



# TRABAJO FINAL DE ESTADISTICA

Características socioeconómicas y laborales de los asalariados privados en el empleo informal en el Ecuador, periodo 2021.

## Estadística para científicos de datos

[https://github.com/andresanaluisa/Estadistica\\_andres\\_analuisa.git](https://github.com/andresanaluisa/Estadistica_andres_analuisa.git)

Luis Andres Analuisa  
And\_lui88@hotmail.com

## Contents

<b>2. Introducción</b>	3
<b>3. Variables</b>	3
3.1. Tipo de variables	4
<b>4. Objetico</b>	4
5.1. Poblaciones	4
5.2. Informalidad	5
5.3. Jefe de hogar	5
5.4. Area	6
5.5. Etnia	6
5.6. Dominio geográfico	7
5.7. Actividad	7
5.8. Salarios	8
5.9. Años de estudio	8
5.10. Años de experiencia	9
<b>6. Metodología</b>	10
6.1. Modelo Logit	10
6.2. Odds Ratios	11
<b>7. Análisis post estimación</b>	12
7.1. R2 Mc Fadden	12
7.2. Matriz de confusión	12
<b>8. Diferencia de salarios</b>	14
8.1. Hipótesis	14
8.2. Parámetro estimado (estadístico)	14
8.3. Normalidad	14
8.4. Homocedasticidad	16
8.5. Nivel de significancia	16
8.6. Cálculo de p-value y tamaño del efecto	17
<b>9. Conclusiones</b>	17
<b>10. Bibliografía</b>	18
<b>11. Anexos</b>	18

## Tabla de ilustraciones:

Ilustración 1 Poblaciones proyectadas y mercado laboral .....	4
Ilustración 2 Sexo e informalidad .....	5
Ilustración 3 Jefe de hogar en la informalidad .....	5
Ilustración 4 Area en la informalidad.....	6
Ilustración 5 Etnia en la informalidad .....	7
Ilustración 6 Dominio geográfico en la informalidad.....	7
Ilustración 7 Actividad en la informalidad .....	8
Ilustración 8 Media de empleo de salarios por informalidad .....	8
Ilustración 9 Años de estudio en la informalidad. ....	8
Ilustración 10 Años de experiencia y la informalidad .....	9
Ilustración 11 Odds Ratios .....	11
Ilustración 12 Normalidad.....	14
Ilustración 13 Hocesdasticidad .....	16

## 1. Tema

Características socioeconómicas y laborales de los asalariados privados en el empleo informal en el Ecuador, periodo 2021.

[https://github.com/andresanaluisa/Estadistica\\_andres\\_analuisa.git](https://github.com/andresanaluisa/Estadistica_andres_analuisa.git)

## 2. Introducción

El fin de esta investigación es identificar las características socioeconómicas y laborales de los asalariados privados en el Ecuador. Se plantea la hipótesis que un asalariado en el sector privado que labora en la informalidad no solo depende de características individuales como género, edad y nivel de educación si no por características laborales y otros factores socioeconómicos que son dispuestos en este apartado.

El presente estudio es empírico y estima un modelo probit de empleo informal determinando los resultados de probabilidad de ocurrencia de las variables de selección, la información base ha sido tomada de los microdatos de las encuestas de hogares (ENENDU) donde recoge las recomendaciones de organismos internacionales y nacionales, esta se aplica a una muestra continua a lo largo de todo el año, para este análisis se considera el periodo 2021 con datos de corte transversal, la variables de referencia es Sectorización de los empleados 15 años y más “secemp”.

Las estadísticas del sector en empleo informal en el Ecuador producidas por el Institución encargadas de realizar el proceso estadístico fueran entregadas en julio del año 2015, por recomendaciones de la OIT se encuentran presentes en el año 2022.

De acuerdo a la resolución Resolución 15ª CIET (1993) al sector informal se lo clasifica como “...todas las personas que, durante un período determinado de referencia, están ocupadas en al menos una empresa del sector informal, independientemente de su situación en el empleo...”.

La base de datos cuenta con 361790 observaciones y 151 variables que han pasado por el proceso estadístico y han sido dispuestas por el INEC para la población ecuatoriana.

Adicional se plantea un análisis estadístico de la diferencia de salarios en el mercado laboral de la informalidad.

## 3. Variables

Al ser un estudio empírico se ha considerado las variables dispuestas por en la ENEMDU seleccionadas mediante revisión bibliográfica y valoradas mediante el modelo de regresión logístico logit determinado su significancia, (Coaguila, 2020)

"ciudad"

"conglomerado"

"vivienda"

"hogar"

"informalidad"

"sexo"

"jefeh"

"area"  
"etnia"  
"region"  
"rama"  
"salario"  
"educacion"  
"exper"

### 3.1. Tipo de variables

La encuesta recoge información de microdatos, las variables utilizadas son de tipo en su mayoría son categóricas (ordinal y nominal) excepto el salario que es cuantitativa continua.

