



 Software Engineering |  IT Staffing |  IT Academy |  IT Consulting

Proyecto apoyado por  
**CORFO**



# Especialización en Ciencia de Datos

## Módulo Inferencia Estadística

2021



# Presentación del módulo

Módulo:

“Inferencia Estadística”

Duración:

24 horas 6 días

Competencia del módulo:

“Aprender inferencia estadística para resolver problemas analíticos, resultados con visualizaciones en Python ”



# Presentación del módulo

## Aprendizajes del módulo:

1. Explicar los principales conceptos de probabilidad asociados a un evento aleatorio.
2. Realizar cálculos de probabilidad seleccionando la distribución de probabilidad requerida para resolver un problema.
3. Realizar cálculos de probabilidad utilizando la distribución muestral para resolver un problema.
4. Realiza estimaciones de la media de una población utilizando intervalos de confianza a partir de una muestra aleatoria.
5. Realiza una prueba de hipótesis para probar la validez de una aseveración acerca de un parámetro de la población.



# Presentación del módulo

## Evaluación del Módulo

1. 60% trabajo en clases (tarefas y desafíos)
2. 40% trabajo final



# Presentación del módulo

## Contenidos y sesiones

1. Concepto de Probabilidad
2. Distribución de Probabilidad
3. Muestras y muestreos
4. Inferencia e Intervalos de Confianza
5. Test de Significancia



# Consejos para el éxito!

- Practicar lo más que puedan (recomendación: media hora por día)
- No sentirse frustrado o con miedo, (Es parte de aprender tener errores al inicio)
- Ser parte de una comunidad, compartir errores y soluciones

**“Tu FUTURO se decide por lo que haces HOY, no MAÑANA”**



# Inferencia Estadística





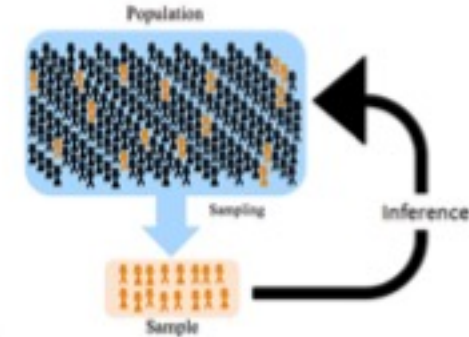
- ❖ Los procesos en nuestras vida y entorno generan datos → los datos representan trazos de procesos que ocurren en el mundo real.
    - ¿cómo describimos y explicamos estos procesos para resolver los problemas que abordamos?
  - ❖ Hay dos fuentes de aleatoriedad e incerteza:
    - La aleatoriedad e incerteza propia del proceso
    - La incerteza asociada a los métodos de recolección de datos.
- Necesitamos procedimientos, métodos y teoremas que nos permitan extraer significado e información a partir de data generada por procesos estocásticos (aleatorios).



- ❖ En DS, buscamos comúnmente generar conclusiones sobre una población, a partir de una muestra (ruidosa).

- ❖ Ejemplos:

- ¿Quién va a ganar las elecciones?
- Pronósticos meteorológicos
- Comportamiento de consumidores o mercados
- ¿Son precisas nuestras estimaciones sobre una cierta población...
- ¿Cuál es el impacto de una determinada política? ¿Hay un efecto real?



→ **Algunas dificultades:**

- ¿Es la muestra representativa de la población sobre la cual queremos inferir conclusiones?
- ¿Hay variables (no) observadas/(des)conocidas que contaminan nuestras conclusiones?
- ¿Hay sesgos (bias) sistemáticos debidos a datos faltantes, o diseño del estudio?
- ¿Qué aleatoriedad existe en los datos, y cómo las consideramos?
- ¿Estamos tratando de estimar un modelo subyacente del fenómeno estudiado?





 Software Engineering |  IT Staffing |  IT Academy |  IT Consulting

Proyecto apoyado por  
**CORFO**

