통계자료처리론 기말프로젝트

방재희

Exhibit 1 Description of 236 Ellery Street Condominium

Rooms:	Five (two bedroom, one bath)
Location:	Mid-Cambridge
Interior Space:	1040 square feet of interior space
Monthly Condominium Fee:	$175
Annual Property Taxes:	$1,121
Price:	$169,000
Put on market:	April 11, 1994
Rent Control Status:	In the event that owner wishes to let condominium, restrictions apply on monthly rent owner may charge.

1. 지역이 M인 자료 중, 1994년~1992년도 자료로 각각 따로 뽑음

```
> data94=read.table("clipboard")
> names(data94) <-
c("AR","FP","LP","SP","CD","DAYS","IT","BED","BATH","ROOM","CONDO","TAX","RC")

> data93=read.table("clipboard")
> names(data93) <-
c("AR","FP","LP","SP","CD","DAYS","IT","BED","BATH","ROOM","CONDO","TAX","RC")

> data92=read.table("clipboard")
> names(data92) <-
c("AR","FP","LP","SP","CD","DAYS","IT","BED","BATH","ROOM","CONDO","TAX","RC")
```

2. 94, 93, 92년별로,
** 1) 지역=M, 2) 940<=Interior94<=1140, 3) BED >=2인 관측값 추출**

```
> filter(data94, IT>=940 & IT<=1140 & BED>=2)
   AR     FP     LP     SP      CD DAYS   IT BED BATH ROOM CONDO  TAX RC
1   M 189000 169000 162000  1/14/94 174 1058   3    1    6   135  966  1
2   M 154975 149000 140000  2/8/94  171  981   3    1    6   125 1788  1
3   M 187000 169000 166000  2/22/94 166  960   3    1    6    50 1545  1
4   M 189000 179000 172000  3/1/94  171 1125   2    1    5   250 1626  1
> filter(data93, IT>=940 & IT<=1140 & BED>=2)
    AR     FP     LP     SP       CD DAYS   IT BED BATH ROOM CONDO  TAX RC
1    M 235000 219000 212500  1/15/93   55 1022   2    2    5   364 1963  0
2    M 139000 129000 118000  2/26/93  241  969   2    2    4   223 1142  0
3    M 179000 179000 170000  6/29/93   74 1100   3    1    6   312 1495  1
4    M 199000 199000 195000  7/16/93   83 1124   2    1    5   100 1596  1
5    M 194000 194000 189000   8/6/93   80 1058   3    1    5   163 1682  0
6    M 189000 153000 148000  8/31/93  185 1060   2    1    5    96 1239  1
7    M 235000 225000 210000   9/1/93  130 1012   2    2    4   435 1811  0
8    M 169500 169500 165000  9/24/93   55 1042   2    2    5   120 1929  0
9    M 189000 189000 177000  9/30/93  133 1120   2    1    5   108 1335  1
10   M 179000 179000 181000 10/29/93   43 1022   2    1    6   100  895  1
> filter(data92, IT>=940 & IT<=1140 & BED>=2)
   AR     FP     LP     SP       CD DAYS   IT BED BATH ROOM CONDO  TAX RC
1   M 132000 132000 125000  1/24/92  140  940   2    1    4   123  412  1
2   M 169900 175000 160000  4/17/92  199 1000   2    2    5   100  830  0
3   M 144900 129000 119500  6/27/92  119  959   2    1    5   130  466  1
```

```
4  M  199000 189000 180000  8/14/92  166  950   2   2   5   100 1483  1
5  M  158000 158000 155000 10/30/92   69  990   2   1   6   160 1119  1
6  M  225000 165000 165000  11/2/92  515 1050   2   2   4   160 2676  0
7  M  190000 165000 160000  11/5/92  237 1060   2   1   5   100 1298  1
```

➔ 94, 93, 92년도 기준으로 보면,

LP가 169,000과 비슷한 관측값 중, SP는 160,000 ~ 165,000 값을 활용할 수 있다.

3. 지역을 M으로 한정하고, 94, 93, 92년별로,
   Last Price를 독립변수로, SalePrice를 종속변수로 하여 회귀분석을 해보자.

> lm(SP~LP, data=data94) -> 회귀선 y=0.9867x-4688 (SP는 LP에서 1.33% 할인)

```
Call:
lm(formula = SP ~ LP, data = data94)

Coefficients:
(Intercept)          LP
 -4688.9092       0.9867
```

> lm(SP~LP, data=data93) -> 회귀선 y=0.9716x-2509 (SP는 LP에서 2.84% 할인)

```
Call:
lm(formula = SP ~ LP, data = data93)

Coefficients:
(Intercept)          LP
 -2509.4949       0.9716
```

> lm(SP~LP, data=data92) -> 회귀선 y=0.9716x-2509 (SP는 LP에서 7.12% 할인)

```
Call:
lm(formula = SP ~ LP, data = data92)

Coefficients:
(Intercept)          LP
  3430.7172       0.9288
```

➔ 94, 93, 92년도 기준으로 보면,

94년 최근일수록 Last Price 대비 SalePrice 차이가 줄어들었고,
SP는 156,967 ~ 166,752 정도로 제안해볼 수 있다.

4. 모든 자료를 data3에 저장하고, 단계적 회귀분석을 해보자.
   독립변수를 모두 추가할 때, AIC 값이 가장 작기 때문에 설명력이 높다고 볼 수 있다.
> start.model=lm(SP~AR+CD, data=data3)
> start.model

```
Call:
lm(formula = SP ~ AR + CD, data = data3)

Coefficients:
(Intercept)         ARAH           ARC         ARCP          ARE
  2.983e+05    -4.994e+04    -4.578e+04    -9.483e+04    5.204e+04
      ARFP         ARHS           ARK          ARM          ARN
 -3.662e+04    -3.662e+04    -7.735e+03    -3.828e+04   -4.372e+04
      ARNW         ARPS          ARRA         ARRS         ARSH
 -1.954e+04    -7.565e+04    -2.649e+04    -5.812e+04   -1.196e+05
       ARW     CD1/12/93     CD1/12/94    CD1/14/93    CD1/14/94
```

