

오늘 할 일

2주차가 어려웠다는 반응이 많았습니다. 오늘은 통째로 리마인드와 퀴즈로 채웁니다.

- 1부 — 1차원 데이터 정리 리마인드: 대표값·산포도·범위와 사분위수·표준화·도수분포표를 개념 → 코드 → 퀴즈 순서로 되짚습니다 (퀴즈 5문항)
- 2부 — LaTeX 입문: 수식을 코드로 쓰는 법. 오늘부터 배운 지표를 직접 타이핑해 봅니다

- 3부 — LaTeX 퀴즈: 2주차 수식을 보고 코드를 쓰거나, 코드를 보고 수식을 맞춥니다 (퀴즈 6문항)

1부는 새 개념이 없습니다. 2주차에서 이미 본 것을 다른 각도로 확인하는 시간입니다.

대표값 세 가지 — 언제 갈라 쓰나

평균은 데이터 전부를 더해 나눕니다. 그래서 극단값 하나에도 확 끌려갑니다.

평균 (mean) — 전부 더해 개수로 나눈다

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

중앙값 (median) — 정렬해서 가운데

$$\text{중앙값} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{이 홀수} \\ \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & n \text{이 짝수} \end{cases}$$

1 중앙값 — 정렬해서 가운데 값. 극단값이 있어도 순서만 밀릴 뿐 값 자체는 안 끌려갑니다

2 최빈값 — 가장 많이 나온 값. 양적 데이터보다 질적 데이터에서 주로 씁니다

최빈값(mode)은 가장 많이 나온 값이라 수식보다 세어 보는 것으로 정합니다. 2주차에서 평균·중앙값은 10명의 영어 점수 scores 로 구했지만, 최빈값만은 [1, 1, 1, 2, 2, 3] 같은 별도 예로 구했습니다 — 점수 10개는 값이 다 달라 최빈값이 잡히지 않기 때문입니다.

평균·중앙값·최빈값 — 코드

```
np.mean(scores)           # 평균
np.median(scores)         # 중앙값
pd.Series([1, 1, 1, 2, 2, 3]).mode() # 최빈값
```

mode() 는 가장 많이 나온 값이 여럿일 수 있어 늘 Series 로 돌아옵니다.

scores_df.mean() 처럼 pandas 도 같은 이름의 함수를 제공합니다. 교재는 둘을 나란히 보여 주는데, 분산에서는 나누는 수가 달라 값이 갈립니다 — 9번 장에서 다시 짚습니다.

이상값이 있을 때 대푯값으로 무엇을 쓰나

10명 중 한 명의 점수가 유난히 낮게 찍혀 있습니다. 양적 데이터라 값이 겹치는 일은 거의 없습니다.

- ① `np.mean(scores)`
- ② `np.median(scores)`
- ③ `pd.Series(scores).mode()`
- ④ `np.std(scores)`

이상값이 있을 때 대푯값으로 무엇을 쓰나

`np.median(scores)`

평균은 모든 값을 더해 나누므로 극단값 하나에도 크게 움직입니다. 중앙값은 정렬 후 가운데 순서를 보므로 그 값이 얼마나 극단적이든 순서만 밀립니다. 최빈값도 이상값에 강하지만, 양적 데이터는 똑같은 값이 거의 없어 대푯값으로 쓰기 어렵습니다 — 그래서 이 경우의 답은 중앙값입니다.

정리 — 대표값에서 산포도로

평균·중앙값·최빈값은 데이터를 한 점으로 요약합니다. 그 점 주변에 얼마나 퍼져 있는지는 아직 안 나옵니다.

1 평균이 같은 두 데이터라도 흩어진 정도는 다를 수 있습니다

2 흩어진 정도를 재는 지표를 산포도(measure of dispersion)라 합니다

다음 묶음은 편차 → 분산 → 표준편차 순서로 짚습니다.

분산은 왜 제곱인가

편차를 그냥 더하면 늘 0 입니다. 흩어진 정도를 재려면 부호를 없애야 합니다.

분산

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

표준편차 — 단위를 원래대로

$$S = \sqrt{S^2}$$

2주차에서 편차의 평균이 0 이 되는 것을 직접 확인했습니다. -14점과 +14점을 같은 크기로 취급하려고 제공합니다.

