

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS

DEPARTAMENTO DE POSTGRADO

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

**“MAGÍSTER EN ESTADÍSTICA CON MENCIÓN EN GESTIÓN DE LA
CALIDAD Y PRODUCTIVIDAD”**

TEMA:

***MODELOS LINEALES GENERALIZADOS EN EL ANÁLISIS DE LAS
DEFUNCIONES FETALES DE LITORAL ECUATORIANO***

AUTOR:

ALEXANDRA MARIBEL ARGUELLO PAZMIÑO

Guayaquil – Ecuador

2020

RESUMEN

El presente proyecto trata de “Modelos lineales generalizados en el análisis de las defunciones fetales de litoral Ecuatoriano”, el cual contempla un estudio tanto descriptivo como analítico de la situación de Ecuador considerando que existe mayor concentración de defunciones fetales en la región litoral con respecto a las defunciones fetales en las regiones sierra y oriente. A partir de modelos de regresión lineal se elaboraron modelos generalizados como el modelo de Regresión Logística Multinomial con datos agrupados en tres grupos que representan las causas de las defunciones fetales. Así mismo se desarrolló un modelo de regresión de Poisson con el cual se observó qué variables influyen directamente con las muertes fetales, entre estas variables explicativas se consideró la edad y la región donde habita la madre. Para la elaboración de los modelos se utilizó la base de datos del INEC del año 2017, en donde se infiere que las variables que tienen mayor influencia con respecto al número de muertes fetales son las variables: causa fetal agrupada, región de inscripción, zona y nacidos puesto que su p valor es menor que el nivel de significancia $\alpha=0.05$, por lo tanto, se concluyó que los modelos son significativos. Por otro lado, la variable que no influye al número de muertes fetales es la edad de la madre pues su valor p es mayor que el nivel de significancia se obtuvo que su valor p es 0.2327 por lo tanto se concluyó que el modelo no es significativo.

Palabras claves: Modelos lineales Generalizados, binomial, Poisson, defunciones fetales

ABSTRACT

This project deals with "Generalized linear models in the analysis of fetal deaths on the Ecuadorian coast", which includes a descriptive and analytical study of the situation in Ecuador, considering that there is a higher concentration of fetal deaths in the coastal region with respect to fetal deaths in the sierra and eastern regions. From linear regression models, generalized models such as the Multinomial Logistic Regression model were elaborated with data grouped in three groups that represent the causes of fetal deaths. Likewise, a Poisson regression model was developed with which it was observed which variables directly influence fetal deaths. Among these explanatory variables, age and the region where the mother lives were considered. For the elaboration of the models, the INEC database of the year 2017 was used, where it is inferred that the variables that have the greatest influence with respect to the number of fetal deaths are the variables: grouped fetal cause, region of registration, area and born since their p value is less than the level of significance $\alpha = 0.05$, therefore, it was concluded that the models are significant. On the other hand, the variable that does not influence the number of fetal deaths is the mother's age since its p-value is greater than the level of significance, it was obtained that its p-value is 0.2327, therefore it was concluded that the model is not significant .

Key words: Generalized linear models, binomial, poisson, fetal deaths

DEDICATORIA

La presente investigación dedico en primer lugar a Dios por haber permitido llegar a una meta más percibida en mi vida, en segundo lugar, a mis padres que sin su apoyo incondicional no hubiese podido lograr tan anhelado sueño.

A mis hermanos Verónica y Andrés que han sido un apoyo moral y emocional para seguir con esta meta, y a mis amigos que siempre han estado pendientes de lo que estoy haciendo en el transcurso de mi vida.

AGRADECIMIENTO

Agradezco en primera instancia a Dios porque con su amor me ha permitido llegar esta fase de culminación de mi vida profesional, a mis padres que con su apoyo incondicional siempre han estado pendiente de mi porvenir, mis hermanos y amigos que con sus palabras de aliento me han acompañado en este camino, en especial a la Dr. Sandra García Bustos por sus criterios en la realización del proyecto.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.



Alexandra Maribel Arguello Pazmiño

TRIBUNAL DE GRADUACIÓN



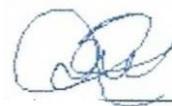
M.Sc. Heydi Roa López

PRESIDENTE



M.Sc Sandra González Camba

VOCAL 1



Sandra García Bustos, Ph.D.

DIRECTOR



Omar Ruíz Barzola, Ph.D.

VOCAL 2

ABREVIATURAS O SIGLAS

OMS= Organización Mundial de la Salud

ENSANUT = Estudio de Nacional de Salud y Nutrición

GLM= Modelos Lineales Generalizados

IMC= Índice de masa corporal

CIUR = Crecimiento uterino retardado

AIC= Criterio de información Akaike

TABLA DE CONTENIDO

CAPÍTULO 1	1
1. INTRODUCCIÓN	1
1.1. Antecedentes.	1
1.2. Descripción del problema.....	2
1.3. Objetivos	4
1.3.1. Objetivo General.	4
1.3.2. Objetivos Específicos.....	4
1.4. Hipótesis	5
1.5. Alcance	5
CAPÍTULO 2.....	6
2. MARCO TEÓRICO	6
2.1. Defunciones fetales.....	6
2.2. Diagnósticos de la muerte fetal.....	8
2.3. Factores de riesgo asociados a la muerte fetal.....	8
2.4. Causas de la muerte fetal.	12
2.5. Modelos regresión lineal	14
2.5.1. Regresión Lineal Simple	14

2.5.2.	Inferencia de Regresión lineal para β_0 y β_1	19
2.5.3.	Modelo de Regresión lineal Múltiple	21
2.6.	Modelos Lineales Generalizados (GLM)	28
2.7.	Modelo Lineal Generalizado.....	30
2.8.	Modelo Binomial.....	34
2.8.1.	Regresión Logística Multinomial	36
2.9.	Modelo Poisson.....	40
2.9.1.	Distribución de Poisson	42
2.9.2.	Propiedades de la Distribución de Poisson.....	42
CAPÍTULO 3.....		43
3.	METODOLOGÍA.....	43
3.1.	Diseño de la investigación.....	43
3.1.1.	Tipo de investigación	43
3.1.2.	Métodos de investigación.....	44
3.1.3.	Técnicas de investigación	44
3.1.4.	Datos	44
CAPÍTULO 4.....		50
4.	RESULTADOS	50
4.1.	División por rangos de edad.....	50

4.2.	Análisis de Causas de defunciones fetales	52
4.3.	Análisis de las variables para el Modelo	54
4.4.	Modelos de regresión lineal	57
4.4.1.	Modelo uno de regresión lineal: considerando la variable edad .	58
4.4.2.	Modelo dos de Regresión lineal: considerando la variable Enfermedad	59
4.4.3.	Modelo tres de Regresión lineal: considerando la variable Región 61	
4.4.4.	Modelo cuatro de Regresión lineal: considerando la variable Nacidos 63	
4.4.5.	Modelo cinco de Regresión lineal: considerando la variable Zona 65	
4.5.	Modelo de Regresión lineal Múltiple	67
4.6.	Modelo lineal Generalizado (Distribución Binomial).....	70
4.6.1.	Modelo de Regresión logística Multinomial.....	70
4.7.	Modelo Lineal Generalizado (Distribución Poisson).....	73
CAPÍTULO 5.....		80
5.	CONCLUSIONES Y RECOMENDACIONES.....	80
5.1.	Conclusiones.....	80
5.2.	RECOMENDACIONES.....	82

6.	Referencias.....	83
7.	Apéndices y Anexos.....	88

ÍNDICE DE FIGURAS

Figura 4.2-1 Análisis Causas Fetales a partir del Diagrama de Pareto.....	53
Figura 4.2-2 Comparación de defunciones Fetales por Región.....	54
Figura 4.3-1 Barplot de observaciones por grupos-edad - madre.....	55
Figura 4.3-2 Barplot de frecuencias de las observaciones por Región	57
Figura 4.4.1-1 Coeficientes del Modelo de Regresión lineal- edad.....	58
Figura 4.4.1-2 Comportamiento del modelo visualmente.....	59
Figura 4.4.2-1 Coeficientes – Modelo-Regresión lineal- Enfermedad.....	60
Figura 4.4.2-2 Comportamiento del modelo dos visualmente	61
Figura 4.4.3-1 Coeficientes-Modelo-Regresión lineal-variable Región	62
Figura 4.4.3-2 Comportamiento del modelo tres visualmente.....	63
Figura 4.4.4-1 Coeficientes- Modelo de Regresión lineal-Nacidos	64
Figura 4.4.4-2 Ajuste del modelo cuatro visualmente	65
Figura 4.4.5-1 Coeficientes-Modelo de Regresión lineal-variable Zona.....	65
Figura 4.4.5-2 Comportamiento del modelo cinco visualmente	66
Figura 4.5-1 Coeficientes del Modelo de Regresión lineal Múltiple.....	67
Figura 4.5-2 Función-Autocorrelación.....	68
Figura 4.5-3 Matriz de varianzas y covarianza robusta.....	69
Figura 4.6.1-1 Coeficientes – Modelo-Regresión Logística Multinomial	71

Figura 4.7-1 Función de Autocorrelación y Parcial78

ÍNDICE DE TABLAS

Tabla 2.3-1 Factores de Riesgo.....	9
Tabla 2.4-1 Causas de Hipoxia intrauterina	12
Tabla 2.4-2 Análisis de Varianzas.....	12
Tabla 2.7-1 Modelos Lineales Generalizados	33
Tabla 4.1-1 Rango por Categorías.....	50
Tabla 4.1-2 Edad de la Madre Agrupada	51
Tabla 4.3-1 Tabla de contingencia Región -área	55
Tabla 4.3-2 Tabla de contingencia Región -área	56
Tabla 4.4.1-1 Resumen del Modelo Uno	59
Tabla 4.4.3-1 Resumen del modelo tres	62
Tabla 4.4.4-1 Resumen del modelo cuatro	64
Tabla 4.4.5-1 Resumen del modelo cinco.....	66
Tabla 4.5-1 Verificación del supuesto de normalidad-Residuos	68
Tabla 4.6-1 Datos para Modelos Binomiales	70
Tabla 4.6.1-1 Distribución logística Multinomial estimación exponencial.....	72
Tabla 4.7-1 Distribución de Poisson en Rstudio	74
Tabla 4.7-2 Distribución de Poisson a partir de estimación exponencial	75
Tabla 4.7-3 Verificación del supuesto de normalidad - residuos.....	79

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Antecedentes.

La defunción fetal es considerada como un problema social desde hace muchas décadas a nivel nacional y mundial. Habría que decir también que las políticas económicas y de salud aplicadas en el Ecuador han encaminado la disminución en la tasa de mortalidad. Dicho lo anterior en el año 1997 fue de 10 por cada 100 nacidos vivos y en el 2017 esta tasa descendió a 7 según el Instituto Nacional de Estadísticas y Censos. (INEC, 2017)

Por otra parte, según la Organización Mundial de la Salud (2017), 2.6 millones de bebés mueren antes de nacer. Por otro lado, define a la defunción fetal como la muerte antes de la expulsión completa de su madre, independientemente de la duración del embarazo, el feto no respira y no presenta algún otro signo de vida. De modo que la tasa de mortalidad fetal en los países desarrollados se encuentra de 5 por cada mil nacidos vivos mientras que en los países en vías de desarrollo va de 30 por cada mil nacidos vivos. El bajo peso al nacer, el tiempo de gestación, las enfermedades maternas como hipoxia intrauterina, son algunos de los factores de riesgo más relevantes en la defunción fetal, donde tanto el feto como la mujer embarazada están expuestos a diversos factores de riesgo y amenazas que pueden poner en peligro de muerte (OMS, 2017).

Las defunciones fetales en un país o región pueden tomarse como un indicador para expresar el riesgo sobre las expectativas de vida y sobrevivencia de un recién nacido durante sus primeros 28 días. Los indicadores socioeconómicos y sociales muestran las condiciones de atención de salud brindadas a la madre por parte de los servicios de salud de un país. El adecuado control y tratamiento de enfermedades por gestación y complicaciones del parto conlleva a determinar si el país está preocupado por este tipo de riesgo. Por otra parte, el embarazo a pesar de ser un proceso biológico natural está acompañado de ciertos riesgos que pueden amenazar la vida de la madre e hijo como pueden ser infecciones y

problemas con nutrición que representan el 70% de la morbilidad y que pueden ocurrir en el primer año de vida del neonato.(Fernández, Castillo Núñez, & Herrarte Llamas, 2013).

Según (María J et al., 2016) señalan que una de las causas de mayor frecuencia de muerte fetal en América Latina es la hipoxia intrauterina, condición materna que genera más muertes fetales, la segunda causa es la compresión del cordón umbilical. La hipoxia intrauterina es el desprendimiento prematuro de la placenta (DPP) que representa el 24% de muertes seguida por la compresión del cordón umbilical 18.5% de un total de 7300 muertes fetales. Es así como “La tasa de mortalidad fetal fue 56.28 muertes fetales por cada 1000 nacimientos y abortos en América Latina. En cuanto al ajustar por edad gestacional se determinó una tasa de mortalidad fetal temprana 12.84; mortalidad fetal intermedia 15.48 y mortalidad fetal tardía 27.78” (María J et al., 2016). Según el (INEC, 2012), en su estudio Nacional de Salud y Nutrición- ENSANUT señala que la salud materna e infantil, así como la salud sexual y reproductiva se encuentran en riesgo, puesto que la tasa de fecundidad global muestra una disminución de 5.4 en los años comprendidos de 1975 – 1979, a 3.00 en el año 2012. Es decir que existe una disminución en el promedio de hijos que tendría una mujer durante su vida fértil, también señala que la mayor concentración de la tasa de fecundidad se concentra en el sector urbano con un indicador de número de hijos promedio de 2.9 por familia.

Con estos precedentes se consideró la región Costa como referencia de las defunciones fetales puesto que cuenta con mayor número de defunciones fetales es así como de 1950 datos analizados 1122 se agrupan en la Región Costa es decir el 58% de los datos de defunciones fetales se encuentran en esta región, además la causa que mayor índice origina las defunciones fetales es hipoxia intrauterina considerada.

1.2. Descripción del problema.

La muerte fetal es un riesgo desde el primer momento de la concepción, tanto la mujer embarazada como el feto están expuestos a diversos factores de peligro que amenazan sus vidas y representa una de las situaciones más relevantes de salud para los países en vías de desarrollo.

Según (Bedoya & Velasquez, 1998), manifiestan que la muerte fetal es un parámetro no medible a nivel mundial esto por no haber estadísticas estandarizadas en cada país. Más aún no se hace una diferencia entre mortalidad fetal temprana, intermedia y tardía. Las tasas de mortalidad perinatal es uno de los parámetros que se utiliza para medir el grado de desarrollo de un país, siendo este parámetro uno de los más importantes. Habría que decir que la tasa de mortalidad fetal es un indicador socioeconómico que visualiza el riesgo de vida de una nación. Cabe mencionar que la muerte fetal es un problema de salud, según la OMS citado en (Calvillo & Noe, 2014), señalan una tasa de muerte tardía de 23.3/1000 recién nacidos a nivel mundial así mismo indican que las causas de muerte se debe a: infarto, necrosis, calcificaciones excesivas, corangioma, transfusión feto-feto, insuficiencia placentaria, circular, nudo, procidencia, hematoma, hipertensión arterial sistémica, diabetes mellitus ,infección placentaria, anomalía congénita, malformación de un órgano ,infección ascendente ,trauma de parto, no determinadas.

La defunción fetal según la Organización Mundial de la Salud es la pérdida de una gestación in útero a partir de las 22 semanas y un peso mayor a 500g, constituye un tercio de toda la mortalidad fetal e infantil y más del 50% de todas las muertes perinatales en los países desarrollados. Se debe agregar que por cada 1000 nacimientos se presentan 5.3 muertes fetales en los países desarrollados y 25.5 en los países en desarrollo (Paladines Chicango , 2014).

En el año 2015 la revista científica *The Lancet*, reveló que Ecuador ocupa el décimo lugar de América Latina con menos muertes intrauterinas (7.7) por cada mil nacidos vivos, siendo superado por Venezuela (7.1) y Nicaragua (7.4) (Red Bioética, 2017). Habría que decir también que en el año 2017 Ecuador redujo esa tasa a (6.6), a nivel nacional se registró 1950 defunciones fetales, de las cuales el 58% a muerte fetal tardía es decir superior a 28 semanas de gestación (INEC, 2017). Datos que resultan alarmantes puesto que la defunción fetal es un indicador incuestionable de la calidad en la atención prenatal y de trabajo obstétrico (Rodríguez, 2014). Ecuador es uno de los países participes en el plan “Todos los Recién Nacidos” de la OMS que tiene como objetivo reducir las defunciones fetales, es así como en el año 2035

habrá 10 muertes fetales por 1000 nacidos y 14 en el 2020 a nivel mundial (Gavilanez Ocampo & Paucar Chalán , 2019). Por otro lado, según (Gavilanez Ocampo & Paucar Chalán , 2019) aluden que Ecuador en el año 2017 redujo la tasa de muerte fetal a 6.6 correspondiente a muerte fetal tardía, teniendo en cuenta que una mujer embarazada debe recibir una adecuada atención primaria encaminada hacia la atención secundaria.

En cuanto al papel importante que tienen los registros de defunciones fetales para identificar la magnitud y distribución de enfermedades que son esenciales en el diseño, implementación, monitoreo y evaluación de políticas de salud (INEC;, 2018). Es por ello por lo que las defunciones fetales evidencian la eficacia de las políticas públicas establecidas en un país en cada período de tiempo.

Con estos antecedentes las causas de la defunción fetal son complejas y difíciles de definir, este indicador coadyuva a la eficacia de políticas públicas que pueden adoptar los gobiernos de un país en cada período de tiempo, es por ello que resulta contar con documentos que contengan estadísticas a partir de modelos estadísticos generalizados que ayuden a dar respuesta a las variables que influyen en las defunciones fetales y que aporten a la toma de decisiones.

1.3. Objetivos

1.3.1. Objetivo General.

Identificar los factores que influyen en las defunciones fetales registradas en el litoral ecuatoriano en el año 2017, aplicando modelos lineales generalizados.

1.3.2. Objetivos Específicos.

- Argumentar teóricamente el desarrollo de los modelos lineales Generalizados en el análisis de las defunciones fetales
- Identificar las variables de estudio que contribuyen a la obtención del modelo a partir de un análisis de tablas de contingencia
- Analizar la defunción fetal a partir del modelo de regresión lineal con la aplicación de modelos lineales generalizados, modelo binomiales y modelo de distribución Poisson

1.4. Hipótesis

H0: “Hay independencia entre las variables predictoras (edad, región, zona, enfermedad, muertes, nacidos) respecto al número de muertes fetales”

H1: “No hay independencia entre las variables predictoras (edad, región, zona, enfermedad, muertes, nacidos) respecto al número de muertes fetales”

1.5. Alcance

El proyecto se realizó sobre la base de datos de las defunciones fetales del litoral ecuatoriano recogidas del Instituto Ecuatoriano de Estadísticas y Censo, donde se identificaron las variables de análisis como: edad, zona, enfermedad muertes y nacidos a partir de modelos de regresión y modelos lineales generalizados (modelo binomial y modelo de Poisson) sobre datos recolectados del año 2017. Sin embargo, también se ha analizado las otras regiones para efectos de comparación con la costa.

CAPÍTULO 2

2. MARCO TEÓRICO

2.1. Defunciones fetales

Uno de los acontecimientos que deben afrontar los profesionales que atienden a una mujer en estado de gestación, es encontrarse con la muerte del feto, en ciertas ocasiones se trata de un proceso diagnosticado con antelación. Lo dicho hasta aquí permite de alguna forma preparar a los padres para ese momento, sin embargo, en la mayoría de los casos se trata de un hallazgo y, por lo tanto, de un evento no previsto. Con respecto al óbito fetal se define como la muerte en el útero del feto en desarrollo. Generalmente va desde las 20 a 22 semanas de gestación con un peso de 500 gramos aproximadamente, es así que (Paladines Chicago, 2014) señala que la edad de gestación que se debe considerar como muerte fetal es de 22 semanas, lo cual concuerda internacionalmente.

Con respecto a la muerte fetal la OMS (2012) define a la muerte fetal como aquella que ocurre antes de la expulsión o extracción completa del producto del embarazo. Por otro lado “La muerte fetal se produce dado que después de dicha separación, el feto tiene ausencia de latidos cardiacos, pulsaciones del cordón, respiraciones espontáneas y movimientos voluntarios” (Guerrero , 2014, p.7).

Para (Valenti & Otegro, 2007) manifiestan que la muerte fetal es uno de los accidentes obstétricos más difíciles de enfrentar, debido a que las madres les significan un desencanto que desvanece todas las expectativas forjadas en ese futuro niño. La pérdida del embarazo es un fenómeno muy frecuente que se encuentra alrededor del 30% de las gestaciones, aunque muchas de estas pérdidas se tratan de patologías que no se llega a detectar clínicamente.

Con otro nombre que se puede conocer una defunción fetal es el síndrome de muerte fetal que se define como la muerte del feto que se presenta a partir de la semana 22 de gestación en ausencia de trabajo de parto. (Salud, 2010).

