**ESCUELA POLITÉCNICA NACIONAL**

**MODELOS LINEALES Y DISEÑO DE EXPERIMENTOS**

**REGRESIÓN LINEAL SIMPLE**

**PATRICIA GUERRERO**

1. **Objetivos**

* Construir un modelo de regresión lineal simple que describa los datos dados de una variable X(Ventas) sobre otra variable Y(Utilidad)
* Obtener los estimadores de los parámetros del modelo creado.
* Construir la tabla ANOVA

1. **Marco Teórico**

El método de análisis llamado *análisis de regresión,* investiga y modela la relación entre una variable *Y* dependiente o *de respuesta* en función de otras variables de *predicción X*, a través del método de mínimos cuadrados.

Al tomar observaciones de ambas variables *Y respuesta y X predicción o regresor*, se puede representar cada punto en un *diagrama de dispersión.*

Y \*

\*\* \* \*

\* \* \*

\*\*\*

X

Fig. 1.1 Diagrama de dispersión y recta de ajuste

El modelo de ajuste o *modelo de regresión lineal* es:

 (1.1)

Donde los coeficientes β1 y β2 son parámetros del modelo denominados *coeficientes de regresión,* son constantes, a pesar de que no podemos determinarlos exactamente sin examinar todas las posibles ocurrencias de X y Y, podemos usar la información proporcionada por una muestra para hallar sus estimadores , El error es difícil de determinar puesto que cambia con cada observación *Y*. Se asume que los errores tienen media cero, varianza desconocida σ2 y no están correlacionados (el valor de uno no depende del valor de otro). Por esto mismo las respuestas tampoco están correlacionadas.

Conviene ver al predictor ***X*** como la variable controlada por el analista y evaluada con el mínimo error, mientras que la variable de respuesta ***Y*** es una variable aleatoria, es decir que existe una distribución de ***Y*** con cada valor de ***X***.

La media de esta distribución es:

 (1.1 a)

y su varianza es:

 (1.1b)

De esta forma la media de ***Y*** es una función lineal de ***X*** a pesar de que la varianza de **Y** no dependa de los valores de ***X***.

Estimación de los parámetros por mínimos cuadrados

El método de mínimos cuadrados se usa para estimar β1 y β2 se estimará β1 y β2 de manera que la suma de cuadrados de las diferencias entre las observaciones *yi* y la línea recta sea mínima.

Los parámetros β0 y β1 son desconocidos y deben ser estimados usando datos de una muestra. Supongamos que se tienen ***n*** pares de datos:

(y1, x1), (y1, x1), (y2, x2),....., (yn, xn) de un experimento o por historia.

De la ecuación modelo de regresión de la población



Usando los pares de datos se puede establecer el criterio de mínimos cuadrados como:



Los estimadores de mínimos cuadrados de β0 y β1 por decir debe satisfacer es:



y



Simplificando estas dos ecuaciones se obtienen las *ecuaciones de mínimos cuadrados*:

La solución a las ecuaciones normales anteriores:





Donde los promedios para X y para Y son los siguientes:

Aplicando el método de mínimos cuadrados del error, se obtiene el modelo que nos da un valor estimado Y en función de X, denominado ecuación de predicción o de regresión lineal, como sigue:

  (1.2)

Donde:

 (1.3)

 (1.4)

Por tanto:

 (1.5)

Cuando  se tiene el punto  que se encuentra en la línea ajustada y representa el centro de gravedad de los datos.

Finalmente la regresión es: , 

Regresión sin término constante: 

**PRUEBA DE SIGNIFICANCIA**

La prueba de significancia del modelo nos permite determinar estadísticamente si las variables independientes (en conjunto) tienen efecto o no sobre la variable dependiente.

Para realizar esta prueba se requiere descomponer la suma total de cuadrados, representada por STC, en dos componentes: SEC y SRC

STC= SEC + SRC

Donde:

STC es la suma total de cuadrados

SEC es la suma de cuadrados de la regresión

SRC es la suma de cuadrados del error

Las ecuaciones apropiadas para calcular las expresiones anteriores son:





SRC= STC -SEC

Partimos de las hipótesis:



Utilizamos la tabla de análisis de varianza la tabla ANOVA:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fuente de Variación** | **Grados**  **de libertad** | **Suma de cuadrados** | **Media de cuadrados** | **Estadístico de prueba** |
| Regresión | K=1 | SEC |  |  |
| Error | n – p, p=2 | SSE |  |
| Total | n – 1 | STC |  |  |

El estadístico de prueba F0 tiene una distribución F (Fisher) con v1 = k y v2 = n – p grados de libertad en el numerador y el denominador, respectivamente.

