06 강

데이터분석방법론2

로지스틱회귀모형 (2)

통계·데이터과학과 이기재 교수

학습목차

학습개요 및 목표

지난 강의에서는 이항자료 분석에서 널리 사용되는
로지스틱회귀분석을 GLM 관점에서 살펴보고,
로지스틱회귀모형의 해석과 통계적 추론에 대해서 공부했습니다.
이번 강의는 지난 강의에 이어서 로지스틱회귀모형의 구축과 활용에 대해 살펴보겠습니다.

1. 범주형 설명변수를 갖는 로지스틱회귀모형을 적용할 수 있다.
2. 다중 로지스틱회귀모형을 적용할 때
   효과에 대한 요약, 예측력 요약 등을 설명할 수 있다.
3. 다중 로지스틱회귀모형을 실제 데이터에 적용하고
   그 결과를 해석할 수 있다.

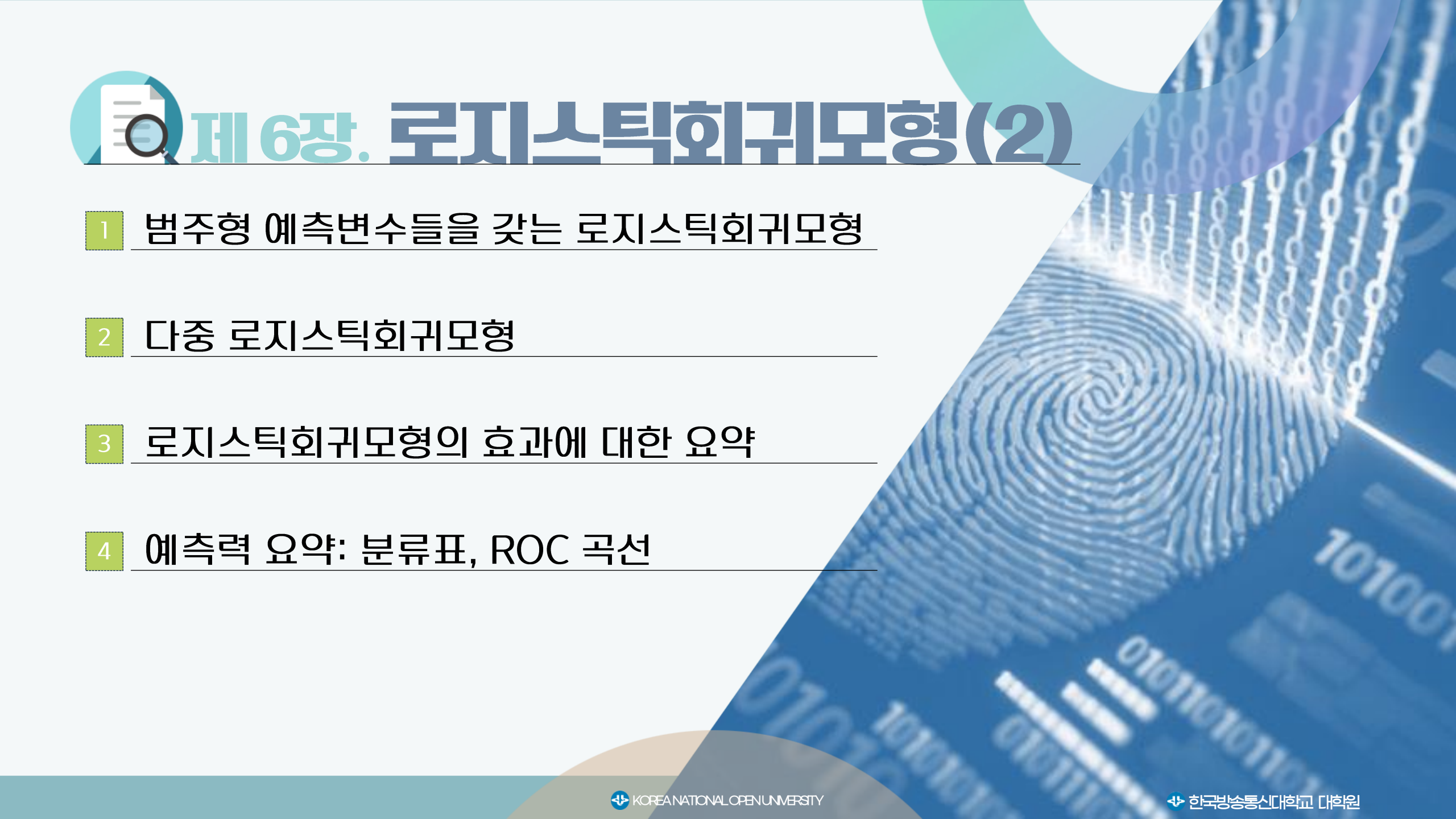
제 6장. 로지스틱회귀모형(2)

1 범주형 예측변수들을 갖는 로지스틱회귀모형

2 다중 로지스틱회귀모형

3 로지스틱회귀모형의 효과에 대한 요약

4 예측력 요약: 분류표, ROC 곡선



01

제 6장. 로지스틱회귀모형(2)

범주형 예측변수들을 갖는 로지스틱회귀모형

1. 범주형 예측변수를 표현하는 지시변수

로지스틱회귀모형이란?

