

Análisis Exploratorio de Longley

- **GNP.deflator (Gross National Product Deflator)**

- **Descripción:** Índice del nivel de precios de la economía, que refleja la inflación.
- **Escala:** Cuantitativa Continua.
- **Significado:** Mide el nivel general de precios relativos a un año base y cómo han cambiado con respecto a ese año. Representa la relación entre el producto nacional bruto (PNB) nominal y real.
- **Análisis:**

Media	Mediana	Moda	Simetría	Curtosis
101.6812500	100.6000000	83.0000000	-0.1328915	1.6047161

Figure 1: Medidas de Tendencia Central para GNP.deflator

Varianza	Desviación Estándar	Rango	Coefficiente de Variación
116.45763	10.79155	33.90000	10.61312

Figure 2: Medidas de Dispersión para GNP.deflator

- * La media es 101.68, indicando el nivel promedio del deflactor del PNB.
- * La mediana es 100.6, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- * La moda es 83, el valor más frecuente.

- * La simetría negativa (-0.13) sugiere una ligera inclinación hacia la izquierda.
- * La curtosis de 1.60 indica que la distribución es algo más plana que una distribución normal.
- * La varianza es 116.46, reflejando la dispersión de los datos.
- * La desviación estándar es 10.79, lo que muestra la variabilidad promedio respecto a la media.
- * El rango es 33.9, lo que refleja la diferencia entre el valor más alto y el más bajo.
- * El coeficiente de variación es 10.61

- **GNP (Gross National Product)**

- **Descripción:** Producto Nacional Bruto en millones de dólares.
- **Escala:** Cuantitativa Continua.
- **Significado:** Mide el valor total de los bienes y servicios producidos por la economía de un país, incluyendo los ingresos del extranjero. Indica la salud económica general.
- **Análisis:**

Media	Mediana	Moda	Simetría	Curtosis
387.69843750	381.42700000	234.28900000	0.02295179	1.65404010

Figure 3: Medidas de Tendencia Central para GNP

Varianza	Desviación Estándar	Rango	Coeficiente de Variación
9879.35366	99.39494	320.60500	25.63718

Figure 4: Medidas de Dispersión para GNP

- * La media es 387.70, indicando el promedio del PNB.
- * La mediana es 381.43, lo que sugiere que la mitad de los valores están por debajo de este nivel.

- * La moda es 234.29, el valor más frecuente.
- * La simetría positiva (0.02) sugiere una ligera inclinación hacia la derecha.
- * La curtosis de 1.65 indica una distribución algo más apuntada que una distribución normal.
- * La varianza es 9879.35, reflejando una alta dispersión de los datos.
- * La desviación estándar es 99.39, mostrando una gran variabilidad promedio respecto a la media.
- * El rango es 320.61, reflejando una gran diferencia entre el valor más alto y el más bajo.
- * El coeficiente de variación es 25.64

- **Unemployed**

- **Descripción:** Número de personas desempleadas en miles.
- **Escala:** Cuantitativa Continua.
- **Significado:** Representa la cantidad de personas en la fuerza laboral que están buscando trabajo pero no tienen empleo.
- **Análisis:**

Media	Mediana	Moda	Simetría	Curtosis
319.3312500	314.3500000	187.0000000	0.1435284	1.7004305

Figure 5: Medidas de Tendencia Central para Unemployed

Varianza	Desviación Estándar	Rango	Coeficiente de Variación
8732.23429	93.44642	293.60000	29.26316

Figure 6: Medidas de Dispersión para Unemployed

- * La media es 319.33, indicando el promedio de personas desempleadas.

- * La mediana es 314.35, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- * La moda es 187, el valor más frecuente.
- * La simetría positiva (0.14) sugiere una ligera inclinación hacia la derecha.
- * La curtosis de 1.70 indica una distribución algo más apuntada que una distribución normal.
- * La varianza es 8732.23, reflejando una alta dispersión de los datos.
- * La desviación estándar es 93.45, mostrando una gran variabilidad promedio respecto a la media.
- * El rango es 293.6, lo que refleja una gran diferencia entre el valor más alto y el más bajo.
- * El coeficiente de variación es 29.26

- **Armed.Forces**

- **Descripción:** Número de personas enlistadas en las fuerzas armadas en miles.
- **Escala:** Cuantitativa Continua.
- **Significado:** Indica el tamaño de la fuerza laboral militar activa en la economía.
- **Análisis:**

