

## Informe Tarea N° 1 Estadística Avanzada

### Rendimiento Vehicular.

El problema a analizar estadísticamente es la relación existente entre ciertas variables predictoras como los son: Capacidad en volumen de vehículo (vol), Caballos de fuerza del motor (hp), Velocidad máxima (sp), Peso del vehículo (wt) y el rendimiento promedio de millas por galón (mpg) ,que será nuestra variable respuesta. Esta información a estudiar fue obtenida de la EPA (U.S. Environmental Protection Agency) y consta de 82 vehículos de diferentes modelos y marcas.

Es apropiado decir que los datos son analizados bajo el programa estadístico R, por ser este de libre distribución.

A modo de conclusión y sin profundizar en detalles, los cuales se explicaran mas adelante, podemos decir que el modelo queda explicado de una manera satisfactoria por

$$mpg = 192.4377 + 0.4052 \text{ hp} - 1.32 \text{ sp} - 1.9221 \text{ wt} + e$$

Con un coeficiente de determinación del  $R^2 = 87,25\%$  , lo cual dice que el ajuste del modelo es bastante bueno, es decir , que el 87,25% de la variación del millaje promedio por galón es explicada por su relación lineal con las variables predictoras. Con un coeficiente de determinación ajustado igual a 86,76% .Por otro lado con una desviación estándar del error igual a 3.64.

A continuación se junta la tabla summary entregada por R. Donde se puede apreciar

los p-valúes de la tabla  $t$  y donde todos son menores a 0.05.

```
> summary(reg)
```

Call:

```
lm(formula = mpg ~ hp + sp + wt, family = gaussian)
```

Residuals:

Min 1Q Median 3Q Max

−9.1633 −2.8387 0.2464 1.7889 12.5566

Coefficients:

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

(Intercept) 194.12962 23.32213 8.324 2.22e−12 \*\*\*

hp 0.40518 0.07891 5.135 2.03e−06 \*\*\*

sp −1.32000 0.24118 −5.473 5.19e−07 \*\*\*

wt −1.92210 0.19238 −9.991 1.31e−15 \*\*\*

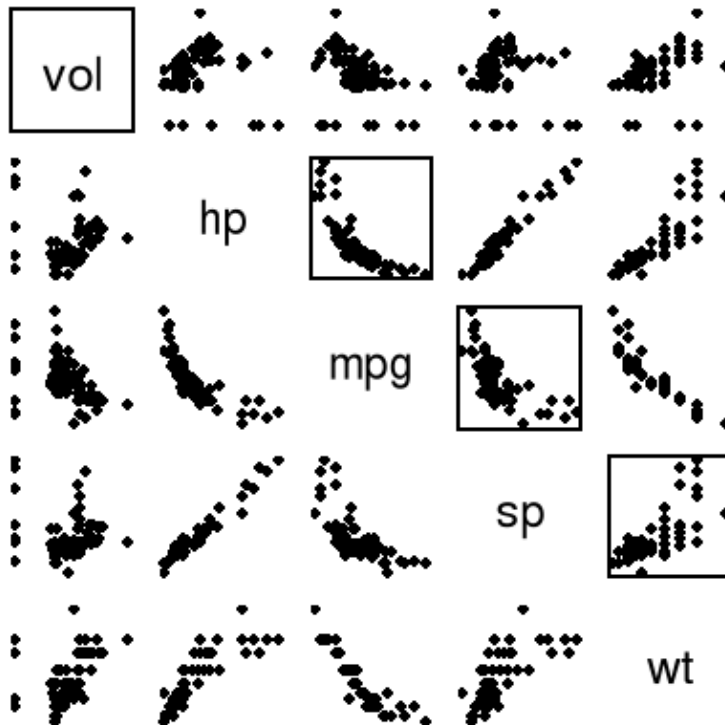
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.64 on 78 degrees of freedom

Multiple R-Squared: 0.8725, Adjusted R-squared: 0.8676

F-statistic: 177.9 on 3 and 78 DF, p-value: < 2.2e-16

Entrando un poco mas en detalle realizamos una exploración de las variables mediante un grafico



