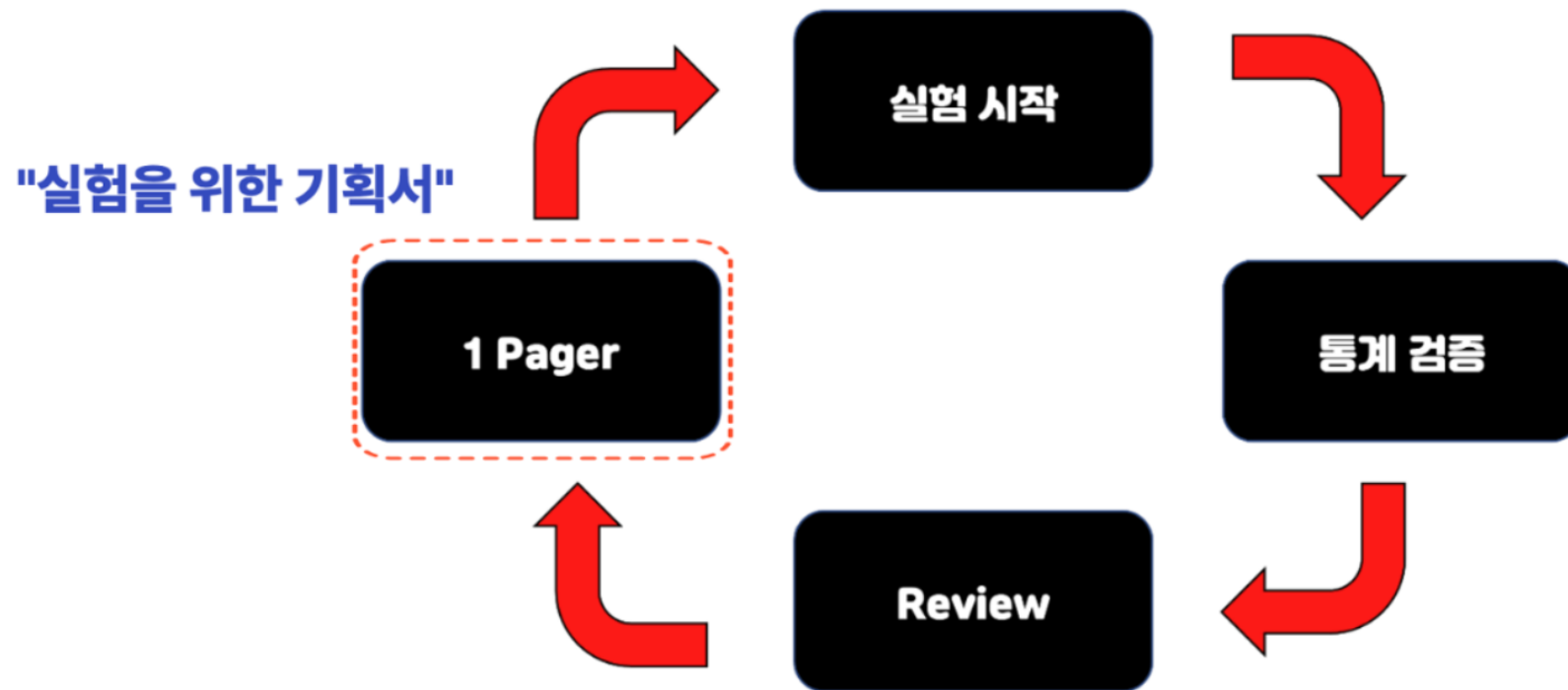




A/B Test 실험 계획

AB Testing Basic Course For PO & DA

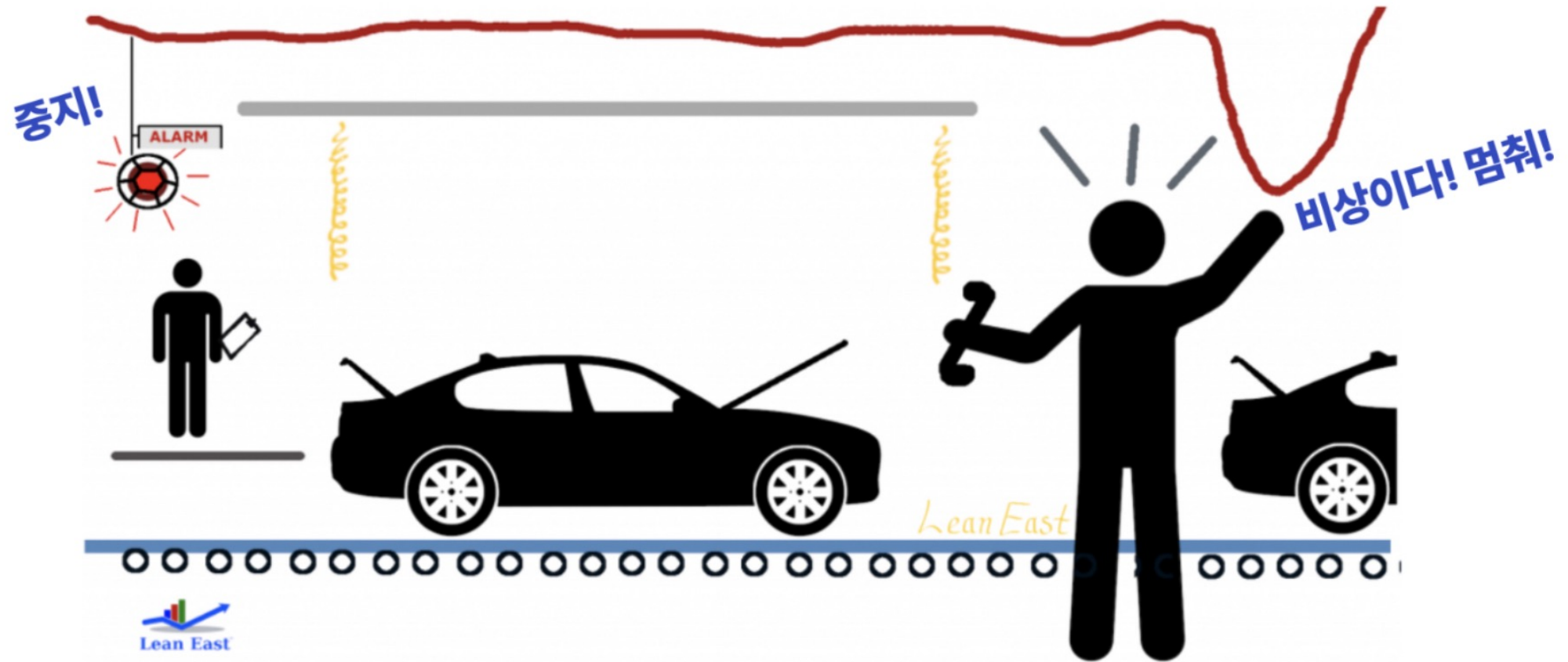
A/B Test 실험 계획서 작성법 - 실험 사이클



A/B Test 실험 계획서 작성법 - 1 Pager 설명

| 항목 | 설명 | 예시 |
|-------------------|--------------------------------------|--|
| Target User | 타겟 고객은 누구인가? | 마지막 결제구간에 있어서 결제 선택에 신중을 기하는 고객 |
| Observation | 데이터적으로 관찰되는 현상은 무엇인가? | 1) 전체 퍼널 Flow 구간중 결제 구간에서 구매까지 퍼널 전환율이 떨어진다는 것을 확인 2) 결제 버튼에서의 이탈률이 다른 Scroll 구간보다 높다는 것을 비율적으로 확인 |
| Problem Statement | 현상으로써 발생하는 문제는 무엇인가? | 고객이 받은 전체 할인 금액을 정확히 인지하지 못해, 좋은 금액으로 구매하고 있다는 느낌을 주지 못해 고객을 이탈시킨다. |
