

Chapter 5 지시변수

다중 회귀 모형에서 설명 변수의 형태가 분류형 (범주형) 변수인 경우 이 변수들을 질적 변수 (qualitative variable), 지시변수(indicator variable), 혹은 가변수(dummy variable)라 하고 자료 입력 방법이나 결과 해석 방법이 측정형 설명 변수와는 다르다. 모든 설명변수가 범주형 변수인 경우 이에 대한 분석은 분산분석(ANOVA)을 실시하면 된다. 이 경우 설명 변수들을 요인이라 한다.

측정형 설명 변수의 경우 회귀 계수는 다른 설명 변수들이 고정인 경우 설명 변수 한 단위 증가가 종속 변수를 얼마나 증가(혹은 감소) 시키는가를 나타내는 값이나 지시 설명 변수는 그룹에 따른 절편의 차이나 기울기의 차이를 볼 수 있다. 우선 그룹에 따른 절편의 차이만을 나타내는 지시 설명 변수 모형을 살펴 보기로 하자.

회귀모형에서 지시변수는 여러 개 사용할 수 있으나 너무 많아지면 모형 해석이 용이하지 않으므로 1~2개가 적절하다. 만약 여러 개 지시변수가 있는 경우에는 절편의 차이만을 보거나 관심의 대상이 되는 1-2개의 설명변수와 기울기 차이를 보는 것이 적절하다.

5.1 지시 변수가 절편 항에만 영향을 미치는 경우

$$\text{회귀모형: } Y_i = \alpha + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \dots + \gamma_k D_{ki} + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + e_i$$

$$\text{○가정: } e_i \sim \text{Normal}(0, \sigma^2)$$

○ D_k 는 지시변수를 의미하며 일반적으로 (0, 1) 두 값 중 하나를 갖는다.



EXAMPLE

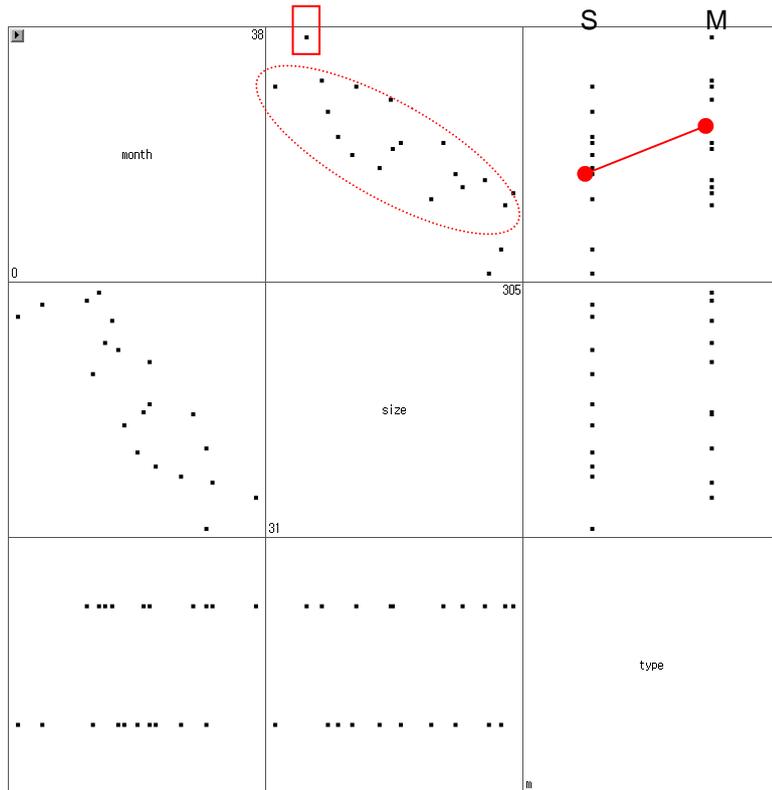
지시 변수(수준 하나)

보험 회사의 외형적인 크기(size, 단위:백만\$), 회사의 형태(지시변수, mutual/stock)가 새로운 기술 혁신을 받아들이는 기간(month: 0의 의미는 최초 시작 기업 의미)에 영향을 주는 지 알아보고자 한다. INSURANCE.txt 회귀모형: $\text{Month} = \beta_0 + \beta_1 \times \text{Type} + \beta_2 \times \text{Size} + e$

우선 산점도 행렬을 그려 보자. 지시변수도 설명변수이므로 함께 넣고 그려야 한다. 사이즈가 클수록 새로운 기술 받아들이는 기간이 길어진다. 그리고 회사 형태에 따라 받아들이는 기간이 달라진다(오른쪽이 Mutual이다). 네모 상자 안의 관측치는 이상치처럼 보인다. (외형적 크기 설명변수 면에서) 이 모두는 likely(일 것 같다)이다. 정확한 결론은 추후 통계량을 이용하여 검정한 후 내려야 한다.

```

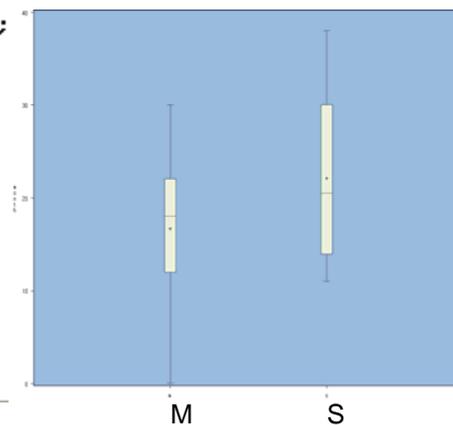
month size type
17 151 m
26 92 m
21 175 m
30 31 m
22 104 m
0 277 m
12 210 m
19 120 m
4 290 m
16 238 m
28 164 s
15 272 s
11 295 s
38 68 s
31 85 s
21 224 s
20 166 s
13 305 s
30 124 s
14 246 s
    
```



