

Sesgo/Varianza Regresión Lineal Simple

Mathias Bourel

4/8/2021

#Simulacion de datos

Simulamos 50 datos con $f(x) = 2 + 3x$ es decir $y_i = 2 + 3x_i + \epsilon_i$ donde $\epsilon_i \sim N(0, 2)$ para todo $i = 1, \dots, 50$.

```
X=runif(50,0,5)
Y=2+3*X+rnorm(50,0,2)
```

```
modelo=lm(Y~X)
modelo
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Coefficients:
## (Intercept)          X
##      2.150         3.067
```

```
x11()
plot(X,Y,main="Una primera simulacion")
abline(b=modelo$coefficients[2],a=modelo$coefficients[1],col='red',lwd=1)
abline(b=3,a=2,col='green',lty=1,lwd=2)
```

```
Y1=2+3*X+rnorm(50,0,10)
points(X,Y1)
modelo2=lm(Y1~X)
modelo2
```

```
##
## Call:
## lm(formula = Y1 ~ X)
##
## Coefficients:
## (Intercept)          X
##      -1.087         4.566
```

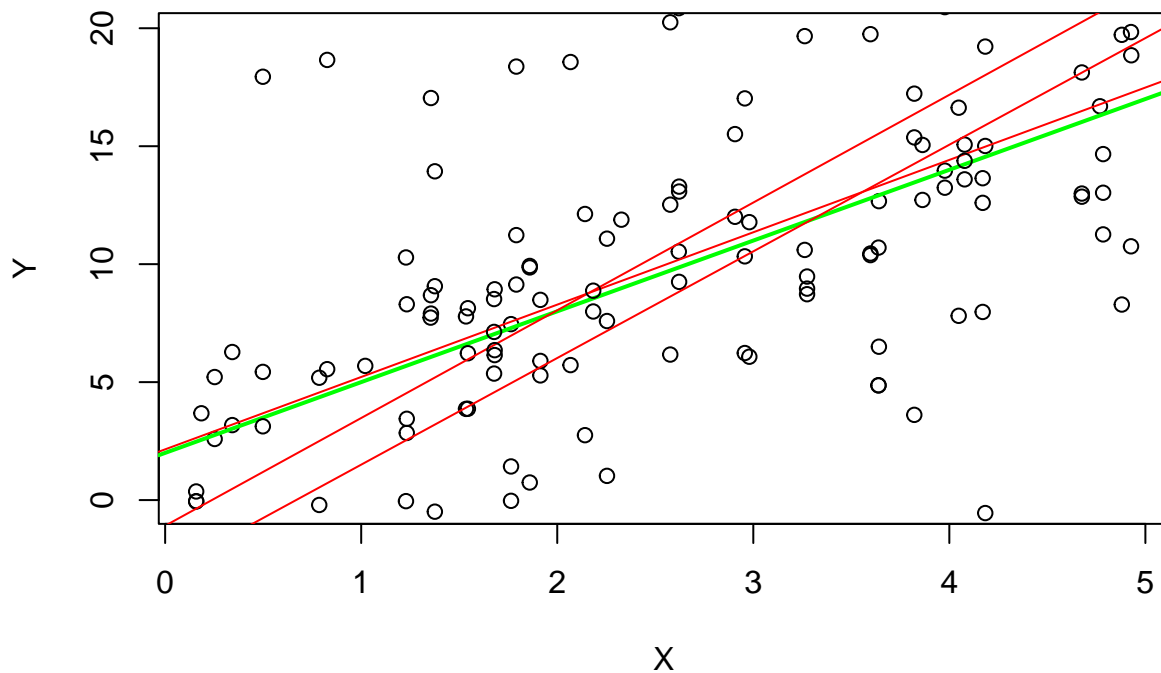
```
abline(b=modelo2$coefficients[2],a=modelo2$coefficients[1],col='red',lwd=1)
```

```
Y2=2+3*X+rnorm(50,0,10)
points(X,Y2)
modelo3=lm(Y2~X)
modelo3
```

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

```
## lm(formula = Y2 ~ X)
##
## Coefficients:
## (Intercept)          X
##      -3.019         4.520
abline(b=modelo3$coefficients[2],a=modelo3$coefficients[1],col='red',lwd=1)
```

