

UNIVERSIDAD DE BARCELONA

FACULTAD DE MEDICINA

DEPARTAMENTO DE ANATOMÍA Y EMBRIOLOGÍA HUMANA

**Estudio de los cambios morfológicos del epitelio
corneal en un modelo animal de ojo seco**

GEMMA JULIO MORÁN

2. MATERIAL Y MÉTODO

2.1. Obtención y preparación de las muestras

Para realizar este trabajo de investigación se han empleado 33 córneas de conejo albino de Nueva Zelanda. El conejo posee un epitelio corneal muy parecido al humano: el mismo número de capas, tamaños celulares similares (Doughty 1990b) y microproyecciones y uniones intercelulares morfológicamente muy semejantes (Rigal 1993). Además, la inervación de la córnea del conejo no posee diferencias significativas respecto de la que presenta el epitelio corneal humano. (Ojeda et al 2001)

Los animales utilizados en esta tesis doctoral, 8 hembras y nueve machos de edad inferior a un año, pesaban entre 1,7 y 2,5 Kg y fueron suministrados por el Estabulario de la Facultad de Medicina de la Universidad de Barcelona. En todo momento se ha seguido las normas de estabulación y manipulación de animales de experimentación dictadas por la *Generalitat de Catalunya* y las normas ARVO (Resolution on the Use of Animals in Research).

Para limitar los movimientos de la cabeza, cada animal fue colocado en un inmovilizador en el que permaneció, bajo anestesia general, hasta su muerte. La pauta de anestesia empleada consiste en administrar maleato de Acepromacina (Calmoneosan®) (0,5 ml/10kg) por vía intramuscular y, tres o cuatro minutos después, Atropina (Atropina Braun®) (0,05 mg/Kg) y Ketamina (Imalgène®) (20 mg/Kg), también por vía intramuscular. La dosis de mantenimiento corresponde a la mitad de la dosis inicial. Este es uno de los protocolos de anestesia del estabulario de la Facultad de Medicina de la Universidad de Barcelona.

Una vez anestesiado el animal, se le colocó un blefarostato en cada ojo para impedir el parpadeo. Después, se realizó una evaluación con biomicroscopio con el fin de asegurar que las córneas eran, inicialmente, sanas y normales. Los blefarostatos se mantuvieron un tiempo determinado, el mismo para los dos ojos, que se ha definido como: tiempo de apertura palpebral (TAP). Por lo tanto, el TAP representa el periodo comprendido desde la colocación del

blefarostato hasta la muerte del animal. En la tabla 2.1 aparecen las córneas agrupadas según el TAP al que han sido sometidas.

TAP	Córneas
Control	C0i; C01; C02; C03; C04
TAP 1 (< de 1h)	C12 (12'), C18(40'); C22(50')
TAP 2 (1 a 2h)	C11(1h 25'); C13(1h 45'); C17 (1h 7')
TAP 3 (2 a 3h)	C14(2h 12'); C15(2h 10'); C16(2h)
TAP 4 (≥ 3h)	C19(3h 7'); C20(3h 15'); C21(3h)

Tabla 2.1. Tiempos de apertura palpebral de cada una de las córneas. Entre paréntesis aparece el TAP exacto y cada dato corresponde tanto a la córnea del ojo derecho como a la del ojo izquierdo de un animal, excepto la córnea C0i que corresponde al ojo izquierdo del conejo número 0. (El ojo derecho fue desestimado)

Para practicar la eutanasia se empleó T-61[®] (Embutramida 200mg, Mebezonio yoduro 50 mg, tetracaína clorhidrato 5mg csp 1ml) y, una vez muerto el animal, se irrigaron los ojos con solución fijadora de Karnovsky 0,1 M y se procedió a su enucleación, de manera que el tiempo post-mortem de todas las córneas es nulo. La composición del mencionado fijador es la siguiente (por 100 ml):

Tampón Fosfato 0,2 M a pH 7,39*.....50 ml
 Paraformaldehido al 8%..... 25 ml
 Glutaraldehido al 25%..... 10 ml
 Agua destilada.....15 ml

*resultante de una disolución acuosa de fosfato monosódico 3,21% y fosfato disódico 3,56% en proporción 2:8

Las maniobras para la enucleación podemos resumirlas como sigue: con un pequeño bisturí se procede a seccionar el saco conjuntival inferior y superior. A continuación mediante una erina de Cushing (de punta roma) se tracciona el tendón de inserción del músculo oblicuo mayor y, una vez exteriorizado, se

secciona con unas tijeras curvas de Mayo. Posteriormente, sujetando el tendón seccionado, se procede a cortar los tendones de los músculos restantes. Finalmente, mediante una tracción y rotación del globo ocular se visualiza el polo posterior y se procede a seccionar el nervio óptico y los vasos sanguíneos que pasan por la zona. De este modo, con una ligera tracción queda enucleado el globo ocular.

Tras esta primera fase de obtención de las muestras se inicia la fase de fijación. Para ello se sumerge, durante un cuarto de hora, el globo ocular en solución fijadora de Karnovsky 0,1M. Así, se consigue que dicha solución se absorba y penetre en la cámara anterior. Una vez todas las capas de la córnea han entrado en contacto con el fijador se puede proceder a seccionar el perímetro con una tijera quirúrgica de Castroviejo (dejando un anillo de esclera de, aproximadamente, 2mm). Acto seguido, se separa el iris, el cuerpo ciliar y el cristalino y se sumerge la córnea en un vial con solución de Karnovsky durante cinco días, conservándola en el refrigerador. Siguiendo este procedimiento, la superficie corneal se mantiene lisa, sin arrugas que puedan afectar al tejido.

El paso siguiente es la deshidratación de la muestra. Para ello, se lava con agua destilada y se sumerge, durante periodos de 15 minutos, en soluciones de alcohol progresivamente más concentradas (50°, 60°, 70° 80° 90° y finalmente alcohol absoluto).

Cuando la córnea se halla completamente deshidratada y sumergida en alcohol absoluto, se inicia la desecación, sustituyendo progresivamente el alcohol del medio por hexametildisilazano. Para ello, la muestra se sumergen, seriadamente, en soluciones anhidras de mezcla alcohol absoluto-hexametildisilazano en las siguientes proporciones: 3:1, 2:2, 1:3, de manera que permanece en cada solución por un período de tiempo de 15 minutos.

Acto seguido, la córnea se sumerge en un recipiente abierto con hexametildisilazano puro, hasta que este se evapora totalmente. Así, la

deseccación se realiza reduciendo al mínimo los efectos distorsionantes originados por la tensión superficial. (Braet et al 1997) (Araujo et al 2003)

Finalmente, la córnea completamente deshidratada y desecada se monta sobre stubs de acero, utilizando plata coloidal para su sujeción. Además de pegar la córnea a la superficie del stub, la plata asegura una perfecta conductividad eléctrica. Todas las córneas se han montado con la cara epitelial hacia arriba para que esta pueda ser explorada a través del microscopio.

Una vez ha finalizado el montaje, y sólo cuando la solución de plata está seca, se procede al recubrimiento de la superficie corneal con una finísima capa de oro mediante el procedimiento de sputtering, asegurando así la conductividad eléctrica de toda la superficie de la muestra. De este modo, la córnea queda preparada para su observación.

2.2. Método de observación y morfometría

La observación del epitelio corneal se ha realizado con un microscopio electrónico de barrido (SEM) modelo Hitachi 2300, del servicio de microscopia de la Universidad de Barcelona, que lleva incorporado un sistema de registro de imágenes. La muestra se ha colocado perpendicular al haz de electrones con una distancia de trabajo de 8 a 15 mm (la más frecuente ha sido de 10 mm) y se ha empleado un voltaje de aceleración de 15 Kw.

Con dicho microscopio se han obtenido imágenes digitalizadas de la corneal central que se han grabado en formato "tif" para su posterior análisis. Este formato no comprime la imagen y, por lo tanto, no produce ninguna modificación en ella. A cada una de estas imágenes se le ha asignado un código en el que queda representada la córnea de la que procede. Por ejemplo, la imagen C01i12 es la imagen número 12 de la córnea izquierda del conejo C01.

La selección de las zonas a fotografiar se realizó al azar dentro de un radio de 2 mm desde el ápice corneal. Cada zona estudiada viene delimitada por el campo de visión a 500x y representa una superficie aproximada de 0,052 mm². Es importante puntualizar que, las zonas escogidas no presentan lesiones mecánicas, producto de la manipulación de las muestras.

Por término medio, se han estudiado cinco zonas por córnea. Dicho estudio se ha basado en una descripción morfológica de visu y un análisis estadístico de los datos morfométricos de una media de seis células en cada zona. En la tabla 2.2 aparecen las córneas con el número exacto de células analizadas en cada caso.

TAP	CÓRNEAS									Total Células
	Nº de células/córnea									
TAP 0	C0i	C01d	C01i	C02d	C02i	C03d	C03i	C04d	C04i	233
	13	31	34	28	22	25	30	25	25	
TAP 1	C12d	C12i	C18d	C18i	C22d	C22i				168
	31	29	28	30	27	23				
TAP 2	C11d	C11i	C13d	C13i	C17d	C17i				191
	29	29	19	42	33	39				
TAP 3	C14d	C14i	C15d	C15i	C16d	C16i				205
	46	35	42	23	30	29				
TAP 4	C19d	C19i	C20d	C20i	C21d	C21i				191
	32	31	32	29	38	29				

Tabla 2.2. Número de células por córnea y TAP.

Por tanto, analizando las imágenes de microscopía electrónica, se ha llevado a cabo una observación minuciosa del estrato escamoso, en aproximadamente un 3% del área central (anteriormente delimitada) de todas las córneas. Gracias a esta observación se han descrito las características morfológicas de las células del epitelio corneal tanto desde el punto de vista cualitativo como cuantitativo.

Para efectuar un estudio riguroso y ordenado, la descripción del epitelio se ha realizado siguiendo un procedimiento sistematizado. Con dicho procedimiento se pretende uniformizar el criterio descriptivo evitando, en la medida de lo posible, la subjetividad del observador.

De este modo, se han agrupado las imágenes corneales atendiendo al tiempo de apertura palpebral (TAP), describiendo primero las características generales de la superficie epitelial de cada grupo. Los aspectos morfológicos estudiados han sido: continuidad del epitelio, presencia de células exfoliadas, contraste y textura del mosaico celular, contornos celulares, forma y disposición de las microproyecciones, aspecto del núcleo y de la membrana celular.

En segundo lugar, se han escogido las células cuyas características morfológicas serán cuantificadas. El criterio de selección es sencillo. Entran a formar parte del estudio morfométrico aquellas células que aparecen en una imagen, a 5000x, del centro de cada una de las zonas, previamente seleccionadas al azar. Cabe señalar que, las imágenes a 5000x son necesarias pues permiten obtener ciertos datos que no se pueden medir a 500x (como se verá más adelante).

A cada célula seleccionada se le ha asignado un código identificativo. Por ejemplo, la célula C04i1_2 es una célula de la córnea izquierda del conejo C04. Seguidamente, aparece el número asignado a la célula (en este caso el 1) y, finalmente, el número de la imagen donde aparece dicha célula. De este modo, el código no sólo identifica a la célula sino que sirve, también, para localizarla en el archivo de imágenes.

