

# Parcial de Bioestadística

María Eugenia Fontecha

09 noviembre, 2020

La infección por el virus del dengue resulta ser sistémica y comportarse en forma dinámica, pasando por la aparición de cuadros clínicos asintomáticos o formas sintomáticas con distintos espectros de gravedad. Las guías de atención para pacientes con dengue en la región de las Américas (OPS 2010) recomiendan abundante hidratación para aquellos pacientes que puedan ser tratados en su hogar. Sin embargo, algunos hallazgos clínicos en Argentina sugieren que los pacientes con dengue podrían presentar niveles de sodio plasmático por debajo de los valores normales (hiponatremia), lo que indicaría una estrategia de hidratación inicial distinta a la habitualmente sugerida. Se diseñó un estudio con el objetivo de determinar la prevalencia de hiponatremia al momento de la primera consulta en pacientes atendidos en un hospital general con infección por el virus del dengue y los factores de riesgo asociados a dicho evento. Participaron 150 pacientes con diagnóstico de dengue confirmado por serología o PCR, mayores a 15 años. Fueron excluidos los pacientes con insuficiencia renal crónica estadio V en tratamiento sustitutivo de la función renal. Los pacientes fueron reclutados enero del 2016 y abril del 2016 en el Hospital Posadas. En cada paciente se determinó el sodio, potasio y cloro plasmático (mEq/L), creatinina plasmática (mg/dl), urea (g/L), edad y género. La hiponatremia fue definida como una concentración sérica de sodio  $\leq 135$  mEq/L. En los pacientes con hiperglucemia (glucemia  $>110$  mg/dl) se corrigió el valor de sodio plasmático (sodio corregido = sodio sin corregir + glucemia/36). Los datos se encuentran en el archivo *base.csv*.

```
dengue <- read.xlsx('base.xls', 'final', header=TRUE)
```

## 1. ¿A qué tipo de diseño experimental cree que corresponde el trabajo planteado?

El trabajo planteado corresponde a un estudio de corte transversal, ya que todos los datos de un individuo son tomados en el mismo momento, y no hay un seguimiento de los individuos en el tiempo. Tampoco hay una intervención ni una comparación entre casos y controles. Además, el objetivo de este trabajo es medir la prevalencia de la hiponatremia en casos de infección por el virus del dengue, y los estudios de corte transversal son los que se utilizan para esto.

## 2. Defina la pregunta (PICO) de la investigación.

- **P:** 150 pacientes del Hospital Posadas reclutados entre enero y abril de 2016 con diagnóstico de dengue confirmado por serología o PCR, mayores a 15 años, que no presentaban insuficiencia renal crónica estadio V con tratamiento sustitutivo de la función renal.
- **I:** no hay una intervención, ya que se trata de un estudio de corte transversal, donde solamente se toman medidas.
- **C:** se compara el género, la edad, el potasio plasmático, el cloro plasmático, la creatinina plasmática, la urea y la hiperglucemia.

- O: si se padece de hiponatremia o no.

### 3. ¿Qué tipo de sesgos le parece que puede presentar el diseño experimental?

- **Sesgo de selección:** ya que todos los individuos fueron tomados del mismo hospital.
- **Sesgo de información:**
  - no se tiene en consideración otras posibles causas de la hiponatremia, como la dieta de los individuos, medicamentos, consumo diario de agua, entre otras.
  - no se indica el estadio de la infección de cada paciente ni la gravedad
  - tampoco se indica el serotipo de dengue, ni si se trata de una primera infección o no.

### 4. Identifique las variables y clasifíquelas según rol (respuesta o explicatoria), tipo, unidad y distribución de probabilidades. Plantee qué test utilizaría para realizar la comparación entre los grupos. (Justifique sus respuestas)

```
summary(dengue)
```

```
##      Id      genero      edad      urea
## Min.   : 1.00  Length:150   Min.   :15.00  Min.   :0.0700
## 1st Qu.: 38.25  Class :character 1st Qu.:26.00  1st Qu.:0.1900
## Median : 75.50  Mode  :character  Median :37.00  Median :0.2500
## Mean   : 75.50          Mean   :39.60  Mean   :0.2737
## 3rd Qu.:112.75          3rd Qu.:51.75  3rd Qu.:0.3200
## Max.   :150.00          Max.   :81.00  Max.   :1.0300
##
##      creatinina      sodio      potasio      cloro
## Min.   :0.0900  Min.   :128.0  Min.   :3.000  Min.   : 85.3
## 1st Qu.:0.7000  1st Qu.:135.0  1st Qu.:3.700  1st Qu.: 96.7
## Median :0.8000  Median :137.0  Median :4.000  Median : 98.4
## Mean   :0.9134  Mean   :136.6  Mean   :3.975  Mean   : 98.6
## 3rd Qu.:1.0000  3rd Qu.:138.0  3rd Qu.:4.200  3rd Qu.:100.5
## Max.   :7.3000  Max.   :144.0  Max.   :5.500  Max.   :109.3
## NA's   :1
## hiperglucemia      NA.
## Min.   :197.0  Mode:logical
## 1st Qu.:200.0  NA's:150
## Median :208.0
## Mean   :237.8
## 3rd Qu.:288.0
## Max.   :296.0
## NA's   :145
```

Acondicionamiento de los datos.

```
dengue$genero <- as.factor(dengue$genero)
dengue$hiponatremia[dengue$sodio <= 135] <- 1
dengue$hiponatremia[dengue$sodio > 135] <- 0
dengue$hiponatremia <- as.factor(dengue$hiponatremia)
dengue$hiperglucemia[is.na(dengue$hiperglucemia)] <- 0
dengue$hiperglucemia[dengue$hiperglucemia > 0] <- 1
dengue$hiperglucemia <- as.factor(dengue$hiperglucemia)
```

Separo los datos en dos grupos:

- Grupo 1: hiponatremia
- Grupo 0: normonatremia

```
dengue1 <- dengue[dengue$hiponatremia == 1,]  
dengue0 <- dengue[dengue$hiponatremia == 0,]
```

## Variables cuantitativas

Evalúo normalidad de las variables numéricas.

- Edad
- – Grupo 1

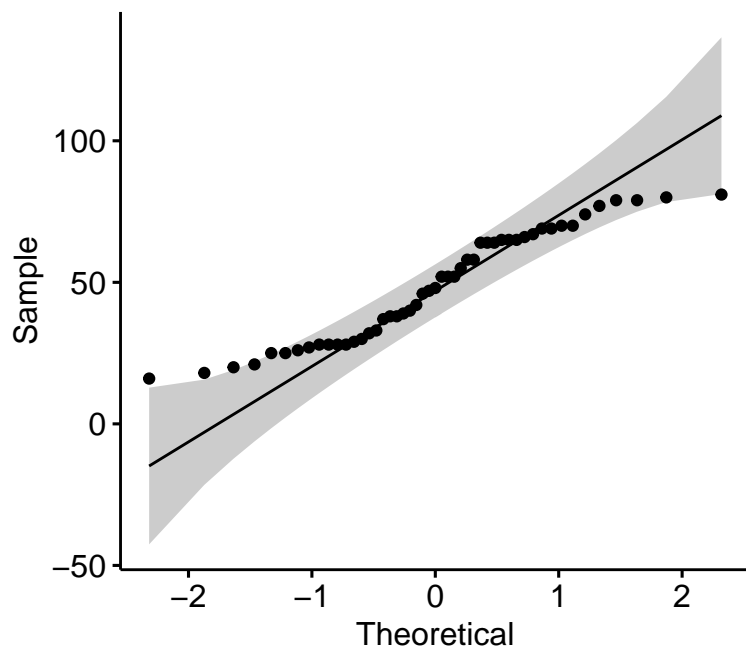
```
x = dengue1$edad  
print(shapiro.test(x))
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  x  
## W = 0.93369, p-value = 0.008452
```

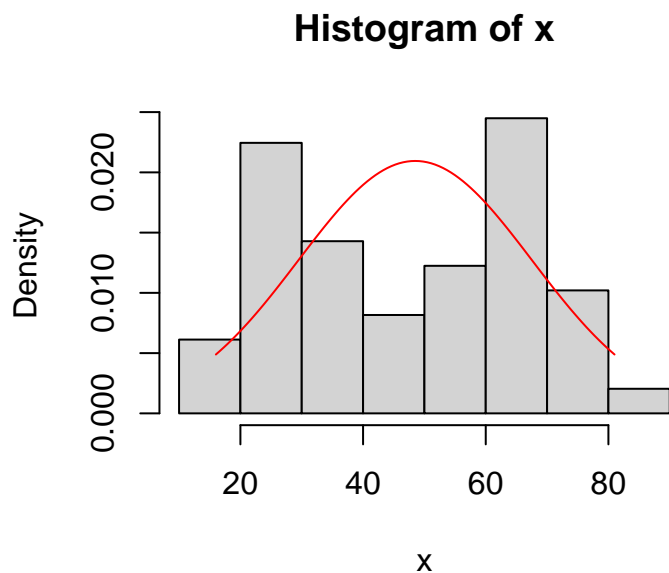
```
print(lillie.test(x))
```

```
##  
##  Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data:  x  
## D = 0.14898, p-value = 0.008261
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```



En el histograma se puede ver que la variable Edad para el grupo 1 no presenta una distribución normal, ya que no tiene la forma acampanada característica. También podemos ver en el QQ-Plot que hay algunos datos que se apartan de la línea de tendencia esperada para una distribución normal, especialmente en los extremos. Además, ambos tests estadísticos arrojaron un p-valor menor a 0.05, por lo que la hipótesis nula de que la variable tiene distribución normal se rechaza. Por ende, la edad no tiene distribución normal para el grupo 1.

- – Grupo 0

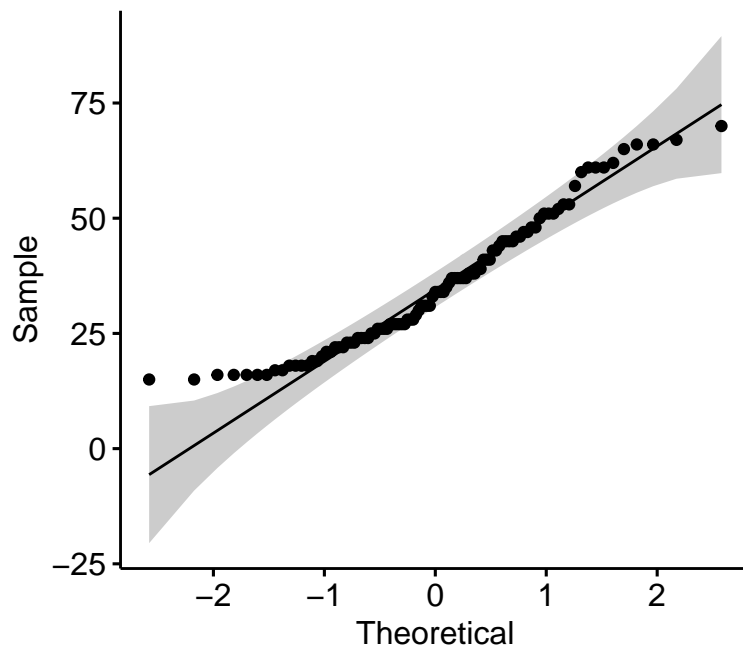
```
x = dengue0$edad
print(shapiro.test(x))
```

```
##
## Shapiro-Wilk normality test
##
## data:  x
## W = 0.94376, p-value = 0.0003056
```

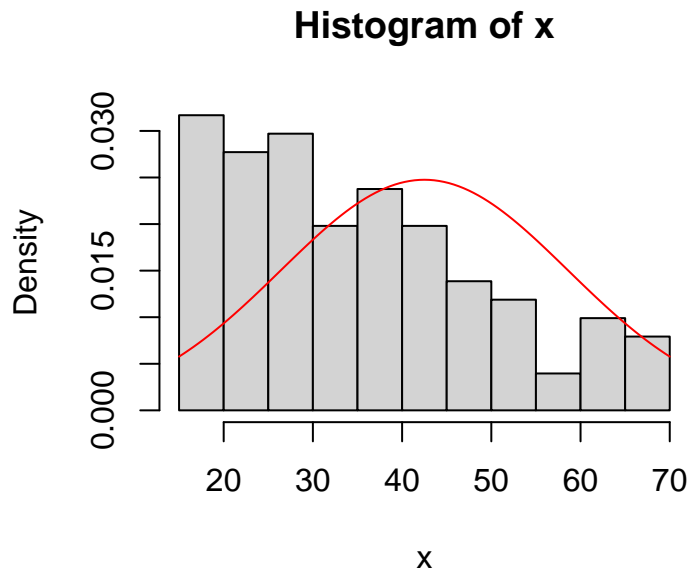
```
print(lillie.test(x))
```

```
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.11763, p-value = 0.00152
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```



Para el grupo 0 podemos ver que el histograma tampoco tiene una distribución que se asemeje a la de la normal. El QQ-Plot muestra algunos datos que no siguen la tendencia esperada para una distribución normal. Por otro lado, en los tests estadísticos de Lilliefors y Shapiro-Wilk se rechaza la hipótesis nula de que los datos tienen distribución normal, ya que el p-valor en ambos casos fue menor a 0.05. En conclusión, la edad tampoco tiene distribución normal para el grupo 0.

