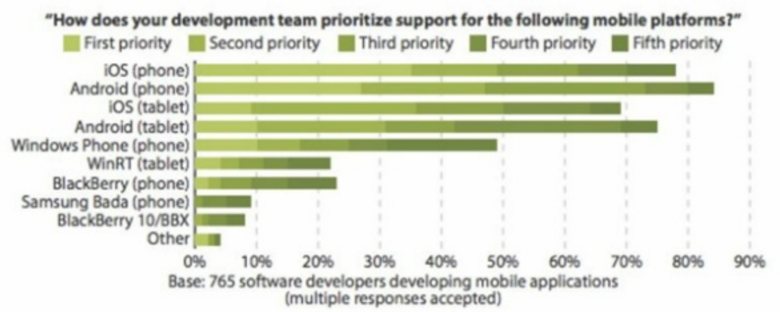
Chapter 10 Hypothesis test for comparing two populations

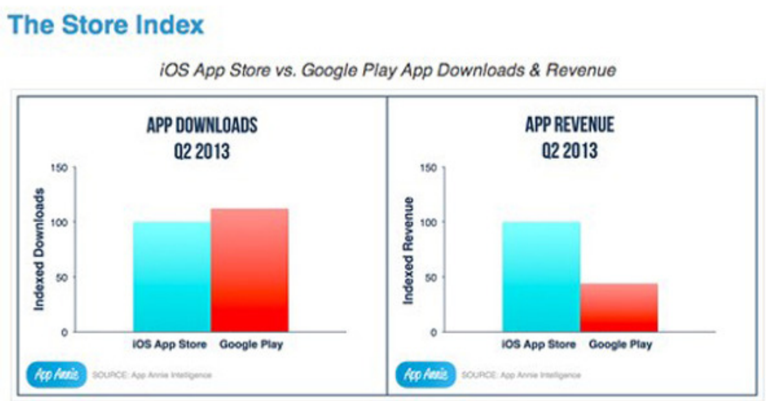
리서치회사인 포레스터의 발표에 따르면 개발자 765명을 대상으로 한 조사에서 새 어플리케이션 개발시 35%가 아이폰 용을 가장 먼저 내놓을 것이라 답한 반면 안드로이드라고 답한 사람은 27%에 불과했습니다.(기타 스마트폰용 OS 및 태블릿 기기 포함 설문)



어플 개발자 765명 대상. 어플 개발시 선호도 조사. 출처 : 포레스터

개발자들의 경우 안드로이드 폰의 압도적인 사용자 비중에도 불구하고 아이폰을 선호했는데요. 단순히 애플에 대한 사랑 때문이 아니라 실제로 iOS가 그들에게 많은 돈을 벌어주기 때문입니다.

안드로이드 사용자 비중이 아이폰 대비 4배 이상 높고, 어플 다운로드 수도 안드로이드가 더 높지만 개발자에게 돌아가는 수익 면에서는 오히려 iOS가 2~3배 가량 높기 때문입니다.(2013. 7. 31. 모바일 어플 리서치업체 앱애니 발표)



iOS 앱스토어와 구글플레이의 다운로드 수와 매출액 비교. 출처 : 앱애니

하나 N 소설 2013/08/30

많은 종류의 스마트폰이 나왔지만 현재 승자는 Apple 의 iPhone 과 Google 의 Android 입니다. 이러한 시장에 성공에 큰 기여를 한 것은 Application 을 판매하는 store 를

통해 개발자들이 수 많은 프로그램을 거래 할 수 있기 때문입니다. 그런데 이 두 스마폰에서 발생되는 현상은 단순히 다운로드된 수로 비교를 하는 것보다 이로 부터 얻은 수익에 차에서 두드러진 차이가 나타납니다.

이와 같이 어떤 두 제품을 비교하기 위해서 비교하는 내용에 따라서 다른 해석이 나올 수 있어 어떤 내용을 비교할 것인가에 대한 고찰이 요구됩니다. 하지만 이러한 통계로 결과를 보여 줄 때 잘못된 점은 2013 년 02 월의 결과만을 갖고 모든것을 설명하고 있습니다. 이것에 대한 보다 정확한 정보를 위해서는 몇 달동안 발생된 App download 수와 수익을 갖고 비교를 해야 더욱 신뢰된 정보를 얻을 수 있습니다.

비교는 우리가 사는 일상 생활 속에서도 수 많이 이행됩니다. 예를 들어 마트에서 같은 종류의 물건을 선택하거나, 여행을 가기 위해서 각 여행정보를 비교 분석합니다. 만일 비교 대상이 없다면 단순할 수 도 있지만, 비교가 있기 때문에 문제점을 알 수 있고 더 좋은 결과를 만들 수 있는 기능도 작용합니다.

비교의 사고를 직감적으로 하는것이 아닌 통계적 사고로 판단하는 방법을 이번 장에서 알아 보도록 하겠습니다.

두 개의 population 의 비교의 타당성을 검증하는 예들은 다음과 같은 것들이 있습니다.

- 두 업체가 만든 냉방기에 전력 소비량을 비교하여 두 업체의 차이점이 있는가 조사
- 두 개의 영화의 평가 점수중 한 영화가 다른 영화에 비해서 평가가 좋은가 조사
- 같은 제품들을 파는 두개의 온라인 스토어에 가격이 같은가 다른가 조사
- 네일샾 상점 주인이 온라인 홍보한 결과 전과 후에 매출액 증가 조사
- 앱카드 사용자에 남성의 비율과 여성의 비율의 차이가 있는가 조사
- 전달과 현재달의 카드사용량의 비교

이러한 비교를 통해서 어떤 요인이 차이를 만드는가를 알아내는 수단으로 활용됩니다. 예를 들어 펩시에서 실시한 blind test 가 대표적인 예입니다. 한쪽 group 에게는 펩시와 코카콜라를 알려주고 맛에 대한 선호도를 물어보고 다른 group 에는 알지 못한 상태에서 음료를 마시게 하여 두 음료의 선호도를 조사하는 방식입니다. 결과 두 group 에 선호도

차이가 없이 코카콜라를 선호한다면 펩시는 문제가 되지만, 차이가 있어 어떤 음료인지 모르는 상태에서 펩시를 선호했다면 펩시콜라의 맛을 인정 받게 됩니다. 실제 결과들을 보면 모르는 상태에서는 펩시가 코카콜라보다 선호도 조사에서 앞서지만, 그렇지 않은 경우 제품의 브랜드로 인해서 실제로 코카콜라를 선호했습니다. 결과 기업은 돈을 사용하는데 음식의 맛을 높이는 것을 중요하지만 브랜드를 높이기 위한 광고 및 마케팅 전략이 더 중요하다는 사실을 알려 줍니다.

