

# 생물학적동등성시험의 통계분석

- I. 기초 통계이론
- II. 생물학적동등성의 통계분석
- III. 참고문헌

통계학은 단순히 자료를 정리하는 것만을 뜻하는 것은 아니며, 여러 가지 예측이나 판단을 위한 매우 강력한 근거를 합리적으로 만들기 위해 많은 분야에서 중요하게 활용되고 있다. 생물학적동등성시험에서 동등성 여부의 판정은 통계분석결과에 의존하게 되므로, 이에 대해 기초통계이론과 생물학적동등성의 경우에 필요한 분산분석과 신뢰구간 추정, 집단약동학과 관련된 사항을 간단히 소개하기로 한다.

## I. 기초 통계이론

### 1. 통계학의 분류

통계학은 크게 기술 (descriptive) 통계학과 추론 (inferential) 통계학으로 나뉘어 진다. 기술 통계는 평균 (mean), 표준편차 (standard deviation), 중앙값 (median), 최빈값 (mode), 범위 (range), 사분위수 (quartiles) 등을 계산하고 자료를 요약하여 표현하는 대한 것이다. 추론 (추측) 통계는 확률 변수 (random variable)와 통계량 (statistic)을 이용하여 여러 가지 추론을 하는 것이다. 추론에는 크게 검정 (test)과 추정 (estimation)이 있다. 검정의 대상은 가설 (hypothesis)이며 추론의 대상은

파라미터 (parameter)이다. 대부분의 가설 검정은 Wald 형태의 검정 (Wald-type test)으로서 1회의 계산 (적합, fitting)에서 나온 파라미터의 점 추정치와 그 오차(error)를 이용하여 가설을 검정하기 때문에 비교적 단순하다. 반면에 우도비 형태의 검정 (likelihood ratio type test)은 2회 이상의 계산에서 나온 우도 (likelihood)들을 이용하여 검정을 하는 것이다. 후자는 모형간의 비교를 통하여 우수한 모형을 선택할 때 유용하다. 모형에 사용된 상수를 파라미터라고 하는데 이 파라미터에 대해서는 점 추정 (point estimation ; PE)을 할 수도 있고, 좀 더 바람직하게는 표준오차 (standard error ; SE) 등을 이용한 구간 추정 (interval estimation)을 할 수도 있다. 구간 추정의 가장 단순한 형태는 t 분포에서의 해당 자유도와 신뢰도 (유의수준)를 이용한 t값과 표준오차 (standard error ; SE)를 곱한 값을 점 추정치에 빼고 더하는 것이다. 이는 대체로 'PE  $\pm$  2 SE' 정도를 많이 사용한다.

### 2. 통계학 기본 개념과 용어

#### 가. 기호사용의 기본 규칙

여기에서는 일반적인 통계학 교과서 또는 이번 교재에서 사용하는 기호표기형식의 기본규칙을 설명한다. 그런데, 모든 교재가 꼭

이러한 원칙을 지켜서 작성된 것은 아니라는 점에 주의해야 한다.

모형은 흔히 수식을 사용하여 나타내고, 수식은 기호 (symbol)를 포함하게 된다. 기호는 함수라든지 여러 가지를 의미할 수 있지만, 수 (number)를 의미한다면 이 수는 상수 (constant, 값 value)가 아니면 변수 (variable)이다. 파라미터는 특별히 관심의 대상이 되는 상수를 의미하며 수식에서는 계수 (coefficient)로 나타나기도 한다. 일반적인 상수는 영어 소문자로 표시되며, 파라미터는 그리스 소문자 (일변수) 또는 대문자 (벡터, 행렬)로 표시한다. 용량, 시간과 같은 수학적 변수 (또는 시간에 따라 변하는 영향인자, covariates)는 영어 소문자로 표시한다. 통계학의 변수인 확률 변수는 대문자로 표시한다.

표 4.1. 기호표기

표시대상	표시의 예
확률변수	X, Y
파라미터	$\mu, \sigma, \rho, \alpha, \beta$
수, 관찰값, 수학적변수	a, b, c, x, y
행렬	$\Omega, \Sigma$

표시 예:  $X \sim N(\mu, \sigma^2)$

위의 의미는 확률변수 X는 평균이  $\mu$ 이고 분산이  $\sigma^2$ 인 정규분포를 따른다는 뜻이다.

나. 확률 (probability)

확률이란 우연 (chance)에 의해 일어나는 현상으로 실험자 또는 관찰자의 의지와 결과

는 무관하며 완벽한 통제가 불가능한 상황에서 이를 제대로 기술하고 예측하는 데 유용하다. 수학적인 의미에서 독립변수 (independent variable, 주로 x로 표기)가 연속인 값을 갖는 경우 연속함수로 확률을 표현할 수 있는데, 가장 대표적인 것이 확률밀도함수 (probability density function ; pdf)이다. 확률밀도 함수는 확률 변수 X가 구간 (a, b)의 값을 가질 확률은 확률 밀도함수를 해당구간에서 적분한 값과 같다. 이를 기호로 표시하면 다음과 같다.

$$P(a < X < b) = \int_a^b f(x)dx$$

확률밀도함수를  $-\infty$ 에서  $+\infty$ 까지 적분한 경우에 1을 가지며, 모든 확률은 0에서 1까지의 값만 가질 수 있다. 확률변수가 불연속인 값을 갖는 경우 불연속 함수로 나타내고 이 함수를 확률질량함수 (probability mass function ; pmf)라 부른다. 확률 변수는 누적 확률밀도함수 (cumulative distribution function ; cdf)나 적률생성함수 (moment generating function ; mgf)를 이용해서 나타낼 수도 있다. 이러한 함수들을 통칭해서 분포함수 (distribution function)라고도 한다.

다. 분포 (Distribution)

분포란 자료가 흩어져 있는 것을 의미한다. 분포를 표현하는 방법에는 자료를 그다지 가공하지 않고 거의 그대로 나타내는 도수분포표 (frequency table), 줄기-잎 그림 (stem-leaf graph)이나 각종 도표가 있다. 인간의 두뇌가 동시에 파악할 수 있는 숫자는 몇 개가

되지 않는다. 따라서, 분포를 요약하는 어떤 숫자 즉 대표값이 필요한데 이들 대표값들은 크게 분포의 중심 (center, 위치, location)을 표현하는 숫자들과 퍼진 (산포, dispersion) 정도를 표현하는 숫자들로 나뉘 수 있다. 평균, 중앙값, 최빈값, 왜도 등은 분포의 중심을 표현하는 대표값들이며, 분산, 표준편차, 범위, 사분위수, 첨도 등은 분포의 퍼진 정도를 표현하는 대표값들이다. 분포를 표현하는 또 다른 방법은 이상적인 형태의 수학함수로 나타내는 것이다. 이들 수학함수들은 확률밀도함수 (probability density function ; pdf), 확률질량함수 (probability mass function ; pmf), 누적확률함수 (cumulative probability function), 적률생성함수 (moment generating function) 등으로 대별할 수 있다. 수학 함수의 형태로 표현할 수 있는 분포의 종류에는 수십 가지가 있는데, 가장 대표적인 정규분포 (normal distribution)의 확률밀도함수를 표현하면 다음과 같다.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

동일한 분포함수라도 정의하는 방법이 최소한 2가지 이상 있는데, 많이 사용되는 분포함수를 가장 간단하게 정의하자면 다음과 같다.

$Z$ 라는 확률변수가 표준정규분포 (평균 0, 분산 1인 정규분포)를 따른다면, 다음의 확률변수  $V$ 는 자유도  $n$ 인 카이제곱분포를 따른다.

$$V = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi^2(n)$$

$V_1$ 이 자유도  $n_1$ 인 카이제곱분포를 따르고,  $V_2$ 가 자유도  $n_2$ 인 카이제곱분포를 따를 때, 그 비는 자유도  $(n_1, n_2)$ 인 F분포를 따른다. F분포는 자유도를 2개 가진다.

$$F = \frac{V_1/n_1}{V_2/n_2} \sim F(n_1, n_2)$$

위 확률변수의 역수는 F분포에서 자유도의 순서만 바꾸어주면 된다.

