14강

SAS, SPSS 활용 (일반화선형모형)

정보통계학과 김성수교수

학습목차

1. SAS를 이용한 일반화선형모형
2. SPSS를 이용한 일반화선형모형

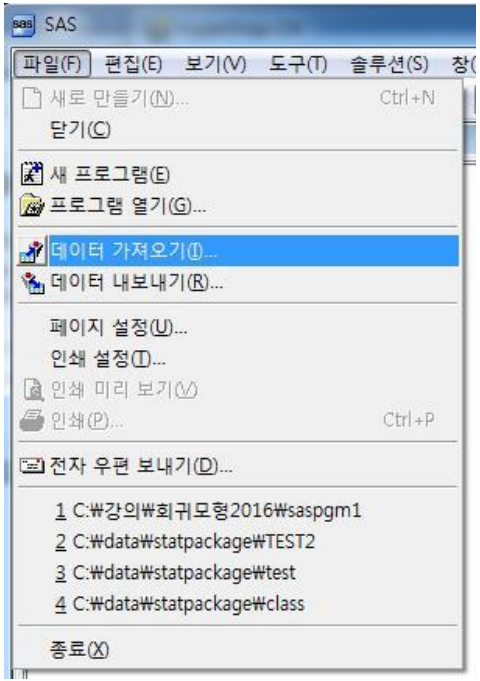
1 SAS를 이용한 일반화선형모형

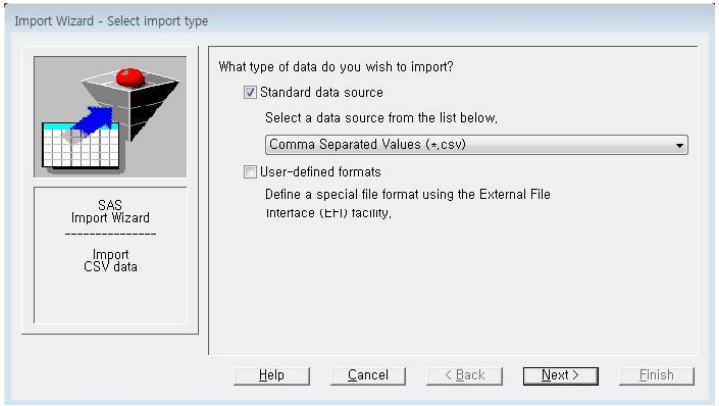
로지스틱 회귀모형 : 이항자료

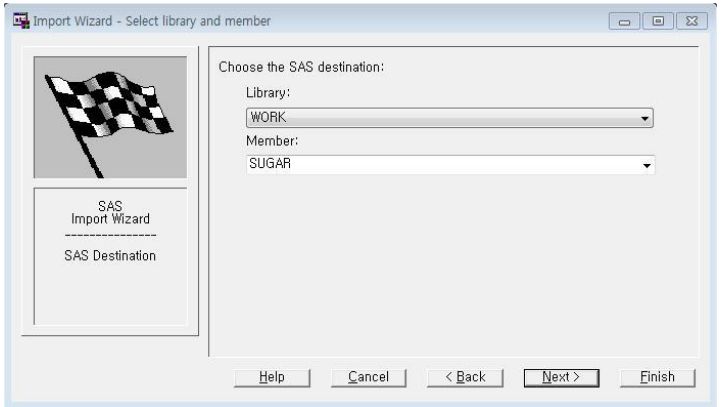
〈날다람쥐 Sugar Glider의 출현자료〉

p_no	occurr	con_metric	p_size_km
1	1	0.650	130.9
2	0	0.610	104.1
3	0	0.744	132.3
4	1	0.213	225.6
5	1	0.723	83.0
6	0	0.678	48.8
7	0	0.733	61.0
8	1	0.522	39.6
9	1	0.552	193.1
10	0	0.245	155.6

데이터읽기







로지스틱 회귀모형 : 이항자료

〈날다람쥐 Sugar Glider의 출현자료〉

p_no	occurr	con_metric	p_size_km
1	1	0.650	130.9
2	0	0.610	104.1
3	0	0.744	132.3
4	1	0.213	225.6
5	1	0.723	83.0
6	0	0.678	48.8
7	0	0.733	61.0
8	1	0.522	39.6
9	1	0.552	193.1
10	0	0.245	155.6

```
data nsugar ;
  set  sugar ;
   if occurr = 0 then  noccurr = 1;
   else if occurr = 1 then noccurr = 0;
   run;
proc logistic data=nsugar;
  model noccurr = con_metric p_size_km ;
  run;
```

반응변수 y=occurr, 1=yes, 0=no 이므로 이항분포를 가정

로지스틱 회귀모형 :

$$\eta = \log it(\pi) = \log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\pi = E(Y \mid x) = \Pr(y = 0 \mid x)$$

※SAS에서는 작은 값 0을 기준으로 함

출력 결과(R & SAS)

```
> logit_m1 <- glm(occurr~p_size_km+con_metric, family=binomial(link=logit), data=glider)
> summary(logit_m1)
...
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4969  -0.8829  -0.3884   0.8766   2.0515

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.606207   1.436391  -2.511  0.01205 *
p_size_km    0.023566   0.007462   3.158  0.00159 **
con_metric   1.631800   1.642758   0.993  0.32055
---
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68.994  on 49  degrees of freedom
Residual deviance: 54.661  on 47  degrees of freedom
AIC: 60.661
Number of Fisher Scoring iterations: 4
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.6062	1.4364	6.3030	0.0121
con_metric	1	1.6318	1.6428	0.9867	0.3206
p_size_km	1	0.0236	0.00746	9.9726	0.0016

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	70.994	60.661
SC	72.906	66.397
-2 Log L	68.994	54.661

R 결과 : 모형의 유의성 검정

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68.994  on 49  degrees of freedom
Residual deviance: 54.661  on 47  degrees of freedom
AIC: 60.661
Number of Fisher Scoring iterations: 4
```

