

Análisis Principal

Manuela Lopez Cambron, 1673688

2026-01-27

```
library(pROC)
```

```
## Warning: el paquet 'pROC' es va construir amb la versió d'R 4.5.2
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
```

```
## S'està adjuntant el paquet: 'pROC'
```

```
## Els següents objectes estan emmascarats des de 'package:stats':
```

```
##
```

```
##      cov, smooth, var
```

```
library(dplyr)
```

```
##
```

```
## S'està adjuntant el paquet: 'dplyr'
```

```
## Els següents objectes estan emmascarats des de 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## Els següents objectes estan emmascarats des de 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(car)
```

```
## S'està carregant el paquet requerit: carData
```

```
##
```

```
## S'està adjuntant el paquet: 'car'
```

```
## L'objecte següent està emmascarat per 'package:dplyr':
```

```
##
```

```
##      recode
```

```
library(marginaleffects)
```

```
## Warning: el paquet 'marginaleffects' es va construir amb la versió d'R 4.5.2
```

```
library(broom)
library(margins)
```

```
## Warning: el paquet 'margins' es va construir amb la versió d'R 4.5.2
```

```
library(ggplot2)
library(forcats)
library(modelsummary)
```

```
## Warning: el paquet 'modelsummary' es va construir amb la versió d'R 4.5.2
```

```
df <- read.csv("prado_variables.csv", stringsAsFactors = TRUE)

df$sop_montaje <- as.factor(df$sop_montaje)
df$serie <- as.factor(df$serie)

str(df)
```

```
## 'data.frame': 7035 obs. of 12 variables:
## $ exito : int 0 0 0 0 0 0 0 0 0 1 ...
## $ area : num 81423 81423 52600 37000 44955 ...
## $ tam_cat : Factor w/ 3 levels "grande","mediano",...: 1 1 1 1 1 1 1 1 2 1 ...
## $ orientacion: Factor w/ 3 levels "cuadrado","horizontal",...: 2 2 2 3 2 2 1 2 3 2 ...
## $ soporte_grp: Factor w/ 5 levels "Lienzo","Metal",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ sop_montaje: Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ tecnica : Factor w/ 3 levels "mixta","oleo",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ tipo_autor : Factor w/ 4 levels "anonimo","hombre",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ serie : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ fecha_est : int 1150 1150 1150 1150 1150 1150 1150 1150 1150 1150 ...
## $ fecha_ancho: int 99 99 99 99 99 99 99 99 99 99 ...
## $ tema : Factor w/ 10 levels "bodegon_floral",...: 8 8 8 8 8 8 8 8 8 8 ...
```

ANÁLISIS PRINCIPAL

Primeramente generamos la submuestra de la población aplicando la regla estructural de 'cuadrado'.

```
# 0) Regla estructural: excluir "cuadrado"
df <- subset(df, orientacion != "cuadrado")
df$orientacion <- droplevels(df$orientacion)
dim(df)
```

```
## [1] 7002 12
```

Continuamos con un total de 7002 pinturas

Fijamos también los niveles de referencia para las variables factor. El nivel más frecuente como referencia para todos los factores a excepción de “tam_cat” donde por interpretabilidad se define la referencia en ‘pequeno’ y las variables binarias “sop_montaje” y “serie” donde se fija el valor “no”

```
# 1) Niveles de los factores binarios
df$sop_montaje <- factor(df$sop_montaje, levels = c(0, 1), labels = c("no", "si"))
df$serie <- factor(df$serie, levels = c(0, 1), labels = c("no", "si"))

# 2) tam_cat nominal con referencia "pequeno"
df$tam_cat <- factor(df$tam_cat, levels = c("pequeno", "mediano", "grande"))
df$tam_cat <- relevel(df$tam_cat, ref = "pequeno")

# 3) Binarios con referencia "no"
df$sop_montaje <- relevel(df$sop_montaje, ref = "no")
df$serie <- relevel(df$serie, ref = "no")

# 4) Referencias = nivel más frecuente (para el resto de factores)
ref_orientacion <- names(sort(table(df$orientacion), decreasing = TRUE))[1]
df$orientacion <- relevel(df$orientacion, ref = ref_orientacion)

ref_soporte <- names(sort(table(df$soporte_grp), decreasing = TRUE))[1]
df$soporte_grp <- relevel(df$soporte_grp, ref = ref_soporte)

ref_tecnica <- names(sort(table(df$tecnica), decreasing = TRUE))[1]
df$tecnica <- relevel(df$tecnica, ref = ref_tecnica)

ref_autor <- names(sort(table(df$tipo_autor), decreasing = TRUE))[1]
df$tipo_autor <- relevel(df$tipo_autor, ref = ref_autor)

ref_tema <- names(sort(table(df$tema), decreasing = TRUE))[1]
df$tema <- relevel(df$tema, ref = ref_tema)

str(df)

## 'data.frame': 7002 obs. of 12 variables:
## $ exito : int 0 0 0 0 0 0 0 0 1 0 ...
## $ area : num 81423 81423 52600 37000 44955 ...
## $ tam_cat : Factor w/ 3 levels "pequeno","mediano",...: 3 3 3 3 3 3 3 2 3 3 ...
## $ orientacion: Factor w/ 2 levels "vertical","horizontal": 2 2 2 1 2 2 2 1 2 2 ...
## $ soporte_grp: Factor w/ 5 levels "Lienzo","Metal",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ sop_montaje: Factor w/ 2 levels "no","si": 2 2 2 2 2 2 2 2 2 2 ...
## $ tecnica : Factor w/ 3 levels "oleo","mixta",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ tipo_autor : Factor w/ 4 levels "hombre","anonimo",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ serie : Factor w/ 2 levels "no","si": 2 2 2 2 2 2 2 2 2 2 ...
## $ fecha_est : int 1150 1150 1150 1150 1150 1150 1150 1150 1150 1150 ...
## $ fecha_ancho: int 99 99 99 99 99 99 99 99 99 99 ...
## $ tema : Factor w/ 10 levels "religioso","bodegon_floral",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Efectos principales

Para determinar los efectos principales del modelo se seguirá la estructura por bloques definida en la sección de metodología

Modelo nulo

Como punto de partida se ajustó un modelo nulo (solos intercepto), sin covariables. Este modelo proporcionará la referencia sobre el cual iremos cuantificando el aporte de los bloques que se añadirán sucesivamente. En un modelo binario como el nuestro, el intercepto del modelo nulo estima la probabilidad media de éxito en la muestra. Recordemos que estaremos trabajando en todo momento con la sub-muestra no-cuadrado.

```
# Modelo nulo (solo intercepto)
m0 <- glm(exito ~ 1, data = df, family = binomial(link = "logit"))

# Resumen del modelo
cat("\nResumen del modelo:\n")

##
## Resumen del modelo:

summary(m0)

##
## Call:
## glm(formula = exito ~ 1, family = binomial(link = "logit"), data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.98064    0.03661  -54.1    <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: 5173.2  on 7001  degrees of freedom
## Residual deviance: 5173.2  on 7001  degrees of freedom
## AIC: 5175.2
##
## Number of Fisher Scoring iterations: 4
```

```
# Probabilidad media estimada de éxito (a partir del intercepto)
cat("\nProbabilidad estimada de éxito:\n")
```

```
##
## Probabilidad estimada de éxito:
```

```
p0 <- plogis(coef(m0)[1])
p0
```

```
## (Intercept)
##    0.1212511
```

Bloque 1: Datación y control de incertidumbre

Se incorpora la variable “fecha_est” y su covariable de incertidumbre “fecha_ancho”, comparando especificación lineal vs flexible (spline) en ambas variables.

Comenzaremos decidiendo la especificación del control “fecha_ancho”.

```

library(splines)

# Transformación loglp(fecha_ancho)
df$log_ancho <- loglp(df$fecha_ancho)

# Modelo A: control lineal en log_ancho
m1_lin <- glm(exito ~ ns(fecha_est, 3) + log_ancho,
              data = df, family = binomial(link = "logit"))

# Modelo B: control flexible (spline) en log_ancho
m1_spl <- glm(exito ~ ns(fecha_est, 3) + ns(log_ancho, 3),
              data = df, family = binomial(link = "logit"))

## Warning in ns(log_ancho, 3): els nusos «interior» que concorden amb els nusos
## del límit els empenyen cap a dins

# Resúmenes (por si aparecen warnings / coeficientes raros / SE enormes)
cat("\nResumen del modelo m1_lin (control lineal):\n")

##
## Resumen del modelo m1_lin (control lineal):

summary(m1_lin)

##
## Call:
## glm(formula = exito ~ ns(fecha_est, 3) + log_ancho, family = binomial(link = "logit"),
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.196862    0.566571  -3.877 0.000106 ***
## ns(fecha_est, 3)1  0.281381    0.368057   0.765 0.444567
## ns(fecha_est, 3)2  0.558190    1.192596   0.468 0.639751
## ns(fecha_est, 3)3  1.153682    0.297870   3.873 0.000107 ***
## log_ancho       0.005883    0.022782   0.258 0.796244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 5120.5  on 6997  degrees of freedom
## AIC: 5130.5
##
## Number of Fisher Scoring iterations: 4

cat("\nResumen del modelo m1_spl (control flexible spline):\n")

##
## Resumen del modelo m1_spl (control flexible spline):

```

```
summary(m1_spl)
```

```
##
## Call:
## glm(formula = exito ~ ns(fecha_est, 3) + ns(log_ancho, 3), family = binomial(link = "logit"),
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.235466   0.571676  -3.910 9.22e-05 ***
## ns(fecha_est, 3)1  0.276625   0.369477   0.749 0.454041
## ns(fecha_est, 3)2  0.730706   1.206027   0.606 0.544596
## ns(fecha_est, 3)3  1.135879   0.300691   3.778 0.000158 ***
## ns(log_ancho, 3)1 -0.283268   0.340571  -0.832 0.405553
## ns(log_ancho, 3)2  0.009094   0.236265   0.038 0.969298
## ns(log_ancho, 3)3  0.431797   0.283258   1.524 0.127410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 5117.3  on 6995  degrees of freedom
## AIC: 5131.3
##
## Number of Fisher Scoring iterations: 4
```

Comprobamos la aportación del bloque comparando ambas especificaciones con el modelo nulo:

```
# Comparación --> aporte del Bloque 1

# Información
cat("\nAporte de información del Bloque 1 (control lineal):\n")
```

```
##
## Aporte de información del Bloque 1 (control lineal):
```

```
anova(m0, m1_lin, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ 1
## Model 2: exito ~ ns(fecha_est, 3) + log_ancho
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1      7001      5173.2
## 2      6997      5120.5  4    52.682 9.933e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\nAporte de información del Bloque 1 (control flexible spline):\n")
```

```
##  
## Aporte de información del Bloque 1 (control flexible spline):
```

```
anova(m0, m1_spl, test="Chisq")
```

```
## Analysis of Deviance Table  
##  
## Model 1: exito ~ 1  
## Model 2: exito ~ ns(fecha_est, 3) + ns(log_ancho, 3)  
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)  
## 1      7001      5173.2  
## 2      6995      5117.3  6   55.875 3.085e-10 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Complejidad
```

```
cat("\nAporte de complejidad:\n")
```

```
##  
## Aporte de complejidad:
```

```
AIC(m0, m1_lin, m1_spl)
```

```
##      df      AIC  
## m0      1 5175.221  
## m1_lin  5 5130.540  
## m1_spl  7 5131.346
```

```
BIC(m0, m1_lin, m1_spl)
```

```
##      df      BIC  
## m0      1 5182.075  
## m1_lin  5 5164.809  
## m1_spl  7 5179.324
```

Comparamos también las dos formas funcionales del control mediante modelos anidados:

```
# Comparación --> formas funcionales
```

```
cat("\nAporte de información entre opciones (m1_lin vs m1_spl)\n")
```

```
##  
## Aporte de información entre opciones (m1_lin vs m1_spl)
```

```
anova(m1_lin, m1_spl, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ ns(fecha_est, 3) + log_ancho
## Model 2: exito ~ ns(fecha_est, 3) + ns(log_ancho, 3)
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6997      5120.5
## 2      6995      5117.3  2    3.1933    0.2026
```

El aporte del bloque 1 es claramente significativo. Ambas especificaciones mejoraron significativamente el modelo nulo (LRT $p = 9.933e - 11$, $p = 3.085e - 10$). Sin embargo, la especificación flexible para el control “fecha_ancho” no proporcionó mejora adicional frente a la lineal (LRT $p = 0.2026$) y presentó peor ajuste penalizado por complejidad (AIC y BIC mayores). Por esta razón, se adoptó para los modelos posteriores la especificación lineal $\log(1 + fecha_ancho)$ como ajuste definitivo del control de incertidumbre en la datación.

Finalmente comprobaremos las formas funcionales de “fecha_est” para ver si la especificación flexible de esta variable es realmente necesaria para nuestro modelo o no.

```
m1_lin_fecha <- glm(exito ~ fecha_est + log_ancho,
                    data = df, family = binomial(link = "logit"))
```

```
# Comparación
# Información
cat("\nAporte de la opción fecha_est felxible:\n")
```

```
##
## Aporte de la opción fecha_est felxible:
```

```
anova(m1_lin_fecha, m1_lin, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho
## Model 2: exito ~ ns(fecha_est, 3) + log_ancho
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6999      5126.6
## 2      6997      5120.5  2    6.0146  0.04943 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Complejidad
cat("\nComplejidad de la opción fecha_est flexible:\n")
```

```
##
## Complejidad de la opción fecha_est flexible:
```

```
AIC(m1_lin,m1_lin_fecha)
```

```
##           df      AIC
## m1_lin      5 5130.540
## m1_lin_fecha 3 5132.554
```



```
BIC(m1_lin,m1_lin_fecha)
```

```
##           df      BIC
## m1_lin      5 5164.809
## m1_lin_fecha 3 5153.116
```

Vemos que la especificación flexible es preferible frente a la lineal en términos de LRT ($p = 0.049$) como AIC ($\Delta AIC \approx -2$). Sin embargo la mejora de AIC es débil frente a una fuerte penalización en BIC ($\Delta AIC \approx -2$ vs $\Delta BIC \approx 12$). Teniendo en cuenta nuestro objetivo descriptivo, y no predictivo, decidimos seleccionar la versión lineal por facilidad interpretativa. Por otro lado, también seleccionamos esta opción por ser más conservadora, ya que añadimos menos parámetros al modelo (2 parámetros menos), y es preferible dado que la mayoría de las futuras covariables son factores y nuestra variable respuesta está fuertemente desbalanceada. Sin embargo, dejamos constancia del hecho que la especificación flexible con spline por fecha_est podría ser considerada para análisis más exhaustivos.

Modelo resultante después de añadir Bloque 1:

```
m1 <- m1_lin_fecha
```

Bloque 2: Morfología

Se incorporan las variables morfológicas de tamaño i formato, “area” y “orientación” respectivamente. Se evaluará si aportan información adicional, una vez controlado el efecto de datación (bloque 1). La variable “area” se introducirá mediante una transformación logarítmica $\log(area)$, pudiendo llegar a ser tratada de manera flexible si fuera necesario. También se proporcionará un modelo secundario sustituyendo “area” por “tam_cat” (categorización de área), y se compararán. Se decidirá el mejor modelo siguiendo las indicaciones de la sección de metodología, que resumidamente dictan lo siguiente:

- si “area” presenta problemas o no mejora sustancialmente más que “tam_cat”, permitimos ajuste flexible (spline) y si no funciona elegimos “tam_cat”
- si “area” mejora sustancialmente más que “tam_cat”, elegimos “area”
- si “area” y “tam_cat” proporcionan resultados cualitativamente iguales, elegimos “area” pero manteniendo el modelo secundario “tam_cat” para interpretaciones claras.

```
# Transformación log(area)
df$log_area <- log(df$area)

# BLOQUE 2 (principal candidato): log(area) + orientacion
m2_area <- update(m1, . ~ . + log_area + orientacion)

# BLOQUE 2 (secundario): tam_cat + orientacion
m2_tamcat <- update(m1, . ~ . + tam_cat + orientacion)

# Resúmenes (por si aparecen warnings / coeficientes raros / SE enormes)
cat("\nResumen del modelo m2_area (+ log_area + orientacion):\n")

##
## Resumen del modelo m2_area (+ log_area + orientacion):
```

```
summary(m2_area)
```

```
##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + log_area + orientacion,
##      family = binomial(link = "logit"), data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -7.4602560   0.6658251  -11.205 < 2e-16 ***
## fecha_est       0.0020017   0.0003269    6.124 9.15e-10 ***
## log_ancho       0.0059186   0.0227367    0.260  0.795
## log_area        0.1989546   0.0278371    7.147 8.86e-13 ***
## orientacionhorizontal 0.3267555   0.0748131    4.368 1.26e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 5048.6  on 6997  degrees of freedom
## AIC: 5058.6
##
## Number of Fisher Scoring iterations: 5
```

```
cat("\nResumen del modelo m2_tamcat (+ tam_cat + orientacion):\n")
```

```
##
## Resumen del modelo m2_tamcat (+ tam_cat + orientacion):
```

```
summary(m2_tamcat)
```

```
##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + tam_cat + orientacion,
##      family = binomial(link = "logit"), data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -6.0159849   0.6000728  -10.025 < 2e-16 ***
## fecha_est       0.0020633   0.0003304    6.244 4.26e-10 ***
## log_ancho       0.0101612   0.0228402    0.445  0.656
## tam_catmediano  -0.0594074   0.1034091   -0.574  0.566
## tam_catgrande   0.7738667   0.0898896    8.609 < 2e-16 ***
## orientacionhorizontal 0.2962533   0.0751562    3.942 8.09e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4988.1  on 6996  degrees of freedom
```

```

## AIC: 5000.1
##
## Number of Fisher Scoring iterations: 5

# Comparación --> aporte del Bloque 2

# Información
cat("\nAporte de información del Bloque 2 (tamaño continuo):\n")

##
## Aporte de información del Bloque 2 (tamaño continuo):

anova(m1, m2_area, test="Chisq")

## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho
## Model 2: exito ~ fecha_est + log_ancho + log_area + orientacion
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         6999      5126.6
## 2         6997      5048.6  2    77.907 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\nAporte de información del Bloque 2 (tamaño categórico):\n")

##
## Aporte de información del Bloque 2 (tamaño categórico):

anova(m1, m2_tamcat, test="Chisq")

## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         6999      5126.6
## 2         6996      4988.1  3    138.45 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Complejidad
cat("\nAporte de complejidad:\n")

##
## Aporte de complejidad:

AIC(m1, m2_area, m2_tamcat)

```

```
##          df      AIC
## m1          3 5132.554
## m2_area     5 5058.647
## m2_tamcat   6 5000.103
```

```
BIC(m1, m2_area, m2_tamcat)
```

```
##          df      BIC
## m1          3 5153.116
## m2_area     5 5092.917
## m2_tamcat   6 5041.227
```

Ambas opciones del tamaño ($\log(\text{area})$ continua vs. tam_cat categórica) aportan información adicional tras controlar la datación (LRT $p < 2.2e - 16$ en ambos casos). Sin embargo, la especificación categórica presenta mejor ajuste-complejidad con AIC y BIC sustancialmente menores (AIC: 5058 vs 5000; BIC: 5092 vs 5042).

Antes de seleccionar “ tam_cat ” debemos tener en cuenta que podría haber una relación no lineal que actualmente $\log(\text{area})$ no está pudiendo capturar. Por esta razón y dado que “ area ” ha demostrado mejorar el ajuste, frente al modelo anterior ($m1$), flexibilizaremos su especificación (spline) y entonces volveremos a comparar con “ tam_cat ”, para asegurar una decisión justa y cerrada.

```
# Version flexible de log(area)
m2_area_spl <- update(m1, . ~ . + splines::ns(log_area, 3) + orientacion)

# Resúmenes (por si aparecen warnings / coeficientes raros / SE enormes)
summary(m2_area_spl)
```

```
##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + splines::ns(log_area,
##      3) + orientacion, family = binomial(link = "logit"), data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.6151418   0.7103089  -6.497 8.17e-11 ***
## fecha_est         0.0021365   0.0003311   6.453 1.10e-10 ***
## log_ancho         0.0092053   0.0228401   0.403  0.6869
## splines::ns(log_area, 3)1  1.0901352   0.2248946   4.847 1.25e-06 ***
## splines::ns(log_area, 3)2 -4.0254449   0.9980720  -4.033 5.50e-05 ***
## splines::ns(log_area, 3)3 -2.0573173   0.8145897  -2.526  0.0116 *
## orientacionhorizontal    0.3319577   0.0759994   4.368 1.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 5007.8  on 6995  degrees of freedom
## AIC: 5021.8
##
## Number of Fisher Scoring iterations: 5
```

```

# Comparaciones
cat("\nAporte de la opción log(area) flexible")

##
## Aporte de la opción log(area) flexible

anova(m1, m2_area_spl, test="Chisq")

## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho
## Model 2: exito ~ fecha_est + log_ancho + splines::ns(log_area, 3) + orientacion
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1         6999      5126.6
## 2         6995      5007.8   4   118.74 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\nComplejidad de la opción log(area) flexible")

##
## Complejidad de la opción log(area) flexible

AIC(m2_area, m2_area_spl, m2_tamcat)

##           df      AIC
## m2_area      5 5058.647
## m2_area_spl  7 5021.812
## m2_tamcat     6 5000.103

BIC(m2_area, m2_area_spl, m2_tamcat)

##           df      BIC
## m2_area      5 5092.917
## m2_area_spl  7 5069.789
## m2_tamcat     6 5041.227

```

El modelo con $\log(\text{area})$ lineal mejoró el ajuste, y al permitir no linealidad, el ajuste mejoró frente al lineal (AIC: 5021 vs 5058). Sin embargo, la especificación categórica “tam_cat” presentó el mejor compromiso ajuste-complejidad, con AIC y BIC claramente inferiores (AIC: 5000 vs 5021); BIC: 5041 vs 5069), superando también a la versión flexible del tamaño continuo. Por ello, se seleccionó “tam_cat” como representación principal del tamaño para los modelos posteriores.

También se considera manejar en un modelo alternativo la opción de especificación continua del tamaño en versión lineal ($\log(\text{area})$ sin spline). El único objetivo de mantener esta opción como alternativa, en vez de la flexible, es el de aportar un modelo más parsimonioso. Aunque su versión flexible mostró un mejor ajuste, no cumple con nuestro objetivo. Vemos que $\log(\text{area})$ añade 1 parámetro (uno menos que “tam_cat”) mientras que $\text{ns}(\log(\text{area}), 3)$ añade 3 parámetros (uno más que “tam_cat”).

Modelo resultante después de añadir Bloque 2:

```
m2 <- m2_tamcat
```

Opción alternativa: especificación continua del tamaño en versión lineal ($\log(\text{area})$)

Bloque 3: Material y técnica

Se incorporan las variables “soporte” y “técnica” para evaluar si aportan información adicional en conjunto, una vez controlados los efectos de datación (Bloque 1) y morfología (Bloque 2). Se evaluará adicionalmente la inclusión de “sop_montaje” como extensión del bloque 3, comparando el modelo con y sin dicha covariable.

