

UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INFORMÁTICA



Laboratorio 1 - Análisis estadístico

Integrantes: Gonzalo Cuevas.
Sofia Castro .
Curso: Análisis de Datos.
Sección A-1.
Profesor: Dr. Max Chacón Pacheco.

7 de Septiembre de 2022

Tabla de contenidos

1. Introducción	1
1.1. Objetivos	1
2. Descripción del Problema	2
2.1. Descripción de la base de datos.	3
2.2. Descripción de clases y variables.	3
3. Análisis Estadístico e Inferencial	10
3.1. Estadística Descriptiva	10
3.1.1. Variables Cuantitativas	10
3.1.2. Variables Cualitativas	13
3.2. Estadística Inferencial	15
3.2.1. Correlación	15
3.2.2. Métodos Estadísticos	17
4. Conclusiones	22
Bibliografía	24

Índice de figuras

1.	Ubicación de la glándula tiroidea, Stock (2020).	2
2.	Correlación entre variables numéricas.	16
3.	Correlación N°1 entre variables categóricas.	17
4.	Correlación N°2 entre variables categóricas.	17

Índice de cuadros

1.	Tipo de atributos.	3
1.	Tipo de atributos.	4
1.	Tipo de atributos.	5
2.	Descripción de atributos.	5
2.	Descripción de atributos.	6
2.	Descripción de atributos.	7
2.	Descripción de atributos.	8
2.	Descripción de atributos.	9
3.	Valores de centralización y dispersión para T3.	11
4.	Valores de centralización y dispersión para T4.	12
5.	Valores de centralización y dispersión para TSH	12
6.	Tabla de contingencia para las variables Sex y Goitre.	14
7.	Tabla de contingencia para las variables 'On Thyroxine' y los tipos de diagnósticos.	14
8.	Tabla de contingencia para las variables Sick y los tipos de diagnósticos.	15

1. Introducción

Las técnicas de estadística descriptiva e inferencial permiten la obtención, tratamiento y la interpretación de un conjunto de datos de observación, actuando como un puente entre los modelos matemáticos y los fenómenos reales según Borrego (2008).

En el presente informe se lleva a cabo la implementación de las técnicas mencionadas en la base de datos *allhypo*, motivado por la indagación de posibles relaciones mediante resultados coherentes.

El documento consta de una breve descripción al problema planteado, la relación que posee con las variables pertenecientes a la base de datos, para luego explicar el significado de cada una, permitiendo tener un conocimiento previo de biología, con el fin de poder implementar las técnicas de estadísticas, razonando de manera adecuada los resultados obtenidos mediante análisis o inferencia. Posteriormente, se presentan las conclusiones respecto al problema y las complicaciones que se presentaron durante el desarrollo del trabajo.

1.1. Objetivos

Los objetivos de este laboratorio consideran lo siguiente:

1. Estudiar e interpretar los datos correspondientes a la base de datos.
2. Explicar el significado de clases, atributos y sus valores.
3. Implementar métodos de estadística descriptiva e inferencial al problema.

2. Descripción del Problema

En el siguiente capítulo se expone el problema planteado y la relación existente entre la realidad y el contenido de la base de datos utilizada para la investigación, al igual que el origen y definición de sus atributos de esta. Cabe destacar que su información se relaciona con la afectación *hipotiroidismo*.

La glándula tiroidea es un órgano endocrino localizado en la parte anterior e inferior del cuello, por delante de la tráquea cervical inmediatamente (Figura 1), su función es la síntesis de las hormonas tiroideas que controlan el metabolismo.

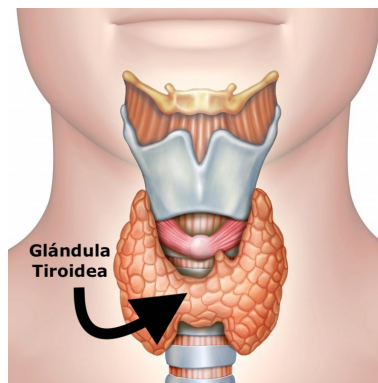


Figura 1: Ubicación de la glándula tiroidea, Stock (2020).

