

# Student Depression: A Statistical Analysis

Gavin B. Jack B. Nathan C.

# Background - The Ideas Behind the Project

- We chose to use a data set on student depression:  
<https://www.kaggle.com/datasets/adilshamim8/student-depression-dataset>
- This data set has 18 variables all about students with the main category being a binary answer on whether or not a student has depression
- We wanted to do something original so rather than looking at gpa we looked at work hours, gender, and financial situations

# Background - The Data Set

We shortened the dataset to fit our needs down to just 4 variables:

- **Quantitative:**
  - **Work Study Hours** (Continuous variable representing the number of hours worked/studied, containing numbers 0-12)
  - **Financial Stress** (Continuous variable representing the financial stress level, containing integers 1-5)
- **Qualitative:**
  - **Gender** (Categorical variable: Male/Female)
  - **Depression** (Binary categorical variable: 1 = Depressed, 0 = Not Depressed)

In total there are about 27900 total observations

# Header of the data set

Gender	Work Study Hours	Financial Stress	Depression
Male	3	1	1
Female	3	2	0
Male	9	1	0
Female	4	5	1
Male	1	1	0
Female	4	1	0

# Goal of our Study

**Main Goal: To investigate how different factors - specifically financial stress, work study hours, and gender-relate to depression among students in a large academic dataset.**

Specific Objectives:

- Understand the relationship between financial stress and the likelihood of experiencing depression
- Explore whether hours spent in work study programs are associated with depression levels
- Examine if gender plays a role in the prevalence of depression

Why this Matters:

- Student mental health is a growing concern in academic settings
- Identifying key risk factors can help guide better support systems policy decisions and mental health interventions on campuses.

# Hypothesis

## Hypothesis 1: Financial Stress

- **Null ( $H_0$ ):** There is no significant difference in financial stress between students who are depressed and those who are not.
- **Alternative ( $H_1$ ):** Students who are depressed report significantly higher financial stress than those who are not.

## Hypothesis 2: Work Study Hours

- **Null ( $H_0$ ):** There is no significant difference in work study hours between students who are depressed and those who are not.
- **Alternative ( $H_1$ ):** Students who are depressed work significantly more hours in work study programs than those who are not.

## Hypothesis 3: Gender

- **Null ( $H_0$ ):** There is no association between gender and depression.
- **Alternative ( $H_1$ ):** Depression rates differ significantly between male and female students.

# Model Assumptions

## T-tests Assumptions:

- Data is approximately normally distributed
- No significant outliers
- Independence of observations
- Homogeneity of variance (Welch's t-test used if violated)
- Variables are continuous or treated as continuous

## Logistic Regression Assumptions:

- Binary dependent variable
- Independence of observations
- No multicollinearity
- Linear relationship between continuous independent variables and the logit of the outcome
- Large sample size ensures model stability

## Chi-Square Test Assumptions

- **1. Categorical Variables:** Both variables (Gender and Depression) are categorical.
- **2. Independence:** Each observation is independent of others (no student is counted more than once).
- **3. Expected Frequencies:** Expected counts in each cell of the contingency table should generally be  $\geq 5$ .
- **4. Random Sampling:** The data should come from a random or representative sample.

# Quantitative Data - Descriptive statistics via R

```
>summary(data$Financial.Stress)
```

Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
1.00	2.00	3.00	3.14	4.00	5.00

```
>summary(data$Work.Study.Hours)
```

Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
0.00	4.00	8.00	7.16	10.00	12.00



# Qualitative Data - Descriptive Stats via R

Table of Gender

Female	Male
12354	15547

Table of Depression

0	1
11565	16336

Proportion Table of Gender

Female	Male
0.4427798	0.5572202

Proportion Table of Depression

0	1
0.4145013	0.5854987

# Descriptive Statistics Recap

## Financial Stress:

- Mean: 3.14
- Median: 3
- Range: 1 to 5

## Gender Distribution:

- Female: 44.3%
- Male: 55.7%

## Work Study Hours:

- Mean: 7.16
- Median: 8
- Range: 0 to 12

## Depression Prevalence:

- Depressed: 58.5%
- Not Depressed: 41.5%

# Regression Model - Hypo 1

- **Hypothesis:** Higher financial stress increases the likelihood of depression among students
- **Findings:**
  - Financial Stress Effect: Each unit increases in financial stress raises the odds of depression
    - ~73.9% (odds ratio 1.739,  $p < 0.001$ ), strongly supporting the hypothesis
  - Statistical Significance: financial stress is highly sufficient ( $p < 2e-16$ ) confirming a robust association with depression
  - Model Fit: Financial stress explains substantial variability in depression
    - Deviance reduced by 3,82, improving model fit (AIC = 34,038)
- **Implication:** Higher financial stress is a significant risk factor for depression, suggesting interventions like financial aid could help reduce depression among students

