

ANÁLISIS DE DATOS
HIPERTIROIDISMO

ADOLFO GUZMÁN
IGNACIO IBAÑEZ

Profesor: Felipe Bello
Ayudantes: Fernanda Lobos
Alfonso Guzmán

Santiago - Chile
April 9, 2017

TABLA DE CONTENIDOS

CAPÍTULO 1. INTRODUCCIÓN	5
CAPÍTULO 2. DESCRIPCIÓN DEL PROBLEMA.....	7
2.1 DESCRIPCIÓN DE LA BASE DE DATOS	8
2.2 DESCRIPCIÓN DE CLASES Y VARIABLES	8
CAPÍTULO 3. ANÁLISIS	13
3.1 VARIABLES BOOLEANAS	13
3.2 VARIABLES NUMÉRICAS	16
CAPÍTULO 4. CONCLUSIÓN.....	21
CAPÍTULO 5. REFERENCIAS	23

CAPÍTULO 1. INTRODUCCIÓN

El sistema endócrino junto al sistema nervioso constituyen los dos principales sistemas de comunicación y control del organismo. El sistema nervioso se caracteriza por la rapidez en que es capaz de realizar cambios en el cuerpo, mientras que el sistema endocrino, puede cambiar la disposición corporal a través de señales químicas, hormonas o neurotransmisores, que surten efecto de manera casi inmediata o de manera tardía día o incluso semanas, provocando grandes cambios en el organismo como la pubertad o controlando procesos metabólicos y anabólicos, cuidando la homeostasis.[1] Sin embargo, existen cuadros en los que hay desequilibrios en estas señales, en el presente informe se estudia el cuadro conocido como hipertiroidismo.

El hipertiroidismo, en términos sucintos, es un cuadro provocado por el exceso de secreciones de hormonas tiroideas tales como la tiroxina(T4), triyodotironina(T3) que puede generar síntomas como dificultad para concentrarse, fatiga e incluso problemas cardiacos y osteoporosis.[2] Por lo tanto, se vuelve esencial entender cuáles son las características comunes a los casos de esta enfermedad para que se pueda diagnosticar efectivamente y evitar más daño de la enfermedad.

En el presente informe se utiliza un conjunto de datos con una serie de resultados de exámenes en cuanto a los niveles de hormonas tiroideas, neurotransmisores y otros índices, ellos son explicados y caracterizados detalladamente en la descripción del problema, tanto a nivel de base de datos como a nivel de clases y variables. Luego se realiza el análisis estadístico de los datos y finalmente se exponen las conclusiones del estudio.

CAPÍTULO 2. DESCRIPCIÓN DEL PROBLEMA

Como se mencionaba anteriormente el hipertiroidismo es un cuadro que se presenta como una anomalía del sistema endocrino, por lo tanto, es necesario recabar información sobre las diferentes hormonas secretadas, pero que también inciden en el normal funcionamiento de la tiroides además de caracterizaciones demográficas que también pueden tener incidencia en la aparición de esta enfermedad.

Por otro lado, también es necesario realizar una breve explicación del modo en que son secretadas las hormonas en el organismos para entender de manera general cuál es el papel que toma cada una de ellas.

La regulación de las hormonas tiroides funcionan en lo que se conoce como un circuito cerrado. En un primer momento el hipotálamo secreta TRH, la hormona que es responsable posteriormente de que la hipófisis secreta la TSH, hormona estimulante de la tiroides, esta finalmente secreta T4 y T3, y en la medida que aumentan sus concentraciones en la sangre inhiben la secreción de TSH y de TRH.

