

ESTATÍSTICA

Tabelas e Gráficos

Elisabeti Kira - IME USP

betikira@ime.usp.br

- Coletar dados
censo, técnicas de amostragem, amostra aleatória, amostra não-aleatória, questionários, simulação
- Organizar dados
→ banco de dados ou planilha
- Resumir informação proveniente de dados
medidas resumo
→ tabelas e gráficos
→ diagrama de dispersão e coeficiente de correlação
- Distribuições de probabilidade
- **Inferência Estatística**
intervalo de confiança
→ teste de hipóteses

Recomendações

- Indicar claramente a codificação utilizada em dados qualitativos e as unidades de medida em dados quantitativos.
- Diferenciar observação não coletada (*missing*) de observação nula.
- Em português, a decimal é expressa por ", "
Em inglês, a decimal é expressa por "."
- Ordem de grandeza, precisão e arredondamento para dados quantitativos.

Ex.: R\$ 2.459.841,53 ou R\$ 2,46 milhões ou R\$ 2,5 milhões

Ex.: 70,248 g ou 70,25 g ou 70,2 g

Variável $\left\{ \begin{array}{l} \text{qualitativa (nominal ou ordinal)} \\ \text{quantitativa (discreta ou contínua)} \end{array} \right.$

Dados **qualitativos**

- resumir em tabela de frequência, tabela de contingência (ou dupla entrada).
- representar em gráfico de setores (pizza) ou gráfico de barras.

Dados **quantitativos**

- resumir em tabela de frequência, tabela de contingência (ou dupla entrada) com ou sem intervalos de frequência, calcular medidas resumo (média, desvio padrão, quartis, mediana).
- representar em gráfico de barras, histograma, *boxplot*, diagrama de dispersão, gráfico de perfis.

Tabelas

Recomendações na elaboração de tabelas

1. Organize as tabelas de forma a realçar o que é necessário.
2. Proponha um **título autoexplicativo** e inclua as unidades de medida.

O título deve dizer o que representam os números do corpo da tabela e não deve conter informações que possam ser obtidas diretamente dos rótulos de linhas e colunas.

Compare *Intenção de voto (%) por candidato para diferentes meses* com o título da Tabela 1 abaixo

Tabela 1: Intenção de voto (%)

Candidato	janeiro	fevereiro	março	abril
<i>A</i>	20	18	21	24
<i>B</i>	39	41	40	38
<i>C</i>	8	15	18	22

3. Ordene colunas e/ou linhas quando possível.

Se não houver impedimentos ordene-as segundo os valores, crescente ou decrescentemente.

Compare a Tabela 1 com a Tabela 2

Tabela 2: Intenção de voto (%)

Candidato	janeiro	fevereiro	março	abril
<i>B</i>	39	41	40	38
<i>A</i>	20	18	21	24
<i>C</i>	8	15	18	22

4. Tente trocar de orientação (linhas por colunas) para melhorar a apresentação. É mais fácil fazer comparações ao longo das linhas do que das colunas.

5. Altere a disposição e o espaçamento das linhas e colunas para facilitar a leitura. Inclua um maior espaçamento a cada grupo de linhas e/ou colunas em tabelas muito extensas.

Exemplo: Ehrenberg (1981)

Tabela 3: Vendas quadrimestrais do produto *XXX* em 8 cidades

Libras x1000	1	2	3	4
Bolton	31,3	29,1	25,2	29,3
Edinburgh	135,1	126,9	132,1	208,3
Hull	70,3	81,3	70,9	84,0
Leeds	276,8	258,6	223,0	336,2
Luton	23,5	27,5	22,7	27,1
Plymouth	41,4	44,0	33,2	50,2
Sheffield	233,4	220,1	193,6	220,9
Swansea	62,3	66,4	61,8	76,7

Compare com a Tabela 4 a seguir

Tabelas

Refazendo a tabela com as cidades ordenadas pelo tamanho populacional e arredondando os valores.

Tabela 4: Vendas quadrimestrais, em libras ($\times 1000$), do produto *XXX*. Cidades ordenadas pelo tamanho populacional.

Cidade	Quadrimestre				Média
	1	2	3	4	
Sheffield	233	220	194	221	220
Leeds	277	259	223	336	270
Edinburgh	135	127	132	208	150
Hull	70	81	71	84	76
Swansea	62	66	62	77	67
Plymouth	41	44	33	50	42
Luton	24	28	23	27	25
Bolton	31	29	25	29	29

6. Inclua totais de linhas e/ou colunas para facilitar as comparações. É sempre bom ter um padrão contra o qual os dados possam ser avaliados.