Las hormonas tiroideas llevan varios átomos de yodo en su composición, tres la triyodotironina (T3), y una más la tetrayodotironina (T4). Las enfermedades mas conocidas a causa de las hormonas tiroideas son: hipotiroidismo e hipertiroidismo, donde la primera consiste en la desaprovechamiento de el yodo por parte de la glándula, por otro lado, el hipertiroidismo ocurre cuando la glándula excede su capacidad de captación, siendo sometida a

cargas de yodo más altas. Algunos efectos adicionales de la hormona T3 y T4 recaen en: crecimiento y desarrollo, elevación del gasto cardiaco, aumento de la síntesis y degradación de las proteínas, intervienen en la síntesis el glucógeno y en la utilización de la glucosa, imprescindibles para el desarrollo del sistema nervioso, central y periférico, entre otras acciones.

Por lo tanto, las hormonas tiroideas son esenciales para el metabolismo de nuestro cuerpo al intervenir en la mayoría de las funciones orgánicas.

2.1. Descripción de la base de datos.

Consiste en una base de datos multivariable, creada por Ross Quinlan, perteneciente al instituto Garavan de Sydney, Australia. En específico, la base de datos allhypo posee 29 atributos, ya sea valores numéricos, booleanos y categóricos, además de atributos con valores desconocidos representados por el símbolo ?, también contiene 2800 instancias de cada set. La información es obtenida por Dua and Graff (2017).

2.2. Descripción de clases y variables.

A fin de entender la base de datos entregada, se ha dispuesto en el **cuadro 1** el tipo de dato de cada variable y en el **cuadro 2** la descripción o definición de la variable.

Variables	Tipo (Rango)
Age	Real (2-94)

Cuadro 1: Tipo de atributos.

Variables	Tipo (Rango)
Sex	Booleano (M ó F)
On Thyroxine	Booleano (f ó t)
Query on Thyroxine	Booleano (f ó t)
On antithyroid medication	Booleano (f ó t)
Sick	Booleano (f ó t)
Pregnant	Booleano (f ó t)
Thyroid Surgery	Booleano (f ó t)
I131 treatment	Booleano (f ó t)
Query hypothyroid	Booleano (f ó t)
Query hyperthyroid	Booleano (f ó t)
Lithium	Booleano (f ó t)
Goitre	Booleano (f ó t)
Tumor	Booleano (f ó t)
Hypopituitary	Booleano (f ó t)
Psych	Booleano (f ó t)
TSH measured	Booleano (f ó t)
TSH	Real (2-264)
T3 measured	Booleano (f ó t)
T3	Real (2-65)
TT4 measured	Booleano (f ó t)
TT4	Real (2-218)
T4U measured	Booleano (f ó t)

Cuadro 1: Tipo de atributos.

Variables	Tipo (Rango)
T4U	Real (2-139)
FTI measured	Booleano (f ó t)
FTI	Real (2-210)
TBG measured	Booleano (f ó t)
TBG	Real (NA)
Referral source	STMW, SVHC, SVHD, SVI, otro

Cuadro 1: Tipo de atributos.

Variables	Descripción
Age	Tiempo que ha vivido una persona o ciertos animales o vegetales R.A.E (2022b).
Sex	Condición orgánica, masculina o femenina, de los animales y las plantas R.A.E (2022e).

Cuadro 2: Descripción de atributos.

Variables	Descripción
On Thyroxine	La tiroxina (T4), es una hormona producida por la glándula tiroidea. Las personas que padecen Hipotiroidismo poseen insuficiencia de esta hormona, por lo que requieren de un tratamiento hormonal (T4) para suplir su falta Association (2022).
Query on Thyroxine	Consulta del tratamiento de tiroxina para el hipotiroidismo.
On antithyroid medication	La medicación de antitiroideos es requerida por personas con hipertiroidismo, pues tiene la capacidad de equilibrar sus hormonas Hengameh Abdi (2019).
Sick	Que padece enfermedad R.A.E (2022d).
Pregnant	Estado en que se halla la mujer gestante R.A.E (2022c).
Thyroid Surgery	Esta cirugía consiste en la remoción total o parcial de la glándula tiroidea, consecuencia de un cáncer, hipertiroidismo y/o nódulos ASSOCIATION (2022).

Cuadro 2: Descripción de atributos.

