

Modelos de Regresión y Series de Tiempo (MRyST)

2025 - 02

Clase 1 – Introducción del curso e introducción a los MR

Docente: Natalia Jaramillo Quiceno

Escuela de Ingenierías

natalia.jaramilloq@upb.edu.co

Atención a estudiantes:

Plataforma TEAMS

Un recordaris antes de comenzar...

- **Estadística descriptiva**

Método para describir una muestra “Pasado”

Métodos gráficos y numéricos

- **Probabilidad**

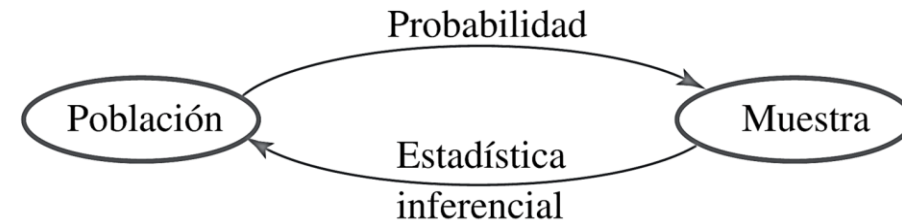
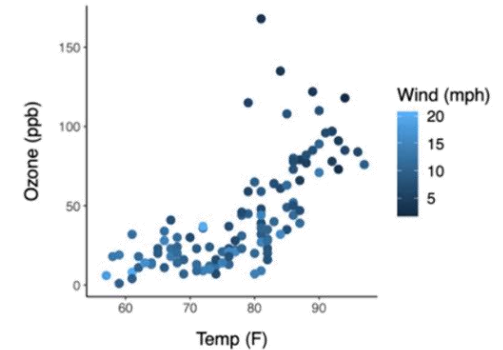
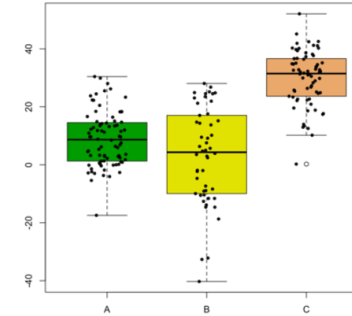
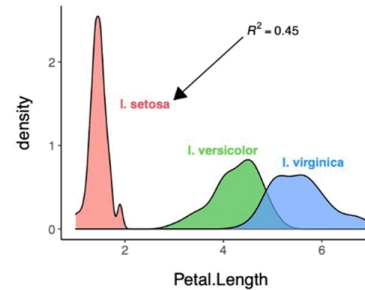
Teoría que permite anticipar un evento aleatorio “Futuro”

**Variables aleatorias y
distribuciones de probabilidad**

- **Inferencia**

Métodos para sacar conclusiones de una población a partir de una muestra (o más)

**Estimación puntual o por intervalos de confianza
Pruebas de hipótesis**



¿Qué veremos en este curso?



Técnicas fundamentales para **pronosticar** el valor de una **variable dependiente (Y)**, que se encuentra asociada a una o más **variables independientes (X)**.

Modelos de regresión

Series de tiempo

Variable
independiente



Cantidad

Una o más

Única

Tipo

Numérica o
categórica

Tiempo
(intervalos fijos)

Análisis de regresión ¿a qué se refiere?



- El **análisis de regresión** es una técnica para investigar y modelizar la **relación** entre variables.

- Se usa para explorar y *cuantificar* la relación de dependencia entre:

Una **variable dependiente o respuesta (Y)**

Una o más variables **independientes** llamadas **variables predictoras o explicativas (X)**

- Aplicaciones en casi todos los campos: ingeniería, física, ciencias económicas, ciencias biológicas y de la salud, ciencias sociales...



