

Estadística Aplicada 3 - Examen 2

Marcelino 191654

Invalid Date

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** (CITAI) 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

Haga analisis de componentes principales para el archivo turtles.txt. Utilice solamente los datos correspondientes a las tortugas hembras.

```
# Cargamos base de datos
data2 <- read.xlsx("turtles.xlsx")
turtlesF <- data2 |>
  filter(sex == 'f ')

turtlesF <- turtlesF |>
  mutate(length = log(length),
         width = log(width),
         height = log(height))

# Media de las variables

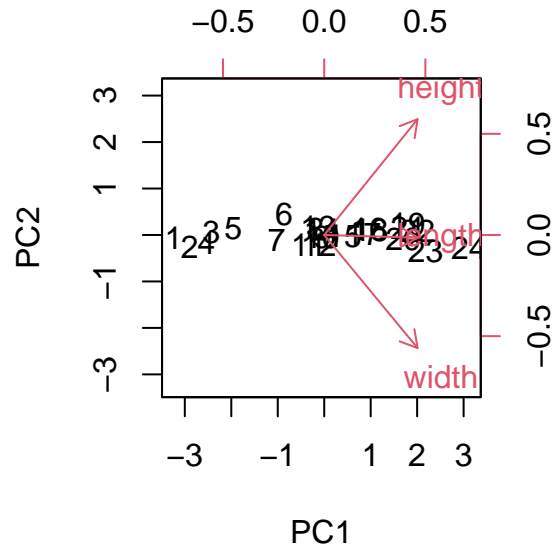
summaryTurtlesF <- turtlesF |>
  summarise(mediaLength = mean(length),
            mediaWidth = mean(width),
            mediaHeight = mean(height))

covariance <- cov(turtlesF[,c('length', 'width', 'height')])

eigenvalues <- covariance |> eigen()

#PCA
#calculate principal components
resultsF <- prcomp(turtlesF[2:4], scale = TRUE)

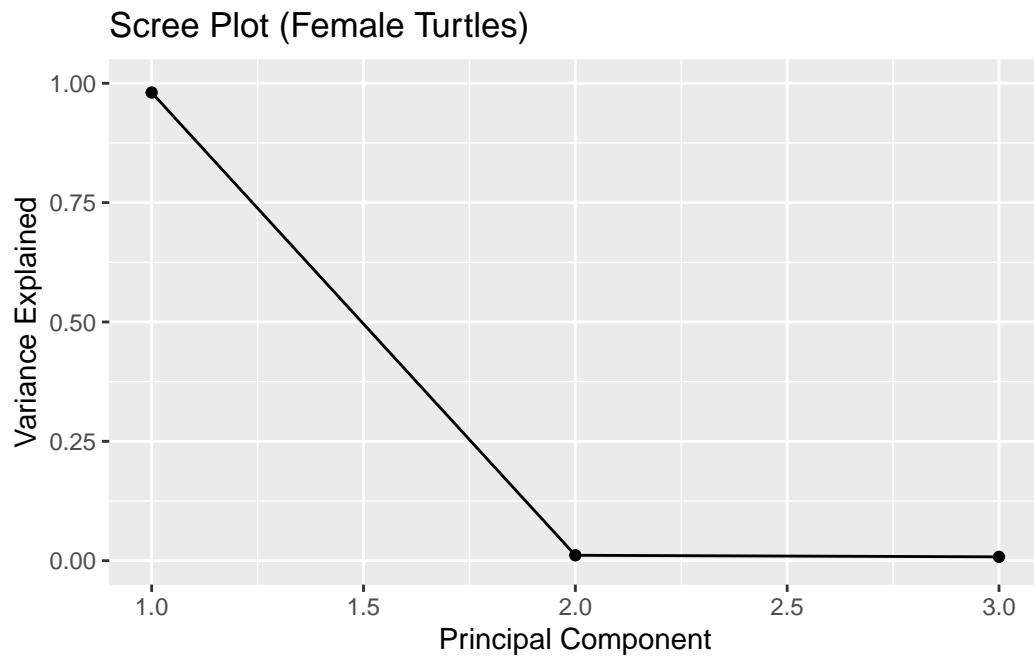
biplot(resultsF, scale = 0)
```



```
var_explainedF = resultsF$sdev^2 / sum(resultsF$sdev^2)

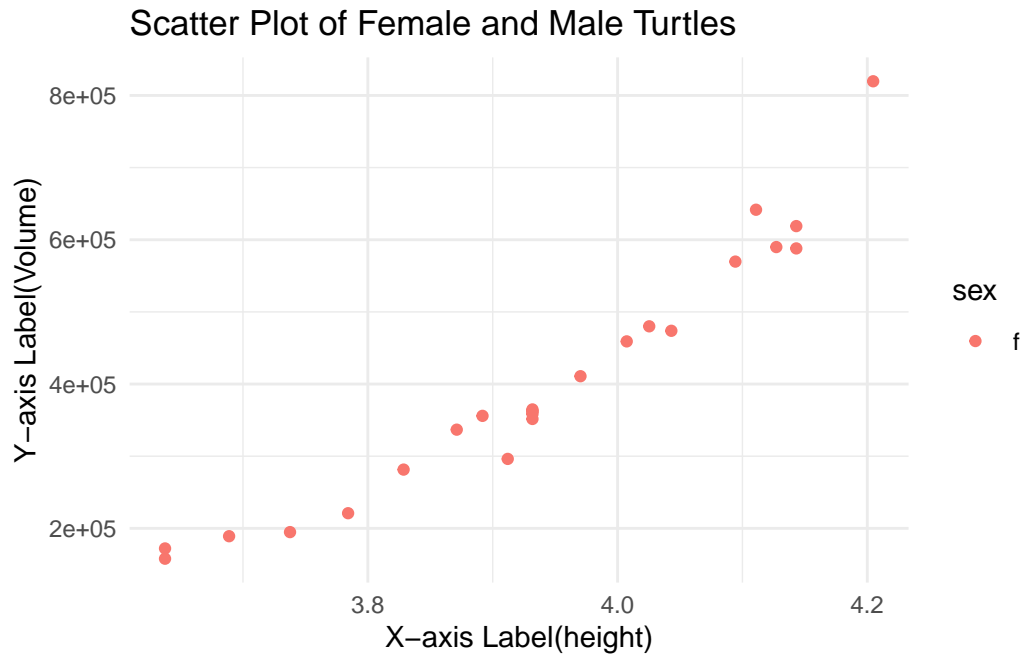
#create scree plot
qplot(c(1:3), var_explainedF) +
  geom_line() +
  xlab("Principal Component") +
  ylab("Variance Explained") +
  ggtitle("Scree Plot (Female Turtles)") +
  ylim(0, 1)
```

