

Estadística Aplicada 3 - Examen 2

Marcelino 191654

2/12/23

Ejercicio 1

En este análisis se trabajó con la base de datos de percepción humana de colores, que según el libro de *Modern Multidimensional Scaling* (Borg, I., & Groenen, P. J. F. (2005)) se le pidió a 14 personas que calificaran la similitud de 14 colores (que solo diferían en sus longitudes de onda pero no en brillo o saturación). Los promedios de estas calificaciones revelaron un patrón de interrelación, un gradiente peculiar de similitudes, en lugar de categorías de color discretas. Esto sugiere que nuestra percepción del color es más un reflejo de un gradiente de similitudes que de categorías de color distintas y separadas. Además, según lo que el autor del libro refiere no es sencillo analizar cómo es que los humanos clasificamos los colores pero que de alguna forma siempre terminamos con cierto orden de similitud entre los colores que refleja una percepción implícita de escala de colores gradual en forma de herradura o círculo, por lo que en este análisis se intenta dar una explicación a este fenómeno.

Para explicarla, debemos poder entender cuál es nuestra percepción de orden o estructura de los colores solo conociendo lo que percibimos como similares. Para esto se utilizará una técnica estadística llamada **MDS** (Multidimensional Scaling) que permitirá representar en un espacio de n dimensiones los datos de similitud de los colores (Hay muchos métodos para poder realizarlo pero como una aproximación simple usaremos el clásico). Por lo general, se utilizan dimensiones muy bajas como 2 o 3 para poder visualizar e interpretar la estructura de los datos, por medio de distancias euclidianas (que representan la similitud) entre los datos analizados. De esta forma aunque la estructura sea muy compleja podrían verse ciertos patrones geométricos que nos darían una idea de cómo percibimos los colores.

Con lo cual realizamos el **MDS** en dos dimensiones y obtuvimos la Figura 1 que se encuentra en los anexos de este reporte. Notamos que los colores se encuentran en un orden de similitud que va de los azules (casi morados o violetas) a los rojos, pasando por los verdes, amarillos y naranja, en forma de herradura o círculo. Lo cual nos lleva a pensar que implícitamente percibimos los colores de esta forma, como una escala de colores, y esta escala no es solamente lineal sino que es circular, porque en cierta forma colores como el violeta podría ser similar al rojo o azul dependiendo de cómo lo veas. Además, observamos algo que hay cierta agrupación de colores que la gente los percibe muy similares, lo cual tiene mucho sentido, dado que luego existen discusiones de si el color aguamarina es verde o azul, etc. Por último es interesante encontrar que la forma de círculo también nos permite ver que hay colores diametralmente opuestos, en el sentido de que una persona siempre los identificaría distintos, como el azul oscuro y el amarillo, o el rojo y verde.

En conclusión, el análisis de escalamiento dimensional nos permite ver que la percepción de los colores es más un reflejo de un gradiente de similitudes que de categorías de color distintas y separadas, y que esta percepción de similitud de colores es en forma de herradura o círculo, además cabe señalar que si realizamos este ejercicio en 3 dimensiones, casi no habrían diferencias significativas en la interpretación de los resultados, por que la disposición de los datos se asemejan a una dona.

Ejercicio 2

En este reporte analizamos la base de datos tortugas hembras, que contiene las variables `length`, `width`, `height`. En este caso realizamos un análisis de componentes principales para poder encontrar índices que nos permitan explicar la variabilidad de los datos de forma resumida.

Primero checamos los de covarianzas y eigenvalores de la matriz de covarianzas de las variables. Encontramos que hay correlación positiva entre las variables, lo que nos lleva a pensar que hay cierta dependencia entre ellas la cual podría ser provechosa para explicar la variabilidad de las mismas a través de un solo índice o variable como se pretende en componentes principales. Además, en los eigenvalores, notamos que hay una dirección de mayor variabilidad en los datos que domina sobre los demás, lo cual confirma lo que decíamos sobre la dependencia entre las variables.

