Assignment 1

Question 1: "wage1"

1. Average education level - 12.56.
   Lowest year of education - 0.00.
   Highest year of education -18.00.
2. Average hourly wage - $5.909, average hourly wage seems to be low.
3. As per the Economic Report of the President report: **cpi_1976 = 56.9, cpi_2010 = 218.1**
4. Average hourly wage in 2010 = $22.6494. Yes, the new average hourly wage rate seems reasonable.
5. Women: 252. Men :274

Question 2: "meap01"

1. Largest value of math4 = 100 and Smallest values of math4 = 0.
   Range [0,100]. The Range makes sense.
2. Schools have a perfect pass rate on the math test = 38, which is 2.08%.
3. Schools have math pass rates of exactly 50% = 17.
4. Average math_prate =71.909%, Average read_prate= 60.06%.
   Read is harder to pass.
5. Correlation = 0.8427 or 84.27%.
   Math and Read are highly correlated.
6. Expenditure per pupil = 5194.865, Standard deviation = 1091.89.
   Yes, there is wide variation in per pupil spending.
7. School A's spending exceeds School B's by 9.09%, and
   their logarithmic difference: 8.701%.

Question 3: "401k"

1. Average Participation rate: 87.36%, Average match rate: 0.7315 %.
2. prate = 83.0755 + 5.8611 mrate
   df=1532, Sample = 1534, R-squared = 0.0747 or 7.47%
3. When mrate =0, intercept = 83.0755. The predicted participation rate is 83.07%
   For every 1 percent increase in mrate, the participation rate increases by 5.86%.
4. When mrate= 3.5, the Participation rate is 103.58%, which make no sense. Because Participation rate cannot be more than 100%.
5. R-Squared = 7.47%, which is very low.
   The variation in mrate does not explain the variation in prate.

Question 4: "ceosal2"

1. Average CEO salary: $ 865.56K
   Average CEO tenure: 7.95 years.
2. CEOs in first year = 5,
   Longest tenure as CEO =37 years.
3. ln[salary] = 6.505498 + 0.0097 ceoten
   R-Squared =0.01316, Sample= 177, df=175
   For one-year increase as CEO, the predicted percentage increase in salary is 0.97%.

Question 5: "wage2"

1. Average salary = $ 957.94545, Average IQ= 101.28235, IQ Standard deviation = 15.05264
2. wage = 116.992+8.303 IQ.
   Sample = 935, df = 933, R-Squared = 0.0955 or 9.55%
   15 points increase in IQ, the wage will be increased by 125.53.
   But R-square is 9.5%, the variation in IQ does not explain the variation in Wage.
3. log(wage) = 5.886 + .0088IQ
   Sample = 935, df = 933, R-Squared = 0.09909 or 9.909%
   15 points increase in IQ, the wage will be increase by 13.21 %

Question 6: "meap93"

1. Diminishing effect seem more appropriate.
   Expend has very less effect on the math pass rate.
2. In the given model: math10 = b0 + b1 ln[expend] + u, the expend coefficient is in log form that gives percentage, and math10 is also in percentage format. So, 10% increase in expend means a 1/10 increase in log value.
   b1/10 implies one percent increase in math10.
3. math10 = -69.341 + 11.164*ln[expend]
   N=408; df= 406; R-square = 2.966%
4. 10% increase in expend -> 1.064% increase in math rate.
5. Range of predicted math10 [21.22,30.15]. The maximum predicted value is 30.15.
   Let's assume math10 value be 101.
   The equation becomes: 101 =-69.34+11.164*ln(expend)
   ln(expend)= (101+69.34)/ (11.164)
   expend is approximately greater than $4.2m (expend > $4.2m)

Question 7: "hprice1"

1. price = -19.315 + 0.1284sqrft + 15.198bdrms
   Sample =88, df =85, R-Squared = 0.6315 or 63.15%.
2. Estimated increase in House= $15.198k or $15198.
3. Estimated increase in House= $33.174k or $33174.
4. R-squared is 63.19% (variation is percentage is explained).
5. Predicted selling price will be $ 354K.
6. Residual = -54k, Underpaid.

Question 8: "ceosal2"

1. ln(salary)=4.6209+ 0.1621 sales+ 0.1067 mktval
   R-squared: 0.2991, Sample =177, df =174.
2. Profit values range from -463 to 2700, some are negatives. We can't use log(profits). Therefore, log (Profit) attribute can't be added to this model. The R-squared is 0.2993, approximately 70% of the variation in log(salary) is unexplained. Profits seems to add very little to the model, suggesting that profits have very mild influence on log(salary).
3. One-year increase in tenure is associated with a 1.167% increase in a CEO's salary.
4. Correlation = 0.776, Market value and Profits are highly correlated.
   Both needs to be included.