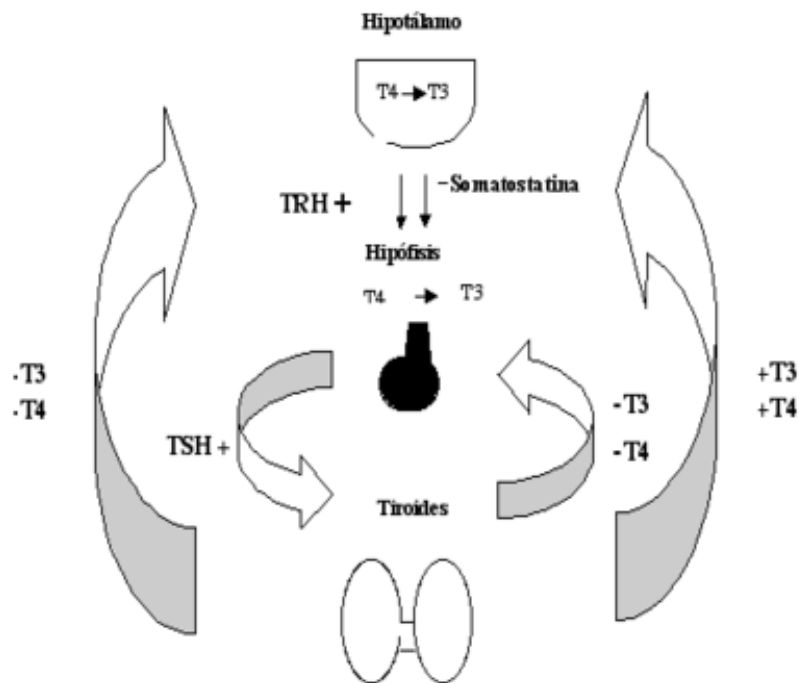


Figura 2-1: Esquema de circuito cerrado de la secreción de hormonas tiroideas[7]

2.1 DESCRIPCIÓN DE LA BASE DE DATOS

Acorde a lo expresado en la descripción general ofrecida por el repositorio de Machine Learning, el archivo que contiene el dataset, están compuesto por 2800 observaciones, las cuales contienen 29 atributos, los cuales pueden ser numéricos o de tipo booleano.

Estos datos están formateados para el algoritmo de decisión de Quila, por lo tanto el número que se encuentra después de la barra vertical de cada ejemplo, el cual corresponde a la id del paciente, es ignorado. Por otro lado el archivo con la extensión .name, era utilizado para la caracterización del dataset .

2.2 DESCRIPCIÓN DE CLASES Y VARIABLES

Finalmente la base de datos contiene 29 variables, que corresponde a los atributos de cada uno de los ejemplos. Todas pertenencientes al dominio de las hormonas tiroides o un poco de demografía de los pacientes, al incluir la edad.

A continuación se presenta una breve descripción de cada una de las variables:

- 1.- Edad: indica la edad del paciente, según la descripción de la base de datos corresponde a una variable numérica y continua.
- 2.- Sex: corresponde al sexo del paciente, puede tomar los valores M y F.
- 3.- On thyroxine: muestra si el paciente está o no en alguna medicación de tiroxina. Puede tomar valores booleanos.
- 4.- Query on thyroxine: muestra si el paciente está o no en una consulta sobre los niveles de tiroxina. Puede tomar valores booleanos.
- 5.- On antithyroid medication: muestra si el paciente está o no en un tratamiento que contrarestre los efectos de las hormonas tiroideas. Puede tomar valores booleanos.
- 6.- Sick: muestra si el paciente enfermo o no. Puede tomar valores booleanos.
- 7.- Pregnant: muestra si el paciente está embarazado o no. Puede tomar valores booleanos.
- 8.- tThyroid surgery: Muestra si el paciente fue tratado con una cirugía a la tiroides. Toma valores booleanos.

- 9.- I131 treatment: Muestra si el paciente está o no en un tratamiento de I131. Este corresponde a un tratamiento para paciente con hipertiroidismo en el cuál se utilizan pequeñas cantidades de un isótopo de la yodo, que emite radiación y permite destruir las células de la tiroides.[3]
- 10.- Query hypothyroid: muestra si el paciente está en consulta por hipotiroidismo. Toma valores booleanos.
- 11.- Query hyperthyroid: muestra si el paciente está en consulta por hipertiroidismo. Toma valores booleanos.
- 12.- Lithium: muestra si el paciente está o no en tratamiento de litio, que permite controlar los episodios maníacos.[4] Toma valores booleanos.
- 13.- Goitre: muestra si el paciente posee o no bocio, este corresponde a un crecimiento anormal del tamaño de la tiroides. Toma valores booleanos.
- 14.- Tumor: muestra si el paciente tiene o no tumores en la tiroides. Toma valores booleanos.
- 15.- Hypopituitary: muestra si existe o no una carencia de hormonas hipofisarias, una o varias. Toma valores booleanos, verdadero si falta al menos una, falso en otro caso.
- 16.- Psych: plantea si el paciente está o no en tratamiento psiquiátrico. Toma valores booleanos, verdadero si está en un tratamiento, falso en caso contrario.
- 17.- TSH measured: Expresa si el paciente se sometió a pruebas de la hormona TSH. Esta hormona es clave en el estudio, ya que es la encargada de realizar la estimulación de la tiroides. Toma valores booleanos.
- 18.- TSH: muestra los niveles de hormona en el torrente sanguíneo. Es una variable numérica y continua.