Variables	Descripción
I131 treatment	Tratamiento médico nuclear para el hipertiroidismo y/o el cáncer de tiroides. Consiste en destruir suficiente tejido de la tiroides, obteniendo un paciente eutiroides o hipotiroideo Malik Mumtaz (2009).
Query hypothyroid	Consulta de hipotiroidismo, condición que se produce cuando las hormonas producidas por la tiroides son inferiores a lo requerido por el cuerpo Shahid et al. (2019a).
Query hyperthyroid	Consulta de hipertiroidismo, condición que se produce cuando las hormonas producidas por la tiroides son mas de las que requiere el cuerpo Shahid et al. (2019a)
Lithium	Tratamiento utilizado para inhibir la liberación de la hormona tiroidea, previniendo el desarrollo de hipertiroidismo Bocchetta and Loviselli (2006).
Goitre	Aumento, difuso o nodular, de la glándula tiroidea. R.A.E (2022a)

Cuadro 2: Descripción de atributos.

Variables	Descripción
Tumor	Masa de células transformadas, con crecimiento y multiplicación anormales R.A.E (2022f).
Hypopituitary	Síndrome clínico resultante de la deficiencia de una o varias hormonas hipofisarias. Pascual-Corrales et al. (2020)
Psych	Posibilidad de disfunción psicológica causada por hipotiroidismo o hipertiroidismo Whybrow P.C. (1969).
TSH	Hormona estimulante de la tiroides, controla la cantidad de T4 y T3 en la sangre Shahid et al. (2019a).
T3	Triyodotironina, hormona producida por la glándula de la tiroides, fundamental en la conversión de los nutrientes en energía y regular la temperatura del cuerpo Shahid et al. (2019a).

Cuadro 2: Descripción de atributos.

Variables	Descripción
TT4	Concentración total de la hormona tiroxina en el paciente, tanto de la T4 ligada y T4 libre Shahid et al. (2019b).
T4U	Capacidad total de unión de la hormona tiroxina a las proteínas (T4 ligada), impidiendo el ingreso a los tejidos del cuerpo Shahid et al. (2019b).
FTI	Índice de T4 libre, utilizado para determinar disfunciones en la glándula tiroidea Shahid et al. (2019b).
TBG	Globulina fijadora de tiroxina, proteína transportadora de hormonas tiroideas en el humano, es sintetizada por el hígado y secretada al torrente sanguíneo Refetoff et al. (1996).

Cuadro 2: Descripción de atributos.

3. Análisis Estadístico e Inferencial

Luego de comprender la base de datos se implementa el análisis de los resultados obtenidos, siendo visualizado en este apartado.

3.1. Estadística Descriptiva

Los métodos empleados pertenecientes a las técnicas de estadística descriptiva corresponden a la medidas de centralización (media y mediana), medidas de dispersión, tales como varianza y desviación estándar. Además se proyectan gráficos con los resultados obtenidos, facilitando la comprensión de los datos.

3.1.1. Variables Cuantitativas

Como variables cuantitativas, son consideradas aquellas que describen el posible comportamiento que pueden tener las hormonas, proteínas y la edad. Asimismo, son identificadas cuatro clases que contemplan el tipo de enfermedad existente en la población, donde estas corresponden a las personas con Hipotiroidismo negativo, compensado, primario y secundario. La cantidad de personas para cada clase son 2800, 2580, 154, 64 y 2 respectivamente.

Respecto a la hormona T3 (Cuadro 3), es posible inferir el rango de valores que debe poseer para pertenecer a cierta clase, por ejemplo, si una persona no posee hipotiroidismos su nivel de T3 debe ser $T3 \geq 2,07$, mientras que para padecer hipotiroidismo compensado corresponde a $1,74 \geq T3 > 1,004$, por otro lado, si el nivel paciente pertenece a $1,850 \geq T3 >$

1,74 significa que padece de hipotiroidismo secundario, finalmente, adolece de hipotiroidismo primario si su valor de T3 es menor a 1.004. Se debe destacar que el promedio de la clase Negativa es acorde a lo esperado, ya que se necesita un valor alto de T3 para obtener un resultado negativo de hipotiroidismo.

Clase	Media	Mediana	Desviación Estándar
Población	2.025	2.000	0.8246004
Negative	2.07	2.00	0.8137529
Compensated	1.74	1.75	0.7134667
Primary	1.004	0.850	0.6542072
Secondary	1.850	1.850	0.9192388

Cuadro 3: Valores de centralización y dispersión para T3.