Para efectuar el análisis cuantitativo, se ha definido una serie de variables que representan los rasgos morfológicos celulares a estudiar. Dichas variables se han clasificado como principales o subordinadas.

Las variables principales son: *superficie ocupada por microproyecciones (SOM)*, *área*, *forma*, *tono*, *holes* y *uniones intercelulares*. (A partir de ahora, aparecerán en cursiva al ser nombradas). Estas variables pueden ser objetivas

o subjetivas. Las objetivas (*SOM, área, forma, tono*) se han calculado mediante el programa de procesamiento de imágenes ImageJ (Rasband, WS., ImageJ, U.S. National Institutes of Health, Bethesda, Maryland. USA, <http://rsb.info.nih.gov/ij/>, 1997-2005) versión 1.33u y pueden adoptar cualquier valor dentro del rango de las dimensiones celulares.

Por su parte, las variables principales subjetivas (*holes y uniones intercelulares*) se han cuantificado en base a las observaciones de un estudio previo del epitelio corneal realizado visualmente, de ahí su carácter subjetivo. Los valores de estas variables son datos numéricos que representan categorías.

Por otro lado, las variables subordinadas se obtienen a partir de la agrupación por categorías de los datos obtenidos en las variables principales de carácter objetivo (ver tabla 2.3). En consecuencia, este grupo de variables sólo puede adoptar ciertos valores concretos, que se corresponden con los números asignados previamente a cada categoría.

Variables principales	Definición	Variables subordinadas	Número de Categorías
Superficie ocupada por las microproyecciones (SOM)	Tanto por ciento de la superficie celular ocupada por microproyecciones	Cat SOM	5
Área	Área celular en μm^2	Cat tamaño	3
Forma	Circularidad de la célula	Cat forma	3
Tono	Media del tono de gris del área celular a 500x	Cat tono	3

Tabla 2.3. Variables principales de carácter objetivo y sus correspondientes variables subordinadas.

A continuación, se describe el proceso de medición empleado para obtener las variables principales y sus correspondientes subordinadas.

Superficie ocupada por microproyecciones

Esta variable representa la relación del área celular con microproyecciones respecto del área celular analizada, expresada en tanto por ciento.

Para realizar el cálculo se binariza la imagen celular a 5000x en base al histograma de tonos de gris de una zona de la célula, previamente seleccionada (comandos Image/ Adjust/ Threshold/ Auto del programa ImageJ 1.33u) (ver figura 2.1).

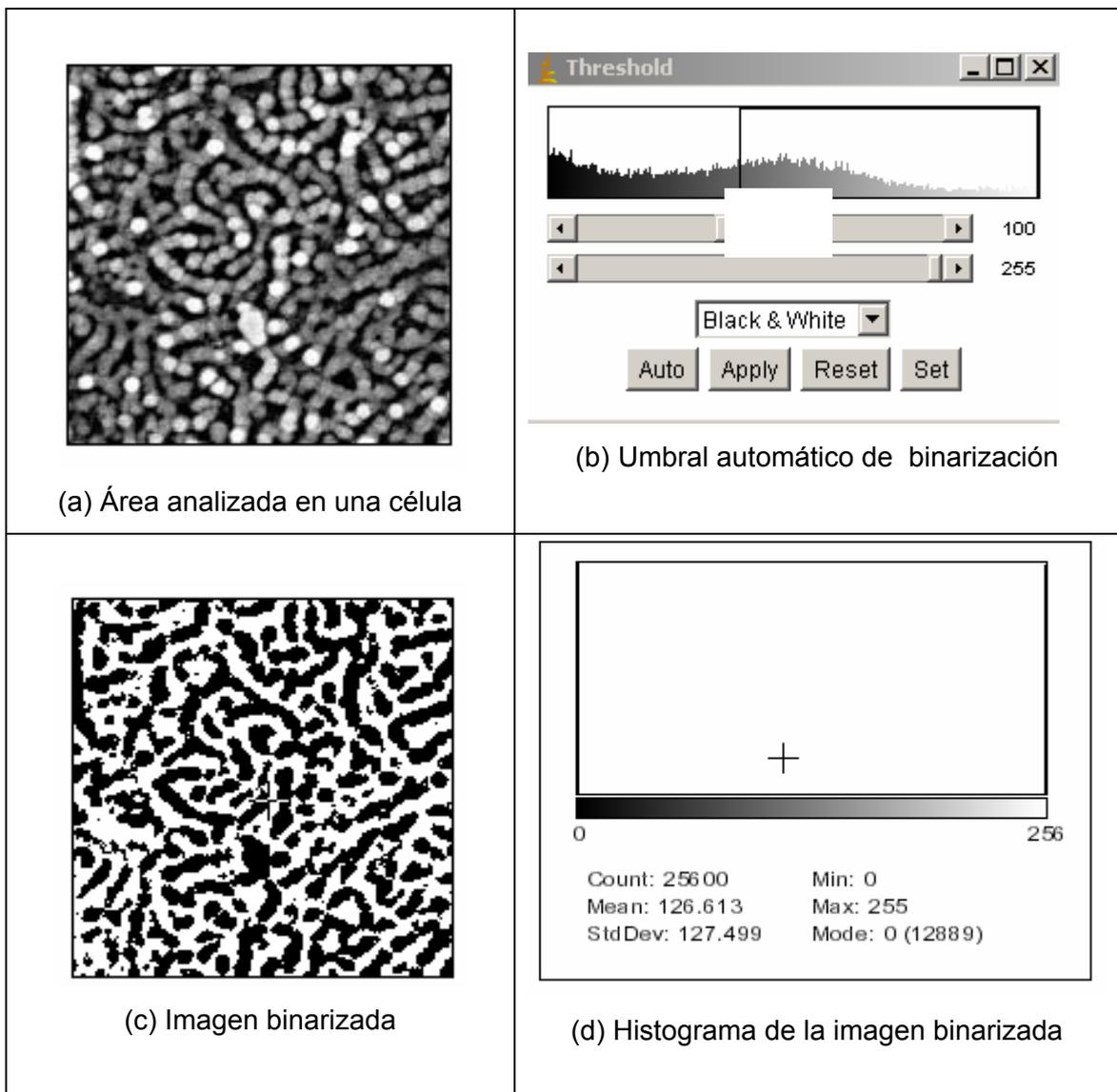


Figura 2.1. Binarización de las imágenes. El umbral de binarización (b) es un valor propio de cada imagen celular y establece, el límite entre los tonos de gris que pasarán a ser negros y los que serán blancos en la imagen binarizada.

De este modo, se consigue una imagen equivalente a la original pero sólo con dos tonos de gris: el blanco (nivel 255) y el negro (nivel 0). (La imagen debe ser a 5000x porque estos aumentos discriminan perfectamente el contorno de la microproyección)

La binarización, que realiza el programa de forma automática, divide la imagen en objetos y fondo. Mediante un proceso iterativo calcula el promedio de píxeles cuyo valor está por encima de un umbral de binarización, escogido inicialmente, y el promedio de los que están por debajo. En función de estos promedios reajusta el umbral de binarización, repitiendo el proceso iterativamente, hasta conseguir que este sea mayor que el valor medio de los promedios obtenidos. Esto se expresa con la siguiente ecuación:

$$Umbral \geq \frac{(promedio\ fondo + promedio\ objetos)}{2}$$

Seguidamente, y siguiendo con el procedimiento de medida de la variable, una vez binarizada la imagen se obtiene el histograma de un área celular seleccionada (ver, de nuevo, figura 2.1). En los datos del histograma aparece el nivel de gris más frecuente (la moda) y el número de píxeles con ese nivel. Comparando la imagen original con la binarizada se puede establecer si estos píxeles representan a las microproyecciones o a la membrana.

En el ejemplo de la figura 2.1 si se comparan las dos imágenes (a) y (c) se puede concluir que en la imagen binarizada (c), los píxeles en negro corresponden a las microproyecciones y los píxeles blancos a la membrana. Como la moda del histograma (d) coincide con el color negro (nivel cero) el número de píxeles negros (c) son los píxeles ocupados por las microproyecciones.

En este caso, con el número de píxeles de la moda y el número total de píxeles incluidos en el área seleccionada (en el histograma (d) este valor aparece

como *Count*) se calcula, finalmente, el tanto por ciento correspondiente que será el valor de la superficie ocupada por microproyecciones.

La equivalencia entre la moda (tono de gris más frecuente) y las microproyecciones aparece siempre que la densidad de estas prominencias sea elevada. Por el contrario, si una célula posee un escaso número de microproyecciones la moda del histograma binarizado representará a la membrana celular. En este último caso, para obtener el número de píxeles que corresponden a las microproyecciones se deben restar los píxeles de la moda al total de píxeles analizados (ver figura 2.2).

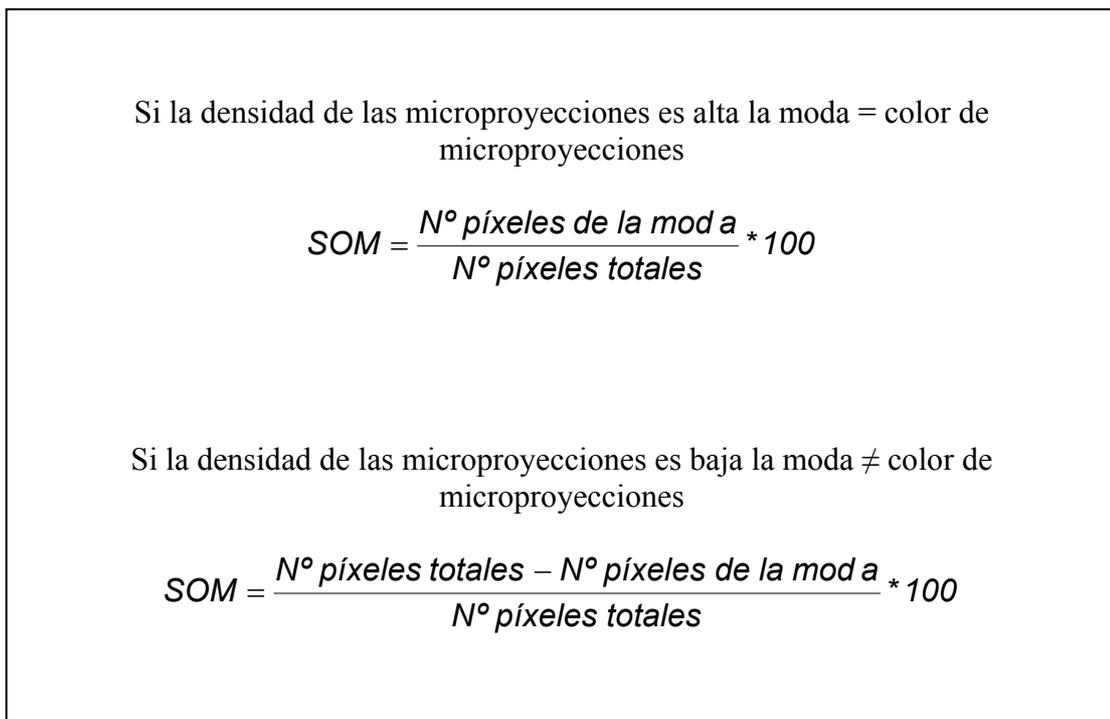


Figura 2.2. Cálculo de la superficie ocupada por las microproyecciones (SOM) en base a los datos del histograma binarizado.