Variable dependiente

Variable explicativa



Variable explicada

Variable independiente



Predicha

Predictora



Regresada

Regresora



Respuesta

Estímulo

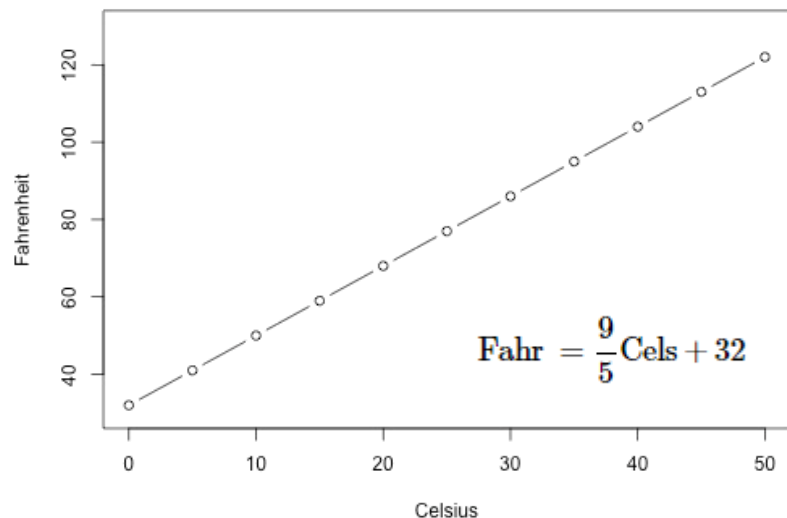


Endógena

Exógena

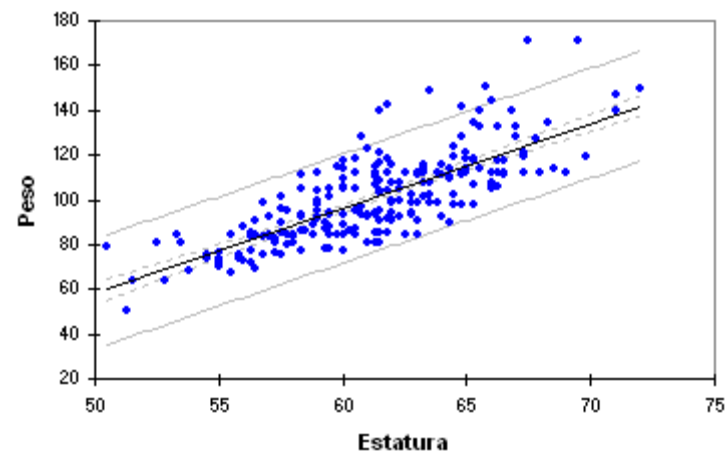
Modelos determinísticos vs. Modelos probabilísticos

Determinístico



- Relación **determinista**.
- Los puntos se ajustan *perfectamente* a una recta
- La ecuación describe *exactamente* la relación entre dos variables

Probabilístico



- Se puede esperar que a mayor estatura haya más peso
- Se evidencia una **tendencia** en los datos, pero también se evidencia cierta **dispersión**
- La relación entre las variables *no es perfecta*

Tipos de regresión



- Según el número de variables independientes:
 - **Regresión simple:** la variable Y depende únicamente de una única variable X .
 - **Regresión múltiple:** la variable Y depende de varias variables (X_1, X_2, \dots, X_r)
- Según el tipo de relación (función $f(X)$) :
 - **Regresión lineal:** Cuando $f(X)$ es una función lineal.
 - **Regresión no lineal:** Cuando $f(X)$ no es una función lineal (cúbica, cuadrática, exponencial, logística etc).
- Según la naturaleza de la relación que exista entre las dos variables:
 - La variable X puede ser la **causa** del valor de la variable Y .
Ej., en toxicología, si x =dosis de la droga, y =mortalidad \rightarrow la mortalidad se atribuye a la dosis administrada y no a otras causas.
 - Puede haber simplemente **relación** entre las dos variables.
Ej., en un estudio en que se estudian las variables x =peso, y =altura de un grupo de individuos, puede haber relación entre las dos, aunque difícilmente una pueda considerarse causa de la otra.

Series de tiempo ¿a qué se refieren?



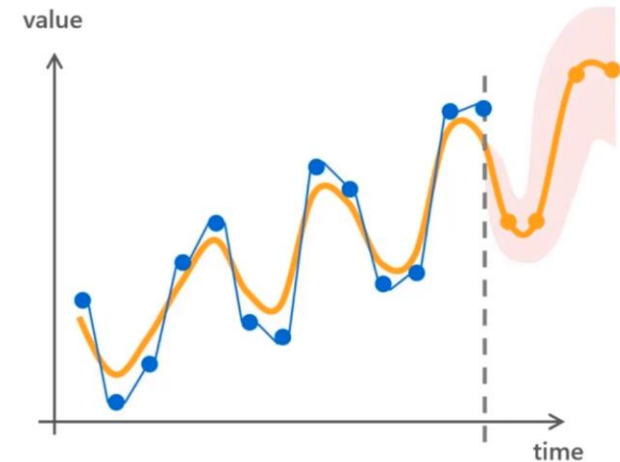
Serie de tiempo = serie temporal = serie

Definida como una secuencia **ordenada** de puntos de datos que son mediciones **sucesivas** de una variable (o varias) realizadas a **intervalos de tiempo igualmente espaciados**.

