



**Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

Dirección General de Estudios de Posgrado

Facultad de Ciencias Matemáticas

Unidad de Posgrado

**Factores sociodemográficos determinantes de la  
depresión en mujeres de 15 a 49 años – Encuesta  
Demográfica y de Salud Familiar 2015**

**TESIS**

Para optar el Grado Académico de Magíster en Bioestadística

**AUTOR**

Luis BAUTISTA LÓPEZ

**ASESOR**

Mg. Emma Norma CAMBILLO MOYANO

Lima, Perú

2019



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Referencia bibliográfica

---

Bautista, L. (2019). *Factores sociodemográficos determinantes de la depresión en mujeres de 15 a 49 años – Encuesta Demográfica y de Salud Familiar 2015*. Tesis para optar el grado de Magíster en Bioestadística. Unidad de Posgrado, Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

---

## HOJA DE METADATOS COMPLEMENTARIOS

**CODIGO ORCID DEL AUTOR:**

**CODIGO ORCID DEL ASESOR:**

0000-0003-3173-9425

DNI: 44061052

**GRUPO DE INVESTIGACIÓN:**

**INSTITUCIÓN QUE FINANCIA PARCIAL O TOTALMENTE LA INVESTIGACIÓN:**

Autofinanciado

**UBICACIÓN GEOGRÁFICA DONDE SE DESARROLLÓ LA INVESTIGACIÓN. DEBE INCLUIR LOCALIDADES Y COORDENADAS GEOGRÁFICAS**

12°03'30"S 77°05'00"O

**AÑO O RANGO DE AÑOS QUE LA INVESTIGACIÓN ABARCÓ:**

2015

## ACTA DE SUSTENTACIÓN DE TESIS DE GRADO ACADÉMICO DE MAGÍSTER

Siendo las, 14:10 horas del día martes veintinueve de octubre del dos mil diecinueve, en el Aula 105 de la Facultad de Ciencias Matemáticas, el Jurado Evaluador de Tesis Presidido por la Mg. Rosa Ysabel Adriazola Cruz e integrado por los siguientes miembros, Mg. María Estela Ponce Aruneri (Jurado Informante), Mg. Rosa Fátima Medina Merino (Jurado Informante), y la Mg. Emma Norma Cambillo Moyano como Miembro Asesor, se reunieron para la sustentación de la tesis titulada: «FACTORES SOCIODEMOGRÁFICOS DETERMINANTES DE LA DEPRESIÓN EN MUJERES DE 15 A 49 AÑOS – ENCUESTA DEMOGRÁFICA Y DE SALUD FAMILIAR 2015» presentada por el Bachiller Luis Bautista López para optar el Grado Académico de Magíster en Bioestadística.

Luego de la exposición del graduando, los Miembros del Jurado hicieron las preguntas correspondientes, así como las observaciones e inquietudes acerca del trabajo de tesis, a las cuales el Bachiller Luis Bautista López respondió con acierto y solvencia, demostrando pleno conocimiento del tema.


A continuación se realizó la calificación correspondiente, según tabla adjunta, resultando el Bachiller Luis Bautista López aprobado con el calificativo de BUENO  
.....(16).....


Habiendo sido aprobada la sustentación de la Tesis, el Jurado Evaluador recomienda para que el Consejo de Facultad apruebe el otorgamiento del Grado Académico de **Magíster en Bioestadística al Bachiller Luis Bautista López.**

Siendo las 17:12 horas, se levantó la sesión, firmando para constancia la presente Acta.

  
Mg. María Estela Ponce Aruneri  
**Miembro**

  
Rosa Ysabel Adriazola Cruz  
**Presidenta**

  
Mg. Rosa Fátima Medina Merino  
**Miembro**

  
Mg. Emma Norma Cambillo Moyano  
**Miembro Asesor**

***Dedicatoria***

*A mis padres por darme la vida,  
a mis familiares por sus apoyos incondicionales y  
a mis docentes que me guiaron con su enseñanza*

## ***Agradecimiento***

*A Dios por brindarme la dicha de tener vida y ayudarme a superar momentos difíciles en mi vida cotidiana.*

*A mis abuelos, padres y familiares por brindarme apoyo durante mi carrera profesional.*

*A mis docentes de la Maestría y mi asesora Mg. Emma Norma Cambillo Moyano, por compartirme sus conocimientos y apoyarme constante hasta lograr mis objetivos profesionales.*

*A mis amigos y compañeros por brindarme ánimos de superación y de compartir momentos difíciles.*

# ÍNDICE GENERAL

<b>ÍNDICE GENERAL</b>	<b>ii</b>
<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. Situación problemática . . . . .	1
1.2. Formulación de problema . . . . .	3
1.3. Preguntas de investigación . . . . .	4
1.3.1. Pregunta general . . . . .	4
1.3.2. Preguntas específicos . . . . .	4
1.4. Justificación . . . . .	4
1.5. Objetivos . . . . .	5
1.5.1. Objetivo general . . . . .	5
1.5.2. Objetivos específicos . . . . .	5
<b>2. MARCO TEÓRICO</b>	<b>6</b>
2.1. Antecedentes de investigación . . . . .	6
2.2. Bases teóricas . . . . .	8
2.2.1. Reseña histórica de la depresión . . . . .	8
2.2.2. Depresión . . . . .	10
2.2.3. Tipos de depresión . . . . .	11
2.2.4. Causas de la depresión en las mujeres . . . . .	13
2.2.5. Síntomas de la depresión . . . . .	16



<b>3. ANÁLISIS DE REGRESIÓN LOGÍSTICA ORDINAL</b>	<b>17</b>
3.1. Modelo logit acumulativo restringido. . . . .	21
3.1.1. Odds ratios e intervalos de confianzas . . . . .	22
3.1.2. Ajuste del modelo de odds proporcionales . . . . .	23
3.1.3. Supuesto de proporcionalidad de los odds . . . . .	24
3.1.4. Residuos . . . . .	25
<b>4. MÉTODOS</b>	<b>27</b>
4.1. Tipo y diseño de investigación . . . . .	27
4.2. Unidad de análisis . . . . .	27
4.3. Población de estudio . . . . .	27
4.4. Tamaño de muestra . . . . .	28
4.5. Selección de muestra . . . . .	28
4.6. Variables . . . . .	29
4.7. Técnicas de recolección de datos . . . . .	31
4.8. Instrumento de depresión . . . . .	31
4.8.1. Método de puntuación para el diagnóstico de depresión PHQ-9 . . . . .	31
4.9. Análisis e interpretación de la información . . . . .	32
4.9.1. Procesamiento y análisis estadístico . . . . .	32
4.10. Aspectos éticos . . . . .	32
<b>5. RESULTADOS Y DISCUSIONES</b>	<b>33</b>
5.1. Análisis univariado . . . . .	33
5.2. Análisis bivariado . . . . .	34
<b>CONCLUSIONES</b>	<b>55</b>
<b>RECOMENDACIONES</b>	<b>56</b>
<b>REFERENCIAS</b>	<b>58</b>

<b>ANEXOS</b>	<b>62</b>
<b>A. Cuestionario sobre la salud del paciente - 9 (PHQ-9)</b>	<b>63</b>
<b>B. Distribución de frecuencias</b>	<b>64</b>
<b>C. Evaluación del modelo de regresión logística ordinal</b>	<b>66</b>
C.1. Evaluación del modelo con todas las variables en estudio . . . . .	66
C.1.1. Evaluación del modelo con las variables significativas, para el Factor sociodemográfico . . . . .	69
C.1.2. Evaluación del modelo con las variables significativas, para el Factor demográfico . . . . .	72
C.1.3. Evaluación del modelo, para el Factor social . . . . .	73
<b>D. REGRESIÓN LOGÍSTICA</b>	<b>75</b>
D.1. Breve reseña histórica de Regresión logística . . . . .	75
D.2. Regresión logística binaria . . . . .	77
D.2.1. Modelo de regresión logística binaria . . . . .	78
D.2.2. Estimación del modelo . . . . .	80
D.2.3. Evaluación de la bondad de ajuste . . . . .	81
D.2.4. Supuestos del modelo . . . . .	82
D.3. Modelo de regresión logística múltiple . . . . .	83
<b>E. Otros Modelos Ordinales</b>	<b>85</b>
E.1. Modelo de categorías adyacentes . . . . .	85
E.2. Modelo razón de continuidad . . . . .	88
E.3. El modelo estereotipo . . . . .	88
E.4. Modelo de odds proporcionales parciales . . . . .	89
E.4.1. Modelo de odds proporcionales parciales no restringido . .	89
E.4.2. Modelo de odds proporcionales parciales restringido . . . .	91

# LISTA DE TABLAS

2.1. Síntomas de la depresión . . . . .	16
5.1. Nivel de depresión de las mujeres de 15 a 49 años, ENDES - 2015	33
5.2. Características demográficas de mujeres de 15 a 49 años, ENDES - 2015 . . . . .	34
5.3. Depresión según características demográficas . . . . .	35
5.4. Factores sociodemográficos asociados a la depresión. . . . .	36
5.5. Estimación de parámetros para la regresión logística ordinal del factor sociodemográfico. . . . .	44
5.6. Estimación de parámetros para la regresión logística ordinal del factor demográfico. . . . .	47
5.7. Estimación de parámetros para la regresión logística ordinal del factor social. . . . .	51
B.1. Distribución de frecuencia de las variables en estudio . . . . .	64
C.1. Historial de iteraciones . . . . .	66
C.2. Prueba de Omnibus . . . . .	66
C.3. Prueba de efectos del modelo . . . . .	67
C.4. Estimación de parámetros con todas las variables de estudio . . .	68
C.5. Historial de iteraciones del modelo con las variables significativas, para el Factor sociodemográfico . . . . .	69
C.6. Prueba ómnibus del modelo con las variables significativas, para el Factor sociodemográfico . . . . .	69

C.7. Pruebas de efectos del modelo con las variables significativas, para el Factor sociodemográfico . . . . .	70
C.8. Prueba de líneas paralelas del modelo con las variables significativas, para el Factor sociodemográfico . . . . .	71
C.9. Historial de iteraciones del modelo para el Factor Demográfico . .	72
C.10. Prueba ómnibus del modelo para el Factor Demográfico . . . . .	72
C.11. Pruebas de efectos del modelo para el Factor Demográfico . . . .	72
C.12. Prueba de líneas paralelas del modelo para el Factor Demográfico	72
C.13. Historial de iteraciones del modelo para el Factor Social . . . . .	73
C.14. Prueba ómnibus del modelo para el Factor Social . . . . .	73
C.15. Pruebas de efectos del modelo para el Factor Social . . . . .	73
C.16. Prueba de líneas paralelas . . . . .	74

# LISTA DE FIGURAS

- 3.1. Relación entre la variable observada y variable latente  $Y^*$  en el modelo de regresión ordinal con una sola variable independiente . 18

## RESUMEN

La investigación tuvo por objetivo determinar los factores sociodemográficos que inducen a la depresión en la población de mujeres peruanas de 19 a 45 años. Siendo la Encuesta Demográfica de Salud Familiar 2015 la base de datos. La investigación es cuantitativa de nivel correlacional, diseño no experimental de corte transversal. La selección y depuración de las variables fue en base a los objetivos, considerándose 13763 casos. A partir de ello, se determina la prevalencia de las mujeres con depresión 33.2 % (19.4 % leve, 7.1 % moderada, 3.8 % severa y 2.9 % muy severa). Por medio del modelo odds proporcionales se determinó los factores sociodemográficos que inducen a la depresión: nivel de estudios de la pareja hasta nivel secundaria, nivel de riqueza pobre, tener dos o más parejas sexuales diagnóstico de diabetes o hipertensión, haber fumado en los últimos 12 meses, haber bebido alcohol alguna vez y haber sufrido violencia por parte de su pareja siendo ésta la característica de mayor riesgo para sufrir depresión.

**Palabras clave:** ENDES- 2015, Niveles de depresión, PHQ-9, Factores sociodemográficos, Regresión logística ordinal, Modelos de odds proporcionales.

## ABSTRACT

The research aimed to determine the sociodemographic factors that induce depression in the population of Peruvian women aged 19 to 45 years. The Demographic Survey of Family Health 2015 being the database. The research is quantitative correlational level, non-experimental design of cross section. The selection and debugging of the variables based on the objectives, considering 13763 cases. From this, the prevalence of women with depression is determined 33.2% (19.4% mild, 7.1% moderate, 3.8% severe and 2.9% very severe).

By means of the proportional odds model, the sociodemographic factors that induce depression were determined: level of studies of the couple until secondary level, level of poor wealth, having two or more sexual partners diagnosing diabetes or hypertension, having smoked in the last 12 months, having drunk alcohol at some time and having suffered violence from your partner, this being the characteristic of increased risk for depression.

**Keywords:** ENDES- 2015, Depression levels, PHQ-9, Sociodemographic factors, Ordinal logistic regression, Proportional odds models.

# Capítulo 1

## INTRODUCCIÓN

### 1.1. Situación problemática

La depresión es un trastorno mental de alteración patológica del estado de ánimo, vinculada con el descenso del humor llegando a niveles de tristeza (Kaplan y Sadock, 2002), y puede estar acompañada de diversos síntomas y signos tales como: trastorno emocional y del pensamiento alterando el comportamiento y ritmos vitales que persisten por un tiempo habitualmente prolongado en promedio de 2 semanas (Ministerio de Salud de Chile, 2006). Según la OMS en los últimos años la depresión fue creciendo intensamente, ubicándose así en el tercer lugar en todo el mundo, en el octavo lugar en los países que tienen ingresos altos y el primer lugar en los países que tienen ingresos bajos y medios, siendo ésta la principal causa de poseer dicha enfermedad y agrede con mayor frecuencia a las mujeres indistintamente de su edad (World Health Organization [WHO], 2008) .

En el 2012 la Organización Mundial de salud (OMS), calculó que la depresión afecta a 350 millones de personas en el mundo, y entre los años 1990 y 2013, el número de personas con depresión o ansiedad incrementó hasta un 50%, específicamente pasando de 416 a 615 millones, representando entre 10 y 12% de la población mundial para la depresión (Ministerio de Sanidad y Consumo-Madrid,



2007). La prevalencia de los trastornos de la depresión están entre 10 y 25 % para las mujeres y entre 5 y 12 % para los varones en algún momento de sus vidas, observándose que los intentos suicidas son más frecuentes en las mujeres, y los suicidios consumados se presentan más en los varones (Instituto Nacional de Salud Mental “Honorio Delgado-Hideyo Noguchi” [INSM “HD-HN”], 2006) .

En el Perú Según Estudios Epidemiológico en Salud Mental (EESM) del INSM “HD-HN”, en los años 2002-2007, el episodio depresivo es el más frecuente en adolescentes, mostrando prevalencias de 8.6 % en Lima; 5.7 % en la Sierra; 4.7 % en la Selva y 4.4 % en la Costa sin Lima (Instituto Nacional de Salud Mental [INSM], 2012) . Distintas investigaciones revelan prevalencias de depresión en algunos departamentos del Perú tales como: 12.08 % para niños y adolescentes en Huancayo (Salazar, 2004), en Junín escolares adolescentes hasta 36.4 % (Freyre, 2004), y en aquellas afectadas por la violencia política en las décadas de los 80 y 90 la prevalencia de depresión fue 17.5 % en Lima y 24.5 % en Ayacucho (INSM, 2012).

Diversos Estudios muestran que las prevalencias de depresión de las mujeres duplican o son mayores a los hombres (Murray y López, 1996). En Chile 2008, 10 % corresponde a las mujeres y 4 % para hombres (Ministerio de Salud Chile, 2008), en Colombia 12 % frente a 6.2 % (Gómez-Restrepo y otros, 2004). La distribución de la depresión se presenta desigualmente en la población peruana, afectada por algunos factores vinculados a bajos recursos económicos, consumir ketamina o sustancias adictivas, tener dependencia alcohólica, discapacidad física y/o psicológica (Gómez-Restrepo y otros, 2004; Torija y otros, 2007; Montoya y otros, 2007), como también factores sociodemográficos; sexo, escolaridad, ocupación, estado civil y entre otras (INSM, 2002). Por todo lo expuesto, y dada la elevada proporción de depresión en las mujeres frente a los varones en distintos países, se justifica la evaluación de los principales factores asociados a la depresión de

las mujeres, utilizando la información de la Encuesta Demográfica y de Salud Familiar (ENDES) del año 2015.

## **1.2. Formulación de problema**

Usualmente, toda persona experimenta instantes de tristeza, abandono, falta de interés y entre otras actitudes negativas ue pueden mostrar dicha persona, ya sea después de fracasos, daños, decepciones y otros. Si la persona muestra estas actitudes de manera excesiva en intensidad y duración, compromete el desempeño de la persona o le nace un pensamiento de terminar con su vida, en este caso nos encontramos frente a un caso de depresión muy severa y si ésto no es tratada adecuadamente, la persona puede llegar a suicidarse (INSM "HD-HN" , 2012).

El estudio realizado por el Banco Mundial y de la Organización Mundial de la Salud, en el año 2008, indican que las causas de discapacidad que sufren las personas que se encuentran entre las edades de 15 y 44 años, el 50% de ellos está relacionado con los problemas de salud mental (OMS, 2008) . En el Perú la prevalencia de salud mental (ansiedad y depresión) se distribuyen de la siguiente manera: en Lima Metropolitana 18.2%, en la sierra 16.2% y en la Selva 21.4% (Rondón, 2006). La depresión mundial es 12% (Ministerio de Sanidad y Consumo-Madrid, 2007) y 5% en la población adulta de América Latina (Organización de Naciones Unidas [ONU],2016) ; sin embargo seis de cada diez personas no buscan o no reciben tratamiento además, según la Organización Panamericana de la Salud (OPS), la depresión podría llegar a ser considerada en el año 2020 la segunda causa de discapacidad en los países industrializados (Chahua y otros, 2014), por todo lo expuesto se considera la siguiente pregunta.

## **1.3. Preguntas de investigación**

### **1.3.1. Pregunta general**

¿Cuáles son los factores sociodemográficos que inducen a la depresión en la población de mujeres peruanas de 15 a 49 años, usando la Encuesta Demográfica de Salud Familiar (ENDES, 2015)?

### **1.3.2. Preguntas específicos**

- ¿Cuales son los factores sociales que inducen a la depresión en la población de mujeres peruanas de 15 a 49 años, tomando como base la Encuesta Demográfica de Salud Familiar (ENDES, 2015)?
- ¿Cuales son los factores demográficos que inducen a la depresión en la población de mujeres peruanas de 15 a 49 años, tomando como base la Encuesta Demográfica de Salud Familiar (ENDES, 2015)?

## **1.4. Justificación**

La depresión es un problema mundial de salud pública que afecta a las personas, no sólo por su alta morbilidad sino también por las secuelas que pueden ocasionar (OMS, 2008), por ende, es importante aportar evidencia que permita explicar los factores asociados a la depresión. Asimismo, con los resultados obtenidos, se podrán plantear modelos de prevención/intervención frente a la depresión.

Los resultados de esta investigación beneficiarán a todas las mujeres peruanas y sobre todo a las mujeres en edad fértil (según la ENDES de 15 años a 49 años) que presentan depresión. Las instituciones de salud y su personal, la comunidad en general, pues al determinar los factores asociados a la depresión, permitirá un mejor abordaje y seguimiento de los casos. Además, el demostrar los factores

probables que pudieren afectar la salud mental (depresión), se podrá implementar medidas de prevención en el caso de los factores modificables.

La propuesta planteada en el presente estudio, profundizará la determinación de los factores con las limitaciones del caso, porque se utilizará la información nacional de la ENDES, 2015.

## **1.5. Objetivos**

### **1.5.1. Objetivo general**

Determinar los factores sociodemográficos que inducen a la depresión en la población de mujeres peruanas de 19 a 45 años, tomando como base la Encuesta Demográfica de Salud Familiar (ENDES, 2015).

### **1.5.2. Objetivos específicos**

- Determinar los factores sociales que inducen a la depresión en la población de mujeres peruanas de 19 a 45 años, tomando como base la Encuesta Demográfica de Salud Familiar (ENDES, 2015).
- Determinar los factores demográficos que inducen a la depresión en la población de mujeres peruanas de 19 a 45 años, tomando como base la Encuesta Demográfica de Salud Familiar (ENDES, 2015).

