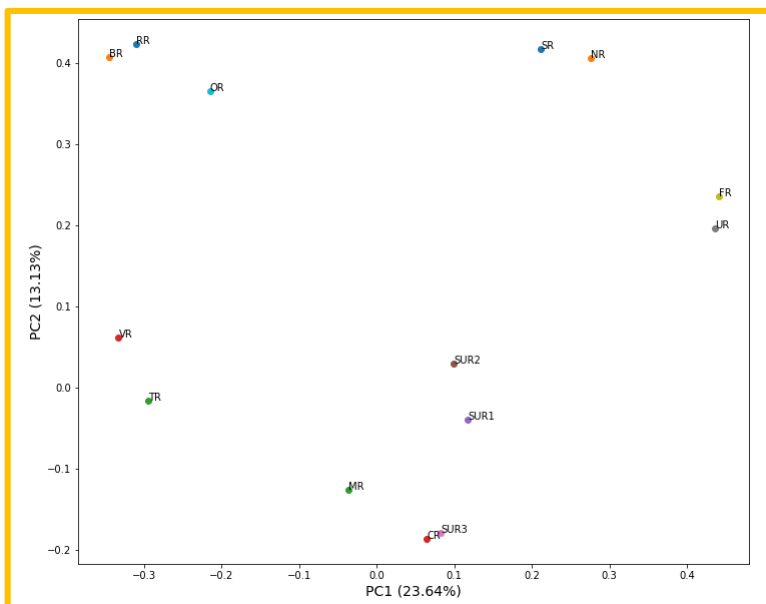
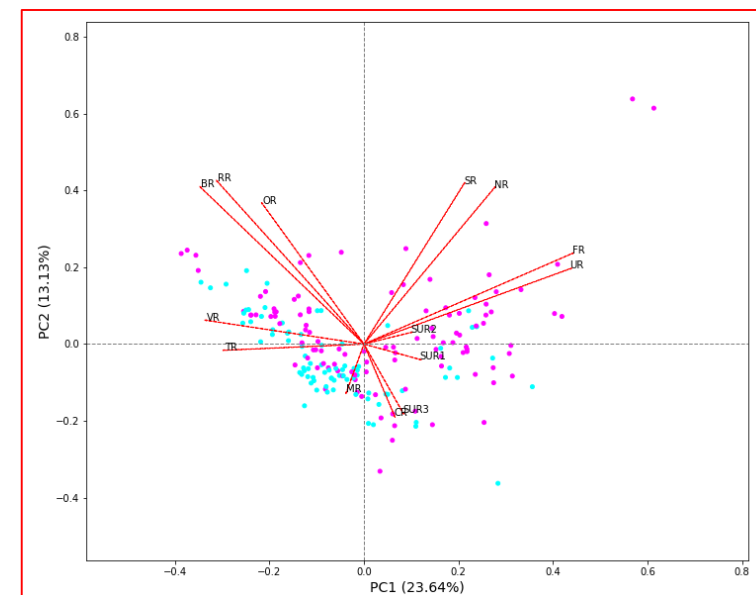


Grafico de carga PCA 2D



A través del grafico bivariado en 2D de la primera componente versus segunda componente principal, se evidencia de manera visual que las variables con más cargas dentro de estas componentes son las relacionadas con porcentaje de acceso desde los bordes del embalse a áreas no desarrolladas (OR), Distancia mínima del depósito de agua a las carreteras (RR), Urbanización - Distancia mínima a los edificios (BR), los valores más atípicos son Tipo de orilla (CR) y Entorno 3: el tercer tipo más dominante de cobertura terrestre que rodea el depósito de agua (SUR3).

Grafico BIPLLOT 2D



Se evidencia que la urbanización, distancia mínima del deposito de agua y los bordes del embalse a áreas no desarrolladas presentes en el primer componente principal están estrechamente relacionado o mide de manera negativa en su mayoría las características del ambiente, y el componente dos mide tales variables pero de manera positiva. Asimismo, se observa que el uso de depósitos de agua, la presencia de la pesca, superficie del deposito de agua numérico y el segundo tipo mas dominante de cobertura terrestre que rodea el depósito de agua, son variables que influyen de manera mayoritaria en la presencia de ranas verdes.

Modelo de regresión

NOTA: La data en estudio era de anfibios contaba con 184 instancias y 23 atributos. Tomada de: <https://archive.ics.uci.edu/ml/datasets/Amphibians>

OLS Regression Results

Dep. Variable:	Green frogs	R-squared:	0.305
Model:	OLS	Adj. R-squared:	0.249
Method:	Least Squares	F-statistic:	5.445
Date:	Sat, 04 Jun 2022	Prob (F-statistic):	1.34e-08
Time:	20:27:17	Log-Likelihood:	-100.89
No. Observations:	189	AIC:	231.8
Df Residuals:	174	BIC:	280.4
Df Model:	14		

Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
const	0.8173	0.379	2.154	0.033	0.069	1.566
SR	-1.823e-07	9.18e-07	-0.199	0.843	-1.99e-06	1.63e-06
NR	0.0380	0.028	1.340	0.182	-0.018	0.094
TR	-0.0316	0.006	-4.878	0.000	-0.044	-0.019
VR	-0.0070	0.031	-0.223	0.824	-0.069	0.055
SUR1	0.0083	0.010	0.842	0.401	-0.011	0.028
SUR2	0.0059	0.009	0.626	0.532	-0.013	0.025
SUR3	-0.0294	0.010	-2.954	0.004	-0.049	-0.010
UR	0.0759	0.047	1.623	0.106	-0.016	0.168
FR	-0.0017	0.047	-0.036	0.971	-0.095	0.091
OR	0.0006	0.002	0.342	0.732	-0.003	0.004
RR	0.0354	0.021	1.674	0.096	-0.006	0.077
BR	0.0013	0.020	0.064	0.949	-0.039	0.042
MR	0.1916	0.109	1.762	0.080	-0.023	0.406
CR	-0.2386	0.264	-0.904	0.367	-0.759	0.282
Omnibus:	20.288	Durbin-Watson:	1.483			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7.907			
Skew:	-0.241	Prob(JB):	0.0192			
Kurtosis:	2.121	Cond. No.	6.68e+05			

El R-squared o R cuadrado no es el apropiado o no es el más factible estadísticamente para mis datos, es decir que el modelo no predice bien la presencia o ausencia de ranas verdes en los lugares en estudio.

```
[162] confusion_matrix(Y,pred)/189*100  
  
array([[29.1005291 , 13.75661376],  
       [10.58201058, 46.56084656]])
```

Teniendo en cuenta la matriz de confusión se logra evidenciar que el modelo tendría un 75,66% de factibilidad para los datos, es decir que la mayor parte de veces se predice de manera correcta la ausencia o presencia de ranas verdes.

Andrea Menco Tovar
ACP- Análisis de Componentes Principales
04/06/2022