-4.017e+04	-1.572e+05	-1.750e+05	4.457e+03	-9.800e+04
CD1/15/93	CD1/16/90	CD1/17/92	CD1/18/94	CD1/19/94
-5.275e+04	-1.213e+05	-7.466e+04	-8.932e+04	-7.963e+04
CD1/2/91	CD1/20/94	CD1/21/94	CD1/24/92	CD1/25/94
1.383e+05	-1.870e+05	-6.179e+04	-1.350e+05	-5.200e+04
CD1/26/94	CD1/27/93	CD1/3/94	CD1/9/90	CD10/1/92
-1.271e+05	-7.645e+04	-1.800e+05	4.467e+04	1.083e+05
CD10/1/93	CD10/17/92	CD10/18/91	CD10/19/93	CD10/2/91
-2.573e+05	-1.897e+05	-1.464e+05	-1.968e+05	-1.127e+05
CD10/20/93	CD10/21/92	CD10/22/93	CD10/25/91	CD10/28/93
-1.066e+05	-1.500e+05	-8.718e+04	-1.297e+05	-1.468e+05
CD10/29/93	CD10/30/92	CD10/31/90	CD10/4/90	CD10/4/91
-8.153e+04	-9.370e+04	-1.518e+05	-5.583e+04	8.421e+03
CD10/5/92	CD10/8/92	CD10/9/92	CD11/1/91	CD11/10/93
3.533e+05	-1.318e+05	-1.847e+05	-8.345e+04	-7.613e+04
CD11/12/92	CD11/12/93	CD11/14/91	CD11/15/90	CD11/15/93
-1.000e+04	-1.024e+05	2.896e+03	-9.266e+04	-8.983e+04
CD11/2/92	CD11/22/91	CD11/23/92	CD11/23/93	CD11/24/93
-9.500e+04	-1.283e+05	-7.379e+04	-9.400e+04	-1.263e+05
CD11/26/91	CD11/30/90	CD11/30/93	CD11/5/92	CD11/5/93
-1.405e+05	-1.687e+05	2.350e+03	-1.000e+05	-1.810e+05
CD11/6/92	CD11/7/90	CD12/10/90	CD12/10/93	CD12/11/90
-1.181e+05	-1.603e+05	-6.316e+04	-1.173e+05	-1.436e+05
CD12/13/90	CD12/13/93	CD12/14/90	CD12/14/93	CD12/15/92
-1.767e+05	4.968e+04	-7.955e+04	-1.612e+05	-1.462e+05
CD12/15/93	CD12/16/91	CD12/18/91	CD12/19/91	CD12/20/93
-8.233e+04	-1.000e+05	1.440e+04	-9.000e+04	-1.601e+05
CD12/21/93	CD12/22/93	CD12/23/91	CD12/23/93	CD12/27/91
-9.432e+04	-7.763e+04	-9.110e+04	-1.088e+05	-1.040e+05
CD12/27/93	CD12/28/92	CD12/28/93	CD12/29/92	CD12/29/93
5.247e+05	8.940e-10	-7.310e+04	-8.510e+04	-1.390e+05
CD12/3/93	CD12/30/93	CD12/31/90	CD12/31/92	CD12/6/93
-1.307e+05	-1.181e+05	-6.850e+04	-5.849e+04	2.247e+05
CD12/7/90	CD2/1/91	CD2/1/94	CD2/12/93	CD2/13/92
-1.590e+05	-1.141e+05	3.834e+04	-5.068e+04	-9.310e+04
CD2/14/91	CD2/14/92	CD2/18/91	CD2/18/94	CD2/22/91
-1.400e+05	-1.446e+05	-1.215e+05	-1.402e+05	-1.582e+05
CD2/22/94	CD2/23/94	CD2/24/93	CD2/26/93	CD2/27/92
-1.191e+05	-8.963e+04	-1.082e+05	-1.526e+05	-1.161e+05
CD2/28/92	CD2/3/92	CD2/3/93	CD2/4/91	CD2/4/92
1.472e+05	-1.617e+05	-1.390e+05	-9.595e+04	-1.227e+05
CD2/5/90	CD2/5/93	CD2/7/94	CD2/8/94	CD2/9/90
-1.410e+05	-1.245e+04	-8.932e+04	-1.200e+05	-1.300e+04
CD3/1/91	CD3/1/94	CD3/11/91	CD3/14/93	CD3/14/94
-1.070e+05	-8.800e+04	-1.356e+05	-2.843e+05	-6.000e+04
CD3/15/93	CD3/15/94	CD3/16/92	CD3/18/92	CD3/19/93
-1.417e+05	-7.054e+04	-1.477e+05	-1.257e+05	-1.550e+05
CD3/2/93	CD3/21/92	CD3/22/91	CD3/22/93	CD3/24/94
-1.005e+05	-1.201e+05	-1.540e+05	1.797e+05	-8.532e+04
CD3/25/94	CD3/26/93	CD3/28/90	CD3/28/94	CD3/3/93
-7.916e+04	-1.231e+05	-1.284e+05	-1.767e+05	-5.983e+04
CD3/30/90	CD3/30/94	CD3/31/92	CD3/31/94	CD3/4/93
-4.300e+04	-7.850e+04	-1.277e+05	-7.831e+04	-1.630e+05
CD3/4/94	CD3/6/92	CD4/1/92	CD4/1/93	CD4/10/91
-1.230e+05	9.092e+04	-2.338e+05	-7.566e+04	-1.267e+05
CD4/11/90	CD4/13/92	CD4/14/94	CD4/15/91	CD4/17/92
-1.037e+05	-2.591e+05	-1.193e+05	3.345e+04	-1.000e+05
CD4/19/90	CD4/2/90	CD4/2/91	CD4/2/92	CD4/20/90
-1.508e+05	-6.955e+04	-1.433e+05	-1.298e+05	-2.060e+04
CD4/22/93	CD4/22/94	CD4/23/93	CD4/25/91	CD4/26/91
-1.340e+05	-1.611e+05	-1.190e+05	-1.077e+05	-1.603e+05
CD4/27/90	CD4/27/94	CD4/28/92	CD4/29/93	CD4/29/94
-3.657e+03	-1.055e+05	-6.183e+04	-1.215e+05	-9.742e+04
CD4/3/94	CD4/30/92	CD4/4/94	CD4/5/94	CD4/6/90
-1.417e+05	-1.531e+05	-1.503e+05	-1.553e+05	-1.103e+05
CD4/8/91	CD5/10/93	CD5/11/90	CD5/11/92	CD5/12/93
-9.833e+04	1.968e+04	-1.340e+05	-1.153e+05	-1.396e+05
CD5/14/91	CD5/14/92	CD5/16/90	CD5/18/90	CD5/18/92

