

Análisis e Interpretación de Datos

MÁSTER UNIVERSITARIO EN ANÁLISIS Y VISUALIZACIÓN DE DATOS
MASIVOS / VISUAL ANALYTICS AND BIG DATA

Miller Janny Ariza Garzón

Tema 5. Probabilidad Condicional y Variables Aleatorias I

Tabla de contenido

□ Tema 5: Probabilidad condicional y variables aleatorias

- Introducción a la teoría de la probabilidad.
- Principios de la teoría de la probabilidad.
- Probabilidad condicional e independencia

□ Actividad grupal: Definición de un problema estadístico: modelización y propuesta de soluciones

Contenido

Probabilidad condicional y variables aleatorias

Probabilidad condicional e independencia

$$P(A \text{ sabiendo que ha ocurrido } B) = P(A|B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

sucesos independientes
 $P(A|B)=P(A)$

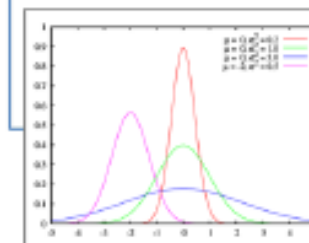
Distribución Binomial

$$X \sim Bi(n, p)$$

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

Distribución Normal

$$X \sim N(\mu, \sigma)$$



Variables aleatorias

Discretas

(un dado, una moneda etc.)

$$P(X = x_i)$$

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

$$E(X) = \sum x_i p_i$$

$$\sigma_x^2 = \sum (x_i - E(X))^2 p_i$$

Continuas

(peso, tiempo etc.)

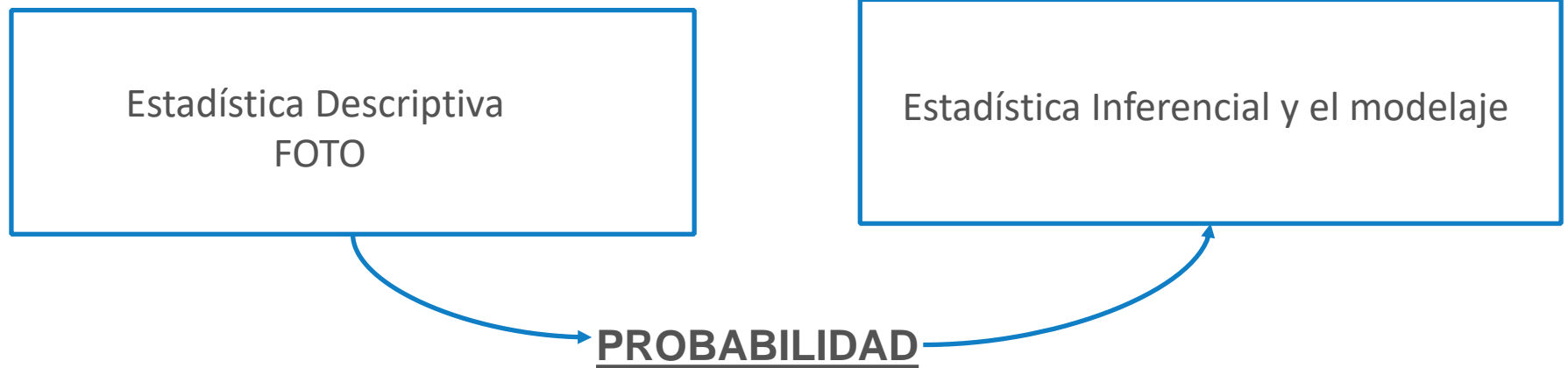
$$P(a < x < b) = \int_a^b f(x) dx$$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

$$E(X) = \int x f(x) dx$$

$$V(X) = \int (x - \mu)^2 f(x) dx$$

Introducción a la teoría de la probabilidad



Introducción a la teoría de la probabilidad

- **Experimento:** un procedimiento o proceso destinado a descubrir, estudiar o comprobar los resultados de un fenómeno.
- **Experimento aleatorio:** toda acción (procedimiento o proceso) definida donde se “conocen” o se pueden definir todos los resultados posibles de antemano pero existe **incertidumbre** sobre la ocurrencia del resultado. Aunque se repita de la misma manera el experimento, puede dar lugar a resultados diferentes
- **Espacio muestral:** es el conjunto de todos los resultados posibles de un experimento aleatorio, “ Ω ”.
- **Evento o suceso:** Un suceso “A”, es cualquier subconjunto del espacio muestral.

