<5-1>

```
LIBNAME MYLIB '/folders/myfolders/mylib';
DATA MYLIB.FP;
        INPUT CR $ FP @@;
CARDS;
G 0.58 G 2.80 G 2.77 G 3.50 G 2.67 G 2.97 G 2.18
G 3.24 G 1.49 G 2.19 G 2.70 G 2.57
B 2.28 B 1.06 B 1.08 B 0.07 B 0.16 B 0.70 B 0.75
B 1.61 B 0.34 B 1.15 B 0.44 B 0.86
;
RUN;
PROC TTEST DATA=MYLIB.FP COCHRAN;
        CLASS CR;
        VAR FP;
RUN;
```

The TTEST Procedure

Variable: FP

CR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
B		12	0.8750	0.6287	0.1815	0.0700	2.2800
G		12	2.4717	0.7920	0.2286	0.5800	3.5000
Diff (1-2)	Pooled		-1.5967	0.7150	0.2919		
Diff (1-2)	Satterthwaite		-1.5967		0.2919		

CR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
B		0.8750	0.4756	1.2744	0.6287	0.4453	1.0674
G		2.4717	1.9684	2.9749	0.7920	0.5611	1.3448
Diff (1-2)	Pooled	-1.5967	-2.2020	-0.9913	0.7150	0.5530	1.0120
Diff (1-2)	Satterthwaite	-1.5967	-2.2039	-0.9895			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	22	-5.47	<.0001
Satterthwaite	Unequal	20.922	-5.47	<.0001
Cochran	Unequal	11	-5.47	0.0002

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	11	1.59	0.4558

(가) Equality of Variance 에 대한 검정으로 F value가 1.59, 유의확률(Pr 〉 F)이 0.4558(〉 0.05) 이므로 등분산 검정을 시행하여야 한다. 즉, ‘Pooled’에 대한 method에 의하여 두 모평균의 차이에 대한 유의학률이 매우 작으므로 귀무가설을 기각한다. 따라서 신용등급이 FP 지표에 영향을 준다고 결론지을 수 있다.
(나) Pooled Method에 의한 모평균 차이의 95% 신뢰구간은 (-2.2020, -0.9913) 이다. 신뢰구간에 0이 포함되지 않음을 추가적으로 확인할 수 있다.

<5-3>

```
DATA MYLIB.PEANUT;
        INPUT RAWROASTED $ PROTEIN @@;
CARDS;
RA 61 RA 60 RA 56 RA 63 RA 56 RA 63 RA 69 RA 56
RA 44 RA 61
RO 56 RO 54 RO 47 RO 59 RO 51 RO 51 RO 57 RO 54
RO 63 RO 58
;
RUN;
PROC TTEST DATA=MYLIB.PEANUT COCHRAN;
        CLASS RAWROASTED;
        VAR PROTEIN;
RUN;
```

The TTEST Procedure

Variable: PROTEIN

RAWROASTED	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
RA		10	58.9000	6.6072	2.0894	44.0000	69.0000
RO		10	55.0000	4.6188	1.4606	47.0000	63.0000
Diff (1-2)	Pooled		3.9000	5.7004	2.5493		
Diff (1-2)	Satterthwaite		3.9000		2.5493		

RAWROASTED	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
RA		58.9000	54.1735	63.6265	6.6072	4.5447	12.0622
RO		55.0000	51.6959	58.3041	4.6188	3.1770	8.4321
Diff (1-2)	Pooled	3.9000	-1.4559	9.2559	5.7004	4.3073	8.4299
Diff (1-2)	Satterthwaite	3.9000	-1.5015	9.3015			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	18	1.53	0.1434
Satterthwaite	Unequal	16.101	1.53	0.1455
Cochran	Unequal	9	1.53	0.1604

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	2.05	0.3010

T TEST 결과 등분산가정을 따르며, Pooled Method에 의한 유의확률이 0.05보다 크므로 귀무가설을 채택한다. 따라서 생 땅콩과 볶은 땅콩의 단백질 함량의 평균에는 차이가 없다고 결론지을 수 있다.

<5-6>

```
DATA MYLIB.DEER;
        INPUT hindleg foreleg @@;
        diff = hindleg - foreleg;
CARDS;
142 138  140 136  144 147  144 139  142 143
146 141  149 143  150 145  142 136  148 146
;
RUN;

PROC UNIVARIATE DATA=MYLIB.DEER NORMAL PLOT;
        VAR DIFF;
RUN;

DATA MYLIB.DEER2;
        INPUT HIND_FORE $ LEG @@;
CARDS;
H 142 H 140 H 144 H 144 H 142 H 146 H 149 H 150
H 142 H 148
F 138 F 136 F 147 F 139 F 143 F 141 F 143 F 145
F 136 F 146
;
RUN;

PROC TTEST DATA=MYLIB.DEER2 COCHRAN;
        CLASS HIND_FORE;
        VAR LEG;
RUN;
```

