Probabilidad y Estadística

MSC RENZO CLAURE ARACENA

1

Ejemplo motivador:

¿Quién va a ganar las elecciones?

En cada elección importante, a los encuestadores les gustaría saber, antes de la elección real, quién va a ganar.

Aquí, el objetivo de estimación (el estimador) es claro, el porcentaje de personas en un grupo particular (ciudad, estado, provincia, país u otro grupo electoral) que votará por cada candidato. No podemos encuestar a todos. Incluso si pudiéramos, algunos encuestados podrían cambiar su voto para cuando se realice la elección.

¿Cómo recopilamos un subconjunto razonable de datos y cuantificamos la incertidumbre en el proceso para producir una buena estimación de quién ganará?

MSC RENZO CLAURE ARACENA

2

Algunos apuntes importantes

- ¿Es la muestra, una parte representativa de la población sobre la cual deseamos hacer inferencias?

- ¿Existen variables conocidas y observadas, conocidas y no observadas o desconocidas y no observadas que contaminen nuestras conclusiones?

- ¿Existe sesgo sistemático creado por datos faltantes o por el diseño o ejecución del estudio?

- ¿Qué aleatoriedad existe en los datos y cómo la utilizamos o la ajustamos? Aquí, la aleatoriedad puede ser explícita a través de la aleatorización o el muestreo aleatorio, o implícita como la agregación de muchos procesos complejos desconocidos.

La inferencia estadística requiere navegar a través de un conjunto de suposiciones y herramientas, y posteriormente pensar en cómo obtener conclusiones a partir de los datos.

Objetivos de la inferencia

- Estimar y cuantificar la incertidumbre de un estimador de una cantidad poblacional (la proporción de personas que votarán por un candidato).
- Determinar si una cantidad poblacional es un valor de referencia ("¿es efectivo el tratamiento?").
- Determinar el impacto de una política, promoción, acciones, etc. ("¿si reducimos los niveles de contaminación, disminuirán las tasas de asma?").
- Estimar sobre la probabilidad de que ocurra algo.

Preguntas

1. ¿El objetivo de la inferencia estadística es?
   - Inferir hechos sobre una población a partir de una muestra.
   - Inferir hechos sobre la muestra de una población.
   - Calcular cantidades de muestra para comprender sus datos.
   - Torturar a los estudiantes de la maestrìa.
2. ¿El objetivo de la aleatorización de un tratamiento en un ensayo aleatorio es?
   - Realmente no hace nada.
   - Obtener una muestra representativa de sujetos de la población de interés.
   - Equilibrar las covariables no observadas que puedan contaminar la comparación entre los grupos tratados y de control.
   - Para agregar variación a nuestras conclusiones.
3. La probabilidad es?
   - Cantidad de población que potencialmente podemos estimar a partir de los datos.
   - Una cantidad de datos que no requiere la idea de una población.

Trabajo entregable:

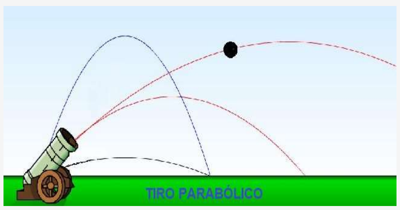
- Debe conseguir una base de datos en formato CSV.
- Describir la fuente y las características de la información que contiene (Año, Cantidad de Registros y Variables, Fuente, Objetivo).
- Describir el tipo de variables:
  - Cualitativas
  - Cuantitativa

- Variables Cualitativas
- Variables Cuantitativas

Revisión de probabilidad

Tipos de Experimentos

Experimento determinístico



Experimento aleatorio

Espacio muestral de lanzar un dado

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Formas de calcular la probabilidad

Método teórico

- División de la cantidad de casos favorables y casos posibles
- Por ejemplo: calcule la probabilidad de que el resultado de un dado sea menor o igual que dos

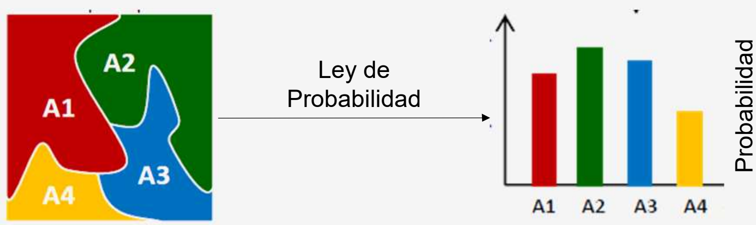
Método empírico

- Se realiza el experimento una cantidad finita de veces y se obtiene la frecuencia relativa de ocurrencia
- Por ejemplo: calcule la probabilidad de que el resultado de un dado sea menor o igual que dos

$$\Pr\{E_2\} = \lim_{rep\to\infty} fr\{éxitos\} = \lim_{rep\to\infty} \frac{éxitos}{rep}$$

Revisión de teoría de la probabilidad

- La probabilidad es una ley que asigna números a la ocurrencia a largo plazo de fenómenos aleatorios después de repeticiones consecutivas
- El espacio muestral de un experimento aleatorio es el grupo de todas las posibles salidas, ocurrencias o eventos

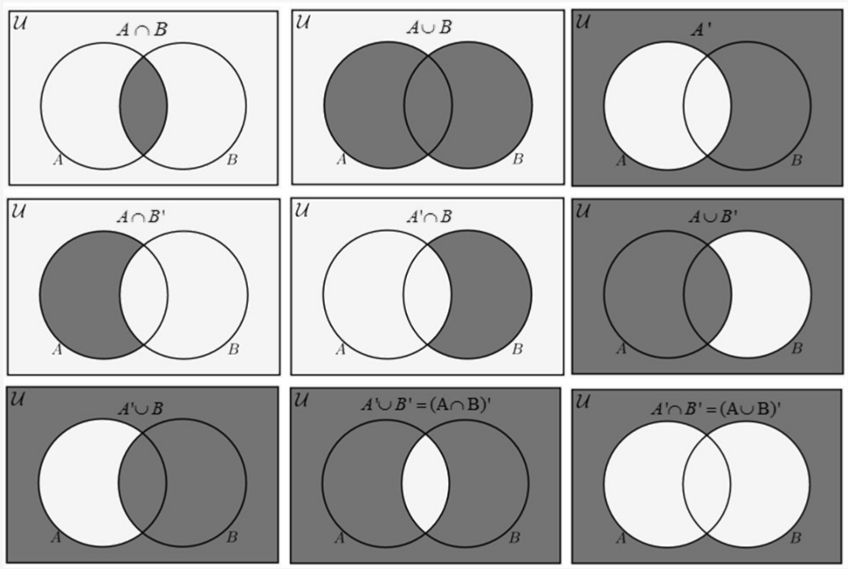


Espacio muestral (S) Eventos

Las tres reglas de Kolmogorov

- Según el brillante matemático ruso, padre de la probabilidad, Kolmogorov, se requieren solo tres reglas para entender cono se comporta la probabilidad:
- Considere un experimento con una salida aleatoria, la probabilidad toma que el resultado de un posible evento y:
  - Se le asigna número entre 0 y 1, [0,1]
  - Requiere que la probabilidad de que algún evento ocurra en cualquier momento es de 1
  - Requiere que la probabilidad de que dos eventos disjuntos ocurran es la suma de ambos eventos