## 4. Objetico

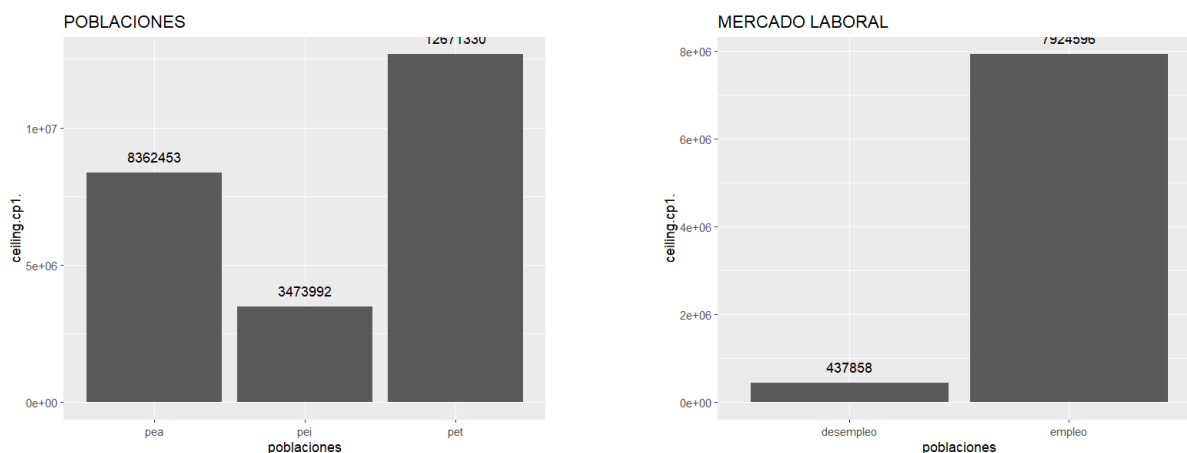
- Identificar las características socioeconómicas y laborales de los asalariados privados en el Ecuador en el periodo 2021, mediante la regresión logística.
- Se quiere determinar si existen evidencias significativas de que el salario de los trabajadores con empelo formal difieren del empleo formal.

## 5. Exploración de datos:

### 5.1. Poblaciones

En el año 2021 se contó con una Población en edad de trabajar PET de 12671330, una Población con económicamente activa (PEA) de 8362453, una población económicamente inactiva (PEI) de 3473992, de igual manera se observa el mercado laboral con una población con empleo de 7924596 y una población en desempleo de 437858.

*Ilustración 1 Poblaciones proyectadas y mercado laboral*



5.2. Informalidad

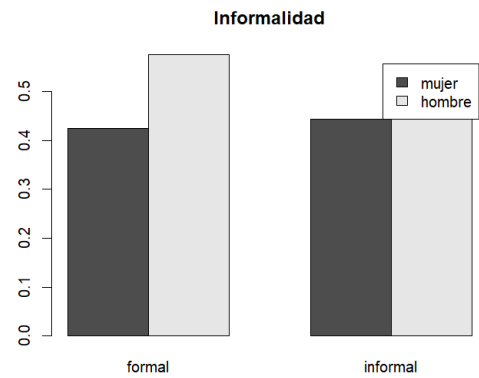
En el mercado laboral las personas que trabajan bajo empleo formal representan el 53% y las personas que se encuentran en informalidad el 47%.

Sexo e informalidad

Al realizar un análisis de la informalidad por sexo se puede decir que la participación del hombre es superior en el empleo formal e informal.

Ilustración 2 Sexo e informalidad

	formal	informal
mujer	42%	44%
hombre	58%	56%

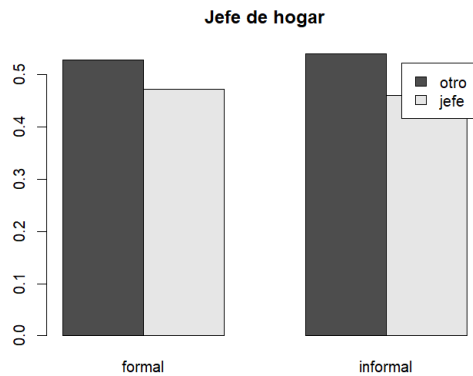


5.3. Jefe de hogar

Al realizar un análisis de la informalidad por jefe de hogar se puede decir que la participación del jefe de hogar en la informalidad es menor con el 46% en el empleo informal en relación a las otras caracterizaciones.

