

Actividad 1.2 Normalidad univariada

Franco Mendoza Muraira A01383399

2023-11-01

```
library(nortest)
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.2.3
```

Pruebas de normalidad variable velocidad:

Prueba de normalidad univariada Kolmogorov Smirnov en velocidad:

```
print(lillie.test(cars$speed))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  cars$speed
## D = 0.068539, p-value = 0.8068
```

Debido al alto pvalue de 0.8, con esta prueba rechaza H_0 , por lo que se concluye que la variable speed tiene una distribución normal, el pvalue tendría que ser menor que 0.1 aproximadamente dependiendo de el alfa establecido, para que la distribución no sea considerada normal, en nuestro caso el nivel de confianza es de 90% por lo que el alfa es 0.1.

Prueba de normalidad univariada Anderson Darling en velocidad:

```
print(ad.test(cars$speed))
```

```
##
##  Anderson-Darling normality test
##
## data:  cars$speed
## A = 0.26143, p-value = 0.6927
```

Con la prueba Anderson-Darling se concluye lo mismo que la anterior, la distribución de la variable speed es normal.

Pruebas de normalidad variable distancia:

Prueba de normalidad univariada Kolmogorov Smirnov en distancia:

```
print(lillie.test(cars$dist))

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  cars$dist
## D = 0.12675, p-value = 0.04335
```

Con la variable distancia, en la prueba Smirnov se tiene un pvalue de 0.04 el cual es muy bajo, por lo que no se rechaza H_0 , y se concluye que la distribución no es normal.

Prueba de normalidad univariada Anderson Darling en distancia:

```
print(ad.test(cars$dist))

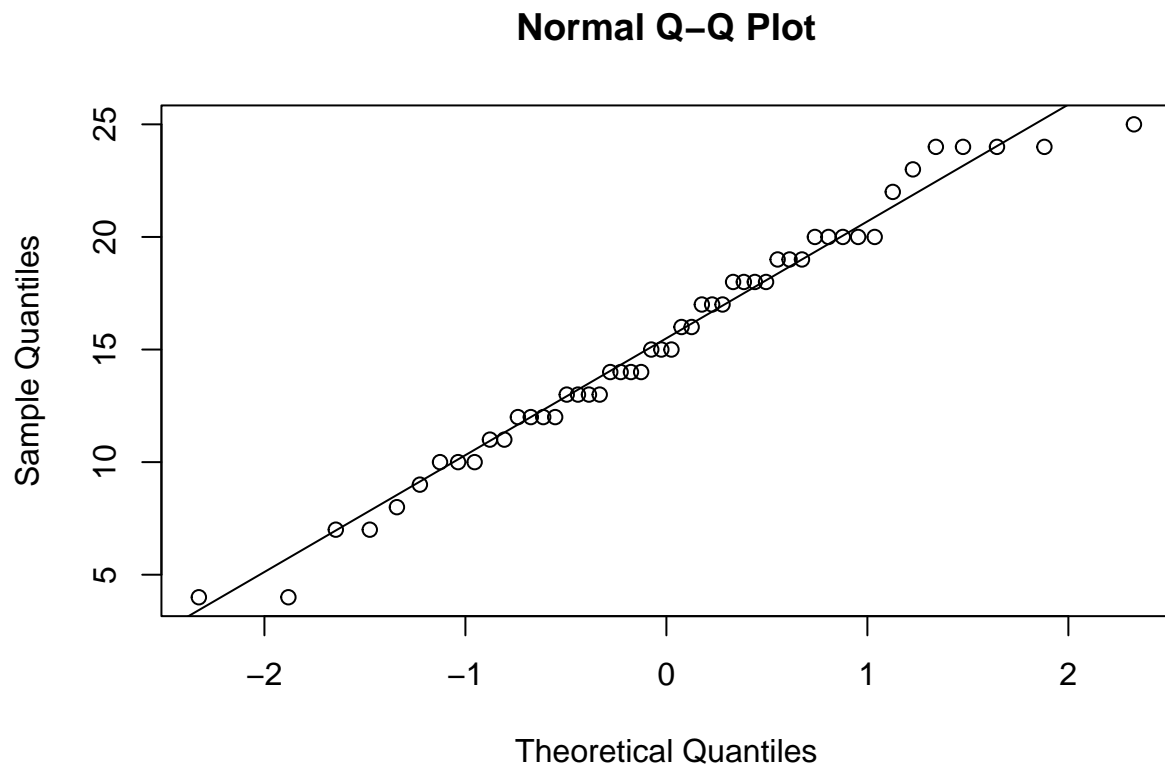
##
##  Anderson-Darling normality test
##
## data:  cars$dist
## A = 0.74067, p-value = 0.05021
```

Al igual que en la prueba pasada, el pvalue es muy chico por lo que la distribución no es considerada normal.

