

Modelos Logit y Probit Binarios

2025-09-16

- 1 Instalación de Paquetes Necesarios
- 2 Teoría Econométrica de Modelos de Elección Binaria
- 3 Simulación de Datos
- 4 Estimación del Modelo Logit
- 5 Estimación del Modelo Probit
- 6 Efectos Marginales
- 7 Probabilidades Estimadas y Elasticidades
- 8 Modelos con Parámetros Heterogéneos por Ingreso
- 9 Visualización de Probabilidades Estimadas
- 10 Utilidad Marginal Según Nivel de Ingreso
- 11 Comparación Final de Resultados
 - 12.1 Comparación de Modelos Logit y Probit
 - 12.2 Utilidad Marginal según Ingreso
- 13 Visualización: Sensibilidad Según Ingreso
- 14 Conclusiones

1 Instalación de Paquetes Necesarios

Antes de ejecutar este documento, asegúrate de tener instalados los siguientes paquetes:

```
packages <- c("dplyr", "ggplot2", "margins")
install_if_missing <- function(pkg) {
  if (!requireNamespace(pkg, quietly = TRUE)) install.packages(pkg)
}
lapply(packages, install_if_missing)
```

2 Teoría Econométrica de Modelos de Elección Binaria

Los modelos logit y probit binarios son herramientas fundamentales para analizar decisiones discretas. Bajo el marco de **utilidad aleatoria**, se modela la elección entre dos alternativas como función de variables explicativas económicas:

- Utilidad total: $U_{ni} = V_{ni} + \varepsilon_{ni}$
- Decisión: elegir la opción con mayor utilidad
- Probabilidad de elección: $P_{ni} = \Pr(U_{ni} > U_{n0})$

La distribución de los términos de error ε define el modelo:

- **Logit**: $\varepsilon \sim \text{Gumbel} \Rightarrow$ distribución logística
- **Probit**: $\varepsilon \sim N(0, 1) \Rightarrow$ distribución normal estándar

3 Simulación de Datos

```
set.seed(123)
n <- 1000
precio <- runif(n, 400, 700)
costo_operativo <- runif(n, 100, 300)
ingreso <- runif(n, 20000, 90000) / 1000

utilidad <- 5 - 0.003 * precio - 0.015 * costo_operativo + rnorm(n)
prob_compra <- 1 / (1 + exp(-utilidad))
compra <- rbinom(n, 1, prob_compra)

datos <- tibble(compra, precio, costo_operativo, ingreso)
```

4 Estimación del Modelo Logit

```
modelo_logit <- glm(compra ~ precio + costo_operativo, family = binomial("logit"), data = datos)
summary(modelo_logit)
```

```
##
## Call:
## glm(formula = compra ~ precio + costo_operativo, family = binomial("logit"),
##      data = datos)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    4.1798144   0.5370250   7.783 7.07e-15 ***
## precio         -0.0023699   0.0007955  -2.979  0.00289 **
## costo_operativo -0.0131372   0.0012644 -10.390 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1373.7  on 999  degrees of freedom
## Residual deviance: 1246.9  on 997  degrees of freedom
## AIC: 1252.9
##
## Number of Fisher Scoring iterations: 4
```

5 Estimación del Modelo Probit

```
modelo_probit <- glm(compra ~ precio + costo_operativo, family = binomial("probit"), data = datos)
summary(modelo_probit)
```

```
##
## Call:
## glm(formula = compra ~ precio + costo_operativo, family = binomial("probit"),
##      data = datos)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.5513160   0.3207429   7.954  1.8e-15 ***
## precio         -0.0014331   0.0004829  -2.968  0.003 **
## costo_operativo -0.0080577   0.0007515 -10.723 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1373.7  on 999  degrees of freedom
## Residual deviance: 1247.1  on 997  degrees of freedom
## AIC: 1253.1
##
## Number of Fisher Scoring iterations: 4
```

