

3.1 - diag

17 de julho de 2024



historico_estratificado.csv
Planilha

Analise os dados e limpe as colunas da seguinte forma:
período_ingresso: de 2006.1 até 2019.2
tipo_matricula: remover "Dispensa"
situacao: remover "Em Curso"
o novo dataframe após a limpeza deve se chamar df_filtrado



Vamos começar lendo o arquivo CSV que você enviou para entender a estrutura dos dados. Em seguida, farei as limpezas solicitadas nas colunas ``período_ingresso``, ``tipo_matricula``, e ``situacao``.

Agora que temos uma visão dos dados, vamos aplicar os filtros que você especificou:

1. **período_ingresso**: Mantenha registros de 2006.1 até 2019.2.
2. **tipo_matricula**: Remova registros onde o valor seja "Dispensa".
3. **situacao**: Remova registros onde o valor seja "Em Curso".

Vamos proceder com essas limpezas e verificar o resultado. [\[↗\]](#)

A limpeza foi concluída com sucesso. O dataframe ``df_filtrado`` agora contém 13.831 registros e manteve todas as 34 colunas originais, mas os registros foram filtrados conforme suas especificações:

- **Período de ingresso** entre 2006.1 e 2019.2.
- **Tipo de matrícula** excluindo registros com "Dispensa".
- **Situação** excluindo registros com "Em Curso".

Se precisar de mais ajustes ou análises, estou à disposição para ajudar! [\[↗\]](#)

O tipo de ingresso influencia de forma estatisticamente significativa na quantidade de períodos até a graduação?



Para responder à sua pergunta sobre a influência do tipo de ingresso na quantidade de períodos até a graduação, podemos usar uma análise estatística. Uma forma comum é realizar um teste de hipótese para comparar as médias dos grupos. Vamos usar o Teste ANOVA, que é adequado para comparar as médias de três ou mais grupos independentes, para ver se existem diferenças estatisticamente significativas entre os diferentes tipos de ingresso em relação à quantidade de períodos até a graduação.

Primeiro, vamos verificar a distribuição e os valores únicos da coluna ``forma_ingresso`` e garantir que temos dados suficientes para a coluna que representa a quantidade de períodos até a graduação. Assumindo que esta coluna seja representada por ``período_relativo``, vamos também verificar sua disponibilidade.

A coluna ``forma_ingresso`` tem várias categorias, com as mais frequentes sendo SISU, VESTIBULAR / ENEM, e VESTIBULAR. Existem 16 valores ausentes na coluna ``período_relativo``, que vamos considerar antes de prosseguir com a análise estatística.

Vamos primeiro tratar esses valores ausentes e então aplicar um teste ANOVA para verificar se as diferenças nas médias de ``período_relativo`` entre os diferentes tipos de ingresso são estatisticamente significativas. Podemos optar por remover essas linhas com valores ausentes, dado que representam uma pequena fração do total. [\[↗\]](#)

O resultado do teste ANOVA indica um valor de estatística F de aproximadamente 42.45 e um valor-p extremamente baixo (próximo de 1.10×10^{-67}), o que é significativamente menor que o nível convencional de significância ($\alpha = 0.05$).

Portanto, podemos concluir que existem diferenças estatisticamente significativas entre os grupos de

`forma_ingresso` em relação à quantidade de períodos até a graduação. Isso sugere que o tipo de ingresso tem um impacto significativo no tempo que leva para um estudante se graduar. Se precisar de mais detalhes ou análises, posso ajudar! [\[>-\]](#)

Comece com Data Analyst

[Denunciar conteúdo](#) | [Termos de uso](#) | [Política de privacidade](#) |

?