모형의 유의성 검정

$$H_0 : \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 \quad vs. \quad H_1 : \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

즉,

$$H_0 : \beta_1 = 0,\ \beta_2 = 0 \quad vs. \quad H_1 : \text{적어도 하나는 } 0\text{이 아니다.}$$

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14,3332	2	0,0008
Score	12,7906	2	0,0017
Wald	9,9764	2	0,0068

p-값 계산

```
> 1-pchisq(68.994-54.661,2)
[1] 0.0007720201
```

=> p-값이 매우 작으므로 대립가설의 모형이 유의함을 알 수 있음

모형의 선택 : 변수선택방법 (R & SAS)

```
> library(MASS)
> stepAIC(logit_m1, direction='both')
...
Call:  glm(formula = occurr ~ p_size_km, family =
binomial(link = logit),
    data = glider)

Coefficients:
(Intercept)    p_size_km
   -2.52830      0.02173

Degrees of Freedom: 49 Total (i.e. Null);  48 Residual
Null Deviance:      68.99
Residual Deviance:  55.72        AIC: 59.72
```

변수 x_1 **(p_size_km) 이 선택됨.**

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 = -2.528 + 0.022^* x_1$$

```
proc logistic data=nsugar;
   model noccurr = con_metric p_size_km
      /selection=stepwise ;
run;
```

Step 1. Effect p_size_km entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	70.994	59.716
SC	72.906	63.540
-2 Log L	68.994	55.716

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
1.0082	1	0.3153

모형의 선택 : 변수선택방법 (R & SAS)

```
> library(MASS)
> stepAIC(logit_m1, direction='both')
...
Call:  glm(formula = occurr ~ p_size_km, family =
binomial(link = logit),
    data = glider)

Coefficients:
(Intercept)     p_size_km
   -2.52830       0.02173

Degrees of Freedom: 49 Total (i.e. Nul
Null Deviance:      68.99
Residual Deviance:  55.72        AIC:
```

변수 x_1 (p_size_km) 이 선택됨.

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 = -2.528 + 0.022^* x_1$$

```
proc logistic data=nsugar;
   model noccurr = con_metric p_size_km
      /selection=stepwise ;
run;
```

Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	p_size_km		1	1	12,1579		0,0005

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2,5281	0,8202	9,4999	0,0021
p_size_km	1	0,0217	0,00689	9,9352	0,0016

승산비(R & SAS)

```
> glider <- read.csv('c:/data/reg/sugar_glider_binomial.csv')
> logit_m2 <- glm(occurr ~ p_size_km, family=binomial(link=logit), data=glider)
> exp(coef(logit_m2))
(Intercept)   p_size_km
 0.07979473  1.02196464
> exp(confint(logit_m2, parm="p_size_km", level=0.95))
Waiting for profiling to be done...
   2.5 %   97.5 %
1.009424 1.037535
```

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
p_size_km	1.022	1.008	1.036

결과해석 : 구획의 크기가 1km **증가할 때** Sugar Glider**가 출현할 승산은 약** 1.022**배 증가하는 것으로 추정되며,** 95% **신뢰수준에서 승산은** 1.009~1.038**배 사이에서 증가할 것으로 추정됨.**

$\pi(x)$의 추정

$$\log\left(\frac{\hat{\pi}(x)}{1-\hat{\pi}(x)}\right) = -2.528 + 0.022 \times x$$

$$\hat{\pi}(x) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x)}{1+\exp(\hat{\beta}_0 + \hat{\beta}_1 x)} = \frac{\exp(-2.528 + 0.022x)}{1+\exp(-2.528 + 0.022x)}$$

x=150 **에서** $\pi(x)$ **추정값 구하기**

```
> x <- 150
> predict(logit_m2, list(p_size_km=x),
    type="response")
         1
0.6749669
```

```
proc logistic data=nsugar;
        model noccurr = p_size_km  ;
        output out=lresult p=pred  ;
 run;
proc print data=lresult;
run;
```

SAS 시스템

OBS	p_no	occurr	con_metric	p_size_km	noccurr	_LEVEL_	pred
1	1	1	0,65	130,9	0	0	0,57828
2	2	0	0,61	104,1	1	0	0,43376
3	3	0	0,744	132,3	1	0	0,58568
4	4	1	0,213	225,6	0	0	0,91475
5	5	1	0,723	83	0	0	0,32631
6	6	0	0,678	48,8	1	0	0,18726
7	7	0	0,733	61	1	0	0,23097
8	8	1	0,522	39,6	0	0	0,15872
9	9	1	0,552	193,1	0	0	0,84118
10	10	0	0,245	155,6	1	0	0,70106

정리된 자료의 로지스틱 회귀모형 적합

<구획 크기의 계급구간에서 구획 수, Sugar Glider 출현 구획 수, 표본비율>

p_size_km	구간의 중앙값	출현 구획 수	구획 수	표본비율
≦ 50.0	35.3	3	10	0.30
50.0 ~ 100.0	79.55	3	14	0.21
100.0 ~ 150.0	123.6	6	14	0.43
150.0 ~ 200.0	177.65	9	10	0.90
200.0 <	214.55	2	2	1.00

```
proc print data=sugar_g ;
run;
proc logistic data=sugar_g ;
  model  cases/count = p_size_med ;
run ;
```

```
> glider_g <- read.csv('c:/data/reg/sugar_glider_binomial_g.csv')
> head(glider_g)
  p_size_med count cases
1      35.30    10     3
2      79.55    14     3
3     123.60    14     6
4     177.65    10     9
5     214.55     2     2
> y <- cbind(glider_g$cases, glider_g$count-glider_g$cases)
> logit_mg <- glm(y~glider_g$p_size_med,
family=binomial(link=logit))
```

