

```
/*-----*/
/* 6장 가설검정 */
/*-----*/
/* 모평균에 대한 가설검정 */
/*-----*/
/* 하나의 모집단에서 모평균에 대한 가설검정 */
/*-----*/
/* 대립가설 (H1) 귀무가설 (H0) 한쪽검정 양쪽검정 */
/*-----*/

DATA Ex6;
/* 가설설정 */

run;

/*-----*/
/* proc ttest data=(데이터) h0=(귀무가설) sides=(l 좌측, u 우측) */
/* 유의수준 > p-value => H0 reject , H1 */
/* 유의수준 < p-value => H0 accept , H0 */
/*-----*/
/* 방향을 안정하면 양측 검정이다. */
```

```
data Ex6_1;
input height @@;
cards;
169 168 164 165 159 158 158 169 160 162
163 160 160 164 160 161 164 160 161 167
158 164 160 169 161 167 169 168 164 161
;

proc ttest data=Ex6_1 H0=163 ;
var height;
run; /* H0 */
```

SAS 시스템

The TTEST Procedure

Variable: height

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|----|-------|---------|---------|---------|---------|
| 30 | 163.1 | 3.6986 | 0.6753 | 158.0 | 169.0 |

| Mean | 95% CL Mean | Std Dev | 95% CL Std Dev |
|-------|-------------|---------|----------------|
| 163.1 | 161.7 164.5 | 3.6986 | 2.9456 4.9720 |

| DF | t Value | Pr > t |
|----|---------|---------|
| 29 | 0.15 | 0.8833 |

```
/*-----*/
/* 두 모집단에서 모평균의 차이에 대한 가설검정 */
/*-----*/
/* 분산의 동일성 검정 */
/* H0 : 분산이 같다 , H1 : 분산이 다르다 */
```

<모평균 검정>

위 결과에 따라 다음의 검정통계량을 적용한다.

i) 두 모분산이 같을 때(동일 모분산 귀무가설 채택)

통계량 : $T = \frac{\bar{x} - \bar{y}}{s_{pooled} \sqrt{\frac{1}{n} + \frac{1}{m}}}$ 여기서 s_{pooled} 는 합동분산 추정량으로 다음과 같다.

$$s_{pooled}^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$$

그리고 이 때 자료들이 정규분포를 따른다는 가정을 만족할 경우 검정통계량은 자유도 $(n+m-2)$ 를 갖는다.

ii) 두 모분산이 같지 않을 때(동일 모분산 귀무가설 기각)

통계량 : $T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$

$$df = \frac{(s_x^2/n + s_y^2/m)^2}{(s_x^2/n)^2/(n-1) + (s_y^2/m)^2/(m-1)}$$

이 때 자료들이 정규분포를 따른다는 가정을 만족할 경우 자유도는 i)의 경우와 조금 다른 값을 갖는데,

$$(\bar{x} - \bar{y}) \sim T(\mu_1 - \mu_2, \sigma_p \sqrt{\frac{1}{n} + \frac{1}{m}})$$
$$t = \frac{(\bar{x} - \bar{y})}{\sigma_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$
$$t = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

/*-----*/

```
data Ex6_2;
input process $ duration @@ ;
cards;
old 245 old 278 old 260 old 210 old 242
new 270 new 286 new 254 new 310
new 285 new 277
;
proc ttest data=Ex6_2;
    class process;
    var duration;
run;
```

/* **중요** 분산은 양측검정으로 들어가지만 평균은 단측이기 때문에 나누기 2를 해줘야한다. H0 */

/*-----*/

```
data Ex6_3;
input method $ tension @@;
cards;
tape 93 tape 112 tape 85 tape 70 tape 96
tape 97 tape 110 tape 89
tape 92 tape 85 thread 83 thread 102
thread 86 thread 65 thread 98 thread 99
thread 95 thread 102 thread 75 thread 90
;
proc ttest data=Ex6_3;
    class method;
    var tension;
run; /* H0 */
```