Diagramas de Ven

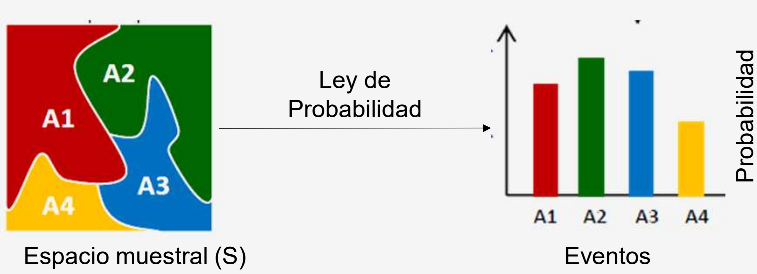


MSC RENZO CLAURE ARACENA

13

Principales axiomas

- Axioma 1: $1 \geq P[A_i] \geq 0$
- Axioma 2: P[S]=1
- Axioma 3: $A_i \cap A_j = \emptyset \rightarrow P[A_i \cup A_j] = P[A_i] + P[A_j]$



MSC RENZO CLAURE ARACENA

14

Bibliografía

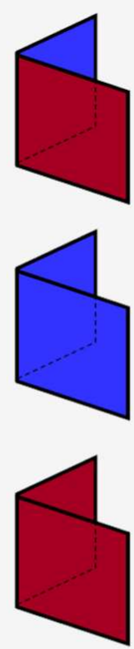
- Estadística, de Murray y Spieguel, cuarta Edicion
- Estadística Matemática con Aplicaciones, 6ta Edición - John E. Freund LEP
- Introduction to Probability, Dimitri P. Bertsekas and John N. Tsitsiklis, SECOND EDITION

15

Ejercicio

- Se cuenta con tres tarjetas. La primera tiene ambas caras de color Rojo, la segunda tiene ambas caras de color Azul y la tercera tiene una cara Roja y la otra Azul.

1. Cuál es el espacio muestral para la extracción de una tarjeta a la vez.
2. Cuál es el espacio muestral de las caras de tres tarjetas extraídas a la vez
3. Cuál es la probabilidad de que en la primera extracción extraigamos una tarjeta con colores distintos en ambos lados
4. Cuál es la probabilidad de obtener una cara de color rojo en el primer intento



16

Propiedades

- $P[A^C] = 1 - P[A]$
- $P[A] \leq 1$
- $P[\emptyset] = 0$
- $Dado\ \{A_1 \dots A_N\}, \{A_i \cap A_j = \emptyset, \forall ij\} \Rightarrow P\left[\bigcup_{k=1}^{N} A_k\right] = \sum_{k=1}^{N} P[A_k]$
- $P[A_1 \cup A_2] = P[A_1] + P[A_2] - P[A_1 \cap A_2]$
- $P\left[\bigcup_{k=1}^{N} A_k\right] =$
  $\sum_{k=1}^{N} P[A_k] - \sum_{j<k}^{N} P\left[A_j \cap A_k\right] + \cdots + (-1)^{N+1} P[A_1 \cap A_2 \dots \cap A_N]$
- $A_1 \subset A_2 \Rightarrow P[A_1] \leq P[A_2]$

17

18

Ejercicio 1:

- En general la población tiene 3% de probabilidad de tener apnea de sueño y el 10% tiene probabilidad de tener temblores en las piernas cuando duerme RLS. ¿se puede deducir que el 13% de la población tiene alguno de estos problemas de sueño?, ¿Se pueden sumar estas probabilidades?

Ejercicio 2:

- Un dado de seis caras está numerado del 1 al 6. Si lanzas el dado dos veces, ¿cuál es la probabilidad de obtener un número par en el primer lanzamiento y un número impar en el segundo lanzamiento?

Ejercicio 3:

- En una baraja estándar de 52 cartas, se extrae una carta al azar. ¿Cuál es la probabilidad de obtener una carta roja o una carta con un número par?

Ejercicio 4:

- En una urna hay 5 bolas rojas y 3 bolas azules. Se extraen dos bolas al azar sin reemplazo. ¿Cuál es la probabilidad de obtener una bola roja y una bola azul en ese orden?

Ejercicio 5:

- En una caja hay 8 bolas, de las cuales 3 son rojas y 5 son verdes. Se extraen dos bolas al azar con reemplazo. ¿Cuál es la probabilidad de obtener dos bolas verdes?

MSC RENZO CLAURE ARACENA

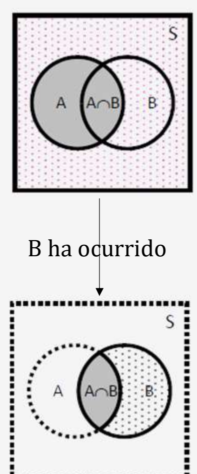
23

Probabilidad Condicional

- Sean A y B dos eventos, la probabilidad del evento A cuando ya es conocido que el evento B ha sucedido es:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \qquad P[B] > 0$$

  - La probabilidad de A dado B
  - La probabilidad de A condicionado a B
- La evidencia de que B ha ocurrido tiene las siguientes implicaciones:
  - El espacio muestral S cambia, ahora es B
  - Entonces A se transforma en A ∩ B
  - Entonces P[B] re-normaliza los eventos que ocurren en B



MSC RENZO CLAURE ARACENA

24

Ejercicios

- Ejercicio 1: El 76 % de los estudiantes de Ingeniería Civil han aprobado resistencia de materiales y el 45 % aprobaron estática. Además, el 30 % aprobaron resistencia de materiales y estática. Si Camilo aprobó resistencia de materiales, ¿qué probabilidad tiene de haber aprobado también estática?
- Ejercicio 2: Si un hincha compra una camiseta y se da cuenta que está defectuosa , ¿Cuál es la probabilidad de que sea del Manchester?

