

안정성시험 데이터의 풀링과 유효기간 추정



경천수 이사 (주)바이오큐어팜 생산본부장

세포배양의약품, 세포치료제, 유전자치료제 등 생물의약품의 안정성 시험에 있어서, 모든 인자 또는 인자조합에서 추정된 유효기간 중 적어도 1개 이상이 제안한 기간 보다 짧게 추정된 경우, 유효기간을 더 길게 추정하기 위하여 모든 배치의 데이터를 하나로 통합할 수 있는지를 결정하기 위하여 풀링테스트를 실시한다.

하나의 유효기간을 추정하기 위하여 여러 배치의 데이터를 풀링하기 전에 예비 통계분석(검정)을 하여 여러 배치의 회귀직선이 공통의 기울기와 공통의 절편을 갖는 지를 결정한다. 배치들 간의 회귀직선의 기울기와 절편의 차이를 검정하기 위하여 시간을 공변량으로 하는 공분산분석(ANOCOVA)을 적용할 수 있다. 통상 상대적으로 제한된 샘플 크기를 갖는 안정성시험 설계에 기인한 작은 검출력을 보상하기 위하여, 기울기(Time항)의 유의성과 절편(Batch항)의 유의성 기준은 0.25의 p값을 기준으로 한다.

공분산분석(ANOCOVA = ANOVA + Regression Analysis)은 실험의 정밀도(Precision)를 높이기 위한 방법으로 어떤 실험에서 반응변수 y 외에 반응변수와 선형 관계(Linear Relationship)를 갖는 x 가 존재 하는 경우, x 는 통제할 수 없는 변수(Uncontrollable Variable)이며, x 에 따라 y 가 관측된다. 이러한 x 를 공변량(Covariate 또는 Concomitant Variable)이라 부른다. 공분산분석 (Analysis of Covariance, ANOCOVA)는 이러한 공변량의 영향을 고려하여, 관측된 반응변수 y 를 조정해 준다. 이러한 조정(Adjustment)이 수행되지 않으면 공변량이 MSE(Mean Squares for Errors)를 변질시켜, 처리(Treatments)에 기인한 반응(Response) 간의 차이(True Differences)를 검출하기 어렵게 만든다. 결국 공분산분석은 통제할 수 없는 잡음 변수(Uncontrollable Nuisance Variable)의 영향을 조정해 주는 방법이라 할 수 있다.

예를 들어, 3개의 서로 다른 기계로 생산하는 섬유의 강도 차를 검출하는 것을 가정하자. 그런데, 섬유의 강도는 그것의 두께에 영향을 받기 때문에(일반적으로 두꺼운 섬유가 얇은 섬유보다 강도가 높다), 기계간의 순수한 강도의 차이를 검출하기 위해서는 공분산분석을 사용하여 섬유의 강도(y)에 대한 두께(x)의 영향을 제거시켜야 한다.

공분산분석을 위하여 하나의 공변량이 존재하는 단일인자(Single Factor) 설계(Design)에 국한하여 설명하기로 하자. 1개의 반응변수만이 존재하는 경우, 반응치 y 에 대한 모형은 다음과 같이 쓸 수 있다.

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases} \dots\dots\dots (1)$$

y_{ij} : i 번째 Treatment(어떤 인자의 특정수준 i)하에서의 j 번째 반응치

x_{ij} : y_{ij} 에 대응하는 공변량

τ_i : i 번째 Treatment의 효과. 단, x_{ij} 는 Treatment의 영향을 받지 않음

μ : 전체 평균

$\bar{x}$: x_{ij} 의 평균

여기서, 처리(Treatment)의 효과인 τ_i 만을 따로 분리해서 평가해야만 처리의 효과를 참되게 평가했다고 할 수 있다.