SAS 시스템

OBS	p_size_med	count	cases
1	35.3	10	3
2	79.55	14	3
3	123.6	14	6
4	177.65	10	9
5	214.55	2	2

정리된 자료

```
> summary(logit_mg)

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 **
---

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16.6058  on 4  degrees of freedom
Residual deviance:  4.1477  on 3  degrees of freedom
AIC: 18.547

Number of Fisher Scoring iterations: 4
```

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 = -2.539 + 0.022^* x_1$$

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates	
		Log Likelihood	Full Log Likelihood
AIC	70,994	60,536	18,547
SC	72,906	64,360	22,371
-2 Log L	68,994	56,536	14,547

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12,4581	1	0,0004
Score	11,4693	1	0,0007
Wald	9,4792	1	0,0021

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2,5392	0,8393	9,1524	0,0025
p_size_med	1	0,0218	0,00707	9,4792	0,0021

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
p_size_med	1,022	1,008	1,036

로그선형모형

< 고속도로 속도제한여부와 교통사고 건수 >

year	day	limit	y	year	day	limit	y
1961	1	no	9	1962	1	no	9
1961	2	no	11	1962	2	no	20
1961	3	no	9	1962	3	no	15
1961	4	no	20	1962	4	no	14
1961	5	no	31	1962	5	no	30
1961	6	no	26	1962	6	no	23
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

```
> library(MASS)
> data(Traffic)
> head(Traffic, 3)
  year day limit  y
1 1961   1    no  9
2 1961   2    no 11
3 1961   3    no  9
> write.csv(Traffic, file="c:/data/reg/Traffic.csv")
```

주요관심 내용 : 고속도로의 속도제한이 평균 사고건수에 어떤 영향을 주는가

분석모형 : $\log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{92} x_{92} + \beta_{93} x_{93}$

$$x_1 = \begin{cases} 0 & no \\ 1 & yes \end{cases}, \quad x_i = \begin{cases} 1 & day = i \\ 0 & \text{나머지} \end{cases}, \ i = 2,3,\cdots,92, \quad x_{93} = \begin{cases} 0 & year = 1961 \\ 1 & year = 1962 \end{cases}$$

SAS : 로그선형모형

```
data ntraffic ;
  set traffic;
   if limit = "no" then nlimit=0;
   else if limit="ye" then nlimit=1;

   if year=1961 then nyear=0 ;
   else if year=1962 then nyear=1;

   day2=0 ; day3=0; ... ; day91=0; day92=0;

   if day=2 then day2=1;
   else if day=3 then day3=1;
    ....
   else if day=92 then day92=1;
Run;
proc genmod  data=ntraffic ;
      model  y = nlimit  day2-day92 / dist=poi link=log;
run;
```

$$x_1 = \begin{cases} 0 & no \\ 1 & yes \end{cases}, \; x_i = \begin{cases} 1 & day = i \\ 0 & \text{나머지} \end{cases}, \; i = 2,3,\cdots,92, \; x_{93} = \begin{cases} 0 & year = 1961 \\ 1 & year = 1962 \end{cases}$$

로그선형모형 (R & SAS)

```
> log_m1 <- glm(y~limit+day, family=poisson(link=log), da
> summary(log_m1)
--- log(μ̂) = β̂0 + β̂1x1 + β̂2x2 + ··· + β̂92x92 = 2.20 − 0.30x1 + 0.54a
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.19722    0.23570   9.322  < 2e-16 ***
limityes    -0.29627    0.03978  -7.448 9.46e-14 ***
day2         0.54362    0.29633   1.834 0.066584 .
...
day91        0.38232    0.31515   1.213 0.225077
day92        0.64803    0.29862   2.170 0.030004 *
---
(Dispersion parameter for poisson fa
    Null deviance: 625.25  on 183  d
Residual deviance: 107.64  on  91  d
AIC: 1183.6
Number of Fisher Scoring iterations:
```

107.64/91=1.18 로 모형적합

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	91	107.6440	1.1829
Scaled Deviance	91	107.6440	1.1829
Pearson Chi-Square	91	106.7283	1.1728
Scaled Pearson X2	91	106.7283	1.1728
Log Likelihood		8467.6407	
Full Log Likelihood		-498.7999	
AIC (smaller is better)		1183.5998	
AICC (smaller is better)		1377.8665	
BIC (smaller is better)		1482.5888	

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	2.1972	0.2357	1.7353	2.6592	86.90	<.0001
nlimit	1	-0.2963	0.0398	-0.3742	-0.2183	55.48	<.0001
day2	1	0.5436	0.2963	-0.0372	1.1244	3.37	0.0666
day3	1	0.2877	0.3118	-0.3234	0.8988	0.85	0.3562

