:

## Coeficiente de Correlação Linear

### Definição

O coeficiente de correlação linear de Pearson é expresso na seguinte forma:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_X s_Y},$$

## Coefficiente de Correlação Linear

### Definição

O coeficiente de correlação linear de Pearson é expresso na seguinte forma:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_X s_Y},$$

em que

## Coefficiente de Correlação Linear

### Definição

O coeficiente de correlação linear de Pearson é expresso na seguinte forma:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_X s_Y},$$

em que

- $\bar{x}$  e  $\bar{y}$  denotam as médias amostrais

## Coefficiente de Correlação Linear

### Definição

O coeficiente de correlação linear de Pearson é expresso na seguinte forma:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_X s_Y},$$

em que

- $\bar{x}$  e  $\bar{y}$  denotam as médias amostrais
- $s_X$  e  $s_Y$  denotam os respectivos desvios padrão amostrais

## Coeficiente de Correlação Linear

### Definição

O coeficiente de correlação linear de Pearson é expresso na seguinte fórmula alternativa:

## Coefficiente de Correlação Linear

### Definição

O coeficiente de correlação linear de Pearson é expresso na seguinte fórmula alternativa:

$$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{(n-1)s_X s_Y}.$$



## Coeficiente de Correlação Linear

### Propriedades

O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

## Coeficiente de Correlação Linear

### Propriedades

O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

$$-1 \leq r \leq 1.$$

## Coefficiente de Correlação Linear

### Propriedades

O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

$$-1 \leq r \leq 1.$$

### Casos particulares

## Coefficiente de Correlação Linear

### Propriedades

O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

$$-1 \leq r \leq 1.$$

### Casos particulares

- $r = 1$ : correlação linear positiva e perfeita

## Coefficiente de Correlação Linear

### Propriedades

O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

$$-1 \leq r \leq 1.$$

### Casos particulares

- $r = 1$ : correlação linear positiva e perfeita
- $r = -1$ : correlação linear negativa e perfeita

## Coefficiente de Correlação Linear

### Propriedades

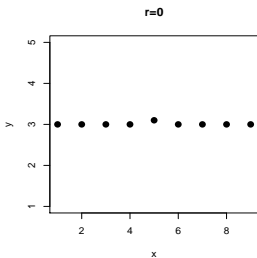
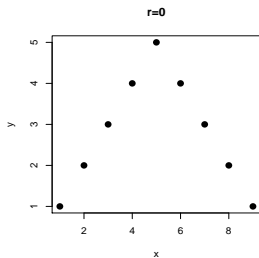
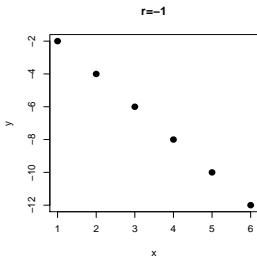
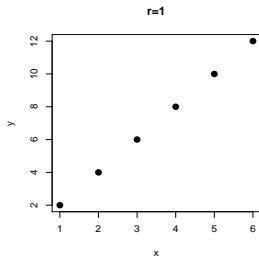
O coeficiente de correlação linear de Pearson apresenta a seguinte propriedade:

$$-1 \leq r \leq 1.$$

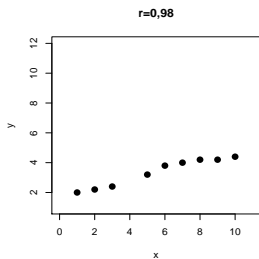
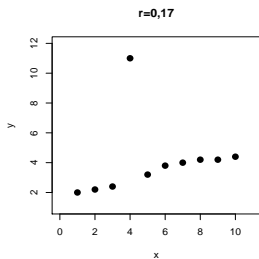
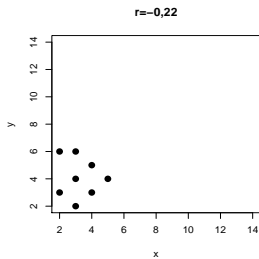
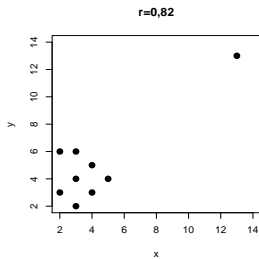
### Casos particulares