| Benchmarks | 적용하려는 기능과 유사하게 벤치마킹할 것이 있는가? | A사 모바일 앱의 결제 화면 |
| Hypothesis | 기능 개선을 위한 가설 설립 | 결제 페이지에서 고객이 받은 전체 할인 금액을 노출시키면 구매전환이 늘어날 것이다. |
| Experiments Group | 실험을 위한 집단 구분 | Group A(대조군) : 기존 결제 페이지(단순 할인된 가격 노출) Group B(실험군) : 기존 결제 페이지(단순 할인된 가격 노출) + 전체 할인받은 금액 노출 |
| Metric | 실험 성과를 평가할 수 있는 정량적 지표 설계 | Input Metrics(= 투입 지표) : 결제 페이지 유입 수, 최종 결제 금액 도달 수 Output Metrics(=산출 지표) : 전환율(결제 페이지 -> 결제 완료 페이지) |
| Trade-Off | 기능 도입으로 인해 발생하는 문제 상황 및 예의주시가 필요한 부분 | 동일한 카테고리 상품 중, 더 좋은 기능보다 더 저렴한 상품을 구매하여 전체 매출에 negative할 수 있음 |
| Andon | 실험 진행 중 심각한 문제 발생 여부를 파악할 수 있는 기준 정의 | 구매전환율이 3일 연속 하락 서비스의 주요 OEC에 심각한 부정적 영향 |

A/B Test 실험 계획서 작성법 - Andon이란?



