Домашнее задание

## Задача 1

Для маркетингового исследования случайным образом необходимо отобрать N людей. Этим людям далее демонстрируется кофе марки Nescafe в упаковке с новым дизайном: вместо классической этикетки в коричнево-золотых тонах – новогодняя этикетка в красно-зеленых тонах. После этого фиксируется, сколько людей захотели купить этот кофе, а точнее, доля таких людей. Рассчитайте необходимый объем выборки для тестовой группы, если известно, что \* выбран 95%-ный уровень доверия; \* погрешность не должна превышать 3%; \* нет определенных ожиданий относительно доли людей, которые захотят кофе с измененным дизайном этикетки (считается, что человек с равной вероятностью выберет кофе или не выберет). Для вычислений можно использовать базовые средства R или Python, но в таком случае необходимо привести код в тексте с решением.

p <- 0.5 # стандартная доля ожидания  
z <- 1.96 # значение уровня доверия 95%  
errors <- 0.03 # доля ошибки исходя из 3 %  
  
n <- z\*\*2 \* p \* (1 - p) / (errors\*\*2)  
n

## [1] 1067.111

*Объем выборки респондентов для тестовой группы должен составлять не менее 1067 опросов*

## Задача 2

Решите задачу 1, учитывая, что теперь ожидаемая доля людей, которые захотят купить кофе с измененным дизайном этикетки, известна и равна 0.67 (группа аналитиков нашла старый отчет с результатами аналогичного эксперимента). Какой из необходимых объемов выборки больше: в случае, когда мы считаем, что шансы выбрать и не выбрать кофе одинаковы или в случае, когда ожидаемая доля выбравших кофе нам известна и равна 0.67? Сделайте вывод о связи между ожидаемой долей и необходимым объемом выборки при прочих равных условиях (тот же уровень доверия).

p2 <- 0.67 # ожидаемая доля людей  
z2 <- 1.96 # значение уровня доверия 95%  
errors2 <- 0.03 # доля ошибки исходя из 3 %  
  
n2 <- z2\*\*2 \* p2 \* (1 - p2) / (errors2\*\*2)  
n2

## [1] 943.7531

*Как видно при ожидаемой доли людей, которые захотят новый продукт в 0.67 объем требуемой выборки для исследевания меньше нежели при доле равной 0.5 доля 0.5 объем выборки 1067 доля 0.67 объем выборки 943*

## Задача 3

В файле experiment.csv сохранены результаты эксперимента, описанного в задаче 1. Данные пред- ставляют собой последовательности из 0 и 1, где 1 соответствуют людям, которые захотели выбрать кофе с измененным дизайном этикетки, а 0 - людям, которые не захотели.

*Описание данных:* test: результаты по тестовой группе (людям предлагали кофе с новым дизайном этикетки); control: контрольная группа (людям предлагали кофе со старым дизайном этикетки).

1. Загрузите в R данные из файла experiment.csv с помощью функции read.csv().

df <- read.csv('G:/Data sainse/R/homeworke/5.7 Domashnee\_zadanie/experiment.csv')  
head(df)

## X control test  
## 1 1 0 0  
## 2 2 1 1  
## 3 3 0 0  
## 4 4 0 0  
## 5 5 0 0  
## 6 6 1 1

1. Постройте в R 90%-ный доверительный интервал для доли людей, которые захотели купить кофе, в тестовой группе. Проинтерпретируйте полученный доверительный интервал. В решении приведите используемый код, границы доверительного интервала и текст с интерпре- тацией.

library(DescTools)# подгружаем нужную библиотеку  
  
BinomCI(x = sum(df$test), n = length(df$test), conf.level = 0.9)

## est lwr.ci upr.ci  
## [1,] 0.577 0.5511289 0.6024556

*Доверительный интегвал для тестовой группы предположительно находится в диапазоне: нижняя граница 0.551, верхняя граница 0.602 и середина 0.577 Где 0,577 это доля людей, которые захотели купить кофе*

1. Постройте в R 90%-ный доверительный интервал для доли людей, которые захотели купить кофе, в контрольной группе. Проинтерпретируйте полученный доверительный интервал. В решении приведите используемый код, границы доверительного интервала и текст с интерпре- тацией.

BinomCI(x = sum(df$control), n = length(df$control), conf.level = 0.9)

## est lwr.ci upr.ci  
## [1,] 0.464 0.4381921 0.4900022

*Из построенных доверительных интервалов по тестовой и контрольной группе видно, что диапазоны не накладываются друг на друга, Это является статистически значимо и говорит о присутствии разницы в долях. Можно предположить, что новая упаковка является более привлекательной.*

1. Сравните полученные доверительные интервалы для тестовой и контрольной группы. Можно ли на основании полученных результатов сделать определенный вывод о том, что новый дизайн этикетки лучше (привлекает больше потенциальных покупателей)? Приведите ваши выводы и объяснения.

*Опираясь на проведенные исследования и полученные доверительные интервалы можно предположить, что новый дизайн более привлекателен. Об этом свидетельствует отсутствие в пересечении доверительных интервалов, а это говорит о присутствии значимой разницы в долях, если же обратить внимание долю желающих купить кофе, то значения с новым дизайном будут выше*