**2022학년도 1학기 중간과제물(온라인제출용)**

**교과목명 : 통계패키지**

**학 번 : 202135-368864**

**성 명 : 홍원표**

**연 락 처 : 101-5343-4341**

**과제유형(공통형/지정형) : 공통형**

- 이하 과제 작성

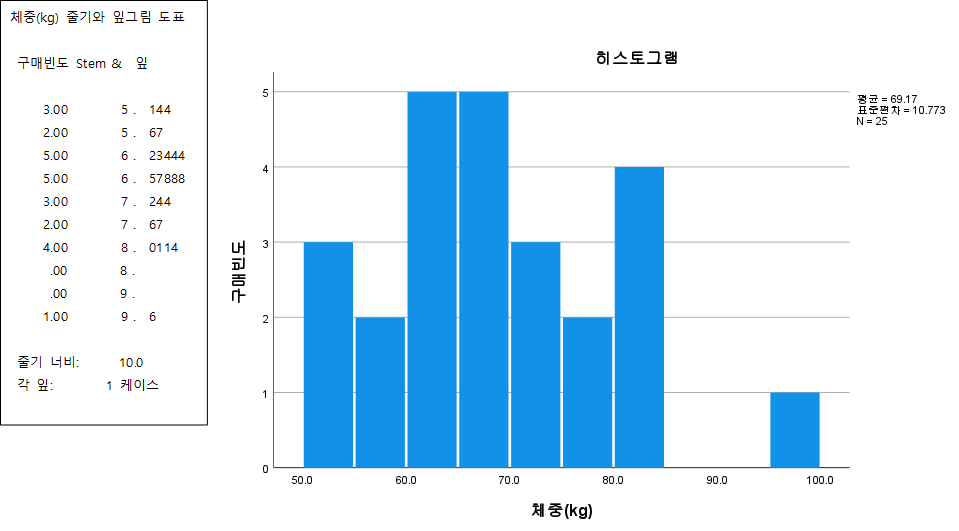
1. (15점) 다음을 SPSS를 이용하여 작성하시오.

1) 교재 5장 연습문제 7번

**(5장)7.** 진천선수촌에서 올림픽에 대비하여 연습하고 있는 육상선수 중에서 25명을 뽑아 체격과 50m 달리기 기록을 수집하여 보니 다음과 같았다. 성별은 (1) 남자, (2) 여자이고, 키와 하체길이는 cm로, 체중은 kg으로, 50m 달리기는 초 단위로 측정한 것이다.

