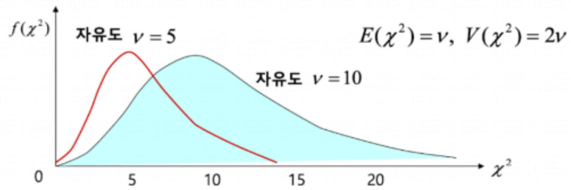
카이제곱 분포와 t 분포

1. 카이제곱 분포

   표준 정규 분포를 따르는 확률 변수 z를 제곱한 것을 r개 더하면 새로운 확률 변수가 됨

   새로운 확률변수의 분포를 자유도가 r인 카이제곱 분포라고 함



- 사용하는 이유 - 데이터의 분산을 기반 → 모집단의 분산과 표본 분산을 비교하거나 또는 분산의 어떤 특정 값인지를 점검하는데 用

카이제곱분포 x^2

- 영국의 수리통계학자 칼 피어슨에 의해 개발된 분포
- 모분산에 대한 가설 검정에 사용
- 교차분석에서 독립성 검정, 동일성 검정 등에 사용
  - 두 범주형 변수 간의 연관성 검정에 사용
  - 모집단이 정규분포인 대표본에서 모분산이 특정한 값은 갖는지 여부 검정
- 심슨의 패러독스
  - 에드워드 심슨의 이름을 따서 만든 용어
  - 교차표를 분석할 경우 각각의 부분자료에 성립하는 대소관계가 그 부분의 자료를 종합한 저네 자료에 대해서는 성립하지 않는 경우가 있음

성별	지원자 수	합격률
남성	8442	44%
여성	4321	35%

학부	남성		여성	
	지원자 수	합격률	지원자 수	합격률
A	825	62%	108	82%
B	325	37%	593	34%
C	417	33%	375	35%
D	373	6%	341	7%
E	560	63%	25	68%
F	191	28%	393	24%

- 카이제곱분포의 특징
  1. 연속형 확률분포로서 확률 변수는 항상 + 값을 가짐

2. 오른쪽 꼬리를 가진 비대칭 분포

3. 자유도에 따라 모양 결정

4. 확률변수 X가 자유도 v인 카이제곱 분포를 따를 경우 $X \sim \chi^2(v)$

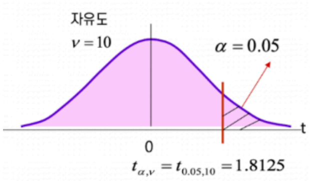
5. $X \sim \chi^2(n)$ 일 때, $\mathrm{E(X) = n, Var(X) = 2n}$이다.

2. 스튜던트 t-분포

- 1908년 추측통계학자의 창시자인 고셋이 영국의 수리통계학 Biometrica에 Student 라는 익명으로 발표
- 소표본에 적용
- 특징
  1. 표준정규분포와 같이 평균이 0, 평균을 중심으로 좌우대칭인 종형 모양
  2. 표준정규분포보다 봉우리가 완만
  3. 자유도에 따라 분포형태가 다름
  4. 표본의 수가 30개 미만인 소표본을 사용해서도 모수의 추측이 가능
  5. 모집단의 표준편차를 모를 경우 모집단 평균에 대한 검정과 추정에 적용

6. $X \sim t(n)$ 일 때, $\mathrm{E(X) = 0, Var(X)} = \frac{n}{n-2}$이다.

$$P(t \ge t_{\alpha,\nu}) = \alpha$$



< t 분포표(부분)> $P(t \ge t_{\alpha,\nu}) = \alpha$

자유도 ν	확률 α 0.05	0.025
9	1.8331	2.2622
10	1.8125	2.2281
11	1.7959	2.2010
12	1.7823	2.1788
13	1.7709	2.1604

이산형 확률 분포

- 베르누이 시행

- 두 가지 결과만을 가지는 단순한 확률 실험을 의미
- 성공확률 p , 실패 확률 1-p

- 베르누이 분포

$$X = \begin{cases} 1 & (\text{성공}) \\ 0 & (\text{실패}) \end{cases} \quad \text{with} \quad P(X=1) = p, \quad P(X=0) = 1-p$$

이때 X는 **베르누이 분포**를 따른다고 하며, 표기법은:

$$X \sim \text{Bernoulli}(p)$$

- 이항분포
  - 성공 확률이 p인 베르누이 시행을 서로 독립적으로 n번 반복 수행할 경우, 전체 성공 횟수를 나타내는 확률변수 X는 이항분포를 따름
  - 수식

수식 : $X = \sum_{i=1}^{n} Z_i \quad \text{where} \quad Z_i \sim \text{Bernoulli}(p)$ 또는 $X \sim \text{Binomial}(n,p)$ 또는 $X \sim B(n,p)$

- 이항분포의 확률질량함수 (PMF)

수식 : $X = \sum_{i=1}^{n} Z_i \quad \text{where} \quad Z_i \sim \text{Bernoulli}(p)$ 또는 $X \sim \text{Binomial}(n,p)$ 또는 $X \sim B(n,p)$

- 이항분포의 기댓값과 분산

  - 기대값

$$\mathbb{E}[X] = np$$

  - 분산:

$$\text{Var}(X) = np(1-p)$$