



# Cookies & ~~Milk~~...Soda?

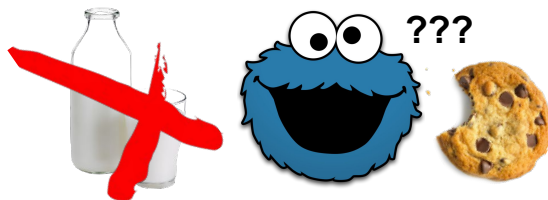
Matthew Lau  
Professor Imielinski  
March 28, 2017  
Data 101 WALMART Assignment

**You might think Cookies and  
Milk are bought together...**

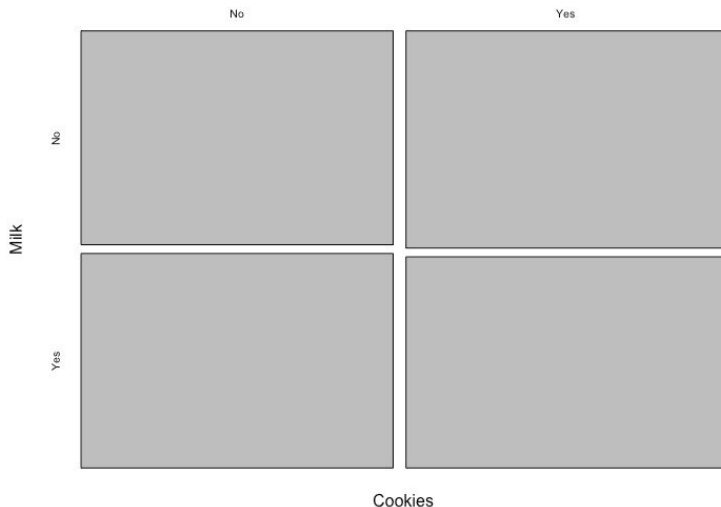


# You would be wrong.

*There are no specific associations between Walmart customers buying milk and cookies together.*



Cookies and Milk Mosaic Plot



## Permutation Test (5x):

```
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "MILK", 1000, 1, 0)
[1] 0.214
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "MILK", 1000, 1, 0)
[1] 0.226
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "MILK", 1000, 1, 0)
[1] 0.218
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "MILK", 1000, 1, 0)
[1] 0.219
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "MILK", 1000, 1, 0)
[1] 0.199
```