편차에서 표준편차까지 — 무엇이 리스트이고 무엇이 숫자 하나인가

2주차의 10명 점수로 단계마다 실제 값을 따라갑니다. 리스트로 가다가 평균을 내는 순간 숫자 하나가 됩니다.

단계	무엇을 하나	결과 모양	10명 실제 값
점수	원본 데이터	리스트 10개	42, 69, 56, 41, 57 ...
평균	전부 더해 10으로 나눔	숫자 하나	55.0
편차	각 점수 - 평균	리스트 10개	-13, 14, 1, -14, 2 ...
편차 ²	각 편차를 제곱	리스트 10개	169, 196, 1, 196, 4 ...
분산	편차 ² 의 평균 ← 평균은 여기서	숫자 하나	86.0
표준편차	분산의 제곱근	숫자 하나	9.27

편차의 평균은 늘 0 이라 그 자리에서는 평균을 낼 수 없습니다. 제곱해 부호를 없앤 뒤에 냅니다.

분산과 표준편차 — 코드

```
np.var(scores)          # 분산
np.std(scores)           # 표준편차
np.sqrt(np.var(scores))  # 같은 값
```

ddof=0 이 기본입니다. 이 10명을 모집단으로 보므로 n 으로 나눕니다.

ddof=1 로 주면 n-1 로 나눕니다 — 표본에서 모분산을 추정할 때 씁니다.

분산을 구하는 함수는?

10명의 영어 점수가 scores 에 들어 있습니다.

- ① `np.var(scores)`
- ② `np.std(scores)`
- ③ `np.mean(scores)`
- ④ `np.median(scores)`

분산을 구하는 함수는?

`np.var(scores)`

var 는 variance /'vɛəriəns/ 베어리언스, 곧 분산입니다. std 는 standard deviation /ˌstændərd ˌdiːvi'eiʃn/ 스탠더드 디비에이션, 표준편차입니다. 표준편차는 분산에 제곱근을 씌운 값이라 `np.sqrt(np.var(scores))` 와 같습니다.

정리 — 산포도에서 범위로

분산·표준편차는 모든 데이터를 하나하나 제공해 더합니다. 값이 많으면 계산이 무겁고, 극단값 하나에도 여전히 영향을 받습니다.

1 값을 딱 두 개만 보고 흩어진 정도를 대충 재는 방법도 있습니다 — 범위 (range)

2 이상값에 강하게 만들려면 최댓값·최솟값 대신 사분위수를 씁니다 — IQR

다음 묶음에서 범위와 사분위수 범위를 짚습니다.

최댓값-최솟값은 이상값에 약하다

범위는 값을 딱 두 개만 봅니다. 그중 하나가 극단값이면 범위가 그 값 하나에 좌우됩니다.

범위

$$Rg = x_{\max} - x_{\min}$$

사분위수 범위

$$IQR = Q3 - Q1$$

1 Q1 — 하위 25% 지점, Q2 — 중앙값(하위 50%), Q3 — 하위 75% 지점

2 가장 크고 작은 값 대신, 이상값이 없을 법한 중간 50% 구간의 폭을 봅니다

범위가 이상값에 흔들리는 문제를 IQR 이 완화합니다.

범위와 사분위수 범위 — 코드

```
np.max(scores) - np.min(scores)          # 범위
scores_Q1 = np.percentile(scores, 25)
scores_Q3 = np.percentile(scores, 75)
scores_IQR = scores_Q3 - scores_Q1        # 사분위수 범위
```

np.percentile 은 몇 번째 백분위인지를 0~100 사이의 수로 받습니다 — 25, 75.

np.percentile(scores, 50) 은 np.median(scores) 와 같은 값입니다.

Q3 - Q1 이 뜻하는 것은?

$\text{scores_IQR} = \text{scores_Q3} - \text{scores_Q1}$ 를 구했습니다.

- ① 데이터 중간 50% 가 퍼진 범위
- ② 전체 데이터의 범위
- ③ 이상값의 개수
- ④ 최빈값과 중앙값의 차이

Q3 - Q1 이 뜻하는 것은?

데이터 중간 50% 가 퍼진 범위

Q1 아래 25% 와 Q3 위 25% 를 잘라내고 남는 가운데 50% 구간의 폭이 IQR 입니다. 양 끝의 극단값을 잘라내고 보므로 범위(Rg)보다 이상값에 덜 흔들립니다.

