Tercera\_Evaluacion

Garcia Rios Santiago

Victor

Alejandra

2024-04-23

Table of contents

## 1 Tilapias

En un ejido en Veracruz un grupo de familias ha montado un sistema de crianza de tilapias y desean ***conocer el efecto de la densidad de peces en el encierro y de la estación del año en el crecimiento*** de los individuos. Para probar el efecto de estos factores, realizan un experimento en encierros en donde colocan 10, 18 o 24 individuos (niveles dentro del factor densidad). Pesan 9 peces antes y después del experimento (marcándolos para recapturar al mismo pez) y registran el incremento en peso al cabo de dos semanas. Realizan estas mediciones en verano y en primavera (factor estación del año). ¿Cuál es el efecto de la época del año y de la densidad de peces en el encierro en el crecimiento de las tilapias? Utiliza todos los recursos vistos en clase para presentar los resultados de este ejercicio. NOTA: Las mediciones en cada estación vienen de peces distintos.

1. Primero, cambiar a factores los datos y exploramos los datos:

Data frame: tilapias

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Name | Label | missings | Values | Value Labels | Freq. | % |
| 1 | densidad |  | 0 (0.00%) |  | 10 18 24 | 18 18 18 | 33.33 33.33 33.33 |
| 2 | estacion |  | 0 (0.00%) |  | primavera verano | 27 27 | 50.00 50.00 |
| 3 | aumento |  | 0 (0.00%) | *range: 60.0-402.0* | |  |  |

Ahora, para determinar si hay un efecto significativo de nuestros dos factores (densidad y estación), podemos utilizar un ANOVA de dos vías:

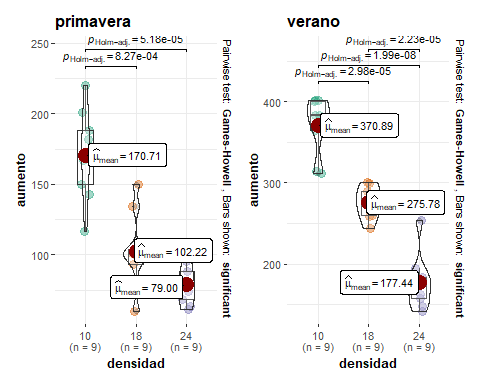
Df Sum Sq Mean Sq F value Pr(>F)   
densidad 2 184282 92141 119.2 < 2e-16 \*\*\*  
estacion 1 334428 334428 432.6 < 2e-16 \*\*\*  
densidad:estacion 2 25050 12525 16.2 4.2e-06 \*\*\*  
Residuals 48 37104 773   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

|  |  |
| --- | --- |
|  | aumento |
| Predictors | p |
| densidad | **<0.001** |
| estacion | **<0.001** |
| densidad:estacion | **<0.001** |
| Residuals |  |
| Observations | 54 |
| R2 / R2 adjusted | 0.936 / 0.929 |
| \* p<0.05   \*\* p<0.01   \*\*\* p<0.001 | |

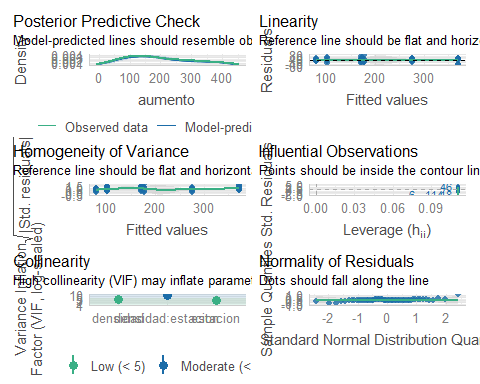
El ANOVA (fórmula aumento ~ densidad \* estación) sugiere que:

* El efecto principal de la densidad es estadísticamente significativo y grande (F(2, 48) = 119.20, p < .001; Eta2 (partial) = 0.83, 95% CI [0.76, 1.00]).
* El efecto principal de estación es estadísticamente significativo y grande (F(1, 48) = 432.63, p < .001; Eta2 (partial) = 0.90, 95% CI [0.86, 1.00]).
* La interacción entre densidad y estación es estadísticamente significativa y grande (F(2, 48) = 16.20, p < .001; Eta2 (partial) = 0.40, 95% CI [0.22, 1.00]).

El siguiente gráfico (post-hoc ajustada con corrección de Holm) se realizó para ejemplificar la diferencia de medias:



Supuestos del modelo:



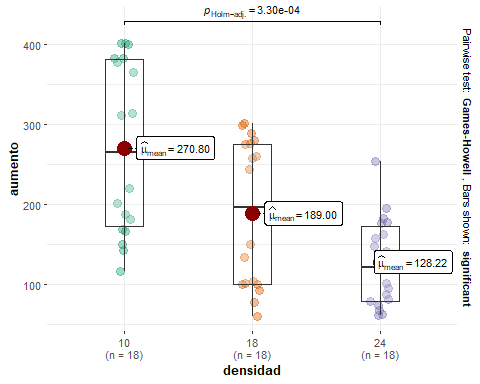
Parce ser que no tenemos violaciones graves a nuestro modelo.

## 2 Tilapias por separado

Analiza los mismos datos del ejercicio anterior pero probando el efecto de la estación y de la densidad por separado. ¿Qué diferencia encuentras en la interpretación de los resultados? Dado el diseño experimental y la pregunta central, ¿qué tipo de análisis es el más conveniente y por qué? No te olvides de revisar los supuestos.

1. **Densidad**:

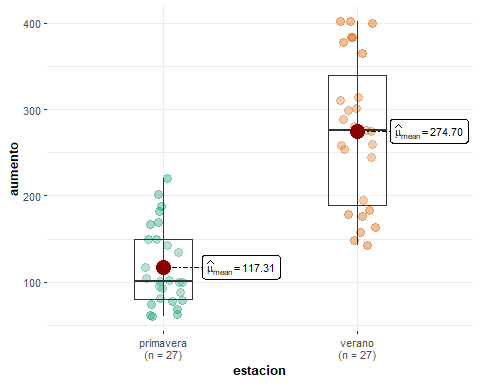
Df Sum Sq Mean Sq F value Pr(>F)   
densidad 2 184282 92141 11.85 5.94e-05 \*\*\*  
Residuals 51 396582 7776   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



|  |  |
| --- | --- |
|  | aumento |
| Predictors | p |
| densidad | **<0.001** |
| Residuals |  |
| Observations | 54 |
| R2 / R2 adjusted | 0.317 / 0.290 |
| \* p<0.05   \*\* p<0.01   \*\*\* p<0.001 | |

El ANOVA de densidad (formula: aumento ~ densidad) sugiere que el efecto principal de la densidad es estadísticamente significativo y grande (F(2, 51) = 11.85, p < .001; Eta2 = 0.32, 95% CI [0.14, 1.00]),.

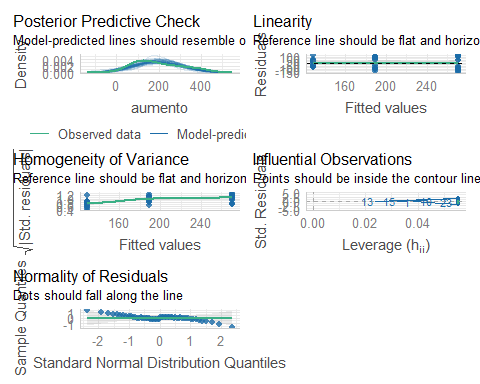
Df Sum Sq Mean Sq F value Pr(>F)   
estacion 1 334428 334428 70.57 2.98e-11 \*\*\*  
Residuals 52 246436 4739   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

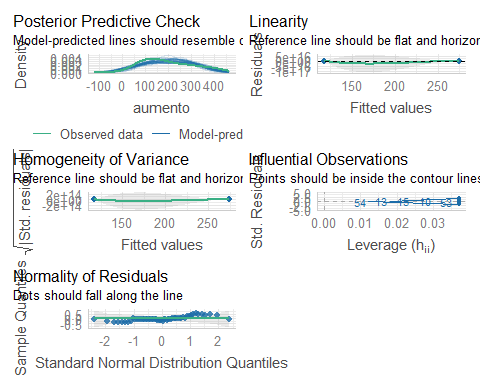


|  |  |
| --- | --- |
|  | aumento |
| Predictors | p |
| estacion | **<0.001** |
| Residuals |  |
| Observations | 54 |
| R2 / R2 adjusted | 0.576 / 0.568 |
| \* p<0.05   \*\* p<0.01   \*\*\* p<0.001 | |

El ANOVA de estación (formula: aumento ~ estación) sugiere que el efecto principal de la estación es estadísticamente significativo y grande (F(1, 52) = 70.57, p < .001; Eta2 = 0.58, 95% CI [0.43, 1.00]).