6 Efectos Marginales

```
ef_marginales_logit <- margins(modelo_logit)
ef_marginales_probit <- margins(modelo_probit)
summary(ef_marginales_logit)
```

```
##              factor      AME      SE      z      p      lower upper
## costo_operativo -0.0028 0.0002 -13.3903 0.0000 -0.0033 -0.0024
## precio          -0.0005 0.0002 -3.0277 0.0025 -0.0008 -0.0002
```

```
summary(ef_marginales_probit)
```

```
##              factor      AME      SE      z      p      lower upper
## costo_operativo -0.0029 0.0002 -13.2038 0.0000 -0.0033 -0.0024
## precio          -0.0005 0.0002 -3.0063 0.0026 -0.0008 -0.0002
```

7 Probabilidades Estimadas y Elasticidades

```
datos <- datos |>
  mutate(
    utilidad_estimada_logit = predict(modelo_logit),
    probabilidad_logit = 1 / (1 + exp(-utilidad_estimada_logit)),
    utilidad_estimada_probit = predict(modelo_probit),
    probabilidad_probit = pnorm(utilidad_estimada_probit),
    elasticidad_precio_logit = coef(modelo_logit)["precio"] * precio * (1 - probabilidad_logit),
    elasticidad_costo_logit = coef(modelo_logit)["costo_operativo"] * costo_operativo * (1 - probabilidad_logit)
  )
```

8 Modelos con Parámetros Heterogéneos por Ingreso

```
modelo_logit_ingreso <- glm(compra ~ I(precio / ingreso) + I(costo_operativo / ingreso), family = binomial("logit"), data = datos)
modelo_probit_ingreso <- glm(compra ~ I(precio / ingreso) + I(costo_operativo / ingreso), family = binomial("probit"), data = datos)
summary(modelo_logit_ingreso)
```

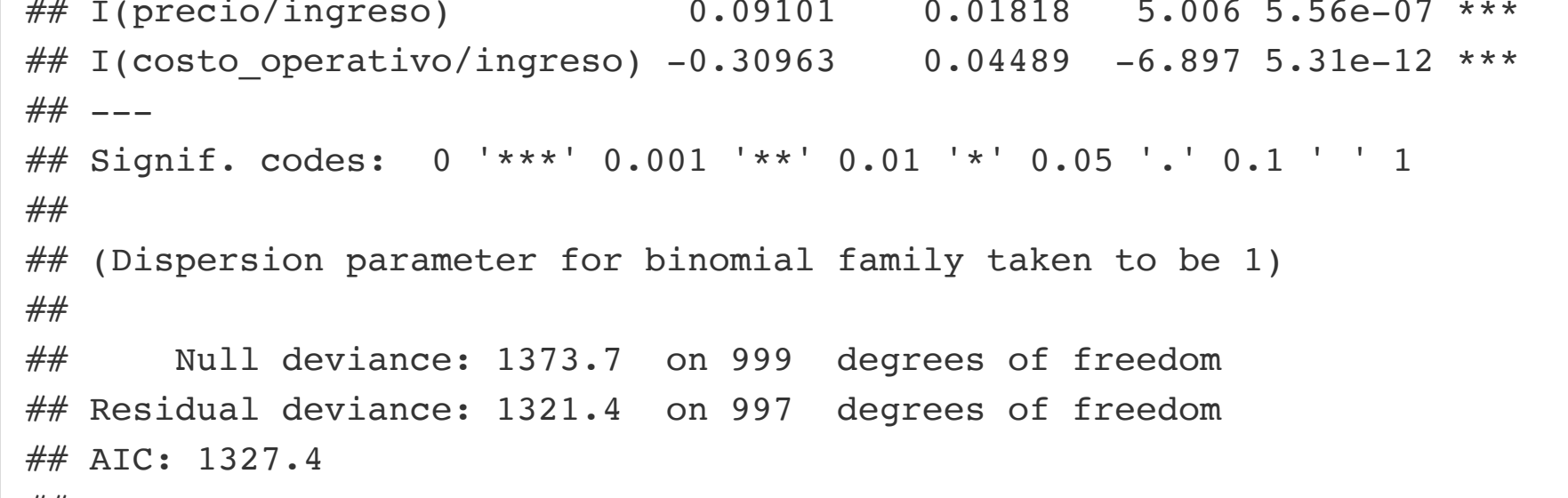
```
##
## Call:
## glm(formula = compra ~ I(precio/ingreso) + I(costo_operativo/ingreso),
##      family = binomial("logit"), data = datos)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.48428   0.15552   3.114  0.00185 **
## I(precio/ingreso)  0.09101   0.01818   5.006 5.56e-07 ***
## I(costo_operativo/ingreso) -0.30963   0.04489 -6.897 5.31e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1373.7  on 999  degrees of freedom
## Residual deviance: 1321.4  on 997  degrees of freedom
## AIC: 1327.4
##
## Number of Fisher Scoring iterations: 4
```

```
summary(modelo_probit_ingreso)
```

