



## Taller 6: PCA

Caso: implementación de PCA

Análisis de datos – Prof. David Porta

### **Maestría en Estadística Aplicada**

Presentado por:

**DIEGO HERRERA MALAMBO** - [malambod@utb.edu.co](mailto:malambod@utb.edu.co)



**TRANSFORMAMOS VIDAS  
CON EDUCACIÓN DE EXCELENCIA**

# Implementación de PCA

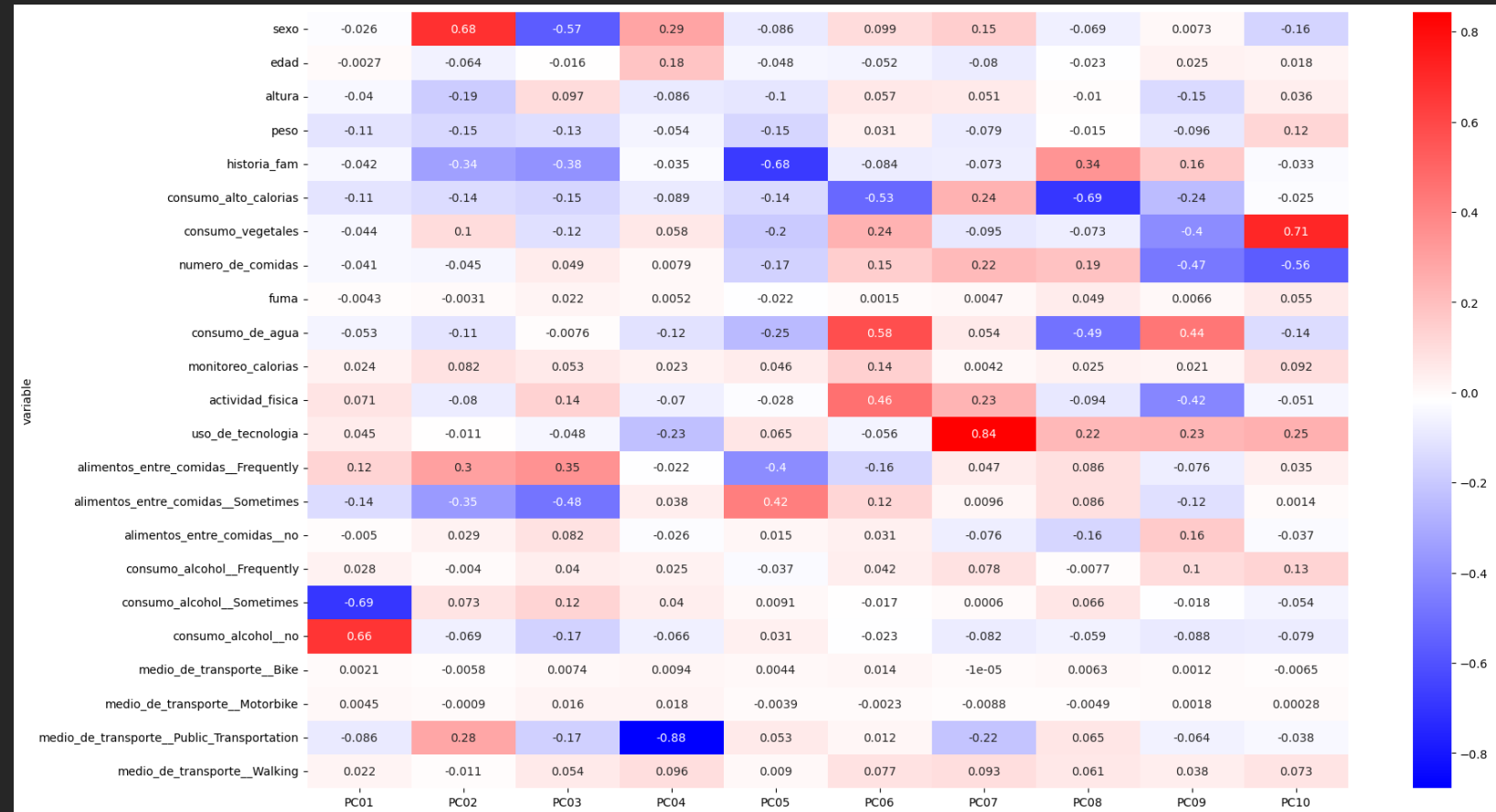
El dataset usado para el ejercicio de PCA corresponde a datos antropometricos y de habitos de personas orientados a la clasificacion de sobrepeso.

El dataset consta de 15 variables independientes y 1 dependiente.