QQPlots

QQPlot de variable velocidad

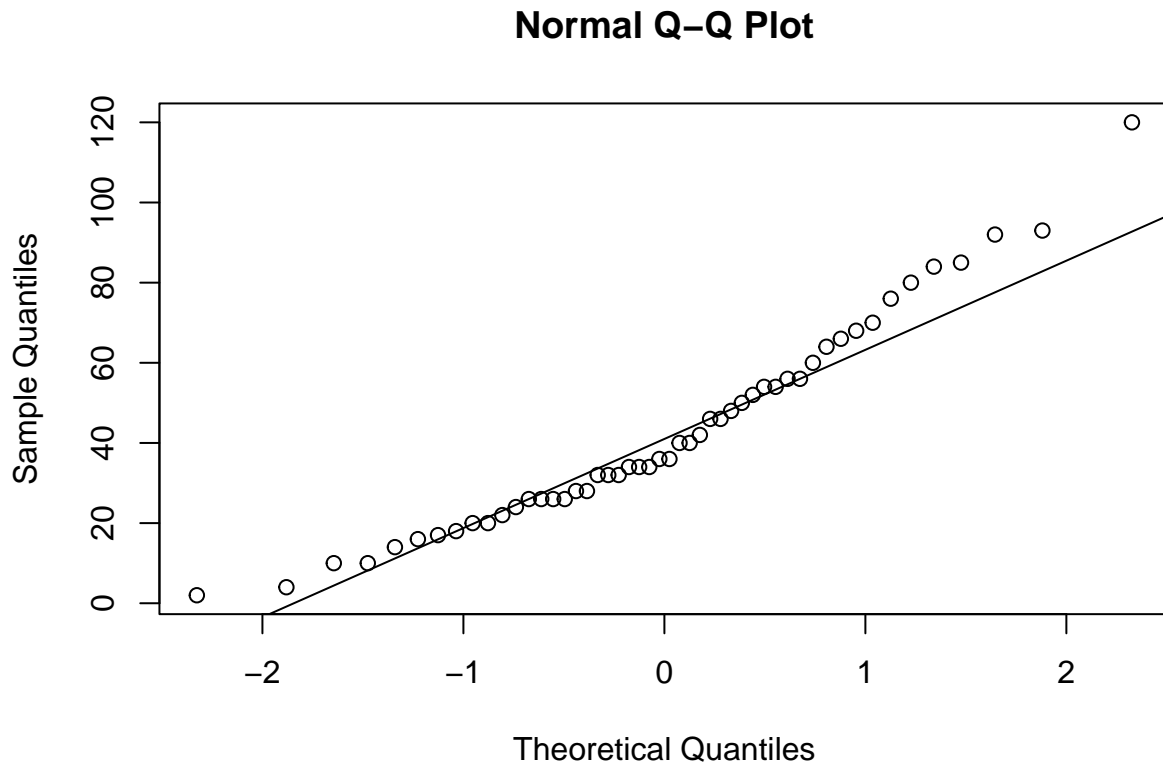
```
qqnorm(cars$speed)
qqline(cars$speed)
```



En la grafica se ve la distribucion de los datos a traves de los cuartiles, y se ve un plot ideal para la distribucion normal como nos dijeron las pruebas, pero debido a la esquina superior del plot, se ve que tendra alta curtosis ya que se acerca a una distribución con colas delgadas.

QQPlot de variable distancia

```
qqnorm(cars$dist)
qqline(cars$dist)
```



En este qqplot de la distancia, se puede ver una asimetría positiva lo que nos dice que los datos tienen un sesgo a la derecha.

Coeficientes de sesgo y curtosis variable velocidad:

Coeficiente de sesgo de la variable de velocidad:

```
skewness(cars$speed)
```

```
## [1] -0.1105533
```

Debido al coeficiente de sesgo bajo, se puede concluir que la distribución es moderadamente simétrica.

