데이터분석의 정의

크게 예측, 분류, 시각화

그 중에서 제일 중요한것 예측.

데이터

정형/비정형/반정형

분석과정에는 통계가 포함되어있다.

예측을 잘하면 올바른 의사결정을 할 수 있다.

그러면 경쟁우위가 생긴다.

경쟁우위는 생존에 필요하다.

측우기 만들어진 것 과학적인 이유는

현재랑 직경이 굉장히 유사하다

기업경영에서의 데이터 분석

Ex) Value chain

PPT 2쪽 참고

ERP에 들어가는 4개

8.MKT/Sales, 6.Operation, Financial Accounting, 2.HRM

CRM에 들어가는 것

8,9번

가설

Ex) 그럴거야~ 부부관계가 좋으면 이혼을 안할거야~

모집단이란 전체

귀무가설

대립가설

Ex)칼로커트

귀무가설 : 칼로커트는 체중감소에 효과가 없을 것이다.

대립가설 : 칼로커트는 통계적으로 유의한 체중감량 효과가 있을 것이다.

* 통계기반 검증할 때는 대부분 대립가설
* 대립가설 = 연구가설
* 통계적으로 유의하다 -> 95% 신뢰수준에서 p값이 0.05 미만일 경우 (p<0.05)

채택 / 기각

가설 세웠는데 맞으면 채택, 아니면 기각

1종오류 ex) 정상인데 불량으로 낸 경우

2종오류 ex) 불량인데 정상으로 낸 경우

통계적으로 유의하다 -> 95% 신뢰수준에서 p값이 0.05 미만일 경우 (p<0.05)

가설 검정 vs 가설 검증

대부분 검정으로 사용한다.

Ex) 수학과외를 했는데 비싸게 돈 주고 한 과외를 듣고 수학점수가 높게 나왔을 때

사후의 결과값이 더 높게 나왔으므로 t값이 –이다?

사후의 결과에 따라 t값이 -t인지 +t인지 다르다.

t값이 –t값

t-검정: 쌍체 비교

영화 나는 전설이다.

csv파일

평균선에서 멀어진 거리

평균

분산

관측수

피어슨 상관 계수

가설 평균차

자유도

t통계량

P(T<=t)단측 검정

T기각치 단측 검정

P(T<=t)양측 검정

T기각치 양측 검정