

[Entrar](#)[Cadastrar](#)

Esta é uma cópia de uma conversa entre ChatGPT e Anônimo.  
[Denunciar conteúdo](#)

📎 Arquivo carregado

Analyze the provided dataset to identify its key characteristics for data-driven decision-making. Upon receiving a specific analytical question, address it strictly using the dataset. Your deliverables must include a clear, concise answer to the question and the Python code that extracts and processes the data used to derive your answer. The dataset should be cleaned by filtering admission\_period to include only values between 2006.1 and 2019.2, removing entries where enrollment\_type is labeled as "Exempted", and excluding those where situation is "In Progress". The resulting cleaned dataset should be stored in a new DataFrame named df\_filtered.

I'll first inspect the dataset to understand its structure and then apply the necessary cleaning steps. Let's load and examine the data.

The dataset contains 30,130 entries and 34 columns. Key columns for filtering include:

- `admission_period` (float64): We will keep only values between 2006.1 and 2019.2.
- `enrollment_type` (object): We will exclude entries labeled as "Exempted."
- `situation` (object): We will exclude entries labeled as "In Progress."

Now, I'll apply these filters and create the `df_filtered` DataFrame. [↗]

↓  
The dataset has been successfully filtered down to 16,500 entries. It now includes only:

- Admission periods between 2006.1 and 2019.2.
- Excludes students with "Exempted" enrollment type.
- Excludes entries where the situation is "In Progress".

Let me know your specific analytical question, and I'll analyze the dataset accordingly. [↗]

Using appropriate statistical tests, determine whether the mode of admission significantly impacts graduation and dropout rates.

To determine whether the **mode of admission** (`admission_form`) significantly impacts **graduation and dropout rates**, we'll follow these steps:

#### 1. Define Key Variables

- **Independent variable:** `admission_form` (categorical)
- **Dependent variable:** `exit_form`, categorized into:
  - "Graduate" (successful completion)
  - "Dropout" (includes all non-graduation exit reasons)

#### 2. Prepare the Data

- Create a summary table of graduation and dropout counts for each `admission_form`.
- Convert the data into proportions to analyze trends.

#### 3. Conduct a Statistical Test

- Since we have categorical data, we'll use a **Chi-Square Test of Independence** to check if `admission_form` significantly affects `exit_form`.
- The null hypothesis ( $H_0$ ) assumes **no association** between `admission_form` and `exit_form`.
- A significant result (p-value < 0.05) would indicate that the mode of admission **does influence** graduation/dropout rates.

Now, let's perform the analysis.

#### Results:

##### Graduation & Dropout Counts by Admission Form:

Admission Form	Graduated	Dropped Out
Entrance Exam	2,657	1,491
Entrance Exam / ENEM	2,800	2,524
Unified Selection System (SISU)	1,348	5,029
Transfer	127	171
Reoption	64	172
Reentry	17	0
Judicial Decision	66	77
PEC-G Agreement	0	10
Graduate (Special Case)	0	10

##### Statistical Test Results:

- **Chi-Square Test p-value = 0.0** (extremely small, effectively zero)

##### Interpretation:

Since the p-value < 0.05, we **reject the null hypothesis**, meaning there is a **significant association** between the mode of admission and the likelihood of graduating or dropping out. In other words, how a student is admitted **does impact** their chances of completing their program.

Would you like a deeper breakdown of specific admission modes or trends? [↗]

Envie uma mensagem para o ChatGPT

📎 Anexar

🗣️ Voz

