Inference Testing I

- Hypothesis 검정

  Hypothesis is a statement about a population parameter

  가설

  - Null hypothesis (H_0) 귀무가설 =
  - Alternative hypothesis (H_1) 대립가설 $>$ $<$ $\neq$

- Testing 검정 통계량

  Is a rule for which sample values the decision is made to accept H_0 as true

한양대학교 HANYANG UNIVERSITY

Inference Testing I

- Accept or reject

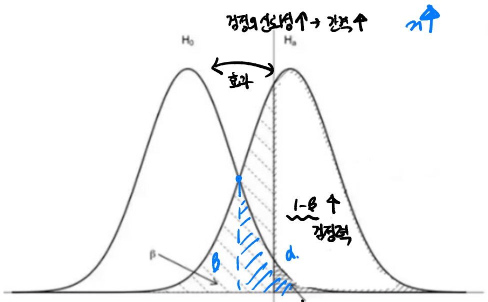
  - Rejection region : The area of the distribution where H_0 can be rejected
  - Acceptance region : The area of the distribution where H_0 can be accepted

한양대학교 HANYANG UNIVERSITY

Inference Testing I

- Statistical errors
  - Type I error (α) : also called the significance level, false positive
  - Type II error (β): false negative

검정의 신뢰성↑ → 간격↑ 기↑



α= 0.05, 0.01, 0.1 → 일반적

유의수준 ~ (1-α)%

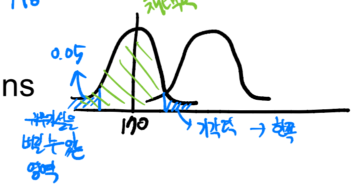
결정	H0 true	H0 False
H0 reject	X → α	O → 옳은 결정.
H0 acc	O	β X → 둘째 오류

(제 1종 오류
  " 2종 오류

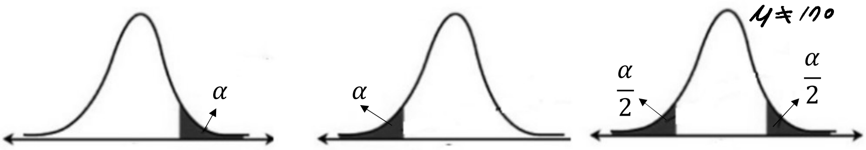
Inference Testing I

- Testing types
  - One-sided testing : left or right tailed rejection region
  - Two-sided testing : left and right tailed rejection regions

H0 $\mu=170$
H1 $\mu>170$



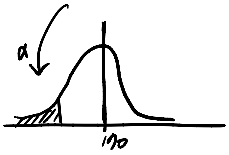
$\mu>170$ $\mu<170$ $\mu\neq170$



One sided testing

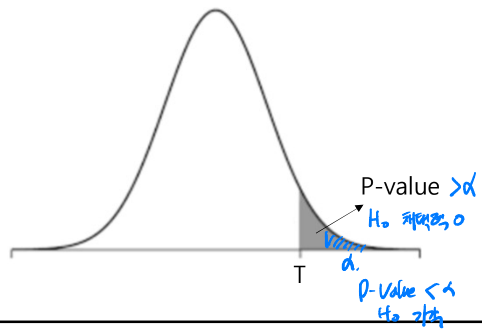
Two sided testing

$\mu<170$



Inference Testing I

- Test statistic
  - Test statistic : A statistic from a sample to make a decision about hypothesis on inference testing
- P-value



Inference Testing I

- Testing about μ
  - With large sample

$$H_0\!: \mu = \mu_0$$

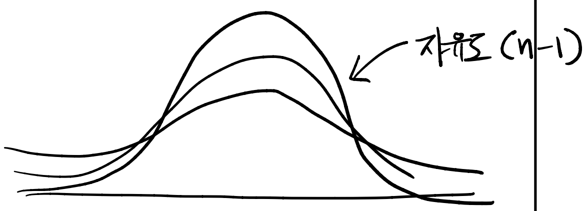
$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim N(0,1)$$

- Rejection region:

$$H_1\!: \mu > \mu_0, \qquad T \geq Z_\alpha$$
$$H_1\!: \mu < \mu_0, \qquad T \leq -Z_\alpha$$
$$H_0\!: \mu \neq \mu_0, \qquad |T| \geq Z_{\frac{\alpha}{2}}$$

Inference Testing I

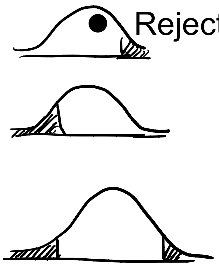
- Testing about μ 평균
  - With small sample



$$H_0 \colon \mu = \mu_0$$

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{(n-1)}$$

(0, 1)



- Rejection region:

기각역.

