



UNIVERSIDAD PERUANA  
**CAYETANO HEREDIA**

**FACULTAD DE CIENCIAS E INGENIERÍA**

**FACTORES SOCIODEMOGRÁFICOS Y CLÍNICOS ASOCIADOS A HIPERTENSIÓN  
EN UNA POBLACIÓN DE LA COSTA DEL PERÚ: MODELO DE REGRESIÓN  
LOGÍSTICA**

**TESIS PARA OPTAR EL TÍTULO DE ESPECIALISTA EN ESTADÍSTICA EN  
INVESTIGACIÓN**

**AUTOR:**

**GINA FIORELLA LEON UNTIVEROS**

**ASESOR**

**Dra. LUZ AURORA CARBAJAL ARROYO**

**LIMA-PERÚ**

**2024**

# **JURADO**

**MG. DANIEL BLANCO VICTORIO  
PRESIDENTE**

**MG. NESTOR CARLOS FLORES RODRIGUEZ  
VOCAL**

**MG. VALERIA MARCELA PAZ APARICIO  
SECRETARIA**

## DEDICATORIA

Este trabajo se lo dedico al gran poderoso Señor de los cielos, señor de los ejércitos mi protector, mi roca firme y mi más fiel compañero, el que me da la fuerza y me cuida con la ternura de un padre perfecto. A mi gran poder superior que se lo debo todo.

A mi madre Emilia, ejemplo de perseverancia, cautela y trabajo, de nunca rendirse y siempre tener un objetivo en la vida, una maquina incansable de proyectos y sueños, una mujer que ama de verdad y se entrega a sus sueños, gracias mami por ser tan inteligente de poder comprender mis puntos de vista diferentes a los tuyos y sabia de poder guiarme y ser siempre ese motor.

A mi padre Telésforo, por inculcarme siempre ese deseo de explorar más allá “de la caja”, esa curiosidad por las cosas diversas, y el deporte, por haber sido mi cómplice cariñoso de lejos, por enseñarme el orden, la organización, el esfuerzo y lo que no se debe hacer, gracias a ti siempre busco ser una mejor persona, buscar aventuras y disfrutar más de la vida, gracias por siempre apoyar mis ideas y sentirte siempre tan orgulloso de mi.

A mi abuela Emilia, por ser mi madre en mis primeros años, por ser mi angelito que me cuida, por ser tan amorosa conmigo y haberme querido tanto, te extraño abuelita, y siempre recibo tus regalitos con mucho agradecimiento, gracias por criar a una mamá tan buena y por ser la mejor abuela que pude tener. Te admiro por haber permitido tanto, y sé que en el cielo estás sonriendo al verme, aunque no te esté haciendo mucho caso.

A mis compañeritos de vida Albi y Luna esos dos nobles y puros corazoncitos que me acompañan siempre.

## **AGRADECIMIENTO**

Mi más sincero agradecimiento a mi asesora de Tesis Luz Carbajal, por su paciencia y su habilidad nata para poder compartir sus conocimientos, por su paciencia y gran amabilidad.

A Dr. Hernando Torres Zevallos por su importante aporte en este trabajo ya que sin su ayuda desinteresada no hubiera sido posible el desarrollo del mismo.

A mis colegas y amigos que me ayudaron e incentivaron a terminar esta tesis que fue la que más tiempo me tomó en culminar, muchas gracias por apoyarme ya que solo se llega lejos, pero acompañado se llega mucho más lejos.

## FACTORES SOCIODEMOGRÁFICOS Y CLÍNICOS ASOCIADOS A HIPERTENSIÓN EN UNA POBLACIÓN DE LA COSTA DEL PERÚ: MODELO DE REGRESIÓN LOGÍSTICA

### INFORME DE ORIGINALIDAD

<b>20%</b>	<b>18%</b>	<b>7%</b>	<b>8%</b>
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

### FUENTES PRIMARIAS

<b>1</b>	<b>static.elsevier.es</b> Fuente de Internet	<b>1%</b>
<b>2</b>	<b>es.scribd.com</b> Fuente de Internet	<b>1%</b>
<b>3</b>	<b>dspace.esPOCH.edu.ec</b> Fuente de Internet	<b>1%</b>
<b>4</b>	<b>R. García Pérez, G. García Pino, D. González Ballester, R. García Moreno. "Modelo de regresión logística para estimar la dependencia según la escala de Lawton y Brody", SEMERGEN - Medicina de Familia, 2010</b> Publicación	<b>1%</b>
<b>5</b>	<b>eprints.uanl.mx</b> Fuente de Internet	<b>&lt;1%</b>
<b>6</b>	<b>www.coursehero.com</b> Fuente de Internet	<b>&lt;1%</b>

## ÍNDICE

Resumen.....	1
Abstract.....	2
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA.....	3
1.1 DETERMINACIÓN DEL PROBLEMA.....	3
1.2 FORMULACIÓN DEL PROBLEMA.....	5
1.3 IMPORTANCIA, ALCANCE Y LIMITACIONES.....	5
1.4 OBJETIVOS.....	8
CAPÍTULO II: MARCO REFERENCIAL Y TEÓRICO.....	9
2.1 ANTECEDENTES DE LA INVESTIGACIÓN.....	9
2.2 BASES TEÓRICAS.....	17
2.3 DEFINICIÓN DE TÉRMINOS BÁSICOS.....	26
2.4. HIPÓTESIS GENERAL.....	27
CAPÍTULO III: METODOLOGÍA.....	28
3.1 ENFOQUE, NIVEL Y METODOLOGÍA DE LA INVESTIGACIÓN.....	28
3.2 DISEÑO DE LA INVESTIGACIÓN.....	28
3.3 POBLACIÓN Y MUESTRA.....	28
3.4 VARIABLES.....	29
3.5 INSTRUMENTOS DE LA INVESTIGACIÓN.....	30
3.6 PROCESAMIENTO Y ANÁLISIS DE DATOS.....	31
CAPÍTULO IV: RESULTADOS.....	32
4.1. CARACTERÍSTICAS DE LA MUESTRA.....	32
4.2. ANALISIS DESCRIPTIVO.....	32
4.3. ANÁLISIS BIVARIADO.....	34
4.4. ANALISIS MULTIVARIADO:.....	36
MODELO DE REGRESIÓN LOGISTICA MULTIPLE.....	36
DISCUSIÓN.....	50
CONCLUSIONES.....	54
RECOMENDACIONES.....	55
REFERENCIAS BIBLIOGRÁFICAS.....	56
ANEXOS.....	62

## **Resumen**

**Objetivo:** Evaluar los factores sociodemográficos y clínicos asociados a hipertensión arterial mediante un modelo de regresión logística en una población de la costa del Perú.

**Métodos:** Investigación observacional, retrospectiva, de base secundaria. Los datos se registraron en historias clínicas de pacientes que fueron atendidos en el Área de Prevención de la Clínica Internacional sede-Lima, de enero a diciembre del 2013. La variable dependiente fue el diagnóstico de hipertensión arterial. Se realizaron modelos de regresión logística múltiple para ajustar los factores sociodemográficos y clínicos asociados a hipertensión arterial. Se calcularon los Odds Ratio (OR) y los intervalos de confianza del 95% (IC95%).

**Resultados:** Se evaluaron datos de 3535 pacientes, con una media de edad de  $46.3 \pm 12.8$ . La prevalencia de hipertensión fue del 5.05% (183pacientes). El modelo de regresión logística final indicó a la edad (OR, 1.07 IC95%, 1.05-1.08), colesterol total (OR, 1.01 IC95%, 1.001-1.01), frecuencia cardiaca (OR, 1.03 IC95%, 1.01-1.05) y peso (OR, 1.06 IC95%, 1.05-1.08) como principales factores de riesgo de presentar hipertensión arterial.

**Conclusión:** La personas con valores altos en edad, colesterol total, frecuencia cardiaca y peso tienen mayor riesgo de presentar hipertensión arterial.

**Palabras clave:** Hipertensión arterial; factores de riesgo; regresión logística; comorbilidades; obesidad.

## **Abstract**

**Objective:** To evaluate by means of a logistic regression model socio-demographic and clinical factors associated with hypertension in a population of the coast of Peru.

**Methods:** Observational, retrospective research, the data were recorded in medical records of patients who were treated in the Prevention Area of the International Clinic, Lima, from January to December 2013, and the dependent variable was the diagnosis of hypertension. Multiple logistic regression models were performed to adjust for sociodemographic and clinical factors associated with Hypertension. Odds Ratio (OR) and 95% confidence intervals (95%CI) were calculated.

**Results:** Data from 3535 patients were evaluated, with a mean age of  $46.63 \pm 12.08$ ; 183 (5.05%) presented hypertension. In the final logistic regression model age was considered (OR, 1.07 IC95%, 1.05-1.08), total cholesterol (OR, 1.01 IC95%, 1.001-1.01), heart rate (OR, 1.03 IC95%, 1.01-1.05), weight (OR, 1.06 IC95%, 1.05-1.08) were considered as main risk factors for hypertension.

**Conclusion:** People with high values for age, total cholesterol, heart rate and weight have a higher risk of developing hypertension.

**Keywords:** Arterial hypertension, risk factors, logistic regression, Comorbidities, obesity

## **CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA**

### **1.1 DETERMINACIÓN DEL PROBLEMA**

Las enfermedades cardiovasculares (ECV) son uno de los mayores problemas de salud pública en los países de bajos y medianos ingresos (1). Se estima que para el año 2030, la mortalidad por ECV alcanzará a 23 millones en la población adulta, alrededor del 85% de la población (2). La hipertensión arterial es un factor de riesgo importante para las ECV y su incidencia está incrementando rápidamente, debido al aumento de la longevidad y a la creciente prevalencia de factores de riesgo tales como la dieta poco saludable y la inactividad física que contribuyen a la obesidad (3).

Según los datos reportados por la Encuesta Demográfica y de Salud Familiar - ENDES en el 2020, se encontró que el 21.7% de las personas en el Perú mayores a 15 años presentaron hipertensión arterial. Asimismo, se reportó una mayor de hipertensión arterial en hombres (24.5%) en relación a las mujeres (19.1%) (4). Además, la región de la Costa presentó una mayor prevalencia con 23.9%, seguido por la región de la Sierra con 18.2% y por último la región de la Selva con 16.6% (4). Un estudio de revisión sistemática y metaanálisis sobre la prevalencia e incidencia de hipertensión arterial en el Perú, señaló que uno de cada cinco peruanos presenta hipertensión arterial y que cada año aumentan cuatro casos por cada 100 personas, además solo el 50% de hipertensos tienen un diagnóstico de la enfermedad (5).

Los factores de riesgos cardiovasculares asociados a la hipertensión arterial pueden coexistir entre ellos y contribuir a un mayor riesgo para la vida de las personas. Entre los factores cardiovasculares asociados con la hipertensión arterial, tenemos los factores de riesgo clínicos y sociodemográficos como la diabetes, la dislipidemia, la obesidad y sobrepeso, la edad, el sexo, los antecedentes familiares y los que comprenden al estilo de vida como el consumo de tabaco y alimentos no saludables, la dependencia al alcohol, la inactividad física y el estrés (6). La

modificación de los factores de riesgo cardiovasculares relacionados con el estilo de vida contribuyen en la prevención de la hipertensión arterial; así como, en la carga de enfermedad y la mortalidad asociadas a la presión arterial elevada.

Analizar los factores de riesgo significa de algún modo buscar la explicación multicausal del proceso salud-enfermedad, y bajo este punto de vista, se busca desglosar las llamadas redes causales para la identificación de los factores que pueden estar involucrados (7). El término riesgo tiene diferentes acepciones. Una de estas acepciones es la probabilidad de que una persona sana, pero expuesta a ciertos factores, adquiera o desarrolle una enfermedad dada. Tales factores, llamados factores de riesgo, son características que van acompañadas de un aumento en la probabilidad de que ocurra un proceso patológico; lo cual significa que se encuentran asociadas estadísticamente con la ocurrencia de un daño, aunque dicha asociación pueda ser o no de tipo causal (8).

La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de predictores. Es similar a un modelo de regresión lineal, pero está adaptado para modelos en los que la variable dependiente es dicotómica. Los coeficientes de regresión logística pueden utilizarse para estimar la razón de probabilidad de cada variable independiente del modelo; además, se puede aplicar a un rango más amplio de situaciones de investigación que el análisis discriminante (9).

La regresión logística no se basa en supuestos distribucionales en el mismo sentido en que lo hace el análisis discriminante; a su vez, la solución puede ser más estable si los predictores tienen una distribución normal multivariante. Sin embargo, al igual que otras formas de regresión, la multicolinealidad entre los predictores puede llevar a estimaciones sesgadas y a errores estándar inflados. El procedimiento es más eficaz cuando la pertenencia a grupos es una variable categórica auténtica; si la pertenencia al grupo se basa en valores de una variable continua, se deberá considerar

el utilizar la regresión lineal para aprovechar la información mucho más rica de la propia variable continua (9).

Por lo tanto, mediante este trabajo se podrá observar, describir y cuantificar los factores de riesgo cardiovasculares asociados a hipertensión arterial, los cuales se pueden utilizar como herramienta de discusión para mejorar el manejo de pacientes con esta enfermedad con el fin de hacer posible la prevención primaria de la hipertensión arterial mediante la reducción de dichos factores de riesgo para evitar complicaciones severas como el accidente cerebro vascular, enfermedad coronaria, entre otras.

En ese sentido, el objetivo del estudio fue evaluar los factores sociodemográficos y clínicos asociados a hipertensión arterial mediante un modelo de regresión logística en una población de la costa del Perú.

## **1.2 FORMULACIÓN DEL PROBLEMA**

¿Cuáles son los factores sociodemográficos y clínicos asociados a hipertensión arterial en una población de la costa del Perú evaluados mediante un modelo de la regresión logística?

## **1.3 IMPORTANCIA, ALCANCE Y LIMITACIONES**

Importancia en la epidemiología e investigación de la HTA:

El presente estudio es de gran importancia en el campo de la epidemiología y la investigación de la hipertensión arterial (HTA) por varias razones: En cuanto a Identificación de Factores de Riesgo: respecto a los Factores Sociodemográficos: La investigación permite identificar variables como la edad, el género que influyen en la prevalencia de la HTA. Esto es crucial para comprender las

disparidades en salud y diseñar intervenciones específicas para distintos grupos poblacionales.

**Factores Clínicos:** El estudio de factores clínicos como el índice de masa corporal (IMC), la y obesidad, ayuda a identificar individuos en mayor riesgo y a personalizar estrategias de prevención y tratamiento. También, el desarrollo de modelos predictivos: El uso de modelos de regresión logística permite cuantificar la relación entre los factores de riesgo identificados y la probabilidad de desarrollar HTA. Esto facilita la creación de herramientas predictivas que pueden ser utilizadas en la práctica clínica para evaluar el riesgo individual de los pacientes y tomar decisiones informadas sobre su manejo. Otro punto es respecto a las Políticas de Salud Pública; **Diseño de Intervenciones:** Conocer los factores asociados a la HTA permite a los responsables de políticas de salud diseñar programas de prevención y control más efectivos, dirigidos a las poblaciones más vulnerables y **Asignación de Recursos:** La identificación de áreas geográficas o grupos demográficos con mayor prevalencia de HTA permite una asignación más eficiente de los recursos de salud pública, enfocando los esfuerzos en las áreas de mayor necesidad. Asimismo, **Educación y Concienciación;** **Informar a la Población:** Los resultados de la investigación pueden ser utilizados para campañas de concienciación sobre la HTA, educando a la población sobre los factores de riesgo y promoviendo hábitos de vida saludables; **Capacitación de Profesionales de la Salud:** Proporciona a los profesionales de la salud información basada en evidencia sobre los factores de riesgo específicos para la población de la costa del Perú, mejorando así la prevención, el diagnóstico temprano y el manejo de la HTA. También, **Base para Investigaciones Futuras;** mediante la **Ampliación del Conocimiento:** Este estudio puede servir como base para investigaciones futuras que exploren la interacción de múltiples factores de riesgo y evalúen la efectividad de intervenciones específicas y la **Comparación Internacional:** Los resultados pueden ser comparados con estudios similares en otras regiones y países, contribuyendo a un entendimiento global de la epidemiología de la HTA y permitiendo adaptaciones de estrategias exitosas de un contexto a otro.

El siguiente trabajo tiene importancia clínica, debido a que genera evidencia científica que refuerza la medicina preventiva al analizar los factores de riesgo cardiovasculares asociados a hipertensión arterial en una población típica de nuestra ciudad capital. Cuando el análisis es realizado a través de modelos de regresión logística aplicados a las ciencias de la salud, nos permite describir los resultados en términos explicativos y predictivos, pudiendo conocer la fuerza de asociación mediante los *odds ratio* (OR) de los factores de riesgo con el efecto estudiado de una manera independiente y conocer el valor predictivo de cada uno de ellos o bien del modelo en su conjunto.

Se sabe que la regresión logística es de utilidad para la investigación clínica y epidemiológica, modela la probabilidad de un evento en forma de factores de riesgos. Además, la cuantificación de los riesgos (con significación estadística, mediante intervalos de confianza: IC) permite estimar riesgo relativo (RR) y *odds ratio* ("razón de productos cruzados": OR) entre otros.

Además, la regresión logística no solo proporciona una medida de cuán apropiado es un predictor (tamaño del coeficiente), sino también su dirección de asociación (positiva o negativa). Sin embargo, la principal limitación de la regresión logística es el supuesto de linealidad entre la variable dependiente y las variables independientes.

No se tuvieron datos que podrían estar asociados a hipertensión, tales como obesidad, consumo excesivo de alcohol y sal, presencia previa de alguna enfermedad, estilos de vida y comportamientos individuales como consumo de algún medicamento que altere su presión arterial, o tiempo de la enfermedad y depresión.

Entre limitaciones del modelo, la variable dependiente se está dicotomizando y se puede estar perdiendo información valiosa al utilizar el modelo de regresión logística para evaluar los factores asociados a presión arterial aumentada.

