Método Inductivo e Inferencia Estadística

El método inductivo tiene por finalidad la generalización del conocimiento que el científico obtiene a nivel particular, ya sean los resultados de la observación de un fenómeno o los datos obtenidos experimentalmente en el laboratorio. Una vez que los resultados son transformados en conocimiento, y este es generalizado, dicho conocimiento pasa a formar parte del conocimiento científico.

Desde un planteamiento estadístico la inducción o generalización es posible gracias a una expresión llamada estadístico. Un estadístico es una función θ de los datos o valores de la muestra aleatoria $\{x_1, x_2, \ldots, x_n\}$, es decir $\theta = f(x_1, x_2, \ldots, x_n)$. Por ejemplo, la media aritmética, la varianza muestral o una proporción son estadísticos. Ahora bien ¿cuál es el fundamento de la inferencia estadística?

En el contexto experimental el valor de un parámetro poblacional (θ) es desconocido, mientras que los valores de los parámetros muestrales (θ) son conocidos al obtenerse experimentalmente. A partir de los parámetros muestrales podemos obtener información acerca de los valores de los parámetros poblacionales. En particular, cada parámetro muestral será útil para inferir información de su correspondiente parámetro poblacional, tal y como se muestra en la Tabla I.

Tabla I.- Parámetros poblacionales y muestrales

Parámetros poblacionales	Parámetros muestrales
μ (media poblacional)	$\overline{x}$ (media aritmética)
σ^2 (varianza poblacional)	s^2 (varianza muestral)
P (proporción en la población)	p (proporción en la muestra)

En general, la inferencia estadística consta de tres clases de técnicas, las dos primeras tienen como finalidad la estimación de un parámetro poblacional mientras que la tercera permite evaluar la aceptación o rechazo de una hipótesis:

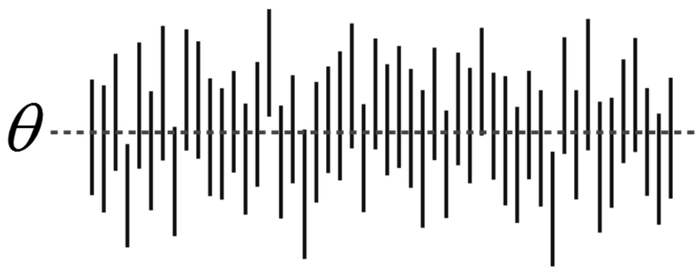
- Estimación puntual: La información suministrada por los datos, es decir la muestra aleatoria $\{x_1, x_2, ..., x_n\}$, es utilizada para obtener un valor estimado de la media, varianza o proporción poblacionales (μ, σ^2, P) . Para cada uno de estos parámetros poblacionales la estimación es el valor numérico obtenido de la media aritmética, la varianza muestral y la proporción muestral, aplicando en cada caso la correspondiente expresión o estadístico:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2 \quad , \quad p = \frac{x}{n}$$

En una proporción muestral x es el número de individuos o unidades de análisis que presentan una cierta característica o atributo.

Las expresiones anteriores o estadísticos se denominan estimadores puntuales.

- Intervalo de confianza: Se trata de una técnica de estimación que consiste en construir un intervalo, popularmente llamado horquilla, obteniendo los valores de los extremos del intervalo.



Un intervalo de confianza tiene la forma:

$$(\text{LRL} \leq \theta \leq \text{URL})$$

siendo LRL y URL los límites reales inferior y superior respectivamente. Los límites del intervalo se obtienen a partir del parámetro muestral (θ), la distribución muestral del estadístico, es decir $f(\theta)$ y el error estándar del parámetro (σ_θ):

$$\theta \pm f(\theta)\,\sigma_\theta$$

- Contraste de hipótesis: La estadística descriptiva aplicada a los datos experimentales permite al investigador obtener una primera impresión sobre el fenómeno objeto de estudio. Sin embargo lo habitual es que aún no se disponga de una hipótesis o conjetura sobre el fenómeno, realizándose un experimento piloto en campo o laboratorio, a partir del que obtendremos un conjunto de datos u observaciones. A partir de los resultados del análisis estadístico descriptivo realizado con dichos datos el investigador establece una afirmación o conjetura que representa una hipótesis científica **H**.

