

05

Apresentação do projeto

Transcrição

[0:00] Pessoal, como eu prometi, um notebook de exercícios está aí para você fazer o download. Eu deixei ele com o nome de Teste_de_hipótese, então você faz o download.

[0:09] Vem aqui em File, upload Notebook e abre o seu notebook, escolher arquivo, aí vai na pasta, e abre ele.

[0:16] É esse cara aqui, é um trabalho de teste de hipóteses. A gente vai usar informações do nosso Dataset, que a gente vem trazendo ele desde lá do começo. É só para ter um arquivo para brincar.

[0:27] E você não fique preso a esses exercícios, você pode pensar na sua cabeça alguns testes e testar as hipóteses aqui em cima desses dados. Tem muita coisa aqui que a gente pode brincar.

[0:39] Eu estou propondo um, o nosso Dataset já está aí para você fazer o download, você já deve ter ele. Aquela mesma coisa, as informações básicas do Dataset.

[0:49] Aqui, esse notebook está careca porque ele está só com algumas informações e você vai preenchendo conforme a gente vem fazendo nos outros cursos.

[0:56] Então, aqui as bibliotecas, importar o Dataset, visualizar o Dataset, aí você fica à vontade para fazer do jeito que você quiser. Se precisar de mais células, vem aqui.

[1:08] Ou aqui, Code, ele vai abrir uma célula nova. Você não quer mais uma célula, Ctrl + M D e ele vai apagar a célula.

[1:16] Deixa eu ver, Ctrl + M D. Aqui ele não está funcionando, mas em editar, Delete, Ctrl + M D. Ele não está funcionando aqui no meu teclado.

[1:25] Mas enfim, aqui o probleminha que eu deixei, que é também um teste de hipótese de comparação de duas médias, porque a gente está usando o nosso arquivo, vai ficar uma coisa mais legal de fazer.

[1:38] Então vamos ler rapidamente e depois eu deixo para você resolver aí.

[1:42] "Você é um pesquisador que estuda o mercado de trabalho brasileiro e resolve estudar as diferenças salariais dos trabalhadores dos estados do Rio de Janeiro e São Paulo".

[1:52] Então a gente vai trabalhar com uma variável diferente no nosso Dataset, que é a UF, que é a unidade da federação.

[1:59] "Durante a sua pesquisa você verifica que, aparentemente, os rendimentos dos trabalhadores no estado do Rio de Janeiro são mais baixos que os rendimentos dos trabalhadores no estado de São Paulo."

[2:11] "Para confirmar esta conclusão realize um teste de hipóteses de comparação de médias", como eu tinha dito, "em cima de duas amostras de trabalhadores dos dois estados".

[2:20] Então é você que vai selecionar as amostras, está bem? "Siga as seguintes etapas".

[2:26] Primeiro, justamente a seleção da amostra. "Selecione duas amostras de 500 trabalhadores para cada um dos estados."

[2:33] 500 para o Rio de Janeiro e 500 para São Paulo, é isso que você tem que selecionar para realizar o nosso teste.

[2:39] "Utilize o parâmetro random_state 101", que é o que a gente vem utilizando. Não sei porque eu estou usando o 101, mas eu achei esse número interessante.

[2:48] Utilize isso para a gente poder depois, quando eu for fazer a correção no próximo vídeo, poder comprar os resultados e a gente ter resultados idênticos. Só por isso.

[2:56] Depois, você pode cortar isso e fazer do seu jeito. Variável UF, vamos voltar aqui em cima.

[3:03] UF está aqui. Para o Rio de Janeiro a UF vai ser igual a 33 e para São Paulo a UF é igual a 35.

[3:09] Você já tem um exemplo de como fazer isso para o sexo masculino e feminino, então é só ir lá, se você tiver alguma dúvida, e ver como vai fazer isso mudando a variável de sexo para unidade da federação, UF.

[3:21] "Considere o nível de significância de 5%". "Teste a hipótese de que a renda média dos trabalhadores do Rio de Janeiro é menor que a renda média dos trabalhadores de São Paulo."

[3:33] Lembra que eu falei que geralmente o que a gente está testando fica na hipótese alternativa? Aqui a gente repara que a igualdade, então, não está aqui, está se uma é menor que a outra.

[3:42] Ou seja, já deixei uma dica aí no meio do caminho. Então corre atrás, faz o exercício. É legal você tentar fazer antes de ver a correção porque você fixa melhor o conteúdo.

[3:54] Aqui, espaço para selecionar as amostras, para você colocar os dados do problema. Não esqueça de obter média e desvio padrão para as duas amostras porque a gente vai estar usando isso nas estatísticas de teste.

[4:04] Deixei de presente aqui aquela tabelinha para você consultar de forma mais fácil. Como eu já disse, eu já falei que teste que eu quero, então a tabela é essa aqui. Teste paramétrico.

[4:19] Aqui, formulação das hipóteses. Fica livre para desenhar do jeito que você quiser. Faz no papel, digita aqui, pega os exemplos que eu fiz com o Latex, corre atrás aí que você vai conseguir.

[4:31] Aqui, responder as informações para que tipo de amostra a gente tem, que tipo de distribuição amostral a gente vai utilizar, as fixações.

[4:40] Depois eu vou colocar as figuras que eu venho colocando durante o nosso curso, aquela da área de rejeição e aceitação, onde fica a estatística.

[4:51] Mas eu quis deixar você sozinho aqui um pouco para você treinar, depois as figuras vão aparecer aí no próximo vídeo.

[4:59] Critério de aceitação do valor crítico, o P valor, isso tudo está aqui na nossa tabela, já está pronto aqui para você.

[5:05] E aqui eu deixei para você fazer de forma mais simples, utilizando o Python, usando o DescrStatsW, você já conhece, e o CompareMeans também.

[5:13] Espaço para digitar a conclusão, só clicar aqui. Apaga esse cara aqui, digita a sua conclusão.

[5:19] Próximo vídeo a gente faz a correção, mas faz aí antes de ver o vídeo. Até lá.