- 로지스틱회귀모형은 회귀모형과 마찬가지로
  여러 설명변수를 갖는 모형으로 확장 가능함

- 예측변수 中
  일부 또는 전부가 질적변수(범주형 변수)일 수 있음

1. 범주형 예측변수를 표현하는 지시변수

- 두 개의 예측변수 X, Z 와
  반응변수 Y 가 각각 (0,1)의 값을 갖는 이항변수인 경우

- X 와 Z 의 주효과를 갖는 로지스틱회귀모형

$$\mathrm{logit}[P(Y=1)] = \alpha + \beta_1 x + \beta_2 z$$

x, z : 지시변수 또는 가변수(Dummy Variable)

1. 범주형 예측변수를 표현하는 지시변수

- 모형 $logit[P(Y=1)] = \alpha + \beta_1 x + \beta_2 z$ 에서 가변수 값에 따른 로짓값

x	z	로짓
0	0	α
1	0	$\alpha + \beta_1$
0	1	$\alpha + \beta_2$
1	1	$\alpha + \beta_1 + \beta_2$

- Z 가 주어졌을 때 $x = 1$ 에서 "성공"일 오즈는 $x = 0$ 에서 "성공"일 오즈의 $\exp(\beta_1)$ 배임

$$\frac{\exp(\alpha+\beta_1)}{\exp(\alpha)} = \exp(\beta_1) \quad \frac{\exp(\alpha+\beta_1+\beta_2)}{\exp(\alpha+\beta_2)} = \exp(\beta_1)$$

1. 범주형 예측변수를 표현하는 지시변수

- 모형에서 교호작용이 없다는 것은
  Z의 두 수준에서 구한 부분 분할표에 대한
  오즈비 값들이 동일하다는 것을 의미함

동질연관성 만족

2. 예제 : 마리화나 사용에 대한 조사

Note

- 오하이오 데이톤 근처의 고등학교 고학년 학생 대상 조사
- 그룹화 자료를 통해 마리화나 사용여부(1=예, 0=아니오)를 예측 목적
- 성별과 인종의 주효과를 갖는 로지스틱회귀모형을 적합시킨 결과

2. 예제 : 마리화나 사용에 대한 조사

인종	성별	마리화나 사용	
		예	아니오
백인	여자	420	620
	남자	483	579
다른 인종	여자	25	55
	남자	32	62

X = 성별

Z = 인종

Y = 마리화나 사용 (1=예, 0=아니오)

2. 예제 : 마리화나 사용에 대한 조사

SAS 프로그램

```
DATA marijuana;
INPUT RACE $ GENDER $ yes n @@;
CARDS;
white M 483 1062 white F 420 1040
black M  32 94 black F  25  80
RUN;

PROC GENMOD order=data;
   class RACE GENDER;
   model yes/n = GENDER RACE/ dist=bin type3 lrci residuals obstats;
RUN;
```

2. 예제 : 마리화나 사용에 대한 조사

분석 결과

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-0.8303	0.1685	-1.1669	-0.5050	24.27	<.0001
GENDER	M	1	0.2026	0.0852	0.0358	0.3697	5.66	0.0174
GENDER	F	0	0.0000	0.0000	0.0000	0.0000	.	.
RACE	white	1	0.4437	0.1677	0.1199	0.7783	7.00	0.0081
RACE	black	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
GENDER	1	5.67	0.0173
RACE	1	7.28	0.0070

2. 예제 : 마리화나 사용에 대한 조사

적합 결과

$$\text{logit}(\hat{\pi}) = -0.8303 + 0.2026 \times \text{GENDER} + 0.4437 \times \text{RACE}$$

마리화나 사용과 성별 사이의 조건부 오즈비는
$\exp(0.2026) = 1.22$ 로 추정됨

2. 예제 : 마리화나 사용에 대한 조사

R 프로그램

```
------------------------------------------------------------------------------
> Marijuana <- read.table("http://www.stat.ufl.edu/~aa/cat/data/Marijuana.dat",
+                         header=TRUE)
> Marijuana
   race gender yes  no
1 white female 420 620
2 white   male 483 579
3 other female  25  55
4 other   male  32  62
> fit <- glm(yes/(yes+no) ~ gender + race, weights = yes + no,
+            family=binomial, data=Marijuana)
> summary(fit)
             Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)   -0.83035     0.16854   -4.927  8.37e-07
gendermale     0.20261     0.08519    2.378   0.01739
racewhite      0.44374     0.16766    2.647   0.00813
---
    Null deviance: 12.7528  on 3  degrees of freedom
Residual deviance:  0.0580  on 1  degrees of freedom
------------------------------------------------------------------------------
```