	Calidad			
Equipo		Buenas	Defectuosas	Total
	Juventus	508	92	600
	Manchester	315	85	400
	Total	823	177	1000

Eventos independientes

- Dos eventos son independientes si es verdadero cualquiera de los siguientes enunciados :
  - $P(A\backslash B)=P(A)$
  - $P(B\backslash A)=P(B)$
  - $P(A\cap B)=P(A)P(B)$

Ejercicio

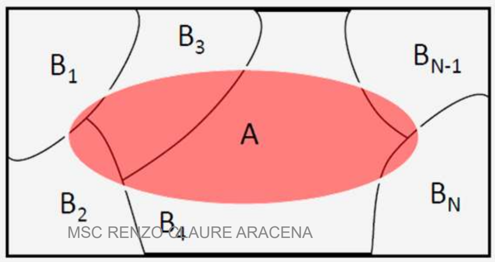
- Se sabe que E1 y E2 son conjuntos de un mismo espacio S, tal que $P(E1)=0.35$ y $P(E2)=0.45$ y $P(E1 \cap E2)=0.18$, ¿Son eventos independientes?

Ejercicio

- Utilizando el set de datos de autos responda la siguiente pregunta.
- ¿Cuál es la probabilidad de que un automóvil use gasolina como combustible dado que tiene tracción trasera?

Teorema de la probabilidad total

- Sean $B_1, B_2,.. B_n$, partes de S, mutuamente excluyentes.
- Sea A un conjunto de eventos dentro de S, que puede representarse como:
$$A = A \cap S = A \cap (B_1 \cup B_2 \dots B_N) = (A \cap B_1) \cup (A \cap B_2) \dots (A \cap B_N)$$
- Dado que $B_1, B_2,.. B_n$, son mutuamente excluyentes, entonces:
$$P[A] = P\{A \cap B_1] + P\{A \cap B_2] + \cdots + P\{A \cap B_N]$$
- Y por lo tanto:
$$P[A] = P[A|B_1]P[B_1] + \cdots P[A|B_N]P[B_N] = \sum_{k=1}^{N} P[A|B_k]P[B_k]$$



Ejercicios

- Ejemplo 1: En una tienda de ropa, el 60% de los clientes son mujeres y el 40% son hombres. El 25% de las mujeres y el 35% de los hombres compran zapatos. Si un cliente al azar compra zapatos, ¿cuál es la probabilidad de que sea un hombre?
- Ejemplo 2: Supongamos que en una universidad el 60% de los estudiantes son mujeres y el 40% son hombres. Además, se sabe que el 70% de las mujeres y el 80% de los hombres han aprobado un examen de matemáticas. Si elegimos un estudiante al azar y se sabe que aprobó el examen de matemáticas, ¿cuál es la probabilidad de que sea un hombre?

Teorema de Bayes

- Se asume que $\{B_1, B_2, .. B_n, \}$ son particiones de S
- Suponiendo que un evento A ocurre
- Cuál es la probabilidad que ocurra el evento B_j,
- Usando la definición de la probabilidad condicional y el Teorema de la probabilidad total, obtenemos:

$$P[B_j|A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A|B_j]P[B_j]}{\sum_{k=1}^{N} P[A|B_k]P[B_k]}$$

- Que es conocido como el Teorema de Bayes, uno de los principales enunciados de probabilidad y estadística

Cáncer

- Un día te sientes mal y vas al médico, después de numerosos exámenes te dan el diagnóstico.
- Te dicen que es Cáncer, un tipo de cáncer que solo afecta al 0,1% de la población.
- Preocupado le preguntas al medico que tan certero es el examen
- El médico te dice, que el 99% de las personas que, efectivamente tienen la enfermedad salen positivas en el examen, e identificará de manera errónea solo al 1,5% que no tiene la enfermedad.
- ¿Cuál es la probabilidad de que tengas la enfermedad?
- Si se realiza un segundo examen y también sale positivo ¿Cuál es la probabilidad de que tengas la enfermedad?

Teorema de Bayes y el reconocimiento estadístico de patrones

- La expresión del Teorema de Bayes, usada generalmente en el reconocimiento de patrones.

$$P[\omega_j|x] = \frac{P[x|\omega_j]P[\omega_j]}{\sum_{k=1}^{N} P[x|\omega_k]P[\omega_k]} = \frac{P[x|\omega_j]P[\omega_j]}{P[x]}$$

  - Donde ω_j es la j-ésima categoría y x es el vector de observaciones/características
- Una regla común decisión es escoger la clase ω_j cuyo $P[\omega_j/x]$ es el mayor.
  - De forma intuitiva elegimos la clase que es más probable de ocurrir, dada una observación x
- Los términos en el Teorema de Bayes son:
  - $P[\omega_j]$, probabilidad a priori (de la clase ω_j)
  - $P[\omega_j/x]$, probabilidad a posteriori (de la clase ω_j dad la observación x)
  - $P[x/\omega_j]$, verosimilitud (probabilidad de la observación x dada la clase ω_j)
  - $P[x]$, constante de normalización, no afecta a la predicción

Resolviendo el Problema

La matriz de confusión

CASOS TOTALES = P + N		PREDICCION Negativo (PN)	PREDICCION Positivo (PP)
REAL	Negativo (N)	Verdadero Negativo, VN, TN	Falso Positivo, FP
REAL	Positivo (P)	Falso negativo, FN	Verdadero Positivo, VP, TP

ACIERTO GLOBAL $\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{P}+\mathrm{N}} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$

SENSIBILIDAD
RECALL, TPR $\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = 1 - \mathrm{FNR}$

PRECISIÓN $\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = 1 - \mathrm{FDR}$

ESPECIFICIDAD $\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} = 1 - \mathrm{FPR}$

FNR $\mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{P}} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}} = 1 - \mathrm{TPR}$

F1 SCORE $\mathrm{F}_1 = 2 \times \frac{\mathrm{PPV} \times \mathrm{TPR}}{\mathrm{PPV}+\mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$

- CARACTERÍSTICAS
- Contrasta modelos vs realidad
- Ayuda a identificar las fuentes de desviación
- Tipos de error
- Permite extraer varios indicadores de precisión
  - Sensibilidad: Acierto de positivos
  - Especificidad: Acierto de negativos

MSC. RENZO CLAURE

35

Ejercicio

Supongamos que una empresa de telecomunicaciones ofrece dos planes de servicio: Plan A y Plan B. El 60% de los clientes tienen el Plan A, mientras que el 40% tienen el Plan B. Se sabe que el 80% de los clientes con el Plan A están satisfechos con el servicio, mientras que el 70% de los clientes con el Plan B están satisfechos.

Ahora, supongamos que un cliente se acerca a la empresa con una queja y se descubre que está insatisfecho con el servicio. Queremos calcular la probabilidad de que el cliente tenga el Plan B.