# Hypothesis 1 in R

## Welch Two Sample t-test

data: Financial Stress by Depression

t = -65.076    df = 24739    p-value < 2.2e-16

Alternative Hypothesis:

true difference in means between group 0 and  
group 1 is not equal to 0

95% Confidence Interval:

-1.092781    -1.028878

Sample Estimates:

Mean Group 0	Mean Group 1
2.518724	3.579553

Coefficients:

	Estimate	Standard Error	Z Value	Pr(> z )
(Intercept)	-1.344531	0.031176	-43.13	<2e-16
Financial Stress	0.553700	0.009516	58.19	<2e-16
Signif Codes	0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1			
Dispersion parameter for binomial family taken to be 1				

Null Deviance:

37855 on 27897 df

Residual Deviance:

34034 on 27896 df

(3 observations deleted due to missingness)

AIC: 34038

Number of Fisher Scoring Iterations: 4

# Hypothesis 1 - Financial Stress and Depression

## T-test Results:

- Mean Financial Stress (Depressed): 3.58
- Mean Financial Stress (Not Depressed): 2.52
- $t = -65.08$ ,  $p < 2.2e-16$

## Logistic Regression:

- Coefficient: 0.554
- Odds Ratio:  $\exp(0.554) \approx 1.74$
- Interpretation: Each unit increase in financial stress increases the odds of depression by approximately 74%.

## Interpretation:

- Significant difference in financial stress levels between depressed and non-depressed students.

# Regression Hypothesis 2

- **Hypothesis:** Higher work study hours increase the likelihood of depression among students
- **Findings:**
  - Effect: each additional hour of work-study increase the odds of depression by ~12.3 (odds ratio = 1.123,  $p < 0.001$ ) which supports the hypothesis
  - Statistical Significance: Work-study hours are highly significant ( $p < 2e-166$ ) confirming a strong association with depression
  - Model Fit: Work-study hours explain some variability in depression (deviance reduced by 1,218) improving model fit (AIC = 36,515)
- **Implication:** More work-study hours are a significant risk factor for depression, suggesting that reducing work study demands could help lower depression rates among students

# Hypothesis 2 in R

## Welch Two Sample t-test

data: Financial Stress by Depression

t = -34.994    df = 23107    p-value < 2.2e-16

Alternative Hypothesis:

true difference in means between group 0 and  
group 1 is not equal to 0

95% Confidence Interval:

-1.657688

-1.481599

Sample Estimates:

Mean Group 0

6.237959

Mean Group 1

7.807603

Coefficients:

	Estimate	Standard Error	Z Value	Pr(> z )
(Intercept)	-0.470281	0.026564	-17.70	<2e-16
Work Study Hours	0.115767	0.003373	34.32	<2e-16
Signif Codes	0 '***' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1			
Dispersion parameter for binomial family taken to be 1				

Null Deviance:

37859 on 27900 df

Residual Deviance:

36641 on 27899 df

AIC: 36645

Number of Fisher Scoring Iterations: 4

# Hypothesis 2 - Work Study Hours and Depression

## T-test Results:

- Mean Work Study Hours (Depressed): 7.81
- Mean Work Study Hours (Not Depressed): 6.24
- $t = -34.94$ ,  $p < 2.2e-16$

## Logistic Regression:

- Coefficient: 0.116
- Odds Ratio:  $\exp(0.116) \approx 1.12$
- Interpretation: Each additional work study hour increases the odds of depression by approximately 12%.

## Interpretation:

- Significant difference in work study hours between depressed and non-depressed students.



# Regression Hypo 3

- **Hypothesis:** Being male increases the likelihood of depression among students
- **Findings:**
  - Effect: Being male increases the odds of depression by ~0.7% (odds ratio = 1.007,  $p = 0.764$ ), but the effect is negligible and not significant
    - Failing to support the hypothesis
  - Statistical Significance: Gender is not significant ( $p = 0.764$ ) indicating no association with depression
  - Model Fit: Gender does not explain any variability in depression showing no improvement over a null model
- **Implication:** Gender is not a meaningful risk factor for depression among students
  - Suggests other factors should be prioritized

# Hypothesis 3 in R

	0	1
Female	5133	7221
Male	6432	9115

Coefficients:				
	Estimate	Standard Error	Z Value	Pr(> z )
(Intercept)	0.341303	0.018257	18.70	<2e-16
Gender Male	0.007333	0.024464	0.3	0.764
Signif Codes	0 '***' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1			
Dispersion parameter for binomial family taken to be 1				

Pearson's Chi-Square Test with Yates Continuity Correction

Data: table\_gender\_depression

$X^2 = 0.082658$ , df = 1, p-value = 0.7737

Null Deviance:

37859 on 27900 df

Residual Deviance:

32859 on 27899 df

AIC: 37863

Number of Fisher Scoring Iterations: 4

# Hypothesis 3 - Gender and Depression

## Chi-Square Test:

- Contingency Table:
  - Female: 5133 Not Depressed, 7221 Depressed
  - Male: 6432 Not Depressed, 9115 Depressed
- $\chi^2 = 0.082658$ ,  $df = 1$ ,  $p = 0.7737$

## Logistic Regression:

- Coefficient for Gender (Male): 0.00733
- $p = 0.764$
- Interpretation: Gender is not a significant predictor of depression.