Coeficiente de curtosis de la variable de velocidad:

```
kurtosis(cars$speed)
```

```
## [1] -0.6730924
```

El coeficiente de curtosis nos dice que la forma de la forma de la distribución tiene un pico más plano y colas más ligeras, platicúrtica, en la variable de velocidad.

Coeficientes de sesgo y curtosis variable velocidad:

Coeficiente de sesgo de la variable de distancia:

```
print(skewness(cars$dist))
```

```
## [1] 0.7591268
```

En la variable de distancia se puede ver una curtosis mucho más alta de 0.75, esto nos dice que los datos están sesgados a la derecha.

Coeficiente de curtosis de la variable de distancia:

```
print(kurtosis(cars$dist))
```

```
## [1] 0.1193971
```

La curtosis en esta variable es muy baja, por lo que se acerca mucho a una forma de distribución mesocúrtica.

Estadísticas de la velocidad de los carros

Media

```
mean(cars$speed)
```

```
## [1] 15.4
```

Mediana

```
median(cars$speed)
```

```
## [1] 15
```

Moda

```
IQR(cars$speed)
```

```
## [1] 7
```

Estadísticas de las distancias de los carros

Media

```
mean(cars$dist)
```

```
## [1] 42.98
```

Mediana

```
median(cars$dist)
```

```
## [1] 36
```

Moda

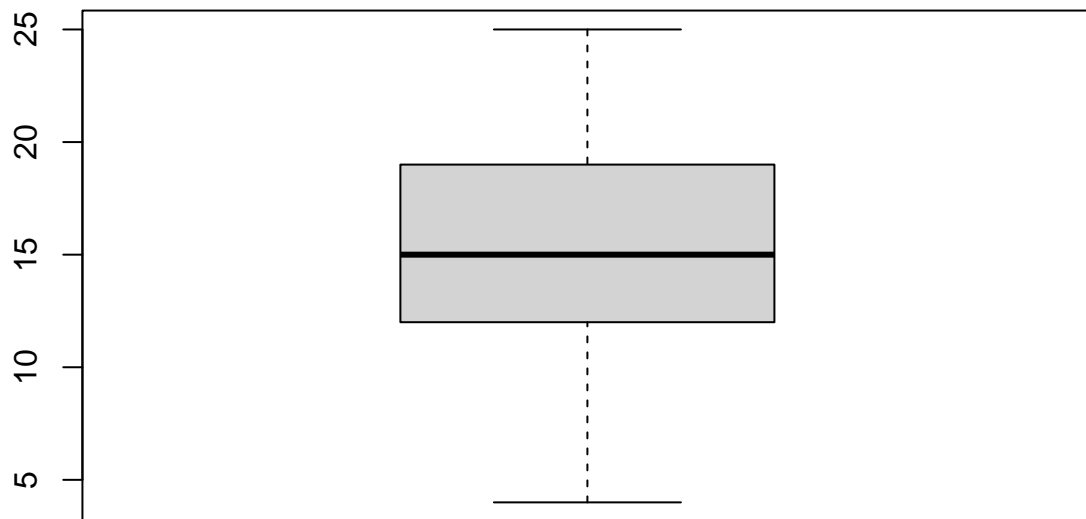
```
IQR(cars$dist)
```

```
## [1] 30
```

```
#Boxplots
```