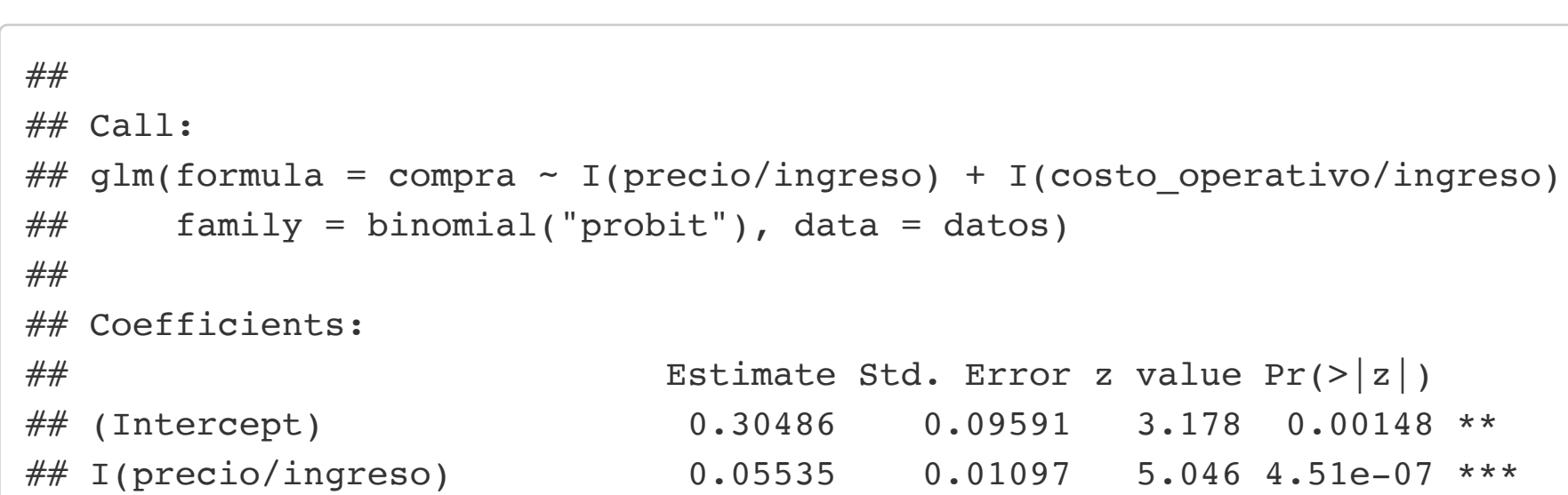
```
##
## Call:
## glm(formula = compra ~ I(precio/ingreso) + I(costo_operativo/ingreso),
##      family = binomial("probit"), data = datos)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.30486   0.09591   3.178  0.00185 **
## I(precio/ingreso)  0.05535   0.01097   5.046 4.51e-07 ***
## I(costo_operativo/ingreso) -0.18970   0.02704 -6.974 2.31e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1373.7  on 999  degrees of freedom
## Residual deviance: 1321.6  on 997  degrees of freedom
## AIC: 1327.6
##
## Number of Fisher Scoring iterations: 4
```

9 Visualización de Probabilidades Estimadas

```
ggplot(datos, aes(x = probabilidad_logit)) +
  geom_density(fill = "steelblue", alpha = 0.5) +
  labs(title = "Probabilidad estimada (Logit)", x = "Probabilidad", y = "Densidad")
```



```
ggplot(datos, aes(x = probabilidad_probit)) +
  geom_density(fill = "darkorange", alpha = 0.5) +
  labs(title = "Probabilidad estimada (Probit)", x = "Probabilidad", y = "Densidad")
```