두 population 들을 비교하기 위한 방법은 앞장에서 한 population 을 조사할 때와 같이 population 표준편차를 알고 모를 때 방법에 따른 계산 방법이 다르고, 두 population 에 sample 이 의존성이 있는가 없는가에 따른 계산 방법이 다르게 됩니다. 예를 들어 어느 한 상점 주인이 온라인 홍보 전후로 매출액 증가를 조사하는 경우는 전과 후의 매출액은 서로 연관성이 있기 때문에 두 데이타는 의존성이 존재합니다. 하지만 두 업체가 만든 냉방기의 전력 소비량의 비교의 경우는 두 제품의 전력 소비량은 서로 연관성이 없기 때문에 독립적인 데이타 입니다.

두 population 들의 비교 검사에 특징 중 하나는 null hypothesis 로 사용되는 두 population 의 평균은 알 필요가 없습니다. Population 의 평균의 차를 0 으로 가정하고 level of significance 에 따라서 one-tail 과 two-tail 로 설정하여 비교를 하게 됩니다.

1. 두 population 들의 평균 비교:

- 독립 sample
- population 표준편차를 알고 있음

설명을 돕기 위해서 두 업체가 만든 냉방기의 전력 소비량 비교의 경우로 설명을 드리겠습니다.

1) 두 업체의 생산된 냉방기 몇개씩 선출하여 sample 을 만들게 됩니다.
2) Sample 크기가 30 개 이상이면 sample 평균의 분포는 normal distribution 으로 형성이 됩니다. 그럼 1)에서 얻어는 두개의 sample 평균을 각각 $\mu_{\overline{x}_1}$과 $\mu_{\overline{x}_2}$로 값을 이용하고 sample 평균 분포의 표준 편차는 각각 $\sigma_{\overline{x}_1}$과 $\sigma_{\overline{x}_2}$됩니다.
3) 비교 검증을 위해 사용될 분포는 sample 평균의 차이에 대한 분포로 $N(\overline{x}_1 - \overline{x}_2|\mu_{\overline{x}_1-\overline{x}_2}, \sigma_{\overline{x}_1-\overline{x}_2})$대한 조사를 하면 됩니다.

즉 3)번 내용을 보시면 한개의 population 을 검증하는 것과 같은 절차로 테스트가 가능함을 알 수 있습니다. 이 계산을 위한 평균과 표준편차는 다음과 같습니다.

$$\mu_{\overline{x}_1-\overline{x}_2} = \mu_{\overline{x}_1} - \mu_{\overline{x}_2} \tag{10.1}$$

$$\sigma_{\overline{x}_1-\overline{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}} \tag{10.2}$$

한 제품의 30 개의 sample 의 평균$\mu_{\overline{x}_1}$를 36.4kWh 이고 다른 제품의 35 개의 sample 평균 $\mu_{\overline{x}_2}$가 38.5kWh 로 측정이 되었고 각 회사가 주장하는 표준 편차가 $\sigma_{\overline{x}_1}$는 4kWh 와$\sigma_{\overline{x}_2}$는 5kWh 라고 했을 때 두 제품에 차이가 있는가를 검사하는 과정은 수식 10.1 과 10.2 를 이용하여 한개의 population 에 검사방식과 동일하게 하면 됩니다.

1) Hypothesis 설정

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

같은지 같지 않은지를 판단하는 것이기 때문에 null hypothesis 는 같다는 것이고 alternative hypothesis 는 업체의 성능에는 차이가 있다는 것입니다.

2) Significance level: 0.05 (experiment-wide error rate)

3) Critical value: $-1.96 < z_{\bar{x}} < 1.96$

4) Test statistic

$$N(\bar{x}_1 - \bar{x}_2 | \mu_{\bar{x}_1-\bar{x}_2}, \sigma_{\bar{x}_1-\bar{x}_2})$$

Population 에 표준편차를 알고 있다는 가정이기 때문에 z-score 를 이용합니다.

$$z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1-\bar{x}_2}} \tag{10.3}$$

이번 예제는 null hypothesis 에 값이 0 이기 때문에 population 평균의 차는 제거가 됩니다. 예제의 값을 수식 10.3 에 적용을 하면 다음과 같습니다.

$$\frac{36.4 - 38.5}{\sqrt{\frac{16}{30} + \frac{25}{35}}} = -1.88$$

5) null hypothesis reject 검사

α가 0.05 이고 two-tail 검사를 했기 때문에 test statistic 값의 절대 값은 $|z_{0.025}|$보다 커야 합니다. 이 값은 1.96 으로 test statistic 값의 절대 값은 1.88 으로 작기 때문에 null hypothesis 를 reject 할 충분한 근거가 없게 됩니다. 즉, 두 제품의 차이가 있다는 근거가 없습니다.

6) Confidence Interval

$$UCL_{\bar{x}_1-\bar{x}_2} = z_{1-\alpha/2}\sigma_{\bar{x}_1-\bar{x}_2}$$
$$LCL_{\bar{x}_1-\bar{x}_2} = z_{\alpha/2}\sigma_{\bar{x}_1-\bar{x}_2} \tag{10.4}$$

10.4 에 적용을 하게 되면 CI 는 -2.1892 ~ 2.1892 는 $\bar{x}_1 - \bar{x}_2$ 인 -2.1 을 포함하고 있기 때문에 null hypothesis 를 reject 하지 못합니다.

