2022학년도 1학기 기말시험(과제물)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **교과목명** | : | 통계패키지 |
|  | **학번** | : | 202135-368864 |
|  | **성명** | : | 홍원표 |
|  | **연락처** | : | 010-5343-4341 |

- 이하 작성

1. (25점) 다음을 SAS를 이용하여 작성하시오.

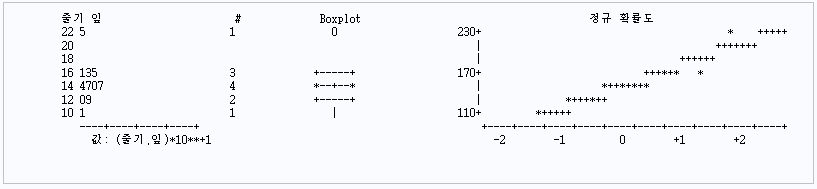
1) 교재 5장 연습문제(p.208) 2, 10번

**(5장)2.** 다음은 상장된 금융사 중 외국기업(F)과 국내기업(K)의 주가이다. 외국기업과 국내기업별로 기술통계량을 구하고 줄기-잎 그림, 상자그림을 그려서 비교 • 분석하라.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| F 120  K 165  K 147  F 144 | K 135  K 161  K 102  K 165 | K 170  F 147  F 235  F 161 | F 139  F 150  F 157  K 173 | K 114  F 175  K 129  F 111 | F 163  K 145  K 129  K 145 |
|  |  |  |  |  |  |

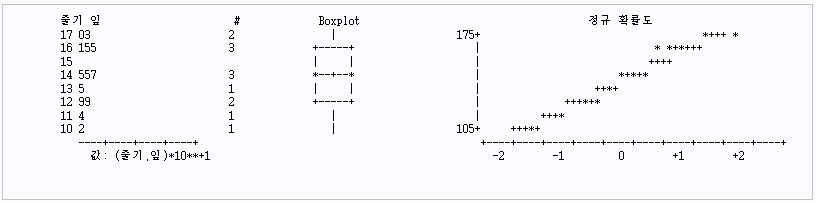


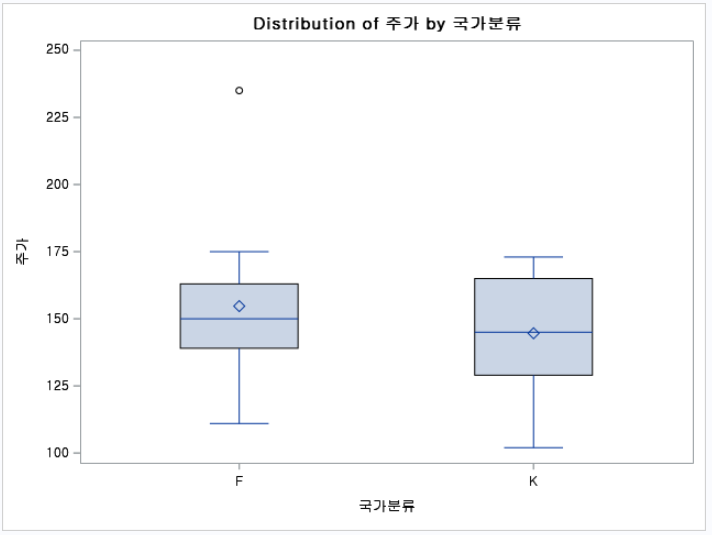




해외 주가 데이터에는 결측치로 보이는 동떨어져 있는 값이 존재한다. 따라서 결측치로 인한 평균값이 높아졌을 것이다. 결측치를 제외하고 평균을 구하면 국내의 평균과 비슷한 수준으로 확인된다.



