

# What sells together? Data Puzzle

Yu-Hon Lin

# Methodology



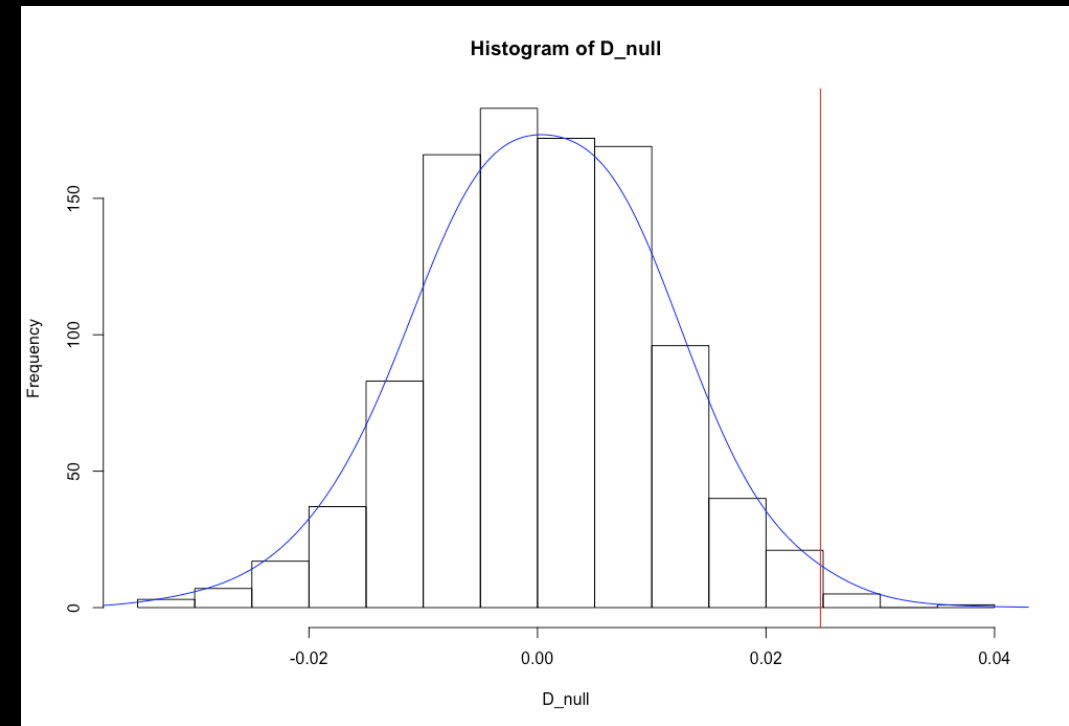
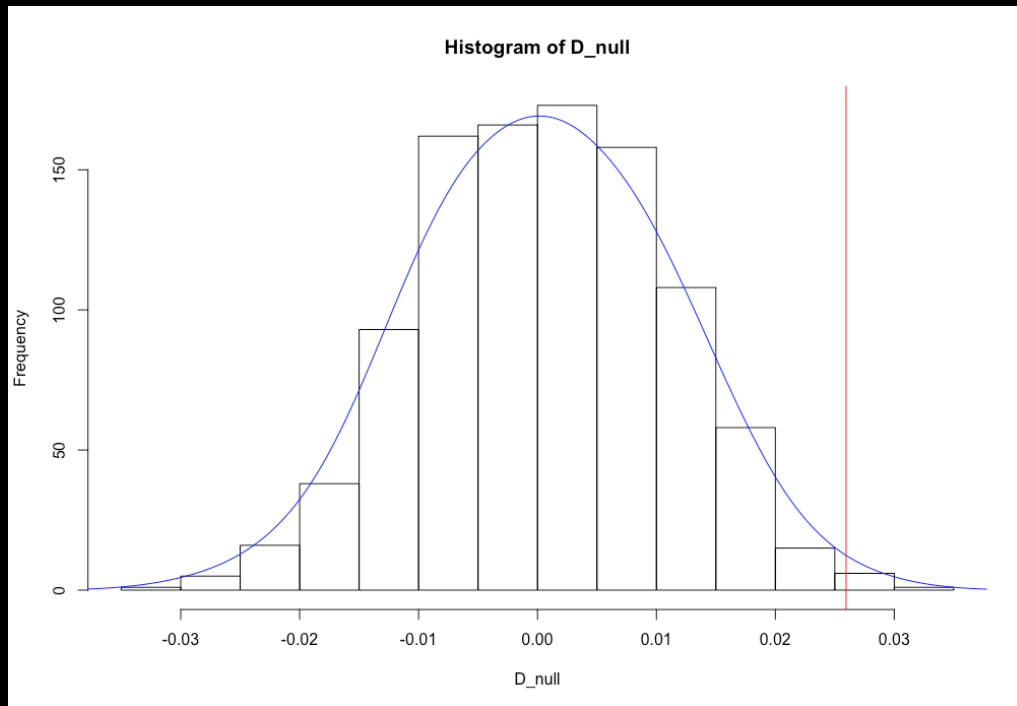
1. Use table function to identify the most common pair of items in the dataset. The most popular pair I found is soda + cookies. For example:

```
> table(WALMART1$SODA, WALMART1$COOKIES)
```

```
  0   1  
0 2419 2340  
1 2380 2542
```

