



PROBABILIDAD Y ESTADÍSTICA

Introducción

Sandra Cano

Facultad de Ingeniería
spcano@usbcali.edu.co

Por qué hay que adorar la estadística?

[https://www.ted.com/talks/
alan_smith_why_we_re_so_bad_at_statistics?language=es](https://www.ted.com/talks/alan_smith_why_we_re_so_bad_at_statistics?language=es)

Objetivo

- ✓ Conocer
- ✓ Explicar
- ✓ Definir

Vocabulario Probabilidad y Estadística

Estadística

Es una disciplina que proporciona la metodología, fundada en la matemática, para **obtener**, **recopilar**, **procesar**, **resumir** y presentar datos referentes a un estudio de interés, **transformándolos** en estadística con el fin de **interpretarlas** y **obtener** conclusiones.

A tener en cuenta:

- **Describir información**
- **Hacer inferencias**

Definición

¿Qué son los datos científicos?

¿Qué es la estadística?

¿Qué es la estadística inferencial?

¿Qué son métodos estadísticos?

¿Cuál es la importancia de la estadística en el mundo?

Casos

Caso 1:

Un ingeniero está interesado en saber la **eficiencia** de un instrumento que sirva para medir el monóxido de azufre en estudios relacionados con la contaminación atmosférica.

Cómo lo resuelve?

Se debe tomar en cuenta dos fuentes de variación:

1. Variación en los valores del monóxido de azufre que se encuentran en el mismo lugar y el mismo día.
2. Variación entre los valores observados y la cantidad real de monóxido de azufre que haya en el aire en ese momento.

Si cualquiera de las dos fuentes es excesivamente grande (definido el valor por el ingeniero(a)), quizás necesite reemplazar el instrumento.

Eficiencia: Capacidad para realizar o cumplir adecuadamente una función. [Wikipedia]

Eficiencia

En la disciplina de la estadística se usa el término eficiencia para referirse al **margen de error de una estadística**.

Para tomar la decisión se elige el que tienen menor margen de error para un determinado estudio.

Cómo se relaciona?

- **Agricultura** se relaciona con el sistema de riego. Cuando resulta eficaz se utiliza menor cantidad de agua. Esta eficiencia permite que se cultive una mayor cantidad de tierras y su producción sea superior.
- **Proyectos de desarrollo de Software** se relaciona con el tiempo que se invierte en terminar un proyecto (coste, funcionalidad).
- **Eficiencia** (uso de un producto): Tiempo que se necesita para completar una tarea.

Casos

Caso 2:

Un estudio biomédico de un nuevo fármaco para reducir la hipertensión, 85% experimentaron alivio. Solo se conoce que el medicamento actual o “viejo” alivia a 80% de los pacientes que sufren hipertensión crónica. El nuevo fármaco es más caro de elaborar y podría tener algunos efectos colaterales.

Se debería adoptar el nuevo medicamento?

Se debe tomar en cuenta necesidades de variación:

1- el 85% está relacionado con un determinado número de pacientes que se han seleccionado para el estudio.

Tal vez si se repitiera el estudio con NUEVOS PACIENTES, el número observado de “Éxitos sería el 75%”.

(variación natural de un estudio a otro que se debe tomar en cuenta para la toma de decisiones)

Variabilidad

- En la variación que tienen los datos. Si el dispositivo para medir el monóxido de azufre siempre diera el mismo valor y fuera exacto no se necesita de un **análisis estadístico**.
- Si entre un paciente y el otro no hubiera variabilidad de respuesta al medicamento (es decir si siempre aliviara o nunca aliviara).
- Variabilidad de un proceso en un **proceso de producción**.

Análisis Estadístico

- Es un componente de análisis de datos. El objetivo es identificar tendencias. El análisis involucra un conjunto de fases:
 - Describir la naturaleza de los datos a ser analizados.
 - Explorar la relación de los datos con la población subyacente.
 - Crear un modelo para resumir la comprensión de cómo los datos se relacionan con la población subyacente.
 - Probar la validez del modelo.
 - Emplear el análisis predictivo para ejecutar escenarios que ayuden a orientar acciones futuras.