¿Cómo surgen las hipótesis científicas?

A continuación, esta hipótesis será expresada en términos cuantitativos, constituyendo lo que se conoce como hipótesis estadística. En estadística hay dos clases de hipótesis, la hipótesis nula (H0) y la alternativa (Ha):

H0: "Los datos apoyan la hipótesis H"

Ha: "Los datos no apoyan la hipótesis H"

La H0 es la hipótesis principal que es objeto de evaluación por medio del método científico o método hipotético-deductivo.

Si la hipótesis es cuantitativa entonces se construye con los parámetros poblacionales (θ), por ejemplo la media, varianza o proporción poblacionales (μ, σ^2, P) . En tales casos la hipótesis puede ser simple cuando el parámetro es igual a un valor concreto ($\theta = \theta_0$) o compuesta cuando es mayor o menor a dicho valor ($\theta > \theta_0, \theta < \theta_0$). En muchos experimentos el valor θ_0 es el valor propuesto o conjeturado.

En otros casos la H0 es una afirmación de tipo estadístico o sobre alguna teoría en particular, por ejemplo "los datos se ajusta a la primera ley de Mendel".

Cuando el contraste es sobre uno o más parámetros, sea cual sea la situación experimental, el contraste será uno de los tres que se muestran a continuación:

[1]	H0: $\theta = \theta_0$ Ha: $\theta \neq \theta_0$	[2]	H0: $\theta = \theta_0$ Ha: $\theta < \theta_0$	[3]	H0: $\theta = \theta_0$ Ha: $\theta > \theta_0$

El contraste [1] es de tipo bilateral, y es útil para responder a preguntas generales, por ejemplo "¿hay diferencias significativas". Por el contrario, los contrastes [2] y [3] son de tipo unilateral, siendo útiles cuando la pregunta es direccional. Por ejemplo, en [3] un investigador "ha observado que el parámetro poblacional tiene un valor que es como mínimo θ_0" .

Técnica del contraste de hipótesis

La técnica estadística con la que llevar a cabo dicha evaluación, y por tanto la decisión de aceptar o rechazar la hipótesis nula H0, se llama contraste de hipótesis.

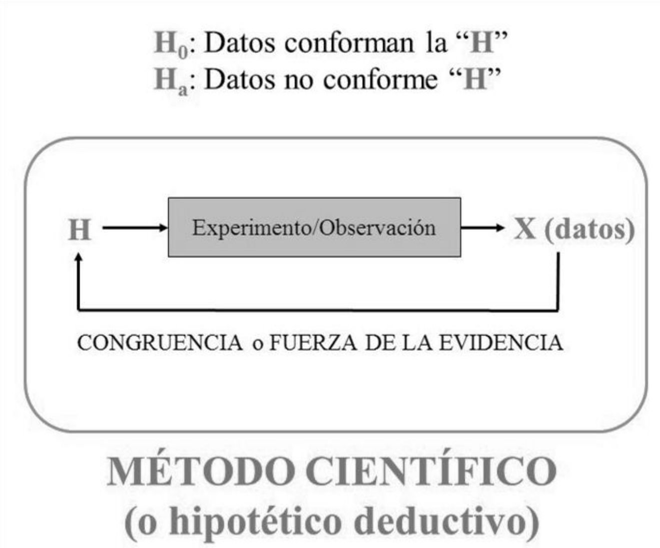
El contraste de hipótesis utiliza la información suministrada por los datos, es decir la muestra aleatoria $\{x_1, x_2, \ldots, x_n\}$, ya sea a través del valor de la media aritmética, la varianza muestral o el valor de una proporción muestral. Además, y de forma similar a un intervalo de confianza, la técnica permite al investigador especificar el riesgo que conlleva aceptar o rechazar la hipótesis nula H0. Este riesgo se fija por medio del concepto de nivel de confianza o valor $1-\alpha$. Este valor representa la probabilidad de aceptar la H0 cuando dicha hipótesis es verdadera. El valor α se conoce como nivel de significación, siendo la probabilidad de rechazar H0 cuando la hipótesis es verdadera.

El método científico y el contraste de hipótesis

El método científico o método hipotético-deductivo tiene sus orígenes en la figura del filósofo Descartes, y desde un planteamiento estadístico consiste en el diseño de un experimento con el que obtener datos u observaciones que nos permitan aceptar o rechazar la hipótesis nula o H0.