H_0: $\beta_1 = \beta_2 = 0$을 검정하기 위한 가능도비 검정통계량
➔ 영이탈도와 잔차 이탈도의 차이값, 12.75-0.06, 자유도 df=3-1

The Null deviance and the residual deviance are defined as:

$$NullDeviance = 2(ll(SaturatedModel) - ll(NullModel))$$

with $df = df_{Sat} - df_{Null}$.

$$ResidualDeviance = 2(ll(SaturatedModel) - ll(ProposedModel))$$

with $df = df_{Sat} - df_{Res}$.

3. 2 X 2 X K 분할표에 대한 검정법

Cochran-Mantel-Haenszel 검정법

- Multi-center clinical trials 사례

Center(Z)	처리(X)	반응(Y)	
		S	F
1	1		
	2		
2	1		
	2		
⋮	⋮	⋮	
K	1		
	2		

$$logit[P(Y=1)] = \alpha + \beta x + \beta_1 c_1 + \beta_2 c_2 + \cdots + \beta_{k-1} c_{k-1}$$

여기서 x 는 X의 두 수준에 대한 가변수임

3. 2 X 2 X K 분할표에 대한 검정법

Cochran-Mantel-Haenszel 검정법

- 모형에 대한 다른 형태 표현
  - $\mathrm{logit}[P(Y=1)] = \alpha + \beta x + \beta_k^z$
  - β_k^z : 센터 K 의 효과
    (보통 마지막 센터에 대한 상대적 크기로 표현)
  - x : X 의 두 수준에 대한 가변수

- $\exp(\beta)$: K 개 분할표에서 Z 를 통제했을 때의 X, Y 의 공통 오즈비
- Z 를 통제했을 때 X, Y 간 조건부독립성 성립
  $\Leftrightarrow \beta = 0$ (X, Y의 오즈비 = 1)

3. 2 X 2 X K 분할표에 대한 검정법

Cochran-Mantel-Haenszel 검정법

- "$H_0 : \beta = 0$"에 대한 검정을 통해서 조건부 독립성 검정 가능
  - 가능도비 검정
  - Wald 검정
  - Cochran-Mantel-Haenszel 검정 (CMH검정)

02

제 6장. 로지스틱회귀모형(2)

다중 로지스틱회귀모형

개요

다중 로지스틱회귀모형이란?

- Y : 이항반응변수, $\pi = P(Y=1)$
- $x_1, x_2, \cdots, x_k$ 개의 설명변수가능도비 검정

$$logit[P(Y=1)] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

$$\Leftrightarrow \pi = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}$$

- β_i : 다른 변수들을 통제할 경우, x_i 가 미치는 효과
- e^{β_i} : 다른 설명변수가 고정되었을 때
  X_i 가 한 단위 증가할 때 오즈의 증가 비(Ratio)

1. 참게 자료에 대한 로지스틱회귀모형 적합

$$Y=\begin{cases}1, & \text{한 마리 이상의 부수체를 보유한 경우}\\ 0, & \text{부수체가 없는 경우}\end{cases}$$

x : 너비

색깔에 대한 가변수 : 밝은색, 약간 밝은색, 중간색,
약간 어두운색, 어두운 색

- c_2 = 1 ➔ 중간색, 그렇지 않은 경우 “0”
- c_3 = 1 ➔ 약간 어두운색, 그렇지 않은 경우 “0”
- c_4 = 1 ➔ 어두운색, 그렇지 않은 경우 “0”

참고

- 밝은 색의 참게 : 없음
  ➔ 나이가 많을수록 어두운 색을 띔
- 약간 밝은 색인 경우는 $c_2 = c_3 = c_4 = 0$ 인 경우임

1. 참게 자료에 대한 로지스틱회귀모형 적합

```
> fit <- glm(y ~  width + factor(color), family=binomial, data=Crabs)
> summary(fit)

                Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)    -11.38519     2.87346   -3.962  7.43e-05
width            0.46796     0.10554    4.434  9.26e-06
factor(color)2   0.07242     0.73989    0.098     0.922
factor(color)3  -0.22380     0.77708   -0.288     0.773
factor(color)4  -1.32992     0.85252   -1.560     0.119
---
    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 187.46  on 168  degrees of freedom
```

