

Задание 2

1. Скачать данные diabetes.csv и выполнить с ними задание.
2. Загрузить данные diabetes.csv
3. Переформатировать качественные переменные в факторные
4. Сформировать две подвыборки по возрасту – старше 45 и младше 45
5. Проверить признак BMI (индекс массы тела) в двух сформированных подвыборках на нормальность распределения, сделать вывод о подчинении признака нормальному распределению.
6. Провести тесты на сравнения сформированных выборок:
 - 6.1 тест Колмогорова-Смирнова для сравнения медиан двух независимых выборок, сделать вывод
 - 6.2 тест Мана-Уитни для двух независимых выборок, сделать вывод
 - 7.1 Добавить к данным переменную ins, равную 1, если переменная insulin>0, 0 в противном случае с помощью команды mutate
data_n <- data %>% mutate(ins=ifelse(data\$ Insulin>0, 1, 0))
ins = as.factor(ins)
 - 7.2 Провести тест Хи-квадрат по признакам Outcome и ins, сделать вывод
 - 7.3 Сделать тест Краскелла-Уоллиса по признаку количество родов Pregnancies (требуется преобразовать в факторную*) и признаку DiabetesPedigreeFunction, сделать вывод
8. Оформить отчет, где статистические тесты оформлять по примеру:
H0: переменная согласуется с нормальным распределением
Shapiro-Wilk normality test
data: data_y\$chol
W = 0.98507, p-value = 0.1052
Вывод: $p > 0,1$, следовательно H0 принимаем, распределение признака согласуется с нормальным.
9. Прикрепить в СДО отчет по результатам проведения статистических тестов и скрипт в R.