Chapter 9 Hypothesis Test for a Single Population

가설(hypothesis)을 만들고 sample 로 부터 가설이 합당함을 추론하는 것을 hypothesis test 라고 합니다. Test 하는 대상은 통계에서 sample 을 통해서 population 을 이해하는 것이기 때문에 population 의 parameter 들 입니다. 이 장에서는 하나의 population 만을 다루는 것을 배우게 됩니다. 이러한 상황에서 hypothesis test 는 sample 로 부터 얻은 정보를 갖고 가설로 만든 population 의 parameter 가 타당한가를 추론하는 것입니다.

예를 들어

- 휴대폰을 만드는 업체가 새로 개발된 휴대폰이 충전이 완료된 후에 연속 대기 시간이 300 시간 이상이라고 주장을 할 때 이를 확인할 때
- PC 방에 새로 구매한 의자에 팔걸이를 갈아야 하는 시점이 평균 200 일이라 하여 이에 맞춰 수리할 수 있도록 준비하려는데 이것이 맞는 정보인지 확인할 때
- 오피스텔 및 상가 주변에서 음식점을 운영하려는 상점 주인이 배달의 비율이 전체 매출에 60%일 것이라 신문 기사를 보고 매장 인테리어 및 규모를 결정하로고 할 때 주변 음식점들의 배달률을 sample 로 하여 기사가 맞는지 확인 할 때.

1. Hypothesis test 개론

두개의 가설(hypothesis)이 있습니다. 첫 번째 것은 null hypothesis 라 하여 H_0 으로 표시하는데 이것은 일반적인 값입니다. 즉 population parameter 값에 대한 예측 혹은 기대값입니다. 다음으로 alternative hypothesis 라 하여 H_1 으로 표시하는 값으로 null hypothesis 를 반박하는 내용입니다. 예를 들어 PC 방에 의자 팔걸이 교체 시기를 평균 200 일로 본것은 null hypothesis 가 되고 그렇지 않는 것이 alternative hypothesis 가 됩니다.

$$H_0: \mu = 200 \text{ 일}$$

$$H_1: \mu \neq 200 \text{ 일}$$

여기서 주의 할 점은 hypothesis test 는 population 을 위한 것이지 sample 에 대한 test 가 아니기 때문에 평균값에 대한 hypothesis 는 population mean(μ)이지 sample mean(s)가 아닙니다. 즉 hypothesis test 를 통해서 population 에 정보를 추론하는 것입니다.

다른 예로 일반 휴대폰에 충전지들의 연속 대기 시간이 300 시간일 때 어느 업체가 만든 휴대폰이 이것 보다 에너지 효율성이 좋다는 것을 입증을 하려고 합니다. 이 검증에서 null hypothesis 는 일반적인 population 에 평균값인 300 시간보다 작은가를 보는 것이고 alternative hypothesis 는 이 시간보다 더 큰가를 알아 보는 것입니다.

H_0: μ ≤ 300 시간

H_1: μ > 300 시간

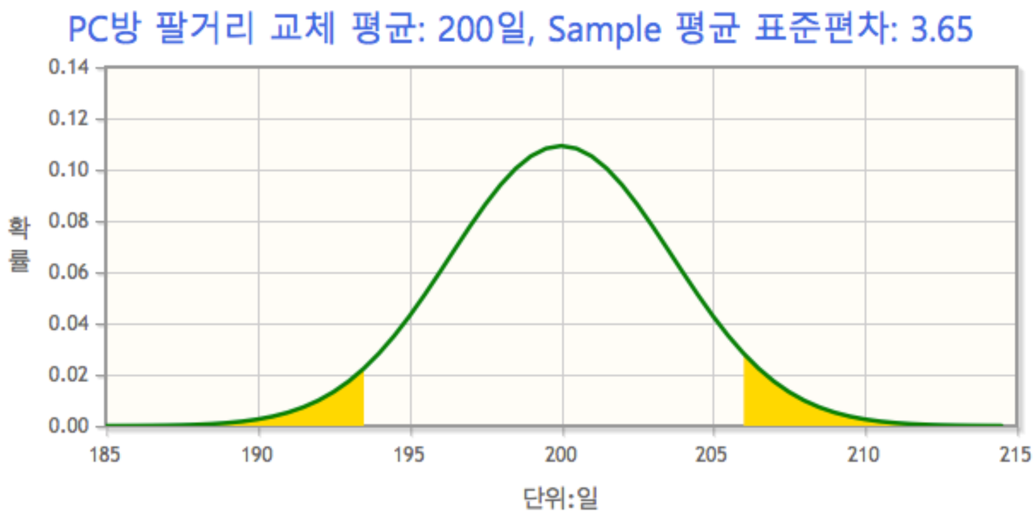
여기서 =를 사용하지 않는 이유는 =는 이 휴대폰의 연속 대기 시간이 300 시간보다 매우 적게 나오거나 많게 나오면 무조건 null hypothesis 인 300 시간이 아니라는 타당성을 제시하기 때문에 이 업체가 주장하는 더 오랜 연속 대시 시간을 확실하게 증명할 수 없기 때문입니다.

이 두가지 경우 처럼 가설을 검증하는 방법은 두가지로 나타납니다.

Two-tail hypothesis test

PC 방 의자 팔걸이와 같이 null hypothesis 에서 같다 "=", alternative hypothesis 는 같지 않다 "≠"로 되어 있을 경우 입니다. 이 경우 90%확실성을 갖고 test 한다면 다음과 결과를 얻을 수 있습니다.

chapter09/9_twotail.html

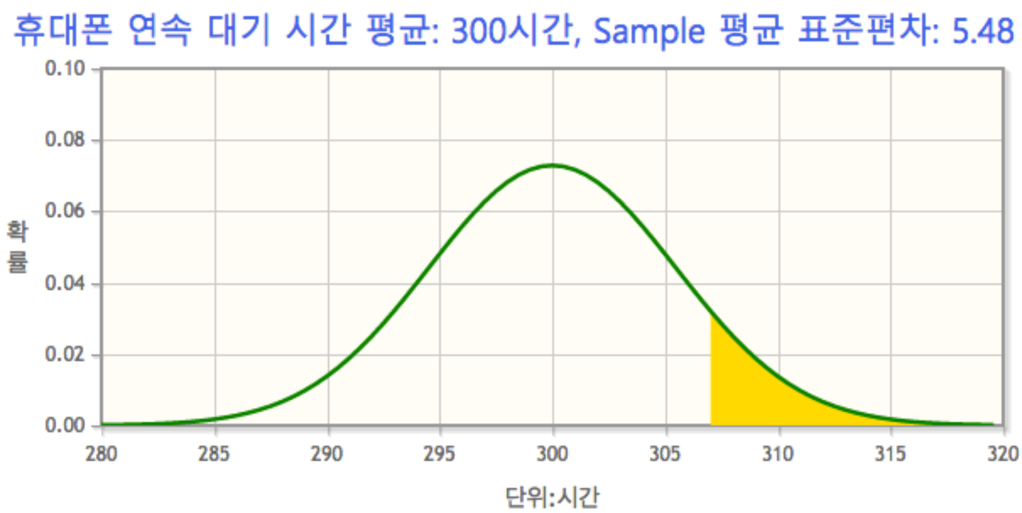


노랑색 영역은 각각 5%에 해당하는 영역으로 null hypothesis 가 부적합한 영역입니다. 만일 sample 로 부터 얻은 값이 이 영역에 있게 된다면 null hypothesis 는 reject 된다고 합니다.