- Si medimos una variable → series univariadas
- Si son más variables → multivariado

Ejemplos:

- las ventas *semanales* de un producto en una tienda (frecuencia: 52)
- el rendimiento de un equipo en los últimos *años*
- el valor *diario* de cierre del Dow Jones o Colcap.
- Número de casos nuevos positivos COVID por *día*



Objetivo del análisis de las series de tiempo:

Fuente: https://www.youtube.com/watch?app=desktop&v=gxoZ-vWUlh4&ab_channel=ArcGIS

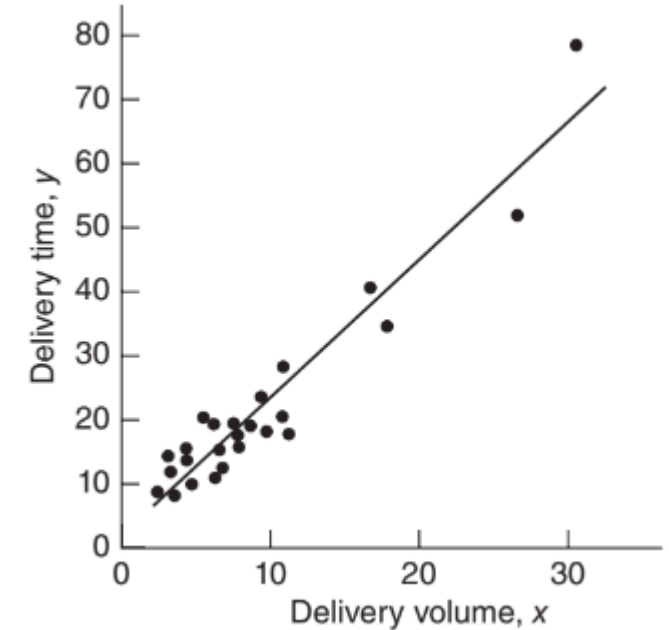
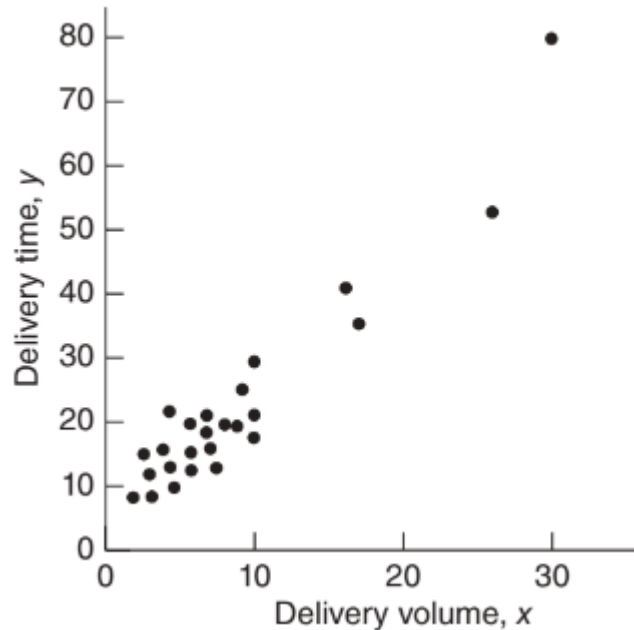
Describir el comportamiento de la serie de tiempo, mediante un **modelo**, que permita realizar **predicciones** con precisión.

