**Capstone Project 1**

**Thông tin chung:**

1. Họ và tên: Bùi Thị Thúy Hồng
2. Mã số dữ liệu: Data\_Hong

**KẾT QUẢ PHÂN TÍCH**

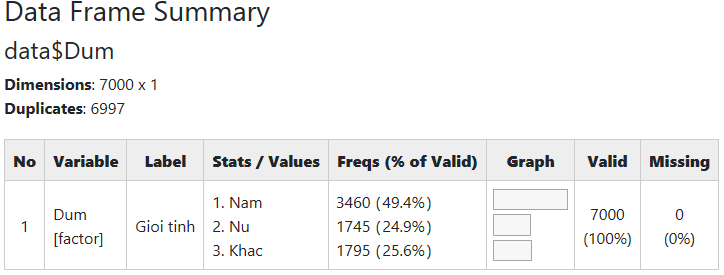
**Phần 1. Làm sạch dữ liệu**

1. Import số liệu
2. Xóa dữ liệu thiếu logic (Ghi rõ biến số và dòng): Biến Dum, dòng 56
3. Tìm dữ liệu bị missing (Ghi rõ biến số và dòng) và thực hiện thay thế dữ liệu bị missing (imputation): Biến Dum, dòng 99. Biến Team , dòng 113. Biến Y dòng 121
4. Gán các thuộc tính cho biến số như label, value label, measure (Không viết báo cáo tại đây):
   * Biến số định lượng (X1, X2, X3, X4, Y): Gán numeric
   * Biến số định tính (Có tên là Dum trong file dữ liệu): Gán label (Giới tính), value label (1: Nam, 2: Nữ, 3: Khác), chuyển thành factor (lưu ý: levels = c(0, 1))

**Phần 2. Thống kê mô tả**

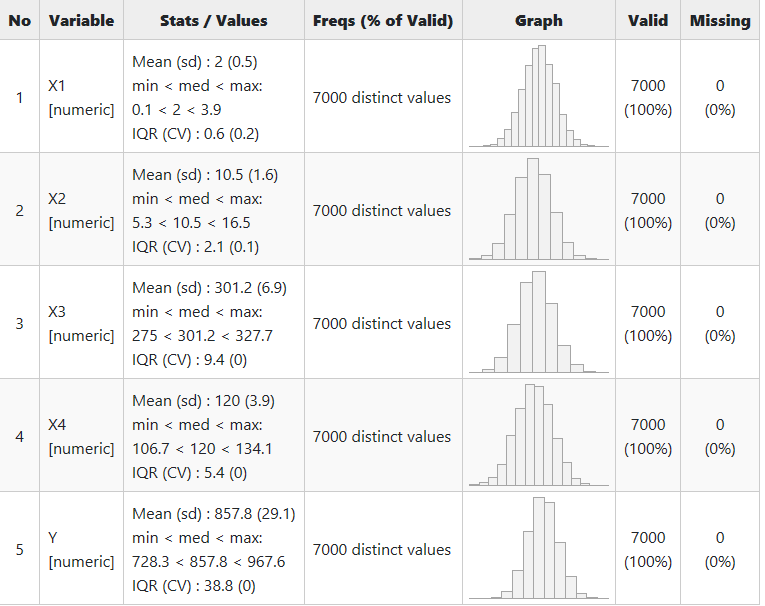
1. Thống kê mô tả biến số định tính:

**Bảng 1. Thống kê mô tả biến định tính**

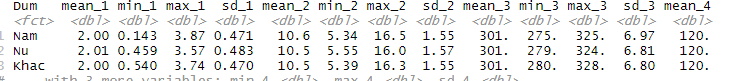


1. Thống kê mô tả biến số định lượng

**Bảng 2. Thống kê mô tả biến định lượng**

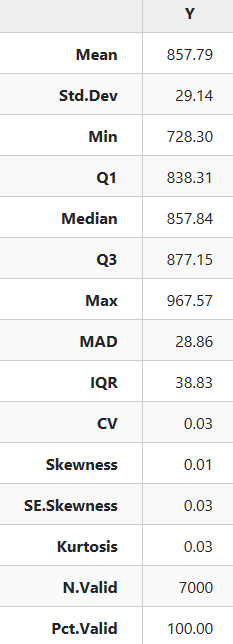
1. Thống kê bảng chéo biến định được mô tả phía trên theo tất cả các biến định lượng

