

# Ejercicios de Correlación y Regresión Camila y Roberto

2025-11-24

```
library(ggplot2)
library(ggpubr)
```

## Ejercicio 18. Tamaño de casinos e ingresos.

A continuación se presentan los tamaños (en miles de pies cuadrados) y los ingresos (en millones de dólares) de casinos de Atlantic City (según datos del New York Times). ¿Existe evidencia suficiente para concluir que existe una correlación lineal entre el tamaño de los casinos y sus ingresos?

### Crear vectores

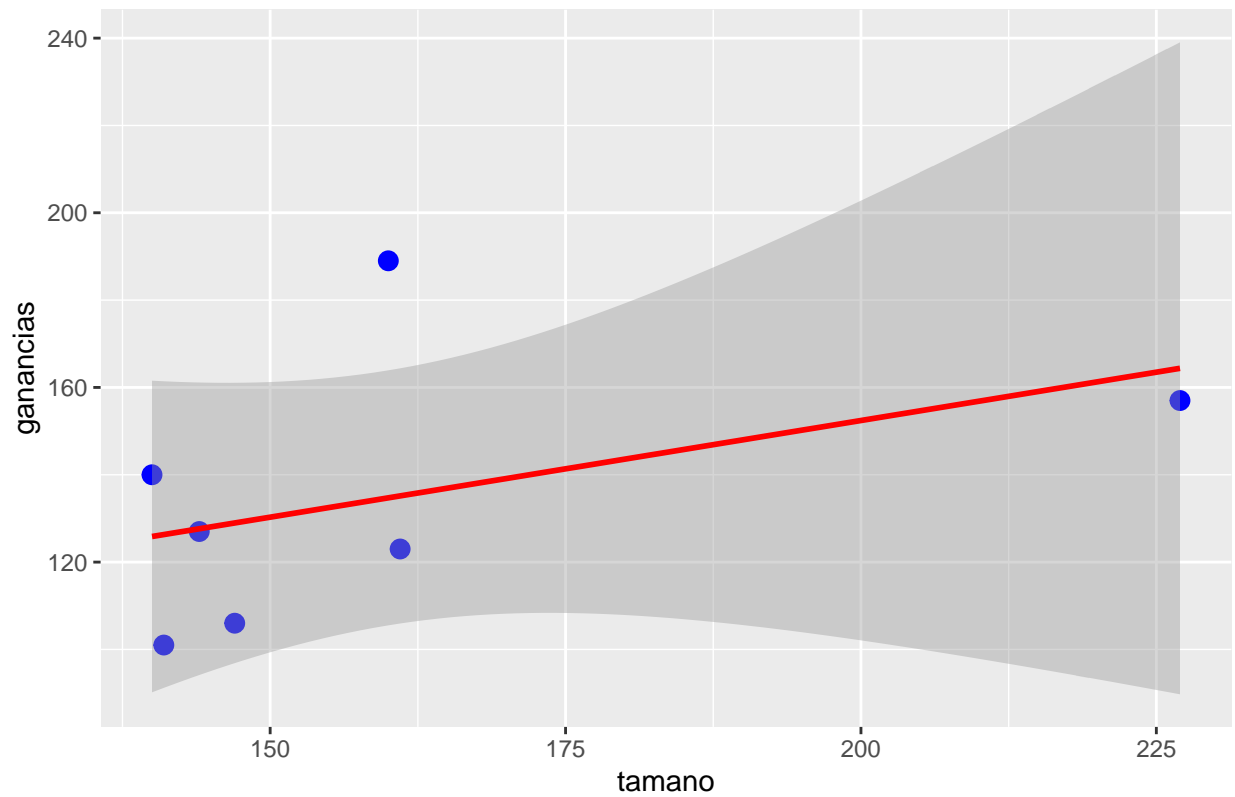
```
tamano <- c(160, 227, 140, 144, 161, 147, 141)
ganancias <- c(189, 157, 140, 127, 123, 106, 101)
```

### # Gráfico de dispersión + línea de regresión

```
ggplot(data.frame(tamano, ganancias),
  aes(x = tamano, y = ganancias)) +
  geom_point(color = "blue", size = 3) +
  geom_smooth(method = "lm", se = TRUE, color = "red") +
  ggtitle("Ejercicio 18: Tamaño vs Ganancias")
```