En este caso, si el estadístico de prueba es mayor que el valor de tablas F, k, n – p, se rechaza la hipótesis nula; concluiremos que la variable independiente está relacionada con al menos una de las variables independientes.

**PRUEBAS SOBRE COEFICIENTES INDIVIDUALES**

En la prueba de significancia determinamos si existe o no relación entre la variable dependiente y las variables independientes en conjunto, es decir, no se puede determinar la relación entre Y y cada una de las variables independientes.

Se pueden realizar pruebas individuales para analizar la relación entre la variable dependiente y cada una de las variables independientes.

Partimos de las hipótesis siguientes:

 para j = 1, 2, ..., k

el estadístico de prueba apropiado es:



donde  es el valor de la diagonal principal de la matriz inversa ( (**X´X**)-1 ):



El estadístico de prueba t0 anterior sigue una distribución t-student con v = n – p grados de libertad.

Entonces, si el valor absoluto del estadístico de prueba es mayor que el valor de tablas t/2, n – p, se rechaza la hipótesis nula. Como conclusión diremos que la variable independiente Y sí está relacionada con la variable independiente Xj.

Intervalos de confianza para ****

En base al error estándar para los parámetros se tiene:

 (1.14)

 (1.15)

1. **Desarrollo**

En nuestro caso tenemos que encontrar , es decir k=2.

Realizamos la regresión lineal de los Datos en R Studio con los siguientes códigos y análisis:

data <- read.table("data\_rls\_uti.txt", header = TRUE, dec=",", sep="\t")

#View(data)

str(data)

data

#Realizamos una regresión lineal de los datos

reg<-lm(Utilidad~Ventas,data)

summary(reg)

#las ventas aumentan en una unidad entonces la utilidad aumenta 0.43

#B1=137.08 Y B2=0.43

#Y=B1+B2X

#RAZON T t1=B1^/ee(B1^)=0.485

# |t1|>tn-2(alfa/2)=t(40-2)(0.025\_)=2.024394

qt(0.975,df=38) # valor percentil t student

#en este caso 0.48> 2.024 no se cumple, no rechazamos Ho,asi que puede centrar los datos ya que beta1 no es significativo

#valor p<0.05 rechazo la hipotesis nula valor p=0.63 solo rechazamos

#Son t2=23.663 aqui si rechazo ya que 23.663>2.024 , rechazo Ho a un 95% de confianza la variable ventas influye en la utilidad

#R^2 el 93% de los dtos son explicados por la regresion

str(reg)

anova<-aov(reg)

summary(anova)

#df grados de libertad

#Ho:B2=0 Y H1:B2!=0

#Ho:B2=B3=...

#RECHAZO H0 si F>F(1,n-2\_)(alfa) a favor de H1:B2!=0

#F=559

qf(0.95,df1=1,df2=38) #valor percentil F fisher

#4.098172

#RECHAZO Ho Y CONCLUIMOS QUE LA REGRESION ES SIGNIFICATIVA

#intervalos de confianza

confint(reg,level=0.95)

#solo analizando el intervalo de confianza rechazamos la hipotesis nula de beta 2, a un 95% de confianza

names(reg)

str(reg[["residuals"]])

res<-reg[["residuals"]]

prediccion<-reg[["fitted.values"]]

data2<-data.frame(data,prediccion,res)

data2

#View(data2)

#fitted values son las predicciones y^

#análisis de los residuos

hist(res,15)

mean(res)

#Prueba de normalidad

qqnorm(res)

qqline(res,col="red")

#Gráficas

plot(data[,"Ventas"],data[,"Utilidad"])

plot(res,prediccion)

#---------------Regresion sin B1-----------------------------------------------------------------

# como vimos gracias a las pruebas de hipotesis NO RECHAZAMOS Ho, beta1 no es significativo por lo tanto centro los datos

#Centramos los datos

media\_u<-mean(data[,"Utilidad"])

media\_v<-mean(data[,"Ventas"])

utilidad\_1<-data[,"Utilidad"]-media\_u

ventas\_1<-data[,"Ventas"]-media\_v

data1<-data.frame (Utilidad=utilidad\_1,Ventas=ventas\_1)

