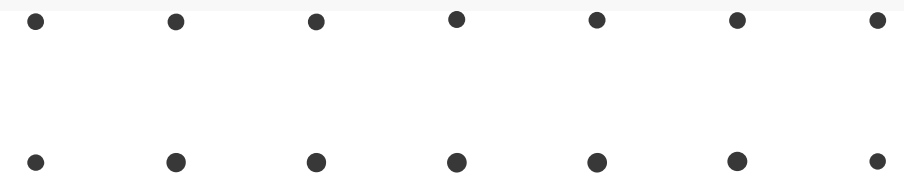
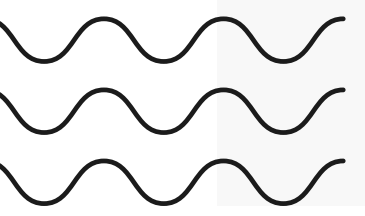


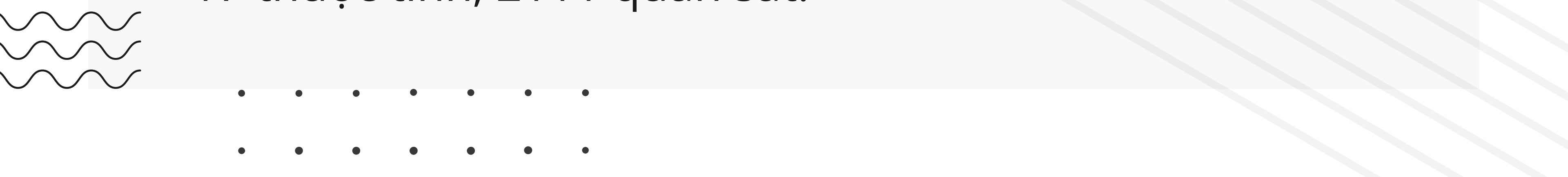
# Obesity Levels

LÂM THY NHÃ  
VÕ YẾN NHI  
NGUYỄN LÊ THANH OANH  
LÊ ĐÌNH PHONG  
VƯƠNG KIẾN PHÁT





# *Tổng quan về bộ dữ liệu*

- Ước lượng về mức độ béo phì dựa trên chế độ ăn uống và lối sống của người dân đến từ Mexico, Peru, Colombia.
  - 17 thuộc tính, 2111 quan sát.
- 

# *Kiểu dữ liệu của các thuộc tính*

Age	float64
Gender	object
Height	float64
Weight	float64
CALC	object
FAVC	object
FCVC	float64
NCP	float64
SCC	object
SMOKE	object
CH20	float64
family_history_with_overweight	object
FAF	float64
TUE	float64
CAEC	object
MTRANS	object
NObeyesdad	object
dtype:	object

# Tiền xử lý dữ liệu:

1

Kiểm tra biến có dữ liệu số nguyên

2

Kiểm tra biến nhị phân

3

Kiểm tra định dạng các biến định tính

4

Kiểm tra định dạng các biến liên tục

5

Thay đổi tên các thuộc tính của dữ liệu

6

Kiểm tra và xử lý Missing Values

7

Kiểm tra các giá trị bất thường của biến định lượng

8

Kiểm tra và xử lý giá trị trùng lặp

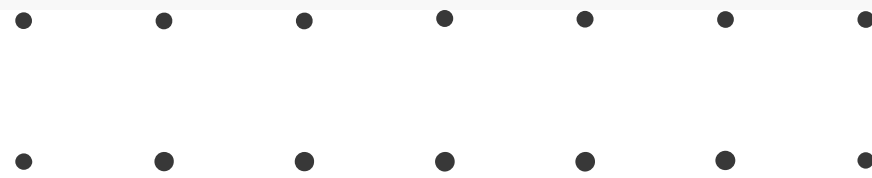
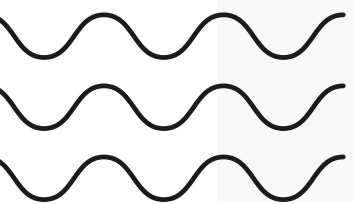
9