- 19.- T3 measured: Expresa si el paciente fue sometido a pruebas de medición de la hormona T3. Esta hormona es producida por la tiroides y participa en la regulación de procesos metabólicos, al crecimiento, temperatura corporal y el ritmo cardiaco.[5]
- 21.- T3: Corresponde al nivel encontrado en la sangre de la hormona. Es una variable numérica y continua.
- 22.- TT4 measured: Expresa si el paciente fue sometido a pruebas de medición de la hormona T4. Esta corresponde a la prohormona de la T3, es decir, tras un proceso de transformación en tejido especializado y enzimas de yodidasas se transforma en T4. Tiene los mismo efectos que la T3, pero es menos potente. Esta variable puede tomar valores booleanos.
- 22.- TT4: Corresponde al nivel encontrado en la sangre de la hormona. Es una variable numérica y continua.
- 23.- T4U measured: Expresa si fue realizado el examen T4U, el cuál consiste en determinar cuál es el consumo de la hormona, a través de un test con indicadores radioactivos.
- 24.- T4U: valores de consumo de la hormona. Es una variable numérica
- 25.- FTI measured: muestra si fue realizado o no un test de T4 libre. Este examen se caracteriza por utilizar la porción de hormona T4 que no está asociada a proteínas y por tanto refleja mejor el funcionamiento de la tiroides. [6]
- 26.- FTI: Niveles de T4 libres en la sangre. Toma valores numéricos continuos.
- 27.- TBG measured: Expresa si fueron realizados exámenes sobre la globulina fijadora de tiroxina. Esta es clave en la captura de excesos de las hormonas T3 y T4, para que las porciones que queden libres realicen las actividades biológicas.
- 28.- TBG: valores que pueden tomar los niveles de la proteína. Es una variable numérica y continúa.

- 29.- referral source: Corresponde a las fuentes de los datos que son utilizados para conformar el dataset.

CAPÍTULO 3. ANÁLISIS

Para realizar el análisis se toman en consideración los dos tipos de variables predominantes en el dataset, las cuales corresponde a las variables booleanas y a las variables numéricas.

Por otro lado, tampoco se consideran todas las variables, consecuencia de que en su mayoría las variables booleanas corresponden a si se hizo o no el examen de cierta hormona y por lo tanto definen valores perdidos en la hormona medida. Sin embargo, como un primer acercamiento a la problemática simplemente esos valores se omitieron.

3.1 VARIABLES BOOLEANAS

Como se observa en la Figura 3-1, la relación entre el síndrome hipopituitario y el hipertiroidismo no presenta una dirección apreciable en los gráfico, como se observa la mayoría de los pacientes se encuentran dentro del grupo Negative, ya sea posean o no la característica.

Por lo tanto, se puede decir que una función disminuida de la hipófisis, no tiene una relación directa con el hipertiroidismo y que por lo tanto, debe existir un tercer factor de riesgo para esta enfermedad.

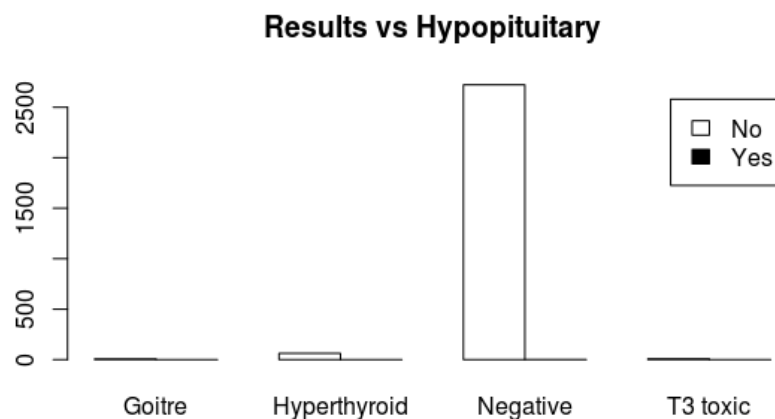


Figura 3-1: Relación entre Síndrome hipopituitario e hipertiroidismo[6].

En la tabla a modo intuitivo se puede decir que las mujeres sufren menos de hipertiroidismo, sin embargo, cuando se toma en consideración la proporción hombres-mujeres que participó del estudio, se observa como la cantidad de mujeres es mayor que la de hombres y por lo tanto los valores nominales mostrados en el gráfico no aportan información relevante desde ese punto de vista.

Ahora, lo que si es posible observar es que este mal no tiene relación directa con el sexo del paciente, es decir consecuencia de que las hormonas participantes en el control metabólico son trascendentes al sexo los problemas de regulación de ellas también lo serán, en este caso el hipertiroidismo.

