Universidad Simón Bolívar

C03321

Trimestre: Sep-Dic 2016

Estadística para ingenieros

Pablo Betancourt 13-10147

José Luis Acevedo 13-10006

**PROYECTO I (10%)**

**Resumen:**

**(Despues mejoro la portada)**

**(Marico el que lo lea)**

**PLANTEAMIENTO DEL PROBLEMA**

Según Wackerly, Mendenhall y Scheaffer (2008), la estadística es la teoría de la información cuyo objetivo es la inferencia de las propiedades de la muestra que se considera. Esta ciencia es utilizada en muchas áreas de trabajo actualmente debido a que permite obtener información un poco más realista que para fines prácticos resulta funcionar bastante bien en la mayoría de los casos. En este curso, estaremos trabajando con dos enfoques de la estadística: El enfoque descriptivo y el enfoque inferencial.

Para este caso, se realizó un proyecto con el objetivo principal de poner en práctica la teoría vista hasta el momento y terminar de afianzar los conocimientos adquiridos. El caso de estudio es el estado mundial de infancia para el año 2003 (Datos recopilados por la UNICEF).

Para poder alcanzar el objetivo principal, se deben cumplir ciertos objetivos específicos representados como sub-problemas a resolver durante el desarrollo de este proyecto.

Primero que nada, se trabajará en el enfoque descriptivo de la estadística haciendo un análisis de este tipo sobre la muestra en general. Para esto, se deben tener claras cuáles son las variables y en caso de que las mismas sean cuantitativas se deben calcular ciertas cosas como el mínimo, máximo, media, mediana, etc. Luego se interpretarán los resultados de esta primera actividad.

Luego, se hará una evaluación por continente de las variables con el objetivo de comparar los tipos de gráficos descriptivos y realizar las respectivas interpretaciones de la información que se refleja en los resultados obtenidos.

Por otro lado, se usará la información dada de una variable en específico para hacer un análisis descriptivo y comparar dichos resultados con el fin de responder ciertas preguntas y, además, sobre la información obtenida se hará uso del enfoque inferencial de la estadística para poder calcular intervalos de confianza sobre 2 parámetros (La media y la proporción)

Para realizar estas actividades y cumplir los objetivos prácticos, se utilizarán R y RStudio como herramientas de apoyo para generar las tablas y los gráficos correspondientes a cada etapa del desarrollo de este proyecto. Los mismos serán expuestos en este informe para generar una mejor comprensión del análisis y las interpretaciones que serán expuestas.

**DESARROLLO**

A continuación, se procederá a responder las preguntas planteadas para cumplir los objetivos específicos y de esta forma cumplir el general.

**Análisis descriptivo de cada una de las variables**

**Tabla 1.** Análisis descriptivo completo para cada variable



En la **Tabla 1** se presenta el análisis descriptivo global de los datos de los indicadores básicos del estado mundial de infancia para el año 2003 según la Unicef. En la misma se pueden observar las 6 variables que fueron consideradas para el estudio:

**Tasa de mortalidad menores de 5 años:**

Para esta variable se puede observar que el mínimo es 3, lo cual implica que hay uno o varios países con buenas condiciones en cuanto a salud infantil. Sin embargo, el máximo es de 284, con lo que se tiene que existe un país con condiciones no óptimas para infantes. Por otro lado, se puede inferir de la tabla que al menos el 25% de los países presentan una tasa menor o igual a 10.25 (Definición de primer cuartil) pero otro 25% presenta una tasa mayor o igual a 107.8 (Tercer cuartil). Gracias a la mediana, también se puede inferir que al menos el 50% de la población tiene una tasa menor o igual a 35. Sin embargo, tenemos que la diferencia entre el mínimo y el máximo es muy grande, lo cual le da sentido a el valor arrojado por la desviación estándar de 67.99. Esto último implica que, en este conjunto de datos, los valores están muy dispersos con respecto a la media.