1. Revisar supuestos (parece que el modelo con interacción cumple mejor con los supuestos).





**Conclusión**

Si analizamos cada factor por separado, solo podemos concluir sobre el efecto individual de cada variable predictora, sin considerar una interacción entre estas variables. Es decir, no podríamos modelar correctamente si la densidad influye sobre el crecimiento de las tilapias dependiendo de la estación del año (o viceversa).

Nuestro diseño experimental sugiere que tanto la densidad como la estación podrían interaccionar en su efecto sobre el crecimiento de las tilapias. En este caso, el ANOVA de dos vías permite evaluar tanto los efectos principales de cada factor como su interacción. Este análisis permite comprender mejor el fenómeno y puede servir mejor en la practica (por ejemplo, para gestionar la crianza de tilapias en diferentes condiciones).

|  |  |
| --- | --- |
|  | **Con Dummy variables**  NOTA, también se puede hacer el ejercicio como un modelo lineal con variables *dummy*:  [1] 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 18 18 18 18 18 18 18 [26] 18 18 18 18 18 18 18 18 18 18 18 24 24 24 24 24 24 24 24 24 24 24 24 24 24 [51] 24 24 24 24 Levels: 10 18 24  # A tibble: 3 × 2  densidad n  <fct> <int> 1 10 18 2 18 18 3 24 18  Call: lm(formula = aumento ~ 1 + dummy\_densidad\_b + dummy\_densidad\_c,   data = tilapias\_dummy)  Residuals:  Min 1Q Median 3Q Max  -154.100 -79.550 -6.222 82.250 131.200   Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 270.80 20.78 13.029 < 2e-16 \*\*\* dummy\_densidad\_b -81.80 29.39 -2.783 0.00754 \*\*  dummy\_densidad\_c -142.58 29.39 -4.851 1.2e-05 \*\*\* --- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  Residual standard error: 88.18 on 51 degrees of freedom Multiple R-squared: 0.3173, Adjusted R-squared: 0.2905  F-statistic: 11.85 on 2 and 51 DF, p-value: 5.937e-05  Call: lm(formula = aumento ~ densidad, data = tilapias)  Residuals:  Min 1Q Median 3Q Max  -154.100 -79.550 -6.222 82.250 131.200   Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 270.80 20.78 13.029 < 2e-16 \*\*\* densidad18 -81.80 29.39 -2.783 0.00754 \*\*  densidad24 -142.58 29.39 -4.851 1.2e-05 \*\*\* --- Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  Residual standard error: 88.18 on 51 degrees of freedom Multiple R-squared: 0.3173, Adjusted R-squared: 0.2905  F-statistic: 11.85 on 2 and 51 DF, p-value: 5.937e-05 |

## 3 Producción Maíz 2009

Estos datos provienen de un estudio del maíz en un área periurbana. Aunque hay más variables en la base, para este ejercicio queremos construir un modelo que permita predecir la producción de 2009 (variable produccion\_2009) a partir de las hectáreas sembradas (ha\_maiz). ¿Qué tan bueno es este modelo para realizar predicciones? Utiliza todos los recursos vistos en clase para contestar. NOTA. Elimina el dato con producción de 2009 igual a cero, pues el entrevistador ha reportado que se trata de un error de captura

1. Eliminar datos de error:
2. Modelado estadístico:

int [1:145] NA 2400 3000 1800 120 1320 1080 NA 1800 1200 ...

num [1:145] NA 1 2 0.33 0.02 0.17 1 NA 1 0.5 ...

Call:  
lm(formula = produccion\_2009 ~ ha\_maiz, data = toluca)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-12733.2 -965.9 123.3 1026.6 16236.8   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 3.408 539.744 0.006 0.995   
ha\_maiz 1969.973 189.743 10.382 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 3618 on 80 degrees of freedom  
 (63 observations deleted due to missingness)  
Multiple R-squared: 0.574, Adjusted R-squared: 0.5687   
F-statistic: 107.8 on 1 and 80 DF, p-value: < 2.2e-16

|  |  |  |  |
| --- | --- | --- | --- |
|  | produccion 2009 | | |
| Predictors | Estimates | CI | p |
| (Intercept) | 3.41 | -1070.72 – 1077.53 | 0.995 |
| ha maiz | 1969.97 \*\*\* | 1592.37 – 2347.57 | **<0.001** |
| Observations | 82 | | |
| R2 / R2 adjusted | 0.574 / 0.569 | | |
| \* p<0.05   \*\* p<0.01   \*\*\* p<0.001 | | | |

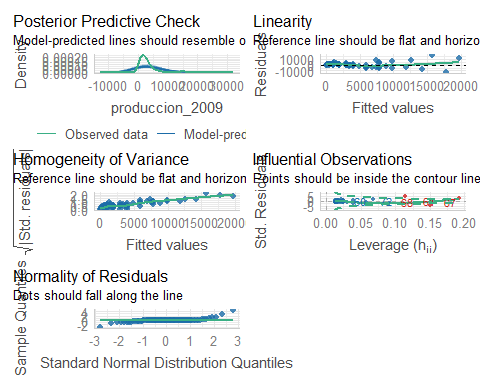
Ajustamos un modelo lineal (estimado usando OLS) para predecir la producción produccion de 2009 a partir de las hectáreas de maíz (fórmula: produccion\_2009 ~ ha\_maiz). El modelo explica una proporción de varianza estadísticamente significativa y sustancial (R2 = 0.57, F(1, 80) = 107.79, p < .001, R2 ajustado = 0.57).

El intercepto del modelo, correspondiente a ha\_maiz = 0, está en 3.41 (IC del 95% [-1070.72, 1077.53], t(80) = 6.31e-03, p = 0.995).

Dentro de este modelo:

El efecto de ha maiz es estadísticamente significativo y positivo (beta = 1969.97, IC del 95% [1592.37, 2347.57], t(80) = 10.38, p < .001; beta estandarizada = 0.76, IC del 95% [0.61, 0.90])

Los parámetros estandarizados se obtuvieron ajustando el modelo en una versión estandarizada del conjunto de datos. Los intervalos de confianza del 95% (IC) y los *p-values* se calcularon utilizando una aproximación de la distribución t de Wald.



## 4 Maíz II,

Utilizando de nuevo los datos del estudio de maíz periurbano, realiza una comparación entre la producción del 2009 entre las diferentes comunidades. Sabemos que la producción está asociada con el área sembrada, por lo que quisiéramos tomar esto en cuenta en nuestro análisis al comparar las comunidades, i.e. queremos incorporar la covariable hectáreas sembradas de maíz (ha\_maiz). ¿Hay diferencias en producción entre los sitios tomando en cuenta las hectáreas sembradas de maíz? Aunque no se detecte una diferencia significativa entre comunidades, ¿cuál sería la ecuación de regresión para cada comunidad (la finalidad es ejercitar el uso de las variables dummy)? Genera la gráfica correspondiente con tres rectas, una para cada comunidad a partir de las ecuaciones que generaste para cada comunidad.

R automáticamente toma una de las categorías como referencia, por lo que necesitamos generar manualmente las variables *dummy*.

Factor w/ 3 levels "chapultepec",..: NA 3 3 3 3 3 3 NA 3 3 ...