2. Run permutation tests and find the associations between items sold in Wal-Mart

# Soda



Permutation test

```
> soda <- WALMART1
```

```
> soda[soda$SODA == 1, 1] = 'Yes'
```

```
> soda[soda$SODA == 0, 1] = "No"
```

```
> PermutationTestSecond::Permutation(soda, "SODA", "FISH", 1000, "No", "Yes")
```

```
[1] 0.006
```

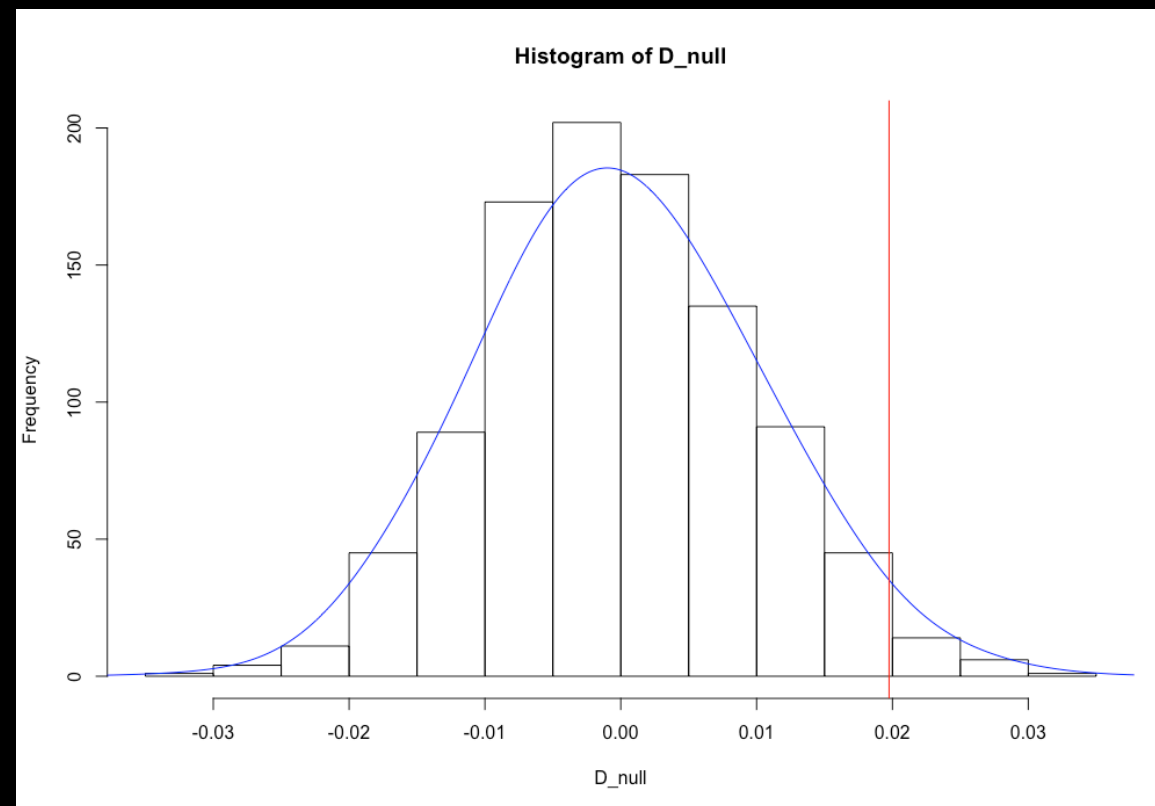
```
> PermutationTestSecond::Permutation(soda, "SODA", "COOKIES", 1000, "No", "Yes")
```

```
[1] 0.006
```