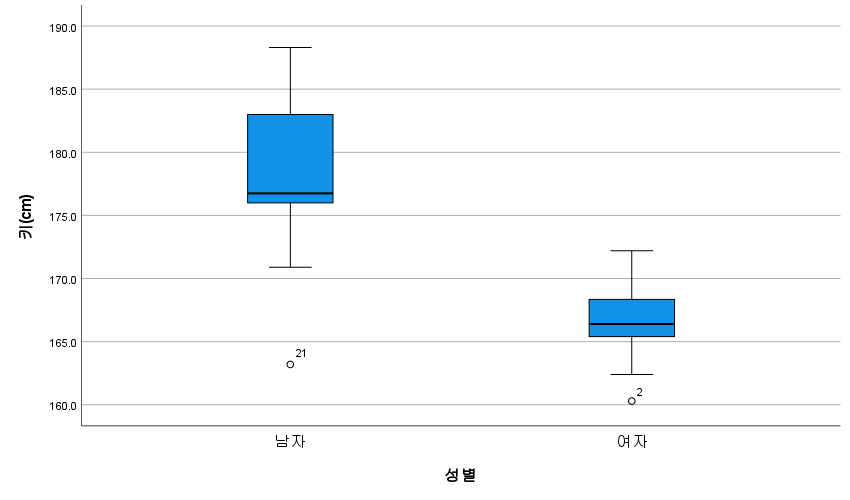
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 개인번호 | | 성별 | 키 | 체중 | 하체길이 | 50m 달리기 |
| 1 | 10010 | 1 | 184.0 | 76.4 | 101.6 | 6.17 |
| 2 | 10012 | 2 | 160.3 | 57.2 | 90.2 | 6.87 |
| 3 | 10015 | 1 | 179.3 | 74.2 | 99.4 | 6.39 |
| 4 | 10017 | 1 | 176.2 | 68.2 | 97.1 | 6.77 |
| 5 | 10021 | 2 | 166.4 | 56.6 | 91.0 | 6.93 |
| 6 | 10023 | 2 | 168.0 | 64.8 | 92.9 | 7.15 |
| 7 | 10044 | 1 | 177.0 | 67.5 | 103.6 | 7.68 |
| 8 | 10055 | 2 | 162.4 | 51.2 | 95.0 | 7.50 |
| 9 | 10059 | 1 | 170.9 | 65.8 | 79.5 | 6.70 |
| 10 | 10060 | 1 | 188.3 | 77.5 | 103.1 | 6.58 |
| 11 | 10065 | 1 | 174.3 | 64.2 | 102.7 | 6.39 |
| 12 | 10070 | 2 | 171.7 | 62.6 | 99.6 | 6.92 |
| 13 | 10072 | 1 | 185.3 | 80.8 | 101.2 | 6.38 |
| 14 | 10074 | 2 | 165.5 | 64.5 | 93.5 | 6.91 |
| 15 | 10079 | 2 | 172.2 | 81.6 | 97.5 | 7.35 |
| 16 | 10080 | 2 | 168.6 | 68.0 | 94.0 | 7.12 |
| 17 | 10090 | 1 | 176.0 | 81.3 | 95.6 | 6.55 |
| 18 | 10093 | 2 | 168.1 | 72.3 | 95.4 | 7.26 |
| 19 | 10096 | 2 | 165.9 | 54.1 | 92.6 | 6.96 |
| 20 | 10101 | 1 | 183.0 | 84.0 | 98.4 | 6.48 |
| 21 | 10103 | 1 | 163.2 | 63.0 | 86.7 | 6.84 |
| 22 | 10118 | 1 | 176.5 | 68.3 | 102.6 | 6.00 |
| 23 | 10123 | 2 | 165.3 | 54.7 | 96.5 | 7.48 |
| 24 | 10125 | 1 | 180.9 | 96.0 | 103.5 | 6.71 |
| 25 | 10126 | 1 | 176.5 | 74.4 | 95.1 | 6.73 |

(1) 체중의 줄기-잎 그림과 히스토그램을 그리고 설명하라.

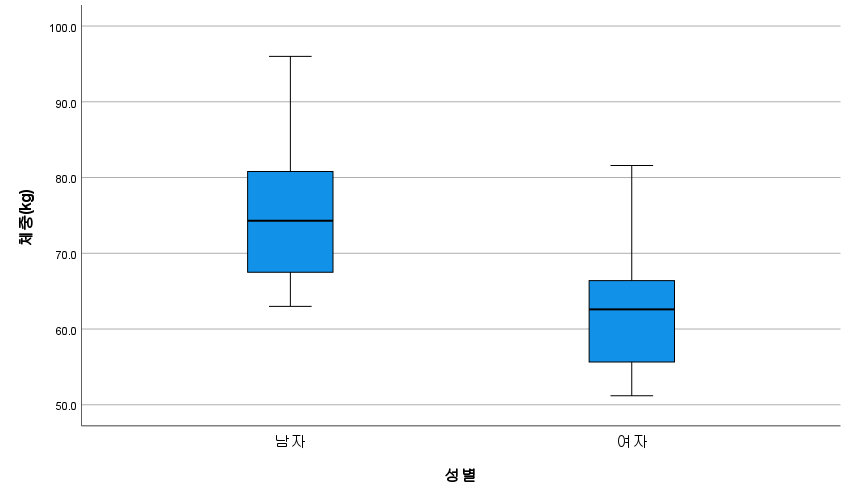


체중은 평균은 69.17이고 표준편차 10.0773으로 분포되어 있고 체중이 많이 나가는 학생이 한 명 있다.