- $r = 1$ : correlação linear positiva e perfeita
- $r = -1$ : correlação linear negativa e perfeita
- $r = 0$ : ausência de correlação linear

## Exemplos



## Exemplos





# Exemplos

Comentários

## Exemplos

### Comentários

- correlação linear próxima de zero significa ausência de relação linear entre as variáveis quantitativas, contudo pode haver outro tipo de relação entre as variáveis

## Exemplos

### Comentários

- correlação linear próxima de zero significa ausência de relação linear entre as variáveis quantitativas, contudo pode haver outro tipo de relação entre as variáveis
- pontos extremos no diagrama de dispersão podem influenciar o valor do coeficiente de correlação linear. Assim, deve-se olhar conjuntamente o diagrama de dispersão e o coeficiente de correlação linear

## Tempo de Estudo e Nota na Prova

### Descrição das Variáveis

Vamos considerar dados hipotéticos para as variáveis quantitativas

## Tempo de Estudo e Nota na Prova

### Descrição das Variáveis

Vamos considerar dados hipotéticos para as variáveis quantitativas

- X: Tempo de Estudo (em horas)

## Tempo de Estudo e Nota na Prova

### Descrição das Variáveis

Vamos considerar dados hipotéticos para as variáveis quantitativas

- X: Tempo de Estudo (em horas)
- Y: Nota na Prova

## Tempo de Estudo e Nota na Prova

### Descrição das Variáveis

Vamos considerar dados hipotéticos para as variáveis quantitativas

- X: Tempo de Estudo (em horas)
- Y: Nota na Prova

para uma amostra de  $n = 5$  estudantes.

## Tempo de Estudo e Nota na Prova

### Descrição das Variáveis

Vamos considerar dados hipotéticos para as variáveis quantitativas

- X: Tempo de Estudo (em horas)
- Y: Nota na Prova

para uma amostra de  $n = 5$  estudantes.

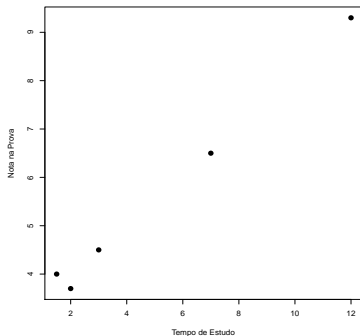
### Descrição dos Dados

X	3,0	7,0	2,0	1,5	12,0
Y	4,5	6,5	3,7	4,0	9,3



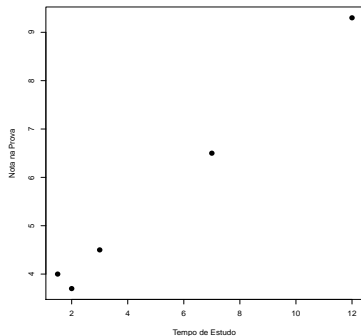
# Tempo de Estudo e Nota na Prova

## Diagrama de Dispersão



## Tempo de Estudo e Nota na Prova

### Diagrama de Dispersão



Podemos notar que, conforme aumenta o **Tempo de Estudo (X)**, a **Nota na Prova (Y)** tende a aumentar. Nota-se também uma tendência linear.

## Tempo de Estudo e Nota na Prova

### Algumas Quantidades

Temos que:  $\bar{x} = 5,1$ ;  $\bar{y} = 5,6$ ;  $\sum x_i^2 = 208,25$ ;  $\sum y_i^2 = 178,68$  e  $\sum x_i y_i = 184$ .