A partir de los resultados obtenidos en el cuadro 4, se pueden deducir los posibles valores que debe considerar la hormona T4 para padecer o no hipotiroidismo. Si se debe comprobar el padecimiento de hipotiroidismo compensado, primario o secundario el nivel de T4 debe ser: $93,82 \geq T4 > 35,88$; $35,88 \geq T4 > 54$ y $35 \leq T4$ respectivamente, por otra parte, si se obtiene un resultado negativo de hipotiroidismo, esto significa que $112 \geq T4$. Para la clase Negativa, ocurre el mismo efecto que en el cuadro 3, donde esta posee el menor valor en promedio de T4, siendo correcto para no padecer hipotiroidismo.

Clase	Media	Mediana	Desviación Estándar
Población	109.1	104.0	35.39244
Negative	112	106	34.04901
Compensated	93.82	89.00	21.27994
Primary	35.88	37.00	22.85336
Secondary	35.0	35.0	18.38478

Cuadro 4: Valores de centralización y dispersión para T4.

Finalmente, a diferencia del cuadro 3 y 4, en el cuadro 5 las personas con resultados negativos de hipotiroidismo poseen el menor promedio equivalente a $1,8626 \leq TSH$, afirmando los efectos que posee la hormona TSH con los niveles de T4 y T3, pues al disminuir su valor, aumenta los valores de T4 y T3 (relación inversa), recayendo en un posible diagnóstico de hipertiroidismo al paciente. El nivel de TSH que debe tener un paciente para padecer hipotiroidismo compensado, primario o secundario corresponde a $13,846 \geq TSH > 2,308$; $83,46 \geq TSH$ y $2,308 \geq TSH > 1,8626$ respectivamente.

Clase	Media	Mediana	Desviación Estándar
Población	4.672	1.400	21.44945
Negative	1.8626	1.200	5.009411
Compensated	13.846	9.850	14.1932
Primary	83.46	45.50	100.8121
Secondary	2.308	2.308	3.242085

Cuadro 5: Valores de centralización y dispersión para TSH .

Cabe considerar que la muestra con personas afectadas por hipotiroidismo secundario, primario y compensado son ínfimas y desproporcionadas entre ellas en comparación con la muestra de hipotiroidismo negativo, causando que esta influya negativamente en las inferencias y deducciones obtenidas de los resultados, pues no se provee de una fuente de dato diversa e imprecisa al no representar de manera adecuada una población con la afección mencionada. Por el contrario, los resultados de hipotiroidismo negativo se consideran los más precisos, ya que su muestra es la mas grande, recayendo en una mejor representación de esta población al igual que las deducciones efectuadas a partir de los resultados.

3.1.2. Variables Cualitativas

La base de datos contiene un total de 24 variables cualitativas o categóricas, las cuales en su mayoría son dicotómicas compuestas por valores booleanos. Para poder evaluar este tipo de dato, se han implementado tablas de contingencia entre algunas de las variables con el fin de visualizar posibles relaciones entre las mismas.

Los resultados del cuadro 6 permiten contemplar las variables sexo y bocio, de las cuales se puede asumir que no existe una relación directa entre ellas, si bien las mujeres poseen mayor probabilidad de desarrollar Bocio, no es un resultado definitivo, pues es muy desproporcional en comparación de la totalidad de mujeres en la población.

-	FALSE	TRUE
FEMALE	1797	33
MALE	854	6

Cuadro 6: Tabla de contingencia para las variables Sex y Goitre.

La tabla de contingencia en el cuadro 7 contiene los resultados para las personas que se encuentran consumiendo T4 y la clase de hipotiroidismo que poseen. Los valores obtenidos resultan coherentes, pues las personas que obtuvieron negativo no deben estar en un tratamiento de T4, así como también quienes padecen de hipotiroidismo compensado tampoco deben consumir dicha hormona, pues su nivel de T4 es alto, visualizado en el cuadro 4.

-	Compensated	Negative	Primary	Secondary
FALSE	154	2255	59	2
TRUE	0	0	5	0

Cuadro 7: Tabla de contingencia para las variables 'On Thyroxine' y los tipos de diagnósticos.