Toyota's Andon cord allows each assembly line worker to stop the entire line if there is a problem.

집단 크기 결정(최소 샘플 사이즈)

샘플 사이즈를 계산하기 위해, 필요한 4가지의 정보

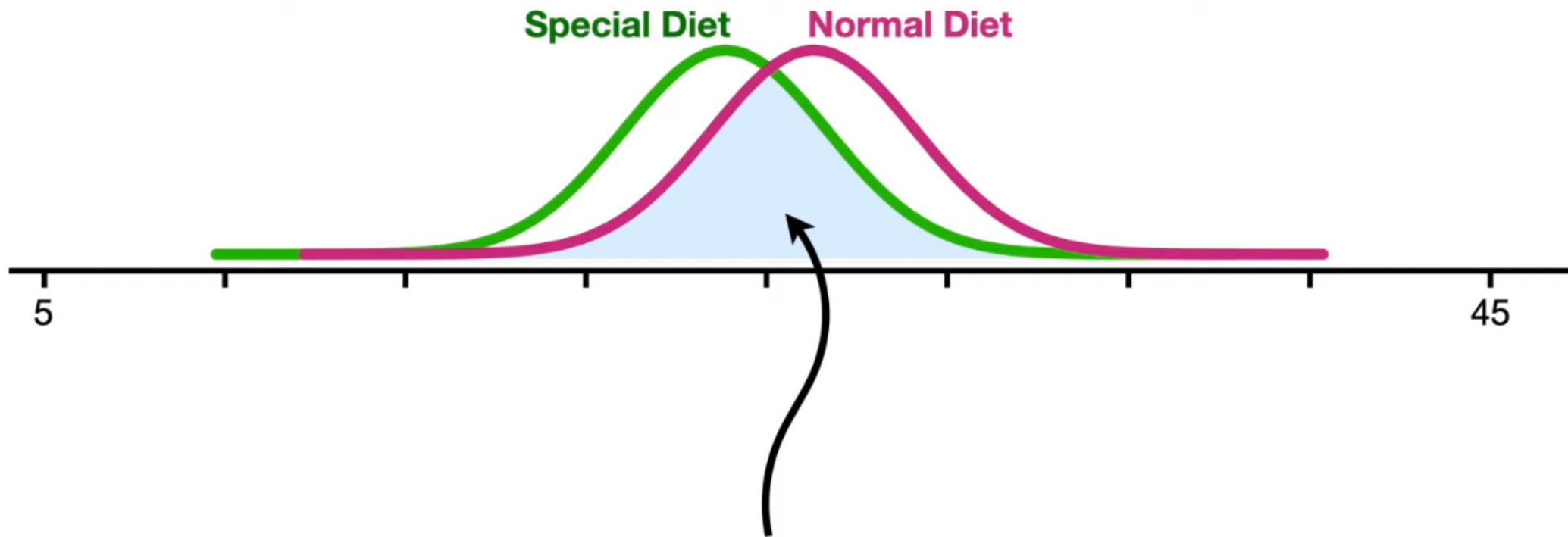
1. μ_0 (Base Value) : 실험 하고자 하는 지표의 과거의 일반적인 평균 데이터

