

Análisis Principal

Manuela Lopez Cambron, 1673688

2026-01-18

!!!!!!!!!!!!!! REESCRIBIR VALORES LRT/AIC/BIC (conclusiones iguales pero diferentes valores)

```
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 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 ...
## $ 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 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 ...
## $ 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” mediante y su covariable de incertidumbre “fecha_ancho”, comparando especificación lineal vs flexible (spline) en ambas variables.

```

library(splines)

# Transformación log1p(fecha_ancho)
df$log_ancho <- log1p(df$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): moviendo los nudos 'interiores' que coinciden con
## los nudos de límite hacia adentro

# 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:

```
# 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 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 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 de la opción fecha_est flexible:\n")
```

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

```
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 en términos generales la especificación lineal es preferible frente a la flexible en términos de LRT como AIC. 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 consideramos esta opción mejor en términos de parsimonia, ya que añadimos menos parámetros al modelo, y dado que la mayoría de las variables son factores y nuestra respuesta fuertemente desbalanceada. Sin embargo, dejamos constancia del hecho que la especificación flexible con spline para fecha_est mejora sustancialmente el análisis.

Modelo resultante después de añadir Bloque 1:

```
m1 <- m1_lin_fecha
```

Bloque 2: Morfología

Se incorporan las variables morfológicas de tamaño y 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 $\log(area)$ (puedo ser tratada de manera flexible, si se observaran problemas. 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 elejimos “tam_cat”
- si “area” mejora sustancialmente más que “tam_cat”, elejimos “area”
- si “area” y “tam_cat” proporcionan resultados cualitativamente iguales, elejimos “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: 5055.282 vs 4995.578; BIC: 5103.259 vs 5050.410).

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("\nApunte de la opción log(area) flexible")
```

```
##
## Apunte 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(area)* lineal mejoró el ajuste, y al permitir no linealidad, el ajuste mejoró frente al lineal (AIC: 5017.2 vs 5055.3). Sin embargo, la especificación categórica “tam_cat” presentó el mejor compromiso ajuste-complejidad, con AIC y BIC claramente inferiores (AIC: 4995.6 vs 5017.241; BIC: 5050.4 vs 5078.9), 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(area)* sin spline). El único objetivo de mantener esta opción como alternativa es el de aportar un modelo más parsimonioso de manera que, aunque su versión flexible mostró un mejor ajuste, no cumple con nuestro objetivo. Vemos que *log(area)* añade 1 parámetro (uno menos que “tam_cat”) mientras que *ns(log(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(area)*)

Bloque 3: Material y técnica

Se incorporan las variables “soporte” y “tecnica” para evaluar si aporta información adicional, 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 covariante.

```

# 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("\nApunte de información del Bloque 3:\n")

## 
## Apunte 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 $p = 0.0016$) también mejoró el ajuste penalizado por AIC ($\Delta AIC \approx -9.4$), aunque el BIC aumentó ($\Delta BIC \approx +31.7$) (mayor penalización por el número de parámetros añadidos). Debido al objetivo descriptivo de nuestro estudio, se necesita 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.

```

# 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("\nApunte 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 $p = 0.009$). El AIC disminuyó ($\Delta AIC \approx -4.8$), indicando una mejora del ajuste penalizado por complejidad, aunque el BIC aumentó ligeramente ($\Delta BIC \approx +2$), 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 ($\Delta_{total} BIC \approx +33.76$). 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_alegoria         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("\nApunte 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("\nApunte de complejidad:\n")

```

```

##
## Apunte 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 $p = 0.0018$) y redujo el AIC ($\Delta AIC \approx -8.3$), 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("\nApunte de información del Bloque 5:\n")
```

```
##  
## Apunte 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")

```

```

##
## Apporte 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 $p = 0.0056$) y redujo el AIC ($\Delta AIC \approx -6.6$), aunque incrementó fuertemente el BIC ($\Delta BIC \approx +20.8$), reflejando un aumento de complejidad. En este caso podemos observar algo interesante, los coeficientes pertenecientes a “tipo_autor” no serultaron 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")

