

Análisis Multivariado

William Gomez

2024-08-15

Tarea 1

1. El archivo ‘anexo1.csv’ contiene los datos de una muestra de vectores aleatorios de 3 componentes.

$$X_1, X_2, \dots, X_n$$

- a. Evalúe la normalidad multivariada de la muestra dada.
b. Si el vector de medias poblacionales es $\mu = [0.1 \quad -0.2 \quad 0.05]^T$ y S es la matriz de varianzas-covarianzas muestrales. ¿Cuál es la distribución aproximada de

$$40 \left(\bar{X} - [0.1 \quad -0.2 \quad 0.05]^T \right) S^{-1} \left(\bar{X} - [0.1 \quad -0.2 \quad 0.05]^T \right)?$$

- c. Usando la **distancia cuadrada generalizada** establezca si existen valores atípicos.
-

```
X1 <- read.csv2("anexo1.csv")
n <- nrow(X1)
n
```

Solución 1

```
## [1] 40
```

```
p <- ncol(X1)
p
```

```
## [1] 3
```

```
head(X1)
```

```
##           X1           X2           X3
## 1 -2.3932177 -2.1584719 -1.5865685
## 2  2.7740539  2.8226023  3.2069833
## 3  0.5384641  0.1086596 -0.4906402
## 4  2.0426030  1.8232092  1.3567716
## 5  1.7732396  2.2194384  1.7299784
## 6 -0.9932595 -0.9960262 -0.8165494
```

a. Para evaluar la normalidad multivariada de la muestra, usaremos el test de Mardia.

```
#Cargar la Librería para el Test de Mardia
library(MVN)

mardia_test <- mvn(X1, mvnTest = "mardia")
mardia_test

## $multivariateNormality
##           Test           Statistic           p value Result
## 1 Mardia Skewness    3.44263279082527 0.969010425206423    YES
## 2 Mardia Kurtosis -0.959974119510589 0.337068240452895    YES
## 3           MVN              <NA>              <NA>    YES
##
## $univariateNormality
##           Test Variable Statistic    p value Normality
## 1 Anderson-Darling    X1      0.1901    0.8931    YES
## 2 Anderson-Darling    X2      0.1688    0.9296    YES
## 3 Anderson-Darling    X3      0.3570    0.4388    YES
##
## $Descriptives
##      n      Mean Std.Dev   Median      Min      Max    25th    75th
## X1 40  0.013463668 1.816194 -0.06227406 -3.952481 3.673581 -1.2041901 1.196851
## X2 40 -0.001685572 1.768519 -0.02411968 -3.656188 3.713670 -1.0884103 1.043540
## X3 40 -0.082345327 1.659509 -0.45863078 -3.778099 3.206983 -0.9503632 1.102676
##           Skew   Kurtosis
## X1 -0.02735881 -0.5332226
## X2  0.02255044 -0.4845468
## X3 -0.14809282 -0.3814142
```

Podemos observar que el test de normalidad multivariada nos arroja resultados afirmativos tanto en el 'Mardia Skewness' como en el 'Mardia Kurtosis', lo que sugiere que existe normalidad multivariada en la muestra. Es decir, las variables consideradas en la muestra se distribuyen de manera normal multivariada por el test de Mardia.

b.

La estadística de prueba

$$n(X - \mu)^\top S^{-1}(X - \mu) \approx \chi_p^2$$

con $\mu = [0.1 \quad -0.2 \quad 0.05]^\top$, S y $n = 40$, tiene una distribución

$$\chi_p^2$$

Esto significa que la estadística de prueba sigue una distribución chi-cuadrada con p grados de libertad.

```
Xbarra<-colMeans(X1)
S<-cov(X1)
S_inv<-solve(S)
mu <- c(0.1, -0.2, 0.05)
result <- 40 * t(Xbarra - mu) %*% S_inv %*% (Xbarra - mu)
result
```

```
##           [,1]
## [1,] 73.3465
```

```
alpha <- 0.05
df <- 3 # grados de libertad
chi_square_critical <- qchisq(1 - alpha, df)
chi_square_critical
```

```
## [1] 7.814728
```

La estadística de prueba calculada (73.3465) es mucho mayor que el valor crítico (7.814728) para $\alpha=0.05$ y 3 grados de libertad. Por tanto, se rechaza la hipótesis nula y se indica que hay una diferencia significativa entre la media muestral y el vector de medias poblacionales especificado.

c.

```
# Calcular la distancia cuadrada generalizada para cada observación
distances <- apply(X1, 1, function(x) {
  (x - Xbarra) %*% S_inv %*% (x - Xbarra)
})

distances
```

```
## [1] 2.536907504 4.811323299 2.644067334 1.522594023 5.308019728 0.412991046
## [7] 3.363817514 1.457424912 5.522028495 2.526179095 5.649669018 2.929268078
## [13] 6.040574759 4.189424293 3.843993707 1.764055796 5.306311722 1.500701698
## [19] 4.500066237 1.184538732 6.252197434 4.782488434 0.003357339 0.691094899
## [25] 2.570044223 1.173329974 0.098347107 0.147513309 6.855704467 2.768477762
## [31] 0.325131091 1.172094524 0.538101578 1.886409407 6.815650410 3.077184194
## [37] 3.160333807 5.005251987 1.606795282 1.056535781
```

Una vez calculadas las distancias cuadradas generalizadas, calculamos el valor crítico de una distribución chi-cuadrada con p grados de libertad, con el fin de comparar estos valores y decidir sobre los outliers.

```
alpha <- 0.05 # Nivel de significancia
df <- ncol(X1) - 1 # Grados de libertad (número de p - 1)
quantile_chi2 <- qchisq(1 - alpha, df)
quantile_chi2
```

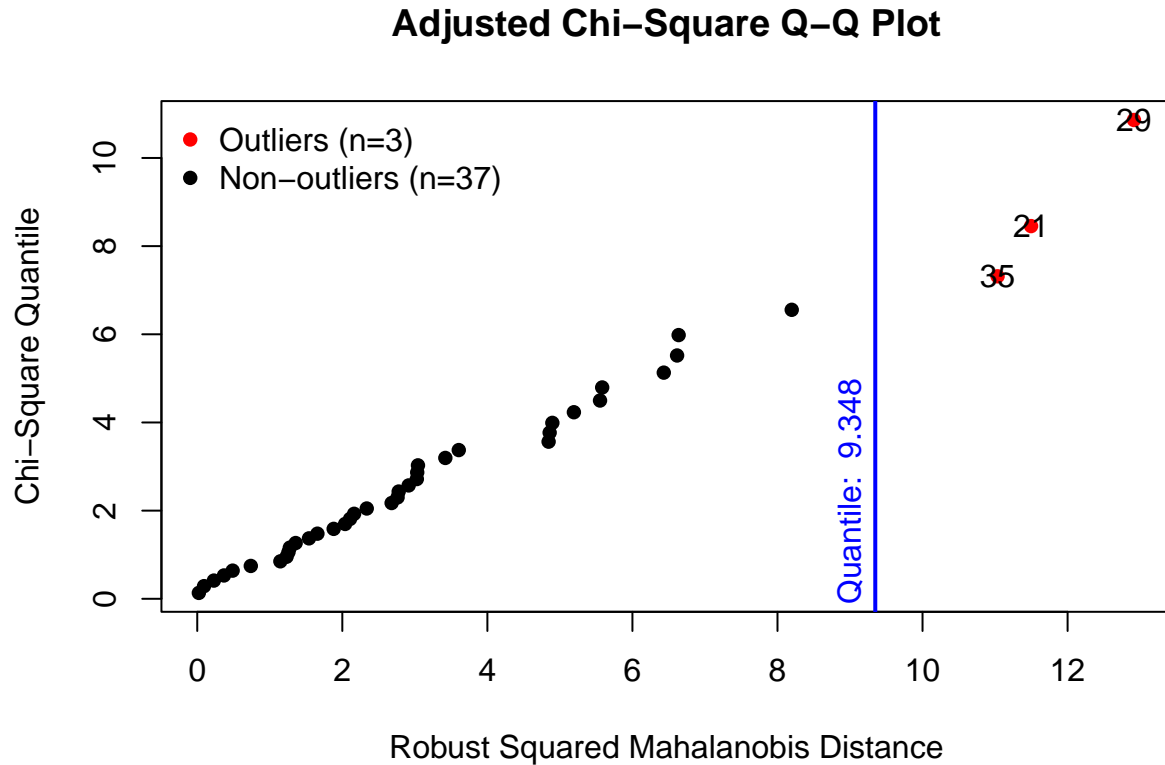
```
## [1] 5.991465
```

```
# Paso 3: Comparar distancias con el cuantil
outliers <- distances[distances > quantile_chi2] # Valores atípicos
outlier_indices <- which(distances > quantile_chi2) # Índices de los valores atípicos
outlier_indices
```

```
## [1] 13 21 29 35
```

Podemos también utilizar un método gráfico.

```
Out<-mvn(X1, mvnTest = "mardia",
multivariateOutlierMethod = "adj")
```



En conclusión, al observar las distancias cuadradas generalizadas, calcular un nivel de significancia y compararlo con el valor crítico de la distribución chi-cuadrada, así como al utilizar el método gráfico para corroborar, los valores atípicos son los índices: [13, 21, 29, 35].