Es importante señalar que la binarización es un método válido partiendo de la premisa de que en la imagen original sólo existen dos grupos de tonos de gris, véase, los que corresponden a las microproyecciones y los que corresponden a la membrana. El tono de las microproyecciones siempre será más claro que el de la membrana ya que están por encima de ella. (Las imágenes de

microscopía electrónica de barrido presentan diferentes tonos de gris atendiendo a la topografía de la muestra)

De lo comentado hasta ahora se puede deducir que la aparición de restos de células, artefactos, grietas o arrugas en la superficie celular podría alterar los resultados. Por este motivo, la superficie ocupada por microproyecciones se ha medido sobre un área celular seleccionada y no sobre toda el área celular, escogiéndose una zona representativa y lo más limpia posible.

Cabe resaltar que, como la binarización se ha realizado célula a célula, el umbral escogido está influido por el nivel de brillo y contraste de cada imagen. Por lo tanto las imágenes binarizadas compensan de forma eficaz el efecto distorsionante de la diferencia de brillo y contraste entre las imágenes originales.

Finalmente, una vez calculados, los valores de la variable principal *SOM* se han categorizado, para obtener los valores de la variable subordinada *cat SOM*, en base a un estudio previo realizado con todas las células. En dicho estudio se compararon los datos de la variable *SOM* con la observación visual de las imágenes celulares a 5000x, buscando cierta concordancia entre los valores de la variable principal y el aspecto de las microproyecciones de visu. En la tabla 2.4 aparecen las características morfológicas de las categorías establecidas visualmente. De este modo, la categorización de la variable *cat SOM* es la siguiente:

- a) Categoría 1: *superficie ocupada por microproyecciones (SOM) \geq 50%*. Dicha categoría representa la máxima densidad de microproyecciones por lo que el límite correspondiente se establece en aquel valor a partir del cual la superficie ocupada por microproyecciones es mayor que la superficie vacía.
- b) Categoría 0: *superficie ocupada por microproyecciones (SOM) $<$ 10%*. La categoría cero no puede incluir células con microproyecciones y debe ser tolerante con las irregularidades de la membrana celular. Por lo

tanto, para esta categoría el límite superior es algo mayor que 0% pues no existe ninguna célula cuya membrana sea perfectamente uniforme.

c) Categoría 0,8: *superficie ocupada por microproyecciones (SOM) ≥ 40%*.

d) Categoría 0,5: *superficie ocupada por microproyecciones (SOM) ≥ 30%*

e) Categoría 0,2: *superficie ocupada por microproyecciones (SOM) ≥ 10%*

Microproyecciones	
Categoría	Descripción
1	Densidad de estructuras vermiformes muy elevada. Los espacios libres de prominencias son iguales o más pequeños que los ocupados por estas.
0.8	Densidad de estructuras vermiformes elevada. Los espacios libres de prominencias son mayores que los ocupados por estas.
0.5	Densidad de microproyecciones baja. Estructuras de formas irregulares. Los espacios libres de prominencias son mucho más grandes que los ocupados por estas.
0.2	Densidad de microproyecciones muy baja. Las prominencias aparecen como estructuras puntiformes en medio de grandes zonas vacías.
0.0	Sin microproyecciones.

Tabla 2.4. Clasificación de las microproyecciones en categorías de visu.

Las tres categorías intermedias se han establecido dividiendo en tres grupos el intervalo de valores restante. Sin embargo, la franja de valores que incluye la categoría 0,2 es mayor que la del resto de categorías intermedias. Ello es debido a que la densidad de microproyecciones en este caso es muy baja y un pequeño aumento en la variable *SOM* no modifica la clasificación visual. De forma recíproca, si la densidad no es tan baja (categoría 0,5) las diferencias en el valor de *SOM* son cualitativamente más significativas y determinan, en consecuencia, un cambio de categoría. En la figura 2.3 aparece un ejemplo de cada una de las categorías.

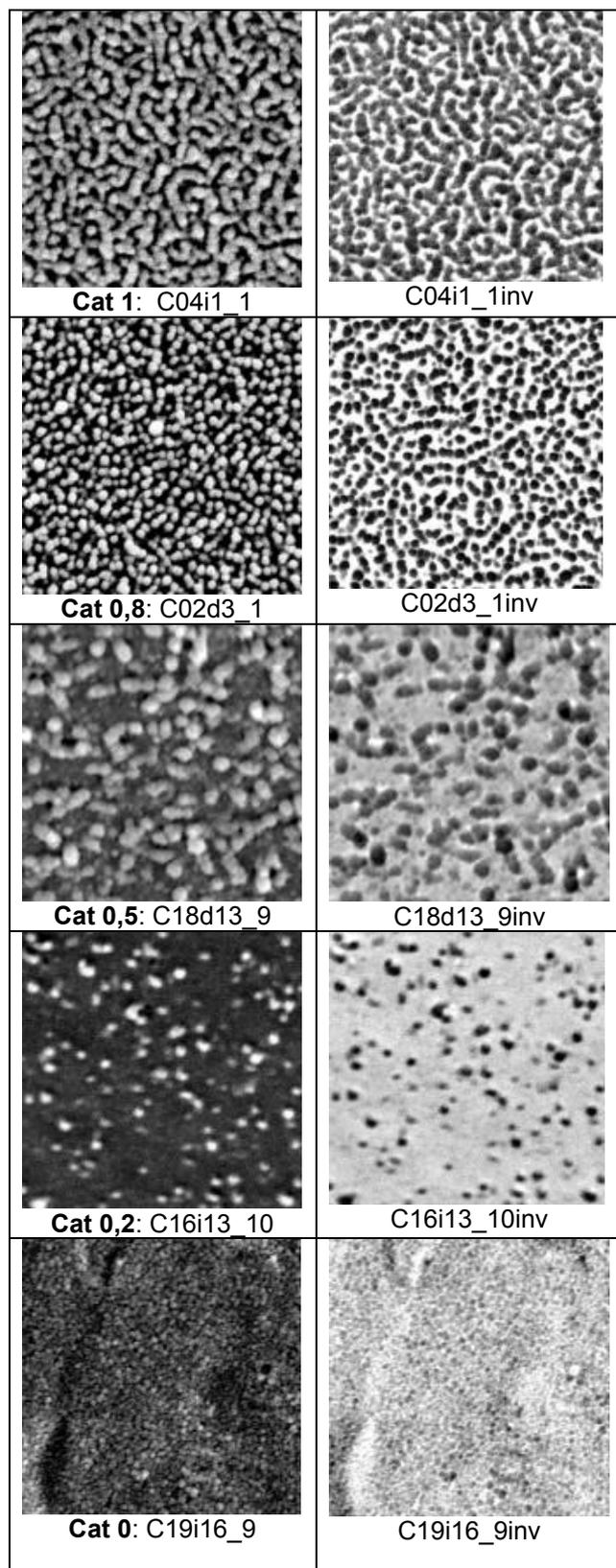


Figura 2.3. Ejemplos de categorías de la variable subordinada *catSOM*. Las imágenes invertidas (columna de la derecha) son equivalentes al negativo de una la fotografía y ayudan a asignar la categoría de visu. El código que aparece debajo de la imagen es el código de la célula. Todas las imágenes de esta figura son a 5000x.

Área

Esta variable, como su nombre indica, se obtiene en base a una medida del área de las células empleando, de nuevo, el programa de procesado de imágenes *ImageJ*. Para establecer dicho valor se marca el contorno celular en una imagen a 500x (adaptando dicho marcaje a las diferentes morfologías celulares) y, aplicando los comandos *Analyze / Measure*, el programa calcula directamente el valor del área delimitada.

Cabe comentar que, *ImageJ* presenta el área celular en píxeles² que deben convertirse, más tarde, en unidades de superficie (μm^2) mediante un factor de conversión. Dicho factor se ha obtenido estableciendo la superficie real que cubre un píxel, gracias a la escala que aparece en todas las imágenes.

Una vez obtenidos los datos de la variable principal, *área*, se ha establecido las tres categorías de su variable subordinada *cat tamaño* (tamaño pequeño, mediano y grande) mediante un estudio preliminar. En este nuevo estudio se seleccionaron nueve imágenes a 500x, una por cada una de las córneas control. De cada una de las imágenes se escogieron las cinco células más grandes y las cinco más pequeñas, según un criterio de visu. (En total se seleccionaron 45 células grandes y el mismo número de células pequeñas de un total de 991 células visualizadas)

Seguidamente, se calculó el área de las 90 células escogidas. De modo que, el área más pequeña de todas las células grandes se ha considerado el límite inferior de la categoría de tamaño grande. De manera recíproca, se escogió el área más grande de entre las células más pequeñas como límite superior, por debajo del cual se considera que una célula es pequeña. El valor de los límites se redondeó a la baja en las células grandes y a la alza en las pequeñas. Los límites de la categoría de tamaño mediano quedan automáticamente definidos entre los de la categoría grande y pequeña. (Ver tabla 2.5 y 2.6)

Imagen	nº de células por imagen	nº de la célula seleccionada	Área en píxeles ²	Área en μm^2
c0i1a	126	1	14189	869.164931
c0i1a		2	10801	661.628756
c0i1a		3	10450	640.127813
c0i1a		4	10030	614.400188
c0i1a		5	11718	717.800738
c01d13	128	6	15454	946.654088
c01d13		7	12964	794.126025
c01d13		8	13887	850.665544
c01d13		9	13619	834.248869
c01d13		10	12859	787.694119
c01i12	114	11	13131	804.355819
c01i12		12	13226	810.175163
c01i12		13	12206	747.693788
c01i12		14	14601	894.402506
c01i12		15	12919	791.369494
c02d8	95	16	18538	1135.56836
c02d8		17	22109	1354.31443
c02d8		18	19256	1179.55035
c02d8		19	14835	908.736469
c02d8		20	18208	1115.3538
c02i16	104	21	18668	1143.53168
c02i16		22	18785	1150.69866
c02i16		23	15316	938.200725
c02i16		24	17972	1100.89733
c02i16		25	18653	1142.61283
c03d08	98	26	18500	1133.24063
c03d08		27	18813	1152.41383
c03d08		28	17272	1058.01795
c03d08		29	18921	1159.02951
c03d08		30	20899	1280.19437
c03i24	120	31	18566	1137.28354
c03i24		32	17203	1053.79127
c03i24		33	20015	1226.04384
c03i24		34	12393	759.148706
c03i24		35	13347	817.587169
c04d21	106	36	14036	859.792725
c04d21		37	11939	731.338369
c04d21		38	12780	782.854875
c04d21		39	11582	709.469888
c04d21	100	40	10640	651.7665
c04i21		41	18162	1112.53601
c04i21		42	16082	985.123013
c04i21		43	16236	994.556475
c04i21		44	17906	1096.85441
c04i21		45	12128	742.9158

Tabla 2.5. Células grandes. El área más pequeña aparece en negrita.