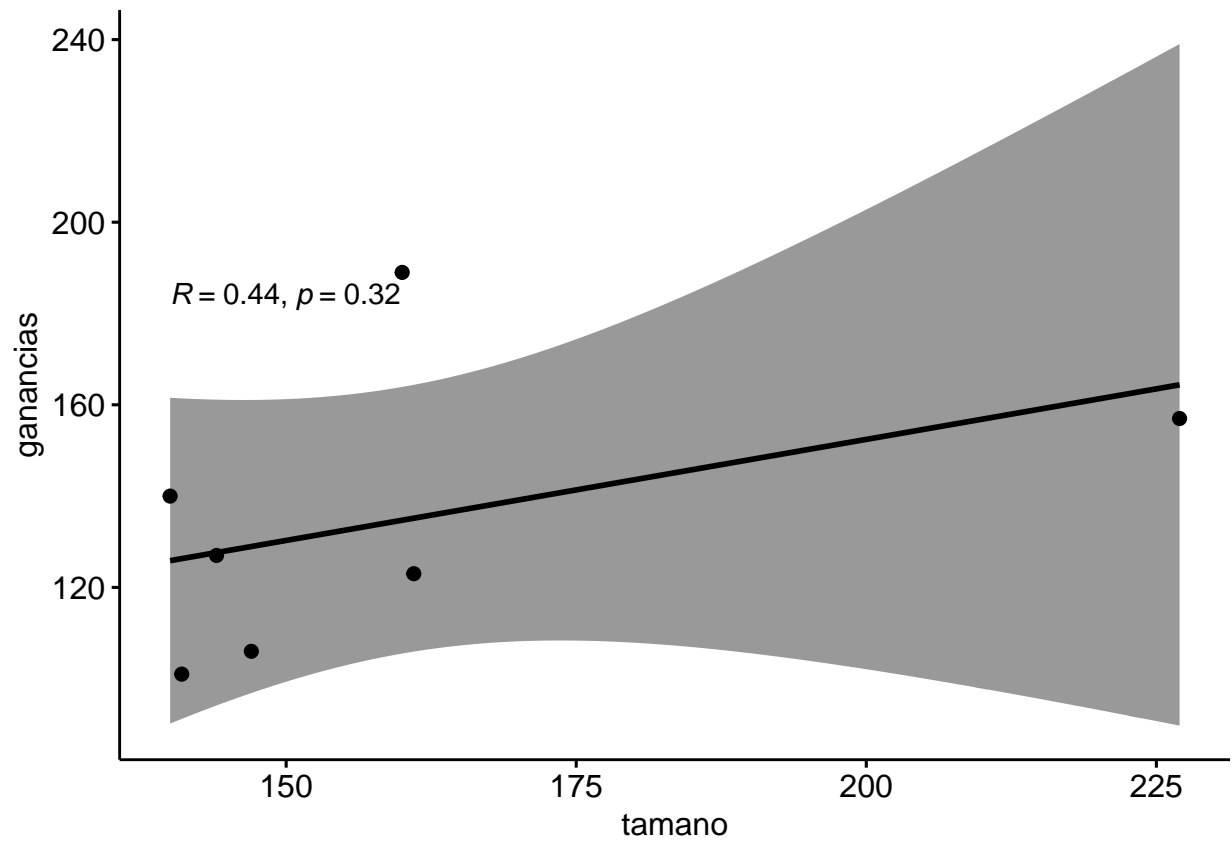
```
## 'geom_smooth()' using formula = 'y ~ x'
```

## Ejercicio 18: Tamaño vs Ganancias



## Correlacion Grafica

```
ggscatter(data.frame(tamano, ganancias),  
  x = "tamano", y = "ganancias",  
  add = "reg.line", conf.int = TRUE,  
  cor.coef = TRUE, cor.method = "pearson")
```



## Correlacion numerica

```
cor18 <- cor(tamano, ganancias)
cor18
```

```
## [1] 0.444569
```

## Regresion

```
modelo18 <- lm(ganancias ~ tamaño)
summary(modelo18)
```

```
##
## Call:
## lm(formula = ganancias ~ tamaño)
##
## Residuals:
##      1      2      3      4      5      6      7
## 54.2857 -7.3780 14.1405 -0.6304 -12.1570 -22.9586 -25.3022
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.8757    64.8338   0.985   0.370
## tamaño      0.4427     0.3989   1.110   0.318
##
## Residual standard error: 30.06 on 5 degrees of freedom
## Multiple R-squared:  0.1976, Adjusted R-squared:  0.03717
## F-statistic: 1.232 on 1 and 5 DF,  p-value: 0.3176
```