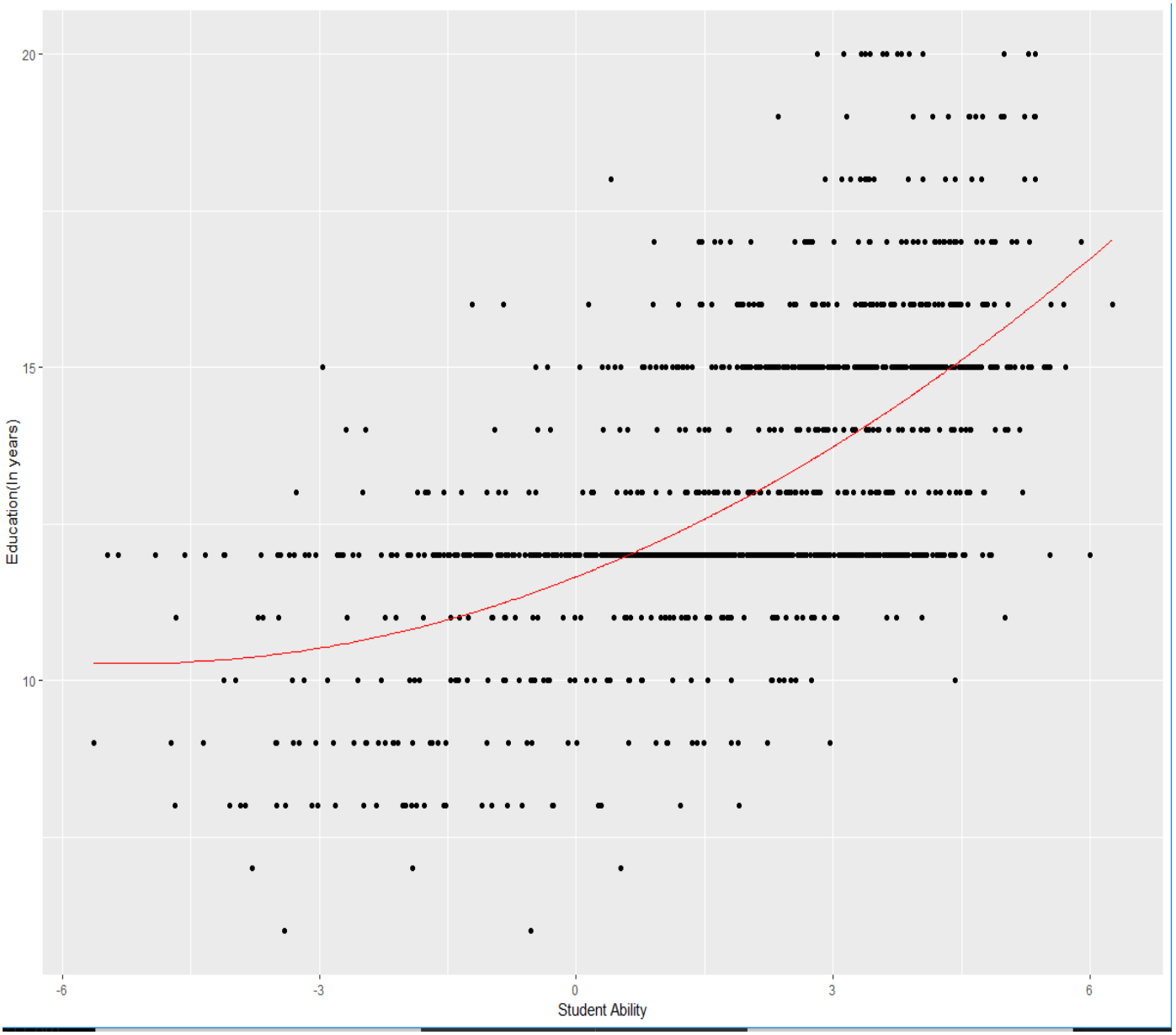
Question 9: "attend"

1. atndrte: min :6.25%, max: 100%, average: 81.71%
   PriGPA : min :0.857, max: 3.930, average: 2.587
   ACT : min :13.00, max: 32.00, average: 22.51
2. antdrte = 75.70+17.261 priGPA - 1.717 ACT
   Sample size = 680, df=677 R-squared = 0.2906
   The predicted attend rate is 75.70% for a student who has Zero ACT score and Zero PriGPA. No, it does not have any useful meaning.
3. # For every point increase in "priGPA", the attendance rate is predicted to increase by 17.26% which makes sense.
   For every point increase in "ACT", the attendance rate is predicted to decrease by 1.717% which does not make any sense.
4. When priGPA=3.65 and ACT=20, the attend rate is over 100 %, which is impossible and makes no sense.
   There is one student in these criteria.
   Details: index number is 568, the original "attendance" is "87.5%".
5. The predicted difference between the attendance between student A and student B is 25.846%

Question 10: "htv"

1. Range [6,20].
   41.62% has completed 12th but no higher degree.
   Mean education levels of Men, Father, and Mother are 13.03740, 12.44715, 12.17805 respectively. So, men have higher education level than their parents.
2. Educ = 6.96 + 0.3042 motheduc + 0.1902 fatheduc.
   Sample = 1230, df =1227, R-Squared =0.2493. So, 24.93% of sample vaariation in educ is explained by the parents' education.
   For one-year increase in mother's education, 0.3042 years increase in son's year in education.
3. Educ = 8.4486 + 0.1891 motheduc + 0.111 fatheduc + 0.5024 abill ;
   Sample = 1230, df =1226, R-square = 0.4275 or 42.75%.
   For one-year increase in ability, 0.5024 years increase in student years in education.
4. educ = 8.2302 + 0.1901 motheduc + 0.1089 fatheduc + 0.4014 abil + 0.0505 abil^2
   Sample = 1230, df =1225, R-square = 0.4444or 44.44%.
   #1st derivative = 0, d(Educ)/d(abil)= 0.4014+ 2*0.0505*abil = 0
   abil = -3.974.
   2nd derivative, d2(Educ)/d2(abil)=0.10.
   The second derivative is positive, so there is a minimum.
5. Out of 1230 only 15 students are predicted to have ability less than -3.974.
   1.21% is very low and we don't have enough data for low ability students.
6. When motheduc=12.18 and fatheduc = 12.45,
   educ = 8.2302 + 0.1901*12.18 + 0.1089*12.45 + 0.4014 abil + 0.0505 abil^2
   educ=11.901+ 0.4014 abil + 0.0505 abil^2

   (*Graph next page)


20
15
10
Education(In years)
-6
-3
0
3
6
Student Ability