

REGRESIÓN LOGÍSTICA EN EL YACIMIENTO DE “LAS VEGAS” EN ECUADOR



Juan Antonio Fernández Morales

Arqueología

Aplicaciones Informáticas en Arqueología y Estadística

Curso 2022/23

1. Introducción

Los comienzos de la arqueología se dan durante el siglo XIX, estos inicios están vinculados al coleccionismo, durante este momento se guardan los elementos de la cultura material del pasado que tienen un valor o por su importancia histórica o por su valor material o por ambas cosas. Como disciplina no nace hasta el siglo XX donde poco a poco se le va dando más importancia al contexto y se comienza a hacer historia con estos elementos de la cultura material. Desde que apareció la disciplina siempre se ha ido dando más importancia al trabajo de campo que al posterior trabajo de análisis en laboratorio, ejemplo claro de esto son las excavaciones de urgencia urbanas, en el que en la mayoría de los casos no existe un estudio posterior de estos elementos extraídos en estas excavaciones y simplemente se almacenan. Estos objetos se van transformando en datos y datos que se van acumulando, llegando hasta un punto en la actualidad en la que existen una cantidad de datos demasiado cuantiosa y que se debe de analizar para seguir realizando historia. Aquí es donde entra una de las herramientas más poderosa, el análisis de datos mediante computación. En el presente trabajo se realizará un análisis de datos computacionales de los datos taxonómicos de los especímenes de aves del yacimiento arqueológico OGSE-80.

2. Contexto

La cultura de Las Vegas se localizaba en la costa de la actual Ecuador, más en concreto sobre la península de Santa Elena, cronológicamente se sitúa entre el 10.000 a.C y el 4600 a.C. Esta cultura es precerámica y destaca por ser uno de los primeros sitios en Sudamérica donde se desarrollaron las primeras técnicas de domesticación de vegetales por parte de los seres humanos, estos vegetales eran principalmente la calabaza y el maíz. También destaca por ser un sitio donde ha sido encontrado uno de los entierros más grandes de toda América, los arqueólogos encontraron este enterramiento en el que consta un número mínimo de individuos humanos de 192. Han sido hallados un total de 32 yacimientos pertenecientes a esta cultura, estos yacimientos ocupan un área de 300 km² y uno de los yacimientos mejor estudiados es el protagonista de este presente trabajo, este yacimiento es el OGSE-80. Este yacimiento es donde encontraron el gran enterramiento que se acaba de mencionar anteriormente, el cual se hizo especialmente conocido tras el descubrimiento de los conocidos como “Amantes de Sumpa”, fue llamada así debido a que se encontraron dos individuos que reposaban uno al lado del otro en una posición que

parecía amorosa. Este descubrimiento tuvo una gran repercusión en prensa y ayudó a que el público general se interesara por este sitio y por la cultura de Las Vegas en general.

3. Metodología y datos

Los datos han sido extraídos de la base de datos de arqueología comparativa de la Universidad de Pittsburgh. Estos datos tratan sobre los especímenes de aves que han sido hallados en el yacimiento OGSE-80 y se encuentra en formato Excel. Este Excel representa todos los especímenes de aves identificados en el yacimiento, hay 279 líneas, es decir, 279 de especímenes identificados, cada una con 14 variables separadas por comas. Las variables se enumeran en el siguiente orden: Orden, Familia, cf, Género, Especie, Elemento óseo, Parte del hueso, Lado, Zanja, Sección, Profundidad, Número de muestra, Fecha de recogida y Notas adicionales.

De los datos disponibles en este trabajo nos centraremos en dos variables, por una parte, el tipo de hueso que se ha identificados y por otra parte la parte del hueso. La parte del hueso es una variable binaria, es decir, solo tiene dos opciones, que en este caso es distal o no distal. Con estas dos variables se realizará una regresión logística. La regresión logística es un método estadístico utilizado para analizar la relación entre una variable de respuesta categórica binaria y una o más variables predictoras, que pueden ser categóricas o numéricas. El objetivo de la regresión logística es modelar la probabilidad de que la variable de respuesta pertenezca a una de las categorías en función de los valores de las variables predictoras.