Boxplot variable velocidad

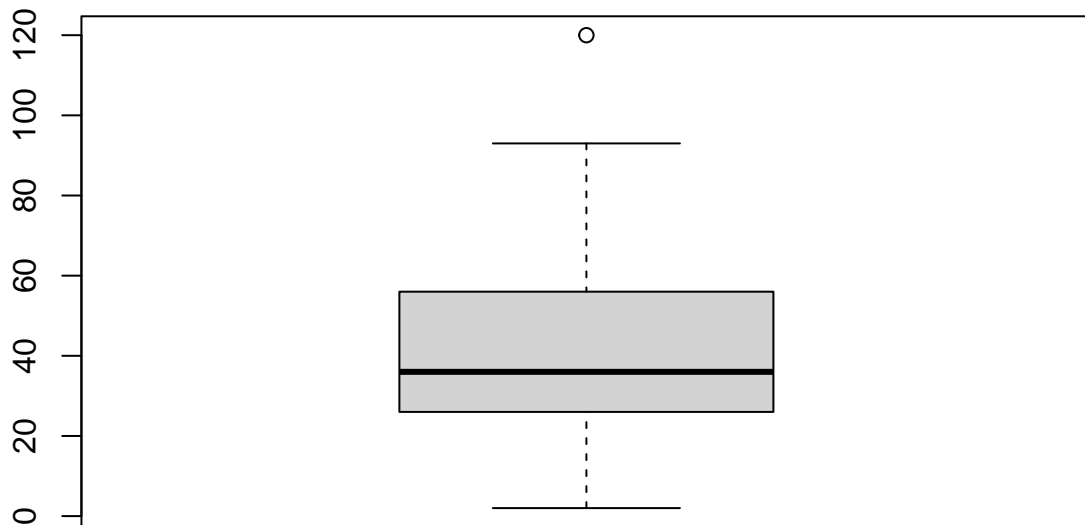
```
boxplot(cars$speed)
```



Esta es la distribución de los datos de la variable velocidad.

Boxplot variable distancia

```
boxplot(cars$dist)
```



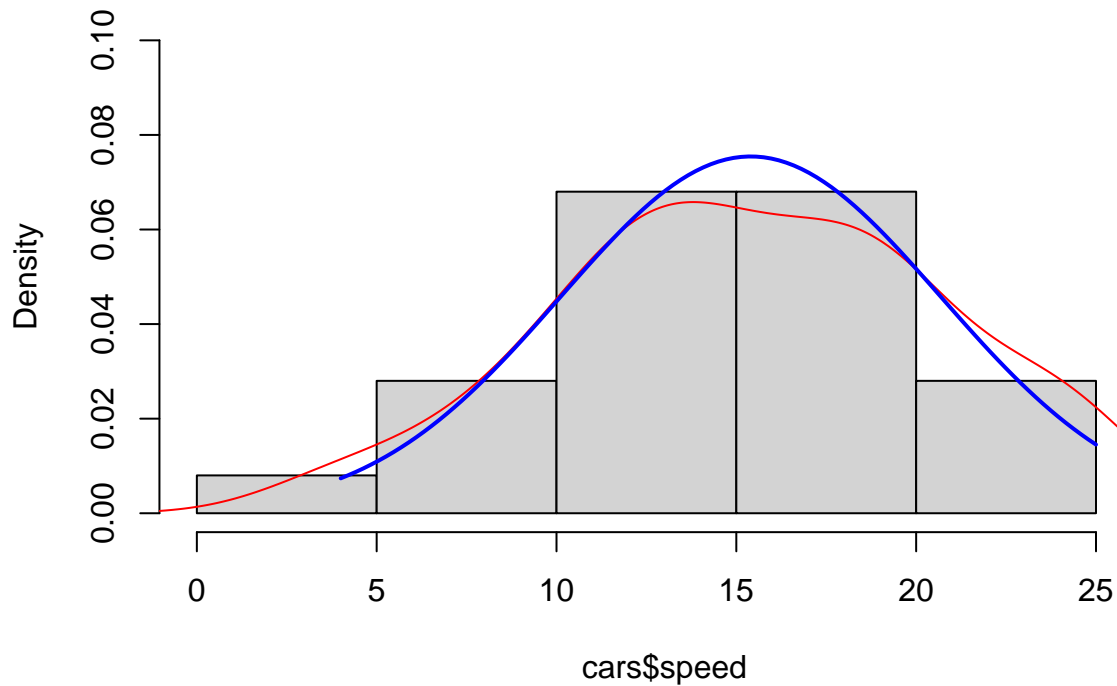
Esta es la distribución de los datos de la variable distancia.

