

# 3강 회귀모형 II

통계·데이터과학과 장영재 교수



# 01 회귀모형관련 R 함수

## 02 R 사용 예제



# 1. 회귀모형관련 R 함수



# 01 회귀모형관련 R 함수

## 1 선형회귀모형 함수

### ■ 함수의 구조

`lm(formula, data, na.action)`

### ■ 기능

함수 `lm`은 선형모형을 적합하기 위해 사용

### ■ 옵션

- `formula`: 적합할 모형의 공식으로 목표변수 ~ 입력변수의 선형결합 형식으로 구성. 입력변수의 선형결합은 `입력변수1 + 입력변수2 + ...` 형태로 나열한다. 목표변수를 제외한 입력 데이터의 모든 변수를 입력변수로 사용할 경우 `목표변수 ~ .` 형태로 표현
- `data`: 사용할 데이터 프레임의 이름을 지정
- `na.action`: 결측치 처리 옵션. `na.fail`는 결측치가 있으면 오류 메시지 발생, `na.omit`와 `na.exclude`는 결측치가 있는 관측치 제외하지만 `na.exclude`는 잔차 또는 예측 값을 계산할 때 NAs를 제공



# 01 회귀모형관련 R 함수

## 2 로지스틱회귀모형 함수

### ■ 함수의 구조

`glm(formula, family, data, na.action)`

### ■ 기능

함수 `glm`은 일반화선형모형을 적합하기 위해 사용

### ■ 옵션

- `formula`: 적합할 모형의 공식으로 목표변수 ~ 입력변수의 선형결합 형식으로 구성. 입력변수의 선형결합은 입력변수1 + 입력변수2 + ... 형태로 나열한다. 목표변수를 제외한 입력 데이터의 모든 변수를 입력변수로 사용할 경우 목표변수 ~ . 형태로 표현
- `family`: 분포와 연결함수 설정할 때 사용. 선형모형은 `family = gaussian` 또는 `family = gaussian(link="identity")`, 로지스틱 회귀모형은 `family = binomial` 또는 `family = binomial(link="logit")`을 사용. 다른 연결함수 `link="probit"` 또는 `link="cloglog"`를 사용.



# 01 회귀모형관련 R 함수

## 2 로지스틱회귀모형 함수

### ■ 함수의 구조

`glm(formula, family, data, na.action)`

### ■ 기능

함수 `glm`은 일반화선형모형을 적합하기 위해 사용

### ■ 옵션

- `data`: 사용할 데이터 프레임의 이름을 지정
- `family`: 분포와 연결함수를 설정할 때 사용. 선형모형은 `family= gaussian` 또는 `family= gaussian(link= "identity")`, 로지스틱회귀모형은 `family= binomial` 또는 `family= binomial(link= "logit")`을 사용. 다른 연결함수 `link= "probit"` 또는 `link= "cloglog"`를 사용할 수도 있음
- `na.action`: 결측치 처리 옵션. `na.fail`는 결측치가 있으면 오류 메시지 발생, `na.omit`와 `na.exclude`는 결측치가 있는 관측치 제외하지만 `na.exclude`는 잔차 또는 예측 값을 계산할 때 NAs를 제공



## 01 회귀모형관련 R 함수

### 3 모형 선택 함수

#### ■ 함수의 구조

`step(object, direction = c("both", "backward", "forward"))`

#### ■ 기능

함수 `step`은 AIC를 이용한 모형 선택을 위해 사용

#### ■ 옵션

- `object`: 함수 `lm`, `glm` 등에서 생성한 `object`를 입력
- `direction`: 변수 선택 방법 ("`both`"=단계적변수선택, "`backward`"=후진소거법, "`forward`"=전진선택법)을 지정





## 01 회귀모형관련 R 함수

### 4 예측 함수

#### ■ 함수의 구조

`predict(object, newdata, type)`

#### ■ 기능

함수 `predict`은 다양한 모형 적합 결과로부터 예측값을 계산할 때 사용

#### ■ 옵션

- `object`: 함수 `lm`, `glm` 등에서 생성한 `object`를 입력
- `newdata`: 예측할 변수들로 구성된 데이터 프레임을 입력
- `type`: 예측 형태를 입력하는 것으로, 목표 값을 예측할 때 `type="response"` 사용





## 2. R 사용 예제



다음시간 안내

# 4강. 의사결정나무 I



한국방송통신대학교