## **1.4 OBJETIVOS**

### **Objetivo general:**

- Evaluar factores sociodemográficos y clínicos asociados a hipertensión arterial mediante un modelo de regresión logística en una población de la costa del Perú.

### **Objetivos específicos:**

- Determinar el modelo de regresión que explique la asociación entre hipertensión arterial y factores clínicos.
- Determinar el modelo de regresión que explique la asociación entre hipertensión arterial y factores sociodemográficos.
- Estimar el modelo de regresión logística multivariado considerando los factores de riesgo asociados a hipertensión arterial.

## **CAPÍTULO II: MARCO REFERENCIAL Y TEÓRICO**

### **2.1 ANTECEDENTES DE LA INVESTIGACIÓN**

En el 2020, Saeed et al. Evaluaron la prevalencia de hipertensión y sus factores de riesgo asociados en 600 personas de Gadarif. Analizaron características sociodemográficas como la edad, sexo, altura, peso, estado civil, nivel educativo, hábito de fumar, consumo de alcohol y antecedentes familiares de hipertensión. El 40.8% de los participantes presentó hipertensión, donde el 7.3% fueron diagnosticados previamente con hipertensión y 33.5% fueron diagnosticados con hipertensión por primera vez. Se utilizó la prueba de chi-cuadrado para comparar las proporciones entre los participantes que fueron diagnosticados con hipertensión y los participantes que no tenían hipertensión. Se utilizaron las pruebas t y Mann-Whitney para comparar los datos paramétricos y no paramétricos continuos, respectivamente, entre los dos grupos (hipertensos y no hipertensos). Se realizaron análisis de regresión logística con hipertensión arterial como variable dependiente y la edad, sexo, estado civil, nivel educativo, índice de masa corporal (IMC) y perímetro de cintura como variables independientes. Las variables independientes ingresaron al modelo si su valor de p era  $<0.20$ . Odds Ratios (OR) e intervalos de confianza del 95 % (IC). Luego, los ajustes se realizaron utilizando la razón de verosimilitud hacia atrás (LR) en los diferentes modelos. Concluyeron que existe una alta tasa de hipertensión en el este de Sudán, especialmente entre las personas mayores y obesas, recomendando la implementación de medidas preventivas, como las medidas dietéticas (10).

En el 2020, Nam et al. Evaluaron la prevalencia hipertensiva y determinaron los factores de riesgo asociados a la hipertensión. Se entrevistó a 197 participantes  $\geq$  de 18 años del distrito de Moc Chau. Se midió la presión arterial y todos los datos se ingresaron al software Epidat 3.1 y luego se analizaron con el software Stata 16.0. Se utilizó el método descriptivo para describir el estado de la

hipertensión por características seleccionadas (como porcentaje e intervalo de confianza del 95%). Para estudiar los factores de riesgo de hipertensión se aplicaron análisis bivariados y multivariados con regresión logística. La significación estadística se consideró como  $p < 0,05$ . Se encontró que el 30% tenía prevalencia hipertensiva y que los factores asociados independientemente con la hipertensión de forma significativa ( $p < 0,05$ ) fueron: edad  $> 55$  años (52.3%), consumo de sal  $> 5g$  por día (32.4%), consumo de alcohol (40.3%), antecedentes familiares de hipertensión (31.7%) y al menos una enfermedad preexistente (44.3%). Concluyeron que la edad, el consumo de mucha sal, el consumo de alcohol, los antecedentes familiares de hipertensión y al menos la presencia de una enfermedad fueron factores de riesgo de hipertensión en el grupo de estudio (11).

En el año 2020, Alvarenga et al. Describieron la prevalencia y los factores de riesgo asociados con la hipertensión arterial sistémica en adultos y ancianos de Krenak, Brasil. El estudio fue de tipo transversal con 183 participantes. Se realizaron análisis bivariado mediante la prueba de chi-cuadrado de Pearson para verificar las asociaciones entre la variable dependiente hipertensión arterial (categorizada dicotómicamente (sí/no)) y las variables independientes. Los cocientes de prevalencia y sus intervalos de confianza del 95 % (IC del 95 %) también se estimaron mediante regresión de Poisson con varianzas robustas. Finalmente, se construyeron modelos multivariados con regresión de Poisson con varianzas robustas según un modelo jerárquico que siguió los supuestos teóricos propuestos por un estudio anterior (Victora CG, 1997) y fue probado previamente en un estudio sobre factores asociados a la hipertensión arterial en una población indígena. En tal modelo, hubo tres bloques de variables (bloque 1 - demográficas; bloque 2 - socioeconómicas y conductuales; bloque 3 - morbilidades). Así, inicialmente, se introdujeron en el modelo las variables demográficas que presentaban un nivel de significancia de  $p < 0,20$  en el análisis bivariado, mientras que aquellas con valor de  $p < 0,05$  permanecían en el modelo. En el segundo paso se incluyeron

variables socioeconómicas y de comportamiento, utilizando los mismos valores de significación estadística que en el primer paso. En la tercera etapa se agregaron las variables de morbilidad, adoptando los mismos criterios que en las etapas anteriores. La investigación halló una prevalencia de 31.2%; se evaluó la altura, circunferencia de la cintura, presión arterial sistólica, presión arterial diastólica y glucosa en sangre capilar; se recolectó información sociodemográfica y de estilo de vida a través de un cuestionario. Se determinó la existencia de una asociación independiente en el modelo final al aumento de la edad ( $p < 0.001$ ), la obesidad abdominal ( $p < 0.001$ ) y la hiperglucemia ( $p = 0.003$ ). Concluyendo que la edad, la obesidad abdominal y la hiperglucemia son factores de riesgo de la hipertensión arterial y que se destaca la necesidad de acciones efectivas de prevención, diagnóstico y seguimiento en la comunidad indígena Krenak (12).

En el 2020 Tshepo et al. Realizaron un estudio transversal cuyo objetivo fue investigar la prevalencia de prehipertensión, e hipertensión, y la asociación entre la presión arterial, el IMC y el área de superficie corporal en 218 adolescentes entre 13 y 19 años de la región de Polokwane al norte de Sudáfrica. Se calcularon las frecuencias de distribución para todas las variables y se verificaron los valores atípicos con los datos sin procesar y, cuando fue necesario, se eliminaron los valores atípicos. Para describir y caracterizar las muestras se calcularon estadísticos descriptivos para todas las variables con el fin de indicar frecuencias (expresadas en porcentajes), medias y desviaciones estándar. Esto sirvió como una indicación de la prevalencia. La estimación de prevalencia se presentó por sexo. La asociación entre la presión arterial y los datos antropométricos (IMC y área de superficie corporal) se probó mediante la correlación de Pearson. Los participantes se dividieron en diferentes categorías de edad y sexo para un análisis más detallado de los resultados. El valor de probabilidad de significación estadística para todas las pruebas se fijó en un valor  $p < 0.05$ . La prevalencia de prehipertensión e hipertensión en varones fue de 28.9% y 10.3% en comparación con

25.6% y 7.4% en mujeres, y la prevalencia de prehipertensión e hipertensión en adolescentes con sobrepeso / obesidad fue de 7.3% y 2.7% con una débil asociación positiva con el IMC. En conclusión, la prevalencia de prehipertensión e hipertensión fue mayor en hombres que en mujeres (13).

En el 2019, Princewel et al. Determinaron la prevalencia e identificaron los factores de riesgo asociados con la hipertensión en 243 adultos mayores a 21 años en un entorno rural de Camerún. Se usó una encuesta, donde se obtuvo datos sociodemográficos, nivel de conocimientos relacionados a la hipertensión, datos antropométricos (IMC) y datos sobre la presión arterial actual. Se determinaron frecuencias y porcentajes para variables categóricas. Se determinaron medias y desviaciones estándar (media  $\pm$  DE). Se utilizaron análisis de regresión logística bivariado y multivariado para investigar los factores asociados con la presión arterial alta (hipertensión). La edad mayor de 40 años ( $p=0.002$ ), el consumo nocivo de alcohol durante más de 10 años ( $p=0.014$ ) y la inactividad física ( $p=0.021$ ) fueron variables asociadas con la hipertensión de forma significativa. Concluyeron que los adultos son más prevalentes a presentar presión arterial alta (19.8%) y que la vejez, el consumo de alcohol y la inactividad física son factores de riesgo independientes para la hipertensión (14).

En el año 2019, Pinto et al. Analizaron la asociación de la hipertensión y su limitada correlación con factores sociodemográficos y epidemiológicos en 44.271 adultos brasileños de 30 años o más. Para estudiar la asociación entre hipertensión arterial y características sociodemográficas, factores de estilo de vida y estado de salud, se estimó la razón de prevalencia (RP) de hipertensión arterial y sus intervalos de confianza al 95% (IC 95%), utilizando modelos de regresión de Poisson. El desenlace fue la presencia de hipertensión arterial y las variables

independientes fueron las siguientes: sexo, grupo de edad, raza/color de piel, nivel educativo, IMC, circunferencia de la cintura, tabaquismo, consumo adecuado de frutas y verduras, actividad física, autopercepción de salud, dolor torácico al caminar, cardiopatía, ictus, diabetes, depresión según PHQ-9 y al menos una consulta médica en el último año. El análisis estadístico se hizo por separado para cuatro grupos de variables: características sociodemográficas, medidas antropométricas, estilos de vida y situación de salud. Para cada grupo de variables se calculó la razón de prevalencia ajustada según sexo y edad (modelo 1) y según todas las demás covariables (modelo 2). Los datos fueron analizados utilizando el software Stata, versión 11.0, a través del módulo de encuestas, que incorpora los efectos de muestras complejas. La hipertensión fue el factor de riesgo más frecuente entre los adultos (40,7%); además estuvo fuertemente asociado con el proceso de envejecimiento (RP 3.51), obesidad (RP 1.73), enfermedad cardíaca (RP 1.67) y accidente cerebrovascular (RP 1.86). Las limitaciones asociadas con la hipertensión fueron más frecuentes entre aquellos con comorbilidades de enfermedades no transmisibles relacionadas con la hipertensión (accidente cerebrovascular RP 1.47; enfermedad cardíaca RP 1.69) y con educación primaria incompleta (RP 1.19) (15).

Para el año 2019, Odame et al. Estudiaron la prevalencia y factores de riesgo de la obesidad y la hipertensión en conductores de autobuses en Ghana, África. La muestra fue de 527 conductores a quienes se le aplicó un cuestionario sobre características sociodemográficas y factores relacionados con el estilo de vida. Los análisis de datos se realizaron mediante el SPSS 20 y utilizaron estadísticas descriptivas como frecuencias, porcentajes y gráficos. Para determinar la asociación entre las variables categóricas se usó la prueba de Fisher o chi-cuadrado. Se realizó análisis de regresión logística para determinar los factores de riesgo de obesidad e hipertensión. Las pruebas de significación estadística ( $p < 0,05$ ) se informaron en función de la probabilidad del test de dos colas (two-tailed probability). La prevalencia de hipertensión fue del 38.7%. También evidenciaron que el envejecimiento, fumar, la ingesta de bebidas alcohólicas, el uso de inhibidores del sueño, ingesta

alta en calorías, estar sentado durante mucho tiempo, comer tarde y estar bajo condiciones estresantes fueron factores de riesgo independientes para la hipertensión ( $p < 0.05$ ). Concluyeron que existe una alta prevalencia de hipertensión y obesidad en conductores de autobuses que se asocian con estilos de vida y comportamientos individuales (16).

En el 2019, Jakab et al. Evaluaron la prevalencia de hipertensión en niños y adolescentes húngaros obesos y con sobrepeso. El estudio estuvo conformado por 8624 participantes entre 3 a 18 años que se dividieron en tres grupos: normo peso, sobrepeso y obesidad en base al IMC, a los cuales se les realizaron diferentes exámenes para diagnosticar hipertensión. Se utilizó SPSS 22 para el procesamiento de datos y las pruebas estadísticas, con un nivel de significación de  $p < 0,05$ . Se encontró que la prevalencia de sobrepeso y obesidad fue de 23.5% en general, 26.4% en niños y 20% en niñas. La prevalencia de la hipertensión fue del 9,8% en pacientes con sobrepeso y de 4,6% en pacientes con obesidad. En conclusión, el riesgo cardiovascular está aumentando en pacientes con sobrepeso y obesidad, y es esencial establecer una adecuada estrategia de prevención primaria (17).

En el 2019, Kumar y Verma. Estudiaron la prevalencia de hipertensión y sus factores asociados en adolescentes que acuden a la escuela en Delhi, India. Se utilizó un cuestionario para recopilar información de 600 adolescentes entre 9 a 12 años de cuatro escuelas. Se midió la presión arterial, la altura, el peso y la circunferencia de la cintura. Se utilizó el software SPSS 21.0 para el análisis. Se realizaron tablas simples y tablas cruzadas para variables cuantitativas la prueba de chi-cuadrado ( $\chi^2$ ) para variables cualitativas. Los resultados mostraron que 44 (7.3%) adolescentes tenían prehipertensión y 26 (4.3%) tenían hipertensión, siendo la hipertensión significativamente mayor en aquellos con antecedentes familiares de hipertensión y con sobrepeso / obesidad. Se concluyó que la prevalencia de hipertensión fue significativamente alta, siendo el

sobrepeso/obesidad y los antecedentes familiares de hipertensión los factores de riesgos más prevalentes (18).

En el 2019, Ghosh y Kumar, evaluaron la prevalencia y factores de riesgo asociados con la hipertensión entre 779649 personas entre 15 a 49 años de la India. Los datos los obtuvieron mediante la Encuesta Nacional de Salud Familiar (NFHS). Se tomaron tres lecturas de presión arterial y las promediaron para decidir si una persona era hipertensa. Emplearon el software estadísticos STATA v14, hicieron análisis bivariado y emplearon un modelo de regresión logística multinivel (primer nivel: individual; segundo nivel: distrito; tercer nivel: estado) con intercepto aleatorios y pendientes fijas para calcular los ORs multinivel con el correspondiente IC del 95%. Los datos revelaron que la prevalencia de hipertensión fue de 11,3%, siendo los hombres los que presentaron mayor prevalencia con 13,8%. Por otro lado, la zona urbana presentó 12,5% de hipertensión y la rural de 10,6%. Se concluyó que la prevalencia de hipertensión ha aumentado de manera considerable en las personas de bajos recursos y que la edad avanzada, la obesidad/sobrepeso, el estado socioeconómico y el consumo de alcohol influyeron en gran medida a desarrollar hipertensión (19).

En el 2019, El Achhab et al. Determinaron la prevalencia y el perfil clínico de la hipertensión en personas de Marruecos. La muestra fue de 10714 y los datos se recolectaron a través de un examen médico y un cuestionario que cubrió la demografía del paciente, el historial médico y los factores de riesgo cardiovascular. Los análisis estadísticos se realizaron con SPSS (versión 17.0). Se utilizó análisis descriptivo para determinar la prevalencia de hipertensión en toda la muestra. Las asociaciones entre varias facetas de la hipertensión (antecedentes de hipertensión, hipertensión recién detectada, gravedad de la hipertensión, hipertensión controlada), y la demografía del paciente y el perfil médico personal (edad, sexo, antecedentes médicos personales y familiares, índice de

masa corporal (IMC), tabaquismo y tratamiento actual) fueron investigadas. El análisis estadístico utilizó en pruebas estadísticas convencionales (prueba  $\chi^2$ , prueba t de Student y análisis de varianza). Una prueba se consideró significativa cuando  $p < 0,05$ . Cuando las comparaciones se realizan dentro de una categoría que contiene más de dos subcategorías, los valores de p se indicaron como "dentro de la categoría". La prevalencia de hipertensión fue del 26,6% (26,3% en hombres y 28,0% en mujeres). Entre los pacientes con antecedentes de hipertensión, al 85,9% se les había recetado medicación antihipertensiva y consejos sobre el estilo de vida y alimentación saludable. Se concluyó que la prevalencia de hipertensión es alta y que puede ser una enfermedad infra diagnosticada y tratada de manera ineficaz (20).

En el 2019, Mokgwathi y Chacha determinaron la prevalencia de hipertensión y factores de riesgo cardiovascular entre adolescentes de escuelas rurales y urbanas en Botswana. Los datos fueron recolectados a través de un cuestionario, mediciones y pruebas de glucosa en sangre en ayunas, y se analizaron utilizando SPSS 23.0. Las variables continuas (glucemia en ayunas, talla, peso, circunferencia de cintura, presión arterial sistólica, presión arterial diastólica y edad) se resumieron por medias ( $\pm$  SD) y las variables categóricas se resumieron por conteos y porcentajes. Se utilizó la prueba de chi-cuadrado de Pearson para comparar la prevalencia de factores de riesgo cardiovascular seleccionados (hipertensión arterial, diabetes mellitus, tabaquismo, obesidad/sobrepeso, nivel de actividad física y consumo de alcohol). Para el análisis bivariado de las variables continuas (glucemia en ayunas, talla, peso, circunferencia de la cintura, edad) se utilizó la prueba T de Student. Un valor de p inferior a 0,05 se consideró estadísticamente significativo. Las variables que tuvieron  $p < 0,25$  en el análisis bivariado se incluyeron como variables independientes para la regresión logística multivariado. La muestra fue de 252 estudiantes con una edad media de  $17.1 \pm 0.9$  años, la prevalencia de hipertensión fue 13.1% y de prehipertensión fue 15.5%. Sobre los

factores de riesgo, la inactividad física fue la más alta (37,7%), luego el sobrepeso y obesidad (10.3%) y, por último, la ingesta de alcohol (9,1%). Los datos mostraron que la hipertensión fue más prevalente en los hombres y que las mujeres eran más propensas a tener sobrepeso u obesidad. En conclusión, la hipertensión, sobrepeso / obesidad y alcohol fue común en la población estudiada y se consideran factores de riesgo para las ECV (21).