```
# BLOQUE 3 (versión base): soporte + técnica
m3_base <- update(m2, . ~ . + soporte_grp + tecnica)

# Resúmenes (por si aparecen warnings / coeficientes raros / SE enormes)
cat("\nResumen del modelo m3_base (soporte_grp + tecnica):\n")
```

```
##
## Resumen del modelo m3_base (soporte_grp + tecnica):
```

```
summary(m3_base)
```

```
##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + tam_cat + orientacion +
##      soporte_grp + tecnica, family = binomial(link = "logit"),
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -7.0808549   0.6673046  -10.611 < 2e-16 ***
## fecha_est        0.0025950   0.0003592    7.224 5.05e-13 ***
## log_ancho        0.0194664   0.0230144    0.846  0.3976
## tam_catmediano  -0.0004402   0.1083753   -0.004  0.9968
## tam_catgrande    0.8764280   0.1014370    8.640 < 2e-16 ***
## orientacionhorizontal  0.3073077   0.0757811    4.055 5.01e-05 ***
## soporte_grpMetal  -0.4347888   0.3736719   -1.164  0.2446
## soporte_grpMural   0.2739620   0.5241350    0.523  0.6012
## soporte_grpOtros  -0.6011812   0.3581133   -1.679  0.0932 .
## soporte_grpTabla/Panel  0.4854756   0.1241065    3.912 9.16e-05 ***
## tecnicamixta       0.3096232   0.3225271    0.960  0.3371
## tecnicaotras      -0.1599554   0.3953591   -0.405  0.6858
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4963.0  on 6990  degrees of freedom
## AIC: 4987
##
## Number of Fisher Scoring iterations: 5
```

```
# Comparación--> aporte del Bloque 3
```

```
# Información
```

```
cat("\nAporte de información del Bloque 3:\n")
```

```
##
```

```
## Aporte de información del Bloque 3:
```

```
anova(m2, m3_base, test = "Chisq")
```

```
## Analysis of Deviance Table
```

```
##
```

```
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion
```

```
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
```

```
##      tecnica
```

```
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
```

```
## 1      6996      4988.1
```

```
## 2      6990      4963.0  6   25.117 0.0003249 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Complejidad
```

```
cat("\nAporte de complejidad:\n")
```

```
##
```

```
## Aporte de complejidad:
```

```
AIC(m2, m3_base)
```

```
##           df          AIC
```

```
## m2           6 5000.103
```

```
## m3_base     12 4986.987
```

```
BIC(m2, m3_base)
```

```
##           df          BIC
```

```
## m2           6 5041.227
```

```
## m3_base     12 5069.234
```

La inclusión conjunta de “soporte_grp” y “tecnica” produjo una mejora significativa respecto al modelo con datación y morfología (LRT $p = 0.0003$) también mejoró el ajuste penalizado por AIC ($\Delta AIC \approx -13$), aunque el BIC aumentó ($\Delta BIC \approx +28$) (mayor penalización por el gran número de parámetros añadidos, 6 añadidos). Debido al objetivo descriptivo de nuestro estudio, se decide mantener el bloque por su relevancia teórica y por la evidencia global de aporte de información.

Extenderemos este bloque añadiendo ahora la variable “sop_montaje”, de menos interés conceptual pero con posibles implicaciones en el modelo a nivel de control.

```

# BLOQUE 3 (extendido): + sop_montaje
m3_montaje <- update(m3_base, . ~ . + sop_montaje)

# Resúmenes (por si aparecen warnings / coeficientes raros / SE enormes)
cat("\nResumen del modelo m3_montaje (+ sop_montaje):\n")

##
## Resumen del modelo m3_montaje (+ sop_montaje):

summary(m3_montaje)

##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + tam_cat + orientacion +
##      soporte_grp + tecnica + sop_montaje, family = binomial(link = "logit"),
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.1200183   0.6700046 -10.627 < 2e-16 ***
## fecha_est         0.0025881   0.0003602   7.184 6.76e-13 ***
## log_ancho         0.0208326   0.0230627   0.903 0.366364
## tam_catmediano    0.0505267   0.1104297   0.458 0.647279
## tam_catgrande     0.9323677   0.1039707   8.968 < 2e-16 ***
## orientacionhorizontal 0.2891818 0.0762205   3.794 0.000148 ***
## soporte_grpMetal  -0.3862326   0.3741296  -1.032 0.301908
## soporte_grpMural  -0.2720792   0.5551693  -0.490 0.624075
## soporte_grpOtros  -1.0478701   0.3926692  -2.669 0.007617 **
## soporte_grpTabla/Panel 0.5158106 0.1246809   4.137 3.52e-05 ***
## tecnicamixta       0.1796075   0.3279779   0.548 0.583952
## tecnicaotras      -0.2057277   0.3960271  -0.519 0.603427
## sop_montajesi      0.6621291   0.2273603   2.912 0.003588 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4955.2  on 6989  degrees of freedom
## AIC: 4981.2
##
## Number of Fisher Scoring iterations: 5

# Comparación--> aporte de sop_montaje

# Información
cat("\nAporte de informacion de sop_montaje (m3_montaje vs m3_base):\n")

##
## Aporte de informacion de sop_montaje (m3_montaje vs m3_base):

```



```
anova(m3_base, m3_montaje, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         6990      4963.0
## 2         6989      4955.2  1    7.7395 0.005403 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\nComplejidad de sop_montaje:\n")
```

```
##
## Complejidad de sop_montaje:
```

```
AIC(m2, m3_base, m3_montaje)
```

```
##           df      AIC
## m2           6 5000.103
## m3_base      12 4986.987
## m3_montaje   13 4981.247
```

```
BIC(m2, m3_base, m3_montaje)
```

```
##           df      BIC
## m2           6 5041.227
## m3_base      12 5069.234
## m3_montaje   13 5070.348
```

Se observó una mejora significativa del ajuste respecto al modelo sin esta covariable (LRT $p = 0.005$). El AIC disminuyó ($\Delta AIC \approx -5$), indicando una mejora del ajuste penalizado por complejidad, aunque el BIC aumentó ligeramente ($\Delta BIC \approx +1$), el incremento fue pequeño. Por tanto, se retuvo “sop_montaje” en el modelo para los bloques posteriores.

Sin embargo, debemos recordar que aunque el incremento en BIC para la inclusión de “sop_montaje” fue pequeño, la inclusión de Bloque 3 ya produjo aumento fuerte en BIC por lo que la complejidad añadida de todo el bloque más el extra sí representa un valor sustancial ($\Delta_{total} BIC \approx +29$). Por esta razón se ha decidido contemplar como modelo alternativo la opción sin este bloque.

Modelo resultante después de añadir Bloque 3:

```
m3 <- m3_montaje
```

Opción alternativa: modelo sin Bloque 3 (y sin “sop_montaje”)

Bloque 4: Iconografía

Se incorpora la variable “tema” para evaluar si la iconografía de la pintura aporta información adicional sobre la probabilidad de éxito, una vez controlados los efectos de datacion (Bloque 1), morfología (Bloque 2) y material/técnica (Bloque 3).

```

# BLOQUE 4: + tema
m4 <- update(m3, . ~ . + tema)

# Resumen (por si aparecen warnings / coeficientes raros / SE enormes)
cat("\nResumen del modelo m4 (+ tema):\n")

##
## Resumen del modelo m4 (+ tema):

summary(m4)

##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + tam_cat + orientacion +
##      soporte_grp + tecnica + sop_montaje + tema, family = binomial(link = "logit"),
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -6.4473096   0.6956496  -9.268 < 2e-16 ***
## fecha_est       0.0021170   0.0003842   5.509 3.60e-08 ***
## log_ancho       0.0161148   0.0234213   0.688 0.49143
## tam_catmediano  0.0611689   0.1114788   0.549 0.58321
## tam_catgrande  0.9698516   0.1070041   9.064 < 2e-16 ***
## orientacionhorizontal 0.1920851  0.0823649   2.332 0.01969 *
## soporte_grpMetal -0.3647577  0.3754039  -0.972 0.33123
## soporte_grpMural -0.1447200  0.5596334  -0.259 0.79595
## soporte_grpOtros -1.0758535  0.3947047  -2.726 0.00642 **
## soporte_grpTabla/Panel 0.5114591  0.1257582   4.067 4.76e-05 ***
## tecnicamixta     0.1919487  0.3278557   0.585 0.55823
## tecnicaotras     -0.2149724  0.4000111  -0.537 0.59098
## sop_montajesi     0.5423510  0.2317816   2.340 0.01929 *
## temabodegon_floral 0.2161149  0.2055642   1.051 0.29311
## temacaza_animales 0.2186608  0.2932870   0.746 0.45594
## temahistoria_allegoria 0.4163296  0.1945265   2.140 0.03234 *
## temamitologia    -0.1780479  0.2441280  -0.729 0.46580
## temaotros         0.3379986  0.1154324   2.928 0.00341 **
## temapaisaje_lugares 0.5350909  0.1279326   4.183 2.88e-05 ***
## temaproceso_obra  -0.1095154  0.3453552  -0.317 0.75116
## temaretrato_corte  0.1259847  0.1287362   0.979 0.32776
## temavida_cotidiana -0.0399076  0.1869047  -0.214 0.83092
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4928.5  on 6980  degrees of freedom
## AIC: 4972.5
##
## Number of Fisher Scoring iterations: 5

```

```
# Comparación--> aporte del Bloque 4
# Información
cat("\nAporte de información del Bloque 4:\n")
```

```
##
## Aporte de información del Bloque 4:
```

```
anova(m3, m4, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje + tema
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6989      4955.2
## 2      6980      4928.5  9   26.763  0.00153 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Complejidad
cat("\nAporte de complejidad:\n")
```

```
##
## Aporte de complejidad:
```

```
AIC(m3, m4)
```

```
##      df      AIC
## m3 13 4981.247
## m4 22 4972.484
```

```
BIC(m3, m4)
```

```
##      df      BIC
## m3 13 5070.348
## m4 22 5123.271
```

El bloque iconográfico (tema) mejoró significativamente el ajuste (LRT $p = 0.001$) y redujo el AIC ($\Delta AIC \approx -9$), pero incrementó fuertemente el BIC ($\Delta BIC \approx +53$), reflejando un aumento importante de complejidad por el número de niveles de tema (se añaden 9 parámetros). Dado que el objetivo principal del estudio es caracterizar el éxito con un modelo parsimonioso y fácilmente interpretable, se decidió mantener como modelo principal el que excluye tema.

Sin embargo, dado el interés interpretativo de la iconografía, se mantiene el modelo con tema como análisis complementario específico para interpretar relaciones temáticas.

Por lo tanto, el modelo resultante después del Bloque 4 es el 'm3' y la opción alternativa contempla la inclusión de "tema"

Bloque 5: autoría y serie

Se incorporan las variables “tipo_autor” y “serie” para evaluar si la información de autoría y pertenencia a serie aporta información adicional sobre la probabilidad de éxito, una vez controlados los efectos de datación (Bloque 1), morfología (Bloque 2) y material/técnica (Bloque 3)

```
# BLOQUE 5: + tipo_autor + serie
m5 <- update(m3, . ~ . + tipo_autor + serie)

# Resumen (por si aparecen warnings / coeficientes raros / SE enormes)
cat("\nResumen del modelo m5 (+ tipo_autor + serie):\n")
```

```
##
## Resumen del modelo m5 (+ tipo_autor + serie):
```

```
summary(m5)
```

```
##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + tam_cat + orientacion +
##      soporte_grp + tecnica + sop_montaje + tipo_autor + serie,
##      family = binomial(link = "logit"), data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.7087803   0.6984622 -11.037 < 2e-16 ***
## fecha_est         0.0028765   0.0003735   7.701 1.35e-14 ***
## log_ancho         0.0197692   0.0247866   0.798  0.42512
## tam_catmediano    0.0439033   0.1105690   0.397  0.69132
## tam_catgrande     0.9209665   0.1045045   8.813 < 2e-16 ***
## orientacionhorizontal 0.3027081 0.0766217   3.951 7.79e-05 ***
## soporte_grpMetal  -0.3951589 0.3745363  -1.055  0.29140
## soporte_grpMural  -0.4827246 0.5611106  -0.860  0.38962
## soporte_grpOtros  -1.1192487 0.3988507  -2.806  0.00501 **
## soporte_grpTabla/Panel 0.5486309 0.1257558   4.363 1.28e-05 ***
## tecnicamixta       0.1537854 0.3294323   0.467  0.64063
## tecnicaotras      -0.2591605 0.3985112  -0.650  0.51548
## sop_montajesi      0.6867481 0.2288773   3.001  0.00270 **
## tipo_autoranonimo  0.1971008 0.1470093   1.341  0.18001
## tipo_autormujer    0.2748502 0.3898887   0.705  0.48084
## tipo_autorvarios   0.3445604 0.2444371   1.410  0.15866
## seriesi           0.2714728 0.0939767   2.889  0.00387 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4943.1  on 6985  degrees of freedom
## AIC: 4977.1
##
## Number of Fisher Scoring iterations: 5
```

```
# Comparación--> aporte del Bloque 5
# Información
cat("\nAporte de información del Bloque 5:\n")

##
## Aporte de información del Bloque 5:

anova(m3, m5, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje + tipo_autor + serie
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         6989      4955.2
## 2         6985      4943.1  4   12.161   0.0162 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Complejidad
cat("\nAporte de complejidad:\n")
```

```
##
## Aporte de complejidad:
```

```
AIC(m3, m5)
```

```
##      df      AIC
## m3 13 4981.247
## m5 17 4977.086
```

```
BIC(m3, m5)
```

```
##      df      BIC
## m3 13 5070.348
## m5 17 5093.603
```

El Bloque 5 (autoría y serie) mejoró significativamente el modelo previo (LRT $p = 0.01$) y redujo el AIC ($\Delta AIC \approx -4$), aunque incrementó fuertemente el BIC ($\Delta BIC \approx +23$), reflejando un aumento de complejidad. En este caso podemos observar algo interesante, los coeficientes pertenecientes a “tipo_autor” no resultaron significativos, mientras que el de “serie=1” sí. Esto reflexa que el efecto significativo dentro del bloque parece concentrarse en serie. Para confirmar este hecho, se evaluará la contribución independiente de cada variable, mediante modelos parciales, antes de tomar una decisión formal sobre el modelo principal.

```
# modelos parciales del Bloque 5
m5_serie <- update(m3, . ~ . + serie)
m5_autor <- update(m3, . ~ . + tipo_autor)

# Comparaciones --> aportes de los mdoelos parciales dle Bloque 5
# Información
cat("\nAporte de información de 'serie':\n")
```

```
##
## Aporte de información de 'serie':
```

```
anova(m3, m5_serie, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje + serie
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         6989      4955.2
## 2         6988      4946.9  1    8.3182 0.003925 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\nAporte de información de 'tipo_autor':\n")
```

```
##
## Aporte de información de 'tipo_autor':
```

```
anova(m3, m5_autor, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje + tipo_autor
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         6989      4955.2
## 2         6986      4951.2  3    4.0594  0.2551
```

```
# Complejidad
```

```
cat("\nAporte de complejidad:\n")
```

```
##
## Aporte de complejidad:
```

```
AIC(m3, m5_serie, m5_autor, m5)
```

```
##           df      AIC
## m3          13 4981.247
## m5_serie    14 4974.929
## m5_autor    16 4983.188
## m5          17 4977.086
```

```
BIC(m3, m5_serie, m5_autor, m5)
```

```
##           df      BIC
## m3          13 5070.348
## m5_serie    14 5070.884
## m5_autor    16 5092.851
## m5          17 5093.603
```

Efectivamente la variable “serie” sí demostró aportar información adicional (LRT $p = 0.003$) mejorando el AIC ($\Delta AIC \approx -6$) y manteniendo aproximadamente estable BIC, respecto al anterior modelo aceptado ‘m3’ y también reduciendo ambos criterios frente al modelo completo ‘m5’ ($\Delta AIC \approx -2$, $\Delta BIC \approx -23$). Por otro lado, “tipo_autor” no demostró aportar información adicional (LRT $p = 0.26$) además de el que aporta los valores más elevados tanto en AIC como en BIC, superando incluso los del modelo completo ‘m5’. Se decide prescindir de la variable “tipo_autor” y conservar el modelo únicamente con la inclusión de “serie”.

Modelo resultante después de añadir Bloque 5:

```
m5 <- m5_serie
```

Resumen

Después de evaluar los efectos principales, el modelo principal es el siguiente:

```
m_final <- glm(
  exito ~ fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie,
  data = df, family = binomial(link = "logit")
)
summary(m_final)
```

```
##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + tam_cat + orientacion +
##      soporte_grp + tecnica + sop_montaje + serie, family = binomial(link = "logit"),
##      data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -7.5545967   0.6911174 -10.931 < 2e-16 ***
## fecha_est        0.0027956   0.0003694   7.567 3.81e-14 ***
## log_ancho        0.0290842   0.0232929   1.249  0.21180
## tam_catmediano   0.0473151   0.1104536   0.428  0.66838
## tam_catgrande    0.9178941   0.1042498   8.805 < 2e-16 ***
## orientacionhorizontal 0.3041996   0.0764413   3.980 6.91e-05 ***
## soporte_grpMetal -0.4025129   0.3743727  -1.075  0.28230
## soporte_grpMural -0.4518929   0.5600464  -0.807  0.41973
## soporte_grpOtros -1.0662903   0.3942547  -2.705  0.00684 **
## soporte_grpTabla/Panel 0.5493312   0.1251844   4.388 1.14e-05 ***
## tecnicamixta      0.1657706   0.3292548   0.503  0.61463
## tecnicaotras     -0.2221573   0.3943927  -0.563  0.57324
## sop_montajesi     0.6860477   0.2283316   3.005  0.00266 **
```

```
## seriesi                0.2744989  0.0937574   2.928  0.00341 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4946.9  on 6988  degrees of freedom
## AIC: 4974.9
##
## Number of Fisher Scoring iterations: 5
```

Interacciones

Para decidir que variables son candidatas de entrar al modelo primero traeremos las conclusiones obre estas obtenidas en el análisis descriptivo. En dicha sección se determinó que las más plausibles era “soporte_grp x fecha_est”, “tecnica x fecha_est”, “soporte_grp x tam_cat”, “tecnica x tam_cat” y “soporte_grp x orientacion”. Graficaremos los correspondientes gráficos de interacción para estas opciones y seleccionaremos las mejores, que seguidamente serán comprobadas con pruebas formales.

Gráficos de interacción

Graficamos $P(\text{exito} = 1)$ predicha por el modelo, fijando el resto de covariables en valores de referencia/mediana.

```
# niveles de referencia factores ([1] porque ya se especificó al inicio del análisis)
ref_tam_cat      <- levels(df$tam_cat)[1]
ref_orientacion  <- levels(df$orientacion)[1]
ref_soporte_grp  <- levels(df$soporte_grp)[1]
ref_tecnica      <- levels(df$tecnica)[1]
ref_sop_montaje  <- levels(df$sop_montaje)[1]
ref_serie        <- levels(df$serie)[1]

# niveles de referencia numéricas (mediana)
fecha0          <- median(df$fecha_est)
log_ancho0      <- median(df$log_ancho)
```

1) soporte_grp x fecha_est

```
# =====
# 1) soporte_grp x fecha_est
# =====
m_int_SopFecha <- glm(
  exito ~ fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie +
    soporte_grp:fecha_est,
  data = df, family = binomial(link = "logit")
)

x <- seq(min(df$fecha_est, na.rm = TRUE), max(df$fecha_est, na.rm = TRUE), length.out = 200)
lev <- levels(df$soporte_grp)
```



```

nd <- expand.grid(
  fecha_est = x,
  soporte_grp = lev,
  KEEP.OUT.ATTRS = FALSE,
  stringsAsFactors = FALSE
)

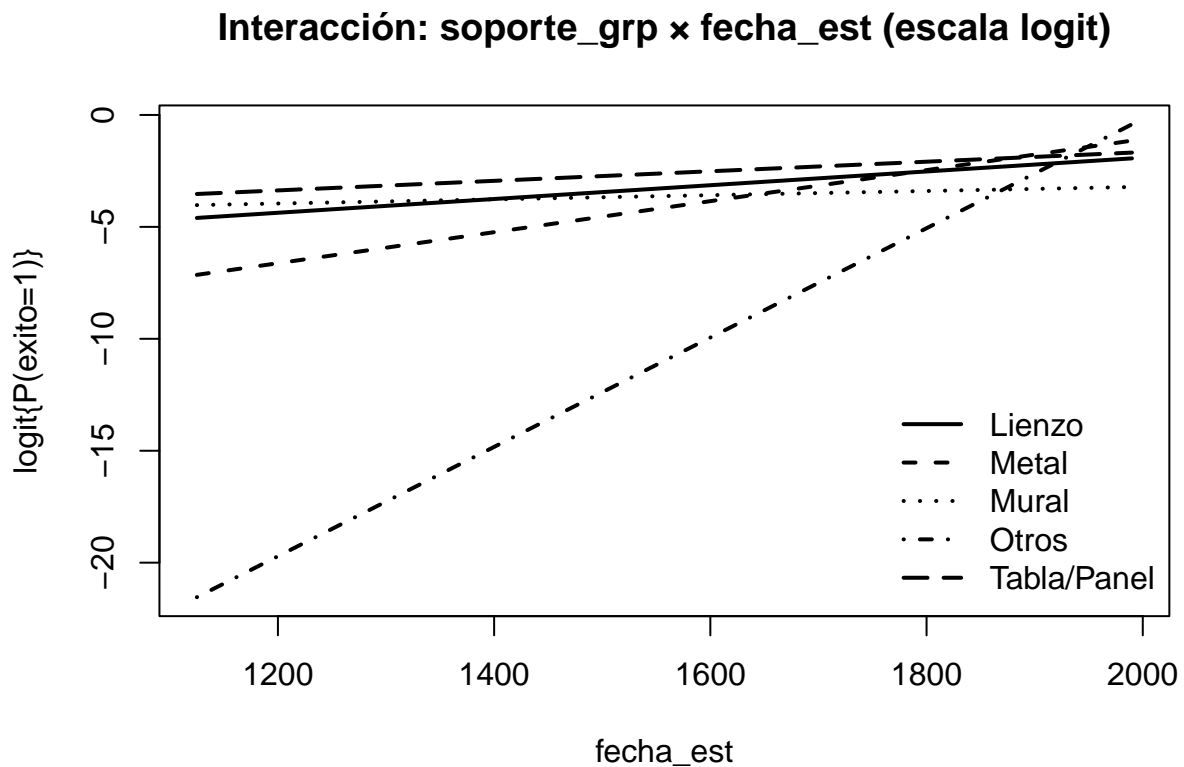
nd$log_ancho <- log_ancho0
nd$tam_cat <- ref_tam_cat
nd$orientacion <- ref_orientacion
nd$tecnica <- ref_tecnica
nd$sop_montaje <- ref_sop_montaje
nd$serie <- ref_serie

eta <- predict(m_int_SopFecha, newdata = nd, type = "link")

plot(range(x), range(eta), type = "n",
      xlab = "fecha_est", ylab = "logit{P(exito=1)}",
      main = "Interacción: soporte_grp x fecha_est (escala logit)")

for (i in seq_along(lev)) {
  idx <- nd$soporte_grp == lev[i]
  lines(nd$fecha_est[idx], eta[idx], lty = i, lwd = 2)
}
legend("bottomright", legend = lev, lty = seq_along(lev), lwd = 2, bty = "n")

```



Esta interacción plantea algunas preguntas, ya que aunque podemos observar algunos indicios vemos que existe una categoría, Tabla/Panel, que presenta una fuerte diferencia. Esto nos hace pensar en que puede tratarse de una escasez de datos en el extremo.

Comprobamos frecuencias y recuento de éxitos por combinación:

```
# Rangos observados por soporte
tapply(df$fecha_est, df$soporte_grp, range)
```

```
## $Lienzo
## [1] 1286 1990
##
## $Metal
## [1] 1356 1880
##
## $Mural
## [1] 1125 1822
##
## $Otros
## [1] 1515 1952
##
## $'Tabla/Panel'
## [1] 1205 1957
```

```
# Cortes de fecha_cat
grp_fecha <- cut(df$fecha_est,
breaks = c(1100, 1400, 1700, 2000),
include.lowest = TRUE)

cat("\nFrecuencias por combinación:\n")
```

```
##
## Frecuencias por combinación:
```

```
xtabs(~ soporte_grp + grp_fecha, data = df)
```

```
##           grp_fecha
## soporte_grp [1.1e+03,1.4e+03] (1.4e+03,1.7e+03] (1.7e+03,2e+03]
## Lienzo           9           2533           2974
## Metal            1           129           24
## Mural           25           10           14
## Otros            0           35          102
## Tabla/Panel     20          884          242
```

```
cat("\nRecuento de éxitos por combinación:\n")
```

```
##
## Recuento de éxitos por combinación:
```

```
xtabs(exito ~ soporte_grp + grp_fecha, data = df)
```

##	grp_fecha			
## soporte_grp	[1.1e+03,1.4e+03]	(1.4e+03,1.7e+03]	(1.7e+03,2e+03]	
## Lienzo	0	266	433	
## Metal	0	5	3	
## Mural	2	2	2	
## Otros	0	0	9	
## Tabla/Panel	2	91	34	

Estos resultados explican la forma rara del gráfico “soporte_grp x fecha_est”. Vemos como en el tramo [1100, 1400] todo son frecuencias bajas y celdas vacías. Por esta razón el modelo interpreta que esas fechas la probabilidad de éxito es tan baja. Consecuentemente, decidimos descartar esta interacción por inconsistencia de resultados resultantes de extrapolación del modelo en regiones sin datos.

2) tecnica x fecha_est

```
# =====
# 2) tecnica x fecha_est
# =====

m_int_TecFecha <- glm(
  exito ~ fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie +
    tecnica:fecha_est,
  data = df, family = binomial(link = "logit")
)

x <- seq(min(df$fecha_est, na.rm = TRUE), max(df$fecha_est, na.rm = TRUE), length.out = 200)
lev <- levels(df$tecnica)

nd <- expand.grid(
  fecha_est = x,
  tecnica = lev,
  KEEP.OUT.ATTRS = FALSE,
  stringsAsFactors = FALSE
)

nd$log_ancho <- log_ancho0
nd$tam_cat <- ref_tam_cat
nd$orientacion <- ref_orientacion
nd$soporte_grp <- ref_soporte_grp
nd$sop_montaje <- ref_sop_montaje
nd$serie <- ref_serie

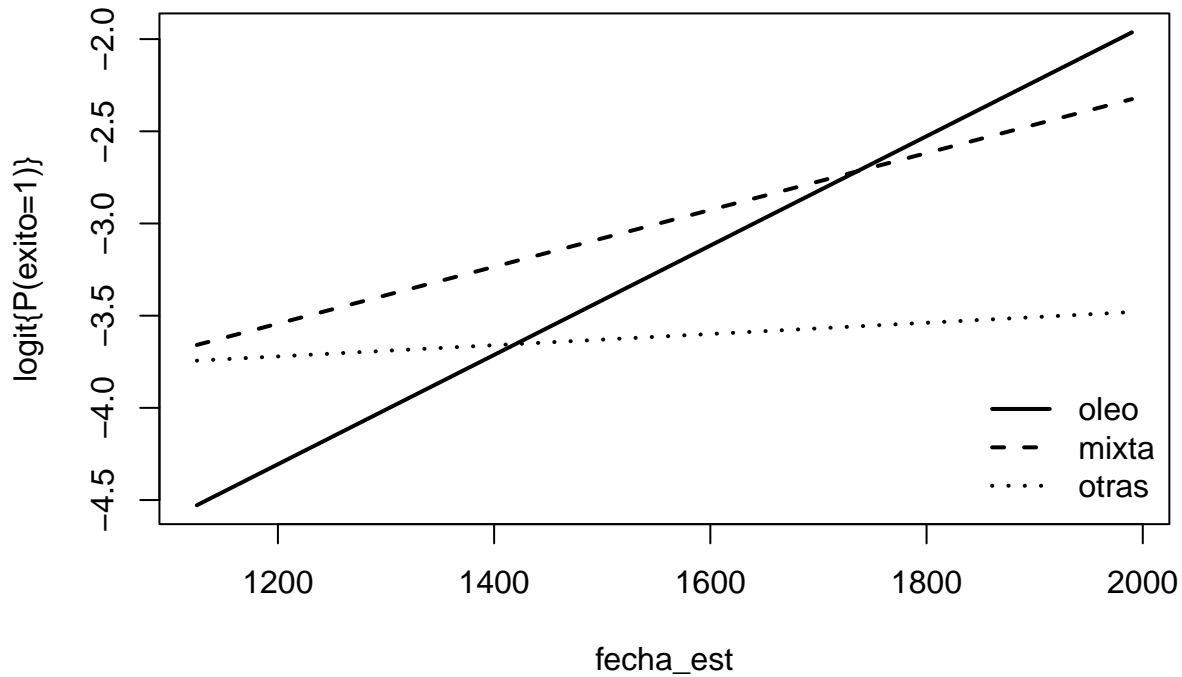
eta <- predict(m_int_TecFecha, newdata = nd, type = "link")

plot(range(x), range(eta), type = "n",
  xlab = "fecha_est", ylab = "logit{P(exito=1)}",
  main = "Interacción: tecnica x fecha_est (escala logit)")

for (i in seq_along(lev)) {
  idx <- nd$tecnica == lev[i]
  lines(nd$fecha_est[idx], eta[idx], lty = i, lwd = 2)
}
```

```
}
legend("bottomright", legend = lev, lty = seq_along(lev), lwd = 2, bty = "n")
```

Interacción: tecnica x fecha_est (escala logit)



Vemos indicios claros de interacción y un claro cruce: ólea pasa de estar por debajo a por encima. Sin embargo, estas observaciones podrían estar de nuevo sesgadas por falta de datos en algunos periodos. comprobamos recuentos:

```
cat("\nFrecuencias por combinación:\n")
```

```
##
## Frecuencias por combinación:
```

```
xtabs(~ tecnica + grp_fecha, data=df)
```

```
##      grp_fecha
## tecnica [1.1e+03,1.4e+03] (1.4e+03,1.7e+03] (1.7e+03,2e+03]
##  oleo          13          3440          3319
##  mixta          5           98           19
##  otras         37           53           18
```

```
cat("\nRecuento de éxitos por combinación:\n")
```

```
##
## Recuento de éxitos por combinación:
```

```
xtabs(exito ~ tecnica + grp_fecha, data=df)
```

```
##          grp_fecha
## tecnica [1.1e+03,1.4e+03] (1.4e+03,1.7e+03] (1.7e+03,2e+03]
##   oleo                0                349                477
##   mixta                0                 10                 4
##   otras                4                 5                 0
```

” Efectivamente podemos ver como la variable “tecnica” está fuertemente sesgada debido al nivel ‘oleo’. Aunque esta variable forma parte de nuestras hipótesis principales, su inclusión ya se aceptó de manera separada asumiendo un aumento de complejidad elevado (incremento significativo en BIC en Bloque 2 de efectos principales), por lo que su interés conceptual e interpretativo ya fue considerado. Por esta razón preferimos descartar su interacción

3) tecnica x tam_cat

```
# =====
# 3) tecnica x tam_cat
# =====
m_int_TecTam <- glm(
  exito ~ fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie +
    tecnica:tam_cat,
  data = df, family = binomial(link = "logit")
)

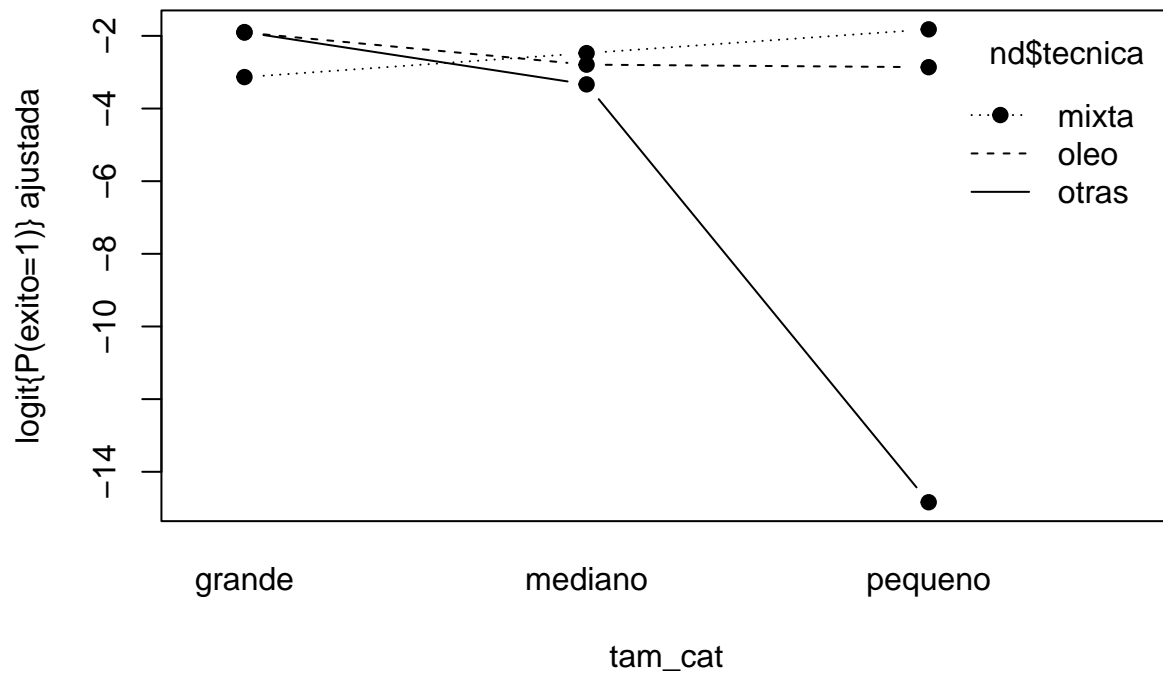
nd <- expand.grid(
  tam_cat = levels(df$tam_cat),
  tecnica = levels(df$tecnica),
  KEEP.OUT.ATTRS = FALSE,
  stringsAsFactors = FALSE
)

nd$fecha_est <- fecha0
nd$log_ancho <- log_ancho0
nd$orientacion <- ref_orientacion
nd$soporte_grp <- ref_soporte_grp
nd$sop_montaje <- ref_sop_montaje
nd$serie <- ref_serie

nd$eta <- predict(m_int_TecTam, newdata = nd, type = "link")

interaction.plot(x.factor = nd$tam_cat, trace.factor = nd$tecnica,
  response = nd$eta, type = "b", pch = 19,
  xlab = "tam_cat", ylab = "logit{P(exito=1)} ajustada",
  main = "Interacción: tecnica x tam_cat (escala logit)")
```

Interacción: tecnica x tam_cat (escala logit)



De nuevo vemos puntos extremos de manera que comprobaremos recuentos:

```
cat("\nFrecuencias por combinación:\n")
```

```
##
## Frecuencias por combinación:
```

```
xtabs(~ tecnica + tam_cat, data=df)
```

```
##          tam_cat
## tecnica pequeno mediano grande
##   oleo      2259      2216      2297
##   mixta       32       64       26
##   otras       15       37       56
```

```
cat("\nRecuento de éxitos por combinación:\n")
```

```
##
## Recuento de éxitos por combinación:
```

```
xtabs(exito ~ tecnica + tam_cat, data=df)
```

```
##          tam_cat
```

```
## tecnica pequeno mediano grande
##   oleo      217      189      420
##   mixta       6       6       2
##   otras       0       2       7
```

Obtenemos los mismo resultados en esta interacción con “tecnica”: nivel ‘oleo’ fuertemente predominante. Se descarta esta interacción.

4) soporte_grp x tam_cat

```
# =====
# 4) soporte_grp x tam_cat
# =====
m_int_SopTam <- glm(
  exito ~ fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie +
    soporte_grp:tam_cat,
  data = df, family = binomial(link = "logit")
)

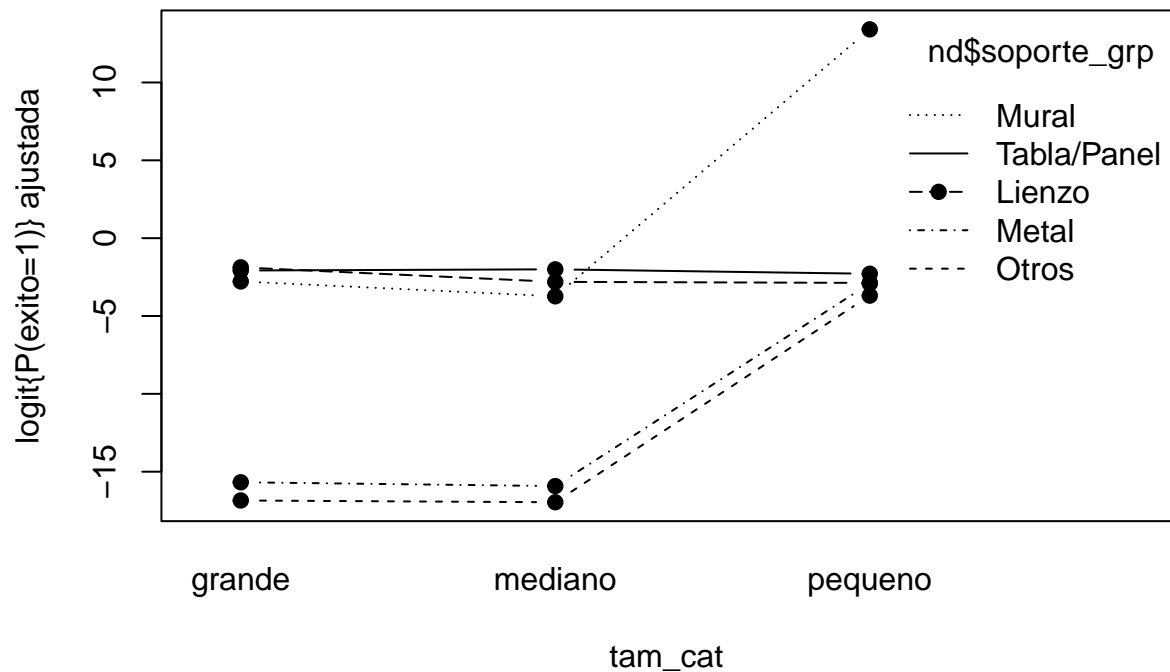
nd <- expand.grid(
  tam_cat      = levels(df$tam_cat),
  soporte_grp  = levels(df$soporte_grp),
  KEEP.OUT.ATTRS = FALSE,
  stringsAsFactors = FALSE
)

nd$fecha_est   <- fecha0
nd$log_ancho   <- log_ancho0
nd$orientacion <- ref_orientacion
nd$tecnica     <- ref_tecnica
nd$sop_montaje <- ref_sop_montaje
nd$serie       <- ref_serie

nd$eta <- predict(m_int_SopTam, newdata = nd, type = "link")

interaction.plot(x.factor = nd$tam_cat, trace.factor = nd$soporte_grp,
  response = nd$eta, type = "b", pch = 19,
  xlab = "tam_cat", ylab = "logit{P(exito=1)} ajustada",
  main = "Interacción: soporte_grp x tam_cat (escala logit)")
```

Interacción: soporte_grp x tam_cat (escala logit)



A primera vista podemos ver algunos incidios pero no determinantes de modificación del efecto. Podrían explicarse por la baja frecuencia de algunas combinaciones:

```
cat("\nFrecuencias por combinación:\n")
```

```
##
## Frecuencias por combinación:
```

```
xtabs(~ soporte_grp + tam_cat, data=df)
```

```
##           tam_cat
## soporte_grp  pequeno mediano grande
## Lienzo       1402    1884   2230
## Metal         120      31      3
## Mural          1      12     36
## Otros         107      28      2
## Tabla/Panel   676     362    108
```

```
cat("\nRecuento de éxitos por combinación:\n")
```

```
##
## Recuento de éxitos por combinación:
```



```
xtabs(exito ~ soporte_grp + tam_cat, data=df)
```

```
##           tam_cat
## soporte_grp  pequeno mediano grande
## Lienzo      129      156    414
## Metal        8        0      0
## Mural        1        1      4
## Otros        9        0      0
## Tabla/Panel  76       40     11
```

Efectivamente vemos que categorías como Metal-grande o Mural-pequeño presentan frecuencias realmente bajas, además los exitos se concentran alrededor de las categoría 'Lienzo' y 'Tabla/Panel' lo cual es lógico ya que són las categorías mayoritarias. Sin embargo se decide aceptar esta interacción como candidata para las posteriores pruebas formales, ya que es una hipótesis central del estudio.

5) soporte_grp x orientacion

```
# =====
# 5) soporte_grp x orientacion
# =====

m_int_SopOri <- glm(
  exito ~ fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie +
    soporte_grp:orientacion,
  data = df, family = binomial(link = "logit")
)

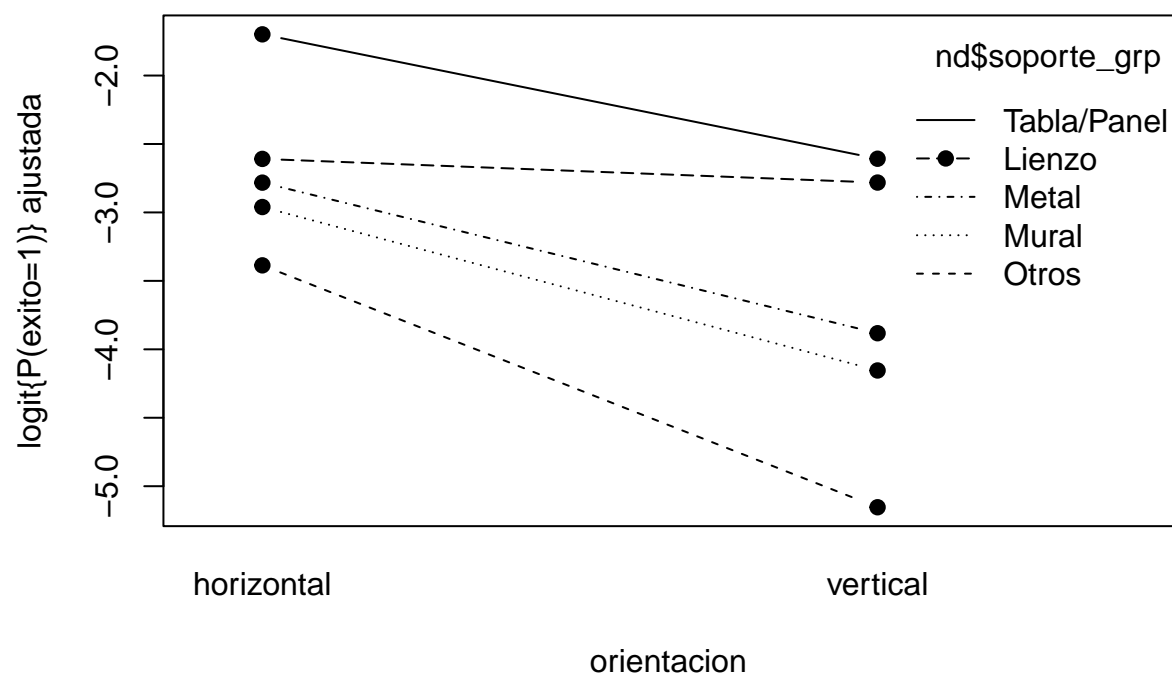
nd <- expand.grid(
  orientacion = levels(df$orientacion),
  soporte_grp = levels(df$soporte_grp),
  KEEP.OUT.ATTRS = FALSE,
  stringsAsFactors = FALSE
)

nd$fecha_est <- fecha0
nd$log_ancho <- log_ancho0
nd$tam_cat <- ref_tam_cat
nd$tecnica <- ref_tecnica
nd$sop_montaje <- ref_sop_montaje
nd$serie <- ref_serie

nd$eta <- predict(m_int_SopOri, newdata = nd, type = "link")

interaction.plot(x.factor = nd$orientacion, trace.factor = nd$soporte_grp,
  response = nd$eta, type = "b", pch = 19,
  xlab = "orientacion", ylab = "logit{P(exito=1)} ajustada",
  main = "Interacción: soporte_grp x orientacion (escala logit)")
```

Interacción: soporte_grp × orientacion (escala logit)



Esta interacción ha resultado la menos relevante, pero aun con incidios de posible interacción. Comprobaremos también los recuentos:

```
cat("\nFrecuencias por combinación:\n")
```

```
##
## Frecuencias por combinación:
```

```
xtabs(~ soporte_grp + orientacion, data=df)
```

```
##          orientacion
## soporte_grp  vertical horizontal
## Lienzo          3159         2357
## Metal            49          105
## Mural            20           29
## Otros            53           84
## Tabla/Panel      720          426
```

```
cat("\nRecuento de éxitos por combinación:\n")
```

```
##
## Recuento de éxitos por combinación:
```

```
xtabs(exito ~ soporte_grp + orientacion, data=df)
```

```
##          orientacion
## soporte_grp  vertical horizontal
## Lienzo          350          349
## Metal            1            7
## Mural            1            5
## Otros            1            8
## Tabla/Panel      54            73
```

Vemos el mismo patrón para la variable respuesta: los éxitos se concentran al rededor de ‘Lienzo’, sin embargo no vemos fuertes desbalances para los grupos de orientación. La mantenemos como posible candidata a pruebas

Pruebas formales

Decidimos testear las siguientes interacciones: “soporte_grp x tam_cat” y “soporte_grp x orientacion”. Ambas opciones parecen plausibles tanto por su representación gráfica como por interpretación conceptual, además concretamente “soporte x tam_cat” se incluía en nuestras hipótesis, de manera que consideramos muy apropiada esta selección.

Comprobaremos en primer caso la inclusión de cada interacción de manera separada para estudiar si cada una por separado aporta información al modelo y su compromiso ajuste-complejidad.

```
# soporte_grp x tam_cat
cat("\n===== \n soporte_grp x tam_cat \n===== \n")
```

```
##
## =====
## soporte_grp x tam_cat
## =====
```

```
m_soporte_tamcat <- update(m_final, . ~ . + soporte_grp:tam_cat)
anova(m_final, m_soporte_tamcat, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##          tecnica + sop_montaje + serie
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##          tecnica + sop_montaje + serie + tam_cat:soporte_grp
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6988      4946.9
## 2      6980      4925.6  8    21.367 0.006233 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(m_final, m_soporte_tamcat)
```

```
##          df          AIC
## m_final      14 4974.929
## m_soporte_tamcat 22 4969.561
```

```
BIC(m_final, m_soporte_tamcat)
```

```
##                df        BIC
## m_final        14 5070.884
## m_soporte_tamcat 22 5120.348
```

```
# soporte_grp x orientacion
```

```
cat("\n===== \n soporte_grp x orientacion \n===== \n")
```

```
##
## =====
##  soporte_grp x orientacion
## =====
```

```
m_soporte_orientacion <- update(m_final, . ~ . + soporte_grp:orientacion)
anova(m_final, m_soporte_orientacion, test = "Chisq")
```

```
## Analysis of Deviance Table
```

```
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##   tecnica + sop_montaje + serie
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##   tecnica + sop_montaje + serie + orientacion:soporte_grp
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         6988      4946.9
## 2         6984      4931.6  4    15.335 0.004054 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(m_final, m_soporte_orientacion)
```

```
##                df        AIC
## m_final        14 4974.929
## m_soporte_orientacion 18 4967.594
```

```
BIC(m_final, m_soporte_orientacion)
```

```
##                df        BIC
## m_final        14 5070.884
## m_soporte_orientacion 18 5090.965
```

Vemos que ambas interacciones demuestran mejorar significativamente el modelo aportando informacion (LRT(99) $p < 0.01$) i reduciendo el AIC ($\Delta AIC \approx -7; -10$). Aunque el BIC aumento en los dos casos ($\Delta BIC \approx +47; +17$). Por interés interpretativo decidimos mantener el modelo con la interacción “soporte_grp x tam_cat” como base y procedemos a examinar el modelo completo anidado con la otra interacción.

