Estadística II

Universidad de Guayaquil

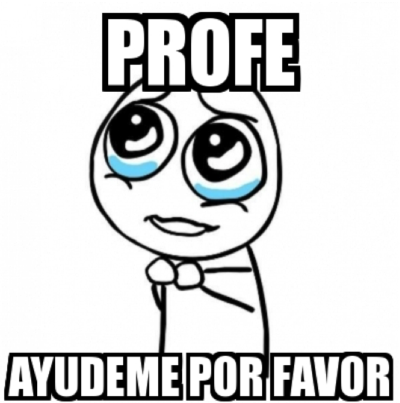
October 23, 2018

Calificaciones

<table>
<tr><th colspan="2">GESTIÓN FORMATIVA , GESTIÓN PRÁCTICA
50%
(NOTA SOBRE 10)</th><th rowspan="5">EXAMEN
50%
(NOTA SOBRE 10)</th></tr>
<tr><td>TAREAS</td><td rowspan="3">8</td></tr>
<tr><td>TALLERES (PRESENCIAL O EDMODO)</td></tr>
<tr><td>LECCIONES</td></tr>
<tr><td>PROYECTO</td><td>2</td></tr>
</table>

Código de Honor

- Ser honesto: no copiar, plagiar ni mentir en ninguna Forma
- Está prohibido llorar



Edmodo

- www.edmodo.com
- Tutorial: https://www.youtube.com/watch?v=oMJcBjzJwiI

POR FAVOR : PONER NOMBRES NO SEUDÓNIMOS

- GRUPO #5 (Martes) : c9j9wj
- GRUPO #4 (Miércoles) : rnz279

<table>
<tr><th colspan="7">D) UNIDADES TEMÁTICAS O DE ANÁLISIS:</th></tr>
<tr><td colspan="7">UNIDAD # 1: Métodos de muestreo, teorema central del límite y estimación e intervalos de confianza</td></tr>
<tr><td colspan="7">OBJETIVO: Definir y contruir una distribución muestral, comprender el teorema central del límite y contruir el intervalo de confianza.</td></tr>
<tr><th rowspan="2">CONTENIDOS: CONOCIMIENTOS A DESARROLLAR</th><th colspan="3">ACTIVIDADES DE ORGANIZACIÓN DEL APRENDIZAJE</th><th rowspan="2">TIEMPO DE APRENDIZAJE</th><th colspan="2">AMBIENTES DE APRENDIZAJE</th></tr>
<tr><th>INTERACCIÓN DIRECTA CON EL PROFESOR</th><th>APLICACIÓN Y EXPERIMENTACIÓN</th><th>TRABAJO AUTÓNOMO</th><th>TIPO</th><th>MÉTODOS, TECNICAS E INSTRUMENTOS</th></tr>
<tr><td>1. Métodos de muestreo:
1.1 Muestreo aleatorio simple.
1.2 Muestreo aleatorio sistemático.
1.3 Muestreo aleatorio estratificado.
1.4 Muestreo por conglomerados.
2. Teorema central del límite.</td><td rowspan="2">Clase magistral, en forma presencial o virtual, con participación activa de estudiantes mediante presentaciones orales</td><td rowspan="2">Presentación de objetivos e introducción

Tormenta de ideas y debate

Talleres, Interpretación de datos</td><td>Búsqueda de información</td><td rowspan="2">9HORAS</td><td rowspan="2">De generación del saber</td><td rowspan="2">• Proyecto de investigación.
• Presentaciones orales.</td></tr>
<tr><td>3. Estimación e intervalos de confianza.</td><td>Búsqueda de información</td></tr>
</table>

Syllabus 2

UNIDAD # 2: Pruebas de hipótesis de una y dos muestras, y análisis de la varianza						
OBJETIVO:						
CONTENIDOS: CONOCIMIENTOS A DESARROLLAR	ACTIVIDADES DE ORGANIZACIÓN DEL APRENDIZAJE			TIEMPO DE APRENDIZAJE	AMBIENTES DE APRENDIZAJE	
	INTERACCIÓN DIRECTA CON EL PROFESOR	APLICACIÓN Y EXPERIMENTACIÓN	TRABAJO AUTÓNOMO		TIPO	MÉTODOS, TECNICAS E INSTRUMENTOS
1. Qué es una hipótesis y qué es la prueba de hipótesis. 2 Pruebas de hipótesis de una muestra. 3 Pruebas de hipótesis de dos muestras.	Clase magistral, en forma presencial o virtual, con participación activa de estudiantes	Presentación de objetivos e introducción Tormenta de ideas y debate Talleres, Interpretación de datos	Búsqueda de información, Trabajo de investigación	9HORAS	De generación del saber	• Resolución de problemas. • Proyecto de investigación. • Talleres.
4. Análisis de la varianza. 4.1 Distribución F. 4.2 La prueba de Anova.			Cuestionarios, resolución de problemas			