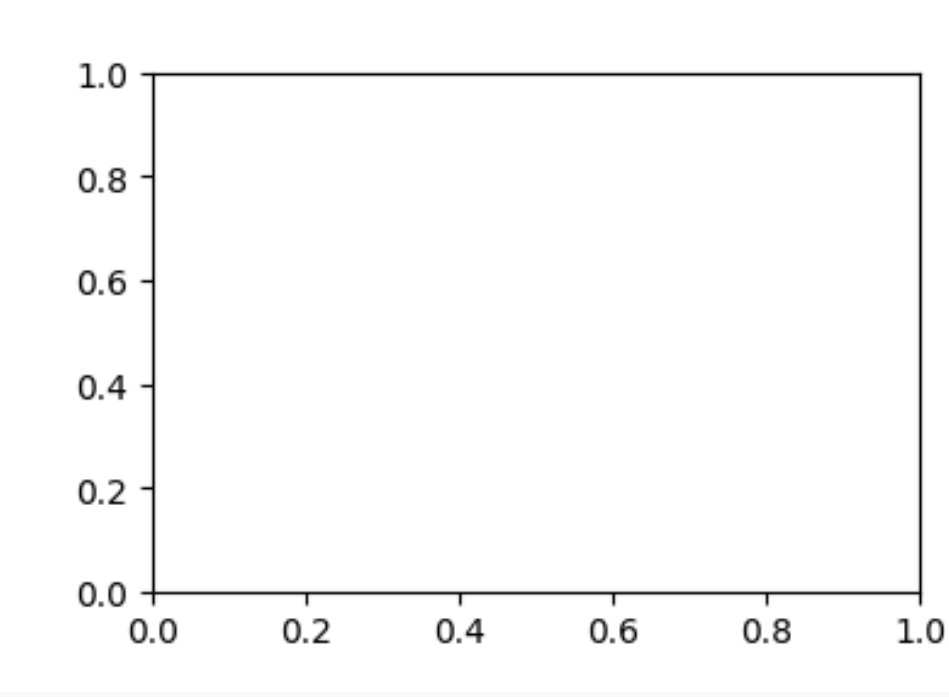
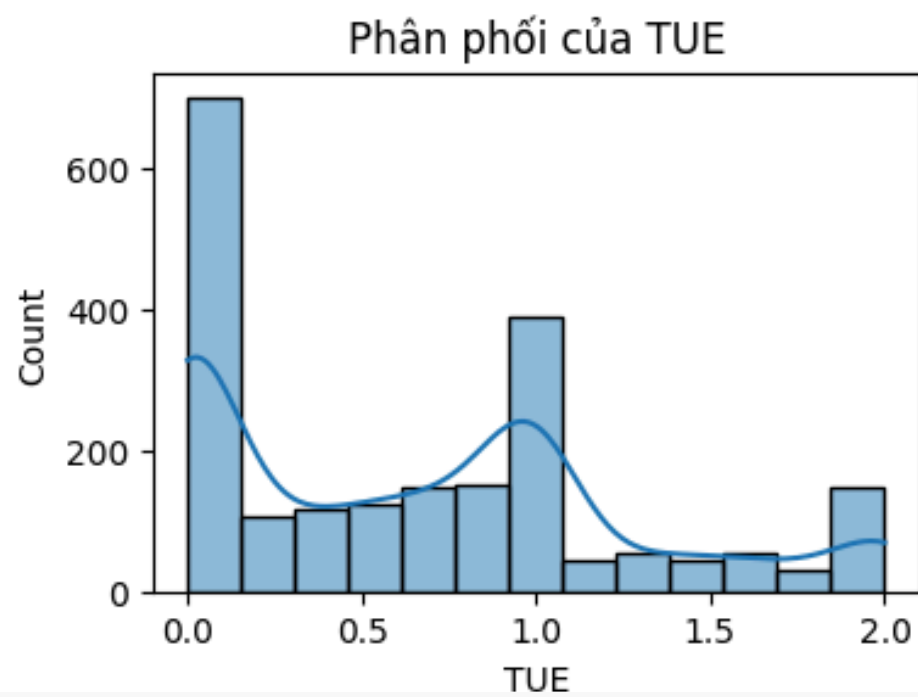
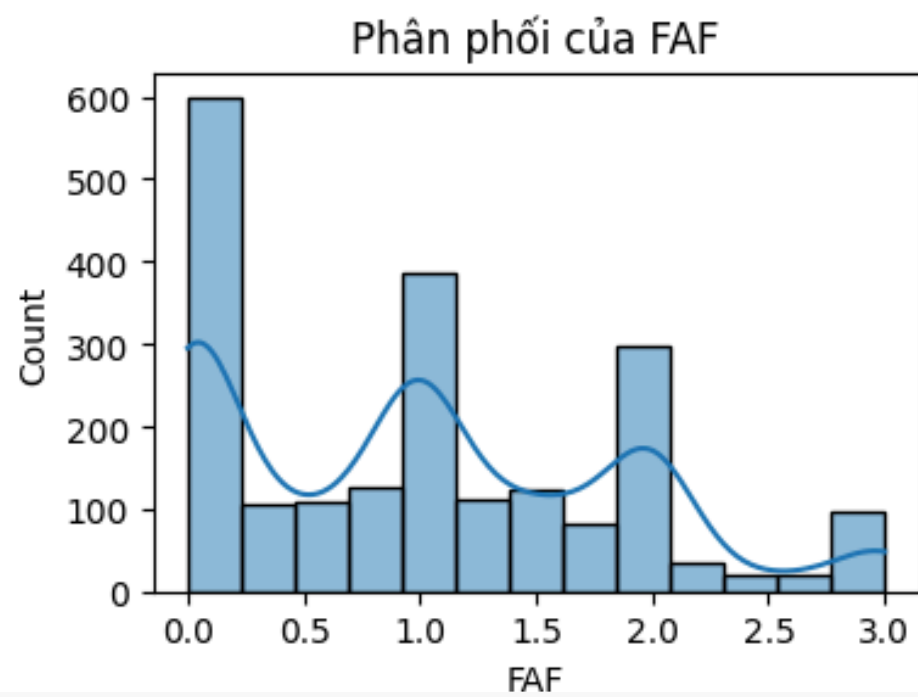
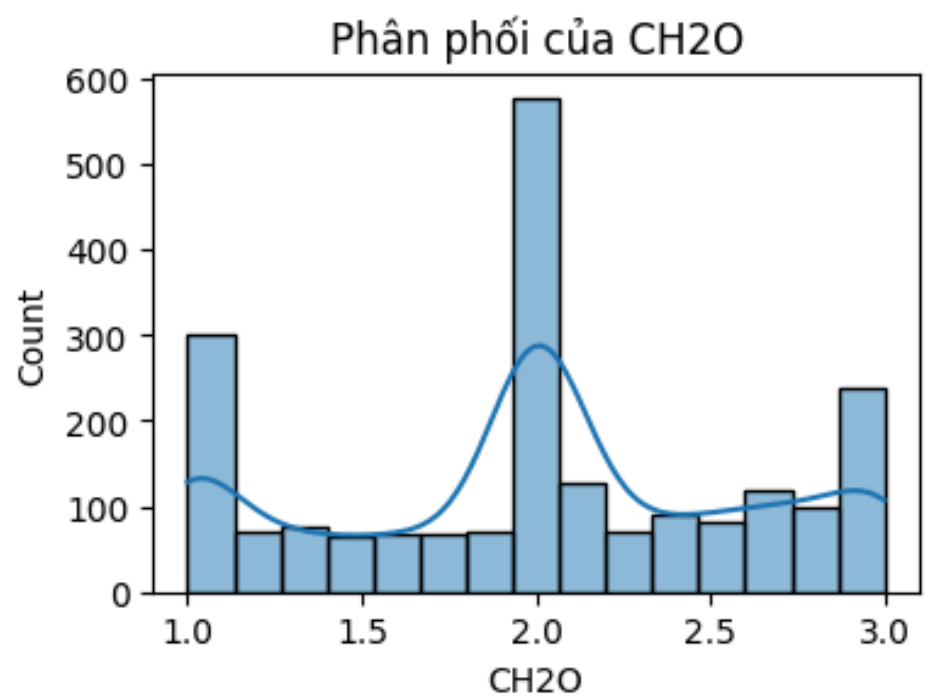
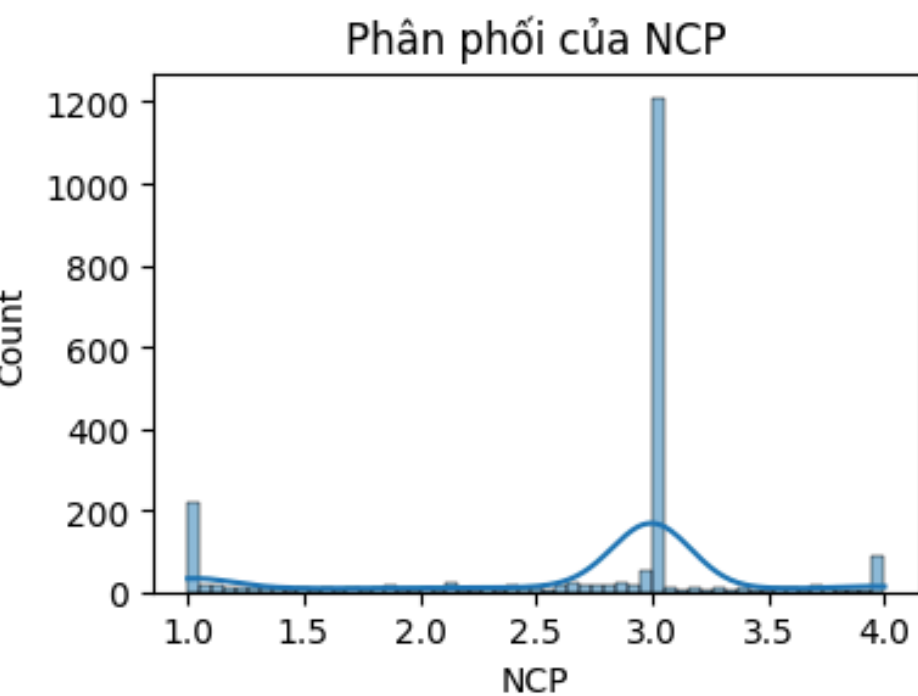
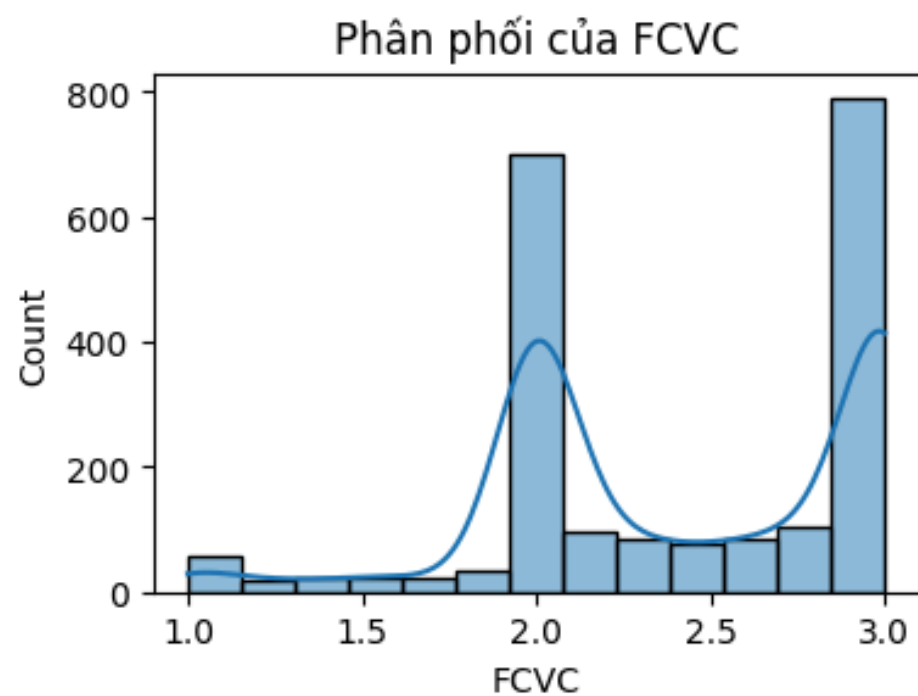
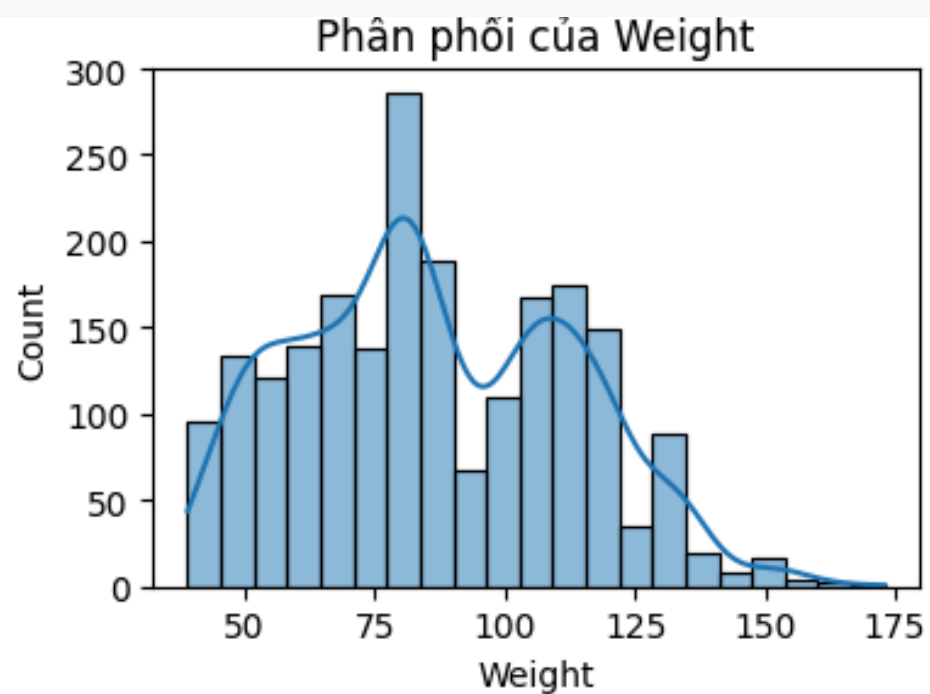
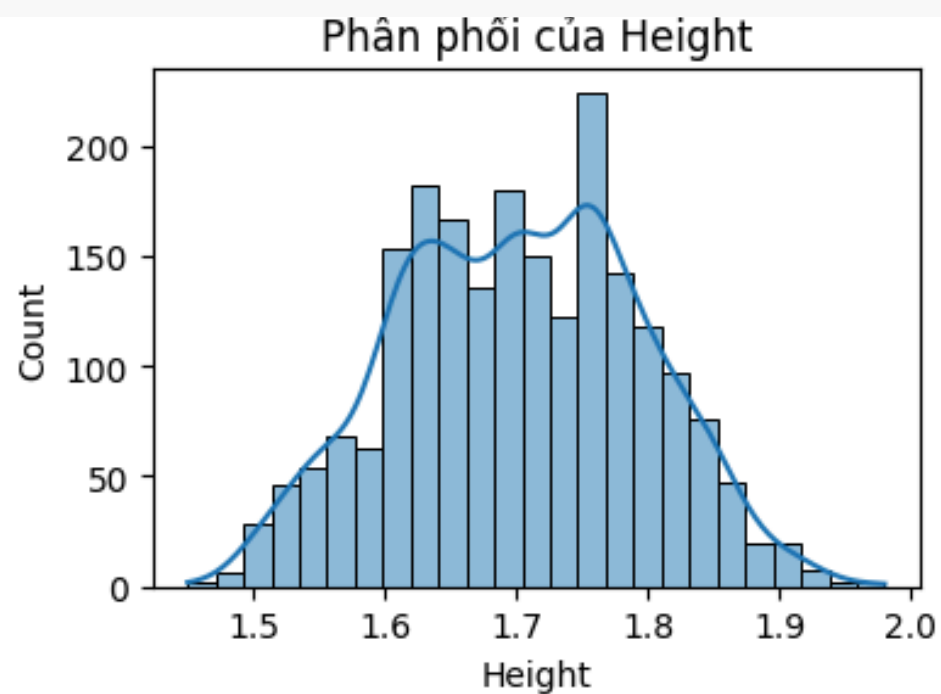
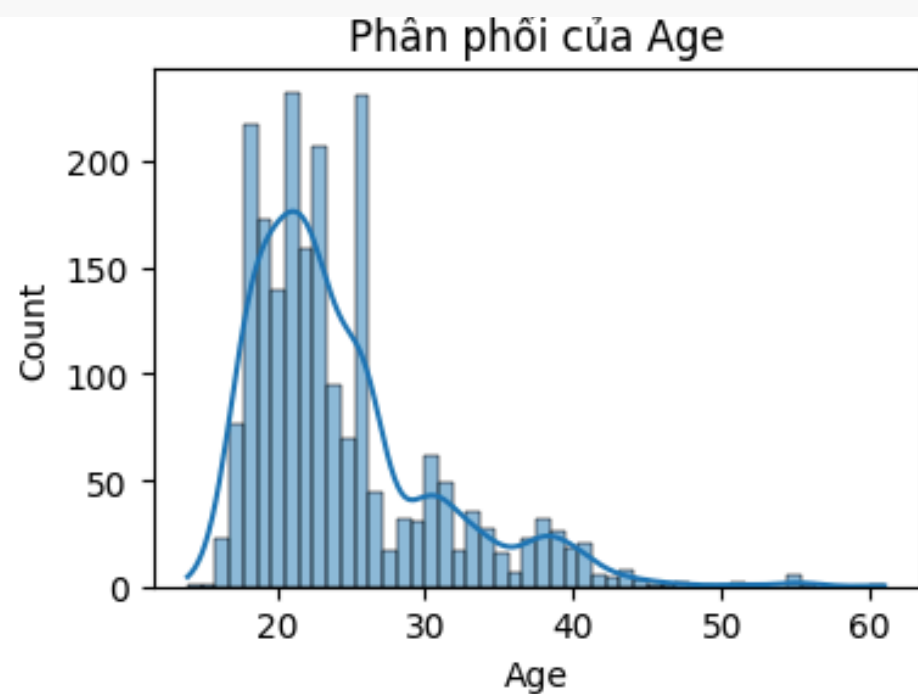
Thêm thuộc tính BMI

