CSC 423 Project 3

Part A:

1.

Age Education Income HomeVal Wealth Balance

1 35.9 14.8 91033 183104 220741 38517

2 37.7 13.8 86748 163843 223152 40618

3 36.8 13.8 72245 142732 176926 35206

4 35.3 13.2 70639 145024 166260 33434

5 35.3 13.2 64879 135951 148868 28162

6 34.8 13.7 75591 155334 188310 36708

7 39.3 14.4 80615 181265 201743 38766

8 36.6 13.9 76507 149880 189727 34811

9 35.7 16.1 107935 276139 211085 41032

10 40.5 15.1 82557 182088 220782 41742

11 37.9 14.2 58294 123500 132432 29950

12 43.1 15.8 88041 194369 267556 51107

13 37.7 12.9 64597 119305 186156 34936

14 36.0 13.1 64894 141011 160017 32387

15 40.4 16.1 61091 194928 113559 32150

16 33.8 13.6 76771 159531 197264 37996

17 36.4 13.5 55609 123085 105582 24672

18 37.7 12.8 74091 143750 217869 37603

19 36.2 12.9 53713 112649 117441 26785

20 39.1 12.7 60262 126928 161322 32576

21 39.4 16.1 111548 230893 331009 56569

22 36.1 12.8 48600 105737 106671 26144

23 35.3 12.7 51419 104149 111168 24558

24 37.5 12.8 51182 106898 88370 23584

25 34.4 12.8 60753 95869 143115 26773

26 33.7 13.8 64601 103737 134223 27877

27 40.4 13.2 62164 114257 144038 28507

28 38.9 12.7 46607 94576 114799 27096

29 34.3 12.7 61446 122619 161538 28018

30 38.7 12.8 62024 134430 149351 31283

31 33.4 12.6 54986 105647 126929 24671

32 35.0 12.0 48182 114436 102732 25280

33 38.1 12.7 47388 92820 118016 24890

34 34.9 12.5 55273 102468 126959 26114

35 36.1 12.9 53892 92968 129176 27570

36 32.7 12.6 47923 104539 88384 20826

37 37.1 12.5 46176 92654 101964 23858

38 23.5 13.6 33088 105430 44223 20834

39 38.0 13.6 53890 108446 95013 26542

40 33.6 12.7 57390 111836 134434 27396

41 41.7 13.0 48439 100788 124474 31054

42 36.6 14.1 56803 149138 101695 29198

43 34.9 12.4 52392 93875 133101 24650

44 36.7 12.8 48631 95490 105202 23610

45 38.4 12.5 52500 105377 139199 29706

46 34.8 12.5 42401 106478 94867 21572

47 33.6 12.7 64792 116071 185714 32677

48 37.0 14.1 59842 106949 135329 29347

49 34.4 12.7 65625 129688 175000 29127

50 37.2 12.5 54044 108654 140726 27753

51 35.7 12.6 39707 89552 80124 21345

52 37.8 12.9 45286 108431 91928 28174