Introducción a la teoría de la probabilidad

Ejemplos de Experimentos:

- Lanzamiento de una moneda
- Lanzamiento de un dado
- Lanzamiento de dos dados
- Anotar tiempo de duración en un tipo de transporte. El espacio muestral del experimento es el intervalo $[0, T]$
- Registrar los posibles rendimientos que pueda generar una inversión. El espacio muestral podría ser un intervalo $[-A\%, B\%]$
- El contagiarse o no de coronavirus al viajar en metro. El espacio muestral tiene dos opciones: contagiarse y no contagiarse.
- Seleccionar al azar en una muestra de usuarios de Netflix y registrar el TIEMPO que pasó ayer viendo Netflix $[0, T]$
- Realizar una votación por la alcaldía de una ciudad con cuatro candidatos (A, B, C, D). Podría ganar **A**, **B**, **C** o **D**.

Introducción a la teoría de la probabilidad

Ejemplo clásico 1:

Experimento: Lanzamiento de un dado



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

Eventos:

- Obtener menos de 7: Ω , evento seguro
- Obtener más de 7: $\{ \} = \phi$, evento imposible
- Obtener un número primo par: $\{\boxed{2}\}$, evento simple, con $\boxed{2}$ un punto muestral
- Obtener un número par: $\{\boxed{2}, \boxed{4}, \boxed{6}\}$, eventos compuestos

es un

Los eventos por lo general se notan como los conjuntos con letras mayúsculas A, E, B,

...

Introducción a la teoría de la probabilidad



Ejemplo clásico 2:

Lanzo un par de dados y observo los resultados.

$$\Omega = \{(i, j): 1 \leq i \leq 6, 1 \leq j \leq 6\}$$

A: Obtener 8 en la suma de los valores resultantes:

$$\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\}$$

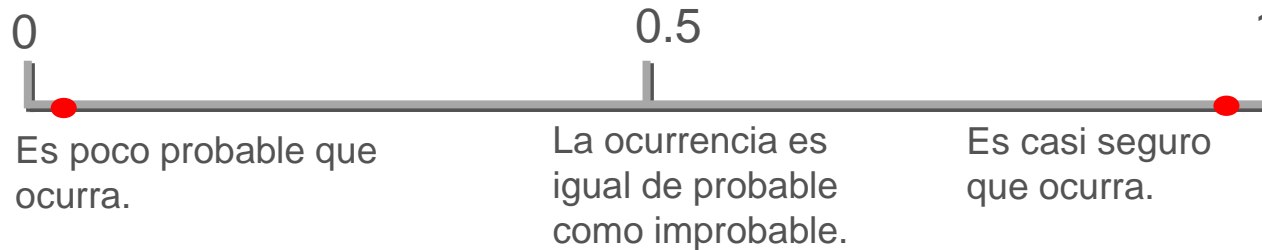
+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Introducción a la teoría de la probabilidad

➤ Probabilidad de un evento “ E ”, $P(E)$:

$$0 \leq P(E) \leq 1$$

La probabilidad describe numéricamente la posibilidad (oportunidad, certidumbre, “riesgo” o chance) de que ocurra el evento “ E ”.



➤ ¿Cómo hallar la probabilidad $P(E)$?

- aproximación frecuentista,
- aproximación subjetiva
- definición Laplaciana,

La elección depende de la disponibilidad de la información para un experimento estudiado.

Introducción a la teoría de la probabilidad

Aproximación Frecuentista

- Se basa en el número de veces que ocurre un evento E como proporción del número de intentos n .

$$fr(E) = \frac{n(E)}{n}, \text{ donde}$$

$n(E)$: número de veces que ocurre el evento E .

- Cuando un experimento aleatorio se realiza un número suficientemente grande de veces, bajo condiciones similares, la frecuencia relativa $fr(E)$ se estabiliza alrededor de un valor específico entre 0 y 1.

