

AI기초통계

담당교수 : 김 승 재

< 공지 사항 >

- 개강 : 9월 01일(수)
- ‘코로나19로 **대면-비대면** 교차수업 진행
<1주차~7주차: **대면**, 8~15주차: **비대면**>
 - 9주차 : 10월 27일(수) ~ 11월 02일(화)
 - 10주차 : 11월 03일(수) ~ 11월 09일(화)
- **대면-비대면** 수업 참여 방법
 - 대면 : 지정된 강의실 입실(시간준수)
 - 비대면 : 호남대학교 **LMS 시스템** 이용하여 참여
- 수업 및 **대학생활**에 필요한 사항은 별도 **공지**

대면 강의 내용(9주차)

○ 9주차 강의내용

● 9-1교시

- T-검정
- 일변량(단일표본) 분석해보기(1)

● 9-2교시

- T-검정
- 일변량(단일표본) 분석해보기(2)

● 9-3교시

- T-검정
- 이변량(독립표본) 분석해보기

단일표본 t 검정

• 단일표본 t 검정(One sample t-test)

• ‘단일표본 t검정’은 ‘일변량 t검정’이라고도 함

- ‘일변량’이란 하나의 변수를 의미하며, 하나의 변수는 하나의 정보만 가짐

• 검증방법

- 모집단의 분산 또는 표준편차를 알지 못하는 상황에서 모집단을 대표하는 표본의 평균과 연구자가 이론적 또는 경험적 지식에 근거하여 설정된 특정한 평균값을 기준으로 비교하며 검증하는 방법

• 단일표본 t-검정은 아래와 같은 상황에서 진행되는 가설검정 방식으로..

- 한 집단의 평균을 검정(test)하고 싶은 경우
- 해당 집단의 (모)평균은 알고 있지만, (모)분산을 모르고 있다고 가정할 경우
 - 귀무가설(H_0) : 표본평균과 모집단 평균은 같다(차이가 없음).
 - 대립가설(H_1) : 표본평균과 모집단 평균은 같지 않다(차이가 있음).

• 또한, 단일표본 t-검정은 정규분포를 가정하기 때문에, 표본크기가 30개 이상 있을 때를 전제하고 있다(중심극한정리).

• William에 의해 발명 되었으며, 필명이 student 라서 student t검정이라 함.

단일표본 t 검정

• 단일표본 t 검정(One sample t-test)

• 표준오차(Standard Error of the Mean: SE)

- ‘단일표본 t검정’을 이해하기 위해 반드시 알고 있어야 할 개념
- 표집분포의 표준편차가 **표준오차**이다.

• 표준오차의 해석

- 표본평균의 분산에 **루트**를 씌워준 것
- 의미상으로 표준오차를 접근할 때는 **표본평균의 분산**의 의미로 생각하면 됨
- **표본평균의 분산**: 표본평균들이 표본평균들의 평균으로부터 **얼마나 떨어져 있는지 알려주는 척도**

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

\bar{x} = 표본 평균

μ_0 = 영가설하에서 모집단 평균

s = 표본의 표준편차

n = 표본의 크기

단일표본 t 검정

• 단일표본 t 검정(One sample t-test)

• Z 분포와 T분포의 차이

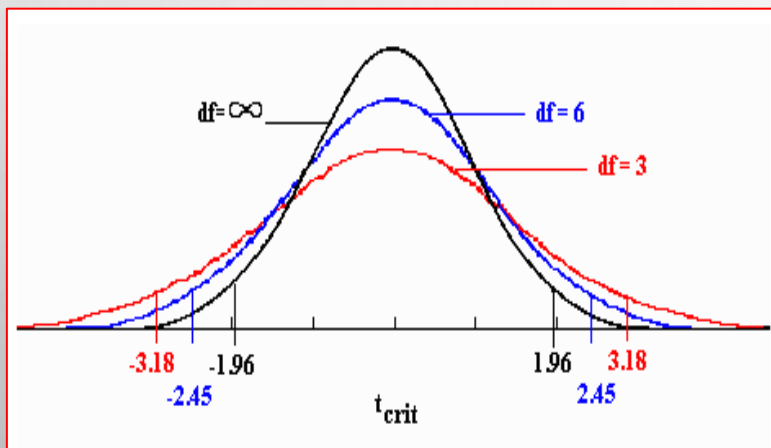
- Z분포는 '표준정규분포'로 나타내지만, T분포는 z분포에 자유도(df)를 고려함.
 - 자유도(degree of freedom, df)

• 양측검정을 해야하는 경우라면...