4. Exposición de datos

Comenzaremos analizando estas variables mediante datos básicos respecto a porcentajes, en la figura1 podemos observar una gráfica tipo tarta en el que se representa el porcentaje de tipo de huesos identificados. En la figura2 se representa una gráfica del mismo tipo que la anterior, pero con la variable binaria de la parte del hueso.

Para elaborarlo se ha escrito el siguiente código r:

```
Fhueso<- table(LAS_VEGAS$Hueso)

porc_hueso <- prop.table(Fhueso)*100

Fpartehueso<- table(LAS_VEGAS$Parte_del_hueso)

porc_partehueso <- prop.table(Fpartehueso)*100
```

```

porc_partehueso

round(porc_partehueso)

porcentajehueso<- round(porc_hueso)

porc_hueso

percentpartehueso <- round(porc_partehueso, 2)

cols=c("purple", "yellow")

pie(percentpartehueso, labels=percentpartehueso, col=cols, radius=1)

legend("topright", c("Distal", "No distal"), title = "Parte del hueso",

      fill=c("purple", "yellow"), cex=1)

percenthueso <- round(porc_hueso,1)

cols=c("purple", "yellow","green","blue","red","grey", "white","orange","gold",
"brown", "darkgreen")

pie(percenthueso, labels=percenthueso, col=cols, radius=1)

legend("topright", c("Carpometacarpo", "Coracoide", "Escapula",
"Esternon","Falange","Femur","Húmero","Os premaxillary","Radio","Tarsometatarso",
"Tibiotarso"), title = "Parte del hueso",

      fill=c("purple", "yellow","green","blue","red","grey", "white","orange","gold",
"brown", "darkgreen"), cex=0.9)

```

Hemos observado como la proporción de no distal (54,84%) es mínimamente superior a la de distal (45,16%). Mientras que en los tipos de hueso, encontramos una variedad más grande, destacando el húmero que es el hueso mayoritario (36,2%), junto con el coracoide (30,1%), luego obtenemos un porcentaje similar en escapula, esternón, falange, fémur, carpometacarpo, tibiotarso y tarsometatarso. Finalmente obtenemos porcentaje muy pequeños en os premaxillary y radio.

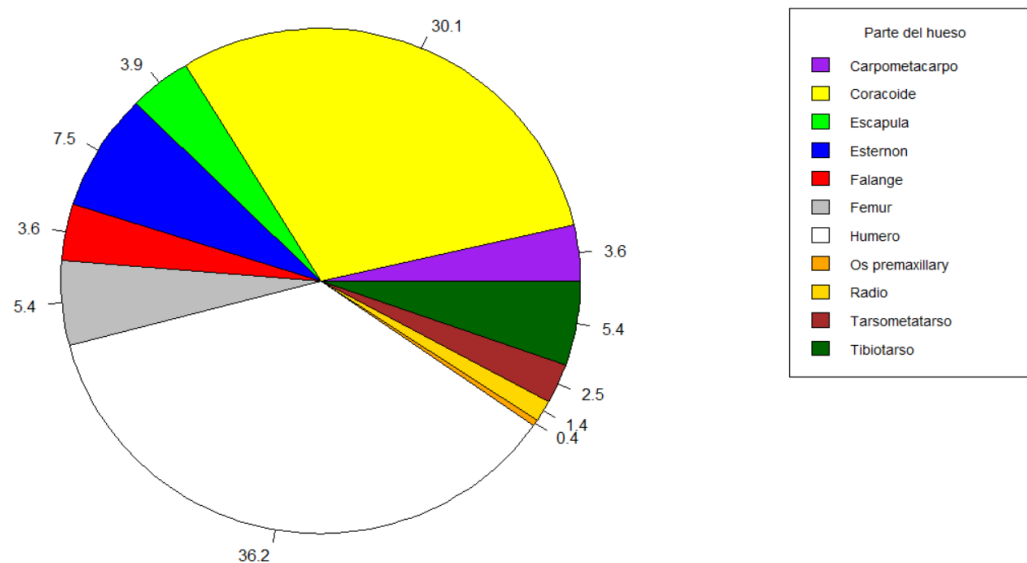


Figura 1. Grafica estilo “tarta” de los porcentajes de los tipos de hueso que encontramos en el yacimiento. Elaboración propia a partir del programa Rstudio.