**Bảng 3. Bảng thống kê chéo**



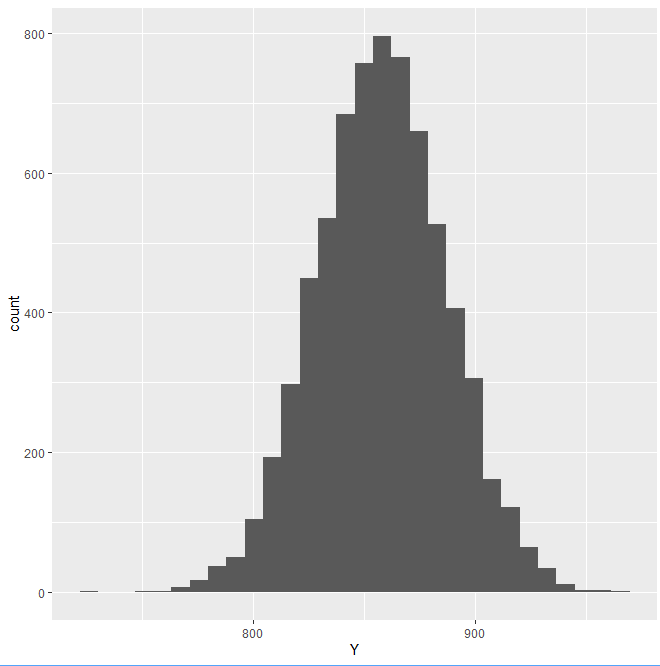
1. Thống kê tứ phân vị biến phụ thuộc:

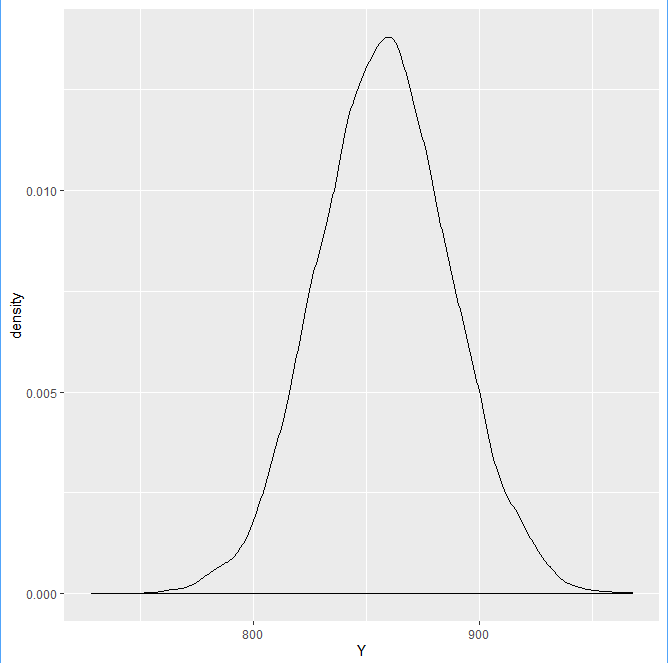
**Bảng 4. Bảng thống tứ phân vị**

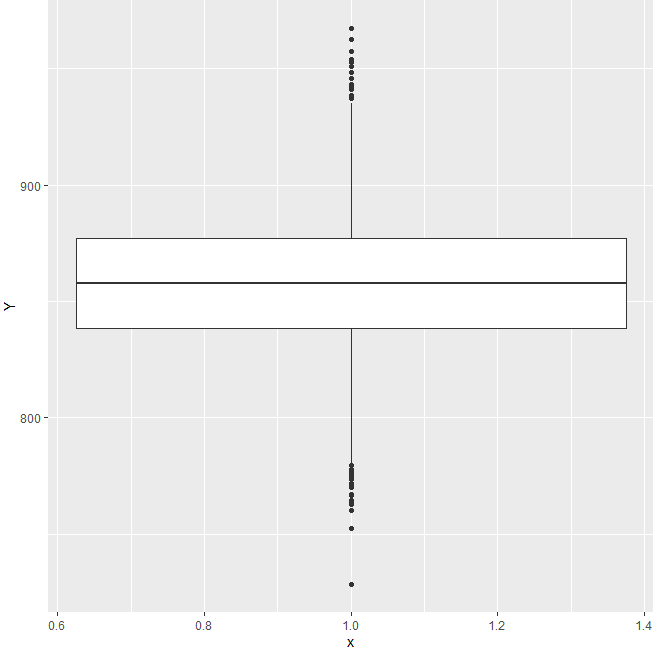


1. Vẽ đồ thị histogram, density và đồ thị box – plot biến phụ thuộc

**Hình 1. Đồ thị histogram biến phụ thuộc**

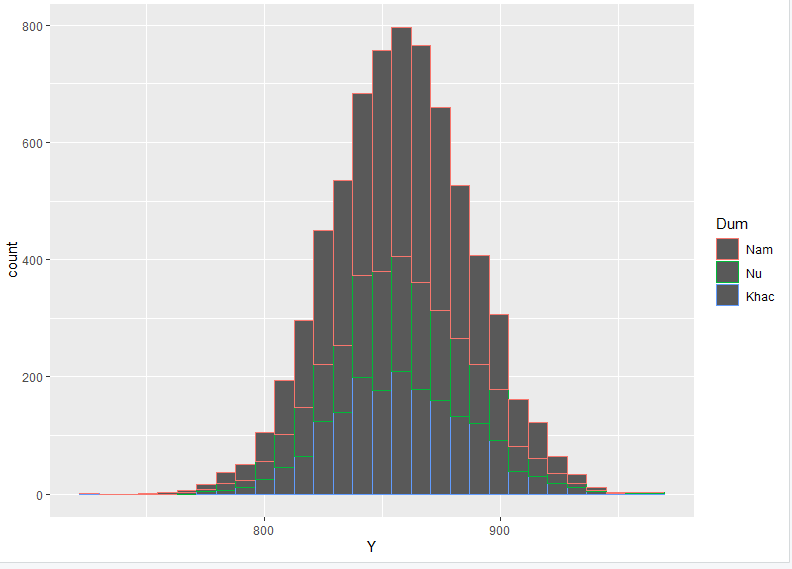
**Hình 2. Đồ thị density biến phụ thuộc**

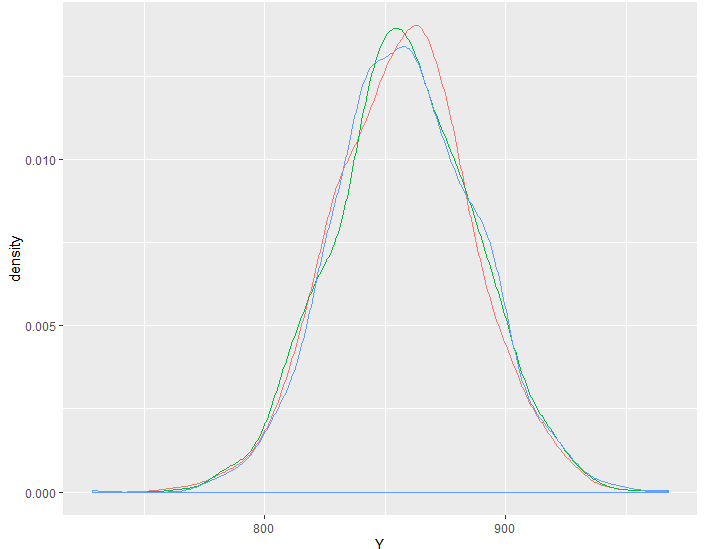
**Hình 3. Đồ thị box - plot biến phụ thuộc**

