Permutation Tests and ANOVA

Ethan Tsao

2024-10-16

Part 1: Two sample permutation test

a) Permutation test function

```
perm_test <- function(data, reps){

  perm_t <- rep(NA, reps)

  names(data) <- c("y", "x")

  for(i in 1:reps){
    permdata <- data
    permdata$y <- sample(permdata$y)
    perm_t[i] <- t.test(y ~ x,
                        data=permdata)$statistic
  }

  t0 <- t.test(y ~ x, data=data)$statistic

  p_value <- (sum(perm_t >= abs(t0)) + sum(perm_t <= -abs(t0)))/reps

  return(p_value)
}
```

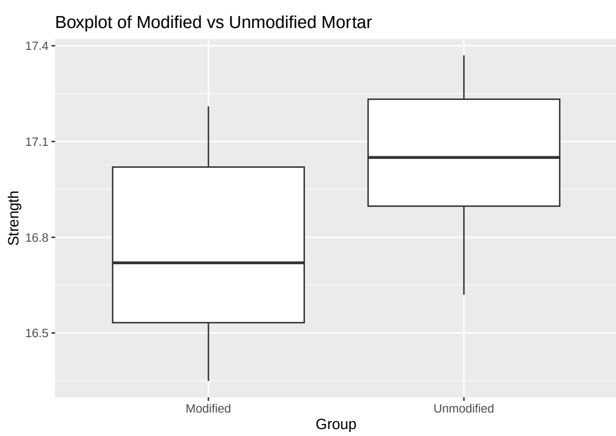
b) Testing your function

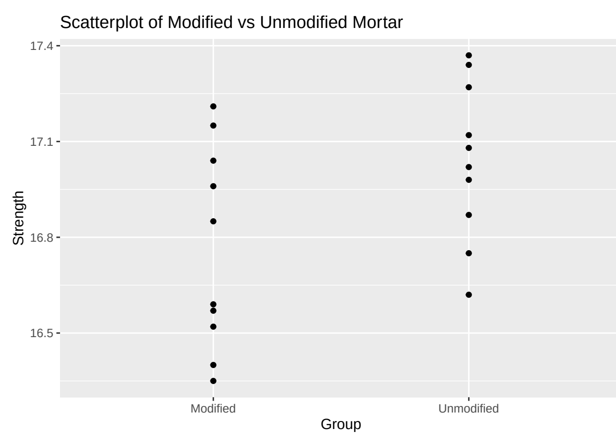
(i) Table 2.1 on page 23 in our textbook

```
Modified_Mortar <- c(16.85, 16.40, 17.21, 16.35, 16.52, 17.04, 16.96, 17.15, 16.59, 16.57)
Unmodified_Mortar <- c(16.62, 16.75, 17.37, 17.12, 16.98, 16.87, 17.34, 17.02, 17.08, 17.27)

mortar.df <- data.frame(
  y = c(Modified_Mortar, Unmodified_Mortar),
  x = factor(rep(c("Modified", "Unmodified"), each = 10))
)

mortar.df
```

```
##        y          x
## 1  16.85   Modified
## 2  16.40   Modified
## 3  17.21   Modified
## 4  16.35   Modified
## 5  16.52   Modified
## 6  17.04   Modified
## 7  16.96   Modified
## 8  17.15   Modified
## 9  16.59   Modified
## 10 16.57   Modified
## 11 16.62 Unmodified
## 12 16.75 Unmodified
## 13 17.37 Unmodified
## 14 17.12 Unmodified
## 15 16.98 Unmodified
## 16 16.87 Unmodified
## 17 17.34 Unmodified
## 18 17.02 Unmodified
## 19 17.08 Unmodified
## 20 17.27 Unmodified
```

```r
# Boxplot
ggplot(data=mortar.df, aes(x = x, y = y)) +
  geom_boxplot() +
  ggtitle("Boxplot of Modified vs Unmodified Mortar") +
  ylab("Strength") +
  xlab("Group")

# Scatterplot
ggplot(data=mortar.df, aes(x = x, y = y)) +
  geom_point() +
  ggtitle("Scatterplot of Modified vs Unmodified Mortar") +
  ylab("Strength") +
  xlab("Group")
```






```r
set.seed(123)

perm_test(mortar.df, reps = 1000)
```

```
## [1] 0.049
```

```r
t.test(y ~ x, data = mortar.df)
```

```
##
##  Welch Two Sample t-test
##
## data:  y by x
## t = -2.1869, df = 17.025, p-value = 0.043
## alternative hypothesis: true difference in means between group Modified and group Unmodified is not e
## 95 percent confidence interval:
##  -0.546174139 -0.009825861
## sample estimates:
##   mean in group Modified mean in group Unmodified
##                   16.764                   17.042
```