## **2.2 BASES TEÓRICAS**

La detección y tratamiento de la hipertensión arterial son estrategias claves de prevención para evitar la muerte y discapacidad relacionada con la enfermedad cardiovascular (ECV) (22). Por ejemplo, un metaanálisis reciente de ensayos controlados aleatorizados, mostró que un paciente con hipertensión puede reducir el riesgo de mortalidad cardiovascular en un 25-40% al reducir la presión sanguínea en 10 mm Hg (22). Sin embargo, muchas preguntas críticas respecto al manejo de la hipertensión arterial siguen sin resolverse, especialmente en países subdesarrollados.

La mayoría de los estudios relacionados con el conocimiento de la hipertensión arterial, el tratamiento y el control se han llevado a cabo en China, Malasia, Tailandia, México e Irán, donde el manejo de la hipertensión arterial está mejorando, debido a un mejor conocimiento y acceso a los servicios de salud (22-25). Estos estudios informaron que la edad, sexo, IMC, educación, condición económica y lugar de residencia son potenciales predictores de diagnóstico, tratamiento y control de la hipertensión arterial. Al igual que en otros países en desarrollo de bajos ingresos económicos, (22-27) la obesidad, diabetes y la hipertensión arterial son un emergente problema de salud pública.

En algunos estudios sobre prevalencia de hipertensión en China como encuestas nacionales realizadas en 1959, 1979 y 1991 se muestran estimaciones de prevalencia de 5,1, 7,7 y 13,6%, respectivamente (27), mientras que encuestas realizadas en los últimos años mostraron un aumento de prevalencia alrededor del 18% en el 2002, 25% en el 2007 y 33% en el 2010 (28, 29). Además,

las encuestas regionales en el este, noroeste y sur de China, mostraron incrementos en la prevalencia de la hipertensión arterial (29-31). Sin embargo, debido a diferencias de la edad y sexo entre los estudios, las prevalencias globales son difíciles de comparar.

Las directrices actuales para el diagnóstico y tratamiento de hipertensión arterial recomiendan la estratificación del riesgo basado en que la presión se mide convencionalmente; es decir, la presión arterial se mide en un entorno médico (32,33). Por otro lado, las directrices europeas (32) categorizan a la presión arterial como óptima, normal, normal alta y grados 1-3 de hipertensión; y las directrices estadounidenses (33) clasifican a la presión arterial como presión arterial normal, prehipertensión y etapas 1 y 2 de la hipertensión.

La hipertensión arterial es una de las enfermedades más comunes que afectan a los seres humanos en todo el mundo. El nivel de la presión arterial ha sido reconocido como un factor determinante de riesgos para varias ECV, incluyendo la enfermedad coronaria y cerebrovascular. La ECV es una crisis de salud pública que afecta a más de 60 millones de estadounidenses con 50 millones de hipertensos, 12 millones con enfermedad coronaria, 7 millones de infarto de miocardio y con más de 4 millones de accidente cerebrovascular (1, 33). Muchas de estas enfermedades están creciendo aún más con la prevalencia de la hipertensión arterial que aumenta a medida que la población envejece.

El control adecuado sigue siendo un enfoque terapéutico esencial para prevenir complicaciones y muerte prematura; a su vez, es una manera eficaz para retardar la progresión de la diabetes y enfermedades renales no diabéticas. Es evidente que el tratamiento, independientemente del enfoque, ha disminuido la mortalidad y la morbilidad. Sin embargo, los informes muestran que los pacientes con la presión arterial bien controlada representan sólo un pequeño porcentaje de la

población hipertensa. Por lo tanto, existe una clara necesidad de desarrollar estrategias eficaces para mejorar el manejo de la hipertensión arterial (34).

### **Modelo de regresión logística**

La regresión logística es una de las técnicas más utilizadas para modelar la asociación entre una variable dependiente dicotómica y una o más variables independientes. Cuando la variable de respuesta  $Y$  es una variable binaria, los parámetros importantes asociados con la distribución de  $Y$  son  $p = P(Y = 1)$  y  $1 - p = P(Y = 0)$  (35).

Los estudios clínicos usan con frecuencia variables con resultados dicotómicos (por ejemplo, dependencia o independencia) por su frecuencia y utilidad para representar múltiples estados de salud. El análisis estadístico de estas variables se realiza a menudo mediante modelos de regresión logística multivariada de uso tanto explicativo como predictivo y con muy buena capacidad de análisis de datos en la investigación clínica y epidemiológica e ideales en los estudios de casos y controles, de factores pronósticos y de pruebas diagnósticas (sensibilidad y especificidad)(36).

Para enfatizar, cualquier modelo en la variable de resultado dicotómica  $Y$  (es decir,  $y = 0$  o  $1$ ) con una o más variables independientes (continuas o categóricas), la regresión lineal a las  $p$  variables independientes:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

ya no es apropiado debido a los valores de resultado binarios de 0 y 1. Para modelar correctamente los datos binarios (o binomiales), se utiliza la regresión logística para modelar la probabilidad con transformación logit de la siguiente manera:

$$\log(p / 1 - p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (36)$$

Las estimaciones de probabilidad estarán siempre entre 0 y 1, por tanto, el valor de la variable respuesta se definirá como la probabilidad de que ocurra un evento bajo determinadas variables independientes.

Al evaluar la validez y calidad de un modelo de regresión logística múltiple, se analiza tanto el modelo en su conjunto como los predictores que lo forman. Se considera que el modelo es útil si es capaz de mostrar una mejora respecto al modelo nulo, el modelo sin predictores(37,38).

El modelo de regresión logística es una herramienta estadística ampliamente utilizada en epidemiología y medicina para identificar y cuantificar la relación entre factores de riesgo y la probabilidad de desarrollar una enfermedad. Además de la hipertensión, este modelo tiene múltiples aplicaciones en el estudio de otras enfermedades. A continuación, se mencionan algunos ejemplos:

### 1. Diabetes Mellitus

**Identificación de Factores de Riesgo:** Se usa para determinar cómo factores como el índice de masa corporal, la edad, la actividad física, la dieta y la historia familiar influyen en el riesgo de desarrollar diabetes tipo 2.

**Predicción del Riesgo:** Ayuda a desarrollar modelos predictivos para identificar individuos en riesgo de desarrollar diabetes.

### 2. Enfermedades Cardiovasculares

**Evaluación de Factores de Riesgo:** Se utiliza para analizar la influencia de factores como el colesterol alto, el tabaquismo, la hipertensión, la diabetes y el sedentarismo en la incidencia de enfermedades cardíacas y accidentes cerebrovasculares.

**Prevención Primaria y Secundaria:** Facilita la creación de programas de prevención y la estratificación del riesgo en pacientes.

### 3. Cáncer

**Determinación de Factores de Riesgo:** Permite investigar la relación entre factores como el tabaquismo, la exposición a carcinógenos, la dieta, la genética y la incidencia de distintos tipos de cáncer (por ejemplo, cáncer de pulmón, mama, colon).

**Evaluación de Resultados de Tratamiento:** Ayuda a evaluar la efectividad de diferentes tratamientos y a predecir la supervivencia y la recurrencia del cáncer.

#### 4. Enfermedades Infecciosas

**Epidemiología del VIH/SIDA:** Se utiliza para identificar factores de riesgo asociados con la infección por VIH y para evaluar la efectividad de intervenciones preventivas.

**Tuberculosis:** Permite analizar cómo factores como la coinfección con VIH, el estatus socioeconómico y el acceso a servicios de salud afectan la probabilidad de desarrollar tuberculosis activa.

#### 5. Enfermedades Respiratorias

**Asma y Enfermedad Pulmonar Obstructiva Crónica (EPOC):** Se emplea para estudiar la influencia de factores ambientales, ocupacionales y genéticos en el desarrollo de enfermedades respiratorias crónicas.

**Neumonía:** Ayuda a identificar factores de riesgo para el desarrollo y la severidad de la neumonía, especialmente en poblaciones vulnerables como niños y ancianos.

#### 6. Enfermedades Neurológicas

**Accidente Cerebrovascular (ACV):** Se utiliza para identificar factores de riesgo como la hipertensión, la fibrilación auricular, la diabetes y el estilo de vida, y para predecir el riesgo de eventos cerebrovasculares.

Demencia y Enfermedad de Alzheimer: Permite investigar la relación entre factores como la edad, la genética, el nivel educativo y el riesgo de desarrollar demencia.

## 7. Enfermedades Metabólicas

Obesidad: Se utiliza para analizar cómo factores como la dieta, la actividad física, los factores genéticos y el entorno influyen en el desarrollo de la obesidad.

Síndrome Metabólico: Ayuda a identificar la relación entre distintos componentes del síndrome metabólico (hipertensión, dislipidemia, hiperglucemia, obesidad abdominal) y su impacto en la salud.

## 8. Salud Mental

Depresión y Ansiedad: Se emplea para identificar factores de riesgo asociados con trastornos mentales, como el estrés, el apoyo social, los antecedentes familiares y los eventos traumáticos.

Trastornos del Sueño: Ayuda a estudiar la influencia de factores como el estilo de vida, las condiciones médicas subyacentes y el entorno en la prevalencia de trastornos del sueño.

## 9. Enfermedades Autoinmunes

Artritis Reumatoide y Lupus: Se utiliza para identificar factores de riesgo y gatillantes ambientales y genéticos asociados con enfermedades autoinmunes.

## 10. Salud Maternoinfantil

Complicaciones del Embarazo: Ayuda a identificar factores de riesgo para condiciones como la preeclampsia, la diabetes gestacional y el parto prematuro(37,38).

**Estimación de los coeficientes:** aunque existen otros métodos, se suele realizar por el método de máxima verosimilitud, que consiste en maximizar la función de verosimilitud de la muestra. Cuando

el tamaño de la muestra es lo suficientemente grande y el modelo se ha especificado correctamente, las estimaciones de máxima verosimilitud son casi insesgadas; sus errores estándar se pueden calcular fácilmente y sus distribuciones de muestreo se distribuirán aproximadamente normal(35,37).

### Supuestos de un modelo de regresión logística

Modelo de regresión logit:

$$\text{logit}(p / 1 - p) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Donde:

1. Las observaciones  $Y_1, Y_2, \dots, Y_n$  son independientes entre sí,
2. Se miden las variables explicativas sin error,
3. No existe multicolinealidad entre las variables explicativas,
4. La variación en la variable de respuesta sigue la función de varianza binomial

$$\text{Var}(Y|X) = p(X)(1 - p(X)),$$

5. el tamaño de la muestra debe ser suficiente para realizar procedimientos de cálculo (35,37)

La interpretación de la regresión logística comparte algunas similitudes con la de la regresión lineal; por ejemplo, las variables a las que se les da la mayor importancia pueden ser predictores confiables, pero en realidad pueden no ser causales. Los parámetros de regresión logística se pueden usar para comprender el poder predictivo relativo de diferentes variables, suponiendo que las variables ya se han normalizado para tener una media de 0 y una varianza 1 (39)

Después de desarrollar un modelo a través de la prueba de bondad de ajuste, el siguiente paso es examinar los residuos y los diagnósticos basado en estos residuos para identificar cualquier punto en el que el modelo no se ajuste bien. Recuerde que los residuos de un modelo de regresión ajustado contienen la información relevante sobre qué tan bien se ajusta el modelo a los datos

observados. De manera similar, los residuos de un modelo de regresión logística ajustado también se pueden usar para evaluar el ajuste de un modelo de regresión logística. Los residuales asociados a un modelo de regresión logística se basan en comparar el valor de la variable respuesta con la probabilidad estimada de la respuesta en base al modelo ajustado. Un residuo que se usa a menudo para evaluar el ajuste de un modelo de regresión logística es el residuo de Pearson que se denotará por  $r$  (35,37,38).

Una vez obtenido los coeficientes del modelo logit ( $p/1-p$ ), se obtendrá la función de regresión logística, bajo el siguiente modelo:

$$P(X_i) = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{(1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)})}$$

Lo que se interpreta como la probabilidad de tener el evento dado que está expuesto a las variables  $X_1, X_2, \dots, X_p$

### **Odds Ratio (OR)**

Es una medida de asociación entre una exposición y un evento. El OR representa la razón de las probabilidades de que ocurra un evento dada una exposición en particular, en comparación con las probabilidades de que el evento ocurra en ausencia de esa exposición. Las proporciones de probabilidades se usan más comúnmente en los estudios de casos y controles; sin embargo, también se pueden usar como una aproximación en diseños de estudios transversales y de cohortes (con algunas modificaciones y / o suposiciones)(39,40).

### **Odds Ratios y regresión logística**

Cuando se calcula una regresión logística, el coeficiente de regresión ( $\beta_1$ ) es el aumento estimado en las probabilidades logarítmicas del *resultado por unidad de aumento* en el valor de la *exposición*. En otras palabras, la función exponencial del coeficiente de regresión ( $e^{\beta_1}$ ) es el índice de probabilidades asociado con un aumento de una unidad en la exposición (41).

### **Multicolinealidad**

Como en el caso de la regresión lineal, las relaciones colineales entre las variables explicativas hacen que los coeficientes de regresión logística estimados, los errores estándar estimados y los valores de  $p$  asociados con una prueba de hipótesis no sean confiables y pueden conducir a inferencias inexactas y engañosas. Afortunadamente, la mayoría de los paquetes estadísticos incluyen el factor de inflación de la varianza (VIF) junto con las estimaciones de los coeficientes en el resumen de salida del modelo ajustado. Al igual que en la configuración de regresión múltiple, un valor de VIF superior a 10 indica un problema de colinealidad que debe abordarse(42,43).

### **La curva ROC**

El análisis en base a las curvas ROC (por sus siglas en inglés *Receiver Operating Characteristic*), constituye un método estadístico para determinar la exactitud diagnóstica del test que utilizan escalas continuas, siendo utilizadas con 3 propósitos específicos:

1. Determinar el punto de corte en el que se alcanza la sensibilidad y especificidad más altas, el cual presenta el mayor índice de Youden, calculado según la fórmula (sensibilidad + especificidad – 1). Gráficamente, este corresponde al punto de la curva ROC más cercano al ángulo superior izquierdo del gráfico (punto 0,1), es decir, más cercano al punto del gráfico cuya sensibilidad y especificidad son del 100%.
2. Evaluar la capacidad discriminativa de un test diagnóstico. En la curva ROC, existe un área denominada área bajo la curva (AUC por sus siglas en inglés) que mide la facultad discriminativa

del test, es decir, su capacidad de diferenciar sujetos sanos versus enfermos. El AUC la podemos definir como la probabilidad de clasificar correctamente a un par de individuos (uno sano y otro enfermo) seleccionados al azar al aplicarles la prueba. Representa la probabilidad de que el resultado del ensayo para un caso positivo elegido aleatoriamente supere el resultado para un caso negativo elegido aleatoriamente. La significación asintótica es menor que 0,05, lo que significa que usar el ensayo es mejor que adivinar.

3. Comparar la capacidad discriminativa de 2 o más test diagnósticos. Para lograr este fin es necesario, en primer lugar, comprobar que se cumplen los siguientes requisitos: los test a comparar deben ser medidos simultáneamente, aplicados sobre los mismos sujetos y contrastados con el mismo patrón de oro. La comparación consiste en confrontar las respectivas AUC de cada test, siendo más discriminativo el que tenga una mayor AUC. Sin embargo, para poder afirmar que un test tiene superior capacidad discriminativa sobre otro, es preciso comparar estadísticamente ambas(44,45).

## 2.3 DEFINICIÓN DE TÉRMINOS BÁSICOS

**Hipertensión:** enfermedad crónica donde los niveles de la presión arterial se encuentran elevados en un registro cuantificado por encima de valores normales de la presión sistólica ( $PAS \geq 140$  mmHg) y/o diastólica ( $PAD \geq 90$  mmHg) (19,4).

**Factores de riesgo:** condición, comportamiento u otro factor que aumenta la posibilidad de desarrollar una enfermedad (46).

**Prevalencia:** es la proporción de una población que tiene una característica específica en un período de tiempo determinado, independientemente de cuándo desarrolló la característica por primera vez (47).

**Variable binaria o dicotómica:** es una variable categórica que solo puede tomar 2 valores o niveles diferentes, como “muerto o vivo” (48).

**Regresión logística:** Es un modelo estadístico para datos dicotómicos, se utiliza para estimar la relación entre una o más variables independientes (predictoras) y una variable de resultado binaria (dicotómica). La regresión logística es una extensión de la regresión lineal (49).

**Chi cuadrado:** Es una prueba estadística no paramétrica, se utiliza para evaluar asociación entre variables categóricas (cualitativas). Se realiza una prueba de independencia bajo la hipótesis nula y alternativa, H0 y H1, respectivamente. H0: Independiente (sin asociación) y H1: No independiente (asociación). Se usa a menudo para comprobar la independencia entre dos factores con niveles nominales (variables independientes) (49).

**Prueba Exacta de Fisher:** es una de las pruebas no paramétricas exactas y se utiliza para ver asociaciones entre variables categóricas (cualitativas), especialmente cuando más del 20% de las celdas tienen frecuencias esperadas  $< 5$ . Se usa esta prueba porque la aplicación del método de aproximación es inadecuada y porque evalúa la hipótesis nula de independencia aplicando una distribución hipergeométrica de los números en las celdas de la tabla (49).

## **2.4. HIPÓTESIS GENERAL**

Existen factores sociodemográficos y clínicos asociados a hipertensión arterial en una población de la costa del Perú.

## **CAPÍTULO III: METODOLOGÍA**

### **3.1 ENFOQUE, NIVEL Y METODOLOGÍA DE LA INVESTIGACIÓN**

Es una investigación con enfoque cuantitativo, nivel observacional, metodología retrospectiva, transversal.

### **3.2 DISEÑO DE LA INVESTIGACIÓN**

Investigación de tipo observacional, retrospectiva, transversal.

### **3.3 POBLACIÓN Y MUESTRA**

La población está compuesta por los pacientes que fueron atendidos en el Área de Prevención de la Clínica Internacional sede-Lima de enero a diciembre durante el período 2013, que corresponde a un total de 3621 pacientes.