Imagen	nº de células por imagen	nº de la célula escogida	Area en píxeles ²	Area en μm^2
c0i1a	126	1	237	14.5281
c0i1a		2	457	28.0141
c0i1a		3	987	60.5031
c0i1a		4	1304	79.9352
c0i1a		5	1092	66.9396
c01d13	128	6	1708	104.7004
c01d13		7	578	35.4314
c01d13		8	488	29.9144
c01d13		9	1365	83.6745
c01d13		10	2601	159.4413
c01i12	114	11	2106	129.0978
c01i12		12	2879	176.4827
c01i12		13	1649	101.0837
c01i12		14	1251	76.6863
c01i12		15	1177	72.1501
c02d8	95	16	2440	149.572
c02d8		17	2082	127.6266
c02d8		18	2187	134.0631
c02d8		19	1628	99.7964
c02d8		20	2465	151.1045
c02i16	104	21	605	37.0865
c02i16		22	1478	90.6014
c02i16		23	1196	73.3148
c02i16		24	1617	99.1221
c02i16		25	1904	116.7152
c03d08	98	26	1554	95.2602
c03d08		27	1638	100.4094
c03d08		28	401	24.5813
c03d08		29	1793	109.9109
c03d08		30	968	59.3384
c03i24	120	31	1557	95.4441
c03i24		32	456	27.9528
c03i24		33	2106	129.0978
c03i24		34	732	44.8716
c03i24		35	1825	111.8725
c04d21	106	36	2476	151.7788
c04d21		37	952	58.3576
c04d21		38	1539	94.3407
c04d21		39	1335	81.8355
c04d21		40	1467	89.9271
c04i21	100	41	693	42.4809
c04i21		42	1562	95.7506
c04i21		43	1846	113.1598
c04i21		44	370	22.681
c04i21		45	1153	70.6789

Tabla 2.6. Células pequeñas. El área más grande aparece en negrita.

De este modo, se considera que una célula es pequeña cuando su área es menor o igual a $200 \mu\text{m}^2$ y es grande cuando es superior o igual a $600 \mu\text{m}^2$. El resto de células, cuya área celular está comprendida entre 200 y $600 \mu\text{m}^2$, se han catalogado como células de tamaño intermedio.

Forma

Como en los casos anteriores, el parámetro forma celular se obtiene empleando el programa de procesamiento de imágenes *ImageJ*. Dicho programa permite preestablecer (*Analyze/Set measurements*) una serie de medidas a realizar sobre un contorno celular señalado y entre ellas está la circularidad.

En concreto, la medida de la forma se basa en calcular la relación que existe entre un contorno celular concreto y una circunferencia perfecta. Si la célula es circular el valor del parámetro será 1 o muy cercano a uno mientras que si la célula es un polígono irregular (como sucede con la mayoría de las células del epitelio corneal) (Doughty 1990c) se obtendrán valores cercanos a cero.

Para establecer los límites entre las tres categorías de la variable *cat forma* (poligonal, pseudopoligonal y circular) se realizó, como en los casos anteriores, un estudio preliminar. En concreto, se han seleccionado treinta células de formas variadas obtenidas de la imagen control designada como C0i1a. (El total de células que aparecen en esta imagen es de 126) (Ver figura 2.4)

Seguidamente, se ha medido la circularidad de cada uno de los contornos de las células escogidas aplicando los comandos *Analyze / Measure* del programa de procesamiento de imágenes Image J. (Ver tabla 2.7)

Esta es la fórmula utilizada por el programa para valorar la circularidad:

$$4\pi (\text{Área celular} / \text{Perímetro celular}^2)$$

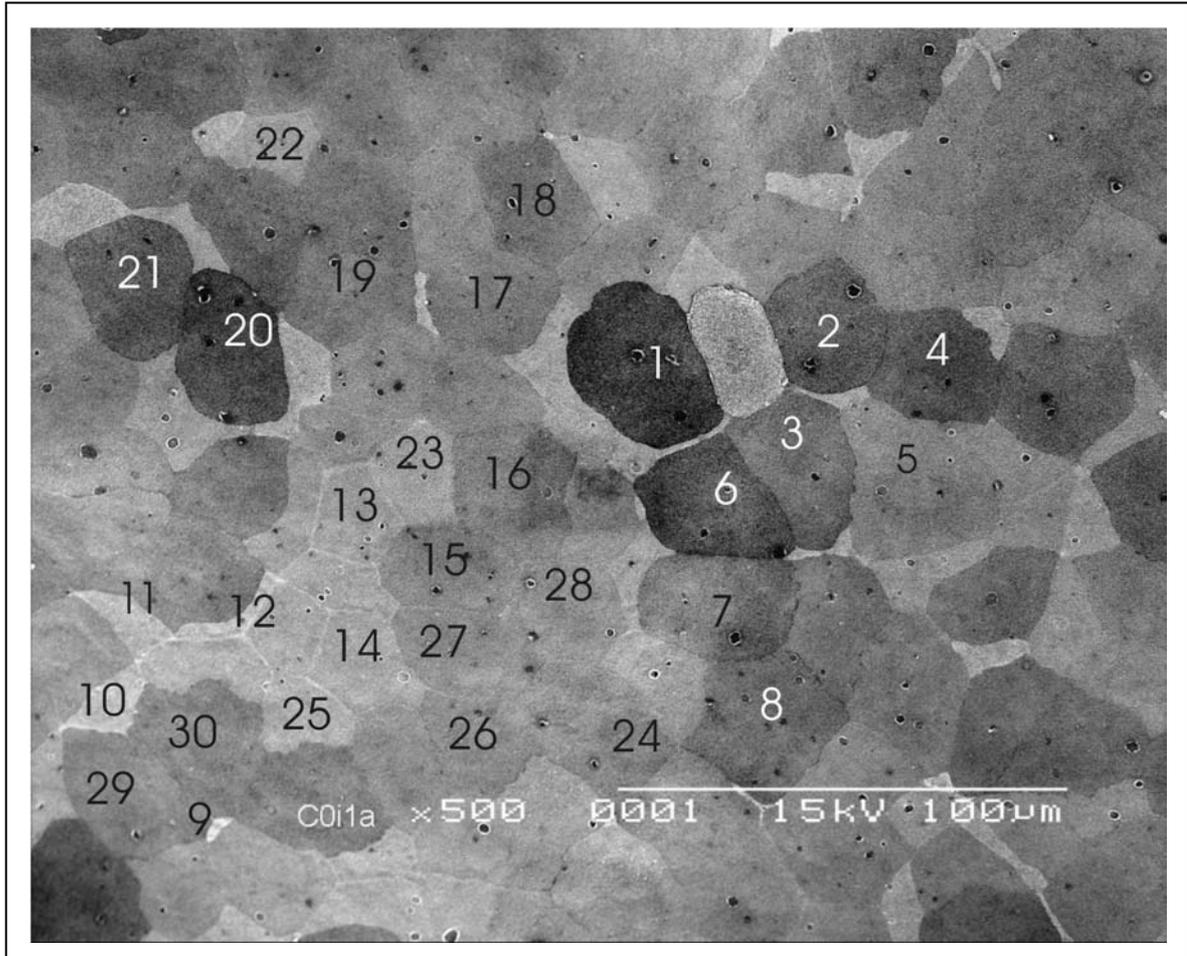


Figura 2.4. Imagen C0i1a con las células seleccionadas para este estudio

Para el caso de una forma perfectamente circular el valor del parámetro es 1 como puede observarse al sustituir los valores del área y el perímetro por la fórmula del área del círculo y la longitud de la circunferencia:

$$4\pi (\pi r^2 / (2 \pi r)^2) = 4\pi r^2 / 4\pi r^2$$

Este método, utilizado anteriormente por otros autores (Doughty 2004), mejora claramente la catalogación visual que presenta muchas dificultades debido a la gran variedad de pequeños matices en las formas celulares existentes (ver la segunda columna de la tabla 2.7 de resultados de este estudio)

Las células que poseen formas poligonales muy regulares, como es el caso de la célula 13 y 14 de este estudio (señaladas con un asterisco en la tabla 2.7),

pueden presentar valores desviados hacia la unidad debido a que su área se circunscribe dentro de una circunferencia. De este modo, dichas células se catalogan como pseudopoligonales con este método, aunque visualmente no lo sean. Estas discrepancias son muy poco frecuentes ya que, como se ha comentado anteriormente, en el epitelio corneal la inmensa mayoría de las células poligonales son irregulares.

Célula	Forma de <i>visu</i>	Area	Circ.
1	Pseudopoligonal	14416	0.867
2	Pseudopoligonal/circular	9450	0.896
3	Poligonal	10255	0.770
4	Pseudopoligonal	10167	0.867
5	Poligonal	10528	0.701
6	Pseudopoligonal	10434	0.818
7	Pseudopoligonal/circular	11313	0.891
8	Poligonal	13169	0.787
9	Poligonal	247	0.567
10	Poligonal	2643	0.666
11	Poligonal	2196	0.480
12	Poligonal	1761	0.302
13	Poligonal*	6069	0.834
14	Poligonal*	6004	0.877
15	Poligonal	6680	0.791
16	Poligonal	9781	0.797
17	Circular/Pseudopoligonal	9671	0.927
18	Pseudopoligonal	8496	0.851
19	Pseudopoligonal	8463	0.847
20	Pseudopoligonal	11088	0.879
21	Pseudopoligonal	10292	0.870
22	Poligonal	4492	0.695
23	Poligonal	4497	0.796
24	Poligonal	5673	0.710
25	Pseudopoligonal/poligonal	4357	0.829
26	Pseudopoligonal	8050	0.889
27	Pseudopoligonal	7009	0.834
28	Pseudopoligonal/poligonal	8966	0.811
29	Poligonal/pseudopoligonal	10845	0.786
30	Pseudopoligonal/poligonal	10198	0.801

Tabla 2.7. Valores de la variable *forma* para las células escogidas en el estudio preliminar que establece las categorías de *cat forma*. *Células de catalogación errónea con el método utilizado.

Finalmente, en base a unos criterios visuales se han establecido los límites entre categorías evitando así grandes discrepancias entre la observación de visu y la categoría asignada. Dichos criterios establecen que, una célula se considerará poligonal si, visualmente, posee vértices marcados y será circular

si no presenta vértices. Las células con vértices redondeados estarán excluidas de los dos grupos anteriores y constituirán la categoría pseudopoligonal.

De este modo, la categoría poligonal de la variable subordinada *cat forma* incluye a todas las células cuya variable *forma* tiene un valor menor de 0,8. Por su parte, la categoría pseudopoligonal incluye a las células con valores entre 0,8 y 0,9 y, finalmente, la categoría circular se asigna a aquellas células cuya variable *forma* es igual o mayor que 0,9.

Tono

Como ya se ha comentado en el capítulo de la introducción, todos los autores que han descrito el epitelio corneal han agrupado a las células, según su tono, en células claras, intermedias y oscuras. En concreto, el tono de gris se ha establecido siempre de forma subjetiva, mediante la observación visual de las imágenes a 500 aumentos del SEM (Doughty 1990b y Rigal 1993, entre otros), sin ninguna referencia o patrón.

Por otro lado, el rango de grises de una imagen puede estar afectado por muy diversas causas (Doughty 1990a), algunas de las cuales nada tienen que ver con las características celulares. Estos factores ajenos a los propiamente celulares son: la consistencia del baño de oro y el ajuste de brillo y contraste del microscopio electrónico, por nombrar los más importantes.

Para establecer el tono celular no sería correcto utilizar imágenes a aumentos más altos que 500x ya que, al visualizarse mejor los detalles de la superficie celular, se originaría una mayor dispersión de los tonos de gris. De este modo, una alta resolución hace que la célula pierda la uniformidad de tono que poseía en las imágenes a 500x. En la figura 2.5 aparecen varios ejemplos de histogramas de nivel de gris de una célula en imágenes a 500 y a 5000x.