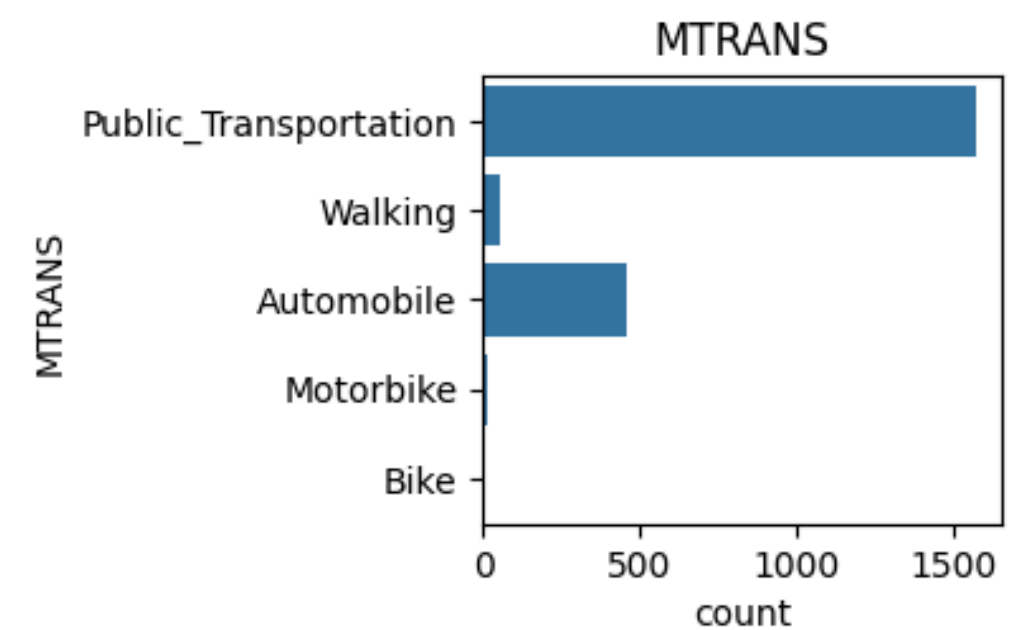
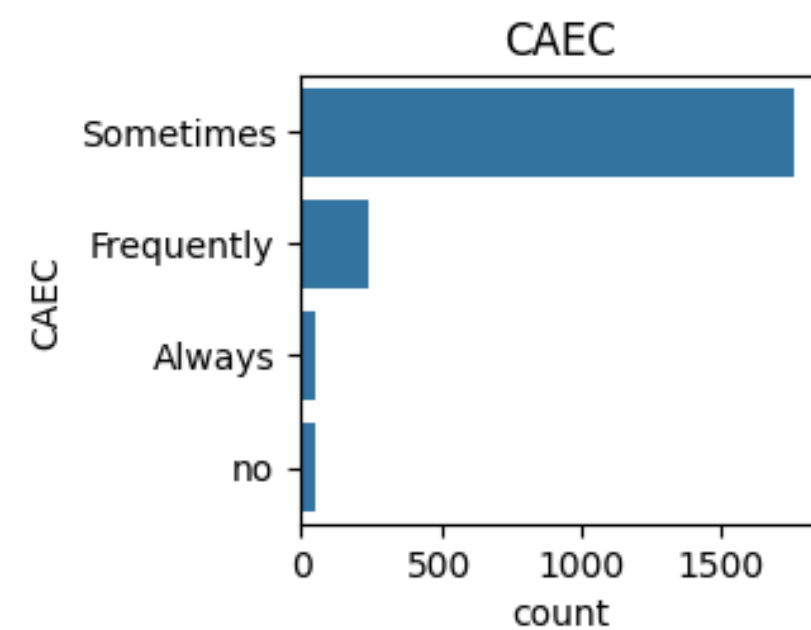
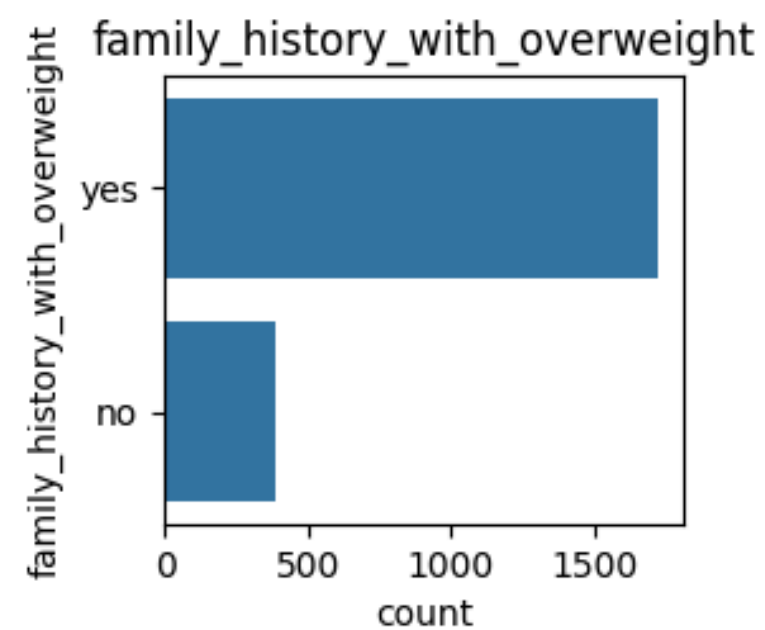
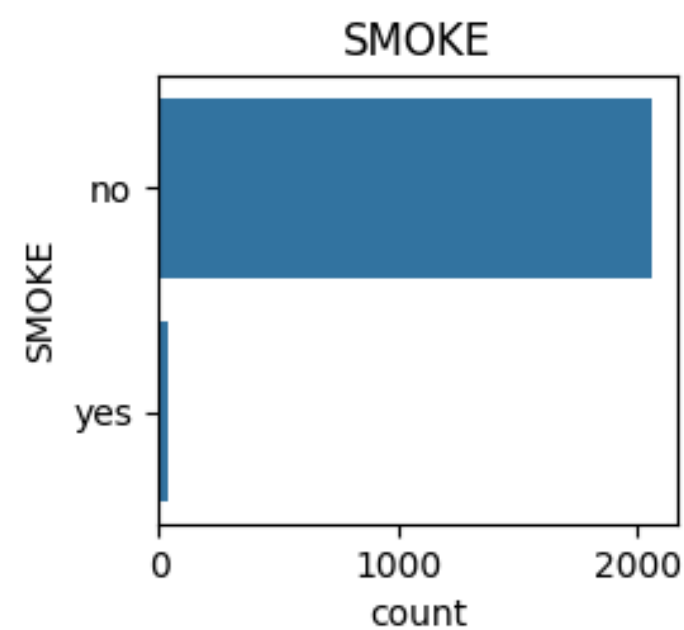
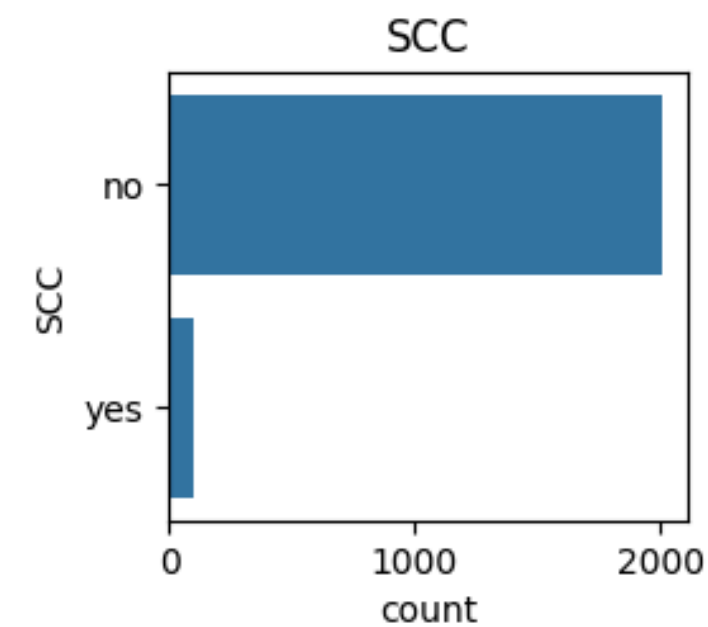
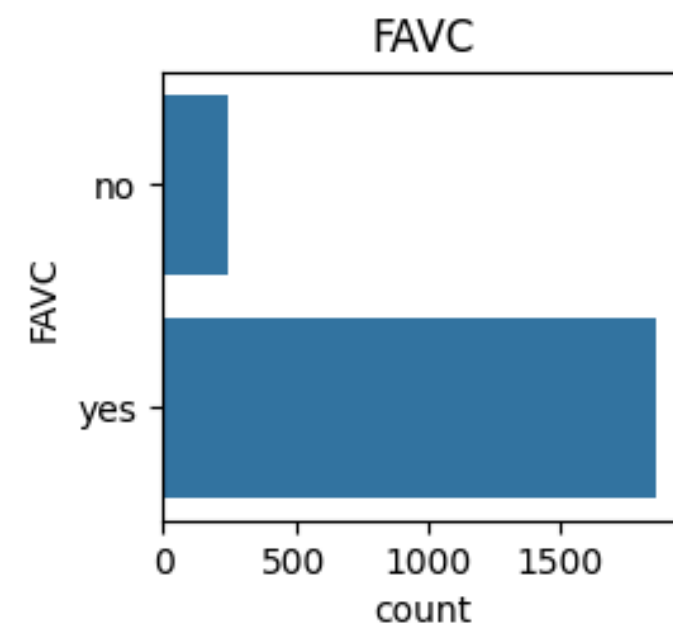
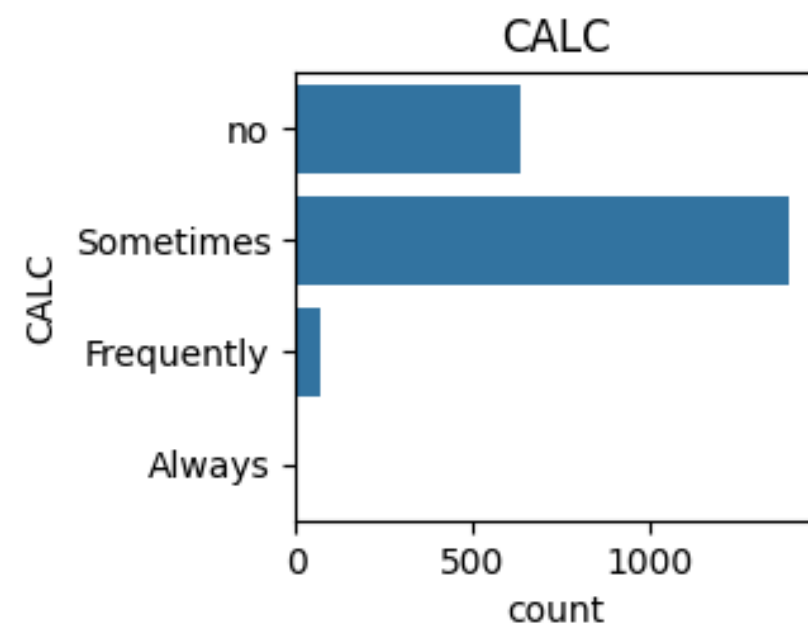
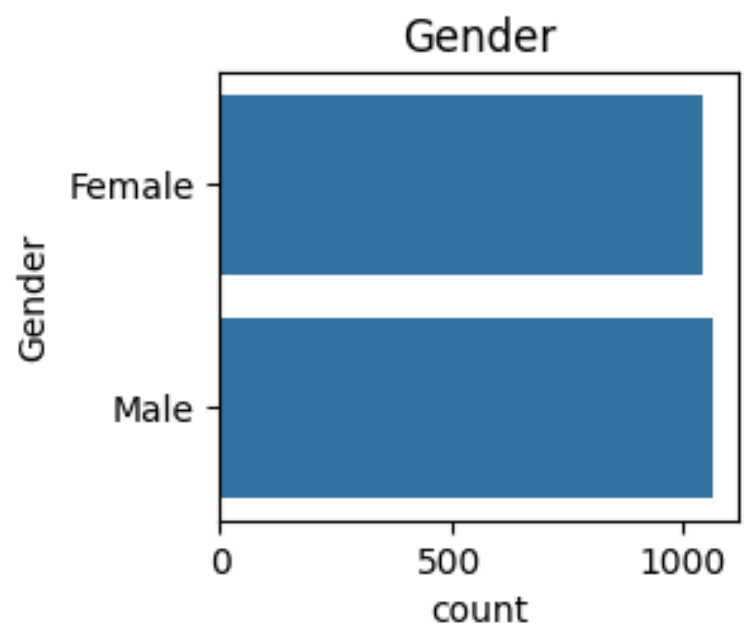


# *Phân phối của các biến độc lập*

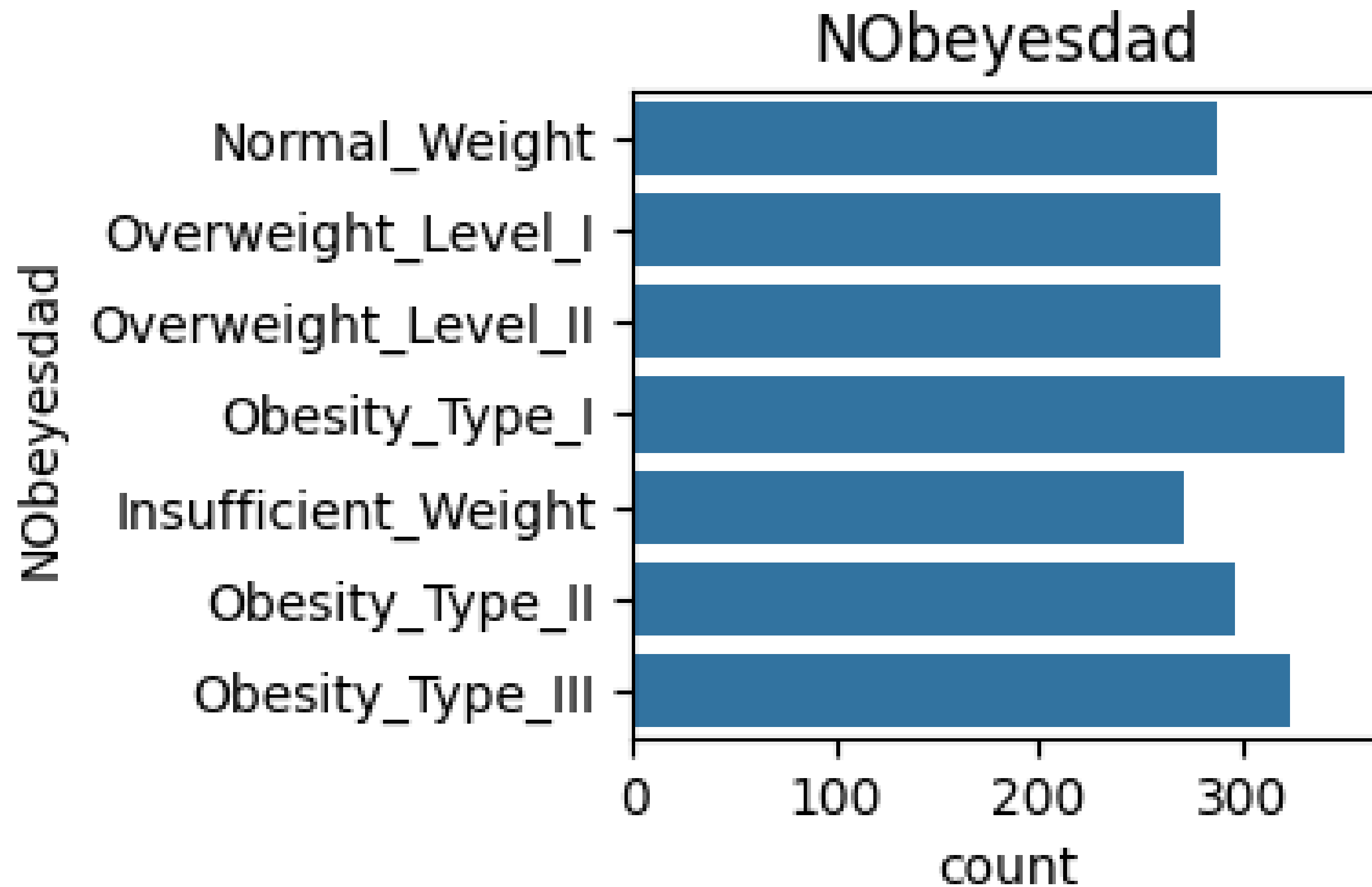
---







# *Phân phối của biến phụ thuộc*

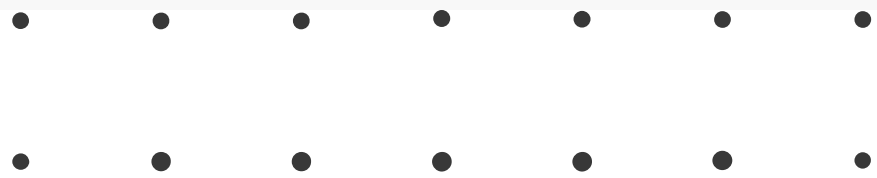
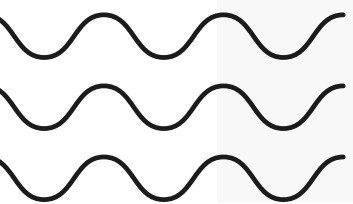




