Introducción al Curso
Clase 1

Nicolás Mejía M.
n.mejia10@uniandes.edu.co

Probabilidad y Estadística II
Departamento de Ingeniería Industrial
Universidad de Los Andes, Bogotá, Colombia

2020-19

Outline

Contexto

Para tomar decisiones se necesita información. Dicha información es desconocida, y por tanto es necesario utilizar datos recolectados de alguna fuente que ayuden a obtener dicha información. Es decir se tiene una muestra que corresponde a una fracción de la población total.

La inferencia, es la proceso en el cual se pasa de observaciones particulares de un fenómeno a la generalización de una teoría.

Contexto

El papel de la Estadística es asegurarse de que el proceso con el cual se sacan conclusiones generales de un fenómeno queden bien hechas.

El principio fundamental, es que las observaciones que vemos son solo información parcial del fenómeno, por lo cual, hay RIESGO de equivocarse (ERROR).

El papel de la estadística es asegurar que los procesos de inferencia sean los mejores posibles, esto es, que la probabilidad de cometer errores sea la más pequeña.

Contexto

Supuesto Fundamental:

Los DATOS son un número finito de realizaciones de un FENÓMENO ALEATORIO que se puede repetir de manera independiente un número indefinido de veces.

Objetivo Fundamental:

A partir de los datos se quiere inferir al modelo de PROBABILIDAD que explica el fenómeno aleatrio de interés.

Outline

Inferencia Paramétrica

En general suponemos que los datos son números y por lo tanto el fenómeno de interés (Población) se entiende como una Variable Aleatoria.

$$\text{Poblacion} \;=\; Y \;\sim f_Y(y)$$

Ahora el problema es aproximarse a la función de probabilidad (masa o densidad) $f_y(y)$.

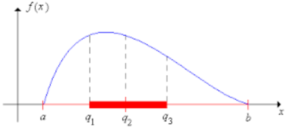
Ejemplo 1: Se quiere saber si una moneda está balanceada. Para esto se lanza la moneda $n = 100$ veces. Qué son los datos?: n realizaciones independientes de una variable con distribución $Y \sim Bernouli(p)$
Qué implica conocer el valor de p?

Inferencial Paramétrica

El caso de la moneda es muy sencillo.

Ejemplo 2: Suponga que Usted quiere emprender un negocio de Hamburgesas. Según sus cuentas, el negocio será rentable si en el largo plazo las ganancias mensuales superan $100 (en miles de pesos).
Si cada dato son las ganancias de cada mes, entonces creemos que:

$$X_1, X_2, \ldots, X_n \sim_{i.i.d.} f_X(x)$$



Para concentrar la información contenida en los datos, es mucho más conveniente asumir un modelo paramétrico. Por ejemplo:

$$X_1, X_2, \ldots, X_n \sim_{i.i.d.} f_X(x) \sim Normal(\mu, \sigma^2)$$

Inferencia Paramétrica

Si se asume que los datos obtenidos siguen una distribución conocida, el problema se reduce a estimar las constantes o parametros desconocidos de dicha distribución. En el ejemplo anterior que nuestros datos siguen una distribución Normal, se deben estimar dos parametros: (μ, σ^2).

Estimadores

Son funciones de los datos que estiman los parámetros de interés. Es decir, nos permiten aproximarnos al valor real de dicho parametro:

$$\hat{\mu} = g_1(X_1, X_2, \ldots, X_n) = \bar{X}$$

$$\hat{\sigma}^2 = g_2(X_1, X_2, \ldots, X_n) = S^2$$

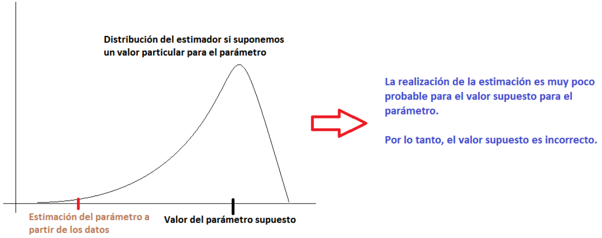
Como se sabe que los datos son realizaciones de un fenomeno aleatorio, los estimadores tambien resultan ser Variables Aleatorias

Outline

Inferencia con medición de Riesgo

Como variables aleatorias, los estimadores tienen asociados una función de distribución de probabilidad. Al conocer dicha distribución, podemos cuantificar el riesgo correspondiente a la veracidad de nuestras conclusiones.

Las dos formas típicas de inferir con medición del riesgo (probabilidad) de equivocarnos en nuestras generalizaciones son: Intervalos de Confianza y Pruebas de Hipótesis. Las dos se basan en el mismo principio:



Inferencia con medición de Riesgo

Intervalos de Confianza

Los intervalos de confianza nos dan unos limites en los cuales se puede encontrar el parámetro:

$$IC(\theta, 1-\alpha) = [a, b]$$

Con una confianza del $100 * (1-\alpha)\%$ el parámetro θ se encuentra entre a y b.

Pruebas de Hipótesis

Las pruebas de hipótesis nos permiten verficar la certeza sobre el valor de un parámetro

$$H_0 : \theta = c$$
$$H_1 : \theta \neq c \quad or \quad \theta > c \quad or \quad \theta < c$$

Inferencia con medición de Riesgo

Ejemplo: Suponga que Usted quiere emprender un negocio de venta de Hamburguesas. Según sus cuentas, el negocio será rentable si en el largo plazo las ganancias mensuales superan $100 (en miles de pesos).
Si cada dato son las ganancias de cada mes, entonces creemos que:

$$Y_1, Y_2, \dots, Y_n \sim_{i.i.d.} f_Y(y) \sim Normal(\mu, \sigma^2)$$

Dado que se miden las ganacias durante 9 meses, cuando los datos se realizan (se convierten en números) se obtiene que:

$$\bar{Y} = 103 \qquad s^2 = 100$$

¿Es el negocio rentable?