설명변수가 지시변수인 경우 다음 그래프(산점도와 같은 개념)를 그리면 더 많은 정보를 얻을 수 있다. 위의 산점도와 유사하지만 상자-나무 그림에서는 이상치 존재 여부를 알 수 있다.

```

proc boxplot data=insurance;
  plot month*type /
  boxwidthscale = 1
  boxstyle=schematic
  nohlabel
  cframe=vlight
  cboxes=dagr
  cboxfill=ywh
  idcolor=salmon
  vaxis=axis1;
run;
    
```

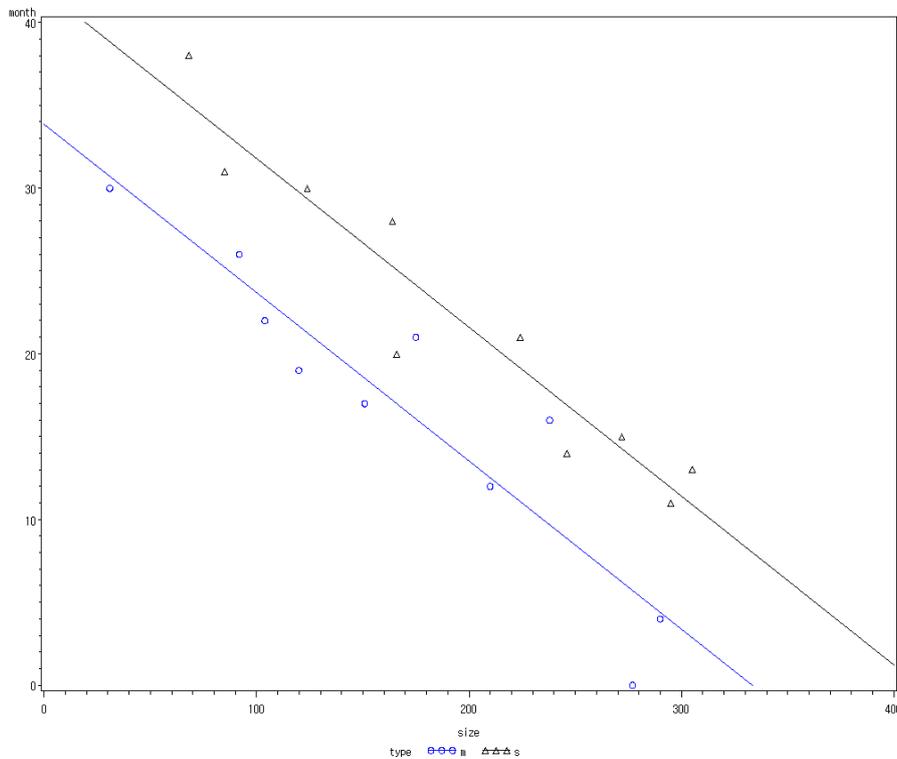


다중 회귀에서 이상치는 어떻게 처리할까? 마지막 단계에서 이상치(영향치) 진단 통계량을 사용하여 판단하면 된다. 그러나 산점도 행렬에서 이상치가 존재함이 확연히 드러나면 미리 제외하고 회귀모형을 추정하는 것이 적절하다.

지시변수가 있는 경우 산점도 행렬과 함께 다음 산점도도 유용하다. 아래 산점도는 지시변수가 절편 항에만 영향을 미치는지 설명변수의 기울기에 영향을 미치는지 미리 진단할

수 있어 반드시 필요한 산점도이다. 측정형 설명변수가 두 개 이상인 경우에는 각 설명변수에 대해 개별적으로 그리면 된다. **SYMBOL** 문장은 점들의 속성을 설정한다. **Symbol**의 변호는 그려지는 그래프의 순서대로 설정하게 된다. **Interpolate** 옵션의 **RL**은 **Regression Line**을 그리라는 옵션이다. **Value** 옵션은 점들의 모양을 설정하며 **Color**는 색을 설정한다. “=TYPE”의 의미는 변수 **type**을 구분하여 점을 표시하라는 의미이고 **Overlay** 옵션은 그래프를 겹쳐 그리라는 옵션이다.

```
proc gplot data=insurance;
  goptions reset=all;
  symbol1 i=r1 v=circle c=blue;
  symbol2 i=r1 v=triangle c=black;
  plot month*size=type/overlay;
run;
```



관측치를 보면 설명변수(지시변수) “Type”에 따라 회귀식의 차이가 있어 보인다. 그러므로 이를 모형에 삽입해야 할 이유가 있다. 우선 회사의 형태에 따라 절편이 다른 모형을 생각해 보자. SAS를 이용하여 지시변수가 있는 회귀모형 추정은 **GML procedure**와 **REG procedure**를 사용할 수 있는데 모형 추정 시는 **GLM**을 사용하는 것이 적절하다. 옵션 **SOLUTION**은 회귀계수를 추정하라는 말이다. 이 옵션이 없으면 회귀계수를 추정하지 않는다.

```
proc glm data=insurance;
  class type;
  model month=size type/solution;
run;
```

SIZE와 TYPE 변수는 모두 유의하다. 회사 크기가 커질수록 받아들이는 기간이 짧아짐을 알 수 있다.(회귀계수 -0.1017, 음의 부호) 그리고 type이 mutual인 경우 받아들이는 기간이 짧아짐을 알 수 있다.