- Potasio
- – Grupo 1

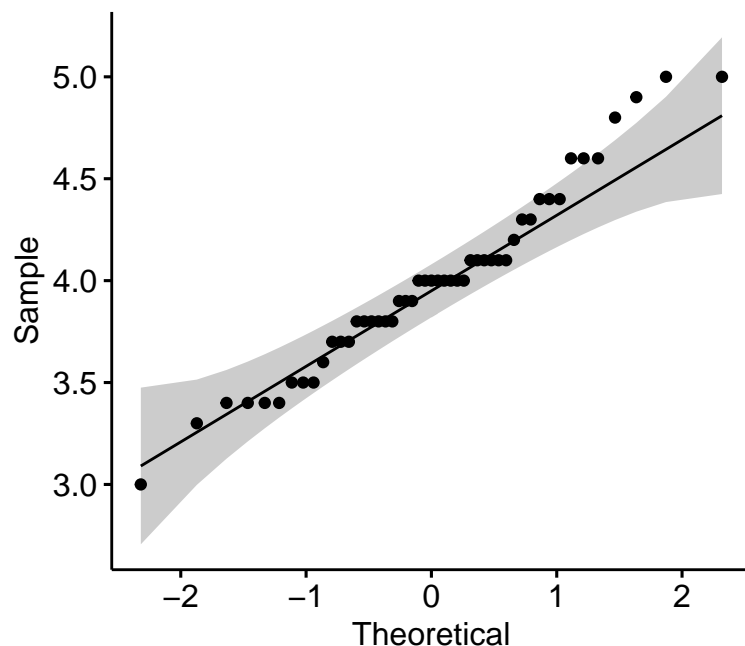
```
x = dengue1$potasio
print(shapiro.test(x))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.96943, p-value = 0.2299
```

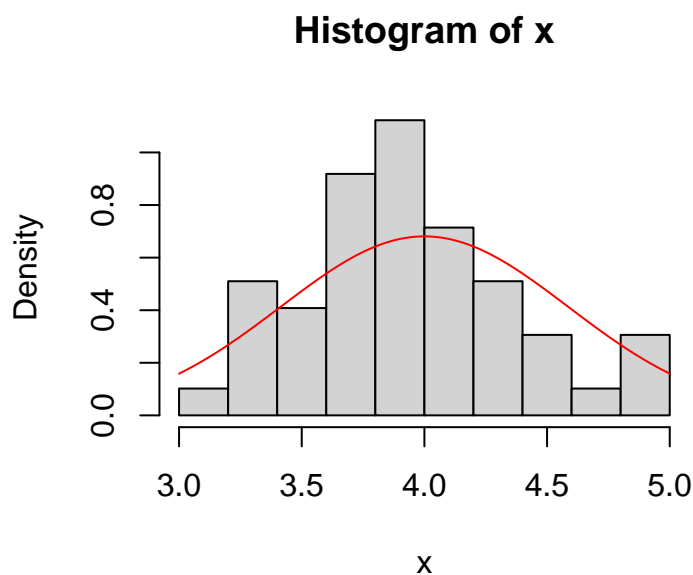
```
print(lillie.test(x))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.14123, p-value = 0.01582
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE, breaks = 12)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```



Los métodos gráficos no muestran normalidad para la variable potasio del grupo 1; por un lado en el QQ-Plot se puede ver que algunos datos están escalonados y que otros se alejan de la recta de tendencia teórica, y por el otro el histograma presenta una forma que no se asimila tanto a la distribución normal. El test de Lilliefors dio un p-valor menor a 0.05, por lo que se rechaza la hipótesis nula, mientras que el test de Shapiro-Wilk dio un p-valor mayor a 0.05, por lo que la hipótesis nula no se rechaza. Como de los 4 métodos, 3 no indicarían normalidad, concluyo que la variable potasio no es normal para el grupo 1.

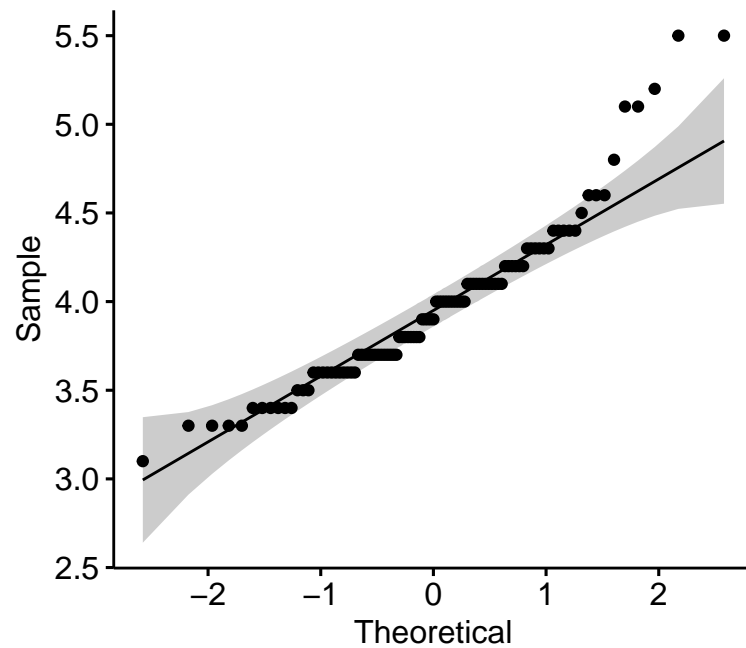
```
x = dengue0$potasio
print(shapiro.test(x))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.92697, p-value = 3.083e-05
```

```
print(lillie.test(x))
```

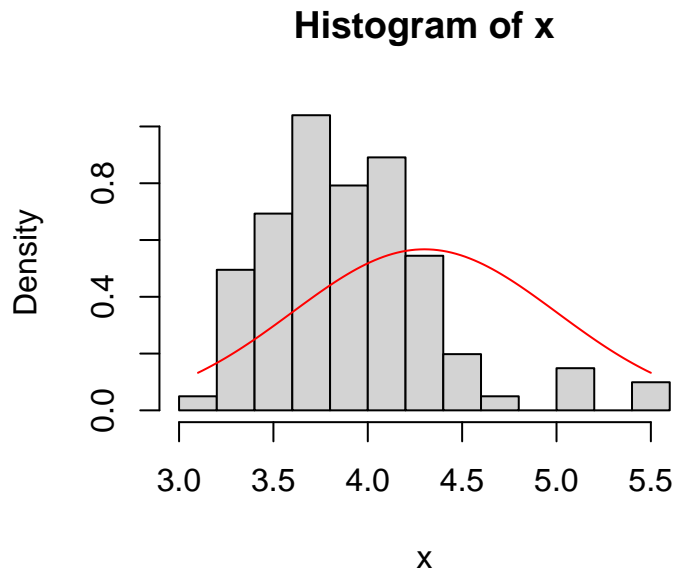
```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.11682, p-value = 0.001705
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE, breaks = 10)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```





Para el grupo 0, el QQ-Plot dio similar al grupo 1, en el sentido de que se ve cierto escalonamiento y que hay datos que se apartan de la recta de normalidad. Además, el histograma tampoco muestra una distribución normal. Por otro lado, a partir de los dos test estadísticos también se llega a la conclusión de que el potasio no tiene distribución normal para este grupo, ya que los p-valores dieron menores a 0.05. Por lo tanto, el potasio no tiene distribución normal.

- Cloro
- – Grupo 1

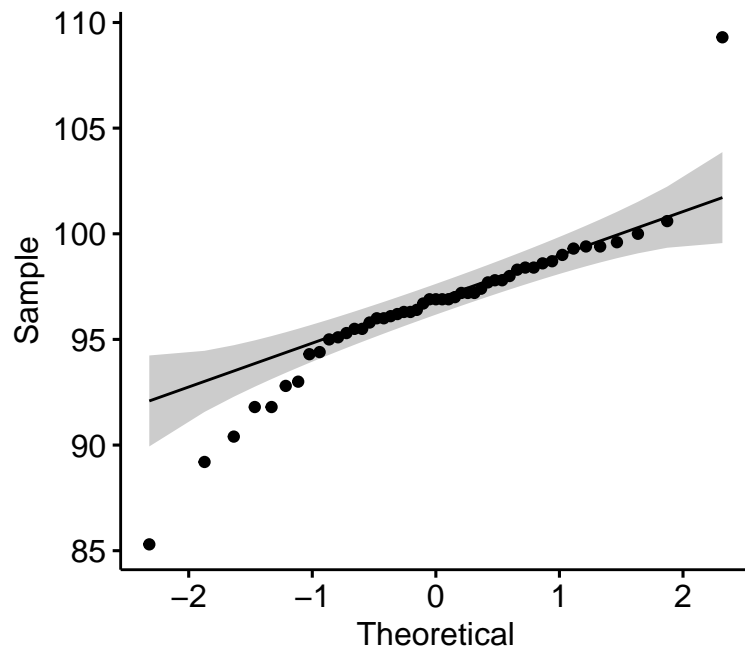
```
x = dengue1$cloro
print(shapiro.test(x))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.88334, p-value = 0.000164
```