Syllabus 3

UNIDAD # 3: Regresión lineal, múltiple y correlación						
• **OBJETIVO:** El estudiante podrá emplear la regresión lineal como método predictivo						
CONTENIDOS: CONOCIMIENTOS A DESARROLLAR	ACTIVIDADES DE ORGANIZACIÓN DEL APRENDIZAJE			TIEMPO DE APRENDIZAJE	AMBIENTES DE APRENDIZAJE	
	INTERACCIÓN DIRECTA CON EL PROFESOR	APLICACIÓN Y EXPERIMENTACIÓN	TRABAJO AUTÓNOMO		TIPO	MÉTODOS, TECNICAS E INSTRUMENTOS
1. Regresión lineal. 2. Intervalos de confianza de regresión. 3. Análisis de correlación. 4. Análisis de regresión múltiple. 5. Inferencias en la regresión lineal múltiple. 6. Variables independientes cualitativas.	Clase magistral, en forma presencial o virtual, con participación activa de estudiantes	Presentación de objetivos e introducción Tormenta de ideas y debate Talleres, resolución de problemas	Búsqueda de información, Trabajo de investigación Cuestionarios, resolución de problemas	9HORAS	De generación del saber	• Resolución de problemas. • Proyecto de investigación. • Talleres.

Syllabus 4

UNIDAD # 4: Prueba de Bondad de Ajuste						
OBJETIVO: Determinar si los datos disponibles se ajustan a una determinada distribución						
CONTENIDOS: CONOCIMIENTOS A DESARROLLAR	**ACTIVIDADES DE ORGANIZACIÓN DEL APRENDIZAJE**			**TIEMPO DE APRENDIZAJE**	**AMBIENTES DE APRENDIZAJE**	
	INTERACCIÓN DIRECTA CON EL PROFESOR	**APLICACIÓN Y EXPERIMENTACIÓN**	**TRABAJO AUTÓNOMO**		**TIPO**	**MÉTODOS, TECNICAS E INSTRUMENTOS**
1. Bondad de Ajuste: Frecuencias Iguales. 2. Bondad de Ajuste: Frecuencias desiguales. 3. Análisis de tablas de contingencia	Clase magistral, en forma presencial o virtual, con participación activa de estudiantes	Presentación de objetivos e introducción. Tormenta de ideas y debate: ejemplos prácticos y experiencias Talleres y resolución de casos.	Búsqueda de información, Trabajo de investigación Cuestionarios, resolución de problemas	15HORAS	De generación del saber	• Resolución de problemas. • Proyecto de investigación. • Talleres.

Bibliografía

Estadística. Décima edición

TRIOLA, MARIO F. PEARSON EDUCACIÓN, México, 2009 ISBN: 978-970-26-1287-2 Área: Matemáticas

ESTADÍSTICA PARA INGENIEROS

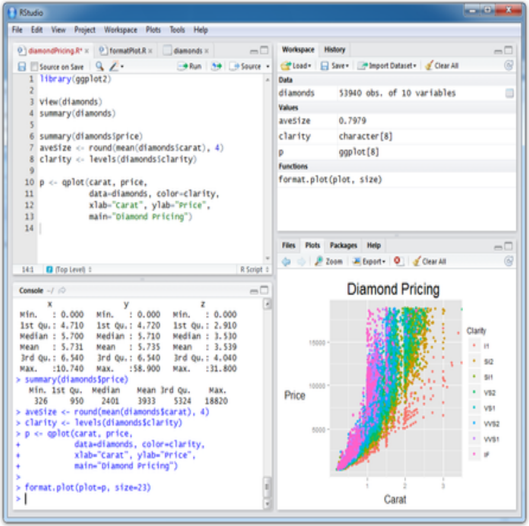
William Navidi McGRAW-HILL/INTERAMERICANA EDITORES, S.A. DE C.V. ISBN 970-10-5629-9