La muestra final se obtuvo mediante un muestreo no probabilístico por conveniencia, teniendo en cuenta aquellos que habían completado la ficha de recolección de datos, estos fueron un total de 3535 pacientes.

#### **3.3.1 Criterios de Elegibilidad**

##### **Criterio de inclusión**

- Historias clínicas de pacientes de sexo femenino y masculino.
- Historias clínicas que cuenten con información completa.
- Historias clínicas de pacientes mayores de 18 años.

##### **Criterios de exclusión**

- Historias clínicas de pacientes con hipertensión que fueron atendidos en el área de Prevención de la Clínica Internacional sede-Lima en un período diferente al 2013.

### 3.4 VARIABLES

Variable dependiente: Hipertensión Arterial / Variables independientes: Factores Sociodemográficos y Factores Clínicos

#### Operacionalización de Variables

	<b>Variables</b>	<b>Definición Operacional</b>	<b>Indicadores</b>	<b>Tipo</b>	<b>Escala de medición</b>	<b>Valores</b>
Factores Clínicos cardiovasculares asociados a hipertensión	Presión arterial	Valor numérico obtenido a través de la medición de la presión sanguínea	Tensiómetro	Cuantitativa	De razón	- Presión Sistólica - Presión Diastólica
	Hipertensión Arterial	Medida según la clasificación PSIT >120		Cualitativa Dicotómica	Nominal	- No tiene: 0 - Si tiene: 1
	Colesterol	Cantidad de colesterol en el organismo	Prueba bioquímica mg/dL	Cuantitativa	De Razón	Niveles de colesterol en sangre
	Frecuencia cardiaca	Número de contracciones del corazón o pulsaciones por unidad de tiempo	Suceso cíclico	Cuantitativa	De Razón	lat/min
	Glicemia	Medida de concentración de glucosa libre en la sangre	Examen de glucosa	Cuantitativa	De Razón	mg/dL
	Talla	Crecimiento en metros de la altura del ser humano	tallímetro	Cuantitativa	De Razón	Metro
	Peso	Fuerza que ejerce un cuerpo sobre un punto de apoyo	balanza	Cuantitativa	De Razón	Kilo
Factores Sociodemográficos asociados a la Hipertensión	Edad	Número de años cronológicos vividos	DNI	Cuantitativa	De razón	Desde 18 a más
	Sexo	Características fenotípicas		Cualitativa dicotómica	Nominal	- Masculino - Femenino
	Hábito de fumar	Consumo de tabaco	Veces que consume tabaco al día	Cualitativo dicotómico	Nominal	- Sí: 1 o más - No: 0
	Práctica deportiva	Presencia de la actividad deportiva	Práctica o no deportes	Cualitativo dicotómico	nominal	- Sí: 1 vez o más - No: 0

### **3.5 INSTRUMENTOS DE LA INVESTIGACIÓN**

Se envió el proyecto al Comité de Ética de la Universidad Peruana Cayetano Heredia, SIDISI, para su respectiva evaluación y aprobación (Anexo 1). Posteriormente, se solicitó autorización a la Sede Administrativa y Oficina de Investigación de la Clínica Internacional sede-Lima para el acceso a la base de datos de las historias clínicas, las cuales fueron proporcionadas de manera electrónica.

Se analizó las historias clínicas de pacientes con hipertensión que fueron atendidos en el área de Prevención de la Clínica Internacional sede-Lima período 2013, y fueron seleccionadas aquellas con información completa, en la ficha de recolección de datos (Anexo 2).

Dichas historias clínicas poseen su sustento teórico-científico para registro de datos según la Guía Técnica: guía de práctica clínica para la prevención y control de la enfermedad hipertensiva en el primer nivel de atención (R.M.N° 491-2009/MINSA) presentada por la Dirección General de Salud del Ministerio de Salud, Lima-Perú 2011; donde se especifica las características técnicas del instrumento para medir la presión arterial, el acondicionamiento previo del paciente y el manejo óptimo de la técnica para hacer uso del instrumento de medición (Anexo 3).

Sin embargo según la Guía Técnica: guía de práctica clínica para la prevención y control de la enfermedad hipertensiva en el primer nivel de atención (R. M. N° 491-2009/MINSA) presentada por la Dirección General de Salud del Ministerio de Salud, Lima-Perú 2011; existen otros factores que se deben tener en cuenta durante la valoración del riesgo cardiovascular, ya que la presión arterial debe de ser considerada como un perfil más de riesgo de la personas (39) (Anexo 4).

Se elaboró una base de datos en Excel después de recibir la base original de parte de la Clínica Internacional sede-Lima, se eliminaron ciertos datos en base a los criterios de selección y se procedió al análisis estadístico con 3535 historias clínicas.

### **3.6 PROCESAMIENTO Y ANÁLISIS DE DATOS**

Todos los datos fueron almacenados en una base de datos de Excel evaluando los datos faltantes, datos no correspondientes y posteriormente fueron analizados con el programa Stata® versión 16 según la siguiente secuencia:

- a. Análisis descriptivo: Medidas de tendencia central y dispersión para las variables cuantitativas, y frecuencias y porcentajes para las variables cualitativas.
- b. Análisis Bivariado: para evaluar las posibles asociaciones clínicas y sociodemográficas con hipertensión arterial, se utilizó la prueba de chi cuadrado y U de Mann Whitney según corresponda.
- c. Análisis Multivariado: para evaluar posibles factores asociados a la hipertensión arterial se utilizó el modelo de regresión logística múltiple con un nivel de significancia de 0.05

Se realizó la prueba de bondad de ajuste para escoger el mejor modelo que identifique los factores asociados a hipertensión arterial.

## CAPÍTULO IV: RESULTADOS

### 4.1. CARACTERÍSTICAS DE LA MUESTRA

### 4.2. ANALISIS DESCRIPTIVO

**Tabla 1. Descripción de los pacientes atendidos en Clínica Internacional 2013. Lima – Perú**

Variables	n	%
Sexo		
Masculino	1425	40.31
Femenino	2110	59.69
Práctica deportiva		
SI	1599	45.23
NO	1936	54.77
Hábito de fumar		
SI	994	28.12
NO	2541	71.88
Hipertensión		
SI	178	5.04
NO	3357	94.96

Se evaluaron datos de 3535 pacientes, de los cuales 1425 (40.3%) fueron de sexo masculino y 2110 (59.7%) de sexo femenino. Realizan Práctica Deportiva 1599 (45.2%). Si tiene Hábito de fumar 994 (28.1%). 3357 (94.96%) no presentan hipertensión y sólo 178 (5.04%) presentan hipertensión como se representa en la tabla 1.

**Tabla 2. Descripción de las variables clínicas de los pacientes atendidos en Clínica Internacional 2013.Lima – Perú**

Variables	media $\pm$ DS	Mediana	RIC	Mín.	Máx.	Skewness	Kurtosis
Edad	46.63 $\pm$ 12.08	44.2	17.4	22	95.7	0.77	3.26
Presión arterial							
Sistólica	112.85 $\pm$ 14.98	110	20	80	200	1.04	4.87
Diastólica	73.43 $\pm$ 9.72	70	10	50	130	0.73	4.26
Frecuencia cardíaca	71.52 $\pm$ 6.04	72	8	48	112	0.63	6.04
Talla	1.64 $\pm$ 0.09	1.63	0.13	1.35	1.96	0.27	2.75
Peso	68.87 $\pm$ 14.17	67	20	40	143.2	0.77	3.85
Glicemia	90.15 $\pm$ 16.54	88	11	57	392	8.88	124.31
Colesterol total	189.45 $\pm$ 38.47	186	51	19	435	0.59	3.87

RIC: es el rango intercuartil, obtenido como diferencia entre el tercer y el primer cuartil de una distribución.

DS: desviación estándar, Skewness: sesgo, Kurtosis: distribución de la variable

La edad media  $\pm$  DS de los pacientes evaluados fue de 46.63  $\pm$  12.08. Presión sistólica media 112.85 mmHg  $\pm$  14.98, presión diastólica media 73.43 mmHg  $\pm$  9.72, frecuencia cardiaca promedio 71.52  $\pm$  6.04, talla promedio 1.64 m  $\pm$  0.09, peso medio 68.87 kg.  $\pm$  14.17, glicemia promedio 90.15 mg/dL  $\pm$  16.54, colesterol total promedio 189.45 mg/dL  $\pm$  38.47 como se presenta en la tabla 2.

### 4.3. ANÁLISIS BIVARIADO

**Tabla 3. Distribución de las variables demográficas según hipertensión arterial. Clínica Internacional 2013, Lima Perú.**

Variable	HTA		Total	P*
	Si	No		
Sexo				
Masculino	117 (3.23)	1340 (37.01)	1457(40.24)	0.000
Femenino	66 (1.82)	2098 (57.94)	2164(59.76)	
Hábito de fumar				
Si	41 (1.16)	953 (26.96)	994 (28.12)	0.121
No	137 (3.88)	2404 (68.01)	2541(71.89)	
Práctica deportiva				
Si	91 (2.51)	1547 (42.72)	1638 (45.23)	0.210
No	92 (2.54)	1891 (52.22)	1983(54.76)	

\* Prueba de Chi Cuadrado.  $P < 0.05$ . indica asociación significativa

Al evaluar la asociación de hipertensión arterial con cada una de las variables demográficas. solo se encontró asociación significativa con la variable sexo. ( $p < 0.0001$ ). importante para considerar en el modelo de regresión logística múltiple. No se encontró asociación entre HTA con las variables hábito de fumar y práctica deportiva. ( $P > 0.05$ ) Tabla 3

**Tabla 4. Comparación de edad, colesterol, frecuencia cardiaca, glicemia, talla y peso según hipertensión arterial. Clínica Internacional 2013. Lima - Perú.**

	media	DS	mediana	RIC	Min-Max	skewness	kurtosis	p
<b>Edad<sup>1</sup></b>								0.00
<b>No HTA</b>	46.32	13.34	58.2	17.9	29.5-95.7	0.08	2.6	
<b>HTA</b>	46.01	11.7	43.7	16.5	22-94.9	0.79	3.35	
<b>Total</b>	46.63	12.08	44.2	17.4	22-95.7	0.77	3.26	
<b>Colesterol<sup>2</sup></b>								0.00
<b>No HTA</b>	188.56	38.25	185	50	19-435	0.61	3.95	
<b>HTA</b>	206.14	38.95	206	58	19-435	0.28	3.27	
<b>Total</b>	189.45	38.47	186	51	112-353	0.59	3.87	
<b>Frecuencia cardiaca<sup>3</sup></b>								0.01
<b>No HTA</b>	71.45	5.93	72	7	48-112	0.6	5.96	
<b>HTA</b>	72.79	7.71	72	8	48-104	0.63	5.51	
<b>Total</b>	71.52	6.04	72	8	48-112	0.63	6.04	
<b>Glicemia<sup>4</sup></b>								0.00
<b>NO HTA</b>	89.64	15.73	88	11	57-392	9.49	143.91	
<b>HTA</b>	99.72	25.93	94	14	75-320	5.1	36.37	
<b>Total</b>	90.15	16.54	88	11	57-392	8.88	124.31	
<b>Talla<sup>5</sup></b>								0.00
<b>NO HTA</b>	1.63	0.09	1.63	0.13	1.37-1.96	0.31	2.79	
<b>HTA</b>	1.66	0.1	1.67	0.15	1.35-1.91	-0.37	2.68	
<b>Total</b>	1.64	0.09	1.63	0.13	1.35-1.96	0.27	2.75	
<b>Peso<sup>6</sup></b>								0.00
<b>NO HTA</b>	68.23	13.77	66.4	19	40-137	0.74	3.73	
<b>HTA</b>	80.91	16.06	80	21.6	50.5-143.2	0.69	3.91	
<b>Total</b>	68.87	14.17	67	20	40-143.2	0.77	3.85	

DS: Desviación estándar <sup>1</sup>Prueba U Mann-Whitney. z = -11.342, <sup>2</sup>Prueba U Mann-Whitney. z = -11.342 <sup>3</sup>Prueba U Mann-Whitney. z = -2.723, <sup>4</sup>Prueba U Mann-Whitney. z = -9.235, <sup>5</sup>Prueba U Mann-Whitney. z = -4.230 <sup>6</sup>Prueba U Mann-Whitney. z = -10.506 p<0.05 significativo

Al comparar cada una de las variables clínicas y edad según hipertensión arterial. se encontró diferencias significativas entre los hipertensos y no hipertensos respecto a edad (p=0.001). colesterol

(p=0.001). frecuencia cardiaca (p=0.006). glicemia(p=0.001). talla (p=0.001). y peso (p=0.001). resultados importantes para considerar en el modelo de regresión logística múltiple. (Tabla 4).

#### 4.4. ANALISIS MULTIVARIADO:

#### MODELO DE REGRESIÓN LOGISTICA MULTIPLE

**Tabla 5: Determinación de los factores sociodemográficos asociados a hipertensión usando regresión logística múltiple. Clínica Internacional 2013. Lima Perú**

Hipertensión	Coef.	OR <sub>crudo</sub>	IC 95%		p	OR <sub>ajustado</sub>	IC 95%		p
Sexo	-0.91	0.4	0.26	0.49	0.00	0.4	0.29	0.56	0.00
Edad	0.06	1.1	1.06	1.08	0.00	1.1	1.05	1.08	0.00
Hábito de fumar	-0.02	0.8	0.53	1.08	0.00	0.9	0.68	1.43	0.94
Práctica deportiva	-0.09	1.3	0.93	1.69	0.14	0.9	0.66	1.26	0.57
constante	-5.78								

Coef: Coeficiente de regresión logística

Para evaluar factores sociodemográficos en base al modelo de regresión logística múltiple, se muestra en la tabla los resultados obtenidos. El modelo logit, se resume en la siguiente formula:

$$\text{logit} \frac{p}{1-p} = -5.78 - 0.91 (\text{sexo}) + 0.06 (\text{edad}) - 0.02 (\text{hábito de fumar}) - 0.09 (\text{deporte})$$

Que también se genera el modelo de regresión logística siguiente:

$$P(X_i) = \frac{e^{(-5.78 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{habito de fumar} - 0.09 \text{ deporte})}}{1 + e^{(-5.78 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{habito de fumar} - 0.09 \text{ deporte})}}$$

En cuanto a los factores sociodemográficos. este modelo nos explica que las variables sexo y edad están asociadas con hipertensión arterial (P< 0.001), manteniendo constante las demás variables.

Para el análisis multivariado. se consideraron en el modelo de regresión logística las variables demográficas edad. sexo. hábito de fumar. practica deporte para evaluar si están asociadas con

hipertensión. encontrando asociación significativa en las variables sexo y edad ajustadas por las demás variables ( $p < 0.001$ ) sin embargo con las variables hábito de fumar y practica deporte no se encontró asociación significativa ( $P > 0.05$ ). La variable sexo, con  $OR = 0.4$  se interpreta como un factor protector IC 95% (0.29; 0.56) lo que nos indica que hay 2.5 veces de riesgo de tener HTA en los hombres respecto a las mujeres. La variable edad. con  $OR = 1.1$  se interpreta como un factor de riesgo significativo IC 95% (1.05-1.08). Lo que nos indica que, por cada año de incremento en la edad, el riesgo de HTA se incrementa en un 10%. La variable hábito de fumar con un  $OR = 0.9$  se interpreta como un factor protector no significativo IC 95% (0.68-1.43) Lo que nos indica que hay un riesgo de 1.1 de tener HTA en las personas que fuman respecto a las personas que no fuman. La variable práctica deportiva con  $OR = 0.9$  se interpreta como un factor protector. no significativo IC 95% (0.66-1.26), Lo que nos indica que las personas que realizan deporte tienen 1.11 menor riesgo de HTA con respecto a las que no practican deporte.

Salida cruda en Stata® v.17. Anexo 5

**Tabla 6: Evaluación de los factores clínicos asociados a hipertensión según el modelo de regresión logística múltiple. Clínica Internacional 2013. Lima Perú**

Hipertensión	Coefficiente	OR <sub>crudo</sub>	IC 95%	p	OR <sub>ajustado</sub>	IC 95%	p		
Colesterol total	0.01	1	1	1.01	0.00	1.01	1.004	1.01	0.00
Frecuencia cardiaca	0.04	1	1.01	1.06	0.00	1.02	0.99	1.05	0.06
Glicemia	0.02	1.01	1.01	1.02	0.00	1.01	1.003	1.01	0.00
Talla	3.31	27.6	5.2	146.05	0.00	0.02	0.003	0.24	0.00
Peso	0.05	1.05	1.04	1.06	0.00	1.06	1.05	1.08	0.00
constante	-5.7								

(Prob > chi2 = 0.0000). Pseudo R2 (0. 1180)

El modelo logit de la regresión logística es la siguiente en base a los resultados de la tabla:

$$\ln\left(\frac{p}{1-p}\right) = -5.7 + 0.01(\text{Colesterol total}) + 0.04(\text{Frecuencia cardiaca}) + 0.02(\text{glicemia}) + 3.31 (\text{talla}) + 0.05 (\text{peso})$$

y el modelo de regresión logística seria:

$$P(X_i) = \frac{e^{(-5.7 + 0.01 \text{colesterol total} + 0.04 \text{frecuencia cardiaca} + 0.02 \text{glicemia} + 3.31 \text{talla} + 0.05 \text{peso})}}{(1 + e^{(-5.7 + 0.01 \text{colesterol total} + 0.04 \text{frecuencia cardiaca} + 0.02 \text{glicemia} + 3.31 \text{talla} + 0.05 \text{peso})})}$$

En cuanto a los factores clínicos, este modelo nos explica que las variables colesterol (p=0.00), glicemia (p=0.00), talla (p=0.00) y peso (p=0.00) están asociadas con la hipertensión arterial. Este modelo es válido, ya que presenta un valor estadísticamente significativo que nos indicaría que el modelo explica un 11.8% la asociación entre dichas variables con la hipertensión arterial.