-8.333e+04	-8.800e+04	-1.090e+05	-9.427e+04	-1.500e+05
CD5/19/93	CD5/20/93	CD5/21/91	CD5/21/93	CD5/25/90
-5.979e+04	-5.500e+04	-1.855e+05	-1.455e+05	-1.607e+05
CD5/26/92	CD5/26/93	CD5/27/93	CD5/28/93	CD5/30/90
-1.064e+05	-1.287e+05	-1.201e+05	-7.179e+04	-1.087e+05
CD5/30/91	CD5/31/91	CD5/7/91	CD5/7/93	CD5/8/91
-7.417e+04	-1.034e+05	-4.366e+04	-1.403e+05	-1.380e+05
CD6/10/91	CD6/12/91	CD6/12/92	CD6/13/92	CD6/15/91
-1.828e+05	-1.233e+05	-8.195e+04	-2.054e+04	-1.260e+05
CD6/15/93	CD6/16/92	CD6/17/93	CD6/18/90	CD6/18/91
-6.789e+04	1.190e+04	-1.325e+05	-1.014e+05	-1.268e+05
CD6/19/92	CD6/21/91	CD6/21/93	CD6/24/92	CD6/24/93
-1.188e+05	-1.389e+05	-1.303e+05	-9.204e+04	-1.617e+05
CD6/25/91	CD6/26/91	CD6/27/92	CD6/28/93	CD6/29/90
-1.033e+05	-1.111e+05	-1.405e+05	-6.445e+04	-1.528e+04
CD6/29/93	CD6/3/90	CD6/3/92	CD6/30/92	CD6/30/93
-1.069e+05	-1.557e+05	-1.103e+05	2.828e+04	-4.031e+04
CD6/4/93	CD6/5/90	CD6/5/91	CD6/7/91	CD6/8/92
-6.827e+04	-1.199e+05	1.155e+04	-4.806e+04	-1.502e+05
CD6/9/93	CD7/1/91	CD7/1/92	CD7/10/91	CD7/12/93
-8.045e+04	-7.645e+04	-6.136e+04	-1.292e+05	2.167e+04
CD7/15/91	CD7/15/92	CD7/15/93	CD7/16/91	CD7/16/93
-7.872e+04	-6.996e+04	-8.221e+04	-8.233e+04	-7.152e+04
CD7/17/92	CD7/19/90	CD7/19/91	CD7/2/93	CD7/23/93
9.946e+04	-9.417e+04	-1.003e+05	-1.761e+05	-1.217e+04
CD7/25/90	CD7/25/91	CD7/27/90	CD7/27/93	CD7/28/92
-1.068e+05	-1.098e+05	-1.264e+05	-1.263e+05	-1.280e+05
CD7/28/93	CD7/29/91	CD7/29/92	CD7/29/93	CD7/30/92
-6.560e+04	-1.828e+05	-9.747e+04	-1.007e+05	-1.439e+05
CD7/30/93	CD7/31/90	CD7/31/91	CD7/31/93	CD7/6/91
-6.487e+04	-1.228e+05	-5.033e+04	-1.140e+05	-8.116e+04
CD7/6/93	CD7/7/92	CD7/9/92	CD7/9/93	CD8/1/90
-2.285e+05	9.468e+04	-1.827e+05	-1.019e+05	-1.267e+05
CD8/1/91	CD8/10/93	CD8/11/93	CD8/12/93	CD8/13/90
-1.205e+05	-1.788e+05	-1.166e+04	-1.052e+05	-8.166e+04
CD8/13/93	CD8/14/91	CD8/14/92	CD8/15/90	CD8/15/91
-1.670e+05	-1.587e+05	-8.000e+04	-3.166e+04	-1.567e+05
CD8/16/91	CD8/16/93	CD8/17/90	CD8/19/93	CD8/2/91
-1.327e+05	-8.816e+03	-1.202e+05	-1.550e+05	-9.258e+04
CD8/2/93	CD8/20/92	CD8/20/93	CD8/21/90	CD8/21/92
9.034e+04	-7.683e+04	-1.648e+05	-1.128e+05	-9.639e+03
CD8/22/90	CD8/22/91	CD8/23/91	CD8/24/92	CD8/24/93
-7.800e+04	-8.266e+04	-5.400e+04	-1.450e+05	-1.423e+05
CD8/26/91	CD8/26/92	CD8/27/91	CD8/27/93	CD8/28/90
-1.500e+05	-7.810e+04	-9.816e+04	-5.164e+04	-1.053e+05
CD8/28/91	CD8/29/91	CD8/3/93	CD8/30/90	CD8/31/92
-1.250e+05	-8.219e+04	-1.348e+05	-1.570e+05	-1.278e+05
CD8/31/93	CD8/5/92	CD8/5/93	CD8/6/90	CD8/6/93
-1.120e+05	-1.353e+05	-9.950e+04	-9.379e+04	-9.579e+04
CD8/7/92	CD8/8/90	CD8/8/91	CD8/9/91	CD8/9/93
-5.166e+04	-1.175e+05	-1.140e+05	-9.763e+04	-1.251e+05
CD9/1/92	CD9/1/93	CD9/10/90	CD9/10/92	CD9/12/91
-7.060e+04	-6.073e+04	3.690e+04	-1.317e+05	-1.430e+05
CD9/13/91	CD9/15/92	CD9/15/93	CD9/16/91	CD9/17/92
-1.045e+05	-8.500e+04	-1.482e+05	-4.045e+04	-1.318e+05
CD9/18/91	CD9/18/92	CD9/22/93	CD9/23/91	CD9/23/92
-1.852e+05	-1.537e+05	-4.666e+04	-1.135e+05	-1.531e+05
CD9/24/93	CD9/26/90	CD9/27/90	CD9/27/91	CD9/27/93
-9.500e+04	-1.000e+05	-1.427e+05	-1.733e+05	-6.245e+04
CD9/28/90	CD9/28/93	CD9/29/92	CD9/3/91	CD9/3/93
-1.483e+05	-1.447e+05	-5.845e+04	-1.285e+05	-1.267e+05
CD9/30/91	CD9/30/92	CD9/30/93	CD9/4/90	CD9/4/91
-2.500e+04	1.033e+05	-1.536e+05	-1.263e+05	-1.496e+05
CD9/4/92	CD9/5/90	CD9/7/90	CD9/8/90	CD9/9/93
-2.500e+04	-5.500e+04	1.334e+04	-1.050e+05	-3.963e+04

```
> stepAIC(start.model,
scope=list(upper=~AR+CD+DAYS+IT+BED+BATH+ROOM+CONDO+TAX+RC, lower=~AR+CD),
direction="both")
Start:  AIC=-764.18
log(SP) ~ AR + CD

         Df Sum of Sq     RSS      AIC
+ TAX     1   13.0623   5.3236 -1327.36
+ IT      1   11.5159   6.8700 -1211.07
+ ROOM    1    7.5497  10.8362 -1003.26
+ BED     1    7.0762  11.3097  -983.76
+ CONDO   1    6.3009  12.0850  -953.52
+ BATH    1    5.9861  12.3998  -941.79
+ RC      1    1.5440  16.8419  -802.17
<none>                 18.3859  -764.18
+ DAYS    1    0.0007  18.3851  -762.19

Step:  AIC=-1327.36
log(SP) ~ AR + CD + TAX

         Df Sum of Sq     RSS      AIC
+ IT      1    1.5836   3.7400 -1486.35
+ ROOM    1    0.7330   4.5905 -1392.91
+ BED     1    0.5434   4.7802 -1374.45
+ DAYS    1    0.2014   5.1222 -1342.94
+ BATH    1    0.1802   5.1434 -1341.06
<none>                  5.3236 -1327.36
+ CONDO   1    0.0117   5.3119 -1326.36
+ RC      1    0.0003   5.3233 -1325.38
- TAX     1   13.0623  18.3859  -764.18

Step:  AIC=-1486.35
log(SP) ~ AR + CD + TAX + IT

         Df Sum of Sq     RSS      AIC
+ DAYS    1   0.14627  3.5937  -1502.5
+ CONDO   1   0.10658  3.6334  -1497.5
+ RC      1   0.02335  3.7166  -1487.2
+ ROOM    1   0.02293  3.7171  -1487.2
+ BED     1   0.02250  3.7175  -1487.1
<none>                 3.7400  -1486.3
+ BATH    1   0.00374  3.7363  -1484.8
- IT      1   1.58359  5.3236  -1327.4
- TAX     1   3.12998  6.8700  -1211.1

Step:  AIC=-1502.55
log(SP) ~ AR + CD + TAX + IT + DAYS

         Df Sum of Sq     RSS      AIC
+ CONDO   1    0.1287  3.4650  -1517.2
+ RC      1    0.0238  3.5699  -1503.6
+ ROOM    1    0.0208  3.5729  -1503.2
+ BED     1    0.0200  3.5738  -1503.1
<none>                 3.5937  -1502.5
+ BATH    1    0.0016  3.5921  -1500.8
- DAYS    1    0.1463  3.7400  -1486.3
- IT      1    1.5285  5.1222  -1342.9
- TAX     1    3.2486  6.8423  -1210.9

Step:  AIC=-1517.18
log(SP) ~ AR + CD + TAX + IT + DAYS + CONDO

         Df Sum of Sq     RSS      AIC
+ ROOM    1   0.05547  3.4096  -1522.5
+ BED     1   0.04677  3.4183  -1521.4
+ RC      1   0.02673  3.4383  -1518.7
<none>                 3.4650  -1517.2
+ BATH    1   0.00022  3.4648  -1515.2
```

```
- CONDO  1   0.12870 3.5937 -1502.5
- DAYS   1   0.16838 3.6334 -1497.5
- TAX    1   1.44913 4.9142 -1359.8
- IT     1   1.63506 5.1001 -1342.9
```

<mark>Step: AIC=-1522.53</mark>
<mark>log(SP) ~ AR + CD + TAX + IT + DAYS + CONDO + ROOM</mark>

```
         Df Sum of Sq    RSS     AIC
+ RC      1   0.03605 3.3735 -1525.4
<none>                 3.4096 -1522.5
+ BED     1   0.00729 3.4023 -1521.5
+ BATH    1   0.00256 3.4070 -1520.9
- ROOM    1   0.05547 3.4650 -1517.2
- CONDO   1   0.16338 3.5729 -1503.2
- DAYS    1   0.16841 3.5780 -1502.5
- IT      1   0.84963 4.2592 -1423.1
- TAX     1   1.23463 4.6442 -1383.6
```

<mark>Step: AIC=-1525.38</mark>
<mark>log(SP) ~ AR + CD + TAX + IT + DAYS + CONDO + ROOM + RC</mark>

```
         Df Sum of Sq    RSS     AIC
<none>                 3.3735 -1525.4
+ BED     1   0.00254 3.3710 -1523.7
+ BATH    1   0.00054 3.3730 -1523.5
- RC      1   0.03605 3.4096 -1522.5
- ROOM    1   0.06479 3.4383 -1518.7
- DAYS    1   0.16933 3.5428 -1505.0
- CONDO   1   0.17078 3.5443 -1504.9
- IT      1   0.85973 4.2333 -1423.9
- TAX     1   1.01902 4.3925 -1407.0
```