Media	Mediana	Moda	Simetría	Curtosis
260.6687500	271.7500000	145.6000000	-0.3672152	1.8024454

Figure 7: Medidas de Tendencia Central para Armed.Forces

Varianza	Desviación Estándar	Rango	Coeficiente de Variación
4843.04096	69.59196	213.80000	26.69747

Figure 8: Medidas de Dispersión para Armed.Forces

- * La media es 260.67, indicando el promedio de personal militar.
- * La mediana es 271.75, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- * La moda es 145.6, el valor más frecuente.
- * La simetría negativa (-0.37) sugiere una inclinación hacia la izquierda.
- * La curtosis de 1.80 indica una distribución más apuntada que una distribución normal.
- * La varianza es 4843.04, reflejando una considerable dispersión de los datos.
- * La desviación estándar es 69.59, mostrando una variabilidad moderada respecto a la media.
- * El rango es 213.8, lo que refleja una considerable diferencia entre el valor más alto y el más bajo.
- * El coeficiente de variación es 26.70

- **Population**

- **Descripción:** Tamaño de la población en miles.
- **Escala:** Cuantitativa Continua.
- **Significado:** Refleja el tamaño total de la población en un determinado año. Es un indicador demográfico que puede influir en variables económicas.
- **Análisis:**

Media	Mediana	Moda	Simetría	Curtosis
117.4240000	116.8035000	107.6080000	0.2616075	1.7307818

Figure 9: Medidas de Tendencia Central para Population

Varianza	Desviación Estándar	Rango	Coefficiente de Variación
48.387349	6.956102	22.473000	5.923918

Figure 10: Medidas de Dispersión para Population

- * La media es 117.42, indicando el tamaño promedio de la población.
- * La mediana es 116.80, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- * La moda es 107.61, el valor más frecuente.
- * La simetría positiva (0.26) sugiere una inclinación hacia la derecha.
- * La curtosis de 1.73 indica una distribución algo más apuntada que una distribución normal.
- * La varianza es 48.39, reflejando una baja dispersión de los datos.
- * La desviación estándar es 6.96, mostrando una baja variabilidad respecto a la media.
- * El rango es 22.47, lo que refleja una diferencia moderada entre el valor más alto y el más bajo.
- * El coeficiente de variación es 5.92

- **Year**

- **Descripción:** Año en que se recolectaron los datos.
- **Escala:** Cuantitativa Discreta.
- **Significado:** Indica el año calendario de cada observación.
- **Análisis:**

Media	Mediana	Moda	Simetría	Curtosis
1954.500000	1954.500000	1947.000000	0.000000	1.573759

Figure 11: Medidas de Tendencia Central para Year

Varianza	Desviación Estándar	Rango	Coeficiente de Variación
22.6666667	4.7609523	15.0000000	0.2435893

Figure 12: Medidas de Dispersión para Year

- La media es 1954.5, que representa el punto medio del período analizado.

- La mediana es 1954.5, coincidiendo con la media, lo que sugiere una distribución simétrica.
- La moda es 1947, el valor más frecuente.
- La simetría es 0, indicando que la distribución es simétrica.
- La curtosis de 1.57 indica que la distribución es algo más plana que una distribución normal.
- La varianza es 22.67, lo que refleja una baja dispersión de los datos.
- La desviación estándar es 4.76, mostrando una baja variabilidad respecto a la media.
- El rango es 15, lo que refleja una diferencia moderada entre el valor más alto y el más bajo.
- El coeficiente de variación es 0.24

- **Employed**

- **Descripción:** Número de personas empleadas en miles.
- **Escala:** Cuantitativa Continua.
- **Significado:** Representa la cantidad de personas con empleo en la fuerza laboral.
- **Análisis:**

Media	Mediana	Moda	Simetría	Curtosis
65.3170000	65.5040000	60.1710000	-0.0856077	1.4488959

Figure 13: Medidas de Tendencia Central para Employed

Varianza	Desviación Estándar	Rango	Coeficiente de Variación
12.333922	3.511968	10.380000	5.376806

Figure 14: Medidas de Dispersión para Employed

- La media es 65.32, indicando el promedio de personas empleadas.