```
m_completo <- update(m_soporte_tamcat, . ~ . + soporte_grp:orientacion)
anova(m_final, m_soporte_tamcat, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje + serie
## Model 2: exito ~ fecha_est + log_ancho + tam_cat + orientacion + soporte_grp +
##      tecnica + sop_montaje + serie + tam_cat:soporte_grp
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6988      4946.9
## 2      6980      4925.6  8   21.367 0.006233 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(m_final, m_soporte_tamcat)
```

```
##              df      AIC
## m_final      14 4974.929
## m_soporte_tamcat 22 4969.561
```

```
BIC(m_final, m_soporte_tamcat)
```

```
##              df      BIC
## m_final      14 5070.884
## m_soporte_tamcat 22 5120.348
```

Vemos que la inclusión de la interacción “soporte_grp x orientación” sigue siendo significativa una vez controlado el efecto de “soporte_grp x tam_cat”, sin embargo provoca un brave problema de complejidad ($\Delta BIC \approx +74$) que no consideramos aceptable ni necesario en este punto análisis. En consecuencia descartamos la interaccion con orientación una vez controlado por tamaño. Además ya detectamos anteriormente existencia de celdas problemáticas en soporte, que pueden provocar separación, por lo que algunos coeficientes pueden volverse fuertemente inestables. Por esta razón, debemos mencionar que la interacción con tamaño se mantendrá pero con interpretación principalmente en ‘Lienzo’ y ‘Tabla/Panel’, además de tratará de minimizar esta problematica en la siguiente sección.

Finalmente se decide definir el modelo principal con únicamente la interacción “soporte_grp x tam_cat” con el obojetivo de dar respuesta a nuestra hipótesis, pero teniendo en cuenta el fuerte aumento de BIC y la existencia de posible inestabilidad, se considera conservar el modelo sin interacciones como alternativo con el fin de explorar más rigurosamente los efectos principales, si se considera oportuno.

Resumen

Después de evaluar las interacciones, el modelo principal es el siguiente:

```
m_completo <- glm(
  exito ~
    fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie +
    soporte_grp:tam_cat,
  data = df,
  family = binomial(link = "logit")
)

summary(m_completo)
```

```
##
## Call:
## glm(formula = exito ~ fecha_est + log_ancho + tam_cat + orientacion +
##      soporte_grp + tecnica + sop_montaje + serie + soporte_grp:tam_cat,
##      family = binomial(link = "logit"), data = df)
##
## Coefficients:
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.549e+00  6.939e-01 -10.880 < 2e-16
## fecha_est         2.767e-03  3.733e-04   7.411 1.25e-13
## log_ancho         2.947e-02  2.332e-02   1.264 0.206323
## tam_catmediano    6.620e-02  1.303e-01   0.508 0.611445
## tam_catgrande     9.977e-01  1.158e-01   8.615 < 2e-16
## orientacionhorizontal 2.869e-01  7.685e-02   3.734 0.000189
## soporte_grpMetal  -3.127e-02  3.834e-01  -0.082 0.934991
## soporte_grpMural   1.629e+01  1.455e+03   0.011 0.991071
## soporte_grpOtros   -8.162e-01  3.957e-01  -2.063 0.039139
## soporte_grpTabla/Panel 5.858e-01  1.620e-01   3.615 0.000300
## tecnicamixta       3.942e-01  3.423e-01   1.152 0.249498
## tecnicaotras      -1.102e-01  3.933e-01  -0.280 0.779314
## sop_montajesi       7.855e-01  2.335e-01   3.364 0.000768
## seriesi            2.799e-01  9.416e-02   2.973 0.002954
## tam_catmediano:soporte_grpMetal -1.309e+01  2.604e+02  -0.050 0.959920
## tam_catgrande:soporte_grpMetal -1.377e+01  8.390e+02  -0.016 0.986902
## tam_catmediano:soporte_grpMural -1.721e+01  1.455e+03  -0.012 0.990563
## tam_catgrande:soporte_grpMural -1.719e+01  1.455e+03  -0.012 0.990575
## tam_catmediano:soporte_grpOtros -1.334e+01  2.726e+02  -0.049 0.960978
## tam_catgrande:soporte_grpOtros -1.416e+01  1.023e+03  -0.014 0.988954
## tam_catmediano:soporte_grpTabla/Panel 2.191e-01  2.513e-01   0.872 0.383262
## tam_catgrande:soporte_grpTabla/Panel -7.876e-01  3.773e-01  -2.088 0.036839
##
## (Intercept)      ***
## fecha_est        ***
## log_ancho
## tam_catmediano
## tam_catgrande    ***
## orientacionhorizontal ***
## soporte_grpMetal
## soporte_grpMural
## soporte_grpOtros  *
## soporte_grpTabla/Panel ***
## tecnicamixta
## tecnicaotras
## sop_montajesi    ***
## seriesi          **
## tam_catmediano:soporte_grpMetal
## tam_catgrande:soporte_grpMetal
## tam_catmediano:soporte_grpMural
## tam_catgrande:soporte_grpMural
## tam_catmediano:soporte_grpOtros
## tam_catgrande:soporte_grpOtros
## tam_catmediano:soporte_grpTabla/Panel
## tam_catgrande:soporte_grpTabla/Panel *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4925.6  on 6980  degrees of freedom
## AIC: 4969.6
##
## Number of Fisher Scoring iterations: 14
```

Y el modelo reducido conservado como alternativo, sin interacciones, es el siguiente:

```
m_reducido <- glm(
  exito ~
    fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie,
  data = df,
  family = binomial(link = "logit")
)
```

Diagnóstico de ajuste y correcciones

En esta sección se presentan diagnósticos preliminares del modelo, centrados en la calidad del ajuste y en la estabilidad de los parámetros, especialmente considerando la baja prevalencia del evento y el uso de múltiples factores e interacciones. No se llevarán a cabo aún procedimientos de validación formales, ya que se abordarán en secciones posteriores.

```
y <- model.response(model.frame(m_completo))
E <- sum(y == 1)
N <- length(y)
p <- length(coef(m_completo))

cat("N =", N, "  Eventos (1) =", E, "  Prevalencia =", round(E/N, 4), "\n")
```

```
## N = 7002  Eventos (1) = 849  Prevalencia = 0.1213
```

```
cat("Num coeficientes (incluye dummies) =", p, "\n")
```

```
## Num coeficientes (incluye dummies) = 22
```

```
cat("EPV aprox (eventos por coef) =", round(E/p, 3), "\n")
```

```
## EPV aprox (eventos por coef) = 38.591
```

Con 7002 observaciones y 849 evento, el modelo dispone de información suficiente para estimar los 22 parámetros que contiene. El coeficiente de ($EPV \approx 38.6$) sugiere que el desnivel en la respuesta no plantea una limitación para el modelo seleccionado.

```
Y <- xtabs(exito ~ soporte_grp + tam_cat, data = df)
N <- xtabs(~ soporte_grp + tam_cat, data = df)
P <- Y / N
```

```
cat("\nFrecuencias de soporte\n")
```

```
##
## Frecuencias de soporte
```

```
table(df$soporte_grp)
```

```
##
##      Lienzo      Metal      Mural      Otros Tabla/Panel
##      5516      154      49      137      1146
```

```
cat("\nFrecuencias de tamaño\n")
```

```
##
## Frecuencias de tamaño
```

```
table(df$tam_cat)
```

```
##
## pequeno mediano grande
##      2306      2317      2379
```

```
cat("\nÉxitos Y:\n"); print(Y)
```

```
##
## Éxitos Y:
```

```
##           tam_cat
## soporte_grp pequeno mediano grande
##   Lienzo      129      156      414
##   Metal        8        0        0
##   Mural        1        1        4
##   Otros        9        0        0
##   Tabla/Panel  76       40       11
```

```
cat("\nTotales N:\n"); print(N)
```

```
##
## Totales N:
```

```
##           tam_cat
## soporte_grp pequeno mediano grande
##   Lienzo      1402      1884      2230
##   Metal        120        31        3
##   Mural         1        12       36
##   Otros        107        28        2
##   Tabla/Panel  676       362      108
```



```
cat("\nProporción P=Y/N:\n"); print(round(P, 3))
```

```
##
```

```
## Proporción P=Y/N:
```

```
##           tam_cat
## soporte_grp  pequeno mediano grande
## Lienzo      0.092   0.083  0.186
## Metal       0.067   0.000  0.000
## Mural       1.000   0.083  0.111
## Otros      0.084   0.000  0.000
## Tabla/Panel 0.112   0.110  0.102
```

```
cat("\nCeldas con 0 éxitos (Y==0):\n")
```

```
##
```

```
## Celdas con 0 éxitos (Y==0):
```

```
print((Y == 0) & (N > 0))
```

```
##           tam_cat
## soporte_grp  pequeno mediano grande
## Lienzo      FALSE  FALSE  FALSE
## Metal       FALSE   TRUE   TRUE
## Mural       FALSE  FALSE  FALSE
## Otros      FALSE   TRUE   TRUE
## Tabla/Panel FALSE  FALSE  FALSE
```

```
cat("\nCeldas con todos éxitos (Y==N):\n")
```

```
##
```

```
## Celdas con todos éxitos (Y==N):
```

```
print((Y == N) & (N > 0))
```

```
##           tam_cat
## soporte_grp  pequeno mediano grande
## Lienzo      FALSE  FALSE  FALSE
## Metal       FALSE  FALSE  FALSE
## Mural       TRUE   FALSE  FALSE
## Otros      FALSE  FALSE  FALSE
## Tabla/Panel FALSE  FALSE  FALSE
```

Las tablas de contingencia de éxitos y totales por combinación de la interacción muestra celdas con respuestas deterministas lo que puede provocar separación. En concreto observamos todo éxitos en Mural-pequeño, lo cual tiene sentido conceptualmente ya que los murales están asociados a grandes obras de arte; y observamos ausencia total de éxitos en Metal-mediano, grande y Otros-mediano, grande. Estos resultados eran esperables debido a la baja frecuencia de las categorías involucradas Metal/Mural/Otros. Recordamos entonces la importancia de centrar la interpretación al rededor de las categorías estables Tabla/Panel y Lienzo, aun sabiendo que no están ajenas a la problemática.

```
library(car)
vif(m_completo)
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

```
##              GVIF Df GVIF^(1/(2*Df))
## fecha_est      1.662901e+00  1      1.289535
## log_ancho      1.283442e+00  1      1.132891
## tam_cat        2.123473e+00  2      1.207151
## orientacion    1.061100e+00  1      1.030097
## soporte_grp    2.718990e+07  4      8.497691
## tecnica        1.465668e+00  2      1.100294
## sop_montaje    1.614959e+00  1      1.270810
## serie          1.103178e+00  1      1.050323
## tam_cat:soporte_grp 2.256196e+07  8      2.881284
```

Los indicadores de colinealidad muestran valores muy altos para soporte_grp ($GVIF \approx 8.50$) y elevados para tam_cat:soporte_grp ($GVIF \approx 2.88$), en contraste con el resto de covariables, cuyos valores permanecen cercanos a 1. Esta evidencia afirma una dependencia fuerte entre los bloques de parámetros de los efectos principales y los de la interacción, especialmente en presencia de un diseño como el nuestro, desbalanceado entre niveles y con celdas con baja frecuencia.

```
n <- nobs(m_completo)
p <- length(coef(m_completo))

cat("Devianza ajustada:", deviance(m_completo) / (n - p))
```

```
## Devianza ajustada: 0.7056678
```

La evaluación del ajuste global mediante devianza ajustada no aporta indicios de sobredispersión ($\varphi \approx 0,706$), es proxima a 1. Por lo que estos resultados sugieren que no existe un exceso de variabilidad no explicada que justifique adoptar estrategias orientadas a corregir sobredispersión.

Después de este primer análisis vemos que existen dos problemáticas que debemos manejar, la separación por celdas vacías y la colinealidad excesiva en soporte.

Primeramente abordaremos la problemática provocada por las celdas deterministas de la interacción y más adelante, si la colinealidad siguiera presente, se adoptarán nuevas medidas.

Separación

Contrastaremos formalmente la existencia de separación en nuestro modelo:

```
df2 <- df
df2$exito <- if (is.factor(df2$exito)) as.integer(df2$exito == levels(df2$exito)[2]) else as.integer(df2$exito)

m_df2 <- update(m_completo, data = df2)

X <- model.matrix(m_df2)
y <- model.response(model.frame(m_df2))
ds <- detectseparation::detect_separation(X, y)
```

```
## Warning in detectseparation::detect_separation(X, y): 'detect_separation' has
## been developed for use with binomial-response GLMs
```

```
print(ds)
```

```
## Implementation: ROI | Solver: lpsolve
## Separation: TRUE
## Existence of maximum likelihood estimates
##              (Intercept)              fecha_est
##              0              0
##              log_ancho              tam_catmediano
##              0              0
##              tam_catgrande              orientacionhorizontal
##              0              0
##              soporte_grpMetal              soporte_grpMural
##              0              Inf
##              soporte_grpOtros              soporte_grpTabla/Panel
##              0              0
##              tecnicamixta              tecnicaotras
##              0              0
##              sop_montajesi              seriesi
##              0              0
##      tam_catmediano:soporte_grpMetal      tam_catgrande:soporte_grpMetal
##              -Inf              -Inf
##      tam_catmediano:soporte_grpMural      tam_catgrande:soporte_grpMural
##              -Inf              -Inf
##      tam_catmediano:soporte_grpOtros      tam_catgrande:soporte_grpOtros
##              -Inf              -Inf
##      tam_catmediano:soporte_grpTabla/Panel      tam_catgrande:soporte_grpTabla/Panel
##              0              0
## 0: finite value, Inf: infinity, -Inf: -infinity
```

Efectivamente nos enfrentamos a un problema de separación real, lo cual puede traducirse en inexistencia de coeficientes finitos. En consecuencia, decidimos reestimar el mismo modelo mediando regresión logística con reducción de sesgo con un enfoque Firth.

```
library(brglm2)
```

```
## Warning: el paquet 'brglm2' es va construir amb la versió d'R 4.5.2
```

```
m_firth <- glm(
  formula(m_completo),
  data = df,
  family = binomial("logit"),
  method = "brglmFit",
)
```

```
## Warning: brglmFit: algorithm did not converge. Try changing the optimization
## algorithm defaults, e.g. the defaults for one or more of 'maxit', 'epsilon',
## 'slowit', and 'response_adjustment'; see '?brglm_control' for default values
## and available options
```

```
## Warning: brglmFit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(m_firth)
```

```
##
## Call:
## glm(formula = formula(m_completo), family = binomial("logit"),
##      data = df, method = "brglmFit")
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1877  -0.5361  -0.4259  -0.3022   8.4904
##
## Coefficients:
##              Estimate Std. Error  z value Pr(>|z|)
## (Intercept)    -8.346e+00  7.506e-01 -1.112e+01 < 2e-16
## fecha_est       3.188e-03  4.030e-04  7.909e+00 2.59e-15
## log_ancho       4.195e-02  2.470e-02  1.698e+00 0.089477
## tam_catmediano  8.759e-02  1.311e-01  6.680e-01 0.503954
## tam_catgrande  1.025e+00  1.170e-01  8.762e+00 < 2e-16
## orientacionhorizontal 3.051e-01  7.978e-02  3.825e+00 0.000131
## soporte_grpMetal -1.464e+15  6.126e+06 -2.389e+08 < 2e-16
## soporte_grpMural  3.251e+15  6.711e+07  4.844e+07 < 2e-16
## soporte_grpOtros  -8.363e-01  3.934e-01 -2.126e+00 0.033532
## soporte_grpTabla/Panel 6.109e-01  1.643e-01  3.719e+00 0.000200
## tecnicamixta    1.167e+00  4.718e-01  2.474e+00 0.013375
## tecnicaotras   -1.255e+00  9.861e-01 -1.273e+00 0.203130
## sop_montajesi    8.506e-01  2.454e-01  3.467e+00 0.000527
## seriesi         3.333e-01  9.881e-02  3.373e+00 0.000744
## tam_catmediano:soporte_grpMetal -1.172e+15  1.352e+07 -8.672e+07 < 2e-16
## tam_catgrande:soporte_grpMetal -3.365e+15  3.923e+07 -8.579e+07 < 2e-16
## tam_catmediano:soporte_grpMural -6.346e+15  6.985e+07 -9.086e+07 < 2e-16
## tam_catgrande:soporte_grpMural -3.414e+15  6.803e+07 -5.018e+07 < 2e-16
## tam_catmediano:soporte_grpOtros -1.902e+16  1.268e+07 -1.500e+09 < 2e-16
## tam_catgrande:soporte_grpOtros -1.689e+15  4.745e+07 -3.559e+07 < 2e-16
## tam_catmediano:soporte_grpTabla/Panel -1.994e+15  3.527e+06 -5.654e+08 < 2e-16
## tam_catgrande:soporte_grpTabla/Panel -1.393e+15  6.458e+06 -2.157e+08 < 2e-16
##
## (Intercept)          ***
## fecha_est            ***
## log_ancho            .
## tam_catmediano
## tam_catgrande        ***
## orientacionhorizontal ***
## soporte_grpMetal      ***
## soporte_grpMural      ***
## soporte_grpOtros      *
## soporte_grpTabla/Panel ***
## tecnicamixta          *
## tecnicaotras
## sop_montajesi         ***
## seriesi               ***
## tam_catmediano:soporte_grpMetal ***
## tam_catgrande:soporte_grpMetal ***
```

```
## tam_catmediano:soporte_grpMural      ***
## tam_catgrande:soporte_grpMural      ***
## tam_catmediano:soporte_grpOtros      ***
## tam_catgrande:soporte_grpOtros      ***
## tam_catmediano:soporte_grpTabla/Panel ***
## tam_catgrande:soporte_grpTabla/Panel ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 9122.7  on 6980  degrees of freedom
## AIC:  9166.7
##
## Type of estimator: AS_mixed (mixed bias-reducing adjusted score equations)
## Number of Fisher Scoring iterations: 100
```

En un primer intento nuestro modelo no alcanzó convergencia, lo que se traduce en estimaciones inestables, de manera que se procede a ajustar parámetros de control de las iteraciones:

```
library(brglm2)

ctrl <- brglm2::brglmControl(
  maxit = 2000,          # + iteraciones (default 100)
  slowit = 0.5,          # pasos más pequeños
  response_adjustment = 0.5 # arranque con ajuste tipo 0.5 en binomial
)

m_firth2 <- glm(
  formula(m_completo),
  data = df,
  family = binomial("logit"),
  method = "brglmFit",
  type = "AS_mean",
  control = ctrl
)

cat("Convergencia:", m_firth2$converged)
```

```
## Convergencia: TRUE
```

```
cat("\nNúmero de iteraciones:", m_firth2$iter)
```

```
##
## Número de iteraciones: 24
```

```
summary(m_firth2)
```

```
##
## Call:
## glm(formula = formula(m_completo), family = binomial("logit"),
```

```

##      data = df, control = ctrl, method = "brglmFit", type = "AS_mean")
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -0.9587   -0.5496   -0.4409   -0.3547    2.4983
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -7.5188015   0.6902160 -10.893 < 2e-16
## fecha_est                      0.0027539   0.0003714   7.415 1.21e-13
## log_ancho                      0.0294263   0.0232197   1.267 0.205048
## tam_catmediano                 0.0643331   0.1299291   0.495 0.620501
## tam_catgrande                 0.9929847   0.1154692   8.600 < 2e-16
## orientacionhorizontal          0.2860374   0.0765986   3.734 0.000188
## soporte_grpMetal              0.0189324   0.3747131   0.051 0.959704
## soporte_grpMural              1.7860819   2.3433219   0.762 0.445940
## soporte_grpOtros             -0.7732117   0.3885090  -1.990 0.046569
## soporte_grpTabla/Panel        0.5866802   0.1614587   3.634 0.000279
## tecnicamixta                  0.4170149   0.3341610   1.248 0.212051
## tecnicaotras                 -0.0504376   0.3820347  -0.132 0.894966
## sop_montajesi                 0.7901158   0.2314649   3.414 0.000641
## seriesi                      0.2797309   0.0937933   2.982 0.002860
## tam_catmediano:soporte_grpMetal -1.7264476   1.4984077  -1.152 0.249244
## tam_catgrande:soporte_grpMetal -0.2107848   1.7900398  -0.118 0.906262
## tam_catmediano:soporte_grpMural -2.3924557   2.4965511  -0.958 0.337909
## tam_catgrande:soporte_grpMural -2.6305508   2.3924610  -1.100 0.271543
## tam_catmediano:soporte_grpOtros -1.9023583   1.5254641  -1.247 0.212372
## tam_catgrande:soporte_grpOtros -0.2635839   1.9476521  -0.135 0.892348
## tam_catmediano:soporte_grpTabla/Panel 0.2242579   0.2499618   0.897 0.369629
## tam_catgrande:soporte_grpTabla/Panel -0.7513967   0.3711931  -2.024 0.042942
##
## (Intercept)                  ***
## fecha_est                    ***
## log_ancho
## tam_catmediano
## tam_catgrande                ***
## orientacionhorizontal        ***
## soporte_grpMetal
## soporte_grpMural
## soporte_grpOtros             *
## soporte_grpTabla/Panel       ***
## tecnicamixta
## tecnicaotras
## sop_montajesi                ***
## seriesi                      **
## tam_catmediano:soporte_grpMetal
## tam_catgrande:soporte_grpMetal
## tam_catmediano:soporte_grpMural
## tam_catgrande:soporte_grpMural
## tam_catmediano:soporte_grpOtros
## tam_catgrande:soporte_grpOtros
## tam_catmediano:soporte_grpTabla/Panel
## tam_catgrande:soporte_grpTabla/Panel *
## ---

```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4929.9  on 6980  degrees of freedom
## AIC:  4973.9
##
## Type of estimator: AS_mixed (mixed bias-reducing adjusted score equations)
## Number of Fisher Scoring iterations: 24
```

Gracias a las especificaciones de control el modelo ahora sí ha alcanzado convergencia. Todavía observamos algunos SE grandes, lo que indica estimaciones débiles para esos niveles, pero una mejora considerable frente al modelo anterior sin este ajuste Firth.

Veamos una comparación de SE, y otro test de separación, para el modelo sin ajuste 'm_completo' y el modelo ajustado 'm_firth2':