El cuadro 8, involucra la clase de hipotiroidismo y la variable enfermedad, sin embargo, las relaciones obtenidas pueden ser consideradas imprecisas, en especial con la clase negativo, pues se debe inferir que verdadero significa el padecimiento de otra enfermedad no relacionada con hipotiroidismo, lo que recae en la sensibilidad de los resultados al tener que asociar esta variable con una enfermedad específica. También se considera incoherente que

la variable enfermedad sea falsa cuando padece hipotiroidismo compensado, primario o secundario, pues estos se consideran enfermedades. Por lo tanto, se puede deducir que la variable enfermedad representa a otra afectación sin considerar el hipotiroidismo.

-	Compensated	Negative	Primary	Secondary
FALSE	143	2481	64	2
TRUE	11	99	0	0

Cuadro 8: Tabla de contingencia para las variables Sick y los tipos de diagnósticos.

3.2. Estadística Inferencial

Para este apartado, se han implementado diversos muestreos de la población descrita en la base de datos, estableciendo hipótesis nulas y alternativas junto con sus veracidades, y el uso de probabilidades por medio de las proporciones de variables. Para las variables numéricas se mide la correlación con Pearson mientras que, se utiliza V de Cramér para las variables categóricas.

3.2.1. Correlación

La correlación de Pearson entrega rangos entre $[-1,1]$, mientras más negativo es el resultado mayor distancia hay entre las variables, por lo que son inversamente proporcionales, por el contrario, si el valor es positivo se considera directamente proporcional. También, se puede interpretar mediante los extremos de sus valores, mientras más cerca de -1 o 1, mayor es

su relación y si es equivalente a 0, significa que las variables son independientes. Sin embargo, un valor distinto a 0 no siempre significa que existe una relación causa y efecto entre las variables, pues ambas variables pueden ser afectadas por una tercera variable, lo cual no puede ser evaluado, pues la prueba permite comparación entre pares de variables Soetewey (2020).

En la figura 2 se permite comprender en la fila 2 que todas exceptuando a T4U de las variables son inversamente proporcional a la hormona TSH, en cambio la TT4 con las variables FTI, T3 y T4U tienen una relación directa, lo que es esperado dado el conocimiento previo en el análisis de estadística descriptiva.

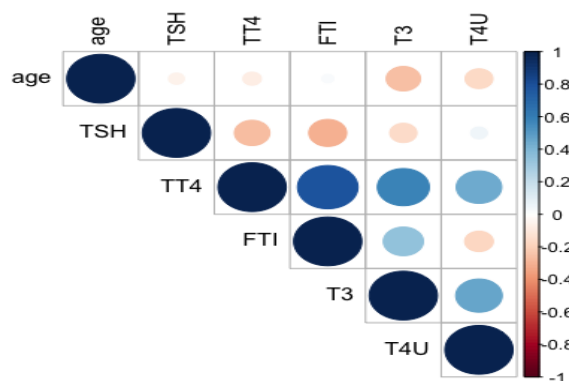


Figura 2: Correlación entre variables numéricas.

La correlación V de Cramér permite verificar si existe una asociación entre variables nominales, donde sus valores poseen un rango de $[0,1]$, con 0 equivalente a una inexistente asociación y 1 significa una asociación fuerte entre las variables Soetewey (2020).

Los colores mas claros corresponden a una asociación débil entre las variables, para la figura 3 y 4. Se puede visualizar que predomina el valor

0, lo cual significa que entre las variables categóricas no existen relaciones, esto puede deberse a los valores ? o NA en la base de datos, influyendo en los resultados negativos.

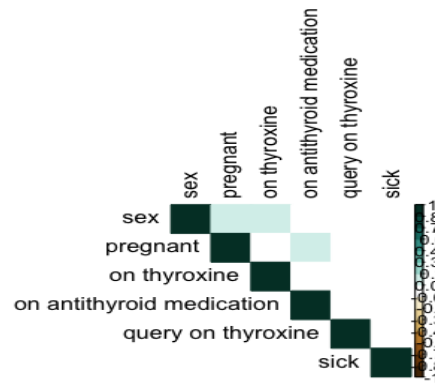


Figura 3: Correlación N°1 entre variables categóricas.