R para Principiantes

Emmanuel Paradis Institut des Sciences de l'E´volution Universit Montpellier II F-34095 Montpellier cdex 05 France

Probabilidad y Estadística, Fundamentos y Aplicaciones

Segunda Edición Gaudencio Zurita



- https://www.youtube.com/watch?v=1WXgaa2Spp0

Objetivos

- La inferencia estadística consiste en extraer una muestra de una población y analizar sus datos con el propósito de aprender acerca de ello.
- Estadística Inferencial: enfatizado en construcción de intervalos de confianza, contrastes de hipótesis estadísticas aplicados a técnicas como bondad ajustes, tabla de contingencia, anova y modelos de regresión lineal.
- Comprobar enunciados respecto de grandes conjuntos de datos, en base a un subconjunto pequeño de los mismos.

Ejemplo Ilustrativo

Considere una máquina que hace varillas de acero para su uso en dispositivos ópticos de almacenamiento. La especificación del diámetro de las varillas es $0{,}45 \pm 0{,}02$ cm. En la última hora, la máquina ha hecho mil varillas. El ingeniero que supervisa la calidad quiere saber cuántas de estas varillas satisfacen la especificación. No tiene tiempo para medir todas. Toma una muestra aleatoria de 50 varillas, las mide y encuentra que 46 de éstas (92 %) satisfacen la especificación del diámetro

- no es probable que la muestra de 50 varillas represente perfectamente a la población de mil??
- La proporción de buenas varillas en la población probablemente es un poco diferente que la proporción de la muestra de 92 %.?
- ¿es admisible que los porcentajes de población sean superiores a 95 %? ¿y de 98 %? ¿O menores de 90 %?, ¿o de 85 %?

Introducción - Definiciones

Datos

Son las observaciones recolectadas (como mediciones, géneros, respuestas de encuestas).

Población N

Es el conjunto completo de todos los elementos (puntuaciones, personas, medidas, etcétera) que se va estudiar. El conjunto es completo porque incluye a todos los sujetos que se estudiarán.

Muestra n

Es un subconjunto de miembros seleccionados de una población. $N > n$

Introducción - Definiciones

Censo

Es el conjunto de datos de cada uno de los miembros de la población.

Parámetro

Es una medición numérica que describe algunas características de una población.

Estadístico

Es una medición numérica que describe algunas características de una muestra.

Población y muestra

Muestreo

"Existen tres clases de mentiras: mentiras, viles mentiras y estadísticas". Se ha dicho que "las cifras no mienten, pero los mentirosos también usan cifras"

- Los datos muestrales deben reunirse de una forma adecuada, como a través de un proceso de selección aleatoria.
- Si los datos muestrales no se reúnen de forma apropiada, resultarán tan inútiles que ninguna cantidad de tortura estadística podrá salvarlos.
- Una muestra aleatoria simple de tamaño n es una muestra elegida por un método en el que cada colección de n elementos de la población tiene la misma probabilidad de formar la muestra, de la misma manera que en una lotería.

Muestras aleatorias simples (I)

Un nuevo proceso químico se realiza diez veces cada mañana durante cinco días consecutivos. Una gráfica de los resultados en el orden en que aparecieron no presenta ningún patrón o tendencia obvia. Si el nuevo proceso se pone en producción, haciéndolo funcionar diez horas todos los días, desde las 7 a.m. hasta las 5 p.m. ¿Es razonable considerar que los 50 resultados sean una muestra aleatoria simple? ¿Qué ocurre si el proceso está siempre funcionando por la mañana?

Muestras aleatorias simples (II)

Un ingeniero civil acaba de recibir una remesa de mil bloques de hormigón, que pesan aproximadamente 50 libras cada uno. Los bloques se han entregado en una gran pila. El ingeniero quiere investigar la fuerza de compresión de los bloques midiendo las fuerzas en una muestra de diez bloques. Para tomar una muestra aleatoria simple se requeriría sacar bloques del centro y de la parte inferior de la pila, lo que puede ser muy difícil. Por esta razón, el ingeniero puede tomar una muestra simplemente tomando diez bloques de la parte superior de la pila.

