Notas sobre el Curso
Estadística I

$$
\begin{aligned}
Profesor &: \text{René Iral Palomino}\\
Oficina &: 43-320\\
Correo &: riral@unal.edu.co
\end{aligned}
$$

Contenido

- Estadística Descriptiva
- Probabilidad y distribuciones de Probabilidad
- Estimación Puntual
- Intervalos de Confianza
- Pruebas de Hipótesis

Evaluación

- 5 parciales

Parcial 1	Parcial 2	Parcial 3	Parcial 4	Parcial 5
10 %	15 %	15 %	15 %	20 %

- 2 trabajos usando R

Trabajo I	Trabajo II
10 %	15 %

- Documentos de interés, programa del curso, horarios de asesorías de profesores y de talleres: **http://unvirtual.medellin.unal.edu.co**, en el curso Estadística I, código 3006914.
- Inscripción al curso: Ingreso a Moodle, usando el login del correo institucional y el respectivo pasword.

Introducción

En muchas áreas del quehacer científico es común la realización de experimentos con la finalidad de verificar afirmaciones acerca de comportamientos de ciertos procesos de la naturaleza, la realización de encuestas para determinar la aceptación hacia un nuevo producto o persona en particular, la realización de estudios de seguimiento a través del tiempo o la recolección de datos de diversas fuentes.

- Un ingeniero Agrónomo puede estar interesado en demostrar que un cierto tipo nuevo de plaguicida es más efectivo que un plaguicida convencional; para esto compara la efectividad de ambos por medio de la realización de un experimento controlado de laboratorio.
- Un investigador está interesado en determinar el porcentaje de personas que se benefician con un nuevo sistema de seguridad social. Para ello toma una muestra aleatoria de potenciales beneficiarios en la población y determina cuantos de la muestra se benefician de dicho programa.
- Un ingeniero Geológico puede recolectar datos cuando estudia las características de un cierto tipo de roca
- Un ingeniero administrativo puede generar observaciones en un estudio de mercado o en un censo sobre una población
- Un médico puede registrar los tiempos de supervivencia de algunos pacientes que padecen cáncer
- Un especialista en vías puede estar interesado en el número de vehículos que llegan a un carril de giro en una hora pico.

Un conjunto de datos puede estar compuesto de solamente algunas pocas observaciones o de cientos y tal vez miles de observaciones, pero este por si solo no proporciona información alguna. Las preguntas más frecuentes que surgen son:

- ¿Qué hacer con los datos producto de una investigación?
- ¿Cómo obtener información de ellos?
- ¿Cual es la distribución de las observaciones registradas?
- ¿Qué tipo de herramientas metodológicas son aplicables a los datos, de manera que permitan responder a las inquietudes iniciales del investigador?

La Estadística se relaciona estrechamente con experimentos, encuestas y estudios de seguimiento, ya que estos generan datos.

Uno de los principales objetivos de la estadística es obtener información confiable de un conjunto de datos producto de un proceso de experimentación o de recolección. Por esta razón la Estadística se considera una disciplina que se relaciona con diversas áreas, tales como economía, sociología, ingeniería, demografía, salud pública, mercadeo, política, deportes, entre otras.

La Estadística se ocupa, entre otras cosas, del manejo de información que pueda ser cuantificada. Implica entonces la descripción de conjuntos de datos y la inferencia a partir de la información recolectada de un fenómeno de interés.

La mejor manera de recopilar la información es a través de la realización de un **Experimento**. Un experimento se entiende como cualquier procedimiento que genera datos ó información.

Un experimento, así se realice bajo condiciones similares, difícilmente generará resultados iguales. Por ejemplo, suponga que se someten a prueba dos componentes electrónicos para determinar su duración en un equipo. Aunque ambos componentes se prueban en equipos con características similares, sería muy casual que duraran exactamente el mismo tiempo.

Esto quiere decir que el resultado de cualquier experimento se ve afectado por un factor de **incertidumbre**, es decir, al momento de realizar la prueba no es posible saber con exactitud cual será el resultado a obtener, y el hecho de que den resultados diferentes quiere decir que hay presente un factor de **variabilidad**. El tipo de experimento en el cual no se sabe de antemano su resultado pero si se conocen los posibles resultados se denomina **Experimento aleatorio**.