UNIVARIATE 프로시저
변수: diff

적률			
N	10	가중합	10
평균	3.3	관측값 합	33
표준편차	3.0568684	분산	9.34444444
왜도	-1.3466487	첨도	0.78935489
제곱합	193	수정 제곱합	84.1
변동계수	92.6323759	평균의 표준 오차	0.96666667

기본 통계 측도			
위치측도		변이측도	
평균	3.300000	표준편차	3.05687
중위수	4.500000	분산	9.34444
최빈값	5.000000	범위	9.00000
		사분위수 범위	3.00000

위치모수 검정: Mu0=0				
검정	통계량		p 값	
스튜던트의 t	t	3.413793	Pr > |t|	0.0077
부호	M	3	Pr >= |M|	0.1094
부호 순위	S	23.5	Pr >= |S|	0.0117

Univariate으로 확인한 결과 양측검정의 $p-value$가 0.0077으로 유의수준 0.05보다 작기 때문에 기각역에 존재한다. 따라서 H_0을 기각한다. 즉 '모평균은 0이 아니다.' 라고 말할 수 있다. 즉 사슴의 앞다리와 뒷다리 길이의 평균에는 차이가 있음을 알 수 있다.

The TTEST Procedure

Difference: hindleg - foreleg

N	Mean	Std Dev	Std Err	Minimum	Maximum
10	3.3000	3.0569	0.9667	-3.0000	6.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
3.3000	1.1132	5.4868	3.0569	2.1026	5.5807

DF	t Value	Pr > |t|
9	3.41	0.0077

PROC TTEST도 같은 결과를 보여준다. T 값이 3.41로 이에 대한 유의확률이 0.0077이므로 귀무가설 $H_0 : \mu_1 = \mu_2$을 기각할 수 있다. 따라서 사슴의 뒷다리가 앞다리보다 길이가 더 길다고 볼 수 있다. 또한 대립가설을 $H_1 : \mu_1 - \mu_2 > 0$ 로 하는 경우 유의확률은 0.0077/2 = 0.00385으로 마찬가지로 대립가설을 채택한다. 따라서 사슴의 뒷다리의 평균 길이가 앞다리보다 길다고 결론지을 수 있다.

<5-7>

(가)

$$\overline{x_1} = 32.5,\ \overline{x_2} = 36.4$$

$$\left(32.5 - Z_{0.0125}\frac{3}{\sqrt{20}}, 32.5 + Z_{0.0125}\frac{3}{\sqrt{20}}\right) = (30.99, 34.00)$$

$$\left(36.4 - Z_{0.0125}\frac{4}{\sqrt{25}}, 36.4 + Z_{0.0125}\frac{4}{\sqrt{25}}\right) = (34.60, 38.19)$$

(나)

$$\left((32.5 - 36.4) - Z_{0.0125}\sqrt{\frac{3^2}{20}+\frac{4^2}{25}},\ (32.5 - 36.4) + Z_{0.0125}\sqrt{\frac{3^2}{20}+\frac{4^2}{25}}\right) = (-6.24, -1.55)$$

(다) $H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 < \mu_2$

$$-3.7355 = \frac{-3.9}{\sqrt{\frac{3^2}{20}+\frac{4^2}{25}}} < -Z_{0.0125} = -2.2414$$

검정통계량이 기각치에 존재하므로 대립가설을 채택한다. 따라서 전기처리를 한 후의 철사가 유의수준 5% 하에서 평균 강도가 더 높다고 볼 수 있다. 이 때의 $p-value$는 0.0000930이다.

<5-9>

```
DATA MYLIB.RICE;
        INPUT KIND $ CROP @@;
CARDS;
A 31 A 34 A 29 A 26 A 32 A 35 A 38 A 34 A 30 A
29 A 32 A 31
B 26 B 24 B 28 B 29 B 30 B 29 B 32 B 26 B 31 B
29 B 32 B 28
;
RUN;
PROC TTEST DATA=MYLIB.RICE COCHRAN;
        CLASS KIND;
        VAR CROP;
RUN;
```

The TTEST Procedure

Variable: CROP

KIND	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
A		12	31.7500	3.1945	0.9222	26.0000	38.0000
B		12	28.6667	2.4618	0.7107	24.0000	32.0000
Diff (1-2)	Pooled		3.0833	2.8518	1.1642		
Diff (1-2)	Satterthwaite		3.0833		1.1642		

KIND	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		31.7500	29.7203	33.7797	3.1945	2.2629	5.4238
B		28.6667	27.1025	30.2308	2.4618	1.7439	4.1799
Diff (1-2)	Pooled	3.0833	0.6689	5.4978	2.8518	2.2055	4.0363
Diff (1-2)	Satterthwaite	3.0833	0.6598	5.5069			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	22	2.65	0.0147
Satterthwaite	Unequal	20.659	2.65	0.0152
Cochran	Unequal	11	2.65	0.0227

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	11	1.68	0.4009

(가) 분산의 동일성에 대한 검증을 통해 $p\text{-}value$가 0.4 이므로 등분산 가정을 사용해야 함을 알 수 있다. Pooled method에 의한 단측검정 $p\text{-}value$가 0.0147/2으로 0.05보다 작으므로 유의수준 5% 하에서 귀무가설을 기각하고 대립가설을 채택한다. 따라서 A품종이 B품종 보다 평균 수확량이 더 많다고 볼 수 있다.