Ilustración 3 Jefe de hogar en la informalidad

	formal	informal
otro	53%	54%
jefe	47%	46%

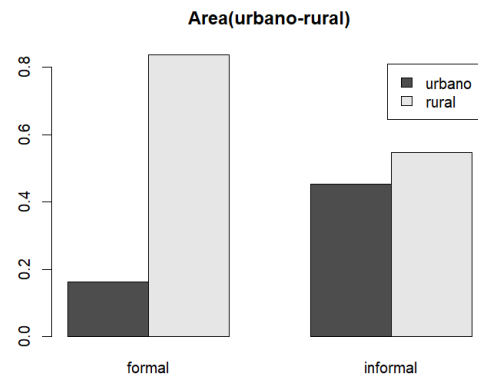


#### 5.4. Area

Al realizar un análisis de la informalidad por área se puede decir que el empleo formal tiene mayor representatividad 84% en el sector rural y el empleo informal tiene mayor representatividad en el sector rural 55%.

*Ilustración 4 Area en la informalidad*

	formal	informal
urbano	16%	45%
rural	84%	55%

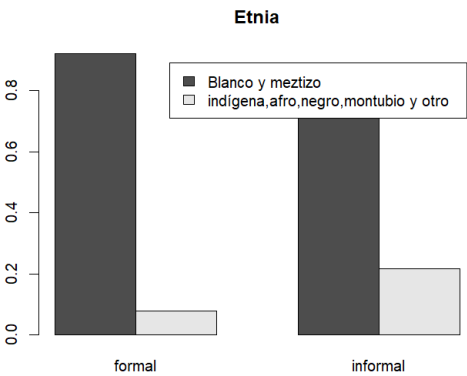


#### 5.5. Etnia

Al realizar un análisis de la informalidad por etnia se puede ver que el detalle Blanco y mestizo tiene mayor 62% de presencia en el empleo formal mientras que en el empleo informal la población indígena, afro, negro, montubio y otro tiene una presencia en la población informal con el 78%.

Ilustración 5 Etnia en la informalidad

	formal	informal
Blanco y meztizo	92%	78%
indígena,afro,negro,montubio y otro	8%	22%

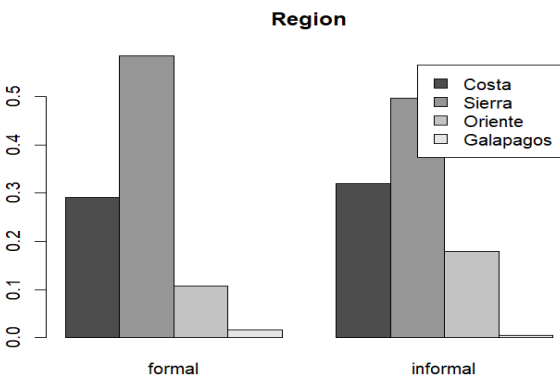


5.6. Dominio geográfico

Al realizar un análisis de la informalidad por dominio geográfico se puede ver que en la región sierra existe 58 % de presencia en el empleo formal mientras que en el empleo informal la región sierra también tiene la mayor presencia con el 50%.

Ilustración 6 Dominio geográfico en la informalidad

	formal	informal
Costa	29%	32%
Sierra	58%	50%
Oriente	11%	18%
Galapagos	2%	1%



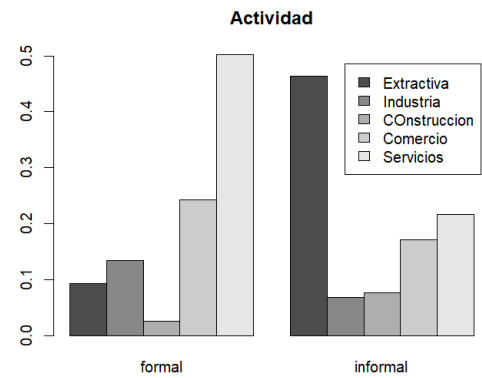
5.7. Actividad

Al realizar un análisis de la informalidad por rama de actividad se puede ver que en servicios tiene el 58 % de presencia en el empleo formal mientras las actividades extractivas superan con el 46% el empleo informal.



Ilustración 7 Actividad en la informalidad

	formal	informal
Extractiva	9%	46%
Industria	14%	7%
Construccion	3%	8%
Comercio	24%	17%
Servicios	50%	22%



5.8. Salarios

Al realizar un análisis de los salarios se observa una media de salarios en el empleo formal de usd 886 para el empleo formal y una media de usd de 274 en el empleo informal.