Donde es fácil ver las relaciones existente entre la variable **mpg** y las predictoras, notando la baja relación lineal existente entre **vol** y **mpg**, esta apreciación queda confirmada al realizar el ajuste con las variables regresoras:

```
> summary(reg)
```

Call:

```
lm(formula = mpg ~ vol + hp + sp + wt, family = gaussian)
```

Residuals:

Min 1Q Median 3Q Max

-9.0108 -2.7731 0.2733 1.8362 11.9854

Coefficients:

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

(Intercept) 192.43 23.53161 8.178 4.62e-12 \*\*\*

vol -0.01565 0.02283 -0.685 0.495

hp 0.39221 0.08141 4.818 7.13e-06 \*\*\*

sp -1.29482 0.24477 -5.290 1.11e-06 \*\*\*

wt -1.85980 0.21336 -8.717 4.22e-13 \*\*\*

----

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

Residual standard error: 3.653 on 77 degrees of freedom

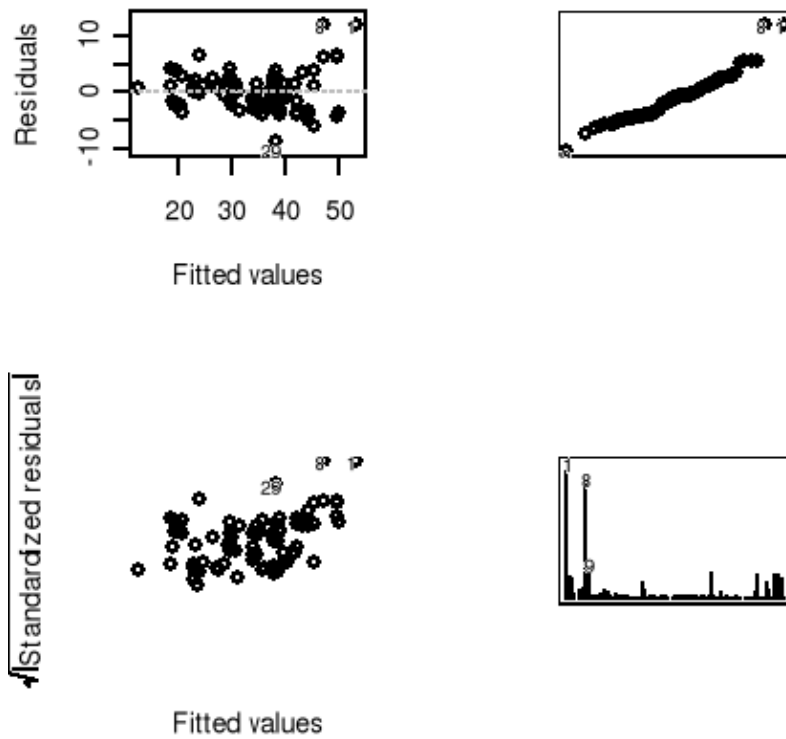
Multiple R-Squared: 0.8733, Adjusted R-squared: 0.8667

F-statistic: 132.7 on 4 and 77 DF, p-value: < 2.2e-16

Notar el p-value de la prueba t de la variable **vol**, superior a 0.05 por lo que esta variable no pacer estar considerada en el modelo final al contrario de las restantes.

Tambien podemos destacar que el coeficiente de determinación es superior el presentado anteriormente, asi como el coeficiente de determinación ajustado tambien es menor habiendo una mayor distancia entre ambos; pero asubes la desviación estimada del error es mayor.

Ademas podemos observar diferentes grafico de residuos los cuales tienes un comportamiento aleatorio:



Es importante destacar que se hicieron otras regresiones bajo diferentes Hipótesis:

como  $H_0: \mathbf{hp, vol=0}$ ,  $H_1: \mathbf{wt, sp=0}$ ,  $H_2: \mathbf{hp, sp=0}$ , siendo en todas estas hipótesis los coeficientes de determinación inferiores en gran magnitud a los ya presentados existiendo gran diferencia entre éste y el coeficiente ajustado, la desviación estimada de los errores son muy superiores a los obtenidos en las dos regresiones presentadas anteriormente.