

# Actividad Integradora 1

Cleber Perez

2024-10-22

## Introducción

El diseño de obras hidráulicas, como presas derivadoras, requiere un análisis riguroso de los datos de precipitación para asegurar su eficacia y durabilidad ante eventos de lluvia extrema. El análisis de la precipitación extrema basado en periodos de retorno es fundamental en este contexto, ya que permite estimar la probabilidad de ocurrencia de eventos de magnitud significativa. Según fuentes confiables como el Centro Nacional de Prevención de Desastres (CENAPRED, 2019) y la Comisión Nacional del Agua (CONAGUA, 2022), el uso de un periodo de retorno adecuado es crucial para mitigar riesgos de inundación y proteger la infraestructura y las poblaciones cercanas a dichas obras.

Para nuestro caso se analizarán los datos históricos (1994-2023) de las precipitaciones máximas mensuales en Coahuila para cumplir el objetivo principal de este estudio que consiste en calcular la precipitación más extrema que se logra con un periodo de retorno seleccionado después de un análisis riguroso para poder saber su mejor ajuste para obtener resultados exitosos.

## Metodología

En este informe se documentarán todos los pasos realizados para justificar la selección del mejor ajuste de distribución para nuestro estado objetivo, Coahuila. La elección se fundamenta en la correcta caracterización de las precipitaciones máximas anuales, lo que permitirá desarrollar proyecciones más fiables en términos de probabilidad de excedencia y periodo de retorno. Iniciamos nuestro análisis ajustando el dataframe para que contenga únicamente los datos de Coahuila y, posteriormente, visualizamos las precipitaciones máximas registradas año con año mediante un gráfico inicial, que nos proporciona una visión general de las tendencias. En este gráfico observamos que, en los años más recientes, los valores de precipitación parecen significativamente elevados en comparación con años anteriores. Esta observación plantea la hipótesis de que los eventos de precipitación extrema han aumentado en frecuencia e intensidad en los últimos tiempos, lo cual merece un análisis más profundo.

Para sustentar esta hipótesis, procedemos a calcular valores de centralización, así como a construir un histograma y un diagrama de caja o boxplot de las precipitaciones. Estos gráficos permiten observar la dispersión de los datos y corroborar si las precipitaciones extremas realmente se han vuelto más frecuentes o si el aumento se debe a algún fenómeno puntual. La visualización de los valores de centralización, en particular la media y la mediana, nos ayuda a observar la tendencia central de los datos, mientras que el boxplot nos da una idea clara sobre los

valores atípicos y el rango intercuartílico, indicando si los valores extremos recientes son estadísticamente significativos o simplemente variaciones menores.

Una vez concluida la exploración inicial de los datos, nos enfocamos en dos gráficos que resultan esenciales para el análisis descriptivo del problema: el gráfico de probabilidad de excedencia y el gráfico de periodo de retorno. La probabilidad de excedencia representa la probabilidad de que un evento específico, como una precipitación extrema, iguale o supere un determinado valor en un año específico. En otras palabras, si en un año se observa que la probabilidad de excedencia para una precipitación de 100 mm es del 1%, esto significa que existe un 1% de probabilidad de que la precipitación anual sea igual o mayor a 100 mm en ese año. Este tipo de análisis es útil para comprender la frecuencia con la que se pueden esperar eventos extremos en el corto plazo y es fundamental para gestionar los riesgos asociados a eventos de precipitación alta.

Por otro lado, el periodo de retorno indica el intervalo de tiempo esperado entre eventos de una magnitud específica o mayor. Por ejemplo, una lluvia con un periodo de retorno de 100 años implica que, en promedio, dicho evento de precipitación extrema ocurrirá una vez cada 100 años. Sin embargo, es importante aclarar que esto no significa que tal evento sucederá solo una vez en un siglo. El periodo de retorno es una herramienta estadística basada en la probabilidad; por lo tanto, en cualquier año existe un 1% de probabilidad de que se presente un evento de precipitación igual o mayor al correspondiente a un periodo de retorno de 100 años. Este concepto es crucial en la planificación y diseño de infraestructuras de control y manejo de agua, tales como presas, diques y canales. Estas estructuras deben estar diseñadas para resistir eventos de precipitación excepcional, especialmente en el caso de obras hidráulicas críticas, donde se suelen utilizar periodos de retorno largos, como 100, 200 o incluso 500 años, para asegurar la durabilidad de las infraestructuras y minimizar el riesgo de fallos.

Al diseñar una obra hidráulica, es común establecer una probabilidad de excedencia baja con el objetivo de reducir al mínimo la probabilidad de que el evento de diseño se supere. Por lo tanto, es preferible seleccionar valores bajos de probabilidad de excedencia, como 1% o 0.5%, ya que indican que el evento de diseño tiene una baja probabilidad de ocurrencia en un año dado. En nuestro análisis observamos que, en la medida en que los datos se actualizan, esta probabilidad parece incrementarse, lo cual subraya la importancia de ajustar adecuadamente los modelos y distribuciones.

Para modelar y pronosticar tanto las precipitaciones máximas como los periodos de retorno, probamos seis distribuciones diferentes: Distribución Normal, Distribución Log-Normal, Distribución Exponencial, Distribución Gamma, Distribución Weibull y Distribución Gumbel. Con cada una de estas distribuciones, nos enfocamos en tres elementos clave para evaluar su adecuación al fenómeno observado: el histograma de los datos, la ojiva acumulativa y una prueba de normalidad. Este proceso es necesario porque, aunque algunas distribuciones parecían ajustarse visualmente a los datos, los valores obtenidos en la prueba de normalidad mostraban significativas desviaciones, sugiriendo que el modelo no era el ideal.

Una de las principales dificultades en este proceso de ajuste es que la mayoría de las distribuciones no cumplen adecuadamente con los criterios estadísticos esperados. Por ejemplo, en algunos casos los gráficos indicaban que la distribución parecía ajustarse bien a los datos observados, pero los valores  $p$  obtenidos en la prueba de normalidad eran demasiado bajos, lo que obligaba a rechazar la hipótesis nula de ajuste adecuado. En particular, la Distribución Log-Normal mostró una buena aproximación visual a los datos, aunque su valor  $p$  de 0.0003727 nos llevó a descartar este ajuste. Esto sugiere que, aunque la Distribución Log-Normal refleja en parte la tendencia de los datos, no ofrece la precisión estadística necesaria para nuestro propósito.

Después de evaluar las seis distribuciones, seleccionamos la Distribución Weibull como el ajuste más adecuado para modelar la precipitación extrema en Coahuila. La Distribución Weibull mostró una representación visual sólida y un valor  $p$  de 0.1924 en la prueba de normalidad, un valor considerablemente mejor en comparación con las demás opciones. Esto implica que la Weibull se acerca más a un comportamiento de los datos aceptable en términos de normalidad, haciendo de esta la mejor candidata para realizar las proyecciones de periodo de retorno de manera confiable.

Con la distribución elegida, procedemos a realizar los cálculos finales de excedencia y periodo de retorno usando la fórmula correspondiente. Esta fórmula nos permite proyectar los niveles de precipitación que podríamos esperar en diferentes horizontes de tiempo, facilitando una evaluación más precisa de los riesgos. Al analizar los resultados obtenidos, podemos observar que, a medida que el periodo de retorno aumenta, también lo hace el valor de la precipitación máxima estimada. Este comportamiento es consistente con la naturaleza de los eventos extremos, que son menos frecuentes pero más intensos, y subraya la importancia de utilizar periodos de retorno adecuados para garantizar la seguridad y resiliencia de las infraestructuras hidráulicas.

En conclusión, este análisis exhaustivo nos ha permitido seleccionar la Distribución Weibull como la mejor opción para caracterizar y proyectar las precipitaciones máximas anuales en Coahuila, basándonos en los datos históricos. Este ajuste proporciona una herramienta fundamental para diseñar infraestructuras que puedan soportar eventos de precipitación extremos en un contexto de cambio climático, donde es probable que dichos eventos se vuelvan más frecuentes e intensos en el futuro.

## **Discusion y conclusiones**

El análisis realizado confirma que la distribución Weibull es adecuada para modelar la precipitación extrema en el estado de nuestra elección que fue Coahuila. Nuestro análisis permite para un periodo de retorno de 200 años representa un valor seguro y práctico para el diseño de la presa derivadora en esta zona de riego, minimizando riesgos de fallos ante eventos extremos. Este trabajo subraya la importancia de usar distribuciones de probabilidad precisas y periodos de retorno adecuados para una planificación hidráulica efectiva.

## Referencias bibliográficas

CENAPRED (2019). Análisis de Riesgos por Inundaciones en México.

CONAGUA (2022). Lineamientos Técnicos para el Diseño Hidráulico de Obras de Infraestructura.