10 Utilidad Marginal Según Nivel de Ingreso

```
ingresos_miles <- c(30, 60, 90)

utilidad_marginal_logit <- sapply(ingresos_miles, function(inc) coef(modelo_logit_ingreso)[2:3] / inc)
utilidad_marginal_probit <- sapply(ingresos_miles, function(inc) coef(modelo_probit_ingreso)[2:3] / inc)

utilidad_marginal_logit <- matrix(utilidad_marginal_logit, nrow = 3, byrow = TRUE)
utilidad_marginal_probit <- matrix(utilidad_marginal_probit, nrow = 3, byrow = TRUE)

colnames(utilidad_marginal_logit) <- c("Precio/Ingreso", "Costo/Ingreso")
colnames(utilidad_marginal_probit) <- c("Precio/Ingreso", "Costo/Ingreso")
```

11 Comparación Final de Resultados

```
library(knitr)

# Creamos tabla comparativa
comparacion_modelos <- tibble::tibble(
  Modelo = c("Logit", "Probit"),
  LogLikelihood = c(logLik(modelo_logit)[1], logLik(modelo_probit)[1]),
  AIC = c(AIC(modelo_logit), AIC(modelo_probit)),
  BIC = c(BIC(modelo_logit), BIC(modelo_probit))
)
```

```
knitr::kable(comparacion_modelos, digits = 2, caption = "Comparación de desempeño entre modelos Logit y Probit")
```

Comparación de desempeño entre modelos Logit y Probit

Modelo	LogLikelihood	AIC	BIC
Logit	-623.47	1252.94	1267.66
Probit	-623.53	1253.05	1267.78

12 Interpretación de Resultados

12.1 Comparación de Modelos Logit y Probit

Ambos modelos muestran un ajuste adecuado para explicar la probabilidad de compra como función de las variables **precio** y **costo_operativo**. En la tabla siguiente se observa que:

- El modelo **Logit** presenta una log-verosimilitud ligeramente mayor.
- Los valores de **AIC** y **BIC** son también menores para el modelo Logit, lo que sugiere un mejor ajuste relativo bajo estos criterios de información.

```
library(knitr)

comparacion_modelos <- tibble::tibble(
  Modelo = c("Logit", "Probit"),
  LogLikelihood = c(logLik(modelo_logit)[1], logLik(modelo_probit)[1]),
  AIC = c(AIC(modelo_logit), AIC(modelo_probit)),
  BIC = c(BIC(modelo_logit), BIC(modelo_probit))
)
```

```
knitr::kable(comparacion_modelos, digits = 2, caption = "Comparación de desempeño entre modelos Logit y Probit")
```

Comparación de desempeño entre modelos Logit y Probit

Modelo	LogLikelihood	AIC	BIC
Logit	-623.47	1252.94	1267.66
Probit	-623.53	1253.05	1267.78

12.2 Utilidad Marginal según Ingreso

Para interpretar la utilidad marginal ajustada por ingreso, se ha calculado la sensibilidad del consumidor ante variaciones en el precio y el costo operativo según diferentes niveles de ingreso.

```
rownames(utilidad_marginal_logit) <- paste("Ingreso:", ingresos_miles * 1000)
rownames(utilidad_marginal_probit) <- paste("Ingreso:", ingresos_miles * 1000)

knitr::kable(utilidad_marginal_logit, digits = 4, caption = "Utilidad marginal Logit ajustada por ingreso")
```

Utilidad marginal Logit ajustada por ingreso

	Precio/Ingreso	Costo/Ingreso
Ingreso: 30000	0.0030	-0.0103
Ingreso: 60000	0.0015	-0.0052
Ingreso: 90000	0.0010	-0.0034

```
knitr::kable(utilidad_marginal_probit, digits = 4, caption = "Utilidad marginal Probit ajustada por ingreso")
```

Utilidad marginal Probit ajustada por ingreso

	Precio/Ingreso	Costo/Ingreso
Ingreso: 30000	0.0018	-0.0063
Ingreso: 60000	0.0009	-0.0032
Ingreso: 90000	0.0006	-0.0021

La utilidad marginal negativa confirma que los consumidores perciben una desutilidad ante aumentos en el precio o en los costos operativos. Además:

- La magnitud de los cambios de precio decrece con el ingreso, lo que implica que los consumidores con mayor ingreso son menos sensibles a los cambios de precio o costo (menor pendiente de la utilidad).
- Esto es consistente con la teoría económica, donde consumidores con mayor ingreso tienden a tener menor elasticidad-precio.

