Chapter 8 Confidence Intervals

7 장으로 부터 sampling 을 배웠고 이를 토대로 얻은 값은 central limit theorem 에 의하여 sample 크기가 30 개 이상이거나 population 자체가 normal distribution 인 경우는 측정하고자하는 값은 normal distribution 을 근거로 얻을 수 있었고 값이 존재할 범위를 얻어 습니다. 이와 같이 범위로 값을 얻는 이유는 Sample 로 부터 얻은 정보가 불확신하기 때문입니다. 다시말해, population 의 정확한 정보를 알려 주기 보다는 특정 범위내에 원하는 정보가 있다라는 것으로 표현하지 않으면 안됩니다.

이렇듯 통계는 population 전체를 보기 보다는 sample 로 부터 정보를 얻어 일반화를 하려고 합니다. 즉, 부분만 보고 전체에 대한 결론을 끌어내도록 하기 위합니다. 그러다 보니 일반화된 정보들를 조립하여 새로운 것을 만드는 deductive reasoning 보다는 부분만으로 일반화를 만드는 추론(inductive reasoning)을 해야 합니다. 따라서 추론 과정에서 오류는 당연히 존재할 수 밖에 없습니다.

공통된 특징을 갖고 있는 대상인 population 에서 공통된 특징이 무엇인가를 찾기 위해서 앞서 배운 sample 과정 즉 일 부분만을 갖고 일반화 작업을 하는데, 여기서 population 의 정보값을 측정하는 것을 parameter estimation 이라 합니다. 다음에 두가지 측정 방법이 있습니다.

- 평균, 표준편차와 같은 한개의 정보값을 측정하는 point estimation(점측정)
- 평균, 표준편차의 범위를 측정하는 interval estimation(구간측정)

Point estimation 의 단점은 한값만을 알려 주기 때문에 잃어 버리는 정보가 많습니다. 반면 Interval estimation 은 point estimation 값이 같다 하더라도 데이타의 상황에 따라서 추론되는 범위가 다르게 나타나기 때문에 더 정확한 정보를 제공할 수 있습니다. 이렇게 측정된 범위를 confidence interval(신뢰구간)라고 합니다.

이 장에서 다룰 내용이 confidence interval 들은 다음과 같습니다.

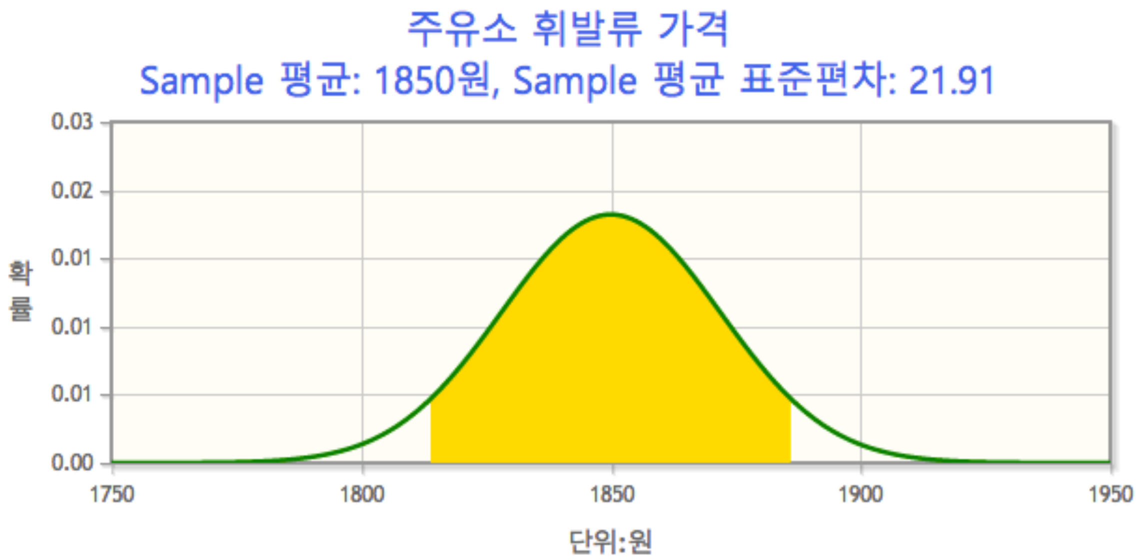
- 평균값 범위
  - Population 의 표준편차를 알고 있을 때
  - Population 의 표준편차를 모르 때
- 비율값의 범위
- 표준편차의 범위

이러한 confidence interval 을 알기 위해서 중요한 내용이 sample 크기로 원하는 결과를 얻기 위해 얼마나 많은 sample 이 필요한가를 계산 하는 방법도 학습하게 될 것입니다.

