

Homework 3 for STA 5166 (Assigned, Oct. 8)

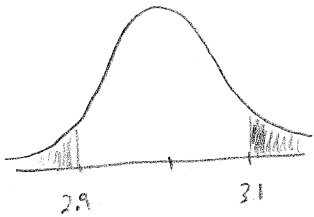
Statistics in Applications I

Due: Oct. 17, 2007 (Wednesday)

- 1: BHH Ch.2; Problems 10, 12, 13(a,b,c,d); Pages 62-65. (40)
- 2: BHH Ch.3; Problem 2 (Pages 124-125). Submit both your summary results and R/Splus program for the problem. (20)
- 3: BHH Ch.3; Problems 4, 7, and 13 (Pages 125-128). For each of the three problems, perform a t -test on the difference of the two means and perform a test based on a randomization distribution (use R/Splus to generate 10000 samples and plot the histogram of the differences). Submit both your summary results and R/Splus program for each of the problems. (40)

10

c) $X \sim N(3.06, 0.03)$

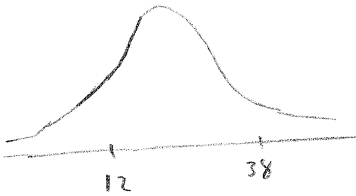


$$P(X \leq 2.9) + P(X \geq 3.1)$$

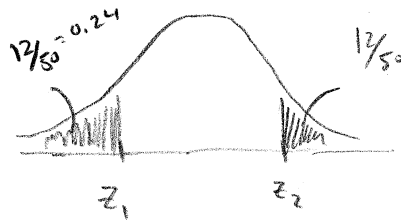
$$0 + 0.0912 = 9.12\%$$

if $n = 397$ then $n_1 = 0$
 $n_2 \approx 36.6 = 37 = (397)(9.12\%)$

b) $X \sim N(3.06, 0.03)$



$$Z \sim N(0, 1)$$



$$P(X \leq 2.9) = P\left(Z \leq \frac{2.9 - \mu}{\sigma}\right) = \frac{12}{50} \rightarrow -0.7063 = \frac{2.9 - \mu}{\sigma}$$

$$P(X \geq 3.1) = P\left(Z \leq \frac{3.1 - \mu}{\sigma}\right) = \frac{12}{50} \rightarrow 0.7063 = \frac{3.1 - \mu}{\sigma}$$

$$3.1 - 0.7063\sigma = \mu = 2.9 + 0.7063\sigma$$

$$\sigma = 0.1415$$

$$\mu = 2.9 + 0.7063(0.1415)$$

$$\mu = 2.9 \approx 3$$

Assumed that randomly
 selected bolts which
 lengths are iid. normally
 distributed.

⑫

$$X \sim N(1000, 10)$$

$$a) P(X < 985) = P\left(Z < \frac{985 - 1000}{10}\right) = 0.066807 = 6.68\%$$

$$b) P(X > 1020) = P\left(Z > \frac{1020 - 1000}{10}\right) = 0.0227 = 2.27\%$$

$$c) P(985 < X < 1020) = P\left(\frac{985 - 1000}{10} < Z < \frac{1020 - 1000}{10}\right)$$

$$= 0.9104 = 91.04\%$$

⑬

$$P(1 \text{ or } 2) = p$$

| x | f | $P(x)$ |
|-----|-----|---------|
| 0 | 0 | 0 |
| 1 | 4 | $4/64$ |
| 2 | 19 | $19/64$ |
| 3 | 15 | $15/64$ |
| 4 | 17 | $17/64$ |
| 5 | 5 | $5/64$ |
| 6 | 2 | $2/64$ |

$$a) P(1 \text{ or } 2) = P(1) + P(2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$E(X) = p = \frac{1}{3}$$

$$Var(X) = pq = \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{2}{9}$$

(13)

for one die

$$b) E(x) = 1\left(\frac{4}{64}\right) + 2\left(\frac{15}{64}\right) + 3\left(\frac{15}{64}\right) + 4\left(\frac{17}{64}\right) + 5\left(\frac{7}{64}\right) + 6\left(\frac{2}{64}\right)$$

$$= 3.15625$$

Since 6 die

$$E(x) = \frac{3.15625}{6} = 0.526$$

$$s^2 = \text{Var}(x) = E(x^2) - [E(x)]^2 =$$

$$= 1^2\left(\frac{4}{64}\right) + 2^2\left(\frac{15}{64}\right) + \dots + (6)^2\left(\frac{2}{64}\right) - [0.526]^2$$

