MSIN0096 Second Assignment

Due 10 am, 24 Nov for SORA students

23/11/2021

Question 1

This question is about comparing means from paired data, 1-sided test.

Step 1: Define 2 competing hypotheses. Suppose mu0 - mothers average sleeping hours per day before having a baby. mu1 - mothers average sleeping hours per day after having a baby. Let mud = mu1 - mu0 H0: mu1 = mu0 <=> H0: mud = 0 H0: mud >= 0, H1: mud < 0 (hence, our alternative hypothesis H1 = mothers' sleeping hours have been significantly reduced after having a baby; while our null hypothesis H0 = mothers' sleeping hours have stayed the same or significantly increased after having a baby.)

Step 2: Find the testing statistic and its distribution. From the data question, x_bard = 2.5 and sample standard deviation sd = 4, so T = (x_bard-mud)/sd/sqrt(n)

```
mud <- 0
sd <- 4
n <- 25
x_bar <- 2.5
t <- (x_bar - mud)/(sd/sqrt(n))
t
```

```
## [1] 3.125
```

Without proceeding to the next stage, we can tell that 3.125 is already slightly bigger than 2, so we're likely to reject H0.

Step 3: Find the critical value C at 5% significance level. At the significance level a = 5%, and df = n-1 = 25-1 = 24, we find t0.95,24 = 1.710882.

```
qt(0.95, 24)
```

```
## [1] 1.710882
```

Step 4: Make the decision. Since $|T| = 3.125 > 1.71$, we can reject H0 at 5% significance level, or 95% confidence level.

Step 5: p-value calculation Below are 2 ways in whcih p-value can be calculated in this case. It is equal to 0.002301319.

```
pt(q=t, df=24, lower.tail=FALSE)
```

```
## [1] 0.002301319
```

```
pt(-abs(t),df=24)
```

```
## [1] 0.002301319
```

Question 2

This question is about testing a proportion.

Step 1: Define 2 competing hypotheses. Let p0 be the rate of severe symptoms caused by the flu season (in general population?), equal to 12%:

```
p0 <- 0.12
```

Let p_hat be the rate of severe symptoms caused by the flu season among undergraduate students:

```
p_hat <- 25/526
```

Hypothesis: H0: p>=p0, H1: p<p0, where p0 = 0.12.

Step 2: Find the testing statistic and its distribution. Testing statistic: T = (p_hat - p0)/(sqrt(p0 x (1-p0)/n)) = (25/526 - 0.12)/(sqrt(0.12 x (1-0.12)/526)), t(df = 525)

```
n <- 526
T <- (p_hat - p0)/(sqrt(p0*(1-p0)/n))
T
```

```
## [1] -5.114793
```

Step 3: Find the critical value C at 0.01% significance level. Critical value at 99.99% confidence level is t0.9999 = -3.745448 (calculated below)

```
qt(1-0.9999, 525)
```

```
## [1] -3.745448
```

|T| > |t0.9999| as 5.114793 > 3.745448, hence we reject H0. Therefore, we can conclude that the drug is a success at 99.99% confidence interval. However, it is impossible to infer that the new anti-flu drug is effective for the general population, as the sample is restricted only to undergraduate students.

Question 3

(a) This question is about assumptions of linear regression model. The equation makes it possible to show how an age group (independent/explanatory variable) affects a game app's revenue (dependent variable). The equation does not violate assumption 1, as it specifies a linear relationship between revenue and age, where each coefficient of B stands by itself, i.e. all components are linear. It seems to be no perfect collinearity, as none of the independent variables is redundant, hence, assumption 2 is not violated.Since the equation explores different user group's contribution, it is likely that we have a random sampling and the sample is from relatively homogeneous group (those interested in the game), satisfying assumptions 3 and 5. However, it is unclear whether the error term E is uncorrelated with independent variables, as there is no dataset given, hence we cannot be sure that assumption 4 holds. Therefore, the model is possible to estimate but is may be not the best linear unbiased estimator, according to Gauss-Markov theorem. The model only takes gamers' age as an independent variable, omitting others, like income, whether the account is shared or not (e.g. an 11-year-old could share their parent's account), years spent on the gaming platform, the achieved level at the game.

