제 3 장 기술통계분석

# 도수분포표(frequency distribution table)

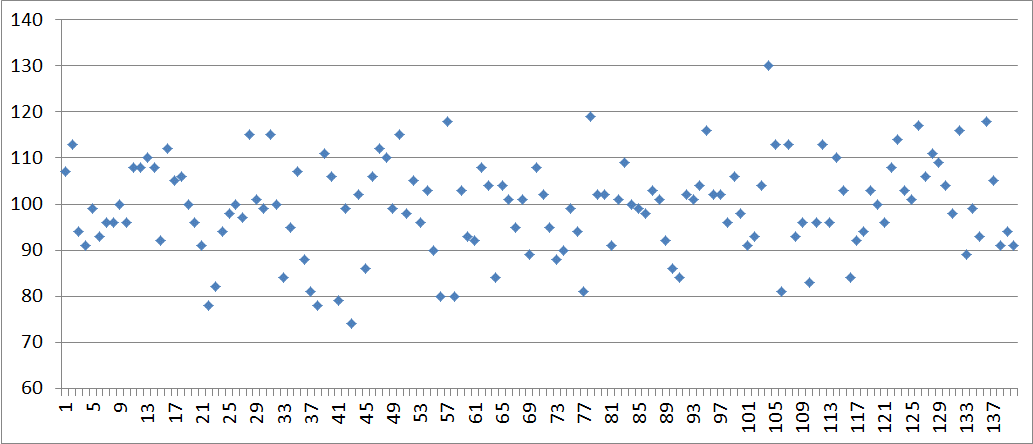
수집된 자료를 적절한 계급(class)으로 분류하여 정리한 표

계급(class)와 빈도수/도수(frequency)로 구성된다.

|  |  |
| --- | --- |
| 연령 | 인원(천명) |
| 10세 미만  10세 이상 ~ 20세 미만  20세 이상 ~ 30세 미만  ……………………………..  50세 이상 ~ 60세 미만  60세 이상 | 25  40  108  ………….……  46  32 |
| 합계 | 457 |

Example:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Data | No | Data | No | Data | No | Data | No | Data | No | Data | No | Data |
| 1 | 107 | 21 | 91 | 41 | 79 | 61 | 92 | 81 | 91 | 101 | 91 | 121 | 96 |
| 2 | 113 | 22 | 78 | 42 | 99 | 62 | 108 | 82 | 101 | 102 | 93 | 122 | 108 |
| 3 | 94 | 23 | 82 | 43 | 74 | 63 | 104 | 83 | 109 | 103 | 104 | 123 | 114 |
| 4 | 91 | 24 | 94 | 44 | 102 | 64 | 84 | 84 | 100 | 104 | 130 | 124 | 103 |
| 5 | 99 | 25 | 98 | 45 | 86 | 65 | 104 | 85 | 99 | 105 | 113 | 125 | 101 |
| 6 | 93 | 26 | 100 | 46 | 106 | 66 | 101 | 86 | 98 | 106 | 81 | 126 | 117 |
| 7 | 96 | 27 | 97 | 47 | 112 | 67 | 95 | 87 | 103 | 107 | 113 | 127 | 106 |
| 8 | 96 | 28 | 115 | 48 | 110 | 68 | 101 | 88 | 101 | 108 | 93 | 128 | 111 |
| 9 | 100 | 29 | 101 | 49 | 99 | 69 | 89 | 89 | 92 | 109 | 96 | 129 | 109 |
| 10 | 96 | 30 | 99 | 50 | 115 | 70 | 108 | 90 | 86 | 110 | 83 | 130 | 104 |
| 11 | 108 | 31 | 115 | 51 | 98 | 71 | 102 | 91 | 84 | 111 | 96 | 131 | 98 |
| 12 | 108 | 32 | 100 | 52 | 105 | 72 | 95 | 92 | 102 | 112 | 113 | 132 | 116 |
| 13 | 110 | 33 | 84 | 53 | 96 | 73 | 88 | 93 | 101 | 113 | 96 | 133 | 89 |
| 14 | 108 | 34 | 95 | 54 | 103 | 74 | 90 | 94 | 104 | 114 | 110 | 134 | 99 |
| 15 | 92 | 35 | 107 | 55 | 90 | 75 | 99 | 95 | 116 | 115 | 103 | 135 | 93 |
| 16 | 112 | 36 | 88 | 56 | 80 | 76 | 94 | 96 | 102 | 116 | 84 | 136 | 118 |
| 17 | 105 | 37 | 81 | 57 | 118 | 77 | 81 | 97 | 102 | 117 | 92 | 137 | 105 |
| 18 | 106 | 38 | 78 | 58 | 80 | 78 | 119 | 98 | 96 | 118 | 94 | 138 | 91 |
| 19 | 100 | 39 | 111 | 59 | 103 | 79 | 102 | 99 | 106 | 119 | 103 | 139 | 94 |
| 20 | 96 | 40 | 106 | 60 | 93 | 80 | 102 | 100 | 98 | 120 | 100 | 140 | 91 |



|  |  |
| --- | --- |
| 계급 | 빈도 |
| ~75 | 1 |
| 76~80 | 5 |
| 81~85 | 9 |
| 86~90 | 8 |
| 91~95 | 23 |
| 96~100 | 29 |
| 101~105 | 28 |
| 106~110 | 18 |
| 111~115 | 12 |
| 116~120 | 6 |
| 121~125 | 0 |
| 126~130 | 1 |