정리 — 범위에서 표준화로

지금까지는 흠어진 정도를 숫자 하나로 요약했습니다. 이번엔 '이 값 하나'가 전체에서 어디쯤 있는지를 잡니다.

1 같은 60점도 어려운 시험과 쉬운 시험에서 의미가 다릅니다

2 평균·표준편차에 안 흔들리는 상대적 위치가 필요합니다 — 표준화

다음 묶음에서 표준화와 편차값을 짚습니다.

다른 시험 점수를 건주는 법

점수 자체는 시험마다 평균·분산이 달라 그대로 비교할 수 없습니다. 평균 0, 표준편차 1 인 척도로 바꾸면 비교가 가능해집니다.

표준화

$$z_i = \frac{x_i - \bar{x}}{S}$$

편차값 — 표준화한 z 를 10 배 하고 50 을 더한 것

$$T_i = 50 + 10z_i$$

표준화한 데이터는 평균이 0, 표준편차가 1 이 됩니다. 편차값은 그 z 를 평균 50, 표준편차 10 짜리 척도로 옮긴 값이라 서로 다른 값입니다 — 그래서 여기서는 좌변을 T 로 구분해 적었습니다. 다만 교재와 노트북은 둘 다 z 로 적으니, 노트북을 채울 때는 교재 표기를 그대로 따르십시오.

표준화·편차값 — 코드

```
z = (scores - np.mean(scores)) / np.std(scores)          # 표준화
z = 50 + 10 * (scores - np.mean(scores)) / np.std(scores)  # 편차값
```

두 줄 다 z 라는 같은 이름을 씁니다 — 표준화 결과에 50 을 더하고 10 을 곱해 편차값을 만듭니다.

표준화한 z 는 평균 0, 표준편차 1 입니다. 편차값은 그 z 를 10배 하고 50 을 더한 것이라 평균 50, 표준편차 10 이 됩니다. 위 코드는 같은 이름 z 에 덮어쓰므로, 두 번째 줄을 실행한 뒤 np.mean(z) 를 찍으면 50 이 나옵니다.

편차값 70 의 뜻은?

$z = 50 + 10 * (\text{scores} - \text{np.mean(scores)}) / \text{np.std(scores)}$ 로 계산했습니다.

- ① 평균보다 표준편차 2 배만큼 위
- ② 평균보다 2 배 높은 점수
- ③ 원점수 70점을 그대로 나타낸 것
- ④ 상위 30등이라는 뜻

편차값 70 의 뜻은?

평균보다 표준편차 2 배만큼 위

편차값 = $50 + 10z$ 이므로 70 이면 $10z = 20$, 즉 $z = 2$ 입니다. 표준편차의 2 배만큼 평균 위에 있다는 상대적 위치를 말할 뿐, 원점수나 등수를 직접 알려주지 않습니다.

정리 — 지표에서 그림으로

숫자 지표는 정확하지만 여러 개를 한꺼번에 보면 눈에 잘 안 들어옵니다. 표와 그림으로 분포 모양 전체를 보는 방법이 남았습니다.

1 값을 구간으로 잘라 세면 도수분포표가 됩니다

2 그 표를 막대로 그리면 히스토그램, 산포를 상자로 그리면 상자 그림입니다

마지막 묶음에서 도수분포표와 그림을 짝꿍합니다.

계급·도수·계급값·상대도수

값을 그대로 늘어놓으면 흩어진 모양이 안 보입니다. 구간으로 잘라 세면 분포 모양이 드러납니다.

계급값 — 구간의 가운데

$$\text{계급값} = \frac{\text{계급의 아래 끝} + \text{계급의 위 끝}}{2}$$

상대도수 — 전체에서 차지하는 비율

$$\text{상대도수} = \frac{\text{그 계급의 도수}}{\text{전체 도수}}$$

1 계급(class) — 나눈 구간. 예: 0~10, 10~20, ...

2 도수(frequency) — 그 계급에 속한 데이터 개수

3 계급값 — 그 계급을 대표하는 값. 구간의 한가운데(아래 끝과 위 끝의 평균)를 씁니다

4 상대도수 — 전체에서 그 계급이 차지하는 비율(도수 / 도수의 합)

5 누적상대도수 — 그 계급까지 상대도수를 누적해 더한 값

도수분포표가 있으면 최빈값도 '도수가 가장 큰 계급의 계급값'으로 구할 수 있습니다 — 값이 다 달라 최빈값이 안 잡히는 양적 데이터에서 특히 쓸모 있습니다.

