Segundo Corte: Regresión lineal Múltiple

Andrés Martínez

2019

Índice

1. Regresión Lineal Múltiple

En un modelo lineal múltiple se asume qye Y es una función lineal de sus predictores mas un ruido blanco:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_p X_p + \epsilon \tag{1}$$

1.1. Supuestos

1. Todos los parámetros son lienales
2. La muestra de n observaciones sigue el modelo poblacional del primer supuesto
3. No existe perfecta colinearidad
4. EL error tiene un valor esperado de cero dado por variables independientes.
5. Homocedasticidad: El error tiene la misma varianza que los valores de las variables explicativas.

No hacemos suposiciones sobre las distribuciones (marginales o conjuntas) de X_i, pero suponemos que $E[\epsilon|X] = 0$, $V[\epsilon|X] = \sigma^2$, y ese ϵ no está correlacionado entre las mediciones. La forma matricial del modelo es:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{2}$$

Donde **X** incluye una columna inicial de 1.

Cuando se incluye se asume que este debe tener un comportamiento normal como se ve en la siguiente ecuación:

$$\epsilon \sim MVN(\mathbf{0}, \sigma^2\mathbf{I}) \tag{3}$$

independiente de **X**.

Los coeficientes se estiman de la siguiente forma:

$$\widehat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{4}$$

Bajo el supuesto de ruido blanco, este es el mismo resultado que la estimación por máxima verosimilitud

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\beta} = \mathbf{H}\mathbf{Y} \tag{5}$$

Donde $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$

Que es simetrico e idempotente. Los residuales se obteienen así:

$$SSE = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{6}$$

donde $\mathbf{I} - \mathbf{H}$ es simetrico e idempotente

1.2. Derivación de los Estimadores

$$\widehat{\beta}_{OLS} = argmin_{\beta \in \mathbb{R}^2}(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) \tag{7}$$

$$\frac{\partial}{\partial \beta} = (\mathbf{Y}^T\mathbf{Y} - 2\mathbf{Y}^T(\mathbf{X}\beta + \beta^T\mathbf{X}^T\mathbf{X}\beta)) \tag{8}$$

$$\frac{\partial}{\partial \beta} = -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\hat{\beta} = 0 \tag{9}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{10}$$

1.3. Propiedades de los Estimadores

Insesgamiento del estimador mínimo cuadrado

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{11}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}^T(\mathbf{X}\beta + \epsilon) \tag{12}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})\beta \tag{13}$$

$$+(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon) \tag{14}$$

$$= \beta + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon) \tag{15}$$

$$E[\hat{\beta}|x] = \beta \tag{16}$$

Varianza de los errores estandar: Suponiendo Homocedasticidad

$$V[\hat{\beta}|\mathbf{X}] = V[\beta + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon)/\mathbf{X}] \tag{17}$$

$$V[\hat{\beta}|\mathbf{X}] = V[(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon)/\mathbf{X}] \tag{18}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}V[\mathbf{X}^T\epsilon/\mathbf{X}](\mathbf{X}^T\mathbf{X})^{-1} \tag{19}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T V[\epsilon/\mathbf{X}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{20}$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{21}$$

$$\hat{\beta} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{22}$$

$$\hat{\beta} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \tag{23}$$

Consistencia

- Ley de los grandes números
- Teorema Límite Central

Tenemos que $\hat{\beta} = \beta + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\epsilon)$ y se desea ver si $\hat{\beta}$ es consistente. $\lim_{n\to p}(\hat{\beta} - \beta) = 0$

Eficiencia: Bajo los supuestos de linealidad, muestreo aleatorio, no colinealidad perfecta, media condicional cero de los errores y homocedasticidad, $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ es el mejor estimador lineal insesgado.

Además $\hat{\beta}_1 = \beta_1 + (\mathbf{X}_1^T\mathbf{X}_1)^{-1}(\mathbf{X}_1^T\epsilon_2)$.

1.4. Colinealidad

Asumiendo que $(\mathbf{X}^T\mathbf{X})^{-1}$ existe y es invertible o no singular, existen un número equivalente de condiciones para una matriz que debe ser invertible:

- Su determinante no es cero
- Su rango de columnas es completo, lo que quiere decir que todas las columnas son lienalmente independientes.
- Su rango de filas es completo, lo que quiere decir que las filas son linealmente independientes.