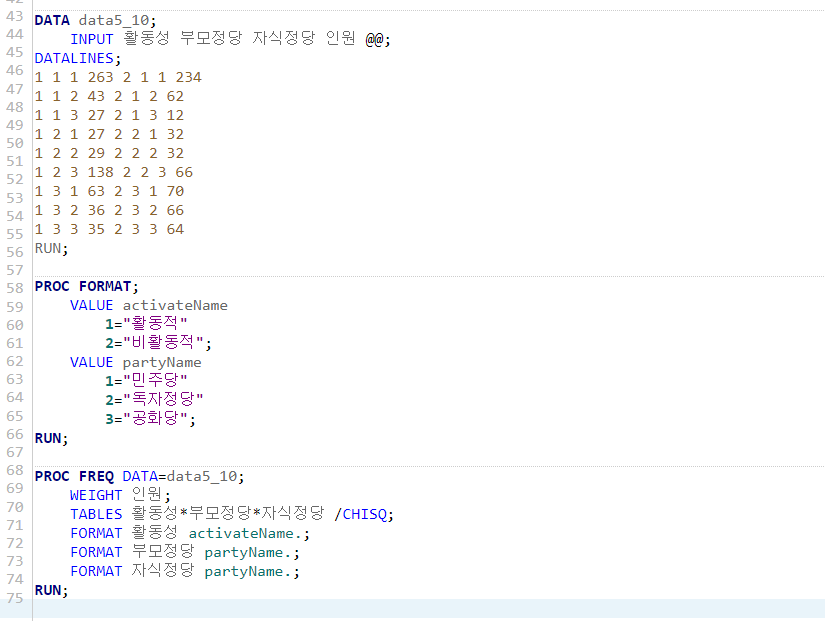


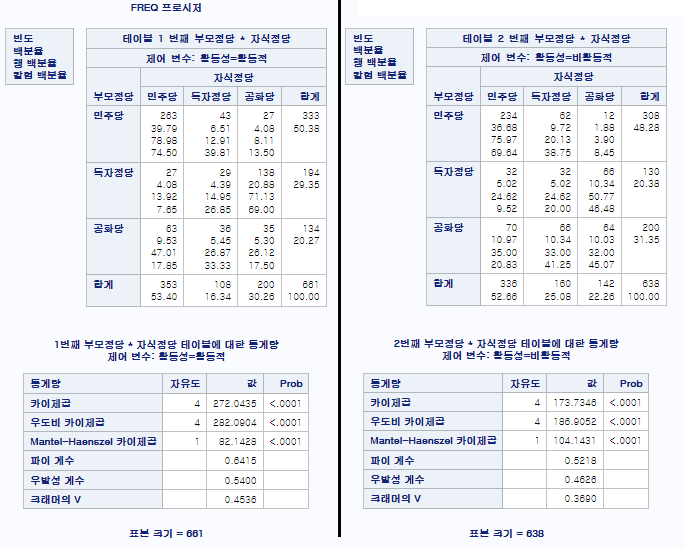


상자그림으로 비교해보면 해외 주가에 결측치를 쉽게 확인할 수 있고 결측치를 이외의 해외와 국내의 주가 분포는 크게 차이가 나지 않아 보인다.

**(5장)10.** 부모와 자식 간 정치 성향의 관련성에 관한 미국 연구자료이다. 부모의 정치활동 수준이 활동적인 경우와 비활동적인 경우 각각에 대해서 분할표를 만든 다음 카이제곱 검정을 시행하라. 어떤 결론을 내릴 수 있는가? 민주당은 진보적, 공화당은 보수적이다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 부모의 정치활동 | 부모의 정당 | 자식의 정당 | | |
| 민주당 | 독자정당 | 공화당 |
| 활동적 | 민주당  독자정당  공화당 | 263  27  63 | 43  29  36 | 27  138  35 |
| 비활동적 | 민주당  독자정당  공화당 | 234  32  70 | 62  32  66 | 12  66  64 |