$$= 0.24937$$

c) Testing hypothesis of mean

$$H_0: p = p_0 = 1/3$$

$$H_a: p > p_0 = 1/3$$

$$Z = \frac{0.526 - 1/3}{\sqrt{\frac{2/9}{64.6}}} = 8.009 > Z_{0.05}$$

$$P(Z > 8.009) = p\text{-value} = 0$$

Reject null H_0 accept $p > p_0$

mean and variance

don't have theoretical values

for Bernoulli

Testing variance

$$H_0: \sigma^2 = \sigma_0^2 \quad p_0 q_0 = \frac{2}{9}$$

$$H_a: \sigma^2 \neq \sigma_0^2 \quad s^2 = 0.2493$$

$$\chi^2 = \frac{(n-1)s^2}{p_0 q_0}$$

$$= \frac{(63)(0.2493)}{2/9}$$

$$= 70.6 \sim \chi_{64}^2$$

$$P(\chi_{64}^2 > 70.6)$$

Reject H_0

13 d) $X \sim \text{Bin}(1/3, 6)$

$$P(X=x_i) = \binom{6}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{6-x}$$

$$P(X=0) = \binom{6}{0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 = 0.088$$

$$P(X=1) = \binom{6}{1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^5 = 0.263$$

$$P(X=2) = \binom{6}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^4 = 0.329$$

$$P(X=3) = 0.219$$

$$P(X=4) = 0.082$$

$$P(X=5) = 0.016$$

$$P(X=6) = 0.001$$

| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|------------------|--------------|--------|--------------|--------|-------|-------|-------|
| expected freq | 5.632 | 16.832 | 21.056 | 14.016 | 5.248 | 1.024 | 0.064 |
| | ↑ | | ↓ | | | | |
| | $64(P(X=0))$ | | $64(P(X=1))$ | | | | |

Chapter 3.2 (三。二)

Summary:

Will try to test the hypothesis to see if there exists a significant difference between the mean values of levels of asbestos fiber in the air of the industrial plant with and without S-142 chemical. From the comparative trail in the plant, the four consecutive readings had a mean difference of -3.5. The null hypothesis is that with or without S-142, the asbestos levels will not change, the alternative is that with S-142, the level will decrease since the mean difference is negative. To test this, used as a reference the past observations of asbestos levels without S-142. From the dataset, obtained a probability that $1/109$ ($=0.0091743119$) that there exists a mean difference less than the comparative trail. Since this probability is less than 5%, we reject the null hypothesis and accept the salesman claim that S-142 is beneficial to reduce the level of asbestos levels in the air of the industrial plant.

CODE

```
data=scan("C:/Documents and Settings/Jaime/Desktop/FALL07/STA5166/BHH2-
Data/datahw3.dat")
data
n1=0
Mean1wout = mean(c(8,6))
Mean2with = mean(c(3,4))
diff_means = Mean2with-Mean1wout
y = c(rep(NA, (109)))
x = c(rep(NA, (109)))
for(i in 1:109){ y[i] = (data[i]+data[i+1])/2}
for(j in 1:109){ x[j] = y[j+2] - y[j];
if(x[j]<= diff_means) n1=n1+1}
x
sort(x)
n1
diff_means
n1/109
```