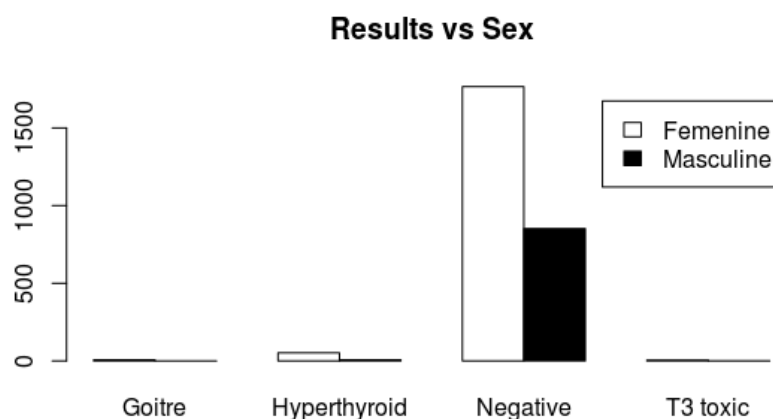


Figura 3-2: Relación entre el diagnóstico final y el sexo del paciente

Esta relación es especialmente importante, ya que una cirugía a la tiroides podría haber sido consecuencia del tratamiento de un cáncer y, por lo tanto, la presencia de hipertiroidismo podría estar relacionada con a un crecimiento anómalo del tejido tiroideo.

Ahora en el gráfico se observa como la existencia de la cirugía permite que el hipertiroidismo no sea una opción como enfermedad, ya que la mayoría de los casos se encuentran en Negative, por lo tanto las operaciones realmente sirven para rehabilitar a las personas de problemas tiroideos asociados a la hiperactividad. Sin embargo, eso no asegura la presencia de otros desordenes como el hipotiroidismo.

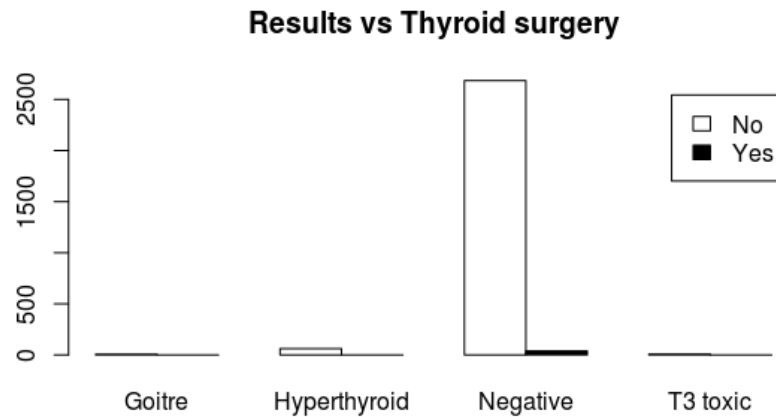


Figura 3-3: Resultado del diagnóstico y la existencia o no de una cirugía a la tiroides

El bocio corresponde al desarrollo descontrolado de la glándula tiroides y puede estar asociado al hipertiroidismo. Sin embargo, como se observa en el grupo Negative hay una pequeña porción que no está asociado a esta enfermedad y por lo tanto los exámenes de sangre para detectar los niveles hormonales se vuelve necesarios. A demás, esta idea se ve apoyada, porque se observa que existe una porción de las personas con hipertiroidismo que no poseen bocio.

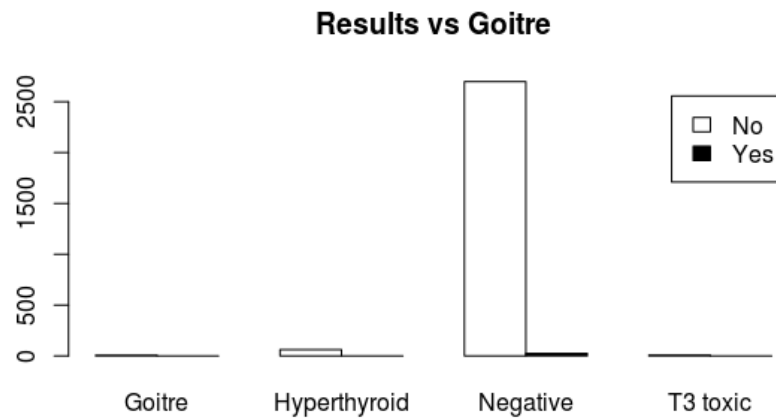


Figura 3-4: Relación entre Goitre y los resultados

El tratamiento con el isótopo de yodo es utilizada para atacar el tejido canceroso de la tiroides, ya que este absorbe el yodo radioactivo y termina destruyendo a las células anómalas. Sin embargo, al ser un tratamiento que destruye tejido intuitivamente se puede pensar que una sobre actividad de la glándula es paleada, cuestión que se ve apoyo con la visualización del gráfico en el que la pequeña porción de paciente bajo este tratamiento se ubica dentro del grupo que no posee la enfermedad.