Observations from Side-by-Side Boxplots and Scatter Diagrams:

- From the boxplot, we observe that both groups (Modified and Unmodified Mortar) have fairly similar central tendencies, but the Unmodified Mortar group seems to have slightly higher values overall.
- The scatterplot further confirms that while there is some overlap in the strength values, the Unmodified Mortar group generally has slightly higher measurements than the Modified Mortar group.

Permutation Test Results:

- The permutation test returned a p-value of 0.049, which is very close to the conventional significance level of 0.05. This suggests that there might be a statistically significant difference between the two groups.

t-Test Results:

- The t-test resulted in a t-statistic of -2.19 and a p-value of 0.043, which is also statistically significant at the 5% level, indicating that there is likely a difference between the means of the two groups.
- The 95% confidence interval for the difference in means is [-0.546, -0.0098], meaning the difference is small but statistically significant.

Comparison of Results:

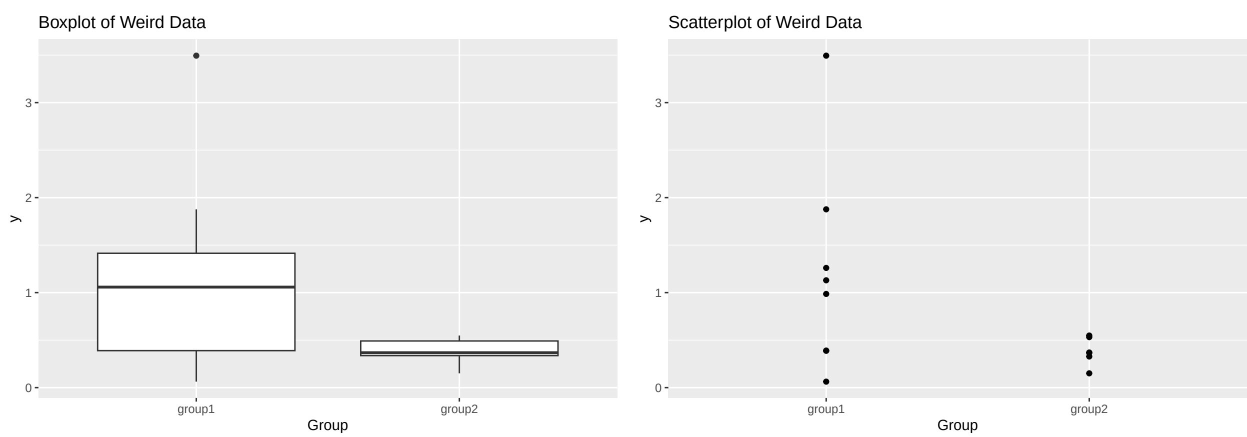
- Both the permutation test and t-test yield similar conclusions: the p-values are below 0.05, suggesting a statistically significant difference between the two groups.
- The slight differences in p-values (0.049 for the permutation test and 0.043 for the t-test) could be due to the inherent differences in how each test works. The permutation test does not make assumptions about normality, whereas the t-test assumes normality and equal variances which might not be entirely true for these data.