## Capítulo 2

# MARCO TEÓRICO

### 2.1. Antecedentes de investigación

**Chahua y otros (2014)**, realizaron un estudio transversal “Depresión en jóvenes usuarios regulares de cocaína reclutados en la comunidad”, cuyo objetivo fue: Identificar la prevalencia de depresión y factores asociados en usuarios regulares de cocaína, en 630 jóvenes. Para medir la depresión usaron los instrumentos Composite International Diagnostic Interview, Duke-Functional Social Support Questionnaire y Severity Dependence Scale. Encontraron que el 33,3% había tenido depresión alguna vez y el 14,6% depresión reciente, usando el OR encontraron asociación con: sexo femenino, no tener hogar, consumir ketamina y menor apoyo confidencial

**Niño-Avenidaño y otros (2014)**, en su estudio transversal “Episodios de ansiedad y depresión en universitarias de Tunja (Colombia): probable asociación con uso de anticoncepción hormonal. 2012”, cuyo objetivo fue identificar la asociación entre el uso de anticonceptivos hormonales y la ocurrencia de depresión y ansiedad. La muestra fue no probabilístico (secuencial) de 538 mujeres. Aplicaron la Escala de Ansiedad de Hamilton y la Escala de Depresión de Beck para la determinación de la depresión. La prevalencia global de depresión fue de 18.95%

utilizaron el Odds Ratio (OR) para determinar la asociación, encontraron así asociación estadísticamente significativa entre el uso de métodos hormonales y la edad (superior a los 20 años) .

**Ramírez-Ruiz y otros (2011)**, en su estudio “Factores asociados a la depresión en las grandes urbes. El caso del distrito Federal en México”, usaron la Encuesta Nacional de salud y Nutrición (ENSANUT) para determinar si el género, la edad, la pobreza y el nivel de escolaridad son los factores que influyen en el trastorno depresivo. Encontraron así que el 31 % tenían depresión baja, 57 % moderada y 12 % alta. Utilizando la regresión lineal múltiple, cuya variable dependiente fue el índice de depresión, mostraron que la escolaridad y el género resultan ser variables significativas .

**Olmedilla y otros (2008)**, en su estudio “Variables sociodemográficas, ejercicio físico, ansiedad y depresión en mujeres: un estudio correlacional”, cuyo objetivo fue estudiar las relaciones que se establecen entre variables sociodemográficas y práctica de ejercicio físico con niveles de ansiedad y depresión en mujeres. El estudio se ha realizado con 200 mujeres que trabajan en la municipalidad de Cartagena (Murcia) cuyas edades varían de 18 a 65 años. La prevalencia de depresión encontrada fue 31 %. En este estudio encontraron que las mujeres con edades de 45 hasta 54 años, presentan altos niveles de ansiedad y depresión frente al resto de las edades; mientras que las mujeres viudas presentan altos niveles de ansiedad frente a las solteras y altos niveles de depresión frente a las casadas como también a las solteras; y por último las mujeres que tienen dos o más hijos tienen altos niveles de ansiedad y depresión, frente a los que no lo tienen .

**Gómez-Restrepo y otros (2004)**, realizaron un estudio descriptivo transversal “Prevalencia de depresión y factores asociados con ella en la población colombiana”, cuyo objetivo fue analizar el síndrome depresivo de acuerdo a la prevalencia

presentada y los factores asociados a ella, utilizando una muestra polietápico seleccionaron a 1116 participantes con edades de 18 o más años. La asociación de las variables seleccionadas con episodios depresivos se evaluó mediante un modelo de regresión logística con múltiples variables. encontraron que el 10.0% de los participantes tuvieron al menos algún nivel depresivo en los últimos 12 meses antes del estudio realizado y el 8.5% el mes último. encontrándose así mayor prevalencia de depresión en mujeres y en las dos etapas mencionadas.

Los factores que se asocian a la depresión en el mes último fueron: el sexo femenino, tener actitudes negativas y buscar adversidades a las personas, padecer de alguna enfermedad o presentar dolores, calificarse el estado de salud de si mismo como regular o malo, el consumo de marihuana o sustancias toxicas, así mismo el consumo de alcohol y sobre todo ser una persona con discapacidad o estar desempleado .

## **2.2. Bases teóricas**

### **2.2.1. Reseña histórica de la depresión**

El término depresión proviene del latín “depressio”; cuyo significado viene hacer “hundimiento” es decir la persona se siente afligido llevando un peso emocional en su vida, la cual le priva vivir con plenitud (Aguirre, 2008).

Hipócrates (siglos V y IV a. C), realizó la primera descripción, quien usó el término Melaina chole (atra bilis, bilis negra) o melancolía proveniente de las alteraciones del equilibrio humoral (Aguirre, 2008). Desde entonces las definiciones a cerca de la depresión fueron evolucionando a medida que transcurría el tiempo. Es por ello que actualmente encontramos distintos enunciados etiológicos en la literatura de cada época y cultura, así mismo diferentes formas de explicación de esta patología.

Areteo de Capadoccia, (siglo II d. C), rechaza la etiología hipocrática y atribuye el cambio de la etiología de la depresión a factores psicológicos, asociándola con la manía y valorizando la importancia de las relaciones interpersonales. Robert Burton (1621), menciona que la depresión también es hereditario, como ejemplo se dice que si un padre es melancólico entonces los hijos también lo serán. En la Edad Media Alejandro Traller (525-605), menciona que la etiología de la depresión no siempre son la tristeza y el temor las dominantes, sino que hay ocasiones que pueden dominar la ira, hilaridad y la ansiedad (Vijay, 2012). Se dice que en el periodo del renacimiento la causa de la depresión fue la posesión diabólica, influencia de los astros o castigo divino y entre otras.

Posteriormente la etiología de la depresión fue psicodinámica, definidos por distintos autores con la finalidad de definir cuadros clínicos. Así con los avances médicos, desde 1960 fue posible extenderse la palabra depresión. Ésta es “provocada por un déficit de los neurotransmisores noradrenalina y serotonina” (Vijay, 2012), tuvo mucha posición en la ciencia médica llegando a impulsar al uso de psicofármacos antidepresivos. Los resultados del uso de estos psicofármacos fueron favorables para las personas en tratamiento ya que aproximadamente el 70 % de los pacientes mejoraron. Posteriormente las investigaciones demostraron que no es suficiente la práctica del uso de este psicofármaco, ya que el 30 % representa un porcentaje alto de los enfermos que no respondían favorable a éste producto pasaron a la búsqueda de nuevos componentes terapéuticos más efectivos y tenga un mínimo efecto secundario. Nemeroff y Plotsky en su publicación a cerca de los trastornos de ansiedad y depresión mencionan a los factores determinantes las cuales son: genéticos, ambientales, consecuencias del sistema nervioso central, resultados adversos que se dieron recientemente o a corto tiempo a algún fármaco. También indican que los antecedentes familiares y los que sufrieron violencia infantil no son determinantes de la depresión. De estos factores se deducen elementos terapéuticos de solución a tomando en cuenta los factores ambientales,



biológicas, psicológicas y de más factores vinculadas al cuadro depresivo (Torres, s.f.).

### **2.2.2. Depresión**

La depresión hasta el momento no se ha podido definir claramente y muchos lo interpretan erróneamente ya que se utiliza con excesiva frecuencia para describir estados anímicos negativos, cambios en la memoria y pensamiento, el funcionamiento físico y en el comportamiento (Ignasio, 2011).

En ocasiones suelen confundir el trastorno depresivo con personas que pasan momentos de tristeza o se encuentran deprimidos o abatidos, pero solo por un tiempo corto, en este sentido la depresión es más que una etapa de desánimo pasajero o un lapso de tiempo de estrés, a menudo el trastorno depresivo tiene como característica: una duración de los estados mencionados de al menos 2 semanas pero usualmente el tiempo de duración es más, además muestra indicios particulares vinculados con el estado anímico, ya sea en el pensamiento, comportamiento además no tiene aptitud para el desarrollo de sus actividades del día a día y entre otras actitudes no necesariamente notables por otros. Si la persona presenta los indicios mencionados entonces necesita la atención profesional ya sea de un psicólogo, médico u otros (Garay, 2005).

La OMS, define a la depresión como un trastorno mental, se distingue por la existencia de tristeza, sentimientos de culpa y pérdida de autoestima, problemas de sueño y falta de apetito, sensación de cansancio y carencia de concentración, pérdida de interés. En múltiples ocasiones esta enfermedad puede hacerse crónica, e impedir el desempeño íntegro ya sea en la escuela o el trabajo y sobre todo dificulta la capacidad de afrontar el día a día. Lo peor es caer en el suicidio. Mientras es leve, se puede tratar sin necesidad de medicamentos, pero si tiene carácter moderado o grave se pueden necesitar medicamentos y psicoterapia pro-

fesional. La depresión también están acompañadas por el cambio de peso, lentitud psicomotora y entre otras (Alberdi, 2006).

### 2.2.3. Tipos de depresión

Los tipos de trastornos depresivos tanto en mujeres y hombres, se dan según su intensidad las cuales pueden ser: trastorno grave, trastorno distímico y la depresión menor.

1. **El trastorno depresivo grave**, este trastorno es caracterizado por la presencia de indicios que interfieren a la capacidad de estudiar, trabajar, comer, dormir y gozar de las ocupaciones que antes eran actividades agradables. El trastorno depresivo grave inhabilita al paciente además le imposibilita realizar sus actividades cotidianas con normalidad. Estudios muestran que el trastorno depresivo grave puede ocurrir únicamente una vez en la vida del individuo, pero en su mayoría, este trastorno puede volver a mostrarse durante su vida.
2. **El trastorno de la distimia o distímico**, la característica de este trastorno es que tiene una duración larga (como mínimo dos años), estos trastornos son menos intensos frente a la depresión grave. Estudios muestran que este trastorno posiblemente no incapacite a la persona, pero si le impide en su desarrollo de actividades con normalidad además de sentirse conforme consigo mismo. Las personas que tienen este trastorno tienen la posibilidad de sufrir uno o más acontecimientos de depresión grave durante su vida.
3. **Depresión menor**, este trastorno se caracteriza por presentar síntomas similares a la depresión grave y distimia, en este caso los síntomas son menos intensos frente a los mencionados y/o generalmente tienen una poca duración (INSM, 2011) .

También existen otras clasificaciones o tipos de depresión, pero sus características

son ligeramente diferentes a las mencionadas anteriormente. Con estas clasificaciones o tipos de depresión que se han determinado, no todos los investigadores concuerdan con dicha caracterización y definición. A continuación desarrollaremos algunos:

1. **Depresión psicótica**, Esta enfermedad sucede cuando el trastorno depresivo viene acompañado con algo de psicosis, tal como, una ruptura con la existencia afectando los órganos de los sentidos como: la visión, el olfato, el oído y entre otras, sentir o palpar cosas que otra gente no tienen la posibilidad de sentir (alucinación); y pensar con firmeza que algunas actitudes falsas son verdaderas, como ejemplo tenemos que la persona piensa que el mejor futbolista del mundo (delirio).
2. **Trastorno afectivo estacional o TAE**, este trastorno se caracteriza por la presencia de una patología depresiva cuando hay poca luz natural o por lo general a lo largo del invierno. Este trastorno depresivo generalmente disminuye o desaparece en los meses de primavera y verano. El trastorno por lo general es tratado eficientemente con terapias de luz, sin embargo, la mitad de los afectados con esta enfermedad no mejoran solamente con este tratamiento de la luz. No obstante, la psicoterapia y los antidepresivos también tienen la posibilidad de disminuir los síntomas de esta enfermedad, solos o juntamente con el tratamiento de luz (Rohana y otros, 2003).

Los tipos de depresión según la OMS.

1. **Trastorno depresivo recurrente**, este trastorno se caracteriza por repetidos episodios de depresión. Durante este episodio la persona presenta un estado de ánimo deprimido, dificultades de concentración, pérdida de interés y de la capacidad de disfrutar, reducción de energía disminuyendo la actividad física durante dos semanas como mínimo. Muchas personas que presentan depresión también padecen de síntomas de ansiedad, muchas veces estos síntomas son difíciles de detectar por los médicos.

2. **Trastorno afectivo bipolar**, Esta enfermedad se caracteriza por presentar sucesos maníacos separados por intermedios de estado de ánimo normal. Estos trastornos siguen un estado de ánimo alto, hiperactivo y en ocasiones irritable, no tienen muchas necesidades de dormir además presentan una autoestima muy alta.

#### **2.2.4. Causas de la depresión en las mujeres**

Estudios sobre la depresión encontraron multitudes factores que ocasionan el incremento del riesgo de padecer esta enfermedad sobretodo en las mujeres. Investigaciones diversas indican que los factores hormonales, químicos, genéticos, ambientales, biológicos, psicológicos, demográficos, sociales y entre otras, son los determinantes de la depresión (INSM, 2011).

##### **1. Las hormonas y las sustancias químicas**

Las hormonas femeninas son las que cambian a la mujer durante su existencia. Investigadores han demostrado que las hormonas intervienen de manera directa con el cerebro, ya que esta es la que controla las el estado de ánimo y las emociones. La mujer pasa muchas etapas peculiares a lo largo de su vida, estos etapas y cambios se dan gracias a las hormonas que hacen un cambio a la mujer ya sea emocionalmente o físicamente. Estas etapas inician con mayor visibilidad en la pubertad; siendo ésta una etapa de cambio brusco ya que inician los ciclos menstruales, pasando luego a un embarazo (posparto), y finalmente la llegan a la menopausia (perimenopausia). Los investigadores también han encontrado que los componentes químicos en el cerebro es un factor determinante de la depresión.

La resonancia magnética (RM) cuya función es mostrar las imágenes del cerebro, han comprobado que el cerebro de la gente que sufre depresión es muy

distinto al cerebro de quienes no la sufren. Las partes del cerebro encargado de regular el estado de ánimo, apetito, sueño, pensamiento, conducta y en entre otras sufren el buen funcionamiento. Además, los principales neurotransmisores (sustancias químicas que las células del cerebro utilizan para comunicarse) se encuentran fuera del equilibrio de funcionamiento. Estos investigadores mencionan lo negativo de éstas imágenes, la causa o motivo del porque se enferman con la depresión.

## 2. **La genética**

Estudios demuestran también que hay alto riesgo de padecer esta enfermedad si existe una historia familiar con este trastorno depresivo, indicando que probablemente halla heredado una tendencia biológica. Cuyo riesgo es alto para los pacientes que sufren trastorno bipolar. Pese a lo mencionado, existe posibilidades de que no todos los pacientes con esta historia familiar llegan a tener cuya enfermedad. Estudios también indican que los que llegan a tener trastorno depresivo grave no necesariamente tienen historias de sus familias con dicha enfermedad. Esto indica que existen factores complementarios que tienen la posibilidad de provocar la depresión, ya sean factores ambientales, bioquímicos y entre otros factores (National institute of mental health, s.f.).

Esta enfermedad causa a que la mujer sufra de depresión grave.

## 3. **El Trastorno disfórico premenstrual (TDPM)**

Muchas mujeres afectadas por el TDPM, A lo largo de la semana antes de la menstruación, por lo general llegan a sufrir el trastorno depresivo, ansiedad, irritabilidad y cambios de humor, de tal forma que obstaculizan en su desarrollo habitual. Se dice que las mujeres que sufren el TDPM no siempre tienen cambios hormonales, son poco comunes, pero si presentan otras actitudes a estos cambios. También, tienen la posibilidad de poseer

antecedentes de otros trastornos relacionados con el estado anímico y presenta diferencia en la química del cerebro, que hacen más sensibles a los cambios hormonales vinculados a la menstruación.

#### **4. Depresión posparto**

Las mujeres fundamentalmente son vulnerables a los trastornos depresivos luego de dar a luz, ya que se vuelven más sensibles soportando los cambios corporales, físicos y hormonales y la reciente obligación de proteger al nuevo ser posiblemente llegue a carga fatigoso. Frente a ello muchas mamás primerizas llegan a sufrir un corto episodio de cambios ligeros en el estado anímico, a esta se le conoce como “tristeza posparto”, en algunos casos las mamás llegan a sufrir depresión posparto. Esta es una enfermedad que muchas veces llega a agravarse, entonces en este caso se necesita un control activo y de acompañamiento emocional para la pronta recuperación. Existen estudios que dieron a conocer que las mujeres en etapa de posparto tienen una posibilidad alta de sufrir numerosos trastornos mentales, inclusive la depresión, a lo largo de varios meses luego de dar a luz.

Muchos estudios indican que las mujeres que sufren depresión posparto, en su mayoría ya han enfermado de trastornos depresivos previos. Muchas de ellas en el transcurso del embarazo, sin embargo, estos trastornos depresivos muchas veces pasa sin que lo sientan.

#### **5. La Menopausia**

Como se ha estado mencionando las mujeres tienen etapas de cambios bruscos, esto debido a cambios de hormonales. Las hormonas aumentan a lo largo de la transición de la premenopausia y menopausia. Existen mujeres que tienen la posibilidad de ingresar a la menopausia sin tener inconveniencias de estado de humor, sin embargo, existen mujeres que tienen alta probabilidad de sufrir esta enfermedad. Muchas veces ocurre hasta en mujeres que

nunca presentaron antecedentes de depresión. Habitualmente el trastorno depresivo es menos en las mujeres que ya pasaron la menopausia.

### 2.2.5. Síntomas de la depresión

Las personas que presentan el trastorno depresivo, aquejan a menudo su estado anímico tanto como; afectivo, somático y cognitivo. (Garibay, 2009) A continuación se muestra un cuadro de los síntomas de depresión.

TABLA 2.1: Síntomas de la depresión

<b>Somático</b>	<b>Afectivo</b>	<b>Cognitivo</b>
- Dificultad para conciliar el sueño	- Ansiedad	- Sentimiento culpa excesiva o inutilidad.
- Cansancio	- Melancolía y tristeza	- Pérdida de memoria
- Cefalea	- Desesperanza	- Preocupación constante
- Enlentecimiento motora	- Irritabilidad	- Visión negativa
- Pérdida o aumento de apetito	- Cambios de humor	- Dificultad para concentrarse
	- Aislamiento social, soledad	
	- Pérdida de placer en todas las actividades	

## Capítulo 3

# ANÁLISIS DE REGRESIÓN LOGÍSTICA ORDINAL

El modelo fue mencionado por primera vez por McKelvey y Zavoina en el año 1975 y fue aplicado en la bioestadística, conocido como modelo de *odds proporcionales* (*Proportional Odds Model*), "modelo de regresión paralela" (*Parallel Regression Model*) o "modelo continuo agrupado" (*Grouped Continuous Model*) (Mc Cullagh, 1980).

Este modelo trata de un método de regresión, cuando la variable respuesta tiene tres o más categorías de tipo ordinal. A modo de una ilustración del modelo de regresión logística ordinal, consideremos una variable respuesta con cinco categorías a estas categorías dividiremos en sólo dos categorías teniendo en cuenta el orden natural. Se muestra a continuación una imagen.

cat 0	<	cat1	cat2	cat 3	cat4
cat 0	cat 1	<	cat 2	cat 3	cat 4
cat 0	cat 1	cat 2	<	cat 3	cat 4
cat 0	cat 1	cat 2	cat 3	<	cat 4

los modelos de regresión logística ordinal puede expresarse como un modelo de variable latente,  $Y^*$  cuyos valores pertenecen a los números reales  $(-\infty, \infty)$ , la



variable observada  $Y$ , toma  $k$  categorías(3.1)

$$Y_i = m \quad \text{si} \quad \tau_{m-1} \leq Y^* < \tau_m; \quad m = 1, 2, \dots, k; \quad i = 1, 2, \dots, n \quad (3.1)$$

Sea  $\tau_0, \tau_1, \dots, \tau_k$ , umbrales o puntos de corte.

A manera de ilustrar con  $k = 5$ , la relación (3.1) será equivalente a:

$$f(x) = \begin{cases} 1 & \text{si} \quad \tau_0 = -\infty \leq Y_i^* < \tau_1 \\ 2 & \text{si} \quad \tau_1 \leq Y_i^* < \tau_2 \\ 3 & \text{si} \quad \tau_2 \leq Y_i^* < \tau_3 \\ 4 & \text{si} \quad \tau_3 \leq Y_i^* < \tau_4 \\ 5 & \text{si} \quad \tau_4 \leq Y_i^* < \tau_5 = \infty \end{cases}$$

Así, cuando la variable latente  $Y^*$  cruza un punto de corte, la categoría observada cambia (Borooah, 2002) .