**Tasa de mortalidad menores de 1 año:**

Para esta variable se puede observar que el mínimo es 3, lo cual implica que hay uno o varios países con buenas condiciones en cuanto a salud infantil. Sin embargo, el máximo es de 166, con lo que se tiene que existe un país con condiciones no óptimas para infantes. Por otro lado, se puede inferir de la tabla que al menos el 25% de los países presentan una tasa menor o igual a 8 (Definición de primer cuartil) pero otro 25% presenta una tasa mayor o igual a 77.5 (Tercer cuartil). Gracias a la mediana, también se puede inferir que al menos el 50% de la población tiene una tasa menor o igual a 28. Sin embargo, tenemos que la diferencia entre el mínimo y el máximo es muy grande, lo cual le da sentido a el valor arrojado por la desviación estándar de 41.38. Esto último implica que, en este conjunto de datos, los valores están muy dispersos con respecto a la media.

**Población total (Miles)**

En este caso, el mínimo y el máximo siguen guardando una distancia considerablemente alta. Además, el tercer cuartil se ubica en 24330, el máximo en 1304000, pero la media se ubica en 37230. Con estos valores se puede inferir que la mayoría de los países poseen una población total cuyos valores rondan a la media, pero el hecho de incluir países con una población muy alta le da sentido al valor de la desviación estándar ya que, como son datos atípicos, se encuentran muy lejanos a la media y aumentan la varianza de los datos.

**Nacimientos anuales (Miles):**

En este caso nos encontramos con algo que hemos visto anteriormente: los datos se encuentran dispersos con respecto a la media y la distancia entre el mínimo y el máximo es muy alta. Esto guarda sentido con los resultados anteriores debido a que la cantidad de nacidos en un año no toma en cuenta la realidad demográfica de los países pertenecientes a la muestra, por ende, es de esperarse que el comportamiento de esta variable cuantitativa vaya de la mano con el de la población total.

**INB per capita (Dolares):**

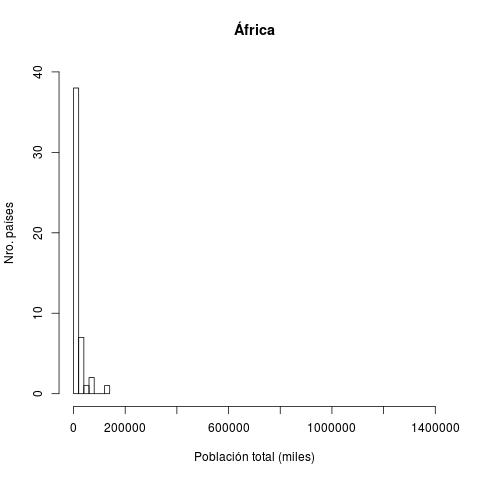
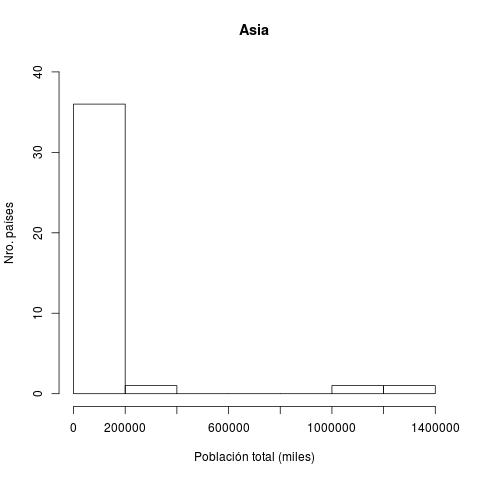
Con respecto al ingreso nacional bruto se puede inferir de la tabla que el comportamiento de esta variable se ve influido por ciertos datos atípicos, lo que genera que la media este por encima del tercer cuartil. Esto implica que al menos el 75% de los países participantes tienen un ingreso menor o igual a 5492.