Algunos software, como: SPSS, R (lenguaje de programación), y Excel

Muestras: Recolección de datos

- Las muestras se relacionan con una **población, conjunto de individuos, o elementos.**
- Una población puede representarse por un sistema científico.

Ejemplo:

Fabricante de tarjetas para computadoras desea eliminar defectos.

Población:

tarjetas de computadoras

Restricción:

- (1) Tarjeta de la empresa
- (2) Período determinado



Muestreo : proceso

- Muestreo aleatorio simple

- Tamaño de la muestra
 - Inferencias
- Muestra Sesgada: muestra pequeña que no representa el total.

Ejemplo: Reunir opiniones de una población respecto del referéndum. Por lo que, hay que tener en cuenta : grupos étnicos, estratos, etc.

- Diseño Experimental

- Unidades experimentales
 - Medicamento (género, edad y peso)

Ejemplos

Datos de corrosión

Recubrimiento	Humedad	Promedio de corrosión en miles de ciclos hasta la ruptura
Sin recubrimiento	20%	975
	80%	350
Con recubrimiento químico contra la corrosión	20%	1750
	80%	1550

Efectividad

- Es un equilibrio entre dos variables eficiencia y eficacia (lograr un resultado).
- Se logra el resultado producir tarjetas del fabricante en el período establecido. (Eficacia)
- El coste que se invierte para fabricar estas tarjetas. (eficiencia)

Ejemplo

- Un ingeniero se encuentra con datos de un proceso de producción donde se **muestran** 100 artículos y se obtienen 10 defectuosos.
- La empresa solo puede tolerar 5% de artículos defectuosos en el proceso. (proceso aceptable)

Cuál es la probabilidad de obtener 10 o más artículos defectuosos de una muestra de 100 artículos?

Papel de la probabilidad

- Cuantificar la confianza en nuestras conclusiones.
- Los elementos de la probabilidad permiten expresar la conclusión en el lenguaje que requieren los científicos e ingenieros.

Inferencia Estadística

- Emplea conceptos de probabilidad.
- Muestra + Inferencia estadística= conclusiones acerca de la población

Estadística Descriptiva

- Conjunto de procedimientos que resumen los datos en diferentes tipos de representaciones, como: **texto, gráficas y tablas.**

Instrumentos

- Un elemento que puede tomarse como un fenómeno que se desea medir y se emplea para ello.
- Examen, una encuesta, pruebas psicométricas, entre otros.

Tipos de Variables

- Cualitativas : Nominales, ordinales
- Cuantitativas : discretas, continuas

Tener en cuenta

- **Variable cualitativa**, mide una cualidad.
- **Variable cuantitativa**, mide una cantidad.

Nominal: Ejemplos: Sexo, color de piel, nacionalidad, religión, estado civil.

Ordinal: Grado, infestación (sana, leve, moderada), calificación (excelente, bueno, regular, malo)

Medidas Cuantitativas: Localización

- Son medidas que sirven de apoyo para hacer un análisis de los valores.
 - Media (promedio de los datos)
 - Media recortada (corta el valor + mayor y - menor)
 - Mediana (centroide de los datos)
 1. Ordenar de Menor a Mayor
 2. Aplicar lo siguiente:

$$\bar{x} = \begin{cases} x_{(n+1)/2} & n : \textit{impar} \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}) & n : \textit{par} \end{cases}$$

