

Actividad 3

Stefania Astudillo Bello

Problema

Con base en los datos de ofertas de vivienda descargadas del portal Fincaraiz para apartamento de estrato 4 con área construida menor a 200 m^2 (vivienda4.RDS) la inmobiliaria A&C requiere el apoyo en la construcción de un modelo que lo oriente sobre los precios de inmuebles.

Con este propósito el equipo de asesores a diseñado los siguientes pasos para obtener un modelo y así poder a futuro determinar los precios de los inmuebles a negociar

```
viviendas <- readRDS(file = "vivienda4.RDS")
head(viviendas)
```

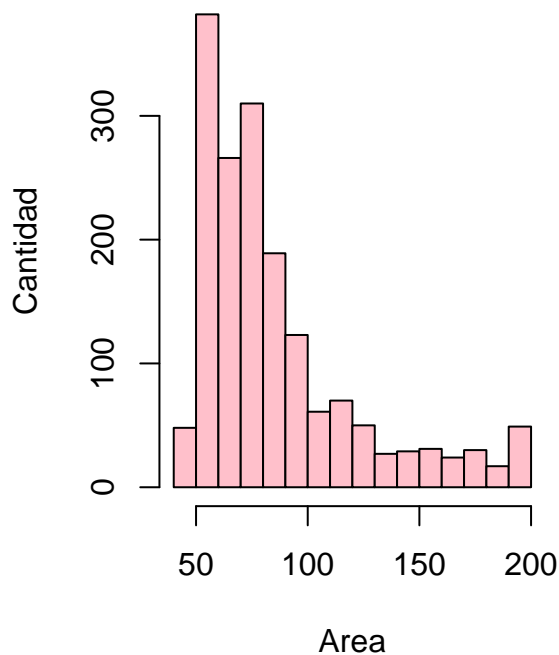
```
## # A tibble: 6 x 5
##   zona      estrato preciom areaconst tipo
##   <fct>    <fct>    <dbl>    <dbl> <fct>
## 1 Zona Norte 4      220      52 Apartamento
## 2 Zona Norte 4      600     160 Casa
## 3 Zona Norte 4      320     108 Apartamento
## 4 Zona Sur   4      290      96 Apartamento
## 5 Zona Norte 4      220      82 Apartamento
## 6 Zona Norte 4      305     117 Casa
```

1. Realice un análisis exploratorio de las variables precio de vivienda (millones de pesos COP) y área de la vivienda (metros cuadrados) - incluir gráficos e indicadores apropiados interpretados.

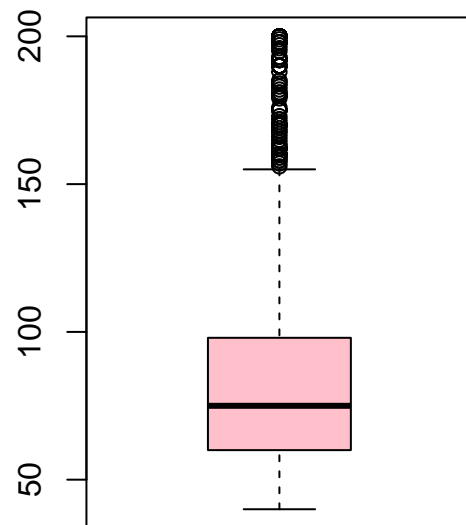
```
par(mfrow=c(1,2))
hist(viviendas$areaconst, xlab = "Area", ylab = "Cantidad", col="pink",
     main="Histograma de la variable areaconst")

boxplot(viviendas$areaconst, col="pink",
        main="Caja de la variable areaconst")
```

Histograma de la variable areacor



Caja de la variable areaconst



```
datosArea <- summary(viviendas$areaconst)
datosArea
```