Call:
lm(formula = log(SP) ~ AR + CD + TAX + IT + DAYS + CONDO + ROOM +
    RC, data = data3)

Coefficients:

(Intercept)	ARAH	ARC	ARCP	ARE	ARFP
ARHS					
11.0629886	-0.0642240	0.0007143	-0.2775281	0.0777248	-0.0741587
-0.0425579					
ARK	ARM	ARN	ARNW	ARPS	ARRA
ARRS					
-0.1316608	-0.0117789	-0.1189313	0.0006830	-0.2994903	0.0958937
-0.2109864					
ARSH	ARW	CD1/12/93	CD1/12/94	CD1/14/93	CD1/14/94
CD1/15/93					
-0.4039203	-0.1587977	-0.1454392	-0.3386144	0.1159089	0.0808019
-0.0137846					
CD1/16/90	CD1/17/92	CD1/18/94	CD1/19/94	CD1/2/91	CD1/20/94
CD1/21/94					
0.1923748	0.1857666	0.0289810	0.3824956	0.2473794	-0.3156132
0.0597507					
CD1/24/92	CD1/25/94	CD1/26/94	CD1/27/93	CD1/3/94	CD1/9/90
CD10/1/92					
0.0592693	0.0869708	0.2120023	0.0922170	-0.1258494	0.1286782
0.1175307					
CD10/1/93	CD10/17/92	CD10/18/91	CD10/19/93	CD10/2/91	CD10/20/93
CD10/21/92					
-0.2568241	-0.1587494	-0.0275291	-0.4619100	0.0919043	-0.0052499
-0.1632349					
CD10/22/93	CD10/25/91	CD10/28/93	CD10/29/93	CD10/30/92	CD10/31/90
CD10/4/90					
0.2527119	-0.2031604	-0.0444788	0.1935496	-0.0971664	0.0597688
-0.0205909					
CD10/4/91	CD10/5/92	CD10/8/92	CD10/9/92	CD11/1/91	CD11/10/93
CD11/12/92					

-0.0672475   -0.1090407   0.1663747   -0.1790718   0.1981062   0.0215837
-0.2932051
CD11/12/93 CD11/14/91 CD11/15/90 CD11/15/93 CD11/2/92 CD11/22/91
CD11/23/92
0.1180871   0.2938435   -0.0340281   0.0828691   -0.1493548   -0.0139832
-0.1342436
CD11/23/93 CD11/24/93 CD11/26/91 CD11/30/90 CD11/30/93 CD11/5/92
CD11/5/93
-0.3205964   -0.1438234   -0.0719525   -0.2659579   0.0851906   0.0728270
-0.1546120
CD11/6/92 CD11/7/90 CD12/10/90 CD12/10/93 CD12/11/90 CD12/13/90
CD12/13/93
0.2992476   0.0411823   0.2816975   0.1974323   0.0077476   -0.4711339
0.2712564
CD12/14/90 CD12/14/93 CD12/15/92 CD12/15/93 CD12/16/91 CD12/18/91
CD12/19/91
0.0749186   -0.0675747   -0.0938485   0.3319167   0.1040058   0.0691690
-0.0796132
CD12/20/93 CD12/21/93 CD12/22/93 CD12/23/91 CD12/23/93 CD12/27/91
CD12/27/93
-0.1764838   -0.0699780   0.0870008   0.2427305   0.0345552   -0.0324197
-0.3586114
CD12/28/92 CD12/28/93 CD12/29/92 CD12/29/93 CD12/3/93 CD12/30/93
CD12/31/90
-0.0652787   0.2906804   0.0361277   -0.0248275   0.1289870   -0.0671462
0.1282224
CD12/31/92 CD12/6/93 CD12/7/90 CD2/1/91 CD2/1/94 CD2/12/93
CD2/13/92
0.1782452   0.1063955   -0.1596445   -0.2337385   0.0647146   -0.1639718
-0.0578581
CD2/14/91 CD2/14/92 CD2/18/91 CD2/18/94 CD2/22/91 CD2/22/94
CD2/23/94
-0.2451286   -0.1564609   -0.0115123   -0.3353019   0.0318777   0.0054766
0.1777225
CD2/24/93 CD2/26/93 CD2/27/92 CD2/28/92 CD2/3/92 CD2/3/93
CD2/4/91
0.1022204   -0.2425150   -0.1101920   -0.1729939   -0.0678937   0.0175616
0.1047555
CD2/4/92 CD2/5/90 CD2/5/93 CD2/7/94 CD2/8/94 CD2/9/90
CD3/1/91
-0.1867023   -0.1235678   0.0931971   0.0907676   -0.1996079   0.2620520
-0.0388496
CD3/1/94 CD3/11/91 CD3/14/93 CD3/14/94 CD3/15/93 CD3/15/94
CD3/16/92
-0.0604883   0.2523837   -0.5939257   0.1435524   -0.0270389   0.4210424
-0.0797472
CD3/18/92 CD3/19/93 CD3/2/93 CD3/21/92 CD3/22/91 CD3/22/93
CD3/24/94
0.0398060   -0.2029868   0.3788453   0.0231917   -0.2007531   0.0706221
0.0265793
CD3/25/94 CD3/26/93 CD3/28/90 CD3/28/94 CD3/3/93 CD3/30/90
CD3/30/94
0.0617067   -0.1041453   0.1796762   -0.2393272   0.1217543   0.0761009
-0.0402272
CD3/31/92 CD3/31/94 CD3/4/93 CD3/4/94 CD3/6/92 CD4/1/92
CD4/1/93
-0.2185368   0.0745291   -0.5562275   0.1336844   -0.0285678   -0.2867541
-0.2038489
CD4/10/91 CD4/11/90 CD4/13/92 CD4/14/94 CD4/15/91 CD4/17/92
CD4/19/90
0.0160575   0.0640307   -0.3465888   -0.1088808   -0.0383332   0.1384740
-0.1378733
CD4/2/90 CD4/2/91 CD4/2/92 CD4/20/90 CD4/22/93 CD4/22/94
CD4/23/93
0.1417536   0.1407769   0.0568239   0.4266645   -0.2046356   -0.0840118
-0.1377020
CD4/25/91 CD4/26/91 CD4/27/90 CD4/27/94 CD4/28/92 CD4/29/93
CD4/29/94

