

Actividad 1.2 Normalidad univariada

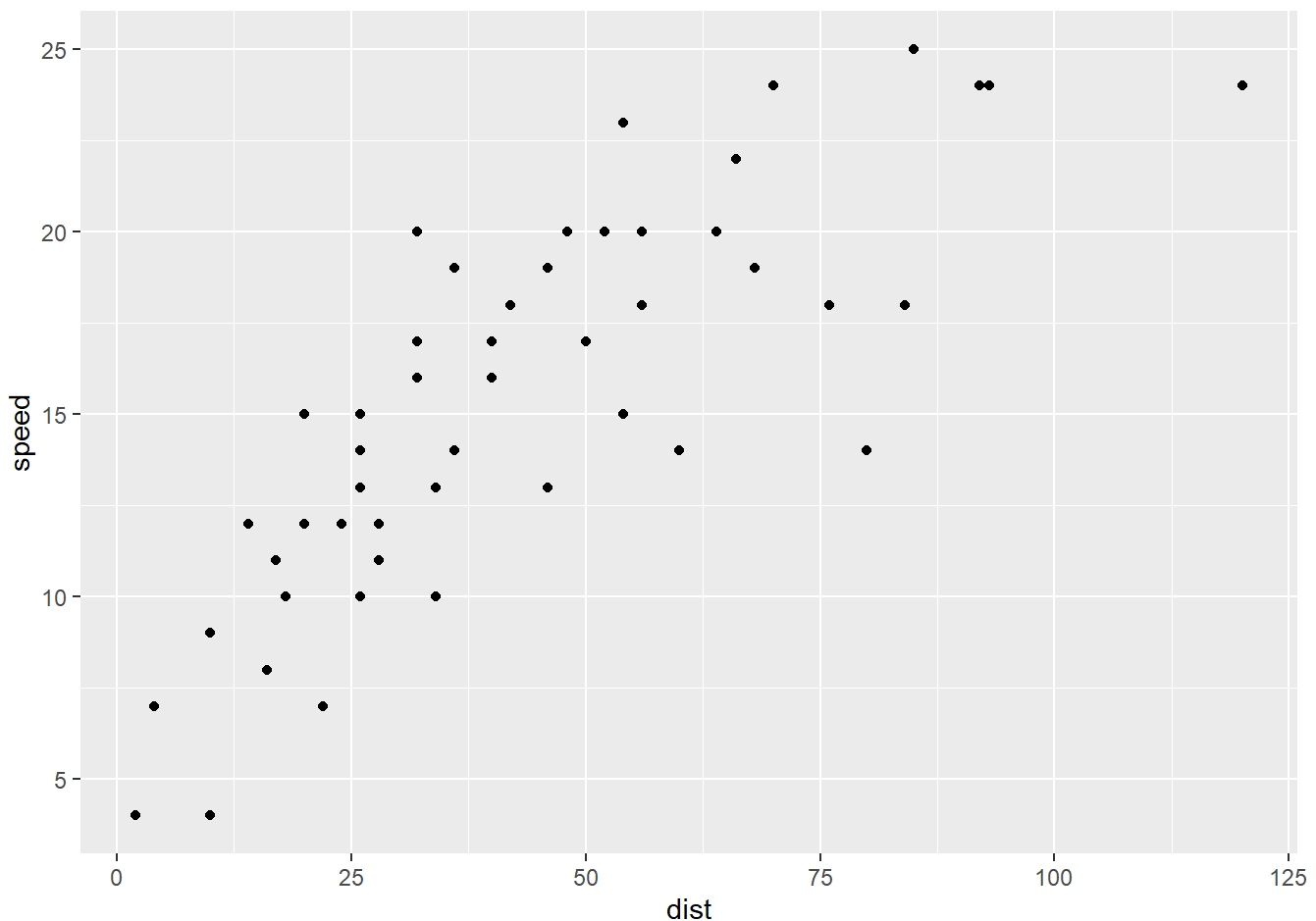
Raúl Correa Ocañas

2024-08-08

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
ggplot(cars, aes(x=dist, y=speed)) +  
  geom_point()
```



1. Prueba de Normalidad Univariada

```
# Rapidez  
shapiro.test(cars$speed)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: cars$speed  
## W = 0.97765, p-value = 0.4576
```

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

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

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

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

```
jarque.test(cars$speed)
```

```
##  
## Jarque-Bera Normality Test  
##  
## data: cars$speed  
## JB = 0.80217, p-value = 0.6696  
## alternative hypothesis: greater
```

Según los resultados obtenidos para las pruebas de normalidad de los datos para la variable de rapidez, todas las hipótesis nulas son aceptadas. Por lo tanto, no se tiene suficiente evidencia estadística para negar que la distribución es normal.

