

Universidade Federal de Lavras

Departamento de Estatística

Prof. Daniel Furtado Ferreira

2ª Aula Prática

Coleta, organização e apresentação de dados

- 1) Os dados apresentados a seguir referem-se ao tempo que uma bateria levou para apresentar uma falha grave, em anos, ou seja, para descarregar completamente. A amostra de tamanho $n = 20$ foi obtida com objetivo de caracterizar a robustez das baterias e é dada por:

8,52	4,19	2,52	1,91	8,78	5,91	1,86	12,04	2,60	1,89
5,63	6,36	5,07	3,03	1,83	1,84	12,58	2,03	1,80	1,95

- a) Agrupar os dados do tempo até a falha das baterias em uma distribuição de frequências, determinando o número de classes pelo critério $k = \sqrt{n}$.
- Dado:** $A = X_{(n)} - X_{(1)}$ (amplitude total), $c = A/(k-1)$ (amplitude de classe), $LI_1 = X_{(1)} - c/2$, $LS_1 = LI_1 + c$, $LI_2 = LS_1$, etc.; computar as frequências a partir dos dados originais (de preferência dos dados elaborados, ordenados).
- b) Obter o histograma (classes na abscissa e frequências na ordenada) e o polígono de frequência (linha poligonal unindo os centros dos retângulos) em um mesmo gráfico.
- c) Construir as distribuições de frequências acumuladas: utilizar os limites de classes, calcular as frequências acumuladas abaixo e acima destes limites e construir a tabela correspondente.
- d) Traçar as ogivas no mesmo plano cartesiano. As ogivas são os gráficos correspondentes às frequências acumuladas abaixo e acima (ordenada) dos limites de classes (abscissa).
- e) Qual é a porcentagem de baterias com tempo até falhar superior a 3 anos? Utilizar as ogivas (leitura gráfica) e a interpolação algébrica na distribuição de frequência para responder a esta pergunta. Comparar e discutir os resultados obtidos com a proporção obtida diretamente na amostra.
- f) Acima de qual valor em anos estão 50% das baterias?
- g) Qual a porcentagem de baterias com tempo até falhar inferior a 8 anos?
- h) Obter o tempo em que 20% das baterias falham antes deste valor? Determinar, também, o tempo em que, além dele, apenas 20% das baterias falham.

Obs. Utilize nos casos **1f**, **1g** e **1h** a distribuição de frequência para realizar os cálculos.

- 2) Os dados a seguir referem-se ao número de empresas/ano que decretaram falência observadas em $n = 85$ anos na cidade de Lavras, MG.

Empresas	Frequências
0	36
1	19
2	16
3	7
4	4
5	2
6	1

- a) Obter o gráfico da ocorrência de empresas falidas.
- b) É possível, em sua opinião, encontrar um ano em que mais de seis empresas venham a falir? Justificar sua resposta.
- c) Qual é a natureza da distribuição de frequências? (simétrica, assimétrica à direita ou à esquerda)
- d) Existe diferença entre a variável apresentada neste exercício e a do exercício 1? Se afirmativo, qual é a diferença?

Resolução

1) Antes de realizar-se qualquer análise, ordena-se o conjunto de dados, obtendo:

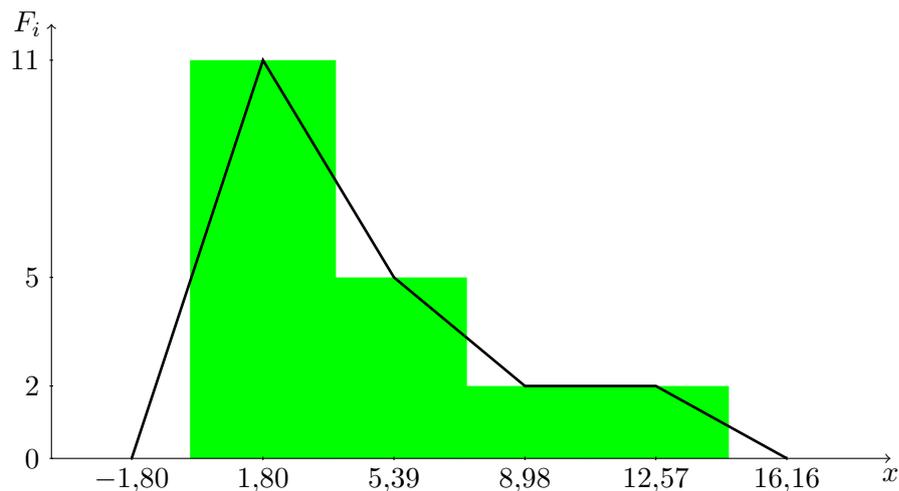
1,80	1,83	1,84	1,86	1,89	1,91	1,95	2,03	2,52	2,60
3,03	4,19	5,07	5,63	5,91	6,36	8,52	8,78	12,04	12,58

a) Para agrupar os dados deve-se obter:

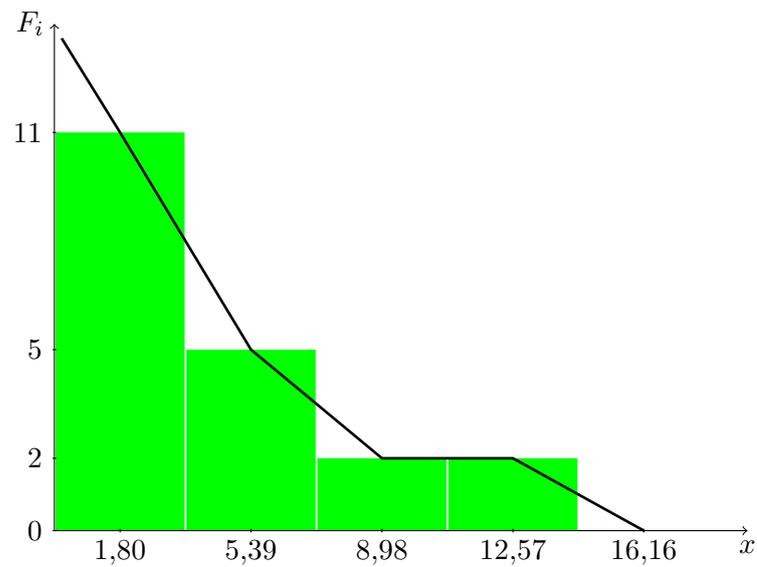
O número de classe é dado por $k = \sqrt{n} = \sqrt{20} \approx 4$ e amplitude total por $A = x_{(20)} - x_{(1)} = 12,58 - 1,80 = 10,78$. Assim, a amplitude de classe é dada por $c = A/(k-1) = 10,78/3 \approx 3,59$ e o limite inferior da primeira classe por $LI_1 = x_{(1)} - c/2 = 1,80 - 3,59/2 = 0,00$. Os demais limites de classe são obtidos somando-se $c = 3,59$ aos limites anteriormente obtidos. A distribuição de frequências obtida desta forma é:

Classes dos tempos	\bar{X}_i	F_i	Fr_i	$Fp_i(\%)$
0,00 † 3,59	1,80	11	0,55	55
3,59 † 7,18	5,39	5	0,25	25
7,18 † 10,77	8,98	2	0,10	10
10,77 † 14,36	12,57	2	0,10	10

b) O histograma e o polígono de frequências foram plotados em um mesmo gráfico, obtendo:



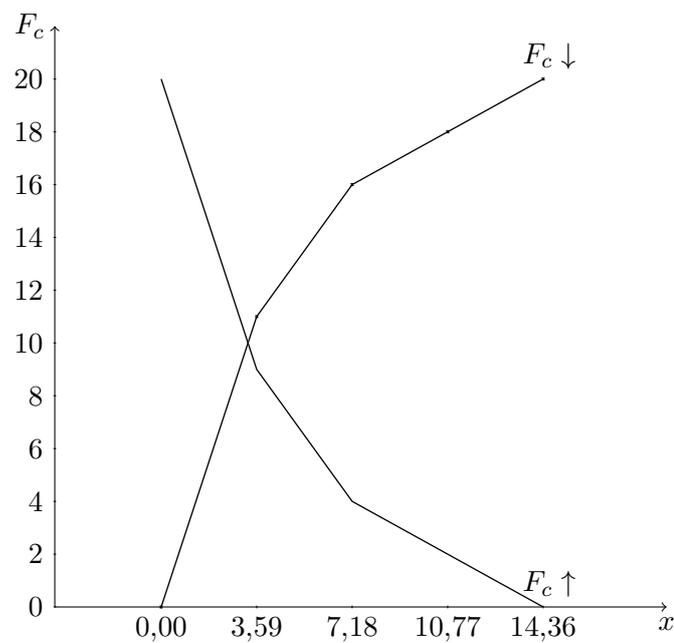
Obs. Esse é um caso típico em que a variável não assume valores negativos e o histograma iniciou-se em zero (limite inferior da primeira classe), portanto, a rigor, o polígono de frequências no seu lado esquerdo não deve tocar na abscissa, caracterizando uma distribuição assimétrica à direita no seu caso mais extremo, ficando assim:



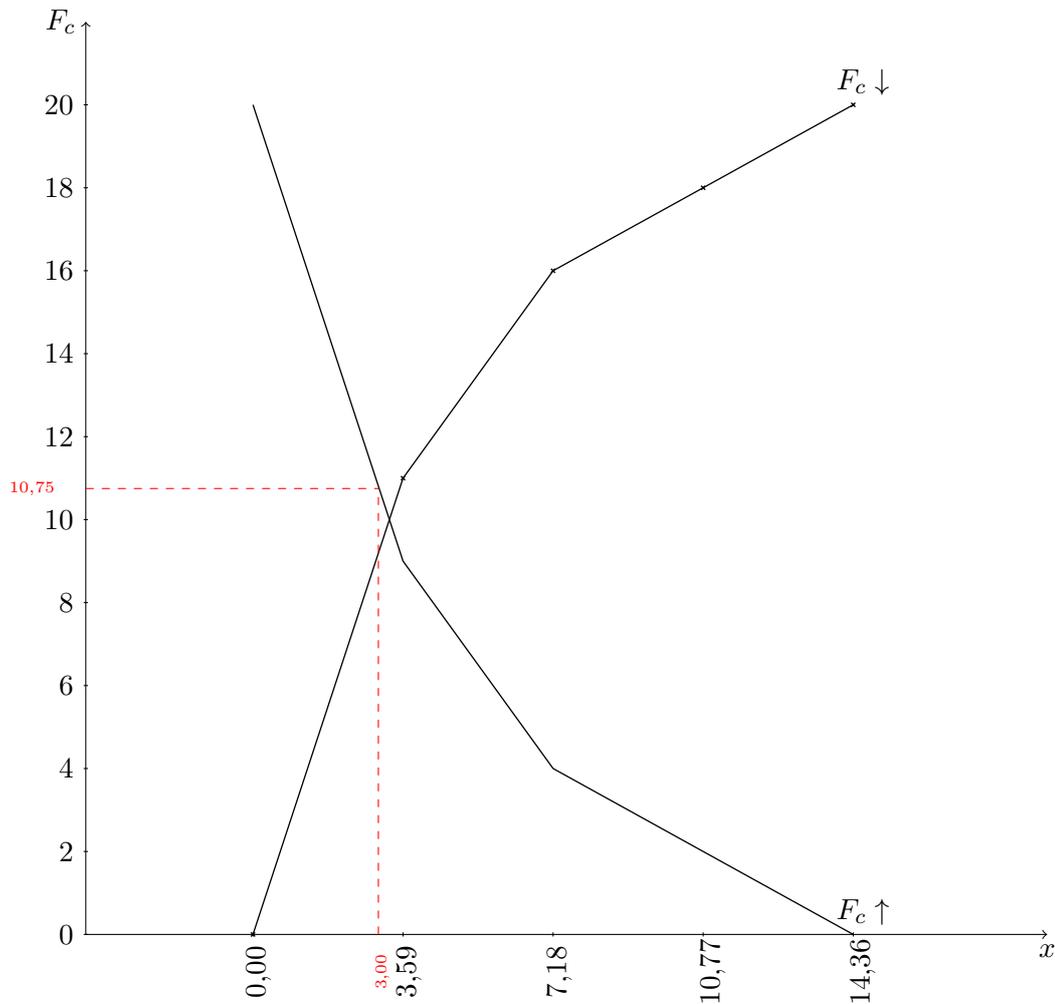
c) A distribuição de frequência acumuladas abaixo de e acima de é