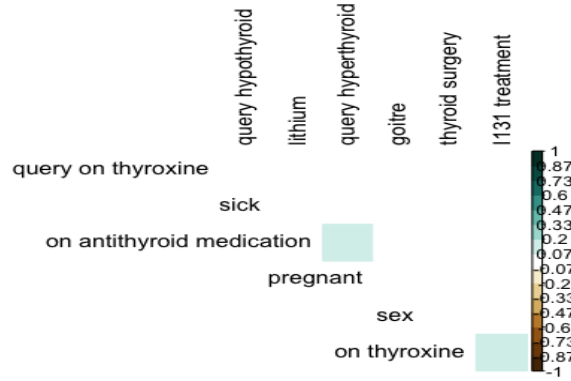


Figura 4: Correlación N°2 entre variables categóricas.

3.2.2. Métodos Estadísticos

Para analizar los datos, estos se han filtrado a través de la selección de variables categóricas de interés y el uso de variables numéricas

con el fin de analizar sus valores promedios. En términos generales, el procedimiento para cada inferencia fue similar, en primer lugar se genera una muestra de valores aleatorios que represente a la población con el objetivo de que estos sigan una distribución normal, sin embargo en varias ocasiones existe la presencia así, mediante el uso de una prueba t obtener un p-value que permita verificar la veracidad de las hipótesis descritas.

En términos generales, el procedimiento utilizado para cada inferencia fue similar, primero es generada una muestra de valores aleatorios que represente la población y siga una distribución normal, sin embargo, debido a la existencia de valores atípicos, todas las pruebas fueron aplicadas para muestras pequeñas utilizando la prueba t con un nivel de significación de 0.05.

Una de las variables con mayor peso que afectan a las personas con hipotiroidismo es el nivel de TSH en su torrente sanguíneo, pues esta se encarga de la producción de hormonas tiroideas.

Por otro lado, tratamientos como el yodo radiactivo o de litio son utilizados para disminuir los niveles de T4 en las personas con hipotiroidismo, por lo que el uso de estos debería afectar los niveles de TSH.

Para las personas que han recibido tratamiento de yodo radiactivo, se esperaría que en promedio, presentaran mayores niveles de TSH en comparación de quienes no han recibido el tratamiento. Las variables en juego y sus hipótesis de estudio serían: $H_0 : \mu_A = \mu_B$, $H_A : \mu_A > \mu_B$; μ_A : TSH promedio de quienes recibieron el tratamiento, μ_B : TSH promedio de quienes no recibieron el tratamiento.

A partir de la prueba, se obtuvo un p-value de 0.3624, indicando

que se falla al rechazar la hipótesis nula, por lo que se podría considerar con un 99 de confianza que el tratamiento no genera variación notable de TSH en los individuos.

Del mismo modo, se analizó el nivel promedio de TSH en los individuos que recibieron tratamiento de litio, obteniendo un resultado similar en el que no se presenta variación en esta hormona.

Un factor que afecta directamente a la producción de TSH es la cirugía realizada para contrarrestar el hipertiroidismo, pues esta genera una desestabilización en la producción de hormonas, causando que se disminuyan los niveles de T4 en la sangre. Con esto, se esperaría que los individuos que recibieron esta cirugía presenten altos niveles de TSH pues esto implicaría que se genera una deficiencia de T4 en comparación de quienes no han sido operados. Así, las variables en juego y las hipótesis a contrastar son: $H_0 : uA = uB$, $H_A : uA > uB$; uA: TSH promedio de quienes fueron operados de la tiroides. uB: TSH promedio de quienes no fueron operados de la tiroides.

En este caso se obtuvo un p-value de 0.02 por lo que se acepta la hipótesis alternativa, lo que nos permite inferir con un 95 de confianza que aquellas personas que fueron operadas de la tiroides, presentan menores niveles de TSH y posiblemente sean individuos que desarrollen hipertiroidismo o contengan sus niveles de T4 estables.

Resulta de interés saber que ocurre con la hormona T4 cuando los individuos presentan ciertas características, ya que esto nos podría ayudar a determinar quienes son mas propensos a desarrollar hipotiroidismo o hipertiroidismo.

En las mujeres embarazadas, la producción de esta hormona resul-

ta fundamental para evitar complicaciones durante la gestación, en concreto se busca que sus niveles de TT4 se mantengan estables, sin embargo, no siempre es posible. Con esto, se ha propuesto verificar si el promedio de TT4 en las mujeres embarazadas difieren con el promedio en las no embarazadas. Haciendo uso de las siguientes hipótesis: $H_0 : \mu_A = \mu_B$, $H_A : \mu_A \neq \mu_B$; μ_A : TT4 promedio de las mujeres embarazadas. μ_B : TT4 promedio de las mujeres no embarazadas.