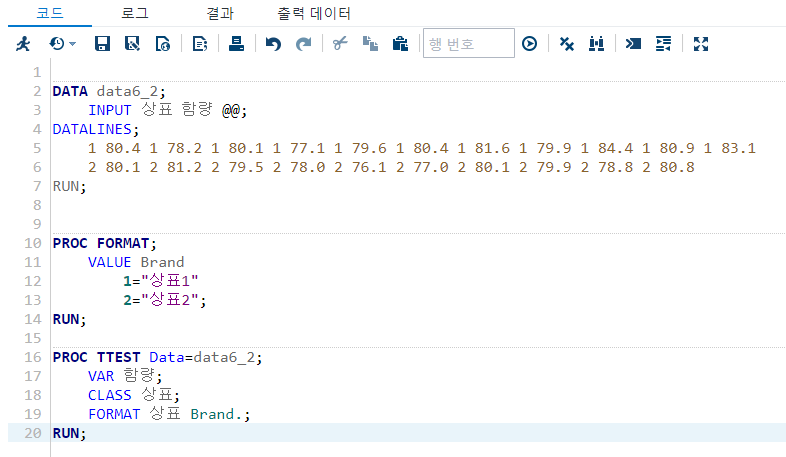


활동적인 부모와 비활동적인 부모 모두에서 부모정당과 자식정당 변수와의 관계는 카이제곱( 통계량의 자유도가 4인 에 대한 *P­*값은 0.0001이므로 유의수준 0.05보다 작기 때문에 귀무가설을 기각하고 대립가설을 채택하여 부모정당과 자식정당의 두 변수는 연관이 있다고 할 수 있다.

2) 교재 6장 연습문제(p.235) 3번, 5번

**(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 |  |





분산에 대한 검정

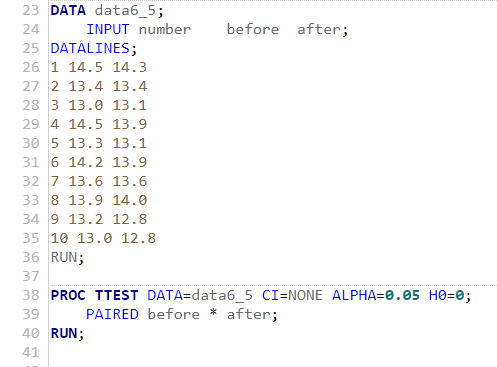
,

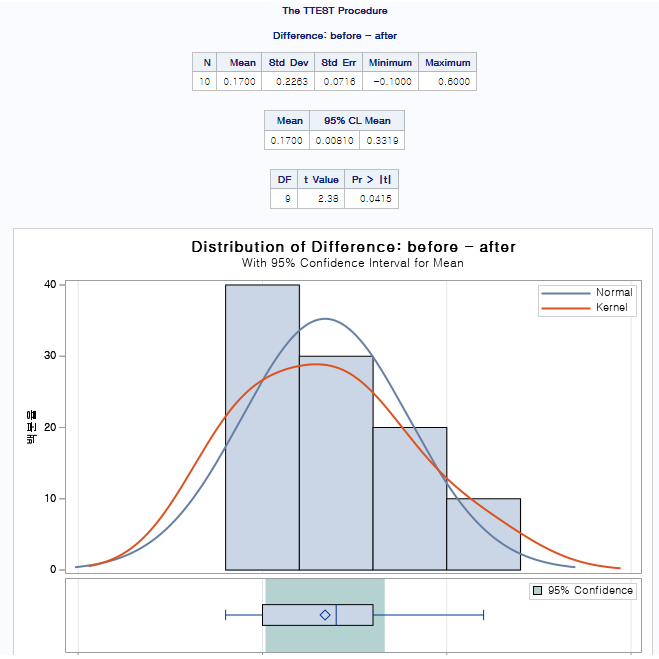
F통계량()에 대한 이 보다 값이 크므로 두 집단의 분산이 동일하다는 가정을 기각할 수 없다는 것을 알 수 있다. 따라서 분산이 동일하다는 가설을 받아들이고 ‘등분산을 가정함’의 결과를 이용한다. 만약 분산이 다를 경우 COCHRAN 통계량을 구해서 검정을 실시해야 한다.

