

Estatística Básica

Lista 1 GABARITO - Introdução à Estatística e Amostragem

Luan Fiorentin

2019-03-05

1. Defina o que é população e o que é amostra. Qual a diferença entre parâmetro populacional e estatística amostral?

*# População é uma coleção de indivíduos, itens ou objetos
que possuem uma característica de interesse.
Amostra é subconjunto da população, mas que possui uma característica
de interesse também. Em geral, a amostra possui dimensões bem menores que o tamanho
da população.
Parâmetro populacional é uma característica numérica que descreve a população.
Estatística é uma característica numérica calculada com base na amostra.*

2. Nos itens abaixo indique se na situação temos uma estatística (E) ou um parâmetro (P).
- (a) Tem-se interesse na altura média dos jogadores de basquete profissional no Brasil. São medidas as alturas de 30% dos jogadores profissionais dos diversos times nacionais, e então a média é calculada.
 - (b) Tem-se interesse na altura média dos jogadores de basquete profissional no Brasil. São medidas as alturas de todos os jogadores profissionais dos diversos times nacionais, e então a média é calculada.
 - (c) O objetivo é obter a biomassa (peso total) de determinada espécie de planta em uma floresta. São pesadas todas as folhas caídas desta espécie, coletados em conglomerados amostrais com 1 m² cada um distribuídos aleatoriamente de forma a cobrir 1% da área da floresta.
 - (d) Tem-se interesse no comprimento médio dos peixes criados em um tanque de cultivo. Na despesca o tanque é esgotado, os peixes são abatidos e medidos, e finalmente se calcula uma média.
 - (e) Foi perguntado para 100 estudantes, selecionados aleatoriamente, o tempo médio de espera na fila do RU. O resultado foi de 20 minutos.
 - (f) Em um estudo de todos os 2223 passageiros do Titanic, verificou-se que 706 sobreviveram quando ele afundou.
 - (g) Há interesse em se estudar a altura de uma espécie de árvore e, após medir 10 delas entre várias presentes no local da coleta, calculamos que, em média, uma árvore mede 4,14 metros.
 - (h) O Senado atual de um país é composto por 87 homens e 13 mulheres.

(resp2)

##	Alternativa	Resposta
## 1	a	E
## 2	b	P
## 3	c	E
## 4	d	P
## 5	e	E
## 6	f	P
## 7	g	E
## 8	h	P

3. Indique que tipo de amostragem (aleatória simples, sistemática, estratificada ou conglomerado) está sendo utilizada em cada um dos casos descritos abaixo.

- (a) Um bairro com 1000 quadras é dividido em 1000 parcelas com 1 quadra cada. São sorteadas 40 parcelas e todos os moradores (maiores de idade) são questionados sobre a eficácia das obras da prefeitura naquele bairro.

- (b) Em uma área arborizada são escolhidas dez árvores ao acaso, e os diâmetros delas são então medidos.
- (c) Em uma indústria de mineração, sedimentos são transportados sobre uma esteira a uma determinada velocidade. Um pesquisador parado ao lado da esteira interrompe seu andamento e coleta uma amostra de 5 em 5 minutos.
- (d) Interessado em estudar os pesos de uma espécie de ave, um pesquisador seleciona ao acaso 30 fêmeas e também 30 machos.
- (e) Uma Universidade fez um estudo sobre o hábito de bebida dos estudantes, selecionando aleatoriamente 10 classes diferentes, e entrevistando todos os alunos de cada uma das classes.
- (f) Um Economista está estudando o efeito da educação sobre o salário, e realiza uma pesquisa com 150 trabalhadores selecionados aleatoriamente de cada uma das seguintes categorias: menos do que o ensino médio, ensino médio, mais do que o ensino médio.

resp3

##	Alternativa	Resposta
## 1	a	Conglomerado
## 2	b	Aleatória Simples
## 3	c	Sistemática
## 4	d	Estratificada
## 5	e	Conglomerado
## 6	f	Estratificada

4. Classifique as variáveis em nominal (N), ordinal (O), discreta (D) ou contínua (C):
- (a) Medidas de diâmetro (mm) de um rolamento industrial.
 - (b) Número de ações da Petrobrás disponíveis no mercado financeiro em 2013.
 - (c) Nomes de empresas de Curitiba com mais de 100 funcionários.
 - (d) Número de estudantes que frequentam o RU diariamente
 - (e) Temperatura das câmaras frias de uma indústria de pescado.
 - (f) Peso de algas coletadas em um determinado banco submerso.
 - (g) Avaliação em graus de qualidade (“bom”, “razoável”, “ruim”) de um equipamento industrial.
 - (h) Tempo que você dispenderá estudando para a prova de estatística.
 - (i) Grau de satisfação da população brasileira com relação ao trabalho de seu presidente (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmenet satisfeito).