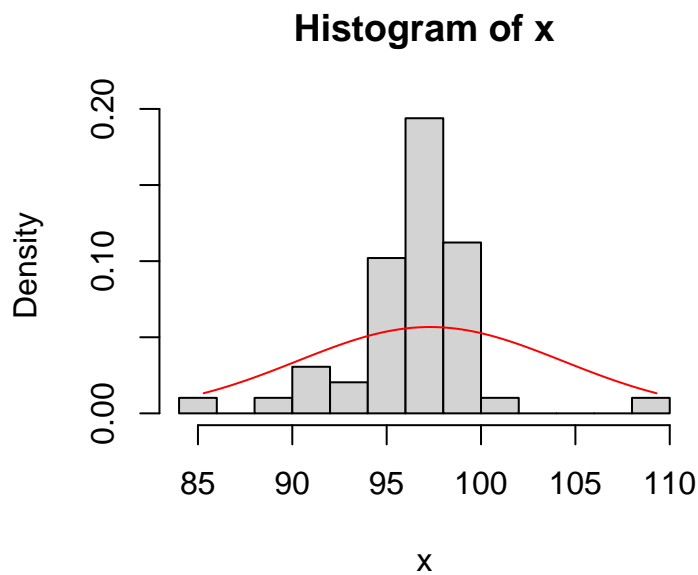
```
print(lillie.test(x))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.14639, p-value = 0.01031
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE, breaks = 9)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```



Tanto el histograma como el QQ-Plot muestran que en los datos centrales hay cierta tendencia normal, mientras que en los extremos se observa un desvío de esta tendencia. Además, ambos tests estadísticos dieron un p-valor menor a 0.05, por lo que se rechaza la hipótesis nula. Entonces, la variable cloro para el grupo de hiponatremia no tiene distribución normal.

- – Grupo 0

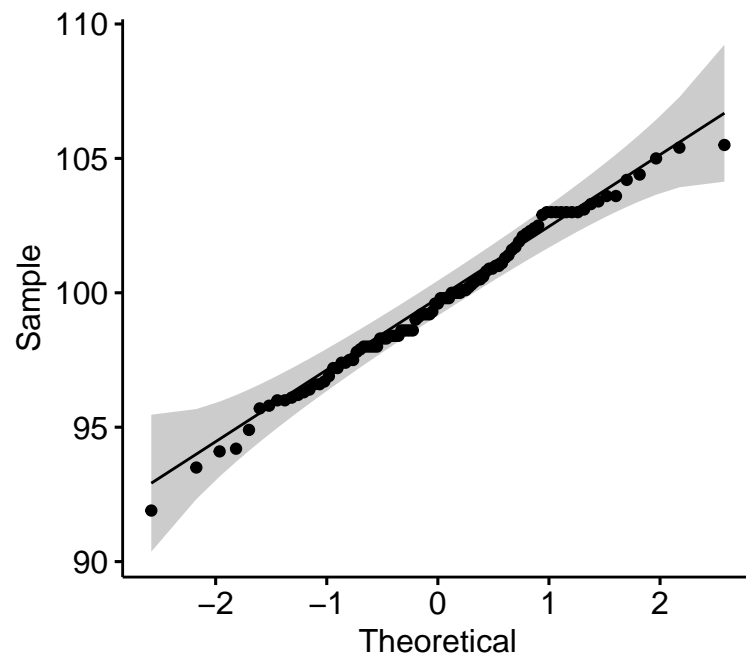
```
x = dengue0$cloro
print(shapiro.test(x))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.99018, p-value = 0.6726
```

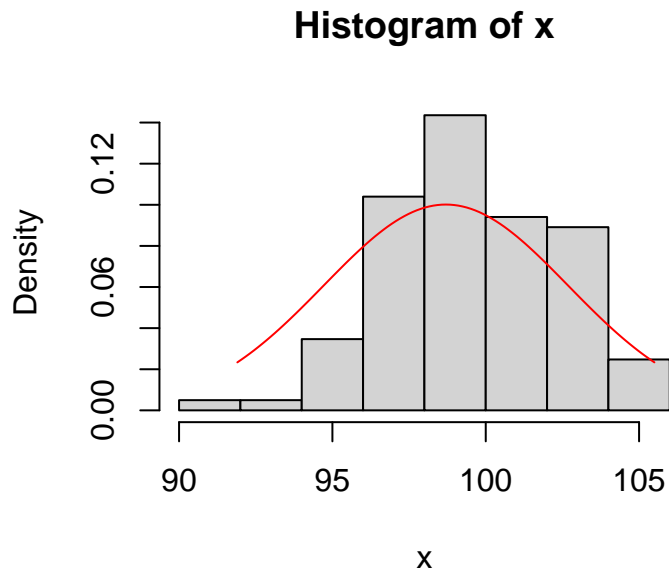
```
print(lillie.test(x))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.061967, p-value = 0.4441
```

```
ggqqplot(x)
```



```
hist(x, freq = FALSE, breaks = 9)
curve( dnorm(x, mean(x), sd(x)), min(x), max(x), add= T, col = 'red')
```



Tanto el test de Shapiro-Wilk como el de Lilliefors dieron un p-valor mayor a 0.05, por lo que no se rechaza la hipótesis nula de distribución normal de la variable. A su vez, el QQ-Plot también muestra esto, ya que todos los datos caen cerca de la línea de tendencia normal. El histograma no muestra una distribución tan concluyente como los demás métodos, ya que si bien se observa cierta forma acampanada, esta no se parece tanto a la de la distribución normal. De todas formas, concluyo que para el grupo 0 el cloro tiene distribución normal debido a los resultados de los demás métodos.

- Creatinina
- – Grupo 1

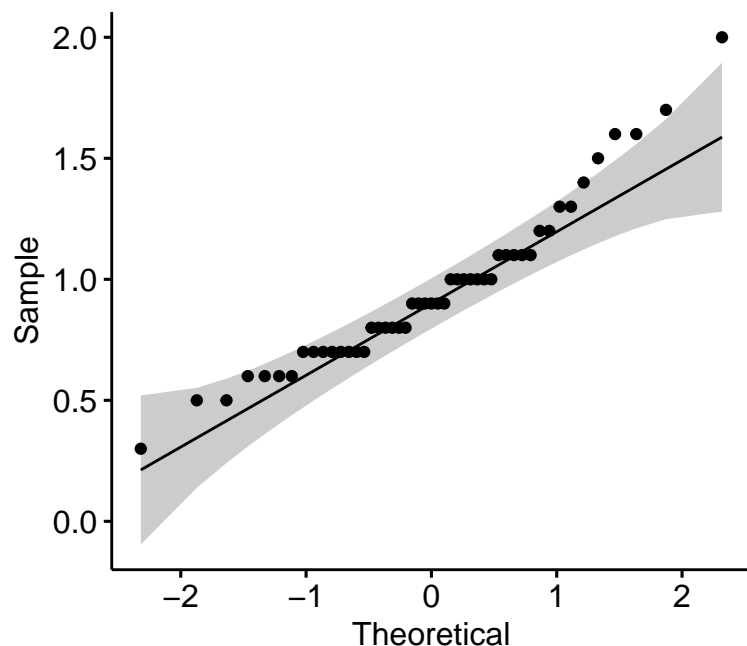
```
x = dengue1$creatinina
print(shapiro.test(x))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.93979, p-value = 0.01451
```

```
print(lillie.test(x))
```

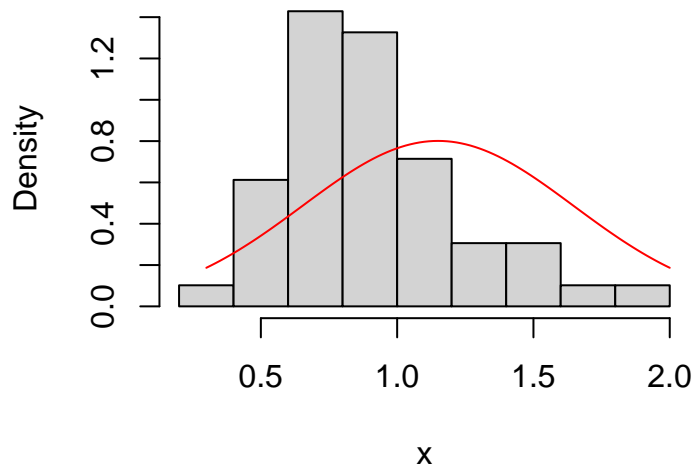
```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.14103, p-value = 0.01608
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```

**Histogram of x**



El histograma para la variable creatinina del grupo 1 muestra una forma acampanada pero que no es simétrica, ya que la cola derecha se ve que es más larga que la izquierda, lo cual se ve como un desplazamiento hacia la izquierda. El otro método gráfico, el Q-Q-Plot, muestra que hay algunos datos que se alejan de la recta de normalidad, y hay otros que tienen una distribución escalonada. Además, ambos tests estadísticos dieron un p-valor menor a 0.05, por lo que se rechaza la hipótesis nula. Entonces, la creatinina para el grupo 1 no tiene distribución normal.

- – Grupo 0

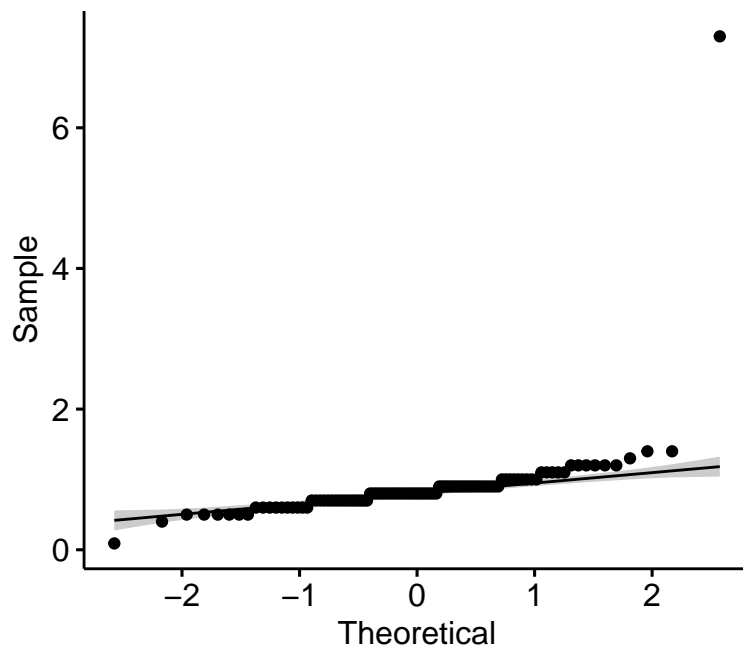
```
x = na.omit(dengue0$creatinina)
print(shapiro.test(x))
```

```
##
## Shapiro-Wilk normality test
##
## data: x
## W = 0.31139, p-value < 2.2e-16
```