One-tail hypothesis test

크다 작다로 alternative hypothesis 를 형성을 할 경우로 휴대폰 연속 대기 시간에 대한 hypothesis test 가 이 경우에 해당합니다. 90%의 확신성을 갖게 test 를 하려고 할 경우 alternative 가 ">"로 한다면 상위 10%영역에 포함되면 null hypothesis 는 reject 되고 반대로 "<"인 경우는 하위 10%에 포함되면 null hypothesis 는 reject 됩니다.

chapter09/9_onetail.html



alternative hypothesis test 가 μ > 300 시간으로 노랑색으로 나타난 10%에 영역에 sample mean 이 포함되면 null hypothesis 는 reject 됩니다.

이런 검증에서 주의할 점은 결과에 대한 해석입니다. 예를 들로 오피스텔 및 상가 주변 음식점에 배달률에 따른 매장 규모 및 인테리어 결정을 하려고 할 때 신문 기사에서 60% 기반으로 결정을 위한 가설은 다음과 같습니다.

$$H_0: \pi \geq 0.6$$
$$H_1: \pi < 0.6$$

즉, 상점 주인이 알고 싶은 내용은 0.6 보다 작으면 매장 규모 및 인테리어를 다시 고려해야 되기 때문에 이러한 방식으로 테스트해야 합니다. 이러한 hypothesis test 결과 null hypothesis 를 reject 하는데 실패를 했다면 의미는 0.6 보다 작다는 것을 증명하기 위한 충분한 증거를 못찾었다는 의미이기도 하기만 또한 0.6 이상이라는 확실한 증거 또한 존재하지 않기 때문에 무조건 결과가 옳다고 판단하기에는 어려움 점이 있습니다.

이말을 정리를 하면 Test 결과에 대한 null hypothesis 의 해석은 null hypothesis 가 참이다라고 말하지 못합니다. 대신 null hypothesis 를 reject 할 충분한 증거가 없다는 것이고, 또한 null hypothesis 이 참이라고 확실하게 말할 수 있는 충분한 증거도 없습니다.

이러한 내용을 토대로 본다면 hypothesis test 에 주 내용은 일반적인 사실과 다르다라는 것을 알리기 위한 목적으로 test 를 하기 때문에 null hypothesis 는 일반적으로 알려진 내용이 되고 alternative hypothesis 는 주장이나 일반적인 것과 다른 내용이면 됩니다. 그래서 null hypothesis 를 reject 하는 경우 주장이 맞을 확률이 높다는 것을 알리는데 목적을 두고 있습니다.

Hypothesis 에서 one-tail 과 two-tail 중 선택하는 방법에 대해서 정리하면, 어느 특정 값에 대한 조사를 하는 경우 two-tail 로 같다 그렇지 않다로 판단하는 경우 입니다. 예를 들어 나무를 잘라서 판매를 할 때 길이가 정확하게 잘라지는지를 확인 해아 하는 경우 잘려진 나무의 길이는 특정 값이기 때문에 two-tail 로 hypothesis test 를 하면 됩니다. 하지만,

자동차 연비 향상, 냉장고 전력 소비 감소등 일반적인 값보다 상향되거나 하향되었다는 것을 검사할 때는 one-tail 로 검사를 해야 합니다.

Sample error 종류: Type I & Type II error

Sample 로 부터 얻어진 정보는 sample error 가 존재하기 때문에 그대로 population 의 정보로 사용을 못합니다. 이러한 sample error 는 두 가지로 구분이 됩니다.
우선 Type I error 는 null hypothesis 를 reject 되었지만, 실제로 null hypothesis 가 맞을 확률을 말하는 것으로 level of significance 와 같은 의미로 α로 표기를 합니다.

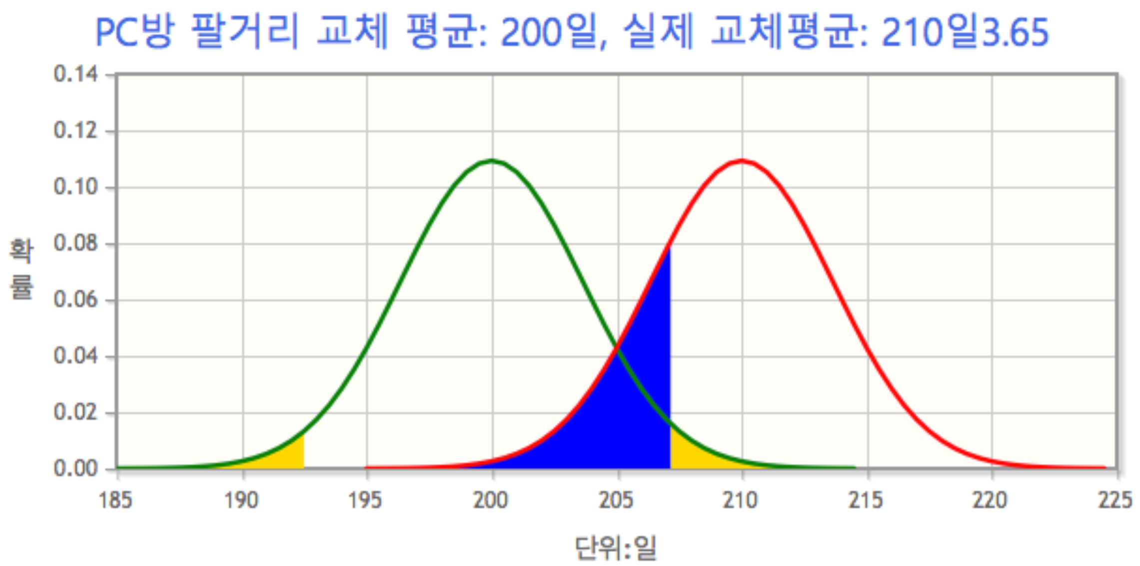
$$\alpha = P(Reject\ H_0|H_0\ true)$$

다른 오류로 Type II error 는 null hypothesis 를 reject 하는 것을 실패 했지만 실제로 null hypothesis 가 잘못 되었을 확률로 β로 표기 합니다.

$$\beta = P(Accept\ H_0|H_1\ true)$$

예를 들어 PC 방 의자 팔걸이 교체 시기를 평균 200 일로 잡고 test 를 하려고 하는데 실제 평균이 210 일 경우 발생되는 상황을 보겠습니다.

chapter09/9_errors.html



이 경우 sample 평균값이 208 일로 나타났다면 null hypothesis 로 생각한 200 일로 처리하였을 때 노랑색 영역인 207.16 이상에 포함되므로 null hypothesis 를 reject 를 하게 됩니다. 하지만 sample error 로 실제로 200 일인데 우연히 sample 로 사용된 것이 208 일이 나온 경우라면 type I error 에 해당 합니다.