Tabela 5: Número de alunos

Estado civil	Bebida preferida			Total
	não alcoólica	cerveja	outra alcoólica	
solteiro	19 (52,8%)	7 (19,4%)	10 (27,8%)	36 (100,0%)
casado	3 (25,0%)	4 (33,3%)	5 (41,7%)	12 (100,0%)
outros	1 (50,0%)	0 (0,0%)	1 (50,0%)	2 (100,0%)
Total	23 (46,0%)	11 (22,0%)	16 (32,0%)	50 (100,0%)

7. Não utilize mais casas decimais do que o necessário para não mascarar as comparações de interesse.

Compare a Tabela 5 com a Tabela 6 a seguir

Tabela 6: Número de alunos

Estado civil	Bebida preferida			Total
	não alcoólica	cerveja	outra alcoólica	
solteiro	19 (53%)	7 (19%)	10 (28%)	36 (100%)
casado	3 (25%)	4 (33%)	5 (42%)	12 (100%)
outros	1 (50%)	0 (0%)	1 (50%)	2 (100%)
Total	23 (46%)	11 (22%)	16 (32%)	50 (100%)

8. Ao analisar uma tabela comente as principais tendências sugeridas pelos dados, não se limite apenas em descrevê-la.

Por exemplo,

"Os dados da Tabela 6 indicam que a preferência por bebidas alcoólicas é maior entre os alunos casados do que entre os solteiros; além disso, há indicações de que a cerveja é menos preferida que outras bebidas alcoólicas, tanto entre solteiros quanto entre casados."

9. Para variáveis **qualitativas ordinais** é comum incluir uma coluna correspondente às frequências relativas acumuladas nas tabelas de frequências

Tabela 7: Distribuição de frequências de Domínio da Língua Inglesa

Domínio da Língua Inglesa	Frequência observada	Frequência relativa (%)	Frequência relativa acumulada (%)
nenhum	2	4	4
algum	26	54	58
fluente	20	42	100
Total	48	100	

Pela frequência acumulada podemos concluir que 58% dos alunos não tem fluência no idioma inglês.

10. Tabelas de frequências para dados quantitativos **discretos** é similar ao de dados qualitativos nominais.
Se há muitos valores distintos, agrupar em classes.
11. Tabelas de frequências para dados quantitativos **contínuos** requerem agrupamento em classes de frequência.
Não existem regras bem definidas para a divisão em classes.
É razoável utilizar entre 5 e 10 classes.
As amplitudes das classes podem ou não podem ser iguais, dependendo da situação.
12. Para dados contínuos pode-se também construir a tabela de coeficientes de correlação.

Tabelas

Compare as tabelas abaixo

Tabela 8: Audiência entre adultos: correlação (programas ordenados alfabeticamente dentro de canal)

Rede Tupi	Hora do esporte	1.0000	0.1064	0.4741	0.0915
Rede Tupi	Amazing Boxe	0.1064	1.0000	0.3221	0.1885
Rede Torto	Notícias do dia	0.4741	0.3221	1.0000	0.9387
Rede Torto	Noite em destaque	0.0915	0.1885	0.9387	1.0000

Tabela 9: Correlações da audiência para quatro programas de TV

Programa	HoEsp	AmBoxe	NotDia	NoiDest
Hora do esporte		0,1	0,5	0,1
Amazing Boxe	0,1		0,3	0,2
Notícias do dia	0,5	0,3		0,9
Noite em destaque	0,1	0,2	0,9	

Recomendações gerais na construção de gráficos

- Escolha o tipo apropriado para os dados.
- Proponha um **título autoexplicativo**.
- Rotule os eixos apropriadamente, incluindo unidades de medida.
- Escolha as escalas dos eixos para não distorcer a informação que se pretende transmitir. Se o objetivo for comparar as informações de dois ou mais gráficos, use a mesma escala.
- Inclua indicações de "quebra" nos eixos para mostrar que a origem (zero) está deslocada.
- Altere as dimensões do gráfico até encontrar o formato adequado.
- Inclua uma legenda.
- Tome cuidado com a utilização de áreas para comparações, pois elas variam com o quadrado das dimensões lineares.
- Não exagere nas ilustrações do gráfico para evitar "poluição" visual, mascarando seus aspectos mais relevantes.

