Trabajo Multivariado

Parcial 3

Ana María López - Pedro Pablo Villegas - Esteban Tabares

Noviembre, 2017

PUNTO 5.4

Use los datos del sudor de la tabla 5.1:

Table 5.1 Sweat Data

Individual	X_1 (Sweat rate)	X_2 (Sodium)	X_3 (Potassium)
1	3.7	48.5	9.3
2	5.7	65.1	8.0
3	3.8	47.2	10.9
4	3.2	53.2	12.0
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14.0
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5	71.6	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Source: Courtesy of Dr. Gerald Bargman.

Definimos entonces las variables de las siguiente manera:

X1: Sweat Rate (Tasa de Sudor)
X2: Sodium (Contenido de Sodio)
X3: Potasssium (Contenido de Potasio)

- Determine los ejes de la elipsoide del 90% de confidencia para μ. Determine la longitud de los ejes.

Para el caso de $p = 3$, los ejes de la región de confianza o Elipse de confianza y sus respectivas longitudes relativas, son determinados a partir de los eigen-valores y eigen-vectores de S. Para los datos de la tabla 5.1 tenemos $\underline{\overline{X}}$ y S definida de la siguiente manera:

```
##        [,1]
## [1,]  4.640
## [2,] 45.400
## [3,]  9.965
```

```
##           X1       X2        X3
## X1  2.879368  10.0100 -1.809053
```

```
## X2 10.010000 199.7884 -5.640000
## X3 -1.809053  -5.6400  3.627658
```

Para los eigen-valores y eigen-vectores se tiene:

```
## eigen() decomposition
## $values
## [1] 200.462464   4.531591   1.301392
##
## $vectors
##             [,1]        [,2]        [,3]
## [1,] -0.05084144 -0.57370364  0.81748351
## [2,] -0.99828352  0.05302042 -0.02487655
## [3,]  0.02907156  0.81734508  0.57541452
```

Iniciando en el centro $\underline{X}$, los ejes del elipsoide de confianza son:

$$\pm\sqrt{\lambda_i}\sqrt{\frac{(n-1)p}{(n-p)n}F_{\alpha;p,n-p}e_i}$$

con $Se_i = \lambda_i e_i$ para $i = 1, 2, 3$.

Para el calculo de las semi-longitudes tenemos: $\pm\sqrt{\lambda_i}\sqrt{\frac{(n-1)p}{(n-p)n}F_{\alpha;p,n-p}}$

Tenemos entonces para los datos que las semi-longitudes son las siguientes:

```
##           [,1]
## [1,] 9.0506741
## [2,] 1.3607857
## [3,] 0.7292367
```

Para el calculo de los ejes, usaremos los eigen-vectores, teniendo como resultado lo siguiente:

$$\pm(9.0506741)\begin{bmatrix}-0.05084144\\0.99828352\\0.02907156\end{bmatrix}$$

$$\pm(1.3607857)\begin{bmatrix}-0.57370364\\0.05302042\\0.81734508\end{bmatrix}$$

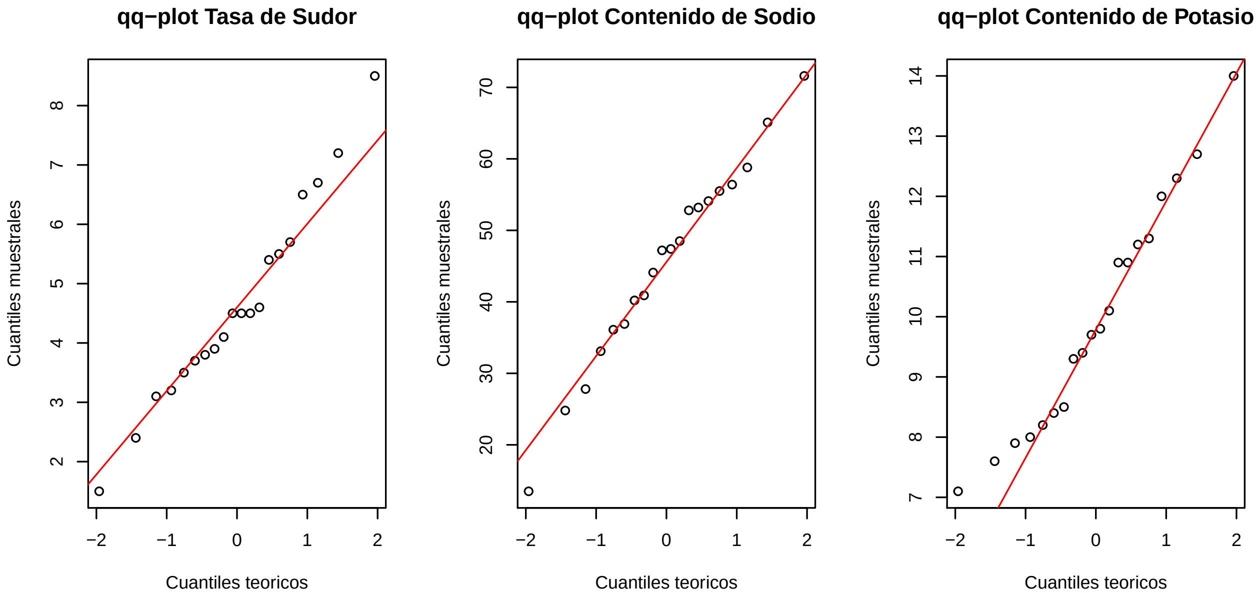
$$\pm(0.7292367)\begin{bmatrix}0.81748351\\-0.02487655\\0.57541452\end{bmatrix}$$