53 35.6 12.8 37784 92712 60721 19125

54 35.7 12.4 52284 92143 146028 29763

55 34.3 12.4 42944 86192 98778 22275

56 39.8 13.4 46036 99508 98343 27005

57 36.2 12.3 50357 90750 126613 24076

58 35.1 12.3 45521 82720 105346 23293

59 35.6 16.1 30418 139739 24999 16854

60 40.7 12.7 52500 94792 147222 28867

61 33.5 12.5 41795 94456 91806 21556

62 37.5 12.5 66667 78906 143750 31758

63 37.6 12.9 38596 95364 54453 17939

64 39.1 12.6 44286 93103 110465 22579

65 33.1 12.2 37287 75561 86591 19343

66 36.4 12.9 38184 80099 76438 21534

67 37.3 12.5 47119 88958 102993 22357

68 38.7 13.6 44520 96112 93915 25276

69 36.9 12.7 52838 101705 75040 23077

70 32.7 12.3 34688 82870 93750 20082

71 36.1 12.4 31770 74525 47446 15912

72 39.5 12.8 32994 89223 50592 21145

73 36.5 12.3 33891 72739 81880 18340

74 32.9 12.4 37813 86667 69643 19196

75 29.9 12.3 46528 88889 96591 21798

76 32.1 12.3 30319 67083 34367 13677

77 36.1 13.3 36492 172768 24999 20572

78 35.9 12.4 51818 80357 135185 26242

79 32.7 12.2 35625 64737 76321 17077

80 37.2 12.6 36789 86563 69764 20020

81 38.8 12.3 42750 77717 95192 25385

82 37.5 13.0 30412 138911 24999 20463

83 36.4 12.5 37083 70909 95833 21670

84 42.4 12.6 31563 81597 71759 15961

85 19.5 16.1 15395 67500 24999 5956

86 30.5 12.8 21433 83456 24999 11380

87 33.2 12.3 31250 91049 52976 18959

88 36.7 12.5 31344 77541 36510 16100

89 32.4 12.6 29733 60252 27531 14620

90 36.5 12.4 41607 76270 98455 22340

91 33.9 12.1 32813 40313 79167 26405

92 29.6 12.1 29375 52096 24999 13693

93 37.5 11.1 34896 65357 81818 20586

94 34.0 12.6 20578 113239 24999 14095

95 28.7 12.1 32574 50244 49662 14393

96 36.1 12.2 30589 69375 48890 16352

97 30.6 12.3 26565 64038 42543 17410

98 22.8 12.3 16590 67850 24999 10436

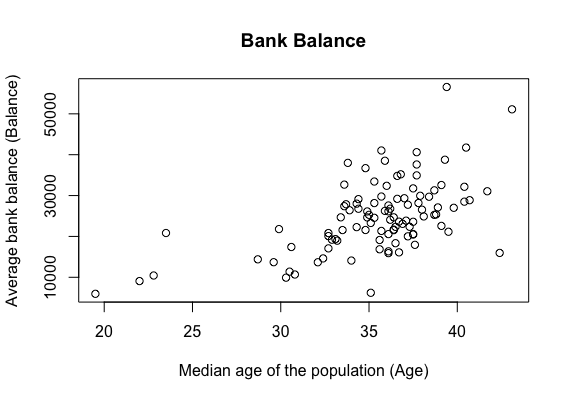
99 30.3 12.2 9354 91708 24999 9904

100 22.0 12.0 14115 53923 24999 9071

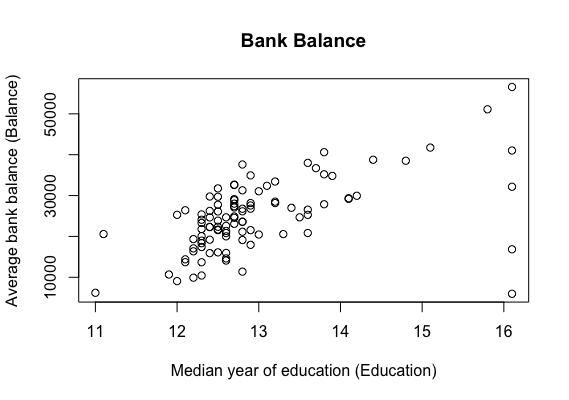
101 30.8 11.9 17992 46885 24999 10679

102 35.1 11.0 7741 99375 24999 6207

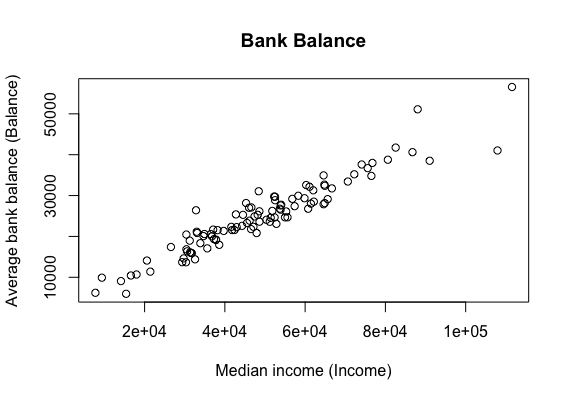
2.



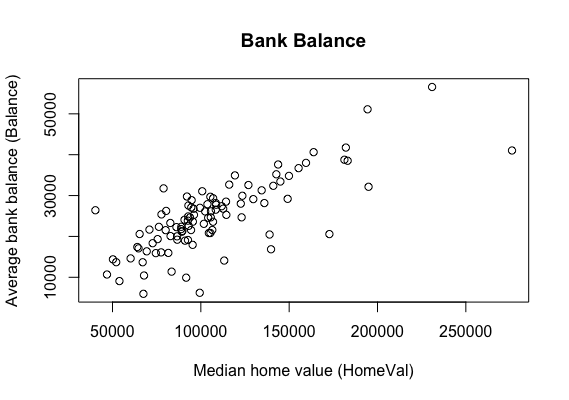
The pattern is random, not close to a straight line, balance and age are not linear.