다음과 같이 표준정규분포를 따르는 확률변수  $Z$ 와 자유도  $k$ 인 카이제곱분포를 따르는 확률변수  $V$ 로 정의되는 확률변수  $T$ 는 자유도  $k$ 인 t-분포를 따른다.

$$T = \frac{Z}{\sqrt{V/k}} \sim t(k)$$

따라서, 위와 같이 정의되는 확률변수의 제곱은 자유도  $(1, k)$ 인 F분포를 따르게 된다.

$$T^2 = \frac{Z^2}{V/k} \sim F(1, k)$$

왜냐하면  $Z^2$ 은 자유도가 1인 카이제곱분포를 따르기 때문이다.

누적확률함수 (cumulative probability function)는 다음과 같이 정의된다.

$$F(x) = P(X < x) = \int_{-\infty}^x f(x)dx$$

따라서,  $F(\infty) = 1$  이 된다.

적률 (moment)은 2가지 형태가 있는데 하나는 평균으로 centering한 centered moment 들이 있고, 하나는 그렇지 않은 형태이다. 적률을 정의하면 다음과 같이 할 수 있다.

**1차 적률** (first moment = 평균, mean)

$$m_1 = \int_{-\infty}^{+\infty} x^1 f(x)dx$$

**2차 적률**

- uncentered form

$$m_2 = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

- centered form (= 분산, variance)

$$m_2 = \int_{-\infty}^{+\infty} (x - m_1)^2 f(x)dx$$

**3차 적률**

- uncentered form

$$m_3 = \int_{-\infty}^{+\infty} x^3 f(x)dx$$

- centered form (= 왜도, skewness)

$$m_3 = \int_{-\infty}^{+\infty} (x - m_1)^3 f(x)dx$$

**4차 적률**

- uncentered form

$$m_4 = \int_{-\infty}^{+\infty} x^4 f(x)dx$$

- centered form (= 첨도, kurtosis)

$$m_4 = \int_{-\infty}^{+\infty} (x - m_1)^4 f(x)dx$$

정규분포의 적률생성함수는 다음과 같다.

$$m.g.f(t) = e^{(\mu t + \frac{1}{2}\sigma^2 t^2)}$$

적률생성함수의 특징은 t에 대해 n 회 미분한 후 t에 0을 대입하면 그것이 곧 n차 적률이 된다는 것이다. 확률밀도함수, 누적확률함수, 적률생성함수는 같은 것을 다르게 표현하는 방법이며 서로는 원칙적으로 일대일 대응을 한다.

라. 확률 변수 (Random variable)

확률 변수란 사건을 숫자로 대응시키는 사상 (mapping) 또는 함수 (function)로서 수학에서의 일반적인 변수와는 의미가 다르다. 확률 변수는 대문자로 나타내는데, 표본의 추출 단위들은 개개가 하나의 확률변수로 표현된다. 모든 확률 변수는 분포를 가지며, 실험 후에는 어떤 상수 (관측값)를 가지게 된다. 확률 변수에 연산을 가하면 새로운 확률변수가 생성된다. 즉 확률 변수의 함수는 또 다른 확률

변수가 된다.

확률변수로서 표본평균의 정의는 아래와 같다.

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

위의  $\bar{X}$ 는 확률변수로서의 표본평균이지만, 소문자를 써서 다음과 같이  $\bar{x}$ 를 쓰게 되면 확률변수에 관찰값들을 대입한 후의 상수로서의 표본평균이 된다.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

표본표준편차의 경우에도 마찬가지이다.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

위와 같이 표기한 경우는 확률변수로서의 표본표준편차를 의미하고 아래와 같이 쓰면 관찰 값 (상수)으로서의 표본표준편차를 의미한다.

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

마. 통계량 (Statistic)

확률 변수 중에 추론에 유용하게 사용되는 확률 변수를 통계량 (statistic)이라 한다. 검정에 사용되는 검정 통계량 (test statistic)은 혼

히 줄여서 통계량 (statistic)이라고 부르며, 추정에 사용되는 통계량은 특별히 추정량 (estimator)이라고 부른다. 추정량의 값 (value)을 추정치 (estimate)라고 한다.

바. 표준 오차 (Standard error)

통계량의 표준편차 (standard deviation)를 특별히 표준오차 (standard error)라 한다.

사. 표본 크기 (Sample size)

모집단 (population)의 일부를 선택 (추출, 관찰)한 것을 표본 (sample)이라고 하는데, 모든 가능한 표본의 집합을 표본 공간 (sample space)이라 한다. 표본이 가진 원소 (element, 대개 근원사건-elementary event) 또는 추출하여 관찰하는 단위 (sampling, observation unit)의 개수를 표본 크기 (sample size)라고 한다. 임상시험에서는 ‘표본 크기’ 대신에 ‘피험자 수 (number of subjects)’라는 용어를 사용할 수 있다. 추론 통계학에서는 우리가 가진 자료 집합 (data set)에서 산출한 값 (통계량, statistic)이 표본 공간 (sample space)에서 어떠한 위치 (location, 여기서는 점 추정, point estimation)에 있으며, 그것의 정밀도 (precision, 여기서는 표준오차, standard error)를 알고자 하는 것이다.

아. 자유도 (Degree of freedom ; DF)

분포함수 (distribution function)의 모양을 결정하는 상수를 통계적 파라미터 (statistical parameter)라고 한다. 정규분포의 경우에는

평균 ( $\mu$ )과 분산 ( $\sigma^2$ )만 알면 함수가 하나로 결정되고, 나머지 적률들은 이에 따라 저절로 구해지게 된다. 분포함수가 가진 특징은 매우 적은 수 (2~3개)의 파라미터로 상당히 복잡한 형태의 곡선을 표현한다는 점이다. 만약 누군가가 새로운 파라미터 개수가 적으면서도 비교적 복잡한 모양을 가지고 있고, 전체 구간을 적분하였을 때 1의 값을 가진 함수를 발견한다면 새로운 분포함수가 탄생하는 것이다. 어떤 분포함수들은 기존의 분포함수를 몇 개씩 더하거나 다른 연산을 통해서 만들기도 하는데, 이 때 사용된 분포함수의 개수를  $n$ 으로 표현하면 이 수는 파라미터라고 하기에는 미흡하지만 분포곡선의 모양을 결정하고 있다는 면에서 파라미터와 유사한 특성이 있다. 즉  $n$ 에 따라서 분포함수의 모양이 달라지는데, 연구에서 이는 표본 크기와 관련된 숫자이며, 이를 자유도 (degree of freedom)라 한다. 일반적으로 어떤 통계량의 자유도는 표본크기에서 이전에 추정된 (추정 공식 우변에서 추정해야하는) 파라미터의 수를 차감하여 계산한다. 따라서 자유도는 수리통계학이나 분포함수 자체를 연구하는 입장에서는 분포함수의 모양을 결정하므로 파라미터와 다를 바가 없고, 연구를 실행하고 결과를 해석하는 입장에서는 표본크기와 밀접히 연관된 개념이다. 원래 파라미터는 상수이면서 분포함수의 모양을 결정하는 것인데, 표본크기는 개별 연구에 따라 바뀌지만 분포의 모양을 결정하므로, 자유도는 파라미터와 변수의 중간적인 성격을 지니고 있다고도 할 수 있다.

### 3. 추정 (Estimation)

관측값이 우연에 의해 결정된다면 추출단위 (sampling unit)를 확률변수로 표현할 수 있으며, 이를 이용한 새로운 함수의 종속변수 역시 확률변수이다. 즉, 개개 확률변수들의 평균을 표본평균이라는 확률변수로 표현한다면, 이 또한 개개 확률변수처럼 분포를 가진다. 이들의 분포를 알면 어떤 구간내에 어느 정도의 확률 (빈도)로 포함되는지 구할 수 있게 된다.