## 질적 자료

상대도수(relative frequency)

|  |
| --- |
| 상대도수 = ,  여기서, = 총 관찰개수, = 번째 계급의 빈도수 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 학점 | 빈도수 | 상대도수 | 백분율 | EMB0000283465d8EMB0000283465d9 |
| A | 6 | 0.12 | 12% |
| B | 22 | 0.44 | 44% |
| C | 15 | 0.30 | 30% |
| D | 5 | 0.10 | 10% |
| F | 2 | 0.04 | 4% |
| 합계 | 50 | 1.00 | 100% |

|  |  |  |  |
| --- | --- | --- | --- |
| 학점 | 빈도수 | 상대도수 | 백분율 |
| A | = 6 | = 0.12 | 12% |
| B | = 22 | = 0.44 | 44% |
| C | = 15 | 0.30 | 30% |
| D | = 5 | 0.10 | 10% |
| F | = 2 | 0.04 | 4% |
| 합계 | = 50 | 1.00 | 100% |

첫번째 계급은?

세번째 계급의 빈도를 기호로 표현하시오.

= ?

다섯 번째 계급의 상대도수 = ?

## 양적 자료

계급구간의 계산

|  |
| --- |
| 계급구간 = |

## 누적도수분포(cumulative frequency distribution)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 체중 | 빈도수 | 누적  도수 | 누적  상대도수 | EMB0000283465daEMB0000283465db |
| 45 미만 | 3 | 3 | 5% |
| 45 이상~ 50 미만 | 8 | 11 | 18% |
| 50 이상~ 55 미만 | 12 | 23 | 38% |
| 55 이상~ 60 미만 | 18 | 41 | 68% |
| 60 이상~ 65 미만 | 10 | 51 | 85% |
| 65 이상~ 70 미만 | 6 | 57 | 95% |
| 70 이상~ 75 미만 | 3 | 60 | 100% |
| 합계 | 60 |  |  |

도수분포그래프에서 수평축과 수직축의 의미는?

누적도수분포그래프에서 수평축과 수직축의 의미는?

수강생 10명인 어떤 학급의 성적은 다음과 같다.   
{100, 48, 57, 95, 75, 85, 81, 88, 71, 79}

이 학급에 관련된 도수분포표를 완성하시오. (빈도는 정수로, 나머지는 소수점 1까지 표기한다.) 여기서 계급은 0~59, 60~69, 70~79, 80~89, 90~100으로 한다.

도수분포 그래프, 상대빈도 그래프, 누적상대빈도 그래프를 완성하시오.

도수분포 그래프 높이의 최소값(빈도의 최소값)은 \_\_\_\_이다.

이 학급의 경우, 도수분포 그래프 높이의 합은 \_\_\_\_\_이다.

일반적으로 상대빈도 그래프 높이의 최소값은 \_\_\_이다.

일반적으로 상대빈도 그래프 높이의 최대값은 \_\_\_이다.

상대빈도 그래프 높이의 합은 (1.00)이다. Ans. True, False

누적상대빈도 그래프의 최우측 그래프 높이는 항상 1이다. Ans. True, False

누적상대빈도 그래프 높이는 우측으로 갈수록 ① 감소한다. ② 감소하지 않는다. ③ 증가하지 않는다. ④ 증가한다.

수강생이 10명인 어떤 과목의 학점이다. 10명 중 7명의 학점은 C, D, F로 확인되었으나, 나머지 3명의 학점은 A 또는 B로만 확인되었을 뿐 A학점 몇 명, B학점 몇 명으로까지는 확인되지 않았다. A학점 받는 학생들의 수는 최소 몇 명에서 최대 몇 명인가?

어떤 도수분포표의 부분 정보이다. 계급 10~19의 상대빈도값은?

|  |  |  |
| --- | --- | --- |
| 계급 | 상대빈도 | 누적상대빈도 |
| 0~9 | 0.150 | 0.150 |
| 10~19 |  | 0.200 |
| 20~29 |  |  |

위의 표에서 계급 20~29의 누적상대빈도값의 최소값과 최대값은?

