

Modelos Lineales

Analisis del Comportamiento del Precio de Venta del Vehiculo Mazda 2

Ana María López - Pedro Pablo Villegas

September, 2017

INTRODUCCIÓN

En el presente documento se presenta el resultado del análisis descriptivo realizado a una base de datos con información de carros Mazda 2 en venta. Esta información se tomó directamente desde la página de www.tucarro.com.co, la información de los carros es ingresada por la persona interesada en vender diligenciando un formulario donde se captura los siguientes datos:

Descripción Campo	Campo Obligatorio	Valor Libre
Departamento	Si	No
Ciudad	Si	No
Barrio	No	No
Año	Si	No
Color	Si	No
Dirección	Si	No
Negociable	Si	No
Placa	Si	Si
Recorrido	Si	Si
Transmisión	Si	Si
Airbag	No	No
Aire Acondicionado	No	No
Asientos	No	No
Combustible	No	No
Doy financiamiento	No	No
Frenos ABS	No	No
Motor	No	No
Motor recién reparado	No	No
Nro. de cilindros	No	No
Sonido	No	No
Tracción	No	No
Único dueño	No	No
Versión	No	No
Vidrios	No	No
Alarma con control	No	No
Asegurado	No	No
Rastreo satelital	No	No
Bloqueo central	No	No
Forro del volante	No	No
Forros de asientos	No	No
Volante deportivo	No	No
Caja de CD's	No	No
DVD	No	No
Planta	No	No
Sub-Buffer (Bajos)	No	No
Estribos	No	No
Forro llanta de repuesto	No	No

Descripción Campo	Campo Obligatorio	Valor Libre
Llantas nuevas	No	No
Luces anti niebla	No	No
Película de Seguridad	No	No
Retrovisores eléctricos	No	No
Revisión tecnomecánica	No	No
Rines de lujo	No	No
Spoiler	No	No
Sun Roof (Quemacoco)	No	No
Precio	Si	Si

La información con la que se trabajo se extrajo el día 26 de Agosto de 2017 y consta de 631 carros que se encontraban para ese día a la venta.

Definición de Variables

A continuación se relacionan cada una de las variables a analizar:

- **Precio:** Representa el valor de mercado del vehículo. Variable numérica continua en pesos
- **Modelo / año:** Representa el modelo del vehículo. Variable numérica discreta en años
- **Departamento / Ciudad / Barrio:** Representan la ubicación del vehículo al momento de publicar el vehículo en la página. Variable categórica nominal departamento, ciudad y barrio
- **Color:** Representa el color del vehículo. Variable categórica nominal.
- **Combustible:** Representa el tipo de combustible que usa el vehículo. Variable categórica nominal binomial
- **Recorrido:** Representa los kilómetros que ha recorrido el vehículo. Variable numérica continua.
- **Único dueño:** Da información sobre si el vehículo ha tenido dueños anteriores al que está vendiendo. variable categórica binomial.
- **Versión:** Representa la versión del vehículo. Variable categórica ordinal (puede ser ordinal por las versiones full media y básica).
- **Frenos ABS:** Representa si tiene frenos ABS. Variable categórica binomial.
- **Aire:** Representa si tiene aire acondicionado. Variable categórica binomial.
- **Airbag:** Representa si tiene airbag. Variable categórica binomial.
- **Asientos:** Representa el material de la cojinería tela, cuero, etc. Variable categórica nominal.
- **Cilindros:** Representa el número de cilindros del motor del vehículo. Variable categórica nominal.
- **Dirección:** Representa el tipo de dirección del vehículo. Variable categórica
- **Financiamiento:** Representa si quien vende el vehículo da opciones de financiamiento para el vehículo. Variable categórica binomial
- **Motor:** Representa el tipo de motor que tiene el vehículo. Variable categórica nominal
- **Negociable:** Representa si quien vende el vehículo da opciones de negociación del vehículo. Variable categórica binomial
- **Motor Reparado:** Representa si el motor del vehículo ha sido reparado. Variable categórica binomial
- **Sonido:** Representa el tipo de sonido que tiene el vehículo. Variable categórica
- **Tracción:** Representa el tipo de tracción del motor del vehículo. Variable categórica
- **Transmisión:** Representa tipo de transmisión del vehículo. Variable categórica
- **Vidrios:** Representa si el vehículo tiene vidrios eléctricos Variable categórica
- **Alarma con control:** Representa si el vehículo tiene incorporado algún sistema de alarma o bloqueo central. Variable categórica
- **Asegurado:** Representa si el vehículo se encuentra asegurado al momento de la publicación. Variable categórica binomial
- **Rastreo Satelital:** Representa si el vehículo cuenta con algún sistema de seguimiento satelital. Variable categórica binomial