개개의 확률변수가 따르는 분포가 다음과 같으며, 그것들이 서로 독립이며 동일할 때 (*i.i.d.*: independent and identically distributed)

$$X_i \sim N(\mu, \sigma^2)$$

표본평균 ( $\bar{X}$ )이라는 확률변수 (통계량)가 따르는 분포는 다음과 같다.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

표본크기  $n$ 이 커질수록 개개  $X$ 의 분포와 상관없이 표본평균 ( $\bar{X}$ )의 분포는 정규분포에 가까워지는데 이를 중심극한 정리 (central limit theorem)라 부른다. 일반적으로는  $n$ 이 30보다 크면 (경우에 따라서는 25정도) 개별  $X$ 의 분포에 상관없이 표본평균 ( $\bar{X}$ )의 분포를 정규분포에 근사하여 계산한다.

$$P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95$$

$$P(\mu - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1.96 \frac{\sigma}{\sqrt{n}}) = 0.95$$

즉, 어떤 구간에 포함될 확률은 분포를 표준화한 경우의  $x$ 값 (여기에서 1.96은 표준정규분포에서의  $Z_{1-0.05/2}$  값이다.)

통계분석에서는 어떤 분포에 대한 가정을 하는 경우가 많은데, 그 가정이 대개 잔차(residual)에 대한 가정이며 잔차를 표현하는 확률변수는 대개 다음과 같은 하나의 분포를 따르고 서로 독립이라고 가정한다.

$$\epsilon_i \sim N(0, \sigma^2)$$

(위에서  $\epsilon_i$ 는 확률변수이지만 예외적으로 영어 대문자로 표현하지 않았음에 주의한다.)

일반적으로 다른 요인들은 모두 어떤 미지의 상수에 의해 결정되는 고정효과(fixed effect)로 간주한다.

예를 들면 다음과 같다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

와 같은 관계가 있으며 우리는 다만 개별  $\beta$ 값들만 모른다고 했을 때, 위의 관계는 어떠한 불확실성도 없는 결정적인 관계(deterministic relation)이며 따라서 확률적인 관계(probabilistic relation)가 아니다. 하지만, 실질적으로는 어떤 원인 때문에 무작위효과가 들어가며 다음과 같은 식으로 표현하는데,  $\epsilon_i$ 가 포함되면서 좌측변수  $y$ 는 확률변수인 대문자  $Y$ 로 표현하는 것이 옳다.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

(여기에서 우측의  $x$ 들은 관측값이며 이것은 우리가 마음대로 선택하는 것이 가능한 값들이어서 확률변수가 아니다.)

위에서  $Y$ 의 평균과 분산을 구하면 다음과 같다.

$$E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\begin{aligned} V(Y) &= V(\beta_0 + \beta_1 x_1 + \beta_2 x_2) + V(\epsilon) \\ &= 0 + \sigma^2 = \sigma^2 \end{aligned}$$

어떤 상수의 기댓값은 그 상수 그대로이며, 분산은 0이기 때문이다.

우리가 추정하고자 하는 미지의 상수를 통계적 모수(statistical parameter)라고 하는데, 이 모수의 추정에 사용하는 추정량(estimator, 확률변수의 일종으로서 추정에 사용되는 통계량)들이 가져야 할 바람직한 특성은 다음과 같은 것들이 있다.

- ① 불편성(unbiasedness):  $E(\hat{\theta}) = \theta$ ,  $\hat{\theta}$ 의 기댓값 또는 평균은  $\theta$ 이다.
- ② 최소분산성(minimum variance, maximum efficiency):  $V(\hat{\theta})$ 가 다른 추정량들에 비해 가장 작다.
- ③ 일치성(consistency):  $n$ 을 증가시킬수록  $\hat{\theta}$ 는  $\theta$ 에 가까워지며,  $n$ 을 증가시키면 오차를 원하는 범위내로 만들 수 있다.

최소분산불편추정량을 MVUE(minimum variance unbiased estimator)라는 약어로 쓰기도 한다.



모평균의 추정량( $\hat{\mu}$ )은 다음과 같이 표본평균 ( $\bar{X}$ )을 주로 사용한다.

$$\begin{aligned}\hat{\mu} &= \bar{X} \\ \text{왜냐하면,} \\ E(\bar{X}) &= \mu \quad (\rightarrow \text{불편}) \\ V(\bar{X}) &= \frac{\sigma^2}{n} \quad (\rightarrow \text{최소분산}) \\ \text{즉, } \bar{X} &\sim N(\mu, \frac{\sigma^2}{n}) \\ \text{따라서,} \\ SE(\hat{\mu}) &= SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}\end{aligned}$$

하지만, 실제로는 대부분의 경우에  $\sigma^2$ 을 알 수 없기 때문에  $\sigma^2$ 에 대한 추정치를 사용해야 한다. 즉,

$$\begin{aligned}\hat{\sigma}^2 &= S^2 \\ \text{왜냐하면,} \\ E(S^2) &= \sigma^2 \quad (\rightarrow \text{불편}) \\ \text{하지만,} \\ E(S) &< \sigma \text{이며 (S는 } \sigma \text{에 대해 편의추정량),} \\ S^2 \text{의 분포는 정규분포가 아니라 아래와 같} \\ &\text{은 카이제곱분포를 따른다.} \\ S^2 &\sim \frac{\sigma^2}{n-1} \chi^2(n-1) \\ \text{따라서, 신뢰구간도 대칭신뢰구간이 아니} \\ &\text{어서 표준오차로 신뢰구간을 표시하기 어} \\ &\text{려워진다.}\end{aligned}$$

미지의 모수에 대한 추정을 할 때 어느 한

점으로 추정을 하게 되면 점 추정 (point estimation)이지만, 추정량의 분포를 알면 구간을 구하고 그 구간 안에 참값이 포함될 것이라 추측하게 된다. 이때 우리가 가지는 확신의 정도를 표현하고자 할 때 신뢰구간과 유의수준 (confidence level,  $\alpha$ )이라는 용어를 사용하게 된다. 엄밀히는 신뢰구간이란 이번과 같은 방식으로 추정을 반복한다면 100 (1- $\alpha$ )% 정도의 횟수로는 이 구간 안에 참값을 포함시킬 수 있다는 뜻이다. 주어진 유의수준에서의 신뢰구간 (confidence interval)을 구할 때 신뢰구간이 포함하지 않는 면적을 유의수준 ( $\alpha$ )이라고도 할 수 있다.

지금까지의 과정을 정리하여, 가장 보편적으로 사용되는 모평균과 모분산에 대한 100(1- $\alpha$ )% 신뢰구간은 다음과 같다.

모평균의 신뢰구간 추정

$$\left( \bar{X} - t(n-1, \frac{\alpha}{2}) \frac{S}{\sqrt{n}}, \bar{X} + t(n-1, \frac{\alpha}{2}) \frac{S}{\sqrt{n}} \right)$$

모분산의 신뢰구간 추정

$$\left( \frac{(n-1)S^2}{\chi^2(n-1, \frac{\alpha}{2})}, \frac{(n-1)S^2}{\chi^2(n-1, 1-\frac{\alpha}{2})} \right)$$

하지만, 모평균의 신뢰구간 추정에 비해 모분산의 신뢰구간 추정은 모집단의 정규분포 가정에 깊게 의존하므로 부정확하기 쉽다는 점을 유의해야 한다.

#### 4. 가설검정 (Hypothesis Test)

통계적으로 입증하고자 하는 가설을 대립가설 (alternative hypothesis,  $H_a$  or  $H_1$ )이라고 하고, 우리가 입증하지 못한 경우에 채택하게 되는 가설을 귀무가설 (null hypothesis,  $H_0$ )이라 한다. 즉, 귀무가설은 우리가 통계적으로 입증하는데 실패하는 경우에 채택하는 것이기 때문에 귀무가설이 채택된 경우에는 강력하게 그 가설이 옳다고 주장할 수 없다는 점에 유의해야 한다. 물론 검정력 (power, 대립가설이 참이면서 이를 채택할 확률)이 큰 경우에는 어느 정도 선호할 수 있지만, 이 경우에도 유의확률 0.05를 사용하는 대립가설처럼 강력히 주장할 수 있는 것은 아니다. 또 다른 유의점은 대부분의 통계교재에는 대립가설은 차이가 있음을 보이는 것이고, 귀무가설을 차이가 없는 경우만 표현하고 있어서 오해의 여지가 있는데, 우리가 입증하고자 하는 것이 동등성일 때는 어느 범위 내 (동등성 한계)에서는 차이가 없음 (동등함)을 입증해야 하므로 귀무가설과 대립가설의 방향이 바뀔 수 이해해야 한다. 전술하였듯이 동등성 입증에 실패한 경우에도 차이가 있음을 입증한 것은 아니다. 연구자가 가장 두려워하는 경우는 어느 쪽으로도 결론을 내리지 못하게 되는 (inconclusive) 경우이다.