Temáticas del curso



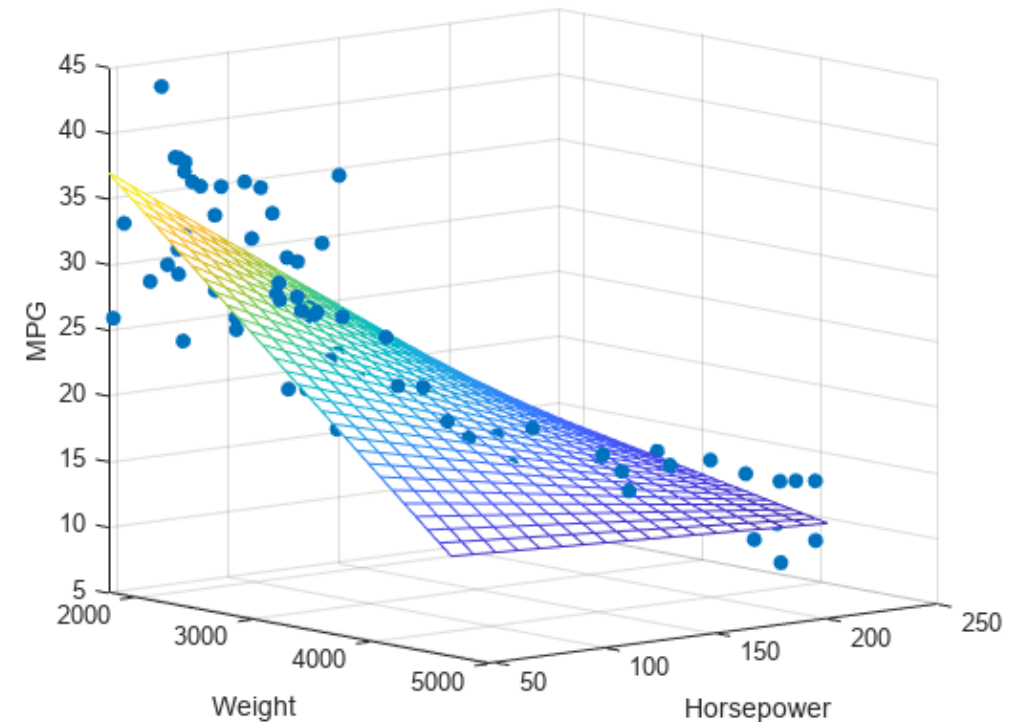
1. Modelos de Regresión Lineal Simple (MRLS)

- Introducción a los MRLS
- Correlación
- Residuales
- Estimación de parámetros
- Utilidad del MRLS
- Análisis e inferencia con MRLS
- Validación de supuestos



2. Modelos de Regresión Lineal Múltiple (MRLM)

- Introducción a los MRLM
- Estimación de parámetros
- Prueba de utilidad del MRLM
- Análisis e inferencia con MRLM
- Validación de supuestos y puntos atípicos
- Multicolinealidad
- Métodos para la selección de variables

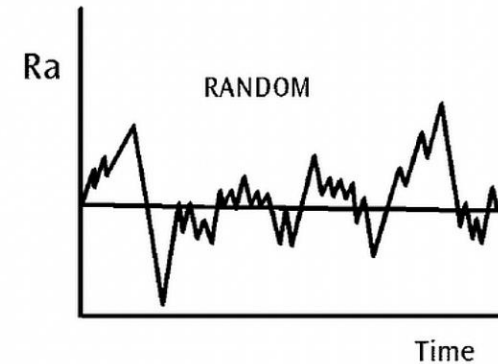
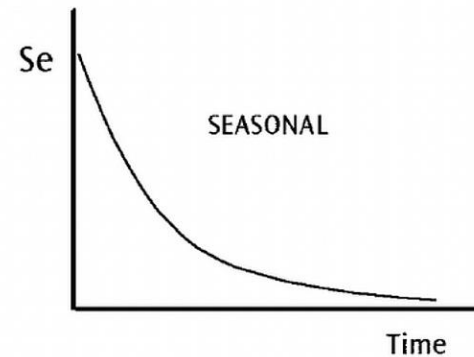
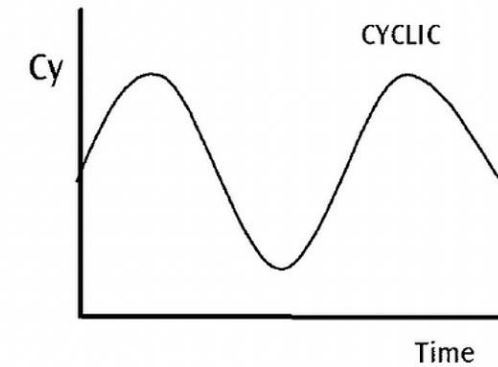
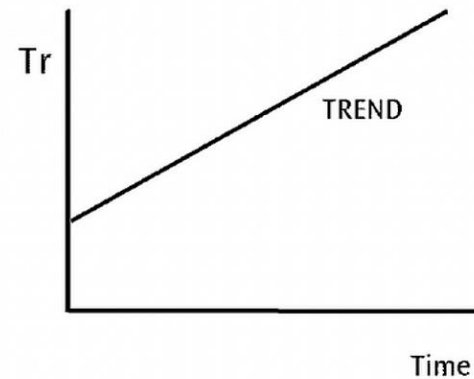


