

UNIVERSIDAD NACIONAL AGRARIA

LA MOLINA

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN



**“VARIABLES QUE EXPLICAN LOS RANGOS REMUNERATIVOS
DEL PRIMER EMPLEO DE LOS EGRESADOS UNIVERSITARIOS
DEL PERÚ APLICANDO REGRESIÓN LOGÍSTICA ORDINAL”**

PRESENTADO POR:

JUAN CARLOS AQUINO GAMBOA

TESIS PARA OPTAR EL TÍTULO DE
INGENIERO ESTADÍSTICO E INFORMÁTICO

Lima – Perú

2019

UNIVERSIDAD NACIONAL AGRARIA

LA MOLINA

FACULTAD DE ECONOMÍA Y PLANIFICACIÓN

**“VARIABLES QUE EXPLICAN LOS RANGOS REMUNERATIVOS
DEL PRIMER EMPLEO DE LOS EGRESADOS UNIVERSITARIOS
DEL PERÚ APLICANDO REGRESIÓN LOGÍSTICA ORDINAL”**

PRESENTADO POR

JUAN CARLOS AQUINO GAMBOA

TESIS PARA OPTAR EL TÍTULO DE
INGENIERO ESTADÍSTICO E INFORMÁTICO

SUSTENTADA Y APROBADA ANTE EL SIGUIENTE JURADO

Mg.Sc. Clodomiro Fernando Miranda Villagómez
Presidente

Dr. César Higinio Menacho Chiok
Asesor

MS. Carlos López de Castilla Vásquez
Miembro

Dr. Jorge Chue Gallardo
Miembro

Lima – Perú
2019

DEDICATORIA

Esta tesis está dedicada a:

A mis padres Mary y Luis Alberto quienes me inculcaron el valor del estudio lo cual me ha permitido llegar a cumplir hoy un sueño más.

A mi novia Juana quien siempre me brindó su aliento, que fue mi soporte y que fue mi gran apoyo, gracias por acompañarme en el cumplimiento de todos mis sueños.

A mis hermanos Jorge, Luis Fernando y Raúl Alberto por su cariño y por todas sus palabras de aliento. A toda mi familia porque con sus consejos y palabras de aliento hicieron de mí una mejor persona y de una u otra forma me acompañan en todos mis sueños y metas.

Juan Carlos Aquino Gamboa

AGRADECIMIENTO

Agradezco a mi madre y mi padre quienes siempre fueron mi soporte, quienes me impulsaron a continuar mis estudios, quienes me enseñaron que todo esfuerzo tiene recompensas.

Finalmente quiero expresar mi más grande y sincero agradecimiento al Dr. César Higinio Menacho Chiok, principal colaborador durante todo este proceso, quien con su dirección, conocimiento y enseñanza permitió el desarrollo de este trabajo.

Índice de Figuras

	Pag.
Figura 1. Tipos de modelos logísticos	06
Figura 2. Tablas para la base de datos del estudio	26

Índice de Cuadros

	Pag.
Cuadro 1. Rangos de ingreso de los egresados universitarios	27
Cuadro 2. Distribución de egresados por rango de ingreso	33
Cuadro 3. Medidas estadísticas para el ingreso	34
Cuadro 4. Cálculo de los intervalos de clase	35
Cuadro 5. Distribución de los egresados por rango de los ingresos	35
Cuadro 6. Distribución de egresados por rango de ingresos y tipo de universidad	36
Cuadro 7. Coeficientes estimados y prueba de Wald (Modelo 1)	38
Cuadro 8. Estimación de los coeficientes de regresión (Modelo 1)	40
Cuadro 9. Prueba de hipótesis global (Modelo 1)	45
Cuadro 10. Prueba de bondad de ajuste de la Deviance y Pearson (Modelo 1)	46
Cuadro 11. Cálculo del Pseudo-R ² (Modelo 1)	47
Cuadro 12. Prueba de hipótesis de líneas paralelas (Modelo 1)	47
Cuadro 13. Coeficientes estimados y prueba de Wald (Modelo 2)	49
Cuadro 14. Estimación de los coeficientes de regresión (Modelo 2)	51
Cuadro 15. Prueba de hipótesis global (Modelo 2)	54
Cuadro 16. Prueba de bondad de ajuste de la Deviance y Pearson (Modelo 2)	55
Cuadro 17. Cálculo del Pseudo-R ² (Modelo 2)	55
Cuadro 18. Prueba de hipótesis de líneas paralelas (Modelo 2)	56
Cuadro 19. Coeficientes estimados y prueba de Wald (Modelo 3)	57
Cuadro 20. Estimación de los coeficientes de regresión (Modelo 3)	59
Cuadro 21. Pruebas de hipótesis global (Modelo 3)	62
Cuadro 22. Prueba de bondad de ajuste de la Deviance y Pearson (Modelo 3)	62
Cuadro 23. Cálculo del Pseudo-R ² (Modelo 3)	63
Cuadro 24. Prueba de hipótesis de líneas paralelas (Modelo 3)	64
Cuadro 25. Ingreso promedio de los egresados universidades públicas y privadas	65
Cuadro 26. Calificación de la preparación recibida en la universidad para el desarrollo de las siguientes competencias	66
Cuadro 27. Importancia de las competencias recibidas para su experiencia profesional	67

ÍNDICE

RESUMEN.....	2
ABSTRACT	3
I. INTRODUCCIÓN	1
II. REVISIÓN DE LITERATURA.....	3
2.1 Panorama de los ingresos de los egresados universitarios	3
2.2 Modelos de la regresión logística.....	6
2.2.1 Modelo de regresión logística binaria	7
2.2.2 Modelos de regresión logística politómica.....	9
2.3 Inferencia estadística en los modelos logísticos.....	10
2.4 Análisis de los coeficientes en los modelos logísticos	15
2.5 La regresión logística ordinal	17
2.5.1 Formulación del modelo logístico ordinal.....	17
2.5.2 Modelos logísticos ordinales	18
2.5.3 Prueba de líneas paralelas.....	19
2.6 Comparación de medias para muestras independientes (Prueba t).....	20
III. MATERIALES Y MÉTODOS	22
3.1 Materiales	22
3.2 Metodología.....	22
3.2.1 Población y Muestra.....	22
3.2.2 Descripción de las variables	22
3.2.3 Modelos logísticos ordinales propuestos.....	24
3.2.4 Análisis de la regresión logística ordinal.....	26
IV. RESULTADOS Y DISCUSIÓN.....	33
4.1 Recopilación de datos.....	33
4.2 Pre procesamiento de datos	33
4.3 Análisis de regresión para los modelos logísticos ordinales ajustados	36
4.3.1 Análisis del Modelo 1.....	37
4.3.2 Análisis del Modelo 2.....	48
4.3.3 Análisis del Modelo 3.....	57
4.4 Comparación de los ingresos entre universidad pública y privada.....	65
V. CONCLUSIONES	68
VI. RECOMENDACIONES	70
VII. REFERENCIA BIBLIOGRÁFICA	71
VIII. ANEXOS	73

RESUMEN

En la presente tesis se aplica la regresión logística ordinal, con la finalidad de identificar las variables que mejor explican los rangos de los ingresos de los egresados universitarios del Perú. Para el estudio se usa los datos de la encuesta de egresados universitarios del año 2014 realizada por el INEI. Se ajustan los datos a tres modelos aplicando la regresión logística ordinal con función de enlace logit proporcional acumulativo y para las universidades públicas y privadas. La variable dependiente son los rangos de ingresos (Bajo, Medio, Alto y Muy alto) y los conjuntos de variables independientes agrupadas en cuatro categorías: Grupo A: socio-académicas (7), Grupo B: referidas a la evaluación de las competencias recibidas en la universidad (12), Grupo C: referidas a la importancia de las competencias para su desarrollo profesional (12) y Grupo D: respecto a los profesores de la carrera (5). Para explicar los rangos de los ingresos, la regresión logística ordinal identificó para los tres modelos y para las universidades públicas y privadas, variables significativas socio-académicas: sexo, pertenencia al cuadro de méritos, si obtuvo o no el título profesional, su primer empleo relacionado con la formación profesional. Respecto a la calificación sobre la preparación recibida en la universidad para el desarrollo de las competencias: para coordinar actividades, para los conocimientos básicos de otros campos (públicas) y para el dominio del área de disciplina, el utilizar herramientas informáticas básicas, el utilizar software específico de la carrera (privadas). Respecto a la importancia de las competencias para su experiencia laboral: redactar informes o documentos, tener conocimientos básicos de otros campos o disciplinas (públicas) y rendir bajo presión y cumplir con los objetivos y coordinar actividades (privadas).

Palabras claves: Encuesta de ingresos de egresados, Regresión logística ordinal, Logit acumulativo, Inferencia estadística.

ABSTRACT

In this thesis the ordinal logistic regression is applied, in order to identify the variables that best explain the income ranges of the university graduates of Peru. For the study, the data from the survey of university graduates of the year 2014 conducted by the INEI used. The data is adjusted to three models applying the ordinal logistic regression with cumulative proportional logit link function and for both public and private universities. The dependent variable is the income ranges (Low, Medium, High and Very high) and the sets of independent variables grouped into four categories: Group A: socio-academic (7), Group B: referred to the evaluation of the competences received in the university (12), Group C: referred to the importance of the competences for their professional development (12) and Group D: regarding the professors of the career (5). To explain the income lag, the ordinal logistic regression identified for the three models and for the public and private universities, significant socioeconomic variables: sex, pertenence to the table of merits, obtained the professional title, their first job related to professional training. On the qualification on the preparation received in the university for the development of the competences: the preparation to coordinate activities, the preparation in basic knowledge of other fields (public) and the preparation in the domain of the discipline area, the use of basic computer tools, the use of specific career software (private). On the importance of competencies for your work experience: write reports or documents, have basic knowledge of other fields or disciplines (public) and under pressure and meet the objectives and coordinate activities (private).

Key words: Graduate income survey, Ordinal logistic regression, Cumulative logit, Statistical inference.

I. INTRODUCCIÓN

La literatura sobre la educación superior y el mercado de trabajo señala que entre más alto es el nivel académico de una universidad, mayor es el éxito de sus egresados en el mercado laboral en términos del salario percibido. En muchos estudios relacionados con esta problemática se identifican las variables que explican el sueldo que obtienen los egresados universitarios, utilizando para ellos diferentes enfoques o perspectivas. Por esto, es necesario realizar estudios cuantitativos, que permitan analizar los factores más importantes que expliquen e incidan en la formación profesional y por ende en el nivel remunerativo de los egresados universitarios. Los niveles remunerativos que pueden percibir los egresados, pueden ser modelados con técnicas de regresión que permitan predecir e identificar las variables (factores) que expliquen mejor dichos niveles remunerativos. Así mismo, al considerar las variables relacionadas con el ingreso de los egresados, se deben considerar además de las variables académicas (rendimiento académico), las variables que midan las competencias adquiridas en la universidad (habilidades, destrezas, etc.), socioeconómicas (ingreso familiar, nivel educativo de los padres, etc.) y demográficas (sexo, edad, lugar de residencia, tipo de institución educativa, etc.).

En el tratamiento del problema del empleo de los egresados universitarios se busca identificar los factores más influyentes por lo cual se ha convertido en un campo de investigación en el dominio educativo. En algunos estudios las variables son agrupadas en factores académicos, factores socioeconómicos como el nivel educativo de los padres, el nivel económico de la familia, el tipo de institución educativa (pública o privada), el estado civil del egresado; así también factores demográficos como el sexo, la edad, lugar de residencia, lugar de nacimiento, etc.

Según (Martínez Zamora, 2015), las remuneraciones de los egresados está en función del rendimiento académico como un efecto positivo y que es medido por un conjunto de capacidades, destrezas y conocimientos específicos que adquieren los estudiantes durante sus estudios de pre grado; teniendo un efecto directo en el desempeño laboral y en el salario obtenido del egresado. También la literatura sugiere que los componentes de la calidad de la universidad, tales como el prestigio en el mercado laboral, la adecuación de la formación profesional y el conjunto de habilidades cognitivas y no cognitivas desarrolladas durante la etapa universitaria están relacionadas con las condiciones de empleo alcanzadas como graduado (Cohodes, S. R. & Goodman, J. S., 2012), (Dale & Krueger, A. B., 2011), (Hoekstra, 2009) y (Blac & Smith, J.A., 2006). En (Forero, N. & Ramírez, M., 2008),

analizan variables socioeconómicas como la edad, el género, el lugar de residencia, el nivel educativo del padre o madre, área de conocimiento; características laborales como el tipo de vinculación con la empresa (contrato o prestación de servicios), actividad económica, posición ocupacional; y las características de la institución como el origen de la institución (oficial o no oficial), la ubicación de la institución y la acreditación de la institución; influyen en los ingresos laborales.

En la presente investigación se utilizarán los datos de la Encuesta Nacional de Egresados Universitarios y Universidades 2014, realizada por el Instituto Nacional de Estadística e Informática - INEI. En la presente investigación se define como la variable dependiente los rangos remunerativos del primer empleo de los egresados universitarios a nivel nacional. Dado que en la encuesta del INEI los datos de las remuneraciones del primer empleo se encuentran agrupados por rangos de ingresos, se hace uso de la técnica Regresión Logística Ordinal; que permitirá explicar la variable dependiente definida como el rango de ingresos del primer empleo de los egresados de naturaleza ordinal politómica en función de un conjunto de variables independientes, las cuales están agrupadas en cuatro categorías. Se ajustan tres modelos logísticos ordinales con función de enlace logit proporcional acumulado, según las categorías de las variables independientes y separadas para los egresados de universidades públicas y privadas del Perú. Se aplica el análisis de regresión logística ordinal para cada uno de los tres modelos, obteniendo los respectivos coeficientes estimados; se aplica pruebas de bondad de ajuste y procedimiento de inferencia estadística.

Los objetivos de la investigación son los siguientes:

1. Ajustar los rangos de ingresos del primer empleo de los egresados a una regresión logística ordinal para las universidades públicas y privadas.
2. Identificar las variables que más influyen para explicar los rangos de ingresos del primer empleo de los egresados de las universidades públicas y privadas.
3. Realizar el estudio comparativo de los ingresos del primer empleo de los egresados entre las universidades públicas y privadas.

II. REVISIÓN DE LITERATURA

Los diversos estudios han indagado la relación entre los niveles educativos, la experiencia de los egresados y los ingresos que perciben, otros estudios han explorado consideraciones estructurales como la cobertura. Las características socioeconómicas de la familia y las posibilidades que esta tiene para invertir en capital humano, más precisamente en educación anterior a la universitaria, son reconocidas como fundamentales en el ciclo de la vida que afectan el aprendizaje, la productividad y el retorno de la educación.

2.1 Panorama de los ingresos de los egresados universitarios

El aumento reciente de las tasas de subempleo entre los egresados universitarios de todo el mundo ha motivado dudas sobre los rendimientos académicos reales que adquieren durante su formación universitaria. Un ejemplo es el subempleo en el Perú, el cual ha mostrado una persistente tendencia al alza equivalente a once puntos porcentuales en el periodo de análisis 2004-2012, a pesar del contexto de expansión económica del país, con una tasa de crecimiento del PBI de alrededor del 5,6% durante más de 10 años (2000-2012). Si bien la economía ha tenido un comportamiento global positivo, se han vuelto más comunes las dificultades que enfrentan los egresados universitarios a fin de obtener puestos de trabajo que les permitan ejercer su formación profesional o que estén a la altura de las expectativas remunerativas de un profesional en el medio peruano.

En (McGuinness, 2006) se realizó un estudio sobre la relación entre el subempleo y la calidad universitaria con información de los ingresantes universitarios de Irlanda del Norte, estableciendo un índice de calidad universitario. Los resultados señalan que las ganancias promedio de asistir a una universidad de mejor calidad tienden a ser limitadas. Asimismo, los impactos hallados sobre la calidad del empleo e ingresos salariales se limitaron a graduados de disciplinas particulares o a aquellos que obtuvieron títulos de universidades consideradas de “Buena Calidad”. Se identificó para la mayoría de graduados que la carrera estudiada y la clasificación del título obtenido en la universidad son los que más influyen para conseguir un nivel de ingreso en el mercado laboral.

En (Milla, 2012), se estiman los retornos salariales según el prestigio y la calidad de las universidades canadienses usando dos rankings: Best Overall Reputation de la revista Maclean y un ranking de calidad de las universidades. Se utiliza la información de los salarios por hora percibidos en el primer trabajo de graduados de pre grado a través de la

Encuesta de Jóvenes en Transición. La investigación concluye que el ingreso por graduarse en una universidad ubicada en el 25% superior del ranking de reputación, en comparación a una ubicada en el 25% inferior, es de 15.2% mayores. Así, los ingresos en universidades con mayor prestigio son en general mayores para los varones que para las mujeres.

En (Forero & Ramírez, M., 2008) se realizó un primer estudio en Colombia con el fin de identificar cuáles variables socioeconómicas (la edad, el género, el lugar de residencia, el nivel educativo del padre o madre, área de conocimiento); las características laborales, la actividad económica, la posición ocupacional y las características de la institución como el origen de la institución (oficial o no oficial), la ubicación de la institución y la acreditación de la institución que influyen en los ingresos laborales. Como resultado de este estudio se identificó que las variables más significantes son el género, la edad, el lugar de residencia, el nivel educativo de los padres, el área de conocimiento del egresado, la actividad económica de la empresa donde labora, el tipo de vinculación con la empresa y si la institución de egreso se encuentra acreditada.

En (Baron, 2010), se analizó la probabilidad que tiene un recién egresado de educación superior de pregrado en Colombia para encontrar un empleo formal y su posibilidad de un nivel de salario, haciendo énfasis en diferencias regionales por área del conocimiento y considerando las variables género y tipo de universidad. El estudio se basa en la información que arroja la encuesta de los graduados en educación del Ministerio de Educación Nacional en el Observatorio Laboral para la Educación (OLE), que utiliza únicamente información sobre aquellos egresados de programas presenciales que obtuvieron el título en 2007 y que hicieron aportes al sistema de seguridad social en 2008. La muestra para el análisis de probabilidad de empleo es de 86.846 registros y para el de salario 49.268. Los resultados del estudio mostraron que hay diferencias económicas y estadísticamente significativas en la probabilidad de encontrar un empleo formal dependiendo del área del saber y de la región donde esté ubicada la Institución de Educación Superior (IES) de la cual se gradúa. Estudiar en una IES ubicada en las costas colombianas significó una menor probabilidad de encontrar un empleo formal, 20% en relación con Bogotá, a su vez los graduados en la Costa Norte devengan un 27% menos que los egresados de IES bogotanas, de otra parte, se encontró que los egresados en ciencias de la educación devengan un 28% menos que los de Ciencias Administrativas. Además se encontró que las mujeres ganan el 8.8% menos que los hombres y que los egresados de IES privadas devengan un 4.1% más que los de universidades públicas.