유의수준 (significance level)  $\alpha$ 인 검정법이란 제 1종의 오류를 범할 확률이  $\alpha$  이하인 검정법을 뜻한다. 검정력(power)란  $1-\beta$ 로 계산하는데 대립가설이 참일 때 이를 채택할 확률을 의미한다.

표 4.2. 오류의 종류

검정후 선택	미지의 진실	
	귀무가설	대립가설
귀무가설	옳은 채택	제2종의 오류 (Type 2 error, $\beta$ error)
대립가설	제1종의 오류 (Type 1 error, $\alpha$ error)	옳은 채택

유의확률 (significance probability, p-value)은 현재 관찰된 경우보다 극단의 경우로 나올 확률을 의미하는데, 유의확률이 유의수준보다 낮을 때 우리는 대립가설을 채택하게 된다.

구간 추정을 이용해서 신뢰구간을 구하게 되면 가설 검정은 쉬운 경우가 많다. 예를 들어, 100  $(1-\alpha/2)\%$ 신뢰구간이 0을 포함하지 않는 경우 우리는 유의수준  $\alpha$ 에서 양측 검정하는 경우 대립가설을 채택하게 된다.

#### 5. 두 모평균의 비교

한 피험자가 하나의 약만 복용한 경우 독립인 표본이 되며, 한 피험자가 2개의 약을 모두 복용한 경우에는 대응표본으로 볼 수 있다.

독립인 경우이면서 모분산들을 모르는 경우에는 평균의 차이에 대한 신뢰구간을 다음과 같은 공식으로 구한다.

$$(\bar{X} - \bar{Y}) \pm t(n_1 + n_2 - 2, \frac{\alpha}{2}) S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\text{단, } S_p^2 = \frac{\sum_i^{n_1} (X_i - \bar{X})^2 + \sum_i^{n_2} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2}$$

$$\text{또는 } S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$\text{검정통계량: } T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

위에서  $\delta_0$ 는 미지의 진실로 생각하는 실제 차이를 의미하는데, 대부분의 경우에 0으로 가정한다.

가교시험의 경우에 동일한 약을 서로 다른 민족에서 시험한 경우이므로 두 민족간의 평균의 차이에 대한 신뢰구간을 구하는데 위의 공식이 유용하게 사용될 수 있다. 또는 다른 문헌에 나온 서로 다른 제품간의 평균값 비교에도 사용할 수 있지만, 연구 간의 변이(inter-study variability)도 무시하지 못할 정도이므로 해석 시에는 이를 고려해야 한다.

한 피험자에서 두 제품의 약물을 먹고 AUC를 구한 경우는 대응비교를 생각할 수 있다. 하지만, 실제 상황에서는 이러한 공식을 사용하지 않고 이에 대해서는 다음에 설명하므로 사용에 유의해야 한다.

대응비교의 경우에는 다음과 같은 공식을 사용한다.

한 개인에서의 측정치의 차이를  $D_i$ 라 하면,  $D_i$ 들의 표본평균과 표본분산은 다음과 같다.

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$$

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$$

이 때 차이에 대한 신뢰구간은 다음과 같다.

$$\text{신뢰구간: } \bar{D} \pm t(n-1, \frac{\alpha}{2}) \frac{S_D}{\sqrt{n}}$$

여기에서 사용하는 검정통계량은 다음과 같다.

$$\text{검정통계량: } T = \frac{\bar{D} - \delta_0}{S_D \sqrt{n}}$$

일반적인 생물학적동등성시험에서는 한 사람이 2가지 제품의 약을 복용하므로 대응자료의 경우에 두 평균 비교방법을 이용하면 될 것처럼 보이나 실제로는 그러하지 않다. 실제로는 생물학적동등성시험의 설계에서 보듯이 투약순서를 무작위배정하여 교차시험을 행하는데 순서효과(sequence effect), 시기효과(period effect), 개인간의 차이(subject effect, intersubject variability)를 모두 고려한 분석을 시행한다. 이러기 위해서는 일단 분산분석을 시행하고, 분산분석 시행후에 구해지는 MSE(mean squared error)를 이용해서 신뢰구간을 추정하게 된다. 즉, 독립표본인 경우의  $S_p$  대신에 MSE의 제곱근을 이용해서 신뢰구간을 구한다. 분산분석에 대해서는 다

음에 설명한다.

생물학적동등성시험과 같은 교차시험에서 두 평균의 차이에 대한 신뢰구간:

$$(\bar{X} - \bar{Y}) \pm t(n_1 + n_2 - 2, \frac{\alpha}{2}) \sqrt{MSE}$$

위에서 MSE는 분산분석을 통해서 구해진다.

생물학적동등성시험에서는 1990년대 초반 이후부터 일단 로그변환한 후에 위의 공식들을 적용하기 때문에 최종 결과 값에 대한 해석은 역 로그 변환을 해서 하는 것이 보기가 좋은데, 이때는 원래 척도에서 기하평균의 비 (ratio of geometric means)에 대한 신뢰구간이라 할 수 있다.

## 6. 분산분석 (ANOVA, Analysis of Variance)

일반적인 기초통계 강의에서는 두 개의 평균을 비교할 때는 t-검정 (또는 t-분포를 이용한 신뢰구간)을 사용하고, 3개 이상의 평균을 비교할 때는 분산분석을 사용한다고 가르친다. 일단 평균을 비교하는 데 왜 분산을 사용할까 하는 점도 의문이 될 수 있고, ‘평균분석 (ANOM, analysis of means)’라는 용어도 인터넷에서 찾을 수 있으니 더욱 혼란스러울 수 있겠다.

기본적으로 분산분석은 군간 변동 (between group variability)이 군내변동 (within group variability)보다 큰가를 비교하는 것이다. 군

간 변동이든 군내변동이든 모두 분산의 형태로 표현되며, 결국 군간 변동이란 ‘군간 평균의 변이’를 의미하기 때문에 분산분석결과 유의하다는 것은 군간에 평균의 차이가 존재한다는 것을 의미한다. 개개 분산에 대한 추정량 (표본분산)들은 카이제곱분포를 따르지만, 표본분산들의 비는 F-분포를 따르기 때문에 F-분포표를 이용하게 된다.

가장 단순한 형태의 분산분석모형은 다음과 같다.

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

$$(i = 1, 2, \dots, k; j = 1, 2, \dots, n)$$

$\mu$ : 총평균

$\alpha_i$ : 처리효과로서 i번째 처리효과

$\epsilon_{ij}$ : 처리효과 i번째 군내에서 j번째 피험자의 오차 항으로서  $N(0, \sigma^2)$ 을 따름

분산 분석 내에서 각 변동들은 다음과 같은 관계를 가지며 총 변동은 치료 간 효과와 치료 내 효과로 분해될 수 있다.

- 총변동 (total sum of squares ; SST):

$$SST = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y})^2$$

- 치료군간 변동 (treatment sum of squares ; SST):

$$SST_T = n \sum_{i=1}^k (\bar{y}_i - \bar{y})^2$$

- 치료로 설명이 안되는 잔차제곱합 (residual sum of squares ; SSE)

$$SSE = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_i)^2$$

수학적으로 풀어보면

$SST = SS_T + SSE$  의 관계를 갖는다.

일반적인 분산분석결과는 표 4.3에서 제시하였다.