## R Markdown

### Lectura de archivo

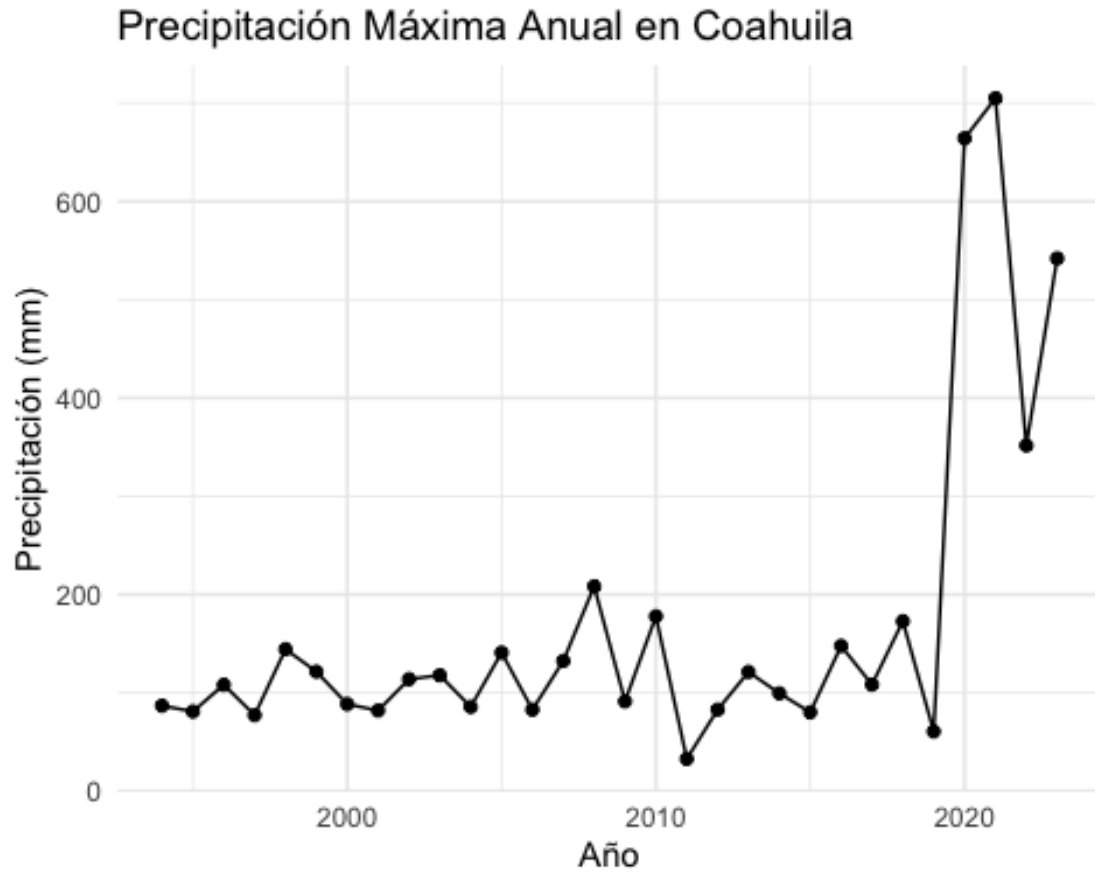
##	Anio	Mes	Estado	Lluvia
## 1	1994	Ene	Aguascalientes	8.3
## 2	1994	Ene	Baja.California	10.3
## 3	1994	Ene	Baja.California.Sur	0.0
## 4	1994	Ene	Campeche	85.4
## 5	1994	Ene	Ciudad.de.México	17.7
## 6	1994	Ene	Coahuila	12.8
## 7	1994	Ene	Colima	0.0
## 8	1994	Ene	Chiapas	46.1
## 9	1994	Ene	Chihuahua	1.2
## 10	1994	Ene	Durango	3.5

### Seleccion = (Coahuila)

##	Anio	Mes	Estado	Lluvia
## 6	1994	Ene	Coahuila	12.8
## 39	1994	Feb	Coahuila	3.1
## 72	1994	Mar	Coahuila	10.2
## 105	1994	Abr	Coahuila	2.6
## 138	1994	May	Coahuila	8.5
## 171	1994	Jun	Coahuila	86.8
## 204	1994	Jul	Coahuila	50.7
## 237	1994	Ago	Coahuila	23.4
## 270	1994	Sep	Coahuila	56.2
## 303	1994	Oct	Coahuila	31.9

