

Alcohol EDU Statistical Significance

Julian DeGroot-Lutzner

12/7/2017

```
require(ggplot2)

# Abstainer (A)
a_prob_before = 0.36
a_prob_after = 0.34
nat_change_a = -0.06
pz_change_a = -0.13

# Nondrinker (N)
n_prob_before = 0.28
n_prob_after = 0.22
nat_change_n = -0.21
pz_change_n = -0.46

# Moderate drinker (M)
m_prob_before = 0.19
m_prob_after = 0.21
nat_change_m = 0.11
pz_change_m = 0.08

# Heavy Episodic Drinker (H)
h_prob_before = 0.13
h_prob_after = 0.17
nat_change_h = 0.31
pz_change_h = 0.43

# Problematic Drinker (P)
p_prob_before = 0.05
p_prob_after = 0.07
nat_change_p = 0.40
pz_change_p = 2.50

drinking_types = c("A", "N", "M", "H", "P")
probs_before = c(a_prob_before, n_prob_before, m_prob_before,
                 h_prob_before, p_prob_before)
probs_after = c(a_prob_after, n_prob_after, m_prob_after,
               h_prob_after, p_prob_after)

pz_sample_size_before = 175
pz_sample_size_after = 173

results <- data.frame()
p_values <- c(0,0,0,0,0)
num_simulations = 50000
```

```

set.seed(4747)
for(s in 1:num_simulations){
  # random sample of drinking types with prob from national averages
  rd_samp_before <- sample(drinking_types,
                           pz_sample_size_before,
                           replace = TRUE,
                           prob = probs_before)

  # random sample of drinking after with prob from national averages
  rd_samp_after <- sample(drinking_types,
                          pz_sample_size_after,
                          replace = TRUE,
                          prob = probs_after)

  # The chart measures the percent change in the number of drinking types before
  # survey 3 to the same number after survey 3.
  # Mathematically, this is (after/before - 1)
  a_change = sum(rd_samp_after == "A")/sum(rd_samp_before == "A") - 1
  n_change = sum(rd_samp_after == "N")/sum(rd_samp_before == "N") - 1
  m_change = sum(rd_samp_after == "M")/sum(rd_samp_before == "M") - 1
  h_change = sum(rd_samp_after == "H")/sum(rd_samp_before == "H") - 1
  p_change = sum(rd_samp_after == "P")/sum(rd_samp_before == "P") - 1

  # keep track of changes each simulation
  results <- rbind(results, c(a_change, n_change, m_change, h_change, p_change))

  # keep track of the number of times simulated change is bigger/smaller
  # (respectively) then the PZ test changes
  p_values <- p_values + c(a_change <= pz_change_a,
                           n_change <= pz_change_n,
                           m_change <= pz_change_m,
                           h_change >= pz_change_h,
                           p_change >= pz_change_p)

}
names(results) <- c("Change.Abstainers", "Change.Nondrinkers",
                   "Change.Moderate.Drinkers", "Change.Heavy.Episodic.Drinkers",
                   "Change.Problematic.Drinkers")
p_values <- data.frame(p_values/num_simulations)
p_values <- cbind( p_values, Type.Drinker = c("Abstainers", "Nondrinkers",
                                              "Moderate.Drinkers", "Heavy.Episodic.Drinkers",
                                              "Problematic.Drinkers" ))
p_values

##   p_values.num_simulations      Type.Drinker
## 1                0.31588      Abstainers
## 2                0.03340      Nondrinkers
## 3                0.47274 Moderate.Drinkers
## 4                0.35036 Heavy.Episodic.Drinkers
## 5                0.03464 Problematic.Drinkers

ggplot(results, aes(x = Change.Abstainers)) +
  geom_histogram(binwidth=.1, colour="black", fill="white") +
  ggtitle("Distribution of Abstainer 'Change' Value",