Colinearidad perfecta:

Usando uno de los ejemplos del libro, que tiene la base de datos vote1, se hace una regresión entre los gastos de la campaña de A y B más los totales, con respecto a el voto de A.

```
> library(wooldridge)
> data("vote1")
> total=vote1$expendA +vote1$expendB
> model1=lm(vote1$voteA~ vote1$expendA+vote1$expendB+total)
> summary(model1)

Call:
lm(formula = vote1$voteA ~ vote1$expendA + vote1$expendB + total)

Residuals:
     Min       1Q   Median       3Q      Max
-27.5661  -8.6919   0.1641   9.0789  27.0131

Coefficients: (1 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   49.618995   1.426147   34.79   <2e-16 ***
vote1$expendA  0.038331   0.003387   11.32   <2e-16 ***
vote1$expendB -0.036127   0.003107  -11.63   <2e-16 ***
total                NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.58 on 170 degrees of freedom
Multiple R-squared:  0.5299,	Adjusted R-squared:  0.5244
F-statistic: 95.83 on 2 and 170 DF,  p-value: < 2.2e-16
```

El resultado evidencia que cuando existe colinealidad perfecta, no es posible obtener el coeficiente para ese valor, porque este está directamente relacionado con los gastos en A y B. Esto se abordará más adelante cuando se profundice más en multicolinealidad.

```
> votop=1.2*vote1$voteA+rnorm(1)
> model2=lm(vote1$voteA~ vote1$expendA+vote1$expendB+votop)
> summary(model2)

Call:
lm(formula = vote1$voteA ~ vote1$expendA + vote1$expendB + votop)

Residuals:
       Min         1Q     Median         3Q        Max
-2.710e-13 -1.129e-15  1.352e-15  4.333e-15  7.269e-14

Coefficients:
                Estimate Std. Error    t value Pr(>|t|)
(Intercept)   -1.557e+00  8.260e-15 -1.885e+14  < 2e-16 ***
vote1$expendA  4.789e-17  8.870e-18  5.399e+00 2.24e-07 ***
vote1$expendB -2.289e-17  8.234e-18 -2.780e+00  0.00606 **
votop          8.333e-01  1.264e-16  6.592e+15  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.289e-14 on 169 degrees of freedom
Multiple R-squared:      1,        Adjusted R-squared:      1
F-statistic: 3.082e+31 on 3 and 169 DF,  p-value: < 2.2e-16
```

Si no existe correlación perfecta, pero hay un alto grado de colinealidad como en el modelo 2, puede haber un problema de especificación probocando que los errores de los coeficientes aumenten. Para evitar la colinealidad entre las variables, se debe tener claro desde antes la relación entre las variables, pero también se deben probar los modelos de tal forma que los errores en los coeficientes sean mínimos y estas variables sean significativas.

1.5. Goodness-of-fit

El objetivo más importante de estos modelos es obtener los errores, pues permiten establecer la eficiencia del modelo y las posibles falencias de especificación que este pueda tener. Para esto definimos rapidamente los tres errores que siempre se buscan en una regresión lineal múltiple o simple:

$$SCT = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \mathbf{Y}^T(\mathbf{I} - \mathbf{I}_{n\times n})\mathbf{Y}. \tag{24}$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y})^2 = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}. \tag{25}$$

$$SSR = \sum_{i=1}^{n}(\hat{y} - \overline{y})^2 = \mathbf{Y}^T(\mathbf{H} - \mathbf{I}_{n\times n})\mathbf{Y}. \tag{26}$$

Con estos resultados se pouede obtener el coeficiente de determinación múltiple R^2.

$$R^2 = \frac{SSR}{SCT} = 1 - \frac{SSE}{SST}. \tag{27}$$

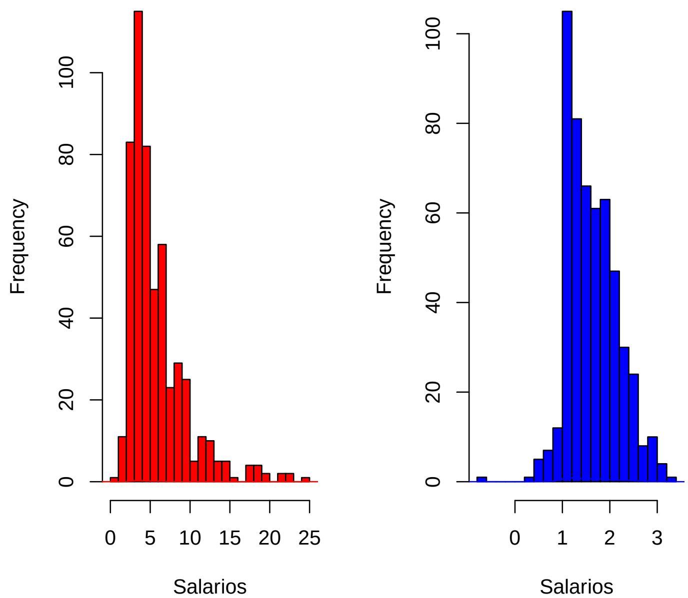
1.6. Propiedades de los estimadores de mínimos cuadrados

- $E[\hat{\beta}_i] = \beta_i$
- $V[\hat{\beta}_i] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$
- Un estimado insesgado de σ^2 es $S^2 = SSE/[n-(k+1)]$, donde $SSE = (\mathbf{Y}-\mathbf{X}\beta)^T(\mathbf{Y}-\mathbf{X}\beta)$
- Cada β_i está distribuido normalmente

1.7. Ejemplo

Usando el ejemplo 3.2 del libro de wooldrige, se hace la regresión del logaritmo de los salarios con respecto a la educación, la experiencia y el tiempo empleado. El proposito de obtener el logaritmo de los salarios es para generar una distribución simétrica y mejorar la escala.

```
> data("wage1")
> par(mfrow=c(1,2))
> hist(as.numeric(wage1$wage),breaks=20,main="",xlab="Salarios", col="red")
> lines(density(wage1$wage), col = 10)
> hist(as.numeric(wage1$lwage),breaks=20,main="",xlab="Salarios", col="blue")
> lines(density(wage1$lwage), col = 4)
>
>
```



La idea es trabajar con variables que tengan una distribución simétrica y al generar el logaritmo, los salarios adquieren una mejor forma como se ve en la gráfica anterior.

