Cho Bảng dữ liệu đính kèm (Bảng 1)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Chăm sóc (Chamscoc)** | **Tuổi (Tuoi)** | **Chủng tộc (Chungtoc)** | **Giới tính (Gioitinh)** | **Chỉ số cơ thể (MBI)** | **Tỷ lệ Creatinine** (SC) | **Huyết áp ngưỡng trên (SBP)** |
| 4995 | 0 | 55 | 1 | 1 | 19.435 | 1.6 | 150 |
| 2312 | 0 | 78 | 2 | 1 | 22.503 | 2.682 | 104 |
| 896 | 0 | 50 | 1 | 1 | 27.406 | 1.3 | 140 |
| 3103 | 0 | 60 | 1 | 1 | 29.867 | 1.091 | 140 |
| 538 | 1 | 31 | 1 | 1 | 27.025 | 1.159 | 120 |
| 1426 | 0 | 70 | 1 | 1 | 19.04 | 1.25 | 150 |
| 4787 | 1 | 46 | 1 | 1 | 28.662 | 1.307 | 140 |
| 5663 | 0 | 59 | 2 | 1 | 27.406 | 1.705 | 152 |
| 1109 | 0 | 68 | 1 | 2 | 27.532 | 1.534 | 144 |
| 666 | 0 | 65 | 1 | 1 | 28.058 | 2 | 120 |
| 2705 | 1 | 66 | 1 | 2 | 28.762 | 0.9 | 150 |
| 5668 | 0 | 74 | 1 | 1 | 29.024 | 1.227 | 116 |
| 999 | 1 | 47 | 1 | 2 | 30.506 | 1.386 | 120 |
| 1653 | 1 | 63 | 1 | 1 | 28.399 | 1.1 | 105 |
| 764 | 1 | 63 | 2 | 2 | 28.731 | 0.9 | 122 |
| 3640 | 0 | 79 | 1 | 1 | 18.957 | 2.239 | 150 |
| 1254 | 1 | 73 | 1 | 1 | 26.545 | 1.3 | 144 |
| 2217 | 1 | 65 | 1 | 1 | 23.739 | 1.614 | 170 |
| 4326 | 0 | 65 | 1 | 1 | 29.34 | 1.2 | 170 |
| 5750 | 1 | 76 | 1 | 1 | 39.837 | 1.455 | 140 |
| 6396 | 0 | 83 | 1 | 1 | 26.156 | 1.489 | 116 |
| 2289 | 0 | 76 | 1 | 1 | 30.586 | 1.7 | 130 |
| 1322 | 1 | 45 | 1 | 2 | 43.269 | 0.9 | 115 |
| 4554 | 1 | 58 | 1 | 2 | 28.192 | 1.352 | 130 |
| 6719 | 1 | 34 | 1 | 1 | 20.426 | 1.886 | 116 |
| 1954 | 1 | 77 | 1 | 1 | 26.545 | 1.307 | 140 |
| 5001 | 1 | 70 | 1 | 1 | 19.044 | 1.2 | 110 |
| 1882 | 0 | 50 | 1 | 1 | 25.712 | 1.034 | 140 |
| 5368 | 1 | 38 | 1 | 1 | 30.853 | 0.9 | 134 |
| 787 | 0 | 58 | 2 | 2 | 27.369 | 0.909 | 100 |
| 4375 | 0 | 61 | 1 | 1 | 32.079 | 1.273 | 128 |
| 5753 | 1 | 75 | 1 | 1 | 37.59 | 1.3 | 138 |
| 6745 | 0 | 45 | 1 | 1 | 22.85 | 1.398 | 130 |
| 6646 | 0 | 61 | 1 | 1 | 27.718 | 1.659 | 128 |
| 5407 | 1 | 50 | 1 | 2 | 24.176 | 1 | 130 |
| 4181 | 0 | 44 | 2 | 2 | 26.37 | 1.148 | 124 |
| 3403 | 0 | 55 | 1 | 2 | 21.79 | 1.17 | 130 |
| 2439 | 1 | 49 | 1 | 1 | 15.204 | 1.307 | 140 |
| 4055 | 0 | 71 | 1 | 1 | 22.229 | 1.261 | 100 |
| 3641 | 0 | 64 | 1 | 1 | 21.228 | 0.9 | 130 |

1. Xử lý dữ liệu bao gồm.

* Hãy tạo biến mới để chia độ tuổi thành các lớp tuổi: 31-40, 41-50, 51-60, 61-70, trên 70 tuổi tương ứng với các giá trị 1, 2, 3, 4, 5.
* Tạo cột biến mới: Nhóm chăm sóc: 0 = Placebo, 1 = Digoxin;
* Chủng tộc: 1 = white, 0 = Nonwhite;
* Giới tính: 1 = male, 2 = female;
* Chỉ số cơ thể: Underweight (), Normal [18.5-25), Overweight [25-30), Obese ().

1. Lập bảng tần suất và ứng dụng.

* Lập bảng tần suất của độ tuổi chia lớp, vẽ biểu đồ Histogram của dữ liệu. Giả sử tổng thể có 1,500,000 người. Dựa vào mẫu đã cho hãy ước lượng số người trong cộng đồng trong mỗi nhóm tuổi.
* Lập bảng tần suất của chỉ số cơ thể, vẽ Histogram. Giả sử tổng thể cộng đồng có 1,500,000 người. Dựa và bảng phân phối tần suất hãy ước tính số người dưới ký, chuẩn, quá ký, và béo phì trong cộng đồng này.

1. Dữ liệu chéo và giá trị xác suất.

* Tìm xác suất chọn một người trong mẫu thỏa mãn điều kiện là người nam, độ tuổi từ 41 đến 60 bị quá ký.
* Tìm xác suất để người nam gia trắng có chỉ số cơ thể trên chuẩn.
* Tìm xác suất để một người nữ trên 40 tuổi bị béo phì.

1. Các giá trị trung tâm.

* Xác định giá trị trung bình, phương sai, độ lệch chuẩn của huyết áp ngưỡng trên và nêu ý nghĩa của các đại lượng này.
* Xác định khoảng huyết áp ngưỡng trên chứa 95% nhóm đối tượng đang xét.
* So sánh độ phân tán dữ liệu giữa SC và SBP.
* Xác định vị trí % của đối tượng có lượng SC là 1.15.
* Xác định giá trị tương ứng với vị trí % của SC là 80%