What Gets Counted Counts Dataset Analysis

Github: WillDBM/DATA102---Student-Stress-Group-Project
Dataset Source: Kaggle (Student Stress Monitoring Datasets)
William Armentrout, David Rayle, Katelyn Ledbetter, Ethan Meadows

Introduction

The dataset provides a look into the causes of stress among college students aged between 18 and 21, totaling 843 respondants. The data was derived from a nationwide survey and contains hundreds of anonymous responses. Responses were gathered with a survey using a five point Likert scale ranging from “Not at all” to “Extremely”.

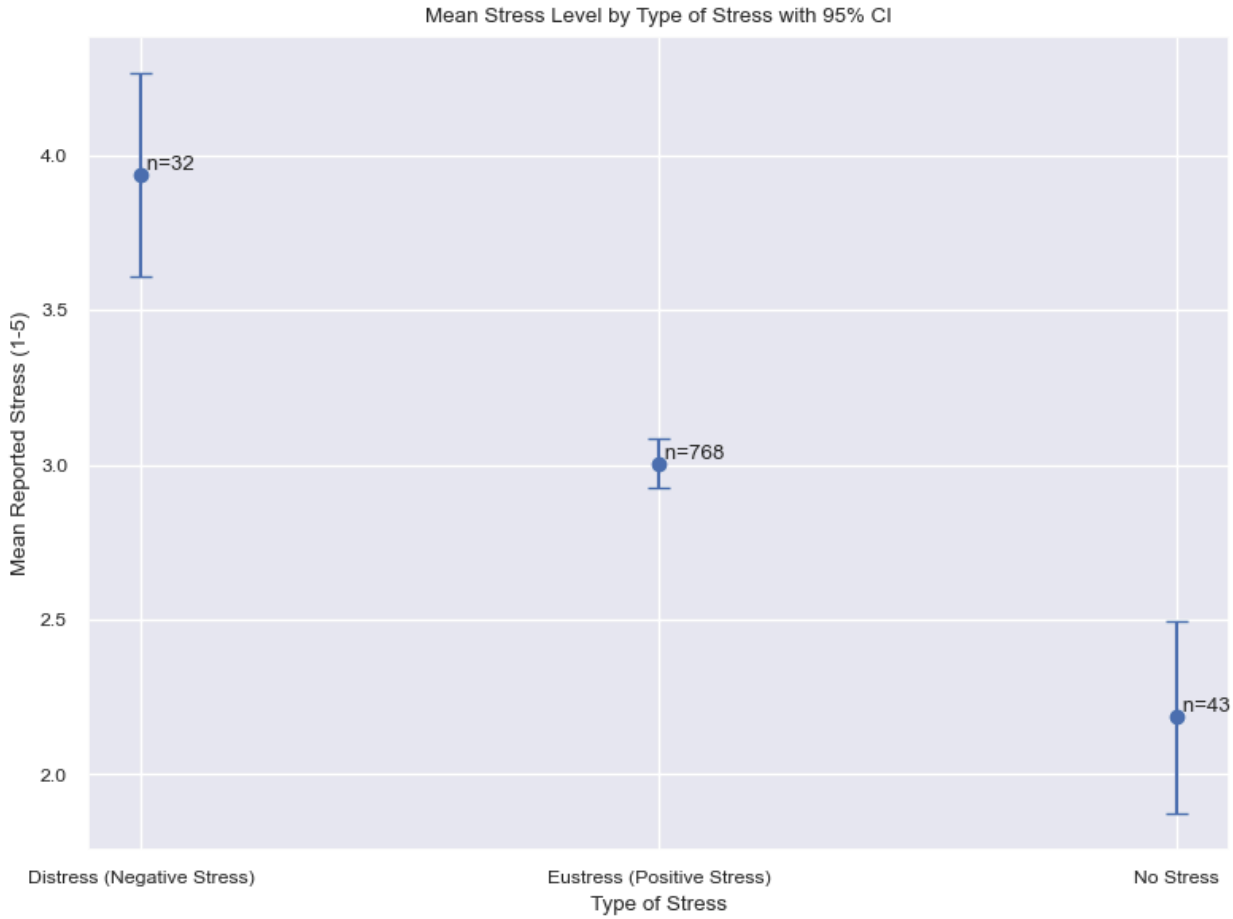
The dataset aims to identify the sources and types of stress affecting college students and to explore correlations between emotional states, physical health, academic pressures, and social factors. The dataset also includes a target variable on what type of stress each respondent primarily experiences as eustress, distress, or no stress, which makes the dataset suitable for both descriptive and predictive analyses.

Data Analysis / Visualizations

Analysis of the *Student Stress Dataset* revealed distinct differences in stress levels depending on the type of stress experienced.