(b) I disagree with student A. *b1* measures how much the revenue will change as the share of teenage users increases by 1 unit. Hence, when *x1* (number of teenage users) increases by 1 person, then predicted Y (revenue) goes up by *b1* value.

(c) AGE1 = age1; AGE2 = age1 + age 2; AGE3 = age1 + age2 +age3. rev = 0.87 + 1.20AGE1 + 1.08AGE2 + 0.67AGE3 = 0.87 + 1.20age1 + 1.08(age1+age2) + 0.67(age1+age2+age3) = 0.87 + 1.20age1 + 1.08age1 + 1.08age2 + 0.67age1 + 0.67age2 + 0.67age3 = 0.87 + 2.95age1 + 1.75age2 + 0.67age3.

Question 4

This question is about quadratic functional form, capturing non-monotonic impact of house age on house price. According to our estimation, after living in the property for a year (i.e., when age increases by 1 unit), modern house price bought by "fam1" will depreciate by 11.79332 units(e.g., £000) given everything unchanged. For another family "fam2" who bought a 50-year old property, the house price will drop only by 7.201805 in a year.

Comparing those 2 values, we can conclude that house value eventually increases over time, albeit depreciating first. Since b1 = I(age^2) = 0.0459152 > 0, the parabola's ends are upwards. It means, that the house price will be decreasing until turning point, and will be increasing again after the the curve's bottom point.

```
price1_fam1 <- 1441.263
price2_fam1 <- 1441.263 + 0.0459152*(1)**2 - 11.83924*1
depreciation_fam1 <- price2_fam1 - price1_fam1
depreciation_fam1
```

```
## [1] -11.79332
```

```
price1_fam2 <- 1441.263 + 0.0459152*(50)**2 - 11.83924*50
price2_fam2 <- 1441.263 + 0.0459152*(51)**2 - 11.83924*51
depreciation_fam2 <- price2_fam2 - price1_fam2
depreciation_fam2
```

```
## [1] -7.201805
```

```
# For fun as it makes no sense, age's units may be in decades, rather than years. Please ignore it if i
turning_point_prep <- -(-11.83924/2*0.0459152)
turning_point_in_days <- 365/(1/turning_point_prep)
turning_point_in_days
```

```
## [1] 99.2072
```

```
turning_point_in_months <- turning_point_in_days/30
turning_point_in_months
```

```
## [1] 3.306907
```

Question 5

a) *b_wave_1* = 95.14

```
house <- read.csv(file="./houseprice.csv", header=T,sep=",")
q5a <- lm(price ~ rooms, data=house)
summary(q5a)
```

```
##
## Call:
## lm(formula = price ~ rooms, data = house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -153.15  -51.47  -13.01   61.99  226.57
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -253.81     102.35  -2.480     0.02 *
## rooms          95.14      15.11   6.297 1.15e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 90.79 on 26 degrees of freedom
## Multiple R-squared:  0.604,  Adjusted R-squared:  0.5887
## F-statistic: 39.65 on 1 and 26 DF,  p-value: 1.15e-06
```

(b) *b_hat_1* = 1.03633, *b_hat_2* = 0.23633.

```
q5b <- lm(price ~ rooms+area, data=house)
summary(q5b)
```

```
##
## Call:
## lm(formula = price ~ rooms + area, data = house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -89.345 -45.493  -0.957  43.489 102.790
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.39711   77.10707   0.226    0.823
## rooms        1.03633   17.53363   0.059    0.953
## area         0.23633    0.03701   6.385  1.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 57.08 on 25 degrees of freedom
## Multiple R-squared:  0.8495, Adjusted R-squared:  0.8374
## F-statistic: 70.54 on 2 and 25 DF,  p-value: 5.255e-11
```

