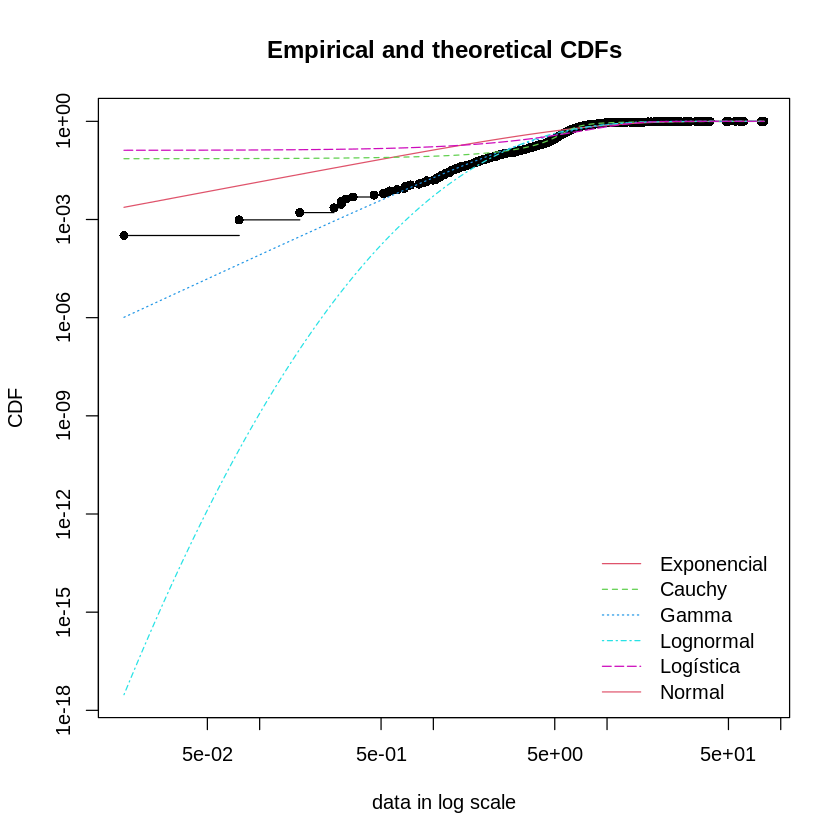
Grupo 4.

Integrantes:

* Luisa María De La Hortúa Pulido 201712785
* David Santiago Moreno Medina 201728270

1. Al llevar a cabo un análisis de posibles distribuciones adecuadas para los datos completos, se observa que ninguna de las distribuciones enumeradas en la tabla a continuación logra capturar el comportamiento completo de los datos. Esta discrepancia se refleja en los valores p obtenidos a través de la prueba de Chi-cuadrado y la decisión para la prueba de KS para cada una de las distribuciones siguientes:

| **Distribución** | **valor p Chi-cuadrado** | **Decisión prueba KS** |
| --- | --- | --- |
| Exponencial | 1.530121\*e^(-320) | Rechazado |
| Logística | 0 | Rechazado |
| Cauchy | 4.123016\*e^(-40) | Rechazado |
| Gamma | 8.616366\*e^(-138) | Rechazado |
| Normal | 0 | Rechazado |
| Lognormal | 3.78984\*e^(-124) | Rechazado |



Teniendo en cuenta la tabla y la grafica presentadas anteriormente, se puede concluir que ninguna de las distribuciones conocidas es apropiada para los datos y no logran abarcar el comportamiento completo de los datos cuando se consideran todas las franjas en conjunto.

Para determinar si las franjas se comportan diferente entre ellas, se llevaron a cabo pruebas de diferencias de varianzas y de diferencias de medias entre las franjas presentes en la base de datos, donde se obtuvieron los siguientes resultados:

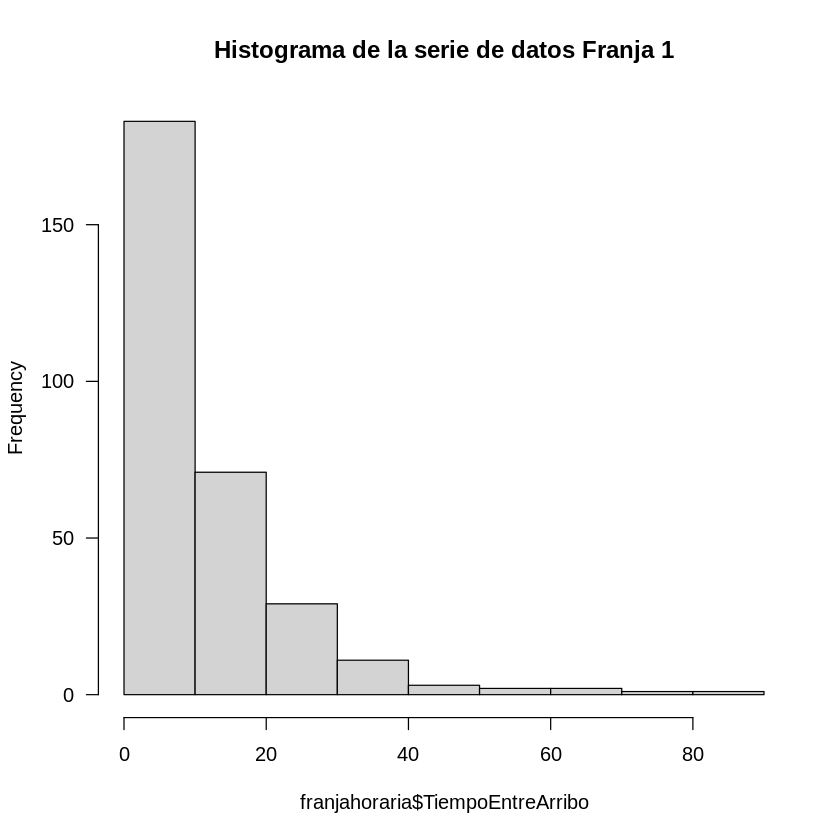
| **Franjas** | **Prueba F para varianzas** | **Prueba t para medias** |
| --- | --- | --- |
| 1-2 | F = 22.736, p-value < 2.2\*e^(-16) | t = 8.7192, p-value < 2.2\*e^(-16) |
| 2-3 | F = 6.8362, p-value < 2.2\*e^(-16) | t = -3.2966, p-value = 0.00102 |
| 1-3 | F = 155.43, p-value < 2.2\*e^(-16) | t = 8.284, p-value = 3.847\*e^(-15) |

Esto muestra que, estadísticamente, ninguna de las franjas comparte una varianza o media en común con las otras dos franjas, ya que son significativamente distintas. Lo que justifica un análisis individual para comprender mejor el comportamiento de cada una.