MSC RENZO CLAURE ARACENA

36

Breve introducción a Python

Jupyter

- Presentación
- Trabajo en Notebboks
- Markdowns

Strings y números

- Operaciones con cadenas de texto
- Operaciones aritméticas y consideraciones especiales

MSC RENZO CLAURE ARACENA

39

Librerías

- Las librerías que casi siempre usaremos:
  - Numpy
  - Pandas
  - Matplot
  - Seaborn
  - Scikitlearn
  - scipy

MSC RENZO CLAURE ARACENA

40

Listas vs Numpy Array vs Diccionarios

- Listas:
  - Son contenedores que pueden tener distintos tipos de datos como: Cadena, Enteros, Real, etc.
  - Se puede acceder a sus ítems a través de índices
- Numpy Array:
  - Son contenedores que solo pueden almacenar un solo tipo de datos
  - Se puede acceder a sus ítems a través de índices
  - Son más eficientes para cálculos ya que solo almacena el tipo de dato del "array" y no de cada valor.
- Diccionarios:
  - Son contenedores que pueden guardar una llave+valor(es)
  - Las llaves son únicas, per los valores pueden ser varios
  - Son muy útiles para crear Tablas de Datos (Dataframes) y para almacenar propiedades

Pandas

- Es una librería que facilita el manejo de datos, para prepararlos, manipularlos y/o transformarlos
- Lo hace a través de DataFrames, que se asemejan a una hoja de cálculo de Excel

Laboratorio

MSC RENZO CLAURE ARACENA

43

Introducción a Matplotlib

MSC RENZO CLAURE ARACENA

44

MatplotLib

- Módulo popular para gráficos basados en Matlab
- Jhon Hunter
- Arquitectura:

ARQUITECTURA

Backend
- FigureCanvas
Configura el área de dibujo
- Render
Para dibujar en el lienzo
- Event
Administra eventos, como clicks o botones de teclado

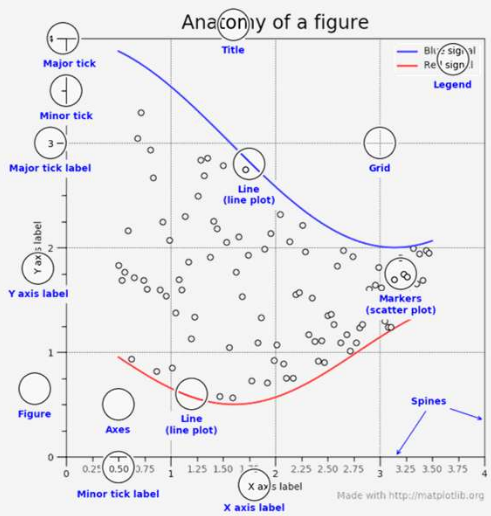
Artist
- FigureCanvas
Coloca los valores de los diferentes objetos del lienzo: Titulo, ejes, etiquetas
- Objeto Primitivo:
Líneas, rectángulos, Circulos y texto
- Objeto compuesto:
Axis, Tick, Axes y Figure

Scripting
- Creado por cientistas para hacer gráficos rápidos para análisis
- Requierre menos pasos o código
- Es el más recomendado para análisis por que genera gráficos buenos, de forma simple y rápida

Jupyter

- Interface Web para utilizar o programar en Python
- Compatible con Matplotlib
- Tiene funciones mágicas para presentar los gráficos
  - %Matplotlib inline
  - %Matplotlib notebook
- Pandas tiene funciones integradas que permiten una fácil ejecución de gráficos
- También utilizaremos Seaborn, otra librería de gráficos más digeridos, que utiliza a Matplot como base

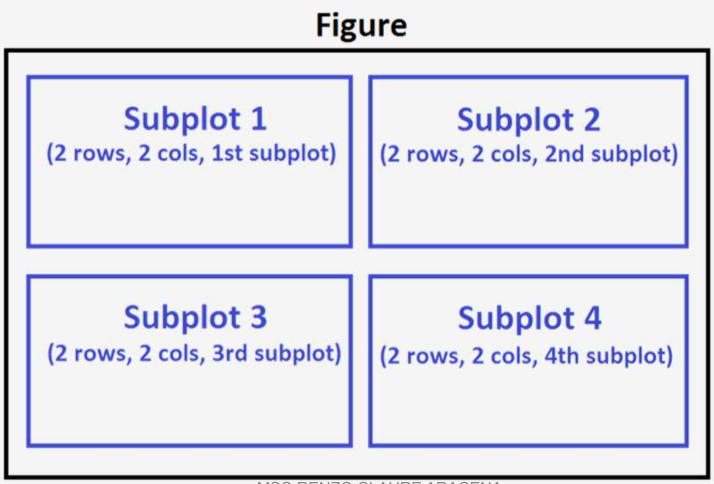
Terminología de un gráfico



Tipos de gráficos a revisar

- Líneas
- Dispersión
- Histogramas
- Diagramas de caja
- Pasteles
- Mapas de calor

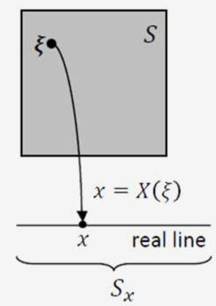
Subgráficos



Variables aleatorias

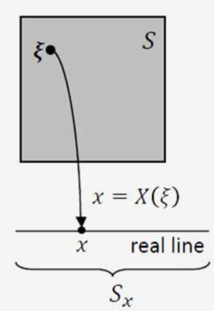
Variable aleatoria

- Cuando realizamos un experimento aleatorio, generalmente estamos interesados en alguna medición o atributo numérico del resultado, Ej:
  - Los pesos en una población de sujetos,
  - Los tiempos de ejecución al realizar pruebas de referencia de CPU
  - Los parámetros de forma al realizar reconocimiento automático de objetivos (ATR).
- Estos ejemplos llevan al concepto de variable aleatoria.
  - Una variable aleatoria X depende de una función que asigna un número real $X(\xi)$ a cada resultado ξ en el espacio muestral de un experimento aleatorio.
  - $X(\xi)$ se mapea desde todos los posibles resultados en el espacio muestral a la línea de números reales (real line).



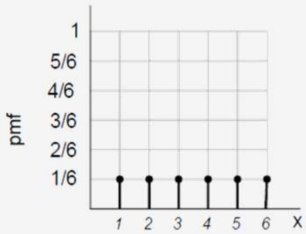
Variable aleatoria

- La función que asigna valores a cada resultado es fija y determinista, como en la regla: "contar el número de caras en tres lanzamientos de moneda". La aleatoriedad en X se debe a la aleatoriedad subyacente del resultado ξ del experimento.
- Las variables aleatorias pueden ser discretas, como el número resultante después de lanzar un dado, o continuas, como el peso de un individuo muestreado.