En (De Vries, 2014), se analizó la diferencia de la remuneración de la fuerza laboral del Reino Unido con grado universitario, en función de la universidad en la que estudiaron y el área de conocimiento; utilizando los datos provenientes de la encuesta de los egresados de educación superior, realizada por la Agencia de Estadísticas de Educación superior, considerando a todos los graduados del Reino Unido con seis meses después de su grado y pasados 3.5 años de ejercicio profesional. Para el estudio a los seis meses se dispuso de una base completa de 53.770 profesionales y corresponde a egresados 2012 - 2013 y para el de 3.5 años de 9.330 graduados en 2008 - 2009 . Para el estudio emplearon modelos de regresión de Mínimos Cuadrados Ordinarios (MCO) y modelaron los ingresos mediante regresión intervalo. Se identificaron en el trabajo grandes diferencias en los ingresos de los graduados respecto a su universidad y al área de conocimiento, los graduados de Oxford y Cambridge devengan salarios iniciales superiores en un 42% comparado con universidades denominadas post 92 (incluye escuelas politécnicas y aquellas que adquirieron el estatus de universidad después de 1992); la diferencia por área de estudio resultó ser significativa, los graduados en medicina y odontología devengan aproximadamente £12.200 más que los que estudiaron diseño y artes creativas, siendo estas las profesiones con menores ingresos.

(Adarve Paz, 2011), analiza la relación entre salarios y rendimiento académico en profesionales universitarios, controlando por factores personales, académicos y laborales. El rendimiento académico se mide por el promedio de calificaciones obtenido al finalizar el plan de estudios. Se propone que el promedio académico tiene un efecto positivo y significativo en el salario individual puesto que mide un conjunto de capacidades, destrezas y conocimientos específicos de cada programa de formación profesional, que tienen un efecto directo en el desempeño laboral del individuo, el cual se ve reflejado en el salario obtenido como profesional. La hipótesis se corroboró estadísticamente y se concluyó que existe una relación positiva entre el salario y el promedio académico. En (Xiaowei Xul & Weiwei Zhang, 2015), se menciona que el problema del empleo de los graduados universitarios es el tema candente de los estudiantes universitarios de hoy e incluso de toda la sociedad en China.

En el caso del Perú, los estudios cuantitativos de (Yamada & Castro, J.F., 2013) sugieren que a pesar que una mayor proporción de la población accede a la educación superior universitaria en el país, la calidad de los servicios educativos ha disminuido. Así, los autores no logran separar el aumento exagerado de la cobertura de la evolución de la calidad universitaria. Adicionalmente se aborda la selectividad de las universidades como una

aproximación a la calidad educativa. En (Yamada, Castro, J.F., Bacigalupo, J., & Velarde, L., 2013), se señala que a partir del año 2000 se percibe un deterioro de habilidades de los estudiantes universitarios, medidas en la Encuesta sobre habilidades y funcionamiento del mercado laboral peruano (ENHAB) 2010. Esta tendencia, unida a la menor selectividad en el ingreso por parte de instituciones superior, se propone como un importante cómplice del empeoramiento de la calidad universitaria. En (Diaz & Jaramillo, M., 2008), se estudia la relación entre el subempleo y la calidad de la educación superior. Los autores señalan que los estudios post secundarios son importantes mecanismos para el aumento de la productividad así como para la movilización social, dadas las limitaciones de la educación regular básica. No obstante, la heterogeneidad de calidad de universidades e institutos hace que persista una discrepancia entre la calificación de la fuerza laboral y la demanda de trabajo, lo cual dificulta obtener los beneficios de la educación superior.

2.2 Modelos de la regresión logística

La regresión logística se utiliza cuando se desea modelar una variable dependiente de naturaleza cualitativa que tiene dos o más categorías, en función de un conjunto de variables independientes o predictoras que pueden ser cuantitativas o cualitativas. Cuando la variable dependiente tiene sólo dos categorías se usa la regresión logística dicotómicas (binaria) y si presenta más de dos categorías se usa la regresión logística politómica. En el caso que en el modelo politómico la variable dependiente esté asociada a una variable tipo nominal (categorías sin jerarquía) se usan la regresión logística nominal y del tipo ordinal (categorías con jerarquía) la regresión logística ordinal. La regresión logística pertenece a los llamados modelos de respuesta discreta, que hace alusión a la variable dependiente de naturaleza cualitativa. Los modelos de respuesta discreta, tienen la posibilidad de usar una variedad de funciones de enlace; tales como la logit (logística, probit, log-log, etc). En los modelos logísticos, la función de enlace que se usa es la logística (logit).

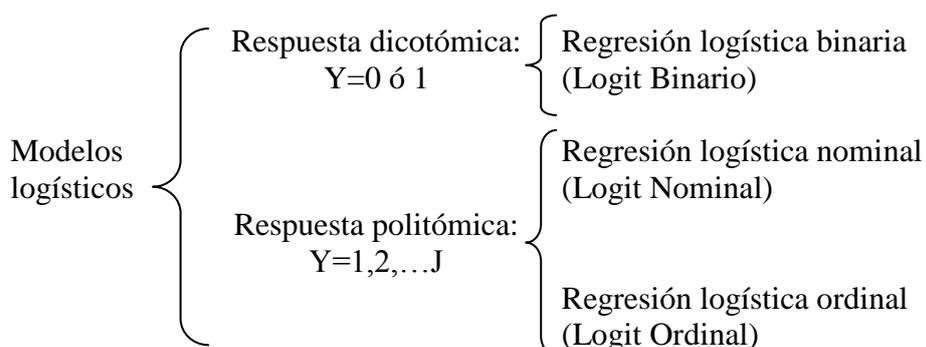


Figura 1. Tipos de modelos logísticos

En estos modelos los datos están agrupados (clasificados) en tantas categorías tenga la variable dependiente. En los modelos de respuesta nominales, se tiene (J-1) modelos logit con estimadores diferentes que deben ser interpretados (intercepto y coeficientes de regresión diferentes). Mientras en los modelos ordinales, los (J-1) modelos logit estimados se diferencian solo en el valor del intercepto. Cuando una variable independiente es cualitativa con “a” categorías, se tienen que crear (a-1) variables binarias (dummy).

Los modelos logísticos pertenecen a los llamados modelos lineales generalizados (MLG), que son una extensión del modelo lineal general y fueron propuestos por Nelder y Wedderburn (1972). Los MLG, son una unificación de los modelos de regresión lineales y no lineales, permitiendo modelar variables dependientes (respuestas) con distribución no Normal. En los MLG la variable respuesta puede ser de tipo cuantitativa (numérica) como cualitativa (categórica), por tanto pueden asumir aparte de la Normal, otras distribuciones tales como; la Binomial, Poisson, Logística, Gamma, etc. En los MLG la distribución que se asume tiene la variable respuesta pertenece a la familia exponencial.

Características de los MLG:

- La formulación de un MLG se basan en tres componentes: *Componente aleatorio* (variable respuesta), *Componente sistemático* (predictor lineal) y *Función de enlace*.
- En los MLG, el supuesto de homocedasticidad del componente aleatorio no necesariamente se debe de cumplir.
- En los MLG se supone que las observaciones son independientes (no correlacionadas). Los MLG supone un único término de error (componente aleatorio).
- Los MLG permiten un modelamiento más complejo entre la variable respuesta y las variables explicativas; a través de la selección de la función de enlace que permite relacionar el predictor lineal con la media de la variable respuesta.

2.2.1 Modelo de regresión logística binaria

Los modelos de regresión logística binaria, son aquellos donde la variable dependiente Y es cualitativa que puede tomar sólo dos valores posibles 1 o 0 (Y=1: éxito; Y=0: fracaso); generalmente Y=1 es el evento de interés y por lo cual las observaciones están clasificadas en dos categorías o grupos.

Los tres componentes definidos para el MLG son:

- **El componente aleatorio.** Está formado por el vector aleatorio $Y'=(Y_1, Y_2, \dots, Y_N)$ que representa a la variable dependiente binaria, se asume independencia entre las “N” observaciones y que pertenecen a la distribución *Binomial*(n, π) .
- **El componente sistemático.** Conocido como el predictor lineal, está compuesto por el vector de parámetros y el conjunto de variables explicativas (factores o covariables). Se define: $\eta = X' \underline{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$
- **La función de enlace.** Permite relacionar el parámetro π con el predictor lineal. La función de enlace usada es la logística (logit):

$$\log it(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \eta = X' \underline{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

El modelo de regresión de respuesta binaria, trata de explicar la variable respuesta binaria Y en términos de que tan probable suceda el evento de interés (Y=1), en función de un conjunto de predictores X_1, X_2, \dots, X_p (factores o covariables). El modelo logístico binario usa como ecuación funcional la función logística. La función Logit permite relacionar y explicar la variable respuesta binaria Y en función de un conjunto de variables predictoras. El modelo logit relaciona el logaritmo de la razón de la probabilidad Y=1 con respecto a Y=0 y con el predictor lineal. Así para, $P(Y = 1) = \pi$ y $P(Y = 0) = 1 - \pi$, se tiene:

$$\log it(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = X' \beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad \dots (2.1)$$

El modelo de regresión logístico binario, modela la probabilidad Y=1, usando la función de enlace Logit para relacionarla con el predictor lineal. El modelo logístico se expresa por:

$$P(Y = 1 / X_1, X_2, \dots, X_p) = \pi = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} \quad \dots (2.2)$$

También se puede expresar el modelo logístico por:

$$P(Y = 1 / X_1, X_2, \dots, X_p) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p))} \quad \dots (2.3)$$

El modelo estimado será:

$$P(Y = 1) = \hat{\pi}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_p X_p)} \quad \dots (2.4)$$

2.2.2 Modelos de regresión logística politómica

El modelo de regresión logística politómica, es una generalización del modelo de regresión logística binaria, donde la variable dependiente presenta más de dos categorías. Son conocidos también como modelos logísticos multinomiales, por ser una extensión de los modelos de regresión logística binaria (modelos binomiales). En estos modelos se tiene tantas ecuaciones de regresión como categorías tenga la variable dependiente. Se tiene una regresión logística ordinal, cuando las categorías correspondientes a la variable dependiente muestran un orden de importancia o jerarquía; de lo contrario se tiene una regresión logística nominal.

Formulación del modelo lineal generalizado

El modelo de una regresión de respuesta politómica con J categorías de la variable dependientes Y, está compuesto por J ecuaciones logísticas. Se tiene la variable aleatoria $Y' = (Y_1, Y_2, \dots, Y_J)$ con J categorías y $\pi_1, \pi_2, \dots, \pi_J$ que denotan las respectivas probabilidades asociadas a cada categoría, con $\pi_1 + \pi_2 + \dots + \pi_J = 1$. Considerando n el total de observaciones independientes para el vector observado $y' = (y_1, y_2, \dots, y_J)$, tal que y_1 es el número de observaciones de la categoría 1, y_2 de la categoría 2, etc. Se define el vector de observaciones: $y' = [y_1, y_2, \dots, y_J]'$, con $\sum_{j=1}^J y_j = n$. Se asume que el número de observaciones de las categorías de Y tienen una distribución multinomial. Esto se expresa como:

$$f(y/n) = \frac{n!}{y_1! y_2! \dots y_J!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_J^{y_J}$$

Además se tiene:

$$E(Y_j) = n\pi_j, \quad Var(Y_j) = n\pi_j(1 - \pi_j)$$

$$Cov(Y_j, Y_k) = -n\pi_j\pi_k$$

Los tres componentes para el MLG son:

- **El componente aleatorio.** Formado por el vector aleatorio $Y' = (Y_1, Y_2, \dots, Y_N)$ de la variable dependiente politómica y que pertenece a la distribución $Multinomial(n, \pi_1, \pi_2, \dots, \pi_J)$.
- **El componente sistemático.** Definido por el llamado predictor lineal $\eta = X' \underline{\beta}$, compuesto por el vector de parámetros y un conjunto de variables explicativas (factores o covariables).
- **La función de enlace.** Permite relacionar el vector definido por $\underline{\pi} = (\pi_1, \pi_2, \dots, \pi_J)$ con el predictor lineal. La función de enlace usada es la logit (logística).

Modelo de regresión logística nominal

El modelo de una regresión logística nominal con J categorías, está compuesto por J ecuaciones logísticas asociadas a cada π_j ; por lo cual generalmente se selecciona la primera categoría como la referencial. La categoría referencial, sirve para definir el logit de una categoría, definida como el cociente entre la categoría de interés (numerador) y la referencial (denominador). El modelo logístico nominal está compuesto por las siguientes expresiones:

$$\left. \begin{aligned} \pi_1 &= P(Y = 1 / X_1, \dots, X_p) = \frac{1}{1 + \sum_{j=2}^J \exp(x'_j \beta_j)}, \quad j = 1 \text{ (Categoría referencial)} \\ \pi_j &= P(Y_j = J / X_1, \dots, X_p) = \frac{\exp(x'_j \beta_j)}{1 + \sum_{j=2}^J \exp(x'_j \beta_j)}, \quad j = 2, \dots, J \\ \text{Con } \pi_1 + \pi_2 + \dots + \pi_J &= 1 \end{aligned} \right\} \dots (2.5)$$

Se modelan tantos modelos logísticos como categorías tenga la variable respuesta Y. Estos modelos son simultáneamente usados para estimar los parámetros, estimándose para cada variable. Se modelan (J-1) ecuaciones logit. Considerando generalmente la primera categoría como la referencial, los (J-1) ecuaciones logit son respectivamente:

$$\text{Logit}(\pi_j) = \log\left(\frac{\pi_j}{\pi_1}\right) = x'_j \beta_j, \quad \text{para } j = 2, 3, \dots, J \quad \dots (2.6)$$

2.3 Inferencia estadística en los modelos logísticos

En los modelos logísticos se pueden aplicar un conjunto de pruebas de inferencia estadística, con la finalidad de probar y evaluar la significancia de la bondad de dichos modelos.

1) Prueba de Máxima Verosimilitud

Esta prueba se basa en comparar la máxima verosimilitud entre dos modelos. La prueba es conocida también como la razón de verosimilitud y evalúa la significancia entre dos modelos, seleccionando el que mejor se ajuste a los datos. Sean los modelos L_1 (modelo reducido) y L_2 (modelo completo), el procedimiento para aplicar la prueba de máxima verosimilitud es:

- **Formulación de las hipótesis.** Las hipótesis correspondientes son:

H_0 : Los parámetros diferentes del modelo L_2 y del modelo L_1 son iguales a 0.

H_1 : Los parámetros diferentes del modelo L_2 del modelo L_1 son diferentes de 0.

- **Prueba estadística.** La prueba se basa en el estadístico G , calculado según la fórmula:

$$G = 2 \times (\ln L_1 - \ln L_2) = -2 \ln \left(\frac{L_1}{L_2} \right) \quad \dots (2.7)$$

Dónde:

L_1 = Máxima verosimilitud del modelo reducido

L_2 = Máxima verosimilitud del modelo con todas las variables independientes

La prueba G , es considerada como una prueba global donde el modelo reducido (L_1) está formado por un número menor de variables que las que están en el modelo completo (L_2). El estadístico G , sigue una distribución Chi-Cuadrado cuyo grado de libertad es igual a la diferencia del número de parámetros entre los dos modelos. Cuando el valor de G es positivo y grande podemos inferir que al menos una de las variables independientes del modelo L_2 que no se encuentran en el modelo L_1 tiene efecto sobre la variable dependiente.

2) Prueba de bondad de ajuste de la Deviance

La medida más usada para evaluar la bondad de ajuste en los modelos logísticos es conocida como la Deviance, esta es una medida de distancia entre los valores ajustados o estimados y los datos observados, y siempre es mayor o igual a cero. El procedimiento para aplicar esta prueba es:

- **Formulación de las hipótesis.** Las hipótesis que se formulan para la estadística de deviance son:

H_0 : El modelo logístico se ajusta a los datos

H_1 : El modelo logístico no se ajusta a los datos

- **Prueba estadística.** Se usa la Deviance, la cual se define como la diferencia entre el logaritmo de la función verosimilitud del modelo saturado (con N parámetros) y el modelo en investigación (con p parámetros). Se calcula con la siguiente expresión:

$$D_c = 2 \left[l(b_{\max}; y) - l(b; y) \right] \quad \dots (2.8)$$

Dónde:

b_{\max} Es la función log-verosimilitud para el modelo saturado evaluado para el estimador

$l(b_{\max}; y)$ que corresponde al vector de parámetros β_{\max} .

$l(b; y)$ Es la función de log-verosimilitud para el modelo de interés del vector estimado b para el vector de parámetros β .

El estadístico de Deviance sigue una distribución de Chi Cuadrado con grados de libertad igual a la diferencia del número de parámetros entre el modelo saturado (N) y el modelo ajustado (p).

- **La decisión estadística.** Se rechaza H_0 , si: $D_c \geq \chi^2_{(N-p)}$

Un valor pequeño de la Deviance, indica que para un número menor de parámetros, se obtiene un ajuste tan bueno como cuando se ajuste con un modelo saturado.

3) Prueba de bondad de ajuste de Pearson

Esta prueba se basa en construir una tabla de contingencia cuyas filas representan los valores observados y las columnas, los valores predichos con el modelo logístico. Esta prueba de Pearson se basa en comparar las frecuencias observadas con las predichas, de tal manera que si difieren significativamente se puede evidenciar una falta de ajuste del modelo.

- **Formulación de las hipótesis.** Las hipótesis que se formulan para la estadística de deviance son:

H_0 : El modelo logístico se ajusta a los datos

H_1 : El modelo logístico no se ajusta a los datos

- **Prueba estadística.** Se calcula el estadístico Chi-cuadrado de Pearson. Se considera que la variable dependiente tiene k categorías y que se forman m combinaciones de valores con las variables independientes. Se calcula con la siguiente expresión:

$$\chi_c^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(y_{ij} - m_j p_{ij})^2}{m_j p_{ij} (1 - p_{ij})} \sim \chi_{(N-p)}^2 \quad \dots (2.9)$$

Dónde:

- y_{ij} Es la frecuencia observada de la i-ésima categoría de la variable dependiente en la j-ésima combinación de los valores de las variables explicativas.
- p_{ij} Es la probabilidad estimada con el modelo para la i-ésima categoría de la variable dependiente en la j-ésima combinación de los valores de las variables explicativas.
- m_j Es la cantidad de elementos en la j-ésima combinación de los valores de las variables explicativas.

- **La decisión estadística.** Se rechaza H_0 , si: $\chi_c^2 \geq \chi_{(N-p)}^2$

Cuando mayor sea el valor del estadístico Chi-Cuadrado, mayor será la sospecha de una falta de ajuste del modelo.

4) Coeficiente de determinación.