El intervalos T^2 con un nivel de confianza del 90% seria el siguiente:

```
##     N  Media Desv_Estandar    T2_Low     T2_Up
## X1 20  4.640      2.879368  3.555292  5.724708
## X2 20 45.400    199.788421 36.364555 54.435445
## X3 20  9.965      3.627658  8.747476 11.182524
```

- Construya un grafico QQ para las observaciones en tasa de sudor, contenido de sodio y contenido de potasio respectivamente. Construir las tres posibles graficas de dispersión para las parejas de observaciones. Es la suposición de normal multivariada aceptada en este caso?

Se realiza el grafico QQ para cada una de las variables:

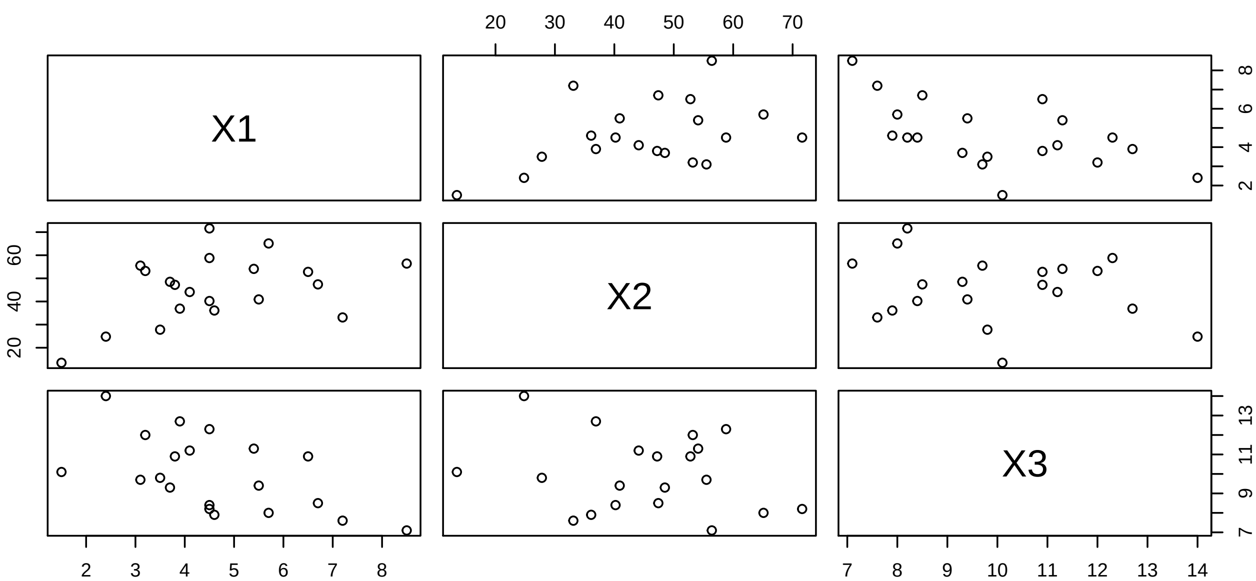


Del gráfico qqplot se observa que X1 y X2 son claramente normales, para X3 observamos que al principio hay un desvío de los quantiles teóricos, sin embargo podemos concluir que son normales pues solo tres observaciones se desvían de los cuantiles teóricos.

```
##                        p-valor
## Tasa de Sudor        0.8689242
## Contenido de Sodio   0.9861883
## Contenido de Potasio 0.6232620
```