En el estudio de (Ottersen, Frenk, & Horton, 2011) señalan que cerca de 2 millones de muertes fetales ocurren durante el trabajo de parto, y en la mayoría de los casos

se trata de fetos de término. A nivel mundial no existen estadísticas estandarizadas de las muertes fetales y estas son manejadas en cada país con sus propias tasas e indicadores de mortalidad fetal, esto debido a que la mayoría de los casos no se discriminan las tasas de mortalidad perinatal, así como el tiempo de diferencia significativa entre la mortalidad fetal temprana, intermedia y tardía.

Según (Fernández , Castillo, Núñez , & Herrarte Llamas , 2013) en su estudio encontraron que la tasa de mortalidad fetal es de 39.16 por cada 1000 nacimientos, a partir de estudios realizados en Colombia en los departamentos de Medellín, Itagüí y Bello en donde se determinó que el 60% de las muertes fetales aportan a los indicadores nacionales de un país. Habría que decir también que la mortalidad perinatal está relacionada a los riesgos obstétricos directos, hipoxia y asfixia antes o durante el parto, otros factores relacionados con la mortalidad perinatal son, el estado biológico de la madre, factores sociodemográficos, accesibilidad y calidad en los servicios de salud.

En un estudio realizado en México las defunciones han disminuido en los últimos 15 años de 30.1 por cada 1000 nacidos vivos en 1984 hoy se encuentra en 28 por cada 1000 nacidos vivos, pero sin embargo existen estratos donde las condiciones socioeconómicas y de subdesarrollo pueden ser parámetros de influencia para que la tasa de mortalidad fetal incremente. (Rivera, Fuentes Román, Albores Esquinca, Abarca, & Hernández Girón, 2003)

Con respecto a los tipos de defunción fetal, existen tres tipos que se dividen según el tiempo en el que acontece, estas son:

- **Defunción fetal temprana o aborto:** Acontece antes de cumplirse 20 semanas de gestación, con su peso <500g.
- **Defunción fetal intermedia:** Se produce entre las semanas de 20 y 27, con peso entre 501-999g
- **Defunción fetal tardía:** Ocurre a partir de la semana 28 de gestación, con un peso >100g. (Tinedo, y otros, 2016)

2.2. Diagnósticos de la muerte fetal.

La incidencia de la muerte fetal si bien según (Fernández , Castillo, Núñez , & Herrarte Llamas , 2013) explican en su estudio se ha reducido en los países desarrollados por implementación de políticas de salud adecuadas y por el desarrollo de normas específicas de manejo de embarazadas de alto riesgo sin que esto signifique la disminución o desaparición de estos indicadores. Pero entre las razones que explica este estancamiento de la tasa de mortalidad fetal es el desconocimiento de la etiología, favoreciéndose entre las principales causas conocidas como son infecciones bacterianas ascendentes, anomalías congénitas preeclampsia, y otras causas que corresponden a infartos placentarios, infecciones transplacentarias, traumas fetales, consumo de tabaco, insuficiencia placentaria, sensibilización del factor Rh y diabetes.

Con respecto al diagnóstico, los datos clínicos relevantes en el diagnóstico de muerte fetal son: ausencia de movimientos fetales, ausencia de crecimiento uterino y ausencia de frecuencia cardíaca fetal. En pacientes con antecedentes de muerte fetal se debe realizar evaluación clínica y pruebas de laboratorio para conocer el bienestar materno e identificar la patología relacionada con la muerte fetal. Por otro lado, que incluyan búsqueda de procesos infecciosos, hiperglucemia siempre enfocándose a identificar la causa de muerte, posibilidad de recurrencia y complicaciones en futuros embarazos. A través de estudios ultrasonográficos se puede documentar la ausencia de actividad cardíaca fetal y otros datos agregados como: cabalgamiento de huesos, maceración, observación de una masa fetal irreconocible, gas intracavitario lo que puede limitar la calidad de las imágenes en tiempo real. En otras palabras, el estudio más indicado para el diagnóstico de muerte fetal es la ultrasonografía puesto que la realización de la está en tiempo real tiene como fin visualizar el latido cardíaco fetal y los movimientos fetales activos. (Valdés Vargas, González , & Vásquez López , 2012)

2.3. Factores de riesgo asociados a la muerte fetal.

Los factores asociados son diversos los cuales conllevan a desenlaces adversos en la gestación. Se puede apreciar dentro de ellos los factores socioeconómicos y ambientales así como las enfermedades médicas que dificultan la problemática,

entre ellas se destacan: la hipertensión arterial y la diabetes mellitus que tienen una relación con hasta la mitad de los óbitos fetales .(Rangel Calvillo, 2014)

Además, (Salud, 2010) considera que los factores como: el número de embarazos, raza, obesidad, extracto socioeconómico, condiciones médicas, acceso a los servicios de salud y complicaciones médicas están dentro de los factores de riesgo para las mujeres y causas de muerte perinatal. Alguno de estos factores es modificable, sin embargo, hay que considerar que los óbitos fetales son efectos obstétricos adversos que incluyen un daño neurológico. Las mujeres con embarazos complicados que resultaron en nacimientos de niños vivos y aquellas con un primer embarazo complicado finalizaron por cesárea. Por otra parte, el alto riesgo por enfermedad de origen vascular tiende a incrementar el riesgo de muertes fetales. Las anomalías de la anatomía o de la longitud del cordón umbilical y su torsión juegan un papel importante en la detección de estas anomalías, por ello se han relacionado con el orbito fetal tras un análisis retrospectivo los factores que influyen en estas anatomías que se visualiza en la tabla 2.3-1.

Tabla 2.3-1 Factores de Riesgo.

Tipo de factor de riesgo	Evidencia	Nivel / grado
Los factores de riesgo relacionados con la muerte fetal	<ul style="list-style-type: none"> • Maternos. • Fetales • Otros 	Clínicos obstétricos y ginecológicos
Factores maternos que se clasifican en modificables y no modificables	No modificables: <ul style="list-style-type: none"> • Antecedentes de pérdida fetal. • Factor Rh negativo en la mujer. • Enfermedades sistémicas (hipertensión arterial crónica, diabetes) 	

	enfermedades renales, cardiacas	
	Modificables: <ul style="list-style-type: none"> • Tabaquismo. • Alcoholismo • Drogas • Obesidad 	Lineamientos técnicos médicos para la atención obstétrica durante el embarazo parto y puerperio.
Las causas de muerte fetal ante parto de origen materno más comunes son:	<ul style="list-style-type: none"> • Diabetes mellitus • Preeclampsia • Hemorragias • Retraso de crecimiento • Infecciones • Anomalías congénitas 	

Fuente: Adaptado de Salud, (2010)

En cuanto a las mujeres de raza negra (afroamericanas) han mostrado tener dos veces más riesgo con respecto a su óbito fetal, también tienen mayor tendencia a presentar diabetes, hipertensión y obesidad, lo que conlleva a ruptura prematura de membranas, desprendimientos placentarios y por ende la muerte fetal. Habría que decir también que las condiciones socioeconómicas de los países en vías de desarrollo incrementan el factor vulnerable de riesgo para muerte fetal, obstáculos relacionados al acceso de servicios de salud y atenciones médicas deficientes por factores sociales, económicos, geográficos, culturales, políticos y gubernamentales. (Paladines Chicago, 2014)

Valentí y Otero (2007) manifiestan los siguientes factores:

Edad Materna Avanzada.

Uno de los factores de riesgo se centra en la edad avanzada de una mujer, puesto que por su edad se incrementa la tendencia de padecer condiciones médicas asociadas que están relacionadas con mayores tasas de efectos perinatales adversos. Así como las mal formaciones congénitas que generalmente se dan en

las mujeres de edad avanzada respecto a su gestación. Es así como la edad es un factor de riesgo para el óbito fetal, con el diagnóstico temprano de las anomalías congénitas, las muertes fetales se han reducido, sin embargo, esta mortalidad se ha trasladado hacia las primeras 20 semanas de gestación. (Valenti & Otero, 2007)

La Obesidad.

Los altos índices de obesidad que sufre la población y que en la actualidad está en aumento también involucra a las mujeres en edad de gestación, este indicador social se ha convertido en un problema de salud pública. Las mujeres con obesidad han puesto en riesgo al óbito fetal. Más aún se ha relacionado con factores de comportamiento, socioeconómicos y obstétricos, aumentando el índice de masa corporal (IMC). Las mujeres con índices altos de masa corporal tienden a percibir menos los movimientos fetales, por consiguiente, no mantienen un seguimiento constante de sus fetos; muchas son hiperlipidemias, lo cual causa alteraciones endoteliales con ella, agresión plaquetaria y aterosclerosis. Se debe agregar que las mujeres gestantes obesas presentan más episodios de ronquidos, apnea del sueño y disminución de la saturación del oxígeno, lo cual se ha relacionado con mayor hipertensión inducida por el embarazo y restricción del crecimiento intrauterino. (Valenti & Otero, 2007)

Hipertensión arterial.

La hipertensión arterial es la complicación de alto riesgo más frecuente en la mujer embarazada. El manejo clínico compromete a diversos campos de la medicina moderna y los planes terapéuticos aún son controvertidos debido al desconocimiento de los verdaderos factores etiológicos. (Voto, 2013)

Las muertes relacionadas con la hipertensión arterial sistémica cada vez son más frecuentes, las mejoras en el servicio médico y obstétrico se relacionan con los estados más graves y obedecen a complicaciones como desprendimientos prematuros de la placenta, hemorragia materna, insuficiente o infarto placentario. (Valenti & Otero, 2007)

2.4. Causas de la muerte fetal.

De acuerdo Medina Castro y Castro Llamas (2012) mencionan que las causas para el óbito fetal son diversas se las pueden clasificar en tres grupos: a) Maternas. B) Fetales. C) Placentarias y de cordón umbilical. Muestran que entre el 25 a 35% tiene defectos congénitos que incluyen defectos aislados (40 %), múltiples (40%) y deformaciones o displasias (20%) esto en proporción de muertes fetales que se les atribuye causas genéticas. De estas la mayor parte son alteraciones cromosómicas, que representan entre el 6 y el 15 por ciento de los casos estudiados. (Medina Castro & Castro Llamas, 2012).

De acuerdo con datos proporcionados por la Organización Mundial de la Salud (OMS) las tres principales causas de muerte de recién nacidos en el mundo son las infecciones, la prematuridad y la asfixia perinatal o complicaciones del parto que constituyen el 23% de la mortalidad neonatal global. La hipoxia intrauterina no es más que la asfixia perinatal y está puede ocurrir antes del nacimiento, durante el embarazo, el trabajo de parto, el parto y después de él. (Rivera Miranda, Lara Latamblé, & Baró Bouly, 2017).

Tabla 2.4-1 Causas de Hipoxia intrauterina

Causas	Características
Maternas	<ul style="list-style-type: none">• Hemorragia en el tercer trimestre.• Infecciones (urinarias, corioamnionitis, sepsis).• Hipertensión inducida por el embarazo.• Hipertensión crónica.• Anemia.• Colagenopatías.• Intoxicación por drogas.• Comorbilidad obstétrica.
Obstétricos	<ul style="list-style-type: none">• Líquido amniótico meconial

	<ul style="list-style-type: none"> • Incompatibilidad cefalopélvica • Uso de medicamentos: oxitocina • Presentación fetal anormal • Trabajo de parto prolongado o anormal • Parto instrumentado o cesárea • Ruptura prematura de membrana • Oligoamnioso polihidramnios
Uteroplacentarias	<ul style="list-style-type: none"> • Anormalidades de cordón: Circular del cordón irreductible, procúbito, prolapso del cordón umbilical. • Anormalidades placentarias: Placenta previa, desprendimiento prematuro de placenta, alteraciones de la contractilidad uterina, hipotonía uterina, hipertonia uterina. • Anormalidades uterinas anatómicas: Útero bicorne
Fetales	<ul style="list-style-type: none"> • Alteraciones de la frecuencia cardíaca fetal: Bradicardia, taquicardia, arritmia, percepción de disminución de movimientos fetales por la madre, retardo del crecimiento intrauterino, prematuridad, bajo peso, macrosomía fetal, posmadurez, malformaciones congénitas, eritrosblastosis fetal, fetos múltiples, retraso en el crecimiento intrauterino

Fuente: Adaptado de Rivera Miranda, Lara Latamblé, y Baró Bouly, (2017)

Por otra parte Cañamero de León (2015) señala que la muerte fetal es una situación terriblemente desconcertante para la madre, los familiares y los profesionales sanitarios. Se han llevado numerosos estudios que intentan buscar la causa de esta situación con el fin de poder identificarla y prevenirla. Sin embargo, en los países desarrollados 1 de cada 160 bebés nacen muertos, o cesa su latido en el último trimestre del embarazo; con lo que es una situación muy dolorosa para los padres y no tan infrecuente. Todavía cabe señalar las causas que originan la muerte del Bebé en vientre:

- Gestación múltiple (de dos o más bebés).
- CIUR (crecimiento uterino retardado): los bebés que crecen muy lentamente tienen un riesgo incrementado de morir durante el embarazo.
- Anomalías congénitas o genéticas: un defecto físico o genético en el bebé. Las malformaciones múltiples en los bebés pueden ser por causa genética, del medio ambiente o desconocida y siempre será importante consultar a un especialista en genética en casos de bebés que fallecen en estas circunstancias.
- Entre el 5 y el 10 por ciento de los bebés nacidos sin vida tienen anomalías relacionadas con sus cromosomas, que son las estructuras que contienen el material genético que determina nuestras características físicas. Sobre todo, en pérdidas antes de la semana 20, son particularmente frecuentes las anomalías de los cromosomas, aunque pueden provocar la muerte del feto en cualquier momento del embarazo.
- Las infecciones bacterianas que afectan al feto o a la placenta son una causa importante de las muertes fetales que ocurren entre las semanas 24 y 27 de gestación. A veces la mujer embarazada puede tener una infección que puede pasar inadvertida (como las infecciones del tracto genital y urinario y ciertos virus como el parvovirus) hasta que ha provocado serias complicaciones, tales como la muerte del feto o su nacimiento en forma prematura (antes de terminar la semana 37 del embarazo).

Por otra parte, después del parto es posible demostrar si la muerte del feto fue provocada por una infección bacteriana mediante la realización de pruebas especiales a la placenta. (Cañamero de León, 2015)

2.5. Modelos regresión lineal

2.5.1. Regresión Lineal Simple

La regresión lineal consiste en generar una ecuación (modelo) que, basándose en la relación existente entre ambas variables, permita predecir el valor de una a partir de la otra.

El modelo varía según qué variable se considere dependiente de la otra (lo cual no implica causa-efecto). (Amat Rodrigo, 2016)

Definición

Según Zurita Herrera (2010) la regresión lineal simple consiste en generar un modelo de regresión (ecuación de una recta) que permita explicar la relación lineal que existe entre dos variables. A la variable dependiente o respuesta se le identifica como **Y** y a la variable predictora o independiente como **X**. El modelo de regresión lineal simple se describe de acuerdo con la ecuación:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

$\beta_0 =$ *Es la ordenada en el origen*

$\beta_1 =$ *Pendiente*

$\varepsilon_i =$ *Error aleatorio*

$X_i =$ *Variable de explicación, condiciona el valor de Y*

Dado el modelo condicional se trabaja en los siguientes supuestos:

$$E(Y_i|X = X_i) = B_0 + B_1 X_i$$

$$E(\varepsilon_i) = 0; \text{Var}(\varepsilon_j) = \sigma^2; \text{cov}(\varepsilon_i, \varepsilon_j) = 0; i \neq j$$

Donde:

$$E(Y_i|X = X_i) = \text{Parte sistemática o determinística del modelo}$$

Los valores B_0, B_1 y σ^2 , son constantes desconocidas, pero estadísticamente estimables. El hecho de que la varianza σ^2 del error sea constante durante todo el proceso es un supuesto fuerte y hace que el modelo utilizado sea considerado como homocedástico, es decir que la variabilidad es constante (Zurita Herrera, 2010). El error aleatorio representa la diferencia entre el valor ajustado por la recta y el valor real. Recoge el efecto de todas aquellas variables que influyen en **Y** pero que no se incluyen en el modelo como predictores (Amat Rodrigo, 2016).

En la gran mayoría de casos, los valores β_0 y β_1 poblacionales son desconocidos, por lo que, a partir de una muestra, se obtienen sus estimaciones como:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x$$

Estas estimaciones se conocen como coeficientes de regresión o *least square coefficient estimates*, ya que toman aquellos valores que minimizan la suma de cuadrados residuales, dando lugar a la recta que pasa más cerca de todos los puntos (Amat Rodrigo, 2016).

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

Donde:

$$S_{xx} = \frac{\sum(x_i - \bar{x})^2}{n - 1} ; \quad S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Dando como resultado:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Mientras que $\hat{\beta}_0$ es:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Donde S_{xx} y S_{xy} son las desviaciones típicas de cada variable. $\hat{\beta}_0$ es el valor esperado la variable Y cuando $X = 0$, es decir, la intersección de la recta con el eje y . Una recta de regresión puede emplearse para diferentes propósitos y dependiendo de ellos es necesario satisfacer distintas condiciones. Siendo \bar{x} y \bar{y} las medias aritméticas de los valores observados de X y de Y ; \hat{y}_i es el valor que el modelo estima para Y_i , dado $X=x_i$ con lo efectuado anteriormente se puede obtener el coeficiente de Correlación r_{xy} siendo:

$$S_{yy} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$

Dando como resultado:

$$r_{xy} = \hat{\beta}_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$$

Por otro lado, el Coeficiente de determinación R^2 se define como el cociente de la suma cuadrática de Regresión para la suma Total:

Donde la suma cuadrática de Regresión (**SCR**) se define de la siguiente manera:

$$SCR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

La suma cuadrática del Error (**SCE**) se define de la siguiente manera:

$$SCE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

La suma cuadrática Total (**SCT**) se define de la siguiente manera:

$$SCT = SCR + SCE$$

$$SCT = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Dando como resultado:

$$R^2 = \frac{SCR}{SCT} = \frac{SCT - SCE}{SCT} = 1 - \frac{SCE}{SCT}$$

Por otro lado, se puede comprobar en la regresión lineal simple la relación entre el coeficiente de Correlación de X con Y y el coeficiente de determinación que viene dado por:

$$r_{xy} = \pm \sqrt{R^2}$$

Tabla de Análisis de Varianza

De acuerdo Zurita Herrera (2010) menciona el teorema de Cochran que establece que si cada una de las n observaciones y_i son tomadas de una misma Población Normal con parámetros μ_i y σ^2 , la SCT se descompone en K sumas Cuadráticas

que se denota SC_q , $q = 1; 2; \dots; k$. Cada una de ellas con gl_q grados de libertad; por lo que el cociente SC_q/σ^2 tiene distribución Ji-Cuadrado con gl_q grados de libertad, teniendo la SCT, $(n-1)$ grados de libertad esto significa:

$$\sum_{q=1}^k gl_q = n - 1$$

Se ha efectuado una partición de la SCT, donde $K=2$, la una parte de la SCT es la SCR ($q=1$) y la otra es la SCE, con $gl_1 = (p - 1)$ y la SCE ($q = 2 = k$) con $gl_2 = (n - p)$.

En la tabla 2.4-2 se visualiza la tabla de análisis de varianza o ANOVA que consiste en un arreglo rectangular cuyos componentes son: las fuentes de variación, grados de libertad, las sumas y medias cuadráticas y un valor adicional que es el estadístico de Prueba F

Tabla 2.4-2 Análisis de Varianzas

Fuentes de Variación	de	Grados de libertad	de	Sumas Cuadráticas	Medias Cuadráticas	Estadístico de Prueba F
Regresión		$p - 1$		$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{SCR}{p - 1}$	$F_0 = \frac{MCR}{MCE}$
Error (residuos)		$n - p$		$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$\frac{SCE}{n - p}$	
Total		$n - 1$		$\sum_{i=1}^n (y_i - \bar{y})^2$		

Fuente: Adaptado de Zurita Herrera, (2010)

Distribución F de Fisher

Es el cociente de:

$$F_0 = \frac{MCR}{MCE} = \frac{SCR/1}{SCE/n(n-2)}$$

Es una variable aleatoria F con $(p-1) = 1$ gl en el numerador y $(n-p)$ gl en el denominador. Si la Hipótesis Nula es verdadera ($H_0: \beta_1 = 0$) entonces, $E(MCR) = E(MCE) = \sigma^2$, por lo que un valor del estadístico de prueba F, cercano a “uno” sería evidencia de que de que la Hipótesis Nula es verdadera, en tanto que un valor alejado de uno, evidencia la necesidad de rechazar H_0 .

Por lo tanto, establecido el contraste:

$$H_0: \beta_1 = 0 \text{ Vs. } H_1: \beta_1 \neq 0$$

Con $(1-\alpha)$ 100% de confianza la Hipótesis Nula debe ser rechazada si el Estadístico de Prueba:

$$F_0 = \frac{MCR}{MCE} > F_{(\alpha; p-1, n-p)}$$

Una recta de regresión puede emplearse para diferentes propósitos y dependiendo de ellos es necesario satisfacer distintas condiciones. Sin embargo, en caso de querer predecir el valor de una variable en función de la otra, no solo se necesita calcular la recta, sino que además hay que asegurar que el modelo se ajusta a los datos (Amat Rodrigo, 2016).