Ilustración 8 Media de empleo de salarios por informalidad

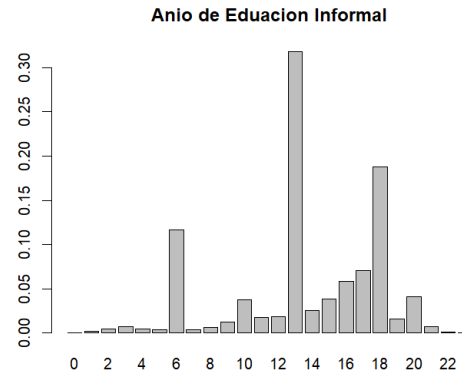
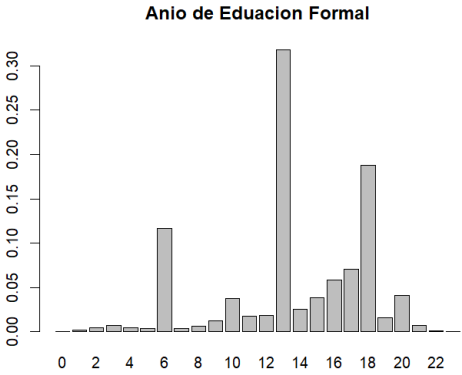
	Sueldo
formal	886
informal	274

5.9. Años de estudio

Al realizar un estudio de la formalidad y los años de estudio y se observa que en la formalidad a mayor años de estudio existe mayor empleo formal , mientras que en la informalidad se observa que a mayor edad mayor participación en empelo de informalidad.

Ilustración 9 Años de estudio en la informalidad.

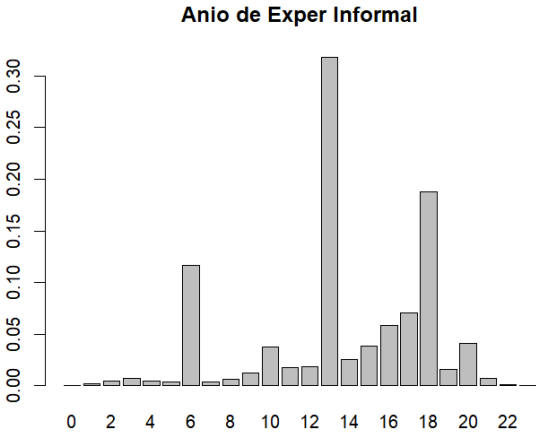
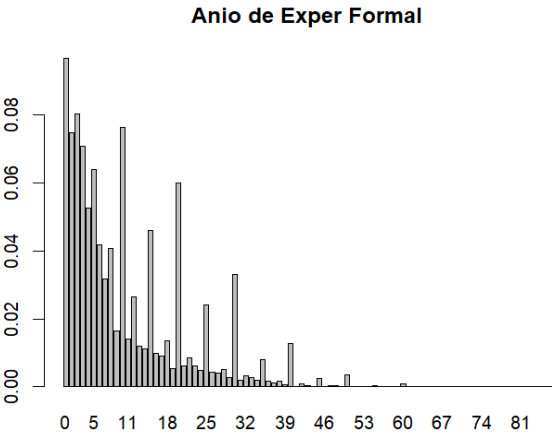
Anios	formal	informal
0	0,0%	0,0%
1	0,2%	1,7%
2	0,4%	2,9%
3	0,7%	3,8%
4	0,4%	2,3%
5	0,3%	1,5%
6	11,7%	29,3%
7	0,4%	1,6%
8	0,7%	1,8%
9	1,3%	2,9%
10	3,8%	6,9%
11	1,8%	3,6%
12	1,8%	3,4%
13	31,8%	27,3%
14	2,5%	1,5%
15	3,9%	1,9%
16	5,8%	2,2%
17	7,0%	1,7%
18	18,8%	3,4%
19	1,6%	0,2%
20	4,1%	0,2%
21	0,7%	0,0%
22	0,1%	0,0%
23	0,0%	0,0%



5.10. Años de experiencia

Al analizar los años de experiencia en la informalidad se observa que la participación porcentual en el formalidad por años es experiencia se alcanza con la menor experiencia, mientras que en la informalidad se observa que la participación porcentual en la población informal se alcanza a mayor edad de experiencia.

Ilustración 10 Años de experiencia y la informalidad



## 6. Metodología

En función a los objetivos del presente documento y a la información disponible, para identificar los determinantes del empleo informal en los asalariados privados del mercado laboral ecuatoriano, se utilizó un modelo probit de regresión logística para Identificar las características socioeconómicas y laborales de los asalariados privados en el Ecuador en el periodo 2021, mediante la regresión logística y un análisis anova para determinar la influencia que tiene la región en los salarios promedio de la población de los empleados informales.