Como se puede apreciar el OR<sub>crudo</sub>: 1 para la variable Colesterol total y OR<sub>ajustado</sub>: 1.01 no se ve afectada por las otras variables, Frecuencia cardiaca, Glicemia, talla y peso. Cabe resaltar que los intervalos de confianza al 95% no incluye a la unidad. Así mismo se observa en el OR<sub>crudo</sub>: 1 para la variable Frecuencia Cardiaca y el OR ajustado: 1.02, interpretándose como un factor de riesgo no significativo IC 95% (1; 1.05).

**Tabla 7: A: Factores sociodemográficos y clínicos asociados a hipertensión según el modelo de regresión logística múltiple. Clínica Internacional 2013. Lima Perú**

Hipertensión	Coefficiente	OR <sub>crudo</sub>	IC 95%		p	OR <sub>ajustado</sub>	IC 95%		p
Sexo	-0.91	0.4	0.26	0.49	0.00	0.78	0.49	1.25	0.31
Edad	0.06	1.1	1.06	1.08	0.00	1.07	1.05	1.08	0.00
Hábito de fumar	-0.02	0.8	0.53	1.08	0.00	0.83	0.56	1.23	0.36
Práctica deportiva	-0.09	1.3	0.93	1.69	0.14	1.09	0.78	1.51	0.61
Colesterol total	0.01	1.0	1.0	1.01	0.00	1.01	1.001	1.009	0.01
Frecuencia cardiaca	0.04	1.0	1.01	1.06	0.00	1.03	1.01	1.05	0.04
Glicemia	0.02	1.01	1.01	1.02	0.00	1.00	1.00	1.01	0.09
Talla	3.31	27.6	5.2	146.05	0.00	0.09	0.01	1.39	0.09
Peso	0.05	1.05	1.04	1.06	0.00	1.06	1.05	1.08	0.00
Constante	0.00								

Según el análisis multivariado, utilizando el modelo de regresión logística. se aplicó el método de máxima verosimilitud y la estrategia pasos hacia atrás (stepwise) para eliminar variables que no contribuyen al modelo ( $\alpha = 0.05$ ) es decir elegimos un modelo parsimonioso.

El modelo final de factores clínicos y sociodemográficos calculando los coeficientes con el comando logit asociados a Hipertensión es

$$\ln\left(\frac{p}{1-p}\right) = 0.00 - 0.91 (\text{sexo}) + 0.06 (\text{edad}) - 0.02 (\text{hábito de fumar}) - 0.09 (\text{practica deporte}) + 0.01 (\text{colesterol total}) + 0.04 (\text{frecuencia cardiaca}) + 0.02 (\text{glicemia}) + 3.31 (\text{talla}) + 0.05 (\text{peso})$$

Y el modelo de regresión logística final sería:

$$P(X_i) =$$

$$\frac{e^{(0.00 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{hábito de fumar} - 0.09\text{practica deporte} + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso})}}{(1 + e^{(0.00 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{hábito de fumar} - 0.09\text{practica deporte} + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso})})}$$

En el modelo de regresión logística múltiple final ( $p = 0.001$ ). Se consideró la edad (OR. 1.07 IC95%. 1.05-1.08). Colesterol total (OR. 1.01 IC95%. 1.001-1.01). Frecuencia cardiaca (OR. 1.03 presentar hipertensión arterial).

Como se puede apreciar el  $OR_{\text{crudo}}$ : 0.4 para la variable sexo  $OR_{\text{ajustado}}$ : 0.78 se ve afectada por las otras variables edad, hábito de fumar, práctica deportiva, colesterol total, frecuencia cardiaca, glicemia, talla y peso. La variable sexo representa un factor protector; interpretándose como la probabilidad de 23% menos riesgo de los hombres de presentar HTA respecto a las mujeres. Se observa que el  $OR_{\text{crudo}}$ : 1.1 para la variable edad  $OR_{\text{ajustado}}$ : 1.07 no se ve afectada por las otras variables sexo, hábito de fumar, práctica deportiva, colesterol total, frecuencia cardiaca, glicemia, talla y peso. La variable edad representa un factor de riesgo interpretándose como la probabilidad de presentar HTA un 7 % más por cada incremento en un año de edad. Además. se puede apreciar el  $OR_{\text{crudo}}$ : 0.8 para la variable Hábito de fumar el  $OR_{\text{ajustado}}$ : 0.83 no se ve afectada por las otras variables sexo, edad, práctica deportiva, colesterol total, frecuencia cardiaca, glicemia, talla y peso. La variable hábito de fumar representa un factor protector no significativo IC 95% (0.53 - 1.23)

interpretándose como la probabilidad del 17% menos riesgo de las personas que no fuman de presentar HTA respecto a las que fuman. Así mismo. se puede apreciar el  $OR_{\text{crudo}}$ : 1.3 para la variable Práctica deportiva el  $OR_{\text{ajustado}}$ : 1.09 se ve afectada por las otras variables sexo, edad, hábito de fumar, colesterol total, frecuencia cardiaca, glicemia, talla y peso. La variable práctica deportiva es un factor de riesgo no significativo IC 95% (0.78 - 1.51) lo que interpreta como la probabilidad de presentar HTA un 9% más por cada incremento en un año para aquellas personas que no practican deporte. Asimismo. se puede apreciar el  $OR_{\text{crudo}}$ : 1 para la variable Colesterol total el  $OR_{\text{ajustado}}$ : 1.01 no se ve afectada por las otras variables sexo, edad, hábito de fumar, práctica deportiva, frecuencia cardiaca, glicemia, talla y peso. La variable colesterol representa un factor de riesgo significativo IC 95% (1.001 - 1.009) lo que se interpreta como la probabilidad de presentar HTA un 1% más por cada incremento en un año.

Como se puede apreciar el  $OR_{\text{crudo}}$ : 1 para la variable Frecuencia cardiaca el  $OR_{\text{ajustado}}$ : 1.03 no se ve afectada por las otras variables sexo, edad, hábito de fumar, práctica deportiva, colesterol total, Glicemia, talla y peso. La variable frecuencia cardiaca de riesgo no significativo IC 95% (1.01 – 1.05). Lo que se interpreta como la probabilidad de presentar HTA en 3% más por cada incremento en un año para aquellas personas valores alterados de frecuencia cardiaca. Se observa que. el  $OR_{\text{crudo}}$ : 1.01 para la variable glicemia el  $OR_{\text{ajustado}}$ : 1 no se ve afectada por las otras variables sexo, edad, hábito de fumar, práctica deportiva, colesterol total, frecuencia cardiaca, talla y peso. Se observa que. el  $OR_{\text{crudo}}$ : 27.6 para la variable talla el  $OR_{\text{ajustado}}$ : 0.09 se ve considerablemente afectada por las otras variables sexo, edad, hábito de fumar, práctica deportiva, colesterol total, frecuencia cardiaca, glicemia y peso. Finalmente se aprecia que. el  $OR_{\text{crudo}}$ : 1.05 para la variable peso el  $OR_{\text{ajustado}}$ : 1.06 se ve considerablemente afectada por las otras variables sexo, Edad, hábito de fumar, práctica deportiva, colesterol total, frecuencia cardíaca, glicemia y talla. Lo que se interpreta como la probabilidad de presentar HTA en un 6%. Por lo que se recomienda considerar todas las variables asociadas a HTA y no solamente los análisis bivariados.

Cabe resaltar que si bien es cierto q al final incorpora todas las variables predictoras independientemente de su significancia estadística, prima el criterio epidemiológico dentro del modelo final.

### **Evaluación de supuestos del modelo:**

En el modelo logit:

$$\ln\left(\frac{p}{1-p}\right) = \underline{0.00 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{hábito de fumar} - 0.09\text{practica deporte} + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso}}$$

Y el modelo de regresión logística final:

$$P(X_i) = \frac{e(\underline{0.00 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{hábito de fumar} - 0.09\text{practica deporte} + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso}})}{(1 + e(\underline{0.00 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{hábito de fumar} - 0.09\text{practica deporte} + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso}))}$$

### **Prueba de bondad de ajuste:**

La prueba de bondad de ajuste en el contexto de la regresión logística se utiliza para evaluar la adecuación del modelo de regresión logística a los datos observados. La prueba de bondad de ajuste específica para la regresión logística se basa en la comparación entre las frecuencias observadas y las frecuencias esperadas según el modelo ajustado.

Si la estadística de prueba es significativa, se puede concluir que el modelo de regresión logística no se ajusta adecuadamente a los datos observados, lo que sugiere que puede haber deficiencias en el modelo o que la relación entre las variables no se está modelando correctamente. Si la estadística de prueba no es significativa, no hay suficiente evidencia para rechazar la hipótesis nula, indicando que el modelo de regresión logística es adecuado para explicar los datos observados.

Es importante destacar que existen varias formas de realizar la prueba de bondad de ajuste en el contexto de la regresión logística. y la elección específica puede depender del software estadístico utilizado y de los detalles del análisis.

### **La prueba de bondad de ajuste de Pearson**

Evalúa la discrepancia entre el modelo actual y el modelo completo.

Se obtiene en base a los valores observados y esperados entre todas las variables.

Es una prueba estadística que se utiliza para determinar si una muestra observada proviene de una población con una distribución esperada específica. Esta prueba compara las frecuencias observadas con las frecuencias esperadas para ver si hay diferencias significativas entre ellas. Aquí tienes una descripción general del proceso:

Formulación de hipótesis:

Hipótesis nula ( $H_0$ ): Las frecuencias observadas no difieren significativamente de las frecuencias esperadas. En otras palabras, los datos siguen la distribución esperada.

Hipótesis alternativa ( $H_1$ ): Las frecuencias observadas difieren significativamente de las frecuencias esperadas.

Comparar el valor calculado de  $\chi^2$  con el valor crítico de  $\chi^2$  de la tabla de distribución chi-cuadrado, utilizando el nivel de significancia ( $\alpha$ ) y los grados de libertad (df), que se calculan como el número de categorías menos 1.

Decisión:

- Si el valor calculado de  $\chi^2$  es mayor que el valor crítico, se rechaza la hipótesis nula, indicando que hay una diferencia significativa entre las frecuencias observadas y las esperadas.
- Si el valor calculado de  $\chi^2$  es menor o igual al valor crítico, no se rechaza la hipótesis nula, indicando que no hay una diferencia significativa.

**Utilizando stata, se tiene la siguiente salida:**

Modelo logístico para la hipertensión. prueba de bondad de ajuste

número de observaciones	=	3535
número de patrones de covariables	=	1971
Chi2 de Pearson (1966)	=	1829.73
Prob > chi2	=	0.9866

Prob > chi2 = 0.9866

No se rechaza la Ho.

La prueba no es útil cuando el número de valores distintos (patrones) es aproximadamente igual al número de observaciones, En este caso se obtuvo 1971 patrones, son demasiados, por lo que se recomienda utilizar la bondad de ajuste de Hosmer y Lemeshow

### **Prueba de bondad de ajuste de Hosmer-Lemeshow**

La prueba de Hosmer-Lemeshow es una herramienta útil para evaluar la bondad de ajuste en modelos de regresión logística. pero también tiene sus limitaciones y alternativas. por lo que es importante considerar otras medidas de ajuste del modelo y explorar la validez del modelo en su conjunto.

**Cálculo del estadístico de Hosmer-Lemeshow:**

- El estadístico de chi-cuadrado de Hosmer-Lemeshow se calcula utilizando la siguiente fórmula:

$$\chi_{HL}^2 = \sum_{g=1}^G \left( \frac{(O_{g1} - E_{g1})^2}{E_{g1}} + \frac{(O_{g0} - E_{g0})^2}{E_{g0}} \right)$$

- Aquí,  $O_{g1}$  y  $O_{g0}$  son las frecuencias observadas de eventos y no eventos en el grupo  $g$ , y  $E_{g1}$  y  $E_{g0}$  son las frecuencias esperadas de eventos y no eventos en el grupo  $g$ .

• **División de los datos en grupos:**

- Se dividen los datos en  $G$  grupos (generalmente deciles) basados en las probabilidades predichas. Cada grupo debería contener aproximadamente el mismo número de observaciones.

• **Cálculo de frecuencias observadas y esperadas:**

- Para cada grupo  $G$  (donde  $g = 1, 2, \dots, G$ ), se cuentan las frecuencias observadas de eventos ( $O_{g1}$ ) y no eventos ( $O_{g0}$ ).
- Las frecuencias esperadas de eventos ( $E_{g1}$ ) y no eventos ( $E_{g0}$ ) se calculan multiplicando el número total de observaciones en cada grupo por la probabilidad promedio predicha en ese grupo.

• **Formulación de hipótesis:**

- **Hipótesis nula ( $H_0$ ):** El modelo se ajusta bien a los datos observados, es decir, no hay diferencias significativas entre las frecuencias observadas y esperadas.
- **Hipótesis alternativa ( $H_1$ ):** El modelo no se ajusta bien a los datos observados, es decir, hay diferencias significativas entre las frecuencias observadas y esperadas.

• **Determinación del valor crítico y comparación:**

- El estadístico  $\chi_{HL}^2$  se compara con un valor crítico de la distribución chi-cuadrado con  $G-2$  grados de libertad, donde  $G$  es el número de grupos.
- Si el estadístico calculado es mayor que el valor crítico, se rechaza la hipótesis nula, indicando que el modelo no se ajusta bien a los datos.
- Si el estadístico calculado es menor o igual que el valor crítico, no se rechaza la hipótesis nula, indicando que no hay evidencia suficiente para afirmar que el modelo no se ajusta bien a los datos.
- Usando stata, la salida es la siguiente:
- 
- La tabla siguiente de los grupos formados para esta prueba es la siguiente:

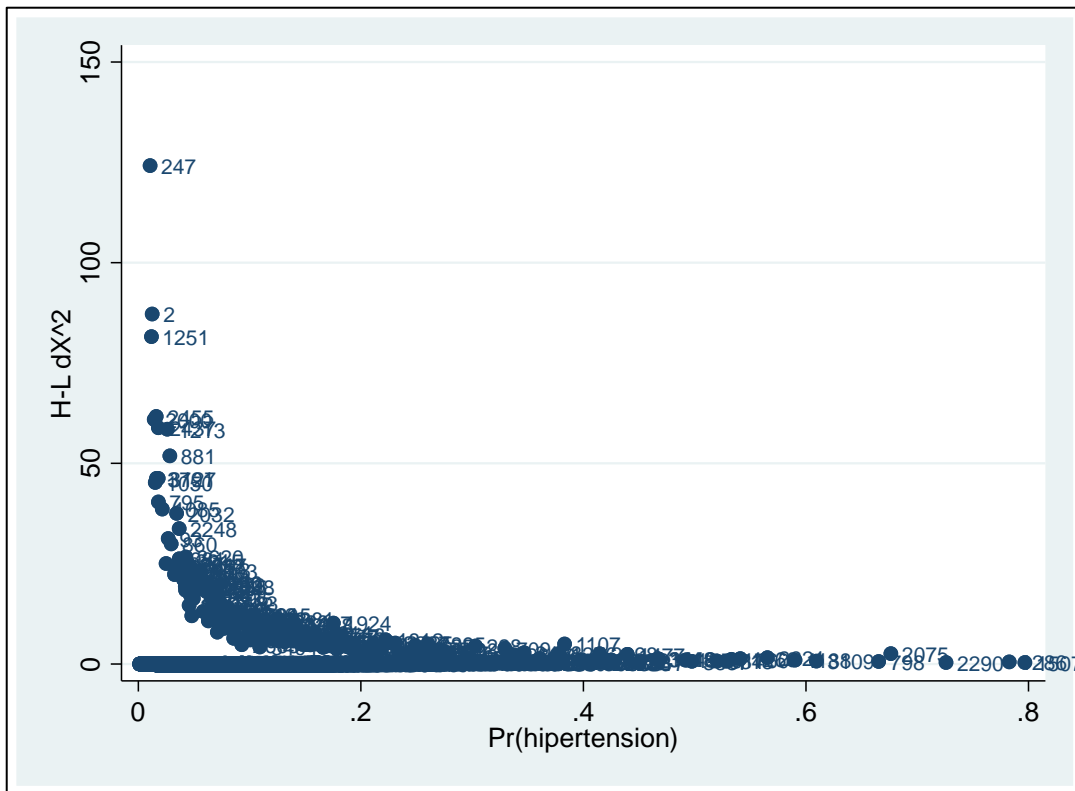
(Tabla colapsada en cuantiles de probabilidades estimadas)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.0114	1	3.4	354	351.6	355
2	0.0144	1	4.5	352	348.5	353
3	0.0190	4	5.9	352	350.1	356
4	0.0253	9	7.9	342	343.1	351
5	0.0316	5	10.1	352	346.9	357
6	0.0391	14	12.3	336	337.7	350
7	0.0493	20	15.6	335	339.4	355
8	0.0665	20	20.1	331	330.9	351
9	0.1140	47	30.8	308	324.2	355
10	0.6065	57	67.3	295	284.7	352

Número de observaciones = 3535  
 Número de grupos = 10  
 Hosmer-Lemeshow chi2(8) = 20.69  
 Prob > chi2 = 0.0080

Se puede apreciar dos frecuencias esperadas menores a 5. lo que nos indica que sería una limitación para aceptar como válidos los resultados de esta prueba chi2.

## Grafica de Hosmer y Lemeshow



El delta-Chi-cuadrado ( $dx^2$ ) es un estadístico de influencia. que nos permite ubicar la presencia de los puntos de influencia. Al revisar el gráfico. destaca entre todos los casos, el número 247 que está muy extremo.

### Prueba de máxima verosimilitud:

La regresión logística se ajusta a los datos utilizando el método de máxima verosimilitud (Maximum Likelihood Estimation. MLE). La idea central del método de máxima verosimilitud es encontrar los valores de los parámetros del modelo que maximizan la probabilidad de observar los datos observados bajo el modelo especificado. En el contexto de la regresión logística, se busca encontrar los valores de los coeficientes que maximizan la probabilidad de observar los resultados de la variable dependiente condicionada a las variables independientes.