Similarmente como en los modelos lineales para evaluar la bondad de ajuste se proponen en los modelos logísticos los coeficientes de determinación, denominados *Pseudo R²*. Mide la reducción proporcional en la incertidumbre debido a la inclusión de los regresores. Se han propuesto varias medidas basadas en los residuales, una medida preferida es aquella que se basa en el desvío del modelo mínimo (intercepto) y el modelo de interés. El Pseudo R² es definido por:

$$Pseudo R^2 = 1 - \frac{D(y, \hat{y})}{D(y, \hat{y}_0)} = \frac{l(b_{\min}; y) - l(b; y)}{l(b_{\min}; y)} \quad \dots (2.10)$$

Dónde:

$D(y, \hat{y})$ y $D(y, \hat{y}_0)$ son las funciones desvío de los modelos ajustados y nulo (mínimo) respectivamente. Se tiene $0 \leq R^2 \leq 1$. No decrece a medida que se adicionen regresores. Se interpreta en términos del contenido de los datos.

Otros Pseudo R^2 que no usan la función de máxima verosimilitud, sino la razón de verosimilitud (Λ). Se tiene los Pseudo R^2 de:

- Mc-Fadden, dado por la expresión: $R_{MF}^2 = 1 - \frac{\Lambda_f}{\Lambda_0}$. Su rango teórico es $0 \leq R_{MF}^2 \leq 1$, pero muy raramente su valor se aproxima a 1. Se considera un ajuste aceptable cuando se tiene valores $0.2 \leq R_{MF}^2 \leq 0.4$ y muy buenos para valores mayores a 0.4.
- Pseudo R^2 de Cox-Snel, se define: $R_{CS}^2 = 1 - \frac{n \overline{L_0}^2}{n \overline{L_f}^2} = 1 - \exp\left(\frac{\Lambda_f - \Lambda_0}{n}\right)$. Su rango teórico es $0 \leq R_{CS}^2 \leq 1 - \frac{n \overline{L_0}^2}{n \overline{L_f}^2}$.
- Pseudo R^2 de Nagelkerke, se define: $R_N^2 = \frac{R_{CS}^2}{1 - \frac{n \overline{L_0}^2}{n \overline{L_f}^2}} = \frac{1 - \exp\left(\frac{\Lambda_f - \Lambda_0}{n}\right)}{1 - \exp\left(-\frac{\Lambda_0}{n}\right)}$. Su rango de valores es $0 \leq R_N^2 \leq 1$, por lo que su interpretación es similar al clásico.
- Pseudo R^2 ajustado de Mc-Fadden, se define por: *Ajustado* $- R_{MF}^2 = 1 - \frac{0.5x\Lambda_f + k + 1}{0.5x\Lambda_0 + 1}$, siendo el número de regresores. Este permite comparar modelos con diferente número de regresores.

5) Prueba de Wald

Es una prueba empleada para evaluar individualmente si alguna variable independiente tiene una influencia estadísticamente significativa sobre la variable dependiente; a través de realizar una prueba la hipótesis de su coeficiente de regresión.

- **Formulación de la hipótesis:**

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

- **Estadístico de prueba.** Se usa la Z.

$$Z_c = \frac{b_j}{S_{b_j}}, \text{ se aproxima a una distribución normal estándar}$$

Dónde b_j y S_{b_j} son el coeficiente de regresión estimado y su error estándar asociada a la variable independiente j.

Cabe resaltar que el uso de esta prueba es recomendado cuando se tiene una muestra grande, cuando la muestra es pequeña, se recomienda el uso de la prueba de Máxima Verosimilitud.

6) **Residuales de Pearson.** Son calculadas para realizar el análisis de residuales. Son calculados por la expresión:

$$r_i = \frac{O_i - e_i}{\sqrt{e_i}} = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}$$

Dónde: \hat{y} es el valor estimado con el modelo logístico.

O_i = Frecuencia observada

E_i = Frecuencia esperada o estimada

7) **Criterio de información Akaike (AIC= Akaike information criterios).** Es una medida relativa para evaluar la bondad de ajuste de los modelos estadísticos. Se selecciona el modelo con menor AIC. En general, el AIC se calcula:

$$AIC = 2k - 2\ln(L)$$

Donde: k= Números de parámetros del modelo

L = Valor máximo de la función verosimilitud para el modelo estimado.

8) Intervalo de confianza

Construir intervalos de confianza con un nivel del (1- α)100% para cada coeficiente de regresión y para los Odds-ratio.

$$\text{Para } \beta_j : IC(\beta_j) = b_j \pm Z_{1-\frac{\alpha}{2}} S_{b_j} \quad (\text{simétrico})$$

$$\text{Para } OR_j : IC(e^{\beta_j}) = e^{b_j \pm Z_{1-\frac{\alpha}{2}} S_{b_j}} \quad (\text{asimétricos})$$

2.4 Análisis de los coeficientes en los modelos logísticos

En los modelos de regresión logísticos la interpretación de los coeficientes estimados no es directa como el caso de la regresión lineal; debido a la naturaleza del modelo no lineal. La interpretación de los coeficientes dependerá de los cambios en la escala de medida y en el tipo de la variable explicativa (factor, dicotómicas, categórica, covariable, continua, etc). Los conceptos de **Odds y Odds ratios**, también son utilizados para interpretar la explicación que aporta cada variable explicativa a la variables respuesta politómica. La interpretación de las variables explicativas, dependerán si son factores (cualitativas) o covariables (cuantitativas).

Tipos de variables explicativas:

- Variables cuantitativas (covariable)
- Variables dummy que representan variables cualitativas (factor, categórica, ordinal, etc)
- Variables dummy que representan interacciones entre variables cualitativas o cuantitativas.

Ventaja (Odds). Es la razón de las probabilidades que el evento de interés ocurra ($Y=1$) respecto a que no ocurra ($Y=0$). Indica la preferencia de elegir la opción ($Y=1$) en la variable dependiente respecto a no elegirla ($Y=0$). Se denota por:

$$Odds = \frac{P(Y=1)}{P(Y=0)} = \frac{P(Y=1)}{1-P(Y=1)} = \frac{\pi}{1-\pi} = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad \dots (2.11)$$

El logit es el logaritmo de los Odds. Tomando logaritmo neperiano, se obtiene una expresión lineal para el modelo:

$$Logit(Odds) = Logit\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad \dots (2.12)$$

Cociente de ventaja (Odds ratios). El cociente o razón de ventaja, es una medida que indica que tan probable (o improbable) para la variable dependiente, se presente el evento de interés ($Y=1$) con respecto a que no se presente ($Y=0$), y considerando un valor determinado de una variable explicativa.

$$\text{El Odds ratios para la variable } X_j : OR_j = e^{\hat{\beta}_j}$$

Indica como la razón de ventaja al observar $Y=1$, cambia ante un incremento unitario en la variable X_j .

$$\begin{cases} \text{Si } \hat{\beta}_j > 0 \Rightarrow e^{\hat{\beta}_j} > 1, \text{ la variable } X_j \text{ incrementa la razón de ventaja para el valor } Y = 1. \\ \text{Si } \hat{\beta}_j < 0 \Rightarrow e^{\hat{\beta}_j} < 1, \text{ la variable } X_j \text{ disminuye la razón de ventaja para el valor } Y = 1. \end{cases}$$

La determinación del cociente de ventaja, para el caso de una variable explicativa dicotómica ($Y=1$ y $Y=0$) y una variable explicativa dicotómica ($X=1$ y $X=0$), se define por:

$$Odds_1 = \frac{P(Y=1/X=1)}{P(Y=0/X=1)} = \frac{1/e^{-(\hat{\beta}_0 + \hat{\beta}_1)}}{1 - 1/e^{-(\hat{\beta}_0 + \hat{\beta}_1)}} = e^{\hat{\beta}_0 + \hat{\beta}_1}$$

$$Odds_0 = \frac{P(Y = 1 / X = 0)}{P(Y = 0 / X = 0)} = \frac{1 / e^{-(\hat{\beta}_0)}}{1 - 1 / e^{-(\hat{\beta}_0)}} = e^{\hat{\beta}_0}$$

$$Entonces : OR = \frac{Odds_1}{Odds_0} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1}}{e^{\hat{\beta}_0}} = e^{\hat{\beta}_1}$$

El valor de OR, indica que tan probable (aumente) o improbable (disminuya) es que se presente el evento de interés $Y=1$, cuando la variable explicativa toma el valor $X=1$ con respecto a que tome el valor $X=0$.

2.5 La regresión logística ordinal

La regresión logística ordinal se aplica cuando la variable cualitativa dependiente es de naturaleza ordinal; esto es, se puede establecer una jerarquía entre sus diversas categorías.

2.5.1 Formulación del modelo logístico ordinal

En el modelo logístico ordinal, se emplea logit como función de enlace que permite relacionar el predictor lineal a las variables explicativas con la razón de probabilidad entre la probabilidad acumulada hasta la categoría i de la variable ordinal, y la probabilidad que la variable tome un valor mayor que la categoría i . Sea la variable respuesta Y , definiendo J el número de categorías ordinales. En (McCullagh, P. & Nelder, J.A., 1980) se plantea que la función de enlace logit es más adecuada para analizar datos ordinales, cuya distribución de frecuencia es uniforme a lo largo de todas las categorías.

La regresión logística ordinal, se aplica cuando existe un orden o jerarquía entre las distintas categorías de la variable dependiente Y . Se modela un único logit que recoge la relación de tendencia entre la variable dependiente y las variables explicativas. Existen varios modelos logísticos ordinales posibles que se pueden aplicar, según interese modelar la tendencia de la variable dependiente.

Cuando la variable dependiente Y es ordinal, se usan los modelos logit acumulados. Para una variable dependiente Y con J categorías y con sus respectivas probabilidades $\pi_1, \pi_2, \dots, \pi_j$; la probabilidad acumulada de Y es la probabilidad de que Y sea menor o igual a un valor determinado de sus categorías. Así, para una categoría dada j , se define la probabilidad acumulada de Y :

$$P(Y \leq j) = \pi_1 + \pi_2 + \dots + \pi_j, \quad \text{para: } j = 1, 2, \dots, J \quad \dots (2.13)$$

Las probabilidades acumuladas para cada categoría indican el orden entre las categorías:

$$P(Y \leq 1) \leq P(Y \leq 2) \leq \dots P(Y \leq J) = 1$$

Los logits de las probabilidades acumuladas son expresadas por:

$$\log \text{it } \mathbf{P}(Y \leq j) \equiv \log \left(\frac{P(Y \leq j)}{1 - P(Y \leq j)} \right) = \log \left(\frac{\pi_1 + \pi_2 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right) = \log \left(\frac{\sum_{i=1}^j \pi_i}{\sum_{i=j+1}^J \pi_i} \right), \quad j = 1, \dots, J - 1$$

El modelo logit acumulativo, cuando se tiene p variables independientes $x = (x_1, x_2, \dots, x_p)$, se expresa como:

$$\log \text{it } \mathbf{P}(Y \leq j) / x \equiv \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p = \alpha_j + \sum_{i=1}^p \beta_i x_i, \quad j = 1, 2, \dots, J - 1 \quad \dots (2.14)$$

La expresión anterior, indica que cada variable independiente tiene sólo coeficiente que no depende del valor j. La dependencia con el valor j se establece sólo en el coeficiente o constante α_j .

$$\log \left(\frac{P(Y \leq j)}{P(Y > j)} \right) = \log \left(\frac{\pi_1 + \pi_2 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p; \quad j = 1, 2, \dots, J - 1 \quad \dots (2.15)$$

El modelo (2.15) refleja las probabilidades acumuladas de la variable dependiente, es conocido también como razón de probabilidades proporcional, puesto que los coeficientes de regresión son independientes de las categorías de la variable dependiente, teniendo los mismos coeficientes de regresión y con un intercepto diferente para cada ecuación una de las J-1 ecuaciones.

2.5.2 Modelos logísticos ordinales

A partir de la expresión (2.15) y según (Dobson, 2002) se pueden definir varios modelos logísticos ordinales, que se diferencian en la forma que se relacionan las probabilidades asignadas a cada categoría de la variable dependiente. Entre los modelos se tiene:

- 1) **Modelo logit acumulado.** El modelo logístico ordinal más usado para modelar la variable dependiente con j categorías, es usar el logit acumulativo. Para la categoría j, el logit acumulativo se define:

$$\log\left(\frac{\pi_1 + \pi_2 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J}\right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p; \quad j = 1, 2, \dots, J-1 \quad \dots (2.16)$$

2) Modelo de Odd proporcional. Suponiendo que en el predictor lineal solo el intercepto depende de la categoría j. El modelo de odds proporcionales es:

$$\log\left(\frac{\pi_1 + \pi_2 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J}\right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p; \quad j = 1, 2, \dots, J-1 \quad \dots (2.17)$$

Este modelo se basa en el supuesto que los efectos de las variables predictoras son iguales para cualquiera de las categorías.

3) Modelo de categorías adyacentes. Se puede considerar ratios de probabilidad para categorías consecutivas:

$$\frac{\pi_1}{\pi_2}, \frac{\pi_2}{\pi_3}, \dots, \frac{\pi_{J-1}}{\pi_J}$$

El modelo logit de categorías adyacentes es:

$$\log\left(\frac{\pi_j}{\pi_{j+1}}\right) = X'_j \beta_j = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p; \quad j = 1, 2, \dots, J-1 \quad \dots (2.18)$$

Se asume que el efecto de cada variable predictora es el mismo para categorías adyacentes.

2.5.3 Prueba de líneas paralelas

El modelo de regresión logística ordinal no supone normalidad, homocedasticidad ni incorrelación de los residuos. La única suposición del modelo sobre los datos a modelar es la de odds proporcionales o líneas paralelas ya que del modelo se deduce:

$$OR_{\Delta X_k=c} Y \leq y_j = e^{\beta k^c}, \quad \forall j = 1, \dots, g-1 \quad \dots (2.19)$$

La expresión (2.18) representa para un modelo logístico con J categorías, J-1 ecuaciones (siendo la categoría mayor la referencial). Este modelo de regresión logístico ordinal se denomina “modelo logit acumulado” puesto que representa las probabilidades acumuladas de la variable respuesta dado los valores de las variables explicativas. También es llamado “modelo de razón de probabilidad proporcional” puesto que los coeficientes de regresión son independientes de las categorías de la variable dependiente, siendo las mismas en la J-1 ecuaciones que se modelan. Esto implica suponer que la relación entre las variables explicativas y la variable dependiente ordinal, es independiente de las categorías de la

última, y por lo tanto los cambios en las variables explicativas provocan el mismo cambio en la razón de probabilidad acumulada de todas las categorías. Este supuesto debe comprobarse a través de la prueba de líneas paralelas. Por lo tanto, existen $J-1$ ecuaciones con coeficientes de regresión iguales y diferente intercepto.

Se puede probar la suposición de odds proporcionales a través de la siguiente hipótesis:

$$H_0: \beta_{1r} = \dots = \beta_{J-1r} \quad \forall r = 1, \dots, k$$

H_1 : *Al menos uno es diferente*

Como estadísticos de prueba se pueden utilizar la prueba de razón de verosimilitud y la prueba de Chi cuadrado de Wald.

2.6 Comparación de medias para muestras independientes (Prueba t)

La distribución T-student se asemeja en gran medida a la distribución normal. Tiene como parámetros la media, la varianza y además incorpora a través de los grados de libertad una modificación que permite flexibilizar las colas en función del tamaño que tenga la muestra. A medida que se reduce el tamaño muestral, la probabilidad acumulada en las colas aumenta, siendo así menos estricta de lo cabría esperar en una distribución normal.

- **Supuestos**

- **Independencia:** Las observaciones tienen que ser independientes, para ello el muestreo debe ser aleatorio.
- **Normalidad:** Las poblaciones de donde se extrae las muestras deben seguir una distribución normal $N(\mu, \sigma^2)$. En caso de cierta asimetría los t-test son considerablemente robustos cuando el tamaño de las muestras es mayor o igual a 30.
- **Igualdad de varianza (homocedasticidad):** la varianza de las poblaciones e donde se extraen las muestras deben ser homogéneas.

- **Formulación de la hipótesis:** La hipótesis que se formula para la prueba T es:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

- **Prueba estadística.** Se calcula el estadístico T. Este estadístico es el valor que se calcula a partir de la muestra y que se quiere extrapolar a la población de origen, en este caso la diferencia de medias ($\bar{x}_1 - \bar{x}_2$)

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\widehat{S}_1^2}{n_1} + \frac{\widehat{S}_2^2}{n_2}}} \sim t_{(n_1+n_2-2)} \quad \dots (2.20)$$

Dónde:

\widehat{S} Es la cuasi desviación típica muestral o desviación típica muestral corregida.

- **La decisión estadística.** Se rechaza H_0 , si: $T_c \geq t_{(n_1+n_2-2, \alpha/2)}^2$

III. MATERIALES Y MÉTODOS

Los materiales y métodos que se usarán en la presente tesis son los siguientes.

3.1 Materiales

Los materiales y equipos de los cuales se usó en la presente tesis son los siguientes:

- 1) Una computadora personal Intel® Core™ i7. CPU 3.5 GHz. RAM de 4.00 GB.
- 2) Programa estadístico R y SPSS Ver. 20.0
- 3) Una impresora inyectora HP.

3.2 Metodología

La metodología propuesta para presente tesis es la siguiente:

3.2.1 Población y Muestra

Población. Todos los estudiantes egresados de las universidades públicas y privadas del Perú en el año 2014 (INEI: Encuesta Nacional de Egresados Universitarios y Universidades 2014)

Muestra: 10564 estudiantes egresados de las universidades públicas y privadas del Perú en el año 2014.

3.2.2 Descripción de las variables

Las variables para el presente estudio estarán definidas en el cuestionario que se aplicó en la Encuesta de Egresados Universitarios 2014 por el INEI-Perú. La variable dependiente (Y) se define como el rango de los ingresos del primer empleo de los egresados. El conjunto de variables independientes para el modelar los rangos de ingresos de los egresados de la educación superior, están categorizadas en cuatro grupos. El grupo A está compuesto por las variables socio-académicas que pueden influir en el ingreso de los egresados. En el cuestionario se han establecido 12 competencias que influyen en el ingreso de los egresados. El grupo B se refiere a la calificación sobre la preparación recibida en la universidad para desarrollar las competencias y el grupo C se refiere a la importancia de las competencias para su experiencia laboral. El grupo D consideran la calificación sobre aspectos de los profesores de su carrera profesional. A continuación se especificarán los grupos y sus

respectivas variables, con la finalidad de facilitar la lectura y la descripción se está siguiendo la nomenclatura que está indicada en el mismo cuestionario.

