**Bài 1: Với mục đích so sánh trọng lượng của hai loại nước Coca-Cola thường và Coca-Cola ăn kiêng, người ta lấy ngẫu nhiên 8 lon nước mỗi loại, sau đó đo trọng lượng của từng lon. Khối lượng của 16 lon nước này được liệt kê trong bảng sau:**

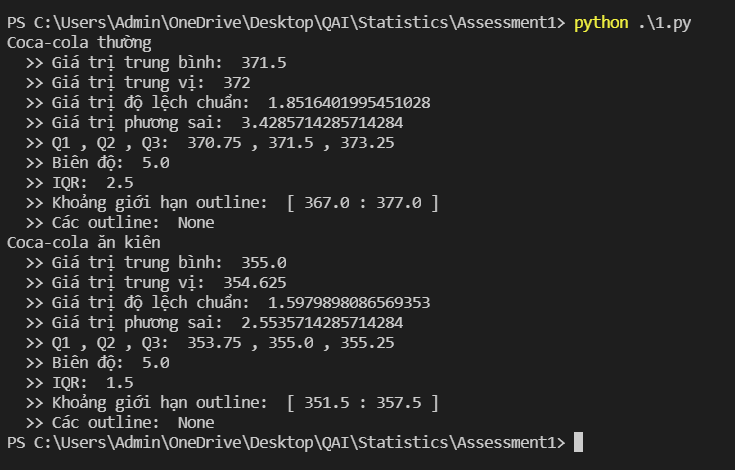
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Thông thường** | **371** | **370** | **370** | **373** | **374** | **372** | **375** | **371** |
| **Ăn kiêng** | **353** | **355** | **352** | **354** | **355** | **356** | **355** | **357** |

1. **Tìm trọng lượng trung bình của Coca-cola thông thường và Coca-cola ăn kiêng, sau đó so sánh kết quả. Trọng lượng trung bình  của hai loại nước này có bằng nhau không?**
2. **Tìm các tứ phân của hai mẫu dữ liệu, sau đó so sánh kết quả thu được, đồng thời chỉ ra các outlier (nếu có) và xây dựng boxplot mô tả dữ liệu.**
3. **Tìm phương sai và độ lệch chuẩn của hai mẫu dữ liệu, sau đó so sánh kết quả.**
4. **Dựa vào các kết quả trên, bạn có thể kết luận rằng trọng lượng của Coca-cola thông thường và Coca-cola ăn kiêng có giống nhau hay không?**

**Trả lời:**

**Cách 1: Dùng code python (có đính kèm theo)**

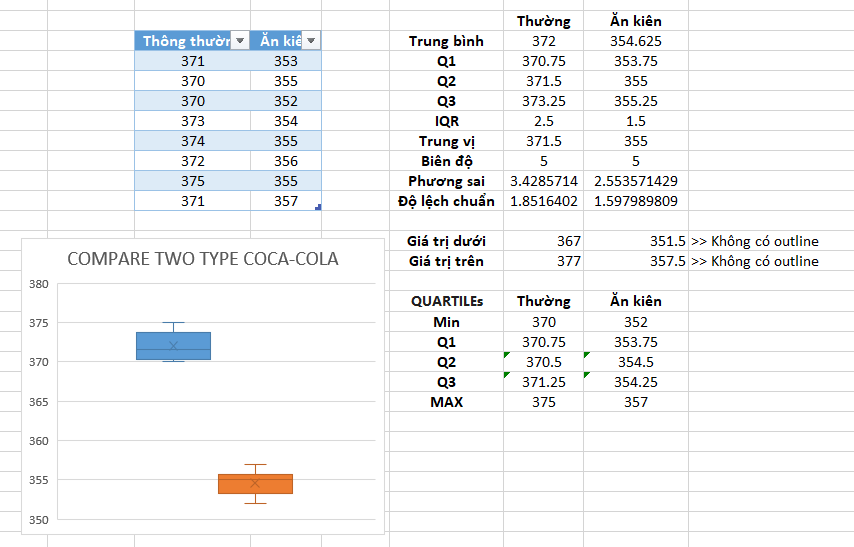
Kết quả thu được:



Nhận xét:

* Trọng lượng trung bình của hai loại coca-cola là không giống nhau
* Từ hình boxplot ta thấy được trọng lượng của hai loại này phân bố khác nhau hoàn toàn.
* Không có outline nên độ tin cây của dữ liệu khá tốt.
* Kết luận luôn là trọng lượng của hai loại này là khác nhau

**Cách 2: Dùng Excel (có file đính kèm theo)**



Nhận xét:

* Trọng lượng trung bình của hai loại coca-cola là không giống nhau.
* Từ hình boxplot ta thấy được trọng lượng của hai loại này phân bố khác nhau hoàn toàn.
* Không có outline nên độ tin cây của dữ liệu khá tốt.
* Kết luận luôn là trọng lượng của hai loại này là khác nhau.