## Summary:

**Both fish and cookies sell together well with soda**

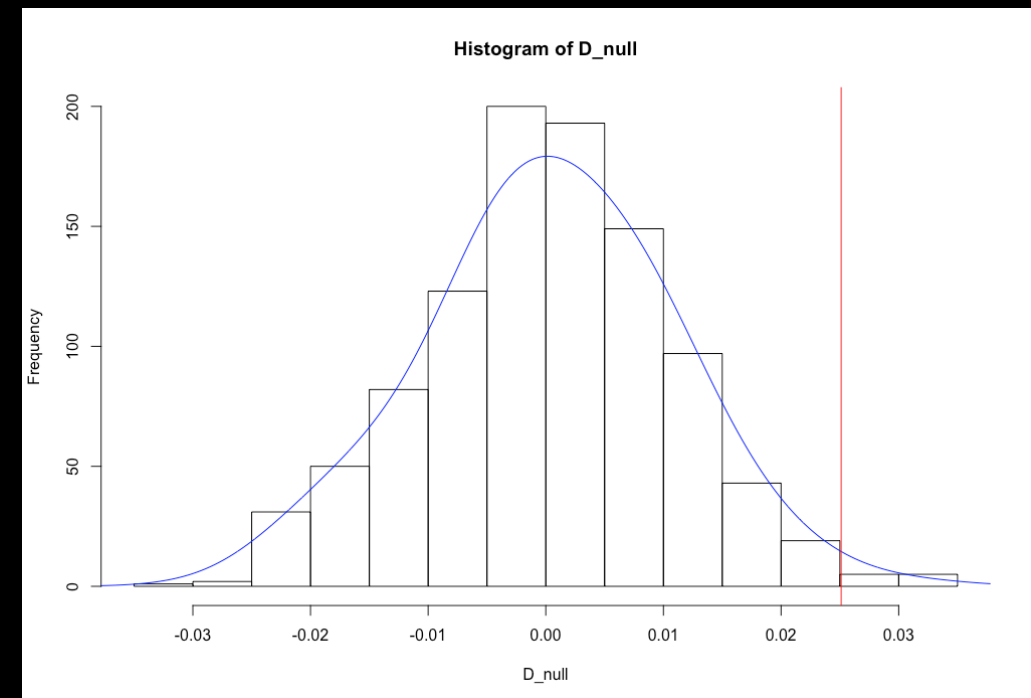
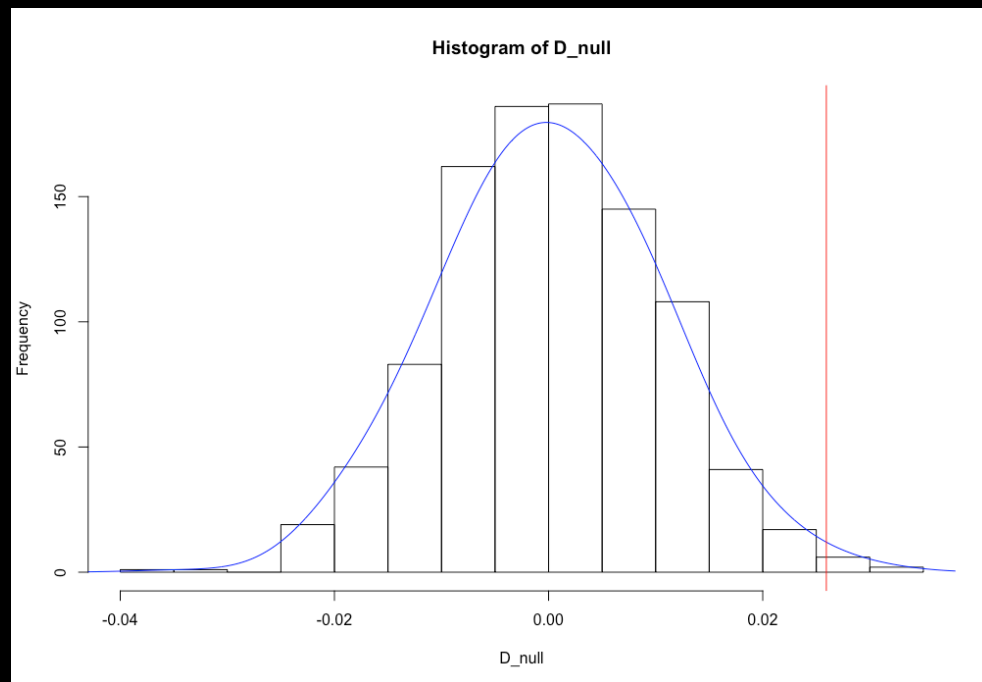
# Steak



```
Permutation test
> steak <- WALMART1
> steak[steak$STEAK == 1, 2] = 'Yes'
> steak[steak$STEAK == 0, 2] = "No"
> PermutationTestSecond::Permutation(steam, "STEAK", "SALT", 1000, "No", "Yes")
[1] 0.021
```