```

```

##
## Apporte 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 $p = 0.0011$) mientras que “autor” no lo hizo (LRT $p = 0.2277$). También se observa como el modelo que solo incluye serie mejoró tanto el AIC como el BIC respecto al anterior modelo aceptado ‘m3’ ($\Delta AIC \approx -8.6$, $\Delta BIC \approx -1.7$), como al modelo completo ‘m5’. Como ver además como el modelo que incluye “tipo_autor” presenta 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

!!!!!! explicar un poco que ha pasado !!!!!!!

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 de 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
# =====
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")
```

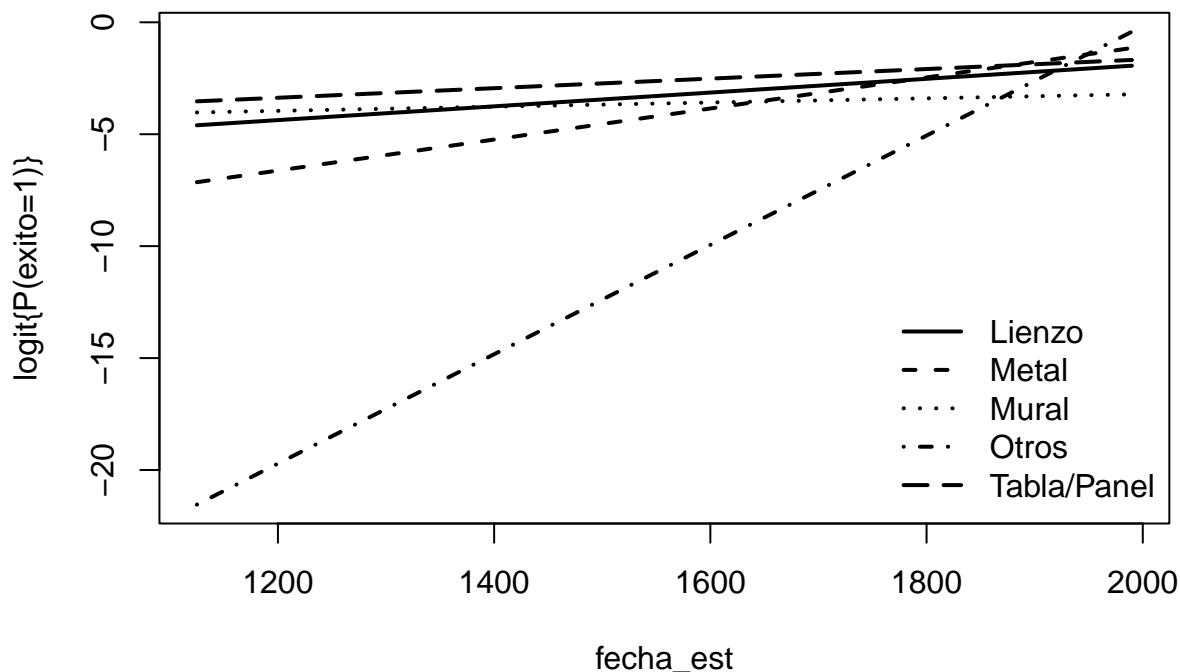
```

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