- **Placa:** Representa la placa del vehículo.

Se cuentan con otras variables que no son obligatorias que indican atributos adicionales de Seguridad, Equipamiento, Sonido y Exterior, estas variables son opcionales se chequean indicando que si cuentan con esa característica, estas son variables categorías binomiales. Se agrupan de la siguiente manera:

- **Seguridad:**
 - Alarma con control
 - Asegurado
 - Rastreo satelital
- **Equipamiento:**
 - Bloqueo central
 - Forro del volante
 - Forros de asientos
 - Volante deportivo
- **Sonido:**
 - Caja de CD's
 - DVD
 - Planta
 - Sub-Buffer (Bajos)
- **Exterior:**
 - Estribos
 - Forro llanta de repuesto
 - Llantas nuevas
 - Luces anti niebla
 - Película de Seguridad
 - Retrovisores eléctricos
 - Revisión tecnomecánica
 - Rines de lujo
 - Spoiler
 - Sun Roof (Quemacoco)

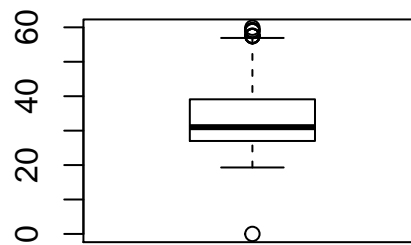
Analisis Descriptivo

Se cuentan con 47 variables, teniendo como variable de interés el Precio del vehículo, de las 46 restantes se seleccionaron: Recorrido y Año del Vehículo. Se pretende identificar que relación se encuentra entre los kilómetros recorridos y el precio, y entre el año del vehículo y el precio.

Variable	Mínimo	Media	Máximo	Q10	Q25	Q50	Q75	Q90
Precio	0.0	34.2	59.9	24.5	27.0	31.0	39.1	49.9
Recorrido	0.0	213.0	100000.0	7.0	28.0	52.0	76.0	96.0
Año	2008	2013	2018	2009	2011	2012	2015	2017

Encontramos en el análisis descriptivo los datos relacionados a vehículos nuevos, estos son los que cuentan con 0KM de Recorrido y son del Año 2018, también para el precio nos encontramos con vehículos donde no se definió el Precio quedando este en 0, se realiza un gráfico de caja para identificar si en el precio se tienen valores atípicos:

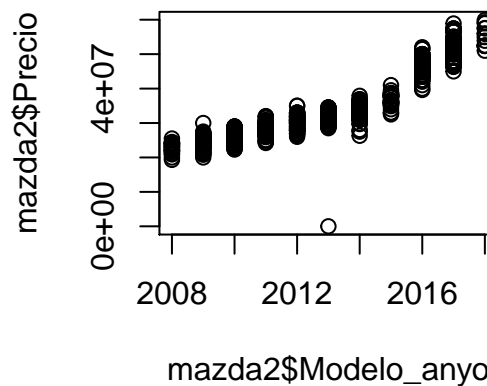
```
boxplot(mazda2$Precio/100000)
```



Se evidencia en el diagrama de cajas que se tiene un atípico con precio cero y adicional se marcan algunos atípicos aproximados a los 60 Millones.

Ahora se desea visualizar que relación se evidencia del Precio vs el Año:

```
plot(mazda2$Modelo_ano,mazda2$Precio)
```



De la gráfica podemos deducir que existe una relación lineal positiva entre el Precio y el Año del vehículo, por lo cual calculamos la Covarianza y la Varianza para determinar que la deducción es correcta:

- Covarianza

```
cov(mazda2$Modelo_ano,mazda2$Precio)
```

```
## [1] 25071705
```

De la covarianza obtenemos un valor Mayor a Cero, lo cual nos indica que existe una asociación lineal positiva.