평균제곱 (mean sum of squares ; MS)은 제곱합 (sum of squares ; SS)에다 자신의 자유도로 나누어 준 값들이다. 이 값의 비가 F 분포를 따른다는 것이 이미 알려져 있으므로 이를 이용해 검정이나 신뢰구간 추정이 가능해진다. 여기에서 MSE는  $\epsilon_{ij}$ 가 따르는 정규 분포의  $\sigma^2$ 에 대한 추정치로 사용되며, 또한 차 후 두 군간 평균의 차이에 대한 신뢰구간을 구할 때도 이용한다.

표 4.3. 분산분석표

요인	제곱합 (SS)	자유도 (df)	평균제곱 (MS)	F비 (F ratio)
처리	$SS_T$	$k-1$	$MS_T = \frac{SS_T}{k-1}$	$F = \frac{MS_T}{MSE}$
잔차	$SSE$	$k(n-1)$	$MSE = \frac{SSE}{k(n-1)}$	
계	$SST$	$kn-1$		

## II. 생물학적동등성의 통계분석

### 1. 생물학적동등성시험설계

가장 보편적으로 널리 사용되는 방법은  $2 \times 2$  교차설계 (crossover design)이다.  $2 \times 2$ 란 피험자를 무작위로 두 군으로 나눈 후, RT 군은 대조약을 투약하고 휴약기간 (washout period)이 지난 후에 시험약을 투약하며, TR군은 반대의 순서로 투약한다. 즉,  $2 \times 2$ 란 2 sequence, 2 period란 뜻이며,

표 4.4.  $2 \times 2$  교차 설계 (Crossover Design)

	순서군	시기 1	시기 2
피험자모집 → 무작위배정 ↗ ↘	RT	대조약 ( $\mu_{R1}$ )	시험약 ( $\mu_{T2}$ )
	TR	시험약 ( $\mu_{T1}$ )	대조약 ( $\mu_{R2}$ )

표 4.5. Mean response by fixed effects

Sequence	Period I	Period II
RT	$\mu_{R1} = \mu + S_1 + P_1 + F_R$	$\mu_{T2} = \mu + S_1 + P_2 + F_T + C_R$
TR	$\mu_{T1} = \mu + S_2 + P_1 + F_T$	$\mu_{R2} = \mu + S_2 + P_2 + F_R + C_T$

S: Sequence Effect, P: Period Effect, F: Formulation Effect, C: Carryover Effect

crossover란 한 피험자가 두 가지 약제를 모두 투여 받는다는 것을 의미한다.

피험자수는 총 24명 내외이며 군당으로는 12명 내외이다. 휴약기간은 대개 1주일이지만 반감기가 길거나, P-glycoprotein을 유도하는 약물의 경우에는 2주 이상 소요될 수 있다.

위와 같은 설계에서는 fixed effect로 순서 (군)효과 [Sequence (Group) Effect], 시기효과 (Period Effect, P), 이월효과 (Carryover effect, C), 제품효과 (Product/Formulation Effect, F)가 있을 수 있다. 이런 효과를 고려할 때 각 경우에 관찰할 수 있는 평균은 표 4.5와 같이 구성할 수 있다.

$2 \times 2$ 설계로는 이월효과를 산출할 수 없기 때문에 이월효과가 없거나 최소한 두 약물사이에 동일하다는 가정을 하고 분석해야 한다.

또한, 최근 미국 FDA가 규정한 개인 생물학적 동등성 (individual bioequivalence)을 구할 수 없으므로 개인 생물학적동등성을 구하기 위해서는 다음과 같은 좀 더 복잡한 설계를 사용해야 한다.

2 sequence, 3 period ( $2 \times 3$ ):

$$\begin{bmatrix} \text{TRR} \\ \text{RTT} \end{bmatrix} \begin{bmatrix} \text{TRT} \\ \text{RTR} \end{bmatrix}$$

$2 \times 4$  Design:

$$\begin{bmatrix} \text{TRRT} \\ \text{RTTR} \end{bmatrix} \begin{bmatrix} \text{TRTR} \\ \text{RTRT} \end{bmatrix}$$

Baalam's Design ( $4 \times 2$ ):

$$\begin{bmatrix} \text{TT} \\ \text{RR} \\ \text{TR} \\ \text{RT} \end{bmatrix}$$

표 4.6. 순서효과, 시기효과, 잔차효과에 따른 설계의 분류

Uniform within Sequence	Uniform within Periods	Balanced	Strongly Balanced	Examples
N	N	N	N	AAB ABB, ABCC BCAA
Y	N	N	N	ABB BAB, ABC CBA
N	Y	N	N	ABC BCA CAB AAA
N	N	Y	N	ABAA BAAB
N	N	Y	Y	AABBA BAABB
Y	Y	N	N	ABC BCA CAB
Y	N	Y	N	AABA ABAA
N	Y	Y	N	ABA BAB
Y	N	Y	Y	AABBA ABBAA
N	Y	Y	Y	ABB BAA, AB BA AA BB
Y	Y	Y	N	AB BA
Y	Y	Y	Y	ABBA BAAB AABB BBAA

4 × 4 Design:

RTTR
TRRT
RRTT
TTRR

3가지 이상의 제품을 동시에 비교하기 위한 아래와 같은 설계도 있다.

3 × 3 Design:

ABC
BCA
CAB

6 × 3 Design:

ABC
BCA
CAB
ACB
CBA
BAC

4 × 4 Design:

ABCD	ABCD
BCDA	BDAC
CDAB	CADB
DABC	DCBA

순서군 간에 제품 종류 및 반복이 동일할 때 ‘uniform within sequences’라 하고, 시기군의 제품 종류 및 반복이 동일할 때 ‘uniform within periods’라 한다. 이월효과가 군 간에 동일할 때 ‘balanced with respect to carryover effects’라 하고, 각 제품별로 자신의 시기 앞에 다른 제품들이 동일한 횟수로 모두 나올 때 ‘strongly balanced with respect to carryover effects’라 한다. 마지막 설

계는 그 이전 설계에 비해 각 치료 앞에는 다른 치료가 1회씩 동일하게 나오므로 잔차효과 (carryover effect)에 대해 strongly balanced 되었다고 한다. 이러한 기준에 따른 설계 예들을 표 4.6에 나타내었다.

## 2. 생물학적동등성 규정상의 통계

미국 규정에서는 만약 두 약이 남녀구분 없이 사용되는 약이면 남녀가 동수로 포함되고, 주로 노인에게 사용될 약이면 가능하면 노인 환자를 많이 포함하길 권하고 있다. 하지만 세부군 별로 통계 처리하는 것은 권장되지 않으며, 일반적으로 건강자원자를 이용하나 필요한 경우 환자를 대상으로 할 수도 있게 하였다. 농도 측정 방법은 반드시 타당성을 입증하고 실시하여야 한다. 신속한 작용시간이 요구되는 약물의 경우는 최고혈장농도 도달시점 ( $T_{max}$ ) 이전에 2회 이상 채혈하고 부분 AUC (partial AUC, 예를 들어  $AUC_{0-2hr}$ )를 비교하여야 한다. 일반적인 약물의 경우에는 약물에 대한 노출정도를  $AUC_{last}$ 와  $AUC_{inf}$  모두를 비교하여야 한다 (우리나라 현 규정에서는  $AUC_{last}$ 만 비교하는 것으로 되어있다). 반복투여인 경우에는 투여간격의  $AUC$  ( $AUC_{\tau}$ )를 비교한다. 경구 투여하여 전신 흡수되는 약물로서 농도측정이 가능한 경우가 아닌 경우 약력학적 변수를 이용하여 동등성을 입증할 수도 있다. 이때는 엄밀히 말하면 생물학적동등성이 아니라 약효동등성시험이라 하겠다. 위 두 가지로도 안 되는 경우 비교 임상시험을 할 수도 있겠으나 약효가 동등하



다해서 꼭 생물학적으로 동등하다는 것은 아니다.

