

# Permutation Test

Clare Cruz

4/13/2021

```
# Now we create two groups according to the group labels in the dataset and calculate a t-statistic
# on the original group labels
# set.seed(114)
p.vector1 <- NULL
p.vector2 <- NULL
t.vector <- NULL

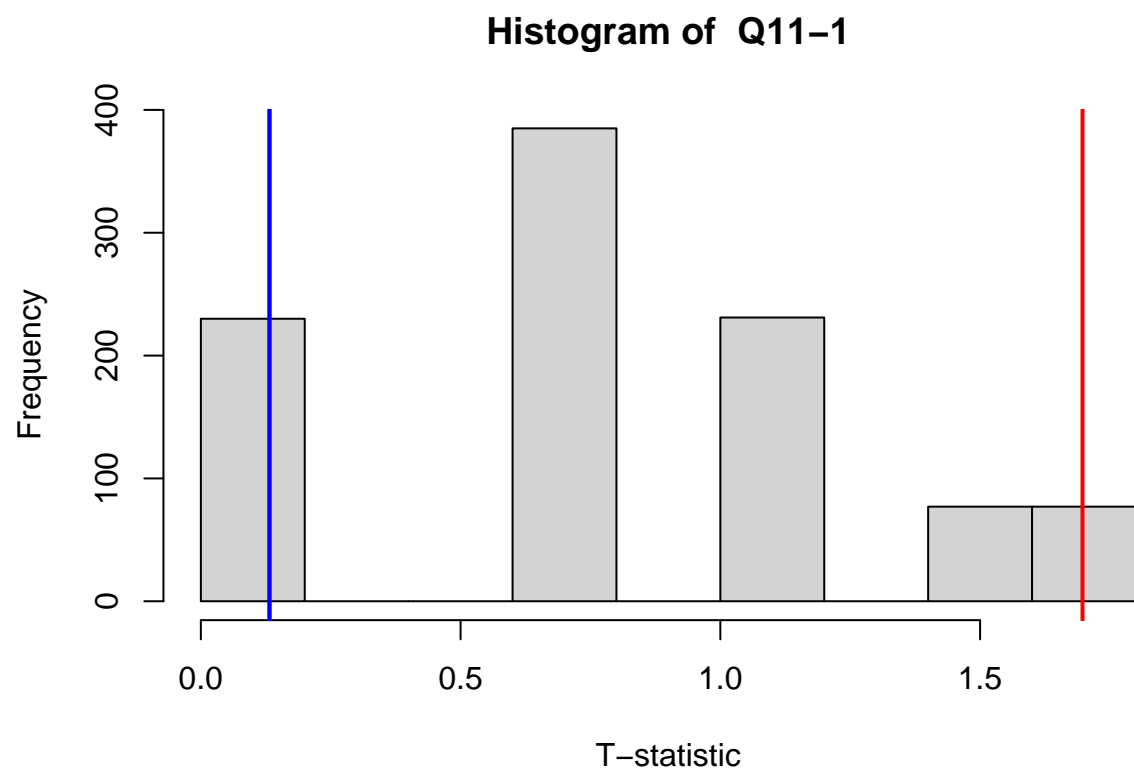
for(j in 2:17)
{
  Group1 <- df[df$Group == 'Student',j]
  Group2 <- df[df$Group == 'Supervisor',j]
  t0 <- abs(t.test(Group1,Group2)$statistic)
  t.vector <- c(t.vector,t0)

# Now we'll do 1000 permutations of the group labels and calculate a new t-statistic each time:
  nperm <- 1000
  t.perm <- rep(0,nperm)

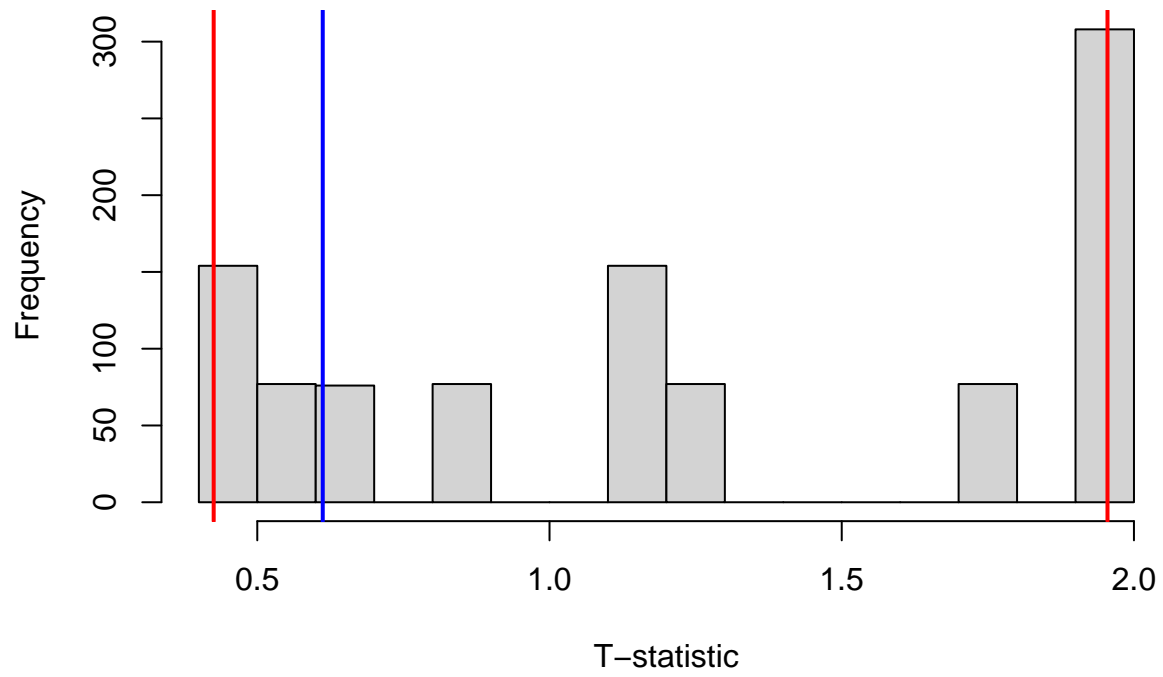
  for (i in 1:nperm) {
    set.seed(141)
    ind <- sample(df$Group)
    df$Group <- ind
    Group1 <- df[df$Group == 'Student',j]
    Group2 <- df[df$Group == 'Supervisor',j]