La prueba de máxima verosimilitud en la regresión logística se centra en comparar modelos alternativos al evaluar si un modelo más complejo (con más parámetros) mejora significativamente la verosimilitud en comparación con un modelo más simple. Esto se hace a menudo utilizando la prueba de razón de verosimilitudes (Likelihood Ratio Test. LRT).

En resumen, la estimación de parámetros en la regresión logística mediante máxima verosimilitud implica encontrar los valores que maximizan la probabilidad de observar los datos dados los parámetros del modelo. La prueba de razón de verosimilitudes luego compara modelos para evaluar la adecuación del modelo a los datos.

Medidas Fit por logit de hipertensión:

Log-Lik Intercept Only:	-705.427	Log-Lik Full Model:	-620.237
D(3530):	1240.474	LR(4):	170.381
		Prob > LR:	0.000
McFadden's R2:	0.121	McFadden's Adj R2:	0.114
Maximum Likelihood R2:	0.047	Cragg & Uhler's R2:	0.143
McKelvey and Zavoina's R2:	0.206	Efron's R2:	0.046
Variance of y*:	4.144	Variance of error:	3.290
Count R2:	0.949	Adj Count R2:	-0.006
AIC:	0.354	AIC*n:	1250.474
BIC:	-27601.280	BIC':	-137.699

De la tabla presentada lo más resaltante del modelo ajustado podemos mencionar:

Log-Lik Full Model = -620.237, en términos de verosimilitud, el modelo ajustado tiene una probabilidad relativamente alta de producir los datos observados.

McFadden's Adjusted R<sup>2</sup> de 0.114 indica que el modelo explica una proporción moderada de la variabilidad en los datos en comparación con el modelo nulo.

Prob > LR=0.000 es la probabilidad asociada a la prueba de razón de verosimilitud en la comparación de modelo. Esto implica que hay diferencias significativas entre el modelo ajustado y el modelo nulo.

AIC normalizado (AIC\*n) =1250.474 se utiliza para comparar modelos alternativos. Al igual que el AIC estándar, un valor más bajo de AIC\*n indica un mejor ajuste del modelo

BIC' = -137.699, BIC (Criterio de Información Bayesiano) ajustado o estandarizado. El BIC es una medida utilizada para la selección de modelos en estadísticas, en comparación con otros modelos, el modelo ajustado tiene un mejor equilibrio entre el ajuste del modelo y la penalización por complejidad.

### Modelo de clasificación:

Modelo logístico para la hipertensión.

```

----- Verdad -----
Clasificada |          D          ~D |          Total
-----+-----+-----+-----
          + |          1          2 |          3
          - |         177        3355 |         3532
-----+-----+-----+-----
Total      |         178        3357 |         3535
    
```

Classified + if predicted Pr(D) >= .5

True D defined as hipertension != 0

```

-----
Sensibilidad                Pr( +| D)    0.56%
Especificidad              Pr( -| ~D)   99.94%
Valor predictivo positivo   Pr( D| +)   33.33%
Valor predictivo negativo   Pr(~D| -)   94.99%
-----
Falso + tasa para verdadero ~D Pr( +| ~D)   0.06%
Falso - tasa para verdadero D  Pr( -| D)   99.44%
Falso + tasa para clasificados + Pr(~D| +)   66.67%
Falso - tasa para clasificados - Pr( D| -)   5.01%
-----
Correctamente clasificado                                     94.94%
-----
    
```

En esta prueba muestra nuevamente la bondad de ajuste del modelo. esta vez considerando un punto de corte para la valoración de las predicciones (0.5). Aquí se está evaluando su capacidad de clasificar con una alta precisión; se observa que el 94.94% de los casos se clasifican correctamente). (Proporción de verdaderos positivos + verdaderos negativos entre el total de la muestra).

## DISCUSIÓN

La metodología empleada en el presente estudio cumple con las normas de ética requeridas para trabajar con los datos de pacientes que han sido atendidos en un Centro de Salud particular. Primero se realizó la evaluación del proyecto por el Comité de Ética de la UPCH. SIDISI. Después se solicitó autorización de la Clínica Internacional sede-Lima para el acceso a la base de datos de las historias clínicas de pacientes con hipertensión que fueron atendidos en el área de Prevención de esta institución.

En los estudios realizados por Saeed et al (10) y Nam et al (11). en el 2020. se utilizó el análisis de regresión logística. considerando a la hipertensión arterial como variable dependiente y a los factores de riesgos como variables independientes. En ambos estudios. se encontró una alta prevalencia de hipertensión arterial. Sin embargo. en el primer estudio. la hipertensión arterial se asoció a la variable edad y obesidad. coincidiendo con el segundo estudio en la variable edad. más no en las otras variables asociadas encontradas en el segundo estudio como son: el consumo excesivo de alcohol y sal. antecedentes de hipertensión arterial y presencia previa de alguna enfermedad. Coincidiendo en parte con los resultados en nuestro estudio en una población peruana.

Otro estudio realizado por Odame et al (16). en el 2019. que también utilizó el análisis de regresión logística. encontró que la hipertensión arterial se asocia a estilos de vida y comportamientos individuales. Estos resultados no coincidieron con los estudios realizados por Ghosh y Kumar (19) y Mokgwathi y Chacha (21) en el 2019. donde se encontraron factores de riesgos asociados a la hipertensión arterial similares tales como obesidad/sobrepeso y consumo de alcohol. Los estudios mencionados tienen resultados similares a los encontrados en el presente estudio.

En el año 2020 Alvarenga et al (12). describen la prevalencia y los factores asociados con la hipertensión arterial sistémica en adultos y ancianos; utilizando como análisis multivariado la regresión de Poisson con varianzas robustas. y coinciden con los estudios realizados por Saeed et al (10). y Nam et al (11). que también evalúan la edad. la obesidad abdominal y la hiperglucemia como

factores de riesgo arterial. Por otro lado. Pinto et al (15). con el mismo método. analizaron la asociación de la hipertensión arterial y su baja asociación con factores sociodemográficos. medidas antropométricas y estilos de vida. analizando a la falta de educación como variable con mayor asociación. Según metodología no coinciden con los resultados de nuestro estudio en población peruana. En el estudio de Pinto se estudió la asociación de la hipertensión y las limitaciones asociadas (definidas como aquellas complicaciones de la hipertensión que no te dejan desarrollar tus actividades usuales) con factores sociodemográficos (sexo. raza. edad de 30 a 60 años. nivel de educación) y epidemiológicos en pacientes brasileños de 30 años de edad o más. en nuestro estudio no vemos factores epidemiológicos. Además, que analizan otros temas tales como depresión evaluados con una encuesta validada. dolores de pecho al caminar. diagnósticos de enfermedad cardíaca. entre otros. En el presente estudio nos hemos basado en una base de datos que difiere en la metodología aplicada por la de Pinto. que nos marca una pauta para en futuros estudios poder realizar investigaciones con mayores variables e interacción con el paciente.

Asimismo. Thepo et al (13). investigaron la prevalencia de prehipertensión. hipertensión arterial y la asociación de la presión arterial con el IMC y área de la superficie corporal mediante la correlación de Pearson. concluyendo que la prehipertensión y la hipertensión arterial es mayor en varones que en mujeres y esta ligeramente asociada a IMC. Además. en trabajos como el de Jakab et al (17).; Kumar y Verma (18); El Achab et al (20). en el 2019. se utilizó el análisis estadístico de prevalencia en el primer caso y el análisis bivariado cualitativo como el chi-cuadrado en los demás casos, hallando cierta relación importante entre el peso la hipertensión y factores asociados entre los adolescentes. respectivamente. Coincidiendo en nuestro estudio que se encontró también dichos factores como edad y peso. Los que nos hace inferir que si se hubiese utilizado el análisis multivariado de regresión podrían haber encontrado resultados más específicos.

En este estudio como aspirante a la especialidad de estadística, deseo saber cuáles son las variables que mejor predicen la ocurrencia o no de la hipertensión arterial (y). bajo ese objetivo se determinó y midió las siguientes variables independientes del modelo ( $x$ 's): factores clínicos y sociodemográficos y declaro independencia entre evaluados, mediciones válidas y confiables, y que su muestreo es no probabilístico por conveniencia.

El modelo de regresión logística tiene que cumplir supuestos que nos indican y nos van dando la ruta de cuál de los modelos planteados predice mejor la ocurrencia o no de la hipertensión arterial. Para lograr esto y obtener el modelo de regresión logística final se ha utilizado el análisis de regresión múltiple (modelado inicial – logit) con la inclusión de todas las variables independientes, la evaluación de supuestos, la estimación de los puntos de influencia, el modelado final (bajo contraste del mejor ajuste y considerando los puntos de influencia detectados) y los algoritmos de selección “backward stepwise” y LR Test.

Una vez teniendo el modelo inicial se procedió a la evaluación de supuestos: la bondad del ajuste de Pearson: en éste se tuvo en cuenta la  $H_0$ : El modelo de regresión logística se ajusta bien a los datos y la  $H_a$ : el modelo de regresión logística no se ajusta bien a los datos; se encontró un valor de  $p > 0.05$ , por lo que no se rechaza la  $H_0$ . Además, la tabla de los grupos formados para las pruebas que dio como resultado un valor de  $\text{prob} > \chi^2 = 0.089$ ;  $p$  con un valor  $< 0.05$  lo que nos indica que se rechaza la  $H_0$  y nos quedamos con la  $H_a$ . Para finalmente obtener un resumen con las pruebas de bondad de ajuste; la Prueba de Hosmer – Lemeshow indicó el ajuste apropiado para el modelo. Los gráficos de la curva de ROC (0.7889) como la de LSens nos permitieron valorar visualmente la especificidad y sensibilidad del modelo versus el corte de la probabilidad del evento evaluado (aproximadamente 0.75). La multicolinealidad es un supuesto que involucra solo a las variables independientes, la prueba resulta funcional para el objetivo de este estudio. Los resultados del VIF, teniendo en cuenta los resultados del VIF (1.05) mayores a 10 y con media baja e inversas balanceadas, permite concluir la no colinealidad de las variables explicativas. Los estadísticos de

influencia permiten generar la probabilidad estimada por el modelo logístico inicial. Los gráficos como: los de residuos extremos que pueden estar afectando al modelo. los del delta-Chi-cuadrado que es un estadístico de influencia desarrollado por Hosmer y Lemeshow (1989). Los del DBeta de Pregibon (1981) que es considerado un equivalente logístico de la Distancia de Cook. permiten la ubicación de puntos de influencia (caso 247. 2840) y son supuestos que se deben cumplir para exponer modelos alternativos. que se tuvieron y tomaron en cuenta en el modelo del presente estudio. Sin embargo, siempre se debe tener en cuenta las desventajas del análisis de la Regresión Logística ya que muestra la imposibilidad de resolver directamente problemas no lineales. Además, cuando la prevalencia de una enfermedad es baja, Schiaffino A y col en un estudio realizado, encontró que en estudios transversales, la aplicación de la regresión logística presenta los valores del OR similares a la razón de prevalencias, en este estudio es lo que hemos aplicado.(50).

En este estudio se plantea el modelo de regresión logística porque estamos dicotomizando la variable dependiente, aunque también se podría plantear la variable dependiente politómica y realizar una regresión logística multivariada, además dejando la variable dependiente continua se puede proceder analizarla como una regresión lineal.

## CONCLUSIONES

1. Las variables clínicas y demográficas que se encontraron asociadas a hipertensión arterial fueron: sexo, edad, hábito de fumar, practica deporte, colesterol total, frecuencia cardiaca, glicemia, talla y peso.

2. El modelo de regresión que explica la asociación entre hipertensión arterial y factores clínicos que se determinó en termino de probabilidades fue el siguiente:

$$P(X_i) = \frac{e^{(-5.7 + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso})}}{(1 + e^{(-5.7 + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso})})}$$

3. El modelo de regresión logística final donde se consideraron los factores sociodemográficos y clínicos asociados a hipertensión arterial en termino de probabilidades fue el siguiente:

$$P(X_i) = \frac{e^{(-5.78 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{habito de fumar} - 0.09 \text{deporte})}}{(1 + e^{(-5.78 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{habito de fumar} - 0.09 \text{deporte})})}$$

4. Los factores sociodemográficos y clínicos asociados a hipertensión arterial mediante un modelo de regresión logística en una población de la costa del Perú se evaluaron mediante el modelo de regresión logística que explica la asociación entre hipertensión arterial y factores clínicos que se determinó en termino de probabilidades fue el siguiente:

$$P(X_i) = \frac{e^{(0.00 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{hábito de fumar} - 0.09\text{practica deporte} + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso})}}{(1 + e^{(0.00 - 0.91\text{sexo} + 0.06\text{edad} - 0.02\text{hábito de fumar} - 0.09\text{practica deporte} + 0.01\text{colesterol total} + 0.04\text{frecuencia cardiaca} + 0.02\text{glicemia} + 3.31\text{talla} + 0.05\text{peso})})}$$

## RECOMENDACIONES

- Retirar los datos palanca o posibles datos atípicos, aun así, habiendo teniendo buenos resultados en la evaluación de supuestos para estudios futuros en busca de la mejora continua del modelo,
- Evaluar valores clínicos asociados a hipertensión según nivel del mar y diferente altura en nuestro país por la diversidad y hábitos de los pobladores peruanos.
- Evaluar modelos de regresión logística diferenciados por grupos etarios.
- Considerar otras variables tales como obesidad, consumo excesivo de alcohol y sal y presencia previa de alguna enfermedad, estilos de vida y comportamientos individuales como consumo de algún medicamento que altere su presión arterial, o tiempo de la enfermedad y depresión.

## REFERENCIAS BIBLIOGRÁFICAS

1. Kearney P. Whelton M. Reynolds K. Whelton P. He J. Worldwide prevalence of hypertension: a systematic review. *J Hypertens.* 2004; 22(1):11–19.
2. Mathers C. Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 2006; 3(11): e442.
3. Seedat Y. Recommendations for hypertension in sub-Saharan Africa. *Cardiovasc J S Afr.* 2004; 15(4):157–8.
4. Instituto Nacional de Estadística e Informática. Perú: Enfermedades transmisibles y no transmisibles. 2020. Lima. Perú: INEI; 2020. Disponible en [https://www.inei.gov.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1796/cap01.pdf](https://www.inei.gov.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1796/cap01.pdf)
5. Ruiz A. Carillo R. Bernabé A. Prevalencia e incidencias de hipertensión arterial en Perú: Revisión sistemática y metaanálisis. *Rev. Peru. Med. Exp. Salud Publica.* 2021; 39(4): 521-29
6. Van Oort S. Beulens J. van Ballegooijen A. Grobbee. D. Larsson. S. Association of Cardiovascular Risk Factors and Lifestyle Behaviors With Hypertension: A Mendelian Randomization Study. *Hypertension.* 2020; 76(6). 1971–9.
7. García-García JJ. Medición del riesgo en epidemiología. Primera parte. *Revista Mexicana de Pediatría* 1998;65(2):76-83
8. Plaut R. Análisis de riesgo. Alcance y limitaciones para el administrador de salud. Panamá 1984. *Boletín de Sanit Panamá* 96(4). 1984 [citado 20 de diciembre del 2022]. Disponible en: [URL chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://iris.paho.org/bitstream/handle/10665.2/17016/v96n4p296.pdf?sequence=1&isAllowed=y](https://iris.paho.org/bitstream/handle/10665.2/17016/v96n4p296.pdf?sequence=1&isAllowed=y)

9. IBM 2021. Disponible en: <https://www.ibm.com/docs/es/spss-statistics/SaaS?topic=regression-logistic>
10. Saeed O et al. Prevalence and associated factors of hypertension among adults in Gadarif in eastern Sudan: a community-based study. BMC Public Health. 2020; 20:291.
11. Nam et al. Hypertension in a mountainous province of Vietnam: prevalence and risk factors. Heliyon. 2020; 6: 1-6.
12. Alvarenga et al. Prevalência estimada e fatores associados à hipertensão arterial em indígenas adultos Krenak do Estado de Minas Gerais. Brasil. Cad. Saúde Pública 2020; 36(1): 1-15.
13. Thato Tshepo. Marlise Van. Winnie Maletladi. Kotsedi Daniel. A Non-Invasive Investigation into the Prevalence of Higher than Normal Blood Pressure. Hypertension and the Association between Blood Pressure and Body Weight in Male and Female Adolescents in the Polokwane Local Municipality. Limpopo-South Africa: A Cross-Sectional Study. Children. 2020; 7.
14. Princewel et al. Prevalence and risk factors associated with hypertension among adults in a rural setting: the case of Ombe. Cameroon. Pan African Medical Journal. 2019;34:147.
15. Pinto et al. Prevalence of arterial hypertension in Brazilian adults and its associated factors and activity limitations: a cross-sectional study. Sao Paulo Med J. 2019; 137(4):312-21
16. Odame et al. Prevalence and lifestyle-related risk factors of obesity and unrecognized hypertension among bus drivers in Ghana. Helion 6. 2019.
17. Jakab et al. Prevalence of hypertension in overweight and obese Hungarian children and Adolescents. EREDETI KÖZLEMÉNY. 2019.
18. Kumar S. Verma A. Prevalence of hypertension among school going adolescent boys in Najafgarh. Delhi. India. International Journal of Adolescent Medicine and Health. 2019.

