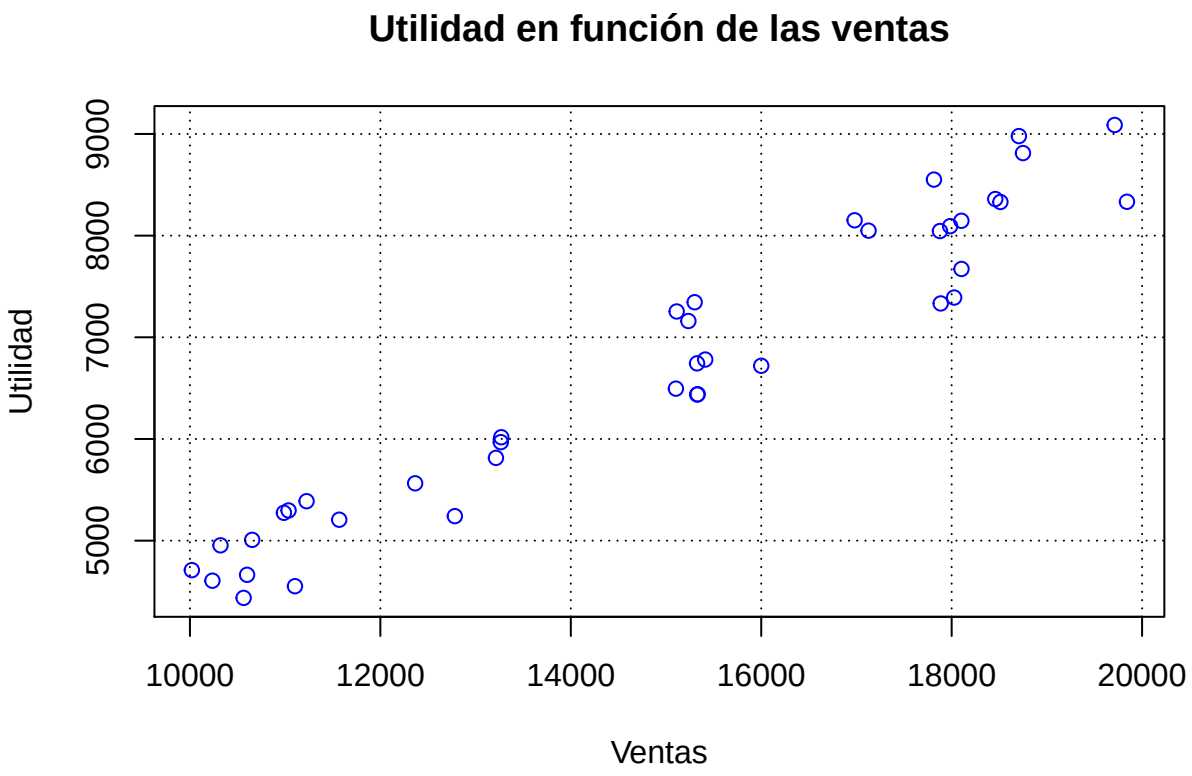
Repaso Modelos Lineales

Boris Polanco

Lunes, 20 de julio, 2015

Utilidad neta en función de las ventas

se cuentan con datos mensuales desde Enero del 2001 hasta Abril del 2004. Se tiene los datos siguientes: Realizamos un gráfico de dispersión



Para encontrar la ecuación de la recta de regresión vamos a utilizar la función *lm()*. Se obtienen los siguientes resultados:

```
##
## Call:
## lm(formula = datos$U ~ datos$V)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -679.22 -303.20   40.75  303.79  609.33
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 132.38632  283.05174   0.468    0.643
```

```
## datos$V       0.44035    0.01862  23.656   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 367.9 on 38 degrees of freedom
## Multiple R-squared:  0.9364, Adjusted R-squared:  0.9347
## F-statistic: 559.6 on 1 and 38 DF,  p-value: < 2.2e-16
```

La ecuación de la recta de regresión viene dada entonces por la expresión

$$U = 137.09 + 0.44V \tag{1}$$

Dado que las ventas como las utilidades están expresadas en dólares, diremos que si las ventas se incrementan en un dólar, la utilidad neta se incrementa en promedio 44 centavos, también se puede decir que la utilidad media es el 44 por ciento de las ventas.

Verifiquemos que los estimadores encontrados $\hat{\beta}_1$ y $\hat{\beta}_2$, son no correlacionados, para ello calculamos la covarianza entre ambos estimadores. la cual viene dada por:

$$cov(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\sigma^2}{S_{xx}} \tag{2}$$