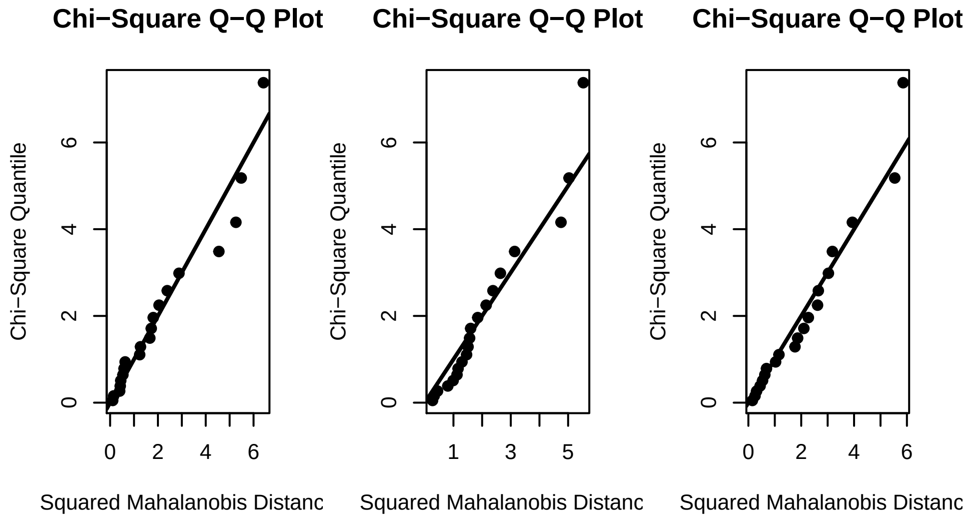
Revisando el resultado de las pruebas formales, se obtiene que los datos son normales ya que tenemos un valor de mayor de 0.5.

Graficos de dispersión:



En los gráficos de dispersión bi-variados se observa que X1-X2 y X1-X3 tienen una clara dispersión elipsoidal lo que nos puede indicar una normalidad bi-variada. Para X2-X3 no es tan claro en la gráfica, pero da la impresión de normalidad, esto lo comprobaremos con los gráficos chi-cuadrado y pruebas de hipótesis de Mardia.

Ahora se realiza el analisis bivariado por las diferentes combinaciones entre las 3 variables:



Se observa que los datos para X1-X2, X1-X3 y X2-X3 siguen claramente la línea de los cuantiles teóricos de una normal bi-variada.

```
##   bivariados p.value.kurt p.value.skew p.value.small
## 1    X1 - X2    0.8133465    0.3263974     0.2070319
## 2    X1 - X3    0.3224551    0.7077270     0.6029315
## 3    X2 - X3    0.4622371    0.8863096     0.8334699
```

De la prueba de Mardia, y los valores p de la kurtosis y la asimetría y con un nivel de confianza del 95%, tenemos que ninguno rechaza la hipótesis nula de que los datos provienen de una distribución normal bi-variada

PUNTO 5.9

Harry Roberts, un naturalista para el departamento de Alaska Fish and Game, estudio los osos pardos con la meta de manterner una población saludable. Mediciones en $n = 61$ de osos proveen el siguiente resumen de estadisticas:

Variable	Peso (Kg)	Longitud Cuerpo (cm)	Cuello (cm)	Cintura (cm)	Longitud Cabeza (cm)	Ancho Cabeza (cm)
Media ($\bar{x}$)	95.52	164.38	55.69	93.39	17.98	31.13

Matriz de covarianza:

$$S = \begin{bmatrix} 3266.46 & 1343.97 & 731.54 & 1175.50 & 162.68 & 238.37 \\ 1343.97 & 721.91 & 324.25 & 537.35 & 80.17 & 117.73 \\ 731.54 & 324.25 & 179.28 & 281.17 & 39.15 & 56.80 \\ 1175.50 & 537.35 & 281.17 & 474.98 & 63.73 & 94.85 \\ 162.68 & 80.17 & 39.15 & 63.73 & 9.95 & 13.88 \\ 238.37 & 117.73 & 56.80 & 94.85 & 13.88 & 21.26 \end{bmatrix}$$

- Obtener la muestra grande de intervalos de confianza simultaneos al 95% para las seis mediciones de las medias de

Obtenga la muestra grande Intervalos de confianza simultáneos al 95% para las seis mediciones medias de población del poblamiento

```
## [1]  67.32099 151.12326  49.08366  82.63692  16.42365  28.85502
```

```
## [1] 123.71901 177.63674  62.29634 104.14308  19.53635  33.40498
```

