Análisis de Regresión Logística, Probit y Bootstrap

Carlos

2025-04-03

Contents

0.1 Carga de Datos

```
auto_time <- c(52.90, 4.10, 4.10, 56.20, 51.80, 0.20, 27.60, 89.90, 41.50,
               95.00, 99.10, 18.50, 82.00, 8.60, 22.50, 51.40, 81.00, 51.00,
               62.20, 95.10, 41.60)
bus_time <- c(4.4, 28.5, 86.9, 31.6, 20.2, 91.2, 79.7, 2.2, 24.5,
              43.5, 8.4, 84, 38, 1.6, 74.1, 83.8, 19.2, 85, 90.1, 22.2, 91.5)
y <- c(0, 0, 1, 0, 0, 1, 1, 0, 0,
       0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1)
x <- auto_time - bus_time
```

0.2 Modelo de Regresión Logística

```
modelo <- glm(y ~ x, family = binomial)
summary(modelo)
```

```
##
## Call:
## glm(formula = y ~ x, family = binomial)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.23758    0.75048  -0.317   0.7516
## x           -0.05311    0.02064  -2.573   0.0101 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 29.065  on 20  degrees of freedom
## Residual deviance: 12.332  on 19  degrees of freedom
## AIC: 16.332
##
## Number of Fisher Scoring iterations: 6
```

```
plot(x, y, pch = 19, col = ifelse(y == 1, "blue", "red"),
     main = "Logistic Regression Fit", xlab = "x = Auto Time - Bus Time", ylab = "P(car)")
curve(predict(modelo, data.frame(x = sort(x)), type = "response"),
      add = TRUE, col = "darkgreen", lwd = 2)
```

```r
legend("bottomright", legend = c("Observado 1", "Observado 0", "Curva logística"),
       col = c("blue", "red", "darkgreen"), pch = c(19, 19, NA), lty = c(NA, NA, 1))
```

regresion_logit_probit_con_apendices_limpios_files/figure-latex/unnamed-chunk-2-1.pdf

0.3 Ajuste por Mínimos Cuadrados No Lineales (NLS)

```r
b0_init <- coef(modelo)[1]
b1_init <- coef(modelo)[2]
modelo_nls <- nls(
  y ~ 1 / (1 + exp(-(b0 + b1 * x))),
  start = list(b0 = b0_init , b1 = b1_init)
)
summary(modelo_nls)
```

```
##
## Formula: y ~ 1/(1 + exp(-(b0 + b1 * x)))
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## b0   -188.344 744010.283       0        1
## b1     -7.197  27886.344       0        1
##
## Residual standard error: 0.2294 on 19 degrees of freedom
##
## Number of iterations to convergence: 16
## Achieved convergence tolerance: 4.887e-06
```

```r
plot(x, y, pch = 19, col = ifelse(y == 1, "blue", "red"),
     main = "Comparación de Métodos de Ajuste", xlab = "x = Auto Time - Bus Time", ylab = "P(car)")
x_sorted <- sort(x)
lines(x_sorted, predict(modelo, type = "response")[order(x)], col = "darkgreen", lwd = 2)

b0_nls <- coef(modelo_nls)["b0"]
b1_nls <- coef(modelo_nls)["b1"]
p_nls <- 1 / (1 + exp(-(b0_nls + b1_nls * x_sorted)))
lines(x_sorted, p_nls, col = "purple", lwd = 2, lty = 2)

legend("bottomleft", legend = c("GLM (Verosimilitud)", "NLS (Mínimos Cuadrados)"),
       col = c("darkgreen", "purple"), lty = c(1, 2), lwd = 2, cex = 0.5)
```

regresion_logit_probit_con_apendices_limpios_files/figure-latex/unnamed-chunk-3-1.pdf

0.4 Modelo Probit

```r
modelo_probit <- glm(y ~ x, family = binomial(link = "probit"))
summary(modelo_probit)
```

```
##
## Call:
## glm(formula = y ~ x, family = binomial(link = "probit"))
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.06443    0.40068  -0.161  0.87224
## x           -0.03000    0.01029  -2.915  0.00355 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 29.065  on 20  degrees of freedom
## Residual deviance: 12.330  on 19  degrees of freedom
## AIC: 16.33
##
## Number of Fisher Scoring iterations: 7
```

```r
plot(x, y, pch = 19, col = ifelse(y == 1, "blue", "red"),
     main = "Ajuste de Regresión Probit", xlab = "x = Auto Time - Bus Time", ylab = "P(car)")

x_seq <- seq(min(x), max(x), length.out = 100)
lines(x_seq, predict(modelo_probit, newdata = data.frame(x = x_seq), type = "response"),
      col = "orange", lwd = 2)

legend("bottomleft", legend = c("Observado 1", "Observado 0", "Curva Probit"),
       col = c("blue", "red", "orange"), pch = c(19, 19, NA), lty = c(NA, NA, 1), lwd = 2, cex = 0.8)
```

regresion_logit_probit_con_apendices_limpios_files/figure-latex/unnamed-chunk-4-1.pdf