```
se_ml <- summary(m_completo)$coefficients[,2]
se_fi <- summary(m_firth2)$coefficients[,2]

cat("\nTop 10 SE (MLE):\n")
```

```
##
## Top 10 SE (MLE):
```

```
print(head(sort(se_ml, decreasing=TRUE), 10))
```

```
## tam_catmediano:soporte_grpMural tam_catgrande:soporte_grpMural
##      1455.3979311      1455.3976728
##      soporte_grpMural tam_catgrande:soporte_grpOtros
##      1455.3975872      1022.7042152
## tam_catgrande:soporte_grpMetal tam_catmediano:soporte_grpOtros
##      838.9572783      272.6487034
## tam_catmediano:soporte_grpMetal (Intercept)
##      260.4130626      0.6938752
##      soporte_grpOtros tecnicaotras
##      0.3956836      0.3932892
```

```
cat("\nTop 10 SE (Firth/AS):\n")
```

```
##
## Top 10 SE (Firth/AS):
```

```
print(head(sort(se_fi, decreasing=TRUE), 10))
```

```
## tam_catmediano:soporte_grpMural tam_catgrande:soporte_grpMural
##      2.4965511      2.3924610
##      soporte_grpMural tam_catgrande:soporte_grpOtros
##      2.3433219      1.9476521
## tam_catgrande:soporte_grpMetal tam_catmediano:soporte_grpOtros
```

```
##              1.7900398              1.5254641
## tam_catmediano:soporte_grpMetal      (Intercept)
##              1.4984077              0.6902160
##              soporte_grpOtros        tecnicaotras
##              0.3885090              0.3820347
```

```
cat("Separación:\n")
```

```
## Separación:
```

```
m_firth2_df2 <- update(m_firth2, data = df2)

X <- model.matrix(m_firth2_df2)
y <- model.response(model.frame(m_firth2_df2))
ds <- detectseparation::detect_separation(X, y)
```

```
## Warning in detectseparation::detect_separation(X, y): 'detect_separation' has
## been developed for use with binomial-response GLMs
```

```
print(ds)
```

```
## Implementation: ROI | Solver: lpsolve
## Separation: TRUE
## Existence of maximum likelihood estimates
##              (Intercept)              fecha_est
##              0              0
##              log_ancho              tam_catmediano
##              0              0
##              tam_catgrande              orientacionhorizontal
##              0              0
##              soporte_grpMetal              soporte_grpMural
##              0              Inf
##              soporte_grpOtros              soporte_grpTabla/Panel
##              0              0
##              tecnicamixta              tecnicaotras
##              0              0
##              sop_montajesi              seriesi
##              0              0
##      tam_catmediano:soporte_grpMetal      tam_catgrande:soporte_grpMetal
##      -Inf      -Inf
##      tam_catmediano:soporte_grpMural      tam_catgrande:soporte_grpMural
##      -Inf      -Inf
##      tam_catmediano:soporte_grpOtros      tam_catgrande:soporte_grpOtros
##      -Inf      -Inf
## tam_catmediano:soporte_grpTabla/Panel tam_catgrande:soporte_grpTabla/Panel
##      0      0
## 0: finite value, Inf: infinity, -Inf: -infinity
```

Podemos ver como la presencia de separación no se ha eliminado, cual era esperable ya que la reducción de sesgo Firth no cambia la estructura de separación de los dato, pero sí asegura que nuestro modelo es robusto frente a ella. Podemos comprobarlo viendo la significativa reducción de los SE.

Colinealidad

Una vez controlada la separación observamos volvamos a realizar las pruebas de colinealidad:

```
library(car)
vif(m_firth2)

## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif

##              GVIF Df GVIF^(1/(2*Df))
## fecha_est      1.671291  1      1.292784
## log_ancho      1.284864  1      1.133518
## tam_cat        2.146515  2      1.210413
## orientacion    1.063647  1      1.031332
## soporte_grp    99.891232  4      1.778038
## tecnica        1.575092  2      1.120280
## sop_montaje     1.675569  1      1.294438
## serie          1.108229  1      1.052725
## tam_cat:soporte_grp 86.923917  8      1.321893
```

Una vez manejada la separación, se elimina la existencia de colinealidad grave y los términos afectados presentan ahora colinealidad moderada que no refleja una preocupación real. Dado uestro objetivo descriptivo, se mantiene el modelo con todos sus términos.

Resumen

Finalmente se concluye con el modelo final de tipo regresión logística incluyendo reducción de sesgo Firth.

```
ctrl <- brglm2::brglmControl(
  maxit = 2000,
  slowit = 0.5,
  response_adjustment = 0.5
)

m_principal <- glm(
  formula(exito ~
    fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie +
    soporte_grp:tam_cat),
  data = df,
  family = binomial("logit"),
  method = "brglmFit",
  type = "AS_mean",
  control = ctrl
)

summary(m_principal)

##
## Call:
## glm(formula = formula(exito ~ fecha_est + log_ancho + tam_cat +
```

```

##      orientacion + soporte_grp + tecnica + sop_montaje + serie +
##      soporte_grp:tam_cat), family = binomial("logit"), data = df,
##      control = ctrl, method = "brglmFit", type = "AS_mean")
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -0.9587   -0.5496   -0.4409   -0.3547    2.4983
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.5188015   0.6902160 -10.893 < 2e-16
## fecha_est         0.0027539   0.0003714   7.415 1.21e-13
## log_ancho         0.0294263   0.0232197   1.267 0.205048
## tam_catmediano    0.0643331   0.1299291   0.495 0.620501
## tam_catgrande     0.9929847   0.1154692   8.600 < 2e-16
## orientacionhorizontal 0.2860374   0.0765986   3.734 0.000188
## soporte_grpMetal   0.0189324   0.3747131   0.051 0.959704
## soporte_grpMural   1.7860819   2.3433219   0.762 0.445940
## soporte_grpOtros  -0.7732117   0.3885090  -1.990 0.046569
## soporte_grpTabla/Panel 0.5866802   0.1614587   3.634 0.000279
## tecnicamixta      0.4170149   0.3341610   1.248 0.212051
## tecnicaotras     -0.0504376   0.3820347  -0.132 0.894966
## sop_montajesi     0.7901158   0.2314649   3.414 0.000641
## seriesi          0.2797309   0.0937933   2.982 0.002860
## tam_catmediano:soporte_grpMetal -1.7264476   1.4984077  -1.152 0.249244
## tam_catgrande:soporte_grpMetal -0.2107848   1.7900398  -0.118 0.906262
## tam_catmediano:soporte_grpMural -2.3924557   2.4965511  -0.958 0.337909
## tam_catgrande:soporte_grpMural -2.6305508   2.3924610  -1.100 0.271543
## tam_catmediano:soporte_grpOtros -1.9023583   1.5254641  -1.247 0.212372
## tam_catgrande:soporte_grpOtros -0.2635839   1.9476521  -0.135 0.892348
## tam_catmediano:soporte_grpTabla/Panel 0.2242579   0.2499618   0.897 0.369629
## tam_catgrande:soporte_grpTabla/Panel -0.7513967   0.3711931  -2.024 0.042942
##
## (Intercept)          ***
## fecha_est            ***
## log_ancho
## tam_catmediano
## tam_catgrande          ***
## orientacionhorizontal ***
## soporte_grpMetal
## soporte_grpMural
## soporte_grpOtros      *
## soporte_grpTabla/Panel ***
## tecnicamixta
## tecnicaotras
## sop_montajesi          ***
## seriesi                **
## tam_catmediano:soporte_grpMetal
## tam_catgrande:soporte_grpMetal
## tam_catmediano:soporte_grpMural
## tam_catgrande:soporte_grpMural
## tam_catmediano:soporte_grpOtros
## tam_catgrande:soporte_grpOtros
## tam_catmediano:soporte_grpTabla/Panel

```

```
## tam_catgrande:soporte_grpTabla/Panel *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 5173.2  on 7001  degrees of freedom
## Residual deviance: 4929.9  on 6980  degrees of freedom
## AIC:  4973.9
##
## Type of estimator: AS_mixed (mixed bias-reducing adjusted score equations)
## Number of Fisher Scoring iterations: 24
```

Validación

```
summary(m_principal)$null.deviance
```

```
## [1] 5173.222
```

```
summary(m_principal)$deviance
```

```
## [1] 4929.932
```

```
nobs(m_principal)
```

```
## [1] 7002
```

```
# Pseudo R2 de McFadden
1 - summary(m_principal)$deviance / summary(m_principal)$null.deviance
```

```
## [1] 0.04702875
```

Observamos que la deviance es menor que la del modelo nulo, lo que nos está indicando que al introducir las variables explicativas tenemos una mejora del modelo. Las covariables introducidas aportan información para estimar la probabilidad. Por lo que el modelo está bien ajustado

Capacidad discriminativa

```
# Probabilidades predichas
prob_principal <- predict(m_principal, type = "response")

# Curva ROC
roc_principal <- roc(df$exito, prob_principal)
```

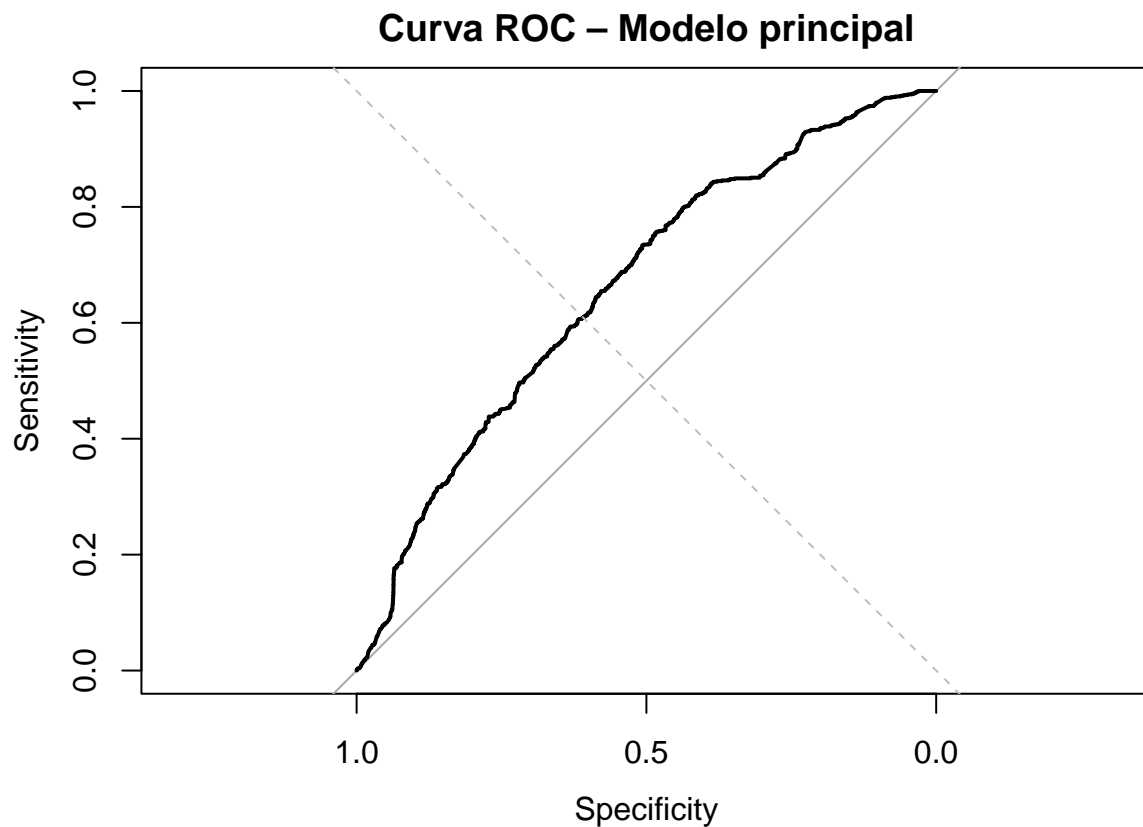
```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
# AUC
auc(roc_principal)
```

```
## Area under the curve: 0.66
```

```
# Gráfico ROC
plot(roc_principal,
     main = "Curva ROC - Modelo principal",
     col = "black",
     lwd = 2)
abline(a = 0, b = 1, lty = 2, col = "gray")
```



```
library(dplyr)

calibracion <- data.frame(
  obs = df$exito,
  pred = prob_principal
) %>%
  mutate(decil = ntile(pred, 10)) %>%
  group_by(decil) %>%
  summarise(
    prob_predicha = mean(pred),
    prob_observada = mean(obs),
    n = n()
```

```
)

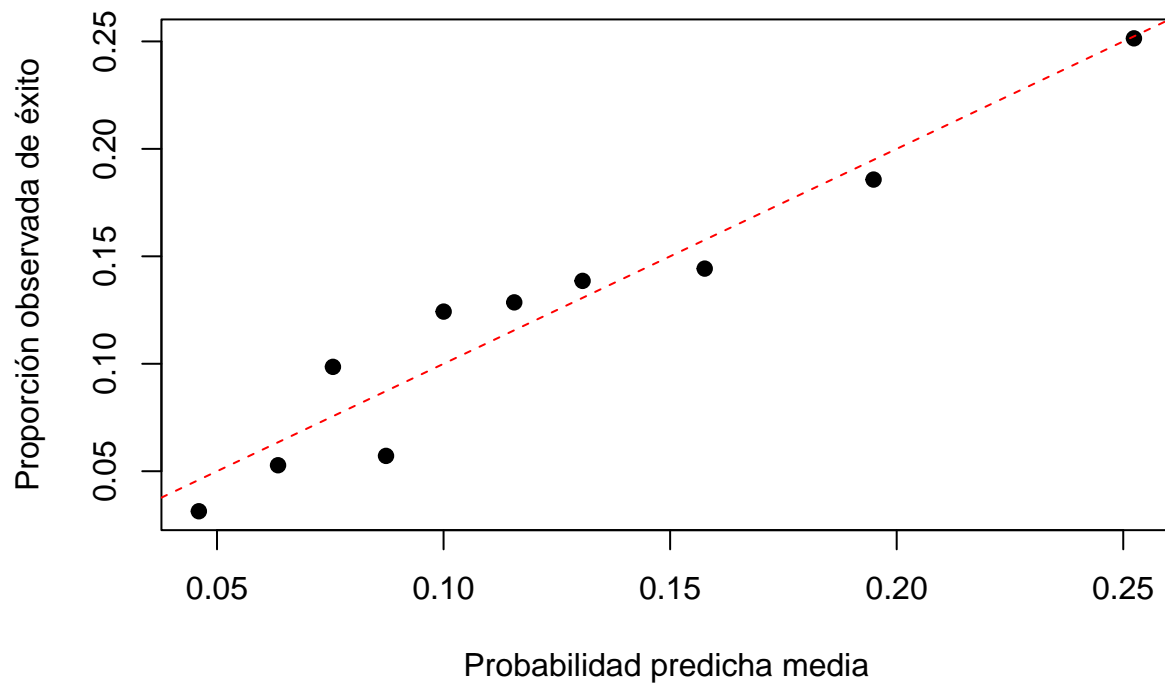
calibracion

## # A tibble: 10 x 4
##   decil prob_predicha prob_observada     n
##   <int>      <dbl>      <dbl> <int>
## 1     1      0.0460      0.0314   701
## 2     2      0.0635      0.0528   701
## 3     3      0.0756      0.0986   700
## 4     4      0.0873      0.0571   700
## 5     5      0.100      0.124   700
## 6     6      0.116      0.129   700
## 7     7      0.131      0.139   700
## 8     8      0.158      0.144   700
## 9     9      0.195      0.186   700
## 10    10      0.252      0.251   700

plot(calibracion$prob_predicha,
     calibracion$prob_observada,
     xlab = "Probabilidad predicha media",
     ylab = "Proporción observada de éxito",
     main = "Calibración del modelo principal",
     pch = 19)

abline(0, 1, lty = 2, col = "red")
```

Calibración del modelo principal



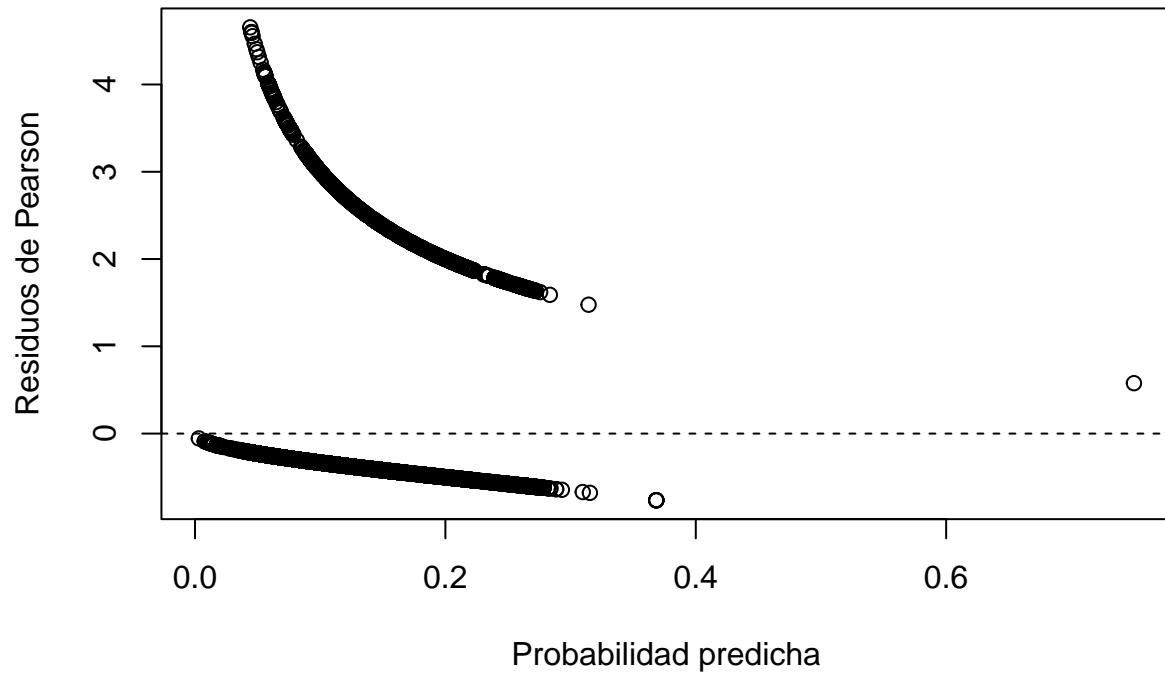
Residuos del modelo

```
res_pearson <- residuals(m_principal, type = "pearson")
summary(res_pearson)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -0.763779 -0.403751 -0.319497 -0.005352 -0.254803  4.654661
```

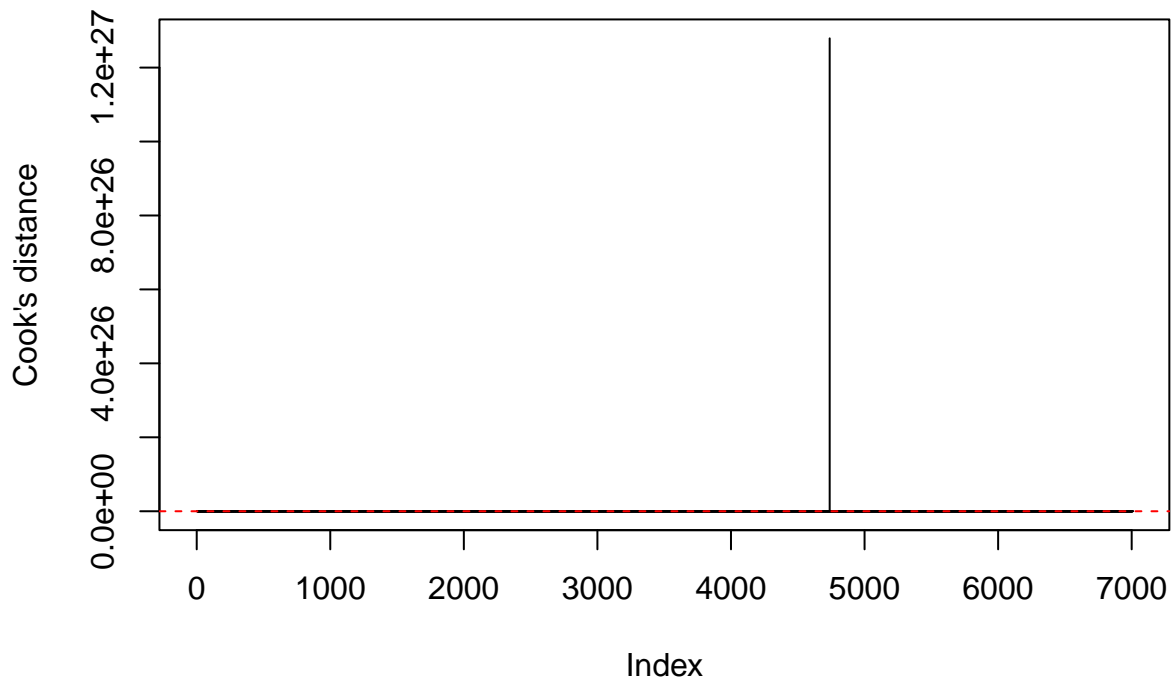
```
plot(prob_principal, res_pearson,
     xlab = "Probabilidad predicha",
     ylab = "Residuos de Pearson",
     main = "Residuos de Pearson vs probabilidad predicha")
abline(h = 0, lty = 2)
```

Residuos de Pearson vs probabilidad predicha



```
cook <- cooks.distance(m_principal)
plot(cook, type = "h",
     main = "Distancia de Cook - Modelo principal",
     ylab = "Cook's distance")
abline(h = 4 / length(cook), lty = 2, col = "red")
```

Distancia de Cook – Modelo principal



```
which(cook > 4 / length(cook))
```

```
## 10 14 33 34 62 106 111 120 129 140 143 144 153 154 158 159
## 9 13 32 33 61 105 110 119 128 139 142 143 151 152 156 157
## 162 174 187 195 196 206 209 210 211 212 222 225 273 302 313 314
## 160 172 185 193 194 204 207 208 209 210 220 223 271 300 311 312
## 323 334 343 361 376 407 418 419 453 473 508 511 546 550 623 638
## 321 332 341 359 374 405 416 417 451 471 506 509 544 548 620 635
## 645 654 656 675 694 729 747 784 785 790 810 822 898 921 986 1002
## 642 651 653 672 691 726 744 781 782 787 807 819 894 917 982 998
## 1003 1012 1053 1071 1094 1124 1136 1184 1190 1192 1193 1194 1204 1207 1220 1246
## 999 1008 1049 1067 1089 1119 1131 1179 1185 1187 1188 1189 1199 1202 1215 1241
## 1248 1273 1275 1298 1302 1311 1317 1334 1356 1363 1415 1417 1420 1431 1452 1479
## 1243 1267 1269 1291 1295 1304 1310 1327 1349 1356 1408 1410 1413 1424 1444 1471
## 1520 1611 1668 1676 1677 1692 1700 1776 1777 1794 1802 1825 1829 1843 1848 1849
## 1512 1603 1660 1668 1669 1684 1692 1768 1769 1786 1794 1817 1821 1835 1840 1841
## 1858 1897 1941 1946 1989 1990 2000 2052 2074 2097 2103 2127 2135 2158 2159 2164
## 1849 1888 1932 1937 1980 1981 1991 2043 2065 2088 2094 2118 2126 2149 2150 2155
## 2171 2203 2219 2224 2231 2234 2249 2261 2281 2293 2295 2673 2677 2689 2702 2716
## 2162 2194 2210 2215 2222 2225 2240 2252 2272 2284 2286 2663 2667 2679 2692 2706
## 2722 2723 2724 2825 2842 2860 2875 2918 2952 3097 3099 3100 3102 3120 3144 3166
## 2712 2713 2714 2812 2829 2845 2860 2903 2937 3082 3084 3085 3087 3105 3129 3151
## 3167 3289 3290 3292 3347 3513 3743 3744 3745 3746 3754 3779 3895 3896 3897 3898
## 3152 3274 3275 3277 3332 3498 3725 3726 3727 3728 3736 3761 3877 3878 3879 3880
## 3922 3941 3943 4284 4335 4423 4714 4720 4722 4740 4743 4747 4748 4750 4754 4755
```