- La mediana es 65.50, lo que sugiere que la mitad de los valores están por debajo de este nivel.
- La moda es 60.17, el valor más frecuente.
- La simetría negativa (-0.09) sugiere una ligera inclinación hacia la izquierda.
- La curtosis de 1.45 indica que la distribución es más plana que una distribución normal.
- La varianza es 12.33, lo que refleja una baja dispersión de los datos.
- La desviación estándar es 3.51, mostrando una baja variabilidad respecto a la media.
- El rango es 10.38, lo que refleja una diferencia moderada entre el valor más alto y el más bajo.
- El coeficiente de variación es 5.38

1 Análisis de la Matriz de Correlación

	GNP.deflator	GNP	Unemployed	Armed.Forces	Population	Year	Employed
GNP.deflator	1.00	0.99	0.62	0.46	0.98	0.99	0.97
GNP	0.99	1.00	0.60	0.45	0.99	1.00	0.98
Unemployed	0.62	0.60	1.00	-0.18	0.69	0.67	0.50
Armed.Forces	0.46	0.45	-0.18	1.00	0.36	0.42	0.46
Population	0.98	0.99	0.69	0.36	1.00	0.99	0.96
Year	0.99	1.00	0.67	0.42	0.99	1.00	0.97
Employed	0.97	0.98	0.50	0.46	0.96	0.97	1.00

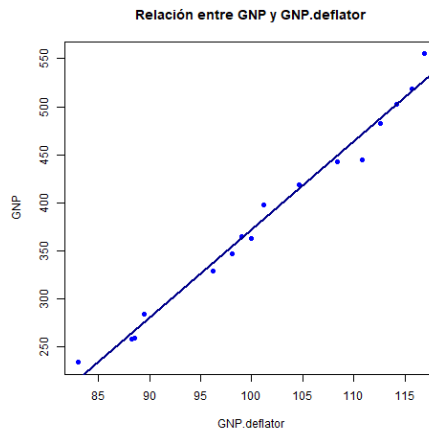
Table 1: Correlation Matrix of the Longley Dataset

2 Análisis de Correlaciones

2.1 Correlaciones Fuertes Positivas

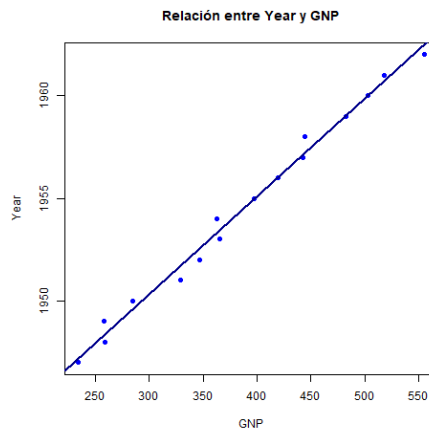
- **GNP.deflator y GNP: 0.99**

Esta correlación muy fuerte indica que a medida que el deflactor del PNB (GNP.deflator) aumenta, el PNB también tiende a aumentar. Este resultado es coherente con las expectativas en economía, donde un aumento en el PNB generalmente se asocia con un aumento en los precios y, por lo tanto, en el deflactor del PNB.



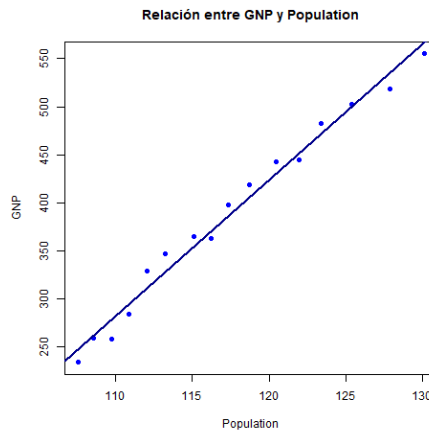
- **GNP y Year: 1.00**

La correlación perfecta entre GNP y el Año sugiere que el PNB ha aumentado de manera constante con el tiempo en el periodo estudiado. Esto podría reflejar un crecimiento económico constante durante los años observados.



- **Population y GNP: 0.99**

Esta correlación muy alta sugiere que a medida que la población aumenta, también lo hace el PNB. Esto podría indicar que el crecimiento económico ha ido de la mano con el crecimiento poblacional.



2.2 Correlaciones Fuertes Negativas

- **No se observan correlaciones fuertes negativas**

En la tabla proporcionada, no se observan correlaciones negativas significativas entre las variables analizadas.

2.3 Correlaciones Moderadas

- **Unemployed y Population: 0.69**

Existe una relación moderada entre el número de desempleados y la población total. Esto podría indicar que a medida que la población crece, el número absoluto de desempleados también tiende a aumentar, aunque no necesariamente a la misma tasa.