**Esperanza de vida al nacer (Años):**

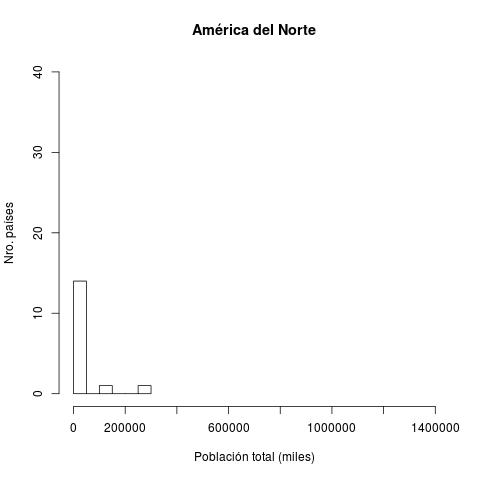
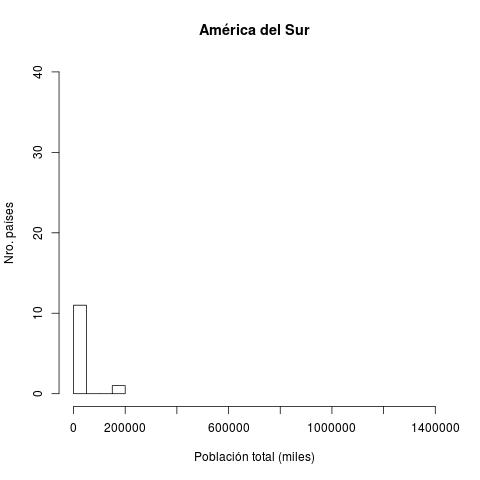
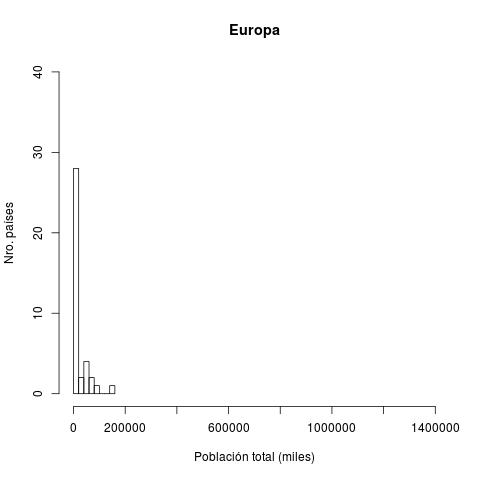
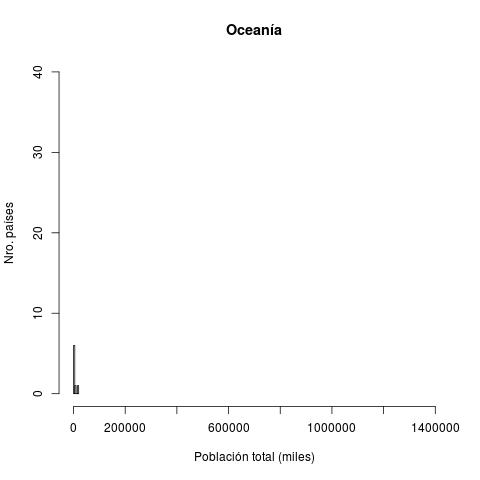
Finalmente, con esta variable, se busca medir la cantidad de años de vida que se esperan de la población durante ese año. Usualmente para calcular esta medida, se asume como fija la tasa de mortalidad y en base a ello se calcula la esperanza. En este caso, se puede observar que la media muestral se ubica en 65 años aproximadamente, lo cual entra en el rango de datos típicos y demuestra que no se ve tan afectada por las variables anteriormente mencionadas.