```

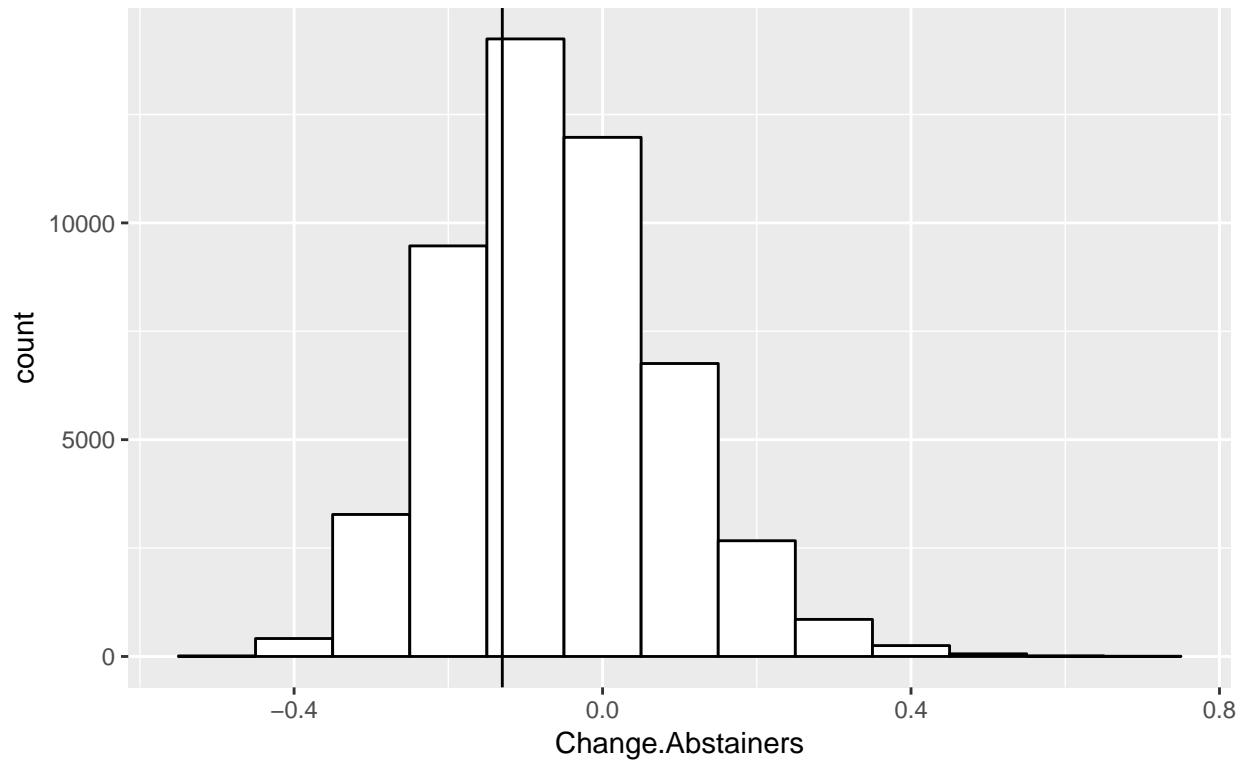
```

    subtitle = "P Value = 0.31588 ") +
  geom_vline(xintercept = pz_change_a)

```

Distribution of Abstainer 'Change' Value

P Value = 0.31588



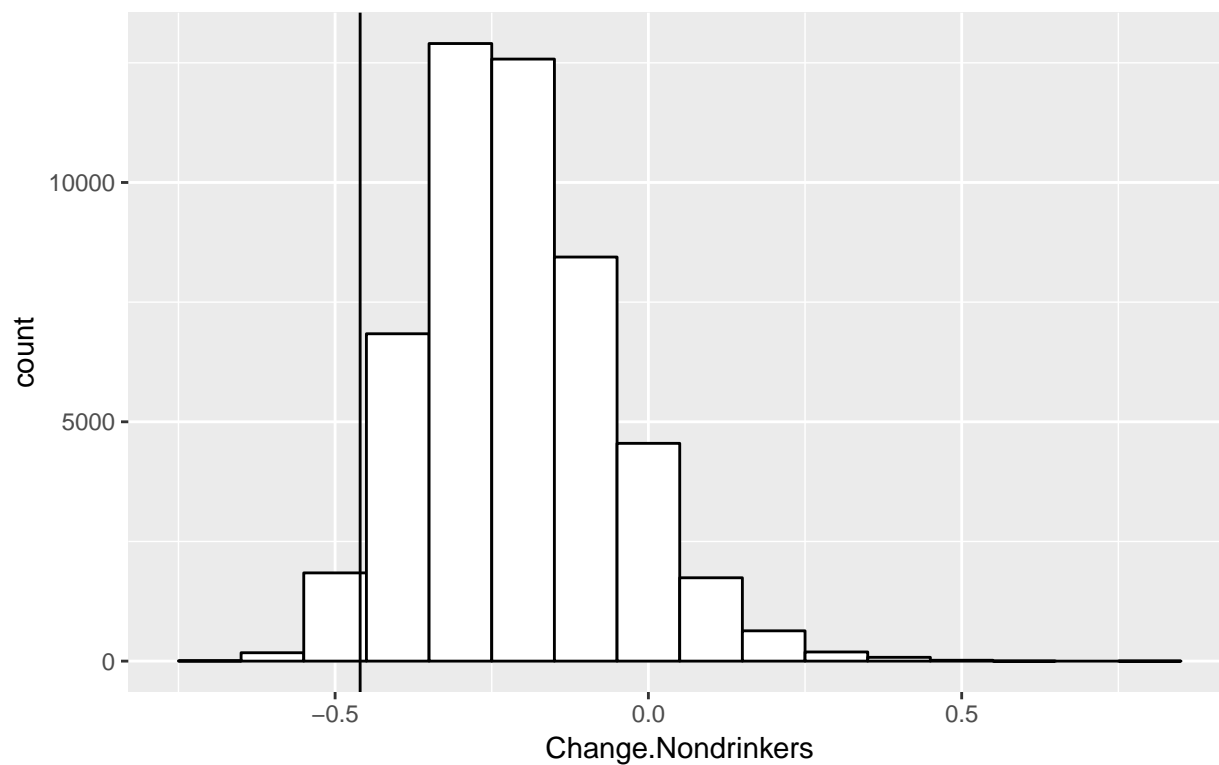
```

ggplot(results, aes(x = Change.Nondrinkers)) +
  geom_histogram(binwidth=.1, colour="black", fill="white") +
  ggtitle("Distribution of Nondrinkers 'Change' Value",
    subtitle = "P Value = 0.03340*") +
  geom_vline(xintercept = pz_change_n)