(나) 위의 결과에서 Pooled method에 의한 95% 신뢰구간은 (0.6689, 5.4978) 이며 0이 이 구간에 존재하지 않음을 추가적으로 확인할 수 있다.

<5-12>

```
DATA MYLIB.DIABETES;
        INPUT ID PRERUN POSTRUN @@;
CARDS;
1 1.45 0.19 2 2.37 1.03
3 0.79 0.15 4 0.77 0.35
5 0.93 1.19 6 0.77 0.27
7 1.35 0.67 8 0.33 0.58
9 0.80 0.39 10 1.42 0.68
11 2.50 1.58 12 0.51 0.82
13 0.41 1.58 14 2.83 1.28
15 1.56 0.91 16 1.01 2.02
17 1.93 0.28 18 0.81 0.24
19 1.35 0.12 20 3.48 0.36
;
RUN;
PROC TTEST DATA=MYLIB.DIABETES;
        PAIRED PRERUN*POSTRUN;
RUN;
```

The TTEST Procedure

Difference: PRERUN - POSTRUN

N	Mean	Std Dev	Std Err	Minimum	Maximum
20	0.6340	0.9717	0.2173	-1.1700	3.1200

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.6340	0.1792	1.0888	0.9717	0.7390	1.4192

DF	t Value	Pr > $\lvert t\rvert$
19	2.92	0.0088

T 값이 2.92이고 이에 대응되는 유의확률이 0.0088이므로, 귀무가설 $H_0: \mu_1 = \mu_2$을 기각할 수 있다. 따라서 5% 유의수준 하에서 500m 달리기를 한 후의 당뇨의 양이 하기 전보다 줄어든다고 결론지을 수 있다.

<5-13>

```
DATA MYLIB.ENERGY;
        INPUT SLEEP $ ENERGY @@;
CARDS;
U 35.3 U 35.9 U 37.2 U 33.0 U 31.9 U 33.7 U 36.0
U 35.0
U 33.3 U 36.6 U 37.9 U 35.6 U 29.0 U 33.7 U 35.7
L 32.4 L 34.0 L 34.4 L 31.8 L 35.0 L 34.6 L 34.6
L 33.5
L 33.6 L 31.5 L 33.8
;
RUN;
PROC TTEST DATA=MYLIB.ENERGY COCHRAN;
        CLASS SLEEP;
        VAR ENERGY;
RUN;
```

The TTEST Procedure

Variable: ENERGY

SLEEP	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
L		11	33.5636	1.1784	0.3553	31.5000	35.0000
U		15	34.6533	2.2778	0.5881	29.0000	37.9000
Diff (1-2)	Pooled		-1.0897	1.8987	0.7537		
Diff (1-2)	Satterthwaite		-1.0897		0.6871		

SLEEP	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
L		33.5636	32.7720	34.3553	1.1784	0.8233	2.0680
U		34.6533	33.3919	35.9147	2.2778	1.6676	3.5923
Diff (1-2)	Pooled	-1.0897	-2.6453	0.4659	1.8987	1.4826	2.6414
Diff (1-2)	Satterthwaite	-1.0897	-2.5147	0.3353			

Method	Variances	DF	t Value	Pr > $\lvert t\rvert$
Pooled	Equal	24	-1.45	0.1612
Satterthwaite	Unequal	21.984	-1.59	0.1270
Cochran	Unequal	.	-1.59	0.1375

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	14	10	3.74	0.0420

(가) Equality of Variances에 대한 검정으로 유의확률 0.0420이 0.05보다 작으므로 모분산이 같다는 귀무가설을 기각해야 한다. 따라서 Method에서 Satterthwaite 혹은 Cochran를 선택해야 한다. 한편, $H_0: \sigma_1^2 \geq \sigma_2^2$, $H_1: \sigma_1^2 < \sigma_2^2$로 둘 때,

$$F = \frac{s_1^2}{s_2^2} = 0.2676 < F_{1-0.05}(11-1, 15-1) = 0.3490$$

검정통계량이 기각치에 존재하므로 대립가설을 채택한다. 따라서 5% 유의수준 하에서 7시간 이상 수면을 취하는 집단의 분산이 7시간 미만 수면을 취하는 집단보다 크다고 할 수 있다.

(나)

$$\left(\frac{1}{F_{0.0125}(n_1-1, n_2-1)}\frac{s_1^2}{s_2^2},\ F_{0.0125}(n_1-1, n_2-1)\frac{s_1^2}{s_2^2}\right)$$

$$\left(\frac{1}{F_{0.0125}(11-1, 15-1)}\frac{1.17^2}{2.27^2}, F_{0.0125}(11-1, 15-1)\frac{1.17^2}{2.27^2}\right)$$

$$\left(\frac{1}{3.7379}\frac{1.17^2}{2.27^2},\quad 3.7379\frac{1.17^2}{2.27^2}\right)$$

$$(0.071601,\quad 1.000439)$$