2. Los datos Protein del paquete MultBiplotR contiene información sobre datos nutricionales de 9 diferentes fuentes de proteínas para los habitantes de 25 países europeos alrededor de 1970:

Consumo de Proteínas por Fuente

- **RedMeat:** Consumo de proteínas provenientes de carnes rojas.
- **WhiteMeat:** Consumo de proteínas provenientes de carnes blancas.
- **Eggs:** Consumo de proteínas del huevo.
- **Milk:** Consumo de proteínas de la leche.
- **Fish:** Consumo de proteínas provenientes del pescado.
- **Cereals:** Consumo de proteínas procedentes de cereales.
- **Starch:** Consumo de proteínas provenientes de carbohidratos.
- **Nuts:** Consumo de proteínas procedentes de cereales, frutos secos y semillas oleaginosas.
- **FruitVeg:** Consumo de proteínas procedentes de frutas y verduras.

Estos datos fueron colectados inicialmente para entender las diferencias nutricionales entre los países europeos.

- Determine y analice el vector de medias y la matriz de covarianzas muestrales para las diferentes regiones.
- Calcule la media de las variables por regiones. ¿Qué puede decir al respecto?
- Intente construir grupos de países usando representaciones pictóricas (gráficos de estrellas o caras de Chernoff).
- Utilice las herramientas de gráficas más adecuadas para verificar normalidad multivariada.
- Realice la prueba de Mardia para verificar las hipótesis:
 - H_0 : Los datos provienen de una población Normal Multivariada.
 - H_1 : Los datos NO provienen de una población Normal Multivariada.
- Verifique si hay outliers (multivariados) e identifíquelos.
- Pruebe las hipótesis:
 - $H_0 : \mu = \mu_0$
 - $H_1 : \mu \neq \mu_0$
 donde $\mu_0 = [9 \ 7 \ 2 \ 15 \ 5 \ 30 \ 4 \ 3 \ 4]^T$.
 - De forma univariada.
 - De forma multivariada.

Comente los resultados.

```
library(MultBiplotR)
```

```
data("Protein")

region_Sur <- subset(Protein, Region == "South")
region_centro <- subset(Protein, Region == "Center")
region_norte <- subset(Protein, Region == "North")
```

- En el dataset existen 3 regiones.

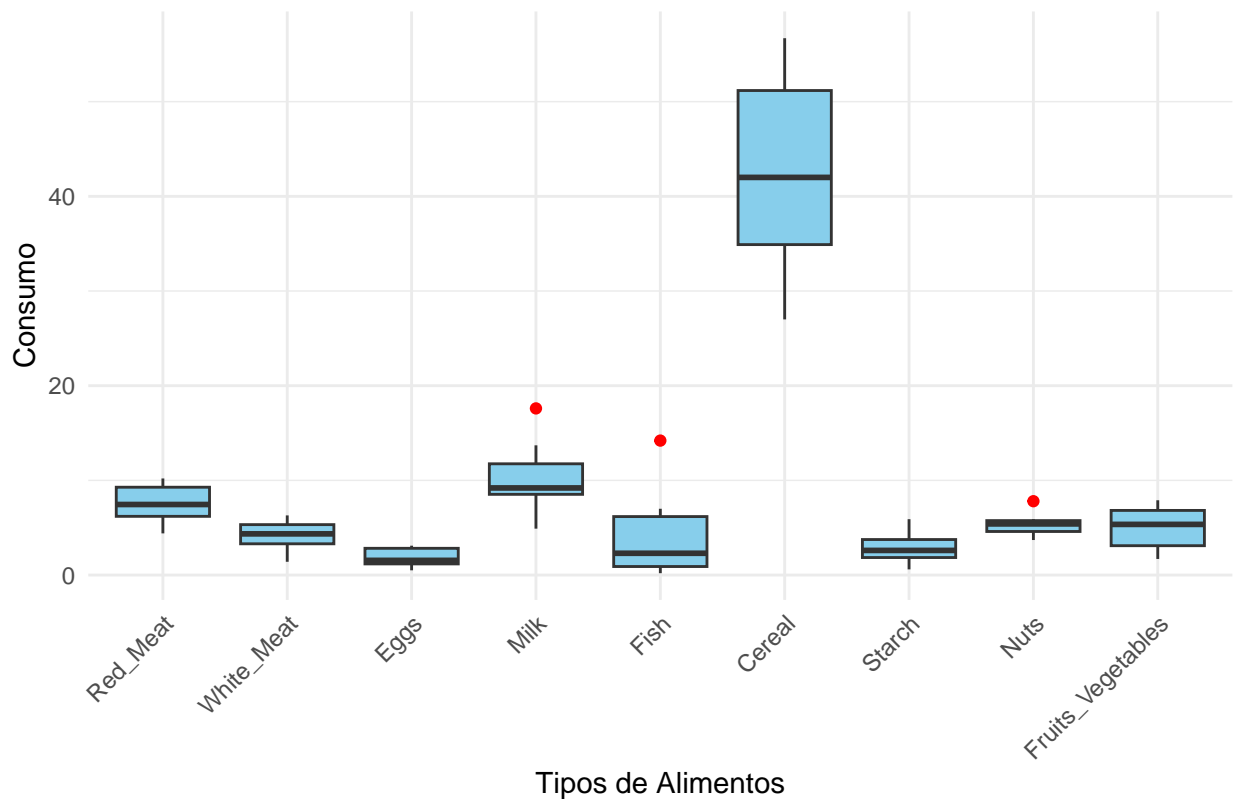
```
summary(region_Sur)
```

Region Sur:

##	Comunist	Region	Red_Meat	White_Meat	Eggs
##	No :4	North :0	Min. : 4.400	Min. :1.400	Min. :0.500
##	Yes:4	Center:0	1st Qu.: 6.200	1st Qu.:3.300	1st Qu.:1.175
##		South :8	Median : 7.450	Median :4.350	Median :1.550
##			Mean : 7.625	Mean :4.237	Mean :1.837
##			3rd Qu.: 9.275	3rd Qu.:5.325	3rd Qu.:2.825

```
##           Max. :10.200  Max. : 6.300  Max. : 3.100
##      Milk           Fish           Cereal           Starch
## Min.   : 4.900   Min.   : 0.200   Min.   :27.00   Min.   :0.600
## 1st Qu.: 8.525   1st Qu.: 0.900   1st Qu.:34.90   1st Qu.:1.850
## Median : 9.200   Median : 2.300   Median :42.00   Median :2.600
## Mean   :10.325   Mean   : 4.188   Mean   :42.40   Mean   :2.962
## 3rd Qu.:11.750   3rd Qu.: 6.175   3rd Qu.:51.17   3rd Qu.:3.750
## Max.   :17.600   Max.   :14.200   Max.   :56.70   Max.   :5.900
##      Nuts      Fruits_Vegetables
## Min.   :3.700   Min.   :1.700
## 1st Qu.:4.600   1st Qu.:3.100
## Median :5.400   Median :5.350
## Mean   :5.362   Mean   :5.025
## 3rd Qu.:5.750   3rd Qu.:6.825
## Max.   :7.800   Max.   :7.900
```

Boxplot de Consumo de Alimentos en la Región Sur



La región Sur se caracteriza por un elevado consumo de cereal por encima que cualquier otra cosa. El consumo de leche tiende a ser también algo común en esta región.

```
summary(region_centro)
```

Region Centro:

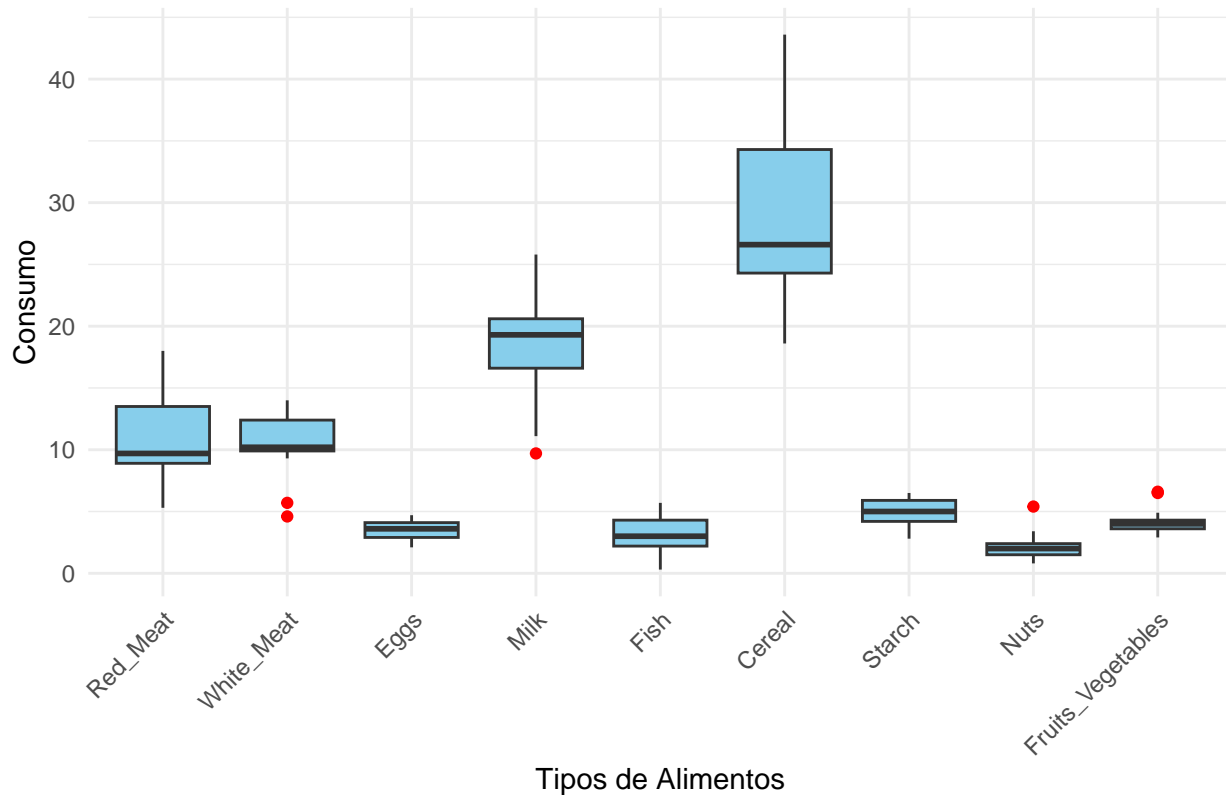
```
## Communist      Region      Red_Meat      White_Meat      Eggs
```

```

## No :8      North : 0      Min.   : 5.30      Min.   : 4.60      Min.   :2.100
## Yes:5      Center:13      1st Qu.: 8.90      1st Qu.: 9.90      1st Qu.:2.900
##                               South : 0      Median : 9.70      Median :10.20      Median :3.600
##                               Mean   :11.18      Mean   :10.41      Mean   :3.546
##                               3rd Qu.:13.50      3rd Qu.:12.40      3rd Qu.:4.100
##                               Max.   :18.00      Max.   :14.00      Max.   :4.700
##      Milk      Fish      Cereal      Starch      Nuts
## Min.   : 9.70   Min.   :0.300   Min.   :18.60   Min.   :2.8     Min.   :0.800
## 1st Qu.:16.60   1st Qu.:2.200   1st Qu.:24.30   1st Qu.:4.2     1st Qu.:1.500
## Median :19.30   Median :3.000   Median :26.60   Median :5.0     Median :2.000
## Mean   :18.35   Mean   :3.131   Mean   :28.95   Mean   :5.0     Mean   :2.246
## 3rd Qu.:20.60   3rd Qu.:4.300   3rd Qu.:34.30   3rd Qu.:5.9     3rd Qu.:2.400
## Max.   :25.80   Max.   :5.700   Max.   :43.60   Max.   :6.5     Max.   :5.400
## Fruits_Vegetables
## Min.   :2.900
## 1st Qu.:3.600
## Median :4.000
## Mean   :4.208
## 3rd Qu.:4.300
## Max.   :6.600

```

Boxplot de Consumo de Alimentos en la Región Centro



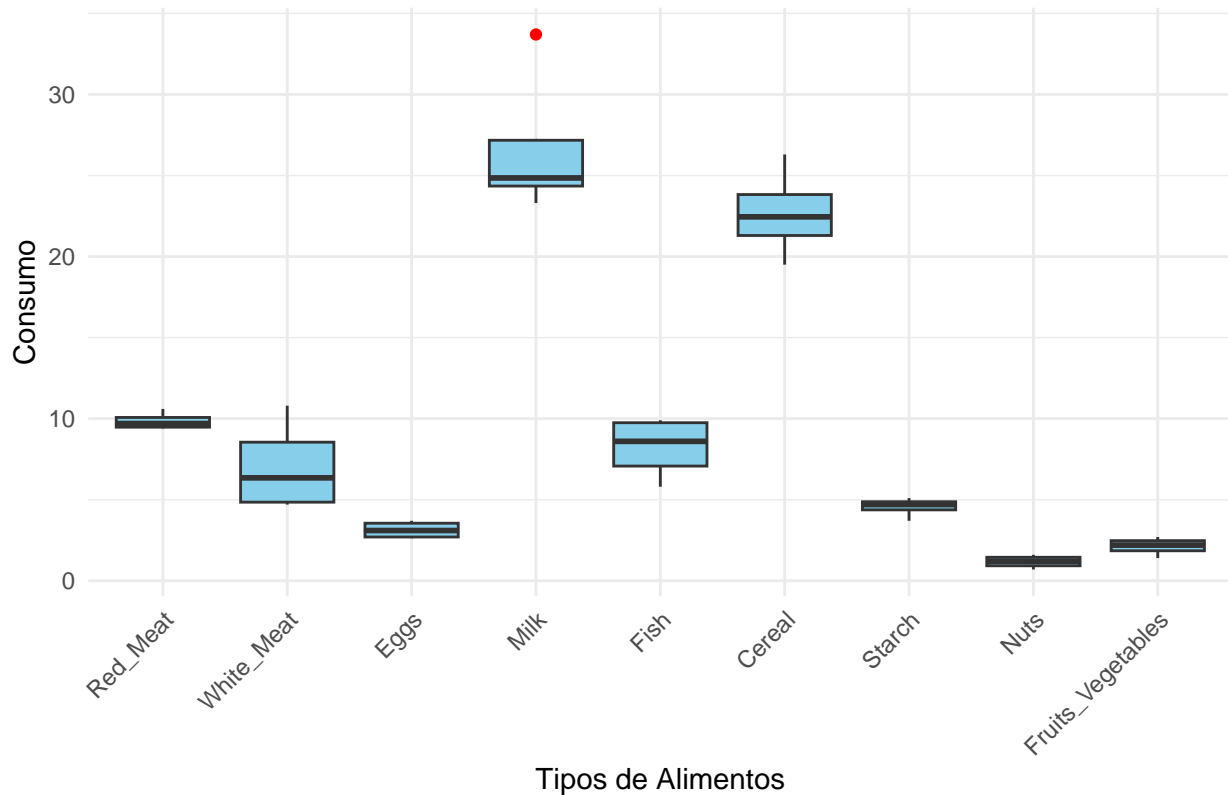
La región Centro se caracteriza también por un consumo elevado de Cereal, pero en menor cantidad. Sin embargo, su población muestra una preferencia mayor a la leche que ne la región Sur. Las carnes rojas y blancas son igualmente una parte importante en la dieta de esta región.

```
summary(region_norte)
```

Region Norte:

```
##  Comunist      Region      Red_Meat      White_Meat      Eggs
##  No :4      North :4      Min.   : 9.400      Min.   : 4.70      Min.   :2.70
##  Yes:0      Center:0      1st Qu.: 9.475      1st Qu.: 4.85      1st Qu.:2.70
##                      South :0      Median : 9.700      Median : 6.35      Median :3.10
##                      Mean   : 9.850      Mean   : 7.05      Mean   :3.15
##                      3rd Qu.:10.075      3rd Qu.: 8.55      3rd Qu.:3.55
##                      Max.   :10.600      Max.   :10.80      Max.   :3.70
##      Milk      Fish      Cereal      Starch
##  Min.   :23.30      Min.   :5.800      Min.   :19.50      Min.   :3.700
##  1st Qu.:24.35      1st Qu.:7.075      1st Qu.:21.30      1st Qu.:4.375
##  Median :24.85      Median :8.600      Median :22.45      Median :4.700
##  Mean   :26.68      Mean   :8.225      Mean   :22.68      Mean   :4.550
##  3rd Qu.:27.18      3rd Qu.:9.750      3rd Qu.:23.82      3rd Qu.:4.875
##  Max.   :33.70      Max.   :9.900      Max.   :26.30      Max.   :5.100
##      Nuts      Fruits_Vegetables
##  Min.   :0.700      Min.   :1.400
##  1st Qu.:0.925      1st Qu.:1.850
##  Median :1.200      Median :2.200
##  Mean   :1.175      Mean   :2.125
##  3rd Qu.:1.450      3rd Qu.:2.475
##  Max.   :1.600      Max.   :2.700
```