**Bài 2. Bạn muốn rút tiền mặt từ cây ATM, nhưng do trời tối nên bạn không thể nhìn thấy thẻ khi chèn thẻ vào cây. Thẻ được đưa vào với mặt trước (có in tên bạn) ở phía trên và được chèn vào cây ATM sao cho từ đầu tiên trong tên của bạn được chèn vào trước là đúng quy định.**

1. **Giả sử rằng bạn chèn thẻ một cách ngẫu nhiên, tính xác suất bạn chèn thẻ đúng theo quy định?**
2. **Giả sử rằng bạn chèn thẻ một cách ngẫu nhiên, tính xác suất thẻ được chèn không đúng quy định vào lần thử đầu tiên, nhưng được chèn đúng quy định vào lần thử thứ hai?**
3. **Bạn sẽ cần bao nhiêu lần thử ngẫu nhiên để chắc chắn rằng thẻ hoạt động (thẻ được chèn đúng quy định)?**

**Trả lời:**

1. Không gian mẫu là : trường hợp gồm : ngửa-thuận, ngửa-ngược, sấp-thuận, sấp-ngược.

Xác suất để bạn chèn thẻ đúng ứng với ngửa-thuận là : 

2. Xác suất thẻ được chèn không đúng quy định vào lần thử đầu tiên: 

Xác suất thẻ được chèn lần thứ 2 đúng và biết lần đầu bị sai là : 

Áp dụng nhân xác suất ta được kết quả là: 

3. Xác suất lần đầu chèn thì xác suất đúng là : 

Để lần 2 đúng thì biết lần 1 sai, do đó xác suất đúng là: 

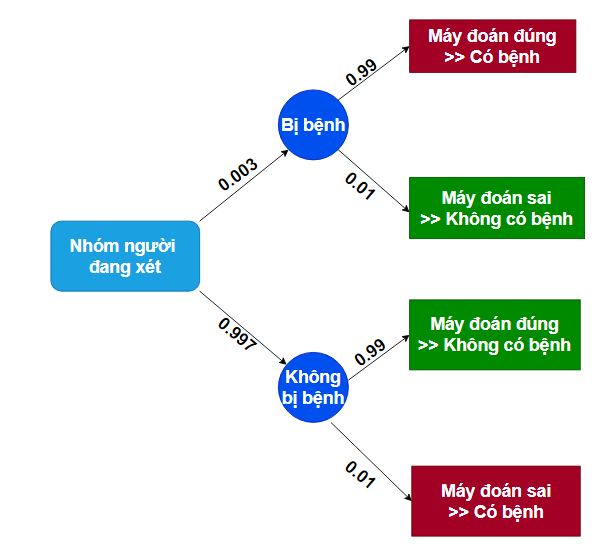
Để lần 3 đúng thì biết 2 lần trước sai, do đó xác suất đúng là: 

Để lần 4 đúng thì biết 3 lần trước sai, do đó xác suất đúng là: 

Vậy cần tối thiểu 4 lần chèn thì biết chắc chắn thẻ hoạt động hay không.

**BÀI 3: Xét nghiệm một căn bệnh hiếm gặp luôn mang lại kết quả chính xác đến 99%. Nếu một người bị mắc bệnh, kết quả xét nghiệm sẽ dương tính với xác suất 0.99 và nếu không mắc bệnh, kết quả xét nghiệm sẽ âm tính với xác suất 0.99. Chọn một người ngẫu nhiên trong một nhóm người nhất định có xác suất bị bệnh là 0.003. Giả sử rằng người đó vừa được xét nghiệm dương tính, vậy xác suất mắc bệnh của người đó là bao nhiêu?**

**Trả lời:**



Từ bài toán ta có sơ đồ cây như trên hình. Như vậy có hai khả năng để máy đoán bị bệnh là: Máy đoán sai khi người đó không bị bệnh và Máy đoán đúng khi người đó bị bệnh.