## EJERCICIO 19 – Tarifas aéreas (30 días vs 1 día)

A continuación se presentan los precios (en dólares) de tarifas aéreas de diferentes aerolíneas desde la ciudad de Nueva York (JFK) a San Francisco. Los precios se basan en boletos comprados con 30 días de anticipación y un día de anticipación; las aerolíneas son US Air, Continental, Delta, United, American, Alaska y Northwest. ¿Hay evidencia suficiente para concluir que existe una correlación lineal entre los precios de los boletos comprados con 30 días de anticipación y los boletos comprados con un día de anticipación?

### Crear vectores

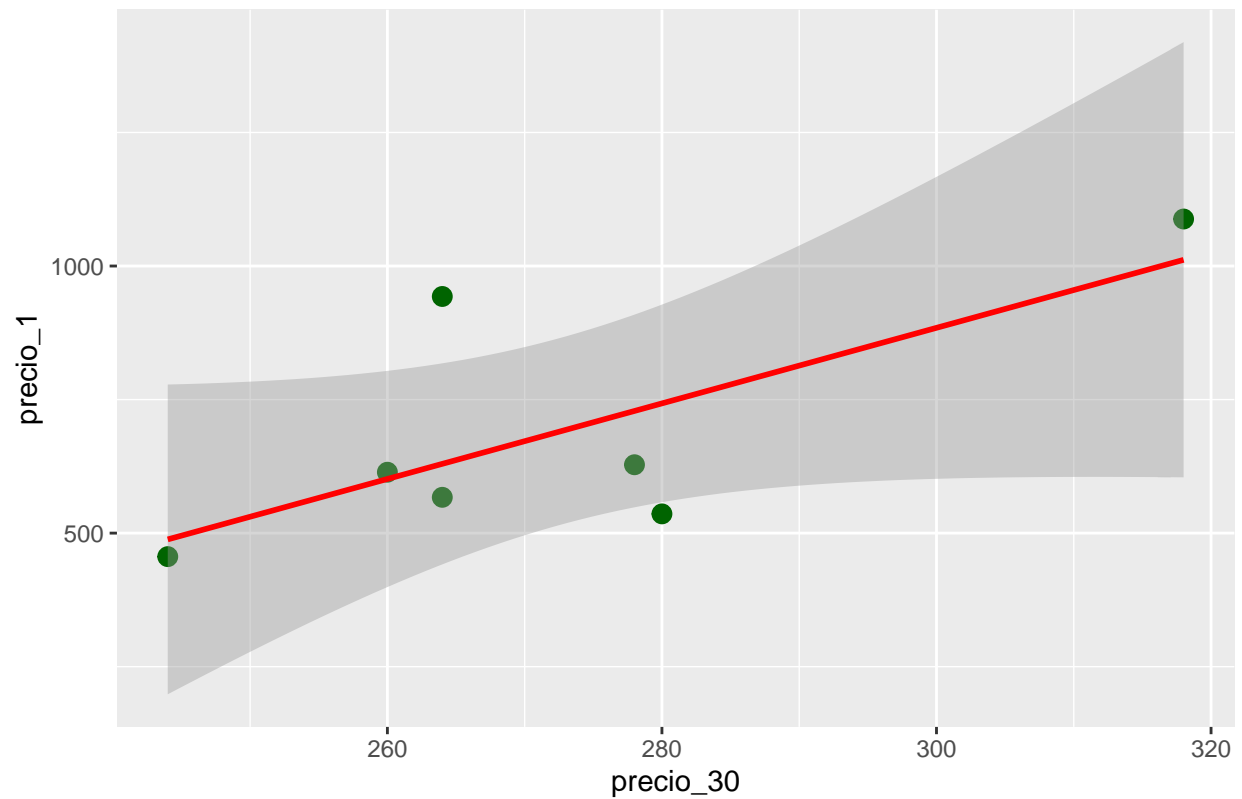
```
precio_30 <- c(244, 260, 264, 264, 278, 318, 280)
precio_1  <- c(456, 614, 567, 943, 628, 1088, 536)
```

