

# Alimentación en Erizos de Mar

## 1. Introducción

En esta práctica se calculará una regresión lineal en donde la variable dependiente es el ancho de las suturas de los erizos de mar y la variable independiente el tipo de alimentación y el peso inicial. Debido que hasta ahora solo se ha visto regresiones lineales con variables numéricas, se considerará únicamente en el análisis el peso inicial.

Adicionalmente se calculará: TSS, ESS, RSS,  $R^2$  y  $\hat{\sigma}^2$

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
library(tidymodels) # package that imports useful packages for modeling
library(readr)      # for importing data
library(dotwhisker) # visualize regression results

# 72 urchins ==
# experimental feeding regime group
urchins <-
  read_csv("https://raw.githubusercontent.com/savrgg/class_ITAM_metodos/main/notas_r/urchins.csv") %>%
  setNames(c("food_regime", "initial_volume", "width")) %>%
  mutate(food_regime = factor(food_regime, levels = c("Initial", "Low", "High")))

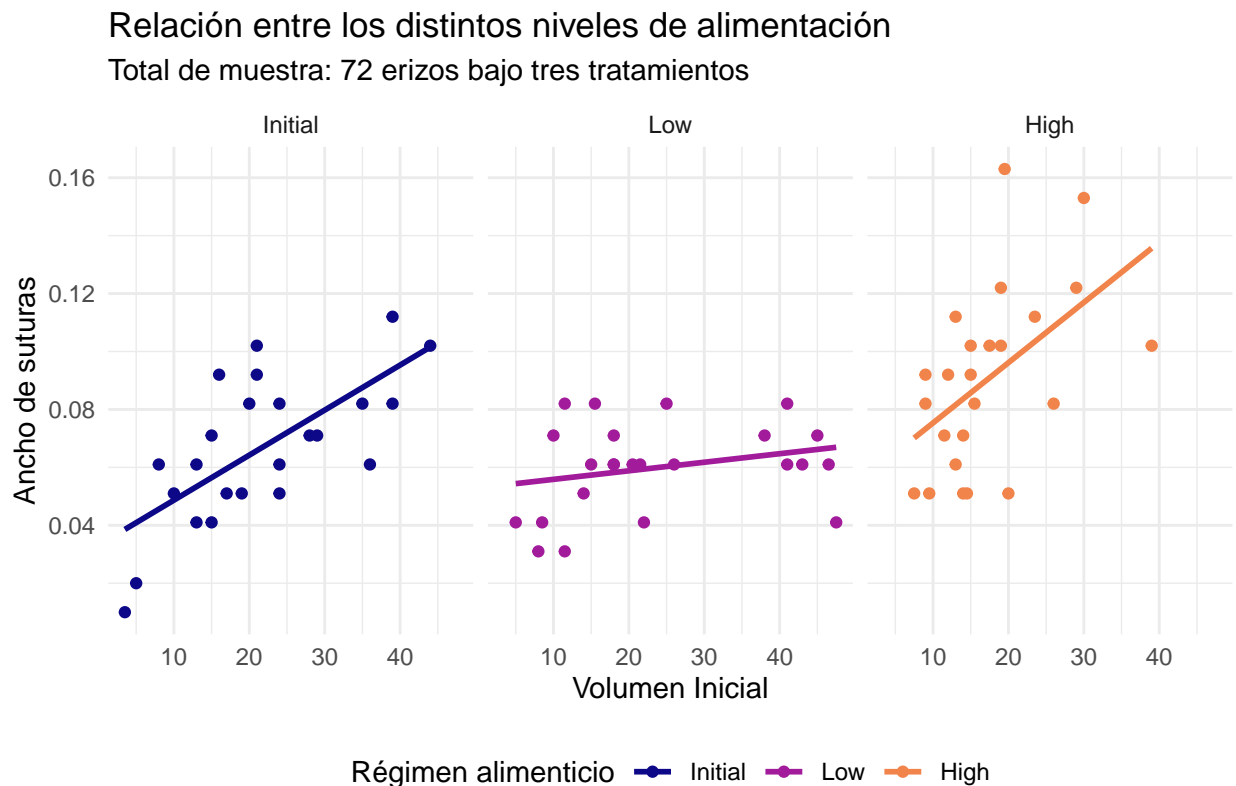
urchins
```

```
## # A tibble: 72 x 3
##   food_regime initial_volume width
##   <fct>          <dbl> <dbl>
## 1 Initial          3.5  0.01
## 2 Initial          5    0.02
## 3 Initial          8    0.061
## 4 Initial         10    0.051
## 5 Initial         13    0.041
## 6 Initial         13    0.061
## 7 Initial         15    0.041
## 8 Initial         15    0.071
## 9 Initial         16    0.092
## 10 Initial        17    0.051
## # ... with 62 more rows
```

### 1.1 Exploratory Data Analysis

Como primer paso se realizará un análisis exploratorio de datos, donde se pondrá por tipo de alimentación los puntos entre volumen inicial y el peso del erizo:

```
urchins %>%
  ggplot(aes(x = initial_volume, y = width,
             group = food_regime, col = food_regime)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
  facet_wrap(~food_regime)+
  theme_minimal()+
  theme(legend.position = "bottom")+
  labs(title = "Relación entre los distintos niveles de alimentación",
       subtitle = "Total de muestra: 72 erizos bajo tres tratamientos",
       caption = "Información recopilada de Constable, A.J., 1993",
       color = "Régimen alimenticio",
       x = "Volumen Inicial",
       y = "Ancho de suturas")+
  scale_color_viridis_d(option = "plasma", end = .7)
```



Información recopilada de Constable, A.J., 1993

En este sentido vemos que la pendiente con alimentación alta es mayor, lo cual hace sentido debido a que mejor alimentación más ancho será la sutura, en cambio, para una alimentación baja, la pendiente es cercana a 0.

## 1.2 Regresión lineal para width de erizos

Ahora se realizará la regresión con tidymodels:

```
# experimental feeding regime group (food_regime: either Initial, Low, or High),
# size in milliliters at the start of the experiment (initial_volume)
# suture width at the end of the experiment (width).
```

```
# 1 - entrenar modelo
lm_fit <-
  linear_reg() %>%
  fit(width ~ initial_volume, data = urchins)

# 2 - visualizar y graficar
glance(lm_fit)
```

```
## # A tibble: 1 x 12
##   r.squared adj.r.squared sigma statistic p.value    df logLik   AIC   BIC
##   <dbl>      <dbl> <dbl>    <dbl>   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    0.0712    0.0579 0.0270     5.37 0.0235     1   159. -312. -305.
## # ... with 3 more variables: deviance <dbl>, df.residual <int>, nobs <int>
```

```
tidy(lm_fit)
```

```
## # A tibble: 2 x 5
##   term          estimate std.error statistic  p.value
##   <chr>          <dbl>    <dbl>    <dbl>   <dbl>
## 1 (Intercept)    0.0587    0.00672     8.73 8.24e-13
## 2 initial_volume 0.000657    0.000283     2.32 2.35e- 2
```

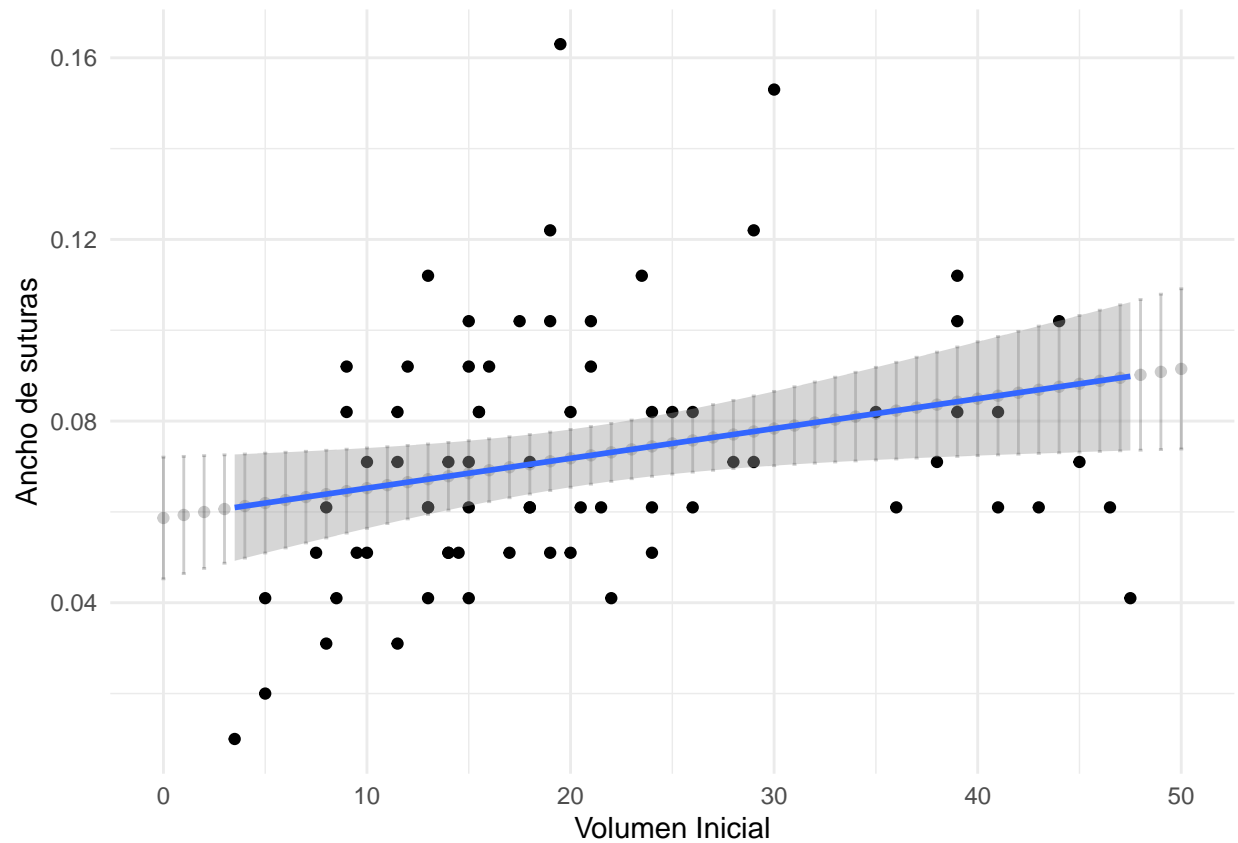
La recta (estimación puntual y por intervalos) la podemos ver en la siguiente gráfica:

```
# 3 - graficar predicciones
new_points <- expand.grid(initial_volume = 0:50)
mean_pred <- predict(lm_fit, new_data = new_points)
conf_int_pred <-
  predict(
    object = lm_fit,
    new_data = new_points,
    type = "conf_int",
    level = .95
  )

plot_data <-
  new_points %>%
  bind_cols(mean_pred) %>%
  bind_cols(conf_int_pred)

# 4 - graficar datos
ggplot(plot_data, aes(x = initial_volume)) +
  geom_point(data = urchins, aes(x = initial_volume, y = width)) +
  geom_point(aes(y = .pred), alpha = 0.2) +
  geom_errorbar(aes(ymin = .pred_lower,
                    ymax = .pred_upper),
                width = .2, alpha = 0.2) +
  geom_smooth(data = urchins, aes(x = initial_volume, y = width), method = "lm") +
```

```
labs(y = "Ancho de suturas", x = "Volumen Inicial") +
theme_minimal()
```



### 1.3 Coeficientes de $\beta_0$ y $\beta_1$

Podemos calcular los coeficientes de  $\beta_0$  y  $\beta_1$  con las fórmulas:

```
# 5 - ¿cómo se ven los coeficientes con las fórmulas?
b1 <- cov(urchins$width,urchins$initial_volume)/var(urchins$initial_volume)
b0 <- mean(urchins$width)-b1*mean(urchins$initial_volume)
b1
```

```
## [1] 0.0006565835
```

```
b0
```

```
## [1] 0.05866426
```

Ahora para poder calcular la  $R^2$  necesitamos calcular:

### 1.4 TSS, ESS y RSS

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
yhat <- predict(lm_fit, new_data = urchins %>% select(-width))
```

```
# TSS (Total Sum of Squares)
tss <- sum((urchins$width-mean(urchins$width))**2)
# RSS (Residual Sum of Squares)
rss <- sum((urchins$width-yhat)**2)
# ESS: (Explained Sum of Squares)
ess <- sum((urchins$width-yhat)**2)
# sigma_hat
rse <- round(sqrt(ess/(72-2)),4)
r2 <- (tss-rss)/tss
```

```
tss
```

```
## [1] 0.05492088
```

```
rss
```

```
## [1] 0.05100981
```

```
ess
```

```
## [1] 0.05100981
```

```
rse
```

```
## [1] 0.027
```

```
r2
```

```
## [1] 0.07121268
```