학습계획서

팀	호성이와 아이들	구성원	서호성, 최홍용

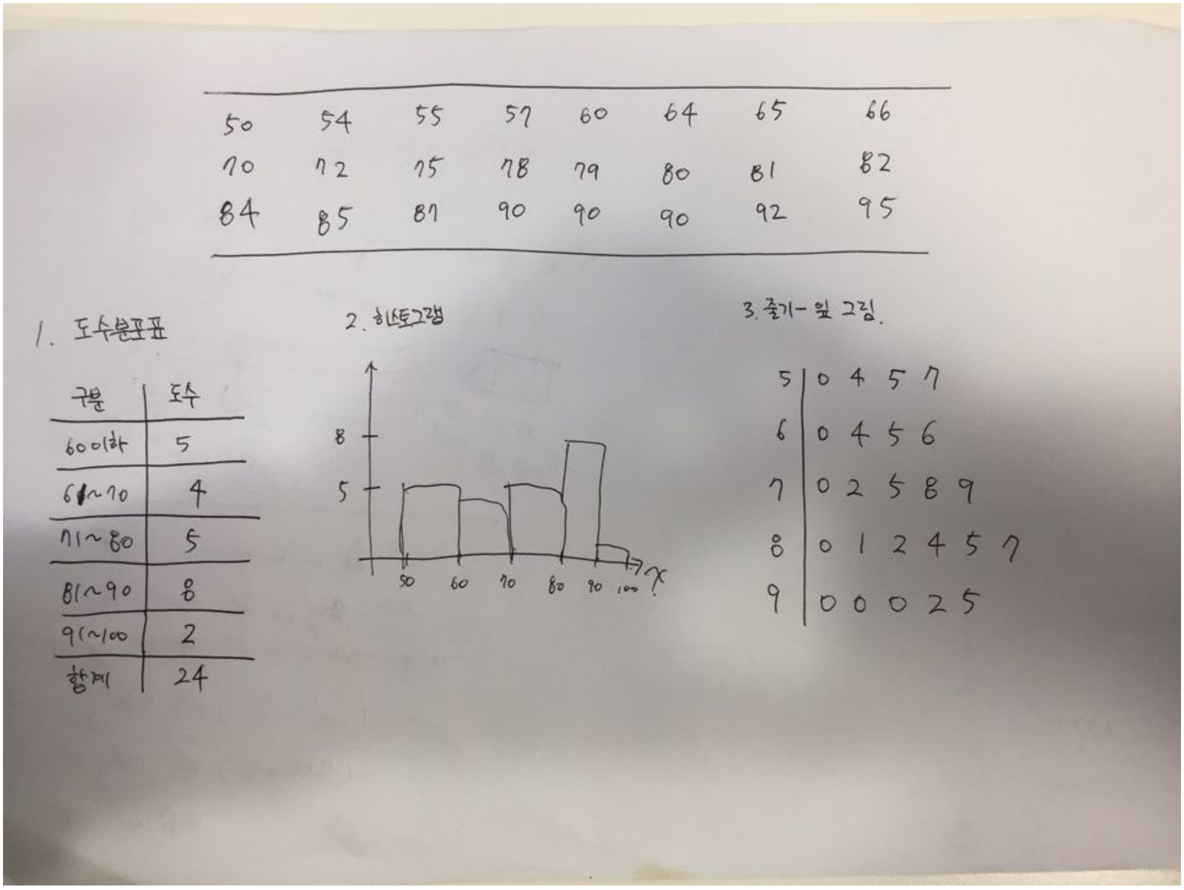
일정	발제자	주제	주요내용
1일차 (5 / 27)	서호성	Big Data Learning Packet : Basics of Statistics	1. data form 2. random variable & distribution 1 3. random variable & distribution 2
2일차 (5 / 28)	최홍용	Big Data Learning Packet : Basics of Statistics	4. normal distribution 5. sampling distribution & central limit theorem 6. statistical inference
3일차 (5 / 29)	서호성	Big Data Learning Packet : Basics of Statistics	7. statistical testing 8. population mean testing 9. correlation analysis
4일차 (5 / 30)	최홍용	Big Data Learning Packet : Basics of Statistics	10. simple linear regression 11. analysis of variance
5일차 (5 / 31)	서호성	Computational Thinking and Data Science	Introduction and Optimization Problems
6일차 (6 / 3)	최홍용	Computational Thinking and Data Science	Optimization Problems
7일차 (6 / 4)	서호성	Computational Thinking and Data Science	Graph-theoretic Models
8일차 (6 / 5)	최홍용	Computational Thinking and Data Science	Stochastic Thinking
9일차 (6 / 7)	서호성	Computational Thinking and Data Science	Random Walks
10일차 (6 / 10)	최홍용	Computational Thinking and Data Science	Monte Carlo Simulation