### Gráfico con regresión

```
ggplot(data.frame(precio_30, precio_1),
       aes(x = precio_30, y = precio_1)) +
  geom_point(color = "darkgreen", size = 3) +
  geom_smooth(method = "lm", se = TRUE, color = "red") +
  ggtitle("Ejercicio 19: Precios 30 días vs 1 día")
```

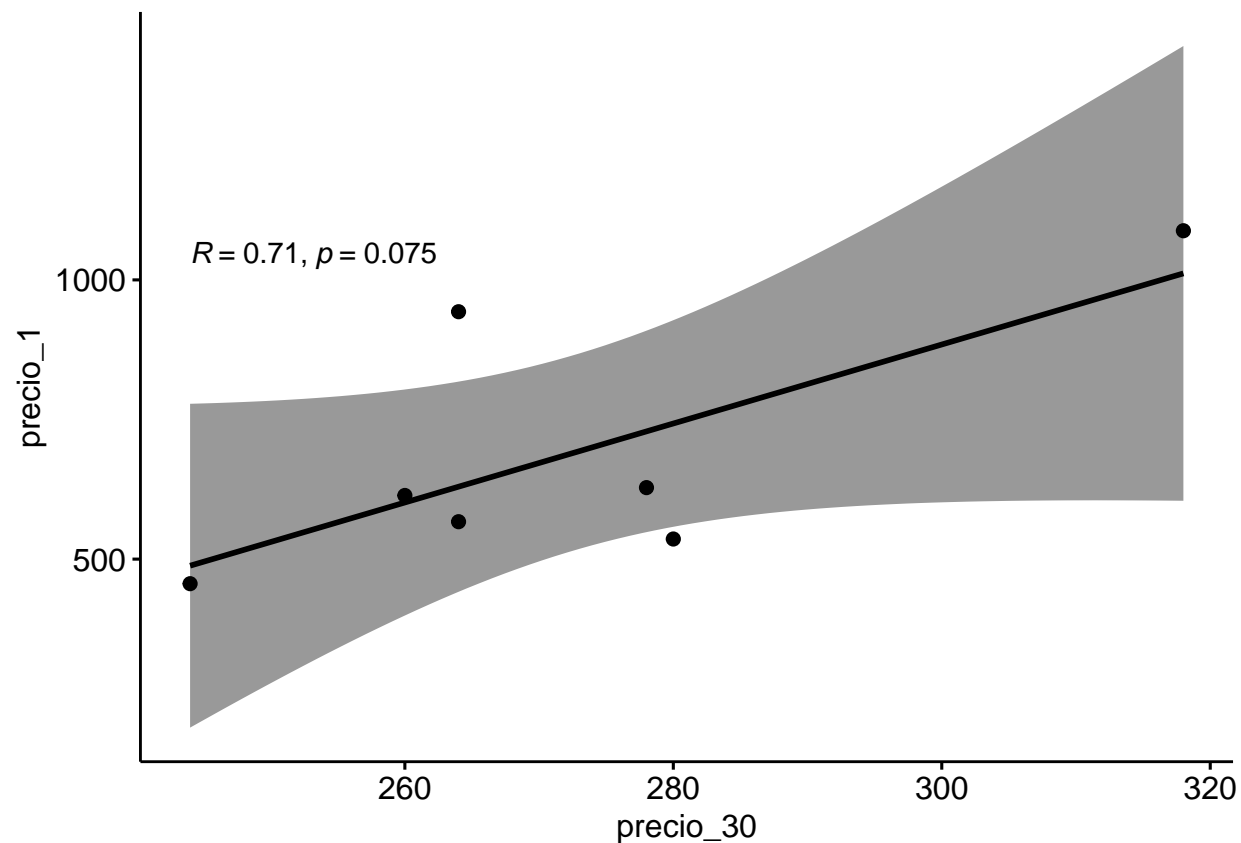
```
## 'geom_smooth()' using formula = 'y ~ x'
```

### Ejercicio 19: Precios 30 días vs 1 día



### Correlación gráfica

```
ggscatter(data.frame(precio_30, precio_1),  
  x = "precio_30", y = "precio_1",  
  add = "reg.line", conf.int = TRUE,  
  cor.coef = TRUE, cor.method = "pearson")
```



