CE-003: Estatística II - Turma: K/O, Prova Final (19/12/2016)

GRR: _______________ **Nome:** ____ **Turma:** ________

1. (15) Considere um jogo com um baralho (52 cartas) no qual em uma primeira rodada retira-se duas cartas e em uma segunda rodada retira-se uma carta. O interesse é se as cartas são figuras (valete, dama ou rei) de qualquer naipe. Obter:
   (a) o espaço amostral;
   (b) a probabilidade de cada ponto amostral;
   (c) a distribuição de probabilidades do número de figuras obtidas nas três cartas.

   Deve-se considerar duas situações, com e sem reposição das cartas entre a primeira e a segunda rodada.

Solução:
Notação:

$$F : \text{a carta é uma figura}$$
$$N = \overline{F} : \text{a carta não é uma figura}$$

- O espaço amostral para as duas situações (com e sem reposição) é o mesmo.

$$\Omega = \{(FF, F); (FF, N); (FN, F); (NF, F); (FN, N); (NF, N); (NN, F); (NN, N)\}$$

- Já as probabilidades são afetadas por repor ou não as cartas

Ponto amostral	(FF,F)	(FF,N)	(FN,F)	(NF,F)	(FN,N)	(NF,N)	(NN,F)	(NN,N)
Com reposição	$\frac{12}{52}\frac{11}{51}\frac{12}{52}$	$\frac{12}{52}\frac{11}{51}\frac{40}{52}$	$\frac{12}{52}\frac{40}{51}\frac{12}{52}$	$\frac{40}{52}\frac{12}{51}\frac{12}{52}$	$\frac{12}{52}\frac{40}{51}\frac{40}{52}$	$\frac{40}{52}\frac{12}{51}\frac{40}{52}$	$\frac{40}{52}\frac{39}{51}\frac{12}{52}$	$\frac{40}{52}\frac{39}{51}\frac{40}{52}$
Sem reposição	$\frac{12}{52}\frac{11}{51}\frac{10}{50}$	$\frac{12}{52}\frac{11}{51}\frac{40}{50}$	$\frac{12}{52}\frac{40}{51}\frac{11}{50}$	$\frac{40}{52}\frac{12}{51}\frac{11}{50}$	$\frac{12}{52}\frac{40}{51}\frac{39}{50}$	$\frac{40}{52}\frac{12}{51}\frac{39}{50}$	$\frac{40}{52}\frac{39}{51}\frac{12}{50}$	$\frac{40}{52}\frac{39}{51}\frac{38}{50}$

-

$$X : \text{número de figuras obtidas nas três cartas}$$
$$x \in \{0, 1, 2, 3\}$$

Com reposição

x	0	1	2	3
P[X=x]	P[(NN,N)]	P[(FN,N)] + P[(NF,N)] + P[(NN,F)]	P[(FF,N)] + P[(FN,F)] + P[(NF,F)]	P[(FFF)]
	$\frac{40}{52}\frac{39}{51}\frac{40}{52}$	$\frac{12}{52}\frac{40}{51}\frac{40}{52} + \frac{40}{52}\frac{12}{51}\frac{40}{52} + \frac{40}{52}\frac{39}{51}\frac{12}{52}$	$\frac{12}{52}\frac{11}{51}\frac{40}{52} + \frac{12}{52}\frac{40}{51}\frac{12}{52} + \frac{40}{52}\frac{12}{51}\frac{12}{52}$	$\frac{12}{52}\frac{11}{51}\frac{12}{52}$
	0,4525	0,4142	0,1218	0,01149

Sem reposição

x	0	1	2	3
P[X=x]	P[(NN,N)]	P[(FN,N)] + P[(NF,N)] + P[(NN,F)]	P[(FF,N)] + P[(FN,F)] + P[(NF,F)]	P[(FFF)]
	$\frac{40}{52}\frac{39}{51}\frac{38}{50}$	$\frac{12}{52}\frac{40}{51}\frac{39}{50} + \frac{40}{52}\frac{12}{51}\frac{39}{50} + \frac{40}{52}\frac{39}{51}\frac{12}{50}$	$\frac{12}{52}\frac{11}{51}\frac{40}{50} + \frac{12}{52}\frac{40}{51}\frac{11}{50} + \frac{40}{52}\frac{12}{51}\frac{11}{50}$	$\frac{12}{52}\frac{11}{51}\frac{10}{50}$
	0,4471	0,4235	0,1195	0,009955

