**SAS 데이터 분석 입문 5장**

**2019020650 김형욱**

**\* 5장 예제문제**

**<예 5-1>**

**data** sasadv.edu;

input group score @@;

cards;

1 65 1 70 1 76 1 63 1 72 1 71 1 68 1 68

2 75 2 80 2 72 2 77 2 69 2 81 2 71 2 78

;

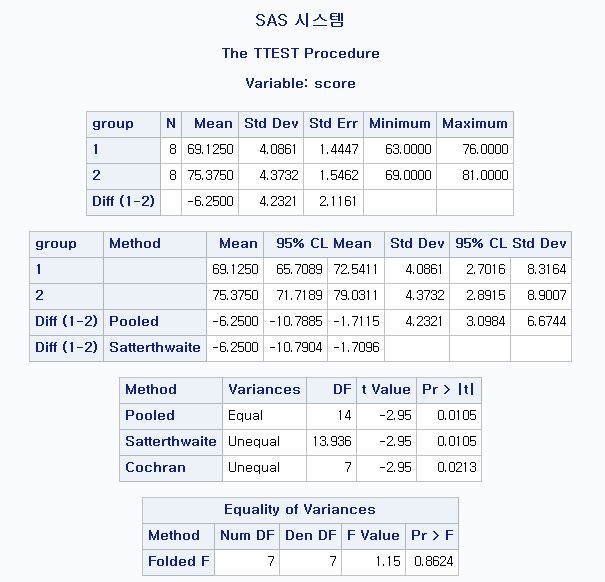
**run**;

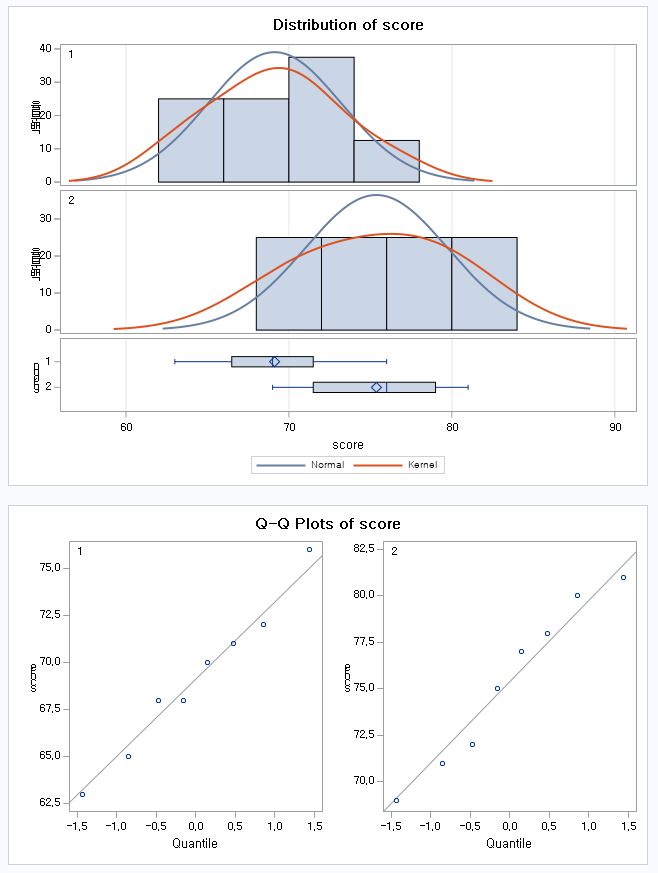
**proc** **ttest** data=sasadv.edu cochran;

class group;

var score;

**run**;





해석 : 위 문제는 두 가지 강의방법에 따른 독해력의 평균 차이를 검정하는 것으로, 귀무가설은 “새로운 강의방식이 기존의 방식보다 독서력 향상에 도움이 되지 않는다”이고 대립가설은 “새로운 강의방식이 기존의 방식보다 독서력 향상에 도움이 된다”이다. 먼저 등분산 가정(두 집단의 모분산이 같다는 가정)에 따라서 t-검정 방법이 달리 적용되어야 하므로, ‘분산의 동일성’에 대한 검정 결과를 살펴보면 F-값이 1.15이고 유의확률이 0.8624로 매우 크다. 따라서 두 집단의 모분산이 동일하다고 할 수 있다. 그러므로 ‘Pooled’에 해당하는 결과를 보면, T-값이 -2.95이고 유의확률이 0.0105이므로 유의수준 5% 하에서 귀무가설을 기각한다. 따라서 새로운 강의방식이 기존의 방식보다 독서력 향상에 도움을 준다고 할 수 있다.