```
## 3904 3923 3925 4263 4313 4401 4691 4697 4699 4717 4720 4724 4725 4727 4731 4732
## 4756 4757 4758 4759 4760 4761 4762 4763 4764 4765 4766 4770 4855 5036 5102 5162
## 4733 4734 4735 4736 4737 4738 4739 4740 4741 4742 4743 4747 4831 5010 5076 5135
## 5224 5410 5418 5422 5492 5494 5497 5499 5503 5505 5514 5515 5516 5520 5564 5638
## 5197 5383 5391 5395 5465 5467 5470 5472 5476 5478 5487 5488 5489 5493 5537 5611
## 5646 5681 5683 5688 5692 5693 5694 5722 5781 5857 5863 5866 5899 5905 5907 5928
## 5619 5653 5655 5660 5664 5665 5666 5694 5753 5828 5834 5837 5870 5876 5878 5899
## 5958 5973 5977 5978 5987 5988 5989 5991 6087 6221 6282 6325 6340 6341 6362 6379
## 5929 5944 5948 5949 5958 5959 5960 5962 6058 6192 6253 6296 6311 6312 6333 6350
## 6419 6502 6508 6519 6597 6648 6661 6684 6711 6730 6738 6739 6788 6894
## 6390 6472 6478 6489 6567 6617 6630 6653 6680 6699 6707 6708 6756 6861
```

El análisis de los residuos nos indica que no tenemos patrones ni desviaciones que nos sugieran que el modelo esta mal especificado. Con las medidas de influencia observamos que no hay observaciones individuales que ejerzan una gran influencia sobre la estimación de los parametros.

Los resultados nos presentan estabilidad y robustez en las estimaciones del modelo.

Multicolinealidad

```
vif(m_principal)
```

```
## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif
```

```
##              GVIF Df GVIF^(1/(2*Df))
## fecha_est      1.671291  1      1.292784
## log_ancho      1.284864  1      1.133518
## tam_cat        2.146515  2      1.210413
## orientacion    1.063647  1      1.031332
## soporte_grp   99.891232  4      1.778038
## tecnica        1.575092  2      1.120280
## sop_montaje    1.675569  1      1.294438
## serie          1.108229  1      1.052725
## tam_cat:soporte_grp 86.923917  8      1.321893
```

Observamos que las variables explicativas no presentan problemas de multicolinealidad.

El modelo `m_principal` se considera adecuado y fiable para predecir la probabilidad de éxito.

Modelos alternativos

En esta sección se proponen estructuras de modelos que fueron descartadas como opciones a modelo principal pero que por interés interpretativo se ajustan igualmente. Primero presentaremos las propuestas y después serán reestimadas utilizando reducción de sesgo Firth cuando sea necesario.

Modelo simplificado

Dada la complejidad y posible separación debida a la interacción “soporte_grp x tam_cat”, aunque ya se controló adecuadamente, se ha decidido estudiar también la posibilidad del modelo simplificado sin interacciones:

```
m_reducido <- glm(
  exito ~
    fecha_est + log_ancho +
    tam_cat + orientacion +
    soporte_grp + tecnica + sop_montaje +
    serie,
  data = df,
  family = binomial(link = "logit")
)
```

Variable tema

Dado el interés interpretativo de tema, y visto que sí aporta información adicional pero incrementa fuertemente la complejidad (muchos niveles), su inclusión se analiza en esta sección de modelos alternativos siguiendo las siguientes estrategias:

- 1) Un modelo con tema ajustado por el conjunto completo de covariables del modelo principal
- 2) Un modelo con tema sin el bloque de material/técnica/montaje para estimar una asociación más global
- 3) Adicionalmente, una versión parsimoniosa usando log(area) lineal para cada caso anterior

De esta manera se explorará el efecto de tema tanto controlando las covariables de técnica/soporte/montaje como sin ellas para, de esta manera, poder obtener también una estimación global de su asociación con la respuesta. Esta decisión viene dada por consecuencia de nuestros objetivos e hipótesis principales. En un inicio se planteó la posibilidad de que la iconografía de la pintura podía estar relacionada con el cumplimiento de la razón áurea, por esta razón queremos explorar el efecto global de esta variable “tema”, aun sabiendo que el efecto de técnica y soporte pudiera estar incluido en la relación, ya que este modelo no pretende estimar un efecto causal directo sino capturar asociación total que puede incluir diferencias mediadas por técnica/soporte. Por otra parte. Con el objetivo de no sesgar totalmente la interpretación, también se estudiará el efecto en conjunto con estas variables: “soporte_grp” y “tecncia”, las cuales también eran de interés. Finalmente se contempla la posibilidad de cambiar la especificación del tamaño por su versión continua lineal (log(area)) con el objetivo de simplemente controlar el efecto del tamaño, pero sin interés interpretativo en esta sección.

Ajsute de modelos

Procedemos a ajustar todos los modelos candidatos

```
# =====
# MODELOS ALTERNATIVOS CON TEMA (resumen compacto)
# =====
# 1) ESPECIFICACIÓN DE MODELOS (sin summary)
# =====
cat("\n===== \n1) ESPECIFICACION DE MODELOS \n===== \n")

##
## =====
## 1) ESPECIFICACION DE MODELOS
## =====

# Modelo final principal (sin tema, con Bloque 3, tamaño categórico)
m_final <- glm(
  exito ~ ns(fecha_est, 3) + log_ancho + tam_cat + orientacion +
```

```

    soporte_grp + tecnica + sop_montaje + serie,
    data = df, family = binomial(link = "logit")
)
cat("\n[m_final] creado: ajuste completo + tam_cat (SIN tema)\n")

##
## [m_final] creado: ajuste completo + tam_cat (SIN tema)

# Ajuste completo + tema (tam_cat)
mT_full_tam <- update(m_final, . ~ . + tema)
cat("[mT_full_tam] creado: ajuste completo + tam_cat (CON tema)\n")

## [mT_full_tam] creado: ajuste completo + tam_cat (CON tema)

# Modelo final equivalente con log_area (sin tema)
m_final_area <- glm(
  exito ~ ns(fecha_est, 3) + log_ancho + log_area + orientacion +
    soporte_grp + tecnica + sop_montaje + serie,
  data = df, family = binomial(link = "logit")
)
cat("[m_final_area] creado: ajuste completo + log_area (SIN tema)\n")

## [m_final_area] creado: ajuste completo + log_area (SIN tema)

# Ajuste completo + tema (log_area)
mT_full_area <- update(m_final_area, . ~ . + tema)
cat("[mT_full_area] creado: ajuste completo + log_area (CON tema)\n")

## [mT_full_area] creado: ajuste completo + log_area (CON tema)

# Sin Bloque 3 + tam_cat (sin tema)
m_noB3_tam <- glm(
  exito ~ ns(fecha_est, 3) + log_ancho + tam_cat + orientacion + serie,
  data = df, family = binomial(link = "logit")
)
cat("[m_noB3_tam] creado: SIN Bloque 3 + tam_cat (SIN tema)\n")

## [m_noB3_tam] creado: SIN Bloque 3 + tam_cat (SIN tema)

# Sin Bloque 3 + tema (tam_cat)
mT_noB3_tam <- update(m_noB3_tam, . ~ . + tema)
cat("[mT_noB3_tam] creado: SIN Bloque 3 + tam_cat (CON tema)\n")

## [mT_noB3_tam] creado: SIN Bloque 3 + tam_cat (CON tema)

# Sin Bloque 3 + log_area (sin tema)
m_noB3_area <- glm(
  exito ~ ns(fecha_est, 3) + log_ancho + log_area + orientacion + serie,
  data = df, family = binomial(link = "logit")
)
cat("[m_noB3_area] creado: SIN Bloque 3 + log_area (SIN tema)\n")

```

```
## [m_noB3_area] creado: SIN Bloque 3 + log_area (SIN tema)
```

```
# Sin Bloque 3 + tema (log_area)
mT_noB3_area <- update(m_noB3_area, . ~ . + tema)
cat("[mT_noB3_area] creado: SIN Bloque 3 + log_area (CON tema)\n")
```

```
## [mT_noB3_area] creado: SIN Bloque 3 + log_area (CON tema)
```

```
# =====
# 2) COMPARACIONES ANOVA (mínimas, aporte de informacion)
# =====
cat("\n===== \n2) COMPARACIONES (LRT - ANOVA) \n===== \n")
```

```
##
## =====
## 2) COMPARACIONES (LRT - ANOVA)
## =====
```

```
cat("\n[2.1] Aporte de tema (ajuste completo, tam_cat): m_final vs mT_full_tam\n")
```

```
##
## [2.1] Aporte de tema (ajuste completo, tam_cat): m_final vs mT_full_tam
```

```
print(anova(m_final, mT_full_tam, test = "Chisq"))
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ ns(fecha_est, 3) + log_ancho + tam_cat + orientacion +
## soporte_grp + tecnica + sop_montaje + serie
## Model 2: exito ~ ns(fecha_est, 3) + log_ancho + tam_cat + orientacion +
## soporte_grp + tecnica + sop_montaje + serie + tema
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6986      4940.8
## 2      6977      4913.6  9      27.22 0.001286 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("\n[2.2] Aporte de tema (ajuste completo, log_area): m_final_area vs mT_full_area\n")
```

```
##
## [2.2] Aporte de tema (ajuste completo, log_area): m_final_area vs mT_full_area
```

```
print(anova(m_final_area, mT_full_area, test = "Chisq"))
```

```
## Analysis of Deviance Table
##
## Model 1: exito ~ ns(fecha_est, 3) + log_ancho + log_area + orientacion +
## soporte_grp + tecnica + sop_montaje + serie
## Model 2: exito ~ ns(fecha_est, 3) + log_ancho + log_area + orientacion +
```

```
## soporte_grp + tecnica + sop_montaje + serie + tema
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6987      5001.1
## 2      6978      4974.3  9    26.863 0.001473 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n[2.3] Aporte de tema (SIN Bloque 3, tam_cat): m_noB3_tam vs mT_noB3_tam\n")

##
## [2.3] Aporte de tema (SIN Bloque 3, tam_cat): m_noB3_tam vs mT_noB3_tam

print(anova(m_noB3_tam, mT_noB3_tam, test = "Chisq"))

## Analysis of Deviance Table
##
## Model 1: exito ~ ns(fecha_est, 3) + log_ancho + tam_cat + orientacion +
##      serie
## Model 2: exito ~ ns(fecha_est, 3) + log_ancho + tam_cat + orientacion +
##      serie + tema
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6993      4971.0
## 2      6984      4942.6  9    28.394 0.0008193 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

cat("\n[2.4] Aporte de tema (SIN Bloque 3, log_area): m_noB3_area vs mT_noB3_area\n")

##
## [2.4] Aporte de tema (SIN Bloque 3, log_area): m_noB3_area vs mT_noB3_area

print(anova(m_noB3_area, mT_noB3_area, test = "Chisq"))

## Analysis of Deviance Table
##
## Model 1: exito ~ ns(fecha_est, 3) + log_ancho + log_area + orientacion +
##      serie
## Model 2: exito ~ ns(fecha_est, 3) + log_ancho + log_area + orientacion +
##      serie + tema
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      6994      5031.8
## 2      6985      5003.8  9      28 0.0009538 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para comenzar podemos ver que tema siempre resulta significativo en cuanto aporte de información, es decir, independientemente de como controlemos el tamaño y de si incluimos o no el BLOQUE 3 (tecnica/soporte/montaje), tema añade información. Este resultado avala nuestra intención de realizar esta reacción para interpretar la asociación de la iconografía con el cumplimiento de la razón aurea.

Comparación de modelos

Para decidir que modelo o modelos interpretar utilizaremos los criterios AIC/BIC. Se muestran sus valores para cada modelo, y una tabla de incrementos, de manera que sea legible.

```

# =====
# 3) TABLA AIC/BIC (todos los modelos)
# =====
cat("\n===== \n3) TABLA AIC / BIC \n===== \n")

##
## =====
## 3) TABLA AIC / BIC
## =====

mods <- list(
  m_final      = m_final,
  mT_full_tam  = mT_full_tam,
  m_final_area = m_final_area,
  mT_full_area = mT_full_area,
  m_noB3_tam   = m_noB3_tam,
  mT_noB3_tam  = mT_noB3_tam,
  m_noB3_area  = m_noB3_area,
  mT_noB3_area = mT_noB3_area
)

ic_tab <- data.frame(
  modelo = names(mods),
  df      = sapply(mods, function(m) attr(logLik(m), "df")),
  AIC     = sapply(mods, AIC),
  BIC     = sapply(mods, BIC),
  row.names = NULL
)

print(ic_tab[order(ic_tab$AIC), ])

```

```

##      modelo df      AIC      BIC
## 2  mT_full_tam 25 4963.570 5134.919
## 1      m_final 16 4972.789 5082.452
## 6  mT_noB3_tam 18 4978.623 5101.994
## 5   m_noB3_tam  9 4989.017 5050.703
## 4 mT_full_area 24 5022.257 5186.752
## 3 m_final_area 15 5031.120 5133.930
## 8 mT_noB3_area 17 5037.819 5154.336
## 7   m_noB3_area  8 5047.819 5102.650

```

Dado que pretendemos explorar la asociacion de tema tanto controlando por las covariables del BLoque 3 como sin ellas, seleccionaremos de cada una la que proporcione mejores valores de AIC/BIC.

```

# =====
# Dataframe de deltas (con tema - sin tema) para AIC y BIC
# =====

delta_ic <- data.frame(
  situacion = c(
    "Ajuste completo + tam_cat",
    "Ajuste completo + log_area",

```

```

    "Sin Bloque 3 + tam_cat",
    "Sin Bloque 3 + log_area"
  ),
  modelo_sin_tema = c("m_final", "m_final_area", "m_noB3_tam", "m_noB3_area"),
  modelo_con_tema = c("mT_full_tam", "mT_full_area", "mT_noB3_tam", "mT_noB3_area"),
  AIC_sin = c(AIC(m_final),      AIC(m_final_area),  AIC(m_noB3_tam),  AIC(m_noB3_area)),
  AIC_con = c(AIC(mT_full_tam),  AIC(mT_full_area),  AIC(mT_noB3_tam), AIC(mT_noB3_area)),
  BIC_sin = c(BIC(m_final),      BIC(m_final_area),  BIC(m_noB3_tam),  BIC(m_noB3_area)),
  BIC_con = c(BIC(mT_full_tam),  BIC(mT_full_area),  BIC(mT_noB3_tam), BIC(mT_noB3_area)),
  stringsAsFactors = FALSE
)

# Deltas: (con tema - sin tema)
delta_ic$delta_AIC <- delta_ic$AIC_con - delta_ic$AIC_sin
delta_ic$delta_BIC <- delta_ic$BIC_con - delta_ic$BIC_sin

# (Opcional) Redondeo para informe
delta_ic$AIC_sin <- round(delta_ic$AIC_sin, 3)
delta_ic$AIC_con <- round(delta_ic$AIC_con, 3)
delta_ic$BIC_sin <- round(delta_ic$BIC_sin, 3)
delta_ic$BIC_con <- round(delta_ic$BIC_con, 3)
delta_ic$delta_AIC <- round(delta_ic$delta_AIC, 3)
delta_ic$delta_BIC <- round(delta_ic$delta_BIC, 3)

# Mostrar tabla ordenada (por mejora AIC: más negativo = mejor)
delta_ic <- delta_ic[order(delta_ic$delta_AIC), ]

delta_ic

```

```

##              situacion modelo_sin_tema modelo_con_tema  AIC_sin  AIC_con
## 3      Sin Bloque 3 + tam_cat      m_noB3_tam      mT_noB3_tam 4989.017 4978.623
## 4      Sin Bloque 3 + log_area      m_noB3_area      mT_noB3_area 5047.819 5037.819
## 1  Ajuste completo + tam_cat          m_final      mT_full_tam 4972.789 4963.570
## 2  Ajuste completo + log_area      m_final_area      mT_full_area 5031.120 5022.257
##      BIC_sin  BIC_con delta_AIC delta_BIC
## 3 5050.703 5101.994   -10.394   51.291
## 4 5102.650 5154.336   -10.000   51.685
## 1 5082.452 5134.919    -9.219   52.466
## 2 5133.930 5186.752    -8.863   52.822

```

```

delta_ic[c("situacion", "delta_AIC", "delta_BIC")]

```

```

##              situacion delta_AIC delta_BIC
## 3      Sin Bloque 3 + tam_cat   -10.394   51.291
## 4      Sin Bloque 3 + log_area   -10.000   51.685
## 1  Ajuste completo + tam_cat    -9.219   52.466
## 2  Ajuste completo + log_area    -8.863   52.822

```

Podemos ver que en ambos casos, con la inclusión del Bloque 3 (tecnica/soprote/montaje) o sin ella, los modelos preferibles son aquellos con una especificación categórica del tamaño, es decir, utilizando “tam_cat” en vez de “area”. Por esta razón queda eliminada la opción alternativa que se había planteado en relación al tamaño.

Los modelos finales que se utilizarán para interpretar la asociación de “tema” son ‘mT_full_tam’ y ‘mT_noB3_tam’.

Modelos alternativos finales

Modelos alternativos finales:

```
# modelo reducido (sin interacciones)
mA1 <- m_reducido

# tema ajustado
mA2 <- mT_full_tam

# tema global
mA3 <- mT_noB3_tam
```

Para evaluar la el posibles ajuste mediante Firth, estudiaremos la separación de los 3 modelos:

```
cat("Separación para mA1 (sin interacciones):\n")
```

```
## Separación para mA1 (sin interacciones):
```

```
mA1_df2 <- update(mA1, data = df2)

X <- model.matrix(mA1_df2)
y <- model.response(model.frame(mA1_df2))
ds <- detectseparation::detect_separation(X, y)
```

```
## Warning in detectseparation::detect_separation(X, y): ‘detect_separation’ has
## been developed for use with binomial-response GLMs
```

```
print(ds)
```

```
## Implementation: ROI | Solver: lpsolve
## Separation: FALSE
## Existence of maximum likelihood estimates
##      (Intercept)      fecha_est      log_ancho
##           0           0           0
##      tam_catmediano      tam_catgrande      orientacionhorizontal
##           0           0           0
##      soporte_grpMetal      soporte_grpMural      soporte_grpOtros
##           0           0           0
##      soporte_grpTabla/Panel      tecnicamixta      tecnicaotras
##           0           0           0
##      sop_montajesi      seriesi
##           0           0
## 0: finite value, Inf: infinity, -Inf: -infinity
```

```
cat("Separación para mA2 (tema ajustado):\n")
```

```
## Separación para mA2 (tema ajustado):
```



```
mA2_df2 <- update(mA2, data = df2)
```

```
X <- model.matrix(mA2_df2)
```

```
y <- model.response(model.frame(mA2_df2))
```

```
ds <- detectseparation::detect_separation(X, y)
```

```
## Warning in detectseparation::detect_separation(X, y): 'detect_separation' has  
## been developed for use with binomial-response GLMs
```

```
print(ds)
```

```
## Implementation: ROI | Solver: lpsolve
```

```
## Separation: FALSE
```

```
## Existence of maximum likelihood estimates
```

```
##           (Intercept)      ns(fecha_est, 3)1      ns(fecha_est, 3)2  
##                0                0                0  
##      ns(fecha_est, 3)3      log_ancho      tam_catmediano  
##                0                0                0  
##      tam_catgrande  orientacionhorizontal  soporte_grpMetal  
##                0                0                0  
##      soporte_grpMural      soporte_grpOtros  soporte_grpTabla/Panel  
##                0                0                0  
##      tecnicamixta      tecnicaotras      sop_montajesi  
##                0                0                0  
##      seriesi      temabodegon_floral      temacaza_animales  
##                0                0                0  
##      temahistoria_allegoria      temamitologia      temaotros  
##                0                0                0  
##      temapaisaje_lugares      temaproceso_obra      temaretrato_corte  
##                0                0                0  
##      temavida_cotidiana  
##                0
```

```
## 0: finite value, Inf: infinity, -Inf: -infinity
```

```
cat("Separación para mA3 (tema global):\n")
```

```
## Separación para mA3 (tema global):
```

```
mA3_df2 <- update(mA3, data = df2)
```

```
X <- model.matrix(mA3_df2)
```

```
y <- model.response(model.frame(mA3_df2))
```

```
ds <- detectseparation::detect_separation(X, y)
```

```
## Warning in detectseparation::detect_separation(X, y): 'detect_separation' has  
## been developed for use with binomial-response GLMs
```

```
print(ds)
```

```
## Implementation: ROI | Solver: lpsolve
## Separation: FALSE
## Existence of maximum likelihood estimates
##      (Intercept)      ns(fecha_est, 3)1      ns(fecha_est, 3)2
##              0              0              0
##      ns(fecha_est, 3)3      log_ancho      tam_catmediano
##              0              0              0
##      tam_catgrande orientacionhorizontal      seriesi
##              0              0              0
##      temabodegon_floral      temacaza_animales temahistoria_allegoria
##              0              0              0
##      temamitologia      temaotros      temapaisaje_lugares
##              0              0              0
##      temaproceso_obra      temaretrato_corte      temavida_cotidiana
##              0              0              0
## 0: finite value, Inf: infinity, -Inf: -infinity
```

Vemos que no existe presencia de separación en ninguna de los modelos por lo que continuamos, para ellos, mediante el enlace logit sin necesidad de ajuste adicional

Validación

Validacion modelos alternativos,

```
prob_mT_full <- predict(mT_full_tam, type = "response")
prob_mT_noB3 <- predict(mT_noB3_tam, type = "response")
```

```
c(mT_full = summary(mT_full_tam)$deviance,
  mT_noB3 = summary(mT_noB3_tam)$deviance)
```

```
## mT_full mT_noB3
## 4913.570 4942.623
```

```
pseudoR2 <- function(m) {
  1 - summary(m)$deviance / summary(m)$null.deviance
}
```

```
c(
  mT_full = pseudoR2(mT_full_tam),
  mT_noB3 = pseudoR2(mT_noB3_tam)
)
```

```
## mT_full mT_noB3
## 0.05019149 0.04457542
```

Observamos que ambos modelos alternativos presentan valores de desviación y pseudo- R^2 similares, lo que indica que no alteran la capacidad del modelo. Son prácticamente iguales por lo que la mejora es muy pequeña.

```
AIC(mT_full_tam, mT_noB3_tam)
```

```
##      df      AIC
## mT_full_tam 25 4963.570
## mT_noB3_tam 18 4978.623
```

```
BIC(mT_full_tam, mT_noB3_tam)
```

```
##           df      BIC
## mT_full_tam 25 5134.919
## mT_noB3_tam 18 5101.994
```

Con los valores AIC i BIC detectamos lo mismo, que el modelo `mT_full_tam` es ligeramente mejor que el modelo `mT_noB3_tam`.