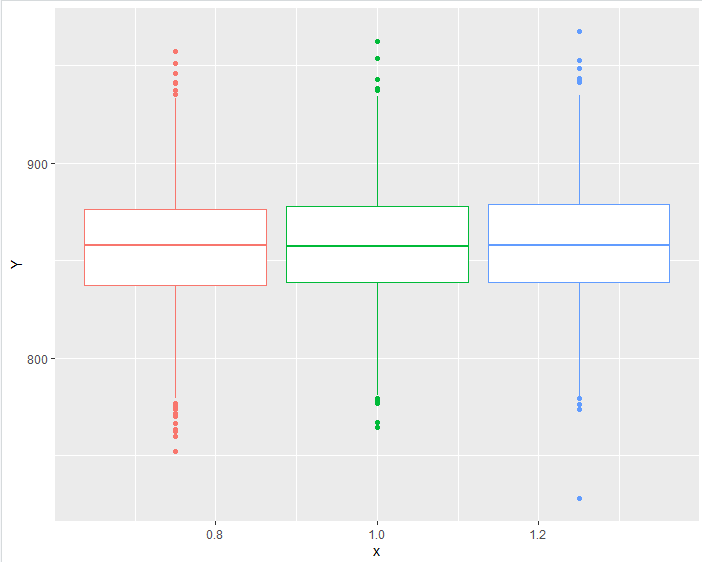


1. Vẽ đồ thị density, histogram và đồ thị box – plot biến phụ thuộc sử dụng colour = Dummy

**Hình 4. Đồ thị histogram biến phụ thuộc**

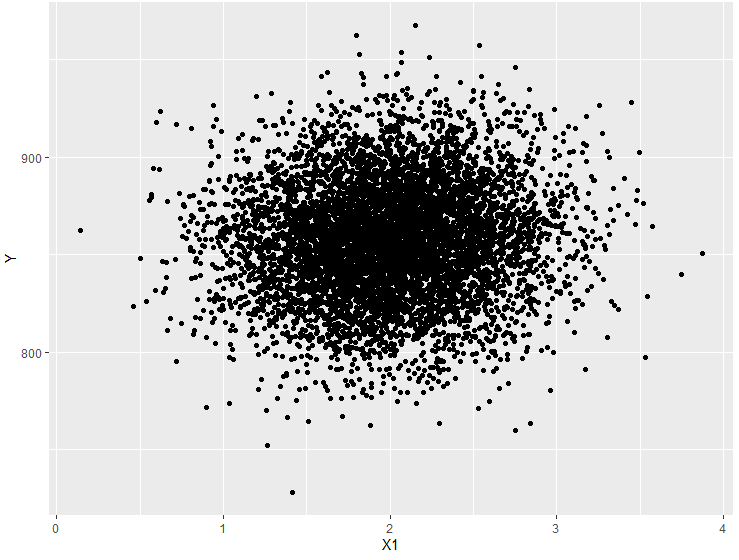
**Hình 5. Đồ thị density biến phụ thuộc**

**Hình 6. Đồ thị box - plot biến phụ thuộc**



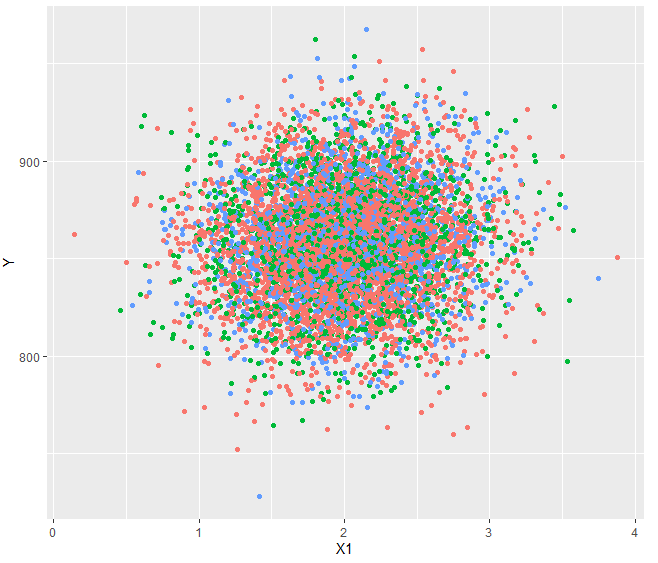
1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1