2. μ_1 (Expected Value) : μ_0 (Base Value)를 얼마나 상승을 기대하는지에 대한 값

3. α (유의수준) : 귀무가설을 기각할 수 있는지 없는지를 판단하는 기준

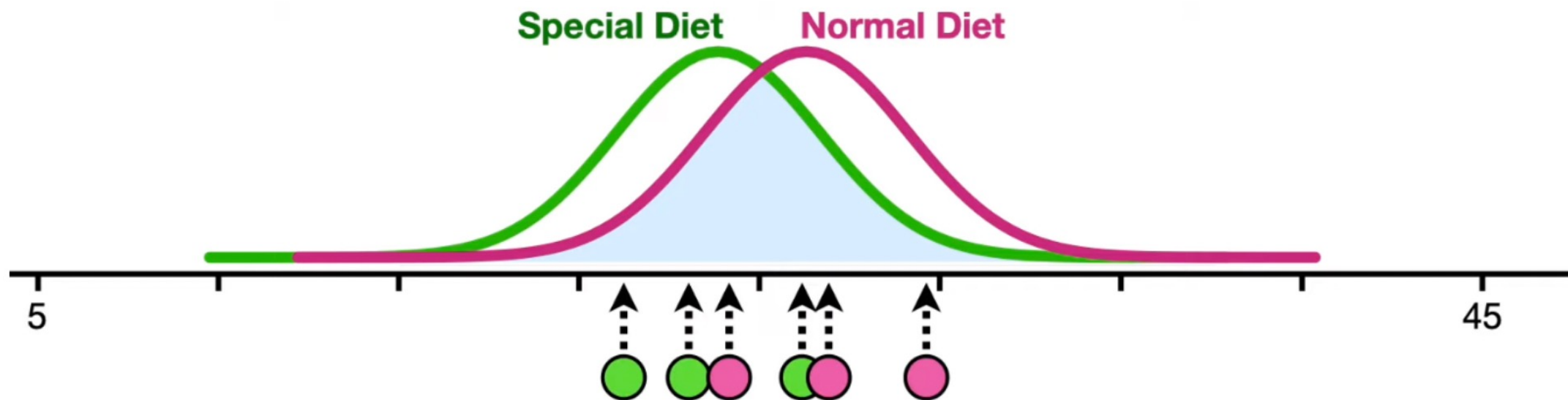
4. $1-\beta$ (검정력) : 귀무가설을 올바르게 기각할 확률

집단 크기 결정(최소 샘플 사이즈) - 1-Beta(검정력)에 대한 추가설명



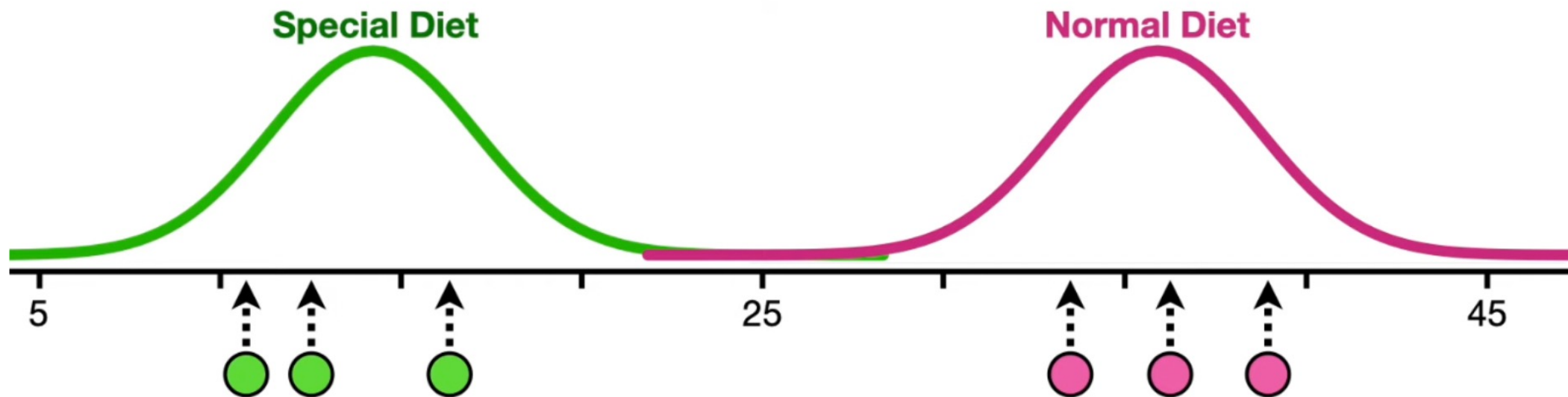
시각적으로만 봐도 두 분포는 많이 겹쳐있기 때문에
유의미한 차이를 보기 어렵게 느껴진다.