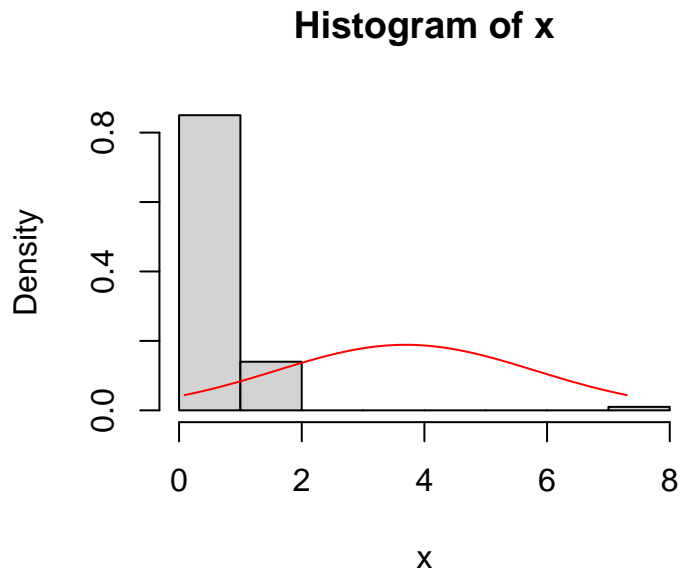
```
print(lillie.test(x))
```

```
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data: x
## D = 0.28763, p-value < 2.2e-16
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```



La creatinina para el grupo 0 no tiene distribución normal. Tanto el QQ-Plot como el histograma muestran que los datos no tienen una distribución normal ya que no presentan el comportamiento esperado. Además, los dos métodos estadísticos dieron un p-valor mucho menor a 0.05, por lo que se rechaza la hipótesis nula. En este grupo se puede observar un outlier, ya que hay un dato que difiere mucho de los demás, y además hay un dato NA.

- Urea
- – Grupo 1

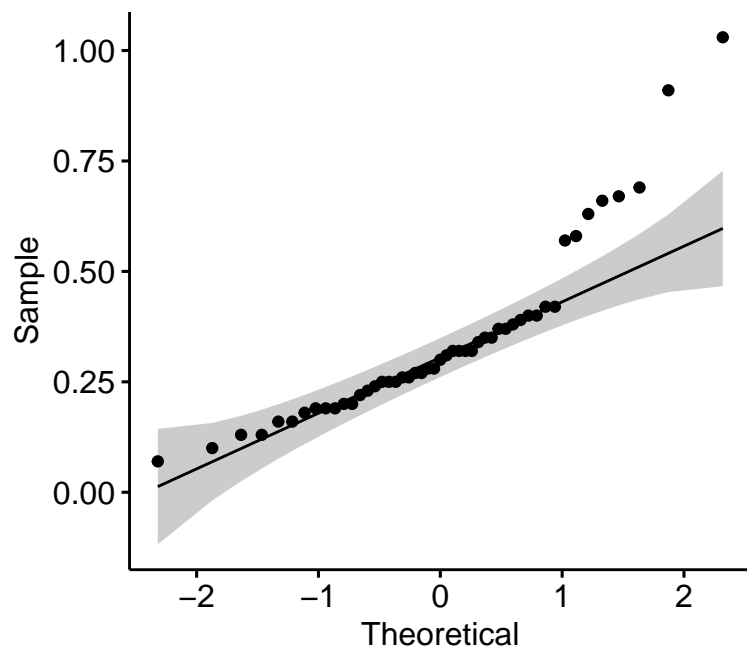
```
x = dengue1$urea
print(shapiro.test(x))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.85421, p-value = 2.381e-05
```

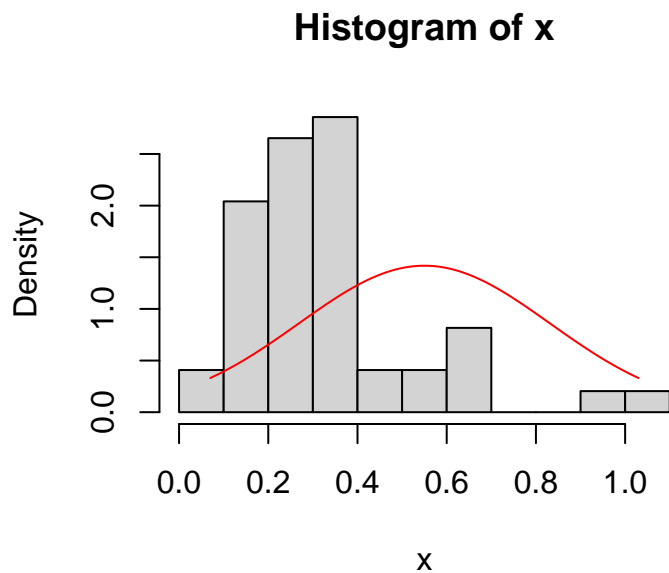
```
print(lillie.test(x))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.18376, p-value = 0.0002615
```

```
ggqqplot(x)
```



```
hist(x, freq = FALSE)
curve( dnorm(x, mean(x), sd(x)), min(x), max(x), add= T, col = 'red')
```



En el histograma de la urea del grupo 1 no se observa la campana de Gauss y en el QQ-Plot hay varios puntos que se alejan mucho de la línea de tendencia de normalidad, en particular para los valores más altos. Por otro lado ambos p-valores de los tests estadísticos dieron mucho más chicos que 0.05, por lo que se rechaza la hipótesis nula de normalidad de la variable. Por lo tanto, la urea no tiene distribución normal para el grupo 1.

- – Grupo 0



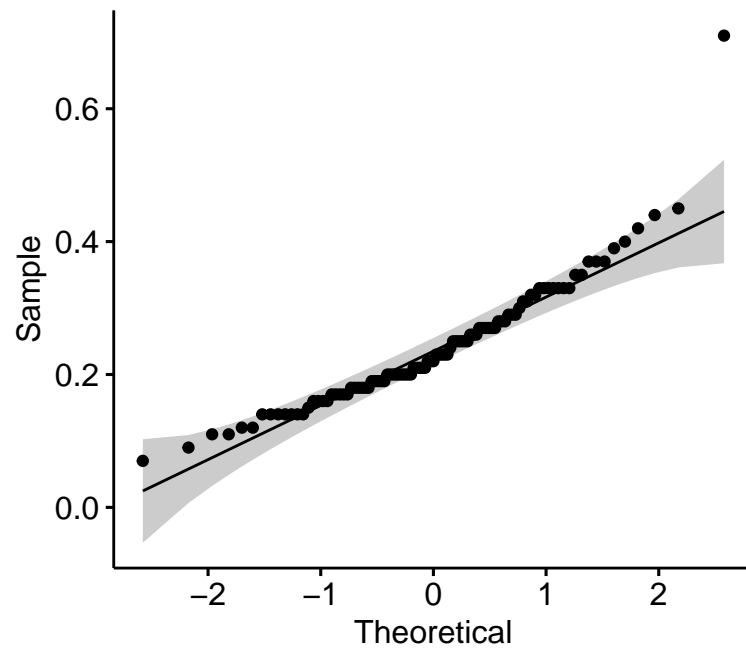
```
x = dengue0$urea
print(shapiro.test(x))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.90825, p-value = 3.2e-06
```

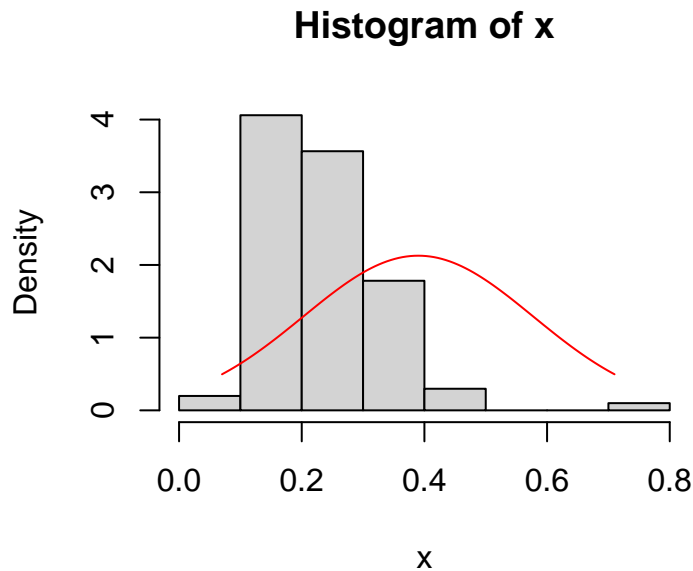
```
print(lillie.test(x))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  x
## D = 0.10332, p-value = 0.009811
```

```
ggqqplot(x)
```



```
hist(x, freq =FALSE)
curve( dnorm(x,mean(x),sd(x)), min(x), max(x), add= T, col = 'red')
```



Para el grupo 0, el QQ-Plot de la urea muestra que la mayoría de los datos siguen la línea de tendencia, aunque hay algunos que se alejan. El histograma no muestra la forma acampanada característica de la distribución normal. Por último, tanto el test de Lilliefors como el de Shapiro-Wilk dieron un p-valor menor a 0.05, rechazando entonces la hipótesis nula. Por ende, la urea tampoco tiene distribución normal para el grupo 0.

### Variables dicotómicas

Las variables dicotómicas tienen distribución binomial.

### Tabla de resumen de las variables

```
tabla <- data.frame(Variables = c('Hiponatremia','Edad', 'Género', 'Potasio','Cloro','Urea',
                                'Creatinina','Hiperglucemia','Edad', 'Género',
                                'Potasio','Cloro','Urea','Creatinina','Hiperglucemia'),

                    Rol = c('Respuesta','Explicativa','Explicativa','Explicativa',
                            'Explicativa',
                            'Explicativa','Explicativa','Explicativa','Explicativa',
                            'Explicativa','Explicativa','Explicativa',
                            'Explicativa','Explicativa','Explicativa'),

                    Tipo = c('Categórica Dicotómica','Cuantitativa Continua',
                            'Categórica Dicotómica','Cuantitativa Continua',
                            'Cuantitativa Continua','Cuantitativa Continua',
                            'Cuantitativa Continua','Categórica Dicotómica',
                            'Cuantitativa Continua','Categórica Dicotómica',
                            'Cuantitativa Continua',
                            'Cuantitativa Continua','Cuantitativa Continua',
                            'Cuantitativa Continua', 'Categórica Dicotómica'),

                    Unidad = c('-', 'Años', '-', 'mEq/L', 'mEq/L', 'g/L', 'mg/dl', '-',
```

```

'Años', '-', 'mEq/L', 'mEq/L', 'g/L', 'mg/dl', '-'),
Distribución = c('Binomial', 'No normal', 'Binomial', 'No normal',
'No Normal', 'No normal', 'No normal', 'Binomial',
'No normal', 'Binomial', 'No normal', 'Normal',
'No normal', 'No normal', 'Binomial'))

kable(tabla, booktabs = T) %>%
  kable_styling(latex_options = 'hold_position') %>%
  group_rows("Grupo 1", 2, 8) %>%
  group_rows("Grupo 0", 9, 15)

```

Variables	Rol	Tipo	Unidad	Distribución
Hiponatremia	Respuesta	Categórica Dicotómica	-	Binomial
<b>Grupo 1</b>				
Edad	Explicativa	Cuantitativa Continua	Años	No normal
Género	Explicativa	Categórica Dicotómica	-	Binomial
Potasio	Explicativa	Cuantitativa Continua	mEq/L	No normal
Cloro	Explicativa	Cuantitativa Continua	mEq/L	No Normal
Urea	Explicativa	Cuantitativa Continua	g/L	No normal
Creatinina	Explicativa	Cuantitativa Continua	mg/dl	No normal
Hiperglucemia	Explicativa	Categórica Dicotómica	-	Binomial
<b>Grupo 0</b>				
Edad	Explicativa	Cuantitativa Continua	Años	No normal
Género	Explicativa	Categórica Dicotómica	-	Binomial
Potasio	Explicativa	Cuantitativa Continua	mEq/L	No normal
Cloro	Explicativa	Cuantitativa Continua	mEq/L	Normal
Urea	Explicativa	Cuantitativa Continua	g/L	No normal
Creatinina	Explicativa	Cuantitativa Continua	mg/dl	No normal
Hiperglucemia	Explicativa	Categórica Dicotómica	-	Binomial

## Tests para realizar la comparación entre los grupos

A partir de los resultados anteriores, puedo definir qué test utilizaría para realizar la comparación entre los grupos.

- Para las variables dicotómicas se debe utilizar el test de Chi Cuadrado si más del 80% de las celdas de la tabla de contingencia tienen una frecuencia esperada mayor o igual a 5, y ninguna menor a 1, y el test exacto de Fisher cuando no se cumple esto.
- Para las variables cuantitativas se debe utilizar el test Z si estas tienen distribución normal y si se tiene una cantidad de individuos mayor a 30, el test T si las variables tienen distribución normal y la muestra es menor a 30, o el test de Wilcoxon-Mann-Whitney si no se cumple normalidad.

Genero la tabla de contingencia para la variable **genero**.