OUTPUT

```
> data=scan("C:/Documents and Settings/Jaime/Desktop/FALL07/STA5166/BHH2-
Data/datahw3.dat")
Read 112 items
> data
 [1] 9 10 8 9 8 8 8 7 6 9 10 11 9 10 11 11 11 11 10 11 12 13 12 13 12
[26] 14 15 14 12 13 13 12 13 13 13 13 13 10 8 9 8 6 7 7 6 5 6 5 6 4
[51] 5 4 4 2 4 5 4 5 6 5 5 6 5 6 7 8 8 8 7 9 10 9 10 9 8
[76] 9 8 7 7 8 7 7 7 8 8 8 8 7 6 5 6 5 6 7 6 6 5 6 6 5
[101] 4 3 4 4 5 5 6 5 6 7 6 5
> n1=0
> Mean1wout = mean(c(8,6))
> Mean2with = mean(c(3,4))
> diff_means = Mean2with-Mean1wout
> y = c(rep(NA, (109)))
> x = c(rep(NA, (109)))
> for(i in 1:109){ y[i] = (data[i]+data[i+1])/2}
> for(j in 1:109){ x[j] = y[j+2] - y[j];
+ if(x[j]<= diff_means) n1=n1+1}
Error in if (x[j] <= diff_means) n1 = n1 + 1 :
  missing value where TRUE/FALSE needed
> x
 [1] -1.0 -0.5 -0.5 -0.5 -0.5 -1.5 0.0 3.0 3.0 0.5 -1.0 0.5 1.5 0.5 0.0
[16] -0.5 -0.5 1.0 2.0 1.0 0.0 0.0 0.5 2.0 1.5 -1.5 -2.0 0.0 0.0 -0.5
[31] 0.5 0.5 0.0 0.0 -1.5 -4.0 -3.0 -0.5 -1.5 -2.0 0.0 0.0 -1.5 -1.0 0.0
[46] 0.0 -0.5 -1.0 -0.5 -0.5 -1.5 -1.0 1.5 1.5 0.0 1.0 1.0 -0.5 0.0 0.5
```

```

[61] 0.0 1.0 2.0 1.5 0.5 -0.5 0.0 2.0 1.5 0.0 0.0 -1.0 -1.0 0.0 -1.0
[76] -1.5 0.0 0.5 -0.5 -0.5 0.5 1.0 0.5 0.0 -0.5 -1.5 -2.0 -1.0 0.0 0.0
[91] 1.0 1.0 -0.5 -1.0 -0.5 0.5 0.0 -1.5 -2.0 -1.0 0.5 1.0 1.0 1.0 0.5
[106] 0.0 1.0 NA NA
> sort(x)
[1] -4.0 -3.0 -2.0 -2.0 -2.0 -2.0 -1.5 -1.5 -1.5 -1.5 -1.5 -1.5 -1.5 -1.5 -1.5
[16] -1.0 -1.0 -1.0 -1.0 -1.0 -1.0 -1.0 -1.0 -1.0 -1.0 -1.0 -1.0 -0.5 -0.5 -0.5 -0.5
[31] -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 -0.5 0.0
[46] 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
[61] 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.5 0.5 0.5 0.5 0.5 0.5
[76] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
[91] 1.0 1.0 1.0 1.0 1.0 1.5 1.5 1.5 1.5 1.5 1.5 1.5 2.0 2.0 2.0 2.0
[106] 3.0 3.0
> n1
[1] 1
> diff_means
[1] -3.5
> n1/109
[1] 0.009174312

```


Chapter 3.4 (三。四)

Assumptions:

Ratings are both approximately normal distributed. Two samples, A and B, are independent. Ratings in each brand are i.i.d.

Summary:

To test the hypothesis that $\eta_A = \eta_B$, against the $\eta_A \neq \eta_B$. I used a t-test to check if the difference in means is not equal to zero. The p-value obtained 0.3316. Therefore there is not sufficient evidence to reject the null hypothesis. The assumptions were needed to conduct test.

Using a randomization distribution, I also tested the above hypothesis. Here, I did not make assumptions about the distributions of the ratings. The mean of brand A: (3.875) and brand B: (5.285714), to obtain a difference of 1.4107.

There exist 6435 possible permutations of 8 ratings of brand A and 7 ratings of brand B. Assuming that the null hypothesis, then there exist no difference in the ratings of brand A and brand B. Can arrange and for each calculate the differences that are less than 1.4107. Count the number of occurrences and this will lead to a calculation of the p-value.

The p-value obtained after a large number of observations should be approximately equal to the p-value obtained from the t-test above. I obtained the p-value: 0.3551. This also leads to the conclusion that one cannot reject the null hypothesis.

CODE

```
brandA = c(2,4,2,1,9,9,2,2)
brandB = c(8,3,5,3,7,7,4)

y = t.test(brandA, brandB)
y
n1=0
h1=0
y1= c(2,4,2,1,9,9,2,2,8,3,5,3,7,7,4)
c1=c(rep("A", 8), rep("B", 7))
d1 = c(rep(0,10000))
diff= 5.285714-3.875
for(i in 1:10000){
  c2=sample(c1);
  x1=y1[c2=="A"];
  x2=y1[c2=="B"];
  m1 = mean(x1);
  m2 = mean(x2);
  d1[i] = m2-m1;
  h1=c(h1,d1[i]);
  if(abs(d1[i]) >= 1.4107)n1=n1+1
}
n1
hist(h1, main="Randomization Distribution")
pvalue= n1/10000
pvalue
```

OUTPUT

```
> brandA = c(2,4,2,1,9,9,2,2)
> brandB = c(8,3,5,3,7,7,4)
>
> y = t.test(brandA, brandB)
> y
```