```

Interacción: soporte_grp × fecha_est (escala logit)



```

# -----
# 2) tecnica × 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

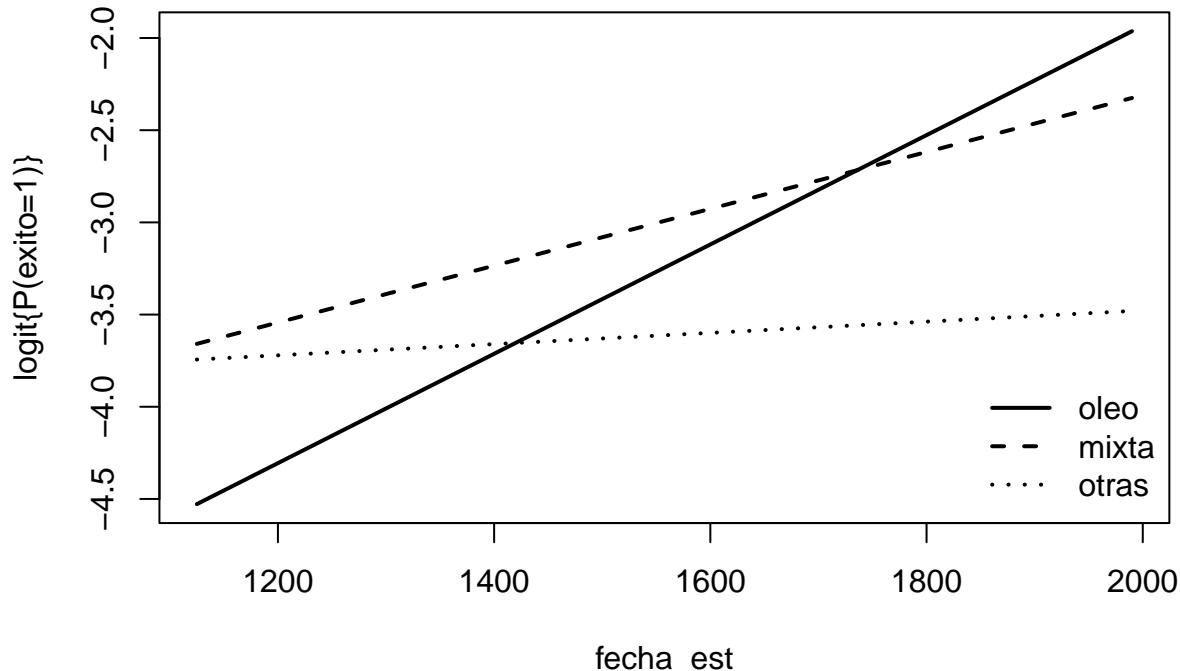
eta <- predict(m_int_TecFecha, newdata = nd, type = "link")

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)



```

# =====
# 3) 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
)

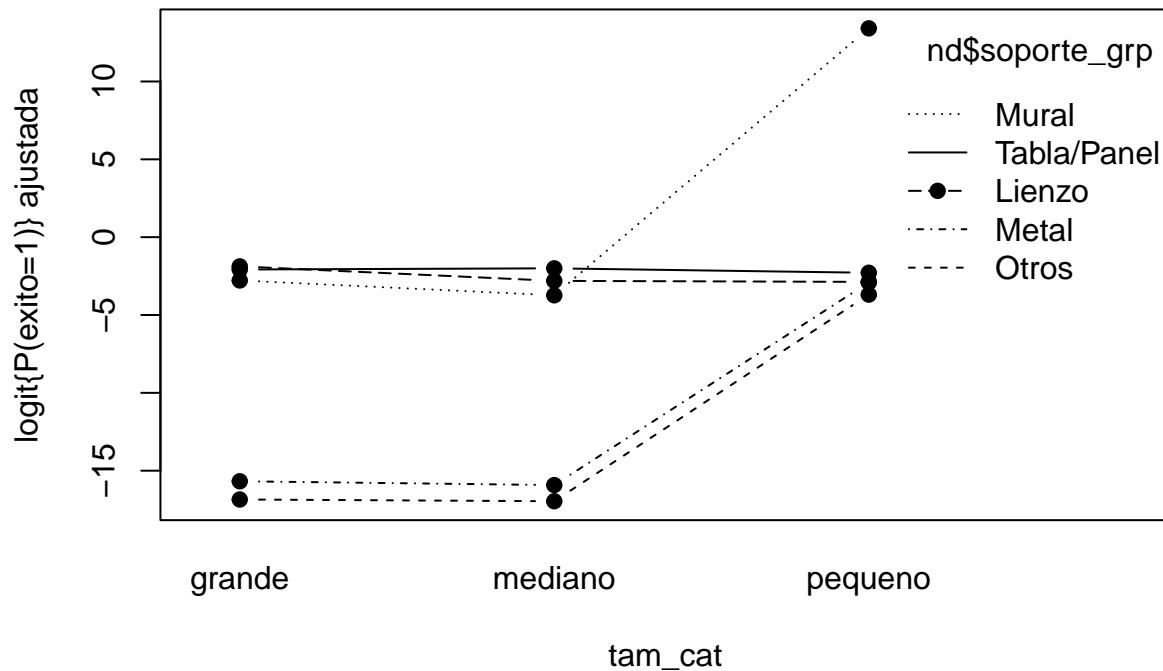
nd$fecha_est     <- fecha0
nd$log_ancho    <- log_ancho0
nd$orientacion  <- ref_orientacion
nd$tecnica      <- ref_tecnica
nd$sop_montaje  <- ref_sop_montaje
nd$serie        <- ref_serie