Ahora, cuando realizamos las componentes principales, encontramos que la primera componente principal explica casi toda la variabilidad de los datos (según la Figura 3), lo cual es muy bueno, porque nos permite explicar casi toda la variabilidad de los datos con un solo índice. Cabe resaltar que escalamos los datos para que la varianza de las variables no afectara el análisis, y además antes de escalar transformamos las variables a logaritmo porque encontramos una relación interesante con respecto al volumen. Y es que para calcular el volumen de una tortuga uno pensaría que es multiplicar esas tres variables y multiplicarla por una constante; de esa forma se podría suponer que si alguna componente principal es una combinación lineal de las $\log(\text{variables})$ entonces en cierta forma podríamos encontrar una relación lineal entre la componente principal de las $\log(\text{variables})$ y el $\log(\text{volumen})$. Y en efecto, encontramos que la primera componente principal es una combinación lineal de las $\log(\text{variables})$ y que además tiene una relación lineal con el $\log(\text{volumen})$. De hecho, la primera componente principal, según el biplot de la Figura 2, es una combinación lineal (casi con la misma proporción) de las tres variables. Y por último, al checar la relación entre la primera componente principal y el $\log(\text{volumen})$ encontramos en la Figura 4 que hay una relación lineal positiva entre ambas variables, lo cual nos permite explicar el volumen de las tortugas con solo un índice.

Los datos de esta relación lineal vienen dadas por la siguiente ecuación: (donde el volumen fue calculado como $V = e^{\text{length} + \text{width} + \text{height}}$ tomando en cuenta las $\log(\text{variables})$)

$$\log(\text{volumen}) = 1.732\text{PCA1}$$

Y estos fueron calculados con una regresión lineal simple que dio coeficientes súper significativos y un R^2 ajustada de 1. Lo cual termina de confirmar la relación lineal entre ambas variables. (véase la Tabla 2)

Por lo tanto, el análisis revela que mediante el uso de una única componente principal, es posible explicar con precisión el volumen de las tortugas hembras, simplificando significativamente el

análisis de sus dimensiones físicas y ofreciendo valiosas aplicaciones en estudios biológicos más profundos con esta relación encontrada.

Tabla 1: Resumen de resultados

(a) Medias			(b) Covarianzas			
mediaLength	mediaWidth	mediaHeight		length	width	height
136	102.5833	51.95833	length	451.3913	271.1739	168.69565
			width	271.1739	171.7319	103.28623
			height	168.6957	103.2862	66.65036

(c) Eigenvalores	
eigenvalues	
680.412306	
6.499624	
2.861621	

Tabla 2: Resumen de resultados regresión

(a) Resumen de la Regresión				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0001943	3.23e-04	-0.6015764	0.5536054
PCA1	1.7319384	4.15e-05	41693.8071127	0.0000000

(b) Adj R ²	
x	
1	

Anexos

Tablas

Figuras

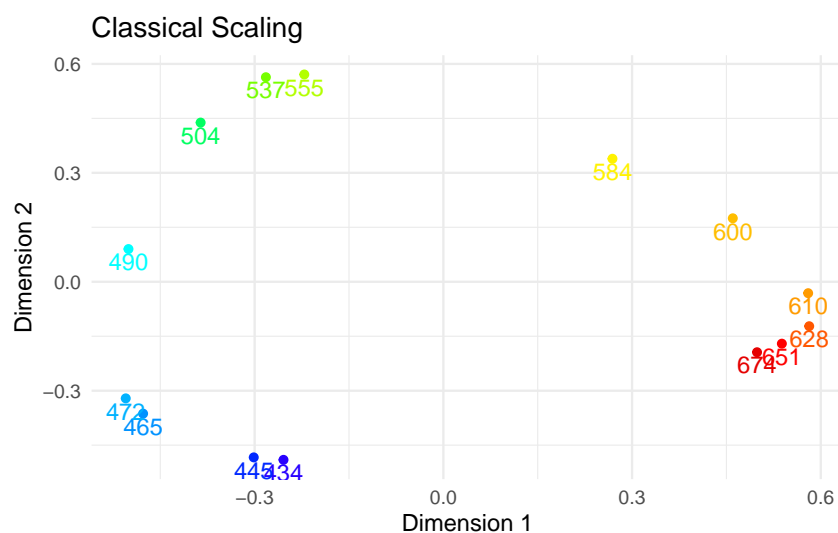


Figura 1: MDS Clásico de la percepción de los colores.