## 1. Análisis estadístico descriptivo de las precipitaciones históricas máximas mensuales de un estado

### Grafica Coahuila



### Medias de centralizacion y variacion

## [1] 47.90861

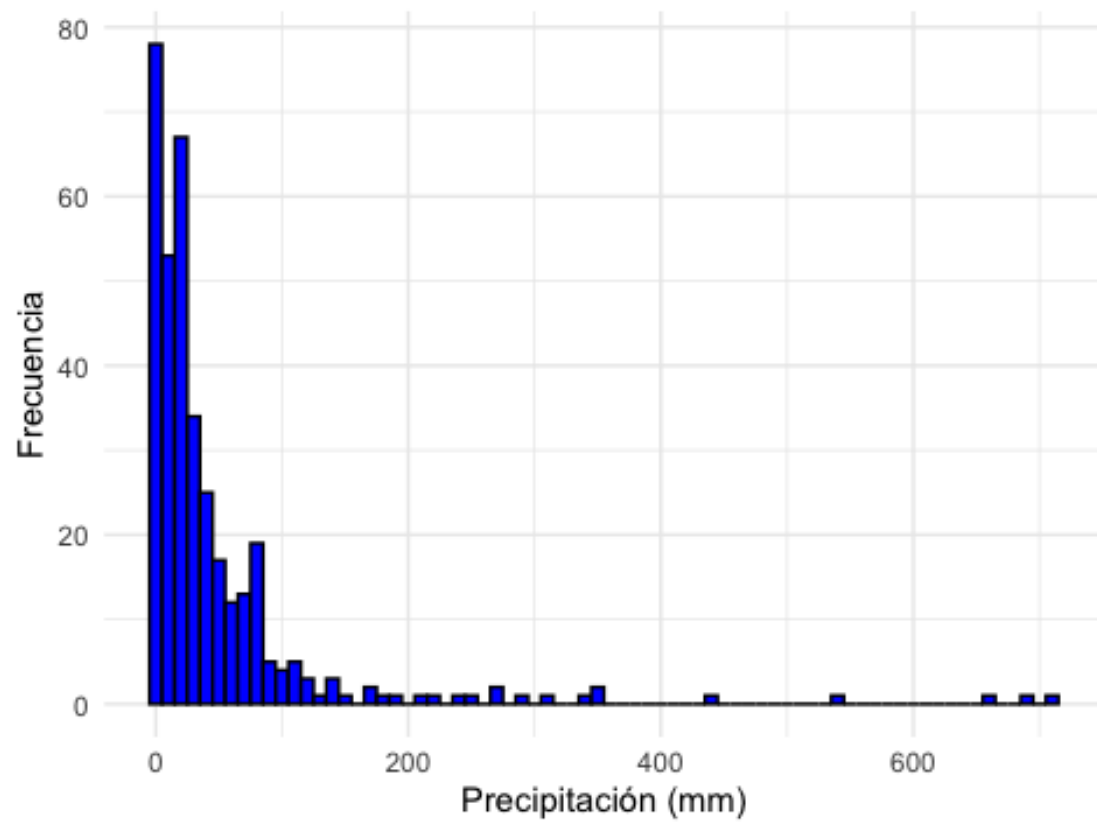
## [1] 22.45

## [1] 87.77857

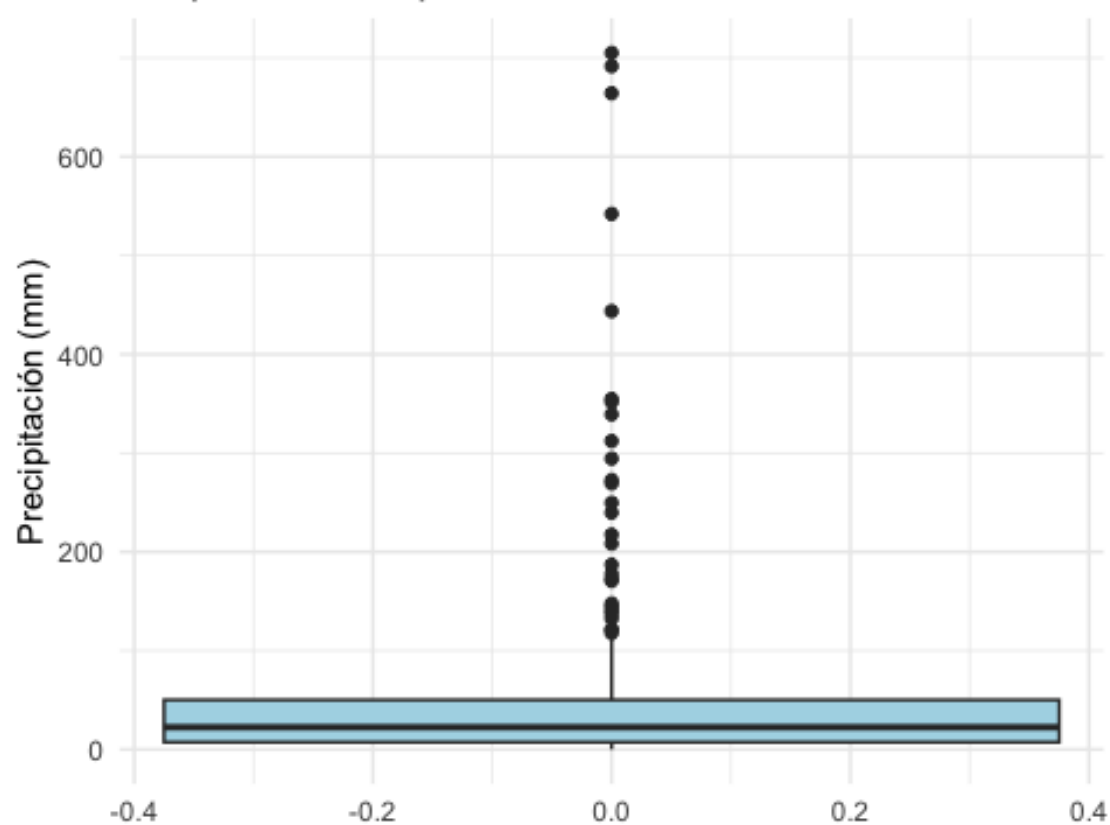
## [1] 0.0 705.1

## Histograma y Boxplot

Histograma de Precipitaciones Máximas Mensuales en l



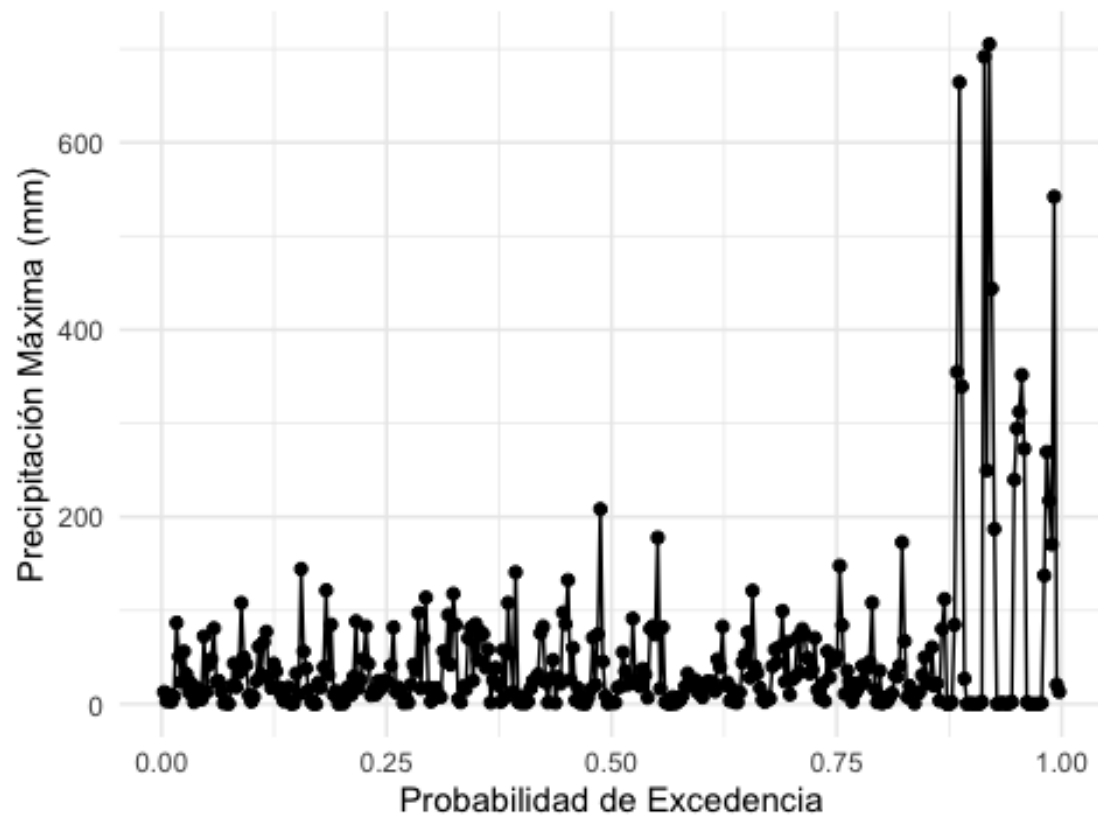
Boxplot de Precipitaciones Máximas Mensuales en Coa

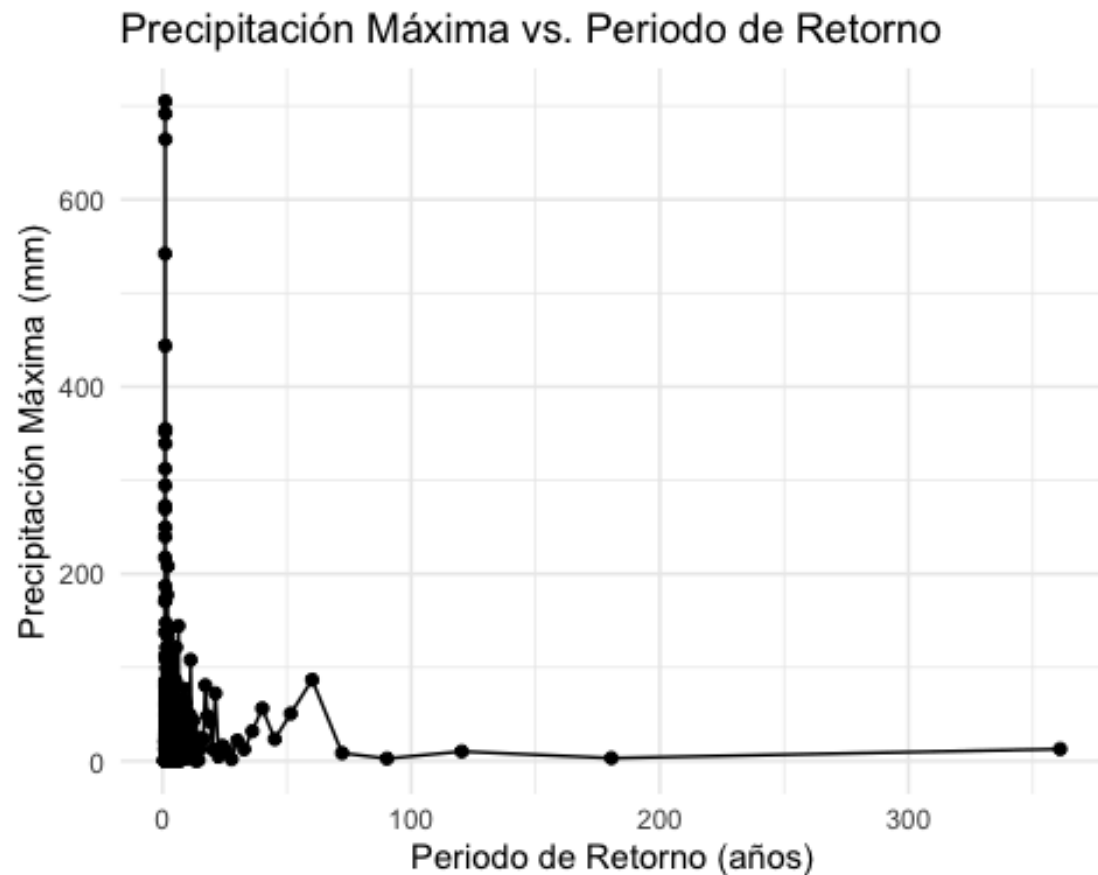




## 2. Análisis de Frecuencias Método Gráfico

Precipitación Máxima vs. Probabilidad de Excedencia





### 3. Análisis de Frecuencias Método Analítico

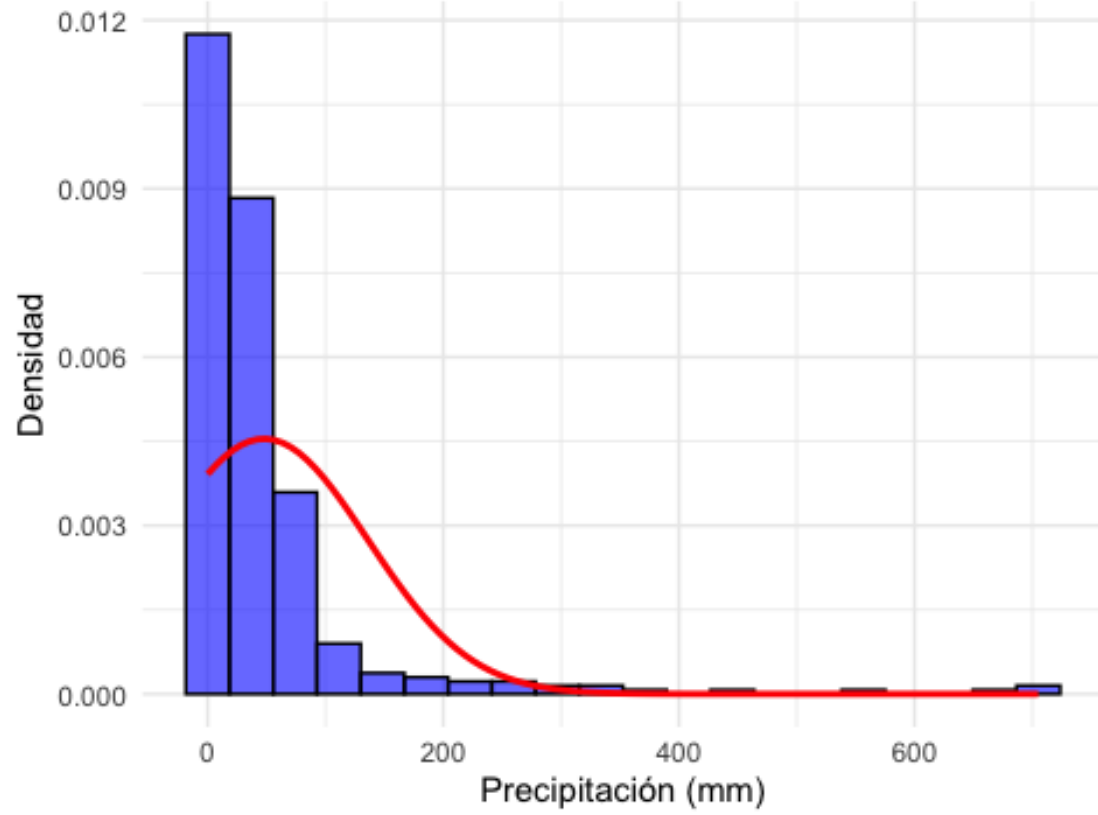
#### Ajuste a una Distribución Normal.