등분산을 가정했을 때, 자유도 19인 검정 통계량 t값은 1.676이고, 이에 대한 유의 확률 으로 유의수준 0.05보다 크므로 이 두 원료의 주성분 A의 함량은 같다는 귀무가설을 기각하지 못한다.

**(6장)5.** 고등학교 육상선수에게 체중감량을 시키면 달리기에 어떤 변화가 일어나는지 조사하기 위하여 어떤 고등학교 육상선수 10명에게 감량훈련(2~4kg 정도)을 시킨 후 100m 달리기를 실시하여 다음의 데이터를 얻었다. 감량전과 후에 달리기 속도에 차이가 있는지 대응 • 비교하라.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 선수 번호 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 감량 전 | 14.5 | 13.4 | 13.0 | 14.5 | 13.3 | 14.2 | 13.6 | 13.9 | 13.2 | 13.0 |
| 감량 후 | 14.3 | 13.4 | 13.1 | 13.9 | 13.1 | 13.9 | 13.6 | 14.0 | 12.8 | 12.8 |





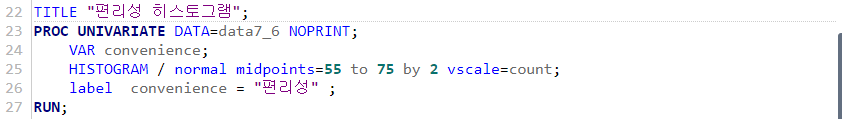
자유도는 9, t값은 2.38이고 P값은 0.0415로 감량전과 후의 속도는 차이가 있다고 볼 수 있다.

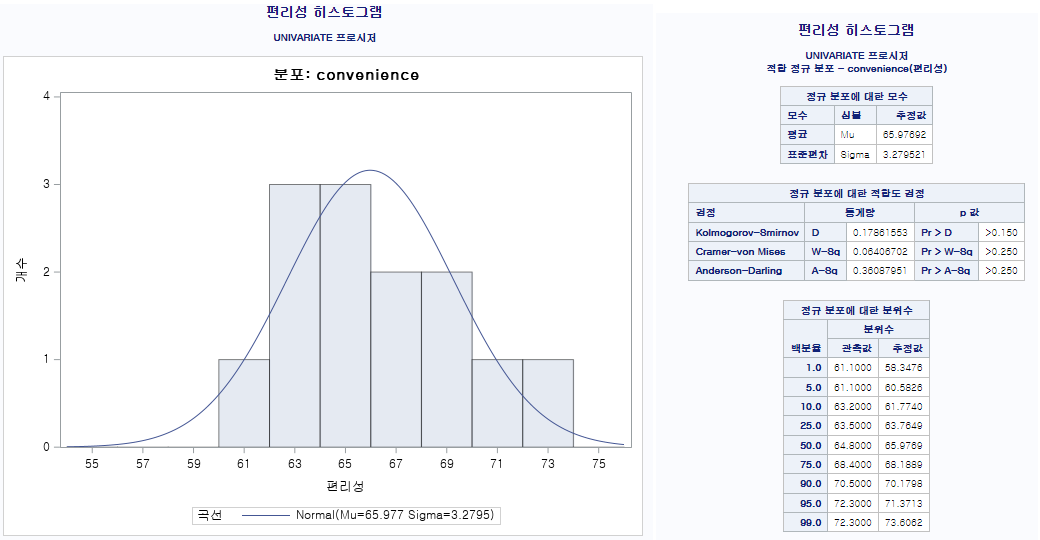
3) 교재 7장 연습문제(p.277) 6번

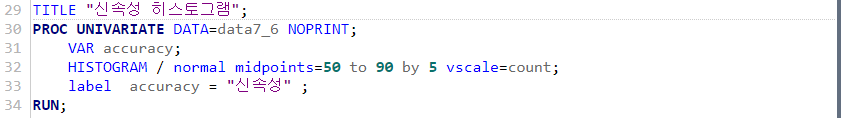
**(7장)6.** 다음은 13개의 시중은행에 대한 편리성, 신속성, 친절, 능률, 쾌적, 자동화 등의 점수를 나타내고 있는 자료이다.