- Correlación

```
cor(mazda2$Modelo_ano,mazda2$Precio)
```

```
## [1] 0.9213659
```

La correlación nos indica que existe una asociación lineal positiva imperfecta pero por su cercanía a 1 podríamos decir que es una asociación positiva fuerte. Esto quiere decir que entre mas nuevo es el modelo del carro su precio es mas alto.

En el recorrido se evidencian problemas de datos atípicos, en el análisis descriptivo nos muestra un valor de 100000Km el cual se omitirá en los siguientes gráficos, para determinar si aún contamos con datos atípicos se realiza el diagrama de cajas para esta variable:

Diagrama de Cajas teniendo en cuenta el valor atípico identificado en el análisis descriptivo:

```
boxplot(Var$Recorrido)
```

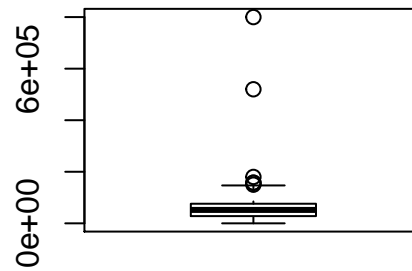
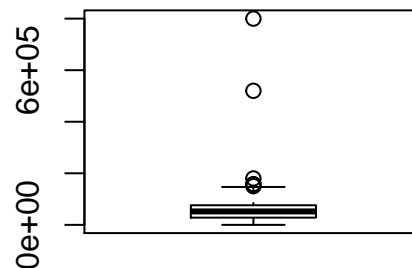


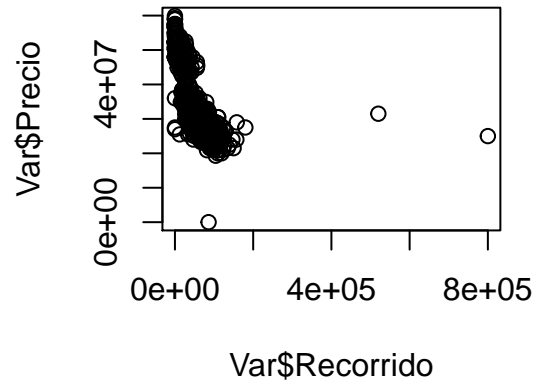
Diagrama de Cajas omitiendo el valor atípico:

```
boxplot(Var$Recorrido)
```



Ahora se desea visualizar que relación se evidencia del Precio vs el Recorrido:

```
plot(Var$Recorrido,Var$Precio)
```



En la gráfica se visualiza que entre menos kilómetros se tengan el valor del precio es mayor, pareciera que se tiene una asociación lineal negativa, para esto se realizará el calculo de la covarianza y correlación para validarlo:

- Covarianza

```
cov(Var$Recorrido,Var$Precio)
```

```
## [1] -275477633776
```

De la covarianza obtenemos un valor Menor a Cero, lo cual nos indica que existe una asociación lineal negativa.