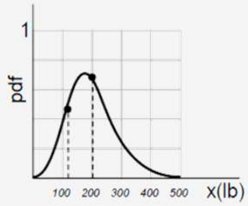


Función de densidad de probabilidad (pdf)

pmf para el lanzamiento de un dado



pdf para el Peso de una persona



- La función de densidad de probabilidad $f_X(x)$ de una variable aleatoria continua X, si existe, se define como la derivada de $F_X(x)$. Matemáticamente, se puede expresar como:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

- Para variables aleatorias discretas, el equivalente de pdf se denomina: función de masa de probabilidad

$$f_X(x) = \frac{\Delta F_X(x)}{\Delta x}$$

- Propiedades
  - $f_X(x) > 0$
  - $P[a < x < b] = \int_a^b f_X(x)dx$
  - $F_X(x) = \int_{-\infty}^x f_X(x)dx$
  - $1 = \int_{-\infty}^{\infty} f_X(x)dx$

- $f_X(x|A) = \frac{d}{dx}F_X(x|A)$ $where$ $F_X(x|A) = \frac{P[\{X<x\}\cap A]}{P[A]}$ if $P[A] > 0$

MSC RENZO CLAURE ARACENA

53

Función de distribución acumulada (cdf)

cdf para el lanzamiento de un dado

cdf para el Peso de una persona

- La función de distribución acumulada $FX(x)$ de una variable aleatoria X se define como la probabilidad del evento $\{X \leq x\}$. Matemáticamente, se puede expresar como

$$F_X(x) = P[X \leq x] \ , \ -\infty < x < \infty.$$

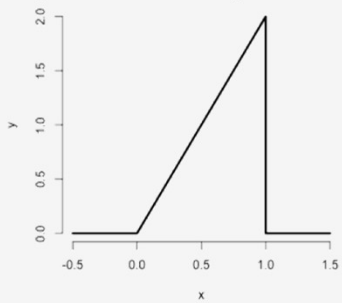
- Intuitivamente, $F_X(b)$ es la proporción a largo plazo de veces cuando $X(\xi) \leq b$.
- Propiedades de la cdf:
  - $0 \leq F_X(x) \leq 1$
  - $\lim_{x\to\infty} F_X(x) = 1$
  - $\lim_{x\to-\infty} F_X(x) = 0$
  - $F_X(a) \leq F_X(b)$ if $a \leq b$
  - $F_X(b) = \lim_{h\to 0} F_X(b+h) = F_X(b^+)$

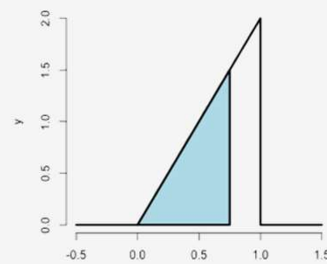
MSC RENZO CLAURE ARACENA

54

Ejemplo

- Suponga que la proporción de llamadas en un centro de llamadas, que pueden ser atendidas un día cualquiera, está dada por la función: $f(x) = 2x$ para $0 < x < 1$, x∈R.
- ¿Es esta una función de densidad de probabilidad?
- ¿Cuál es la probabilidad de que el 75% o menos de las llamadas sean atendidas?





Preguntas

- ¿Se puede sumar las probabilidades de los eventos para obtener la probabilidad de que ocurra al menos uno?
- Una variable aleatoria, X, es uniforme, una caja de 0 a 1 de altura 1. (De modo que su densidad es f(x) = 1 para 0<= x <= 1.) ¿Cuál es su mediana expresada con dos decimales?

Distribución de Probabilidad Acumulada y Sobrevivencia

- Como ya hemos visto, la función de probabilidad acumulada CDF es igual a:

$$F(x) = P(X \leq x)$$

- La función de sobrevivencia es el complemento, es decir es la probabilidad de que X sea mayor a x:

$$S(x) = P(X > x)$$

- Ejemplo: calcule la función de sobrevivencia de la distribución del ejemplo anterior:

Otro ejemplo

- Cuál es la probabilidad de que al lanzar un dado se obtenga el valor de 4 o menos.

Asintopia

Asintopia

- Asintótica es el término para el comportamiento de las estadísticas cuando el tamaño de la muestra tiende al infinito y son increíblemente útiles para inferencias y aproximaciones estadísticas simples.
- Ley de los grandes números: mientras más grande sea una muestra, mejor estima a la población. Ejemp, el lanzamiento de una moneda (Python)

Distribución muestral

- Distribución muestral:
  - Distribución de frecuencias de los estadísticos muestrales, con un mismo tamaño de muestra y extraídas aleatoriamente de la población
  - La naturaleza aleatoria de las muestras indica que las medias de las muestras oscilan alrededor de la media poblacional
  - La media del conjunto de las muestras posibles es igual a la media poblacional
  - Mientras más grandes son las muestras, las medias muestrales se acercan más a la media poblacional
- Error estándar EE:
  - Es la desviación estándar de las muestras o de la distribución muestral
  - Mide la desviación absoluta del valor verdadero desconocido

Valores esperados

- Son valores que caracterizan una distribución, el mas conocido y útil es la media, que representa el centro de masa de una distribución.
- Otro valor esperado es la varianza, que caracteriza la dispersión de una distribución
- Otro valor calculado es el sesgo, que identifica como o cuanta densidad esta presente en altos o bajos valores de una variable
- Pyhton

La media poblacional para variables aleatorias discretas

- Es el centro de su distribución. Para una variable aleatoria discreta *X* con *PMF p(x)* está definida según:

$$E[X] = \sum_x xp(x). \qquad \{x, p(x)\}$$

- Es también llamada el centro de masa, que es producto de los valores discretos por su proporción
- La media muestral es el estimador de la media poblacional, es el centro de masa de los datos disponibles:

$$\bar{X} = \sum_{i=1}^{n} x_i p(x_i), \qquad p(x_i) = 1/n.$$

MSC RENZO CLAURE ARACENA

63

La media poblacional para variables aleatorias continuas

- Si X es una variable aleatoria continua, la fórmula matemática incluye en este caso a la función de densidad:

$$\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx.$$

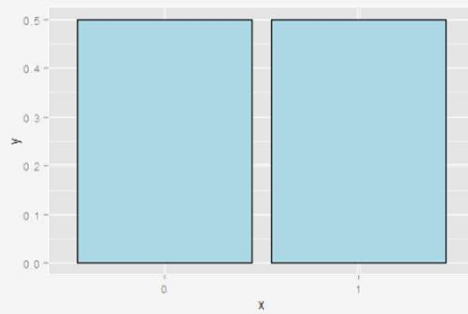
MSC RENZO CLAURE ARACENA

64

Ejemplo de cálculo de media poblacional

- Una moneda equilibrada:

$$E[X] = \sum_x xp(x). \qquad E[X] = .5 \times 0 + .5 \times 1 = .5$$



Función de masa de probabilidad

MSC RENZO CLAURE ARACENA

65

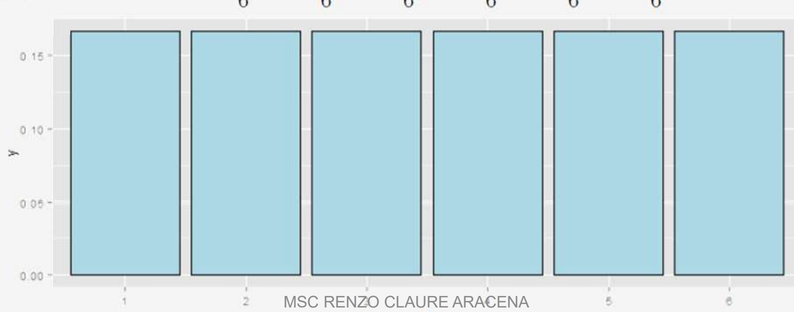
Ejemplo de cálculo de media poblacional

- Una moneda desbalanceada

$$E[X] = 0 * (1 - p) + 1 * p = p$$

- Ejemplo de lanzamiento de un dado:

$$E[X] = \sum_x xp(x). \qquad E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$