#View(data1)

str(data)

#Regresión

reg1<-lm(Utilidad~Ventas,data1)

summary(reg1)

#B1=0 Y B2=0.4399

#Y=B2X

str(reg1)

anova1<-aov(reg1)

summary(anova1)

#Ho:B2=0 Y H1:B2!=0

#Rechazo Ho si F>F(1,n-2\_)(alfa) a favor de H1:B2!=0

#F=559

#Valor percentil

qt(0.975 , df =38)

qf(0.95 , df1=1,df2=38)

#F(1,n-2)(alfa)=4.0987

#Por lo tanto se rechaza Ho, así que se tiene que B2 es significativo, es decir la utilidad depende linealmmente de las Ventas

#Intervalos de confianza

confint(reg1,level=0.95)

names(reg1)

res1<-reg1[["residuals"]]

prediccion1<-reg[["fitted.values"]]

data1\_2<-data.frame(data1,Predicciones=prediccion1,Residuos=res1)

View(data1\_2)

#análisis de los residuos

hist(res1,15) #se asemeja a la gráfica de una distribución normal

mean(res1)

#Prueba de normalidad

qqnorm(res1)

qqline(res1,col="red")

#Gráficas

plot(data1[,"Ventas"],data1[,"Utilidad"])#se ve que tiene tendencia lineal

plot(res1,prediccion1)#no existe evidencia de violación de hipótesis

Corriendo el código se obtuvo :

> data <- read.table("data\_rls\_uti.txt", header = TRUE, dec=",", sep="\t")

> View(data)

> str(data)

'data.frame': 40 obs. of 2 variables:

$ Utilidad: int 6017 8049 8551 6720 7391 8045 5814 4954 5564 7160 ...

$ Ventas : int 13270 17127 17814 16000 18026 17877 13214 10321 12365 15235 ...

> data

Utilidad Ventas

1 6017 13270

2 8049 17127

3 8551 17814

4 6720 16000

5 7391 18026

6 8045 17877

7 5814 13214

8 4954 10321

9 5564 12365

10 7160 15235

11 7345 15301

12 8333 19841

13 5241 12782

14 4606 10236

15 6744 15326

16 5274 10987

17 8331 18512

18 4710 10020

19 5297 11036

20 6440 15333

21 7333 17885

22 8306 18458

23 6781 15412

24 7254 15112

25 5969 13265

26 5008 10654

27 8979 18706

28 6437 15327

29 7672 18103

30 8146 18102

31 4664 10600

32 8151 16981

33 5206 11568

34 9088 19712

35 5380 11225

36 4437 10564

37 4552 11103

38 8093 17984

39 6495 15104

40 8812 18749

> #Realizamos una regresiÃ³n lineal de los datos

> reg<-lm(Utilidad~Ventas,data)

> summary(reg)

Call:

lm(formula = Utilidad ~ Ventas, data = data)

Residuals:

Min 1Q Median 3Q Max

-676.35 -302.04 42.59 303.67 612.49

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 137.08270 282.69543 0.485 0.631

Ventas 0.43994 0.01859 23.663 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 367.4 on 38 degrees of freedom

Multiple R-squared: 0.9364, Adjusted R-squared: 0.9348

F-statistic: 559.9 on 1 and 38 DF, p-value: < 2.2e-16

> #las ventas aumentan en una unidad entonces la utilidad aumenta 0.43

> #B1=137.08 Y B2=0.43

> #Y=B1+B2X

> #RAZON T t1=B1^/ee(B1^)=0.485

> # |t1|>tn-2(alfa/2)=t(40-2)(0.025\_)=2.024394

> qt(0.975,df=38) # valor percentil t student

[1] 2.024394

> #en este caso 0.48> 2.024 no se cumple, no rechazamos Ho,asi que puede centrar los datos ya que beta1 no es significativo

> #valor p<0.05 rechazo la hipotesis nula valor p=0.63 solo rechazamos

> #Son t2=23.663 aqui si rechazo ya que 23.663>2.024 , rechazo Ho a un 95% de confianza la variable ventas influye en la utilidad

> #R^2 el 93% de los dtos son explicados por la regresion

> str(reg)

List of 12

$ coefficients : Named num [1:2] 137.08 0.44

..- attr(\*, "names")= chr [1:2] "(Intercept)" "Ventas"

$ residuals : Named num [1:40] 42 377 577 -456 -676 ...