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	41.92953824 B	2.01010079	20.86	<.0001
size	-0.10174212	0.00889122	-11.44	<.0001
type m	-8.05546921 B	1.45910570	-5.52	<.0001
type s	0.00000000 B	.	.	.

물론 아직 잔차분석 및 이상치 진단을 하지 않았지만 모두 문제 없었다면 추정 회귀모형은 다음과 같다.

$$\text{Month} = 41.93 - 0.1017\text{Size} \quad (\text{기업이 stock})$$

$$\text{Month} = 33.87 - 0.1017\text{Size} \quad (\text{기업이 mutual})$$

잔차분석 및 이상치 진단을 위해서는 REG procedure를 사용하는 것이 정답이다. REG에서는 모든 설명변수는 숫자이어야 한다. 그럼 어떤 값으로 넣을 것인가? 만약 Mutual인 경우 $D=1$, Stock이면 $D=0$ 으로 한다고 하자. $\text{Month} = \beta_0 + \beta_1 \times D + \beta_2 \times \text{Size} + e$ β_1 이 유의하고 $\hat{\beta}_1$ 으로 추정되었다면, Mutual 경우 추정 회귀 모형은 $\hat{Y} = (\hat{\beta}_0 + \hat{\beta}_1) + \hat{\beta}_2 \times \text{Size}$ Stock 경우 추정 회귀 모형은 $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_2 \times \text{Size}$ 가 된다.

```
data insurancel;
  set insurance;
  if (type='s') then d=0;
  if (type='m') then d=1;
run;

proc reg data=insurancel;
  model month=size d;
run;
```

D=1(기업이 stock)이면 절편이 β_1 만큼 늘어난다.
그러나 음수이므로 8.06만큼 줄어든다.

$$\text{Month} = \beta_0 + \beta_1 \times D + \beta_2 \times \text{Size} + e$$

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	41.92954	2.01010	20.86	<.0001
size	1	-0.10174	0.00889	-11.44	<.0001
d	1	-8.05547	1.45911	-5.52	<.0001

Mutual인 경우 $D=1$, Stock이면 $D=2$ 으로 한다고 하자. $\text{Month} = \beta_0 + \beta_1 \times D + \beta_2 \times \text{Size} + e$ 모형은 동일하다. 그러나 $D=1$ (기업이 stock)이면 절편이 β_1 만큼 늘어나고 $D=2$ (mutual) 절편이 $2\beta_1$ 만큼 늘어난다. 즉 추정 회귀계수는 위와 다를 것이다. 그러나 최종 회귀모형은 이전과 동일하다.

$$\text{Month} = (49.99 - 8.056) - 0.1017\text{Size} \quad (\text{기업이 stock, } D=1)$$

$$\text{Month} = (49.99 - 2 * 8.056) - 0.1017\text{Size} \quad (\text{기업이 mutual, } D=2)$$

```

data insurancel;
  set insurance;
  if (type='s') then d=1;
  if (type='m') then d=2;
run;

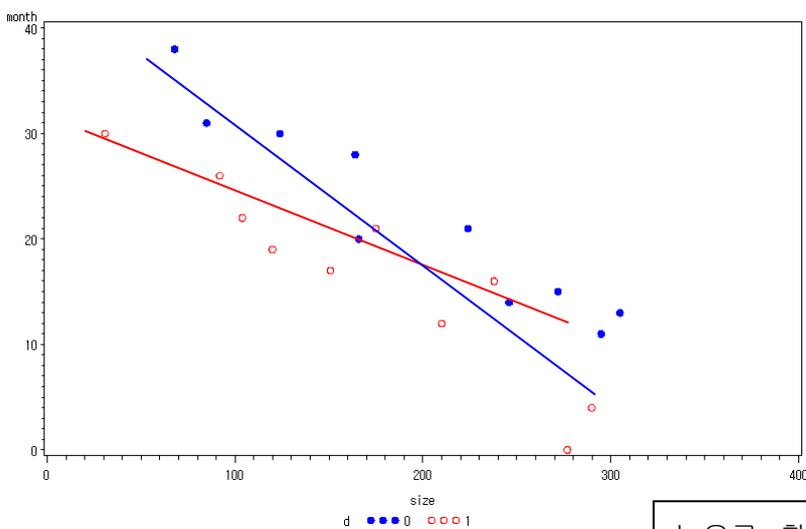
proc reg data=insurancel;
  model month=size d;
run;