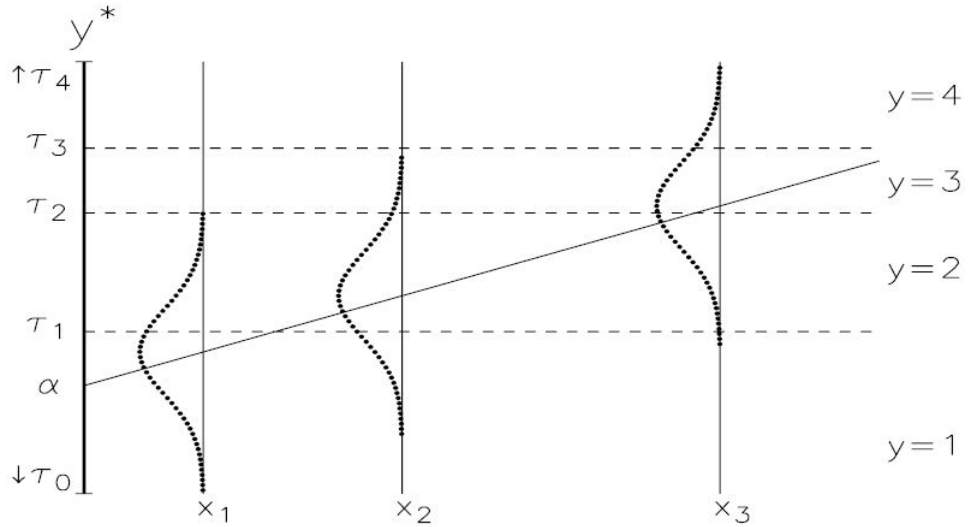


FIGURA 3.1: Relación entre la variable observada y variable latente  $Y^*$  en el modelo de regresión ordinal con una sola variable independiente

Se asigna valores discretos en  $Y$ , mientras que los valores de  $Y^*$  se encuentran en intervalos. Cuyo modelo estructural para la variable latente se muestra a conti-

nuación.

$$Y^* = x_i\beta + \epsilon_i \quad (3.2)$$

Donde:  $x_i$  es el vector de observaciones de las variables independientes, cuyo primer valor es 1 para el intercepto,  $\beta$  viene hacer el vector de coeficientes estructurales que corresponde a las variables independientes y  $\epsilon_i$  viene hacer el error. Para este modelo asumimos un distribución logística de media cero y varianza  $\pi^2/3$ , cuya función de distribución de probabilidad acumulada es:

$$F(\epsilon) = \frac{\exp(\epsilon)}{1 + \exp(\epsilon)} \quad (3.3)$$

Sea  $Y_i = m$  la probabilidad de una observación, entonces de (3.1) se tiene la siguiente ecuación

$$P(Y_i = m/X_i = x_i) = P(\tau_{m-1} \leq Y_i^* < \tau_m/X_i = x_i) \quad (3.4)$$

Del mismo modo la ecuación estructural en (3.2) equivale a

$$P(Y_i = m/X_i = x_i) = P(\tau_{m-1} \leq x_i\beta + \epsilon \leq \tau_m)$$

Restando  $x_i\beta$  en la desigualdad, se obtiene

$$P(Y_i = m/X_i = x_i) = P(\tau_{m-1} - x_i\beta \leq \epsilon \leq \tau_m + x_i\beta)$$

La probabilidad de que una variable aleatoria esté en un intervalo, se puede calcular como la diferencia de las probabilidades acumuladas de los límites del intervalo; como se muestra

$$\begin{aligned} P(Y_i = m/X_i = x_i) &= P(\epsilon \leq \tau_m + x_i\beta) - P(\epsilon \leq \tau_{m-1} - x_i\beta) \\ &= F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta); \quad m = 1, \dots, k \end{aligned} \quad (3.5)$$

En consecuente, la probabilidad de que la variable observada, tome valores de  $m$

o inferiores, es

$$P(Y_i \leq m/X_i = x_i) = \sum_{j=1}^m P(Y_i = j/X_i = x_i); \quad m = 1, \dots, k-1 \quad (3.6)$$

La manera extendida es:

$$\begin{aligned} P(Y_i \leq m/X_i = x_i) = & P(Y_i = 1/X_i = x_i) + P(Y_i = 2/X_i = x_i) + \dots \\ & + P(Y_i = m-1/X_i = x_i) + P(Y_i = m/X_i = x_i) \end{aligned}$$

con respecto al enunciado de probabilidad para la variable observada  $Y_i$  mostrada en (3.5), la probabilidad acumulada definida anterior equivale a:

$$\begin{aligned} P(Y_i \leq m/X_i = x_i) = & [F(\tau_i - x_i\beta) - F(\tau_0 - x_i\beta)] + [F(\tau_2 - x_i\beta) - F(\tau_1 - x_i\beta)] + \dots + \\ & [F(\tau_{m-1} - x_i\beta) - F(\tau_{m-2} - x_i\beta)] + [F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta)] \end{aligned}$$

Realizando algunas simplificaciones, se tiene

$$P(Y_i \leq m/X_i = x_i) = -F(\tau_0 - x_i\beta) + F(\tau_m - x_i\beta)$$

En la relación a (3.1), la primera categoría ( $m = 1$ ), muestra un intervalo abierto a la izquierda, en otras palabras  $\tau_0 = -\infty$ , de la misma forma, la última categoría ( $m = k$ ), muestra un intervalo abierto a la derecha, osea  $\tau_k = \infty$ .

De esta forma,  $F(\tau_0 - x_i\beta) = F(-\infty - x_i\beta) = 0$ , por la propiedad de distribución de probabilidad acumulada (Wackerly y otros, 2010)

De esta manera, la probabilidad acumulada de (3.6) es equivalente a

$$P(Y_i \leq m/X_i = x_i) = F(\tau_m - x_i\beta); \quad m = 1, \dots, k-1 \quad (3.7)$$

La chance (“odds”) a que la variable observada sea igual o menor a  $m$ , esta dado por

$$odds = \frac{P(Y_i \leq m/X_i = x_i)}{1 - P(Y_i \leq m/X_i = x_i)} = \frac{F(\tau_m - x_i\beta)}{1 - F(\tau_m - x_i\beta)} \quad (3.8)$$

De la distribución de probabilidad acumulada dado en (3.3), se tiene que (3.8) es equivalente a

$$odds = \frac{P(Y_i \leq m/X_i = x_i)}{1 - P(Y_i \leq m/X_i = x_i)} = \frac{\frac{\exp(\tau_m - x_i\beta)}{1 + \exp(\tau_m - x_i\beta)}}{1 - \frac{\exp(\tau_m - x_i\beta)}{1 + \exp(\tau_m - x_i\beta)}}$$

Desarrollando lo anterior tenemos

$$odds = \frac{P(Y_i \leq m/X_i = x_i)}{1 - P(Y_i \leq m/X_i = x_i)} = \exp(\tau_m - x_i\beta) \quad (3.9)$$

Insertando el logaritmo en la expresión anterior, definimos el modelo logístico ordinal,

$$\ln(odds) = \ln\left(\frac{P(Y_i \leq m/X_i = x_i)}{1 - P(Y_i \leq m/X_i = x_i)}\right) = \tau_m - x_i\beta; \quad m = 1, \dots, k - 1 \quad (3.10)$$

Del mismo modo, de (3.9), expresamos el modelo logístico ordinal para odds mayores de  $m$

$$\ln\left(\frac{P(Y_i > m/X_i = x_i)}{P(Y_i \leq m/X_i = x_i)}\right) = -\tau_m + x_i\beta; \quad m = 1, \dots, k - 1 \quad (3.11)$$

### 3.1. Modelo logit acumulativo restringido.

Este modelo también es llamado modelo de probabilidades proporcionales o modelos de odds proporcionales. Este modelo compara la probabilidad de la variable respuesta que es igual o menor a la categoría  $m$ , es decir  $P(Y \leq m)$ , con la probabilidad que es mayor que la categoría  $m$ , en otras palabras  $P(Y > m)$ , teniendo en cuenta a los valores de las variables explicativas  $(X_0, X_1, \dots, X_p)$  con  $X_0 = 1$ ;

$$c_m(X) = \ln\left[\frac{P(Y \leq m/X = x)}{P(Y > m/X = x)}\right] = \tau_m - x^T\beta; \quad m = 0, 1, \dots, k - 1. \quad (3.12)$$

Los parámetros  $\tau_m$ , son denominados umbrales,  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$  son llamados

vector de parámetros relacionados con las variables independientes y  $X$  matriz con  $n$  observaciones de las  $p$  variables independientes (Scott Long, 1997).

### 3.1.1. Odds ratios e intervalos de confianzas

Sea  $X$ , una variable independiente, única y a la vez dicotómica ( $X = 0$  ó  $X = 1$ ), entonces por las definiciones anteriores sabemos que el *odds* hace la comparación de  $Y \geq m$  con  $Y < m$  ( $\exp(\alpha_m + \beta_i X_i)$ ). Por lo tanto para hacer la evaluación del efecto de la variable independiente sobre la variable dependiente el *odds ratio* de  $Y \geq m$  para comparar  $X_1 = 0$  y  $X_1 = 1$  es:

$$OR(Y \geq y_i | X_1) = \frac{\text{odds}(Y \geq y_i | X_1 = 1)}{\text{odds}(Y \geq y_i | X_1 = 0)} = \frac{\exp(\alpha + \beta_1)}{\exp(\alpha)} = e^{\beta_1}$$

Por otro lado indicamos que el *odds ratio* en cualquier punto de corte de  $y_i$  será constante.

Así mismo  $\beta_1$  es:

$$\beta_1 = \log OR(Y \geq | X_1)$$

para todo  $y_i$

En el caso de que,  $X_1$  y  $X_1^*$  son variables numéricas (cuantitativas), la comparación entre estos valores esta dada por:

$$OR(Y \geq y_i | X_1, X_1^*) = \exp \beta_1 (X_1^* - X_1)$$

Entonces, la ecuación general para un intervalo de confianza al 95 % de confianza para dos niveles de cualquier variable independiente  $X_1$  y  $X_1^*$  es:

$$IC95\% = \exp \left[ \hat{\beta}_1 (X_1^* - X_1) \pm 1.96 (X_1^* - X_1) s_{\hat{\beta}_1} \right]$$

Donde:  $\hat{\beta}_1$  viene hacer el estimador de máxima verosimilitud del modelo y  $s_{\hat{\beta}_1}$  el error de estimación (Agresti, 2002).

### 3.1.2. Ajuste del modelo de odds proporcionales

El método que se utiliza para realizar el ajustar del modelo de odds proporcionales se basa en la adaptación de la verosimilitud multinomial y su logaritmo (Hosmer y Lemeshow, 2000), para  $k = 2$  se sigue los siguientes pasos básicos.

1. Se Utiliza las ecuaciones que definen los logits particulares del modelo para la creación de una ecuación determinando  $\phi_k(X)$  como función de los parámetros desconocidos.
2. Creación de los valores de una variable respuesta multinomial cuya dimensión es  $k + 1$ ,  $z' = (z_0, z_1, \dots, z_k)$  mientras que para la variable respuesta ordinal sea  $z_j = 1$  si  $y = j$  y  $z_j = 0$  si no cumple la condición anterior, de aquí tenemos que, únicamente uno de los valores de  $z$  será igual a 1 para cada elemento. Entonces la expresión general de la función de verosimilitud para una muestra de  $n$  observaciones independientes,  $(y_i, x_i) : i = 1, 2, \dots, n$  es dado por

$$l(\beta) = \prod_{i=1}^n [\phi_0(x_i)^{z_{0i}} \phi_1(x_i)^{z_{1i}} \times \dots \times \phi_k(x_i)^{z_{ki}}]$$

Donde,  $\beta$  es el vector de coeficientes de regresión y los interceptos (umbrales) de los  $k$  modelos particulares.

Por lo tanto función de verosimilitud del logaritmo se presenta a continuación :

$$L(\beta) = \sum_{i=1}^n \{z_{0i} \ln[\phi_0(x_i)] + z_{1i} \ln[\phi_1(x_i)] + \dots + z_{ki} \ln[\phi_k(x_i)]\} \quad (3.13)$$

Para la estimación de los parámetros se utiliza el método de los mínimos cuadrados de manera reiterada, indicada en la sección D.2

Mientras que la matriz de covarianzas de los parámetros estimados, se determinará a partir de la matriz de información de Fisher.

En el caso de la determinación del efecto de cambio de  $X$ , consideremos  $X = x_i$

y  $X = x_l$  dos valores de  $X$ , . El “odds ratio” para  $X = x_i$  y  $X = x_l$  es igual a:

$$OR = \frac{\exp(\tau_m - x_i\beta)}{\exp(\tau_m - x_l\beta)} = \exp([x_l - x_i]\beta) \quad (3.14)$$

Entonces la ecuación (3.14) es necesario si una variable realiza un cambio entre dos valores posibles. A modo de ejemplo tenemos, si  $X_s$  cambia de  $X_s = x_s$  a  $X_s = x_s + \delta$ , entonces la ecuación es la siguiente:

$$OR = \frac{\exp(\tau_m - [x_s + \delta]\beta_s)}{\exp(\tau_m - x_s\beta_s)} = \exp(-\delta\beta_s) \quad (3.15)$$

La interpretación del OR es de la siguiente manera: Para un incremento  $X_s$ , en  $\delta$ , el odds que la respuesta sea menor o igual a  $m$  es cambiado por el factor  $\exp(-\delta\beta_k)$ , mientras que las otras variables permanecen constante (Scott Long, 1997).

### 3.1.3. Supuesto de proporcionalidad de los odds

La ecuación anterior (3.15), implica que el OR de  $Y \leq m$  y  $Y > m$  para un cambio de  $X_s = x_s$  a  $X_s = x_s + 1$ , es el mismo para cualquier valor  $m$  de  $Y$ . Tal como se muestra a continuación

$$OR = \frac{\exp(\tau_m - [x_s + 1]\beta_s)}{\exp(\tau_m - x_s\beta_s)} = \exp(-\beta_s)$$

La ecuación anterior es conocido como la suposición de regresión paralela o proporcionalidad de los odds (Scott Long, 1997). De acuerdo a esta suposición, decimos que cada variable independiente insertada en el modelo, presentarían un único odds ratio en toda la regresiones logísticas binarias.

La aplicación del modelo de odds proporcionales es válido y aplicado precisamente bajo este cumplimiento del supuesto de proporcionalidad de los odds.

Para la comprobación del supuesto de regresión paralela, se puede utilizar el test Score o el test de Wald (Scott Long, 1997). Así mismo, la opción “brant” o

“autofit” en STATA, que permite probar el cumplimiento del supuesto de proporcionalidad de odds para la variables explicativas.

Lo interesante de este modelo esta en la facilidad de interpretación al suponer constante el efecto de las variables sea cual sea el punto de corte elegido en la variable dependiente, es necesario estimar un coeficiente por cada variable dependiente para conocer su efecto. Por lo que, este modelo podría ser suficiente para explicar la relación de la variable respuesta ordinal y el conjunto de variables explicativas. Otra propiedad interesante también es que, cuando la codificación de la variable respuesta  $Y$  son invertidas (es decir  $Y = 1$  como  $Y = k$ ,  $Y = 2$  como  $Y = k - 1$  y así sucesivamente) los valores de los coeficientes de cada variable explicativa se mantienen, sólo los signos son invertidos. Otra característica de este modelo, es que, cuando algunas categorías de la variable respuesta son excluidas o reagrupadas, los coeficientes de las variables explicativas permanecen inalterados, aunque los umbrales sean afectados (Quispe, 2016) .

### 3.1.4. Residuos

Para el modelo de odds proporcional se puede realizar el análisis de la contribución individual en la primera derivada de la función de log-probabilidad con relación a  $\beta_m$ , promediándolos por separado de acuerdo a los niveles de  $Y$ , la variable respuesta, y analizando el comportamiento de las gráficas de los residuos, pero el método de las gráficas es complejo ya que es complicado la interpretación de las gráficas de los residuos.

Los residuos parciales para el  $i$ -ésimo sujeto y la  $m$ -ésima variable explicativa es definida de la siguiente manera:

$$r_{im} = \hat{\beta}_m X_{im} + \frac{Y_i - \hat{P}_i}{\hat{P}_i(1 - \hat{P}_i)}$$

, donde

$$\hat{P}_i = \frac{1}{1 + \exp[-(\alpha + X_i \hat{\beta})]}$$



Una gráfica más suave, (es decir más regular en el sentido de diferencialidad) de  $X_{im}$  y  $r_{im}$  proporciona una estimación no paramétrica de cuanto influye  $X_m$  al log-odds relativo, que asume  $Y = 1|X_m$ . Para una variable dependiente ordinal  $Y$ , se necesita solo repetir en cada corte de nivel  $k$ ,

$$r_{im} = \hat{\beta}_m X_{im} + \frac{[Y_i \geq k] - \hat{P}_{ik}}{\hat{P}_{ik}(1 - \hat{P}_{ik})}$$

Luego se debe realizar una gráfica para cada  $m$  mostrando así una curva suave en cada  $k$  y luego de debe buscar pendientes o formas similares para cada  $k$  con una variable predictora

fija, cada curva da una estimación de cuanto influye  $X_m$  al log-odds relativo tal que  $Y \geq k$ . Dado que los residuos parciales permiten el estudio de las transformaciones en las variables predictoras (linealidad), al mismo tiempo que permiten el estudio sobre el modelo de odds proporcional (paralelismo), generalmente se prefieren las gráficas de residuos parciales en lugar de las gráficas de residuos, para modelos ordinales.

# Capítulo 4

## MÉTODOS

### 4.1. Tipo y diseño de investigación

La presente investigación es cuantitativa, de nivel correlacional, diseño no experimental de corte transversal. Para este estudio se utilizó la Encuesta Demográfica y de Salud Familiar (ENDES) del año 2015, llevada a cabo por el Instituto Nacional de Estadística e Informática del Perú, esta constituye una data secundaria.

### 4.2. Unidad de análisis

La unidad de análisis es una mujer que se encuentra en las edades de 15 a 49 años habitualmente residente en la vivienda particular ya sean rurales o urbanas del país.

### 4.3. Población de estudio

Todas las mujeres peruanas de 15 a 49 años habitualmente residente en la vivienda particular ya sean rurales o urbanas del país.

## 4.4. Tamaño de muestra

El tamaño de la muestra en el año 2015 fue 35 900 viviendas, correspondientes a 14 140 viviendas al área sede (Capitales de departamento y distritos de Lima Metropolitana), 9 310 viviendas al resto Urbano y 12 450 viviendas al área rural.

## 4.5. Selección de muestra

El diseño muestral que usó la ENDES fue probabilístico, de áreas, estratificada y bietápico. La selección de unidades de muestreo en cada etapa:

1. **Pimera etapa:** selección de Unidad Primaria de Muestreo (UPM) (conglomerados).
  - a) **Área urbana:** representada por áreas geográficas que está conformada mínimamente por una manzana que en total tienen aproximadamente 140 viviendas específicas.
  - b) **Área rural:** están dadas por dos tipos:
    - i El conglomerado está constituido mínimamente por una manzana que en promedio tienen 120 viviendas específicas.
    - ii El sector de empadronamiento rural está conformado mínimamente por un centro poblado rural, en conjunto éstas tienen 120 viviendas específicas (En muchas ocasiones las viviendas están dispersas).
2. **segunda etapa:** En esta etapa se selecciona la viviendas. La integra a la unidad primaria de muestreo seleccionada, ya sea en el área urbano o rural (Encuesta demográfica y de salud familiar 2012, 2014).

## 4.6. Variables

Las variables demográficas y sociales de acuerdo a nuestros objetivos y definidas en el cuestionario de la ENDES, se muestra a continuación:

### \* Variable dependiente

Variable	Código	categoria
Depresión	0	Si
	1	No
Nivel de depresión	0	Sin depresión
	1	Depresión leve
	2	Depresión moderada
	3	Depresión severa
	4	depresión muy severa

### \* Variable independiente

#### Factor Demográfico

Variable	Código	categoria
Edad	0	15 - 19
	1	20 - 29
	2	30 - 39
	3	40 - 49
Estado civil	0	soltera
	1	casada
	2	conviviente
	3	viuda/divorciada
Región natural	0	Sierra
	1	Selva
	2	Resto Costa
	3	Lima metropolitana
Niños menores de 5 años	0	0
	1	1 - 2
	2	3 +
Lugar de residencia	0	Campo
	1	Pueblo
	2	Ciudad pequeña
	3	Capital, ciudad grande

## Factor Social

Variable	Código	Categoría
Nivel de estudios	0	ninguno/preescolar
	1	primaria
	2	Secundaria
	3	superior no universitaria
	4	superior universitaria
Nivel de estudios de la pareja	0	Ninguno/preescolar
	1	Primaria
	2	Secundaria
	3	Superior
	4	No tiene pareja
Alfabetismo	0	No puede leer en absoluto
	1	Lee con dificultad
	2	Lee facilmente
índice de riqueza	0	Muy pobre
	1	Pobre
	2	Medio
	3	Rico
	4	Muy rico
violencia	0	No
	1	Si
Número de pareja sexual	0	ninguno
	1	1
	2	2 +
nivel de anemia	0	No tiene
	1	Severo
	2	Moderado
	3	Leve
Le diagnosticaron diabetes o azúcar Alta	0	No
	1	Si
Le diagnosticaron hipertensión arterial o Presión Alta	0	No
	1	Si
En los últimos 12 meses ha fumado Cigarrillos	0	No
	1	Si
Ha consumido bebidas alcohólicas alguna Vez	0	No
	1	Si
Actividad última semana	0	Estudiante/buscó trabajo
	1	Trabajo esporádico
	2	Trabajo doméstico
	3	Otros
Trabaja actualmente	0	No
	1	Si
Actualmente está embarazada	0	No, no está seguro
	1	Si

## 4.7. Técnicas de recolección de datos

La técnica de recolección de datos fue la entrevista personal, ésta fue ejecutado por un personal oportunamente preparado para dicha actividad, quienes fueron los visitantes a las viviendas seleccionadas para aplicar los cuestionarios.