Variable Dependiente:

Y = Rango de ingreso del primer empleo de los egresados universitarios (1=Bajo, 2=Medio, 3=Alto, 4=Muy alto)

Variables independientes:

Grupo A. Variables Socio-académicas

X1=P11: Sexo (1=Hombre, 2=Mujer)

X2=P205: Padre con el estudio más alto alcanzado (1=Primaria, 2=Secundaria y 3=Superior)

X3=P324: ¿Pertenece a algún cuadro de méritos? (1=Si, 2=No)

X4=P330: ¿Ha obtenido el título profesional? (1=Si, 2=No)

X5=P349: ¿Realizó prácticas pre profesionales? (1=Si, 2=No)

X6=P456: ¿Su primer empleo estuvo relacionado con su formación profesional? (1=Si, 2=No)

X7=P471: ¿Formación de la universidad ayudó a obtener primer empleo? (1=Si, 2=No)

Grupo B. P603: ¿Cómo calificaría Ud. la preparación recibida en la universidad para el desarrollo de las competencias (1 = Excelente, 2 = Bueno, 3 = Regular, 4 = Malo).

P603_i = La calificación de la preparación recibida en la universidad para el desarrollo de la competencia i-ésima. Para $i=1, \dots, 12$.

Grupo C. P604: ¿Qué tan importantes considera Ud. las siguientes competencias en su experiencia laboral (1 = Muy importante, 2 = Importante, 3 = Poco importante, 4 = Nada importante).

P604_i = Importancia que considera en la competencia i-ésima en su experiencia laboral. Para $i=1, \dots, 12$.

Las variables (competencias) que se han definido para P603_i y P604_i son las siguientes:

1. Dominar tu área de disciplina?
2. Rendir bajo presión y cumplir ciertos objetivos?

3. Coordinar actividades?
4. Trabajar en equipo?
5. Tener liderazgo?
6. Utilizar herramientas informáticas básicas?
7. Utilizar software específico de la carrera?
8. Presentar en público productos, ideas o informes?
9. Redactar informes o documentos?
10. Leer, escribir y hablar en idioma(s) extranjero(s)?
11. Conocimientos básicos de otros campos o disciplinas?
12. Tener pensamiento crítico?

Grupo D. P605: ¿Cómo calificaría Ud. los siguientes aspectos de la universidad de la cuál egresó (1 = Excelente, 2 = Bueno, 3 = Regular, 4 = Malo), sobre los profesores de tu carrera.

P605_i = La calificación en el aspecto i-ésimo sobre los profesores de su carrera. Para $i=1, \dots,$

5. Las variables definidas son:

1. La calidad de los profesores?
2. El grado de acceso a ellos?
3. El grado de actualización de conocimientos?
4. La preparación de clases?

3.2.3 Modelos logísticos ordinales propuestos

Se propone ajustar los datos del rango de ingreso del primer empleo de los egresados de las universidades públicas y privadas del Perú a tres modelos logísticos. Se ajustarán los datos a una regresión logística ordinal con logit acumulados. Los logits de las probabilidades acumuladas son expresadas por:

$$\log it \mathbf{P}(Y \leq j) \equiv \log \left(\frac{P(Y \leq j)}{1 - P(Y \leq j)} \right) = \log \left(\frac{\pi_1 + \pi_2 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right) \quad j = 1, 2, \dots, 3 \quad \dots (3.1)$$

$$Con: \quad \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$$

El modelo logit acumulativo, cuando se tiene p variables independientes $x = (x_1, x_2, \dots, x_p)$, se expresa como:

$$\log it \mathbf{P}(Y \leq j / x) \equiv \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p = \alpha_j + \sum_{i=1}^p \beta_i x_i, \quad j = 1, 2, 3 \quad \dots (3.2)$$

La expresión (3.2), indica que cada variable independiente tiene sólo coeficiente que no depende del valor j . La dependencia con el valor j se establece sólo en el coeficiente o constante α_j , los cuales aumentan junto con los valores de j . Usando las expresiones (3.1) y (3.2), se obtienen las $J-1$ ecuaciones logísticas ordinales, las cuales se expresan por:

$$\frac{P(Y \leq j/x)}{P(Y > j/x)} = \begin{cases} J = 1, & \frac{P(Y \leq 1/x)}{P(Y > 1/x)} = \exp(\alpha_1 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \\ J = 2, & \frac{P(Y \leq 2/x)}{P(Y > 2/x)} = \exp(\alpha_2 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \\ J = 3, & \frac{P(Y \leq 3/x)}{P(Y > 3/x)} = \exp(\alpha_3 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \end{cases} \quad \dots (3.3)$$

La expresión (3.3), muestra las $J-1$ ecuaciones predictivas que serán estimadas. Se modelan la razón de las probabilidades acumuladas de la categoría j . Los coeficientes de regresión (pendientes) son similares pero con interceptos diferentes. Se hace referencia a la variable dependiente, siendo las probabilidades condicionales, es decir, dado los valores de la variable independiente. Además, los coeficientes cuantifican el efecto de las variables independientes sobre el logaritmo de la razón de probabilidades. Los tres modelos logísticos ordinales son formulados a continuación:

Modelo 1. Explicar el rango de ingreso de los egresados universitarios (Y) en función de las variables socio-académicas (Grupo A) y la calificación sobre la preparación recibida en la universidad para el desarrollo de las competencias (Grupo B). La variable Y está representada por los cuatro rangos de ingreso.

$$\log it \mathbf{P}(Y \leq j/x) \bar{=} \alpha_j + \sum_{i=1}^7 \beta_i X_i + \sum_{i=1}^{12} \beta_i P603_i, \quad para : j = 1,2,3 \quad \dots (3.4)$$

Modelo 2. Explicar el rango de ingreso de los egresados universitarios (Y) en función de las variables socio-académicas (Grupo A) y la importancia sobre las competencias en su experiencia laboral (Grupo C).

$$\log it \mathbf{P}(Y \leq j/x) \bar{=} \alpha_j + \sum_{i=1}^7 \beta_i X_i + \sum_{i=1}^{12} \beta_i P604_i, \quad para : j = 1,2,3 \quad \dots (3.5)$$

Modelo 3. Explicar el rango de ingreso de los egresados universitarios (Y) en función de las variables socio-académicas (Grupo A) y la calificación de aspectos de la universidad sobre los profesores de la carrera (Grupo D).

$$\log it \mathbf{P}(Y \leq j/x) = \alpha_j + \sum_{i=1}^7 \beta_i X_i + \sum_{i=1}^5 \beta_i P605_i, \quad \text{para } j = 1, 2, 3 \quad \dots (3.6)$$

3.2.4 Análisis de la regresión logística ordinal

Se propone ejecutar las siguientes etapas para el proceso del análisis de regresión con los datos.

1. Recopilación de datos

Se recuperará la base de datos de la Encuesta Nacional de Egresados Universitarios y Universidades 2014 desde el portal del INEI. Luego se procederá a fusionar las tablas de datos en una base de datos que integre todas las variables relevantes para de estudio. Los módulos de la base de datos de la ENEU y sus respectivas variables que se consideran en el estudio se presentan a continuación:

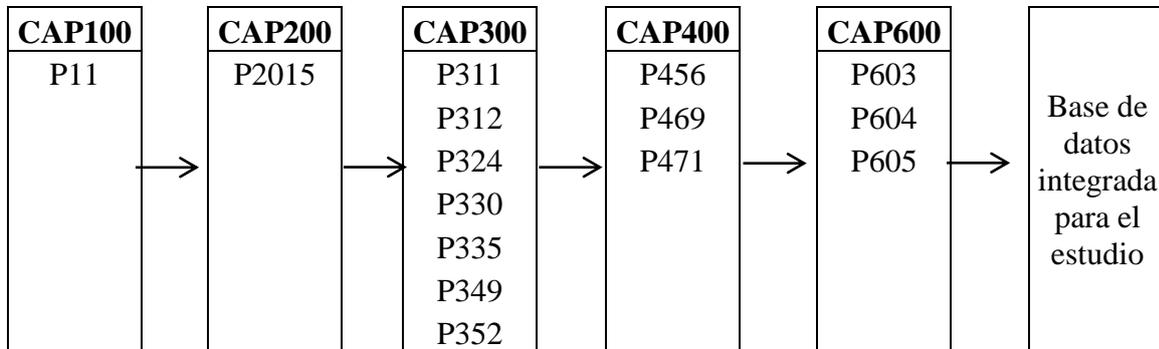


Figura 2. Tablas para la base de datos del estudio

2. Pre procesamiento de datos

Con la finalidad de preparar los datos para aplicar la regresión logística ordinal, se aplicarán técnicas para la limpieza (manejos de datos atípicos y datos faltantes) y la transformación de los datos (discretización y recodificación de datos).

- **Manejo de datos faltantes.** Para los datos faltantes se procederá a eliminar aquellos datos missing y con respuesta “No Sabe” en las variables independientes.

- **Manejo de datos atípicos.** Se aplicará el procedimiento de diagrama de cajas. Se calcularán los límites de seguridad inferior y superior para la variable ingreso y luego se eliminarán los datos outliers inferiores y superiores respectivamente.
- **Transformación de datos.** Para la variable ingreso del egresado, se procederá a obtener cuatro intervalos de clase con igual tamaño que definirán los rangos de ingreso y las respectivas categoría de la variable dependiente Y, ver Cuadro 1. También se procederá a recodificar algunas variables con el propósito de evitar categorías con porcentajes muy bajos, como el nivel de educación alcanzado por el padre y otras variables necesarias.

Cuadro 1. Rangos de ingreso de los egresados universitarios

Categoría J	Rango de ingreso
1	Bajo
2	Medio
3	Alto
4	Muy alto

3. Análisis de la regresión para los modelos logísticos ordinales

Se realizará el análisis de regresión para los tres modelos logísticos ordinales propuestos, con la finalidad de identificar las variables más significativas que expliquen los rangos de ingresos de los egresados universitarios.

- **Estimación de los coeficientes de regresión**

Se estima los coeficientes de regresión y sus respectivos Odds aplicando la regresión logística ordinal con logit proporcional acumulativo. Para la interpretación de los coeficientes de regresión se considera el concepto de ventaja (Odds) y razón de ventaja (Odds Ratios).

Ventaja (Odds). Se define como la razón entre la probabilidad de que el rango de ingreso de un egresado sea menor o igual a la categoría j ($P(Y \leq j)$) sobre la probabilidad que sea mayor a la categoría j ($P(Y > j)$). Esto es:

$$Odds = \frac{P(Y \leq j)}{P(Y > j)} = \frac{P(Y \leq j)}{1 - P(Y \leq j)} = \frac{\pi_1 + \pi_2 + \dots + \pi_j}{1 - (\pi_1 + \pi_2 + \dots + \pi_j)}, \quad \dots j = 1, 2, 3, 4$$

Cociente de ventaja (Odds ratios). Mide que tan probable (o improbable) que el rango de ingreso del egresado sea menor o igual a la categoría j ($Y \leq j$) con respecto a que sea mayor a dicha categoría ($Y > j$), cuando la variable independiente toma el valor $X=1$ con respecto a que tome el valor $X=0$ (categoría referencial). Se define el cociente de ventaja o cociente de Odds acumulado (OR) para la variable X_i , como:

$$OR_i = \frac{Odds_1}{Odds_0} = \frac{\frac{P(Y \leq j / X = 1)}{P(Y > j / X = 1)}}{\frac{P(Y \leq j / X = 0)}{P(Y > j / X = 0)}} = e^{\beta_i}, \quad \dots j = 1, 2, 3, 4$$

Si $\hat{\beta}_i > 0 \Rightarrow e^{\hat{\beta}_i} > 1$ Entonces es más probable que el rango de ingreso del egresado sea mayor o igual a la categoría j , cuando la variable independiente toma el valor $X=1$ que cuando es $X=0$.

Si $\hat{\beta}_i < 0 \Rightarrow e^{\hat{\beta}_i} < 1$ Entonces es menos probable que el rango de ingreso del egresado sea mayor o igual a la categoría j , cuando la variable independiente toma el valor $X=1$ que cuando es $X=0$.

▪ Inferencia estadística

Se realizan pruebas de hipótesis a nivel global para todo el modelo con todas las variables independientes e individualmente para cada coeficiente de regresión, para lo cual se obtienen las estadísticas de prueba de máxima verosimilitud (prueba G) y prueba Wald (pruebas Z). Así mismo, se aplicaron y calcularon pruebas para evaluar la bondad de ajuste de cada uno de los modelos propuestos. A continuación se describen los métodos de inferencia que se aplicarán en términos del dominio de estudio.

Prueba de Máxima Verosimilitud

La prueba de máxima verosimilitud se aplicará con la finalidad de probar la hipótesis de la significación global de los modelos de regresión logística ordinal propuestos en esta investigación. Esto es, evaluar si el conjunto de variables independientes tienen un efecto estadísticamente significativo sobre el rango de ingreso de los egresados.

- **Formulación de las hipótesis.** Las hipótesis correspondientes son:

Ho: El modelo logístico ordinal con sólo intercepto

H₁: El modelo logístico ordinal con las variables independientes

- **Prueba estadística.** La prueba se basa en el estadístico G, calculado según la fórmula:

$$G = -2Ln \left(\frac{\text{Verosimilitud del Modelo}_i \text{ sólo con intercepto}}{\text{Verosimilitud del Modelo}_i \text{ con todas las variables}} \right) \approx \chi_p^2$$

Cuando las predicciones de la variable dependiente con el modelo logístico ordinal con todas las variables independientes superan las predicciones el modelo sólo con el intercepto, el valor de G tiende a ser grande y conlleva a concluir que al menos una de las variables explicativas tiene efecto sobre la variable dependiente del rango de ingreso y por lo tanto la probabilidad de que ocurra algún rango de ingreso, varía para alguna de las combinaciones de los valores de las variables independientes.

Prueba de bondad de ajuste de la Deviance

La prueba de bondad de ajuste, se basa en medir las diferencias entre el conjunto de los valores observados y los predichos con el modelo logístico ordinal.

- **Formulación de las hipótesis.** Las hipótesis que se formulan para la estadística de deviance son:

H₀ : El modelo logístico ordinal se ajusta a los datos

H₁ : El modelo logístico ordinal no se ajusta a los datos

- **Prueba estadística.** Se usa la Deviance, la cual se define como la diferencia entre el logaritmo de la función verosimilitud del modelo saturado (con N parámetros) y el modelo en investigación (con p parámetros) (Dobson, 2002). Se calcula con la siguiente expresión:

$$D_c = 2 \left[l(b_{\max}; y) - l(b; y) \right] \sim \chi_{(N-p)}^2$$

Dónde:

b_{\max} Es la función log-verosimilitud para el modelo logístico ordinal saturado evaluado para el estimador $l(b_{\max}; y)$ que corresponde al vector de parámetros β_{\max} .

$l(b; y)$ Es la función de log-verosimilitud para el modelo de interés del vector estimado b para el vector de parámetros β .

El estadístico de Deviance sigue una distribución de Chi Cuadrado con grados de libertad igual a la diferencia del número de parámetros entre el modelo saturado (N) y el modelo ajustado (p).

- **La decisión estadística.** Se rechaza H_0 , si: $D_c \geq \chi^2_{(N-p)}$

Un valor pequeño de la Deviance, indica que para un número menor de parámetros, se obtiene un ajuste tan bueno como cuando se ajuste con un modelo saturado.

Prueba de bondad de ajuste de Pearson

Esta prueba se basa en construir una tabla de contingencia cuyas filas representan los valores de los rangos de ingreso y las columnas los valores predichos con el modelo logístico ordinal.

- **Formulación de las hipótesis.** Las hipótesis que se formulan para probar los modelos logísticos propuestos es:

H_0 : El modelo logístico ordinal se ajusta a los datos

H_1 : El modelo logístico ordinal no se ajusta a los datos

- **Prueba estadística.** Se calcula el estadístico Chi-cuadrado de Pearson. Se considera que la variable dependiente como el rango de ingreso con k categorías y que se forman m combinaciones de valores con las variables independientes (Dobson, 2002). Se calcula con la siguiente expresión:

$$\chi_c^2 = \sum_{i=1}^k \sum_{j=1}^m \frac{(y_{ij} - m_j p_{ij})^2}{m_j p_{ij} (1 - p_{ij})} \sim \chi^2_{(N-p)}$$

Dónde:

y_{ij} Es la frecuencia observada de la i-ésima categoría de la variable dependiente en la j-ésima combinación de los valores de las variables explicativas.

p_{ij} Es la probabilidad estimada con el modelo para la i-ésima categoría de la variable dependiente en la j-ésima combinación de los valores de las variables explicativas.

m_j Es a cantidad de elementos en la j-ésima combinación de los valores de las variables explicativas.

- **La decisión estadística.** Se rechaza H_0 , si: $\chi_c^2 \geq \chi_{(N-p)}^2$

Cuando mayor sea el valor del estadístico Chi-Cuadrado, mayor será la sospecha de una falta de ajuste del modelo.

Pseudo coeficiente de determinación.

Conocido como un *Pseudo R^2* . Mide la reducción proporcional en la incertidumbre debido a la inclusión de los regresores. Se han propuesto varias medidas basadas en los residuales, una medida preferida es aquella que se basa en el desvío del modelo mínimo (intercepto) y el modelo de interés.

$$Pseudo R^2 = 1 - \frac{D(y, \hat{y})}{D(y, \hat{y}_0)} = \frac{l(b_{\min}; y) - l(b; y)}{l(b_{\min}; y)}$$

Dónde:

$D(y, \hat{y})$ y $D(y, \hat{y}_0)$ son las funciones desvío de los modelos ajustados y nulo (mínimo) respectivamente.

Se tiene $0 \leq R^2 \leq 1$. No decrece a medida que se adicionen regresores. Se interpreta en términos del contenido de los datos.

Prueba de Wald

Es una prueba empleada para evaluar individualmente si alguna variable independiente tiene una influencia estadísticamente significativa sobre la variable dependiente; a través de realizar una prueba la hipótesis de su coeficiente de regresión.

- **Formulación de la hipótesis:**

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

- **Estadístico de prueba.** Se usa la Z.

$$Z_c = \frac{b_j}{S_{b_j}} \sim N(0,1)$$

Dónde b_j y S_{b_j} son el coeficiente de regresión estimado y su error estándar asociada a la variable independiente j.

Cabe resaltar que el uso de esta prueba es recomendado cuando se tiene una muestra grande, cuando la muestra es pequeña, se recomienda el uso de la prueba de Máxima Verosimilitud.

Intervalo de confianza

Construir intervalos de confianza con un nivel del $(1-\alpha)100\%$ para cada coeficiente de regresión de los modelos logísticos ordinales y los respectivos Odds ratio.

$$\text{Para } \beta_j : IC(\beta_j) = b_j \pm Z_{1-\frac{\alpha}{2}} S_{b_j} \quad (\text{simétrico})$$

$$\text{Para } OR_j : IC(e^{\beta_j}) = e^{b_j \pm Z_{1-\frac{\alpha}{2}} S_{b_j}} \quad (\text{asimétricos})$$

IV. RESULTADOS Y DISCUSIÓN

La presente tesis tiene como finalidad identificar las variables más relevantes que permitan explicar los rangos remunerativos del primer empleo de los egresados universitarios del Perú, usando la Encuesta Nacional de Egresados Universitarios realizada en el año 2014 por el INEI. Con esta finalidad, se propone ajustar los datos usando el modelo de logístico ordinal. Para realizar el análisis de la regresión logístico ordinal, se usará el SPSS para obtener los resultados de las estimaciones de los coeficientes de regresión y medidas para realizar pruebas de bondad de ajuste de dichos modelos. A continuación se presenta el procedimiento estadístico que se ha propuesto en la metodología para satisfacer los objetivos definidos en esta investigación.

