UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE DESENVOLVIMENTO TECNOLÓGICO
ESTATÍSTICA BÁSICA PARA COMPUTAÇÃO – 1ª PROVA

Nome: Gabarito Data: 23/04/2014

Questão 1 (3,0). Considere os dados da tabela abaixo.

Tabela 1. Precipitação pluviométrica (em mm) ocorrida em Pelotas, nos meses de janeiro e dezembro, num período de cinco anos (1951-55).

Mês	**Ano**				
	1951	**1952**	**1953**	**1954**	**1955**
Janeiro	174	65	149	162	24
Dezembro	49	22	9	69	40

a) (0,2) Classifique a série estatística apresentada na tabela.

Série mista temporal temporal (ou histórica histórica)

b) (1,3) Calcule a média, a amplitude total, a variância, o desvio padrão e o coeficiente de variação para os valores referentes ao mês de **Dezembro**, e **interprete** o valor do **desvio padrão**. (Mostrar expressões matemáticas, cálculos e as unidades de medida).

Média = 37,8 mm
a_t = 60 mm
s^2 = 545,7 mm^2
s = 23,36 mm
CV = 61,8%
O desvio padrão expressa a variação média dos dados em torno da média aritmética. A precipitação pluviométrica média no mês de dezembro foi de 37,8 mm, com variação média de 23,36 mm abaixo e acima desta média.

c) (0,4) Enuncie uma propriedade aritmética da média e uma propriedade da variância.

- Se somarmos uma constante **c** a todos os valores de um conjunto de dados, sua média também é somada (ou subtraída) por esta constante.

- Se somarmos uma constante **c** a todos os valores de um conjunto de dados, a variância destes dados não se altera.

d) (0,4) Sabe-se que no mês de **Janeiro** a precipitação média foi de 114,8 e a variância foi 4406,7. Em qual dos dois meses a precipitação pluviométrica **variou mais** ao longo dos anos? Justifique sua resposta.

$CV_{Janeiro}$ = 57,8%
$CV_{Dezembro}$ = 61,8%
Variou mais no mês de Dezembro, pois apresentou o maior coeficiente de variação.

e) (0,7) Construa um gráfico para representar a precipitação pluviométrica ao longo dos anos nos meses de janeiro e dezembro. (Siga as normas de representação gráfica.)

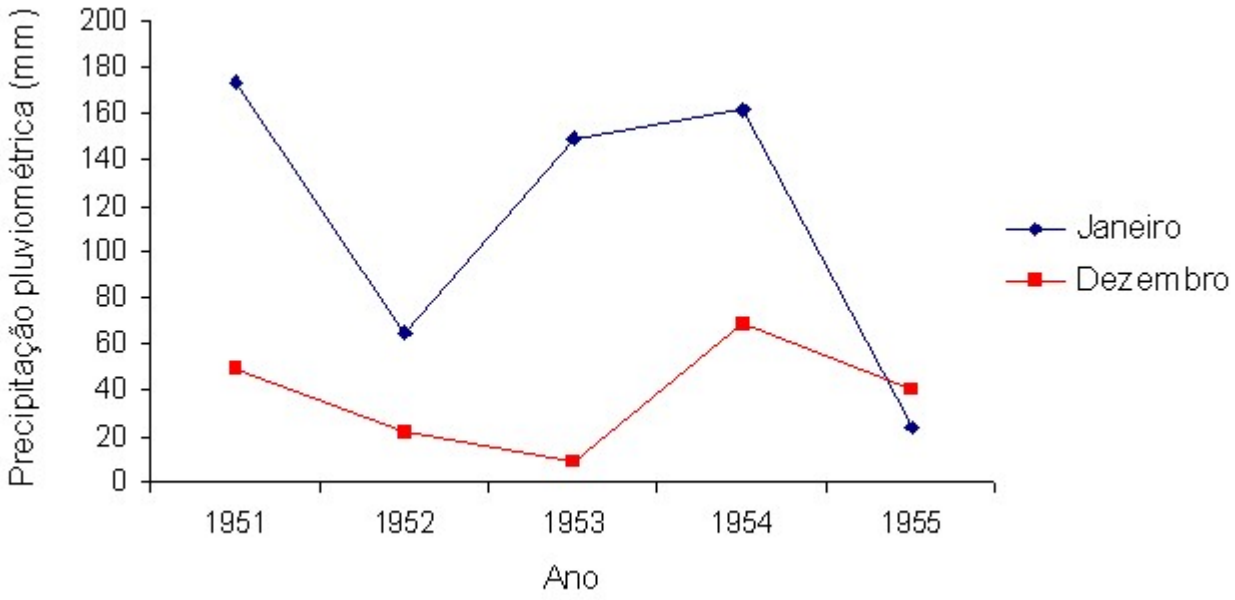


Figura 1. Precipitação pluviométrica (em mm) ocorrida em Pelotas, nos meses de janeiro e dezembro, num período de cinco anos (1951-55).