그런데 만일 실제 평균 교체가 200 일이 아닌 210 일 경우 sample 로 부터 얻은 평균 교체일이 202 일로 나타날 경우 문제가 됩니다. 200 일로 보았을 때 null hypothesis 를 reject 하지 못한 상황이지만 210 일에서 보았을 때는 reject 영역에 있는 경우가 됩니다.
위의 그림에서 210 일이 실제 값인데도 불구하고 null hypothesis 인 200 일에서는 reject 되지 못하는 영역을 파랑색으로 표시 했는데 이것을 type II error 라 합니다.

$$\beta = P(Accept\ H_0 | \mu = 210)$$

다른 예를 들어 충전지에 충전 횟수가 1000 번 가능하다고 했을 경우 실험 결과 type I error 가 발생 하면 1000 번이 맞는데도 그렇지 않다고 나왔기 때문에 생산자에게 문제를 발생합니다. 반면에 만일 type II error 가 발생을 하여 실제 970 번만 사용이 가능한데 1000 번으로 설정한 테스트 결과에서 맞는다고 나타난다면 구매한 고객은 손해를 보게 됩니다. 그래서 type I error 를 producer's risk 라고 하고 type II error 를 customer's risk 라고도 합니다. 이 내용에 대해서는 나중에 자세히 다룰 것입니다.

가설 검증을 위한 절차는 다음과 같습니다.

1) Null 과 alternative hypothesis 를 설정합니다.
2) Level of significance 를 설정합니다. 일반적으로 5%를 사용합니다.
3) 검사에서 사용 가능한 score 로 level of significance 에 해당하는 Critical value 를 결정합니다.
4) 검사에서 사용 가능한 test statistic 인 값을 구합니다.
5) Test statistic 과 critical value 와 비교하여 hypothesis 에 대한 결정을 합니다.

검사에 상황에 따른 분류를 하여 hypothesis test 를 하는 방법을 소개 하겠습니다.

2. Hypothesis test for Population mean : σ 알고 있음.

z-score 값을 이용해서 null hypothesis 가 reject 시킬 수 있는지를 확인합니다. Hypothesis test 에서 사용하는 test statistic 은 다음과 같습니다.

$$z_{\overline{x}} = \frac{\overline{x} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}} \tag{9.1}$$

수식 9.1 을 이용해서 Null hypothesis 를 reject 하기 위한 조건은 다음과 같습니다.

Hypothesis Test	Critical Sample mean($\bar{x}_\alpha$) 범위	Reject 조건
Two tail H_0: μ = μ_0 H_1: μ ≠ μ_0	$\mu_{H_0} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \overline{x}_\alpha \le \mu_{H_0} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$	$\lvert z_{\overline{x}}\rvert > \lvert z_{\alpha/2}\rvert$
Upper tail H_0: μ ≤ μ_0 H_1: μ > μ_0	$\overline{x}_\alpha \le \mu_{H_0} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$	$z_{\overline{x}} > \lvert z_\alpha\rvert$
Lower tail H_0: μ ≥ μ_0 H_1: μ < μ_0	$\overline{x}_\alpha \ge \mu_{H_0} + z_\alpha\frac{\sigma}{\sqrt{n}}$	$z_{\overline{x}} < z_\alpha$

(9.2)

예를 이용해서 test 를 하는 절차를 설명하도록 하겠습니다.

기존에 냉장고의 한달 소비전력량이 35.4kWh 에 표준편차 8kWh 라고 했을 때 새로 개발된 냉장고의 소비전력이 32.2kWh 라고 하였을 때 30 개의 sample 을 조사하여 보니 33.2kWh 라고 나타났습니다. 이 결과만으로는 성능이 향상되어 보입니다만 정확한 판단을 위해 가설 검사를 할 수 있습니다.

1) Hypothesis 설정: lower tail

소비전력의 향상은 기존의 값보다 작게 나타나야 되기 때문에 one-tailed hypothesis 에서 lower tail 로 설정을 맞춰야 합니다.

H_0: μ ≥ 35.4kWh

H_1: μ < 35.4kWh

2) Level of significance α = 0.05
3) Critical value

$$z_{\overline{x}} < z_{0.05}(= -1.65)$$

4) test statistic

$$z_{\overline{x}} = \frac{33.2 - 35.4}{\frac{8}{\sqrt{30}}} = -0.15062$$

5) Null hypothesis 를 reject 검사

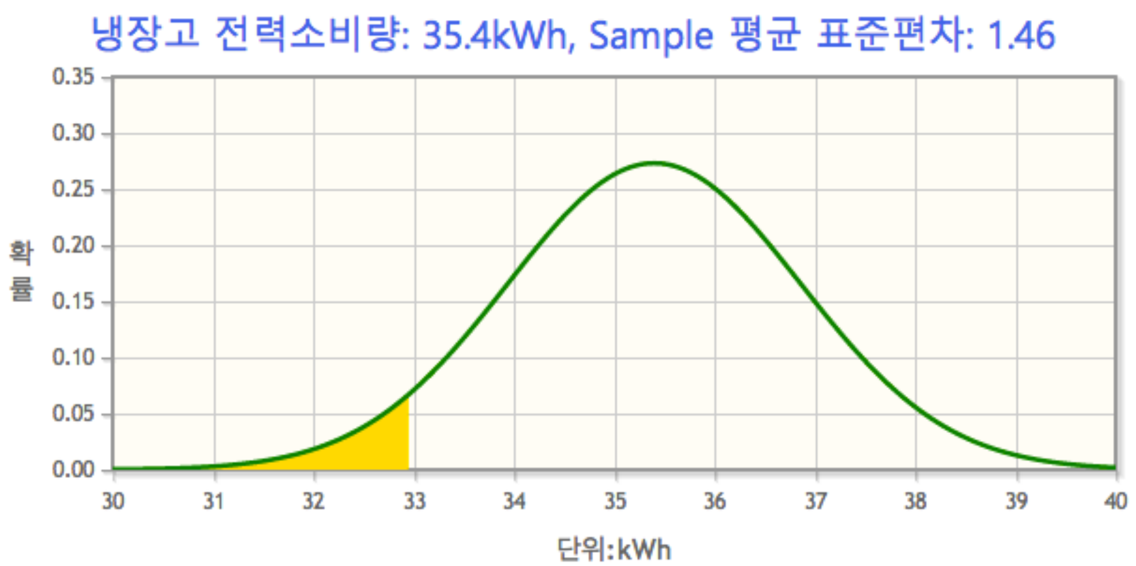
Lower tail 의 검사의 경우 $z_{\overline{x}} < z_{\alpha}$ 조건이 설립되는지 검사하는데 -0.15062 은 critical value 인 -1.65 보다 크기 때문에 null hypothesis 를 reject 할 충분한 근거를 만들지 못합니다.