위의 (1)의 반응치 모형을 가정하고, 실험에 의해 반응치를 얻은 경우, 각 변수에 대한 제곱합을 정리하면 아래와 같다. 아래의 제곱합 식에서 알 수 있듯이, 일반적으로 $S = T + E$ 이다.

$$\begin{aligned} S_{yy} &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{an} & S_{xx} &= \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^a \sum_{j=1}^n x_{ij}^2 - \frac{x_{..}^2}{an} \\ S_{xy} &= \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) = \sum_{i=1}^a \sum_{j=1}^n x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{an} \\ T_{yy} &= \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^a \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an} & T_{xx} &= \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^a \frac{x_{i.}^2}{n} - \frac{x_{..}^2}{an} \\ T_{xy} &= \sum_{i=1}^a (\bar{x}_{i.} - \bar{x}_{..})(\bar{y}_{i.} - \bar{y}_{..}) = \sum_{i=1}^a \frac{(x_{i.})(y_{i.})}{n} - \frac{(x_{..})(y_{..})}{an} \\ E_{yy} &= \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = S_{yy} - T_{yy} & E_{xx} &= \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2 = S_{xx} - T_{xx} \\ E_{xy} &= \sum_{i=1}^a (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}) = S_{xy} - T_{xy} \end{aligned}$$

위와 같은 제곱합이 주어졌을 때, 회귀계수에 대한 최소 제곱 추정치를 구하면 아래와 같다.

$$\hat{\beta} = E_{xy} / E_{xx}$$

이 때, 회귀식으로 설명될 수 없는 y 의 변동인 SSE는 다음의 식으로 계산된다.

$$SSE = E_{yy} - (E_{xy})^2 / E_{xx}$$

위 제곱합에 대한 자유도는 $a(n-1)-1$ 이며, $(E_{xy})^2/E_{xx}$ 는 얼마나 회귀식이 y 를 설명하는데 적절한가에 대한 하나의 지표이다. 참고적으로, 위의 식은 일반적인 회귀분석 실시 시 계산되는 오차항에 대한 제곱합과 유사한 형태를 갖게 된다. 따라서, 오차항에 대한 평균 제곱합은 다음과 같이 계산된다.

$$MSE = \frac{SSE}{a(n-1)-1}$$

처리(Treatment), 즉 인자의 유의성 판정 방법을 살펴보기로 하자. (1)의 모형에서 y 에 대한 처리의 영향, 즉 효과가 없다고 가정하면 아래의 식으로 다시 쓸 수 있다.

$$y_{ij} = \mu + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij}$$

이 때, 회귀식의 기울기에 대한 최소 제곱 추정치는 앞서 소개한 식과는 달리 다음의 식으로 계산된다.

$$\hat{\beta} = S_{xy} / S_{xx}$$

아울러, 오차항의 제곱합 $SSE' = S_{yy} - (S_{xy})^2/S_{xx}$ 로 계산될 수 있다. 여기서, $SSE < SSE'$ 임을 알 수 있다.

$SSE' - SSE$ 는 처리의 효과 τ_i 가있고 없고의 차이이므로, 이 값이 크면 처리의 효과 τ_i 가 없다고 할 수 없다. 즉 처리가 유의하다고 할 수 있다. 이에 대한 통계적 검정을 위해 아래의 검정 통계량을 사용한다.

$$F_0 = \frac{(SSE' - SSE)/(a-1)}{SSE/\{a(n-1)-1\}}$$

여기서, F_0 의 계산치가 $F_{\alpha, a-1, a(n-1)-1}$ 보다 크면 $(SSE' - SSE)$ 는 무시할 수 없을 정도로 크다고 할 수 있으며, 결국 $\tau_i = 0$ 이다라고 할 수 없다. 안정성시험에서의 기울기(공변량: 시간)의 유의성 판단 기준값인 p 값은 F_0 값을 확률로 변환한 값으로, p 값이 0.25보다 큰 경우, 기울기는 유의하지 않다고 볼 수 있다. 아래 표1은 F_0 를 이용한 처리의 유의성 검정을 위한 분산 분석표이다.

