Taller 1

Taller 1

Vectores

1. Considerar los siguientes vectores:

- $\vec{u} = (4, -2)$
- $\vec{v} = (2, 1)$

    a. Calcular la proyección de $\vec{u}$ sobre $\vec{v}$

    b. Calcular la proyección de $\vec{v}$ sobre $\vec{u}$

Matrices

2. Considerar la siguiente matriz:

- $A = \begin{pmatrix} 185 & -80 \\ -80 & 65 \end{pmatrix}$

    a. Calcular los autovalores y autovectores

    b. Expresar A en su forma diagonalizada ($A = V \times D \times V^{-1}$)

    c. Dividir cada columna de V por su norma, obteniendo la matriz V_o. Comprobar que $A = V_o \times D \times V_o^T$. ¿Cómo haría para invertir la matriz A?

3. Considerar la matriz $B = \begin{pmatrix} 10 & -7 & 6 \\ -5 & 6 & 2 \end{pmatrix}$

    a. Calcular los autovalores y autovectores de la matriz $B \times B^T$. Obtener una matriz U disponiendo los autovectores como columnas y dividiendo por su norma.

    b. Calcular los autovalores y autovectores de la matriz $B^T \times B$ (Ayuda: los autovalores son los mismos que los de $B \times B^T$, al que se le agrega un cero). Obtener una matriz V disponiendo los autovectores como columnas y dividiendo por su norma.

    c. Descomponer la matriz B según sus valores singulares ($B = U \times D \times V^T$)

Vectores aleatorios

4. Considerar los siguientes datos:

x	y
5820	5880
6030	5970
5850	6060
6120	5940
6240	6000
6000	6240

a. Calcular el vector de medias, la matriz de covarianza y la matriz de correlación para estos datos.

b. Calcular la distancia de Mahalanobis y la densidad normal multivariada de los primeros 3 datos utilizando el vector de medias $\mu = (6000, 6000)$ y matriz de covarianza $\Sigma = \begin{pmatrix} 185 & -80 \\ -80 & 65 \end{pmatrix}$. ¿Qué observa?

Bases de datos

5. Considerar las bases `df_a.csv`, `df_b.csv` y `df_c.csv`

    a. Completar el siguiente código para hacer un scatterplot de ambas variables en cada base.

```
scatGG=function(Base,x.Ch,y.Ch){
  GG=_____ %>%
    ggplot(aes(x=get(______),y=get(____)))+
    geom_point(alpha=0.5)
  return(GG)
}

textGG=function(GG,tL,xL,yL,tS,aTS,atS){
  GG=GG+
    labs(title = ______)+
    xlab(___)+
    ylab(____)+
    theme_bw()+
    theme(title = element_text(size=_____,face = "bold"),
          axis.title = element_text(size=____,face = "plain"),
          axis.text = element_text(size=___),
          plot.title = element_text(hjust=0.5))
  return(GG)
}
```

```
b. En cada base, calcular el vector de medias, la matriz de covarianza y la matriz de correlación.

c. Completar el siguiente código para hacer gráficos de densidad conjunta de ambas variables,
utilizando contornos y heatmaps.
```

```
dCont2DGG=function(Base,x.Ch,y.Ch){
  GG=_____ %>%
      ggplot(aes(x=get(____),y=get(____)))+
      stat_density_2d(aes(color = ..level..)) +
      scale_color_viridis_c()
  return(GG)
}

dFill2DGG=function(Base,x.Ch,y.Ch){
    GG=Base %>%
      ggplot(aes(x=get(_____),y=get(_____)))+
      stat_density_2d(aes(fill = ..density..), geom = "raster", contour = FALSE) +
      scale_fill_viridis_c()
  return(GG)
}
```

d. Completar el siguiente código para hacer histogramas y qq-plots de cada variable, identificar cuáles de las variables tienen distribución parecida a una normal.

```
GG_qqNorm=function(Base,y.Ch){
  my=mean(Base[,_____])
  sy=sd(_____[,____])
  GG=Base %>%
    ggplot(aes(sample=get(y.Ch)))+
    stat_qq(distribution = qnorm,dparams = list(mean=____,sd=____))+
    stat_qq_line(distribution = qnorm,dparams = list(mean=____,sd=___),linetype="dashed",color="red")
  return(GG)
}

histGG_Norm=function(Base,x.Ch){
  mx=mean(____[,_____])
  sx=_____(Base[,x.Ch])
  xs=seq(mx-2.5*sx,mx+2.5*sx,length.out=500)
  ds=dnorm(xs,mx,sx)
  df_d=data.frame(x=xs,d=ds)
  GG=Base %>%
    ggplot(aes(x=get(_____)))+
    geom_histogram(mapping=aes(y=..density..),color="blue",fill="white")+
    geom_line(data = _____,mapping = aes(x=x,y=d),color="red")
  return(GG)
}
```

e. ¿Cuál de las bases tiene una distribución de sus columnas que se asemeja a una normal multivariada?