```

t_genero <- table(dengue$genero, dengue$hiponatremia)
t_genero = t_genero[rev(order(rownames(t_genero))) ,
                    rev(order(colnames(t_genero)))]

kable(t_genero, booktabs = T) %>%
kable_styling(latex_options = "striped")

```

	1	0
m	25	51
f	24	50

Como todas las frecuencias son mayores a 5 puedo utilizar el test de Chi cuadrado.

Genero la tabla de contingencia para la variable hiperglucemia.

```
t_hglucemia <- table(dengue$hiperglucemia, dengue$hiponatremia)
t_hglucemia = t_hglucemia[rev(order(rownames(t_hglucemia))) ,
                           rev(order(colnames(t_hglucemia)))]
kable(t_hglucemia, booktabs = T) %>%
kable_styling(latex_options = "striped")
```

	1	0
1	3	2
0	46	99

Para la variable hiperglucemia no se cumple que el 80% de las celdas tengan frecuencia mayor a 5, por lo que tengo que realizar el test exacto de Fisher.

En cuanto a las variables cuantitativas, edad, potasio, urea y creatinina tienen distribución no normal, por lo que se debe utilizar el test de Wilcoxon. La variable cloro dio no normal para un grupo y normal para el otro, por lo que también utilice el test de Wilcoxon.

En resumen:

```
tabla_test <- data.frame(Variables = c('Edad', 'Género', 'Potasio', 'Cloro', 'Urea',
                                       'Creatinina', 'Hiperglucemia'),
                        Test = c('Wilcoxon', 'Chi cuadrado', 'Wilcoxon', 'Wilcoxon', 'Wilcoxon',
                                'Wilcoxon', 'Fisher exacto'))
kable(tabla_test, booktabs = T) %>%
kable_styling(latex_options = "striped")
```

Variables	Test
Edad	Wilcoxon
Género	Chi cuadrado
Potasio	Wilcoxon
Cloro	Wilcoxon
Urea	Wilcoxon
Creatinina	Wilcoxon
Hiperglucemia	Fisher exacto

5. Analice en forma univariada las diferencias entre ambos grupos (Hiponatremia, normonatremia) con respecto a sexo, edad, urea, creatinina, cloro, potasio e hiperglucemia.

## Sexo

### Test de Chi Cuadrado

Las hipótesis del test de Chi Cuadrado son:

$H_0$ : el sexo y la hiponatremia son variables independientes

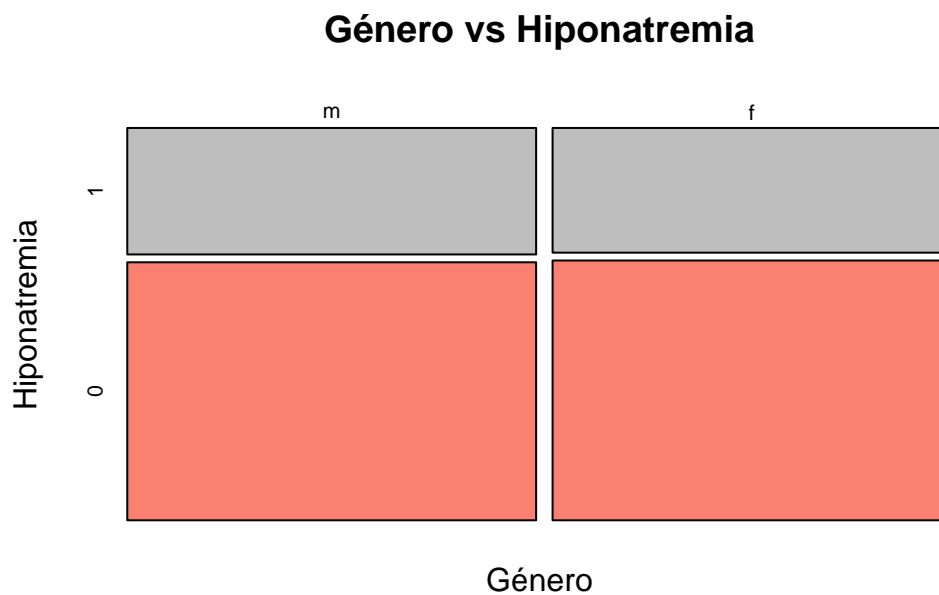
$H_1$ : el sexo y la hiponatremia no son variables independientes.

```
chisq.test(t_genero)
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  t_genero  
## X-squared = 0, df = 1, p-value = 1
```

El p-valor del test de Chi cuadrado dio mayor a 0.05, por lo que no se rechaza la hipótesis nula. Entonces, el sexo y la hiponatremia en los casos de infección por dengue son variables independientes y no hay una asociación estadísticamente significativa.

```
mosaicplot(t_genero, main = 'Género vs Hiponatremia', xlab = 'Género',  
            , ylab = 'Hiponatremia', color = c('grey', 'salmon'))
```



En el gráfico de mosaicos podemos ver que las proporciones de individuos con hiponatremia son muy similares entre el grupo de hombres y el grupo de mujeres.

## Edad

### Test de Wilcoxon

Las hipótesis del test de Wilcoxon son:

- $H_0: \theta_1 - \theta_0 = 0$
- $H_1: \theta_1 - \theta_0 \neq 0,$

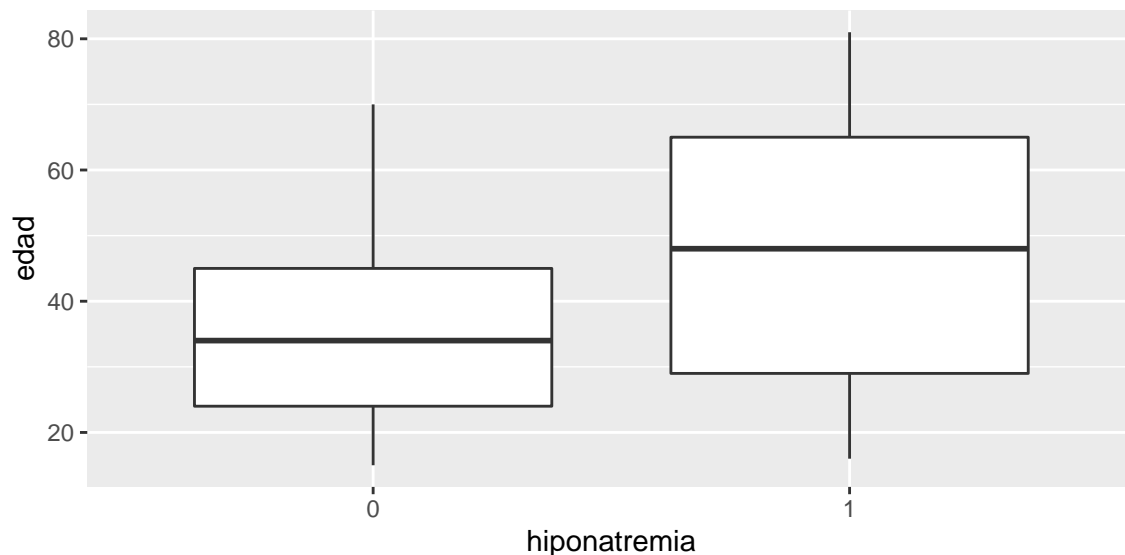
donde  $\theta_1$  y  $\theta_0$  son las medianas de la edad del grupo 1 (hiponatremia) y del grupo 0 (normonatremia), respectivamente.

```
wilcox.test(dengue1$edad,dengue0$edad, paired = FALSE, conf.int = TRUE)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: dengue1$edad and dengue0$edad
## W = 3473, p-value = 6.318e-05
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  6.000046 20.000078
## sample estimates:
## difference in location
##                               13
```

Con un p-valor menor a 0.05 se rechaza la hipótesis nula, por lo que hay una diferencia estadísticamente significativa entre la edad del grupo que no presentó hiponatremia y del que sí presentó hiponatremia. La diferencia entre las edades de los dos grupos es de 13 años, siendo mayores las del grupo 1, con un intervalo de confianza del 95% entre 6 y 20 años.

```
bplt <- ggplot(dengue, aes(x=hiponatremia ,y=edad)) +
  geom_boxplot()
bplt
```



Mediante el boxplot podemos ver que efectivamente la edades de los individuos del grupo de hiponatremia son mayores a las de grupo que no presentó hiponatremia, ya que el boxplot del grupo 1 se encuentra más arriba que el del grupo 0.

## Urea

### Test de Wilcoxon

Las hipótesis del test de Wilcoxon son:

- $H_0: \theta_1 - \theta_0 = 0$
- $H_1: \theta_1 - \theta_0 \neq 0,$

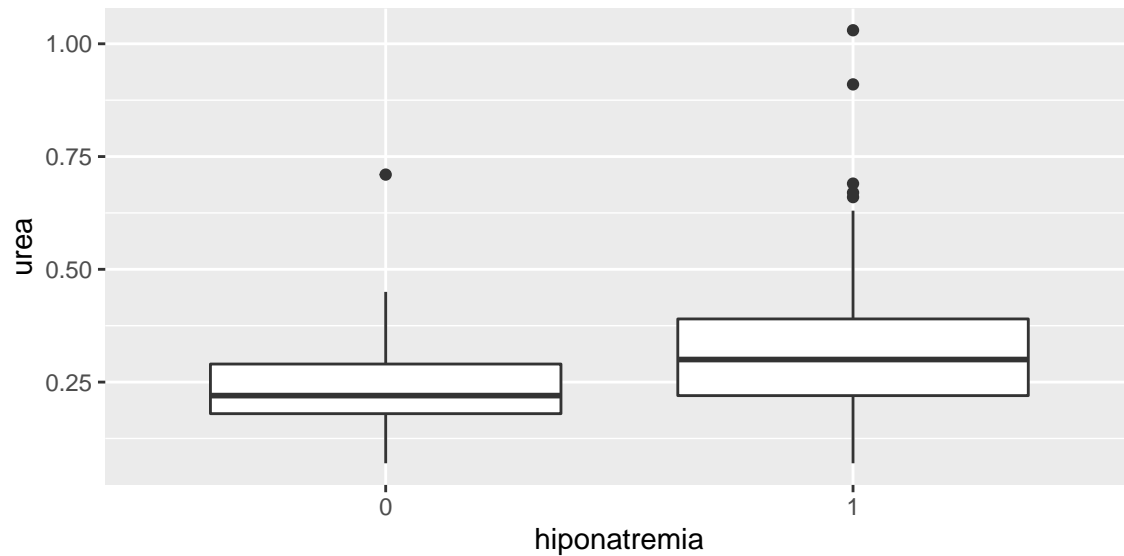
donde  $\theta_1$  y  $\theta_0$  son las medianas de la urea del grupo 1 (hiponatremia) y del grupo 0 (normonatremia), respectivamente.

```
wilcox.test(dengue1$urea,dengue0$urea, paired = FALSE, conf.int = TRUE)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: dengue1$urea and dengue0$urea
## W = 3323.5, p-value = 0.0006667
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  0.02993336 0.10995324
## sample estimates:
## difference in location
##                0.06997629
```

El p-valor dio menor a 0.05, por lo que se rechaza la hipótesis nula del test de Wilcoxon: las medianas de la urea para el grupo con hiponatremia y el grupo con normonatremia son significativamente diferentes. La urea del grupo 0 es 0.06997629 (g/L) menor que la del grupo 1, con un intervalo de confianza del 95% que va de 0.02993336 (g/L) a 0.10995324 (g/L).

