

판별분석

판별분석(Discriminant Analysis)은 두 개 이상의 모집단에서 추출된 표본들의 정보를 이용하여 이 표본들이 어느 모집단에서 추출된 것인지를 결정해주는 기준을 찾는 분석 방법입니다. 판별분석에는 선형판별분석(Linear Discriminant Analysis) 그리고 이차판별분석(Quadratic Discriminant Analysis)가 있습니다. 판별분석에서 종속변수는 두개 혹은 제한적인 계급을 가지는 질적 데이터여야 합니다. 설명변수는 성향에 관계없이 사용할 수 있습니다. 선형판별분석에서는 설명변수들의 벡터는 각각 다변량 정규분포를 따르고, 변수들 간의 분산-공분산행렬은 동일해야 합니다. 이차판별분석에는 변수들 간의 분산-공분산 행렬은 동일하지 않아도 됩니다.

메뉴 호출하기

- 고급분석 > 분류분석 > 지도 학습 > 판별분석



• 변수설정 탭

판별분석

변수설정

분석옵션

자료분할

출력옵션

데이터

전체변수

id

bweight

lowbw

gestwks

preterm

matage

hyp

sex

>

<

① 종속변수(필수)

설명변수

② 질적변수(선택-1개이상가능)

>

<

③ 양적변수(선택-1개이상가능)

>

<

도움말

재설정

확인

취소

메뉴 요소	설명
1. 종속변수	종속변수에 해당하는 변수를 전체변수로부터 선택할 수 있습니다. 한 개의 변수가 필수적으로 선택되어야 하며 질적변수만 사용이 가능합니다. 종속변수에 결측치가 존재하는 관측치는 분석에서 제외됩니다.
② 질적변수	설명변수 중 질적변수에 해당하는 변수를 전체변수로부터 선택할 수 있습니다. 종속변수와 중복하여 선택할 수 없습니다. 질적변수와 양적변수 중 적어도 하나 이상의 변수를 선택해야 분석이 가능합니다.
3. 양적변수	설명변수 중 양적변수에 해당하는 변수를 전체변수로부터 선택할 수 있습니다. 종속변수와 중복하여 선택할 수 없습니다. 질적변수와 양적변수 중 적어도 하나 이상의 변수를 선택해야 분석이 가능합니다. 설명변수가 양적일 때는 낮은 예측능력을 보일 수 있습니다.

• 분석옵션 탭

판별분석

변수설정
분석옵션
자료분할
출력옵션

1
분석방법

☒ LDA
☐ QDA

2
사전확률

☒ 훈련데이터 내의 집단의 상대빈도
☐ 모든 집단에 대하여 동일한 확률
☐ 사용자정의

집단
확률
적용목록

>

확률의 합 :0
삭제

도움말
재설정

확인
취소

메뉴 요소	설명
1. 분석옵션 > 분석방법	<p>판별분석 기법 중 하나를 선택합니다.</p> <ul style="list-style-type: none"> LDA (Default) : 선형판별분석을 수행합니다. K개의 클래스에 대해 공통 분산을 가정하게 됩니다. 훈련 자료 수가 적을 경우, 변동을 줄이는 것에 유리한 LDA를 주로 사용합니다. QDA : LDA와 달리 k개의 클래스 각각 다른 공분산행렬을 가지게 합니다. 데이터 수가 많아서 변동에 대한 우려가 적거나, LDA의 공분산 가정이 비현실적으로 판단될 때 QDA를 사용합니다.
2. 사전확률	<p>사후확률을 계산할 때 사용되는 종속변수의 각 집단에 대한 사전확률을 어떻게 정의할지 선택합니다.</p> <ul style="list-style-type: none"> 훈련데이터 내의 집단의 상대빈도 (Default) : 사전확률을 훈련자료 내 종속변수 간 상대빈도를 사용해 정의합니다. 모든 집단에 대하여 동일한 확률 : 사전확률로 각 범주에 동일한 확률을 부여합니다. 사용자정의 : 사용자가 직접 각 집단에 대한 확률을 부여합니다. '사용자정의'를 선택할 경우, '집단' 콤보박스가 활성화됩니다. 콤보박스에 종속변수의 범주가 나타나게 되며 각 범주에 대해 사용자가 직접 확률을 적고 [>] 버튼을 클릭하면 적용목록에 사전확률이 추가됩니다. 각 범주에는 0에서 1 사이의 확률을 부여할 수 있으며, 적용된 확률의 합은 1이 되어야합니다. 입력된 확률 값은 소수점 16자리에서 반올림하여 적용됩니다.