resp4

##	Alternativa	Resposta
## 1	a	C
## 2	b	D
## 3	c	N
## 4	d	D
## 5	e	C
## 6	f	C
## 7	g	O
## 8	h	C
## 9	i	O

5. Um professor pergunta a cada um de seus alunos que ramo do conhecimento prefere estudar: Línguas e Literatura (LL), Ciências Exatas (CE), Ciências Físicas e Naturais (FN), Artes e Música (AM). Organize a distribuição das frequências com as respostas abaixo e construa um gráfico de barras.

AM	LL	FN	AM	FN
LL	LL	LL	LL	AM
FN	AM	CE	FN	
CE	FN	FN	AM	
CE	FN	CE	LL	

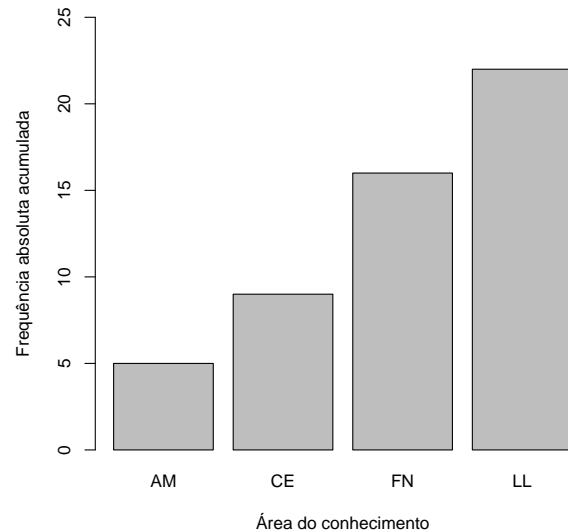
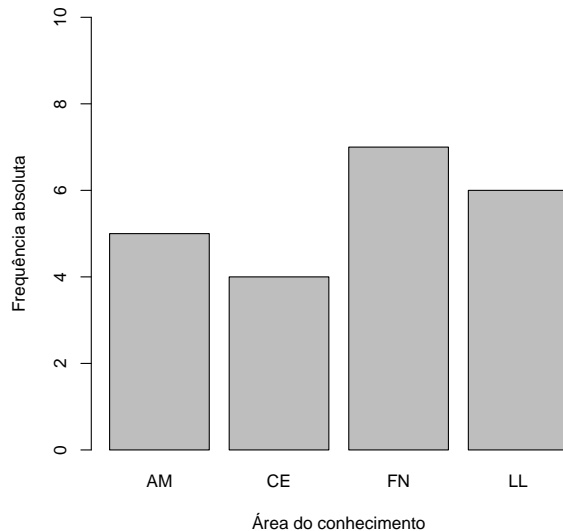
```
# Frequência absoluta
(fa <- addmargins(table(da5)))
```

```
## da5
## AM CE FN LL Sum
## 5 4 7 6 22
```

```
# Frequência absoluta acumulada
(FA <- cumsum(round(table(da5), 2)))
```

```
## AM CE FN LL
## 5 9 16 22
```

```
# Gráfico de barras para frequência absoluta
# e absoluta acumulada
par(mfrow = c(1,2))
barplot(table(da5),
        ylim = c(0,10),
        ylab = "Frequência absoluta",
        xlab = "Área do conhecimento")
barplot(FA, ylim = c(0,25),
        ylab = "Frequência absoluta acumulada",
        xlab = "Área do conhecimento")
```



6. A tabela de distribuições de frequência abaixo apresenta o tempo (em minutos) que uma pessoa leva para encontrar um livro na estante de uma biblioteca, após consultar o sistema e saber o número de referência do livro.
- Complete a tabela com as frequências: acumulada, relativa e relativa acumulada
 - Em quantos minutos (intervalo de minutos) é mais comum as pessoas encontrarem o livro?

- (c) Qual a porcentagem de pessoas que levam menos de 3 minutos para encontrar o livro?
 (d) Quantas pessoas levam menos de 2 minutos para encontrar o livro?
 (e) Qual o percentual de pessoas que levam entre 4 minutos (inclusive) e 4,5 minutos?
 (f) Qual o percentual de pessoas que levam no máximo 1 minuto para encontrar o livro?