MSC RENZO CLAURE ARACENA

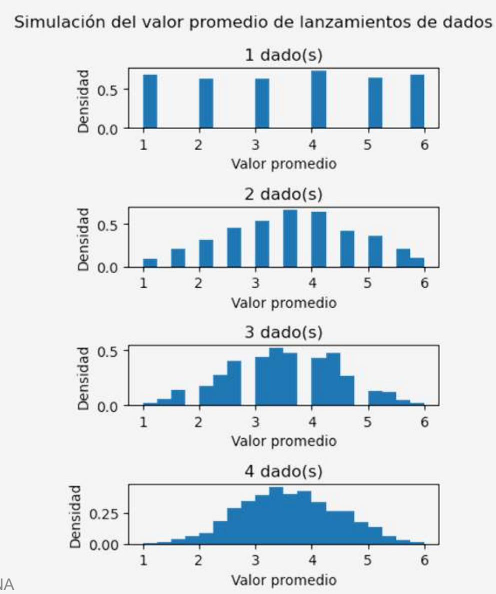
66

Ejemplo

- Calcule la esperanza matemática de la suma de dos dados

Distribución gausiana

- Dados: Suponga que arroja un dado, después 2 dados, 3 dados y 4 dados. La distribución de los promedios del resultado del lanzamiento tendría la siguiente forma:
- Código disponible en python



En resumen

- Los valores esperados son propiedades de una distribución
- La media poblacional es el centro de masa de la población.
- La media muestral es el centro de masa de los datos observados.
- La media muestral es una estimación de la media poblacional.
- La media muestral no está sesgada: la media poblacional de su distribución es la media que se intenta estimar.
- Cuantos más datos entren en la media de la muestra, más. concentrada su función densidad/masa y está alrededor de la media poblacional.

Ejercicio

- Lance diez dados y tome su promedio, luego repita este proceso una y otra vez y construya un histograma ¿en que estaría centrado y cuál seria su varianza?.
- Si una población tiene media μ, ¿cuál es la media de la distribución de promedios de 20 observaciones de esta distribución?

Varianza

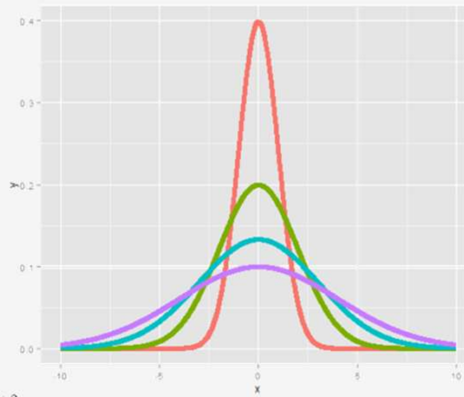
- La varianza es una medida de dispersión
- Si X es una variable aleatoria con media μ, la varianza de X se define como:

$$Var(X) = E[(X-\mu)^2] = E[X^2] - E[X]^2.$$

- Ejemplo: ¿cuál sería la varianza del lanzamiento de un dado?, sabemos que $E[X] = 3.5,$

$$E[X^2] = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = 15.17$$

$$Var(X) = E[X^2] - E[X]^2 \approx 2.92.$$



- La varianza muestral viene dada por: $S^2 = \frac{\sum_{i=1}(X_i - \bar{X})^2}{n-1}$

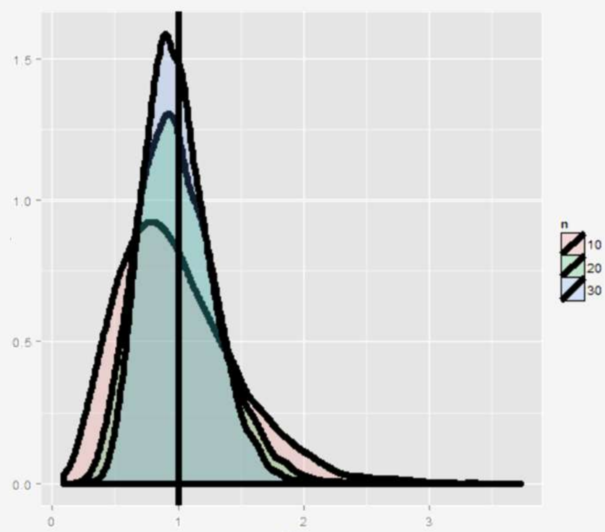
[Python]

MSC RENZO CLAURE ARACENA

71

Simulaciones

- Simulaciones de distribuciones normales

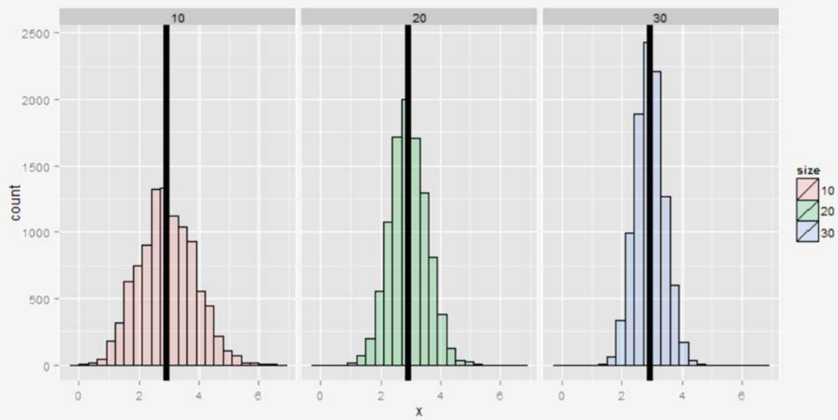


MSC RENZO CLAURE ARACENA

72

Simulaciones

- Simulación del lanzamiento de un dado, obteniendo la varianza y revisando su distribución



73

Preguntas

- ¿Qué es la desviación standard?
- ¿Por qué suele utilizarse más la desviación estándar para describir la dispersión?
- ¿Por qué se utiliza entonces la desviación cuadrática?