2.5.2. Inferencia de Regresión lineal para β_0 y β_1

Significa que el modelo generado es una estimación de la relación poblacional a partir de la relación que se observa en la muestra y, por lo tanto, está sujeta a variaciones. Para cada uno de los parámetros de la ecuación de regresión lineal simple (β_0 y β_1) se puede calcular su significancia (p-value) y su intervalo de confianza. La prueba estadística más empleada es el t-test (existen alternativas no paramétricas).

La prueba de significancia para la pendiente (β_1) del modelo lineal considera como hipótesis:

H_0 : No hay relación lineal entre ambas variables por lo que la pendiente del modelo lineal es cero. $\beta_1 = 0$

H1: Sí hay relación lineal entre ambas variables por lo que la pendiente del modelo lineal es distinta de cero. $\beta_1 \neq 0$

Cálculo del estadístico T y del p-value:

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}, t = \frac{\hat{\beta}_0 - 0}{SE(\hat{\beta}_0)}$$

El error estándar de β_0 y β_1 se calcula con las siguientes ecuaciones:

$$SE(\hat{\beta}_0)^2 = \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

$$SE(\hat{\beta}_1)^2 = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

La varianza del error σ_ε^2 se estima a partir del error estándar residual (RSE), que no es más que la raíz cuadrada de la media de la suma de los residuos al cuadrado:

$$RSE = \sqrt{\frac{RSS}{n-2}}$$

Donde RSS suma de residuos al cuadrado se determina:

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Por otro lado, en la tabla de varianzas se lo representa como SCE

Grados de libertad (df)= número de observaciones -2=número observaciones – número predictores -1

$$p - value = P(|t| > valor calculado de t)$$

Los grados de libertad (df) son la cantidad de información suministrada por los datos que se puede dilapidar para estimar los valores de parámetros de población desconocidos y calcular la variabilidad de esas estimaciones. Este valor se determina según el número de observaciones de la muestra y el número de parámetros del modelo. Si incrementa el tamaño de la muestra, obtendrá más

información sobre la población y, por consiguiente, los grados de libertad de los datos aumentarán (Zurita Herrera, 2010).

Los intervalos de confianza para cada β_i :

Con $(1-\alpha)$ 100% de confianza para β_i , el intervalo para cada β_i es:

$$\hat{\beta}_i \pm SE(\hat{\beta}_i)t_{(\frac{\alpha}{2}, n-2)}$$

El intervalo quedaría como:

$$\hat{\beta}_i - SE(\hat{\beta}_i)t_{(\frac{\alpha}{2}, n-2)} \leq \beta_i \leq \hat{\beta}_i + SE(\hat{\beta}_i)t_{(\frac{\alpha}{2}, n-2)}$$

donde $t_{(\frac{\alpha}{2}, n-2)}$ es el percentil de la distribución t de Student con n-2 grados de libertad que deja a su derecha un área de $\alpha/2$. Cuanto menor es el número de observaciones n, menor la capacidad para calcular el error estándar del modelo. Como consecuencia, la exactitud de los coeficientes de regresión estimados se reduce.

2.5.3. Modelo de Regresión lineal Múltiple

La regresión lineal múltiple permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta (Y) se determina a partir de un conjunto de variables independientes llamadas predictores ($X_1, X_2, X_3 \dots$). Es una extensión de la regresión lineal simple, por lo que es fundamental comprender esta última. Los modelos de regresión múltiple pueden emplearse para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores sobre ella (esto último se debe que analizar con cautela para no malinterpretar causa-efecto) (Amat Rodrigo, 2016).

Los modelos lineales múltiples siguen la siguiente ecuación:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \dots + \beta_n X_{ni} + \varepsilon_i, \text{ donde}$$

$$\varepsilon_i \sim N(0, \sigma^2 I_n) \quad cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j; i, j = 1, 2, \dots, n$$

β_0 : es la ordenada en el origen, el valor de la variable dependiente Y cuando todos los predictores son cero.

β_i : es el efecto promedio que tiene el incremento en una unidad de la variable X_i sobre la variable dependiente Y , manteniéndose constantes el resto de las variables. Se conocen como coeficientes parciales de regresión.

ε_i : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

Esta vez con los parámetros $\beta_0, \beta_1, \beta_2$, dos variables de explicación X_1 y X_2 y el error ε_i . Es importante tener en cuenta que la magnitud de cada coeficiente parcial de regresión depende de las unidades en las que se mida la variable predictora a la que corresponde, por lo que su magnitud no está asociada con la importancia de cada predictor. Para poder determinar qué impacto tienen en el modelo cada una de las variables, se emplean los coeficientes parciales estandarizados, que se obtienen al estandarizar (sustraer la media y dividir entre la desviación estándar) las variables predictoras previo ajuste del modelo (Amat Rodrigo, 2016).

Supuestos del Modelo

Según Joaquin Amat (2016) menciona que los modelos de correlación lineal múltiple requieren de los siguientes supuestos:

1. No colinealidad o multicolinealidad:

En los modelos lineales múltiples los predictores deben ser independientes, no debe de haber colinealidad entre ellos. La colinealidad ocurre cuando un predictor está linealmente relacionado con uno o varios de los otros predictores del modelo o cuando es la combinación lineal de otros predictores. Como consecuencia de la colinealidad no se puede identificar de forma precisa el efecto individual que tiene cada una de las variables colineales sobre la variable respuesta, lo que se traduce en un incremento de la varianza de los coeficientes de regresión estimados hasta el punto de que resulta prácticamente imposible establecer su significancia estadística. Además, pequeños cambios en los datos provocan grandes cambios en las estimaciones de los coeficientes. Si bien la colinealidad propiamente dicha existe solo si el coeficiente de correlación simple o múltiple entre algunas de las variables independientes es 1, esto raramente ocurre en la realidad. Sin embargo, es frecuente encontrar la llamada casi-colinealidad o multicolinealidad no perfecta.

No existe un método estadístico concreto para determinar la existencia de colinealidad o multicolinealidad entre los predictores de un modelo de regresión, sin embargo, se han desarrollado numerosas reglas prácticas que tratan de determinar en qué medida afecta a la estimación y contraste de un modelo. Los pasos recomendados a seguir son:

- Si el coeficiente de determinación R^2 es alto, pero ninguno de los predictores resulta significativo, hay indicios de colinealidad.
- Calcular una matriz de correlación en la que se estudia la relación lineal entre cada par de predictores. Es importante tener en cuenta que, a pesar de no obtenerse ningún coeficiente de correlación alto, no está asegurado que no exista multicolinealidad. Se puede dar el caso de tener una relación lineal casi perfecta entre tres o más variables y que las correlaciones simples entre pares de estas mismas variables no sean mayores que 0.5.
- Generar un modelo de regresión lineal simple entre cada uno de los predictores frente al resto. Si en alguno de los modelos el coeficiente de determinación R^2 es alto, estaría señalando a una posible colinealidad.
- Tolerancia (TOL) y Factor de Inflación de la Varianza (VIF). Se trata de dos parámetros que vienen a cuantificar lo mismo (uno es el inverso del otro). El VIF de cada predictor se calcula según la siguiente fórmula:

$$VIF_{\hat{\beta}_j} = \frac{1}{1 - R^2}$$
$$Tolerancia_{\hat{\beta}_j} = \frac{1}{VIF_{\hat{\beta}_j}}$$

Donde R^2 se obtiene de la regresión del predictor X_j sobre los otros predictores. Esta es la opción más recomendada, los límites de referencia que se suelen emplear son:

- $VIF = 1$: Ausencia total de colinealidad
- $1 < VIF < 5$: La regresión puede verse afectada por cierta colinealidad.
- $5 < VIF < 10$: Causa de preocupación

- El termino tolerancia es $1/VIF$ por lo que los límites recomendables están entre 1 y 0.1.

En caso de encontrar colinealidad entre predictores, hay dos posibles soluciones. La primera es excluir uno de los predictores problemáticos intentando conservar el que, a juicio del investigador, está influyendo realmente en la variable respuesta. Esta medida no suele tener mucho impacto en el modelo en cuanto a su capacidad predictiva ya que, al existir colinealidad, la información que aporta uno de los predictores es redundante en presencia del otro. La segunda opción consiste en combinar las variables colineales en un único predictor, aunque con el riesgo de perder su interpretación. Cuando se intenta establecer relaciones causa-efecto, la colinealidad puede llevar a conclusiones muy erróneas, haciendo creer que una variable es la causa cuando en realidad es otra la que está influenciando sobre ese predictor.

2. Independencia

Para verificar que los errores son independientes entre observaciones, se aplica el test Durbin- Watson. Este test puede variar entre 0 y 4, donde un valor cercano a 2 significa que los residuos no están correlacionados, por otro lado, un valor superior a 2 indica una correlación positiva entre errores subyacentes, mientras que un valor menor a 2 representa una correlación negativa entre errores.

3. Distribución normal de los residuos

Los residuos se deben distribuir de forma normal con media cero. Para comprobarlo se recurre a histogramas, a los cuantiles normales o a test de hipótesis de normalidad.

4. Homocedasticidad

La varianza de los residuos debe de ser constante en todo el rango de observaciones. Para comprobarlo se representan los residuos. Si la varianza es constante, se distribuyen de forma aleatoria manteniendo una misma dispersión y sin ningún patrón específico. Una distribución cónica es un claro

identificador de falta de homocedasticidad. También se puede recurrir a contrastes de homocedasticidad como el test de Breusch-Pagan.

Estimación de los coeficientes del Modelo

De acuerdo Zurita Herrera (2010) señala que se debe recurrir al criterio de Mínimos Cuadrados para estimar los p coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$, del modelo de regresión múltiple.

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{ip-1} + \varepsilon_i$ a través de $b_0, b_1, b_2, \dots, b_{p-1}$, en su orden.

Se considera la desviación de Y_i respecto de su valor esperado $E(Y_i)$ o función de respuesta $E(Y_i)$, siendo:

$$E(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{ip-1} \text{ esto es:}$$

$$[Y_i - E(Y_i)] = Y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{ip-1})$$

Aplicando la derivación de la ecuación:

$$Q = \sum_{i=1}^n [Y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{ip-1})]^2$$

Al derivar Q con respecto a los p parámetros se obtienen los estimadores $b_0, b_1, b_2, \dots, b_{p-1}$ las ecuaciones previas se sintetizan en la igualdad:

$$x^T x b = x^T Y$$

El vector de estimadores es b donde:

$$\hat{\beta} = b = (x^T x)^{-1} x^T Y$$

En caso de que no exista $(x^T x)^{-1}$ se debe recurrir a la inversa generalizada de la matriz $x^T x$.

Estimación de la matriz covarianzas de b

Si se tiene un vector Aleatorio p -variado x con valor esperado $E[x^T] = \mu^T = [\mu_1 \mu_2 \dots \mu_p]$ y cuya matriz de varianzas y covarianzas es Σ_x ; la varianza de cualquier combinación lineal $U = a^T x$ es:

$$Var(U) = Var(a^T x) = \sigma_u^2 = a^T \Sigma_x a \text{ mientras que:}$$

$$E[a^T x] = [a^T E(x)] = a^T \mu$$

$$\text{Siendo } a^T = [a_1 \ a_2 \ \dots \ a_p]$$

Por otro lado, se puede probar que si A es una matriz $n \times p$, las n combinaciones lineales que representa el vector AX tienen matriz de varianzas y covarianzas $\Sigma_{AX} = A \Sigma_x A^T$ (Zurita Herrera, 2010).

Variabilidad de los errores en la regresión

Según Escalante Cortina, (2012) menciona :

- 1.- Variabilidad total: $SCT = \sum (y_i - \bar{y})^2$; sumatoria de los cuadrados totales
- 2.- Variabilidad explicada: $SCE = \sum (\hat{y}_i - \bar{y})^2$; sumatoria de los cuadrados de errores
- 3.- Variabilidad no explicada: $SCR = \sum u_i^2$; sumatoria de los cuadrados de la regresión

En el caso en el que el modelo hay una constante la $SCT = SCE + SCR$

La bondad del ajuste muestra que tan bien se ajusta la línea de regresión a los datos. El coeficiente de determinación para la regresión lineal simple R y el R^2 para la regresión múltiple, es una medida que refleja que tan bien se ajusta la línea de regresión muestral de los datos:

$$R^2 = 1 - \frac{SCE}{SCT} = 1 - \frac{\text{Suma de residuales cuadrados}}{\text{Suma total de Cuadrados}}$$

Por otro lado, una propiedad del R^2 es que es una función no decreciente del número de variables explicativas o de regresores presentes en el modelo; a medida que aumenta el número de regresores, el R^2 aumenta. Es por esto que al comparar dos modelos de regresión con la misma variable dependiente pero con un número diferente de variables X , se debe tener en cuenta el número de variables presentes en el modelo. Esto puede hacerse con el coeficiente de determinación ajustado \bar{R}^2 , el cual está ajustado por los grados de libertad asociados con la suma de los cuadrados:

$$\bar{R}^2 = 1 - \frac{\sum U_i^2 / (n - k)}{\sum (y_i - \bar{y})^2 / (n - 1)}$$

Pruebas estadísticas para medir la significancia del modelo y Parámetros

- La prueba “t” de student es utilizada para medir la significancia estadística de los parámetros del modelo, es decir de las betas. El estadístico “t” (t-statistic) que se calcula como cociente entre el estimador y su error estándar

$\hat{\beta}/\hat{\sigma}$ permite contrastar la hipótesis de que el coeficiente es igual a cero ($H_0: \beta_i = 0$ frente a $H_a: \beta_i \neq 0$).

- La prueba estadística “F” de Fisher puede medir la significancia global del modelo es decir que el modelo de regresión múltiple es estadísticamente significativo.

En ambos casos se busca un nivel de confianza del 95% con un p-valor entre cero y 0.05 para que el modelo y los parámetros sean estadísticamente significativos (Escalante Cortina, 2012).

Pruebas de Bondad de Ajuste

Las pruebas de hipótesis más empleados para analizar la normalidad de los datos. En todos ellos, se considera como hipótesis nula que los datos sí proceden de una distribución normal y como hipótesis alternativa que no lo hacen.

El p-value de estas pruebas indica la probabilidad de obtener una distribución como la observada si los datos proceden realmente de una población con una distribución normal. Cuando estas pruebas se emplean con la finalidad de verificar las condiciones de métodos paramétricos, es importante tener en cuenta que, al tratarse de p-valúes, cuanto mayor sea el tamaño de la muestra más poder estadístico tienen y más fácil es encontrar evidencias en contra de la H_0 (normalidad).

Al mismo tiempo, cuanto mayor sea el tamaño de la muestra, menos sensibles son los métodos paramétricos a la falta de normalidad. Por esta razón, es importante no basar las conclusiones únicamente en el p-value de la prueba, sino también considerar la representación gráfica y el tamaño de la muestra, entre ellos está la prueba de Kolmogórov-Smirnov (Amat Rodrigo, 2016).

Test de Kolmogorov-Smirnov

Permite estudiar si una muestra procede de una población con una determinada distribución (media y desviación típica), no está limitado únicamente a la distribución normal. En R Se ejecuta con la función `ks.test()` (Amat Rodrigo, 2016).

Prueba de Hipótesis de Bondad de Ajuste para la normalidad de los errores:

$H_0 =$ Los errores siguen una distribución Normal

Vs

$H_1 =$ Los errores no siguen una distribución Normal

Rechazar H_0 en favor de H_1 si el p-value < 0.05 . En esta prueba de hipótesis nos interesa en general no rechazar H_0 (p_values > 0.05), debido a que uno de los supuestos de los errores es que estos siguen una distribución normal.

Permite estudiar si una muestra procede de una población con una determinada distribución (media y desviación típica), no está limitado únicamente a la distribución normal, aunque en regresión lineal se utiliza para el contraste de hipótesis del supuesto de la normalidad de los errores. En R Se ejecuta con la función `ks.test()` (Amat Rodrigo, 2016).

Función de Autocorrelación

En ocasiones en una serie de tiempo acontece, que los valores que toma una variable en el tiempo no son independientes entre sí, sino que un valor determinado depende de los valores anteriores. Entendemos por correlación el valor que nos indica si dos variables están o no relacionadas. Cuantifica en una escala de -1 a 1 si la relación existente entre dos variables es negativa o positiva.

La función de autocorrelación de un proceso estocástico $(Y_t)_{t \in I}$ es una función de dos instantes que describe las correlaciones entre las variables en un par de instantes t, s cualesquiera:

$$\rho_{t,s} = \text{Cor}(Y_{t,t+s}) = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,s} \gamma_t}}, t, s \in I$$

Si el proceso es estacionario, los momentos de segundo orden no dependen de t . En R, con la función ACF podremos calcular tanto la función de autocovarianza como la función de autocorrelación (Molina & Agulló, 2018).

2.6. Modelos Lineales Generalizados (GLM)

Modelo

De acuerdo Judd y McClelland (1989,p.1) citado por (López González & Ruíz Soler, 2011) mencionan que un modelo es la función que se introduce con objeto de explicar los datos , puesto que la variabilidad recogida en datos no termina de estar explicada, se incluye el término error que hace referencia a la discrepancia o la

falta de ajuste entre los datos y modelo. A todo esto, se puede señalar que un modelo estudia la relación entre la realidad empírica y la explicación teórica.

Modelado Estadístico

Un modelo estadístico es un proceso consustancial al ser humano para comprender lo que sucede a nuestro alrededor, a partir de la observación detallada de los hechos, fenómenos. Por consiguiente cuando estos fenómenos a explicar son probabilísticos antes que determinísticos (López González & Ruíz Soler, 2011). Un modelo pretende explicar la variación de una respuesta a partir de la relación conjunta de dos fuentes de variabilidad, una de carácter determinista y otra aleatoria lo que revela al enunciado:

$$\text{Respuesta} = \text{componente sistemático} + \text{componente aleatorio}$$

De manera que un modelo estadístico se encarga de la construcción, formulación y ajustes de modelos a los datos empíricos respondiendo a tres criterios:

- 1. Criterio estadístico.-** Hace referencia a la inclusión de parámetros del modelo en beneficio de una mejor representación de los datos con la correspondiente disminución del error según menciona (Mc Clelland,1997) citado por (López González & Ruíz Soler, 2011).
- 2. Criterio lógico.** - La selección de los parámetros que formen parte del modelo de tal modo que éste se convierta en representación simple y sobria de la realidad según (Judd y Mc Clelland,1989,3; y Ruíz-Soler, Peregrina y López – González,2000) citado por (Ibídem,2011).
- 3. Criterio sustantivo.** - Conocido también como criterio de integración teórica del modelo en la red conceptual que lo generó conforme (Alto y Vallejo, 2007,47) citado por (Ibídem, 2011).

Etapas de un modelo estadístico

López González y Ruíz Soler (2011) establecen las siguientes etapas que un modelo parsimonioso puede explicar la variable de respuesta con el menor error posible.

1. **Especificación del modelo teórico.** - Establece las variables que son de interés, así como cuáles son las relaciones entre ellas, es decir que el modelo describa de la forma más simple posible la concordancia entre el modelo y los datos con un mínimo error.
2. **Estimación de parámetros.** - Evidencia el cálculo del valor de los coeficientes del modelo examinado a partir del conjunto de datos observados al objeto de determinar si el modelo teórico propuesto es aceptable como representación aproximada de los datos.
3. **Selección del modelo.** - Se refiere a la valoración, si el nivel de discrepancia entre los datos observados y los datos ajustados es suficientemente bajo como para optar por el modelo, o por el contrario suficientemente elevado como para rechazarlo.
4. **Evaluación del modelo.** - Consiste en examinar las observaciones individuales los datos influyentes y los datos anómalos, así como comprobando los supuestos de normalidad, linealidad, homocedasticidad e independencia.
5. **Interpretación del modelo.** - Se refiere a las implicaciones con respecto a la variable de respuesta, es decir conlleva a una explicación detallada de los parámetros del modelo para comprobar si se cumplen los criterios: estadístico, lógico y sustantivo.

2.7. Modelo Lineal Generalizado

Con respecto al Modelo lineal generalizado (Lopez, 2004) menciona que este tipo de modelo se utiliza para datos independientes son de una distribución de la familia exponencial (binomial, Poisson, gamma entre otras). Modela $E[y]$ como una función no lineal de $X\beta$. Así mismo (Cayuela, 2010) señala que una alternativa a la transformación de la variable respuesta y a la falta de normalidad es el uso de los modelos lineales generalizados (GLM *Generalized Linear Models*) son una extensión de los modelos lineales que permiten utilizar distribuciones no normales de los errores y varianzas no constantes. Se considera a utilizar estos modelos cuando las variables de respuesta son: un conteo de casos, un conteo de casos expresados como porciones y una respuesta binaria.

Construcción y evaluación de un GLM

En la construcción de modelos lineales generalizados es importante tener en cuenta una cosa: no existe un único modelo que sea válido. Este es uno de los errores más comunes implícitos en el uso de regresión o ANOVA, en donde el mismo modelo se usa una y otra vez, muchas veces sin una perspectiva crítica. En la mayoría de los casos, habrá un número variable de modelos plausibles que puedan ajustarse a un conjunto determinado de datos. Parte del trabajo de construcción y evaluación del modelo es determinar cuál de todos estos modelos son adecuados, y entre todos los modelos adecuados, cual es el que explica la mayor proporción de la varianza sujeta a la restricción de que todos los parámetros del modelo deberían ser estadísticamente significativos. (Cayuela , 2010).