Para observar mejor el cambio, primer se hace la regresión con el logaritmo de los salarios, luego se hace con los salarios y se compara el resultado de cada uno de los coeficientes y el coeficiente de correlación

```
> modelo=lm(wage1$lwage~ wage1$educ+wage1$exper+wage1$tenure)
> summary(modelo)

Call:
lm(formula = wage1$lwage ~ wage1$educ + wage1$exper + wage1$tenure)

Residuals:
     Min       1Q   Median       3Q      Max
-2.05802 -0.29645 -0.03265  0.28788  1.42809

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.284360   0.104190   2.729  0.00656 **
wage1$educ   0.092029   0.007330  12.555  < 2e-16 ***
wage1$exper  0.004121   0.001723   2.391  0.01714 *
wage1$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

>
>
```

El segundo modelo es

```
> modelo1=lm(wage1$wage~ wage1$educ+wage1$exper+wage1$tenure)
> summary(modelo1)

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

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 ***
wage1$educ    0.59897    0.05128  11.679  < 2e-16 ***
wage1$exper   0.02234    0.01206   1.853   0.0645 .
wage1$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

>
```

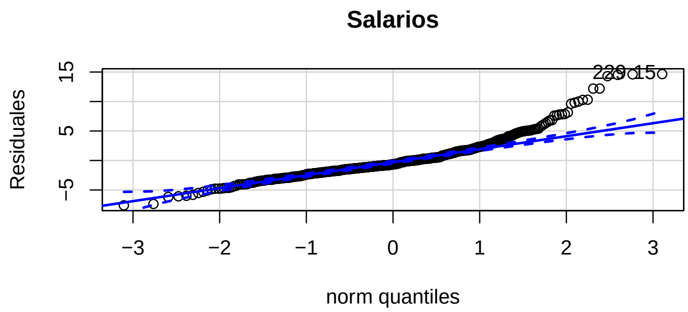
Como se puede ver el error en los coeficientes de cada variable aumenta, si bien el coeficiente de correlación aun no es óptimo para generar la regresión, si se puede ver una mejora cuando se trabaja con el logaritmo de los salarios.

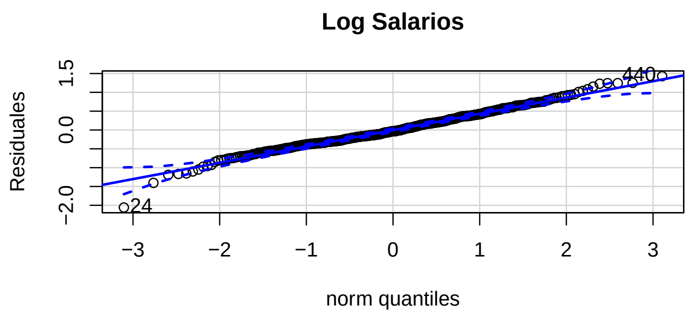
Para continuar con el análisis se hace un gráfico qq plot de los residuales con el objetivo de observar cuál de los dos cumple con el suspuesto de normalidad.

```
> library(car)
> par(mfrow=c(2,1))
> qqPlot(modelo1$residuals,main = "Salarios",ylab="Residuales")

[1]  15 229
```

```
> qqPlot(modelo$residuals,main = "Log Salarios",ylab="Residuales")
[1]  24 440
```





Al observar los residuales de ambos modelos, es mucho más claro que transformar la variable de salarios genera un mejor proceso de la regresión.

En esta sección se va a calcular el modelo del logaritmo de los salarios a partir de la forma matricial.

Primero preparamos las matrices.

```
> Y=wage1$lwage
> X=cbind(wage1$educ,wage1$exper,wage1$tenure)
> X=cbind(rep(1,length(Y)),X)
>
```

Recuerde que se debe aderir una columna de unos para el intercepto. Luego se obtienen los betas con ayuda del método de mínimos cuadrados.

```
> library(MASS)
> Beta=ginv(t(X)%*%X)%*%t(X)%*%Y
> Beta
            [,1]
[1,] 0.284359541
```

```
[2,] 0.092028988
[3,] 0.004121109
[4,] 0.022067218
```

```
> modelo$coefficients
```

```
 (Intercept)   wage1$educ  wage1$exper wage1$tenure 
 0.284359541  0.092028988  0.004121109  0.022067218 
```