En consecuencia, y con la idea de realizar una valoración más objetiva del tono celular, se ha establecido que la variable *tono* se obtendrá hallando la media

del tono de gris de todos los píxeles circunscritos dentro del área celular de una imagen a 500x.

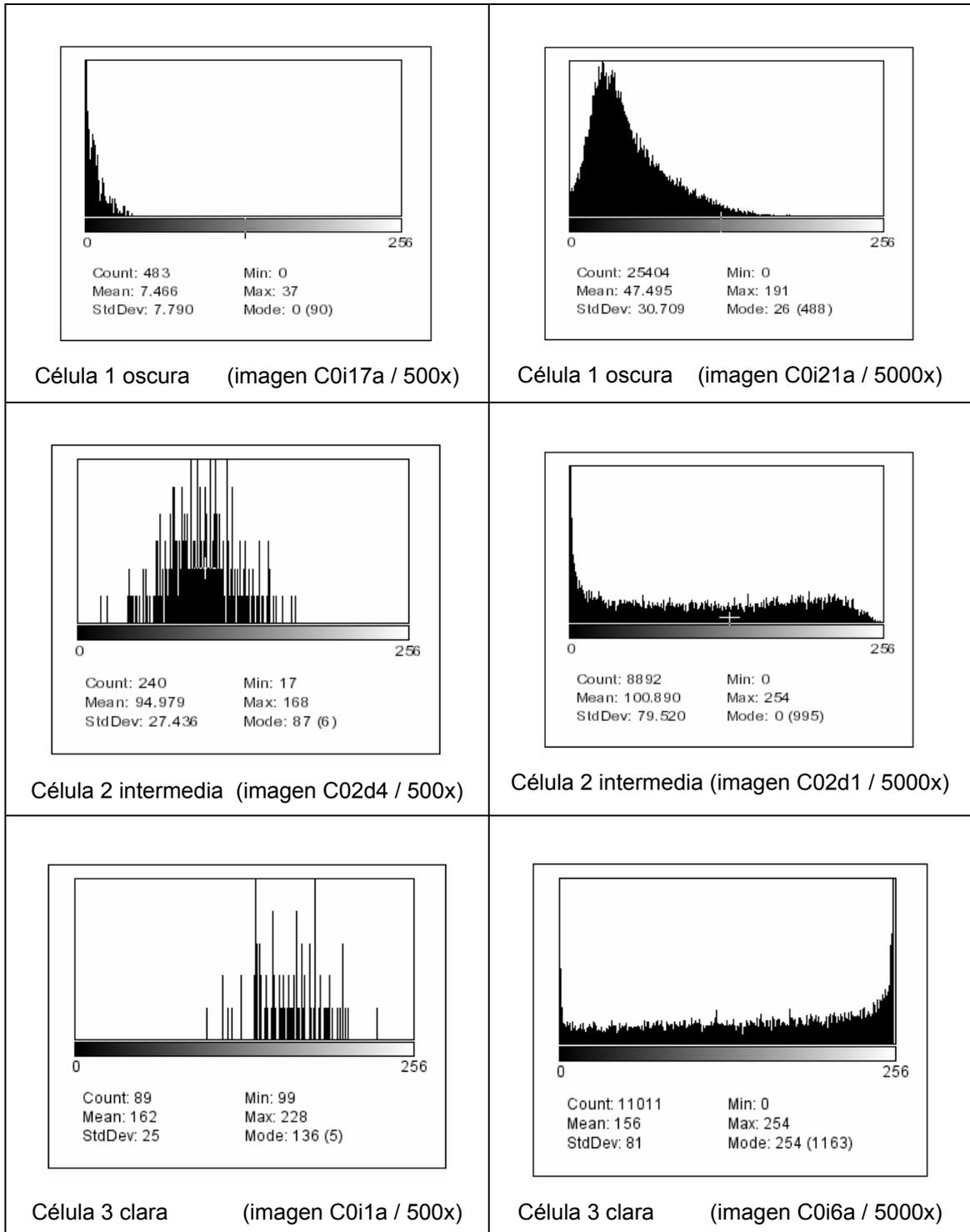


Figura 2.5. Efecto de los aumentos en el histograma de niveles de gris de las imágenes celulares. Las tres células de esta figura se han clasificado visualmente.

Para obtener este valor se perfila el contorno celular, con el programa de procesado de imágenes *ImageJ* y se calcula la media del tono de gris del área celular señalada (*Image J/ Analyze/ Measure*), estableciendo previamente la orden de medida (*Analyze/set Measurements*).

Se ha elegido la media porque es un valor que representa al total de los tonos de gris de la superficie celular. Además, se ha realizado un estudio previo de los parámetros del histograma (media, moda, máximo y mínimo) de la imagen de doce células visualizadas en la fotografía C0i1a. Se escogió dicha imagen a 500x, en la que se observan un total de 126 células, porque es nítida y posee un buen nivel de brillo y contraste, mostrando una gran variedad de tonos celulares. (Ver figura 2.6)

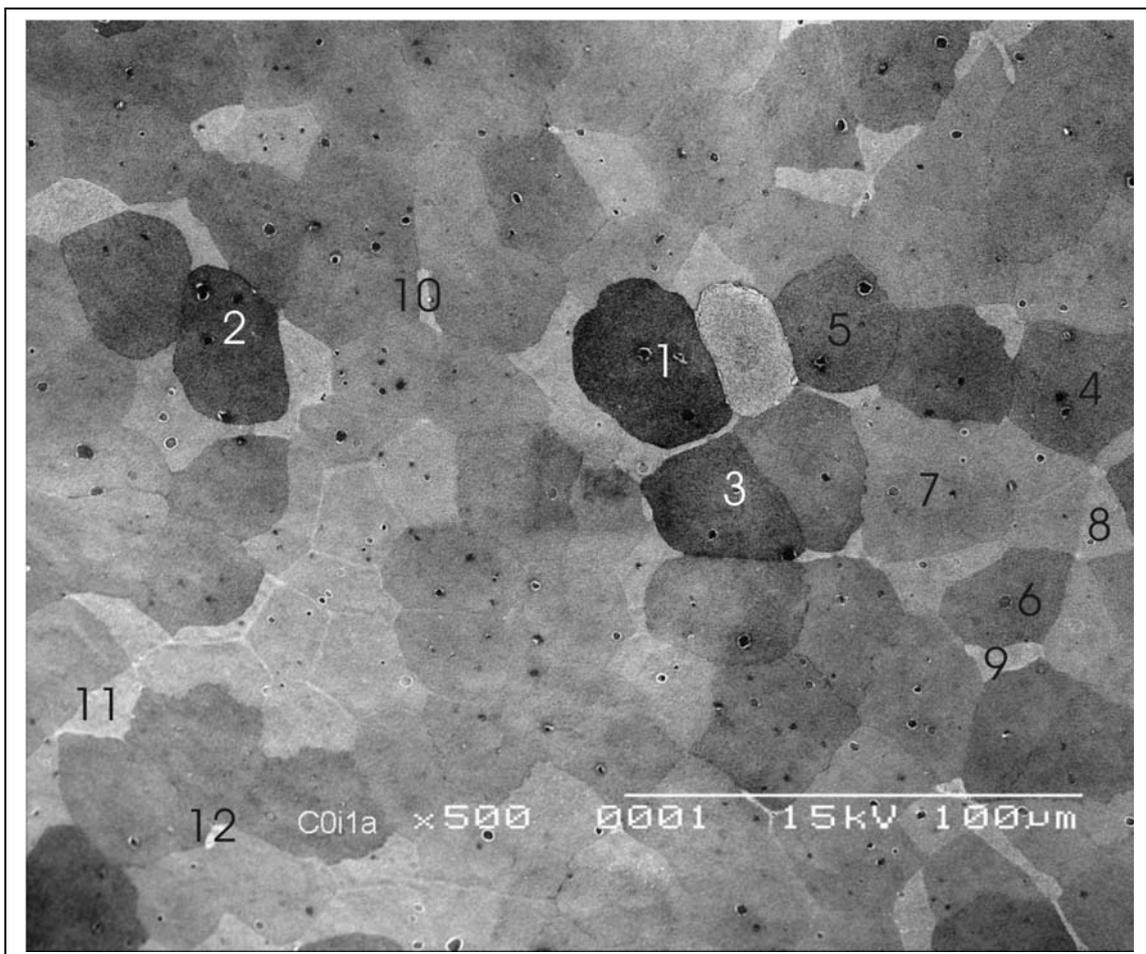


Figura 2.6. Imagen C0i1a. En ella aparecen numeradas las células seleccionadas en el estudio preliminar de la variable *tono*.

Este estudio previo muestra que la media de niveles de gris de una imagen celular a 500x presenta una gradación que concuerda con la catalogación mediante el tono celular, de *visu*. (Ver figura 2.7 y tabla 2.8)

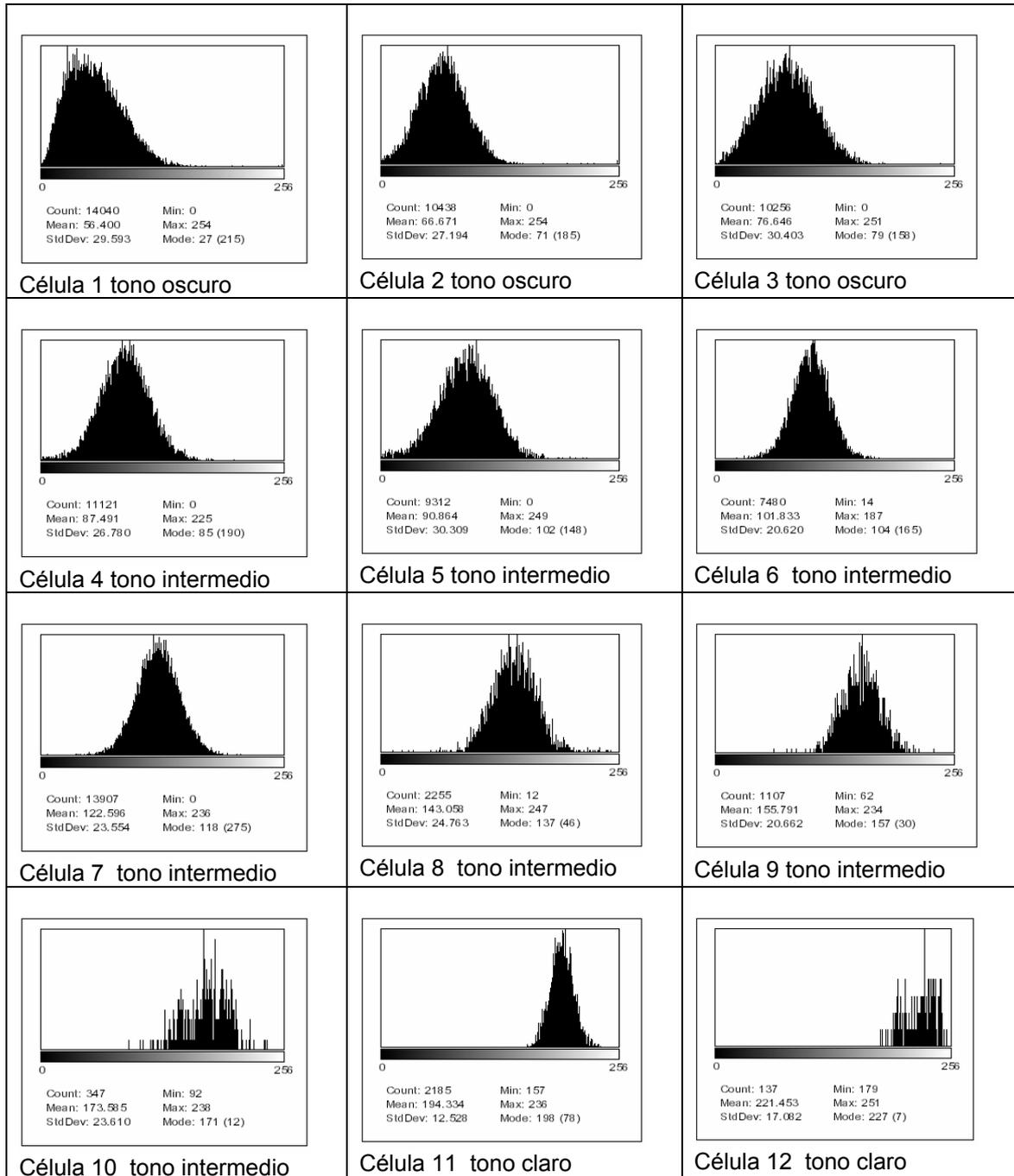


Figura 2.7. Histogramas tonales de las imágenes celulares incluidas en el estudio de la variable *tono*. Obsérvese como el rango del histograma va desplazándose desde la izquierda del gráfico (valores oscuros) hacia la derecha (valores claros), en concordancia con la gradación del tono de las células seleccionadas.