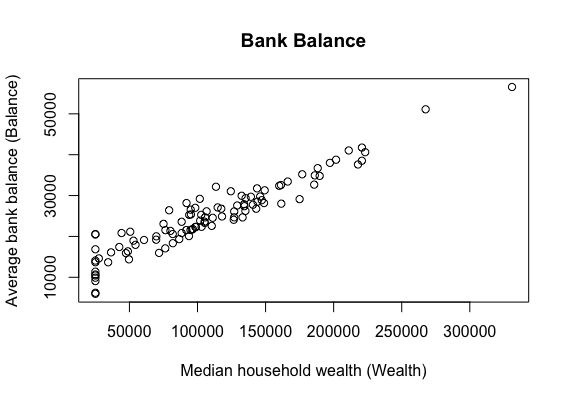
The pattern is random, not a straight line, balance and education are not linear.



The pattern is close to a straight line, so balance and income are linear.



The pattern is almost close to straight line, so balance and home value is linear.



The pattern is close to a straight line, so balance and wealth are linear.

3.

Correlation of Age and Balance: 0.5654668

Correlation of Education and Balance: 0.5521889

Correlation of Income and Balance: 0.9516845

Correlation of Home Value and Balance: 0.7663871

Correlation of Wealth and Balance: 0.9487117

From the r value of these correlations, since r value of age, education is 0.5, they are moderate positive linear relationships with balance; since r value of income, home value and wealth is larger than 0.7, they are strong positive linear relationships with balance.

4.

Call:

lm(formula = Balance ~ Age + Education + Income + HomeVal + Wealth,

data = Banking)

Residuals:

Min 1Q Median 3Q Max

-5365.5 -1102.6 -85.9 868.9 7746.5

Coefficients:

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

(Intercept) -1.033e+04 4.219e+03 -2.449 0.016160 \*

Age 3.175e+02 6.104e+01 5.201 1.12e-06 \*\*\*

Education 5.903e+02 3.151e+02 1.873 0.064085 .

Income 1.468e-01 4.083e-02 3.596 0.000512 \*\*\*

HomeVal 9.864e-03 1.099e-02 0.898 0.371591

Wealth 7.414e-02 1.120e-02 6.620 2.06e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2059 on 96 degrees of freedom

Multiple R-squared: 0.9468, Adjusted R-squared: 0.944

F-statistic: 341.4 on 5 and 96 DF, p-value: < 2.2e-16

The regression model is: y = 317.5xAge + 590.3xEducation + 0.1468xIncome + 0.009864xHomeValue + 0.07414xWealth – 0.0001033

5.

dfb.1\_ dfb.Age dfb.Edct dfb.Incm dfb.HmVl dfb.Wlth dffit cov.r cook.d hat

9 -0.22 0.43 0.27 -0.68 -0.66 0.78 -1.51\_\* 1.15 0.37 0.32\_\*

12 -0.58 0.31 0.53 -0.59 0.03 0.67 1.02\_\* 1.02 0.17 0.19\_\*

15 -0.13 0.07 0.10 -0.02 0.04 -0.03 0.19 1.25\_\* 0.01 0.16

21 -0.12 -0.05 0.14 -0.18 0.06 0.27 0.47 1.27\_\* 0.04 0.20\_\*

38 0.43 -1.12\_\* 0.11 0.11 0.21 -0.18 1.34\_\* 0.57\_\* 0.27 0.12

52 -0.05 0.17 -0.02 0.01 0.05 -0.07 0.30 0.81\_\* 0.01 0.02

59 0.29 -0.10 -0.30 0.07 0.01 0.03 -0.40 1.34\_\* 0.03 0.23\_\*

77 0.18 0.04 -0.29 0.02 0.58 -0.30 0.74 1.17 0.09 0.20\_\*

84 0.44 -0.75 -0.20 0.36 0.06 -0.20 -0.88\_\* 0.72\_\* 0.12 0.09

85 1.12\_\* 1.14\_\* -2.17\_\* 0.75 0.91 -0.77 -2.81\_\* 1.22\_\* 1.22\_\* 0.48\_\*

91 -0.16 0.05 0.33 -0.03 -0.74 0.22 1.00\_\* 0.41\_\* 0.14 0.05

99 -0.05 0.03 0.03 0.13 -0.10 -0.09 -0.16 1.24\_\* 0.00 0.15

100 0.10 -0.15 -0.02 -0.05 0.03 0.05 0.19 1.23\_\* 0.01 0.14

102 -0.72 -0.14 0.83 1.20\_\* -1.33\_\* -0.76 -1.71\_\* 0.84 0.45 0.25\_\*

So they are some influence points in this model.