7) p-value

$$2 \times P(z_{\bar{x}} < -1.88) = 0.06$$

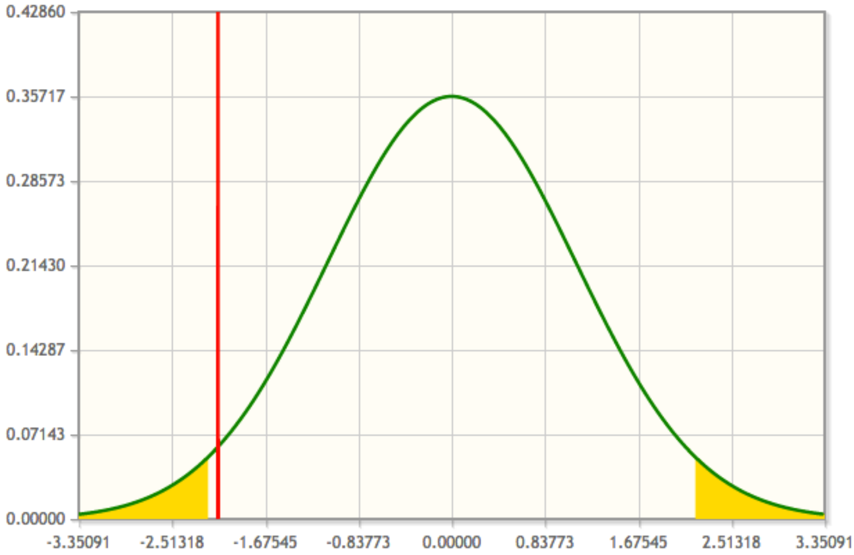
이 값 역시 α값 0.05 보다 크기 때문에 null hypothesis 를 reject 하지는 못합니다.

jMath 사용법

```
jMath.stat.compare( alpha, type, s1, s2, H0)
```

- alpha: significant level
- type: 1 upper tail, 0 two tail, -1 lower tail
- s1 과 s2: sample 에 대한 정보: { mean: , sigma: , n: }. sigma 는 population 표준편차값이고 n 은 sample 크기 입니다.
- H0: 차에 대한 null hypothesis 값으로 입력하지 않으면 기본값 0 입니다.

```
jMath.stat.compare( 0.05, 0, { mean: 36.4, sigma: 4, n:30}, {mean:38.5, sigma:5, n:35})
Object {H0: 0, alpha: 0.05, popSigma: Array[2], type: 0, mean: -2.1000000000000014…}
H0: 0
alpha: 0.05
ci: [ -2.1892183982741873, 2.1892183982741873 ]
mean: -2.1000000000000014
meanStd: 1.1169686869465265
popSigma: [ 4, 5 ]
pvalue: 0.06009601546360288
samples: Array[2]
testscore: -1.8800885150512159
type: 0
zscores: [-1.9599639845400547, 1.9599639845400547]
```


0.42860
0.35717
0.28573
0.21430
0.14287
0.07143
0.00000
-3.35091
-2.51318
-1.67545
-0.83773
0.00000
0.83773
1.67545
2.51318
3.35091

2. 두 population 들의 평균 비교

- 독립 sample
- population 표준 편차 알지 못함

Sample 들만들 갖고 hypothesis test 를 해야 되기 때문에 Student's t-distribution 을 이용해 z-score 가 아닌 t-score 를 사용 해야 합니다.
이 방식 역시 sample 크기가 30 개 이상이거나 그렇지 않으면 population 의 distribution 이 normal distribution 이어야 합니다.

2.1. Population 편차가 같을 경우

검사를 위한 절차는 같고 차이점은 test statistic 인 t-score 를 계산하는 방식입니다.

$$t_{\overline{x}} = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{10.5}$$

Degree of freedom 은 n_1+n_2-2 가 됩니다. 여기서 s_p^2는 pooled variance 로 두 sample 의 편차의 weighted average 값입니다.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} \tag{10.6}$$

예를 들어서 두 개의 영화의 10 점 만점에 평가를 하여 한 영화가 다른 영화에 비교해서 좋은 평가를 받았는가를 알고 싶을 경우에 두개의 sample 정보는 다음과 같습니다.

	평균	표준편차	Sample 크기
영화 1	8.0	1.2	45
영화 2	7.4	1.8	40

1) Hypothesis 설정

$$H_0: \mu_1 - \mu_2 \le 0$$
$$H_1: \mu_1 - \mu_2 > 0$$

영화 1 이 영화 2 보다 높은 평가를 받았는가를 알기 위해서는 같다 같지 않다가 아니라 한쪽이 더 높게 나와야 합니다. 이를 위한 두 영화의 평점에 차가 0 보다 커야 합니다.

2) Significance level: 0.05

3) Critical value: $t_{\overline{x}} < 1.6634$

Degree of Freedom 45+40-2=83 에 대한 95%에 대한 t-score 보다 작게 나와야 null hypothesis 를 reject 하는데 충분한 사유가 되지 못합니다.

4) Test statistic

수식 10.5 를 적용한 결과

$$\frac{(8.0 - 7.4)}{\sqrt{\frac{44 \times 1.2^2 + 39 \times 1.8^2}{83}\left(\frac{1}{45} + \frac{1}{40}\right)}} = 1.826$$

5) null hypothesis reject 검사

Test statistic 값이 Critical value 보다 크기 때문에 null hypothesis 를 reject 할 근거가 되므로 이는 영화 1 이 영화 2 보다 평가가 좋다는 것을 알수 있습니다.

6) Confidence Interval

Two-tail 검사에 null hypothesis 를 만족하는 CI 는 다음과 같습니다.

$$UCL_{\overline{x}_1 - \overline{x}_2} = t_{1-\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
$$LCL_{\overline{x}_1 - \overline{x}_2} = t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{10.8}$$

하지만 이 경우는 one-tail 검사를 해야 하므로 UCL 만 알면 됩니다.

$$UCL_{\bar{x}_1-\bar{x}_2} = t_{1-\alpha}\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \tag{10.9}$$

수식 10.9 를 적용하여 UCL 을 계산을 하면 0.5465 로 두 영화의 차 0.6 보다 작기 때문에 null hypothesis 를 reject 할 근거가 됩니다.

7) p-value

$$P(t_{\bar{x}} > 1.826) = 0.0357$$

α값 0.05 보다 작기 때문에 이 또한 null hypothesis 를 reject 할 근거로 영화 1 이 영화 2 보다 좋은 평가를 받았다는 근거가 됩니다.

jMath 사용법

```
jMath.stat.compare_t( alpha, type, s1, s2, eqPvar, H0)
jMath.prototype.compare(alpha, type, sampleObj, style, H0 )
```