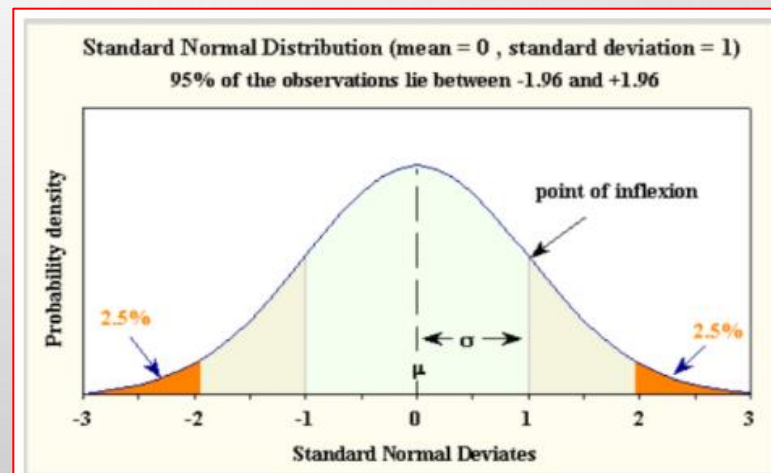
- 해당 분포 양측 5%에 해당하는 critical value(임계값)을 설정해줘야 함

• 주의해야 할 점...

- T분포는 자유도에 따라 양측 5%에 해당하는 critical value 값들이 다르다는 것을 유의해야 함(반드시 자유도에 따라 값을 확인해야 함).



이미지 출처: <https://enook.jbnu.ac.kr/contents/44/#!/p/17>



이미지 출처: <https://bioinformaticsandme.tistory.com/186>

단일표본 t 검정

• 단일표본 t 검정(One sample t-test)

- 예1) '단일표본 t 검정'의 가설설정 및 유의수준과 유의확률 비교하기..
- 가설 설정
 - 전국 대학교 대학생 평균 키 = 178.5
 - 우리 대학교 대학생 평균 키 = 179.9
 - 귀무가설:
 - 전국 대학교의 대학생 평균 키와 우리 대학교 대학생 평균 키는 같다고 할 수 있다.
 - 대립가설:
 - 전국 대학교의 대학생 평균 키와 우리 대학교 대학생 평균 키는 다르다고 할 수 있다.
 - 우리 대학교 대학생 키 표준편차(s) = 표본표준편차(s) = 7.05
 - 표본크기 (n) = 101
 - 자유도 = $n-1 = 100$
 - 임계값 = 1.984
 - T분포의 임계값은 T-table을 참고하여 해석함
 - 그림에서...95% 신뢰수준, 자유도 100, 양측검정일 경우 임계값은 1.984

단일표본 t 검정

• 단일표본 t 검정(One sample t-test)