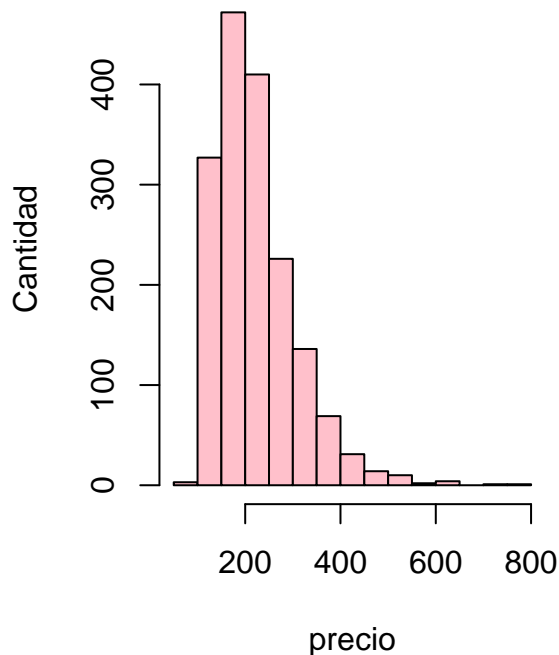
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   40.00   60.00   75.00   87.63   98.00  200.00
```

el promedio del área construida de la oferta de viviendas es de 87.63 m^2 para áreas que oscilan entre los 40 m^2 (mínimo) y los 200 m^2 (máximo). La distribución del área de las viviendas si muestra un sesgo bastante pronunciado a la derecha; podemos observar que el 50% de las viviendas tienen entre 40 m^2 y 75 m^2 mientras que el otro 50% tiene entre 75 m^2 y 200 m^2 .

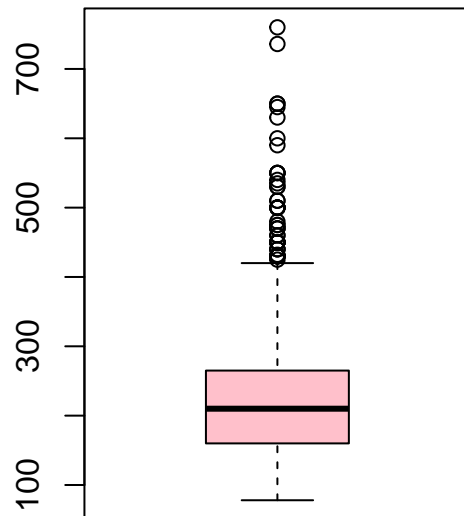
```
par(mfrow=c(1,2))
hist(viviendas$preciom, xlab = "precio", ylab = "Cantidad", col="pink",
     main="Histograma de la variable preciom")

boxplot(viviendas$preciom, col="pink",
       main="Caja de la variable preciom")
```

Histograma de la variable precio



Caja de la variable preciom



```
datosPrecio <- summary(viviendas$preciom)
datosPrecio
```

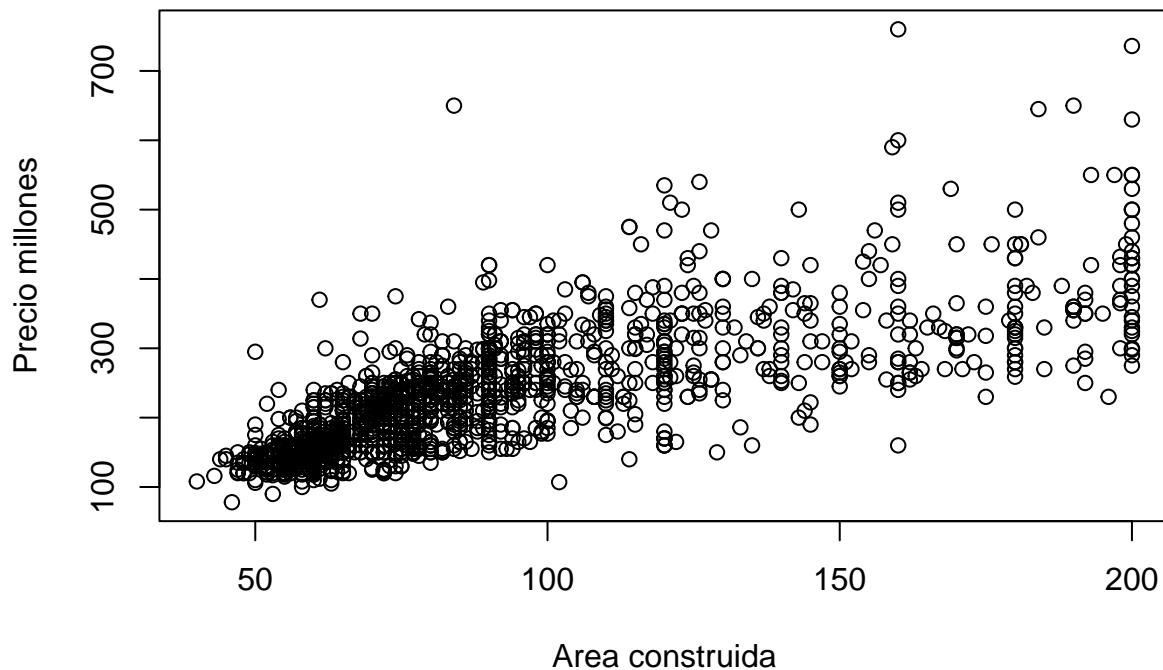
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      78.0  160.0   210.0   225.4  265.0   760.0
```

Los datos nos revelan que el promedio del precio de la oferta de viviendas es de 225.4 millones, para valores que oscilan entre los 78 millones (mínimo) y los 760 millones (máximo). podemos precisar que la mitad de las viviendas se oferta a un precio menor o igual a los 210 millones y sólo un 25% están en un rango de costosas superando los 265 millones de pesos colombianos.