Por su parte, la moda del histograma, al ser el nivel de gris más frecuente, no representa al total de tonos de la imagen celular y podría catalogar de forma incorrecta algunas células con elevada dispersión de niveles de gris. El nivel máximo y el mínimo se han descartado ya que poseen una dependencia muy elevada del ajuste de brillo y contraste de la imagen.

Célula	Categoría tonal de <i>visu</i>	Media	Desv típica	Moda	Mín	Máx
1	Oscura	56.4	29.593	27	0	254
2	Oscura	66.671	27.194	71	0	254
3	Oscura	76.646	30.403	79	0	251
4	Intermedia	87.491	26.78	85	0	225
5	Intermedia	90.864	30.309	102	0	249
6	Intermedia	101.833	20.62	104	14	187
7	Intermedia	122.596	23.554	118	0	236
8	Intermedia	143.058	24.763	137	12	247
9	Intermedia	155.791	20.662	157	62	234
10	Intermedia/clara	173.585	23.61	171	92	238
11	Clara	194.334	12.528	198	157	236
12	Clara	221.453	17.082	227	179	251

Tabla 2.8. Parámetros del histograma tonal de las células incluidas en el estudio de la variable *tono*. La media define a dicha variable (ver texto). Desv típica: desviación típica.

Una vez definido el proceso de medición de la variable principal *tono*, se han establecido las tres categorías de su variable subordinada, *cat tono* (véase, clara, intermedia y oscura). Para ello, se ha tomado como patrón la escala de grises (figura 2.8), propia de las imágenes digitalizadas con un rango dinámico de 8 bits (como es el caso de todas las imágenes de esta tesis). Dicha escala posee 256 tonos de gris (véase, también, histogramas de figuras 2.5 y 2.7), de los que el tono 0 es el más negro y el 255 es el más blanco.

Para establecer las tres categorías de *cat tono*, la escala se ha dividido en tres intervalos. Los límites de los cuales se han designado utilizando el criterio empleado, anteriormente, en otros estudios (Doughy 1990-2004) según el cual las células claras y las oscuras son, sencillamente, los extremos de la escala

de tonos de gris. De este modo, el número de tonos incluidos en la categoría clara ha de ser igual al de la categoría oscura y estos límites nunca deben contradecir a la observación de visu de las doce células de la imagen C0i1a, escogidas para el estudio preliminar, sino, más bien facilitarla.

Por otro lado las categorías no se pueden establecer dividiendo la escala de grises en tres partes iguales ya que las células intermedias poseen una variedad de tonos mucho más amplia que las otras dos categorías (Doughty 1990a) (Rigal 1993).

Teniendo en cuenta estas premisas, se ha establecido que el límite entre una célula oscura y una intermedia será el nivel 77 de la escala patrón de grises (la franja de tonos oscuros delimitada por este valor corresponde al 30% de la escala patrón). Por su parte, el límite a partir del cual una célula se considerará clara será el nivel 179 (así, los tonos claros también se corresponden con un 30% de la escala). La franja de tonos intermedios se ha establecido por exclusión.

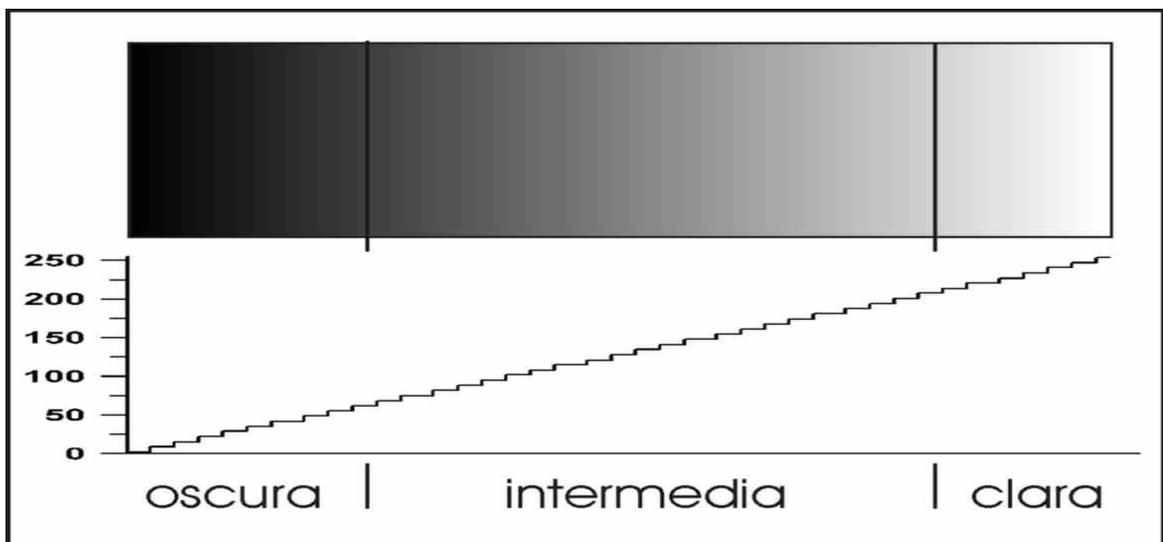


Figura 2.8. Escala patrón de tono celular.

Cabe especificar que, para el estudio del origen del tono celular (estudio que se ha realizado exclusivamente en las células control) se han utilizado tres

variables que miden el nivel de gris de la imagen, por lo que se les ha llamado variables tonales.

La primera es *tono*, cuyo cálculo ya se ha detallado en las páginas anteriores. Las dos restantes son *tono medio de las microproyecciones* (*tono medio_{mcp}*) y *tono medio de la membrana* (*tono medio_{mb}*).

El *tono medio de la membrana* (*tono medio_{mb}*), representa la media del tono de gris de los píxeles del área celular a 5000x que se han considerado integrantes del fondo de dicha imagen en el proceso de binarización. Por su parte, el *tono medio_{mcp}* se obtiene como resultado de promediar el tono de aquellos píxeles que se consideran “objetos” en el proceso de binarización de un área celular a 5000x.

Para el cálculo de estas dos variables se deben llevar a cabo los siguientes pasos:

1. Binarizar la imagen celular a 5000x en base al histograma de tonos de gris de una zona de la célula, previamente seleccionada (comandos *Image/ Adjust/ Threshold/ Auto* del programa *ImageJ* 1.33u).
2. Sin cerrar la ventana del umbral de binarización, se activa, de nuevo, la ventana de la imagen (donde permanece marcada la zona de estudio).
3. Seguidamente, se obtiene el histograma de los píxeles considerados como objetos en la imagen original (por encima del mural de binarización), siguiendo la ruta *Analyze/Histograma* del menú principal. En dicho histograma aparece la media de nivel de gris de este grupo de píxeles, o sea, la variable *tono medio_{mcp}*. (Ver figura 2.9)
4. Una vez obtenido este valor, se cierra la ventana del histograma y se activa la del umbral de binarización invirtiendo los valores (mediante el comando “*set*” de la ventana) pero manteniendo el mismo umbral, de modo que los píxeles que antes eran negros ahora serán blancos y viceversa.
5. Una vez realizado este paso, y sin cerrar la ventana del umbral, se activa de nuevo la imagen y se sigue la ruta (*Analyze/Histograma*) para obtener el histograma de los píxeles considerados como fondo (por debajo del umbral de

binarización), o lo que es lo mismo, los que representan a la membrana celular. En este nuevo histograma aparece el nivel de gris medio que poseen dichos píxeles, obteniéndose, así, el valor de la variable tono medio de la membrana (*tono medio_{mb}*). (Ver figura 2.9)