### 6.1. Modelo Logit

Cabe resaltar que los coeficientes estimados no son directamente interpretables debido a que en esencia es un modelo probit corregido por sesgo de selección, caracterizado solo por evaluar el sentido de impacto mas no la magnitud del mismo. Por tal razón, se realizó la estimación de los efectos marginales -evaluados en la media de cada variable continua- que permiten cuantificar el efecto que las variables explicativas sobre la probabilidad de tener empleo informal. Dado que la metodología contempla la estimación simultánea de un modelo de selección o participación laboral, se utilizó como variables control al género, la edad y la condición de jefes de hogar; tal y como fue considerado por Garavito (2016) en la estimación de la ecuación de participación del mercado laboral. La estimación obtuvo los signos esperados. Se verifica que el hombre tiene menor probabilidad de participación que la mujer, también a mayor educación menor son las posibilidades de inserción laboral formal. Caso contrario, se registró en los jefes de hogar quienes tuvieron mayor probabilidad de participar en el mercado de trabajo, sobre todo, por las responsabilidades que acarrea tener una familia. Es decir, se confirma lo encontrado por Garavito (2016). Adicional se observa la significancia de los predictores en el modelo.

Coefficients:					
	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	3.105e+00	4.405e-02	70.491	< 2e-16	***
sexohombre	-2.842e-01	1.462e-02	-19.436	< 2e-16	***
jefehjefe	1.420e-01	1.554e-02	9.136	< 2e-16	***
area	3.004e-01	1.887e-02	15.921	< 2e-16	***
etniaindígena, afro, negro, montubio y otro	4.715e-01	2.068e-02	22.796	< 2e-16	***
regionsierra	-4.954e-01	1.518e-02	-32.634	< 2e-16	***
regionOriente	-4.118e-01	2.319e-02	-17.760	< 2e-16	***
regionGalapagos	-1.191e+00	7.432e-02	-16.023	< 2e-16	***
ramaIndustria	-1.671e+00	2.657e-02	-62.916	< 2e-16	***
ramaConstruccion	1.795e-01	3.455e-02	5.196	2.04e-07	***
ramaComercio	-1.363e+00	2.332e-02	-58.426	< 2e-16	***
ramaServicios	-1.477e+00	2.200e-02	-67.141	< 2e-16	***
salariom	-1.279e-03	1.990e-05	-64.275	< 2e-16	***
educacion	-1.365e-01	1.752e-03	-77.870	< 2e-16	***
exper	-4.058e-02	1.646e-03	-24.657	< 2e-16	***
exper2	1.052e-03	3.924e-05	26.810	< 2e-16	***
---					
signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

## 6.2. Odds Ratios

Para cuantificar los resultados del modelo logit es importante analizar los odds ratios que determinan un valor cuantitativo comparativo. En esta etapa se observa un modelo con menos ruido en sus resultados, es importante considerar que esta información muestra el comportamiento del año 2021.

*Ilustración 11 Odds Ratios*

Modelo 1 - Odds Ratios	
sexohombre	0.75 *** (0.01)
jefehjefe	1.15 *** (0.02)
area	1.35 *** (0.02)
etniaindígena, afro, negro, montubio y otro	1.60 *** (0.02)
regionSierra	0.61 *** (0.02)
regionOriente	0.66 *** (0.02)
regionGalapagos	0.30 *** (0.07)
ramaIndustria	0.19 *** (0.03)
ramaConstruccion	1.20 *** (0.03)
ramaComercio	0.26 *** (0.02)
ramaServicios	0.23 *** (0.02)
salariom	1.00 *** (0.00)
educacion	0.87 *** (0.00)
exper	0.96 *** (0.00)
exper2	1.00 *** (0.00)
AIC	134356.89
BIC	134514.40
Log Likelihood	-67162.45
Deviance	134324.89
Num. obs.	139299
*** p < 0.001; ** p < 0.01; * p < 0.05	

Al interpretar el modelo se puede decir que el hombre tiene la probabilidad de 1.333333 veces más que la mujer de pertenecer a la población de informalidad

Cuando la persona es jefe de hogar tiene una probabilidad de 1.15 más de ser empleado informal que las otras condiciones de pertenecer al empleo informal

Cuando la persona se auto percibe como Indígena, afro, negro, montubio y otro tiene una probabilidad de 1.6 veces mas de pertenecer al empleo informal que las personas que se autoidentifican como blancos o mestizos de pertenecer al empleo informal

Cuando la persona se encuentra laborando en la región sierra tiene una probabilidad de 1.64 veces más que cuando se encuentra en la región costa de pertenecer al empleo informal

Cuando la persona se encuentra en laborando en la región oriente tiene una probabilidad de 1.66 veces más que cuando se encuentra en la región costa de pertenecer al empleo informal

Cuando la persona se encuentra laborando en la regios Galápagos tiene una probabilidad de 3.33 veces más que cuando se encuentra en la región costa de pertenecer al empleo informal

Cuando la persona se encuentra laborando en la industria tiene una probabilidad de 5.26 veces mas que cuando se encuentra laborando en la rama extractiva de pertenecer al empleo informal.