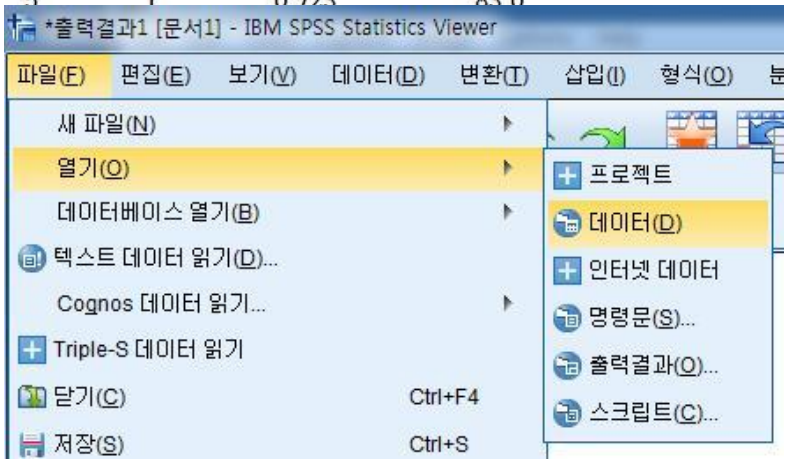
2 SPSS를 이용한 일반화선형모형

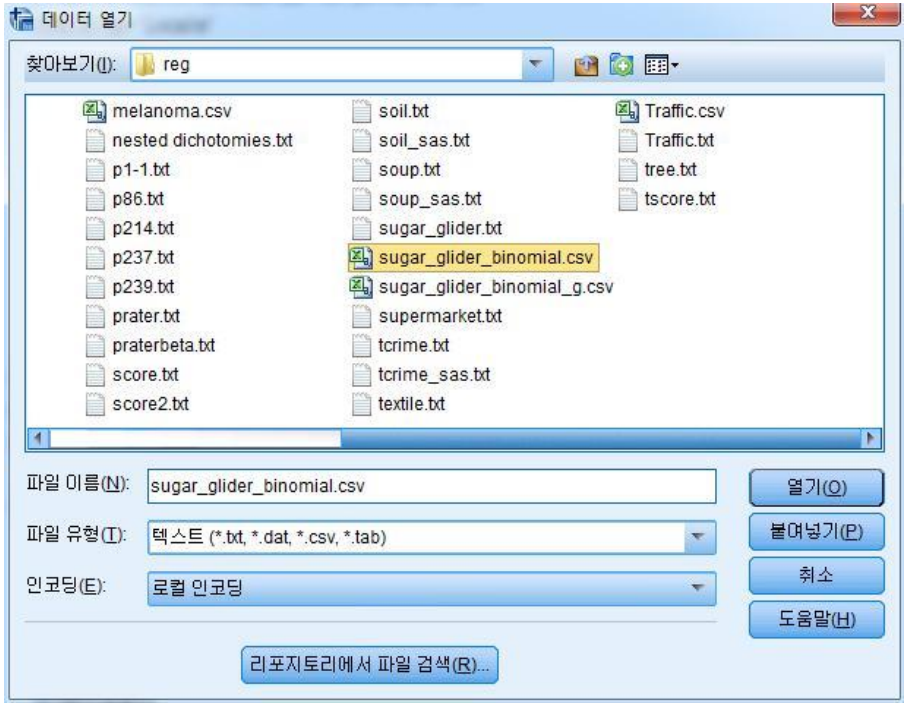
로지스틱 회귀모형 : 이항자료

〈날다람쥐 Sugar Glider의 출현자료〉

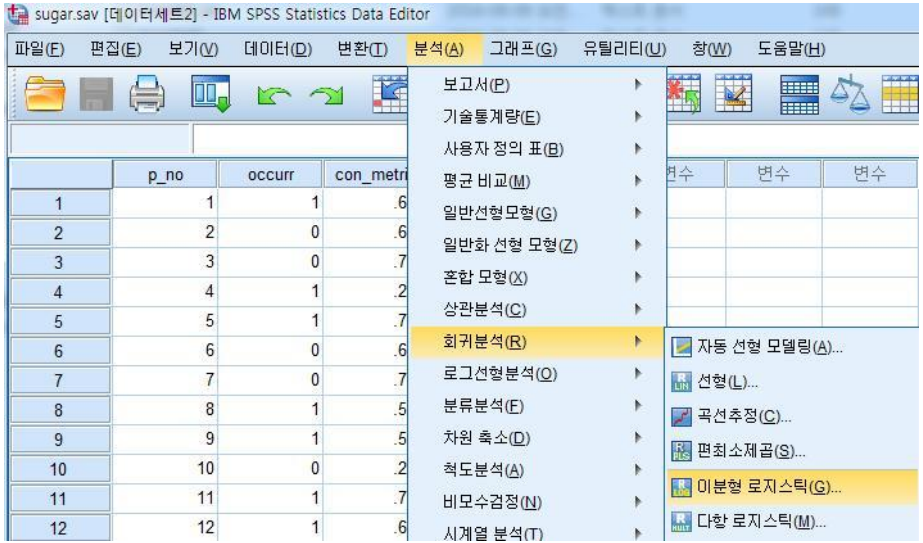
p_no	occurr	con_metric	p_size_km
1	1	0.650	130.9
2	0	0.610	104.1
3	0	0.744	132.3
4	1	0.213	225.6
5	1	0.723	83.0

데이터읽기





로지스틱 회귀모형 : 이항자료



반응변수 y=occurr, 1=yes, 0=no 이므로 이항분포를 가정

로지스틱 회귀모형 :

$$\eta = \log it(\pi) = \log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\pi = E(Y \mid x) = \Pr(y = 1 \mid x)$$