**Summary:**  
**Salt sells together well with steak**

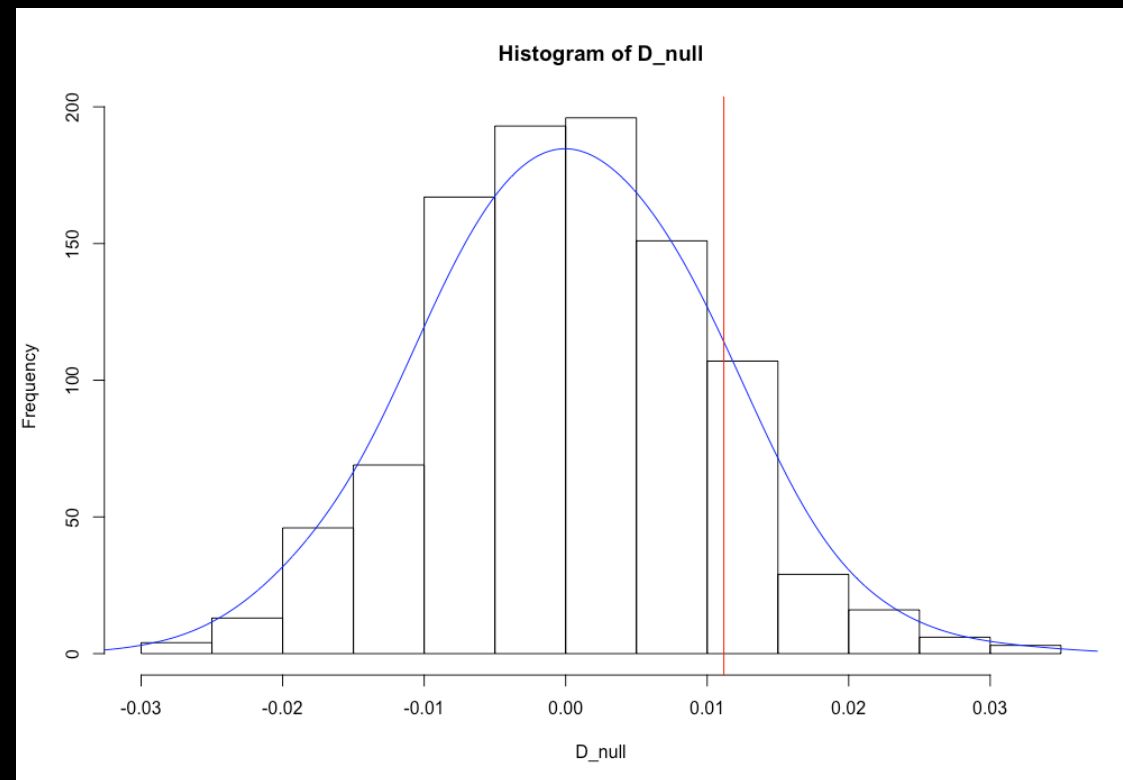
# Fish



```
Permutation test
> fish <- WALMART1
> fish[fish$FISH == 1, 3] = "Yes"
> fish[fish$FISH == 0, 3] = "No"
> PermutationTestSecond::Permutation(fish, "FISH", "SODA", 1000, "No", "Yes")
[1] 0.006
> PermutationTestSecond::Permutation(fish, "FISH", "TEA", 1000, "No", "Yes")
[1] 0.009
```

