**엑셀데이터분석 실습과제물 (2023학년도)**

**(줌강의: 3/27일)**

♣ 과제물은 정해진 기한까지 온라인제출하시기 바랍니다.

질문이 있으면 게시판을 글을 올리기 바랍니다.

1. 첨부된 강수량 자료는 A 지역과 B 지역의 1990년부터 2022년까지의 연강수량 자료이다. 이 자료를 엑셀과 KESS로 분석하여 다음 물음에 답하시오.

(1) 두 지역의 연도별 강수량 자료에 대해 꺾은선형 차트를 이용하여 전체적인 경향을 설명하시오.

차트이(가) 표시된 사진

자동 생성된 설명

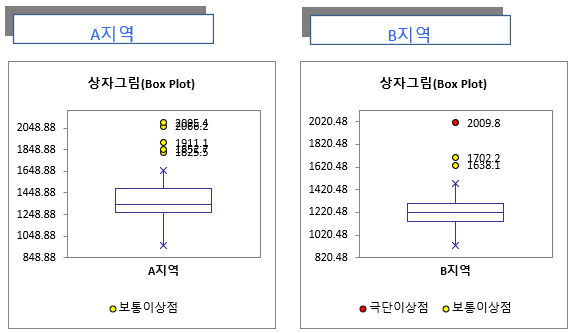
A지역과 B지역 모두 연간강수량의 표준편차가 2007년 이후부터는 점차 줄어들고 있고 A지역과 B지역의 연간 강수량의 차이도 2018년 이후부터는 거의 없음을 알 수 있다.

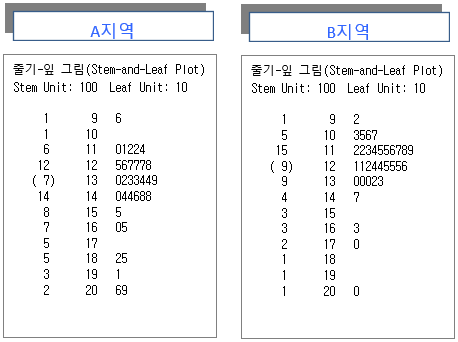
(2) 각 지역의 강수량에 대한 기술통계량을 구해 두 지역의 연강수량을 비교하시오.

A지역이 B지역보다 평균적으로 더 많은 연강수량을 기록하고 있다. A지역의 연강수량의 표준편차도 더 크지만 B지역보다는 연강수량은 더 크다.

|  |  |  |  |
| --- | --- | --- | --- |
| A지역 |  | B지역 |  |
|  |  |  |  |
| 평균 | 1420.969 | 평균 | 1251.176 |
| 표준 오차 | 47.8119 | 표준 오차 | 35.85728 |
| 중앙값 | 1342.7 | 중앙값 | 1218.83 |
| 최빈값 | #N/A | 최빈값 | #N/A |
| 표준 편차 | 274.6584 | 표준 편차 | 205.9844 |
| 분산 | 75437.25 | 분산 | 42429.58 |
| 첨도 | 0.611789 | 첨도 | 5.400731 |
| 왜도 | 0.977249 | 왜도 | 1.994002 |
| 범위 | 1133.2 | 범위 | 1081.2 |
| 최소값 | 962.2 | 최소값 | 928.6 |
| 최대값 | 2095.4 | 최대값 | 2009.8 |
| 합 | 46891.97 | 합 | 41288.82 |
| 관측수 | 33 | 관측수 | 33 |

(3) 각 지역의 연 강수량에 대한 줄기-잎 그림과 상자그림을 그려서 비교하시오.





(4) 두 지역의 연간 강수량에 대한 분석할 때 어느 지역의 강수량이 많다고 할 수 있는가?

2018년 이전에는 A 지역의 연간 강수량이 평균적으로 더 많았지만 2018년 이후에는 거의 비슷한 연간 강수량을 보이고 있다.

2. 다음 물음에 답하시오.