- Correlación

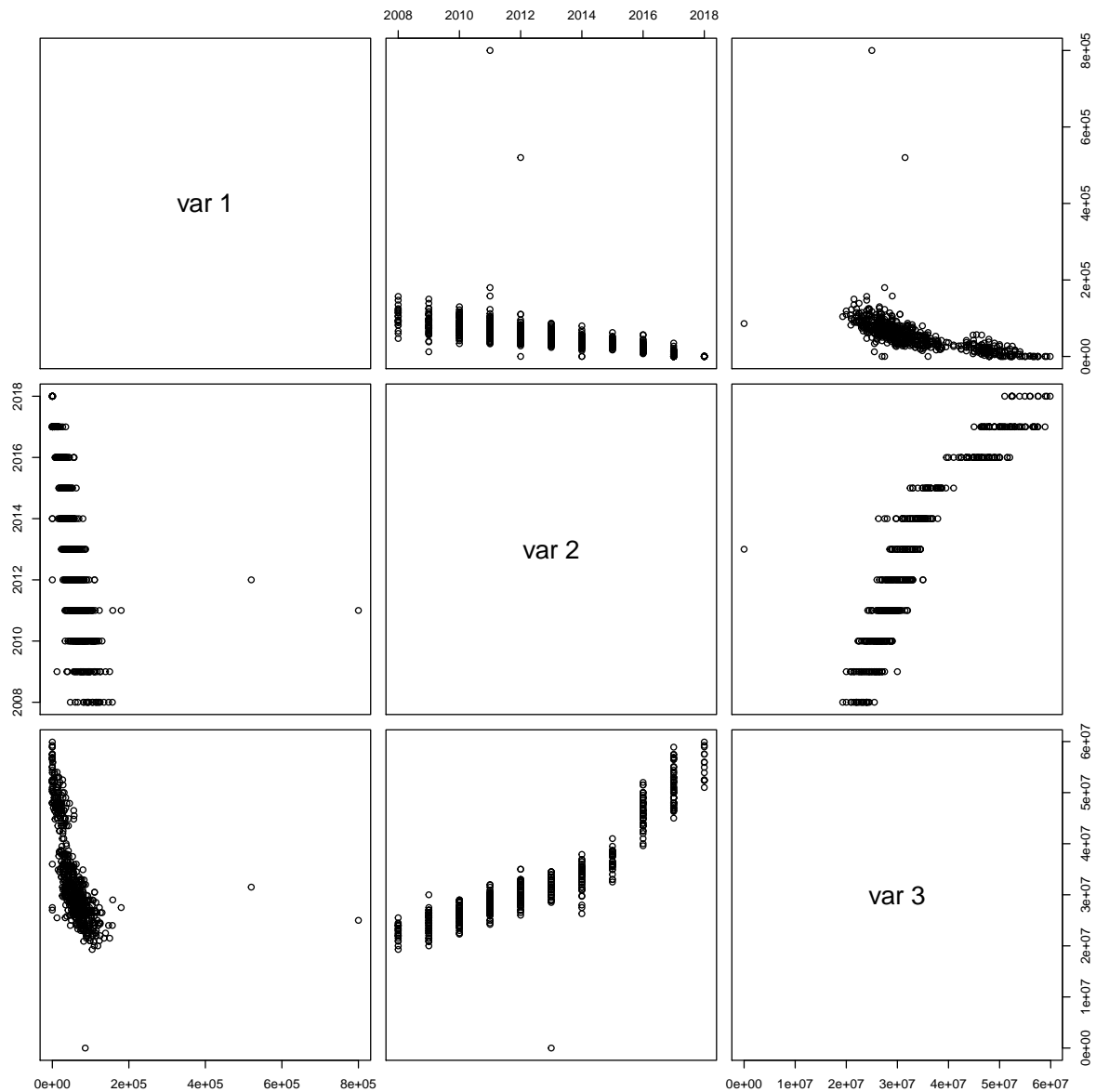
```
cor(Var$Recorrido,Var$Precio)
```

```
## [1] -0.5852163
```

La correlación nos indica que existe una asociación lineal negativa imperfecta.