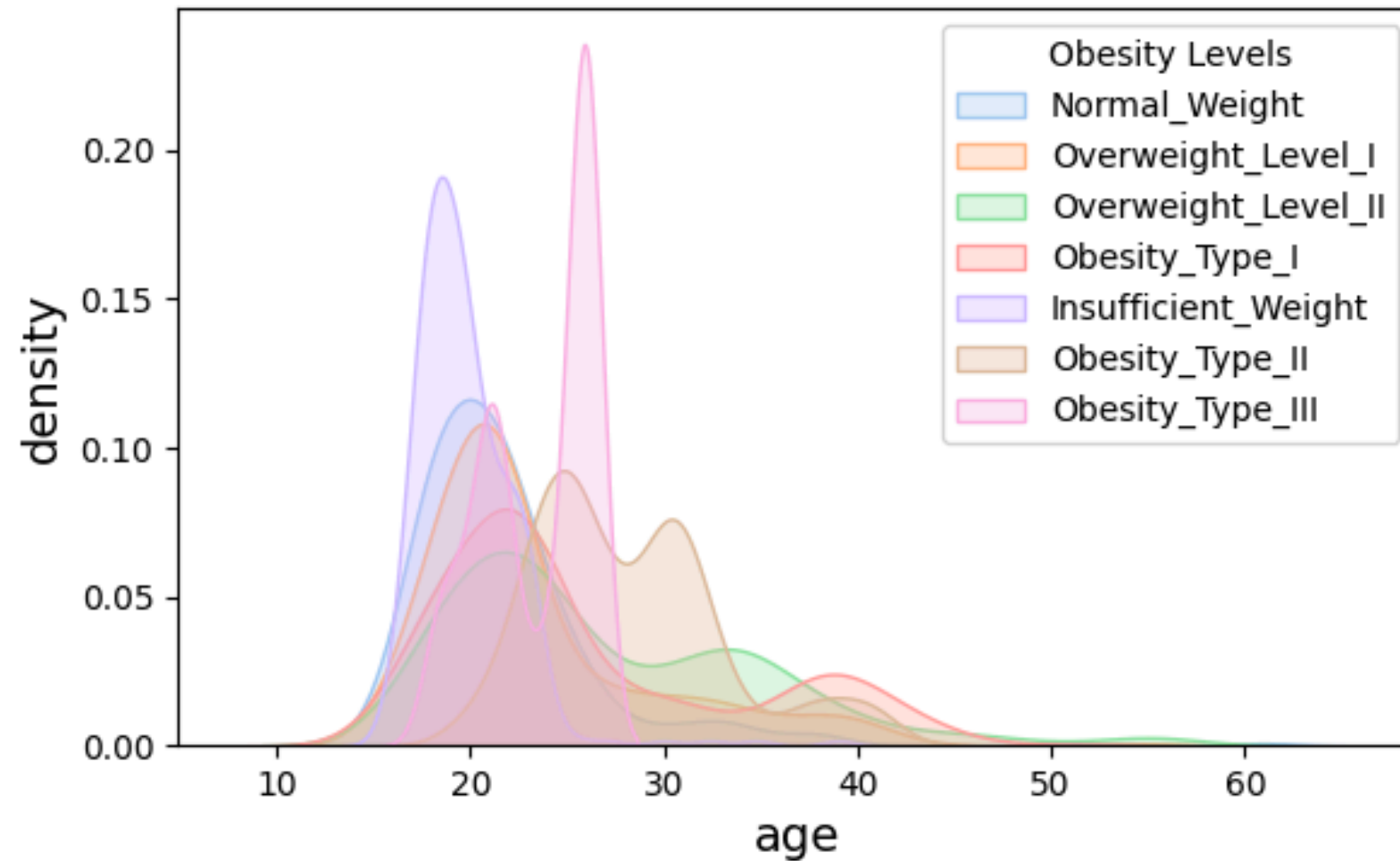


# *Phân phối của các biến độc lập với biến phụ thuộc*

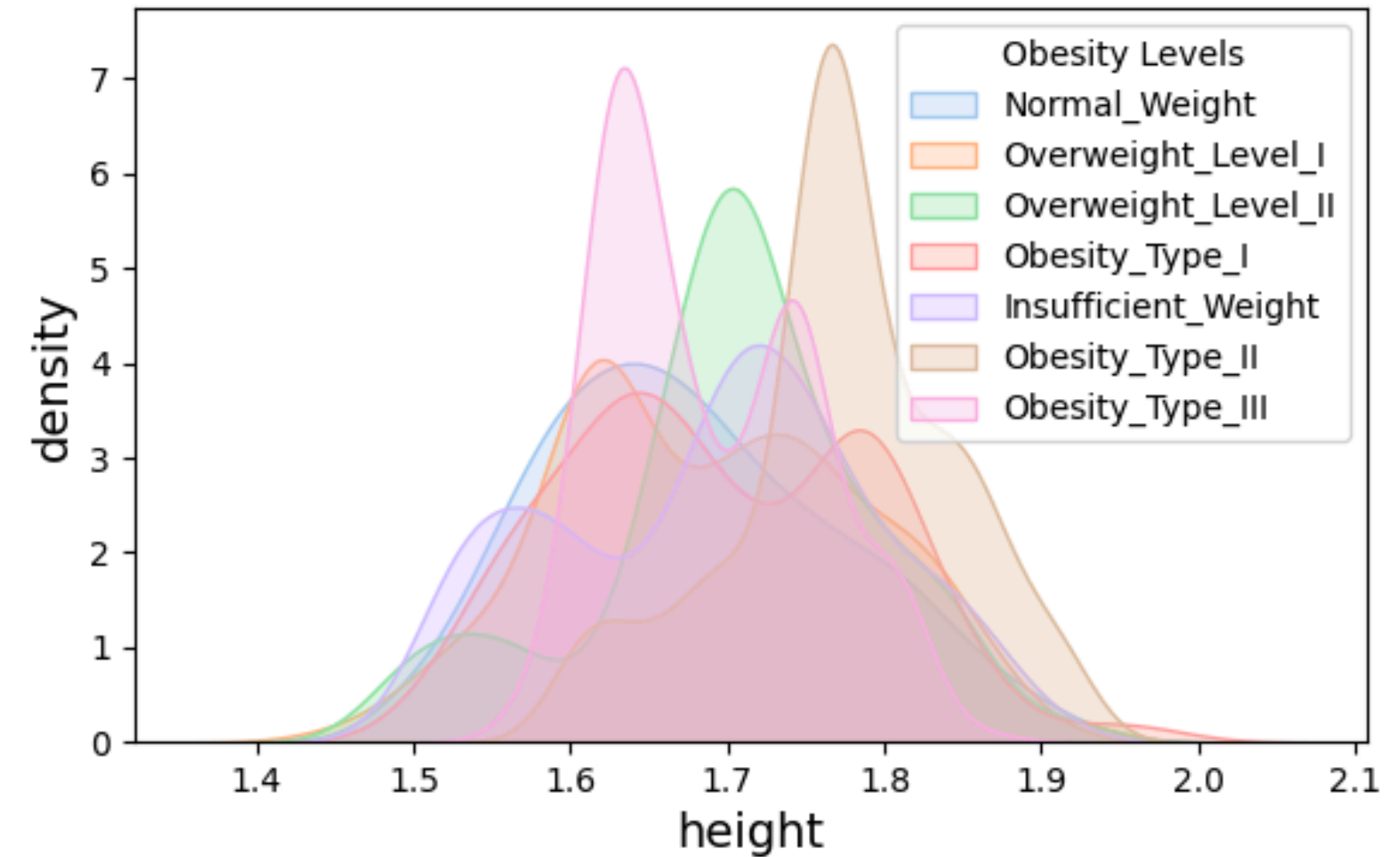
---



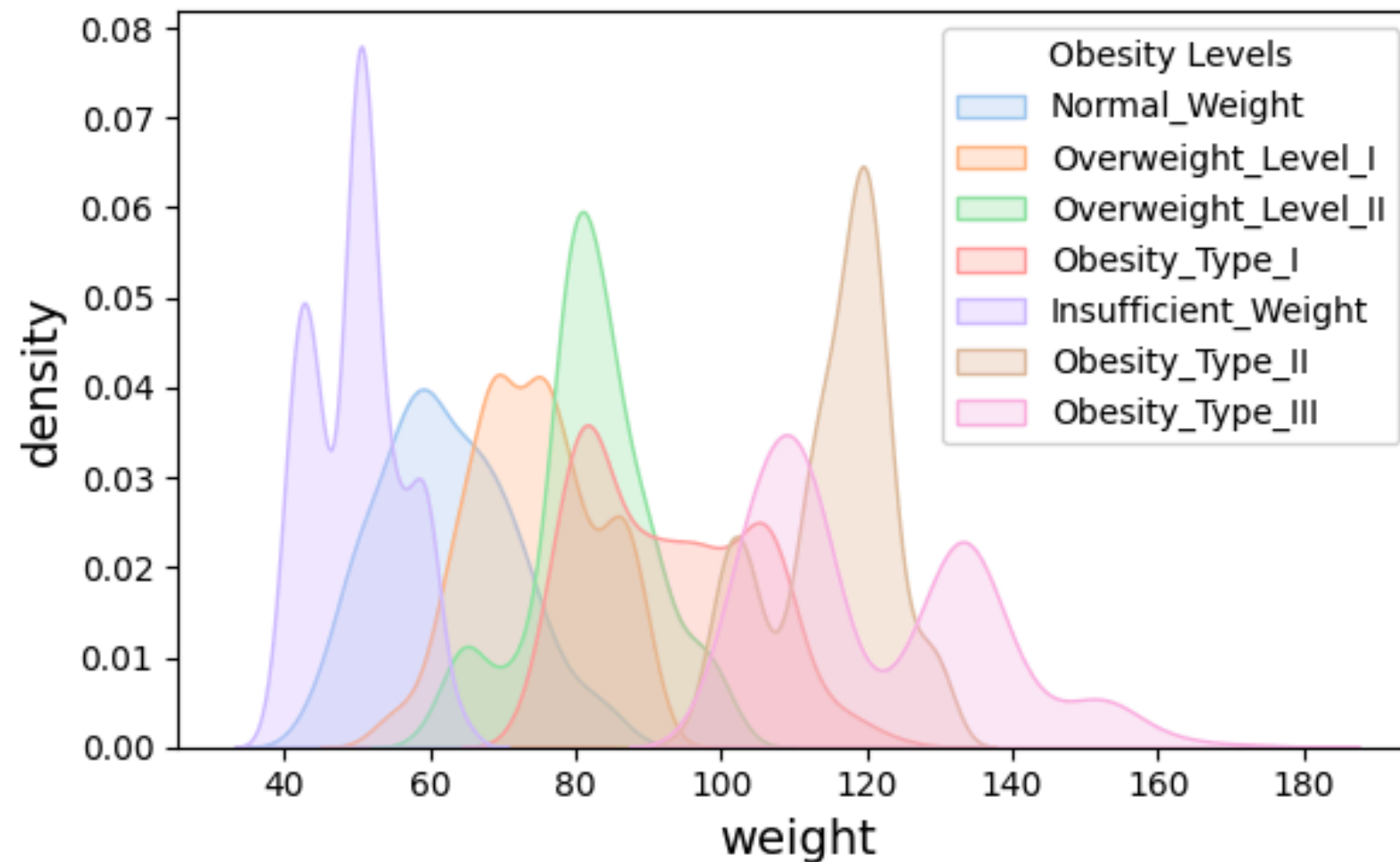
age density for obesity levels



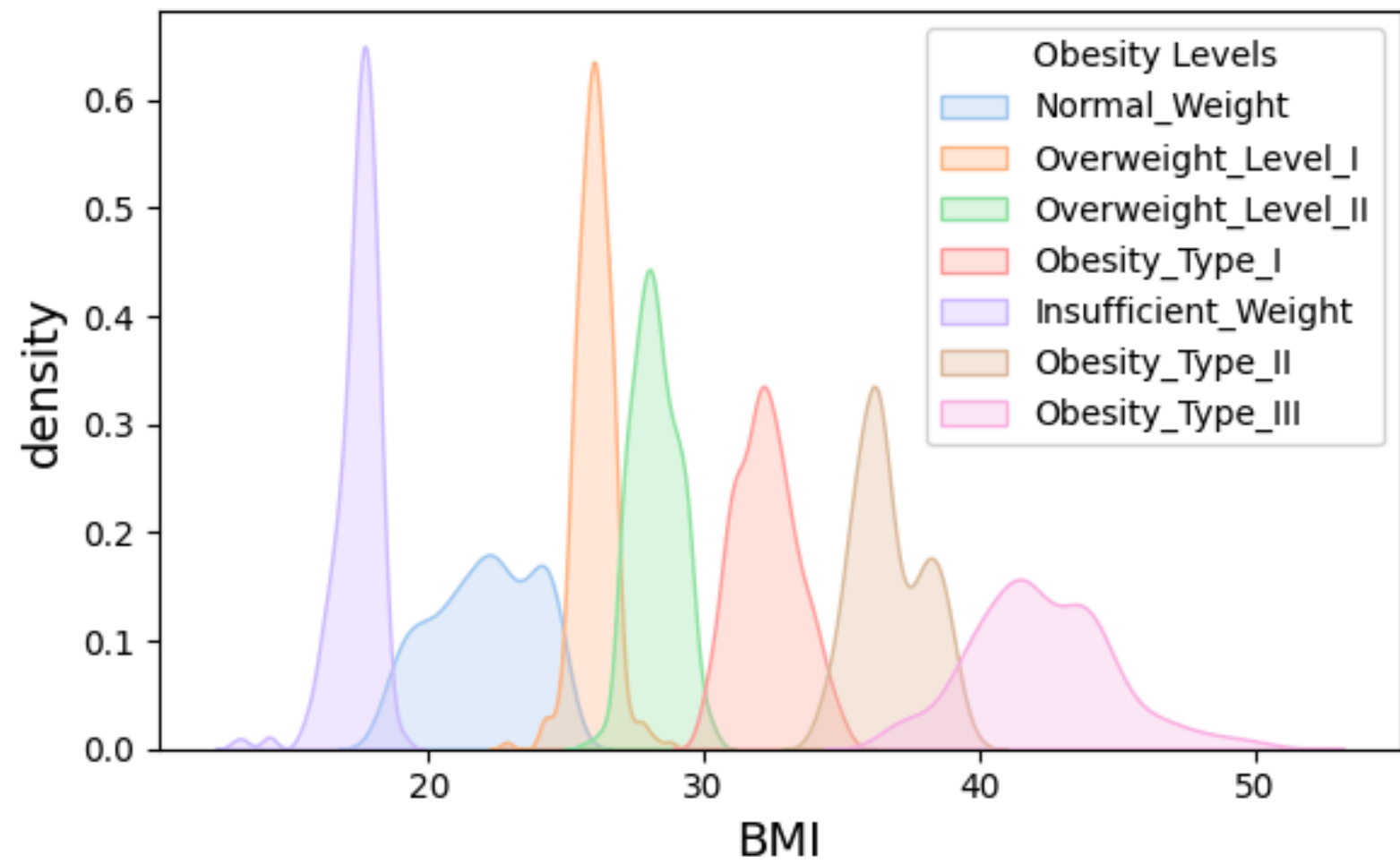
height density for obesity levels



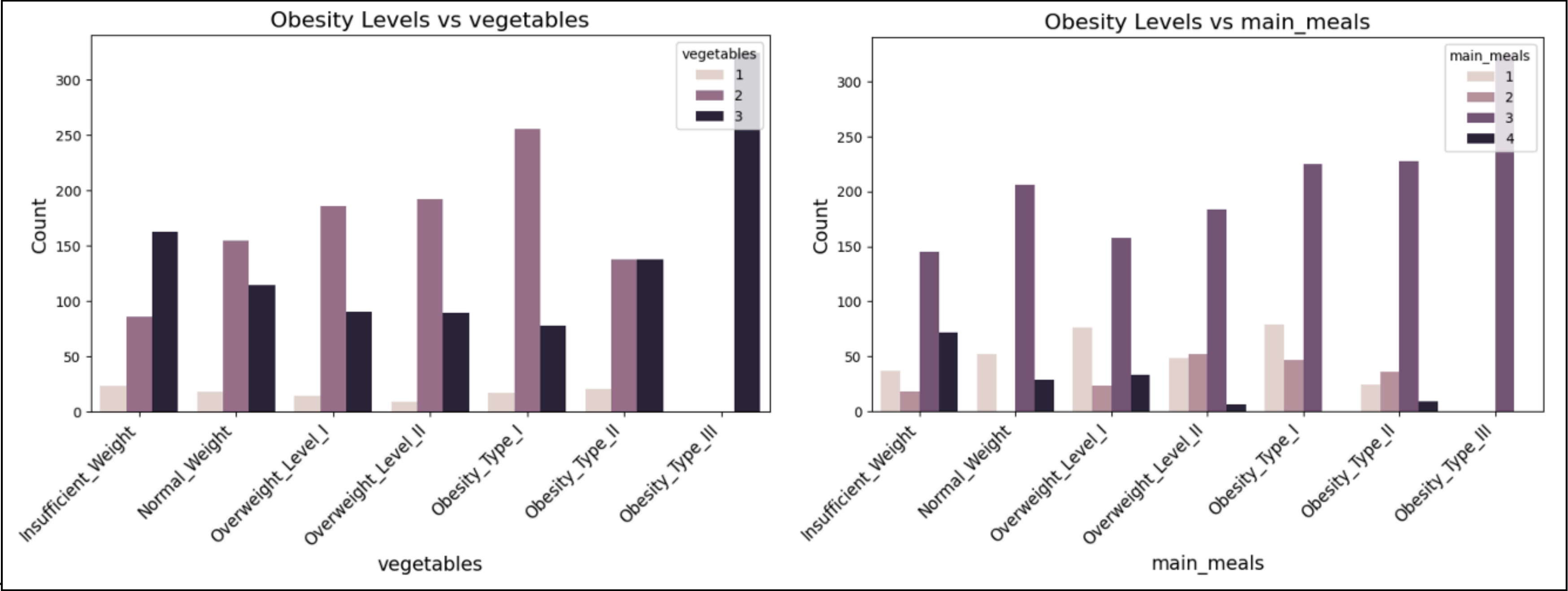
weight density for obesity levels



BMI density for obesity levels



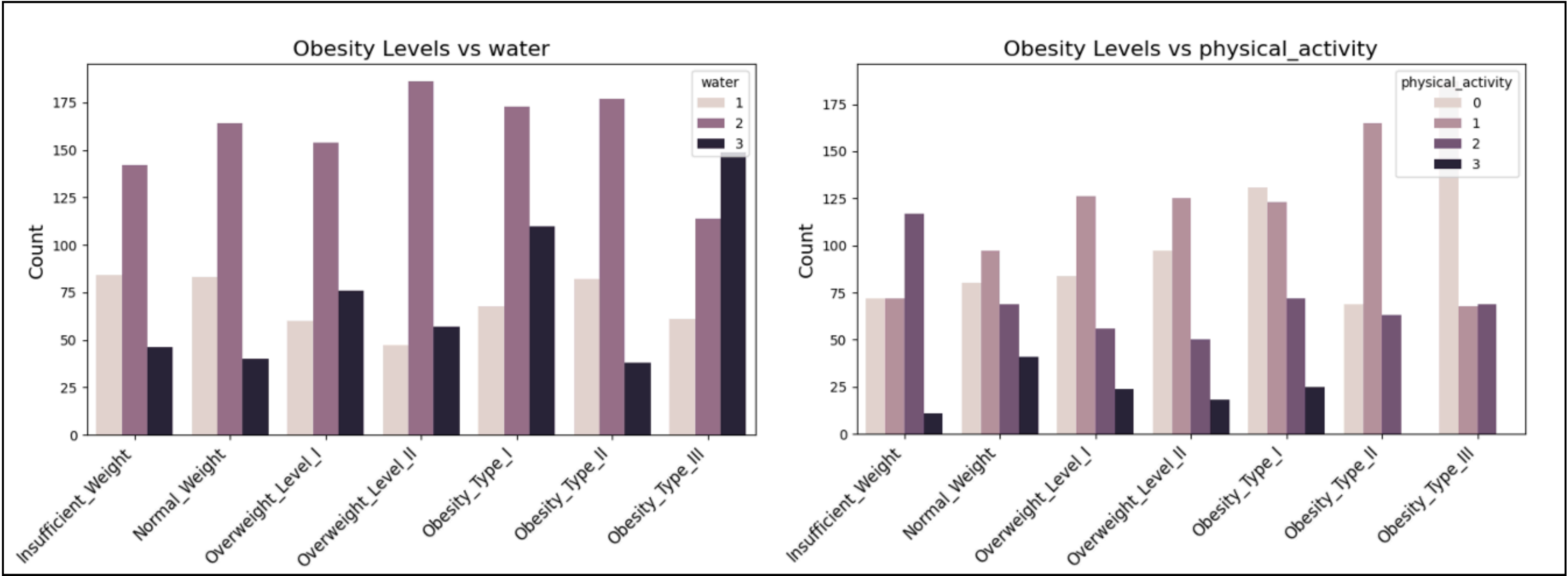
