

1. In the MASS data set named immer, the barley yield in years 1931 and 1932 of the same field are recorded. The yield data are presented in the data frame columns Y1 and

Y2.

```
> library(MASS)      # load the MASS package
> head(immer)
  Loc Var  Y1  Y2
1  UF  M 81.0 80.7
2  UF  S 105.4 82.3
.....
```

Problem

Assuming that the data in immer follows the normal distribution, find the 95% confidence interval estimate of the difference between the mean barley yields between years 1931 and 1932.

```
> t.test(immer$Y1, immer$Y2, paired=T)
```

Paired t-test

```
data: immer$Y1 and immer$Y2
t = 3.324, df = 29, p-value =
0.002413
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 6.121954 25.704713
sample estimates:
mean of the differences
      15.91333
```

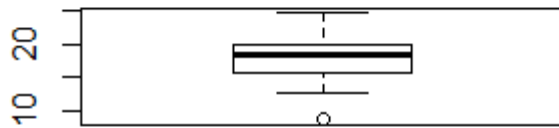
95%신뢰구간은 6.12~25.7 이다.

2 Load the UsingR dataset homework. This measures study habits of students from private and public high schools. Make a side-by-side boxplot. Use the appropriate test to test for equality of centers.

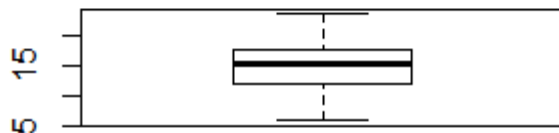
```
> data(homework)
> head(homework)
```

	Private	Public
1	21.3	15.3
2	16.8	17.4
3	8.5	12.3
4	12.6	10.7
5	15.8	16.4
6	19.3	11.3

```
> boxplot(homework$Private)
```



```
> boxplot(homework$Public)
```



```
> t.test(homework$Private,homework$Public)
```

Welch Two Sample t-test

```
data: homework$Private and homework$Public
t = 1.7134, df = 27.727, p-value =
0.09779
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
-0.5345123  5.9878456
sample estimates:
mean of x mean of y
17.63333  14.90667
```

귀무가설: 두 집단의 평균은 같다.

대립가설: 두 집단의 평균은 같지 않다.

P-value 가 0.09 로 유의수준 5% 이하이므로 대립가설을 채택하여 두 집단의 평균이 같지 않음을 알 수 있다.

3 Load the UsingR data set corn. Twelve plots of land are divided into two and then one half of each is planted with a new corn seed, the other with the standard. Do a two-sample t-test on

the data. Do the assumptions seem to be met. Comment why the matched sample test is more appropriate, and then perform the test. Did the two agree anyways?

```
> data(corn)
```

```
> head(corn)
```

	New	Standard
1	110	102
2	103	86
3	95	88
4	94	75
5	87	89
6	119	102

```
> t.test(corn$New, corn$Standard, paired = T)
```

Paired t-test

data: corn\$New and corn\$Standard

t = 3.8308, df = 11, p-value =
0.00279

alternative hypothesis: true difference in means is not equal
to 0

95 percent confidence interval:

3.297258 12.202742

sample estimates:

mean of the differences
7.75

귀무가설: 두 집단의 평균은 같다.

대립가설: 두 집단의 평균은 다르다.

P-Value 가 0.002 로 유의수준 5% 이하이므로 두 집단의 평균이 다르다는 대립가설을 채택한다.

4 Load the UsingR dataset blood. Do a significance test for equivalent centers. What was the p-value?

```
> data(blood)
```

```
> head(blood)
```

	Machine	Expert
1	68	72
2	82	84
3	94	89
4	106	100
5	92	97
6	80	88

```
> t.test(blood$Machine, blood$Expert, paired = T)
```

Paired t-test

```

data: blood$Machine and blood$Expert
t = 0.68162, df = 14, p-value =
0.5066
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
-2.146615 4.146615
sample estimates:
mean of the differences
1

```

귀무가설: 두 집단의 평균은 같다.

대립가설: 두 집단의 평균은 차이가 있다.

P-value 는 귀무가설이 참일때 현재와 같은 자료 내지 대립가설쪽으로 치우친 결과가 나올 확률을 말하며 여기서는 0.5066 으로 유의수준 5% 이상이므로 귀무가설을 채택한다.

5 미국의 빈곤률은 2000 년에 11.3% 였다. 2001 년에 조사해 보니 5 만명 중 5850 명이 빈곤하였다. 즉, 11.7%. 미국의 빈곤률은 증가하였는가?

```
> prop.test(x=5850, n=50000, p=0.113, alt='greater')
```

```

1-sample proportions test with
continuity correction

```

```

data: 5850 out of 50000, null probability 0.113
X-squared = 7.9417, df = 1, p-value
= 0.002415
alternative hypothesis: true p is greater than 0.113
95 percent confidence interval:
0.1146464 1.0000000
sample estimates:
p
0.117

```

귀무가설: 2000 년과 2001 년의 빈곤률은 같거나 2001 의 빈곤률이 줄었다.

대립가설: 2001 년의 빈곤률이 증가했다.