## Correlación numérica

```
cor19 <- cor(precio_30, precio_1)
cor19
```

```
## [1] 0.7087568
```

## Regresión

```
modelo19 <- lm(precio_1 ~ precio_30)
summary(modelo19)
```

```
##
## Call:
## lm(formula = precio_1 ~ precio_30)
##
## Residuals:
##      1      2      3      4      5      6      7
## -32.20  12.63 -62.66 313.34 -100.68  76.40 -206.83
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1237.626    860.863  -1.438   0.2100
## precio_30     7.073      3.148   2.247   0.0746 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 180 on 5 degrees of freedom
## Multiple R-squared:  0.5023, Adjusted R-squared:  0.4028
## F-statistic: 5.047 on 1 and 5 DF,  p-value: 0.0746
```

## 20. Pasajeros y espacios de estacionamiento.

A continuación se presentan los números de pasajeros y los números de espacios de estacionamiento en diferentes estaciones del tren Metro-North (según datos de MetroNorth). ¿Existe una correlación lineal entre los números de pasajeros y los números de espacios de estacionamiento?

### Crear vectores

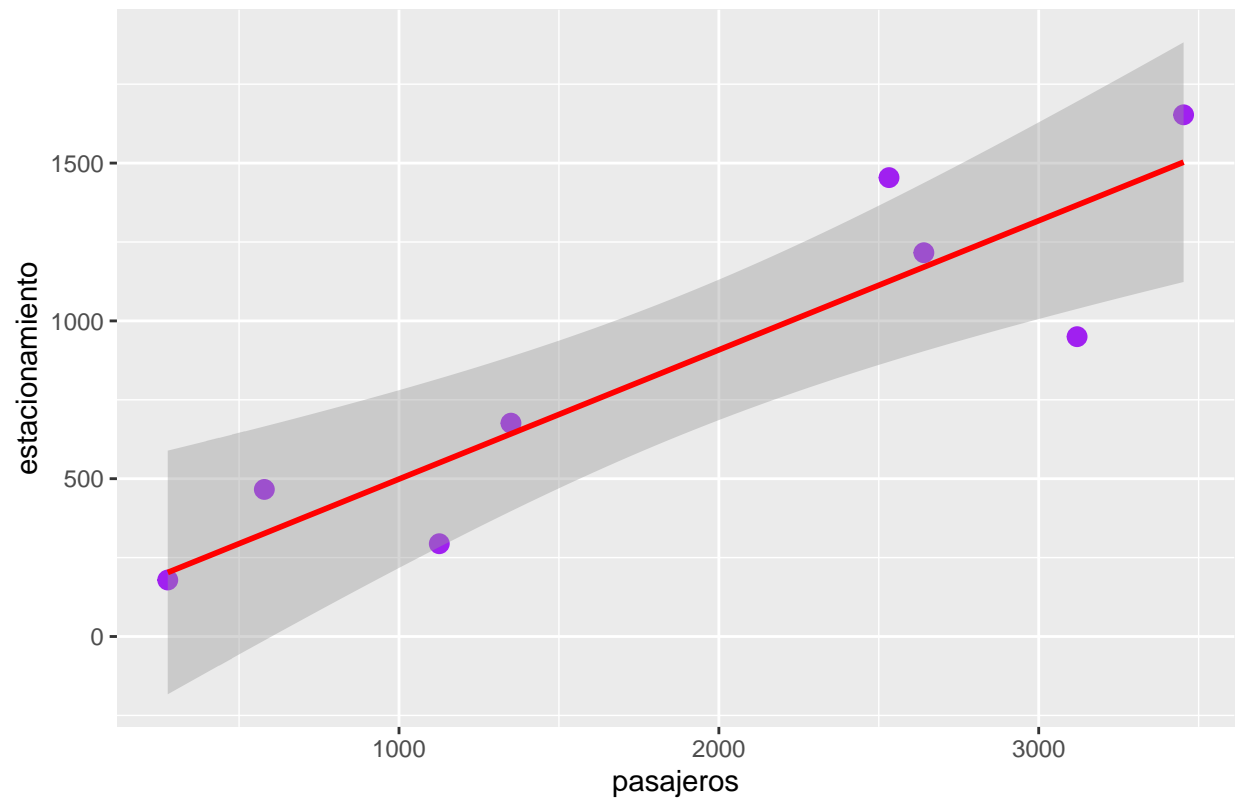
```
pasajeros <- c(3453, 1350, 1126, 3120, 2641, 277, 579, 2532)
estacionamiento <- c(1653, 676, 294, 950, 1216, 179, 466, 1454)
```