/*-----*/

```
data Ex6_4;
input raw $ materia @@;
cards;
a 8.1 a 7.8 a 8.1 a 7.7 a 7.9 a 8.0
a 8.1 a 7.9 a 8.4 a 8.3 a 7.8 a 8.3
a 9.9 a 8.2 a 7.9 b 8.0 b 7.6 b 7.9
b 7.6 b 7.7 b 8.0 b 7.9 b 8.1 b 7.6
b 7.7 b 7.6 b 8.0 b 7.5
;
proc ttest data=Ex6_4;
    class raw;
    var materia;
run; /* H1 */
```

/*-----*/

| SAS 시스템 | | | | | | |
|---------------------|---|---------|---------|---------|---------|---------|
| The TTEST Procedure | | | | | | |
| Variable: duration | | | | | | |
| process | N | Mean | Std Dev | Std Err | Minimum | Maximum |
| new | 6 | 280.3 | 18.6833 | 7.6274 | 254.0 | 310.0 |
| old | 5 | 247.0 | 25.1396 | 11.2428 | 210.0 | 278.0 |
| Diff (1-2) | | 33.3333 | 21.7902 | 13.1946 | | |

| process | Method | Mean | 95% CL Mean | Std Dev | 95% CL Std Dev |
|------------|---------------|---------|----------------|---------|-----------------|
| new | | 280.3 | 260.7 299.9 | 18.6833 | 11.6623 45.8230 |
| old | | 247.0 | 215.8 278.2 | 25.1396 | 15.0620 72.2401 |
| Diff (1-2) | Pooled | 33.3333 | 3.4850 63.1817 | 21.7902 | 14.9881 39.7805 |
| Diff (1-2) | Satterthwaite | 33.3333 | 1.4679 65.1988 | | |

| Method | Variances | DF | t Value | Pr > t |
|---------------|-----------|--------|---------|---------|
| Pooled | Equal | 9 | 2.53 | 0.0324 |
| Satterthwaite | Unequal | 7.2935 | 2.45 | 0.0425 |

| Equality of Variances | | | | |
|-----------------------|--------|--------|---------|--------|
| Method | Num DF | Den DF | F Value | Pr > F |
| Folded F | 4 | 5 | 1.81 | 0.5285 |

| Method | Variances | DF | t Value | Pr > t |
|---------------|-----------|--------|---------|---------|
| Pooled | Equal | 18 | 0.62 | 0.5440 |
| Satterthwaite | Unequal | 17.999 | 0.62 | 0.5440 |

| Equality of Variances | | | | |
|-----------------------|--------|--------|---------|--------|
| Method | Num DF | Den DF | F Value | Pr > F |
| Folded F | 9 | 9 | 1.02 | 0.9815 |

| Method | Variances | DF | t Value | Pr > t |
|---------------|-----------|--------|---------|---------|
| Pooled | Equal | 26 | 2.43 | 0.0225 |
| Satterthwaite | Unequal | 18.647 | 2.56 | 0.0192 |

| Equality of Variances | | | | |
|-----------------------|--------|--------|---------|--------|
| Method | Num DF | Den DF | F Value | Pr > F |
| Folded F | 14 | 12 | 6.62 | 0.0023 |

```
data Ex6_5;
input x @@;
cards;
20 18 23 20 22 19 25 23 24 15
;
proc ttest data=Ex6_5 h0=20 sides=u;
var x;
run; /* H0 */

/*-----*/
/* 독립이 아닌 대응인 표본일때 */
```

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|----|---------|---------|---------|---------|---------|
| 10 | 20.9000 | 3.0714 | 0.9713 | 15.0000 | 25.0000 |

| Mean | 95% CL Mean | Std Dev | 95% CL Std Dev |
|---------|---------------|---------|----------------|
| 20.9000 | 19.1196 Infty | 3.0714 | 2.1126 5.6071 |

| DF | t Value | Pr > t |
|----|---------|--------|
| 9 | 0.93 | 0.1891 |

- 그러나 자료가 한 개체의 쌍을 이루므로, 가설 검정을 위한 과정에서는 한 개체의 두 자료
값의 차이

$$d_i = x_i - y_i$$

를 구하여, 차이의 평균값 μ_d 를 '0'과 비교하는 일표본 t-test를 실시한다. 차이의 평균값 μ_d 를 사용하여 가설을 표현하면 다음과 같다.