# SỰ TƯƠNG QUAN GIỮA CÁC BIẾN ĐỘC LẬP VỚI BIẾN PHỤ THUỘC



BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA MỨC ĐỘ TIÊU THỤ RAU CỦ MỖI BỮA ĂN VÀ CÁC MỨC ĐỘ BÉO PHÌ.

BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA SỐ BỮA ĂN CHÍNH MỖI NGÀY VÀ CÁC MỨC ĐỘ BÉO PHÌ.

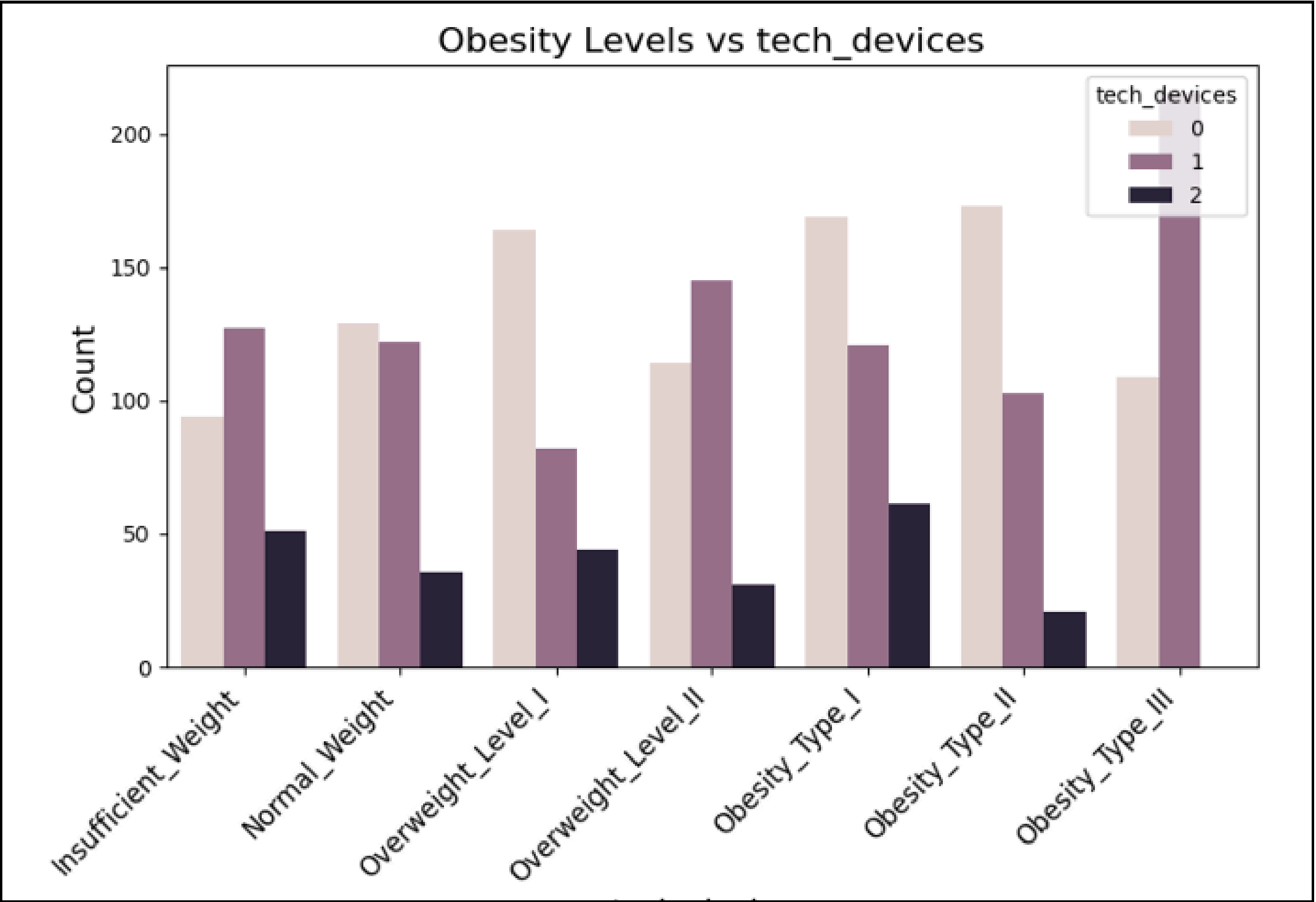
# SỰ TƯƠNG QUAN GIỮA CÁC BIẾN ĐỘC LẬP VỚI BIẾN PHỤ THUỘC



BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA LƯỢNG NƯỚC UỐNG MỖI NGÀY VÀ CÁC MỨC ĐỘ BÉO PHÌ.