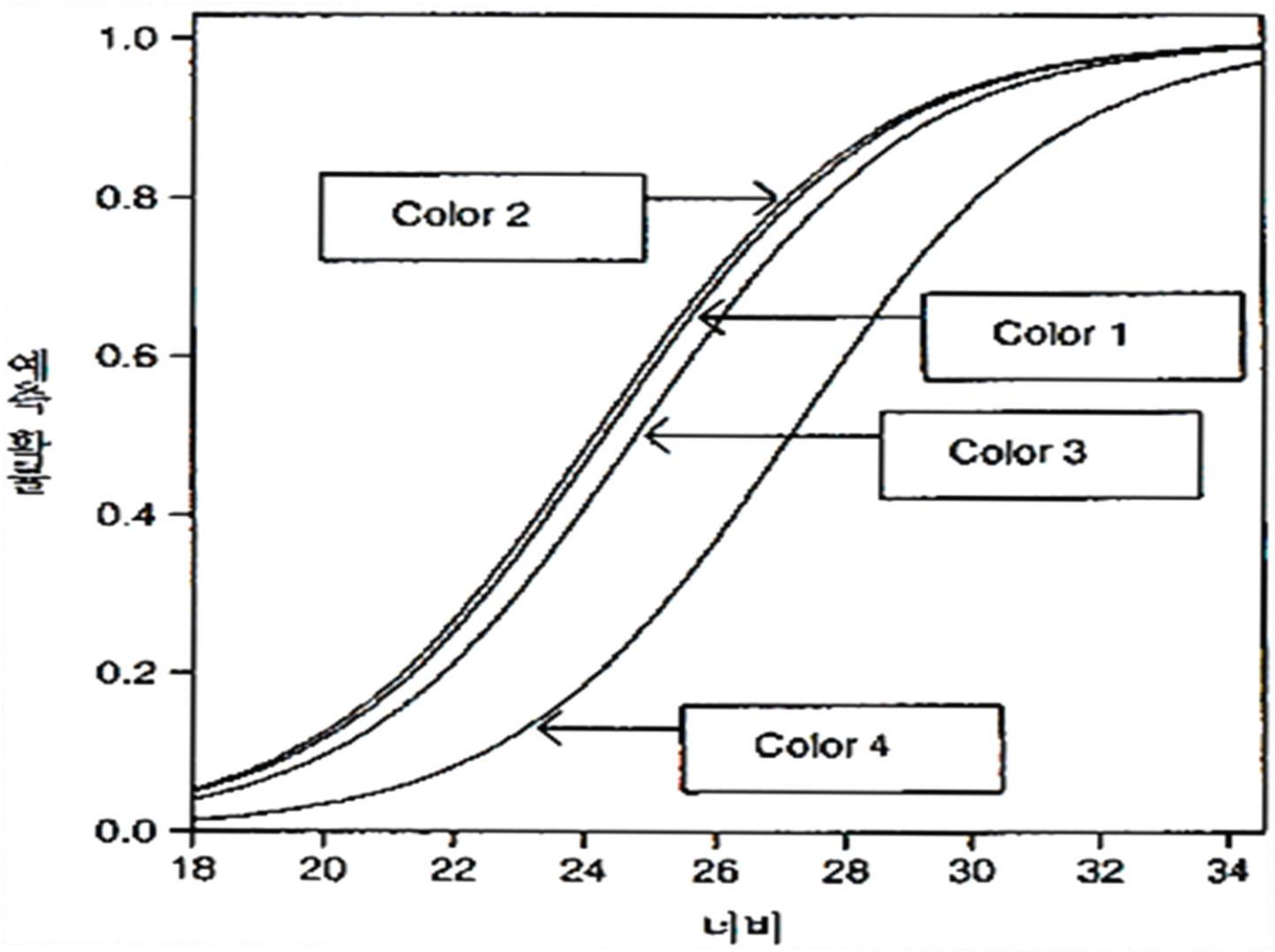
1. 참게 자료에 대한 로지스틱회귀모형 적합

- $logit[P(Y=1)] = \alpha + \beta_1 x + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4$
  - $logit[\hat{P}(Y=1)] = -11.385 + 0.468x + 0.072c_2 - 0.224c_3 - 1.330c_4$
  - 어두운 색 참게에 대한 예측결과 ($c_2 = c_3 = 0,\ \ c_4 = 1$)
  - $logit[\hat{P}(Y=1)] = (-11.385 - 1.330) + 0.468x$
  - $x = \bar{x} = 26.3$인 어두운 색 참게가 부수체를 가질 확률

$$\hat{P}(Y=1) = \frac{\exp[-12.715 + 0.468(26.3)]}{1 + \exp[-12.715 + 0.468(26.3)]} = 0.399$$

1. 참게 자료에 대한 로지스틱회귀모형 적합

- 너비와 색깔을 예측변수로 갖는 로지스틱회귀모형 분석결과



1. 참게 자료에 대한 로지스틱회귀모형 적합

- 참게의 너비가 주어진 경우
  “중간색” 참게가 부수체를 가질 오즈는
  “약간 밝은 색” 참게에 비해서 1.075배 큼

$$\hat{\beta}_2 = 0.0724, \quad e^{0.0724} = 1.075$$

Note

- $logit[P(Y=1)] = \alpha + \beta_1 x + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4$
- 색깔과 너비 사이의 교호작용이 없는 것으로 간주하여 분석한 것임
- 즉, 참게의 너비(x)가 π에 미치는 영향은 색깔과 무관하게 동일하다고 가정한 것임