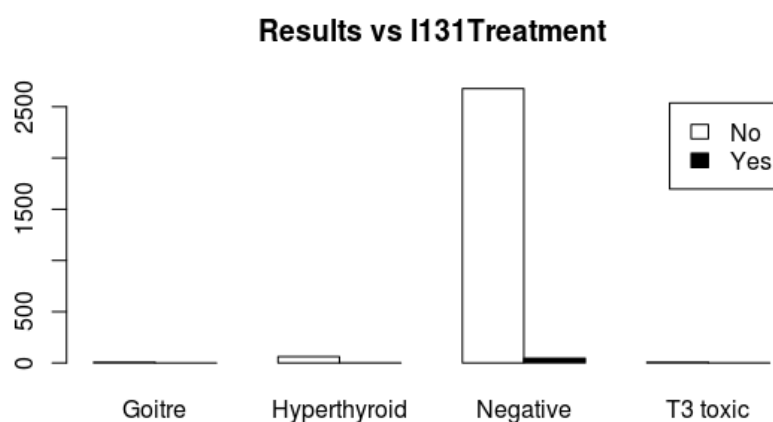


Figura 3-5: Relación entre los resultados del estudio y el tratamiento con isótopo I31 de yodo

3.2 VARIABLES NUMÉRICAS

A continuación se muestran los histogramas resultantes de los datos numéricos presentados en el dataset. Estos no incluyen la columna TBG, ya que esta contenía exclusivamente valores nulos y por lo tanto, se puede llegar a dos ideas sobre esta situación uno es que no es una variable relevante para el estudio o que solo en caso muy singulares es necesario realizar ese examen para poder realizar el diagnóstico con precisión, ya que esta proteína es la más importante en el proceso de fijación de hormonas T3 y T4.

En la figura 3-6, se observa como la edad de los pacientes se acumula en la tercera edad y, por lo tanto, se puede pensar que existe una desregulación hormonal dada los problemas naturales de la edad.