74

El error estándar de la media

- Hemos visto que la esperanza de las medias muestrales es igual a la media poblacional: $E[\bar{X}] = \mu$.
- La varianza de las medias muestrales es: $Var(\bar{X}) = \sigma^2/n$
- No podemos calcular la varianza de la medias, pero la varianza de la muestra es un buen estimador de la varianza poblacional
- Entonces podemos tomar la raíz cuadrada de la varianza, dividida por la cantidad de casos, como un estimador del error estándar de la media muestral

PROPIEDADES DE LA ESPERANZA

PROPIEDADES DE LA VARIANZA

Resumiendo

- La varianza de una muestra es una medida de dispersión específica de una muestra, siendo la varianza modificada una estimación de la varianza de toda la población basada en una muestra y se calcula de manera ligeramente diferente
- La varianza muestral S^2 estima la varianza poblacional σ^2
- La varianza de la distribución muestral de la media es: $\sigma^2/n.$
  - Cuál es su estimador lógico?
  - Cuál es el estimador lógico de su error estándar?
- Para la varianza de la muestra se divide la suma cuadrática de las diferencias por n-1, para corregir el sesgo
- La desviación muestral S, nos da una idea de que tan variable es la población
- El error estándar $S/\sqrt{n},$ indica como varían las medias muestrales alrededor de la media poblacional

Ejercicios

- Si tengo una muestra aleatoria de una población, ¿la varianza de la muestra es una estimación de?
  - La desviación estándar de la población.
  - La varianza de la población.
  - La varianza muestral.
  - La desviación estándar de la muestra.
- Obtienes una muestra aleatoria de n observaciones de una población y calculas su promedio. Deseas estimar la variabilidad de los promedios de n observaciones de esta población para comprender mejor qué tan precisa es esta estimación. ¿Necesitas recolectar promedios repetidamente para hacer esto?
  - No, podemos multiplicar nuestra estimación de la varianza de la población por 1/n para obtener una buena estimación de la variabilidad del promedio.
  - Sí, debes obtener promedios repetidos.

Distribución muestral, demostración

- X, S^2 son estimadores insesgados de μ y σ^2
- Demostración E[X]= μ

$$E[\overline{X}] = E[\tfrac{1}{n}\sum_{i=1}^{n} X_i] = \tfrac{1}{n}E[\sum_{i=1}^{n} X_i] = \tfrac{1}{n}\sum_{i=1}^{n} E[X_i] = \tfrac{1}{n}n \times \mu = \mu.$$

- Demostración E[S^2]= σ^2

$$E[S^2] = E[\tfrac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2)] = \tfrac{1}{n-1}\sum_{i=1}^{n} E[X_i^2] - \tfrac{n}{n-1}E[\overline{X}^2].$$

$$E[X_i^2] = \mathrm{var}(X_i) + E[X_i]^2 = \sigma^2 + \mu^2$$
$$E[\overline{X}^2] = \mathrm{var}(\overline{X}) + E[\overline{X}]^2 = \tfrac{\sigma^2}{n} + \mu^2.$$

$$E[S^2] = \tfrac{n}{n-1}(\sigma^2 + \mu^2) - \tfrac{n}{n-1}\left(\tfrac{\sigma^2}{n} + \mu^2\right) = \sigma^2.$$

Propiedades de la media

- La media de una constante es la misma constante
- La media de una variable por una constante, es igual a la constante por la media de una variable

Propiedades de la varianza

1. La varianza de una constante es cero
2. La varianza de una constante por una variable es igual a la constante al cuadrado por la varianza de la variable
3. La varianza de la suma de una variable con una constante es igual a la varianza de la variable
4. La Varianza Total de una variable dividida en dos (o más) grupos es igual a la suma entre la variación que existe al interior de los grupos (Inter) y la variación que existe entre los grupos (Intra)

Teorema central del límite

- La distribución de las medias muestrales tiene una distribución Normal, para tamaños muestrales suficientemente grandes.
- Los requisitos son:
  - Las muestras son independientes
  - Variables de tipo Cuantitativo: El tamaño muestral mínimo es de 30 casos (sólo para asimetrías moderadas)
  - Variables de tipo Categórico: Si n es la cantidad de individuos y k la proporción de una variable, entonces los tamaños mínimos son de:

    n*k>=5 o n*(1-k)>=5

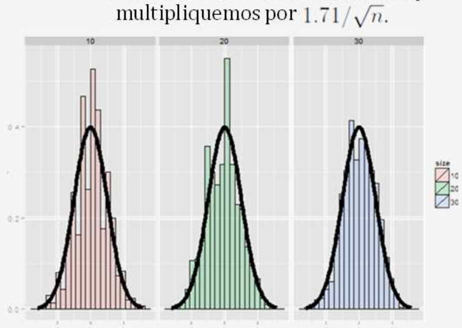
    (Python)

MSC RENZO CLAURE ARACENA

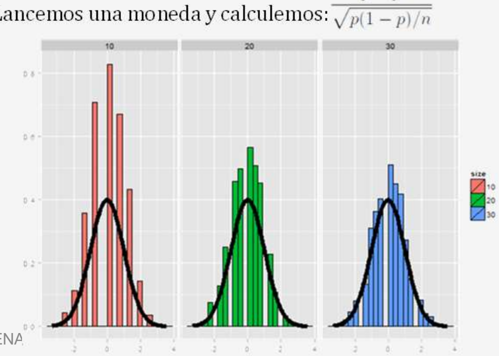
83

TCL Experimentos

- Lanzamiento de un dado
  - $\mu = E[X_i] = 3.5$
  - $Var(X_i) = 2.92.$
  - SE $\sqrt{2.92/n} = 1.71/\sqrt{n}.$
  - Lancemos un dado, restemos 3,5 y multipliquemos por $1.71/\sqrt{n}.$



- Lanzamiento de una moneda
  - $E[X_i] = p,$
  - $Var(X_i) = p(1-p),$
  - $\sqrt{Var(\hat{p})} = \sqrt{p(1-p)/n}.$
  - Lancemos una moneda y calculemos: $\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$



MSC RENZO CLAURE ARACENA

84

Teoría de la estimación estadística

85

Teoría de la estimación estadística

- Busca hallar el valor de un parámetro a partir de la información aportada por una muestra representativa de la población, sujeto a un valor determinado de precisión.
- El parámetro desconocido sigue una distribución aleatoria
- El error depende del proceso de medición:
  - Pese a que la muestra es parte de la población, es imposible que la represente de forma exacta
  - El azar de la toma de la muestra
  - Las circunstancias de la toma de muestra, errores, defectos, sesgos