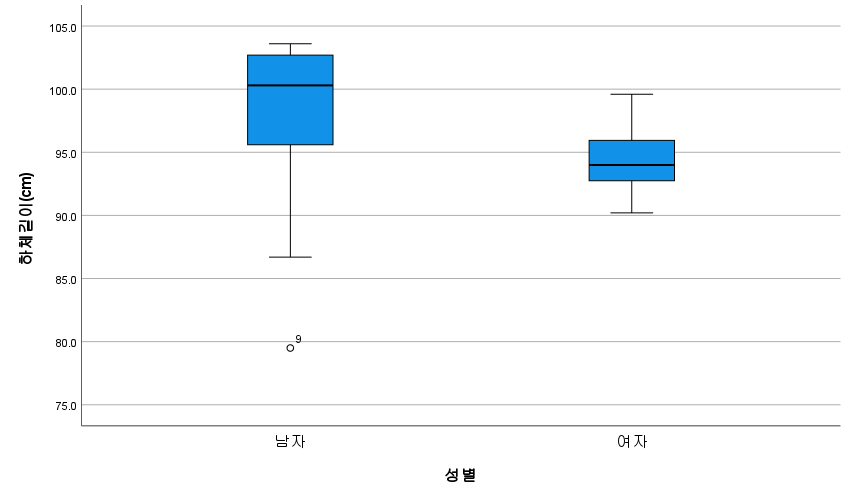
(2) 남자와 여자별로 키, 하체길이, 체중, 50m 달리기의 상자그림을 그리고 비교하라.



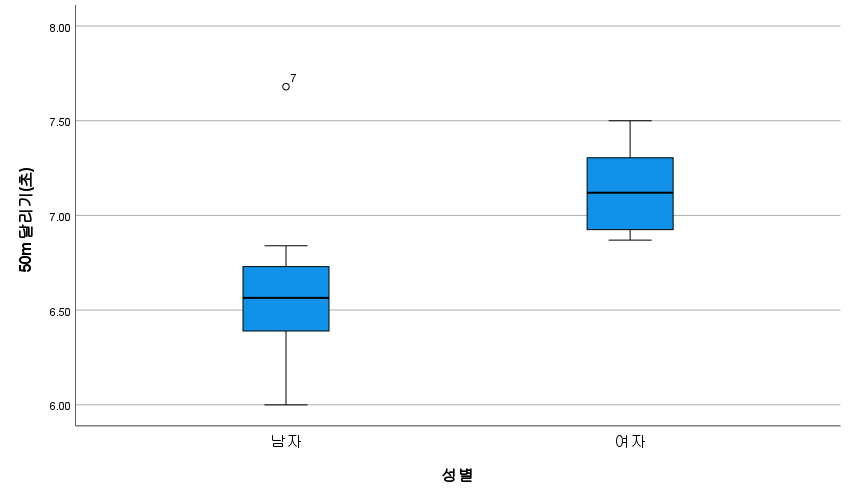
남자의 키는 여자보다 크고 넓게 분포하고 있고 키가 165이하의 결측치 데이터도 존재하기 때문에 1사분위수에 가깝게 평균이 위치하게 된 것 같다.



남녀의 체중도 남자의 체중이 여자보다는 많이 나가지만 체중의 분포는 비슷한 형태로 나타난다.



하체의 길이는 남자가 여자보다 더 길고 분포가 넓게 분포하고 있다. 남자의 경우에는 결측치(outlier)값이 하나 포함되어 있다.

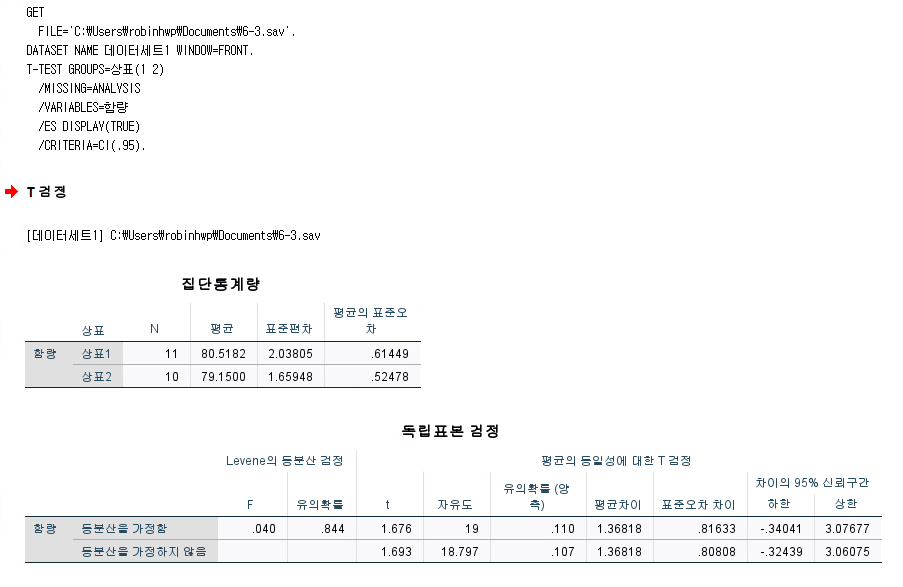


달리기는 남자가 여자보다 빠르고 평균보다 더 빠른 남자들이 분포하고 많이 느린 결측치 값이 존재한다. 반면 여자는 남자보다 느리고 대부분 평규에서 차이가 많이 나지는 않는다.