### Grafico

```
ggplot(data.frame(pasajeros, estacionamiento),
  aes(pasajeros, estacionamiento)) +
  geom_point(color = "purple", size = 3) +
  geom_smooth(method = "lm", se = TRUE, color = "red") +
  ggtitle("Ejercicio 20: Pasajeros vs Estacionamiento")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

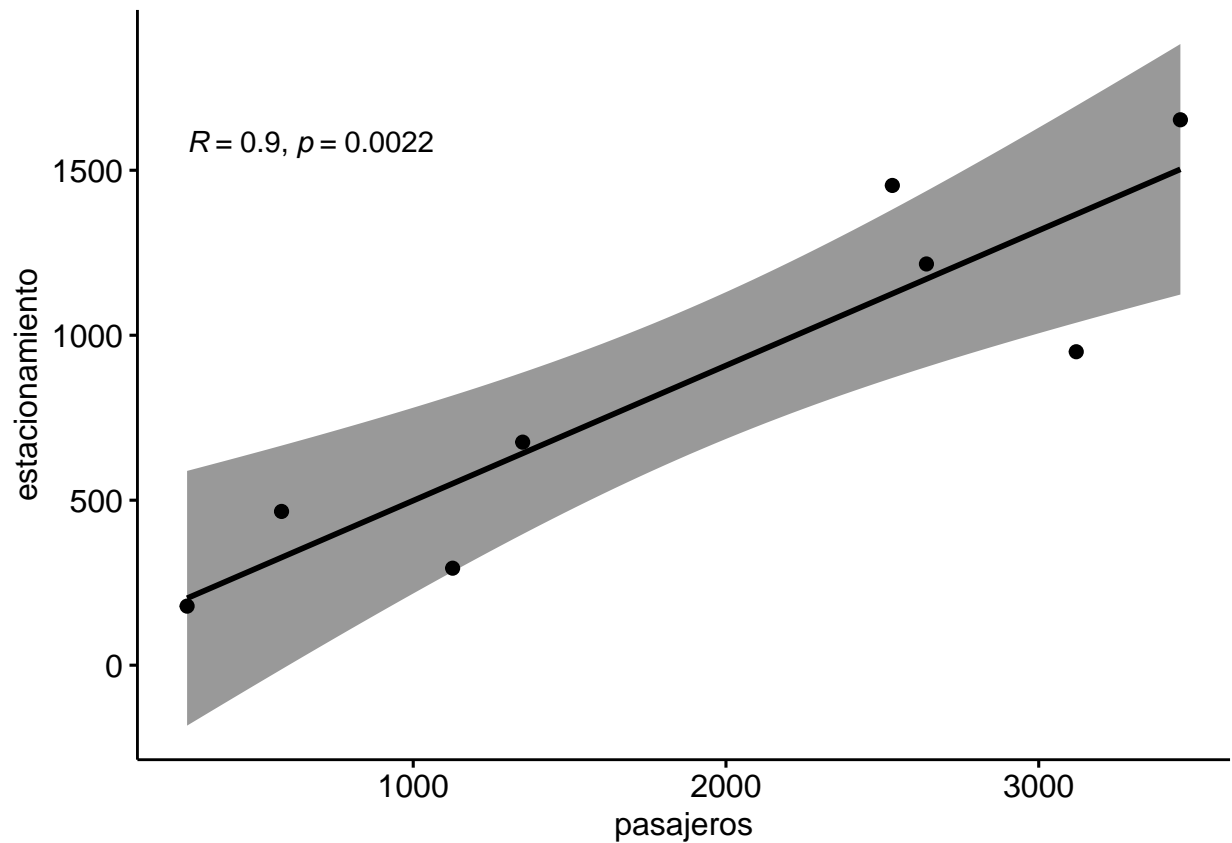
## Ejercicio 20: Pasajeros vs Estacionamiento



### Gráfico de correlación

```
ggscatter(data.frame(pasajeros, estacionamiento),  
  x = "pasajeros", y = "estacionamiento",  
  add = "reg.line", conf.int = TRUE,  
  cor.coef = TRUE, cor.method = "pearson")
```





## Correlacion