```
# Distancia  
shapiro.test(cars$dist)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: cars$dist  
## W = 0.95144, p-value = 0.0391
```

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

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

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

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

```
jarque.test(cars$dist)
```

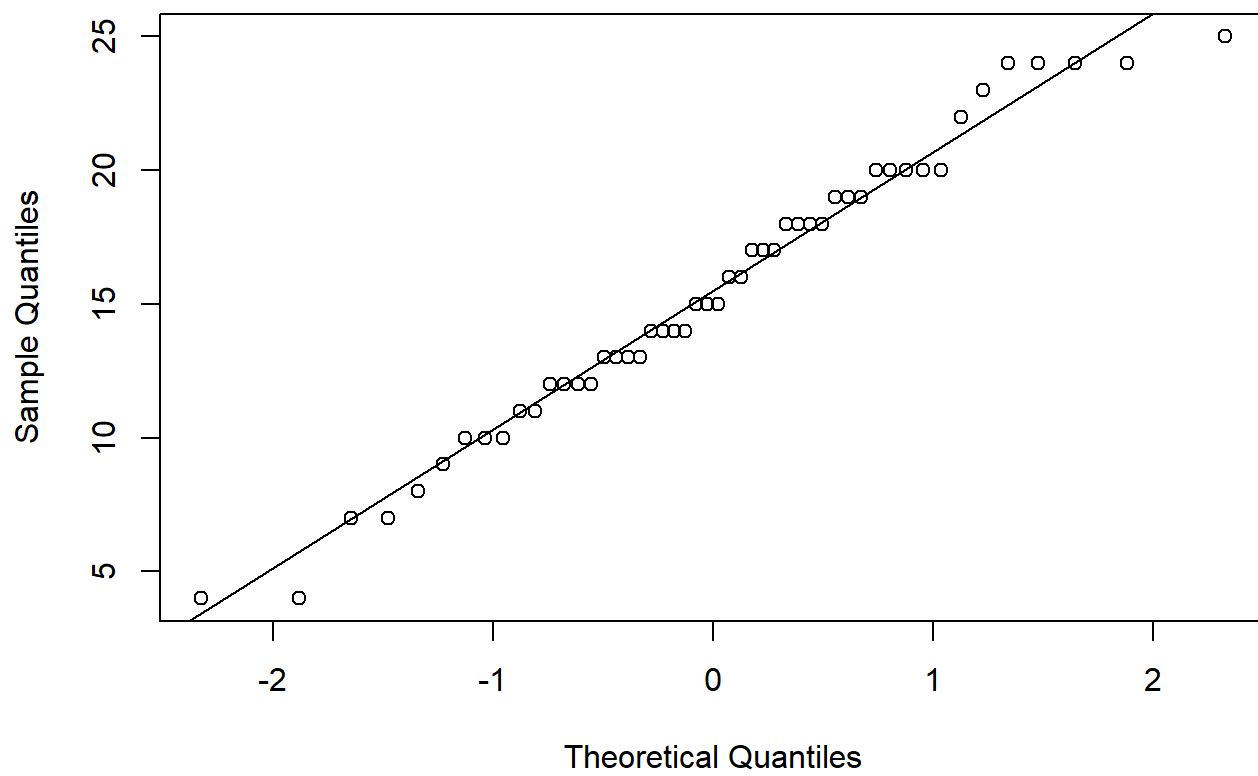
```
##  
## Jarque-Bera Normality Test  
##  
## data: cars$dist  
## JB = 5.2305, p-value = 0.07315  
## alternative hypothesis: greater
```

Por otro lado, las mismas pruebas de normalidad sobre la distribución de la distancia indican que se rechaza la hipótesis nula. Por lo tanto, se tiene suficiente evidencia estadística para afirmar que la distribución de los datos no es normal.