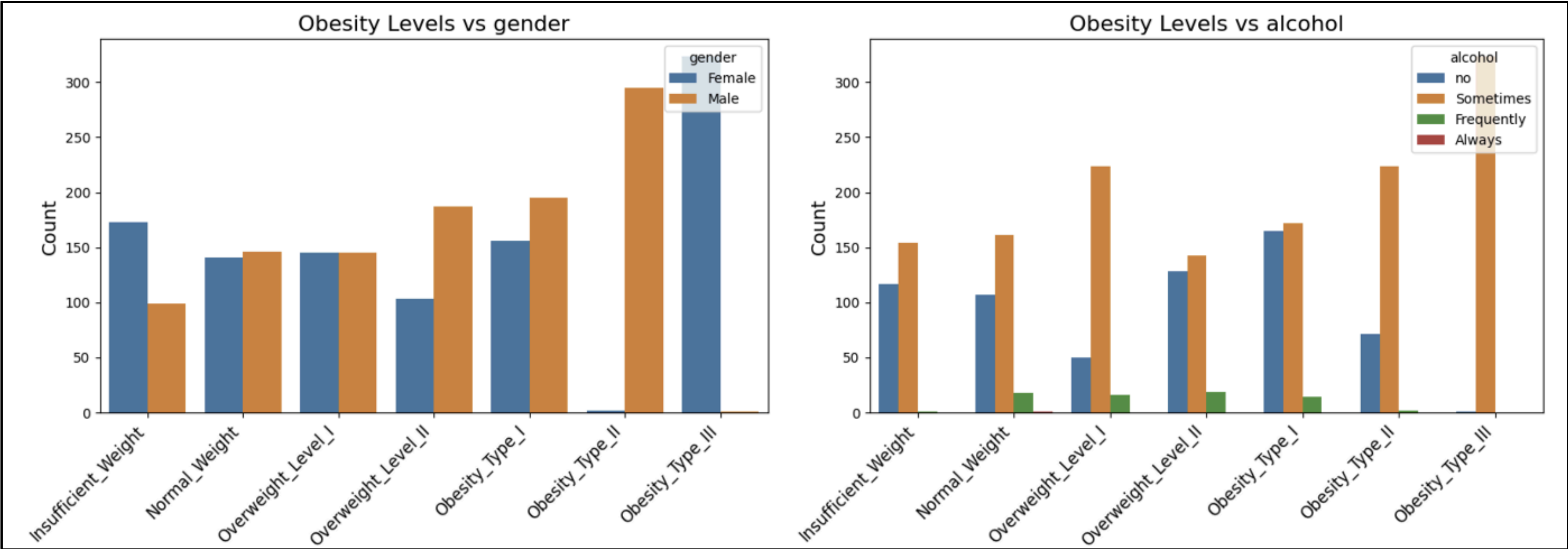
BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA TUẦN SUẤT TẬP THỂ DỤC MỖI TUẦN VÀ CÁC MỨC ĐỘ BÉO PHÌ.

# SỰ TƯƠNG QUAN GIỮA CÁC BIẾN ĐỘC LẬP VỚI BIẾN PHỤ THUỘC



BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA TẦN SUẤT SỬ DỤNG THIẾT BỊ ĐIỆN TỬ VÀ CÁC MỨC ĐỘ BÉO PHÌ.

# SỰ TƯƠNG QUAN GIỮA CÁC BIẾN ĐỘC LẬP VỚI BIẾN PHỤ THUỘC

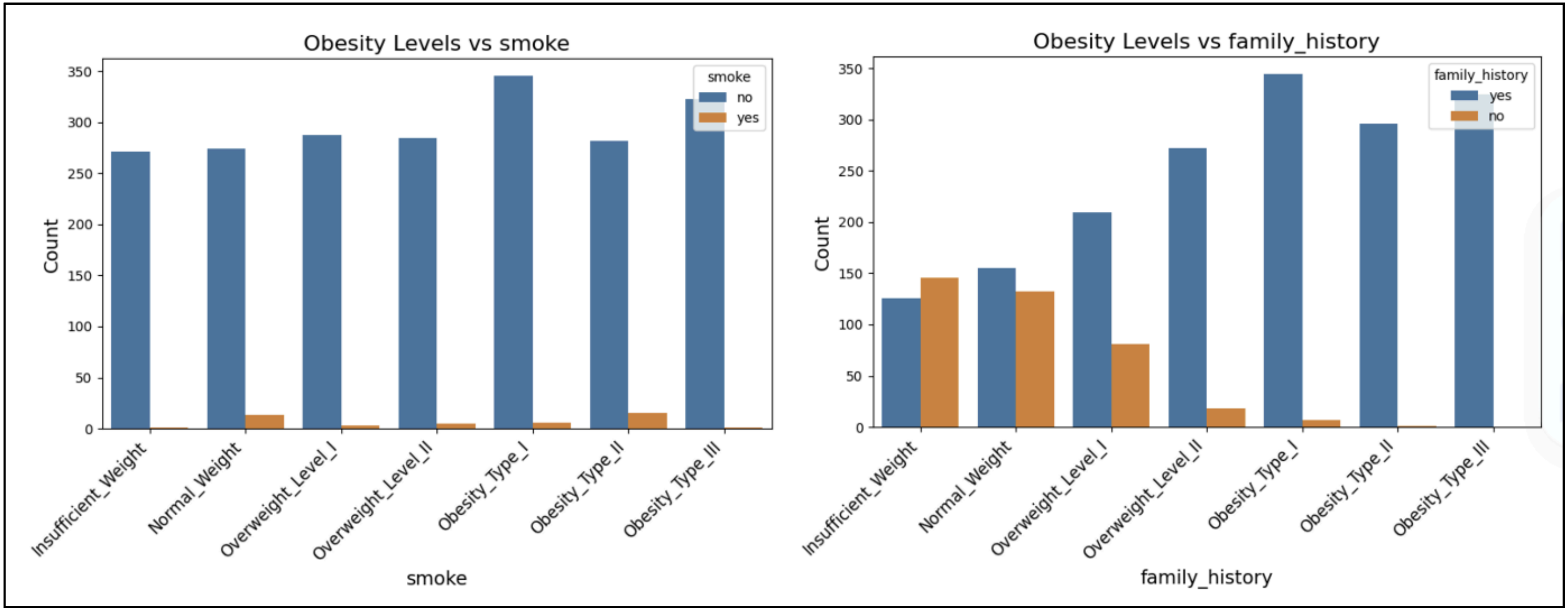


BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA GIỚI TÍNH VÀ CÁC MỨC ĐỘ BÉO PHÌ.

BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA TẦN SUẤT SỬ DỤNG ĐỒ UỐNG CÓ CỒN VÀ CÁC MỨC ĐỘ BÉO PHÌ.



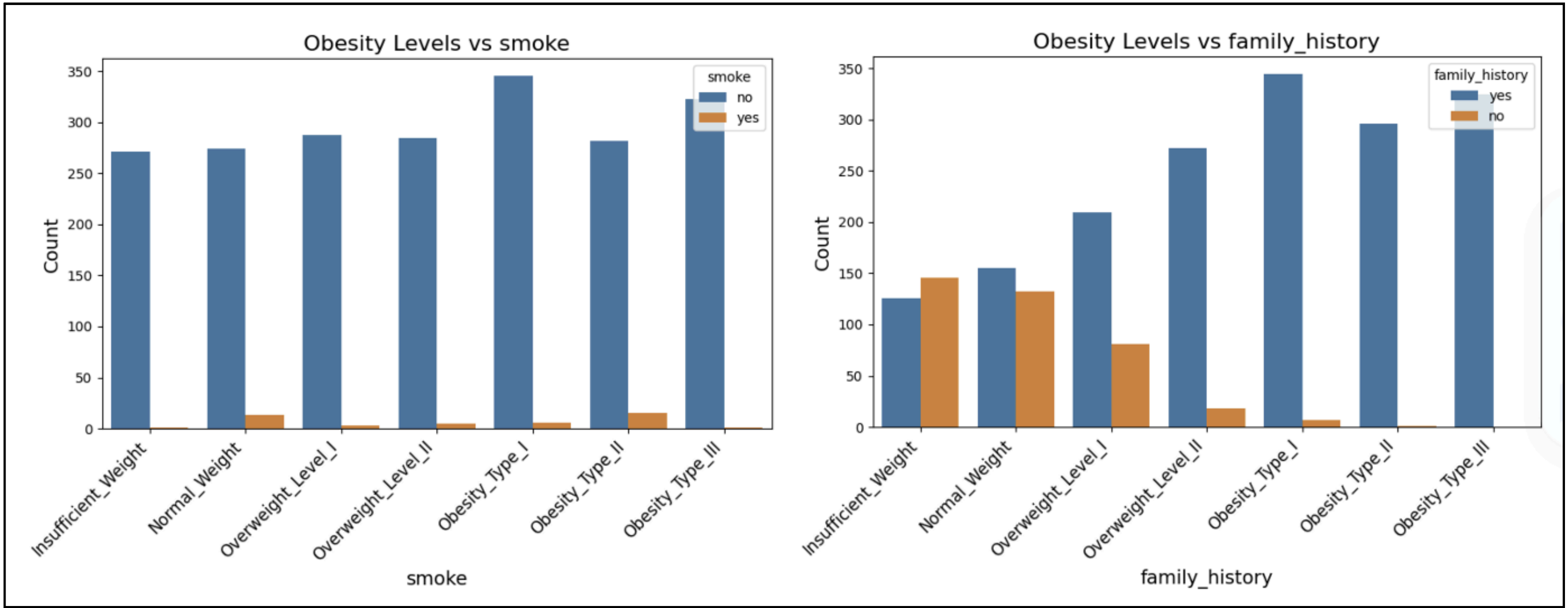
# SỰ TƯƠNG QUAN GIỮA CÁC BIẾN ĐỘC LẬP VỚI BIẾN PHỤ THUỘC



**BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ VIỆC HÚT THUỐC LÁ VÀ CÁC MỨC ĐỘ BÉO PHÌ.**

**BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA TIỀN SỬ BỆNH LÝ CỦA GIA ĐÌNH VÀ CÁC MỨC ĐỘ BÉO PHÌ.**

# SỰ TƯƠNG QUAN GIỮA CÁC BIẾN ĐỘC LẬP VỚI BIẾN PHỤ THUỘC

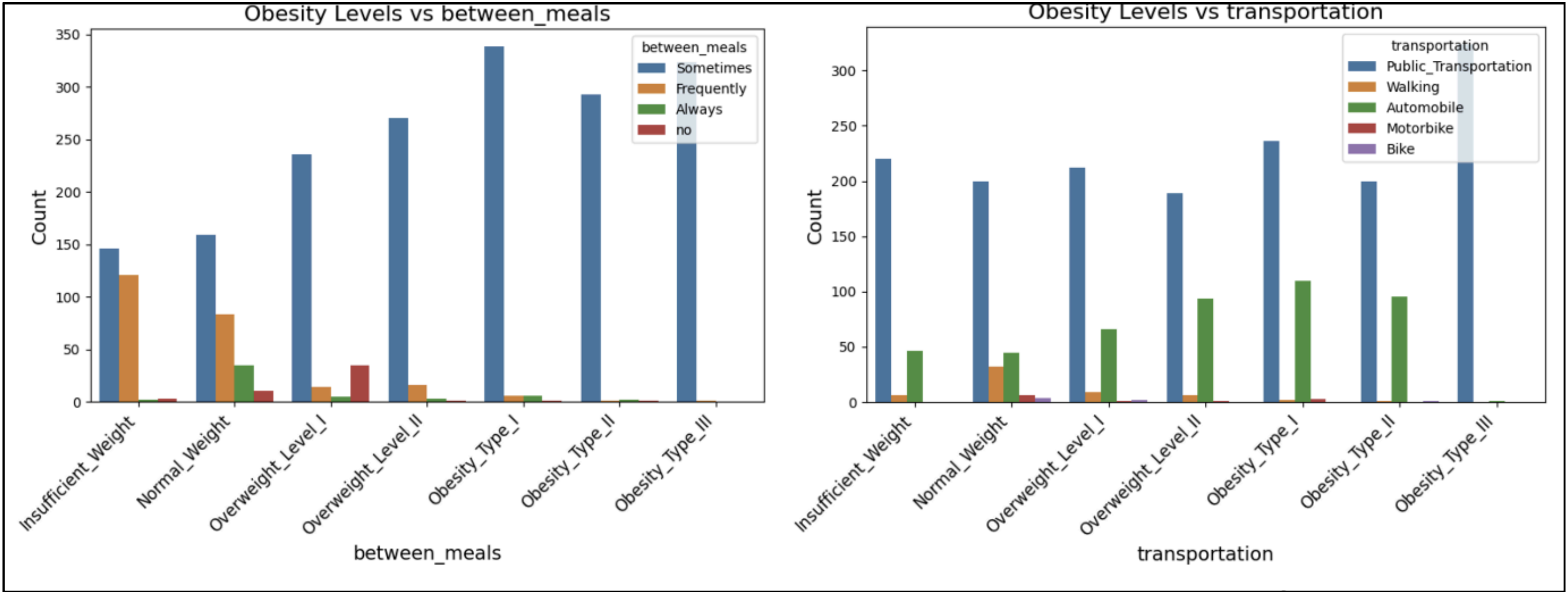


BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA VIỆC HÚT THUỐC LÁ VÀ CÁC MỨC ĐỘ BÉO PHÌ.

BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA TIỀN SỬ BỆNH LÝ CỦA GIA ĐÌNH VÀ CÁC MỨC ĐỘ BÉO PHÌ.



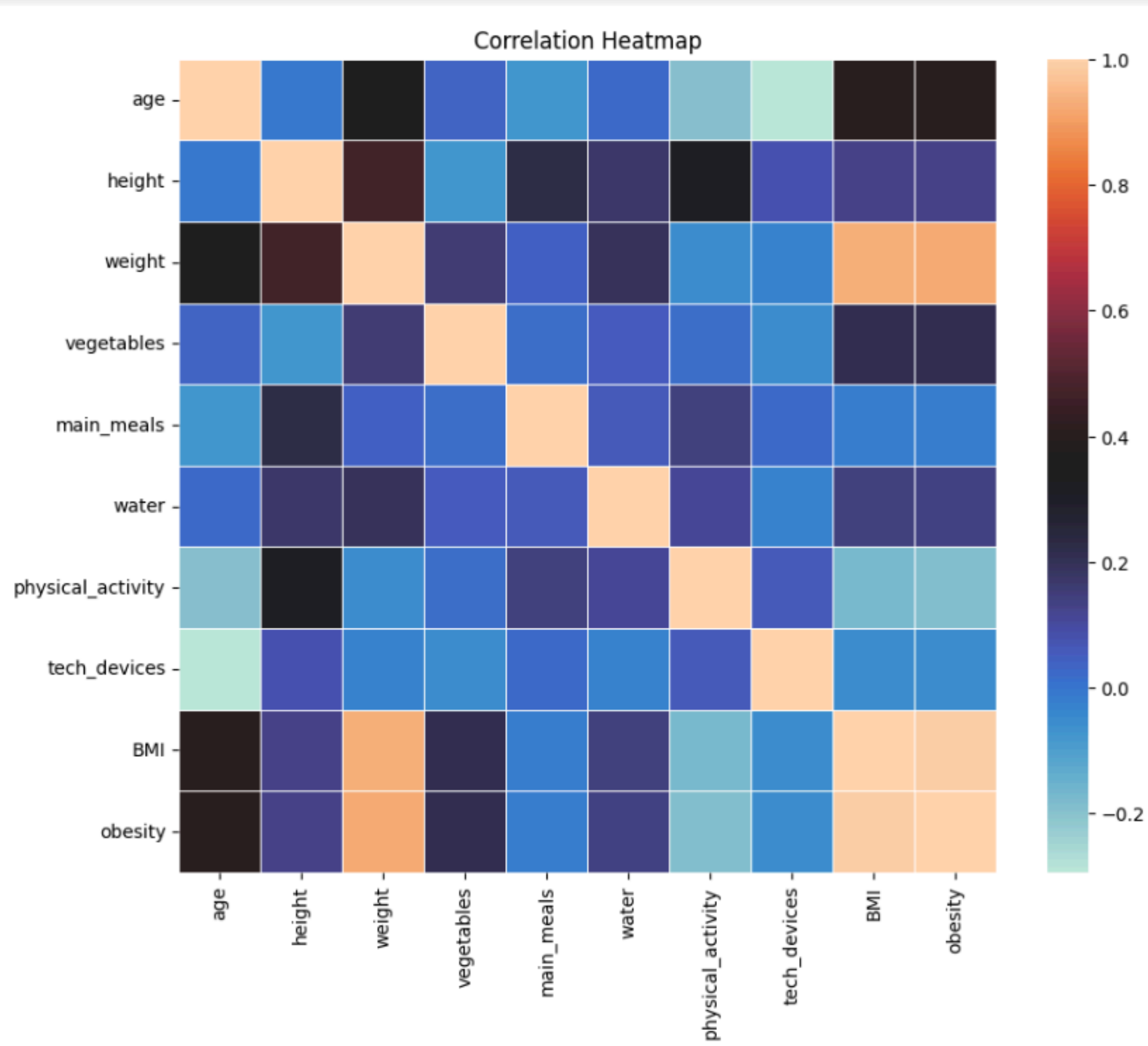
# SỰ TƯƠNG QUAN GIỮA CÁC BIẾN ĐỘC LẬP VỚI BIẾN PHỤ THUỘC



BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA TẦN SUẤT ĂN VẬT VÀ CÁC MỨC ĐỘ BÉO PHÌ.

BIỂU ĐỒ THỂ HIỆN MỐI QUAN HỆ GIỮA PHƯƠNG TIỆN ĐI LẠI VÀ CÁC MỨC ĐỘ BÉO PHÌ.

# KIỂM ĐỊNH



Các biến có hệ số tương quan trên  0.7	
weight và BMI	0.93
weight và obesity	0.92
BMI và obesity	0.99



# *KIỂM ĐỊNH BIÊN ĐỊNH LƯỢNG*

**Các biến định lượng được kiểm định:** 'age', 'height', 'main meals', 'water', 'physical activity', 'tech devices'

## **Giả thuyết:**

- **H0** :  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ : Trung bình của tất cả các nhóm là như nhau (với k là số nhóm)
- **H1** : Tồn tại ít nhất một nhóm i sao cho  $\mu_i \neq \mu$

## **Ví dụ:**

**Kiểm định: Độ tuổi (Age) không ảnh hưởng đến Obesity**

- **H0** : Trung bình độ tuổi là như nhau giữa các mức độ béo phì.
- **H1** : Có sự khác biệt về trung bình độ tuổi giữa các mức độ béo phì với nhau.



# KIỂM ĐỊNH BIÊN ĐỊNH LƯỢNG

```
a_Obesity_Type_I = df[df['obesity'] == 5]['age']  
a_Obesity_Type_II = df[df['obesity'] == 6]['age']  
a_Obesity_Type_III = df[df['obesity'] == 7]['age']  
a_Normal_Weight = df[df['obesity'] == 2]['age']  
a_Overweight_Level_I = df[df['obesity'] == 3]['age']  
a_Overweight_Level_II = df[df['obesity'] == 4]['age']  
a_Insufficient_Weight = df[df['obesity'] == 1]['age']
```

Mỗi nhóm chứa độ tuổi (age) của các cá nhân thuộc mức độ béo phì tương ứng:

- 1: Thiếu cân.
- 2: Cân nặng bình thường.
- 3, 4: Thừa cân cấp độ I và II.
- 5, 6, 7: Béo phì cấp độ I, II và III.

# KIỂM ĐỊNH BIÊN ĐỊNH LƯỢNG

```
f, p =  
stats.f_oneway(a_Obesity_Type_I, a_Obesity_Type_II, a_Obesity_Type_III,  
a_Normal_Weight, a_Overweight_Level_I, a_Overweight_Level_II,  
a_Insufficient_Weight )  
print(f"ANOVA Test: F-statistic = {f:.2f}, p-value = {p:.3f}")
```

Thực hiện kiểm định ANOVA -->  
Tính giá trị F và p-value

so sánh giá trị p với mức ý nghĩa = 0.05:

- Nếu  $p < 0.05$ : Bác bỏ  $H_0$
- Nếu  $p \geq 0.05$ : Không bác bỏ  $H_0$

```
alpha = 0.05
```

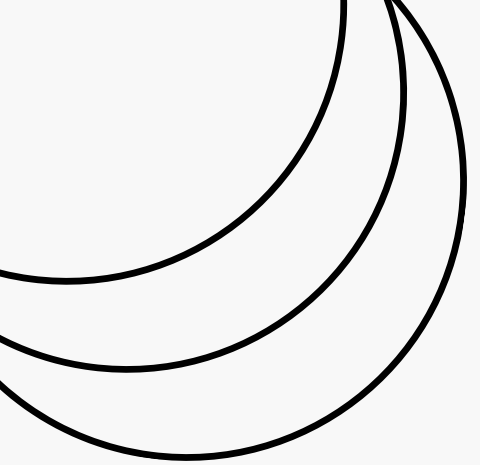
Kết luận theo phương pháp p-value (trị số p)

```
(p < alpha):
```

```
print(f'Trị số p = {p:.4f} < {alpha:.4f} cho nên bác bỏ  $H_0$  ==> Có sự  
khác biệt giữa trung bình độ tuổi và các mức độ béo phì với nhau')
```

```
else:
```

```
print(f'Trị số p = {p:.4f} >= {alpha:.4f} cho nên không có sự khác  
biệt trung bình độ tuổi giữa các mức độ béo phì với nhau')
```



# *KIỂM ĐỊNH BIẾN PHÂN LOẠI*

**Các biến phân loại được kiểm định – Ví dụ:** 'alcohol', 'high\_caloric', 'family\_history', 'between\_meals'

**Ví dụ – biến alcohol:**

**Kiểm định: Tần suất sử dụng đồ uống có chứa cồn không có ảnh hưởng tới mức độ béo phì**

- **H0:** Tần suất sử dụng đồ uống có chứa cồn và mức độ béo phì độc lập.
- **H1:** Tần suất sử dụng đồ uống có chứa cồn có tác động đến mức độ béo phì.

# KIỂM ĐỊNH BIÊN PHÂN LOẠI

**Bước 1:** Tạo bảng tần suất (Contingency Table)

Contingency Table:

	alcohol	Always	Frequently	Sometimes	no
obesity					
1		0	1	154	117
2		1	18	161	107
3		0	16	224	50
4		0	19	143	128
5		0	14	172	165
6		0	2	224	71
7		0	0	323	1

# KIỂM ĐỊNH BIÊN PHÂN LOẠI

**Bước 2:** Kiểm định Chi-bình phương (Chi-square) cho biến alcohol

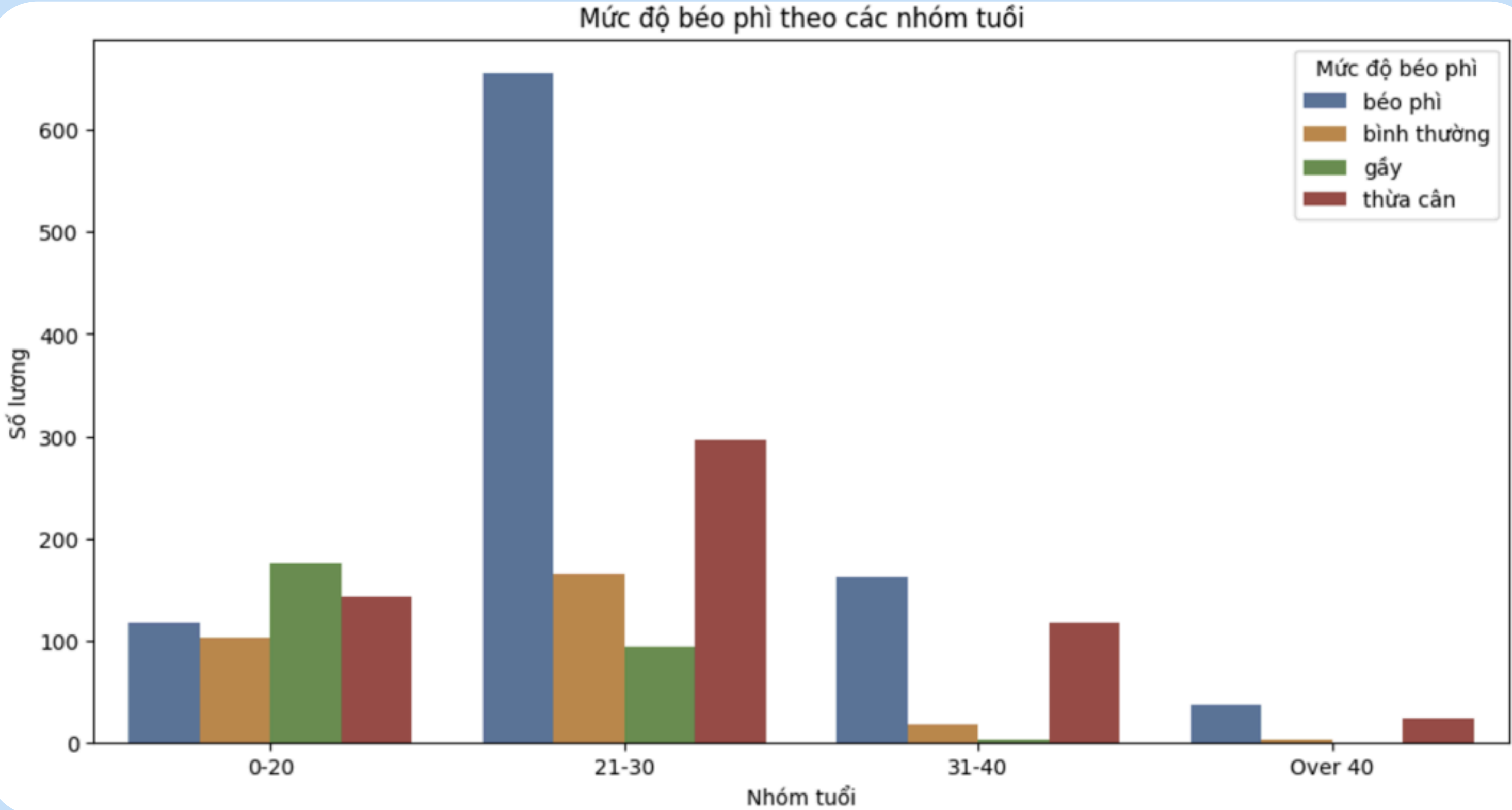
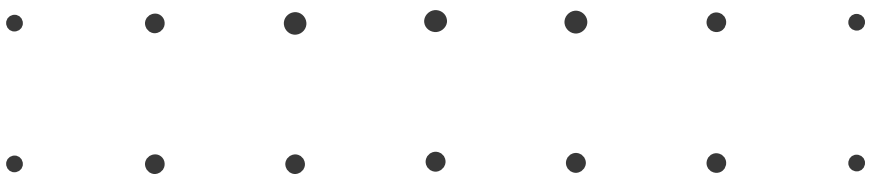
```
## Kiểm định Chi-square  
stat, p, ddof, expected = stats.chi2_contingency(contingency_table_2)
```