**Gráficos descriptivos por continente**

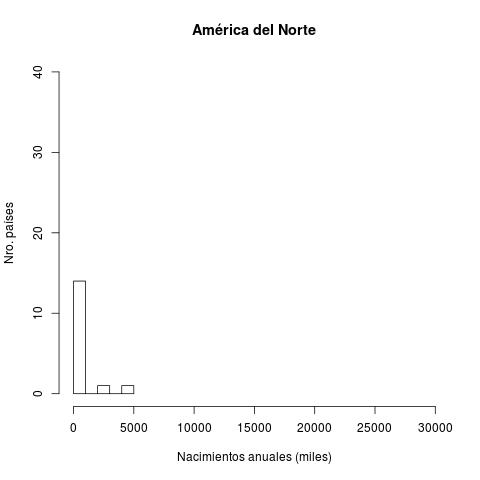
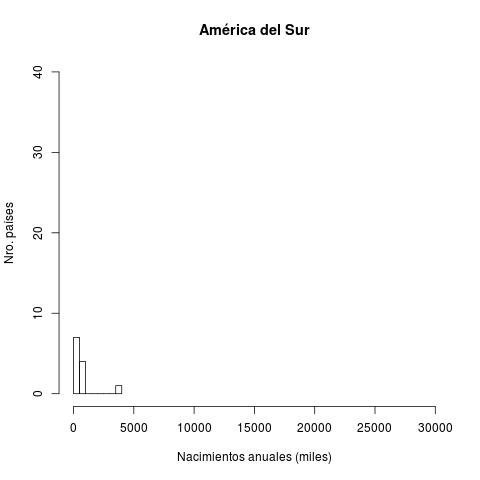
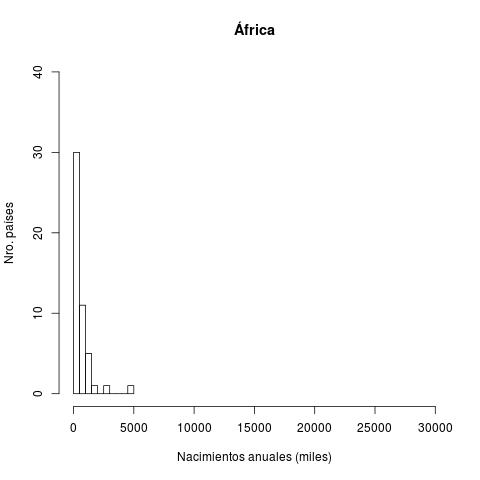
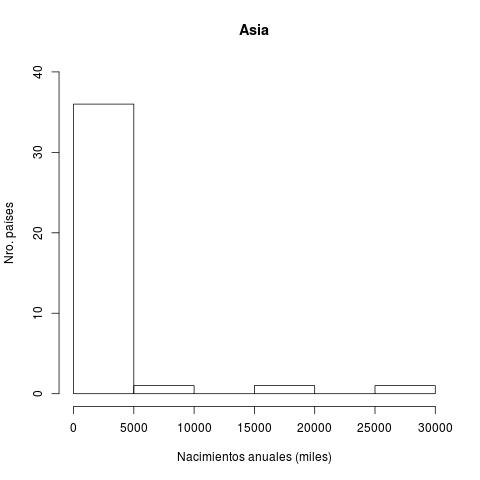
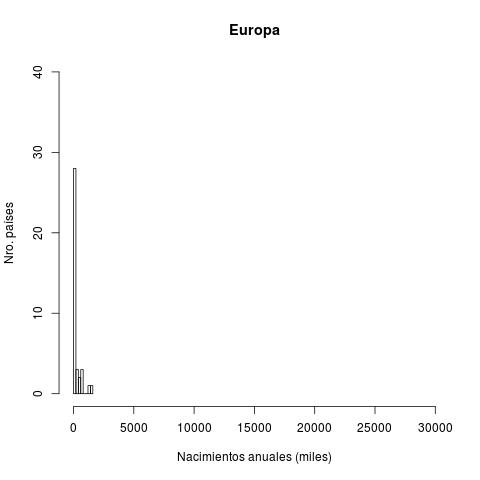
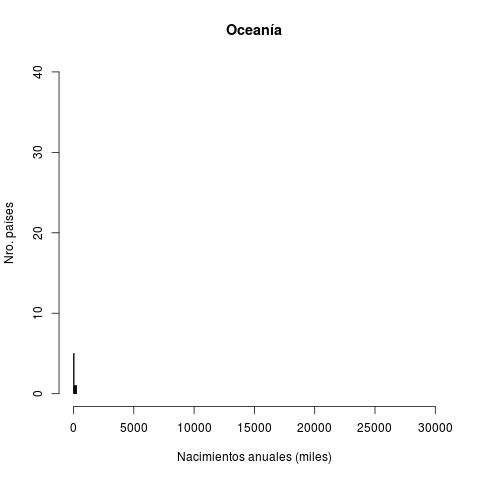
**Gráfico 1.** Población total por continente.



