SLEEP HEALTH AND LIFESTYLE

GROUP 4
Elisa Bonomi
Emma Didelon
Flavia Leotta
Negin Kafee

Bioinformatics for Computational Genomics 2023/24
Statistics project

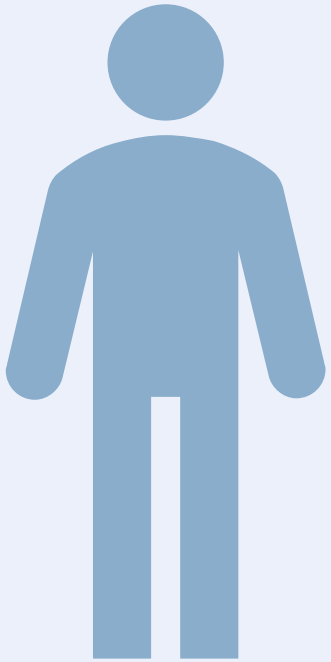
WHY FOCUSING ON SLEEP HEALTH?

- **Good sleep** is essential to control daily physiological pattern but also to maintain a baseline mental health.

- **Sleep disorders** may thus affect dramatically our mental and physical health and may be **linked** to many pathological conditions (diabetes, anxiety, depression).

GENERAL DATASET INFORMATIONS

- **Author**: Laksika Tharmalingam.
- **374 rows**: people participating in the study.
- **13 columns**: aside from PersonID, variables related to characteristics, quality and quantity of sleep and daily habits.
- Made for illustration purposes, not scientific ones.



Link : Sleep Health and Lifestyle Dataset (kaggle.com)

VARIABLES OF THE DATASET

DISCRETE / CATEGORICAL VARIABLES

GENDER	MALE	FEMALE	
OCCUPATION	11 UNIQUE JOBS		
QUALITY OF SLEEP	SCALE 1-10		
STRESS LEVEL	SCALE 1-10		
BMI CATEGORY	NORMAL	OVER-WEIGHT	OBESE
SLEEP DISORDERS	NONE	SLEEP APNEA	INSOMNIA

VARIABLES OF THE DATASET

DISCRETE / CATEGORICAL VARIABLES

GENDER	MALE, FEMALE
OCCUPATION	11 UNIQUE JOBS
QUALITY OF SLEEP	SCALE 1-10
STRESS LEVEL	SCALE 1-10
BMI CATEGORY	NORMAL, OVER-WEIGHT, OBESE
SLEEP DISORDERS	YES, NO

VARIABLES OF THE DATASET

CONTINUOUS VARIABLES

AGE	IN YEARS
SLEEP DURATION	IN HOURS
PHYSICAL ACTIVITY LEVEL	IN MINUTES
BLOOD PRESSURE (SYSTOLIC/DIASTOLIC)	mmHg
HEART RATE	IN BPM
DAILY STEPS	ADIMENSIONAL

DO PEOPLE WITH SLEEP DISORDERS HAVE A WORSE QUALITY OF SLEEP?

1st QUESTION

1.1 Are the two populations normally distributed?

Asymptotic one-sample Kolmogorow-Smirnov test

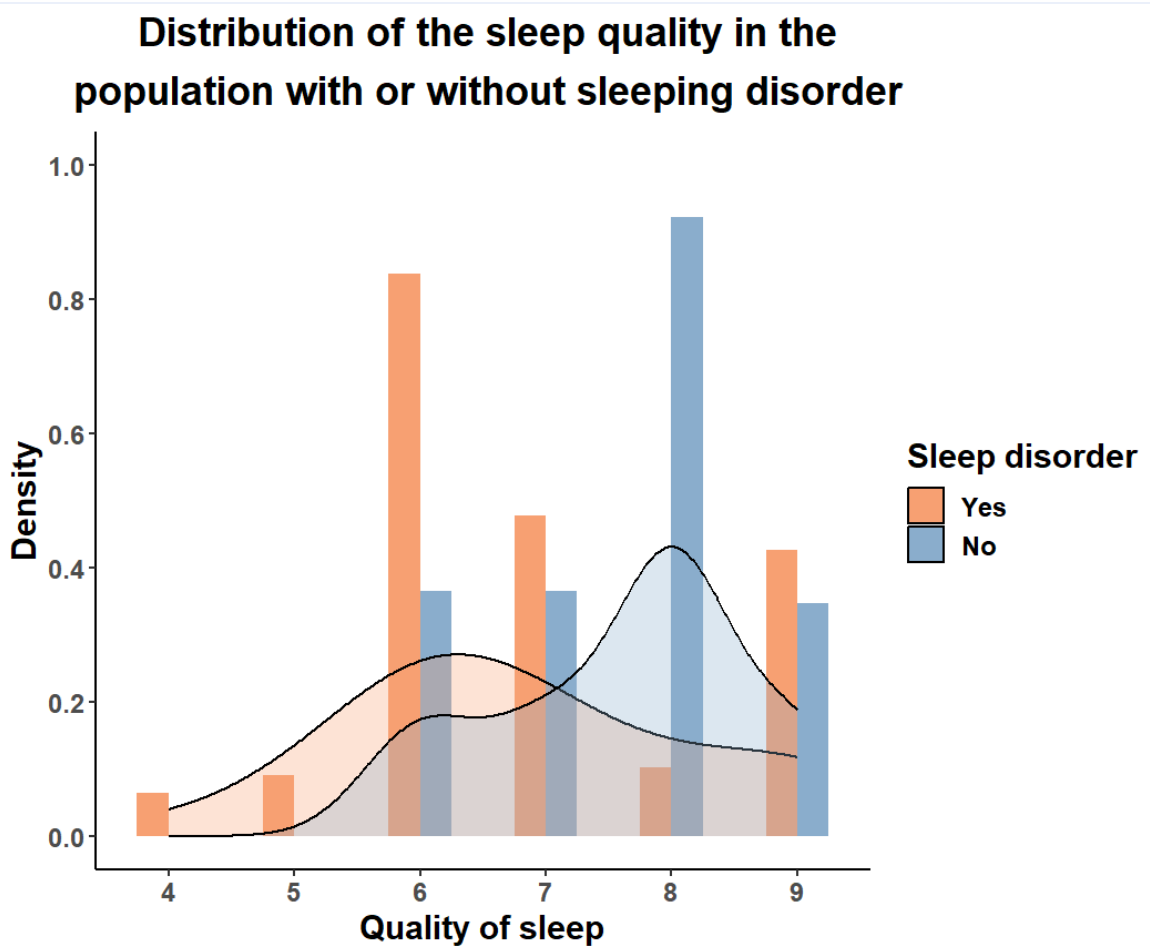
$$H_0 : F(x) \sim N(\mu, \sigma^2)$$
$$H_1 : F(x) \nsim N(\mu, \sigma^2)$$