Result

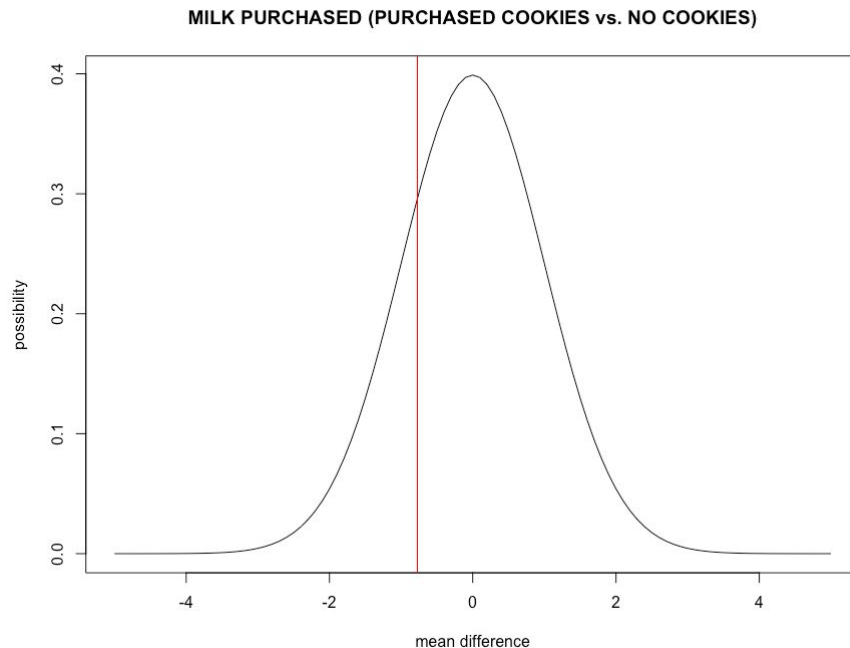
P-value > .05

# Z-TEST: Milk and Cookies

*I decided to perform a Z-TEST on this as well.*

## R Code:

```
> YES.Cookies <- subset(WALMART, WALMART$COOKIES == 1)
> NO.Cookies <- subset(WALMART, WALMART$COOKIES == 0)
>
>
> YES.Cookies.Milk <- YES.Cookies$MILK
> NO.Cookies.Milk <- NO.Cookies$MILK
> sd.YES.Cookies.Milk <- sd(YES.Cookies.Milk)
> sd.NO.Cookies.Milk <- sd(NO.Cookies.Milk)
> mean.YES.Cookies.Milk <- mean(YES.Cookies.Milk)
> mean.NO.Cookies.Milk <- mean(NO.Cookies.Milk)
> len_YES.Cookies.Milk <- length(YES.Cookies.Milk)
> len_NO.Cookies.Milk <- length(NO.Cookies.Milk)
> sd.YES.NO.Cookies.Milk <- sqrt(sd.YES.Cookies.Milk^2/len_YES.Cookies.Milk + sd.NO.Cookies.Milk^2/len_NO.Cookies.Milk)
>
> zeta <- (mean.YES.Cookies.Milk - mean.NO.Cookies.Milk)/sd.YES.NO.Cookies.Milk
> zeta
[1] -0.7729675
> p <- pnorm(zeta)
> p
[1] 0.2197708
>
```



## Result

P-value = 0.21 > .05 NO correlation



TEA

```
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "TEA", 1000, 1,0)
[1] 0.125
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "TEA", 1000, 1,0)
[1] 0.119
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "TEA", 1000, 1,0)
[1] 0.128
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "TEA", 1000, 1,0)
[1] 0.107
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "TEA", 1000, 1,0)
[1] 0.14
> |
```

Result  
P-value > .05



COFFEE

```
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "COFFEE", 1000, 1,0)
[1] 0.3
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "COFFEE", 1000, 1,0)
[1] 0.317
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "COFFEE", 1000, 1,0)
[1] 0.328
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "COFFEE", 1000, 1,0)
[1] 0.319
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "COFFEE", 1000, 1,0)
[1] 0.33
> |
```

Result  
P-value > .05

Neither tea nor coffee show correlations either.

# SODA AND COOKIES

## Permutation Test (x5)

### R Code

```
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "SODA", 1000, 1,0)
[1] 0.011
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "SODA", 1000, 1,0)
[1] 0.015
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "SODA", 1000, 1,0)
[1] 0.01
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "SODA", 1000, 1,0)
[1] 0.013
> PermutationTestSecond::Permutation(WALMART, "COOKIES", "SODA", 1000, 1,0)
[1] 0.007
> |
```

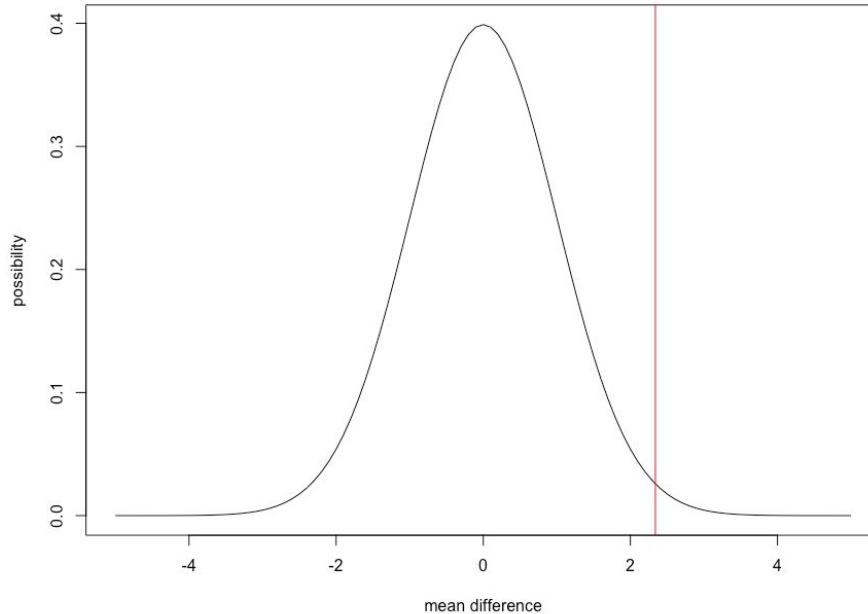
Result  
P-value > .05  
(shows association)