## 4.8. Instrumento de depresión

Para diagnosticar la depresión, la ENDES incluye en el cuestionario de salud en la sección 7 el instrumento Patient Health Questionnaire-9 (PHQ-9), ver anexo A. Instrumento que consta de 9 ítems y es válido para diagnosticar la depresión en el Perú (Calderón y otros , 2012).

La puntuación es de 0 a 27 (nunca = 0, varios días = 1, la mayoría de los días = 2, casi todos los días = 3) (Baader y otros , 2012).

### 4.8.1. Método de puntuación para el diagnóstico de depresión PHQ-9

La puntuación total de gravedad puede ser usado para rastrear cambios en la gravedad a través del tiempo y puede ser interpretada como sigue (Montenegro y Kilstein, 2000):

Código	Puntaje	Categoría
0:	0 – 4	sin depresión
1:	5 – 9	depresión leve
2:	10 – 14	depresión moderada
3:	15 – 19	depresión severa
4:	20 – 27	depresión muy severa

## **4.9. Análisis e interpretación de la información**

Previo a todo se realizó la limpieza de datos, posteriormente se seleccionaron los sujetos y variables de interés. Las variables seleccionadas fueron categorizadas y codificadas de acuerdo al objetivo de la investigación.

### **4.9.1. Procesamiento y análisis estadístico**

Se realizó un análisis exploratorio univariado, de todas las variables de interés usando medidas de frecuencia y porcentaje para todas las variables puesto que se tienen variables categorizadas. Posteriormente se realizó un análisis bivariado, de los factores sociodemográficos y la depresión a través de la prueba de chi-cuadrado, ésto con el objetivo de encontrar asociación y finalmente se realizó análisis de regresión logística ordinal cuyo modelo aplicado es odds proporcionales ya que cumple el supuesto de la líneas paralelas teniendo como variable respuesta la depresión, con las variables: edad, nivel de instrucción, estado civil, región, tabaquismo, alcoholismo, ocupación y entre otras variables de interés. Se tuvo en cuenta un nivel de confianza del 95 % y un nivel de significancia menor de 0.05 %.

## **4.10. Aspectos éticos**

La presente investigación realiza un análisis de datos secundarios, tomando como base la Encuesta Demográfica de Salud Familiar correspondiente al año 2015. Como se realiza análisis de datos secundarios no se realizaron entrevistas personales y el riesgo de ser afectados son mínimos, además estos datos se basan fundamentalmente en la brecha de confidencialidad ya que no contamos con registros que tengan identificación personal como: DNI, nombres, número de teléfono, número de historia clínica o dirección de vivienda. Estos datos se obtuvieron en la base de datos publicadas por el Instituto Nacional de Estadística e Informática (INEI), <http://iinei.inei.gob.pe/microdatos/>.

# Capítulo 5

## RESULTADOS Y DISCUSIONES

### 5.1. Análisis univariado

La investigación se realizó con 13763 mujeres encuestadas en la ENDES-2015, y en la Tabla 5.1, se presenta los niveles de depresión de las mujeres de 15 a 49 años, según el PHQ-9, donde se observa que el 19.44 % presentan depresión leve, 7.07 % depresión moderada 3.83 % depresión severa y 2.91 % depresión muy severa; Con la finalidad de aplicar el análisis de regresión logística ordinal, estas categorías fueron posteriormente recategorizados en: sin depresión (66.75 %), depresión leve (19.44 %) y depresión modera/severa (13.81 %) porque algunas de éstas presentan un porcentaje bajo.

TABLA 5.1: Nivel de depresión de las mujeres de 15 a 49 años, ENDES - 2015

Variable	Categoría	Frecuencia	%
Categorías según PHQ-9	Sin depresión	9187	66.75
	Depresión leve	2676	19.44
	Depresión moderada	973	7.07
	Depresión severa	527	3.83
	Depresión muy severa	400	2.91
Depresión en categorías	Sin depresión	9187	66.75
	Depresión leve	2676	19.44
	Depresión severa	1900	13.81

La mayoría de las mujeres encuestadas se encuentran entre los 20 a 34 años



(55.0 %), estado civil conviviente (50.7 %), nivel de estudios alcanzado secundaria (43.8 %), lugar de residencia en su mayoría son de la sierra (34.5 %) y nivel de estudios de la pareja secundaria (58.3 %), 5.2

TABLA 5.2: Características demográficas de mujeres de 15 a 49 años, ENDES - 2015

Variable	Categoría	Frecuencia	%
Edad	15 - 19	1666	12.1
	20 - 34	7573	55.0
	35 - 49	4524	32.9
Estado civil	soltera	1376	10.0
	casada	3648	26.5
	conviviente	6973	50.7
	viuda/divorciada	1766	12.8
Nivel de estudios	hasta primaria	4101	29.8
	Secundaria	6030	43.8
	superior	3632	26.4
Nivel de estudios de la pareja	hasta primaria	2842	20.6
	Secundaria	8022	58.3
	superior	2899	21.1
Región natural	Sierra	4755	34.5
	Selva	3388	24.6
	Resto Costa	4084	29.7
	Lima metropolitana	1536	11.2

## 5.2. Análisis bivariado

La Tabla 5.3 muestra resultados bivariados de la depresión y las variables demográficas en estudio. Por ende se presentan porcentajes altos en todas las características demográficas sin depresión, seguido por la depresión leve y porcentajes mínimos en depresión moderada. Siendo el porcentaje alto de 23.5 % en el grupo de las mujeres viudas o divorciadas con depresión leve y el mínimo de 17.3 % en el grupo del nivel de estudio superior también con depresión leve. mientras que el Porcentaje máximo en el grupo de mujeres que son viudas o divorciadas fue 21.0 % con depresión moderada o severa y el mínimo porcentaje en el grupo de los que viven en la costa menos Lima metropolitana con depresión moderada o severa fue 43.1 % .

TABLA 5.3: Depresión según características demográficas

Variables demográficas		Depresión							
		No		Leve		Moderada/severa		Total	
		n	%	n	%	n	%	n	%
Edad	15 - 19	1107	66.4	340	20.4	219	13.1	1666	100.0
	20 - 34	5224	69.0	1436	19.0	913	12.1	7573	100.0
	35 - 49	2856	63.1	900	19.9	768	17.0	4524	100.0
Estado civil	soltera	908	66.0	280	20.3	188	13.7	1376	100.0
	casada	2504	68.6	671	18.4	473	13.0	3648	100.0
	conviviente	4795	68.8	1310	18.8	868	12.4	6973	100.0
	viuda/divorciada	980	55.5	415	23.5	371	21.0	1766	100.0
Nivel de estudio	hasta primaria	2691	65.6	806	19.7	604	14.7	4101	100.0
	secundaria	3963	65.7	1241	20.6	826	13.7	6030	100.0
	superior	2533	69.7	629	17.3	470	12.9	3632	100.0
Nivel de estudio de la pareja	hasta primaria	1909	67.2	559	19.7	374	13.2	2842	100.0
	secundaria	5299	66.1	1581	19.7	1142	14.2	8022	100.0
	superior	1979	68.3	536	18.5	384	13.2	2899	100.0
Región natural	sierra	2932	61.7	1041	21.9	782	16.4	4755	100.0
	selva	2370	70.0	604	17.8	414	12.2	3388	100.0
	resto costa	2864	70.1	734	18.0	486	11.9	4084	100.0
	Lima metropolitana	1021	66.5	297	19.3	218	14.2	1536	100.0
Total		9187	66.8	2676	19.4	1900	13.8	13763	100.0

Las características demográficas y sociales está asociadas significativamente al nivel de depresión a excepción del nivel de estudios de la pareja, lugar de residencia, actualmente está embarazada y alfabetismo no están asociados con la depresión ( $p - valor > 0.05$ ). ( ver Tabla 5.4)

TABLA 5.4: Factores sociodemográficos asociados a la depresión.

Variable independiente	Depresión	
	$\chi^2$	p-valor
Edad	66.99	0.000**
Estado civil	136.53	0.000**
Nivel de estudios	23.70	0.000**
Nivel de estudios-pareja	5.97	0.202
Región natural	95.51	0.000**
Lugar de residencia	5.80	0.446
Trabaja actualmente	32.85	0.000**
Actualmente está embarazada	0.45	0.800
Alfabetismo	1.29	0.525
Índice de riqueza	36.09	0.000**
Niños menores de 5 años	41.71	0.000**
Violencia por parte de la pareja	686.88	0.000**
Número de parejas sexuales	87.50	0.000**
Anemia	6.03	0.049*
Diagnóstico de diabetes	29.37	0.000**
Diagnóstico de presión alta	105.98	0.000**
En los últimos 12 meses ha fumado cigarro	73.84	0.000**
Ha consumido bebidas alcohólicas alguna vez	78.54	0.000**

\* $p < 0.05$ , \*\*  $p < 0.01$

### Regresión logística ordinal

Los niveles de depresión es recategorizada en 1: sin depresión 2: depresión leve 3: depresión moderada/severa presentando un orden creciente que permite alcanzar el objetivo de la investigación para lo cual se requiere ajustar el modelo de regresión logística ordinal de odds proporcionales basado en la prueba de hipótesis de líneas paralelas requeridas para este modelo

$H_0$ : Los  $\beta_i$  son los mismos para todos los niveles de respuesta.

$H_1$ : Los  $\beta_i$  no son los mismos para todos los niveles de respuesta

Supuesto que no es rechazado, dado que el valor del estadístico fue  $\chi^2 = 23.81$  ( $P - valor = 0.472$ ), detalles en Anexo C.8 de tal manera que se puede concluir que los parámetros de ubicación son los mismos para todas las categorías de respuesta (no tiene depresión, depresión leve y depresión moderada/severa), de esta manera podemos indicar el cumplimiento del supuesto de líneas paralelas.

Además, previo a ello se ha realizado las evaluaciones correspondientes para dicho ajuste. Las Tablas que a continuación mencionaremos lo encontraremos en el Anexo C. Primeramente se hizo la evaluación del modelo insertando todas las variables sociales y demográficas mencionadas en este estudio.

Para ello se realizó el registro de las iteraciones; en la iteración (iteración 0), se ajustó a primer modelo llamado modelo nulo (sin predictores), en efecto el modelo solo con el intercepto. Seguidamente se ajusta el modelo completo; de acuerdo al proceso de las iteraciones, estas iteraciones se ha detenido cuando la diferencia en la probabilidad de registro entre estas iteraciones repetidas se conviertan muy pequeñas, mostrándonos así un log de verosimilitud de -11262.268 en la iteración 4, ver la Tabla C.1, siendo las hipótesis las siguiente.

$$H_0 : -2\ln \left[ \frac{\text{verosimilitud del modelo sin variables explicativas}}{\text{verosimilitud del modelo sin variables explicativas}} \right] = 1$$

$$H_1 : -2\ln \left[ \frac{\text{verosimilitud del modelo sin variables explicativas}}{\text{verosimilitud del modelo sin variables explicativas}} \right] \neq 1$$

También se realizó la prueba de Omnibus, cuya hipótesis de prueba es:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_m = 0 .$$

$$H_1: \exists! \beta_i \neq 0 \quad i = 0, \dots, m$$

Entonces el estadístico chi-cuadrado de razón de verosimilitud fue de 1191.39 con 28 grados de libertad cuyo  $p - valor = 0.000 < 0.05$ . Éste estadístico indica que el modelo obtenido es estadísticamente significativo, frente al modelo nulo. De esta manera hemos comprobado que todos los coeficientes del modelo no son nulos

y que las variables explicativas obtenidas fueron participes en la predicción del modelo de manera significativa, tabla C.2.

Así mismo se ha evaluado la prueba de efectos del modelo, cuya hipótesis es:

$H_0$ : Los efectos del modelo especificado conjuntamente son iguales a cero.

$H_1$ : Los efectos del modelo especificado conjuntamente no son iguales a cero.

en la Tabla C.3, el estadístico de prueba es  $\chi^2$  de Wald, y muestra algunos  $p - valor > 0.05$ , de esto podemos indicar que las variables; actualmente trabaja, está embarazada, alfabetismo y diagnóstico de anemia, son no significativas para el modelo, es decir no tienen algún efecto perceptible en el modelo. De esta manera se han extraído éstas variables del modelo

De esta manera se desarrolló un primer análisis de la regresión logística ordinal insertando todas las variables demográficas y sociales en estudio con sus respectivas categorías, tal como se muestran en la Tabla C.4, de ellas podemos indicar que las variables: actualmente trabaja, el embarazo, alfabetismo y la anemia presentan un  $p - valor > 0.05$  (no significativas) en todas sus categorías, por esta razón se realiza un segundo análisis insertando sólo las variables que al menos una categoría tuvo un  $p - valor < 0.05$ , obteniéndose así los resultados en la Tabla 5.5.

Para realizar el ajuste con las variables demográficas y sociales significativas también se realizó las pruebas correspondientes: como el registro de la iteraciones; una vez más en la iteración 0, donde se ajusta un modelo nulo, luego continuamos con el ajuste del modelo completo, hasta obtener diferencia mínima de las probabilidades de registro entre estas iteraciones sucesivas, esta diferencia encontramos en la iteración 4, cuyo valor de log de verosimilitud es  $-11265.806$ , Tabla C.5.

También se ha realizado la prueba del ómnibus cuyo chi-cuadrado de razón de verosimilitud fue  $1184.31$  con 24 grados de libertad con un  $p - valor = 0.000 < 0.05$ , indicándonos así que nuestro modelo en su totalidad es estadísticamente

significativo, frente al modelo nulo. Además comprobamos que la totalidad de los coeficientes del modelo son distintos a cero, por lo que las variables demográficas y sociales participantes en el modelo predicen significativamente, ver la tabla C.6. Además se realizó las pruebas de efectos del modelo de cada variable, en este caso todas las variables son significativas ( $p - valor < 0.05$ ). Indicándonos así que todas las variables entran al modelo, ver Tabla C.7.

Por último se realizó el supuesto de líneas paralelas (razones de probabilidades u odds proporcionales), el  $p - valor = 0.472 > 0.05$  para el modelo general por tanto cumple con el supuesto del paralelismo, además observamos que se cumple para cada una de las variables y sus categorías; ésto nos indica que los coeficiente de la variable es el mismo para cada una de las categorías de respuesta, Anexo, Tabla C.8.

La Tabla 5.5, la estimación de los parámetros de la regresión logística ordinal cuyo modelo aplicado es odds proporcionales o también llamado modelo de rectas paralelas. Por medio del paquete STATA hemos obtenido el modelo odds proporcionales, se tiene entonces:

$$\begin{aligned}
 f(\gamma_1(X)) &= \text{logit}(P(\gamma \leq y_1|X)) \\
 &= \alpha_1 - \beta^T X \\
 &= \alpha_1 - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m) \\
 &= 0,828 - (-0,278(E20 - 34) - 0,99(E35 - 49) - 0,443(casada) \\
 &\quad - 0,483(conviviente) - 0,23(viuda/divorciada) - 0,077(secundaria) \\
 &\quad - 0,239(superior) + 0,126(parsecundaria) + 0,036(parsuperior) \\
 &\quad - 0,445(selva) - 0,451(restocosta) - 0,325(Limametropolitana) \\
 &\quad + 0,161(pobre) + 0,215(medio) + 0,045(rico) + 0,017(muyrico) \\
 &\quad - 0,198(niños < 5años) + 0,932(violencia) + 0,232(2 + Parejassex) \\
 &\quad + 0,578(diabetes) + 0,579(hipertensión) + 0,514(fuma) \\
 &\quad + 0,449(alcohol) + 0,035(miembrosdela familia))
 \end{aligned}$$

De manera análoga tenemos para el modelo del siguiente nivel, donde solamente varia en la constante, presentamos a continuación.

$$\begin{aligned}
f(\gamma_2(X)) &= \text{logit}(P(\gamma \leq y_2|X)) \\
&= \alpha_2 - \beta^T X \\
&= \alpha_2 - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m) \\
&= 2,043 - (-0,278(E20 - 34) - 0,99(E35 - 49) - 0,443(\text{casada}) \\
&\quad -0,483(\text{conviviente}) - 0,23(\text{viuda/divorciada}) - 0,077(\text{secundaria}) \\
&\quad -0,239(\text{superior}) + 0,126(\text{parsecundaria}) + 0,036(\text{parsuperior}) \\
&\quad -0,445(\text{selva}) - 0,451(\text{restocosta}) - 0,325(\text{Limametropolitana}) \\
&\quad +0,161(\text{pobre}) + 0,215(\text{medio}) + 0,045(\text{rico}) + 0,017(\text{muyrico}) \\
&\quad -0,198(\text{niños} < 5\text{años}) + 0,932(\text{violencia}) + 0,232(2 + \text{Parejassex}) \\
&\quad +0,578(\text{diabetes}) + 0,579(\text{hipertensi3n}) + 0,514(\text{fuma}) \\
&\quad +0,449(\text{alcohol}) + 0,035(\text{miembrosdelafamilia})
\end{aligned}$$

En el lado derecho de los modelos se muestra un modelo lineal simple con pendiente  $\beta$  y una intersecci3n que cambia de acuerdo a  $y_i$  y  $\alpha_i$  donde  $y_i$  viene hacer el nivel de una categor3a ordenada con los niveles de  $j = 1, 2, \dots, k - 1$ . En nuestro caso tuvimos 3 categor3as por lo tanto  $j = 1, 2$ , no debemos confundir la notaci3n de la respuesta de la variable  $Y$  con las categor3as. En nuestro caso cuando  $y = 0$  entonces decimos que estamos en la primera categor3a es decir:  $y_1 = 1$  es la categor3a no tiene depresi3n. Cuando  $P(Y \leq 1)$  nos referimos a la probabilidad de ser una persona con el nivel de que *no tiene depresi3n* y si  $P(Y \leq 2)$  nos referimos a la probabilidad de ser una persona con el nivel de que *no tiene depresi3n o tiene depresi3n leve*. En este modelo el nivel m3s alto responde a una probabilidad de 1, es decir  $P(Y \leq y_3) = P(Y \leq 3) = 1$  por lo que no es necesario modelar.

Si el odds ratio es menor que uno, esto sucede cuando el coeficiente de la variable regresora es negativo, esto indica que, si las otras variables independientes permanecen constantes, los cambios en la variable independiente analizada aumentan la probabilidad de obtener categor3as de mayor valor en la variable objetivo. Mientras que, si los Odds Ratio son mayores que uno, esto demuestra que las variaciones en la variable explicativa disminuye el riesgo de obtener categor3as de

mayor valor de la variable objetivo. Mientras que si el odds ratio es 1 entonces diremos que el efecto es indiferente sobre los grupos expuestos. seguidamente procederemos a interpretar los Odds Ratio más relevantes. Las categorías señaladas en rojo no presentan diferencias significativas respecto a la categoría de referencia porque el valor el intervalo del odds ratio incluye a la unidad.

