학습계획서

팀	슈퍼빅데이터	구성원	최재림, 황성욱

일정	발제자	주제	주요내용
1일차 (5 / 27)	황성욱	- 오리엔테이션	향후 학습방향 조정 및 교육컨텐츠 선정
2일차 (5 / 28)	최재림	- 자료의 형태와 요약1, 2	1-1. 자료(변수)의 두 가지 형태 1-2. 범주형 자료의 요약 1-3. 양적 자료의 요약 2-1. 대표값 2-2. 산포도 2-3. 사분위범위 2-4. 상자그림
3일차 (5 / 29)	황성욱	- 확률변수와 분포	1. 확률과 임의성 2. 이산확률변수 3. 연속확률변수 4. 평균과 분산
4일차 (5 / 30)	최재림	- 정규분포	1. 정규분포의 개념 2. 정규분포의 형태 3. 정규분포의 표준화
5일차 (5 / 31)	황성욱	- 표본분포와 중심극한정리	1. 모집단과 표본 2. 표본분포 3. 중심극한정리
6일차 (6 / 3)	최재림	- 통계적 추론 및 검정	1-1. 통계적 추론 1-2. 통계적 추정 1-3. 표본크기의 결정 1-4. t-분포 1-5. 일표본 t-신뢰구간 2-1. 통계적 검정의 개념 2-2. 두 종류의 가설 2-3. 두 종류의 오류 2-4. P값 2-5. P값을 이용한 유의성 검정의 단계 2-6. 단측검정과 양측검정
7일차 (6 / 4)	황성욱	- 모평균에 대한 검정	1. 모평균의 검정: Z-검정 2. 검정통계량 3. 양측검정과 단측검정의 경우 P값 4. 신뢰구간과 가설검정 5. 일표본 t-검정 6. 유의성 검정에 대한 주의점 7.정규분포가 아닐 때의 추론
8일차 (6 / 5)	최재림	- 상관분석	1. 상관분석의 개념 2. 상관계수 3. 상관계수 행렬
9일차 (6 / 7)	황성욱	- 단순선형 회귀분석	1. 단순선형회귀분석의 개념 2. 회귀직선의 적합 3. 최소제곱회귀직선 4. Y값의 예측 : 내삽법(보간법)과 외삽법(보외법) 5. 결정계수 6. 변수변환 7. 회귀계수의 검정 8. 회귀직선의 유의성검증 9. 잔차의 분석
10일차 (6 / 10)	최재림	- 분산분석	1. 분산분석의 개념 2. 일원분류분산분석의 모형 3. ANOVA F-검정 4. 분산분석표

학습 정리

팀	슈퍼빅데이터	구성원	최재림, 황성욱

일정	발제자	주제
1일차 (5 / 27)	황성욱	- 오리엔테이션
주요 내용 요약		

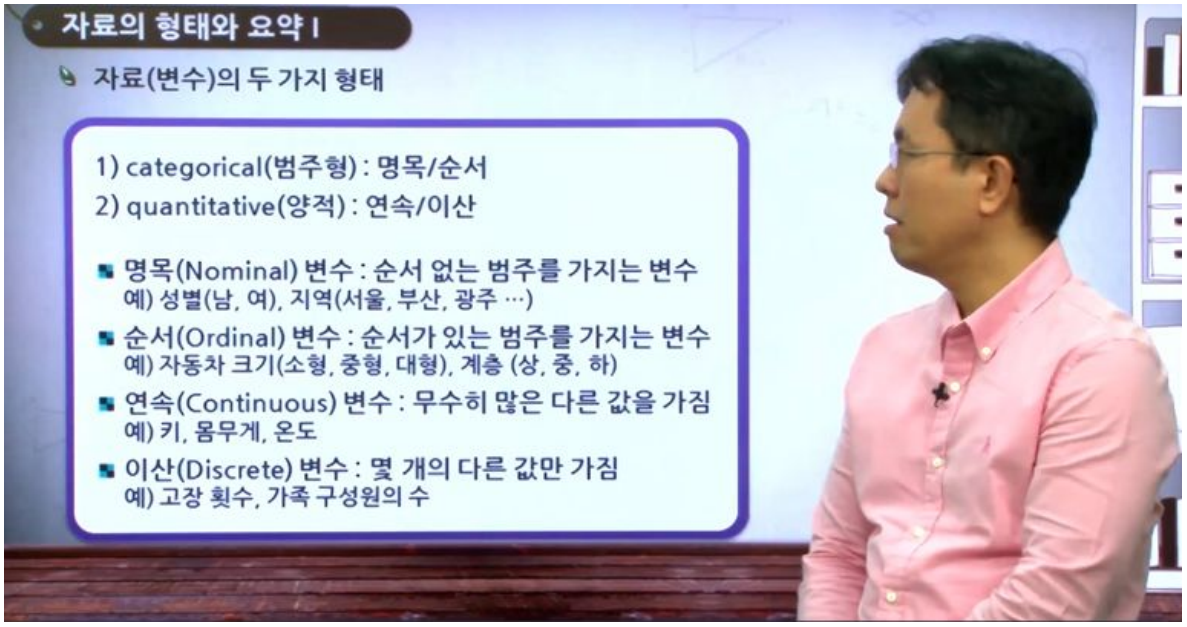
1. 학습 주제 선정
2. 팀 구성
3. 빅데이터 학습 플랫폼 선정
4. 구체적인 학습 계획 수립