Warning: `qplot()` was deprecated in ggplot2 3.4.0.



```
turtlesF$Volume <- pi/6*exp(turtlesF$length+turtlesF$width+turtlesF$height)

ggplot(turtlesF, aes(x = height, y = Volume, color = sex)) +
  geom_point() +
  labs(title = "Scatter Plot of Female and Male Turtles",
       x = "X-axis Label(height)", y = "Y-axis Label(Volume)") +
  theme_minimal()
```



```
# Transformar variables

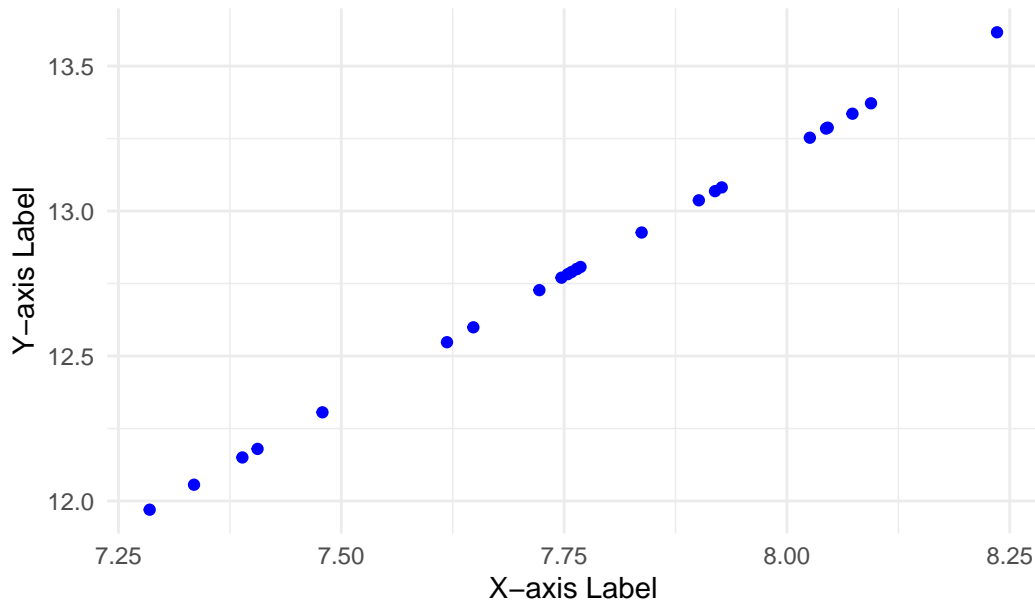
pca_scores <- as.matrix(turtlesF[2:4]) %*% resultsF$rotation

# Extract the first principal component (PCA1)
pca2 <- pca_scores[, 1]
# Create a new data frame with only PCA1
df_pca2 <- data.frame(PCA1 = pca2)

df_pca2$Volume <- turtlesF$Volume

plot <- ggplot()+
  geom_point(data=df_pca2, aes(x = PCA1, y = log(Volume)),color="blue") +
  labs(title = "Scatter Plot of Female and Male Turtles",
       x = "X-axis Label", y = "Y-axis Label") +
  theme_minimal()
print(plot)
```

