HW7

Dandong Tu

2017/10/19

1(4.1)

```r
library(alr4)
```

```
## Loading required package: car
## Loading required package: effects
##
## Attaching package: 'effects'
## The following object is masked from 'package:car':
##
##     Prestige
```

```r
m1=lm(BMI18~WT2+WT9+WT18, data = BGSgirls)
summary(m1)
```

```
##
## Call:
## lm(formula = BMI18 ~ WT2 + WT9 + WT18, data = BGSgirls)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1037 -0.7432 -0.1240  0.8320  4.3485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.30978    1.65517   5.020 4.16e-06 ***
## WT2         -0.38663    0.15145  -2.553    0.013 *
## WT9          0.03141    0.04937   0.636    0.527
## WT18         0.28745    0.02603  11.044  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 66 degrees of freedom
## Multiple R-squared:  0.7772, Adjusted R-squared:  0.767
## F-statistic: 76.73 on 3 and 66 DF,  p-value: < 2.2e-16
```

This is the summary for the original regressors.

```r
BGSgirls$ave=with(BGSgirls, (WT2+WT9+WT18)/3)
BGSgirls$lin=with(BGSgirls, WT18-WT2)
BGSgirls$quad=with(BGSgirls, WT2-2*WT9+WT18)
m2=lm(BMI18 ~ ave+lin+quad,data = BGSgirls)
summary(m2)
```

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

```
## lm(formula = BMI18 ~ ave + lin + quad, data = BGSgirls)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1037 -0.7432 -0.1240  0.8320  4.3485
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.30978    1.65517   5.020 4.16e-06 ***
## ave         -0.06778    0.12751  -0.532    0.597
## lin          0.33704    0.07466   4.514 2.68e-05 ***
## quad        -0.02700    0.03976  -0.679    0.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.333 on 66 degrees of freedom
## Multiple R-squared:  0.7772, Adjusted R-squared:  0.767
## F-statistic: 76.73 on 3 and 66 DF,  p-value: < 2.2e-16
```

This is the summary for the revised regressors.

Comparing two summaries, we found that 1.$\hat{\sigma}^2$ are the same. 2.Intercept are the same 3. Residual standard error are the same. 4.All residuals are the same

We analyze the mean function for first model as following:

$$E(BMI18|Weight) = \beta_0 + \beta_1 WT2 + \beta_1 WT9 + \beta_3 WT18$$

While we have the mean function for the second model as :

$$\begin{aligned}
E(BMI18|Weight) &= \beta_0 + \beta_1 ave + \beta_2 lin + \beta_3 quad \\
&= \beta_0 + \beta_1 (WT2 + WT9 + WT18)/3 + \beta_2 (WT18 - WT2) + \beta_3 (WT2 - 2WT9 + WT18) \\
&= \beta_0 + (\beta_1/3 + \beta_2 + \beta_3) WT2 + (\beta_1/3 - 2\beta_3) WT9 + (\beta_1/3 - \beta_2 + \beta_3) WT18
\end{aligned}$$

The new mean function shows the relationships between old βs and new βs. We observed large **p-value** for **ave** and **quad** while the small **p-value** for **lin**. Therefore, we can conclude that the second model looks easier since only the linear trend has a small**p-value**, which is helpful for us to describe the changes in **BMI18** over time by increasing same amount on average.

2(4.2)

4.2.1

```r
Transact$a=(Transact$t1+Transact$t2)/2
Transact$d=(Transact$t1-Transact$t2)
summary(lm(Transact$time~Transact$t1+Transact$t2))
```

```
##
## Call:
## lm(formula = Transact$time ~ Transact$t1 + Transact$t2)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4652.4  -601.3     2.4   455.7  5607.4
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 144.36944  170.54410   0.847    0.398
## Transact$t1   5.46206    0.43327  12.607   <2e-16 ***
## Transact$t2   2.03455    0.09434  21.567   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1143 on 258 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9083
## F-statistic:  1289 on 2 and 258 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Transact$time~Transact$a+Transact$d))
```

```
##
## Call:
## lm(formula = Transact$time ~ Transact$a + Transact$d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4652.4  -601.3     2.4   455.7  5607.4
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 144.3694   170.5441   0.847    0.398
## Transact$a    7.4966     0.3654  20.514  < 2e-16 ***
## Transact$d    1.7138     0.2548   6.726 1.12e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1143 on 258 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9083
## F-statistic:  1289 on 2 and 258 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Transact$time~Transact$t2+Transact$d))
```

```
##
## Call:
## lm(formula = Transact$time ~ Transact$t2 + Transact$d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4652.4  -601.3     2.4   455.7  5607.4
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 144.3694   170.5441   0.847    0.398
## Transact$t2   7.4966     0.3654  20.514   <2e-16 ***
## Transact$d    5.4621     0.4333  12.607   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1143 on 258 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9083
## F-statistic:  1289 on 2 and 258 DF,  p-value: < 2.2e-16
```

```r
summary(lm(Transact$time~Transact$t1+Transact$t2+Transact$a+Transact$d))
```

```
##
## Call:
## lm(formula = Transact$time ~ Transact$t1 + Transact$t2 + Transact$a +
##     Transact$d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4652.4  -601.3     2.4   455.7  5607.4
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 144.36944  170.54410   0.847    0.398
## Transact$t1   5.46206    0.43327  12.607   <2e-16 ***
## Transact$t2   2.03455    0.09434  21.567   <2e-16 ***
## Transact$a         NA         NA      NA       NA
## Transact$d         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1143 on 258 degrees of freedom
## Multiple R-squared:  0.9091, Adjusted R-squared:  0.9083
## F-statistic:  1289 on 2 and 258 DF,  p-value: < 2.2e-16
```