Temáticas del curso



3. Series de Tiempo (ST)

- Introducción
- Función de autocorrelación
- Series estacionarias
- Modelos ARMA
- Series No estacionarias
- Modelos ARIMA y SARIMA
- Suavizados y filtros
- Validación de supuestos



Metodología y evaluación



CLASES

- Componente teórico
- Talleres
- Actividades en clase (seguimiento)
- Manejo de software: R y Excel

**Talleres en R
30 %**

5 de agosto – Modelos RLS

18 de septiembre – Modelos RLM

9 y 28 de octubre – Series de Tiempo

**Seguimiento
10 %**

Talleres previos a los parciales, quizzes y actividades en clase

**10 % por
tema
Incluye
sustentación**

Metodología y evaluación



EVALUACIONES
60 %

3 evaluaciones

19 de agosto

- Modelos RLS

25 de septiembre

- Modelos RLM

4 de noviembre

- Series de Tiempo

**20 % por
tema**

Metodología y evaluación



Otras fechas importantes

Registro de notas en el sistema:

9 de agosto – Ingreso del 10%

6 de septiembre – Ingreso del 30%

3 de octubre – Ingreso del 40%

31 de octubre – Ingreso del 60 %

18 de noviembre – Ingreso del 100%

24 de octubre - Último día para cancelar materias o semestre

Proceso para la atención de reclamos acerca de las evaluaciones

1. En clase se entregará la evaluación para su revisión por parte del estudiante.
2. Si existen dudas sobre la calificación, devolver la evaluación a la docente.
3. Los reclamos se atenderán en el horario de asesoría.

Herramientas computacionales



[R Project Webpage](#)

[R Studio Desktop](#)



Excel

[Real Statistics Using Excel](#)

Ojo, vamos a usar estas herramientas de manera práctica, pero ¡el curso NO es sobre ellas!

Bibliografía



Modelos de regresión

- Devore, J.L., *Probabilidad y estadística para ingeniería y ciencias*. Cenage Learning Editores, 2008.
- Montgomery, D., Peck, E., Vining, G., *Introducción al análisis de regresión lineal*. CECSA, 2006.

Series de tiempo

- Giraldo, N.D., *Técnicas de pronósticos: aplicaciones con R*. Universidad Nacional de Colombia, 2006. [Disponible aquí](#)
- Shumway, R & Stoffer, D., *Time series análisis and its applications*. Springer, 2017.

Introducción a los modelos de regresión (MR)

Introducción a los modelos de regresión lineal



El primer paso para hacer un modelo de regresión lineal es identificar si las variables parecen tener relación o asociación → Diagrama de dispersión y análisis de correlación

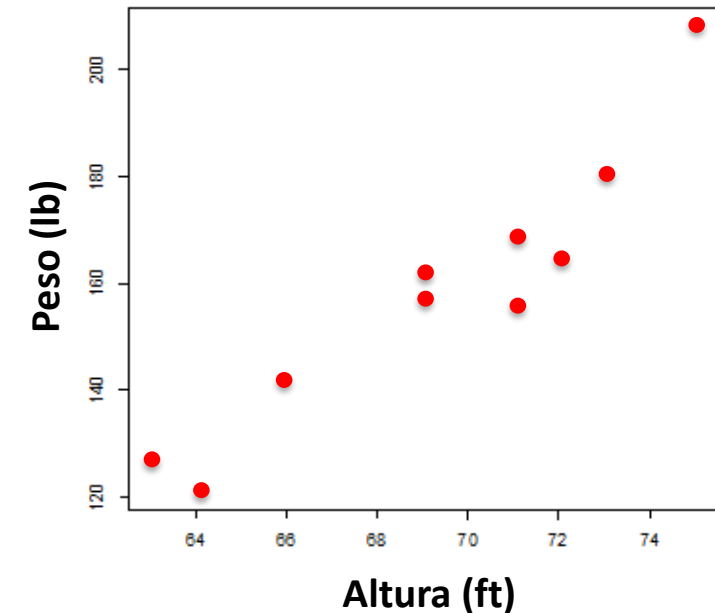
Ejemplo

Altura (en pies) y peso (en libras) de una muestra de 10 estudiantes en USA

Variable respuesta: peso (lb)

Variable explicativa: altura (ft)

Altura (ft)	Peso (lb)
63	127
64	121
66	142
69	157
69	162
71	156
71	169
72	165
73	181
75	208



¿Se puede decir que hay relación entre las variables altura y peso?