OBS: no caso sem reposição a v.a. X segue uma distribuição hipergeométrica e as probabilidades podem ser obtidas pela função de probabilidade desta distribuição.

$$X \sim \text{HG}(N = 52, n = 3, k = 12)$$

$$P[X = x] \sim \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$P[X = 0] = \frac{\binom{12}{0}\binom{52-12}{3-0}}{\binom{52}{3}} = 0,4471$$

$$P[X = 1] = \frac{\binom{12}{1}\binom{52-12}{3-1}}{\binom{52}{3}} = 0,4235$$

$$P[X = 2] = \frac{\binom{12}{2}\binom{52-12}{3-2}}{\binom{52}{3}} = 0,1195$$

$$P[X = 3] = \frac{\binom{12}{3}\binom{52-12}{3-3}}{\binom{52}{3}} = 0,009955$$

2. (15) Um vendedor consegue vender, em média, 0,5 unidades de um produto por dia. Calcule as probabilidades de:
   (a) vender alguma unidade em um particular dia;
   (b) não efetuar nenhuma venda em uma semana (considere a semana tendo cinco dias úteis);
   (c) em uma semana (cinco dias úteis) efetuar vendas em ao menos três dias.

Solução:

(a)

$$
\begin{aligned}
X_1 &: \text{número de vendas em um dia} \\
x_1 &\in \{0, 1, 2, \ldots\} \\
X_1 &\sim \mathrm{P}(\lambda = 0,5) \\
P[X_1 = 0] &= \frac{\mathrm{e}^{0,5} 0,5^0}{0!} = 0,6065 \\
P[X_1 \geq 1] &= 1 - P[X_1 = 0] = 0,3935
\end{aligned}
$$

(b)

$$
\begin{aligned}
X_2 &: \text{número de vendas em uma semana (cinco dias)} \\
x_2 &\in \{0, 1, 2, \ldots\} \\
X_2 &\sim \mathrm{P}(\lambda = 2,5) \\
P[X_2 = 0] &= \frac{\mathrm{e}^{2,5} 2,5^0}{0!} = 0,08208
\end{aligned}
$$

Solução alternativa (sob independência):

$$
P[X_1 = 0 \cap X_1 = 0 \cap X_1 = 0 \cap X_1 = 0 \cap X_1 = 0] \overset{ind.}{=} (P[X_1 = 0])^5 = 0,3935 = 0,08208
$$

(c)

$$
\begin{aligned}
X_3 &: \text{número de dias com vendas em uma semana (cinco dias)} \\
x_3 &\in \{0, 1, 2, 3, 4, 5\} \\
X_3 &\sim \mathrm{B}(n = 5, p = P[X_1 = 0] \\
P[X_3 \geq 3] &= P[X_3 = 3] + P[X_3 = 4] + P[X_3 = 5] = 0,6938
\end{aligned}
$$

Soluções computacionais (linguagem R):

```
> (q1 <- ppois(0, lambda=0.5, lower=FALSE))

[1] 0,3935

> (q2 <- ppois(0, lambda=0.5*5))

[1] 0,08208

> (q2a <- (ppois(0, lambda=0.5)^5))

[1] 0,08208

> (q3 <- pbinom(2, size=5, prob=dpois(0, lambda=0.5), lower=FALSE))

[1] 0,6938
```

3. (20) O diagrama ramo-e-folhas abaixo mostra medidas do fluxo anual do rio Nilo próximo à cidade de Ashwan no período de 1871-1970.

```
A casa decimal esté 2 digitos à direita de |

   4 | 6
   5 |
   6 | 5899
   7 | 000123444455667778
   8 | 0000112222333445555556667779
   9 | 001122224446667889 9
  10 | 0122234455
  11 | 00012244566678
  12 | 112356
  13 | 7
```