Sello conmemorativo a la figura de Descartes

Por consiguiente, asumiremos que los datos experimentales que analizamos con métodos de estadística descriptiva procedían de un experimento piloto o preliminar, a partir de cuyo resultado se construyó una hipótesis científica **H**. Es entonces cuando **H** es expresada cuantitativamente en una hipótesis estadística, ya sea H0 o Ha, diseñándose un experimento *ad hoc* con el que evaluar H0.



El maridaje de dos metodologías: el método científico y el contraste de hipótesis

Por lo general H0 representa la hipótesis que es aceptada por la comunidad científica, ubicando el investigador en Ha la hipótesis candidata, o conjetura propuesta. En tal caso, y desde el punto de vista del investigador, lo que éste desea es que el contraste de hipótesis conduzca al rechazo de H0, aceptándose la Ha. En tal caso la Ha se convertirá en la nueva H0 que será admitida por la comunidad científica hasta que surja una nueva conjetura o hipótesis candidata a sustituir a H0. En otros casos, por ejemplo en una situación experimental de control de calidad siempre ubicaremos el apto de un producto en la Ha, evitándose así el llamado error de tipo II.

Por consiguiente, la ciencia a través del método científico construye el conocimiento por medio de la sustitución sucesiva de unas hipótesis nulas por otras candidatas o alternativas que llegado el caso serán las nuevas H0. El contraste de hipótesis es la técnica estadística con la que llevar a cabo dicha decisión, la aceptación o rechazo de H0, en función de la muestra obtenida y el nivel de confianza fijado por el investigador.

El experimento u observación que realicemos debe estar diseñado cumpliéndose que (1) su protocolo experimental pueda ser *reproducible* por otros investigadores, y (2) que en base a dicho protocolo los datos obtenidos en el experimento puedan conducir a *refutar* la hipótesis H0.

En resumen, en un contraste de hipótesis medimos la congruencia de la H0, es decir evaluamos cuantitativamente hasta que grado los datos experimentales apoyan la aceptación de la hipótesis nula.

Un aspecto importante del método científico/contraste de hipótesis es que sólo permite aceptar o rechazar una hipótesis, por tanto nunca sabremos si lo que aceptamos o rechazamos es verdadero o falso. Esto significa que al "hacer ciencia", es decir investigación, hay siempre tres personajes: (a) el científico que aplicando el contraste de hipótesis sólo puede aceptar o rechazar H0, (b) un ente superior o "demiurgo" que conoce si H0 es verdadera o falsa, y finalmente (c) la ciencia que acepta provisionalmente la H0 hasta que una hipótesis candidata, Ha, la sustituya. Es decir, la veracidad o falsedad de una hipótesis es algo cuya evaluación no es posible aplicando el método científico, y por tanto que no es objeto de la ciencia ni esta a su alcance. Por consiguiente, y de acuerdo con la Tabla 2 que se muestra a continuación, experimentalmente hay cuatro situaciones posibles.

Tabla 2.- Escenarios posibles en un contraste de hipótesis

	H0 Verdadera	H0 Falsa
Aceptar H0	$1-\alpha$ (1)	β (3)
Rechazar H0	α (2)	$1-\beta$ (4)

En el escenario (1) los datos de un experimento y el contraste de hipótesis nos conducen a aceptar la H0 cuando H0 es verdadera, siendo la probabilidad de que esto ocurra lo que se llama nivel de confianza. Se suele trabajar a un 90%, 95% o 99% de nivel de confianza (0.90, 0.95, 0.99). La situación (2) se conoce como error de tipo I o falso positivo ya que rechazamos la H0 cuando dicha hipótesis es en realidad verdadera. La probabilidad de error de tipo I se llama nivel de significación, siendo los valores habituales 0.10, 0.05 y 0.01. Sin embargo (3) es el error más peligroso en términos científicos y al que se conoce como error de tipo II o falso negativo, ya que aceptamos una H0 que en realidad es falsa. Finalmente, (4) es la sensibilidad de una prueba o test estadístico, y cuya probabilidad se denomina potencia de una prueba.