Results:

Null hypothesis is rejected in both cases.

Conclusions:

Since both samples do not follow a normal distribution, it's not recommended to perform a t-test.



DO PEOPLE WITH SLEEP DISORDERS HAVE A WORSE QUALITY OF SLEEP?

1st QUESTION

1.2 Do they follow a similar non-parametric distribution?

Wilcoxon rank sum test with continuity correction.

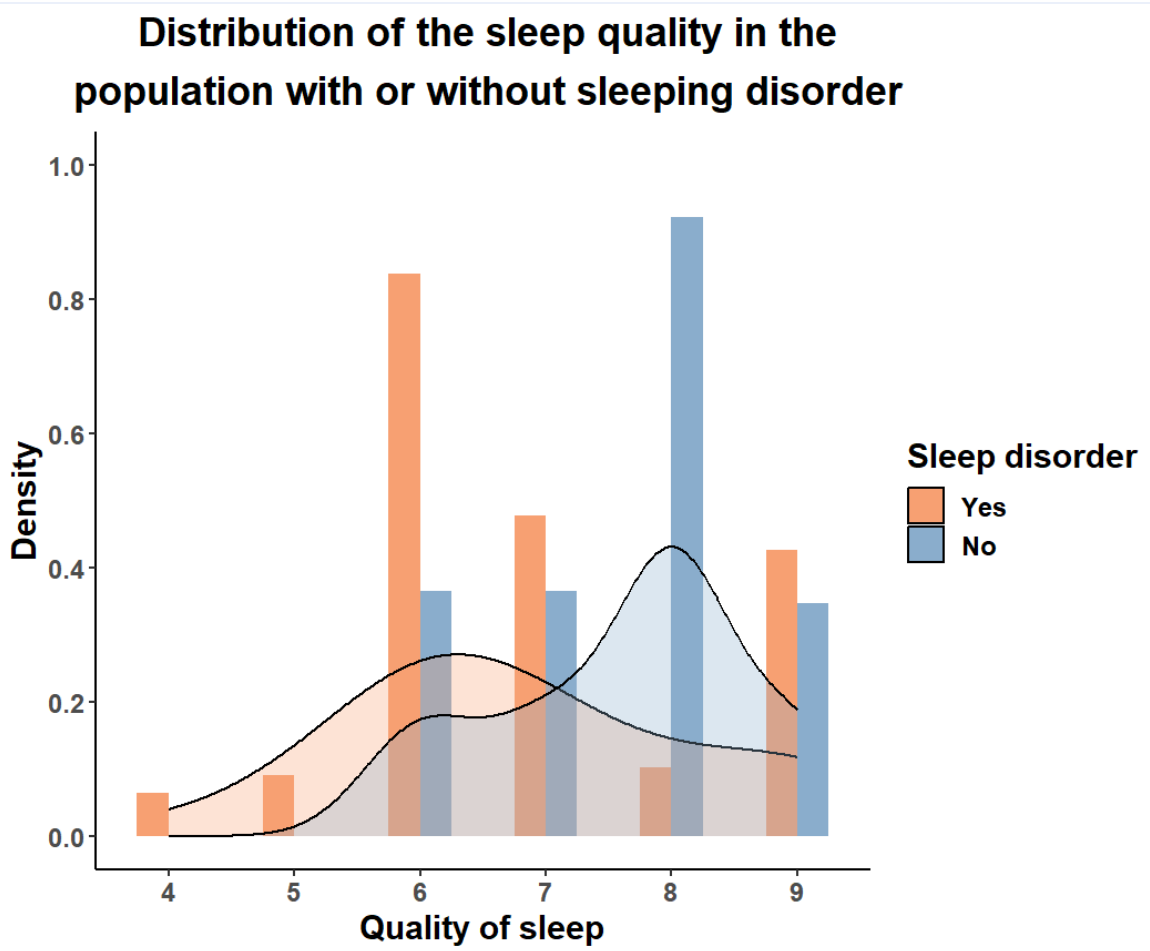
$$H_0 : The\ samples\ have\ similar\ distribution$$
$$H_1 : The\ samples\ do\ not\ have\ similar\ distribution$$