13 Visualización: Sensibilidad Según Ingreso

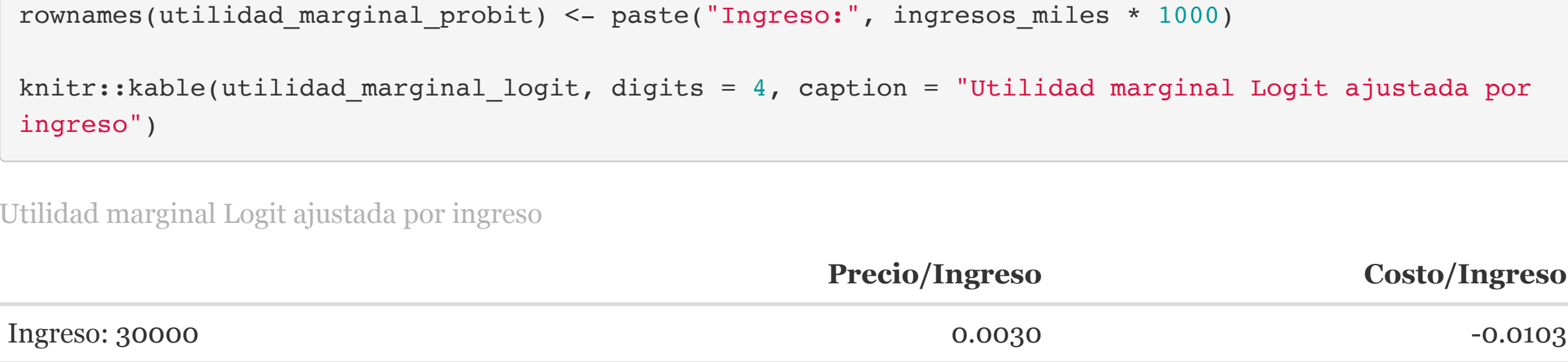
```
# Transformamos a data frame largo para graficar
library(tidyrr)

logit_df <- as.data.frame(utilidad_marginal_logit)
logit_df$Ingreso <- ingresos_miles * 1000
logit_df$Modelo <- "Logit"

probit_df <- as.data.frame(utilidad_marginal_probit)
probit_df$Ingreso <- ingresos_miles * 1000
probit_df$Modelo <- "Probit"

data_long <- bind_rows(logit_df, probit_df) |>
  pivot_longer(cols = c("Precio/Ingreso", "Costo/Ingreso"), names_to = "Variable", values_to = "Valor")

# Visualización
library(ggplot2)
ggplot(data_long, aes(x = Ingreso, y = Valor, color = Modelo, linetype = Variable)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(
    title = "Utilidad marginal según nivel de ingreso",
    x = "Ingreso (en miles)",
    y = "Utilidad marginal",
    color = "Modelo",
    linetype = "Variable"
  ) +
  theme_minimal()
```



14 Conclusiones

Los modelos Logit y Probit binarios permiten analizar la probabilidad de que ocurra un evento discreto (en este caso, la compra de un bien o servicio) en función de variables explicativas económicas como el precio y el costo operativo. A partir de los resultados obtenidos:

- Ambos modelos muestran efectos marginales consistentes: a mayor precio o costo operativo, la probabilidad de compra disminuye.
- Las elasticidades indican que la demanda es sensible a cambios en el precio, con valores negativos esperados.
- El modelo Logit proporciona interpretaciones basadas en log-odds, mientras que el Probit lo hace a partir de la distribución normal acumulada. En la práctica, sus resultados tienden a ser similares.
- La inclusión del ingreso en los modelos permite capturar heterogeneidad no observable, ajustando las decisiones al poder adquisitivo.

En general, ambos enfoques son robustos para analizar elecciones binarias. La elección entre Logit y Probit puede depender de supuestos teóricos sobre la distribución del error, preferencias de interpretación, o conveniencia empírica.