• 임계값 = 1.984

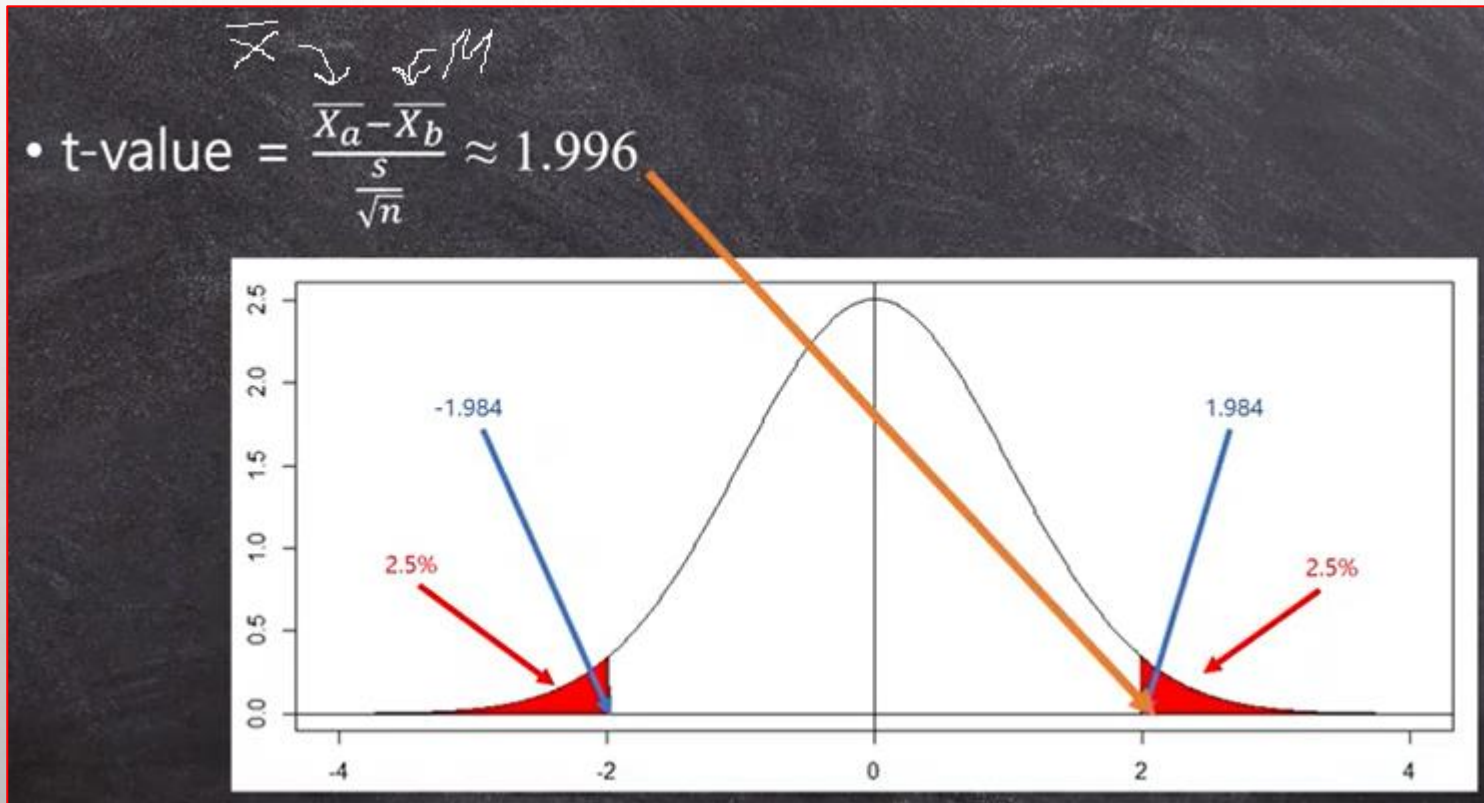
- T분포의 임계값은 T-table을 참고하여 해석함
- 그림에서...95% 신뢰수준, 자유도 100, 양측검정일 경우 임계값은 1.984

cum. probl	95% 신뢰수준										
	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.998}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
om	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%

단일표본 t 검정

• 단일표본 t 검정(One sample t-test)

- T-value는 기각역에 포함하므로 대립가설이 채택된다.
- 즉, T검정 결과...
 - '전국 대학교의 대학생 평균 키와 우리 대학교 대학생 평균 키는 다르다'고 할 수 있다.



<9-1차시, 자기주도적 학습정리>

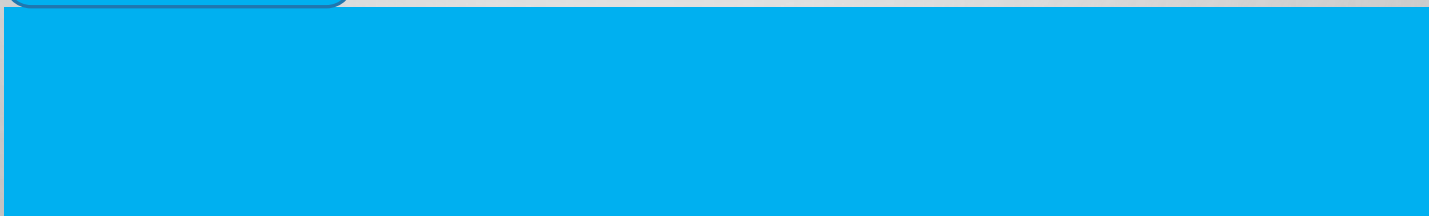
단일표본 t검정



Z분포와 T분포



양측검정



대면 강의 내용(9주차)

○ 9주차 강의내용

• 9-2교시

- T-검정
- 일변량(단일표본) 분석해보기(2)

• 9-3교시

- T-검정
- 이변량(독립표본) 분석해보기

‘코로나19’로 **대면-비대면** 강의 진행

수고하셨습니다