Recomendações de gráficos de acordo com o tipo de variável (qualitativa ou quantitativa), caso univariado

- Dados **qualitativos nominais**: gráfico de setores (“pizza”) ou gráfico de barras.
- Dados **qualitativos ordinais**: gráfico de barras, ou ainda, gráfico de setores (“pizza”).
- Dados quantitativos **discretos**: gráfico de barras.
- Dados **contínuos**: histograma ou *boxplot*.
O eixo vertical do histograma deve ser a densidade de frequência

Recomendações de gráficos - caso bivariado

Para ver relação entre uma variável qualitativa e outra quantitativa

- histogramas na mesma escala para as categorias da variável qualitativa
- *boxplots* na mesma escala para as categorias da variável qualitativa
- gráfico de perfis

Para ver relação entre duas variáveis quantitativas:

- diagrama de dispersão e gráfico de perfis

Diagrama de Dispersão e Coeficiente de Correlação

Considere as 11 observações das variáveis abaixo

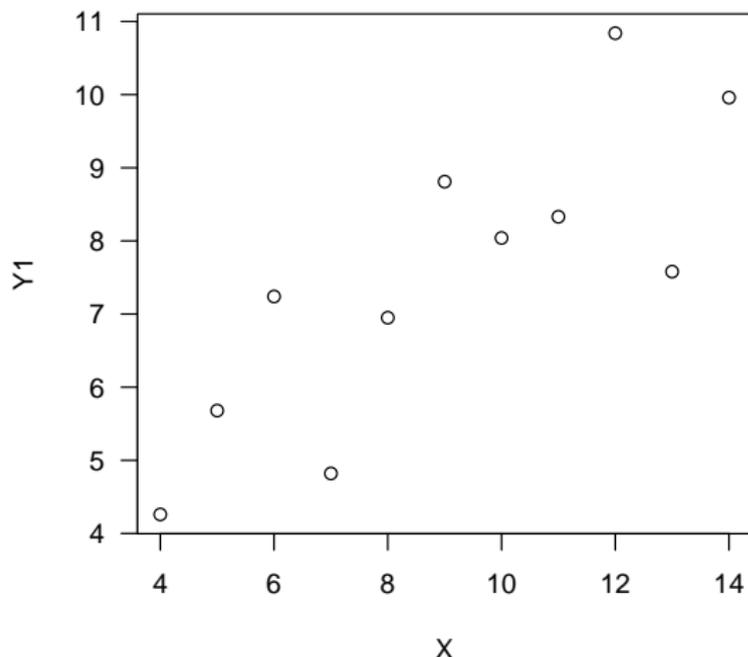
	X	Y_1	Y_2	Y_3	X_4	Y_4
1	10	8,04	9,14	7,46	8	6,58
2	8	6,95	8,14	6,77	8	5,76
3	13	7,58	8,74	12,74	8	7,71
4	9	8,81	8,77	7,11	8	8,84
5	11	8,33	9,26	7,81	8	8,47
6	14	9,96	8,10	8,84	8	7,04
7	6	7,24	6,13	6,08	8	5,25
8	4	4,26	3,10	5,39	19	12,50
9	12	10,84	9,13	8,15	8	5,56
10	7	4,82	7,26	6,42	8	7,91
11	5	5,68	4,74	5,73	8	6,89

Correlação

Calculando o coeficiente de correlação entre as variáveis:

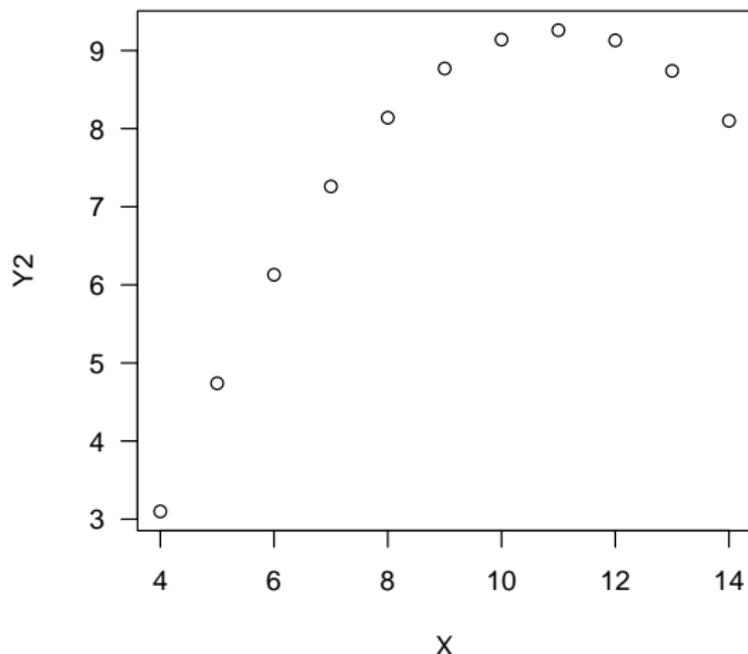
	Correlação
X e Y_1	0,816
X e Y_2	0,816
X e Y_3	0,816
X_4 e Y_4	0,817