6.

Wealth, income and age are all significant, but p-value of wealth is the smallest, so wealth has a significant effect on balance.

7.

Home Value has the largest p value, then remove.

Call:

lm(formula = Balance ~ Age + Education + Income + Wealth, data = Banking)

Residuals:

Min 1Q Median 3Q Max

-5403.9 -1234.1 -75.0 998.6 7430.7

Coefficients:

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

(Intercept) -1.214e+04 3.704e+03 -3.278 0.00145 \*\*

Age 3.242e+02 6.051e+01 5.358 5.68e-07 \*\*\*

Education 7.498e+02 2.600e+02 2.884 0.00484 \*\*

Income 1.615e-01 3.738e-02 4.321 3.75e-05 \*\*\*

Wealth 7.265e-02 1.106e-02 6.566 2.57e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2057 on 97 degrees of freedom

Multiple R-squared: 0.9463, Adjusted R-squared: 0.9441

F-statistic: 427.4 on 4 and 97 DF, p-value: < 2.2e-16

The regression model is: y = 324.2xAge + 749.8xEducation + 0.1615xIncome + 0.07265xWealth – 0.0001033

8.

The rest four predictors are all significant association with balance.

9.

The regression coefficients for independent variables, Age: 324.2, Education: 749.8, Income: 0.1615, Wealth: 0.07265.

10.

Age: 0.1448635; Education: 0.08702261; Income: 0.3595092; Wealth: 0.4997957;

Since standardized Coefficients of wealth is the largest, it is the strongest influence.

11.

R-squared for the final model is 0.9463, it is close to 1, so it means the model is fitted to the data.

12.

i.

H0: β1 = β2 = , ... , = βp-1 = 0

H1: βj ≠ 0 for some j

ii.

F = 427.4

iii.

DFE = n – p = 102 – 5 = 97, DFM = 5 – 1 = 4, confidence interval is [0, 2.46548]

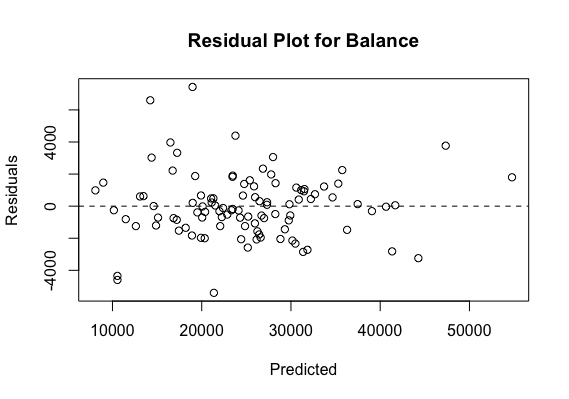
iv.

427.4∉ [0, 2.46548], so reject H0.

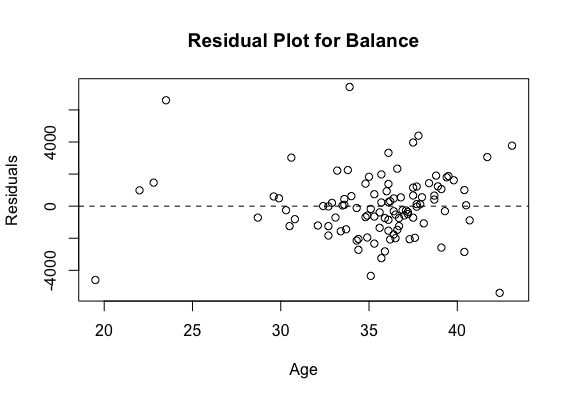
v.

p-value < 2.2e-16

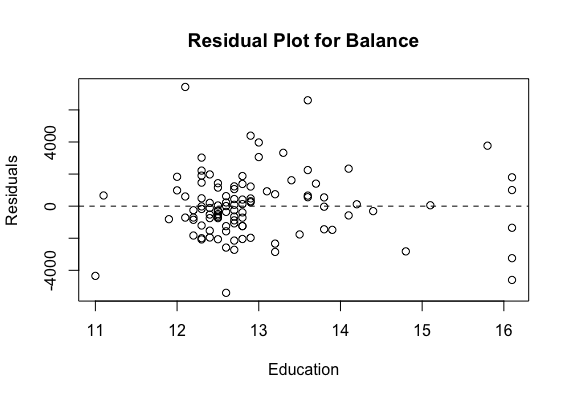
13.