-0.0374713   -0.1495046   0.1504571   0.1305457   0.0382356   -0.0702184
0.0484309
CD4/3/94   CD4/30/92   CD4/4/94   CD4/5/94   CD4/6/90   CD4/8/91
CD5/10/93
0.1079930   0.1936481   -0.0007081   -0.0772729   0.3517276   0.1902733
0.1340209
CD5/11/90   CD5/11/92   CD5/12/93   CD5/14/91   CD5/14/92   CD5/16/90
CD5/18/90
-0.0681249   0.0801742   -0.0070770   0.1031932   0.0825481   -0.0946256
0.0984793
CD5/18/92   CD5/19/93   CD5/20/93   CD5/21/91   CD5/21/93   CD5/25/90
CD5/26/92
-0.1787130   0.0069798   -0.2461335   -0.6529771   -0.3417692   0.1610541
0.1158520
CD5/26/93   CD5/27/93   CD5/28/93   CD5/30/90   CD5/30/91   CD5/31/91
CD5/7/91
0.1499421   -0.0204399   0.0372455   0.1626891   -0.0172329   0.0428045
-0.0256454
CD5/7/93   CD5/8/91   CD6/10/91   CD6/12/91   CD6/12/92   CD6/13/92
CD6/15/91
-0.0277404   -0.1571568   -0.2652539   0.2115875   0.1808341   -0.0842075
-0.0033283
CD6/15/93   CD6/16/92   CD6/17/93   CD6/18/90   CD6/18/91   CD6/19/92
CD6/21/91
0.0934269   0.0725397   0.1219654   0.0641154   -0.0205011   -0.1065640
-0.3616989
CD6/21/93   CD6/24/92   CD6/24/93   CD6/25/91   CD6/26/91   CD6/27/92
CD6/28/93
0.2424492   0.1287218   -0.0855296   0.1815726   0.0525640   -0.0518062
0.2401693
CD6/29/90   CD6/29/93   CD6/3/90   CD6/3/92   CD6/30/92   CD6/30/93
CD6/4/93
-0.2260980   -0.0962485   0.0298431   0.0627737   0.1005228   0.1268497
0.0695962
CD6/5/90   CD6/5/91   CD6/7/91   CD6/8/92   CD6/9/93   CD7/1/91
CD7/1/92
0.2421722   0.0845901   0.1129449   0.0076622   0.2150471   -0.0231535
0.0586940
CD7/10/91   CD7/12/93   CD7/15/91   CD7/15/92   CD7/15/93   CD7/16/91
CD7/16/93
-0.0934447   0.4260190   -0.0266888   -0.0153438   0.0921158   0.0332569
0.2269640
CD7/17/92   CD7/19/90   CD7/19/91   CD7/2/93   CD7/23/93   CD7/25/90
CD7/25/91
-0.0075605   0.1128031   0.4637626   -0.1203885   0.1036543   0.0302552
-0.0272410
CD7/27/90   CD7/27/93   CD7/28/92   CD7/28/93   CD7/29/91   CD7/29/92
CD7/29/93
0.1812800   0.1151833   -0.0231661   0.0579782   -0.3792353   0.0439743
0.1184848
CD7/30/92   CD7/30/93   CD7/31/90   CD7/31/91   CD7/31/93   CD7/6/91
CD7/6/93
0.0192951   0.1779224   -0.0031881   0.2756650   -0.0397278   0.3555311
-0.5226190
CD7/7/92   CD7/9/92   CD7/9/93   CD8/1/90   CD8/1/91   CD8/10/93
CD8/11/93
0.0912361   -0.1406755   0.1259932   0.3032504   0.1414545   -0.0967458
-0.1740616
CD8/12/93   CD8/13/90   CD8/13/93   CD8/14/91   CD8/14/92   CD8/15/90
CD8/15/91
0.1171823   0.2204841   -0.2190701   -0.1965569   0.1790479   0.1410708
0.1573116
CD8/16/91   CD8/16/93   CD8/17/90   CD8/19/93   CD8/2/91   CD8/2/93
CD8/20/92
0.2534068   0.1643264   0.1315124   0.0223573   0.1837789   -0.3023604
0.1503172
CD8/20/93   CD8/21/90   CD8/21/92   CD8/22/90   CD8/22/91   CD8/23/91
CD8/24/92

-0.1241371	0.2603062	0.0414549	-0.0233140	0.1897048	0.0327786

0.0988781

CD8/24/93	CD8/26/91	CD8/26/92	CD8/27/91	CD8/27/93	CD8/28/90

CD8/28/91

-0.0524080	0.1166478	-0.0097536	0.3265905	0.0106681	0.0902358

-0.1696356

CD8/29/91	CD8/3/93	CD8/30/90	CD8/31/92	CD8/31/93	CD8/5/92

CD8/5/93

0.0445913	-0.0822468	-0.0014616	-0.1610504	-0.0053615	0.0548829

-0.8053367

CD8/6/90	CD8/6/93	CD8/7/92	CD8/8/90	CD8/8/91	CD8/9/91

CD8/9/93

0.0654794	-0.2384035	0.2073452	0.1926070	0.0221886	0.0559349

-0.1771644

CD9/1/92	CD9/1/93	CD9/10/90	CD9/10/92	CD9/12/91	CD9/13/91

CD9/15/92

0.1498057	-0.0168385	0.3903059	0.0340257	-0.1175231	0.0573609

-0.2773978

CD9/15/93	CD9/16/91	CD9/17/92	CD9/18/91	CD9/18/92	CD9/22/93

CD9/23/91

-0.2985088	0.0012858	0.0561002	-0.1017210	-0.0292466	-0.0386322

-0.3129537

CD9/23/92	CD9/24/93	CD9/26/90	CD9/27/90	CD9/27/91	CD9/27/93

CD9/28/90

-0.3390298	-0.1339471	0.0618113	0.1105176	-0.2621948	0.0540571

0.0450866

CD9/28/93	CD9/29/92	CD9/3/91	CD9/3/93	CD9/30/91	CD9/30/92

CD9/30/93

0.0668502	0.2586879	-0.0913407	0.2218630	-0.2049047	0.3582393

-0.2811916

CD9/4/90	CD9/4/91	CD9/4/92	CD9/5/90	CD9/7/90	CD9/8/90

CD9/9/93

-0.1319136	-0.1215142	0.2059735	0.2484556	-0.1918739	0.1139717

-0.2324524

TAX	IT	DAYS	CONDO	ROOM	RC
0.0002105	0.0004355	-0.0002890	0.0005946	0.0362022	-0.0480394

> lm(log(SP)~LP+ROOM+RC, data=data94)

Call:
lm(formula = log(SP) ~ LP + ROOM + RC, data = data94)

Coefficients:
(Intercept)	LP	ROOM	RC
1.068e+01	6.319e-06	3.666e-02	NA

> p=10.68+0.000006*169000+0.04*5
> p
[1] 11.894
> exp(p)
[1] 146385.7

➔ -> Area=M, Year=1994년도로 정리한 data94 데이터 세트를 이용하여

위 단계적 회귀분석에서 coefficients 값이 상대적으로 큰 ROOM과 RC, 그리고 Last Price 변수로 계산하여, 예상가격을 146,385로 찾았다.

결론,
✓ **94, 93, 92**년별로,

지역=M, 2) 940<=Interior94<=1140, 3) BED >=2 인 관측값 추출

➔ 예상가격: 160,000 ~ 165,000

✓ 지역을 M 으로 한정하고, 94, 93, 92 년별로,
Last Price 를 독립변수로, SalePrice 를 종속변수로 하여 회귀분석

➔ 예상가격: 156,967 ~ 166,752

✓ **lm(log(SP)~LP+ROOM+RC, data=data94)**

➔ 예상가격: 146,385

안지훈

> Reyeme2 <- read_excel("C:/Users/k16hy/Desktop/통계자료처리론/새 폴더/Reyeme2.xls")

> View(Reyeme2)

Multiple Regression

> reg=lm(SalePrice~Area+Days)