Histogramas

##Histograma variable velocidad

```
hist(cars$speed, freq = FALSE, ylim = c(0,.1), xlim=c(0,26))  
lines(density(cars$speed), col = "red")  
curve(dnorm(x, mean = mean(cars$speed), sd = sd(cars$speed)), from = min(cars$speed), to = max(cars$speed))
```

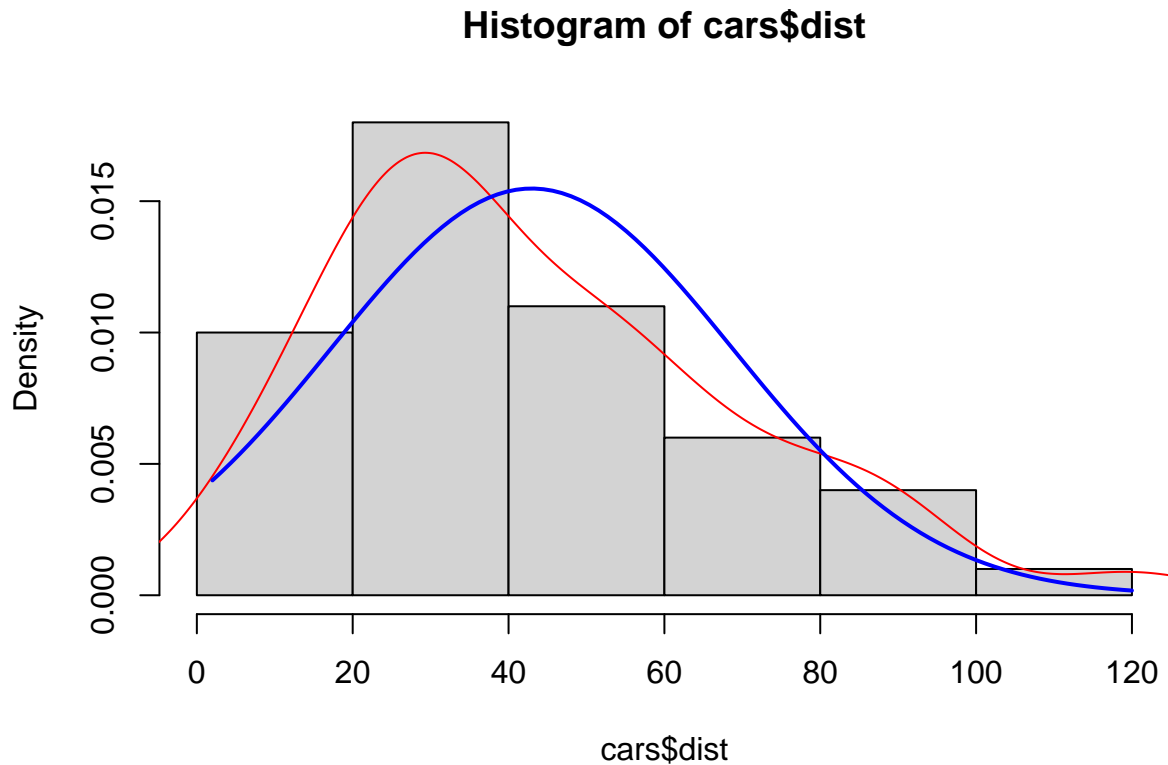
Histogram of cars\$speed



En el histograma de la variable de velocidad se pueden ver 2 líneas, la azul mostrándonos como se verían los datos si tuvieran una distribución normal, y la roja mostrando la distribución real. En este histograma se puede ver que las 2 líneas son muy parecidas, que demuestra lo que se ha visto desde nuestra prueba principal que es la variable más cercana a la distribución normal.

Histograma variable distancia

```
hist(cars$dist, freq = FALSE)
lines(density(cars$dist), col = "red")
curve(dnorm(x, mean = mean(cars$dist), sd = sd(cars$dist)), from = min(cars$dist), to = max(cars$dist),
```

En este histograma podemos ver el sesgo a la derecha que mencionamos previamente, y que las 2 líneas están mas lejanas en esta variable de distancia que en la de velocidad. Este histograma nos confirma igual lo de la prueba principal, que la variable está lejana de la distribución normal.

Conclusión

Con todas las pruebas, pudimos ver como cada dato y calculo hecho nos llevó a lo mismo, que fue la distribución normal de la variable de velocidad y la distribución no normal de la variable distancia, pudimos ver como la curtosis y sesgo trabajan juntos para solos poder darnos una visualización de como se distribuirán los datos, también en las estadísticas de las variables, se pudo ver una cercanía entre la mediana y moda de la velocidad, cuando se vió una lejanía en la distancia, estos son grandes indicadores y cualquiera puede servir para conseguir información útil sobre los datos, aunque no toda. Aún así se necesitan diferentes pruebas y análisis para poder asegurar lo que se hipotetiza.