(ii) Non-normal data with a true difference in means

```r
library(readr)
weird_data <- read_csv("/Users/ethantsao/PSTAT 122/weird_data.csv")
```

```
## Rows: 14 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): x
## dbl (1): y
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
ggplot(data=weird_data, aes(x = x, y = y)) +
  geom_boxplot() +
  ggtitle("Boxplot of Weird Data") +
  ylab("y") +
  xlab("Group")

ggplot(data=weird_data, aes(x = x, y = y)) +
  geom_point() +
  ggtitle("Scatterplot of Weird Data") +
  ylab("y") +
  xlab("Group")
```




```r
set.seed(123)

perm_test(weird_data, reps = 1000)
```

```
## [1] 0.033
```

```r
t.test(y ~ x, data = weird_data)
```

```
##
##  Welch Two Sample t-test
##
## data:  y by x
```

```
## t = 2.0846, df = 7.3315, p-value = 0.07377
## alternative hypothesis: true difference in means between group group1 and group group2 is not equal t
## 95 percent confidence interval:
##  -0.1012245  1.7335747
## sample estimates:
## mean in group group1 mean in group group2
##            1.1983885            0.3822134
```

Observations from Side-by-Side Boxplots and Scatter Diagrams:

- The boxplot shows a notable difference between the two groups. Group 1 has higher median values than Group 2, and there is less overlap between the groups compared to the mortar dataset.
- The scatterplot also shows a clearer separation between Group 1 and Group 2, indicating a true difference in distributions.

Permutation Test Results:

- The permutation test returned a p-value of 0.033, indicating that the difference between the two groups is statistically significant at the 5% level.

t-Test Results:

- The t-test produced a t-statistic of 2.08 with a p-value of 0.0738, which is not significant at the 5% level but is close.
- The 95% confidence interval for the difference in means is [-0.101, 1.734], suggesting that while the data shows a difference, the t-test is not as confident due to the non-normality of the data.

Comparison of Results:

- The permutation test yielded a significant result, whereas the t-test did not. This occurred because the t-test relies on assumptions of normality, and since the data is highly non-normal, the t-test underestimates this difference.
- The permutation test, on the other hand, does not rely on these assumptions and is more accurate in this case, hence the significant p-value.
- Since we know that there is a true difference in means, the t-test is likely underpowered which could lead to a Type II error (failing to reject the null hypothesis when it is false).

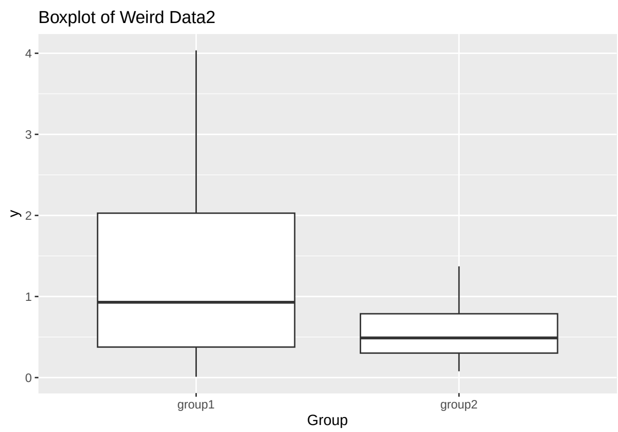
(iii) Non-normal data with no true difference in means