Call:  
lm(formula = produccion\_2009 ~ ha\_maiz + comunidad, data = toluca)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-12704.3 -1197.1 224.6 1079.8 16232.1   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 121.8 849.0 0.143 0.886   
ha\_maiz 1936.4 223.6 8.659 4.96e-13 \*\*\*  
comunidadparedon 154.8 1209.1 0.128 0.898   
comunidadsanfrancisco -220.5 1000.5 -0.220 0.826   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 3661 on 78 degrees of freedom  
 (63 observations deleted due to missingness)  
Multiple R-squared: 0.5747, Adjusted R-squared: 0.5583   
F-statistic: 35.13 on 3 and 78 DF, p-value: 1.799e-14

Sin embargo, para hacer las variables *dummy* manualmente (a manera de un *ANCOVA*), podríamos hacer lo siguiente:

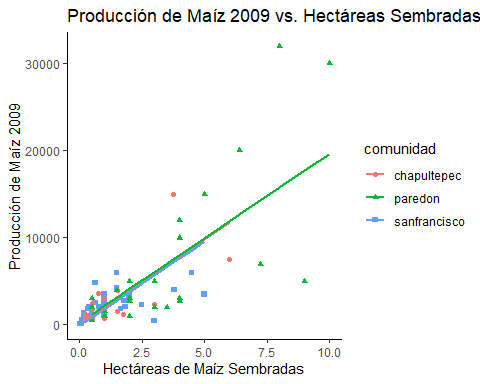
[1] <NA> sanfrancisco sanfrancisco sanfrancisco sanfrancisco  
 [6] sanfrancisco sanfrancisco <NA> sanfrancisco sanfrancisco  
 [11] sanfrancisco sanfrancisco <NA> sanfrancisco sanfrancisco  
 [16] <NA> sanfrancisco <NA> sanfrancisco <NA>   
 [21] <NA> sanfrancisco sanfrancisco sanfrancisco sanfrancisco  
 [26] sanfrancisco sanfrancisco sanfrancisco sanfrancisco sanfrancisco  
 [31] <NA> sanfrancisco sanfrancisco <NA> <NA>   
 [36] <NA> <NA> <NA> <NA> <NA>   
 [41] sanfrancisco sanfrancisco sanfrancisco sanfrancisco sanfrancisco  
 [46] sanfrancisco sanfrancisco <NA> sanfrancisco sanfrancisco  
 [51] sanfrancisco sanfrancisco <NA> sanfrancisco <NA>   
 [56] chapultepec chapultepec chapultepec chapultepec chapultepec   
 [61] <NA> chapultepec chapultepec chapultepec <NA>   
 [66] <NA> chapultepec <NA> <NA> chapultepec   
 [71] chapultepec chapultepec chapultepec chapultepec chapultepec   
 [76] chapultepec <NA> <NA> <NA> <NA>   
 [81] <NA> chapultepec <NA> <NA> <NA>   
 [86] <NA> <NA> <NA> <NA> <NA>   
 [91] <NA> <NA> <NA> <NA> <NA>   
 [96] <NA> <NA> <NA> chapultepec chapultepec   
[101] chapultepec chapultepec paredon paredon paredon   
[106] paredon paredon paredon paredon <NA>   
[111] paredon <NA> paredon paredon <NA>   
[116] paredon paredon <NA> paredon <NA>   
[121] <NA> <NA> <NA> <NA> <NA>   
[126] <NA> paredon paredon <NA> <NA>   
[131] <NA> <NA> <NA> paredon paredon   
[136] <NA> <NA> <NA> paredon paredon   
[141] paredon paredon paredon paredon paredon   
Levels: chapultepec paredon sanfrancisco

Call:  
lm(formula = produccion\_2009 ~ ha\_maiz + dummy\_comunidad\_b +   
 dummy\_comunidad\_c, data = toluca\_dummy\_casewhen)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-12704.3 -1197.1 224.6 1079.8 16232.1   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 121.8 849.0 0.143 0.886   
ha\_maiz 1936.4 223.6 8.659 4.96e-13 \*\*\*  
dummy\_comunidad\_b 154.8 1209.1 0.128 0.898   
dummy\_comunidad\_c -220.5 1000.5 -0.220 0.826   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 3661 on 78 degrees of freedom  
Multiple R-squared: 0.5747, Adjusted R-squared: 0.5583   
F-statistic: 35.13 on 3 and 78 DF, p-value: 1.799e-14

Como se observa, ambos enfoques dan exactamente el mismo resultado.

Los coeficientes de los modelos corresponden a las hectáreas sembradas y para las diferencias entre las comunidades (ajustadas por las hectáreas). Esto nos permite interpretar cómo difiere la producción entre comunidades, controlando por el tamaño del área sembrada.

Visualización del modelo con las Rectas de Regresión para cada comunidad:

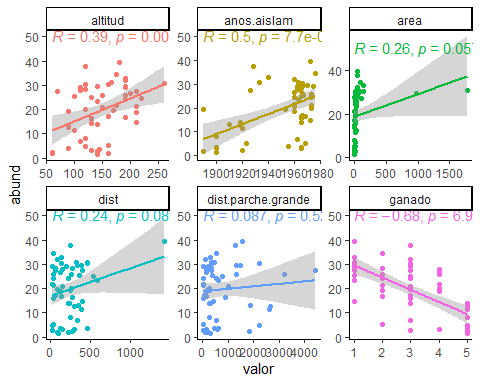


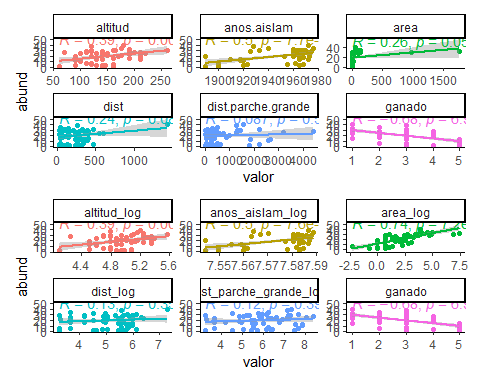
## 5 Aves de Australia

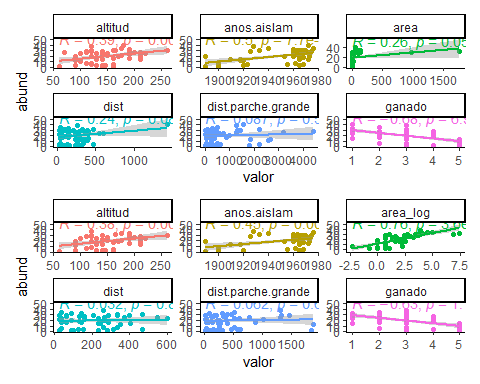
Entender qué aspectos del ambiente y de las actividades humanas afectan la abundancia de organismos es un aspecto muy importante para la conservación de áreas naturales. En 1987 se colectaron datos de aves en 56 parches de vegetación natural en Australia. En este estudio se registró la abundancia de aves (abundancia) y algunas variables que serán utilizadas como predictoras de esta abundancia, incluyendo el área del parche medido (area), el tiempo en años en que dicho parche ha quedado aislado del resto de la vegetación natural (anos.aislam), la distancia al parche de vegetación más cercano (dist), la distancia al parche más grande de vegetación en el área (dist.parche.grande), la cantidad de ganado presente en el parche (ganado, medido de 1 a 5 donde 1 es poco ganado y 5 es abundante ganado), y la altitud. Se tienen en total seis variables predictoras. OJO: Revisa que exista una relación lineal entre la variable de respuesta y las predictoras. Es posible que algunas requieran una transformación.

### 5.1 Relación Lineal y transformaciones

1. Hacemos un gráfico con las 6 variables predictoras y añadimos una correlación de Pearson para evaluar la relación lineal.







'data.frame': 55 obs. of 9 variables:  
 $ abund : num 5.3 2 1.5 17.1 13.8 14.1 3.8 2.2 3.3 3 ...  
 $ area : num 0.1 0.5 0.5 1 1 1 1 1 1 1 ...  
 $ anos.aislam : int 1968 1920 1900 1966 1918 1965 1955 1920 1965 1900 ...  
 $ dist : int 39 234 104 66 246 234 467 284 156 311 ...  
 $ dist.parche.grande: int 39 234 311 66 246 285 467 1829 156 571 ...  
 $ ganado : int 2 5 5 3 5 3 5 5 4 5 ...  
 $ altitud : int 160 60 140 160 140 130 90 60 130 130 ...  
 $ area\_log : num -2.303 -0.693 -0.693 0 0 ...  
 $ ganado\_factor : Factor w/ 5 levels "1","2","3","4",..: 2 5 5 3 5 3 5 5 4 5 ...