# 분포의 특성

중심위치(central location) ... 자료가 집중되어 있는 위치(평균값)  
산포경향(dispersion) ............ 자료의 흩어진 상태(분산, 표준편차)  
비대칭도(skewness) ............. 중앙대칭에서 왼쪽/오른쪽으로 치우친 정도  
첨도(kurtosis) ....................... 분포의 모양이 뾰족한 정도

## 중심위치(central location)

산술평균(arithmetic mean), 최빈값(mode), 중앙값(median)

### **산술평균(arithmetic mean)**

|  |
| --- |
| 모집단평균 = =  여기서, = 모집단 평균  = 모집단 크기  = 번째 관찰값 |

|  |
| --- |
| 표본평균 = =  여기서, = 표본평균  = 표본 크기 |

#### 에 관한 학습

|  |  |  |
| --- | --- | --- |
|  |  | (1)  (2)  (3)  (4)  (5)  (6) |
| 1 | 10 |
| 2 | 30 |
| 3 | 20 |
| 4 | 40 |
| 5 | 30 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 결과 |
|  |  |  |  |  |  | + + + |
|  | 10 | 10 | 10 |  |  | 10 + 10 + 10 |
|  |  |  |  |  |  | 해당 없음 |
|  |  |  |  |  |  |  |
|  |  | 20 | 20 | 20 |  | 20 + 20 + 20 |
|  |  |  |  |  |  |  |

성적은 로 표기하고, 번째 성적은 로 표기하자. 이렇게 표기하면, 3번째 학생의 성적은 으로, 여덟 번째 학생의 성적은 로 표기된다.

를 덧셈 형식으로 풀어서 표기하시오.

를 덧셈 형식으로 풀어서 표기하시오.

를 덧셈 형식으로 풀어서 표기하시오.

+ + 를 형식으로 표기하시오.

+ + + 를 형식으로 표기하시오.

+ + + 를 형식으로 표기하시오.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 번호 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 성적 | 75 | 70 | 71 | 63 | 90 | 76 | 90 | 65 |

8명이 수강하는 반의 평균을 알고자 한다. 8명의 성적을 조사하니 위의 표와 같았다. (이 경우, 8명이 모집단이다.) 모집단의 평균 (mu라고 읽는다.)의 공식은?

값은?

500명이 수강하는 반의 평균을 알고자 한다. 500명 중 8명을 추출하여 성적을 조사하니 위의 표와 같았다. (이 경우, 8명은 표본이 된다.) 표본의 평균 의 공식은?

=

20개 학급의 성적을 조사하였다. 학급별로 수강인원은 매우 다르지만, 학급 별로 조사한 학생의 수는 20명으로 동일하다. 20명의 성적의 평균을 구했다. 그 성적은 모집단 평균인가 아니면 표본의 평균인가? 당연히 수강생이 20명인 학급이라면 조사한 학생의 성적 평균은 모집단 평균이고, 수강생이 20명을 넘는 학급이라면 조사한 학생의 성적 평균은 표본의 평균이다.  
학급 1과 2에서 표기한 방식으로, 학급 3에서 10까지의 빈 칸을 모두 채우시오.

| 학급 | 수강생의 수 | 조사한 학생의 수 | 조사한 학생의 성적 평균 | 모집단  평균 | 표본 평균 |
| --- | --- | --- | --- | --- | --- |
| 1 | 89 | 20 | 43 | 해당 없음 | = 43 |
| 2 | 20 | 20 | 93 | = 93 | 해당 없음 |
| 3 | 37 | 20 | 32 |  |  |
| 4 | 20 | 20 | 32 |  |  |
| 5 | 74 | 20 | 98 |  |  |
| 6 | 43 | 20 | 87 |  |  |
| 7 | 89 | 20 | 84 |  |  |
| 8 | 20 | 20 | 55 |  |  |
| 9 | 60 | 20 | 64 |  |  |
| 10 | 97 | 20 | 73 |  |  |

#### 가중평균

|  |
| --- |
| 가중평균  = = (since = 1)  여기서, = 번째 관찰치의 가중치 |

에 관한 학습

① 0 1, = 1, 2, ...,

② = 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | 수강생(명) | 평균 | 반 총점 |
| A | 40 | 75 | 3,000 |
| B | 50 | 73 | 3,650 |
| C | 20 | 80 | 1,600 |
| 합계 | 110 |  | 8,250 |

전체 평균 = = = 75

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 반 | 수강생(명) | 가중치(E) | 평균(D) | DE |
| A | 40 | 0.36 (40/110) | 75 | 27.27 |
| B | 50 | 0.45 (50/110) | 73 | 33.18 |
| C | 20 | 0.18 (20/110) | 80 | 14.55 |
| 합계 | 110 | 1.00  (.36+.45+.18) |  | 75.00 (27.27+33.18+14.55) |

(가중 평균 구하기) 아래 표를 완성하시오.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 반 | 수강생(명) | 평균(D) | 가중치(E) | D\*E |
| A | 50 | 80 |  |  |
| B | 30 | 70 |  |  |
| C | 20 | 85 |  |  |
| 합계 | 100 |  |  |  |

(가중평균 구하기) 50명의 평균 학습시간은 6.5시간, 30명의 평균 학습시간은 6.0시간, 10명의 평균 학습시간은 5.5시간이다. 총 90명의 평균 학습시간은?

#### 집단자료 산술평균

|  |
| --- |
| 집단자료 산술평균 = =  여기서, = 번째 계급의 빈도수  = 번째 계급의 중간점  = 총관찰수 (=) |

[주의] 계급구간별 평균을 알지 못해 (계급구간별) 평균을 계급의 중간점으로 추정한 경우이다. 그러므로 위의 공식에 의한 산술평균은 정확한 값이 아니라 추정치다.