..- attr(\*, "names")= chr [1:40] "1" "2" "3" "4" ...

$ effects : Named num [1:40] -42272 -8694 518 -484 -738 ...

..- attr(\*, "names")= chr [1:40] "(Intercept)" "Ventas" "" "" ...

$ rank : int 2

$ fitted.values: Named num [1:40] 5975 7672 7974 7176 8067 ...

..- attr(\*, "names")= chr [1:40] "1" "2" "3" "4" ...

$ assign : int [1:2] 0 1

$ qr :List of 5

..$ qr : num [1:40, 1:2] -6.325 0.158 0.158 0.158 0.158 ...

.. ..- attr(\*, "dimnames")=List of 2

.. .. ..$ : chr [1:40] "1" "2" "3" "4" ...

.. .. ..$ : chr [1:2] "(Intercept)" "Ventas"

.. ..- attr(\*, "assign")= int [1:2] 0 1

..$ qraux: num [1:2] 1.16 1.12

..$ pivot: int [1:2] 1 2

..$ tol : num 1e-07

..$ rank : int 2

..- attr(\*, "class")= chr "qr"

$ df.residual : int 38

$ xlevels : Named list()

$ call : language lm(formula = Utilidad ~ Ventas, data = data)

$ terms :Classes 'terms', 'formula' length 3 Utilidad ~ Ventas

.. ..- attr(\*, "variables")= language list(Utilidad, Ventas)

.. ..- attr(\*, "factors")= int [1:2, 1] 0 1

.. .. ..- attr(\*, "dimnames")=List of 2

.. .. .. ..$ : chr [1:2] "Utilidad" "Ventas"

.. .. .. ..$ : chr "Ventas"

.. ..- attr(\*, "term.labels")= chr "Ventas"

.. ..- attr(\*, "order")= int 1

.. ..- attr(\*, "intercept")= int 1

.. ..- attr(\*, "response")= int 1

.. ..- attr(\*, ".Environment")=<environment: R\_GlobalEnv>

.. ..- attr(\*, "predvars")= language list(Utilidad, Ventas)

.. ..- attr(\*, "dataClasses")= Named chr [1:2] "numeric" "numeric"

.. .. ..- attr(\*, "names")= chr [1:2] "Utilidad" "Ventas"

$ model :'data.frame': 40 obs. of 2 variables:

..$ Utilidad: int [1:40] 6017 8049 8551 6720 7391 8045 5814 4954 5564 7160 ...

..$ Ventas : int [1:40] 13270 17127 17814 16000 18026 17877 13214 10321 12365 15235 ...

..- attr(\*, "terms")=Classes 'terms', 'formula' length 3 Utilidad ~ Ventas

.. .. ..- attr(\*, "variables")= language list(Utilidad, Ventas)

.. .. ..- attr(\*, "factors")= int [1:2, 1] 0 1

.. .. .. ..- attr(\*, "dimnames")=List of 2

.. .. .. .. ..$ : chr [1:2] "Utilidad" "Ventas"

.. .. .. .. ..$ : chr "Ventas"

.. .. ..- attr(\*, "term.labels")= chr "Ventas"

.. .. ..- attr(\*, "order")= int 1

.. .. ..- attr(\*, "intercept")= int 1

.. .. ..- attr(\*, "response")= int 1

.. .. ..- attr(\*, ".Environment")=<environment: R\_GlobalEnv>

.. .. ..- attr(\*, "predvars")= language list(Utilidad, Ventas)

.. .. ..- attr(\*, "dataClasses")= Named chr [1:2] "numeric" "numeric"

.. .. .. ..- attr(\*, "names")= chr [1:2] "Utilidad" "Ventas"

- attr(\*, "class")= chr "lm"

> anova<-aov(reg)

> summary(anova)

Df Sum Sq Mean Sq F value Pr(>F)

Ventas 1 75578286 75578286 559.9 <2e-16 \*\*\*

Residuals 38 5129142 134977

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #df grados de libertad

> #Ho:B2=0 Y H1:B2!=0

> #Ho:B2=B3=...