2) 교재 6장 연습문제 3,4번

**(6장)3.** 어떤 화학 약품의 제조에 상표가 다른 두 종류의 원료가 사용되고 있다. 각 원료에서 주성분 A의 함량은 다음과 같다. 단, 함량은 정규분포를 따른다고 가정한다. 이 두 원료의 주성분 A의 함량이 다른지를 분석하라.

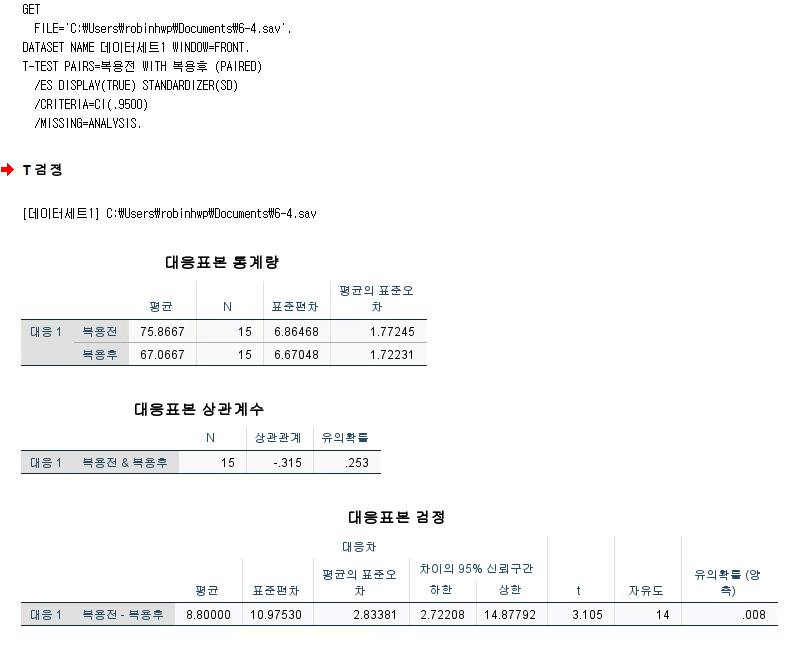
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 상표1 | 80.4 | 78.2 | 80.1 | 77.1 | 79.6 | 80.4 | 81.6 | 79.9 | 84.4 | 80.9 | 83.1 |
| 상표2 | 80.1 | 81.2 | 79.5 | 78.0 | 76.1 | 77.0 | 80.1 | 79.9 | 78.8 | 80.8 |  |



유의확률이 0.844로 값이 크므로 두 집단의 분산이 동일하다는 가정을 기각할 수 없다는 것을 알수 있다. 따라서 분산이 동일하다는 가설을 받아들이고 ‘등분산을 가정함’의 결과를 이용한다. 검정 통계량 t값은 1.676, 이에 대한 유의 확률은 0.110으로 유의수준 0.05보다 크므로 두 집단의 모평균이 같다는 귀무가설을 기각하지 못한다.

**(6장)4.** 특정 피임약이 사용자의 혈압을 저하시키는지 조사하고자 한다. 이를 위해 부인 15명을 대상으로 평상시 혈압을 측정한 뒤, 이들에게 이 피임약을 일정 기간 복용하게 한 후 이들의 혈압을 다시 측정한 결과를 기록했다. 얻어진 데이터는 다음과 같다. 피임약 복용이 혈압에 영향을 주는지 분석하라.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 부인 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 복용 전 | 70 | 80 | 72 | 76 | 76 | 76 | 72 | 78 | 82 | 64 | 74 | 92 | 74 | 68 | 84 |
| 복용 후 | 68 | 72 | 62 | 70 | 58 | 66 | 68 | 52 | 64 | 72 | 74 | 60 | 74 | 72 | 74 |