```
cor20 <- cor(pasajeros, estacionamiento)
cor20
```

```
## [1] 0.9011516
```

## Regresion

```
modelo20 <- lm(estacionamiento ~ pasajeros)
summary(modelo20)
```

```
##
## Call:
## lm(formula = estacionamiento ~ pasajeros)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -416.64  -82.01   39.67  142.14  328.05
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  89.48817  176.46254   0.507  0.63016
## pasajeros   0.40934   0.08039   5.092  0.00224 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 255.8 on 6 degrees of freedom
## Multiple R-squared:  0.8121, Adjusted R-squared:  0.7808
## F-statistic: 25.93 on 1 and 6 DF,  p-value: 0.002239
```

## EJERCICIO 21 – Costos frontales vs traseros

A continuación se presentan los costos de reparación (en dólares) para automóviles que participaron en pruebas de choques frontales a una velocidad de 6 mi/h y en pruebas de choques traseros a una velocidad de 6 mi/h (según datos del Insurance Institute for Highway Safety). Los automóviles son Toyota Camry, Mazda 6, Volvo S40, Saturn Aura, Subaru Legacy, Hyundai Sonata y Honda Accord. ¿Hay evidencia suficiente para concluir que existe una correlación lineal entre los costos de reparación de los choques frontales y de los choques traseros?

### Crear vectores

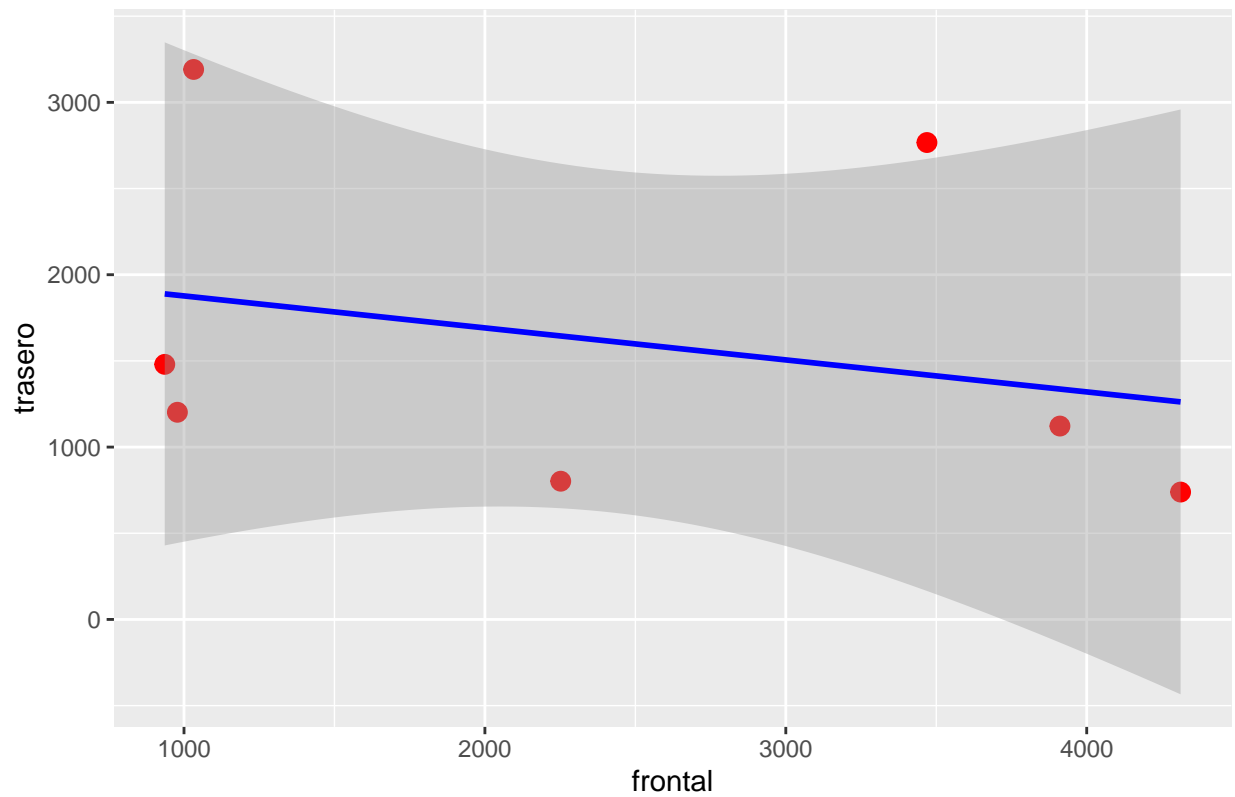
```
frontal <- c(936, 978, 2252, 1032, 3911, 4312, 3469)
trasero <- c(1480, 1202, 802, 3191, 1122, 739, 2767)
```

### Gráfico

