Regression Specification| R programming language

Preliminaries

Script Editor

```
setup
m(list = ls())
irectory <- "C:/Users/amalz/OneDrive/Desktop/"

Install packages
ackageNames <- c("tidyverse", "stargazer", "magrittr", "car")
or(i in PackageNames){
 if(!require(i, character.only = T)){
   install.packages(i, dependencies = T)
   require(i, character.only = T)
 }
```

RESET

- *The coefficient significance or adjusted R-squared can be used to choose a correctly specified model among many alternatives.*
- *The RESET (regression specification error test) can also be used to choose*
- *among these models.*
- *RESET includes squares and cubes of the fitted values in the regression model and tests for joint coefficient significance.*
- *In this example we will be using 4 models, model on wage, log wage, wage with squared terms and log wage with squared terms.*
- *The common steps for RESET for these models are running the initial regression, using mutate function to generate squares and cubes of the fitted values in the regression model and then testing for joint significance of coefficients on yhatsq and yhatcube using* linearHypothesis *command.*
- *The null hypothesis is the model is well-specified. If p-value<0.05 then the model is misspecified.*

Script Editor

```
wage1 <- read.csv(paste0(directory, "wage1.csv"))
model1 <- lm(wage ~ educ + exper + tenure, wage1)
summary(model1)
wage1 %<>% mutate(yhat = fitted(model1),
                  yhatsq = yhat^2,
                  yhatcube = yhat^3)

model1_RESET <- update(model1, ~ . + yhatsq + yhatcube)
summary(model1_RESET)
linearHypothesis(model1_RESET, c("yhatsq = 0", "yhatcube = 0"))
```

MODEL 1

Regression Specification| R programming language

Console

```
Call:
lm(formula = wage ~ educ + exper + tenure, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.6068 -1.7747 -0.6279  1.1969 14.6536 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.87273    0.72896  -3.941 9.22e-05 ***
educ         0.59897    0.05128  11.679  < 2e-16 ***
exper        0.02234    0.01206   1.853   0.0645 .  
tenure       0.16927    0.02164   7.820 2.93e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.084 on 522 degrees of freedom
Multiple R-squared:  0.3064,	Adjusted R-squared:  0.3024 
F-statistic: 76.87 on 3 and 522 DF,  p-value: < 2.2e-16
Call:
lm(formula = wage ~ educ + exper + tenure + yhatsq + yhatcube, 
    data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.0655 -1.5770 -0.6431  1.1461 15.0432 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)   
(Intercept)  5.309407   2.196706   2.417  0.01599 * 
educ        -0.364355   0.274468  -1.327  0.18493   
exper       -0.015075   0.016312  -0.924  0.35583   
tenure      -0.150291   0.082359  -1.825  0.06860 . 
yhatsq       0.204119   0.078237   2.609  0.00934 **
yhatcube    -0.006243   0.004173  -1.496  0.13526   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.024 on 520 degrees of freedom
Multiple R-squared:  0.336,	Adjusted R-squared:  0.3296 
F-statistic: 52.62 on 5 and 520 DF,  p-value: < 2.2e-16
Linear hypothesis test:
yhatsq = 0
yhatcube = 0

Model 1: restricted model
Model 2: wage ~ educ + exper + tenure + yhatsq + yhatcube

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)    
1    522 4966.3                                  
2    520 4754.8  2    211.52 11.566 1.217e-05 *** p-value<0.05, the model is misspecified.
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Regression Specification| R programming language

Script Editor

```
model2 <- lm(lwage ~ educ + exper + tenure, wage1)
summary(model2)
wage1 %<>% mutate(lyhat = fitted(model2),
                  lyhatsq = lyhat^2,
                  lyhatcube = lyhat^3)
model2_RESET <- update(model2, ~ . + lyhatsq + lyhatcube, wage1)
summary(model2_RESET)
linearHypothesis(model2_RESET, c("lyhatsq = 0", "lyhatcube = 0"))
```

MODEL 2

Console

```
Call:
lm(formula = lwage ~ educ + exper + tenure, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.05802 -0.29645 -0.03264  0.28788  1.42809 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.284359   0.104190   2.729  0.00656 ** 
educ        0.092029   0.007330  12.555  < 2e-16 ***
exper       0.004121   0.001723   2.391  0.01714 *  
tenure      0.022067   0.003094   7.133 3.29e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4409 on 522 degrees of freedom
Multiple R-squared:  0.316,	Adjusted R-squared:  0.3121 
F-statistic: 80.39 on 3 and 522 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = lwage ~ educ + exper + tenure + lyhatsq + lyhatcube, 
    data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.02663 -0.27034 -0.04341  0.27080  1.46412 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.582470   0.379563   4.169 3.58e-05 ***
educ        -0.347462   0.140312  -2.476  0.01359 *  
exper       -0.016092   0.006703  -2.401  0.01671 *  
tenure      -0.085481   0.033550  -2.548  0.01113 *  
lyhatsq      2.733926   0.959701   2.849  0.00456 ** 
lyhatcube   -0.494222   0.197001  -2.509  0.01242 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4362 on 520 degrees of freedom
Multiple R-squared:  0.3328,	Adjusted R-squared:  0.3264 
F-statistic: 51.88 on 5 and 520 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test:
lyhatsq = 0
lyhatcube = 0