도수분포표.히스토그램.상자 그림 — 코드

```
freq, _ = np.histogram(english_scores, bins=10, range=(0,100))
```

```
ax.hist(english_scores, bins=10, range=(0,100))
```

```
ax.boxplot(english_scores, tick_labels=['english'])
```

bins 는 계급수. np.histogram 은 도수만 세고, ax.hist 는 그 도수를 그림으로 그립니다.

freq_dist_df.loc[freq_dist_df['frequency'].idxmax(), 'class value'] 로 도수가 가장 큰 계급의 계급값(최빈값)을 찾습니다.

계급폭을 바꾸면 도수분포표의 최빈값이 바뀌는가?

bins=10 이던 히스토그램을 bins=25 로 다시 그렸습니다. 여기서 묻는 것은 '도수가 가장 큰 계급의 계급값', 곧 도수분포표에서 구한 최빈값입니다.

- ① 그렇다 — 계급을 다시 나누면 도수가 가장 큰 계급 자체가 달라질 수 있다
- ② 아니다 — 계급을 어떻게 나누든 같은 계급값이 나온다
- ③ 계급수가 짝수일 때만 바뀐다
- ④ 계급값을 바꾸지 않는 한 바뀌지 않는다

계급폭을 바꾸면 도수분포표의 최빈값이 바뀌는가?

그렇다 — 계급을 다시 나누면 도수가 가장 큰 계급 자체가 달라질 수 있다

도수분포표에서 최빈값은 '도수가 가장 큰 계급의 계급값'으로 정의됩니다. 50명 영어 점수에서 bins=10 일 때는 60~70 계급이 도수 18 로 가장 커 최빈값이 65 였는데, bins=25 로 잘게 나누면 64~68 계급이 도수 10 으로 가장 커 최빈값이 66 이 됩니다. 원자료의 최빈값(57)은 그대로입니다 — 바뀌는 것은 우리가 그은 선 위에서 구한 값 쪽입니다.

2부 · LaTeX 이란

문서 조판 프로그래밍 언어입니다. 수식이 많은 논문·과제·주피터 마크다운에서 널리 씁니다.

1 논문·리포트에서 수식을 깔끔하게 조판합니다

2 Jupyter 마크다운 셀에서 $\$...\$$ 로 바로 렌더링됩니다

3 오늘부터 2주차에서 배운 지표를 LaTeX 코드로 직접 쳐 봅니다

발음은 '라텍'이 보편적으로 권장됩니다. 1984년 레슬리 램포트가 발표했습니다.

노트북에서 수식 쓰는 세 가지 방법

감싸는 방식만 다르고, 안에 쓰는 LaTeX 본체는 셋 다 같습니다.

방법	어디에 쓰나	화면에서
\$ 수식 \$	마크다운 셀 · 문장 속에 섞어	글줄 안에 그대로
\$\$ 수식 \$\$	마크다운 셀 · 문단으로 떼어	가운데 정렬, 크게
display(Math(r'수식'))	코드 셀 · 실행 결과로	Out[n] · \$ 안 붙임

`\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i`

세 방법 모두 이 부분은 똑같이 씁니다 · 노트북에서 직접 확인합니다

→ 화면에 이렇게 나옵니다

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

왜 r' 을 붙이나 — 이걸 빼면 수식이 깨진다

파이썬이 따옴표 안의 백슬래시를 특수문자로 먼저 해석합니다.

쓴 것	RAW STRING 일 때	그냥 따옴표일 때
<code>\frac</code>	<code>\frac</code> 그대로	폼피드 문자 + <code>rac</code>
<code>\nu</code>	<code>\nu</code> 그대로 (그리스 문자 뉴)	줄바꿈 + <code>u</code>
<code>\times</code>	<code>\times</code> 그대로	탭 + <code>imes</code>

```
\frac{1}{n}
```

이 LaTeX 을 `Math()` 로 넘기면

→ r 을 붙였을 때

$$\frac{1}{n}$$

위첨자 ^ 아래첨자 _

x^2

x_i

x_{ij}

글자 하나면 중괄호 없이 써도 되지만, 두 글자 이상이면 중괄호로 묶어야 합니다.