$$\begin{matrix} H_0: \mu_d = 0 \\ H_1: \mu_d > 0 \end{matrix}$$

또는

$$\begin{matrix} H_0: \mu_d = 0 \\ H_1: \mu_d < 0 \end{matrix}$$

(한쪽 검정)

또는

$$\begin{matrix} H_0: \mu_d = 0 \\ H_1: \mu_d \neq 0 \end{matrix}$$

(양쪽 검정)

/*-----*/

```
data Ex6_6;
input pretest nextttest @@;
cards;
80 82 73 71 70 95 60 69 88 100
84 71 65 75 37 60 91 95 98 99
52 65 78 83 40 60 79 86 59 62
;
proc ttest data=Ex6_6;
paired pretest*nextttest ;
run; /* H1 */
```

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|----|---------|---------|---------|----------|---------|
| 15 | -7.9333 | 9.9317 | 2.5643 | -25.0000 | 13.0000 |

| Mean | 95% CL Mean | Std Dev | 95% CL Std Dev |
|---------|------------------|---------|----------------|
| -7.9333 | -13.4333 -2.4334 | 9.9317 | 7.2712 15.6632 |

| DF | t Value | Pr > t |
|----|---------|---------|
| 14 | -3.09 | 0.0079 |

/*-----*/

```
data Ex6_7;
input m10 m12 @@;
cards;
0.04 0.06 0.12 0.10 0.08 0.11
0.06 0.10 0.05 0.06 0.09 0.11
0.03 0.06
;
proc ttest data=Ex6_7;
paired m10*m12;
run; /* **중요** 단측이기 때문에 나누기 2를 해줘야 한다. H1 */
```

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|---|---------|---------|---------|---------|---------|
| 7 | -0.0186 | 0.0195 | 0.00738 | -0.0400 | 0.0200 |

| Mean | 95% CL Mean | Std Dev | 95% CL Std Dev |
|---------|------------------|---------|----------------|
| -0.0186 | -0.0366 -0.00052 | 0.0195 | 0.0126 0.0430 |

| DF | t Value | Pr > t |
|----|---------|---------|
| 6 | -2.52 | 0.0454 |

/*-----*/

```
data Ex6_8;
input pretest nextttest @@;
cards;
1.45 0.19 0.79 0.15 0.93 1.19 1.35 0.67 0.80 0.39
0.41 1.58 1.56 0.91 1.93 0.28 0.33 0.58 1.42 0.68
2.83 1.28 1.01 2.02 2.48 0.36
;
proc ttest data=Ex6_8;
    paired pretest*nextttest ;
run; /* **중요** 단측이기 때문에 나누기 2를 해줘야 한다. H1 */

/*-----*/
/* 분산분석 */
/*-----*/
/* 세집단 이상의 평균비교 */
/*-----*/
```

| N | Mean | Std Dev | Std Err | Minimum | Maximum |
|----|--------|---------|---------|---------|---------|
| 13 | 0.5392 | 0.9982 | 0.2768 | -1.1700 | 2.1200 |

| Mean | 95% CL Mean | Std Dev | 95% CL Std Dev |
|--------|----------------|---------|----------------|
| 0.5392 | -0.0640 1.1424 | 0.9982 | 0.7158 1.6477 |

| DF | t Value | Pr > t |
|----|---------|---------|
| 12 | 1.95 | 0.0752 |

```
data Ex6_9;
input fertil $ yield @@;
cards;
F1 148 F1 76 F1 134 F1 98 F2 166 F2 153 F2 255 F3 254
F3 214 F3 327 F3 304 F4 335 F4 436 F4 423 F4 380 F4 465
;
proc anova data=Ex6_9;
    class fertil;
    model yield = fertil;
run; /* H1 */

/*-----*/
/* 분산 다중 비교 */
/*-----*/
/* 평균값이 같지 않다는 결과가 나왔을대 사용 , 예 data 6_9 */
/* 차이가 많이 날땐 tukey, 차이가 적게 날땐 duncan */
/* 정확도는 duncan이 높다, 미세한 요인까지 고려하기 때문 */
```