### 5.2 lm con todos los predictores

Modelo para predecir abundancia a partir de todas las variables:

Call:  
lm(formula = abund ~ altitud + anos.aislam + area\_log + dist +   
 dist.parche.grande + ganado\_factor, data = aves\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-16.3890 -1.5523 0.2463 2.9852 10.0335   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 5.271e+01 1.131e+02 0.466 0.6435   
altitud -2.915e-03 2.431e-02 -0.120 0.9051   
anos.aislam -1.857e-02 5.723e-02 -0.325 0.7471   
area\_log 3.257e+00 6.300e-01 5.170 5.22e-06 \*\*\*  
dist -3.615e-03 7.000e-03 -0.516 0.6081   
dist.parche.grande -2.391e-04 1.101e-03 -0.217 0.8290   
ganado\_factor2 8.519e-01 3.225e+00 0.264 0.7929   
ganado\_factor3 3.395e-03 2.899e+00 0.001 0.9991   
ganado\_factor4 -1.199e+00 3.222e+00 -0.372 0.7115   
ganado\_factor5 -1.197e+01 4.576e+00 -2.615 0.0121 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.895 on 45 degrees of freedom  
Multiple R-squared: 0.7362, Adjusted R-squared: 0.6834   
F-statistic: 13.95 on 9 and 45 DF, p-value: 2.227e-10

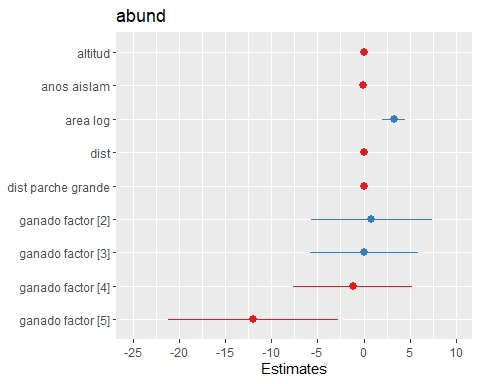
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Aves   |  |  |  |  | | --- | --- | --- | --- | |  | abund | | | | Predictors | Estimates | CI | p | | (Intercept) | 52.71 | -175.10 – 280.51 | 0.643 | | altitud | -0.00 | -0.05 – 0.05 | 0.905 | | anos.aislam | -0.02 | -0.13 – 0.10 | 0.747 | | area\_log | 3.26 \*\*\* | 1.99 – 4.53 | **<0.001** | | dist | -0.00 | -0.02 – 0.01 | 0.608 | | dist.parche.grande | -0.00 | -0.00 – 0.00 | 0.829 | | ganado\_factor2 | 0.85 | -5.64 – 7.35 | 0.793 | | ganado\_factor3 | 0.00 | -5.84 – 5.84 | 0.999 | | ganado\_factor4 | -1.20 | -7.69 – 5.29 | 0.712 | | ganado\_factor5 | -11.97 \* | -21.18 – -2.75 | **0.012** | | Observations | 55 | | | | R2 / R2 adjusted | 0.736 / 0.683 | | | | \* p<0.05   \*\* p<0.01   \*\*\* p<0.001 | | | |   **Table** **:** Aves   | **Characteristic** | **Beta** | **95% CI**1 | **p-value** | | --- | --- | --- | --- | | altitud | 0.00 | -0.05, 0.05 | 0.905 | | anos.aislam | -0.02 | -0.13, 0.10 | 0.747 | | area\_log | 3.3 | 2.0, 4.5 | **<0.001** | | dist | 0.00 | -0.02, 0.01 | 0.608 | | dist.parche.grande | 0.00 | 0.00, 0.00 | 0.829 | | ganado\_factor |  |  |  | | ganado\_factor2 - ganado\_factor1 | 0.85 | -8.3, 10 | 0.999 | | ganado\_factor3 - ganado\_factor1 | 0.00 | -8.2, 8.2 | >0.999 | | ganado\_factor3 - ganado\_factor2 | -0.85 | -8.8, 7.1 | 0.998 | | ganado\_factor4 - ganado\_factor1 | -1.2 | -10, 8.0 | 0.996 | | ganado\_factor4 - ganado\_factor2 | -2.1 | -11, 6.8 | 0.964 | | ganado\_factor4 - ganado\_factor3 | -1.2 | -9.2, 6.8 | 0.993 | | ganado\_factor5 - ganado\_factor1 | -12 | -25, 1.0 | 0.085 | | ganado\_factor5 - ganado\_factor2 | -13 | -26, 0.57 | 0.066 | | ganado\_factor5 - ganado\_factor3 | -12 | -23, -1.4 | **0.020** | | ganado\_factor5 - ganado\_factor4 | -11 | -24, 2.4 | 0.156 | | 1CI = Confidence Interval | | | | |

Ajustamos un modelo lineal (estimado usando OLS) para predecir abundancia a partir de altitud, anos.aislam, area\_log, dist, dist.parche.grande y ganado (fórmula: abund ~ altitud + anos.aislam + area\_log + dist + dist.parche.grande + ganado). El modelo explica una proporción de varianza estadísticamente significativa y sustancial (R2 = 0.70, F(6, 48) = 18.36, p < .001, R2 ajustado = 0.66). El intercepto del modelo, correspondiente a altitud = 0, anos.aislam = 0, area\_log = 0, dist = 0, dist.parche.grande = 0 y ganado = 0, está en -115.32 (IC del 95% [-286.79, 56.16], t(48) = -1.35, p = 0.183). Dentro de este modelo:

* El efecto de la ***altitud*** es estadísticamente **no** significativo y positivo (beta = 1.39e-03, IC del 95% [-0.05, 0.05], t(48) = 0.06, p = 0.955; beta estandarizada = 5.80e-03, IC del 95% [-0.20, 0.21])
* El efecto de ***anos aislam*** es estadísticamente **no** significativo y positivo (beta = 0.07, IC del 95% [-0.02, 0.15], t(48) = 1.63, p = 0.110; beta estandarizada = 0.17, IC del 95% [-0.04, 0.38])
* El efecto de ***area log*** es estadísticamente significativo y positivo (beta = 3.53, IC del 95% [2.30, 4.76], t(48) = 5.78, p < .001; beta estandarizada = 0.63, IC del 95% [0.41, 0.85])
* El efecto de ***dist*** es estadísticamente **no** significativo y negativo (beta = -0.01, IC del 95% [-0.02, 2.93e-03], t(48) = -1.56, p = 0.125; beta estandarizada = -0.15, IC del 95% [-0.33, 0.04])
* El efecto de ***dist parche grande*** es estadísticamente **no** significativo y negativo (beta = -9.26e-04, IC del 95% [-3.13e-03, 1.27e-03], t(48) = -0.85, p = 0.402; beta estandarizada = -0.08, IC del 95% [-0.27, 0.11])
* El efecto de ***ganado*** es estadísticamente **no** significativo y negativo (beta = -1.72, IC del 95% [-3.53, 0.09], t(48) = -1.91, p = 0.062; beta estandarizada = -0.24, IC del 95% [-0.49, 0.01])

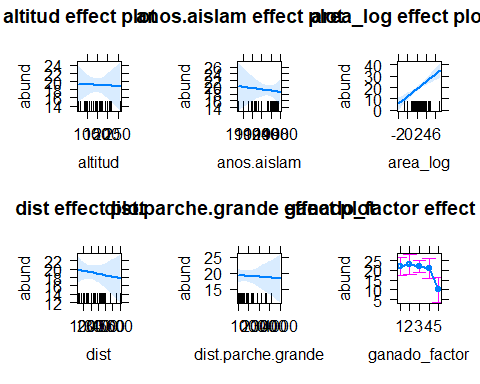
Los parámetros estandarizados se obtuvieron ajustando el modelo en una versión estandarizada del conjunto de datos. Los intervalos de confianza del 95% (IC) y los valores p se calcularon utilizando una aproximación de la distribución t de Wald.

**Grafico de los estimados del modelo con IC**



* Podemos observar cómo el logarítmo del área del parche medido tiene un efecto positivo y significativo sobre la abundancia.
* También se observa cómo el nivel 5 del factor de ganado tiene un efecto negativo y significativo sobre la abundancia.

Podemos graficar los efectos de cada predictor sobre la variable de respuesta:



* Igual se observa cómo area\_log y el nivel 5 de ganado tienen una mayor influencia sobre la abundancia.