<b>Variable (Vs)</b>	<b>V. Categoría: Descripción</b>	<b>OR</b>	<b>Interpretación (Conservando las demás variables)</b>
Edad (15-19)	20 - 34	0.757	Una mujer que tiene la edad de 15 a 19 años tiene mayor riesgo de tener depresión ya sea leve o moderada/severa, frente a las mujeres de edad de 20 a 34 años.
Estado civil (Soltera)	Casada	0.642	Una mujer soltera tiene mayor riesgo de tener depresión ya sea leve o moderada/severa, respecto a una mujer casada.
	Conviviente	0.617	Una mujer soltera tiene mayor riesgo de tener depresión ya sea leve o moderada/severa, respecto a una mujer conviviente.
	viuda o divorciada	0.794	Una mujer soltera tiene mayor riesgo de tener depresión ya sea leve o moderada/severa, respecto a una mujer viuda o divorciada .
Nivel de estudio (primaria)	superior	0.787	Una mujer que tiene estudios hasta la primaria tiene mayor riesgo de tener depresión ya sea leve o moderada/severa, que alguien que tenga estudios de nivel primaria.
Nivel de estudio de la pareja (primaria)	secundaria	1,134	La mujer que tiene pareja con estudios hasta la secundaria tiene 1.134 veces más posibilidades de tener depresión ya sea leve o moderada/severa, que alguien que tenga pareja con estudios de nivel primaria.



<b>Variable (Vs)</b>	<b>V. Categoría: Descripción</b>	<b>OR</b>	<b>Interpretación</b> (Conservando las demás variables)
Región (sierra)	Selva	0.641	Una mujer que vive en la sierra tiene mayor riesgo a sufrir depresión ya sea leve o moderada/severa, respecto a una que vive en la selva.
	Resto de la costa	0.637	Una mujer que vive en la sierra tiene mayor riesgo de sufrir depresión ya sea leve o moderada/severa, respecto a una que vive en la costa menos Lima.
	Lima metropolitana	0.722	Una mujer que vive en la sierra presenta mayor riesgo de sufrir depresión ya sea leve o moderada/severa, respecto a una mujer que vive Lima metropolitana.
Riqueza (Muy pobre)	Pobre	1.174	Una mujer pobre tiene 1.174 veces más posibilidades de sufrir depresión ya sea leve o moderada/severa, respecto a una mujer muy pobre.
	Medio	1.240	Una mujer de riqueza medio tiene 1.24 veces más posibilidades de sufrir depresión ya sea leve o moderada/severa, respecto a una mujer muy pobre.
Niños < 5 años años (No)	Si	0.820	Tener niños menores de 5 años en casa presenta mayor riesgo de sufrir depresión ya sea leve o moderada/severa, frente a los que no tienen niños menores de 5 años en casa.
Violencia (No)	Si	2.540	Sufrir violencia en casa es 2.54 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que no sufren violencia.
Nº parejas sexuales (< 2)	2+	1.261	Tener 2 o más parejas sexuales es 1.261 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que tienen menos de dos parejas sexuales.

<b>Variable</b> (Vs)	<b>V. Categoría:</b> <b>Descripción</b>	<b>OR</b>	<b>Interpretación</b> (Conservando las demás variables)
Diabetes (No)	Si	1.782	El diagnóstico positivo de diabetes es 1.782 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que no se diagnosticaron.
Hipertensión (No)	Si	1.784	El diagnóstico positivo de hipertensión es 1.784 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que no se diagnosticaron.
Fuma (No)	Si	1.672	El hábito de fumar es 1.672 veces más probable de tener depresión ya sea leve o moderada/severa, frente a los que no fuman.
Alcohol (No)	Si	1.556	La frecuencia de beber alcohol es 1.566 veces más probable de tener depresión ya sea leve o moderada/severa, frente a los que no beben.

TABLA 5.5: Estimación de parámetros para la regresión logística ordinal del factor sociodemográfico.

Parámetro	$\hat{\beta}$	SE	$\chi^2$	Sig.	$e^{\hat{\beta}}$	IC. 95 % $e^{\hat{\beta}}$	
						Inf.	Sup.
<b>Depresión</b>							
Leve	.828	.1251	43.795	.000	2.288	1.791	2.924
Moderado/severa	2.043	.1264	261.285	0.000	7.717	6.023	9.887
<b>Edad</b>							
20 - 34	-.278	.0614	20.535	.000	.757	.671	.854
35 - 49	-.099	.0645	2.356	.125	.906	.798	1.028
<b>Estado civil</b>							
Casada	-.443	.0904	23.992	.000	.642	.538	.767
Conviviente	-.483	.0877	30.336	.000	.617	.519	.732
Viuda/divorciada	-.230	.0972	5.605	.018	.794	.657	.961
<b>Nivel de estudio</b>							
Secundaria	-.077	.0502	2.367	.124	.926	.839	1.021
Superior	-.239	.0654	13.428	.000	.787	.692	.895
<b>Nivel de estudio-pareja</b>							
Secundaria	.126	.0535	5.517	.019	1.134	1.021	1.259
Superior	.036	.0829	.187	.666	1.036	.881	1.219
<b>Región</b>							
Selva	-.445	.0494	80.973	0.000	.641	.582	.706
Resto de la costa	-.451	.0498	82.011	0.000	.637	.578	.702
Lima metropolitana	-.325	.0692	22.075	.000	.722	.631	.827
<b>Riqueza</b>							
Pobre	.161	.0542	8.801	.003	1.174	1.056	1.306
Medio	.215	.0628	11.707	.001	1.240	1.096	1.402
Rico	.045	.0733	.371	.542	1.046	.906	1.207
Muy rico	.017	.0885	.036	.849	1.017	.855	1.210
<b>Niños menores de 5 años</b>							
Si	-.198	.0480	17.097	.000	.820	.747	.901
<b>Violencia</b>							
Si	.932	.0405	529.284	0.000	2.540	2.346	2.750
<b>Número de parejas sexuales</b>							
2+	.232	.0389	35.518	.000	1.261	1.169	1.361
<b>Diabetes</b>							
Si	.578	.1726	11.191	.001	1.782	1.270	2.499
<b>Hipertensión</b>							
Si	.579	.0788	53.943	.000	1.784	1.528	2.082
<b>Fuma</b>							
Si	.514	.0710	52.343	.000	1.672	1.455	1.922
<b>Alcohol</b>							
Si	.449	.0623	51.825	.000	1.566	1.386	1.770
<b>Miembros de la familia</b>							
	.035	.0118	8.913	.003	1.036	1.012	1.060

El análisis e interpretación de la Tabla 5.6 es análogo a la Tabla anterior, en este caso se ha realizado el modelo del odds ratio sólo para el factor demográfico, realizando las pruebas estadísticas adecuados y básicamente el supuesto de las rectas paralelas ( $p - valor = 0.252$ ), Anexo C.1.2.

Entonces los modelo para el factor demográfico serán:

$$\begin{aligned}
 f(\gamma_1(X)) &= \text{logit}(P(\gamma \leq y_1|X)) \\
 &= \alpha_1 - \beta^T X \\
 &= \alpha_1 - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m) \\
 &= 0.372 - (-0.71(E20 - 34) + 1.165(E35 - 49) - 0.112(\text{casada}) \\
 &\quad - 0.034(\text{conviviente}) + 0.492(\text{viuda/divorciada}) - 0.381(\text{selva}) \\
 &\quad - 0.398(\text{restocosta}) - 0.253(\text{Limametropolitana}))
 \end{aligned}$$

De manera análoga tenemos el modelo del siguiente nivel, donde solamente varia en la constante:

$$\begin{aligned}
 f(\gamma_2(X)) &= \text{logit}(P(\gamma \leq y_2|X)) \\
 &= \alpha_2 - \beta^T X \\
 &= \alpha_2 - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m) \\
 &= 1.525 - (-0.71(E20 - 34) + 1.165(E35 - 49) - 0.112(\text{casada}) \\
 &\quad - 0.034(\text{conviviente}) + 0.492(\text{viuda/divorciada}) - 0.381(\text{selva}) \\
 &\quad - 0.398(\text{restocosta}) - 0.253(\text{Limametropolitana}))
 \end{aligned}$$

A continuación se realiza la interpretación de los Odds ratio.

<b>Variable (Vs)</b>	<b>V. Categoría: Descripción</b>	<b>OR</b>	<b>Interpretación</b> (Conservando las demás variables)
Edad (15-19)	35 - 49	1.179	Una mujer entre las edades de 35 a 49 años es 1.179 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que tienen la edad de 15 a 19 años.
Estado civil (Soltera)	viuda o divorciada	1.635	Una mujer que es viuda/divorciada tiene 1.635 veces más posibilidad de sufrir depresión ya sea leve o moderada/severa, frente a los que son solteras.
Región (sierra)	Selva	0.683	Una mujer que vive en la sierra tiene mayor riesgo a sufrir depresión ya sea leve o moderada/severa, respecto a una que vive en la selva.
	Resto de la costa	0.671	Una mujer que vive en la sierra tiene mayor riesgo de sufrir depresión ya sea leve o moderada/severa, respecto a una que vive en la costa menos Lima.
	Lima metropolitana	0.777	Una mujer que vive en la sierra presenta mayor riesgo de sufrir depresión ya sea leve o moderada/severa, respecto a una mujer que vive Lima metropolitana.
Niños < 5 años (No)	Si	0.846	Tener niños menores de 5 años en casa presenta mayor posibilidad de sufrir depresión ya sea leve o moderada/severa, frente a las que no tienen niños menores de 5 años en casa.

TABLA 5.6: Estimación de parámetros para la regresión logística ordinal del factor demográfico.

Parámetro	$\hat{\beta}$	SE	$\chi^2$	Sig.	$e^{\hat{\beta}}$	IC. 95% $e^{\hat{\beta}}$	
						Inf.	Sup.
<b>Depresión</b>							
Leve	.372	.075	24.810	.000	1.450	1.253	1.679
Moderado/severa	1.525	.976	400.449	0.000	4.597	3.959	5.338
<b>Edad</b>							
20 - 34	-.071	.058	1.533	.216	.931	.832	1.042
35 - 49	1.165	.060	7.486	.006	1.179	1.048	1.327
<b>Estado civil</b>							
Casada	-.112	.068	2.686	.101	.894	.782	1.022
Conviviente	-.034	.064	0.258	.593	.966	.852	1.096
Viuda/divorciada	.492	.074	44.266	.000	1.635	1.415	1.890
<b>Región</b>							
Selva	-.381	.048	64.003	0.000	.683	.622	.750
Resto de la costa	-.398	.045	79.223	0.000	.671	.615	.733
Lima metropolitana	-.253	.061	17.189	.000	.777	.689	.875
<b>Niños menores de 5 años</b>							
Si	-.167	.044	14.558	.000	.846	.777	.922

En este sentido también realizamos el análisis e interpretación de la Tabla 5.7 lo cual también es análogo a las interpretaciones anteriores, una vez más hemos realizado los pasos adecuados para realizar este modelo de odds proporcionales ahora sólo para el factor social, realizando las pruebas estadísticas adecuados y básicamente el supuesto de las rectas paralelas ( $p - valor = 0.428$ ), Anexo C.1.3. Entonces los modelo para el factor social serán:

$$\begin{aligned}
 f(\gamma_1(X)) &= \text{logit}(P(\gamma \leq y_1|X)) \\
 &= \alpha_1 - \beta^T X \\
 &= \alpha_1 - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m) \\
 &= 1.549 - (-0.127(\text{secundaria}) - 0.313(\text{superior}) + 0.139(\text{trabaja}) \\
 &\quad + 0.926(\text{violencia}) + 0.192(2 + \text{Parejassex}) + 0.593(\text{diabetes}) \\
 &\quad + 0.604(\text{hipertensión}) + 0.493(\text{fuma}) + 0.345(\text{alcohol}))
 \end{aligned}$$

y el modelo para el siguiente nivel es:

$$\begin{aligned}
f(\gamma_2(X)) &= \text{logit}(P(\gamma \leq y_2|X)) \\
&= \alpha_2 - \beta^T X \\
&= \alpha_2 - (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m) \\
&= 2.748 - (-0.127(\textit{secundaria}) - 0.313(\textit{superior}) + 0.139(\textit{trabaja}) \\
&\quad + 0.926(\textit{violencia}) + 0.192(2 + \textit{Parejassex}) + 0.593(\textit{diabetes}) \\
&\quad + 0.604(\textit{hipertensi3n}) + 0.493(\textit{fuma}) + 0.345(\textit{alcohol}))
\end{aligned}$$

Interpretaci3n del odds ratio

<b>Variable</b> (Vs)	<b>V. Categoría:</b> <b>Descripción</b>	<b>OR</b>	<b>Interpretación</b> (Conservando las demás variables)
Nivel de estudio (Hasta primaria)	secundaria	0.881	Una mujer con estudios hasta la primaria tiene mayor posibilidad de sufrir depresión ya sea leve o moderada/severa, respecto a una mujer con estudios de nivel secundaria.
	Superior	0,686	Una mujer con estudios hasta la primaria tiene mayor posibilidad de sufrir depresión ya sea leve o moderada/severa, frente a una mujer con estudios superiores.
Nivel de estudio de la pareja (Hasta primaria)	secundaria	1.139	El tener una pareja con estudios de nivel secundaria es 1.139 veces más probable de sufrir depresión ya sea leve o moderada/severa, respecto a una mujer que tenga pareja con nivel de estudio hasta primaria
	Superior	1.368	El tener una pareja con estudios superiores es 1.368 veces más probable de sufrir depresión ya sea leve o moderada/severa, a una mujer con educación primaria o no.
¿Trabaja? (No)	Si	1.149	El tener un trabajo es 1.149 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que no trabajan.
Violencia (No)	Si	2.524	Sufrir violencia en casa es 2.524 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que no sufren violencia.
Nº parejas sexuales (<2)	2+	1.212	Tener 2 o más parejas sexuales es 1.212 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que tienen menos de dos parejas sexuales.
Diabetes (No)	Si	1.810	El presentar diagnóstico positivo de diabetes es 1.810 veces más probable de tener depresión ya sea leve o moderada/severa, frente a que no lo tengan.



<b>Variable</b> (Vs)	<b>V. Categoría:</b> <b>Descripción</b>	<b>OR</b>	<b>Interpretación</b> (Conservando las demás variables)
Hipertensión (No)	Si	1.829	El diagnóstico positivo de hipertensión es 1.829 veces más probable de sufrir depresión ya sea leve o moderada/severa, frente a los que no se diagnosticaron.
Fuma (No)	Si	1.429	El hábito de fumar es 1.429 veces más probable de tener depresión ya sea leve o moderada/severa, frente a los que no fuman.
Alcohol (No)	Si	1.412	La frecuencia de beber alcohol es 1.412 veces más probable de tener de presión ya sea leve o moderada/severa, frente a los que no beben.

TABLA 5.7: Estimación de parámetros para la regresión logística ordinal del factor social.

Parámetro	$\hat{\beta}$	SE	$\chi^2$	Sig.	$e^{\hat{\beta}}$	IC. 95% $e^{\hat{\beta}}$	
						Inf.	Sup.
<b>Depresión</b>							
Leve	1.549	.0703	485.996	.000	4.707	4.101	5.402
Moderado/severa	2.748	.0733	1403.981	.000	15.611	13.521	18.024
<b>Nivel de estudio</b>							
Secundaria	-.127	.0472	7.243	.007	.881	.803	.966
Superior	-.376	.0581	41.923	.000	.686	.612	.769
<b>Nivel de estudio-pareja</b>							
Secundaria	.130	.0518	6.323	.012	1.139	1.029	1.261
Superior	.313	.0668	21.986	.000	1.368	1.200	1.559
<b>¿Trabaja actualmente?</b>							
Si	.139	.0389	12.838	.000	1.149	1.065	1.240
<b>Violencia</b>							
Si	.926	.0389	566.951	.000	2.524	2.339	2.724
<b>Número de parejas sexuales</b>							
2+	.192	.0372	26.684	.000	1.212	1.127	1.304
<b>Diabetes</b>							
Si	.593	.1703	12.149	.000	1.810	1.297	2.527
<b>Hipertensión</b>							
Si	.604	.0775	60.661	.000	1.829	1.571	2.129
<b>Fuma</b>							
Si	.493	.0693	50.509	.000	1.637	1.429	1.875
<b>Alcohol</b>							
Si	.345	.0596	33.513	.000	1.412	1.257	1.587

# DISCUSIÓN

La presente investigación tuvo como objetivo determinar los factores sociodemográficos que inducen a la depresión en la población de mujeres peruanas de 15 a 49 años participantes en la encuesta demográfica de salud en el 2015.

La prevalencia de depresión en mujeres peruanas fue de 33.2 %, de ellas 58.5 % tienen depresión leve, mientras que la depresión moderada fue 21.3 %, severa 11.5 % y muy severa 8.7 %. Además hemos encontrado asociación bivariada estadísticamente significativa de la depresión con: la edad, estado civil, nivel de educación, región natural, trabaja actualmente, índice de riqueza, niños menores de 5 años, violencia por parte de la pareja, número de parejas sexuales, anemia, diagnóstico de diabetes, diagnóstico de presión alta, el haber fumado en los últimos 12 meses y el haber consumido alcohol alguna vez.

Por otra parte con el análisis multivariante, se han encontrado los incidentes a la depresión tales como: la edad, estado civil, educación de la misma y la pareja, región, riqueza, niños menores de 5 años, violencia por parte de la pareja, número de parejas sexuales, diagnóstico de diabetes, diagnóstico de hipertensión, el haber fumado en los últimos 12 meses, el haber consumido alcohol alguna vez y número de miembros de la familia, siendo la más significativa la violencia por parte de la pareja con Odds Ratio=2.540, intervalo de confianza al 95 %: 2.346 – 2.750.

Según Chahua y otros, la prevalencia de la depresión fue de 14.6 %, en comparación a nuestra investigación es menos en más del 50 % ya que en nuestra

investigación encontramos un 33.2 %. las predominantes que llevan a la depresión en su estudio es el consumo de cocaína, en el análisis multivariante (regresión logística binomial) encontraron asociación con el sexo femenino, no tener hogar, consumir ketamina y menor apoyo confidencial.

También hemos hecho una revisión literaria a la investigación de Ramírez-Ruiz y otros, ellos encontraron la prevalencia global de depresión; baja, media y alta de 31.7 %, 56.3 % y 12.0 % respectivamente entre los que tienen depresión, además desglosaron la depresión de acuerdo al sexo, mostrando así para los hombres la prevalencia de depresión baja, media y alta de 47.3 %, 40.9 % y 11.8 %, mientras que en las mujeres fue 25 %, 63 % y 12 %, respectivamente, evidenciando así, que la depresión se da más en las mujeres. Utilizando la regresión lineal múltiple, cuya variable dependiente fue el índice de depresión, mostraron que la escolaridad y el género resultan ser factores significativas, en cuanto a la escolaridad en nuestra investigación también hemos encontrado significancia estadística.

Por otra parte también se debe considerar otros factores determinantes de la depresión. Niño-Avendaño y otros, 2014, encontraron una prevalencia de depresión de 18.95 %, y asociación estadísticamente significativa entre el uso de métodos hormonales y la depresión.

Alvarado-Moreno y otros, encontraron que la prevalencia de depresión en el grupo de los controles fue de 83.3 % entre leve, moderada y severa. Siendo la depresión leve más alta (57.7 %), y sólo un 3.8 % de depresión moderada. Mientras que en el grupo de casos fue 96.2 % con depresión entre leve, moderada y severa. siendo la más alta la depresión severa (38.4 %), seguido por la depresión leve (32.7 %) y moderada con un 25 %. Con la regresión logística hallaron asociación entre el síndrome de fibromialgia y ansiedad.

Gómez-Restrepo y otros, indicaron en su investigación que la prevalencia de episodios depresivos es 10 % (intervalo de confianza al 95 %: 9.2 % a 10.7 %), ésto en los últimos 12 meses antes de la encuesta realizada y 8.5 % (intervalo de confianza al 95 %: 7.8 % a 9.2 %) durante el último mes. Indican también que en ambos periodos las mujeres tuvieron una mayor proporción, que el más del 50 % de los episodios fueron moderados, tanto en hombres y mujeres.

Los factores relacionados a tener depresión en el último mes fueron: el sexo femenino, padecer de molestias y dolores, Considerar como regular o malo su estado de salud, dificultarse en el vínculo interpersonal, ser dependiente del alcohol, tomar medicamentos calmantes o estimulantes, consumo de sustancias adictivas o marihuana, como también el estar desempleado o tener alguna discapacidad.

Tenemos también la investigación desarrollada por Cabrera-Gutiérrez y otros, donde llegan a mostrar que el factor sociocultural (sentimiento) y el diagnóstico de cáncer están asociados a la depresión.