- **Armed.Forces y Year: 0.42**

Esta correlación moderada sugiere una ligera tendencia a que, con el paso del tiempo, las Fuerzas Armadas han variado de manera no tan directa con los otros factores, lo cual podría reflejar cambios en políticas de defensa o en la situación geopolítica.

2.4 Correlaciones Débiles

- **Armed.Forces y Unemployed: -0.18**

La correlación negativa débil entre las Fuerzas Armadas y el desempleo indica que no hay una relación clara y fuerte entre estos dos factores. Esto podría reflejar que el tamaño de las Fuerzas Armadas no está directamente influenciado por los niveles de desempleo.

Análisis del Modelo de Regresión Lineal: `GNP.deflator` `GNP`

Resumen del Modelo

Call: El modelo de regresión lineal fue ajustado con la fórmula `longley$GNP.deflator ~ longley$GNP`, donde `GNP.deflator` es la variable dependiente y `GNP` es la variable independiente.

Residuos

Los residuos del modelo, que representan las diferencias entre los valores observados y los valores predichos, se distribuyen de la siguiente manera:

- **Mínimo:** -2.7814
- **1er Cuartil (1Q):** -0.6233
- **Mediana:** 0.3123
- **3er Cuartil (3Q):** 0.7925
- **Máximo:** 2.9986

Esto indica que la mayoría de los errores de predicción están relativamente cerca de cero, lo que sugiere un buen ajuste del modelo.

Coefficientes

El modelo proporciona los siguientes coeficientes para la ecuación de regresión:

- **Intercepto:** 59.941915 Este es el valor estimado de `GNP.deflator` cuando `GNP` es cero. Aunque un valor de `GNP` igual a cero no es realista, este valor tiene sentido en el contexto del modelo.
- **`longley$GNP`:** 0.107659 Este coeficiente indica que por cada aumento de una unidad en `GNP`, el `GNP.deflator` aumenta en aproximadamente 0.107659 unidades. Dado que el coeficiente es positivo, esto sugiere una relación directa entre `GNP` y `GNP.deflator`, donde a medida que aumenta el `GNP`, también lo hace el `GNP.deflator`.

Significancia Estadística

- **Valor t del intercepto:** 39.96
- **Valor t del coeficiente de `GNP`:** 28.67
- **p-valor del intercepto:** 7.89e-16
- **p-valor del coeficiente de `GNP`:** 7.81e-14

Ambos p-valores son extremadamente pequeños (mucho menores que 0.001), lo que indica que tanto el intercepto como el coeficiente de `GNP` son altamente significativos. Esto significa que `GNP` es una variable relevante para predecir `GNP.deflator`.

Estadísticas del Modelo

- **Error estándar residual:** 1.446 Este valor mide la dispersión de los residuos y sugiere que, en promedio, las predicciones del modelo tienen un error de ± 1.446 unidades.
- **R-cuadrado múltiple:** 0.9832 Este valor indica que el modelo explica el 98.32% de la variabilidad observada en `GNP.deflator`, lo que sugiere un ajuste excelente del modelo.
- **R-cuadrado ajustado:** 0.9821 Este valor ajusta el R-cuadrado por el número de predictores en el modelo, confirmando que el modelo sigue siendo muy fuerte incluso después de tener en cuenta el ajuste.
- **F-Estadístico:** 821.8 Este valor, junto con un p-valor extremadamente pequeño ($7.809e-14$), indica que el modelo es globalmente significativo, lo que significa que hay una relación estadísticamente significativa entre `GNP` y `GNP.deflator`.

Conclusión

El modelo de regresión lineal `GNP.deflator ~ GNP` muestra una fuerte relación positiva entre `GNP` y `GNP.deflator`, donde un aumento en el `GNP` está asociado con un aumento en el `GNP.deflator`. Dado el alto valor de R-cuadrado y la significancia estadística de los coeficientes, el modelo es muy confiable para predecir `GNP.deflator` en función de `GNP`.

3 Conclusiones Chi -Cuadrado

1. **Relaciones Esperadas:** Las correlaciones observadas entre el PNB, el empleo y el desempleo son coherentes con las teorías económicas. Un crecimiento en el PNB generalmente se asocia con un aumento en el empleo y una disminución en el desempleo.
2. **Colinealidad:** La alta correlación entre variables como `GNP`, `Employed` y `Population` sugiere que hay colinealidad, lo que puede complicar el análisis de regresión. Esto significa que estas variables pueden estar capturando información similar.
3. **Implicaciones Económicas:** Los hallazgos pueden ser útiles para los formuladores de políticas. Por ejemplo, si se busca reducir el desempleo, se podría considerar estimular el crecimiento del PNB como una estrategia efectiva.