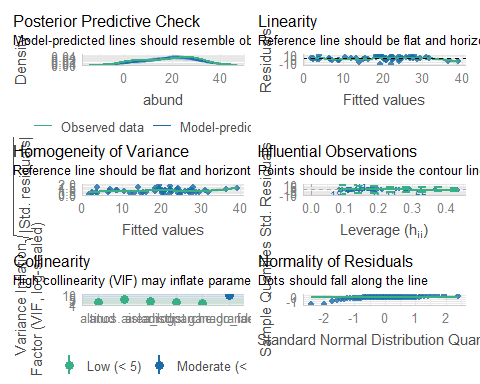
Los valores predichos para estos efectos son los siguientes:

$altitud  
# Predicted values of abund  
  
altitud | Predicted | 95% CI  
------------------------------------  
 60 | 19.40 | [14.91, 23.89]  
 90 | 19.31 | [16.15, 22.47]  
 120 | 19.22 | [17.19, 21.26]  
 140 | 19.17 | [17.54, 20.79]  
 150 | 19.14 | [17.52, 20.75]  
 165 | 19.09 | [17.23, 20.95]  
 175 | 19.06 | [16.91, 21.22]  
 260 | 18.82 | [12.99, 24.64]  
  
$anos.aislam  
# Predicted values of abund  
  
anos.aislam | Predicted | 95% CI  
----------------------------------------  
 1890 | 20.25 | [13.23, 27.28]  
 1900 | 20.07 | [14.16, 25.97]  
 1910 | 19.88 | [15.07, 24.69]  
 1920 | 19.69 | [15.95, 23.44]  
 1940 | 19.32 | [17.39, 21.25]  
 1950 | 19.14 | [17.53, 20.74]  
 1960 | 18.95 | [16.93, 20.97]  
 1980 | 18.58 | [14.70, 22.46]  
  
$area\_log  
# Predicted values of abund  
  
area\_log | Predicted | 95% CI  
-------------------------------------  
 -3 | 2.49 | [-4.19, 9.18]  
 -2 | 5.75 | [ 0.29, 11.21]  
 0 | 12.26 | [ 9.14, 15.39]  
 1 | 15.52 | [13.38, 17.66]  
 2 | 18.78 | [17.17, 20.38]  
 4 | 25.29 | [22.41, 28.17]  
 5 | 28.55 | [24.55, 32.55]  
 8 | 38.32 | [30.68, 45.96]  
  
$dist  
# Predicted values of abund  
  
dist | Predicted | 95% CI  
---------------------------------  
 25 | 19.85 | [16.68, 23.02]  
 95 | 19.60 | [17.23, 21.96]  
 170 | 19.33 | [17.58, 21.07]  
 240 | 19.07 | [17.44, 20.70]  
 310 | 18.82 | [16.77, 20.87]  
 380 | 18.57 | [15.79, 21.35]  
 450 | 18.31 | [14.68, 21.95]  
 595 | 17.79 | [12.25, 23.33]  
  
$dist.parche.grande  
# Predicted values of abund  
  
dist.parche.grande | Predicted | 95% CI  
-----------------------------------------------  
 0 | 19.32 | [17.06, 21.58]  
 550 | 19.19 | [17.55, 20.83]  
 1100 | 19.06 | [17.25, 20.87]  
 1650 | 18.93 | [16.31, 21.54]  
 2250 | 18.78 | [15.03, 22.54]  
 2800 | 18.65 | [13.77, 23.54]  
 3350 | 18.52 | [12.47, 24.57]  
 4450 | 18.26 | [ 9.83, 26.68]  
  
$ganado\_factor  
# Predicted values of abund  
  
ganado\_factor | Predicted | 95% CI  
------------------------------------------  
1 | 22.01 | [17.31, 26.70]  
2 | 22.86 | [17.81, 27.90]  
3 | 22.01 | [18.74, 25.27]  
4 | 20.81 | [15.77, 25.84]  
5 | 10.04 | [ 3.77, 16.30]  
  
attr(,"class")  
[1] "ggalleffects" "list"   
attr(,"model.name")  
[1] "lm\_aves\_todos"

Ahora, vamos a ver el desempeño de nuestro modelo:

# Indices of model performance  
  
AIC | AICc | BIC | R2 | R2 (adj.) | RMSE | Sigma  
---------------------------------------------------------------  
362.193 | 368.332 | 384.273 | 0.736 | 0.683 | 5.332 | 5.895

Ver los supuestos:



### 5.3 Backward

Utiliza un procedimiento de eliminación backward comenzando con el modelo saturado (con las seis variables) para proponer un modelo múltiple que prediga la abundancia a partir de un subconjunto de las seis variables predictoras.

1. Modelo Saturado

Call:  
lm(formula = abund ~ altitud + anos.aislam + area\_log + dist +   
 dist.parche.grande + ganado\_factor, data = aves\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-16.3890 -1.5523 0.2463 2.9852 10.0335   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 5.271e+01 1.131e+02 0.466 0.6435   
altitud -2.915e-03 2.431e-02 -0.120 0.9051   
anos.aislam -1.857e-02 5.723e-02 -0.325 0.7471   
area\_log 3.257e+00 6.300e-01 5.170 5.22e-06 \*\*\*  
dist -3.615e-03 7.000e-03 -0.516 0.6081   
dist.parche.grande -2.391e-04 1.101e-03 -0.217 0.8290   
ganado\_factor2 8.519e-01 3.225e+00 0.264 0.7929   
ganado\_factor3 3.395e-03 2.899e+00 0.001 0.9991   
ganado\_factor4 -1.199e+00 3.222e+00 -0.372 0.7115   
ganado\_factor5 -1.197e+01 4.576e+00 -2.615 0.0121 \*   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.895 on 45 degrees of freedom  
Multiple R-squared: 0.7362, Adjusted R-squared: 0.6834   
F-statistic: 13.95 on 9 and 45 DF, p-value: 2.227e-10

1. Hacemos la eliminación Backward utilizando el criterio AIC:
   * step() evalúa la contribución de cada variable al modelo
   * en cada paso, intenta eliminar la variable que menos contribuye a la explicación de la variabilidad

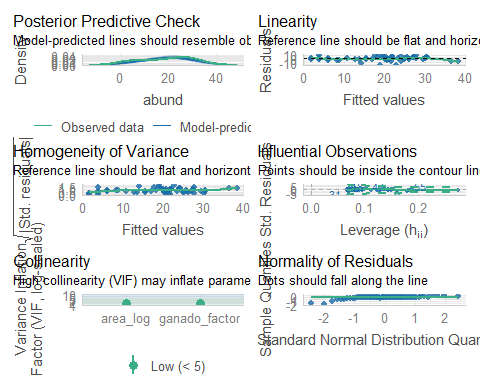
Start: AIC=204.11  
abund ~ altitud + anos.aislam + area\_log + dist + dist.parche.grande +   
 ganado\_factor  
  
 Df Sum of Sq RSS AIC  
- altitud 1 0.50 1564.2 202.13  
- dist.parche.grande 1 1.64 1565.3 202.17  
- anos.aislam 1 3.66 1567.3 202.24  
- dist 1 9.26 1572.9 202.43  
<none> 1563.7 204.11  
- ganado\_factor 4 372.53 1936.2 207.86  
- area\_log 1 928.69 2492.3 227.75  
  
Step: AIC=202.13  
abund ~ anos.aislam + area\_log + dist + dist.parche.grande +   
 ganado\_factor  
  
 Df Sum of Sq RSS AIC  
- dist.parche.grande 1 1.24 1565.4 200.17  
- anos.aislam 1 3.28 1567.4 200.24  
- dist 1 8.79 1572.9 200.44  
<none> 1564.2 202.13  
- ganado\_factor 4 386.97 1951.1 206.28  
- area\_log 1 973.44 2537.6 226.74  
  
Step: AIC=200.17  
abund ~ anos.aislam + area\_log + dist + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
- anos.aislam 1 3.92 1569.3 198.31  
- dist 1 10.01 1575.4 198.52  
<none> 1565.4 200.17  
- ganado\_factor 4 441.68 2007.1 205.84  
- area\_log 1 1116.52 2681.9 227.78  
  
Step: AIC=198.31  
abund ~ area\_log + dist + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
- dist 1 14.39 1583.7 196.81  
<none> 1569.3 198.31  
- ganado\_factor 4 966.85 2536.2 216.71  
- area\_log 1 1180.81 2750.1 227.16  
  
Step: AIC=196.81  
abund ~ area\_log + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
<none> 1583.7 196.81  
- ganado\_factor 4 1117.1 2700.8 218.17  
- area\_log 1 1170.0 2753.7 225.24

Call:  
lm(formula = abund ~ area\_log + ganado\_factor, data = aves\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-16.0684 -1.9310 -0.0048 3.1527 10.0316   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 14.6537 2.7205 5.386 2.03e-06 \*\*\*  
area\_log 3.1700 0.5269 6.017 2.21e-07 \*\*\*  
ganado\_factor2 1.4033 2.8549 0.492 0.625242   
ganado\_factor3 0.8320 2.5103 0.331 0.741732   
ganado\_factor4 -0.5752 2.9155 -0.197 0.844403   
ganado\_factor5 -10.8489 2.8747 -3.774 0.000434 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.685 on 49 degrees of freedom  
Multiple R-squared: 0.7328, Adjusted R-squared: 0.7055   
F-statistic: 26.88 on 5 and 49 DF, p-value: 5.708e-13

abund ~ area\_log + ganado\_factor

Validación del Modelo

# Indices of model performance  
  
AIC | AICc | BIC | R2 | R2 (adj.) | RMSE | Sigma  
---------------------------------------------------------------  
354.893 | 357.276 | 368.945 | 0.733 | 0.706 | 5.366 | 5.685



### 5.4 Forward

Utiliza un procedimiento de selección forward comenzando con un modelo que solamente contenga al intercepto para proponer un modelo múltiple que prediga la abundancia a partir de un subconjunto de las seis variables predictoras.