Boxplot de Consumo de Alimentos en la Región Norte



La región Norte se caracteriza por el consumo elevado de leche seguido a su vez por un gran consumo de cereal. Entre los alimentos con importancia en su población están las carnes rojas, blancas y el pescado.

```
matriz_covarianza_por_region <- lapply(split(Protein[, -c(1, 2)], Protein$Region), cov)
matriz_covarianza_por_region
```

```
## $North
##           Red_Meat White_Meat      Eggs      Milk      Fish
## Red_Meat    0.2966667  1.5533333  0.2633333 -0.7650000  0.4683333
## White_Meat  1.5533333  8.2566667  1.4500000 -4.9783333  2.4950000
## Eggs        0.2633333  1.4500000  0.2766667 -1.0850000  0.3650000
## Milk       -0.7650000 -4.9783333 -1.0850000 22.4825000 -7.7958333
## Fish         0.4683333  2.4950000  0.3650000 -7.7958333  3.7958333
## Cereal      -0.7183333 -4.6150000 -1.1050000 10.6458333 -2.4358333
## Starch      -0.0233333 -0.3333333 -0.1433333  1.6516667 -0.0750000
## Nuts        -0.1583333 -0.7450000 -0.0983333 -0.7708333  0.0308333
## Fruits_Vegetables 0.0650000  0.3816667  0.0583333 -2.4158333  1.0525000
##           Cereal      Starch      Nuts Fruits_Vegetables
## Red_Meat   -0.7183333 -0.0233333 -0.1583333      0.0650000
## White_Meat -4.6150000 -0.3333333 -0.7450000      0.3816667
## Eggs       -1.1050000 -0.1433333 -0.0983333      0.0583333
## Milk       10.6458333  1.6516667 -0.7708333     -2.4158333
## Fish       -2.4358333 -0.0750000  0.0308333      1.0525000
## Cereal      7.9758333  1.5050000 -0.2808333     -0.7525000
## Starch      1.5050000  0.3633333 -0.1283333     -0.0650000
## Nuts       -0.2808333 -0.1283333  0.1625000      0.0708333
```

```

## Fruits_Vegetables -0.7525000 -0.06500000 0.07083333 0.31583333
##
## $Center
##           Red_Meat White_Meat      Eggs      Milk      Fish
## Red_Meat      14.6735897 -4.5664744  1.51615385  9.7978205  3.3441026
## White_Meat     -4.5664744  7.6274359  0.37628205 -1.1695513 -1.3427564
## Eggs           1.5161538  0.3762821  0.65935897  1.6885256  0.2859615
## Milk           9.7978205 -1.1695513  1.68852564 23.8326923  0.2542949
## Fish           3.3441026 -1.3427564  0.28596154  0.2542949  2.2839744
## Cereal        -14.0588462 -6.8528846 -4.71480769 -18.2364744 -4.2882051
## Starch         -0.2300000 -1.2300000 -0.03666667 -1.2075000  0.7283333
## Nuts           -0.3980128 -1.2287179 -0.30147436 -1.4564744 -0.6632051
## Fruits_Vegetables 0.1743590  0.6391026 -0.31621795  0.2037821  0.3364103
##           Cereal      Starch      Nuts Fruits_Vegetables
## Red_Meat      -14.058846 -0.23000000 -0.39801282      0.17435897
## White_Meat     -6.852885 -1.23000000 -1.22871795      0.63910256
## Eggs          -4.714808 -0.03666667 -0.30147436     -0.31621795
## Milk          -18.236474 -1.20750000 -1.45647436      0.20378205
## Fish          -4.288205  0.72833333 -0.66320513      0.33641026
## Cereal         54.344359  1.26000000  4.96185897      1.14711538
## Starch         1.260000  1.29666667 -0.33250000     -0.31750000
## Nuts           4.961859 -0.33250000  1.51769231     -0.02288462
## Fruits_Vegetables 1.147115 -0.31750000 -0.02288462      1.38243590
##
## $South
##           Red_Meat White_Meat      Eggs      Milk      Fish
## Red_Meat      4.2021429 -1.84392857  0.5017857  4.24071429 -0.9567857
## White_Meat    -1.8439286  2.75125000  0.1798214 -0.04107143 -1.9708929
## Eggs           0.5017857  0.17982143  0.9369643  2.15178571  0.9119643
## Milk           4.2407143 -0.04107143  2.1517857 14.87642857 -5.3067857
## Fish          -0.9567857 -1.97089286  0.9119643 -5.30678571 22.7955357
## Cereal        -4.5957143  9.42857143 -3.7471429  6.85000000 -43.1157143
## Starch        -2.1803571  0.00875000  0.4530357 -3.12892857  7.2708929
## Nuts           0.6467857 -1.03125000  0.3430357  2.80821429  0.4708929
## Fruits_Vegetables 0.2207143 -0.24964286  1.4832143  0.46071429  9.3532143
##           Cereal      Starch      Nuts Fruits_Vegetables
## Red_Meat      -4.595714 -2.1803571  0.6467857      0.2207143
## White_Meat     9.428571  0.0087500 -1.0312500     -0.2496429
## Eggs          -3.747143  0.4530357  0.3430357      1.4832143
## Milk           6.850000 -3.1289286  2.8082143      0.4607143
## Fish          -43.115714  7.2708929  0.4708929      9.3532143
## Cereal        125.977143 -14.4157143 -1.7614286     -19.3600000
## Starch        -14.415714  3.7883929  0.1941071      2.9782143
## Nuts          -1.761429  0.1941071  1.5283929      0.2167857
## Fruits_Vegetables -19.360000  2.9782143  0.2167857      5.4278571

```

Por otro lado, la matriz de covarianzas muestra alta varianza en 'Milk' (22.4825), lo que indica un consumo muy variable. Existe una covarianza positiva entre 'Red_Meat' y 'White_Meat' (1.5533), sugiriendo que a medida que aumenta el consumo de carne roja, también lo hace el de carne blanca. Sin embargo, la covarianza negativa entre 'Milk' y 'Fish' (-7.7958) indica que los hogares que consumen más leche tienden a consumir menos pescado.

En la region Centro se observa una alta varianza en 'Milk' (23.8327) y una covarianza negativa entre 'Red_Meat' y 'White_Meat' (-4.5665), sugiriendo que un aumento en el consumo de carne roja está asociado

con una disminución en el de carne blanca. Además, la covarianza negativa en Cereal con otras variables sugiere relaciones inversas en el consumo, lo que podría reflejar preferencias alimentarias distintivas.

En la región Sur la varianza en 'Fish' (22.7955) es notablemente alta, lo que indica un consumo muy variable. La covarianza negativa entre 'Cereal' y 'Fish' (-43.1157) sugiere que los hogares que consumen más pescado tienden a consumir menos cereal, mientras que la covarianza positiva entre 'Milk' y 'Eggs' (2.1518) sugiere que ambos productos son consumidos conjuntamente en esta región.

b.

```
media_por_region <- aggregate(. ~ Region, data = Protein, FUN = mean)
media_por_region <- media_por_region[, -c(1, 2)]
media_por_region
```

