HW9

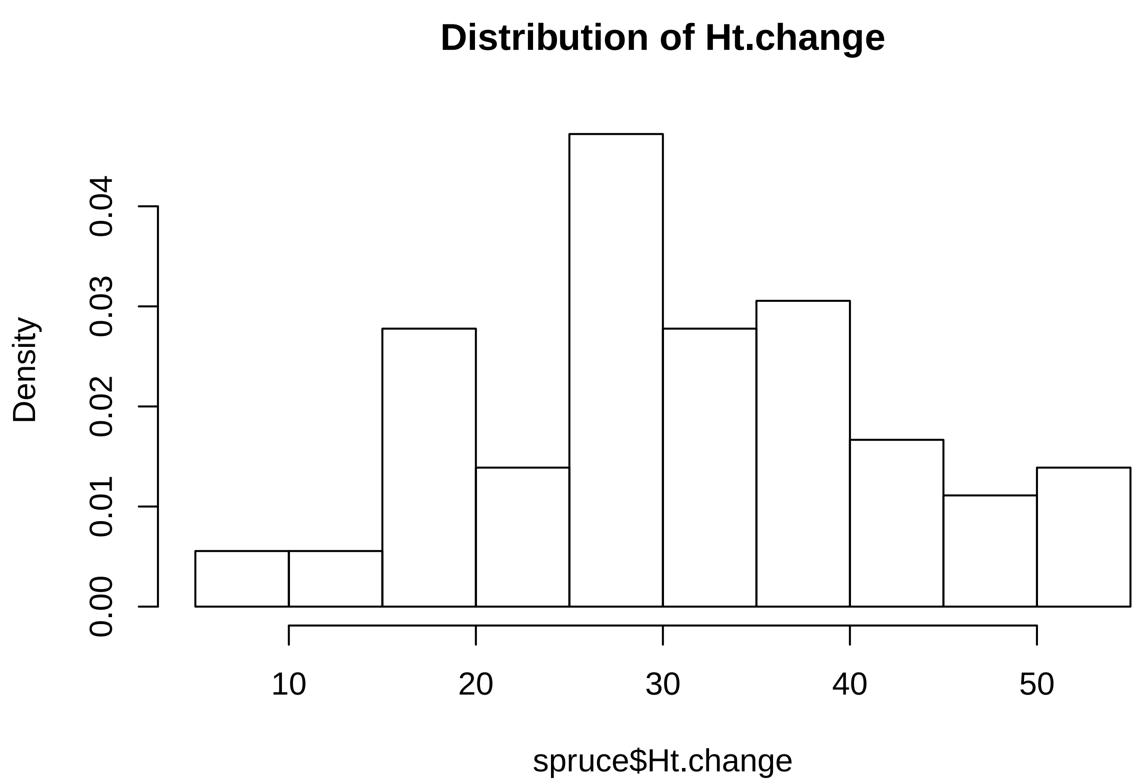
Yigao Li

November 28, 2017

Problem 1 (7-9)

(a)

```
spruce <- read.csv("D:/Courses/ANLY 511/Spruce.csv")
hist(spruce$Ht.change, prob = TRUE, main = "Distribution of Ht.change")
```



(b)

```
dHeight <- spruce$Height5 - spruce$Height0
n <- length(dHeight)
dHeight.s <- sd(dHeight)
dHeight.mu <- mean(dHeight)
error <- dHeight.s/sqrt(n)
talpha <- qt(0.975, n-1)
cat("95% t confidence interval for the mean height change over the 5-year period is [",
    dHeight.mu-talpha*error, ",", dHeight.mu+talpha*error, "].")
```

```
## 95% t confidence interval for the mean height change over the 5-year period is [ 28.33685 , 33.52982
```

We are 95% confident that the true height change is between 28.33685 and 33.52982.

Problem 2 (7-10)

```r
n1 <- 43
x1 <- 5.8
s1 <- 8.6
n2 <- 36
x2 <- 1.9
s2 <- 4.2
talpha <- qt(0.975, n1+n2-2)
me <- talpha * sqrt(s1^2/n1 + s2^2/n2)
cat("95% t confidence interval for the mean difference in weight loss between 2 groups is [",
    x1-x2-me, ",", x1-x2+me, "].")
```

```
## 95% t confidence interval for the mean difference in weight loss between 2 groups is [ 0.9397875 , 6
```

We are 95% confident that the true difference in weight loss between 2 groups is between 0.9397875 and 6.860212.

Problem 3 (7-20)

(a)

```r
nf <- 980
mf <- 459
nm <- 759
mm <- 426
pf_hat <- mf/nf
zalpha <- qnorm(0.975)
lowerbound <- pf_hat - zalpha * sqrt(pf_hat*(1-pf_hat)/nf)
upperbound <- pf_hat + zalpha * sqrt(pf_hat*(1-pf_hat)/nf)
cat("95% confidence interval for the proportion of women who voted for Bush is [", lowerbound, ",",
    upperbound, "].")
```

```
## 95% confidence interval for the proportion of women who voted for Bush is [ 0.4371257 , 0.499609 ].
```

(b)

```r
pm_hat <- mm/nm
lowerbound <- pm_hat - zalpha * sqrt(pm_hat*(1-pm_hat)/nm)
upperbound <- pm_hat + zalpha * sqrt(pm_hat*(1-pm_hat)/nm)
cat("95% confidence interval for the proportion of men who voted for Bush is [", lowerbound, ",",
    upperbound, "].")
```

```
## 95% confidence interval for the proportion of men who voted for Bush is [ 0.5259618 , 0.5965679 ].
```

No, confidence intervals for men and women do not overlap. The proportion of men voting for Bush is higher than that of women.