2. QQPlot

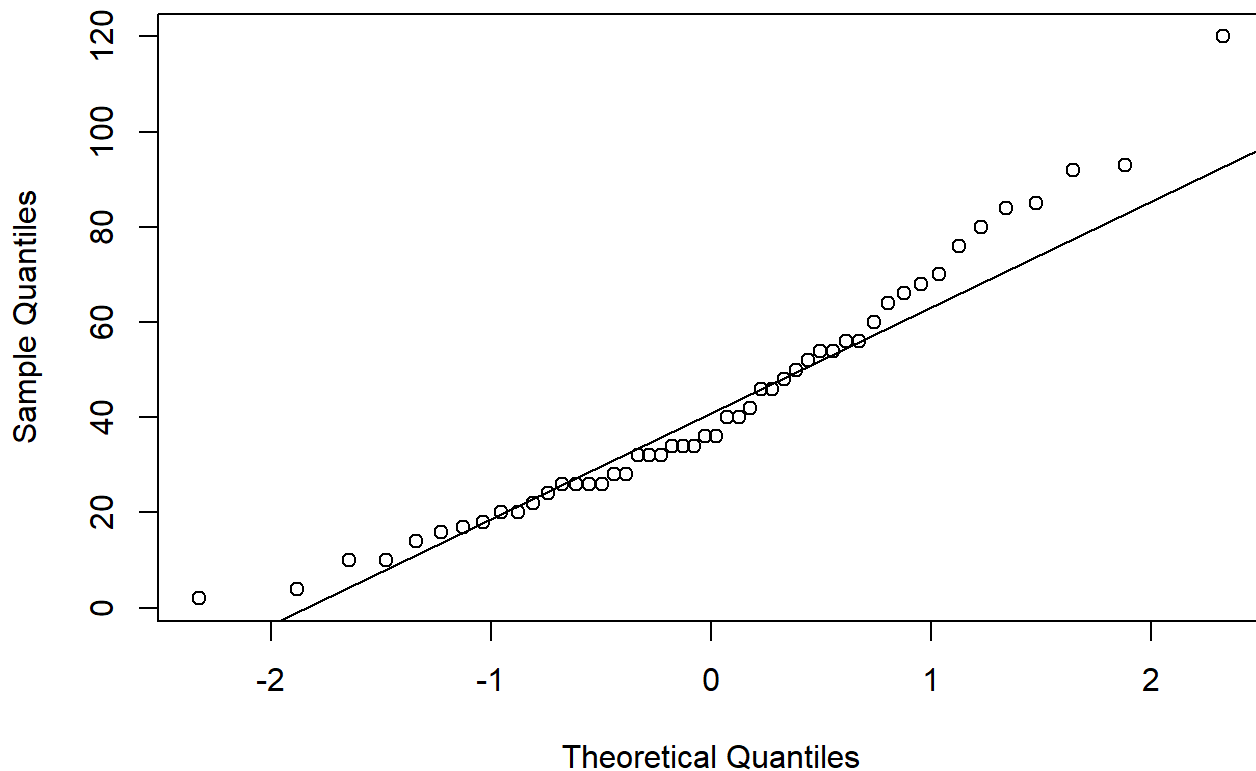
```
qqnorm(cars$speed, main = "Rapidez")  
qqline(cars$speed)
```

Rapidez



```
qqnorm(cars$dist, main = "Distancia")  
qqline(cars$dist)
```

Distancia



Se observa que los puntos se apegan cercanamente a la linea que representan los cuartiles teóricos para el caso de la rapidez, lo cual reafirma los resultados obtenidos con las pruebas estadísticas de normalidad. La curtosis negativa no es muy evidente, mas si está presente a partir de la primera desviación estandar hacia la derecha. En el caso de los datos de la distancia, es clara la presencia de curtosis positiva en las orillas del diagrama. La forma de como se distribuyen los puntos aparenta no ser una linea recta, por lo que esto confirma la sospecha de no normalidad de los datos.

3. Sesgo y Curtosis

