5장 학습 정리

5.1 자료의 정리

5.1.1 기본 개념

① (평균)= $\frac{(\text{전체 자료의 합})}{(\text{전체 자료의 개수})}$

② (비율)= $\frac{(\text{비교하는 양})}{(\text{기준량})}$

③ (백분율)(%)=(비율)×100

④ (변화량)=(비교 시점의 자료)-(기준 시점의 자료)

5.1.2 자료의 수집과 정리

(1) 자료 수치에 따른 분류

자료 수치에 따른 분류는 절대 자료, 비율 자료, 지수 자료, 대비 자료의 4가지로 나눌 수 있다.

(2) 자료 측정 시점에 따른 분류

자료 측정 시점에 따른 분류는 일반적 자료와 시계열 자료로 나눌 수 있다.

5.1.3 자료의 특성에 알맞은 그래프로 나타내기

① **선 그래프**: 어떤 항목에 대한 시간적 추이를 나타낼 때 사용한다.

② **막대그래프:** 어떤 항목을 영역별로 비교하여 나타낼 때 사용한다.

③ **원 그래프**: 어떤 항목을 전체에 대한 영역별 비교를 나타낼 때 사용된다.

④ **방사형 그래프**: 여러 가지 항목에 대한 척도를 비교하여 나타낼 때 사용된다.

⑤ **누적막대그래프**: 한 구간이 몇 개의 세부 항목으로 나뉘면서 전체의 합이 의미가 있을 때 사용한다.

5.2 줄기와 잎 그림과 도수분포표

5.2.1 줄기와 잎 그림

줄기와 잎 그림: 줄기와 잎을 이용하여 자료를 나타낸 그림

5.2.2 도수분포표

도수분포표: 주어진 자료를 몇 개의 계급으로 나누고 각 계급에 속하는 도수를 조사하여 나타낸 표

도수분포표에서의 평균

$$(\text{평균}) = \frac{\{(\text{계급값})\times(\text{도수})\}\text{의 총합}}{(\text{도수})\text{의 총합}}$$

5.2.3 히스토그램과 도수분포다각형

히스토그램: 도수분포표의 각 계급의 끝 값을 가로축에, 도수를 세로축에 적고 계급의 크기를

가로로, 도수를 세로로 하는 직사각형으로 나타낸 그래프
도수분포다각형: 히스토그램에서 각 직사각형의 윗변의 중앙의 점과 그래프의 양 끝에 도수가 0인 계급이 하나씩 있는 것으로 생각하여 그 중앙의 점을 선분으로 연결하여 그린 그래프

5.3 상대도수와 누적도수

5.3.1 상대도수

상대도수 : $(\text{어떤 계급의 상대도수}) = \dfrac{(\text{그 계급의 도수})}{(\text{전체 도수})}$

5.3.2 누적도수

누적도수 : 도수분포표에서 처음 계급의 도수부터 어떤 계급까지의 도수를 모두 더한 합

5.4 대푯값과 산포도

5.4.1 대푯값

평균, 중앙값, 최빈값

⑴ $(\text{평균}) = \dfrac{(\text{변량의 총합})}{(\text{변량의 수})}$

⑵ 변량의 개수가 n인 자료의 중앙값

① n이 홀수인 경우 : 변량을 크기순으로 나열하였을 때, 중앙값은 $\dfrac{n+1}{2}$번째 변량이다.

② n이 짝수인 경우 : 변량을 크기순으로 나열하였을 때, 중앙값은 $\dfrac{n}{2}$번째와 $\left(\dfrac{n}{2}+1\right)$번째 변량의 평균이다.

⑶ 최빈값 : 변량 중에서 가장 많이 나타난 값

5.4.2 산포도

분산과 표준편차

⑴ $(\text{분산}) = \dfrac{\{(\text{편차})^2\text{의 총합}\}}{(\text{변량의 개수})}$ ⑵ $(\text{표준편차}) = \sqrt{(\text{분산})}$

5.4.3 도수분포표에서 분산과 표준편차 구하기

도수분포표에서의 분산과 표준편차

❶ $(\text{분산}) = \dfrac{\{(\text{편차})^2 \times (\text{도수})\}\text{의 총합}}{(\text{도수})\text{의 총합}}$

❷ $(\text{표준편차}) = \sqrt{(\text{분산})}$

5.5 상관관계

5.5.1 산점도와 상관관계

산점도: 두 가지 자료를 순서쌍으로 하여 좌표평면 위에 나타낸 그림

상관관계 : 두 자료 사이에 어떤 관계가 있음을 나타내는 것

5.5.2 상관표

상관표: 두 변량의 도수분포표를 함께 나타낸 표

상관계수(표본상관계수)