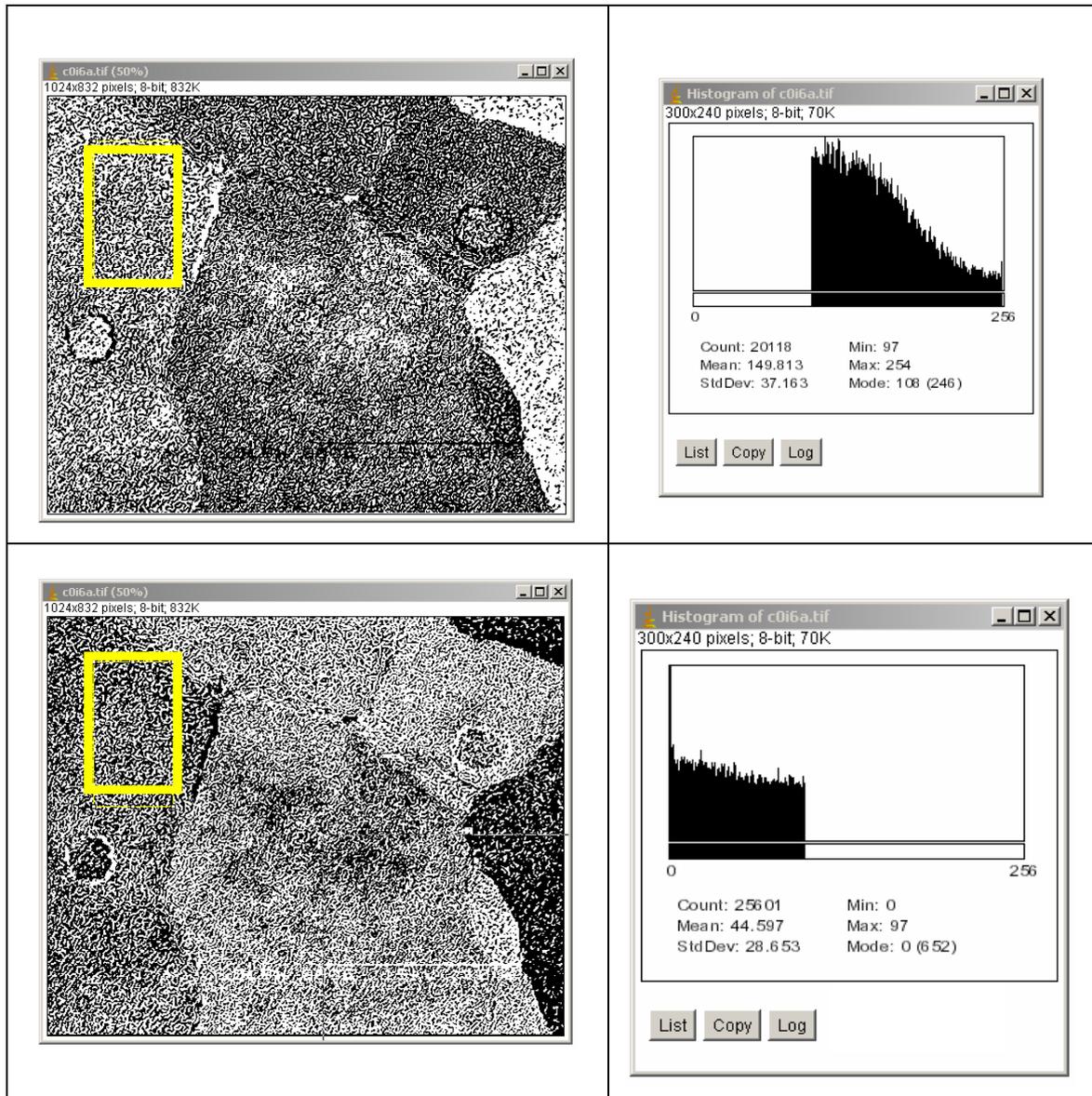


Figura 2.9. Calculo de las variables *tono medio_{mcp}* y *tono medio_{mb}*. En la columna de la izquierda aparecen la imagen C01i6 binarizada automáticamente (arriba) y la inversión de la binarización (abajo). En la columna de la derecha se muestran los histogramas de los píxeles por encima y por debajo del umbral de binarización (nivel 97). En el recuadro amarillo aparece la zona analizada.

Holes

Esta variable principal representa la presencia o ausencia de cráteres o *holes* en la superficie celular. Es, por lo tanto, una variable binaria (todo o nada), que posee cierto carácter subjetivo ya que se establece visualmente, sin emplear ningún proceso objetivo para su medición.

Al ser una variable principal cuyos valores se agrupan en dos categorías no se ha establecido ninguna variable subordinada a ella.

Uniones intercelulares

La variable principal, *uniones intercelulares*, define el estado de la adhesión entre células epiteliales vecinas, valorando su continuidad y aspecto. Como la valoración se realiza mediante la observación visual, esta variable es, también, de carácter subjetivo.

Mediante un análisis visual de las imágenes del epitelio corneal se han establecido cinco categorías (0; 0,2; 0,5; 0,8 y 1) para las diferentes tipologías observadas. En la figura 2.10 aparecen imágenes representativas de cada una de ellas. Al igual que *holes*, *uniones intercelulares* no tiene una variable subordinada ya que ella misma está categorizada. La descripción de cada categoría aparece en la tabla 2.9.

Uniones intercelulares	
Categoría	Definición
1	Uniones continuas, muy estrechas y planas
0.8	Uniones continuas, elevadas o muy marcadas
0.5	Uniones con discontinuidades o grietas
0.2	Perdida de la unión en alguna zona de la interfaz célula/célula
0.0	No hay uniones (La célula más externa se ha desprendido dejando una concavidad a través de la cual se observa la célula subyacente)

Tabla 2.9. Definición de las categorías de la variable *uniones intercelulares*.

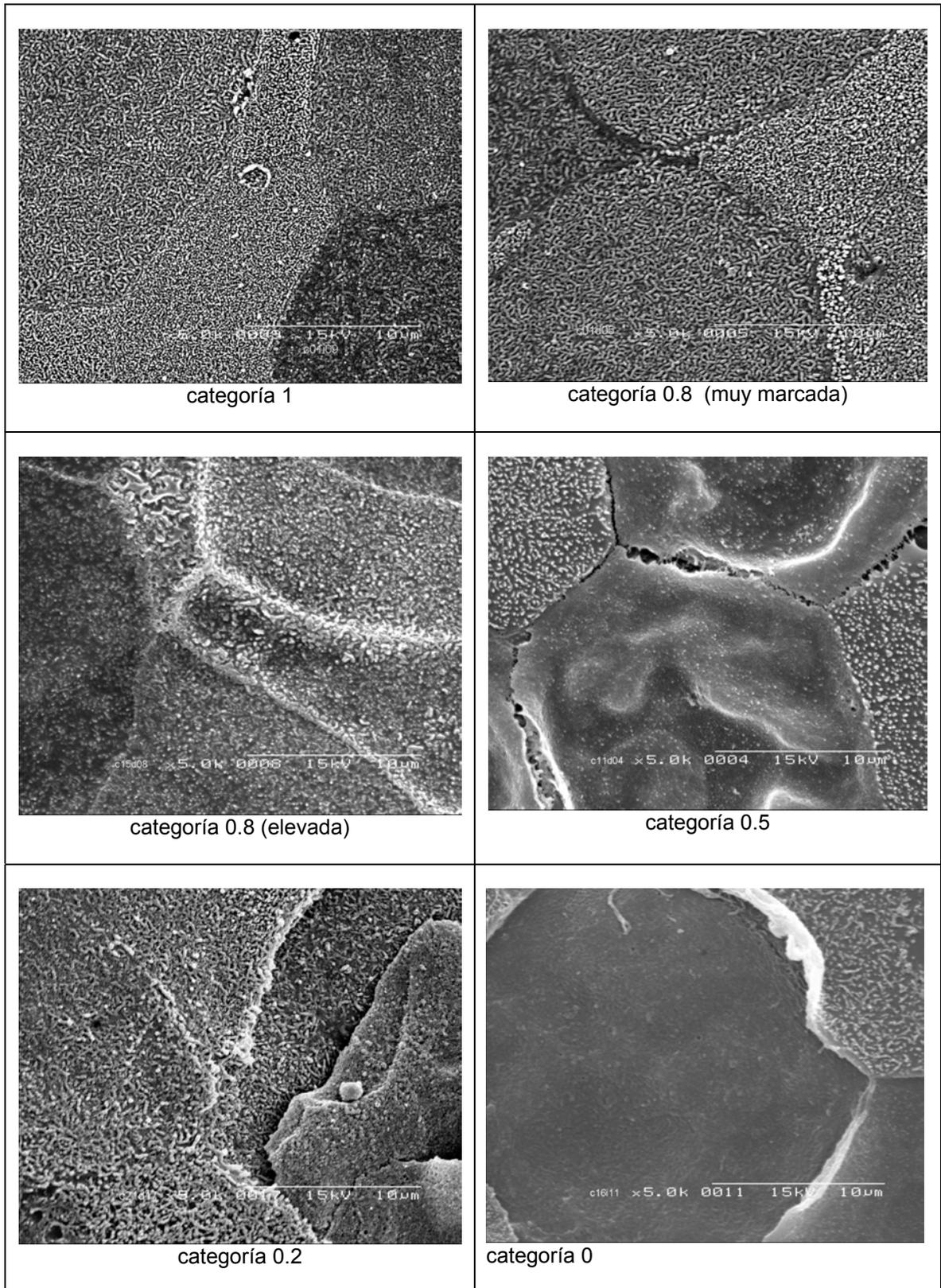


Figura 2.10. Categorías de la variable principal *uniones intercelulares*.

Hasta aquí la definición de las variables empleadas en esta tesis doctoral y la descripción de su método de cálculo. Con los valores adoptados por dichas variables, en cada una de las células estudiadas, se han confeccionado tres archivos de datos: uno con los datos de las células control, otro con los datos de todas las células y otro con los datos de las variables zonales. Los datos de este último archivo se han calculado promediando los valores obtenidos en las células que integran una zona. Posteriormente, todos estos datos se han analizado estadísticamente.

2.3. Métodos estadísticos

El análisis de datos de cualquier índole exige la utilización de métodos estadísticos en casi todos sus pasos. Los estadísticos básicos correspondientes al análisis exploratorio que se han utilizado se describen brevemente en el primer subapartado. El segundo subapartado se dedica a la presentación de los contrastes de hipótesis sobre igualdad de tendencias centrales de diversas poblaciones. Finalmente, el subapartado tercero describe el modelo de regresión logística que ha sido utilizado repetidas veces a lo largo de este estudio. Todas las técnicas descritas pueden considerarse estándar y, en particular, están disponibles en SPSS 12.0S “para Windows” versión 12.01 (SPSS INC. 1989-2003). El contenido del primer apartado puede hallarse en cualquier texto de estadística. Los textos de análisis estadístico multivariante como Seber (1984), Rice (1995), Peña (2002) o Gómez-Villegas (2005) contienen información más detallada de los otros dos apartados.

Estadísticos básicos

Las técnicas de descripción de muestras se basan en el uso de estadísticos o funciones de la muestra. Los que han sido utilizados pueden clasificarse en medidas de tendencia central y medidas de dispersión. Las medidas de tendencia central utilizadas son: la media o promedio y la mediana. Si se denota la muestra de extensión n como x_i , $i = 1; 2; \dots; n$, se define media muestral o promedio como

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

La media muestral representa el centro de masas de la muestra cuando sus elementos se consideran como unidades de masa y se distribuyen según su valor sobre un eje.

La mediana también describe el centro de la muestra pero prescinde de los valores concretos de cada elemento. Se define como el valor tal que el número de elementos de la muestra inferiores a él es igual al número de superiores. Si n es impar la mediana coincide con un elemento de la muestra; si n es par hay un intervalo de valores que satisfacen la condición, entonces se suele tomar el punto medio de los dos elementos de la muestra más cercanos.

La mediana coincide con el cuantil 0.5 o percentil 50 %. La diferencia más importante entre la media y la mediana es que la primera es sensible a los valores extremos de la muestra, mientras que la mediana no tiene en cuenta estos valores. Por ello se dice que la mediana es una medida de tendencia central más robusta que la media.

Las medidas de dispersión de una muestra que se utilizan son: la varianza, la desviación típica y el coeficiente de variación muestrales, diversos percentiles, el máximo y el mínimo.

Se define variabilidad de una muestra respecto a un punto ξ como

$$V(\xi) = \frac{1}{n} \sum_{i=1}^n (x_i - \xi)^2$$

El valor de ξ que hace mínimo esta cantidad es precisamente la media muestral \bar{x} y esa variabilidad mínima es la varianza muestral. Es frecuente modificar la definición de varianza muestral cambiando el factor $1/n$ a $1/(n-1)$.

Así la varianza muestral que se ha utilizado es

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

La desviación típica muestral utilizada es la raíz cuadrada S que tiene las unidades de la variable que se estudia.

El coeficiente de variación muestral representa otra manera de expresar la desviación típica. Es un porcentaje que se define como 100 veces la desviación típica dividida por la media. Su principal aportación estriba en relacionar la desviación típica y el nivel al que se han hecho las medidas y, por lo tanto, permite comparar las dispersiones de dos variables diferentes.