1. Population 의 표준편차를 알고 mean 의 confidence intervals(CI)계산

Sample mean 으로 부터 Population 의 표준표차를 통해서 Population mean 의 범위를 계산하기 위해서 우선 범위를 설정해야 하는데 일반적으로 90%, 95%, 99%와 같이 사용을 합니다. 이러한 범위는 평균을 중심으로 양쪽으로 반반씩 나뉘어야 하므로 90%일 경우 왼쪽에 45%, 오른쪽에 45%만을 찾이하고 양쪽 끝에 5%에 해당하는 범위를 배제 해야 합니다.

예를 들어 주유소 휘발류 가격의 평균을 조사하기 위해서 30 군대의 주유소의 가격을 조사한 결과 1850 원이고 population 의 표준편차가 120 원일 때 평균 값이 나타날 수 있는 범위를 보시면 다음과 같습니다.



노랑색 영역이 90%의 영역을 나타내는 범위로 그 외의 부분을 합하면 10%의 영역이 됩니다. 이 90%의 영역의 시작과 끝이 Confidence Intervals(CI)로 양 끝 포함되지 않은 영역을 significance level 라 합니다. 표기는 α (alpha)로 표기 하여, 90%의 CI 의 경우 α는 0.1 이 됩니다. 이를 이용해서 CI 를 계산하는 공식은 다음과 같습니다.

$$UCL_{\bar{x}} = \bar{x} - z_{\alpha/2}\sigma_{\bar{x}}$$
$$LCL_{\bar{x}} = \bar{x} + z_{\alpha/2}\sigma_{\bar{x}} \tag{8.1}$$

여기서 UCL 은 Upper Confidence Limit 의 약자이고 LCL 은 Lower Confidence Limit 의 약자입니다. $\bar{x}$는 sample mean 이고 $\sigma_{\bar{x}}$값은 수식 7.4 와 같습니다.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

jMath 에서 CI 를 계산하는 방법은 다음과 같습니다.

jMath.stat.ci(alpha, mu, sigma, n)

여기서 σ(sigma)는 population 표준편차이고 n 은 sample 의 크기 입니다.

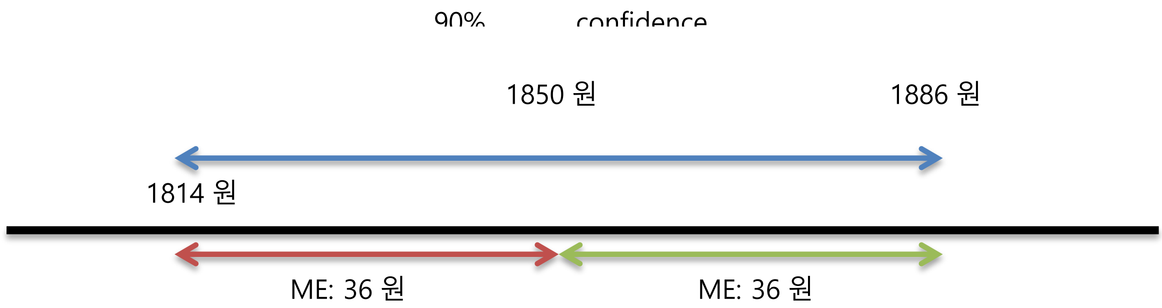
마지막으로 $z_{\alpha/2}\sigma_{\bar{x}}$에 해당하는 값을 Margin Error(ME)라고 합니다. ME 를 줄이는 방법은 significance level 을 늘리거나 sample 크기를 늘리면 됩니다.

주유소 휘발류 가격을 수식 8.1 에 적용하여 CI 를 계산하면 다음과 같습니다.

$$UCL = 1850 - z_{0.05}21.91 = 1850 + 1.649 \times 21.91 \approx 1886$$

$$LCL = 1850 + z_{0.05}21.91 = 1850 - 1.649 \times 21.91 \approx 1814$$

여기서 ME 값은 36 이 됩니다.



jMath 로 계산하는 예는

```
> jMath.stat.ci(0.1, 1850, 120,30 )
[1813.9630625889797, 1886.0369374110203]
```

이 결과의 의미는 30 개 주유소의 평균 가격 1850 원을 갖고 1814 원~1886 원 사이에 전체 주유소 평균 가격이 있을 확률이 90%라는 것을 말하는 것입니다. 그렇지 않을 확률이 10%입니다.