Results:

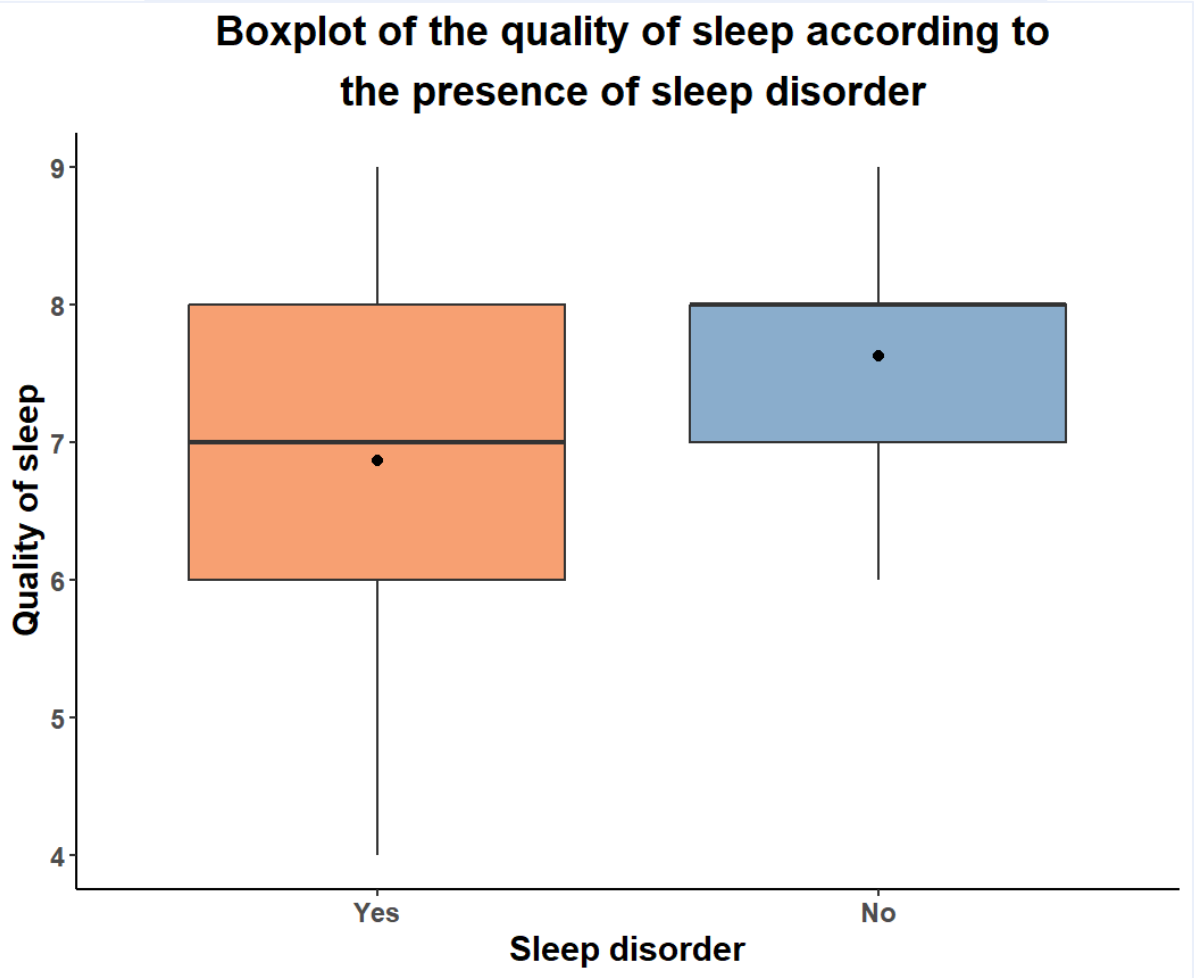
The null hypothesis is rejected.

Conclusions:

The two populations follow different distributions.



ONE-TAILED Z TEST



$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Population 1 : people without sleep disorders

Population 2: people with sleep disorders

Assumptions:

- **CLT**: n is big enough to use z-test
- True variance is the same

Results: p-value low, we reject H_0

Conclusion: People with sleep disorders have a worse quality of sleep.

WHICH LIFESTYLE FACTORS TO CHANGE?

2nd QUESTION

Which lifestyle factors are linked to the sleeping disorders and can be changed to achieve a better quality of sleep?