생물학적동등성 판정기준은 AUC,  $C_{max}$ 에 대해서는 개개 값을 로그 변환하여 평균의 차이에 대한 90% 신뢰구간을 구하고 이것이  $\log(0.8)$ 에서  $\log(1.25)$  사이이어야 한다. 이것은 로그변환하지 않은 상태에서 두 제품의 기하평균 (geometric mean)의 비 (ratio)가 0.8에서 1.25사이임을 보이는 것과 같다. 식사의 영향을 보는 생물학적동등성에서는 AUC의 기준은 공복 시와 동일하고  $C_{max}$ 의 경우에는  $\log(0.7)$ 에서  $\log(1.43)$ 까지로 약간 더 넓다. 이는 식사에 의해서는  $C_{max}$ 변이가 커지기 때문이다. 로그 변환의 근거는 여러 가지가 있으나 통계적으로는 AUC,  $C_{max}$ 는 로그 변환 시 좀 더 정규분포에 가까워지기 때문이다. 즉, 대부분의 약물에서 로그 변환 값이 정규분포를 따르며, 변환이전에는 분포의 꼬리가 우측으로 길게 치우쳐진 (right skewed) 것으로 보기 때문이다.

실제 통계처리 과정에서는 이러한 신뢰구간을 구하기 전에 시험 진행/결과에 문제가 없었는지 확인하기 위해 교차설계시의 분산분석 (ANOVA for crossover study)을 실시하여 군 (순서)효과, 시기효과, 약제 (제품)효과가 어떠한지를 분석한다 (유의수준은 0.05). 시기효과는 p-glycoprotein의 영향, 농도측정일간의 변이, 대사효소의 변화 등으로 인해 흔히 발견되므로, 시기효과가 있었다고 해서 반드시 시험이 잘못되었다거나, 동등성을 입증할 수 없는 것은 아니다. 분산분석에서 구한 MSE (mean squared error)를 신뢰구간 추정시의 표

준오차 (standard error)에 사용된다. 최근에는 피험자효과 (subject effect, 피험자 간 변이)를 고정효과 (fixed effect)로 보지 않고 무작위효과 (random effect)로 보는 선형혼합효과모형 (linear mixed effects model)을 사용하여 분석하는 경향이 있으며, 통계적으로 더욱 바람직하다고 할 수 있다. 이미 개인 생물학적동등성 (individual BE)을 위해서 사용하는 미국 FDA의 표준적 SAS 코드는 PROC GLM이 아니라 PROC MIXED를 사용하는 것이다.

### 3. 통계분석 예

다음은 어느 한 생물학적동등성시험의 AUC자료를 자연로그변환한 후 분산분석결과를 구하는 과정을 예시하였다. 실제로는

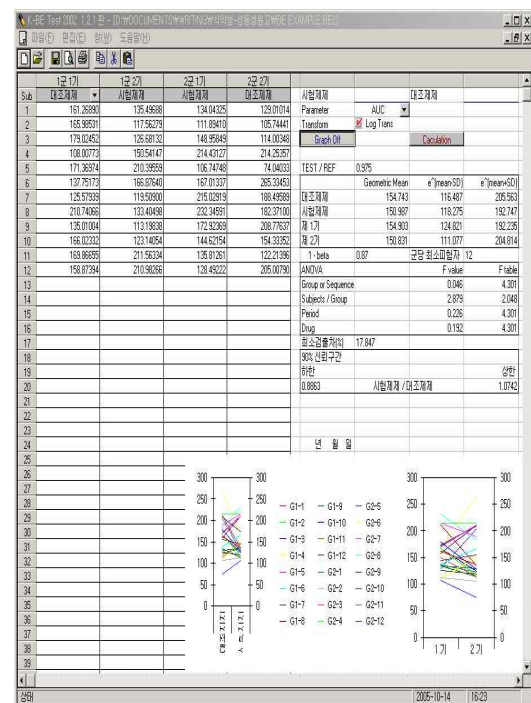


그림 4.1. KBE test로 통계분석을 실시한 예



KBE<sup>®</sup> test, EquivTest<sup>®</sup>, WinNonlin<sup>®</sup> 등의 편리한 소프트웨어를 사용하는 것이 대부분이며, 관련 통계 지식이 충분하다면 SAS<sup>®</sup>나 MS-Excel<sup>®</sup>을 사용할 수도 있다.

분산분석표에 필요한 제곱 합들을 계산하기 위하여, 개별 관측 값을  $y_{ijk}$ 라 하고  $i$ 를 순서군 번호 (1, 2),  $j$ 를 순서군내 피험자번호 (1, ...,  $n$ ),  $k$ 를 시기번호 (1, 2)라 하면,

- 군당 피험자 수( $n$ ): 12

- 대조약 평균( $\hat{\mu}_R$ ):

$$\hat{\mu}_R = \left( \frac{1}{n} \sum_{j=1}^n y_{1j1} + \frac{1}{n} \sum_{j=1}^n y_{2j2} \right) / 2 = 5.042$$

- 시험약 평균 ( $\hat{\mu}_T$ ):

$$\hat{\mu}_T = \left( \frac{1}{n} \sum_{j=1}^n y_{1j2} + \frac{1}{n} \sum_{j=1}^n y_{2j1} \right) / 2 = 5.017$$

- 평균의 차 ( $\hat{\mu}_T - \hat{\mu}_R$ ): -0.025

- 총 제곱 합 (SST<sub>u</sub>):

$$SST_u = \sum_{i=1}^2 \sum_{j=1}^n \sum_{k=1}^2 y_{ijk}^2 = 1217.426$$

- 총합 (ST):

$$ST = \sum_{i=1}^2 \sum_{j=1}^n \sum_{k=1}^2 y_{ijk} = 241.415$$

- 교정항 (CF): (계산의 편의를 위해 도입)

$$\frac{ST^2}{4n} = 1214.192$$

표 4.7. 자연 로그 변환 후 AUC 값

순서군	피험자 번호	시기 1	시기 2	계
1 (RT)		대조약	시험약	
	A-1	5.083	4.909	9.992
	A-2	5.112	4.767	9.879
	A-3	5.188	4.842	10.029
	A-4	4.682	5.014	9.696
	A-5	5.144	5.349	10.493
	A-6	4.925	5.117	10.043
	A-7	4.833	4.783	9.616
	A-8	5.351	4.893	10.244
	A-9	4.905	4.729	9.634
	A-10	5.112	4.813	9.925
	A-11	5.135	5.355	10.490
	A-12	5.068	5.352	10.420
	소계	60.538	59.924	120.462
2 (TR)		시험약	대조약	
	B-1	4.898	4.860	9.758
	B-2	4.718	4.661	9.379
	B-3	5.004	4.736	9.740
	B-4	5.368	5.367	10.735
	B-5	4.670	4.305	8.975
	B-6	5.118	5.581	10.699
	B-7	5.371	5.239	10.610
	B-8	5.448	5.206	10.654
	B-9	5.153	5.341	10.494
	B-10	4.974	5.039	10.013
	B-11	4.911	4.806	9.717
	B-12	4.856	5.323	10.179
	소계	60.489	60.464	120.953
계		121.027	120.388	241.415

- 총 변동 (SST):

$$SST = SST_u - CF = 3.234$$

- 피험자 간 변동 제곱 합 (SS<sub>s</sub>):

$$\frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^n (y_{ij1} + y_{ij2})^2 - CF = 2.390$$

- 순서군간 변동 제공 합 ( $SS_g$ ):

$$\frac{1}{2n} \sum_{i=1}^2 \left[ \left( \sum_{j=1}^n (y_{ij1} + y_{ij2}) \right)^2 \right] - CF = 0.005$$

- 순서군내 피험자 간 변동 제공 합 ( $SS_{s/g}$ ):  
 $SS_s - SS_g = 2.385$

- 시기 간 변동 제공 합 ( $SS_p$ ):

$$SS_p = \frac{1}{2n} \sum_{k=1}^2 \left( \sum_{i=1}^2 \sum_{j=1}^n y_{ijk} \right)^2 - CF = 0.009$$