```
bplt <- ggplot(dengue, aes(x=hiponatremia ,y=urea)) +
  geom_boxplot()
bplt
```



Con el boxplot podemos ver que los valores de urea del grupo 0 son menores a los del grupo 1.

## Creatinina

### Test de Wilcoxon

Las hipótesis del test de Wilcoxon son:

- $H_0: \theta_1 - \theta_0 = 0$
- $H_1: \theta_1 - \theta_0 \neq 0$ ,

donde  $\theta_1$  y  $\theta_0$  son las medianas de la creatinina plasmática del grupo 1 (hiponatremia) y del grupo 0 (normonatremia), respectivamente.

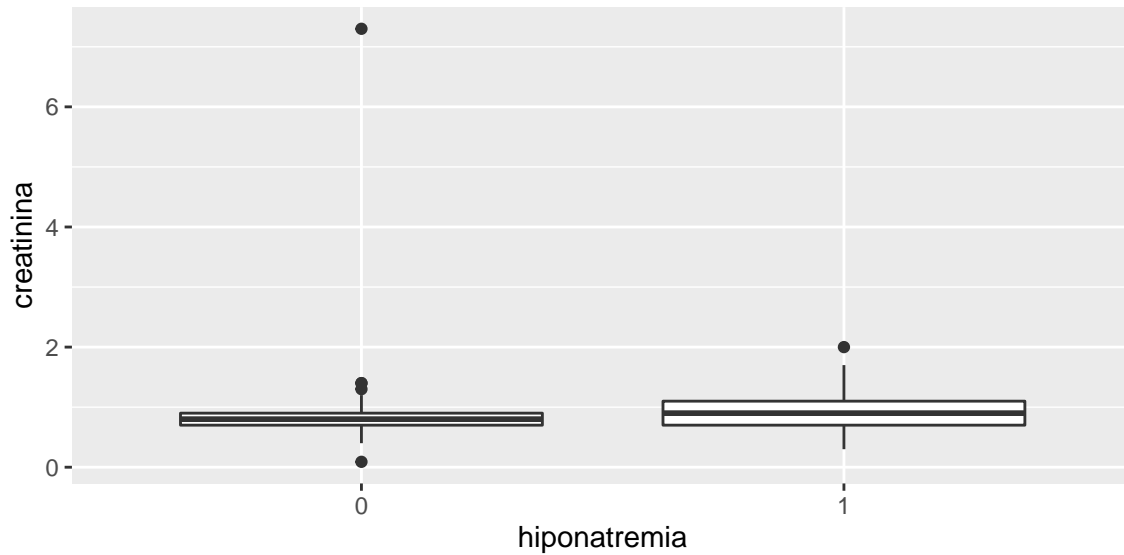
```
wilcox.test(dengue1$creatinina,dengue0$creatinina, paired = FALSE, conf.int = TRUE)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: dengue1$creatinina and dengue0$creatinina
## W = 2916, p-value = 0.05757
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -0.0000260809 0.1999763509
## sample estimates:
## difference in location
## 0.09996035
```

La creatinina no es significativamente diferente entre ambos grupos, ya que el test de Wilcoxon dio un p-valor mayor a 0.05, por lo que no se rechaza la hipótesis nula de que las medianas son iguales. Por lo tanto, no hay una asociación estadísticamente significativa entre el nivel de creatinina y la hiponatremia en individuos con el virus del dengue.



```
bplt <- ggplot(dengue, aes(x=hiponatremia ,y=creatinina)) +
  geom_boxplot()
bplt
```



Si bien el grupo 0 presenta un valor que se aleja mucho de todos los demás, se puede ver que los grupos tienen valores de creatinina similares.

## Cloro

### Test de Wilcoxon

Las hipótesis del test de Wilcoxon son:

- $H_0: \theta_1 - \theta_0 = 0$
- $H_1: \theta_1 - \theta_0 \neq 0$ ,

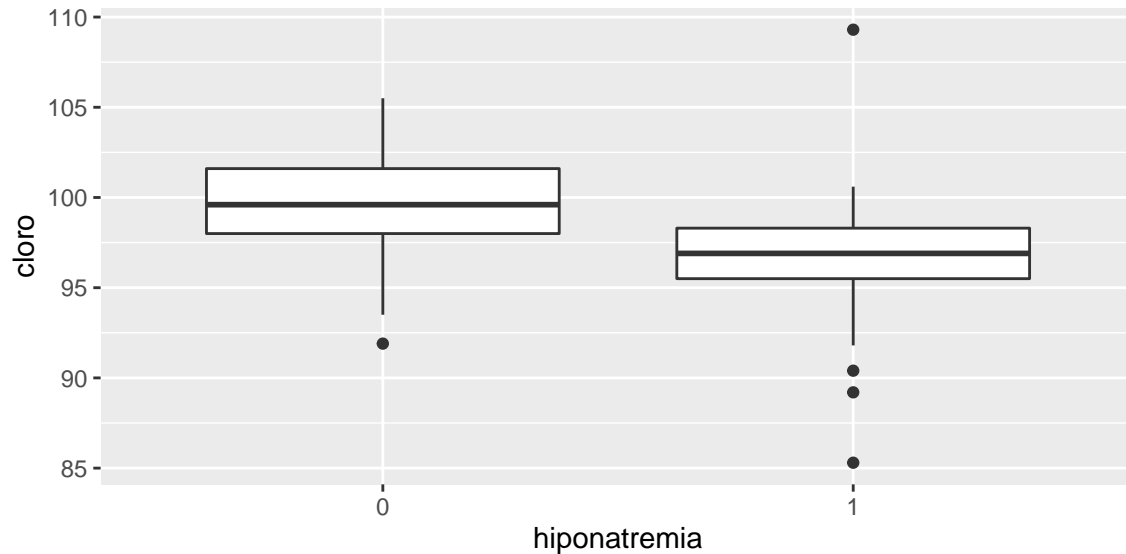
donde  $\theta_1$  y  $\theta_0$  son las medianas del cloro plasmático del grupo 1 (hiponatremia) y del grupo 0 (normonatremia), respectivamente.

```
wilcox.test(dengue1$cloro,dengue0$cloro, paired = FALSE, conf.int = TRUE)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: dengue1$cloro and dengue0$cloro
## W = 1029.5, p-value = 7.041e-09
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
## -3.900039 -2.000046
## sample estimates:
## difference in location
## -2.900066
```

El test de Wilcoxon dio un p-valor menor a 0.05, por lo que se rechaza la hipótesis nula. Entonces, hay una diferencia significativa entre la mediana del cloro plasmático del grupo sin hiponatremia y la mediana del grupo con hiponatremia. Por lo tanto, existe una asociación estadísticamente significativa entre el cloro plasmático y la hiponatremia. En el grupo con normonatremia el nivel de cloro plasmático es 2.900066 (mEq/L) mayor que el nivel del cloro del grupo con hiponatremia, con un intervalo de confianza al 95% entre 2.000046 (mEq/L) y 3.900039 (mEq/L).

```
bplt <- ggplot(dengue, aes(x=hiponatremia ,y=cloro)) +  
  geom_boxplot()  
bplt
```



En el boxplot se puede ver que el grupo 0 tiene niveles de cloro plasmático mayores que los del grupo con hiponatremia.

## Potasio

### Test de Wilcoxon

Las hipótesis del test de Wilcoxon son:

- $H_0: \theta_1 - \theta_0 = 0$
- $H_1: \theta_1 - \theta_0 \neq 0$ ,

donde  $\theta_1$  y  $\theta_0$  son las medianas del potasio plasmático del grupo 1 (hiponatremia) y del grupo 0 (normonatremia), respectivamente.

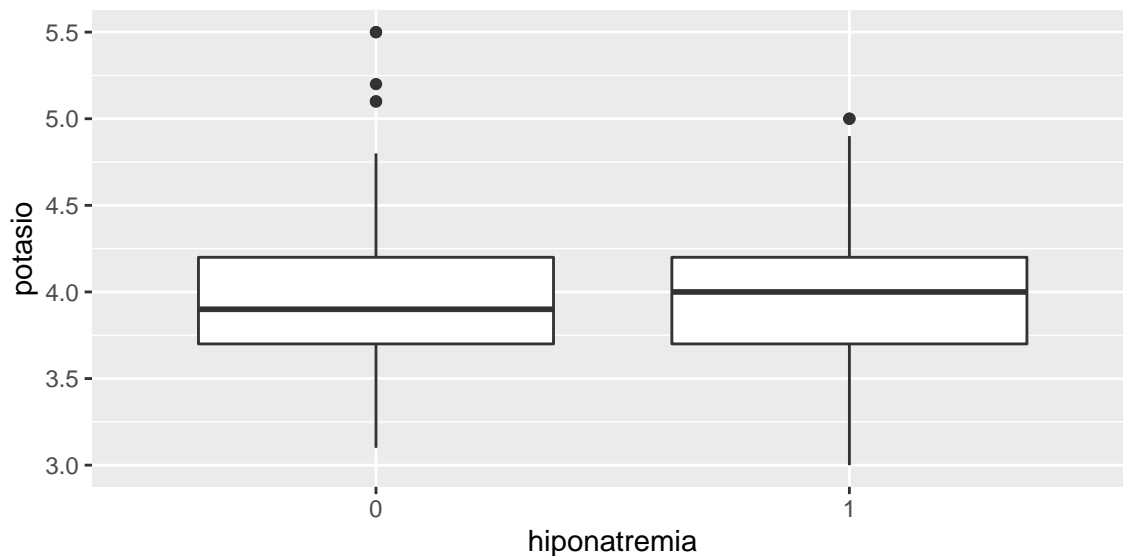
```
wilcox.test(dengue1$potasio,dengue0$potasio, paired = FALSE, conf.int = TRUE)
```

```
##  
## Wilcoxon rank sum test with continuity correction  
##  
## data: dengue1$potasio and dengue0$potasio  
## W = 2616.5, p-value = 0.5693
```

```
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##  -0.1000781  0.1999464
## sample estimates:
## difference in location
##          1.918249e-05
```

No se rechaza la hipótesis nula del test de Wilcoxon ya que el p-valor dio mayor a 0.05. Entonces, el potasio plasmático no presenta diferencias significativas entre el grupo sin hiponatremia y el grupo con hiponatremia. Por lo tanto, no hay una asociación estadísticamente significativa entre el potasio y la hiponatremia en individuos infectados con el virus del dengue.

```
bplt <- ggplot(dengue, aes(x=hiponatremia ,y=potasio)) +
  geom_boxplot()
bplt
```



En el gráfico de los boxplot se puede ver que para ambos grupos los valores de potasio son similares.

## Hiperglucemia

### Test exacto de Fisher

Las hipótesis del test exact de Fisher son:

$H_0$ : la hiperglucemia y la hiponatremia son variables independientes

$H_1$ : la hiperglucemia y la hiponatremia no son variables independientes.

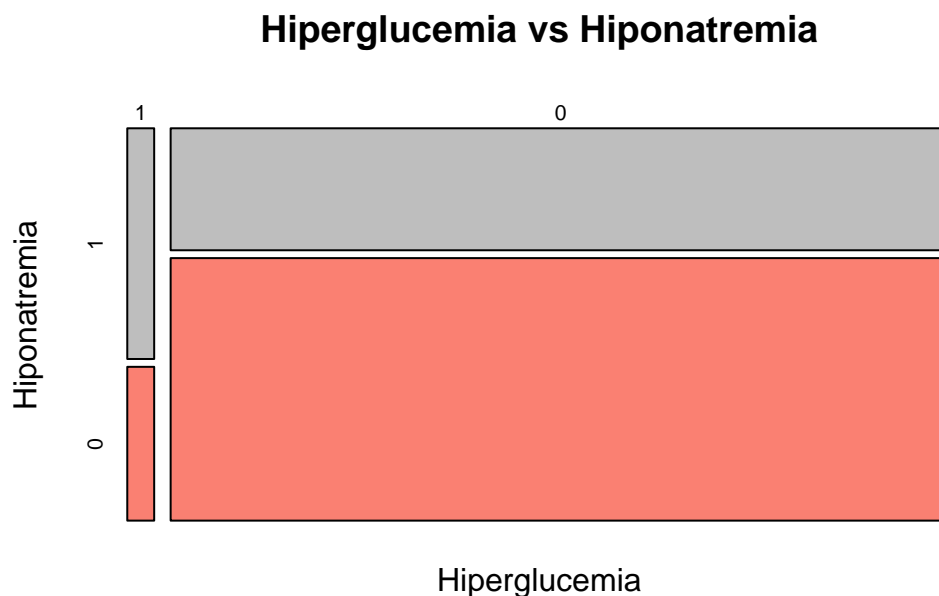
```
fisher.test(t_hglucemia)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  t_hglucemia
## p-value = 0.3306
```

```
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##    0.3540808 39.5461133
## sample estimates:
## odds ratio
##    3.200408
```

El test arrojó un p-valor mayor a 0.05, por lo que no se rechaza la hipótesis nula. Por lo tanto, la hiperglucemia y el desarrollo de hiponatremia en individuos infectados con el virus del dengue son variables independientes, es decir, que no hay una asociación significativa entre ellas.