Medidas Cuantitativas: Variabilidad

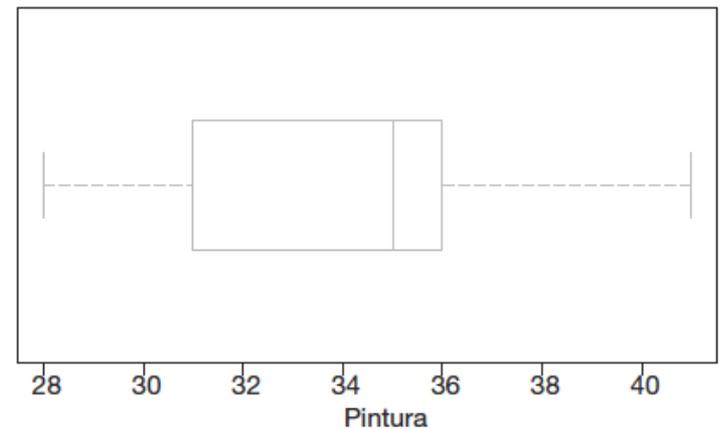
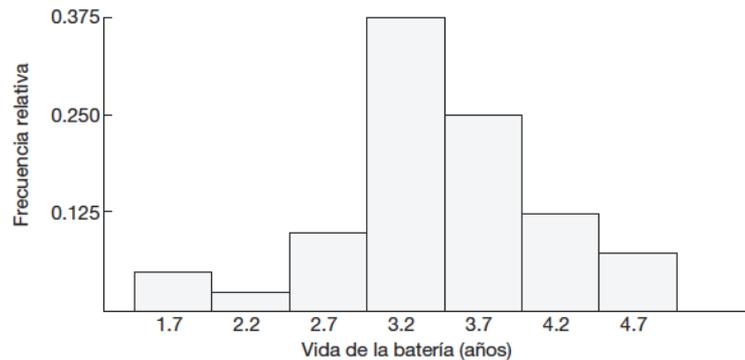
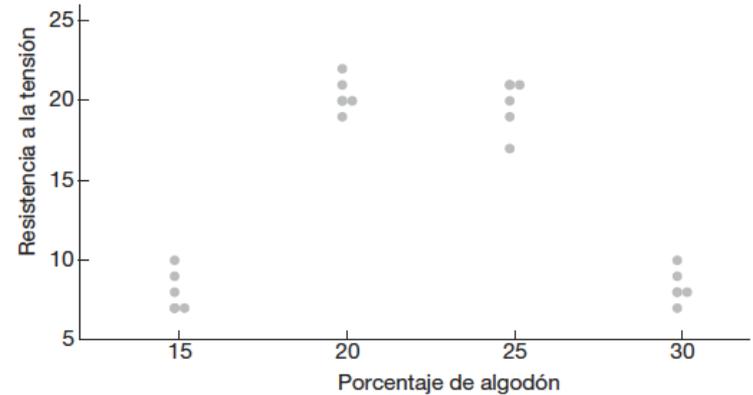
Dispersión de los datos

- Rango ($X_{\max} - X_{\min}$)
- Desviación Estándar (s) (varianza de la muestra)

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}, \text{ entonces, } s = \sqrt{s^2}$$

Gráficos

- Diagrama de Barras (Histograma)
- Diagrama de dispersión
- Grafica de caja y bigote



Ejercicio No 1

Se registran las siguientes mediciones para el tiempo de secado (en horas) de cierta marca de pintura esmaltada.

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

- Cuál es el tamaño de la muestra?
- Cuál es la media de la muestra?
- Grafique los datos usando una gráfica de puntos
- Calcule la media recortada al 20% para el conjunto de datos. (elimina el mínimo y máximo valor del conjunt de datos)

Ejercicio No 2

Se utiliza cierto polímero para los sistemas de evaluación de los aviones. Es importante que el polímero sea resistente al proceso de envejecimiento. Se utilizaron veinte especímenes del polímero en un experimento. Diez se asignaron aleatoriamente para exponerse a un proceso de envejecimiento acelerado del lote, el cual implica la exposición de altas temperaturas durante 10 días. Se hicieron las mediciones de resistencia

Sin envejecimiento	227	222	218	217	225
	218	216	229	228	221
Con envejecimiento	219	214	215	211	209
	218	203	204	201	205

- **Cuál es el tamaño de la muestra?**
- **Cuál es la media de la muestra?**
- **Grafique los datos usando una gráfica de puntos**
- **Calcule la media recortada al 20% para el conjunto de datos. (elimina el mínimo y máximo valor del conjunt de datos)**

Ejercicio No 3

La resistencia a la tensión del caucho de silicio se considera una función de la temperatura de vulcanizado. Se llevó a cabo un estudio en el que se prepararon muestras de 12 especímenes del caucho utilizando temperaturas de vulcanización de 20°C y 45°C. Los siguientes datos se presentan los valores de resistencia a la tensión en megapascales.

20°C	2.07	2.14	2.22	2.03	2.21	2.03
	2.05	2.18	2.09	2.14	2.11	2.02
45°C	2.52	2.15	2.49	2.03	2.37	2.05
	1.99	2.42	2.08	2.42	2.29	2.01

- Elabore una gráfica de puntos con los datos, tanto de los valores de resistencia a la tensión a temperatura alta como los de temperatura baja.