NOT TAKEN INTO CONSIDERATION

- GENDER
- AGE
- BLOOD PRESSURE
- DAILY STEPS

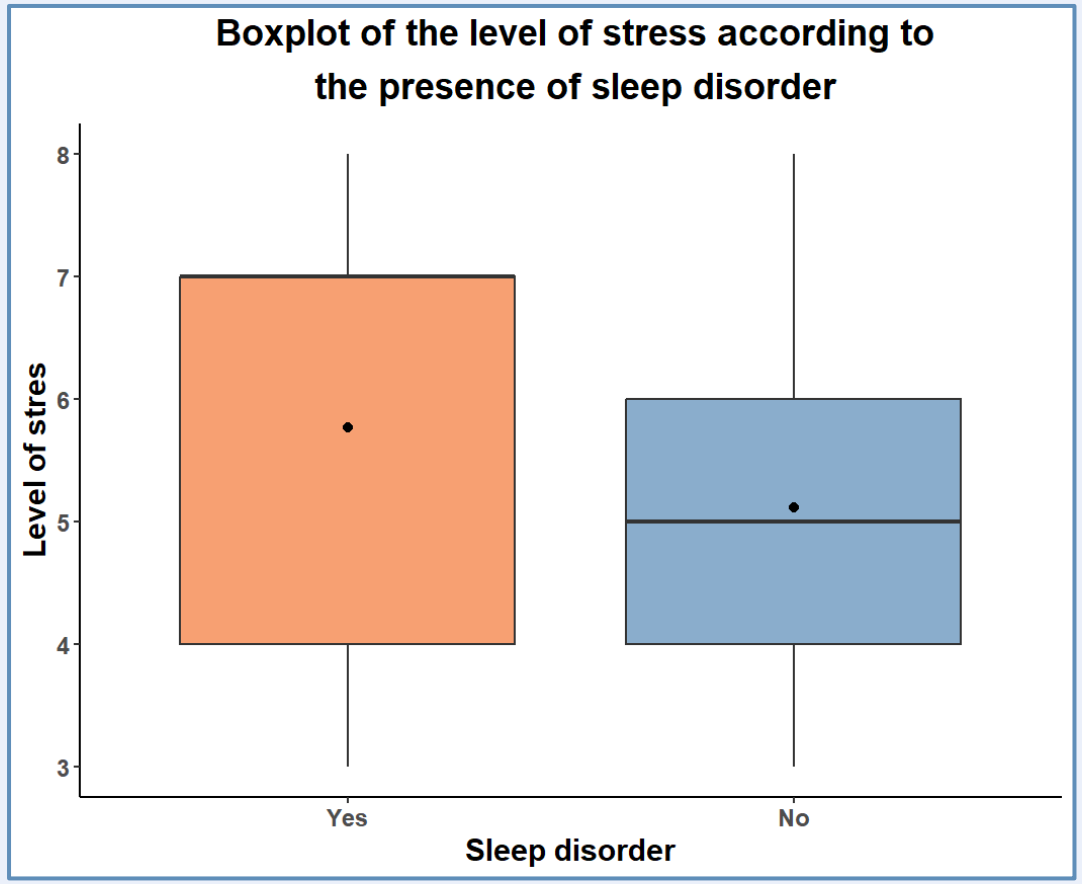
TAKEN INTO CONSIDERATION

- LEVEL OF STRESS
- PHYSICAL ACTIVITY
- BMI
- JOB OCCUPATION
- SLEEP DURATION

WHICH LIFESTYLE FACTORS TO CHANGE?

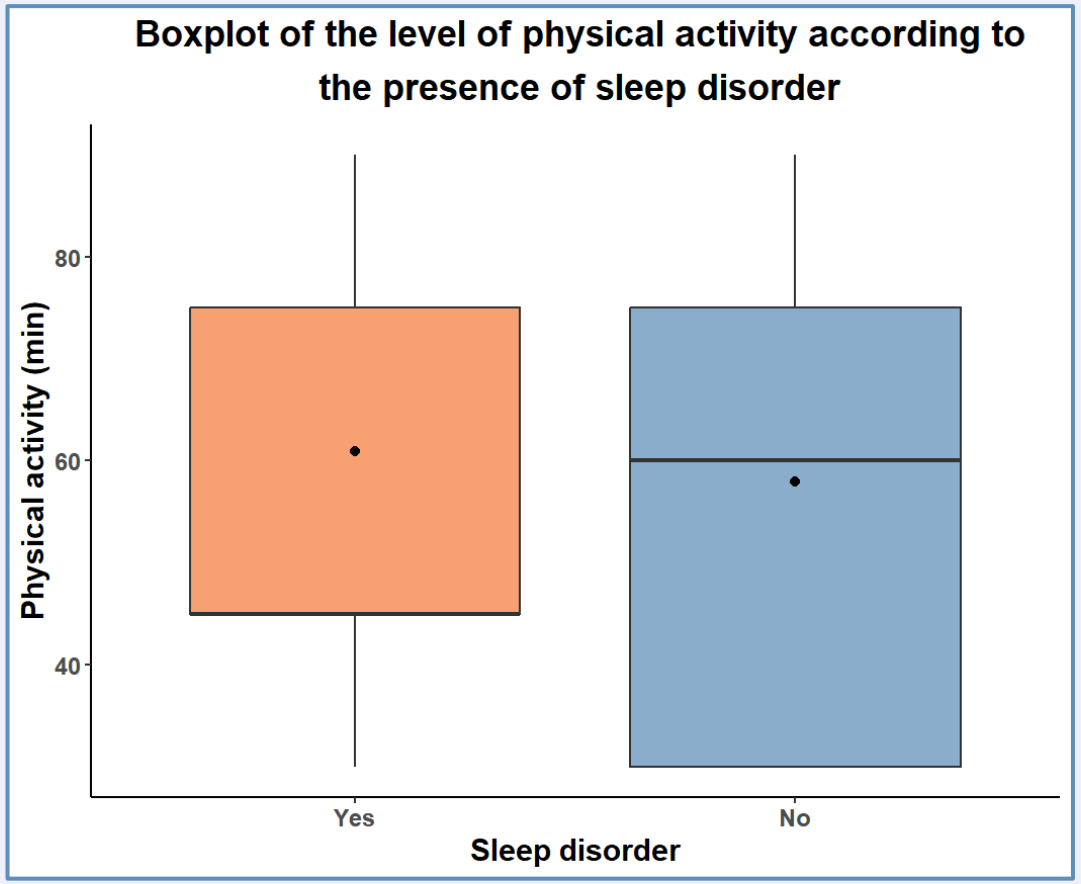
STRESS ◦ PHYSICAL ACTIVITY

z-test



$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$

Conclusion: People with sleep disorders have an higher level of stress.