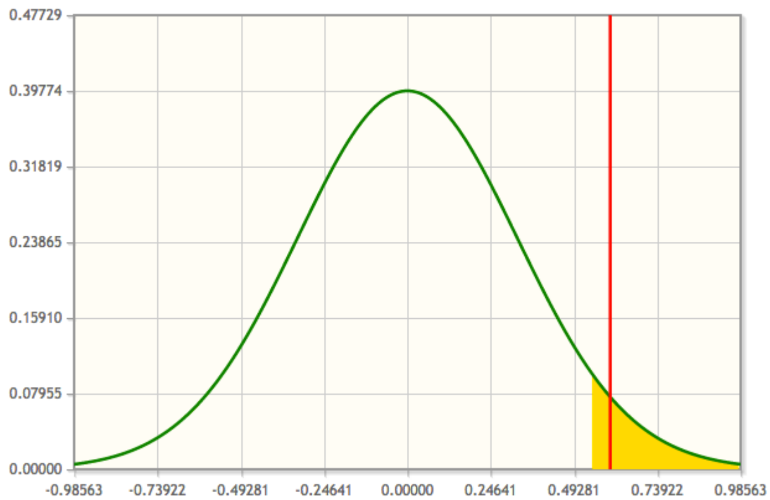
- alpha: Significant level
- type: 1 upper tail, 0 two tail, -1 lower tail
- s1 과 s2: sample 에 대한 정보: { mean: , sigma: , n: }. sigma 는 sample 표준편차이고 n 은 sample 크기 입니다.
- eqVar: 두 population 표준편차가 같은가
- style: 0: 두 population 표준편차가 다를 경우, 1: 같을 경우, 2: dependent sample 비교
- H0: 차에 대한 null hypothesis 값으로 입력하지 않으면 기본값 0 입니다.

첫번째 경우는 sample 에 대한 평균,표준편차값을 직접 넣지만 두 번째 경우는 jMath object 에 sample 정보가 있기 때문에 첫 번째 경우와 같이 직접 평균, 표준편차, sample 크기를 넣을 필요가 없습니다.

여기서는 jMath.stat.compare_t 만을 예를 들어 보겠습니다.

```
> jMath.stat.compare_t( 0.05, 1, { mean: 8.0, sigma: 1.2, n: 45 }, { mean: 7.4, sigma: 1.8, n: 40 }, true)
Object {H0: 0, alpha: 0.05, type: 1, eqPvar: true, samples: Array[2]…}
H0: 0
alpha: 0.05
ci: [undefined, 0.5465031518160691]
ci_diff: [-0.053456549343617876, 1.2534565493436172]
df: 83
eqPvar: true
mean: 0.5999999999999996
meanStd: 0.32854186801098056
pvalue: 0.035704034429678844
samples: Array[2]
testscore: 1.8262512587282982
tscores: [undefined, 1.663420114838465]
type: 1
```

여기서 mean 값이 두 영화의 평가 점수의 차로 ci 에 범위에 포함되지 않음을 알 수 있습니다.



2.2.Population 편차가 다를 경우

일반적인 경우에 우리는 population 에 정보를 갖고 있지 않기 때문에 이 방식이 population 비교에 적합합니다. 이 경우 test statistic 을 계산하는 방식과 degree of freedom 계산 방식에 차이가 있습니다.

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\hat{\sigma}_{\bar{x}_1}^2 + \hat{\sigma}_{\bar{x}_2}^2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \tag{10.10}$$

$$df = \left\lfloor \frac{\left(\hat{\sigma}_{\bar{x}_1}^2 + \hat{\sigma}_{\bar{x}_2}^2\right)^2}{\frac{\left(\hat{\sigma}_{\bar{x}_1}^2\right)^2}{n_1 - 1} + \frac{\left(\hat{\sigma}_{\bar{x}_2}^2\right)^2}{n_2 - 1}} \right\rfloor \tag{10.11}$$

예를 들어 네일 샵 주인이 온라인 홍보전후로 매주 일일 매출액을 비교하려고 합니다. Population 1 은 온라인 홍보전 일일 매출액이고 Population 2 는 온라인 홍보후 일일 매출액일 경우 데이타는 다음과 같습니다.

홍보전 일일 매출액
352 298 297 220 402 287 229 435 278 241 271 215 188 553 466 331 174 213 282 379 167 67 155 333 339 345 287 318 252 386
홍보후 일일 매출액
227 391 274 320 400 444 249 463 409 344 332 330 323 352 355 424 487 392 331 406 366 257 435 378 362 396 374 265 335 337

단위: 천원

1) Hypothesis test

매출액이 같은가 변화가 있는가를 확인 하기 위해서는 two-tail 검사를 하면 됩니다.

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 \neq 0$$

2) Significance level: 0.05

3) Critical value

$t_{\bar{x}}$의 범위를 구하기 위해서 우선 degree of freedom 을 알아야 합니다. Sample 들에 표준편차들을 계산을 하면 63.64 와 101.03 이고 이를 sample 평균에 standard error 로 전환하면 $63.64/\sqrt{30}$, $101.03/\sqrt{30}$ 으로 이것을 수식 10.11 에 적용을 하면 됩니다.

$$df = \left\lfloor \frac{(11.6190^2 + 18.4455^2)^2}{\frac{(11.6190^2)^2}{29} + \frac{(18.4455^2)^2}{29}} \right\rfloor = 48$$

이 degree of freedom 에 0.025~0.975 에 해당하는 t-score 는 -2.01 ~ 2.01 이 됩니다.

4) Test statistic

$\bar{x}_1$는 292 이고 $\bar{x}_2$값은 358.6 으로 이 null hypothesis 는 차이가 없는 것이기 때문에 10.10 에 분자로 두 sample 평균의 차를 사용하면 됩니다.

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\hat{\sigma}^2_{\bar{x}_1} + \hat{\sigma}^2_{\bar{x}_2}}} = \frac{(292 - 358.6)}{\sqrt{(11.6190^2 + 18.4455^2)}} = -3.0551$$

5) Null hypothesis reject

Test statistic 값이 critical value 범위 밖에 있기 때문에 두 값에 차이점이 있다는 것을 확실히 알 수 있습니다. 다시 말해, 이는 차이가 없다는 null hypothesis 를 reject 할 수 있는 근거로 사용이 될 수 있습니다.

6) CI

두 sample 평균값의 차는 -66.6 으로 null hypothesis 가 참이라고 했을 경우 CI 가 -43.83 ~48.83 사이인데 이 범위 밖에 있기 때문에 또한 null hypothesis 를 reject 할 수 있습니다.