(a) Obtenha a mediana e quartis dos dados

(b) Obtenha o máximo, mínimo, primeiro e nono decis

(c) Faça um diagrama *box-plot* dos dados

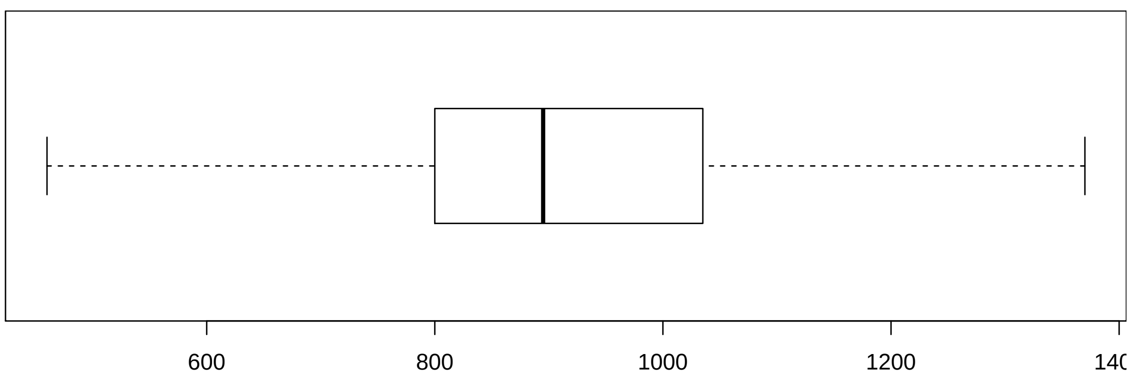
(d) O que pode ser dito da distribuição dos dados baseando-se nos gráficos e medidas?

Solução:

(a)
```
Mediana 1o Quartil 3o Quartil
    895        800       1035
```

(b)
```
  Min      Max 1o decil 9o decil
  460     1370      725     1160
```

(c) boxplot:



4. (25) Considere o seguinte problema (Magalhães & Lima, 2006):
Um fabricante afirma que sua vacina contra gripe imuniza em 80% dos casos. Uma amostra de 25 indivíduos entre os que tomaram a vacina foi sorteada e testes foram feitos para verificar a imunização ou não desses indivíduos. Se o fabricante estiver corrreto, qual é a probabilidade de proporção de imunizados ser inferior a 0,75? E superior a 0,85?

(a) No contexto do problema identifique:
- a população
- o parâmetro de interesse
- o estimador
- a estimativa
- a distribuição amostral

(b) Responda as perguntas propostas no problema

(c) Quais seriam as estimativas pontuais e intervalares se fosse observados 18 imunizados dentre os 25 avaliados?

(d) Qual deveria ser o tamanho da amostra em um novo estudo para que a margem de erro fosse de no máximo 0,03?

(e) Suponha que se adote o critério de refutar a afirmativa do fabricante caso sejam observados 17 ou menos não imunizados. Qual a probabilidade de refutar a afirmativa mesmo quando ela é verdadeira (a vacina de fato imuniza 80%)?

(f) Suponha agora que a imunização real seja de apenas 70%. Qual a probabilidade de mesmo assim não refutar a afirmativa do fabricante?

Solução:

$$X: \text{ imunizado (0: não, 1: sim)}$$
$$x \in \{0,1\}$$
$$X \sim \text{Ber}(p)$$

(a)
- Os indivíduos que receberam a vacina.
- A proporção (p) de indivíduos imunizados entre todos os que receberam a vacina (*na população*).
- O cálculo da proporção de indivíduos imunizados na amostra $\hat{p} = \sum_{i=1}^{n} X_i/n$.
- A proporção observada em uma determinada amostra (no caso na amostra de $n = 25$ indivíduos).
- A distribuição amostral do estimador, ou seja, a distribuição que seria obtida caso fossem obtidas estimativas de *diversas* amostras.

(b)

(c)
- Usando $p = 0,80$

$$\hat{p} = \frac{18}{25} = 0,72$$
$$(\text{I.C.95\%})\hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$$
$$0,72 \pm 1,96\sqrt{\frac{0,80(1-0,80)}{25}}$$
$$0,72 \pm 0,157$$
$$(0,563 \quad ; \quad 0,877)$$

-

$$\hat{p} = \frac{18}{25} = 0,72$$
$$(\text{I.C.95\%}) : \hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$$
$$0,72 \pm 1,96\sqrt{\frac{0,72(1-0,72)}{25}}$$
$$0,72 \pm 0,176$$
$$(0,544 \quad ; \quad 0,896)$$