**Summary:**  
**Both soda and tea sell together well with fish**

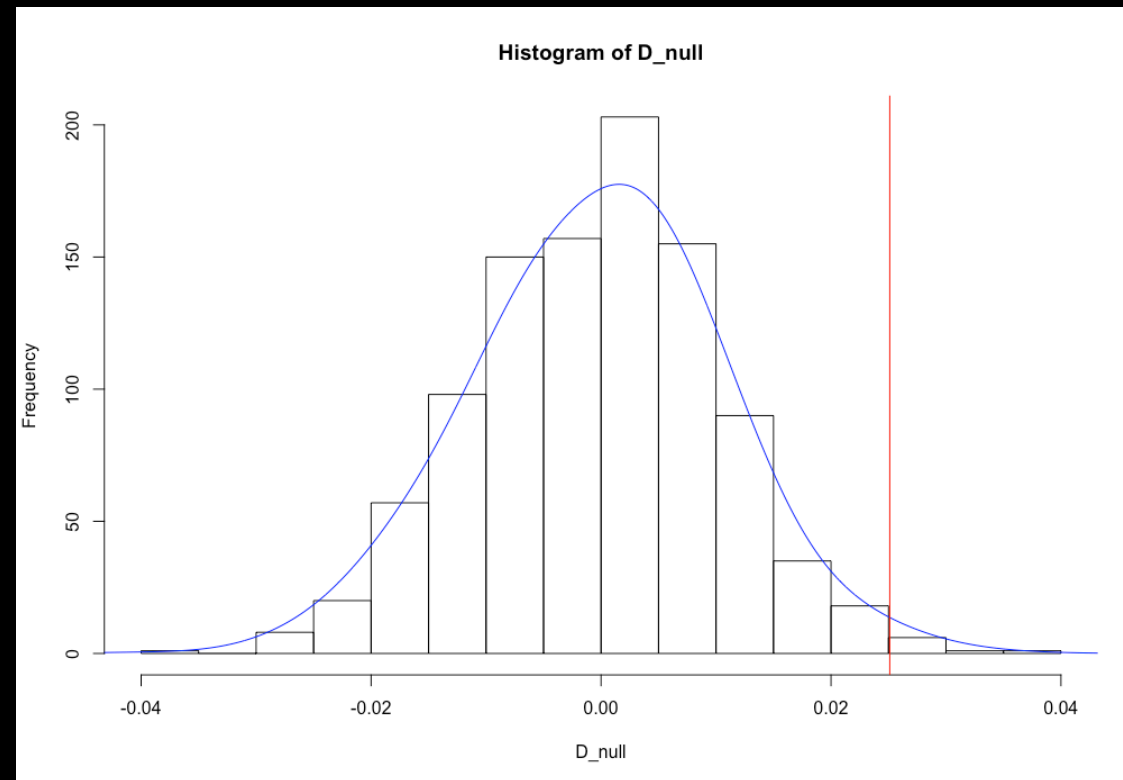
# Coffee



```
Permutation test
> coffee <- WALMART1
> coffee[coffee$COFFEE == 1, 4] = "Yes"
> coffee[coffee$COFFEE == 0, 4] = "No"
> PermutationTestSecond::Permutation(coffee, "COFFEE", "TEA", 1000, "No", "Yes")
[1] 0.124
```