> summary(reg)

```
Call:
lm(formula = SalePrice ~ Area + Days)

Residuals:
    Min      1Q  Median      3Q     Max
-239228  -43357  -11321   24477  569652

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 170342.552  24724.230   6.890 1.95e-11 ***
AreaAH       -9856.448  28221.051  -0.349   0.7271
AreaC       -42347.331  47694.229  -0.888   0.3751
AreaCP      -37794.315  30068.498  -1.257   0.2094
AreaE       135184.153  28658.897   4.717 3.22e-06 ***
AreaFP      -25416.368  62614.196  -0.406   0.6850
AreaHS        1766.706  26226.247   0.067   0.9463
AreaK        41901.839  30559.763   1.371   0.1710
AreaM       -11896.850  25763.990  -0.462   0.6445
AreaN       -19228.674  32043.781  -0.600   0.5488
AreaNW       -9165.521  43947.163  -0.209   0.8349
AreaPS      -45336.019  32303.655  -1.403   0.1612
AreaRA       -5603.165  27783.423  -0.202   0.8403
AreaRS      -28495.583  32288.256  -0.883   0.3780
AreaSH      -86068.430  41344.479  -2.082   0.0379 *
AreaW         5290.708  28225.982   0.187   0.8514
Days            -1.841     28.027  -0.066   0.9477
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81340 on 439 degrees of freedom
Multiple R-squared: 0.2065,   Adjusted R-squared: 0.1776
F-statistic: 7.141 on 16 and 439 DF, p-value: 8.919e-15

Call:
lm(formula = SalePrice ~ Area + Days)

Residuals:
    Min   1Q Median   3Q   Max
-239228 -43357 -11321   24477 569652

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 170342.552 24724.230   6.890 1.95e-11 ***
AreaAH   -9856.448 28221.051 -0.349   0.7271
AreaC   -42347.331 47694.229 -0.888   0.3751
AreaCP   -37794.315 30068.498 -1.257   0.2094
AreaE   135184.153 28658.897   4.717 3.22e-06 ***
AreaFP   -25416.368 62614.196 -0.406   0.6850
AreaHS   1766.706 26226.247   0.067   0.9463
AreaK   41901.839 30559.763   1.371   0.1710
AreaM   -11896.850 25763.990 -0.462   0.6445
AreaN   -19228.674 32043.781 -0.600   0.5488
AreaNW   -9165.521 43947.163 -0.209   0.8349
AreaPS   -45336.019 32303.655 -1.403   0.1612
AreaRA   -5603.165 27783.423 -0.202   0.8403
AreaRS   -28495.583 32288.256 -0.883   0.3780
AreaSH   -86068.430 41344.479 -2.082   0.0379 *
AreaW   5290.708 28225.982   0.187   0.8514
Days   -1.841   28.027 -0.066   0.9477

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81340 on 439 degrees of freedom
Multiple R-squared: 0.2065,   Adjusted R-squared: 0.1776
F-statistic: 7.141 on 16 and 439 DF, p-value: 8.919e-15

> reg=lm(SalePrice~Area+Interior)

> summary(reg)

Call:
lm(formula = SalePrice ~ Area + Interior)

Residuals:
    Min   1Q Median   3Q   Max
-284450 -21516   -1849   17523 438905

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -17899.01   19483.90 -0.919 0.35878
AreaAH   21994.67   19972.44   1.101 0.27139
AreaC   -90959.61   33600.23 -2.707 0.00705 **
AreaCP   -31643.93   21202.97 -1.492 0.13630
AreaE   115902.20   20169.44   5.746 1.71e-08 ***

```
AreaFP      -59937.97   44163.31  -1.357  0.17542
AreaHS       38262.64   18431.97   2.076  0.03849 *
AreaK        39131.59   21551.72   1.816  0.07010 .
AreaM         -854.56   18155.67  -0.047  0.96248
AreaN       -49421.36   22532.05  -2.193  0.02880 *
AreaNW      -13183.48   30966.42  -0.426  0.67051
AreaPS      -23833.55   22813.34  -1.045  0.29673
AreaRA       31051.18   19624.46   1.582  0.11431
AreaRS      -31343.41   22790.67  -1.375  0.16975
AreaSH      -27975.54   29269.20  -0.956  0.33970
AreaW         9731.05   19846.31   0.490  0.62415
Interior       174.72       8.31  21.027  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 57410 on 439 degrees of freedom
Multiple R-squared:  0.6047,    Adjusted R-squared:  0.5903
F-statistic: 41.97 on 16 and 439 DF,  p-value: < 2.2e-16
```

> reg=lm(SalePrice~Area+Interior+Condo+Tax+RC)
> summary(reg)

```
Call:
lm(formula = SalePrice ~ Area + Interior + Condo + Tax + RC)

Residuals:
    Min      1Q  Median      3Q     Max
-116070  -21276     516   19646  219179

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -25481.078  13863.962  -1.838 0.066752 .
AreaAH       18210.269  12792.005   1.424 0.155287
AreaC       -37070.022  21598.245  -1.716 0.086809 .
AreaCP      -10754.777  13779.153  -0.781 0.435514
AreaE        48031.380  13810.281   3.478 0.000556 ***
AreaFP       -6484.645  28285.097  -0.229 0.818774
AreaHS        3479.072  11935.578   0.291 0.770817
AreaK       -11777.514  14539.314  -0.810 0.418356
AreaM          105.031  11694.775   0.009 0.992838
AreaN       -12930.541  14602.561  -0.885 0.376376
AreaNW         878.484  20024.644   0.044 0.965028
AreaPS      -10307.589  15127.882  -0.681 0.496003
AreaRA       10905.235  12596.245   0.866 0.387102
AreaRS       -1975.399  14697.997  -0.134 0.893149
AreaSH       -4204.201  18815.207  -0.223 0.823292
AreaW       -18714.637  12768.412  -1.466 0.143451
Interior        76.767      6.974  11.007  < 2e-16 ***
Condo          177.496     26.294   6.750 4.71e-11 ***
Tax             51.924      3.574  14.528  < 2e-16 ***
RC           11376.706   4540.865   2.505 0.012595 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36660 on 436 degrees of freedom
Multiple R-squared:  0.8399,    Adjusted R-squared:  0.833
F-statistic: 120.4 on 19 and 436 DF,  p-value: < 2.2e-16
```

*회귀모형 변수 생성

<pre>
> reg<-lm(SalePrice~.,data=Reyeme2)

> summary(null_reg)

> null_reg<-lm(SalePrice~-1,data=Reyeme2)

> summary(null_reg)
</pre>

*전진선택법

```
> fs<-step(null_reg,direction="forward",scope=list(lower=null_reg,upper=reg))
Start:  AIC=11099.73
SalePrice ~ -1

            Df  Sum of Sq        RSS      AIC
+ LastPrice   1 1.6667e+13 3.4235e+10  8134.0
+ FirstPrice  1 1.6474e+13 2.2785e+11  8983.1
+ Tax         1 1.5608e+13 1.0937e+12  9685.9
+ Address   234 1.6118e+13 5.8306e+11  9870.1
+ Interior    1 1.4848e+13 1.8540e+12  9922.3
+ Condo       1 1.4563e+13 2.1389e+12  9986.4
+ Bath        1 1.4009e+13 2.6928e+12 10089.5
+ Rooms       1 1.3706e+13 2.9953e+12 10137.2
+ Bed         1 1.3393e+13 3.3084e+12 10181.8
+ Area       15 1.3516e+13 3.1851e+12 10192.8
+ CloseDate   1 1.3088e+13 3.6136e+12 10221.3
+ FirstDate 311 1.5754e+13 9.4731e+11 10241.5
+ Unit      179 1.4539e+13 2.1626e+12 10347.3
+ AreaCode    1 1.0237e+13 6.4643e+12 10481.8
+ Days        1 8.4636e+12 8.2380e+12 10590.5
+ RC          1 6.2859e+12 1.0416e+13 10695.6
<none>                     1.6702e+13 10905.1

Step:  AIC=8275.79
SalePrice ~ LastPrice - 1

            Df  Sum of Sq        RSS      AIC
+ Unit      179 2.4394e+10 9.8409e+09 7933.5
+ Area       15 5.2704e+09 2.8965e+10 8089.1
+ Interior    1 1.9137e+09 3.2321e+10 8110.2
+ Rooms       1 1.8044e+09 3.2431e+10 8111.7
+ Bed         1 1.7620e+09 3.2473e+10 8112.3
+ CloseDate   1 1.2784e+09 3.2957e+10 8118.9
+ RC          1 1.2629e+09 3.2972e+10 8119.1
+ AreaCode    1 1.2113e+09 3.3024e+10 8119.8
+ Condo       1 7.2157e+08 3.3514e+10 8126.4
+ Tax         1 5.7333e+08 3.3662e+10 8128.4
+ FirstPrice  1 5.5978e+08 3.3675e+10 8128.6
+ Bath        1 2.9864e+08 3.3936e+10 8132.0
<none>                     3.4235e+10 8134.0
+ Days        1 9.7217e+06 3.4225e+10 8135.8
+ Address   234 1.8227e+10 1.6008e+10 8261.4
+ FirstDate 311 2.1448e+10 1.2787e+10 8314.8
```

```
install.packages("car")
library("car")
library("MASS")

at=as.data.frame(read_excel("Reyeme2.xls"))
at
nrow(at) #456개 데이터
ncol(at) #17개 변수
str(at)
```