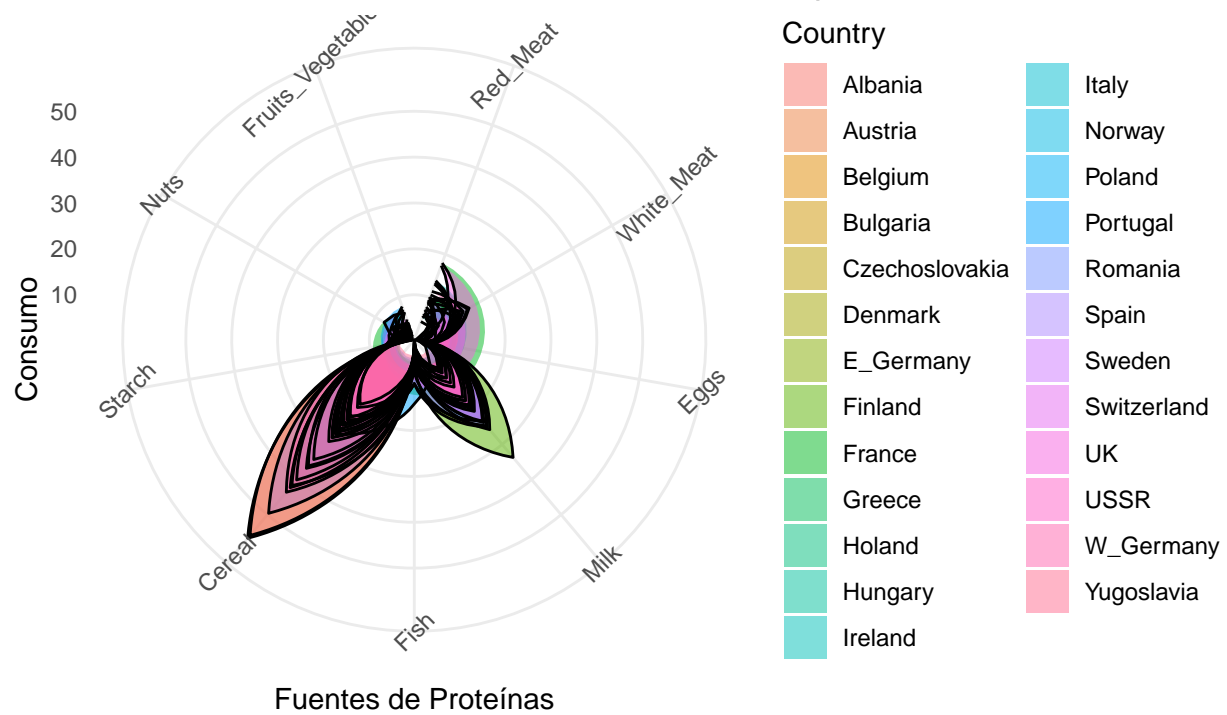
```
##   Red_Meat White_Meat    Eggs    Milk    Fish  Cereal Starch    Nuts
## 1  9.85000    7.05000 3.150000 26.67500 8.225000 22.67500 4.5500 1.175000
## 2 11.17692   10.40769 3.546154 18.34615 3.130769 28.94615 5.0000 2.246154
## 3  7.62500    4.23750 1.837500 10.32500 4.187500 42.40000 2.9625 5.362500
##   Fruits_Vegetables
## 1          2.125000
## 2          4.207692
## 3          5.025000
```

La región que mayor consumo de cereal presenta es la región Sur. Esta región a su vez es la que mayor consumo de frutas y verduras presenta y menor consumo de Nuts. En la región Norte se da el mayor consumo de Leche y Pescado. Mientras que en la region Centro se da el mayor consumo de carnes rojas y blancas.

c.

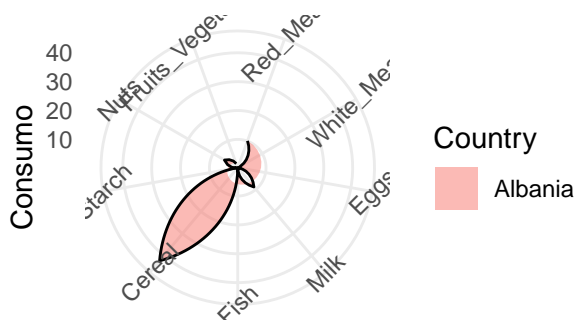
```
library(ggplot2)
library(reshape2)
library(gridExtra)
```

Gráfico de Estrellas de Consumo de Proteínas por País



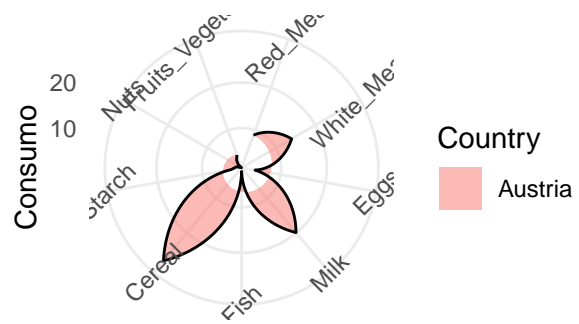
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Consumo de Proteínas – Albania



