



Capítulo VII

Análisis de la variable perimeter en cáncer de mama

DOI:<https://doi.org/10.58299/utp.268.c933>

Gerardo Martínez Guzmán
gerardo.martinezgu@correo.buap.mx
<https://orcid.org/0000-0002-0861-7932>

Jorge Alejandro Fernández Pérez
jorge.fernandez@correo.buap.mx
<https://orcid.org/0000-0002-5275-3309>

Carmen Cerón Garnica
carmen.ceron@correo.buap.mx
<https://orcid.org/0000-0001-6480-6810>

Benemérita Universidad Autónoma de Puebla



Resumen

El diagnóstico de células cancerosas a partir biopsias, se basa principalmente en el análisis de cambios morfológicos de la estructura nuclear como el aumento del tamaño nuclear, que ocurre posiblemente debido a la desregulación del ciclo celular, así como del crecimiento celular. El incremento del tamaño nuclear se observa en biopsias de pacientes con diagnóstico benigno y maligno. Una variable perimeter relacionada con el incremento del tamaño nuclear en pacientes con diagnóstico benigno y maligno, es estudiada en este trabajo. Un análisis de esta variable prueba, mediante el algoritmo de aprendizaje no supervisado, maximización de la esperanza (EM), que dicha variable tiene un comportamiento tipo mezcla de normales con dos componentes. Tal algoritmo es capaz de discriminar los datos en dos grupos (malignos y benignos). Dicho modelo proyecta una clasificación con un porcentaje alto de coincidencia con los datos observados.

Introducción

El diagnóstico de células cancerosas hecho con base en las biopsias se basa principalmente en el análisis de cambios morfológicos de la estructura nuclear (Hanahan & Weinberg, 2011). Un aspecto importante es el aumento del tamaño nuclear, que ocurre posiblemente debido a la desregulación del ciclo celular, así como del crecimiento celular (Denais & Lammerding, 2014).

Las distribuciones de mezclas finitas se han empleado para la modelización de datos heterogéneos, puesto que, en muchos casos no es suficiente explicar la distribución de unos datos mediante una única distribución estadística, por lo que es necesario la utilización de una combinación de distribuciones. Es decir, las combinaciones mixtas son utilizadas para la modelización de datos que en muchas situaciones experimentales pueden interpretarse como procedentes de dos o más subpoblaciones. La obtención de estas componentes conduce necesariamente a la estimación de los parámetros y a las proporciones en las que cada una de las componentes contribuye a la distribución general, (McLachlan & Peel, 2000; Mengerser *et al.*, 2011).

Ya que se requiere del agrupamiento del conjunto de observaciones en grupos que comparten algunas características, los modelos de mezclas son utilizados para aplicaciones médicas (Schlattmann, 2009).

El algoritmo Maximización de la Esperanza (ME) citado por Levine, R. & Casella, G. (2001) y por McCulloch (1998) es una herramienta habitual iterativa para la estimación de máxima verosimilitud de las distribuciones mixtas. La idea es introducir una variable indicadora multinomial que identifica la pertenencia a un grupo de cada observación del conjunto de datos. El algoritmo EM fue expuesto por Arthur Dempster, Nan Laird y Donald Rubin de la Royal Statistical Society en una publicación de 1977.

Análisis de los datos

Los datos que se tienen para el estudio es una muestra de 569 mujeres, creada por el Dr. William H. Wolberg, médico del Hospital de la Universidad de Wisconsin en Madison, Wisconsin, EE. UU. (Wolberg *et al.* 1995). Cada una de las

mujeres de la muestra se identifica con el ID, el diagnostico si el tumor es benigno o maligno y el valor de la variable perimeter, entre otras. Del total de la muestra existen 357 casos que resultaron benignos y 212 de los casos maligno. Mostramos algunos datos de la base en la Tabla 1.