6) Confidential interval : 32.9975 이상

33.2 는 32.9975 보다 크기 때문에 이 또한 null hypothesis 를 reject 하지 못함을 알려주는 값입니다.

이 결과를 그래프로 확인하면 다음과 같습니다.

chapter09/9_refrige.html



그래프에 노랑색 영역 α값이 0.05 인 영역으로 이 안에 sample mean 이 존재 하여야 null hypothesis 를 reject 할 근거가 됩니다. 하지만 33.2kWh 는 노랑색 영역에 있지 않기 때문에 성능이 향상되었다고 말할 충분한 근거가 없게 됩니다.

- **p-value: observed level of significance**

null hypothesis 가 참이라는 가정하에 sample mean 이 관측될 최악의 확률을 말합니다. 예를 들어 PC 방 의자 팔걸이가 평균 200 일 교체 시기를 갖는지를 확인하려고 할 때 null hypothesis 는 200 일이고 population 표준편차를 20 일이라고 했을 경우 sample mean 값이 208 일로 나타났다면 이 값이 나올 확률이 p-value 입니다.

One-tailed : $\mu > \mu_0$ $$P\left(x > \overline{x} | \mu_{H_0}, \frac{\sigma}{\sqrt{n}}\right) = 1 - P\left(x < \overline{x} | \mu_{H_0}, \frac{\sigma}{\sqrt{n}}\right)$$

One-tailed: $\mu < \mu_0$ $$P\left(x < \overline{x} | \mu_{H_0}, \frac{\sigma}{\sqrt{n}}\right) \tag{9.3}$$

Two tailed $$2 \times \min\left(P\left(x < \overline{x} | \mu_{H_0}, \frac{\sigma}{\sqrt{n}}\right), P\left(x > \overline{x} | \mu_{H_0}, \frac{\sigma}{\sqrt{n}}\right)\right)$$

PC 방 팔걸이 교체 시기에 대해서 보시면 208 일은 수식 9.3 에 적용을 했을 때 약 2x0.0142 로 약 2.8%가 됩니다.

```
> var sm = 208;
> var H0 = 200;
```

```
> var n = 30
> var sigma = 20;
> 2*(1-jMath.stat.normcdf(sm,H0,sigma/Math.sqrt(n)))

0.02845973691631043
```

이 값은 level of significance 값 α값인 5%보다 작기 때문에 null hypothesis 를 reject 할 근거가 됩니다.

주의할 점은 이러한 결과가 나타났다고 sample mean 에 해당하는 208 일이 population mean 으로 생각하면 안됩니다.. 이 값은 단지 null hypothesis 에 200 일을 반박하는 근거 자료일 뿐 더 이상의 의미가 없습니다.

이 가설 검사를 위한 jMath 사용법은 다음과 같습니다.

```
jMath.stat.test( alpha, type, x, H0, sigma, n, N )
```

여기서

type 의 값은 -1: lower tail, 0: two tail, 1: upper tail

x 는 sample 평균값

sigma 는 population 의 표준편차

n 은 sample 크기

N 은 population 크기입니다.

여기서 N 은 표준편차 수식 8.9 인 correction factor 계산을 위한 것으로 모를 경우 넣지 않아도 됩니다.

```
> jMath.stat.test( 0.05, 0, 208, 200, 20, 30 )
Object {H0: 200, alpha: 0.05, popSigma: 20, sampleSize: 30, popSize: undefined…}
H0: 200
alpha: 0.05
ci: [ 192.84322342513136, 207.15677657486864]
meanStd: 3.6514837167011076
popSigma: 20
popSize: undefined
pvalue: 0.02845973691631043
sampleSize: 30
testscore: 2.1908902300206643
type: 0
```

```
zscores: [-1.9599639845400547,1.9599639845400547]
```

3. Hypothesis test for Population mean : σ 모름.

실제 상황에서 population 의 표준편차를 알지는 못하기 때문에 sample 만을 갖고 모든 것을 처리하는 경우 Student's *t*-distribution 을 활용해야 합니다. 그런데 이를 활용하기 위한 전제 조건은 population 은 normal distribution 으로 되어 있어야 합니다. 앞으로의 예제들은 population 이 normal distribution 으로 되어 있다는 가정으로 설명을 하겠습니다.

검사 절차는 σ를 알고 있을 때와 같고 단지 z-score 대신 t-score 를 사용한다는 것만 다릅니다.

예를 들어 커피가계를 운영하는 상점 주인이 매장에 두 명 이상이 같이 온 손님들이 커피를 주문하고 마시고 나가는데 소요되는 시간을 평균 30 분 이상이라고 생각을 하고 손님 15 명을 대상으로 조사를 해 보았습니다.

```
32.1507,  37.3355,  20.9646,  33.4487,  31.2751,  24.7692,  28.2656,  31.3705,
44.3136, 41.0777, 24.6005, 42.1397, 32.9016, 29.7478, 32.859
```

1) Hypothesis 설정: upper tail

상점 주인이 알고 싶은 것은 30 분 이상이 맞는지 그렇지 않는지를 알고 싶은 것이기 때문에 null hypothesis 는 30 분 미만입니다.

H_0: μ ≤ 30 분

H_1: μ > 30 분

즉 만약 null hypothesis 를 reject 할 사유가 나타나지 못한다면 상점 주인의 주장은 받아 드리기 어렵습니다.

2) Level of significance α = 0.05
3) Critical value

$$t_{\bar{x}} > t_{0.95}(= 1.7613)$$

4) test statistic

$$t_{\bar{x}} = \frac{32.48 - 30}{\frac{6.623}{\sqrt{15}}} = 1.451$$

5) Null hypothesis reject 검사

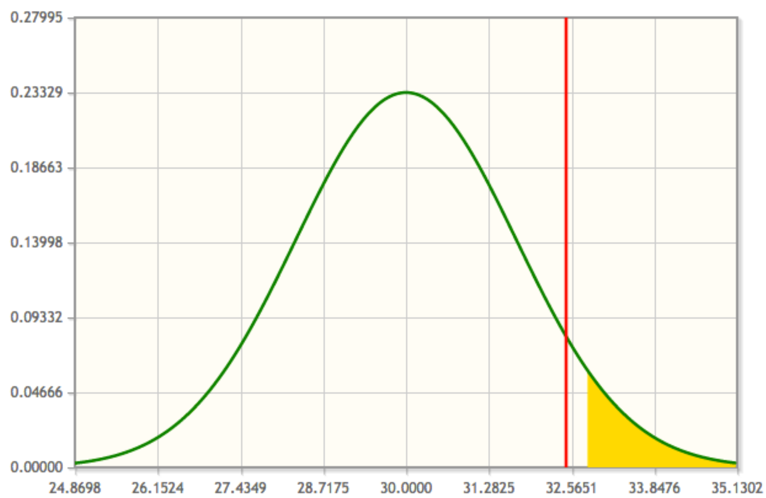
Upper tail 의 검사의 경우 $t_{\bar{x}} > t_{\alpha}$조건이 설립니 되어야 되는데 1.451 이 critical value 인 1.7613 보다 작기 때문에 null hypothesis 를 reject 할 충분한 근거를 만들지 못합니다.