Cuando la persona se encuentra laborando en el construcción tiene una probabilidad de 1.20 veces más que cuando se encuentra laborando en la rama extractiva de pertenecer al empleo informal.

Cuando la persona se encuentra laborando en el comercio tiene una probabilidad de 3.846154 veces más que cuando se encuentra laborando en la rama extractiva de pertenecer al empleo informal.

Cuando la persona se encuentra laborando en el servicio tiene una probabilidad de 4.347826 veces más que cuando se encuentra laborando en la rama extractiva de pertenecer al empleo informal.

Cuando la persona se encuentra laborando en el servicio tiene una probabilidad de 4.347826 veces más que cuando se encuentra laborando en la rama extractiva de pertenecer al empleo informal.

En tanto a la educación y experiencia se observa que a mayor educación y experiencia se reduce la probabilidad de estar en la informalidad.

## 7. Análisis post estimación

### 7.1. R2 Mc Fadden

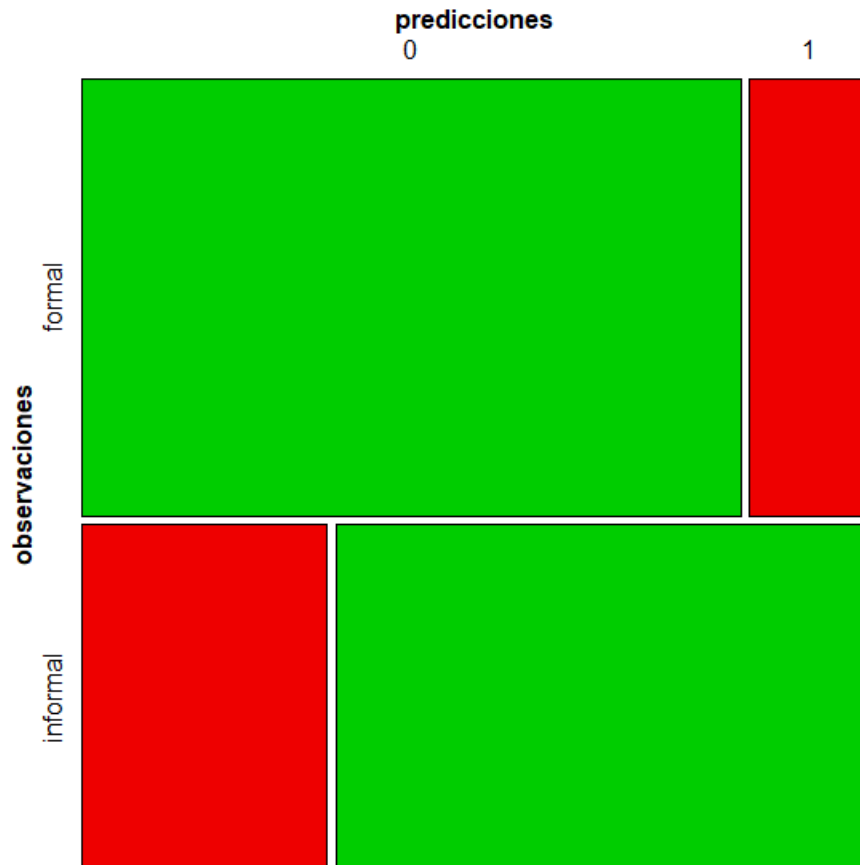
Para el análisis post estimación analizamos el R2 que Básicamente, el **rho-cuadrado** se puede interpretar como R2, los valores de 0.2-0.4 indican (en palabras de McFadden) un excelente ajuste del modelo, apara este caso se cuenta con 0.2971436 considerado como un modelo aceptable.

### 7.2. Matriz de confusión

En el campo de la inteligencia artificial y en especial en el problema de la clasificación estadística, una matriz de confusión es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado. Cada columna de la matriz representa el número de predicciones de

cada clase, mientras que cada fila representa a las instancias en la clase real. Uno de los beneficios de las matrices de confusión es que facilitan ver si el sistema está confundiendo dos clases.

		Predicciones	
		0	1
Observaciones	formal	65904	12074
	informal	19228	42093



## 8. Diferencia de salarios

Se quiere determinar si existe evidencia significativa de la diferencia de salarios en el empleo formal e informal.