## Tempo de Estudo e Nota na Prova

### Algumas Quantidades

Temos que:  $\bar{x} = 5,1$ ;  $\bar{y} = 5,6$ ;  $\sum x_i^2 = 208,25$ ;  $\sum y_i^2 = 178,68$  e  $\sum x_i y_i = 184$ .

### Desvios Padrão Estimados

$$s_X^2 = \frac{208,25 - 5 \times (5,1)^2}{4} = 19,55 \rightarrow s_X = \sqrt{19,55} = 4,42.$$

$$s_Y^2 = \frac{178,68 - 5 \times (5,6)^2}{4} = 5,47 \rightarrow s_Y = \sqrt{5,47} = 2,34.$$

## Tempo de Estudo e Nota na Prova

### Cálculo da Correlação

Portanto, o coeficiente de correlação linear de Pearson fica dado por

## Tempo de Estudo e Nota na Prova

### Cálculo da Correlação

Portanto, o coeficiente de correlação linear de Pearson fica dado por

$$\begin{aligned} r &= \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) s_X s_Y} \\ &= \frac{184 - 5 \times 5,1 \times 5,6}{4 \times 4,42 \times 2,34} \\ &= \frac{184 - 142,8}{41,37} \\ &= \frac{41,20}{41,37} = 0,996. \end{aligned}$$

## Criminalidade e Analfabetismo

### Descrição das Variáveis

Vamos considerar como ilustração a relação entre

## Criminalidade e Analfabetismo

### Descrição das Variáveis

Vamos considerar como ilustração a relação entre

- X: Taxa de Analfabetismo (em %)



## Criminalidade e Analfabetismo

### Descrição das Variáveis

Vamos considerar como ilustração a relação entre

- X: Taxa de Analfabetismo (em %)
- Y: Taxa de Criminalidade (por 100.000 habitantes)

## Criminalidade e Analfabetismo

### Descrição das Variáveis

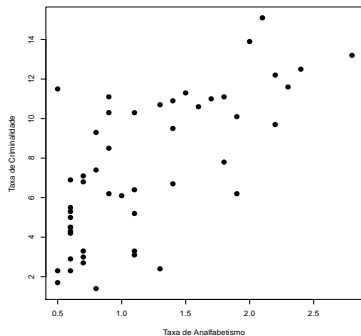
Vamos considerar como ilustração a relação entre

- X: Taxa de Analfabetismo (em %)
- Y: Taxa de Criminalidade (por 100.000 habitantes)

nos 50 estados norte americanos no início dos anos 70.

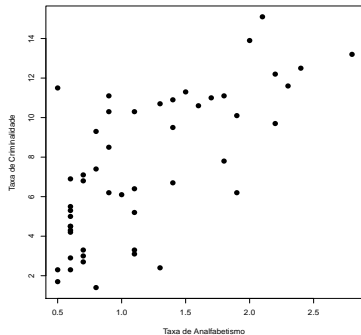
# Criminalidade e Analfabetismo

## Diagrama de Dispersão



# Criminalidade e Analfabetismo

## Diagrama de Dispersão



Podemos notar que, conforme aumenta a **Taxa de Analfabetismo (X)**, a **Taxa de Criminalidade (Y)** tende a aumentar. Nota-se também uma tendência linear.

## Criminalidade e Analfabetismo

### Algumas Quantidades

Temos que:  $\bar{x} = 1,17$ ;  $\bar{y} = 7,38$ ;  $s_X = 0,609$ ;  $s_Y = 3,692$  e  
 $\sum x_i y_i = 509,12$ .

## Criminalidade e Analfabetismo

### Algumas Quantidades

Temos que:  $\bar{x} = 1,17$ ;  $\bar{y} = 7,38$ ;  $s_X = 0,609$ ;  $s_Y = 3,692$  e  $\sum x_i y_i = 509,12$ .