Capacidad discriminativa

```
roc_mT_full <- roc(df$exito, prob_mT_full)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
roc_mT_noB3 <- roc(df$exito, prob_mT_noB3)
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
auc(roc_mT_full)
```

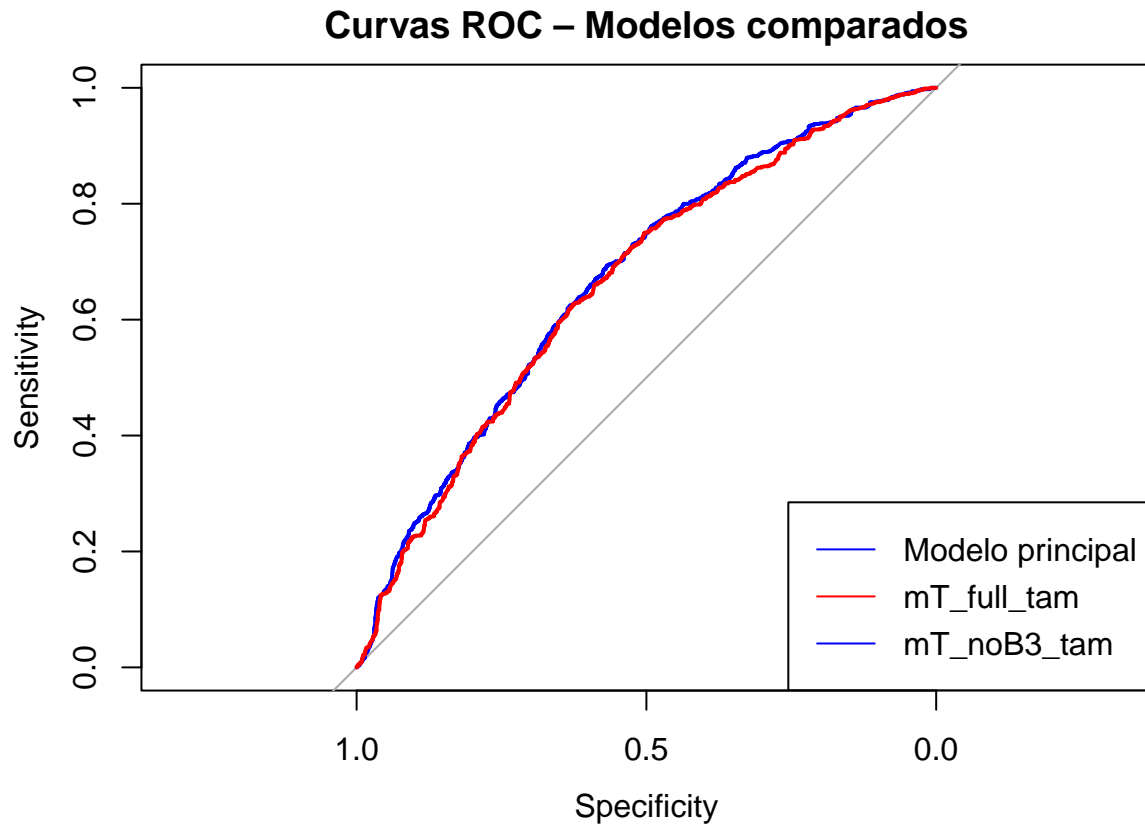
```
## Area under the curve: 0.6686
```

```
auc(roc_mT_noB3)
```

```
## Area under the curve: 0.6609
```

```
plot(roc_mT_full, col = "blue", lwd = 2, main = "Curvas ROC - Modelos comparados")
plot(roc_mT_noB3, col = "red", lwd = 2, add = TRUE)
```

```
legend("bottomright",
      legend = c("Modelo principal", "mT_full_tam", "mT_noB3_tam"),
      col = c("blue", "red"),
      lwd = 1)
```



Con la grafica ROC vemos que ambos modelos se sobrepone a la media, lo que nos sugiere que no introduzcamos variables adicional ni excluyamos variables, ya que no generaran mejoras apreciables para el modelo.

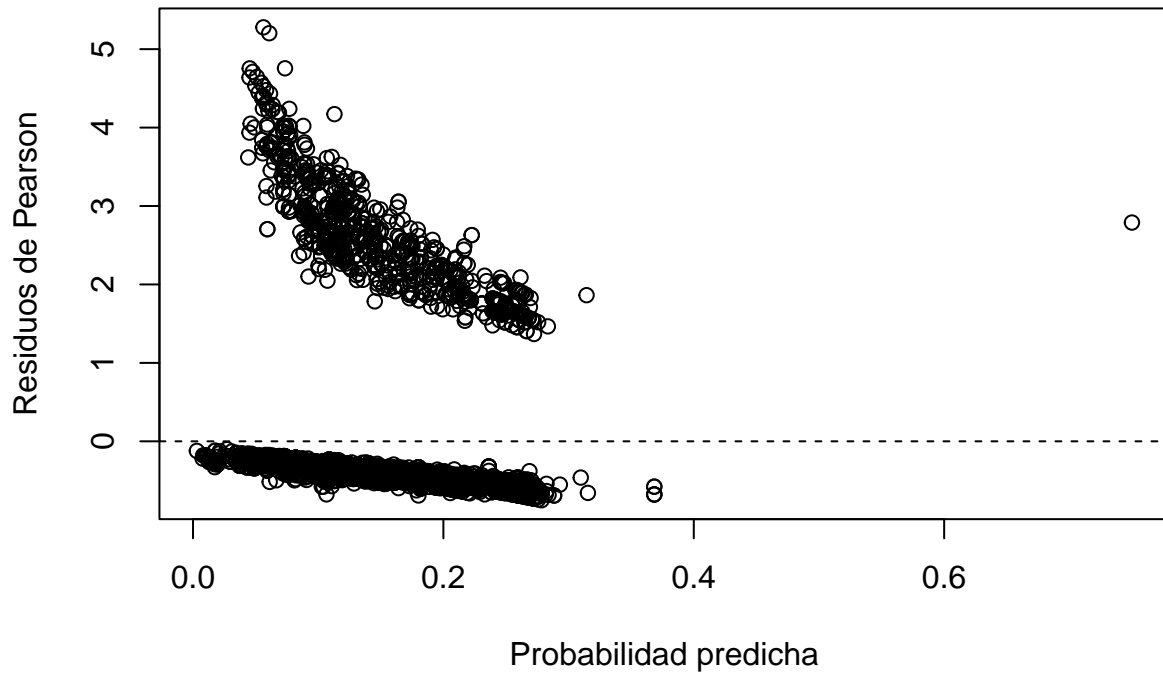
Residuos del modelo

```
res_pearson <- residuals(mT_full_tam, type = "pearson")
summary(res_pearson)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -0.751367 -0.404426 -0.317567 -0.001605 -0.249178  5.277871
```

```
plot(prob_principal, res_pearson,
     xlab = "Probabilidad predicha",
     ylab = "Residuos de Pearson",
     main = "Residuos de Pearson vs probabilidad predicha")
abline(h = 0, lty = 2)
```

Residuos de Pearson vs probabilidad predicha

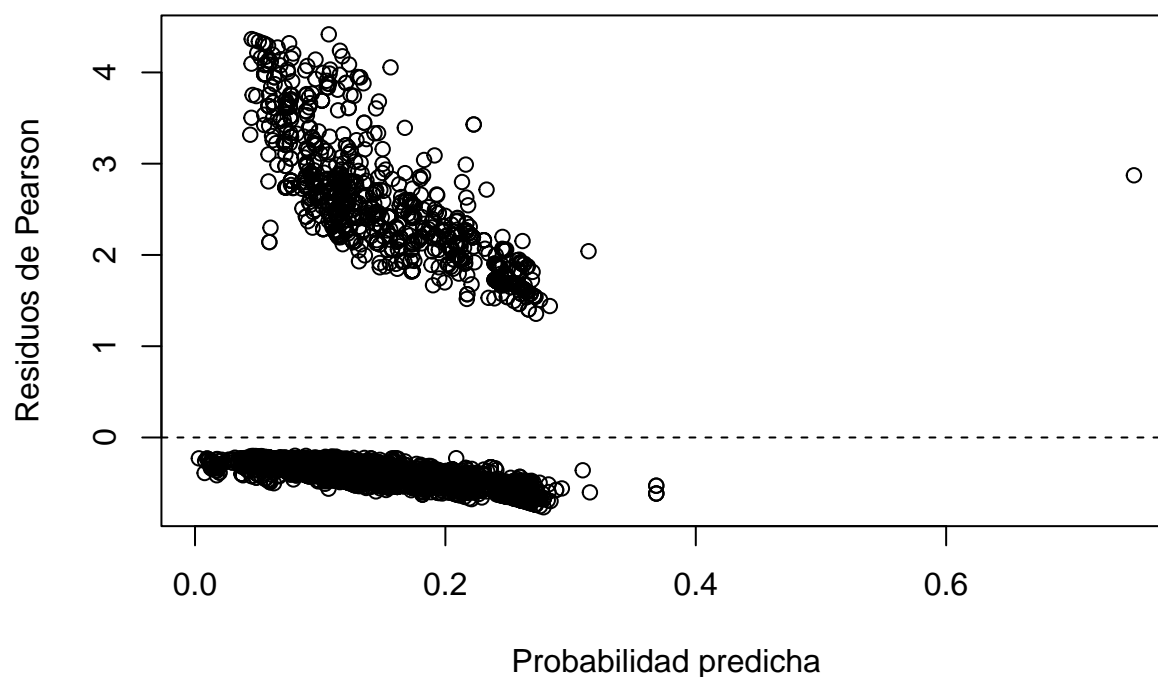


```
res_pearson <- residuals(mT_noB3_tam, type = "pearson")
summary(res_pearson)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
## -0.764288 -0.406849 -0.318889 -0.001467 -0.249956  4.417080
```

```
plot(prob_principal, res_pearson,
     xlab = "Probabilidad predicha",
     ylab = "Residuos de Pearson",
     main = "Residuos de Pearson vs probabilidad predicha")
abline(h = 0, lty = 2)
```

Residuos de Pearson vs probabilidad predicha



Los residuos no muestran patrones ni grandes desviaciones que nos indiquen problemas de especificacion de los modelos.

Multicolinealidad

```
vif(mT_full_tam)
```

```
##                GVIF Df  GVIF^(1/(2*Df))
## ns(fecha_est, 3) 4.186079 3      1.269505
## log_ancho        1.358991 1      1.165758
## tam_cat          1.471191 2      1.101330
## orientacion      1.228459 1      1.108359
## soporte_grp      3.677936 4      1.176794
## tecnica          1.981844 2      1.186499
## sop_montaje      1.662388 1      1.289336
## serie            1.173471 1      1.083269
## tema             1.781400 9      1.032598
```

```
vif(mT_noB3_tam)
```

```
##                GVIF Df  GVIF^(1/(2*Df))
## ns(fecha_est, 3) 1.798475 3      1.102768
## log_ancho        1.348681 1      1.161327
## tam_cat          1.113872 2      1.027327
## orientacion      1.208745 1      1.099429
## serie            1.158856 1      1.076502
## tema             1.653687 9      1.028339
```

Los modelos alternativos no presentan de problemas de multicolinealidad.

Confirmamos que los modelos no aportan ventajas claras respecto al modelo principal en la predicción de la probabilidad del éxito del cuadro.

RESULTADOS

En esta sección presentamos los resultados correspondientes al modelo escogido (`m_principal`), después de analizar múltiples modelos y las diferentes variables de la base de datos este es el modelo que presenta un mejor equilibrio entre la capacidad explicativa, parsimonia e interpretación,

El modelo se trata de un modelo de regresión logística binaria, estimado sobre una submuestra de 7002 pinturas. La variable dependiente es el éxito (binaria). Las variables explicativas las hemos distribuido en 3 bloques según sus características.

Para identificar cual es el mejor modelo para predecir el éxito del cuadro, nos hemos basado en variables explicativas separadas en cinco grupos, las referentes a la datación y el control de incertidumbre (`fecha_est` y `fecha_ancho`), las referentes a la morfología (área y orientación) y las referentes a al material y la técnica usada (soporte y técnica), la referente a la Iconografía (tema) y las referentes a autoría y serie (`tipo_autor` y `serie`). De todo el conjunto de bloques hemos analizado su nivel de significación en el modelo y hemos descartado del modelo final las variables `tema` y `tipo_autor`. Otras variables significativas las hemos transformado a logarítmicas o categorizado.

```
sum<-summary(m_principal)
```

Interpretación de coeficientes

Variables continuas

La variable `fecha_est` tiene un coeficiente positivo y significativo (0.00275), indicando que las obras realizadas en fechas más recientes presentan una mayor probabilidad de éxito.

```
(OR_f<-exp(100*sum$coefficients["fecha_est", "Estimate"]))
```

```
## [1] 1.317047
```

Cada 100 años que pasan, la pintura tiene un 31.7% más de probabilidad de tener éxito, siempre comparandola con 100 años menos.

En cambio, referente también a la datación, la variable `log_ancho` no es estadísticamente significativa, lo que nos sirve para tener un criterio de control del efecto temporal.

Variables categóricas

El tamaño de la obra `tam_cat` coge como referencia la categoría pequeño, en base a ella las otras medidas estimadas son mediano y grande. El tamaño del cuadro mediano no proporciona una diferencia significativa respecto a los cuadros pequeños. En cambio, las obras grandes sí que presentan un efecto en el éxito (0.918).

```
(OR_t<-exp(sum$coefficients["tam_catgrande", "Estimate"]))
```

```
## [1] 2.699279
```

El odds ratio nos indica que las obras de tamaño grande multiplican por 2.69 las probabilidades de éxito respecto a las pequeñas. Por lo que el tamaño si que es un factor importante para el modelo.

En lo que respecta a la orientación, tomamos como referencia la orientación vertical.

```
(OR_o<-exp(sum$coefficients["orientacionhorizontal","Estimate"]))
```

```
## [1] 1.331142
```

Las obras con una orientación horizontal presentan un 33.11% más de probabilidades de éxito que las verticales.

La variable soporte hace referencia al tipo de soporte usado, como referencia se coge el lienzo, en comparación con este tenemos Tabla/Panel y Otros que son significativos y Metal y Mural que no son estadísticamente significativos.

```
(OR_stp<-exp(sum$coefficients["soporte_grpTabla/Panel","Estimate"]))
```

```
## [1] 1.798009
```

```
(OR_so<-exp(sum$coefficients["soporte_grpOtros","Estimate"]))
```

```
## [1] 0.4615284
```

Los cuadros realizados sobre tabla o panel presentan un 79.8% más de probabilidades de éxito que el lienzo mientras que los soportes otros tienen una probabilidad de -53.9% de éxito respecto los lienzos.

La técnica con la que se pintaron las obras no presentan significación entre ellas, lo que nos indica que es independiente como se ha pintado el cuadro para nuestro modelo.

La presencia o no de soporte de montaje `sop_montaje` es significativo para el modelo.

```
(OR_so<-exp(sum$coefficients["sop_montajesi","Estimate"]))
```

```
## [1] 2.203652
```

Las pinturas que cuentan con soporte de montaje tienen 2.2 veces más probabilidades de éxito que aquellas que no lo tienen.

Si pertenecen a una serie es también importante en nuestro modelo. Las obras que pertenecen a una serie presentan un 32.28% más de probabilidades de éxito que las que son independientes.

```
(OR_ser<-exp(sum$coefficients["seriesi","Estimate"]))
```

```
## [1] 1.322774
```

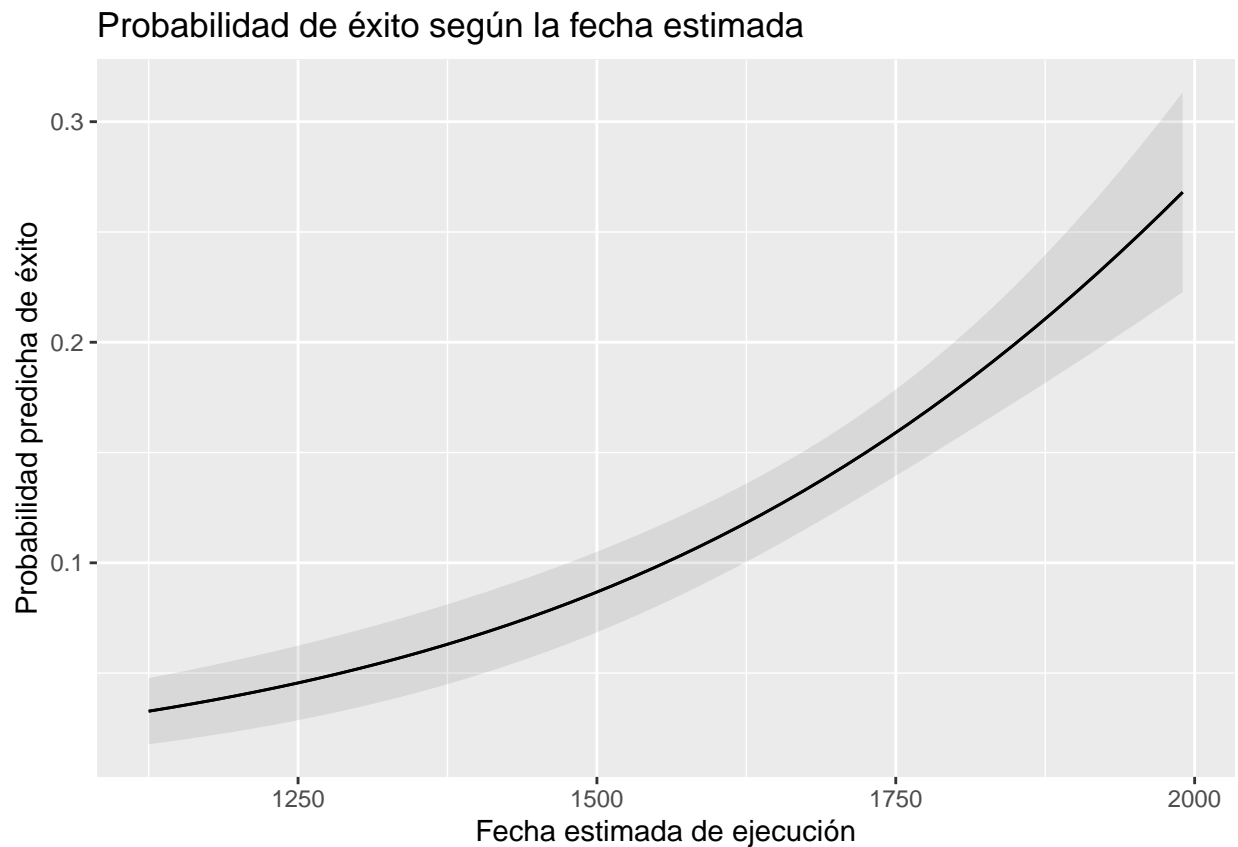
Finalmente tenemos las interacciones entre las variables tamaño y tipo de soporte de las cuales tan solo es significativa la interacción grande con Tabla/Panel, la cual reduce la probabilidad de éxito.

En conjunto podemos interpretar que en la estimación de la probabilidad del éxito intervienen diversos factores, temporal, morfológico, material y estructural.

Resumen de los resultados

Analizamos ahora gráficamente las predicciones de la probabilidad de éxito para cada variable donde las demás covariables las estamos obviando. Así comparamos como afecta cada una en el modelo de predicción.

```
plot_predictions(  
  m_principal,  
  condition = "fecha_est",  
  vcov = TRUE  
) +  
  labs(  
    x = "Fecha estimada de ejecución",  
    y = "Probabilidad predicha de éxito",  
    title = "Probabilidad de éxito según la fecha estimada"  
  )
```



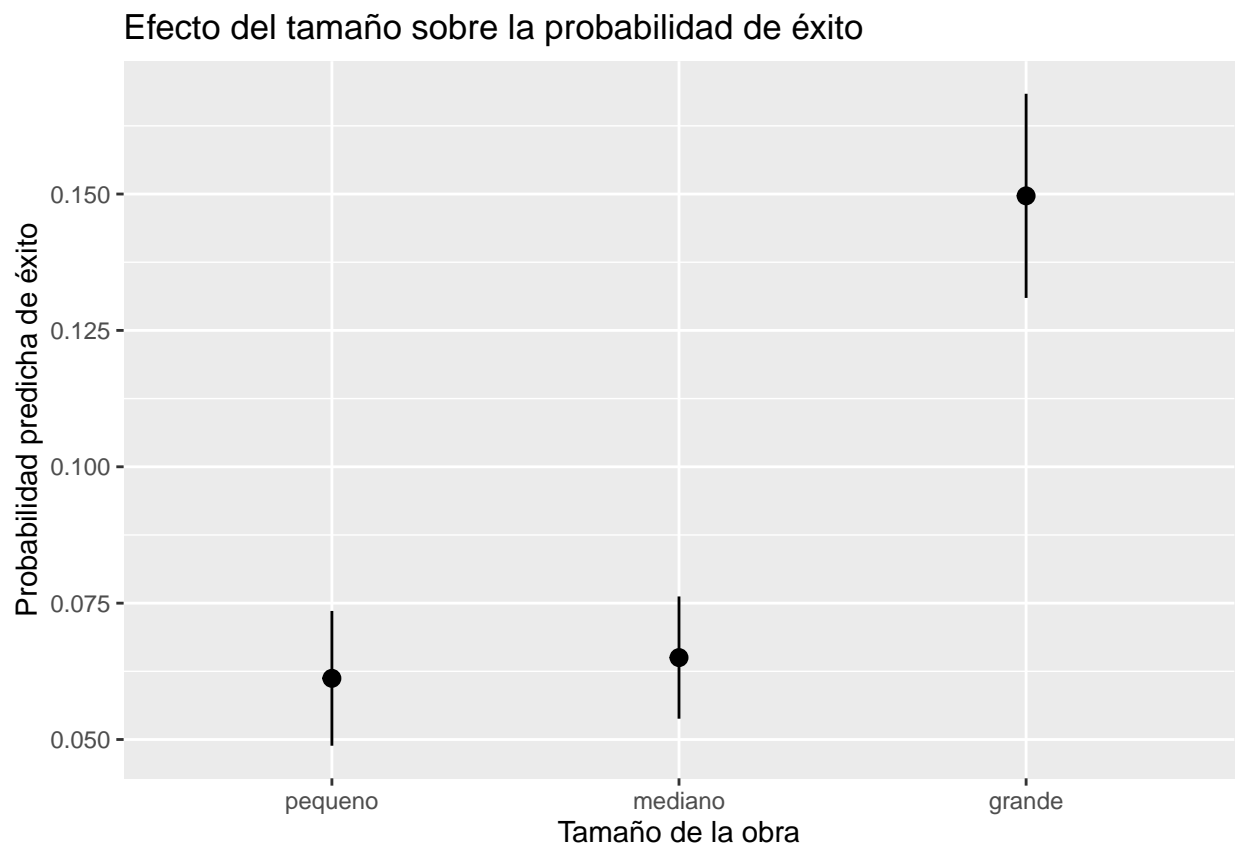
Observamos que a más actual es la obra, mayor probabilidad de éxito.

```
plot_predictions(  
  m_principal,  
  condition = "tam_cat",  
  vcov = TRUE  
) +  
  labs(  
    x = "Tamaño de la obra",  
    y = "Probabilidad predicha de éxito",  
  )
```

```

title = "Efecto del tamaño sobre la probabilidad de éxito"
)