Permutation tests reveal no significant correlation between other items purchased by coffee buyers, with the lowest p value being the coffee-tea pair.

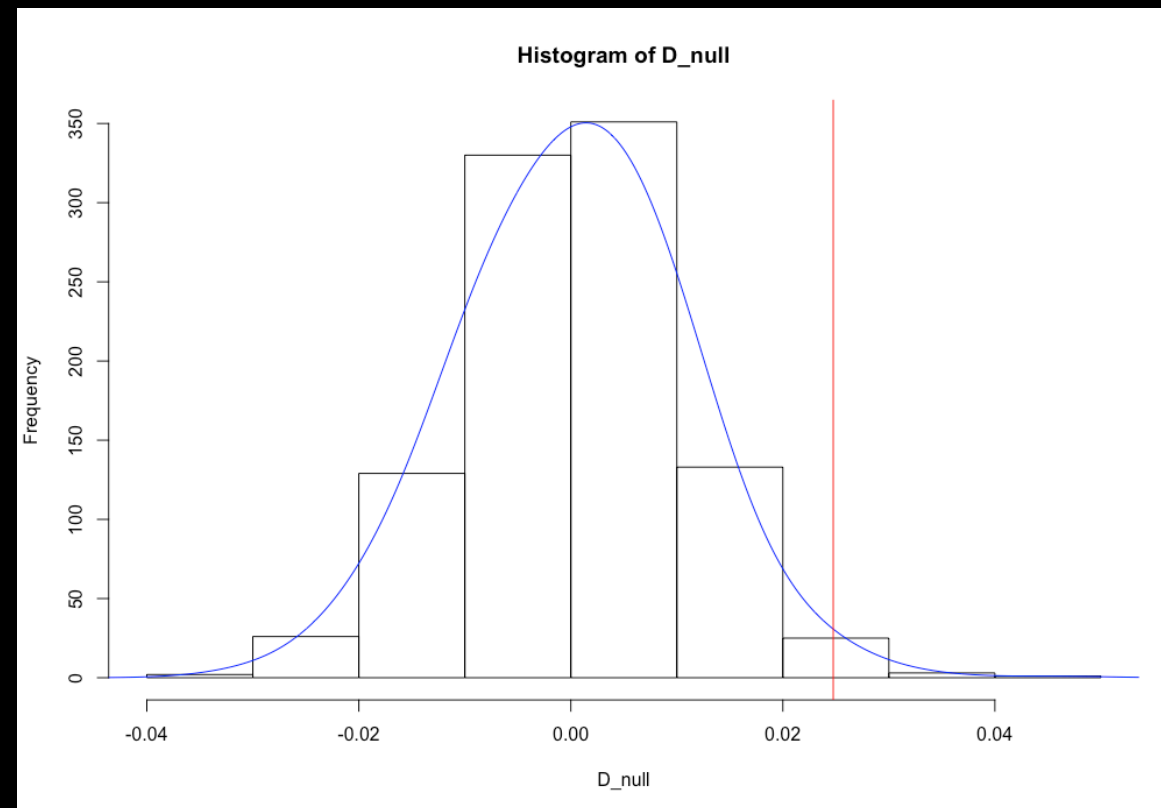
# Tea



```
Permutation test
> tea <- WALMART1
> tea[tea$TEA == 1, 5] = "Yes"
> tea[tea$TEA == 0, 5] = "No"
> PermutationTestSecond::Permutation(tea, "TEA", "FISH", 1000, "No", "Yes")
[1] 0.008
```