### Cálculo da Correlação

Portanto, o coeficiente de correlação linear de Pearson fica dado por

## Criminalidade e Analfabetismo

### Algumas Quantidades

Temos que:  $\bar{x} = 1,17$ ;  $\bar{y} = 7,38$ ;  $s_X = 0,609$ ;  $s_Y = 3,692$  e  $\sum x_i y_i = 509,12$ .

### Cálculo da Correlação

Portanto, o coeficiente de correlação linear de Pearson fica dado por

$$\begin{aligned}
 r &= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_X s_Y} \\
 &= \frac{509,12 - 50 \times 1,17 \times 7,38}{49 \times 0,609 \times 3,692} \\
 &= \frac{77,39}{110,17} = 0,702.
 \end{aligned}$$

## Expectativa de Vida e Analfabetismo

### Descrição das Variáveis

Vamos considerar como ilustração a relação entre



## Expectativa de Vida e Analfabetismo

### Descrição das Variáveis

Vamos considerar como ilustração a relação entre

- X: Taxa de Analfabetismo (em %)

## Expectativa de Vida e Analfabetismo

### Descrição das Variáveis

Vamos considerar como ilustração a relação entre

- X: Taxa de Analfabetismo (em %)
- Y: Expectativa de Vida (em anos)

## Expectativa de Vida e Analfabetismo

### Descrição das Variáveis

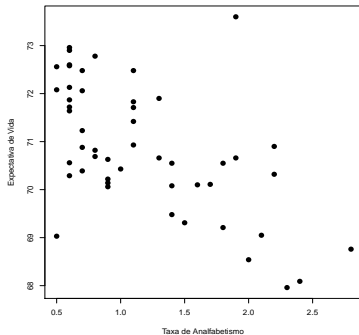
Vamos considerar como ilustração a relação entre

- X: Taxa de Analfabetismo (em %)
- Y: Expectativa de Vida (em anos)

nos 50 estados norte americanos no início dos anos 70.

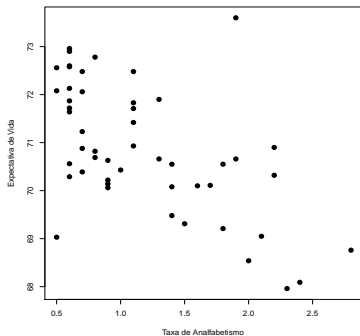
# Expectativa de Vida e Analfabetismo

## Diagrama de Dispersão



# Expectativa de Vida e Analfabetismo

## Diagrama de Dispersão



Podemos notar que, conforme aumenta a **Taxa de Analfabetismo (X)**, a **Expectativa de Vida (Y)** tende a diminuir. Nota-se também uma tendência linear.

## Expectativa de Vida e Analfabetismo

### Algumas Quantidades

Temos que:  $\bar{x} = 1,17$ ;  $\bar{y} = 70,88$ ;  $s_X = 1,342$ ;  $s_Y = 0,609$  e  
 $\sum x_i y_i = 4122,8$ .

## Expectativa de Vida e Analfabetismo

### Algumas Quantidades

Temos que:  $\bar{x} = 1,17$ ;  $\bar{y} = 70,88$ ;  $s_X = 1,342$ ;  $s_Y = 0,609$  e  
 $\sum x_i y_i = 4122,8$ .

### Cálculo da Correlação

Portanto, o coeficiente de correlação linear de Pearson fica dado por

## Expectativa de Vida e Analfabetismo

### Algumas Quantidades

Temos que:  $\bar{x} = 1,17$ ;  $\bar{y} = 70,88$ ;  $s_X = 1,342$ ;  $s_Y = 0,609$  e  $\sum x_i y_i = 4122,8$ .

