**Conceitos Estatísticos para IA – Turma 24IA**

Trabalho 1

**Habilidades desenvolvidas: Analisar tabelas de frequências e gráficos. Análise exploratória dos dados. Detecção de outliers.**

1. Classifique o tipo de variável

Utilize a base de dados do cadastro para classificar o tipo de variável.

|  |  |
| --- | --- |
| **Variável** | **Tipo da variável (natureza)** |
| NUM\_CPF | ID |
| CHAVE\_CONTRATO | ID |
| DATA\_NASCIMENTO |  |
| RENDA\_PRESUMIDA |  |
| CEP |  |
| LOGRADOURO |  |
| NUMERO |  |
| COMPLEMENTO |  |
| CEP\_A |  |
| BAIRRO |  |
| CIDADE |  |
| UF |  |
| LATITUDE |  |
| LONGITUDE |  |
| DDD\_CELULAR |  |
| CELULAR |  |
| DDD\_CELULAR\_2 |  |
| CELULAR\_2 |  |
| COD\_BANCO |  |
| NUM\_AGENCIA |  |
| NUM\_CONTA |  |
| MARCA\_VEICULO |  |
| MODELO\_VEICULO |  |
| PLACA\_VEICULO |  |
| CNPJ\_CREDOR |  |
| VALOR\_DIVIDA |  |
| STATUS\_CONSENTIMENTO |  |
| DATA\_INCLUSAO1 |  |
| IDADE |  |

Trecho do arquivo Cadastro\_PF.csv



1. Relacione a Estatística com a Definição do conceito estatístico de cada medida resumo:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estatísticas |  |  | Definição |
| (a) | Média |  | ( ) | Medida de dispersão que caracteriza o pico ou "achatamento" da curva da função de distribuição de probabilidade |
| (b) | Erro padrão |  | ( ) | É o maior valor de um conjunto de dados. |
| (c) | Mediana |  | ( ) | É o menor valor de um conjunto de dados. |
| (d) | Moda |  | ( ) | Valor numérico que separa a metade superior de uma amostra de dados, uma população ou uma distribuição de probabilidade, a partir da metade inferior. |
| (e) | Desvio padrão |  | ( ) | Valor que detém o maior número de observações, ou seja, o valor ou valores mais frequentes, ou ainda "o valor que ocorre com maior frequência num conjunto de dados, isto é, o valor mais comum |
| (f) | Variância da amostra |  | ( ) | Medida de variabilidade da média amostral. |
| (g) | Curtose |  | ( ) | Medida de dispersão. É a diferença entre o máximo e o mínimo. |
| (h) | Assimetria |  | ( ) | Valor que aponta para onde mais se concentram os dados de uma distribuição. Pode ser considerada o ponto de equilíbrio das frequências, num histograma. |
| (i) | Amplitude |  | ( ) | Medida que se obtém somando os quadrados dos desvios dos dados relativamente à média, e dividindo pelo número de dados menos um. Representa-se por s2. |
| (j) | Mínimo |  | ( ) | É a medida mais comum da dispersão estatística. Ele mostra a quanto de variação ou dispersão em relação à média. |
| (k) | Máximo |  | ( ) | Conjunto constituído pela reunião de diversos subconjuntos; total, conjunto, somatório. |
| (l) | Soma |  | ( ) | Grau de desvio de afastamento da simetria de uma distribuição, pode ser positiva para distribuições a direita e negativa para a esquerda. Para distribuições simetricas seu valor é zero. |
| (m) | Contagem |  | ( ) | Quantidade de registros. |

1. Critique a tabela abaixo:

Salários pagos na empresa XYZ (2o. semestre de 2009)

|  |  |  |
| --- | --- | --- |
| Meses/ano | No. de Empregados | Salários Pagos |
|  |  |  |
| Jul/09 | 57 | 30.127 |
| Ago/09 | 61 | 32.893 |
| Set/09 | 58 | 31.041 |
| Out/09 | 55 | 29.412 |
| Nov/09 | 66 | 35.523 |
| Dez/09 | 75 | 40.080 |
| Total | 372 | 199.076 |

1. UM GRÁFICO PODE LEVAR A CONCLUSÕES ERRADAS? EXPLIQUE POR QUE O GRÁFICO ABAIXO LEVA A CONCLUSÕES ERRADAS.



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

­­­­ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. SEM CALCULAR, OBSERVE OS TRÊS GRÁFICOS E RESPONDA?



1. QUAL É O CONJUNTO DE DADOS COM MAIOR DESVIO PADRÃO? QUAL TEM O MENOR DESVIO PADRÃO? EXPLIQUE SEU RACIOCÍNIO.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

­­­­ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. NO QUE OS CONJUNTOS DE DADOS SÃO IGUAIS? NO QUE ELES DIFEREM?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

­­­­ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. EM QUAL AGÊNCIA VOCÊ IRIA? JUSTIFIQUE?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

­­­­ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## BOX PLOT

A variável RENDA\_PRESUMIDA tem outlier ou ponto extremo?

Use o Box Plot para detectar presença de outlier e ponto extremo. Apresente os cálculos para identificar os limites que definem os outliers e pontos extremos.

|  |
| --- |
| > summary(RENDA\_PRESUMIDA)  Min. 1st Qu. Median Mean 3rd Qu. Max.  0 2068 2432 10388 2924 1890000  Mínimo =  Máximo =  Quartil 1 (Q1) =  Quartil 2 (Q2) = mediana =  Quartil 3 (Q3) =  Interquartil (IQ) = Q3 – Q1 =  Limite 1 = Q3 + 3\*IQ =  Limite 2 = Q3 + 1;5\*IQ =  Limite 3 = Q1 – 1.5\*IQ =  Limite 4 = Q1 – 3\*IQ = |

Entrega do exercício no formato word.

Data de entrega: 13/05/2022

Regina Bernal

03/05/2022