다른 의미로 30 개 주유소들의 평균 휘발류 가격을 나열을 하였을 때 계산된 CI 안에 모든 주유도의 평균 휘발류 가격이 있을 확률이 90%이고 그렇지 않을 확률이 10%라는 뜻이기도 합니다. 예를 들어 휘발류 값에 population 평균(μ) 가격이 1860 원 일 경우 앞의 1850 원으로 조사 된 결과의 CI 는 이 평균가격을 포함하고 있습니다. 이러한 과정을 반복하게 되면 다음과 같은 결과를 얻을 수 있습니다.

Sample 번호	Sample Mean	CI: ME 는 36	μ 포함여부
1	1800	1764 ~ 1836	X
2	1850	1814 ~ 1886	O

3	1830	1794 ~ 1866	O
4	1840	1804 ~ 1876	O
5	1870	1834 ~ 1906	O
6	1880	1844 ~ 1916	O
7	1890	1854 ~ 1926	O
8	1865	1829 ~ 1901	O
9	1855	1819 ~ 1891	O
10	1845	1809 ~ 1881	O

만일 첫번재 sample 을 갖고 조사를 했다면 CI 에 population mean 이 포함되지 않을 것이고 잘못된 결론을 만들수 있습니다. 하지만 이러한 반복을 많이 해보면 이렇게 포함되지 않을 확률이 10%가 된다는 것입니다.

Significance level 에 따라서 CI 범위가 달라지는데 일반적으로 많이 사용되는 것이 95%로 α값 5%이고 과학 분야에서는 99%를 많이 사용합니다. CI 가 넓어지게 된다면 population mean 이 존재 범위가 넓어지게 되기 때문에 더 넓은 영역을 빠르게 mean 의 위치를 찾게 되지만 정확하게 찾기 위해서는 CI 를 줄여야 합니다.

만약에 population 의 데이타 분포가 normal distribution 일 경우에는 sample 크기가 30 개 보다 작아도 됩니다. 예를 들어 일반적으로 한 학년에 남학생 평균 키를 얻기 위해서 sample 로 20 개를 얻어 CI 를 계산을 할 경우 sample 평균이 175cm 이고 population 에 표준편차가 10cm 일 경우 일반적으로 사람의 키에 대한 분포는 normal distribution 에 특성을 갖고 있어 20 개의 sample 값으로도 CI 를 계산 할 수 있습니다. 95%인 영역에 대한 ME 값으로 다음과 같습니다.

$$ME = z_{0.025}\frac{10}{\sqrt{20}} = 1.96 \times 2.2361 = 4.3826$$

UCL = 175 + 4.3826 ≈ 179.38cm

LCL = 175 – 4.3826 ≈170.62cm

CI 를 이용한 방법은 7 장과 같이 주장에 대한 타당성을 확인하는데 사용됩니다. 예를 들어 휴대폰을 만드는 업체가 방수폰을 만들었는데 7 등급으로 15cm 에서 1m 깊이에서 장시간

보호가 된다고 하는데 업체 주장은 실제로 80cm 까지는 장시간 보호 한다고 하고, 회사가 주장하는 표준 편차는 20cm 라고 할 때 sample 을 35 개를 조사한 결과 70cm 로 나왔다고 했을 때 95%영역내에 80cm 가 포함 되는가를 확인하도록 하겠습니다. ME 는 6.63 이고 이를 적용한 결과는 80cm 는 포함이 안되는 것을 확인 할 수 있습니다.

$$UCL = 70 - z_{0.025}\ 3.3806 = 70 + 1.96 \times 3.3806 \approx 76.63cm$$

$$LCL = 70 + z_{0.025} 3.3806 = 70 - 1.96 \times 3.3806 \approx 63.37cm$$

물론 운이 좋지 않아 불량품들만 선택된것이라만 이렇게 나올 수도 있기 때문에 몇번 더 테스트를 해볼 수 있지만 만약 10 번더 해서도 이런 결과가 나타난다면 문제가 있는건 확실합니다. 그렇지만 이 결과만 봐도 35 개를 random 으로 선별을 했을 경우 이런 문제가 발생될 확률이 5%이기 때문에 결과를 신뢰할 수 있습니다.

다음 기사 내용은 자동차 연비 검증에 대해 다룬 기사로 허용오차 범위인 Margin Error 를 넘어 서면 다른 것도 그런것인가를 확인 하기위해서 추가 조사를 하게 한다는 내용입니다.

