Problem Set 1 Solutions

Buan/Mis 6356

Fall 2018

Question 1

i. Avg 12.56 years. Range [0,18]
ii. Avg $5.90, seems low
iii. 56.9 in 1976 and 218.1 in 2010
iv. Avg $22.64, seems reasonable
v. 252 women, 274 men

```
summary(wage1$educ)
mean(wage1$wage)
mean(wage1$wage)/56.9*218.056
table(wage1$female)
```

Question 2

i. Range [0,100]. Makes sense
ii. 38 schools or 2% of sample
iii. 17 schools
iv. Math 71.9% Reading 60.1%; reading is harder
v. Corr 84%. This is quite high
vi. Avg $5194 SD $1092. The variance is moderate
vii. Real: 9.1% Log approx.: 8.4%.

```
summary(meap01$math4)
c(sum(meap01$math4==100),mean(meap01$math4==100))
sum(meap01$math4==50)
c(mean(meap01$math4),mean(meap01$read4))
cor(meap01$math4,meap01$read4)
c(mean(meap01$exppp),sd(meap01$exppp))
c((6000-5500)/5500,log(6000)-log(5500))
```

Question 3

i. Avg prate 87% avg mrate: 0.73%
ii. N 1534 Rsq 0.075; prate = 83.0755 + 5.8611*mrate + e
iii. When mrate is zero, the average participation rate is 83.0755%. For every 1% increase in the match rate, the prate increases by 5.86%
iv. Pred prate 103% which is impossible. The best fit line just doesn't make sense this far from the center of the data.
v. Rsq 7.5% which is basically zero. Not a good fit.

```
c(mean(four01k$prate),mean(four01k$mrate))
model <- lm(prate~mrate,data=four01k)
model %>% summary
model %>% predict(data.frame(mrate=3.5))
```

Question 4

i. Avg salary $865k Avg tenure 7.95 years
ii. 5 ceos in 1st year. 37 years is longest
iii. N 177 Rsq 0.013; ln(salary) = 6.51 + 0.009724 ceoten + e; Every 1 year increase in tenure relates to a 0.9% increase in salary.

```
c(mean(ceosal2$salary),mean(ceosal2$ceoten))
c(sum(ceosal2$ceoten==0),max(ceosal2$ceoten))
lm(log(salary)~ceoten,data=ceosal2) %>% summary
```

Question 5

i. Avg salary $957 per month Avg IQ 101 SD 15.1
ii. 15 IQ => $124.5 increase in salary but the r squared is 9.5%
iii. 15 IQ => 13% increase in salary

```
c(mean(wage2$wage),mean(wage2$IQ),sd(wage2$IQ))
lm(wage~IQ,data=wage2) %>% summary
lm(log(wage)~IQ,data=wage2) %>% summary
```

Question 6

i. Diminishing return makes sense
ii. A 1 unit change in log(expend) has a b1 effect on the math pass rate and 1 unit in log(expend) = 100% change in expend so a 10% change in expend will have a b1/10 effect.
iii. N=408 Rsq 2.966%; math10 = -69.341 + 11.164 ln(expend)
iv. Expend inc by 10% => Math10 pass rate increases by 1.164%
v. Pred math10 values are in [21.22,30.15]
169.341/11.164=ln(expend) or expend > $3.8M

```
model <- lm(math10~log(expend),data=meap93)
model %>% summary
model %>% predict %>% summary
```

Question 7

i. Price = -19.315 + 0.12844 sqrft + 15.198 bdrms
ii. $15000 = 15.198*1
iii. $32000 = 0.12844*140 + 15.198
iv. 63% is explained
v. $354K predicted price
vi. Residual -$54; underpaid

```
model <- lm(price~sqrft+bdrms,data=hprice1)
model %>% summary
predict(model)[1]
residuals(model)[1]
```

Question 8

i. Ln(salary) = 4.6 + 0.162*ln(sales) + 0.107*ln(mktval)
ii. You can't take the log of a negative number. Only 29.9% of the variation is explained
iii. 1 year of tenure => 1.17% increase is salary all else equal
iv. Corr 0.78 which is very high, both of these need to be included to avoid omitted variable bias

```
lm(log(salary)~log(sales)+log(mktval),data=ceosal2) %>% summary
lm(log(salary)~log(sales)+log(mktval)+profits,data=ceosal2) %>% summary
lm(log(salary)~log(sales)+log(mktval)+profits+ceoten,data=ceosal2) %>%
summary
cor(log(ceosal2$mktval),ceosal2$profits)
```

