**05 - Apresentação do projeto**

0:00] Como eu prometi, temos um notebook de exercícios, está disponível [aqui](https://caelum-online-public.s3.amazonaws.com/1229-estatistica-parte3/stages/aula-7-inicio.zip) para você fazer o download. Eu deixei com o nome de **Teste\_de\_hipótese**.

[0:09] Basta fazer o download, acessar "File > upload Notebook" e, na próxima tela, apertar "Escolher arquivo", selecionar a pasta dele e abrir. É um trabalho de teste de hipóteses. Vamos usar informações do nosso Dataset, que a gente vem trazendo desde o começo.

[0:25] É só para ter um arquivo para brincar e para que você não fique preso a esses exercícios, você pode pensar em alguns testes e testar as hipóteses em cima desses dados.

[0:39] O nosso Dataset já está disponível para você fazer o download, você já deve ter ele. Aquela mesma coisa, as informações básicas do Dataset.

[0:49] Esse notebook está só com algumas informações e você vai preenchendo, assim como fizemos nos outros cursos, as bibliotecas, o Dataset, visualizar o Dataset, aí você fica à vontade para fazer do jeito que você quiser. Se precisar de mais células, basta selecionar "Code" e ele vai abrir uma célula nova. Se você não quer mais uma célula, basta apertar "Ctrl + M D" e ele vai apagar a célula.

[1:16] Aqui não está funcionando, mas em editar, "Edit > Delete selected cells" podemos verificar o comando, que, de fato é "Ctrl + M D". Ele não está funcionando no meu teclado.

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

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

[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 de São Paulo".

[1:52] Então a gente vai trabalhar com uma variável diferente no nosso Dataset, que é a UF, 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] São 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 que, depois, quando eu for fazer a correção no próximo vídeo, você possa comparar os resultados e termos 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.

| **Código** | **Descrição** |
| --- | --- |
| 11 | Rondônia |
| 12 | Acre |
| 13 | Amazonas |
| 14 | Roraima |
| 15 | Pará |
| 16 | Amapá |
| 17 | Tocantins |
| 21 | Maranhão |
| 22 | Piauí |
| 23 | Ceará |
| 24 | Rio Grande do Norte |
| 25 | Paraíba |
| 26 | Pernambuco |
| 27 | Alagoas |
| 28 | Sergipe |
| 29 | Bahia |
| 31 | Minas Gerais |
| 32 | Espírito Santo |
| 33 | Rio de Janeiro |
| 35 | São Paulo |
| 41 | Paraná |
| 42 | Santa Catarina |
| 43 | Rio Grande do Sul |
| 50 | Mato Grosso do Sul |
| 51 | Mato Grosso |
| 52 | Goiás |
| 53 | Distrito Federal |

[3:09] Você já tem um exemplo de como fazer isso para o sexo masculino e feminino. Se você tiver alguma dúvida, basta fazer o mesmo processo 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 estamos testando fica na hipótese alternativa? Aqui, a gente repara que a igualdade não está aqui, está se uma é menor que a outra.

[3:42] Ou seja, já deixei uma dica. Então, corra atrás, faça o exercício. É legal você tentar fazer antes de ver a correção, porque você fixa melhor o conteúdo.

[3:54] Temos 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 usaremos isso nas estatísticas de teste.

[4:04] Deixei de presente no notebook aquela tabelinha para você consultar de forma mais fácil. Eu já falei qual teste que eu quero, então a tabela é a de teste paramétrico.

[4:19] Temos também a formulação das hipóteses. Fique livre para desenhar do jeito que você quiser. Faz no papel, digita, corre atrás que você vai conseguir.

[4:31] Seguindo, é necessário responder as informações para qual tipo de amostra temos, que tipo de distribuição amostral vamos 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 treinar, depois as figuras vão aparecer, no próximo vídeo.

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

[5:05] E 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 apertar em "Conclusão: (Digite sua conclusão aqui)", apagar o que está escrito no quadro e digitar a sua conclusão.

[5:19] No próximo vídeo, faremos a correção, mas tente fazer antes de passar para ele. Até lá.