집단 크기 결정(최소 샘플 사이즈) - 1-Beta(검정력)에 대한 추가설명



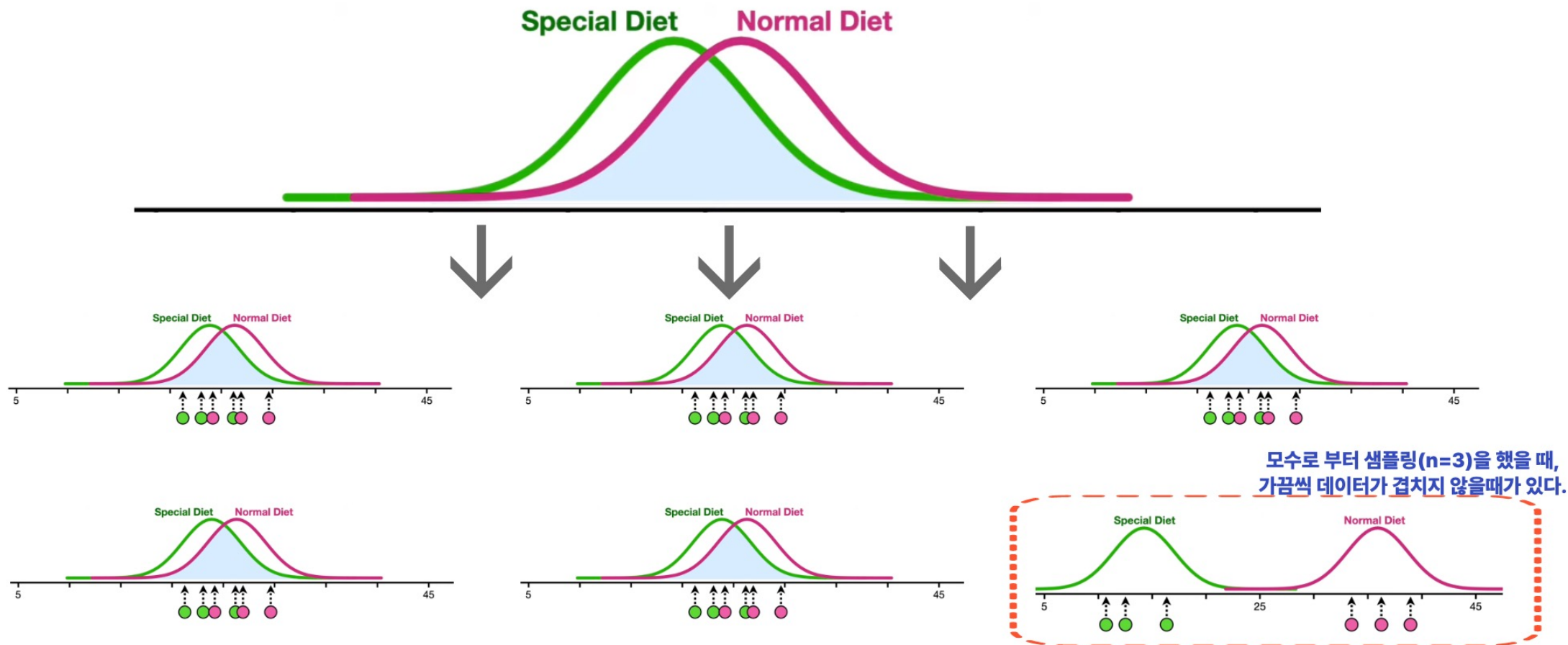
하지만 사실 이 분포의 실제 데이터가
작은 표본의 크기($n=3$)들로 이루어졌다면 좋지 않은 검정력을 가질 것이다.

집단 크기 결정(최소 샘플 사이즈) - 1-Beta(검정력)에 대한 추가설명



만약 작은 크기의 샘플들로 유의미한 지표가 나왔을 때는,
Type I Error (1종 오류) 발생시킬 가능성이 있다.

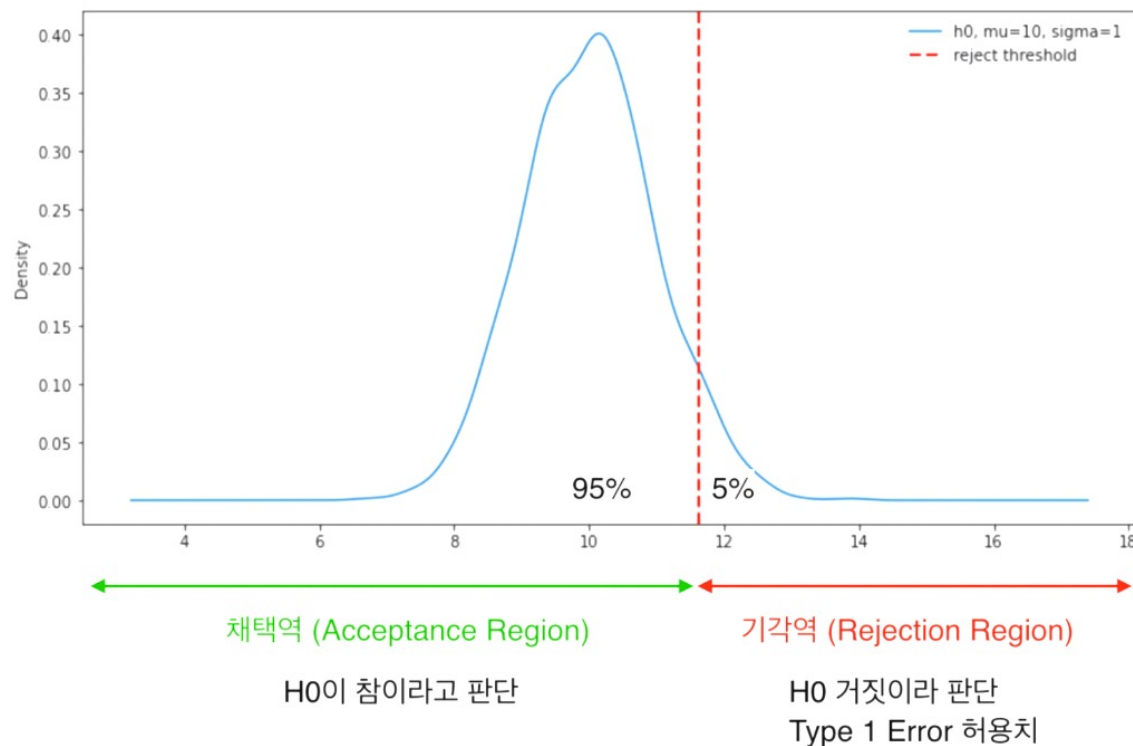
집단 크기 결정(최소 샘플 사이즈) - 1-Beta(검정력)에 대한 추가설명