Los pasos que hay que seguir en la construcción y evaluación de un GLM son muy similares a los de cualquier modelo estadístico. Se detalla a continuación:

- 1. Exploración de los datos.** Conviene conocer nuestros datos. Puede resultar interesante obtener gráficos que nos muestren la relación entre la variable respuesta y cada una de las variables explicativas, o matrices de correlación entre las variables explicativas. El objetivo es: a) Buscar posibles relaciones de la variable respuesta con la(s) variable(s) explicativa(s); b) Considerar la necesidad de aplicar transformaciones de las variables; c) Eliminar variables explicativas que estén altamente correlacionadas.
- 2. Elección de la estructura de errores en función de vínculo.** Resulta en elegir las propiedades del modelo, analizar los residuos para ver su idoneidad en la distribución de errores elegida.
- 3. Ajuste del modelo a los datos.** Hace referencia a: a) Los tests de significación para los estimadores del modelo; b) La cantidad de varianza explicada por el modelo. En los modelos de GLM se determina la devianza D^2 . La devianza da la idea de variabilidad explicada por el modelo, se compara la devianza del modelo nulo (*Null deviance*) con la devianza residual (*Residual deviance*) lo que es lo mismo:

$$D^2 = \frac{\text{Devianza modelo nulo} - \text{Devianza residual}}{\text{Devianza modelo nulo}} * 100$$

- 4. Análisis de los residuos.** Son las diferencias entre los valores estimados por el modelo y los valores observados, sin embargo, muchas veces se utilizan los residuos estandarizados que tienen que seguir una distribución normal. (Cayuela , 2010)

Criterios de evaluación del modelo

El criterio que comúnmente se utiliza para la evaluación de este tipo de modelos es el criterio de Información Akaike (AIC es *Akaike Information Criterion*), es un índice que evalúa tanto el ajuste del modelo a los datos como la complejidad del modelo. Cuanto más pequeño es AIC mejor es el ajuste. El AIC es muy útil para comparar modelos similares con distintos grados de complejidad o modelos iguales. Las funciones `stepAIC()`, `addterm()` y `glmsAIC()` muestran los grados de complejidad en función del AIC. Todo esto se aplica igualmente a los modelos lineales (Cayuela ,2010).

Descripción de la Función glm() en R

La función `glm()` viene especificada por los siguientes argumentos

`>args(glm)`

```
function(formula,family=Gaussian,data,weights,subset,na.action,start=NULL,etast
art,mustart,offset,control model=TRUE, method="glm.fit",x=FALSE, y=TRUE,
contrasts= NULL,...)
```

NULL

Es una fórmula que especifica el modelo siguiendo la lógica de los modelos lineales especificados si se hace referencia a estos modelos por la función `lm()` y `family` es la familia de errores de distribución, especificada de la siguiente forma:

- binomial (link= "logit")
- gaussian (link= "identity")
- gamma (link= "inverse")
- inverse.gaussian (link= "1/mu ^2")
- poisson (link= "log")

- quasi (link= "identity", variance="constant")
- quasibinominal (link= "logit")
- quasipoisson (link= "logit")

Tipos de Modelos Lineales Generalizados

En la tabla se puede visualizar los principales modelos lineales generalizados

Tabla 2.7-1 Modelos Lineales Generalizados

Naturaleza de la variable de Respuesta	Componente		Función de enlace	Modelo Lineal	Siglas
	Sistemático	Aleatorio			
Numérica cuantitativa	<ul style="list-style-type: none"> • Numérico • Categórico • Mixto 	<ul style="list-style-type: none"> • Normal • Normal • Normal 	<ul style="list-style-type: none"> • Identidad • Identidad • Identidad 	<ul style="list-style-type: none"> • Regresión Lineal • ANOVA o de diseño experimental • ANCOVA o de diseño experimental con variables concomitantes 	ML
Categórica binaria	<ul style="list-style-type: none"> • Mixto • Categórico 	<ul style="list-style-type: none"> • Binomial (1) • Bernoulli • Binomial (n) 	<ul style="list-style-type: none"> • Logit • Logit • Probit 	<ul style="list-style-type: none"> • Regresión Logística • Análisis Logit • Regresión probit 	GLM

Categórica politómica • No agrupada • Agrupada (frecuencias)	• Mixto • Categórico	• Multinomial	• Logit Generalizado	• Regresión Logística multinomial • Análisis logit multinomial
Recuento	• Mixto	• Poisson	• Logarítmica	• Regresión de Poisson
Frecuencia	• Categórico	• Poisson	• Logarítmica	• Análisis Loglineal

Fuente: Adaptado de López González y Ruíz Soler, (2011)

2.8. Modelo Binomial

Según Cayuela (2010) menciona con respecto a este modelo muchas variables de respuesta:

- Vivo o muerto
- Hombre o mujer
- Infectado o saludable
- Ocupado o vacío

Además, señala que variables están relacionados con la asignación de un individuo a una clase u otra mediante modelos GLM con una distribución de errores de tipo binaria, siempre y cuando exista al menos una variable explicativa que sea continua. La variable respuesta debe de contener solo 0 o 1, en donde un 0 representaría por ejemplo a un individuo muerto y un 1 a un individuo vivo. La manera en la que R trata datos binarios es asumiendo que los 0 y los 1 provienen de una distribución binomial de tamaño 1. Si la probabilidad de que un individuo este muerto es p , entonces la probabilidad de obtener vendría dada por la forma abreviada de la distribución binomial con $n = 1$, conocida como la distribución de Bernoulli:

$$P(y) = p^y \cdot (1 - p)^{1-y}$$

El objetivo en este caso sería determinar cómo las variables explicativas influyen el valor de p . Para ajustar un modelo binomial en R hay que usar la función `glm()` especificando el argumento `family=binomial` (Cayuela,2010).

En muchos de los casos las respuestas tienen solo dos categorías de tipo si/no de modo que se puede definir una variable Y que tome dos posibles valores 1 (éxito) y 0 (fracaso) es decir $Y \sim Bin(1, \pi)$ (Cayuela,2010).

En este caso:

$$f(y|\pi) = \pi^y(1 - \pi)^{1-y}$$

$$f(y|\pi) = (1 - \pi) \left(\frac{\pi}{1 - \pi} \right)^y$$

$$f(y|\pi) = (1 - \pi) \exp \left[\left(\frac{\pi}{1 - \pi} \right)^y \right]$$

Con $y=0.1$

El parámetro natural es

$$Q(\pi) = \log \left(\frac{\pi}{1 - \pi} \right) = \text{logit}(\pi)$$

En este caso

$$E(Y) = P(Y = 1) = \pi(x)$$

Dependiente de p variables explicativas $x = (x_1, \dots, x_p)$ y

$$\text{Var}(Y) = \pi(x)(1 - \pi(x))$$

En respuestas binarias, un modelo análogo al de regresión lineal es:

$$\pi(x) = \alpha + \beta x$$

Que se denomina modelo de probabilidad lineal, ya que la probabilidad de éxito cambia linealmente con respecto a x .

El parámetro β representa el cambio en probabilidad por unidad de x . Este modelo es un GLM con un componente aleatorio binomial y con función link igual a la identidad. Sin embargo, este modelo puede predecir a veces valores $\pi(x) > 1$ y $\pi(x) < 0$ (Cayuela, 2010)

2.8.1. Regresión Logística Multinomial

La regresión logística multinomial es utilizada en modelos con variable dependiente de tipo nominal con más de dos categorías (politómica) y es una extensión multivariante de la regresión logística binaria clásica. Las variables independientes pueden ser tanto continuas (covariables) como categóricas o mayormente factores. (Gonzalez, 2015)

Tradicionalmente las variables dependientes politómicas han sido modeladas mediante análisis discriminante, pero, gracias al creciente desarrollo de las técnicas de cálculo, cada vez es más habitual el uso de modelos de regresión logística multinomial (Gómez & Palacios, 2013), implementados en paquetes estadísticos, debido a la mejor interpretabilidad de los resultados que proporciona. La regresión logística multinomial es utilizada en modelos con variable dependiente de tipo nominal con más de dos categorías (politómica) y es una extensión multivariante de la regresión logística binaria clásica. Las variables independientes pueden ser tanto continuas (covariables) como categóricas (factores). (Roque Cruz , 2018)

Formulación del Modelo

Consideramos una variable aleatoria dependiente Y categórica nominal politómica con soporte $(Y) = \{1,2,3\}$ y con probabilidades $p_1 = p(Y = 1)$, $p_2 = p(Y = 2)$ y $p_3 = p(Y = 3) = 1 - p_1 - p_2$. Se redefine la variable Y mediante vector (Y_1, Y_2) construido de la siguiente forma:

$$(Y_1, Y_2) = \begin{cases} (1,0) & \text{si } Y = 1 \\ (0,1) & \text{si } Y = 2 \\ (0,0) & \text{si } Y = 3 \end{cases}$$

Las variables Y_1 e Y_2 tienen una distribución de Bernouilli con $E(Y_1) = p_1$ y $E(Y_2) = p_2$, al igual que la variable dependiente en una regresión logística binaria clásica. Estas dos variables no son independientes ya que $Cov(Y_1, Y_2) = -p_1p_2$.

El modelo multivariante queda definido por las siguientes ecuaciones:

$$\left. \begin{aligned} p_1(X_1, X_2) = p_1 = E(Y_1) &= \frac{\exp(Z_1)}{1 + \exp(Z_1) + \exp(Z_2)} \\ p_2(X_1, X_2) = p_2 = E(Y_2) &= \frac{\exp(Z_2)}{1 + \exp(Z_1) + \exp(Z_2)} \end{aligned} \right\}$$

Donde $Z_1 = \beta_{01} + \beta_{11} * X_1 + \beta_{21} * X_2$ y $Z_2 = \beta_{02} + \beta_{12} * X_1 + \beta_{22} * X_2$

Siendo $\beta_{01}, \beta_{11}, \beta_{21}, \beta_{02}, \beta_{12}, \beta_{22}$ parámetros que deseamos estimar.

Al cociente p_1/p_2 se le denomina “odds” de la categoría 1 respecto de la categoría 1 respecto de la categoría 3 y se lo representa por $O_1(X_1, X_2) = O_1(\text{idem. para } O_2)$.

De este modo puede observarse fácilmente que la razón de cambio en O_1 cuando X_1 se incrementa en unidad manteniéndose constante

X_2 viene dada por $\frac{O_1(X_1+1, X_2)}{O_1(X_1, X_2)} = \exp(\beta_{11})$, que recibe el nombre de “odds-ratio” de

la categoría 1 respecto de la variable X_1 y se representa por

$OR_1(X_1)$ (idem, para $OR_1(X_2), OR_2(X_1)$ y $OR_2(X_2)$).

Es interesante observar que estas “odds-ratio” dependen de las unidades en que vengan medidas las variables regresoras (si multiplicamos X_1 por 10, $OR_1(X_1)$ pasaría a ser $\sqrt[10]{\exp(\beta_{11})}$). Por tanto, la importancia de cada variable regresora en el modelo debería medirse por el valor de la odds-ratio suponiendo estandarizadas en las variables regresoras. Cuando más grande sea este valor más relevante es la variable dentro del modelo. También interesa definir las proporciones de cambio en las “odds” con respecto a cada variable regresora, viene dada por:

$$\frac{O_1(X_1 + 1, X_2) - O_1(X_1, X_2)}{O_1(X_1, X_2)} = OR_1(X_1) - 1 = \exp(\beta_{11}) - 1$$

Otra formulación alternativa, y quizás más conocida, se obtiene tomando logaritmos en ambas ecuaciones del modelo:

$$\left. \begin{aligned} \ln\left(\frac{p_1}{p_3}\right) &= Z_1 = \beta_{01} + \beta_{11} * X_1 + \beta_{21} * X_2 \\ \ln\left(\frac{p_2}{p_3}\right) &= Z_2 = \beta_{02} + \beta_{12} * X_1 + \beta_{22} * X_2 \end{aligned} \right\}$$

Donde las expresiones del miembro izquierdo se denominan “logits” (al igual que en la regresión logística binaria) y los parámetros representan las tasas de cambio en los “logits” cuando una de las variables explicativas se incrementa en una unidad manteniéndose constante la otra. (Roque Cruz , 2018)

Estimación de parámetros

Dada una muestra de datos $(Y_{1i}, Y_{2i}, X_{1i}, X_{2i})$ con $i = 1, 2, \dots, n$ podemos definir, en función de los parámetros del modelo, las funciones $Z_{1i}, Z_{2i}, P_{1i}, P_{2i}$ y abordar el problema de la estimación de los mismos mediante el método de máxima verosimilitud como se muestra a continuación:

Con el modelo planteado, la función de verosimilitud de la muestra viene dada la siguiente expresión:

$$L = \prod_{i=1}^n (p_{1i}^{Y_{1i}} * p_{2i}^{Y_{2i}} * p_{3i}^{1-Y_{1i}-Y_{2i}}) = \prod_{i=1}^n \left(\left(\frac{p_{1i}}{p_{3i}}\right)^{Y_{1i}} * \left(\frac{p_{2i}}{p_{3i}}\right)^{Y_{2i}} * p_{3i} \right)$$

En vez de trabajar con esta expresión se utiliza la función auxiliar:

$$\begin{aligned} \Lambda &= -2 * \ln(L) \\ &= -2 \\ &* \sum_{i=1}^n \left(Y_{1i} * \ln\left(\frac{p_{1i}}{p_{3i}}\right) + Y_{2i} * \ln\left(\frac{p_{2i}}{p_{3i}}\right) + \ln(p_{3i}) \right) \\ &= 2 * \sum_{i=1}^n (\ln(1 + \exp(Z_{1i}) + \exp(Z_{2i})) - Y_{1i} * Z_{1i} - Y_{2i} * Z_{2i}) \end{aligned}$$

El problema de maximizar la verosimilitud equivale al de minimizar la función auxiliar Λ y puede resolverse por métodos numéricos de forma iterativa partiendo de la estimación inicial $\beta_{11} = \beta_{21} = \beta_{12} = \beta_{22} = 0$, $\beta_{01} = \ln(n_1) - \ln(n - n_1 - n_2)$ y $\beta_{02} = \ln(n_2) - \ln(n - n_1 - n_2)$ siendo n_1 y n_2 el número de observaciones en las categorías 1 y 2 respectivamente. Estos estimadores iniciales se obtienen

suponiendo que no hay una influencia de las variables regresoras en el modelo planteado y para ellos el valor inicial de la función auxiliar que debemos de minimizar es:

$$\Lambda_0 = -2 * \left(n_1 * \ln\left(\frac{n_1}{n}\right) + n_2 * \ln\left(\frac{n_2}{n}\right) + (n - n_1 - n_2) * \ln\left(\frac{n - n_1 - n_2}{n}\right) \right)$$

Una vez alcanzada la convergencia del método iterativo, designaremos por Λ_0 al mínimo obtenido y por $\widehat{\beta}_{01}, \widehat{\beta}_{11}, \widehat{\beta}_{21}, \widehat{\beta}_{02}, \widehat{\beta}_{12}, \widehat{\beta}_{22}$ a los valores estimados de los parámetros del modelo. (Roque Cruz , 2018)

Significatividad global del modelo

Podemos contrastar la hipótesis de no existencia de un efecto significativo global de las variables regresoras teniendo en cuenta que la diferencia entre el valor inicial y el valor final de la función auxiliar Λ tiene una distribución χ^2 con 4 grados de libertad. El p-valor de la prueba para la hipótesis nula de que no existe efecto de las variables regresoras ($\beta_{11} = \beta_{21} = \beta_{12} = \beta_{22} = 0$) vendrá dado por $p(X_4^2 > \Lambda_0 - \Lambda_f)$

Significatividad del efecto de cada variable regresora

Consiste en llamar Λ_{-1} al mínimo de la función auxiliar que se obtendrá eliminando del modelo la variable X_1 ($\beta_{11} = \beta_{12} = 0$) se verifica que la diferencia entre los mínimos de la función auxiliar en el modelo reducido y en el modelo completo tiene una distribución χ^2 con 2 grados de libertad. Por tanto, el p-valor de la prueba para la hipótesis nula de que no existe efecto de la variable X_1 ($\beta_{11} = \beta_{12} = 0$) vendrá dado por $p(\chi_2^2 > \Lambda_{-1} - \Lambda_0)$. De modo similar podríamos calcular:

Λ_0 (mínimo de la función auxiliar eliminando β_{01} y β_{02} del modelo) y

Λ_{-2} (mínimo de la función auxiliar eliminando del modelo la variable x_2)

y construir tests de hipótesis para $\beta_{01} = \beta_{02} = 0$ y $\beta_{21} = \beta_{22} = 0$ respectivamente. (Roque Cruz , 2018)

Significatividad de cada parámetro

Se tiene en cuenta que el cuadrado de cada estimador dividido por su error estándar tiene una distribución χ^2 con 1 grado de libertad se construye la prueba de hipótesis

para la igualdad de cada parámetro a cero y se determinan los estimadores de los parámetros del modelo son significativamente distintos de cero. La prueba de hipótesis $\beta_{11} = 0$ el p-valor sería $p\left(x_1^2 > \left(\frac{\beta_{11}}{s.e.(\widehat{\beta}_{11})}\right)^2\right)$, siendo $s.e.(\widehat{\beta}_{11})$ el valor correspondiente al error estándar del estimador del parámetro β_{11} . (Roque Cruz , 2018)

2.9. Modelo Poisson

El modelo de regresión de Poisson es un tipo de Modelo Lineal Generalizado que se utiliza para datos de conteo, este modelo es adecuado cuando la varianza muestral es igual a la media. En cuanto a este modelo la media (λ) se explica en términos de las variables explicativas mediante el uso de un enlace. (García, 2014) Por lo que se refiere al modelo de regresión de Poisson surge cuando la variable respuesta es una cantidad discreta que se puede modelizar con una Poisson y se quiere estudiar si ciertas variables explicativas influyen en la variable respuesta y cómo lo hacen. Este tipo de variable respuesta suele representar el recuento de sucesos o hechos. Por tanto, se considera una variable respuesta Y que toma los valores en $\{0, 1, 2, \dots\}$, y se va a estudiar su relación con otras variables explicativas X mediante un análisis de regresión. Se pretende construir un modelo para $\lambda(x) = E(Y|X = x)$ es decir, para la media de Y condicionada a cada valor de la variable explicativa. Como Y nunca toma valores negativos, no procede utilizar un modelo lineal directo, y por tanto se necesita una función de enlace previa a cualquier modelo lineal. Además, la función de regresión está en el intervalo $(0, +\infty)$, el logaritmo parece la función de enlace más adecuada. Se expresaría de la siguiente forma: $g(\lambda(x, \beta)) = x'\beta$ donde lo más habitual es tomar $g(r) = \log(r), r \in (0, +\infty)$ Donde $x'\beta$ representa el producto escalar del vector de variables explicativas por el vector de parámetros. Para incluir un intercepto se considera una primera variable explicativa igual a 1.

Dicho de otra manera, la función de regresión del modelo Poisson se expresa por $\lambda(x, \beta) = e^{x'\beta}$

Los parámetros de este modelo, que se puede denominar modelo log-lineal, se interpretan de la siguiente forma: la exponencial del intercepto es el valor esperado de la respuesta en la categoría de referencia o cuando las variables explicativas numéricas valen cero, y las exponenciales de los coeficientes de cada variable representan tasas de incremento de la respuesta esperada al aumentar una unidad la variable si es numérica, o al pasar a la categoría correspondiente si es cualitativa. Es decir, el modelo supone efectos multiplicativos: si la componente explicativa unidimensional X_j aumenta n unidades, la media para la variable de Poisson se multiplica por la potencia n -ésima de e^{β_j} , supuestas las demás variables explicativas constantes, matemáticamente

$$\frac{\lambda((x_1, \dots, x_j + n, \dots, x_p), \beta)}{\lambda((x_1, \dots, x_j, \dots, x_p), \beta)} = \frac{e^{\beta_0 + x_1\beta_1 + \dots + (x_j+n)\beta_j + \dots}}{e^{\beta_0 + x_1\beta_1 + \dots + x_j\beta_j + \dots}} = e^{n\beta_j} = (e^{\beta_j})^n$$

Si se tiene una muestra aleatoria simple

$(X_1, Y_1), \dots, (X_n, Y_n)$ de (X, Y) , entonces $Y_i \in \text{Poisson}(\lambda(X_i, \beta))$ siendo $\lambda(x, \beta) = e^{x'\beta}$

Para estimar los parámetros del modelo se utiliza la función de máxima verosimilitud, que adopta la siguiente forma:

$$L(\beta) = \prod_{i=1}^n \left[e^{-\lambda(x_i, \beta)} \frac{\lambda(x_i, \beta)^{y_i}}{y_i!} \right], \text{ siendo su logaritmo (en términos dependientes de } \beta)$$

$$L(\beta) = \sum_{i=1}^n (y_i x_i' \beta - e^{x_i' \beta})$$

Derivando dicha función e igualando a cero, se obtienen las ecuaciones de verosimilitud:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^n x_i' [y_i - \lambda(x_i, \beta)] = 0,$$

La matriz hessiana adopta la forma:

$$\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta^t} = - \sum_{i=1}^n x_i x_i' \lambda(x_i, \beta)$$

Igual que ocurre en la regresión logística, las ecuaciones de verosimilitud no son lineales en los parámetros y es necesario recurrir a procedimientos iterativos para el cálculo de sus estimaciones. Newton-Raphson y el IRLS (*Iterative Re-weighted Least Squares*) son los métodos iterativos que se aplican en este modelo. Además, es conocido que los estimadores de máxima verosimilitud son asintóticamente normales y centrados y su matriz de varianzas-covarianzas es la inversa de la matriz de información (la matriz hessiana cambiada de signo), lo cual permite hacer inferencias sobre los parámetros del modelo.