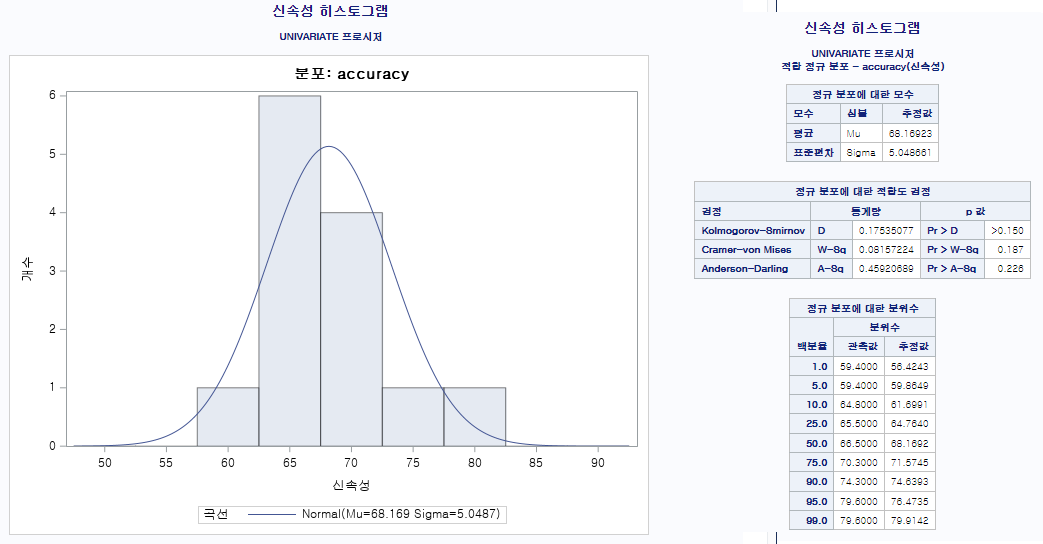
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | convenience | accuracy | kindness | efficiency | pleasant | automatic |
| Kookmin  Enterpr  Boram  Commerce  Seoul  Shinhan  City  Exchange  First  Chohung  Hana  Hanil  House | 70.5  64.8  67.1  61.1  63.4  72.3  64.2  68.4  66.1  63.5  69  63.2  64.1 | 59.4  70.3  79.6  65  66.5  69.1  72  67.5  66.5  65.7  74.3  65.5  64.8 | 63.7  68.6  78.5  65.6  67.9  74.2  71.4  67.3  67.3  64.3  80.5  68.3  67.8 | 54.3  55.2  62.4  54.4  65  60  56.9  51.3  50.7  53.9  63.6  49.8  59.7 | 66.9  68  79.8  64.5  59.7  70.1  72.8  71.3  63.4  61.7  75.7  64.6  65.7 | 62.6  64.1  62.4  63.9  62  68.2  57.8  65.8  63.3  62.7  55.9  59.1  61.8 |

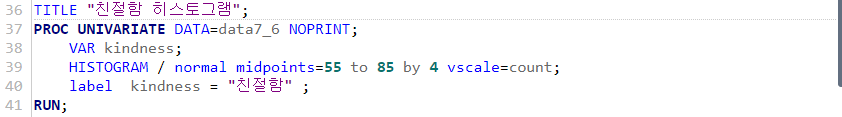
1. 각 변수의 히스토그램을 그리고 설명하라.