19. Ghosh S., Kumar M. Prevalence and associated risk factors of hypertension among persons aged 15–49 in India: a cross-sectional study. *BMJ Open* 2019.
20. El Achhab et al. Prevalence, control and risk factors related to hypertension among Moroccan adults: a multicentre study. *EMHJ*. 2019;25(7).
21. Matshidiso Mokgwathi. Julius Chacha Mwita. Prevalence of hypertension and selected cardiovascular risk factors among adolescents in selected rural and urban secondary schools in Botswana. *Cardiovascular journal of africa*. 2019.
22. Law M. Morris J. Wald N. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *BMJ*. 2019; 338: b1669.
23. Basu S. Millett C. Social epidemiology of hypertension in middle income countries: determinants of prevalence, diagnosis, treatment, and control in the WHO SAGE study. *Hypertension*. 2013; 62 (1):18–26.
24. Ibrahim M. Damasceno A. Hypertension in developing countries. *Lancet*. 2012; 380 (9841): 611-9.
25. Islam A. Majumder A. Hypertension in Bangladesh: a review. *Indian Heart J*. 2012; 64(3) :319-23.
26. Wu YF. A review on the advancement of epidemiology on hypertension and risk factors in China 2003; 32:27–29.
27. Wu Y. Huxley R. Li L. Anna V. Xie G. Yao C. et al. Prevalence, awareness, treatment, and control of hypertension in China: data from the China National Nutrition and Health Survey 2002. *Circulation* 2008; 118:2679–86.
28. Li YC. Wang LM. Jiang Y. Li XY. Zhang M. Hu N. Prevalence of hypertension among Chinese adults in 2010. *Zhonghua Yu Fang Yi Xue Za Zhi* 2012; 46:409–413.

29. Xi B. Liang Y. Reilly KH. Wang Q. Hu Y. Tang W. Trends in prevalence, awareness, treatment, and control of hypertension among Chinese adults 1991–2009. *Int J Cardiol* 2012; 158:326–329.
30. Xu B. Xu Z. Xu X. Cai Q. Xu Y. Prevalence, awareness, treatment, and control of hypertension among residents in Guangdong Province, China, 2004 to 2007. *Circ Cardiovasc Qual Outcomes* 2013; 6:217– 222.
31. Zhao Y. Yan H. Marshall RJ. Dang S. Yang R. Li Q. Qin X. Trends in population blood pressure and prevalence, awareness, treatment, and control of hypertension among middle-aged and older adults in a rural area of Northwest China from 1982 to 2010. *PLoS One* 2013; 8:e61779.
32. Mancia G. De Backer G. Dominiczak A. Cifkova R. Fagard R. et al. Guidelines for the management of arterial hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Eur Heart J* 2007; 28: 1462– 36.
33. Chobanian AV. Bakris GL. Black HR. Cushman WC. Green LA. et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* 2003;42: 1206–52.
34. Obineche E. Management of hypertension: update and review. *Bull Kuwait Inst Med Spec* 2003;2:73-82.
35. Rossi RJ. *Applied biostatistics for the health sciences*. Second edition. Hoboken, NJ: Wiley; 2022. 1 p.
36. García Pérez R. García Pino G. González Ballester D. García Moreno R. Modelo de regresión logística para estimar la dependencia según la escala de Lawton y Brody. *Semergen*. 1 de agosto de 2010;36(7):365-71.


37. Chen DG. Chen JK. Statistical Regression Modeling with R: longitudinal and multi-level modeling. Cham: Springer; 2021. 228 p. (Emerging Topics in Statistics and Biostatistics).
38. MacFarland TW. Yates JM. Using R for biostatistics. Cham. Switzerland: Springer; 2021. 902 p.
39. Regresión logística simple y múltiple [Internet]. [citado 6 de febrero de 2023]. Disponible en: [https://www.cienciadedatos.net/documentos/27\\_regresion\\_logistica\\_simple\\_y\\_multiple](https://www.cienciadedatos.net/documentos/27_regresion_logistica_simple_y_multiple)
40. Lever J. Krzywinski M. Altman N. Logistic regression. Nature Methods. 1 de julio de 2016;13(7):541-2.
41. Dominguez-Lara SA. El odds ratio y su interpretación como magnitud del efecto en investigación. EDUMED. 1 de enero de 2018;19(1):65-6.
42. Szumilas M. Explaining Odds Ratios. J Can Acad Child Adolesc Psychiatry. agosto de 2010;19(3):227-9.
43. Martínez Pérez JA. Pérez Martin PS. La curva ROC. Medicina de Familia SEMERGEN. 1 de enero de 2023;49(1):101821.
44. 1Logistic Regression with Stata [Internet]. [citado 6 de febrero de 2023]. Disponible en: <https://stats.oarc.ucla.edu/stata/seminars/stata-logistic/>
45. Instituto Nacional de Cancer. 2022. <https://www.cancer.gov/espanol>
46. National Institute of Mental Health (NIMH). What is prevalence? [Internet]. [citado el 30 de marzo de 2022]. Disponible en: <https://www.nimh.nih.gov/health/statistics/what-is-prevalence>
47. Schober P. Vetter T. Logistic Regression in Medical Research. Anesth Analg. 2021; 132(2):365-366.
48. Kim H. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restor Dent Endod. 2017;42(2):152-155.

49. Ministerio de Salud. Guía Técnica: guía de práctica clínica para la prevención y control de la enfermedad hipertensiva en el primer nivel de atención R.M.N° 491-2009/MINSA. Dirección General de Salud de las Personas del Ministerio de Salud; 2011.
50. Schiaffino A, Rodríguez M, Pasarín MI, Regidor E, Borrell C, Fernández E. Odds ratio o razón de proporciones: su utilización en estudios transversales. Gac Sanit. 2003;17(1):70-4.

## ANEXOS

### ANEXO 1

#### CARTA DE AUTORIZACIÓN DEL ÁREA OPERATIVA DE CLÍNICA INTERNACIONAL

**UNIVERSIDAD PERUANA  
CAYETANO HEREDIA**  
Vice Rectorado de Investigación

**Declaración del Investigador Principal y  
del Jefe de la Unidad Operativa**

**Título del Proyecto:**

PREVALENCIA Y FACTORES CARDIOVASCULARES ASOCIADOS A  
HIPERTENSION EN UNA POBLACION DE LA COSTA DEL PERU

**Declaración del Investigador Principal:**

Yo, como investigador principal, acepto la responsabilidad de conducir de este estudio de acuerdo a los principios, establecidos en los estatutos y en las normas vigentes de la universidad, en las leyes, regulaciones, y códigos nacionales e internacionales aplicables. Para referencia puede consultar el Manual de Procedimientos de la Oficina de Protección de seres Humanos sujetos a Investigación o el Manual de Procedimientos del Comité Institucional de Ética para el uso de Animales, de acuerdo a los requisitos de su proyecto de investigación.

Certifico que todos los investigadores y el personal involucrado en este estudio, se encuentran calificados y posee la experiencia suficiente para desempeñar adecuadamente su labor en el proyecto.

Nombre del Investigador Principal:

Firma Fecha 09/09/2015

VRI-05  
NO-UPCH  
Página 1 de 2

**Declaración del jefe de la Unidad Operativa en la que se llevara a cabo el estudio:**

Certifico que mi unidad operativa ha revisado y aprobado este proyecto según nuestros procedimientos internos, y nos comprometemos a apoyar y supervisar su realización dentro de las normas vigentes, dentro de la ley y de las normas nacionales e internacionales para la realización de proyectos de investigación.

Certifico además, que el investigador principal y sus colaboradores tienen la competencia necesaria para su realización y por lo tanto me responsabilizo por los actos y consecuencias de la ejecución del proyecto en mi unidad operativa.

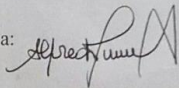
Si su proyecto tiene más de una unidad operativa imprima varias veces esta hoja.

Fecha 09/09/2015

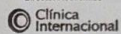
Nombre de la Unidad Operativa: UNIDAD DE INVESTIGACION DE LA CLINICA INTERNACIONAL

Jefe de la Unidad Operativa: DR. ALFREDO GUERREROS BENAVIDES

Firma:



**Dr. Alfredo Guerreros Benavides**  
Director Académico



VRI-05  
NO-UPCH  
Página 2 de 2

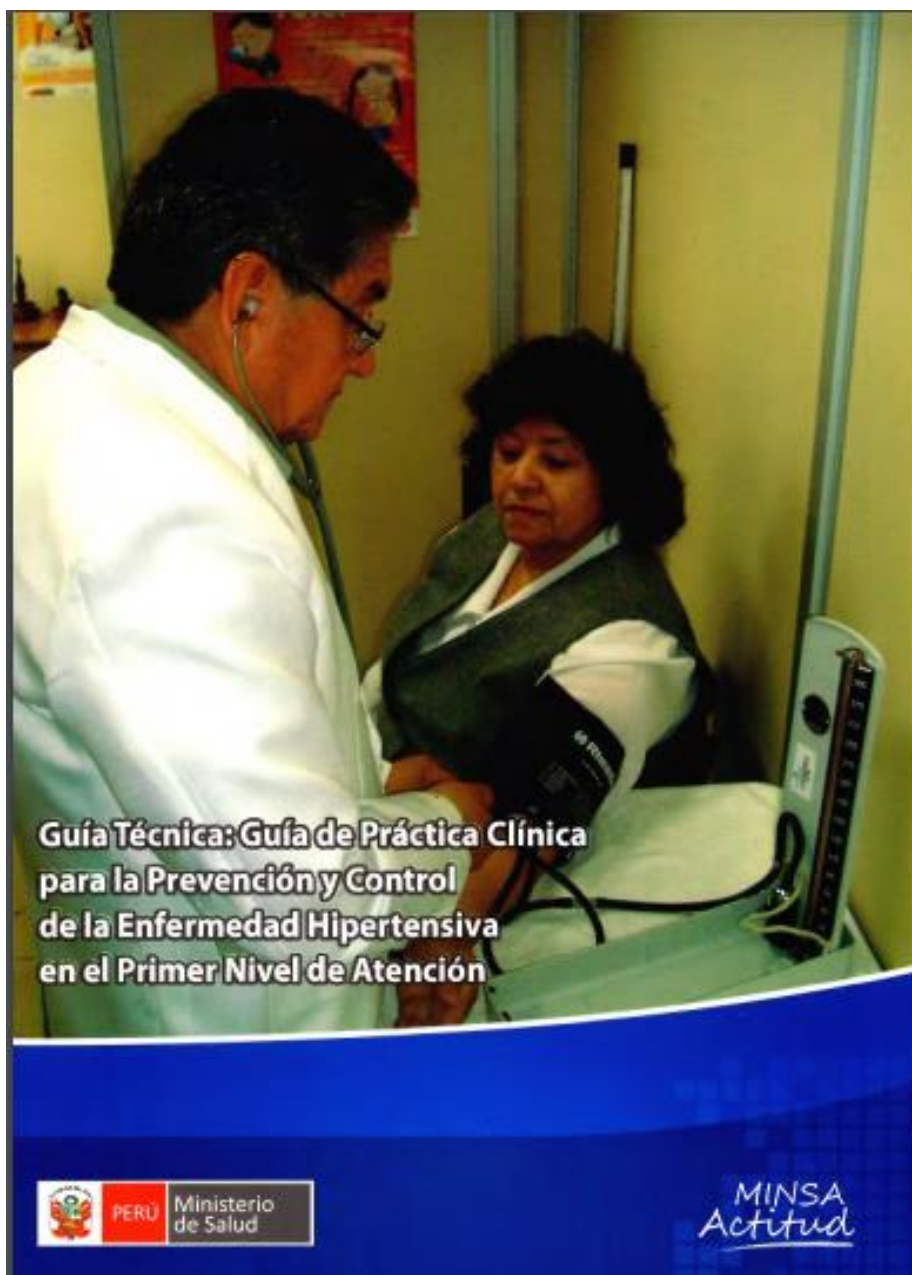
## ANEXO 2

### FICHA DE RECOLECCIÓN DE DATOS

<b>Código del paciente</b>	<b>Dimensiones</b>	<b>Resultado</b>
<b>Sexo</b>	Masculino	M
	Femenino	F
<b>Edad</b>	Categoría	18 - 20
		21 - 25
		26 - 30
		31 - 35
<b>Práctica deportiva</b>	Frecuencia	Si: 1 vez a más No: 0 veces
<b>Hábito de Fumar</b>	Frecuencia	Si: 1 vez a más No: 0 veces
<b>Presión Arterial</b>	Tipo	Presión Sistólica Presión Diastólica
<b>Frecuencia Cardíaca</b>	Medida	lat/min
<b>Talla</b>	Medida	centímetro
<b>Peso</b>	medida	kilogramos
<b>Glicemia</b>	medida	mg/dl
<b>Hipertensión</b>	Presenta	1
	No presenta	0

### ANEXO 3

**Guía Técnica: guía de práctica clínica para la prevención y control de la enfermedad hipertensiva en el primer nivel de atención R.M.N° 491-2009/MINSA**



## ANEXO 4

### Variables para la estratificación de riesgo

<b>Factores de Riesgo Cardiovascular (FRCV)</b>	<ul style="list-style-type: none"> <li>• Varones &gt; 55 años</li> <li>• Mujeres &gt; 65 años o menopausia precoz</li> <li>• Tabaquismo</li> <li>• Dislipidemia               <ul style="list-style-type: none"> <li>○ Colesterol Total &gt; 200 mg/dl o</li> <li>○ Colesterol-LDL &gt; 130 mg/dl o</li> <li>○ Colesterol-HDL en varones &lt;40 mg/dl, mujeres &lt;50 mg/dl o</li> <li>○ Triglicérido &gt;150 mg/dl</li> </ul> </li> <li>• Historia de enfermedad cardiovascular (ECV) prematura en familiar de 1º grado (angina, infarto de miocardio, revascularización miocárdica, muerte súbita):               <ul style="list-style-type: none"> <li>○ En familiares varones &lt; 55 años</li> <li>○ En familiares mujeres &lt; 65 años</li> </ul> </li> <li>• Obesidad abdominal (perímetro abdominal):               <ul style="list-style-type: none"> <li>○ En varones &gt; 102 cm.</li> <li>○ En mujeres &gt; 88 cm.</li> </ul> </li> <li>• Glucosa sérica en ayunas mayor de 100 mg/dl</li> <li>• Síndrome metabólico (SM): 3 ó más de los siguientes criterios:               <ul style="list-style-type: none"> <li>Obesidad abdominal</li> <li>Glucemia mayor de 100 mg/dl</li> <li>PA ≥ 130/85 mmHg</li> <li>Colesterol-HDL en varones &lt;40 mg/dl , Mujeres &lt;50 mg/dl</li> <li>Triglicéridos séricos en ayuna &gt;150 mg/dl</li> </ul> </li> </ul>
<b>Daño de Órgano Blanco (sub-clínico)</b>  Se determina de acuerdo a la capacidad resolutive en el 2do ó 3er nivel de atención	<ul style="list-style-type: none"> <li>• Hipertrofia ventricular izquierda (ECG o ecocardiograma)</li> <li>• Engrosamiento de la pared carotídea (espesor intima-media &gt; 0,9 mm) o placa aterosclerótica</li> <li>• Velocidad onda de pulso carótida-femoral &gt; 12 m/s</li> <li>• Índice tobillo/brazo &lt; 0,9</li> <li>• Incremento ligero de la creatinina sérica (H:1,3-1,5 mg/dl; M:1,2-1,4 mg/dl)</li> <li>• Disminución del filtrado glomerular (&lt; 60 ml/min/1,73 m<sup>2</sup>) o depuración de creatinina en 24 horas (&lt; 60 ml/min)</li> <li>• Microalbuminuria (30-300 mg/24 h; albúmina-creatinina: H: 22, M: 31 mg/g)</li> </ul>
<b>Diabetes</b>	<ul style="list-style-type: none"> <li>• Glucosa plasmática basal &gt;126 mg/dl en medidas repetidas o Glucosa plasmática postsobrecarga oral &gt;200 mg/dl</li> </ul>
<b>Daño de Órgano Blanco (clínico)</b> Enfermedad Cardiovascular o Renal	<ul style="list-style-type: none"> <li>• Enfermedad Cerebrovascular:               <ul style="list-style-type: none"> <li>○ Ictus Isquémico</li> <li>○ Hemorragia cerebral</li> <li>○ Ataque Isquémico Transitorio</li> </ul> </li> <li>• Enfermedad Cardíaca:               <ul style="list-style-type: none"> <li>○ Infarto de Miocardio</li> <li>○ Angina</li> <li>○ Revascularización coronaria</li> <li>○ Insuficiencia cardíaca congestiva</li> </ul> </li> <li>• Enfermedad Renal:               <ul style="list-style-type: none"> <li>○ Nefropatía diabética</li> <li>○ Deterioro renal (creatinina H &gt; 1,5; M &gt; 1,4 mg/dl)</li> <li>○ Proteinuria (&gt; 300 mg/24 h)</li> </ul> </li> <li>• Enfermedad Vasculiar Periférica</li> <li>• Retinopatía avanzada:               <ul style="list-style-type: none"> <li>○ Hemorragias o exudados</li> </ul> </li> <li>• Edema de papila</li> </ul>

## ANEXO 5

### Evaluación del modelo factores sociodemográficos:

Logistic regression Number of obs = 3,535  
LR chi2(4) = 170.38  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.1208  
Log likelihood = -620.23689

hipertension	Odds ratio	Std. err.	z	P>z	[95% conf. interval]
sexo					
1. Femenino	0.4014586	0.0676138	-5.42	0	0.2885899 0.5584706
edad					
	1.067108	0.0063214	10.96	0	1.054789 1.07957
fuma_3					
SI	0.9847666	0.1886916	-0.08	0.936	0.6764476 1.433615
1.deport	0.9113363	0.1488934	-0.57	0.57	0.6616222 1.255299
_cons	0.0030791	0.001145	-15.55	0	0.0014856 0.0063818

Note: \_cons estimates baseline odds.