자동차 연비 2017 년부터 엄격히 검증한다

연비 사후조사는 시험자동차 1 대를 선정해 측정하며 측정한 결과 허용오차범위(연비 -5%, 온실가스 +5%)를 초과하면 차량 3 대를 추가로 측정한다.

관계부처는 시험차량 대수를 놓고 오랜 협의 끝에 예산과 인력 등을 고려해 이같이 결정한 것으로 알려졌다. 이제까지 산업부는 3 대를 조사해 평균을 내고 국토부는 1 대를 조사해왔다.

연합뉴스 2014/07/13

이와 같은 차량의 대수는 sample 로 대수의 증가는 Margin Error 를 줄이는 효과가 있기 때문입니다. 하지만 sample 로 제공된 자동차만 실험에 통과한다면 결과의 신뢰성이 떨어져서 대수를 더 많이 하는 것이 좋겠지만 기사에서 보듯이 sampling 의 기본적인 문제인 시간과 돈때문에 쉽지는 않습니다.

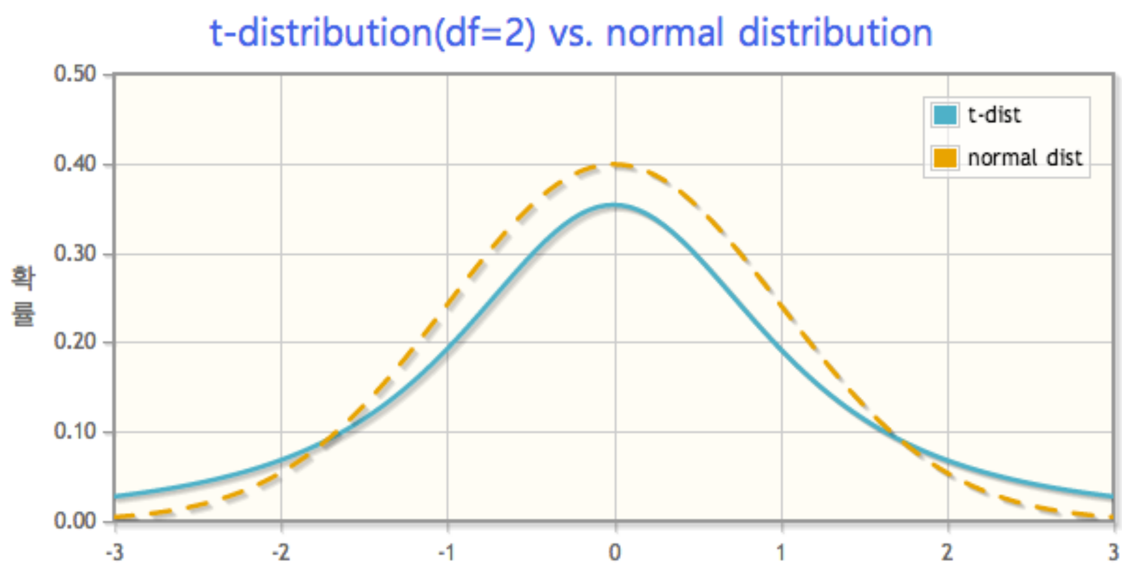
2. Population 의 표준편차를 모르고 mean 의 confidence intervals 계산

실질적으로 Population 의 표준편차를 알지는 못하기 때문에 어떤 주장에 제공되는 population 표준편차로 확인을 위해서 사용은 가능하지만 그렇지 않을 경우에는 population 표준편차를 대신할 것이 필요합니다.

이를 위해서 Student's t-distribution 이라는 것에 대한 소개를 하겠습니다. 이 distribution 은 normal distribution 과 같은 형태를 갖고 있지만, 표준 편차를 sample 의 표준편차를 사용합니다. 특징을 나열하면 다음과 같습니다.