```

Distribution of Nondrinkers 'Change' Value

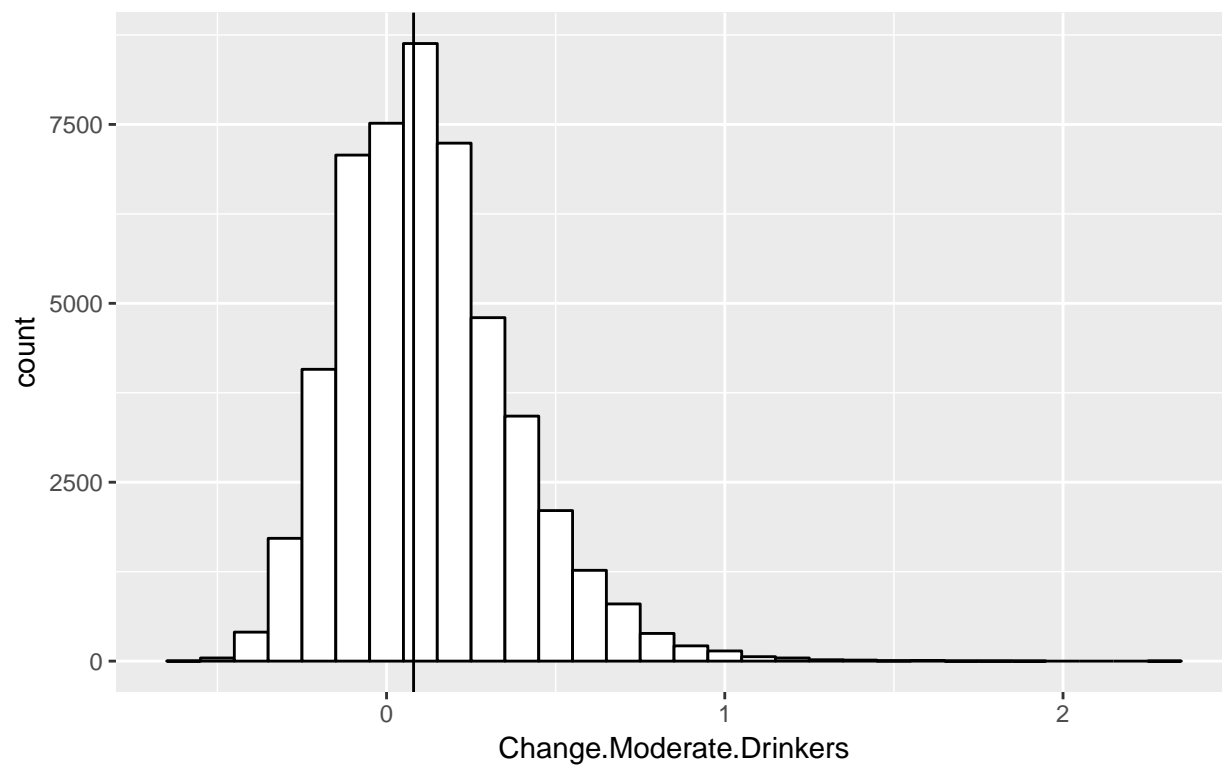
P Value = 0.03340*



```
ggplot(results, aes(x = Change.Moderate.Drinkers)) +  
  geom_histogram(binwidth=.1, colour="black", fill="white") +  
  ggtitle("Distribution of Moderate Drinkers 'Change' Value",  
    subtitle = "P Value = 0.47274") +  
  geom_vline(xintercept = pz_change_m)
```