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 3 | 209499.7833 | 69833.2611 | 30.47 | <.0001 |
| Error | 12 | 27498.2167 | 2291.5181 | | |
| Corrected Total | 15 | 236998.0000 | | | |

| R-Square | Coeff Var | Root MSE | yield Mean |
|----------|-----------|----------|------------|
| 0.883973 | 18.37612 | 47.86980 | 260.5000 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| fertil | 3 | 209499.7833 | 69833.2611 | 30.47 | <.0001 |

```
data Ex6_10;
set Ex6_9;
proc anova data=Ex6_9;
    class fertil;
    model yield = fertil;
    means fertil/tukey duncan;
run;

/*-----*/
/* 이원분산분석, 2가지의 가설 분석을 각요인에 한번씩 2번한다. */
/*-----*/
/* 반복이 없는 경우 */
/*-----*/
```

| Comparisons significant at the 0.05 level are indicated by ***. | | | |
|---|--------------------------|------------------------------------|-------------|
| fertil Comparison | Difference Between Means | Simultaneous 95% Confidence Limits | |
| F4 - F3 | 133.05 | 37.72 | 228.38 *** |
| F4 - F2 | 216.47 | 112.68 | 320.25 *** |
| F4 - F1 | 293.80 | 198.47 | 389.13 *** |
| F3 - F4 | -133.05 | -228.38 | -37.72 *** |
| F3 - F2 | 83.42 | -25.13 | 191.96 |
| F3 - F1 | 160.75 | 60.26 | 261.24 *** |
| F2 - F4 | -216.47 | -320.25 | -112.68 *** |
| F2 - F3 | -83.42 | -191.96 | 25.13 |
| F2 - F1 | 77.33 | -31.21 | 185.88 |
| F1 - F4 | -293.80 | -389.13 | -198.47 *** |
| F1 - F3 | -160.75 | -261.24 | -60.26 *** |
| F1 - F2 | -77.33 | -185.88 | 31.21 |

```
data Ex6_11;
do product='A1', 'A2', 'A3', 'A4' ;
  do customer=1 to 5 by 1;
    input prefer @@;
    output;
  end;
end;
cards;
5 7 9 10 8
2 3 4 5 2
4 7 6 5 7
6 4 2 2 1
;

proc anova data=Ex6_11;
  class product customer;
  model prefer= product customer;
  means product/ tukey duncan;
run; /* product H1 , customer H0 */
```

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 7 | 83.2500000 | 11.8928571 | 3.59 | 0.0253 |
| Error | 12 | 39.7000000 | 3.3083333 | | |
| Corrected Total | 19 | 122.9500000 | | | |

| R-Square | Coeff Var | Root MSE | prefer Mean |
|----------|-----------|----------|-------------|
| 0.677105 | 36.74510 | 1.818882 | 4.950000 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|----------|----|-------------|-------------|---------|--------|
| product | 3 | 78.55000000 | 26.18333333 | 7.91 | 0.0035 |
| customer | 4 | 4.70000000 | 1.17500000 | 0.36 | 0.8356 |

/*-----*/

```
data Ex6_12;
do product='기계A', '기계B', '기계C' ;
  do people='기능공1', '기능공2', '기능공3' ;
    input quality @@;
    output;
  end;
end;
cards;
11 15 20
14 23 16
11 14 15
;

proc anova data=Ex6_12;
  class product people;
  model quality= product people;
run; /* product H0, people H0 */
```

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model | 4 | 81.7777778 | 20.4444444 | 2.02 | 0.2560 |
| Error | 4 | 40.4444444 | 10.1111111 | | |
| Corrected Total | 8 | 122.2222222 | | | |

| R-Square | Coeff Var | Root MSE | quality Mean |
|----------|-----------|----------|--------------|
| 0.669091 | 20.58862 | 3.179797 | 15.44444 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|---------|----|-------------|-------------|---------|--------|
| product | 2 | 28.22222222 | 14.11111111 | 1.40 | 0.3469 |
| people | 2 | 53.55555556 | 26.77777778 | 2.65 | 0.1851 |