- Normal distribution 과 같이 mean 값을 중심으로 벨 모양의 대칭의 형태로 구성됩니다.
- Degree of freedom(df)에 따라서 curve 모양이 다르게 나타납니다. 여기서 df 란 어떠한 값을 알고 있을 때 변화되는 값이 자유로운 값의 개수를 나타납니다. 예를 들어 10 개의 데이타에서 평균값을 알게 되면 9 개의 값을 바꾸면 1 값만 평균값이 나올 수 있도록 조절하면 됩니다. 이런 경우 degree of freedom 은 9 가 됩니다. 이러한 방식으로 N 개의 데이타가 있고 평균을 알 경우 degree of freedom 은 N-1 이 됩니다.
- Student's t-distribution 은 normal distribution 보다 더 넓게 퍼져있고, 만일 sample 개수가 많아지면 normal distribution 과 차이가 없어집니다.

chapter08/8_tdist.html



Student's t-distribution 역시 만일 sample 크기가 작다면 population 이 normal distribution 일 경우에는 상관이 없지만 그렇지 않다면 sample 크기는 30 개 이상이고 sample 로 부터 그려진 distribution 의 형태가 한쪽으로 크게 치우치지 않고 bell 모양을 유지한다면 sample standard deviation 은 population standard deviation 을 대체하여 사용할 수 있습니다.

Sample 평균들의 표준편차 근사치 값은 다음과 같습니다.

$$\hat{\sigma}_{\overline{x}} = \frac{s}{\sqrt{n}} \tag{8.2}$$

여기서 s 는 sample standard deviation 이고 n 은 sample 의 크기 입니다.

$$\begin{aligned} UCL_{\overline{x}} &= \overline{x} - t_{\alpha/2}\hat{\sigma}_{\overline{x}} \\ LCL_{\overline{x}} &= \overline{x} + t_{\alpha/2}\hat{\sigma}_{\overline{x}} \end{aligned} \tag{8.3}$$

수식 8.1 에서는 z-score 를 사용했지만 수식 8.3 은 t-score 를 사용합니다. t-score 를 계산하는 방법은 다음과 같습니다.

jMath.stat.tinv(p,df)

여기서 p 는 α/2 값이 되고 df 는 degree of freedom 으로 8.2 수식에서 sample 크기값에 n 에서 1 을 뺀 값입니다.

CI 값을 얻기 위한 방법은 다음과 같습니다.

jMath([data]).ci_t(alpha)

예를 들어 95%의 CI 를 구할 경우 p 는 0.05/2 로 0.025 가 됩니다.

```
> jMath.stat.tinv(0.025,19)
-2.093024021596347

> jMath.stat.tinv(0.025,29)
-2.0452296111397397
```

```
> jMath.stat.tinv(0.025,99)
-1.9842169002535168

> jMath.stat.norminv(0.025,0,1);
-1.9599639845400547
```

Degree of freedom 값이 증가 할 수록 z-score 값에 가까워 지는 것을 확인 할 수 있습니다.

커피 전문점에서 주중에 하루에 사용되는 종이컵량이 평균 55 개 정도 될 거라 생각하고 30 일 동안의 사용된 종이컵량을 나열해 보았습니다.

```
var obj = jMath('47 68 59 49 63 40 47 64 65 57 55 44 57 42 56
                 53 61 46 56 58 41 52 66 51 47 55 47 54 49 50');
console.log( obj.ci_t(0.05) );

[50.4458446948482, 56.1541553051518]

var m = obj.mean(3)
53.3

var s = obj.std()/Math.sqrt(30)
1.395518278048631

m + s*jMath.stat.tinv(0.975,29)
56.1541553051518

m + s*jMath.stat.tinv(0.025,29)
50.4458446948482
```

결과를 보면 CI 에 55 가 포함되어 있기 때문에 생각했던 55 개가 어느 정도 맞을 수 있습니다.

3. Proportions(Binomial Parameter)에 대한 Confidence Intervals

Proportions 에 대한것은 binomial distribution 으로 population 에 proportion π가 있고 sample 크기 n 일 때

$$\pi n \geq 5 \text{ 이고 } (\pi-1)n \geq 5$$

만족을 하게 된다면 normal distribution 과 대략적으로 맞아 normal distribution 처럼 취급할 수 있습니다.
이를 토대로 proportion 에 대한 standard error 는 다음과 같습니다.

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \tag{8.4}$$

하지만 CI 로 구하는 대상이 π 이기 때문에 계산을 위해서 sample proportion(p)을 이용할 수 있습니다.

$$\hat{\sigma}_p = \sqrt{\frac{p(1-p)}{n}} \tag{8.5}$$

이를 이용하여 CI 계산은 다음과 같습니다.

$$\begin{aligned} UCL_p &= p - z_{\alpha/2}\hat{\sigma}_p \\ LCL_p &= p + z_{\alpha/2}\hat{\sigma}_p \end{aligned} \tag{8.6}$$

여기서 t-score 를 사용하지 않고 z-score 를 사용하는 이유는 sample 의 크기가 크기 때문에 z-score 를 사용해도 되기 때문입니다.