Tabla 1

Muestra de la base de datos

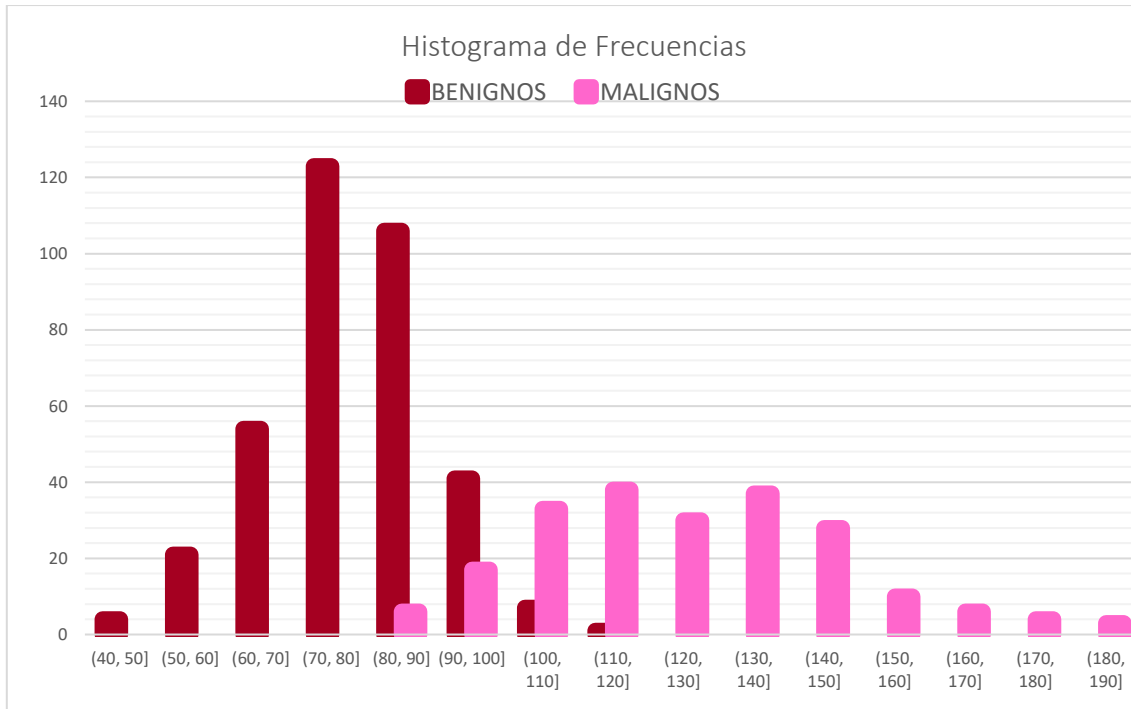
ID	Diagnostico	perimeter
	Benigno (B), Maligno (M)	
859464	B	60.11
859465	B	71.8
859471	B	58.79
859487	B	81.37
859575	M	123.6
859711	B	58.79
859717	M	114.2
859983	M	90.43
8610175	B	79.19
8610404	M	104.1
8610629	B	87.91
8610637	M	120.2
8610862	M	143.7
8610908	B	83.19
861103	B	73.81
8611161	B	86.49
8611555	M	171.5
8611792	M	129.1
8612080	B	76.95
8612399	M	121.1
86135501	M	94.25
86135502	M	122
861597	B	79.78
861598	B	95.77

Fuente: Elaboración propia con datos de. (Wolberg *et al.* 1995).

El histograma de una muestra proporciona una estimación de la forma de la función de densidad. Sin embargo, no es la densidad, pero desde el punto de vista no paramétrico puede ser visto como una estimación razonable de la misma. Entonces si consideramos la variable *perimeter* y tomamos en cuenta que, de las 569 observaciones totales, 357 observaciones indican la ausencia de células cancerosas, mientras que 212 observaciones muestran la presencia de células cancerosas, el histograma de frecuencias de la variable con el diagnóstico, presenta una forma de mezcla de dos normales como se observa en la Figura 1.

Figura 1

Histograma de frecuencias de casos benignos y malignos



Fuente: Elaboración propia con datos de Wolberg *et al.* (1995).