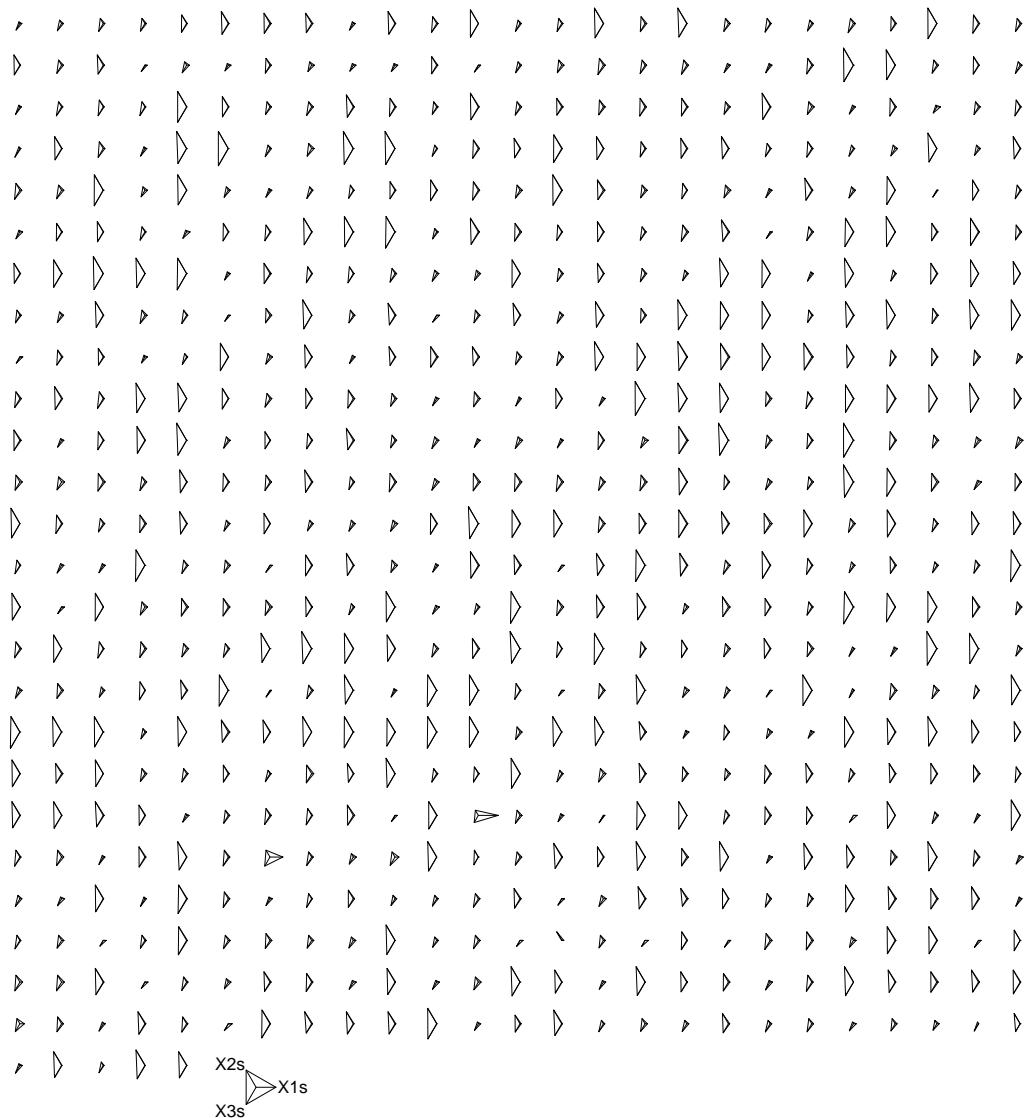
Continuando con el análisis de las variables a analizar, se realiza los siguientes gráficos, matriz de dispersión:

```
x <- as.matrix(cbind(Var$Recorrido,Var$Modelo_anyo,Var$Precio))  
pairs(x)
```



