

Tarea 1

Yhon Paúl Tiahualpa Yucra

10/6/2020

1. Definir todas las variables explicativas que ustedes utilizarán en el modelo. Las variables explicativas que se proponen en el modelo son:

- **x** = La coordenada x de la trampa.
- **y** = La coordenada y de la trampa.

Pero no se uso ambas variables por separado, sino que se uso una interacción entre la variable **x** y **y**

- **alt** = Altura de la trampa sobre el nivel del mar.
- **south** = 1 (La trampa se encuentra en la *zona sur*), 0 (otro caso).

Aparte de usar las variables por separado se uso su interacción, **x y**.

- **dist** = La distancia de la trampa hasta la región más caliente.
- **loc_host** = 1 (Si se encuentra un huésped dentro), 0 (otro caso).
- **A** (Se le realizó una transformación de variable $\log(1 + A)$ = Número de moscas atrapadas en la trampa.
- **W** (La cual se separó en 3 categorías) = Semana de la primera captura
 - **W1** = Si la primera captura fue la semana entre **0** a **15**
 - **W2** = Si la primera captura fue la semana entre **15** a **20**
 - **W3** = Si la primera captura fue la semana entre **20** a **más**

2. Definir el modelo de la regresión lineal adecuado para responder a la pregunta de la investigación (tienen que utilizar interacciones en el modelo). El modelo propuesto:

$$\hat{Y}_i = 2.132 - 0.04804 * S_i - 0.001277 * Alt_i - 0.04559 * Dist_i + 0.9240 * Loc_host_i + 3.552 * W1 + 3.863 * W2 + 2.685 * W3 + 8.962 * 10^{-11} * (X * Y) + 0.00445 * S_i * Alt_i$$

3. Aplicar el modelo (utilizando el software R), definido en el inciso (2) y explicar los resultados obtenidos.

```
modelo <- lm(log(1+medfly$A) ~ medfly$south + medfly$alt + medfly$dist + medfly$loc_host +  
            W1 + W2 + W3 + medfly$x*medfly$y + medfly$south*medfly$alt)
```

```
#Tabla ANVA del modelo  
summary(modelo)
```

```
##  
## Call:  
## lm(formula = log(1 + medfly$A) ~ medfly$south + medfly$alt +  
##     medfly$dist + medfly$loc_host + W1 + W2 + W3 + medfly$x *  
##     medfly$y + medfly$south * medfly$alt)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.7819 -0.7827  0.2037  0.9464  2.7954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.132e+00  3.759e+01  -0.057  0.95492
## medfly$south    -4.804e-01  1.389e+00  -0.346  0.73045
## medfly$alt      -1.277e-03  1.400e-03  -0.912  0.36452
## medfly$dist     -4.559e-02  2.550e-02  -1.788  0.07773 .
## medfly$loc_host  9.240e-01  4.418e-01   2.092  0.03976 *
## W1              3.552e+00  8.859e-01   4.010  0.00014 ***
## W2              3.863e+00  7.693e-01   5.022  3.23e-06 ***
## W3              2.685e+00  9.053e-01   2.966  0.00402 **
## medfly$x        -6.659e-05  1.887e-04  -0.353  0.72512
## medfly$y         9.891e-06  6.556e-05   0.151  0.88048
## medfly$x:medfly$y  8.962e-11  3.304e-10   0.271  0.78689
## medfly$south:medfly$alt 4.450e-03  2.074e-03   2.146  0.03506 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.539 on 77 degrees of freedom
## Multiple R-squared:  0.5719, Adjusted R-squared:  0.5107
## F-statistic: 9.351 on 11 and 77 DF, p-value: 2.041e-10
```

Se obtuvo un $R^2 = 0.5719$.