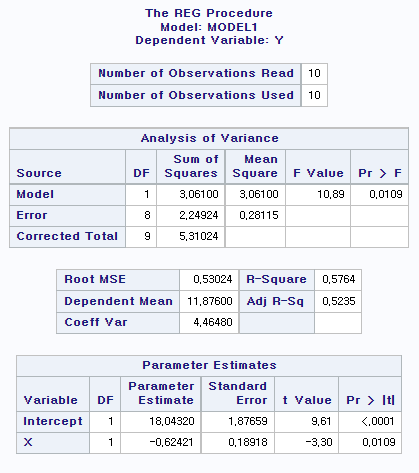
대응표본 검정결과에서 검정통계량 t값이 3.105이고 유의확률 p-값이 0.008로 유의수준 0.05보다 작으므로 차이가 없다는 귀무가설을 기각한다. 따라서 피임약을 복용하는 경우 혈압에 차이가 있다고 판단된다.

2. (15점) 다음을 SAS를 이용하여 작성하시오.

1) 교재 8장 연습문제 5번, 6번

**(8장)5.** 다음 데이터에서 X는 콩 10개의 수분 함유량을 나타내고, Y는 콩의 강도(strength)를 나타낸다.

|  |  |  |  |
| --- | --- | --- | --- |
| 수분 함유량 X | 강도 Y | 수분 함유량 X | 강도 Y |
| 11.1 | 11.14 | 9.9 | 12.60 |
| 8.9 | 12.74 | 10.7 | 11.13 |
| 8.8 | 13.13 | 10.5 | 11.70 |
| 8.9 | 11.51 | 10.5 | 11.02 |
| 8.8 | 12.38 | 10.7 | 11.41 |



DATA dat1;

INPUT X Y;

DATALINES;

11.1 11.14

8.9 12.74

8.8 13.13

8.9 11.51

8.8 12.38

9.9 12.60

10.7 11.13

10.5 11.70

10.5 11.02

10.7 11.41

RUN;

PROC gplot data=dat1;

plot X\*Y;

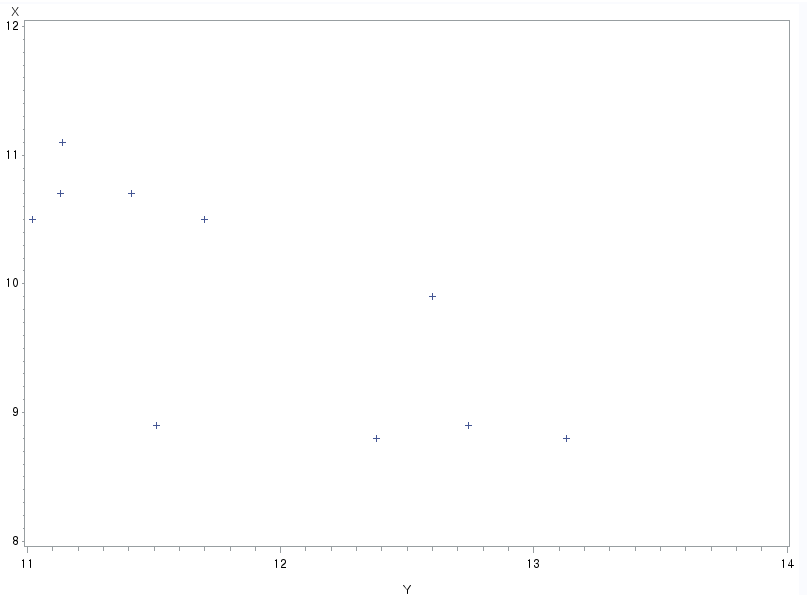
RUN;

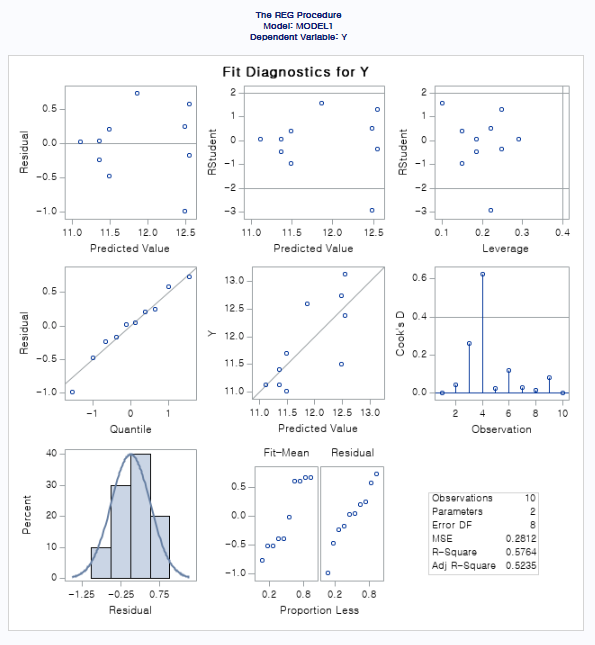
proc reg data=dat1;