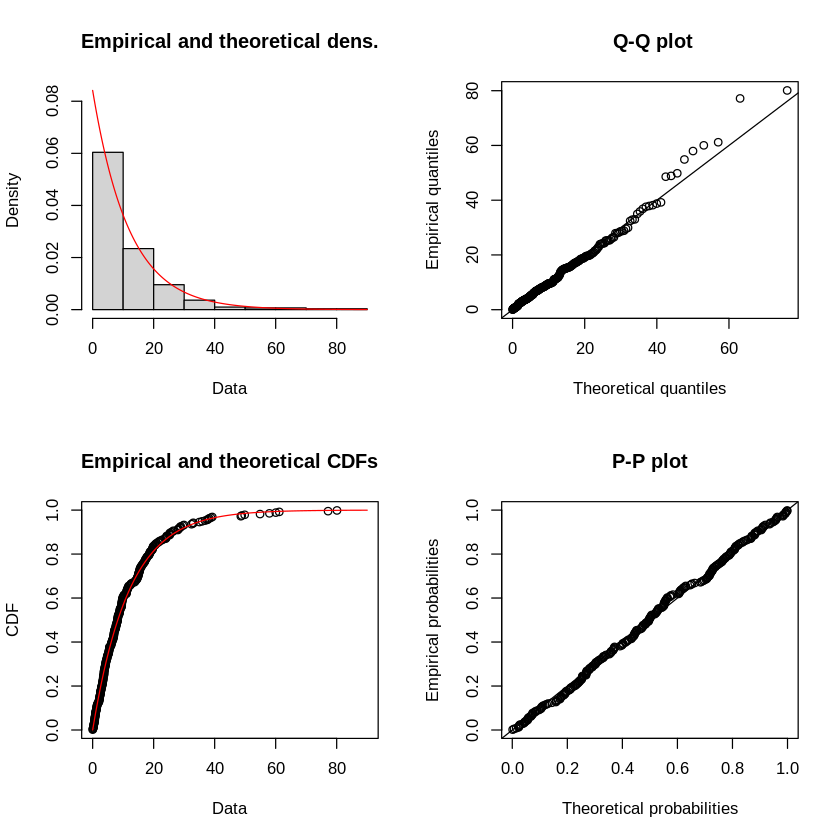
Cuando se examina cada franja por separado, se pueden llegar a las siguientes conclusiones:

* Franja 1:

La franja 1 muestra la siguiente distribución para el tiempo de arribo.



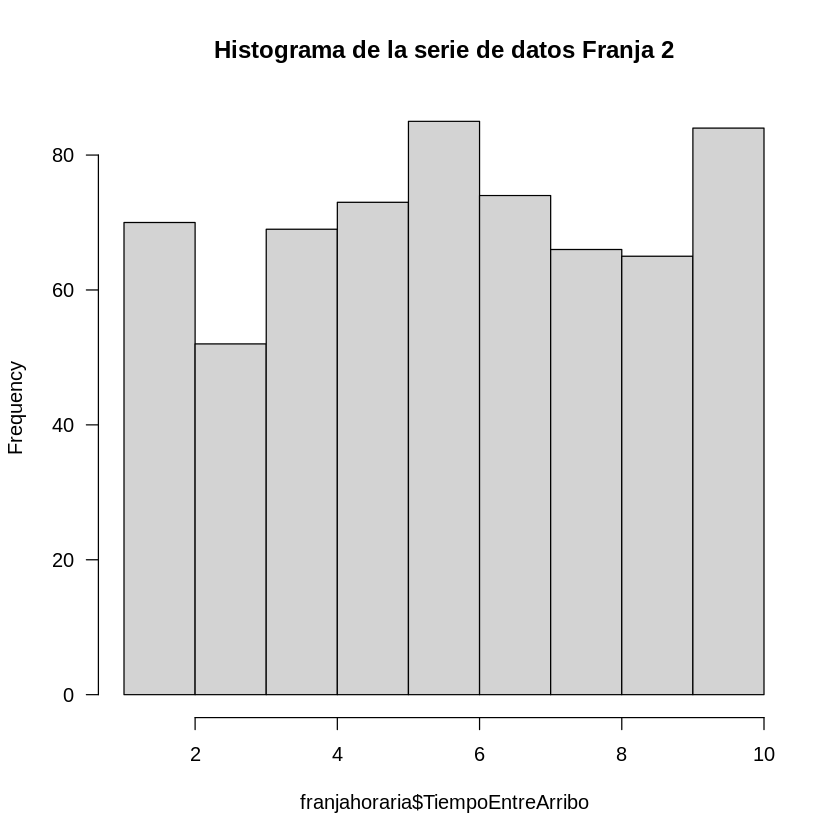
Donde se muestra que los datos están muy concentrados en los valores de 0 a 10 y presentan una cola hacia la derecha. Para esta franja se ajustaron 6 distribuciones (Exponencial, Gamma, Normal, Lognormal, Weibull y Cauchy), de las cuales la distribución Exponencial es la que más se parece a la distribución de los datos para la franja 1, mostrando los siguientes gráficos:



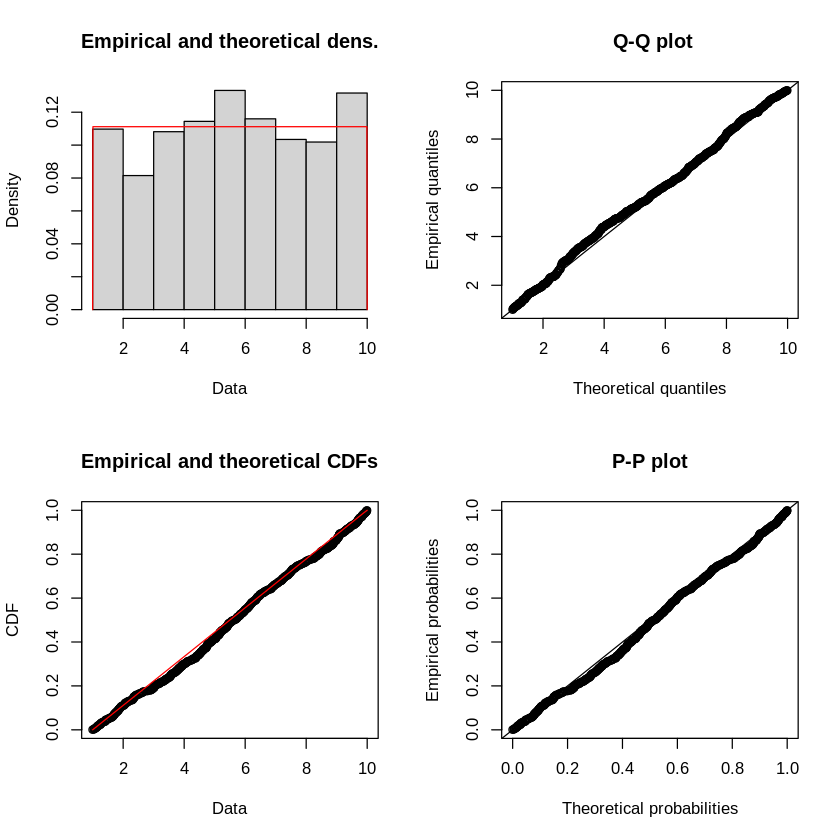
Se evidencia como en el Q-Q plot el ajuste es muy bueno, exceptuando unos pocos datos en la cola derecha de la distribución. Pero las probabilidades en el P-P plot muestran que es una distribución adecuada para los datos. También las pruebas formales como KS y chi-cuadrado muestran que en la dos no se rechaza la hipótesis nula de que los datos del tiempo arribo para la franja 1 provienen de una distribución exponencial.

* Franja 2:

La franja 2 muestra la siguiente distribución para el tiempo de arribo.



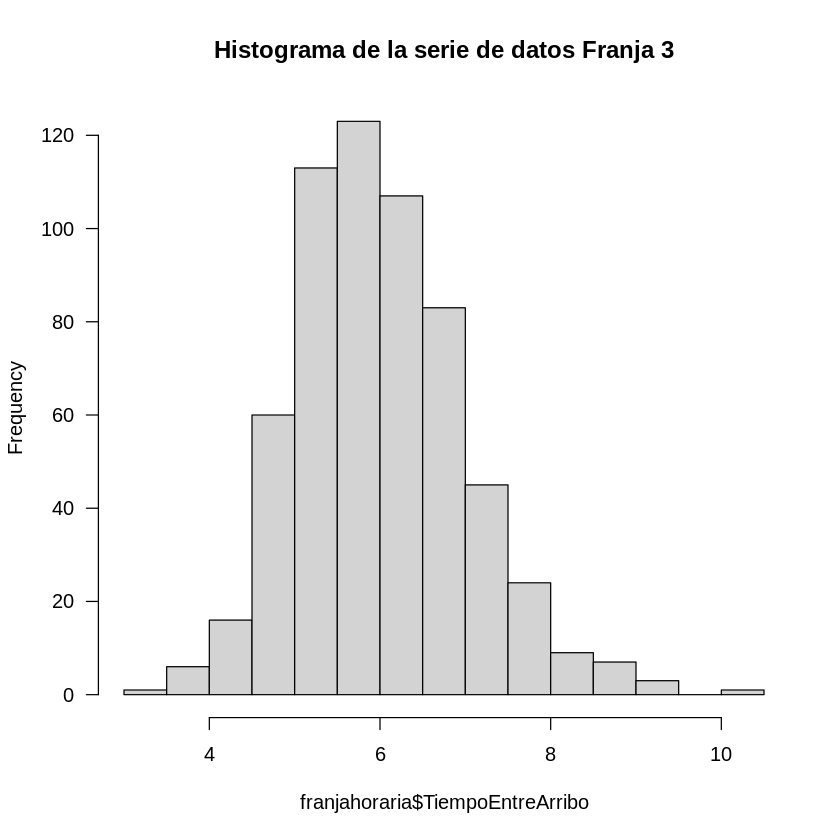
Donde se muestra que los datos están muy bien distribuidos en cada uno de los tiempos. Por lo tanto, se puede llegar a pensar de manera inicial que estos datos pueden provenir de una distribución uniforme. Para esta franja se ajustaron 7 distribuciones (Uniforme, Exponencial, Gamma, Normal, Lognormal, Weibull y Cauchy), de las cuales la distribución Uniforme es la que más se parece a la distribución de los datos para la franja 2, mostrando los siguientes resultados:



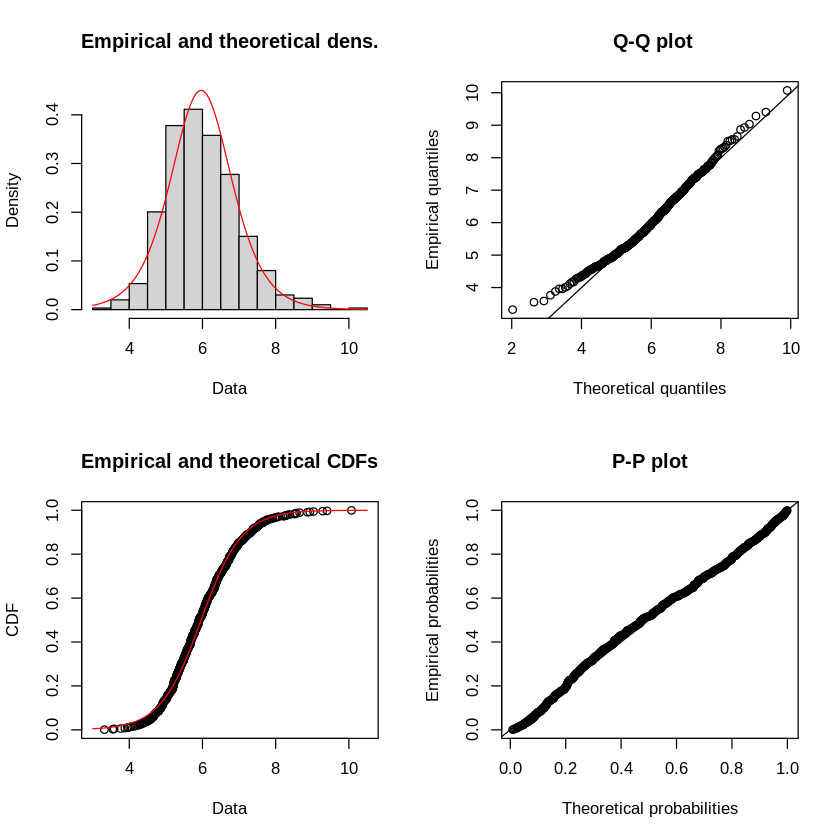
Se evidencia como en el Q-Q plot el ajuste es muy bueno. En cada uno de los otros gráficos muestran que es una distribución adecuada para los datos. También las pruebas formales como KS y chi-cuadrado muestran que en la dos no se rechaza la hipótesis nula de que los datos de arribo para la franja 2 provienen de una distribución Uniforme.