Probabilidad y Distribuciones de Probabilidad

Uno de los objetivos de la estadística es el estudio de la variabilidad. Debido a éste aspecto, se necesita de una medida ó escala que nos permita cuantificar el grado de seguridad ó de incertidumbre respecto a un resultado ó conjunto de resultados, en la realización de un experimento aleatorio. Dicho de otra forma, una medida de la posibilidad de ocurrencia de un resultado.

El término PROBABILIDAD está asociado con el estudio de la aleatoriedad y la incertidumbre.

Definiciones Básicas

- **Experimento Aleatorio**: Es aquel que proporciona diferentes resultados, aún cuando este es repetido bajo las mismas condiciones experimentales.

- **Espacio Muestral**: Es el conjunto de toda la posible información que se obtiene en la realización de un experimento aleatorio. Usualmente denotado con la letra S.

- **Evento**. Debido a que un espacio muestral es un conjunto de información, un Evento puede ser considerado con un subconjunto de información del espacio muestral (esto puede ser visto también, como la información asociada a una pregunta de investigación o una condición particular en el experimento de interés). Estos pueden ser catalogados como *Simples* o *Compuestos*. Son usualmente denotados con letras mayúsculas: A, B, C, D, A_i, etc. Es importante acalra que los eventos

- **Operaciones entre eventos**. Sean E_1, E_2 eventos de un espacio muestral S. Los subconjuntos: $E_1 \cup E_2$, $E_1 \cap E_2$, $E_1 - E_2 = E_1 \cap E_2'$ y E_1', son también eventos de S, donde E_1' representa el complemento del evento E_1.

Nota: El conjunto vacio, denotado ϕ, es un evento de S, llamado el evento *Imposible* y S será tambien un evento de S, llamado el evento *Seguro*.

Ejemplos

- Se lanza una moneda no cargada, $S = \{C, S\}$ donde 'C' es la abreviación para 'Cara' y 'S' para 'Sello'.

- Se lanza un dado cúbico, $S = \{1, 2, 3, 4, 5, 6\}$.

- De la producción diaria de una empresa se seleccionan aleatoriamente tres artículos, los cuales se clasifican como defectuoso (**D**), o no defectuoso (**N**). Por notación el resultado NNN se asocia a que los tres artículos resultaron no defectuosos, mientras que el resultado NDN representa que el primer y tercer artículo evaluados son defectuosos y el segundo no lo es. Así, el espacio muestra será:
  $S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$.

- De una población muy grande se encuestan aleatoriamente personas hasta encontrar la primera con cierta característica de interés.
  $S = \{T, NT, NNT, NNNT, NNNNT, \cdots\}$, donde 'T' representa el caso en el cual la persona tiene la característica de interés y 'N' cuando no la tiene.

- Se elige aleatoriamente una bombilla de la producción diaria y se establece su duración en horas. $S = [0, +\infty\,)$.

- Se lanzan dos dados cúbicos. El espacio muestral en este caso está constituído por todas las posibles parejas de resultados. Por simplicidad, el espacio muestra estará dado por:
  $S = \{(1, 1), (1, 2), \cdots, (6, 5), (6, 6)\}$, donde $(1, 3)$ indica que el primer dado resultó en un 1 y el segundo en un 3.

Definición. Sean A, B eventos de un espacio muestral S. Se dice que A y B son *Excluyentes o Disjuntos* si $A \cap B = \phi$. En general si $E_1, \cdots, E_n$ son eventos de S, se dice que son *Mutuamente excluyentes* si $E_i \cap E_j = \phi$, para $i \neq j$, $i, j = 1, \cdots, n$.

Ejemplo. Se lanzan dos dados cúbicos. Considere los eventos: A: La suma de los resultados es inferior o igual a 3, B: La suma de los resultados es par, C: La suma de los resultados es múltiplo de 5. Escriba por extensión los eventos A, B y C.

$A = \{(1,1), (1,2), (2,1)\}$.
$B = \{(1,1), (1,3), (1,5), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5), (4,2)$
$(4,4), (4,6), (5,1), (5,3), (5,5), (6,2), (6,4), (6,6)\}$.
$C = \{(1,4), (2,3), (3,2), (4,1), (4,6), (5,5), (6,4)\}$.

Observe que $A \cap B = \{(1,1)\}$, $A \cap C = \phi$ y $B \cap C = \{(4,6), (5,5), (6,4)\}$. Esto indica que A y C son exluyentes.