Observando el histograma de frecuencias de la muestra, se tiene un comportamiento del tipo mezcla de normales, por lo que desarrollaremos el trabajo considerando que la variable *perimeter*, tiene una distribución de probabilidad completamente especificada, a saber, una mezcla de dos distribuciones normales, donde encontraremos una estimación de sus parámetros, mediante el método

estadístico maximización de la esperanza (EM). La variable perimenter nos puede dar información para predecir si el diagnóstico del cáncer de seno es benigno o maligno, ya que tal algoritmo es capaz de discriminar los datos en grupos.

Algoritmo EM

En el desarrollo del algoritmo EM, proporcionamos una formulación paramétrica para la representación del modelo. En lo sucesivo, mediante $Y = (Y_1, Y_2, \dots, Y_n)$, se denotará a una muestra aleatoria de tamaño n , donde Y_i es un vector aleatorio p -dimensional con función de densidad $f(y_i)$ donde $y_i \in R^p$. Así $y = (y_1, y_2, \dots, y_n)$ representa una muestra observada de Y .

Definición 1. Si la función de densidad de una variable aleatoria Y_i es de la forma

$$f(y_i|\boldsymbol{\psi}) = \sum_{k=1}^g \pi_k f_k(y_i|\boldsymbol{\theta}_k) \quad y_i \in R^p$$

se dice que posee una distribución de mezcla finita con g componentes, con un vector de parámetros

$$\boldsymbol{\psi} = (\pi_1, \dots, \pi_g, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_g).$$

Aquí, $f_k(y_i|\boldsymbol{\theta}_k)$, $k = 1, 2, \dots, g$, denotan las densidades de las componentes de la mezcla con parámetros $\boldsymbol{\theta}_k$ y parámetros de peso π_1, \dots, π_g . En el caso más general se asume también que las funciones $f_k(y_i|\boldsymbol{\theta}_k)$ pueden pertenecer a diferentes familias paramétricas.

Para que la mezcla sea una función de densidad los pesos deben cumplir con las condiciones,

$$0 \leq \pi_k \leq 1 \quad k = 1, \dots, g \quad y \quad \sum_{k=1}^g \pi_k = 1$$

Note que en la condición anterior uno de los pesos resulta redundante (uno de ellos se expresa en términos de los demás).

Definición 2. Sea $y = (y_1, y_2, \dots, y_n)$ observaciones independientes de una variable aleatoria, cuya función de densidad $f(y|\boldsymbol{\psi})$ es una mezcla, entonces la función

$$L(\boldsymbol{\psi}|y) = \prod_{i=1}^n f(y_i|\boldsymbol{\psi}) = \prod_{i=1}^n \sum_{k=1}^g \pi_k f_k(y_i|\boldsymbol{\theta}_k)$$

recibe el nombre de función de verosimilitud de la mezcla.

Tomando logaritmo natural en $L(\boldsymbol{\psi}|y)$ obtenemos su función log-verosimilitud

$$\begin{aligned} l(\boldsymbol{\psi}|y) &= \log L(\boldsymbol{\psi}|y) = \log \prod_{i=1}^n \left\{ \sum_{k=1}^g \pi_k f_k(y_i|\boldsymbol{\theta}_k) \right\} \\ &= \sum_{i=1}^n \log \left\{ \sum_{k=1}^g \pi_k f_k(y_i|\boldsymbol{\theta}_k) \right\}. \end{aligned}$$

Para calcular el estimador de máxima verosimilitud $\hat{\boldsymbol{\psi}}$ es común utilizar el logaritmo de la función de verosimilitud, pues recordemos que la función y el logaritmo de la función bajo ciertas condiciones de regularidad toman en el mismo punto su máximo. Por lo tanto, debemos resolver la ecuación de verosimilitud,

$$\frac{\partial}{\partial \boldsymbol{\psi}} \sum_{i=1}^n \log \left\{ \sum_{k=1}^g \pi_k f_k(y_i|\boldsymbol{\theta}_k) \right\} = 0$$

Debido a la presencia del logaritmo de una suma, es difícil la resolución de la ecuación, por tal motivo se requiere otro tipo de procedimiento que permitan la maximización de la función log-verosimilitud.

