20231104 DAY05

1. 09:30~10:20 : 통계이론

* <http://www.kocw.net/home/cview.do?cid=9d2e1e8ac755bc24>
* 1장 통계기초
* 2장 자료의 정리 및 표현

1. 10:30~11:20 : 통계이론

* 3장 분포의특성
  + 모집단 분산 : 모든 데이터 샘플 사용 (N)
  + 표본 분산 : 모집단 중 일부를 샘플로 사용 (N-1)
* Z-score : 상대 위치
* 왜도(skewness) : 좌우 치우친 정도, 정규분포 = 0
* 첨도(kurtosis) : 위아래 퍼진 정도, 정규분포 = 1

1. 11:30~12:20 : 확률

* 4장, 5장 확률이론
* 사상(event) : 사건, 예: 동전 던지기
* 확률변수 : 확률로 표시할 수 있는 변수 (주사위, 숫자3, 확률 1/6)
* 조건확률 : P(A|B) : B 가 발생하였다는 조건 하에서 A가 발생할 확률
* 기대값 : 가중평균
* 공분산(Covariance) : 기울기, x가 증가할 때 y의 경향성
* 상관계수 : 공분산의 범위를 -1 ~ 1 사이로 표현

1. 12:30~13:20 : 확률

* 6장 확률분포
* 베르누이 시행 : 동전 던지기, 0 or 1
* 이항분포 : 베르누이 시행을 1번 이상 반복한 경우 시행결과의 합
* 푸아송분포(pmf) : 일정한 단위시간, 단위공간, 단위거리, 단위면적 등에서 발생하는 관심 있는 사건의 발생 가능한 횟수를 변수 값으로 하는 확률변수의 분포
  + 어떤 단위시간에서 발생하는 평균수는 이 단위시간의 크기에 비례한다.
  + 단위시간에서 발생하는 발생횟수는 단위시간 간에 독립적이다.
  + 충분히 작은 단위시간에서 둘 이상의 사건이 동시에 발생할 확률은 0이다.
  + 동일한 간격에서 발생하는 사건의 확률은 동일하다.
* 초기하분포 : 비복원 추출 : 로또 번호
* 7장 연속확률분포
* 정규분포(pdf) : 음의 무한대에서 양의 무한대사이의 연속활류분포, 평균과 표준편차 2개의 모수에 의해 특징지어지는 분포
  + 평균을 중심으로 좌우대칭, 평균=중앙치=최빈치

1. 14:00~14:50 : 확률

* 표준정규분포 : X를 Z-Score값으로 표준화한 분포, 평균 = 0, 표준편차 = 1
* 정규분포표를 이용해 확률을 계산
* 지점에서의 확률은 0 -> 즉 초과, 이상의 확률은 동일함.
* 지수분포
* 8장 표본분포
  + 표본추출오차 : 모집단을 대표할 수 있는 전형적인 구성요소를 표본으로 선택하지 못할 경우

1. 15:00~15:50 : 확률

* 학습: ~8장 표본분포까지

1. 16:00~16:50 : 통계

* 9장 통계적 추정1: 한 모집단

1. 17:00~17:50 : 통계

* 9장 통계적 추정2: 한 모집단

20231105 DAY06

1. 09:30~10:20 : 통계

* 10장 가설검정

1. 10:30~11:20 : 통계이론

* [과정 25: 통계 검정의 거듭제곱 (psu.edu)](https://online.stat.psu.edu/stat415/book/export/html/845)

1. 11:30~12:20 : 교재 378
2. 12:30~13:20 : 분산분석 ANOVA

* 모집단이 3개 이상인 경우,
  + H0 :
  + H1 : 최소 둘중의 하나는 다르다. -> 3번 반복하면 오차가 증가함 (0.95^3 = 0.85)
* ANOVA : ANalysis of VAriance
* 분산분석의 기본 가정
  + 정규성
  + 독립성
* F 통계량 : F = 집단간 분산 / 집단내 분산
* 분산분석의 구분
  + 일원배치 : 요인이 하나인 경우 (예: 타이어 성능 : 온도(-20/0/20/40))
  + 이원배치 : 요인이 둘인 경우 (예: 온도&습도 (-20&25 / -20 &50 / …))
  + 다원배치 : 요인이 셋 이상인 경우 (예: 온도&습도&비 (-20&25&O / -20&25&X / …))
* 사후비교 방법 (있다 정도만 보고 넘어감)
  + Least Square Difference method
  + Benfoerronis
  + Scheff…
  + ???
* Scipy.stats : [Statistics (scipy.stats) — SciPy v1.11.3 Manual](https://docs.scipy.org/doc/scipy/tutorial/stats.html)

1. 14:00~14:50 : Scipy.stats

* 정규성 검정 : Shapiro-Wilk Test
* 상관관계 검정 : Pearson’s Correlation
* ANOVA : kruskal()

1. 15:00~15:50 : 실습, pp.388-407

* Percentile vs quantile
  + Percentile : 0 ~ 100 범위
  + Quantile : 0 ~ 1 범위

1. 16:00~16:50 : 실습



1. 17:00~17:50 : 실습

* before (A) | after (B)
  + H0: D (A-B) = 0
  + H1: D (A-B) > 0
* Difference = before – after
* Ttest\_rel(A, B, alternative=’greater’]
* Difference = after – before
* Ttest\_rel(df[‘after’,df[before’], alternative=’greater’]