# Z-TEST: Soda and Cookies



SODA (PURCHASED COOKIES vs. NO COOKIES)



```
> YES.Cookies <- subset(WALMART, WALMART$COOKIES == 1)
> NO.Cookies <- subset(WALMART, WALMART$COOKIES == 0)
>
>
> YES.Cookies.Soda <- YES.Cookies$SODA
> NO.Cookies.Soda <- NO.Cookies$SODA
> sd.YES.Cookies.Soda <- sd(YES.Cookies.Soda)
> sd.NO.Cookies.Soda <- sd(NO.Cookies.Soda)
> mean.YES.Cookies.Soda <- mean(YES.Cookies.Soda)
> mean.NO.Cookies.Soda <- mean(NO.Cookies.Soda)
> len.YES.Cookies.Soda <- length(YES.Cookies.Soda)
> len.NO.Cookies.Soda <- length(NO.Cookies.Soda)
> sd.YES.NO.Cookies.Soda <- sqrt(sd.YES.Cookies.Soda^2/len.YES.Cookies.Soda + sd.NO.Cookies
.Soda^2/len.NO.Cookies.Soda)
>
> zeta <- (mean.YES.Cookies.Soda - mean.NO.Cookies.Soda)/sd.YES.NO.Cookies.Soda
> zeta
[1] 2.340048
> p = 1 - pnorm(zeta)
> p
[1] 0.009640636
>
```

Result

P-value = 0.009  
(shows association)



# Another Finding: Coffee and Sugar Association

## Permutation Test (x5):

```
> PermutationTestSecond::Permutation(dat, "COFFEE", "SUGAR", 1000, 1,0)
[1] 0.006
> PermutationTestSecond::Permutation(dat, "COFFEE", "SUGAR", 1000, 1,0)
[1] 0.007
> PermutationTestSecond::Permutation(dat, "COFFEE", "SUGAR", 1000, 1,0)
[1] 0.008
> PermutationTestSecond::Permutation(dat, "COFFEE", "SUGAR", 1000, 1,0)
[1] 0.008
> PermutationTestSecond::Permutation(dat, "COFFEE", "SUGAR", 1000, 1,0)
[1] 0.007
```

### Result

P-value ~ 0.008

*(shows association)*

## Z-TEST

```
> YES.Coffee.Sugar <- YES.Coffee$SUGAR
> NO.Coffee.Sugar <- NO.Coffee$SUGAR
> sd.YES.Coffee.Sugar <- sd(YES.Coffee.Sugar)
> sd.NO.Coffee.Sugar <- sd(NO.Coffee.Sugar)
> mean.YES.Coffee.Sugar <- mean(YES.Coffee.Sugar)
> mean.NO.Coffee.Sugar <- mean(NO.Coffee.Sugar)
> len_YES.Coffee.Sugar <- length(YES.Coffee.Sugar)
> len_NO.Coffee.Sugar <- length(NO.Coffee.Sugar)
> sd.YES.NO.Coffee.Sugar <- sqrt(sd.YES.Coffee.Sugar^2/len_YES.Coffee.
Sugar + sd.NO.Coffee.Sugar^2/len_NO.Coffee.Sugar)
>
> zeta <- (mean.YES.Coffee.Sugar - mean.NO.Coffee.Sugar)/sd.YES.NO.Coffee.
Sugar
> zeta
[1] -2.397572
>
> plot(x=seq(from = -5, to= 5, by=0.1),y=dnorm(seq(from = -5, to= 5, by=0
.1),mean=0),type='l',xlab = 'mean difference', ylab='possibility',
main="SUGAR PURCHASED (PURCHASED COFFEE vs. NO PURCHASED COFFEE)")
> abline(v=zeta, col='red')
> p <- pnorm(zeta)
> p
[1] 0.008252064
```



# COOKIES AND SODA



By Matthew Lau

---