Ejercicio de clase 5

Andrés Limón Cruz

2024-03-23

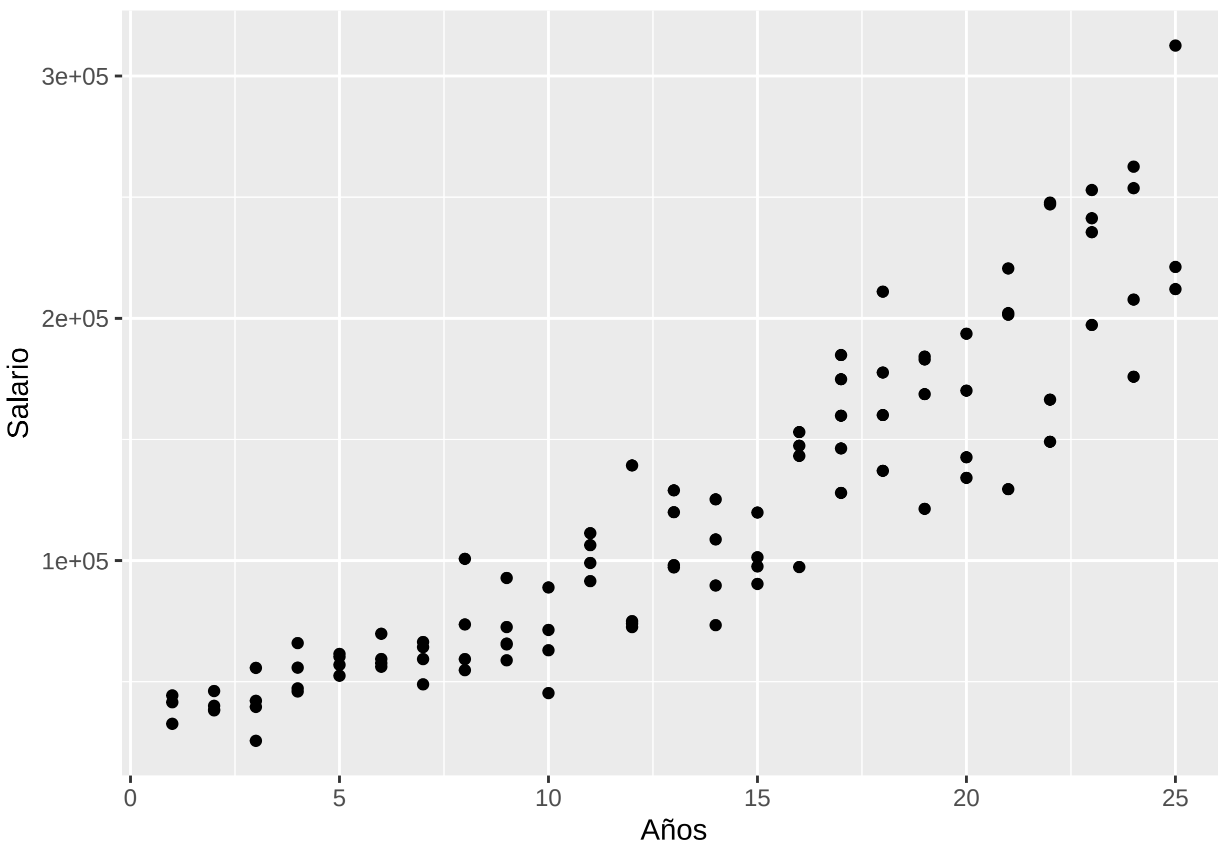
```
library(tidyverse)
```

a) Hacer dos scatterplots. Uno con los datos sin transformar, y otro con alguna transformación a los datos (usen transformaciones tipo Box-Cox o Box-Tidwell)

```
datos <- read_csv("initech.csv")
```

```
## Rows: 100 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (2): years, salary
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
ggplot() + geom_point(mapping = aes(x = datos$years, y = datos$salary)) +
  xlab("Años") + ylab("Salario")
```