**Bước 3:** So sánh p-value với mức ý nghĩa  $\alpha = 0.05$

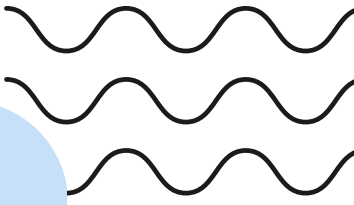
```
if (p < alpha):  
    print(f'Trị số p = {p} < {alpha:.4f} nên bác bỏ H0.',  
          '\n=> Tần suất sử dụng đồ uống có chứa cồn có ảnh hưởng đến trạng thái béo phì.')  
else:  
    print(f'Trị số p = {p} >= {alpha:.4f} nên chấp nhận H0.',  
          '\n=> Tần suất sử dụng đồ uống có chứa cồn không ảnh hưởng đến trạng thái béo phì')
```



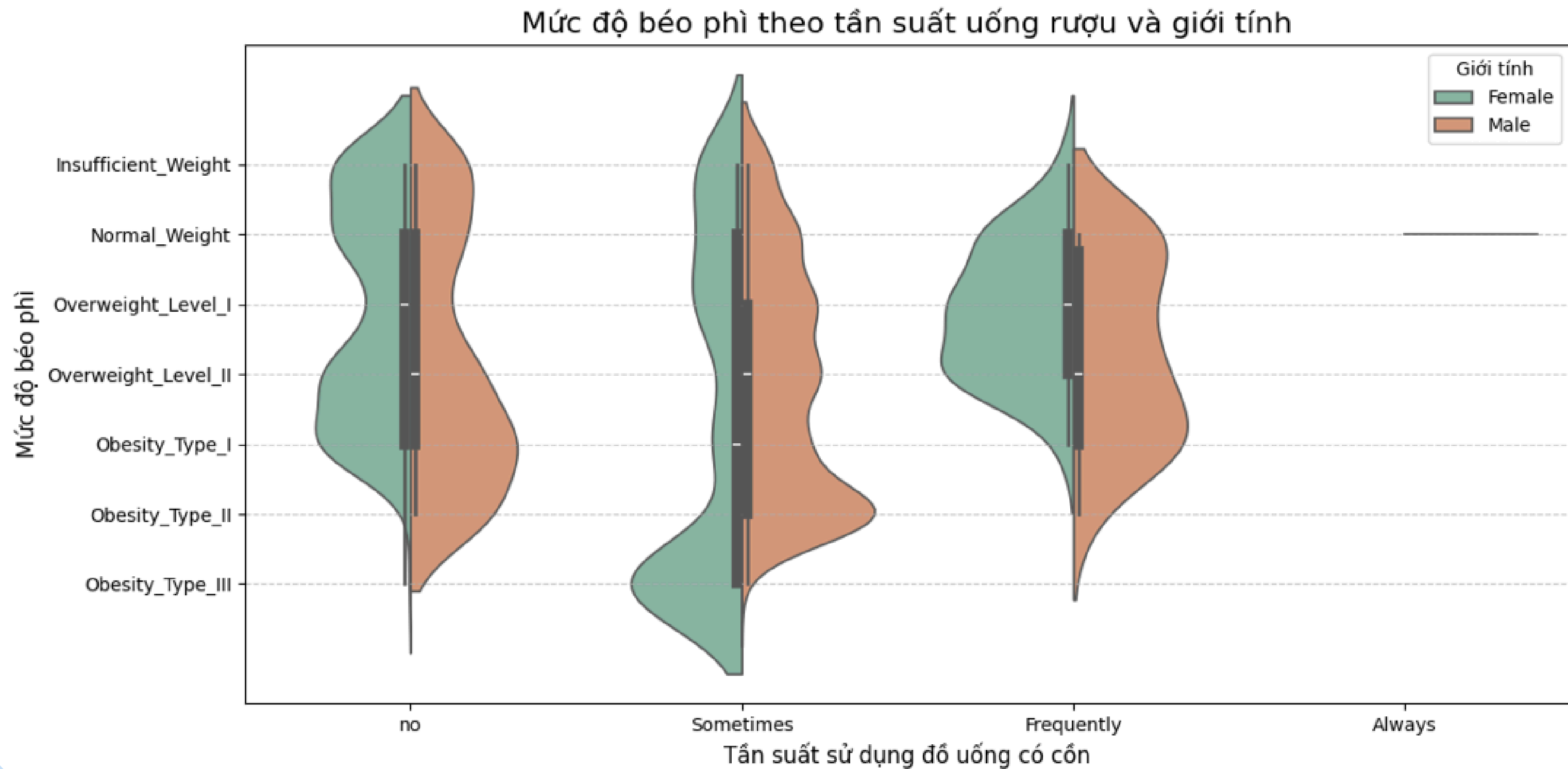
# BIỂU DIỄN TRỰC QUAN:



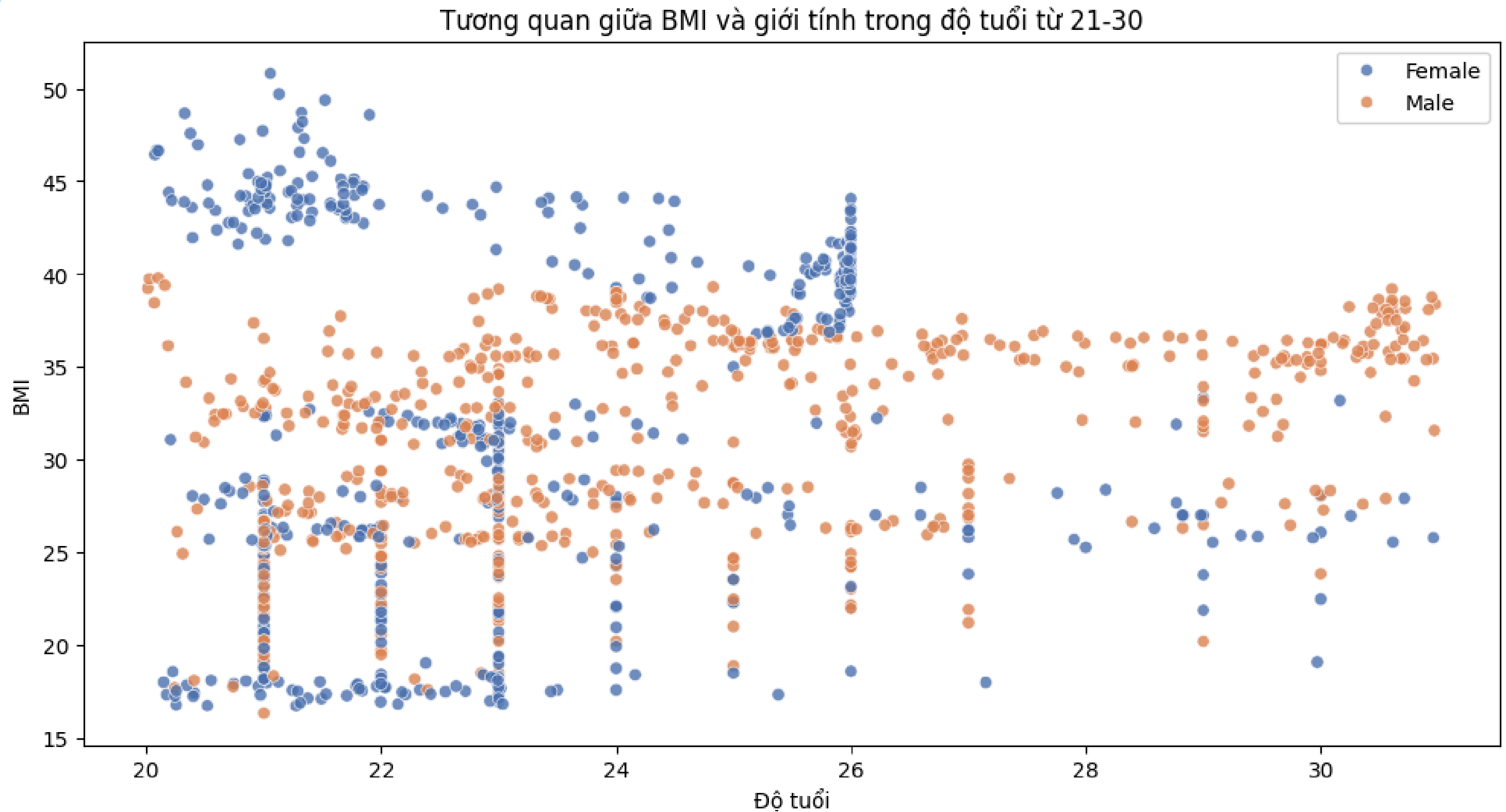
A rectangular array of dots arranged in 2 rows and 7 columns. The top row contains 7 dots, and the bottom row contains 7 dots, for a total of 14 dots.

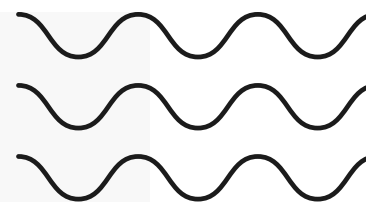
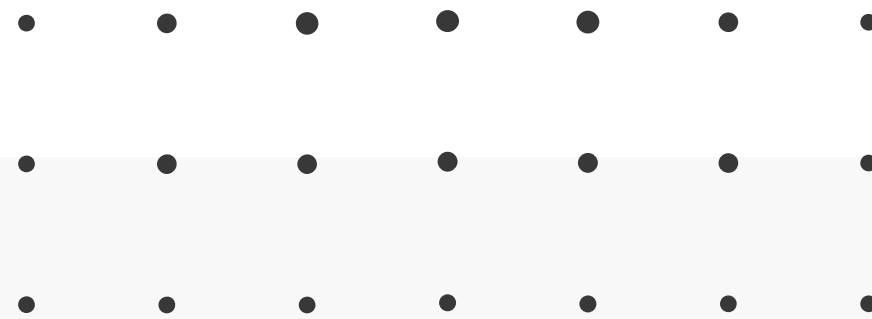
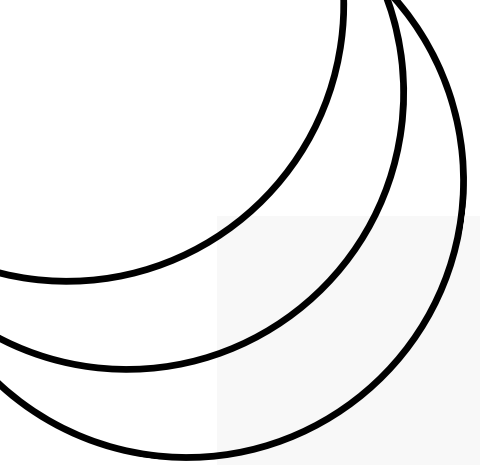


# BIỂU DIỄN TRỰC QUAN:



# BIỂU DIỄN TRỰC QUAN:





*Thanks You*

