고급회귀분석(조절/매개/로지스틱)

# 조절 회귀분석

## 1. 조절 회귀분석에서의 위계적 회귀분석

조절 회귀분석을 위계적 회귀분석으로 설명하는 사람은 거의 없다. 대부분 1단계/2단계/3단계의 회귀분석을 단순/다중 회귀분석이 각각 진행되는 것으로 설명한다. 그러나 위의 1단계/2단계/3단계의 순서대로 회귀분석을 진행하는 과정은 사실 앞에서 학습한 위계적 회귀분석과 동일하다. 그러므로 조절 회귀분석에서의 위계적 회귀분석은 ‘상호작용변수’를 추가하여 실시한 위계적 회귀분석과 동일한 결과를 얻는다.

## 2. F변화량과 유의확률 F변화량의 의미

우리는 유의수준을 학습하면서 회귀분석에서 유의수준을 확인하는 지표로 F, t, p가 있다고 배웠다. 조절회귀분석에서의 모형요약 표에서 ‘F변화량’은 변수투입에 따른 F값의 변화량을 의미하며, ‘유의확률 F변화량’은 변수 투입에 따른 분산들의 변화차이를 F값 대신 p값으로 나타낸다. 이 p값의 변화량이 변수투입에 따른 유의확률의 변화량을 나타낸다. 변수투입에 따른 모형별 분산의 변화량은 p값으로 표시한 [계수]표를 보고 확인해야 한다.

# 매개 회귀분석

## 1. 측정오차가 포함될 가능성이 높다면?

SPSS Statistics 에서 매개회귀분석을 실시할 때는 오차항을 고려하지 않고 분석을 진행한다. 만약 연구자가 측정한 데이터에 측정오차가 포함됐을 가능성이 높다면 역시 정확하다고 할 수 없다. 만약 좀 더 정밀한 분석 결과를 얻고자 한다면 구조방정식모델을 사용하여야 한다.

구조방정식모델에서는 연구모델 자체에 오차(측정오차, 구조오차)를 포함하여 분석을 진행한다.

# 로지스틱 회귀분석

## 1. 로지스틱 회귀분석의 각종 지표

* Cox 와 Snell의 R2, Nagelkerke R2 : 모형의 설명력을 말한다.
* Hosmer 와 Lemeshow 검정 : 로지스틱 회귀모형의 전체적인 적합도를 판단한다.(유의확률이 유의수준(p<.05) 보다 커야함).
* 회귀식에 추정된 계수(B), 표준오차(S.E. : standard error), Wald(혹은 Wals), 자유도, 유의확률, Exp(B)

: 추정된 회귀계수와 유의성에 대한 결과를 확인한다.

## 2. 연구문제(다항 로지스틱 회귀분석)

지역(강남, 강북), 학력(초대졸, 대졸, 대학원졸), 스마트폰의 효용(만족감, 외관, 유용성, 편의성)에 대해 구매의사(구매, 구매안함, 관심없음)의 차이여부에 대해 확인하고자 한다.

## 3. 상호작용의 설정

다항 로지스틱 회귀 : 모형 창의 ‘사용자 정의/단계선택’을 선택하면 더욱 세밀하게 모형을 설정할 수 있다. 독립변수는 6개이지만 하나의 독립변수를 기준으로 삼아 독립변수 간의 상호작용을 나타내므로 ‘모든 5원 효과’ 까지 설정할 수 있다. 상호작용을 확인하는 과정에서 강제적으로 변수를 설정하는 항목과 연구자가 선택적으로 선택하는 항목을 설정할 수 있다.

## 4. 다항 로지스틱 회귀 : 통계량

1. **케이스 처리 요약** : 이 표에서는 지정한 범주형 변수에 대한 정보를 포함한다.
2. **모형** : 전체 모형에 대한 통계
   1. 유사R-제곱 : Cox 및 Snell, Nagelkerke, McFadden R2 통계를 출력한다.
   2. 단계요약 : 단계적 방법의 각 단계에서 입력되거나 제거된 효과를 요약한다.(모형 대화상자에 단계선택 모형을 지정해야 표시된다.)
   3. 모형 적합 정보 : 적합 모형과 절편만 있는 모형 또는 널 모형을 비교한다.
   4. 정보기준 : Akaike 정보기준(AIC) 과 Schwartz 베이지안 정보기준(BIC)을 인쇄한다.
   5. 셀 확률 : 공분산 패턴과 반응 범주에 의한 관측빈도, 기대빈도(잔차 포함) 및 비율 표를 출력한다.
   6. 분류표 : 관측반응 대 예측반응 값 표를 출력한다.
   7. 적합도 : Pearson 및 우도비 카이제곱 통계를 출력한다. 모든 요인과 공분산 또는 사용자가 정의한 요인 및 공분산의 서브세트에서 지정한 공분산 패턴에 대한 통계를 계산한다.
   8. 단조성 측도 : 일치되는 대응, 비일치되는 대응, 그리고 동률한 대응의 수에 대한 정보를 나타내는 표를 표시한다.(Somer의 D.Goodman과 Kruskal의 감마, Kendall의 타우-a와 일치지수 C 또한 표에 표시된다.)
3. **모수** : 모형 모수와 관련된 통계
   1. 추정값 : 사용자가 지정한 신뢰수준으로 모형 모수의 추정값을 출력한다.
   2. 우도비 검정 : 모형의 부분 효과에 대한 우도비 검정을 출력한다. 전체 모형 검정은 자동으로 출력한다.
   3. 근사 상관 : 모수 추정값 상관 행렬을 출력한다.
   4. 근사 공분산 : 모수 추정값 공분산 행렬을 출력한다.
4. **부-모집단 정의**

요인 및 공분산의 서브세트를 선택하여 셀 확률 및 적합도 검정에서 사용한 공분산 패턴을 정의할 수 있다.