El nuevo procedimiento fue introducido por Dempster *et al.* (1977) como un mecanismo para manejar información ausente y consiste en definir una nueva esperanza o expectativa que facilita la maximización, de tal forma que los parámetros que maximizan a esta expectativa en cada iteración convergen a los parámetros que maximizan a la función de verosimilitud.

Sea $y = (y_1, y_2, \dots, y_n)$ una muestra observada de tamaño n , a la que denominaremos vector de datos incompletos, correspondientes a una realización de Y , con función de densidad $f(y|\boldsymbol{\psi})$, donde $\boldsymbol{\psi}$ es el vector de parámetros a estimar. Ahora considere la variable $Z = (Z_1, Z_2, \dots, Z_n)$ que denominaremos latente, que representa a los datos no observados y cuya realización es $z = (z_1, z_2, \dots, z_n)$. Entonces el vector aleatorio $X = (Y, Z)$ recibe el nombre de vector de datos completos y su realización es $x_1 = (y_1, z_1), x_2 = (y_2, z_2), \dots, x_n = (y_n, z_n)$ de tal forma que a cada realización y_i le corresponde siempre una z_i .

En este contexto podemos suponer que Z_i representa una variable indicadora binaria g -dimensional cuyo elemento j -ésimo Z_{ij} indica la pertenencia de la observación y_i a la componente j -ésima de la mezcla donde $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, g$. Así, podemos definir Z_{ij} como,

$$Z_{ij} = z_{ij} = \begin{cases} 1 & \text{Si } y_i \text{ proviene de la} \\ & \text{componente } j \text{ - esima.} \\ 0 & \text{en otro caso} \end{cases}$$

Dada la naturaleza categórica de la variable Z_i al indicar la pertenencia de los puntos muestrales a una componente u otra de la mezcla, se puede interpretar que los pesos π_k es la probabilidad a priori de que la observación y_i pertenezca a la población k , lo que hace suponer que Z_i sigue una distribución multinomial de una sola realización sobre g categorías con probabilidades $\pi = (\pi_1, \pi_2, \dots, \pi_g)$, es decir,

$$P(Z_i = z_i) = \binom{1}{z_{i1}, z_{i2}, \dots, z_{ig}} \pi_1^{z_{i1}} \pi_2^{z_{i2}} \dots \pi_g^{z_{ig}} = \prod_{k=1}^g \pi_k^{z_{ik}}$$

Donde

$$\sum_{k=1}^g z_{ik} = 1 \quad \sum_{k=1}^g \sum_{i=1}^n z_{ik} = n.$$

Entonces,

$$P(z_{ik} = 1) = \pi_k \quad k = 1, 2, \dots, g.$$

Modelo de mezclas Gaussianas

Se puede obtener información acerca de los datos ausentes mediante la observación de las elecciones que se realizaron. Por lo que definimos a $h(z|y, \boldsymbol{\psi})$ como la densidad de los datos ausentes condicionada a las elecciones observadas en la muestra. Así, por el teorema de Bayes, tenemos

$$\begin{aligned} h(z|y, \boldsymbol{\psi}) &= P(z_{ik} = 1 | Y_i = y_i) = \frac{P(z_{ik} = 1)P(Y_i = y_i | z_{ik} = 1)}{P(Y_i = y_i)} \\ &= \frac{\pi_k f_k(y_i | \boldsymbol{\theta}_k)}{\sum_{k=1}^g \pi_k f_k(y_i | \boldsymbol{\theta}_k)}. \end{aligned}$$