(1) 자유투 성공률이 90%인 어느 농구선수가 자유투를 시도하려고 한다. 20번의 자유투에서 성공한 횟수를 확률변수 로 정의할 때 확률변수 가 따르는 분포는 무엇인가? 18번 이상 성공하게 될 확률과 15번 이하 성공하게 될 확률은 각각 얼마인가?

자유투 성공과 실패로 나눠지기 때문에 이항분포이고 성공확률을 0.9, 시행 횟수는 20번 이다. X ~ B(20, 0.9)를 따르는 분포이다.

|  |  |  |  |
| --- | --- | --- | --- |
| 18개이상일 확률 | P(X>=18)=1-P(X<=17) | 0.676926805 | (=1-BINOM.DIST(17,20, 0.9, TRUE)) |
| 15개이하일 확률 | P(X<=15) | 0.043174495 | (=BINOM.DIST(15,20,0.9, TRUE)) |

(2) 10문항의 4지선다형 문제가 있다. 어느 학생이 10문항 전체에 대해서 임의로 답안을 적을 때 정답수를 확률변수 X라고 하자. 정답 수가 2개 이하일 확률과 정답 수가 6개 이상일 확률은 각각 얼마인가?

4지선다형 문제이고 답을 임으로 선택했기 때문에 확률은 1/4=0.25 이다. 따라서 확률분포 X~B(10,0.25)를 따른다.

|  |  |  |  |
| --- | --- | --- | --- |
| 2개이하일 확률 | P(X<=2) | 0.525592804 | (=BINOM.DIST(2,10,0.25, TRUE)) |
| 6개이상일 확률 | P(X>=6)=1-P(X<=5) | 0.019727707 | (=1-BINOM.DIST(5,10, 0.25, TRUE)) |

3. 어느 은행의 콜센터로 한 시간에 평균 3.5통의 상담전화가 걸려온다고 한다. 한 시간 동안에 걸려오는 상담전화의 수는 포아송 분포를 따른다고 가정할 수 있다. 다음 물음에 답하시오.

(1) 한 시간에 6통 이상의 상담전화가 걸려올 확률은?

(2) 한 시간에 2통 이하의 상담전화가 걸려올 확률은?

(3) 한 시간에 3통 이상 6통 이하의 상담전화가 걸려올 확률은?

**※ 문제 4와 5는 엑셀데이터 멀티미디어 7강 내용을 참고하여 해결해 보기 바랍니다. 문제 4의 (2), (3)은 NORM.INV 함수를 이용하는 문제입니다.**

4. 어느 회사의 전체 직원을 대상으로 하는 직무관련 시험에서 시험점수는 평균 80, 표준편차 8인 정규분포를 따른다는 것을 알았다. 다음 물음에 답하시오.

(1) 시험점수가 65점 이하이면 재교육을 받도록 하려고 한다. 재교육을 받는 직원의 비율은 얼마인가?

(2) 만약 전체 직원의 90%만 합격을 시키고 나머지 직원은 재교육을 받도록 하려고 한다면 패스 점수는 얼마로 해야 하는가?

(3) 상위 5%의 직원에게는 포상금을 주려고 한다. 포상금을 받기 위해서는 최소한 몇 점을 얻어야 하는가?

5. 어떤 사람이 자기 집에서 직장까지 차를 몰고 가는데 걸리는 시간이 평균 30분, 표준편차 5분인 정규분포를 따른다고 한다. 다음 물음에 답하시오.

(1) 이 사람이 집에서 아침 8시 20분에 출발하였을 때 아침 9시까지 직장에 도착할 수 있을 확률은?

(2) 이 사람이 아침 9시까지 직장에 도착할 확률이 0.99 이상 되게 하려면 늦어도 몇 시에 집에서 출발해야 하는가?

|  |
| --- |
| <참고> 집에서 직장까지 차를 몰고 가는데 걸리는 시간을 확률변수 X로 정의하였음. 주어진 문제를 확률변수 X를 이용해서 나타낼 수 있으면 문제풀이에 도움이 될 것임. |