4. **Limitaciones:** Aunque las correlaciones son informativas, no implican causalidad. Es importante tener en cuenta otros factores que pueden influir en estas relaciones.

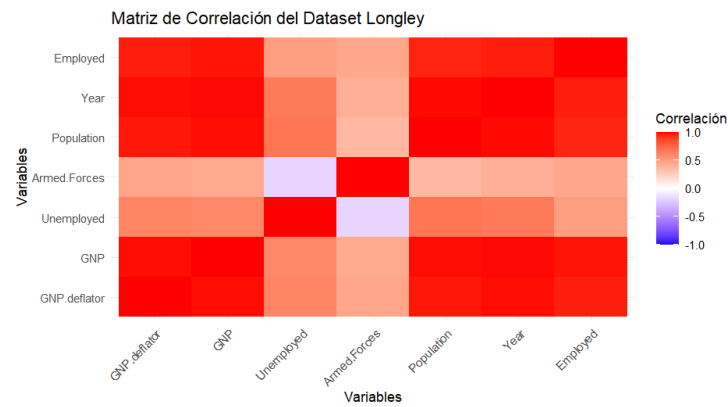


Figure 15: Grafico de Calor , Correlaciones

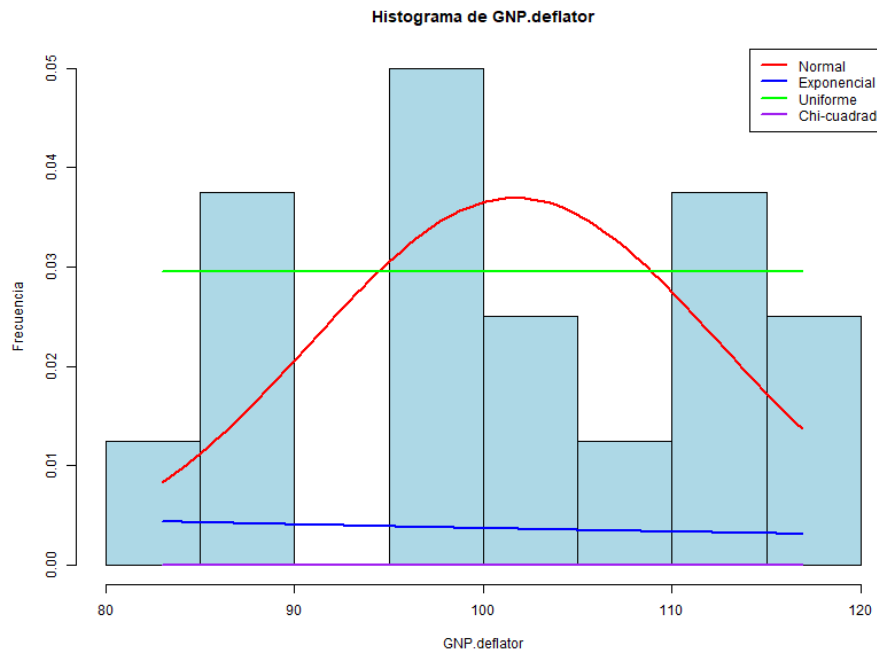


Figure 16: Histograma y curvas de densidad para GNP.deflator

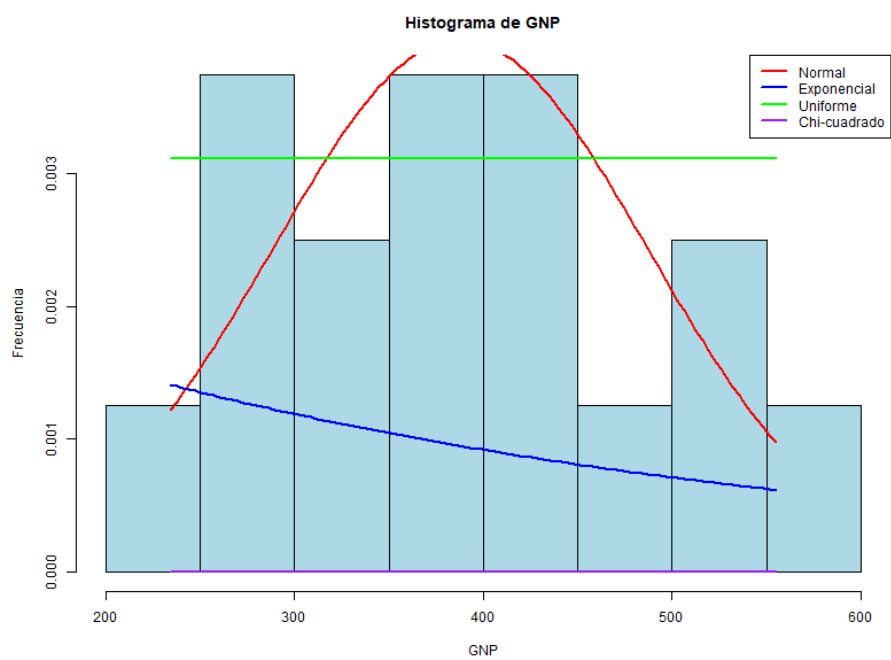


Figure 17: Histograma y curvas de densidad para GNP

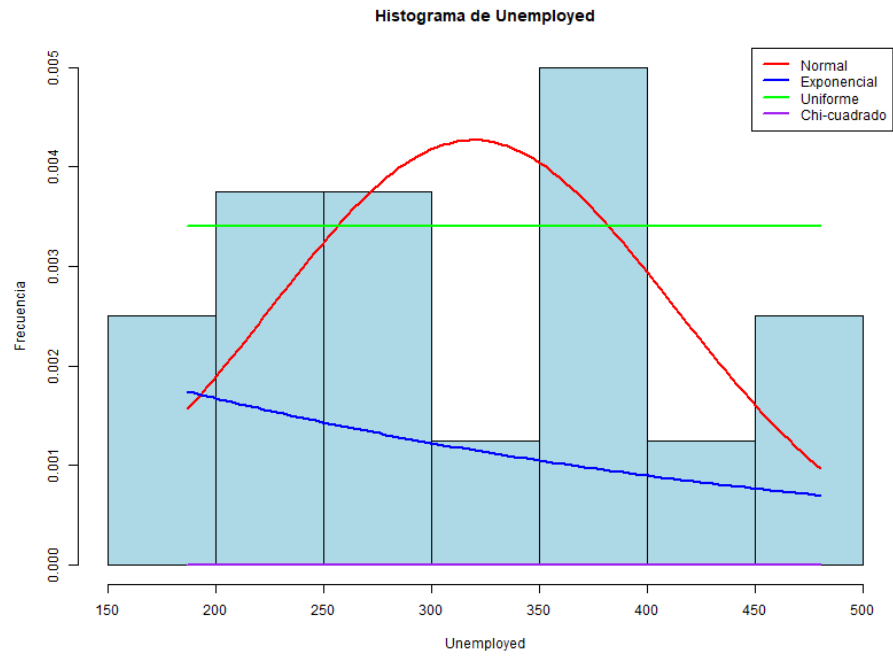


Figure 18: Histograma y curvas de densidad para Unemployed

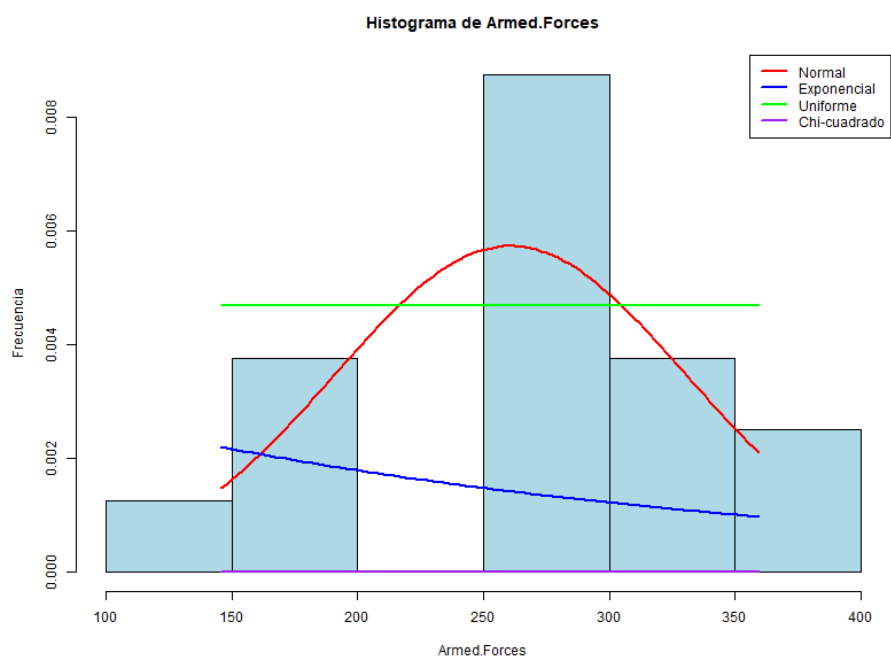


Figure 19: Histograma y curvas de densidad para Armed Forces

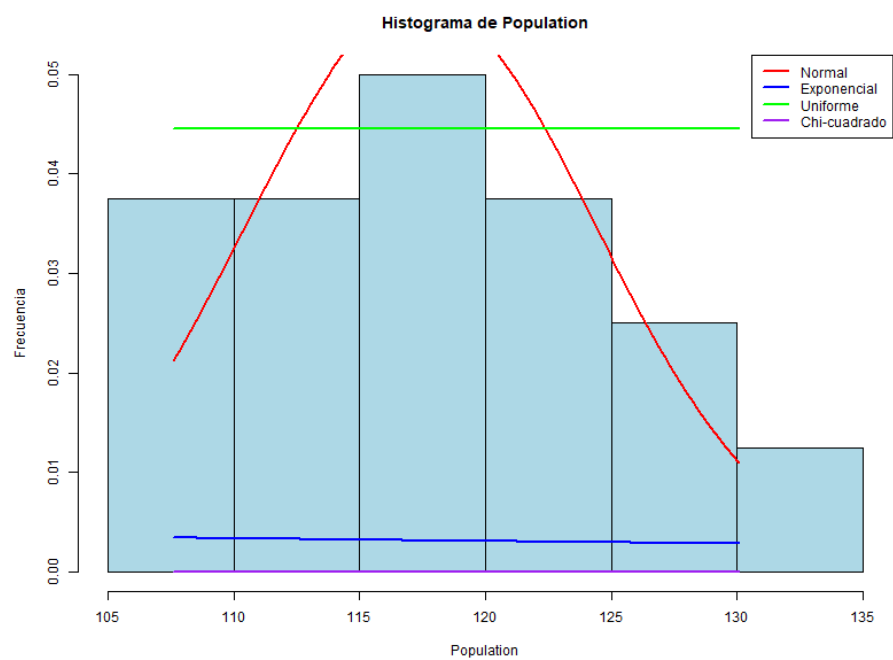