1. Modelo Nulo (solo intercepto)

Call:  
lm(formula = abund ~ 1, data = aves\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-17.649 -7.049 1.751 8.901 18.551   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 19.149 1.413 13.55 <2e-16 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 10.48 on 54 degrees of freedom

1. Selección *Forward*
   * lower = modelo más simple
   * upper el modelo más complejo

Start: AIC=259.4  
abund ~ 1  
  
 Df Sum of Sq RSS AIC  
+ area\_log 1 3226.4 2700.8 218.17  
+ ganado\_factor 4 3173.4 2753.7 225.24  
+ anos.aislam 1 1439.5 4487.6 246.10  
+ altitud 1 825.9 5101.2 253.15  
<none> 5927.2 259.40  
+ dist 1 32.6 5894.5 261.09  
+ dist.parche.grande 1 20.2 5907.0 261.21  
  
Step: AIC=218.17  
abund ~ area\_log  
  
 Df Sum of Sq RSS AIC  
+ ganado\_factor 4 1117.09 1583.7 196.81  
+ anos.aislam 1 557.50 2143.3 207.45  
+ dist.parche.grande 1 220.75 2480.0 215.48  
+ altitud 1 200.91 2499.9 215.92  
+ dist 1 164.63 2536.2 216.71  
<none> 2700.8 218.17  
  
Step: AIC=196.81  
abund ~ area\_log + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
<none> 1583.7 196.81  
+ dist 1 14.3945 1569.3 198.31  
+ anos.aislam 1 8.3041 1575.4 198.52  
+ dist.parche.grande 1 4.2092 1579.5 198.66  
+ altitud 1 2.2214 1581.5 198.73

Call:  
lm(formula = abund ~ area\_log + ganado\_factor, data = aves\_final)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-16.0684 -1.9310 -0.0048 3.1527 10.0316   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 14.6537 2.7205 5.386 2.03e-06 \*\*\*  
area\_log 3.1700 0.5269 6.017 2.21e-07 \*\*\*  
ganado\_factor2 1.4033 2.8549 0.492 0.625242   
ganado\_factor3 0.8320 2.5103 0.331 0.741732   
ganado\_factor4 -0.5752 2.9155 -0.197 0.844403   
ganado\_factor5 -10.8489 2.8747 -3.774 0.000434 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.685 on 49 degrees of freedom  
Multiple R-squared: 0.7328, Adjusted R-squared: 0.7055   
F-statistic: 26.88 on 5 and 49 DF, p-value: 5.708e-13

abund ~ area\_log + ganado\_factor

En este caso, ambos modelos (*Forward* y *Backward*) dieron el mismo resultado.

### 5.5 Modelado II

Utiliza el procedimiento de selección de modelos basado en el AIC para seleccionar un grupo de modelos que predigan la abundancia a partir de un subconjunto de las variables predictoras. Puedes acudir a Burnham et al. 2011 y a Symonds & Moussalli 2011 y al ejercicio que realizamos en clase para revisar de nuevo esta aproximación. Para hacer el ejercicio más sencillo (es decir, para reducir el número de modelos a explorar), utiliza solamente el siguiente conjunto de cuatro variables predictoras en lugar de las seis de los modelos de los incisos previos: *área, ganado, distancia al parche más grande y altitud*. ***Explora todos los modelos posibles*** para estas cuatro variables y realiza las comparaciones que sean necesarias a través del valor de AIC de los modelos. ¿Cuál sería el conjunto que propondrías como buenos modelos? ¿Cuál es la importancia relativa de las diferentes variables tomando en cuenta su aparición en los diferentes modelos candidatos?

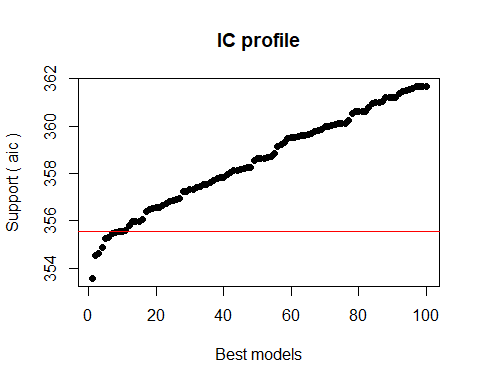
Para explorar TODOS los modelos (incluidas las interacciones), podemos usar el siguiente código a partir del paquete [glmulti](https://www.jstatsoft.org/article/view/v034i12) (Calcagno and Mazancourt 2010).

Initialization...  
TASK: Genetic algorithm in the candidate set.  
Initialization...  
Algorithm started...  
  
After 10 generations:  
Best model: abund~1+altitud+area\_log+area\_log:altitud+ganado\_factor:altitud+ganado\_factor:area\_log  
Crit= 355.307420351296  
Mean crit= 370.996269030846  
Change in best IC: -9644.6925796487 / Change in mean IC: -9629.00373096915  
  
After 20 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 368.163760666855  
Change in best IC: -1.73417773376082 / Change in mean IC: -2.83250836399196  
  
After 30 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 365.482999566207  
Change in best IC: 0 / Change in mean IC: -2.6807611006476  
  
After 40 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 363.748040956646  
Change in best IC: 0 / Change in mean IC: -1.73495860956109  
  
After 50 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 362.641810018036  
Change in best IC: 0 / Change in mean IC: -1.10623093861022  
  
After 60 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 362.033674668219  
Change in best IC: 0 / Change in mean IC: -0.608135349816394  
  
After 70 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 361.513084316441  
Change in best IC: 0 / Change in mean IC: -0.520590351778594  
  
After 80 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 360.962718420006  
Change in best IC: 0 / Change in mean IC: -0.550365896434585  
  
After 90 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 360.541178512359  
Change in best IC: 0 / Change in mean IC: -0.421539907646604  
  
After 100 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 360.281389577253  
Change in best IC: 0 / Change in mean IC: -0.259788935106542  
  
After 110 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 359.745791867125  
Change in best IC: 0 / Change in mean IC: -0.535597710128343  
  
After 120 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 359.487870817113  
Change in best IC: 0 / Change in mean IC: -0.257921050011305  
  
After 130 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 359.465179407351  
Change in best IC: 0 / Change in mean IC: -0.0226914097619897  
  
After 140 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 359.216542817796  
Change in best IC: 0 / Change in mean IC: -0.248636589554962  
  
After 150 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 359.209329100514  
Change in best IC: 0 / Change in mean IC: -0.00721371728235454  
  
After 160 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 359.138868137902  
Change in best IC: 0 / Change in mean IC: -0.0704609626124011  
  
After 170 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 359.021980443156  
Change in best IC: 0 / Change in mean IC: -0.11688769474523  
  
After 180 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.94024160627  
Change in best IC: 0 / Change in mean IC: -0.0817388368861316  
  
After 190 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.863011375598  
Change in best IC: 0 / Change in mean IC: -0.0772302306725692  
  
After 200 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.80356099733  
Change in best IC: 0 / Change in mean IC: -0.0594503782676838  
  
After 210 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.80356099733  
Change in best IC: 0 / Change in mean IC: 0  
  
After 220 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.702519355974  
Change in best IC: 0 / Change in mean IC: -0.101041641355437  
  
After 230 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.663742557969  
Change in best IC: 0 / Change in mean IC: -0.0387767980055287  
  
After 240 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.6242195515  
Change in best IC: 0 / Change in mean IC: -0.0395230064690395  
  
After 250 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.597149075154  
Change in best IC: 0 / Change in mean IC: -0.027070476345898  
  
After 260 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.554853636353  
Change in best IC: 0 / Change in mean IC: -0.0422954388006929  
  
After 270 generations:  
Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log  
Crit= 353.573242617535  
Mean crit= 358.542036065938  
Improvements in best and average IC have bebingo en below the specified goals.  
Algorithm is declared to have converged.  
Completed.

