

# Análise da Participação Feminina no Mercado de Trabalho

Alison Cordeiro Sousa

## Análise da Participação Feminina no Mercado de Trabalho

Este estudo examina os fatores que influenciam a decisão de mulheres casadas participarem do mercado de trabalho, utilizando dados de 753 mulheres. Foram comparados três modelos estatísticos: um modelo linear de probabilidade (LPM), um modelo Probit e um modelo Logit. As variáveis explicativas incluíram renda não salarial da esposa (nwifeinc), anos de educação (educ), experiência profissional (exper), idade (age) e número de filhos menores de 6 anos (kidslt6). O objetivo principal era entender como esses fatores afetam a probabilidade de participação no mercado de trabalho, considerando as limitações do LPM e as vantagens dos modelos não lineares Probit e Logit para variáveis binárias.

## Resultados e Interpretação

Os três modelos mostraram padrões consistentes: a renda não salarial e a presença de filhos pequenos reduzem a probabilidade de participação no mercado, enquanto maior educação e experiência aumentam essa probabilidade. O efeito marginal médio no Probit indicou que cada ano adicional de educação aumenta a probabilidade em 4%, enquanto cada filho pequeno reduz em 27.1%. O modelo Logit apresentou resultados similares, com efeitos marginais de 4.1% para educação e -26.6% para filhos pequenos. Ambos os modelos não lineares tiveram melhor desempenho preditivo (acurácia de 73.8%) comparado ao LPM, confirmando sua adequação para análise de variáveis binárias. O pseudo  $R^2$  de 0.21 sugere que o modelo explica parte significativa, mas não toda, da variação na decisão de participar do mercado de trabalho.

```
# --- Limpar ambiente e carregar pacotes ---
rm(list = ls())
setwd("C:/Users/PC GAMER/Downloads/data")
pacotes <- c("tidyverse", "stargazer", "magrittr", "margins", "ggplot2")
for(p in pacotes){ if(!require(p, character.only = TRUE))
  { install.packages(p, dependencies=TRUE); library(p, character.only=TRUE) } }

# --- Carregar e preparar dados ---
IFW <- read.csv("IFW.csv")
dados <- IFW %>% select(inlf, nwifeinc, educ, exper, age, kidslt6)
summary(dados)
```

```
##          inlf          nwifeinc          educ          exper
## Min.      :0.0000   Min.      :-0.02906   Min.      : 5.00   Min.      : 0.00
```

```
## 1st Qu.:0.0000 1st Qu.:13.02504 1st Qu.:12.00 1st Qu.: 4.00
## Median :1.0000 Median :17.70000 Median :12.00 Median : 9.00
## Mean :0.5684 Mean :20.12896 Mean :12.29 Mean :10.63
## 3rd Qu.:1.0000 3rd Qu.:24.46600 3rd Qu.:13.00 3rd Qu.:15.00
## Max. :1.0000 Max. :96.00000 Max. :17.00 Max. :45.00
## age kidslt6
## Min. :30.00 Min. :0.0000
## 1st Qu.:36.00 1st Qu.:0.0000
## Median :43.00 Median :0.0000
## Mean :42.54 Mean :0.2377
## 3rd Qu.:49.00 3rd Qu.:0.0000
## Max. :60.00 Max. :3.0000
```

```
stargazer(dados, type = "text")
```