Scatter Plot of Female and Male Turtles



```
reg <- lm(log(Volume) ~ PCA1, data = df_pca2)
summary(reg)
```

Call:

```
lm(formula = log(Volume) ~ PCA1, data = df_pca2)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|------------|------------|-----------|-----------|-----------|
| -8.324e-05 | -3.254e-05 | 1.164e-05 | 3.147e-05 | 9.202e-05 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|------------|------------|---------|------------|
| (Intercept) | -6.472e-01 | 3.230e-04 | -2004 | <2e-16 *** |
| PCA1 | 1.732e+00 | 4.154e-05 | 41694 | <2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.133e-05 on 22 degrees of freedom

Multiple R-squared: 1, Adjusted R-squared: 1

F-statistic: 1.738e+09 on 1 and 22 DF, p-value: < 2.2e-16

Anexos

Tablas

Figuras

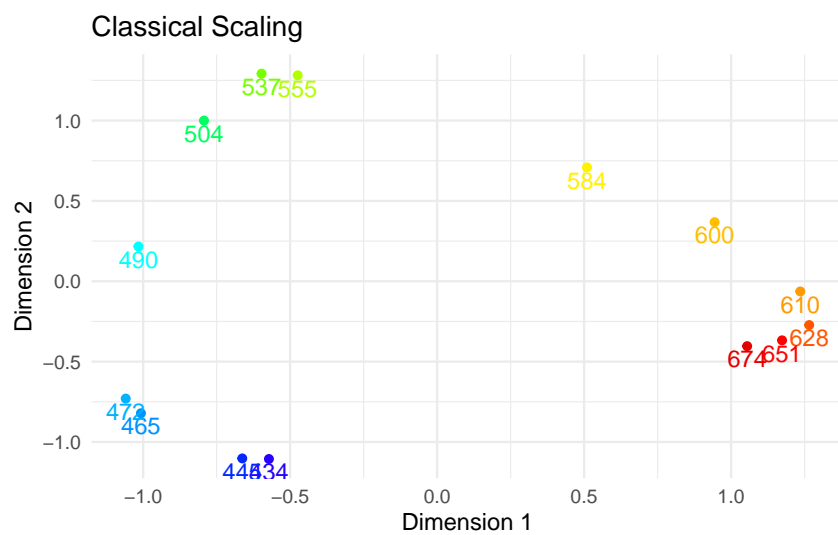


Figura 1: MDS Clásico de la percepción de los colores.