### 8.1. Hipótesis

$H_0$  : no hay diferencia entre las medias poblacionales:  $\mu(nf) - \mu(f) = 0$

$H_a$  : si hay diferencia entre las medias poblacionales:  $\mu(nf) - \mu(f) \neq 0$

### 8.2. Parámetro estimado (estadístico)

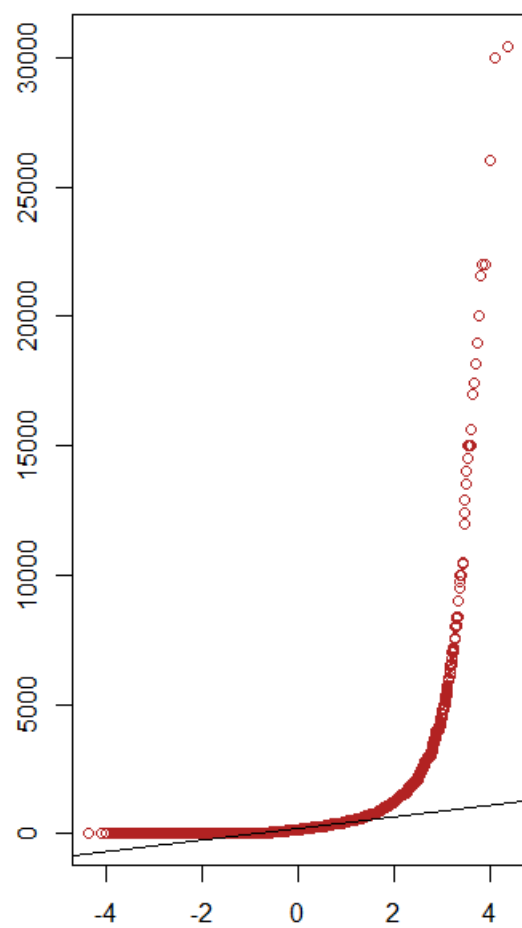
Se realiza una comparación de las medias muestrales: empleo formal e informal.

### 8.3. Normalidad

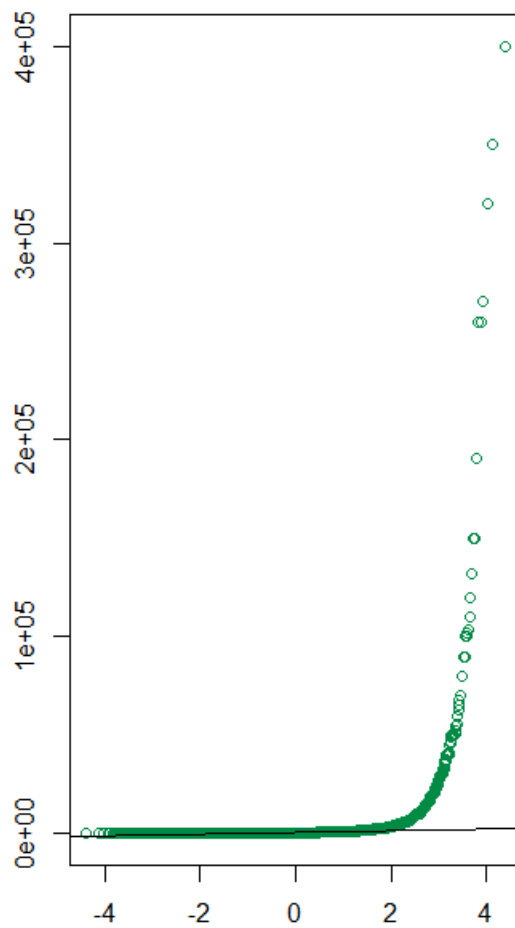
Para determinar la normalidad se utilizó la técnica visual gráficos qqnorm y la prueba estadística estadística de Lilliefors Kolmogorov-Smirnov `lillie.test()` con valores para las muestras de formal: p-value < 2.2e-16 e informal p-value < 2.2e-16, se puede concluir que existe una asimetría hacia la derecha y con un test que encuentra evidencia significativa de que los datos no proceden de poblaciones con distribución normal.