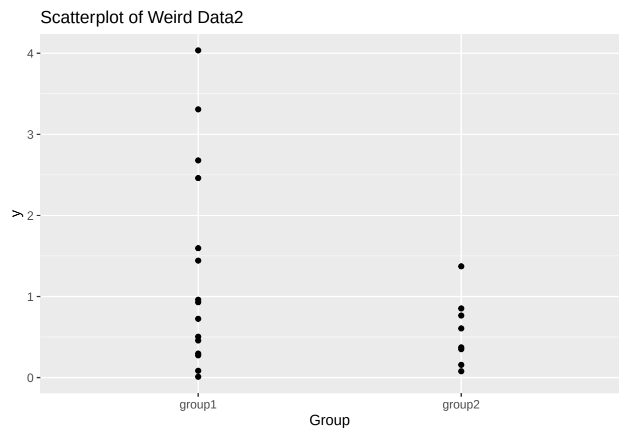
```r
library(readr)
weird_data2 <- read_csv("/Users/ethantsao/PSTAT 122/weird_data2.csv")
```

```
## Rows: 23 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): x
```

```
## dbl (1): y
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
ggplot(data=weird_data2, aes(x = x, y = y)) +
  geom_boxplot() +
  ggtitle("Boxplot of Weird Data2") +
  ylab("y") +
  xlab("Group")

ggplot(data=weird_data2, aes(x = x, y = y)) +
  geom_point() +
  ggtitle("Scatterplot of Weird Data2") +
  ylab("y") +
  xlab("Group")
```






```r
set.seed(123)

perm_test(weird_data2, reps = 1000)
```

```
## [1] 0.062
```

```r
t.test(y~x, data = weird_data2)
```

```
##
##  Welch Two Sample t-test
##
## data:  y by x
## t = 2.0973, df = 18.92, p-value = 0.04963
## alternative hypothesis: true difference in means between group group1 and group group2 is not equal
## 95 percent confidence interval:
##  0.001328876 1.495189455
## sample estimates:
## mean in group group1 mean in group group2
##            1.3173254            0.5690662
```

Observations from Side-by-Side Boxplots and Scatter Diagrams:

- Both the boxplot and scatterplot show that the two groups (Group 1 and Group 2) have highly overlapping distributions, with no clear visual distinction between their central tendencies or spreads.

Permutation Test Results:

- The permutation test yielded a p-value of 0.062, which is slightly above the 0.05 threshold, indicating no statistically significant difference between the two groups.

t-Test Results:

- The t-test resulted in a t-statistic of 2.097 and a p-value of 0.0496, which is just below the 0.05 threshold, suggesting a significant difference between the groups.

Comparison of Results:

- There is a discrepancy between the two tests: the t-test finds a significant difference, while the permutation test does not.
- Since we know there is no true difference in means, the t-test result might be a false positive, likely leading to a Type I error (incorrectly rejecting the null hypothesis when it is true).
- The permutation test's p-value being just above 0.05 suggests it is more conservative in this case, which may be preferable given what we know to be true about the data.

Part 2: ANOVA

Exercise 3.9

Part a)

Null Hypothesis (H0):

The mean tensile strengths of the cement are the same across all four mixing techniques.

H0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

Alternative Hypothesis (H1):

At least one of the mixing techniques leads to a different mean tensile strength.

H1: At least one μ_i is different

Significance Level:

We will use $\alpha = 0.05$ as the significance level.

Conducting the ANOVA Test:

Using an ANOVA test, we can compare the mean tensile strengths across the four groups. ANOVA is the appropriate test because we are comparing more than two group means.

```
data <- data.frame(
  tensile_strength = c(3129, 3000, 2865, 2890,
                       3200, 3150, 2985, 3050,
                       2800, 2900, 2985, 3000,
                       2600, 2765, 2700, 2600),
  technique = factor(rep(1:4, each = 4))
)

anova_result <- aov(tensile_strength ~ technique, data = data)
summary(anova_result)
```