En general (1) y (2) al ser complementarios son elegidos por el investigador, por ejemplo un 95% para situaciones generales y un 99% cuando la aceptación o rechazo de H0 impliquen riesgos para la salud, económicos etc. Por el contrario, (4) no puede ser fijado de antemano por el investigador. El único procedimiento de aumentar la potencia de una prueba, y por tanto la sensibilidad de un test estadístico reduciéndose el error de tipo II, es aumentando el tamaño muestral (n).

Criterio de decisión

El paso final en un test estadístico o contraste de hipótesis consiste en obtener una conclusión estadística, conclusión que se resume en la aceptación o rechazo de la H0. Una vez decidida la aceptación o rechazo de la hipótesis nula traduciremos dicha decisión en una conclusión a nivel experimental, es decir si un tratamiento es o no eficaz, si hay o no diferencias significativas entre dos grupos experimentales o si unos datos se ajustan o no a un cierta ley de probabilidad (por ejemplo la distribución normal) entre otras muchas posibilidades.

Desde un punto de vista metodológico la aceptación o rechazo de la H0 requiere de una regla de decisión, distinguiéndose dos técnicas. Por lado, la decisión se basa en el cálculo del p-valor, método que se utiliza en la práctica en diseño experimental. De otro, la decisión se sustenta en el cálculo de valores críticos de la distribución muestral utilizándose para tal fin tablas estadísticas; método cuya aplicación se limita al ámbito académico, cursos de formación y seminarios pero que no es utilizado en el ámbito experimental.

- Método del p-valor

  Desde un punto de vista experimental el criterio utilizado para aceptar o rechazar la H0 se basa en el cálculo del llamado p-valor. En la actualidad el uso habitual de software estadístico hace que el resultado de un análisis estadístico se resuma finalmente en el valor p. El concepto de p-valor tiene varias definiciones: (1) es el nivel de significación más bajo para el que habría que rechazar la H0, (2) es la probabilidad de obtener la muestra aleatoria $\{x_1, x_2,..., x_n\}$ observada, es decir los datos que hemos obtenido experimentalmente, y (3) puede ser interpretado en términos experimentales como una medida de congruencia de los datos con la hipótesis nula.

Un aspecto interesante del p-valor es que una vez calculado dejamos al lector de nuestro trabajo que sea el que decida si acepta o rechaza la hipótesis H0. Por consiguiente, y de acuerdo con lo anterior, será el lector de un trabajo científico el que aplicará la siguiente regla de decisión:

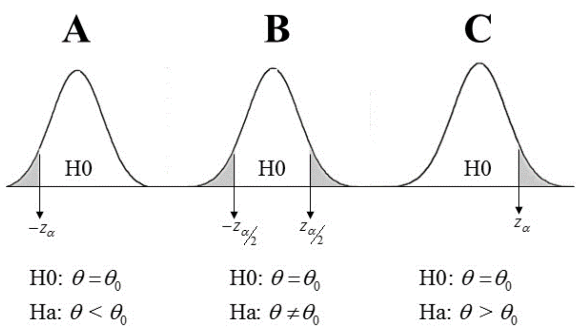
Sí p-valor > α entonces aceptamos H0
Sí p-valor < α entonces rechazamos H0

Según este criterio dejamos que sean otros los que asuman el riesgo de la decisión, eligiendo cada cual el nivel de significación que considere más adecuado. En situaciones experimentales generales elegiremos un nivel de significación igual a 0.05.

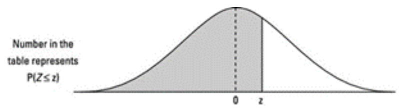
- Método de los valores críticos: usando tablas estadísticas

Es una técnica que no se utiliza en la práctica. La idea es obtener los valores críticos de la distribución muestral, tal que se comprueba si el valor del estadístico cae en la región o no de aceptación de la H0. El área de la región de aceptación es el nivel de confianza, mientras que las regiones en 'gris' corresponden al nivel de significación, y por tanto de rechazo de H0 y aceptación de Ha.

Ejemplo casos A-B-C: En la siguiente figura se muestra un ejemplo para las tres clases de contraste cuando la distribución muestral es la normal estándar o distribución *Z*.