**<예 5-2>**

**data** sasadv.paired;

input id pretest posttest @@;

cards;

01 80 82 02 73 71 03 70 95 04 60 69 05 88 100

06 84 71 07 65 75 08 37 60 09 91 95 10 98 99

11 52 65 12 78 83 13 40 60 14 79 86 15 59 62

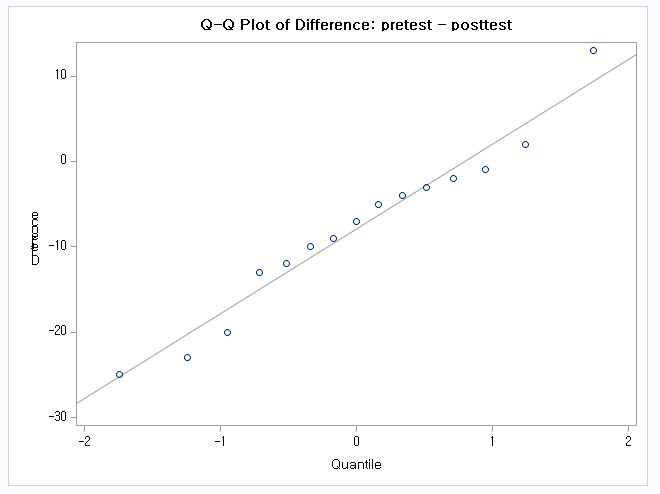
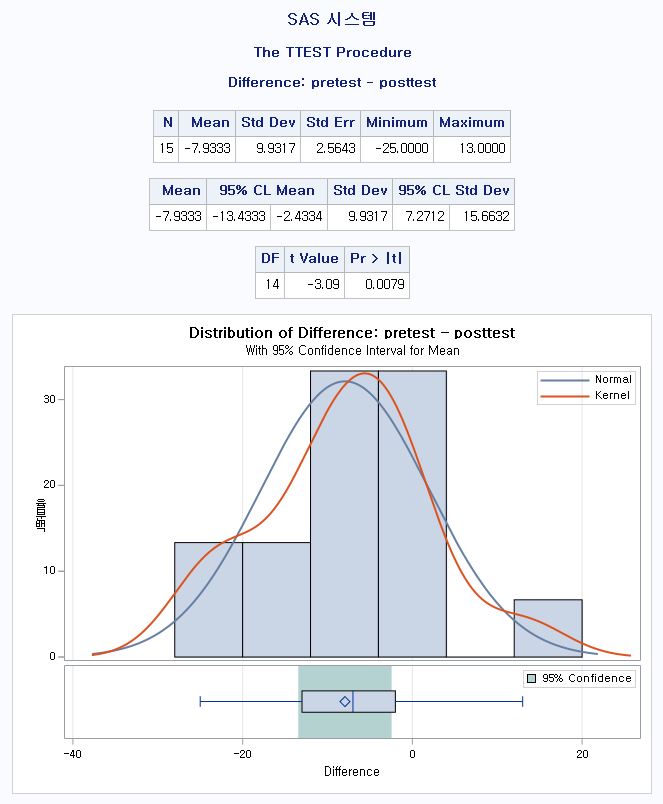
;

**run**;

**proc** **ttest** data=sasadv.paired;

paired pretest\*posttest;

**run**;



해석 : 위 문제는 학생 15명에게 컴퓨터 교육을 실시하기 전의 통계학 시험 성적(pretest)과 교육실시 후의 성적(posttest)간에 차이가 있는지를 확인하려는 것이다. 귀무가설은 “교육을 실시하기 전과 후의 성적간 모평균에 차이가 없다”이고, 대립가설은 “교육을 실시하기 전과 후의 성적간 모평균에 차이가 있다”이다. 검정 결과, T-값이 -3.09이고 유의확률이 0.0079이므로 유의수준 5% 하에서 귀무가설을 기각한다. 따라서 교육을 실시하기 전과 후의 성적간 모평균에 차이가 있다고 할 수 있는데, “95% CL Mean”의 하한 값이 -13.4333이므로 교육을 실시한 후의 성적이 평균적으로 13점 정도 높다는 것을 알 수 있다.