Since **a=(t1+t2)/2** and **d=t1-t2**, these are exact linear relationships with **t1** and **t2**, therefore, only 2 of the 4 terms added after the intercept can be estimated.

4.2.2

```r
m11=lm(Transact$time~Transact$t1+Transact$t2)
m22=lm(Transact$time~Transact$a+Transact$d)
m33=lm(Transact$time~Transact$t2+Transact$d)
m44=lm(Transact$time~Transact$t1+Transact$t2+Transact$a+Transact$d)
compareCoefs(m11,m22,m33,m44,se=FALSE)
```

```
##
## Call:
## 1: lm(formula = Transact$time ~ Transact$t1 + Transact$t2)
## 2: lm(formula = Transact$time ~ Transact$a + Transact$d)
## 3: lm(formula = Transact$time ~ Transact$t2 + Transact$d)
## 4: lm(formula = Transact$time ~ Transact$t1 + Transact$t2 + Transact$a
##   + Transact$d)
##             Est. 1 Est. 2 Est. 3 Est. 4
## (Intercept) 144.37 144.37 144.37 144.37
## Transact$t1   5.46                 5.46
## Transact$t2   2.03          7.50   2.03
## Transact$a           7.50
## Transact$d           1.71   5.46
```

From the compareCoefs table we observed that the estimates for t1 and t2 are same in model 1 and 4. Based on the part1, the intercept, Residual standard error and Multiple R-squared are the same for each model.

4.2.3

In M1, the estimate is the change in response for one unit change in **t2**, while **t1** is fxied. In M3, the estimate is the change in Y for one unit change in **t2**, while d is fixed, where **d=t1-t2**. The way that **t1** can be increased by one unit with d fixed is to increase **t2**, therefore, the coefficient for **t2** in M3, which is **7.50** is the sum of the coefficients for **t1** and **t2** in M1, which is **5.46** and **2.03** relatively.

(4.6,4.7)

4.6

```
m3=lm(log(fertility)~pctUrban,data = UN11)
m3
```

```
##
## Call:
## lm(formula = log(fertility) ~ pctUrban, data = UN11)
##
## Coefficients:
## (Intercept)     pctUrban
##     1.50096     -0.01016
```

```
100*(exp(-0.0102)-1)
```

```
## [1] -1.014816
```

Based on the calculation shows above, we can say that increasing **pctUrban** by 1 unit is associated with **100(exp(-0.0102)-1)=-1.01482%** decrease in **fertility**

4.7

```
summary(lm(log(fertility)~log(ppgdp)+lifeExpF,data=UN11))
```

```
##
## Call:
## lm(formula = log(fertility) ~ log(ppgdp) + lifeExpF, data = UN11)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.61778 -0.16891  0.03731  0.17591  0.61072
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.50736    0.12707  27.601  < 2e-16 ***
## log(ppgdp)  -0.06544    0.01781  -3.675 0.000307 ***
## lifeExpF    -0.02824    0.00274 -10.306  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.248 on 196 degrees of freedom
## Multiple R-squared:  0.6926, Adjusted R-squared:  0.6894
## F-statistic: 220.8 on 2 and 196 DF,  p-value: < 2.2e-16
```

```
a=(exp(log(1.25)*(-0.06544)-1))
b=(exp(log(1)*(-0.06544)-1))
100*(a-b)/b
```

```
## [1] -1.449641
```

We observed the $\hat{\beta}_1 = -0.06544$ and then we plug in to the formula to verify that a 25% increase in **ppgdp** make the changes from **b** to **a** is about 1.4% decrease in expected fertility.