학습 정리

팀	호성이와 아이	구성원	서호성, 최홍용

일정	발제자	주제
(5 / 27)	서호성	Big Data Learning Packet : Basics of Statistics

주요 내용 요약



50	54	55	57	60	64	65	66
70	72	75	78	79	80	81	82
84	85	87	90	90	90	92	95

1. 도수분포표

구분	도수
60이하	5
61~70	4
71~80	5
81~90	8
91~100	2
합계	24

2. 히스토그램

3. 줄기-잎 그림

5	0 4 5 7
6	0 4 5 6
7	0 2 5 8 9
8	0 1 2 4 5 7
9	0 0 0 2 5

1. 어느 전기부품이 고장 날 때까지 걸리는 시간을 조사하기 위하여 24개 부품을 실험한 결과 다음의 자료를 얻었다.

44	48	64	51	32	29	48	39	57	55
101	49	74	59	56	62	60	37	61	73
122	45	69	52						

(1) 이 표본에서 고장 날 때까지 걸린 시간의 평균을 구하라.

$$m = \frac{\Sigma x_i}{n} = \frac{(44+48+\cdots+52)}{24} = 57.79$$

(2) 고장 날 때까지 걸린 시간의 표준편차를 구하라.

$$s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{(44-57.79)^2+\cdots+(52-57.79)^2}{23}} = 20.51$$

1. 세명의 학생이 각각 백화점에서 구두나 운동화 중 하나를 산다.
서로의 구매에 영향을 받지 않고, 모두 반반의 가능성을 가지고 결정한다.
여기서 확률변수 X를 세명 중 구두를 구매한 학생의 수라고 할 때,
평균, 분산을 구하여라.

A : 구두 구매 B : 운동화 구매

value of X	0	1	2	3
사건	BBB	ABB BAB BBA	AAB ABA BAA	AAA

- 확률분포표

X	0	1	2	3	
P_r	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

(1) 평균

$$E(X) = \Sigma\, x \cdot P(X=x) = 0\times\frac{1}{8} + 1\times\frac{3}{8} + 2\times\frac{3}{8} + 3\times\frac{1}{8} = \frac{3}{2}$$

(2) 분산

$$\begin{aligned} V(X) &= \Sigma\,(X-M)^2 \cdot P(X=x) \\ &= E(X^2) - \{E(X)\}^2 \\ &= 0^2\times\frac{1}{8} + 1^2\times\frac{3}{8} + 2^2\times\frac{3}{8} + 3^2\times\frac{1}{8} - \left(\frac{3}{2}\right)^2 \\ &= \frac{3}{4} \end{aligned}$$

2. 확률 밀도 함수가

$$f(x) = \begin{cases} c\,(4x-2x^2) & 0<x<2 \\ 0 & \text{그 이외} \end{cases}$$

(1) c의 값은 얼마인가?

$$1 = \int_{-\infty}^{\infty} f(x)\,dx$$

$$= \int_{-\infty}^{0} f(x)dx + \int_{0}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx$$

$$= \int_{0}^{2} c\,(4x-2x^2)dx = c\left[2x^2 - \frac{2}{3}x^3\right]_0^2 = \frac{8}{3}c.$$

$$\therefore\ c = \frac{3}{8}$$

(2) $P(X>1)$을 구하라.

$$P(X>1) = \int_1^{\infty} f(x)dx = \int_1^2 f(x)dx + \int_2^{\infty} f(x)dx$$

$$= \int_1^2 f(x)dx = \frac{3}{8}\left[2x^2 - \frac{2}{3}x^3\right]_1^2$$

$$= \frac{3}{8}\left\{\left(8 - \frac{16}{3}\right) - \left(2 - \frac{2}{3}\right)\right\} = \frac{1}{2}$$

학습 정리

팀	호성이와 아이들	구성원	서호성, 최홍용

일정	발제자	주제
(5 / 28)	최홍용	Big Data Learning Packet : Basics of Statistics

주요 내용 요약

정규분포 - 연습문제.