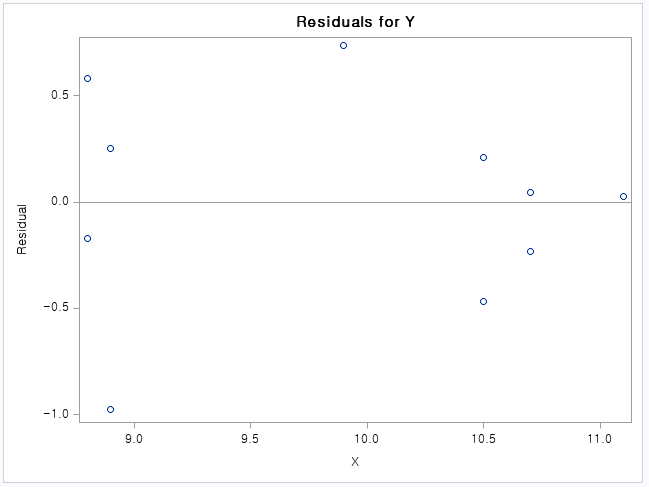
model Y = X;

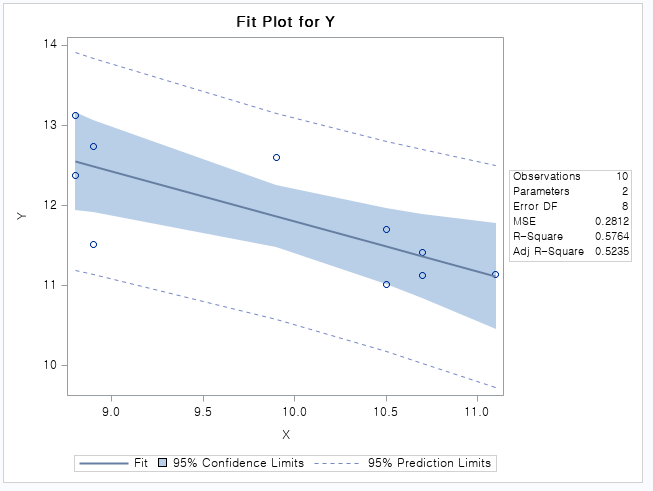
RUN;

산점도









(1) X에 대한 Y의 회귀식을 구하라

**Y = 18.04 -0.62X**

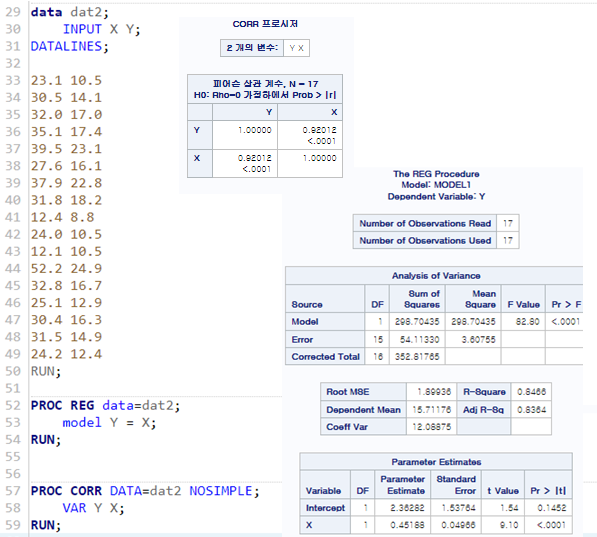
(2) 잔차를 분석하라.

