

# Actividad 6.1

## Regresión Lineal Simple

---

**A01351806** Renata Pilar  
Gómez Castillo

**5 de Mayo de 2023**

Tecnológico de Monterrey, Campus Qro



# Regresión Lineal simple

Para este proyecto se utilizaron dos Data Frames de dos ciudades diferentes, D.F. y Girona. Se analizó la correlación existente, modelo matemático y regresión lineal de 6 variables para 2 tipos diferentes de habitación en el dataset.

## Data Frame D.F. México

|    | Tipo Habitación | Variable X                  | Variable Y             | Coef Determinación | Coef Correlación |
|----|-----------------|-----------------------------|------------------------|--------------------|------------------|
| 0  | Entire home/apt | host_response_rate          | host_acceptance_rate   | 0.479756           | 0.692644         |
| 1  | Entire home/apt | price                       | host_acceptance_rate   | 0.000555           | 0.023561         |
| 2  | Entire home/apt | number_of_reviews           | host_acceptance_rate   | 0.017849           | 0.133599         |
| 3  | Entire home/apt | review_scores_cleanliness   | review_scores_location | 0.806106           | 0.897834         |
| 4  | Entire home/apt | number_of_reviews           | availability_365       | 0.002782           | 0.052747         |
| 5  | Entire home/apt | review_scores_communication | reviews_per_month      | 0.140989           | 0.375485         |
| 6  | Private Room    | host_response_rate          | host_acceptance_rate   | 0.013333           | 0.115468         |
| 7  | Private Room    | price                       | host_acceptance_rate   | 0.000113           | 0.010608         |
| 8  | Private Room    | number_of_reviews           | host_acceptance_rate   | 0.019622           | 0.140079         |
| 9  | Private Room    | review_scores_cleanliness   | review_scores_location | 0.829152           | 0.910578         |
| 10 | Private Room    | number_of_reviews           | availability_365       | 0.073772           | 0.271611         |
| 11 | Private Room    | review_scores_communication | reviews_per_month      | 0.168067           | 0.409959         |

## Data Frame Girona España

|    | Tipo Habitación | Variable X                  | Variable Y             | Coef Determinación | Coef Correlación |
|----|-----------------|-----------------------------|------------------------|--------------------|------------------|
| 0  | Entire home/apt | host_response_rate          | host_acceptance_rate   | 0.296518           | 0.544535         |
| 1  | Entire home/apt | price                       | host_acceptance_rate   | 0.000959           | 0.030962         |
| 2  | Entire home/apt | number_of_reviews           | host_acceptance_rate   | 0.004182           | 0.064665         |
| 3  | Entire home/apt | review_scores_cleanliness   | review_scores_location | 0.944244           | 0.971722         |
| 4  | Entire home/apt | number_of_reviews           | availability_365       | 0.000990           | 0.031469         |
| 5  | Entire home/apt | review_scores_communication | reviews_per_month      | 0.228595           | 0.478117         |
| 6  | Private Room    | host_response_rate          | host_acceptance_rate   | 0.316489           | 0.562573         |
| 7  | Private Room    | price                       | host_acceptance_rate   | 0.225146           | 0.474495         |
| 8  | Private Room    | number_of_reviews           | host_acceptance_rate   | 0.021538           | 0.146759         |
| 9  | Private Room    | review_scores_cleanliness   | review_scores_location | 0.977059           | 0.988463         |
| 10 | Private Room    | number_of_reviews           | availability_365       | 0.698815           | 0.835951         |
| 11 | Private Room    | review_scores_communication | reviews_per_month      | 0.354133           | 0.595091         |

Para ambas ciudades se realizó el modelo de regresión lineal en dos diferentes tipos de habitación ('Entire home/Apt y Private room).

Podemos observar que DF en el tipo de habitación Entire home/apt el modelo que tiene un mayor coeficiente de correlación (positiva considerable) y determinación es el modelo 4 (**review\_scores\_cleanliness & review\_scores\_location**), con un valor de 0.89 y 0.80 correspondientemente.

De igual manera, este es el mismo modelo con los coeficientes más grandes en Girona, obteniendo un coeficiente de correlación de 0.97 (positiva muy fuerte) y de determinación de 0.94. El modelo de Girona obtuvo mejores predicciones que el de DF, pero ambos fueron los coeficientes más altos en el tipo de habitación Entire home/apt.

Esto nos indica que las variable independiente **review\_scores\_cleanliness** tiene poder explicativo sobre la variable independiente **review\_scores\_location**, teniendo así una correlación positiva entre ambas variables. De igual manera, estas dos variables son las que presentan un modelo con el mayor nivel de correlación y determinación en el tipo de habitación Private Room.

Es importante mencionar que el coeficiente de correlación nos permite observar la relación lineal que existe entre dos variables. Mientras que el coeficiente de determinación indica la proporción de la varianza de la variable que es explicada por la regresión, según la variable independiente.

Por el otro lado, podemos observar que el modelo en Entire home/apt con el valor de correlación y determinación más pequeño para DF es el de las variables **price & host\_acceptance\_rate**, mientras que para Girona es con las variables **number\_of\_reviews & availability\_365**. La correlación de estas variables, para DF y Girona, está muy cerca del 0, por lo que su nivel de correlación es casi nulo.