

2025- Lista 2

Exercícios da Seção 1:

1. Classifique cada uma das variáveis abaixo em qualitativa (nominal/ordinal) ou quantitativa (discreta / contínua):

- Ocorrência de hipertensão pré-natal em grávidas com mais de 35 anos (sim ou não são possíveis respostas para esta variável).
- Intenção de voto para presidente (possíveis respostas são os nomes dos candidatos, além de *não sei*).
- Perda de peso de atletas na Corrida de São Silvestre, em quilos.
- Intensidade da perda de peso de maratonistas na Corrida de São Silvestre (leve, moderada, forte).
- 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 totalmente satisfeito).

2. (**sem computador**) Quinze pacientes de uma clínica de ortopedia foram entrevistados quanto ao número de meses previstos de fisioterapia, se haverá (S) ou não (N) sequelas após o tratamento e grau de complexidade da cirurgia realizada: alto (A), médio (M) ou baixo (B). Os dados são apresentados na tabela abaixo:

Pacientes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Fisioterapia	7	8	5	6	4	5	7	7	6	8	6	5	5	4	5
Sequelas	S	S	N	N	N	S	S	N	N	S	S	N	S	N	N
Cirurgia	A	M	A	M	M	B	A	M	B	M	B	B	M	M	A

- Classifique cada uma das variáveis.
- Usando as variáveis (Sequelas e Cirurgia), construa um gráfico de barras e um gráfico de setores.
- Faça um gráfico de ramos-e-folhas (stem) para a variável “Fisioterapia”.
- Faça um gráfico de boxplot usando a variável “Fisioterapia”, somente com os pacientes que ficaram com sequelas. Você acha que essa variável se comporta de modo diferente quando comparado com o grupo de pacientes sem sequelas? Compare os boxplot.
- Verifique se existe relação entre os pacientes que fizeram menos de 6 sessões de fisioterapia e não ficaram com Sequelas. Elabore uma tabela de contingência.
- Elabore uma tabela de contingência relacionando grau de complexidade da cirurgia realizada com a variável Sequelas.

Exercícios da Seção 2:

1. (Use a calculadora/computador) O arquivo "C2-Dados_veneno_(Exercício_23)" contém informações sobre indivíduos que foram expostos ao veneno de um determinado tipo de inseto. Após a exposição, os indivíduos foram submetidos a três diferentes tipos de tratamento, com o objetivo de avaliar a eficácia de cada um na redução dos efeitos do veneno.

Idade: idade do paciente no momento de admissão, em anos.

Diag: tempo, em horas, gasto entre o contato com o inseto e administração do tratamento.

Recup: tempo, em horas, entre a administração do tratamento e recuperação.

Tratam: tipo do tratamento administrado.

Coag: presença de coágulos no momento de admissão.

Importar arquivo "c. 2025--Lista2-dados.veneno_csv"

```
dad=read.csv(file.choose(), header=T, sep=";", dec=".")
```

```
attach(dad);    dim(dad);    str(dad);    head(dad)
```

Pac_No	Idade	Diag	Recup	Tratam	Coag
19	28	7	3	II	nao
4	15	52	45	I	nao
27	76	30	23	III	sim
7	15	53	46	I	sim
14	21	3	2	II	nao
5	11	46	42	I	nao
11	16	55	47	I	nao
10	16	54	47	I	sim
25	47	13	12	III	sim
6	18	59	51	II	nao
16	40	20	11	III	sim
20	24	3	1	II	nao
13	32	9	3	II	nao
15	31	9	3	II	nao
8	10	44	40	I	sim
18	31	9	3	II	sim
12	31	10	4	II	sim
24	46	13	11	III	sim
21	21	1	2	II	sim
22	39	17	8	III	sim
3	15	53	46	I	sim
2	9	42	39	I	nao
23	75	30	22	III	sim
26	54	18	16	III	nao
17	35	12	5	II	sim
9	18	58	50	II	sim

(a). Com base nas informações no arquivo “dados.veneno”, construa um “novo” banco de dados com “**acréscimo**” de **6** indivíduos escolhidos, ao acaso, para todas as variáveis. Use a semente **set.seed(número email)**. Imprima e confira os valores digitados.

Vamos criar um banco de dados “individual” para realização de tarefas.

Para isso, usar número email, NE=01; 02; ... ; 48; 49 (número e-mail)

Exemplo e-mail: ## 123.sergio <sergio@unesp.br>

set.seed(123); sorteio123=sample(1:26, 6,replace=T); sorteio123

[1] 15 19 14 3 10 18 ##Números das linhas selecionadas para 123.sergio

dad2=dad[sorteio123,] ## Novos componentes para amostra do 123.sergio
dad2

	Pac_No	Idade	Diag	Recup	Tratam	Coag
15	8	10	44	40	I	sim
19	21	21	1	2	II	sim
14	15	31	9	3	II	nao
3	27	76	30	23	III	sim
10	6	18	59	51	II	nao
18	24	46	13	11	III	sim

(b). Através do computador, crie uma “nova” planilha com os dados apresentados, **veneno123**.

veneno123=rbind(dad,dad2) ; dim(veneno123) ; head(veneno123)

[1] 32 6

	Pac_No	Idade	Diag	Recup	Tratam	Coag
1	19	28	7	3	II	nao
2	4	15	52	45	I	nao
3	27	76	30	23	III	sim
4	7	15	53	46	I	sim

5 14 21 3 2 II nao
6 5 11 46 42 I não

salvar **veneno123** em arquivo **veneno123.csv**

write.table(veneno123, "D:\\ veneno123_csv.csv", sep = "\t", dec = ".")

veneno123_csv

Pac_No	Idade	Diag	Recup	Tratam	Coag
1	19	28	7	3 II	nao
2	4	15	52	45 I	nao
3	27	76	30	23 III	sim
4	7	15	53	46 I	sim
5	14	21	3	2 II	nao
6	5	11	46	42 I	nao

Marcar a coluna em amarelo → clicar em cima da área com lado direito do mouse → clicar em excluir → clicar opção → deslocar as células à esquerda.

