

CE-003: Estatística II, turma O

1ª Prova - 2º semestre 2011 (19 de outubro de 2011)

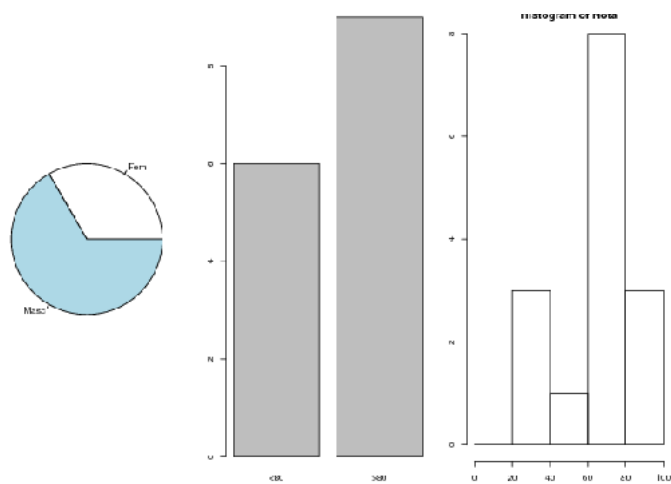
1. Um professor selecionou aleatoriamente 15 alunos. Os alunos foram classificados considerando frequência em 'menor' que 80% e frequência 'maior' que 80%. Além disso, considerou o sexo e a nota na primeira avaliação. Os dados desses 15 alunos estão abaixo:

Sexo	Presença	Nota
Fem	>80%	79
Masc	>80%	69
Masc	>80%	69
Masc	<80%	28
Masc	>80%	83
Fem	<80%	76
Masc	>80%	69
Fem	>80%	38
Masc	<80%	100
Masc	>80%	97
Masc	<80%	48
Masc	<80%	62
Fem	>80%	76
Fem	>80%	72
Masc	<80%	28

- a. Obtenha uma distribuição de frequências de cada uma das variáveis
 b. Faça uma visualização gráfica adequada de cada uma das distribuições de frequências

```
a. > alunos <- read.table("alunos.txt", head=TRUE)
> ## Sexo
> with(alunos, table(Sexo))
Sexo
Fem Masc
 5    10
> with(alunos, prop.table(table(Sexo)))
Sexo
Fem    Masc
0.3333333 0.6666667
> ## Presença
> with(alunos, table(Presença))
Presença
<80 >80
 6    9
> with(alunos, prop.table(table(Presença)))
Presença
<80 >80
0.4 0.6
> ## Nota
> with(alunos, table(cut(Nota, br=(0:5)*20)))
(0,20] (20,40] (40,60] (60,80] (80,100]
 0      3      1      8      3
> with(alunos, prop.table(table(cut(Nota, br=(0:5)*20))))
(0,20] (20,40] (40,60] (60,80] (80,100]
0.00000000 0.20000000 0.06666667 0.53333333 0.20000000

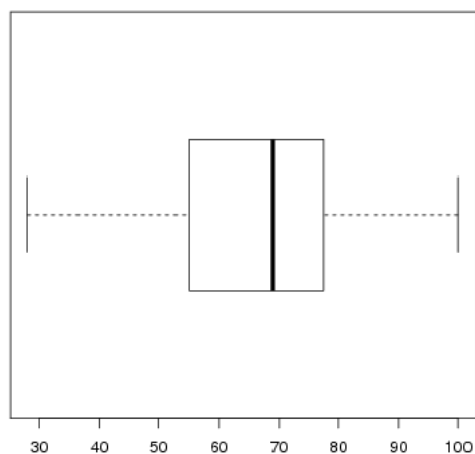
b. > par(mfrow=c(1,3), mar=c(3,3,.5, .5))
> with(alunos, pie(table(Sexo)))
> with(alunos, barplot(table(Presença)))
> with(alunos, hist(Nota, br=(0:5)*20))
```



2. Considere a variável **Nota** dos dados do exercício anterior

- Faça um box-plot
- Faça um ramo-e-folhas
- Calcule a média, variância e coeficiente de variação.

a. `> with(alunos, boxplot(Nota, horizontal=TRUE))`



b. Faça um ramo-e-folhas

`> with(alunos, stem(Nota))`

The decimal point is 1 digit(s) to the right of the |

```

2 | 888
4 | 8
6 | 29992669
8 | 37
10 | 0

```

c. Calcule a média, variância e coeficiente de variação.

`> cat(with(alunos, paste("Média: ", round(mean(Nota), dig=1), " Variância: ", round(var(Nota), dig=1), " CV: ", round(100*sd(Nota)/mean(Nota), dig=1), "%\nMédia: 66.3 Variância: 489.2 CV: 33.4 %`

3. Considere as variáveis **Sexo** e **Presença**

a. Você acha que há associação entre essas duas variáveis? Justifique calculando uma medida de associação e fazendo um gráfico adequado.