    t.perm[i] <- abs(t.test(Group1,Group2)$statistic)
  }

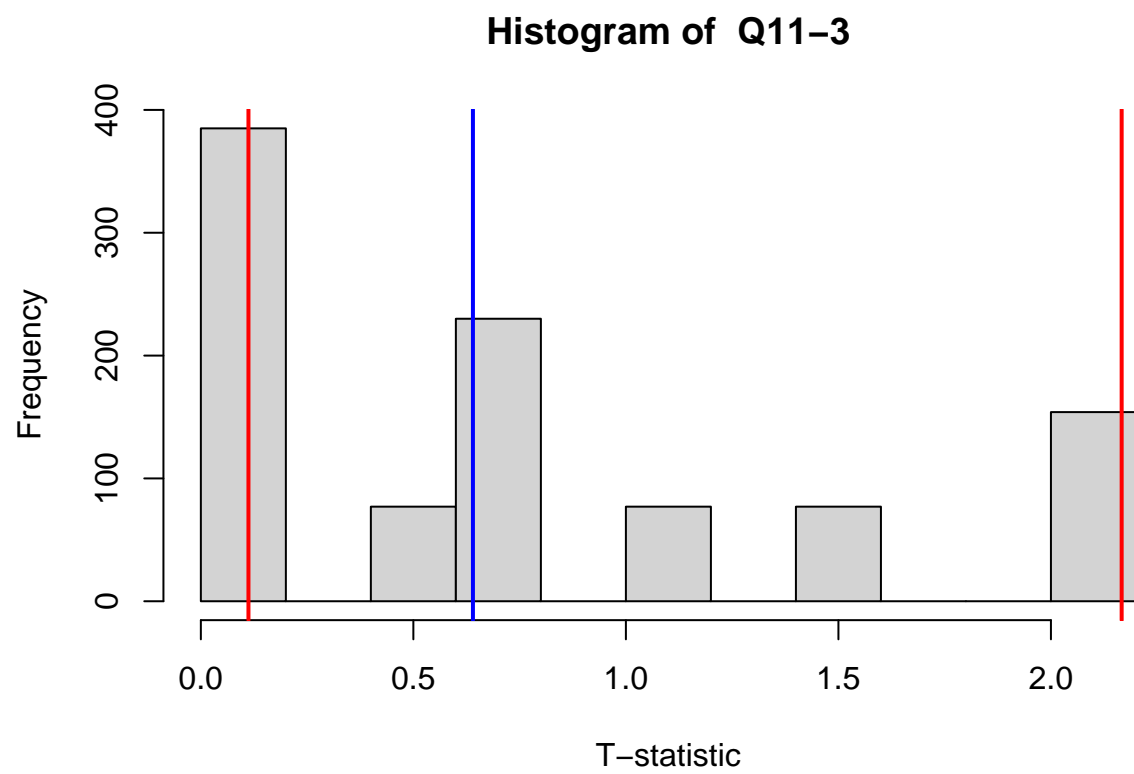
# Now let's make a histogram of our permuted t-statistics and see where
# our original t-statistic falls:
hist(t.perm, main = paste("Histogram of ", colnames(df[j])), xlab = "T-statistic")
abline(v=quantile(t.perm, 0.05),col='red',lwd=2)
abline(v=quantile(t.perm, 0.95),col='red',lwd=2)
abline(v=t0,col='blue',lwd=2)

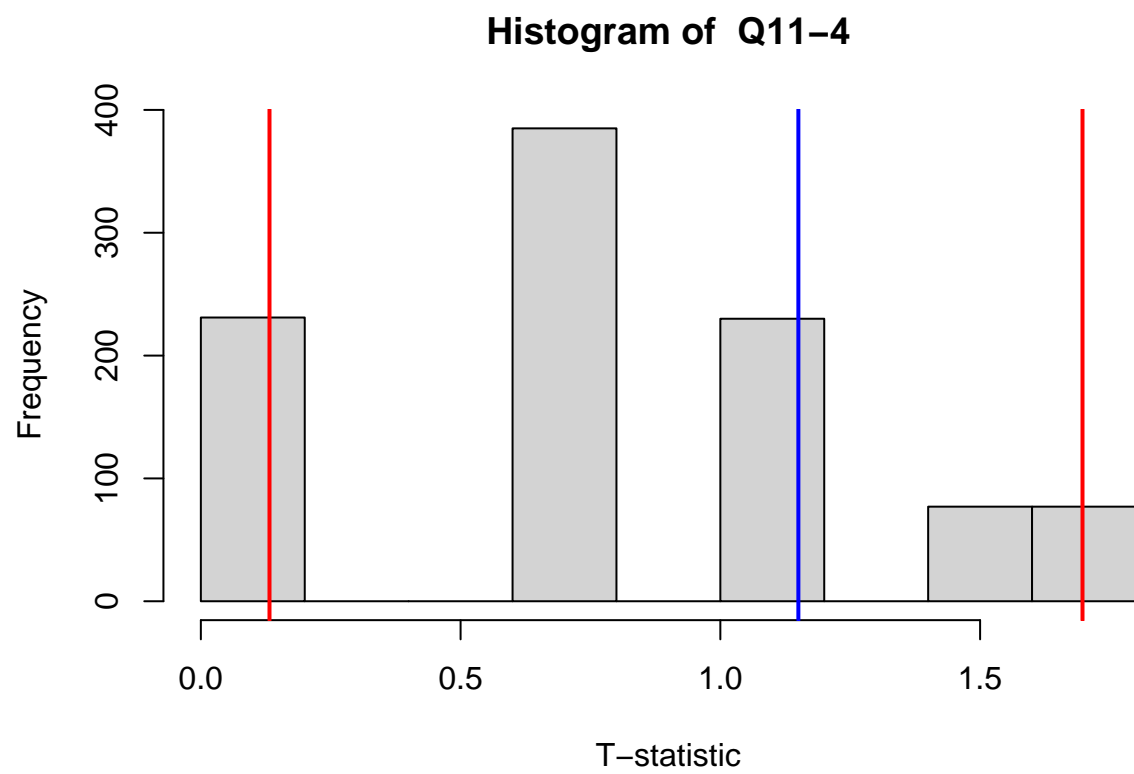
# You should see that it looks like it could have plausibly come from this distribution, so we
# probably not reject the null hypothesis. Let's be sure by calculating an explicit p-value.
# Here, the p-value is just the percentage of permutation statistics that fell above (were more
# extreme than) our original t-statistic.
p.vector1 <- c(p.vector1, mean(t.perm >= t0))
}
```