#### Histograma de la función de densidad empírica

```
## 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.

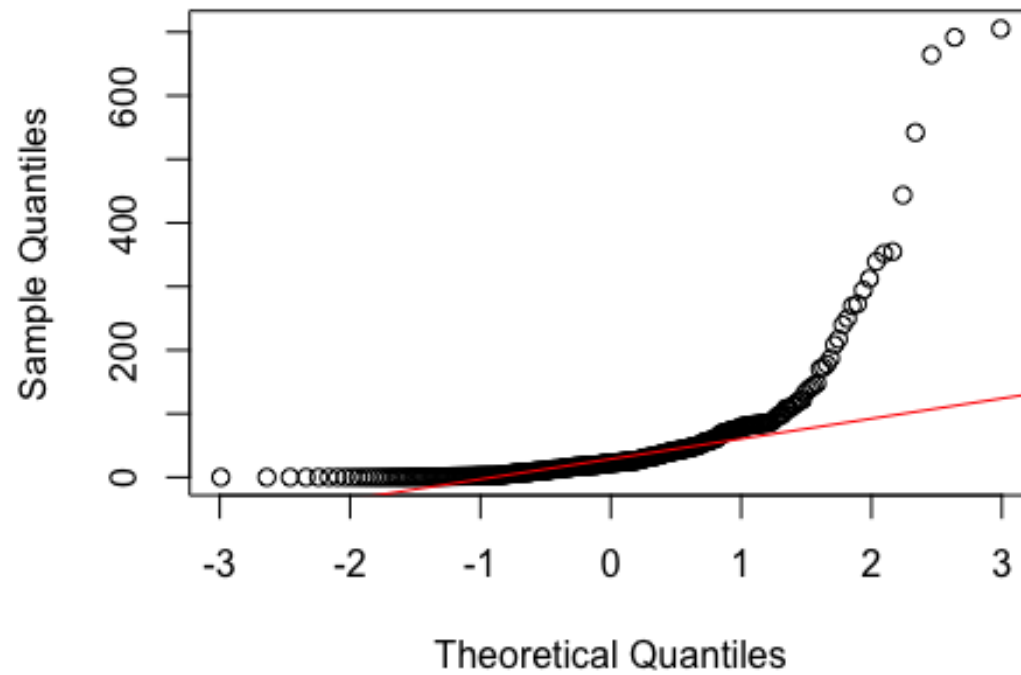
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2
## 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Histograma de Densidad Empírica y Distribución Norn



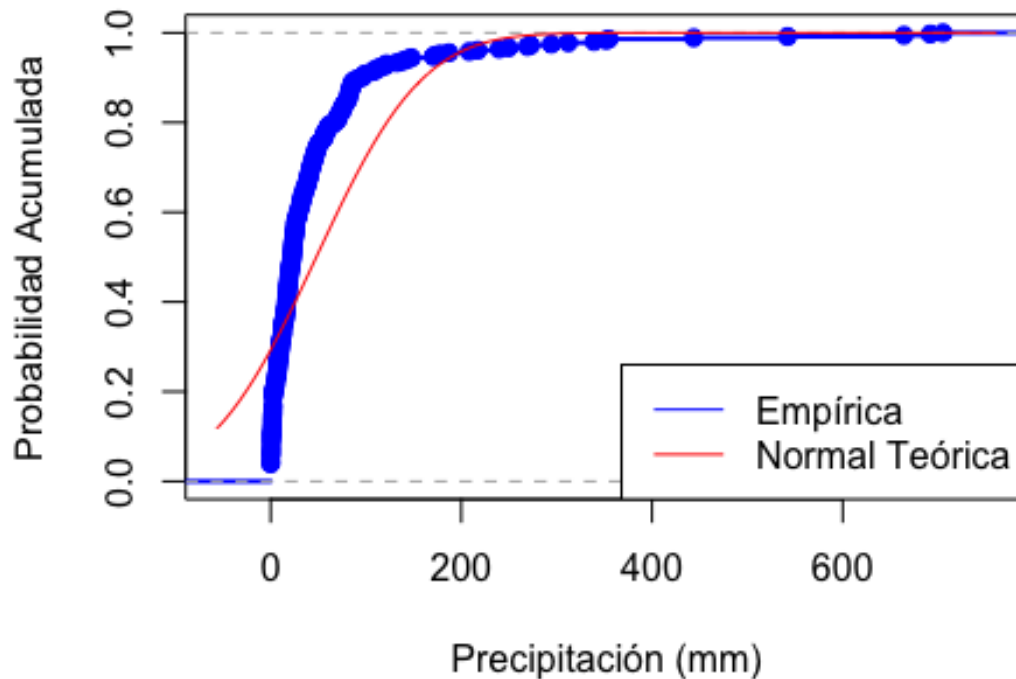
## Grafica QQplot

### Q-Q Plot para la Distribución Normal



## Ojiva

### Ojiva de Distribuciones Empírica y Normal Teórica

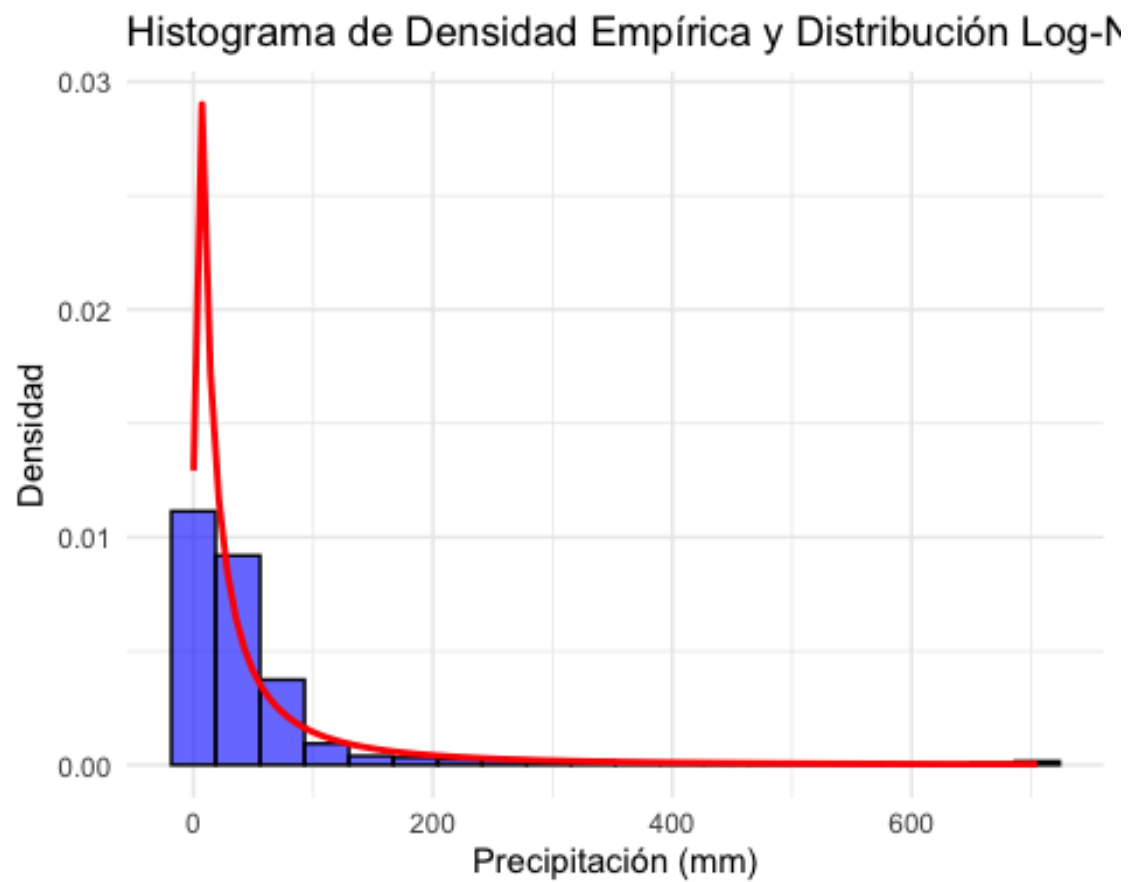


### Pruebas de bondad de ajuste

```
## Warning in ks.test.default(coahuila_data$Lluvia, "pnorm", mean =  
mean_rain, :  
## ties should not be present for the one-sample Kolmogorov-Smirnov test  
  