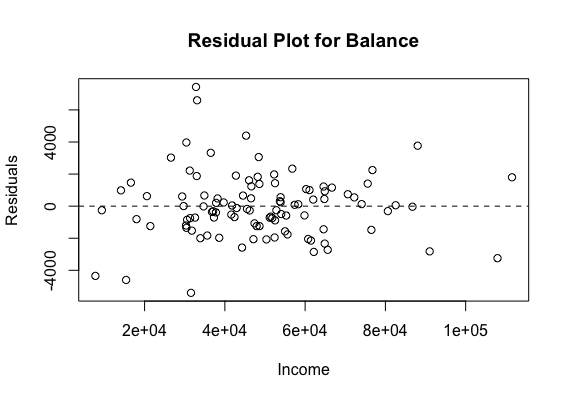
The pattern is unbiased and homoscedastic.



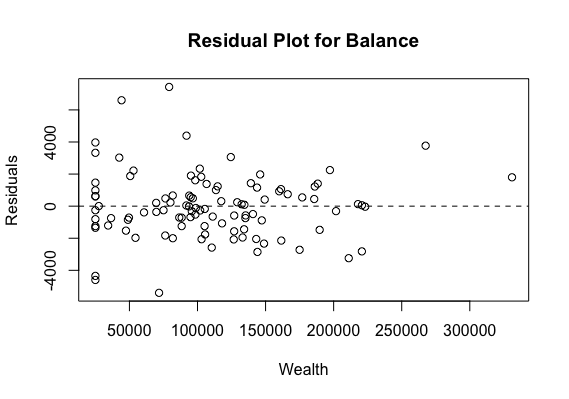
The pattern is unbiased and homoscedastic.



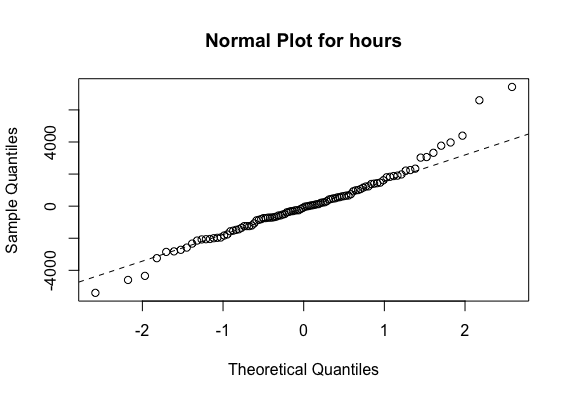
The pattern is unbiased and homoscedastic.



The pattern is unbiased and homoscedastic.



The pattern is unbiased and homoscedastic.



The pattern is close to a straight line.

14.

dfb.1\_ dfb.Age dfb.Edct dfb.Incm dfb.Wlth dffit cov.r cook.d hat

9 0.09 0.28 -0.10 -0.82 0.71 -1.09\_\* 1.19\_\* 0.23 0.26\_\*

12 -0.68 0.32 0.67 -0.64 0.67 1.02\_\* 1.05 0.20 0.19\_\*

15 -0.20 0.09 0.18 0.00 -0.04 0.23 1.23\_\* 0.01 0.15\_\*

21 -0.17 -0.05 0.23 -0.18 0.28 0.49 1.25\_\* 0.05 0.20\_\*

38 0.38 -1.11\_\* 0.27 0.22 -0.22 1.34\_\* 0.63\_\* 0.32 0.12

52 -0.09 0.18 0.01 0.03 -0.08 0.30 0.84\_\* 0.02 0.02

59 0.33 -0.10 -0.36 0.09 0.03 -0.41 1.33\_\* 0.03 0.23\_\*

84 0.47 -0.76 -0.21 0.42 -0.21 -0.89\_\* 0.77\_\* 0.15 0.09

85 0.78 1.27\_\* -2.02\_\* 1.22\_\* -0.92 -2.68\_\* 1.15 1.32\_\* 0.43\_\*

91 0.20 -0.04 -0.10 -0.33 0.31 0.63 0.51\_\* 0.07 0.02

98 0.13 -0.22 0.01 -0.06 0.06 0.28 1.16\_\* 0.02 0.12

100 0.11 -0.17 0.00 -0.05 0.06 0.21 1.21\_\* 0.01 0.14

102 -0.07 -0.21 0.07 0.51 -0.40 -0.75\_\* 0.90 0.11 0.10

There are some influence points in the final model.