Una primera simulación



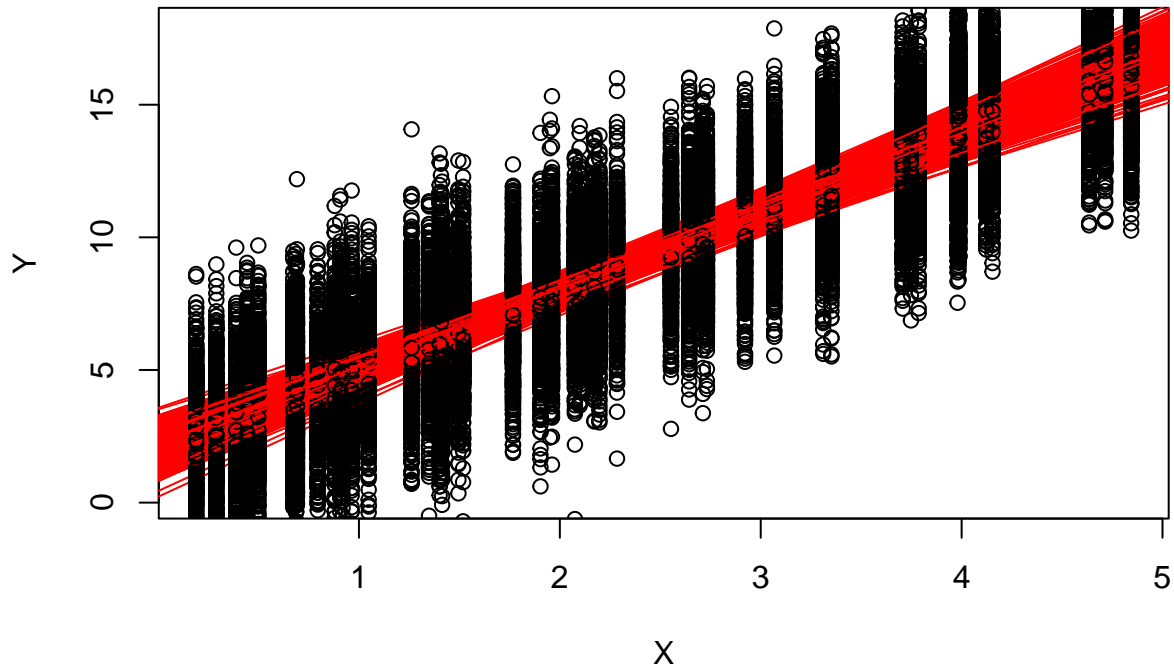
La clase de función elegida es la de los polinomios de grado 1 y con modelo, modelo 2 y modelo 3 veo cuanto varía el modelo (varianza) al cambiar la muestra. El sesgo de la estimación es cero porque la clase de funciones es la misma que la función teórica y los estimadores MC son insesgados.

Acá va una función que hace lo mismo que en las transparencias de clases:

```
simulacion=function(beta0=2,beta1=3,n=50,M=15){
  X=runif(n,0,5)
  Y=beta0+beta1*X+rnorm(n,0,2)
  plot(X,Y)
  abline(b=3,a=2,col='green',lty=1,lwd=4)
  vecbeta0=NULL
  vecbeta1=NULL
  for(i in 1:M){
    Y=beta0+beta1*X+rnorm(n,0,2)
    points(X,Y)
    modelo=lm(Y~X)
    abline(b=modelo$coefficients[2],a=modelo$coefficients[1],col='red',lwd=1)
    vecbeta0[i]=modelo$coefficients[1]
    vecbeta1[i]=modelo$coefficients[2]
  }
  list(vecbeta0=vecbeta0,vecbeta1=vecbeta1)
}
```

Ejecuto la función. Los promedios me dan una estimación de los (verdaderos) parámetros, β_0 aprox 2 y β_1 aprox 3, verificando que la estimación de la regresión lineal es insesgada.

```
a=simulacion(beta0=2,beta1=3,n=50,M=500)
```



```
mean(a$vecbeta0)
```

```
## [1] 1.997223
```

```
mean(a$vecbeta1)
```

```
## [1] 2.990506
```

```
““
```