6) Confidential interval : 33.012 이하

32.48 는 33.012 보다 작기 때문에 이 또한 null hypothesis 를 reject 하지 못함을 알려주는 값입니다.

7) p-value: 0.084

p-value 값도 0.05 보다 크기 때문에 null hypothesis 를 reject 하지 못함을 알려 줍니다.




```
jMath.prototype.test(alpha, type, H0, N)
```

여기서 type 의 값은 -1: lower tail, 0: two tail, 1: upper tail 이고, N 은 population 크기입니다. 여기서 N 은 표준편차 수식 8.9 인 correction factor 계산을 위한 것으로 모를 경우 넣지 않아도 됩니다.

```
var obj = jMath('32.1507 37.3355 20.9646 33.4487 31.2751 24.7692 28.2656 31.3705
44.3136 41.0777 24.6005 42.1397 32.9016 29.7478 32.8590');
obj.test(0.05, 1, 30);

Object {H0: 30,alpha: 0.05,sampleSize: 15,type: 1,mean: 32.48132000…}
H0: 30
alpha: 0.05
ci: [ undefined, 33.012339687397564]
mean: 32.481320000000004
meanStd: 1.7102835335865052
pvalue: 0.08443234092932861
sampleSize: 15
sigma: 6.623899642873308
testscore: 1.450823767680565
tscores: [undefined, 1.7613101151015698]
type: 1
```

Two-tail 검사의 예로 초등학생의 하루 평균 통화시간을 조사하니 10 분으로 나타나는데 이를 확인 하기 위해서는 two-tail hypothesis test 를 수행 합니다.

1) Hypothesis:

H_0: μ = 10 분

H_1: μ ≠ 10 분

20 명을 대상으로 조사를 했을 경우 다음과 같은 결과가 나났습니다.

6.3909	6.3105	13.9528	11.2905	11.2158	17.6772	13.1663	12.7912
18.3508	8.7821	14.7865	15.3404	11.0251	12.8627	7.3366	7.4082
12.4195	14.8890	22.3420	9.3324				

2) Level of significance: 0.05

3) Critical value

$$t_{0.025}(-2.093) < t_{\bar{x}} < t_{0.975}(= 2.093)$$

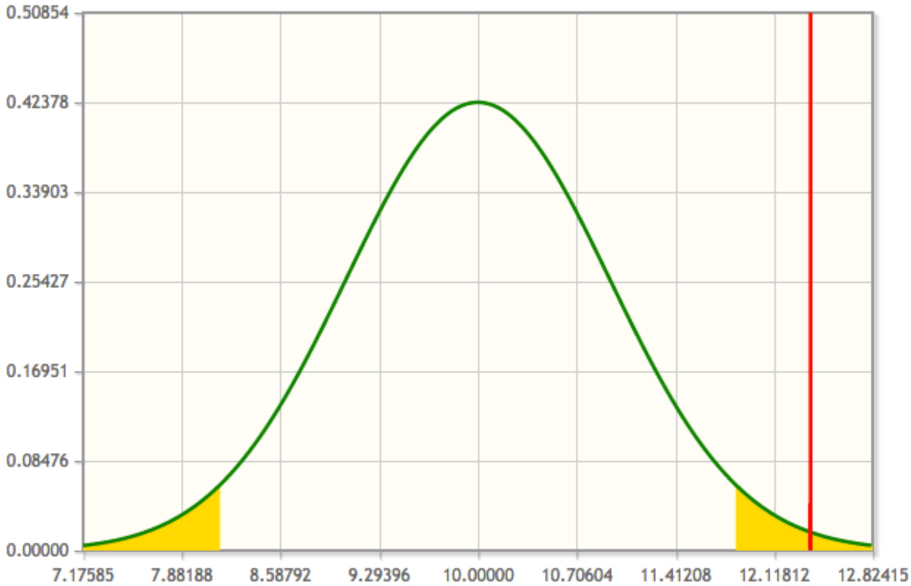
4) test statistic

$$t_{\bar{x}} = \frac{12.38 - 10}{\frac{4.21}{\sqrt{20}}} = 2.53$$

5) Null hypothesis reject 검사 : test statistic 은 critical value 범위 밖에 있기 때문에 null hypothesis 를 reject 할 수 있습니다. 즉 초등학생 평균 통화 시간은 10 분이 아니라는 자료 근거 됩니다.

6) CI: 8.029 ~ 11.970

7) p-value: 0.0203




```
obj = jMath('6.3909 6.3105 13.9528 11.2905 11.2158 17.6772 13.1663 12.7912 18.3508 8.7821
14.7865 15.3404 11.0251 12.8627 7.3366 7.4082 12.4195 14.8890 22.3420 9.3324');
Object {H0: 10,alpha: 0.05,sampleSize: 20,type: 0,mean:12.383525…}
H0: 30
alpha: 0.05
ci: [8.029619348865888,11.970380651134112]
mean: 12.383525
meanStd: 1.7102835335865052
pvalue: 0.9414037444402124
sampleSize: 20
sigma: 4.210085533682309
testscore: 2.531883917051252
tscores: [ -2.093024021596347,2.093024021596347]
type: 0
```

4. Hypothesis test for Proportions

가설을 평균으로 하지 않고 비율값으로 하여 검사를 하는 것으로 binomial distribution 이 가설로 사용되는 비율 π에 sample 크기 n 일 경우 다음의 조건을 만족을 하게 된다면

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

이 조건이 성립되면 normal distribution 으로 취급할 수 있기 때문에 hypothesis test 를 z-score 를 이용해서 처리할 수 있습니다.

$$z_p = \frac{p - \pi_{H_0}}{\sqrt{\frac{\pi_{H_0}(1 - \pi_{H_0})}{n}}} \tag{9.4}$$

여기서 n 은 sample 크기, p 는 sample proportion 이고 π는 null hypothesis 에 해당하는 값입니다.

예를 들어 황사와 미세먼지로 도시내에서 공원과 같은 야외지에서 마스크 착용률을 지난해에 조사한 결과 30%가 착용한 것으로 조사 되었습니다. 마스크 착용율이 높아 졌는지 조사하기 위해서 올해 다시 한 결과 300 명 중 107 명이 착용한 것으로 조사가 되는데 이것이 마스크 착용률이 높았다는 가설이 증명하기 위한 절차는 다음과 같습니다. 우선 z-score 사용을 위해 normal distribution 이 적용되는 가를 확인하면

300x0.3=90, 300x0.7=210

이기 때문에 hypothesis test 를 위해 z-score 를 사용할 수 있습니다.

1) Hypothesis:

$$H_0: \pi \leq 0.3$$
$$H_1: \pi > 0.3$$

향상성을 테스트 하기 때문에 upper one-tail test 를 하면 됩니다.