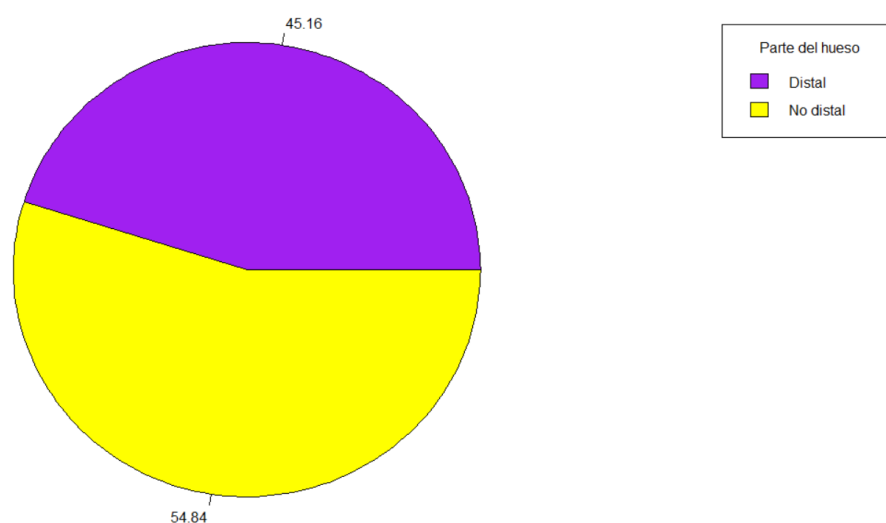


Figura 2. Grafica estilo “tarta” de los porcentajes de las partes de hueso que encontramos en el yacimiento. Elaboración propia a partir del programa Rstudio.

5. Análisis de los datos

Se debe cambiar las variables categóricas binarias por los números 1 y 0, para ello trabajamos en Rstudios utilizando una serie de códigos que hagan sustituir los nombres de las variables binarias por estos dos números:

```
LAS_VEGAS$Parte_del_hueso <- ordered(LAS_VEGAS$Parte_del_hueso, labels =  
c("Distal", "No distal"), levels = c(1,0) )
```

Tras esto hacemos lo mismo con la variable de tipos de hueso, esta variable al tener 11 opciones nos conviene transformar cada una de estas 11 opciones en una variable distinta, para ello trabajamos en Rstudio aplicando el siguiente código:

```
LAS_VEGAS$HuesoCoracoide <- ifelse(LAS_VEGAS$Hueso== "Coracoide", 1,0)  
LAS_VEGAS$HuesoCoracoide  
  
LAS_VEGAS$HuesoTibiotarso <- ifelse(LAS_VEGAS$Hueso== "Tibiotarso", 1,0)  
LAS_VEGAS$HuesoTibiotarso  
  
LAS_VEGAS$HuesoHumero <- ifelse(LAS_VEGAS$Hueso== "Humero", 1,0)  
LAS_VEGAS$HuesoHumero  
  
LAS_VEGAS$HuesoEsternon <- ifelse(LAS_VEGAS$Hueso== "Esternon", 1,0)  
LAS_VEGAS$HuesoEsternon  
  
LAS_VEGAS$HuesoFalange <- ifelse(LAS_VEGAS$Hueso== "Falange", 1,0)  
LAS_VEGAS$HuesoFalange  
  
LAS_VEGAS$HuesoTarsometatarso<-ifelse(LAS_VEGAS$Hueso== "  
Tarsometatarso", 1,0)  
LAS_VEGAS$HuesoTarsometatarso  
  
LAS_VEGAS$HuesoCarpometacarpo <- ifelse(LAS_VEGAS$Hueso==  
"Carpometacarpo", 1,0)  
LAS_VEGAS$HuesoCarpometacarpo  
  
LAS_VEGAS$HuesoFemur <- ifelse(LAS_VEGAS$Hueso== "Femur", 1,0)  
LAS_VEGAS$HuesoFemur  
  
LAS_VEGAS$HuesoEscapula <- ifelse(LAS_VEGAS$Hueso== "Escapula", 1,0)  
LAS_VEGAS$HuesoEscapula  
  
LAS_VEGAS$HuesoRadio <- ifelse(LAS_VEGAS$Hueso== "Radio", 1,0)
```

```
LAS_VEGAS$HuesoRadio
```

```
LAS_VEGAS$HuesoOspremaxillary<-ifelse(LAS_VEGAS$Hueso=="Os  
premaxillary", 1,0)
```

```
LAS_VEGAS$HuesoOspremaxillary
```

Tras esto realizamos una regresión logística para modelar la relación entre una variable de respuesta y una variable predictora. El objetivo es encontrar una función lineal que mejor se ajuste a los datos observados.