De tal forma que:

$$P(E) \cong \lim_{n \rightarrow \infty} fr(E)$$

- Sólo está bien definida para experimentos que pueden repetirse bajo condiciones de independencia.
- ❖ **Muy usada en encuestas y en métodos de simulación.**

Introducción a la teoría de la probabilidad

Aproximación Frecuentista

Ej. Lanzamiento de un dado y observar su resultado.

$$\Omega = \{i: 1 \leq i \leq 6\}$$

E : evento de obtener 3 = {3} (evento simple)



Para hallar la probabilidad de E se lanza el dado 1000 veces y se cuenta el número de veces en que se obtiene 3. Para nuestro caso obtuvimos 165 veces en que el resultado fue 3.

$$P(E) \cong fr(E) = \frac{165}{1000} \simeq 0.165$$

Introducción a la teoría de la probabilidad

Aproximación Subjetiva

Si se cuenta con poca información con la cual sustentar la probabilidad, es posible aproximarla en forma subjetiva (**basada en la experiencia-conocimiento previo del evento**).

Una probabilidad subjetiva no se basa en datos, ni en información histórica, y difiere de una persona a otra.

Se crea a partir de la **opinión** de un individuo. Expresa el grado de **creencia** p que tiene un “individuo” I acerca de que el evento E ocurra.

$$P(E) \cong P_I(E) = p$$

Introducción a la teoría de la probabilidad

Aproximación Subjetiva

Ej. Un inversionista A, de acuerdo con su experiencia, estableció las probabilidades para diferentes escenarios de beneficios en un portafolio de dos proyectos

Beneficios por proyecto	Beneficios total	Probabilidad
(10, 8)	18000 Euros	0.6
(10, -2)	8000 Euros	0.35
(-5, -8)	-13000 Euros	0.05

E : evento de obtener beneficios positivos, $E=\{(10,8), (10,-2)\}$

$$P(E) \cong P_A(E) = 0.35 + 0.6 = 0.95$$

Introducción a la teoría de la probabilidad

Definición Laplaciana

Aplica a experimentos que tienen un número finito de posibles resultados (puntos muestrales w), cada uno con la misma probabilidad de ser obtenido

$$P(w) = \frac{1}{|\Omega|}$$

Para $E \subseteq \Omega$,

$$P(E) = \sum_{w \in E} P(w) = \sum_{w \in E} \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|},$$

en otras palabras:

$$P(E) = \frac{\text{"número de resultados que satisfacen la condición del evento } E\text{"}}{\text{"número de resultados posibles del experimento"}}.$$

- ❖ En experimentos laplacianos se tiene que el cálculo de probabilidades se reduce al conteo del número de elementos de conjuntos finitos.

Esta definición motiva el estudio de la teoría del conteo y el análisis combinatorio (permutaciones, combinaciones, ...)

Introducción a la teoría de la probabilidad

Definición Laplaciana

Ej: Lanzo dos monedas y observo su resultado.

$$\Omega = \{CC, CS, SC, SS\}$$

E1: evento de obtener dos caras,

$$E1 = \{CC\}$$

E2: evento de obtener una sola cara,

$$E2 = \{SC, CS\}$$

$$P(E1) = \frac{|E1|}{|\Omega|} = \frac{1}{4}, \quad P(E2) = \frac{|E2|}{|\Omega|} = \frac{2}{4}$$

Ej3: Se van a realizar las elecciones de presidente en un país X donde los candidatos son: Andrés, Bernardo, Carlos, Dora.

$$\Omega = \{A, B, C, D\}$$



NO es un
experimento
Laplaciano

Principios básicos de la teoría de Probabilidad

Al ser los eventos conjuntos, las propiedades de la teoría de conjuntos se trasladan a la teoría de probabilidad.

Definición Axiomática:

Sea p una función $p: \Omega \rightarrow [0,1]$, talque para cualquier $A \subseteq \Omega$,

- $0 \leq P(A) \leq 1$
- $P(A) = \sum_{w \in A} P(w)$
- $P(\Omega) = \sum_{w \in \Omega} P(w) = 1$
- Si los eventos A y B son disjuntos
$$P(A \cup B) = P(A) + P(B)$$

Si el espacio muestral está conformado por infinitos eventos disjuntos A_i entonces:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Principios básicos de la teoría de Probabilidad

De los axiomas podemos deducir **otros resultados útiles**:

1. $P(\emptyset) = 0$

2. Si \bar{A} es el complementario de A , entonces $P(\bar{A}) = 1 - P(A)$

3. Si $A \subset B$ entonces $P(A) \leq P(B)$

4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Principios básicos de la teoría Probabilidad

Lanzamiento de un dado y observar su resultado.