Logistic regression Number of obs = 3,535  
LR chi2(4) = 170.38  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.1208  
Log likelihood = -620.23689

hipertension	Coefficient	Std. err.	z	P>z	[95% conf. interval]
sexo					
1. Femenino	-0.9126509	0.1684203	-5.42	0	-1.242749 -0.5825533
edad					
	0.0649517	0.0059239	10.96	0	0.0533411 0.0765624
fuma_3					
SI	-0.0153506	0.1916105	-0.08	0.936	-0.3909002 0.3601991
1.deport	-0.0928433	0.1633792	-0.57	0.57	-0.4130606 0.227374
_cons	-5.783112	0.3718492	-15.55	0	-6.511923 -5.054301

### Evaluación del modelo factores clínicos:

Logistic regression Number of obs = 3,535  
LR chi2(5) = 166.91  
Prob > chi2 = 0.0000  
Pseudo R2 = 0.1183  
Log likelihood = -621.97474

hipertension	Odds ratio	Std. err.	z	P>z	[95% conf. interval]
--------------	------------	-----------	---	-----	----------------------

coles_total	1.008123	0.002018	4.04	0	1.004175	1.012086
freccard	1.024203	0.0127406	1.92	0.055	0.999534	1.049481
glicemia	1.008954	0.0027233	3.3	0.001	1.003631	1.014306
talla	0.0263724	0.0294669	-3.25	0.001	0.0029517	0.2356294
peso	1.063907	0.0068342	9.64	0	1.050596	1.077387
_cons	0.003356	0.0064662	-2.96	0.003	0.0000769	0.1465175

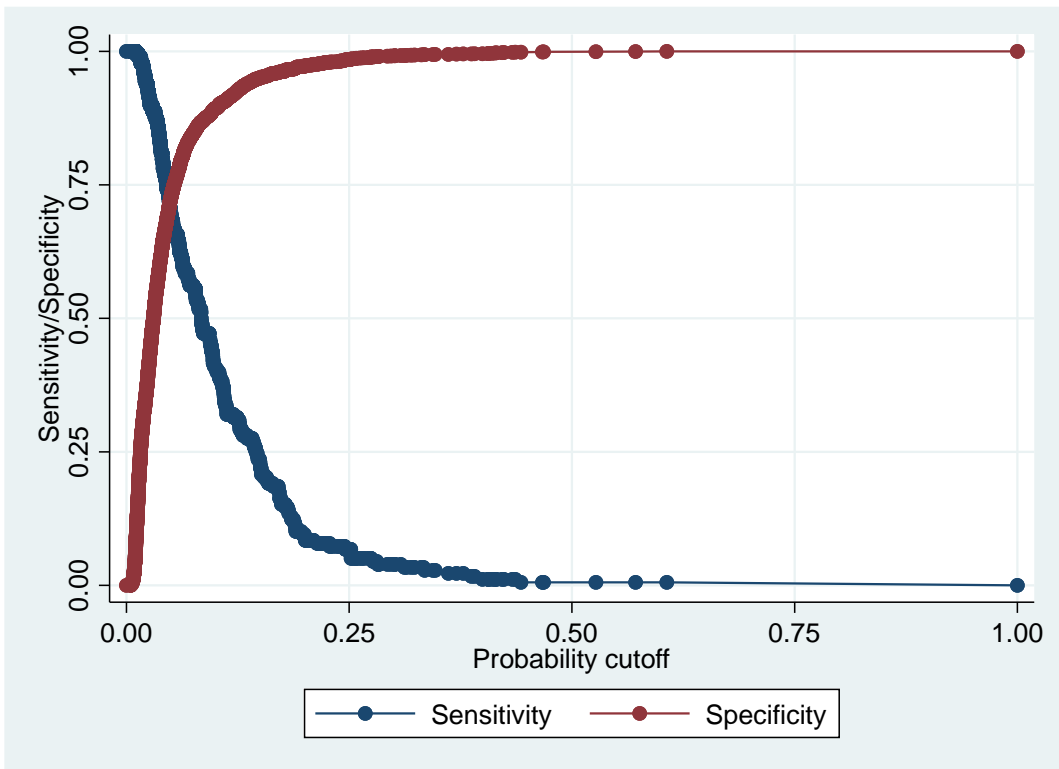
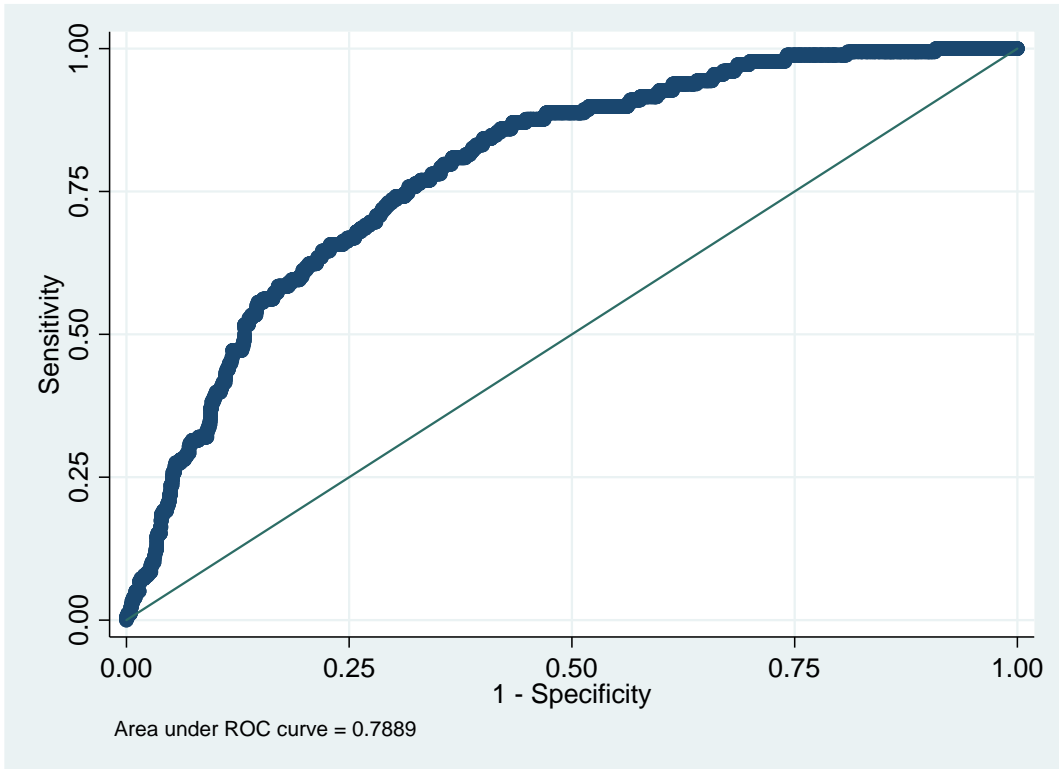
Note: \_cons estimates baseline odds.

Logistic regression  
 Log likelihood = -621.97474  
 Number of obs = 3,535  
 LR chi2(5) = 166.91  
 Prob > chi2 = 0.0000  
 Pseudo R2 = 0.1183

hipertension	Coefficient	Std. err.	z	P>z	[95% conf. interval]
coles_total	0.00809	0.0020018	4.04	0	0.0041667 0.0120134
freccard	0.0239149	0.0124395	1.92	0.055	-0.0004661 0.0482959
glicemia	0.0089143	0.0026991	3.3	0.001	0.0036241 0.0142044
talla	-3.635438	1.117338	-3.25	0.001	-5.82538 -1.445495
peso	0.061948	0.0064237	9.64	0	0.0493578 0.0745383
_cons	-5.697006	1.926768	-2.96	0.003	-9.473403 -1.92061

**Evaluación del modelo final:**

Logistic model for hipertension  
 number of observations = 3535  
 area under ROC curve = 0.7889

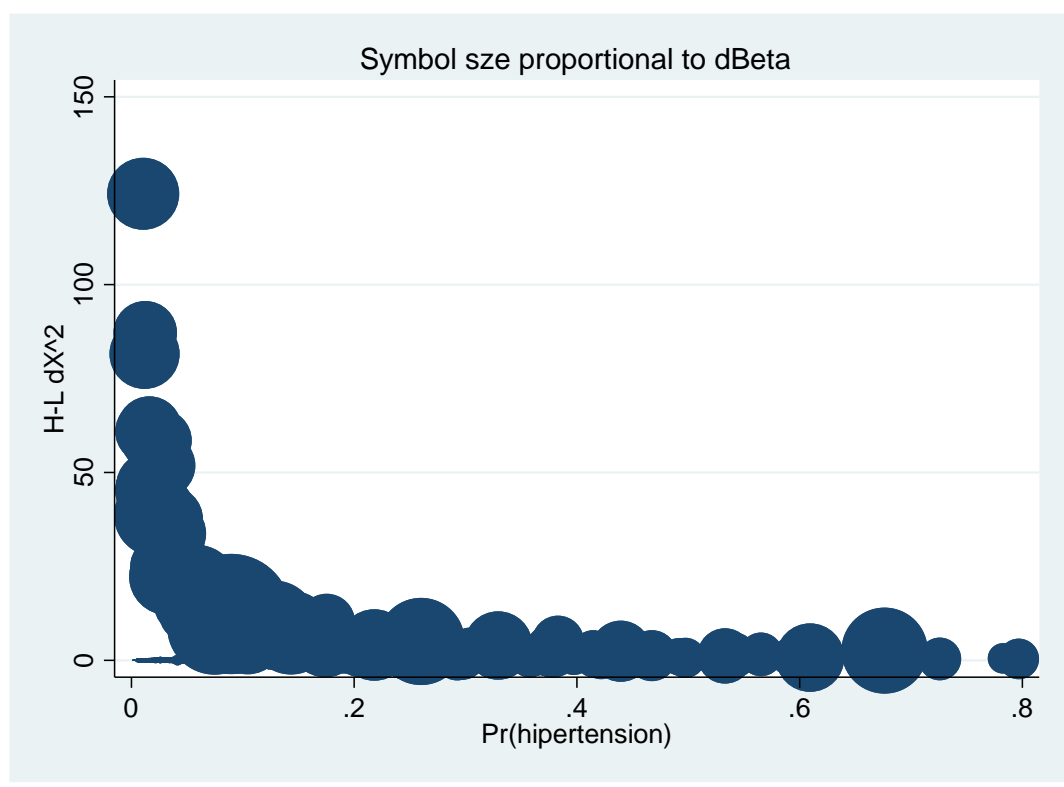


Variable	VIF	1/VIF
sexo	1.07	0.933588
deport	1.06	0.946545
fuma_3	1.04	0.964252

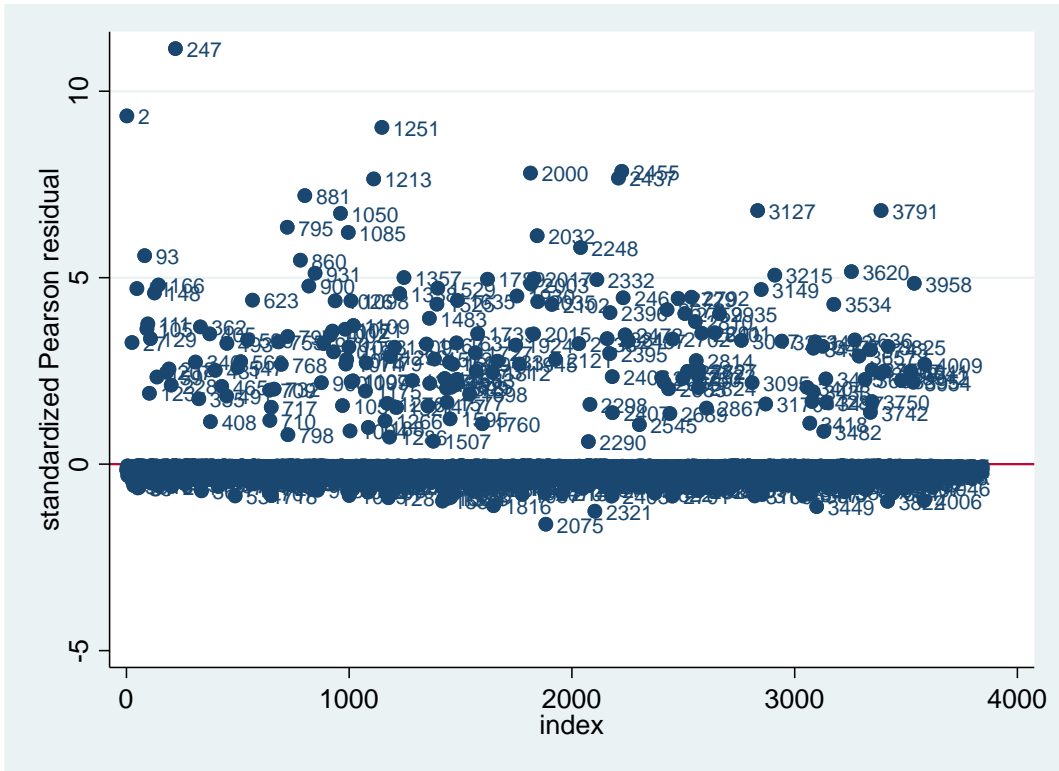


```
3273. |-----+
      | num |
      |-----+
      | . |
      |-----+
```

```
. graph twoway scatter dx2 p [w=db]. t1("Symbol size proportional to dBeta")
(analytic weights assumed)
(analytic weights assumed)
(analytic weights assumed)
```

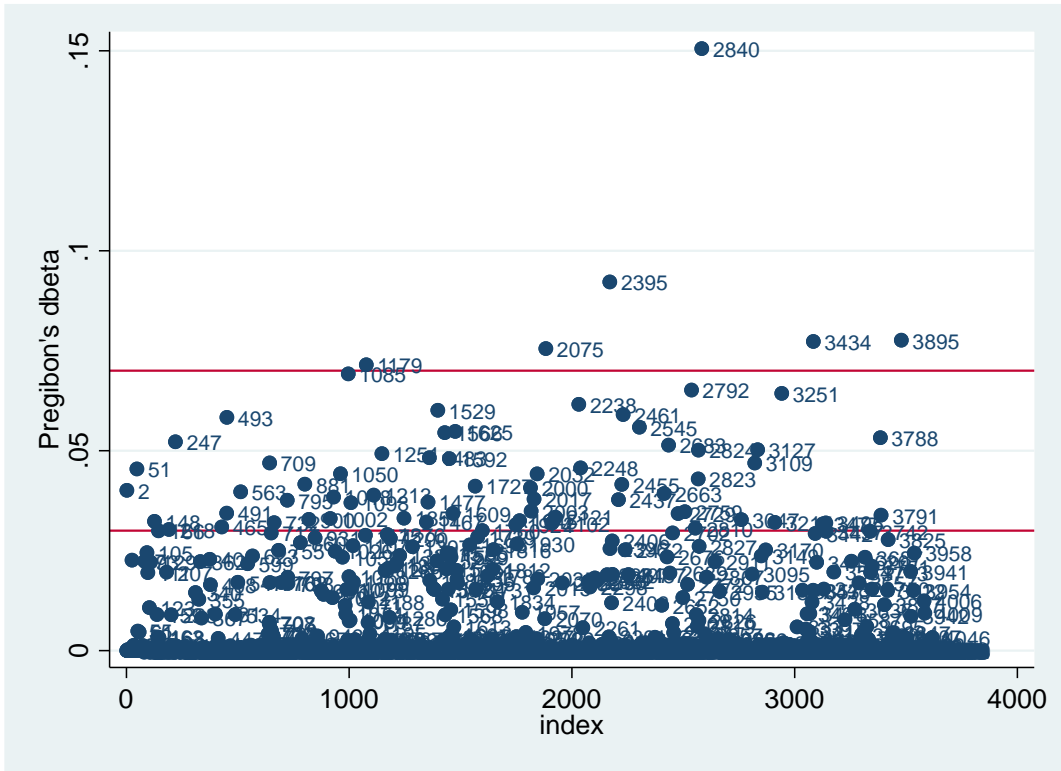


```
variable index already defined
r(110);
```



Variable	Obs	Mean	Std. Dev.	Min	Max
db	3.621	.0017478	.0076842	9.71e-08	.1505594

```
scatter db index. mlabel( num ) yline(0.03) yline(0.07)
```



## ANEXO 6

### Comandos utilizados para el análisis estadístico:

```
tab sexo
tab deport
tab fuma_3
. tabstat edad pa_sis pa_dias freccard talla peso glicemia coles_total, statistics( mean sd median iqr
min max skewness kurtosis ) columns(statistics)
ksmirnov edad = normal(( edad -(46.56303))/(12.02504 ))
ksmirnov pa_sis = normal(( pa_sis -(112.811))/(14.91384))
ksmirnov pa_dias = normal(( pa_dias -(73.44498))/(9.741524))
ksmirnov freccard = normal(( freccard -(71.5389))/(6.033119))
ksmirnov talla = normal(( talla -(1.635477))/(.0882573))
ksmirnov peso = normal(( peso -(68.9058))/(14.18709))
ksmirnov glicemia = normal(( glicemia -(90.17963))/(16.64657))
ksmirnov coles_total = normal(( coles_total -( 189.4371))/(38.49619))
qnorm edad
qnorm pa_sis
qnorm pa_dias
qnorm freccard
qnorm talla
qnorm peso
qnorm glicemia
qnorm coles_total
tabulate sexo hipertension, cell chi2 exact expected
tabulate fuma_3 hipertension, cell chi2 exact expected
tabulate deport hipertension, cell chi2 exact expected
ranksum edad, by(hipertension)
ranksum coles_total , by(hipertension)
ranksum freccard , by(hipertension)
ranksum talla , by(hipertension)
ranksum peso , by(hipertension)
ranksum glicemia , by(hipertension)
logistic hipertension i.sexo edad i.fuma_3 i.deport
logit
logistic hipertension coles_total freccard glicemia talla peso
logit
logistic hipertension i.sexo edad i.fuma_3 i.deport coles_total freccard glicemia talla peso
logit
lfit
lfit, group(10)
lfit, group(10) table
fitstat
lstat
lroc
regress sexo edad fuma_3 deport coles_total freccard glicemia talla peso
vif
logistic hipertension sexo edad fuma_3 deport coles_total freccard glicemia talla peso
predict predict
```

```
sum predict hipertension
predict d, deviance
predict rs, rstandard
predict h, hat
graph twoway scatter h r, xline(0) mlabel(id)
predict dx2, dx2
graph twoway scatter dx2 p, mlabel(id)
predict db, dbeta
gen index = _n
scatter rs index, mlabel(id) yline(0)
scatter db index. mlabel( num ) yline(0.03) yline(0.07)
```