```
mosaicplot(t_hglucemia, main = 'Hiperglucemia vs Hiponatremia', xlab = 'Hiperglucemia'
, ylab = 'Hiponatremia', color = c('grey', 'salmon'))
```



El gráfico de mosaicos muestra que el grupo con hiperglucemia presenta una mayor proporción de individuos con hiponatremia, aunque la diferencia dio no significativa. También se puede ver que el grupo sin hiperglucemia tiene una cantidad mucho mayor de individuos que el grupo con hiperglucemia.

## Resumen de los análisis univariados

```
tabla_1 <- data.frame(Variable = c('Edad (mediana)', 'Hombre', 'Potasio (mediana)',
    'Cloro (mediana)', 'Urea (mediana)', 'Creatinina (mediana)',
    'Hiperglucemia'),
    Normonatremia = c('37 años (RI: 21 años)', '50.49% (101)',
    '3.9 mEq/L (RI: 0.5 mEq/L)', '99.60 mEq/L (RI: 3.60 mEq/L)',
    '0.22 g/L (RI: 0.11 g/L)', '0.8 mg/dl (RI: 0.2 mg/dl)',
    '1.98% (101)'),
```

```

Hiponatremia = c('48 años (RI: 36 años)', '51.02% (49)',
'4.0 mEq/L (RI: 0.5 mEq/L)', '96.90 mEq/L (RI: 2.80 mEq/L)',
'0.30 g/L (RI: 0.17 g/L)', '0.90 mg/dl (RI: 0.4 mg/dl)',
'6.12% (49)'),
P_valor = c('6.318e-05', '1', '0.5693', '7.041e-09', '0.0006667',
'0.05757', '0.3306'))
kable(tabla_1, booktabs = T) %>%
  kable_styling(latex_options = "striped")

```

Variable	Normonatremia	Hiponatremia	P_valor
Edad (mediana)	37 años (RI: 21 años)	48 años (RI: 36 años)	6.318e-05
Hombre	50.49% (101)	51.02% (49)	1
Potasio (mediana)	3.9 mEq/L (RI: 0.5 mEq/L)	4.0 mEq/L (RI: 0.5 mEq/L)	0.5693
Cloro (mediana)	99.60 mEq/L (RI: 3.60 mEq/L)	96.90 mEq/L (RI: 2.80 mEq/L)	7.041e-09
Urea (mediana)	0.22 g/L (RI: 0.11 g/L)	0.30 g/L (RI: 0.17 g/L)	0.0006667
Creatinina (mediana)	0.8 mg/dl (RI: 0.2 mg/dl)	0.90 mg/dl (RI: 0.4 mg/dl)	0.05757
Hiperglucemia	1.98% (101)	6.12% (49)	0.3306

## 6. ¿Para el análisis multivariado qué tipo de modelo utilizaría?

Para el análisis multivariado utilizaría un modelo de regresión logística ya que la variable respuesta es dicotómica. Se busca ajustar un modelo para estimar la probabilidad de ocurrencia del evento, que en este caso es la hiponatremia en individuos infectados por el virus del dengue, según la presencia o no o el nivel de factores que influyen en la ocurrencia del evento.

Los supuestos del modelo de regresión logística son:

- la variable respuesta es dicotómica. Este supuesto se cumple ya que la variable respuesta es 1 si el individuo presenta hiponatremia y 0 si no la presenta.
- las observaciones son independientes. Este supuesto se cumple ya que cada una de las observaciones corresponde a un paciente distinto.
- linealidad de logit(p). Se supone que  $\text{logit}(p) = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_n \cdot x_n$

## 7. ¿Qué variables agregaría al modelo para un primer análisis? Plantee el modelo teórico. Corra el modelo e interprete los resultados.

En un primer análisis agregaría todas las variables al modelo ya que, si bien en el análisis anterior algunas asociaciones dieron no significativas, al incluir varias variables puede verse modificada esta asociación por un efecto de confusión o efecto modificador. A partir de la significatividad de los coeficientes de cada variable puede luego decidirse qué variables conservar y cuáles no.

El modelo teórico es:

$$\text{logit}(p_{\text{hiponatremia}}) = \beta_0 + \beta_{\text{genero}} \cdot \text{genero} + \beta_{\text{edad}} \cdot \text{edad} + \beta_{\text{urea}} \cdot \text{urea} + \beta_{\text{creatinina}} \cdot \text{creatinina} + \beta_{\text{cloro}} \cdot \text{cloro} + \beta_{\text{potasio}} \cdot \text{potasio} + \beta_{\text{hiperglucemia}} \cdot \text{hiperglucemia}$$

Para poder realizar el modelo con todos los datos, reemplazo el dato faltante de la variable creatinina por la mediana de esta variable. Luego, corro el modelo.

```

mediana = median(na.omit(dengue$creatinina))
dengue$creatinina <- na.replace(dengue$creatinina,mediana)
m1 <- glm(hiponatremia ~ genero + edad + urea + creatinina + cloro + potasio +
          hiperglucemia, data = dengue, family = "binomial")
summary(m1)

```

```

##
## Call:
## glm(formula = hiponatremia ~ genero + edad + urea + creatinina +
##      cloro + potasio + hiperglucemia, family = "binomial", data = dengue)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1017  -0.6668  -0.3662   0.5971   2.3621
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   43.20339    9.84371   4.389 1.14e-05 ***
## genero        -0.65896    0.49023  -1.344 0.178892
## edad          0.01292    0.01384   0.933 0.350587
## urea          8.44970    2.41338   3.501 0.000463 ***
## creatinina    -1.27491    0.66269  -1.924 0.054376 .
## cloro         -0.46232    0.10090  -4.582 4.60e-06 ***
## potasio       0.01453    0.53463   0.027 0.978316
## hiperglucemia 0.70590    1.79506   0.393 0.694138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 189.54  on 149  degrees of freedom
## Residual deviance: 130.29  on 142  degrees of freedom
## AIC: 146.29
##
## Number of Fisher Scoring iterations: 5

```

A partir de los p-valores de los coeficientes estimados para cada variable podemos ver que las variables que resultan significativas para el modelo son la urea y el cloro, ya que sus p-valores son menores a 0.05. En cambio, el género, la edad, la creatinina, el potasio y la hiperglucemia no dieron significativas para el modelo ya que sus p-valores son mayores a 0.05. Esto significa que no se rechaza la hipótesis nula del test de Wald que dice que los coeficientes son iguales a cero. En el caso de la creatinina, si bien el p-valor es mayor a 0.05, está en el límite. Si comparamos con los análisis univariados, en general se mantuvieron las significancias de las variables, con la excepción de la edad que sí había dado una asociación significativa con el evento. Esto se puede deber a un efecto confusor que puede haber con alguna otra variable.

## Significancia del modelo

El test omnibus evalúa la hipótesis nula de que los coeficientes de todas las variables incluidas en el modelo son iguales a cero.

```

dengue$hiponatremia<-as.numeric(dengue$hiponatremia)
omnibus<- aov(m1, data = dengue)
summary(omnibus)

```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## genero      1  0.001   0.001   0.005  0.94314
## edad        1  4.347   4.347  27.696 5.13e-07 ***
## urea         1  1.328   1.328   8.463  0.00421 **
## creatinina   1  0.295   0.295   1.878  0.17268
## cloro         1  4.673   4.673  29.771 2.11e-07 ***
## potasio       1  0.000   0.000   0.003  0.95680
## hiperglucemia 1  0.062   0.062   0.397  0.52964
## Residuals    142 22.287   0.157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El test omnibus nos permite ver si todos los coeficientes del modelo, excepto el de Intersección, son significativamente distintos de 0. La hipótesis nula del modelo plantea que los coeficientes del modelo son iguales a 0. La alternativa, en cambio, dice que alguno de los coeficientes es distinto de 0.

Por lo tanto, como para los coeficientes de la *edad*, la *urea* y el *cloro* los p-valores son menores a 0.05, estos coeficientes son significativamente diferentes de 0. Entonces, se rechaza a hipótesis nula y el modelo es significativo.

## Bondad de ajuste

Evalúo la bondad del ajuste del modelo mediante el test de Hosmer-Lemeshow, que compara las frecuencias esperadas con las observadas. Dado que un modelo brinda un buen ajuste cuando las frecuencias esperadas y las observadas son muy similares, y que la hipótesis nula del test plantea que estas frecuencias son iguales, entonces un modelo dará un buen ajuste cuando no se rechace la  $H_0$ .

```
h11 <- hoslem.test(m1$y, m1$fitted.values)
h11
```

```
##
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  m1$y, m1$fitted.values
## X-squared = 4.0634, df = 8, p-value = 0.8514
```

El modelo brinda un buen ajuste ya que el p-valor es mayor a 0.05, y entonces no se rechaza la hipótesis nula que dice que las frecuencias observadas son iguales a las esperadas.

## 8. Del modelo planteado en la pregunta 7 ¿Qué variables conservaría? Corra el modelo e interprete lo resultados. ¿Cómo lo compararía con el modelo anterior?

Dado que las únicas variables que dieron significativas para el modelo fueron el *cloro* y la *urea*, conservaría solamente estas dos.

El modelo teórico es:

$$\text{logit}(p_{\text{hiponatremia}}) = \beta_0 + \beta_{\text{urea}} \cdot \text{urea} + \beta_{\text{cloro}} \cdot \text{cloro}$$

Genero el nuevo modelo.

```
dengue$hiponatremia <- as.factor(dengue$hiponatremia)
m2 <- glm(hiponatremia ~ urea + cloro, data = dengue, family = "binomial")
summary(m2)
```

```
##
## Call:
## glm(formula = hiponatremia ~ urea + cloro, family = "binomial",
##      data = dengue)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8438  -0.7722  -0.4626   0.7912   2.2545
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 35.19422    8.23891   4.272 1.94e-05 ***
## urea         5.46139    1.64143   3.327 0.000877 ***
## cloro        -0.38135    0.08505  -4.484 7.34e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 189.54  on 149  degrees of freedom
## Residual deviance: 144.32  on 147  degrees of freedom
## AIC: 150.32
##
## Number of Fisher Scoring iterations: 5
```

Con el nuevo modelo, se puede ver que las significancias de ambos coeficientes se mantuvieron, ya que los dos p-valores dieron menores a 0.05.

## Significancia del modelo

Realizo el test omnibus para el nuevo modelo.