2.9.1. Distribución de Poisson

Sea Y una variable aleatoria discreta, se dice que Y sigue una distribución de Poisson con parámetro λ si su función de probabilidad viene dada por $f(y) =$

$$P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!}, \text{ para } y = 0, 1, 2, \dots, \text{ donde:}$$

y : representa el número de veces que ocurre un evento de interés

λ : representa al parámetro positivo que representa el número de veces que se espera que ocurra el evento en un período determinado

2.9.2. Propiedades de la Distribución de Poisson

Con respecto a las propiedades de la presente distribución son las siguientes:

1. La media y la varianza son iguales a su parámetro $E(Y) = \lambda = Var(Y)$
2. Si λ crece, la masa de la distribución se desplaza hacia la derecha y $P(Y = 0)$ decrece
3. A medida que λ crece, la distribución de Poisson se aproxima a una distribución normal por el TCL.(García, 2014)

CAPÍTULO 3

3. METODOLOGÍA

3.1. Diseño de la investigación

3.1.1. Tipo de investigación

El diseño de la investigación tuvo un enfoque cuantitativo puesto que responde a la estadística como característica principal en la elaboración de modelos. Se empezó con el análisis de modelos lineales para culminar con la elaboración de modelos lineales generalizados. Con un proceso basado en lo deductivo que analizó la realidad objetiva de las defunciones fetales del litoral cuyas bondades se reflejaron en la generalización de los modelos. Hay que mencionar además que es de tipo observacional descriptiva, debido que el INEC considera que las operaciones estadísticas son basadas en registros administrativos que acceden el aprovechamiento de datos para fines comparables, estandarizados e integrables y analíticos. Por otro lado, el examinar el comportamiento de los datos con la utilización de la estadística analítica hace el uso de modelos generalizados a partir de modelos de regresión lineal. Con el fin de dar respuesta al cumplimiento del primer objetivo específico, se utilizó un marco estadístico donde se argumentó teóricamente los modelos de regresión lineal y lineales generalizados de la investigación objeto de estudio. En lo que respecta al segundo objetivo específico se identificó las variables a partir de un análisis y formulación de hipótesis correlacionales que establecen la relación de variables objeto de estudio. (Hernández Sampieri, Fernández Collado, & Baptista , 2014) . Finalmente, para dar respuesta al tercer objetivo específico se desarrolló modelos de regresión lineal simple, y múltiple que son base para la elaboración de modelos lineales generalizados. Se realizó un análisis de correlación de las variables objeto de estudio para descubrir la asociatividad lineal entre ellas a partir de gráficos de dispersión y análisis de correlación propiamente dicho. Así mismo se construyeron modelos lineales generalizados dentro de los cuales se encuentran: el modelo binomial a partir de las causas que originan las defunciones fetales de respuestas

binaria (regresión logística multinomial) y modelo Poisson (Cayuela , 2010), se utilizó generalmente para representar datos de conteos, es decir, datos enteros positivos.

3.1.2. Métodos de investigación

La presente investigación inició con un método inductivo a partir del análisis de riesgo de la mortalidad fetal a partir de la derivación inductiva del problema desde proposiciones particulares hacia una conclusión general mediante contrastación a partir de datos proporcionados de INEC 2017. Con respecto a su enfoque fue de tipo cuantitativo de acuerdo con la obtención de una base secundaria de datos de acuerdo con las variables estudiadas de carácter descriptivo, y analítico que permitió estimar la media de defunciones fetales objeto de estudio.

3.1.3. Técnicas de investigación

La lectura científica se utilizó para la confección del marco teórico, de esta manera se obtuvo la información necesaria para comprender el comportamiento de los modelos lineales y lineales generalizados objeto de estudio de la presente indagación para ello se revisó artículos científicos, bibliografía web y libros que ayudaron de sustento en el desarrollo del estudio.

3.1.4. Datos

La fuente de los datos que se utilizó para la aplicación de los modelos lineales y lineales generalizados es la base de datos de defunciones fetales del Instituto Nacional de Estadísticas y Censos del año 2017. La presente fuente de datos cuenta con la siguiente información: edad, región, causas, muertes fetales por región y nacidos por región. Además se consideró la Región Costa como referencia debido a que cuenta con mayor número de defunciones fetales es así que de 1950 datos analizados en los modelos 1122 se concentran en la Región Costa es decir el 58%, además la causa que mayor índice origina las defunciones fetales es hipoxia intrauterina considerada en el modelo de regresión logística multinomial, se consideró la variable enfermedad categórica nominal politómica con soporte P1: Hipoxia intrauterina con respecto a P2: Muerte del bebe en el vientre, P3: Otras causas para la elaboración del modelo.

3.1.5 Modelos Lineales

Se procedió a utilizar Rstudio en donde se generó el modelo de regresión lineal, se devuelve junto con el valor de la pendiente y la ordenada en el origen el valor del estadístico t obtenido para cada uno y los p-value correspondientes. Esto permitió saber, además de la estimación de β_0 y β_1 , si son significativamente distintos de 0.

Para la realización del modelo de regresión línea en RStudio se procedió a instalar las siguientes librerías: Library (tidyverse, library(boot), library(car), library(QuantPsyc), library(ggplot2).

Modelos Multinomiales

Se procedió a determinar los modelos multinomiales en Rstudio a partir de la función glm y la family=binomial (link = logit). Se utilizó la librería library(nnet) para plantear el modelo que determine la probabilidad de que un feto muera por una de tres causas previamente clasificadas.

Modelos Poisson

Este tipo de regresión se usa para modelar datos de conteo (UCLA, 2020). Según Morales (2018) este tipo de modelo son frecuentes los datos discretos que provienen de conteos de sucesos que se producen por azar con cierta frecuencia y son modelizables en términos de tasas de incidencia que dependen de ciertas variables explicativas. Para modelizar este tipo de datos se utiliza la distribución de Poisson, $X \sim \text{Po}(\mu)$, donde μ representa el número medio de ocurrencias, de forma que:

$$E(X)=\mu \text{ y } V(X)=\mu$$

El parámetro μ requiere una definición cuidadosa. A menudo es necesario describirlo como una tasa; de manera más general, la tasa se especifica en términos de unidades de “exposición” o en términos de años-persona ‘en riesgo’, en nuestro caso, el número de defunciones fetales para un instante de tiempo

determinado. El efecto de las variables predictoras sobre la respuesta se modeliza valorando su efecto sobre μ . Sea Y_1, Y_2, \dots, Y_n un conjunto de variables aleatorias Poisson, donde Y_i representa el número de eventos observados a partir de la exposición n_i para el i -ésimo patrón de covariables. En esta situación el valor esperado de Y_i se puede escribir como:

$$E(Y_i) = \mu_i = n_i \lambda_i$$

Donde λ_i representa la tasa de incidencia para el patrón i -ésimo. Hay que mencionar además que el modelo lineal generalizado para variables aleatorias Poisson, más conocido como regresión de Poisson se puede especificar siguiendo la definición establecida en el punto anterior como:

La distribución de cada variable aleatoria Y_i viene dada por

$$Y_i \sim Po(\mu_i)$$

El valor medio de cada variable viene dado por:

$$E(Y_i) = \mu_i = n_i \lambda_i$$

Si consideramos un conjunto $x_{i1}, x_{i2}, \dots, x_{ip}$ de p variables predictoras que pueden afectar el comportamiento de Y_i asumimos que:

$$\lambda_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$$

Donde cada β representa el efecto de la correspondiente variable predictora.

En esta situación y tomando como función de enlace el logaritmo tenemos que:

$$\log(E(Y_i)) = \log(\mu_i) = \log(n_i \lambda_i) = \log(n_i) + \log(\lambda_i) = \log(n_i) + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

Donde tenemos la expresión del modelo lineal generalizado. Respecto de la interpretación de dichos coeficientes se puede demostrar que $\exp(\beta_j)$ representa el riesgo relativo (RR) sobre la tasa de incidencia de los sucesos asociado a un incremento de una unidad en la covariable x_j . Es similar a la interpretación que hacíamos en los modelos lineales, salvo el cambio de escala debido a la transformación logaritmo sobre la media de la respuesta. Para una variable de tipo numérico se define el riesgo relativo al incrementar una unidad el valor de la variable predictora X como:

$$RR(X) = \frac{\lambda(X=x+1)}{\lambda(X=x)} = \exp(\beta x)$$

Por otro lado el riesgo relativo para valorar el cambio entre dos niveles de una variable categórica (A y B) se define como:

$$RR(X_{AB}) = \frac{\lambda(X=B)}{\lambda(X=A)} \exp(\text{coeficiente asociado a B} - \text{coeficiente asociado a A})$$

El riesgo relativo aumenta cuando $RR > 1$ y disminuye cuando $RR < 1$. Tenemos el mismo riesgo cuando $RR = 1$ (Morales, 2018).

Distribución de Poisson

Llamada así en honor a Simeón-Denis Poisson, quien la describió por primera vez a finales del siglo XIX Dentro de su trabajo: *Recherches sur la probabilité des jugements en matière criminelle et matière civile* (Investigación sobre la probabilidad de los juicios en materias criminales y civiles) (Zylberberg , 2005).

Arroyo et al (2014) señalan que la distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media λ , la probabilidad que ocurra un determinado número de eventos $k \in x$ durante un intervalo de tiempo dado o una región específica. Sea x una variable aleatoria que representa el número de eventos aleatorios independientes que ocurren a una rapidez constante sobre el tiempo o el espacio. Se dice entonces que la variable aleatoria x tiene una distribución de Poisson con función de probabilidad.

$$p(k, \lambda) = f(k) = \begin{cases} \frac{e^{-\lambda} \lambda^k}{k!}, & \text{si } k = 0, 1, 2 \dots; \lambda > 0 \\ 0, & \text{de otra manera} \end{cases}$$

Propiedades de la distribución de Poisson

La distribución de Poisson desempeña un papel importante, por derecho propio, como un modelo probabilístico apropiado para un gran número de fenómenos aleatorios. Las características mas sobresalientes de esta distribución son:

La distribución de Poisson tiene la particularidad de que la esperanza $E(X)$ y la varianza $\text{Var}(X)$ son iguales, esto es:

$$E(x) = \lambda$$

$$Var(x) = \lambda$$

Los factores de forma de la distribución de Poisson son:

$$\text{Coeficiente de asimetría: } \frac{1}{\sqrt{\lambda}}$$

$$\text{Curtosis relativa : } 3 + \frac{1}{\lambda}$$

Con lo anterior se puede visualizar que la distribución de Poisson es leptocúrtica con un sesgo positivo.

La función generadora de momentos de la variable aleatoria de Poisson X , con valor esperado λ , es $m_x(t) = e^{\lambda(e^t - 1)}$

El espacio muestral en un modelo de Poisson se genera por un número muy grande (puede considerarse infinito) de repeticiones de un experimento cuyo modelo de probabilidad es el de *Bernoulli*, con probabilidad de éxito muy pequeña. Por esta razón, a la distribución de Poisson se le suele llamar de eventos raros. Las repeticiones del experimento de Bernoulli se realizan en cada uno de los puntos de un intervalo de tiempo o espacio. La probabilidad de que se tenga dos o más éxitos en el mismo punto del intervalo es cero. El número promedio de éxitos en un intervalo es una constante λ , que no cambia de intervalo a intervalo.

La distribución de Poisson se puede expresar de forma gráfica, ya que en realidad consiste en un diagrama de barras, con forma asimétrica positiva como sucede con la distribución binomial. Sin embargo, al ir aumentando los valores de λ , va adquiriendo la típica forma de la Campana de Gauss (esto se puede evidenciar analizando el coeficiente de asimetría su objetivo es determinar, sin necesidad de dibujar, la deformación horizontal de la distribución con respecto a un valor central, generalmente la media (García, Solarte, & Imuez, 2009). Pudiendo deducirse, que conforme aumenta λ , las variables de Poisson van a poder aproximarse a la distribución normal, por el Teorema Central del Límite. La aproximación se considera buena para valores de λ iguales o superiores a nueve (Arroyo, Bravo M,

Llinás, & Muñoz, 2014). Se procedió a utilizar Rstudio en donde se generó los modelos de Poisson se especificó `family=poisson(link=log)`.

CAPÍTULO 4

4. RESULTADOS

Antes de definir las variables de estudio se procedió a realizar la división por rangos de edad de la madre de la base de datos proporcionada por el INEC sobre las defunciones fetales. Así como también a partir de un diagrama de Pareto 80 -20 para determinar las causas principales de la muerte de un feto y el resto se las agrupo en otras causas que tienen una ocurrencia mínima. En la primera parte de este análisis se hará una exploración a Ecuador en cuanto a defunciones fetales, para posteriormente resumir lo que sucede en la costa. En los modelos construidos se hará un comparativo de las otras regiones con la costa, por lo que se considerará la variable región siendo la Costa la categoría de referencia.

4.1. División por rangos de edad

Se realizó la división por rangos de edad de la madre de la base de datos proporcionada por el INEC (2017) sobre las defunciones fetales. En la tabla 4.1-1 se visualiza la división de rangos por edad. Sobre salud sexual y reproductiva la fertilidad de la mujer se la puede dividir en 6 etapas la cual inicia a partir de los 14 en el cual ya se encuentra en la mayoría de las jóvenes, la segunda etapa es de los 21 donde son más fértiles las mujeres hasta los 30 años, tercera etapa hasta los 35 comienza a decaer su fertilidad, cuarta etapa a partir de los 37 se complica el que mujeres queden embarazadas, por último la quinta etapa después de tener los 37 años la fertilidad de la mujer decrece aceleradamente hasta los 44 años, en la que las posibilidades biológicas de ser madres son muy bajas.

Tabla 4.1-1 Rango por Categorías

Rango	Categoría
10-17	Adolescencia Temprana-Media
18-20	Adolescencia Tardía
21-24	Joven Adulto

25-29	Adulto
30-35	Adulto
+36	Adulto

En la tabla 4.1-2 con respecto a los rangos se observa que las frecuencias por categoría de estudio en lo que respecta al rango de mayor continuidad en defunciones fetales se encuentra entre edades de 21 a 35 años. Se podría decir que se encuentran entre las etapas dos y tres de las cinco etapas de fertilidad femenina. Según Rangel Calvillo, (2014) considera que la etapa uno es a partir de los 14 años cuando las mujeres han empezado a ovular, la etapa dos que empieza a partir de los 21 años es la etapa más fértil. Por otro lado, en la etapa tres se encuentra en edad de 30 años que ya empieza a decrecer. La etapa cuatro a partir de los 37 años la fertilidad femenina decrece aceleradamente. Por último, la etapa 5 que es de 44 años las probabilidades biológicas de ser madre son muy bajas. De donde resulta que la etapa dos y tres es en donde se concentran mayor número de defunciones fetales debido a que las mujeres tiende a quedarse embarazadas por ser más fértiles, sin embargo, presentan dificultades a los 5 meses y medio en su gestación provocando la muerte del feto en pleno desarrollo, por lo cual este indicador juega un papel importante e incuestionable en lo que respecta en atención prenatal y trabajo obstétrico en un país.

Tabla 4.1-2 Edad de la Madre Agrupada

Rangos	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
10 – 17	162	8.3	8.3	8.3
18 – 20	255	13.1	13.1	21.4
21 – 24	383	19.6	19.6	41.0
25 – 29	426	21.8	21.8	62.9
30 – 35	393	20.2	20.2	83.0
36+	331	17.0	17.0	100.0
Total	1950	100.0	100.0	

4.2. Análisis de Causas de defunciones fetales

En la figura 4.2-1 se visualiza el diagrama de Pareto 80-20 en donde se conoce las causas principales de la muerte de un feto y el resto se las agrupó en otras causas que tienen una incidencia mínima. Se debe agregar que las causas se las agrupó de la siguiente manera:

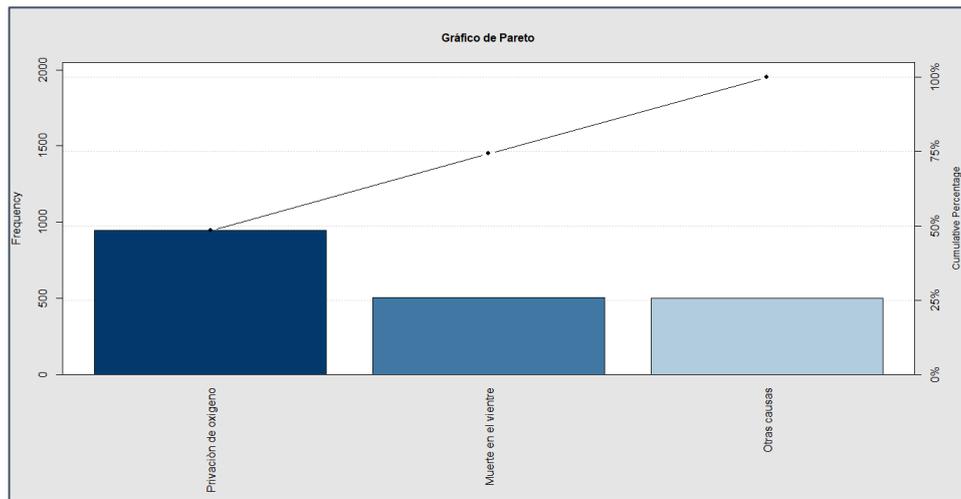
Hipoxia intrauterina no especificada: “Privación de oxígeno en el cuerpo” un proceso patológico, que se caracteriza por una ingesta insuficiente de oxígeno al feto. (OMS, 2018).

Muerte fetal de causa no especificada: “Muerte del bebe en el vientre” es una muerte previa a la expulsión o extracción del feto independientemente de la semana de embarazo y diversas causas. (OMS, 2018)

Otras causas. – Corresponde a defunciones fetales con un menor número de incidencias sobre el total de la base, puede ser defectos congénitos.

Partiendo de la representación del diagrama de Pareto se puede decir que Hipoxia intrauterina es la causa mayor por la cual se originan las defunciones fetales en el Ecuador y en donde se debería centrar mayor cuidado en el ámbito de los centros de salud para de esta manera poder disminuir el índice de defunciones fetales. Mientras que el 20% se encuentra concentrado en otras causas y causas de muerte previa a la expulsión o extracción del feto, considerando que a partir de este tipo de analices el ministerio de salud debería poner atención en estudios de hipoxia intrauterina para poder disminuir las muertes fetales.

Figura 4.2-1 Análisis Causas Fetales a partir del Diagrama de Pareto

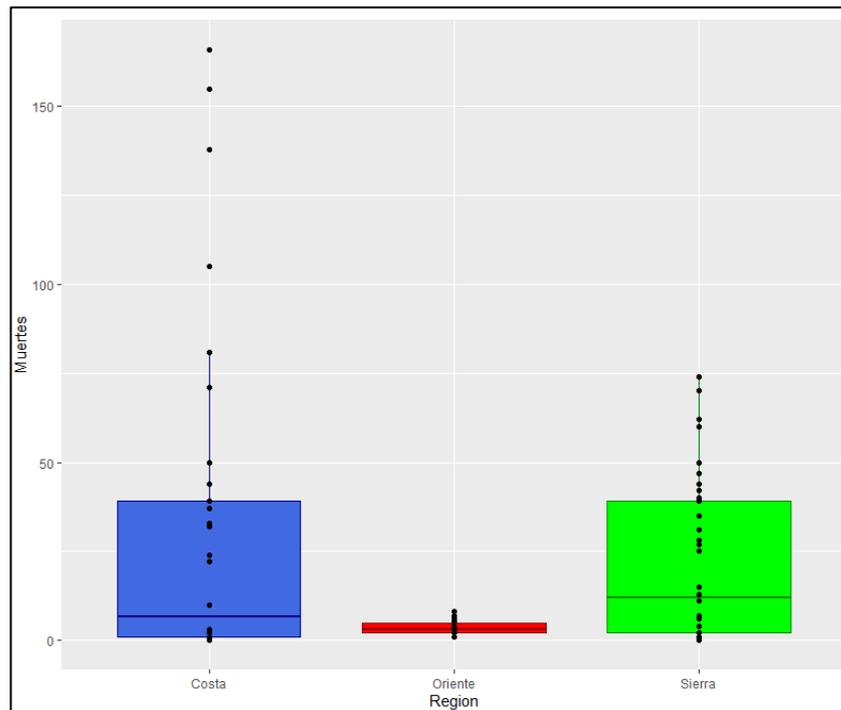


Fuente: Adaptado de datos INEC (2017)

Partición de Defunciones Fetales de acuerdo con la región

Se realizó un análisis de la cantidad de defunciones fetales que se ha obtenido por región durante el año 2017. Siendo la región costa la mayor debido a su densidad poblacional, con respecto a la figura. Con un número de 1122 que en término porcentual es de 57.53% de muertes antes de la expulsión completa del cuerpo de su madre, debido a que en el área rural no reciben atención especializada, de hecho, alcanzan un índice de pobreza del 41.8% de acuerdo los datos del INEC (2019) influyendo en la aglomeración de la región Costa, sin con ello dejar de lado la región Sierra que alcanza un porcentaje de 39.18% de defunciones.