#### 산술평균(arithmetic mean), 기하평균(geometric mean), 조화평균(harmonic mean)

- 덧셈으로 계산하는 값에서의 평균

- 곱셈으로 계산하는 값에서의 평균 (예: 성장률)

- 역수의 산술평균의 역수 (예: 속도)

### **최빈값(mode)**

자료의 분포에서 빈도수가 가장 높은 관찰치

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 3 | 2 |
| 4 | 5 |
| 5 | 12 |
| 6 | 8 |
| 7 | 4 |
| 8 | 2 |
| 9 | 1 |

### **중앙값(median)**

개의 값이 있고 모든 값들이 크기 순서대로 놓여 있을 때,

(1) 이 홀수이면 번째 있는 값  
 - 이 7이면, 4번째 크기의 값

(2) 이 짝수이면 번째 값과 번째 값의 평균

- 이 6이면, 3번째 크기의 값과 4번째 크기의 값의 평균

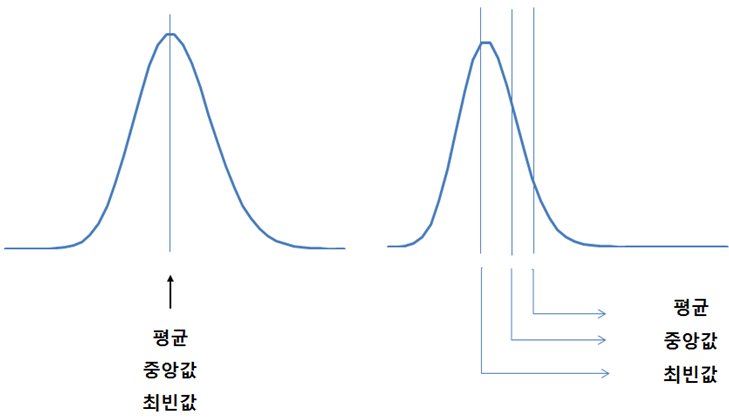
List of countries by median age (CIA World Factbook 2016)

|  |  |  |  |
| --- | --- | --- | --- |
| Indonesia | 29.9 | United Kingdom | 40.5 |
| Vietnam | 30.1 | S. Korea | 41.2 |
| N. Korea | 33.8 | Spain | 42.3 |
| China | 37.1 | Greece | 44.2 |
| United States | 37.9 | German | 46.8 |
| Taiwan | 40.2 | Japan | 46.9 |

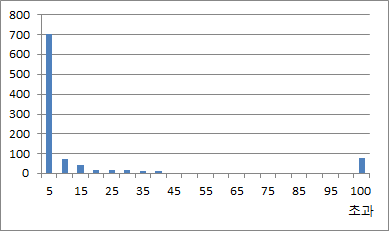
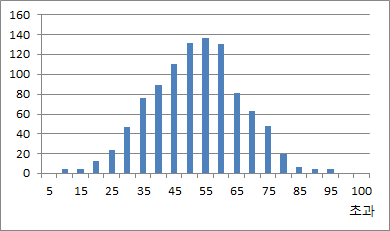
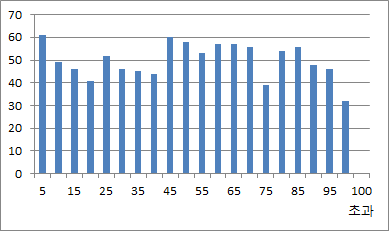
어떤 학급의 학점별 빈도이다. A: 20, B: 30, C: 15, D: 10, F: 5 최빈값과 중앙값은?

{1, 2, 2, 2, 3, 4, 4, 5, 5, 6} 평균, 최빈값, 중앙값을 구하시오.

### **대표값의 비교**



데이터 수가 1,000개인 3개의 자료 비교 I.



평균 50 50 125,131

최빈값 79 46 0

중앙값 50 50 1

최대값 99 98 113,110,156

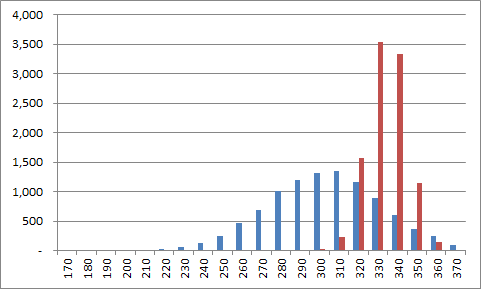
### **백분위수(percentile)**

사분위수(quartiles)