##  
## Shapiro-Wilk normality test  
##  
## data: coahuila_data$Lluvia  
## W = 0.49731, p-value < 2.2e-16  
  
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: coahuila_data$Lluvia  
## D = 0.29261, p-value < 2.2e-16  
## alternative hypothesis: two-sided
```

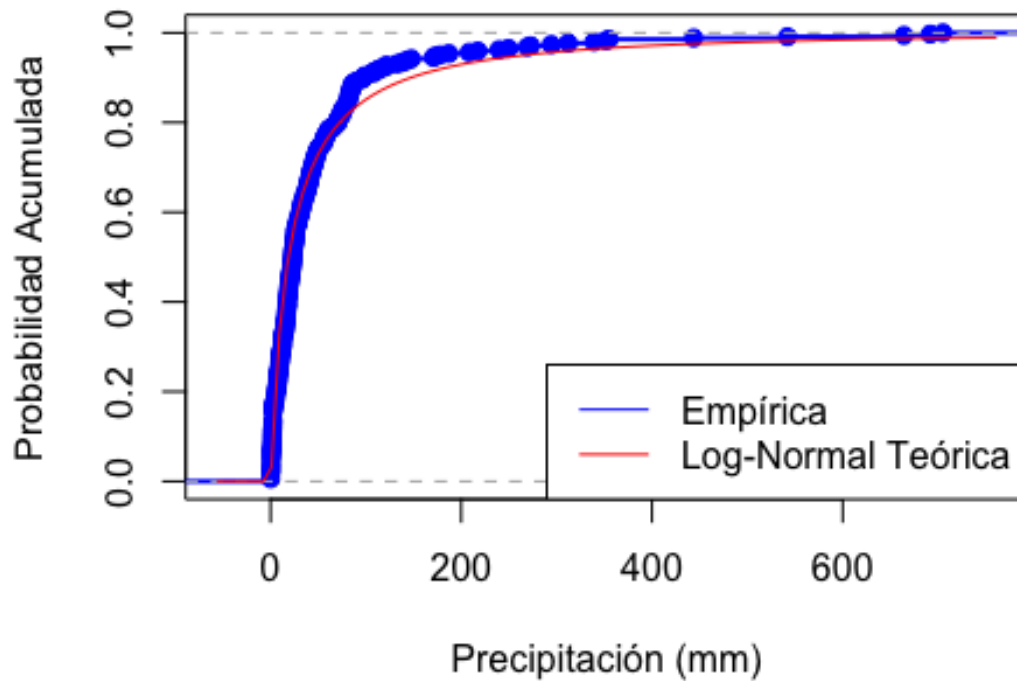
Ajuste a una Distribución Log-Normal.

Histograma de la función de densidad empírica



## Ojiva

### Ojiva de Distribuciones Empírica y Log-Normal Teórica



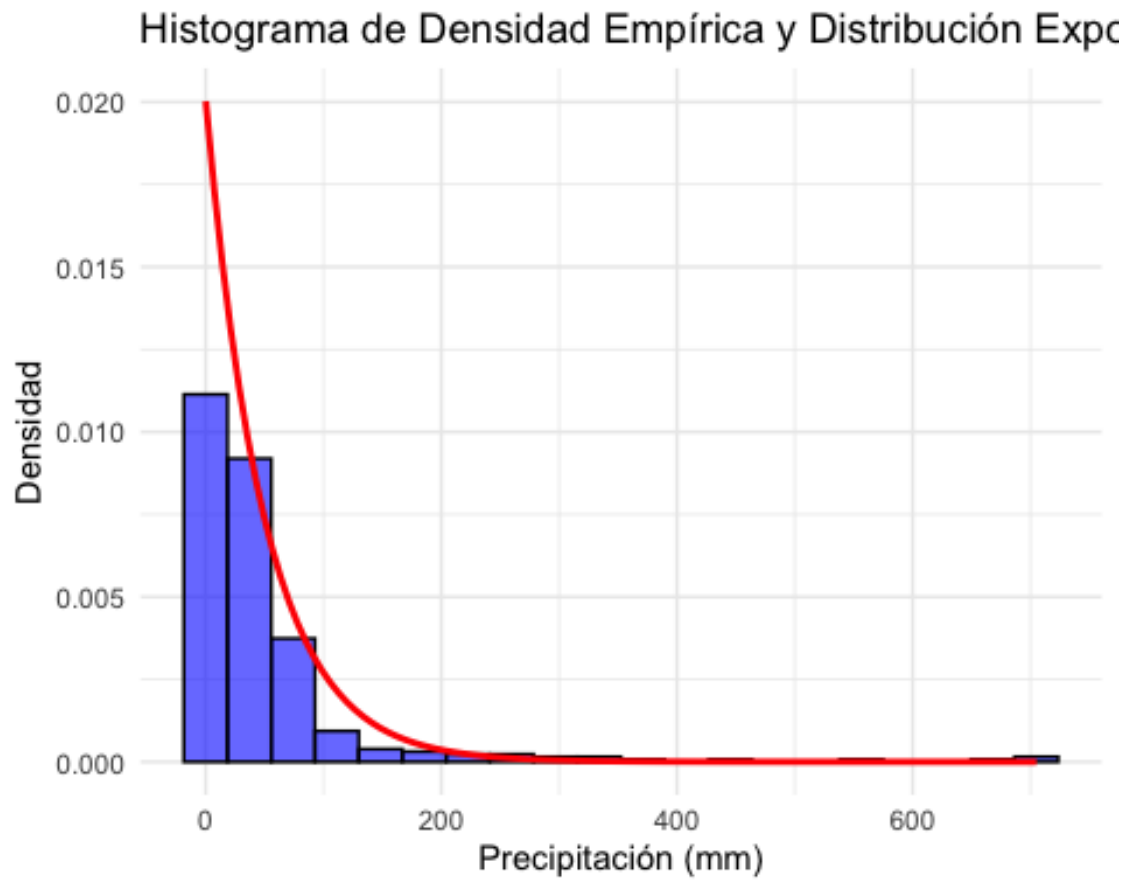
### Pruebas de bondad de ajuste

```
## Warning in ks.test.default(coahuila_data, "plnorm", meanlog = log_mean, :  
ties  
## should not be present for the one-sample Kolmogorov-Smirnov test  
  
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: coahuila_data  
## D = 0.1114, p-value = 0.0003727  
## alternative hypothesis: two-sided
```



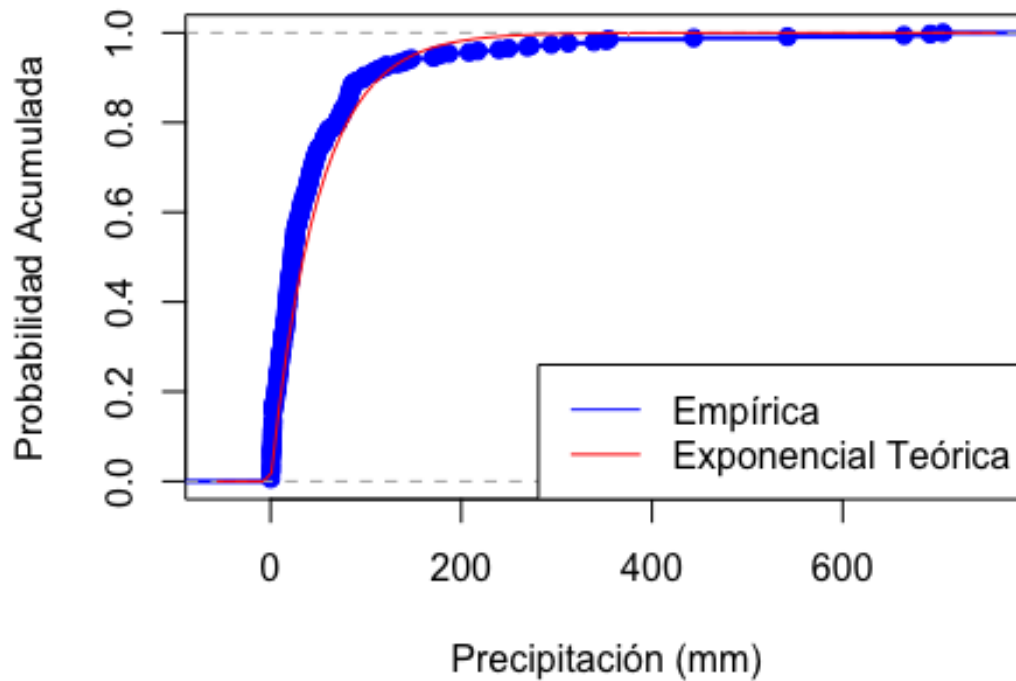
## Ajuste a una Distribución Exponencial.

### Histograma de la función de densidad empírica



## Ojiva

### Ojiva de Distribuciones Empírica y Exponencial Teórica



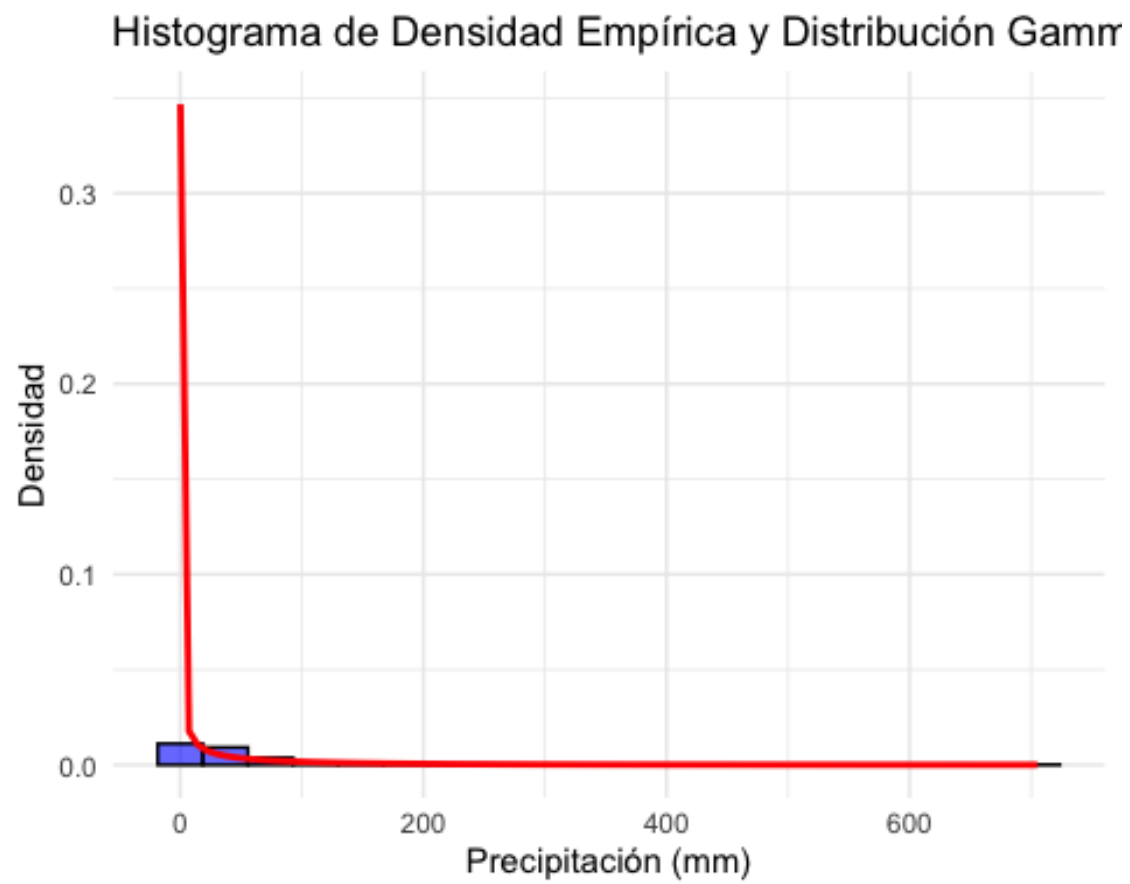
### Pruebas de bondad de ajuste