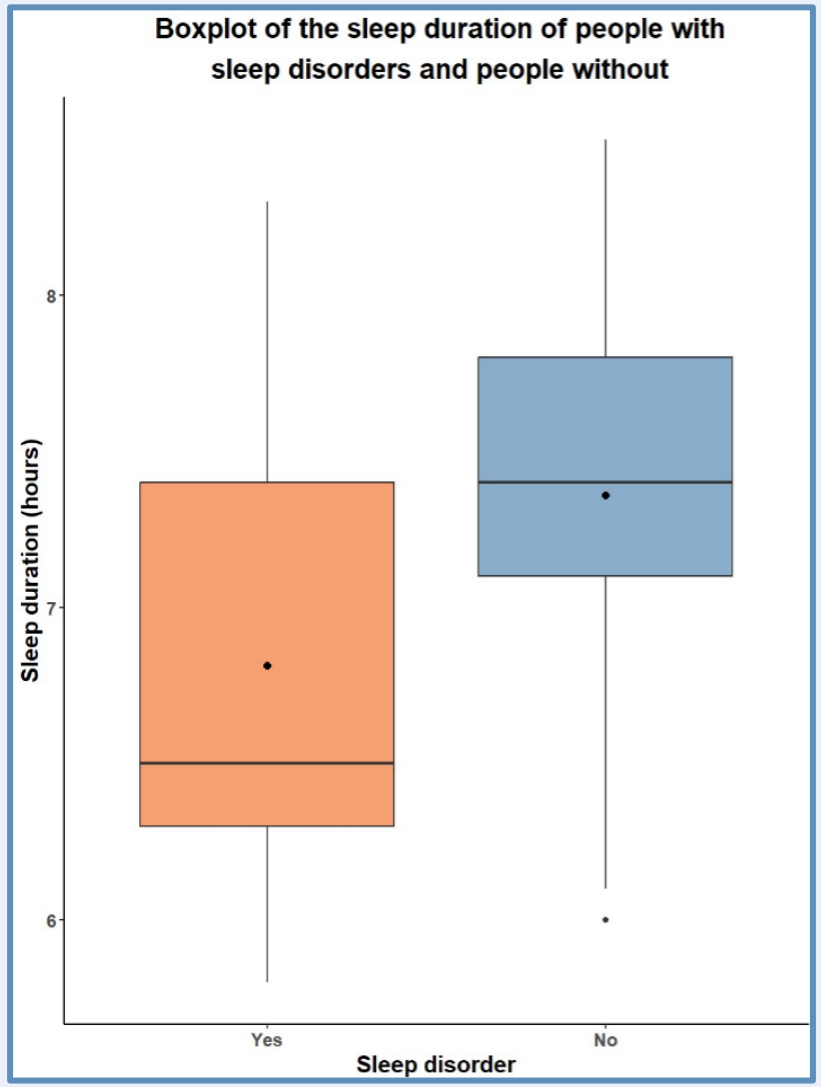
$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 > \mu_2$

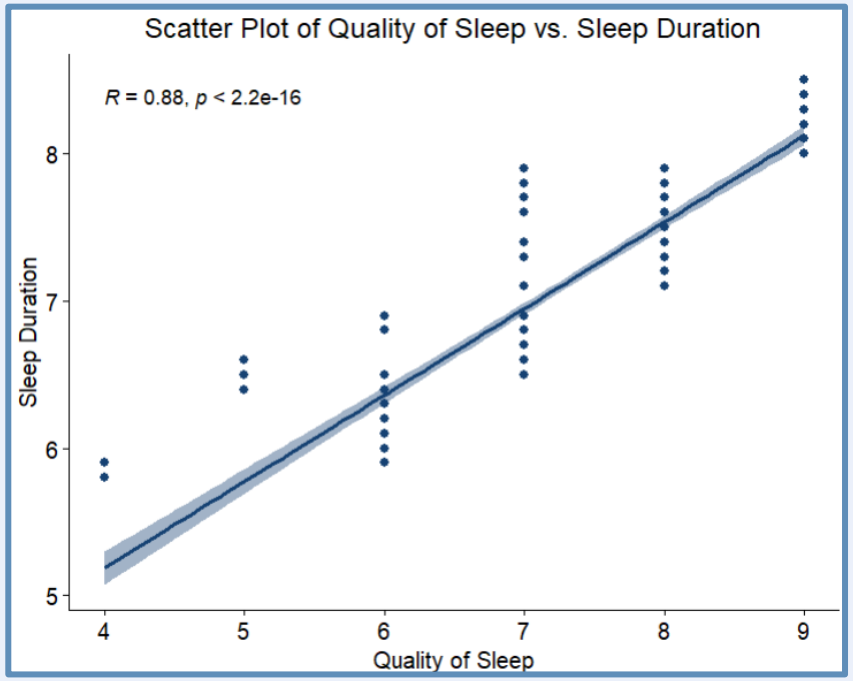
Conclusion: People with and without sleep disorders have **no difference** in physical activity.