Axiomas de probabilidad

El enfóque básico sobre el cuál podemos iniciar el cálculo de probabilidades está vinculado a la frecuencia relativa. Así, la probabilidad de un resultado de un experimento aleatorio puede asociarse a la frecuencia de ocurrencia de dicho resultado en repeticiones sucesivas del experimento.
Si un experimento tiene N posibles resultados, todos con la misma probabilidad de ocurrencia, la probabilidad asociada a cada resultado es $\frac{1}{N}$.
Si en vez de un resultado, se tiene un conjunto de resultados, digamos un evento E, la probabilidad asociada al evento E, después de n repeticiones del experimento aleatorio, es aproximadamente $\frac{n_E}{n}$, donde n_E es el número de resultados contenidos en E de las n repeticiones.

Ejemplos.

- Se lanza un dado cúbico no cargado. El espacio muestral es:
  $S = \{1, 2, 3, 4, 5, 6\}$. La Probabilidad asociada a cada resultado será $\frac{1}{6}$.

- Se lanzan dos dados cúbicos no cargados.
  $S = \{(1,1), (1,2), \cdots, (6,5), (6,6)\}$. Probabilidad asociada a cada resultado (par) será $\frac{1}{36}$. Considere el evento A: La suma de los resultados es 5. El evento A por extensión está dado por:
  $A = \{(1,4), (2,3), (3,2), (4,1)\}$.
  ¿Cómo calcular la probabilidad del evento A?.
  En este caso el experimento no se ha repetido n veces, solo se tiene disponible el conjunto de posibles resultados asociados a S y al evento A. Teniendo en cuenta que los 4 resultados contenidos en A son excluyentes, es razonable pensar que la probabilidad asociada a A se pueda obtener como la suma de las probabilidades de cada uno de los resultados en A. Bajo este supuesto, se tiene que:
  $$P(A) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{4}{36}.$$
- Se lanzan tres monedas no cargadas. Si 'H' denota 'Cara' y 'T' denota 'Sello', entonces el espacio muestral estará dado por:

  $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.

  La probabilidad asociada a cada resultado será $\frac{1}{8}$.
- Se escoge una bombilla aleatoriamente y se registra su duración. La probabilidad de que el tiempo de duración esté entre a y b está dada por: $\int_a^b f_X(t)\, dt$, para cierta función $f_X(x)$.

En espacios muestrales donde el conjunto de resultados es Contable (finito o numerable), dado un evento E de dicho espacio, la probabilidad asociada al evento E, se calcula como la suma de las probabilidades de los resultados contenidos en el evento E.

Ejemplo. Se lanzan tres monedas no cargadas. Defina los eventos A: Al menos una cara es obtenida, B: Se obtiene el doble de sellos que de caras. Calcule $P(A)$, $P(B)$, $P(A \cap B)$ y $P(A \cap B')$. El espacio muestral para este experimento es $S = \{CCC, CCS, CSC, SCC, CSS, SCS, SSC, SSS\}$. Así,

$$P(A) = \frac{7}{8}, \quad P(B) = \frac{3}{8}, \quad P(A \cap B) = \frac{3}{8}, \quad P(A \cap B') = \frac{4}{8}.$$

Definición. Una función $P : S \;\rightarrow\; \Re$, será llamada una *Función de Probabilidad*, si satisface las siguientes propiedades:

I) Si A es cualquier evento de S, entonces $P\,(A) \geq 0$

II) $P\,(S) \;=\; 1$

III) Si E_1 , E_2, ... , E_n , ...es una colección (finita o infinita) de eventos de S, mutuamente excluyentes entonces:

$$P\,(E_1 \;\cup\; E_2 \cup\; \ldots\; \cup\; E_n\; \cup\; \ldots) = P\left(\bigcup_{i\,=\,1}^{\infty} E_i\right) = \sum_{i\,=\,1}^{\infty} P\,(E_i) \;<\; +\,\infty\;.$$

Si $n \;=\; 2 \;\;\Rightarrow\;\; P\,(E_1 \;\cup\; E_2) \;=\; P\,(E_1) \;+\; P\,(E_2)$.

Teorema: Sean A y B eventos de un espacio muestral S; entonces:

- $0 \leq P\,(A) \leq 1$
- $P\,(\phi) = 0$
- $P\,(A') = 1 \;-\; P\,(A)$, donde $A^{'}$ es el complemento de A.
- Si $A \subseteq B$, entonces $P\,(A) \leq P\,(B)$.
- $P\,(A \cup B) = P\,(A) + P\,(B) - P\,(A \cap B)$.