P-value 가 0.002 이므로 유의수준 5% 이하이다. 그러므로 대립가설을 채택하여 빈곤률이 증가했음을 알 수 있다.

6. 자동차 명세서에 연비가 17miles per gallon 으로 표시. 실제 테스트해 보니까,
mpg=c(11.4, 13.1, 14.7, 14.7, 15.0, 15.5, 15.6, 15.9, 16.0, 16.8)

이였다. 마일리지가 실제 17 마일 이하인가?

```

t.test(mpg, mu=17, alt='less')
> t.test(mpg, mu=17, alt='less')

```

One Sample t-test

```

data: mpg
t = -4.2847, df = 9, p-value =
0.001018

```

```
alternative hypothesis: true mean is less than 17
95 percent confidence interval:
 -Inf 15.78127
sample estimates:
mean of x
14.87
```

귀무가설: 마일리지는 17 과 같거나 크다.

대립가설: 마일리지는 17 보다 작다.

P-value 가 0.001 으로 유의수준 5%보다 낮으므로 대립가설을 채택한다. 그러므로 마일리지가 17 보다 작다는 사실을 알 수 있다.

7. 모비율의 차이 검정. 2001 년 5 만명 중 11.7%가 빈곤. 2002 년 6 만명 중 12.1%. 빈곤률이 증가하였는가?

```
phat=c(0.117, 0.121)
n=c(50000, 60000)
prop.test(n*phat, n, alt='less')
> prop.test(n*phat, n, alt='less')
```

```
2-sample test for equality of
proportions with continuity
correction
```

```
data: n * phat out of n
X-squared = 4.1187, df = 1, p-value
= 0.02121
alternative hypothesis: less
95 percent confidence interval:
 -1.00000000000 -0.0007588993
sample estimates:
prop 1 prop 2
0.117 0.121
```

귀무가설: 2001 년의 빈곤률과 2002 년의 빈곤률이 같거나 2001 년이 더 높다.

대립가설: 2002 년의 빈곤률이 증가했다.

p-value 가 0.02 로 유의수준 5%이하이므로 2002 년의 빈곤률이 증가했다는 대립가설을 채택한다.

8.

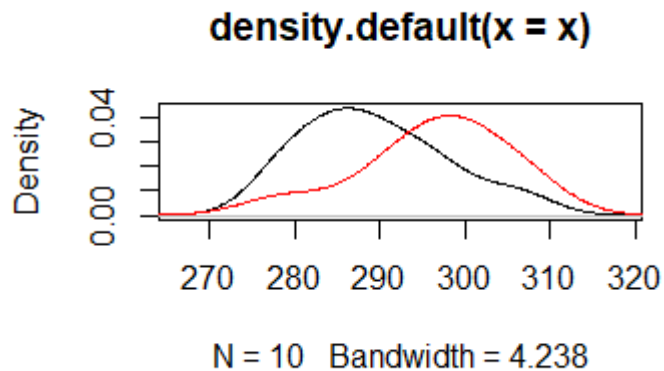
```
x=c(284, 279, 289, 292, 287, 295, 285, 279, 306, 298)
y=c(298, 307, 297, 279, 291, 335, 299, 300, 306, 291)
```

1) x 와 y 의 density 함수를 그려라.

```
plot(density(x),col=1)
```

```
lines(density(y),col=2)
```

plot 은 새 그림을 그리고, lines 는 있는 그림에 그림을 추가한다.



2) 분산이 동일하다고 가정하고 평균에 차이가 있는지 검정.

```
> t.test(x,y,var.equal = T)
```

Two Sample t-test

```
data: x and y
t = -2.034, df = 18, p-value =
0.05696
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
-22.1584072 0.3584072
sample estimates:
mean of x mean of y
289.4 300.3
```

분산이 동일하다고 가정할 때 x의 평균은 289.4 이고 y의 평균은 300.3 이다.

3) 평균에 차이가 있는지 검정

```
> t.test(x,y)
```

Welch Two Sample t-test

```
data: x and y
t = -2.034, df = 14.509, p-value =
0.06065
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
-22.3557409 0.5557409
sample estimates:
mean of x mean of y
289.4 300.3
```

x의 평균은 289.4 이며 y의 평균은 300.3 이다.

x와 y가 분산이 동일하다고 가정할 때와 아닐 때의 평균의 차이는 없다.

9. 두 벡터는 동일한 사람이 강의 전과 강의 후의 점수를 비교한 것이다. 강의가 점수를 높이는 데 도움을 주었는가?

```
pre = c(77,56,64,60,57,53,72,62,65,66)
post = c(88,74,83,68,58,50,67,64,74,60)
> t.test(pre,post,paired = T, alternative = 'less')
```

Paired t-test

```
data: pre and post
t = -1.8904, df = 9, p-value =
0.04564
alternative hypothesis: true difference in means is less than
0
95 percent confidence interval:
      -Inf -0.1635821
sample estimates:
mean of the differences
      -5.4
```

귀무가설: 강의 후의 점수가 더 낮거나 차이가 없다.

대립가설: 강의 후의 점수가 증가했다.

P-value의 값이 0.045으로 유의수준 5%이하이므로 대립가설을 채택한다.