WHICH LIFESTYLE FACTORS TO CHANGE?

SLEEP DURATION

z-test





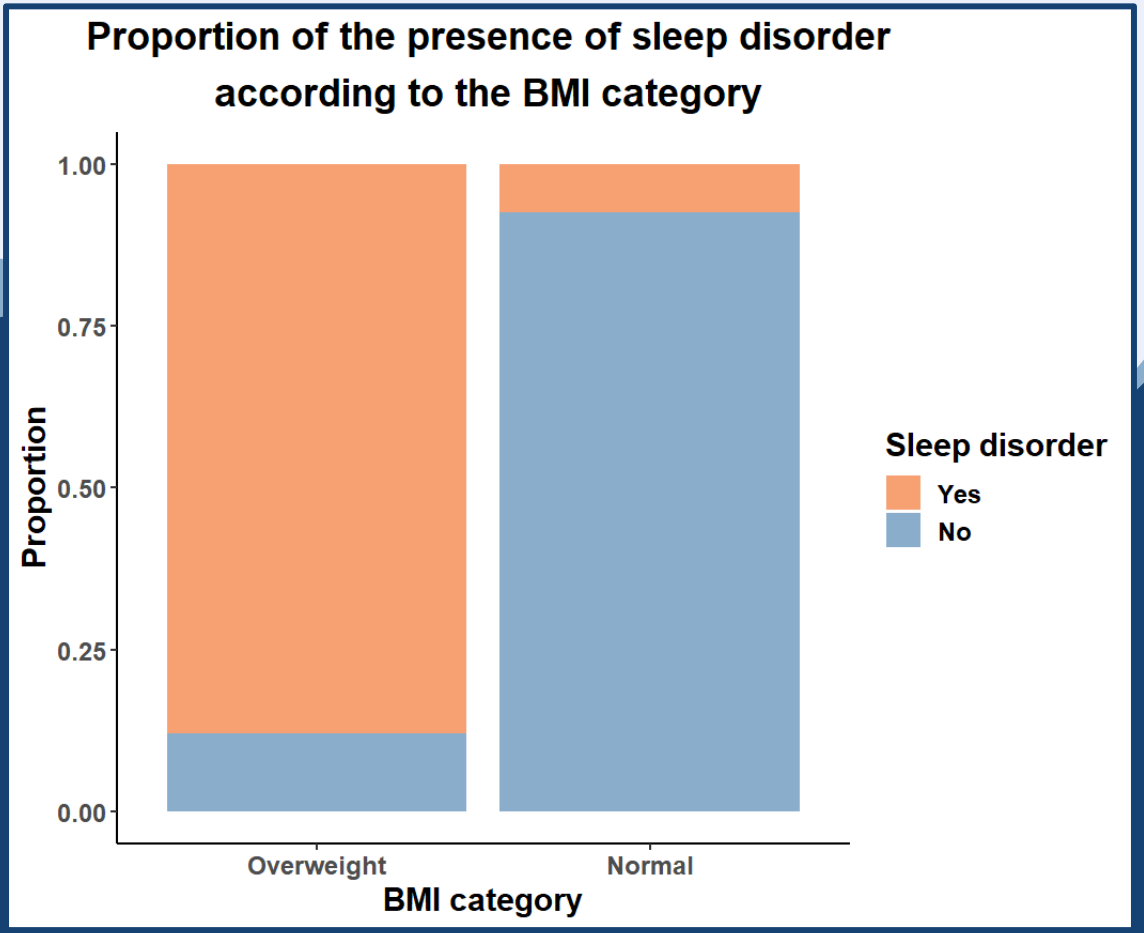
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

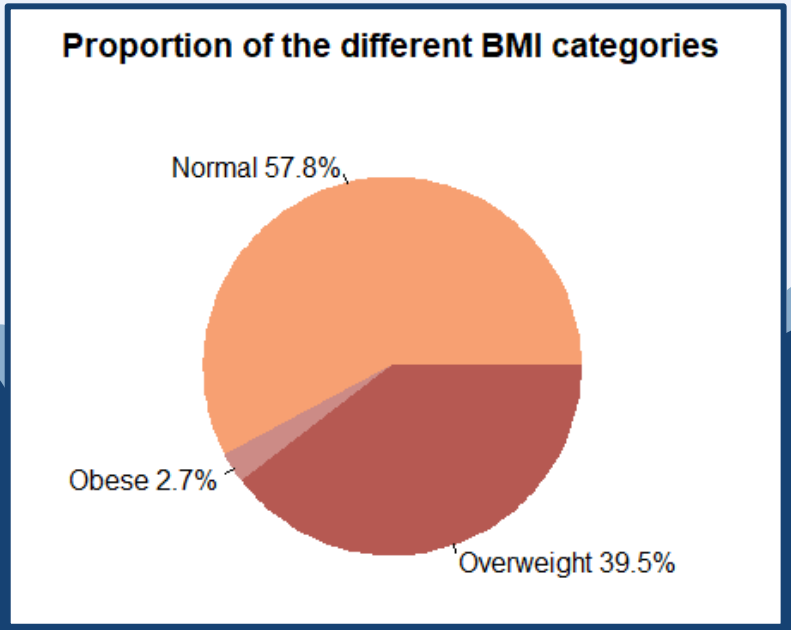
Conclusion: People without sleep disorders sleep more than people with.

WHICH LIFESTYLE FACTORS TO CHANGE?