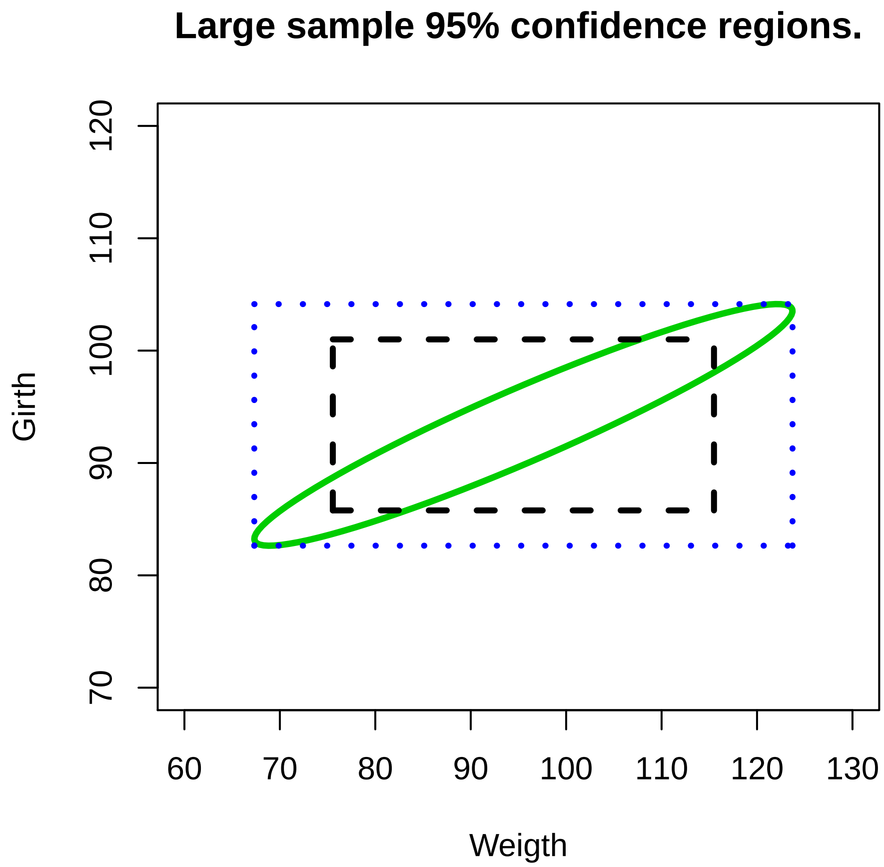
```
##    inferior  superior
## 1  67.32099 123.71901
## 2 151.12326 177.63674
## 3  49.08366  62.29634
## 4  82.63692 104.14308
## 5  16.42365  19.53635
## 6  28.85502  33.40498
```

```
## % latex table generated in R 3.4.2 by xtable 1.8-2 package
## % Fri Nov 10 11:17:17 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{rr}
##   \hline
## inferior & superior \\
##   \hline
## 67.32 & 123.72 \\
##   151.12 & 177.64 \\
##   49.08 & 62.30 \\
##   82.64 & 104.14 \\
##   16.42 & 19.54 \\
##   28.86 & 33.40 \\
##    \hline
## \end{tabular}
## \end{table}
```

```
##       [,1]    [,2]
## x1 3266.46 1175.50
## x4 1175.50  474.98
```

```
## [1] 95.52 93.39
```

```
## eigen() decomposition
## $values
## [1] 3695.51873   45.92127
##
## $vectors
##            [,1]       [,2]
## [1,] -0.9393810  0.3428751
## [2,] -0.3428751 -0.9393810
```



PUNTO 5.21

Usando los datos del contenido mineral de los huesos de la Tabla 1.8, construya el intervalo de Bonferroni al 95% para las medias individuales. Tambien encuentre el T^2-Intervalo simultaneo, compare los dos intervalos hallados.