※SPSS에서는 큰값 1을 기준으로 함

출력 결과(R & SPSS)

```
> logit_m1 <- glm(occurr~p_size_km+con_metric, family=binomial(link=logit), data=glider)
> summary(logit_m1)
...
Deviance Residuals:
    Min       1Q   Median       3Q
-1.4969  -0.8829  -0.3884   0.8766

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.606207   1.436391  -2.511  0.01205 *
p_size_km    0.023566   0.007462   3.158  0.00159 **
con_metric   1.631800   1.642758   0.993  0.32055
---
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68.994  on 49  degrees of freedom
Residual deviance: 54.661  on 47  degrees of freedom
AIC: 60.661
Number of Fisher Scoring iterations: 4
```

방정식의 변수

		B	S.E.	Wald	자유도	유의확률	Exp(B)
1 단계[a]	con_metric	1.632	1.643	.987	1	.321	5.113
	p_size_km	.024	.007	9.973	1	.002	1.024
	상수항	-3.606	1.436	6.303	1	.012	.027

a. 변수가 1: con_metric, p_size_km 단계에 입력되었습니다.

모형 요약

단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	54.661[a]	.249	.333

a. 모수 추정값이 .001보다 작게 변경되어 계산반복수 5에

분류표[a]

관측됨		예측		
		occurr		분류정확 %
		0	1	
1 단계	occurr 0	20	7	74.1
	1	7	16	69.6
	전체 퍼센트			72.0

a. 절단값은 .500입니다.

모형의 유의성 검정

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 68.994  on 49  degrees of freedom
Residual deviance: 54.661  on 47  degrees of freedom
AIC: 60.661
Number of Fisher Scoring iterations: 4
```

모형의 유의성 검정

$$H_0 : \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 \quad vs. \quad H_1 : \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

즉,

$$H_0 : \beta_1 = 0,\ \beta_2 = 0 \quad vs. \quad H_1 : \text{적어도 하나는 } 0\text{이 아니다.}$$

p-값 계산

> 1-pchisq(68.994-54.661,2)

[1] 0.0007720201

=> p-값이 매우 작으므로 대립가설의 모형이 유의함을 알 수 있음

모형 계수의 총괄 검정

		카이제곱	자유도	유의확률
1 단계	단계	14.333	2	.001
	블록	14.333	2	.001
	모형	14.333	2	.001

변수선택 : (R & SPSS)

```
> library(MASS)
> stepAIC(logit_m1, direction='both')
...
Call:  glm(formula = occurr
binomial(link = logit),
    data = glider)

Coefficients:
(Intercept)     p_size_km
   -2.52830       0.02173

Degrees of Freedom: 49 Tota
Null Deviance:      68.99
Residual Deviance:  55.72
```

모형 요약

단계	-2 로그 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
1	55.716[a]	.233	.312

a. 모수 추정값이 .001보다 작게 변경되어 계산반복수 4에서 추정을 종료하였습니다.

분류표[a]

관측됨		예측		
		occurr		
		0	1	분류정확 %
1 단계	occurr 0	21	6	77.8
	1	6	17	73.9
	전체 퍼센트			76.0

a. 절단값은 .500입니다.

$$\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 = -2.528 + 0.022^* x_1$$

변수 x_1 (p_size_km) 이 선택됨

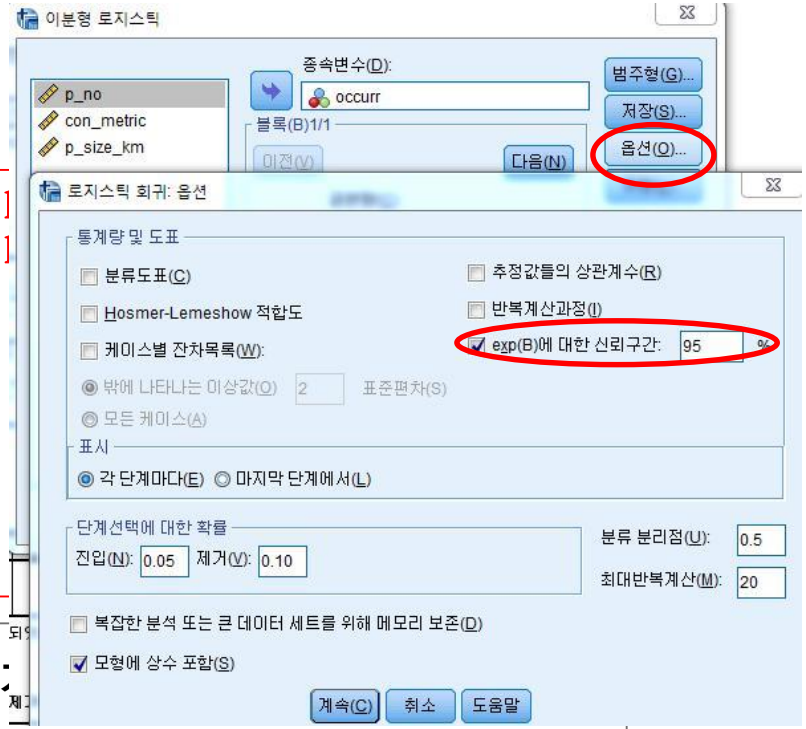
방정식의 변수

		B	S.E.	Wald	자유도	유의확률	Exp(B)
1 단계[a]	p_size_km	.022	.007	9.936	1	.002	1.022
	상수항	-2.528	.820	9.501	1	.002	.080

a. 변수가 1: p_size_km 단계에 입력되었습니다.

승산비(R & SPSS)