Limites	$FC(X < X_i)$	$FC(X \geq X_i)$
0,00	0	20
3,59	11	9
7,18	16	4
10,77	18	2
14,36	20	0

d) As ogivas são:



e) Acima de 3 anos, pela ogiva, estão aproximadamente 10,75 baterias. Se 20 baterias correspondem a 100%, então 10,75 correspondem a $10,75 \times 100/20\% = 53,75\%$.



Utilizando a interpolação na distribuição de frequências, tem-se: as três últimas classes possuem tempos de falhas superiores a 3. Se as suas frequências forem somadas, têm-se 9 baterias, certamente com tempo de falha superior a 3 anos. Assim, a primeira classe possui 11 baterias com valores entre 0 e 3,59. É necessário quantificar quantas das 11 são superiores a 3 anos, ou seja, quantas possuem tempo de falha entre 3 e 3,59. Para isso faz-se a suposição que a distribuição dos dados em cada classe é uniforme e estima-se a frequência de baterias que superam o tempo de 3 anos na classe em questão. A variação na classe toda é de 3,59, que corresponde a uma frequência de 11 baterias. A variação entre 3 a 3,59 corresponde a variação de tempo de falhas das baterias que possuem valores superiores a 3 anos, sendo $3,59 - 3 = 0,59$. Assim,

Varição	Frequências
3,59	11
0,59	x

Logo, $x = (11 \times 0,59)/3,59 = 1,81$. Portanto, têm-se $1,81 + 9 = 10,81$ baterias com tempo de falha superior a 3 anos, o que corresponde a $10,81 \times 100/20\% = 54,05\%$.

Finalmente, pode-se obter a contagem direta na amostra original. O número de baterias com tempo de falha superior a 3 anos é 10, correspondendo a $10 \times 100/20\% = 50,0\%$.

Os três métodos apresentaram resultados parecidos. A contagem direta na amostra é a mais precisa, mas pode-se observar que a utilização da tabela de frequências e o gráfico das ogivas são, relativamente, eficientes, pois a diferença em pontos percentuais não ultrapassou 5%. É natural que, ao se simplificar a informação por meio de tabelas e gráficos, haja uma perda de precisão, mas espera-se que ainda seja confiável utilizar os dados sumariados para extrair informações úteis a respeito do que se está estudando.

- f) A percentagem de 50% corresponde a 10 baterias. Se for utilizado o seguinte raciocínio: acima de 10,77 estão apenas 2 baterias; acima de 7,18, estão 4 baterias, as duas da classe e as duas da classe posterior; acima de 3,59, estão 9 baterias; e acima de 0,00 estão 20 baterias, as 11 da classe e as 9 das classes posteriores. Assim, pode-se concluir que o tempo de falha que deixa 10 baterias acima dele está entre 0 e 3,59. Se este número for denominado de y , conclui-se que entre y e 3,59 tem de haver 1 bateria, pois acima de y , sem limite superior, tem de haver 10 baterias e acima de 3,59 há precisamente 9. Logo, basta realizar uma regra de três simples. Se a variação entre y e 3,59 for denominada de x , o valor de y poderá ser calculado por $y = 3,59 - x$. Entre 0,00 e 3,59 há uma variação de 3,59 (amplitude de classe) e corresponde a 11 baterias e entre y e 3,59 há uma variação de x , correspondendo a 1 bateria. Logo,