2. 어떤 항이 필요한가를 보는 모형 비교

- $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ 검정
  (너비가 주어진 경우, Y는 색깔과 독립)
  - 가능도비 검정 통계량

$$2(L_1 - L_0) = 194.45 - 187.46 = 6.99$$

$$df = 171 - 168 = 3, P-\text{값} = 0.07$$

➔ 참게 색깔의 효과가 뚜렷하지는 않지만 완전히 무시할 수는 없음

➔ 색깔에 대한 예측 변수를 모형에 포함하는 것이 바람직함

2. 어떤 항이 필요한가를 보는 모형 비교

- $H_0 : \beta_1 = 0$ 검정
  - $\hat{\beta}_1 = 0.468, \quad SE = 0.106, \quad P-$ 값 < 0.001
  - 색깔이 주어진 경우, 너비가 $x+1$ 인 경우의 오즈는 너비가 x 인 경우의 오즈에 비해서 $e^{0.468} = 1.597$ 배 큼

2. 어떤 항이 필요한가를 보는 모형 비교

```
> summary(glm(y ~ width, family=binomial, data=Crabs))
             Estimate  Std. Error  z value   Pr(>|z|)
(Intercept)  -12.3508      2.6287   -4.698   2.62e-06
width          0.4972      0.1017    4.887   1.02e-06
---
    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 194.45  on 171  degrees of freedom # deviance=187.46 when
                                                     # color also in model
> library(car)
> Anova(glm(y ~ width + factor(color), family=binomial, data=Crabs))
             LR Chisq  Df  Pr(>Chisq)
width          24.6038   1   7.041e-07 # LR test of width effect
factor(color)   6.9956   3     0.07204 # LR test of color effect  .
```

3. 순서형 예측변수를 양적변수로 분석

1. 색깔의 범주점수로 $c = \{1, 2, 3, 4\}$를 사용하는 경우

- 색깔 범주에 단조성을 만족하는 단조점수를 부여하여 색깔에 대해서 선형효과를 갖도록 함
- $$logit[P(Y=1)] = \alpha + \beta_1 x + \beta_2 c$$
  $$\Rightarrow logit[\hat{P}(Y=1)] = -10.071 + 0.458x - 0.509c$$

3. 순서형 예측변수를 양적변수로 분석

```
> fit2 <- glm(y ~  width + color, family=binomial, data=Crabs)
> summary(fit2) # color treated as quantitative with scores (1, 2, 3, 4)
             Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)  -10.0708      2.8068   -3.588  0.000333
width          0.4583      0.1040    4.406  1.05e-05
color         -0.5090      0.2237   -2.276  0.022860
---
    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 189.12  on 170  degrees of freedom

> anova(fit2, fit, test="LRT") # likelihood-ratio test comparing models
Model 1: y ~  width + color
Model 2: y ~  width + factor(color)
  Resid. Df   Resid. Dev  Df  Deviance  Pr(>Chi)
1       170       189.12
2       168       187.46   2    1.6641    0.4351
```

3. 순서형 예측변수를 양적변수로 분석

1. 색깔의 범주점수로 $c = \{1, 2, 3, 4\}$를 사용하는 경우

주어진 너비에서 색깔 범주가
하나씩 어두운 쪽으로 변할 때마다 부수체를 가질 오즈는
$\exp(-0.509) = 0.60$ 배씩 감소함

- 각 색깔별로 서로 다른 모수를 갖는 복잡한 모형과의 비교
  - 가능도비 통계량 $= 2(L_1 - L_0) = 1.664$

$$df = 2,\ P-\text{값} = 0.44$$

간단한 모형
(색깔을 양적 변수로 간주한 경우)이 적합함

3. 순서형 예측변수를 양적변수로 분석

2. 색깔 변수를 간단하게 정의하여 적용하는 경우

- $c_4 = \begin{cases} 1 & \text{어두운 색 참게} \\ 0 & \text{어두운 색이 아닌 참게} \end{cases}$

- $\mathrm{logit}[P(Y=1)] = \alpha + \beta_1 x + \beta_2 c_4$

  $\hat{\beta}_2 = -1.301 \quad (SE = 0.526)$

  ⇔ 주어진 너비에서 “어두운 색이 아닌” 참게가 부수체를 가질 오즈는 “어두운 색” 참게가 부수체를 가질 오즈의 exp(1.301)=3.673배

3. 순서형 예측변수를 양적변수로 분석

2. 색깔 변수를 간단하게 정의하여 적용하는 경우

- 모형 적합 분석

  H_0 : 간단한 모형(가변수가 1개인 경우)