Part B:

1.

type rpm hours

1 A 243 38.7

2 A 340 33.5

3 A 381 35.1

4 A 415 23.1

5 A 449 27.2

6 A 470 25.0

7 A 529 17.1

8 A 588 11.0

9 A 612 23.0

10 A 672 11.1

11 B 519 40.3

12 B 522 43.1

13 B 624 35.9

14 B 711 36.3

15 B 712 26.4

16 B 791 37.7

17 B 820 20.2

18 B 900 14.6

19 B 947 17.1

20 B 988 11.2

2.

Call:

lm(formula = hour ~ rpm + type\_dum, data = with\_dummy\_B)

Residuals:

Min 1Q Median 3Q Max

-5.8963 -3.6395 -0.3763 1.2834 11.8344

Coefficients:

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

(Intercept) 54.654176 3.858993 14.163 7.67e-11 \*\*\*

rpm -0.064214 0.007547 -8.509 1.56e-07 \*\*\*

type\_dum 22.004679 3.034917 7.251 1.36e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

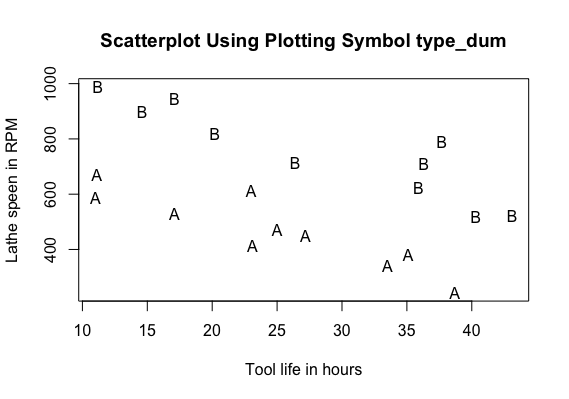
Residual standard error: 4.813 on 17 degrees of freedom

Multiple R-squared: 0.8163, Adjusted R-squared: 0.7946

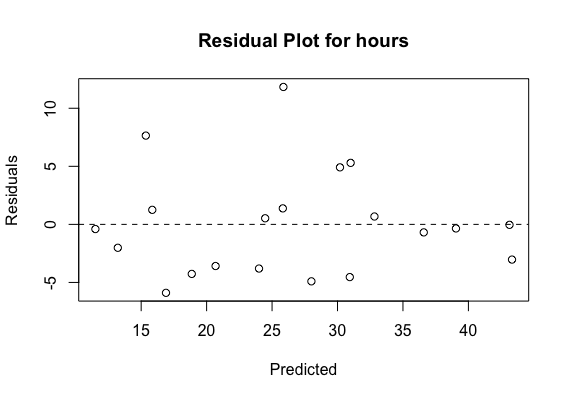
F-statistic: 37.76 on 2 and 17 DF, p-value: 5.569e-07

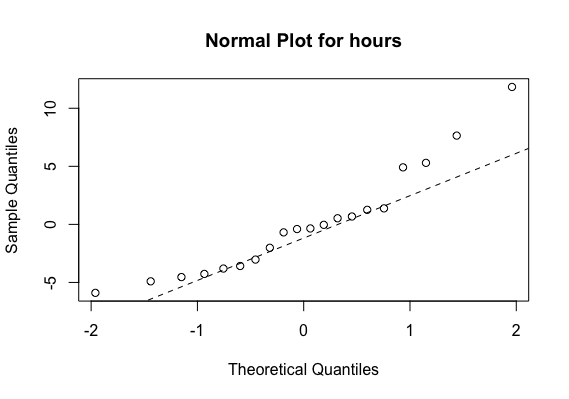
The regression model is: y = 22.004679xtype\_dum – 0.064214xrpm + 54.654176

3.



4.





5.

Yes, it has. Since the type\_dum has a 3 stars on significant, it has a significant effect on hours.