• 자료분할 탭

판별분석

변수설정

분석옵션

자료분할

출력옵션

변수목록

id
bweight
lowbw
gestwks
preterm
matage
hyp
sex

① 훈련 및 검증(필수)

② 분할검증

☒ 모든 데이터를 훈련에 이용

☐ 비율에 따라 임의로 분할

☐ 변수로 분할

훈련(train) 자료

%

시험(test) 자료

%

☐ 교차검증

☐ Leave-one-out 교차검증

☒ K-fold 교차검증

K

10

④ 예측(선택)

분할변수(1-예측, 2-훈련 및 검증)

>

<

도움말

재설정

확인

취소

메뉴 요소	설명
① 훈련 및 검증	<p>판별분석모형 적합에 사용될 데이터를 훈련자료(training data)와 시험자료(test data)로 분할하는 방식으로 다음 2가지 옵션 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none">분할검증 (Default) : 훈련자료와 시험자료로 분할된 자료로 모형을 1회 검증하는 방법입니다.교차검증 : 훈련자료와 시험자료를 변경해가며 여러 차례 반복 검증하는 방법입니다.
② 분할검증	<p>[분할검증]을 선택하는 경우 다음의 3가지 옵션이 활성화되어 이 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none">모든 데이터를 훈련에 이용 (Default) : 시험자료 없이 모든 개체를 모형 적합에 사용합니다.비율에 따라 임의로 분할 : 훈련자료와 시험자료의 비율을 설정하여 임의로 분할하는 방식입니다. Default 값은 훈련자료 70%, 시험자료가 30% 입니다. 사용자는 훈련자료에 0~100을 입력할 수 있으며, 시험자료에는 100에서 입력한 값을 뺀 수치가 자동으로 입력됩니다. 임의로 분할된 개체들 중 훈련자료와 시험자료의 인덱스를 저장하려면 [출력옵션]-[저장]-[자료분할지표]를 선택합니다.변수로 분할 : 훈련자료와 시험자료로 사용될 개체가 결정되어 있는 경우 이 옵션을 선택합니다. 이때, 훈련자료에 해당하는 개체는 1, 시험자료에 해당하는 개체는 2의 값을 갖는 인덱스 변수를 분할변수로 지정해주어야 합니다.

• 자료분할 탭

판별분석

변수설정 분석옵션 **자료분할** 출력옵션

변수목록

id
bweight
lowbw
gestwks
preterm
matage
hyp
sex

① 훈련 및 검증(필수)

☒ 분할검증

② ☒ 모든 데이터를 훈련에 이용

☐ 비율에 따라 임의로 분할

훈련(train) 자료 %

시험(test) 자료 %

☐ 변수로 분할

분할변수(1-훈련, 2-시험)

>

<

③ 교차검증

☐ Leave-one-out 교차검증

☒ K-fold 교차검증 K

④ 예측(선택)

분할변수(1-예측, 2-훈련 및 검증)

>

<

도움말 재설정 **확인** 취소

메뉴 요소	설명
③ 교차검증	<p>[교차검증]을 선택하는 경우 다음의 2가지 옵션이 활성화되어 이 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> Leave-one-out 교차검증 : 한 개체를 시험자료로 사용하고 나머지 개체를 모두 훈련자료로 하여 모형을 적합하는 방식으로 모든 개체에 대해 이 과정을 반복한 뒤, 전체 개체 수만큼의 모형으로부터 얻은 정확도의 평균을 모형의 최종 정확도로 계산합니다. K-fold 교차검증 : 전체 개체를 K개의 그룹으로 임의로 분할하여, 하나의 그룹을 시험자료로 사용하고 나머지 그룹을 모두 훈련자료로 하여 모형을 적합하는 방식으로 K개의 그룹에 대해 이 과정을 반복한 뒤, 그룹 수만큼의 모형으로부터 얻은 정확도의 평균을 모형의 최종 정확도로 계산합니다. <p>- K : [교차검증]-[K-fold 교차검증]을 선택할 경우 활성화됩니다. K-fold 교차검증에 사용할 K의 값을 입력합니다. 2 이상의 정수만 입력 가능하며, 전체 개체 수보다 더 큰 정수가 입력되는 경우 자동으로 Leave-one-out 교차검증을 실시합니다. Default는 10입니다.</p>
④ 예측 > 분할변수	<p>판별분석모형 적합에 사용될 훈련 및 검증 데이터와 해당 모형으로부터 예측값을 얻을 예측 데이터가 분할되어 있는 경우 사용됩니다. 훈련 및 검증에 사용되는 개체는 2, 예측에 사용되는 개체는 1의 값을 갖는 인덱스 변수를 분할변수로 지정해주어야 합니다. 예측분할변수를 지정하지 않아도 분석이 가능합니다. 예측분할변수가 지정된 경우, 예측에 해당하는 개체에 해당하는 예측값이 엑셀 시트에 "Predicted_pred_DA"라는 변수명으로 저장됩니다.</p>

출력옵션 탭

판별분석

변수설정

분석옵션

자료분할

출력옵션

저장

훈련자료

☐ 적합값

시험자료

☐ 예측값

☐ 자료분할지표

도움말

재설정

확인

취소

메뉴 요소	설명
1. 적합값	적합값을 괄호 안의 변수명으로 저장합니다. (Fitted_DA_Train)
2. 예측값	[자료분할] 탭에서 '비율에 따라 임의로 분할' 또는 '변수로 분할' 을 택할 경우 예측값이 활성화됩니다. 예측값을 괄호 안의 변수명으로 저장합니다. (Predicted_test_DA)
③ 자료분할지표	각 관측값이 훈련 혹은 시험자료 중 어떤 자료로 사용되었는지 여부를 괄호 안의 변수명으로 저장합니다. (Partition_idx_DA)