```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	49.98501	3.00814	16.62	<.0001
size	1	-0.10174	0.00889	-11.44	<.0001
d	1	-8.05547	1.45911	-5.52	<.0001

일반적으로 수준이 2개 있을 경우에는 $D=0,1$ 으로 코딩하는 것이 일반적이다. 만약 수준이 3개 이상인(예: 회사의 종류가 mutual, stick, bank) 경우에는 지시변수(가변수)가 2개 필요하다. 즉 수준 수에서 하나 줄어든 개수만큼 지시변수가 필요하다. 회사의 종류가 3개인 경우(mutual, stock, bank) 지시변수 $D=1(m),2(s),3(b)$ 하나가 아니라 다음과 같이 지시변수 두 개가 필요하다. $(d_1, d_2) = \begin{cases} (0,0), & m \\ (0,1), & s \\ (1,0), & b \end{cases}$

5.2 지시 변수가 절편 향은 물론 기울기에도 영향을 미치는 경우



눈으로 확신할 수 없지만 기울기가 다르다고 하자.

만약 기업 형태에 따라 Size의 회귀 계수가 다르면 회귀 모형은 다음과 같이 쓸 수 있을 것이다. $Month = \beta_0 + \beta_1 \times D + \beta_2 \times Size + \beta_3 \times Size \times D + e$ 그러므로 만약 β_1, β_3 모두 유의하다면 회귀 모형은 다음과 같다.

$$\text{(Mutual 기업)} \quad Month = (\hat{\beta}_0 + \hat{\beta}_1) + (\hat{\beta}_2 + \hat{\beta}_3) \times Size$$

$$\text{(Stock 기업)} \quad Month = \hat{\beta}_0 + \hat{\beta}_2 \times Size$$

```
data insurance;
  input month size type $;
  if(type='m') then d=1;
  if(type='s') then d=0;
  sized=size*d;
  cards;
  17 151 m
  26 92 s
```

```
proc reg data=insurance;
  model month=d size sized;
run;
```

Parameter Estimates

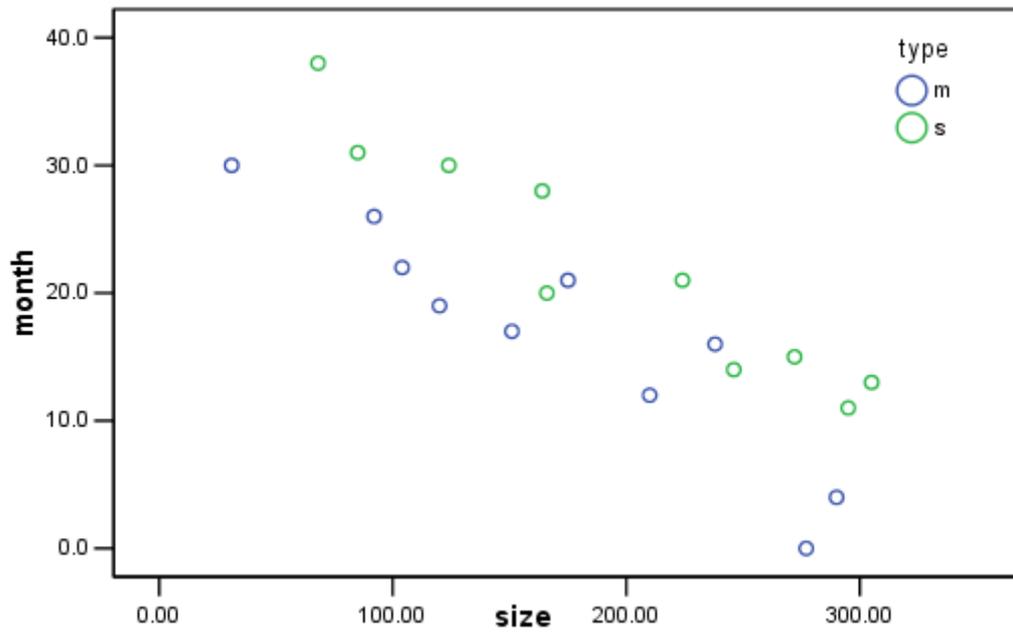
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	41.96962	2.71943	15.43	<.0001
d	1	-8.13125	3.65405	-2.23	0.0408
size	1	-0.10195	0.01287	-7.92	<.0001
sized	1	0.00041714	0.01833	0.02	0.9821

기업 형태별 기울기의 차이는 유의하지 않다. 만약 유의하였다면 Mutual인 경우 Size의 기울기는 -0.1015이고 Stock의 경우는 -0.10195이다.