BMI

Inference on proportion



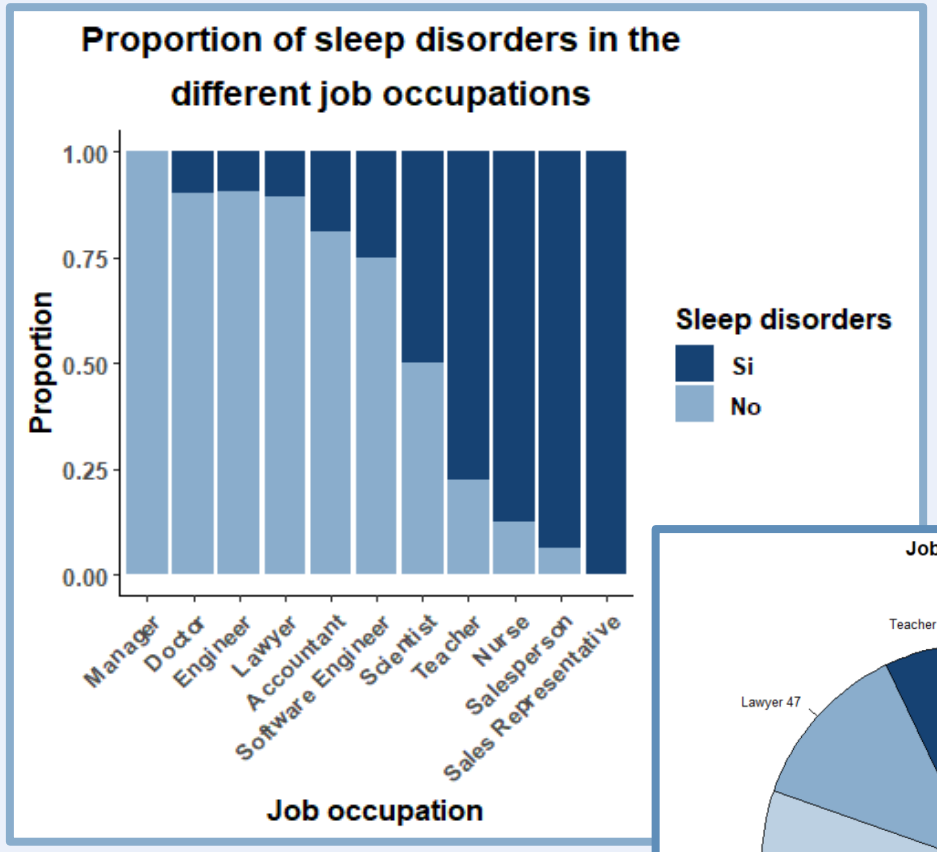


The two categories **Overweight** and **Obese,** will be fused together.

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 < p_2$$

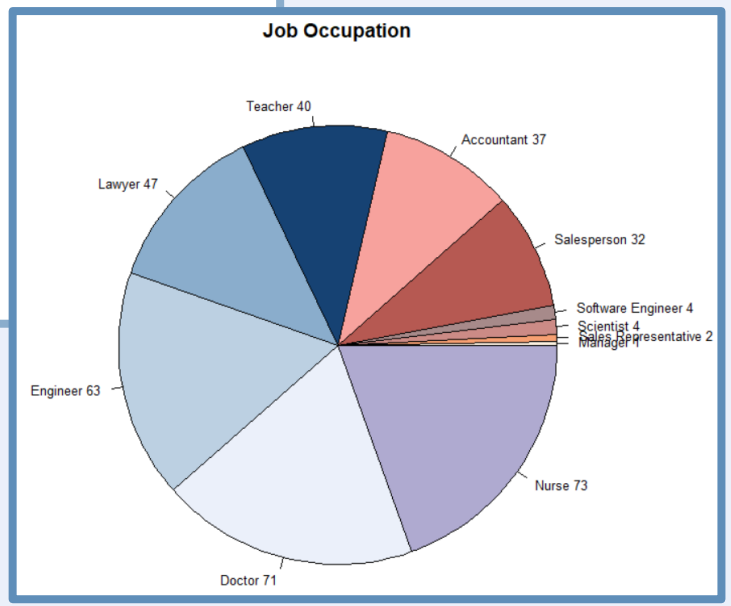
Conclusion: Overweight people show the tendency to have sleep disorders more often than people within the Normal BMI category.

REQUIREMENTS FOR A REGRESSION LOGISTIC MODEL



> variables must be **independent** of one another.
> data should represent **unrepeated** or independent phenomena.
> each variable can be represented using **binary categories**.

Ex: BMI categories: normal weight or not.



> requires a **significant sample size** for each outcome of every variable.

Ex: The choice of job occupation is a hard variable to take into consideration with all its possible outcomes because of the high variability of sample size (minimum: 1 (Manager), maximum: 73 (Nurse))

REGRESSION LOGISTIC MODEL

> **Logistic regression** can estimate the **probabilities** of events, determining a relationship between features and the probabilities of outcomes.