학습 정리

팀	슈퍼빅데이터	구성원	최재림, 황성욱

일정	발제자	주제
2일차 (5 / 28)	최재림	- 자료의 형태와 요약1, 2

주요 내용 요약



자료는 범주형/양적 자료로 나누어 지며, 각 자료형을 요약하기 위해서 사용되는 기법은 다음과 같다.

1.범주형
- 도수분포표
- 막대그래프
- 파이 챠트

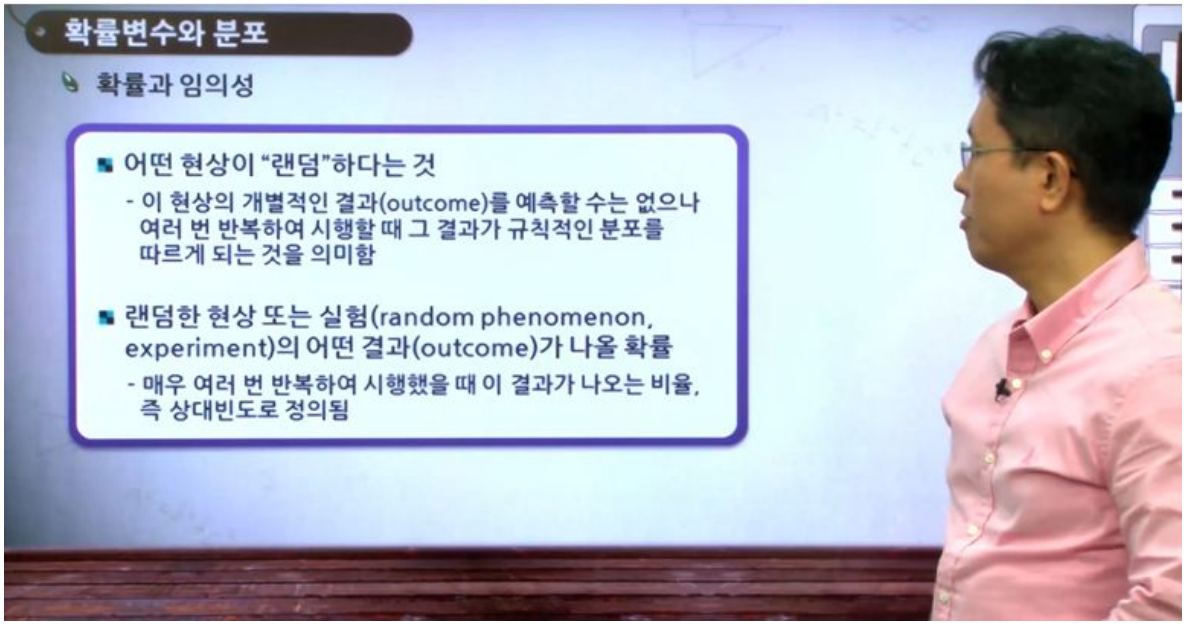
2.양적
- Graphical 요약
  Dotplot, Stamplot, Histogram, Boxplot, Linegraph 등등..
- 수치적 요약
  대표값(산술평균, 중앙값, 최빈값), 산포도(범위, 사분위범위, 표준편차)