Los valores críticos se obtienen utilizando tablas estadísticas:



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

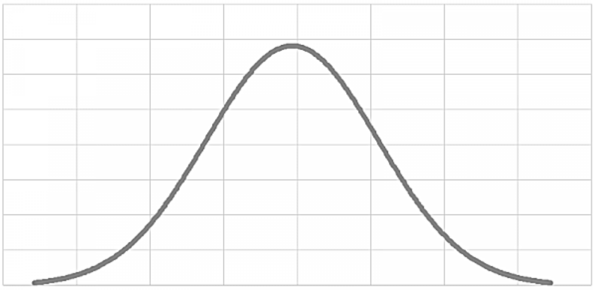
aunque con Python también podemos obtenerlos (**Ejemplo casos A-B-C [ver código en Python]** en esta página web).

Normalidad de la variable

Uno de los supuestos más importantes en inferencia estadística es que la distribución de la variable aleatoria que es objeto de estudio debe tener o aproximarse a la distribución normal $N(\mu,\sigma)$ o de Gauss:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Las pruebas estadísticas que requieren del cumplimiento de este supuesto se llaman test paramétricos. En la práctica las pruebas o test paramétricos tienen una prueba o test equivalente que no requiere dicho supuesto, y que se conoce como test no paramétrico. Por consiguiente, en muchas situaciones experimentales, es costumbre realizar las dos clases de pruebas simultáneamente. En aquellos casos en los que la variable no es continua o es continua pero sin distribución normal, procederemos a realizar el test no paramétrico.



Distribución normal

Con el fin de testar la normalidad de la variable aleatoria, hay varias pruebas o test estadísticos especialmente diseñados para realizar el siguiente contraste:

H0: "La variable X es $N(\mu,\sigma)$"
Ha: "La variable X no es $N(\mu,\sigma)$"

Estos test estadísticos son específicos para evaluar la normalidad de una variable aleatoria. Entre otros, destacamos tres:

- Test de Saphiro-Wilks: Se trata de una prueba muy utilizada para testar la normalidad. El test fue propuesto en 1965, y utiliza el estadístico:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

- Test de D' Agostino.
- Test de Anderson-Darling.

En las pruebas anteriores:

Sí p-valor > α entonces aceptamos H0 y por tanto aceptamos normalidad de X

Sí p-valor < α entonces rechazamos H0 y por tanto rechazamos normalidad de X

En general se elige un nivel de significación igual a 0.05.

Si rechazamos la normalidad entonces recurriremos al test no paramétrico. No obstante, hay otras pruebas o tests generales con los que es posible estudiar el ajuste de una variable aleatoria a cualquier otra distribución distinta de la normal. Entre estas pruebas las más populares son:

- Test de chi-cuadrado: El test de chi-cuadrado es útil cuando se evalúa el ajuste de una variable aleatoria discreta a una distribución discreta. Se utiliza el estadístico con distribución muestral chi-cuadrado con ν grados de libertad:

$$\chi^2 = \frac{(O-E)^2}{E}$$

  siendo O y E las frecuencias observadas y esperadas si los datos se ajustaran a una cierta distribución respectivamente.

- Test de Kolmogorov-Smirnov. Es un test o prueba que es de uso general siendo útil tanto en variables discretas como continuas. La prueba evalúa como los datos se alejan de los valores que deberíamos haber observado si efectivamente la variable se ajustara a una cierta distribución. Elegida por el investigador una cierta función de distribución *F(x)*, deseamos evaluar si los datos se ajustan o no a dicha distribución. El test obtiene los valores que se alejan de forma máxima tanto por exceso como por defecto, es decir los valores D^+ y D^-, obteniéndose el estadístico K-S:

$$D_n = \sup_x \left| F_n(x) - F(x) \right|$$

De forma similar en estas pruebas si p-valor > α entonces aceptaremos H0, y por tanto que la variable X se ajusta a una cierta distribución *F(x)*.

Rafael Lahoz-Beltrá, Pilar López González-Nieto, Mariángeles Gómez Flechoso, María Eugenia Arribas, Mocoroa, Alfonso Muñoz Martín, María de la Luz García Lorenzo, Gloria Cabrera Gómez, Jose Antonio Alvarez Gómez, Andrea Caso Fraile, Jefferson Mark Orosco Dagan, Raul Merinero Palomar. Universidad Complutense de Madrid, 2017.