```
> glider <- read.csv('c:/data/reg/sugar_glider_binomial
> logit_m2 <- glm(occurr ~ p_size_km, family=binomial(l
> exp(coef(logit_m2))
(Intercept)   p_size_km
 0.07979473  1.02196464
> exp(confint(logit_m2, parm="p_size_km", level=0.95))
Waiting for profiling to be done...
    2.5 %    97.5 %
1.009424 1.037535
```



결과해석 : 구획의 크기가 1km **증가할 때** Sugar Glider**가 증가하는 것으로 추정되며,** 95% **신뢰수준에서 승산은** 1.009~1.038**배 사이에서 증가할 것으로 추정됨.**

방정식의 변수

		B	S.E.	Wald	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
								하한	상한
1 단계[a]	p_size_km	.022	.007	9.936	1	.002	1.022	1.008	1.036
	상수항	-2.528	.820	9.501	1	.002	.080		

a. 변수가 1: p_size_km 단계에 입력되었습니다.

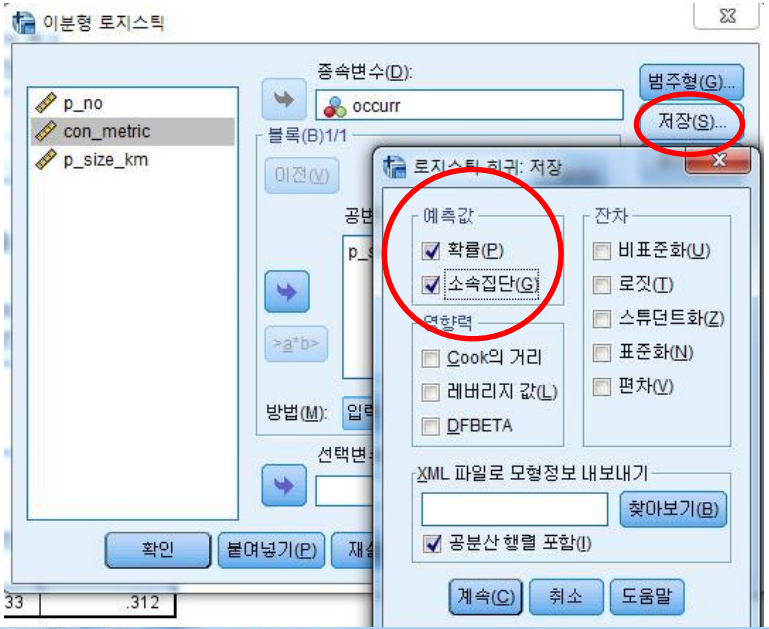
$\pi(x)$ 의 추정

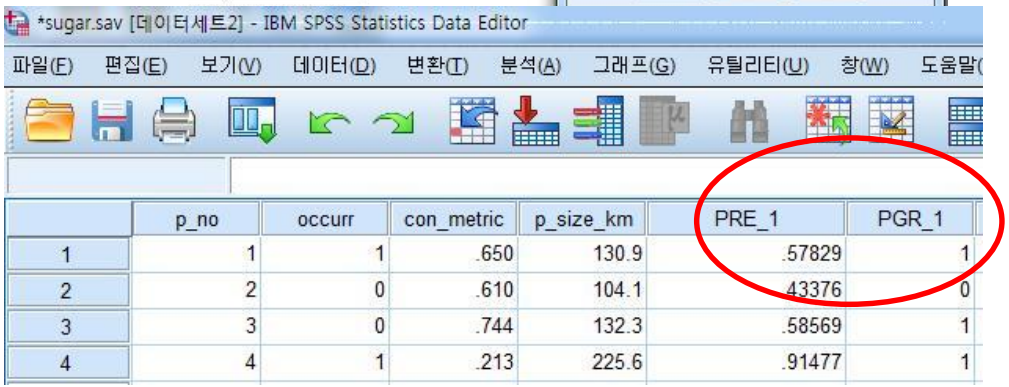
$$\log\left(\frac{\hat{\pi}(x)}{1-\hat{\pi}(x)}\right) = -2.528 + 0.022 \times x$$

$$\hat{\pi}(x) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x)}{1+\exp(\hat{\beta}_0 + \hat{\beta}_1 x)} = \frac{\exp(-2.528 + 0.022x)}{1+\exp(-2.528 + 0.022x)}$$

x=150 **에서** $\pi(x)$ **추정값 구하기**

