19101185 강동현 데이터 분석 Assignment 03

1. 제품에 대한 주별 요구량은 평균이 1000 , 표준 편차가 200인 정규분포를 따른다. 현재 재고가 2200이 있고 , 앞으로 2주간 배송이 없다고 가정하자. 또한 주별 요구량은 독립적이다. 이때 물음에 답하는 문제이다.

(a). 2주를 각각 한주씩 을 따로 보아 한주에 1100 미만이고 두주 모두 조건을 만족할 확률을 구하는 문제이다. 일단 한주에 대해서 1100미만일 확률을 구하면 ( 1100 – 1000 ) / 200 으로 문제에서 주어진 값을 활용하면 0.6915 임을 알 수있다. 두주가 독립이므로 두주 모두 1100 이하일 확률을 0.6915 \* 0.6915 이다. 따라서 답은 0.47817225 이다.

(b). 우리는 2주의 요구량에 대하여 생각해야 한다. 따라서 정규분포를 따르고 두 주의 요구량이 독립적이기 떄문에 두주간의 요구량의 평균은 2000 , 분산은 80000이다. 따라서 표준편차는 200루트2 이다. 이후 1100 미만일 확률을 구하게 되면 ( 2200 – 2000 ) / 200루트2 를 통해 구할 수 있고 , 문제에 주어진 값을 활용하면 구하고자 하는 확률이 0.2399 임을 알 수 있다.

2. 경제 시험에서 표본 평균과 표본 표준 편차는 60과 20 이다. 통계 시험에서는 55와 10 이다. 경제 시험에서 70점을 맞고 통계 시험에서 62점을 받았다. 두 히스토그램이 정규 히스토그램에 근사한다고 가정하고 물음에 답하는 문제이다.

(a). 어떤 시험에서 백분위 점수가 높았는지를 묻는 문제이다. 경제 시험의 경우 ( 70 – 60 ) / 20 = 1/2 이고 통계 시험의 경우는 62 – 55 / 10 = 7/10 이다. 두 시험 모두 정규 히스토그램에 근사하므로 백분위 점수가 더 높은 과목은 통계 라고 볼 수 있다.

(b). 경제 시험에서의 나보다 점수가 낮은 학생들의 백분율을 근사적으로 구하는 문제이다. 위에서 구했듯이 경제 시험의 경우 점수가 낮은 학생들의 백분율은 P(Z<1/2)를 통해서 표현이 가능하다. 따라서 답은 0.6915 이고 대략 69.15% 이다.

(c). 통계 시험에서의 나보다 점수가 낮은 학생들의 백분율을 근사적으로 구하는 문제이다. 위의 (a) 문제에서 구했듯이 통계 시험의 경우 점수가 낮은 학생들의 백분율을 P(Z<0.7)을 통해서 표현이 가능하다. 따라서 답은 0.7580 이고 대략 75.8% 이다.

3. central processing unit 이 작업을 처리하는데 걸리는 시간은 평균 20초 이고 표준편차는 3초로 정규분포를 따른다. 15개의 샘플이 관찰 된다고 할 때 표본 분산이 12를 초과할 확률이 얼마인지에 대하여 묻는 문제이다. 따라서 분산에 관한 식을 활용할 수 있다. 우리는 전체의 평균과 표준편차를 알고 있고 n = 15이기 때문에 이와 문제에서 주어진 점을 활용해 chi-square를 통해 원하는 값을 구할 수 있다. 따라서 (n-1) S^2 / 시그마^2 ~ X(2 , n-1 ) 를 활용 할 수 있다. 따라서 답은 0.1781 이다.

4. 전자저울은 평균이 0이고 표준편차가 0.1mg 인 정규분포를 가진다. 이때 동일한 물체의 연속적인 무게는 3.142, 3.163, 3.155, 3.150, 3.141 라고 하자. 이때 물음에 답하는 문제이다.

(a). 실제 관측값에 대하여 interval estimate를 구하는 문제이다. 표본의 개수는 5개 이고 표준편차의 경우는 0.1 이다. 문제에서 주어진 값을 활용하면 95퍼센트의 경우 표준 정규분포의 경우 평균에서 1.96만큼 떨어진 부분에서 나뉘어진다는 것을 알 수 있다. 따라서 이 경우 샘플 평균 즉 3.1502를 표준 정규분포화 했을때 , 1.96만큼 떨어진 부분을 찾는 것이다. 따라서 3.1502 + 1.96 \* 0.1 / 루트5 가 upper 바운드가 된다. 계산값이 약 3.1502 + 혹은 - 0.0876538647179917 이므로 정리하자면 약 3.06255 에서 약 3.23785 사이의 interval estimate를 구할 수 있다.

(b). (a)번과 같은 방식으로 풀면 된다. 단 99퍼센트의 경우를 묻는 문제 이므로 2.58 \* 0.1 / 루트5 를 표본평균에 더하고 빼면 된다. 따라서 계산값은 약 3.1502 + 혹은 - 0.1153 이고 정리하면 약 3.0349 에서 약 3.2655 사이의 interval estimate를 구할 수 있다.

5. 평균 가정 수도 사용량이 하루에 평균 350갤런이라고 하자. 무작위의 20개 주택중에서 물 사용량 표본을 다음과 같이 추출 하였다. 이 자료가 관계자가 말한 평균에 대한 의견과 모순이 되는지 확인하는 문제이다. 이 경우 H0를 평균이 350이라는 것으로 두고 H1을 평균이 350이 아니라는 것으로 두고 문제를 보면 된다. 표본 평균은 353.8 이다. 따라서 TS는 루트20 ( 353.8 – 350 ) / 시그마 이다. 이때 시그마는 루트453.46 으로 약 21.29 이다. 따라서 루트20 \* 3.8 / 21.29로 약 0.7982 이다. 따라서 1.730 보다 안쪽이므로 모순되지 않았음을 알 수 있다.