## 5. 다항 로지스틱 회귀 : 수렴기준

1. **반복** : 알고리즘 순환의 최대 횟수, 단계 반분에서 최대 단계 수 , 로그-우도 및 모수 값 변경에 대한 수렴 허용 오차, 반복적 알고리즘 과정의 출력 빈도 및 반복에서 완전한 데이터 분리 또는 완전에 가까운 데이터 분리를 확인해야 하는 프로시저를 지정한다.
   1. 로그–우도 수렴 : 수렴은 로그 우도 함수의 절대 변화량이 지정된 값보다 작다고 가정한다. 값이 0인 경우 이 기준은 적용되지 않으며, 음수가 아닌 값을 지정한다.
   2. 모수 수렴 : 수렴은 모수 추정값의 절대 변화량이 지정된 값보다 작다고 가정한다. 값이 0인 경우 이 기준은 적용되지 않는다.
2. **델타** : 안정적인 알고리즘을 만들고 추정값이 편향되지 않도록 한다. 0이상 1미만의 값을 지정할 수 있고, 이 값은 공분산 패턴에 따라 반응 범주 교차 분석표의 빈 셀에 추가된다.
3. **비정칙성 공차** : 비정칙성 검사에 사용되는 허용 오차 범위를 지정할 수 있다.

## 6. 다항 로지스틱 회귀 : 옵션

1. **산포 척도** : 모수 공분산 행렬의 추정값을 수정할 때 사용하는 산포 척도값을 지정한다.
   1. 편차 : 편차 함수(우도비 카이제곱) 통계를 사용하여 척도값을 추정한다.
   2. Pearson : Pearson 카이제곱 통계를 사용하여 척도값을 추정한다.
   3. 사용자 정의 : 사용자가 임의의 척도값을 지정한다.(단, 값은 양수로 지정한다.)
2. **단계선택 옵션** : 단계적 방법을 사용하여 모형을 작성하는 경우에 이러한 옵션을 사용하여 통계 기준을 제어한다. 모형 대화상자에서 단계선택 모형을 지정하지 않으면 이 옵션은 무시된다.
   1. 입력 확률 : 변수 입력에 필요한 우도비 통계 확률(지정한 확률이 클수록 변수를 모형에 입력하기 쉬움)

⇨ 전진입력 , 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시된다.

* 1. 입력 검정 : 단계적 방법으로 항을 입력하는 방법(우도비 검정과 스코어 검정 중 하나를 선택)

⇨ 전진입력 , 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시된다

* 1. 제거 확률 : 변수 제거에 필요한 우도비 통계 확률(지정한 확률이 클수록 변수가 모형에 남아 있기 쉬움)

⇨ 후진 제거법, 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시된다.

* 1. 제거 검정 : 단계적 방법으로 항을 제거하는 방법(우도비 검정과 Wald 검정 중 하나를 선택)

⇨ 후진제거법, 단계적 전진 또는 단계적 후진을 선택하지 않으면 이 기준은 무시된다.

* 1. 모형 내 최소 다단효과 : 후진 제거법 또는 단계적 후진을 사용하는 경우로, 모형에 포함할 최소 항의 개수를 지정한다. 절편은 모형 항으로 계산되지 않는다.
  2. 모형 내 최대 다단효과 : 전진 입력 또는 단계적 전진을 사용하는 경우로, 모형에 포함할 최대 항의 개수를 지정한다. 절편은 모형 항으로 계산되지 않는다.

1. **계층별로 항의 입력 및 제거 제한** : 모형 항의 포함에 제한을 둘지 말 것인지 여부를 선택한다. 포함할 항의 계층을 사용하려면 먼저 포함할 항의 일부인 모든 저차항이 모형에 있어야 한다.

## 7 다항 로지스틱 회귀 : 저장

저장을 클릭하면 작업 파일에 변수를 저장하고 모형정보를 외부 파일로 내보낼 수 있다.

1. **저장된 변수** 
   1. 반응확률 추정 : 요인 / 공분산 패턴을 반응 범주로 분류하는 확률(반응변수 범주 만큼의 확률 추정이 있으며 25개 까지 저장)
   2. 예측 범주 : 요인/공분산 패턴에 대한 기대확률이 가장 큰 반응 범주
   3. 예측 범주 확률 : 반응확률 추정의 최대값
   4. 실제 범주 확률 : 요인/공분산 패턴을 관측 범주로 분류하는 확률
2. **XML 파일에 모형정보 내보내기** : 모수 추정값과 필요에 따라 해당 공분산을 XML(PMML)형식의 지정된 파일로 내보낼 수 있다 .스코어링 목적으로 이 모형 파일을 사용하여 모형정보를 다른 데이터 파일에 적용할 수 있다.