`> with(alunos, addmargins(table(Sexo, Presença)))`

```

Presença
Sexo  <80 >80 Sum
Fem   1   4   5
Masc  5   5  10
Sum   6   9  15

```

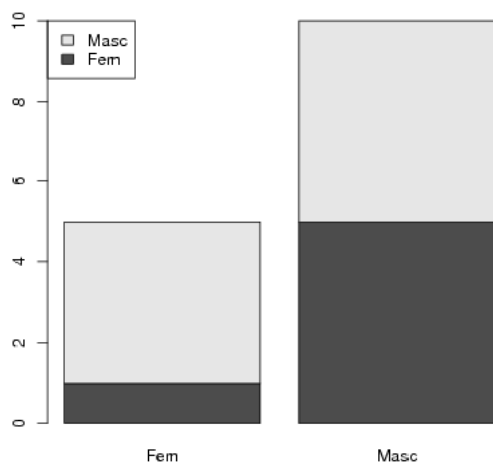
`> with(alunos, prop.table(table(Sexo, Presença), mar=1))`

```

Presença
Sexo  <80 >80
Fem   0.2 0.8
Masc  0.5 0.5

```

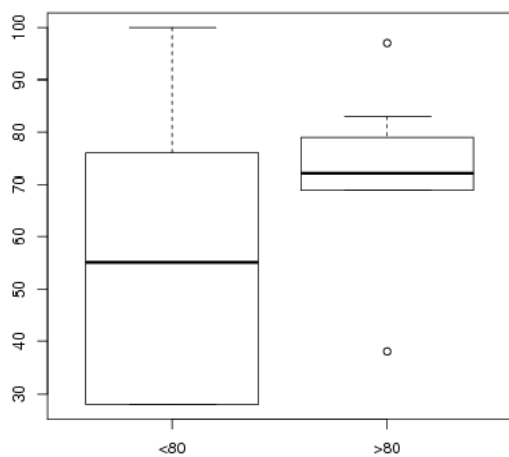
```
> (x2 <- with(alunos, chisq.test(Nota, Presença, sim=T)$stat))
x-squared
12.91667
> ## coeficiente de contingência
> unname(sqrt(x2/nrow(alunos)))
[1] 0.9279607
> with(alunos, barplot(table(Presença, Sexo), legend=levels(Sexo), args.legend=list(x="topleft")))
```



4. Considere as variáveis Presença e Nota

- Faça também um gráfico adequado com as duas variáveis.
- Você acha que há associação entre essas duas variáveis? Justifique calculando medida estatísticas adequadas.

```
> with(alunos, boxplot(Nota ~ Presença))
```



```
> with(alunos, tapply(Nota, Presença, summary))
```

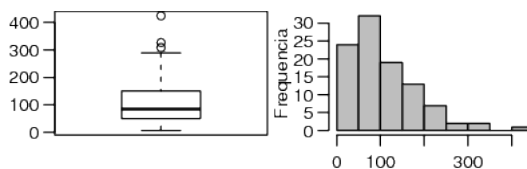
```
 $'<80'
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
28.0	33.0	55.0	57.0	72.5	100.0

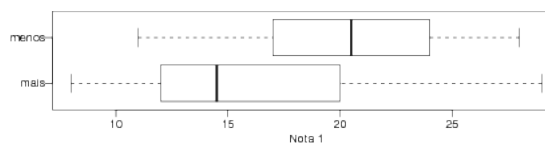
```
 $'>80'
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
38.0	69.0	72.0	72.44	79.0	97.0

5. Considere o histograma e o box-plot a seguir. Comente sobre esses gráficos. A média é maior ou menor que a mediana nesse caso?



6. Considere que um professor selecionou um grupo de alunos e avaliou a nota da primeira prova (cujo valor era 30pts) e a frequência às aulas. Os alunos foram classificados considerando frequência em 'menos' que 80% e frequência em 'mais' que 80%. Após isso, foi construído um box-plot da nota para cada um desses dois grupos. Que se pode concluir analisando os box-plots?



7. Considerando o item anterior, os alunos foram também classificados quanto à nota em dois grupos: maior que 70% e menor que 70% da nota máxima. Após isso, fez-se uma tabela da classificação por frequência do item anterior versus essa classificação por nota. A tabela de 30 alunos sorteados aleatoriamente está apresentada a seguir:

	nota	
presenca	<70%	>70%
<80%	7	4
>80%	5	14

- Calcule uma medida de associação
- Você acha que há associação entre essas duas variáveis? Justifique

```
> chisq.test(matrix(c(7,5,4,14), nc=2), sim=T)$stat
```

```
X-squared
```

```
4.043062
```

```
> prop.table(matrix(c(7,5,4,14), nc=2), mar=1)
```

```
      [,1]      [,2]
```

```
[1,] 0.6363636 0.3636364
```

```
[2,] 0.2631579 0.7368421
```

8. Considerando os 30 alunos do item anterior, a variância da nota desses 30 alunos é 32 e a variância desses 30 alunos com menos e mais de 80% de frequência é 39 e 23 respectivamente. Calcule o percentual da variabilidade de notas que é explicada pela classificação quanto à presença.

```
> 100*var(c(39, 23))/(32*29)
```

```
[1] 13.7931
```