### Cálculo da Correlação

Portanto, o coeficiente de correlação linear de Pearson fica dado por

$$\begin{aligned}
 r &= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_X s_Y} \\
 &= \frac{4122,8 - 50 \times 1,17 \times 70,88}{49 \times 1,342 \times 0,609} \\
 &= \frac{-23,68}{40,047} = -0,59.
 \end{aligned}$$



# Sumário

3.07pt

- 1 Objetivos da Aula
- 2 Fluência da Fala
- 3 Tabelas de Contingência
- 4 Diagrama de Dispersão
- 5 Correlação Linear
- 6 Reta de Regressão**

# Reta de Regressão

## Motivação

Quando há evidências de relação linear entre as duas variáveis  $Y$  (resposta) e  $X$  (preditora), podemos explicitar a forma dessa relação através da **Reta de Regressão** ajustada aos dados pelo **Método de Mínimos Quadrados** (Legendre, 1805).

# Reta de Regressão

## Definição

A reta de regressão ajustada aos dados é expressa na seguinte forma:

# Reta de Regressão

## Definição

A reta de regressão ajustada aos dados é expressa na seguinte forma:

$$\hat{Y} = a + bX,$$

# Reta de Regressão

## Definição

A reta de regressão ajustada aos dados é expressa na seguinte forma:

$$\hat{Y} = a + bX,$$

em que

# Reta de Regressão

## Definição

A reta de regressão ajustada aos dados é expressa na seguinte forma:

$$\hat{Y} = a + bX,$$

em que

- $\hat{Y}$ : valor predito de  $Y$  para  $X$

## Reta de Regressão

### Definição

A reta de regressão ajustada aos dados é expressa na seguinte forma:

$$\hat{Y} = a + bX,$$

em que

- $\hat{Y}$ : valor predito de  $Y$  para  $X$
- $a$ : intercepto

## Reta de Regressão

### Definição

A reta de regressão ajustada aos dados é expressa na seguinte forma:

$$\hat{Y} = a + bX,$$

em que

- $\hat{Y}$ : valor predito de  $Y$  para  $X$
- $a$ : intercepto
- $b$ : coeficiente angular da reta (tangente do ângulo formado entre a reta ajustada e o eixo-x)



## Descrição Reta de Regressão

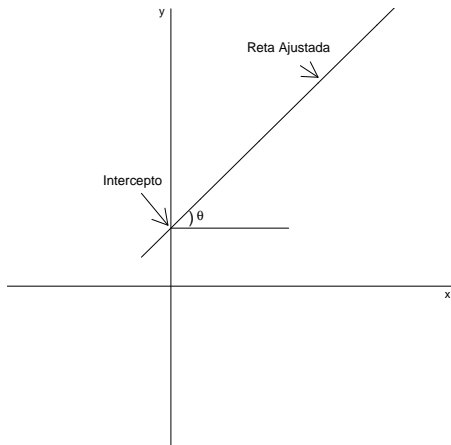
### Descrição Reta de Regressão

Na reta de regressão  $\hat{Y} = a + bX$  descrita abaixo temos que  $b = \text{tg}(\theta)$ .

## Descrição Reta de Regressão

### Descrição Reta de Regressão

Na reta de regressão  $\hat{Y} = a + bX$  descrita abaixo temos que  $b = \text{tg}(\theta)$ .



## Descrição Reta de Regressão

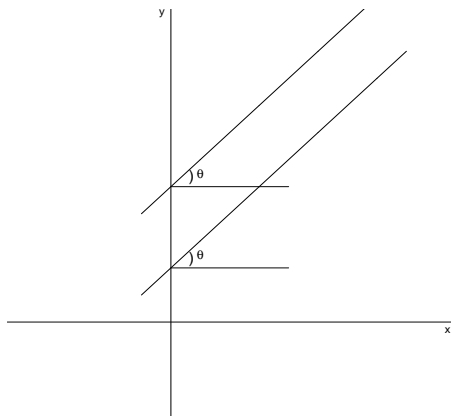
### Retas Paralelas

Nas retas de regressão descritas abaixo temos o mesmo coeficiente angular (mesma inclinação)  $b = \text{tg}(\theta)$ , porém interceptos diferentes.