<표 1> 공분산분석표

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Regression	$(S_{xy})^2 / S_{xx}$	1		
Treatments	$SSE' - SSE$ $= S_{yy} - (S_{xy})^2 / S_{xx}$ $- [E_{yy} - (E_{xy})^2 / E_{xx}]$	$a-1$	$\frac{SSE' - SSE}{a-1}$	$\frac{(SSE' - SSE)/(a-1)}{MSE}$
Error	$SSE = E_{yy} - (E_{xy})^2 / E_{xx}$	$a(n-1)-1$	$MSE = \frac{SSE}{a(n-1)-1}$	
Total	S_{yy}	$an-1$		

공분산분석의 예로서, 다음 <표 2>와 같은 3배치의 안정성시험 결과 2번째 배치의 안정성이 제안한 유효기간 보다 짧다고 가정하자.

<표 2> 18개월의 의약품의 안정성 데이터

Batch	Time Point (Month)					
	0	3	6	9	12	18
Batch 1	100	101	101	100	99	98
Batch 2	99	100	100	99	99	97
Batch 3	101	100	99	99	98	97

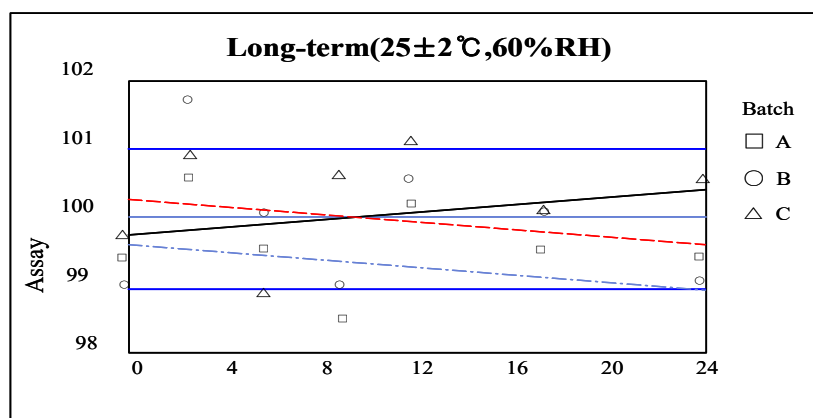
이 경우 좀 더 긴 유효기간을 추정하기 위하여 다른 배치로부터 선택된 데이터가 단일의 유효기간의 추정을 위해 모아질 수 있는가를 판단하기 위하여 다음 <표 3>과 같이 공분산분석을 실시하였다.

<표 3> 공분산분석표

Source	DF	SS	MS	F	P
Time	1	0.1587	0.1587	0.26	0.621
Batch	2	4.8150	2.4075	3.90	0.049
Batch*Time	2	0.4413	0.2206	0.36	0.706
Error	12	7.4000	0.6167		
Total	17	21.7778			

기울기(Time항)의 p값이 0.621로 0.25보다 크므로 세 배치를 통합하여 공통의 기울기를 사용할 수 있으나, 절편(Batch항)의 p값은 0.049로 0.25보다 작으므로 절편은 통합할 수 없다. 따라서, 공통의 기울기와 각각의 절편을 사용한 95% 신뢰한계가 허용기준(규격값)과 교차되는 시기를 재시험기간 또는 유효기간으로 추정하여야 한다.

6개월 가속시험에서 유의적인 변화가 없는 원료의약품의 예로, 공통의 기울기와 절편의 적용 가능성, 시간에 따른 경시적 변화와 데이터 변동을 측정하여 재시험기간 유도를 위한 통계분석의 필요성 결정, 재시험기간 추정 등을 살펴보자. 그림1은 25±2℃, 60%RH 조건에 보관한 원료의약품 3배치의 장기안정성 데이터이다.