```
ggplot(data.frame(frontal, trasero),
       aes(frontal, trasero)) +
  geom_point(color = "red", size = 3) +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  ggtitle("Ejercicio 21: Choques frontales vs traseros")
```

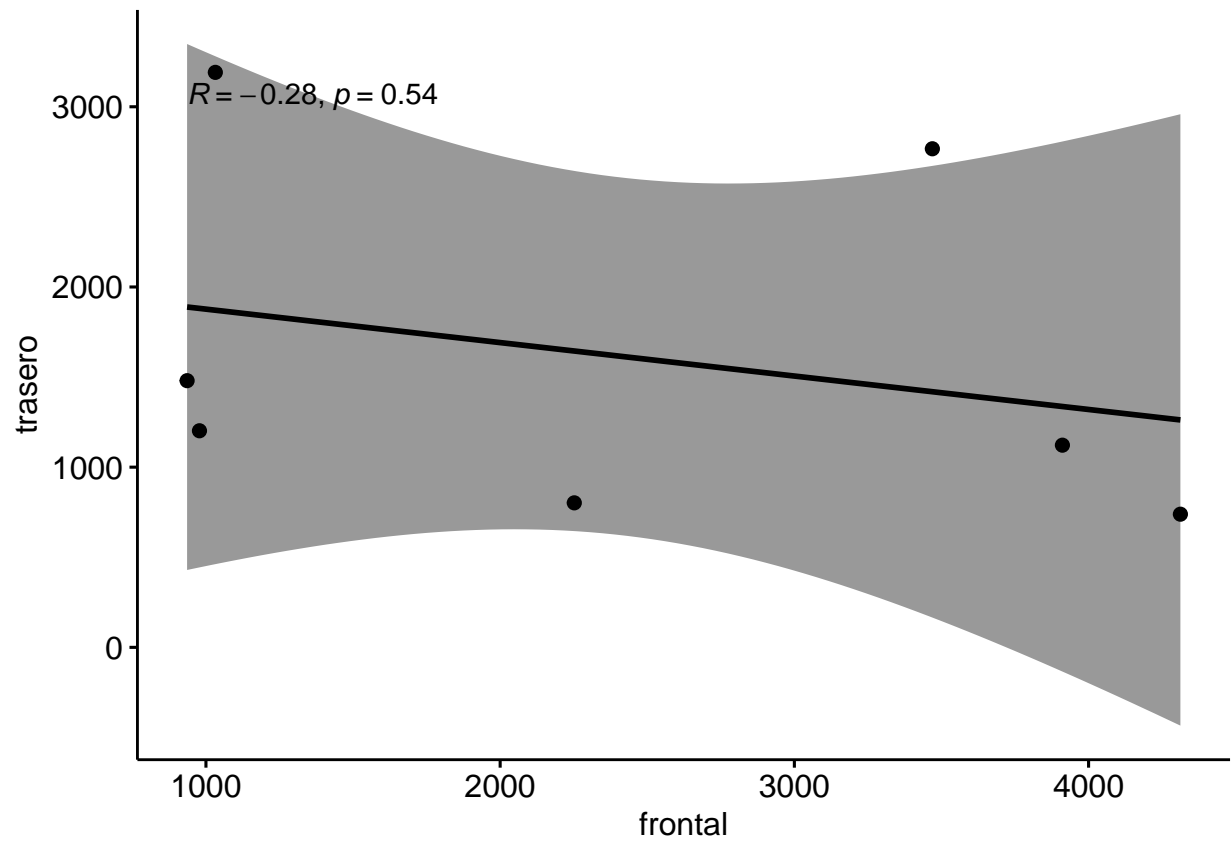
```
## 'geom_smooth()' using formula = 'y ~ x'
```

## Ejercicio 21: Choques frontales vs traseros



## Correlación gráfica

```
ggscatter(data.frame(frontal, trasero),  
  x = "frontal", y = "trasero",  
  add = "reg.line", conf.int = TRUE,  
  cor.coef = TRUE, cor.method = "pearson")
```



## Correlacion

```
cor21 <- cor(frontal, trasero)
cor21
```

```
## [1] -0.2825546
```

## Regresion

```
cor21 <- cor(frontal, trasero)
cor21
```

```
## [1] -0.2825546
```