Hướng giải quyết Question1-2

Bước 1: Thêm package utf8, tidyverse, xlsx, e1071

Bước 2: Đọc file xlsx vào biến data frame DataChart, gán columns names = ("stdid", "stat", "time\_begin", "time\_end", "time\_duration", "total\_score", "Q1", "Q2", "Q3", "Q4", "Q5", "Q6", "Q7", "Q8", "Q9", "Q10")

Xử lý chuỗi ở cột 6 (Total score), các chuỗi ‘-‘ thay thành -1,00, đổi dấu ‘,’ thành ‘.’ Ép kiểu về numeric

Lấy subset lên DataChart để loại các sinh viên không nộp bài (điểm -1.00)

Bước 3: Giải quyết câu hỏi

2a) Lấy thông tin cột 6 của DataChart

2b) Lấy min(cột 6)

2c) Dùng hàm subset lên DataChart với ĐK total\_score == min\_total\_score rồi lấy cột stdid được danh sách sinh viên chứa lần nộp điểm thấp nhất, dung unique() để loại hàm trùng lặp

2d) Tạo data frame mới lowest\_list với 2 column là stdid và times (số lần nộp), dung vòng lặp đếm số lần xuất hiện stdid các sinh viên trong DataChart tương ứng số lần nộp, dung hàm hist để vẽ phổ phân theo số lần xuất hiện của số lượng nộp khác nhau

2e) Lấy cột stdid của DataChart, dung hàm unique để loại giá trị trùng lập, ta được danh sách sinh viên

Tạo data frame mới final\_score\_chart với column stdid là danh sách sinh viên mới lấy, column final\_score là điểm tổng kết sinh viên gán cho vector c(0)

Lặp hai vòng cho mỗi stdid trong DataChart lấy điểm lớn nhất trong các lần nộp làm điểm tổng kết

Lấy min() trong cột final\_score được điểm tổng kết thấp nhất

2f) Lấy subset trên final\_score\_chart điều kiện final\_score == giá trị min ở câu trên

2g) Tạo data frame lowest\_final\_list với col stdid của các sinh viên có điểm tổng kết thấp nhất và column times là số lần nộp, dung vòng lặp trên DataChart để đếm số lần lặp mỗi sinh viên

Dùng hàm hist để vẽ theo phổ số lần xuất hiện số lần nộp khác nhau

2h) 2i) 2j) trùng với 2k) 2l) 2m) bởi vì những ai có lần nộp điểm cao nhất cũng sẽ có điểm tổng kết cao nhất

2h) Lấy max của cột final\_score trong final\_score\_chart

2i) Lấy subset lên final\_score\_chart điều kiện final\_score == max final\_score lấy ở trên, lấy cột stdid

2j) Tạo data frame highest\_final\_score\_chart với column stdid vừa lấy ở câu i) và column times là số lần nộp bài, dung vòng lặp lên DataChart để đếm và dung hist để vẽ phổ các số lần lặp khác nhau

2k,l,m) giống như trên

2n) Lấy sum cột final\_score của final\_score\_chart rồi chia cho nrow(final\_score\_chart)

2o)Subset final\_score\_chart với điều kiện final\_score == ĐTB lấy ở câu 2.n)

2p) Dùng hàm min,max,median lên cột DataChart để lấy theo điểm tổng kết hoặc final\_score\_chart để lấy theo điểm submit

2q) Độ phân tán: đo phương sai, độ lệch chuẩn với range của dữ liệu hàm

2r) Dùng hàm skewness() và kurtosis

2s) Lấy hàm quantile() lên final\_score\_chart ở cột final\_score với tham số 25% tứ phân vị thứ nhất, 75% tứ phân vị thứ hai

2t) Lập data frame result\_chart (score, student\_num), lặp trong final\_score cột final\_score, nếu thấy chưa có điểm số trong result\_chart thì add hàng mới vô, có rồi thì cộng 1 vô số sinh viên. Sort lại thứ tự theo điểm số final\_score giảm dần.

Rồi lấy times(số lượng sinh viên) ở 2 hàng đầu cộng lại

2u) Tạo data frame final\_score\_chart\_first\_second là subset của final\_score\_chart với điểm số bằng một trong hai điểm cao nhất (2 score của 2 hàng đầu trong result\_chart)

Dùng hist() trên final\_score\_chart\_first\_second$times để vẽ phổ là tần số nộp bài của sinh viên ở các phân điểm tổng kết cao nhất và cao nhì

2v) và 2w) Gán result\_chart cho một cột mới là total\_submit thể hiện tổng số lần nộp của các sinh viên thuộc mỗi phân điểm tổng kết khác nhau

Lặp trong final\_score\_chart so sánh điểm số gặp sinh viên có điểm tổng kết bằng với điểm số score thì lấy total\_submit ở cùng hang với score đó cộng cho số lần nộp của sinh viên đó.

Đặt lại giá trị score gán cho c(1:nrow(result\_chart)) đưa về thứ tự k

Dùng barplot vẽ đồ thị biểu diễn tổng số sinh viên theo điểm tổng kết và tổng số lần nộp nhóm theo điểm tổng kết