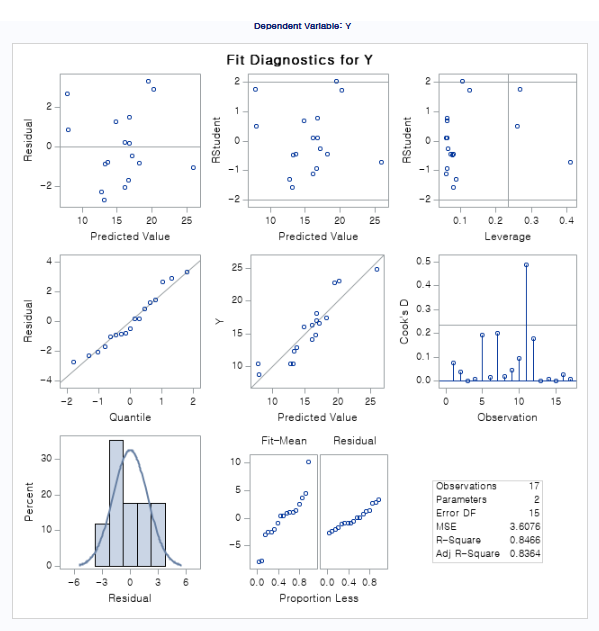
회귀식은 이고, 자유도는 1이고 총데이터는 10개이고 결정계수는 0.5764이므로 58%의 설명력이 있다.

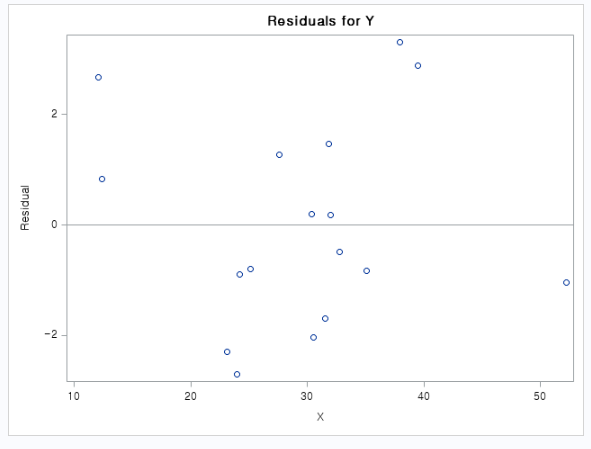
P-값 0.0109가 0.05보다 작기 때문에 귀무가설(을 기각한다.

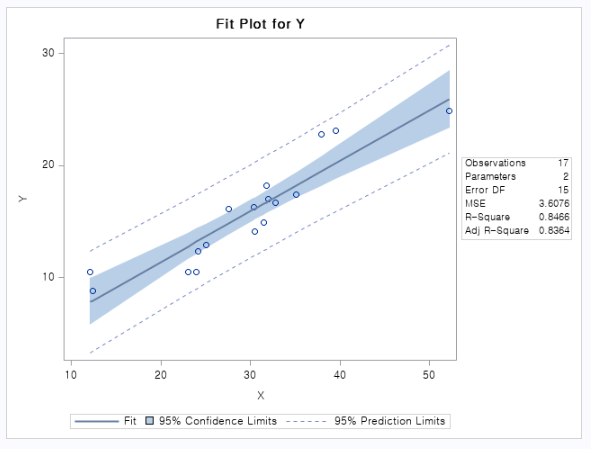
**(8장)6.** 다음 데이터에서 변수 X는 매년 4월 1일에 관측한 눈 속에 포함된 수분 함유량(단위 : %), Y는 그 해 4월부터 7월까지 용수량(water yield, 단위 : t)이다. 적절한 회귀분석을 하라.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | Y | X | Y | X | Y |
| 23.1 | 10.5 | 37.9 | 22.8 | 32.8 | 16.7 |
| 30.5 | 14.1 | 31.8 | 18.2 | 25.1 | 12.9 |
| 32.0 | 17.0 | 12.4 | 8.8 | 30.4 | 16.3 |
| 35.1 | 17.4 | 24.0 | 10.5 | 31.5 | 14.9 |
| 39.5 | 23.1 | 12.1 | 10.5 | 24.2 | 12.4 |
| 27.6 | 16.1 | 52.2 | 24.9 |  |  |