De acuerdo a las investigaciones revisadas se han encontrado la coincidencia de la asociación de muchas variables con nuestra investigación, como también la no coincidencia de otras variables. En muchas investigaciones se han hecho el estudio con otros factores que nosotros no incluimos en nuestra investigación esto porque no se han encontrado dichas variables en el cuestionario de la ENDES, tales como; el uso de métodos hormonales, síndrome de fibromialgia, ansiedad, calificar su estado de salud como malo o regular, soportar algunas molestias o dolores, dificultad con el vínculo interpersonal, consumo de sustancias adictivas o marihuana, consumo de medicamentos calmantes o estimulantes, sufrir de alguna discapacidad, que tengan diagnóstico de cáncer y entre otros.

# CONCLUSIONES

En nuestro estudio el 33.2% de las mujeres han presentado depresión de ellas el 19.4% depresión leve, 7.1% depresión moderada, 3.8% depresión severa y 2.9% depresión muy severa.

- Con respecto a los factores sociodemográficos, se ha encontrado asociación de la depresión con: nivel de estudios de la pareja hasta el nivel secundario, nivel de riqueza pobre, sufrir violencia por parte de su pareja, tener dos o más parejas sexuales, diagnóstico de diabetes o hipertensión, haber fumado en los últimos 12 meses y haber bebido alcohol alguna vez
- Los factores demográficos que inciden en la depresión son: la edad de 35 a 49 años, estado civil viuda o divorciada.
- Los factores sociales que inciden en la depresión son: nivel de estudios de la pareja hasta primaria, secundaria o superior; tener un trabajo; sufrir violencia por parte de su pareja; tener dos o más parejas sexuales; diagnosticado de diabetes o hipertensión; haber fumado en los últimos 12 meses y el haber bebido alcohol alguna vez.

# RECOMENDACIONES

A partir de lo anterior, se sugiere:

Debido al alto porcentaje de depresión que existe en mujeres con edades de 15 a 49 años, que participaron en la encuesta demográfica de salud familiar, 2015, sería oportuno la aplicación de pruebas de depresión más exhaustivas, para determinar los niveles de depresión, como también pruebas diagnósticas, ya que con frecuencia la depresión no se diagnostica y por ende no se trata.

Capacitación al personal quien llega a encuestar, para el abordaje de esta temática de manera apropiada, y el uso del instrumento que faculte la identificación adecuada y rápida de los individuos, además sería lo más loable desarrollar la encuesta de manera continua para la pronta identificación y de esta manera remitir los casos para la toma de decisiones, por otra parte hacer los seguimientos correspondientes a las mujeres que presentan depresión y sobre todo a las más graves.

También se recomienda realizar la evaluación de manera regular a los factores que son determinantes de la depresión, para así proyectar óptimas soluciones para el tratamiento y prevención de esta patología.

Es necesario la participación de los profesionales especialistas y de salud, en la solución progresiva del problema, formando parte de grupos interdisciplinarios que deben planificar programas de prevención e intervención en problemas de

depresión, considerando los factores asociados que se presentaron en la presente investigación.

También se recomienda incluir variables utilizadas en otros países para realizar el estudio de la depresión, en las ENDES.

y por último, se recomienda que el sector salud desarrolle actividades de intervención y prevención, para la toda la población del país, por grupos de edades y sobretodo en la educación escolar, ya que es la población en formación.



# REFERENCIAS

- Agresti, A. (2002). *Categorical data analysis* (Editora Jhon Wiley ed.).
- Aguirre, A. (2008). *Antropología de la depresión* (Vol. 8) (n.º 3). Descargado de [http://pepsic.bvsalud.org/scielo.php?script=sci\\_abstract&pid=S1518-61482008000300002&lng=pt&nrm=iso&tlng=es](http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S1518-61482008000300002&lng=pt&nrm=iso&tlng=es)
- Alberdi, S. (2006). *Depresión*. Descargado de [www.scamfyc.org/documentos/Depresion%20Fisterra.pdf](http://www.scamfyc.org/documentos/Depresion%20Fisterra.pdf)
- Baader, T., Molina, J. L., Venezian, S., Rojas, C., Farías, R., Fierro-Freixenet, C., ... Mundt, C. (2012). Validación y utilidad de la encuesta phq-9 (patient health questionnaire) en el diagnóstico de depresión en pacientes usuarios de atención primaria en Chile. *Revista chilena de neuro-psiquiatría*, 50(1), 10–22.
- Borooh, V. (2002). *Logit and probit*. Thousand Oaks, California. Descargado de <https://methods.sagepub.com/book/logit-and-probit>
- Calderón, M., Gálvez-Buccollini, J. A., Cueva, G., Ordoñez, C., Bromley, C., y Fiestas, F. (2012). Validación de la versión peruana del phq-9 para el diagnóstico de depresión. *Revista Peruana de Medicina Experimental y Salud Pública*, 29(4), 578–578.
- Chahua, M., y otros. (2014). *Depression in young regular cocaine users recruited in the community* (Vol. 28) (n.º 2). Gaceta Sanitaria. Descargado de [http://www.scielosp.org/scielo.php?script=sci\\_abstract&pid=S0213-91112014000200010&lng=en&nrm=iso&tlng=pt](http://www.scielosp.org/scielo.php?script=sci_abstract&pid=S0213-91112014000200010&lng=en&nrm=iso&tlng=pt)
- Encuesta demográfica y de salud familiar 2012. (2014). *Diseño y cobertura de la muestra - APÉNDICE A*. Descargado de [http://proyectos.inei.gob.pe/endes/2012/apendice\\_A.pdf](http://proyectos.inei.gob.pe/endes/2012/apendice_A.pdf)
- Fienberg, E. S. (2007). *The analysis of cross-classified categorical data* (Cengage Learning Editores ed.).
- Freyre, E. (2004). *Síntomas depresivos en escolares adolescentes de cuatro colegios del distrito de Pichanaki, provincia de Chanchamayo, Junín*. Tesis de Bachiller. Lima, Perú. Universidad Peruana Cayetano Heredia. Descargado de [http://apps.who.int/iris/bitstream/10665/41864/1/0965546608\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/41864/1/0965546608_eng.pdf)
- Garay, J. (2005). *Niveles de depresión, autoestima y estrés en mujeres que tienen un trabajo reenumerado y mujeres que tienen un trabajo no reenumerado (ama de casa)*. Descargado de <http://www.bib.uia.mx/tesis/pdf/>

014584/014584.pdf

- Garibay, R. (2009). *Procesos cognitivos en pacientes con depresión: efecto del tratamiento con antidepresivos*.
- Gómez-Restrepo, C., y otros. (2004). *Prevalencia de depresión y factores asociados con ella en la población colombiana*. Descargado de [www.scielo.org/pdf/rpsp/v16n6/23682.pdf](http://www.scielo.org/pdf/rpsp/v16n6/23682.pdf)
- Hosmer Jr. David W., M. S., Lemeshow Stanley. (2000). *Applied survival analysis : regression modeling of time-to-event data* [djvu]. Descargado de [1506000/41b0866e0c2edb9d5bb6e332deb9cbb5](http://1506000/41b0866e0c2edb9d5bb6e332deb9cbb5)
- Ignasio, A. (2011). *Depresión generalidades y particularidades*. Ciencias Médicas. Descargado de [http://newpsi.bvs-psi.org.br/ebooks2010/pt/Acervo\\_files/depresion-cuba.pdf](http://newpsi.bvs-psi.org.br/ebooks2010/pt/Acervo_files/depresion-cuba.pdf)
- Instituto Nacional de Salud Mental. (2002). *Estudio epidemiológico metropolitano en salud mental 2002*. Informe general. Anales de Salud Mental.
- Instituto Nacional de Salud Mental. (2011). *Las mujeres y la depresión*.
- Instituto nacional de salud mental. (2012). *Anales de salud mental, estudio epidemiológico de salud mental de niños y adolescentes en lima metropolitana y callao 2007*. Descargado de [www.inism.gob.pe/investigacion/archivos/estudios/2007-ASM-EESM-NYA.pdf](http://www.inism.gob.pe/investigacion/archivos/estudios/2007-ASM-EESM-NYA.pdf)
- Instituto Nacional de Salud Mental, Honorio Delgado - Hideyo Noguchi. (2006). *La depresión*. Descargado de [www.inism.gob.pe/investigacion/articulos/1.html](http://www.inism.gob.pe/investigacion/articulos/1.html)
- Kaplan, H., y Sadock, B. (2002). *Synopsis of psychiatry: Behavioral sciences/clinical psychiatry* (Vol. 565) (n.º 566).
- Ministerio de Salud Chile. (2008). *Orientación para la programación en red*. Descargado de [www.terapia-ocupacional.cl/documentos/miscelaneo/2008\\_MINSAL\\_2008.pdf](http://www.terapia-ocupacional.cl/documentos/miscelaneo/2008_MINSAL_2008.pdf)
- Ministerio de Salud de Chile. (2006). *Garantías explícitas en salud. guía clínica: Depresión en personas de 15 años y más*. Descargado de [www.enfermeriaaps.com/portal/?wpfb\\_dl=2151](http://www.enfermeriaaps.com/portal/?wpfb_dl=2151)
- Ministerio de Sanidad y Consumo-Madrid. (2007). *Estrategia en salud mental del sistema nacional de salud*.
- Molinero, L. M. (2004). Historia del razonamiento estadístico. *Asociación de la Sociedad Española de Hipertensión*. Descargado de <http://www.seh-lelha.org/pdf/historiastat.pdf>
- Montenegro, M. B., y Kilstein, J. G. (2000). *Utilidad del phq-9 en el diagnóstico de trastorno depresivo en una sala de clínica médica*. Descargado de <http://www.clinica-unr.com.ar/2015-web/Posgrado/Graduados/Maria-Belen-Montenegro.pdf>
- Montoya, L., y otros. (2007). *Depresión en estudiantes universitarios y su asociación con el estrés académico*. Descargado de <http://www.scielo.org.co/pdf/cesm/v24n1/v24n1a02.pdf>
- Murray, C., y López, A. (1996). *The global burden of disease. a comprehensive assessment of mortality and disability from diseases, injuries, and risk fac-*

- tors in 1990 and projected to 2020*. Descargado de [http://apps.who.int/iris/bitstream/10665/41864/1/0965546608\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/41864/1/0965546608_eng.pdf)
- National institute of mental health. (s.f.). *Depresión que toda mujer debe saber*. Niño-Avenidaño, C., y otros. (2014). *Episodes of anxiety and depression in female university students in Tunja (Colombia): possible association with use of hormonal contraception*. 2012 (Vol. 16) (n.º 29). Investigación andina. Descargado de [www.scielo.org.co/scielo.php?script=sci\\_abstract&pid=S0124-81462014000200006&lng=en&nrm=iso&tlng=es](http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0124-81462014000200006&lng=en&nrm=iso&tlng=es)
- Olmedilla Zafra, E. y. M. G. J., A.\*; Ortega Toro. (2008). *Variables socio-demográficas, ejercicio física, ansiedad y depresión en mujeres: un estudio correlacional*.
- Organización Mundial de la Salud. (2008). *Informe sobre los servicios de Salud Mental del Subsector Ministerio de Salud del Perú 2008*. Descargado de [http://www.minsa.gob.pe/dgsp/archivo/salud\\_mental\\_documentos/09\\_saludmental\\_minsa2008.pdf](http://www.minsa.gob.pe/dgsp/archivo/salud_mental_documentos/09_saludmental_minsa2008.pdf)
- Organización Mundial de la Salud. (2016). *Centro de noticias de la ONU en español - OMS afirma que la depresión es el trastorno mental más frecuente*. Descargado de <http://www.un.org/spanish/News/story.asp?NewsID=24705#.V0dwTJF97IU>
- Quispe Flores, R. (2016). Regresión logística ordinal aplicado al estudio de la gravedad de lesiones por accidente de tránsito en la región madre de dios, 2010–2014.
- Ramírez-Ruiz, L., y otros. (2011). *Factores asociados a la depresión en las grandes urbes. El caso del distrito Federal en México. 2006*. Descargado de [www.scielo.org.co/pdf/rgps/v10n20/v10n20a08.pdf](http://www.scielo.org.co/pdf/rgps/v10n20/v10n20a08.pdf)
- Rohana, K., y otros. (2003). *Cognitive-behavioral therapy, light therapy, and their combination in treating seasonal affective disorder*. Descargado de <http://www.sciencedirect.com/science/article/pii/S0165032703000983>
- Rondón, M. (2006). *Salud mental: un problema de salud pública en el Perú* (Vol. 23) (n.º 4). Revista Peruana de Medicina Experimental y Salud Publica. Descargado de [http://www.scielo.org.pe/scielo.php?script=sci\\_abstract&pid=S1726-46342006000400001&lng=es&nrm=iso&tlng=es](http://www.scielo.org.pe/scielo.php?script=sci_abstract&pid=S1726-46342006000400001&lng=es&nrm=iso&tlng=es)
- Salazar, G. (2004). *Prevalencia de depresión mayor en escolares de 12 a 16 años de edad en tres colegios de la ciudad de Huancayo*. Tesis de Bachiller. Lima, Perú. Universidad Peruana Cayetano Heredia.
- Scott Long, J. (1997). *Regression models for categorical and limited dependent variables* (Sage publications ed.).
- Torija, U., y otros. (2007). *Síntomas depresivos en personas mayores: Prevalencia y factores asociados* (Vol. 21) (n.º 1). Gaceta Sanitaria. Descargado de [http://scielo.isciii.es/scielo.php?script=sci\\_abstract&pid=S0213-91112007000100008&lng=es&nrm=iso&tlng=es](http://scielo.isciii.es/scielo.php?script=sci_abstract&pid=S0213-91112007000100008&lng=es&nrm=iso&tlng=es)
- Torres, M. (s.f.). *Estudio Comparativo en Enfermos con Tuberculosis Pulmonar Frotis Positivo de los grupos Nunca tratados, Antes tratados y Crónicos en*

*sus niveles de Depresión y Soporte Social.*

- Vijay, E. (2012). *Depresión en los adolescentes: Diagnóstico y elaboración de una propuesta desde el modelo integrativo. (Estudio a realizarse con los alumnos del noveno de educación básica de los colegios del Cantón Biblián )*.
- Wackerly, D. D., y otros. (2010). *Estadística matemática con aplicaciones* (Cengage Learning Editores ed.).
- World Health Organization. (2008). *The global burden of disease 2004*.

# ANEXOS

# ANEXO A

## Cuestionario sobre la salud del paciente - 9 (PHQ-9)

En los últimos 14 días, es decir entre el ..... y el día de ayer, usted ha sentido molestias o problemas como:

	Nunca	Varios días	Más de la mitad de los días	Casi todos los días
	0	1	2	3
a. Poco interés o agrado al hacer las cosas.				
b. Se ha sentido triste, deprimido o desesperado.				
c. Ha tenido problemas para dormir, mantenerse despierto o duerme demasiado.				
d. Se siente cansado o tiene poca energía.				
e. Tiene poco o excesivo apetito.				
f. Se ha sentido mal consigo mismo, ha sentido que usted es un fracaso o ha sentido que se ha fallado a sí mismo o a su familia.				
g. Ha tenido problemas para concentrarse en actividades como leer el periódico o ver televisión.				
h. Se mueve o habla tan despacio que otras personas pueden darse cuenta. Está tan inquieto o intranquilo que da vueltas de un lugar a otro más que de costumbre.				
i. Ha pensado que estaría mejor muerto o ha deseado hacerse daño de alguna forma.				

# ANEXO B

## Distribución de frecuencias

TABLA B.1: Distribución de frecuencia de las variables en estudio

Variable	Cod/ categoría	Frecuencia	%
Depresión	0/ No	9187	66.8
	1/ Leve	2676	19.4
	2/ Moderada/severa	1900	13.8
Edad	0/ 15 - 19	1666	12.1
	1/ 20 - 34	7573	55.0
	2/ 35 - 49	4524	32.9
Estado civil	0/ soltera	1376	10.0
	1/ casada	3648	26.5
	2/ conviviente	6973	50.7
	3/ viuda/divorciada	1766	12.8
Nivel de estudios	0/ Hasta primaria	4101	29.8
	1/ Secundaria	6030	43.8
	2/ Superior	3632	26.4
Nivel de estudios de la pareja	0/ Hasta primaria	2842	20.6
	1/ Secundaria	8022	58.3
	2/ Superior	2899	21.1

continua →

Variable	Cod/ categoría	Frecuencia	%
Región natural	0/ Sierra	4755	34.5
	1/ Selva	3388	24.6
	2/ Resto Costa	4084	29.7
	3/ Lima metropolitana	1536	11.2
Lugar de residencia	0/ Campo	4328	31.4
	1/ Pueblo	3840	27.9
	2/ Ciudad pequeña	4059	29.5
	3/ Capital. ciudad grande	1536	11.2
Trabaja actualmente	0/ No	4878	35.4
	1/ Si	8885	64.6
Actualmente está embarazada	0/ No. no está seguro	13248	96.3
	1/ Si	515	3.7
Alfabetismo	0/ No lee/con dificultad	1413	10.3
	1/ Lee facilmente	12350	89.7
índice de riqueza	0/ Muy pobre	3522	25.6
	1/ Pobre	3698	26.9
	2/ Medio	2822	20.5
	3/ Rico	2173	15.8
	4/ Muy rico	1548	11.2
Niños. < 5 años	0/ No	3168	23.0
	1/ Si	10595	77.0
violencia	0/ No	9621	69.9
	1/ Si	4142	30.1
Número de pareja sexual	0/ < 1	7394	53.7
	1/ 2 +	6369	46.3
Anemia	0/ No	10872	79.0
	1/ Si	2891	21.0
Le diagnosticaron diabetes o azúcar Alta	0/ No	13637	99.1
	1/ Si	126	.9
Le diagnosticaron hipertensión arterial o Presión Alta	0/ No	13084	95.1
	1/ Si	679	4.9
En los últimos 12 meses ha fumado Cigarrillos	0/ No	12856	93.4
	1/ Si	907	6.6
Ha consumido bebidas alcohólicas alguna Vez	0/ No	1829	13.3
	1/ Si	11934	86.7



# ANEXO C

## Evaluación del modelo de regresión logística ordinal

### C.1. Evaluación del modelo con todas las variables en estudio

TABLA C.1: Historial de iteraciones

Iteración 0:	log verosimilitud	=	-11857.961
Iteración 1:	log verosimilitud	=	-11272.492
Iteración 2:	log verosimilitud	=	-11262.272
Iteración 3:	log verosimilitud	=	-11262.268
Iteración 4:	log verosimilitud	=	-11262.268

TABLA C.2: Prueba de Omnibus

Chi-cuadrado de razón de verosimilitud	gl	Sig.
1191.39	28	0.000

TABLA C.3: Prueba de efectos del modelo

Origen	Tipo III		
	$\chi^2$ de Wald	gl	Sig.
Edad	29.843	2	.000
Estado civil	43.709	3	.000
Nivel de estudio	14.470	2	.001
Nivel de estudio-pareja	6.817	2	.033
Región	115.040	3	.000
Trabaja	2.456	1	.117
Embarazo	1.221	1	.269
Alfabetismo	.002	1	.963
Riqueza	21.417	4	.000
Niños < 5 años	16.191	1	.000
Violencia	526.648	1	.000
N <sup>0</sup> pareja sexual	34.155	1	.000
Anemia	3.353	1	.067
Diabetes	11.044	1	.001
Hipertensión	54.024	1	.000
Fuma	52.796	1	.000
Alcohol	51.156	1	.000
Miembros de la familia	8.687	1	.003