Tomando en cuenta lo anterior, aquí definimos la nueva esperanza o expectativa en ψ que se relaciona con la función de verosimilitud pero que utiliza la distribución condicionada $h(z|y, \psi)$. El procedimiento EM es iterativo y comienza con un valor inicial de los parámetros ψ^0 y en cada iteración se van actualizando. Como es conocido la maximización repetida de esta nueva función converge al máximo de la propia función de verosimilitud. Esta nueva función es,

$$\begin{aligned}\mathcal{E}(\psi|\psi^0) &= E[l(\psi|y, z)|Y = y, \psi^0] \\ \mathcal{E}(\psi|\psi^0) &= E\left[\sum_{i=1}^n \sum_{k=1}^g z_{ik} \log[\pi_k f_k(y_i|\theta_k)]|Y = y, \psi^0\right] \\ &= \sum_{i=1}^n \sum_{k=1}^g E[z_{ik}|Y_i = y_i, \psi^0] [\log \pi_k + \log f_k(y_i|\theta_k)]\end{aligned}$$

Sin embargo:

$$\begin{aligned}E[z_{ik}|Y_i = y_i, \psi^0] &= P(z_{ik} = 1|Y_i = y_i, \psi^0) \\ &= \frac{f_k(Y_i = y_i|z_{ik} = 1)P(z_{ik} = 1)}{P(Y_i = y_i)} \Big|_{\psi^0} \\ &= \frac{\pi_k f_k(y_i|\theta_k)}{\sum_{k=1}^g \pi_k f_k(y_i|\theta_k)} \Big|_{\psi^0} = \hat{t}_{ik}^{(0)}\end{aligned}$$

Por lo tanto,

$$\begin{aligned}\mathcal{E}(\psi|\psi^0) &= \sum_{i=1}^n \sum_{k=1}^g \hat{t}_{ik}^{(0)} [\log \pi_k + \log f_k(y_i|\theta_k)] \\ &= \sum_{i=1}^n \sum_{k=1}^g \hat{t}_{ik}^{(0)} \log \pi_k + \sum_{i=1}^n \sum_{k=1}^g \hat{t}_{ik}^{(0)} \log f_k(y_i|\theta_k).\end{aligned}$$

Después del cálculo anterior se realiza la maximización de la función \mathcal{E} respecto de ψ . Esta maximización se realiza en dos partes dado que π_k aparece únicamente en el primer sumando y θ_k lo hace el segundo sumando.

Comenzamos con la maximización del primer sumando. Para este caso usamos los multiplicadores de Lagrange

$$\frac{\partial}{\partial \pi_k} \left(\sum_{i=1}^n \sum_{k=1}^g \hat{t}_{ik}^{(0)} \log \pi_k + \lambda \left[\sum_{k=1}^g \pi_k - 1 \right] \right) = 0$$

$$\sum_{i=1}^n \hat{t}_{ik}^{(0)} \frac{1}{\pi_k} + \lambda = 0$$

$$\sum_{i=1}^n \hat{t}_{ik}^{(0)} = -\lambda \pi_k$$

Tomando la suma sobre k en ambos lados de la última igualdad obtenemos,

$$n = \sum_{i=1}^n \sum_{k=1}^g \hat{t}_{ik}^{(0)} = \sum_{k=1}^g -\lambda \pi_k = -\lambda$$

Lo cual implica que

$$\hat{\pi}_k^{(1)} = \pi_k = \frac{1}{n} \sum_{i=1}^n \hat{t}_{ik}^{(0)}$$

Para la maximización del segundo sumando respecto de θ_k depende de la función de densidad $f_k(y_i|\theta_k)$ que en nuestro caso son densidades Gaussianas, entonces

$$\begin{aligned} \log f_k(y_i|\theta_k) &= \log \varphi(y_i|\mu_k, \sigma_k^2) = \\ &= -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2} \frac{(y_i - \mu_k)^2}{\sigma_k^2} \\ &= -\frac{1}{2} \log(2\pi) - \log \sigma^2 - \frac{1}{2} \frac{(y_i - \mu_k)^2}{\sigma_k^2} \end{aligned}$$