```
print(paste("Sesgo de Rapidez: ", skewness(cars$speed)))
```

```
## [1] "Sesgo de Rapidez:  -0.110553277648993"
```

```
print(paste("Curtosis de Rapidez: ", kurtosis(cars$speed)))
```

```
## [1] "Curtosis de Rapidez:  -0.673092385955565"
```

```
print(paste("Sesgo de Distancia: ", skewness(cars$dist)))
```

```
## [1] "Sesgo de Distancia:  0.75912677852551"
```

```
print(paste("Curtosis de Distancia: ", kurtosis(cars$dist)))
```

```
## [1] "Curtosis de Distancia: 0.119397118346566"
```

Numericamente verificamos los resultados de la gráfica previa, viendo que la rapidez tiene un sesgo cercano a 0 con leve sesgo a la izquierda. Esto puede ser explicado por la ausencia de datos identificada en la grafica anterior a partir de la primera desviación estandar hacia la derecha. Esta misma ausencia hace que su curtosis se desvie de ser 0. Por el lado de los datos de la distancia, su sesgo es mucho más significativo hacia la derecha, lo cual nos indica nuevamente que la distribución no puede ser normal.

4. Media, Mediana, Rango Medio

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.   :120.00
```

```
print(paste("Rango medio de Rapidez: ", (max(cars$speed) + min(cars$speed))/2))
```

```
## [1] "Rango medio de Rapidez: 14.5"
```

```
print(paste("Rango medio de Distancia: ", (max(cars$dist) + min(cars$dist))/2))
```

```
## [1] "Rango medio de Distancia: 61"
```

La media y la mediana están relativamente cerca para los datos de rapidez, teniendo una diferencia de 0.4 relativamente chica a comparación de su rango medio de 14.5. En cuanto a los datos de la distancia, la diferencia entre media y mediana es de casi 9, mientras su rango medio es de 61. Esto muestra que la primera distribución puede ser normal, mientras que la segunda sería mucho menos probable que sea el caso.

5. Boxplots

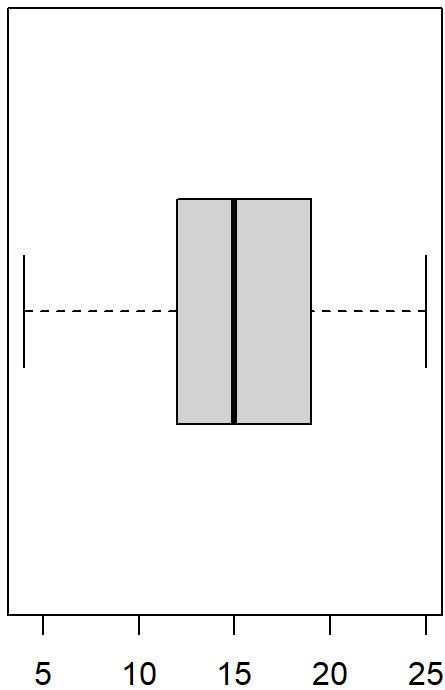
```
par(mfrow = c(1, 2))
```

```
# Create the boxplots
```

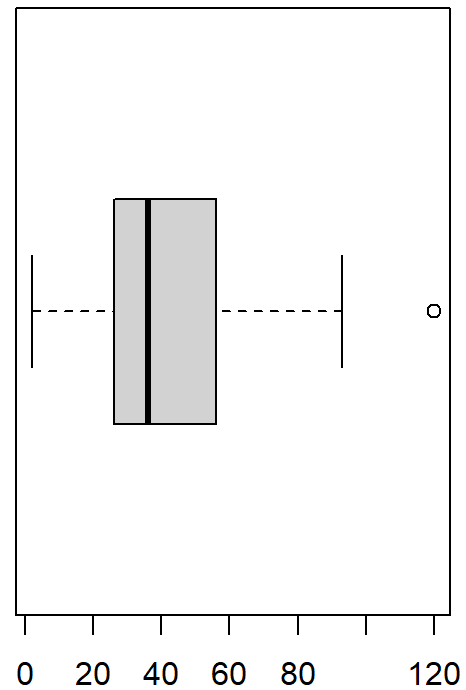
```
boxplot(cars$speed, main = "Rapidez", horizontal = TRUE)
```

```
boxplot(cars$dist, main = "Distancia", horizontal = TRUE)
```