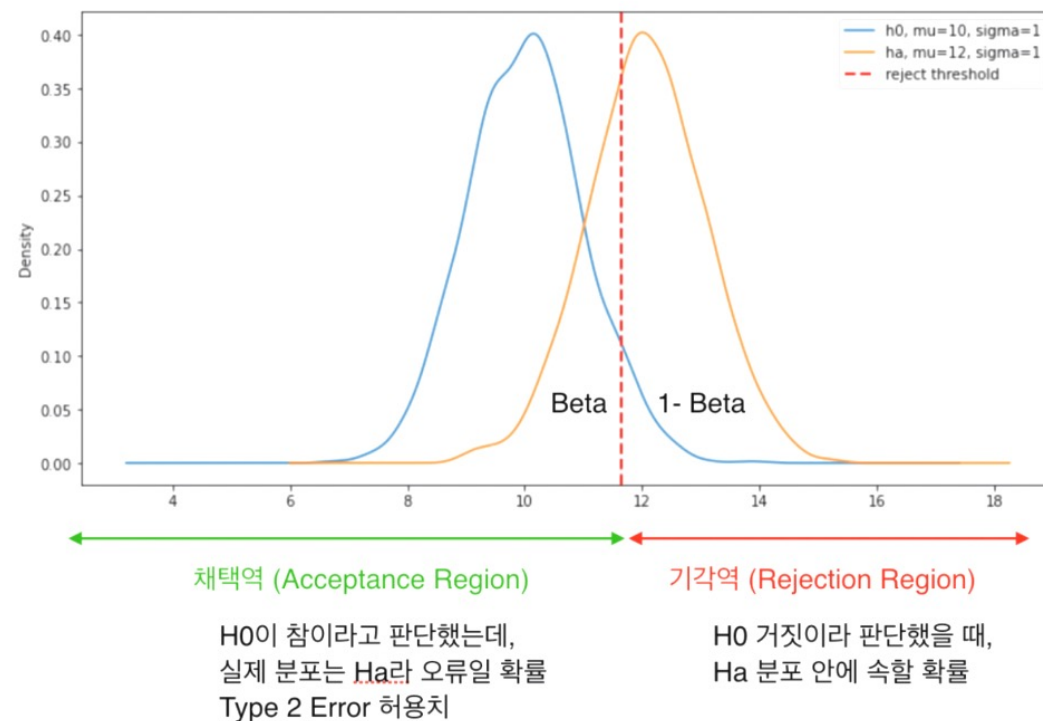
**(1) 두 분포간의 겹치는 부분이 많고, (2) 표본의 크기가 작다면
상대적으로 낮은 검정력을 갖는다는 것을 의미**