Questão 2 (4,7). Os dados da tabela abaixo (ordenação vertical) se referem às quantidades (em toneladas) de óxidos de enxofre emitidas por uma indústria em 70 dias.

6,2	11,8	15,5	18,0	20,0	22,7	25,7
7,7	12,3	15,8	18,0	20,1	22,7	25,9
8,3	12,8	15,9	18,1	20,1	22,9	26,1
9,0	13,2	16,2	18,5	20,4	23,0	26,4
9,4	13,3	16,7	18,7	20,5	23,5	26,6
9,8	13,5	16,9	19,0	20,9	23,7	26,8
10,5	13,9	17,3	19,1	21,4	23,9	27,5
10,7	14,5	17,5	19,2	21,6	24,1	28,5
11,0	14,7	17,6	19,3	21,9	24,6	28,6
11,2	15,2	17,9	19,4	22,3	24,8	31,8

Onde: $\sum x_i = 1302{,}60$; $\sum (x_i - \bar{x})^2 = 2304{,}00$; $\sum (x_i - \bar{x})^3 = -967{,}03$ e $\sum (x_i - \bar{x})^4 = 178430{,}47$.

a) (0,3) Classifique a variável em estudo e identifique a escala de medida e a unidade de observação.

Variável numérica contínua; escala de razão; unidade de observação: dia

b) (0,3) Calcule as seguintes quantidades (indicando as unidades de medida):

$(x_{(5)} - \bar{x})^2 = 84{,}82\ t^2$

$(x_{(5)} - \bar{x})^3 = -781{,}23\ t^3$

$(x_{(5)} - \bar{x})^4 = 7195{,}13\ t^4$

c) (1,1) Calcule os momentos centrados na média (m_2, m_3 e m_4) e os coeficientes de assimetria e curtose, indicando as expressões matemáticas e as unidades de medida. A partir desses coeficientes, classifique a distribuição quanto ao formato.

$m_2 = 32{,}91\ t^2$
$m_3 = -13{,}81\ t^3$
$m_4 = 2549{,}01\ t^4$
$a_3 = -0{,}07315$
$a_4 = 2{,}354$

Distribuição simétrica e platicúrtica

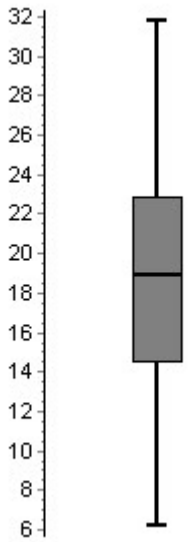
d) (0,5) Obtenha e **interprete** (contextualizando) os seguintes **quantis**: $x_{.25}$, $x_{.75}$ e $x_{.95}$.

$x_{.25} = 14{,}5 \rightarrow$ em pelo menos 0,25 dos dias a quantidade de óxido de enxofre emitida foi menor ou igual a 14,5 t
$x_{.75} = 22{,}9 \rightarrow$ em pelo menos 0,75 dos dias a quantidade de óxido de enxofre emitida foi menor ou igual a 22,9 t.
$x_{.95} = 27{,}5 \rightarrow$ em pelo menos 0,95 dos dias a quantidade de óxido de enxofre emitida foi menor ou igual a 27,5 t.

e) (0,3) Verifique se existem valores discrepantes no conjunto de dados.

$a_q = 8{,}4$ CI = 1,9 t CS = 35,5 t
Não há valores discrepantes.

f) (0,6) Construa o gráfico de caixa (box plot) para representar a distribuição.



g) (0,2) Observando o gráfico é razoável supor que essa amostra provém de uma população com distribuição normal? Por quê?