$$\Omega = \{i: 1 \leq i \leq 6\}$$

- evento de obtener 7 = $\{ \}$. $P(\phi) = 0$
- A : Obtener 4, A^c : No obtener 4

$$P(A^c) = 1 - P(A) = 1 - \frac{1}{6} = \frac{5}{6}$$

- A : Obtener par, B : Obtener un número menor que 3

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$A = \{\boxed{2}, \boxed{4}, \boxed{6}\}$$

$$B = \{\boxed{1}, \boxed{2}\}$$

$$A \cup B = \{\boxed{1}, \boxed{2}, \boxed{4}, \boxed{6}\}$$

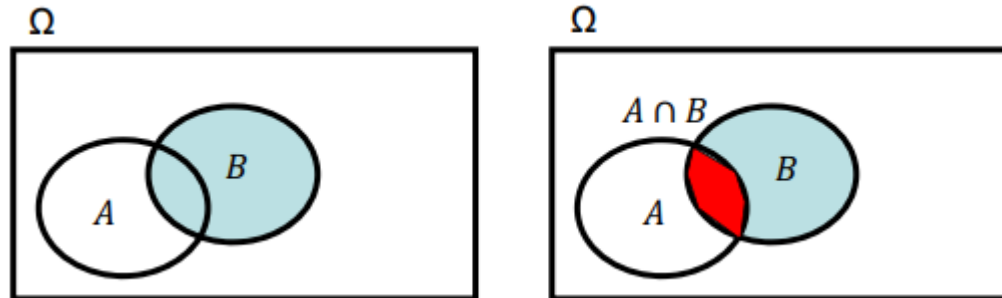
$$A \cap B = \{\boxed{2}\}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6}$$

Probabilidad condicional e independencia

Se llama probabilidad de A condicionada a B a:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



Es la medida de la parte de A condicionada al espacio muestral restringido B

Probabilidad condicional e independencia

Ej. Se lanzan dos dados :

- ¿Cuál es la probabilidad de que la suma de los dados sea 7?
- ¿Cuál es la probabilidad de que la suma de los dados sea 7 sabiendo que...?
 - La suma de los dados es impar.
 - La suma es mayor que 6.
 - El resultado del primer dado es impar.
 - Los dos dados tuvieron el mismo resultado.
 - Los dos dados tuvieron distinto resultado.

$S(2)$	$S(3)$	$S(4)$	$S(5)$	$S(6)$	$S(7)$	$S(8)$	$S(9)$	$S(10)$	$S(11)$	$S(12)$
(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)	(2, 6)	(3, 6)	(4, 6)	(5, 6)	(6, 6)
	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(3, 5)	(4, 5)	(5, 5)	(6, 5)	
		(3, 1)	(3, 2)	(3, 3)	(3, 4)	(4, 4)	(5, 4)	(6, 4)		
			(4, 1)	(4, 2)	(4, 3)	(5, 3)	(6, 3)			
				(5, 1)	(5, 2)	(6, 2)				
					(6, 1)					

Probabilidad condicional e independencia

Ej. Durante el año 2012-13 el número de estudiantes matriculados en cierta universidad española fue el siguiente:

	Derecho	Ingeniería Industrial	CC. Económicas	Total
Hombre	389	283	1156	1828
Mujer	483	52	728	1263
Total	872	335	1884	3091

$$P(\text{Derecho} \cap \text{Hombre}) = \frac{389}{3091} = 0,126$$

$$P(\text{Derecho} \text{ sabiendo que es hombre}) = \frac{389}{1828} = \frac{389/3091}{1828/3091} = \frac{P(D \cap H)}{P(H)}$$

Probabilidad condicional e independencia

Si sabiendo que ha ocurrido uno de los eventos no modifica la probabilidad del otro diremos que son **independientes**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- $P(A|B) = P(A)$, A no se ve afectado por la ocurrencia de B
 - Para que esto ocurra $P(A \cap B) = P(A) * P(B)$
-

¿ Son los eventos elegir Derecho y ser Hombre independientes?

$$P(D|H) = \frac{389}{1828} = 0.212$$

$$P(D) = \frac{872}{3091} = 0.282$$

NO

□ Tema 5: Probabilidad condicional y variables aleatorias.

- Variable aleatoria.
- Modelos discretos.
- Modelos continuos.

Actividad Grupal

Entrega: viernes, 27 de enero de 2023, 23:59.

Actividad Grupal: Todos deben entregar la Actividad. La entrega debe tener la lista de los miembros del grupo.

Tipo de documento: **artículo técnico (paper)** que brevemente comunique el resultado de vuestra investigación.