TABLA C.4: Estimación de parámetros con todas las variables de estudio

Parámetro	$\hat{\beta}$	SE	$\chi^2$	gl	Sig.	$e^{\hat{\beta}}$	IC. 95% $e^{\hat{\beta}}$	
							Inf.	Sup.
Depresión								
Leve	.889	.1351	43.337	1	.000	2.434	1.868	3.172
Moderada/severa	2.105	.1364	238.379	1	0.000	8.211	6.285	10.727
Edad								
20 - 34	-.280	.0614	20.771	1	.000	.756	.670	.853
35 - 49	-.102	.0646	2.502	1	.114	.903	.795	1.025
Estado civil								
Casada	-.447	.0905	24.448	1	.000	.639	.535	.763
Conviviente	-.485	.0879	30.466	1	.000	.616	.518	.731
viuda/divorciada	-.245	.0975	6.325	1	.012	.782	.646	.947
Nivel de estudio								
Secundaria	-.075	.0544	1.918	1	.166	.927	.834	1.032
Superior	-.244	.0684	12.741	1	.000	.784	.685	.896
Nivel de estudio-pareja								
Secundaria	.124	.0537	5.375	1	.020	1.132	1.019	1.258
Superior	.032	.0830	.153	1	.696	1.033	.878	1.215
Región								
Selva	-.444	.0496	80.277	1	0.000	.641	.582	.707
Resto de la costa	-.449	.0501	80.430	1	0.000	.638	.579	.704
Lima metropolitana	-.328	.0694	22.404	1	.000	.720	.629	.825
Trabaja								
Si	.063	.0402	2.456	1	.117	1.065	.984	1.152
Embarazo								
Si	.108	.0977	1.221	1	.269	1.114	.920	1.349
Alfabetismo								
Lee fácilmente	.003	.0710	.002	1	.963	1.003	.873	1.153
Riqueza								
Pobre	.168	.0544	9.530	1	.002	1.183	1.063	1.316
Medio	.224	.0631	12.603	1	.000	1.251	1.105	1.416
Rico	.055	.0736	.555	1	.456	1.056	.915	1.220
Muy rico	.025	.0887	.081	1	.777	1.025	.862	1.220
Niños < 5 años								
Si	-.194	.0481	16.191	1	.000	.824	.750	.905
Violencia								
Si	.930	.0405	526.648	1	0.000	2.535	2.342	2.745
$N^0$ pareja sexual								
2+	.228	.0390	34.155	1	.000	1.256	1.164	1.356

continua →

Parámetro	$\hat{\beta}$	SE	$\chi^2$	gl	Sig.	$e^{\hat{\beta}}$	IC. 95% $e^{\hat{\beta}}$	
							Inf.	Sup.
Anemia								
Si	.082	.0446	3.353	1	.067	1.085	.994	1.184
Diabetes								
Si	.574	.1727	11.044	1	.001	1.775	1.265	2.490
Hipertensión								
Si	.579	.0788	54.024	1	.000	1.785	1.530	2.083
Fuma								
Si	.517	.0711	52.796	1	.000	1.676	1.458	1.927
Alcohol								
Si	.447	.0624	51.156	1	.000	1.563	1.383	1.766
$N^0$ de miembros								
	.035	.0118	8.687	1	.003	1.035	1.012	1.060

### C.1.1. Evaluación del modelo con las variables significativas, para el Factor sociodemográfico

TABLA C.5: Historial de iteraciones del modelo con las variables significativas, para el Factor sociodemográfico

Iteración 0:	log verosimilitud	=	-11857.961
Iteración 1:	log verosimilitud	=	-11275.918
Iteración 2:	log verosimilitud	=	-11265.81
Iteración 3:	log verosimilitud	=	-11265.806
Iteración 4:	log verosimilitud	=	-11265.806

TABLA C.6: Prueba ómnibus del modelo con las variables significativas, para el Factor sociodemográfico

$\chi^2$ de razón de verosimilitud	gl	Sig.
1184.31	24	0.000

TABLA C.7: Pruebas de efectos del modelo con las variables significativas, para el Factor sociodemográfico

Variable	$\chi^2$ de Wald	gl	Sig.
Edad	29.964	2	.000
Estado civil	45.853	3	.000
Nivel de estudio	14.375	2	.001
Nivel de estudio-pareja	6.883	2	.032
Región	116.930	3	0.000
Riqueza	20.556	4	.000
Niños < 5 años	17.097	1	.000
Violencia	529.284	1	0.000
Nº pareja sexual	35.518	1	.000
Diabetes	11.191	1	.001
Hipertensión	53.943	1	.000
Fuma	52.343	1	.000
Alcohol	51.825	1	.000
Miembros de la familia	8.913	1	.003

TABLA C.8: Prueba de líneas paralelas del modelo con las variables significativas, para el Factor sociodemográfico

Modelo	$\chi^2$	Sig.	gl
<b>General</b>	<b>23.81</b>	<b>0.472</b>	24
Edad			
20 - 34	0.00	0.96	1
35 - 49	2.57	0.11	1
Estado civil			
Casada	0.20	0.65	1
Conviviente	0.15	0.70	1
Viuda/divorciada	0.46	0.50	1
Nivel de estudio			
Secundaria	1.11	0.29	1
Superior	0.59	0.44	1
Nivel de estudio-pareja			
Secundaria	2.42	0.12	1
superior	2.77	0.10	1
Región			
Selva	0.44	0.51	1
Resto de la costa	0.48	0.49	1
Lima metropolitana	0.43	0.51	1
Riqueza			
Pobre	1.15	0.28	1
Medio	2.71	0.10	1
Rico	1.60	0.21	1
Muy rico	1.94	0.16	1
Niños < 5 años			
Si	0.04	0.85	1
Violencia			
Si	0.12	0.73	1
$N^0$ pareja sexual			
2+	0.48	0.49	1
Diabetes			
Si	1.39	0.24	1
Hipertensión			
Si	0.08	0.78	1
Fuma			
Si	0.19	0.67	1
Alcohol			
Si	0.33	0.57	1
$N^0$ de miembros	1.54	0.21	1

### C.1.2. Evaluación del modelo con las variables significativas, para el Factor demográfico

TABLA C.9: Historial de iteraciones del modelo para el Factor Demográfico

Iteración 0:	log verosimilitud	=	-11857.961
Iteración 1:	log verosimilitud	=	-11714.344
Iteración 2:	log verosimilitud	=	-11713.448
Iteración 3:	log verosimilitud	=	-11713.448

TABLA C.10: Prueba ómnibus del modelo para el Factor Demográfico

$\chi^2$ de razón de verosimilitud	gl	Sig.
289.026	9	0.000

TABLA C.11: Pruebas de efectos del modelo para el Factor Demográfico

Variable	$\chi^2$ de Wald	gl	Sig.
Edad	34.201	2	.000
Estado civil	120.749	3	.000
Región	102.724	3	0.000
Niños < 5 años	14.558	1	.000

TABLA C.12: Prueba de líneas paralelas del modelo para el Factor Demográfico

Modelo	$\chi^2$	Sig.	gl
<b>General</b>	<b>11.36</b>	<b>0.252</b>	9
Edad			
20 - 34	0.04	0.842	1
35 - 49	4.25	0.039	1
Estado civil			
Casada	0.05	0.828	1
Conviviente	0.00	0.986	1
Viuda/divorciada	0.19	0.666	1
Región			
Selva	0.24	0.625	1
Resto de la costa	0.00	0.991	1
Lima metropolitana	0.10	0.749	1
Niños < 5 años			
Si	0.24	0.622	1

### C.1.3. Evaluación del modelo, para el Factor social

TABLA C.13: Historial de iteraciones del modelo para el Factor Social

Iteración 0:	log verosimilitud	=	-11857.961
Iteración 1:	log verosimilitud	=	-11398.562
Iteración 2:	log verosimilitud	=	-11391.488
Iteración 3:	log verosimilitud	=	-11391.487
Iteración 4:	log verosimilitud	=	-11391.487

TABLA C.14: Prueba ómnibus del modelo para el Factor Social

$\chi^2$ de razón de verosimilitud	gl	Sig.
932.948	11	0.000

TABLA C.15: Pruebas de efectos del modelo para el Factor Social

Variable	$\chi^2$ de Wald	gl	Sig.
Nivel de estudio	44.056	2	.000
Nivel de estudio-pareja	22.698	2	.000
trabaja actualmente	12.838	1	.000
violencia	566.951	1	.000
Nº de parejas sexuales	26.684	1	.000
Diabetes	12.149	1	.000
Hipertensión	60.661	1	.000
Fuma	50.509	1	.000
Alcohol	33.513	1	.000



TABLA C.16: Prueba de líneas paralelas

Modelo	$\chi^2$	Sig.	gl
<b>General</b>	<b>11.19</b>	<b>0.428</b>	11
Nivel de estudio			
Secundaria	3.92	0.048	1
Superior	0.00	0.976	1
Nivel de estudio-pareja			
Secundaria	1.10	0.295	1
superior	0.79	0.374	1
¿Trabaja?			
Si	0.05	0.830	1
Violencia			
Si	0.11	0.746	1
N <sup>0</sup> pareja sexual			
2+	0.94	0.332	1
Diabetes			
Si	1.19	0.276	1
Hipertensión			
Si	0.46	0.496	1
Fuma			
Si	0.10	0.754	1
Alcohol			
Si	0.62	0.433	1

# ANEXO D

## REGRESIÓN LOGÍSTICA

### D.1. Breve reseña histórica de Regresión logística

La regresión logística fue desarrollada como modelo de crecimiento de una población y fue nombrada como "logística" por Pierre François Verhulst entre los años 1830 y 1840, juntamente con Adolphe Quetelet. Verhulst al desarrollar su primer artículo en el año 1838 no mencionó del cómo realizó el ajuste de las curvas a los datos. Posteriormente en su artículo que presentó en el año 1845, Verhulst detalló el desarrollo de tres parámetros del modelo haciendo que la curva pasara por tres puntos observados, pero las predicciones fueron muy malas.

La regresión logística se desarrolló independientemente en el área de química como modelo de autocatálisis (Wilhelm Ostwald, 1883). Una reacción autocatalítica es aquella en la que uno de los productos es en sí mismo un catalizador para la misma reacción, mientras que el suministro de uno de los reactivos es fijo. Esto se presenta como una regresión logística.

La regresión logística fue redescubierta independientemente como un modelo de crecimiento de la población en el año 1920 por Raymond Pearl y Lowell Reed, usando en las estadísticas modernas. Al principio desconocían el trabajo desarrollado por Verhulst, posteriormente, llegaron a enterarse por L. Gustave du Pasquier, pero no tuvo mucha relevancia para ellos y no optaron su terminología. Pearl y Reed inicialmente aplicaron el modelo a la población en los Estados Uni-

dos, y desarrollaron el ajuste a la curva al hacerla pasar por tres puntos; Al igual que Verhulst, nuevamente estos resultados fueron de fracaso. La iniciativa de Verhulst fue reconocida y el término “logístico” fue revivido por Udny Yule en 1925 y fue seguido desde entonces.

En la década 1930, Chester Ittner Bliss desarrolló y sistematizó el modelo probit, y fue quien insertó el término “probit”, en 1934, y Ronald A. Fisher desarrolla el modelo de ajuste a la estimación de máxima probabilidad en 1935, como un aporte al trabajo de Bliss. El modelo probit se usó principalmente en bioensayos, que ya había sido precedido por un trabajo anterior que data de 1860. Posteriormente el modelo probit influyó en el desarrollo del modelo logit y estos modelos compitieron entre sí.

Probablemente en 1943, Edwin Bidwell Wilson y su estudiante Jane Worcester hayan utilizado el modelo logístico por primera vez como una alternativa al modelo probit en bioensayo. Sin embargo, el desarrollo del modelo logístico como alternativa general al modelo probit se debió principalmente al trabajo de Joseph Berkson durante muchas décadas (1944), donde insertó “logit”, por analogía con “probit”, y dando continuidad por Berkson en 1951 y años posteriores. Inicialmente el modelo logit fue descartado por ser inferior al modelo probit, pero logró alcanzar gradualmente un equilibrio con el probit entre 1960 y 1970. En 1970, el modelo logit alcanzó la paridad con el modelo probit llegando a publicar en revistas de estadística, llegando a sobrepasar al modelo probit. Esta relativa popularidad se debió a la adopción del logit fuera del bioensayo, en lugar de desplazar el probit dentro del bioensayo y su uso informal en la práctica; La popularidad del logit se atribuye a la simplicidad computacional, las propiedades matemáticas y la generalidad del modelo logit, lo que permite su uso en diversos campos. Durante ese tiempo ocurrieron varios refinamientos, notablemente por David Cox, el modelo logit multinomial se introdujo de forma independiente por Cox en el año 1970, y podemos señalar a David R. Cox como el principal difusor de la regresión logística publicada en su libro *The Analysis of Binary Data*.

lo que aumentó considerablemente el alcance de la aplicación y la popularidad del modelo logit. En 1973, Daniel McFadden vinculó el logit multinomial con la teoría de la elección discreta, específicamente el axioma de elección de Luce, mostrando que el logit multinomial seguía el supuesto de independencia de alternativas irrelevantes e interpretaba las probabilidades de alternativas como preferencias relativas; esto dio una base teórica para la regresión logística.

En los últimos años, con respecto a la regresión el avance más importante fue la introducción de los modelos lineales generalizados (MLG). En 1972 Los estadísticos británicos R.W.M. Wedderburn y John Nelder, unifican todas las teorías existentes, los modelos logísticos y modelos probit, con los modelos lineales con distribución normal, como también con el análisis de la varianza. El principal algoritmo para realizar ajustes a estos modelos se les conoce como “Fisher scoring”, ya que fue Fisher quien introdujo este ajuste a los modelos probit de máxima verosimilitud (Molinero, 2004).

## D.2. Regresión logística binaria

La regresión logística es un método de análisis de datos empleado para la descripción de una relación entre una variable respuesta categórica binaria y un conjunto de variables explicativas continuas y/o categóricas.

La variable repuesta  $Y$  es categórica dicotómica, cuya descripción se realiza mediante la “proporción” de casos que pertenecen a una categoría.

Sea  $Y = 1$  la categoría de interés, entonces la proporción de casos favorables será  $\pi$  donde  $0 \leq \pi \leq 1$  correspondiente a esta categoría, y es expresado por:

$$\pi = \frac{N_1}{N} \tag{D.1}$$

$N_1$  es el número casos interés ( $Y = 1$ ),  $N_2$  las de no interés ( $Y = 0$ ) y  $N = N_1 + N_2$  es el tamaño poblacional.

El resultado de  $\pi$ , interpretamos como “la frecuencia relativa de ocurrencia de la característica de interés”.

### D.2.1. Modelo de regresión logística binaria

Dada una variable respuesta dicotómica con distribución Bernoulli y media  $\pi$ , la esperanza condicional se expresará de la siguiente manera:

$$E(Y/x) = \pi(x) = \beta_0 + \beta_1 x \quad (\text{D.2})$$

Donde el vector de parámetros  $(\beta_0, \beta_1) \in \mathbb{R}^2$ , entonces el predictor lineal  $\beta_0 + \beta_1 x$  toma cualquier valor real. A su vez,  $\pi$  únicamente toma valores entre  $[0, 1]$ ; En efecto la relación (D.2) no tenía sentido. por consiguiente, se llegó a considerar la transformación  $\ln \left[ \frac{\pi(x)}{1-\pi(x)} \right] \in \mathbb{R}$ , en efecto  $\frac{\pi(x)}{1-\pi(x)}$  toma los valores del intervalo  $(0, \infty)$ . Entonces, el modelo de regresión para respuesta binaria se expresará de la siguiente manera:

$$\ln \left[ \frac{\pi(x)}{1-\pi(x)} \right] = \beta_0 + \beta_1 x \quad (\text{D.3})$$

La transformación  $\ln \left[ \frac{\pi(x)}{1-\pi(x)} \right]$  se denomina transformación logit y se denota por

$$g(x) = \text{logit}(\pi(x)) = \ln \left[ \frac{\pi(x)}{1-\pi(x)} \right]$$

Así, a partir de la ecuación (D.3), el modelo de regresión logística se expresa como:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (\text{D.4})$$

Una vez codificada la variable explicativa  $X$ , con 1 y 0. asumiendo la ecuación (D.3), el logit para  $X = 0$ , será:

$$\ln \left[ \frac{\pi_0}{1-\pi_0} \right] = \beta_0 + \beta_1(0) = \beta_0 \quad (\text{D.5})$$

Donde  $\pi_0$  es la probabilidad de ocurrencia cuya característica es de interés ( $Y = 1$ ) cuando  $X = 0$ . Entonces  $\pi_0/1-\pi_0$  se denomina “odds” (ventaja o chance de que ocurra la característica de interés), una vez que  $X = 0$ . Por lo tanto,  $\beta_0$  es el

logaritmo del “odds”.

Además, si  $X = 1$ , el logit obtenido en la ecuación (D.3) será:

$$\ln \left[ \frac{\pi}{1 - \pi} \right] = \beta_0 + \beta_1(1) = \beta_0 + \beta_1 \quad (\text{D.6})$$

Una vez más  $\pi$  es la probabilidad que ocurra la característica de interés ( $Y = 1$ ) si la variable explicativa toma el valor  $X = 1$ . Después, viene hacer la ventaja o chance de ocurrencia de la característica de interés, si  $X = 1$ .

Restando la ecuación (D.5) de (D.6), y haciendo algunas operaciones tenemos

$$\ln \left[ \frac{\pi_1}{1 - \pi_1} \right] - \ln \left[ \frac{\pi_0}{1 - \pi_0} \right] = \ln \left[ \frac{\frac{\pi_1}{1 - \pi_1}}{\frac{\pi_0}{1 - \pi_0}} \right] = \ln[OR] = \beta_1 \quad (\text{D.7})$$

Luego,  $\beta_1$ , es el logaritmo del cociente de “odds ratio”, tal cual se observan, los parámetros de este modelo no se pueden interpretar de manera inmediata; pero, a partir de la ecuación (D.7) se calculan los “odds ratio” ( $OR$ ) de la siguiente manera

$$OR = e^{\beta_1} \quad (\text{D.8})$$

El valor calculado, es una medida de asociación entre las variables, que en este caso si podemos interpretar de manera inmediata.

Si  $\beta_1 = 0$ ; a partir de (D.4) se puede decir que el valor esperado de  $Y$ , dado un valor de  $X = x$ , es constante  $E(Y/x) = \pi(x) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$ . Es decir, si  $\beta_1 = 0$ , entonces decimos que la variable explicativa  $Y$ , y la variable respuesta  $X$  son independientes.

En el caso de que se reemplace  $\beta_1 = 0$  en (D.8), y el valor de odds ratio sea 1 ( $OR = e^0 = 1$ ), entonces diremos que la variable respuesta  $Y$  y la variable explicativa  $X$  son independientes.

### D.2.2. Estimación del modelo

En modelos de regresión logística simple, en otras palabras, el modelo tiene una sola variable explicativa, entonces  $\beta_0$  y  $\beta_1$  son los parámetros desconocidos y son estimados mediante el método de máxima verosimilitud, este método consiste en ajustar la estimación que proporcione máxima verosimilitud o probabilidad a los datos observados.

De esta manera en la regresión logística binaria simple, y suponemos que las observaciones son independientes, entonces la función de verosimilitud es:

$$L(\beta_0, \beta_1) = \prod_{i=1}^n p_j^{y_i} (1 - p_j)^{1-y_i} \quad (\text{D.9})$$

Donde:  $y_1, \dots, y_n$  son observaciones de Y  $y_j \in \{0, 1\}$ ,  $p_j = \frac{\exp(\beta_0 + \beta_1 x_j)}{1 + \exp(\beta_0 + \beta_1 x_j)}$ ,  $j = 1, \dots, n$  y  $x_1, \dots, x_n$  son observaciones de X.

Para el ajuste de los parámetros se necesita maximizar la función de verosimilitud o lo que es maximizar su logaritmo.

$$\text{Log}(L(\beta_0, \beta_1)) = \sum_{j=1}^n (y_j \log(p_j) + (1 - y_j) \log(1 - p_j)) \quad (\text{D.10})$$

Realizando los cálculos de derivada respecto a cada parámetro  $(\beta_0, \beta_1)$  y luego igualando a cero se obtiene las ecuaciones de verosimilitud que se muestran a continuación:

$$\sum_{j=1}^n (y_j - p_j) = 0$$

y

$$\sum_{j=1}^n x_j (y_j - p_j) = 0$$

Los parámetros estimados de un modelo logit por la máxima verosimilitud son únicos y siempre existirán a excepción de ciertos casos de separación completa esto puede darse debido a la concavidad de la log-verosimilitud. Las ecuaciones que se obtuvieron no son lineales en los parámetros, entonces se requiere métodos

iterativos tal como el de Newton-Raphson para su resolución. A continuación presentaremos la fórmula iterativa de resolución de las ecuaciones de verosimilitud:

$$\begin{pmatrix} \beta_0^t \\ \beta_1^t \end{pmatrix} = \begin{pmatrix} \beta_0^{(t-1)} \\ \beta_1^{(t-1)} \end{pmatrix} + \begin{pmatrix} \sum_{j=1}^n p_j^{(t-1)}(1-p_j)^{(t-1)} & \sum_{j=1}^n x_j p_j^{(t-1)}(1-p_j)^{(t-1)} \\ \sum_{j=1}^n x_j p_j^{(t-1)}(1-p_j)^{(t-1)} & \sum_{j=1}^n x_j^2 p_j^{(t-1)}(1-p_j)^{(t-1)} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n (y_j - p_j)^{(t-1)} \\ \sum_{j=1}^n x_j (y_j - p_j)^{(t-1)} \end{pmatrix}$$

Donde:  $p_j^{(t-1)}$  es la probabilidad estimada en la iteración  $t - 1$  originado desde los parámetros estimados dadas en la iteración  $t - 1$  de la siguiente forma

$$p_j^{(t-1)} = \frac{\exp(\beta_0^{(t-1)} + \beta_1^{(t-1)} x_j)}{1 + \exp(\beta_0^{(t-1)} + \beta_1^{(t-1)} x_j)}, \quad j = 1, \dots, n \quad (\text{D.11})$$

Una vez desarrollado todo el proceso la máxima verosimilitud de  $P_j$  esta dado por:

$$\hat{p}_j = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_1)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1)}, \quad j = 1, \dots, n$$

Donde:  $\hat{\beta}_0$  y  $\hat{\beta}_1$  son los estimadores de máxima verosimilitud de los parámetros.