## Interpretation:

- No significant association between gender and depression.

# Regression Model - All together

- **Hypothesis:** Being male, higher financial stress, and more work-study hours increase the likelihood of depression among students
- **Findings:**
  - **Gender/Male Effect:** Being male increases the odds of depression by ~0.6%
  - **Financial Stress Effect:** Each unit increase in financial stress raises the odds of depression by ~74.0%
  - **Work Study Hours Effect:** Each additional hour of work-study increases the odds of depression by ~12.3%
  - **Statistical Significance:** Financial stress and work-study hours are highly significant, while gender is not
  - **Model Fit:** The model explains substantial variability in depression with financial stress and work-study hours driving the improvement
- **Implication:** Higher financial stress and more work-study hours are significant risk factors for depression among students, warranting interventions like financial aid and reduced work demands

Coefficients:				
	Estimate	Standard Error	Z Value	Pr(> z )
(Intercept)	-2.165933	0.044514	-48.66	<2e-16
Gender Male	0.006276	0.026796	0.234	0.815
Financial Stress	0.553960	0.009716	57.016	<2e-16
Work Study Hours	0.115919	0.003622	32.008	<2e-16
Signif Codes	0 '***' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 '.' 1			
Dispersion parameter for binomial family taken to be 1				

Null Deviance:	37855 on 27897 df
Residual Deviance:	32973 on 27894 df
(3 observations deleted due to missingness)	
AIC: 32981	
Number of Fisher Scoring Iterations: 4	

# Model Diagnostics 1

**Accuracy:** The model correctly classified approximately **60.06%** of observations. This indicates moderate predictive ability, but not strong.

**Pseudo R<sup>2</sup> (McFadden's):** The model had a pseudo R<sup>2</sup> value of **0.129**, suggesting that about 12.9% of the variation in depression status is explained by Gender, Financial Stress, and Work Study Hours.

**Limitation:** The model's predictive power is relatively low and may benefit from additional predictors or better data cleaning.

```
> ll_model <- logLik(model)

> ll_null <- logLik(glm(Depression ~ 1, data = data, family = "binomial"))

> pseudo_r2 <- 1 - as.numeric(ll_model / ll_null)

> pseudo_r2

[1] 0.12905
```

# Model Diagnostics 2 - Confusion Table

Confusion Table	Actual: Not Depressed (0)	Actual: Depressed (1)
Predicted: Not Depressed (0)	6,375 (True Negative)	3,807 (False Negative)
Predicted: Depressed (1)	5,188 (False Positive)	12,528 (True Positive)

## Summary:

- Model performs reasonably well at identifying depression based on gender, financial stress, and work study hours.
- However, the **false positive rate** is high—potentially labeling healthy individuals as depressed.
- **Improvement Ideas:**
  - Adjust classification threshold (currently 0.5).
  - Add more predictors or use a different model.
  - Evaluate sensitivity and specificity to fine-tune accuracy.

# Limitations

**Sample Size & Scope:** The dataset was limited in size and only included a specific population, which may limit generalizability.

**Binary Outcome:** Depression was treated as a binary variable (1 = Depressed, 0 = Not Depressed), which may oversimplify the complexity of mental health.

**Unmeasured Confounders:** Other important factors (e.g., social support, academic stress, pre-existing mental health conditions) were not included in the model.

**Cross-Sectional Data:** The data was cross-sectional, preventing us from making causal inferences between financial stress and depression.

**Self-Reported Measures:** Financial stress and depression were based on self-report, which can introduce reporting bias.

# Conclusion

This study investigated how financial stress and work-study hours relate to depression among college students using logistic regression.

Results showed that financial stress was a significant predictor of depression, indicating that students with higher levels of financial stress were more likely to experience depressive symptoms.

In contrast, work-study hours were not significantly associated, suggesting that the number of hours worked alone may not be a direct driver of depression.

These findings reinforce the idea that subjective financial strain, rather than objective workload, has a stronger impact on student well-being.

The analysis contributes to growing evidence that financial stress is a key risk factor for mental health challenges in higher education settings.

Targeted interventions addressing financial anxiety and access to resources could be effective in reducing depression risk.



## Relevant Links

# kaggle

<https://github.com/gavin-barro/math268-final-proj>

<https://www.kaggle.com/datasets/adilshamim8/student-depression-dataset>



Questions?