집단 크기 결정(최소 샘플 사이즈) - Beta를 결정되는 과정

(1) 기존 귀무가설에 해당하는 분포에 α (유의수준) 지정

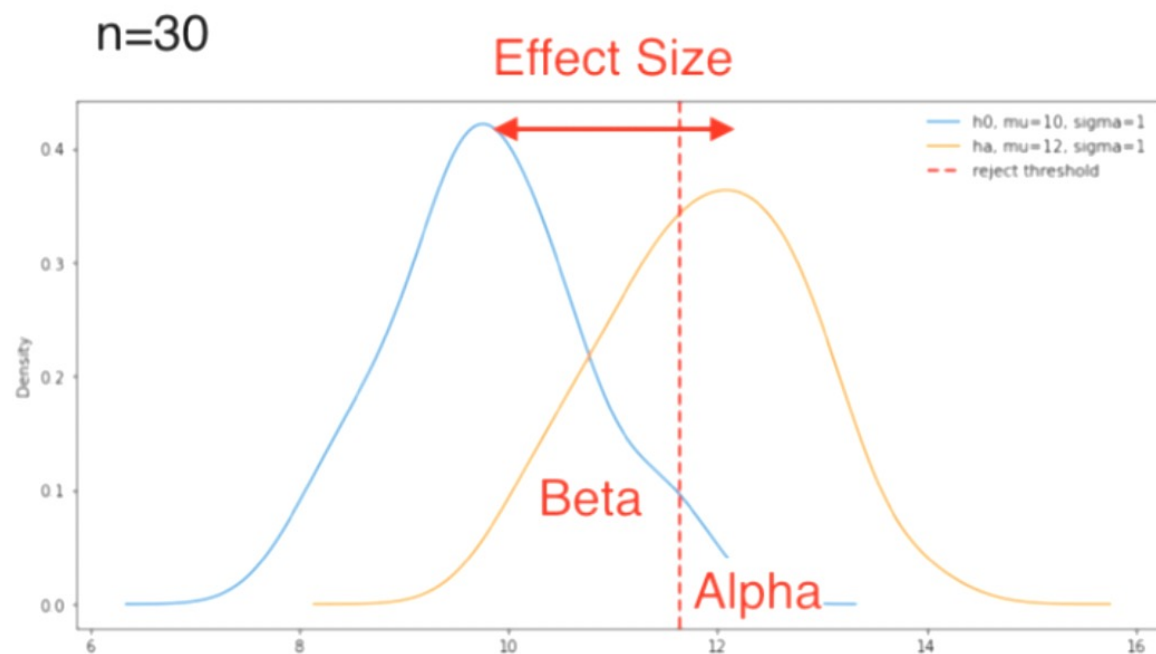


(2) 대립가설이 되는 분포를 올려놓기

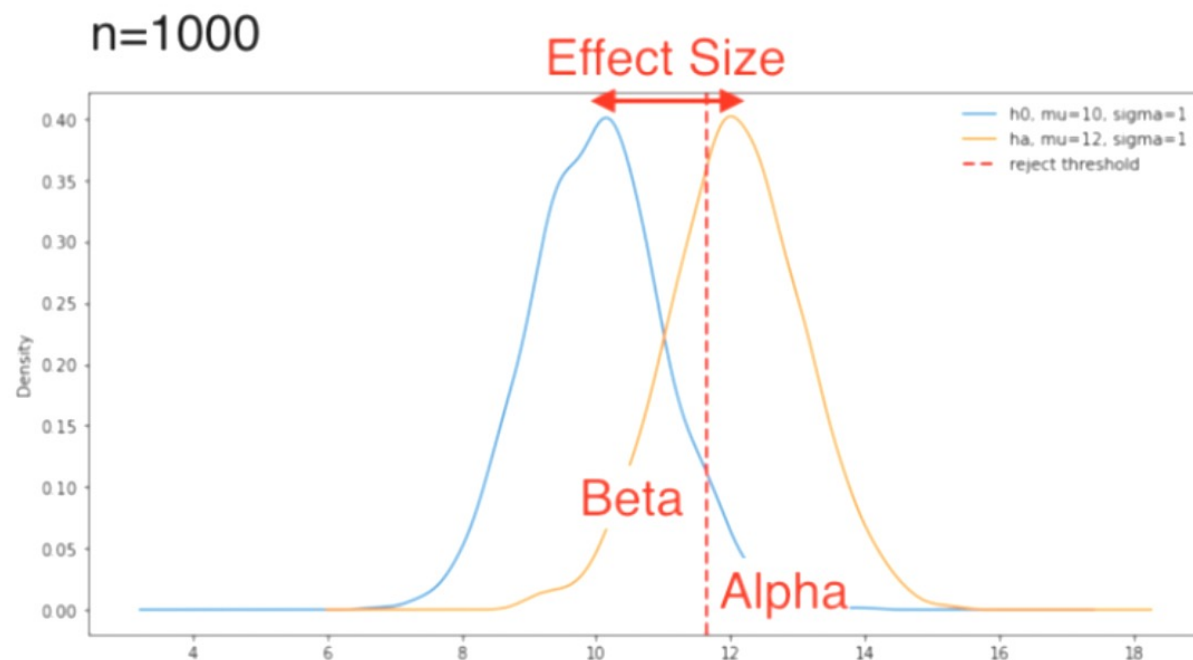


AB 테스트에서는 $1 - \beta$ 의 값인 power를 기준으로 관습적으로 0.8을 주로 사용한다.
여기서 의미는 귀무가설이 틀렸을 때 이를 정말로 틀리다고 할 확률을 80%로 놓고 실험을 진행하겠다는 뜻입니다.

집단 크기 결정(최소 샘플 사이즈) - 그림으로 보는 샘플 사이즈 결정



...



