Hw 7

20150056 국윤범

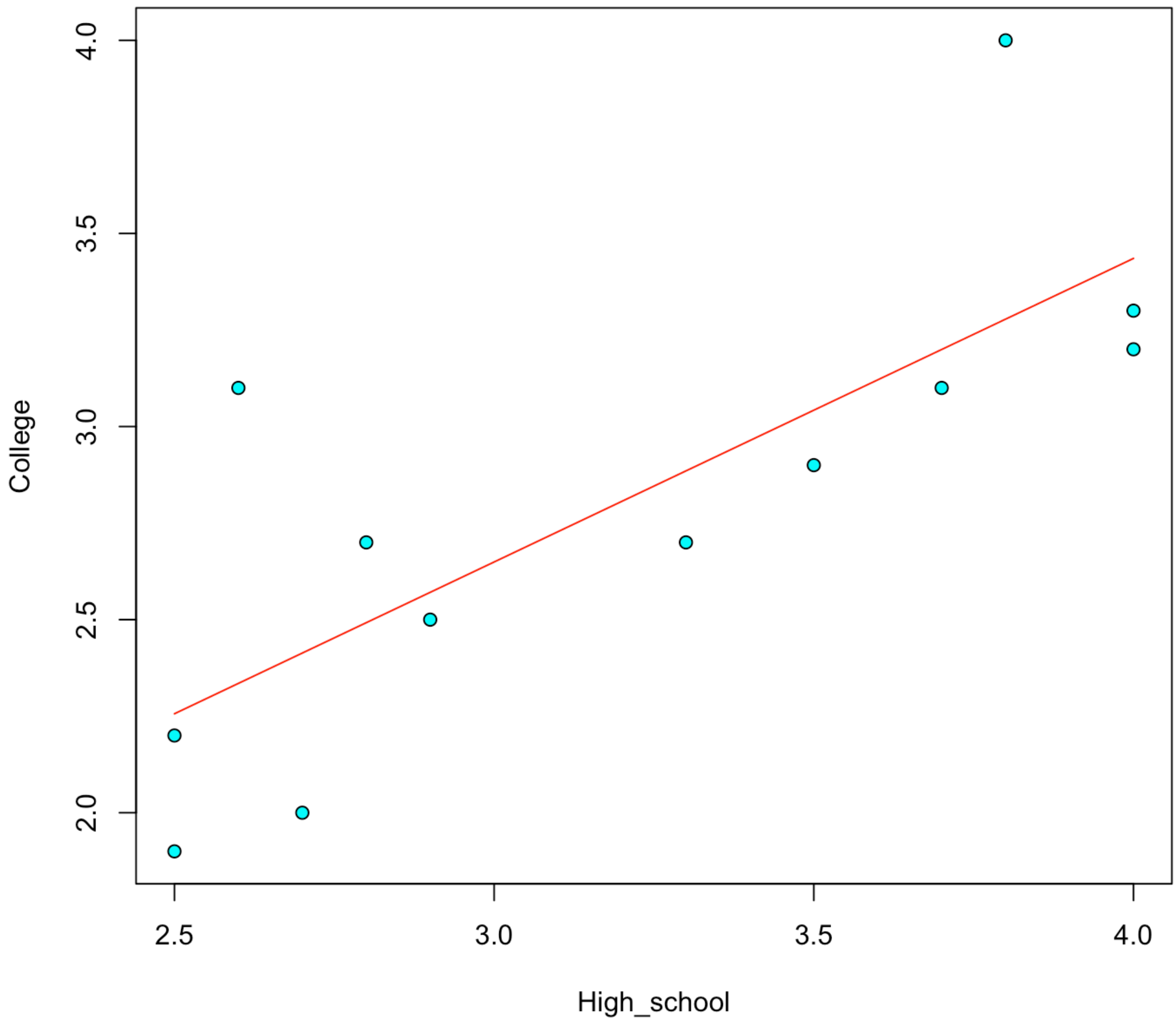
Problem 1

(a)

```
Call:
lm(formula = College ~ High_school, data = df1)

Coefficients:
(Intercept)  High_school
     0.2911       0.7861
```

➔ College's GPA = 0.2911 + 0.7861 * High_school's GPA

(b)