4.1 Recopilación de datos

En esta etapa se procedió a recuperar la base de datos de la Encuesta Nacional de Egresados Universitarios 2014 (ENEU) que está en el portal del INEI. Luego se fusionó las tablas para formar la base de datos que integre todas las variables relevantes para el estudio.

4.2 Pre procesamiento de datos

Con la finalidad de preparar los datos para aplicar la regresión logística ordinal, se aplicarán técnicas del pre procesamiento de datos para la limpieza (manejos de datos atípicos y datos faltantes) y la transformación de los datos (discretización y recodificación de datos). En el Cuadro 2, se presenta la distribución de los egresados por rango de ingreso que está definida en la base de datos del INEI.

Cuadro 2. Distribución de egresados por rango de ingreso

Rangos de ingreso	Número	Porcentaje
Hasta S/. 750	1725	16,3
De S/. 751 a S/. 1000	1524	14,4
De S/. 1001 a S/. 1500	1393	13,2
De S/. 1501 a S/. 2500	623	5,9
De S/. 2501 a S/. 3500	172	1,6
De S/. 3501 a S/. 4500	59	,6
De S/. 4501 a S/. 5500	12	,1
De S/. 5501 a S/. 6500	6	,1
Más de S/. 6500	5	,0
Total	5519	52,2
Datos faltantes	5045	47,8
Total	10564	100,0

Se observa que los ingresos de los egresados universitarios se han establecido en 9 rangos de ingreso que se encuentran distribuidos en forma muy desproporcional. Se identifica que existen aproximadamente un 50% de datos faltantes. Por esta razón no se considerará este rango de ingreso para este estudio. Para el pre procesamiento se analizará la variable que contiene los valores de los ingresos de los egresados, luego se elaborará un rango de ingreso para el análisis de regresión logístico ordinal.

1) Manejo de datos faltantes (missing)

Se procedió a aplicar el procedimiento listwise, que consiste en eliminar los ingresos con valores missing. Para el caso de las variables independientes cualitativas se eliminaron los datos faltantes, con valores indefinidos (NO SABE) y con frecuencia porcentual menores al 5%.

2) Manejo de datos atípicos (outliers)

Para la detección de los datos atípicos del ingreso de los egresados universitarios, se usó el diagrama de cajas. En el Cuadro 3, se presenta medidas estadísticas para la variable ingreso.

Cuadro 3. Medidas estadísticas para el ingreso

Medida	Valor
Media	1842,66
Mediana	1500,00
Moda	1500,00
P ₂₅	1000,00
P ₅₀	1500,00
P ₇₅	2500,00
Rango	8798,00
Rango intercuartil	1500.00
Desviación estándar	1236,95

- **Cálculo de los límites de seguridad inferior y superior:**

$$LI = 1000 - 1.5 \times (2500 - 1000) = - 1250$$

$$LS = 2500 + 1.5 \times (2500 - 1000) = 4750$$

Aplicando el límite superior, se eliminaron los datos atípicos cuyos ingresos sean mayores a 4750.

3) Discretización de la variable ingreso

Como se mostró en el Cuadro 2, el rango de los ingresos de los egresados definidos en la base de datos tiene casi el 50% de datos missing y los 9 rangos de ingresos representan una distribución desproporcionada. Por consiguiente, se decidió crear un rango de ingreso con cuatro intervalos de clase. En el Cuadro 4, se presenta el procedimiento de los cálculos necesarios para hallar los límites inferiores y superiores de los cuatro intervalos de clase.

Cuadro 4. Cálculo de los intervalos de clase

Rango: $r =$	4698
Número clases: $k =$	4
TIC =	1174.5
Límite Inferior	Límite Superior
2.0	1176.5
1176.5	2351.0
2351.0	3525.5
3525.5	4700.0

Fuente: Elaboración propia.

En el Cuadro 5, se presenta la distribución final de los egresados universitarios según el rango de ingreso. El número total de egresados a ser analizados es 5464. Se han categorizado los cuatro rangos como: Bajo, Medio, Alto y Muy Alto y con los porcentajes de los egresados distribuidos en el 31.6%, 42.8%, 19.8% y 5.9% respectivamente.

Cuadro 5. Distribución de los egresados por rango de los ingresos

Rangos	Descripción	Número	Porcentaje
Ingreso ≤ 1176.5	Bajo	1724	31,6
$1176.5 < \text{Ingreso} \leq 2351.0$	Medio	2337	42,8
$2351.0 < \text{Ingreso} \leq 3525.5$	Alto	1082	19,8
$3525.5 < \text{Ingreso} \leq 4700$	Muy Alto	321	5,9
Total		5464	100,0

Fuente: Elaboración propia.

4) Transformación de datos

También se recodificó la variable nivel de estudios alcanzado por el padre, considerando sólo tres categorías (Primaria, Secundaria y Superior).

4.3 Análisis de regresión para los modelos logísticos ordinales ajustados

El análisis de regresión logístico ordinal se aplicará a cada uno de los tres modelos propuestos, según las expresiones (3.4), (3.5) y (3.6), haciendo la comparación entre los egresados de las universidades públicas y privadas del Perú.

Para la regresión logística ordinal se usó como función de enlace la logit proporcional acumulativo. Usando el programa SPSS, se realizó en primer lugar una corrida con la finalidad de identificar las variables independientes que son significativas para explicar el rango de los ingresos de los egresados universitarios. Para esto se verificó las respectivas estadísticas t para cada regresor individual. Luego se obtuvo la segunda corrida con las variables independientes significativas.

En el Cuadro 6, se presenta la distribución de los egresados según los cuatro rangos de ingresos para las universidades públicas y privadas. Se observa que los mayores porcentajes de egresados 44.1% y 41.9% corresponden a los rangos de ingresos de S/. 1176.5 a S/. 2351.0 para las universidades públicas y privadas respectivamente. Siendo los menores porcentajes para el cuarto rango de ingreso con porcentajes de 5.3% y 6.2%. En el análisis comparativo del rango de ingresos entre las públicas y privadas, se nota una ligera superioridad de los sueldos de los egresados de las universidades privadas con un 27.5% entre los S/. 2351.0 y S/ 4700.0, mientras que para los egresados de las universidades públicas corresponde en el mismo intervalo de sueldo sólo el 22.9%.

Cuadro 6. Distribución de egresados por rango de ingresos y tipo de universidad

Rango ingreso	Universidades públicas		Universidades privadas	
	Número	Porcentaje	Número	Porcentaje
Ingreso <= 1175.5	722	33,0	1002	30,6
1175.5 < Ingreso <= 2351.0	965	44,1	1372	41,9
2351.0 < Ingreso <= 3525.5	384	17,6	698	21,3
3525.5 < Ingreso <= 4700	117	5,3	204	6,2
Total	2188	100,0	3276	100,0

Fuente: Elaboración propia.

4.3.1 Análisis del Modelo 1

El respectivo modelo se expresa en (3.4), considerando las universidades públicas y privadas. Se ajusta los datos del rango de ingreso de los egresados universitarios a una regresión logística ordinal con función de enlace logit. La variable dependiente Y, es el rango de los ingresos y las variables independientes son las variables socio-académicas (Grupo A) y la calificación sobre la preparación recibida en la universidad para el desarrollo de las competencias (Grupo B).

$$\log it \mathbf{P}(Y \leq j / x) = \alpha_j + \sum_{i=1}^7 \beta_i X_i + \sum_{i=1}^{12} \beta_i P603_i, \quad \text{para } j = 1, 2, 3 \quad \dots(4.1)$$

Con la finalidad de evaluar y validar los modelos logísticos ordinales, se aplicarán medidas y pruebas de hipótesis para evaluar el buen ajuste de los modelos logísticos ordinales. Para determinar las variables significativas que explique el rango de ingreso de los egresados universitarios, se ha realizado el siguiente procedimiento:

1. Con la finalidad de identificar las categorías de las variables independientes que no sean tan relevantes, se obtuvieron tablas de frecuencias de las categorías para cada variable (Ver Anexo 1). Se identificaron y se eliminaron las categorías con frecuencias menores e iguales al 5%.
2. Con este primer filtro de selección de variables relevantes, se procedió a ajustar a la regresión logística ordinal según el Modelo 1. Con la finalidad de identificar las variables independientes significativas se aplicó la prueba de Wald. Con un nivel de significación del 5%, se seleccionaron las variables significativas para las universidades públicas y privadas.
3. Finalmente se ajustan sólo las variables significativas a la regresión logística ordinal según el Modelo 1 para las universidades públicas y privadas. Se obtuvieron los coeficientes de regresión estimados y su respectivos Odds ratio. Adicionalmente se realizaron las pruebas de bondad de ajuste y de líneas paralelas.

Prueba de Wald para el Modelo 1

Con la finalidad de probar la significación de cada coeficiente de regresión, se aplica pruebas con la estadística Wald. Las hipótesis formuladas son las siguientes:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

En el Cuadro 7 se presenta los coeficientes de regresión estimados y su significación para los datos de las universidades públicas y privadas.

Cuadro 7. Coeficientes estimados y prueba de Wald (Modelo 1)

Variables	Universidades públicas		Universidades privadas	
	Estimación	Sig.	Estimación	Sig.
[R_Ingreso = 1]	,212	,538	-,186	,618
[R_Ingreso = 2]	2,299	,000	1,766	,000
[R_Ingreso = 3]	4,041	,000	3,603	,000
[P11=1]	,706	,000	,581	,000
[P205_1=3]	-,360	,003	-,743	,000
[P205_1=4]	-,327	,000	-,633	,000
[P324=1]	,503	,000	,385	,000
[P330=1]	,588	,000	,600	,000
[P349=1]	-,244	,171	,011	,942
[P456=1]	,304	,031	,207	,089
[P471=1]	-,047	,737	-,109	,369
[P603_1=1]	,004	,984	,413	,012
[P603_1=2]	,175	,180	,262	,055
[P603_2=1]	,174	,359	,139	,397
[P603_2=2]	-,060	,665	,071	,597
[P603_3=1]	,437	,035	,266	,123
[P603_3=2]	,341	,010	,106	,428
[P603_4=1]	-,175	,396	-,431	,017
[P603_4=2]	-,244	,095	-,212	,155
[P603_5=1]	-,028	,889	,101	,547
[P603_5=2]	,070	,607	-,002	,988
[P603_6=1]	-,021	,909	-,291	,050
[P603_6=2]	-,044	,689	-,189	,101
[P603_7=1]	,144	,520	,422	,055
[P603_7=2]	,132	,440	,329	,101
[P603_7=3]	-,012	,942	,079	,696
[P603_8=1]	,114	,546	-,042	,773
[P603_8=2]	,095	,393	-,025	,816
[P603_9=1]	,081	,676	,032	,833
[P603_9=2]	-,029	,806	-,046	,693
[P603_10=1]	-,326	,206	,203	,294
[P603_10=2]	-,113	,427	,124	,448
[P603_10=3]	-,090	,477	-,103	,521
[P603_11=1]	,727	,032	,249	,413
[P603_11=2]	,491	,060	,130	,646
[P603_11=3]	,445	,080	,018	,949
[P603_12=1]	-,333	,074	-,470	,004
[P603_12=2]	-,373	,004	-,356	,006

Para las universidades públicas:

Con un nivel de significación de 0.05, existe fuerte evidencia estadística para afirmar que las variables independientes, “género” (P11), “nivel de estudios del padre” (P205_1), “perteneció al cuadro de méritos” (P324), “obtuvo el título profesional” (P330), “primer empleo relacionado con la formación profesional” (P456), “calificación de la preparación para coordinar actividades” (P603_3), “calificación de la preparación en conocimientos básicos de otros campos o disciplinas” (P603_11) y “calificación de la preparación en pensamiento crítico” (P603_12); explican la variable dependiente: rangos de ingresos del primer empleo de los egresados de las universidades públicas

No existe suficiente evidencia estadística para afirmar que las variables independientes, “realizo prácticas pre profesionales” (P349), “la formación universitaria lo ayudo a conseguir su primer empleo” (P471), “calificación de la preparación en el dominio del área de disciplina” (P603_1), “calificación de la preparación para rendir bajo presión” (P603_2), “calificación de la preparación en trabajo en equipo” (P603_4), “calificación de la preparación para tener liderazgo” (P603_5), “calificación de la preparación para utilizar herramientas informáticas básicas” (P603_6), “calificación de la preparación para utilizar software específico de la carrera” (P603_7), “calificación de la preparación para presentar en público productos, ideas o informes” (P603_8), “calificación de la preparación para redactar informes o documentos” (P603_9) y “calificación de la preparación para leer, escribir y hablar en idioma(s) extranjero(s)” (P603_10), explican la variable dependiente.

Para las universidades privadas:

Con un nivel de significación de 0.05, existe fuerte evidencia estadística para afirmar que las variables independientes, “género” (P11), “nivel de estudios del padre” (P205_1), “perteneció al cuadro de méritos” (P324), “obtuvo el título profesional” (P330), “calificación de la preparación para dominar el área de disciplina” (P603_1) , “calificación de la preparación para trabajar en equipo” (P603_4), “calificación de la preparación para utilizar herramientas informáticas básicas” (P603_6) y “calificación de la preparación para tener pensamiento crítico” (P603_12), explican la variable dependiente: rangos de ingresos del primer empleo de los egresados de las universidades privadas.

No existe suficiente evidencia estadística para afirmar que las variables independientes, “realizó prácticas pre profesionales” (P349), “primer empleo relacionado con la formación

profesional” (P456), “la formación universitaria lo ayudó a conseguir su primer empleo” (P471), “calificación de la preparación para rendir bajo presión” (P603_2), “calificación de la preparación para coordinar actividades” (P603_3), “calificación de la preparación para tener liderazgo” (P603_5), “calificación de la preparación para utilizar software específico de la carrera” (P603_7), “calificación de la preparación para presentar en público productos, ideas o informes” (P603_8), “calificación de la preparación para redactar informes o documentos” (P603_9), “calificación de la preparación para leer, escribir y hablar en idioma(s) extranjero(s)” (P603_10), y “calificación de la preparación en conocimientos básicos de otros campos o disciplinas” (P603_11), explican la variable dependiente.

Análisis de regresión logística con variables significativas para el Modelo 1

Se ajusta los datos con las variables que resultaron significativas en el Cuadro 7. Se realizará pruebas de bondad de ajuste y la prueba de líneas paralelas.

▪ Estimación de los coeficientes de regresión para el Modelo 1

En el Cuadro 8, se presenta la respectiva salida de los coeficientes de la regresión logística ordinal estimados, la significación con la estadística Wald y los respectivos Odds ratios para el Modelo 1 y para las universidades públicas y privadas.

Cuadro 8. Estimación de los coeficientes de regresión (Modelo 1)

Variables	Universidades públicas			Universidades privadas		
	Coefficientes	Sig.	exp(b)	Coefficientes	Sig.	exp(b)
[R_Ingreso = 1]	,524	,049		-,256	,114	
[R_Ingreso = 2]	2,600	,000		1,675	,000	
[R_Ingreso = 3]	4,337	,000		3,499	,000	
[P11=1]	,722	,000	2,058	,586	,000	1,796
[P205_1=3]	-,377	,002	,686	-,801	,000	,449
[P205_1=4]	-,339	,000	,712	-,661	,000	,516
[P324=1]	,515	,000	1,674	,415	,000	1,515
[P330=1]	,586	,000	1,797	,585	,000	1,795
[P456=1]	,270	,001	1,310			
[P603_1=1]				,646	,000	1,909
[P603_1=2]				,345	,006	1,412
[P603_3=1]	,517	,002	1,678			
[P603_3=2]	,340	,003	1,405			

Variables	Universidades públicas			Universidades privadas		
	Coefficientes	Sig.	exp(b)	Coefficientes	Sig.	exp(b)
[P603_4=1]				-,184	,239	,832
[P603_4=2]				-,123	,369	,885
[P603_6=1]				,045	,728	1,046
[P603_6=2]				,001	,991	1,001
[P603_11=1]	,686	,026	1,986			
[P603_11=2]	,485	,048	1,624			
[P603_11=3]	,419	,085	1,520			
[P603_12=1]	-,260	,133	,771	-,147	,312	,863
[P603_12=2]	-,332	,007	,717	-,215	,075	,807

Para las universidades públicas:

Se emplea los Odds ratio (Razón de ventaja) para explicar los rangos de ingresos del primer empleo de los egresados de las universidades públicas:

- P11=1: Para los egresados hombres, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 2.058 veces mayor que para las egresadas mujeres, dado que las otras variables se mantengan constantes.
- P205_1=3: Para los egresados cuyo padre alcanzó el nivel de educación primaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.686 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P205_1=4: Para los egresados cuyo padre alcanzó el nivel de educación secundaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.712 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P324=1: Para los egresados que pertenecieron al cuadro de méritos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.674 veces mayor que para los egresados que no pertenecieron al cuadro de méritos, dado que las otras variables se mantengan constantes.
- P330=1: Para los egresados que obtuvieron su título profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos

“Bajo, Medio y Alto” es 1.797 veces mayor que para los egresados que no obtuvieron su título profesional, dado que las otras variables se mantengan constantes.

- P456=1: Para los egresados cuyo primer empleo estuvo relacionado con su formación profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.310 veces mayor que para los egresados cuyo primer empleo no estuvo relacionado con su formación profesional, dado que las otras variables se mantengan constantes.
- P603_3=1: Para los egresados que calificaron como Excelente la preparación recibida en la universidad para “Coordinar actividades”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.678 veces mayor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Coordinar actividades”, dado que las otras variables se mantengan constantes.
- P603_3=2: Para los egresados que calificaron como Buena la preparación recibida en la universidad para “Coordinar actividades”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.405 veces mayor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Coordinar actividades”, dado que las otras variables se mantengan constantes.
- P603_11=1: Para los egresados que calificaron como Excelente la preparación recibida en la universidad en “Conocimientos básicos de otros campos o disciplinas”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.986 veces mayor que para los egresados que calificaron como Malo la preparación recibida en la universidad en “Conocimientos básicos de otros campos o disciplinas”, dado que las otras variables se mantengan constantes.
- P603_11=2: Para los egresados que calificaron como Bueno la preparación recibida en la universidad en “Conocimientos básicos de otros campos o disciplinas”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.624 veces mayor que para los egresados que calificaron como Malo la preparación recibida en la universidad en “Conocimientos básicos de otros campos o disciplinas”, dado que las otras variables se mantengan constantes.
- P603_11=3: Para los egresados que calificaron como Regular la preparación recibida en la universidad en “Conocimientos básicos de otros campos o disciplinas”, la probabilidad

de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.520 veces mayor que para los egresados que calificaron como Malo la preparación recibida en la universidad en “Conocimientos básicos de otros campos o disciplinas”, dado que las otras variables se mantengan constantes.