(d)

$$z\sqrt{\frac{p(1-p)}{n}} = 0,03$$
$$1,96\sqrt{\frac{0,80(1-0,80)}{n}} = 0,03$$
$$n = \frac{1,96^2}{0,03^2}0,80(1-0,80) = 683$$

(e)

$$P[X \leq 17|p = 0,80] \overset{\text{Binomial}}{=} \sum i = 0^{17}\binom{25}{i}(0,8)^i(1-0,8)^{25-i} = 0,109$$
$$\overset{\text{normal}}{\approx} P[\hat{p} \leq 17,5/25] = 0,106$$

(f) $P[X \geq 18|p = 0,70] \approx P[\hat{p} > 17,5/25] = 0,894$

$$P[X > 17|p = 0,70] \overset{\text{Binomial}}{=} \sum i = 18^{25}\binom{25}{i}(0,7)^i(1-0,7)^{25-i} = 0,512$$
$$\overset{\text{normal}}{\approx} P[\hat{p} \leq 17,5/25] = 0,5$$

5. (25) Um conjunto de dados (*Womenlf)* disponível no pacote **car** do programa estatístico ***R*** possui registros de condições relacionadas apo trabalho de 263 mulheres no Canadá. Os atributos são: o tipo de trabalho (externo em tempo integral *fulltime*, externo em tempo parcial *parttime*, não trabalha fora de casa *not.work*), o salário do marido ((hincome)), se possui ou não filhos (com filho(s) *present*, sem filho(s) *absent*), a região do país (dividido em cinco regiões). Uma última coluna na tabela de dados apresenta uma categorização do salário do marido. Abaixo são mostrados: um extrato da tabela com os 10 primeiros registros e um resumo univariado de todos os 263 dados.

Discuta quais seriam as possíveis questões de interesse a serem examinadas com procedimentos estatísticos com este conjunto de dados relacionando (ao menos) duas variáveis. Verifique, interprete e discuta os resultados fornecidos a seguir.

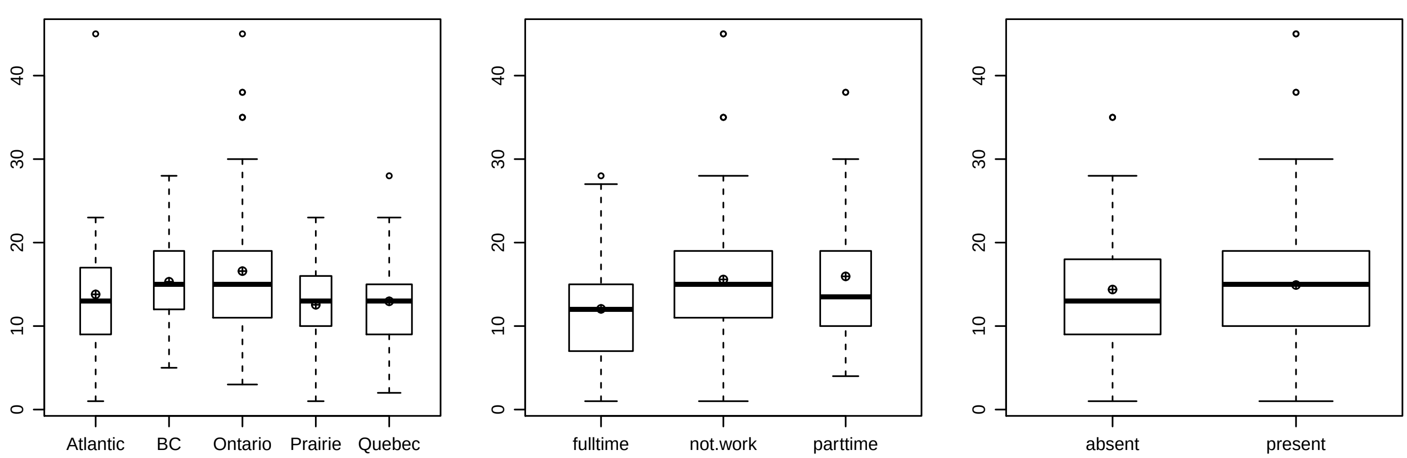
```
     partic hincome children  region classHI
1  not.work      15  present Ontario (10,15]
2  not.work      13  present Ontario (10,15]
3  not.work      45  present Ontario (20,50]
4  not.work      23  present Ontario (20,50]
5  not.work      19  present Ontario (15,20]
6  not.work       7  present Ontario  (5,10]
7  not.work      15  present Ontario (10,15]
8  fulltime       7  present Ontario  (5,10]
9  not.work      15  present Ontario (10,15]
10 not.work      23  present Ontario (20,50]
```