Sample Size를 구한다는 것은 Effect Size로 표시된 평균의 차이가 x로 주어졌을 때,
n을 조정하면서 Alpha로 표시된 구역의 넓이가 0.05이면서, Beta로 표시된 구역의 넓이가 0.2인 값을 찾는 과정

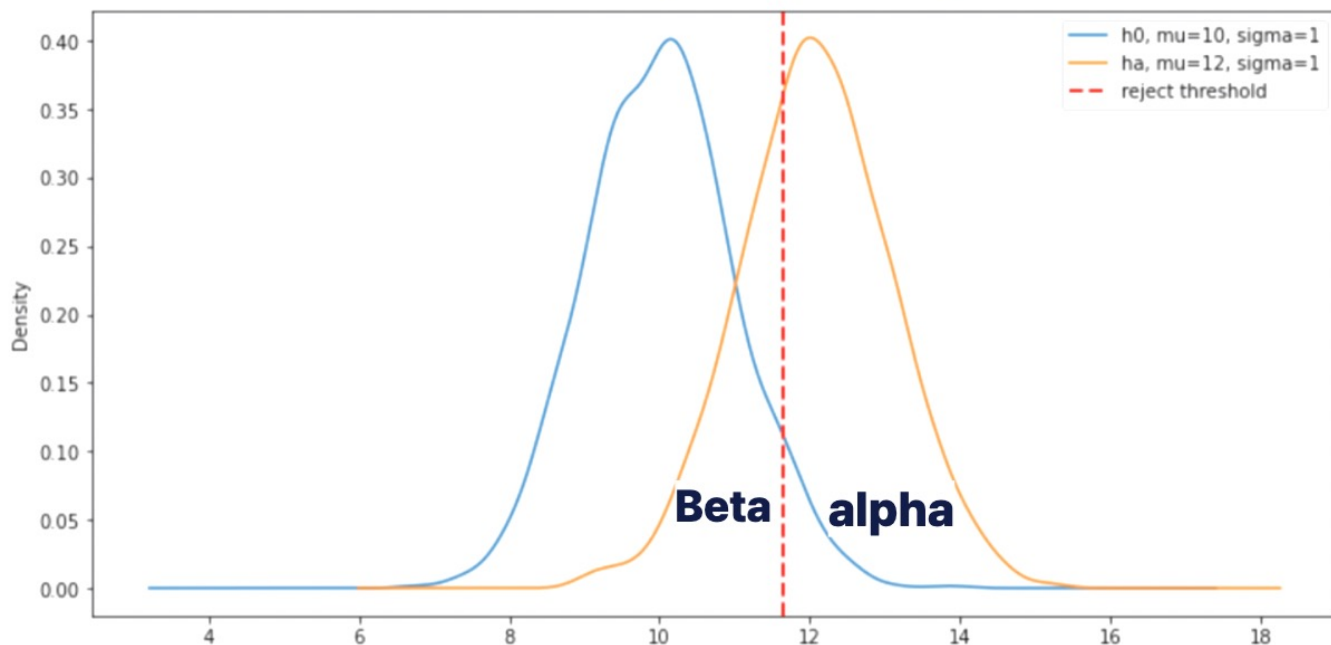
집단 크기 결정(최소 샘플 사이즈) - 실제로 n을 구하는 계산식

기각역과 채택역을 나누는 Threshold : k

$$\alpha = P(\bar{Y} > k, \mu = \mu_0) = P(Z > \frac{(k - \mu_0)}{\frac{\sigma}{\sqrt{n}}}, \mu = \mu_0) = P(Z > \lambda_\alpha)$$

$$\beta = P(\bar{Y} \leq k, \mu = \mu_\alpha) = P(Z \leq \frac{(k - \mu_\alpha)}{\frac{\sigma}{\sqrt{n}}}, \mu = \mu_0) = P(Z \leq -\lambda_\alpha)$$

k(=Threshold)



"" e.g. 기본 전환율이 2%이고, 최소 0.2%의 전환율(MDE=0.1)

alpha = 0.05, beta = 0.8일 때

필요한 최소 샘플사이즈 = 63,219 (집단별) ""

```
>>> ((0.84 + 1.64)**2 * ((0.02)*(1-0.02) + (0.022)*(1-0.022))) / (0.022 - 0.02) **2
63219.9616000011
```

k를 공통 수식으로 정리

$$k = \mu_0 + z_\alpha \left(\frac{\sigma}{\sqrt{n}} \right) = \mu_\alpha + z_\beta \left(\frac{\sigma}{\sqrt{n}} \right)$$

n으로 다시 정리

* (z_beta(0.8) : 0.84, z_alpha(0.95) : 1.64)

$$\sigma = p_0(1 - p_0) + p_\alpha(1 - p_\alpha)$$

두 분포의

Pooled Standard Deviation

$$n = \frac{(z_\alpha + z_\beta)^2 (p_0(1 - p_0) + p_\alpha(1 - p_\alpha))^2}{(\mu_\alpha - \mu_0)^2}$$