제1사분위수 = 제25백분위수

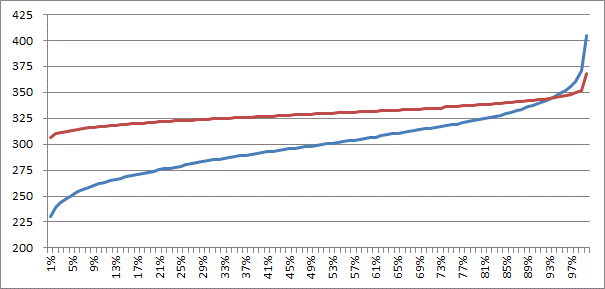
제2사분위수 = 제50백분위수

제3사분위수 = 제75백분위수



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 계열 | 평균 | 최빈값 | 중앙값 | 최대 | 최소 |
| 파란색 | 299 | 295 | 299 | 405 | 181 |
| 붉은색 | 330 | 333 | 330 | 368 | 295 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Blue | Red |  | Blue | Red |  | Blue | Red |  | Blue | Red |
| 1% | 230 | 307 | 15% | 269 | 319 | 65% | 310 | 333 | 90% | 337 | 342 |
| 2% | 239 | 310 | 20% | 274 | 321 | 70% | 315 | 335 | 91% | 339 | 343 |
| 3% | 244 | 311 | 25% | 279 | 323 | 75% | 319 | 336 | 92% | 341 | 343 |
| 4% | 248 | 312 | 30% | 284 | 324 | 80% | 324 | 338 | 93% | 343 | 344 |
| 5% | 251 | 313 | 35% | 288 | 326 | 84% | 328 | 339 | 94% | 346 | 345 |
| 6% | 254 | 314 | 40% | 292 | 327 | 85% | 330 | 340 | 95% | 349 | 346 |
| 7% | 256 | 315 | 45% | 296 | 328 | 86% | 331 | 340 | 97% | 352 | 347 |
| 8% | 258 | 316 | 50% | 299 | 330 | 87% | 333 | 341 | 98% | 356 | 348 |
| 9% | 260 | 316 | 55% | 303 | 331 | 88% | 334 | 341 | 99% | 361 | 350 |
| 10% | 262 | 317 | 60% | 307 | 332 | 89% | 336 | 342 | 100% | 371 | 352 |



## 산포경향(dispersion)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| class | 붉은 색 | 파란 색 | 녹색 | EMB0000225c2b1a |
| 1 |  |  | 1 |
| 2 |  |  |  |
| 3 |  | 1 | 1 |
| 4 |  |  |  |
| 5 | 7 | 5 | 3 |
| 6 |  |  |  |
| 7 |  | 1 | 1 |
| 8 |  |  |  |
| 9 |  |  | 1 |
| **평균** | **5.0** | **5.0** | **5.0** |

세 자료 모두 평균값이 동일하다.

붉은 색의 경우, 7개의 자료 모두 평균과 같은 값이다.

파란 색의 경우, 5개는 평균과 같고 두 개는 평균보다 약간 크거나 작다.

녹색의 경우, 3개만 평균과 같고 나머지는 평균보다 크거나 작다.

**편차(deviation)** - 관찰치의 평균값으로부터 차이

위의 예에서 각 그룹 별 편차를 구하면

붉은 색의 경우: 편차 = {0, 0, 0, 0, 0, 0, 0}

파란 색의 경우: 편차 = {-2, 0, 0, 0, 0, 0, 2}

녹색의 경우: 편차 = {-4, -2, 0, 0, 0, 2, 4}

세 그룹 모두 편차의 합과 평균 모두 0이다. (합이 0이므로, 평균도 0이 된다.)

붉은 색은 편차들이 모두 0이므로, 합이 0이 되며, 파란 색과 녹색의 경우는 음수의 편차가 존재하기 때문에 합이 0이 된다.

정리하면,

모집단의 경우

평균은 이고 번째 관찰값은 이므로, 번째 편차는 이다.

모집단의 크기를 이라 하면, 총 개의 편차가 존재한다.  
개의 편차들의 평균을 구하면, 이 된다.

표본의 경우

평균은 이고 번째 관찰값은 이므로, DRW0000225c2b70번째 편차는 이다.

표본의 크기를 이라 하면, 총 개의 편차가 존재한다.  
개의 편차들의 평균을 구하면, 이 된다.

그런데, 편차들의 합은 항상 0이다. = 0, = 0

편차의 합은 항상 0이므로, 편차들의 평균도 항상 0이다.

### **평균편차(mean deviation)**

|  |
| --- |
|  |

평균편차란 (편차 절대값)의 평균이다.

1. 음수의 편차는 양수의 편차로 변환한다.
2. =   
   =   
   = 편차 절대값을 모두 다 더하고 (더한 값을) 개수로 나눈 값  
   = 편차 절대값의 평균

### **분산과 표준편차**

|  |  |
| --- | --- |
| 모집단분산 | = |
| 모집단 표준편차 | = |

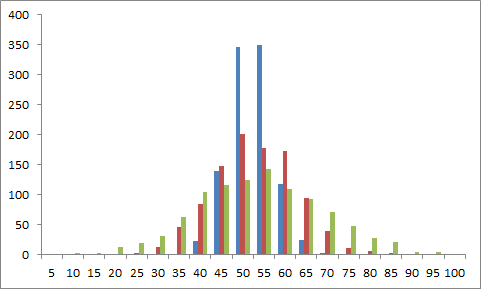
|  |  |
| --- | --- |
| 표본분산 |  |
| 표본의 표준편차 |  |

