STA4211 Question 3

Josh Cha

2025-09-09

```r
group_A <- c(14, 18, 11, 13, 18, 17, 21, 9, 16, 17, 14, 15)  # Sleep (Group 1)
group_B <- c(12, 12, 14, 13, 6, 18, 14, 16, 10, 7, 15, 10)   # Caffeine (Group 2)

# Combine data
data <- c(group_A, group_B)
n_A <- length(group_A)
n_B <- length(group_B)
```

(a)

```r
# Observed difference in means
obs_diff <- mean(group_A) - mean(group_B)
cat("Observed Difference in Means:", obs_diff, "\n")
```

```
## Observed Difference in Means: 3
```

(b)

```r
# Null hypothesis : No difference in mean
# Alternative hypothesis : Two-sided difference in mean
```

(c)

```r
# Randomization test
n_perm <- 1000
perm_diffs <- numeric(n_perm)
set.seed(347)

for (i in 1:n_perm) {
  shuffled <- sample(data)
  perm_group_A <- shuffled[1:n_A]
  perm_group_B <- shuffled[(n_A + 1):(n_A + n_B)]
  perm_diffs[i] <- mean(perm_group_A) - mean(perm_group_B)
}
```

(d)

```r
# One simulation run (seed = 347, 1,000 permutations)

p_value <- mean(abs(perm_diffs) >= abs(obs_diff))
cat("Randomization test p-value", p_value, "\n")
```

```
## Randomization test p-value 0.048
```

```r
cat("This is not surprising because it seems likely that we will reject the null hypothesis")
```

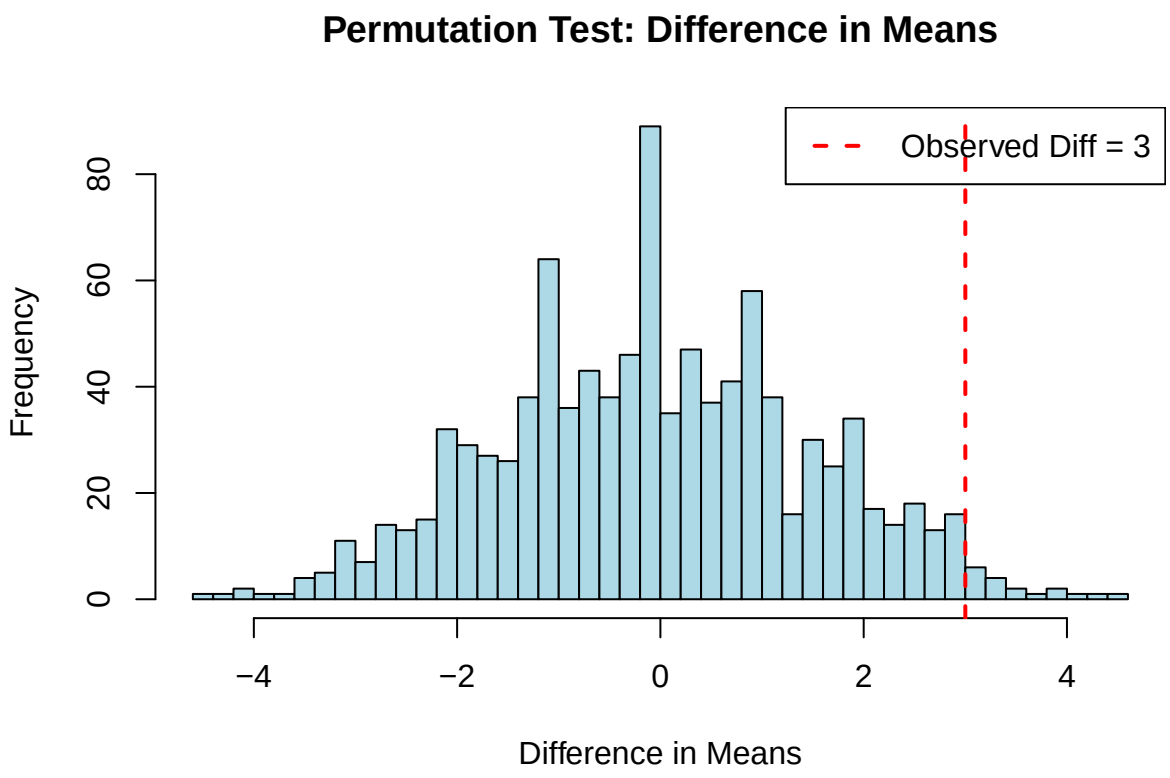
```
## This is not surprising because it seems likely that we will reject the null hypothesis
```

```r
cat("Seed value used: 347\n")
```

```
## Seed value used: 347
```

```r
# Histogram
hist(perm_diffs, breaks = 50, col = "lightblue",
     main = "Permutation Test: Difference in Means",
     xlab = "Difference in Means")

abline(v = obs_diff, col = "red", lwd = 2, lty = 2)
legend("topright", legend = paste("Observed Diff =", round(obs_diff, 3)),
col = "red", lty = 2, lwd = 2)
```