Mal uso de la estadística

- El 33% de las alumnas de un curso de ingeniería se caso con profesores de la universidad. Sin embargo, no se dijo que el total de alumnas en el curso eran **3 estudiantes**.
- Un estudio reveló una alta correlación entre el peso de un niño de básica y su rapidez de lectura, deduciéndose que los niños gordos tienen mayor rapidez de lectura que los flacos.
 - Deducción es mala por qué se asocian el peso con “gordura” y a la vez el peso está altamente relacionado con la edad y la estatura. Lo otro alumnos de mayor peso están asociados a mayor edad y por lo tanto son los últimos cursos de básica.

Probabilidad

- La teoría de probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o **experimentos aleatorios**.

- **Experimento aleatorio**, aquel que cuando se repite bajo las mismas condiciones, no es el mismo siempre.

Probabilidad

- ¿Qué necesitamos conocer?
 - **Espacio muestral:** todos los posibles resultados, y se denota con la letra griega (omega) Ω o también S (sampling space).
 - **Evento:** cualquier subconjunto del espacio muestral. Se denota por letras mayúsculas A,B,C.

Probabilidad: Ejemplo

Un experimento consiste en lanzar un dado. Observar el número que aparece en la cara superior.

¿Espacio muestral?

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

¿Eventos?

$A = \{2, 4, 6\}$ resultado par

$B = \{1, 3, 5\}$ resultado impar

Si lanzamos el dado y obtenemos 4 y luego 1.



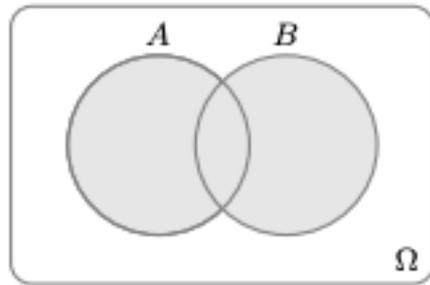
Probabilidad

Sea A un subconjunto de un espacio muestral S o Ω de cardinalidad finita. Se define **probabilidad clásica** del evento A como conciente:

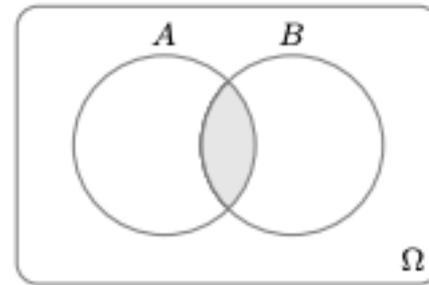
$$P(A) = \frac{\# A}{\# \Omega}$$

Probabilidad

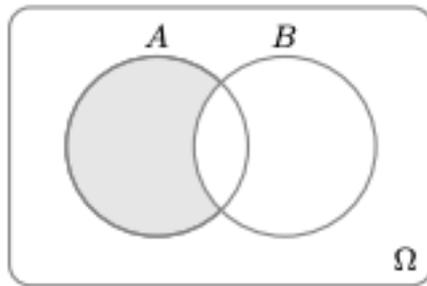
Teoría de conjuntos



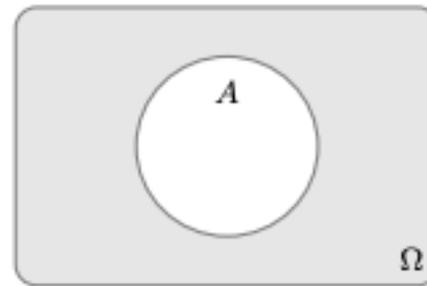
$$A \cup B$$



$$A \cap B$$



$$A - B$$



$$A^c$$

Bibliografía

- Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keyend Ye. Probabilidad y estadística para ingeniería y ciencias. Novena Edición, 2012.
- Antonio Rustom J. Estadística Descriptiva, probabilidad e inferencia .Una visión conceptual y aplicada, 2012.

Recursos en línea:

- http://www.jorgegalbiati.cl/ejercicios_4/ConceptosBasicos.pdf

Tarea

Película:

- “Una mente brillante”. **A Beautiful Mind**