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)



```

# -----
# 4) 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

```

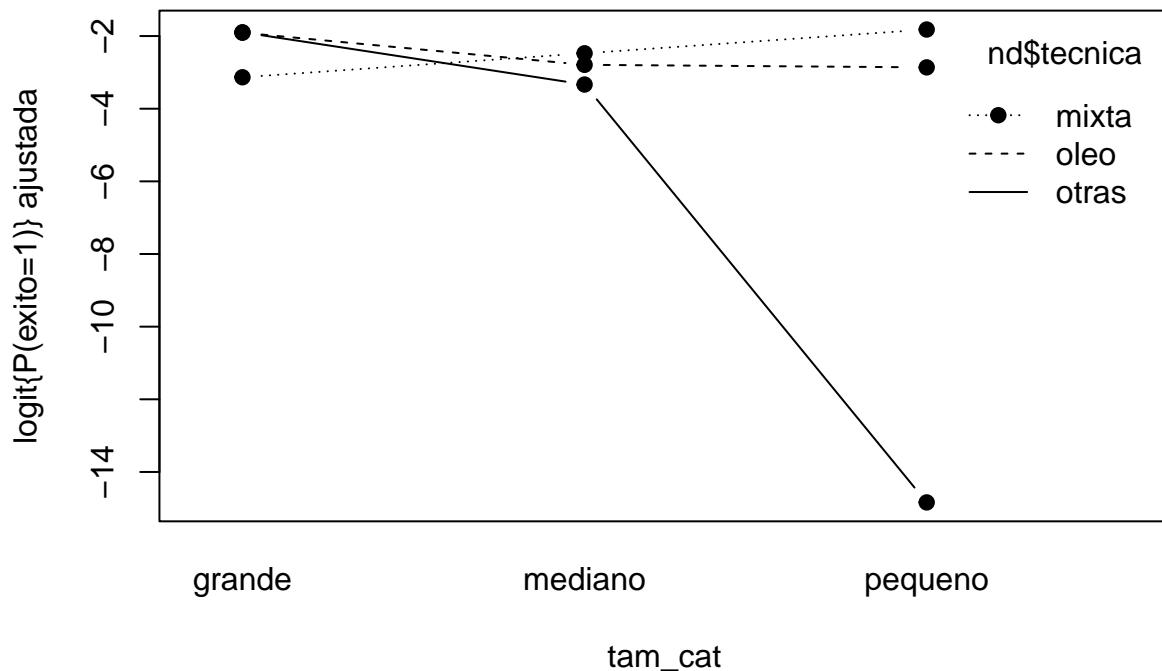
```

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)