- 약제 간 변동 제공 합 ( $SS_d$ ):

$$SS_d = \frac{1}{2n} \left[ \left( \sum_{j=1}^n (y_{1j1} + y_{2j2}) \right)^2 + \left( \sum_{j=1}^n (y_{1j2} + y_{2j1}) \right)^2 \right] - CF = 0.007$$

- 잔차 제공 합 (SSR)

$$\begin{aligned} SSR &= SST - (SS_s + SS_p + SS_d) \\ &= SST - (SS_g + SS_{s/g} + SS_p + SS_d) \\ &= 0.828 \end{aligned}$$

$$MSE = SSR / (2n-2) = 0.069$$

평균제공합 (MS, mean square)들은 제공합에다 자신의 자유도를 나누어 주면 되고, 대부분의 효과들에 대한 유의성 여부는 자신의 평균 제공 합을 잔차의 평균 제공 합 (MSE)으로 나눈 값 (F값)이 F분포를 따르므로 F검정을 실시하여 판단하면 된다. 다만, 이와 같은 교차시험에 대한 분산분석에서 유의할 점은 피험자 효과와 순서군 효과를 동일한 수준 (level)에서 볼 수 없다는 것이다. 한

피험자가 동시에 양쪽 순서 군에 포함될 수 없고, 피험자 번호가 결정되면 순서군은 무조건 결정된다 (A-1 피험자와 B-1 피험자는 피험자 번호가 다르다). 즉, 순서군은 자유도가 없어지며, 역으로 말하자면 피험자는 순서군에 내포되어 (nested) 있다. 따라서, 순서군 효과는 순서군내 피험자 간 변동을 이용해서 순서군에 따라 변동이 있는지 검정해야 한다.

표 4.8. 교차시험의 분산분석표

변동	자유도	제공합	평균제공합	F값
피험자	2n-1	$SS_s$	$SS_s/(2n-1)$	$MS_g/MSE$
순서군	1	$SS_g$	$SS_g/1$	$MS_g/MS_{s/g}$
피험자/순서군	2n-2	$SS_{s/g}$	$SS_{s/g}/(2n-2)$	$MS_{s/g}/MSE$
시기	1	$SS_p$	$SS_p/1$	$MS_p/MSE$
약제	1	$SS_d$	$SS_d/1$	$MS_d/MSE$
잔차	2n-2	SSR	$SSR/(2n-2)$	
총합	4n-1	SST		

표 4.9. 분산분석표 계산결과

변동	자유도	제공합	평균 제공합	F값	p-값
피험자	23	2.390	0.104	2.760	0.010
순서군	1	0.005	0.005	0.046	0.831
피험자/순서군	22	2.385	0.108	2.879	0.008
시기	1	0.009	0.009	0.226	0.639
약제	1	0.007	0.007	0.192	0.665
잔차	22	0.828	0.038		
총합	47	3.234	0.069		

따라서, 순서군 효과를 검정할 때는 MSE가 아니라  $MS_{s/g}$ 를 오차항 (error term)으로 삼아서 나누어 주어야 한다. 또 여기에서 자세히 설명하지는 않았지만, 양쪽 순서군의 피험자 수가 다른 경우에는 각각의 피험자 수를  $n_1, n_2$ 라 할 때,  $2n$  대신에  $n_1 + n_2$ 를 사용하면 되고, 평균을 구할 때 적절한  $n$ 으로 나누어 주면 된다. 두 군의 피험자수가 다르다고 하여 많은 쪽 순서군의 피험자를 임의로 누락시켜서는 안된다.

분산분석결과를 살펴보면 피험자 간 차이 (피험자, 피험자/순서군)는 통계적으로 유의하며 (모두  $p < 0.05$ ), 순서군, 시기, 약제의 효과는 통계적으로 유의하지 않다 (각각  $p = 0.831, 0.639, 0.665$ ). 피험자 간 변동 (개개인의 차이, 피험자 효과)은 대부분의 경우에 유의한 것이 자연적인 현상이다. 따라서, 이 연구에서 시험 진행 과정/결과상의 특별한 문제는 발견되지 않는다. 여기에서 구한 MSE는 0.038로서 다음 과정인 기하평균의 비 ( $\mu_{기하,T}/\mu_{기하,R}$ )에 대한 90% 신뢰구간 (CI, confidence interval)을 구하는데 필요한 표준 오차에 대한 추정치의 제공으로 사용한다.

먼저 위 표에서 각 약제의 기하평균 및 그에 대한 비는 위의 결과에 역 로그 변환 (anti-log transformation, exponentiation)를 하면 된다.

- 대조약의 기하평균:

$$e^{\mu_R} = e^{5.042} = 154.743$$

- 시험약의 기하평균:

$$e^{\mu_T} = e^{5.017} = 150.987$$

- 기하평균의 비 ( $\mu_{기하,T}/\mu_{기하,R}$ ):

$$e^{\mu_T - \mu_R} = e^{5.017 - 5.042} = 0.976$$

- 로그 변환 후 차이에 대한 점 추정치의 표준 오차 (SE):

$$SE = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{\sqrt{\hat{\sigma}^2}}{\sqrt{n}} = \frac{\sqrt{MSE}}{\sqrt{n}}$$

- 90%CI이므로 유의수준 (0.05)에서의 T값 :

$$\begin{aligned} t(2n-1, \text{양측 } \alpha) &= t(2n-1, 1-\alpha/2) \\ &= t(1-0.10/2) = 1.714 \end{aligned}$$

따라서, 로그 변환한 상태에서 평균의 차이에 대한 90% 신뢰구간의 하한과 상한은 다음과 같다.

$$\begin{aligned} &(\widehat{\mu_T - \mu_R}) \pm t(2n-2, 1-\alpha/2) \cdot SE \\ &= (\widehat{\mu_T - \mu_R}) \pm t(2n-2, 1-\alpha/2) \cdot \frac{\hat{\sigma}}{\sqrt{n}} \\ &= (-0.121, 0.072) \end{aligned}$$

로그변환상태에서 차이의 신뢰구간은 역 로그 변환시 이전상태에서 기하평균의 비에 대한 신뢰구간이 되며, 계산 결과는 다음과 같다.

- 기하평균의 비에 대한 90% 신뢰구간:

$$(e^{-0.121}, e^{0.072}) = (0.886, 1.074)$$

따라서, 자연로그 변환 후 90% 신뢰구간이  $-0.22314$  ( $\log_e 0.8$ )와  $0.22314$  ( $\log_e 1.25$ ) 사이 (원래 척도로는 기하평균의 비에 대한 90%신뢰구간이 0.8에서 1.25)에 있으므로 이 제품의 AUC는 생물학적동등성을 만족시킨다고 판단한다.

$C_{max}$ 의 경우에도 AUC와 동일하게 자연 로그 변환 후 분산분석과 90%신뢰구간을 산출하면 된다. 위에서 피험자 간 변동 (피험자 효과)을 고정효과 (fixed effect)로 간주하고 통계분석을 실시하였으나, 피험자는 항상 고정된 것이 아니라 모집단에서 무작위 추출된 것 무작위효과 (random effect)로 보는 것이 더 타당하다. 아직 규정은 그렇게 변화하지 않았지만, 향후 차츰 이러한 방향으로 변화할 것이 예상된다. 그러나, 피험자효과를 고정효과로 간주하거나, 무작위효과로 간주하는 것이 분산분석표 구하는 중간과정에서 약간의 차이가 있을 뿐 순서군, 약제, 시기 효과의 검정이나 신뢰구간추정에는 동일한 결과를 가져

온다. 다만 SAS®와 같은 프로그램에서 피험자 효과를 무작위 효과로 간주하는 경우 다음 예제와 같이 10여줄 정도의 프로그래밍으로 모든 원하는 결과를 얻을 수 있지만, 기존처럼 고정효과로 간주하는 경우에는 수 십 줄에서 수 백 줄의 프로그래밍이 필요하다.