**분산이란 (편차 제곱)의 평균이다.**

1. 각 편차를 제곱한다.
2. =   
   =   
   = 편차 제곱들을 다 더하고 (그 값을) 갯수로 나눈 값  
   = 편차 제곱의 평균
3. 모집단이면 분산은 , 평균은 로, 갯수는 으로 표기하고, 표본이면 분산은 , 평균은 , 갯수는 으로 표기한다.
4. 평균이란 적절한 수로 나누는 것인데, 모집단의 경우는 으로 나누고 표본인 경우는 -1로 나눈다.
5. 분산의 단위는 원래 자료 단위의 제곱이다.
6. 예: 단위가 미터 – 평균, 편차 단위는 미터, 분산(편차 제곱)의 단위는 미터 제곱
7. 예: 단위가 인원 – 평균, 편차 단위는 인원, 분산(편차 제곱)의 단위는 인원 제곱

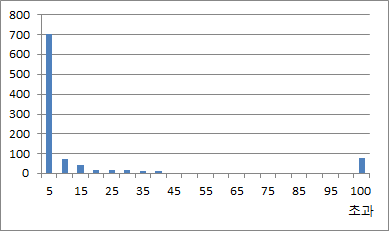
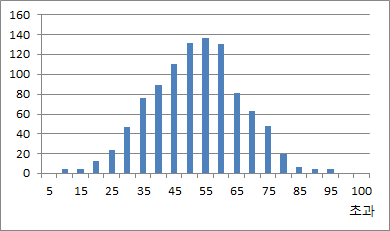
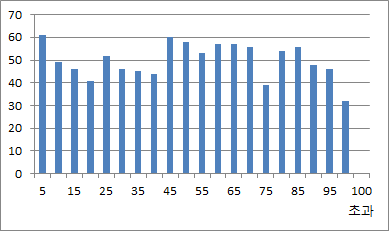
**표준편차란 분산의 제곱근이다.**

1. 원래 자료의 단위와 동일하다.
2. 예: 의 단위가 미터 – 분산(편차 제곱)의 단위는 미터 제곱, 표준편차는 미터
3. 예: 의 단위가 인원 – 분산(편차 제곱)의 단위는 인원 제곱, 표준편차는 인원
4. 모집단의 표준편차는 로, 표본의 표준편차는 로 표기한다.



|  |  |  |  |
| --- | --- | --- | --- |
|  | 파란색 | 붉은색 | 녹색 |
| 평균 | 50.10 | 49.94 | 50.08 |
| 분산 | 24.27 | 98.20 | 231.78 |
| 표준편차 | 4.93 | 9.91 | 15.22 |

**데이터 수가 1,000개인 3개의 자료 비교 II.**



평균 50 50 125,131

최빈값 79 46 0

중앙값 50 50 1

최소값 0 -2 0

최대값 99 98 113,110,156

범위 99 100 113,110,156

분산 797 222 12,929,317,404,095

표준편차 28 15 3,595,736

### 추가 학습 - 편차의 합은 항상 0이다.

=   
 =   
 =   
 = 0

두 소모임(모집단)의 연령별 자료이다. 빈 칸을 모두 채우시오.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 번호 | 나이  Group A | 편차 | 편차  절대값 | 편차  제곱 |  | 나이  Group B | 편차 | 편차  절대값 | 편차  제곱 |
| 1 | 10 |  |  |  |  | 24 |  |  |  |
| 2 | 20 |  |  |  |  | 27 |  |  |  |
| 3 | 30 |  |  |  |  | 30 |  |  |  |
| 4 | 30 |  |  |  |  | 30 |  |  |  |
| 5 | 30 |  |  |  |  | 30 |  |  |  |
| 6 | 40 |  |  |  |  | 33 |  |  |  |
| 7 | 50 |  |  |  |  | 36 |  |  |  |
| 합계 | 210 |  |  |  |  | 210 |  |  |  |
| 평균 | 30 |  |  |  |  | 30 |  |  |  |

두 소모임(모집단) 나이의 표준편차는?

문제 1의 표를 기호로 표기하시오. 번째 자료값은 , 편차, 편차 절대값, 편차 제곱에 사용되는 평균 기호는 로 합니다. 또한 표 하단의 합계는 기호를 사용하고, 표 하단의 평균은 형식으로 표현합니다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 번호 | 나이  Group A | 편차 | 편차  절대값 | 편차  제곱 |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| 7 |  |  |  |  |
| 합계 |  |  |  |  |
| 평균 |  |  |  |  |

표준편차(모집단)의 공식을 기입하시오. 단, 모집단의 크기는 7이다.

두 소모임의 자료가 전체가 아니라 부분이라고 하자. Group A의 분산과 표준편차는?

자료가 모집단이 아니고 표본이다. Group A의 분산과 표준편차 공식은? 단, 표본의 크기는 7이다. (주의: 표본의 평균은 로 표기한다.)

{3, 4, 5, 6}: 답은 소숫점 3자리에서 반올림하여 2자리까지 표기한다.

산술평균은?

중앙값은?

편차의 합은?

평균편차의 개념을 약술하시오.

분산의 개념을 약술하시오.

{3, 4, 5, 6}이 모집단이라면 분산은?