2) Level of significance : α = 0.05

3) Critical value:

$$z_{\bar{x}} > z_{0.95}(= 1.65)$$

4) Test statistic: p 는 107/300 = 0.3567

$$z_p = \frac{107/300 - 0.3}{\sqrt{\frac{0.3(1-0.3)}{300}}} = 2.1418$$

5) Null Hypothesis reject 검사:

Critical value 보다 test statistic 이 크기 때문에 null hypothesis 를 reject 할 근거가 되고 이는 마스크 착용률이 높아졌다는 근거가 될 수 있습니다.

6) Confidence Interval

p > 0.3434 이면 null hypothesis 가 reject 될 수 있음을 알수 있습니다.

7) p-value:

1 – P(p<0.3567) = 0.016 으로 0.05 보다 작기 때문에 null hypothesis 는 reject 됩니다.

```
jMath.stat.test_p( alpha, type, p, H0, n, N )
```

여기서 type 의 값은 -1: lower tail, 0: two tail, 1: upper tail 이고, n 은 sample 크기, N 은 population 크기입니다. 여기서 N 은 표준편차 수식 8.9 인 correction factor 계산을 위한 것으로 알지 못하면 넣지 않아도 됩니다.

```
obj = jMath.stat.test_p(0.05,1,107/300, 0.3,300);
Object {H0: 0.3, alpha: 0.05, popSigma: 0.458257569495584, sampleSize: 300, popSize: undefi
ned…}
H0: 0.3
alpha: 0.05
ci: [ undefined, 0.34351873640016206 ]
meanStd: 0.026457513110645904
popSigma: 0.458257569495584
popSize: undefined
pvalue: 0.016104842836855204
sampleSize: 300
testscore: 2.141798680385622
type: 1
zscores: [ undefined, 1.6448536269514724 ]
```

5. Hypothesis test for the population variances

제품을 생산하거나 정해진 시간내에 작업을 완료하는 과정에서 오류는 항시 발생을 합니다. 그래서 허용되는 오류 범위를 정하여 이 범위를 벗어나지 않기를 바람니다. 이러한 오류 범위의 검증은 지금까지 학습한 평균을 갖고 검증하는 방식이 아닌 variance 에 의한 검증이 요구 됩니다. 이 검사를 위한 전제 조건은 조사 대상인 population 은 normal distribution 이어야 합니다.

고기집에서 1 인분에 200g 을 판매를 하지만 실제로 손님에게 제공되는 무게는 이 보다 크거나 작을 경우가 발생되기 때문에 이를 확인하기 위해서 판매된 고기의 무게를 30 번 측정하여 10g 정도의 표준편차보다 작은가 검사를 하려고 합니다.

```
207 191 192 203 192 210 210 208 194 206
203 214 209 210 201 201 211 192 193 194
201 211 210 192 207 211 213 206 206 197
```

단위: g

1) Hypothesis 설정

$$H_0: \sigma^2 \geq 100$$
$$H_1: \sigma^2 < 100$$

오차값이 작은가를 알고 싶은 경우이므로, 편차가 100 인가 아닌가가 아니라 100 보다 작은가를 알고 싶은 것입니다.

2) Significance level: 0.05
3) Test statistic

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \tag{9.5}$$

먼저 sample 표준 편차값이 7.58g 으로 이를 수식 13.1 에 적용을 하게 되면

$$\chi^2 = \frac{(30-1)57.45}{100} = 16.667$$

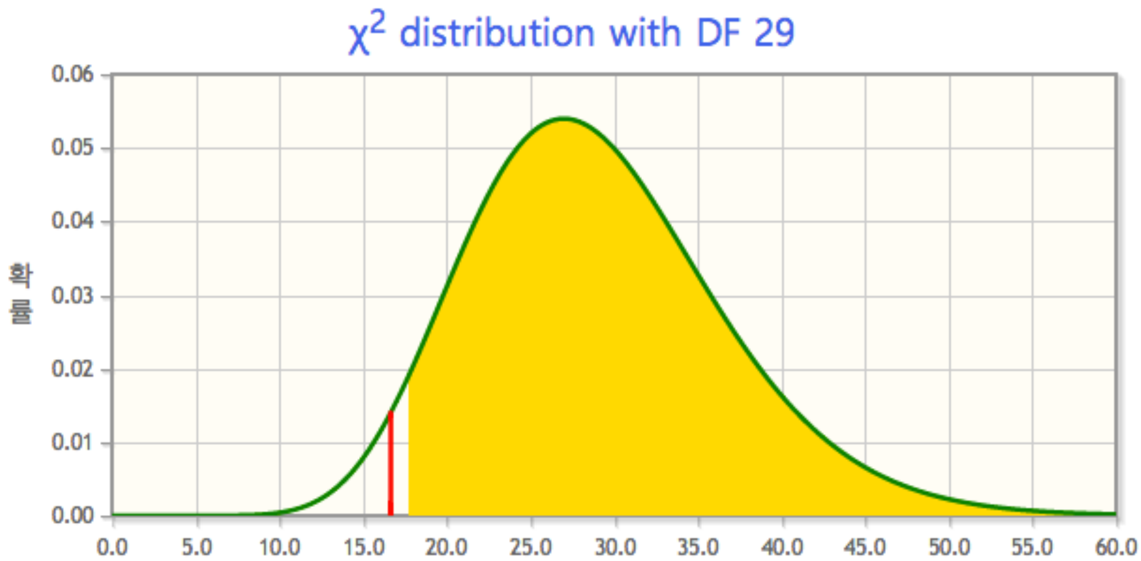
4) Critical value

Degree of freedom 은 30-1 로 29 이고 테스트가 lower tail 에 관한 것이기 때문에 $\chi^2_{0.05}$는 17.71 이 됩니다.

5) Null hypothesis reject 검사

Lower tail 에 대한 검사이기 때문에 Critical value 는 test statistic 보다 작아야 null hypothesis 를 reject 하지 못합니다. 하지만 test statistic 은 Critical value 보다 작기 때문에 null hypothesis 를 reject 한 근거가 됩니다. 이를 명확하게 알 수 있는 것은 p-value 값 0.0328 로 0.05 보다 작아 표준편차가 10 보다 작다는 것에 타당성이 있어 판매되는 고기량의 차이는 10g 보다 작아 손해는 크지 않은 것으로 판단이 될 수 있습니다.

varsingle.html



노랑색 영역이 null hypothesis 를 지지할 95%영역이고 빨간색 라인이 test statistic 입니다.