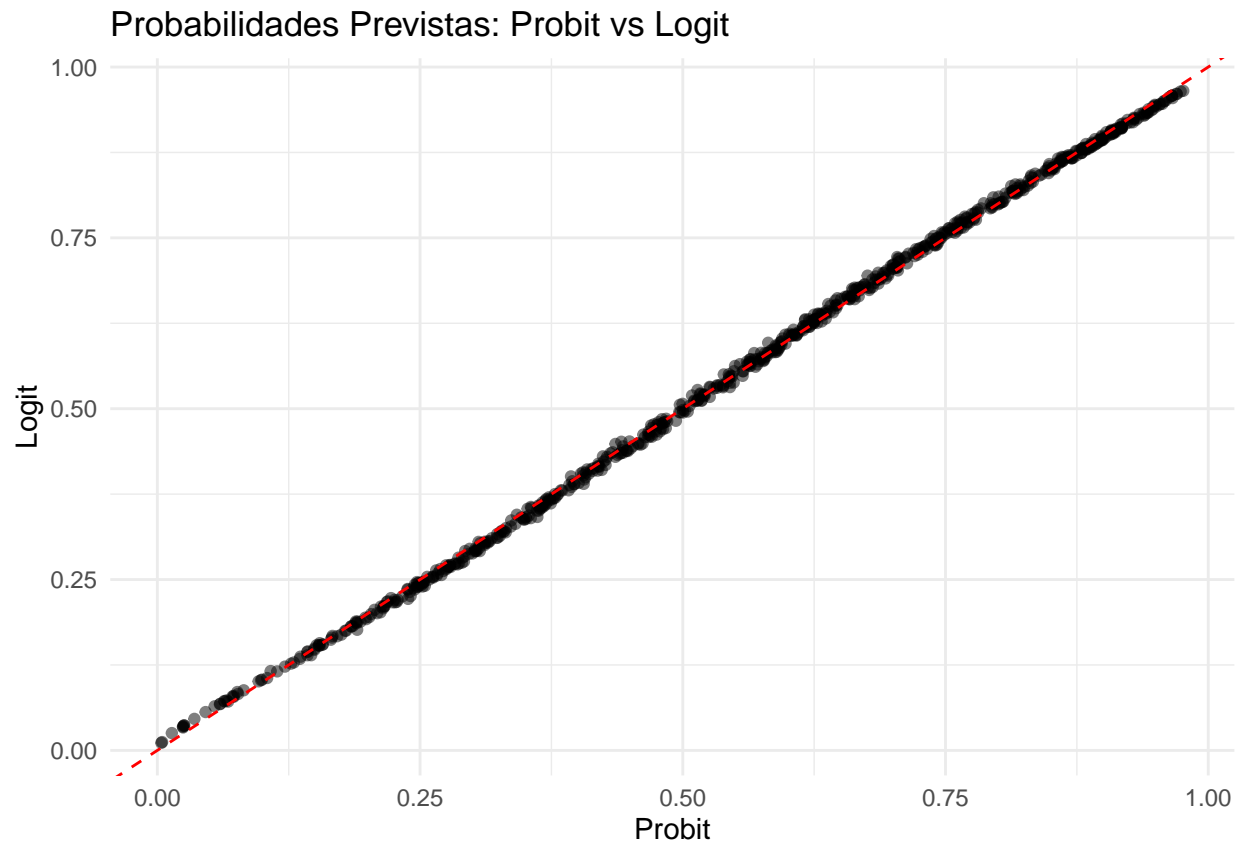
```
##
## =====
## Statistic N Mean St. Dev. Min Max
## -----
## inlf 753 0.568 0.496 0 1
## nwifeinc 753 20.129 11.635 -0.029 96.000
## educ 753 12.287 2.280 5 17
## exper 753 10.631 8.069 0 45
## age 753 42.538 8.073 30 60
## kidslt6 753 0.238 0.524 0 3
## -----
```

```
# --- Modelos estatísticos ---
LPM <- lm(inlf ~ nwifeinc + educ + exper + age + kidslt6, data = dados)
Probit <- glm(inlf ~ nwifeinc + educ + exper + age + kidslt6,
              family = binomial(link = "probit"), data = dados)
Logit <- glm(inlf ~ nwifeinc + educ + exper + age + kidslt6,
             family = binomial(link = "logit"), data = dados)

# --- Adicionar previsões ---
dados <- dados %>% mutate(inlfhat_lpm = fitted(LPM),
                          inlfhat_probit = fitted(Probit), inlfhat_logit = fitted(Logit))
head(dados %>% select(inlf, inlfhat_lpm, inlfhat_probit, inlfhat_logit))
```

```
## inlf inlfhat_lpm inlfhat_probit inlfhat_logit
## 1 1 0.6472128 0.6726774 0.6810984
## 2 1 0.7316488 0.7671096 0.7650845
## 3 1 0.6102180 0.6292476 0.6384870
## 4 1 0.7212125 0.7615327 0.7632913
## 5 1 0.5585574 0.5713338 0.5726343
## 6 1 0.9407706 0.9169831 0.9160665
```

```
# --- Gráfico Probit vs Logit ---
ggplot(dados, aes(x = inlfhat_probit, y = inlfhat_logit)) +
  geom_point(alpha = 0.5) +
  geom_abline(color = "red", linetype = "dashed") +
  labs(title = "Probabilidades Previstas: Probit vs Logit", x = "Probit", y = "Logit") +
  theme_minimal()
```



```
# --- Efeitos marginais ---
summary(LPM)
```

```
##
## Call:
## lm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
##     data = dados)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06854 -0.38575  0.08284  0.35737  0.95335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.769833   0.135007   5.702 1.70e-08 ***
```

```
## nwifeinc      -0.003259   0.001456  -2.239   0.0255 *
## educ          0.039129   0.007364   5.314 1.42e-07 ***
## exper         0.022211   0.002144  10.358 < 2e-16 ***
## age          -0.018508   0.002299  -8.052 3.22e-15 ***
## kidslt6      -0.275306   0.033366  -8.251 7.08e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4298 on 747 degrees of freedom
## Multiple R-squared:  0.253, Adjusted R-squared:  0.248
## F-statistic: 50.61 on 5 and 747 DF,  p-value: < 2.2e-16
```

```
medias <- model.frame(Probit) %>% map_df(mean)
Probit.atmean <- margins(Probit, at = medias)
Probit.ame <- margins(Probit)
Logit.atmean <- margins(Logit, at = medias)
Logit.ame <- margins(Logit)
summary(Probit.atmean)
```

```
##      factor      inlf nwifeinc      educ      exper      age kidslt6      AME      SE
##      age 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.0226 0.0030
##      educ 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377  0.0513 0.0098
##      exper 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377  0.0270 0.0029
##      kidslt6 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.3457 0.0457
##      nwifeinc 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.0044 0.0019
##      z      p      lower      upper
## -7.4251 0.0000 -0.0286 -0.0166
##  5.2427 0.0000  0.0321  0.0705
##  9.1961 0.0000  0.0212  0.0327
## -7.5682 0.0000 -0.4352 -0.2561
## -2.3407 0.0192 -0.0081 -0.0007
```

```
summary(Probit.ame)
```

```
##      factor      AME      SE      z      p      lower      upper
##      age -0.0177 0.0021 -8.3532 0.0000 -0.0219 -0.0136
##      educ  0.0402 0.0073  5.5269 0.0000  0.0260  0.0545
##      exper  0.0211 0.0019 10.9991 0.0000  0.0174  0.0249
##      kidslt6 -0.2709 0.0316 -8.5768 0.0000 -0.3328 -0.2090
##      nwifeinc -0.0035 0.0015 -2.3646 0.0181 -0.0064 -0.0006
```

```
summary(Logit.atmean)
```

```
##      factor      inlf nwifeinc      educ      exper      age kidslt6      AME      SE
##      age 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.0231 0.0033
```

```
##      educ 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377  0.0541 0.0104
##      exper 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377  0.0286 0.0032
##      kidslt6 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.3546 0.0488
##      nwifeinc 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.0048 0.0020
##      z      p      lower      upper
## -7.0705 0.0000 -0.0294 -0.0167
##  5.1986 0.0000  0.0337  0.0745
##  8.9082 0.0000  0.0223  0.0348
## -7.2724 0.0000 -0.4501 -0.2590
## -2.4066 0.0161 -0.0087 -0.0009
```

```
summary(Logit.ame)
```

```
##      factor      AME      SE      z      p      lower      upper
##      age -0.0173 0.0021 -8.0925 0.0000 -0.0214 -0.0131
##      educ  0.0405 0.0073  5.5481 0.0000  0.0262  0.0548
##      exper  0.0214 0.0020 10.9388 0.0000  0.0176  0.0252
##      kidslt6 -0.2655 0.0315 -8.4298 0.0000 -0.3273 -0.2038
##      nwifeinc -0.0036 0.0015 -2.4388 0.0147 -0.0065 -0.0007
```

```
# --- Pseudo R² (McFadden) ---
LLur <- logLik(Probit)
modelo_restrito <- glm(inlf ~ 1, family = binomial(link = "probit"), data = dados)
LL0 <- logLik(modelo_restrito)
pseudo_r2 <- 1 - as.numeric(LLur) / as.numeric(LL0)
cat("Pseudo R² (McFadden):", round(pseudo_r2, 4), "\n")
```

```
## Pseudo R² (McFadden): 0.2104
```

```
# --- Percentual corretamente previsto (Probit e Logit) ---
pred_probit <- as.factor(as.numeric(fitted(Probit) > 0.5))
pred_logit <- as.factor(as.numeric(fitted(Logit) > 0.5))
real <- as.factor(dados$inlf)

# Matriz Probit
confusao_probit <- table(Predito = pred_probit, Real = real)
print(confusao_probit)
```

```
##      Real
## Predito  0  1
##      0 210  82
##      1 115 346
```

```

VP <- ifelse(!is.na(confusao_probit["1", "1"]), confusao_probit["1", "1"], 0)
VN <- ifelse(!is.na(confusao_probit["0", "0"]), confusao_probit["0", "0"], 0)
FP <- ifelse(!is.na(confusao_probit["1", "0"]), confusao_probit["1", "0"], 0)
FN <- ifelse(!is.na(confusao_probit["0", "1"]), confusao_probit["0", "1"], 0)
acc_probit <- (VP + VN) / (VP + VN + FP + FN)
cat("Acurácia Probit:", round(acc_probit * 100, 2), "%\n")

