

STAT1301 Assignment 4

Caleb Yates s49886351

October 14, 2025

1 Question 4

1.1 Part a)

Let X be the random variable for the number of people received the direct mail strategy and completed screening. Let Y be the random variable for the number of people who received the education only outreach and completed screening.

The notation p_X and p_Y represent the population proportion for X and Y respectively.

The null hypothesis is that both population proportions are equal: $H_0 : P_X - P_Y = 0$. The alternative hypothesis is therefore: $H_1 : \mu_X - \mu_Y > 0$.

1.2 Part b)

$X \sim \text{Bin}(n_X, p_X)$ where $n_X = 1415$ so $X \sim \text{Bin}(n_X, p_X)$. It is (implicitly) assumed that samples X_i from X follow the distribution of X , hence:

$$X_i \sim \text{Bin}(n_X, p_X)$$

Since $n_X p_X = 505 \gg 5$ and $n_X(1 - p_X) = 910 \gg 5$, the conditions for the Central Limit Theorem (CLT) to apply are met, as well as a suitably large n_X . Therefore:

$$X_i \overset{\text{approx}}{\sim} N(n_X p_X, n_X p_X (1 - p_X))$$

$$\bar{X} \overset{\text{approx}}{\sim} N(p_X, \frac{p_X(1 - p_X)}{n_X})$$

$$Y \sim \text{Bin}(n_Y, p_Y)$$