

# Hypothesis Testing

April Sang

15/05/2019

```
sensors <- read.csv('sensor.csv', header = T, sep = ",")
#head(sensors, 5)

# Create a dataset filters only those observations where value is between 0 and 100 inclusive
sensors_clean <- subset(sensors, as.numeric(value) >=0 & as.numeric(value) <=100)

# Convert column to numeric
sensors_clean$value <- as.numeric(sensors_clean$value)

# T-test to check if there is a difference between site 1 and site 2
# make a dataframe with just siteid = 1 and 2
new_data <- subset(sensors_clean, siteid==1 | siteid==2)
#head(new_data, 15)

# Null Hypothesis: mean sensor value for site 1 = mean sensor value for site 2
# Alternative Hypothesis: mean sensor value for site 1 != mean sensor value for site 2
t.test(value~siteid, data = new_data)
```

```
##
##  Welch Two Sample t-test
##
## data:  value by siteid
## t = 0.014532, df = 280.42, p-value = 0.9884
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.071862  4.132430
## sample estimates:
## mean in group 1 mean in group 2
##      68.94245      68.91216
```

```
# FAIL TO REJECT NULL HYPOTHESIS.
# 0 is contained in the interval and p-value is 0.9346 which is greater than 0.05.
# There is no strong enough evidence suggesting that there is a difference between the mean sensor v
alue of sites 1 and 2.
```