Los cuantiles o percentiles muestrales pueden considerarse también medidas de dispersión pues definen intervalos que contienen fracciones fijadas de la muestra. El q -cuantil ó $q \times 100$ -percentil es el mayor valor tal que el $q \times 100\%$ de la muestra es inferior o igual que él. Si no coincide con ningún valor de la muestra se toma el punto medio de los dos más cercanos.

Una de las formas más efectivas de mostrar las características de la distribución de una muestra es el diagrama de caja o box-plot. Consiste en una caja a lo largo del eje limitada por los cuantiles 0.25 y 0.75 (o cuartiles) y esta caja se divide a su vez por el valor de la mediana (ver figuras 1,2,3,4,7,10 y 11 del apartado de resultados estadísticos). La longitud de la caja es $2L$. Los datos mayores que el cuantil 0.75 más $3L$ ó menores que el cuantil 0.25 menos $3L$ se los considera erráticos y se representan individualmente. El diagrama se completa con un brazo que parte de la caja central y llega hasta el más grande y más pequeño de los datos que no son erráticos. Este diagrama permite visualizar la forma de la distribución de la muestra (centro, dispersión, asimetría, datos erráticos o extremos) y realizar comparaciones visuales entre diferentes poblaciones (ver, de nuevo, figuras 1,2,3,4,7,10 y 11 del apartado de resultados estadísticos).

Finalmente el estadístico más utilizado en el análisis multivariante de datos es el coeficiente de correlación entre dos variables (X, Y) . Si se dispone de una muestra bivalente, (x_i, y_i) , $i = 1; 2; \dots; n$, se define covarianza muestral

$$C(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Como en el caso de la varianza muestral, también se puede utilizar la versión en que se divide por n .

La covarianza muestral normalizada, para liberarla de la escala de las variables, es el coeficiente de correlación (empírico)

$$\rho(X, Y) = \frac{C(X, Y)}{S_x S_y}$$

donde la normalización se ha realizado dividiendo por la desviación típica muestral de ambas variables. Los valores del coeficiente de correlación están entre -1 y 1. Estos valores extremos indican una relación lineal entre las variables. Por su parte, los valores cercanos a 0 indican la ausencia de relación lineal. El signo negativo corresponde a relaciones lineales opuestas: cuando una variable crece la otra decrece. Y el signo positivo indica crecimiento o decrecimiento de ambas variables en el mismo sentido.

Contrastes de igualdad de tendencia central y bondad de ajuste

Los contrastes de hipótesis son una de las herramientas básicas de la estadística. Se establece una hipótesis, generalmente designada por H_0 , sobre el modelo del que se supone que provienen los datos de los que se dispone. Esta hipótesis se complementa con una hipótesis alternativa, H_1 con la que se compara H_0 . El objetivo es determinar si los datos indican que debe rechazarse H_0 . Generalmente, si no existen esos indicios se acepta H_0 . La cuantificación del nivel de rechazo de H_0 se hace a través del p -valor o nivel de significación del contraste que toma valores entre 0 y 1. Un p -valor cercano a 0 indica un

rechazo contundente de H_0 . Valores mayores que 0.1, pueden asociarse a situaciones en que no hay razón para rechazar H_0 . El nivel del 0.05 ó 5% suele tomarse como referencia, valores mayores llevan a la aceptación de H_0 y valores menores a su rechazo.

Los contrastes de hipótesis se evalúan teniendo en cuenta los dos tipos de errores que pueden cometerse al tomar la decisión de aceptar o rechazar H_0 : son errores de tipo I los que se cometen rechazando H_0 cuando esta hipótesis es cierta; son errores de tipo II los que corresponden a no rechazar H_0 cuando esta hipótesis es falsa. El p -valor es el valor máximo de la probabilidad de error de tipo I para el que puede aceptarse H_0 .

Los contrastes de hipótesis son omnipresentes en los métodos estadísticos. Por ejemplo, cuando se estima un coeficiente de correlación ρ , se suele acompañar el resultado de un p -valor correspondiente al contraste $H_0 : \rho = 0$, $H_1 : \rho \neq 0$. Un p -valor mayor que 0.05 indica que la correlación puede considerarse nula o al menos no existe razón para suponer lo contrario. Un valor 0.01 o menor indica que no hay razones para pensar que $\rho = 0$.

En el estudio que sigue se han utilizado contrastes de hipótesis referentes a la igualdad de medias de varias poblaciones (supuestas con distribución normal). Estos contrastes suelen agruparse bajo el título de análisis de la varianza (ANOVA) y pueden considerarse estándar en sus versiones más corrientes. Pueden formularse como sigue. Se dispone de muestras de variables X_1, X_2, \dots, X_K , donde el subíndice indica poblaciones diferentes que se supone que sólo varían según un factor y la letra del subíndice indica el nivel de ese factor. Se supone que la población X_i tiene distribución normal con media μ_i desviación típica σ_i . Las extensiones de las muestras pueden ser diferentes.

La hipótesis del contraste es

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

siendo la alternativa que, al menos, una de estas medias sea diferente a las demás. Entre los contrastes bajo el epígrafe ANOVA los hay que admiten que la varianza de las diversas variables sean diferentes y, por supuesto, que las extensiones de las muestras sean también diferentes. Se han utilizado los contrastes disponibles en SPSS 12.0S para Windows versión 12.01 SPSS INC en el apartado de comparación de medias. Las bases de los test de tipo ANOVA son la suposición de normalidad de las variables y la descomposición de la varianza de la muestra total en componentes intra-población e inter-poblaciones. Su comparación genera todos los contrastes de este tipo. La descomposición de la varianza da nombre común a este tipo de contrastes y ha motivado el nombre de análisis de la varianza. Las comparaciones entre dos poblaciones (un factor a dos niveles) desembocan en los estadísticos t-student, mientras que las comparaciones de varias poblaciones lo hacen a través de estadísticos de tipo F ó χ^2 .

Sin embargo, puede ocurrir que las diferentes poblaciones no puedan considerarse distribuidas normalmente. Para llegar a esta conclusión pueden aplicarse diversos contrastes de bondad de ajuste. Para poner a prueba la normalidad de cada una de las poblaciones se ha utilizado el contraste de Kolmogorov-Smirnov. La hipótesis H_0 de este contraste es que la distribución que corresponde a la muestra es una normal con parámetros desconocidos, los cuales se estiman a partir de la muestra. El test de Kolmogorov-Smirnov utiliza la diferencia máxima en probabilidad entre la función de distribución hipotética con la distribución empírica: se rechaza H_0 cuando esa diferencia es grande. La normalidad de cada una de las variables no asegura la normalidad conjunta de todas las poblaciones. Sin embargo, se ha renunciado a utilizar contrastes de normalidad multivariante porque para que estos sean efectivos se exigen muestras más extensas que las que se dispone.

En algunos casos no ha sido posible suponer la normalidad de las poblaciones que deben compararse. En estos casos se ha acudido a utilizar contrastes no-paramétricos que comparan la tendencia central de las muestras. En este contexto se han utilizado los test de Mann-Whitney y de Kruskal-Wallis.

El test de Mann-Whitney compara la localización de dos muestras; la localización sustituye a la media de las poblaciones que se utiliza en ANOVA. En este caso se tienen en cuenta cómo se ordenan los elementos de las muestras de ambas poblaciones (*estadístico U*). En este caso la hipótesis H_0 supone igual localización de ambas muestras. En cierto modo, puede considerarse que este test compara las medianas de ambas poblaciones. Su carácter no paramétrico procede de que no se hace ninguna suposición sobre la distribución de las poblaciones, lo que le da una amplia aplicabilidad a costa de una pérdida de potencia (incremento de probabilidad de errores de tipo II) comparado con las técnicas ANOVA.

La comparación de la localización de muestras procedentes de varias poblaciones en versión no paramétrica se ha abordado con el test de Kruskal-Wallis. Las muestras se consideran generadas por un parámetro de localización y perturbaciones aleatorias entorno del mismo para cada población. La hipótesis H_0 es que todos los parámetros de localización coinciden. El test utiliza la ordenación conjunta de las muestras y a cada dato le asigna su rango en la ordenación; entonces se comparan los rangos medios de todas las poblaciones. Aunque no se basa en los mismos principios que el contraste de Mann-Whitney, puede considerarse como una generalización al caso de varias poblaciones. Como en el caso del test de Mann-Whitney, este tipo de test es menos potente que sus homólogos de tipo ANOVA pero tiene la ventaja de su flexibilidad.

Regresión logística

La regresión logística es un modelo estadístico en el que se intenta predecir una respuesta a partir de una o más variables predictoras o covariables.

La variable respuesta es cualitativa y se supone que asigna un individuo a una categoría o grupo. Las covariables pueden ser numéricas o factores a varios niveles que corresponden al individuo. En general se interpreta que la variable respuesta asigna un individuo de la muestra a un grupo o categoría.

Supongamos que esos grupos son G_1, G_2, \dots, G_K , que X representa al individuo y x_1, x_2, \dots, x_m son las covariables que lo describen. El modelo de regresión logística consiste en expresar las probabilidades de que el individuo pertenezca a cada grupo a partir de las covariables según

$$P[X \in G_k] = \frac{\exp\left[\beta_{k0} + \sum_{j=1}^m \beta_{kj} x_j\right]}{1 + \sum_{g=1}^{k-1} \exp\left[\beta_{g0} + \sum_{j=1}^m \beta_{gj} x_j\right]}, \quad k=1,2,\dots,K-1$$

donde el grupo K -ésimo se ha tomado como referencia y su probabilidad toma la forma

$$P[X \in G_k] = \frac{1}{1 + \sum_{g=1}^{K-1} \exp\left[\beta_{g0} + \sum_{j=1}^m \beta_{gj} x_j\right]}$$

Los coeficientes β_{gj} , correspondientes al grupo g y la covariable j , deben determinarse a partir de los datos. La forma que adquiere el modelo se adapta a que la suma de las probabilidades de pertenencia a cada grupo deben ser la unidad.

Los datos correspondientes a un individuo son el grupo al que pertenece y los valores de las covariables. Para estimar los coeficientes se establece la verosimilitud de los datos en función de los coeficientes, es decir, la probabilidad de que cada individuo pertenezca al grupo del que proviene, fijado el valor de sus covariables. Los coeficientes β_{gj} se eligen de forma que la verosimilitud de los datos sea máxima. Estos valores se obtienen mediante optimización numérica.

Una vez estimados los coeficientes, el modelo asigna a cada individuo una probabilidad de pertenencia a cada uno de los grupos según indican sus covariables. Además, puede asignarse el grupo más probable a cada individuo de la muestra como grupo predicho por el modelo. La comparación entre el grupo observado y el grupo predicho permite medir la calidad del modelo.

Finalmente, conocidos los valores de las covariables de individuos ajenos a la muestra, se puede asignar las probabilidades de pertenencia a cada grupo y un grupo predicho y, por tanto, el modelo puede considerarse como un método discriminante o de clasificación de individuos.

La regresión logística admite también la realización de contrastes de hipótesis sobre los coeficientes. La hipótesis que se pretende contrastar es $\beta_{gj} = 0$ contra la alternativa $\beta_{gj} \neq 0$. El contraste se realiza mediante los estadísticos de Wald. Cuando el contraste no permite rechazar que $\beta_{gj} = 0$, se infiere que la covariable χ_j no tiene influencia significativa en la predicción del grupo G_g . Por ello estos contrastes son muy útiles para analizar el papel de las covariables respecto a cada uno de los grupos.