Part C:

1.

salary exper mgt\_dum educ\_dum\_1 educ\_dum\_2

1 58418 1 0 0 0

2 48870 1 1 0 1

3 78731 1 0 0 1

4 47501 1 1 1 0

5 49539 1 1 0 1

6 87871 2 0 1 0

7 49560 2 1 1 0

8 44352 2 1 0 0

9 51341 2 1 0 1

10 51838 3 1 1 0

11 63045 3 0 0 0

12 89972 3 0 1 0

13 83358 3 0 0 1

14 48066 4 1 0 0

15 85307 4 0 0 1

16 55703 4 1 0 1

17 54242 4 1 1 0

18 55761 5 1 1 0

19 57580 5 1 0 1

20 67213 5 0 0 0

21 51935 6 1 0 0

22 89892 6 0 0 1

23 58262 6 1 1 0

24 96342 6 0 1 0

25 71477 7 0 0 0

26 62321 8 1 1 0

27 73271 8 0 0 0

28 93395 8 0 0 1

29 57037 8 1 0 0

30 60906 10 1 0 0

31 67116 10 1 1 0

32 97563 10 0 0 1

33 100114 10 0 1 0

34 106976 11 0 1 0

35 62565 11 1 0 0

36 71073 12 1 1 0

37 101764 12 0 0 1

38 67318 13 1 0 0

39 110849 13 0 1 0

40 75565 14 1 1 0

41 108134 15 0 0 1

42 117194 16 0 1 0

43 79308 16 1 1 0

44 73603 16 1 0 0

45 80861 17 1 1 0

46 81447 20 1 0 0

2.

Call:

lm(formula = salary ~ exper + educ\_dum\_1 + educ\_dum\_2 + mgt\_dum,

data = with\_dummy\_C)

Residuals:

Min 1Q Median 3Q Max

-7934.3 -2751.7 93.6 3556.8 7226.3

Coefficients:

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

(Intercept) 62809.4 1777.4 35.337 < 2e-16 \*\*\*

exper 2299.5 128.5 17.897 < 2e-16 \*\*\*

educ\_dum\_1 13236.3 1523.9 8.686 7.73e-11 \*\*\*

educ\_dum\_2 12614.7 1733.5 7.277 6.71e-09 \*\*\*

mgt\_dum -28980.0 1321.6 -21.928 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

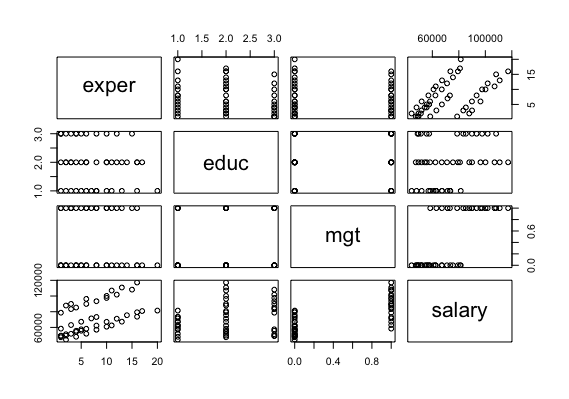
Residual standard error: 4325 on 41 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9526

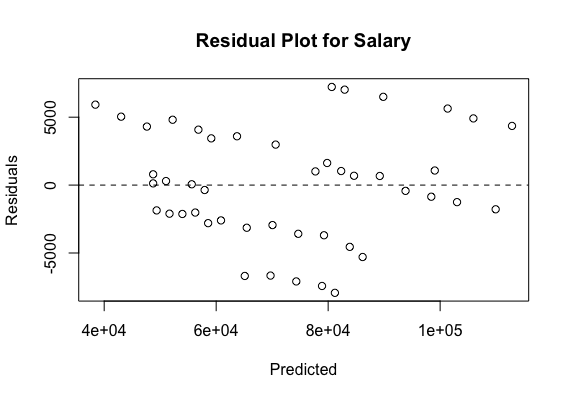
F-statistic: 226.8 on 4 and 41 DF, p-value: < 2.2e-16