학습 정리

팀	슈퍼빅데이터	구성원	최재림, 황성욱

일정	발제자	주제
3일차 (5 / 29)	황성욱	- 확률변수와 분포

주요 내용 요약



- 확률과 임의성 = 현상이 랜덤하다
- 확률변수 = 현상 또는 실험의 결과로 결정되는 수치적인 양 (numerical quantity)
  일정한 확률분포를 가짐(이산형/연속형)

ex) 동전던지기
  동전던지기의 결과는 랜덤하지만 각 시행(던지기)이 독립적이라는 가정하에 여러 번 던졌을 때의 결과는 예측 가능하다
X = 각 시행에서 앞면이 나오는 횟수 (확률변수)
P(X=1) = p
P(X=0) = 1-p (확률분포)

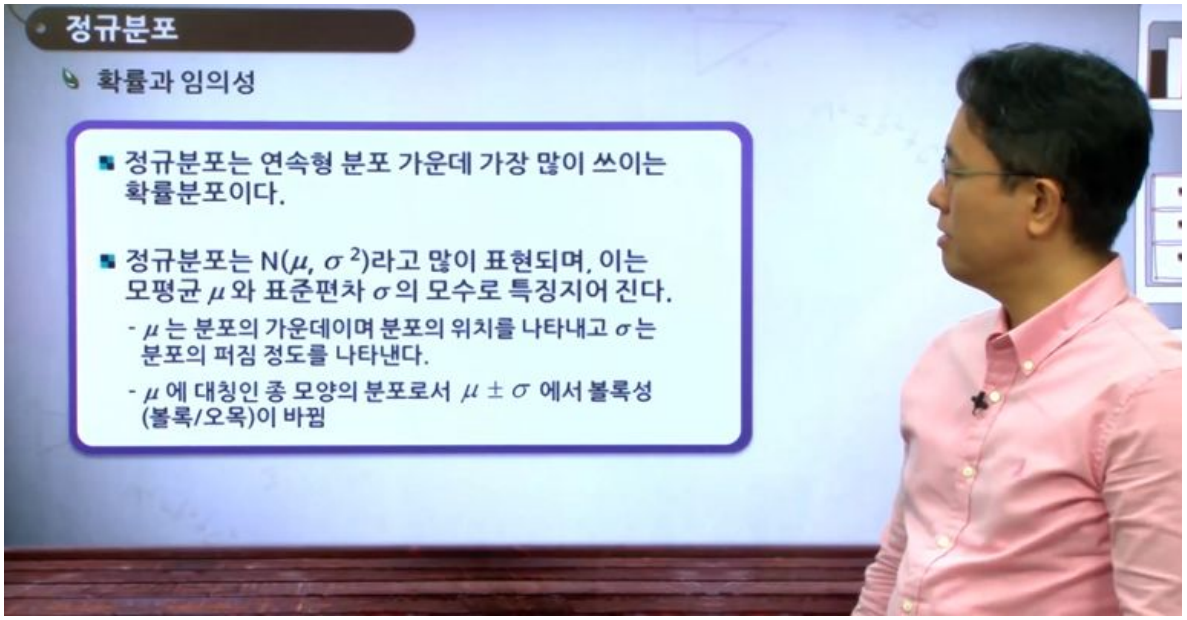
- 이산확률변수 = 유한 또는 셀 수 있는 무한의 값만을 가질 수 있음
- 연속확률변수 = 어떤 구간 안의 모든 값을 다 취할 수 있는 변수
- 평균과 분산

학습 정리

팀	슈퍼빅데이터	구성원	최재림, 황성욱

일정	발제자	주제
4일차 (5 / 30)	최재림	- 정규분포

주요 내용 요약

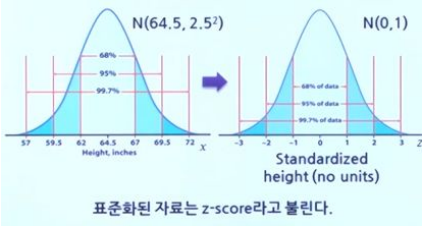


-정규분포의 개념 =
정규분포는 연속형 분포 가운데 가장 많이 쓰이는 확률 분포이다.
모평균과 표준편차의 모수로 특징지어 진다.