Welch Two Sample t-test

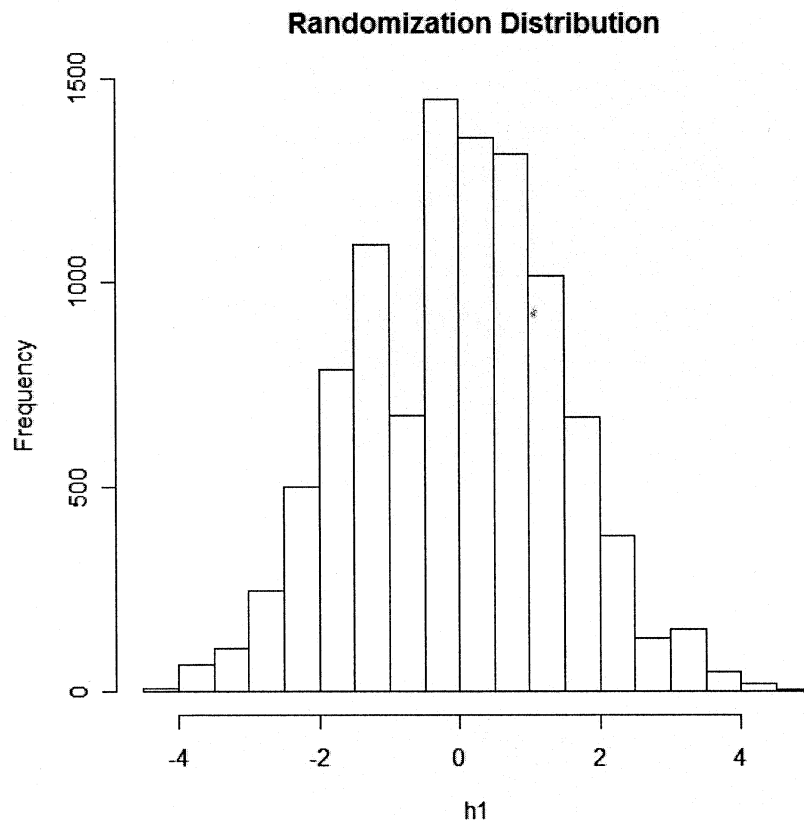
```
data: brandA and brandB
t = -1.0122, df = 11.923, p-value = 0.3316
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.449587  1.628159
sample estimates:
mean of x mean of y
 3.875000  5.285714

> n1=0
```

```

> h1=0
> y1= c(2,4,2,1,9,9,2,2,8,3,5,3,7,7,4)
> c1=c(rep("A", 8), rep("B", 7))
> d1 = c(rep(0,10000))
> diff= 5.285714-3.875
> for(i in 1:10000){
+ c2=sample(c1);
+ x1=y1[c2=="A"];
+ x2=y1[c2=="B"];
+ m1 = mean(x1);
+ m2 = mean(x2);
+ d1[i] = m2-m1;
+ h1=c(h1,d1[i]);
+ if(abs(d1[i]) >= 1.4107)n1=n1+1
+ }
> n1
[1] 3551
> hist(h1, main="Randomization Distribution")
> pvalue= n1/10000
> pvalue
[1] 0.3551

```



Chapter 3.7 (三。七)

Assumptions:

Results are both approximately normal distributed. Two samples, designs A and B, are independent. Results in each design are i.i.d.

Summary:

Will try to test the hypothesis to see if there exists a significant difference between the mean values for the power attainable for the two designs. The null hypothesis assumes there is no difference in the mean values. I used a t-test to check if the difference in means is not equal to zero. The p-value obtained 0.4343. Therefore there is not sufficient evidence to reject the null hypothesis. The assumptions were needed to conduct test.

Using a randomization distribution, I also tested the above hypothesis. Here, I did not make assumptions about the distributions of the ratings. The mean of design A: (1.55) and brand B: (1.75), to obtain a difference of 0.2

The p-value obtained after a large number of observations should be approximately equal to the p-value obtained from the t-test above. I obtained the p-value: 0.4454. This also leads to the conclusion that one cannot reject the null hypothesis.