Diagrama de Dispersão - X versus Y_1



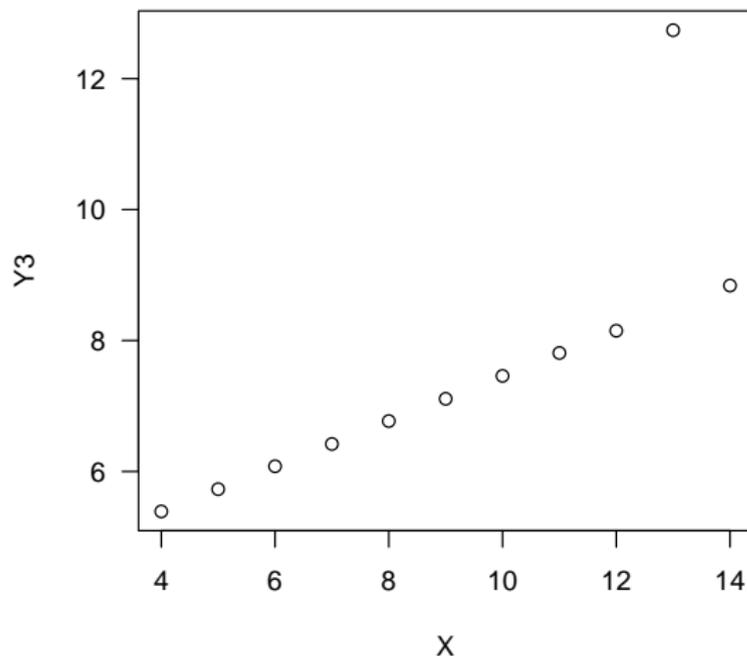
Correlação linear entre X e $Y_1 = 0,816$

Diagrama de Dispersão - X versus Y_2



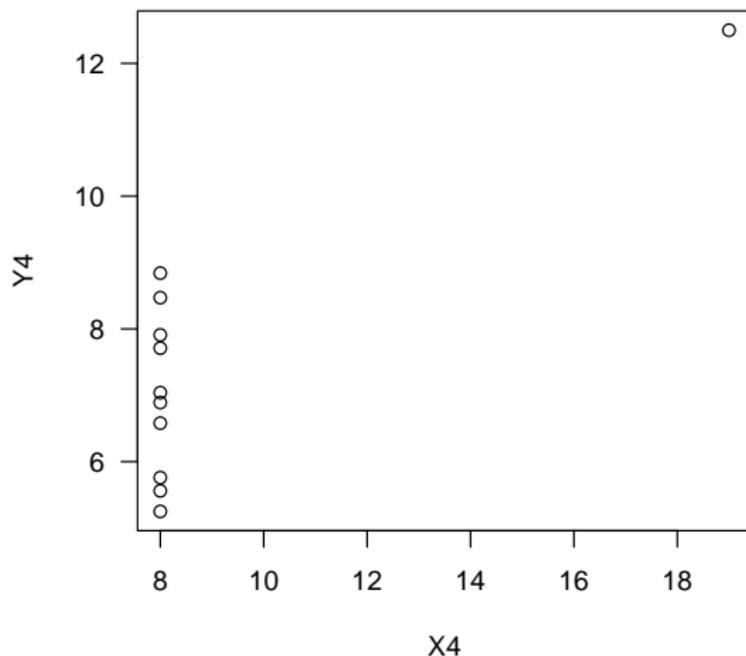
Correlação linear entre X e $Y_2 = 0,816$

Diagrama de Dispersão - X versus Y_3



Correlação linear entre X e $Y_3 = 0,816$

Diagrama de Dispersão - X_4 versus Y_4



Correlação linear entre X_4 e $Y_4 = 0,816$

Teste t

Considere uma amostra aleatória (X_1, \dots, X_n) de $X \sim \mathbf{Normal}(\mu, \sigma^2)$, com μ e σ^2 desconhecidos

Ao se testar as hipóteses

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

a estatística de teste é baseada na média amostral \bar{X} . Usualmente

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}$$

→ verificar se histograma dos dados indica distribuição Normal

→ testes não-paramétricos

Tabela de Contingência e Teste Quiquadrado

Teste de homogeneidade ou independência entre 2 variáveis qualitativas

Tabela 10: Tabela de frequências de *Mudança de Estilo de Aula e Apoio Institucional* em cursos com alunos com deficiência visual

		mudou estilo de aula		Total
		Não	Sim	
Apoio institucional	Não	6	33	39
	Sim	0	8	8
Total		6	41	47

- nem sempre necessita testar
- Aplicação do teste quiquadrado requer caselas com valor esperado maior ou igual a 5.
- Teste exato de Fisher

Referências

Bussab, W. O. & Morettin, P. A., *Estatística Básica*, 8a.edição, Ed. Saraiva, 2013.

Singer, Júlio M., *Sugestões para construção de tabelas e gráficos*, disponível em www.ime.usp.br/~jmsinger

Ehrenberg, A. S. C. (1981). The problem of numeracy. *The American Statistician*, 35, 67-71.