차례로 제곱 · i번째 값 · 중괄호로 묶은 아래첨자

x^2

x_i

x_{ij}

x_{ij} 처럼 중괄호를 빼면 i만 아래첨자가 되고 j는 그냥 옆에 붙어 버립니다.

분수 — `\frac{위}{아래}`

```
\frac{1}{n}
```

```
\frac{x_i - \bar{x}}{S}
```

쓰는 법

차례로 $1/n$ · 표준화 식의 분자분모

$$\frac{1}{n} \quad \frac{x_i - \bar{x}}{S}$$

중괄호 두 쌍이 각각 분자와 분모입니다. 한 글자면 중괄호를 생략해도 되지만 늘 쓰는 편이 안전합니다.

루트 — `\sqrt{}`

`\sqrt{S^2}`

`\sqrt[3]{x}`

대괄호로 몇 제곱근인지 지정합니다 — 생략하면 제곱근입니다.

차례로 제곱근(표준편차) · 세제곱근

$$\sqrt{S^2} \quad \sqrt[3]{x}$$

`\sqrt[n]{}` 의 n 을 비우면 그냥 `\sqrt{}` 와 같은 제곱근이 됩니다.

합 `\sum` 과 적분 `\int`

```
\sum_{i=1}^n x_i  
\int_a^b f(x)\,dx
```

아래첨자로 시작, 위첨자로 끝을 적습니다 — 합과 적분이 같은 문법을 씁니다.

차례로 합 — n 개를 더합니다 · 적분 — a 에서 b 까지

$$\sum_{i=1}^n x_i \qquad \int_a^b f(x) \, dx$$

`\,` 는 좁은 빈 칸입니다. 적분 기호와 dx 사이에 붙여 가독성을 줍니다.

그리스 문자

```
\mu  
\sigma  
\alpha  
\beta  
\lambda
```

역슬래시 + 이름을 그대로 씁니다.

차례로 뮤 · 시그마 · 알파 · 베타 · 람다

μ σ α β λ

첫 글자를 대문자로 쓰면 대문자 그리스 문자가 됩니다 — `\sigma` 는 표준편차, `\Sigma` 는 대문자 시그마입니다. 합 기호는 모양이 비슷해도 명령이 달라 `\sum` 으로 씁니다(위아래 첨자가 기호 위 아래에 붙습니다).

괄호 크기 `\left( \right)` 와 `\bar{\}` `\hat{\}`

```
\left( \frac{x_i - \bar{x}}{S} \right)
\bar{x}
\hat{p}
```

`\left \right` 는 안의 내용 크기에 맞춰 괄호를 자동으로 늘립니다.

차례로 괄호가 분수 높이에 맞춰 커진 모습 · 표본평균 · 추정값

$$\left(\frac{x_i - \bar{x}}{S} \right) \quad \bar{x} \quad \hat{p}$$

괄호를 그냥 () 로 쓰면 안의 분수가 커도 괄호 크기가 그대로라 답답해 보입니다. `\left \right` 를 붙이면 자동으로 늘어납니다.

2주차 수식을 통째로 써 보기 — 분산

```
S^2 = \frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2
```

지금까지 배운 분수·아래첨자·위첨자·합이 모두 들어 있습니다.

분산

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

괄호 ($x_i - \bar{x}$) 다음에 바로 ^{2} 를 붙이면 괄호 전체가 제공됩니다 — 이 순서를 안 지키면 x_i 만 제공되어 버립니다.

3부 · LaTeX 퀴즈

여기서 묻는 것은 LaTeX 문법이지 통계 개념이 아닙니다.

1 앞의 4문항은 코드를 보고 무슨 수식인지 맞힙니다

2 뒤의 2문항은 수식을 보고 코드를 씁니다

3 오늘 다루는 것은 2주차에서 배운 1차원 데이터 수식까지입니다

공분산·상관계수처럼 아직 안 배운 수식은 그 단원을 배우는 주에 함께 씁니다. 배운 것을 손으로 쓰는 것이 순서입니다.