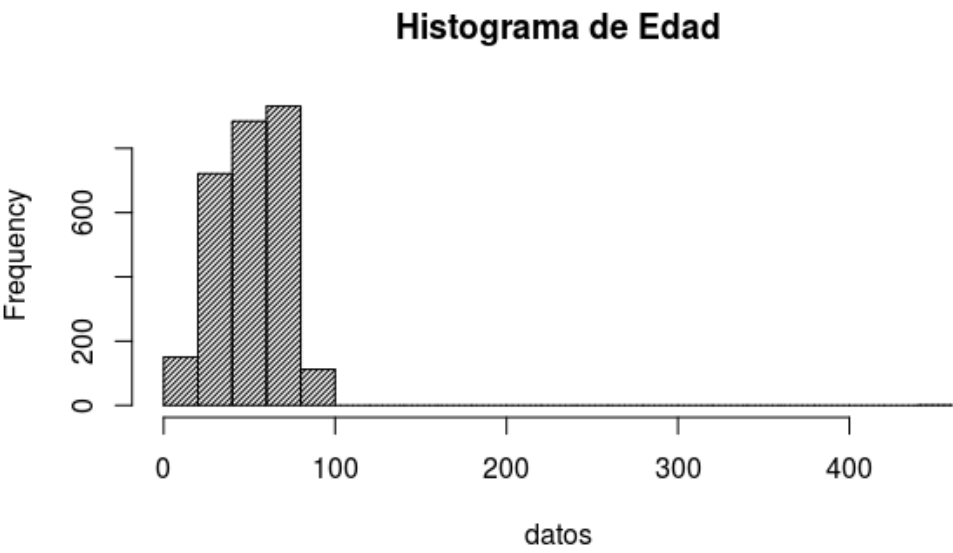


Figura 3-6: Histograma de edad

Este exámen hace relación con los níveles de libertad de TT4

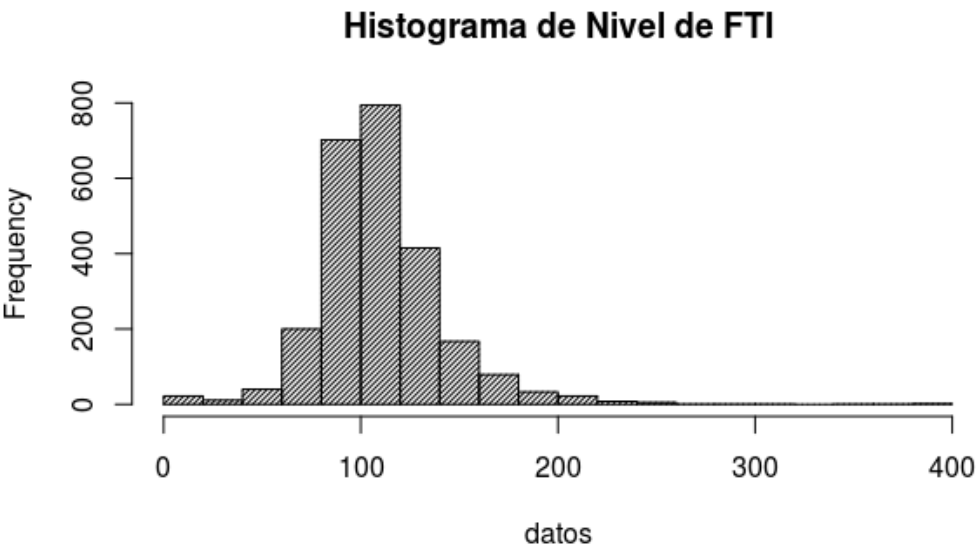


Figura 3-7: Histograma de FTI

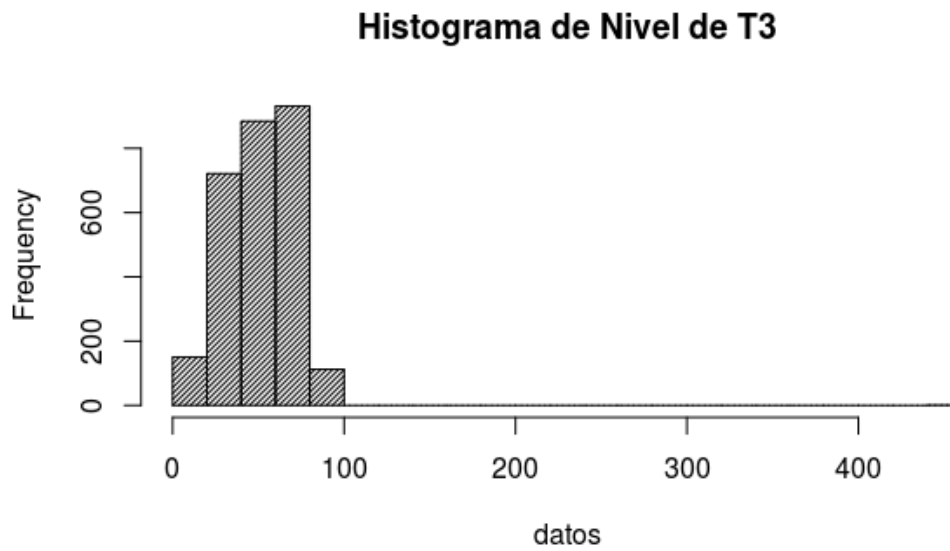


Figura 3-8: Histograma de T3

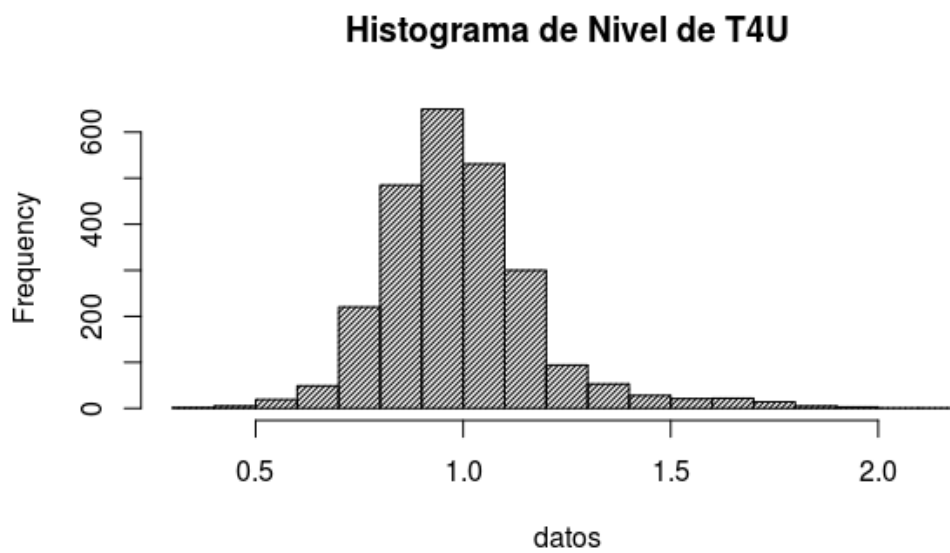


Figura 3-9: Histograma de T4U

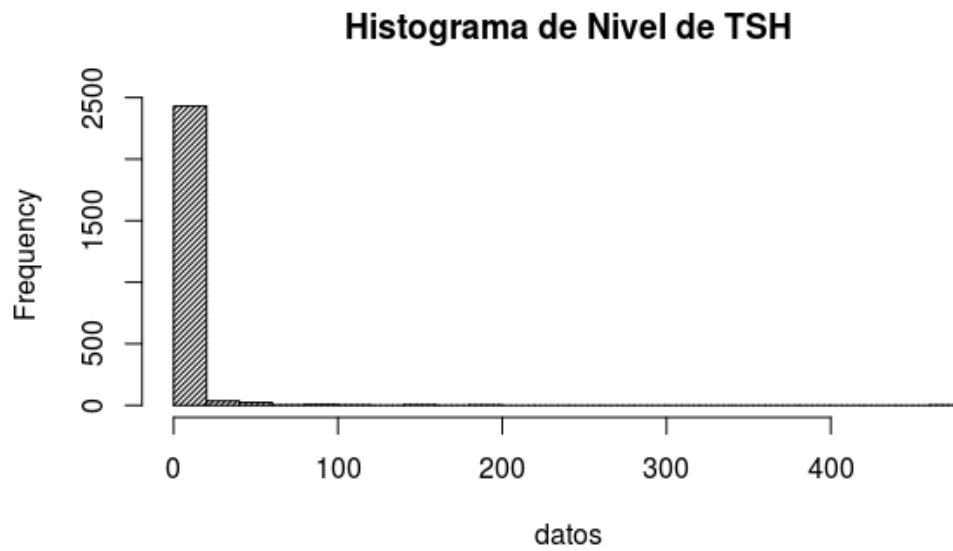


Figura 3-10: Histograma de TSH

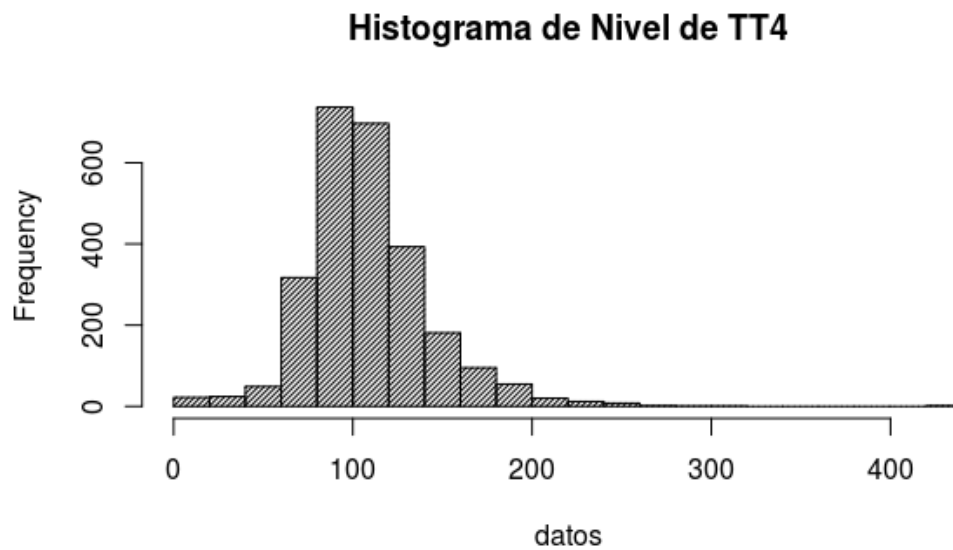


Figura 3-11: Histograma de TT4

Finalmente, se entrega una tabla que resume los valores más importantes, entre ellos medidas de tendencia central y medidas de posición.

	Min	1st Qu	Median	Mean	3rd Qu.	Máx.	Na
Edad	1.00	36.00	54.00	51.84	67.00	455.00	1
Nivel de TSH	0.005	0.440	1.400	4.672	2.600	478.000	284
Nivel de T3	0.050	1.600	2.000	2.025	2.400	10.600	585
Nivel de TT4	2.0	88.0	104.0	109.1	125.0	430.0	184
Nivel de T4U	0.3100	0.8800	0.9800	0.9979	1.0800	2.1200	297
Nivel de FTI	2.0	93.0	107.0	110.8	124.0	395.0	295

Figura 3-12: Tabla Sumario de los datos numéricos utilizados

Como se observa en ella existe una gran cantidad de valores para los cuales no existe un valor, entonces existe una complicación referente a cómo realizar los test estadísticos, ya que dependiendo de si se reconstruyen o no estos datos es seguro que los supuestos que permiten la aplicación de teoremas y por consiguiente, la aplicación de test estadísticos de un gran poder variarán notablemente.

