Mini Market

By: Rohit Manjunath

What Sell Together?

- To determine if item x and y sell together, we can calculate the average of item ‘x’ given whether item ‘y’ was bought or not.

- This dataset has 5 columns all numerical(sort of binary) which contains 0 and 1 values. 1 if the item was bought and 0 otherwise.

- Using the permutation test we can find if the relationship is a result of random data or not

Bonferroni Correction

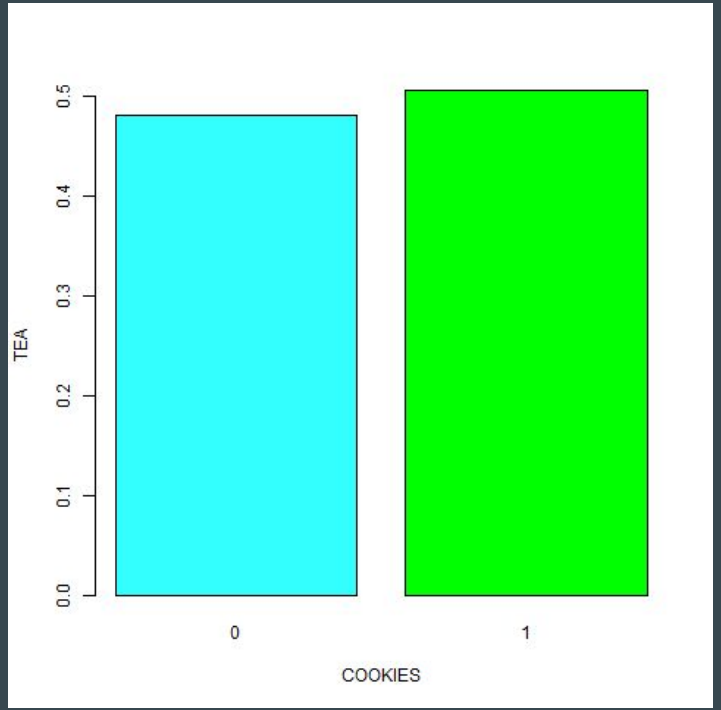
- Since this dataset has a 5 columns and we want to find a relationship between 2 columns, we can create biases to find our desired result. Hence, we need to change the p-value threshold.
- To find all the possible combinations we can make with 5 columns we find the value of 5C2 which is 10.
- Hence, with a p-value threshold of 0.05 we will have to divide this by 10 (5C2). Which gives us a new value of 0.005.
- Any relationship with a p-value below 0.005 will reject the null hypothesis.

Cookies and Tea

- As you can see the bar graph on the right show that people prefer to buy cookies and tea together. The average of people buying tea is higher when people buy cookies. Using the permutation test we get:

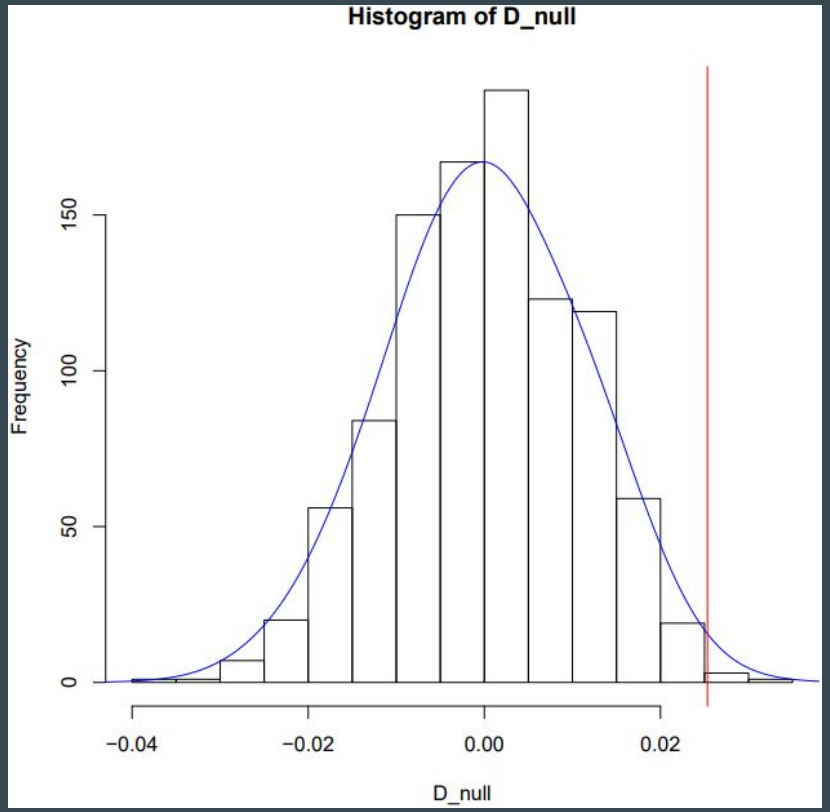
```
p <- Permutation(market, "COOKIES", "TEA",
1000, 1, 0)
p
Output:
p : 0.003
```