- P603_12=1: Para los egresados que calificaron como Excelente la preparación recibida en la universidad para “Tener pensamiento crítico”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.771 veces menor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Tener pensamiento crítico”, dado que las otras variables se mantengan constantes.
- P603_12=2: Para los egresados que calificaron como Bueno la preparación recibida en la universidad para “Tener pensamiento crítico”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.717 veces menor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Tener pensamiento crítico”, dado que las otras variables se mantengan constantes.

Para las universidades privadas:

Se emplea los Odds ratio (Razón de ventaja) para explicar los rangos de ingresos del primer empleo de los egresados de las universidades privadas:

- P11=1: Para los egresados hombres, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.796 veces mayor que para las egresadas mujeres, dado que las otras variables se mantengan constantes.
- P205_1=3: Para los egresados cuyo padre alcanzó el nivel de educación primaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.449 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P205_1=4: Para los egresados cuyo padre alcanzó el nivel de educación secundaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.516 veces menor que para aquellos

egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.

- P324=1: Para los egresados que pertenecieron al cuadro de méritos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.515 veces mayor que para los egresados que no pertenecieron al cuadro de méritos, dado que las otras variables se mantengan constantes.
- P330=1: Para los egresados que obtuvieron su título profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.795 veces mayor que para los egresados que no obtuvieron su título profesional, dado que las otras variables se mantengan constantes.
- P603_1=1: Para los egresados que calificaron como Excelente la preparación recibida en la universidad para “Dominar su área de disciplina”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.909 veces mayor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Dominar su área de disciplina”, dado que las otras variables se mantengan constantes.
- P603_1=2: Para los egresados que calificaron como Buena la preparación recibida en la universidad para “Dominar su área de disciplina”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.412 veces mayor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Dominar su área de disciplina”, dado que las otras variables se mantengan constantes.
- P603_4=1: Para los egresados que calificaron como Excelente la preparación recibida en la universidad para “Trabajar en equipo”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.832 veces menor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Trabajar en equipo”, dado que las otras variables se mantengan constantes.
- P603_4=2: Para los egresados que calificaron como Buena la preparación recibida en la universidad para “Trabajar en equipo”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.885 veces menor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Trabajar en equipo”, dado que las otras variables se mantengan constantes.

- P603_6=1: Para los egresados que calificaron como Excelente la preparación recibida en la universidad para “Utilizar herramientas informáticas básicas”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.046 veces mayor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Utilizar herramientas informáticas básicas”, dado que las otras variables se mantengan constantes.
- P603_6=2: Para los egresados que calificaron como Buena la preparación recibida en la universidad para “Utilizar herramientas informáticas básicas”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.001 veces mayor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Utilizar herramientas informáticas básicas”, dado que las otras variables se mantengan constantes.
- P603_12=1: Para los egresados que calificaron como Excelente la preparación recibida en la universidad para “Tener pensamiento crítico”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.863 veces menor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Tener pensamiento crítico”, dado que las otras variables se mantengan constantes.
- P603_12=2: Para los egresados que calificaron como Bueno la preparación recibida en la universidad para “Tener pensamiento crítico”, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.807 veces menor que para los egresados que calificaron como Regular la preparación recibida en la universidad para “Tener pensamiento crítico”, dado que las otras variables se mantengan constantes.

▪ **Prueba de hipótesis global para el Modelo 1**

Para probar la validez del Modelo1, se procede a formular las siguientes hipótesis respecto a los coeficientes de regresión:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{19} = 0$$

$$H_1: \text{Al menos un } \beta_i \neq 0 ; \forall i = 1, 2, \dots, 19$$

Cuadro 9. Pruebas de hipótesis global (Modelo 1)

Modelo	Universidades públicas			Universidades privadas		
	-2 log de la verosimilitud	Chi-cuadrado	Sig.	-2 log de la verosimilitud	Chi-cuadrado	Sig.

Sólo intersección	2702,072			3226,421		
Final	2480,553	221,519	,000	2903,939	322,483	,000

En el Cuadro 9 se muestra que el modelo global (Modelo 1) resulta significativo tanto para los datos de las universidades públicas como privadas. Con un nivel de significancia de 0.05, se rechaza la hipótesis nula, lo que indica que al menos una de las variables predictoras analizadas permite explicar la variable rangos de ingreso del primer empleo de los egresados de las universidades públicas, y de los egresados de las universidades privadas.

▪ **Prueba de bondad de ajuste de la Deviance y Pearson para el Modelos 1**

En forma similar para evaluar el ajuste del Modelo 1, se aplica la prueba de bondad de ajuste de la Deviance y Pearson que se basan en determinar sus respectivos residuales. Las hipótesis formuladas son las siguientes:

H_0 : El modelo 1 logístico ordinal se ajusta a los datos

H_1 : El modelo 1 logístico ordinal no se ajusta a los datos

Cuadro 10. Prueba de bondad de ajuste de la Deviance y Pearson (Modelo 1)

Prueba	Universidades públicas		Universidades privadas	
	Chi-cuadrado	Sig.	Chi-cuadrado	Sig.
Pearson	1710,259	,719	1907,784	,529
Desviación	1603,973	,993	1765,430	,993

En el Cuadro 10, se muestra los resultados de las pruebas de bondad de ajuste de la Deviance y Pearson para las universidades públicas y privadas. Con niveles de significancia de 0.719 (Pearson) y 0.993 (Deviance), no existe suficiente evidencia estadística para rechazar la hipótesis nula, es decir, se puede afirmar que el modelo logístico ordinal se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades públicas; de igual forma, con niveles de significancia de 0.529 (Pearson) y 0.993 (Deviance), no existe suficiente evidencia estadística para rechazar la hipótesis nula, es decir, se puede afirmar que el modelo logístico ordinal se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades privadas. Se concluye que el modelo logístico ordinal propuesto (Modelo1) se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades públicas, y de los egresados de las universidades privadas.

- **Medidas de bondad de ajuste Pseudo-R² para el Modelo 1**

Se calculan los Pseudo-R² con la finalidad de evaluar la bondad de ajuste.

Cuadro 11. Cálculo del Pseudo-R² (Modelo 1)

Medidas	Universidad pública	Universidad privada
	Valor	Valor
Cox y Snell	,096	,094
Nagelkerke	,106	,103
McFadden	,043	,040

En el Cuadro 11, se presenta los valores de los Pseudos-R² de Cox-Snel, Nagelkerke y McFadden. Se puede indicar que el modelo logístico ordinal (Modelo1) explica según la medida Cox-Snel el 9.6%, con Nagelkerke el 10.6% y con McFadden el 4.3%, de la variabilidad los rangos de ingresos del primer empleo de los egresados de universidades públicas; de igual manera las medidas de bondad de ajuste que indican el porcentaje de la variabilidad, de los rangos de ingresos del primer empleo de los egresados de universidades privadas, explicada por el modelo logístico ordinal (Modelo 1) son 9.4%, 10.3% y 4.0%, respectivamente.

- **Prueba de hipótesis de líneas paralelas para el Modelo 1**

Con esta prueba de la hipótesis nula establece que los parámetros de ubicación (los coeficientes para las pendientes) son los mismos para todas las categorías de respuesta.

Ho: Los coeficientes para las pendientes son los mismos para todas las categorías de Y.

H1: Los coeficientes para las pendientes no son los mismos para todas las categorías de Y.

Cuadro 12. Prueba de hipótesis de líneas paralelas (Modelo 1)

Modelo	Universidades públicas			Universidades privadas		
	-2 log de la verosimilitud	Chi-cuadrado	Sig.	-2 log de la verosimilitud	Chi-cuadrado	Sig.
Hipótesis nula	2480,553			2903,939		
General	2442,424	38,129	,059	2879,798	24,141	,568

En el Cuadro 12, se presenta la prueba de líneas paralelas usando las estadísticas de Log-verosimilitud y Chi-cuadrado:

- Con un nivel de significancia de 0.05, no existe suficiente evidencia estadística para rechazar que los coeficientes de las variables independientes del modelo logístico ordinal para explicar los rangos de ingresos de los egresados de universidades públicas son los mismos para todas las categorías.
- Además, con un nivel de significancia de 0.05, no existe suficiente evidencia estadística para rechazar que los coeficientes de las variables independientes del modelo logístico ordinal para explicar los rangos de ingresos de los egresados de universidades privadas son los mismos para todas las categorías.

4.3.2 Análisis del Modelo 2

El respectivo modelo se expresa en (3.5). Se ajusta los datos del rango de ingreso de los egresados universitarios a una regresión logística ordinal con función de enlace la logit. La variable dependiente Y, es el rango de los ingresos y las variables independientes son las variables socio-académicas (Grupo A) y la importancia de las competencias adquiridas para su experiencia laboral (Grupo C).

$$\log it \mathbf{P}(Y \leq j / x) = \alpha_j + \sum_{i=1}^7 \beta_i X_i + \sum_{i=1}^{12} \beta_i P604_i, \quad \text{para } j = 1, 2, 3$$

Con la finalidad de evaluar y validar los modelos logísticos ordinales, se aplicarán medidas y pruebas de hipótesis. En primer lugar se ajusta los datos a la regresión logística para obtener los coeficientes de regresión significativos aplicando la prueba de hipótesis de Wald.

Prueba de Wald para el Modelo 2

Con la finalidad de probar la significación de cada coeficiente de regresión, se aplica la prueba estadística de Wald. La hipótesis que se formula es la siguiente:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

En el Cuadro 13 se presenta los coeficientes de regresión estimados y su significación para los datos de las universidades públicas y privadas.

Cuadro 13. Coeficientes estimados y prueba de Wald (Modelo 2)

Variables	Universidades públicas		Universidades privadas	
	Estimación	Sig.	Estimación	Sig.
[R_Ingreso = 1]	-,420	,407	-,129	,753
[R_Ingreso = 2]	1,669	,001	1,817	,000
[R_Ingreso = 3]	3,414	,000	3,648	,000
[P11=1]	,740	,000	,611	,000
[P205_1=3]	-,427	,000	-,793	,000
[P205_1=4]	-,338	,000	-,647	,000
[P324=1]	,483	,000	,382	,000
[P330=1]	,567	,000	,585	,000
[P349=1]	-,274	,124	,037	,813
[P456=1]	,310	,028	,167	,165
[P471=1]	-,060	,669	-,091	,452
[P604_1=1]	,167	,107	,013	,882
[P604_2=1]	,147	,149	,353	,000
[P604_3=1]	,194	,064	,261	,003
[P604_4=1]	-,156	,202	-,077	,448
[P604_5=1]	,072	,534	-,021	,819
[P604_6=1]	-,041	,691	-,048	,589
[P604_7=1]	,017	,926	-,111	,449
[P604_7=2]	,245	,158	-,097	,494
[P604_8=1]	-,064	,809	-,122	,563
[P604_8=2]	-,121	,637	-,156	,445
[P604_9=1]	,324	,002	,060	,497
[P604_10=1]	-,058	,672	,344	,003
[P604_10=2]	,029	,807	,128	,235
[P604_11=1]	,052	,784	-,055	,718
[P604_11=2]	-,027	,874	,057	,688
[P604_12=1]	-,646	,114	-,187	,572
[P604_12=2]	-,501	,220	-,088	,790

Para las universidades públicas:

Con un nivel de significación de 0.05, existe fuerte evidencia estadística para afirmar que las variables independientes “género” (P11), “nivel de estudios del padre” (P205_1), “perteneció al cuadro de méritos” (P324), “obtuvo el título profesional” (P330), “primer empleo relacionado con la formación profesional” (P456) e “importancia de redactar informes o documentos” (P604_9), explican la variable dependiente: rangos de ingresos del primer empleo de los egresados de las universidades públicas.

No existe suficiente evidencia estadística para afirmar que las variables independientes, “realizó prácticas pre profesionales” (P349), “la formación universitaria lo ayudó a conseguir su primer empleo” (P471), “importancia del dominio de su área de disciplina” (P604_1), “importancia de rendir bajo presión y cumplir ciertos objetivos” (P604_2), “importancia de coordinar actividades” (P604_3), “importancia de trabajar en equipo” (P604_4), “importancia de tener liderazgo” (P604_5), “importancia de utilizar herramientas informáticas básicas” (P604_6), “importancia de utilizar software específico de la carrera” (P604_7), “importancia de presentar en público productos, ideas o informes” (P604_8), “importancia de leer, escribir y hablar en idioma(s) extranjero(s)” (P604_10), “importancia de conocimientos básicos de otros campos o disciplinas” (P604_11) e “importancia de tener pensamiento crítico” (P604_12), explican la variable dependiente.

Para las universidades privadas:

Con un nivel de significación de 0.05, existe fuerte evidencia estadística para afirmar que las variables independientes “género” (P11), “nivel de estudios del padre” (P205_1), “perteneció al cuadro de méritos” (P324), “obtuvo el título profesional” (P330), “importancia de rendir bajo presión y cumplir con los objetivos” (P604_2) y “importancia de coordinar actividades” (P604_3) e “importancia de leer, escribir y hablar en idioma(s) extranjero(s)” (P604_10), explican la variable dependiente: rangos de ingresos del primer empleo de los egresados de las universidades privadas.

No existe suficiente evidencia estadística para afirmar que las variables independientes, “realizó prácticas pre profesionales” (P349), “primer empleo relacionado con la formación profesional” (P456), “la formación universitaria lo ayudó a conseguir su primer empleo” (P471), “importancia del dominio de su área de disciplina” (P604_1), “importancia de trabajar en equipo” (P604_4), “importancia de tener liderazgo” (P604_5), “importancia de utilizar herramientas informáticas básicas” (P604_6), “importancia de utilizar software específico de la carrera” (P604_7), “importancia de presentar en público productos, ideas o

informes” (P604_8), “importancia de redactar informes o documentos” (P604_9), “importancia de conocimientos básicos de otros campos o disciplinas” (P604_11) e “importancia de tener pensamiento crítico” (P604_12), explican la variable dependiente.

Análisis de regresión logística con variables significativas para el Modelo 2

Se ajusta los datos con las variables que resultaron significativas en el Cuadro 13. Se realizará pruebas de bondad de ajuste y la prueba de líneas paralelas.

▪ **Estimación de los coeficientes de regresión para el Modelo 2**

En el Cuadro 14, se presenta la respectiva salida de los coeficientes de la regresión logística ordinal estimados, la significación con la estadística Wald y los respectivos Odds ratios para el Modelo 2 y para las universidades públicas y privadas.

Cuadro 14. Estimación de los coeficientes de regresión (Modelo 2)

Variables	Universidades públicas			Universidades privadas		
	Coefficientes	Sig.	exp(b)	Coefficientes	Sig.	exp(b)
[R_Ingreso = 1]	,194	,062		,113	,343	
[R_Ingreso = 2]	2,262	,000		2,054	,000	
[R_Ingreso = 3]	3,999	,000		3,882	,000	
[P11=1]	,732	,000	2,079	,611	,000	1,841
[P205_1=3]	-,398	,001	,672	-,789	,000	,454
[P205_1=4]	-,335	,000	,715	-,648	,000	,523
[P324=1]	,497	,000	1,645	,385	,000	1,469
[P330=1]	,571	,000	1,770	,592	,000	1,807
[P456=1]	,265	,001	1,304			
[P604_2=1]				,304	,000	1,355
[P604_3=1]				,192	,013	1,212
[P604_9=1]	,331	,000	1,392			
[P604_10=1]				,248	,016	1,281
[P604_10=2]				,098	,328	1,103

Para las universidades públicas:

Se emplea los Odds ratio (Razón de ventaja) para explicar los rangos de ingresos del primer empleo de los egresados de las universidades públicas:

- P11=1: Para los egresados hombres, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 2.079 veces mayor que para las egresadas mujeres, dado que las otras variables se mantengan constantes.
- P205_1=3: Para los egresados cuyo padre alcanzó el nivel de educación primaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.672 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P205_1=4: Para los egresados cuyo padre alcanzó el nivel de educación secundaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.715 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P324=1: Para los egresados que pertenecieron al cuadro de méritos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.645 veces mayor que para los egresados que no pertenecieron al cuadro de méritos, dado que las otras variables se mantengan constantes.
- P330=1: Para los egresados que obtuvieron su título profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.770 veces mayor que para los egresados que no obtuvieron su título profesional, dado que las otras variables se mantengan constantes.
- P456=1: Para los egresados cuyo primer empleo estuvo relacionado con su formación profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.304 veces mayor que para los egresados cuyo primer empleo no estuvo relacionado con su formación profesional, dado que las otras variables se mantengan constantes.
- P604_9=1: Para los egresados que consideran Muy importante redactar informes o documentos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.392 veces mayor que para

los egresados que consideran Importante redactar informes o documentos, dado que las otras variables se mantengan constantes.

Para las universidades privadas:

Se emplea los Odds ratio (Razón de ventaja) para explicar los rangos de ingresos del primer empleo de los egresados de las universidades privadas:

- P11=1: Para los egresados hombres, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.841 veces mayor que para las egresadas mujeres, dado que las otras variables se mantengan constantes.
- P205_1=3: Para los egresados cuyo padre alcanzó el nivel de educación primaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.454 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P205_1=4: Para los egresados cuyo padre alcanzó el nivel de educación secundaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.523 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P324=1: Para los egresados que pertenecieron al cuadro de méritos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.469 veces mayor que para los egresados que no pertenecieron al cuadro de méritos, dado que las otras variables se mantengan constantes.
- P330=1: Para los egresados que obtuvieron su título profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.807 veces mayor que para los egresados que no obtuvieron su título profesional, dado que las otras variables se mantengan constantes.
- P604_2=1: Para los egresados que consideran Muy importante rendir bajo presión y cumplir ciertos objetivos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.355 veces mayor

que para los egresados que consideran Importante rendir bajo presión y cumplir ciertos objetivos, dado que las otras variables se mantengan constantes.

- P604_3=1: Para los egresados que consideran Muy importante coordinar actividades, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.212 veces mayor que para los egresados que consideran Importante coordinar actividades, dado que las otras variables se mantengan constantes.
- P604_10=1: Para los egresados que consideran Muy importante leer, escribir y hablar en idioma(s) extranjero(s), la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.281 veces mayor que para los egresados que consideran Poco importante leer, escribir y hablar en idioma(s) extranjero(s), dado que las otras variables se mantengan constantes.
- P604_10=2: Para los egresados que consideran Importante leer, escribir y hablar en idioma(s) extranjero(s), la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.103 veces mayor que para los egresados que consideran Poco importante leer, escribir y hablar en idioma(s) extranjero(s), dado que las otras variables se mantengan constantes.