  H_a : 복잡 모형(가변수가 3개인 경우)

가능도비 검정통계량

=187.96 - 187.46=0.501 (df=2)

"간단한 모형이 적합함"

3. 순서형 예측변수를 양적변수로 분석

```
> Crabs$c4 <- ifelse(Crabs$color == 4, 1, 0) # indicator for color cat. 4
> # or could use I(Crabs$color == 4) to directly define indicator var.
> fit3 <- glm(y ~ width + c4, family=binomial, data=Crabs)
> summary(fit3)

             Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)  -11.6790      2.6925   -4.338  1.44e-05
width          0.4782      0.1041    4.592  4.39e-06
c4            -1.3005      0.5259   -2.473    0.0134
---
    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 187.96  on 170  degrees of freedom

> anova(fit3, fit, test="LRT") # likelihood-ratio test comparing models
Model 1: y ~ width + c4
Model 2: y ~ width + factor(color)
  Resid. Df  Resid. Dev  Df  Deviance  Pr(>Chi)
1       170      187.96
2       168      187.46   2   0.50085    0.7785
```

4. 교호 작용을 포함한 모형

$$logit[P(Y=1)] = \alpha + \beta_1 x + \beta_2 c_4 + \beta_3 (x \times c_4)$$

색깔	너비 효과	가변수 값
어두운 색	$\beta_1 + \beta_3$	$c_4 = 1$
약간 어두운 색	β_1	$c_4 = 0$
중간색	β_1	$c_4 = 0$
약간 밝은색	β_1	$c_4 = 0$

4. 교호 작용을 포함한 모형

$$logit[P(Y=1)] = \alpha + \beta_1 x + \beta_2 c_4 + \beta_3 (x \times c_4)$$

```
> glm(y ~ width + c4 + width:c4, family=binomial, data=Crabs)

(Intercept)         width            c4     width:c4
   -12.8117        0.5222        6.9578       -0.3217

Null Deviance:      225.76
Residual Deviance:  186.79
```

$$\Rightarrow logit[\hat{P}(Y=1)] = -12.812 + 0.522x + 0.6958c_4 - 0.322(x \times c_4)$$

4. 교호 작용을 포함한 모형

- H_0 : 교호작용 없음($\beta_3 = 0$) 검정
- 가능도비 검정통계량

=187.96 - 186.79 = 1.17 (df=1, P=0.28)

교호작용에 대한 Weak Evidence!!

4. 교호 작용을 포함한 모형

모형의 선택 과정에서 고려사항

- 자료에 대한 적합성
  : 모형이 복잡해질수록 유리
- 적합된 모형의 해석의 용이성
  : 모형이 간단할수록 유리

확증적(Confirmatory) 연구와
탐색적(Exploratory) 연구

03

제 6장. 로지스틱회귀모형(2)

로지스틱 회귀모형의 효과에 대한 요약

1. 확률에 기초한 해석

Note

- 예측변수 x_j의 효과 설명 방법

  다른 예측변수의 값을 표본평균으로 고정시킨 후 x_j의
  가장 작은 값과 큰 값에서 $\hat{P}(Y=1)$을 구해 비교해 보는 것

$$\Rightarrow logit[\hat{P}(Y=1)] = -12.812 + 0.522x + 0.6958c_4 - 0.322(x \times c_4)$$

2. 주변 효과와 그 평균

Note

- 양적 예측변수의 상대적으로 작은 변화가 확률에 미치는 영향은 근사적으로 직선의 기울기로 사용하여 파악
- 다른 설명변수를 고정했을 때 설명변수 x_j가 한 단위 증가할 때마다 근사적으로 $\hat{\pi}$는 $\hat{\beta}_j \hat{\pi}(1-\hat{\pi})$ 만큼 변화한다.

```
------------------------------------------------------------------------------
> fit3 <- glm(y ~  width + c4, family=binomial, data=Crabs)
> library(mfx)
> logitmfx(fit3, atmean=FALSE, data=Crabs) # with atmean=TRUE, finds
Marginal Effects:                          # effect only at the mean
          dF/dx  Std. Err.         z     P>|z|
width   0.08748    0.02447   3.5748  0.00035
c4     -0.26142    0.10569  -2.4735  0.01338
---
dF/dx is for discrete change for the following variables: "c4"
------------------------------------------------------------------------------
```

3. 표준화된 해석

Note

- 예측변수가 여러 개인 경우, 예측 변수의 효과를 비교하기 위하여 $\{\hat{\beta}_j\}$의 크기를 비교할 수 있음
- 이때 서로 다른 단위를 갖는 양적 예측변수 x_j들의 효과를 비교하고자 할 때 표준화 계수(Standardized Coefficient)를 사용할 수 있음
- 예측변수 대신에 $(x_j - \overline{x_j})/s_{x_j}$로 대체하여 적합하게 구함