Empezamos derivando respecto de μ

$$\frac{\partial}{\partial \mu_k} \sum_{i=1}^n \sum_{k=1}^g \hat{t}_{ik}^{(0)} \left(-\frac{1}{2} \log(2\pi) - \log \sigma_k - \frac{1}{2} \frac{(y_i - \mu_k)^2}{\sigma_k^2} \right) = 0$$

$$2 \sum_{i=1}^n \hat{t}_{ik}^{(0)} \left(\frac{y_i - \mu_k}{2\sigma_k^2} \right) = 0$$

$$\sum_{i=1}^n \hat{t}_{ik}^{(0)} y_i = \sum_{i=1}^n \hat{t}_{ik}^{(0)} \mu_k$$

$$\hat{\mu}_k^{(1)} = \frac{\sum_{i=1}^n \hat{t}_{ik}^{(0)} y_i}{\sum_{i=1}^n \hat{t}_{ik}^{(0)}}$$

Para obtener el estimador de σ_k^2 tenemos,

$$\begin{aligned} \frac{\partial}{\partial \sigma_k^2} \sum_{i=1}^n \sum_{k=1}^g \hat{t}_{ik}^{(0)} \left(-\frac{1}{2} \log(2\pi) - \frac{\log \sigma_k^2}{2} - \frac{1}{2} \frac{(y_i - \mu_k)^2}{\sigma_k^2} \right) &= 0 \\ -\sum_{i=1}^n \hat{t}_{ik}^{(0)} \frac{1}{2\sigma_k^2} + \sum_{i=1}^n \hat{t}_{ik}^{(0)} \frac{(y_i - \mu_k)^2}{2(\sigma_k^2)^2} &= 0 \\ \sum_{i=1}^n \hat{t}_{ik}^{(0)} \frac{(y_i - \mu_k)^2}{\sigma_k^2} &= \sum_{i=1}^n \hat{t}_{ik}^{(0)} \\ \sigma_k^2 &= \frac{\sum_{i=1}^n \hat{t}_{ik}^{(0)} (y_i - \mu_k)^2}{\sum_{i=1}^n \hat{t}_{ik}^{(0)}} \end{aligned}$$

Usando la estimación de μ_k , obtenemos una estimación de σ_k ,

$$\hat{\sigma}_k^{(1)} = \sqrt{\frac{\sum_{i=1}^n \hat{t}_{ik}^{(0)} (y_i - \hat{\mu}_k^{(1)})^2}{\sum_{i=1}^n \hat{t}_{ik}^{(0)}}}$$

Valores iniciales

Los valores iniciales sobre los que el algoritmo comienza a iterar, y que son implementados en muchas situaciones, se toman dividiendo la muestra en g particiones y sobre cada una de ellas se calcula la media de las observaciones. Estos valores se representan por $\hat{\mu}_1^{(0)}, \hat{\mu}_2^{(0)}, \dots, \hat{\mu}_g^{(0)}$. Respecto a los pesos todos se toman semejantes según $\pi_1^{(0)} = \pi_2^{(0)} = \dots = \pi_g^{(0)} = 1/g$. Existen otras formas de tomar los valores iniciales (Finch *et al.*, 1989).

Criterio de parado

Para detener las iteraciones vamos a considerar la diferencia relativa

$$\frac{|l(\boldsymbol{\psi}^{(t+1)}|y) - l(\boldsymbol{\psi}^{(t)}|y)|}{|l(\boldsymbol{\psi}^{(t)}|y)|}$$

Que se utiliza más por su adimensionalidad. Y se detiene el proceso cuando el valor máximo de dicha diferencia es menor que 10^{-6} .

Implementación del algoritmo

Si el algoritmo es capaz de discriminar por medio de la variable *perimeter*, si el tumor es maligno o benigno, debemos considerar una mezcla de normales con dos componentes, por lo que dividimos nuestra base de datos en dos grupos iguales previamente ordenados de menor a mayor, como el número de registros es de $n = 569$, consideramos dos grupos, uno con los primeros $n_1 = 285$ registros y el segundo con los $n_2 = 284$ registros restantes. Los valores iniciales que se determinan en base a los valores de la variable *perimeter* son;