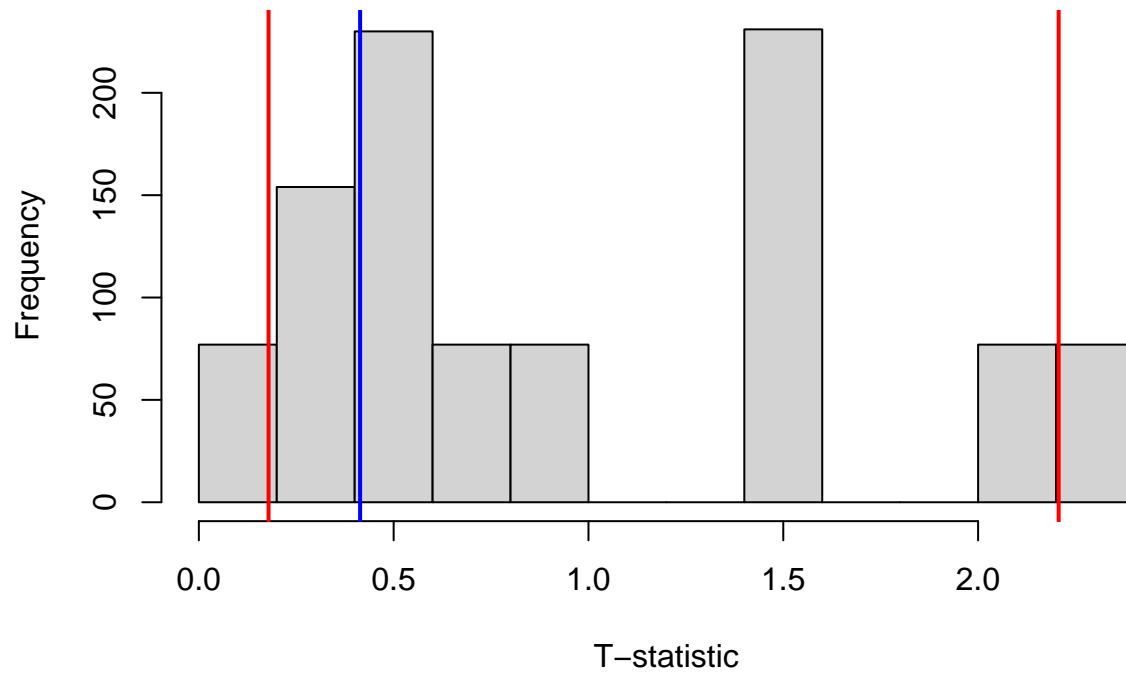
**Histogram of Q11-2**



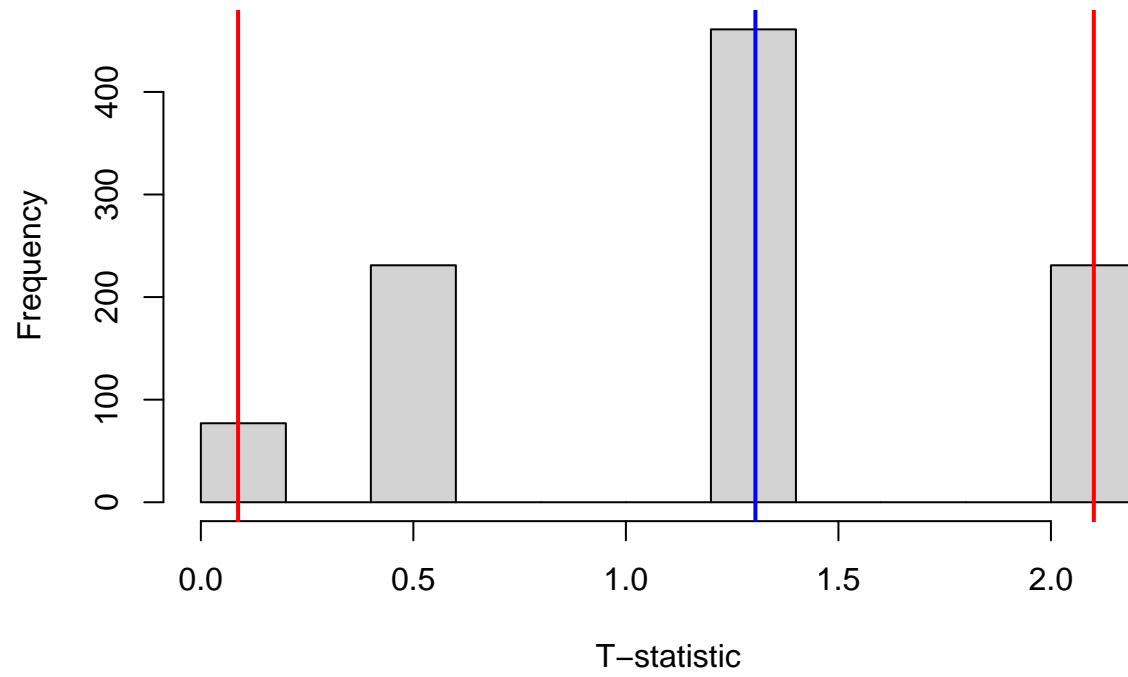




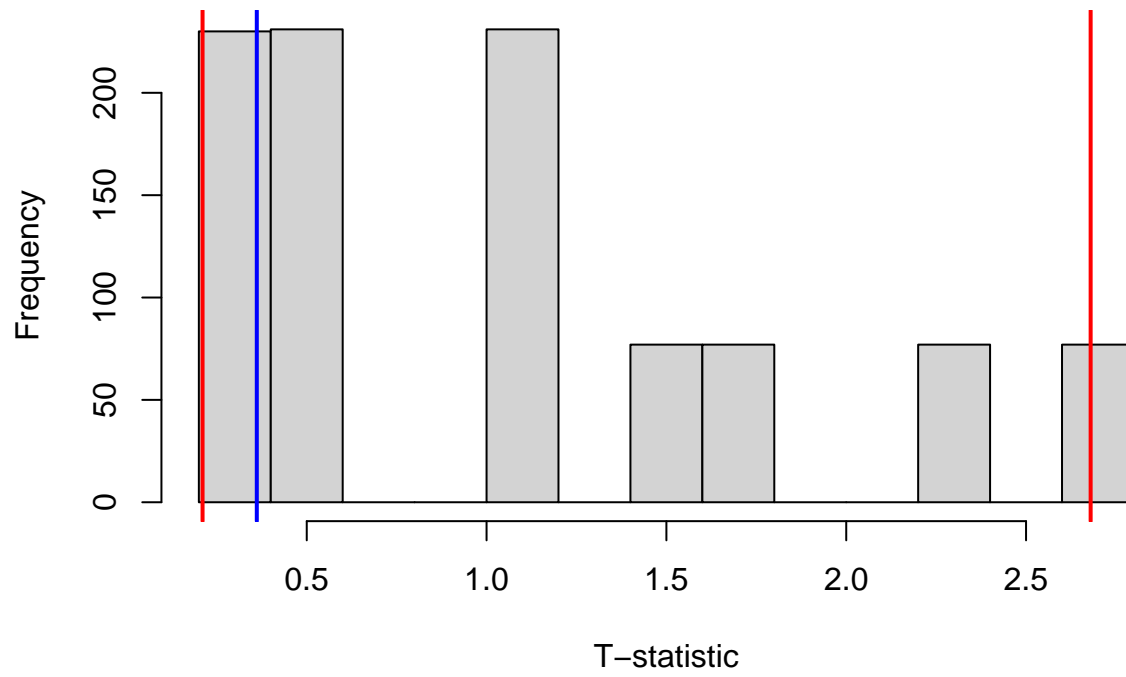
**Histogram of Q11-5**



**Histogram of Q11-6**

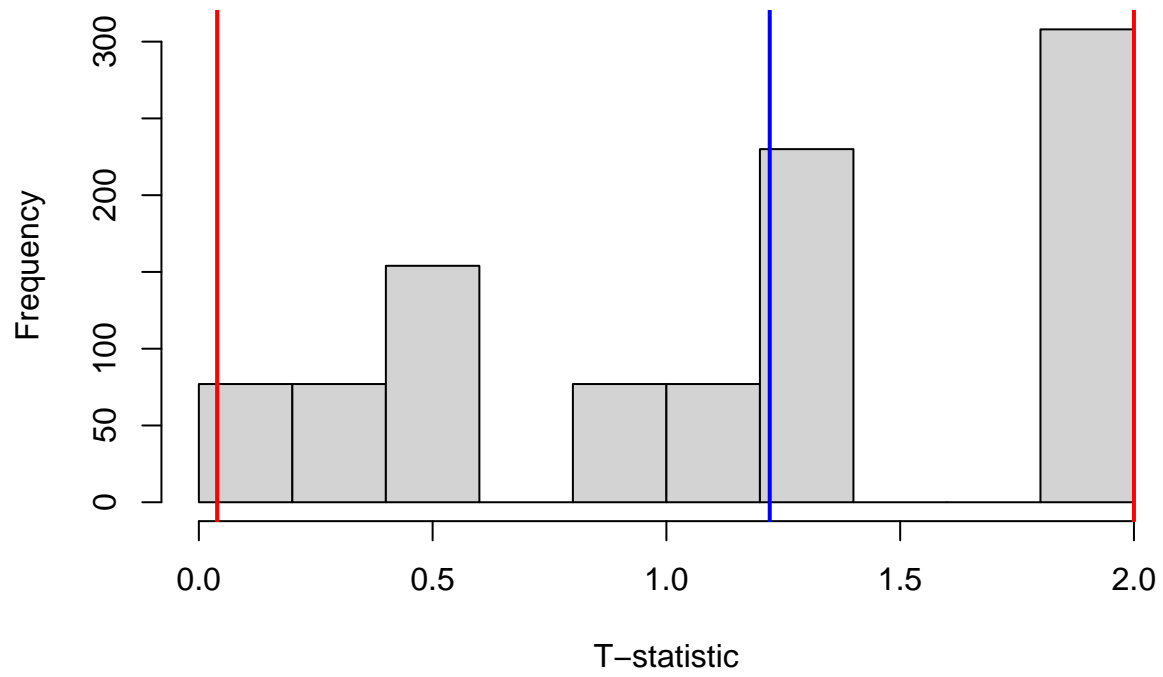


**Histogram of Q11-7**

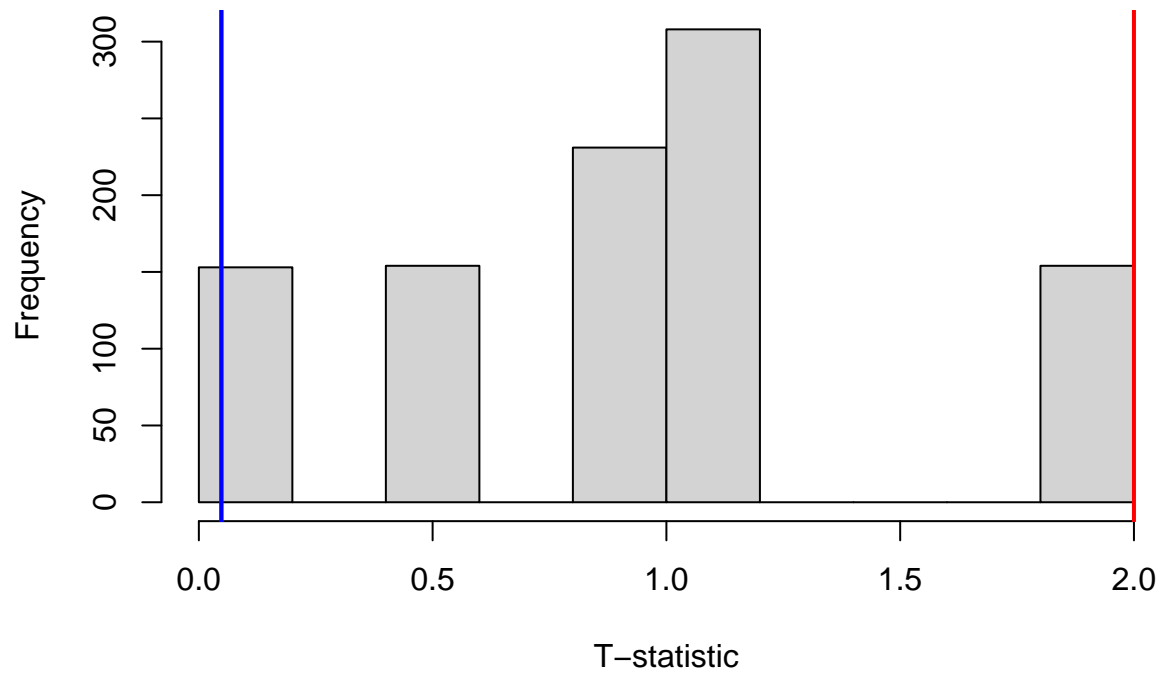




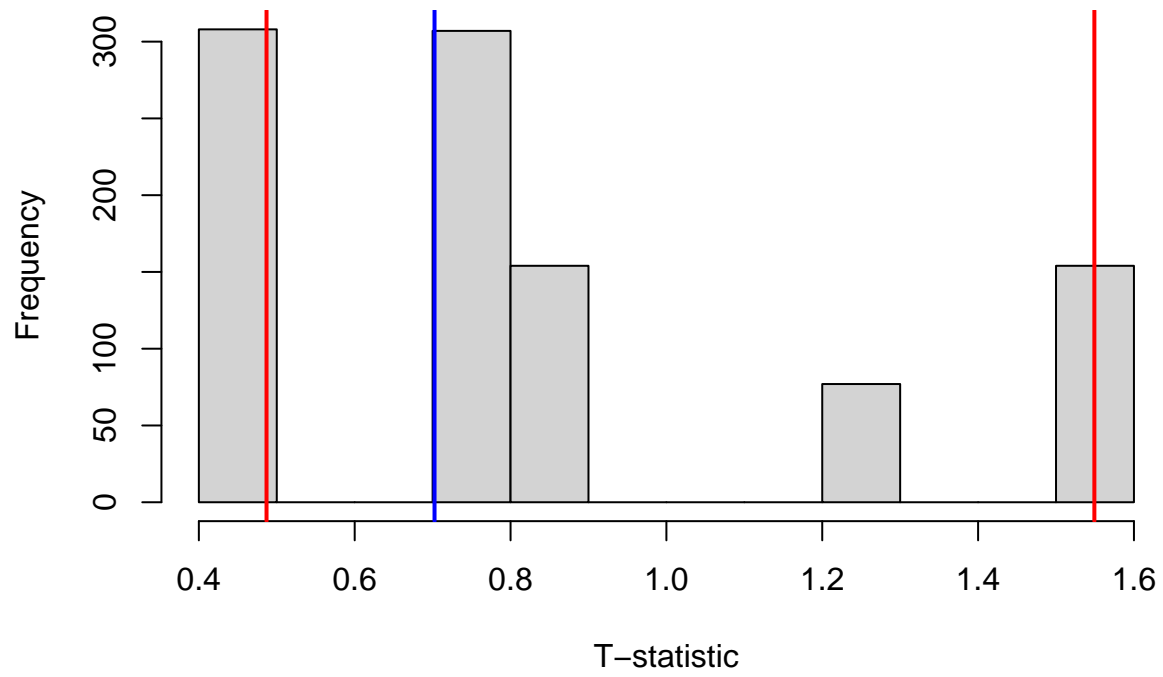
**Histogram of Q11-8**



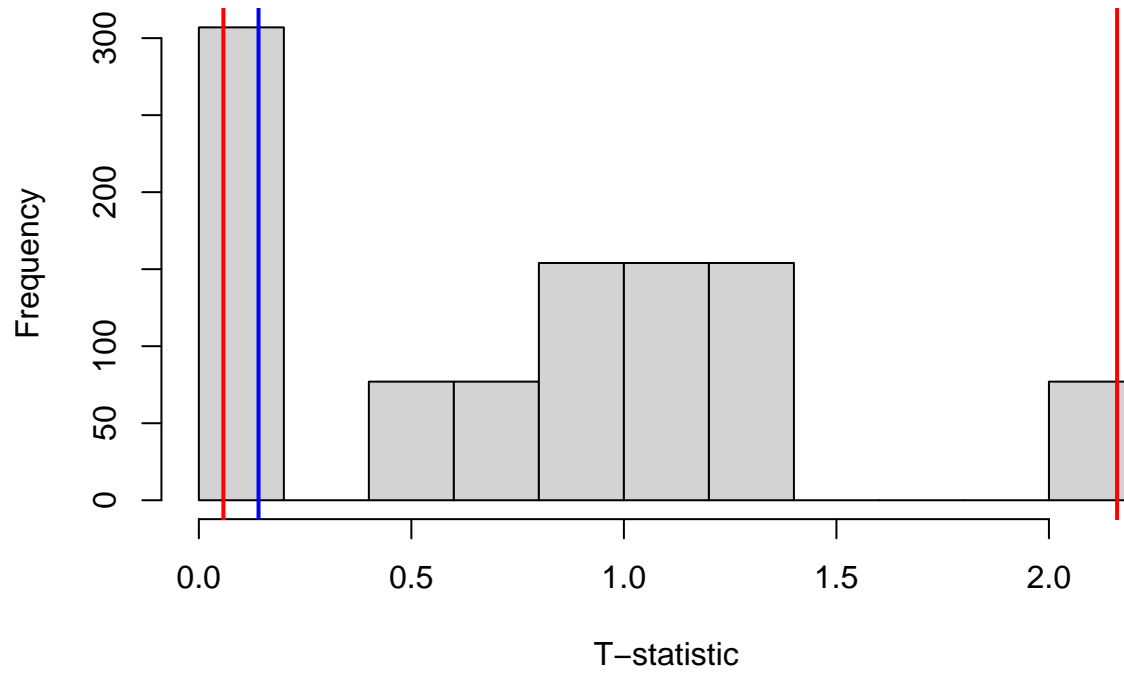