{3, 4, 5, 6}이 모집단이라면 표준편차는?

{3, 4, 5, 6}이 표본이라면 분산은?

{3, 4, 5, 6}이 표본이라면 표준편차는?

{0, 0, 0, 1,000}: 답은 소숫점 3자리에서 반올림하여 2자리까지 표기한다.

산술평균은?

최빈값은?

중앙값은?

편차의 합은?

모집단이라면 분산은?

모집단이라면 표준편차는?

표본이라면 분산은?

표본이라면 표준편차는?

### **체비셰프 부등식(Chebyshev inequality)**

|  |
| --- |
| 어떠한 자료에서도 전체 관찰치의 평균에서 ± 표준편차 이내에 포함될 관찰치의 비율은 최소한 이다. |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 하한값 | 상한값 | 포함 비율 |
| 1.25 | -1.25 | +1.25 | 0.3600 |
| 1.50 | -1.50 | +1.50 | 0.5556 |
| 2.00 | -2.00 | +2.00 | 0.7500 |
| 2.50 | -2.50 | +2.50 | 0.8400 |

는 (단위가 없는) 배수이다. 의 단위는 의 단위와 동일하다.

DRW0000225c2c3a, 모두 자료의 단위와 동일하므로, 덧셈과 뺄셈이 가능하다.

어떤 학교의 1학년 전교생은 400명이다. 이들 학생의 성적 평균은 80점이고, 표준편차는 5점이다.

80점과 90점 사이의 점수를 받은 학생 비율은 얼마인가?

75점~95점 구간 밖의 점수를 받은 학생 비율은 최소 이다. Ans. T, F

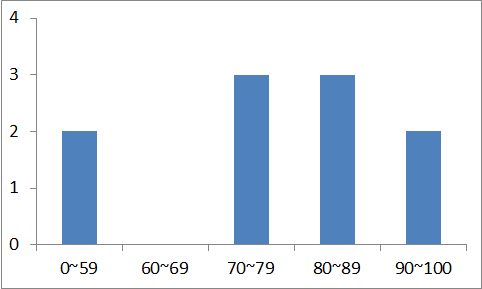
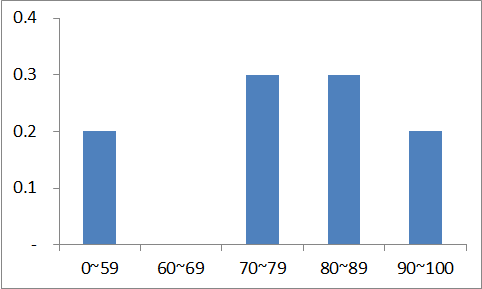
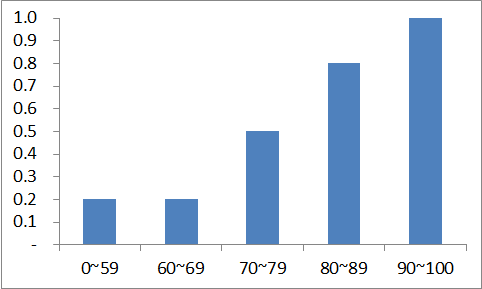
연습문제 정답

1. (1) A, (2) , (3) = 네 번째 계급의 빈도 = 5, (4) = = 0.04

2. (1) 수평축은 계급, 수직축은 도수, (2) 수평축은 계급, 수직축은 누적도수

3. (1)

|  |  |  |  |
| --- | --- | --- | --- |
| 점수 | 빈도수 | 상대빈도수 | 누적상대빈도수 |
| 0~59 | 2 | 0.2 | 0.2 |
| 60~69 | 0 | - | 0.2 |
| 70~79 | 3 | 0.3 | 0.5 |
| 80~89 | 3 | 0.3 | 0.8 |
| 90~100 | 2 | 0.2 | 1.0 |
| 합계 | 10 | 1.0 |  |

(2)   

(3) 0 (해당 계급에 속하는 데이터가 없으면 그 계급의 빈도수는 0이다.)

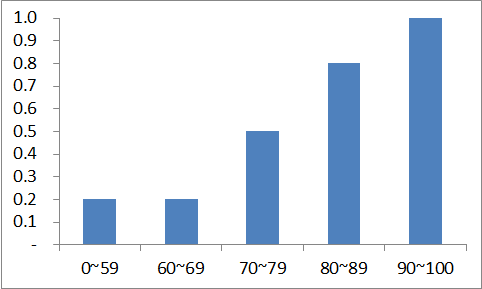
(4) 10 (도수분포그래프에서 그래프의 높이는 빈도와 동일하다. 그러므로 그래프 높이의 합은 데이터의 수와 일치한다.)

(5) (3)에서 본 바와 같이 빈도의 최소값은 0이므로, 빈도를 합계로 나눈 수인 상대빈도도 최소값은 0이 된다.

(6) 1.0. 오직 하나의 계급에서만 빈도가 존재하고 나머지 계급에서는 빈도가 없다고 하자. 예를 들어, 0~89까지는 모두 빈도가 없고 90~100 구간에 10명의 점수가 존재한다고 하자. 이 경우 90~100 구간의 빈도수와 전체 빈도수는 동일하다. 이 경우, 상대빈도값은 1.0이 된다.