04

제 6장. 로지스틱회귀모형(2)

예측력 요약: 분류표, ROC 곡선

1. 예측력 요약: 분류표

분류표 이용

- '$\hat{\pi} \geq 0.50$'이면 $\hat{Y}=1$, '$\hat{\pi} < 0.50$'이면 $\hat{Y}=0$ 으로 예측

실제	예측 $\hat{Y}=1$	예측 $\hat{Y}=0$	계
$Y=1$	94	17	111
$Y=0$	34	28	62

- 민감도(Sensitivity) $= P(\hat{Y}=1 \mid Y=1) = \frac{94}{94+17} = 0.85$
- 특이도(Specificity) $= P(\hat{Y}=0 \mid Y=0) = \frac{28}{62} = 0.45$

2. 예측력 요약: ROC 곡선

ROC 곡선 작성(Receiver Operating Characteristic Curve)

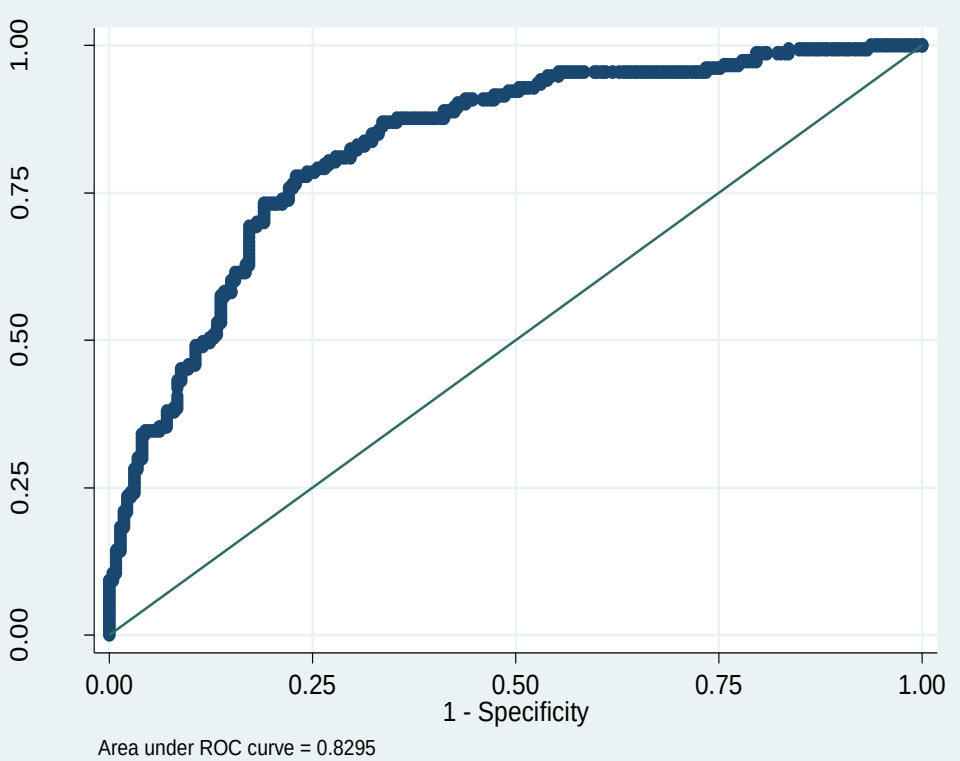
- 분류표

	예측		
실제	$\hat{Y}=1$	$\hat{Y}=0$	계
$Y=1$	True Positives	False Negatives	P
$Y=0$	False Positives	True Negatives	N

- P = True Positives + False Negatives
- TPR = TP / P, FPR = FP/N
- Sensitivity = TPR , Specificity = 1- FPR

2. 예측력 요약: ROC 곡선

ROC 곡선 작성(Receiver Operating Characteristic Curve)



① ROC 곡선
: cutoff 값의 모든 가능한 값에 대해서 민감도를 (1-특이도)의 함수로 나타낸 그림

② Cutoff 값의 모든 가능값에 대해서 예측검정력을 구하기 때문에 분할표를 이용하는 것보다 더 많은 정보를 갖게 됨

③ ROC 곡선
: 일반적으로 좌표 (0,0)과 (1,1)을 연결하는 위로 나타남

④ ROC 곡선 아래 면적
: 예측검정력의 측도로 사용됨(일치성 지수, concordance index)