En RStudio, puedes ajustar un modelo de regresión lineal utilizando la función “*glm*”. Una vez que hemos preparado los datos para el modelo de regresión lineal, se utilizar la función “*summary*” para obtener los resultados del modelo. Esto incluye los coeficientes estimados, los valores p, las estadísticas de ajuste y otras métricas relevantes.

Es importante interpretar los resultados del modelo, prestando atención a los coeficientes estimados y sus valores p. Los coeficientes estimados indican la magnitud y la dirección de la relación entre las variables, mientras que los valores p indican la significancia estadística de los coeficientes. Un valor p menor a 0.05 generalmente se considera como evidencia de una relación significativa.

Se realiza un ajuste del modelo de regresión lineal de cada una de las variables creadas binominalmente con las variable distal o no distal que es la que nos interesa comprobar. En Rstudio representamos esto con el siguiente código:

```
modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoTibiotarso, data  
= LAS_VEGAS, family = binomial)
```

```
summary(modelo)
```

```
modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoHumero, data =  
LAS_VEGAS, family = binomial)
```

```
summary(modelo)
```

```
modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoCoracoide, data  
= LAS_VEGAS, family = binomial)
```

```
summary(modelo)
```

```

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoEsternon, data =
LAS_VEGAS, family = binomial)

summary(modelo)

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoFalange, data =
LAS_VEGAS, family = binomial)

summary(modelo)

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoTarsometatarso,
data = LAS_VEGAS, family = binomial)

summary(modelo)

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoCarpometacarpo, data = LAS_VEGAS, family = binomial)

summary(modelo)

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoFemur, data =
LAS_VEGAS, family = binomial)

summary(modelo)

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoEscapula, data
= LAS_VEGAS, family = binomial)

summary(modelo)

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoRadio, data =
LAS_VEGAS, family = binomial)

summary(modelo)

modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoOspremaxillary,
data = LAS_VEGAS, family = binomial)

summary(modelo)

```

Tras esto comprobamos que las únicas dos opciones que contienen un pvalue que indique una relación significativa entre la variable predictora y la variable de respuesta, es el

hueso coracoide y el tibiotarso. Tras esto nos centramos en los huesos coracoide y tibiotarso, la información que extraemos del ajuste del modelo es la que se representa en la figura 3 y la figura 4.

```
> modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoCoracoide, data = LAS_VEGAS, family = binomial)
> summary(modelo)

Call:
glm(formula = LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoCoracoide,
    family = binomial, data = LAS_VEGAS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3921 -1.3921  0.9769  0.9769  1.3893

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.4917    0.1476   3.332 0.000862 ***
LAS_VEGAS$HuesoCoracoide -0.9772    0.2688  -3.635 0.000278 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 384.16  on 278  degrees of freedom
Residual deviance: 370.53  on 277  degrees of freedom
AIC: 374.53

Number of Fisher Scoring iterations: 4
```

Figura 3. Ajuste del modelo del "HuesoCoracoide". Elaboración propia a partir del programa Rstudio.

```
> modelo <- glm(LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoTibiotarso, data = LAS_VEGAS, family = binomial)
> summary(modelo)

Call:
glm(formula = LAS_VEGAS$Parte_del_hueso ~ LAS_VEGAS$HuesoTibiotarso,
    family = binomial, data = LAS_VEGAS)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.289 -1.289  1.070  1.070  1.626

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.2590    0.1241   2.087 0.0369 *
LAS_VEGAS$HuesoTibiotarso -1.2706    0.5969  -2.129 0.0333 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 384.16  on 278  degrees of freedom
Residual deviance: 378.99  on 277  degrees of freedom
AIC: 382.99

Number of Fisher Scoring iterations: 4
```

Figura 4. Ajuste del modelo del "HuesoTibiotarso". Elaboración propia a partir del programa Rstudio.