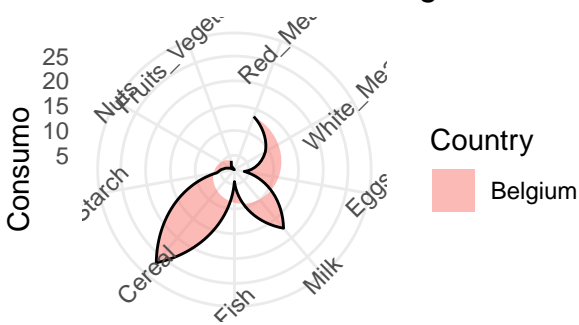
Fuentes de Proteínas

Consumo de Proteínas – Austria



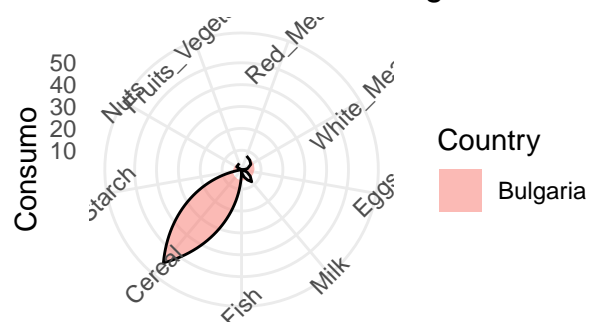
Fuentes de Proteínas

Consumo de Proteínas – Belgium



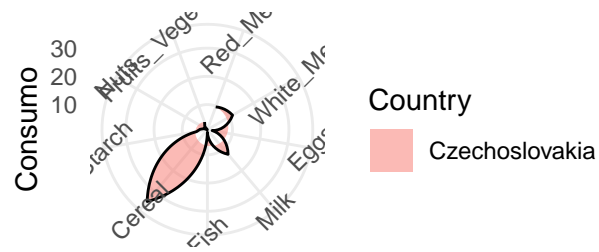
Fuentes de Proteínas

Consumo de Proteínas – Bulgaria



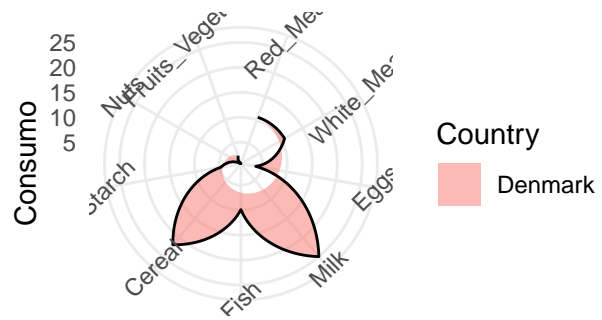
Fuentes de Proteínas

Consumo de Proteínas – Czechoslovakia



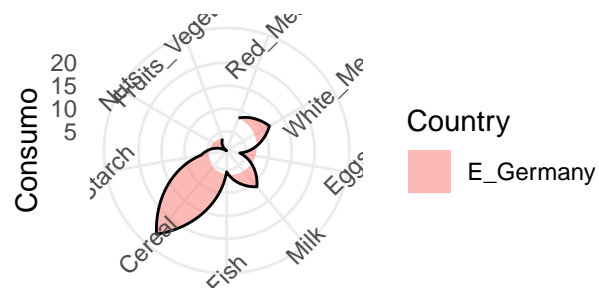
Fuentes de Proteínas

Consumo de Proteínas – Denmark



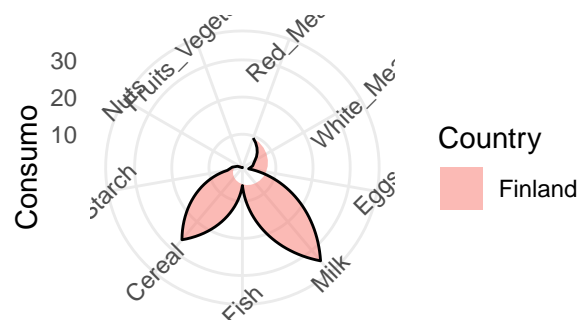
Fuentes de Proteínas

Consumo de Proteínas – E_Germany



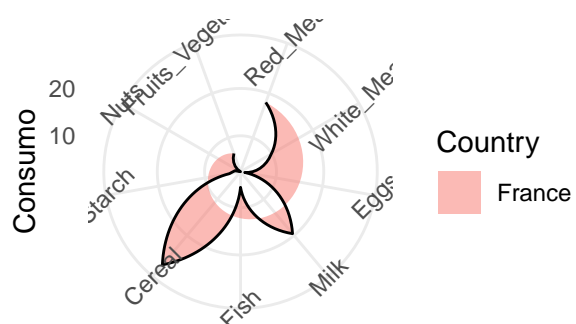
Fuentes de Proteínas

Consumo de Proteínas – Finland



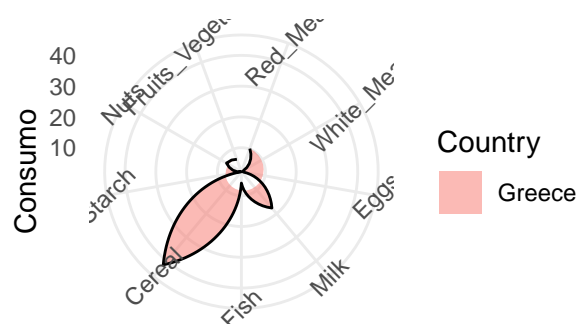
Fuentes de Proteínas

Consumo de Proteínas – France



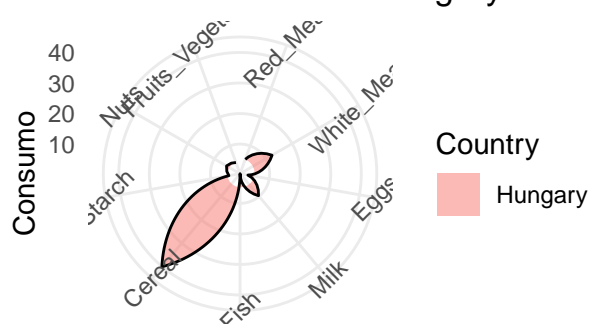
Fuentes de Proteínas

Consumo de Proteínas – Greece



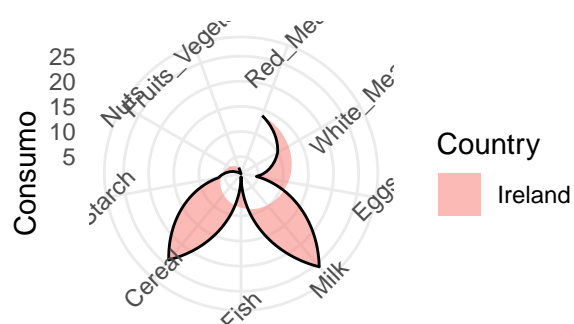
Fuentes de Proteínas

Consumo de Proteínas – Hungary



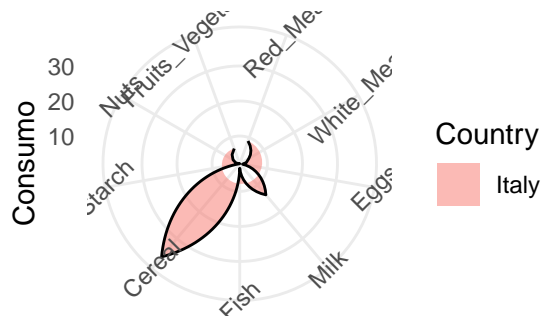
Fuentes de Proteínas

Consumo de Proteínas – Ireland



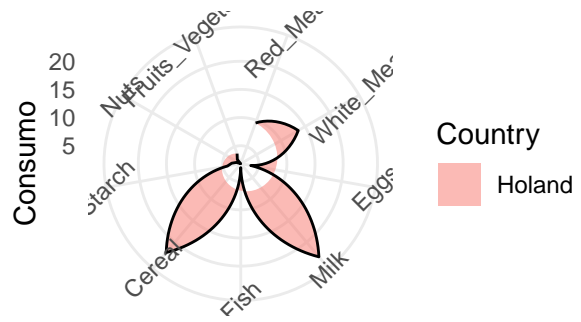
Fuentes de Proteínas

Consumo de Proteínas – Italy



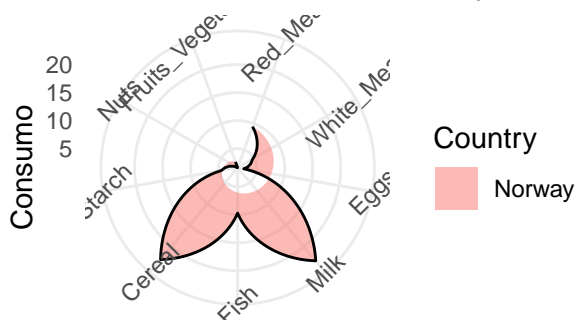
Fuentes de Proteínas

Consumo de Proteínas – Holand



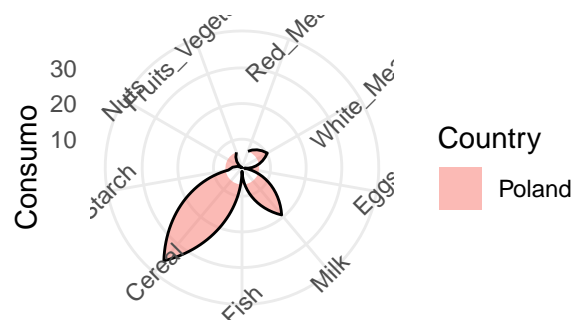
Fuentes de Proteínas

Consumo de Proteínas – Norway



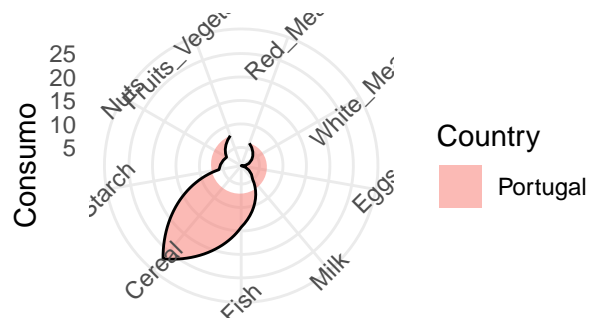
Fuentes de Proteínas

Consumo de Proteínas – Poland



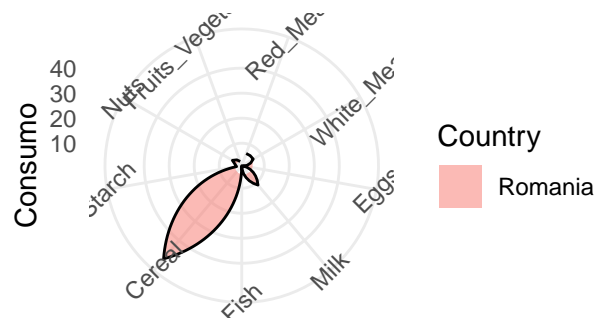
Fuentes de Proteínas

Consumo de Proteínas – Portugal



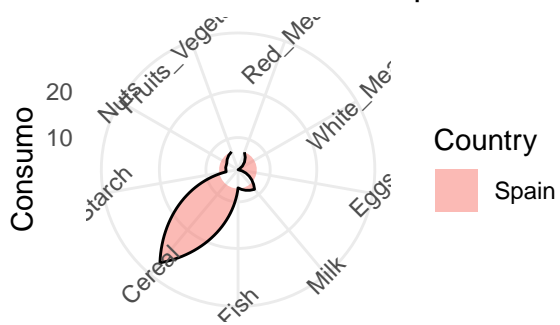
Fuentes de Proteínas

Consumo de Proteínas – Romania



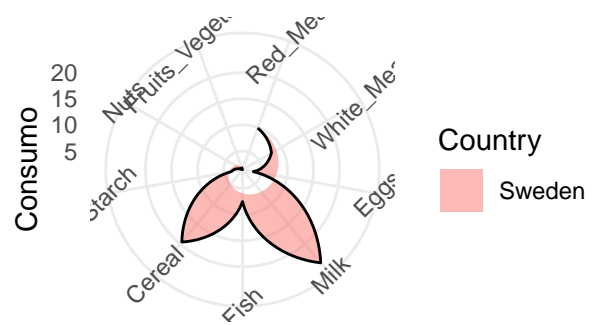
Fuentes de Proteínas

Consumo de Proteínas – Spain



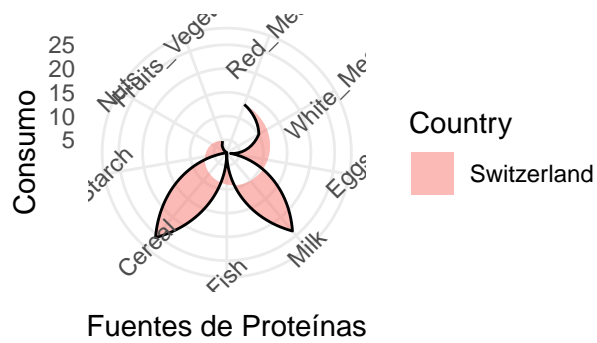
Fuentes de Proteínas

Consumo de Proteínas – Sweden

