ECON 3140 HW 2

Nick Gembs

2/13/2022

```r
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

data <- read.csv("C:/Users/Nick/Downloads/data_17.csv")

x = data$X1
y = data$X0

# Assumption SLR.1 does not provide any restriction on the model as it only
identifies the error of each point fitted on to the model as a residual (u).
The residuals to not have to be of any accuracy to fit SLR.1. For SLR.4, as
long as the linear model is properly fitted, the sum of the residuals will
always be 0. Therefore the expected value of the residual for both x=1 and
x=0 will be 0. Jointly, these to assumptions show that the expected value of
Y given X is equal to the linear model, B0 + B1x. Going back to hw1, the
model will be fitted so that the expected value of the model at x=0 will be
the sample mean of the y values at 0. The same occurs at x=1.
summary(data)

##        X0                X1
##  Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000
##  Mean   :0.1818   Mean   :0.5354
##  3rd Qu.:0.0000   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000

# For SLR.5, the assumption does not always hold as the variance of u is not
always independent of x. In this case, the variance of u is different for x=0
and x=1, suggesting that homoskedasticity is not present. SLR.6 cannot
possibly hold because the distribution of the u's given x cannot be a normal
distribution. This is because the only output y values in this set of data
are 1 and 0, these values will not fit such a strong assumption as normality.
```

```r
lm.fit <- lm(y~x)
lm.fit
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)            x
##     0.17391      0.01477
```

```r
var(data)
```

```
##             X0          X1
## X0 0.150278293 0.003710575
## X1 0.003710575 0.251288394
```

```r
dataf<-split(data, data$X1)

var(dataf$'0')
```

```
##           X0 X1
## X0 0.1468599  0
## X1 0.0000000  0
```

```r
var(dataf$'1')
```

```
##           X0 X1
## X0 0.1560232  0
## X1 0.0000000  0
```

```r
# If in the real model beta1 = 0, homoskedasicity would in fact hold, as the
variance would stay consistent from x=0 to x=1. This is because in order to
make beta1=0, there must be an equal number of 1 and 0 y values for each x
value. This will create a zero slope fitted line. This idea will allow us to
use looser analysis under asymptotal analysis, but is not accurate in real
life because there will almost always be variance and unexpected results in a
random sample.

#1.4

lm.fit <- lm(data)
lm.fit
```

```
##
## Call:
## lm(formula = data)
##
## Coefficients:
```

```
## (Intercept)           X1
##     0.17391      0.01477
```

```r
Confidence <- confint(lm.fit, 'X1',level=.95)
Confidence
```

```
##          2.5 %    97.5 %
## X1 -0.1410443 0.1705767
```

```r
# Beta 1 value of 0 is within the 95% confidence interval
```