```

```
## Acurácia Probit: 73.84 %
```

```

# Matriz Logit
confusao_logit <- table(Predito = pred_logit, Real = real)
print(confusao_logit)

```

```

##          Real
## Predito   0   1
##          0 211  83
##          1 114 345

```

```

VP <- ifelse(!is.na(confusao_logit["1", "1"]), confusao_logit["1", "1"], 0)
VN <- ifelse(!is.na(confusao_logit["0", "0"]), confusao_logit["0", "0"], 0)
FP <- ifelse(!is.na(confusao_logit["1", "0"]), confusao_logit["1", "0"], 0)
FN <- ifelse(!is.na(confusao_logit["0", "1"]), confusao_logit["0", "1"], 0)
acc_logit <- (VP + VN) / (VP + VN + FP + FN)
cat("Acurácia Logit:", round(acc_logit * 100, 2), "%\n")

```

```
## Acurácia Logit: 73.84 %
```

```

# --- Histograma das probabilidades previstas ---
df_preds <- dados %>% pivot_longer(cols = starts_with("inlfhat_"),
                                   names_to = "modelo", values_to = "prob")
ggplot(df_preds, aes(x = prob, fill = modelo)) +
  geom_histogram(alpha = 0.6, position = "identity", bins = 30) +
  labs(title = "Distribuição das Probabilidades Previstas",
       x = "Probabilidade", y = "Frequência") +
  theme_minimal()

```

Distribuição das Probabilidades Previstas