```
# > str(at)
# 'data.frame':    456 obs. of   17 variables:
# $ Address  : chr  "73 Allston St" "239 Allston St" "83 Amory St" "107 Antrim St" ...
# $ Unit     : chr  "1" "2" "3" "1" ...
# $ Area     : chr  "CP" "CP" "M" "M" ...
# $ AreaCode : num  4 4 9 9 9 9 9 9 9 9 ... (지역코드)
# $ FirstPrice: num  109000 127000 123000 219000 199000 ... (초기요구가격)
# $ LastPrice : num  109000 127000 100000 210000 189000 ... (판매전 마지막 요청가격)
# $ SalePrice : num  106000 125000 95000 206000 175000 ... (매매가격)
# $ FirstDate : chr  "34167" "34202" "34034" "33314" ... (판매 제안된 날짜)
# $ CloseDate : POSIXct, format: "1993-10-22" "1993-10-22" "1993-09-30" "1991-08-23" ... (마감일)
# $ Days     : num  97 62 208 159 110 116 160 80 157 123 ... (처음 개시부터 팔릴때까지의 숫
자)
# $ Interior : num  834 850 730 1460 1630 ... (내부면적)
# $ Bed      : num  2 2 2 3 4 2 2 3 3 3 ... (침실수)
# $ Bath     : num  1 1 1 2 1 1 1 1 1 1 ... (욕실수)
# $ Rooms    : num  4 5 4 6 8 6 6 5 6 6 ... (방수)
# $ Condo    : num  70 100 185 100 50 100 100 163 150 100 ... (월 콘도미니엄 또는 조합비)
# $ Tax      : num  490 923 575 1582 1659 ... (세)
# $ RC       : num  0 1 0 1 1 1 1 0 1 1 ... (소유자가 콘도미니엄을 허용하려는 경우 적용되는 임
대료 통제 상태)
```

```
modelW                                                                        =
lm(SalePrice~Unit+Area+AreaCode+FirstPrice+factor(FirstDate)+factor(CloseDate)+LastPrice+Days
+Interior+Bed+Bath+Rooms+Condo+Tax+RC, data=at)
modelW
vif(modelW)
```

vif.default(modelW)에서 다음과 같은 에러가 발생했습니다:

there are aliased coefficients in the model

모든 변수를 넣어 돌려보니, 다중공선성을 살펴볼수 없을 뿐만 아니라 회귀분석의 결과도 다소 이상하게 나오게 됨. 비슷한 변수들을 모아 각자 돌려본 뒤 추후 추려내는 방식을 선택함

● **Bed+Bath+Interior+Rooms(침실, 화장실, 인테리어, 방들)_더미변수화 X**

modelWAdjust1 = lm(SalePrice~Bed + Bath + Interior + Rooms, data=at)

Call:

lm(formula = SalePrice ~ Bed + Bath + Interior + Rooms, data = at)

Coefficients:

(Intercept)	Bed	Bath	Interior	Rooms
-1518.6	-13528.5	39111.9	169.6	-4887.7

➤ Bath, Interior를 제외하고 음의 관계

vif(modelWAdjust1)

Bed	Bath	Interior	Rooms
3.867804	1.614318	3.680481	4.750244

➤ 다중공선성 전부 5이하

stepAIC(modelWAdjust1, method="backward")

Call:

lm(formula = SalePrice ~ Bed + Bath + Interior, data = at)

Coefficients:

(Intercept)	Bed	Bath	Interior
-10373.0	-18140.6	40946.9	162.6

➤ Bed, Bath, Rooms 3개의 변수로 추려짐

anova(modelWAdjust1)

Analysis of Variance Table

Response: SalePrice

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Bed	1	6.0329e+11	6.0329e+11	141.7437	<2e-16 ***

```
Bath         1 6.1949e+11 6.1949e+11 145.5512 <2e-16 ***
Interior     1 5.1499e+11 5.1499e+11 120.9981 <2e-16 ***
Rooms        1 2.9439e+09 2.9439e+09   0.6917  0.406
Residuals 451 1.9195e+12 4.2562e+09
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

> ➢ anova결과 Rooms는 p-value가 높게 나옴

anova결과 Rooms는 제외하는 것이 좋아보인다. 또한, Bath, Interior를 제외하고 음의 관계가 나타나는 것이 다소 이상하여 factor화 할 수 있는 침실, 화장실, 방의 개수를 더미변수화 한 후에 모델을 돌려봄.

● Bed+Bath+Interior+Rooms(침실, 화장실, 인테리어, 방들)_더미변수화 O

modelWAdjust2 = lm(SalePrice~factor(Bed) + factor(Bath) + factor(Rooms) + Interior, data=at)

```
Call:
lm(formula = SalePrice ~ factor(Bed) + factor(Bath) + factor(Rooms) +
    Interior, data = at)



Coefficients:
   (Intercept)      factor(Bed)2    factor(Bed)3    factor(Bed)4    factor(Bath)2
       -6770.1         -32049.9        -27411.7        -46169.7         39029.7
 factor(Bath)3  factor(Rooms)3  factor(Rooms)4  factor(Rooms)5  factor(Rooms)6
       67018.8          -1277.3         -2961.9          3149.6        -10784.2
factor(Rooms)7  factor(Rooms)8  factor(Rooms)9        Interior
      -66231.4         -67848.9       -329751.5          194.4
```

> ➢ Interior는 양의 관계(내부 면적이 커질수록 Price는 높아짐)
> Bed및 Rooms는 많아질수록 대체적으로 음의 관계를 보임

vif(modelWAdjust2)

	GVIF	Df	GVIF^(1/(2*Df))
factor(Bed)	13.219640	3	1.537694
factor(Bath)	1.917167	2	1.176698
factor(Rooms)	18.910464	7	1.233653
Interior	4.215918	1	2.053270

> 다중공선성의 결과 Bath, Interiror를 제외하고 모두 5보다 큼

Bed, Rooms를 뺀 결과 내부면적과 화장실의 개수가 증가할수록 판매가격은 양의 관계로 정상적으로 보임 또한 다중공선성도 두 변수 모두 5이하였으며 backward방식으로 돌려본결과 Interior과 Bath는 유의한 변수로 보여짐

modelWAdjust3 = lm(SalePrice~Interior+factor(Bath), data=at)

```
Call:
lm(formula = SalePrice ~ Interior + factor(Bath), data = at)

Coefficients:
   (Intercept)          Interior   factor(Bath)2   factor(Bath)3
       24852.8             133.3         40901.4         94058.4
```

vif(modelWAdjust3)

```
GVIF Df GVIF^(1/(2*Df))
Interior      1.480845   1          1.216900
factor(Bath) 1.480845   2          1.103132
```

stepAIC(modelWAdjust3, method="backward")

```
Call:
lm(formula = SalePrice ~ Interior + factor(Bath), data = at)

Coefficients:
   (Intercept)          Interior   factor(Bath)2   factor(Bath)3
       24852.8             133.3         40901.4         94058.4
```

==결론적으로, 더미화 한 것이 더 정상적으로 보여졌으며 Interior과 Bath는 유의한 변수로 보여짐==

- **FirstPrice, LastPrice**

==같은 방식으로 진행했으나 다중공선성이 전부 20을 넘는 값이 나오기에 다른 변수와 함께 사용하기로 함.== 회귀분석 결과 FirstPrice는 음의 관계, LastPrice는 양의 관계를 보임

- **Tax, Condo, RC**

세 개의 항목 모두 회귀분석 결과 양의 관계가 나왔고 다중공선성은 전부 5 이하로 보여짐. ==stepAIC결과도 세 개의 변수 모두 유의한 값으로 나옴==

Day 항목은 크면 클수록 판매가격이 높게 나옴(양의관계). 아마도 큰 평수의 집일 수록 안 팔릴 확률이 높기 때문으로 보여짐. 이에 따라 종속변수에 log를 취한 결과는 다음과 같다.

```
Call:
lm(formula = log(SalePrice) ~ Days, data = at)

Coefficients:
(Intercept)          Days
 11.9710409     -0.0001049
```