Tema: De actualidad (definimos actualidad como el período en que transcurre este cuatrimestre en que cursas la asignatura). Se propondrá un tema que se pueda valorar y estudiar críticamente mediante métodos propios de los estudios estadísticos. En este escenario, se debe presentar finalmente un artículo que responda críticamente a las preguntas propias de un estudio estadístico.

(Puede ser uno de los temas seleccionados de la actividad 1)

Actividad Grupal

Normas APA.

Máximo 4 ± 1 págs . (Sin incluir bibliografía)

Formato (preferiblemente): Fuente Georgia 11 e interlineado 1,5.

Códigos y datos: Se adjuntan en archivos separados

*Pueden seguir el formato de la plantillas (Opcional).

`tinytex::install_tinytex()`

Actividad Grupal

Estructura

- Se debe proponer título del trabajo,
- nombre de los autores y afiliación (supondremos para toda una única afiliación: UNIR),
- resumen del trabajo (1 párrafo),
- i. Introducción y Estado del Tema,
- ii. Metodología y Resultados Obtenidos
- iii Conclusiones y Líneas Futuras.
- Estos bloques pueden contener secciones, apartados con otros títulos que consideren más apropiados para transmitir los mensajes técnicos propios de vuestro trabajo.

Actividad Grupal

Debe contener

- **Enunciado del problema que se va a tratar** (indicar claramente la fuente que justifica que es un tema de actualidad).
- **Descripción del Modelo Estadístico** que consideran puede ayudar a dilucidar sobre el problema planteado.
- Buscar e incluir base de datos que permita calcular numéricamente sobre el modelo planteado
- Presentar resultados numéricos
- Discusión relativa a las conclusiones obtenidas.
- Presentar otro modelo estadístico diferente al ya planteado y volver a realizar el estudio numérico y el análisis de conclusiones. Analizar posibles discrepancias de los dos enfoques presentados.

Actividad Grupal

Nota

Tengan en cuenta que para el problema dado se plantean **dos modelos, estudios diferentes**. Esto puede ser, por ejemplo:

- **Dos enfoques descriptivos**
- **Un enfoque descriptivo y otro con modelos de regresión**
- **Un modelo de regresión y un modelo robusto de regresión**
- **Un enfoque descriptivo robusto y otro robusto de regresión**
- ...

Modelos distintos pueden en algunos casos llevar a conclusiones diferentes.

En este sentido, es importante analizar estas discrepancias volviendo al estudio de los modelos y tratando de distinguir cuál de los modelos presentados podría describir más acertadamente el escenario en cuestión.

Guía de la plantilla de entrega final del TFM (UNIR. Máster Universitario en Análisis y Visualización de Datos Masivos)

Resumen

En este apartado se introducirá un breve resumen en español del trabajo realizado (extensión entre 150 y 300 palabras). Este resumen debe incluir el objetivo o propósito de la investigación, la metodología, los resultados y las conclusiones.

El resumen debe contener lo qué se ha pretendido realizar (objetivo o propósito de la investigación), cómo se ha realizado (método o proceso desarrollado) y para qué se ha realizado (resultados y conclusiones).

Tiempo verbal: pasado

Introducción y Estado del Tema

Las ideas principales a transmitir son la identificación del problema a tratar, la justificación de su importancia, los objetivos generales a grandes rasgos y un adelanto de la contribución que esperas hacer.

Se incluye el dominio de aplicación, citando referencias. Debe aportar un buen resumen del conocimiento que ya existe en el campo del problema habituales identificados. Es el contexto general del trabajo

Tiempo verbal: presente

¿Porqué se hizo?

Metodología y Resultados Obtenidos

La metodología debe describir qué pasos se van a dar, el porqué de cada paso, qué instrumentos se van a utilizar, cómo se van a analizar los resultados, modelos, etc.

Deben describir el conjunto de datos.

Seleccionar los principales resultados estadísticos (Tablas y gráficos)

Resultados de los Modelos

Estadísticos comparativos

Tiempo verbal: pasado

¿Qué se hizo?

¿Qué se encontró?

- Datos
- Metodología
- Resultados

Conclusiones y Líneas Futuras

Presenta el resumen final del trabajo y debe servir para informar del alcance y relevancia de su aportación.

Suele estructurarse empezando con un resumen del problema tratado, de cómo se ha abordado y de por qué la solución sería válida.

Es recomendable que incluya también un resumen de las contribuciones del trabajo, discutiendo hasta qué punto se ha conseguido resolver los objetivos planteados.