## Descrição Reta de Regressão

### Retas Paralelas

Nas retas de regressão descritas abaixo temos o mesmo coeficiente angular (mesma inclinação)  $b = \text{tg}(\theta)$ , porém interceptos diferentes.



## Descrição Reta de Regressão

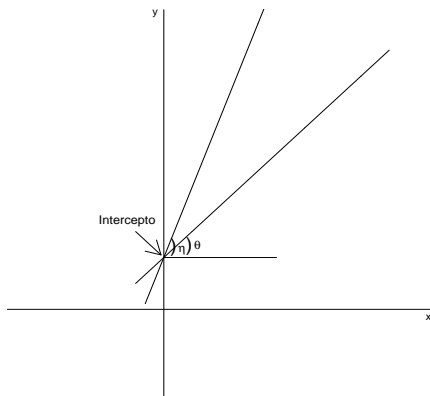
### Retas Concorrentes

Nas retas de regressão descritas abaixo temos o mesmo intercepto, porém coeficientes angulares diferentes,  $b = \text{tg}(\eta)$  e  $b = \text{tg}(\theta)$ , respectivamente.

## Descrição Reta de Regressão

### Retas Concorrentes

Nas retas de regressão descritas abaixo temos o mesmo intercepto, porém coeficientes angulares diferentes,  $b = \text{tg}(\eta)$  e  $b = \text{tg}(\theta)$ , respectivamente.



# Reta de Regressão

## Expressões

O intercepto e o coeficiente angular da reta de regressão ajustada são, respectivamente, expressos nas formas

# Reta de Regressão

## Expressões

O intercepto e o coeficiente angular da reta de regressão ajustada são, respectivamente, expressos nas formas

$$a = \bar{y} - b\bar{x}$$
$$b = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{(n-1)s_X^2}$$



## Reta de Regressão

### Expressões

O intercepto e o coeficiente angular da reta de regressão ajustada são, respectivamente, expressos nas formas

$$a = \bar{y} - b\bar{x}$$
$$b = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{(n-1)s_X^2}$$

### Interpretação de $b$

Para cada aumento de uma unidade de  $X$  espera-se aumento de  $b$  unidades em  $Y$ .

# Reta de Regressão

Relação entre  $b$  e  $r$

$$b = \frac{s_y}{s_x} r.$$

# Reta de Regressão

Relação entre  $b$  e  $r$

$$b = \frac{s_y}{s_x} r.$$

Comentários

## Reta de Regressão

Relação entre  $b$  e  $r$

$$b = \frac{s_y}{s_x} r.$$

Comentários

- o coeficiente angular  $b$  tem o mesmo sinal do coeficiente de correlação linear  $r$

## Reta de Regressão

Relação entre  $b$  e  $r$

$$b = \frac{s_y}{s_x} r.$$

### Comentários

- o coeficiente angular  $b$  tem o mesmo sinal do coeficiente de correlação linear  $r$
- quando  $b = 0$  tem-se  $r = 0$  e vice-versa

## Reta de Regressão

Relação entre  $b$  e  $r$

$$b = \frac{s_y}{s_x} r.$$

### Comentários

- o coeficiente angular  $b$  tem o mesmo sinal do coeficiente de correlação linear  $r$
- quando  $b = 0$  tem-se  $r = 0$  e vice-versa
- se  $X$  e  $Y$  estão padronizadas ( $s_x = s_y = 1$ ) então  $b = r$

## Criminalidade e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

## Criminalidade e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 2,397 + 4,257X,$$



## Criminalidade e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 2,397 + 4,257X,$$

em que

## Criminalidade e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 2,397 + 4,257X,$$

em que

- $\hat{Y}$ : valor predito para a taxa de criminalidade (por 100.000 habitantes)

