**Conceitos Estatísticos –** Turma 1IAR

Trabalho 01

**Habilidades desenvolvidas: Análise descritiva dos dados. Testes de Hipótese Paramétrica e Não Paramétrica. Análise de Correlação. Regressão Linear. Análise de Resíduos.**

1. Escolha a melhor resposta a questão.
2. Selecione a opção que lista as variáveis apenas quantitativas contínuas.

( ) a temperatura do corpo, o número de crianças, o sexo, a área de um campo de futebol

(X) idade, a temperatura corporal, o consumo de combustível, a renda

( ) número de filhos, sexo, consumo de combustível, a renda

( ) gênero, consumo de combustível, área de um campo de futebol, renda

1. Separe as variáveis qualitativas das quantitativas do item a):

|  |  |
| --- | --- |
| Variáveis qualitativas | Variáveis quantitativas |
| Sexo | Idade |
| Gênero | Temperatura do corpo |
|  | Número de crianças |
|  | Temperatura corporal |
|  | Renda |
|  | Consumo de combustível |
|  | Número de filhos |
|  | Área de um campo de futebol |

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. 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.

**Resposta:** Y é o maior desvio padrão. Já o menor é o K, nesta agência está mais concentrada. Na agência Z até encontra-se simétrico, mas está mais achatada e na agência Y é onde encontra-se mais distante da média.

­­­­

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

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

1. EM QUAL AGÊNCIA VOCÊ IRIA? JUSTIFIQUE?

**Resposta:** Eu iria na agência Z porque existe uma probabilidade maior de ser atendimento isso porque nesta agência a probabilidade de ser atendimento de 4 a 5 é de 10%. Já na Y a chance de ser atendido aparenta ser de 40% aproximadamente. E já na K a chance de ser atendido aparenta ser de 30%, e sendo assim a que tem a maior probabilidade de atendimento acaba sendo a agência Z.

­­­­

1. Escolha a melhor resposta a questão.

a) Para uma distribuição assimétrica, qual das seguintes estatísticas é uma boa medida para resumir a tendência central dos dados?

( ) média

(X) mediana

( ) ou média ou mediana

b) Classifique cada estatística a seguir em uma das opções: (a) tendência central; (b) dispersão; (c) forma;

( ) média

( ) curtose

( ) mediana

( ) desvio padrão

( ) variância

( ) moda

( ) assimetria

( ) intervalo interquartil

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 |

Precisamos saber ler a tabela, tem que ser transparente não pode ter dúvidas. Será que todas as linhas fazem sentidos? A forma que está apresentando empregados e salários está claro? Você faria diferente? Poderia estar de outro jeito? Ou seja, ao ler podemos estar entendendo errado?

1. Uma loja de utensílios vende em média 320 peças por dia, com desvio de 40. Após exaustiva campanha publicitária a administração pede para calcular a “nova média” de vendas para avaliar se ocorreu alguma melhora.

* Sabendo-se que o resultado foi de 370 peças vendidas na campanha, você pode dizer que a campanha foi um sucesso com α=5%? Quais são as hipóteses nula e alternativa deste teste e sua conclusão. E sua conclusão sobre o sucesso ou não da campanha publicitária.
* Ainda usando α=5%. Qual deve ser a nova média para que Ho seja rejeitado.

Para resolver isso usar o exemplo da faculdade que

1. Uma amostra de 1250 graus universitários conferidos em uma certa universidade, acusou a seguinte tabela.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sexo \ Grau** | **Bacharel** | **Mestre** | **Doutor** |
| **Masculino** | 505 | 160 | 25 |
| **Feminino** | 405 | 145 | 10 |

 Teste a hipótese de que as variáveis Grau Universitário e Sexo são independentes com α=5%.

**Python teste de hipótese, esse exemplo é do sal – Este é o teste de hipótese verificar se a hipótese é nula**

1. A partir dos dados da empresa XYZ (**Arquivo\_Salarios\_Colaboradores\_2021.csv)** avalie a distribuição da variável Salário dos seus colaboradores. Faça um algoritmo que estime a variável “Salário’ em função demais características do funcionário;
2. Selecione incialmente as variáveis que são candidatas a serem usadas no modelo. Qual a variável dependente e quais as variáveis independentes.
3. Na análise exploratória veja que os dados precisam ser limpos e tratados.
4. Exclusão de outliers, caso necessário (sempre explicando a opção)
5. Matriz de Correlação e gráficos
6. Para cada etapa anterior colocar comentários sobre a técnica utilizada e análise sobre as variáveis utilizadas e seus respectivos “achados”

* Técnicas: Regressão Linear, Árvore de Regressão

1. Com a métrica correta faça a comparação entre os diferentes algoritmos e qual foi o melhor entre eles.

Data de entrega: 18/09/2022

Adelaide Alves

17/08/2022