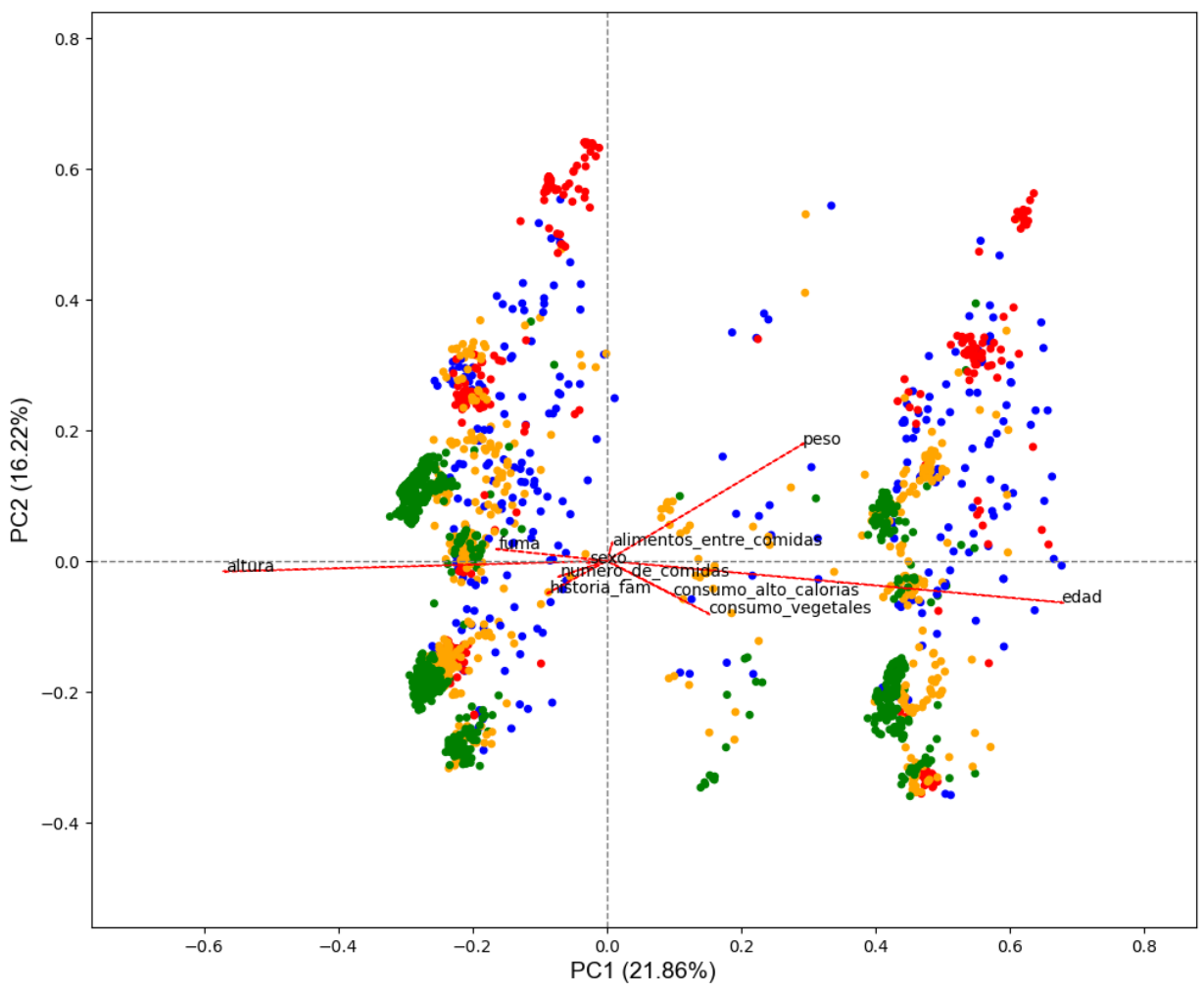
Realice el proceso de normalizacion y de separacion de variables categoricas en dicotomicas.

Al implementar PCA se pudo encontrar que en 5 variables se podia concentrar el 65% del dataset, al usar 10 componentes llegamos al 86.5%

Relación de variables y el impacto sobre cada componente generado.



# Implementación de PCA



Estas son las variables que dan mejores resultados al implementar regression OLS y validarlo con PCA

	coef	std err	t	P> t	[0.025	0.975]
sexo	-0.1886	0.045	-4.194	0.000	-0.277	-0.100
edad	1.0460	0.127	8.251	0.000	0.797	1.295
altura	-1.0758	0.141	-7.622	0.000	-1.353	-0.799
peso	1.7834	0.122	14.607	0.000	1.544	2.023
historia_fam	0.4532	0.051	8.929	0.000	0.354	0.553
consumo_alto_calorias	-0.1794	0.054	-3.339	0.001	-0.285	-0.074
consumo_vegetales	-0.5172	0.067	-7.749	0.000	-0.648	-0.386
numero_de_comidas	-0.4865	0.066	-7.404	0.000	-0.615	-0.358
monitoreo_calorias	0.3390	0.081	4.166	0.000	0.179	0.499
alimentos_entre_comidas_Frequently	-0.3258	0.112	-2.903	0.004	-0.546	-0.106
alimentos_entre_comidas_Sometimes	0.3477	0.105	3.323	0.001	0.142	0.553
alimentos_entre_comidas_no	0.8713	0.148	5.879	0.000	0.581	1.162
consumo_alcohol_Frequently	2.1920	0.163	13.422	0.000	1.872	2.512
consumo_alcohol_Sometimes	1.7643	0.142	12.408	0.000	1.485	2.043
consumo_alcohol_no	1.7066	0.140	12.195	0.000	1.432	1.981