CISC 3225 Extra Credit
Due May 22, 2023 at 11:59 PM
5 points

Statistical Tests

Use the cameras dataset:
https://raw.githubusercontent.com/CUNY-CISC-3225/datasets/main/cameras.csv

Correlations (1 points)

Create a correlation matrix of the dataset. Find 2 interesting strong correlations (>0.5, +/-) and 2 interesting weak correlations (<0.5, +/-). Write a brief discussion about what you see. As part of your discussion, indicate whether each correlation is significant.

Trends (1 points)

Are cameras getting cheaper or more expensive every year? By how much? Is the effect significant?

How much does max resolution increase each year? By how much? Is the effect significant?

Differences (1 points)

Create a new boolean column in the DataFrame which is True if the camera was released on or after 2004, and False if it was released prior to 2004.

On average, is the mean weight of cameras different across these two time periods? Is the difference significant?

On average, is the mean price different across these two time periods? Is the difference significant?

Brands (2 points)

Using the before/after 2004 column you created in the last question, determine how many cameras were released before 2004 and how many were released on or after 2004. Which era saw more cameras released?

Do all camera brands follow this general trend, or do some differ? In other words, does the brand of camera have an effect on the number of cameras released pre- or post-2004? If there are differences among brands, which brands were closest to the general trends in pre-2004/post-2004 camera releases, and which brands differed the most?

Submission

Download any notebooks you created from Colab as .ipynb files and submit them in Blackboard. You may answer written questions in the notebook or in a separate text file (PDF, Word, or plain text).