다음으로 two-tail 검사는 기존과 같이 significance level 을 반반씩 나뉘어 영역을 설정을 하고 내부에 포함되면 null hypothesis 를 reject 하지 못하고 그렇지 않으면 reject 할 근거가 있는 것으로 예를 들어 튀김 닭을 판매하는 가계에서 너무 오래 튀기면 음식이 탈수 있고 너무 빨리 제공하면 닭이 들 익을수 있기 때문에 표준편차가 기대값보다 크다 작다가 아닌 같은가 아닌가가 더욱 중요합니다. 그래서 마구잡이로 30 개의 닭을 튀기는 시간을 측정을 해 기대 표준편차가 2 분인가를 확인하려 합니다.

```
28 31 29 28 29 27 28 32 32 30
27 28 29 31 27 27 28 30 31 30
29 30 28 31 28 30 28 29 30 31
```

단위: 분

1) Hypothesis 설정

$$H_0: \sigma^2 = 4$$
$$H_1: \sigma^2 \neq 4$$

2) Significance level: 0.05
3) Test statistic

Sample 의 편차값은 2.3 으로 수식 13.1 에 적용을 하면

$$\chi^2 = \frac{(30-1)2.3}{4} = 16.7$$

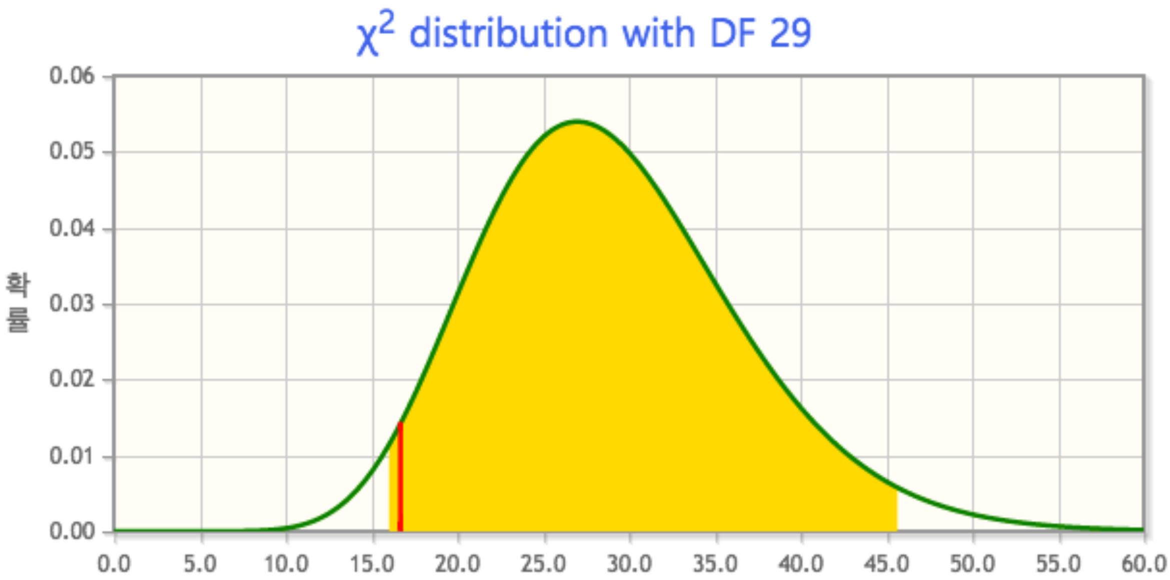
4) Critical value

Degree of freedom 은 29 이고 χ^2범위가 16.047 ~ 45.72 안에 있으면 됩니다.

5) Null hypothesis reject 검사

Test statistic 값은 Critical value 의 범위안에 있고 p-value 역시 0.0667 로 0.05 보다 크기 때문에 null hypothesis 를 reject 하지는 못합니다. 즉, 닭을 튀기는데 소요되는 표준편차가 2 분이 아니라는 것에 실패를 했습니다.

varsingle2.html



빨간색 선이 test statistic 으로 null hypothesis 를 지지할 95%영역인 노랑색 영역안에 표시가 되었기 때문에 null hypothesis 를 reject 하지 못했습니다.

```
jMath.prototype.test_v(alpha,type,H0)
jMath.stat.test_v(alpha,type,sigma2,H0,n);
```

type 은 lowertail(-1), twotail(0), uppertail(1)중 하나이고 H0 는 population variance, sigma2 는 sample variance 입니다. 마지막으로 n 은 sample 크기 입니다.

6. Type II Errors : β

Level of significance 인 α값은 null hypothesis 를 reject 하는 용으로 사용합니다. 여기서 발생되는 오류를 Type I error 라고 하여 null hypothesis 로 설정된 값이 참일 때 발생되는 오류로 즉 reject 을 하지 못하는데도 reject 한 것에 해당 합니다.

반대로 Type II error 는 null hypothesis 가 거짓인데도 null hypothesis 를 reject 하지 못한 경우입니다. 결과 α와 β는 반비례 관계로 α의 감소는 null hypothesis 를 reject 할 기회가 줄어 들기 때문에 β가 늘어 나게 됩니다. 이러한 α와 β를 줄이는 방법은 sample 크기를 늘려야 합니다.

예를 들어 출산을 앞둔 임산부가 스튜디어에 단계별 사진 촬영과 출산 후 돌잔치 까지 사진 촬영을 하는 서비스를 신청하는 비율이 30%라고 했을 때 임산부 200 명에게 물어 보았을 때 69 명이 서비스를 신청했다고 했을 때 30%이상인가를 확인하려고 합니다. Type I error 인 α를 0.05 로 하여 계산을 할 경우 Hypothesis 는 다음과 같습니다.

$$H_0: \pi \le 0.3$$
$$H_1: \pi > 0.3$$

Test-statistic 값을 계산을 하면

$$p = \frac{69}{200} = 0.345$$
$$\sigma = \sqrt{\frac{0.3(1-0.3)}{200}} = 0.0324$$
$$z_p = \frac{p - 0.3}{\sigma} = 1.3887$$

1.3887 로 critical value 인 1.65 보다 작고 p-value 또한 0.0824 로 null hypothesis 를 reject 하기에 충분하지 못합니다. 하지만 만약에 신청 비율값이 사실 30%이 아니라 38%일 경우 30%에 α에 해당되는 값

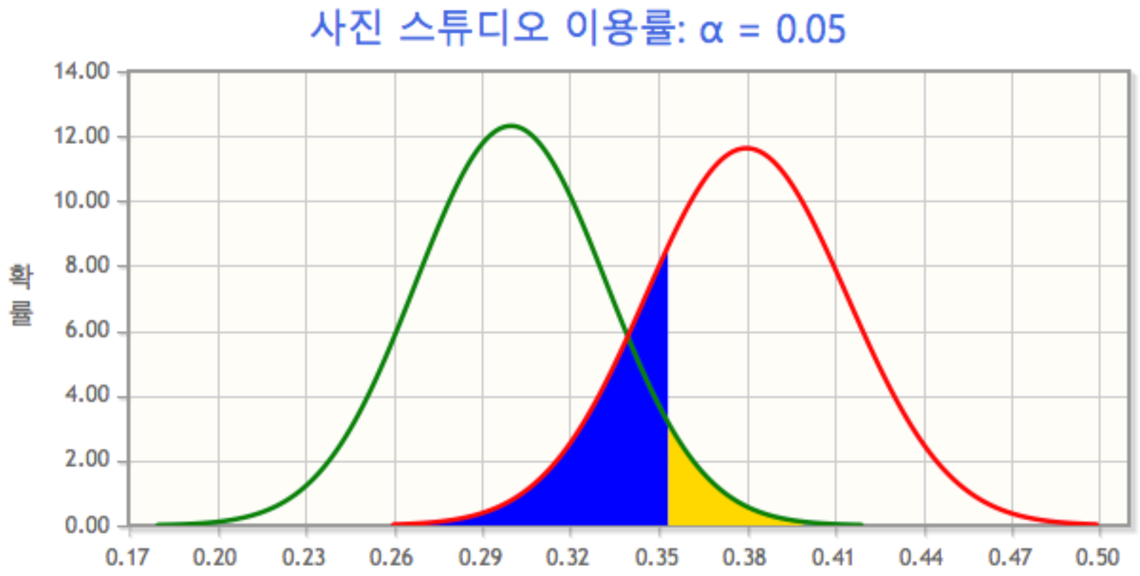
jMath.stat.norminv(0.95, 0.3, 0.0324) = 0.3533