86

Estimaciones

- **Estimación puntual:** Estima el parámetro mediante el cálculo de un único valor, como la media de la muestra. Pero es improbable que un valor muestral coincida de forma perfecta con el valor de la población
- **Estimación por intervalos:** Representa la estimación de los limites entre los cuales puede estar el valor verdadero. Este espacio se conoce como precisión y refleja el error de muestreo

MSC RENZO CLAURE ARACENA

87

Estimación puntual

- Un estimador puntual , es un valor utilizado para estimar los parámetros reales de la población a la que pertenece
- **Definición**: Un ESTIMADOR para θ, basado en una muestra X1, X2, X3.. Xn es un estadístico que toma valores en el espacio estimador Θ idénticamente distribuido. Esta definido por:

$$\hat{\theta} = \hat{\theta}(\underline{X}) = \hat{\theta}(X_1, X_2, \ldots, X_n)$$

- *El valor observado de* $\widehat{\theta}$ *basado en valores observados* $x_1, x_2, \ldots x_n$

$$\hat{\theta}(\underline{x}) = \hat{\theta}(x_1, x_2, \ldots, x_n)$$

- Le llamaremos estimador de θ
- La principal propiedad de un estimador es que “en promedio” proporciona el valor correcto de ϑ, esta propiedad se llama: ”imparcialidad”, “ausencia de sesgo” o en inglés “unbiasedness”

MSC RENZO CLAURE ARACENA

88

Ausencia de Sesgo

$$E\left[\hat{\theta}\right] = \theta \; para \; toto \; \theta \; \in \; \Theta$$

- Si la muestra es representativa de la población, se espera q los estadísticos calculados tengan valores semejantes a los parámetros de la población. La estimación consiste en asignar los valores de los estadísticos muestrales a los valores poblacionales
- Los estadísticos que obtenemos entonces se llaman estimadores
- Ejemplos:
  - Muestras de Notas
  - Encuetas de votación

Error cuadrático medio ECM-MSE

- Mide la variación entre un estimador y el parámetro al que intentar estimar, en otras palabras mide la dispersión del estimador alrededor el valor real del parámetro.
- A partir de ahora definamos al estimador con t mayúscula: $\hat{\theta} = T$
- Se utiliza para comprara estimadores y modelos estadísticos
  - $ECM_{T1} < ECM_{T2}$ entonces el estimador $T1$ es mejor estimador del parámetro θ

Error cuadrático medio ECM-MSE

- Un estimador con $ECM = 0$ es un estimador insesgado o estimador perfecto

$$ECM(T) = E[(T-\theta)]^2 = V(T) + [E(T)-\theta]^2 = V(T) + sesgo^2(T)$$

- El tamaño del ECM se ve influenciado por:
  - La precisión o varianza del estimador
  - La diferencia entre el valor promedio del estimador y el parámetro desconocido

Comparación de estimadores

- Hemos considerado como función de pérdida, o también llamada: función de riesgo cuadrática:

$$l(\theta, \alpha) = (\theta - \alpha)^2.$$

- Pero también existen otras funciones que a veces pueden ser más adecuadas:
  - Error absoluto: $l(\theta, \alpha) = \theta - \alpha.$
  - Función truncada de riesgo cuadrático: $l(\theta, \alpha) = \min((\theta-\alpha)^2, d^2).$
  - Función de riesgo 0-1: $l(\theta, \alpha) = 0$ si $\theta - \alpha \leq \epsilon$ de otro modo: $l(\theta, \alpha) = 1$
- Como ya se comentó previamente, un estimador T1 es mejor que uno T2 si: $R(T, \theta) \leq R(T', \theta), \quad \forall \ \theta \in \Theta$

Criterio de evaluación de estimadores

- Sin sesgo: Cuando el valor de la media muestra coincide con el valor del verdadero parámetro.
  - ¿la media muestral es un estimador /sesgado/insesgado?
  - ¿la varianza muestral es un estimador /sesgado/insesgado?
- Eficiente: Entre varios estimadores sin sesgo de un mismo parámetro el que tiene el error estándar mas pequeño es el más eficiente (mínima varianza)
- Consistente: a medida que aumentar el tamaño de la muestra, el estimador se acerca más al parámetro real que estima: La media y la varianza son estimadores consistentes
- Suficiente: ningún otro estimador puede ofrecer más información sobre el parámetro estimado

La desigualdad de Cramer Rao

- La desigualdad de Cramér-Rao es una herramienta importante en estadística e inferencia, que establece un límite inferior en la varianza de un estimador no sesgado. En otras palabras, establece una cota mínima para la precisión del estimador.
- Ejemplos:
  - En la teoría de la comunicación, la desigualdad de Cramér-Rao se utiliza para determinar la capacidad de un canal de comunicación, es decir, la tasa máxima de información que se puede transmitir a través del canal.
  - En la teoría de la detección y estimación de señales, la desigualdad de Cramér-Rao se utiliza para analizar la calidad y la eficiencia de los estimadores de parámetros, como la frecuencia o la fase de una señal, en presencia de ruido.
  - En la estadística bayesiana, la desigualdad de Cramér-Rao se utiliza para evaluar la calidad de la inferencia bayesiana y para diseñar priors informativos.
- En general, la desigualdad de Cramér-Rao es una herramienta valiosa para evaluar la calidad de los estimadores en presencia de ruido y para diseñar estimadores más precisos y eficientes.

La desigualdad de Cramer Rao

- Ejemplo: Se cuenta con un modelo lineal con la forma: $y = \beta_0 + \beta_1 x + \varepsilon$
- Donde β_0 y β_1 son los parámetros desconocidos del modelo, x es la variable explicativa y y es la variable de respuesta. ε es el error del modelo, que tiene una media cero desviación estándar σ^2.
- El objetivo es estimar los parámetros β_0 y β_1 a partir de una serie de observaciones de la variable de respuesta y y la variable explicativa x.
- La varianza del estimador de un parámetro β_i se puede calcular utilizando la desigualdad de Cramér-Rao de la siguiente manera:

$$\mathrm{Var}(\hat{\beta}_i) \geq 1/I_i$$

- Donde $\hat{\beta}_i$ es el estimador de β_i, I_i es la información de Fisher de β_i que mide la cantidad de información que los datos proporcionan sobre β_i y es igual a:

$$I_i = -\mathbb{E}\left[\frac{\partial^2 \log f(Y;\beta)}{\partial \beta_i^2}\right]$$

- Donde f es la función de densidad conjunta de y y E es el valor esperado

Resumen

- La LGN (Ley de los grandes números) establece que los promedios de muestras iid convergen hacia las medias de la población que están estimando.
- El TLC (Teorema del límite central) establece que los promedios son aproximadamente normales, con distribuciones:
  - Centradas en la media de la población.
  - Con una desviación estándar igual al error estándar de la media.
  - El CLT (Central Limit Theorem) no garantiza que n sea lo suficientemente grande.