$$\pi_1^{(0)} = \pi_2^{(0)} = 1/2$$

$$\hat{\mu}_1^{(0)} = \frac{1}{n_1} \sum_{s=1}^{n_1} y_s = 73.60 \quad \hat{\sigma}_1^{(0)} = \left(\frac{1}{n_1 - 1} \sum_{s=1}^{n_1} (y_s - \hat{\mu}_1^{(0)})^2 \right)^{1/2} = 8.83$$

$$\hat{\mu}_2^{(0)} = \frac{1}{n_2} \sum_{s=n_1+1}^n y_s = 110.27 \quad \hat{\sigma}_2^{(0)} = \left(\frac{1}{n_2 - 1} \sum_{s=n_1+1}^n (y_s - \hat{\mu}_2^{(0)})^2 \right)^{1/2} = 20.73$$

El algoritmo empieza a iterar y se detiene exactamente en la iteración donde se cumple,

$$\frac{|l(\boldsymbol{\psi}^{(t+1)}|y) - l(\boldsymbol{\psi}^{(t)}|y)|}{|l(\boldsymbol{\psi}^{(t)}|y)|} < 10^{-6}.$$

Obteniendo los siguientes valores para los estimadores de la Tabla 2

Tabla 2

Estimadores

θ_k	Valor de $\hat{\theta}_k$
π_1	0.63
π_2	0.37
μ_1	79.25
μ_2	113.82
σ_1	12.14
σ_2	24.35

Fuente: Elaboración propia

Con estos valores tenemos la forma de la mezcla de normales y con los valores de la última iteración, podemos calcular las componentes, algunos datos se muestran en la Tabla 3.

Tabla 3

Muestra de la forma de la mezcla de normales

ID	Diagnostico Benigno (B), Maligno (M)	perimeter	Mezcla de Normales (1) Benigno (2) Maligno
859464	B	60.11	1
859465	B	71.8	1
859471	B	58.79	1
859487	B	81.37	1
859575	M	123.6	2
859711	B	58.79	1
859717	M	114.2	2
859983	M	90.43	1
8610175	B	79.19	1
8610404	M	104.1	2
8610629	B	87.91	1
8610637	M	120.2	2
8610862	M	143.7	2
8610908	B	83.19	1
861103	B	73.81	1
8611161	B	86.49	1
8611555	M	171.5	2
8611792	M	129.1	2
8612080	B	76.95	1
8612399	M	121.1	2
86135501	M	94.25	1
86135502	M	122	2
861597	B	79.78	1
861598	B	95.77	1

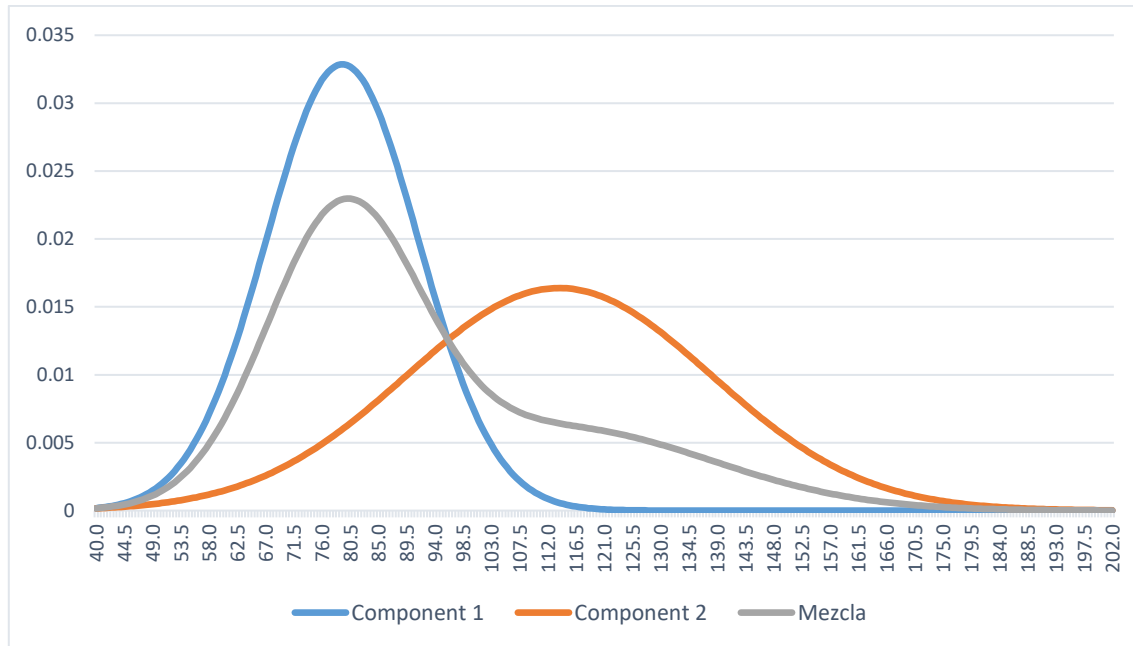
Fuente: Elaboración propia con datos de Wolberg *et al.* (1995).