jMath.stat.ci_p(alpha,p,n)

Margin Error 값인 $z_{\alpha/2}\hat{\sigma}_p$로 예를 들어 선거 기간에 후보 지지율을 조사할 때 2000 명을 조사한 결과 한 후보의 지지율값인 p 가 50%로 나타났을 경우 95%의 CI 를 통해서 ME 를 살펴본다면

$$z_{0.975}\hat{\sigma}_p = 1.96\sqrt{\frac{0.5 \times 0.5}{2000}} \approx 0.022$$

약 2.2%가 됩니다. 이 내용의 의미는 지지율 조사에서 말하는 오차범위 ±2.2%라는 말을 의미하는 것으로 CI 를 통해서 전국민에 지지율을 47.8%~52.2%내로 존재한다는 의미입니다.

```
jMath.stat.ci_p(0.05, 0.5, 2000)
[0.4780869364855855, 0.5219130635144146]
```

4. Sample 크기 결정하기

실제로 조사를 위해서 필요한 중요한 내용 중 하나가 바로 알맞는 sample 크기를 알아내는 것입니다. 이를 계산하는 위한 예를 다음과 같습니다.

피자 가계에서 주중에 하루 평균 온라인 주문된 개수를 오차범위(Margin Error) 2 개로 95%의 확신이 되는 영역을 알고 싶다고 할 때 다음의 수식으로 계산을 할 수 있습니다.

$$ME_{\bar{x}} = z_{\alpha/2}\sigma_{\bar{x}} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
$$n = \left(\frac{z_{\alpha/2}\sigma}{ME_{\bar{x}}}\right)^2 \tag{8.7}$$

CI 가 95%인 경우 z-score 는 1.96 이고 표준편차를 10 으로 했을때 sample 크기는

$$\left(\frac{1.96 \times 10}{2}\right)^2 = 96.04 \approx 97$$

실제로 population 에 표준편차를 알 수가 없기에 sample 표준편차를 대신하여 사용이 가능합니다. 이 경우 z-score 대신 t-score 를 사용해야 하나 t-score 는 degree of freedom 에 따라서 값이 다르기 때문에 사용에 문제가 있지만 만일 sample 크기가 크다면 z-score 와 t-score 는 거의 비슷하기 때문에 z-score 를 대신 활용해도 문제가 되지 않습니다.

이러한 sample 크기 계산에서 수식 8.7 에서 보는 것과 같이 ME 에 반비례하기 때문에 허용 오차범위가 커지게 되면 sample 크기는 줄어 듭니다.

jMath.stat.numSamples(alpha,me,sigma)

```
jMath.stat.numSamples(0.05, 2, 10)
97

jMath.stat.numSamples(0.05, 3, 10)
43

jMath.stat.numSamples(0.05, 4, 10)
25
```

Sample 크기 계산을 Proportion 에 적용을 한다면 $\hat{\sigma}_p = \sqrt{\frac{p(1-p)}{n}}$를 수식 8.7 에 적용하여 계산을 하게 되면 됩니다.

$$n = \left(\frac{z_{\alpha/2}}{ME_{\overline{p}}}\right)^2 p(1-p) \tag{8.8}$$

예를 들어 음식점에서 현금으로 결제 하는 손님이 50 명의 손님중 7 명이 현금으로 결제를 했다면 p 는 0.14 로 이를 토대로 95% CI 로 오차범위 2%로 할 경우 필요한 sample 의 크기는 140 명이 필요합니다.