La prueba entrega un p-value de 1.213e-05, por lo que se concluye que los niveles de TT4 son diferentes para ambos casos y de hecho, la media para las mujeres no embarazadas es inferior a la media de las embarazadas, por lo que estas ultimas estarían predispuestas a desarrollar hipertiroidismo en lugar de hipotiroidismo.

A su vez, resulta de interés saber que ocurre con los niveles de TT4 en las personas con tumores y bocio, pues ambas afecciones tienen una estrecha relación con las variaciones en las hormonas de t4. Con esto, se busca determinar si para alguno de los dos casos, los niveles de TT4 son superiores. las siguientes hipótesis y variables a contrastar son: $H_0 : \mu_A = \mu_B$, $H_A : \mu_A > \mu_B$; μ_A : TT4 promedio de quienes presentan tumores. μ_B : TT4 promedio de quienes presentan bocio.

Se obtiene un p- value de 0.0005, por lo que se acepta la hipótesis alternativa, y en concreto , los individuos con bocio presentan menores niveles de tt4, por lo que estos podrían ser mas propensos a desarrollar hipotiroidismo en lugar de aquellos que presentan tumores.

Analizar la cantidad de FTI circulante también permite determinar si una persona puede sufrir hipotiroidismo en especial si estas se encuen-

tran consumiendo algún medicamento asociado a esto. Con ello, las hipótesis y variables a contrastar son: $H_0 : uA = uB$, $H_A : uA < uB$; uA: FTI promedio de quienes consumen antitiroideos y no presentan hipotiroidismo. uB: FTI promedio de quienes no consumen antitiroideos y no presentan hipotiroidismo.

Las personas que no presentan hipotiroidismo y consumen medicamentos antitiroideos presentan menores niveles de FTI, lo cual puede indicar que el tratamiento utilizando podría llegar a generar en algún momento esta afección.

Es posible inferir sobre los niveles de FTI de las personas que presentan hipotiroidismo primario y compensado. Se esperaría que uno de los dos presentara deficiencias en los niveles de FTI, en concreto, el hipotiroidismo primario debiese presentar menores niveles de FTI al estar directamente relacionado con las variaciones en TSH y su estado elevado. mediante las siguientes hipótesis: $H_0 : uA = uB$, $H_A : uA \neq uB$ uA: FTI promedio de las personas con hipotiroidismo primaria que no se encuentran medicadas . uB: FTI promedio de quienes con hipotiroidismo compensada que no se encuentran medicadas.

Prueba t obtuvo un p-value de 1.794e-11, por lo que se acepta la hipótesis alternativa, analizando los valores de las medias de cada muestra, se concluye que las personas con hipotiroidismo primaria presentan un menor promedio de FTI

4. Conclusiones

A partir de la investigación, fue posible entender que variables juegan un papel fundamental a la hora de determinar como se desarrolla las enfermedades tiroideas.

Con esto, el desarrollo y análisis de la información, se logró identificar el comportamiento de las hormonas en una población de estudio con individuos con características específicas, y así inferir quienes son los individuos que presentan una mayor posibilidad de desarrollar alguna enfermedad tiroidea.

Las relaciones encontradas nos permitieron entender el efecto y causa de las hormonas en el hipotiroidismo, en especial las hormonas T3, T4 y TSH, que juegan un rol fundamental en la salud de las personas, dado que si sus valores varían de extremos el efecto será negativo para el humano.

Si bien los archivos asociados a la base de datos aportaba una idea de como funcionaban sus variables y una breve descripción de las mismas, esta no era suficiente, pues pese a la investigación exhaustiva realizada, en esta oportunidad no fue posible identificar las unidades de medida de las variables numéricas, además, con ciertas variables categóricas, identificar cual era su asociación y el comportamiento que generaba en los demás datos no estaba claro.

Adicionalmente, trabajar con los datos entregados conllevó un mayor número de complicaciones, pues existe una gran cantidad de variables con datos inconclusos o perdidos, lo que implicó trabajo extra al momento de calcular estadísticos con ellas.

Con esto, al examinar ciertos casos y eliminar aquellas instancias

en que las variables contaban con datos perdidos, se producía una reducción notable en los datos causando que la variabilidad se viera afectada y así inferir con respecto a ellas y sus posibles relaciones con las demás variables.

