SPSS를 활용한 연관성분석

# 1. ‘기타 변수의 개입 여부’의 개념

기타 변수의 개입여부는 주로 편상관분석(부분상관분석, partial correlation analysis)에서 생각해야 하는 문제이다. 예를 들어 발표 연습(X), 발표점수(Y) , 발표 울렁증(Z)과 같은 변수가 있다고 가정하자. 이 3가지 변수는 서로 연관이 있다고 할 수 있다. 즉 발표 연습(X)을 많이 했다면 당연히 발표점수(Y)는 높아지겠지만, 발표 울렁증(Z)이 있는 학생이라면 그 결과는 달라질 것이다. 이런 경우에 Z를 제어 변수로 설정하여 배제한, 고유의 상관이라고 할 수 있다.

# 2. 부정확하지만 많이 사용하는 방법 vs. 정확한 분석 방법

이 둘의 차이점을 정리하면 다음과 같다.

|  |  |
| --- | --- |
| **부정확하지만 많이 사용하는 방법** | **정확한 방법** |
| 주성분분석으로 요인 분석한 결과로 분석 | 공통요인분석의 최대우도법과 직접 오블리민으로 요인분석한 결과로 분석 |
| 산술평균값을 변수로 저장하여 분석 | 요인분석의 변수 저장값을 기준으로 분석 |

정확한 방법으로 상관분석을 실시한 뒤 출력결과의 [기술통계량] 표를 확인하면 평균이 모두 0으로 나올 것이다. 이에 연구자들은 무척 당혹스러워 한다. 그래서 통계학 지식이 없는 경우에 산술평균을 직접계산 한 후 상관분석을 실시하는데, 사실은 평균이 0이 되어야 정확한 방법으로 상관분석이 실행된 것이다.

그 이유를 간단히 살펴보자. 상관분석은 서로 다른 변수들 간에서 이루어지므로, 각 변수들을 비교할 수 있는 어떠한 매커니즘이 필요하다. 이러한 매커니즘을 ‘표준화’라고 할 수 있다. 이때 ‘표준화’한다는 의미는 각 표본의 평균을 0으로 조정하고, 다른 수치들(분산, 표준편차 등)을 기준으로 상관관계를 확인한다는 뜻이다. 따라서 문항 간 기계적인 산술평균을 이용한다면 뜻하지 않게 다중공선성 및 데이터의 왜곡을 야기할 수 있다. 이 부분이 잘 이해되지 않는다면 기초 통계학을 참고 해야한다.

또한, 상관분석에서는 최대우도법의 요인분석과 직접 오블리민의 요인회전을 실시하여 얻어진 결과를 이용하여 상관분석을 실시해야 한다. 주요인분석과 공통요인 분석의 차이점, 그리고 직각회전과 사각회전의 차이점을 다시 기억해야 한다.

# 3. 변환 메뉴를 이용하여 산술평균으로 변수를 만드는 방법의 위험성

이 방법은 다수의 통계서적에서 소개되고 있는데, 요인으로 구분된 문항들의 산술평균값으로 상관분석을 진행하는 방식이다. 하지만 요인분석에서 변수의 내재적 특성을 반영한 ‘변수로 저장’ 기능을 이용하는 경우가 아닌, 단순한 산술평균을 이용하는 경우에는 요인의 특성이 무시될 수 밖에 없다. 즉 변인에 내재된 특성을 완전히 무시하기 때문에 변수들 간에 상관성이 존재한다는 특성인 ‘다중공선성’을 고려하지 못한 채 분석 결과를 내놓을 수 밖에 없다.

# 4. 상관분석에 이용하는 변수

요인분석을 할 때 요인점수를 활용하기 위해 ‘변수로 저장’을 하면, 아래와 같이 요인분석을 통한 변수들이 저장되어 있음을 확인할 수 있다.

요인분석에서 이렇게 한 이유는 ‘요인점수’를 별도로 계산하여 변수로 저장한 후 활용하겠다는 의미이다. 요인분석의 요인추출방법은 ‘최대우도’이며, 회전방법은 ‘직접 오블리민’이다.

# 5. 저장된 변수의 확인

변수란의 변수명이 ‘FAC1\_1~FAC5\_1’ 이 아닌 ‘REGR factor score 1 for analysis 1 ~REGR factor score 5 for analysis 1’ 임을 알 수 있다. 그 이유는 ‘변수로 저장’방법에서 ‘회귀분석(regression)’을 선택했기 때문이다. 만약 여기서 ‘Bartlett’을 선택했다면 ‘BART factor score 1 for analysis 2’ 와 같이 나타날 것이다.

이에 대한 설명은 SPSS Statistics의 변수보기 탭을 선택해 보면 더 명확하게 알 수 있다. ‘이름’은 데이터 보기 탭에서 본 것과 같은 ‘FAC1\_1’로 되어있고, ‘설명’ 열은 ‘REGR factor score 1 for analysis 1’ 과 같이 되어 있는 것을 확인할 수 있다.

# 6. [기술통계량]의 평균값

[기술통계량] 표의 평균값이 모두 0인 것을 보고 분석이 잘못된 건 아닌지 의문을 가질 수 있다. 하지만 여기서는 오히려 평균값이 0으로 나타나야 제대로 분석을 마쳤다고 할 수 있다. 그 이유는 [NOTE 25]의 요인분석: 요인점수 창에서 ‘변수로 저장’의 선택 방법에 따라 특성이 다르기 때문이다.

* 회귀분석 : 평균을 0으로 하는 참 요인값과 추정된 요인 간의 차이를 제곱한 값을 최소로 하는 값을 변수로 따로 저장한다.
* Bartlett : 평균을 0으로 하는 변수들 간의 범위에서 고유한 요인들을 제곱한 값의 합을 최소로 하는 값을 따로 저장한다.
* Anderson-Rubin 방법 : (평균) = 0 , (표준편차) = 1, 추정된 요인들간에 상관관계가 없음을 확인하기 위해 Bartlett 값을 수정한 값을 따로 저장한다.