**Histogram of Q14-1**



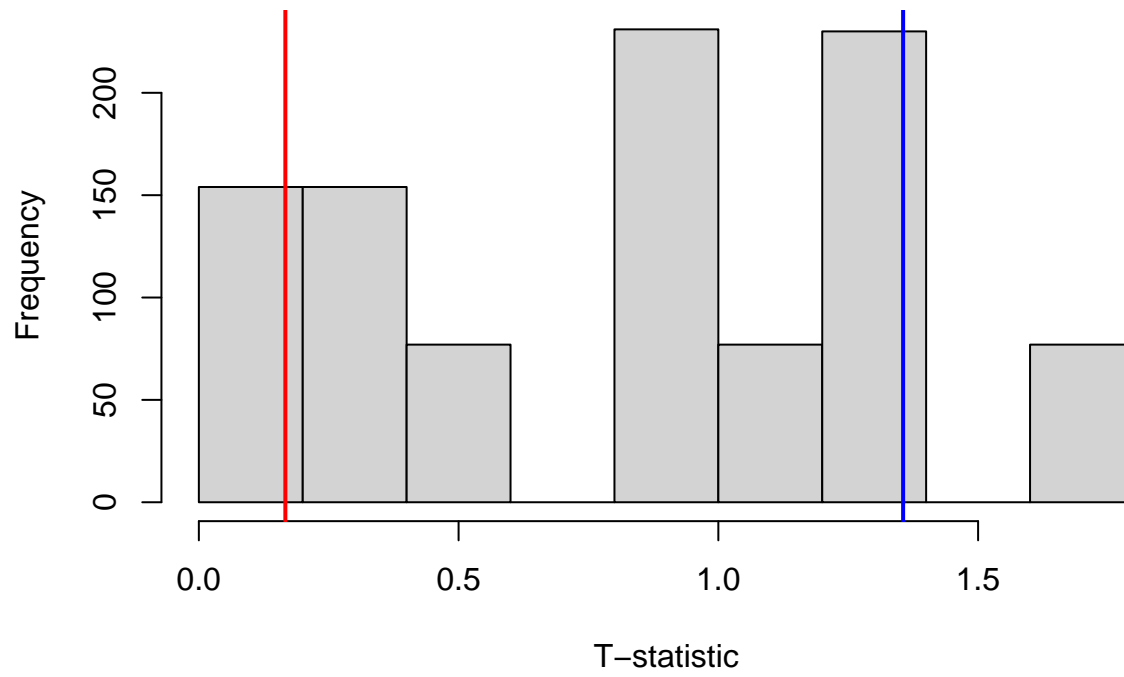
**Histogram of Q14-2**



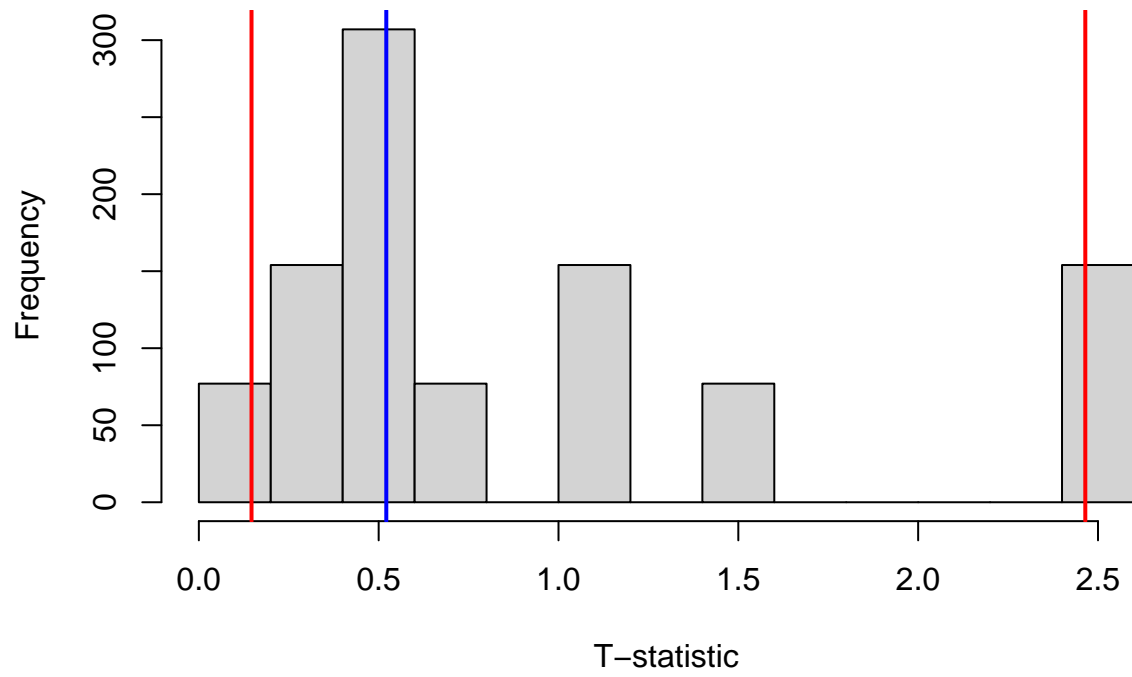
**Histogram of Q14-3**



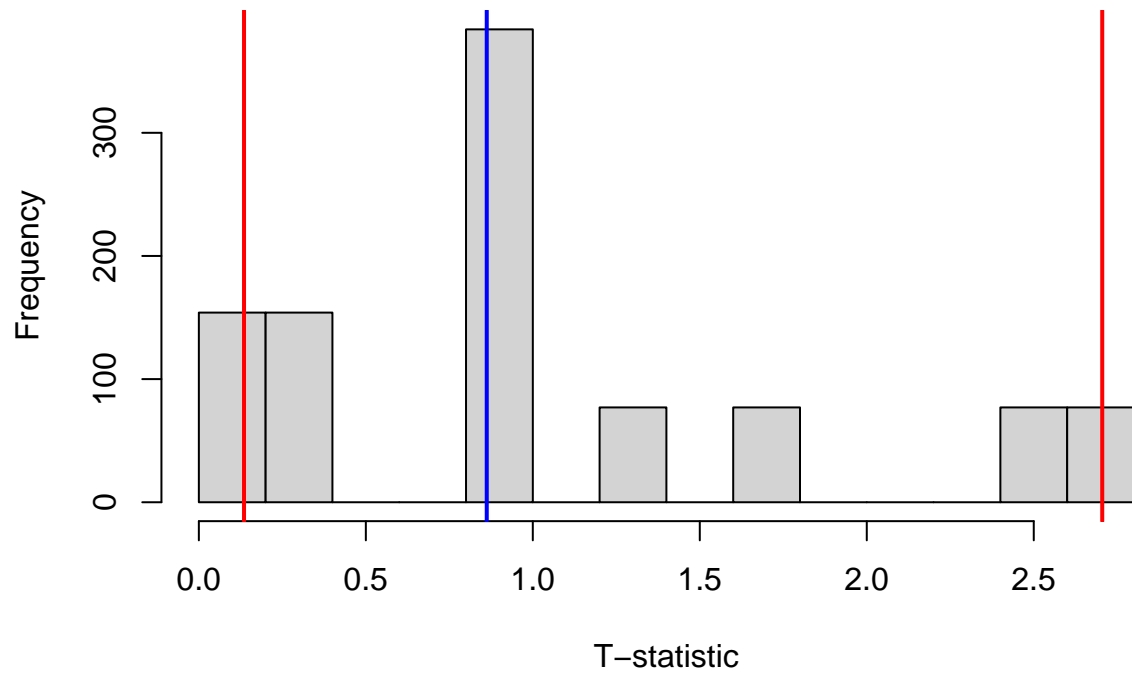
**Histogram of Q14-4**



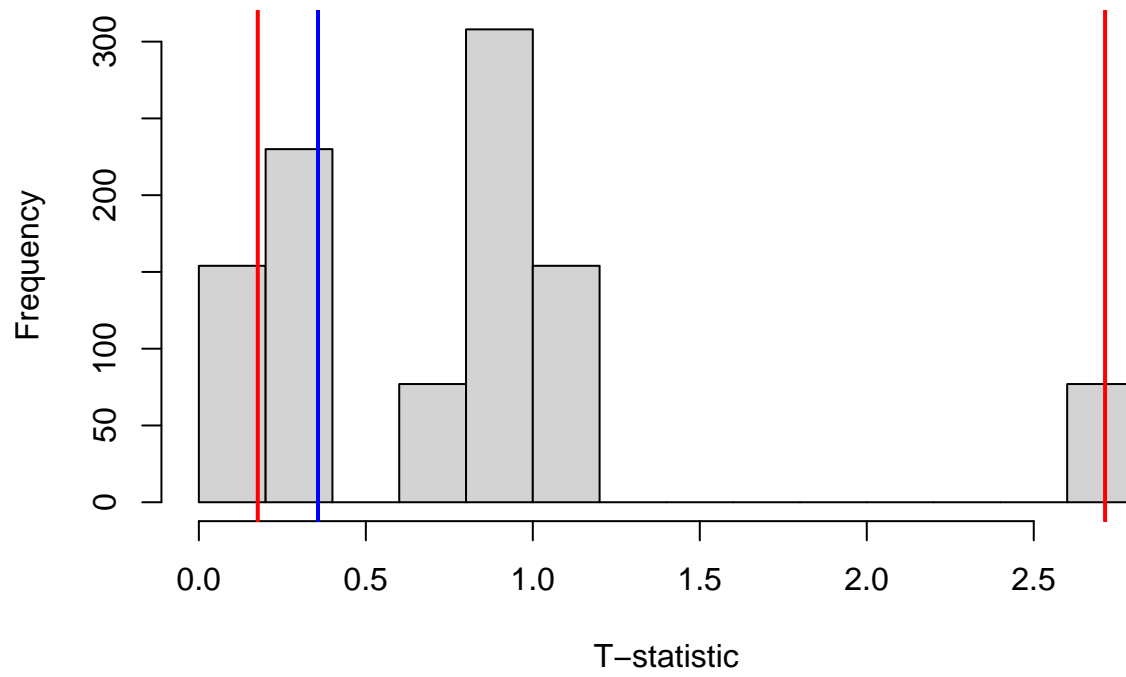
**Histogram of Q14-5**



**Histogram of Q14-6**

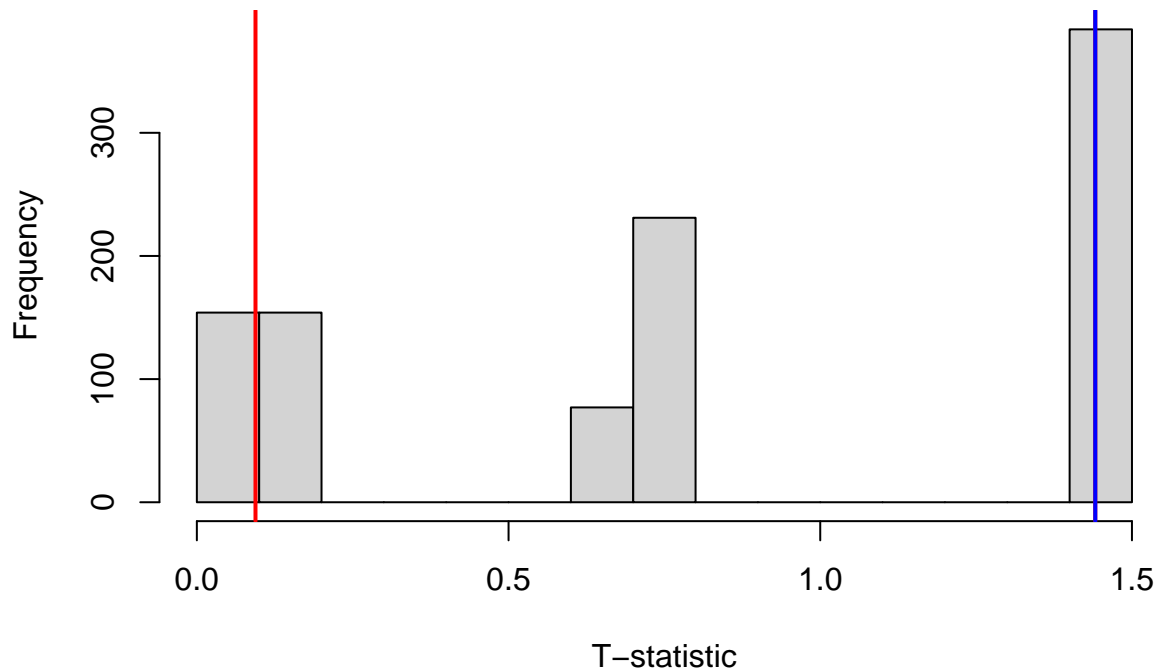


**Histogram of Q14-7**





## Histogram of Q14-8



```
p.vector1