7) p-value

two tail 이기 때문에 p-value 값은 $2P(t_{\bar{x}} < -3.0551) = 0.0037$으로 0.05 보다 훨씬 작은 값입니다. 정리하면 온라인 홍보는 매출에 변화를 주었다는 근거가 되면서 매출이 증가함을 알려주게 됩니다.

jMath 사용법

```
jMath.prototype.compare( alpha, type, sampleObj, style, H0)
```

```
var s1 = jMath('352 298 297 220 402 287 229 435 278 241 271 215 188 553 466 331 174 213
282 379 167 67  155 333 339 345 287 318 252 386');
var s2 = jMath('227 391 274 320 400 444 249 463 409 344 332 330 323 352 355 424 487 392
331 406 366 257 435 378 362 396 374 265 335 337');
console.log(s1.compare(0.05,0,s2,0))

Object {H0: 0, alpha: 0.05, type: 0, eqPvar: 0, samples: Array[2]…}
H0: 0
alpha: 0.05
ci: [-43.830055890993215,43.830055890993215]
ci_diff: [-110.43005589099323,-22.769944109006808]
df: 48
eqPvar: 0
mean: -66.60000000000002
meanStd: 21.799114186368787
pvalue: 0.0036660372447401873
samples: [{df: 29, mean: 292, n: 30, sigma: 101.02543210021169}, { df: 29, mean: 358.6,
n: 30, sigma: 63.638851720280606 }]
testscore: -3.0551700142772633
tscores: [-2.01063472195584,2.01063472195584]
type: 0
```

현재 까지의 예들은 두 population 이 차가 0 을 기준으로 계산을 했지만 차를 특정 값으로 하여 테스트를 할 수 있습니다.

예를 들어 돼지 고기집에서 고기를 팔때 초벌 구이를 해서 팔 때와 그렇지 않을 경우에 1 인당 100g 을 더 팔았다고 하여 결과가 맞는지 확인하려고 합니다.

	초벌구이 할 경우	초벌구이 안했을 경우
Sample 평균	370g	248g
Sample 표준편차	81g	103g
Sample 크기	23	25

1) Hypothesis test

증명하고 싶은 것은 인당 100g 이상을 더 팔았다는 것이기 때문에 upper one-tail 검사가 적합합니다.

$$H_0: \mu_1 - \mu_2 \le 100$$
$$H_1: \mu_1 - \mu_2 > 100$$

여기서 μ_1은 초벌구이 할 경우에 population 의 평균이고 μ_2는 초벌구이 안했을 경우에 1 인당 평균 판매되는 돼기 고기량입니다.

2) Significance level: 0.05

3) Critical value

Degree of freedom 을 수식 10.11 로 부터 계산을 하면 됩니다.

$$df = \left\lfloor \frac{(16.8897^2 + 20.6^2)^2}{\frac{(16.8887^2)^2}{22} + \frac{(20.6^2)^2}{24}} \right\rfloor = 44$$

상위 95%이하의 영역에 해당하는 t-score 값은 1.68 입니다.

4) test statistic

$$t_{\overline{x}} = \frac{(\overline{x}_1 - \overline{x}_2) - 100}{\sqrt{\hat{\sigma}^2_{\overline{x}_1} + \hat{\sigma}^2_{\overline{x}_2}}} = \frac{(370 - 248) - 100}{\sqrt{(16.8897^2 + 20.6^2)}} = 0.8259$$

5) null hypothesis test

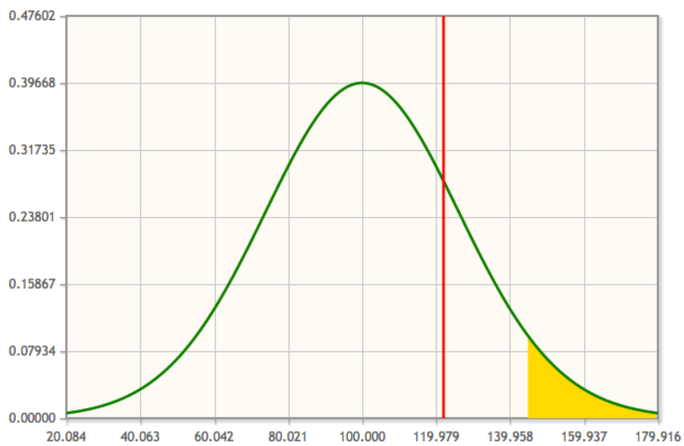
비록 평균의 차이는 122g 이지만 test statistic 값은 critical value 에 비해서 작기 때문에 null hypothesis 를 reject 할 충분한 근거가 되지 못합니다.

6) p-value

$$1 - P(t_{\overline{x}} < 0.8259) = 0.2067$$

0.05 보다 큰 값이기 때문에 null hypothesis 를 reject 하지 못합니다. Null hypothesis 를 reject 하기 위해서는 sample 크기는 늘리고 sample 표준편차가 더 작아져야 가능합니다. 그렇지 않으면 평균대비 표준편차가 크게 작용하기 때문에 비록 평균만 보았을 때 소비량이 100g 이상 증가 한거 같아 보이지만 실제는 그렇지 않은 것을 알 수 있습니다.

```
> jMath.stat.compare_t(0.05, 1,
    {n: 23, mean: 370, sigma: 81 }, {n:25, mean: 248, sigma: 103 }, false, 100 )
```



3. Dependent sample 들의 비교

지금까지의 모든 비교 조사 대상들은 서로 상관 관계가 없는 데이타들 입니다. 상관 관계가 있기 위해서는 같은 조사 대상들에게 각기 다른 조건으로 어떤 작업을 이행한 결과의 차이점을 알아보는 것으로 각 sample 의 값들은 서로 쌍으로 구성됩니다.

예를 들어 매장을 운영하는 의류 상점에서 온라인 판매도 같이 했을 경우 판매 경로에 따른 차별점이 있는가를 알아내기 위해 매일 혹은 매주 매장 판매 매출액과 온라인 판매 매출액을 나열하여 비교 할 수 있습니다.

비교 방법은 비록 두 개의 population 을 비교하는 것이지만 두 집단의 각각의 값의 차를 갖고 따지는 것이기 때문에 single population 을 갖고 계산하는 법과 동일합니다.

예를 들어 12 개월 동안 매장과 온라인 매출액의 차이를 조사한 결과 입니다.

	1 월	2 월	3 월	4 월	5 월	6 월	7 월	8 월	9 월	10 월	11 월	12 월
온라인	443	639	379	470	422	639	599	397	513	534	570	638
매장	427	680	314	475	453	606	618	375	496	478	559	584

단위: 만원

매출액을 눈으로 보면 일반적으로 온라인을 통한 매출액이 더 높게 나타나는 것을 알 수 있습니다. 이것이 통계를 통해서도 같은 결과가 나타나는지를 알아 보겠습니다.

1) Matched-pair difference

계산을 위해서 매 달 매출액 차를 계산 합니다.