Muestras aleatorias simples (III)

Un inspector de calidad prueba 40 pernos de una gran remesa y mide la longitud de cada uno. Descubre que 34 de ellos (85 %) cubre la especificación de longitud. Llega entonces a la conclusión de que exactamente 85 % de los pernos de la remesa satisfacen la especificación. Por otra parte, el supervisor del inspector concluye que la proporción de pernos buenos está cerca de 85 % con cierta probabilidad, pero que no es exactamente igual. ¿Cuál es la conclusión correcta?

Errores de muestreo

No importa lo bien que usted planee y ejecute el proceso de recolección de muestras, es probable que ocurra algún error en los resultados. Por ejemplo, seleccione 1000 adultos al azar, pregúnteles si se graduaron de bachillerato y registre el porcentaje de respuestas afirmativas en la muestra. Si usted elige otra muestra de 1000 adultos al azar, es probable que obtenga un porcentaje diferente en esa muestra. Esto se conoce como **variación del muestreo**

- Un **error de muestreo** es la diferencia entre el resultado de una muestra y el verdadero resultado de la población; este error es consecuencia de las fluctuaciones por el azar.
- Un **error que no es de muestreo** sucede cuando los datos muestrales se obtienen, registran o analizan de forma incorrecta (como cuando se selecciona una muestra sesgada, cuando se usa un instrumento de medición defectuoso o cuando se copian los datos de forma incorrecta).

Conceptos básicos de Probabilidades

Experimento

Proceso de prueba y observación

Experimento aleatorio

No se conoce el resultado del experimento

Espacio muestral Ω

Espacio muestral del experimento que se denota con Ω

Puntos de la muestra

Son los valores de los experimentos y pertenecen al espacio muestral
$\Omega = \{w_1, w_2, w_3, ...w_n\}$

Evento

Es un subconjunto del espacio muestral

Experimento - Ejemplo (I)

Lanzar un dado

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

- Eventos???
- $E_1 = \{2, 4, 6\}$
- $E_2 = \{1, 3\}$

- Lanzar la moneda 3 veces

$$\Omega = \{HHH, HHT, HTH, HTT, \\ THH, THT, TTH, TT\}$$

- Evento

$$A = \{HTT, THT, TTH\}$$

$$B = \{HHT, HTH, THH\}$$

- A y B son subconjuntos de Ω

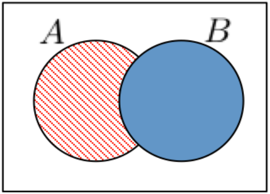
Experimento - Ejemplo (II)

Podemos definir nuevos eventos a partir de los eventos A y B

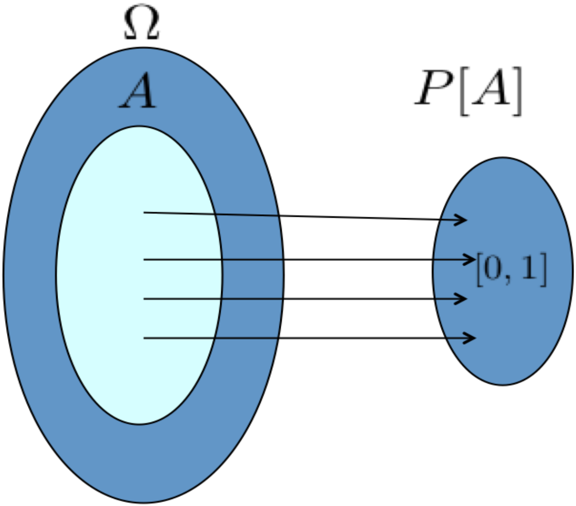
$$A \cup B := \{x : x \in A \vee x \in B\}$$

$$A \cap B := \{x : x \in A \wedge x \in B\}$$

$$A \setminus B := \{x \in A : x \notin B\}$$



Probabilidad



Axioma de Probabilidades

Los axiomas de probabilidad son las condiciones mínimas que deben verificarse para que una función definida sobre un conjunto de sucesos determine consistentemente sus probabilidades. Fueron formulados por Kolmogórov en 1933.