(c)

```
> predict(fit, list(High_school=3.0))
       1
2.649336
```

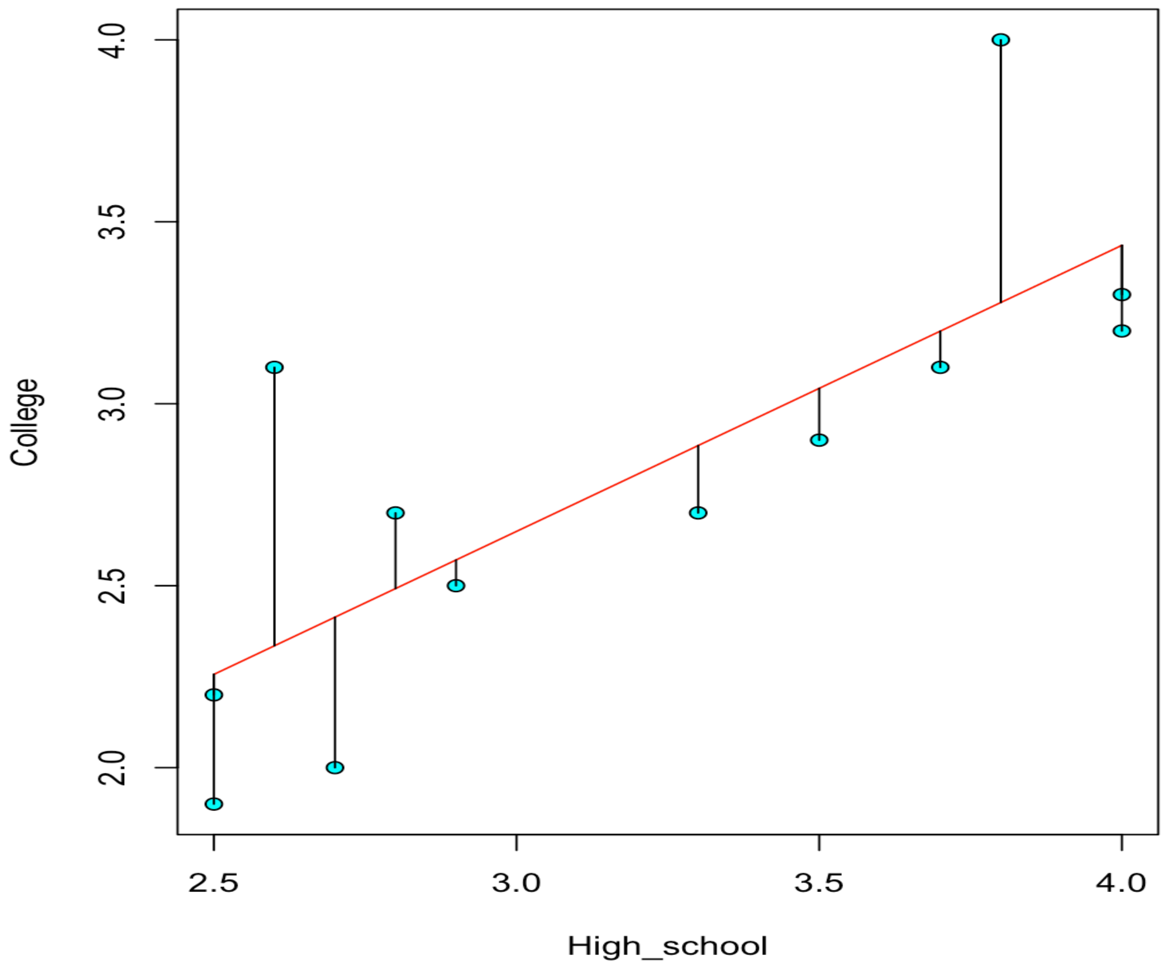
(d)

```
> summary(fit)

Call:
lm(formula = College ~ High_school, data = df1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.41351 -0.19772 -0.11750  0.00974  0.76509

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.2911     0.6613   0.440   0.6691
High_school   0.7861     0.2040   3.853   0.0032 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3992 on 10 degrees of freedom
Multiple R-squared:  0.5975,    Adjusted R-squared:  0.5572
F-statistic: 14.84 on 1 and 10 DF,  p-value: 0.003197
```

Hence, the coefficient of determination is 0.5975 and the coefficient of non-determination is 0.4025.

(e) From above, p-value is 0.0032, which is smaller than 0.05. Hence, we can say that there is a significant relationship between GPA in college and high school.

(f)



Problem 2

```
> summary(mfit2)

Call:
lm(formula = form, data = kfm)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.74201 -0.81173 -0.00926  0.78326  2.52646 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -11.681839   4.361561  -2.678 0.010363 *  
sexgirl      -0.499532   0.312672  -1.598 0.117284    
weight        1.349124   0.322450   4.184 0.000135 ***
ml.suppl     -0.002233   0.001241  -1.799 0.078829 .  
mat.weight    0.006212   0.023708   0.262 0.794535    
mat.height    0.072278   0.030169   2.396 0.020906 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.075 on 44 degrees of freedom
Multiple R-squared:  0.5459,	Adjusted R-squared:  0.4943 
F-statistic: 10.58 on 5 and 44 DF,  p-value: 1.03e-06
```

From above, weight and mat.height are significant factors. For sex in the form of factor, it means that we need to make a dummy variable which is 1 for girl and 0 for boy. Since linear regression is basically working on numeric explanatory variables, we need to change factor columns into numeric column. By the way, the p value 0.117 of sex, we can say sex is not an influence factor.