$H_1 \colon \mu > \mu_0, \qquad T \geq t_{(n-1,\alpha)}$

$H_1 \colon \mu < \mu_0, \qquad T \leq -t_{(n-1,\alpha)}$

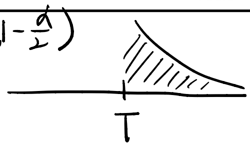
$H_0 \colon \mu \neq \mu_0, \qquad |T| \geq t_{(n-1,\frac{\alpha}{2})}$

(평균 분산 비율

$H_0 : \mu = \mu_0$ / $H_1 :$ $>$ $<$ $\neq$

$\bar{x} \sim (\mu, \frac{\sigma^2}{n})$ np.abs(T)

norm.ppf$(1 - \frac{\alpha}{2})$



→ norm.cdf(?) 찾을 수 0

앞에 하나를 더 넣어야

Inference Testing I

- Testing about p

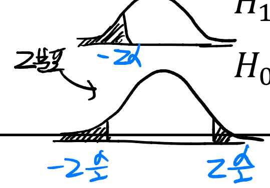
Sample 비율

$$H_0 \colon p = p_0$$

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0.(1-p_0)}{n}}} \sim N(0,1)$$

$$\hat{p} \pm Z \boxed{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

- Rejection region:



$H_1 \colon p > p_0, \qquad T \geq Z_\alpha$

$H_1 \colon p < p_0, \qquad T \leq -Z_\alpha$

$H_0 \colon p \neq p_0, \qquad |T| \geq Z_{\frac{\alpha}{2}}$

Inference Testing I

- Testing about σ^2 분산

$$H_0: \sigma^2 = \sigma_0^2$$

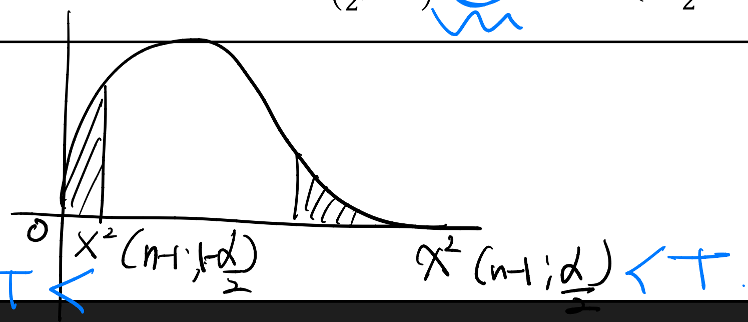
$$T = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{(n-1)} \text{ 카이제곱 분포.}$$

- Rejection region:

$$H_1: \sigma^2 > \sigma_0^2, \qquad T \geq \chi^2_{(\alpha, n-1)}$$

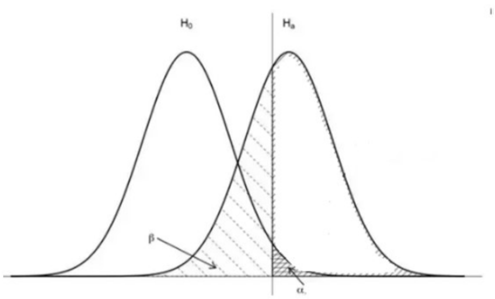
$$H_1: \sigma^2 < \sigma_0^2, \qquad T \leq \chi^2_{(1-\alpha, n-1)}$$

$$H_1: \sigma^2 \neq \sigma_0^2, \qquad T \geq \chi^2_{(\frac{\alpha}{2}, n-1)} \text{ or } T \leq \chi^2_{(1-\frac{\alpha}{2}, n-1)}$$



Inference Testing I

- Practice
  - Calculate the pooled estimator of variance with A, B groups.