```

```
## [1] 1.000 0.769 0.538 0.384 0.769 0.461 0.846 0.538 1.000 0.692 0.846 0.307
## [13] 0.615 0.461 0.846 0.153
```

```
t.vector
```

```
##          t          t          t          t          t          t          t          t          t
## 0.1321637 0.6119990 0.6401436 1.1501723 0.4138915 1.3047422 0.3611576 1.2209822
##          t          t          t          t          t          t          t          t          t
## 0.0482056 0.7024394 0.1404219 1.3556615 0.5212034 0.8620437 0.3567530 1.4411534
```

```
# You should see that the p-value is larger than 0.05, so we cannot reject the null hypothesis that
# the group means are equal
```

```
# Calculating the averages for each group, for each survey question
df2 <- df %>% group_by(Group) %>%
  summarize((across(everything(), list(mean), na.rm = TRUE)))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```
df2 <- as.data.frame(df2)
df2
```

```
##      Group Q11-1_1 Q11-2_1 Q11-3_1 Q11-4_1 Q11-5_1 Q11-6_1 Q11-7_1
## 1 Student 3.666667 3.666667 3.777778 3.666667 3.222222 3.777778 3.555556
## 2 Supervisor 2.800000 3.000000 2.600000 2.800000 2.200000 2.800000 2.200000
## Q11-8_1 Q14-1_1 Q14-2_1 Q14-3_1 Q14-4_1 Q14-5_1 Q14-6_1 Q14-7_1
## 1 3.5 3.625000 3.875000 3.500000 3.375000 3.625 3.375000 3.500000
## 2 3.2 2.333333 2.333333 2.666667 2.333333 2.000 2.333333 2.333333
## Q14-8_1
## 1 3.750000
## 2 2.666667
```

```
#v1 <- as.vector(df2[2,2:17])
#v2 <- as.vector(df2[1,2:17])
#v3 <- v1-v2
#v3$Group = 'Diff'
#p.vector1$Group = 'P-value'
#t.vector$Group = 'T-stat'
#final <- rbind(df2,v3,p.vector1,t.vector)
# write.csv(final, "C:\\Users\\cbrig\\OneDrive - University of Pittsburgh\\Mom\\Permutations Experiment
```