▪ **Prueba de hipótesis global para el Modelo 2**

Para probar la validez del Modelo 2, se procede a formular las siguientes hipótesis respecto a los coeficientes de regresión:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{19} = 0$$

$$H_1: \text{Al menos un } \beta_i \neq 0 ; \forall i = 1, 2, \dots, 19$$

Cuadro 15. Prueba de hipótesis global (Modelo 2)

Modelo	Universidades públicas			Universidades privadas		
	-2 log de la verosimilitud	Chi-cuadrado	Sig.	-2 log de la verosimilitud	Chi-cuadrado	Sig.
Sólo intersección	1165,215			2143,922		
Final	952,524	212,691	,000	1799,029		,000

En el Cuadro 15 se muestra que el modelo global (Modelo 2) resulta significativo tanto para los datos de las universidades públicas como privadas. Con un nivel de significancia de 0.05,

se rechaza la hipótesis nula, lo que indica que al menos una de las variables predictoras analizadas permite explicar la variable rangos de ingreso del primer empleo de los egresados de las universidades públicas, y de los egresados de las universidades privadas.

▪ **Prueba de bondad de ajuste de la Deviance y Pearson para el Modelo 2**

Similarmente con la finalidad de evaluar el ajuste del Modelo 2, se aplica la prueba de bondad de ajuste de la Deviance y Pearson que se basan en determinar sus respectivos residuales. Las respectivas hipótesis que se formulan son:

H_0 : El modelo logístico ordinal se ajusta a los datos

H_1 : El modelo logístico ordinal no se ajusta a los datos

Cuadro 16. Prueba de bondad de ajuste de la Deviance y Pearson (Modelo 2)

Prueba	Universidades públicas		Universidades privadas	
	Chi-cuadrado	Sig.	Chi-cuadrado	Sig.
Pearson	282,954	,406	881,826	,001
Desviación	308,234	,103	787,853	,205

En el Cuadro 16, se muestra los resultados de las pruebas de bondad de ajuste de la Deviance y Pearson para las universidades públicas y privadas. Con niveles de significancia de 0.406 (Pearson) y 0.103 (Deviance), no existe suficiente evidencia estadística para rechazar la hipótesis nula, es decir, se puede afirmar que el modelo logístico ordinal se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades públicas. Con niveles de significancia de 0.001 (Pearson), existe suficiente evidencia estadística para rechazar la hipótesis nula y 0.205 (Deviance), no presenta suficiente evidencia estadística para rechazar la hipótesis nula. Se concluye que el modelo logístico ordinal propuesto (Modelo2) se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades públicas, en el caso del modelo que busca explicar los rangos de ingresos del primer empleo de los egresados de las universidades privadas, los resultados no son concluyentes, puesto que uno de los estadísticos de prueba ofrece evidencia estadística para rechazar H_0 , y el otro estadístico de prueba no rechaza H_0 .

▪ **Medidas de bondad de ajuste de Pseudo- R^2 para el Modelo 2**

Se calculan los Pseudos- R^2 con la finalidad de evaluar la bondad de ajuste.

Cuadro 17. Cálculo del Pseudo- R^2 (Modelo 2)

Medidas	Universidad pública	Universidad privada
---------	---------------------	---------------------

	Valor	Valor
Cox y Snell	,093	,100
Nagelkerke	,102	,109
McFadden	,041	,043

En el Cuadro 17, se presenta los valores de los Pseudos- R^2 de Cox-Snel, Nagelkerke y McFadden. Se puede indicar que el modelo logístico ordinal (Modelo2) explica según la medida Cox-Snel el 9.3%, con Nagelkerke el 10.2% y con McFadden el 4.1%, de la variabilidad los rangos de ingresos del primer empleo de los egresados de universidades públicas; de igual manera las medidas de bondad de ajuste que indican el porcentaje de la variabilidad, de los rangos de ingresos del primer empleo de los egresados de universidades privadas, explicada por el modelo logístico ordinal (Modelo 2) son 10.0%, 10.9% y 4.3%, respectivamente.

▪ **Prueba de hipótesis de líneas paralelas para el Modelo 2**

Con esta prueba de la hipótesis nula establece que los parámetros de ubicación (los coeficientes para las pendientes) son los mismos para todas las categorías de respuesta.

Ho: Los coeficientes para las pendientes son los mismos para todas las categorías de Y.

H1: Los coeficientes para las pendientes no son los mismos para todas las categorías de Y.

Cuadro 18. Prueba de hipótesis de líneas paralelas (Modelo 2)

Modelo	Universidades públicas			Universidades privadas		
	-2 log de la verosimilitud	Chi-cuadrado	Sig.	-2 log de la verosimilitud	Chi-cuadrado	Sig.
Hipótesis nula	952,524			1799,029		
General	927,134	25,390	,031	1777,118	21,910	,236

En el Cuadro 18, se presenta la prueba de líneas paralelas usando las estadísticas Log-verosimilitud y Chi-Cuadrado:

- Con un nivel de significancia de 0.05, existe suficiente evidencia estadística para rechazar que los coeficientes de las variables independientes del modelo logístico ordinal para explicar los rangos de ingresos de los egresados de universidades públicas son los mismos para todas las categorías.

- Además, con un nivel de significancia de 0.05, no existe suficiente evidencia estadística para rechazar que los coeficientes de las variables independientes del modelo logístico ordinal para explicar los rangos de ingresos de los egresados de universidades privadas son los mismos para todas las categorías.

4.3.3 Análisis del Modelo 3

El respectivo modelo se expresa en (3.6). Se ajusta los datos del rango de ingreso de los egresados universitarios a una regresión logística ordinal con función de enlace la logit. La variable dependiente Y, es el rango de los ingresos y las variables independientes son las variables socio-académicas (Grupo A) y la calificación de aspectos de la universidad sobre los profesores de la carrera (Grupo D).

$$\log it \mathbf{P}(Y \leq j / x) = \alpha_j + \sum_{i=1}^7 \beta_i X_i + \sum_{i=1}^5 \beta_i P605_i, \quad para : j = 1,2,3$$

Con la finalidad de evaluar y validar los modelos logísticos ordinales, se aplicarán medidas y pruebas de hipótesis. En primer lugar se ajusta los datos a la regresión logística para obtener los coeficientes de regresión significativos aplicando la prueba de hipótesis de Wald.

Prueba de Wald para el Modelo 3

Con la finalidad de probar la significación de cada coeficiente de regresión, se aplica la prueba estadística de Wald. La hipótesis que se formula es la siguiente:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

En el Cuadro 19 se presenta los coeficientes de regresión estimados y su significación para los datos de las universidades públicas y privadas.

Cuadro 19. Estimación de los coeficientes de regresión (Modelo 3)

Variables	Universidades públicas		Universidades privadas	
	Estimación	Sig.	Estimación	Sig.
[R_Ingreso = 1]	-,103	,620	-,100	,614
[R_Ingreso = 2]	1,966	,000	1,820	,000

Variables	Universidades públicas		Universidades privadas	
	Estimación	Sig.	Estimación	Sig.
[R_Ingreso = 3]	3,700	,000	3,636	,000
[P11=1]	,710	,000	,598	,000
[P205_1=3]	-,408	,001	-,829	,000
[205_1=4]	-,350	,000	-,689	,000
[P324=1]	,517	,000	,405	,000
[P330=1]	,570	,000	,569	,000
[P349=1]	-,214	,228	,061	,698
[P456=1]	,323	,022	,148	,216
[P471=1]	-,052	,711	-,064	,597
[P605_1=1]	,476	,012	,048	,782
[P605_1=2]	,095	,418	,154	,286
[P605_2=1]	-,155	,382	,099	,494
[P605_2=2]	,136	,190	,091	,438
[P605_3=1]	-,284	,158	-,051	,748
[P605_3=2]	-,021	,853	-,150	,236
[P605_4=1]	-,081	,713	,088	,572
[P605_4=2]	-,095	,386	,092	,425

Para las universidades públicas:

Con un nivel de significación de 0.05, existe fuerte evidencia estadística para afirmar que las variables independientes “género” (P11), “nivel de estudios del padre” (P205_1), “perteneció al cuadro de méritos” (P324), “obtuvo el título profesional” (P330), “primer empleo relacionado con la formación profesional” (P456) y “calificación de la calidad de los profesores de su carrera” (P605_1), explican la variable dependiente: rangos de ingresos del primer empleo de los egresados de las universidades públicas.

No existe suficiente evidencia estadística para afirmar que las variables independientes, “realizó prácticas pre profesionales” (P349), “la formación universitaria lo ayudó a conseguir su primer empleo” (P471), “calificación del grado de acceso a los profesores de su carrera” (P605_2), “calificación del grado de actualización de conocimientos de los profesores de su carrera” (P605_3) y “calificación de la preparación de clases de los profesores de su carrera” (P605_4), explican la variable dependiente.

Para las universidades privadas:

Con un nivel de significación de 0.05, existe fuerte evidencia estadística para afirmar que las variables independientes “género” (P11), “nivel de estudios del padre” (P205_1), “perteneció al cuadro de méritos” (P324) y “obtuvo el título profesional” (P330), explican la variable dependiente: rangos de ingresos del primer empleo de los egresados de las universidades privadas.

No existe suficiente evidencia estadística para afirmar que las variables independientes, “realizó prácticas pre profesionales” (P349), “primer empleo relacionado con la formación profesional” (P456), “la formación universitaria lo ayudó a conseguir su primer empleo” (P471), “calificación de la calidad de los profesores de su carrera” (P605_1), “calificación del grado de acceso a los profesores de su carrera” (P605_2), “calificación del grado de actualización de conocimientos de los profesores de su carrera” (P605_3) y “calificación de la preparación de clases de los profesores de su carrera” (P605_4), explican la variable dependiente.

Análisis de regresión logística con variables significativas para el Modelo 3

Se ajusta los datos con las variables que resultaron significativas en el Modelo 1. Se realizará pruebas de bondad de ajuste y la prueba de líneas paralelas.

▪ **Estimación de los coeficientes de regresión para el Modelo 3**

En el Cuadro 20, se presenta la respectiva salida de los coeficientes estimados de la regresión logística ordinal, la significación con la estadística Wald y los respectivos Odds ratio para el Modelo 3, para las universidades públicas y privadas.

Cuadro 20. Estimación de los coeficientes de regresión (Modelo 3)

Variables	Universidades públicas			Universidades privadas		
	Coefficientes	Sig.	exp(b)	Coefficientes	Sig.	exp(b)
[R_Ingreso = 1]	,080	,501		-,350	,000	
[R_Ingreso = 2]	2,139	,000		1,567	,000	
[R_Ingreso = 3]	3,870	,000		3,382	,000	
[P11=1]	,711	,000	2,036	,595	,000	1,813
[P205_1=3]	-,392	,001	,676	-,844	,000	,430
[P205_1=4]	-,337	,000	,714	-,686	,000	,504
[P324=1]	,512	,000	1,669	,421	,000	1,523

[P330=1]	,573	,000	1,774	,582	,000	1,790
[P456=1]	,272	,001	1,313			
[P605_1=1]	,197	,164	1,218			
[P605_1=2]	,046	,631	1,047			

Para las universidades públicas:

Se emplea los Odds ratio (Razón de ventaja) para explicar los rangos de ingresos del primer empleo de los egresados de las universidades públicas:

- P11=1: Para los egresados hombres, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 2.036 veces mayor que para las egresadas mujeres, dado que las otras variables se mantengan constantes.
- P205_1=3: Para los egresados cuyo padre alcanzó el nivel de educación primaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.676 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P205_1=4: Para los egresados cuyo padre alcanzó el nivel de educación secundaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.714 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P324=1: Para los egresados que pertenecieron al cuadro de méritos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.669 veces mayor que para los egresados que no pertenecieron al cuadro de méritos, dado que las otras variables se mantengan constantes.
- P330=1: Para los egresados que obtuvieron su título profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.774 veces mayor que para los egresados que no obtuvieron su título profesional, dado que las otras variables se mantengan constantes.
- P456=1: Para los egresados cuyo primer empleo estuvo relacionado con su formación profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.313 veces mayor que para los

egresados cuyo primer empleo no estuvo relacionado con su formación profesional, dado que las otras variables se mantengan constantes.

- P605_1=1: Para los egresados que calificaron como Excelente la calidad de los profesores de su carrera, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.218 veces mayor que para los egresados que calificaron como Regular la calidad de los profesores de su carrera, dado que las otras variables se mantengan constantes.
- P605_1=2: Para los egresados que calificaron como Buena la calidad de los profesores de su carrera, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.047 veces mayor que para los egresados que calificaron como Regular la calidad de los profesores de su carrera, dado que las otras variables se mantengan constantes.

Para las universidades privadas:

Se emplea los Odds ratio (Razón de ventaja) para explicar los rangos de ingresos del primer empleo de los egresados de las universidades privadas:

- P11=1: Para los egresados hombres, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.813 veces mayor que para las egresadas mujeres, dado que las otras variables se mantengan constantes.
- P205_1=3: Para los egresados cuyo padre alcanzó el nivel de educación primaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.430 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P205_1=4: Para los egresados cuyo padre alcanzó el nivel de educación secundaria, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 0.504 veces menor que para aquellos egresados cuyo padre alcanzó el nivel superior, dado que las otras variables se mantengan constantes.
- P324=1: Para los egresados que pertenecieron al cuadro de méritos, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.523 veces mayor que para los egresados que no pertenecieron al cuadro de méritos, dado que las otras variables se mantengan constantes.

- P330=1: Para los egresados que obtuvieron su título profesional, la probabilidad de obtener ingresos “Muy Altos”, en el primer empleo, frente a la combinación de ingresos “Bajo, Medio y Alto” es 1.790 veces mayor que para los egresados que no obtuvieron su título profesional, dado que las otras variables se mantengan constantes.

▪ **Prueba de hipótesis global para el Modelo 3**

Para probar la validez del Modelo 3, se procede a formular las siguientes hipótesis respecto a los coeficientes de regresión:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{12} = 0$$

$$H_1: \text{Al menos un } \beta_i \neq 0 ; \forall i = 1, 2, \dots, 8$$

Cuadro 21. Prueba de hipótesis global (Modelo 3)

Modelo	Universidades públicas			Universidades privadas		
	-2 log de la verosimilitud	Chi-cuadrado	Sig.	-2 log de la verosimilitud	Chi-cuadrado	Sig.
Sólo intersección	1368,488			628,153		
Final	1170,556	197,931	,000	337,373	290,780	,000

En el Cuadro 21 se muestra que el modelo global (Modelo 3) resulta significativo tanto para los datos de las universidades públicas como privadas. Con un nivel de significancia de 0.05, se rechaza la hipótesis nula, lo que indica que al menos una de las variables predictoras analizadas permite explicar la variable rangos de ingreso del primer empleo de los egresados de las universidades públicas, y de los egresados de las universidades privadas

▪ **Prueba de bondad de ajuste de la Deviance y Pearson para el Modelo 3**

En forma similar para evaluar el ajuste del Modelo 3, se aplica la prueba de bondad de ajuste de la Deviance y Pearson que se basan en determinar sus respectivos residuales. Las respectivas hipótesis formuladas son las siguientes:

$$H_0: \text{El modelo logístico ordinal se ajusta a los datos}$$

$$H_1: \text{El modelo logístico ordinal no se ajusta a los datos}$$

Cuadro 22. Prueba de bondad de ajuste de la Deviance y Pearson (Modelo 3)

Prueba	Universidades públicas		Universidades privadas	
	Chi-cuadrado	Sig.	Chi-cuadrado	Sig.

Pearson	440,794	,158	57,611	,700
Desviación	456,522	,064	61,887	,552

En el Cuadro 22, se muestra los resultados de las pruebas de bondad de ajuste de la Deviance y Pearson para las universidades públicas y privadas. Con niveles de significancia de 0.158 (Pearson) y 0.064 (Deviance), no existe suficiente evidencia estadística para rechazar la hipótesis nula, es decir, se puede afirmar que el modelo logístico ordinal se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades públicas; de igual forma, con niveles de significancia de 0.700 (Pearson) y 0.552 (Deviance), no existe suficiente evidencia estadística para rechazar la hipótesis nula, es decir, se puede afirmar que el modelo logístico ordinal se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades privadas. Se concluye que el modelo logístico ordinal propuesto (Modelo3) se ajusta a los datos de los rangos de ingresos del primer empleo de los egresados de las universidades públicas, y de los egresados de las universidades privadas.

▪ **Medidas de bondad de ajuste Pseudo-R² del Modelo 3**

Se calculan los Pseudo-R² con la finalidad de evaluar la bondad de ajuste.

Cuadro 23. Cálculo del Pseudo-R² (Modelo 3)

Medidas	Universidad pública	Universidad privada
	Valor	Valor
Cox y Snell	,086	,085
Nagelkerke	,095	,093
McFadden	,038	,036

En el Cuadro 23, se presenta los valores de los Pseudos-R² de Cox-Snel, Nagelkerke y McFadden. Se puede indicar que el modelo logístico ordinal (Modelo 3) explica según la medida Cox-Snel el 8.6%, con Nagelkerke el 9.5% y con McFadden el 3.8%, de la variabilidad los rangos de ingresos del primer empleo de los egresados de universidades públicas; de igual manera las medidas de bondad de ajuste que indican el porcentaje de la variabilidad, de los rangos de ingresos del primer empleo de los egresados de universidades privadas, explicada por el modelo logístico ordinal (Modelo 2) son 8.5%, 9.3% y 3.6%, respectivamente.

▪ **Prueba de hipótesis de líneas paralelas para el Modelo 3**

Con esta prueba de la hipótesis nula establece que los parámetros de ubicación (los coeficientes para las pendientes) son los mismos para todas las categorías de respuesta.

Ho: Los coeficientes para las pendientes son los mismos para todas las categorías de Y.

H1: Los coeficientes para las pendientes no son los mismos para todas las categorías de Y.

Cuadro 24. Prueba de hipótesis de líneas paralelas (Modelo 3)

Modelo	Universidades públicas			Universidades privadas		
	-2 log de la verosimilitud	Chi-cuadrado	Sig.	-2 log de la verosimilitud	Chi-cuadrado	Sig.
Hipótesis nula	1170,556			337,373		
General	1145,878	24,678	,076	321,600	15,773	,106

En el Cuadro 24, se presenta la prueba de líneas paralelas usando las estadísticas Log-verosimilitud y Chi-Cuadrado:

- Con un nivel de significancia de 0.05, no existe suficiente evidencia estadística para rechazar que los coeficientes de las variables independientes del modelo logístico ordinal para explicar los rangos de ingresos de los egresados de universidades públicas son los mismos para todas las categorías.
- Además, con un nivel de significancia de 0.05, no existe suficiente evidencia estadística para rechazar que los coeficientes de las variables independientes del modelo logístico ordinal para explicar los rangos de ingresos de los egresados de universidades privadas son los mismos para todas las categorías.

