Estimation

- Estimation
  - Estimator

    : the function of X to estimate the parameters

  - Estimate

    : the computed value from the estimator

추정 량 틀

추정 치 붕어빵

한양대학교 HANYANG UNIVERSITY

Estimation

- Estimation
  - Point estimation 점추정

    : Calculate the value of “the best guess” for the parameters

  - Interval estimation 구간추정

    : Calculate the interval which probably contain the parameters

한양대학교 HANYANG UNIVERSITY

Estimation

- Mean
  - Point estimator about μ

    $X_1, X_2, \dots, X_n$ from the population whose mean is μ and the variance is σ^2

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

	모수	
평균	모평균 ←	표본 평균($\bar{X}$)
분산	모분산 ←	〃 분산(σ^2)
비율	모비율 ←	〃 비율($\hat{p}$)

Estimation

- Mean
  - (1-α)% confidence Intervals (with large sample)

$$\left(\bar{X} - Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \quad \bar{X} + Z_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right)$$

  - (1-α)% confidence Intervals (with small sample)

$$\left(\bar{X} - t_{(\frac{\alpha}{2}, n-1)} \cdot \frac{s}{\sqrt{n}}, \quad \bar{X} + t_{(\frac{\alpha}{2}, n-1)} \cdot \frac{s}{\sqrt{n}}\right)$$

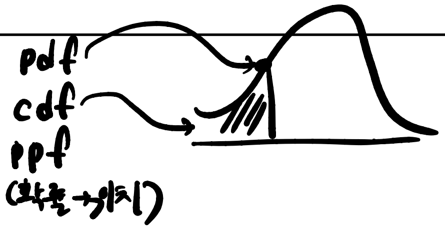
한양대학교 HANYANG UNIVERSITY

Sampling Distributions

```
In [1]: import numpy as np
        import matplotlib.pyplot as plt
        from scipy.stats import norm, t
```

-. *t* : t-distribution
-. *norm*: normal distribution

t (len(x) -1)



Sampling Distributions

.rvs (*size*)
-. *Random sample*

.pdf (*x*) or .pmf (*x*)
-. *Probability density or probability mass*

.cdf(*x*)
-. *Cumulative distribution*

.ppf(*prob*)
-. *point percentile*

Estimation

- Practice

점추정.

  - Compute the point estimator of the mean and 95% confidence Intervals of the mean.

```
In [2]: can = np.array([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, 98.7,101.4, 99.4, 101.7, 99, 99.7, 98.9
                        ,99.5, 100, 99.7, 100.9,99.7, 99, 98.8, 99.7, 100.9, 99.9
                        ,97.5, 101.5, 98.2, 99.2,98.6, 101.4, 102.1, 102.9, 100.8
                        ,99.4, 103.7, 100.3, 100.2, 101.1,101.8, 100, 101.2, 100.5
                        ,101.2, 101.6, 99.9, 100.5, 100.4, 98.1,100.1, 101.6, 99.3
                        ,96.1, 100, 99.7, 99.7, 99.4, 101.5, 100.9,101.2, 99.9, 99.1
                        ,100.7, 100.8, 100.8, 101.4, 100.3, 98.4,97.2])
```

평균 ← 표본평균 : mean.

신뢰구간 찾기. [L, U]

크면 → normal.
작으면 → t
(기준 : 30)

Estimation

- Practice

  - Solution:

```
In [7]: n = len(can)
        mean_can = np.mean(can)
        std_can = np.std(can,ddof=1)
        can_norm = norm(loc=0, scale=1)

        ll = mean_can - can_norm.ppf(0.975)*std_can/np.sqrt(n)
        ul = mean_can + can_norm.ppf(0.975)*std_can/np.sqrt(n)

        print("point estimator of mean is %.2f" % mean_can)
        print("95% confidence intervals is {} and {}".format(np.round(ll,4), np.round(ul,4)))

point estimator of mean is 100.04
95% confidence intervals is 99.7067 and 100.3733
```

