Chapter 2 - The Simple Regression Model

Thalles Quinaglia Liduares

11/02/2022

Exercise 2.8

Upload packages

```
library(lmreg)
```

To complete this exercise you need a software package that allows you to generate data from the uniform and normal distributions.

(i) Start by generating 500 observations x_i – the explanatory variable – from the uniform distribution with range [0,10]. (Most statistical packages have a command for the Uniform[0,1] distribution; just multiply those observations by 10.) What are the sample mean and sample standard deviation of the x_i ?

The mean value and variance are given, respectively by

$$\mathbb{E}(X) = \frac{1}{2}(b+a) \ \ and \ \ \ VAR(X) = \frac{1}{12}(b-a)^2$$

First, generating the uniform distribuition in the range [0,10].

```
set.seed(123) # to generate always the same values

unif<-runif(500, min=0, max=10)

mean1<-round(mean(unif),2)

sd1<-round(sd(unif),2)
```

The `mean` and `sd` is given by 4.95 and 2.84, respectively.

(ii) Randomly generate 500 errors, ui, from the Normal[0,36] distribution. (If you generate a Normal[0,1], as is commonly available, simply multiply the outcomes by six.) Is the sample average of the ui exactly zero? Why or why not? What is the sample standard deviation of the u_i?

```
set.seed(312)

u<-rnorm(500, mean=0, sd=6)

mean_u<-round(mean(u),2)

sd_u<-round(sd(u),2)
```

Now, in this case, the `mean` and `sd` are given by -0.27 and 6, respectively.

(iii) Now generate the y_i as

$$y_i = 1 + 2x_i + u_i = \beta_0 + \beta_1 x_i + u_i$$

that is, the population intercept is one and the population slope is two. Use the **data to run the regression of y_i on x_i. What are your estimates of the intercept *and slope? Are they equal to the population values in the above equation? Explain.**

```
yi<-1+2*unif+u

est_yi<-lm(yi~unif)

summary(est_yi)
```

```
##
## Call:
## lm(formula = yi ~ unif)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.4715  -4.2388   0.0314   4.1798  19.6407
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.58394    0.53929   1.083    0.279
## unif         2.03019    0.09445  21.496   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.001 on 498 degrees of freedom
## Multiple R-squared:  0.4813, Adjusted R-squared:  0.4802
## F-statistic: 462.1 on 1 and 498 DF,  p-value: < 2.2e-16
```

Given the estimates, the coefficients are almost equal to the population values.

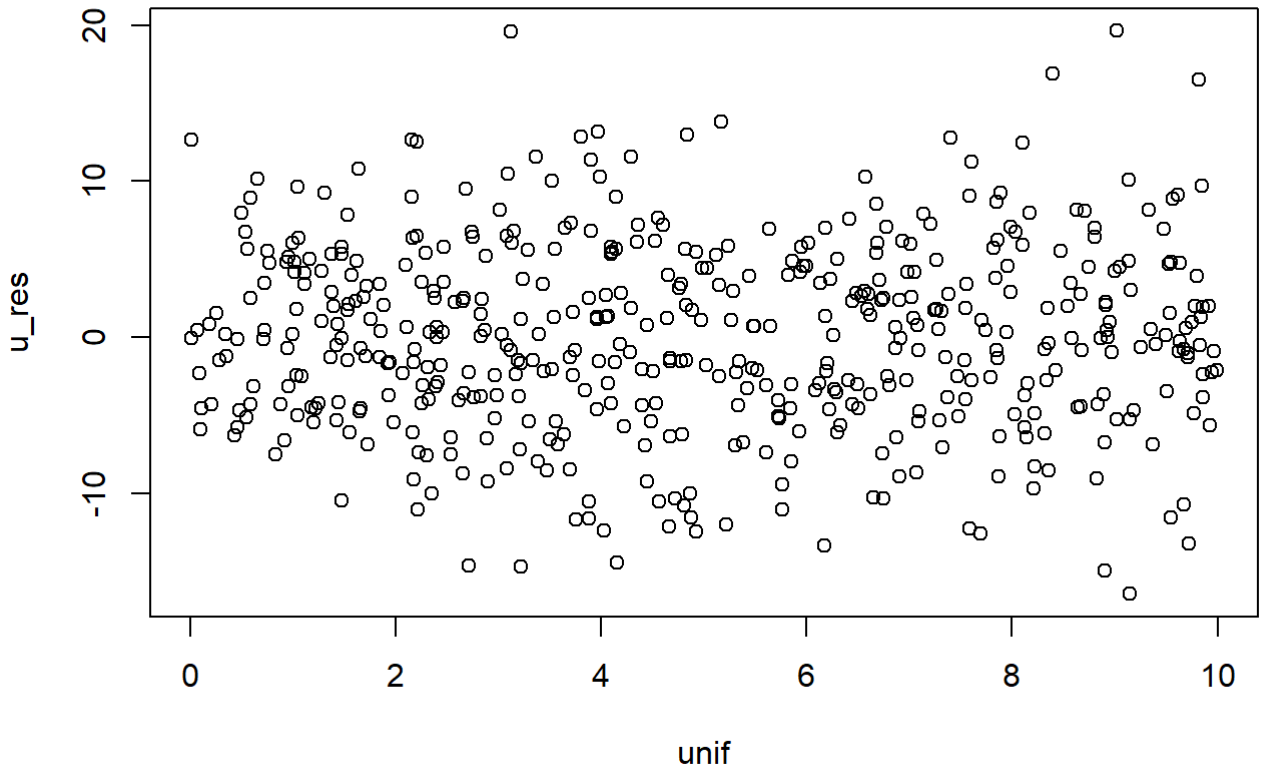
(iv) Obtain the OLS residuals, u_i, and verify that equation (2.60) hold (subject to rounding error).

```
u_res<-resid(est_yi)

summary(u_res)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -16.47154  -4.23883   0.03142   0.00000   4.17980  19.64067
```

```
plot(u_res~unif)
```


u_res
20
10
0
-10
0
2
4
6
8
10
unif