2.31 sec elapsed

Ahora revisamos los modelos

glmulti.analysis  
Method: g / Fitting: lm / IC used: aic  
Level: 2 / Marginality: FALSE  
From 100 models:  
Best IC: 353.573242617535  
Best model:  
[1] "abund ~ 1 + ganado\_factor + area\_log + ganado\_factor:area\_log"  
Evidence weight: 0.0713402729785392  
Worst IC: 361.683502604567  
10 models within 2 IC units.  
72 models to reach 95% of evidence weight.  
Convergence after 270 generations.  
Time elapsed: 2.18971490859985 minutes.



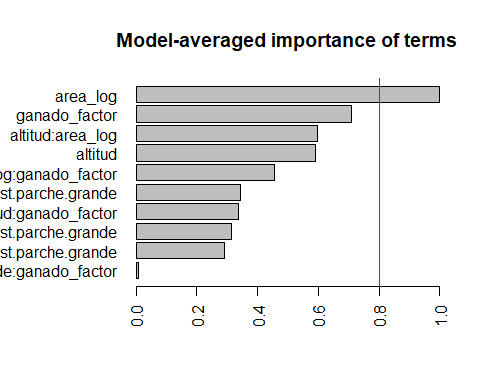
Call:  
fitfunc(formula = as.formula(x), data = data, family = ..1)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-16.3615 -2.2072 -0.2545 2.7473 10.8225   
  
Coefficients:  
 Estimate Std. Error t value Pr(>|t|)   
(Intercept) 20.0335 3.8754 5.169 5.22e-06 \*\*\*  
ganado\_factor2 -4.9559 4.6911 -1.056 0.296397   
ganado\_factor3 -6.0055 4.6704 -1.286 0.205072   
ganado\_factor4 -16.7005 6.4704 -2.581 0.013181 \*   
ganado\_factor5 -15.8335 4.2756 -3.703 0.000579 \*\*\*  
area\_log 1.8637 0.8600 2.167 0.035563 \*   
ganado\_factor2:area\_log 1.8332 1.2989 1.411 0.165007   
ganado\_factor3:area\_log 2.1029 1.4755 1.425 0.160994   
ganado\_factor4:area\_log 6.5526 2.4754 2.647 0.011151 \*   
ganado\_factor5:area\_log 0.8027 1.5251 0.526 0.601217   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5.45 on 45 degrees of freedom  
Multiple R-squared: 0.7745, Adjusted R-squared: 0.7293   
F-statistic: 17.17 on 9 and 45 DF, p-value: 7.743e-12

Best model: [1] "abund ~ 1 + ganado\_factor + area\_log + ganado\_factor:area\_log"

Los mejores modelos (por criterio de AIC) propuestos son:

| model | aic | weights |
| --- | --- | --- |
| abund ~ 1 + ganado\_factor + area\_log + ganado\_factor:area\_log | 353.5732 | 0.07134027 |
| abund ~ 1 + altitud + area\_log + dist.parche.grande + area\_log:altitud + ganado\_factor:altitud | 354.5491 | 0.04379596 |
| abund ~ 1 + altitud + area\_log + area\_log:altitud + dist.parche.grande:area\_log + ganado\_factor:altitud | 354.6147 | 0.04238147 |
| abund ~ 1 + ganado\_factor + area\_log | 354.8935 | 0.03686776 |
| abund ~ 1 + ganado\_factor + altitud + area\_log + ganado\_factor:area\_log | 355.2517 | 0.03082143 |
| abund ~ 1 + altitud + area\_log + area\_log:altitud + ganado\_factor:altitud + ganado\_factor:area\_log | 355.3074 | 0.02997525 |

La importancia relativa de los términos es la siguiente:



|  |  |
| --- | --- |
|  | **Con efectos Mixtos**  También podríamos evaluarlo con efectos mixtos.   * Podríamos considerar altitud como un efecto aletorio; si la influencia de esta variable sobre la abundancia es variada, podría ser un efecto aleatorio. * Tal vez ganado podría considerarse. Aunque suena más a un efecto fijo, si tenemos varios registros por nivel de ganado o si nos interesa la variabilidad de la respuesta de la abundancia de aves a diferentes niveles de manejo de ganado en varios sitios, podríamos tomarlo como un efecto aleatorio.   Initialization... TASK: Exhaustive screening of candidate set. Fitting...  After 50 models: Best model: abund~1+ganado\_factor+area\_log+ganado\_factor:area\_log Crit= 330.904738800118 Mean crit= 385.487585313527 Completed.  0.89 sec elapsed  glmulti.analysis Method: h / Fitting: glmer.glmulti / IC used: aic Level: 2 / Marginality: FALSE From 36 models: Best IC: 330.904738800118 Best model: [1] "abund ~ 1 + ganado\_factor + area\_log + ganado\_factor:area\_log" Evidence weight: 0.994012070053268 Worst IC: 454.757938515849 1 models within 2 IC units. 0 models to reach 95% of evidence weight.    Como vemos, parece ser que en este caso llegamos a la misma fórmula y no es necesario considerar los efectos mixtos para mejorar nuestro modelo.  abund ~ 1 + ganado\_factor + area\_log + ganado\_factor:area\_log |

Ahora comparamos nuestro modelo óptimo con los modelos backward y forward:

Start: AIC=200.64  
abund ~ altitud + area\_log + dist.parche.grande + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
- altitud 1 0.61 1579.5 198.66  
- dist.parche.grande 1 2.60 1581.5 198.73  
<none> 1578.9 200.64  
- ganado\_factor 4 830.76 2409.6 215.90  
- area\_log 1 965.53 2544.4 224.89  
  
Step: AIC=198.66  
abund ~ area\_log + dist.parche.grande + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
- dist.parche.grande 1 4.21 1583.7 196.81  
<none> 1579.5 198.66  
- ganado\_factor 4 900.55 2480.0 215.48  
- area\_log 1 1038.06 2617.6 224.45  
  
Step: AIC=196.81  
abund ~ area\_log + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
<none> 1583.7 196.81  
- ganado\_factor 4 1117.1 2700.8 218.17  
- area\_log 1 1170.0 2753.7 225.24

Start: AIC=259.4  
abund ~ 1  
  
 Df Sum of Sq RSS AIC  
+ area\_log 1 3226.4 2700.8 218.17  
+ ganado\_factor 4 3173.4 2753.7 225.24  
+ altitud 1 825.9 5101.2 253.15  
<none> 5927.2 259.40  
+ dist.parche.grande 1 20.2 5907.0 261.21  
  
Step: AIC=218.17  
abund ~ area\_log  
  
 Df Sum of Sq RSS AIC  
+ ganado\_factor 4 1117.09 1583.7 196.81  
+ dist.parche.grande 1 220.75 2480.0 215.48  
+ altitud 1 200.91 2499.9 215.92  
<none> 2700.8 218.17  
  
Step: AIC=196.81  
abund ~ area\_log + ganado\_factor  
  
 Df Sum of Sq RSS AIC  
<none> 1583.7 196.81  
+ dist.parche.grande 1 4.2092 1579.5 198.66  
+ altitud 1 2.2214 1581.5 198.73

Analysis of Variance Table  
  
Model 1: abund ~ area\_log + ganado\_factor  
Model 2: abund ~ area\_log + ganado\_factor  
 Res.Df RSS Df Sum of Sq Pr(>Chi)  
1 49 1583.7   
2 49 1583.7 0 0

|  |  |  |  |
| --- | --- | --- | --- |
|  | abund | | |
| Predictors | Estimates | CI | p |
| (Intercept) | 14.65 | 9.19 – 20.12 | **<0.001** |
| area log | 3.17 | 2.11 – 4.23 | **<0.001** |
| ganado factor [2] | 1.40 | -4.33 – 7.14 | 0.625 |
| ganado factor [3] | 0.83 | -4.21 – 5.88 | 0.742 |
| ganado factor [4] | -0.58 | -6.43 – 5.28 | 0.844 |
| ganado factor [5] | -10.85 | -16.63 – -5.07 | **<0.001** |
| Observations | 55 | | |
| R2 / R2 adjusted | 0.733 / 0.706 | | |

Ambos modelos convergen en la misma fórmula:

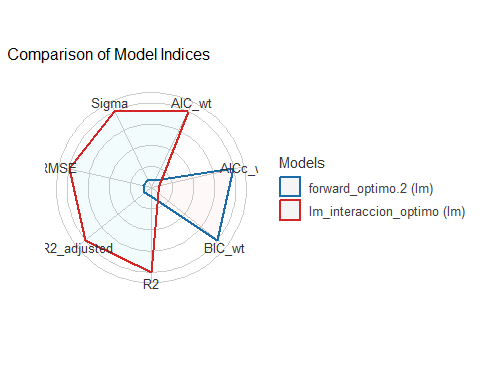
lm(formula = abund ~ area\_log + ganado\_factor, data = aves\_final)

Ahora, comparamos con nuestro modelo (“abund ~ 1 + ganado\_factor + area\_log + ganado\_factor:area\_log”):

|  |  |  |  |
| --- | --- | --- | --- |
|  | abund | | |
| Predictors | Estimates | CI | p |
| (Intercept) | 20.03 | 12.23 – 27.84 | **<0.001** |
| ganado factor [2] | -4.96 | -14.40 – 4.49 | 0.296 |
| ganado factor [3] | -6.01 | -15.41 – 3.40 | 0.205 |
| ganado factor [4] | -16.70 | -29.73 – -3.67 | **0.013** |
| ganado factor [5] | -15.83 | -24.44 – -7.22 | **0.001** |
| area log | 1.86 | 0.13 – 3.60 | **0.036** |
| ganado factor [2] × area log | 1.83 | -0.78 – 4.45 | 0.165 |
| ganado factor [3] × area log | 2.10 | -0.87 – 5.07 | 0.161 |
| ganado factor [4] × area log | 6.55 | 1.57 – 11.54 | **0.011** |
| ganado factor [5] × area log | 0.80 | -2.27 – 3.87 | 0.601 |
| Observations | 55 | | |
| R2 / R2 adjusted | 0.774 / 0.729 | | |

| Name | AIC | AICc | BIC | R2 | R2\_adjusted | RMSE |
| --- | --- | --- | --- | --- | --- | --- |
| forward\_optimo.2 | 354.8935 | 357.2765 | 368.9448 | 0.7328060 | 0.7055414 | 5.366056 |
| lm\_interaccion\_optimo | 353.5732 | 359.7128 | 375.6539 | 0.7744559 | 0.7293471 | 4.930122 |

| Name | R2 | R2\_adjusted | RMSE | Performance\_Score |
| --- | --- | --- | --- | --- |
| lm\_interaccion\_optimo | 0.7744559 | 0.7293471 | 4.930122 | 0.7142857 |
| forward\_optimo.2 | 0.7328060 | 0.7055414 | 5.366056 | 0.2857143 |



conjunto que propondrías como buenos modelos

importancia relativa de las diferentes variables tomando en cuenta su aparición en los diferentes modelos candidatos

### 5.6 Comparar modelos

Compara los resultados obtenidos con los diferentes métodos de selección de modelos. ¿A qué conclusión llegamos que conteste la pregunta principal de investigación.

La principal diferencia entre los dos modelos lineales (obtenido por método forward/backward y nuestro modelado) es la inclusión de un término de interacción en el segundo modelo.

Modelo *forward/backward* **lm(formula = abund ~ area + ganado)**

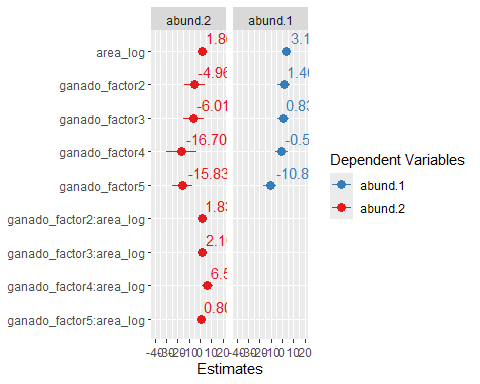
Este modelo es un modelo lineal simple que incluye dos predictores: **area** y **ganado**. Esta fórmula modela la abundancia de aves como una función lineal del área del parche y la cantidad de ganado presente, sin considerar cómo la combinación de estos dos factores podría afectar de manera conjunta la abundancia de aves. En este modelo, los coeficientes para **area** y **ganado** representan el cambio esperado en la abundancia de aves por cada unidad de cambio en **area** y **ganado**, respectivamente, asumiendo que el otro predictor se mantiene constante.

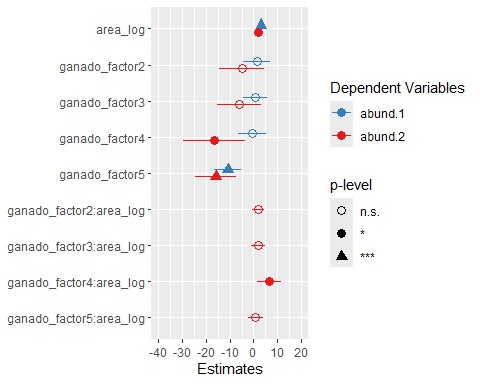
Modelo 2 **lm(formula = abund ~ 1 + ganado + area + ganado:area)**

Este modelo también incluye una interacción entre **ganado** y **area**. El término de interacción **ganado:area** examina cómo el efecto de una variable (por ejemplo, **ganado**) sobre la abundancia de aves cambia dependiendo del nivel de la otra variable (**area**), y viceversa.

* **Efectos principales**: Los coeficientes de **ganado** y **area** todavía representan el efecto de cada uno de estos predictores sobre la abundancia de aves, pero ahora estos efectos son condicionales al nivel de la otra variable siendo en su nivel base (el nivel más bajo).
* **Término de interacción**: El coeficiente de **ganado:area** indica cuánto cambia el efecto de **ganado** sobre la abundancia por cada unidad de incremento en **area**, y viceversa. Si este coeficiente es significativamente diferente de cero, implica que el efecto de una variable sobre la abundancia de aves depende del nivel de la otra variable.

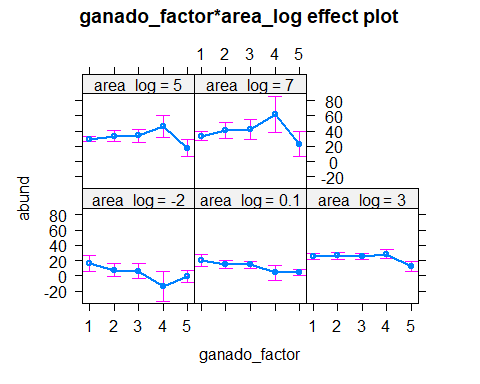
Si bien, al comparar las métricas de ambos modelos (AIC, R^2) no hay mucha diferencia, podemos pensar que el modelo con interacción nos es útil si creemos que la relación entre las variables predictoras y la respuesta no es simplemente aditiva, sino que una variable podría modificar el efecto de la otra. Podría ser que el efecto del ganado sobre la abundancia de aves sea más pronunciado en áreas más grandes o más pequeñas, lo cual no podría captarse en el primer modelo.. Además, la significancia estadística del término de interacción te nos sugiere que es necesario incluirla en nuestro modelo.

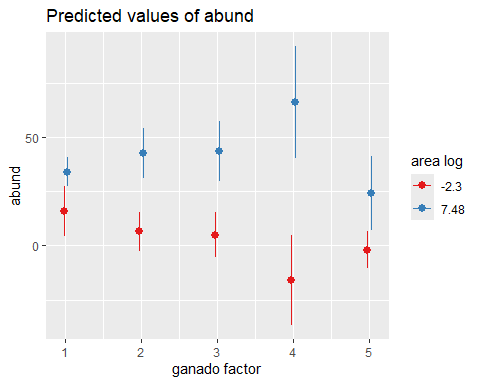




El modelo con interacción es muy similar al aditivo en casi todos los efectos principales. Sin embargo, de manera interesante, se puede observar que el nivel de ganado 4 cambia de signo cuando se considera la interacción. Además, permite ver cómo interaccionan de manera positiva los términos de interacción.

Por último, podemos visualizar los efectos de nuestro modelo de interacción:





Calcagno, Vincent, and Claire de Mazancourt. 2010. “**Glmulti**: An*R*Package for Easy Automated Model Selection with (Generalized) Linear Models.” *Journal of Statistical Software* 34 (12). <https://doi.org/10.18637/jss.v034.i12>.