Table 1.8 Mineral Content in Bones

Subject number	Dominant radius	Radius	Dominant humerus	Humerus	Dominant ulna	Ulna
1	1.103	1.052	2.139	2.238	.873	.872
2	.842	.859	1.873	1.741	.590	.744
3	.925	.873	1.887	1.809	.767	.713
4	.857	.744	1.739	1.547	.706	.674
5	.795	.809	1.734	1.715	.549	.654
6	.787	.779	1.509	1.474	.782	.571
7	.933	.880	1.695	1.656	.737	.803
8	.799	.851	1.740	1.777	.618	.682
9	.945	.876	1.811	1.759	.853	.777
10	.921	.906	1.954	2.009	.823	.765
11	.792	.825	1.624	1.657	.686	.668
12	.815	.751	2.204	1.846	.678	.546
13	.755	.724	1.508	1.458	.662	.595
14	.880	.866	1.786	1.811	.810	.819
15	.900	.838	1.902	1.606	.723	.677
16	.764	.757	1.743	1.794	.586	.541
17	.733	.748	1.863	1.869	.672	.752
18	.932	.898	2.028	2.032	.836	.805
19	.856	.786	1.390	1.324	.578	.610
20	.890	.950	2.187	2.087	.758	.718
21	.688	.532	1.650	1.378	.533	.482
22	.940	.850	2.334	2.225	.757	.731
23	.493	.616	1.037	1.268	.546	.615
24	.835	.752	1.509	1.422	.618	.664
25	.915	.936	1.971	1.869	.869	.868

Source: Data courtesy of Everett Smith.

Definimos entonces las variables de las siguiente manera:

X1: Dominant Radius
X2: Radius
X3: Dominant Humerus
X4: Humerus
X5: Dominant ulna
X6: Ulna

Para hallar los intervalos de Bonferroni y los T^2 se calcula $\underline{\overline{X}}$ y S, para Bonferroni los intervalos estan dados por:

$$\bar{x}_i \pm t_{\frac{\alpha}{2p},\ n-1}\sqrt{\frac{S_{ii}}{n}}$$

Así que para los datos del ejercicio los intervalos de Bonferroni con un nivel de confianza del 95% serian:

```
##         [,1]
## [1,] 0.84380
## [2,] 0.81832
## [3,] 1.79268
## [4,] 1.73484
## [5,] 0.70440
## [6,] 0.69384
```

```
##             X1          X2         X3         X4          X5          X6
## X1 0.013001583 0.010378442 0.02234997 0.02008568 0.009120708 0.007957842
## X2 0.010378442 0.011417893 0.01853519 0.02109951 0.008529783 0.008908512
## X3 0.022349975 0.018535190 0.08035723 0.06677620 0.016836925 0.012847030
## X4 0.020085675 0.021099512 0.06677620 0.06948447 0.017735483 0.016793598
## X5 0.009120708 0.008529783 0.01683692 0.01773548 0.011568417 0.008071150
```

```
## X6 0.007957842 0.008908512 0.01284703 0.01679360 0.008071150 0.010599140
```

```
##         X1         X2         X3         X4         X5         X6
## 0.01300158 0.01141789 0.08035723 0.06948447 0.01156842 0.01059914
```

```
## [1] 2.875094
```

```
##           [,1]      [,2]
## [1,] 0.7782338 0.9093662
## [2,] 0.7568766 0.8797634
## [3,] 1.6296774 1.9556826
## [4,] 1.5832656 1.8864144
## [5,] 0.6425529 0.7662471
## [6,] 0.6346406 0.7530394
```

```
## [1] 4.46317
```

```
##           [,1]      [,2]
## [1,] 0.7420179 0.9455821
## [2,] 0.7229380 0.9137020
## [3,] 1.5396419 2.0457181
## [4,] 1.4995425 1.9701375
## [5,] 0.6083914 0.8004086
## [6,] 0.6019414 0.7857386
```

```
##     N   Media Desv_Estandar    T2_Low     T2_Up Longitud_T2     B_Low
## X1 25 0.84380    0.01300158 0.7420179 0.9455821   0.2035643 0.7782338
## X2 25 0.81832    0.01141789 0.7229380 0.9137020   0.1907640 0.7568766
## X3 25 1.79268    0.08035723 1.5396419 2.0457181   0.5060761 1.6296774
## X4 25 1.73484    0.06948447 1.4995425 1.9701375   0.4705950 1.5832656
## X5 25 0.70440    0.01156842 0.6083914 0.8004086   0.1920173 0.6425529
## X6 25 0.69384    0.01059914 0.6019414 0.7857386   0.1837971 0.6346406
##         B_Up Longitud_B
## X1 0.9093662  0.1311325
## X2 0.8797634  0.1228868
## X3 1.9556826  0.3260052
## X4 1.8864144  0.3031489
## X5 0.7662471  0.1236941
## X6 0.7530394  0.1183988
```

Se observa que los intervalos de Bonferroni siempre son más angostos que los intervalos producidos por el método T2, siendo los de bonferroni más angostos lo que es una característica mejor.

Longitude de los intervalos, practica los boferronio porque son intervalos mas pequeños, longitud de los intervalos

se espera que el valor de las medias individuales con un nivel de confianza tal caigan dentro de los intervalos de confianza generados anteriormente.