(c) $b\_wave\_1 = 95.14$; $b\_hat\_1 = 1.0363$
(d) $Y = 398.18$

```
q5d <- lm(area ~ rooms, data=house)
summary(q5d)
```

```
##
## Call:
## lm(formula = area ~ rooms, data = house)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -641.66 -165.46  -53.99  197.40  585.80
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1147.59     341.00  -3.365  0.00238 **
## rooms         398.18      50.34   7.910 2.18e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 302.5 on 26 degrees of freedom
## Multiple R-squared:  0.7065, Adjusted R-squared:  0.6952
## F-statistic: 62.57 on 1 and 26 DF,  p-value: 2.183e-08
```

```
Y <- 398.18
b_wave_1 <- 95.14
b_hat_1 <- 1.0363
b_hat_2 <- 0.23633
b_wave_1_check = b_hat_1 + Y*b_hat_2
b_wave_1_check
```

```
## [1] 95.13818
```

```
b_wave_1
```

```
## [1] 95.14
```

```
round(b_wave_1_check, digits = 2) == b_wave_1
```

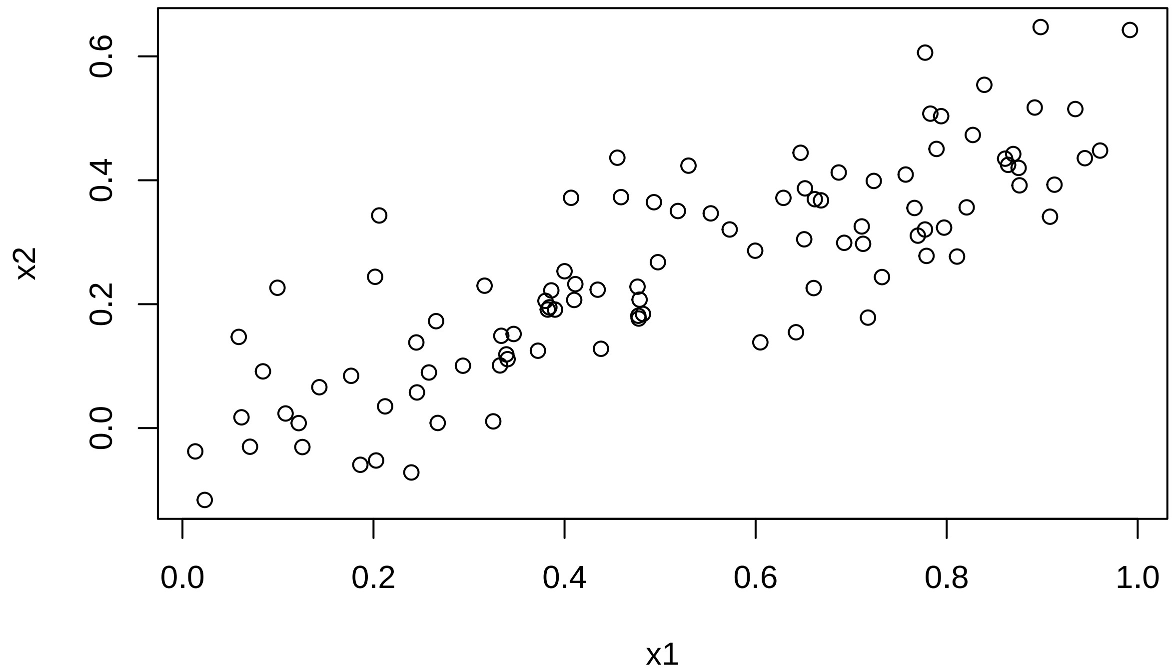
```
## [1] TRUE
```

Question 6

```
 set.seed(1)
  x1=runif(100)
  x2=0.5*x1+rnorm(100)/10
  y=2+2*x1+0.3*x2+rnorm(100)
```

The form of the linear model: **Y = 2 + 2X1 +0.3X2 + rnorm** or **Y = b0 + b1xX1 + b2*X2 + E**. The regression coefficients are: *b0*=2, *b1*=0.5, *b2*=0.3. The correlation is 0.84 (2 d.p.), which means that there is a strong positive correlation between *x1* and *x2*, which is also observable on the plot below.

```
plot(x1, x2)
```




```
cor(x1, x2)
```

```
## [1] 0.8351212
```