**Hình 7. Đồ thị scatter giữa biến phụ thuộc Y và biến X1**



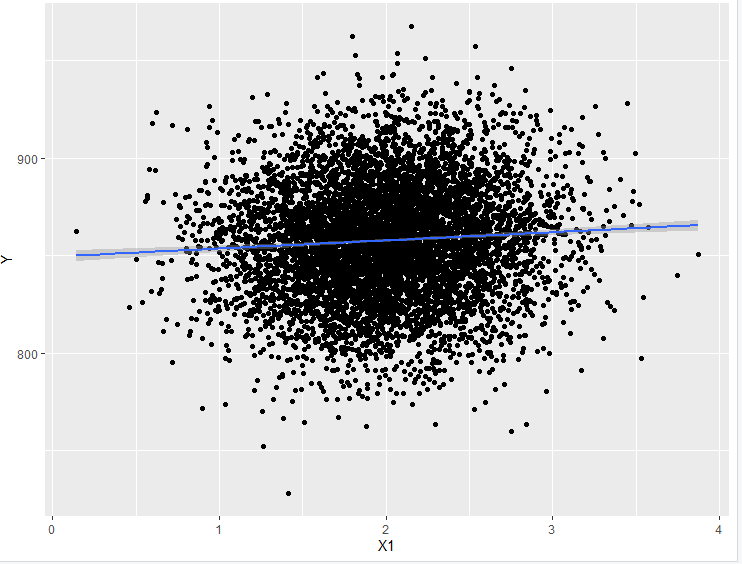
1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1  sử dụng colour = Dummy

**Hình 8. Đồ thị scatter giữa biến phụ thuộc Y và biến X1**



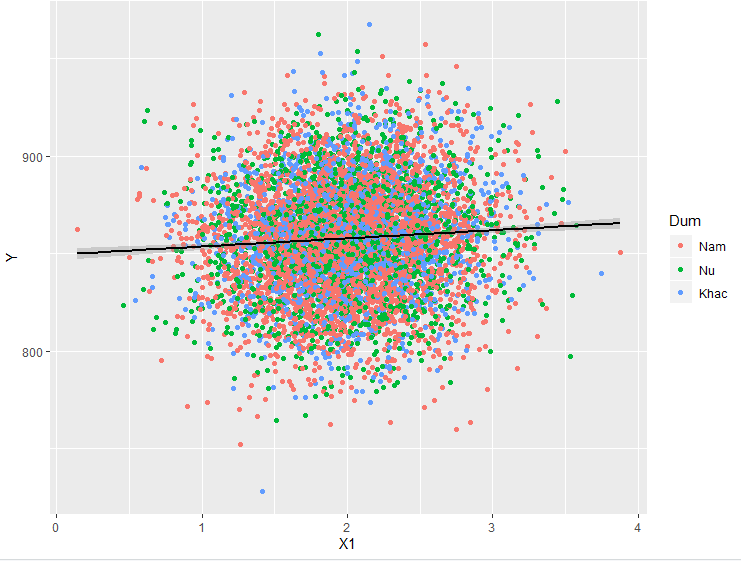
1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1  + vẽ đường hồi quy

**Hình 9. Đồ thị scatter và đường hồi quy giữa biến phụ thuộc Y và biến X1**



1. Vẽ đồ thị scatter giữa biến phụ thuộc Y và biến độc lập X1  sử dụng colour = Dummy + đường hồi quy

**Hình 10. Đồ thị scatter và đường hồi quy giữa biến phụ thuộc Y và biến X1**



**Phần 3. Phân tích ma trận tương quan:**

1. Ma trận tương quan giữa các ***biến độc lập định lượng***

**Bảng 5. Bảng ma trận tương quan**

|  |
| --- |
| X1 X2 X3 X4 |
| X1 1.000000000 -0.0122046230 -0.0134598643 0.005796956 |
| X2 -0.012204623 1.0000000000 0.0001841389 0.017093210 |
| X3 -0.013459864 0.0001841389 1.0000000000 0.011730346 |
| X4 0.005796956 0.0170932097 0.0117303456 1.000000000 |