$$\left(\frac{1.96}{0.02}\right)^2 \times 0.14 \times (1-0.14) \approx 1157$$

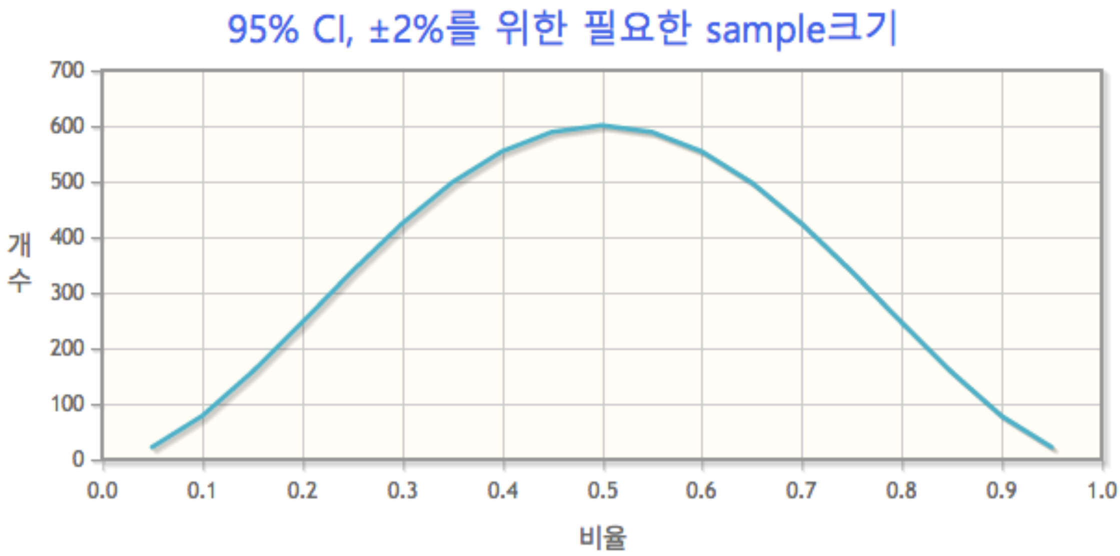
jMath.stat.numSamples(alpha,me,p,true)

```
jMath.stat.numSamples(0.05, 0.02, 0.14, true)
1157
```

하지만 만약 p 값을 결정할 수 없다면 p 를 0.5 로 계산된 sample 크기로 하면 됩니다. 이유는 수식 8.8 에 최대값은 p 가 0.5 이기 때문입니다.

$$\left(p(1-p)\right)' = (1-p) - p = 0$$
$$p = \frac{1}{2}$$

chapter08/8_numsamples.html


95% CI, ±2%를 위한 필요한 sample크기
개수
비율
0
100
200
300
400
500
600
700
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

5. Population 크기가 작을 경우

Sample 크기 n 과 Population 크기 N 에 비율이 0.05 보다 클 경우 7 장에서 Correction Factor 를 적용했습니다. 결과 Confidence Interval 계산은 다음과 같이 변형이 됩니다.

Population 표준편차를 알 경우

$$UCL_{\overline{x}} = \overline{x} - z_{\alpha/2}\sigma_{\overline{x}}\sqrt{\frac{N-n}{N-1}}$$
$$LCL_{\overline{x}} = \overline{x} + z_{\alpha/2}\sigma_{\overline{x}}\sqrt{\frac{N-n}{N-1}} \qquad (8.9)$$

Population 표준편차를 모를 경우

$$UCL_{\overline{x}} = \overline{x} - t_{\alpha/2}\hat{\sigma}_{\overline{x}}\sqrt{\frac{N-n}{N-1}}$$
$$LCL_{\overline{x}} = \overline{x} + t_{\alpha/2}\hat{\sigma}_{\overline{x}}\sqrt{\frac{N-n}{N-1}} \qquad (8.10)$$

Proportion

$$UCL_p = p - z_{\alpha/2}\hat{\sigma}_p\sqrt{\frac{N-n}{N-1}}$$
$$LCL_p = p + z_{\alpha/2}\hat{\sigma}_p\sqrt{\frac{N-n}{N-1}} \qquad (8.11)$$

예를 들어 1000 명의 직원이 있는 보험회사에서 일주일 평균 근무 시간을 측정을 하는데 100 명의 직원을 조사하니 일주일 평균 46.32 시간으로 sample 에 표준편차가 4.93 시간이라고 할 때 95%의 CI 계산은 100/1000 은 0.1 로 0.05 보다 크기 때문에 수식 8.10 을 적용하면 됩니다.

jMath([data]).ci_t(alpha, N)

N 은 population 의 총 크기 입니다.

```
> obj = jMath('44 42 42 40 46 57 42 45 52 52 48 51 50 44 48 45 42 49 46 45 47 51 48 47
47 46 42 50 53 43 56 45 45 42 42 41 48 54 49 50 53 34 45 48 50 39 41 50 45 49 41 38 50
52 54 38 39 39 46 43 53 39 37 47 52 42 40 45 38 46 50 47 36 43 44 55 51 46 53 41 44 53
54 50 48 43 50 42 49 47 52 44 49 50 42 47 54 47 44 48');

> obj.ci_t(0.05, 1000)
```