Podemos observar que existen diferencias, por ejemplo, en los registros 8 y 21 la mezcla de normales los clasifica en la componente uno (benignos), sin embargo, en el diagnostico estos registros están clasificados en la componente de malignos.

En la Figura 2 se muestra la forma de la mezcla de normales y sus dos componentes; componente 1 (benignos) y componente 2 (malignos).

Figura 2

Forma de la mezcla de normales



Fuente: Elaboración propia.

Conclusiones

Podemos observar que, de un total de 569 observaciones, 357 que representan el 62.7% del total, indican la ausencia de células cancerosas, mientras que 212 que representan el 37.3% del total muestran la presencia de células cancerosas. Con el modelo de mezcla de normales, se encontró que, del total de observaciones, 403 que representan el 71%, indican la ausencia de células cancerosas, mientras que 166 que representan 29% del total, muestran la presencia de células cancerosas.

En los datos encontrados mediante la mezcla de normales, observamos que en el caso de tumores benignos existe una coincidencia de 348 casos, que representan el 97.5% del total de casos benignos, con una diferencia de 9 casos.

Mientras que el caso de los malignos existe una coincidencia de 157 casos que representa el 74.1% del total de los casos malignos, con una diferencia de 55 caso.

Podemos concluir que el modelo en el caso de tumores benignos tiene una predicción con buena aceptación, mientras que en el caso de tumores malignos la predicción no sería muy buena. En este último caso debemos tomar en cuenta que en el estudio del Dr. Wolberg nos indica que “El porcentaje es inusualmente grande; el conjunto de datos no representa en este caso una distribución típica de análisis médico. Por lo general, tendremos una cantidad considerable de casos que representan negativos versus una pequeña cantidad de casos que representan tumores positivos (malignos)”. Esta afirmación puede corroborar la diferencia que se encontró en los casos malignos.

Referencias

- Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1), 1-38.
- Denais, C., & Lammerding, J. (2014). Nuclear mechanics in cancer. *Advances in experimental medicine and biology*, 773, 435–470. https://doi.org/10.1007/978-1-4899-8032-8_20
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, 144(5), 646-674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Levine, R. & Casella, G., (2001). Implementation of the monte carlo EM algorithm. *Journal of Computational and Graphical Statistics*, 10(3), 422-439. <https://www.jstor.org/stable/1391097>
- McCulloch, C. E. (1998). [Review of The EM Algorithm and Its Extensions, by G. J. MacLachlan & T. Krishnan]. *Journal of the American Statistical Association*, 93(441), 403–404. <https://doi.org/10.2307/2669645>
- McLachlan, G. & Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons, Inc. <https://doi.org/10.1002/0471721182>
- Mengerser, K., Robert, C. & Titterton, D. (2011). *Mixtures: Estimation and Applications*. John Wiley & Sons, Inc.

Wolberg, W., Mangasarian, O., Street, N., & Street, W. (1995). *Breast Cancer Wisconsin (Diagnostic)*. UCI Machine Learning Repository. <https://doi.org/10.24432/C5DW2B>

Schlattmann, P. (2009). *Medical Applications of Finite Mixture Models*. Springer.