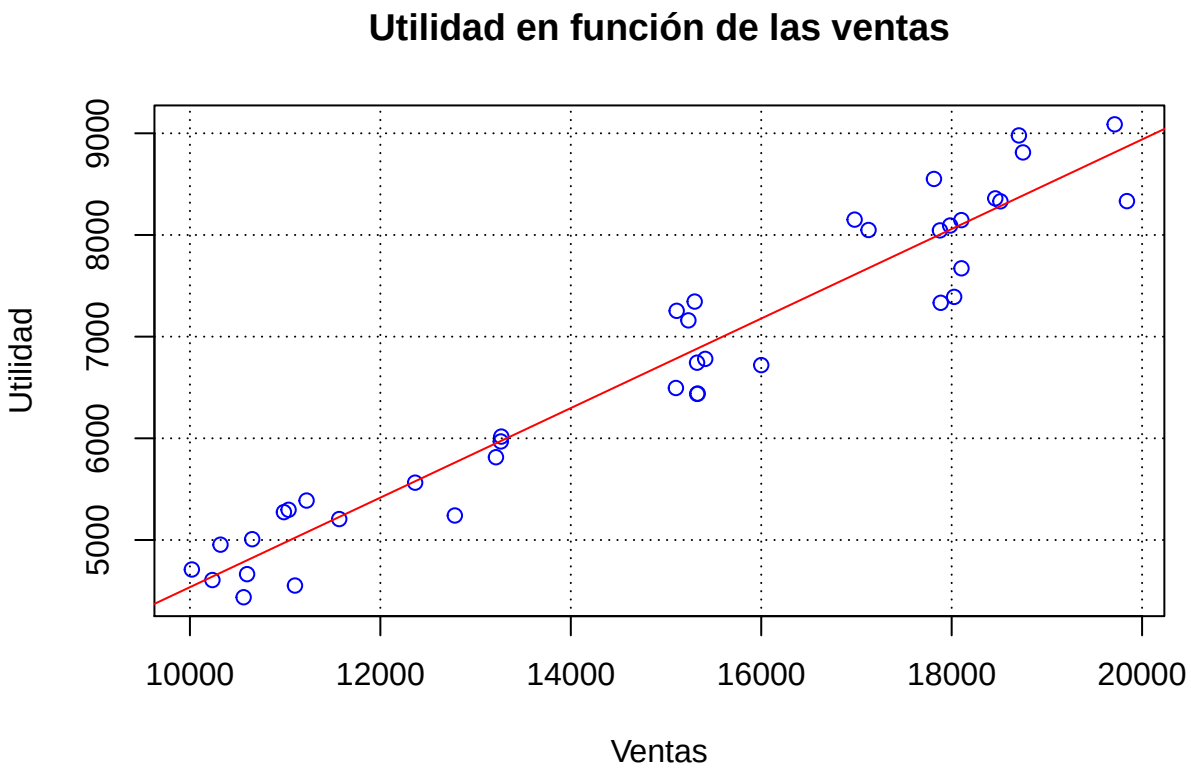
con $S_{xx} = \sum (x_i - \hat{x})^2$.

Intervalos de confianza

Los intervalos de confianza con nivel $1 - \alpha$ para los parámetros β_i vienen dados por la siguiente expresión:

$$\hat{\beta}_i \pm ee(\hat{\beta}_i) t_{n-2}(\alpha/2) \tag{3}$$

donde $t_{n-2}(\alpha/2)$ es el fractil de orden $1 - \alpha/2$ de la ley de student con $n - 2$ grados de libertad. Para nuestro caso utilizando R, podemos encontrar estos intervalos utilizando la función *confint()*.



Coeficiente de determinación

Es un buen indicador de la calidad de la regresión, pero no es determinante ni suficiente para decidir sobre la adecuación del modelo. Su uso es muy difundido pero en general inapropiado porque no tienen presentes las limitaciones y alcances de este indicador. Además no se estudian los residuos, como se indica más adelante.

Coeficiente R^2

Una regresión será buena si la variabilidad explicada por la regresión es relativamente alta con respecto a la variabilidad total de Y, es decir si $SEC \approx STC$, donde $SEC = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ y $STC = \sum_{i=1}^{n}(y_i - \bar{y})^2$.

Definición

Se denomina coeficiente de determinación R^2 al cociente

$$R^2 = \frac{SEC}{STC} \tag{4}$$

Para nuestro caso lo calculamos en R de la siguiente manera:

```
## [1] 0.9364113
```

Para interpretar este coeficiente lo multiplicamos por 100 y lo interpretamos como porcentaje, que es el porcentaje de variabilidad explicada por la regresión con respecto a la variabilidad total. Por lo que el modelo propuesto explica aproximadamente el 94% de la variabilidad de Y.

Gráficos de residuos

No se pueden conocer los errores u_i, pero si se pueden calcular los residuos $\hat{u_i} = y_i - \hat{y_i}$, estos no son estimadores de los errores pero si dan información sobre ellos. Las hipótesis asumidas: linealidad, igualdad de varianzas, no correlación y normalidad deben reflejarse en los gráficos de los residuos. Los residuos pueden indicar

- La función de regresión no es lineal.
- La varianza de los errores no es constante.
- Los errores son correlacionados.
- Existen observaciones singulares.
- Los errores no son normalmente distribuídos.
- Uno o varios regresores han sido omitidos del modelo

Residuos en función de: $\hat{Y}$ y de X

Se grafica los puntos $(\hat{y_i}, \hat{u_i})|i = 1, 2..., n$ y también $(x_i, \hat{u_i})|i = 1, 2..., n$. Si los puntos están dentro de una franja horizontal simétrica al eje X no habrá evidencia de violación de hipótesis.