Week3

PSTAT 122

4/12/2022

Two-Sample t-Test

1. Unpaired Two-Sample t-Test Suppose there are two professors, you would like to know if two professors have same interests in using memes. You record the number of memes for them.

$$x : 3, 7, 11, 0, 7, 0, 4, 5, 6, 2$$

$$y : 4, 7, 8, 0, 10, 11, 6, 9, 11, 0$$

```r
x = c(3,7,11,0,7,0,4,5,6,2)
y = c(4,7,8,0,10,11,6,9,11,0)

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

```
##
##  Two Sample t-test
##
## data:  x and y
## t = -1.2382, df = 18, p-value = 0.2316
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.663319  1.463319
## sample estimates:
## mean of x mean of y
##       4.5       6.6
```

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

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = -1.2382, df = 17.451, p-value = 0.2321
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.671364  1.471364
## sample estimates:
## mean of x mean of y
##       4.5       6.6
```

2. Paired Two-Sample t-Test Suppose we use two machines, A and B, to measure the weight of 9 pieces of metal material and get 9 pairs of observations:

$$A : 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00$$

$$B : 0.10, 0.21, 0.52, 0.32, 0.78, 0.59, 0.68, 0.77, 0.89$$

We want to test $H_0 : \mu_A = \mu_B$ using a paired Two-Sample t-Test.

$$Z_i = X_i - Y_i, \quad i = 1, ..., 9$$

It is equivalent to the test $H_0 : \mu_z = 0$. Suppose $Z_i \sim N(0, \sigma^2)$, then the test statistic is defined as

$$T = \frac{\bar{Z} - 0}{S_z/\sqrt{n}} \overset{H_0}{\sim} t_{n-1}$$

```
x=c(0.20,0.30,0.40,0.50,0.60,0.70,0.80,0.90,1.00)
y=c(0.10,0.21,0.52,0.32,0.78,0.59,0.68,0.77,0.89)
t.test(x, y, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  x and y
## t = 1.4673, df = 8, p-value = 0.1805
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.034299  0.154299
## sample estimates:
## mean of the differences
##                    0.06
```

Two-Sample F-Test

The Two-Sample F-Test tests the null hypothesis that two samples come from two independent populations having the equal variances.

$$H_0 : \sigma_2^2/\sigma_1^2 = 1 \quad vs. \quad H_1 : \sigma_2^2/\sigma_1^2 \neq 1$$

1. F distribution $X_1 \sim \chi^2_m$, $X_2 \sim \chi^2_n$. X_1 and X_2 are independent variables. Then, $F = \frac{X_1/m}{X_2/n}$ follows a F distribution with degree of freedoms m and n.

2. μ_1 and μ_2 are given $\mathbf{X} = (X_1, ..., X_m)$ and $\mathbf{Y} = (Y_1, ..., Y_n)$ are random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. The test statistic is

$$F_* = S^2_{2*}/S^2_{1*} \overset{H_0}{\sim} F_{n,m}$$

where

$$S^2_{1*} = \frac{1}{m}\sum_{i=1}^{m}(X_i - \mu_1)^2$$

$$S^2_{2*} = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu_2)^2$$

We reject the H_0 if $F_* > F_{n,m}(\alpha/2)$ or $F_* < F_{n,m}(1 - \alpha/2)$.

3. μ_1 and μ_2 are unknown The test statistic is

$$F = S_2^2/S_1^2 \overset{H_0}{\sim} F_{n-1,m-1}$$

where

$$S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2$$
$$S_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$$

We reject the H_0 if $F > F_{n-1,m-1}(\alpha/2)$ or $F < F_{n-1,m-1}(1-\alpha/2)$.

4. Suppose $\mathbf{X} = (x_1, ..., x_6)$ and $\mathbf{Y} = (y_1, ..., y_6)$ are random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Two variables are independent and all parameters are unknown. $\bar{X} = 0.14$, $\bar{Y} = 0.139$, $S_1 = 0.0026$ and $S_2 = 0.0024$. Does $\mu_1 = \mu_2$ hold?

Note: This problem is a test of the means of two independent normal distributions, but we first need to determine if $\sigma_1^2 = \sigma_2^2$.

$$H_0 : \sigma_2^2/\sigma_1^2 = 1 \quad vs. \quad H_1 : \sigma_2^2/\sigma_1^2 \neq 1, \qquad \alpha = 0.05$$

Here $F = S_1^2/S_2^2 = 0.0026^2/0.0024^2 = 1.17$ and $F_{m-1,n-1}(\alpha/2) = F_{5,5}(\alpha/2) = 7.15$. Since $F < 7.15$, we failed to reject H_0. Then, consider the following test

$$H_0 : \mu_1 = \mu_2 \quad vs. \quad H_1 : \mu_1 \neq \mu_2, \qquad \alpha = 0.05, \sigma_1^2 = \sigma_2^2$$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_w^2}{n_x} + \frac{S_w^2}{n_y}}} \overset{H_0}{\sim} t_{n_x+n_y-2}$$

$n_x = n_y = 6$, $S_w^2 = \frac{(n_x-1)S_x^2+(n_y-1)S_y^2}{n_x+n_y-2} = 6.26 \times 10^{-6}$. So,

$$|T| = 0.6928 < t_{10}(\alpha/2) = t_{10}(0.025) = 2.228.$$

We need to accept the null hypothesis $\mu_1 = \mu_2$.

```r
x = c(-3,7,11,0,-7,0,4,5,6,2)
y = c(-4,7,8,0,10,-11,6,9,11,0)
var.test(x, y, alternative = "two.sided", conf.level=0.95)
```

5. The var.test( )

```
##
##  F test to compare two variances
##
## data:  x and y
## F = 0.53774, num df = 9, denom df = 9, p-value = 0.3691
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.133567 2.164938
## sample estimates:
## ratio of variances
##            0.53774
```

```
# Plot a F distribution
m = 12
n = 4
input = seq(0, 10, 1/100)
y = df(input, df1=m, df2=n)
plot(input, y, xlab="X",ylab="f(x)", main = "density of a F distribution", type="l")
abline(v=qf(c(0.025, 0.975), df1=m, df2=n), col=c("blue","blue"))
```