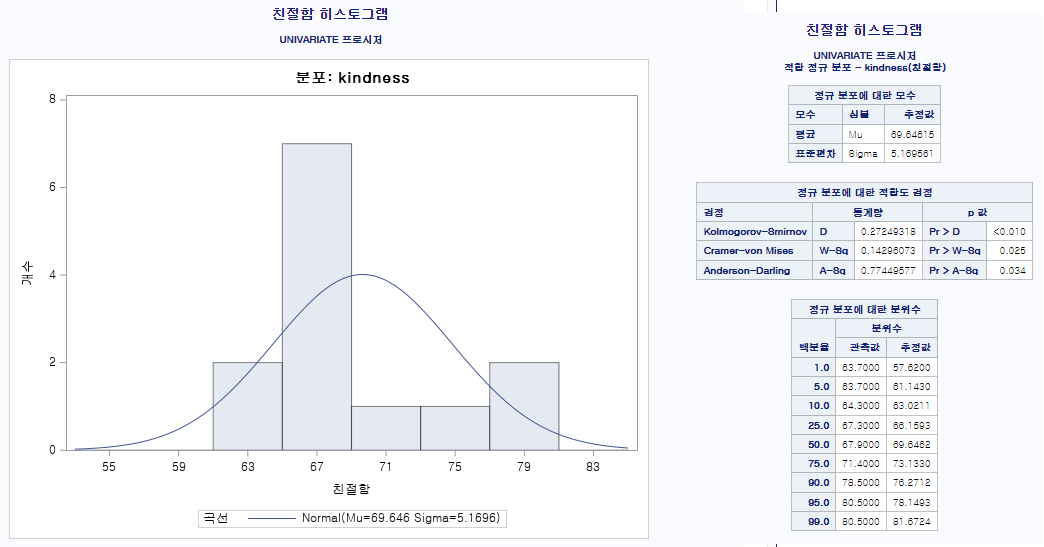


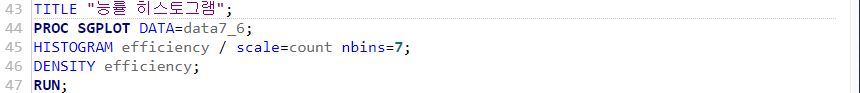


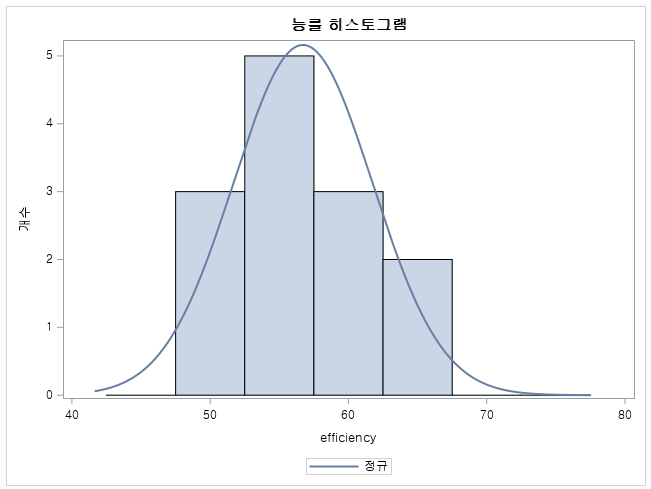




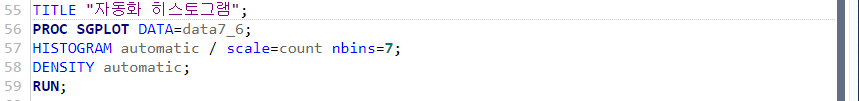


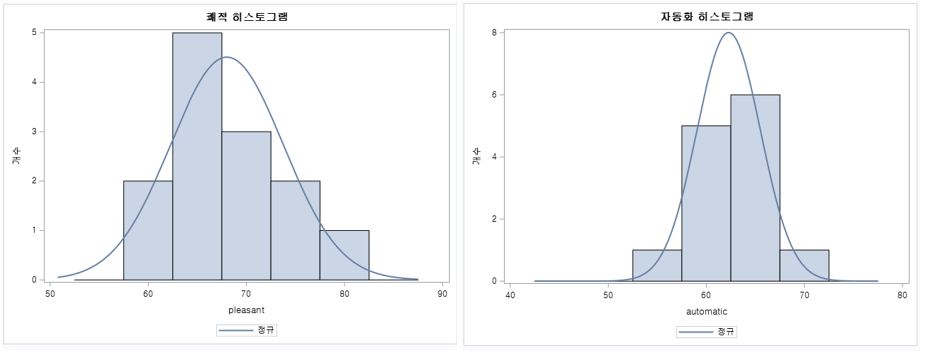






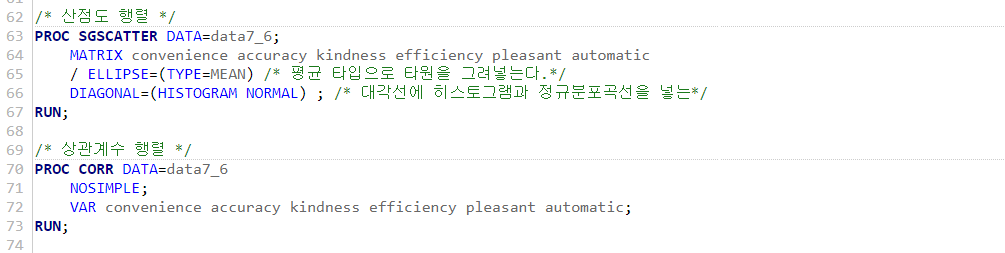


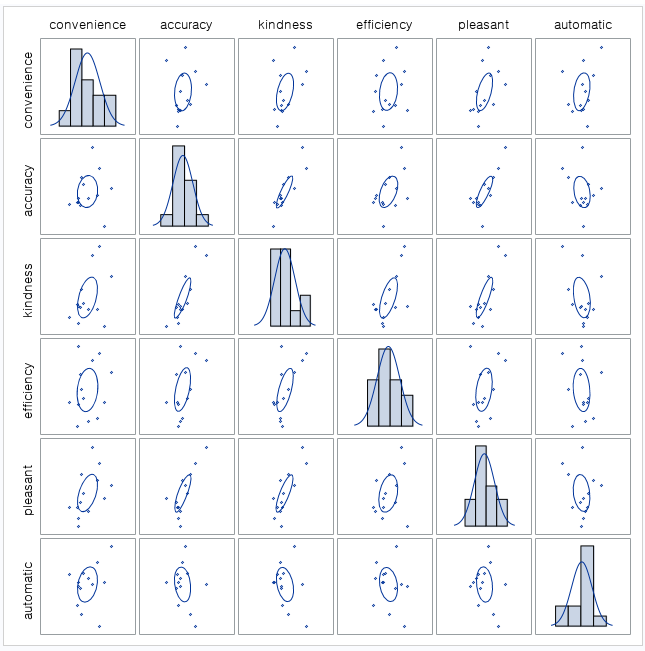




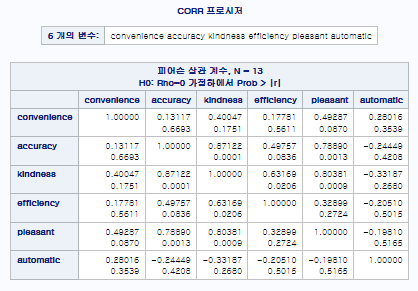
친절 히스트그램을 제외한 편리성, 신속성, 능률, 쾌적, 자동화 등의 전수의 분포는 정규분포에 근접하지만 친절 변수에서는 많은 은행들이 평균보다 낮은 점수를 받고 있다.

1. 산점도 행렬 및 상관계수 행렬을 구하고, 변수들의 관계를 설명하라.





타원이 많이 찌그러진 그림이 있는 변수들의 관계가 선형관계를 쉽게 알 수 있다. 위의 행렬에서 보면 convenience와 automatic은 다른 변수와는 선형관계가 뚜렷이 보이지 않는다. Accuracy\*kindness, accuracy\*pleasant, kindness\*pleasant 변수끼리 선형관계가 뚜렷해 보이고 convenience\* pleasant, kindness\*efficient 변수끼리 선형관계가 약하게 있는 것으로 볼 수 있다.