이 LaTeX 코드는 무슨 수식인가

```
\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i
```

- ① 평균
- ② 분산
- ③ 표준편차
- ④ 범위

이 LaTeX 코드는 무슨 수식인가

평균

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

평균

$\bar{x}$ 는 표본평균을 나타내는 표준 표기입니다. 전체를 더해($\sum$) n 으로 나누는($\frac{1}{n}$) 모양이 평균의 정의 그대로입니다.

이 LaTeX 코드는 무슨 수식인가

```
S^2 = \frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2
```

- ① 분산
- ② 표준편차
- ③ 공분산
- ④ 상관계수

이 LaTeX 코드는 무슨 수식인가

분산

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

분산

편차($x_i - \bar{x}$)를 제곱해 평균낸 식이 분산입니다. 제곱 없이 그냥 더하면 늘 0이 되므로 ^{2} 가 들어간 것이 분산의 표식입니다.

이 LaTeX 코드는 무슨 수식인가

```
S = \sqrt{S^2}
```

- ① 표준편차
- ② 분산
- ③ 범위
- ④ 편차값

이 LaTeX 코드는 무슨 수식인가

표준편차

$$S = \sqrt{S^2}$$

표준편차

분산 S^2 에 `\sqrt` 를 씌워 원래 단위로 되돌린 값이 표준편차입니다. 분산은 제곱 단위라 그대로는 원 데이터와 단위가 안 맞습니다.

이 LaTeX 코드는 무슨 수식인가

$$R_g = x_{\{\max\}} - x_{\{\min\}}$$

- ① 범위
- ② 사분위수 범위
- ③ 표준편차
- ④ 분산

이 LaTeX 코드는 무슨 수식인가

범위

$$Rg = x_{\max} - x_{\min}$$

범위

최댓값에서 최솟값을 뺀 값이 범위(Rg)입니다. `\max`, `\min` 처럼 함수 이름은 `\` 를 붙여 로마체로 씁니다.

이 수식을 LaTeX 으로 쓰시오 — 표준화

$$z_i = \frac{x_i - \bar{x}}{S}$$

이 수식을 LaTeX 으로 쓰시오 — 표준화

```
z_i = \frac{x_i - \bar{x}}{S}
```

표준화

$$z_i = \frac{x_i - \bar{x}}{S}$$

아래첨자 `_i`, 분수 `\frac{ }{ }`, `\bar{ }` 세 가지를 조합하면 됩니다.

이 수식을 LaTeX 으로 쓰시오 — 편차값

$$z_i = 50 + 10 \times \frac{x_i - \bar{x}}{S}$$

이 수식을 LaTeX 으로 쓰시오 — 편차값

```
z_i = 50 + 10 \times \frac{x_i - \bar{x}}{S}
```

편차값

$$z_i = 50 + 10 \times \frac{x_i - \bar{x}}{S}$$

표준화 식 앞뒤로 50 을 더하고 10 을 곱하는 부분만 숫자와 \times 로 그대로 적으면 됩니다. 좌변은 교재 표기를 그대로 따라 z_i 로 씁니다 — 18번 장에서 본 T 는 표준화와 구분해야 할 때 쓰는 기호입니다.

이제 노트북에서 직접 채운다

LaTeX.ipynb 의 빈칸을 오늘 배운 문법으로 채웁니다.

1 맨 위 '수식을 쓰는 세 가지 방법' 셀 — 더블클릭해 \$\$ 원본을 보고, 다시 실행해 렌더를 봅시다

2 `display(Math(r'...'))` 의 따옴표 안에 LaTeX 코드를 넣습니다

3 02장 — 평균부터 편차값까지, 오늘 3부에서 다룬 것과 같은 순서입니다

4 03장 이후는 오늘 비워 둡니다. 그 단원을 배우는 주에 채웁니다

노트북에는 03~08장 수식도 들어 있습니다. 단원을 배울 때 그 칸을 채우면 학기 끝에는 전체가 완성됩니다.

다음 주 — 2차원 데이터 정리

2차원 데이터 정리 — 두 변수의 관계를 재는 지표를 배웁니다.

상관계수 — 다음 주에 배울 식

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

1 공분산 — 두 변수가 같은 방향으로 움직이는지

2 상관계수 — 그 관계의 세기를 -1~1로 표준화한 값

이 식을 LaTeX으로 쓰는 것도 다음 주에 함께 합니다. 오늘 익힌 `\frac`과 아래첨자 문법이 그대로 쓰입니다.