(7) False. 상대빈도 그래프의 합은 1.0이 아닐 수 있다. 연습문제 3의 경우, 상대빈도 그래프 높이의 합은 2.7이다.

(8) True. 상대빈도 그래프 최우측 그래프 높이는 항상 1.0이다.

(9)  이 그래프에서 알 수 있듯이 해당 구간의 빈도가 0이면 누적그래프는 그 구간은 전 구간과 높이가 같다.

4. (1) 최소 0명, 최대 3명 (2) 0.200 – 0.150 = 0.050

(3) 최소값은 0.200 (해당 계급의 상대빈도값이 0인 경우)  
 최대값은 1.000 (해당 계급의 상대빈도값이 0.800인 경우)

5. (1) = + + + (2) =

(3) = + + (4) (5) (6)

6. (1) 5 + 70 + 71 = 216,

(2) +++ = 75 + 70 + 71 + 63 = 279,

(3) = = ,

여기서, = 모집단 평균, = 번째 학생 점수

(4) (75 + 70 + + 65) = 75.00, (5) = = ,

여기서, = 표본평균,

(6) (75 + 70 + + 65) = 75.00

7.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 학급 | 수강생 수 | 조사한 학생의 수 | 조사한 학생의 성적 평균 | 모집단 평균 | 표본 평균 |
| 1 | 89 | 20 | 43 | 해당 없음 | = 43 |
| 2 | 20 | 20 | 93 | = 93 | 해당 없음 |
| 3 | 37 | 20 | 32 | 해당 없음 | = 43 |
| 4 | 20 | 20 | 32 | = 32 | 해당 없음 |
| 5 | 74 | 20 | 98 | 해당 없음 | = 98 |
| 6 | 43 | 20 | 87 | 해당 없음 | = 87 |
| 7 | 89 | 20 | 84 | 해당 없음 | = 84 |
| 8 | 20 | 20 | 55 | = 55 | 해당 없음 |
| 9 | 60 | 20 | 64 | 해당 없음 | = 64 |
| 10 | 97 | 20 | 73 | 해당 없음 | = 73 |

8. (1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 반 | 수강생(명) | 평균(D) | 가중치(E) | D\*E |
| A | 50 | 80 | 0.5 | 40 |
| B | 30 | 70 | 0.3 | 21 |
| C | 20 | 85 | 0.2 | 17 |
| 합계 | 100 |  | 1.0 | 78 |

(2) 가중평균 = 6.5 + 6.0 + 5.5 = 6.22시간

9. (1) 최빈값 = B, 중앙값 = B,

(2) 평균 3.4, 최빈값 2, 중앙값 3.5 (5번째 값인 3과 6번째 값인 4의 평균)

10. (1)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 번호 | 나이  Group A | 편차 | 편차  절대값 | 편차  제곱 |  | 나이  Group B | 편차 | 편차  절대값 | 편차  제곱 |
| 1 | 10 | -20 | 20 | 400 |  | 24 | -6 | 6 | 36 |
| 2 | 20 | -10 | 10 | 100 |  | 27 | -3 | 3 | 9 |
| 3 | 30 | - | - | - |  | 30 | - | - | - |
| 4 | 30 | - | - | - |  | 30 | - | - | - |
| 5 | 30 | - | - | - |  | 30 | - | - | - |
| 6 | 40 | 10 | 10 | 100 |  | 33 | 3 | 3 | 9 |
| 7 | 50 | 20 | 20 | 400 |  | 36 | 6 | 6 | 36 |
| 합계 | 210 | - | 60 | 1,000 |  | 210 | - | 18 | 90 |
| 평균 | 30 | - | 8.57 | 142.86 |  | 30 | - | 2.57 | 12.86 |

(2) Group A의 표준편차 = 11.95, Group B의 표준편차 = 3.59

(3)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 번호 | 나이  Group A | 편차 | 편차  절대값 | 편차  제곱 |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| 7 |  |  |  |  |
| 합계 |  |  |  |  |
| 평균 |  |  |  |  |

(4) (5) Group A의 표본의 분산 = 166.67, 표준편차 = 12.91

(6) 분산 = = , 표준편차 = =

11. (1) 4.5, (2) 4.5, (3) 0.0, (4) 편차 절대값의 평균, (5) 편차 제곱의 평균,

(6) 1.25, (7) 1.12, (8) 1.67, (9) 1.29

12. (1) 250, (2) 0, (3) 0, (4) 0, (5) 187,500, (6) 433, (7) 250,000, (8) 500

13. (1) 80점은 평균보다 1표준편차 작은 값, 90점은 평균보다 1표준편차 큰 값에 해당하므로 체비세프 부등식에서 = 2에 해당한다. 그러므로, 최소 1 – = 0.75, 즉 최소 0.75이다

(2) False. 그 구간에 포함될 확률이 최소 1- = 최소 이므로, 포함되지 않을 비율은 최대 이다.