

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설 설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

단일 표본 T-Test

하나의 모집단 평균이 **이전보다** 커졌는지/작아졌는지/달라졌는지를 통계적으로 알아보기 위해 사용
기준에 알려져 있던 사실이 현재는 어떻게 변화 했는지를 확인하는 방법

#01. 작업준비

패키지 참조

`scipy` 패키지의 설치가 필요하다.

```
from scipy import stats
from pandas import read_excel
```

#02. 예제(1) - 강아지 체온 변화

주어진 데이터는 강아지 25마리의 체온을 측정한 데이터이다. 과거의 자료에서 강아지의 평균 체온은 24.3알려져 있다. 이번에 조사한 데이터에서도 강아지의 평균 체온이 24.3이라고 해도 좋은지 알고 싶다.

데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E02/temperature.xlsx")
df
```

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설 설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

	체온
0	25.8
1	24.6
2	26.1
3	22.9
4	25.1
5	27.3
6	24.0
7	24.5
8	23.9
9	26.2
10	24.3
11	24.6
12	23.3
13	25.5
14	28.1
15	24.8
16	23.5
17	26.3

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

	체온
18	25.4
19	25.5
20	23.9
21	27.0
22	24.8
23	22.9
24	25.4

case 1

강아지의 평균 체온이 이전에 알려져 있던 24.3 과 달려졌는지 여부

가설설정

강아지의 평균 체온 : μ 귀무가설(H_0) : $\mu = 24.3$ 대립가설(H_1) : $\mu \neq 24.3$

```
result = stats.ttest_1samp(df['체온'], 24.3)
result
```

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설 설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

```
TtestResult(statistic=2.7127692953052716, pvalue=0.012145537241648427, c
```

p-value의 기준

- $P\text{-value} < 0.01$: 귀무가설이 옳을 확률이 0.01 이하 → 틀렸다(깐깐한 기준) --> 대립가설 채택
- $P\text{-value} < 0.05$: 귀무가설이 옳을 확률이 0.05 이하 → 틀렸다(일반적인 기준) --> 대립가설 채택
- $0.05 < P\text{-value} < 0.1$: 애매한 경우(샘플링을 다시한다)
- $0.1 < P\text{-value}$: 귀무가설이 옳을 확률이 0.1 이상 → 틀리지 않았다(맞다와 다름)

결과 해석

출력에서 일표본 t검정에 대한 유의확률값이 0.0121 로 통상적인 유의수준 0.05보다 작으므로 유의수준 0.05 에서 강아지 25마라의 체온이 추출된 모집단의 평균인 24.3 이라고 할 수 없다.

case 2

강아지의 평균 체온이 과거의 자료인 24.3보다 작아졌는지 알고자 하는 경우

가설 설정

귀무가설(H_0) : $\mu = 24.3$

대립가설(H_1) : $\mu < 24.3$

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

```
result = stats.ttest_1samp(df['체온'], 24.3, alternative='less')
result
```

```
TtestResult(statistic=2.7127692953052716, pvalue=0.9939272313791758, df=
```

결과 해석

p-value가 0.05 보다 크므로 귀무가설을 기각할 수 없다. 즉, 강아지의 평균 체온은 24.3 과 같다(작지 않다).

case 3

강아지의 평균 체온이 과거의 자료인 24.3 보다 커졌는지 알고자 하는 경우

가설 설정

귀무가설(H_0): $\mu = 24.3$

대립가설(H_1): $\mu > 24.3$

```
t, p = stats.ttest_1samp(df['체온'], 24.3, alternative='greater')
"t-statistic: {:.3f}, p-value: {:.3f}".format(t, p)
```