*Ilustración 12 Normalidad*

informal



formal

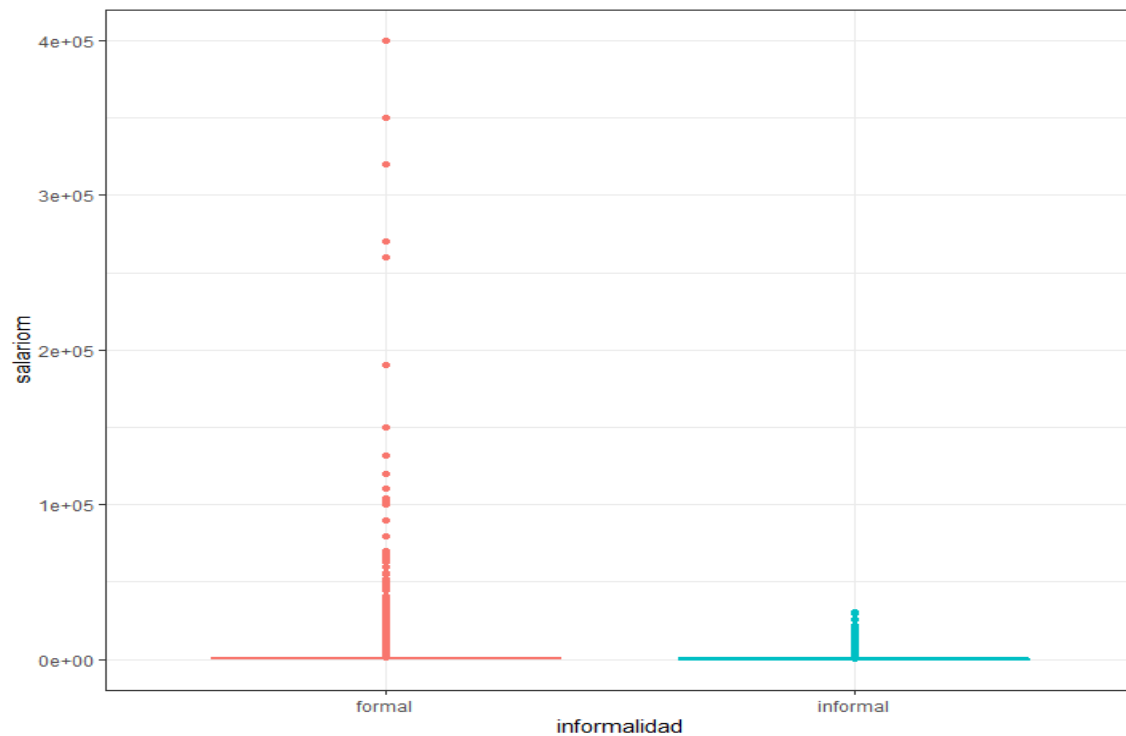




#### 8.4. Homocedasticidad

Dado que no se cumple con el supuesto de normalidad se utiliza el tes basado en la media Levene's (leveneTest ) donde valor p de la prueba es  $2.2e-16$  , que es menor que nuestro nivel de significancia de 0.05. Por lo se concluye que el tes encuentra evidencias significativas (para  $\alpha = 0.05$ ) de que las variancias sean distintas entre ambas poblaciones.

*Ilustración 13 Hocedasticidad*



#### 8.5. Nivel de significancia

$\alpha=0.05$

### 8.6. Cálculo de p-value y tamaño del efecto

Dado que p-value ( $2.2e-16$ ) es menor que alpha (0.05) , se dispone de evidencia suficiente para considerar que existe una diferencia significativa entre el salario promedio de los trabajadores con empleo formal e informal. El tamaño de efecto medido por d-Cohen es pequeño (0.23).

```
ttest
data: formal and informal
t = 45.479, df = 157737, p-value <
2.2e-16
alternative hypothesis: true
difference in means is not equal to 0
95 percent confidence interval:
 585.5647 638.3087
sample estimates:
mean of x mean of y
885.5130 273.5763
```

```
Cohen's d

d estimate: 0.2296386 (small)
95 percent confidence interval:
  lower  upper
0.2197097 0.2395675
```

## 9. Conclusiones

- Con el modelo de regresión logística bajo la literatura seleccionada se ha calculado las determinantes del mercado laboral ecuatoriano y se ha demostrado que las variables son determinantes para el cálculo de la informalidad, en este mismo análisis se encontrando resultados que dejan ver que la informalidad es un problema latente y que exige estrategias estructurales integrales para gestionar esta condición.
- Con los resultados obtenidos se concluye que existe suficiente evidencia estadística para determinar que los salarios promedio de la población en empleo formal e informal son diferentes en las regiones de estudio, es decir la región sí influye en los salarios promedio.

## 10. Bibliografía

Coaguila, J. C. (2020). *Determinantes y duración de la informalidad de los asalariados en Peru*. Pontificia Universidad Católica del Perú.

INEC . (2023). *Boletín Técnico N° 05-2022-ENEMDU* . Quito- Ecuador.

INEC. (2021). *Diccionario de variables* . Quito.

INEC. (2021). *Principales resultados 2021-ENEMDU*. Quito: INEC.

Peru, M. d. (2018). *Informe anual de la mujer en el mercado*. Peru .

## 11. Anexos

[https://github.com/andresanaluisa/Estadistica\\_andres\\_analuisa.git](https://github.com/andresanaluisa/Estadistica_andres_analuisa.git)