Accuracy\*kindness, accuracy\*pleasant, kindness\*pleasant의 *P*­값은 0.001 이하로 나타나고 있으며, convenience\* pleasant *P*­값은 0.0206, kindness\*efficient *P*­값은 0.0836으로 나타나고 있다. Accuracy\*kindness, accuracy\*pleasant, kindness\*pleasant, convenience\* pleasant 변수끼리 상관관계가 있다고 할 수 있다.

2. (25점) 다음을 SPSS를 이용하여 작성하시오.

1) 교재 7장 연습문제(p.277) 3번

**(7장)3.** 다음은 어떤 공정에서 생산되는 기계부품의 길이(mm)를 측정한 값이다. 줄기-잎 그림, 히스토그램, 상자그림을 그리고 설명하라

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 16.22  15.93  16.08  16.12  16.00  16.07  15.94  16.26 | 15.75  15.96  16.19  15.70  16.19  16.11  15.74  15.96 | 15.71  15.96  15.78  15.92  15.93  15.75  16.05  15.94 | 15.80  16.12  16.32  15.89  15.91  16.04  16.21  16.08 | 16.25  15.89  15.83  15.95  15.94  16.07  16.02  15.75 | 16.10  15.95  15.96  16.07  16.02  15.75  15.88  15.96 | 16.15  16.01  15.86  16.10  15.80  15.41  15.94  16.17 |

2) 교재 8장 연습문제(p.305) 1번, 8번

**(8장)1.** 다음은 동물 25마리로부터 얻은 두 변수에 관한 자료이다. 산점도를 그리고 상관분석과 회귀분석을 하라.

X: arterial calcium deposition(mg/100g)

Y: serum cholesterol(mg/200mL)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 번호 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| X  Y | 59  298 | 24  239 | 36  264 | 57  299 | 63  284 | 32  239 | 40  242 | 24  213 | 24  206 | 59  299 | 36  260 | 42  277 | 55  285 |
| 번호 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |  |
| X  Y | 45  272 | 28  231 | 35  265 | 47  256 | 51  271 | 29  233 | 33  248 | 51  266 | 49  234 | 44  265 | 43  251 | 32  245 |  |

**(8장)8.** x, y변수 간에 다음의 데이터를 얻었다. 두 변수 간의 산점도를 그리고, 적합한 회귀모형을 추정하라.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Y | 20 | 14 | 11 | 10 | 9 | 8 | 8 | 7 | 7 | 6 | 6 | 6 |

2) 교재 9장 연습문제(p.339) 1번, 6번

**(9장)1.** 유아들을 대상으로 세 가지 읽는 방법을 비교 • 실험하여 다음과 같은 독서평가 점수자료를 얻었다. 세 방법에 통계적으로 유의한 차이가 있는지 검정하라.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 독서방법 | 점수 | | | | |
| B |  | 6 | 9 | 2 | 16 |
| D | 7 | 7 | 12 | 10 | 16 |
| S | 11 | 7 | 4 | 7 | 7 |

**(9장)6.** 어떤 기계의 소음을 작게 하려고 모터(motor)의 베어링 부분에 대하여 조립 후의 볼 베어링의 유격(play)을 3수준, 조립 후의 진동을 3수준으로 바꾸어 3회 반복하여 3x3x3=27(회)의 실험을 랜덤하게 행하고, 소음계로 소음을 측정한 결과가 아래와 같다.

[실험조건]

|  |  |  |  |
| --- | --- | --- | --- |
| 볼 베어링의 유격 | | 조립후의 진동: | |
|  |  | |  | |  |
|  | 78  76  75 | | 74  76  75 | | 75  76  77 |
|  | 79  80  80 | | 78  79  80 | | 78  77  79 |
|  | 81  83  83 | | 79  77  75 | | 79  80  81 |

<참고자료>

통계패키지 김성수,성내경,이영섭 공저

<https://data-flair.training/blogs/sas-correlation-analysis/>

https://documentation.sas.com/doc/en/vdmmlcdc/1.0/grstatproc/p0lfzklhx36ylln1t9sssgzuf64m.htm