Ejemplo
La siguiente tabla presenta la historia de 940 productos de un proceso de fabricación de semi - conductores. Se elige al azar un producto.

		Revisión Electrónica		
		SI	NO	Total
Contaminación Alta	SI	246	112	358
	NO	68	514	582
	Total	314	626	940

Sea A: el evento en que el producto tenga altos niveles de contaminación y B: el evento en que el producto pasó por un proceso de revisión electrónica. Calcule: $P(A)$, $P(B)$, $P(A \cap B)$, $P(A \cup B)$, $P(A \cap B')$.

Solución

$$P(A) = \frac{358}{940} = 0.38085\,, \quad P(B) = \frac{314}{940} = 0.33404\,.$$

$$P(A \cap B) = \frac{246}{940} = 0.26170\,.$$

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{358 + 314 - 246}{940} = \frac{426}{940} = 0.4319 \end{aligned}$$

$$P\left(A \cap B'\right) = \frac{112}{940} = 0.11915\,.$$

Ejemplo

Un dado cúbico particular está cargado de tal manera que un número par tiene el doble de probabilidad de presentarse que un impar. Si E es el evento en el que se da un número menor que cuatro en un solo lanzamiento, encuentre $P(E)$.

Solución

El espacio muestral para este experimento está dado por;
$S = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$. Ahora bien, sea $P(Par) = w$. Como la probabilidad de par es el doble de la impar, se tiene que $P(Impar) = \frac{w}{2}$.
Por el axioma 2 de probabilidad, la suma de las probabilidades debe ser igual a uno; entonces,

$$\begin{aligned} \frac{w}{2} + w + \frac{w}{2} + w + \frac{w}{2} + w &= 1 \quad \text{despejando a } w \\ w &= \frac{2}{9} \end{aligned}$$

Por lo anterior

$$P(\text{Par}) = \frac{2}{9} \quad , \quad P(\text{Impar}) = \frac{1}{9}\,.$$

Como $E = \{1,\ 2,\ 3\}$, entonces:

$$P(E) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}\,.$$

Ejemplo
Cuando A y B juegan tenis, A siempre gana en proporción de 3 a 5. Si ambos juegan solo tres partidos, ¿Cuál es la probabilidad de que B gane al menos dos partidos?

Solución
Denotemos por ABA, un resultado típico, el cuál representa el orden y el sujeto que ganó el partido. Por ejemplo ABA indica que A ganó el primer y tercer partido y B ganó el segundo. De ésta manera, el espacio muestral está dado por:

$$S = \{AAA, AAB, ABA, BAA, ABB, BAB, BBA, BBB\} \ .$$

Este es un espacio muestral donde los resultados no son equiprobables (es decir, no podemos asignar la misma probabilidad a cada resultado). Por ejemplo es más probable que ocurra AAA que BBB.

Del enunciado se tiene que $P(A) = \frac{3}{5}$ y $P(B) = \frac{2}{5}$. Para calcular las probabilidades asociadas a cada resultado, asumiremos que ABA, por ejemplo, representa la intersección de tres eventos y que la probabilidad se calcula como el producto de las probailidades de los tres eventos involucrados. Por ejemplo:

$$P(ABA) \ = \ P(A) \ P(B) \ P(A) \ = \ \left(\frac{3}{5}\right) \left(\frac{2}{5}\right) \left(\frac{3}{5}\right) = \frac{18}{125} \ .$$

De esta forma tenemos:

$$\underbrace{AAA}_{\frac{27}{125}} \qquad \underbrace{AAB}_{\frac{18}{125}} \qquad \underbrace{ABA}_{\frac{18}{125}} \qquad \underbrace{BAA}_{\frac{18}{125}} \qquad \underbrace{ABB}_{\frac{12}{125}} \qquad \underbrace{BAB}_{\frac{12}{125}} \qquad \underbrace{BBA}_{\frac{12}{125}} \qquad \underbrace{BBB}_{\frac{8}{125}}$$

Sea C el evento donde B gana al menos dos partidos, entonces $C \ = \ \{ABB, BAB, BBA, BBB\}$. Así:

$$P(C) \ = \ P(ABB) \ + \ P(BAB) \ + \ P(BBA) \ + \ P(BBB) \ = \ \frac{44}{125} \ .$$