**<예 5-3>**

**data** sasadv.support;

input gender $ yesno $ count @@;

cards;

남자 YES 110 남자 NO 140

여자 YES 104 여자 NO 96

;

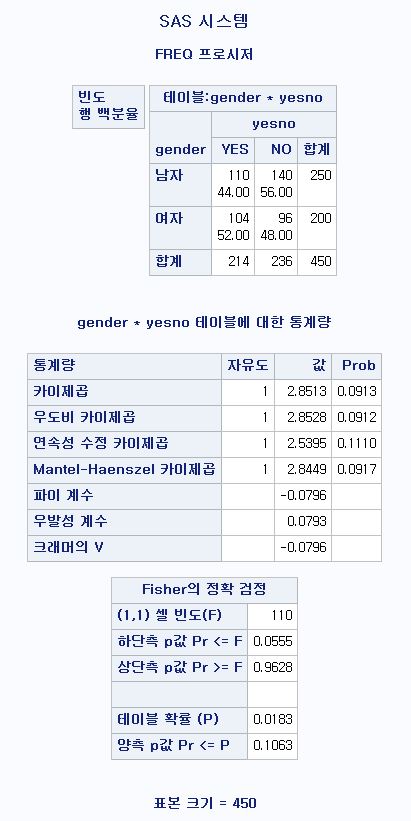
**run**;

**proc** **freq** data=sasadv.support order=data;

weight count;

tables gender\*yesno / nocol nopercent chisq fisher;

**run**;



해석 : 위 문제는 현 정부에 대한 지지율이 성인 남녀별로 차이가 있는가를 알아보기 위한 조사이다. 귀무가설은 “성인 남녀별로 현정부에 대한 지지율이 차이가 없다”이고, 대립가설은 “성인 남녀별로 현정부에 대한 지지율이 차이가 있다”이다. 카이제곱검정 결과, 카이제곱 통계량 값이 2.8513이고 유의확률이 0.0913이므로 유의수준 5% 하에서 귀무가설을 기각하지 못한다. 따라서 성인 남녀별로 현정부에 대한 지지율이 차이가 없다고 할 수 있다.

**<예 5-4>**

**data** sasadv.mcpaired;

input pre $ post $ count @@;

cards;

YES YES 63 YES NO 4

NO YES 21 NO NO 12

;

**run**;

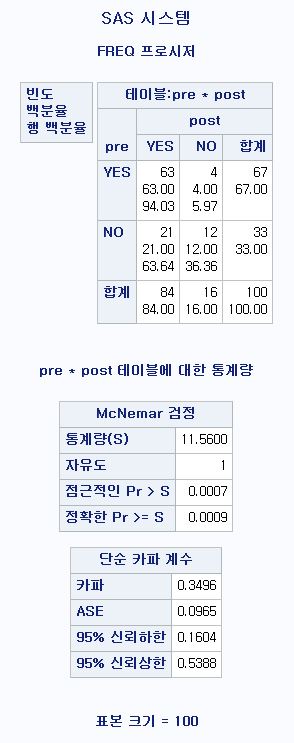
**proc** **freq** data=sasadv.mcpaired order=data;

weight count;

exact mcnem;

tables pre\*post / nocol;

**run**;



해석 : 위 문제는 정부에서 어떤 정책을 발표하기 이전과 이후에 정부에 대한 지지율이 변하였는지를 알아보기 위한 조사이다. 귀무가설은 “정책을 발표하기 이전과 이후에 정부에 대한 지지율이 차이가 없다”이고, 대립가설을 “정책을 발표하기 이전과 이후에 정부에 대한 지지율이 차이가 있다”이다. McNemar 검정 결과를 살펴보면, McNemar의 카이제곱 통계량 값이 11.56이고 근사 유의확률이 0.0007이므로 유의수준 5% 하에서 귀무가설을 기각한다. 따라서 정책을 발표하기 이전과 이후의 정부에 대한 지지율이 달라졌다고 할 수 있다.