로그를 취한결과 Day와 SalePrice는 음의 관계로 보여졌다. anova결과, Sale에 로그를 취한 결과값의 p-value가 더 낮게 나오는 것을 알 수 있었음

- **Area, AreaCode**

anova결과 둘다 함께 돌렸을 경우 AreaCode는 p-value가 너무 높게 형성되고 더미변수화하여 돌렸을 경우 NA 값도 다수 보였다. 따로 돌려본 결과는 다음과 같다.
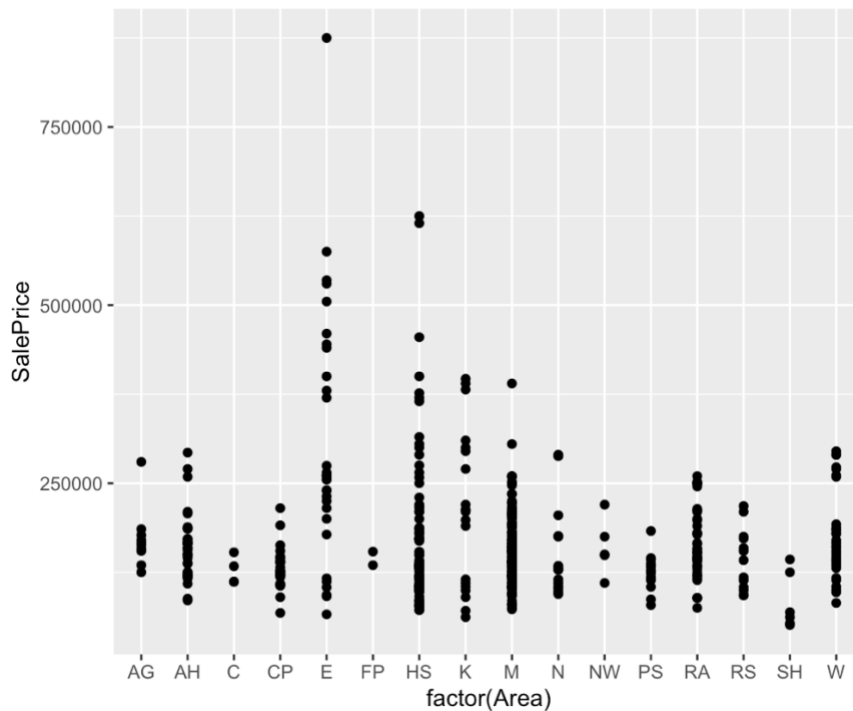
Area를 factor화 후 단순회귀로 돌려본 결과)

```
Call:
lm(formula = SalePrice ~ factor(Area), data = at)

Coefficients:
   (Intercept)   factor(Area)AH    factor(Area)C   factor(Area)CP
        170136            -9899           -42636           -37886
 factor(Area)E   factor(Area)FP   factor(Area)HS    factor(Area)K
        135017           -25636             1539            41817
 factor(Area)M    factor(Area)N   factor(Area)NW   factor(Area)PS
        -12006           -19455            -9336           -45403
factor(Area)RA   factor(Area)RS   factor(Area)SH    factor(Area)W
         -5751           -28510           -86220             5126
```

Reference그룹은 AG로 보여진다. 더미변수가 다소 많아 이를 plot으로 그려본 결과는 다음과 같다.

ggplot으로 그려본 결과 E와 HS의 매매가격이 다소 높게 형성되는 것을 알 수 있음

AreaCode를 factor화 후 단순회귀로 돌려본 결과

| (Intercept) | factor(AreaCode)2 | factor(AreaCode)3 |
|---|---|---|
| 170136 | -9899 | -42636 |
| factor(AreaCode)4 | factor(AreaCode)5 | factor(AreaCode)6 |
| -39351 | 135017 | -25636 |
| factor(AreaCode)7 | factor(AreaCode)8 | factor(AreaCode)9 |
| 1909 | 41817 | -12006 |
| factor(AreaCode)10 | factor(AreaCode)11 | factor(AreaCode)12 |
| -22080 | -9336 | -45403 |
| factor(AreaCode)13 | factor(AreaCode)14 | factor(AreaCode)15 |
| -5751 | -28510 | -86220 |
| factor(AreaCode)16 | | |
| 5126 | | |

Reference 그룹은 1그룹

AreaCode도 다소 더미변수가 많아 plot으로 그려본 결과는 다음과 같다.
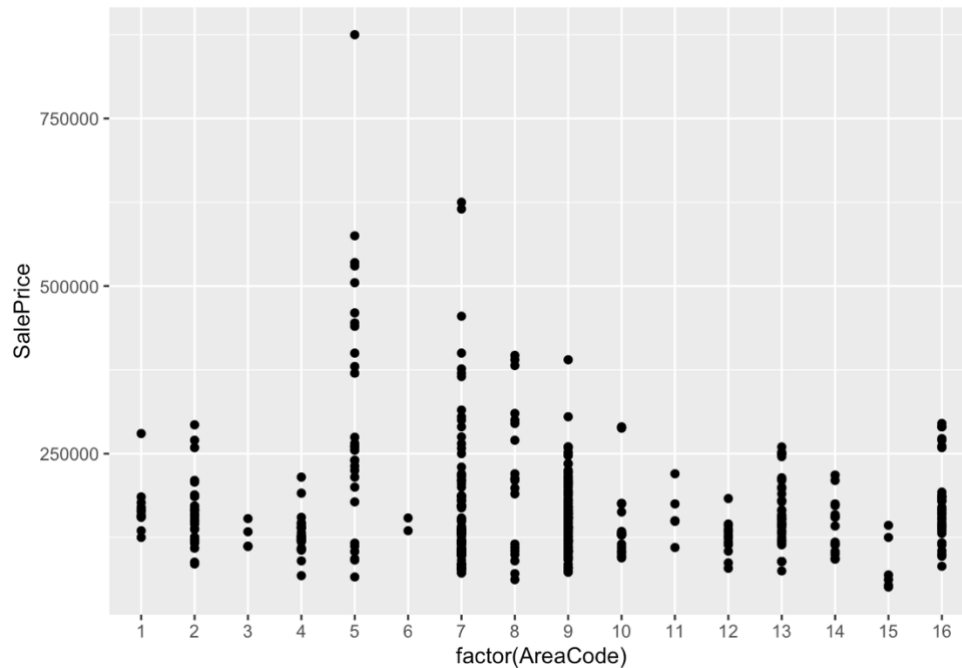
결과, AreaCode와 Area는 같은 변수인데 표기만 다르게 하였음을 알 수 있음. <mark>둘 중에 하나는 빼고 다중회귀를 돌려도 되는 것을 볼 수 있었다.</mark>

● **결론 모델**

결론적으로, Interior, Bath, FirstPrice, LastPrice, Tax, Condo, RC, Area로 stepAIC를 한 결과는 다음과 같았다.

<mark>lm(formula = log(SalePrice) ~ factor(Area) + Interior + factor(Bath) + LastPrice + Condo + RC + Days, data = at)</mark>

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|---|---|---|---|---|---|---|
| factor(Area) | 15 | 14.064 | 0.938 | 55.2082 | < 2.2e-16 | *** |
| Interior | 1 | 39.906 | 39.906 | 2349.6985 | < 2.2e-16 | *** |
| factor(Bath) | 2 | 0.599 | 0.300 | 17.6409 | 4.313e-08 | *** |
| LastPrice | 1 | 18.725 | 18.725 | 1102.5414 | < 2.2e-16 | *** |
| Condo | 1 | 0.077 | 0.077 | 4.5592 | 0.03330 | * |
| RC | 1 | 0.099 | 0.099 | 5.8569 | 0.01593 | * |
| Days | 1 | 0.130 | 0.130 | 7.6688 | 0.00586 | ** |
| Residuals | 433 | 7.354 | 0.017 | | | |