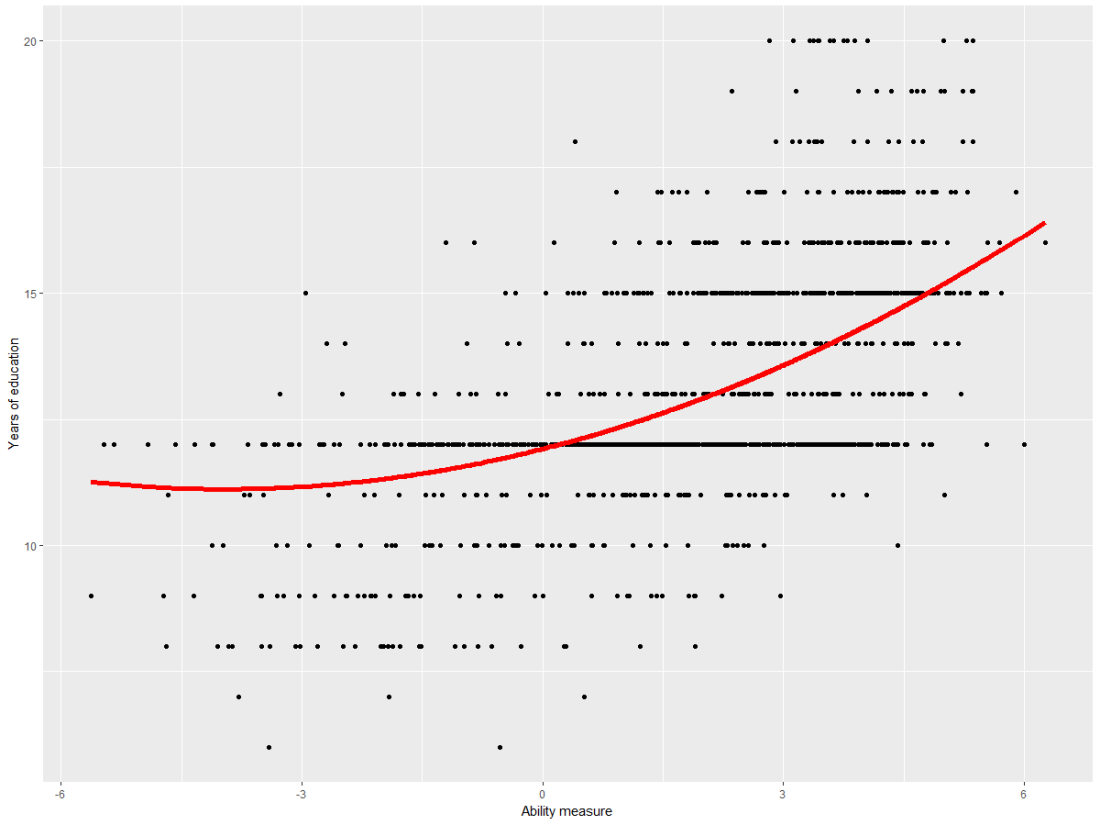
Question 9

i. Attend rate: min 6.25% max 100% avg 81.7%; priGPA: min 0.86 max 3.93 avg 2.59; ACT: min 13 max 32 avg 22.51
ii. Attendrate = 75.7 + 17.2 priGPA – 1.72 ACT. If GPA and ACT were 0 (which is impossible), attendance would be at 75.7%
iii. For every 1 point increase in ACT score, attendance goes down by 1.72% which is surprising. For every 0.1 point increase in GPA, attendance goes up by 1.72% which is not surprising.
iv. Predicted attendance is 104% which makes no sense. There is one student with those values
v. 25.84% is the predicted difference

```
attend %>% select(atndrte,priGPA,ACT) %>% summary
model <- lm(atndrte~priGPA+ACT,data=attend)
model %>% summary
model %>% predict(data.frame(priGPA=3.65,ACT=20))
attend[priGPA==3.65 & ACT==20] %>% nrow
model %>% predict(data.frame(priGPA=3.1,ACT=21))-model %>%
predict(data.frame(priGPA=2.1,ACT=26))
```

Question 10

i. Educ range [6,20] 42% completed 12th but no higher; Men have more education than their parents on average
ii. Educ = 6.96 + 0.304 motheduc + 0.190 fatheduc; 24.9% of variation is explained, For every 1 year increase in mother's education, their sons get 0.304 more years of education
iii. Educ = 8.44 + 0.189 motheduc + 0.111 fatheduc + 0.502 abil; Rsquared 42.75% which is much higher
iv. Set the derivative (with respect to ability) equal to zero and solve for ability. 0.401+2*0.051*abil=0 or abil = -3.931373
v. Only 1.2% of the data has ability lower than -3.93. Probably ability doesn't curve back up, we just don't have good data on low ability persons
vi.




```
htv %>% select(educ,motheduc,fatheduc) %>% summary
mean(htv$educ==12)
lm(educ~motheduc+fatheduc,data=htv) %>% summary
lm(educ~motheduc+fatheduc+abil,data=htv) %>% summary
lm(educ~motheduc+fatheduc+abil+I(abil^2),data=htv) %>% summary
betas <- tidy(lm(educ~motheduc+fatheduc+abil+I(abil^2),data=htv))
-betas[4,2]/2/betas[5,2]
mean(htv$abil<=-betas[4,2]/2/betas[5,2])
```