$T_{max}$ 의 경우는 처리에 주의를 필요로 한다.  $T_{max}$ 는 일견보기에 연속형 변수로 보이지만, 실제로는 투약 후 0.5, 1, 2시간과 같이 예정된 시간에만 관찰하였고, 이 값을 사용하므로 연속형으로 볼 수 없으며, 더군다나 정규분포를 따르지도 않는다. 따라서,  $T_{max}$ 에 대해서 이와 같은 방법으로 신뢰구간을 구하지 않고, 비모수적 검정법을 사용하여야 한다. 대표적인 방법으로 Hodges-Lehmann 방법과 Bootstrap 방법이 있다.

#### 4. 표본크기 (피험자 수)의 산출

다음은 시험약 (test drug)과 대조약 (reference drug)의 실제 차이 (미지의 진실)가 0 (비율로는 1)이라고 가정한 경우 90% CI가 동등성 기준을 만족하기 위한 표본크기를 어떻게 구하는지 알아본다.

표본크기에는 유의수준 (제 1종의 오류)뿐 아니라 검정력 (제 2종의 오류와 관련,  $1-\beta$ )의 크기에 의존적이다. 그런데, 동등성이란 차이가 어느 정도 이하임을 주장하는 것이므로, 차이가 실제로 존재할 때의 p-value를 산출할 수 있어야 한다. 그러기 위해서는 다음과 같은 비심 t-분포 (noncentral t-distribution)의 이용

표 4.10 SAS® 프로그램 예

```
DATA BEDRUG;
  INFILE 'BEDRUG.DAT';
  INPUT SUBJ $ GROUP $ PERIOD $ DRUG $ AUC;
  LNAUC = LOG(AUC);

PROC GLM;
  CLASS SUBJ GROUP PERIOD DRUG;
  MODEL LNAUC = GROUP SUBJ(GROUP) PERIOD
          DRUG;
  RANDOM SUBJ(GROUP);
  ESTIMATE '약제효과' DRUG 1 -1 /ALPHA=0.1;
```

이 필요하다.

원래 t-분포의 정의는 확률변수  $Z$ 가 표준 정규분포 (평균은 0, 분산은 1인 정규분포)를 따르고, 확률변수  $V$ 가 자유도  $k$ 인 카이제곱 분포를 따를 때 다음과 같은 확률변수  $T$ 는 t-분포를 따르게 된다.

$$T = \frac{Z}{\sqrt{V/k}} \sim t(k)$$

하지만, 만약에  $Z$ 가 표준정규분포가 아니라  $N(\lambda, 1^2)$ 의 정규분포를 따를 때, 우리는 비심도 (non-centrality)  $\lambda$ 인 t-분포를 따른다고 한다.

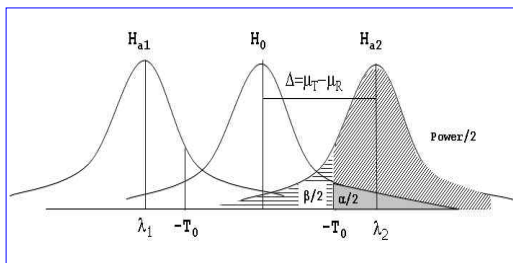


그림 4.2. 신뢰구간과 동등성 판정

우선 양측검정 유의수준  $\alpha$ 와 표본크기  $n$ 에서  $t$ 값을  $T_0$ 로 하여 임시로 다음과 같이 정한다 (세 번째 인자는 비심도이다).

$$T_0 = t(1 - \frac{\alpha}{2}, 2n - 2, 0)$$

유의한 차이가 있는 경우 비심도 ( $\lambda$ )는 다음과 같다.

$$\lambda_1 = \frac{\log_e \rho_L}{\sigma / \sqrt{n}}$$

( $\rho_L$ : 동등성하한, 대개 0.8)

$$\lambda_2 = \frac{\log_e \rho_U}{\sigma / \sqrt{n}}$$

( $\rho_U$ : 동등성상한, 대개 1.25)

표준편차( $\sigma$ )에 대해서는 MSE의 제곱근을 사용하게 되므로 양측검정을 기준으로 귀무가설이 참일 때 대립가설을 선택할 확률( $\beta$ -error)은 아래와 같다.

$$p_1 = P(t(2n - 2, \frac{\log_e 0.8}{\sqrt{MSE/n}}) > -T_0)$$

$$p_2 = P(t(2n - 2, \frac{\log_e 1.25}{\sqrt{MSE/n}}) < T_0)$$

$$\beta = p_1 + p_2$$

따라서, 위 3개의 식을 동시에 만족하는  $n$ 을 낮은 수부터 순차적으로 증가시키면서 구하거나, 반복대입 (iteration)을 통해서 구할 수 있다.

MSE가 가용하지 않고, 다음과 같은 CV가 있는 경우에는 MSE와 CV간에 환산을 할 수 있다.

$$MSE = \log_e (1 + CV^2)$$

여기에서 CV는 피험자 간 변동 (intersubject

variability)을 제외한 부분의 변이를 의미한다 (많은 경우 이를 피험자내 변이 intrasubject variability라 한다).

가장 정확히 표본 수를 구하는 것은 위와 같이 비심 t-분포를 이용해서 표본크기를 구하는 것이지만, 사용이 불편하여 아래와 같은 상대적으로 간단한 공식을 사용하거나 피험자 수 산출 소프트웨어를 이용하는 경우가 많다.

피험자 수가  $n$ 이고 유의수준  $\alpha$ 와 검정력  $(1-\beta)$  0.8에서 다음과 같은 공식으로 어느 정도 피험자 수를 산출할 수 있다.

$$n \geq [(t(2n-2, \alpha/2) + t(2n-2, \beta))]^2 (CV/20)^2$$

$$\text{단, } CV = 100 \times \frac{\sqrt{MSE}}{\mu_R}$$

위 식에서  $n$ 은 군당 피험자 수인데, 양변에 모두 포함되어 있으므로,  $n$ 에 여러 값을 대입하면서 (iteration) 구하게 된다. 조금 더 단순하게 반복대입 (iteration)이 필요 없는 형태의 공식도 있으나, 약간 더 부정확해진다.

문헌상에서 분산분석표의 MSE가 표시되어 있는 경우나 우리가 독자적으로 분산분석을 실시할 만큼 충분히 자료가 나타나있는 경우에는 MSE를 산출하고 이를 토대로 피험자 수를 산출할 수 있겠지만, 대부분의 경우에는 이를 직접적으로 구할 수 없다. 하지만, 피험자들의 전체 변동계수 (CV)만 나와 있는 경

우에는 그 값의 일부가 MSE (또는 개체내 변동을 나타내는 CV)가 되므로 이를 적절히 감하여 피험자 수를 산출할 수 있다. 일반적으로 생물학적동등성시험에는 피험자 탈락시 대체추가를 허용하지 않으므로 탈락을 고려하여 위에서 구한 표본크기에 군당 1명 정도를 추가해야 한다.

### III. 참고문헌

- 1) 식품의약품안전청, 고시 제 2005-31호 생물학적동등성시험기준 (2005. 6. 7).
- 2) US FDA CDER, Guidance for industry: Bioavailability and bioequivalence studies for orally administered drug products - General Considerations (2000. 10).
- 3) US FDA CDER, Guidance for industry: Statistical approaches to establishing bioequivalence (2001. 1).
- 4) 식품의약품안전청 생동성시험 추진반, 생물학적동등성시험 설명회 (2001. 8. 30).
- 5) S.C. Chow and J.P. Liu, Design and Analysis of bioavailability and bioequivalence studies - Revised and expanded, 2nd ed. Marcel Dekker (2000).
- 6) E. Diletti, D. Hauschke and V.W. Steinijans, Sample size determination for bioequivalence assessment by means of confidence intervals, *Int J Clin Pharmacol Ther Toxicol* **29**(1), 1-8 (1991).
- 7) A.M. Mood, D.C. Boes and F.A. Graybill, Introduction to the theory of statistics, 3rd ed. McGraw-Hill (1974).

- 8) Y. Pawitan, In All Likelihood: Statistical modelling and inference using likelihood, Oxford University Press (2001).
- 9) E.F. Vonesh and V.M. Chinchilli, Linear and nonlinear models for the analysis of repeated measurements, Marcel Dekker (1997).