[일반화 선형모형 (1)]

1. 일반화 선형 모형

회귀분석

선형모형 : Linear Model

일반화 선형 모형 : Generalized Linear Model

지수 분포족 : 정규분포, 이항분포, 포아송분포, 감마분포, 역감마분포

선형모형 : 정규분포를 가정함

일반화 선형모형 : 지수분포족을 가정함

2. 연결함수 / 분산함수

연결함수 : link function - 선형 예측치와 평균 모수 사이의 관계를 연결해주는 함수

이항분포의 연결함수 : 로짓(logit), 프라빗(probit), cloglog(complementary log-log)

정준연결함수

분산함수

3. 최대우도법

최소자승법

최대우도법

포아송 분포의 우도

모형과 우도

포화모형 : 모형 예측값들이 관측값들과 모두 동일해지도록 모수의 수를 증가시킨 모형

4. 잔차 / 이탈도

이탈도 : 포화 모형에서의 값이 0이 되도록 하고 두 배의 음로그우도를 보정하여, 산포모수를 곱한 값.

이탈도 잔차

피어슨 잔차

AIDS data : 잔차 - 이탈도잔차/피어슨잔차

확장지수분포족

지수분포족 : 확률밀도함수로 분포를 특정

확장지수분포족 : 평균과 분산 사이의 관계만으로 분포를 특정

의사포아송 분포 (quasi poission) 의사이항(quasi binomial) 분포

잔차제곱합, 이탈도 : 특정 모형에 대한 자료의 적합/부적합 척도

좋은 모형 = 자료 적합도가 높고 가능한 단순한 모형

AIC (Akaike Information Criterion)

5. R에서 GLM 적용법

함수 : lm() , glm()

formula 인자 : (일반화) 선형 모형에서 모형을 설정하는 인자.

절편없는 모형

요인 변수의 처리

특별한 표현

결과 사용 예

glm() 클래스에 대한 메쏘드

: anova, coef, deviance, fiited, print, predict, plot, resid, plot, update

MASS 패키지 함수

addterm : 기존 적합된 모형에 하나의 항을 추가한 모형들을 제시

dropterm : 기존 적합된 모형에 하나의 항을 뺀 모형들을 제시

stepAIC : AIC 를 기준으로 한 stepwise 모형 선택

vcov : 모수 추정값들에 대한 분산 공분산 행렬

[일반화 선형모형 (2)]

일반화 선형 모형 : 정규분포/ 이항분포 / 포아송분포 / 다항분포의 예

1. 정규분포 경우의 예

IRIS data : 붓꽃 자료

꽃받침의 폭과 길이

두 가지 모형

붓꽃 자료 : 모형 A , 모형 B

두 모형의 비교 : 이탈도 / 2*음로그우도 / 잔차의 자유도 / 모형 모수 개수 / AIC

2. 이항분포 경우의 예

이항분포 : 담배나방 자료

담배나방 자료 : 변수

담배나방 자료의 준비

담배나방 자료 : 두 모형의 비교 - 교호작용의 유,무 / 이탈도, AIC

3. 포아송 분포 경우의 예

갈라파고스 자료 : 갈라파고스 제도에 속하는 30개의 섬들에 대한 생태적 특성을 기록한 자료

library(faraway) 의 gala 데이터셋

- *섬이름*
- *섬에서 발견된 식물 종의 수*
- *토착종수*
- *섬의 면적*
- *섬의 최고점 해발고도*
- *가장 가까운 섬과의 거리*
- *산타크루즈 섬과의 거리*
- *인접한 섬의 면적*

갈라파고스 제도의 배치. 과산포성, 산포모수의 추정

의사포아송 분포

4. 다항 분포 경우의 예

다항분포 로지스틱 회귀모형

교육 프로그램 선택자료 progselection. txt

세 모형의 비교 : 독립변수 / 이탈도 / AIC

library(nnet)

* 모형 A : prog ~ ses -1

multinom(prog ~ ses -1, data = hdata)

* 모형 B : prog ~ ses + read -1

multinom(prog ~ ses + read -1, data = hdata)

* 모형 C : prog ~ read

multinom(prog ~ read, data = hdata)