



I have a dataset <https://ximarketing.github.io/data/bankchoice.csv>. The first row is variable names, and each column is a variable. I want to run a multinomial logit model. The dependent variable is Choice, independent variables are Age, Income, Education, Job. Choice and Job are categorical variables.

Xi Li

I have a dataset <https://ximarketing.github.io/data/bankchoice.csv>. The first row is variable names, and each column is a variable. I want to run a multinomial logit model. The dependent variable is Choice, independent variables are Age, Income, Education, Job. Choice and Job are categorical variables.


27/03/2025, 17:53:30

GPT-4o (Tokens used 716)

To run a multinomial logit model using the dataset you've provided, you can follow these steps in Python using the `statsmodels` library. Here's a step-by-step guide to help you:


1. **Install Required Libraries:** Ensure you have the necessary libraries installed. You can install them using pip if you haven't already.

KeyError: 'Choice\_1'

Xi Li

KeyError: 'Choice\_1'

27/03/2025, 17:54:37


GPT-4o (Tokens used 1116)

It looks like there's an issue with how the dependent variable 'Choice' is being referenced in the script. The dataset most likely has the 'Choice' column as a categorical variable, and it needs to be converted correctly.

Here's an improved version of the script that should handle the conversion of the 'Choice' variable correctly:


```
import pandas as pd
import statsmodels.api as sm
from statsmodels.tools import add_constant
```

ValueError: Pandas data cast to numpy dtype of object. Check input data with `np.asarray(data)`.

Xi Li

ValueError: Pandas data cast to numpy dtype of object. Check input data with `np.asarray(data)`.

27/03/2025, 17:55:15


GPT-4o (Tokens used 1625)

The error you're encountering is often due to the presence of non-numeric data in the independent variables. Let's ensure that all the columns used in the model are numeric, as required by statsmodels. We'll also ensure that the dependent variable is correctly processed.

Here's an updated version of the script to handle this:

```
import pandas as pd
import statsmodels.api as sm
```


It still does not work. Can you write a completely different code?



**Xi Li**

It still does not work. Can you write a completely different code?

27/03/2025, 17:57:30



**GPT-4o** (Tokens used 2157)

Sure, let's approach it differently. We can use the **patsy** library to construct the model formula, which can simplify the process of handling categorical variables. We'll also ensure that all necessary conversions are handled correctly.

Here is an alternative approach using **patsy** and **statsmodels** :

```
import pandas as pd
import statsmodels.api as sm
import patsy
```

It works this time!