Model 1: restricted model
Model 2: lwage ~ educ + exper + tenure + lyhatsq + lyhatcube

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)   
1    522 101.46                                
2    520  98.96  2    2.4955 6.5566 0.001541 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

p-value<0.05, the model is misspecified.

Regression Specification| R programming language

Script Editor

```
wage1 %<>% mutate(educsq = educ^2,
                  tenuresq = tenure^2)
model3 <- update(model1, ~ . + educsq + expersq + tenuresq)
summary(model3)
wage1 %<>% mutate(yhat1 = fitted(model3),
                  yhat1sq = yhat1^2,
                  yhat1cube = yhat1^3)
model3_RESET <- update(model3, ~ . + yhat1sq + yhat1cube)
summary(model3_RESET)
linearHypothesis(model3_RESET, c("yhat1sq = 0", "yhat1cube = 0"))
```

MODEL 3

Console

```
Call:
lm(formula = wage ~ educ + exper + tenure + educsq + expersq +
    tenuresq, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1467 -1.7825 -0.3346  1.1666 14.3732

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.4384097  1.3734346   1.775   0.0764 .
educ        -0.4994870  0.2183672  -2.287   0.0226 *
exper        0.1988154  0.0365855   5.434 8.48e-08 ***
tenure       0.2301329  0.0500825   4.595 5.44e-06 ***
educsq       0.0445760  0.0090145   4.945 1.03e-06 ***
expersq     -0.0041949  0.0007904  -5.308 1.65e-07 ***
tenuresq    -0.0027587  0.0017237  -1.600   0.1101
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.934 on 519 degrees of freedom
Multiple R-squared:  0.3762,	Adjusted R-squared:  0.369
F-statistic: 52.17 on 6 and 519 DF,  p-value: < 2.2e-16
Call:
lm(formula = wage ~ educ + exper + tenure + educsq + expersq +
    tenuresq + yhat1sq + yhat1cube, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max
-7.7714 -1.6205 -0.4841  1.0887 14.6685

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.954502   1.501385   1.302    0.194
educ        -0.086399   0.474427  -0.182    0.856
exper        0.152182   0.147161   1.034    0.302
tenure       0.155696   0.181741   0.857    0.392
educsq       0.018489   0.036680   0.504    0.614
expersq     -0.003168   0.003103  -1.021    0.308
tenuresq    -0.002676   0.002672  -1.002    0.317
yhat1sq     -0.040694   0.115303  -0.353    0.724
yhat1cube    0.006336   0.005388   1.176    0.240

Residual standard error: 2.869 on 517 degrees of freedom
Multiple R-squared:  0.4055,	Adjusted R-squared:  0.3963
F-statistic: 44.08 on 8 and 517 DF,  p-value: < 2.2e-16
Linear hypothesis test:
yhat1sq = 0
yhat1cube = 0

Model 1: restricted model
Model 2: wage ~ educ + exper + tenure + educsq + expersq + tenuresq +
    yhat1sq + yhat1cube

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    519 4466.7
2    517 4256.7  2       210 12.753 3.925e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

p-value<0.05, the model is misspecified.

Regression Specification| R programming language

Script Editor

```
model4 <- update(model2, ~ . + educsq + expersq + tenuresq)
summary(model4)
wage1 %<>% mutate(lyhat1 = fitted(model4),
                  lyhat1sq = lyhat1^2,
                  lyhat1cube = lyhat1^3)
model4_RESET <- update(model4, ~ . + lyhat1sq + lyhat1cube, wage1)
summary(model4_RESET)
linearHypothesis(model4_RESET, c("lyhat1sq = 0", "lyhat1cube = 0"))
```

MODEL 4

Console

```
Call:
lm(formula = lwage ~ educ + exper + tenure + educsq + expersq +
    tenuresq, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.92550 -0.24203 -0.03173  0.26922  1.32423