1) $0 \leq P[A] \leq 1$

2) $P[\Omega] = 1$

3) $P[A_1 \cup A_2 \cup ... \cup A_n] = P[A_1] + P[A_2] + ... + P[A_n]$

Para eventos mutuamente excluyentes

Eventos mutuamente excluyentes

$$P[A_1 \cap A_2] = 0$$

A_1: Sacar una as

A_2: Sacar un Rey

A_1: Sacar una carta roja

A_2: Sacar un Rey

$$P[A_1 \cap A_2] \neq 0$$

Espacio de Probabilidades

1) $P[\overline{A}] = 1 - P[A]$

2) $P[\emptyset] = 0$

3) $A \subset B \implies P[A] \leq P[B]$

4) $A = A_1 \cup A_2 \cup ... \cup A_n \quad where \quad A_1, A_2 ..., A_n$

$$P[A] = P[A_1] + P[A_2] + ... + P[A_n]$$

5) $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

Probabilidad condicional

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

A : suma del lanzamiento de dos dados es 7

B: el primer dado es 4

Eventos Independientes (I)

Se dice que dos eventos, A y B, son independentiens si y sólo si la probabilidad del evento B no esta influenciada o cambiada por el suceso A, o viceversa.

Tirar dados Por el contrario, considere tirar un solo dado dos veces y defina dos eventos:

A: observar un 2 en el primer tiro

B: observar un 2 en el segundo tiro

Si el dado es imparcial, la probabilidad del evento A es $P(A) = 1/6$. Considere la probabilidad del evento B. Ya sea que el evento A haya ocurrido o no haya ocurrido, la probabilidad de observar un 2 en el segundo tiro todavía es 1/6. Podríamos escribir:

$P(B$ dado que A ocurrió$) = 1/6$

$P(B$ dado que A no ocurrió$) = 1/6$

Eventos Independientes (II)

Si dos eventos A y B son independientes, la probabilidad de que ocurran A y B es

$$P(A \cap B) = P(A)P(B)$$

Del mismo modo, si A, B y C son eventos mutuamente independientes (todos los pares de eventos son independientes), entonces la probabilidad de que A, B y C ocurran es

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

Tiros de monedas en juegos de fútbol Un equipo de fútbol interviene en dos periodos de tiempo extra durante un juego determinado, de modo que hay tres tiros de monedas al aire. Si la moneda es imparcial, ¿cuál es la probabilidad de que pierdan los tres tiros?

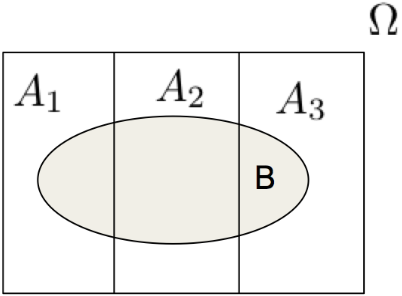
A: perder el primer tiro

B: perder el segundo tiro

C: perder el tercer tiro

$$P(\text{gana}) = P(\text{pierde}) = .5$$

Probabilidad total y Teorema de Bayes



$$P[B] = P[B|A_1]P[A_1] + P[B|A_2]P[A_2] + ... + P[B|A_n]P[A_n]$$

$$P[A_k|B] = \frac{P[A_k \cap B]}{P[B]} = \frac{P[B|A_k]P[A_k]}{\sum_{i=1}^{n} P[B|A_i]P[A_i]}$$

Probabilidad - Ejemplos

Calcule lo siguiente a partir de la tabla:

Tabla 4-1 Resultados de exámenes sobre el consumo de marihuana

	¿Los sujetos realmente consumen marihuana? Sí	No
Resultado de prueba positivo (La prueba indica que la marihuana está *presente*).	119 (verdadero positivo)	24 (falso positivo)
Resultado de prueba negativo (La prueba indica que la marihuana está *ausente*).	3 (falso negativo)	154 (verdadero negativo)

- Si se elige al azar a uno de los 300 sujetos de prueba, calcule la probabilidad de que la persona resulte positiva, dado que en realidad consumió marihuana.
- Si se elige al azar a uno de los 300 sujetos de prueba, calcule la probabilidad de que la persona realmente haya consumido marihuana, dado que tuvo un resultado de prueba positivo

Probabilidad - Respuesta

$$P(\text{positivo} \mid \text{consumo de marihuana}) = \frac{P(\text{consumo de marihuana y positivo})}{P(\text{consumo de marihuana})}$$

$$= \frac{119/300}{122/300} = 0.975$$

$$P(\text{consumo de marihuana} \mid \text{positivo}) = \frac{P(\text{positivo y consumo de marihuana})}{P(\text{positivo})}$$

$$= \frac{119/300}{143/300} = 0.832$$