En este mismo ámbito, cierta elección de variables para filtrar los datos carecía de sentido, como fue en el caso de las clases, específicamente aquella que contempla al hipotiroidismo secundario, pues su población se encuentra formada solamente por dos personas, causando que al asociar estas personas con otra variable categórica no fuera útil para inferir. Lo mismo ocurre con las variables "psych", "hipotuituary" Y TBG, que solo presentaban una instancia diferente a las demás o simplemente todas contenían el mismo valor.

Finalmente, se espera en futuras evaluaciones poder contemplar la unidad de medida de las variables numéricas y adicionalmente inferir y modelar otras características que no fueron contempladas en esta instancia y que son conocidas por la comunidad científica, por lo que el objetivo se concentrara en hacer un descubrimiento entre las variables que pueda ayudar a las personas que padecen de hipotiroidismo o tienen secuelas de esta enfermedad.

-

Bibliografía

- Association, A. T. (2022). Thyroid hormone treatment. https://www.thyroid.org/wp-content/uploads/patients/brochures/HormoneTreatment_brochure.pdf.
- ASSOCIATION, A. T. (2022). Thyroid surgery. <https://www.thyroid.org/wp-content/uploads/patients/brochures/ThyroidSurgery.pdf>.
- Bocchetta, A. and Loviselli, A. (2006). Lithium treatment and thyroid abnormalities. *Clinical Practice and Epidemiology in Mental Health*, 2(1):1–5.
- Borrego, S. (2008). Estadística descriptiva e inferencial. https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero_13/SILVIA_BORREGO_2.pdf.
- Dua, D. and Graff, C. (2017). UCI machine learning repository.
- Hengameh Abdi, Atieh Amouzegar, F. A. (2019). Antithyroid drugs. 18(1):1–12.
- Malik Mumtaz, Lim Shueh Lin, K. C. H. A. S. M. K. (2009). Radioiodine i-131 for the therapy of graves disease. 16(1):25–33.
- Pascual-Corrales, E., Araujo-Castro, M., Ortiz-Flores, A., and Escobar-Morreale, H. (2020). Hipopituitarismo. *Medicine - Programa de Formación Médica Continuada Acreditado*, 13(15):856–864. Enfermedades endocrínicas y metabólicas (III) Enfermedades hipotalámicas e hipofisarias.
- R.A.E (2022a). Bocio. <https://dle.rae.es/bocio?m=form>.

R.A.E (2022b). Edad. <https://dle.rae.es/edad>.

R.A.E (2022c). Embarazo. <https://dle.rae.es/embarazo?m=form>.

R.A.E (2022d). Enfermo. <https://dle.rae.es/enfermo?m=form>.

R.A.E (2022e). Sexo. <https://dle.rae.es/sexo?m=form>.

R.A.E (2022f). Tumor. <https://dle.rae.es/tumor>.

Refetoff, S., Murata, Y., Mori, Y., Janssen, O. E., Takeda, K., and Hayashi, Y. (1996). Thyroxine-binding globulin: organization of the gene and variants. *Hormone Research in Paediatrics*, 45(3-5):128–138.

Shahid, A., Singh, M., Raj, R., Suman, R., Jawaid, D., and Alam, M. (2019a). A study on label tsh, t3, t4u, tt4, fti in hyperthyroidism and hypothyroidism using machine learning techniques. pages 930–933.

Shahid, A., Singh, M., Raj, R., Suman, R., Jawaid, D., and Alam, M. (2019b). A study on label tsh, t3, t4u, tt4, fti in hyperthyroidism and hypothyroidism using machine learning techniques. pages 930–933.

Soetewey, A. (2020). Correlation coefficient and correlation test in r. <https://statsandr.com/blog/correlation-coefficient-and-correlation-test-in-r/>.

Stock, A. (2020). Overview of the position of the thyroid gland in the neck. <https://teachmeanatomy.info/wp-content/uploads/Overview-of-the-Thyroid-Gland-768x1024.jpeg>.

Whybrow P.C., Prange A.J., J. T. C. (1969). Mental changes accompanying thyroid gland dysfunction: A reappraisal using objective psychological measurement. *Archives of General Psychiatry*, 20(1):48–63.