Figura 4.2-2 Comparación de defunciones Fetales por Región



Fuente: Adaptado de datos INEC (2017)

4.3. Análisis de las variables para el Modelo

Para el estudio en lo que respecta a modelos lineales y lineales generalizados se consideraron las siguientes variables:

- Edad
- Región (costa, sierra, oriente)
- Zona (urbano, rural)
- Causas de muertes : (Hipoxia intrauterina no especificada, muerte fetal de causa no especifica, otras Causas)
- Nacidos

Se procedió hacer un análisis de contingencia entre Región y Causa de fallecimiento, respecto a la variable Edad de la madre.

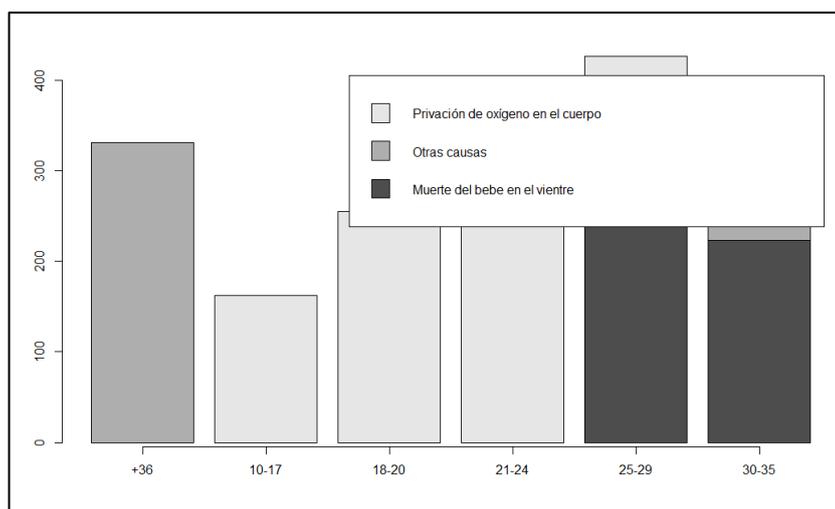
La hipótesis de estudio es:

H0: “La causa fetal, la región de inscripción y la edad de la madre son independientes”

H1: “La causa fetal, la región de inscripción y la edad de la madre no son independientes”

De donde se infiere que con un valor p de 0.00 hay dependencia entre las tres variables.

Figura 4.3-1 Barplot de observaciones por grupos-edad - madre.



Fuente: Adaptado de datos INEC (2017)

Además, se analizó edad y región, la tabla 4.3.1-1 muestra la distribución de frecuencias de estas variables de manera bivariada.

Tabla 4.3-1 Tabla de contingencia Región -área

Edad	Región		
	Costa	Sierra	Oriente
Grupo 1	106	47	9
Grupo 2	153	97	5
Grupo 3	249	120	14
Grupo 4	244	166	16
Grupo 5	216	168	9
Grupo 6	154	166	11

La hipótesis de estudio es:

H0: “La región y la edad de la madre son independientes”

H1: “La región y la edad de la madre no son independientes”

Se aprecia que a mayor edad se presentan más defunciones. En el oriente hay pocos registros de defunciones fetales. La prueba de independencia Ji cuadrado presenta un valor $p=0.000$, por tanto, se rechaza la prueba de independencia entre las variables.

Es importante también entender si hay relación entre área y región. La siguiente tabla muestra los conteos de defunciones fetales considerando la región y el área.

Se resalta el poco registro de defunciones en el área rural, en el oriente no se han registrado defunciones en esa área.

Tabla 4.3-2 Tabla de contingencia Región -área

Región	Rural	urbana
Costa	15	1107
Sierra	49	715
Oriente	0	64

La hipótesis de estudio es:

H0: “La región y la zona son independientes”

H1: “La región y zona no son independientes”

El test exacto de Fisher fue usado para ver si hay relación entre las variables región y área, con un valor $p=0.000$ se concluye que existe evidencia estadística para rechazar la hipótesis nula de independencia, por tanto, hay relación entre estas dos variables.

H0: “La causa de muerte fetal agrupada y la región de inscripción son independientes”

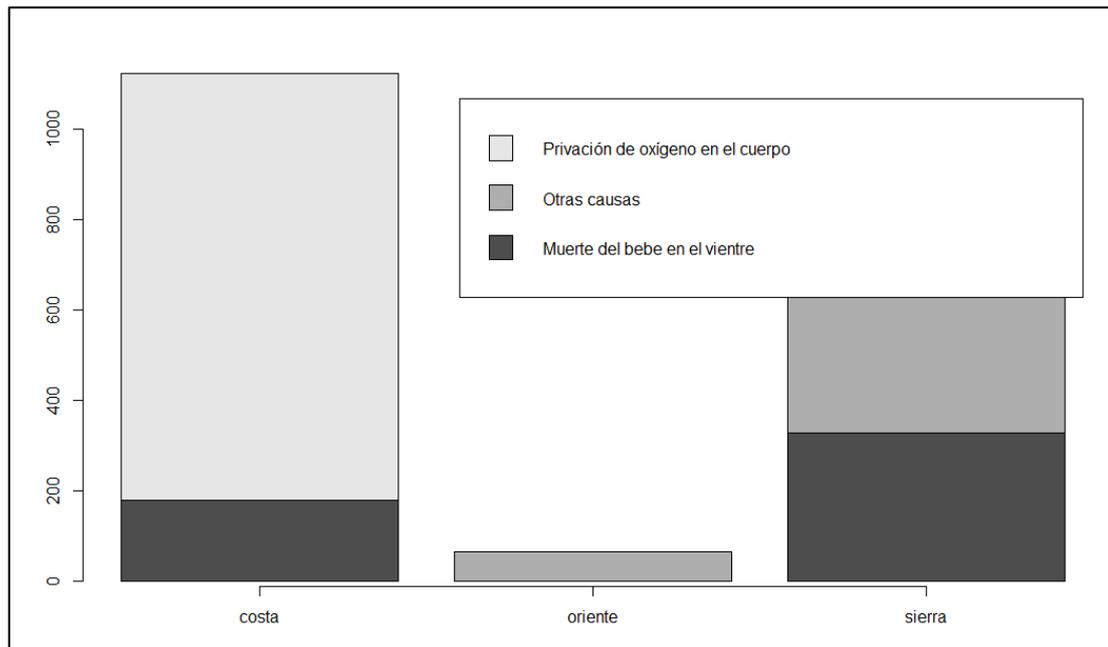
H1: “La causa de muerte fetal agrupada y la región de inscripción no son independientes”

De donde resultó lo siguiente:

Con un valor de Chi-cuadrado de Pearson 276,854 con $gl=4$ y con un valor $p = 0.00$, se concluye que p es menor que el nivel de significancia de 0.05, y se dice que hay evidencia estadísticamente significativa para rechazar H_0 y concluir que entre las

variables causa fetal agrupada y la región de inscripción existe una relación, es decir son dependientes.

Figura 4.3-2 Barplot de frecuencias de las observaciones por Región



Fuente: Adaptado de datos INEC (2017)

Considerando que el tema principal de este proyecto es las defunciones en la costa, se resume de este análisis que la principal causa de muerte en la costa es la privación de oxígeno en el cuerpo, el 98.6% de las muertes están registradas en el área urbana y que cuando la madre tiene una edad comprendida entre los 21 a 35 años, el número de defunciones parece constante y más alto.

4.4. Modelos de regresión lineal

En la primera parte de esta sección se construyeron modelos de regresión lineal simples y se consideró a todo el Ecuador.

H0: “No hay relación lineal entre ambas variables por lo que la pendiente del modelo lineal es cero. $\beta_1 = 0$ ”

H1: “Sí hay relación lineal entre ambas variables por lo que la pendiente del modelo lineal es distinta de cero. $\beta_1 \neq 0$ ”

4.4.1. Modelo uno de regresión lineal: considerando la variable edad

$$Y_i = \beta_0 + \beta_1 X_i$$

La variable Y son las muertes fetales y la variable X es Edad:

$$Y_i = 12.987 + 2.480 * Edad$$

$$H_0: \beta_{Edad} = 0$$

$$H_1: \beta_{Edad} \neq 0$$

H_0 : No existe una relación lineal significativa entre la variable Edad y Muertes

H_1 : Existe una relación lineal significativa entre la variable Edad y Muertes

Figura 4.4.1-1 Coeficientes del Modelo de Regresión lineal- edad

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.987	8.036	1.616	0.110
Edad	2.480	2.064	1.202	0.233

Residual standard error: 33.43 on 88 degrees of freedom
Multiple R-squared: 0.01615, Adjusted R-squared: 0.004968
F-statistic: 1.444 on 1 and 88 DF, p-value: 0.2327

No existe una relación lineal significativa entre la variable Edad y Muertes fetales, pues el valor p es mayor que el nivel de significancia $\alpha=0.05$, ahora al verificar si el modelo es significativo se obtiene que el valor p es 0.2327 por lo tanto se concluye que el modelo no es significativo.

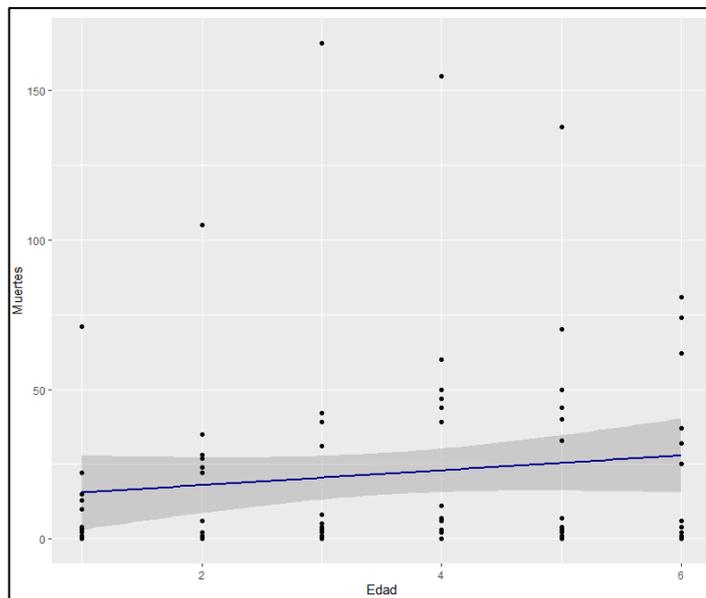
Además, se puede observar que su R^2 no nos proporciona que sea un buen modelo puesto se considera que mientras que R^2 se acerca a uno es mejor en este caso es menor de 1.

Tabla 4.4.1-1 Resumen del Modelo Uno

Modelo Uno	Ecuación	R ²	P valor
Regresión lineal	Muertes fetales = 12.98 + 2.48 Grupos de edad	0.016	0.23

En la figura 4.4.1-2 se observa la pendiente del modelo que se encuentra representado por 2.48 y su intercepto inicia desde 12.98.

Figura 4.4.1-2 Comportamiento del modelo visualmente



Fuente: Adaptado de datos INEC (2017)

4.4.2. Modelo dos de Regresión lineal: considerando la variable Enfermedad

La variable Y son las muertes fetales y la variable X es Enfermedad (causas que originan la muerte fetal):

$$Y_i = 36.433 - 7.383 * Enfermedad$$

$$H_0: \beta_{Enfermedad} = 0$$

$$H_1: \beta_{Enfermedad} \neq 0$$

H₀: No existe una relación lineal significativa entre la variable Enfermedad y Muertes

H_1 : Existe una relación lineal significativa entre la variable Enfermedad y Muertes

Figura 4.4.2-1 Coeficientes – Modelo-Regresión lineal- Enfermedad

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  31.467      6.054   5.198 1.32e-06 ***
Enfermedad2 -14.633      8.562  -1.709  0.0910 .
Enfermedad3 -14.767      8.562  -1.725  0.0881 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.16 on 87 degrees of freedom
Multiple R-squared:  0.04323,    Adjusted R-squared:  0.02123
F-statistic: 1.965 on 2 and 87 DF,  p-value: 0.1463
    
```

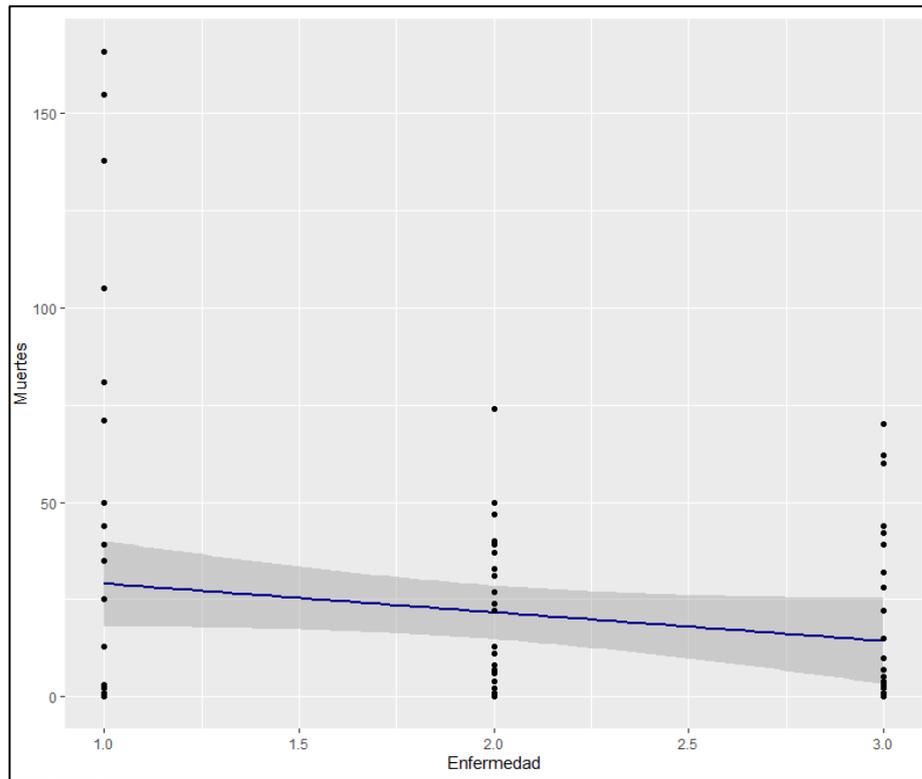
Existe una relación lineal significativa entre la variable Enfermedad y Muertes fetales, pues el valor p es mayor que el nivel de significancia $\alpha=0.05$, ahora al verificar si el modelo es significativo se obtiene que el valor p es 0.1463 por lo tanto se concluye que el modelo es significativo. Su R^2 no nos proporciona que sea un buen modelo sin embargo en este modelo incremento en un 0.016 con respecto al modelo uno.

Tabla 4.4.2-1 Resumen del modelo dos

Modelo Dos	Ecuación	R^2	P valor
Regresión lineal	Muertes fetales =31.46- 14.63Enfermedad	0.043	0.1463

En la figura 4.4.2-2 se observa la pendiente del modelo que se encuentra representado por -7.38 y su intercepto inicia desde 36.43.

Figura 4.4.2-2 Comportamiento del modelo dos visualmente



Fuente: Adaptado de datos INEC (2017)

4.4.3. Modelo tres de Regresión lineal: considerando la variable Región

Como se ha mencionado anteriormente, debido a que el tema del proyecto está centrado en la región costa, las otras regiones del país serán comparadas con la costa. Con respecto al procedimiento de la obtención del presente modelo se procedió a identificar que las variables al ser cualitativas se categorizaron en tres categorías para poder obtener el modelo de regresión lineal tomando como referencia a la región Costa con una valoración de cero, Oriente uno y Sierra dos.

La variable Y son las muertes fetales y la variable X es Región:

$$Y_i = 31.167 - 27.61 \text{ Region Oriente} - 9.94 \text{ Region Sierra}$$

$$H_0: \beta_{Region} = 0$$

$$H_1: \beta_{Region} \neq 0$$

H_0 : No existe una relación lineal significativa entre la variable Region y Muertes

H_1 : Existe una relación lineal significativa entre la variable Region y Muertes

Figura 4.4.3-1 Coeficientes-Modelo-Regresión lineal-variable Región

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    31.167      5.385   5.788 1.11e-07 ***
RegionOriente -27.611      9.327  -2.960 0.00396 **
RegionSierra   -9.944      7.615  -1.306 0.19505
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.31 on 87 degrees of freedom
Multiple R-squared:  0.09162,    Adjusted R-squared:  0.07074
F-statistic: 4.388 on 2 and 87 DF,  p-value: 0.0153

```

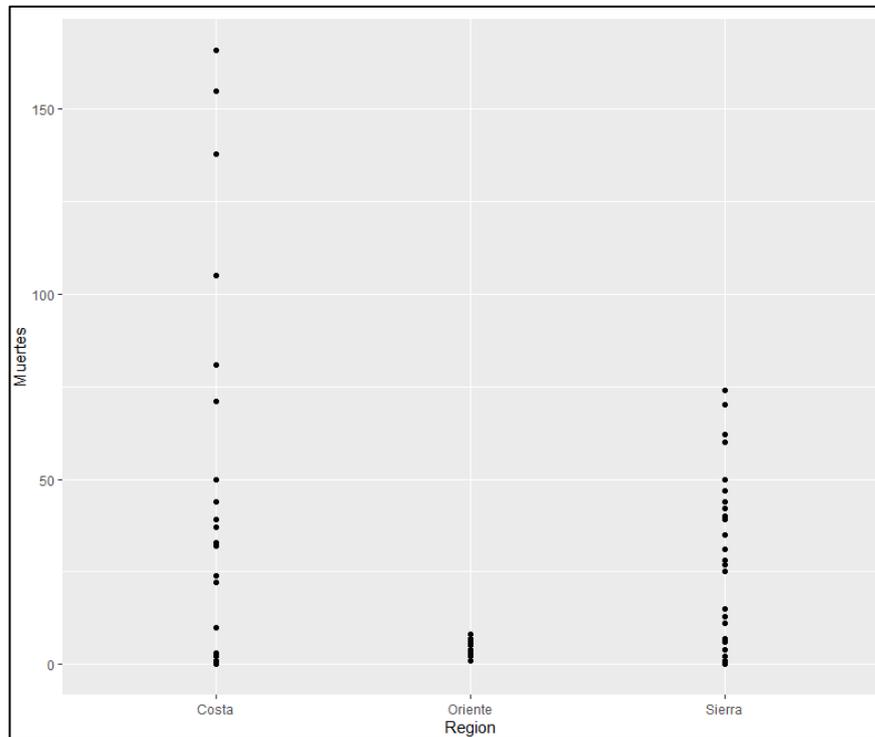
Existe una relación lineal significativa entre la variable región y Muertes fetales, pues el valor p es menor que el nivel de significancia $\alpha=0.05$, ahora al verificar si el modelo es significativo obtenemos que el valor p es 0.0153 por lo tanto se concluye que el modelo es significativo. Sin embargo, su R^2 alcanza un 0.10 mejor que los demás modelos que se ha analizado.

Tabla 4.4.3-1 Resumen del modelo tres

Modelo Tres	Ecuación	R^2	P valor
Regresión lineal	Muertes Fetales=31.17-27.61Región Oriente -9.94Región Sierra	0.091	0.015

En la figura 4.4.3-2 se observa el comportamiento del modelo tres con respecto a la variable región.