CODE

```
designA = c(1.8, 1.9, 1.1, 1.4)
designB = c(1.9, 2.1, 1.5, 1.5)
y = t.test(designA , designB)
y
n1=0
h1=NULL
y1= c(1.8, 1.9, 1.1, 1.4, 1.9, 2.1, 1.5, 1.5)
c1=c(rep("A", 4), rep("B", 4))
diff=1.75-1.55
d1 = rep(0, 10000)
for(i in 1:10000){
  c2=sample(c1);
  x1=y1[c2=="A"]; x2=y1[c2=="B"]
  m1 = mean(x1); m2 = mean(x2);
  d1[i] = m2-m1;
  h1=c(h1,d1[i]);
  if(abs(d1[i]) >= diff)n1=n1+1
}
n1
hist(h1, main="Randomization Distribution")
pvalue= n1/10000
pvalue
```

OUTPUT

```
> designA = c(1.8, 1.9, 1.1, 1.4)
> designB = c(1.9, 2.1, 1.5, 1.5)
>
> y = t.test(designA , designB)
> y
```

Welch Two Sample t-test

```
data: designA and designB
t = -0.8402, df = 5.756, p-value = 0.4343
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7885191  0.3885191
sample estimates:
mean of x mean of y
  1.55      1.75
```

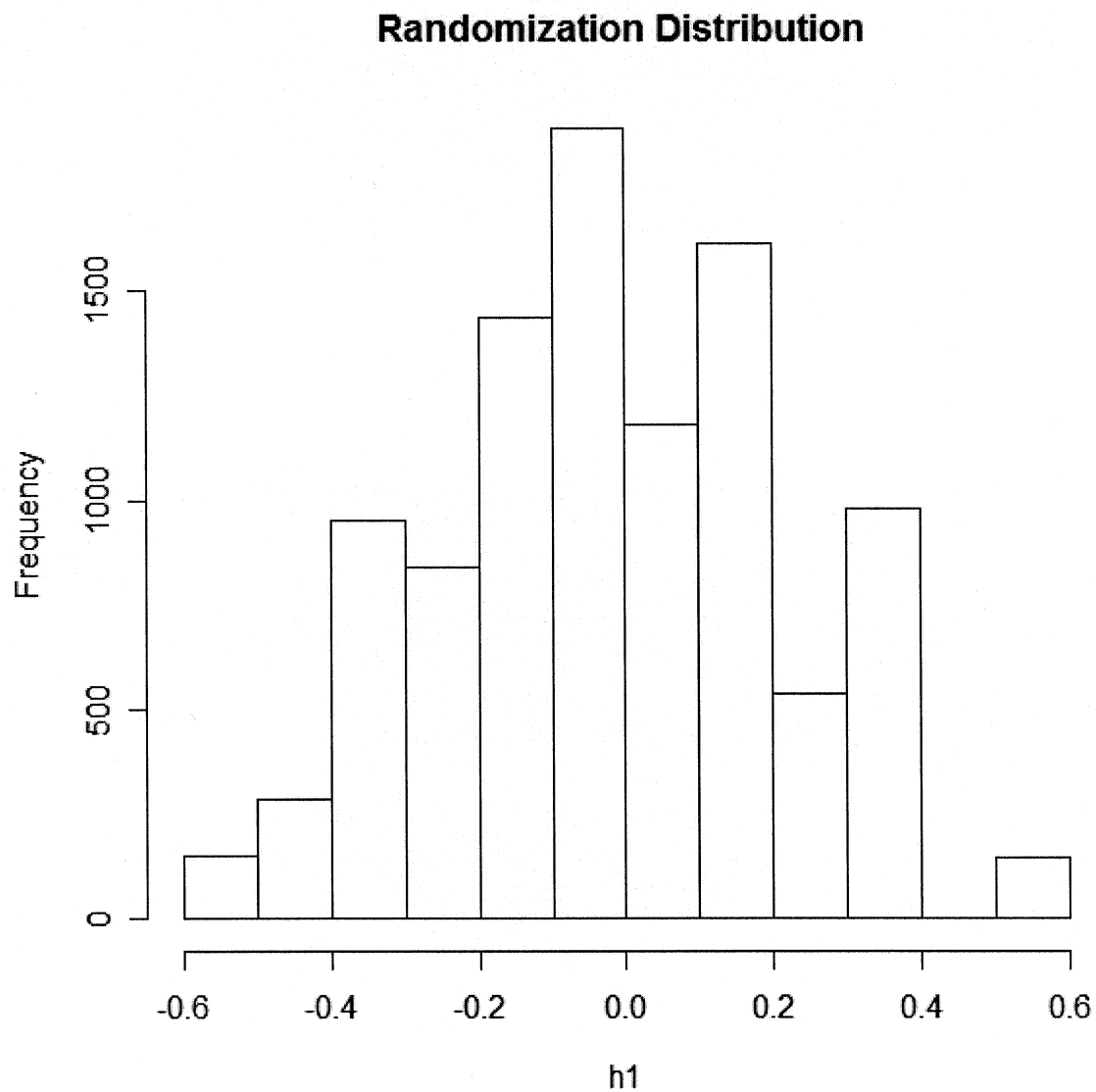
```
> n1=0
> h1=NULL
> y1= c(1.8, 1.9, 1.1, 1.4, 1.9, 2.1, 1.5, 1.5)
> c1=c(rep("A", 4), rep("B", 4))
> diff=1.75-1.55
> d1 = rep(0, 10000)
> for(i in 1:10000){
```

```

+ c2=sample(c1);
+ x1=y1[c2=="A"]; x2=y1[c2=="B"]
+ m1 = mean(x1); m2 = mean(x2);
+ d1[i] = m2-m1;
+ h1=c(h1,d1[i]);
+ if(abs(d1[i]) >= diff)n1=n1+1
+ }
> n1
[1] 4454
> hist(h1, main="Randomization Distribution")
> pvalue= n1/10000
> pvalue
[1] 0.4454

