

UNIVERSIDADE REGIONAL DE BLUMENAU
Pós-Graduação em Data Science
Disciplina Inferência Estatística e Modelos de Previsão
Prof. Moacir Manoel Rodrigues Junior

AVALIAÇÃO FINAL

Considere o cenário descrito e responda o que for pedido. As respostas devem ser formuladas em um documento de texto e carregadas no AVA, na atividade específica da disciplina até a data limite. São critérios de avaliação:

- Exatidão dos resultados.
- Organização do Documento.
- Apresentar o resultado de todos os cálculos.
- Utilizar alguma linguagem para os cálculos que não o Excel, para as respostas colar o código utilizado para o cálculo e o resultado apresentado.
- Profundidade da Discussão Formulada.

Cenário

A planilha “Dados.xlsx” (ou “Dados.csv”) possui dados de um levantamento feito com 1289 pessoas dos Estados Unidos da América. Esta planilha está organizada com as seguintes variáveis:

Rótulo	O que mede
obs	Rótulo das Observações.
salario	Salário dos indivíduos entrevistados – em milhares de reais por ano.
sexo	Se for do sexo feminino recebe valor 1, se for do sexo masculino recebe valor 0.
cor	Recebe 1 se a pessoa for não-branca, 0 caso contrário.
est_civil	Recebe 1 caso a pessoa for casada, 0 caso contrário.
instrucao	Anos de educação formal que a pessoa recebeu.
experiencia	Anos de experiência que a possui na área em que trabalha.
idade	Anos de vida que a pessoa entrevistada possui.

Considerando este contexto, faça o que se pede em cada uma das questões descritas a seguir.

Questão 1 (3,00) – Considerando os dados apresentados acima, construa os testes de hipótese dadas as questões apresentadas. Elaborar: (i) A hipótese nula e a alternativa a ser testada; (ii) Destaque o teste que será realizado; (iii) Descreva os pressupostos necessários para cada teste; (iv) Efetue o teste de hipótese pretendido, e (v) Conclua sobre quais os indícios trazidos pelos dados.

- a) O Salário médio dos homens se difere do salário médio das mulheres?
- b) O Salário médio das pessoas não brancas se difere das pessoas brancas?
- c) O Salário médio das pessoas casadas se difere das pessoas solteiras?
- d) Considerando o sexo das pessoas e seu estado civil conjuntamente, é possível afirmar que algumas dos grupos formados possui média salarial diferente dos demais?

- e) O tempo de experiência médio é diferente para homens e mulheres?
- f) A idade média dos casados é diferente da idade média dos solteiros?

Questão 2 (3,00 pontos) – Analisando o banco de dados apresentado, é possível afirmar que haveria alguma das relações destacadas a seguir? Destaque o grau de associação para cada uma das relações apresentadas e verifique se seria significativo.

- g) Há relação entre Salário e Experiência?
- h) Há relação entre Salário e tempo de Instrução?
- i) Há relação entre Salário e Idade dos indivíduos investigados?
- j) Há relação entre a Experiência e a Idade dos Indivíduos?

Questão 3 (4,00 pontos) – O banco de dados descrito acima, foi utilizado para investigar o poder preditivo dos fatores destacados na base sobre o salário. Desta forma apresente o modelo de regressão que ajude a entender a forma como cada uma destas variáveis explicativas explica o salário das pessoas investigadas. Considere o seguinte modelo e responda as questões que seguem:

$$\text{salario}_i = \beta_0 + \beta_1 \text{sexo}_i + \beta_2 \text{cor}_i + \beta_3 \text{est}_{\text{civil}_i} + \beta_4 \text{instrucao}_i + \beta_5 \text{experiencia}_i + \beta_6 \text{idade}_i + \varepsilon_i$$

- k) Sabendo que as variáveis β_1 , β_2 e β_3 são todas variáveis dummies (dicotômicas), qual deveria ser a interpretação dada destes coeficientes sobre o salário?
- l) Qual é o poder explicativo do modelo (percentual de variância explicada)?
- m) De modelo geral, o modelo foi significativo?
- n) É possível afirmar que o grau de instrução do indivíduo afeta significativamente no seu salário? A que nível de significância?
- o) É possível afirmar que o estado civil do indivíduo afeta significativamente no seu salário? A que nível de significância?
- p) É possível afirmar que a experiência do indivíduo afeta significativamente no seu salário? A que nível de significância?
- q) É possível afirmar que o sexo do indivíduo afeta significativamente no seu salário? A que nível de significância?
- r) Considere o modelo de modo geral, principalmente os sinais de cada um dos coeficientes do modelo de regressão, os resultados fazem sentido conceitualmente para você? Justifique.