El coeficiente estimado para "HuesoCoracoide" es -0.9772. Esto indica que, cuando la variable predictora aumenta en una unidad, el logaritmo de las odds de que la parte del hueso sea distal disminuye en 0.9772 unidades. Además, el valor p asociado a este coeficiente es 0.000278, lo que indica que hay evidencia significativa para afirmar que la

presencia de huesos coracoides está asociada a un aumento en la probabilidad de que la parte del hueso sea distal.

Hipótesis: Existe una asociación estadísticamente significativa entre la presencia de huesos coracoides y el aumento en la probabilidad de que la parte del hueso sea distal en la muestra de datos analizada.

En resumen, según el modelo de regresión logística ajustado, se sugiere que la presencia de huesos coracoides está relacionada con una mayor probabilidad de que la parte del hueso sea distal.

Por último, se calcularán las probabilidades de que las opciones de las variables predichas pertenezcan a cada una de las opciones de la variable predictora, es decir calcular las probabilidades de que las opciones de “HuesoCoracoide” y “HuesoTibiotarso” pertenezcan a las opciones de la variable binaria “parte del hueso”. Para ellos se empleará la función “*exp*” tal y como se muestra en la figura 5. Para transformar los valores de odds en porcentajes, puedes utilizar la función “*plogis*” en R.

```
> prediccionModelotibiotarso <- exp(Modelotibiotarso$coefficients[,1])
> prediccionModelotibiotarso
      (Intercept) LAS_VEGAS$HuesoTibiotarso
      1.2956522      0.2806589
> predccionModelocoracoide <- exp(Modelocoracoide$coefficients[,1])
> predccionModelocoracoide
      (Intercept) LAS_VEGAS$HuesoCoracoide
      1.6351351      0.3763509
> oddscoracoide<- 0.3763509
> probabilidadcoracoide <- plogis(oddscoracoide) * 100
> print(probabilidadcoracoide)
[1] 59.29927
> oddstibiotarso<- 0.2806589
> probabilidadtibiotarso <- plogis(oddstibiotarso) * 100
> print(probabilidadtibiotarso)
[1] 56.97078
```

Figura 5. Predicción de los objetos “HuesoTibiotarso” y “HuesoCoracoide”. Elaboración propia a partir del programa Rstudio.

Los resultados plasmados en el código que se representa en la figX nos da el valor que nos indica cómo la odds de la variable de respuesta cambia por cada unidad de cambio en la variable "HuesoCoracoide" y “HuesoTibiotarso”. Para representar esto en porcentaje usamos la función “*plogis*” que nos dio un porcentaje de efecto de la variable predictora en la variable predicha en que pueda ocurrir un evento o categoría específica. En este caso obtenemos un porcentaje de 59,3% para "HuesoCoracoide" y 56,9% para “HuesoTibiotarso”, lo que quiere decir que la variable de respuesta cambiará en estos

porcentajes por cada unidad de cambio en la variable "HuesoCoracoide" y "HuesoTibiotarso".

6. Conclusiones.

La comparación de las variables hueso y parte del hueso de la base de datos del yacimiento de Las Vegas nos ha permitido llegar a una serie de conclusiones. En la mayoría de los huesos encontrados en el yacimiento no existe una relación clara entre si son o no distales en relación con el tipo de hueso. En concreto solo 2 de las 11 opciones existe una relación clara entre ambas variables. La existencia en general de una relación entre los tipos de huesos y la parte del hueso debería ser estudiada con un conjunto de datos amplios de yacimientos en los que existan amplios datos arqueozoológicos. La realización de este trabajo plasma a la perfección la importancia de los análisis de datos y el conocer como hacerlo para ciencias como la arqueología.

7. Bibliografía:

Kanomata, Y., Tabarev, A. V., Tabarev, J. V., & Stothert, K. E. (2014). Un nuevo acercamiento al estudio de herramientas líticas de la cultura milenaria “Las Vegas” Santa Elena, Ecuador.

Stothert, K. E. (1985). The preceramic Las Vegas culture of coastal Ecuador. *American Antiquity*, 50(3), 613-637.

Stothert, K. E., & Mosquera, A. S. (2011). Culturas del Pleistoceno final y el Holoceno temprano en el Ecuador. *Boletín de Arqueología PUCP*, (15), 81-119.