```
't-statistic: 2.713, p-value: 0.006'
```

결과 해석

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

p-value가 0.05보다 작으므로 귀무가설을 기각하고 대립가설을 채택한다. 즉, 강아지의 평균 체온은 24.3보다 크다

참고: t-statistic (T값)

계산된 차이를 표준 오차 단위로 나타낸 것으로, T의 크기가 클수록 귀무 가설에 대한 증거가 큼니다. 즉, 유의한 차이가 있다는 증거가 더 명확한 것입니다. 반면 T가 0에 가까울수록 유의미한 차이가 없을 가능성이 커집니다.

#03. 예제(2)

주어진 데이터는 3년에 한 번씩 생명보험협회에서 실시하는 생명보험 전국 성향조사의 일부이다.

이 금액의 모평균이 1억원이라고 볼 수 있는지를 유의수준 5%에서 검정하여라

각 변수의 의미

구분	설명
직업	1 육체노동, 2 정신노동(직장인), 3 자영업
교육수준	1 초등학교 이하, 2 중졸 및 고졸, 3 대졸, 4 대학원 졸
성별	1 남자, 2 여자
가입금액	단위=천만원
월수입	단위=만원

데이터 가져오기

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설 설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

```
df = read_excel("https://data.hossam.kr/E02/insur.xlsx")
df
```

	직업	교육수준	성별	나이	가입금액	월수입
0	1	2	1	35	15.0	100
1	2	3	1	40	10.0	150
2	2	3	2	36	8.0	120
3	3	4	2	38	12.0	140
4	1	1	1	45	10.0	150
5	2	2	1	48	2.5	200
6	3	3	2	50	9.0	90
7	3	3	1	53	7.5	180
8	2	2	1	49	5.5	250
9	3	4	1	55	25.0	300
10	2	3	2	52	10.5	220
11	1	1	1	35	3.5	80
12	1	2	1	46	9.7	175
13	2	4	2	41	12.5	280
14	3	4	1	59	30.0	550

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설 설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

	직업	교육수준	성별	나이	가입금액	월수입
15	2	3	1	55	11.0	235
16	1	2	2	43	8.8	110
17	3	1	2	39	4.5	125
18	2	2	1	40	7.8	160
19	1	2	1	29	6.7	99
20	2	3	1	33	7.0	148
21	2	4	1	45	33.0	400
22	3	2	2	32	15.0	115
23	2	2	2	44	20.0	165
24	1	2	1	28	4.0	80
25	2	1	1	38	5.0	210
26	3	3	1	51	15.0	340
27	2	3	2	43	30.0	510
28	1	2	1	42	5.0	130
29	3	3	1	56	10.0	650

case 1

가입금액의 평균이 1억원과 같은지, 다른지 여부

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

가설설정

가입금액: μ

$$H_0 : \mu = 10$$

$$H_1 : \mu \neq 10$$

```
t, p = stats.ttest_1samp(df['가입금액'], 10)
"t-statistic: {:.3f}, p-value: {:.3f}".format(t, p)
```

```
't-statistic: 1.201, p-value: 0.239'
```

결과 해석

단일 표본 검정에서 유의확률이 유의수준 0.05보다 크므로, 보험 가입금액의 모평균이 1억원이라는 가설을 기각할 수 없다.

case 2

가입금액: μ

$$H_0 : \mu < 10$$

$$H_1 : \mu \geq 10$$

```
t, p = stats.ttest_1samp(df['가입금액'], 10, alternative='less')
"t-statistic: {:.3f}, p-value: {:.3f}".format(t, p)
```

단일 표본 T-Test

#01. 작업준비

패키지 참조

#02. 예제(1) - 강아지 체온 변화

데이터 가져오기

case 1

가설 설정

p-value의 기준

결과 해석

case 2

가설 설정

결과 해석

case 3

가설 설정

결과 해석

참고: t-statistic (T값)

#03. 예제(2)

각 변수의 의미

데이터 가져오기

't-statistic: 1.201, p-value: 0.880'

case 3

가입금액 : μ $H_0 : \mu > 10$ $H_1 : \mu \leq 10$

```
t, p = stats.ttest_1samp(df['가입금액'], 10, alternative='greater')
"t-statistic: {:.3f}, p-value: {:.3f}".format(t, p)
```

't-statistic: 1.201, p-value: 0.120'