Agregar un párrafo con limitaciones y perspectivas de futuro que abre el trabajo desarrollado para el campo de estudio definido.

Tiempo verbal: presente para referirse a trabajos ya publicados, y el pasado, para referir sus resultados actuales. La aportación se describe en presente.

¿Qué significan los resultados?

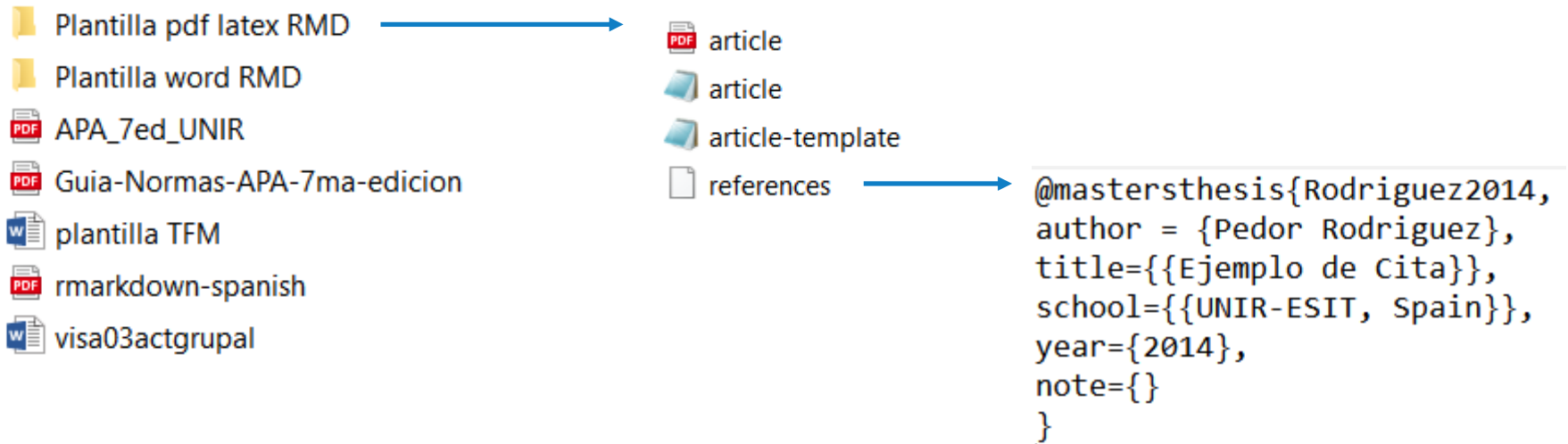
Actividad Grupal

Sección del documento	Tiempo recomendado	Ejemplo
Revisión de literatura (o cuando se discute el trabajo de otros investigadores)	Pasado _____ Pretérito perfecto	Martínez (2020) se refirió a _____ Investigadores han estudiado
Método Descripción del procedimiento	Pasado _____ Pretérito perfecto	Los participantes llenaron una encuesta. _____ Otros han utilizado enfoques similares
Reportando resultados propios o de otros investigadores	Pasado	Los resultados mostraron Los puntajes disminuyeron La hipótesis no fue respaldada
Reacciones personales	Pasado _____ Pretérito perfecto _____ Presente	Me sentí sorprendido _____ He experimentado _____ Yo creo
Discusión de implicaciones de resultados o de enunciados previos	Presente	Los resultados indican Los hallazgos significan que
Presentación de conclusiones, limitaciones, direcciones futuras y otros	Presente	Concluimos que Las limitaciones del estudio son Las futuras investigaciones deben explorar

TABLA CON TIEMPOS VERBALES RECOMENDADOS PARA LA REDACCIÓN DE TRABAJOS CON FORMATO APA SÉPTIMA EDICIÓN.

Actividad Grupal

Material complementario para la actividad:



Criterios. Rúbrica

Título de la actividad	Descripción	Puntuación máxima (puntos)	Peso %
Criterio 1	Definición del problema, base de datos y referencias	2	20
Criterio 2	Planteamiento de dos modelos estadísticos	2	20
Criterio 3	Análisis numérico de los datos	2	20
Criterio 4	Discusión y conclusiones a partir de los resultados obtenidos	2	20
Criterio 5	Escritura del artículo técnico	2	20
		10	100 %

Ejemplos:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0185638>

https://eprints.ucm.es/id/eprint/64153/1/CHI_21.pdf

<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9050779>



www.unir.net