```

=====



Chapter 3.13 (三。十三)

Assumptions:

Results of production from each diet are both approximately normal distributed. Two samples, designs A and B, are independent. Results in each diet are i.i.d.

Summary:

Will try to test the hypothesis to see if there exists a significant difference between the mean values for the power attainable for the two designs. The null hypothesis assumes there is no difference in the mean values. I used a t-test to check if the difference in means is not equal to zero. The p-value obtained 0.07842. Therefore there is not sufficient evidence to reject the null hypothesis. The assumptions were needed to conduct test.

Using a randomization distribution, I also tested the above hypothesis. Here, I did not make assumptions about the distributions of the ratings. The mean of diet A: (166.5) and brand B: (156.6667), to obtain a absolute value of the difference of 9.83

The p-value obtained after a large number of observations should be approximately equal to the p-value obtained from the t-test above. I obtained the p-value: 0.0913. This also leads to the conclusion that one cannot reject the null hypothesis.

A 95% confidence interval for the mean difference: [4.8649, 14.79508]

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$
$$\bar{x}_1 - \bar{x}_2 \pm t_c s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Here, the 95% confidence interval for the difference in mean hen production between diet A and diet B numbers above. Thus, not only do we estimate the difference to be 9.83 mg/dl, but we are 95% confident it is no less than lower bound or greater than upper bound.

CODE

```
dietA = c(166,174,150,166,165,178)
dietB = c(158,159,142,163,161,157)

y = t.test(dietA , dietB)
y
n1=0
h1=NULL
y1= c(166,174,150,166,165,178, 158,159,142,163,161,157)
c1=c(rep("A", 6), rep("B", 6))
diff=166.5 - 156.6667
d1 = rep(0,10000)
for(i in 1:10000){
  c2=sample(c1);
  x1=y1[c2=="A"];
  x2=y1[c2=="B"];
  m1 = mean(x1);
  m2 = mean(x2);
  d1[i] = m2-m1;
  h1=c(h1,d1[i]);
  if(abs(d1[i])>=9.83)n1=n1+1
}
n1
hist(h1, main="Randomization Distribution")
pvalue= n1/10000
pvalue
```

```
> dietA = c(166,174,150,166,165,178)
> dietB = c(158,159,142,163,161,157)
>
> y = t.test(dietA , dietB)
> y
```

Welch Two Sample t-test

```
data: dietA and dietB
t = 1.9735, df = 9.436, p-value = 0.07842
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.359600 21.026267
sample estimates:
mean of x mean of y
166.5000 156.6667
```

```
> n1=0
> h1=NULL
> y1= c(166,174,150,166,165,178, 158,159,142,163,161,157)
> c1=c(rep("A", 6), rep("B", 6))
```



```

> diff=166.5 - 156.6667
> d1 = rep(0,10000)
> for(i in 1:10000){
+ c2=sample(c1);
+ x1=y1[c2=="A"];
+ x2=y1[c2=="B"];
+ m1 = mean(x1);
+ m2 = mean(x2);
+ d1[i] = m2-m1;
+ h1=c(h1,d1[i]);
+ if(abs(d1[i])>=9.83)n1=n1+1
+ }
> n1
[1] 913
> hist(h1, main="Randomization Distribution")
> pvalue= n1/10000
> pvalue
[1] 0.0913

```

