Student Mental Health

1. Data

The dataset consists of demographics, self-reported health status, test results from psychological assessments, and survey findings on the empathy, mental health, and burnout of Swiss medical students. It contains 886 observations and 20 variables. The variables selected for correlation analysis are the CES-D total score (cesd), which assesses depression on a scale from 0 to 60, the STAI score (stai_t), which measures anxiety on a scale from 0 to 63, and the MBI emotional exhaustion score, which measures burnout. Given that all these variables are related to an individual's mental health and are likely to be correlated, they were chosen to enhance the meaningfulness of this analysis (appendix-3). Also, as they are continuous, they are well suited for correlation analysis. (For cleaning and effect of transformations Q-Q plots, tests etc. refer to appendix-1,2).

Column	Definition	Data Type	Explanation	Mean	Min	Max
cesd	CES-D Score	Numeric	Center for Epidemiologic Studies Depression scale of the participant	0.00	18.05	56.00
stai_t	STAI Score	Numeric	State-Trait Anxiety Inventory scale of the participant	5.00	16.88	30.00
mbi_cx	MBI Exhaustion	Numeric	Maslach Burnout Inventory-Exhaustion scale of the participant	20.00	42.90	77.00

Table 1: Overview of the dataset (relevant features)

2. Planning

i. We examine the correlation between CES-D score (depression) and STAI score (anxiety) both of which are interval data.

ii. Our objective is to understand the relationship between the CES-D score (target/outcome) and STAI score (predictor) after controlling the effect of MBI Emotional Exhaustion which is also interval data.

iii. According to the Central Limit Theorem, the mean of a sample with a size greater than 30 is normally distributed, hence all three variables may be considered to be normally distributed.

iv. We use a parametric test: Pearson's product moment correlation (most widely used) which requires interval data and assumes normal distribution for testing significance & obtaining CI.

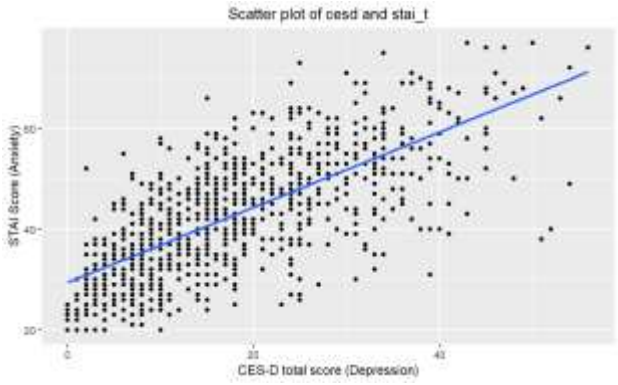

Figure 1: Scatter plot STAI vs CES-D score

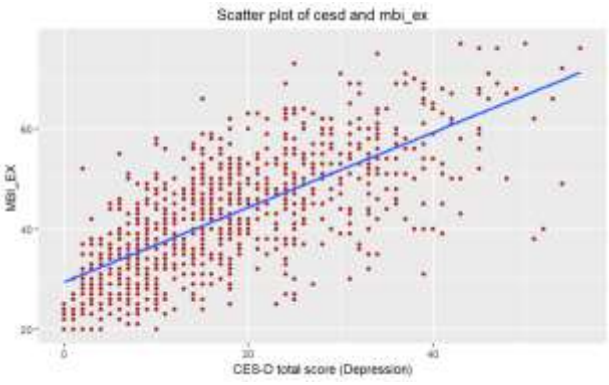

Figure 2: Scatter plot MBI_EX vs CES-D score

3. Analysis

i. Correlations

The table shows pairwise correlation between cesd, mbi_ex and stai_t. All the correlations were found to be statistically significant at 0.05 or 0.01 level of significance (as p-values < 2.2e-16).

Variables	Correlation coefficient r	t-value	p-value	95% CI
cesd and mbi_ex	0.61	22.628	< 2.2e-16	(0.562, 0.645)
stai_t and mbi_ex	0.53	18.606	< 2.2e-16	(0.481, 0.576)
cesd and stai_t	0.72	30.471	< 2.2e-16	(0.682, 0.746)

Table 2: Pairwise Correlations and results from significance tests of Pearson's Correlation

ii. Variability

$R^2 * 100$	cesd	mbi_ex	stai_t
cesd	100.00	36.71	51.39
mbi_ex	36.71	100.00	28.12
stai_t	51.39	28.12	100.00

Table 3: Proportions of Variability in Percentage

51.38% of the variability in CES-D score (depression) can be explained by STAI score (measuring anxiety), while MBI Emotional Exhaustion score (measuring burnout) accounts for 36.7% of the variability in CES-D score.

iii. Partial Correlation

The partial correlation between CES-D score and STAI score when controlling MBI Emotional Exhaustion is 0.585 and is statistically significant at 0.05 or 0.01 level of significance. After controlling the effect of MBI Emotional Exhaustion, the STAI score accounts for 34.22% of variability in CES-D scores.

Partial Correlation	t-value	Degrees of Freedom	p-value	$R^2 * 100$
0.585	21.421	883	< 0.001	34.22

The proportion of variability in CES-D scores explained by STAI scores decreases from 51.38% to 34.22% after controlling the effect of MBI Emotional Exhaustion. This suggests that emotional exhaustion has a substantial impact on the relationship between depression and anxiety. However, it can be established that the level of emotional exhaustion is not the only factor that is responsible for positive correlation that is still statistically significant (as p-value <0.001) between depression and anxiety even after controlling for exhaustion.

4. Conclusion

From the analysis, the r after controlling emotional exhaustion (0.585) was lower than the bivariate correlation (0.72) but still significant, as seen in the plot. The correlation between MBI Emotional Exhaustion and STAI score (r=0.53) suggests some shared variance between the variables, which can make it hard to distinguish their distinctive effects on the relationship between CES-D and STAI. The causality of CES-D score (depression) is therefore uncertain and may depend on other factors as well to explain variability in CES-D beyond STAI and MBI-EX.

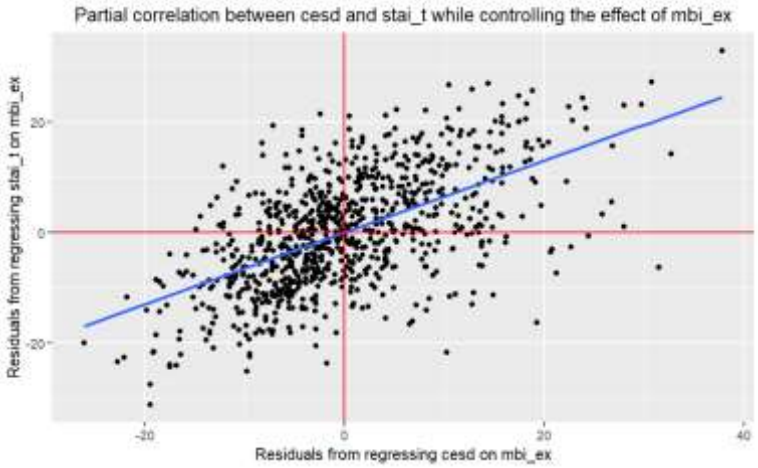

Figure 3: Partial Correlation while controlling mbi_ex