```
## Warning in ks.test.default(coahuila_data, "pexp", rate = lambda): ties
## should
## not be present for the one-sample Kolmogorov-Smirnov test

##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: coahuila_data
## D = 0.15123, p-value = 2.676e-07
## alternative hypothesis: two-sided
```

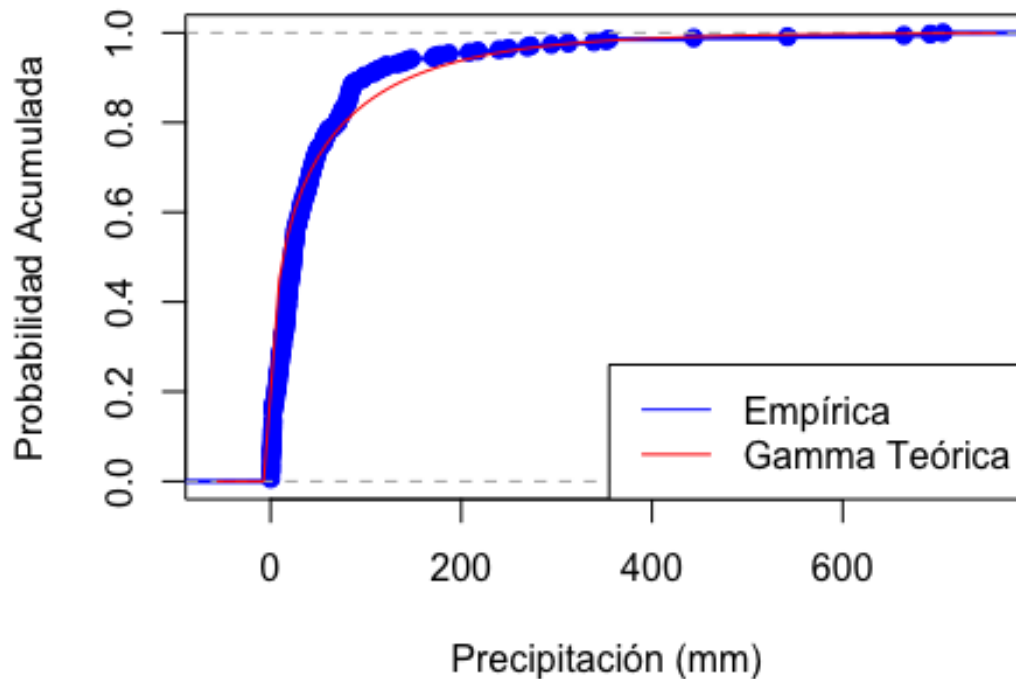
Ajuste a una Distribución Gamma.

Histograma de la función de densidad empírica



## Ojiva

### Ojiva de Distribuciones Empírica y Gamma Teórica



### Pruebas de bondad de ajuste

```
## Warning in ks.test.default(coahuila_data, "pgamma", shape = shape, scale =  
## scale): ties should not be present for the one-sample Kolmogorov-Smirnov  
test
```

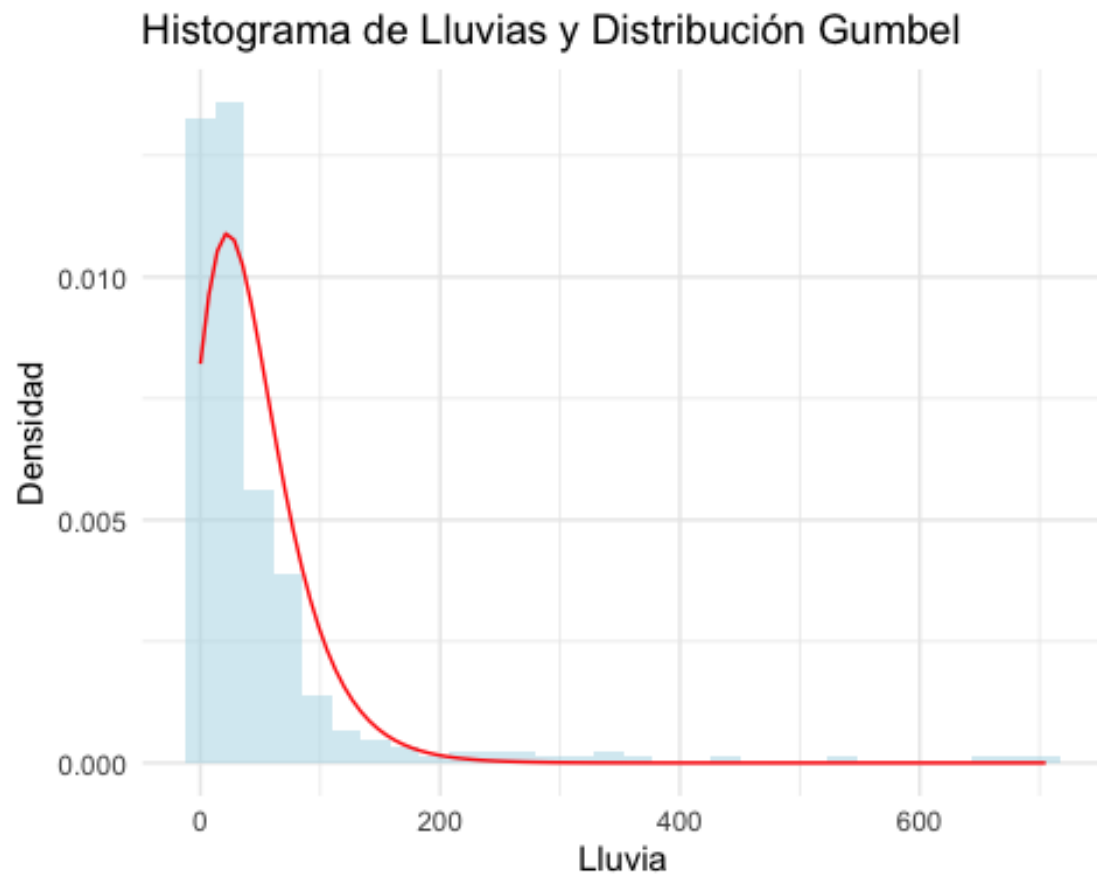
```
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: coahuila_data  
## D = 0.20614, p-value = 3.392e-13  
## alternative hypothesis: two-sided
```

### Ajuste a una Distribución Gumbel.

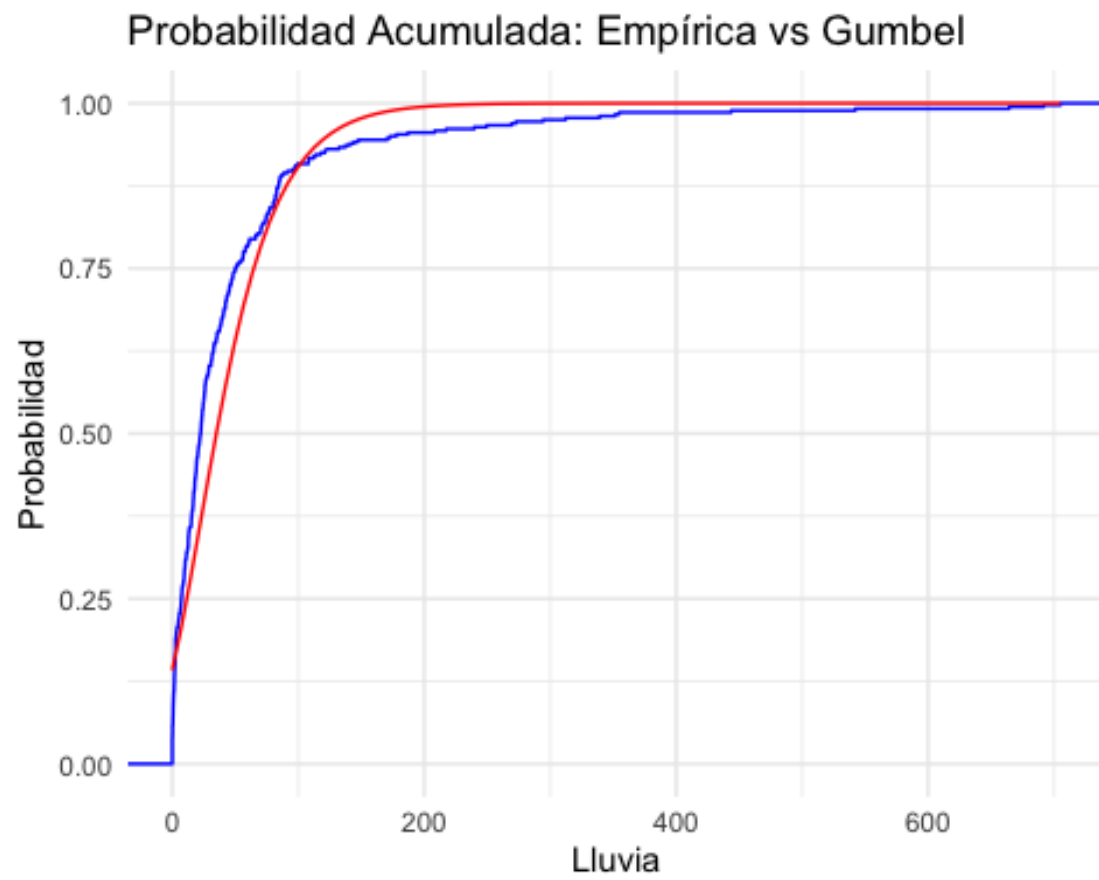
```
## Loading required package: MASS  
  