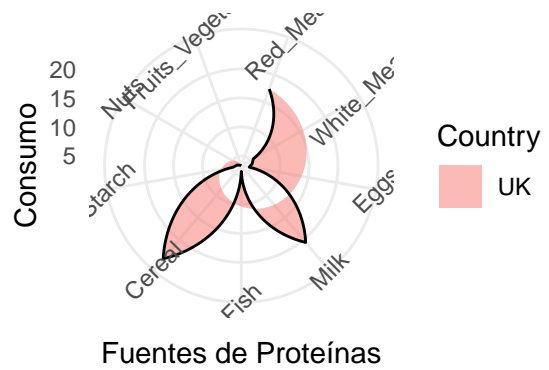


Fuentes de Proteínas

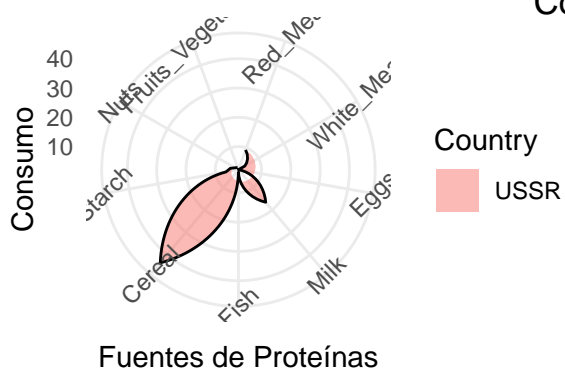
Consumo de Proteínas – Switzerland



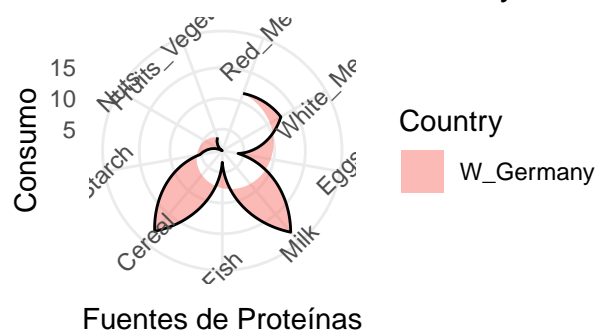
Consumo de Proteínas – UK



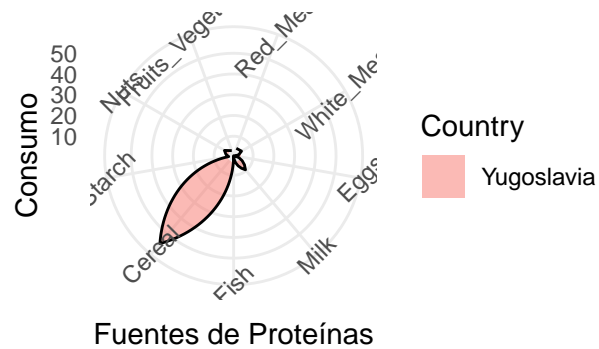
Consumo de Proteínas – USSR



Consumo de Proteínas – W_Germany



Consumo de Proteínas – Yugoslavia



A simple vista puedo sugerir agrupar los países en 2 grupos. Los países de Bulgaria que solo consumen Cereal principalmente pueden agruparse con Albania por ejemplo. Países como Austria que presentan consumo de Cereal, Leche y Carnes Blancas pueden formar un grupo con , Belgium, Denmark, Germany , entre otros.

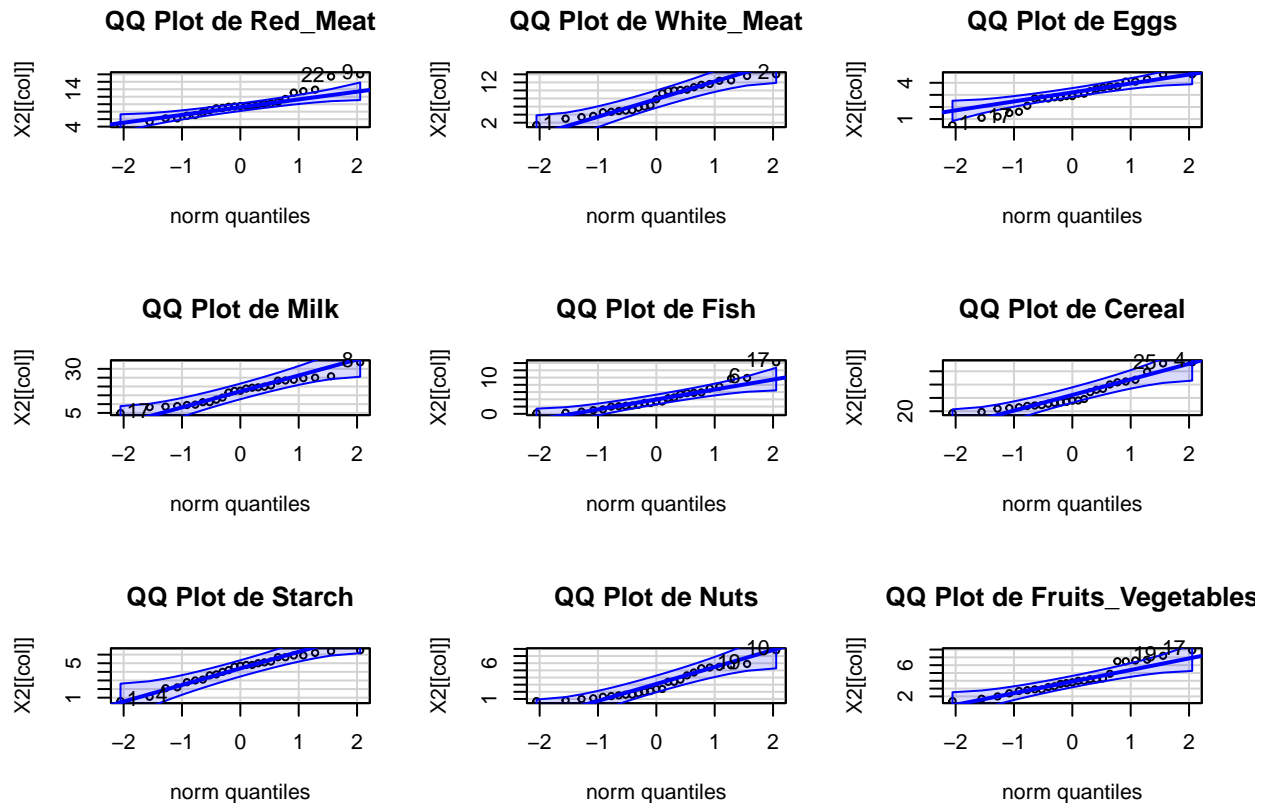
d.

```
##           Red_Meat White_Meat Eggs Milk Fish Cereal Starch Nuts
## Albania          10.1         1.4  0.5  8.9  0.2  42.3   0.6  5.5
## Austria           8.9         14.0  4.3 19.9  2.1  28.0   3.6  1.3
## Belgium          13.5          9.3  4.1 17.5  4.5  26.6   5.7  2.1
## Bulgaria          7.8          6.0  1.6  8.3  1.2  56.7   1.1  3.7
## Czechoslovakia    9.7         11.4  2.8 12.5  2.0  34.3   5.0  1.1
## Denmark          10.6         10.8  3.7 25.0  9.9  21.9   4.8  0.7
##           Fruits_Vegetables
## Albania                   1.7
## Austria                   4.3
## Belgium                   4.0
## Bulgaria                   4.2
## Czechoslovakia            4.0
## Denmark                   2.4
```

```
## Cargando paquete requerido: carData
```

```
##
## Adjuntando el paquete: 'car'
```

```
## The following object is masked from 'package:MultBiplotR':
##
##   logit
```



A simple vista podemos ver cierta discrepancia entre los puntos y la línea recta de color azul. Sin embargo puede que exista normalidad en la mayoría de las variables. e.

Se plantea las siguientes hipótesis para evaluar la normalidad multivariada de los datos:

- **H0:** Los datos provienen de una población Normal Multivariada.
- **H1:** Los datos NO provienen de una población Normal Multivariada.

```
library(MVN)
mardia_result <- mvn(X2, mvnTest = "mardia")
mardia_result
```

```
## $multivariateNormality
##           Test      Statistic      p value Result
## 1 Mardia Skewness 168.086605971262 0.41858425807236 YES
## 2 Mardia Kurtosis -0.523571666842164 0.600576492382074 YES
## 3           MVN           <NA>           <NA> YES
##
## $univariateNormality
##           Test      Variable Statistic      p value Normality
## 1 Anderson-Darling Red_Meat      0.6627      0.0737 YES
```

```

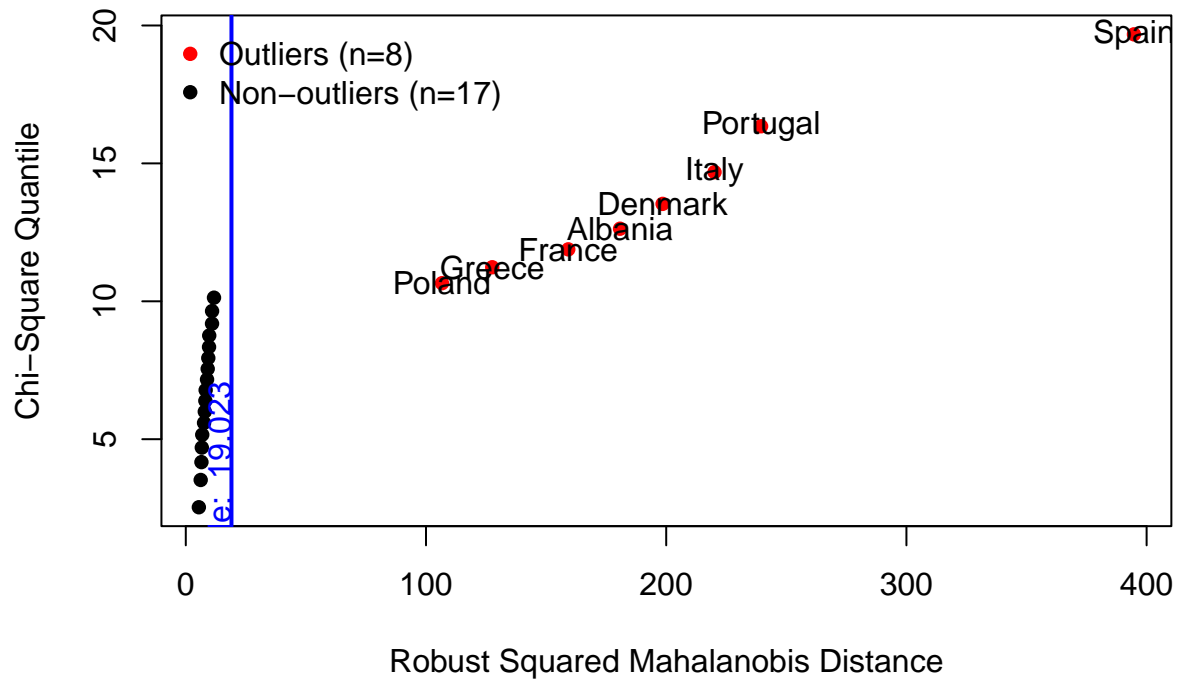
## 2 Anderson-Darling      White_Meat      0.6238      0.0926      YES
## 3 Anderson-Darling      Eggs            0.3929      0.3508      YES
## 4 Anderson-Darling      Milk           0.3966      0.3435      YES
## 5 Anderson-Darling      Fish           0.7257      0.0509      YES
## 6 Anderson-Darling      Cereal        1.0013      0.0101      NO
## 7 Anderson-Darling      Starch        0.4254      0.2921      YES
## 8 Anderson-Darling      Nuts          0.9281      0.0155      NO
## 9 Anderson-Darling      Fruits_Vegetables 0.7753      0.0380      NO
##
## $Descriptives
##              n      Mean      Std.Dev      Median      Min      Max      25th      75th      Skew
## Red_Meat      25      9.828      3.347078      9.5      4.4      18.0      7.8      10.6      0.77847879
## White_Meat    25      7.896      3.694081      7.8      1.4      14.0      4.9      10.8      0.03504453
## Eggs          25      2.936      1.117617      2.9      0.5      4.7      2.7      3.7      -0.39426844
## Milk          25     17.112      7.105416     17.6      4.9     33.7     11.1     23.3      0.21508664
## Fish          25      4.284      3.402533      3.4      0.2     14.2      2.1      5.8      1.09632534
## Cereal        25     32.248     10.974786     28.0     18.6     56.7     24.3     40.1      0.81881605
## Starch        25      4.276      1.634085      4.7      0.6      6.5      3.1      5.7     -0.58880387
## Nuts          25      3.072      1.985682      2.4      0.7      7.8      1.5      4.7      0.63338857
## Fruits_Vegetables 25      4.136      1.803903      3.8      1.4      7.9      2.9      4.9      0.54556050
##
##              Kurtosis
## Red_Meat      0.2225442
## White_Meat    -1.4190709
## Eggs          -0.6646212
## Milk          -0.7787642
## Fish          0.7974802
## Cereal        -0.4981383
## Starch        -0.6590063
## Nuts          -0.8391066
## Fruits_Vegetables -0.8604672

```

Dado que no se rechaza la hipótesis nula (H_0), se concluye que los datos provienen de una población Normal Multivariada. Esto sugiere que las suposiciones de normalidad son válidas para los análisis posteriores.

f.

Adjusted Chi-Square Q-Q Plot



```
## $multivariateNormality
##           Test           Statistic           p value Result
## 1 Mardia Skewness 168.086605971262 0.41858425807236 YES
## 2 Mardia Kurtosis -0.523571666842164 0.600576492382074 YES
## 3 MVN <NA> <NA> YES
##
## $univariateNormality
##           Test           Variable Statistic           p value Normality
## 1 Anderson-Darling Red_Meat 0.6627 0.0737 YES
## 2 Anderson-Darling White_Meat 0.6238 0.0926 YES
## 3 Anderson-Darling Eggs 0.3929 0.3508 YES
## 4 Anderson-Darling Milk 0.3966 0.3435 YES
## 5 Anderson-Darling Fish 0.7257 0.0509 YES
## 6 Anderson-Darling Cereal 1.0013 0.0101 NO
## 7 Anderson-Darling Starch 0.4254 0.2921 YES
## 8 Anderson-Darling Nuts 0.9281 0.0155 NO
## 9 Anderson-Darling Fruits_Vegetables 0.7753 0.0380 NO
##
## $Descriptives
##           n      Mean      Std.Dev Median   Min   Max 25th 75th      Skew
## Red_Meat 25  9.828  3.347078    9.5  4.4 18.0  7.8 10.6  0.77847879
## White_Meat 25  7.896  3.694081    7.8  1.4 14.0  4.9 10.8  0.03504453
## Eggs 25  2.936  1.117617    2.9  0.5  4.7  2.7  3.7 -0.39426844
## Milk 25 17.112  7.105416   17.6  4.9 33.7 11.1 23.3  0.21508664
## Fish 25  4.284  3.402533    3.4  0.2 14.2  2.1  5.8  1.09632534
## Cereal 25 32.248 10.974786   28.0 18.6 56.7 24.3 40.1  0.81881605
```

```
## Starch      25  4.276  1.634085    4.7  0.6  6.5  3.1  5.7 -0.58880387
## Nuts        25  3.072  1.985682    2.4  0.7  7.8  1.5  4.7  0.63338857
## Fruits_Vegetables 25  4.136  1.803903    3.8  1.4  7.9  2.9  4.9  0.54556050
##           Kurtosis
## Red_Meat      0.2225442
## White_Meat    -1.4190709
## Eggs          -0.6646212
## Milk          -0.7787642
## Fish          0.7974802
## Cereal        -0.4981383
## Starch        -0.6590063
## Nuts          -0.8391066
## Fruits_Vegetables -0.8604672
```

8 valores atípicos son encontrados con la prueba gráfico siendo el más alejado de todos ‘España’.

g.

```
mu0 <- c(9, 7, 2, 15, 5, 30, 4, 3, 4)
Xbarra <- colMeans(X2)
S <- apply(X2, 2, sd)

n <- nrow(X2)
n
```

```
## [1] 25
```

```
p <- ncol(X2)
p
```

```
## [1] 9
```

- Univariada:

```
p_values <- sapply(1:p, function(i) {
  t_test <- (Xbarra[i] - mu0[i]) / (S[i] / sqrt(n))
  p_value <- 2 * (1 - pt(abs(t_test), df = n - 1))
  return(p_value)
})
p_values
```

```
##           Red_Meat      White_Meat      Eggs      Milk
##      0.2280913981    0.2370278242    0.0003277511    0.1502482214
##           Fish      Cereal      Starch      Nuts
##      0.3032062741    0.3159698559    0.4067247572    0.8576561693
## Fruits_Vegetables
##      0.7095156149
```

Se observa que solo la variable **Eggs** tiene un p-valor menor a 0.05, lo que indica que se rechaza la hipótesis nula para esta variable, sugiriendo que su media es significativamente diferente de la población de referencia. Las demás variables no presentan diferencias significativas.

- Multivariada:

```
# --- PH multivariadas --- #
T2<-function(mu0,alpha,n,p){
Xbarra<-colMeans(X2)
S<-cov(X2)
InvS<-solve(S)
DifMed<-Xbarra - mu0
T2<-n%*%t(DifMed)%*%InvS%*%DifMed
return(T2)
}
```

```
T2_value <- T2(mu0, alpha = 0.05, n = n, p = p)
T2_value
```

```
##           [,1]
## [1,] 64.84185
```

```
# Comparar con la distribución F
F_value <- (T2_value * (n - p)) / (p * (n - 1))
F_value
```

```
##           [,1]
## [1,] 4.8031
```

```
p_value_multivariado <- 1 - pf(F_value, df1 = p, df2 = n - p)
p_value_multivariado
```

```
##           [,1]
## [1,] 0.00318293
```

El valor de T^2 calculado es **64.84185** y el valor F correspondiente es **4.8031**, lo que da como resultado un p-valor de **0.0032**. Este p-valor es menor que 0.05, lo que indica que se rechaza la hipótesis nula de que los datos provienen de una población Normal Multivariada.