

# WHO Health Data Analysis

## A Presentation by Marjella Ernst

Marjella Ernst

June 2025

Introduction  
○○

Hypotheses to test  
○○

Research Question  
○○

Data  
○○

Methodology  
○○

Challenge: What is Health?  
○○

Regression Results  
○○○○

Marjella Ernst

WHO Health Data Analysis

# Introduction

# Introduction

This presentation explores the relationship of various factors such as: - stress - physical exercise - sleep, and on health.

# Hypotheses to test

# Hypotheses to test

- 1 sleeping is more important to health than exercise
- 2 living a stress free lifestyle has a major impact on health

# Research Question

# Research Question

“Which are the main influencing factors on health and how can we use these insights to improve it?”



# Data

# Data

- The data used for this analysis originates from the **World Health Organization (WHO)**.
- It includes self-reported indicators on:
  - Physical activity
  - Sleep quality
  - Stress levels
  - Caloric intake
- The dataset builds the empirical basis for testing the relationship between lifestyle factors and health outcomes

Introduction  
○○

Hypotheses to test  
○○

Research Question  
○○

Data  
○○

**Methodology**  
●○

Challenge: What is Health?  
○○

Regression Results  
○○○○

L  
○

# Methodology

# Methodology

- A multiple linear regression model is used to estimate the effect of:

- **Exercise** (physical activity),
- **Sleep quality**, and
- **Stress level**

on the dependent variable **Health** (approximated by excess caloric intake).

- The model is specified as:

$$\text{Health}_i = \beta_0 + \beta_1 \cdot \text{Exercise}_i + \beta_2 \cdot \text{Sleep}_i + \beta_3 \cdot \text{Stress}_i + \varepsilon_i$$

- In R, the model is estimated as:

```
lm(Health ~ Exercise + Sleep + Stress, data = processed)
```

# Challenge: What is Health?

# Challenge: What is Health?

- For the purpose of this analysis, health is approximated by excess body weight
  - This serves as a proxy indicator due to data limitations
  - Other common indicators, such as BMI, are not available in the dataset

# Regression Results

# Regression Results

Table 1: Regression Results:  
Health  $\sim$  Exercise +  
Sleep + Stress

(1)	
(Intercept)	246.148** (87.845)
Exercise	285.774*** (21.394)
Sleep	30.220 (20.625)
Stress	-5.357 (8.851)
Num.Obs.	100
R <sup>2</sup>	0.652
R <sup>2</sup> Adj.	0.642
AIC	1370.7



# Interpretation of Results I

- **Exercise** has strong and highly significantly positive effect on health
  - Coefficient: 285.77 (\*\*\*),  $p < 0.001$
  - Interpretation: Higher physical activity is strongly associated with better health (less excess weight)
- **Sleep** shows small positive effect, but is not statistically significant
  - Coefficient: 30.22,  $p > 0.1$
  - No strong evidence that sleep quality alone explains variation in health

# Interpretation of Results II

- **Stress** has a small negative effect, but is also not statistically significant
  - Coefficient:  $-5.36$ ,  $p > 0.5$
  - Cannot conclude that stress significantly affects health in this model
- **Model fit:**
  - $R^2 = 0.652 \rightarrow$  model explains  $\sim 65\%$  of the variation in health
  - Adjusted  $R^2 = 0.642$  is still high after adjusting for model complexity
  - $AIC = 1370.7$  model complexity vs. fit trade-off

# Limitations

# Limitations

- Health is difficult to measure because it includes multiple dimensions:
  - Physical aspects (e.g., number of illnesses, fitness)
  - Mental aspects (e.g., stress resistance, absence of psychological conditions)
- Many influencing factors of health are themselves difficult to observe or quantify:
  - Genetics, sleep, nutrition
  - Environmental and external factors
  - Access to health care
- There is a significant risk of reverse causality:
  - Poor health can increase stress
  - Illness may limit exercise
  - Health problems can disrupt sleep

# Conclusion

# Conclusion

- Exercise is found to be clearly statistically significant for health
- Sleep and stress show expected effects but lack statistical significance
- The model explains a large share of variation in health (~65%)
- However, health is a multifaceted concept, more complex than this model can explain
- Further research should include:
  - Better proxies for health (e.g., BMI, blood pressure, clinical diagnoses)
  - Longitudinal data to address reverse causality
  - Interaction effects (like stress and sleep)