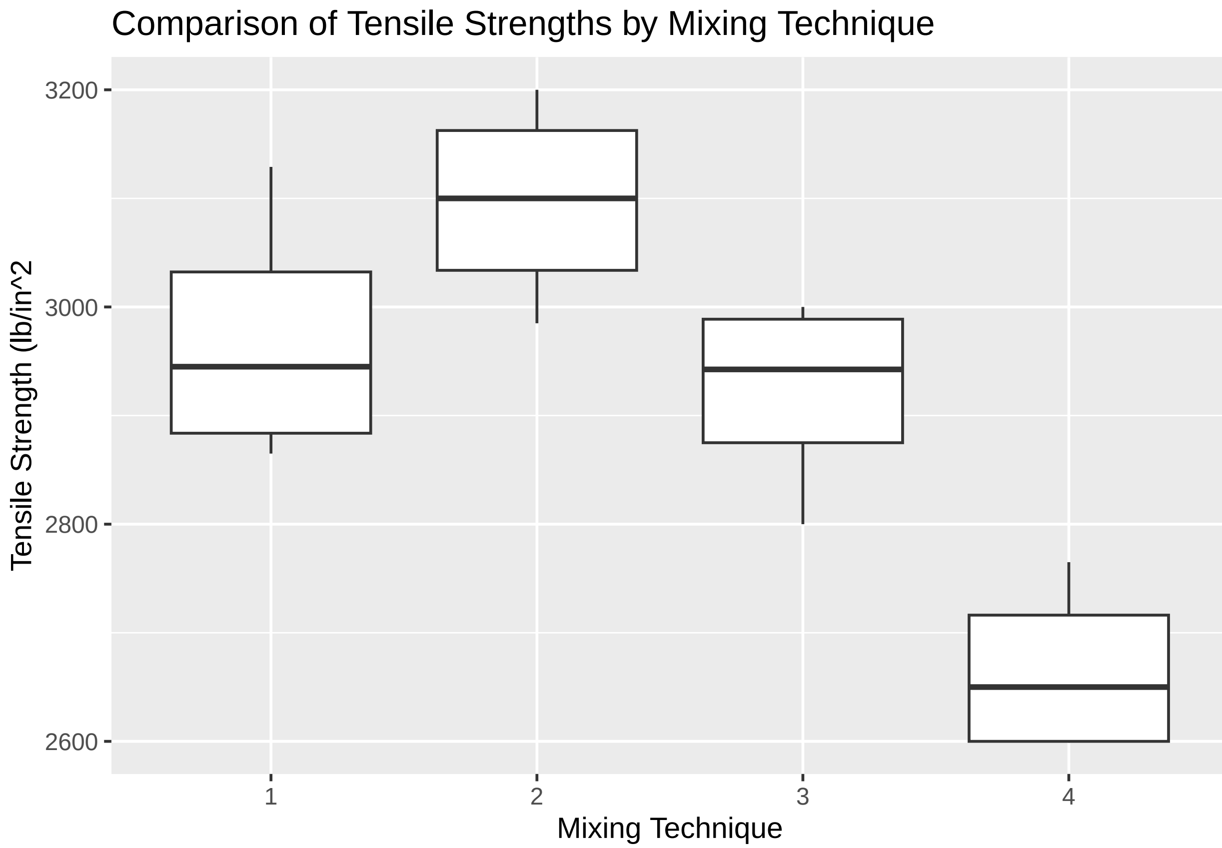
```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## technique    3 391585  130528    13.4 0.000387 ***
## Residuals   12 116858    9738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusion:

- Since the p-value is 0.000387, which is much smaller than the significance level $\alpha = 0.05$, we reject the null hypothesis. This means there is strong evidence that at least one of the mixing techniques significantly affects the tensile strength of the cement.

Part b)

```
ggplot(data, aes(x = technique, y = tensile_strength)) +
  geom_boxplot() +
  ggtitle("Comparison of Tensile Strengths by Mixing Technique") +
  ylab("Tensile Strength (lb/in^2") +
  xlab("Mixing Technique")
```



Mixing Techniques 1 and 2:

Both techniques 1 and 2 exhibit relatively higher tensile strength, with technique 2 having the highest median and range. Technique 1 has a slightly lower median but still a large spread.

Mixing Technique 3:

Technique 3 shows a moderate tensile strength but has a tighter spread compared to techniques 1 and Its median is lower than those two techniques.

Mixing Technique 4:

Technique 4 has the lowest tensile strength among the four techniques. It also shows less variability, with its interquartile range being smaller.

Conclusions:

- There appears to be a clear difference in the distributions of tensile strengths across the different techniques. Techniques 1 and 2 seem to have stronger tensile results, while Technique 4 is the weakest.
- These graphical results are consistent with the findings from the ANOVA test, where we rejected the null hypothesis, showing that there is a significant difference in the tensile strengths across the four mixing techniques.