

15강. 총정리

◆ 담당교수 : 김성수 교수

■ 주요용어

용어	해설
중회귀모형	<p>독립변수의 수가 k개인 중회귀모형은 일반적으로 다음과 같이 표현된다.</p> $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \epsilon_i$ <p>$\beta_0, \beta_1, \dots, \beta_k$: 회귀계수 $\epsilon_i \sim N(0, \sigma^2)$ 이고 서로 독립 $i = 1, 2, \dots, n$</p>
로지스틱 회귀모형	<p>로지스틱회귀모형은 로짓함수를 연결함수로 설정하는 모형으로 다음과 같다.</p> $\eta = \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ <p>여기서 $\pi = \mu = E(Y)$ 이다.</p>

■ 연습문제

1. 완전모형(F)과 축소모형(R)을 비교하고자 한다. 두 모형을 비교하는 검정통계량은 ?

①
$$F_0 = \frac{[SSE(R) - SSE(F)] / (df_R - df_F)}{SSE(F) / df_F}$$

②
$$F_0 = \frac{[SSE(R) - SSE(F)] / (df_R - df_F)}{SSE(R) / df_R}$$

③
$$F_0 = \frac{[SSE(F) - SSE(R)] / (df_R - df_F)}{SSE(R) / df_R}$$

④
$$F_0 = \frac{[SSE(F) - SSE(R)] / (df_R - df_F)}{SSE(F) / df_F}$$

정답 및 해설 : ①

2. 다음은 회귀계수 결과이다. 적합한 회귀모형식은 ?

Variable	Estimate	Standard Error	t value	Pr > t
Intercept	-0.651	2.908	-0.22	0.8293
X1	1.551	0.646	(f)	0.0474
X2	0.760	0.397	(g)	0.0970

- ① $\hat{Y} = 0.548X_1 + 0.437X_2$
- ② $\hat{Y} = 0.548X_1 * 0.437X_2$
- ③ $\hat{Y} = -0.651 + 1.551X_1 + 0.760X_2$
- ④ $\hat{Y} = -0.651 + 1.551X_1 * 0.760X_2$

정답 및 해설 : ③

3. 독립변수가 4개 (X_1, X_2, X_3, X_4) 가 있다. 변수 X_1 에 대한 추가변수그림을 그리는 방법은 ?

- ① Y를 (X_1) 으로 회귀한 후 얻어지는 잔차 $e(Y | X_1)$ 와 X_1 을 (X_2, X_3, X_4) 로 회귀한 후 얻어지는 잔차 $e(X_1 | X_2, X_3, X_4)$ 의 산점도를 그린다.
- ② Y를 (X_2, X_3, X_4) 으로 회귀한 후 얻어지는 잔차 $e(Y | X_2, X_3, X_4)$ 와 X_1 을 (X_2, X_3, X_4) 로 회귀한 후 얻어지는 잔차 $e(X_1 | X_2, X_3, X_4)$ 의 산점도를 그린다.
- ③ Y를 (X_1, X_2, X_3, X_4) 으로 회귀한 후 얻어지는 잔차 $e(Y | X_1, X_2, X_3, X_4)$ 와 X_1 을 (X_2, X_3, X_4) 로 회귀한 후 얻어지는 잔차 $e(X_1 | X_2, X_3, X_4)$ 의 산점도를 그린다.
- ④ Y를 (X_1) 으로 회귀한 후 얻어지는 잔차 $e(Y | X_1)$ 와 YFMF (X_2, X_3, X_4) 로 회귀한 후 얻어지는 잔차 $e(Y | X_2, X_3, X_4)$ 의 산점도를 그린다.

정답 및 해설 : ②

4. 다음은 로지스틱 회귀모형 적합 결과이다. 추정된 로지스틱 회귀모형식을 쓰시오.

```
> y <- cbind(glider_g$cases, glider_g$count-glider_g$cases)
> logit_mg <- glm(y~glider_g$p_size_med, family=binomial(link=logit))
```

```

> summary(logit_mg)
Call:
glm(formula = y ~ glider_g$p_size_med, family = binomial(link = logit))
Deviance Residuals:
    1      2      3      4      5 
1.2452 -0.7897 -0.8196  0.9238  0.6694

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.539381   0.839355 -3.025 0.00248 **
glider_g$p_size_med 0.021776 0.007073  3.079 0.00208 **

Null deviance: 16.6058 on 4 degrees of freedom
Residual deviance: 4.1477 on 3 degrees of freedom

```

정답 및 해설 :

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -2.539 + 0.022 * x_1$$

■ 참고사이트

- 강명옥,김영일,안철환,이용구, 『회귀분석』, 을곡출판사, 1996.
- 박성현, 『회귀분석』 (제3판), 민영사, 2007.
- Faraway, J.J. (2002), Practical Regression and Anova Using R,
(www.google.com에서 검색 후, pdf 파일로 다운받을 수 있음)
- McCullagh, P. and Nelder, J.A., Generalized Linear Models, 2nd ed,
Chapman & Hall / CRC., 1999.
- Neter et al. Applied Linear Statistical Models, 4th ed. IRWIN, 1996.
- R 사이트 <http://www.r-project.org/>
- R Studio 사이트 <https://www.rstudio.com/>