> #RECHAZO H0 si F>F(1,n-2\_)(alfa) a favor de H1:B2!=0

> #F=559

> qf(0.95,df1=1,df2=38) #valor percentil F fisher

[1] 4.098172

> #4.098172

> #RECHAZO Ho Y CONCLUIMOS QUE LA REGRESION ES SIGNIFICATIVA

> #intervalos de confianza

> confint(reg,level=0.95)

2.5 % 97.5 %

(Intercept) -435.2042893 709.3696869

Ventas 0.4022981 0.4775722

> #solo analizando el intervalo de confianza rechazamos la hipotesis nula de beta 2, a un 95% de confianza

> names(reg)

[1] "coefficients" "residuals" "effects" "rank" "fitted.values" "assign"

[7] "qr" "df.residual" "xlevels" "call" "terms" "model"

> str(reg[["residuals"]])

Named num [1:40] 42 377 577 -456 -676 ...

- attr(\*, "names")= chr [1:40] "1" "2" "3" "4" ...

> res<-reg[["residuals"]]

> prediccion<-reg[["fitted.values"]]

> data2<-data.frame(data,prediccion,res)

> data2

Utilidad Ventas prediccion res

1 6017 13270 5975.022 41.977567

2 8049 17127 7671.852 377.147608

3 8551 17814 7974.088 576.912145

4 6720 16000 7176.045 -456.045453

5 7391 18026 8067.354 -676.354111

6 8045 17877 8001.804 43.196229

7 5814 13214 5950.386 -136.386063

8 4954 10321 4677.654 276.346390

9 5564 12365 5576.881 -12.881102

10 7160 15235 6839.495 320.504954

11 7345 15301 6868.531 476.469233

12 8333 19841 8865.836 -532.836449

13 5241 12782 5760.334 -519.334069

14 4606 10236 4640.259 -34.259121

15 6744 15326 6879.529 -135.529147

16 5274 10987 4970.650 303.349565

17 8331 18512 8281.163 49.837395

18 4710 10020 4545.233 164.766877

19 5297 11036 4992.207 304.792742

20 6440 15333 6882.609 -442.608693

21 7333 17885 8005.323 -672.323252

22 8306 18458 8257.406 48.593894

23 6781 15412 6917.364 -136.363572

24 7254 15112 6785.383 468.616980

25 5969 13265 5972.823 -3.822757

26 5008 10654 4824.152 183.847977

27 8979 18706 8366.510 612.489971

28 6437 15327 6879.969 -442.969082

29 7672 18103 8101.229 -429.229120

30 8146 18102 8100.789 45.210815

31 4664 10600 4800.396 -136.395523

32 8151 16981 7607.622 543.378143

33 5206 11568 5226.253 -20.252770

34 9088 19712 8809.085 278.915188

35 5380 11225 5075.355 304.644994

36 4437 10564 4784.558 -347.557857

37 4552 11103 5021.683 -469.682915

38 8093 17984 8048.877 44.123166

39 6495 15104 6781.864 -286.863539

40 8812 18749 8385.427 426.572759

> #View(data2)

> #fitted values son las predicciones y^

> #anÃ¡lisis de los residuos

> hist(res,15)

> mean(res)

[1] -8.14071e-15

> #Prueba de normalidad

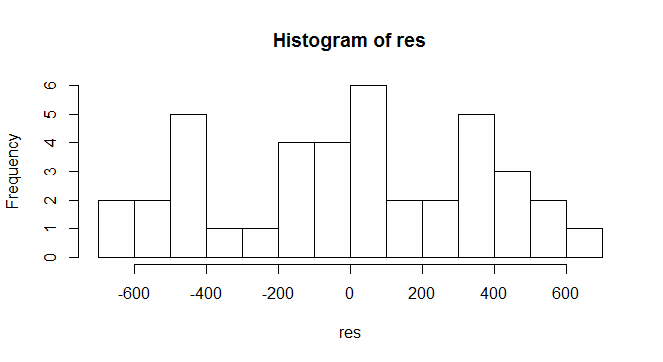
> qqnorm(res)

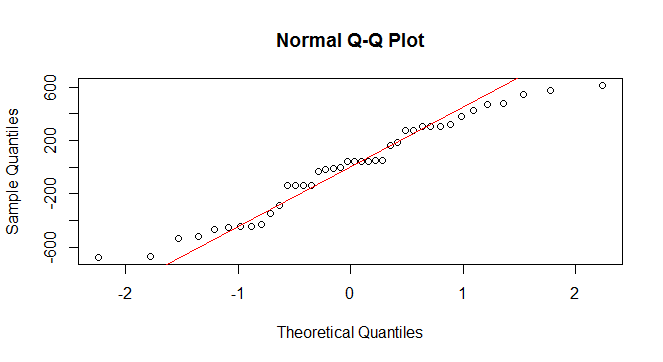
> qqline(res,col="red")