 **SPSS** 위의 작업을 SPSS 이용하여 하는 방법을 사용해 보자.

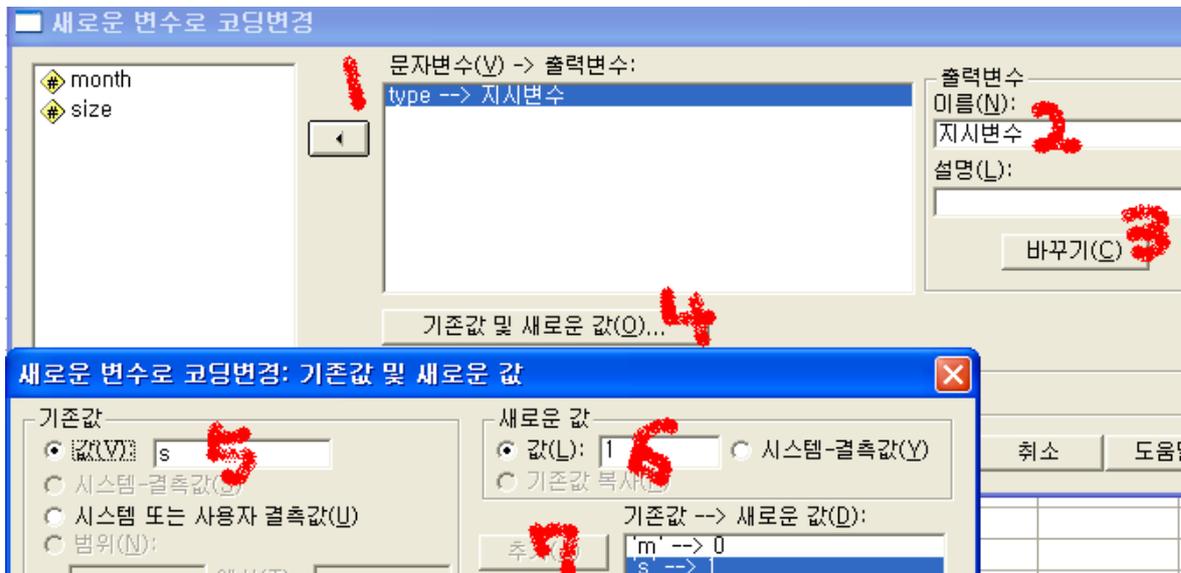
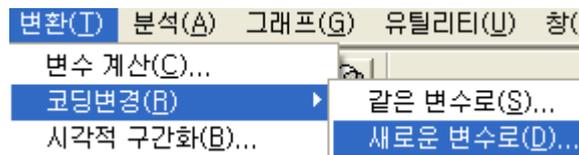
(1) 산점도 그리기 **그래프(G)** ▶ **산점도(S)...**





(2) 회귀모형 추정하기

우선 지시변수를 (0,1)이 되도록 변수 변환해야 한다.



분석(A) ▶ 회귀분석(R) ▶ 선형(L)...



모형		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	33,874	1,814		18,675	.000
	size	-.102	.009	-.911	-11,443	.000
	지시변수	8,055	1,459	.439	5,521	.000

5.3 예제

In_price	hp	type
9.7725	172.9	coupe
10.2819	194.9	sedan
11.0021	174.2	coupe
10.1523	190.4	sedan
10.3169	179.4	sedan
9.988	167.7	coupe
10.1601	176.3	coupe
10.3222	192.6	sedan
10.7547	192.6	coupe
10.9402	191.9	coupe
10.5001	170.2	sedan
10.2972	174.5	sedan
10.5214	185.8	coupe
10.8967	193.3	coupe
11.3314	188.2	sedan
9.6744	189.1	hatchback
9.9002	200	hatchback
10.1674	205.2	hatchback
9.7178	193.6	hatchback
10.1747	198.2	hatchback
10.0237	215.8	hatchback
9.4692	189.1	hatchback
11.0296	178.7	coupe
10.4316	203.3	hatchback
10.4388	202.2	hatchback
10.4388	225.1	hatchback
10.5184	204.4	hatchback
9.5953	176	hatchback
9.3409	183.4	hatchback
9.7978	214.1	hatchback
10.2219	103.1	sedan
8.9158	163.6	sedan
9.5649	184.4	sedan



EXAMPLE

지시 변수(기울기 변환)

마력(Horse Power)이 자동차 가격(가격에 Log 취한 값, 가격은 위로 치우친 형태의 자료)이므로 일반적으로 회귀 분석 시 Log 변환을 해 종속 변수로 사용)에 영향을 미치는 알아보고자 한다. 이 관계에서 자동차의 종류에 따라 달라질 것이라 생각되어 다음 자료를 얻었다.

5.3.1 지시변수 만들기

```
data one;
  input price hp type $;
  cards;
  9.7725 172.9 coupe
  10.2819 194.9 sedan
  11.0021 174.2 coupe
```

```
data one;
  input price hp type $;
  d1=0;d2=0;
  if (type='coupe') then d1=1;
  if (type='sedan') then d2=1;
  cards;
```

지시 변수의 수준이 3개 이상인 경우에는 하나의 지시 변수만으로는 집단을 구별할 수 없다. 즉 $D = 0, 1, 2$ 는 사용할 수 없다. 왜냐하면 0, 1, 2는 숫자가 아니기 때문이다. 지시변수 수준이 3개 이상인 경우 (수준-1)만큼의 지시 변수가 필요하다. 위의 자료의 경우 `type`은 3 수준 (coupe, sedan, hatchback)이므로 회귀 모형에서는 두 개의 지시 변수가 필요하다.