TRAIN 1

CREATE 2

TEST 3

MEASURE 4

1. Training dataset: 299 people
   Test dataset: 75 people

2. Create a logistic model based on the training dataset.

$$\log(odds) = 5.6876 + 1.6256x_{nurse} - 3.8679x_{normal\ BMI} - 0.6359x_{sleep\ duration}$$

$$P(Y = 1) = \frac{1}{1 + e^{-\log(odds)}}$$

$$f(x_{nurse}) = \begin{cases} 1, & person\ is\ a\ nurse \\ 0, & person\ is\ not\ a\ nurse \end{cases}$$

$$f(x_{normal\ BMI}) = \begin{cases} 1, & person\ has\ a\ normal\ BMI \\ 0, & person\ doesn't\ have\ a\ normal\ BMI \end{cases}$$

$$Y = \begin{cases} 1, & person\ has\ sleep\ disorders \\ 0, & person\ doesn't\ have\ sleep\ disorders \end{cases}$$

REGRESSION LOGISTIC MODEL

> **Logistic regression** can estimate the **probabilities** of events, determining a relationship between features and the probabilities of outcomes.

TRAIN 1

CREATE 2

TEST 3

MEASURE 4

3. Test the robustness of the model on the test dataset.

CONFUSION MATRIX

		REAL ONE	
		NO	***YES***
PREDICTED	***NO***	**42**	5
	YES	1	**27**

False negative

False positive

REGRESSION LOGISTIC MODEL

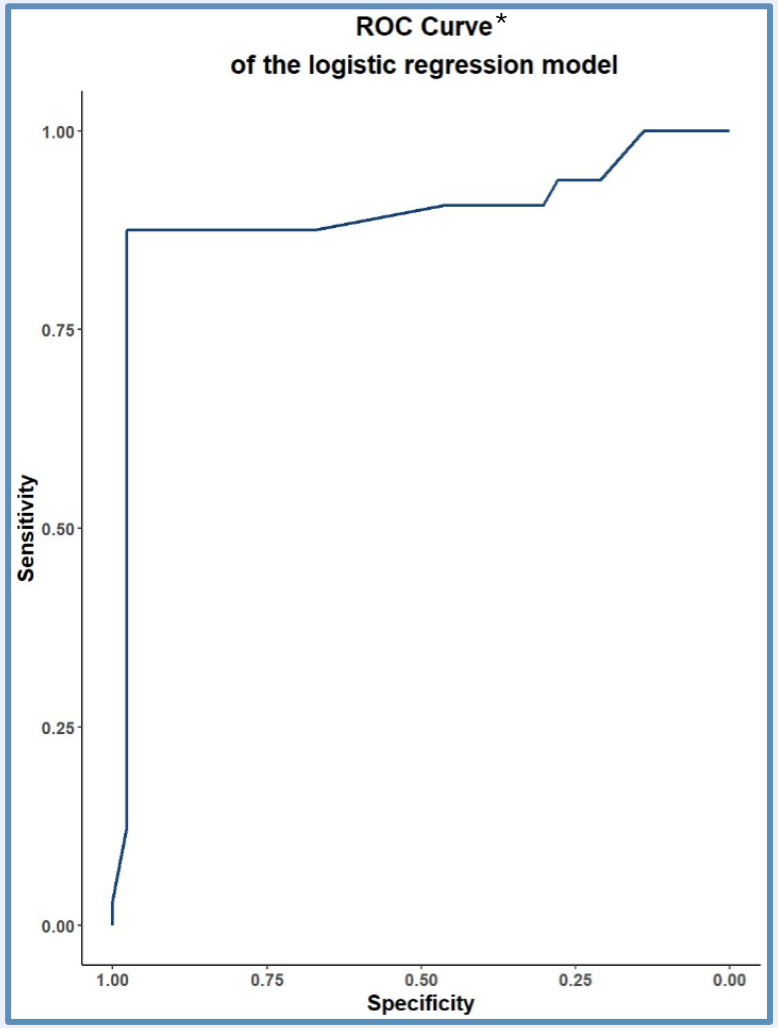
> **Logistic regression** can estimate the **probabilities** of events, determining a relationship between features and the probabilities of outcomes.

TRAIN 1

CREATE 2

TEST 3

MEASURE 4



4. Accuracy, **sensitivity** (true positive rate) and **specificity** (true negative rate).

$$Accuracy = 92\%$$
$$Specificity = 89\%$$
$$Sensitivity = 96\%$$

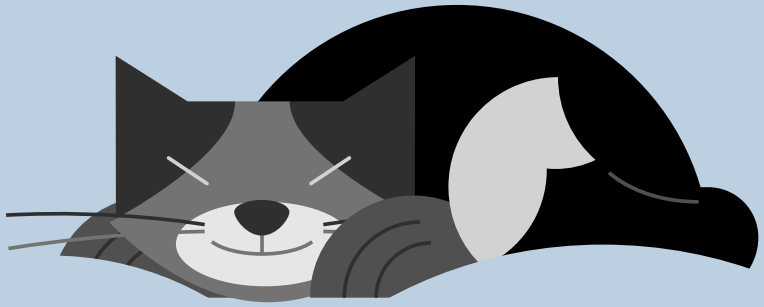
*Receiver Operating Characteristic Curve

LIMITS OF THE STUDY

- **Subjectivity** of the test
- **Illustration** and not scientific **purpose**
- Concerning the job occupation, **not enough people for each job**

CONCLUSIONS

- People with **sleep disorders** have a **worse quality of sleep**.
- Although linked also to stress level, the **probability of the presence of sleeping disorders** is mostly influenced by the **BMI category**, **sleep duration** and by the **job occupation** (**only** in the case of **nurses**).
- A broader study can be more conclusive about the impact of the job occupation on the development of sleep disorders.



THANK YOU

Please take care of your health!