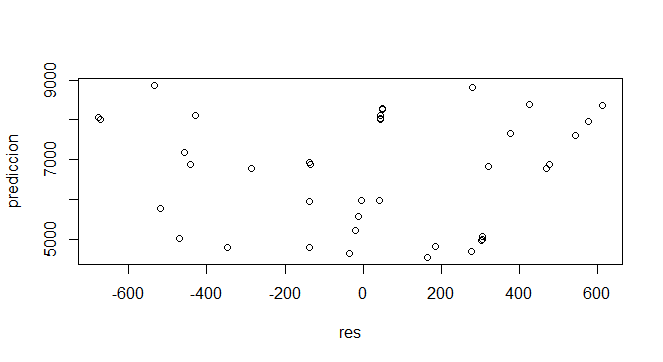
> #GrÃ¡ficas

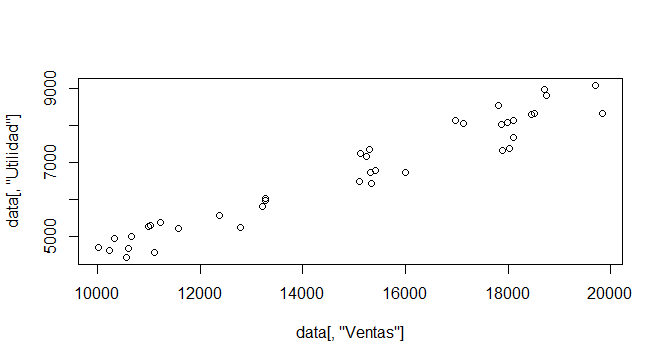
> plot(data[,"Ventas"],data[,"Utilidad"])

> plot(res,prediccion)









> #---------------Regresion sin B1-----------------------------------------------------------------

> # como vimos gracias a las pruebas de hipotesis NO RECHAZAMOS Ho, beta1 no es significativo por lo tanto centro los datos

> #Centramos los datos

> media\_u<-mean(data[,"Utilidad"])

> media\_v<-mean(data[,"Ventas"])

> utilidad\_1<-data[,"Utilidad"]-media\_u

> ventas\_1<-data[,"Ventas"]-media\_v

> data1<-data.frame (Utilidad=utilidad\_1,Ventas=ventas\_1)

> #View(data1)

> str(data)

'data.frame': 40 obs. of 2 variables:

$ Utilidad: int 6017 8049 8551 6720 7391 8045 5814 4954 5564 7160 ...

$ Ventas : int 13270 17127 17814 16000 18026 17877 13214 10321 12365 15235 ...

> #RegresiÃ³n

> reg1<-lm(Utilidad~Ventas,data1)

> summary(reg1)

Call:

lm(formula = Utilidad ~ Ventas, data = data1)

Residuals:

Min 1Q Median 3Q Max

-676.35 -302.04 42.59 303.67 612.49

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -6.442e-13 5.809e+01 0.00 1

Ventas 4.399e-01 1.859e-02 23.66 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 367.4 on 38 degrees of freedom

Multiple R-squared: 0.9364, Adjusted R-squared: 0.9348

F-statistic: 559.9 on 1 and 38 DF, p-value: < 2.2e-16

> #B1=0 Y B2=0.4399

> #Y=B2X

> str(reg1)

List of 12

$ coefficients : Named num [1:2] -6.44e-13 4.40e-01

..- attr(\*, "names")= chr [1:2] "(Intercept)" "Ventas"

$ residuals : Named num [1:40] 42 377 577 -456 -676 ...

..- attr(\*, "names")= chr [1:40] "1" "2" "3" "4" ...

$ effects : Named num [1:40] 2.27e-12 -8.69e+03 5.18e+02 -4.84e+02 -7.38e+02 ...

..- attr(\*, "names")= chr [1:40] "(Intercept)" "Ventas" "" "" ...

$ rank : int 2

$ fitted.values: Named num [1:40] -709 988 1290 492 1384 ...

..- attr(\*, "names")= chr [1:40] "1" "2" "3" "4" ...

$ assign : int [1:2] 0 1

$ qr :List of 5

..$ qr : num [1:40, 1:2] -6.325 0.158 0.158 0.158 0.158 ...