으로 적용하여 38%에 대한 z-score 을 계산을 했을 때 결과는 다음과 같습니다.

$$z_p = \frac{p_\alpha - 0.38}{\sqrt{\frac{0.38(1-0.38)}{200}}} = \frac{0.3533 - 0.38}{\sqrt{\frac{0.38(1-0.38)}{200}}} = -0.7779$$

$$\beta \;=\; P(z < -0.7779) = 0.2183$$

chapter09/9_beta.html

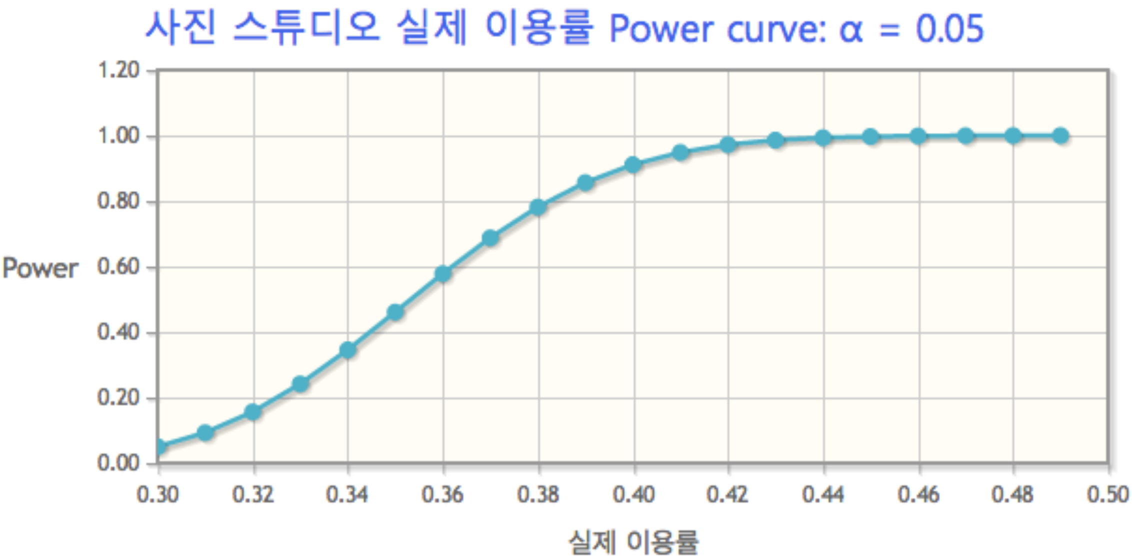


이 결과의 의미는 참 값이 0.38 인데 0.3 으로 조사를 했을 때 21.83%가 스튜디오 이용률이 30%이상이 아니라고 판단이 되게 합니다. 이 말의 의미는 반대로 0.38 이 참일 경우 78.17%는 정상적으로 0.3 대한 null hypothesis 를 reject 합니다. 이 값을 hypothesis test 의 power 라고 합니다.

$$Power = 1 - \beta \tag{9.6}$$

β값에 따라서 power 가 조절이 되는데 이러한 β값은 사진 촬영의 예에서 실제 이용률의 사실 값에 따라서 다르게 나타납니다.

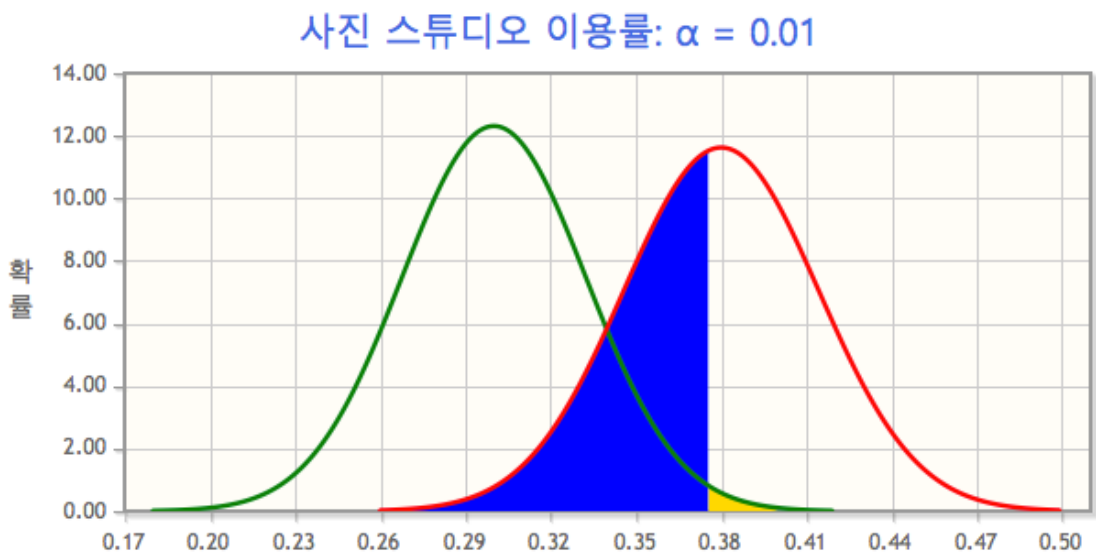
chapter09/9_power.html



그래프에서 보듯이 null hypothesis 로 사용된 0.3 에서 멀어 질수록 power 값이 증가하여 0.43 이후에는 1 로 완벽하게 분리되는 것을 알 수 있습니다. 이 말의 의미는 0.43 이후에는 값이 0.43 을 중심으로 분포된 값들 사이에 이용률 p 값들은 확실하게 null hypothesis 0.3 를 reject 하는데 실패하지 않는 다는 의미 입니다.

다음은 α와 β의 반비례 관계를 보여주는 것으로 9_beta2 예제에서 보듯이 α영역을 0.05 에서 0.01 로 감소는 β영역이 증가되는 것을 볼 수 있습니다.

chapter09/9_beta2.html



Power 값을 증가 시킬 수 있는 다른 방법은 sample 크기를 늘리는 것입니다. 이유는 sample 크기가 증가하면 null hypotheiss 를 나타내는 영역이 작아지기 때문입니다. 다음 예를 보시면 앞의 예제에서 sample 크기만 2 배로 늘렸습니다. 결과를 보시면 β 영역인 파랑색이 줄면서 power 가 증가 되는 것을 볼 수 있습니다.

chapter09/9_beta3.html