**\* 5장 연습문제**

**<연습문제 5-2>**

**data** sasadv.ex5\_2;

input machine $ mercury;

cards;

A 0.95

A 0.82

A 0.78

A 0.96

A 0.71

A 0.86

A 0.99

B 0.89

B 0.91

B 0.94

B 0.91

B 0.90

B 0.89

;

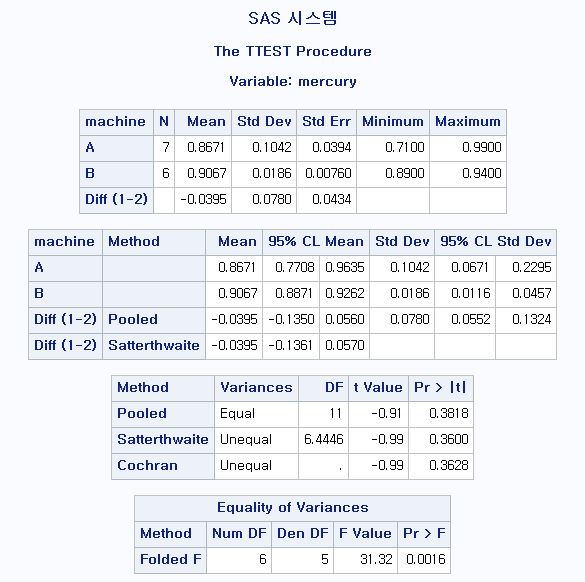
**run**;

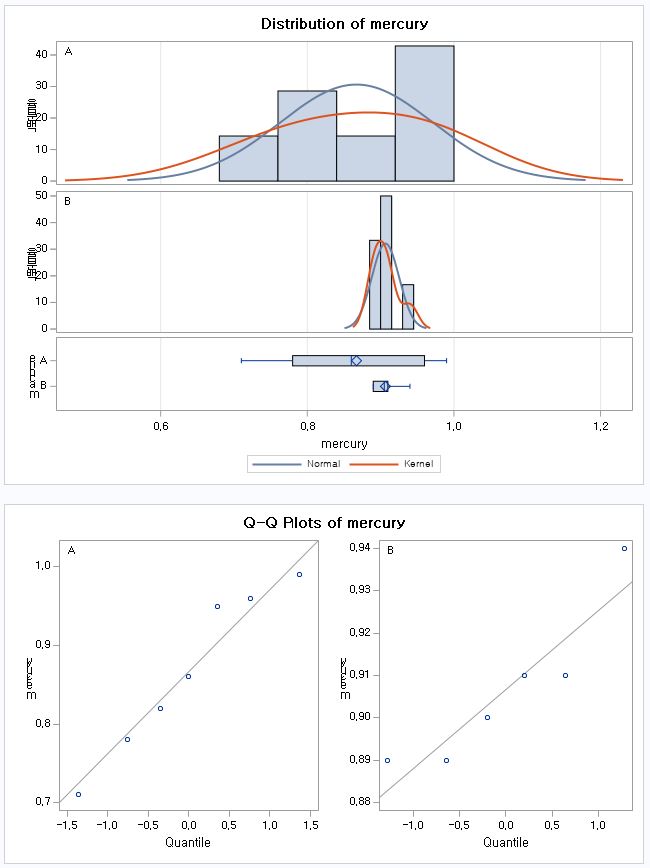
**proc** **ttest** data=sasadv.ex5\_2 cochran;

class machine;

var mercury;

**run**;





해석 : 위 문제에서 귀무가설은 “두 모집단의 분산이 동일하다”이고, 대립가설은 “두 모집단의 분산이 동일하지 않다”이다. 등분산 가정(두 집단의 모분산이 같다는 가정)에 대한 검정 결과를 살펴보면, F-값이 31.32이고 유의확률이 0.0016이므로 유의수준 5% 하에서 귀무가설을 기각한다. 따라서 등분산 가정(두 집단의 모분산이 같다는 가정)을 만족하지 않는다고 할 수 있다. (모분산이 동일하지 않기 때문에 Cochran과 Cox의 근사방법을 이용해야 한다)

**<연습문제 5-4>**

**data** sasadv.ex5\_4;

length contract $ **10.**;

input insurance $ contract $ count @@;

cards;

A 총계약자 2350 A 계약파기자 580

B 총계약자 5210 B 계약파기자 1500

;

**run**;

**proc** **freq** data=sasadv.ex5\_4 order=data;

weight count;

tables insurance\*contract / nocol nopercent chisq fisher;

**run**;



해석 : 위 문제는 두 보험에 대한 계약파기비율이 다르다고 할 수 있는지에 대한 검정이다. 귀무가설은 “두 보험에 대한 계약파기비율이 같다”이고, 대립가설은 “두 보험에 대한 계약파기비율이 다르다”이다. 검정 결과, 카이제곱 통계량 값이 7.8956이고 유의확률이 0.0050이므로 유의수준 5% 하에서 귀무가설을 기각한다. 따라서 두 보험에 대한 계약파기비율이 다르다고 할 수 있다.