Figura 4.4.3-2 Comportamiento del modelo tres visualmente



Fuente: Adaptado de datos INEC (2017)

4.4.4. Modelo cuatro de Regresión lineal: considerando la variable Nacidos

La variable Y son las muertes fetales y la variable X es Nacidos:

$$Y_i = -1.44611 + 0.0022745 \text{ Nacidos}$$

$$H_0: \beta_{\text{Nacidos}} = 0$$

$$H_1: \beta_{\text{Nacidos}} \neq 0$$

H₀: No existe una relación lineal significativa entre la variable Nacidos y Muertes

H₁: Existe una relación lineal significativa entre la variable Nacidos y Muertes

Figura 4.4.4-1 Coeficientes- Modelo de Regresión lineal-Nacidos

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.4461140  3.0682167  -0.471   0.639
Nacidos      0.0022745  0.0002024  11.238  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.6 on 88 degrees of freedom
Multiple R-squared:  0.5893,    Adjusted R-squared:  0.5847
F-statistic: 126.3 on 1 and 88 DF,  p-value: < 2.2e-16
    
```

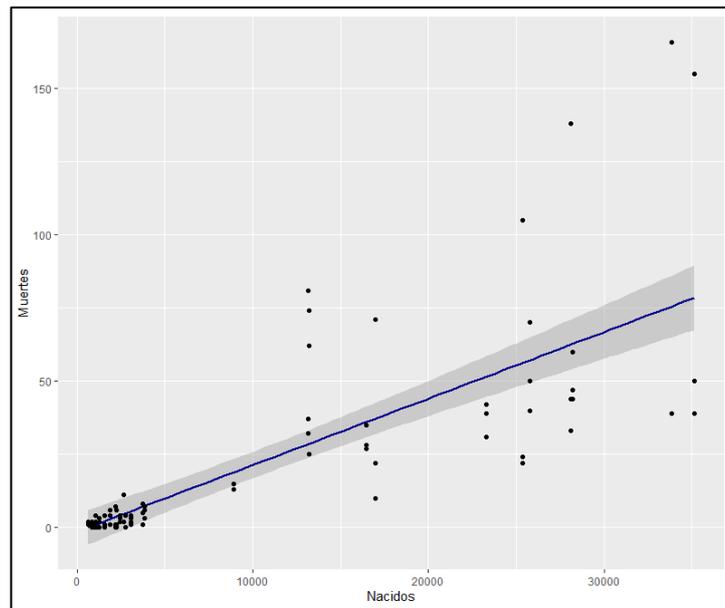
Existe una relación lineal significativa entre la variable Nacidos y Muertes fetales, pues el valor p es menor que el nivel de significancia $\alpha=0.05$, ahora al verificar si el modelo es significativo obtenemos que el valor p es 0.0000 por lo tanto se concluye que el modelo es significativo. Su R^2 nos proporciona que es un mejor modelo que los anteriores.

Tabla 4.4.4-1 Resumen del modelo cuatro

Modelo Cuatro	Ecuación	R ²	P valor
Regresión lineal	Muertes Fetales=-1.446+0.0022Nacidos	0.589	0.00

En la figura 4.4.4-2 se observa la pendiente del modelo que se encuentra representado por 0.0022 y su intercepto inicia desde -1.446.

Figura 4.4.4-2 Ajuste del modelo cuatro visualmente



Fuente: Adaptado de datos INEC (2017)

4.4.5. Modelo cinco de Regresión lineal: considerando la variable Zona

La variable Y son las muertes fetales y la variable X es Zona:

$$Y_i = 1.778 - 33.148 \text{ Zona}$$

$$H_0: \beta_{\text{Zona}} = 0$$

$$H_1: \beta_{\text{Zona}} \neq 0$$

H_0 : No existe una relación lineal significativa entre la variable Zona y Muertes

H_1 : Existe una relación lineal significativa entre la variable Zona y Muertes

Figura 4.4.5-1 Coeficientes-Modelo de Regresión lineal-variable Zona

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.778      4.906    0.362   0.718
ZonaUrbano    33.148      6.333    5.234 1.12e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 29.44 on 88 degrees of freedom
Multiple R-squared:  0.2374,    Adjusted R-squared:  0.2287
F-statistic: 27.39 on 1 and 88 DF,  p-value: 1.116e-06

```

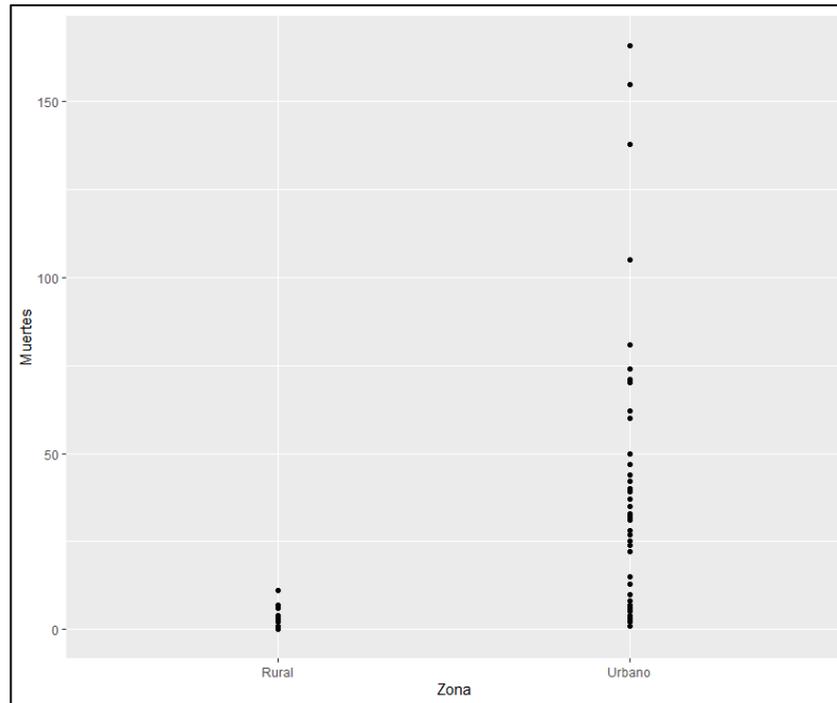
Existe una relación lineal significativa entre la variable Zona y Muertes fetales, pues el valor p es menor que el nivel de significancia $\alpha=0.05$, ahora al verificar si el modelo es significativo obtenemos que el valor p es 0.000 por lo tanto se concluye que el modelo es significativo. Su R^2 nos proporciona que no es un buen modelo puesto que no se acerca a uno.

Tabla 4.4.5-1 Resumen del modelo cinco

Modelo Cinco	Ecuación	R^2	P valor
Regresión lineal	Muertes Fetales=1.778-33.148Zona	0.2374	0.00

En la figura 4.4.5-2 se observa el comportamiento del modelo con respecto a la zona rural y urbana.

Figura 4.4.5-2 Comportamiento del modelo cinco visualmente



Fuente: Adaptado de datos INEC (2017)

4.5. Modelo de Regresión lineal Múltiple

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 X_5$$

La variable Y es Muertes Fetales y las Variables X son Región, Enfermedad, Edad, Zona, Nacidos

$$Y_i = B_0 + \beta_1 Region + \beta_2 Enfermedad + \beta_3 Zona + \beta_4 Nacidos + \beta_5 Edad + \beta_6 Region * Enfermedad + \beta_7 Zona * Enfermedad + \beta_8 Edad * Enfermedad$$

$$H_0: \beta_{Region} = \beta_{Enfermedad} = \beta_{Zona} = \beta_{Nacidos} = \beta_{Edad} = \beta_{Region*Enfermedad} = \beta_{Zona*Enfermedad} = \beta_{Edad*Enfermedad} = 0$$

H_1 : Al menos una es diferente

Figura 4.5-1 Coeficientes del Modelo de Regresión lineal Múltiple

```

Call:
lm.default(formula = Muertes ~ Region + Edad + Enfermedad1 +
  Zona + Nacidos + Region * Enfermedad1 + Zona * Enfermedad1 +
  Edad * Enfermedad1, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-24.487  -9.821  -1.011   4.852  50.061

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    14.162519   7.753816   1.827  0.07180 .
RegionOriente  -59.674499  10.380034  -5.749 1.89e-07 ***
RegionSierra   -38.344075   6.181875  -6.203 2.89e-08 ***
Edad            0.895054   1.607255   0.557  0.57929
Enfermedad1Muerte del bebe en el vientre -26.885000  10.932278  -2.459  0.01626 *
Enfermedad1Otras causas de muerte      -31.615000  10.932278  -2.892  0.00502 **
ZonaUrbano     39.047021   8.968196   4.354 4.23e-05 ***
Nacidos        0.001773   0.000314   5.645 2.88e-07 ***
RegionOriente:Enfermedad1Muerte del bebe en el vientre 67.375000  11.469147   5.874 1.13e-07 ***
RegionSierra:Enfermedad1Muerte del bebe en el vientre 47.914667   8.669860   5.527 4.65e-07 ***
RegionOriente:Enfermedad1Otras causas de muerte      65.291667  11.469147   5.693 2.37e-07 ***
RegionSierra:Enfermedad1Otras causas de muerte      50.750000   8.669860   5.854 1.23e-07 ***
Enfermedad1Muerte del bebe en el vientre:ZonaUrbano -42.750000   8.669860  -4.931 4.88e-06 ***
Enfermedad1Otras causas de muerte:ZonaUrbano      -38.750000   8.669860  -4.470 2.77e-05 ***
Edad:Enfermedad1Muerte del bebe en el vientre      1.502857   2.270302   0.662  0.51005
Edad:Enfermedad1Otras causas de muerte            1.925714   2.270302   0.848  0.39905
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.02 on 74 degrees of freedom
Multiple R-squared:  0.8331,    Adjusted R-squared:  0.7993
F-statistic: 24.62 on 15 and 74 DF,  p-value: < 2.2e-16

```

Todas las variables tienen una relación lineal con la variable Muertes fetales y al observar si el modelo es significativo tengo un valor $p=0.000$ por lo que se concluye que el modelo es significativo, pero al verificar los supuestos del modelo no hay normalidad en residuos como se visualiza en la tabla 4.5-1.

Tabla 4.5-1 Verificación del supuesto de normalidad-Residuos

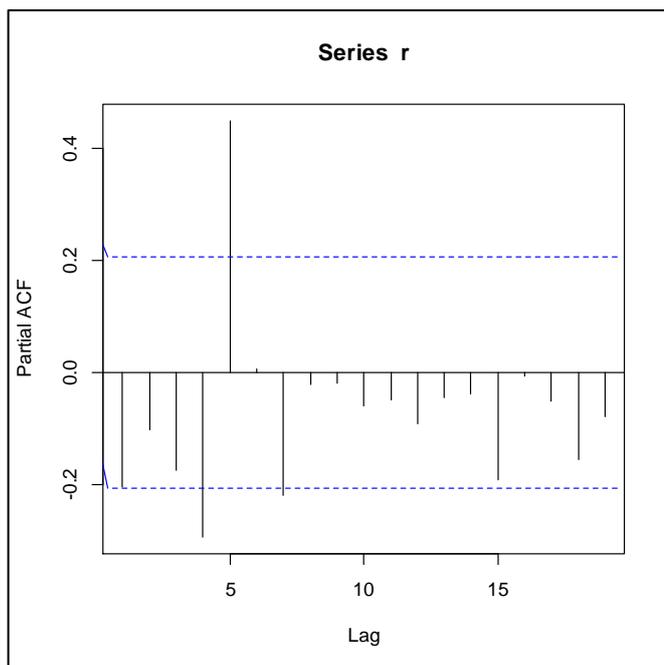
```

Ks.test(r,"prom",0,sd(r))
One-sample Kolmogorov-Smirnov test
data:r
D=0.15949,p-value=0.01818
alternative hypothesis: two-sided

```

En la Figura 4.5-2 se observa que los residuos están autocorrelacionados, por lo tanto, los residuos no son ruido blanco.

Figura 4.5-2 Función-Autocorrelación



Fuente: Adaptado de datos INEC (2017)

Debido a que el incumplimiento de los supuestos afecta a las estimaciones de las varianzas de los estimadores, se interpretan los coeficientes considerando una matriz de varianzas y covarianza robusta como se visualiza en la figura 4.5-3.

Figura 4.5-3 Matriz de varianzas y covarianza robusta

```

t test of coefficients:

                                     Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                          1.4163e+01  1.0388e+01  1.3633  0.1769220
RegionOriente                         -5.9674e+01  1.0480e+01 -5.6940  2.362e-07 ***
RegionSierra                          -3.8344e+01  1.0045e+01 -3.8174  0.0002779 ***
Edad                                   8.9505e-01  2.0674e+00  0.4329  0.6663256
EnfermedadlMuerte del bebe en el vientre -2.6885e+01  1.3819e+01 -1.9455  0.0555161 .
EnfermedadlOtras causas de muerte      -3.1615e+01  1.2892e+01 -2.4523  0.0165546 *
ZonaUrbano                             3.9047e+01  1.1616e+01  3.3616  0.0012290 **
Nacidos                                1.7729e-03  5.1633e-04  3.4336  0.0009789 ***
RegionOriente:EnfermedadlMuerte del bebe en el vientre 6.7375e+01  1.2521e+01  5.3811  8.340e-07 ***
RegionSierra:EnfermedadlMuerte del bebe en el vientre 4.7917e+01  1.2034e+01  3.9819  0.0001583 ***
RegionOriente:EnfermedadlOtras causas de muerte 6.5292e+01  1.2458e+01  5.2408  1.456e-06 ***
RegionSierra:EnfermedadlOtras causas de muerte 5.0750e+01  1.1685e+01  4.3430  4.400e-05 ***
EnfermedadlMuerte del bebe en el vientre:ZonaUrbano -4.2750e+01  1.2034e+01 -3.5526  0.0006682 ***
EnfermedadlOtras causas de muerte:ZonaUrbano -3.8750e+01  1.1685e+01 -3.3161  0.0014167 **
Edad:EnfermedadlMuerte del bebe en el vientre 1.5029e+00  2.7841e+00  0.5398  0.5909501
Edad:EnfermedadlOtras causas de muerte 1.9257e+00  2.5028e+00  0.7694  0.4440917
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Las variables que no resultaron significativas son edad, la interacción entre edad y muerte. Lo que se evidenció cuando se construyeron los modelos de regresión lineal simple.

Por otro lado, en región oriente hay una disminución en la media de muertes de 59.67 en relación con la costa, mientras que en la región sierra se tiene una disminución de 38.34 en la media muertes con relación a la costa. En el caso de pasar de un grupo de edad a otro se tendrá un aumento en la media de muertes de 0.89. Por otro lado, la media de muertes fetales cuando ocurre en el vientre tiene una disminución de 26.88 en relación con cuando la muerte ocurre por privación de oxígeno. Por otra parte, cuando las muertes son por otras causas hay una disminución de en la media de muertes de 31.62 en relación con la causa de privación de oxígeno en el cuerpo. Habría que decir también que la media de muertes en la zona urbano en mayor en 39.05 en relación con la zona rural. Así mismo si se incrementa una unidad en nacidos se tendrá un incremento de muertes de 0.002.

4.6. Modelo lineal Generalizado (Distribución Binomial).

Para la obtención de los modelos binomiales se trabajó con los datos agrupados de en tres grupos que representan las causas de las defunciones fetales. Esta variable es cualitativa de tipo nominal con tres niveles.

Tabla 4.6-1 Datos para Modelos Binomiales

N°	Causas de Muerte
1	Hipoxia intrauterina
2	Muerte del bebe en el vientre
3	Otras Causas

4.6.1. Modelo de Regresión logística Multinomial

En el modelo regresión logística multinomial se consideró a la variable enfermedad como dependiente puesto que es tipo nominal con tres categorías (politómica): Muerte del bebe en el vientre, Hipoxia intrauterina y otra causa de muerte. Se consideró una variable enfermedad categórica nominal politómica con soporte $(Y) = \{1,2,3\}$ y con probabilidades $p_1 = p(Y = 1)$, $p_2 = p(Y = 2)$ y $p_3 = p(Y = 3) = 1 - p_1 - p_2$. Se redefine la variable Y mediante vector (Y_1, Y_2) construido de la siguiente forma:

$$(Y_1, Y_2) = \begin{cases} (1,0) & \text{si } Y = 1 \\ (0,1) & \text{si } Y = 2 \\ (0,0) & \text{si } Y = 3 \end{cases}$$

$$\left. \begin{aligned} \ln\left(\frac{p_2}{p_1}\right) &= Z_1 = \beta_{01} + \beta_{11} * X_1 + \beta_{21} * X_2 \\ \ln\left(\frac{p_3}{p_1}\right) &= Z_2 = \beta_{02} + \beta_{12} * X_1 + \beta_{22} * X_2 \end{aligned} \right\}$$

$$\begin{aligned} \ln\left(\frac{P(\text{Enfermedad} = \text{Muerte del bebe en el vientre})}{P(\text{Enfermedad} = \text{Hipoxia intrauterina})}\right) \\ = -0.08061441 + 0.1445009 * \text{Edad} - 1.7391508 * \text{ZonaU} \\ + 2.222046 * \text{Región Oriente} + 1.392251 * \text{Region Sierra} \end{aligned}$$

$$\begin{aligned} \ln\left(\frac{P(\text{Enfermedad} = \text{Otra causa de muerte})}{P(\text{Enfermedad} = \text{Hipoxia intrauterina})}\right) \\ = -1.16633354 + 0.1625396 * \text{Edad} - 0.7769213 * \text{ZonaU} \\ + 1.967448 * \text{Región Oriente} + 1.582315 * \text{Region Sierra} \end{aligned}$$

En la figura 4.6.1-1 se visualiza el comportamiento del modelo como la obtención de los coeficientes del modelo objeto de estudio.

Figura 4.6.1-1 Coeficientes – Modelo-Regresión Logística Multinomial

```
Call:
multinom(formula = Enfermedad2 ~ Edad + Zona + Region, data = data,
          weights = Muertes)

Coefficients:
(Intercept)      Edad ZonaUrbano RegionOriente RegionSierra
2 -0.08061441 0.1445009 -1.7391508      2.222046      1.392251
3 -1.16633354 0.1625396 -0.7769213      1.967448      1.582315

Std. Errors:
(Intercept)      Edad ZonaUrbano RegionOriente RegionSierra
2  0.3965394 0.03898848  0.3754454      0.3521620      0.1234613
3  0.4453073 0.03920894   0.4234061      0.3696865      0.1230855