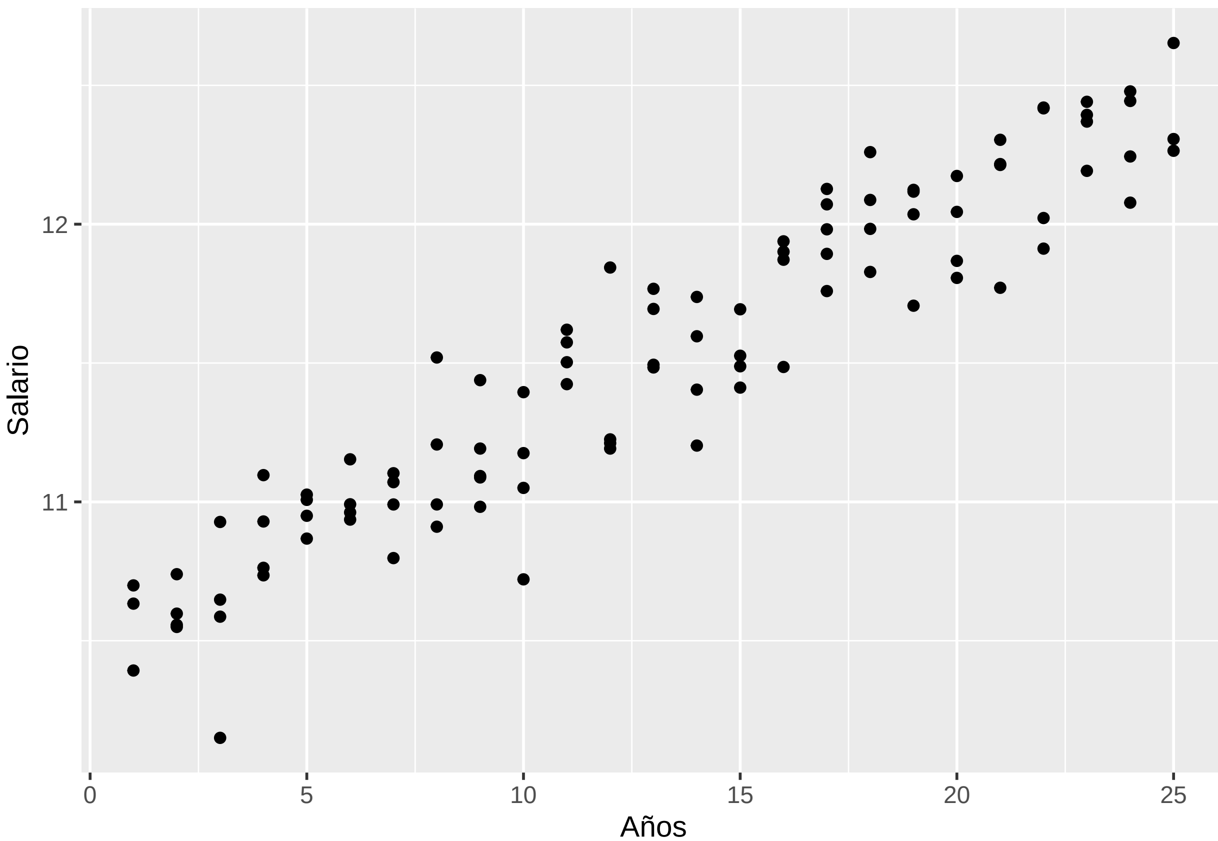
Apliquemos ahora una transformación a los datos, en este caso he escogido usar box-cox:

```
fit <- glm(salary ~ years, data = datos)
car::powerTransform(fit)
```

```
## Estimated transformation parameter
##         Y1
## 0.07897346
```

Entonces tenemos que el mas cercano es 0, entonces usamos log()

```
ggplot() + geom_point(mapping = aes(x = datos$years, y = log(datos$salary))) +
  xlab("Años") + ylab("Salario")
```



b) Ajustar una regresión según la transformación que hayan elegido (o sea creen un objeto de R de tipo lm usando la respectiva función lm() que hayan usado)

```
fit <- lm(log(salary) ~ years, data = datos)
b0 <- coef(fit)[1]
b1 <- coef(fit)[2]
print(b0)
```

```
## (Intercept)
##    10.48381
```

```
print(b1)
```

```
##      years
## 0.07887976
```

c) De ser posible, interpreten los parámetros β_0 y β_1 de su ajuste.

```r
summary(fit)
```

```
##
## Call:
## lm(formula = log(salary) ~ years, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.57022 -0.13560  0.03048  0.14157  0.41366
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.48381    0.04108  255.18   <2e-16 ***
## years        0.07888    0.00278   28.38   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1955 on 98 degrees of freedom
## Multiple R-squared:  0.8915, Adjusted R-squared:  0.8904
## F-statistic: 805.2 on 1 and 98 DF,  p-value: < 2.2e-16
```