4.4 Comparación de los ingresos entre universidad pública y privada

Con la finalidad de realizar un estudio comparativo de los ingresos percibidos por los egresados entre las universidades públicas y privadas, se elaboran cuadros de las variables socio-académicas (Grupo A) y las variables respecto a la calificación de la preparación recibida en la universidad para el desarrollo de la competencia (Grupo B), la importancia que considera Ud. las siguientes competencias en su experiencia laboral (Grupo C) y la calificación de los siguientes aspectos de la universidad de la cuál egresó (Grupo D).

En el Cuadro 25, se presenta los promedios del ingreso de los egresados de las universidades públicas y privadas según las variables socio-académicas. Se observa que los ingresos promedios de los egresados de las universidades privadas son mayores que las públicas para todas las variables en sus diferentes categorías: sexo, nivel de estudios más alto que aprobó el padre, si pertenece algún cuadro de mérito, si ha obtenido título profesional, si realizó prácticas pre profesionales, si su primer empleo estuvo relacionado con su formación profesional y si su formación de la universidad ayudó a obtener primer empleo.

Cuadro 25. Ingreso promedio de los egresados universidades públicas y privadas

Variable	Categoría	Pública	Privada	Sig.
P11. Sexo	Hombre	1838,68	1950,90	0,010
	Mujer	1494,18	1684,79	0,000
P205 Nivel de estudios más alto que aprobó el padre	Primaria	1535,84	1428,53	0,209
	Secundaria	1561,15	1517,10	0,346
	Superior	1770,96	1903,43	0,000
P324 ¿Pertenece a algún cuadro de méritos?	Sí	1812,86	1916,29	0,008
	No	1493,03	1640,58	0,000
P330 ¿Ha obtenido el título profesional?	Sí	1835,41	1960,10	0,004
	No	1540,59	1625,53	0,016
P349 ¿Realizó practicas pre profesionales?	Sí	1659,06	1794,23	0,000
	No	1710,67	1661,45	0,686
P456 ¿Su primer empleo estuvo relacionado con su formación profesional	Sí	1767,69	1838,58	0,086
	No	1571,29	1747,84	0,000
P471 Formación de la universidad ayudó a obtener primer empleo	Sí	1723,29	1818,50	0,018
	No	1603,82	1764,63	0,000

En el Cuadro 26, se presenta los promedios del ingreso de los egresados de las universidades públicas y privadas según las variables respecto a la calificación de la preparación recibida

en la universidad para el desarrollo de sus competencias. Se observa que los ingresos promedios de los egresados de las universidades privadas son mayores que las públicas para todas las variables (competencias recibidas) y en sus diferentes categorías. También se observa un crecimiento en el promedio del ingreso cuando la calificación va de regular a excelente.

Cuadro 26. Calificación de la preparación recibida en la universidad para el desarrollo de las siguientes competencias

Variables	Categorías	Pública	Privada	Sig.
P603_1: Dominar tu área de disciplina	Excelente	1731,44	1953,51	0,001
	Bueno	1684,31	1756,04	0,032
	Regular	1503,73	1531,16	0,704
P603_3: Coordinar actividades	Excelente	1833,96	1951,87	0,102
	Bueno	1664,81	1753,57	0,008
	Regular	1524,09	1617,76	0,187
P603_6: Utilizar herramientas informáticas básicas	Excelente	1774,90	1884,55	0,110
	Bueno	1654,77	1759,56	0,003
	Regular	1623,02	1708,29	0,176
P603_7: Utilizar software específico de la carrera	Excelente	1796,15	1953,82	0,034
	Bueno	1697,04	1789,46	0,022
	Regular	1602,44	1634,35	0,525
	Malo	1543,57	1599,65	0,642
P603_11: Conocimientos básicos de otros campos o disciplinas	Excelente	1774,40	1976,47	0,049
	Bueno	1675,88	1803,49	0,001
	Regular	1643,79	1646,21	0,960
	Malo	1443,90	1803,70	0,042
P603_12: Tener pensamiento crítico	Excelente	1800,45	1903,11	0,126
	Bueno	1627,77	1744,93	0,000
	Regular	1680,53	1746,10	0,399

En el Cuadro 27, se presenta los promedios del ingreso de los egresados de las universidades públicas y privadas según las variables respecto a la importancia de las competencias recibidas para su experiencia profesional. Se observa que los ingresos promedios de los egresados de las universidades privadas son mayores que las públicas para todas las variables (competencias recibidas) y en sus diferentes categorías. También se observa un crecimiento en el promedio del ingreso cuando la calificación va de regular a excelente.

Cuadro 27. Importancia de las competencias recibidas para su experiencia profesional

Variables	Categorías	Pública	Privada	Sig.
P604_2: Rendir bajo presión y cumplir ciertos objetivos	Muy importante	1726,88	1877,22	0,000
	Importante	1560,31	1610,62	0,251
P604_3: Coordinar actividades	Muy importante	1720,93	1868,60	0,000
	Importante	1579,37	1653,53	0,081
P604_9: Redactar informes o documentos	Muy importante	1744,40	1854,99	0,004
	Importante	1570,08	1689,90	0,003
P604_11: Conocimientos básicos de otros campos o disciplinas	Muy importante	1714,07	1831,10	0,011
	Importante	1632,52	1758,22	0,001
	Poco importante	1670,60	1748,28	0,468

V. CONCLUSIONES

Las conclusiones de la presente investigación son:

1. Las variables más importantes, para explicar los rangos de ingreso del primer empleo de los egresados de universidades públicas, según el Modelo 1, evaluando los valores Odds-ratio para las variables socio-académicas son: el sexo (2.058), si su padre alcanzó el nivel de educación primaria (0.686), si su padre alcanzó el nivel de educación secundaria (0.712), si perteneció al cuadro de méritos (1.674), si obtuvo el título profesional (1.797), si su primer empleo está relacionado con la formación profesional (1.310); para las variables sobre la calificación sobre la preparación recibida en la universidad para el desarrollo de las competencias son: la calificación de la preparación para coordinar actividades como excelente (1.678) y como bueno (1.405), la calificación de la preparación en conocimientos básicos de otros campos o disciplinas como excelente (1.986), bueno (1.624) y regular (1.520), y la calificación de la preparación para tener pensamiento crítico como excelente (0.771) y bueno (0.717). Mientras para las universidades privadas, las variables más importantes son: el sexo (1.796), si su padre alcanzó el nivel de educación primaria (0.449), si su padre alcanzó el nivel de educación secundaria (0.516), si perteneció al cuadro de méritos (1.515), si obtuvo el título profesional (1.795), la calificación de la preparación en el dominio del área de disciplina como excelente (1.909) y como bueno (1.412), la calificación de la preparación para trabajar en equipo como excelente (0.832), como bueno (0.885), la calificación de la preparación para utilizar herramientas informáticas básicas como excelente (1.046) y como bueno (1.001), la calificación de la preparación para tener pensamiento crítico como excelente (0.863) y como bueno (0.807).
2. Las variables más importantes, para explicar los rangos de ingreso del primer empleo de los egresados de universidades públicas, según el Modelo 2, evaluando los valores Odds-ratio para las variables socio-académicas son: el sexo (2.079), si su padre alcanzó el nivel de educación primaria (0.672), si su padre alcanzó el nivel de educación secundaria (0.715), si perteneció al cuadro de méritos (1.645), si obtuvo el título profesional (1.770), si su primer empleo está relacionado con la formación profesional (1.304); para las variables sobre la importancia de las competencias es: importancia de redactar informes o documentos como “muy importante” (1.392). Mientras para las universidades privadas, las variables más importantes son: el sexo (1.841), si su padre alcanzó el nivel de educación primaria (0.454), si su padre alcanzó el nivel de educación secundaria (0.523),

si perteneció al cuadro de méritos (1.469), si obtuvo el título profesional (1.807), importancia de rendir bajo presión y cumplir ciertos objetivos como “muy importante” (1.355), importancia de coordinar actividades como “muy importante” (1.212), importancia de leer, escribir y hablar en idioma(s) extranjero(s) como “muy importante” (1.281) e importancia de leer, escribir y hablar en idioma(s) extranjero(s) como “importante” (1.103).

3. Las variables más importantes, para explicar los rangos de ingreso del primer empleo de los egresados de universidades públicas, según el Modelo 3, evaluando los valores Odds-ratio para las variables socio-académicas son: el sexo (2.036), si su padre alcanzó el nivel de educación primaria (0.676), si su padre alcanzó el nivel de educación secundaria (0.714), si perteneció al cuadro de méritos (1.669), si obtuvo el título profesional (1.774), si su primer empleo estuvo relacionado con su formación profesional (1.313); para las variables sobre la calificación a los profesores de su carrera son: calificación sobre la calidad de los profesores de su carrera como “excelente” (1.218) y calificación sobre la calidad de los profesores de su carrera como “bueno” (1.047). Mientras para las universidades privadas, las variables más importantes son: el sexo (1.813), si su padre alcanzó el nivel de educación primaria (0.430), si su padre alcanzó el nivel de educación secundaria (0.504), si perteneció al cuadro de méritos (1.523) y si obtuvo el título profesional (1.790).
4. Respecto a las pruebas de bondad de ajuste, con excepción del Modelo 2 para universidades públicas, los tres modelos propuestos para las universidades públicas y privadas resultaron con evidencia estadística significativa para las pruebas de ajuste global (Log-Verosimilitud y Chi-Cuadrado) y las pruebas de bondad de ajuste (Pearson y Deviance) con un nivel de significancia de 0.05, indicando un ajuste de los modelos. De igual forma, las pruebas de líneas paralelas usando las estadísticas de Log-verosimilitud y Chi-cuadrado, resultaron no significativas con 0.05 con lo cual se concluye que los coeficientes de las variables independientes del modelo logístico ordinal para explicar los rangos de ingresos del primer empleo de los egresados de universidades públicas, y de los egresados de las universidades privadas, son los mismos para todas las categorías.
5. Los ingresos promedios de los egresados de las universidades privadas son mayores que las públicas para las variables socio-académicas, sobre la calificación de la preparación recibida en la universidad para el desarrollo de sus competencias y respecto a la importancia de las competencias recibidas para su experiencia profesional.

VI. RECOMENDACIONES

Las recomendaciones de la presente investigación son:

1. Identificar otras variables relevantes para explicar los ingresos de los egresados universitarios; tales como: ingresos y nivel estudios de los de los padres, rendimiento en el nivel secundario y pre-académico, etc.
2. Aplicar y comparar los modelos logísticos ordinales propuestos, segmentando la base de datos según el lugar de las universidades por región (norte, centro y sur), por Lima y universidades de provincias.
3. Hacer un estudio comparativo aplicando técnicas de clasificación de minería de datos tales como; redes bayesianas, redes neuronales, etc.

VII. REFERENCIA BIBLIOGRÁFICA

- Baron, J. D. (2010). Documentos de trabajo sobre economía regional . *Centro de estudios econométricos regionales-CEER* .
- Blac, D., & Smith, J.A. (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics* 24(3)., pp. 701-728.
- Cohodes, S. R., & Goodman, J. S. (2012). First degree earns: The impact of college quality on college completion rates.
- Dale, S., & Krueger, A. B. (2011). Estimating the return to college selectivity over the career using administrative earnings data. *National Bureau of Economic Research*.
- De Vries, R. (2014). Earning by Degrees Differences in the career outcomes of UK graduates. *Improving social mobility through education*, pp. 2-75.
- Diaz, J., & Jaramillo, M. (2008). Educación superior en el Perú: tendencias de la demanda y la oferta.
- Forero, N., & Ramírez, M. (2008). Determinantes de los ingresos laborales de los graduados universitarios durante el periodo 2001-2004.
- Forero, N., & Ramírez, M. (2008). Determinantes de los ingresos laborales de los graduados universitarios durante el periodo 2001-2004.
- Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics*, 9(14)., pp. 717-724.
- Martínez Zamora, M. (2015). Determinantes del salario de los recién graduados de educación superior a nivel . *Tesis para optar el grado de Maestría en ciencias Económicas*.
- McCullagh, P., & Nelder, J.A. (1980). Generalized Linear Models.
- McGuinness. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, pp. 387-418.
- Milla, J. (2012). University Quality and Labor Market: Outcomes of Canadian Youth. *University of Guelph*.
- Xiaowei Xul, & Weiwei Zhang. (2015). Analysis of the Influence Factors of the Ability of Graduate Employment Based on the Logistic Regression Model. *International Journal of Control and Automation Vol. 8- N° 9.*, pp. 405-412.
- Yamada, G., & Castro, J.F. (2013). Evolución reciente de la calidad de la educación superior en el Perú: no son buenas noticias. *Consejo de Evaluación, Acreditación y Certificación de la Calidad de la Educación*.

Yamada, G., Castro, J.F., Bacigalupo, J., & Velarde, L. (2013). Mayor acceso con menor calidad en la educación superior: Algunas evidencias desde las habilidades de los estudiantes. *Apuntes 72. Educación: Calidad y procesos. Universidad del Pacífico.*

VIII. ANEXOS

Anexo 1. Distribución de frecuencias de las categorías de las variables independientes

Modelo 1. Distribución de las categorías de las variables independientes

Variable	Categorías	N	Porcentaje
X_P11	Hombre	3259	44,4%
	Mujer	4085	55,6%
X_P205	Primaria	708	9,6%
	Secundaria	2114	28,8%
	Superior	4522	61,6%
X_P324	Sí	3816	52,0%
	No	3528	48,0%
X_P330	Sí	3234	44,0%
	No	4110	56,0%
X_P349	Sí	6938	94,5%
	No	406	5,5%
X_P456	Sí	3298	44,9%
	No	4046	55,1%
X_P471	Sí	3266	44,5%
	No	4078	55,5%
X_P603_1	Excelente	1345	18,3%
	Bueno	4565	62,2%
	Regular	1346	18,3%
	Malo	88	1,2%
X_P603_2	Excelente	1573	21,4%
	Bueno	4339	59,1%
	Regular	1302	17,7%
	Malo	130	1,8%
X_P603_3	Excelente	1245	17,0%
	Bueno	4567	62,2%
	Regular	1395	19,0%
	Malo	137	1,9%
X_P603_4	Excelente	1802	24,5%
	Bueno	4342	59,1%
	Regular	1088	14,8%
	Malo	112	1,5%
X_P603_5	Excelente	1697	23,1%
	Bueno	4153	56,5%
	Regular	1326	18,1%
	Malo	168	2,3%
X_P603_6	Excelente	1411	19,2%
	Bueno	3980	54,2%
	Regular	1647	22,4%

	Malo	306	4,2%
X_P603_7	Excelente	1184	16,1%
	Bueno	3198	43,5%
	Regular	2125	28,9%
	Malo	837	11,4%
X_P603_8	Excelente	1253	17,1%
	Bueno	4020	54,7%
	Regular	1745	23,8%
	Malo	326	4,4%
X_P603_9	Excelente	1267	17,3%
	Bueno	4268	58,1%
	Regular	1572	21,4%
	Malo	237	3,2%
X_P603_10	Excelente	688	9,4%
	Bueno	2592	35,3%
	Regular	2893	39,4%
	Malo	1171	15,9%
X_P603_11	Excelente	669	9,1%
	Bueno	3666	49,9%
	Regular	2579	35,1%
	Malo	430	5,9%
X_P603_12	Excelente	1426	19,4%
	Bueno	4521	61,6%
	Regular	1277	17,4%
	Malo	120	1,6%
Total		7344	

Modelo 2. Distribución de las categorías de las variables independientes

Variable	Categorías	N	Porcentaje
X_P11	Hombre	3259	44,4%
	Mujer	4085	55,6%
X_P205	Primaria	708	9,6%
	Secundaria	2114	28,8%
	Superior	4522	61,6%
X_P324	Sí	3816	52,0%
	No	3528	48,0%
X_P330	Sí	3234	44,0%
	No	4110	56,0%
X_P349	Sí	6938	94,5%
	No	406	5,5%
X_P456	Sí	3298	44,9%
	No	4046	55,1%
X_P471	Sí	3266	44,5%
	No	4078	55,5%
X_P604_1	Muy importante	4952	67,4%

	Importante	2286	31,1%
	Poco importante	91	1,2%
	Nada importante	15	,2%
X_P604_2	Muy importante	4391	59,8%
	Importante	2722	37,1%
	Poco importante	214	2,9%
	Nada importante	17	,2%
X_P604_3	Muy importante	4137	56,3%
	Importante	3060	41,7%
	Poco importante	138	1,9%
	Nada importante	9	,1%
X_P604_4	Muy importante	5161	70,3%
	Importante	2070	28,2%
	Poco importante	105	1,4%
	Nada importante	8	,1%
X_P604_5	Muy importante	4798	65,3%
	Importante	2407	32,8%
	Poco importante	131	1,8%
	Nada importante	8	,1%
X_P604_6	Muy importante	4083	55,6%
	Importante	2972	40,5%
	Poco importante	265	3,6%
	Nada importante	24	,3%
X_P604_7	Muy importante	3403	46,3%
	Importante	3000	40,8%
	Poco importante	818	11,1%
	Nada importante	123	1,7%
X_P604_8	Muy importante	3536	48,1%
	Importante	3277	44,6%
	Poco importante	478	6,5%
	Nada importante	53	,7%
X_P604_9	Muy importante	3822	52,0%
	Importante	3134	42,7%
	Poco importante	357	4,9%
	Nada importante	31	,4%
X_P604_10	Muy importante	2527	34,4%
	Importante	3165	43,1%
	Poco importante	1441	19,6%
	Nada importante	211	2,9%
X_P604_11	Muy importante	2526	34,4%
	Importante	3992	54,4%
	Poco importante	774	10,5%
	Nada importante	52	,7%
X_P604_12	Muy importante	4347	59,2%
	Importante	2792	38,0%
	Poco importante	185	2,5%
	Nada importante	20	,3%
Total		7344	

Modelo 3. Distribución de las categorías de las variables independientes

Variable	Categorías	N	Porcentaje
X_P11	Hombre	3259	44,4%
	Mujer	4085	55,6%
X_P205	Primaria	708	9,6%
	Secundaria	2114	28,8%
	Superior	4522	61,6%
X_P324	Sí	3816	52,0%
	No	3528	48,0%
X_P330	Sí	3234	44,0%
	No	4110	56,0%
X_P349	Sí	6938	94,5%
	No	406	5,5%
X_P456	Sí	3298	44,9%
	No	4046	55,1%
X_P471	Sí	3266	44,5%
	No	4078	55,5%
X_P605_1	Excelente	1404	19,1%
	Bueno	4239	57,7%
	Regular	1559	21,2%
	Malo	142	1,9%
X_P605_2	Excelente	1329	18,1%
	Bueno	4164	56,7%
	Regular	1678	22,8%
	Malo	173	2,4%
X_P605_3	Excelente	1385	18,9%
	Bueno	3817	52,0%
	Regular	1844	25,1%
	Malo	298	4,1%
X_P605_4	Excelente	1033	14,1%
	Bueno	3990	54,3%
	Regular	2060	28,1%
	Malo	261	3,6%
Total		7344	

