

위키백과

분산 분석

분산 분석(**analysis of variance**, **ANOVA**, **변량 분석**)은 통계학에서 두 개 이상 다수의 집단을 비교하고자 할 때 집단 내의 분산, 총평균과 각 집단의 평균의 차이에 의해 생긴 집단 간 분산의 비교를 통해 만들어진 **F분포**를 이용하여 가설검정을 하는 방법이다. 통계학자이자 유전학자인 로날드 피셔(**R.A. Fisher**)에 의해 1920년대에서 1930년대에 걸쳐 만들어졌다.

목차

F분포

분산분석 모형

고정효과 모형

무선효과(無選效果) 모형

혼합효과 모형

분산분석의 종류

일원분산분석(one-way ANOVA)


이원분산분석(two-way ANOVA)

다원변량분산분석(MANOVA)

공분산분석(ANCOVA)

같이 보기

F분포

 **F분포** 문서를 참고하십시오.

F분포는 분산의 비교를 통해 얻어진 분포비율이다. 이 비율을 이용하여 각 집단의 모집단분산이 차이가 있는지에 대한 검정과 모집단평균이 차이가 있는지 검정하는 방법으로 사용한다. 즉 $F = (\text{군간변동})/(\text{군내변동})$ 이다. 만약 군내변동이 크다면 집단간 평균차이를 확인하는 것이 어렵다. 분산분석에서는 집단간의 분산의 동질성을 가정하고 하기 때문에 만약 분산의 차이가 크다면 그 차이를 유발한 변인을 찾아 제거해야 한다. 그렇지 못하면 분산분석의 신뢰도는 나빠지게 된다.

■ 가정

1. 정규성 가정

각각의 모집단에서 변인 Y는 정규분포를 따른다. 각각의 모집단에서 Y의 평균은 다를 수 있다.

2. 분산의 동질성 가정

Y의 모집단 분산은 각각의 모집단에서 동일하다. $\sigma_1^2 = \sigma_2^2$

3. 관찰의 독립성 가정

각각의 모집단에서 크기가 각각 n_1, n_2 인 표본들이 독립적으로 표집된다.

각각의 표본에서 산출된 모집단 분산의 추정치의 비율 $F = \frac{s_1^2}{s_2^2}$ 을 구한다. 이를 'F' 또는 'F 통계치'라고 한다. F 값들은 특정한 이론적 확률분포를 따르게 되는데 이것이 F 분포이다.

분산분석 모형

실험계획을 수행할 때 수준을 선택하는 방법에 따라 모수인자와 변량인자가 있으며 각각 다음과 같은 분산 분석 모형이 있다,

고정효과 모형

고정효과 인자는 수준의 선택이 기술적으로 정해져 있고 각 수준이 기술적 의미를 가지고 있는 효과 인자를 말한다. 예로 온도, 압력 같은 것들이다. 모수인자만 사용된 경우 고정효과 모형(Fixed-effects model, 모수인자 모형)이라고 한다. 이 경우 각 수준에서의 모평균 값의 추정에 의미를 두고 있다.

무선효과(無選效果) 모형

무선효과 인자는 수준의 선택이 임의적으로 이루어지며 각 수준이 기술적 의미를 가지고 있지 않은 효과 인자를 말한다. 예를 들어 원료의 종류 같은 것들이다. 무선효과 인자만 사용된 경우 무선효과 모형(Random-effects model, 변량인자 모형)이라고 한다. 이 경우 각 수준은 임의적으로 결정되었기 때문에 각 수준의 모평균값의 추정이 의미가 없으며 단지 인자에 의한 산포의 정도를 추정하는 것에 의미를 두고 있다. 대표적인 예로 Gage R&R이 있다.

혼합효과 모형

고정효과 인자와 무선효과 인자가 함께 사용된 경우 혼합효과 모형(Mixed-effects model, 혼합인자 모형)이라고 한다.

분산분석의 종류

일원분산분석(one-way ANOVA)

종속변인은 1개이며, 독립변인의 집단도 1개인 경우이다.

예1) 가구소득에 따른 식료품소비 정도의 차이이다. 여기서 가구소득은 독립변인으로 가구소득집단의 구분-저소득, 중산층, 고소득층 등으로 2개 이상이다. 독립변인의 집단이 2개 이상이므로 사후분석을 실시한다.

예2) 한/중/일 국가간 10세 남아의 체중비교의 경우이다. 독립변인:10세 남아, 독립변인의 집단 : 3개 (한/중/일), 종속변인 : 1개(체중)이다.

이원분산분석(two-way ANOVA)

독립변인의 수가 두 개 이상일 때 집단 간 차이가 유의한지를 검증하는 데 사용한다.

예1) 독립변인 2개, 종속변인이 동일한 경우로 학력및 성별에 따른 휴대폰요금의 차이를 분석한다면 이때 학력, 성별은 독립변인이고 종속변인은 휴대폰요금이 된다. 이원분산분석은 주효과와 상호작용효과를 분석할 수 있다. 주효과는 학력(a), 성별(b)이라면 상호작용효과는 이를 곱한 $a*b$ 이다. 여기서 상호작용효과가 유의하다면 그래프를 만들어 볼 수 있다.

예2) 한/중/일 국가간 성별과 학력에 따른 체중비교의 경우이다. 독립변인 : 2개(성별/학력), 독립변인의 집단 : 3개 (한/중/일), 종속변인 : 1개(체중)이다.

다원변량분산분석(MANOVA)

단순한 분산분석을 확장하여 두개 이상의 종속변인이 서로 관계된 상황에 적용시킨 것이다. 둘 이상의 집단간 차이를 검증 할 수 있다.

공분산분석(ANCOVA)

다원변량분산분석에서 특정한 독립변인에 초점을 맞추고 다른 독립변인은 통제변수로 하여 분석하는 방법이다. 특정한 사항을 제한을 하여 분산분석을 하는 것이다.

같이 보기

- 분산(variance)
- 회귀 분석(regression analysis)
- F-분포
- t-test

원본 주소 "https://ko.wikipedia.org/w/index.php?title=분산_분석&oldid=26617517"

이 문서는 2020년 5월 26일 (화) 12:56에 마지막으로 편집되었습니다.

모든 문서는 크리에이티브 커먼즈 저작자표시-동일조건변경허락 3.0에 따라 사용할 수 있으며, 추가적인 조건이 적용될 수 있습니다. 자세한 내용은 이용 약관을 참고하십시오.

Wikipedia®는 미국 및 다른 국가에 등록되어 있는 Wikimedia Foundation, Inc. 소유의 등록 상표입니다.