Por otro lado, aquí la opción de reconstrucción es arriesgada, ya que como no expertos en el área la distribución de estas o el valor que tomen será cuestionable y por tanto el conocimiento extraído de ellas es inútil, la única opción es utilizar técnicas de imputación de valores, que se escapan a los alcances de la experiencia.

Para los test de normalidad de utilizó el test de kolmogorov-smirnov, sin embargo, en todas las pruebas se obtuvo un valor $p = 2.2 \times 10^{-16}$, por lo tanto la posibilidad de observar los datos dada una distribución normal es casi nula y por lo tanto no se pueden aplicar los test que se basan en el teorema del límite central.

Por otro lado, el análisis que realmente puede aportar información a la caracterización de cada uno de los grupos resultantes, corresponde al análisis de componentes, sin embargo esa técnica se escapa a los alcances de esta actividad.

CAPÍTULO 4. CONCLUSIÓN

A la luz de los resultados obtenidos se observa como el hipertiroidismo está relacionado una serie de secreciones hormonales que son regulados en un circuito cerrado. Sin embargo, el problema de determinar las causas de esta enfermedad vienen dadas por la múltiples variables las cuáles nos on necesariamente numéricos y por lo tanto el proceso de inferencia estadística cambia totalmente.

Por otro lado, desde el punto de vista de los conocimiento tratados en el ramo se puede ver que una forma de abordar el problema puede ser a través del análisis de agrupamiento, de modo tal de encontrar las distancias que se puedan tratar a las variables binarias a través de las distancias.

También, es importante agregar el punto sobre la normalidad de las variables numéricas que fueron abordadas, ya que si bien el resultado fue negativo en ese caso y por tanto el poder de cualquier prueba de estadística aplicada a ella se ve disminuido, se puede ver explicado por la eliminación de los valores desconocidos, y por lo tanto, a través de algún proceso de imputación o de reconstrucción de estos datos que son perdidos, la distribución pueda variar y lleve el resultado del estudio a otros horizontes.

Otro problema interesante, en este primer acercamiento al dataset, es que hay muchas variables que necesitan ser observadas desde el punto de vista acabado de la biología, de modo tal que se lleven a cabo las interpretación correctas de cada una de ellas y no solo de manera aisladas, sino que también el modo en que ellas interfieren o regulan el comportamiento de otras, como por ejemplo la T4U y la TBG, donde la T4U se entiende como la hormona libre que realmente causará efectos en el cuerpo y la TBG la proteína fijadora que permitirá el proceso de regulación metabólica a nivel celular.

Desde otro punto de vista también es importante destacar la necesidad de aplicar un análisis de componentes principales, para determinar los mayores niveles de variabilidad y por tanto la cantidad de información que cada uno de los exámenes que se realizan en los pacientes cobran más importancia de modo tal que el diagnosticar la enfermedad de vuelva menos tedioso e invasivo para el paciente en caso de que existan falsos positivos que necesiten la repetición de los exámenes.

Finalmente, entre los problemas que de zanjarlos aportarán a un mejor entendimiento de la problemática y por ende a obtener conocimientos más acabados corresponde una comprensión de las unidades de medida de las variables que nacen producto de un examen médico, ya que esta información es compleja de manejar como estudiantes que están fuera del nicho de la biología.

CAPÍTULO 5. REFERENCIAS

[1] *Sistema endocrino*. Recuperado: 03 de Septiembre de 2016, desde <http://med.unne.edu.ar/enfermeria/catedras/fisio/sistema%20endocrino.pdf>

[2] Brent Wisse. *Hipertiroidismo*. Recuperado: 03 de Septiembre de 2016, desde: <https://medlineplus.gov/spanish/ency/article/000356.htm>

[3] Radiology Info.org (17 de Marzo 2016), *Radioactive Iodine (I-131) Therapy*. Recuperado: 03 de Septiembre de 2016, desde: <http://www.radiologyinfo.org/en/info.cfm?pg=radioiodine>

[4] American Society of Health-System Pharmacists(4 de Marzo de 2014). *Litio* Recuperado: 03 de Septiembre de 2016, desde: <https://medlineplus.gov/spanish/druginfo/meds/a681039-es.html>

[5] Boewn, R. *Mechanism of Action and Physiologic Effects of Thyroid Hormones*(24 de Julio de 2010). Recuperado: 03 de Septiembre de 2016, desde: <http://www.vivo.colostate.edu/hbooks/pathphys/endocrine/thyroid/physio.html>

[6] American thyroid association. *Thyroid function test*. Recuperado: 03 de Septiembre de 2016, desde: <http://www.thyroid.org/thyroid-function-tests/>

[7] *Vitamina K*. Recuperado: 03 de Septiembre de 2016, desde: <http://m.exam-10.com/medicina/21331/index.html>