2. Realice un análisis exploratorio bivariado de datos enfocado en la relación entre la variable respuesta (precio) en función de la variable predictora (área construida) - incluir gráficos e indicadores apropiados interpretados.

```
plot(viviendas$areaconst, viviendas$preciom,
     xlab = "Area construida", ylab="Precio millones",
     main="Análisis Exploratorio bivariado Area/Precio")
```

Analisis Exploratorio bivariado Area/Precio



```
cor(viviendas$areaconst,viviendas$preciom)
```

```
## [1] 0.7630166
```

Observando la gráfica de dispersión podemos destacar una tendencia directa o creciente en la relación del área construida con el precio de la vivienda, sin embargo, el comportamiento del precio es más disperso para áreas mayores a $100 m^2$, lo que puede dificultar la representatividad de un modelo de regresión lineal. Observando el Coeficiente de Correlación de Pearson (0.76) podemos determinar que la asociación que mide la relación lineal entre el precio y el área construida es una relación lineal positiva débil.

3. Estime el modelo de regresión lineal simple entre $precio = f(area) + \epsilon$. Interprete los coeficientes del modelo β_0, β_1 en caso de ser correcto.

```
mod = lm(viviendas$preciom~viviendas$areaconst)
summary(mod)
```

```
##
## Call:
## lm(formula = viviendas$preciom ~ viviendas$areaconst)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -195.86  -31.95   -8.95   27.87  431.17
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      67.381      3.510   19.20  <2e-16 ***
## viviendas$areaconst  1.803      0.037   48.73  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.53 on 1704 degrees of freedom
## Multiple R-squared:  0.5822, Adjusted R-squared:  0.5819
## F-statistic: 2374 on 1 and 1704 DF,  p-value: < 2.2e-16
```

El modelo de regresión lineal se estima en: $Precio = 67.381 + 1.803 * Area_{construida}$

$\beta_0 = 67.381$, este coeficiente nos indica el valor de la variable precio cuando no se tiene una vivienda construida, se podría pensar que es el valor base del terreno.

$\beta_1 = 1.803$, indica que por cada metro cuadrado construido el valor de la vivienda incrementara 1.803 millones aproximadamente.

4. Construir un intervalo de confianza (95%) para el coeficiente β_1 , interpretar y concluir si el coeficiente es igual a cero o no. Compare este resultado con una prueba de hipótesis t.

```
confint(mod)
```

```
##              2.5 %      97.5 %
## (Intercept)  60.496208  74.265055
## viviendas$areaconst  1.730404  1.875547
```

Con un nivel de confianza del 95% y un 5% de error α , podemos determinar que el coeficiente β_1 para el caso del área construida podría tomar un valor entre los 1,73 y los 1,87 millones por cada metro cuadrado construido. dado que el 0 no esta incluido en el intervalo de confianza, podemos concluir que se rechaza la hipótesis nula y se concluye que hay suficiente evidencia para no aceptar la nulidad de este coeficiente. Si comparamos este resultado con la prueba t que nos da el output del punto anterior, el p-value para el coeficiente β_1 fue $2.2e - 16$, esto apoya el rechazo de la hipótesis de nulidad y apoya la conclusión obtenida con el intervalo de confianza.

5. Calcule e interprete el indicador de bondad y ajuste R^2 .

```
r2 = summary(mod)$r.squared
print( paste("Coeficiente de determinación (R2): ", r2) )
```

```
## [1] "Coeficiente de determinación (R2):  0.582194373334701"
```

El indicador de bondad y ajuste para esta relación es de 0.58, esto quiere decir que el modelo explica un 58% de la variación del precio de la vivienda. De forma general podemos decir que el modelo tiene poca fuerza para predecir a la variable dependiente.

6. ¿Cuál sería el precio promedio estimado para un apartamento de 110 metros cuadrados? Considera entonces con este resultado que un apartamento en la misma zona con 110 metros cuadrados en un precio de 200 millones sería una atractiva esta oferta? ¿Qué consideraciones adicionales se deben tener?.