Inference Testing I

```
In [7]: import numpy as np
        from scipy.stats import ttest_1samp, chi2, norm
        from statsmodels.stats.proportion import proportions_ztest
```

.ttest_1samp
-. *T-test with one sample*
.proportions_ztest
-. *Z-test for proportions*

Inference Testing I

in data

.ttest_1samp(*x, popmean, alternative*)
-. *x : array like samples*
-. *popmean: the mean of H0* $= \mu_0$
-. *alternative: 'two-sided' or 'less' or 'greater'*

$>, <, \neq$

Inference Testing I

성공횟수 Sample 수

.proportions_ztest(*count*, *nobs*, *value*, *alternative*)

-. *count : number of successes*

-. *nobs: number of trials*

-. *value: the value of H0 equal to the proportion*

-. *alternative: 'two-sided' or 'smaller' or 'larger'*

한양대학교 HANYANG UNIVERSITY

Inference Testing I

- Practice
  - With 'can' data, perform two-sided test whether H_0: $\mu = 100$ or not at $\alpha = 0.05$

T= ttest_1samp(   ) → return (T값, p-value ⇒ 결정 (기각, 채택

α, p-value 비교

↳ t분포

Sample 수 ↑ → z분포

한양대학교 HANYANG UNIVERSITY

Inference Testing I

- Solution

```
In [22]: def z_stat(x, popmean, alpha, alternative):
             # alternative ='two-sided' or 'less' or 'greater'
             z_val = (np.mean(x)-popmean)/(np.std(x, ddof=1)/np.sqrt(len(x)))
             print('Test statistic is {}'.format(np.round(z_val, 4)))

             if alternative =='two-sided':
                 prnt('reject H0') if np.abs(z_val)> norm.ppf(1-alpha/2) else print('accept H0')
             elif alternative == 'less':
                 print('reject H0') if z_val < norm.ppf(alpha) else print('accept H0')
             else:
                 print('reject H0') if z_val > norm.ppf(1-alpha) else print('accept H0')


         z_stat(can, 100, 0.05, 'two-sided')

         Test statistic is 0.2352
         accept H0
```

Inference Testing I

- Practice
  - Use only the first 20 observations in 'can' data, and test H_0: $\mu = 100$ vs. H_1: $\mu > 100$ at $\alpha = 0.05$

```
In [26]: can_small = can[0:20]
         print(can_small)

         [101.8 101.5 102.6 101.  101.8  96.8 102.4 100.   98.8  98.1  98.8  98.
           99.4  95.5 100.1 100.5  97.4 100.2 101.4  98.7]
```

Inference Testing I

- Solution → 소을정

```
In [36]: stat, pval = ttest_1samp(can_small, popmean = 100, alternative = 'greater')
         print('Test statistic is {}'.format(np.round(stat,4)))
         print('p-value is {}'.format(np.round(pval,4)))
         print('H0 is rejected') if pval < 0.05 else print('H0 is accepted')

Test statistic is -0.5934
p-value is 0.72
H0 is accepted
```

Inference Testing I

- Practice
  - A company found that failure rate of A product is 15%, and it could be improved if they changed the material of the product. They produced the A product with new material, and found 54 failures out of 400 pieces of the product. Then do you think the failure rate decreased at $\alpha = 0.05$.

$H_0: p = 0.15$ $H_1: p < 0.15$

$\hat{p} = 54/400$

count nobs.

Inference Testing I

- Solution

```
In [37]: stat, pval = proportions_ztest(54, 400, value = 0.15, alternative = 'smaller')

         print('Test statistic is {}'.format(np.round(stat,4)))
         print('p-value is {}'.format(np.round(pval,4)))
         print('H0 is rejected') if pval < 0.05 else print('H0 is accepted')

Test statistic is -0.8779
p-value is 0.19
H0 is accepted
```


Inference Testing I

- Practice
  - Using ‘can’ data, and test $H_0: \sigma^2 = 2.1$ vs. $H_1: \sigma^2 > 2.1$ at $\alpha = 0.1$

오른쪽 기각역

Inference Testing I

- Solution

```
In [39]: s20 =2.1
n = len(can)
s2 = np.var(can, ddof=1)
stat = (n-1)*s2/s20
pval = 1-chi2.cdf(stat,n-1)

print('Test statistic is {}'.format(np.round(stat,4)))
print('p-value is {}'.format(np.round(pval,4)))
print('H0 is rejected') if pval < 0.05 else print('H0 is accepted')
```

```
Test statistic is 87.0152
p-value is 0.2515
H0 is accepted
```