1 월	2 월	3 월	4 월	5 월	6 월	7 월	8 월	9 월	10 월	11 월	12 월
16	-41	65	-5	-31	33	-19	22	17	56	11	54

단위: 만원

$$\overline{d} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{10.12}$$

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \overline{d}\right)^2}{n-1}} \tag{10.13}$$

이 예제에서 $\overline{d} = 14.83$이고 $s_d = 34.25$

2) Hypothesis test

온라인 매출액이 매장 매출액 보다 좋다는 주장을 증명하기 위한 upper tail 검사를 하면 됩니다.

$$H_0: \mu_d \leq 0$$
$$H_1: \mu_d > 0$$

3) Significance level: 0.05
4) Critical value:

Degree of freedom 은 11 이고 이에 해당하는 $t_{1-\alpha}$값은 1.796 이 됩니다.

5) Test statistic:

$$t_{\overline{d}} = \frac{\bar{d} - \mu_{d_{H_0}}}{\frac{s_d}{\sqrt{n}}} \tag{10.14}$$

수식 10.15 를 적용하여 test statistic 을 계산을 하면

$$t_{\overline{d}} = \frac{14.83 - 0}{\frac{34.25}{\sqrt{12}}} = 1.5$$

6) Null hypothesis reject

Critical value 가 test statistic 보다 크기 때문에 null hypothesis 를 reject 할 충분한 근거를 만들지 못하게 됩니다. 이 말은 온라인 판매가 매장 판매보다 좋다고 말을 하지 못합니다.

7) p-value

$$P(t_{1-\alpha} > 1.5) = 0.081$$

0.05 보다 크기 때문에 이 또한 null hypothesis 를 reject 할 충분한 사유가 되지 못함을 알려 줍니다.

이러한 검사 방식의 장점은 독립적인 sample 들의 비교시 전체 평균과 표준편차로 따지지만 이 방식은 각 sample 별 차를 갖고 하기 때문에 전체의 변화 정도 보다는 개개의 차이만을 집중하여 더욱 정확한 결과를 얻을 수 있습니다. 예를 들어 옷을 판매하는 경우 성수기와 비수기로 인한 월 마다 매출 차이점으로 인한 영향은 제거되고 각 월에 판매량 차이점으로 비교 하기 때문에 온라인 스토어와 실제 매장에서의 매출액 차이를 더욱 확실하게 알아 낼 수 있습니다.

jMath 사용법

```
jMath.prototype.compare(alpha,type,sampleObj,2)
```

```
var s1 = jMath('443 639 379 470 422 639 599 397 513 534 570 638');
var s2 = jMath('427 680 314 475 453 606 618 375 496 478 559 584');
var result = s1.compare(0.05,1,s2,2);
console.log(result);

Object {H0: 0, alpha: 0.05, sampleSize: 12, type: 1, df: 11…}
H0: 0
alpha: 0.05
ci: [undefined,17.756120040119647]
df: 11
mean: 14.833333333333334
meanStd: 9.887115378362415
pvalue: 0.08084466142827917
sampleSize: 12
sigma: 34.24997235123857
testscore: 1.5002690638965874
tscores: [undefined, 1.7958847814174657]
```

4. 독립 sample 의 Population proportion 들의 비교

예를 들어 스마트폰 이용 고객 중 남자와 여자의 앱카드 등록률의 차이점이 있는지 알고 싶을 때 남성의 등록 비율과 여성의 등록 비율을 비교하면 됩니다.

이러한 테스트를 하기 전에 반드시 normal distribution 으로 취급가능한가를 확인을 해야 합니다. Proportion 에 대한 검사는 binomial distribution 을 기본으로 하고 있지만 normal distribution 으로 취급되기 위한 조건만 성립되면 z-score 를 이용해서 검사를 할 수 있습니다.

$$np \geq 5 \text{ and } n(1-p) \geq 5$$

테스트를 위해서 confidence interval 로 조사를 하는 방식을 할 경우와 z-score 를 이용한 hypothesis test 를 할 경우와 다른 방식으로 표준 편차를 다루것에 차이점이 있습니다.

4.1. Confidence interval

Margin error 값으로 두 비율의 차의 범위를 알아 내는 방식입니다. 이를 위해서 비율 평균의 표준 편차값을 알아야 합니다.

$$\sigma_{p_1-p_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}} \tag{10.15}$$

여기서 π는 population 에 비율값이지만 실제로 알 수 없는 경우가 일반적이라 이 값 대신 sample 로 부터 얻은 값을 대신 사용하면 됩니다.

$$\hat{\sigma}_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \tag{10.16}$$

이 값으로 CI 는 다음과 같습니다.

$$UCL_{\bar{x}_1-\bar{x}_2} = (p_1 - p_2) + z_{1-\alpha/2}\hat{\sigma}_{p_1-p_2}$$
$$LCL_{\bar{x}_1-\bar{x}_2} = (p_1 - p_2) + z_{\alpha/2}\hat{\sigma}_{p_1-p_2} \quad (10.17)$$

예를 들어 스마트폰 사용자 남여 각각 300 명에게 앱카드 등록률을 조사한 결과 남성은 138 명 여성은 161 명이 등록을 했을 경우 두 비율의 차에 대한 CI 값을 수식 10.17 을 이용하여

$$\text{여성}: p_1 = \frac{161}{300} = 0.5366, \qquad \text{남성}: p_1 = \frac{138}{300} = 0.46$$
$$p_1 - p_2 = 0.07666$$

α가 0.05 일 때

$$UCL_{\bar{x}_1-\bar{x}_2} = 0.07666 + 1.96 \times 0.0407 = 0.156$$
$$LCL_{\bar{x}_1-\bar{x}_2} = 0.07666 - 1.96 \times 0.0407 = -0.031$$

결과를 보면 95%범위 내에 0 이 보함 되어 있기 때문에 남성과 여성의 앱카드 등록률의 차이가 있다는 근거를 충분히 제시를 못합니다.

다른 방식으로 hypothesis test 를 하게 되면 다음과 같습니다.

1) Hypothesis test 설정

$$H_0: \pi_1 - \pi_2 = 0$$
$$H_1: \pi_1 - \pi_2 \neq 0$$

2) Significance level: 0.05
3) Critical value: $Z_{0.025} \sim Z_{0.975} = -1.96 \sim 1.96$
4) Test statistic:

$$z_p = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)_{H_0}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (10.18)$$

여기서 $\hat{p}$는

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \tag{10.19}$$

수식 10.18 에 적용하여 계산을 하면

$$z_p = \frac{0.07666 - 0}{\sqrt{0.49833(1 - 0.49833)\left(\frac{1}{300} + \frac{1}{300}\right)}} = 1.8779$$

5) null hypothesis reject

test statistic 이 critical value 범위내에 존재하기 대문에 null hypothesis 를 reject 하지 못합니다. 즉, 남녀의 앱카드 등록비율의 차이가 있다는 것을 지지하지는 못하게 됩니다.