Residual Deviance: 3762.452
AIC: 3782.452
```

En la tabla 4.6.1-1 se visualiza el modelo a partir de la estimación exponencial en donde se realiza una comparación de las probabilidades entre las causas de muerte de fetos de las regiones objeto de estudio.

Tabla 4.6.1-1 Distribución logística Multinomial estimación exponencial

EXP(COEF(M1))	Estimate	Exp (estimate)	Interpretación
Intercepto	-0.0806144	0.9225493	
Edad	0.1445009	1.155463	La propensión de fallecer por la causa dos de cambiar de un rango de edad a otro aumenta en 16% con respecto a la causa uno
Zona Urbana	-1.7391508	0.1756695	La propensión de fallecer por la causa dos en la zona urbana disminuye en un 82% con respecto a la causa uno si se compara con la zona rural.
Región Oriente	2.222046	9.226189	La propensión de fallecer por la causa dos en la Región Oriente aumenta en un 822% con respecto a la causa uno si se compara con la costa.
Región Sierra	1.392251	4.023899	La propensión de fallecer por la causa dos en la Región Sierra aumenta en un 302% con respecto a la causa uno, si se compara con la costa.
Enfermedad tres respecto a la Enfermedad uno			
Intercepto	-1.16633354	0.3115070	

Edad	0.1625396	1.176495	La propensión de fallecer por la causa tres de cambiar de un rango de edad a otro aumenta en 18% con respecto a fallecer por la causa uno
Zona Urbana	-0.7769213	0.4598195	La propensión de fallecer por la causa tres en la zona urbana disminuye en un 54% con respecto a fallecer por la causa uno, si se compara con la zona rural.
Región Oriente	1.967448	7.152399	La propensión de fallecer por la causa tres en la Región Oriente aumenta en un 615% con respecto a fallecer por la causa uno, si se compara con la costa
Región Sierra	1.582315	4.866208	La propensión de fallecer por la causa tres en la Región Sierra aumenta en un 487% con respecto a fallecer por la causa uno, si se compara con la costa.

4.7. Modelo Lineal Generalizado (Distribución Poisson)

Para efectuar esta modelación se utilizó una base de datos conformada por 1950 datos la cual fue visualizada en el software Rstudio de donde se obtuvieron los resultados que se pueden visualizar en la tabla.

Tabla 4.7-1 Distribución de Poisson en Rstudio

Coeficientes	Estimate	Std. Error	z. value	Pr.(< z)	
Intercepción	-7.12753	0.32752	-21.762	< 2 e-16	***
Región oriente	-1.95410	0.29107	-6.713	1,90 e-11	***
Región sierra	-0.99104	0.07886	-12,567	< 2 e-16	***
Edad	0.04617	0.02240	2.061	0.03931	*
Muerte del bebe en el vientre	-0.06420	0.39104	-0.164	0.86959	
Otras causas de muertes	-1.19906	0.43660	-2.746	0.00603	**
Zona Urbano	1.60506	0.31838	5,041	4,62 e-07	***
Región oriente: Muerte del bebe en el vientre	2.20552	0.35063	6.290	3,17 e-10	***
Región Sierra: Muerte del bebe en el vientre	1.40584	0.12231	11.494	< 2 e-16	***
Región Oriente: Otras causas de Muerte	1.94315	0.36797	-4.511	6,45 e-06	***
Región sierra: Otras causas de Muerte.	1.58356	0.122555	12.922	< 2 e-16	***
Edad: Muerte del bebe en el vientre	0.14375	0.03866	3.719	0.0002	***
Edad: Otras causas de Muerte.	0.17650	0.03913	4.511	6,45 e-06	***
Muerte del bebe en el vientre: zona urbana	-1.75760	0.36109	-4.868	1.13 e-06	***
Otras causas de Muerte: zona urbana	-0.79850	0.40791	-1.958	0.05028	

En la tabla 4.7-2 se visualiza el modelo a partir de la estimación exponencial en donde se realiza una comparación de medias entre las muertes fetales de las regiones objeto de estudio.

Tabla 4.7-2 Distribución de Poisson a partir de estimación exponencial

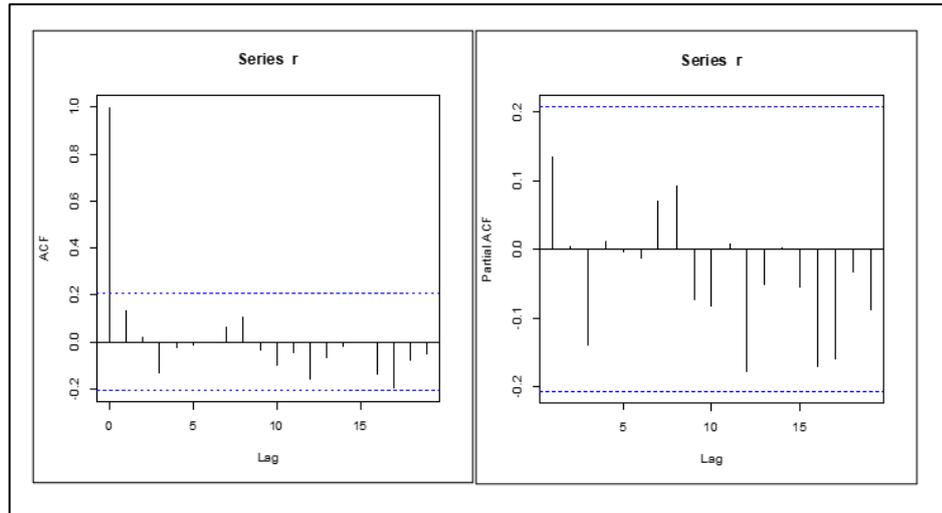
EXP(COEF(M1))	Estimate	Exp (estimate)	Interpretación
Intercepto	-7.12753	0.000802700	
Región Oriente	-1.95410	0.1416919421	La media de muertes fetales en la región oriente es 86% menos que la media de muertes en la región costa
Región Sierra	-0.991040	0.3711904521	La media de muertes fetales en la región sierra es 63% menos que la media de muertes en la región costa
Edad	0.046170	1.0472524287	La media de muertes fetales de cambiar de un rango de edad a otro aumenta en un 5%
Muerte del bebe en el vientre	-0.064200	0.9378174173	La media de muertes fetales por la causa 2 es 7% menos que la media de muertes por la causa 1
Otras causas de muertes	-1.19906	0.301477468	La media de muertes fetales causada por "otras causas" es 70% menor que la media de muertes por la causa 1.
Zona Urbano	1.60506	4.978158283	La media de muertes fetales en la Zona urbano es 397% más que la media de muertes en las Zona Rural

Región Oriente: Muerte del bebe en el vientre	2.20552	9.074969325	La media de muertes del feto cuando ocurre en el vientre versus muerte por privación de oxígeno en la región oriente en relación de la región costa implica un incremento de 807%.
Región sierra: Muerte del bebe en el vientre	1.40584	4.078951622	El cambio en esta ratio indica que la media de muertes del bebe en el vientre versus muerte por privación de oxígeno en la región sierra en relación de la región costa implica un incremento en 307% en la media de muertes.
Región Oriente: Otras causas de muerte	1.94315	6.980705597	El cambio en esta ratio indica que la media de muertes por otras causas versus muerte por privación de oxígeno en la región oriente en relación de la región costa implica un incremento en 598% en la media de muertes.
Región Sierra: Otras causas de Muerte.	1.58356	4.872270255	El cambio en esta ratio indica que la media de muertes por otras causas versus muerte por privación de oxígeno en la región sierra en relación de la región costa implica un cambio en 387% en la media de muertes.
Edad: Muerte del bebe en el vientre	0.14375	1.154595424	El cambio en esta ratio indica que la media de muertes del bebe en el vientre versus muerte por privación de

			oxígeno cuando hay un cambio en el rango de edad implica un cambio en 15% en la media de muertes
Edad: Otras causas de Muerte.	0.1765	1.193034427	El cambio en esta ratio indica que la media de muertes por otras causas versus muerte por privación de oxígeno cuando hay un cambio en el rango de edad implica un cambio en 19% en la media de muertes
Muerte del bebe en el vientre: zona urbana	-1.7576	0.172458267	El cambio en esta ratio indica que la media de muertes del bebe en el vientre versus muerte por privación de oxígeno cuando hay un cambio de la zona urbana a la zona rural implica un cambio en 83% menos en la media de muertes
Otras causas de Muerte: zona urbana	-0.7985	0.450003463	El cambio en esta ratio indica que la media de muertes por otras causas versus muerte por privación de oxígeno cuando hay un cambio de la zona urbana a la zona rural implica un cambio en 55% menos en la media de muertes

En las gráficas de función de autocorrelación se puede apreciar que sus autocorrelaciones muestran que los residuos deviance son ruido blanco.

Figura 4.7-1 Función de Autocorrelación y Parcial



Fuente: Adaptado de datos INEC (2017)

Con respecto al Pseudo R

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})}$$

$$R^2 = 1 - \frac{132.13}{796,46}$$

$$R^2 = 0.834103$$

Con lo cual se concluye lo siguiente: El modelo un pseudo- R^2 de 0,83 es bueno. Esto sugiere que las defunciones fetales se describen mejor con un modelo multivariado que incorpore otras covariantes, que con un modelo univariado que sólo describe las defunciones fetales con una variable. Por último, en la tabla 4.7-3 se presenta el test de que los residuos tengan una distribución normal, los resultados del modelo no discrepan tanto de los datos originales y como p es mayor a 0.05 se dice que los datos no son normales.

Ho: “Los residuos provienen de una distribución normal estándar”

H1: “Los residuos no provienen de una distribución normal estándar”

Tabla 4.7-3 Verificación del supuesto de normalidad - residuos

```
r=residuals(m1,type="deviance")
>ks.test(r,"pnorm",mean(r),sd(r))
One-sample Kolmogorov-Smirnov test
data:r
D=0.097436,p-value=0.3379
alternative hypothesis: two-sided
```

A un 95% de confianza hay suficiente evidencia estadística para concluir que los residuos se pueden modelar como una normal estándar.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES.

5.1. Conclusiones

- De acuerdo con los hallazgos en los modelos lineales, las variables que tienen mayor influencia con respecto al número de muertes fetales son las variables: causa fetal agrupada, región de inscripción, zona y nacidos puesto que su p valor es menor que el nivel de significancia $\alpha=0.05$, por lo tanto, se concluyó que los modelos son significativos. Por otro lado, la variable que no influye al número de muertes es la edad de la madre pues su valor p es mayor que el nivel de significancia se obtuvo que su valor p es 0.2327 por lo tanto se concluyó que el modelo no es significativo.
- En el modelo de regresión múltiple las variables: región, enfermedad, zona y nacidos tienden a tener mayor influencia en el número de muertes fetales pues se tuvo un valor $p=0.000$ por lo que se concluyó que el modelo es significativo, pero al verificar los supuestos del modelo no hay normalidad en residuos, por lo tanto, los residuos no son ruido blanco.
- Como en la Región Costa se han registrado el mayor número de defunciones fetales es decir el 58%, siendo la principal causa de esas muertes la hipoxia intrauterina, en el modelo de regresión logística multinomial se determinó que la propensión de fallecer por la causa de muerte del bebe en el vientre de cambiar de un rango de edad a otro aumenta en 16% con respecto a la causa por hipoxia intrauterina. Así mismo la propensión de fallecer por la causa de muerte del bebe en el vientre de cambiar en la zona urbana disminuye en un 82% con respecto a la causa por hipoxia intrauterina si se compara con la zona rural. Por otra parte, la propensión de fallecer por la muerte del bebe en el vientre de cambiar en la Región Oriente aumenta en un 822% con respecto a hipoxia intrauterina si se compara con la costa. Habría que decir también que la propensión de fallecer por la causa de muerte del bebe en el vientre de cambiar

en la Región Sierra aumenta en un 302% con respecto a hipoxia intrauterina, si se compara con la costa.

- Al examinar la Región Costa que cuenta con mayor índice en las defunciones fetales originada por la causa de hipoxia intrauterina representada en el 76% de las tres causas de enfermedad analizadas en trabajo de investigación, el modelo Lineal Generalizado Distribución de Poisson demostró que las variables que mayor influencia tienen con respecto al número de muertes fetales son las variables región, zona y causa fetal agrupada, no obstante, la variable edad no influye al número de muertes fetales, habría que decir también que el modelo tuvo un pseudo- R^2 de 0,83 es decir bueno, esto sugiere que las defunciones fetales se describen mejor con un modelo multivariado que incorpore otras covariantes, que con un modelo univariado que sólo describe las defunciones fetales con una variable.
- Se puede apreciar poco registro de defunciones en el área rural de todas las regiones, lo que nos hace pensar que ocurren pero que la gente no suele registrar estas defunciones.

5.2.RECOMENDACIONES.

- Se debería incursionar en trabajar con modelos lineales generalizados puesto que son referentes imprescindibles actualmente en el análisis de datos.
- Desde el aspecto médico se debería ahondar en estudios a partir de la creación de modelos predecibles que ayuden a mitigar las consecuencias de las muertes fetales puesto que se podría contar con características métricas de las variables objeto de estudio.
- En lo que respecta al desarrollo de este tipo de modelos se debería utilizar el software R puesto que su adaptabilidad permite su desarrollo como su libre acceso en internet.
- Se debería llevar un control de estas defunciones para mejorar el registro estadístico de estos datos con lo que se obtendrían análisis más certeros con datos más confiables.

6. Referencias

- Amat Rodrigo, J. (Junio de 2016). *rpubs*. Obtenido de https://rpubs.com/Joaquin_AR/223351
- Arroyo, I., Bravo M, L. C., Llinás, H., & Muñoz, F. L. (2014). Distribuciones Poisson y Gamma: Una Discreta y Continua Relación. *Prospect*, 99-107.
- Cañamero de León, S. (15 de Diciembre de 2015). *guiainfantil.com*. Obtenido de <https://www.guiainfantil.com/articulos/embarazo/riesgos/causas-de-muerte-fetal-en-el-embarazo/>
- Cayuela , L. (2010). *Modelos lineales generalizados (GLM)*. Granada : Universidad de Granada- Junta de Andalucía.
- Cayuela, L. (17 de Mayo de 2020). Obtenido de https://s3.amazonaws.com/academia.edu.documents/33538949/3-Modelos_lineales_generalizados.pdf?response-content-disposition=inline%3B%20filename%3DModelos_lineales_generalizados_GLM.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=ASIATUSBJ6BACAD4BL5W
- Escalante Cortina, R. D. (2012). *Manual de aplicación del modelo de regresión lineal múltiple con correcciones de especificación, usos de STATA 9.0, STATA 10.0, EVIEWS 5.0, SSPS 11.0*. S.I. Argentina : B-EUMED
- García, H. A., Solarte, C., & Imuez, M. (2009). *BioEstadística*. Argentina: Universitaria .
- García, E. M. (2014). *Aplicación de Modelos de Regresión de Poisson Bivariados a los resultados de los partidos de la Liga Española de Fútbol*. Retrieved from http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_751.pdf
- Gavilanez Ocampo , J., & Paucar Chalán , E. V. (2019). *Factores de riesgo relacionados con la defunción fetal en el Ecuador durante el período 1997 al 2017*. Quito: Universidad Central del Ecuador.

- Gómez, S., & Palacios, D. (2013). *Modelación logística Multinomial para clasificar los hogares de El Salvador*. Salvador : Universidad de El Salvador
- Gonzalez, A. (2015). *Selección de variables: Una revisión de métodos existentes*. Coruña: Universidad de Coruña.
- Guerrero , R. (2014). Protocolo de estudio en los casos de Muerte Fetal Intrauterina atendidos en el Hospital Regional de Poza Intrauterina atendidos en el Hospital Regional de Poza Intrauterina atendidos en el Hospital Regional de Poza Rica.
- Fanny Patricia, B. D., & Velasquez, V. L. (1998). Mortalidad fetal, frecuencia y factores de riesgo.
- Fernández, E., Castillo Núñez, B., & Llamas Herrarte, H. C. (2016). Caracterización clinicoepidemiológica de las defunciones neonatales. *Revista Médica de Risaralda*, 22(1), 241–248.
<https://doi.org/10.15446/rsap.v19n2.66239>
- Fernández, E. R., Castillo Núñez, B., & Herrarte Llamas, C. H. (2013). Caracterización clinicoepidemiológica de las defunciones neonatales. *MEDISAN*.
- Joaquín Amat , R. (2016). *Introducción a la Regresión Lineal Múltiple*.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista , L. (2014). *Metodología de la Investigación*. México: McGraw-Hill.
- INEC;. (2018). *Registro Estadístico de Defunciones Fetales 2017, Metodología*. Ecuador : Dirección de Estadísticas Sociodemográficas.
- OMS. (2012). *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud Décima Revisión*. Washington .
- OMS. (2017).
- OMS. (23 de Febrero de 2018). *Organización Mundial de la Salud*. Obtenido de <https://www.who.int/es/news-room/fact-sheets/detail/adolescent-pregnancy>

- Paladines Chicango , A. G. (2014). *Factores relacionados a muerte fetal en usuarias atendidas en el área de ginecología del hospital Teófilo Dávila de Machala durante el primer semestre del año 2013*. Machala: Universidad de Machala .
- UCLA. (28 de Febrero de 2020). *Statistical Consulting*. Obtenido de <https://stats.idre.ucla.edu/r/dae/poisson-regression/>
- Fanny Patricia, B. D., & Velasquez, V. L. (1998). Mortalidad fetal, frecuencia y factores de riesgo.
- Fernández, E., Castillo Núñez, B., & Llamas Herrarte, H. C. (2016). Caracterización clinicoepidemiológica de las defunciones neonatales. *Revista Médica de Risaralda*, 22(1), 241–248.
<https://doi.org/10.15446/rsap.v19n2.66239>
- Fernández, E. R., Castillo Núñez, B., & Herrarte Llamas, C. H. (2013). Caracterización clinicoepidemiológica de las defunciones neonatales. *MEDISAN*.
- García, E. M. (2014). *Aplicación de Modelos de Regresión de Poisson Bivariados a los resultados de los partidos de la Liga Española de Fútbol*. Retrieved from http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_751.pdf
- INEC. (2012). Encuesta Nacional de Salud y Nutrición- ENSANUT 2012 Demografía, salud materna e infantil y salud sexual y reproductiva, 1–30.
- López González, E., & Ruíz Soler, M. (n.d.). Análisis de datos con el Modelo Lineal Generalizado . Una aplicación con R, 59–80.
- María J, T., Francisco, S. P., José Alonso, F., Adrián, H., Carla, C. S., & Mardorys, D. (2016). Muerte Fetal: Caracterización epidemiológica. *Salud*.
- Medina Castro, D., & Castro Llamas, J. (2012). Alteraciones genéticas y estrategias diagnósticas en muerte fetal, *80(5)*, 313–319.

- Molina, P., & Agulló, Ó. (9 de Marzo de 2018). *Rpubs*. Obtenido de <https://rpubs.com/PatriciaMolina/378266>
- Ottersen, O. P., Frenk, J., & Horton, R. (2011). The Lancet-University of Oslo Commission on Global Governance for Health, in collaboration with the Harvard Global Health Institute. *The Lancet*, 378(9803), 1612–1613. [https://doi.org/10.1016/S0140-6736\(11\)61617-6](https://doi.org/10.1016/S0140-6736(11)61617-6)
- Paladines Chicago, A. G. (2014). *FACTORES RELACIONADOS A MUERTE FETAL EN USUARIAS ATENDIDAS EN EL ÁREA DE GINECOLOGÍA DEL HOSPITAL TEÓFILO DÁVILA DE MACHALA DURANTE EL PRIMER SEMESTRE DEL AÑO 2013*.
- Rangel Calvillo, M. N. (2014). Análisis de muerte fetal tardía. *Perinatología y Reproducción Humana*, 28(3), 139–145.
- Rivera Miranda, M. A., Lara Latamblé, N. T., & Baró Bouly, T. (2017). Asfixia al nacer: factores de riesgo materno y su repercusión en la mortalidad neonatal. *Revista de Información Científica* , 1143-1152
- Rivera, L., Fuentes Román, M. de L., Albores Esquinca, C., Abarca, F. J., & Hernández Girón, C. (2003). Factores asociados a mortalidad perinatal en el hospital general de Chiapas, México. *Revista de Saude Publica*, 37(6), 687–692. <https://doi.org/10.1590/s0034-89102003000600001>
- Rodríguez, J. M. (2014). Análisis de muerte fetal tardía. *Scielo* .
- Roque Cruz , M. J. (2018). *Modelos de Regresión Logística Multinomial de la Calidad de Fibra de Alpaca Huacaya en función de sus características:Sexo y Edad-Corani,Carabaya,Puno- 2017*. Perú : UNIVERSIDAD NACIONAL DEL ALTIPLANO Universidad Nacional del Altiplano .
- Tinedo, M., Santander, F., José, A., Herrera, A., Colombo, C., & Díaz, M. (2016). Muerte fetal: caracterización epidemiológica . *Scielo* .
- Salud, C. N. de E. T. en. (2010). Guía de práctica clínica Diagnóstico y tratamiento

de muerte fetal con feto único, 51.

UCLA. (28 de Febrero de 2020). *Statistical Consulting*. Obtenido de <https://stats.idre.ucla.edu/r/dae/poisson-regression/>

Valdés Vargas, M. A., González , R., & Vásquez López , H. (2012). *Diagnóstico y muerte fetal con feto único* . México : Copyright CENETEC .

Voto, L. S. (2013). *Hipertensión en el embarazo*. Corpus Editorial.

Valenti, E., & Otero, C. (2007). Mortalidad fetal 2006 en la Maternidad Sardá. *Revista Del Hospital Materno Infantil Ramón Sardá*.

Zurita Herrera, G. (2010). *Probabilidad y Estadística Fundamentos y Aplicaciones*. Guayaquil : ESPOL.

Zylberberg , A. D. (2005). *Probabilidad y Estadística*. Argentina : Nueva Librería.

7. Apéndices y Anexos

Apéndice A base de datos INEC (2017)

Edad	Región		Causa			Muertos por Región	Total	Nacidos por Región	Total	Nacidos Totales por Región	Nacidos Totales	Probabilidad
			Hipoxia intrauterina, no especificada	Muerte fetal de causa no específica	Otras Causas							
10-17	Costa	Urbano	71	22	10		59	16881	30616	16881	30675	0,19%
		Rural	2	1	0	3		845		848		
	Sierra	Urbano	13	13	15	41		8860		8901		
		Rural	0	4	2	6		1021		1027		
	Oriente	Urbano	3	4	2	9		2427		2436		
		Rural	0	0	0	0		582		582		
18-20	Costa	Urbano	105	24	22	151	255	25202	48627	25353	48882	0,52%
		Rural	1	0	1	2		1007		1009		
	Sierra	Urbano	35	27	28	90		16363		16453		
		Rural	0	6	1	7		2215		2222		
	Oriente	Urbano	2	2	1	5		3060		3065		
		Rural	0	0	0	0		780		780		
21-24	Costa	Urbano	166	39	39	244	383	33624	65506	33868	65889	0,58%
		Rural	3	0	2	5		1256		1261		
	Sierra	Urbano	39	31	42	112		23172		23284		
		Rural	0	4	4	8		2746		2754		
	Oriente	Urbano	1	8	5	14		3710		3724		
		Rural	0	0	0	0		998		998		
25-29	Costa	Urbano	155	50	39	244	426	34895	71548	35139	71974	0,59%
		Rural	0	0	0	0		1212		1212		
	Sierra	Urbano	44	47	60	151		28075		28226		
		Rural	2	11	2	15		2610		2625		
	Oriente	Urbano	3	6	7	16		3847		3863		
		Rural	0	0	0	0		909		909		
30-35	Costa	Urbano	138	33	44	215	393	27886	60515	28101	60908	0,65%
		Rural	0	0	1	1		1009		1010		
	Sierra	Urbano	50	40	70	160		25605		25765		
		Rural	1	7	0	8		2165		2173		
	Oriente	Urbano	2	4	3	9		3067		3076		
		Rural	0	0	0	0		783		783		
mas de 36	Costa	Urbano	81	37	32	150	331	13011	31087	13161	31418	1,05%
		Rural	1	1	2	4		647		651		
	Sierra	Urbano	25	74	62	161		13068		13229		
		Rural	0	4	1	5		1547		1552		
	Oriente	Urbano	1	6	4	11		1871		1882		
		Rural	0	0	0	0		943		943		

Apéndice B codificación del diagrama de Pareto en r studio

```
causa <-read.csv("C:/Users/ALEXANDRA/Desktop/causas de
defunciones.csv",header = T,");
```

```
attach(causa)
```

```
names(causa)
```

```
install.packages("qcc")## para diagrama de pareto
```

```
library(qcc)# nos permite utilizar dos variables
```

```
##Se genera una variable etiquetada
```

```
tipo<-Cantidad
```

```
names(tipo)<-Causas
```

```
tipo## variable etiquetada cada una de las observaciones se encuentra  
etiquetada
```

```
##Generar el diagrama
```

```
pareto.chart(tipo)
```