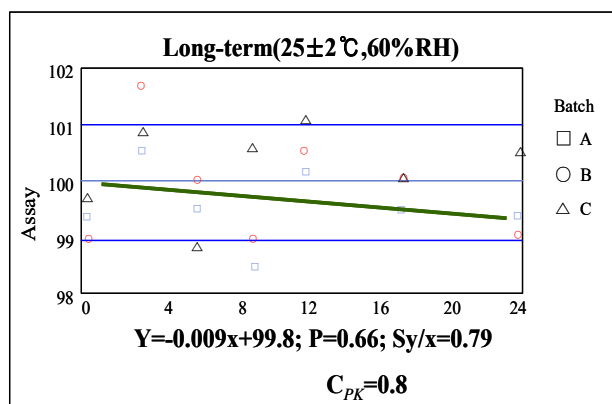


[그림 1] 배치 A, B, C의 장기안정성 데이터

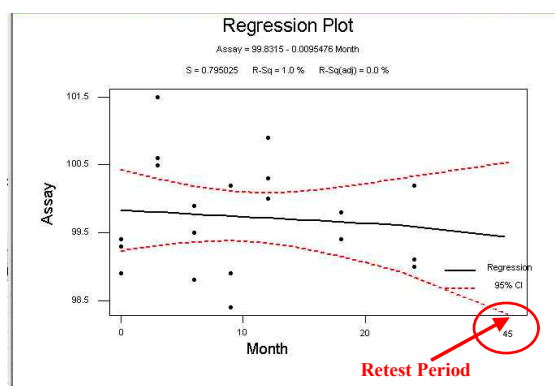
각 배치에 회귀분석을 적용하기 보다는 우선, 대표적인 재시험기간을 유도하기 위하여 3 배치의 안정성 데이터를 통합할 수 있는지 풀링테스트 실시한다. <표 4>는 풀링테스트를 위하여 실시한 공분산분석표이다. 3개의 기율기간의 차이(p 값=0.67)와 3개의 절편간의 차이(p 값=0.48)는 모두 p 값이 0.25보다 크므로 통계적으로 유의하지 않음을 알 수 있다. 따라서, 모든 배치의 데이터를 단일 배치의 데이터로 취급하여 공통의 기율기와 공통의 절편을 가진 모델을 상정할 수 있다. 공통의 기율기와 공통의 절편으로 상정한 회귀직선과 공정능력분석을 [그림 2]에 나타냈다.

<표 4> 공분산분석표

Source	DF	SS	MS	F	P
Month	1	0.1258	0.1258	0.19	0.6727
Batch	2	1.0352	0.5176	0.76	0.4831
Month *Time	2	0.6253	0.3176	0.46	0.6390
Error	15	10.1633	0.6776		
Total	20	11.9496			



[그림 2] 배치 A, B, C 통합 데이터의 회귀분석 및 공정능력분석



[그림 3] 배치 A, B, C 통합 데이터를 이용한 재시험기간의 추정

위의 [그림 2]의 분석에서 보듯이 기율기의 p 값은 0.66으로 0.25보다 크므로 시간에 따른 경시적 변화는 없다고 판단할 수 있다. 그러나, 공정능력지수는 0.8로 2.5보다 작으므로 데이터의 변동은 유의하다. 즉 데이터의 변동이 존재한다고 볼 수 있다.

이러한 상황에서는 통계분석에 의한 재시험기간의 추정이 요구된다. 이러한 통계적 외삽에 의한 재시험기간 추정을 [그림 3]과 같이 실시하였다. 여기서 95% 편측 신뢰 하한선이 45개월에서 규격하한인 98.0과 교차하는 것을 볼 수 있다. 그러나, ICH Q1E에 따라 최대를 제안할 수 있는 재시험기간은 $24 + 12 = 36$ 개월을 초과할 수 없으므로, 식약처에 품목허가 시 제안할 수 있는 재시험기간은 36월로 한다.