Pac_No	Idade	Diag	Recup	Tratam	Coag
19	28	7	3 II	nao	
4	15	52	45 I	nao	
27	76	30	23 III	sim	
7	15	53	46 I	sim	
14	21	3	2 II	nao	
5	11	46	42 I	nao	

salvar em arquivo **veneno123.csv**

Agora importar arquivo "**veneno123.csv**"

dad=read.csv(file.choose(), header=T, sep=";", dec=".")

attach(dad); dim(dad); str(dad);

Baseando-se nesta nova planilha:

(a). Classifique cada uma das variáveis.

(b). Construa uma tabela de distribuição de frequência para a variável Diag. Utilize k classes pelo método de Sturges, ou seja, $k = 1 + 3.3219 \times \log_e(n)$.

```
hist(Recup, "sturges",freq=T,labels=T)
```

(c). Através de representação gráfica adequada, compare a variável **Recup** entre os pacientes dentro dos três tratamentos. Você diria que as recuperações apresentam distribuição simétrica e mesocúrtica dentro dos três tratamentos (I, II, III)?

```
par(mfrow=c(1,3)) ; Recup1= Recup[Tratam=="I"] ; Recup2= Recup[Tratam=="II"] ;  
Recup3= Recup[Tratam=="III"] ; hist(Recup1, "sturges",freq=T,labels=T) ; hist(Recup2,  
"sturges",freq=T,labels=T) ; hist(Recup3, "sturges",freq=T,labels=T)
```

(d). Calcule medida de cinco números para as variáveis Recup1, Recup2 e Recup3. Calcule as médias amostrais. Qual medida de posição recomendaria? Justifique

```
num5.1=fivenum(Recup1) ; num5.1      ## (min(Recup); Q1; Q2; Q3; Max(Recup))  
num5.2=fivenum(Recup2) ; num5.2  
num5.3=fivenum(Recup3) ; num5.3  
mean(Recup1); mean(Recup2); mean(Recup3)
```

(e). Através de representação gráfica adequada e discuta sobre “homogeneidade” dos três tratamentos. Calcule as variâncias e desvios padrões amostrais. Semi amplitude interquartil $(Q3-Q1)/2$.

```
par(mfrow=c(1,3)) ; boxplot(Recup1) ; boxplot(Recup2) ; boxplot(Recup3)  
var(Recup1); var(Recup2); var(Recup3) ; sd(Recup1); sd(Recup2); sd(Recup3)  
IQR(Recup1)/2; IQR(Recup2)/2; IQR(Recup3)/2
```

(f). Ao invés de trabalhar com a variável Idade, criar uma nova variável denominada **Etaria**, assumindo “Jovem” se a Idade for menor ou igual a 25 anos e “Adulto”, caso contrário.

```
Etaria=Idade<=25;Etaria2=factor(Etaria,levels=c("TRUE","FALSE"),  
labels=c("Jovem","Adulto"))  
Etaria; Etaria2
```

(g). Construa um box-plot para a variável Recup. Depois os boxplot da variável Recup para cada grupo **Etaria**. Com base nas informações úteis do gráfico, você diria que, dependendo da faixa etária, o tempo de recuperação é diferente? Justifique sua resposta.

Calcule medida de cinco números para variável Recup.

```
par(mfrow=c(1,1)) ; boxplot(Recup)
```

```
Recup1= Recup[Etaria2=="Jovem"] ; Recup2= Recup[Etaria2=="Adulto"]
```

```
Recup1; Recup2
```

```
par(mfrow=c(1,2)) ; boxplot(Recup1) ; boxplot(Recup2)
```

(g). Crie uma nova variável denominada **Cura**. A Cura será “Rápida” se Recup for menor ou igual a 10, será “Normal” entre ≥ 10 e < 40 e será “Lenta” para $\text{Recup} \geq 40$.

```
Cura=matrix(0,1)
```

```
for (i in 1:32)
```

```
{
```

```
if (Recup[i] < 10) Cura[i]=0
```

```
else if (Recup[i] $\geq$ 10 & Recup[i] $<$ 40) Cura[i]=1
```

```
else if (Recup[i] $\geq$ 40) Cura[i]=2
```

```
}
```

```
Cura11=factor(Cura); Cura11
```

```
Cura2=factor(Cura11, levels=0:2,
```

```
labels=c("Rápida","Normal","Lenta"))
```

```
Cura2
```

(h). Crie uma tabela de dupla entrada com Cura2 nas linhas e Tratam nas colunas. Com base na tabela, você diria que neste caso a rapidez da cura depende do tipo de tratamento considerado? Justifique.

```
tb1=table(Cura2, Tratam); tb1;
```

```
prop.table(tb1) ; tb1; prop.table(tb1, margin=1) ; prop.table(tb1, margin=2)
```

(h). Verifique, graficamente, se pacientes em cada uma das categorias de **Cura2** apresentam diferenças no que se refere ao tempo entre o contato com o inseto e a administração do tratamento (I, II, III).

```
tb2=table(Tratam , Cura2); tb2;
```

```
barplot(tb2,  
# definir a forma do gráfico lado a lado (beside=T)  
# ou sobreposto (beside=F)  
beside=T,  
# comando para legendar o gráfico  
legend.text=rownames(tb2),  
main="Distribuição de frequências",  
ylab="frequências absolutas",  
xlab="Tratamentos",  
ylim=c(0,20),  
col=c(5,6,7))
```