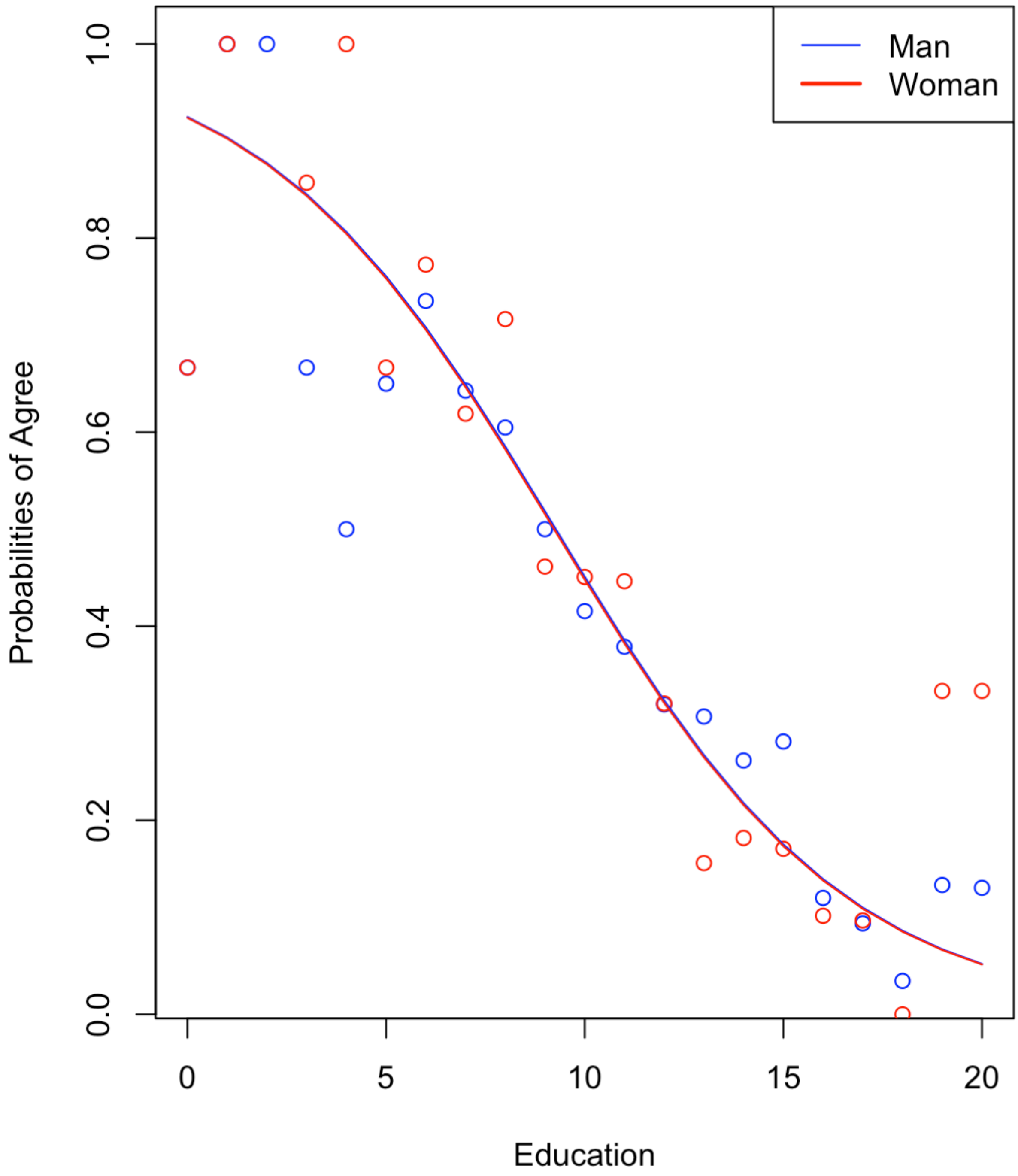
Problem 4 (Problem 3 is optional, so skipped)

(a)

```
Call:
glm(formula = fm, family = binomial(), data = womensrole)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.72544  -0.86302  -0.06525   0.84340   3.13315

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.50937    0.18389  13.646   <2e-16 ***
education    -0.27062    0.01541 -17.560   <2e-16 ***
genderFemale -0.01145    0.08415  -0.136    0.892
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 451.722  on 40  degrees of freedom
Residual deviance:  64.007  on 38  degrees of freedom
AIC: 208.07

Number of Fisher Scoring iterations: 4
```

From the p-value, the response significantly differs by education level and not by gender.

(b)


Man
Woman
Probabilities of Agree
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
Education