2. 예측력 요약: ROC 곡선

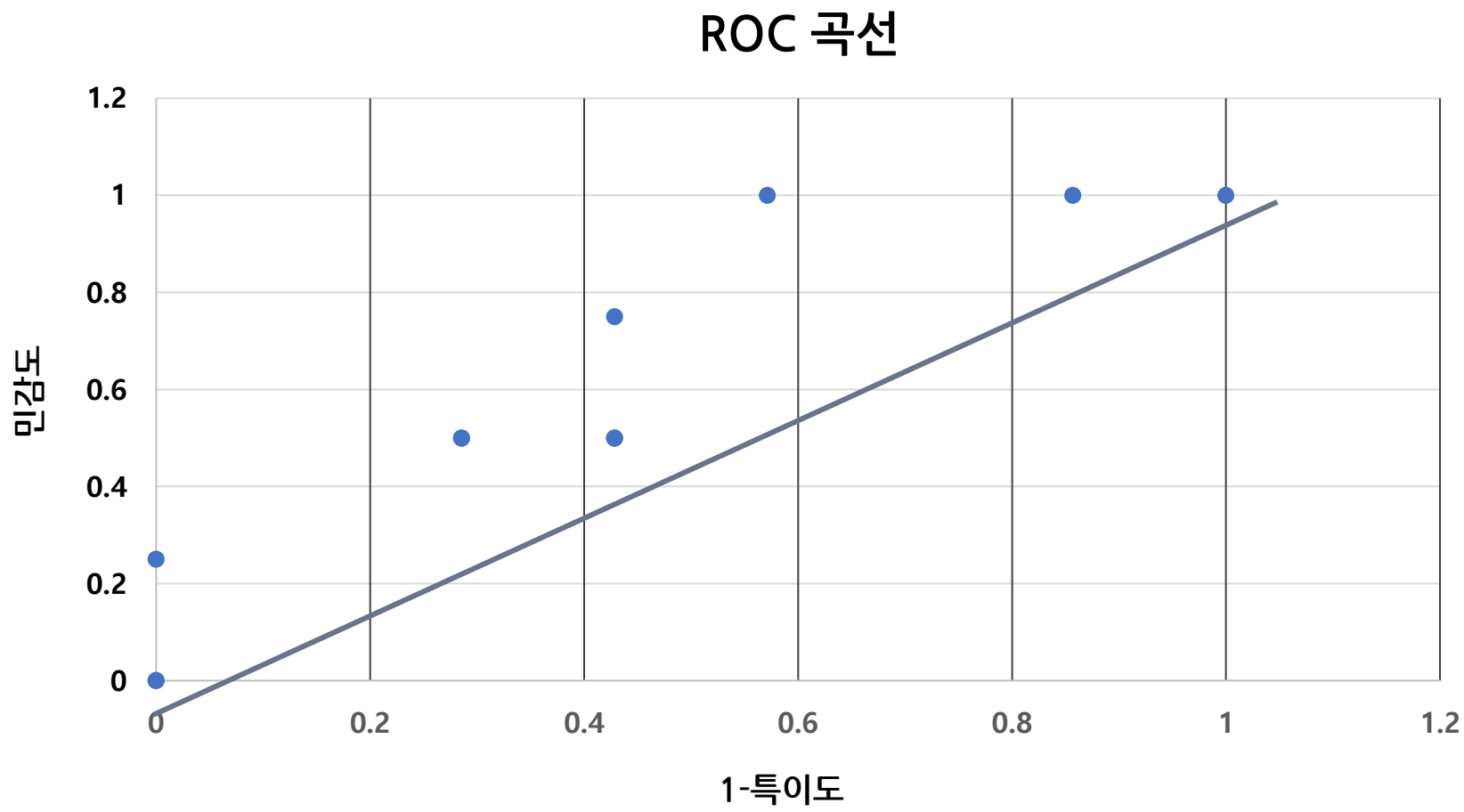
ROC 곡선 작성 (예시)

TRUTH	Score
1	0.7198
0	0.2460
0	0.1219
0	0.1560
0	0.7527
1	0.3064
0	0.7194
0	0.5531
1	0.2173
0	0.0839
1	0.8429

	LOGISTIC REGRESSION	
THRESHOLD	TP-Rate	FP-Rate
0	1	1
0.1	1	0.8571
0.2	1	0.5714
0.3	0.75	0.4286
0.4	0.5	0.4286
0.5	0.5	0.4286
0.6	0.5	0.2857
0.7	0.5	0.2857
0.8	0.25	0
0.9	0	0
1.0	0	0

2. 예측력 요약: ROC 곡선

ROC 곡선 (결과)



3. 예측력 요약: 다중상관성

다중상관성

- 관찰된 반응변수 값 $\{y_j\}$와 모형에서 구한 적합값 $\{\hat{\mu}_j\}$ 간에 구한 상관계수 R
- 선형모형을 적합시키는 경우라면 최소제곱법에서 다중상관계수에 해당함

```
> fit <- glm(y ~ width + factor(color), family=binomial, data=Crabs)
> cor(Crabs$y, fitted(fit))
[1] 0.45221
```

07 강

다음시간안내

로지스틱회귀모형 (3)

수고하셨습니다.