```

# -----
# 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
)

```

```

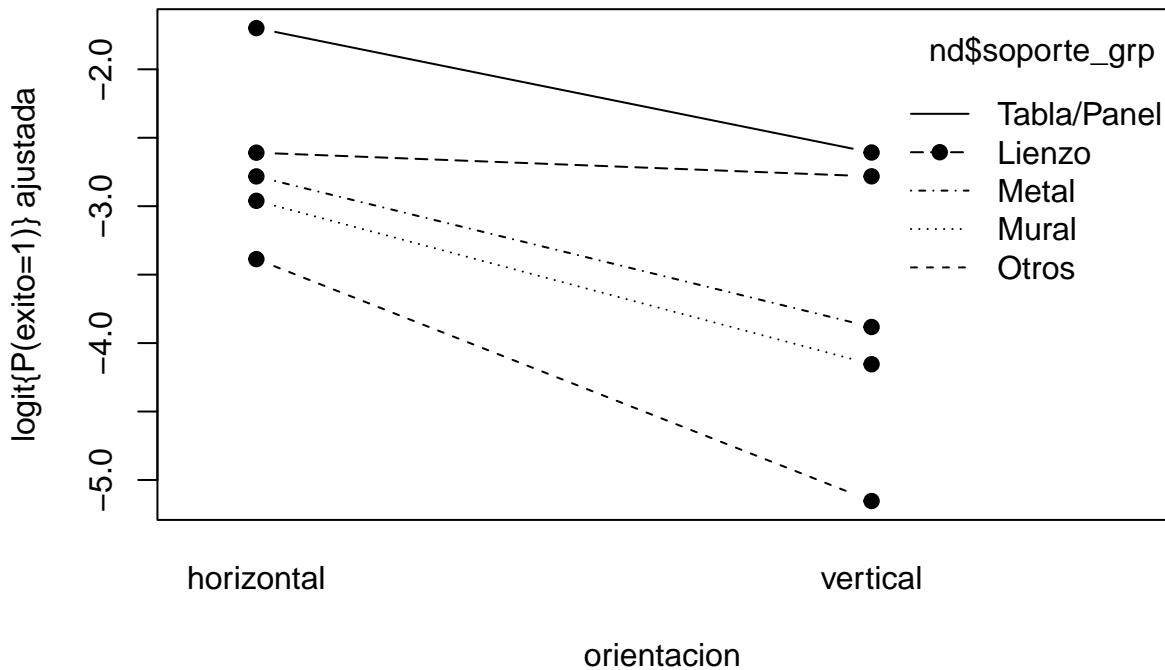
nd$fecha_est      <- fecha0
nd$log_ancho     <- log_ancho0
nd$tam_cat        <- ref_tam_cat
nd$tecnica        <- ref_tecnica
nd$sop_montaje   <- ref_sop_montaje
nd$serie          <- ref_serie

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 x orientacion (escala logit)



1) soporte_grp x fecha_est

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

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

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

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 pequeño 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 éxitos se concentran alrededor de las categoría ‘Lienzo’ y ‘Table/Panel’ lo cual es lógico ya que son 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

Esta interacción ha resultado la menos relevante, pero aun con indicios 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\\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 información (LRT(99) $p < 0.01$) i reduciendo el AIC ($\Delta AIC \approx -7; -10$). Aunque el BIC aumentó 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 grave problema de complejidad ($\Delta BIC \approx +74$) que no consideramos aceptable ni necesario en este punto análisis. En consecuencia descartamos la interacción con orientación una vez controlada 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’.

Finalmente se decide definir el modelo principal con únicamente la interacción “soporte_grp x tam_cat” con el objetivo 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.

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")
)

```

Modelos alternativos

Las alternativas de enlace y penalización ya se evaluaron sobre el modelo principal por lo que los modelos complementarios se presentan bajo la versión base: el enlace logit.

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 $\log(\text{area})$ 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 “tecnica”, las cuales también eran de interés. Finalmente se contempla la posibilidad de cambiar la especificación del tamaño por su versión continua lineal ($\log(\text{area})$) con el objetivo de simplemente controlar el efecto del tamaño, pero sin interés interpretativo en esta sección.

Ajuste 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 (minimas, 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 aparte de información, es decir, independientemente de como controlamos el tamaño y de si incluimos o no el Bloque 3 (técnica/soporte/montaje), tema añade información. Este resultado avala nuestra intención de realizar esta reción 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 (técnica/soporte/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 utilizaran para interpretar la asociación de “tema” son ‘mT_full_tam’ y ‘mT_noB3_tam’.

Modelos alternativos finales

Modelos alternativos finales:

```

# asociación ajustado
mA1 <- mT_full_tam

# asociación global
mA2 <- mT_noB3_tam

```

Validación

!!!!!!!!!!!!!! (angel/marc) !!!!!!!!!!!!!!!11

creo que se tendría que validar estas dos modelos así un poco por encima