$d1 = 0, d2 = 0$: hatchback

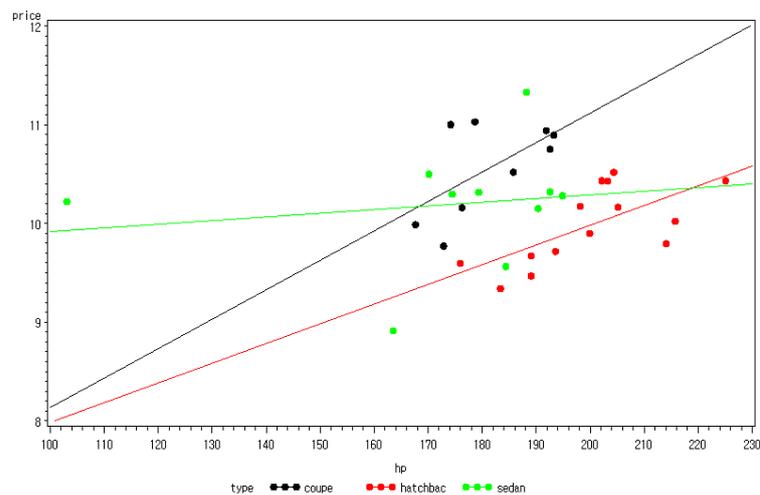
$d1 = 1, d2 = 0$: coupe

$d1 = 0, d2 = 1$: sedan

5.3.2 적절한 산점도를 그리기

```
proc gplot data=one;
  symbol i=r1 v=dot;
  plot price*hp=type;
run;
```

자동차 형태별로 기울기, 절편의 차이가 있어 보인다.



5.3.3 산점도로부터 적절한 회귀 모형 설정

$$\text{Price} = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 HP + \beta_4 D_1 * HP + \beta_5 D_2 * HP + e$$

$d_1 = 0, d_2 = 0$: hatchback

$d_1 = 1, d_2 = 0$: coupe

$d_1 = 0, d_2 = 1$: sedan

5.3.4 모형 및 회귀계수 유의성 검정

```
data one;
  input price hp type $;
  d1=0;d2=0;
  if (type='coupe') then d1=1;
  if (type='sedan') then d2=1;
  hpd1=hp*d1;
  hpd2=hp*d2;
  cards;
9.772.5 172.9 coupe
```

```
proc reg data=one;
  model price=d1 d2 hp hpd1 hpd2;
run;
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	5.98380	1.91944	3.12	0.0043
d1	1	-0.83529	3.60206	-0.23	0.8184
d2	1	3.56719	2.16161	1.65	0.1105
hp	1	0.01997	0.00958	2.08	0.0467
hpd1	1	0.00986	0.01932	0.51	0.6138
hpd2	1	-0.01630	0.01112	-1.47	0.1543

가장 유의하지 않는 D1 변수를 제외하고 회귀 모형을 다시 추정한다.

```
proc reg data=one;
  model price=d2 hp hpd1 hpd2;
run;
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	5.74661	1.59654	3.60	0.0012
d2	1	3.80437	1.87186	2.03	0.0517
hp	1	0.02115	0.00797	2.65	0.0130
hpd1	1	0.00539	0.00131	4.12	0.0003
hpd2	1	-0.01748	0.00972	-1.80	0.0828

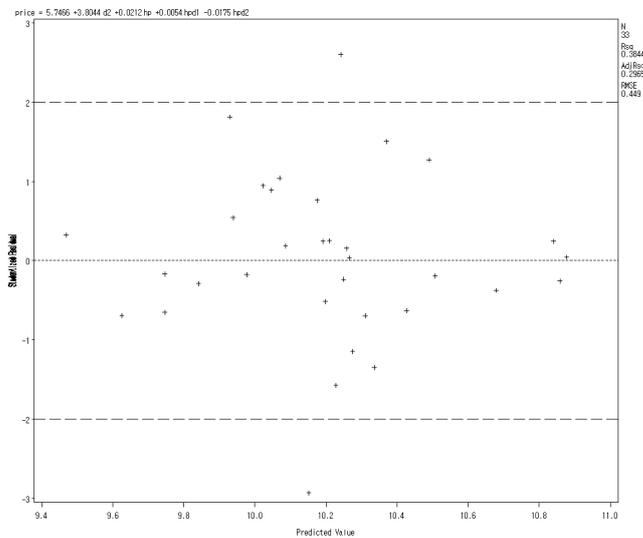
유의수준 0.1 하에서 모든 설명 변수가 유의하다. 이제 추정된 회귀모형을 적어보자.

$$\hat{\text{Price}} = 5.746 + 3.804 * D_2 + 0.0212HP + 0.0054HP * D_1 - 0.0175HP * D_2$$

5.3.5 잔차 분석 실시

```
proc reg data=one;
  goptions reset=all;
  model price=d2 hp hpd1 hpd2/r;
  plot student.*predicted./vref=2 vref=-2;
  output out=out1 r=res;
run;
```

```
proc univariate data=out1 normal plot;
  var res;
run;
```



이상치가 2개 존재한다.(15번째, 32번째) 이를 제외하고 분석해 보자.

3번째, 33번째가 다시 이상치이나 2에 가까우므로 2개만 제외한 모형을 최종 모형으로 한다.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	5.74661	1.16686	4.92	<.0001
d2	1	4.61106	1.37715	3.35	0.0025
hp	1	0.02115	0.00583	3.63	0.0012
hpd1	1	0.00539	0.00095704	5.64	<.0001
hpd2	1	-0.02202	0.00716	-3.08	0.0049 (이상치 2개 제외하고)

별다른 패턴을 보이지 않으므로 회귀모형 가정들을 만족하므로 추정된 회귀 모형은 적절하다. 오차의 정규성을 만족한다.