.. ..- attr(\*, "dimnames")=List of 2

.. .. ..$ : chr [1:40] "1" "2" "3" "4" ...

.. .. ..$ : chr [1:2] "(Intercept)" "Ventas"

.. ..- attr(\*, "assign")= int [1:2] 0 1

..$ qraux: num [1:2] 1.16 1.12

..$ pivot: int [1:2] 1 2

..$ tol : num 1e-07

..$ rank : int 2

..- attr(\*, "class")= chr "qr"

$ df.residual : int 38

$ xlevels : Named list()

$ call : language lm(formula = Utilidad ~ Ventas, data = data1)

$ terms :Classes 'terms', 'formula' length 3 Utilidad ~ Ventas

.. ..- attr(\*, "variables")= language list(Utilidad, Ventas)

.. ..- attr(\*, "factors")= int [1:2, 1] 0 1

.. .. ..- attr(\*, "dimnames")=List of 2

.. .. .. ..$ : chr [1:2] "Utilidad" "Ventas"

.. .. .. ..$ : chr "Ventas"

.. ..- attr(\*, "term.labels")= chr "Ventas"

.. ..- attr(\*, "order")= int 1

.. ..- attr(\*, "intercept")= int 1

.. ..- attr(\*, "response")= int 1

.. ..- attr(\*, ".Environment")=<environment: R\_GlobalEnv>

.. ..- attr(\*, "predvars")= language list(Utilidad, Ventas)

.. ..- attr(\*, "dataClasses")= Named chr [1:2] "numeric" "numeric"

.. .. ..- attr(\*, "names")= chr [1:2] "Utilidad" "Ventas"

$ model :'data.frame': 40 obs. of 2 variables:

..$ Utilidad: num [1:40] -666.7 1365.3 1867.3 36.3 707.3 ...

..$ Ventas : num [1:40] -1611 2246 2933 1119 3145 ...

..- attr(\*, "terms")=Classes 'terms', 'formula' length 3 Utilidad ~ Ventas

.. .. ..- attr(\*, "variables")= language list(Utilidad, Ventas)

.. .. ..- attr(\*, "factors")= int [1:2, 1] 0 1

.. .. .. ..- attr(\*, "dimnames")=List of 2

.. .. .. .. ..$ : chr [1:2] "Utilidad" "Ventas"

.. .. .. .. ..$ : chr "Ventas"

.. .. ..- attr(\*, "term.labels")= chr "Ventas"

.. .. ..- attr(\*, "order")= int 1

.. .. ..- attr(\*, "intercept")= int 1

.. .. ..- attr(\*, "response")= int 1

.. .. ..- attr(\*, ".Environment")=<environment: R\_GlobalEnv>

.. .. ..- attr(\*, "predvars")= language list(Utilidad, Ventas)

.. .. ..- attr(\*, "dataClasses")= Named chr [1:2] "numeric" "numeric"

.. .. .. ..- attr(\*, "names")= chr [1:2] "Utilidad" "Ventas"

- attr(\*, "class")= chr "lm"

> anova1<-aov(reg1)

> summary(anova1)

Df Sum Sq Mean Sq F value Pr(>F)

Ventas 1 75578286 75578286 559.9 <2e-16 \*\*\*

Residuals 38 5129142 134977

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #Ho:B2=0 Y H1:B2!=0

> #Rechazo Ho si F>F(1,n-2\_)(alfa) a favor de H1:B2!=0

> #F=559

> #Valor percentil

> qt(0.975 , df =38)

[1] 2.024394

> qf(0.95 , df1=1,df2=38)

[1] 4.098172

> #F(1,n-2)(alfa)=4.0987

> #Por lo tanto se rechaza Ho, asÃ� que se tiene que B2 es significativo, es decir la utilidad depende linealmmente de las Ventas

> #Intervalos de confianza

> confint(reg1,level=0.95)

2.5 % 97.5 %

(Intercept) -117.5968432 117.5968432

Ventas 0.4022981 0.4775722

> names(reg1)

[1] "coefficients" "residuals" "effects" "rank" "fitted.values" "assign"

[7] "qr" "df.residual" "xlevels" "call" "terms" "model"

> res1<-reg1[["residuals"]]

> prediccion1<-reg[["fitted.values"]]

> data1\_2<-data.frame(data1,Predicciones=prediccion1,Residuos=res1)

> View(data1\_2)

> #anÃ¡lisis de los residuos

> hist(res1,15) #se asemeja a la grÃ¡fica de una distribuciÃ³n normal

> mean(res1)