측정값 $(x,\ y)$에 대하여 n개의 측정값 $(x_1,\ y_1)$, $(x_2,\ y_2)$, $\cdots$, $(x_n,\ y_n)$이 주어졌을 때, $x,\ y$사이의 상관계수는 다음과 같다.

$$r=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{(n-1)s_x s_y}$$

여기서 $\overline{x}$, $\overline{y}$는 각각 $x,\ y$의 평균이고, $s_x,\ s_y$는 각각 $x,\ y$의 표준편차이다.

5.6 통계

5.6.1 확률변수와 확률분포

확률변수: 어떤 시행에서 표본공간의 각 원소에 하나의 실수 값을 대응시키는 것

확률분포: 확률변수 X가 갖는 값과 X가 이 값을 가질 확률의 대응 관계

5.6.2 이산확률변수

이산확률변수: 확률변수 X가 가질 수 있는 값을 셀 수 있을 때,

확률질량함수: 이산확률변수 X가 가질 수 있는 모든 값이 $x_1,\ x_2,\ x_3,\ \cdots,\ x_n$일 때 이 값을 가질 확률 $p_1,\ p_2,\ p_3,\ \cdots,\ p_n$에 대응되는 함수

$$\mathrm{P}(X=x_i)=p_i\ (i=1,\ 2,\ \cdots,\ n)$$

확률질량함수의 성질

이산확률변수 X의 확률질량함수 $\mathrm{P}(X=x_i)=p_i\ (i=1,\ 2,\ \cdots,\ n)$에 대하여

❶ $0\le \mathrm{P}(X=x_i)\le 1$

❷ $\sum_{i=1}^{n}\mathrm{P}(X=x_i)=1$

❸ $\mathrm{P}(x_i\le X\le x_j)=\sum_{k=i}^{j}\mathrm{P}(X=x_k)$ (단, $j=1,\ 2,\ \cdots,\ n,\ i\le j$)

5.6.3 이산확률변수의 기댓값과 표준편차

이산확률변수 X의 기댓값(평균), 분산, 표준편차

이산확률변수 X의 확률질량함수가

$$\mathrm{P}(X=x_i)=p_i \ (i=1,\ 2,\ \cdots,\ n)$$

일 때

❶ 기댓값(평균) $\mathrm{E}(X)=\sum_{i=1}^{n} x_i p_i$

❷ 분산 $\mathrm{V}(X)=\mathrm{E}((X-m)^2)=\sum_{i=1}^{n}(x_i-m)^2 p_i$ (단, $m=\mathrm{E}(X)$)

❸ 표준편차 $\sigma(X)=\sqrt{\mathrm{V}(X)}$

이산확률변수 $aX+b$의 평균, 분산, 표준편차

이산확률변수 X와 두 상수 $a(a\neq 0)$, b에 대하여

❶ $\mathrm{E}(aX+b)=a\mathrm{E}(X)+b$

❷ $\mathrm{V}(aX+b)=a^2\mathrm{V}(X)$

❸ $\sigma(aX+b)=|a|\sigma(X)$

5.6.4 이항분포

이항분포 $\mathrm{B}(n,\ p)$: $\mathrm{P}(X=x)={}_n\mathrm{C}_x p^x q^{n-x}$ $(x=0,\ 1,\ 2,\ \cdots,\ n,\ q=1-p)$

이항분포의 평균, 분산, 표준편차

확률변수 X가 이항분포 $\mathrm{B}(n,\ p)$를 따를 때,

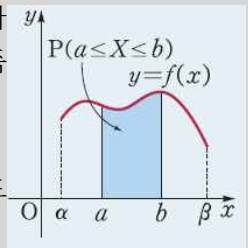
$$\mathrm{E}(X)=np,\ \mathrm{V}(X)=npq,\ \sigma(X)=\sqrt{npq}\ (\text{단, } q=1-p)$$

5.6.5 연속확률분포

연속확률변수: 어떤 연속하는 범위 안에서 모든 실수의 값을 가지는 확률변수

확률밀도함수의 성질

$\alpha \le X \le \beta$에서 모든 실수 값을 가질 수 있는 연속확률변수 X에 대하여 $\alpha \le x \le \beta$에서 정의된 확률밀도함수 $f(x)$는 다음 세 가지를 만족시킨다.