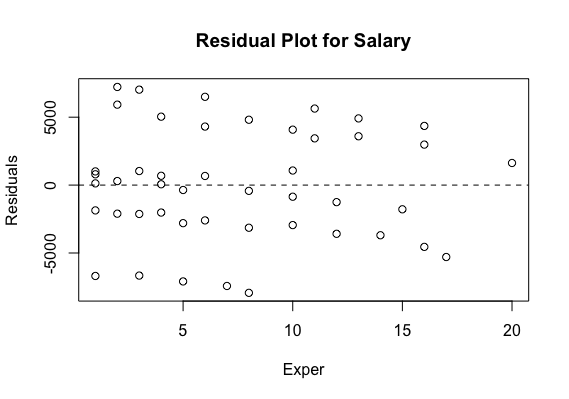
The regression model is: y = 2299.5xexper + 13236.3xeduc\_dum\_1 + 12614.7xeduc\_dum\_2 – 28980xmgt\_dum + 62809.4

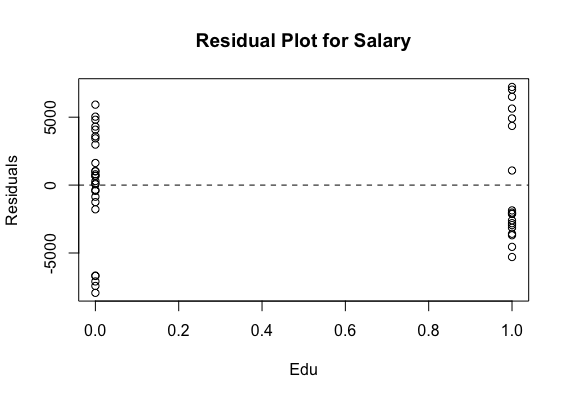
3.

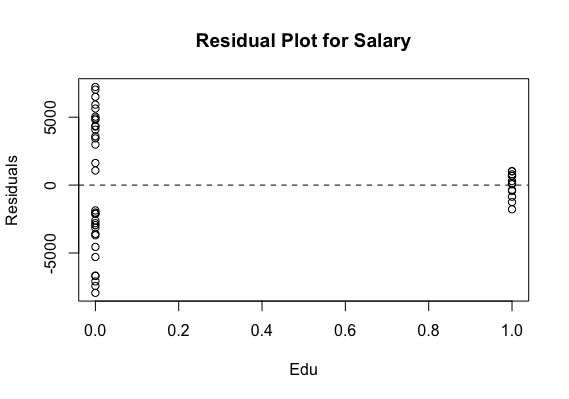


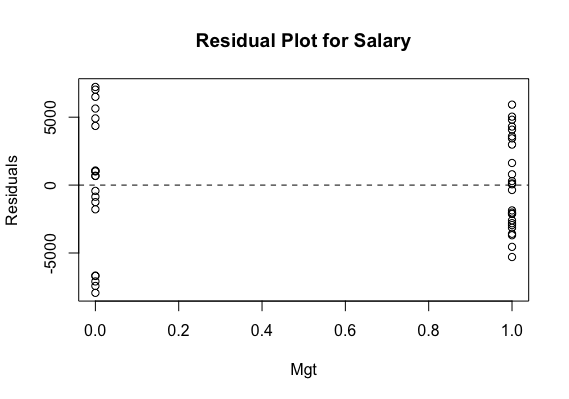
4.

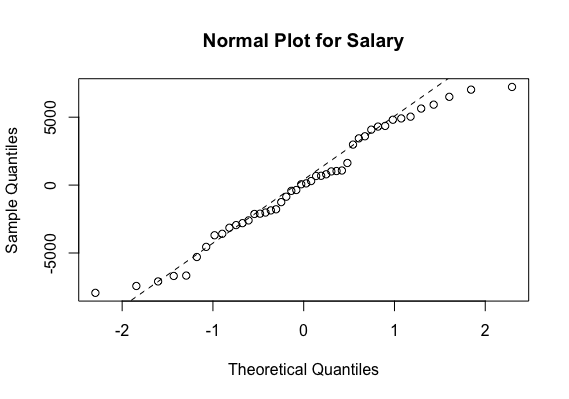












5.

There will be a 2299.5 dollar increasing in salary.

6.

There will be a 13236.3 dollar increasing in salary.

7.

y = 2299.5 \* 3 – 28980 + 62809.4 = 40727.9

The predicted salary is 40727.9 dollar.

8.

No, it is not, since there are also another 2 elements influences the salary, experience and management responsibilities. But if the other two are at same level, salary of advanced degree is still lower than the one of college degree.

9.

fit lwr upr

1 40727.88 31538.8 49916.96

So the prediction interval is [31538.8, 49916.96]