Rapidez



Distancia

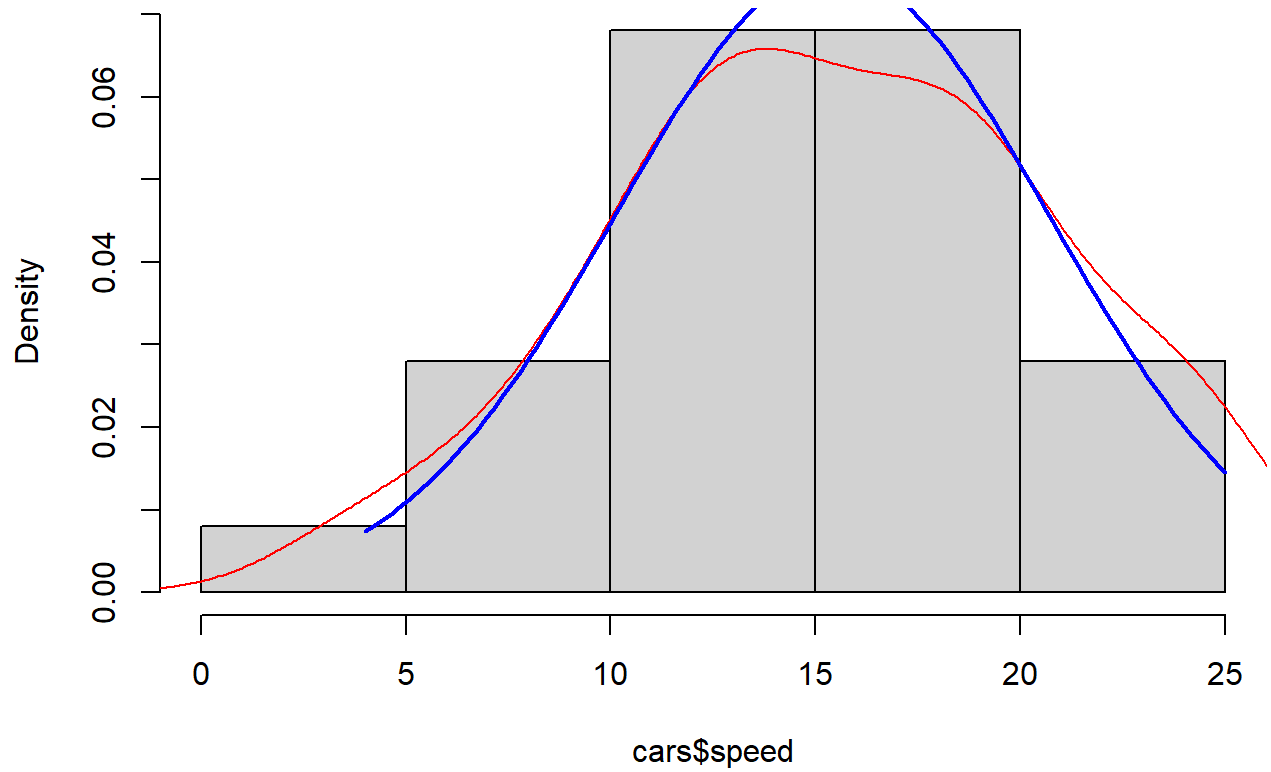


En el boxplot de los datos de velocidad, se puede ver que la mediana esta relativamente centrada, mientras que en el caso de la distancia se ve cargada hacia la izquierda, indicando un sesgo a la derecha. Esto indica nuevamente que la distribucion de los datos de la izquierda se aproximan más a lo esperado de una distribución normal, en contraste de lo observado en el boxplot de la derecha.