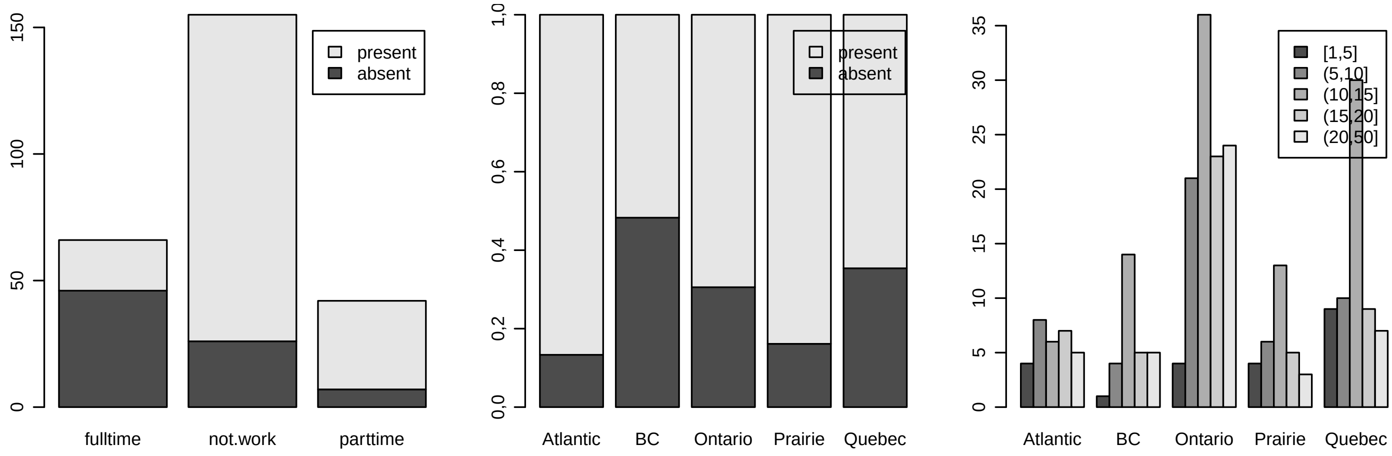
```
      partic       hincome        children         region       classHI
 fulltime: 66   Min.   : 1,0   absent : 79   Atlantic: 30   [1,5]  :22
 not.work:155   1st Qu.:10,0   present:184   BC      : 29   (5,10] :49
 parttime: 42   Median :14,0                 Ontario :108   (10,15]:99
                Mean   :14,8                 Prairie : 31   (15,20]:49
                3rd Qu.:19,0                 Quebec  : 65   (20,50]:44
                Max.   :45,0
```

	fulltime	not.work	parttime	Sum
absent	46,00	26,00	7,00	79,00
present	20,00	129,00	35,00	184,00
Sum	66,00	155,00	42,00	263,00

	Atlantic	BC	Ontario	Prairie	Quebec	Sum
absent	0,05	0,18	0,42	0,06	0,29	1,00
present	0,14	0,08	0,41	0,14	0,23	1,00
Sum	0,19	0,26	0,83	0,20	0,52	2,00

	[1,5]	(5,10]	(10,15]	(15,20]	(20,50]	Sum
Atlantic	0,18	0,16	0,06	0,14	0,11	0,66
BC	0,05	0,08	0,14	0,10	0,11	0,48
Ontario	0,18	0,43	0,36	0,47	0,55	1,99
Prairie	0,18	0,12	0,13	0,10	0,07	0,61
Quebec	0,41	0,20	0,30	0,18	0,16	1,26
Sum	1,00	1,00	1,00	1,00	1,00	5,00






```
        Pearson's Chi-squared test

data:  with(Womenlf, table(partic, region))
X-squared = 5,4, df = 8, p-value = 0,7

        Pearson's Chi-squared test

data:  with(Womenlf, table(children, partic))
X-squared = 66, df = 2, p-value = 5e-15
```