* Franja 3:

La franja 3 muestra la siguiente distribución para el tiempo de arribo.

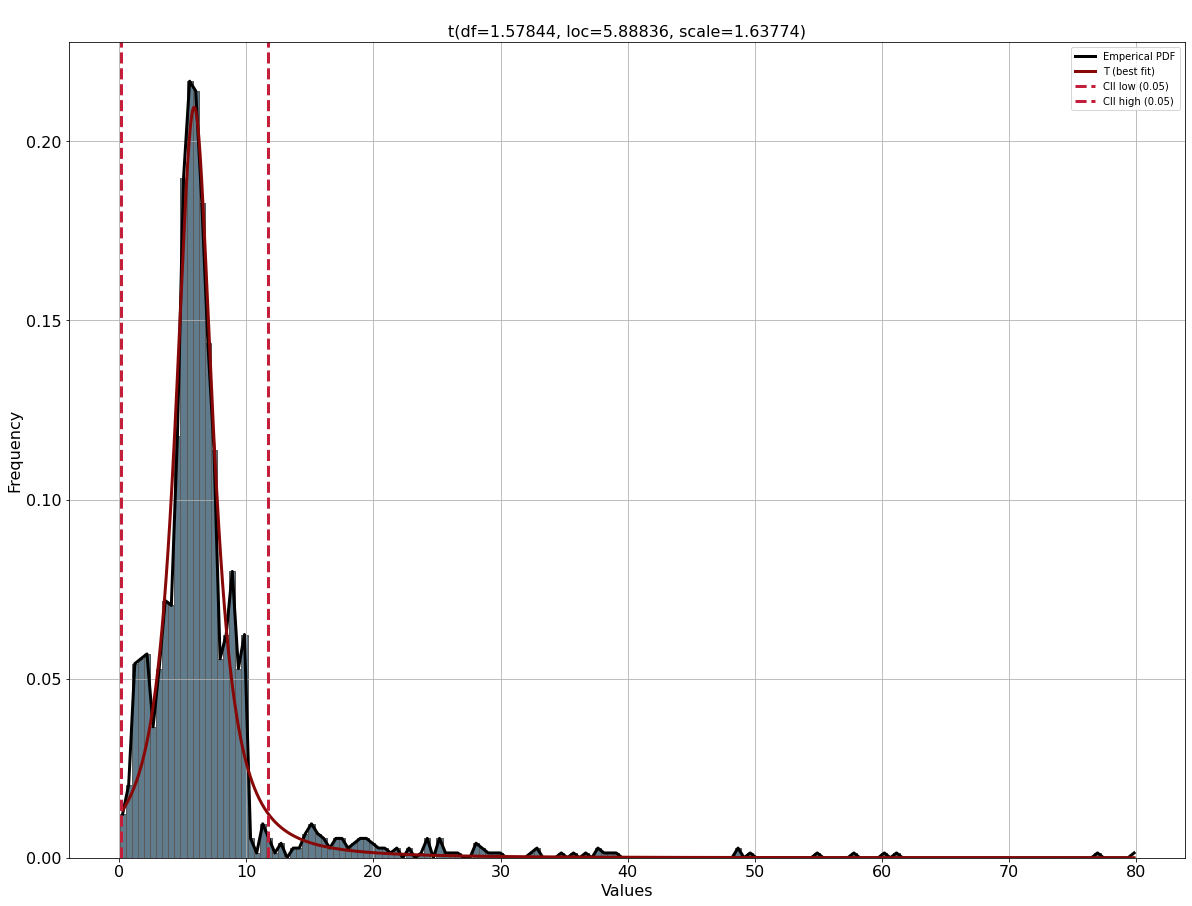


Donde se muestra que los datos están distribuidos muy parecido a una normal, pero con una pequeña asimetría hacia la izquierda y una distribución leptocúrtica. Para esta franja se ajustaron 7 distribuciones (Beta, Exponencial, Gamma, Normal, Lognormal, Weibull y Cauchy), de las cuales la distribución Lognormal es la que más se parece a la distribución de los datos para la franja 3, mostrando los siguientes resultados:

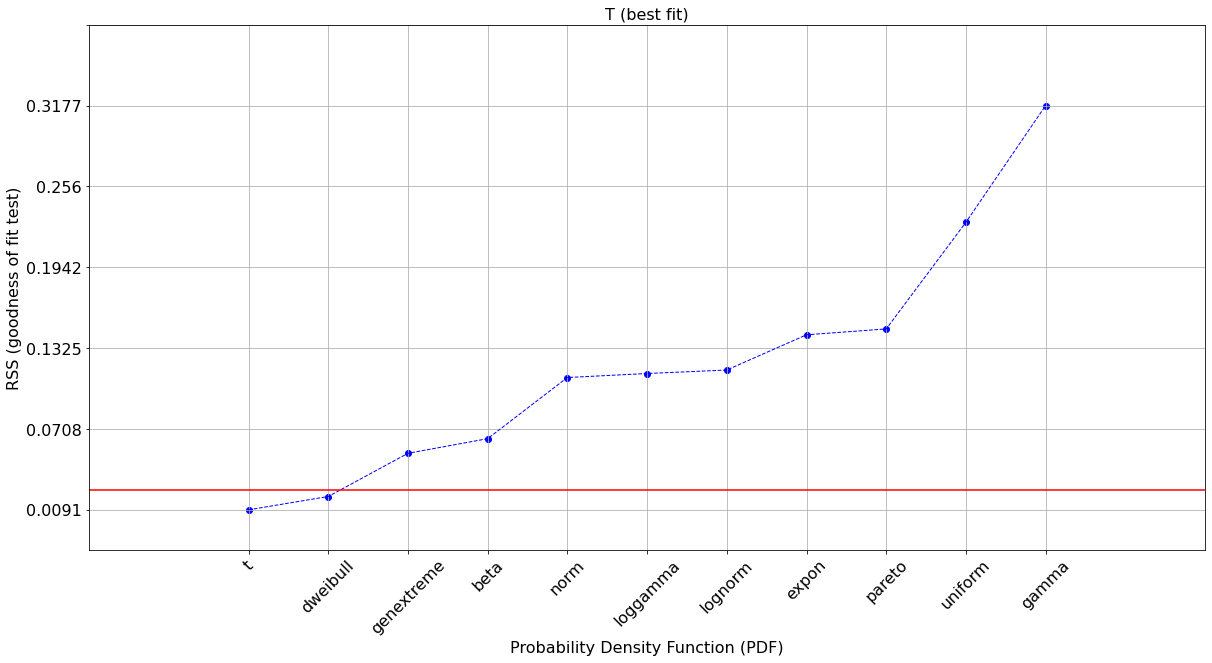


Se evidencia como en el Q-Q plot el ajuste es muy bueno, con pequeñas desviaciones en ambas colas, pero en general ajusta bien. En cada uno de los otros gráficos muestran que es una distribución adecuada para los datos. También las pruebas formales como KS y chi-cuadrado muestran que en la dos no se rechaza la hipótesis nula de que los datos de arribo para la franja 3 provienen de una distribución Lognormal.

1. Cuando se hace uso de una herramienta como Python, esta permite hacer las comparaciones mucho más rápido. De donde se puede evidenciar que sí hay cambios con respecto a lo que se encontró usando el software de R, dado que la forma de buscar distribuciones esa automática. Cuando se tiene en cuenta la totalidad de los datos, se puede ver que la distribución que más se parece al comportamiento real de los tiempos de arribo es la distribución t, como se muestra en la siguiente figura:



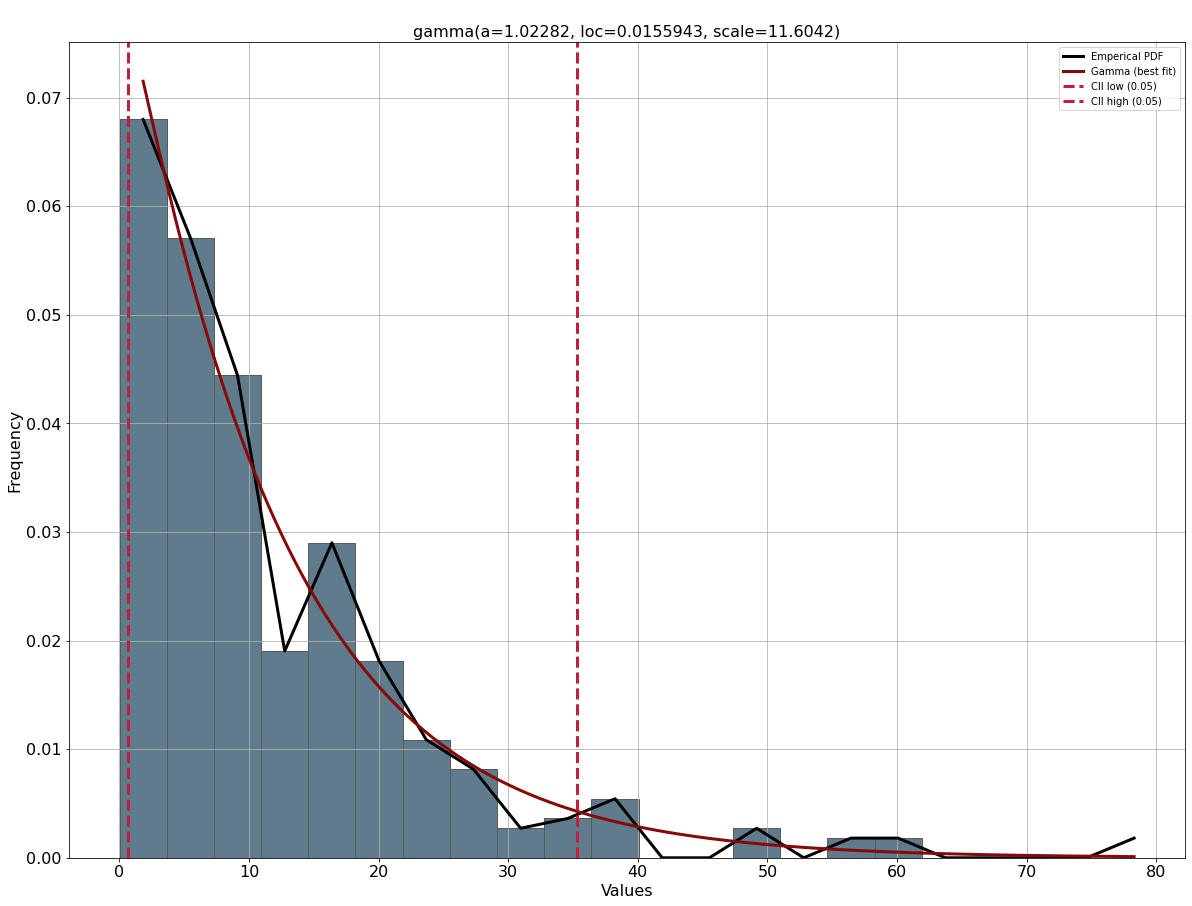
De todas maneras parece indicar que al comienzo de la distribución no lograr ajustarse de manera correcta al comportamiento de los datos. Ya cuando se evalúa con las demás distribuciones, se ve que una distribución Weibull también podría llegar a ajustarse de alguna manera.