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.9589366  0.1953606   4.909 1.23e-06 ***
educ        -0.0518310  0.0310611  -1.669   0.0958 .
exper        0.0308265  0.0052040   5.924 5.74e-09 ***
tenure       0.0352960  0.0071239   4.955 9.82e-07 ***
educsq       0.0057790  0.0012822   4.507 8.13e-06 ***
expersq     -0.0006362  0.0001124  -5.659 2.52e-08 ***
tenuresq    -0.0005676  0.0002452  -2.315   0.0210 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4173 on 519 degrees of freedom
Multiple R-squared:  0.3907,    Adjusted R-squared:  0.3837
F-statistic: 55.47 on 6 and 519 DF,  p-value: < 2.2e-16
Call:
lm(formula = lwage ~ educ + exper + tenure + educsq + expersq +
    tenuresq + lyhat1sq + lyhat1cube, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.93298 -0.24399 -0.03674  0.26180  1.32896

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.1390879  1.0017863   1.137    0.256
educ        -0.0615067  0.1313678  -0.468    0.640
exper        0.0409712  0.0714968   0.573    0.567
tenure       0.0463257  0.0831935   0.557    0.578
educsq       0.0073013  0.0138412   0.528    0.598
expersq     -0.0008444  0.0014746  -0.573    0.567
tenuresq    -0.0007582  0.0013582  -0.558    0.577
lyhat1sq    -0.2828101  1.3557335  -0.209    0.835
lyhat1cube   0.0725223  0.2547436   0.285    0.776

Residual standard error: 0.4178 on 517 degrees of freedom
Multiple R-squared:  0.3915,    Adjusted R-squared:  0.3821
F-statistic: 41.58 on 8 and 517 DF,  p-value: < 2.2e-16
Linear hypothesis test:
lyhat1sq = 0
lyhat1cube = 0

Model 1: restricted model
Model 2: lwage ~ educ + exper + tenure + educsq + expersq + tenuresq +
    lyhat1sq + lyhat1cube

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    519 90.374
2    517 90.254  2    0.1206 0.3454 0.7081
```

p-value>0.05, the model is correctly specified.

Regression Specification| R programming language

Measurement error

Script Editor

```
wage1 %<>% mutate(wage_m = case_when(
  wage > 20 ~ 25,
  wage > 15 ~ 20,
  wage > 10 ~ 15,
  wage > 5 ~ 10,
  wage > 0 ~ 5
))
```

Generate mismeasured variable wage_m, rounded up to next 5 dollars

```
wage1 %<>% mutate(exper_m = case_when(
  exper > 40 ~ 50,
  exper > 30 ~ 40,
  exper > 20 ~ 30,
  exper > 10 ~ 20,
  exper > 0 ~ 10
))
```

Generate mismeasured variable exper_m, rounded up to next 10 years

```
summary(model1)
```

 # Model with no measurement error
 # model1 <- lm(wage ~ educ + exper + tenure, wage1)

```
model8 <- update(model1, wage_m ~ .)
summary(model8)
```

Model with mismeasured dependent variable wage_m

```
model9 <- update(model1, ~ . - exper + exper_m)
summary(model9)
```

Model with mismeasured independent variable exper_m

Console

```
> summary(model1)

Call:
lm(formula = wage ~ educ + exper + tenure, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max
-7.6068 -1.7747 -0.6279  1.1969 14.6536

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.87273    0.72896  -3.941 9.22e-05 ***
educ         0.59897    0.05128  11.679  < 2e-16 ***
exper        0.02234    0.01206   1.853   0.0645 .
tenure       0.16927    0.02164   7.820 2.93e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.084 on 522 degrees of freedom
Multiple R-squared:  0.3064,	Adjusted R-squared:  0.3024
F-statistic: 76.87 on 3 and 522 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = wage_m ~ educ + exper + tenure, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max
-8.0153 -2.0914 -0.3999  2.0669 15.3401

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.18259    0.78010  -1.516    0.130
educ         0.62479    0.05488  11.384  < 2e-16 ***
exper        0.01754    0.01290   1.360    0.175
tenure       0.18675    0.02316   8.063 5.14e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.301 on 522 degrees of freedom
Multiple R-squared:  0.3018,	Adjusted R-squared:  0.2977
F-statistic:  75.2 on 3 and 522 DF,  p-value: < 2.2e-16
```

Coefficients are not biased, but the coefficient on exper became insignificant.

Console ctd...

```
Call:
lm(formula = wage ~ educ + tenure + exper_m, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.7358 -1.7570 -0.6686  1.2200 14.6379 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.72403    0.76739  -3.550  0.00042 ***
educ         0.58833    0.05140  11.446  < 2e-16 ***
tenure       0.17693    0.02153   8.218 1.64e-15 ***
exper_m      0.01472    0.01241   1.186  0.23610    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.09 on 522 degrees of freedom
Multiple R-squared:  0.3037,	Adjusted R-squared:  0.2997 
F-statistic: 75.91 on 3 and 522 DF,  p-value: < 2.2e-16
```

Attenuation bias - coefficient on exper is biased toward zero and became insignificant.