6) p-value

$$1 - 2P(z_p > 1.8779) = 0.06$$

p-value 역시 0.05 보다 크기 때문에 null hypothesis 를 reject 하지 못함을 알 수 있습니다.

jMath 사용법

```
jMath.stat.compare_p( alpha, type, s1, s2, H0)
```

- alpha: Significant level
- type: 1 upper tail, 0 two tail, -1 lower tail
- s1 과 s2: sample 에 대한 정보: { x:, n: }. x 는 성공 개수, n 은 sample 크기 입니다.
- H0: 차에 대한 null hypothesis 값으로 입력하지 않으면 기본값 0 입니다.

```
> Math.stat.compare_p(0.05, 0, { x: 161, n: 300 }, { x: 138, n: 300} )
Object {H0: 0, alpha: 0.05, type: 0, samples: Array[2]…}
H0: 0
alpha: 0.05
ci: [-0.08001475007582813, 0.08001475007582813]
ci_diff: [-0.0031125797629638075, 0.15644591309629702]
mean: 0.0766666666666666
meanStd: 0.040824602241150476
pvalue: 0.06038765391749168
samples: Array[2]
sigma: 0.040704445111705574
testscore: 1.8779525692325783
zscores: [-1.9599639845400547, 1.9599639845400547]
type: 0
```

5. 두 population 의 variance 들의 비교 검사

편차의 비교가 필요한 경우로는

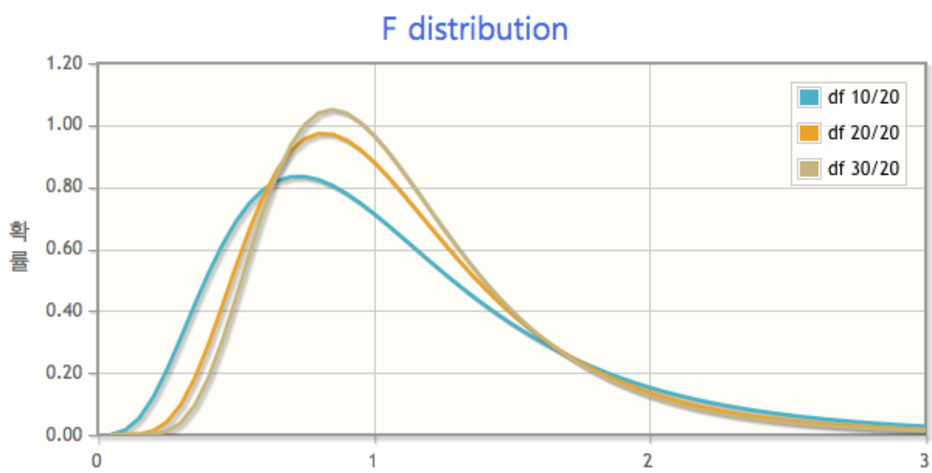
- 두 population 의 평균을 비교할 때 population 의 variance 는 같다는 가정하에 계산을 한 경우
- 다음 장에 배울 여러 group 을 비교분석 하는 ANOVA 검사를 위해서 비교하려는 population 들은 같은 편차를 갖고 있어야 비교를 할 수 있습니다.

이번에 학습할 내용은 이러한 검증을 위한 방법론을 배우겠습니다.

이 검사를 위해서 F-distribution(Variance ratio distribution)을 사용해야 합니다. 이 분포는 크기가 n_1 과 n_2 인 normal distribution 으로 된 두 독립된 sample 들로 부터 편차 비율대한 비율로 수식은 다음과 같습니다.

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \tag{10.20}$$

계산을 위해서 degree of freedom 은 알아야 하는데 분자는 n_1-1, 분모는 n_2-1 이 됩니다. 분포의 형태는 다음과 같습니다.



비교 검사를 위해서 상점에서 같은 물건을 판매하기 위해서 두개의 다른 업체의 전자 저울에 에 오차가 같은가를 알고 싶어 30 개씩 물건을 측정을 하여 표준 편차를 보았습니다.

	저울 A	저울 B
표준편차	10g	8.2g
Sample 크기	30	30

1) Hypothesis 설정

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

두 개의 전자저울의 성능이 같은가 같지 않은가를 알아보는 실험이기 때문에 two-tail 로 검사를 합니다.

주의 할 점은 항상 첫번째 sample 의 편차가 두번째 것보다 커야 합니다.

2) Significance level: 0.05
3) Test statistic

Population 의 표준편차 값이 같다고 가정했기 때문에 수식 10.20 은 다음과 같이 됩니다.

$$F = \frac{s_1^2}{s_2^2} \tag{10.21}$$

표준편차값을 10.21 에 적용을 하면

$$F = \frac{10^2}{8.2^2} = 1.4872$$

4) Critical value

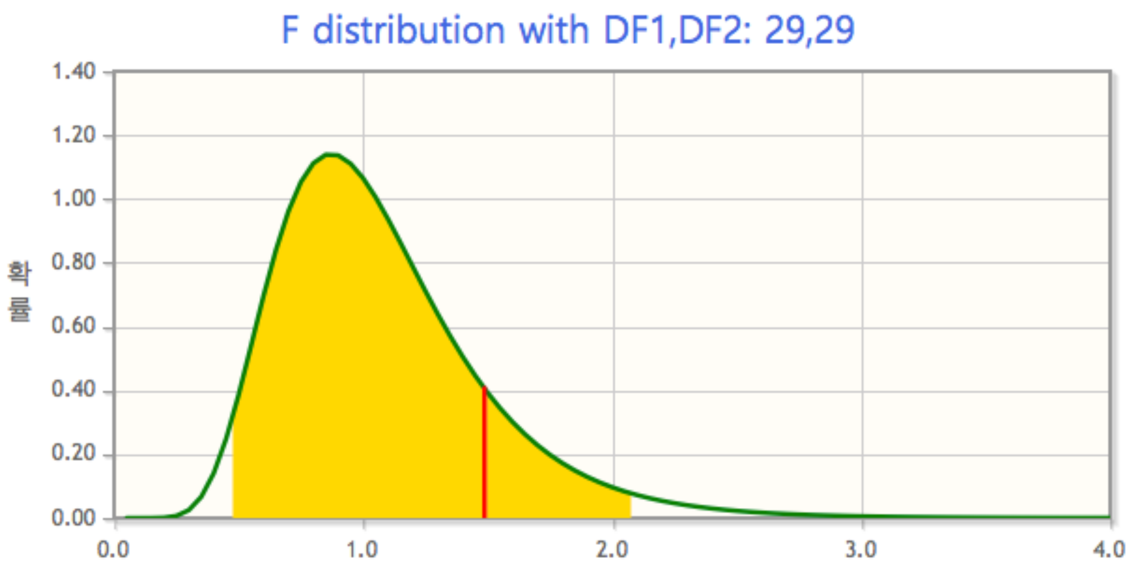
Two tail 이지만 항상 첫번째 sample 에 표준편차가 크기 때문에 반드시 F-distribution 에 오른쪽만 신경쓰면 되고 따라서 0.975 에 해당하는 F 값을 조사하면 됩니다. Degree of freedom 은 sample 크기에 1 을 뺀 값으로 여기서는 29,29 가 됩니다.