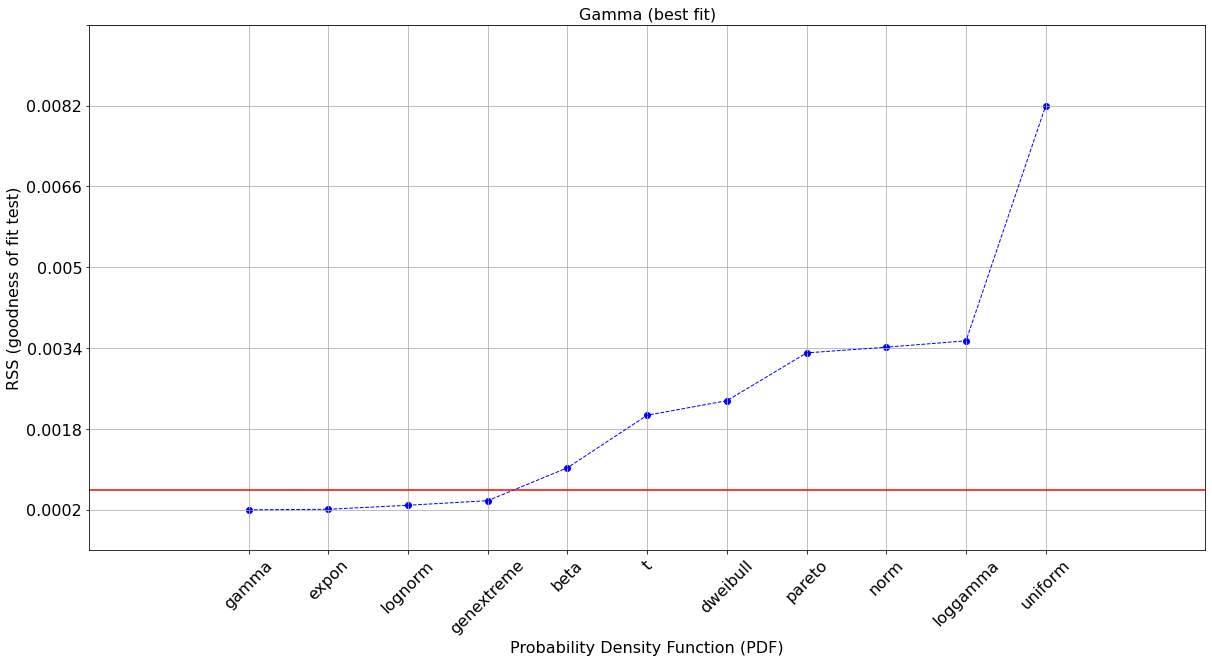
Por otro lado, cuando se analizan los datos por cada una de las franjas presentan, se obtienen los siguientes resultados:

* Franja 1:

La franja 1 muestra la siguiente distribución para el tiempo de arribo.

****

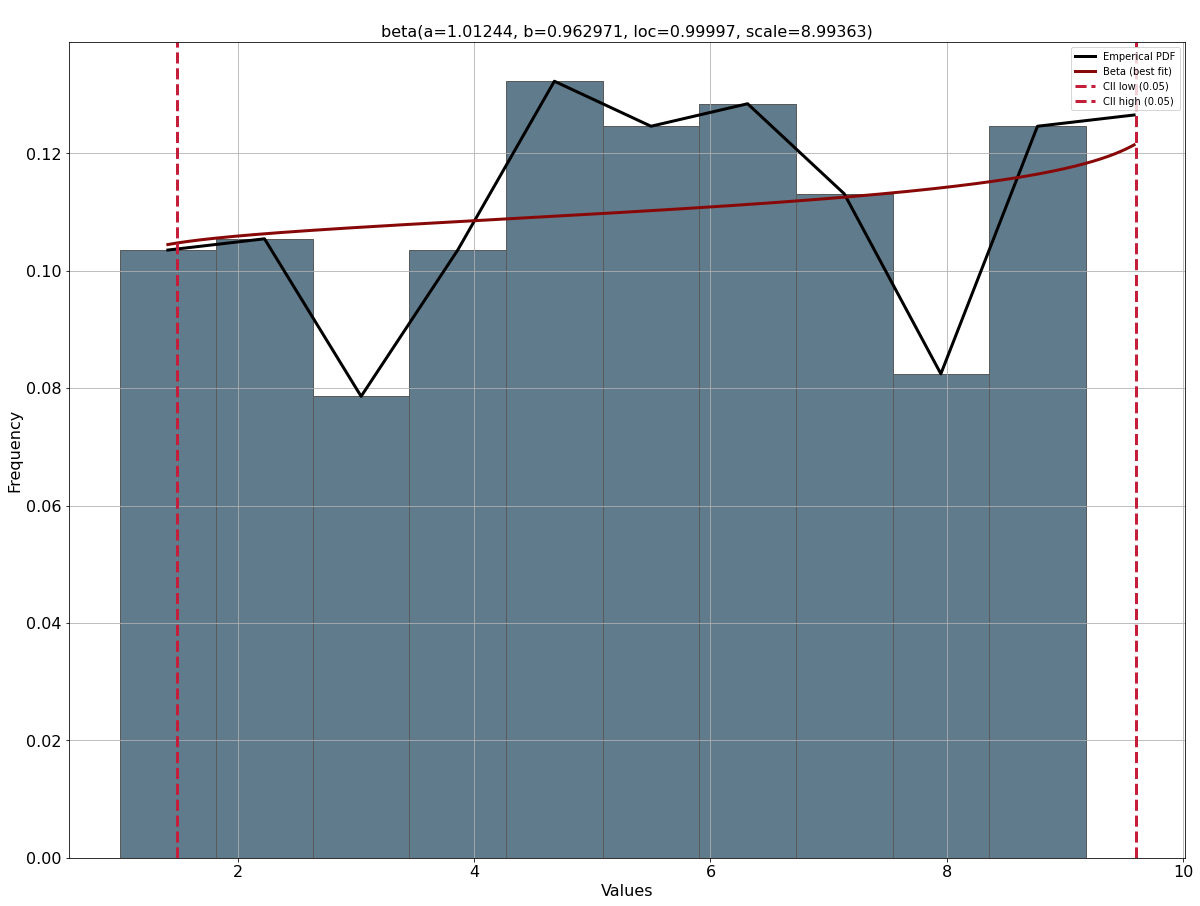
Donde se muestra que los datos están muy concentrados en los valores de 0 a 10 y presentan una cola hacia la derecha. Para esta franja, la distribución Gamma es la que más se parece a la distribución de los datos para la franja 1, mostrando el siguiente gráfico:



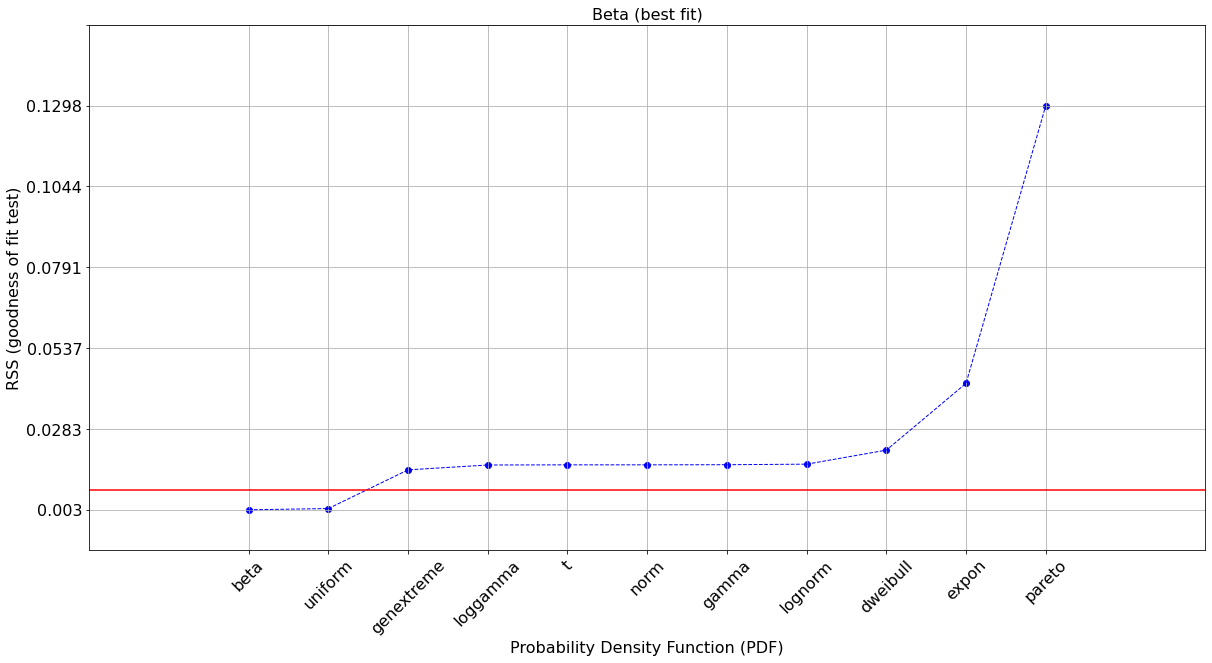
Donde se evidencia que tanto como la gamma, la exponencial, la lognormal y genextreme podrían llegar a considerarse para ser posibles distribuciones para los datos de arribo de esta franja.

* Franja 2:

La franja 2 muestra la siguiente distribución para el tiempo de arribo.



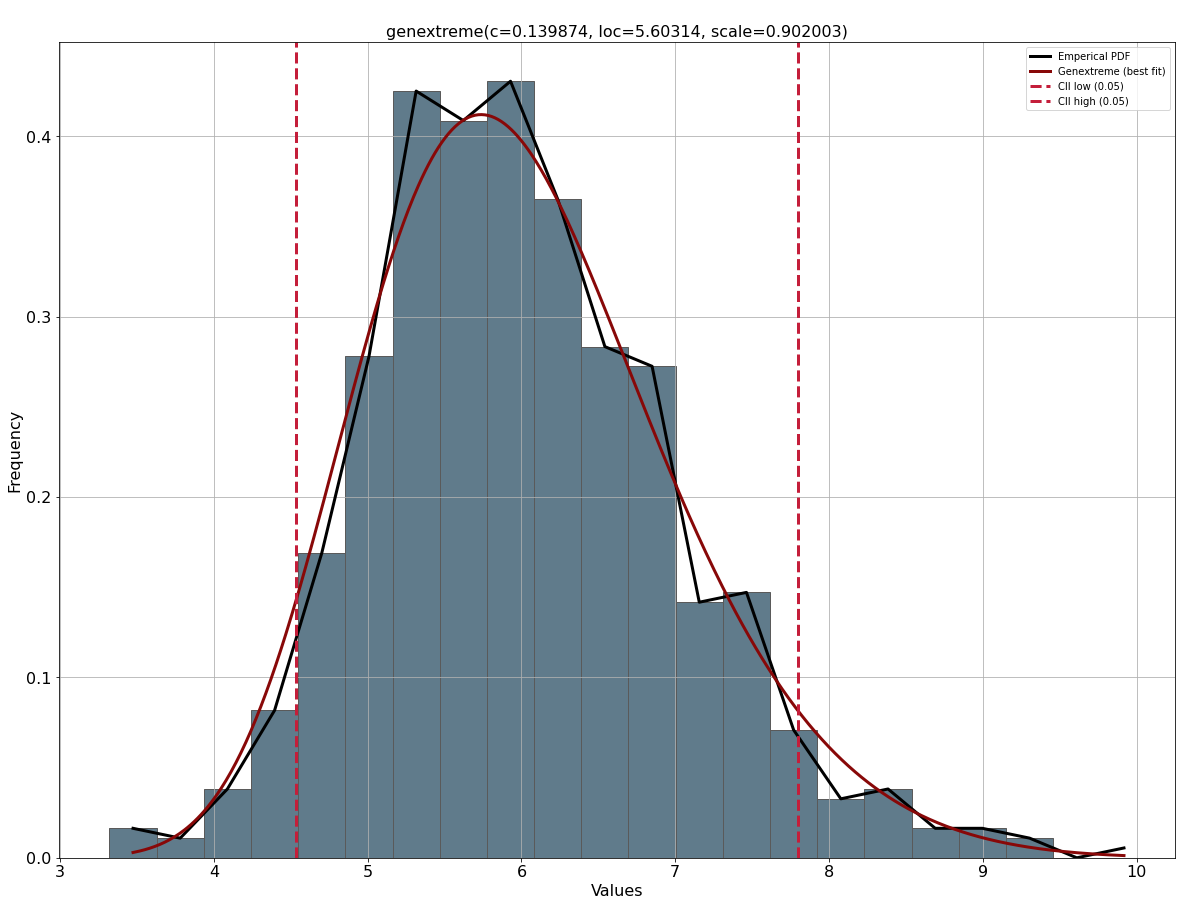
Donde se muestra que los datos están muy bien distribuidos en cada uno de los tiempos. Por lo tanto, se puede llegar a pensar de manera inicial que estos datos pueden provenir de una distribución uniforme, pero en este caso, Python determino que la distribución es beta. , mostrando los siguientes resultados:



Se evidencia como en el Q-Q plot el ajuste es muy bueno. En cada uno de los otros gráficos muestran que es una distribución adecuada para los datos. También las pruebas formales como KS y chi-cuadrado muestran que en la dos no se rechaza la hipótesis nula de que los datos de arribo para la franja 2 provienen de una distribución Uniforme.

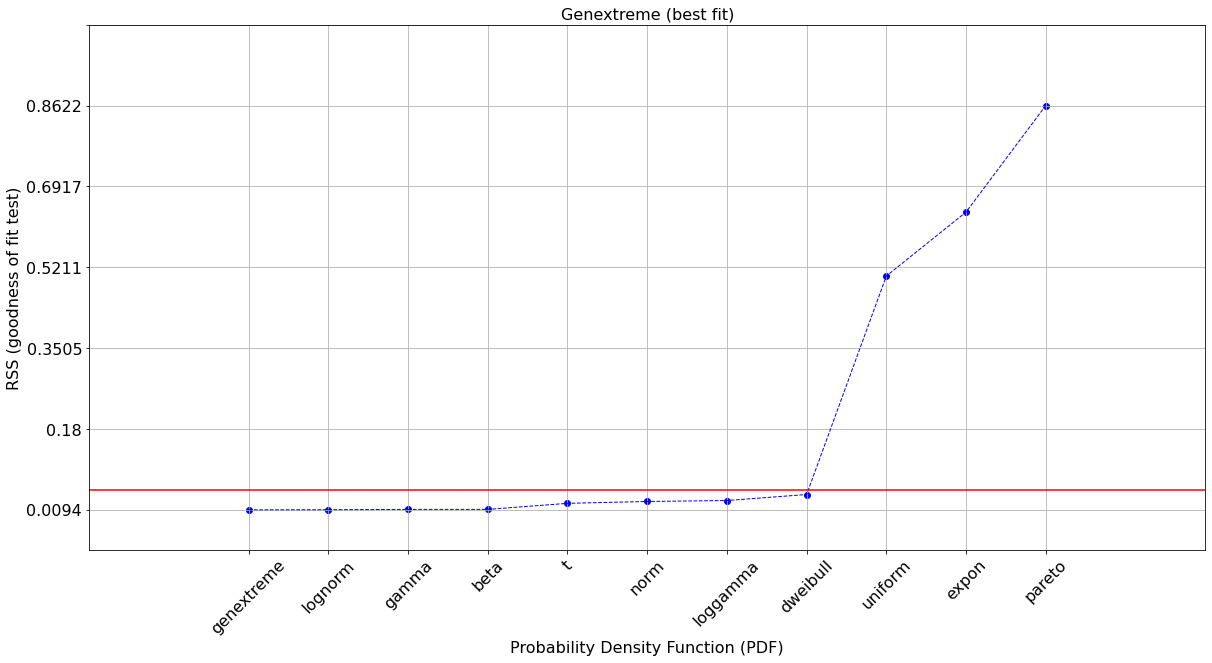
* Franja 3:

La franja 3 muestra la siguiente distribución para el tiempo de arribo.



Donde se muestra que los datos están distribuidos muy parecido a una normal, pero con una pequeña asimetría hacia la izquierda y una distribución leptocúrtica. Para esta franja se ajustaron 7 distribuciones (Beta, Exponencial, Gamma, Normal, Lognormal, Weibull y Cauchy), de las cuales la distribución Lognormal es la que más se parece a la distribución de los datos para la franja

3, mostrando los siguientes resultados:



Se evidencia como en el Q-Q plot el ajuste es muy bueno, con pequeñas desviaciones en ambas colas, pero en general ajusta bien. En cada uno de los otros gráficos muestran que es una distribución adecuada para los datos. También las pruebas formales como KS y chi-cuadrado muestran que en la dos no se rechaza la hipótesis nula de que los datos de arribo para la franja 3 provienen de una distribución Lognormal.

**Respuesta:**

1. Adjunto a este documento encontrará el archivo de jupyter mmps tarea2.ipynb. En este archivo se generan números aleatorios de acuerdo a una distribución normal, luego se grafica el histograma de los datos y la densidad asociada, y se calculan los momentos de la variable. Finalmente, se realiza un ajuste para obtener una variable aleatoria normal ajustada a los parámetros obtenidos. Este ejercicio se repite para otras variables aleatorias: gamma, beta, lognormal y weibull. Revise el cuaderno y a continuación, para cada una de las variables diferentes a la normal:
   1. Modifique alguno de los parámetros y vuelva a ejecutar el código para ver el impacto del cambio. Incluya en su reporte el cambio realizado y gráficas que muestren el cambio.
   2. Utilice la documentación disponible en https://docs.scipy.org/doc/scipy/reference/stats.html para comprender la forma funcional de la distribución de la variable aleatoria asociada y el rol del parámetro modificado. Explique en su reporte el cambio asociado por el cambio del parámetro.
   3. Seleccione una variable aleatoria diferente, disponible en scipy.stats, y realice el mismo procedimiento hecho para las demás variables. Incluya este código en su cuaderno de ipynb. En su reporte incluya una breve descripción de la variable seleccionada.

**Respuesta:**