0.5 Comparación: Logit vs Probit

```r
modelo_logit  <- glm(y ~ x, family = binomial(link = "logit"))

plot(x, y, pch = 19, col = ifelse(y == 1, "blue", "red"),
     main = "Comparación: Logit vs Probit", xlab = "x = Auto Time - Bus Time", ylab = "P(car)", ylim = 

lines(x_seq, predict(modelo_logit, newdata = data.frame(x = x_seq), type = "response"),
      col = "darkgreen", lwd = 2)
lines(x_seq, predict(modelo_probit, newdata = data.frame(x = x_seq), type = "response"),
      col = "orange", lwd = 2, lty = 2)

legend("bottomleft", legend = c("Observado 1", "Observado 0", "Logit", "Probit"),
```

```
       col = c("blue", "red", "darkgreen", "orange"),
       pch = c(19, 19, NA, NA), lty = c(NA, NA, 1, 2), lwd = 2, cex = 0.5)
```

regresion_logit_probit_con_apendices_limpios_files/figure-latex/unnamed-chunk-5-1.pdf

0.6 Bootstrap + Modelos

```
datos <- data.frame(auto_time, bus_time, x = auto_time - bus_time, y)
set.seed(123)
bootstrap_sample <- datos[sample(1:nrow(datos), size = 100, replace = TRUE), ]
x_btsp <- bootstrap_sample$x
y_btsp <- bootstrap_sample$y

modelo_btstrap <- glm(y_btsp ~ x_btsp, family = binomial)
summary(modelo_btstrap)
```

```
##
## Call:
## glm(formula = y_btsp ~ x_btsp, family = binomial)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.52119    0.41264  -1.263    0.207
## x_btsp      -0.06812    0.01300  -5.240 1.61e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 138.589  on 99  degrees of freedom
## Residual deviance:  45.803  on 98  degrees of freedom
## AIC: 49.803
##
## Number of Fisher Scoring iterations: 6
```

```
plot(x_btsp, y_btsp, pch = 19, col = ifelse(y_btsp == 1, "blue", "red"),
     main = "Logistic Regression Fit (Bootstrap)", xlab = "x = Auto Time - Bus Time", ylab = "P(car)")

x_seq_btsp <- seq(min(x_btsp), max(x_btsp), length.out = 100)
lines(x_seq_btsp, predict(modelo_btstrap, newdata = data.frame(x_btsp = x_seq_btsp), type = "response")
      col = "darkgreen", lwd = 2)

legend("bottomleft", legend = c("Observado 1", "Observado 0", "Curva logística"),
       col = c("blue", "red", "darkgreen"), pch = c(19, 19, NA), lty = c(NA, NA, 1), lwd = 2, cex = 0.8
```

regresion_logit_probit_con_apendices_limpios_files/figure-latex/unnamed-chunk-6-1.pdf

0.7 Conclusión General

Ambos modelos, logit y probit, son útiles para clasificación binaria. El modelo logit tiene una curva más empinada en el centro, mientras que el probit es más suave en los extremos. La técnica de bootstrap refuerza la validez de los estimadores obtenidos.

1 Apéndices: Documentos TeX relevantes

1.1 Derivadas Regresion

1.2 Maalouf Logistic Regression Deducciones

1.3 Maalouf Logistic Regression Full

1.4 Resumen Ai Ml Con Referencias Compila

1.5 Resumen Ai Ml Con Referencias Final

1.6 Resumen Ai Ml Con Referencias Tikz

1.7 Resumen Ai Ml Con Referencias

1.8 Resumen Logistic Regression Completo

1.9 Resumen Logistic Regression Formal

2 Apéndices: Documentos TeX (versiones limpias)

2.1 Derivadas Regresion

2.2 Maalouf Logistic Regression Deducciones

2.3 Maalouf Logistic Regression Full

2.4 Resumen Ai Ml Con Referencias Compila

2.5 Resumen Ai Ml Con Referencias Final

2.6 Resumen Ai Ml Con Referencias Tikz

2.7 Resumen Ai Ml Con Referencias

2.8 Resumen Logistic Regression Completo

2.9 Resumen Logistic Regression Formal