Entonces veamos que

$$\mathbb{E}(y|x=0) = \exp(b_0)$$

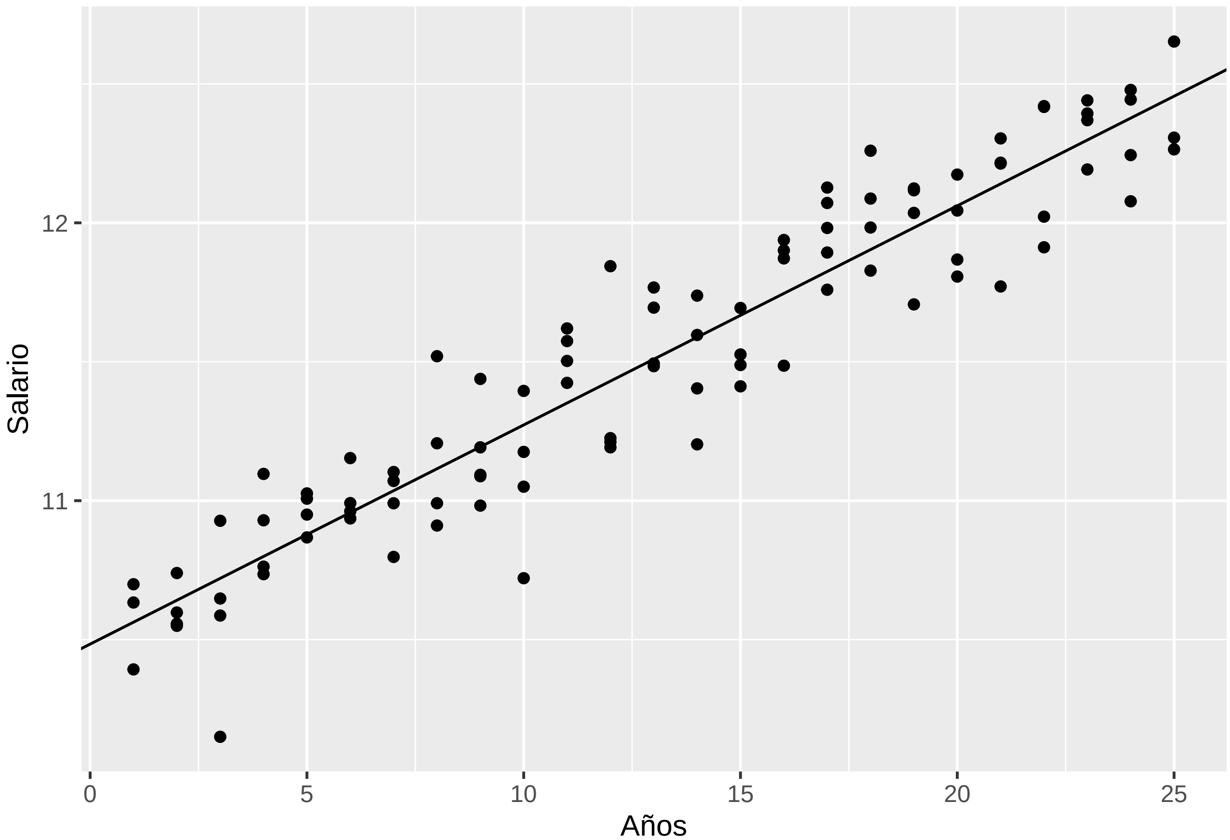
y para b_1 tenemos que

$$\mathbb{E}(y|x+1) = \exp(b_0 + b_1(x+1)) = \exp(b_0 + b_1x + b_1) = \exp(b_0 + b_1x) \cdot \exp(b_1)$$