## Loading required package: survival  
  
## Fitting of the distribution ' gumbel ' by maximum likelihood  
## Parameters :  
## estimate Std. Error  
## location 22.57356 1.822281
```

```
## scale      33.75029    1.588634
## Loglikelihood: -1897.148    AIC: 3798.297    BIC: 3806.069
## Correlation matrix:
##           location      scale
## location 1.0000000 0.2175979
## scale    0.2175979 1.0000000
```

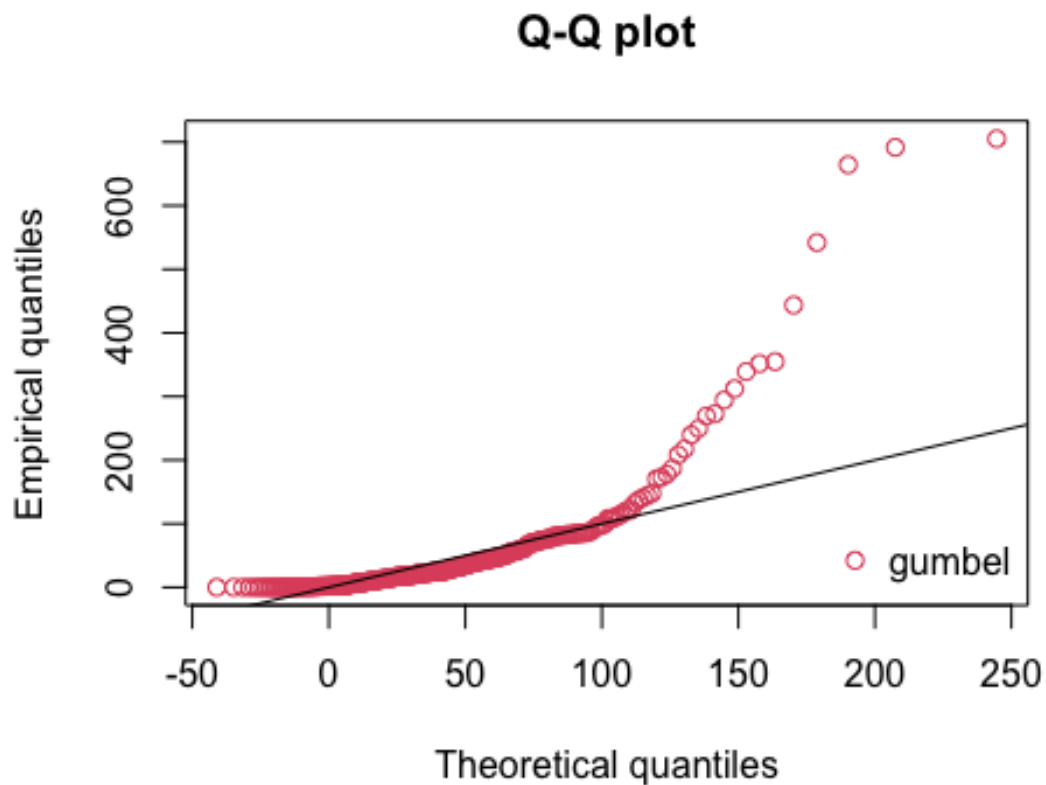
### Histograma de la función de densidad empírica



## Ojiva



## QQplot



## Pruebas de bondad de ajuste

```
## Warning in ks.test.default(coahuila_data$Lluvia, "pgumbel", location =  
## gumbel_fit$estimate[1], : ties should not be present for the one-sample  
## Kolmogorov-Smirnov test  
  
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: coahuila_data$Lluvia  
## D = 0.16997, p-value = 1.85e-09  
## alternative hypothesis: two-sided
```

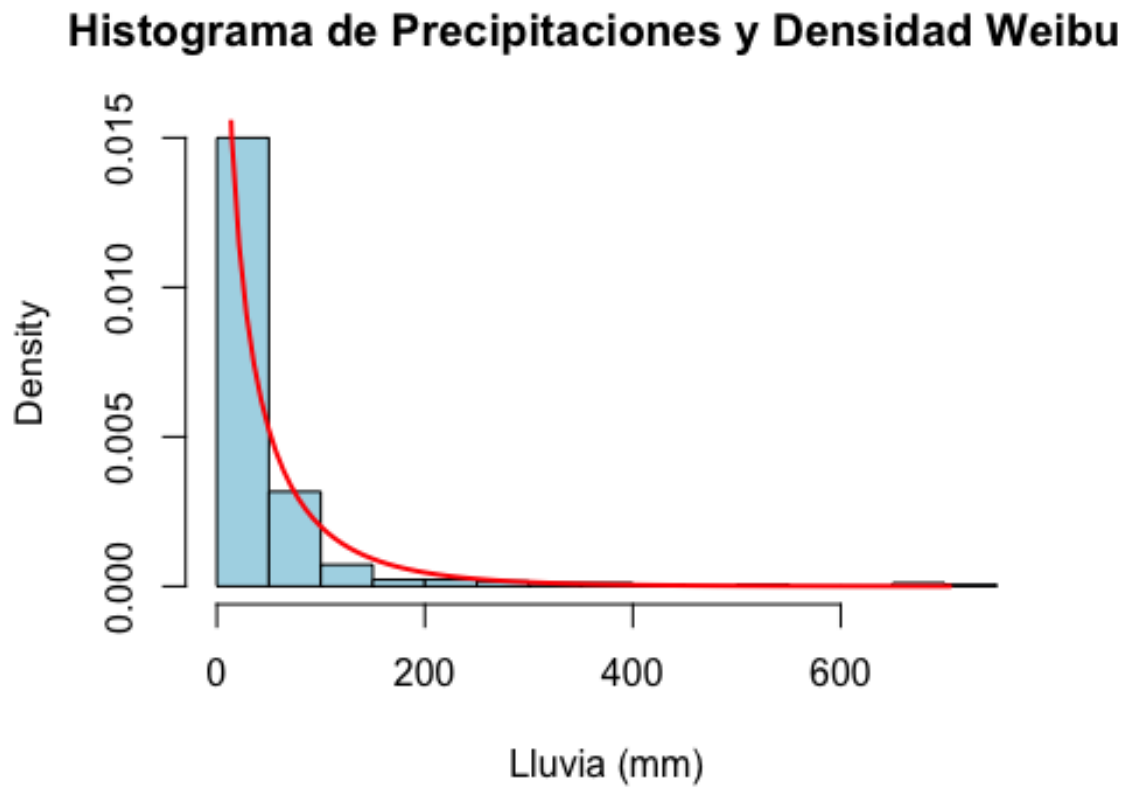
## Including Plots

```
## Fitting of the distribution ' weibull ' by maximum likelihood  
## Parameters :  
##      estimate Std. Error  
## shape  0.7230485 0.02886011  
## scale 39.4720063 3.09906049  
## Loglikelihood: -1657.975 AIC: 3319.95 BIC: 3327.643  
## Correlation matrix:  
##      shape      scale
```