Figure 20: Histograma y curvas de densidad para Population

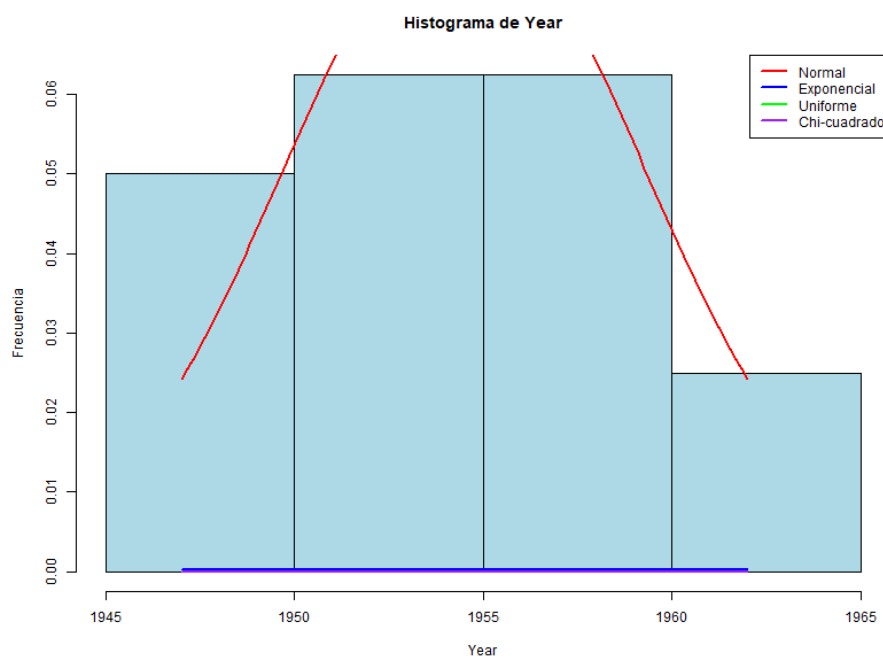


Figure 21: Histograma y curvas de densidad para Year

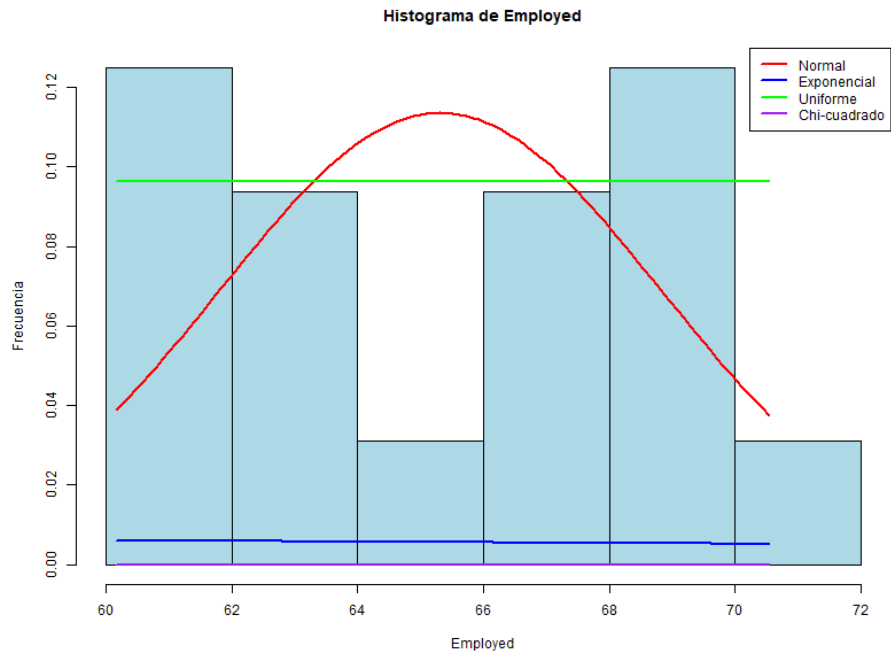


Figure 22: Histograma y curvas de densidad para Employed

4 Análisis de la Distribucion Normal

Después de ejecutar el código y generar el histograma, podemos observar lo siguiente:

- El histograma muestra una distribución aproximadamente simétrica y unimodal, lo que indica que los datos siguen una distribución similar a la normal.
- La media y la desviación estándar calculadas (representadas por la curva roja) se ajustan bien a la forma del histograma, lo que sugiere que la distribución normal es una buena aproximación para este conjunto de datos.