```
> x <- 150
> predict(logit_m2, list(p_size_km=x),
    type="response")
         1
0.6749669
```





	p_no	occurr	con_metric	p_size_km	PRE_1	PGR_1
1	1	1	.650	130.9	.57829	1
2	2	0	.610	104.1	.43376	0
3	3	0	.744	132.3	.58569	1
4	4	1	.213	225.6	.91477	1

로그선형모형

< 고속도로 속도제한여부와 교통사고 건수 >

year	day	limit	y	year	day	limit	y
1961	1	no	9	1962	1	no	9
1961	2	no	11	1962	2	no	20
1961	3	no	9	1962	3	no	15
1961	4	no	20	1962	4	no	14
1961	5	no	31	1962	5	no	30
1961	6	no	26	1962	6	no	23
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

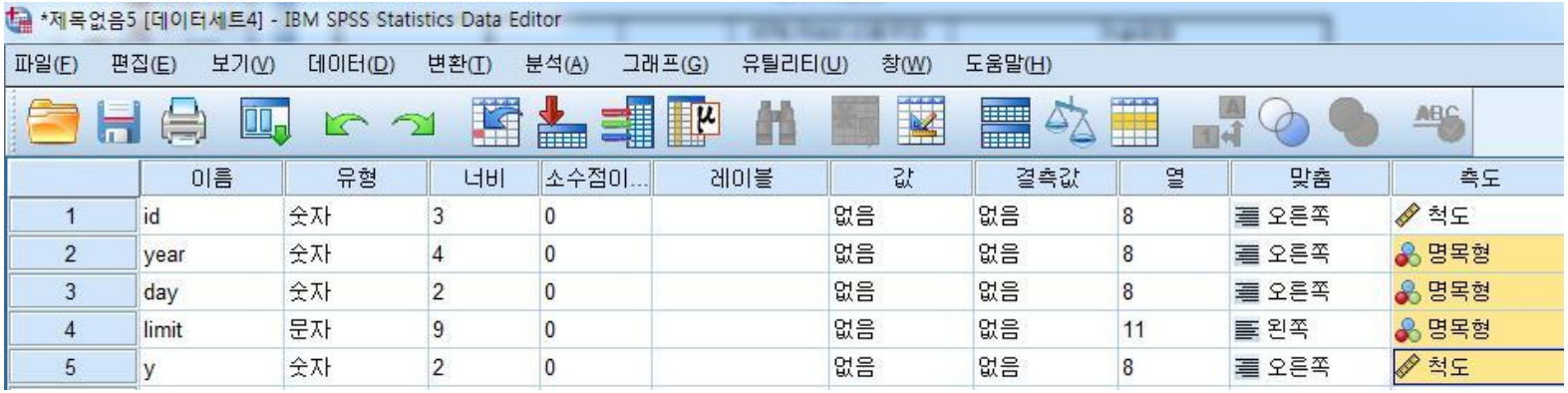
```
> library(MASS)
> data(Traffic)
> head(Traffic, 3)
  year day limit  y
1 1961   1    no  9
2 1961   2    no 11
3 1961   3    no  9
> write.csv(Traffic, file="c:/data/reg/Traffic.csv")
```

주요관심 내용 : 고속도로의 속도제한이 평균 사고건수에 어떤 영향을 주는가

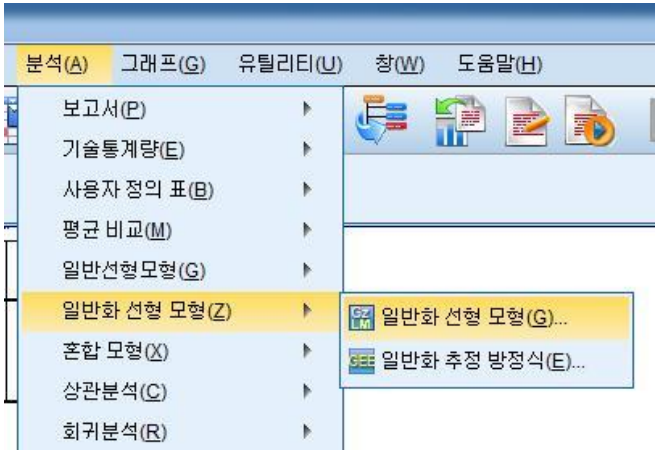
분석모형 : $\log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{92} x_{92} + \beta_{93} x_{93}$

$$x_1 = \begin{cases} 0 & no \\ 1 & yes \end{cases}, \quad x_i = \begin{cases} 1 & day = i \\ 0 & \text{나머지} \end{cases}, \ i = 2,3,\cdots,92, \quad x_{93} = \begin{cases} 0 & year = 1961 \\ 1 & year = 1962 \end{cases}$$

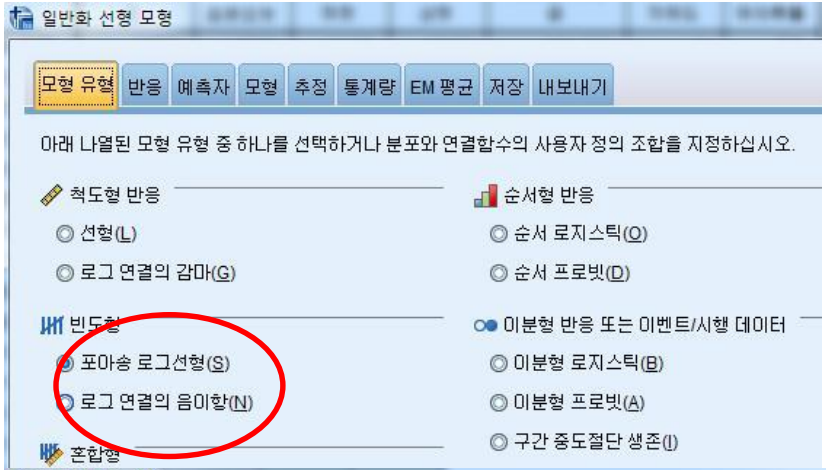
SPSS 절차 : 로그선형모형

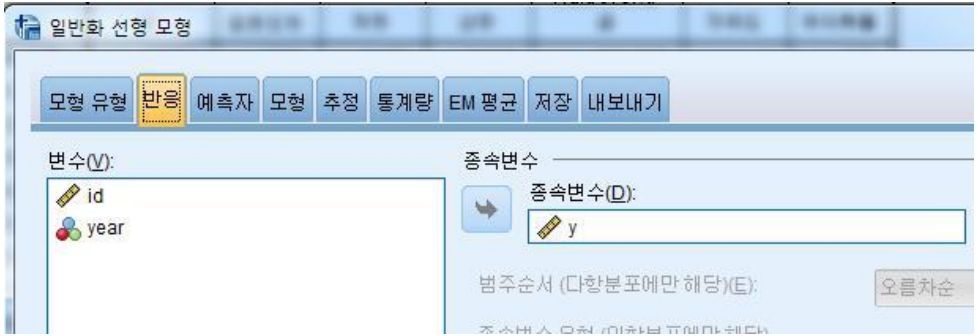


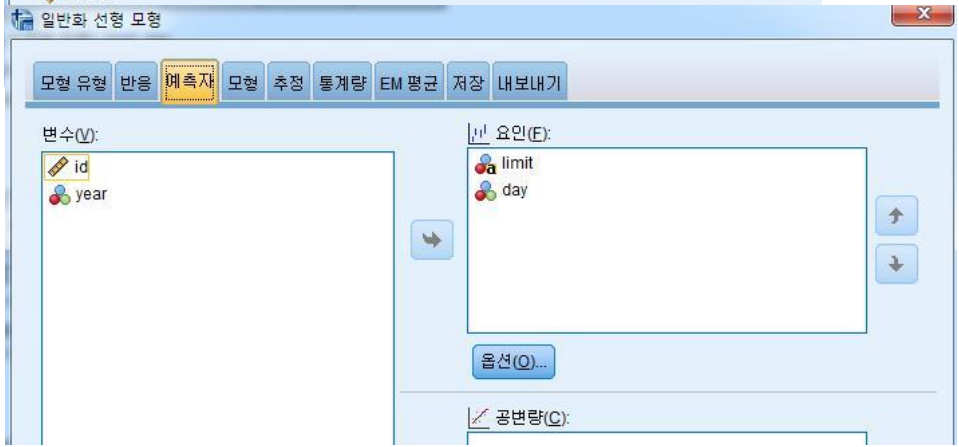
	이름	유형	너비	소수점이...	레이블	값	결측값	열	맞춤	측도
1	id	숫자	3	0		없음	없음	8	오른쪽	척도
2	year	숫자	4	0		없음	없음	8	오른쪽	명목형
3	day	숫자	2	0		없음	없음	8	오른쪽	명목형
4	limit	문자	9	0		없음	없음	11	왼쪽	명목형
5	y	숫자	2	0		없음	없음	8	오른쪽	척도

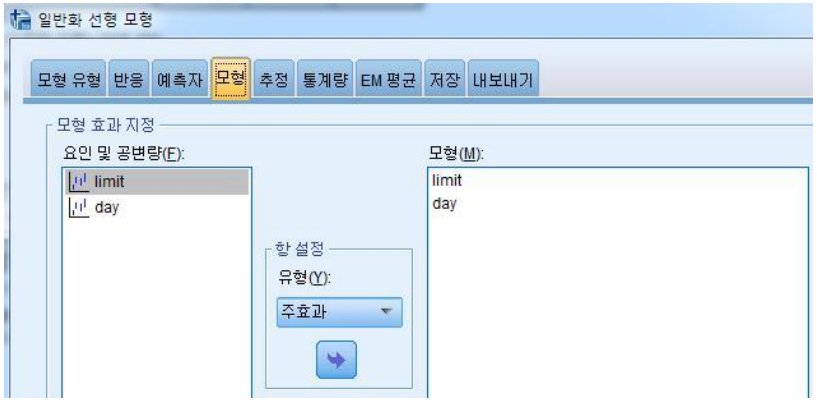


SPSS 절차 : 로그선형모형









로그선형모형 (R & SAS)