**<연습문제 5-6>**

**data** sasadv.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

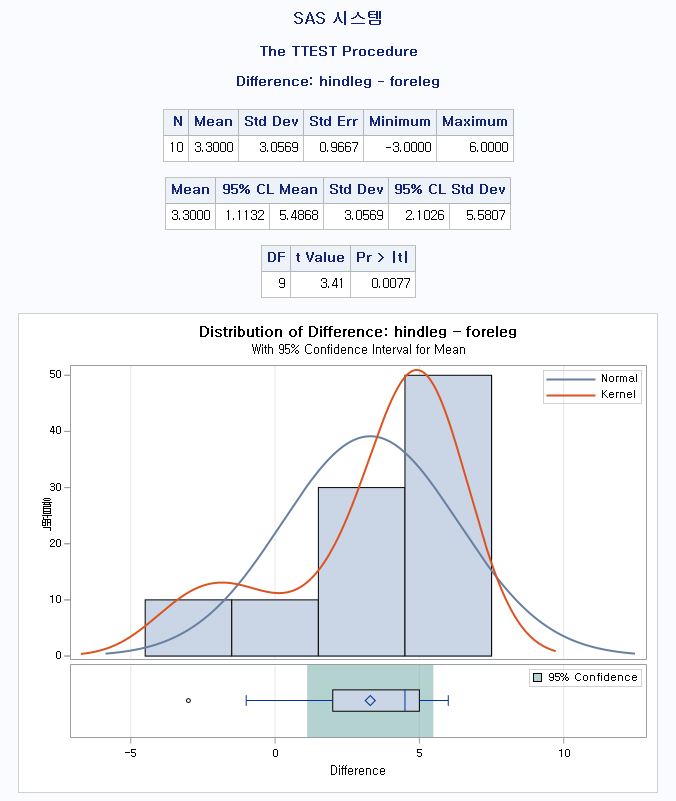
;

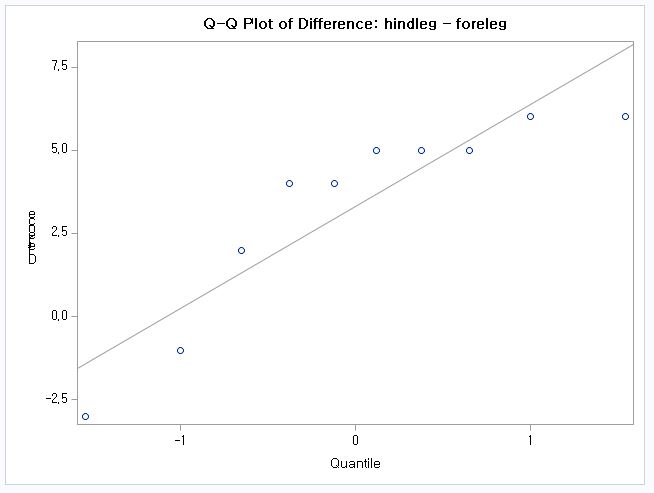
**run**;

**proc** **ttest** data=sasadv.deer;

paired hindleg\*foreleg;

**run**;





해석 : 위 문제는 사슴의 앞다리와 뒷다리 길이의 평균에 차이가 있는지를 검정하는 문제이다. 귀무가설은 “사슴의 앞다리와 뒷다리 길이의 평균에 차이가 없다”이고, 대립가설은 “사슴의 앞다리와 뒷다리 길이의 평균에 차이가 있다”이다. 검정 결과, T-값이 3.41이고 유의확률은 0.0077이므로 유의수준 5% 하에서 귀무가설을 기각한다. 따라서 사슴의 앞다리와 뒷다리 길이의 평균에 차이가 있다고 할 수 있다. 또한 검정 결과가 연습문제 4-2의 결과와 일치함을 확인할 수 있다.

**<연습문제 5-8>**

이 문제는 두 자동차의 연료소비량은 같다고 말할 수 있는지에 대한 검정 문제이다. 귀무가설은 “두 자동차의 연료소비량은 같다”이고, 대립가설은 “두 자동차의 연료소비량은 다르다”이다. 그리고 주어진 값들을 보면, DRW000017a425e0 이므로 검정통계량 값은 다음과 같다.

DRW000017a425e5

검정통계량 값이 1.248로 임계치인 1.96보다 작으므로, 유의수준 5% 하에서 귀무가설을 기각하지 못한다. 따라서 두 자동차의 연료소비량은 같다고 할 수 있다.