## Criminalidade e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 2,397 + 4,257X,$$

em que

- $\hat{Y}$ : valor predito para a taxa de criminalidade (por 100.000 habitantes)
- $X$ : taxa de analfabetismo (em %)

## Criminalidade e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 2,397 + 4,257X,$$

em que

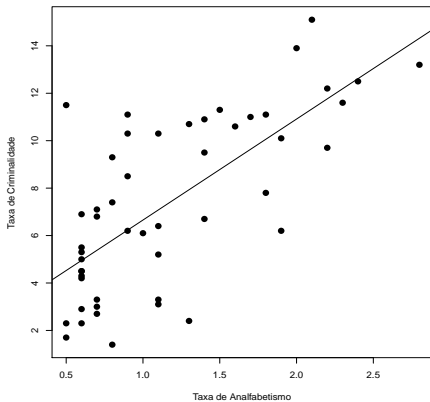
- $\hat{Y}$ : valor predito para a taxa de criminalidade (por 100.000 habitantes)
- $X$ : taxa de analfabetismo (em %)

### Interpretação de $b$

Para um aumento de uma unidade na taxa de analfabetismo ( $X$ ), a taxa de criminalidade ( $Y$ ) aumenta, em média, 4,257 unidades.

# Criminalidade e Analfabetismo

## Reta Ajustada



## Expectativa de Vida e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

## Expectativa de Vida e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 72,395 - 1,296X,$$

## Expectativa de Vida e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 72,395 - 1,296X,$$

em que



## Expectativa de Vida e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 72,395 - 1,296X,$$

em que

- $\hat{Y}$ : valor predito para a expectativa de vida (em anos)

## Expectativa de Vida e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 72,395 - 1,296X,$$

em que

- $\hat{Y}$ : valor predito para a expectativa de vida (em anos)
- $X$ : taxa de analfabetismo (em %)

## Expectativa de Vida e Analfabetismo

### Reta de Regressão

A reta de regressão ajustada fica dada por

$$\hat{Y} = 72,395 - 1,296X,$$

em que

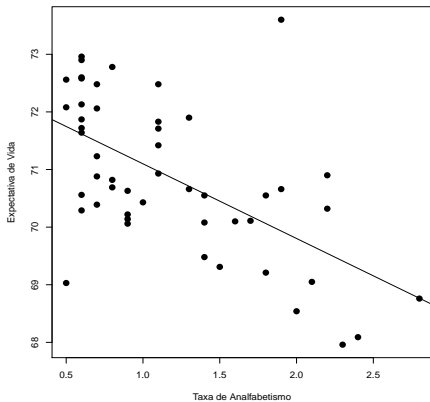
- $\hat{Y}$ : valor predito para a expectativa de vida (em anos)
- $X$ : taxa de analfabetismo (em %)

### Interpretação de $b$

Para um aumento de uma unidade na taxa de analfabetismo ( $X$ ), a expectativa de vida ( $Y$ ) diminui, em média, 1,296 anos.

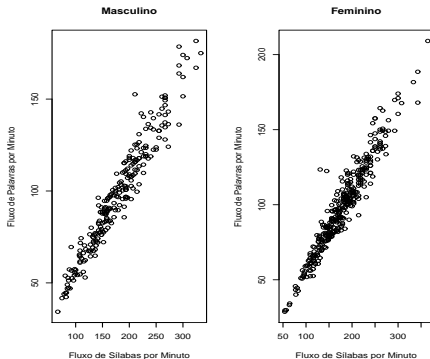
# Expectativa de Vida e Analfabetismo

## Reta Ajustada



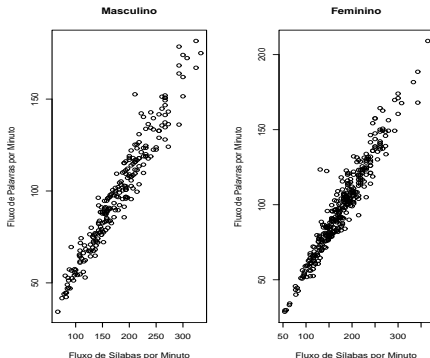
# Relação entre FPM e FSM segundo o Gênero