-표준 정규분포 = 모평균:0, 표준편차:1인 정규분포

-정규분포의 형태 = 표준편차가 작은 경우에는 평균 주위에 가깝게 몰려있게되고,
표준편차가 큰 모집단의 분포는 넓게 퍼져있는 형태를 취한다.

-정규분포의 표준화 =
모든 정규분포는 같은 형태적 성질을 갖기 때문에 N(m,s^2)를 표준화해 표준 정규분포 N(0,1)을 얻을 수 있고, 표준화 후 N(0,1)의 확률표를 이용해 확률계산을 할 수 있다.

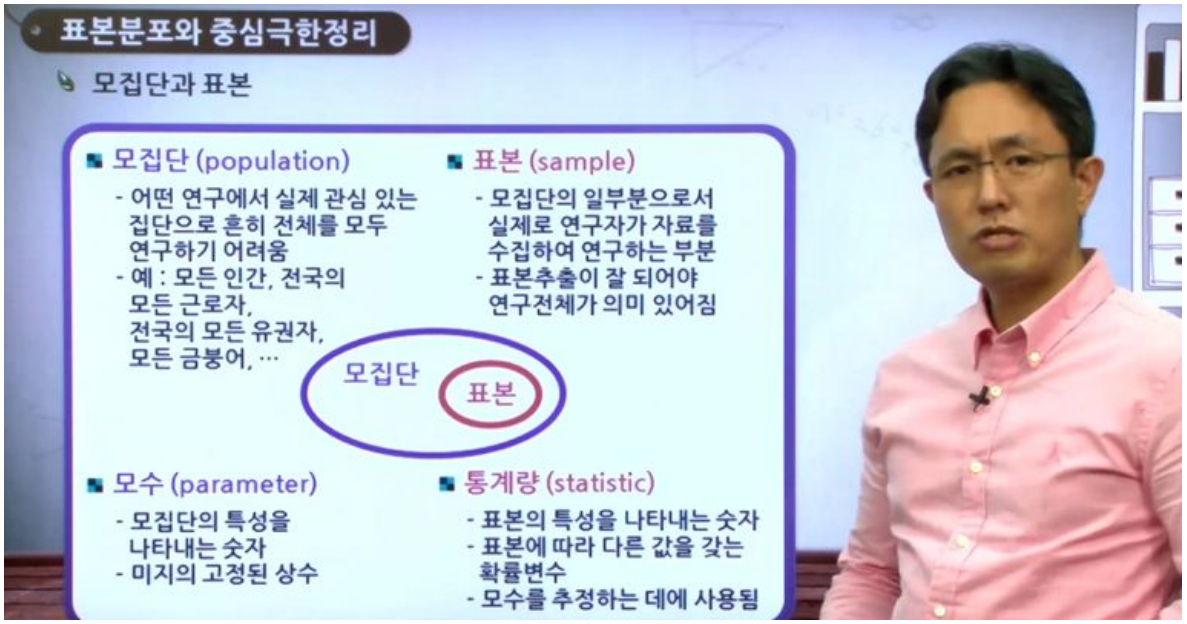


학습 정리

팀	슈퍼빅데이터	구성원	최재림, 황성욱

일정	발제자	주제
5일차 (5 / 31)	황성욱	- 표본분포와 중심극한정리

주요 내용 요약



1. 모집단과 표본
  모집단 : 정보를 얻고자 하는 관심 대상의 전체 집합
  표본 : 모집단의 부분집합
2. 표본분포 = 표본에서 도출되는 통계량에 대한 확률분포
3. 중심극한정리 = 표본의 갯수(N)이 충분하다면 모수를 모르는 상황에서도 표본 통계량으로 정규분포를 구성하여 모수를 추정할 수 있다는 것