Estimation

- Proportion
  - Point estimator about p

    X is the number of success from binomial distribution (n, p)

$$\hat{p} = \frac{X}{n}$$

Binomial→ normal

한양대학교 HANYANG UNIVERSITY

Estimation

- Proportion
  - (1-α)% confidence Intervals (with large sample)

$$\left(\hat{p} - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}, \quad \hat{p} + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right)$$

한양대학교 HANYANG UNIVERSITY

Estimation

- Practice
  - A company randomly select 250 customers, and send the promotion code for a new product. Among these, 70 people responded to buy the product using the promotion code. Then what would be the proportion of the customers to buy the product, and what is 90% confidence interval of that.

한양대학교 HANYANG UNIVERSITY

Estimation

- Practice
  - Solution:

In [10]:
```
n = 250
x = 70
p_hat = x/n

z_norm = norm(loc=0, scale=1)
ll = p_hat - z_norm.ppf(0.95)*np.sqrt(p_hat*(1-p_hat)/n)
ul = p_hat + z_norm.ppf(0.95)*np.sqrt(p_hat*(1-p_hat)/n)

print("the proportion is : ", p_hat)
print("90% confidence intervals of it is {} and {}".format(np.round(ll,4), np.round(ul,4)))
```

```
the proportion is :  0.28
90% confidence intervals of it is 0.2333 and 0.3267
```

한양대학교 HANYANG UNIVERSITY

Estimation

- Variance
  - Point estimator about σ^2

$X_1, X_2, \dots, X_n$ from the population whose mean is μ and the variance is σ^2

$$\widehat{\sigma^2} = s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Estimation

- Variance
  - (1-α)% confidence Intervals

$$\left(\frac{(n-1)\cdot s^2}{x^2_{(\frac{\alpha}{2},n-1)}}, \quad \frac{(n-1)\cdot s^2}{x^2_{(1-\frac{\alpha}{2},n-1)}}\right)$$

Estimation

- Practice
  - Find 95% confidence intervals of the variance and variance estimator with 'can' data.

Estimation

- Practice
  - Solution:

```
In [11]: from scipy.stats import chi2

n = len(can)
s2 = np.var(can, ddof=1)
chi_dist = chi2(n-1)

ll = (n-1)*s2/chi_dist.ppf(0.975)
ul = (n-1)*s2/chi_dist.ppf(0.025)

print("the variance estimator is : ", s2)
print("95% confidence intervals of it is {} and {}".format(np.round(ll,4), np.round(ul,4)))

the variance estimator is :  2.3130632911392417
95% confidence intervals of it is 1.7325 and 3.2452
```

Estimation

- The difference of two means
  - Point estimator of $\mu_1 - \mu_2$

$$\hat{\mu}_1 - \hat{\mu}_2 = \bar{X}_1 - \bar{X}_2$$

Estimation

- The difference of two means
  - (1-α)% confidence Intervals (with large sample)

$$\left(\bar{X}_1 - \bar{X}_2 - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \bar{X}_1 - \bar{X}_2 + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right)$$

Estimation

- The difference of two means
  - (1-α)% confidence Intervals (with small sample and equal variance)

$$s_p^2 = \frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{(n_1 + n_2 - 2)}$$

$$\left(\bar{X}_1 - \bar{X}_2 - t_{(\frac{\alpha}{2}, n_1 + n_2 - 2)} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X}_1 - \bar{X}_2 + t_{(\frac{\alpha}{2}, n_1 + n_2 - 2)} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

한양대학교 HANYANG UNIVERSITY

Estimation

- Practice
  - Calculate the pooled estimator of variance with A, B groups.
    A group is the first 10 observations of 'can' data.
    B group is the last 10 observations of 'can' data.