**Summary:**  
**Fish sells together well with tea**

# Cookies

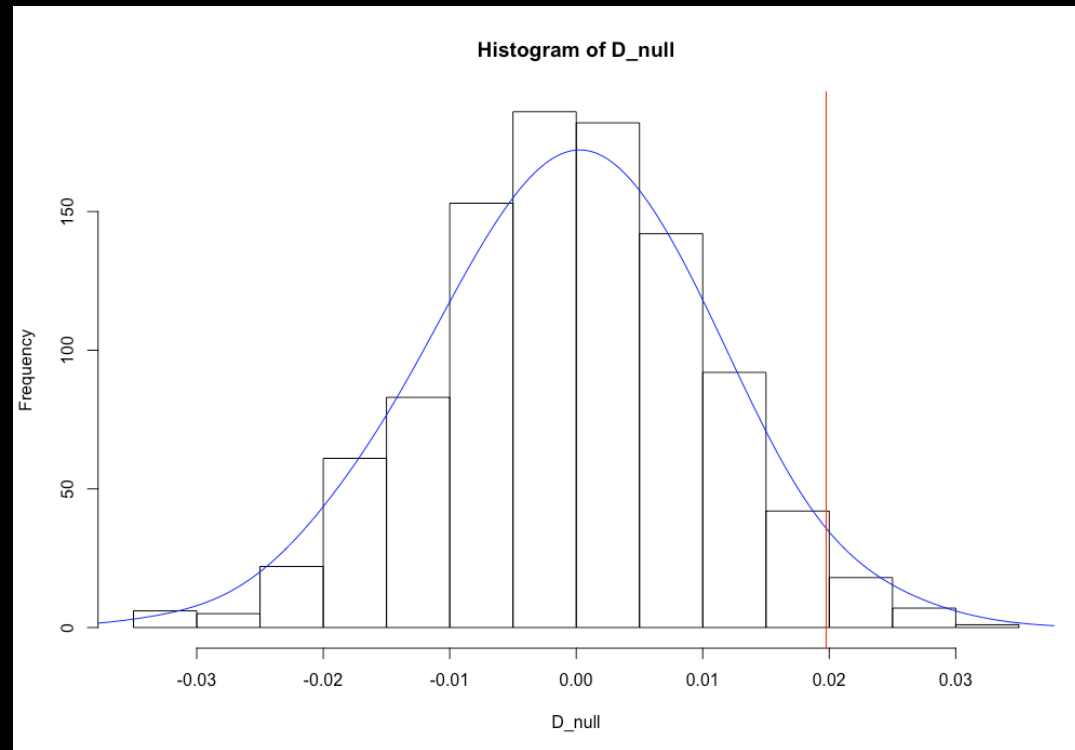


```
Permutation test
> cookies <- WALMART1
> cookies[cookies$COOKIES == 1, 6] = "Yes"
> cookies[cookies$COOKIES == 0, 6] = "No"
> PermutationTestSecond::Permutation(cookies, "COOKIES", "SODA", 1000, "No", "Yes")
[1] 0.012
```

**Summary:**  
**Soda sells together well with cookies**



# Salt



```
Permutation test
> salt <- WALMART1
> salt[salt$SALT == 1, 7] = "Yes"
> salt[salt$SALT == 0, 7] = "No"
> PermutationTestSecond::Permutation(salt, "SALT", "STEAK", 1000, "No", "Yes")
[1] 0.026
```

**Summary:**  
**Steak sells together well with salt**