

BÀI TẬP THỰC HÀNH

THỐNG KÊ MÔ TẢ

1. Đọc file dữ liệu

- Sử dụng file dữ liệu **igfdata**

Ngoài các chỉ số liên quan đến giới tính, độ tuổi, trọng lượng và chiều cao, chúng tôi đo lường các hormone liên quan đến tình trạng tăng trưởng như **igfi**, **igfbp3**, **als**, và các markers liên quan đến sự chuyển hóa của xương **pinp**, **ictp** và **p3np**

- Đọc vào một dataframe

2. Tính toán và hiển thị một số chỉ số thống kê mô tả

- Giá trị trung bình
- Phương sai
- Độ lệch chuẩn
- Giá trị nhỏ nhất, cao nhất, miền giá trị
- Hiển thị toàn bộ thông tin chung về một biến nào đó trong dữ liệu
- Viết hàm tính sai số chuẩn (trong R không cung cấp hàm tính sai số chuẩn)

$$SE = \frac{SD}{\sqrt{|x|}} = \frac{\sigma^2}{N}$$

- Tóm tắt dữ liệu **igfdata** theo **sex**
- Để xem phân phối của các hormones, chỉ số sinh hóa trong bộ dữ liệu này. Hãy vẽ đồ thị cho 6 biến số: **igfi**, **igfbp3**, **als**, **pinp**, **ictp**, **p3np**

3. Kiểm định t (t.test)

3.1. Kiểm định t một mẫu

- Hiển thị tuổi thọ trung bình trong mẫu trên (19.17)
- Kiểm tra giá trị trung bình vừa tính được ở bước trên có thật sự khác với giá trị trung bình 30 của nghiên cứu trước khi không?

3.2. Kiểm định t 2 mẫu

- Hiển thị giá trị trung bình hormones **igfi** của nữ và nam (167.97 và 160.29)
- Kiểm tra hai giá trị trung bình vừa tính được là khác biệt thực sự, có ý nghĩa thống kê hay chỉ là yếu tố ngẫu nhiên?

3.3. Kiểm định biến có phân phối chuẩn?

- Kiểm tra biến **igfi** có phân phối chuẩn không?

3.4. Kiểm định t cho các biến số theo cặp (paired t-test)

Ví dụ: Một nhóm bệnh nhân gồm 10 người được điều trị bằng một thuốc nhằm giảm huyết áp. Huyết áp của bệnh nhân được đo lúc khởi đầu nghiên cứu (lúc chưa điều trị), và sau khi điều trị. Số liệu huyết áp của 10 bệnh nhân như sau:

Trước khi điều trị (x0)	180, 140, 160, 160, 220, 185, 145, 160, 160, 170
Sau khi điều trị (x1)	170, 145, 145, 125, 205, 185, 150, 150, 145, 155

Câu hỏi đặt ra là độ biến chuyển huyết áp trên có đủ để kết luận rằng thuốc điều trị có hiệu quả giảm áp huyết. Để trả lời câu hỏi này, chúng ta dùng kiểm định t cho từng cặp như sau:

```
> # nhập dữ kiện
> before <- c(180, 140, 160, 160, 220, 185, 145, 160, 160, 170)
> after <- c(170, 145, 145, 125, 205, 185, 150, 150, 145, 155)
> bp <- data.frame(before, after)
> # kiểm định t
> t.test(before, after, paired=TRUE)
```

Kết quả

```
Paired t-test

data:  before and after
t = 2.7924, df = 9, p-value = 0.02097
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.993901 19.006099
sample estimates:
mean of the differences
                10.5
```

3.5. Kiểm định Wilcoxon cho các biến số theo cặp (*wilcox.test*)

Thay vì dùng kiểm định t cho từng cặp, chúng ta cũng có thể sử dụng hàm *wilcox.test* cho cùng mục đích:

```
> wilcox.test(before, after, paired=TRUE)
```

Kết quả

```
Wilcoxon signed rank test with continuity correction

data:  before and after
V = 42, p-value = 0.02291
alternative hypothesis: true mu is not equal to 0
```