6. Distribución e Histogramas

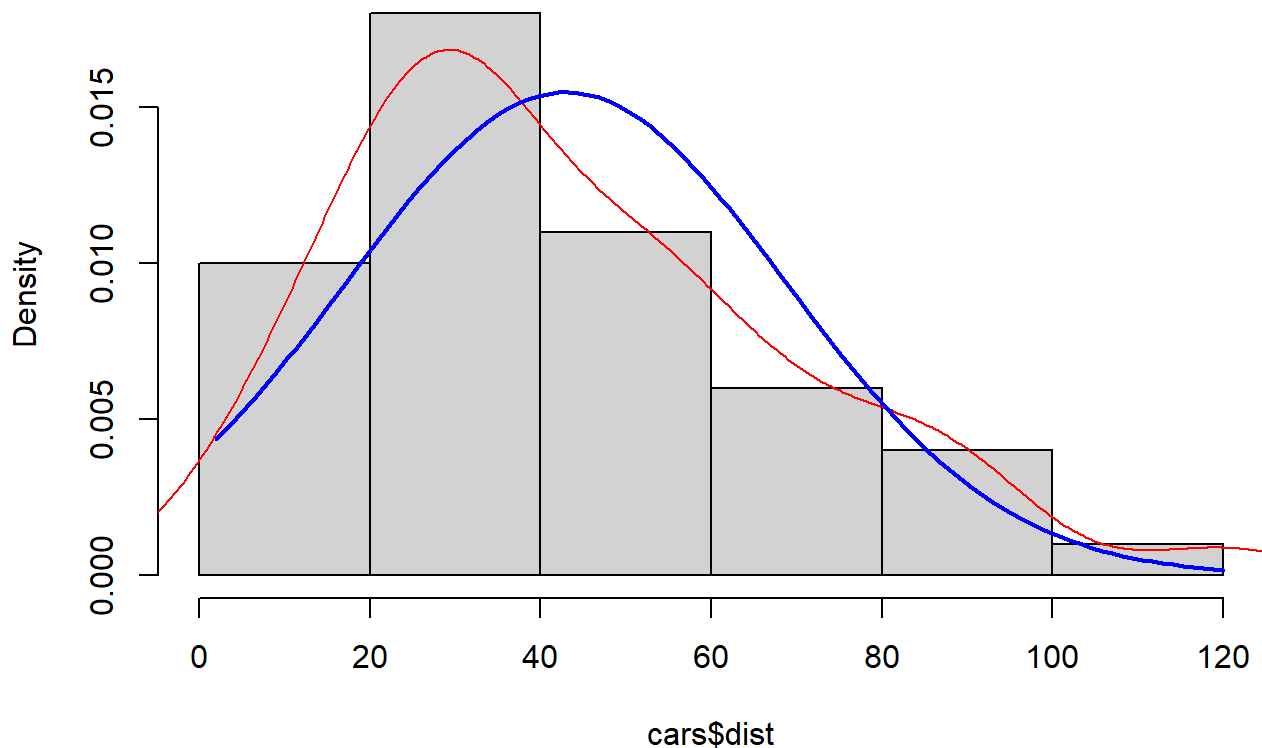
```
hist(cars$speed,freq=FALSE)
lines(density(cars$speed),col="red")
curve(dnorm(x,mean=mean(cars$speed),sd=sd(cars$speed)), from=min(cars$speed), to=max(cars$speed), add=TRUE, col="blue",lwd=2)
```

Histogram of cars\$dist



```
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), add=TRUE, col="blue",lwd=2)
```


Histogram of cars\$dist



En el histograma de la rapidez, se puede ver que la función de densidad de probabilidad esperada de la distribución actual se asemeja a la función de densidad de probabilidad de una distribución normal. Lo mismo no se puede decir para la distribución de la distancia, donde el sesgo a la derecha desvía por completo la forma de la distribución y no se alinea a lo esperado de la FDP de una normal.

7. Conclusiones

El análisis de normalidad univariada para las variables speed y dist del conjunto de datos cars sugiere que speed sigue una distribución normal, respaldado por pruebas estadísticas, QQPlots y un histograma que muestra una densidad de probabilidad cercana a la esperada. En contraste, la variable dist no cumple con los criterios de normalidad, presentando un sesgo positivo y una distribución que se aleja significativamente de la forma normal. En especial, la simetría en speed frente a la evidente asimetría en dist probablemente son la distinción clave entre ser considerado una distribución normal y no ser considerado una.