```
> log_m1 <- glm(y~limit+day, family=poisson(link=log), data=Traffic)
> summary(log_m1)
```

$$\log(\hat{\mu}) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_{92} x_{92} = 2.20 - 0.30 x_1 + 0.54 x_2 + \cdots + 0.65 x_{92}$$

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.19722    0.23570   9.322  < 2e-16 ***
limityes    -0.29627    0.03978  -7.448 9.46e-14 ***
day2         0.54362    0.29633   1.834 0.066584 .
...
day91        0.38232    0.31515   1.213 0.225077
day92        0.64803    0.29862   2.170 0.030004 *
---
(Dispersion parameter for poisson family
```

107.64/91=1.18 로 모형적합

```
    Null deviance: 625.25  on 183  degree
Residual deviance: 107.64  on  91  degree
AIC: 1183.6
Number of Fisher Scoring iterations: 4
```

적합도[a]

	값	자유도	값/자유도
편차	107.644	91	1.183
척도 편차	107.644	91	
Pearson 카이제곱	106.728	91	1.173
척도 Pearson 카이제곱	106.728	91	
로그 우도[b]	-498.800		
Akaike 정보 기준(AIC)	1183.600		
무한 표본 수정된 AIC (AICC)	1377.866		
베이지안 정보 기준(BIC)	1482.589		
일관된 AIC(CAIC)	1575.589		

총괄 검정[a]

우도비 카이제곱	자유도	유의확률
517.601	92	.000

종속변수: y

모형: (수정된 모형), limit, day

a. 적합한 모형을 절편 전용 모형과 비교

모수 추정값

모수	B	표준오차	95% Wald 신뢰구간 하한	95% Wald 신뢰구간 상한	가설검정 Wald 카이제곱	가설검정 자유도	가설검정 유의확률
(수정된 모형)	2.549	.1840	2.188	2.910	191.923	1	.000
[limit=no]	.296	.0398	.218	.374	55.475	1	.000
[limit=yes]	0[a]	.	.	.	.	.	.
[day=1]	-.648	.2986	-1.233	-.063	4.709	1	.030
[day=2]	-.104	.2567	-.607	.399	.165	1	.684
[day=3]	-.360	.2744	-.898	.177	1.725	1	.189

다음시간 안내

15강. 총정리