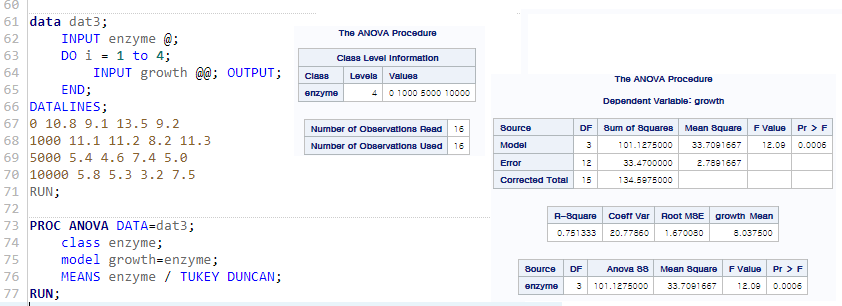
결정계수는 0.8466으로 회귀모형이 매우 적합하다는 것을 알 수 있다. F 검정 통계량이 82.80이고 이에 대한 유의확률 p-값이 0.001보다 작다. 따라서 귀무가설(을 기각하게 되며, 주어진 회귀식이 매우 유의하다는 것이다.

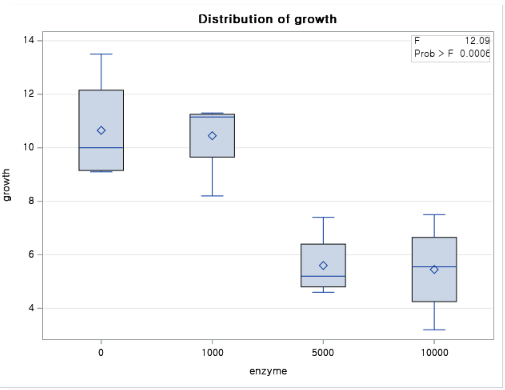
회귀식은 이다.

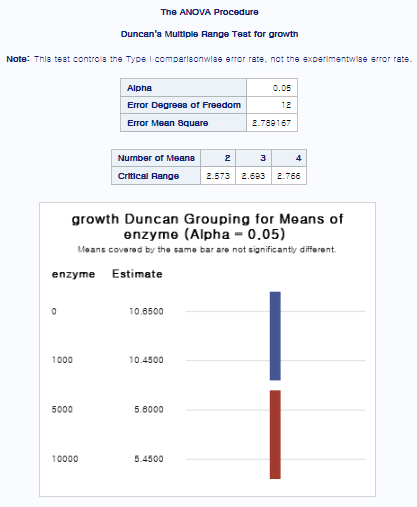
2) 교재 9장 연습문제 2번

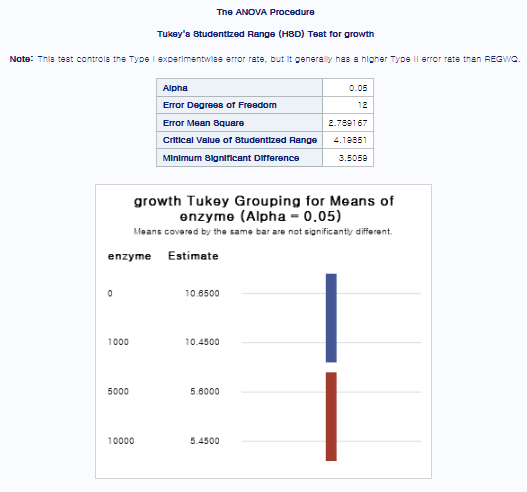
**(9장)2.** 효소 성분의 비료가 효소 함량이 증가함에 따라 토마토 모종 포기의 성장에 어떤 영향을 주는지 실험연구를 하여 다음 자료를 얻었다. 분산분석을 하여 효소 함량에 따라 토마토 모종 성장에 차이가 나는지 검정하라.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 효소 함량 | 모종 성장 | | | |
| 0 | 10.8 | 9.1 | 13.5 | 9.2 |
| 1000 | 11.1 | 11.2 | 8.2 | 11.3 |
| 5000 | 5.4 | 4.6 | 7.4 | 5.0 |
| 10000 | 5.8 | 5.3 | 3.2 | 7.5 |









분류변수 enzyme수준 수는 4 입력된 수준값은 0, 1000, 5000, 10000이고 입력 관측 총 수는 16이다.

유의확률은 0.0005로 일반적인 유의수준 0.05보다도 아주 작음으로 ‘효소함량에 따라 모종의 성장에 차이가 있다’고 결론을 내린다. 결정계수는 0.751333으로 효소 함량에 따른 모종성장 변화 패턴을 약 75%정도 설명하고 있다.

-- 통계패키지 중간과제물 끝 --