In [13]:
```
A_group = can[0:10]
B_group = can[-10:]

print("A group:",A_group)
print("B group:",B_group)
```

```
A group: [101.8 101.5 102.6 101.  101.8  96.8 102.4 100.   98.8  98.1]
B group: [101.2  99.9  99.1 100.7 100.8 100.8 101.4 100.3  98.4  97.2]
```

한양대학교 HANYANG UNIVERSITY

Estimation

- Practice
  - Solution:

```
In [15]: n1 = len(A_group)
         n2 = len(B_group)

         s2_pool = ((n1-1)*np.var(A_group, ddof=1) + (n2-1)*np.var(B_group, ddof=1))/(n1+n2-2)

         print("pooled estimator of the variance is ", np.round(s2_pool,4))

pooled estimator of the variance is  2.8729
```


Estimation

- Practice
  - Calculate 95% of the confidence intervals of the difference of the means.

Estimation

- Practice
  - Solution:

```
In [18]: a_mean = np.mean(A_group)
b_mean = np.mean(B_group)

t_can = t(n1+n2-2)
ll = (a_mean-b_mean) - t_can.ppf(0.975)*np.sqrt(s2_pool)*np.sqrt(1/n1 + 1/n2)
ul = (a_mean-b_mean) + t_can.ppf(0.975)*np.sqrt(s2_pool)*np.sqrt(1/n1 + 1/n2)

print("The difference of two means :", np.round(a_mean-b_mean, 4))
print("95% of CI of the difference of two means is {} and {}".format(np.round(ll,4), np.round(ul,4)))
```

```
The difference of two means : 0.5
95% of CI of the difference of two means is -1.0925 and 2.0925
```


Estimation

- The difference of two means
  - Point estimator of $p_1 - p_2$

$$\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

Estimation

- The difference of two proportions
  - (1-α)% confidence Intervals (with large sample)

$$\left(\hat{p}_1 - \hat{p}_2 - Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1 \cdot (1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}_1 \cdot (1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1-\hat{p}_2)}{n_2}}\right)$$

한양대학교 HANYANG UNIVERSITY

Estimation

- Practice
  - A survey company asked 100 male and 100 female if they married or not. 62 male and 29 female responded that they are married.

Compute the 90% confidence Intervals of the difference of two proportions.

한양대학교 HANYANG UNIVERSITY

Estimation

- Practice
  - Solution:

```
In [19]: n1 = n2 = 100
         p1_hat = 62/100
         p2_hat = 29/100
         z_norm = norm(loc=0, scale=1)

         ll = (p1_hat -p2_hat) - z_norm.ppf(0.975)*np.sqrt(p1_hat*(1-p1_hat)/n1 + p2_hat*(1-p2_hat)/n2)
         ul = (p1_hat -p2_hat) + z_norm.ppf(0.975)*np.sqrt(p1_hat*(1-p1_hat)/n1 + p2_hat*(1-p2_hat)/n2)

         print("The difference of two proportions :", np.round(p1_hat-p2_hat, 4))
         print("95% of CI of the difference of two proportions is {} and {}".format(np.round(ll,4), np.round(ul,4)))

         The difference of two proportions : 0.33
         95% of CI of the difference of two proportions is 0.1998 and 0.4602
```

Estimation

- Practice
  - Solution:

```
In [19]: n1 = n2 = 100
p1_hat = 62/100
p2_hat = 29/100
z_norm = norm(loc=0, scale=1)

ll = (p1_hat -p2_hat) - z_norm.ppf(0.975)*np.sqrt(p1_hat*(1-p1_hat)/n1 + p2_hat*(1-p2_hat)/n2)
ul = (p1_hat -p2_hat) + z_norm.ppf(0.975)*np.sqrt(p1_hat*(1-p1_hat)/n1 + p2_hat*(1-p2_hat)/n2)

print("The difference of two proportions :", np.round(p1_hat-p2_hat, 4))
print("95% of CI of the difference of two proportions is {} and {}".format(np.round(ll,4), np.round(ul,4)))
```

```
The difference of two proportions : 0.33
95% of CI of the difference of two proportions is 0.1998 and 0.4602
```