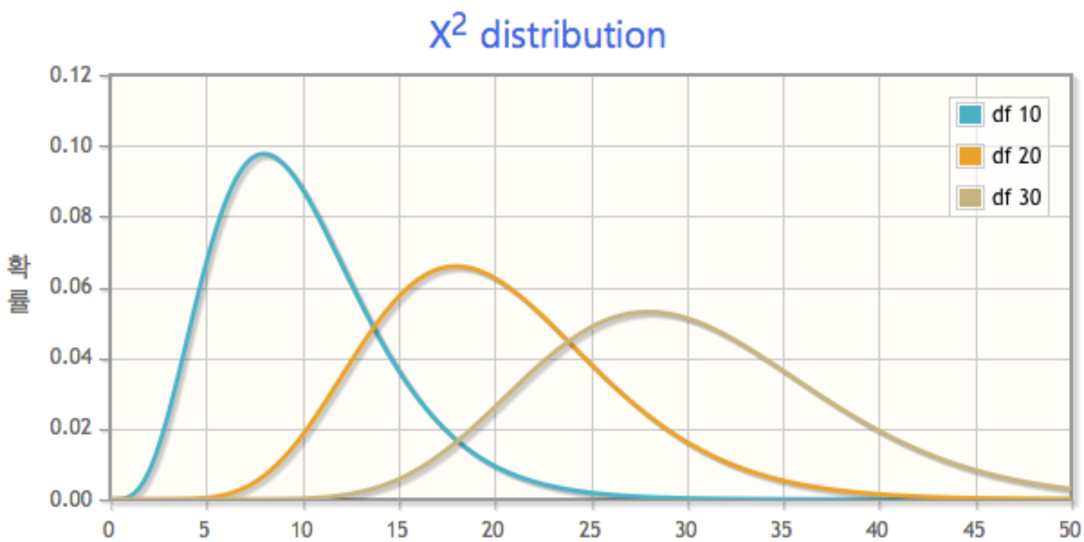
```
[45.39236983771909, 47.24763016228091]
```

6. Normal distribution 에서 variance 의 Confidence Intervals

Population 이 normal distribution 으로 형성되어 있는 상황에서 평균과 표준편차 둘다 모를 때 편차의 confidence interval 을 계산하기 위해서 χ^2 distribution (Chi-square distribution)을 이용해야 합니다.

$$\chi^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \tag{8.12}$$

Chi-square distribution 은 아래의 그래프 처럼 degree of freedom 에 따라서 값이 다르게 나타나는데 이 계산을 위한 degree of freedom 은 sample 의 크기 n 에 1 을 뺀 값이 됩니다.



이 수식을 이용해서 confidence interval 을 계산을 하게 되면 UCL 과 LCL 은 다음과 같습니다.

$$UCL_{\sigma^2} = \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}$$
$$LCL_{\sigma^2} = \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}} \tag{8.13}$$

고양시에 운영하는 유료 자전거 시스템인 피프틴(FIFTEEN)은 자전거의 평균 속도인 시속 15Km 를 의미하는 뜻으로 회원은 1 년에 회비 6 만원 비회원은 한번이용시 1000 원의 이용료로 이용할 수 있습니다. 이용은 한번에 40 분까지(평균 이동거리 10Km) 무제한으로 자전거를 이용할 수 있습니다.

이러한 시스템에서 중요한 것은 항시 대기 자전거 개수 입니다. 이를 위해서 평균도 중요하지만 표준편차도 중요합니다. 예를 들어 24 일 동안 20 대 자전거가 있는 한 장소에 아침 출근 시간에 남아 있는 자전거의 개수를 조사하여 한 결과가 다음과 같다고 하겠습니다.

17	3	6	12	13	13	7	10	6	7	10	15
8	11	9	13	7	5	1	13	15	10	6	8

표준 편차의 95%범위을 구하기 위해서 수식 8.13 을 적용하면 다음과 같습니다.

```
var a = jMath('17 3 6 12 13 13 7 10 6 7 10 15 8 11 9 13 7 5 2 13 15 10 6 8')
> a.ci_std()
[3.115710769846267, 5.623417312168498]
> a.ci_t()
[7.682223026827215, 11.067776973172785]
```

jMath.prototype.ci_std 함수는 표준편차 95% CI 를 계산할 수 있도록 합니다. 결과를 보시면 약 3 대에서 5.6 대의 표준편차가 있습니다. 평균만 보았을 때 최저거 약 7.7 대가 있다고 보지만 표준 편차와 결합을 하면 최저 2 대까지 낮아 질 수 있기 때문에 5 대 이상을 항시 보유하고 있어야 한다면 자전거 대수를 20 대에서 23 대로 늘려야 합니다.