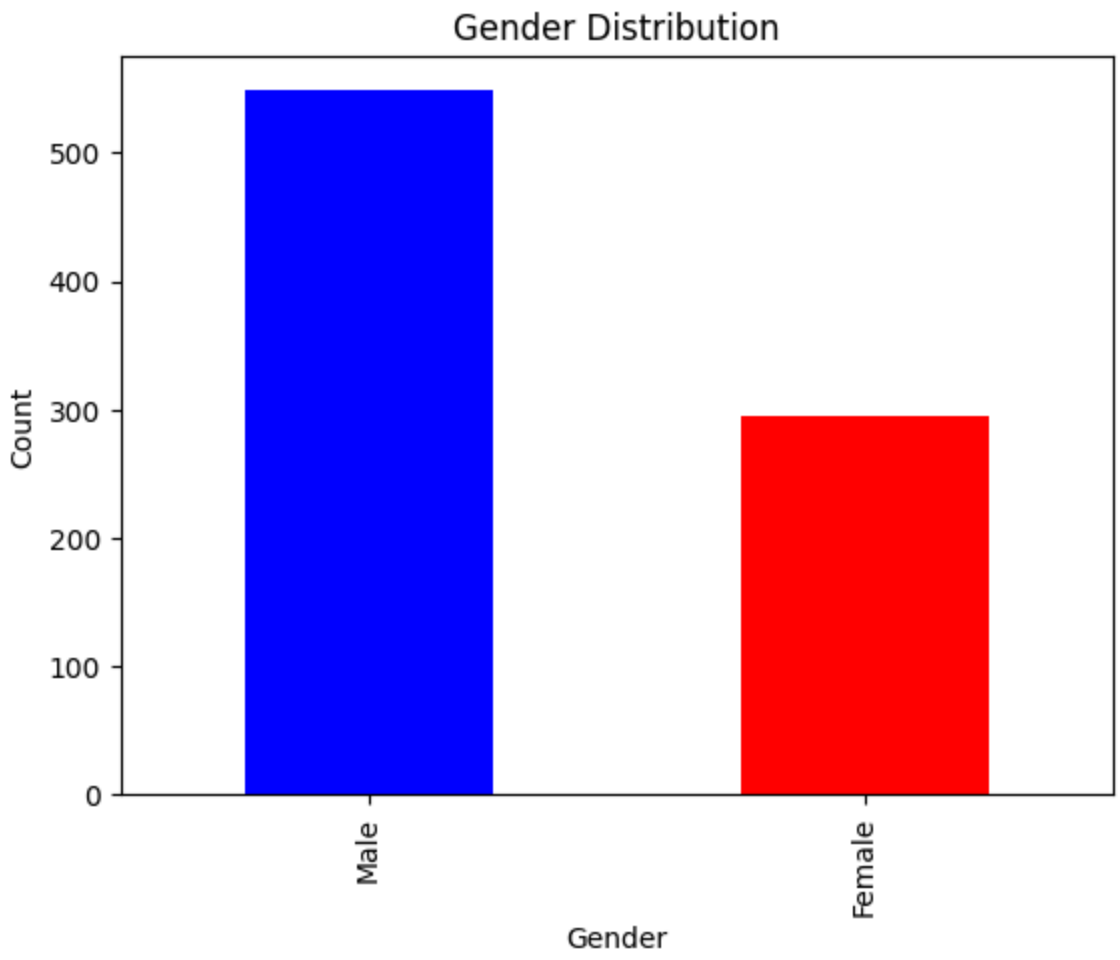
The ANOVA test (supported by visual means and 95% CIs) showed statistically significant differences in mean reported stress by participants and the type of stress they experienced. As shown in the following Mean Stress plot, those who experience higher levels of stress tend to experience Negative Stress, and those who experience moderate stress experience

Positive Stress.



We noticed that the responses were highly uneven, with the vast majority of respondents reporting Positive Stress, as shown by the n=768 in the above figure, which makes sense for a self-reported survey such as this one.

Another uneven thing in the dataset was the difference in the number from each gender, with there being 548 Males and 295 Females. The difference in the Gender Distribution can be best shown in the bar chart below.



A Correlation matrix using Pearson's r reveals a moderate positive correlation between items tied to physical/psychological strain, for example: “feeling sadness,” “difficulty concentrating,” “fatigue,” and “anxiety.” Negative correlations appear between the “type of stress” variable and wellness items (e.g., confidence, relaxation ability), suggesting that greater distress coincides with more negative well-being indicators. (This matrix, 26 by 26, is better shown on the GitHub linked at the top of the report.)

Interpretation and Analysis

These findings indicate that stress type strongly affects perceived intensity and emotional impact.

- Distress correlates with higher self-reported stress and aligns with negative mental/physical indicators in the correlation matrix.

- Eustress, though technically "stress," aligns with healthier outcomes and may reflect motivational pressure rather than strain.
- The statistical significance of group differences (via ANOVA and CI separation) validates that the variance in mean stress level is not random.
- Gender does not substantially alter overall trends, though male students may experience slightly more stress. The fact that there were 548 male respondents and 295 female respondents may also indicate a certain bias.

Conclusion

Overall, the data support the interpretation that the type of stress, not just its presence, matters most for understanding student well-being. Encouraging adaptive forms of eustress while addressing sources of distress could be an effective focus for student support initiatives.

Citation / References

Sultanulislamovi, M. D. (n.d.). *Student stress monitoring datasets* [Data set]. Kaggle. https://www.kaggle.com/datasets/mdsultanulislamovi/student-stress-monitoring-datasets/data

WillDBM. (n.d.). *DATA102 — Student Stress Group Project* [Code repository]. GitHub. https://github.com/WillDBM/DATA102---Student-Stress-Group-Project