Varição	Frequências
3,59	11
x	1

Logo, $x = (1 \times 3,59)/11 = 0,33$. Portanto, têm-se $y = 3,59 - 0,33 = 3,26$ anos. Portanto 50% das baterias possuem tempo de falha superior a 3,26 anos.

- g) A percentagem de baterias com tempo de falha inferior a 8 anos pode ser computada da seguinte forma. O valor 8 anos pertence a terceira classe e todas as baterias das duas primeiras classes, 16, possuem tempos de falha inferiores a 8. É necessário determinar na terceira classe, quantas das duas baterias possuem tempos de falhas inferiores a 8 anos. Assim, realiza-se a seguinte regra de três:

Varição	Frequências
3,59	2
$8 - 7,18 = 0,82$	x

Logo, $x = (2 \times 0,82)/3,59 = 0,46$ baterias possuem tempo de falha entre 7,18 e 8 anos. Portanto, tem-se $16 + 0,46 = 16,46$ baterias com tempo de falha inferior a 8 anos, totalizando $100 \times 16,46/20 = 82,30\%$ das baterias.

- h) Para se determinar tempo de falha que deixa 20% das baterias, 4 baterias, abaixo dele, tem-se que aplicar uma regra de três semelhante às anteriores. Verifica-se que abaixo de 3,49 anos tem-se 11 baterias, indicando que o tempo de falha almejado está na primeira classe. Assim,

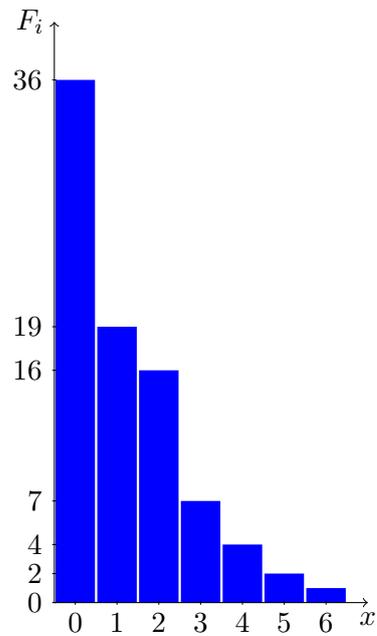
Varição	Frequências
3,59	11
x	4

Logo, $x = (4 \times 3,59)/11 = 1,31$ representa a variação em tempo de falha das baterias entre 0,00 e y anos. Portanto, tem-se 20% das baterias com tempo de falha inferior a $y = 0 + 1,31 = 1,31$ anos.

Para se determinar tempo de falha que deixa 20% das baterias, 4, acima dele, verifica-se que além dos 7,18 anos existem 4 baterias, indicando que o tempo de falha almejado é exatamente este limite. Assim, tem-se que 20% das baterias possuem tempo de falha superior a 7,18 anos.

- 2) A variável número de empresas falidas por ano foi analisada da seguinte forma:

- a) O gráfico da ocorrência de empresas falidas/ano é dado por:



- b) É possível encontrar tal ano, mesmo em uma região (município) que tem baixa incidência de falências de empresas como esse. Para isso é necessário apenas aumentar o tamanho da amostra, pois anos com tal número de falência, pelo que indica os dados e o gráfico anterior, são potencialmente pouco prováveis na população amostrada. Usa-se o padrão de resposta do passado para inferir sobre o futuro.
- c) A natureza da distribuição é assimétrica à direita.
- d) Sim, elas são diferentes. A variável do exercício 1 é quantitativa contínua e a variável do exercício atual é quantitativa discreta.