문제 어느 회사에 입사를 희망한 지원자의 영어점수는 평균이 700이고 표준편차가 100인 정규분포를 따른다고 한다.

1) 합격자 중 영어점수가 최저인 사람의 점수가 870점일 때, 몇 퍼센트의 지원자가 합격하겠는가?

x : 입사를 희망한 지원자의 영어점수

$x \sim N(700, 100^2)$

$$P(x \ge 870) = P\left(Z \ge \frac{870-700}{100}\right)$$
$$= P(Z \ge 1.7)$$
$$= P(Z \le -1.7) = 0.0446$$

2) 상위 15%를 선택하기 위한 커트라인은 얼마인가?

$P(\cancel{z} X \ge x)$

$= P\left(Z \ge \frac{x-700}{100}\right)$

$= 0.15$

$z = 1.036$

$P(Z \ge 1.036) = 0.15$

$\frac{X-700}{100} = 1.036 \Rightarrow x = 803.6$점

표본분포와 중심극한정리 - 연습문제.

문제 어느 도시 근로자의 수입은 평균 250만원이고, 표준편차는 50만원이라고 한다.

(1) 100명을 표본으로 선택했을 때, 표본 평균의 분포는 무엇인가?

단위 : 만원

x : 도시 근로자의 수입.

$\bar{x} \sim N(250, 5^2)$

$E(\bar{x}) = 250$

$\sigma(\bar{x}) = 50/\sqrt{100} = 5$

(2) $P(\bar{x} > 260$만원$)$은 얼마인가?

$P(\bar{x} \geq 260) = P\left(z > \frac{260-250}{5}\right)$

$= P(z > 2) = 0.0228$

문제 모집단의 평균이 550이고, 표준 편차가 70일때, 다음의 각 경우에 표본평균 $\bar{x}$의 분포를 구하라.

(1) 표본의 크기는 16으로 한다.

$\bar{x} \sim N\left(M, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

$E(\bar{x}) = 550$

$\sigma(\bar{x}) = \frac{70}{\sqrt{16}} = 17.5$

$\bar{x} \sim N(550, 17.5^2)$

(2) 표본의 크기는 160으로 한다.

$\bar{x} \sim N\left(M, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

$E(\bar{x}) = 550$

$\sigma(\bar{x}) = \frac{70}{\sqrt{160}} = 5.53$

$\bar{x} \sim N(550, 5.53^2)$

통계적 추론 - 연습문제.

문제. 울산시에 거주하는 성인 1인당 신용카드 보유 개수를 추정하기 위해 25명을 무작위로 추출하여 조사한 결과 1인당 평균 7.3개, 표준편차는 2.6개였다. 울산시민 1인당 모평균 신용카드 보유 개수에 대한 99% 신뢰구간을 구하시오.

$n = 25$ CI $\bar{x} \pm t^{*} \frac{s}{\sqrt{n}}$

$\bar{x} = 7.3$ $7.3 \pm 2.797 \cdot \frac{2.6}{\sqrt{25}}$

$S = 2.6$ $= (5.846, 8.754)$

$5.846 \le \mu \le 8.754$.

문제 기존 컴퓨터의 시분할시스템을 조사한 결과 편집명령에 대한 반응시간이 표준편차가 25밀리초인 정규분포를 따르는 사실을 알았다. 새로운 운영체제가 설치되었다. 이 시스템에 대한 평균 반응시간을 추정하고 싶다. 이 시스템의 반응시간도 표준편차가 25밀리초인 정규분포를 따른다고 가정할 때, 모평균에 대한 95% 신뢰구간의 폭이 최대 10이 되게끔 추정하려면 표본크기를 얼마로 정해야 할까?

$\sigma = 25$

$l = 2 \times \frac{\sigma}{\sqrt{n}} \le 10$

$2 \times 1.96 \times \frac{25}{\sqrt{n}} \le 10$.

$\left(\frac{2 \times 1.96 \times 25}{10}\right)^2 \le n \longrightarrow 96.04 \le n$.

$\therefore n = 97$