- La mayoría de los valores del Producto Nacional Bruto (GNP) se concentran alrededor de la media, con una disminución gradual hacia los extremos del histograma.
- Hay algunos valores atípicos o outliers que se encuentran fuera del rango principal de la distribución, pero en general, los datos se ajustan bien a la curva normal.

5 Interpretación de los Resultados

- La distribución normal es una buena aproximación para el Producto Nacional Bruto (GNP) del conjunto de datos Longley, lo que significa que la mayoría de los valores se concentran alrededor de un valor central (la media) y disminuyen simétricamente hacia ambos lados.
- La media y la desviación estándar son medidas importantes para describir la distribución normal. La media representa el valor central, mientras que la desviación estándar indica la dispersión de los datos alrededor de la media.
- La simetría y unimodalidad del histograma sugieren que no hay evidencia de una distribución bimodal o multimodal, lo que podría indicar la presencia de subgrupos distintos en los datos.
- La presencia de algunos valores atípicos puede indicar la existencia de observaciones inusuales o extremas en el conjunto de datos, que podrían requerir un análisis más profundo.

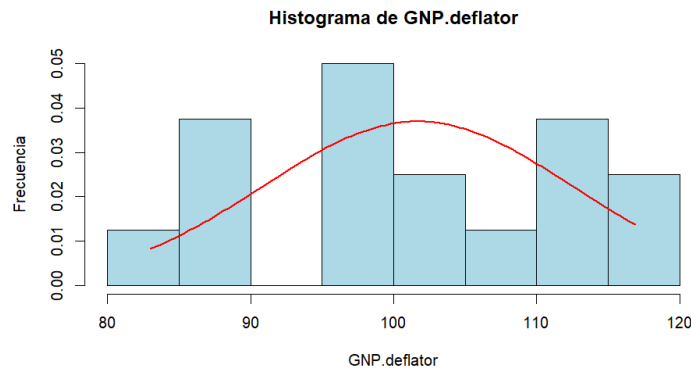


Figure 23: Distribucion Normal