```
dengue$hiponatremia<-as.numeric(dengue$hiponatremia)
omnibus<- aov(m2, data = dengue)
summary(omnibus)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## urea           1  3.700    3.700   22.21 5.61e-06 ***
## cloro          1  4.805    4.805   28.84 3.00e-07 ***
## Residuals     147 24.488    0.167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Podemos ver que todos los coeficientes del modelo son significativamente diferentes de 0, ya que sus p-valores son menores a 0.05. Se rechaza la hipótesis nula del test omnibus que dice que todos los coeficientes del modelo son iguales a 0, por lo que el modelo es significativo.

## Bondad de ajuste

Evalúo la bondad del ajuste del modelo mediante el test de Hosmer-Lemeshow.



```
h12 <- hoslem.test(m2$y, m2$fitted.values)
h12
```

```
##
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data: m2$y, m2$fitted.values
## X-squared = 4.7754, df = 8, p-value = 0.7813
```

Este modelo de regresión logística brinda un buen ajuste, ya que el p-valor del test dio mayor a 0.05, por lo que no se rechaza la hipótesis nula que dice que las frecuencias esperadas y observadas son iguales.

## Comparación entre los modelos

**Deviance.** La *deviance* es una medida de confiabilidad del modelo que habla de la desviación del modelo ajustado con respecto al modelo perfecto, o saturado. Cuanto menor sea esta desviación, mejor explicará el modelo el comportamiento de la variable respuesta. Por ende, si comparamos la *deviance* de los dos modelos generados, el que menor *deviance* tenga será el que mejor ajuste tenga.

Modelo 1

```
m1$deviance
```

```
## [1] 130.2857
```

Modelo 2

```
m2$deviance
```

```
## [1] 144.3231
```

Podemos ver que el modelo que solo incluye las variables **urea** y **cloro** tiene una *deviance* mayor que el modelo que utiliza a todas las variables. Por esta razón, el modelo que tiene mayor capacidad para explicar el comportamiento de la variable respuesta es el que tiene a todas las variables explicativas, es decir, el modelo 1.

**Curva ROC y Área bajo la curva.** También podemos comparar la performance de los modelos mediante la curva ROC, que grafica la sensibilidad del modelo vs. la especificidad. Dado que el modelo de regresión logística devuelve una probabilidad de ocurrencia del evento, se puede generar una función que defina la ocurrencia o no del evento según:

$$Y = \begin{cases} 1, & \text{if } p \geq u \\ 0, & \text{if } p < u, \end{cases}$$

donde  $p$  es la probabilidad obtenida del modelo y  $u$  es un umbral de probabilidad a partir del cual se considera que el evento ocurre. De esta forma, se va variando el umbral desde 0 hasta 1 y para cada caso se calcula la sensibilidad y la especificidad del modelo. Luego, estos puntos se grafican en la llamada curva ROC. Cuanto más por encima de la recta de no discriminación se encuentre la curva, mejor será el modelo. Esto también se puede medir con el área bajo la curva, ya que esta será mayor cuanto más por encima esté la curva de la recta de no discriminación.

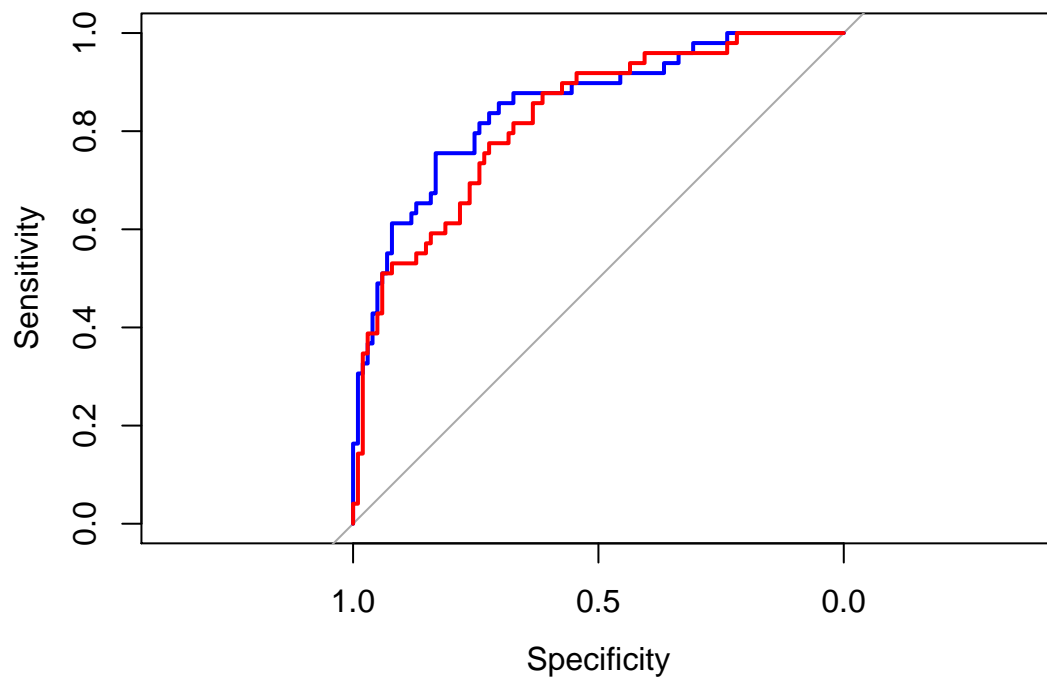
```

#modelo con todas las variables
prob=predict(m1,type=c("response"))
dengue$prob1 = prob
roc1 <- roc(hiponatremia ~ prob1, data = dengue)

#modelo con urea y cloro
prob=predict(m2,type=c("response"))
dengue$prob2 = prob
roc2 <- roc(hiponatremia ~ prob2, data = dengue)

plot(roc1, col = 'blue')
plot(roc2, add=TRUE, col='red')

```



- Azul: modelo con todas las variables
- Rojo: modelo con dos variables.

```

#modelo 1
roc1$auc

```

```
## Area under the curve: 0.8505
```

```

#modelo 2
roc2$auc

```

```
## Area under the curve: 0.8232
```

Con el gráfico de las curvas ROC podemos ver que ambas curvas son bastante similares, aunque la curva azul, correspondiente al modelo que incluye a todas las variables parece alejarse un poco más de la curva de no discriminación. Esto se puede ver mejor con el AUC, que dio un poco mayor en el caso del modelo 1, aunque tampoco se observa una diferencia muy significativa.

Esta comparación también indica que el modelo que mejor se ajusta a los datos es el modelo que tiene a todas las variables explicativas.

## 9. Resuma los datos del análisis multivariado en una tabla y elabore una conclusión.

### Resumen del modelo con todas las variables

Calculo los OR de todas las variables y sus intervalos de confianza.

```
OR1<-exp(coefficients(m1))
IC1<- confint(m1)
c<-exp(IC1)
OR_IC_1<- cbind (OR1, c)
OR_IC_1
```

```
##              OR1          2.5 %          97.5 %
## (Intercept)  5.794182e+18  8.957488e+10  6.936006e+27
## generom     5.173913e-01  1.924115e-01  1.333036e+00
## edad        1.013002e+00  9.858685e-01  1.041230e+00
## urea         4.673649e+03  4.996992e+01  7.963224e+05
## creatinina   2.794575e-01  3.992312e-02  6.930520e-01
## cloro        6.298200e-01  5.075387e-01  7.560895e-01
## potasio      1.014637e+00  3.427751e-01  2.839746e+00
## hiperglucemia1 2.025671e+00  7.712349e-02  6.502001e+01
```

```
tabla_2 <- data.frame(Variable = c('Género', 'Edad', 'Urea', 'Creatinina', 'Cloro',
                                   'Potasio', 'Hiperglucemia'),
                      OR = c('0.5172274', '1.012979', '4629.620', '0.2799056',
                              '0.6303235', '1.014646', '2.024767'),
                      IC_inf = c('0.1924325', '0.9858550', '49.64074',
                                  '0.04002313', '0.5078305', '0.3429473', '0.07734262'),
                      IC_sup = c('1.332166', '1.041201', '788639.6', '0.6940406',
                                  '0.7569227', '2.838558', '64.79183'),
                      P_valor = c('0.178505', '0.351292', '0.000468', '0.054533',
                                   '4.93e-06', '0.978293', '0.693874'))
kable(tabla_2, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

Variable	OR	IC_inf	IC_sup	P_valor
Género	0.5172274	0.1924325	1.332166	0.178505
Edad	1.012979	0.9858550	1.041201	0.351292
Urea	4629.620	49.64074	788639.6	0.000468
Creatinina	0.2799056	0.04002313	0.6940406	0.054533
Cloro	0.6303235	0.5078305	0.7569227	4.93e-06
Potasio	1.014646	0.3429473	2.838558	0.978293
Hiperglucemia	2.024767	0.07734262	64.79183	0.693874

Del modelo con todas las variables explicativas podemos sacar las siguientes conclusiones:

- Por cada 1 mEq/L de cloro que se aumenta en el nivel plasmático de cloro, disminuyen en 0.63 veces las chances de padecer hiponatremia en una persona infectada por el virus del dengue, con un intervalo de confianza del 95% entre 0.51 y 0.76. Esta asociación resultó estadísticamente significativa, con un p-valor menor a 0.05.
- Por cada aumento de 1 g/L en el nivel de urea, aumentan en 4629.62 veces las chances de padecer hiponatremia en una persona infectada por el virus del dengue, con un intervalo de confianza del 95% entre 49.64 y 788639.60. Dicha asociación resulta estadísticamente significativa, con un p-valor menor a 0.05.
- Las variables edad, género, creatinina, potasio e hiperglucemia no resultaron tener una asociación estadísticamente significativa con la hiponatremia en persona infectadas por el virus del dengue.

### Resumen del modelo con dos variables explicativas

Calculo los OR de las dos variables incluidas y sus intervalos de confianza.

```
OR2<-exp(coefficients(m2))
IC2<- confint(m2)
c<-exp(IC2)
OR_IC_2<- cbind (OR2, c)
OR_IC_2
```

```
##                OR2          2.5 %          97.5 %
## (Intercept) 1.925995e+15 5.763961e+08 8.477803e+22
## urea        2.354249e+02 1.039552e+01 6.992910e+03
## cloro       6.829408e-01 5.693643e-01 7.970399e-01
```

```
tabla_3 <- data.frame(Variable = c('Urea', 'Cloro'),
                      OR = c('235.4249', '0.6829408'),
                      IC_inf = c('10.39552', '0.5693643'),
                      IC_sup = c('6992.910', '0.7970399'),
                      P_valor = c('0.000877', '7.34e-06'))
kable(tabla_3, booktabs = T) %>%
  kable_styling(latex_options = "striped")
```

Variable	OR	IC_inf	IC_sup	P_valor
Urea	235.4249	10.39552	6992.910	0.000877
Cloro	0.6829408	0.5693643	0.7970399	7.34e-06

Del modelo que solo tiene en cuenta al cloro y a la urea podemos sacar las siguientes conclusiones:

- Por cada 1 mEq/L de cloro que se aumenta en el nivel plasmático de cloro, disminuyen en 0.68 veces las chances de padecer hiponatremia en una persona infectada por el virus del dengue, con un intervalo de confianza del 95% entre 0.57 y 0.80. Esta asociación resultó estadísticamente significativa, con un p-valor menor a 0.05.

- Por cada aumento de 1 g/L en el nivel de urea, aumentan en 235.42 veces las chances de padecer hiponatremia en una persona infectada por el virus del dengue, con un intervalo de confianza del 95% entre 10.40 y 6992.91. Dicha asociación resulta estadísticamente significativa, con un p-valor menor a 0.05.
- El OR de la urea varía significativamente entre el modelo 1 y el modelo 2. Esto se debe a que existe una variable que actúa como modificador de efecto.