[1] -7.771561e-16

> #Prueba de normalidad

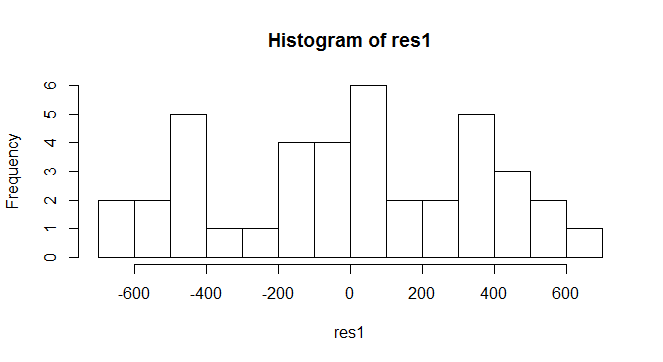
> qqnorm(res1)

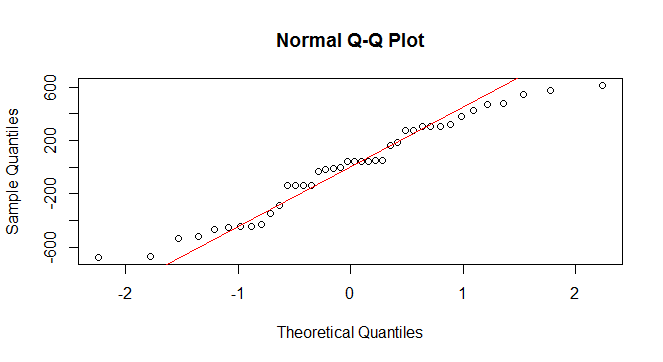
> qqline(res1,col="red")

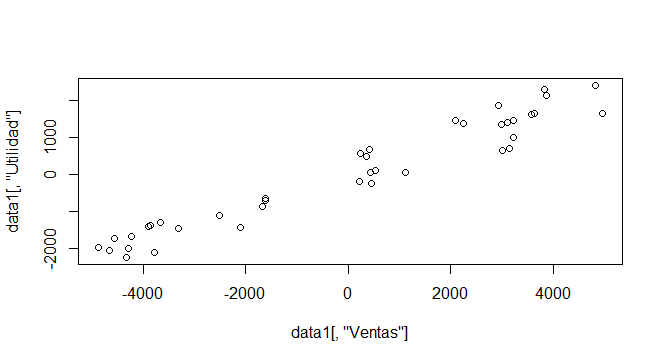
> #GrÃ¡ficas

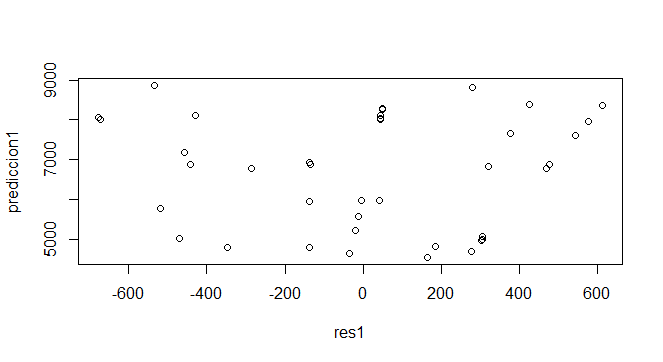
> plot(data1[,"Ventas"],data1[,"Utilidad"])#se ve que tiene tendencia lineal

> plot(res1,prediccion1)#no existe evidencia de violaciÃ³n de hipÃ³tesis









1. **Conclusiones**

* Realizamos el análisis de regresión obteniendo: **U=137.08+0.43V**
* De la recta obtenida se tiene que si las ventas aumentan en una unidad entonces la utilidad aumenta 0.43.
* Gracias a las pruebas de hipótesis con la t-student y Fisher, se obtuvo que se debían centrar los datos (No se rechaza Ho), además de que es significativo y la recta de regresión fue: **U=0.4399V**
* Vemos que la gráfica de los histogramas para ambos casos se tiene una tendencia a una distribución normal.
* Los gráficos de los datos dados y para los centrados, de Utilidad vs Ventas, se ve que se tiene una tendencia lineal.
* La gráfica de los Y (valores predichos) vs res (residuos), para los dos casos no existe evidencia de violación de hipótesis.