```

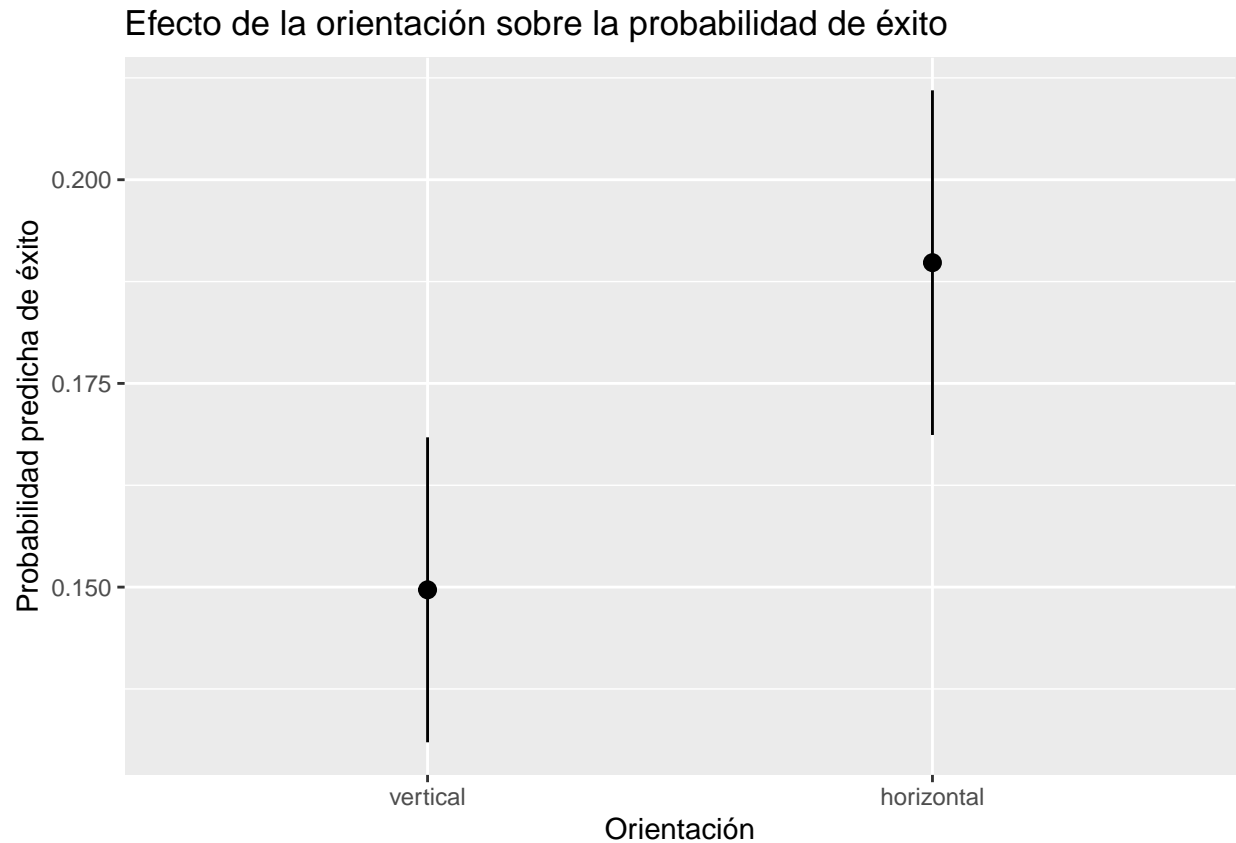


En el caso del tamaño de la obra separado por categorías vemos como la diferencia entre pequeño y mediano es practicamente nula e indicando una probabilidad muy baja, en cambio los cuadros con area grande tienen una probabilidad de éxito mucho mayor.

```

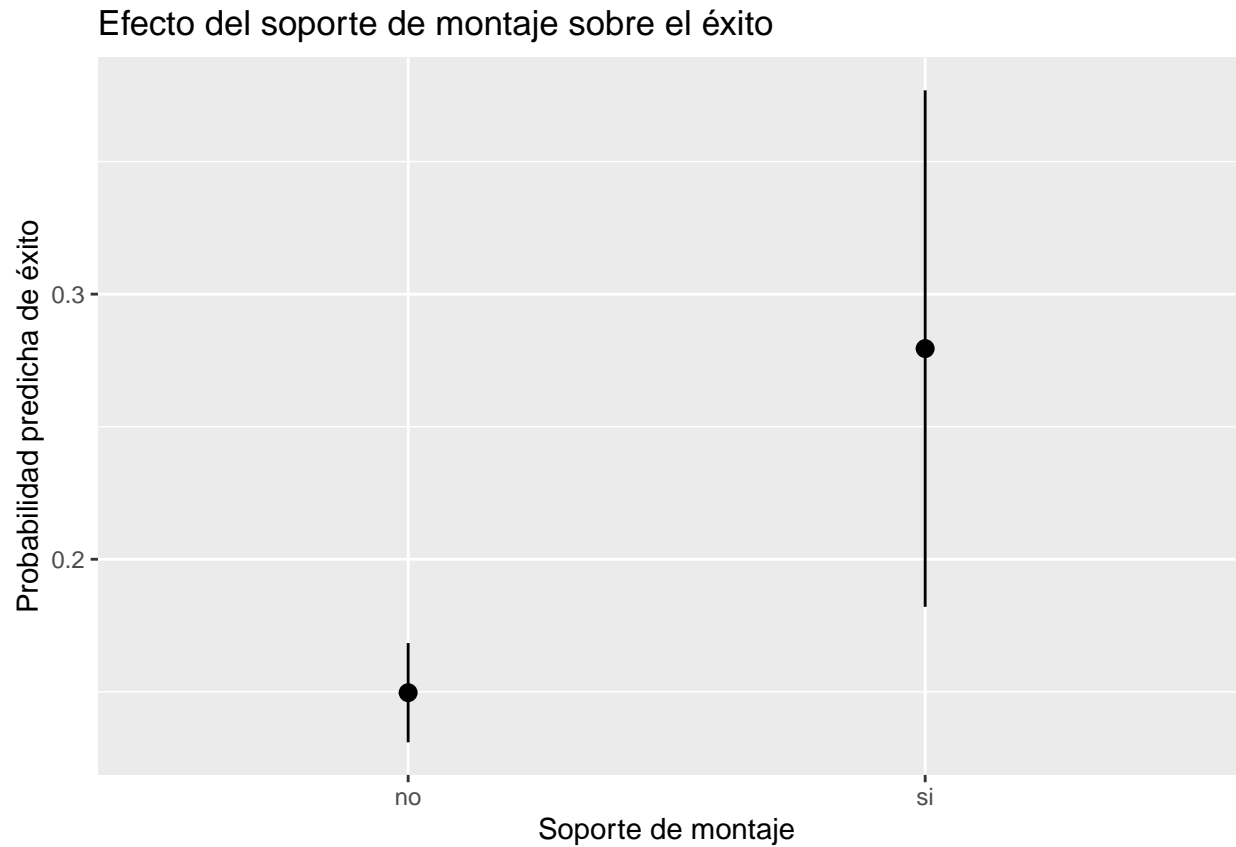
plot_predictions(
  m_principal,
  condition = "orientacion",
  vcov = TRUE
) +
  labs(
    x = "Orientación",
    y = "Probabilidad predicha de éxito",
    title = "Efecto de la orientación sobre la probabilidad de éxito"
  )

```



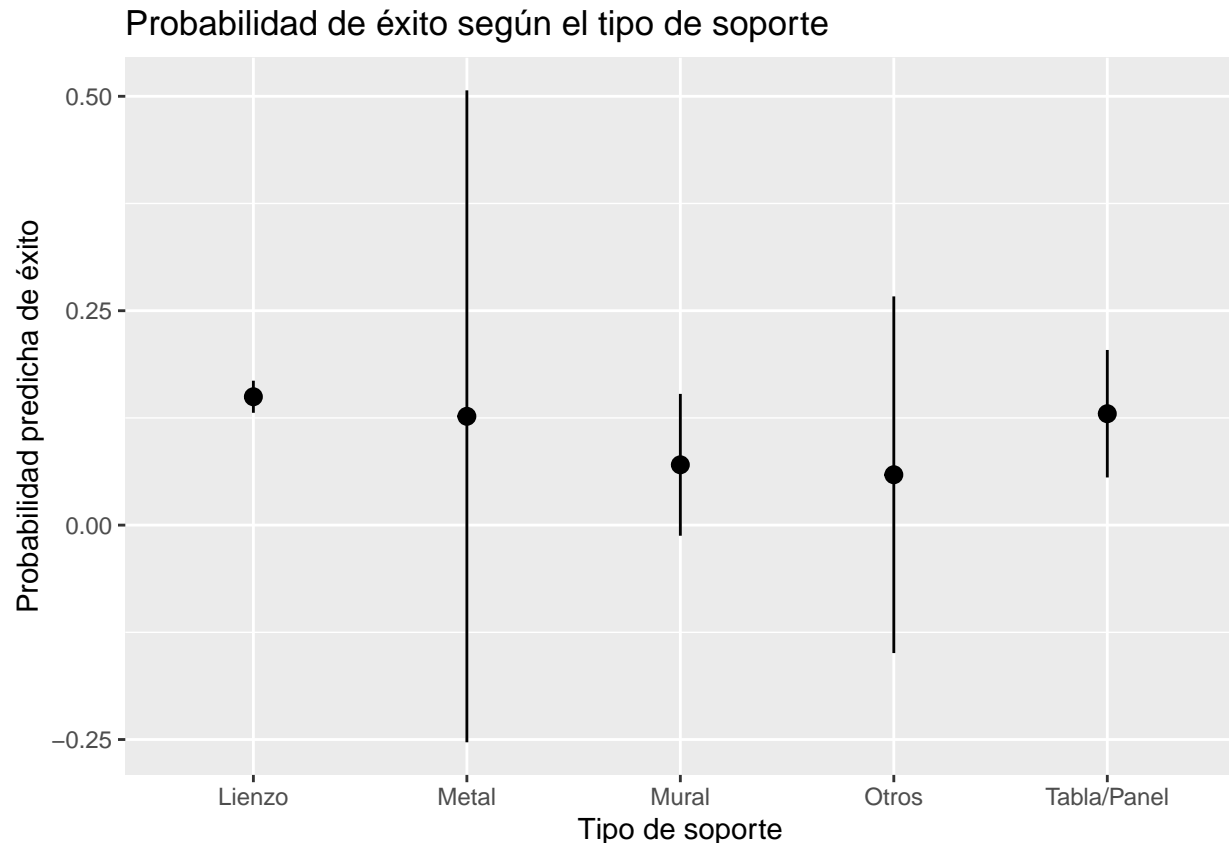
En lo que respecta a la orientación, las obras horizontales tienen una mayor probabilidad de éxito que las verticales.

```
plot_predictions(  
  m_principal,  
  condition = "sop_montaje",  
  vcov = TRUE  
) +  
  labs(  
    x = "Soporte de montaje",  
    y = "Probabilidad predicha de éxito",  
    title = "Efecto del soporte de montaje sobre el éxito"  
  )
```



Aquellos cuadros que presentan soporte de montaje tienen mayor probabilidad de éxito que los que no tienen soporte.

```
plot_predictions(  
  m_principal,  
  condition = "soporte_grp",  
  vcov = TRUE  
) +  
  labs(  
    x = "Tipo de soporte",  
    y = "Probabilidad predicha de éxito",  
    title = "Probabilidad de éxito según el tipo de soporte"  
  )
```



Finalmente vemos que las probabilidades según el tipo de soporte varían bastante. El lienzo tiene un intervalo de confianza muy bajo y probabilidad muy marcada. El metal vemos que no es indicativo del éxito ya que el intervalo de confianza es muy amplio llevándonos de 0.5 a -0.25. Para mural y otros tenemos un impacto menor que el lienzo y en el caso de Tabla/Panel tenemos una media similar al lienzo.

RESULTADOS MODELOS ALTERNATIVOS

Aunque hemos cogido `m_principal` como mejor modelo para explicar la probabilidad de éxito, también hemos creado dos modelos alternativos que nos explican la probabilidad de éxito. Estos modelos se tratan de `mT_full_tam` y `mT_noB3_tam`

```
smod1<-summary(mT_full_tam)
```

En el primer modelo alternativo hemos aplicado un modelo logit en el cual hemos añadido la variable tema de la obra para predecir el éxito y eliminado cualquier tipo de interacción entre variables.

En lo que respecta a las variables observamos que `ns(fecha_est,3)` es la única variable de fecha que tiene significación en el modelo. También vemos que tenemos 3 categorías de tema que sí que tienen diferencias significativas con los demás temas, que son `paisaje_lugares`, `otros` y `historia_allegoria`, todas ellas positivas en la predicción del éxito, los demás temas no son significativos.

```
smod2<-summary(mT_noB3_tam)
```

El segundo modelo alternativo `mT_noB3_tam` proporciona una representación más simple y contiene las mismas variables y estructura que el modelo anterior pero eliminando las referentes al bloque de material i técnica.

Las variables que se repiten en ambos modelos siguen presentando el mismo nivel de significación, magnitud y signo, por lo que la interpretación de las variables repetidas es la misma en ambos modelos.

Comparación entre los modelos

```
modelos_a_comparar <- list(
  "Principal" = m_principal,
  "full_tam" = mT_full_tam,
  "noB3_tam" = mT_noB3_tam
)

modelsummary(modelos_a_comparar,
  stars = TRUE,
  title = "Comparación de coeficientes entre especificaciones",
  coef_rename = c("fecha_est" = "Fecha Estimada", "log_area" = "log(Área)"),
  gof_omit = "AIC|BIC|Log.Lik|RMSE",
  output = "markdown")
```

Table 1: Comparación de coeficientes entre especificaciones

	Principal	full_tam	noB3_tam
(Intercept)	-7.519*** (0.690)	-3.020*** (0.812)	-2.950*** (0.521)
Fecha Estimada	0.003*** (0.000)		
log_ancho	0.029 (0.023)	0.033 (0.024)	0.026 (0.024)
tam_catmediano	0.064 (0.130)	0.041 (0.112)	-0.053 (0.104)
tam_catgrande	0.993*** (0.115)	0.945*** (0.107)	0.812*** (0.093)
orientacionhorizontal	0.286*** (0.077)	0.208* (0.083)	0.195* (0.082)
soporte_grpMetal	0.019 (0.375)	-0.372 (0.376)	
soporte_grpMural	1.786 (2.343)	-0.554 (0.609)	
soporte_grpOtros	-0.773* (0.389)	-1.070** (0.395)	
soporte_grpTabla/Panel	0.587*** (0.161)	0.476*** (0.130)	
tecnicamixta	0.417 (0.334)	0.128 (0.324)	
tecnicaotras	-0.050 (0.382)	-0.547 (0.460)	
sop_montajesi	0.790*** (0.231)	0.515* (0.233)	
seriesi	0.280** (0.094)	0.333*** (0.097)	0.302** (0.096)
tam_catmediano:soporte_grpMetal	-1.726 (1.498)		
tam_catgrande:soporte_grpMetal	-0.211 (1.790)		

	Principal	full_tam	noB3_tam
tam_catmediano:soporte_grpMural	-2.392 (2.497)		
tam_catgrande:soporte_grpMural	-2.631 (2.392)		
tam_catmediano:soporte_grpOtros	-1.902 (1.525)		
tam_catgrande:soporte_grpOtros	-0.264 (1.948)		
tam_catmediano:soporte_grpTabla/Panel	0.224 (0.250)		
tam_catgrande:soporte_grpTabla/Panel	-0.751* (0.371)		
ns(fecha_est, 3)1		0.303 (0.518)	0.126 (0.354)
ns(fecha_est, 3)2		0.495 (1.653)	0.950 (1.089)
ns(fecha_est, 3)3		1.313*** (0.396)	1.324*** (0.312)
temabodegon_floral		0.264 (0.207)	0.236 (0.205)
temacaza_animales		0.281 (0.294)	0.255 (0.294)
temahistoria_allegoria		0.416* (0.195)	0.386* (0.194)
temamitologia		-0.196 (0.245)	-0.251 (0.244)
temaotros		0.316** (0.116)	0.301** (0.116)
temapaisaje_lugares		0.554*** (0.129)	0.552*** (0.127)
temaproceso_obra		-0.146 (0.346)	-0.169 (0.345)
temaretrato_corte		0.136 (0.129)	0.106 (0.128)
temavida_cotidiana		-0.012 (0.188)	-0.023 (0.187)
Num.Obs.	7002	7002	7002
F	10.536	10.258	13.090
• p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001			

Finalmente observamos la comparación y nivel de significación de las variables y sus categorías en los tres modelos planteados, el principal y los alternativos.

ANALISIS SECUNDARIOS

También podemos usar la base de datos existente a fin de predecir otras variables de la base de datos.

Una de las variables que podemos predecir es el tamaño de los cuadros, es decir, el área de cada cuadro. Para hacerlo aplicamos un modelo de regresión simple, también analizamos sus posibles interacciones.

```
m_fecha <- lm(fecha_est ~ tema + soporte_grp + tam_cat + orientacion+tipo_autor+tecnica+serie*tam_cat,
anova(m_fecha)
```

```
## Analysis of Variance Table
##
## Response: fecha_est
##           Df    Sum Sq Mean Sq F value    Pr(>F)
## tema       9 24549173 2727686 259.863 < 2.2e-16 ***
## soporte_grp 4 16133137 4033284 384.245 < 2.2e-16 ***
## tam_cat     2 2919789 1459894 139.082 < 2.2e-16 ***
## orientacion 1  419822  419822  39.996 2.702e-10 ***
## tipo_autor  3 4317586 1439195 137.110 < 2.2e-16 ***
## tecnica     2 1724946  862473  82.166 < 2.2e-16 ***
## serie       1  559768  559768  53.328 3.135e-13 ***
## tam_cat:serie 2  689146  344573  32.827 6.458e-15 ***
## Residuals   6977 73235132 10497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

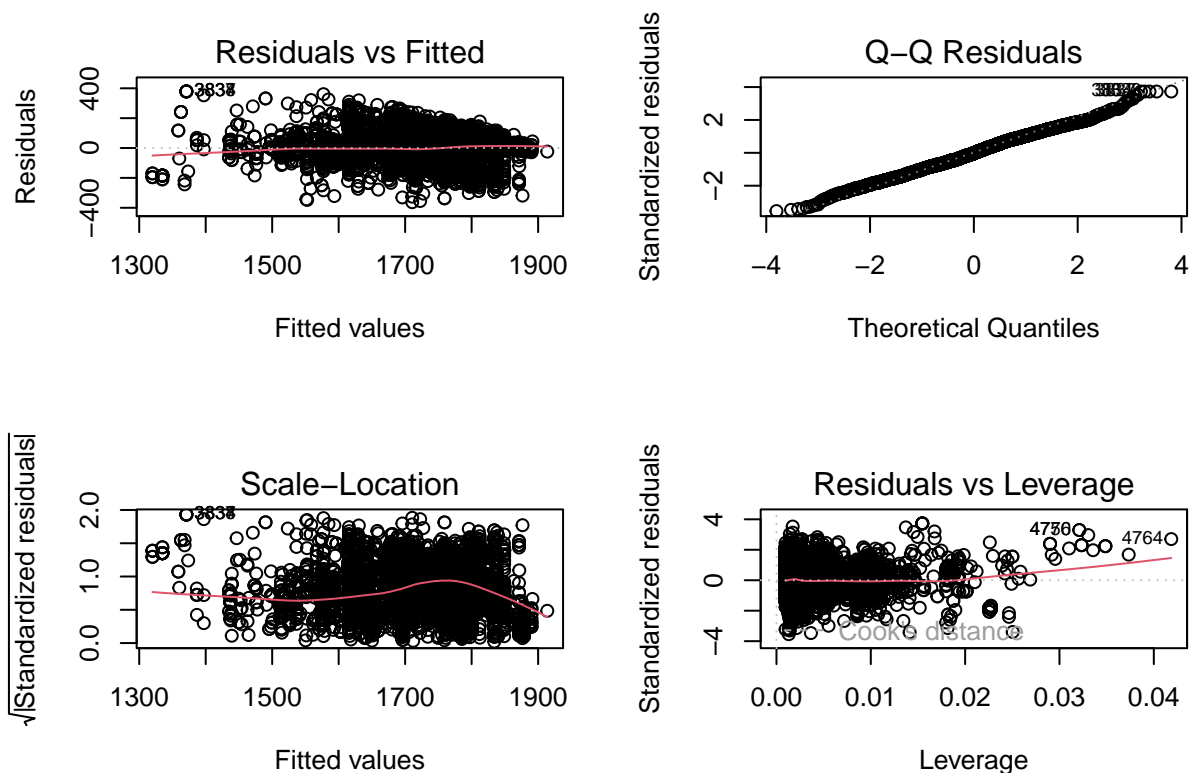
```
summary(m_fecha)
```

```
##
## Call:
## lm(formula = fecha_est ~ tema + soporte_grp + tam_cat + orientacion +
##      tipo_autor + tecnica + serie * tam_cat, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -362.65  -72.24   -3.16    77.53   379.07
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1734.0778     3.5821 484.090 < 2e-16 ***
## temabodegon_floral    11.9712     6.2570   1.913  0.05575 .
## temacaza_animales     27.5582    10.2641   2.685  0.00727 **
## temahistoria_allegoria 62.5315     6.9007   9.062 < 2e-16 ***
## temamitologia      -6.8152     7.2688  -0.938  0.34849
## temaotros          99.9706     3.6901  27.091 < 2e-16 ***
## temapaisaje_lugares   88.7365     4.3713  20.300 < 2e-16 ***
## temaproceso_obra     137.1869     9.6493  14.217 < 2e-16 ***
## temaretrato_corte     63.3937     3.9375  16.100 < 2e-16 ***
## temavida_cotidiana    73.2719     5.6800  12.900 < 2e-16 ***
## soporte_grpMetal    -104.9571     8.6092 -12.191 < 2e-16 ***
## soporte_grpMural    -106.4013    17.4394  -6.101 1.11e-09 ***
## soporte_grpOtros      41.2088     9.4413   4.365 1.29e-05 ***
## soporte_grpTabla/Panel -118.6438     3.7411 -31.714 < 2e-16 ***
## tam_catmediano      -38.2948     3.4641 -11.055 < 2e-16 ***
## tam_catgrande       -36.8353     3.6936  -9.973 < 2e-16 ***
## orientacionhorizontal  14.8703     2.7466   5.414 6.37e-08 ***
## tipo_autoranonimo    -78.9534     4.2730 -18.477 < 2e-16 ***
## tipo_autormujer       42.5075    13.5056   3.147  0.00165 **
## tipo_autorvarios     -50.9824     8.6961  -5.863 4.76e-09 ***
```



```
## tecnicamixta          -63.0984      10.1604   -6.210 5.59e-10 ***
## tecnicaotras          -139.3635     11.6534  -11.959 < 2e-16 ***
## seriesi               0.3808       6.3876   0.060 0.95246
## tam_catmediano:seriesi -3.1478      8.4012   -0.375 0.70790
## tam_catgrande:seriesi -52.9766      7.9775   -6.641 3.36e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 102.5 on 6977 degrees of freedom
## Multiple R-squared:  0.412, Adjusted R-squared:  0.41
## F-statistic: 203.7 on 24 and 6977 DF, p-value: < 2.2e-16
```

```
par(mfrow = c(2, 2))
plot(m_fecha)
```



Observamos que para predecir la fecha estimada del cuadro las variables tema, tipo de soporte, tamaño, orientación, tipo de autor, técnica, serie y la interacción entre tamaño del cuadro y serie son significativas.

CONCLUSIONES