정규성 검정

검정	통계량	p-값
Shapiro-Wilk	W 0.950542	Pr < W 0.1380
Kolmogorov-Smirnov	D 0.122012	Pr < D 0.1482

5.3.6 최종 회귀모형

```
proc reg data=one;
  goptions reset=all;
  model price=d2 hp hpd1 hpd2/r;
  reweight obs.=15;
  reweight obs.=32;
  plot student.*predicted./vref=2 vref=-2;
run;
```

$\hat{Price} = 5.746 + 4.61 * D_2 + 0.021HP + 0.0054HP * D_1 - 0.022HP * D_2$, $d1 = 0, d2 = 0$: hatchback
 $d1 = 1, d2 = 0$: coupe
 $d1 = 0, d2 = 1$: sedan

$\hat{Price} = 5.746 + 0.021HP$ (Hatchback 자동차)

$\hat{Price} = (5.746 + 4.61) + (0.021 - 0.022)HP$ (Sedan 자동차)

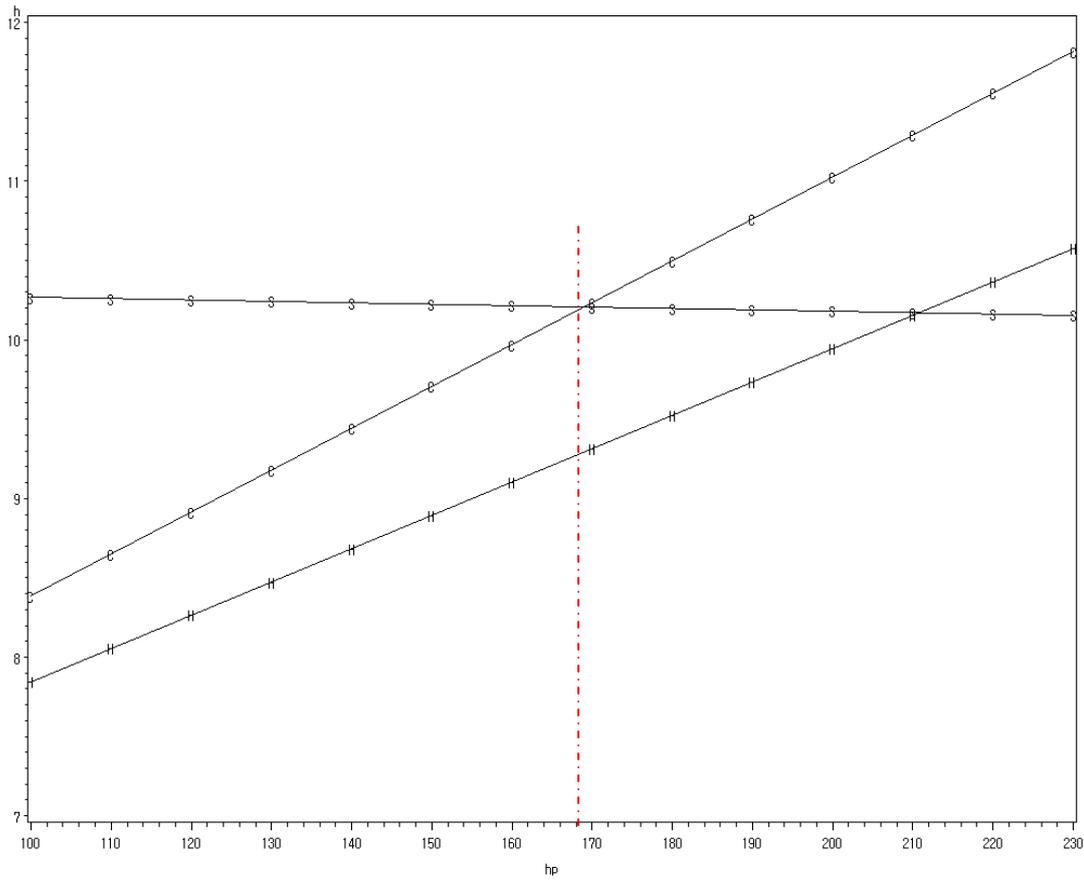
$\hat{Price} = 5.746 + (0.021 + 0.0054)HP$ (Coupe 자동차)

마력은 가격에 양의 영향을 미쳐 마력이 높을수록 자동차 가격은 높아진다. 특히 COUPE의 경우 마력당 자동차 가격 상승률이 다른 차에 비해 높음을 알 수 있다. 이제 자동차 종류에 따른 마력과 자동차 가격을 보기 위하여 그래프를 그려보자. 데이터에서 설명 변수의 마력은 103~225이므로 100~230으로 x-축을 하자.

```
data two;
  do hp=100 to 230 by 10;
    h=5.746+0.021*hp;
    s=10.356-0.00087*hp;
    c=5.746+0.0264*hp;
    output;
  end;
run;

proc gplot data=two;
  goptions reset=all;
  symbol1 i=join v="h" c=black;
  symbol2 i=join v="s" c=black;
  symbol3 i=join v="c" c=black;
  plot (h s c)*hp/overlay;
run;
```

항상 동일 마력인 경우 Hatchback 자동차의 가격이 다른 종류의 차에 비해 항상 낮다. 마력이 170(정확하게 수식으로부터 계산해야 하지만)이후부터 Coupe 차량 가격이 가장 높다. 최종 회귀모형의 산점도는 초기 산점도와 유사함을 알 수 있다. Hatchback 차량의 회귀모형과 Coupe 차량의 추정회귀모형의 절편은 동일하나 마력=0인 곳이 절편이므로 마력=100부터 그린 산점도에서는 만나지 않는다.