Regresión vs. Correlación



Análisis de Correlación

- Indica la fuerza y dirección de la **relación lineal** entre dos o más variables
- Existen *diferentes tipos de correlación*, la correlación simple, la correlación múltiple y la correlación parcial
- **No significa causalidad**

Análisis de Regresión

- Una vez identificada una correlación fuerte entre dos variables se puede hacer análisis de regresión: modelizar con una ecuación la relación causal

Análisis de correlación



Coeficiente de correlación (o de Pearson)

- Denotado por **r** o **R**
- Indica la intensidad, o fuerza, de la relación lineal entre dos variables **Y** y **X**
- No depende de las escalas de medición de las variables
- Los valores de **r** oscilan entre -1 y 1.

Entre más cercano sea el valor a 1 o -1 mejor es el ajuste de la recta de regresión lineal.

¿Cómo calcularlo?

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 * \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad \frac{s_{xy}}{n-1} \rightarrow \text{Covarianza de x y y}$$

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \frac{s_{xx}}{n-1} \rightarrow \text{Varianza de x}$$

$$s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad \frac{s_{yy}}{n-1} \rightarrow \text{Varianza de y}$$

Análisis de correlación

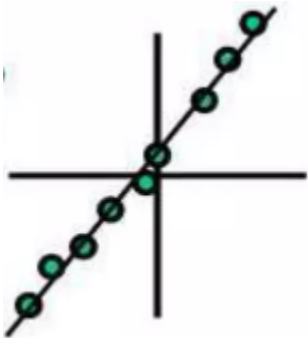


Coeficiente de correlación (o de Pearson)

Entre más cercano sea el valor a 1 o -1 mejor es el ajuste de la recta de regresión lineal.

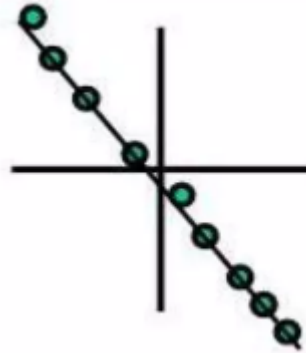
Correlación lineal positiva

$$r \approx 1$$



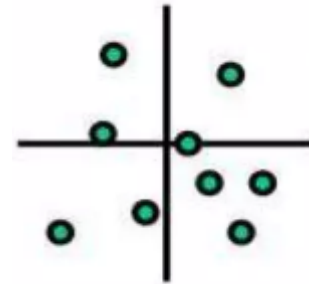
Correlación lineal negativa

$$r \approx -1$$



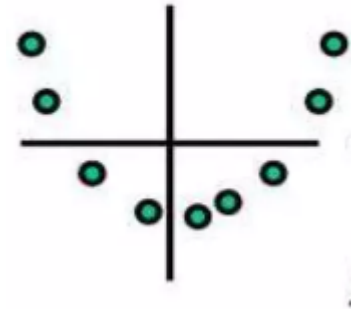
Ninguna correlación lineal

$$r \approx 0$$



Correlación curvilínea

$$r \approx 0$$

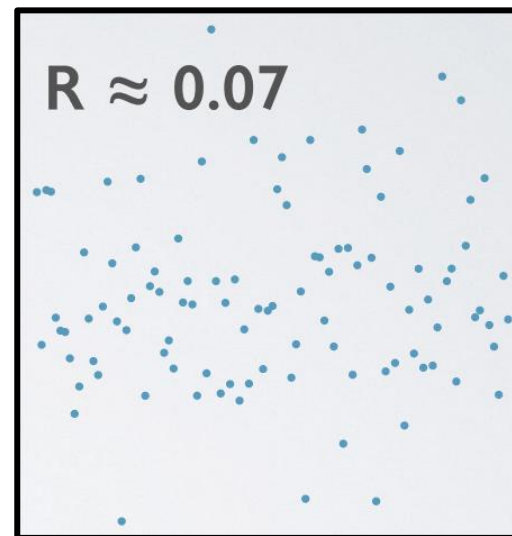
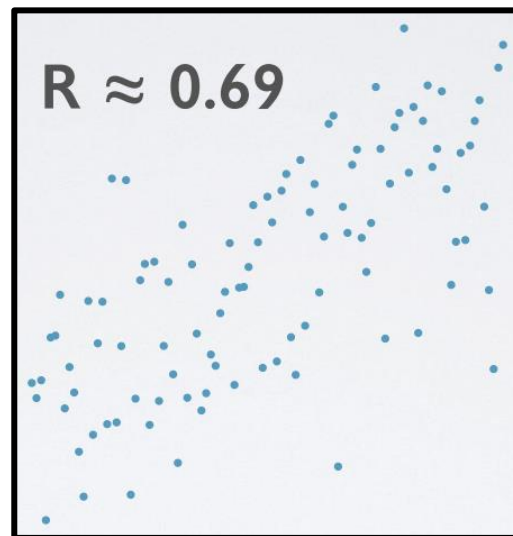
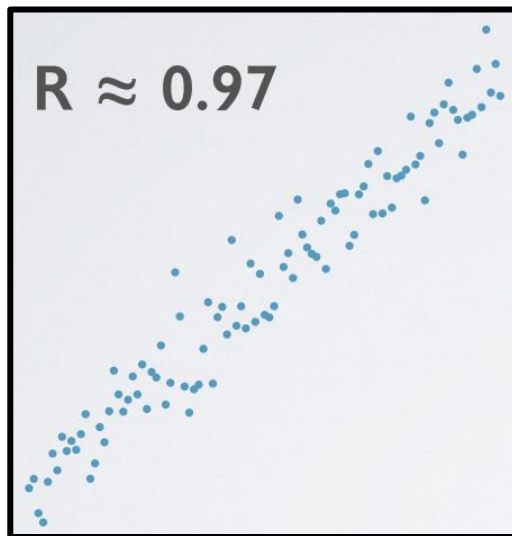