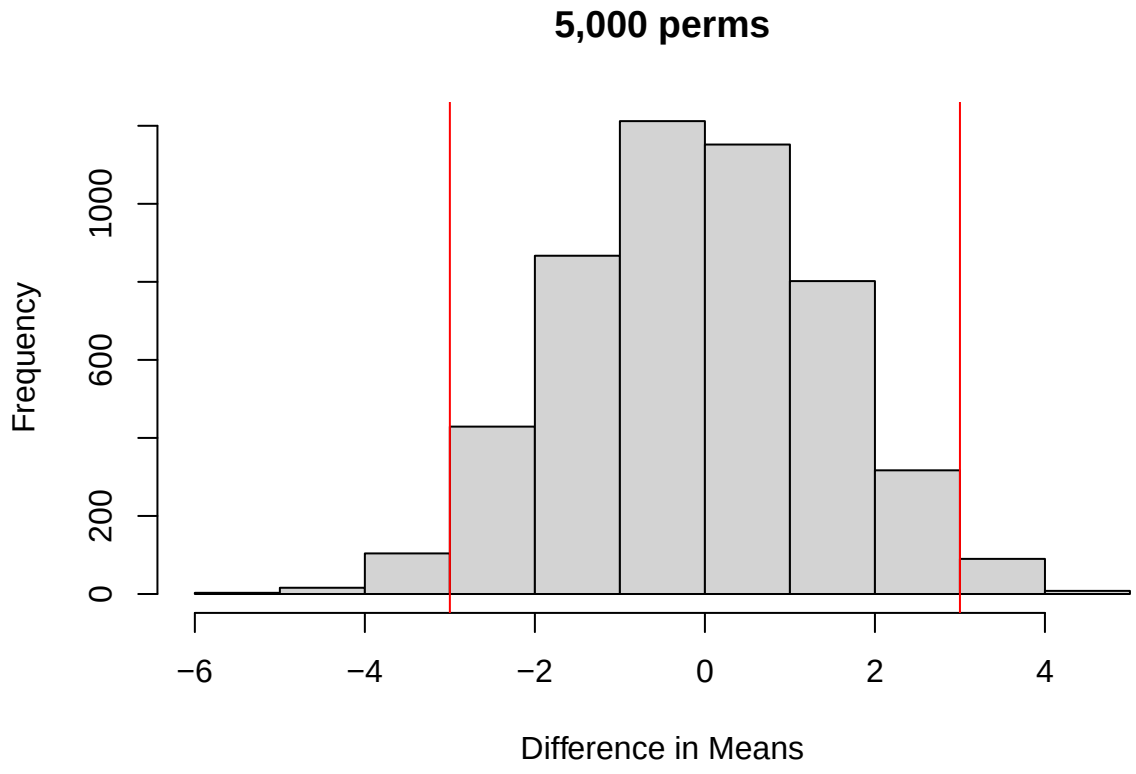
(e)

```
# 5,000 permutations, seed=913
n_perm <- 5000; set.seed(913)
perm_diffs <- replicate(n_perm, {
  s <- sample(data)
  mean(s[1:n_A]) - mean(s[(n_A+1):(n_A+n_B)])
})

cat("Center:", mean(perm_diffs),
    " SD:", sd(perm_diffs),
    " Shape: bell-shaped around 0\n")
```

```
## Center: -0.02406667  SD: 1.520956  Shape: bell-shaped around 0
```

```
hist(perm_diffs,
     main="5,000 perms",
     xlab="Difference in Means")
abline(v=c(-obs_diff,obs_diff), col="red")
```



(f)

```r
p_value <- mean(abs(perm_diffs) >= abs(obs_diff))
cat("(f) P-value (5,000 perms):", p_value, "\n")
```

```
## (f) P-value (5,000 perms): 0.0494
```

```r
# At the 5% level, we reject the null hypothesis but not so convincingly
```

(g)

```r
t_test_result <- t.test(group_A, group_B,
                        alternative = "two.sided",
                        var.equal = FALSE)

p_value
```

```
## [1] 0.0494
```

```r
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  group_A and group_B
## t = 2.1438, df = 21.894, p-value = 0.04342
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.09699633 5.90300367
## sample estimates:
## mean of x mean of y
##     15.25     12.25
```

```r
# Both p-values are very close, so can make the same conclusion
```