Classes	Frequência
[0,5-1,0)	1
[1,0-1,5)	3
[1,5-2,0)	22
[2,0-2,5)	115
[2,5-3,0)	263
[3,0-3,5)	287
[3,5-4,0)	99
[4,0-4,5)	32

(a)

da6

```
##      Classes Freq Fac   fr  Frac
## 1 [0,5-1,0)    1   1 0.001 0.001
## 2 [1,0-1,5)    3   4 0.004 0.005
## 3 [1,5-2,0)   22  26 0.027 0.032
## 4 [2,0-2,5)  115 141 0.140 0.172
## 5 [2,5-3,0)  263 404 0.320 0.492
## 6 [3,0-3,5)  287 691 0.349 0.841
## 7 [3,5-4,0)   99 790 0.120 0.961
## 8 [4,0-4,5)   32 822 0.039 1.000
```

(b)

Entre 3,0 e 3,5 minutos.

(c)

49,2 %

(d)

26 pessoas

(e)

3,9%

(f)

0,1%

7. Construa uma tabela com as distribuições de frequência absoluta, relativa, absoluta acumulada e relativa acumulada usando a amostra do número de páginas de livros infanto-juvenis dada na sequência.

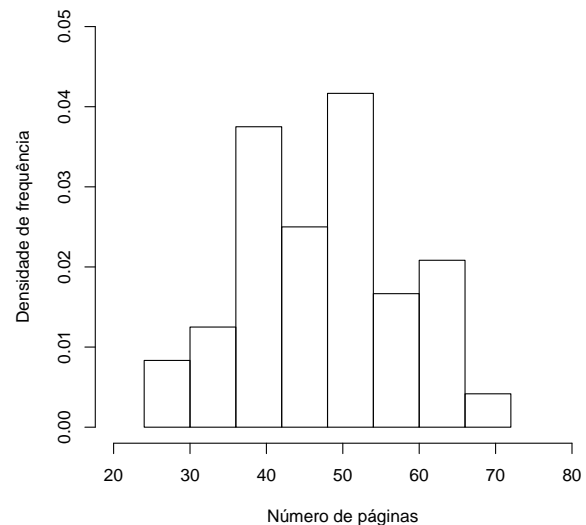
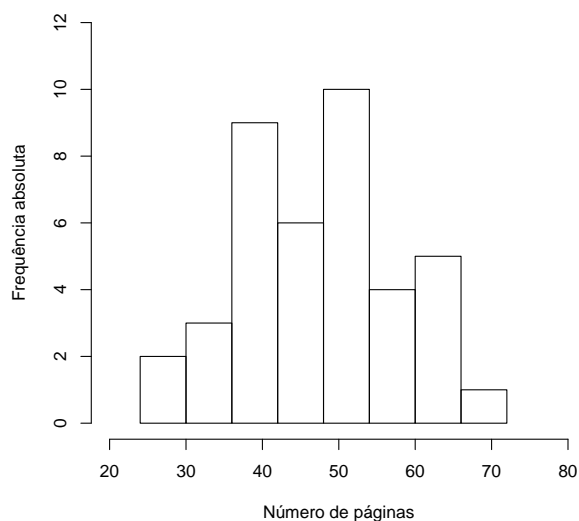
46	46	53	30	62	50	69	49	58	65
62	52	44	38	33	60	50	39	53	50
64	53	45	38	31	41	56	54	38	42
31	38	66	29	41	55	43	50	40	45

Construa um histograma com a densidade de frequência. O que você pode interpretar destes dados a partir da tabela e dos gráficos?

da71

##	Freq.	Freq. ac.	Freq. rel.	Freq. rel. ac.	Dens.
## [24,30)	1	1	0.025	0.025	0.004
## [30,36)	4	5	0.100	0.125	0.017
## [36,42)	8	13	0.200	0.325	0.033
## [42,48)	7	20	0.175	0.500	0.029
## [48,54)	9	29	0.225	0.725	0.038
## [54,60)	4	33	0.100	0.825	0.017
## [60,66)	5	38	0.125	0.950	0.021
## [66,72]	2	40	0.050	1.000	0.008

```
par(mfrow = c(1,2))
hist(da7, breaks = seq(24,72,6),
     ylim = c(0,12), xlim = c(20,80),
     xlab = "Número de páginas",
     ylab = "Frequência absoluta",
     main = "")
hist(da7, breaks = seq(24,72,6), probability = T,
     ylim = c(0,0.05), xlim = c(20,80),
     xlab = "Número de páginas",
     ylab = "Densidade de frequência",
     main = "")
```



8. Qual a diferença entre um gráfico de barras e um histograma?

O gráfico de barras é uma representação gráfica de uma tabela de frequência para dados qualitativos.
O histograma é uma representação gráfica de uma tabela de frequência para dados quantitativos, com intervalos de classes determinados.
O histograma deve ter as barras unidas para representar uma área.