Es decir cuando $x = 0$ tenemos que $\exp(b_0) = \mathbb{E}(y)$ y cada que incrementamos una unidad a x se multiplica el valor por la constante $\exp(b_1)$

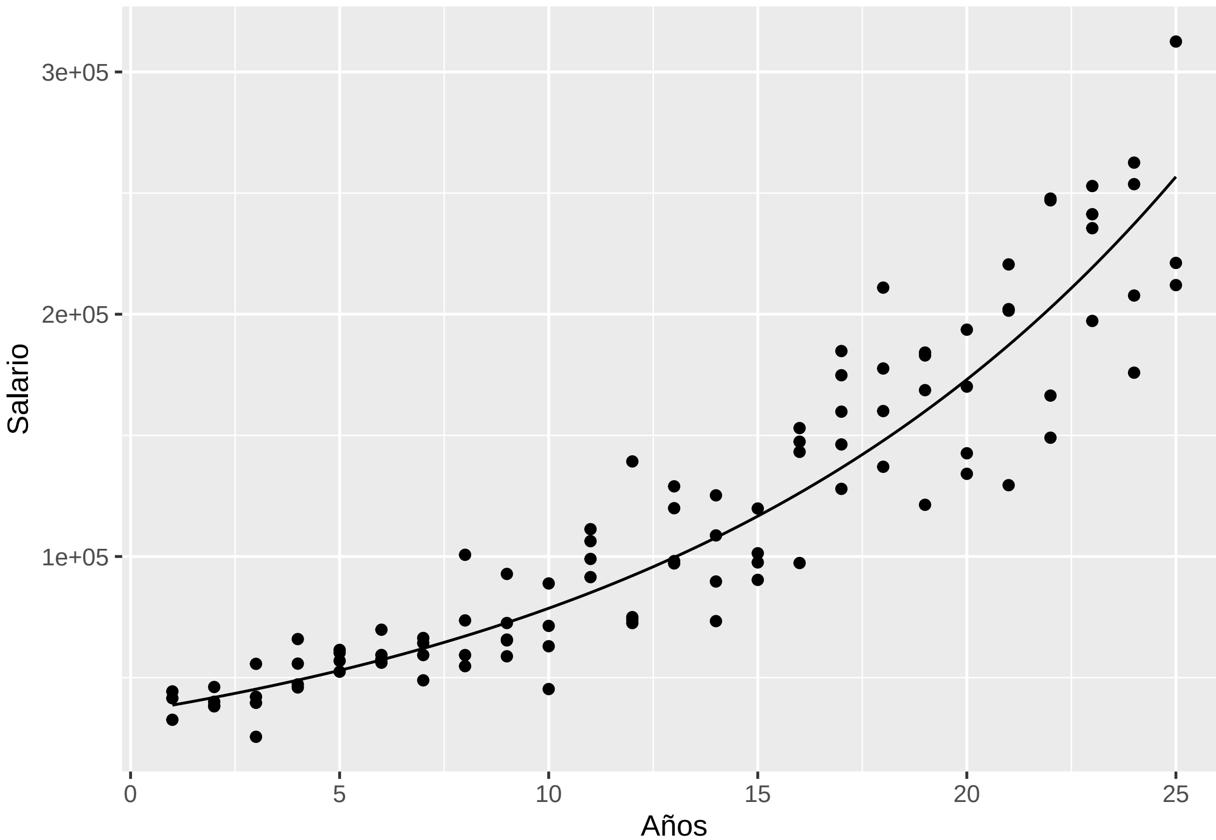
d) Hagan un scatterplot con los datos transformados y su respectiva recta ajustada al modelo

```r
ggplot() + geom_point(mapping = aes(x = datos$years, y = log(datos$salary))) +
  xlab("Años") + ylab("Salario") +
  geom_abline(intercept = b0, slope = b1)
```



e) Hagan un scatterplot con los datos sin transformar y su respectiva curva ajustada al modelo

```
curva_ajustada <- function(x) {exp(b0 + b1*x)}
ggplot() + geom_point(mapping = aes(x = datos$years, y = datos$salary)) +
  xlab("Años") + ylab("Salario") +
  geom_function(fun = curva_ajustada)
```



3e+05
2e+05
1e+05
Salario
0
5
10
15
20
25
Años