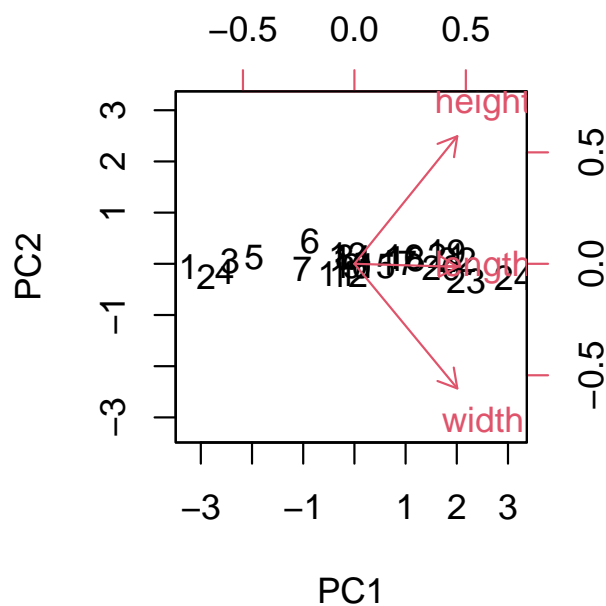


Figura 2: Biplot componentes principales.

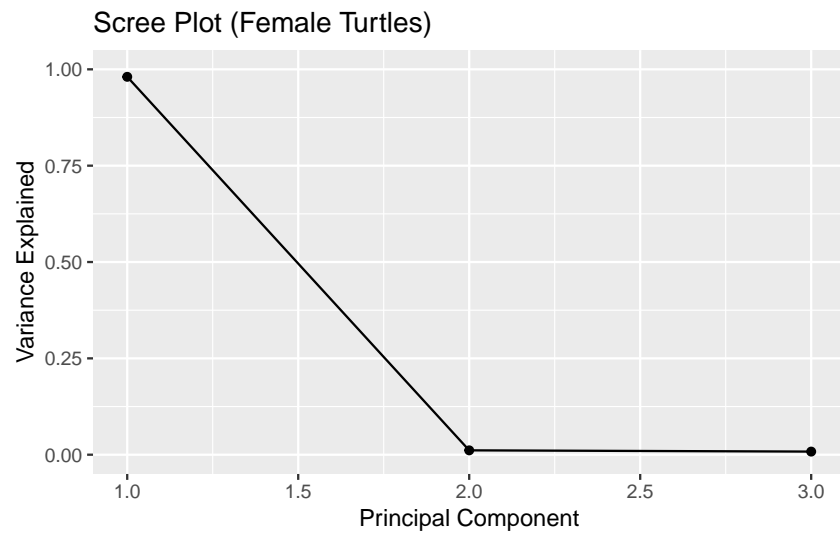


Figura 3: Scree plot.

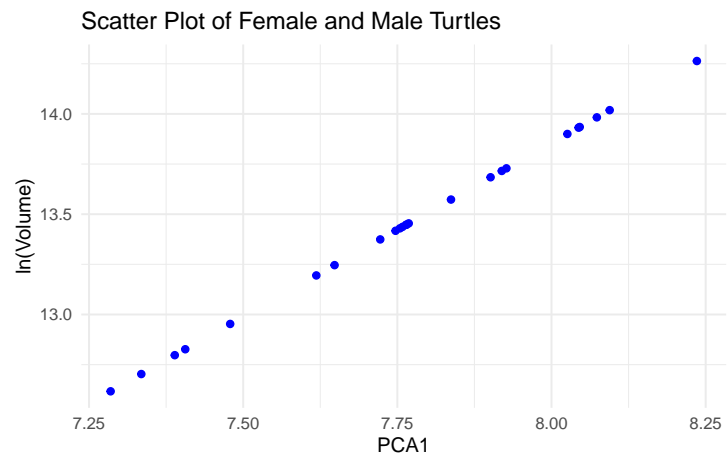


Figura 4: Scree plot.

Referencias

Borg, I., & Groenen, P. J. F. (2005). Modern multidimensional scaling: Theory and applications (2nd ed.). Springer Science + Business Media