Sim é razoável, pois a distribuição é simétrica e também se verifica que $a_q < DI \cong DS$.

h) (1,0) Complete na tabela abaixo as duas primeiras linhas da distribuição de frequências desses dados, utilizando **k = 9**.

i = 2,85

j	Classes	F_j	F'_j	f_j	f'_j
1	6,2 |– 9,05	4	4	0,05714	0,05714
2	9,05 |– 11,9	7	11	0,1	0,1571

i) (0,4) Interprete (contextualizando) F_1 e f'_2.

$F_1 = 4 \rightarrow$ número de dias em que foram emitidos de 6,2 a 9,05 t (exclusive) de óxido de enxofre.
$f'_2 = 0,1571 \rightarrow$ proporção de dias em que foram emitidos de 6,2 a 11,9 t (exclusive) de óxido de enxofre.

Questão 3 (0,8). Os dados apresentados no diagrama de ramo e folhas se referem à temperatura corporal de homens (em ºF).

```
 96 | 3
 96 | 7 9
 97 | 0 1 1 1 2 3 4 4 4 4
 97 | 5 5 6 6 6 7 8 8 8 8 9 9
 98 | 0 0 0 0 0 0 1 1 2 2 2 3 3 4 4 4 4
 98 | 5 5 6 6 6 6 7 7 8 8 8 9
 99 | 0 0 0 1 2 3 4 5
 99 | 6
100 | 0                          Chave: 96 | 3 = 96,3
```

a) (0,3) Classifique a variável em estudo e identifique a escala de medida e a unidade de observação.

Variável numérica contínua; escala intervalar; unidade de observação: homem

b) (0,1) Quantas unidades de observação participaram desta pesquisa?

n=64

c) (0,1) Houve perda de informação neste tipo de representação gráfica? Por quê?

Não, pois todas as observações são apresentadas no diagrama, de modo que os dados originais podem ser recuperados.

d) (0,2) Obtenha o primeiro quartil e interprete-o.

Q_1= 97,6 $\rightarrow$ Pelo menos 0,25 (ou 25%) dos homens teve temperatura menor ou igual a 97,6.

e) (0,1) A partir do diagrama, classifique a distribuição quanto à assimetria.

Distribuição aproximadamente simétrica.

Questão 4 (1,5). Complete as afirmações com V (verdadeiro) ou F (falso) e **corrija** as falsas.

a) (F) O coeficiente de variação só deve ser utilizado para variáveis mensuradas em escala ~~intervalar~~ de razão e com média diferente de zero.

b) (V) Numa distribuição simétrica, unimodal e mesocúrtica podemos esperar aproximadamente 68% dos valores no intervalo entre a média mais e menos um desvio padrão.

c(F) As escalas ~~ordinal~~ intervalar e de razão geram variáveis numéricas, enquanto as escalas ~~intervalar~~ ordinal e nominal geram variáveis categóricas.

d) (F) Os bigodes de um gráfico de caixa ~~sempre~~ vão até os valores extremos, quando não há valores discrepantes.

e) (F) Numa distribuição assimétrica positiva e unimodal a média é ~~menor~~ maior que a mediana.

f) (F) Numa distribuição assimétrica positiva os valores acima da mediana são ~~mais~~ menos homogêneos que os valores abaixo dela.

g) (F) Medidas ~~de localização~~ separatrizes delimitam proporções de observações em um conjunto ordenado.

h) (V) No cálculo da variância o número de graus de liberdade (n-1) é utilizado para torná-la uma medida com melhores propriedades estatísticas.

i) (F) Quando multiplicamos todos os valores de uma variável por uma constante, a média ~~e a variância~~ desta variável fica~~m~~ multiplicada~~s~~ pela constante e a variância pelo quadrado da constante.

j) (F) A simetria da distribuição normal tem a seguinte caraterística: a_q ~~>~~ < DS = DI.

l) (V) É possível obter medidas descritivas a partir de uma tabela de distribuição de frequências, mas essas medidas serão exatas somente se a variável for discreta, se a variável for contínua as medidas serão aproximadas.

m) (F) São exemplos de medidas resistentes: mediana e ~~desvio padrão~~ amplitude interquartílica.

n) (F) Um valor é discrepante quando ~~é igual ou~~ ultrapassa as cercas.

o) (V) Pelo menos 30% dos valores de um conjunto de dados são menores ou iguais ao quantil $x_{.3}$.

p (F) O Q_2 e a mediana sempre são iguais ~~somente quando a distribuição é simétrica~~.