## Diagrama de Dispersão



## Relação entre FPM e FSM segundo o Gênero

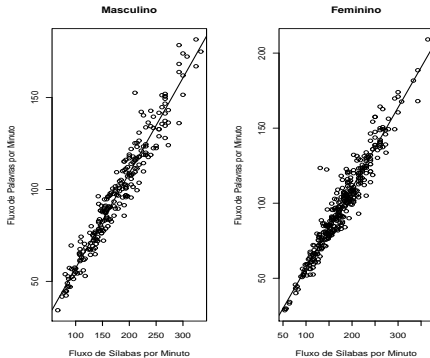
### Diagrama de Dispersão



Coeficiente de correlação linear:  $r = 0,967$  (Masculino) e  $r = 0,962$  (Feminino).

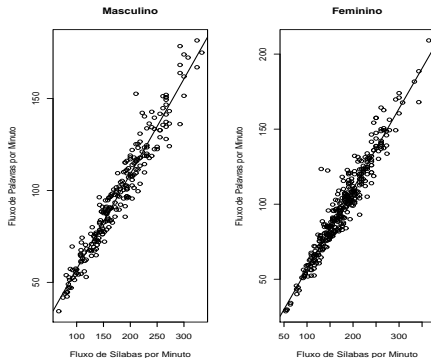
## Relação entre FPM e FSM segundo o Gênero

### Retas Ajustadas



## Relação entre FPM e FSM segundo o Gênero

## Retas Ajustadas

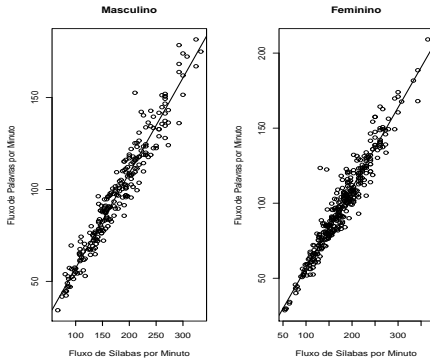


- (Masculino)  $\hat{Y} = 5,503 + 0,517X$



# Relação entre FPM e FSM segundo o Gênero

## Retas Ajustadas



- (Masculino)  $\hat{Y} = 5,503 + 0,517X$
- (Feminino)  $\hat{Y} = 3,135 + 0,535X$

## Relação entre FPM e FSM segundo o Gênero

Interpretação de  $b$

## Relação entre FPM e FSM segundo o Gênero

### Interpretação de $b$

- (Masculino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,517 unidades no FPM

## Relação entre FPM e FSM segundo o Gênero

### Interpretação de $b$

- (Masculino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,517 unidades no FPM
- (Feminino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,535 unidades no FPM

## Relação entre FPM e FSM segundo o Gênero

### Interpretação de $b$

- (Masculino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,517 unidades no FPM
- (Feminino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,535 unidades no FPM

### Predição para $X = 200$

## Relação entre FPM e FSM segundo o Gênero

### Interpretação de $b$

- (Masculino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,517 unidades no FPM
- (Feminino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,535 unidades no FPM

### Predição para $X = 200$

- (Masculino)  $\hat{Y} = 5,503 + 0,517 \times 200 = 108,903$

## Relação entre FPM e FSM segundo o Gênero

### Interpretação de $b$

- (Masculino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,517 unidades no FPM
- (Feminino) para cada aumento de uma sílaba no FSM espera-se aumento de 0,535 unidades no FPM

### Predição para $X = 200$

- (Masculino)  $\hat{Y} = 5,503 + 0,517 \times 200 = 108,903$
- (Feminino)  $\hat{Y} = 3,135 + 0,535 \times 200 = 110,135$