(b) Estimated *b_hat_0*=2.1305, original *b0*=2; Estimated *b_hat_1*=1.4396, original *b1*=0.5; Estimated *b_hat_2*=1.0097, original *b2*=0.3. Hence, we can reject both null hypotheses H0: *b1*=0 and H0: *b2*=0.

```
set.seed(1)
 x1=runif(100)
 x2=0.5*x1+rnorm(100)/10
 y=2+2*x1+0.3*x2+rnorm(100)

 q6b <- lm(y ~ x1+x2)
 summary(q6b)
```

```
##
## Call:
```

```
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8311 -0.7273 -0.0537  0.6338  2.3359
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
## x1            1.4396     0.7212   1.996   0.0487 *
## x2            1.0097     1.1337   0.891   0.3754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.056 on 97 degrees of freedom
## Multiple R-squared:  0.2088, Adjusted R-squared:  0.1925
## F-statistic:  12.8 on 2 and 97 DF,  p-value: 1.164e-05
```

(c) *b_hat_1* = 1.9759 Hence, we can reject the null hypothesis H0: *b1*=0.

```
set.seed(1)
 x1=runif(100)
 x2=0.5*x1+rnorm(100)/10
 y=2+2*x1+0.3*x2+rnorm(100)

 q6c <- lm(y ~ x1)
 summary(q6c)
```

```
##
## Call:
## lm(formula = y ~ x1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.89495 -0.66874 -0.07785  0.59221  2.45560
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1124     0.2307   9.155 8.27e-15 ***
## x1            1.9759     0.3963   4.986 2.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.055 on 98 degrees of freedom
## Multiple R-squared:  0.2024, Adjusted R-squared:  0.1942
## F-statistic: 24.86 on 1 and 98 DF,  p-value: 2.661e-06
```

(d) *b_hat_2* = 2.8996 Hence, we can reject the null hypothesis H0: *b2*=0.

```
set.seed(1)
 x1=runif(100)
 x2=0.5*x1+rnorm(100)/10
```

```r
y=2+2*x1+0.3*x2+rnorm(100)

q6d <- lm(y ~ x2)
summary(q6d)
```

```
##
## Call:
## lm(formula = y ~ x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62687 -0.75156 -0.03598  0.72383  2.44890
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.3899     0.1949   12.26  < 2e-16 ***
## x2            2.8996     0.6330    4.58 1.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.072 on 98 degrees of freedom
## Multiple R-squared:  0.1763, Adjusted R-squared:  0.1679
## F-statistic: 20.98 on 1 and 98 DF,  p-value: 1.366e-05
```

(e) Results in hypothesis testings obtained in (b)-(d) do not contradict each other.

(f) Estimated *b_hat_0*=2.02039, original *b0*=2; Estimated *b_hat_1*=1.99053, original *b1*=0.5; Estimated *b_hat_2*=0.31715, original *b2*=0.3. Hence, we can reject both null hypotheses H0: *b1*=0 and H0: *b2*=0.

Model in (b) is more accurate than model (f), as coefficients are way closer to the original (true) ones.

```r
set.seed(1)
 x1=runif(10000)
 x2=0.5*x1+rnorm(10000)/10
 y=2+2*x1+0.3*x2+rnorm(10000)

 q6f <- lm(y ~ x1+x2)
 summary(q6f)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7029 -0.6707  0.0055  0.6539  3.6299
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.02039    0.01984 101.849  < 2e-16 ***
## x1           1.99053    0.06104  32.609  < 2e-16 ***
## x2           0.31715    0.10088   3.144  0.00167 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9976 on 9997 degrees of freedom
## Multiple R-squared:  0.2828, Adjusted R-squared:  0.2826
## F-statistic:  1971 on 2 and 9997 DF,  p-value: < 2.2e-16
```