❶ $f(x)\ge 0$

❷ $y=f(x)$의 그래프와 x축 및 두 직선 $x=\alpha$, $x=\beta$로 둘러싸인 도형의 넓이가 1이다.

❸ $\mathrm{P}(a\le X\le b)$는 $y=f(x)$의 그래프와 x축 및 두 직선 $x=a$, $x=b$로 둘러싸인 도형의 넓이와 같다. (단, $\alpha \le a \le b \le \beta$)

연속확률변수 X의 기댓값(평균)과 분산

확률변수 X의 확률밀도함수가 $f(x)$ $(a\le x\le b)$일 때, X의 기댓값 $\mathrm{E}(X)$와 분산 $\mathrm{V}(X)$는 각각 다음과 같다.

$$\mathrm{E}(X)=\int_a^b xf(x)dx,\ \mathrm{V}(X)=\int_a^b (x-m)^2 f(x)dx,\ (\text{단, } m=\mathrm{E}(X))$$

5.6.6 정규분포

정규분포 : 연속확률변수 X의 확률밀도함수가

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (-\infty<x<\infty)$$

인 확률분포

■ **정규분포의 표준화**

정규분포의 표준화

확률변수 X가 정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따를 때, 확률변수

$$Z=\frac{X-m}{\sigma}$$

은 표준정규분포 $\mathrm{N}(0,\ 1)$을 따른다.

5.6.7 이항분포와 정규분포의 관계

이항분포와 정규분포의 관계

확률변수 X가 이항분포 $\mathrm{B}(n,\ p)$ $(0<p<1)$을 따를 때, n의 값이 충분히 크면 X는 정규분포 $\mathrm{N}(np, npq)$ $(q=1-p)$를 따른다.

5.6.8 모집단과 표본

표본평균의 평균, 분산, 표준편차

모평균이 m이고 모표준편차가 σ인 모집단에서 임의추출한 크기가 n인 표본의 표본평균 $\overline{X}$에 대하여 기댓값, 분산, 표준편차는 각각 다음과 같다.

$$\mathrm{E}(\overline{X})=m,\ \mathrm{V}(\overline{X})=\frac{\sigma^2}{n},\ \sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}$$

표본평균의 분포

정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는 모집단에서 임의추출한 크기가 n인 표본의 표본평균을 $\overline{X}$라고 할 때, $\overline{X}$는 정규분포 $\mathrm{N}\left(m,\ \frac{\sigma^2}{n}\right)$을 따른다.

5.6.9 모평균의 추정

추정 : 모평균, 모표준편차와 같이 모집단의 특성을 나타내는 값을 표본을 이용하여 추측하는 것

모평균의 신뢰구간

정규분포 $\mathrm{N}(m,\ \sigma^2)$을 따르는 모집단에서 임의추출한 크기가 n인 표본의 표본평균 $\overline{X}$의 값이 $\overline{x}$일 때, 모평균 m의 신뢰구간은 다음과 같다.

❶ 신뢰도 95 %의 신뢰구간: $\overline{x}-1.96\dfrac{\sigma}{\sqrt{n}} \le m \le \overline{x}+1.96\dfrac{\sigma}{\sqrt{n}}$

❷ 신뢰도 99 %의 신뢰구간: $\overline{x}-2.58\dfrac{\sigma}{\sqrt{n}} \le m \le \overline{x}+2.58\dfrac{\sigma}{\sqrt{n}}$

5.6.10 모비율의 추정

표본비율의 분포

모비율이 p이고 표본의 크기 n이 충분히 클 때, 표본비율 $\hat{p}$은 근사적으로 정규분포 $\mathrm{N}\left(p,\ \dfrac{pq}{n}\right)$를 따른다. 따라서 확률변수 $Z=\dfrac{\hat{p}-p}{\sqrt{\dfrac{pq}{n}}}$는 근사적으로 표준정규분포 $\mathrm{N}(0,\ 1)$을 따른다. (단, $q=1-p$)

모비율의 신뢰구간

모집단에서 임의추출한 크기가 n인 표본의 표본비율 $\hat{p}$에 대하여 표본의 크기 n이 충분히 크면 모비율 p의 신뢰구간은 다음과 같다. (단, $\hat{q}=1-\hat{p}$)

❶ 신뢰도 95 %의 신뢰구간: $\hat{p}-1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}} \le p \le \hat{p}+1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$

❷ 신뢰도 99 %의 신뢰구간: $\hat{p}-2.58\sqrt{\dfrac{\hat{p}\hat{q}}{n}} \le p \le \hat{p}+2.58\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$