```
## shape 1.0000000 0.3213667
## scale 0.3213667 1.0000000
```

Ajuste a una Distribución Weibull.

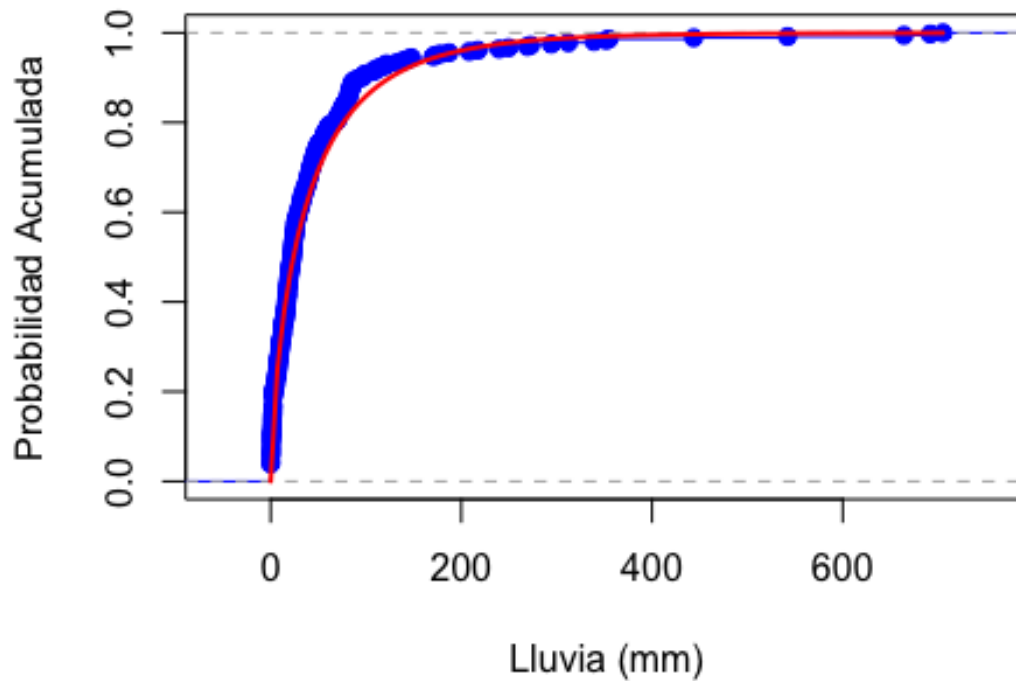
Histograma de la función de densidad empírica





## Ojiva

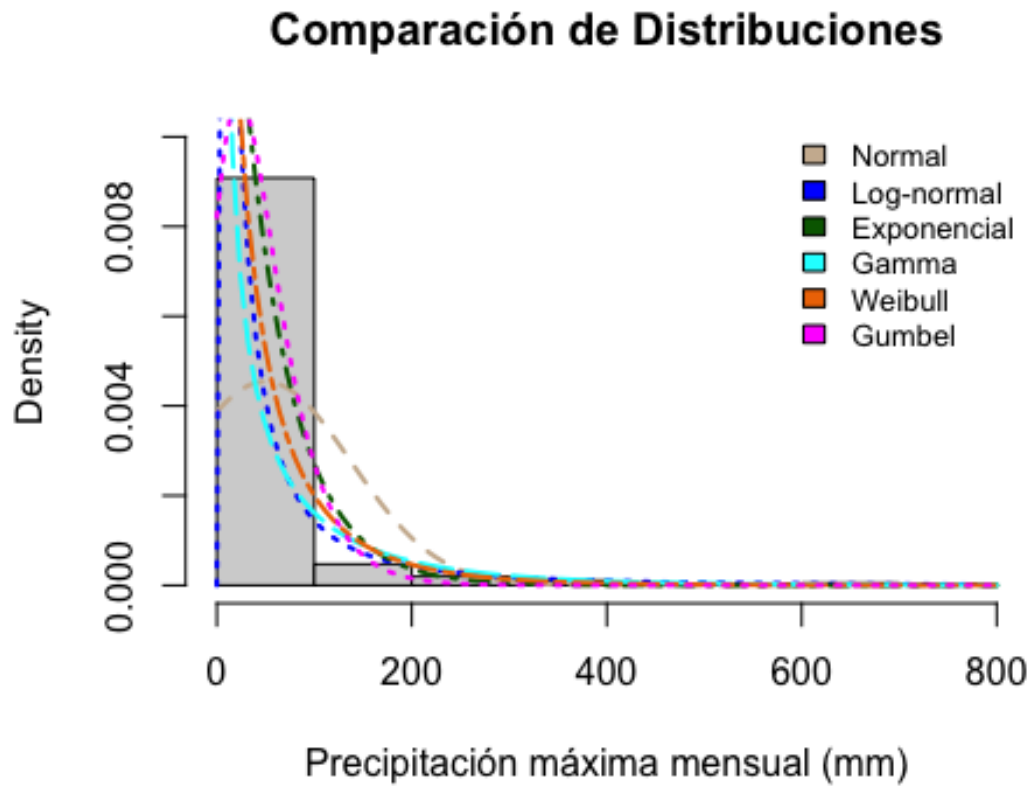
### Distribución Acumulada Empírica vs Teórica



### Pruebas de bondad de ajuste

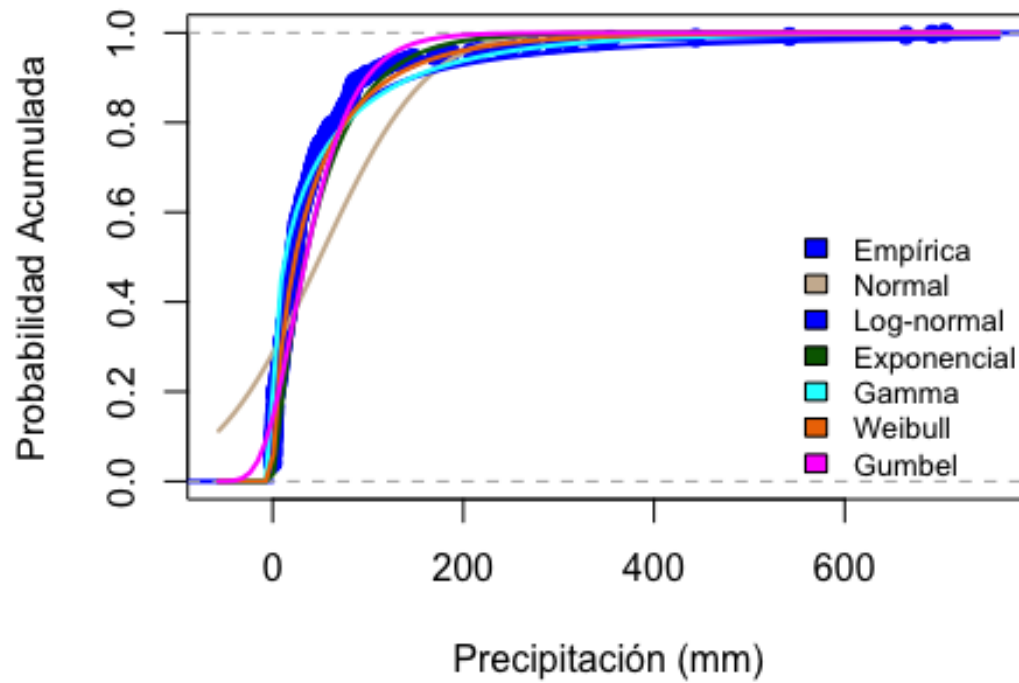
```
## Warning in ks.test.default(coahuila_data$Lluvia, "pweibull", shape =  
## weibull_fit$estimate[1], : ties should not be present for the one-sample  
## Kolmogorov-Smirnov test  
  
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: coahuila_data$Lluvia  
## D = 0.062635, p-value = 0.1186  
## alternative hypothesis: two-sided
```

## Compara los ajustes de las distribuciones

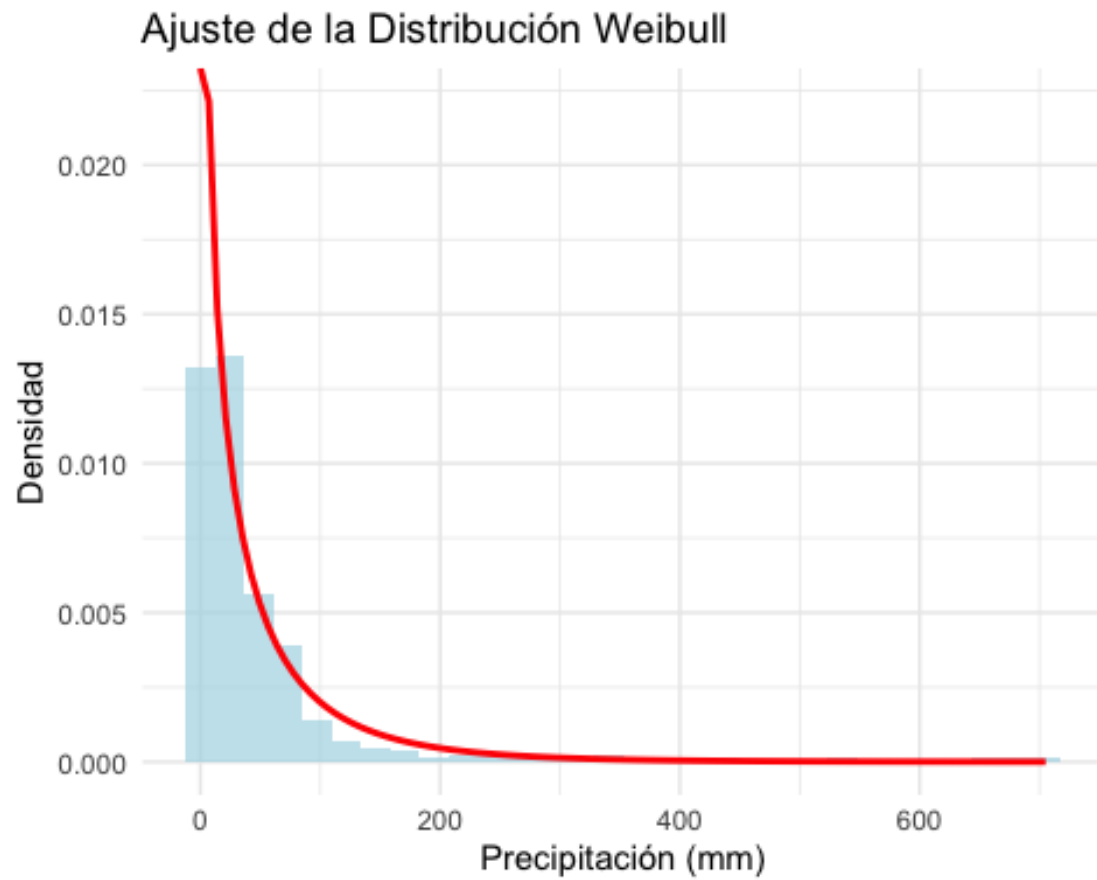


## Including Plots

## Probabilidad Acumulada: Empírica vs Teórica



#### 4. Precipitación de diseño de obras hidráulicas



```
## [1] 396.0919
```

```
## La precipitación máxima mensual esperada para un periodo de retorno de 200  
años es: 396.0919 mm
```