Distribution of Moderate Drinkers 'Change' Value

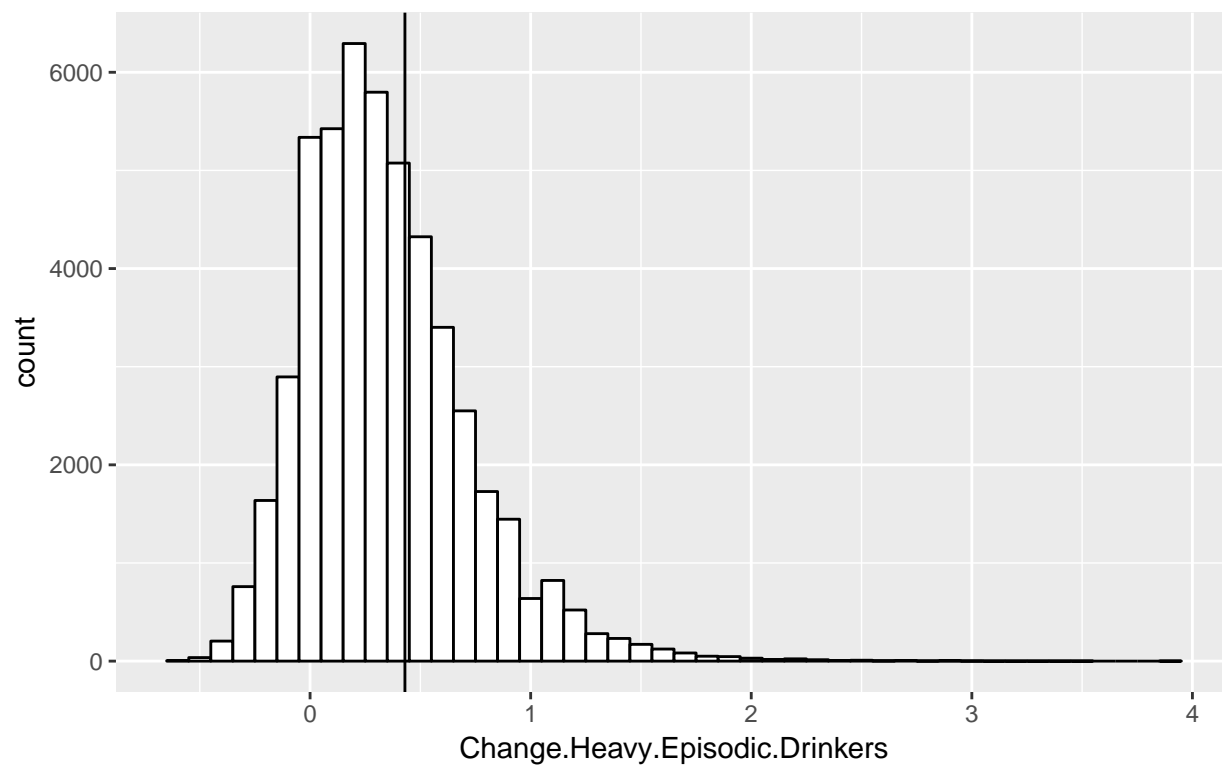
P Value = 0.47274



```
ggplot(results, aes(x = Change.Heavy.Episodic.Drinkers)) +  
  geom_histogram(binwidth=.1, colour="black", fill="white") +  
  ggtitle("Distribution of Heavy Episodic Drinkers 'Change' Value",  
    subtitle = "P Value = 0.35036") +  
  geom_vline(xintercept =pz_change_h)
```

Distribution of Heavy Episodic Drinkers 'Change' Value

P Value = 0.35036



```
ggplot(results, aes(x = Change.Problematic.Drinkers)) +  
  geom_histogram(binwidth=.1, colour="black", fill="white") +  
  ggtitle("Distribution of Problematic Drinkers 'Change' Value",  
    subtitle = "P Value = 0.03464*") +  
  geom_vline(xintercept =pz_change_p)
```

Distribution of Problematic Drinkers 'Change' Value

P Value = 0.03464*