(c)

```r
p_hat <- (mf+mm)/(nf+nm)
se <- sqrt(p_hat*(1-p_hat)*(1/nf+1/nm))
lowerbound <- pm_hat - pf_hat - zalpha * se
upperbound <- pm_hat - pf_hat + zalpha * se
cat("95% confidence interval for the difference in proportions (men - women) is [", lowerbound, ",",
    upperbound, "].")
```

```
## 95% confidence interval for the difference in proportions (men - women) is [ 0.04552076 , 0.1402742 ]
```

We are 95% confident that proportion of men voting for Bush is 4.552076% to 14.02742% higher than proportion of women voting for Bush.

Problem 4 (8-3)

(a) H_0: $\mu = 98.6$
H_a: $\mu > 98.6$

(b)

```r
temperature <- c(98, 98.9, 99, 98.9, 98.8, 98.6, 99.1, 98.9, 98.5, 98.9, 98.9, 98.4, 99, 99.2, 98.6,
                 98.8, 98.9, 98.7)
t.test(temperature, alternative = "greater", mu = 98.6)
```

```
##
##  One Sample t-test
##
## data:  temperature
## t = 2.745, df = 17, p-value = 0.006907
## alternative hypothesis: true mean is greater than 98.6
## 95 percent confidence interval:
##  98.66715      Inf
## sample estimates:
## mean of x
##  98.78333
```

P-value is $0.006907 < \alpha = 0.05$. We reject H_0 and conclude that body temperatures in children in Sober are higher than the standard one.

Problem 5 (8-10)

H_0: $\hat{p} = 29.1\%$
H_a: $\hat{p} < 29.1\%$

```r
prop.test(87, 350, p = 0.291, alternative = "less")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  87 out of 350, null probability 0.291
## X-squared = 2.8517, df = 1, p-value = 0.04564
## alternative hypothesis: true p is less than 0.291
## 95 percent confidence interval:
##  0.0000000 0.2898908
## sample estimates:
##         p
## 0.2485714
```

P-value is $0.04564 < \alpha = 0.05$. We reject H_0 and sample data supports professor's hypothesis.

Problem 6 (7-30)

Without loss of generality, assume that $\hat{\theta}_1 \geq \hat{\theta}_2$.
$(\implies)\hat{\theta}_1 \pm 1.96\hat{SE}_1$ and $\hat{\theta}_2 \pm 1.96\hat{SE}_2$ overlap. Therefore,

$$\hat{\theta}_1 - 1.96\hat{SE}_1 \leq \hat{\theta}_2 + 1.96\hat{SE}_2$$
$$\hat{\theta}_1 - \hat{\theta}_2 - 1.96(\hat{SE}_1 + \hat{SE}_2) \leq 0$$

Since $\hat{\theta}_1 - \hat{\theta}_2 \geq 0$, $\hat{SE}_1, \hat{SE}_2 \geq 0$,

$$\hat{\theta}_1 - \hat{\theta}_2 + 1.96(\hat{SE}_1 + \hat{SE}_2) \geq 0$$

Therefore, $\hat{\theta}_1 - \hat{\theta}_2 \pm 1.96(\hat{SE}_1 + \hat{SE}_2)$ must contain 0.
$(\Longleftarrow)\hat{\theta}_1 - \hat{\theta}_2 \pm 1.96(\hat{SE}_1 + \hat{SE}_2)$ contains 0

$$\hat{\theta}_1 - \hat{\theta}_2 + 1.96(\hat{SE}_1 + \hat{SE}_2) \geq 0$$
$$\hat{\theta}_1 - \hat{\theta}_2 - 1.96(\hat{SE}_1 + \hat{SE}_2) \leq 0$$

So,

$$\hat{\theta}_1 - 1.96\hat{SE}_1 \leq \hat{\theta}_2 + 1.96\hat{SE}_2$$

Because $\hat{\theta}_1 \geq \hat{\theta}_2$, $\hat{\theta}_1 \pm 1.96\hat{SE}_1$ and $\hat{\theta}_2 \pm 1.96\hat{SE}_2$ must overlap.

Problem 7 (7-33)

(a) Let s^2 be sample variance. By Central Limit Theorem, $\lambda \sim N(\bar{X}, \frac{s^2}{n})$. Therefore, 95% confidence interval is $[\bar{X} - 1.96\frac{s}{\sqrt{n}}, \bar{X} + 1.96\frac{s}{\sqrt{n}}]$

(b)

```r
x <- c(4,6,7,9,10,13)
mean(x)
```

```
## [1] 8.166667
```

```r
var(x)
```

```
## [1] 10.16667
```

$\bar{X} = \frac{49}{6}$, $s^2 = \frac{61}{6}$, $n = 6$, 95% confidence interval $[5.6, 10.7]$

Problem 8 (8-22)

(a)

```r
ppois(16/5, 2, lower.tail = FALSE)
```

```
## [1] 0.1428765
```

The probability of a Type I error is 14.28765%.
(b)

```r
ppois(16/5, 4, lower.tail = TRUE)
```

```
## [1] 0.4334701
```

The probability of a Type II error is 43.34701%.