### D.2.3. Evaluación de la bondad de ajuste

Para la evaluación de los modelos lineales generalizados, generalmente la bondad de ajuste que se usa es la desviación. Este estadístico es el cociente de la verosimilitud del modelo propuesto y el modelo saturado (llamamos modelo saturado, al modelo que tiene tantos parámetros como observaciones). la hipótesis estadística esta dada por:

$H_0$  : El modelo ajustado es significativo (No existe diferencia entre los valores observados y los valores pronosticados).

$H_1$  : El modelo ajustado no es significativo (Existe diferencia entre los valores observados y los valores pronosticados).



donde la desviación es:

$$D = -2\ln \left[ \frac{\text{verosimilitud del modelo ajustado}}{\text{verosimilitud del modelo saturado}} \right] \quad (\text{D.12})$$

Para la evaluación de la significancia de la variable explicativa, se realiza la comparación de la desviación ( $D$ ), con y sin la variable independiente en el modelo. Mientras que la desviación debido a la inserción de la variable explicativa en el modelo, es:

$$G = D(\text{modelo sin la variable}) - D(\text{modelo con la variable}) \quad (\text{D.13})$$

Ya que la verosimilitud del modelo saturado es clásico en el argumento de las desviaciones para los modelos con y sin la variable explicativa, cuya ecuación equivalente al cambio de desviación (D.13), es

$$G = -2\ln \left[ \frac{\text{Verosimilitud sin la variable}}{\text{Verosimilitud con la variables}} \right] \quad (\text{D.14})$$

La estadística Wald, también nos ayuda a evaluar la significancia de las variables explicativas, a continuación se menciona la hipótesis estadística:

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

La estadística Wald está dada por:

$$W = \left( \frac{\hat{\beta}_1}{\hat{SE}(\hat{\beta}_1)} \right)^2 \quad (\text{D.15})$$

Donde:  $\hat{\beta}_1$  viene hacer el estimador de máximo verosímil de  $\beta_1$ , cuyo error estándar estimado es  $SE(\hat{\beta}_1)$ . el argumento general se presenta en la ecuación (??). El estadístico dado en (D.15), bajo la hipótesis  $\beta_1 = 0$ , tiene una distribución  $\chi^2$  con una unidad de grado de libertad.

#### D.2.4. Supuestos del modelo

Los supuestos básicos que debe cumplir este modelo son los siguientes:

- a. No linealidad de la variable de respuesta: En el modelo de regresión logística la variable respuesta es categórica por ello es no lineal.
- b. Independencia de los errores: No existe correlación entre el error y las variables independientes. El valor esperado del error es cero.
- c. Multicolinealidad: Las variables regresoras no deben tener una alta correlación. El incumplimiento de este supuesto tiene graves consecuencias. Existen dos alternativas cuando la multicolinealidad es alta:

- Regresión sesgada (“ridge regresión”), intenta estabilizar los parámetros manipulando las varianzas.
- Regresión por componentes principales, que se basa en la alta correlación entre predictores para definir variados que son combinaciones lineales de los predictores y emplear los variados como nuevos predictores del criterio.

### D.3. Modelo de regresión logística múltiple

Supongamos que el vector de variables regresoras es  $X^T = (x_1, x_2, \dots, x_p)$ . entonces el logaritmo del Modelo de Regresión Logística Múltiple esta dado por:

$$g(X) = \ln \left( \frac{\pi(X)}{1 - \pi(X)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (\text{D.16})$$

Por tanto, el Modelo de Regresión Logística Múltiple es:

$$\pi(X) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}} \quad (\text{D.17})$$

Donde:  $\pi(X) = P(Y = 1/X)$  viene hacer la probabilidad condicional para que la variable dependiente sea 1 ( $Y = 1$ ), dado  $X_1 = x_1, X_2 = x_2, \dots, X_p = x_p$

Es necesario incorporar variables artificiales, o también llamadas dummy; es decir hacer la transformación de las categorías de la variable respuesta en 0 y 1,

así entonces el predictor con  $k$  variables regresoras y una variable categórica se expresa como:

$$\eta = \beta_0 + \beta_1 x_1 + \dots + \sum_{l=1}^{k_j-1} \beta_{jl} D_{jl} + \dots + \beta_p x_p \quad (\text{D.18})$$

Donde  $D_{jl}$ ,  $l = 1, \dots, k_j - 1$  son variables dummy.

**A.1.5.1 Ajuste del modelo** Sea  $n$  una muestra de observaciones independientes  $(x_i, y_i)$ , donde  $i = 1, \dots, n$ . La obtención de los estimadores del vector  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$  es el de máxima verosimilitud, cuyos estimadores de los parámetros son  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ . El modelo de regresión logística múltiple ajustado es:

$$\hat{\pi}(x_i) = \frac{e^{\hat{\eta}}}{1 + e^{\hat{\eta}}} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}}} \quad (\text{D.19})$$

Para la estimación de varianzas y covarianzas de los coeficientes estimados se utiliza la teoría de estimación de máxima verosimilitud.

# ANEXO E

## Otros Modelos Ordinales

### E.1. Modelo de categorías adyacentes

Supongamos que necesitamos comparar cada variable con la respuesta más pequeña, esto es  $p_i(x)$  con  $p_{i+1}(x)$ . El logaritmo de la ventaja de  $p_i(x)$  frente a  $p_{i+1}(x)$  será:

$$\ln \left[ \frac{P(Y = Y_i / X = x)}{P(Y = Y_{i+1} / X = x)} \right] = \ln \left[ \frac{p_i(x)}{p_{i+1}(x)} \right] \quad \forall i = 1, \dots, j - 1$$

Se caracterizan porque el *logit* es el logaritmo de la razón entre dos categorías adyacentes:

$$L_i(x) = \ln \left[ \frac{p_i(x)}{p_{i+1}(x)} \right], \quad \forall i = 1, \dots, j - 1 \quad (\text{E.1})$$

donde  $i$  hace referencia a la categoría y  $p_i(x)$  la probabilidad de que la respuesta tome el valor de la categoría  $i$  dada la observación  $X = x$ . Ante la respuesta dada  $Y_i$  para una variable explicativa, el modelo se formula como:

$$L_i(x) = \beta_{0i} + \beta_{1i}x; \quad i = 1, \dots, j$$

Por lo que si

$$\begin{aligned} L_i(x) = \ln \left[ \frac{p_i(x)}{p_j(x)} \right] &= \ln \left[ \frac{p_i(x)}{p_{i+1}(x)} \right] + \ln \left[ \frac{p_{i+1}(x)}{p_{i+2}(x)} \right] + \dots + \ln \left[ \frac{p_{j-1}(x)}{p_j(x)} \right] \\ &= \sum_{l=i}^{j-1} L_l(x) \end{aligned} \quad (\text{E.2})$$

$$L_i(x) = \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) = \sum_{l=i}^{j-1} \beta_{0l} + \sum_{l=i}^{j-1} \beta_{1l}x \quad (\text{E.3})$$

Por tanto, de las ecuaciones E.2 y E.3 se deduce que:

$$\ln \left[ \frac{p_i(x)}{p_j(x)} \right] = \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x)$$

Tomando exponenciales tenemos que:

$$\frac{p_i(x)}{p_j(x)} = \exp \left( \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) \right) \quad (\text{E.4})$$

Operando de la siguiente manera, obtenemos cada una de las probabilidades de respuesta:

$$\begin{aligned} p_i(x) &= p_j(x) \exp \left( \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) \right) = \frac{p_j(x)}{1} \exp \left( \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) \right) \\ &= \frac{p_j(x)}{\sum_{m=1}^j p_m(x)} \exp \left( \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) \right) \\ &= \frac{p_j(x)}{p_j(x) + \sum_{m=1}^{j-1} p_m(x)} \exp \left( \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) \right) \\ &= \frac{1}{1 + \sum_{m=1}^{j-1} \frac{p_m(x)}{p_j(x)}} \exp \left( \sum_{l=i}^{j-1} \beta_{0l} + \beta_{1l}x \right) \\ p_i(x) &= \frac{\exp \left( \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) \right)}{1 + \sum_{m=1}^{j-1} \exp \left( \sum_{l=i}^{j-1} (\beta_{0l} + \beta_{1l}x) \right)} \quad (\text{E.5}) \end{aligned}$$

Los cocientes de ventajas que tienen un cambio de una unidad en la variable independiente provocan un cambio en la variable respuesta de la siguiente forma:

$$\theta_i = \frac{\frac{p_i(x+1)}{p_{i+1}(x+1)}}{\frac{p_s(x)}{p_{s+1}(x)}} = \frac{\exp(\beta_{0i} + \beta_{1i}(x+1))}{\exp(\beta_{0i} + \beta_{1i}x)} = \exp(\beta_{1i}), \quad \forall i = 1, \dots, j-1$$

De esta manera el parámetro  $\beta_i$  se interpreta como el cociente de ventajas de la categoría  $Y_i$  frente a la categoría adyacente  $Y_{i+1}$  cuando la variable explicativa incrementa en una unidad.

Cuando en lugar de existir una variable, el modelo contiene  $p$  variables explicativas, el modelo se representa de la siguiente forma:

$$L_i(x) = \sum_{m=0}^p \beta_{ri} x_r = x' \beta_j$$

donde el vector de observaciones es  $x = (x_0, x_1, \dots, x_p)'$ ,  $\beta_i = (\beta_{0i}, \beta_{1i}, \dots, \beta_{pi})$  y por simplificar la notación  $x_0 = 1$ .

En este caso la interpretación de los parámetros resulta análoga. Para ello incrementemos la variable de la categoría  $k$  en una unidad mientras que el resto permanecen controladas:

$$\begin{aligned} \theta_s(\Delta X_k = 1 | X_m = x_m, m \neq k) &= \frac{P(Y = Y_i | X_k = x_k + 1; X_m = x_m, m \neq k)}{P(Y = Y_{i+1} | X_k = x_k + 1; X_m = x_m, s \neq k)} \\ &= \frac{P(Y = Y_i | X_k = x_k; X_m = x_m, m \neq k)}{P(Y = Y_{i+1} | X_k = x_k; X_m = x_m, m \neq k)} \\ &= \frac{\exp(\beta_{0i} + \beta_{1i}x_1 + \dots + \beta_{ki}(x_k + 1) + \dots + \beta_{pi})}{\exp(\beta_{0i} + \beta_{1i}x_1 + \dots + \beta_{ki}x_k + \dots + \beta_{pi})} \\ &= \frac{\exp(\beta_{ki}x_k) \cdot \exp(\beta_{ki})}{\exp(\beta_{ki}x_k)} \\ &= \exp(\beta_{ki}) \end{aligned}$$

Por tanto la exponencial de  $\beta_{ki}$  resulta ser el cociente de ventajas de las categorías  $Y_i$  frente a la categoría  $Y_{i+1}$  al haber incrementado una unidad la variable  $x_k$  mientras las demás permanecen iguales.

## E.2. Modelo razón de continuidad

El presente modelo nos permite realizar la comparación de la probabilidad de que la respuesta sea igual a la categoría  $m$ , frente a la probabilidad de que la respuesta sea mayor a  $m$  (Fienberg, 2007).

Este modelo presenta varias pendientes como interceptos para la cada contraste Cuyo modelo se presenta a continuación:

$$\begin{aligned} \text{logit}[\pi(m, k)] &= \ln \left[ \frac{P(Y = m/X_1 = x_1, \dots, X_p = x_p)}{P(Y > m/X_1 = x_1, \dots, X_p = x_p)} \right] \\ &= \alpha_m + (\beta_{m1}x_1 + \dots + \beta_{mp}x_p); \quad m = 1, \dots, k \end{aligned} \quad (\text{E.6})$$

La inconveniencia del uso de este modelo es que, a medida que la variable respuesta asuma valores superiores, se pierde datos para los cálculos de las estimaciones. Si se tiene interés solo en una categoría en específico de la variable respuesta entonces es muy apropiado el uso de este modelo.

## E.3. El modelo estereotipo

Este modelo es similar al modelo logit para una categoría base o inicial cuya variable dependiente es asumida como nominal y no como ordinal. El modelo esta dado por

$$\begin{aligned} \text{logit}[\pi(m, k)] &= \ln \left[ \frac{\pi(Y = m/X_1 = x_1, \dots, X_p = x_p)}{\pi(Y = k/X_1 = x_1, \dots, X_p = x_p)} \right] \\ &= \alpha_m - \phi_m(\beta_1x_1 + \dots + \beta_px_p), \quad m = 1, 2, \dots, k - 1 \end{aligned} \quad (\text{E.7})$$

Siendo,  $m = 1, 2, \dots, k - 1$ ,  $k$  viene hacer la categoría de referencia, por ahora tomaremos a la última categoría,  $Y$  es la variable dependiente ordinal con  $k$  categorías,  $\beta_1, \beta_2, \dots, \beta_p$  coeficientes logit para las variables predictoras de los  $X_1, X_2, \dots, X_p$  respectivamente,  $\alpha_m$  los interceptos y  $\phi_m$  limitaciones utilizadas para aseverar la identificabilidad del modelo, la condición que deben cumplir los

coeficientes son.

$$1 = \phi_1 > \phi_2 \dots > \phi_k = 0 \quad (\text{E.8})$$

La primera limitación es cuando  $\phi_1$  es considerado como uno, y la última  $\phi_k$  como cero.

Si algún par de limitaciones son iguales, entonces las categorías son indistinguibles, y tienen la posibilidad unirse en solo una. En otras palabras, si  $\phi_2 = \phi_3$ , entonces las categorías 2 y 3 tienen la posibilidad de juntarse. La ordinalidad de las limitaciones en el modelo puede ser evaluado, tal es así que se puede juntar o reordenar si se diera el caso. El cambio de una unidad en la variable explicativa  $X_j, j = 1, 2, \dots, p$ , del odd ratio de la categoría  $Y = t$  frente a la categoría  $Y = t$  es:

$$OR = \exp(-[\phi_s - \phi_t]\beta_j) \quad (\text{E.9})$$

Supongamos que la categoría de referencia es  $Y = t$ , en otras palabras  $t = k$ , entonces por la limitación en (E.8),  $\phi_t = 0$ ; posteriormente reemplazando este resultado en (E.9), el “odds ratio” de la categoría  $Y = s$  frente a la categoría de referencia  $Y = k$  para el cambio en una unidad en la variable explicativa  $X_j, j = 1, 2, \dots, p$ , es:

$$OR = \exp(-\phi_s\beta_j) \quad (\text{E.10})$$

## **E.4. Modelo de odds proporcionales parciales**

### **E.4.1. Modelo de odds proporcionales parciales no restringido**

Si la variable dependiente  $Y$  es ordinal con categorías de  $1, \dots, k$ , entonces las probabilidades acumuladas podemos definir de la siguiente manera

$$C_{im} = P(Y > m / X_1 = x_{i1}, \dots, X_p = x_{ip}), \quad i = 1, \dots, n; \quad m = 1, \dots, k - 1$$



Este último nos permite definir el modelo logístico como

$$\ln \left( \frac{C_{im}}{1 - C_{im}} \right) = \alpha_m + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n; \quad m = 1, \dots, k - 1 \quad (\text{E.11})$$

Donde  $\alpha_1 > \alpha_2 > \dots > \alpha_{k-1}$  vienen hacer los interceptos del modelo,  $m$  señala los  $k - 1$  posibles probabilidades acumuladas que se obtienen para dicotomizar la variable dependiente  $Y$  utilizando  $k - 1$  puntos de corte.  $\beta_j (j = 1, \dots, p)$  para la  $j - \text{ésima}$  variable independiente  $X_j$ , es el logaritmo del  $OR$  que se calcula la asociación de  $Y$  y  $X_j$  pudiendo controlar las variables independientes restantes. Ya que  $\beta_j$  no depende de  $m$ , el modelo supone que la asociación de la variable dependiente dicotomizada  $Y$  y  $X_j$ , no está sujeto al punto  $m$  del que se tuvo el corte.

Sea  $n$  una muestra aleatoria de observaciones independientes  $(y_i, x_{i1}, x_{i2}, \dots, x_{ip}; \quad i = 1, \dots, n)$ , siendo  $y_i$  valorado con escala ordinal de  $k$  categorías. La variable dependiente posee una distribución multinomial. Entonces el modelo de odds proporcionales parciales no restringido esta definido por::

$$C_{im} = P(Y > m / X_1 = x_{i1}, \dots, X_p = x_{ip}) = \frac{1}{1 + \exp[-\alpha_m - (\beta_1 + \gamma_{m1})x_{i1} - \dots - (\beta_q + \gamma_{mq})x_{iq} - \beta_{q+1}x_{i,q+1} - \dots - \beta_p x_{ip}]} \quad (\text{E.12})$$

de manera equivalente se tiene el enunciado:

$$C_{im} = P(Y > m / X_1 = x_{i1}, \dots, X_p = x_{ip}) = \frac{\exp[\alpha_m + (\beta_1 + \gamma_{m1})x_{i1} + \dots + (\beta_q + \gamma_{mq})x_{iq} + \beta_{q+1}x_{i,q+1} + \dots + \beta_p x_{ip}]}{1 + \exp[\alpha_m + (\beta_1 + \gamma_{m1})x_{i1} + \dots + (\beta_q + \gamma_{mq})x_{iq} + \beta_{q+1}x_{i,q+1} + \dots + \beta_p x_{ip}]} \quad (\text{E.13})$$

donde  $m = 1, \dots, k - 1$  y  $q \leq p$ .

### E.4.2. Modelo de odds proporcionales parciales restringido

El modelo de odds proporcionales parciales mencionado en (E.13), asume  $k - 2$  parámetros que tiene la forma  $(\beta_j + \gamma_{mj})$ ;  $m = 1, \dots, k - 2$  para interpretar los coeficientes de una variable independiente  $X_j$  que no satisface el supuesto de proporcionalidad de los odds. No obstante, si existiera una relación lineal entre el logaritmo de los  $OR$  y los valores  $m$ ; un modelo que limita el parámetro  $\lambda_{mj}$  a ser lineal en  $m$ , requeriría únicamente un parámetro más.

Si lo mencionado fuera válido para todas las variables independientes que no cumplen el supuesto de proporcionalidad de odds, entonces el modelo propuesto en (E.13) se puede escribir como:

$$C_{im} = P(Y_i > m / X_1 = x_{i1}, \dots, X_p = x_{ip}) = \frac{\exp\{\tau_m + (\beta_1 + \alpha_m \gamma_1)x_{i1} + \dots + (\beta_q + \alpha_m \gamma_q)x_{iq} + \beta_{q+1}x_{i,q+1} + \dots + \beta_p x_{ip}\}}{1 + \exp\{\tau_m + (\beta_1 + \alpha_m \gamma_1)x_{i1} + \dots + (\beta_q + \alpha_m \gamma_q)x_{iq} + \beta_{q+1}x_{i,q+1} + \dots + \beta_p x_{ip}\}} \quad (E.14)$$

De esta manera el modelo enunciado en (E.14), se llama modelo de odds proporcionales parciales restringido, donde los escalares  $\alpha_m$ , con  $\alpha_1 = 0$  son pre enunciado y se les conoce como restricciones. Este modelo, es adecuado si las variables independientes requieren limitaciones idénticas.

De manera específica, si  $\alpha_m = m - 1$ , el logaritmo del  $OR$  en la primera ecuación del modelo viene hacer  $\beta_j$ , en el segundo  $\beta_j + \gamma_j$ , en el tercero  $\beta_j + 2\gamma_j$ , así sucesivamente hasta  $\beta_j + (k - 2)\gamma_j$  en la última ecuación.

Las restricciones también tienen la posibilidad de conseguir relaciones no lineales entre los logaritmos de los  $OR$ .

El test de significancia de los coeficientes estimados de cada variable independiente, tanto para el modelo de odds proporcionales parciales no restringido como restringido, se tiene la posibilidad de realizar por la prueba de Wald. Mientras que la significancia del modelo, se evalúa por la prueba de razón de verosimilitud.