Análisis de correlación



Coeficiente de correlación (o de Pearson)

Entre más cercano sea el valor a 1 o -1 mejor es el ajuste de la recta de regresión lineal.



Fuente: <https://www.coursera.org/learn/linear-regression-model>

Pequeña guía para valores de r

$0 \leq r \leq 0.5$: Asociación débil

$0.5 < r < 0.8$: Asociación moderada

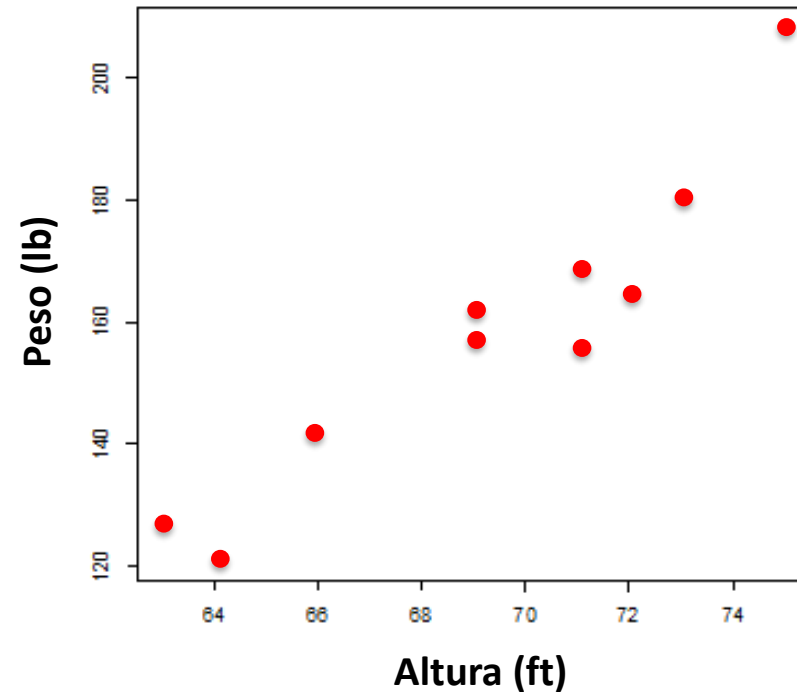
$0.8 \leq r \leq 1$: Asociación fuerte

Análisis de correlación



Coeficiente de correlación (o de Pearson)

¿Cuál de los siguientes valores de R es el más apropiado para describir la correlación entre la altura (en pies) y peso (en libras) de los estudiantes?

