t and Permutation Test Homework

Megha Pandit

April 16, 2019

Snodgrass Problem

```
twain <- c(0.225, 0.262, 0.217, 0.240, 0.230, 0.229, 0.235, 0.217)
snodgrass <- c(0.209, 0.205, 0.196, 0.210, 0.202, 0.207, 0.224, 0.223, 0.220, 0.201)
```

The null hypothesis here is: $H_0 : \mu_t = \mu_s$, and alternate hypothesis is: $H_1 : \mu_t \neq \mu_s$

For the parametric test, the two test statistics we need to test for are:

$$T = \frac{\bar{X}_t - \bar{X}_s}{S_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

and

$$S_{pool}^2 = \frac{(n_1 - 1)S_1^2(n_2 - 1)S_2^2}{n_1 + n_2 + 2}$$

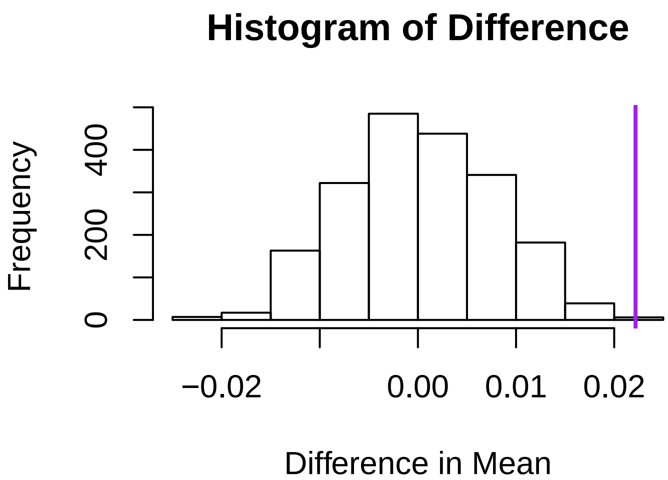
```
spool <- sqrt(((8-1)*var(twain)+(10-1)*var(snodgrass))/(8+10-2))
t <- (mean(twain)-mean(snodgrass))/spool*sqrt((1/8)+(1/10))
```

The t statistic is 0.87, which is less than 2. Therefore, we cannot reject the null hypothesis.

For the permutation test: Considering the large number of permutations we could get from the data, only 2000 are considered here.

```
mean_diff <- mean(twain) - mean(snodgrass)
data <- c(twain, snodgrass)

set.seed(99)
p.test <- function(x, a, b){
s <- sample(x)
mean(s[1:a]) - mean(s[(a+1):b])
}
many <- replicate(2000, p.test(data,8,18))
hist(many, main = "Histogram of Difference", xlab = "Difference in Mean"); abline(v=mean_diff,lwd=2, co
```




```r
n <- sum(abs(many) > abs(mean_diff));n
```

```
## [1] 2
```

There are only two out of the 2000 times where the true difference was exceeded and the p value is < 0.05, therefore, we can reject the null hypothesis.

Hot Dog Problem

```r
hd <- c(186, 181, 176, 149, 184, 190, 158, 139, 175, 148, 152, 111, 141, 153, 190, 157, 131, 149, 135,
hd_bar <- mean(hd)
hd_bar #mean of the observations
```

```
## [1] 156.85
```

```r
hd_sd <- sd(hd)
hd_sd #standard deviation of the observations
```

```
## [1] 22.64201
```

```r
n <- 20
```

The critical value of t is 1.729, for a two-tailed test at an $\alpha = 0.1$ and df $= 19$.

```r
#upper tail
hd_bar + 1.729 * hd_sd/sqrt(n)
```

```
## [1] 165.6038
```

```r
#lower tail
hd_bar - 1.729 * hd_sd/sqrt(n)
```

```
## [1] 148.0962
```

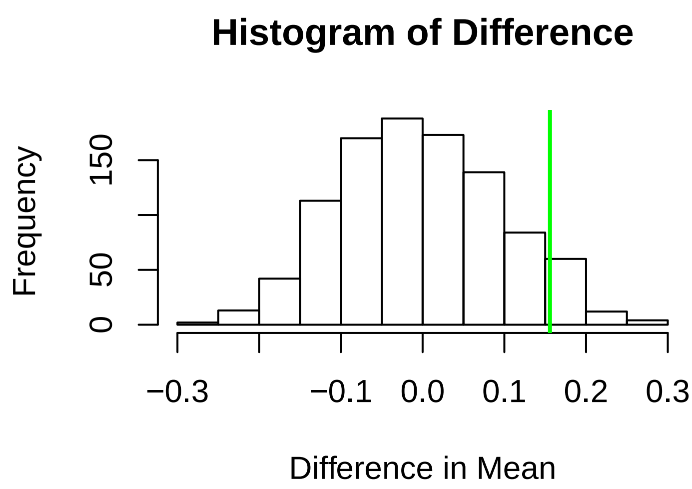
The 90% confidence interval for the mean is [148.0962, 165.6038].

Reading Score Problem

```r
c1 <- c(1.23, 1.42, 1.41, 1.62, 1.55, 1.51, 1.60, 1.76)
c2 <- c(1.76, 1.41, 1.87, 1.49, 1.67, 1.81)
t.test(c1, c2, alternative= "less", var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  c1 and c2
## t = -1.6934, df = 12, p-value = 0.05808
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##         -Inf 0.008184246
## sample estimates:
## mean of x mean of y
##  1.512500  1.668333
```

```r
m_diff <- abs(mean(c1) - mean(c2))
c <- c(c1, c2)
set.seed(222)
diff <- replicate(1000, p.test(c, 8, 14))
hist(diff, main = "Histogram of Difference", xlab = "Difference in Mean"); abline(v=m_diff,lwd=2, col="g
```




```r
num <- sum(abs(diff)>m_diff); num
```

```
## [1] 111
```

In the t-test, $p > 0.05$, therefore, we fail to reject the null hypothesis that the mean of class 1 is higher or equal to the mean of class 2.

In the permutation test, $p = 0.11 > 0.05$, and similar to the t-test, we fail to reject the null hypothesis.