Question 7

(a) Keeping everything else unchanged, if the area's ratio of the minority ethnicity population increases by 1, fast-food restaurants charge 0.06493units (e.g.,0.0£) higher price on soda. Based on the regression result, we can conclude that there is price discrimination against minorities.

```r
soda <- read.csv(file="./sodaprice.csv", header=T,sep=",")
q7a <- lm(psoda ~ prpminor, data=soda)
summary(q7a)
```

```
##
## Call:
## lm(formula = psoda ~ prpminor, data = soda)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30884 -0.05963  0.01135  0.03206  0.44840
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.03740    0.00519  199.87  < 2e-16 ***
## prpminor     0.06493    0.02396    2.71  0.00702 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0881 on 399 degrees of freedom
## Multiple R-squared:  0.01808,    Adjusted R-squared:  0.01561
## F-statistic: 7.345 on 1 and 399 DF,  p-value: 0.007015
```

(b) The discrimination effect becomes smaller when income is controlled: the average price of soda increases by 1.603e-06 while the minority ethnicity population increases by 1 (estimated coefficient is > 0).

```r
q7b <- lm(psoda ~ prpminor+income, data=soda)
summary(q7b)
```

```
##
## Call:
## lm(formula = psoda ~ prpminor + income, data = soda)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29401 -0.05242  0.00333  0.04231  0.44322
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.563e-01  1.899e-02  50.354  < 2e-16 ***
## prpminor    1.150e-01  2.600e-02   4.423 1.26e-05 ***
## income      1.603e-06  3.618e-07   4.430 1.22e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08611 on 398 degrees of freedom
## Multiple R-squared:  0.06422,    Adjusted R-squared:  0.05952
## F-statistic: 13.66 on 2 and 398 DF,  p-value: 1.835e-06
```

(c) We do not want to take logarithm on *prpminor*, as it is a ratio, and taking its logarithm may skew the data.

```r
q7c <- lm(psoda ~ prpminor+log(income) + prppov + log(house), data=soda)
summary(q7c)
```

```
##
## Call:
## lm(formula = psoda ~ prpminor + log(income) + prppov + log(house),
##     data = soda)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26460 -0.04699  0.00387  0.04151  0.43924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.13247    0.30682   0.432  0.66616
## prpminor     0.10066    0.03070   3.279  0.00113 **
## log(income) -0.05509    0.03937  -1.399  0.16253
## prppov       0.05606    0.14112   0.397  0.69140
## log(house)   0.12574    0.01855   6.777 4.46e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08081 on 396 degrees of freedom
## Multiple R-squared:  0.1801, Adjusted R-squared:  0.1718
## F-statistic: 21.74 on 4 and 396 DF,  p-value: 3.105e-16
```

```r
p_change_psoda <- 0.2*0.10066*100
cat("If *prpminor* increases by 0.2, the estimated percentage change in psoda is ", p_change_psoda)
```

```
## If *prpminor* increases by 0.2, the estimated percentage change in psoda is  2.0132
```

(d) f=3.504284, qf=3.01851. Because $f>qf$, we reject the null hypothesis at 5% level: *ln(income)* and *prppov* are jointly significant, i.e. logarithm on median family income and proportion of population in poverty have joint impact on average soda price per unit. The restricted model has weaker explanatory power than the unrestricted model, as SSRr > SSRur, and we have a larger F statistic.

```
#Step 1: fit restricted and unrestricted models
q7d_ur <- lm(psoda ~ prpminor+log(income)+prppov+log(house), data=soda)
q7d_r <- lm(psoda ~ prpminor + log(house), data=soda)
#Step 2: compute SSR
ssr_ur <- sum(q7d_ur$residuals**2)
cat("SSR of the unrestricted model is ", ssr_ur)
```

```
## SSR of the unrestricted model is  2.586079
```

```
cat("\n")
```

```
ssr_r <- sum(q7d_r$residuals**2)
cat("SSR of the restricted model is ", ssr_r)
```

```
## SSR of the restricted model is  2.631849
```

```
cat("\n")
```

```
#Step 3: compute F statistics
f <- ((ssr_r-ssr_ur)/2)/(ssr_ur/(401-4-1))
cat("F statistic is ", f)
```

```
## F statistic is  3.504284
```

```
cat("\n")
```

```
#Step 4: find critical value
cat("Critical value is ", qf(0.95, 2, 396))
```

```
## Critical value is  3.01851
```