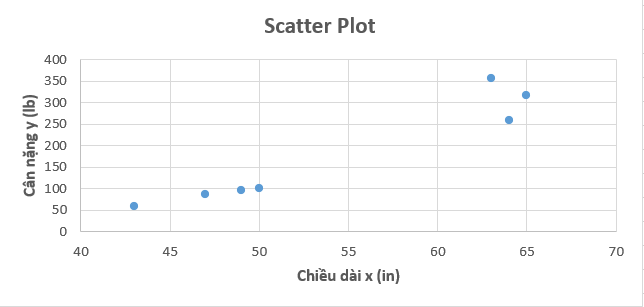
**Bài 4: Dưới đây là chiều dài và cân nặng cơ thể  của một số con gấu được chọn ngẫu nhiên:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Chiều dài x (in)** | **43** | **65** | **63** | **50** | **49** | **64** | **47** |
| **Cân nặng y (lb)** | **60** | **316** | **356** | **100** | **96** | **259** | **86** |

1. **Xây dựng scatter plot mô  tả dữ liệu mẫu. Scatter plot đưa ra gợi ý gì về tương quan tuyến tính giữa chiều dài và cân nặng của các con gấu?**
2. **Tìm hệ số tương quan giữa x và y. Tương quan này có phải là tương quan dương hay không?**
3. **Bạn hãy xây dựng phương trình đường hồi quy tuyến tính mô tả mối quan hệ phụ thuộc của y vào x.**
4. **Sử dung phương trình đường hồi quy, bạn hãy dự đoán cân nặng cho một con gấu có chiều dài 72 in.**

**Trả lời:**

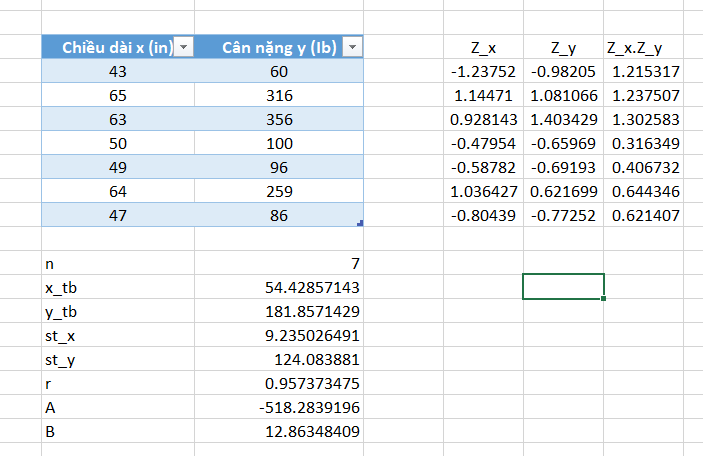
1. Hai kiểu dữ liệu đều là dạng định lượng nên sẽ sử dụng scatter plot để mô ta dữ liệu mẫu.



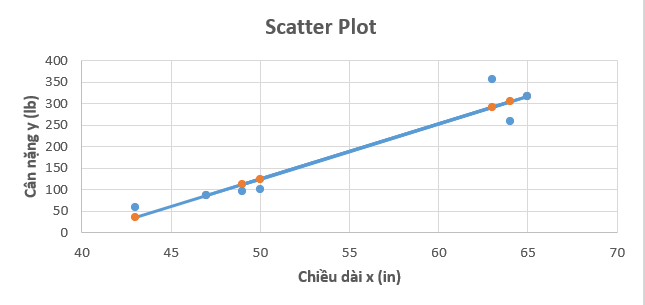
Tính tương quan giữa chiều dài x và cân nặng y có thể là tương quan tuyến tính, chúng ta cần kiểm tra thử tương quan tuyến tính này có mạnh hay không ?

2. Tìm hệ số tương quan tuyến tính giữa x và y.

Hệ số tương quan gần với 1 chứng tỏ quan hệ này là tương quan dương và rất chặt.



3. Phương trình hồi quy tuyến tính là : 



4. Dự đoán tại chiều dài 72 in là: 