$$F_{\alpha/2} = 2.101$$

5) Null hypothesis reject 검사

Test statistic 이 critical value 보다 작기 때문에 null hypothesis reject 는 실패를 했습니다. 이러한 결과는 p-value 를 값이 0.291 로 0.05 보다 큰것으로 나타나 실패했을을 알 수 있습니다. 결과적으로 두개의 저울의 오차가 같지는 않지만 두 제품의 성능이 비슷하다고 판단을 할 수 있습니다.

chapter10/compare_v2.html



6) Confidence Interval

$$\frac{s_1^2}{s_2^2}\frac{1}{F_{1-\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}F_{1-\alpha/2} \tag{10.22}$$

수식 10.22 를 적용하면 95%에 해당하는 두 편차의 비율의 범위는

$$\left[\frac{10^2}{8.2^2}\frac{1}{2.101}, \frac{10^2}{8.2^2}2.101\right] = [0.7079, 3.1246]$$

1 을 포함 하고 있어 두 저울의 차이가 없을 수 있다는 것을 알려 줍니다.

마지막으로 one-tail 검사를 하는 방법을 설명하겠습니다. 예를 들어 가구 생산을 하는 업체에서 제작 방법을 바꾸어 생산시간의 향상 및 생산 시간의 차가 줄어 들었는지를 확인하려고 합니다.

	기존의 방식	새로 적용된 방식
표준편차	10 분	7 분
Sample 크기	30	30

1) Hypothesis 설정

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

새로 적용된 방식으로 생산시간의 차이가 줄어 들기를 바라기 때문에 upper tail 의 검사 방식을 채택합니다.

2) Significance level: 0.05
3) Test statistic

수식 10.21 를 적용하여

$$F = \frac{10^2}{6^2} = 2.78$$

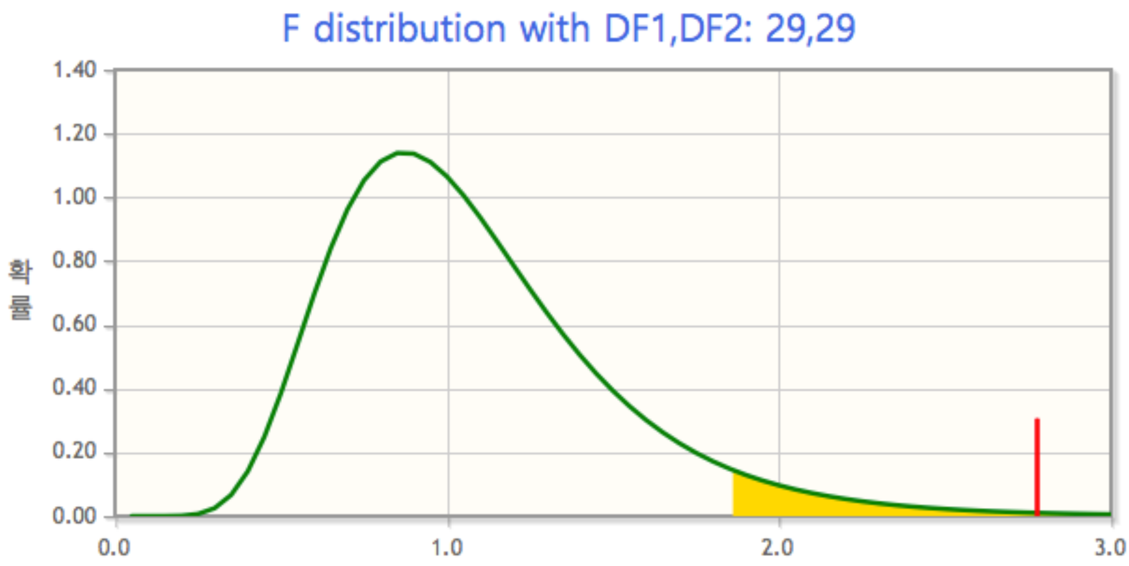
4) Critical value

Sample 크기가 둘 다 30 이기 때문에 F distribution 에 적용할 degree of freedom 은 29,29 입니다. Upper tail 검사 방식이기 때문에 0.95 에 해당하는 값은 1.86 입니다.

5) Null hypothesis reject 검사

Critical value 보다 test statistic 이 크기 때문에 null hypothesis reject 에 성공을 했습니다. 이는 새로 적용된 방식에 가구를 만들때 마다 생산시간에 차이를 줄이는데 성공했다고 판단이 됩니다.

chapter10/compare_v1.html



jMath.stat.compare_v(alpha,type,s1,s2)

type 은 twotail(0), uppertail(1)이어야 하고 s1 과 s2 는 표준편차값과 sample 크기를 갖고 있어야 합니다.

```
jMath.stat.compare_v(alpha, type, s1, s2)
```

- alpha: Significant level
- type: 1 upper tail, 0 two tail, -1 lower tail

https://github.com/handuck/jMath

- s1 과 s2: sample 에 대한 정보: { sigma:, n: }. Sigma 은 sample 표준편차

```
jMath.stat.compare_v(0.05, 'uppertail', {
        sigma: 10,
        n: 30
},{
        sigma: 6,
        n: 30
});
F: 2.7777777777777778
alpha: 0.05
Fcrit: 1.8608114343211721
pvalue: 0.0037733573253969643
samples: Array[2]
reverse: false
type: 1
```

여기서 reverse 는 만일 s2 의 표준편차가 s1 보다 크다면 뒤집어서 계산하기 때문에 true 로 표기해줍니다.