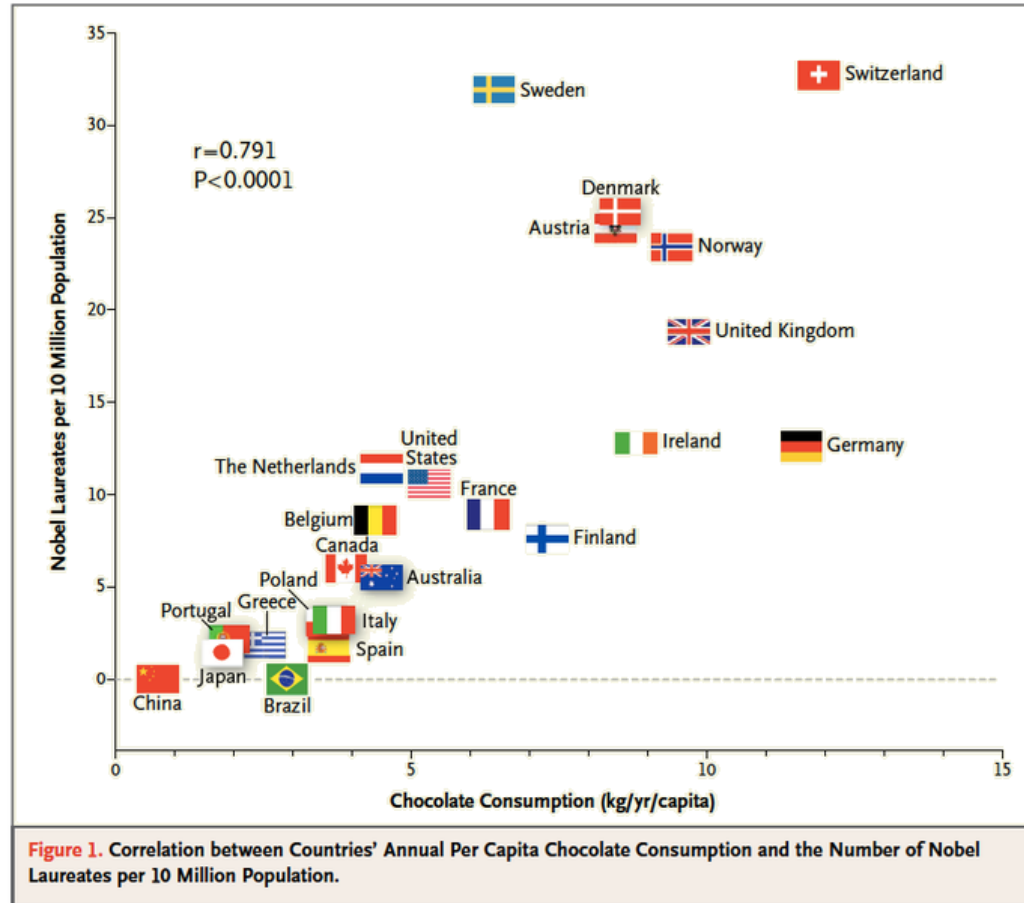


- A) - 0,6
- B) 0,90
- C) 0,1
- D) - 0,02
- E) 1,5

Correlación no implica causalidad



Does chocolate
make you clever?



Correlación

Sólo describe una asociación lineal entre las variables

Causalidad

Basada en la teoría

Necesaria para poder predecir una variable a partir de otra

Es decir, la **regresión** sólo puede utilizarse si se asume una **relación causal**

Fuentes: New England Journal of Medicine.

<https://www.bbc.com/news/magazine-20356613#:~:text=Chocolate%20consumption%20and%20Nobel%20laureates,of%200.0001.%22%20says%20Messerli.>

Otro ejemplo curioso: https://www.instagram.com/reel/C-v5hBXvOKB/?utm_source=ig_web_copy_link

23

Análisis de regresión



- Tras detectar que entre dos (o más) variables existe una relación (gráfico de dispersión y correlación) el siguiente paso es intentar modelizarla.
- El proceso consiste en ajustar la recta al conjunto de datos y crear una ecuación que permita predecir, de forma aproximada, el valor de la **variable dependiente** cuando se conoce el valor de la **variable predictora**
- La forma más sencilla es mediante una **ecuación** lineal $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$
 - El caso más simple para una única variable es una recta y recibe el nombre de regresión lineal *simple*.
 - Cuando $k > 1$ se llama regresión lineal *múltiple*.

Análisis de regresión



- A la **ecuación** ajustada se le llama **modelo de regresión**
 - El ajuste se realiza mediante el método de los **mínimos cuadrados**
 - Se debe evaluar si el modelo es útil para describir la variable Y en función de X
- Se debe comprobar que el modelo cumple unos supuestos:
 - Validación de supuestos sobre los residuos (errores): **normalidad, varianza constante**
- También es necesario evaluar la calidad del ajuste que presenta el modelo:
 - **Coeficiente de determinación r^2 o R^2** : cuánta variabilidad de los datos explica el modelo lineal



MUCHAS GRACIAS

Natalia Jaramillo Quiceno

e-mail: natalia.jaramilloq@upb.edu.co