- The p value is below .005(after Bonferroni Correction). Hence, this relationship is not a result of random data.




```
barplot(tapply(market$TEA, market$COOKIES, mean),
col = c("#33FFFF", "green", "yellow", "orange",
"red"), xlab = "COOKIES", ylab = "TEA")
```

Permutation Test Result for Cookies and Tea

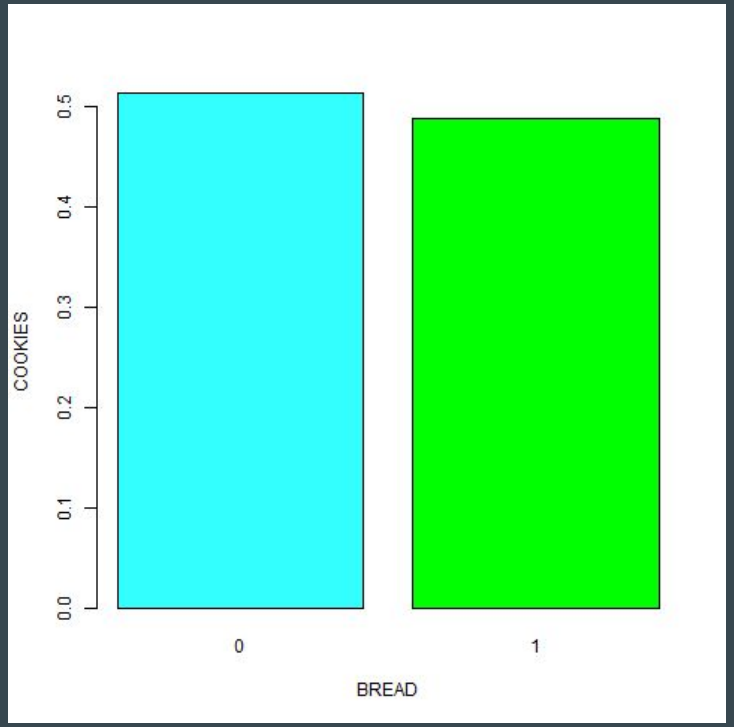


Bread and Cookies

- As you can see the bar graph on the right show that people prefer to either buy bread or cookies not both. The average of people buying cookies is higher when people don't buy bread. Using the permutation test we get:

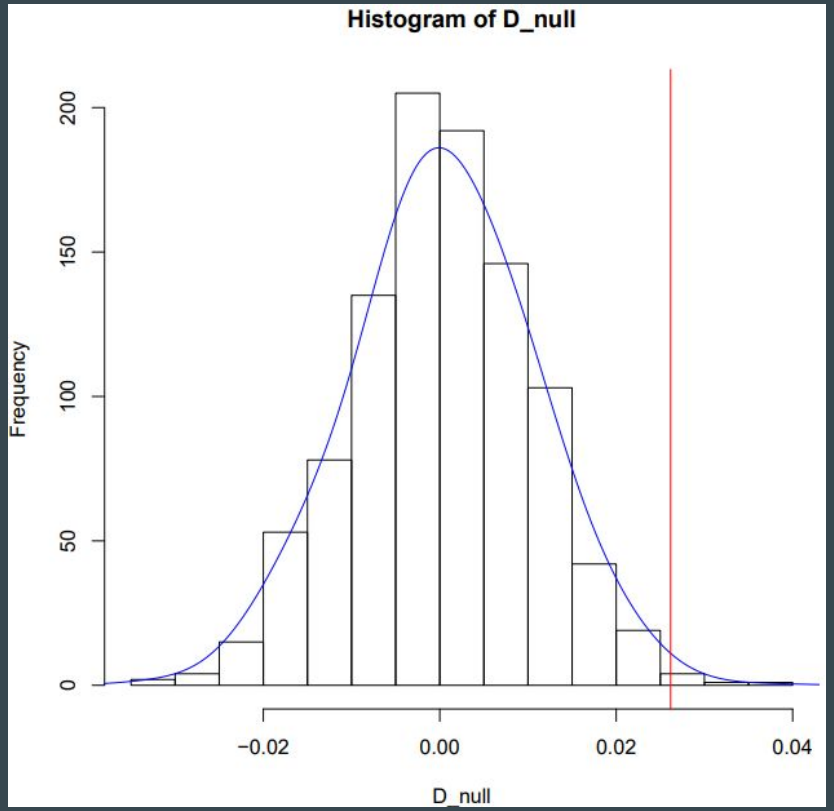
```
p <- Permutation(market, "BREAD", "COOKIES",
1000, 1, 0)
p
Output:
p : 0.003
```

- The p value is below .005(after Bonferroni Correction). Hence, this relationship is not a result of random data.




```
barplot(tapply(market$COOKIES, market$BREAD, mean),
col = c("#33FFFF", "green", "yellow", "orange",
"red"), xlab = "BREAD", ylab = "COOKIES")
```

Permutation Test Result for Bread and Cookies



Thank you!