|  |
| --- |
| Nhận xét tóm gọn về hệ số tương quan:  -Cor(X1, X2): -0.89% -> X1, X2 dao động ngược chiều,  -Cor(X1, X3): -4.39% -> X1, X3 dao động ngược chiều,  -Cor(X1, X4): 1.08% -> X1, X4 dao động cùng chiều  -Cor(X2, X3): 0.19% -> X2, X3 dao động cùng chiều  -Cor(X2, X4): 10.45% -> X2, X4 cùng chiều, dao dộng cùng nhau 10.45%  -Cor(X3, X4): 31.77% ->X3, X4 cùng chiều, dao dộng cùng nhau 31,77% |

**Phần 4. Phân tích hồi quy:**

**4.1. Mô hình 1: Tác động của biến độc lập định lượng tới biến phụ thuộc**

* **Phương trình:**

**Y = 0 + 1X1 + X + …+ nXn**

**Bảng 6. Kết quả hồi quy**

*(Insert table here)*

|  |
| --- |
| **Nhận xét tóm gọn những nội dung sau:**   * R-Saquared: * Kiểm định mức ý nghĩa tổng thể của mô hình: * Mức ý nghĩa thống kê của từng biến độc lập: * Mức tác động biên của từng biến độc lập: |

**4.2. Mô hình 2: Tác động của biến độc lập định tính và định lượng tới biến phụ thuộc (không xuất hiện biến tương tác)**

* **Phương trình:**

**Y = 0 + 1Dum + X + X + …+ kXn**

**Bảng 7. Kết quả hồi quy**

*(Insert table here)*

|  |
| --- |
| **Nhận xét tóm gọn những nội dung sau:**   * R-Saquared: * Kiểm định mức ý nghĩa tổng thể của mô hình: * Mức ý nghĩa thống kê của từng biến độc lập: * Mức tác động biên của từng biến độc lập: |

**4.3. Mô hình 3: Tác động tương tác của biến độc lập định tính và định lượng tới biến phụ thuộc**

* **Phương trình:**

**Y = 0 + 1(Dum\*X1) + Dum + X + X +…+ kXn**

**Bảng 8. Kết quả hồi quy**

*(Insert table here)*

* **Viết phương trình kết quả hồi quy:**

*(Insert here)*

|  |
| --- |
| **Nhận xét tóm gọn những nội dung sau:**   * R-Saquared: * Kiểm định mức ý nghĩa tổng thể của mô hình: * Mức ý nghĩa thống kê của từng biến độc lập: * Mức tác động biên của từng biến độc lập: |

* **Kiểm định đa cộng tuyến:**

**Bảng 9. Hệ số VIF**

*(Insert table here)*

|  |
| --- |
| Nhận xét tóm gọn kết quả kiểm định: |

* **Kiểm định phương sai của sai số thay đổi:**

**Bảng 10. Kết quả kiểm định phương sai của sai số thay đổi**

*(Insert table here)*

|  |
| --- |
| **Nhận xét tóm gọn kết quả kiểm định:** |

* **Kiểm định phân phối chuẩn của sai số** (**nâng cao**):

**Bảng 11. Kết quả kiểm định phân phối chuẩn của sai số thay đổi**

*(Insert table here)*

* Tính biến số được đặt tên là **Y\_fitted** (giá trị dự đoán của Y) và **Residuals** (sai số của mô hình): Không báo cáo ở đây
* Vẽ đồ thị scatter: Y\_fitted vs Y và Y\_fitted vs. Residuals

**Hình 11. Đồ thị scatter Y\_fitted và Y**

*(Insert figure here)*

**Hình 12. Đồ thị scatter Y\_fitted và Residuals**

*(Insert figure here)*

**KẾT THÚC**

***Chúc mừng các em đã hoàn thành capstone project!!!***