

1번 문제

(1)

```
> str(homeData)
```

```
'data.frame':    30 obs. of  4 variables:
```

```
 $ Price  : chr  "505.5" "784.1" "649.0" "689.8" ...
```

```
 $ SqFt   : chr  "2192" "3429" "2842" "2987" ...
```

```
 $ LotSize: chr  "16.4" "24.7" "17.7" "20.3" ...
```

```
 $ Baths  : chr  "2.5" "3.5" "3.5" "3.5" ...
```

```
> homeData$Price <- as.numeric(homeData$Price)
```

```
> homeData$SqFt <- as.numeric(homeData$SqFt)
```

```
> homeData$LotSize <- as.numeric(homeData$LotSize)
```

```
> homeData$Baths <- as.numeric(homeData$Baths)
```

```
> model = lm(formula = Price~SqFt+LotSize+Baths, data=homeData)
```

```
> anova(model)
```

Analysis of Variance Table

Response: Price

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SqFt	1	222330	222330	539.2202	< 2.2e-16 ***
LotSize	1	8947	8947	21.6988	8.286e-05 ***
Baths	1	1173	1173	2.8461	0.1036
Residuals	26	10720	412		

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

(2) 답: 결정계수는 0.9559이며, 수정된 결정계수는 0.9508이다. 결정계수는 독립변수 개수가 많아 질수록 그 값이 커지게 된다. 보다 정확한 추정치를 얻기 위해서는 수정결정계수를 사용하는 것이 적합하다. 특히나 표본이 200개 미만일 때에는 수정결정계수를 고려해야한다.

```
> summary(model)
```

Call:

```
lm(formula = Price ~ SqFt + LotSize + Baths, data = homeData)
```

Residuals:

Min	1Q	Median	3Q	Max
-49.019	-11.781	-2.312	14.839	33.240

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
----------	------------	---------	----------

```
(Intercept) -28.84775    29.71148   -0.971    0.341
SqFt         0.17091     0.01545   11.064 2.48e-11 ***
LotSize      6.77770     1.42129    4.769 6.19e-05 ***
Baths       15.53470     9.20827    1.687   0.104
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.31 on 26 degrees of freedom
Multiple R-squared:  0.9559,    Adjusted R-squared:  0.9508
F-statistic: 187.9 on 3 and 26 DF,  p-value: < 2.2e-16
```

(3) 답: 다중 공선성 문제는 모든 변수에서 관측되지 않음

```
> vif(out)
      SqFt  LotSize   Baths 
2.615542 1.617827 1.900531
```

(4) 답: 659.6985

```
> df = data.frame(SqFt=2955, LotSize=20.2, Baths=3)
> predict(model,new=df)
      1 
659.6985
```

2번 문제

(1)

b5 : -7.27780, b6 : -1.92319 b7 : -5.67146

b5는 유의수준 10%에서 p-value가 유의수준 보다 작으므로 H_0 를 기각한다. 따라서, 유의수준 10%에서 기울기 b5는 0이 아니며 이 모델은 유의하다.

b6는 유의수준 10%에서 p-value가 유의수준 크므로 H_0 를 기각하지 못한다. 따라서, 유의수준 10%에서 기울기 b6는 0이 되며 이 모델은 유의하다 볼 수 없다.

b7는 유의수준 10%에서 p-value가 유의수준 보다 작으므로 H_0 를 기각한다. 따라서, 유의수준 10%에서 기울기 b7는 0이 아니며 이 모델은 유의하다.

```
> model2 = lm(choice~Age65+Urban+ColGrad+Union+Area, data=electionData)
> summary(model2)
```

Call:

```
lm(formula = choice ~ Age65 + Urban + ColGrad + Union + Area,  
    data = electionData)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-10.2291	-2.6059	-0.2298	2.8764	8.2806

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	100.22162	8.24377	12.157	2.42e-15	***
Age65	-1.28688	0.40096	-3.210	0.002549	**
Urban	-0.09827	0.03476	-2.827	0.007158	**
ColGrad	-0.58151	0.19330	-3.008	0.004425	**
Union	-0.72808	0.13274	-5.485	2.18e-06	***
AreaNeast	-7.27780	1.91986	-3.791	0.000474	*** (b5)
AreaSeast	-5.67146	2.03405	-2.788	0.007928	** (b7)
AreaWest	-1.92319	1.88728	-1.019	0.314025	(b6)

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

Residual standard error: 4.287 on 42 degrees of freedom

Multiple R-squared: 0.7925, Adjusted R-squared: 0.758

F-statistic: 22.92 on 7 and 42 DF, p-value: 1.935e-12

(2)

b5 계수 = Neast choice 평균 - MidWest choice 평균

-7.27780 = Neast choice 평균 - 100.22162

즉, Neast choice 평균 = 약 93

b6 계수 = West choice 평균 - MidWest choice 평균

-1.92319 = Neast choice 평균 - 100.22162

즉, West choice 평균 = 약 102

b7 계수 = Seast choice 평균 - MidWest choice 평균

-5.67146 = Seast choice 평균 - 100.22162

즉, Seast choice 평균 = 약 105

(3) 답: 변수는 Age65, Urban, ColGrad, Union, Area, AIC의 값은 152.83

(4) 답: 15.7486