Question 8

(a) Keeping everything else unchanged, when education expenditure increases by 1% per pupil in the district, the math scores increase by 0.35 units (as expenditure is taken in a logarithm form). The result is very statistically significant,as p-value is <2e-16. Hence, education expenditure has a significant impact on math scores.

```
school <- read.csv(file="./schoolscores.csv", header=T,sep=",")
#Step 1: estimate OLS regression ignoring the panel data structure.
q8a <- lm(math4 ~ lrexpp + lenrol+ lunch, data=school)
summary(q8a)
```

```
##
## Call:
## lm(formula = math4 ~ lrexpp + lenrol + lunch, data = school)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -70.722 -11.270   0.129  11.171  59.367
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -231.3499    10.6621 -21.698   <2e-16 ***
## lrexpp        35.1079     1.2678  27.692   <2e-16 ***
## lenrol        -0.6185     0.2488  -2.485    0.013 *
## lunch         -0.3761     0.0169 -22.263   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.83 on 3846 degrees of freedom
## Multiple R-squared:  0.244,  Adjusted R-squared:  0.2434
## F-statistic: 413.8 on 3 and 3846 DF,  p-value: < 2.2e-16
```

(b) Coefficients on *lxepp* and *lunch* in part (b) are smaller than in part (a), with an exclusion of *lenrol*.

```r
#Step 2: Introduce panel data regression without all coefficients listed, with year fixed effects.
library(plm)
q8b <- plm(math4 ~ lrexpp + lenrol+ lunch, data=school, model="within", index=c("year"))
summary(q8b)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = math4 ~ lrexpp + lenrol + lunch, data = school,
##     model = "within", index = c("year"))
##
## Balanced Panel: n = 7, T = 550, N = 3850
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -58.102528  -7.202971  -0.023069   7.502756  76.146570
##
## Coefficients:
##         Estimate Std. Error  t-value  Pr(>|t|)
## lrexpp  8.420889   1.103118   7.6337  2.86e-14 ***
## lenrol  0.476389   0.189754   2.5106    0.0121 *
## lunch  -0.414113   0.012817 -32.3103 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    724250
## Residual Sum of Squares: 551720
## R-Squared:      0.23823
## Adj. R-Squared: 0.23644
## F-statistic: 400.295 on 3 and 3840 DF, p-value: < 2.22e-16
```

(c) Part (b) is more reliable in evaluating the impact of education expenditure on math scores, as it has a smaller std deviation than part (c). In part (c), the new coefficient 68.145429 suggests that when education expenditure increases by 1%, math scores increase by 0.68 units, i.e. by 0.68% of fourth graders who pass a standardized math test.

```r
#Step 3: Introduce panel data regression without all coefficients listed, with year fixed effects and d
library(plm)
q8c <- plm(math4 ~ lrexpp + lenrol+ lunch, data=school, model="within", index=c("distid", "year"))
summary(q8c)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = math4 ~ lrexpp + lenrol + lunch, data = school,
##     model = "within", index = c("distid", "year"))
##
## Balanced Panel: n = 550, T = 7, N = 3850
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -90.201073  -7.294675  -0.021144   6.732179  84.108354
##
## Coefficients:
##          Estimate Std. Error t-value  Pr(>|t|)
## lrexpp  68.145429   1.624155 41.9575 < 2.2e-16 ***
## lenrol -10.007889   1.157228 -8.6482 < 2.2e-16 ***
## lunch    0.525320   0.058285  9.0130 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    814510
## Residual Sum of Squares: 482400
## R-Squared:      0.40774
## Adj. R-Squared: 0.30858
## F-statistic: 756.592 on 3 and 3297 DF, p-value: < 2.22e-16
```