5.4 COMMENT

설명변수가 2개 이상인 경우에는 각 설명변수에 대해 기울기가 다른가 동일한가를 알아 보기 위하여 설명변수와 종속변수의 산점도를 지시변수에 의해 그리면 된다. 지시변수가 2개 이상이거나 설명변수가 2개 이상인 경우, 지시변수의 수준 수가 4개 이상인 경우 다소 복잡해지므로 지시변수가 들어간 모형은 다음 경우에 한정하는 것이 좋다.

- ① 지시변수가 하나이고 수준 수가 3개 이하이면 설명변수가 2개 이상이어도 분석이 용이하다. 그리고 각 설명변수의 기울기가 다른 모형도 고려할 수 있다.
- ② 지시변수가 2개 이상인 경우에는 설명변수 하나인 모형만 다루는 것이 좋다.



HOMEWORK #8-1

DUE 5월 2일(월)

다음은 비누 생산 공정으로부터 측정된 자료이다. 두 생산 라인(1, 2)에서 생산 속도(x_1)가 토막 양(y)에 영향을 미치는지 알아보려고 한다. (SAS 이용하기)

- ① 적절한 산점도를 그리고 적절한 회귀모형을 적으시오.
- ② 제시된 모형의 유의성을 검정하고 잔차 분석을 실시하시오.
- ③ 최종 추정회귀 모형을 제시하고 추정 회귀식 그래프를 해석하시오.

Production Line 1				Production Line 2			
Case i	Amount of Scrap Y_i	Line Speed X_{1i}	X_{12}	Case i	Amount of Scrap Y_i	Line Speed X_{1i}	X_{12}
1	218	100	1	16	140	105	0
2	248	125	1	17	277	215	0
3	360	220	1	18	384	270	0
4	351	205	1	19	341	255	0
5	470	300	1	20	215	175	0
6	394	255	1	21	180	135	0
7	332	225	1	22	260	200	0
8	321	175	1	23	361	275	0
9	410	270	1	24	252	155	0
10	260	170	1	25	422	320	0
11	241	155	1	26	273	190	0
12	331	190	1	27	410	295	0
13	275	140	1				
14	425	290	1				
15	367	265	1				



HOMEWORK #8-2

DUE 5월 2일(월)

INDICATOR.xls 데이터에 주 이름, Y(종속변수), X(설명변수), 그리고 지시변수 G(지역; 1, 2, 3, 4)가 있다. (SPSS 이용하기)

- ① 적절한 산점도를 그리고 적절한 회귀모형을 적으시오.
- ② 제시된 모형의 유의성을 검정하고 잔차 분석을 실시하시오.
- ③ 최종 추정회귀 모형을 제시하고 추정 회귀식을 그리고 해석하시오.

주이름	Y	X	Group
ME	235	508	1
NH	231	564	1
VT	270	322	1
MA	261	846	1
RI	288	871	1

HOMEWORK #7-4

```

proc glm data=one;
  class group;
  model y=x group x*group/solution;
run;

```

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	467.7888116 B	79.3877258	5.89	<.0001
X	-0.1971342 B	0.1104986	-1.78	0.0816
Group 1	-270.3371805 B	103.0935939	-2.62	0.0121
Group 2	-312.8000394 B	117.5835483	-2.66	0.0110
Group 3	-334.8040531 B	100.3153925	-3.34	0.0018
Group 4	0.0000000 B	.	.	.
X*Group 1	0.3246260 B	0.1426255	2.28	0.0280
X*Group 2	0.4010062 B	0.1726935	2.32	0.0252
X*Group 3	0.3853936 B	0.1488929	2.59	0.0132
X*Group 4	0.0000000 B	.	.	.

```

data two;
  set one;
  d1=0;d2=0;d3=0;
  if (group=1) then d1=1;
  if (group=2) then d2=1;
  if (group=3) then d3=1;
  d1x=d1*x;
  d2x=d2*x;
  d3x=d3*x;
run;
proc reg data=two;
  model y=x d1 d2 d3 d1x d2x d3x;
run;

```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	467.78881	79.38773	5.89	<.0001
X	1	-0.19713	0.11050	-1.78	0.0816
d1	1	-270.33718	103.09359	-2.62	0.0121
d2	1	-312.80004	117.58355	-2.66	0.0110
d3	1	-334.80405	100.31539	-3.34	0.0018
d1x	1	0.32463	0.14263	2.28	0.0280
d2x	1	0.40101	0.17269	2.32	0.0252
d3x	1	0.38539	0.14889	2.59	0.0132