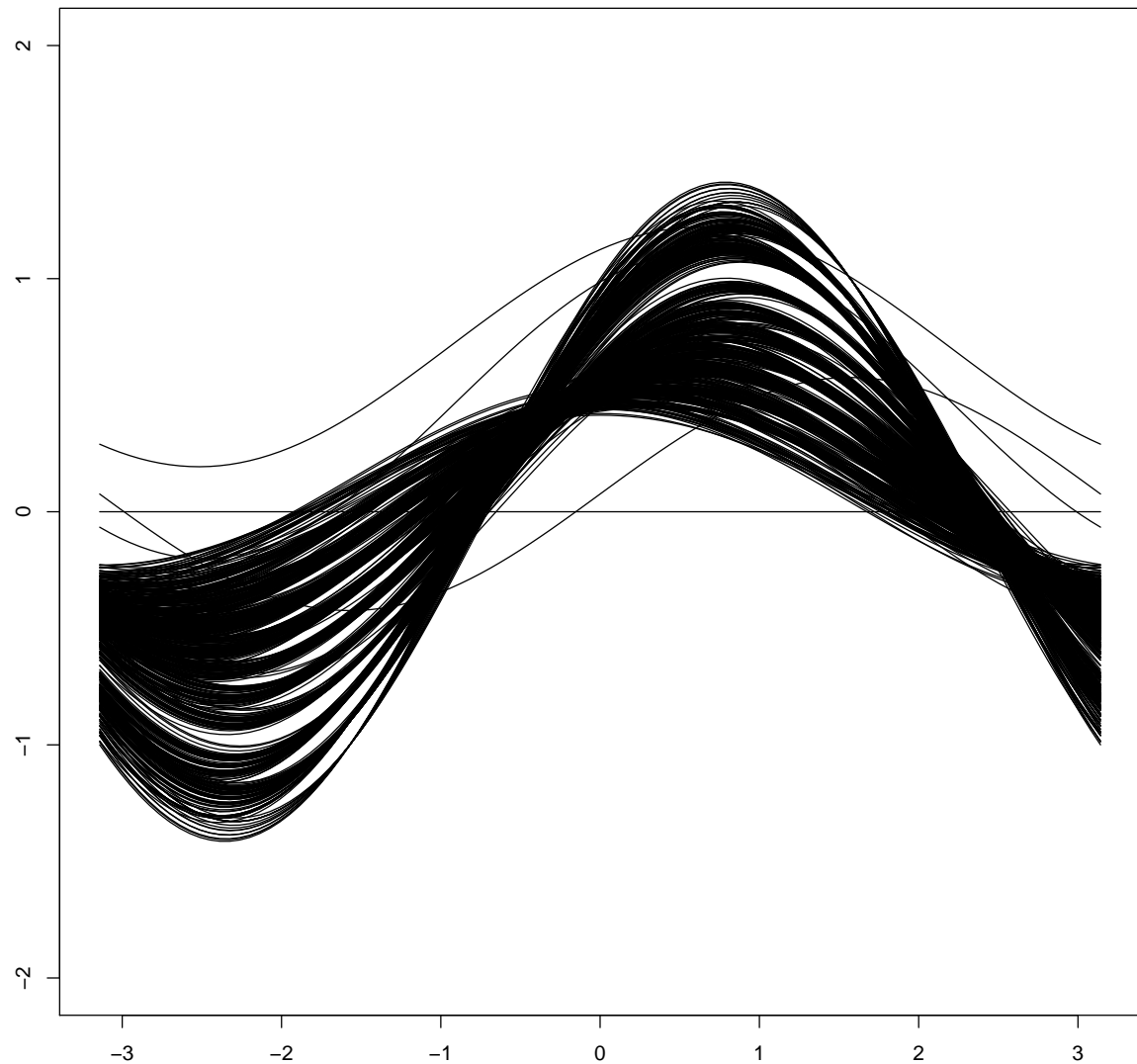
Para poder realizar el Gráfico de estrellas se realiza la normalización de las variables ya que son datos medidos en diferente escala:

```
X1s=(Var$Recorrido-mean(Var$Recorrido))/sd(Var$Recorrido)
X2s=(Var$Modelo_ano-mean(Var$Modelo_ano))/sd(Var$Modelo_ano)
X3s=(Var$Precio-mean(Var$Precio))/sd(Var$Precio)
Xs <- as.matrix(cbind(X1s,X2s,X3s))
stars(Xs,key.loc=c(14,1))
```



Se realiza la gráfica de las Curvas de Andrews para identificar agrupamiento de observaciones:

```
library(andrews)
X1s=(Var$Recorrido-mean(Var$Recorrido))/sd(Var$Recorrido)
X2s=(Var$Modelo_ano-mean(Var$Modelo_ano))/sd(Var$Modelo_ano)
X3s=(Var$Precio-mean(Var$Precio))/sd(Var$Precio)
Xs <- as.matrix(cbind(X1s,X2s,X3s))
andrews(Xs,type=1,ymax=2)
```

Conclusiones

- A pesar de que solo se tuvieron en cuentas las variables numéricas en el análisis se puede determinar que el Precio del Mazda 2 tiene una asociación lineal positiva fuerte con el año del vehículo, al contrario de los kilómetros recorridos, ya que en este caso la asociación lineal es negativa, indicando que a menos kilómetros recorridos el precio es mayor.