En el **Gráfico 1** podemos observar los histogramas de población total para cada uno de los continentes. En estos se puede observar como patrón común: en todos los continentes existe una mayoría de países con una población total menor a 200 millones de personas (Sobre todo en Asia y África). Por otro lado, en Asia y América del norte existen países, que representan una pequeña porción de cada continente, cuya población total supera las 200 millones de personas (Llevandola incluso a 1400 millones en el caso de Asia). Toda esta información respalda lo expuesto anteriormente con respecto a la población total, es decir, la mayoría de los continentes mantienen sus cifras cerca de la media mientras que hay unos pocos países cuyas cantidades generan un aumento abrupto de la desviación estándar dado que son datos atípicos.

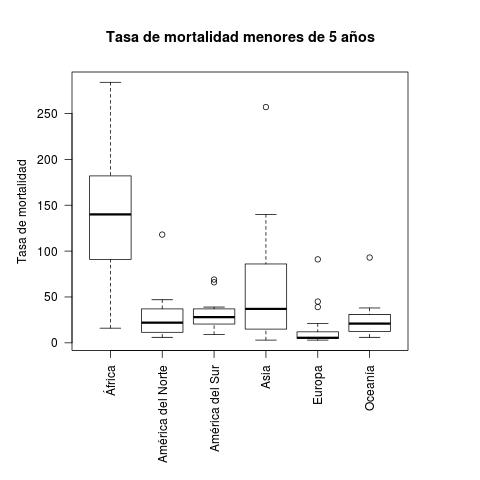


**Gráfico 2.** Nacimientos anuales por continente.



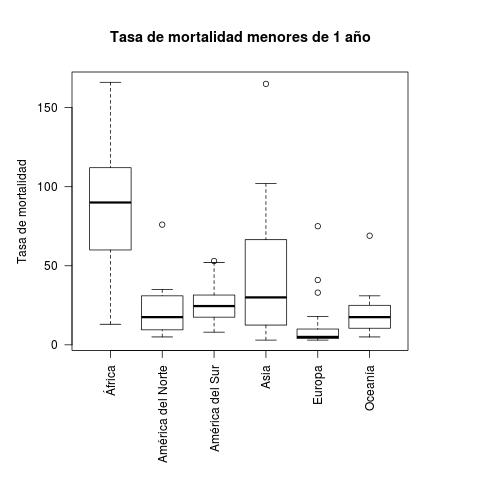
En el **Gráfico 2** se pueden observar los histogramas de nacimientos anuales por continente. Se sigue observando el mismo patrón en el que existen muchos países en los que la cantidad de nacimientos al año no llega a ser más de 5 millones, pero en este caso, Asia posee países que tienen de 5 a 10 millones, de 15 a 20 millones y de 25 a 30 millones de nacimientos anuales. El comportamiento de estos histogramas se asemeja al de la población total, lo cual respalda el hecho de que esta medición depende directamente de la realidad demográfica del país donde se esté haciendo la encuesta.

**Gráfico 3.** Diagrama de caja para la tasa de mortalidad de menores a 5 años.



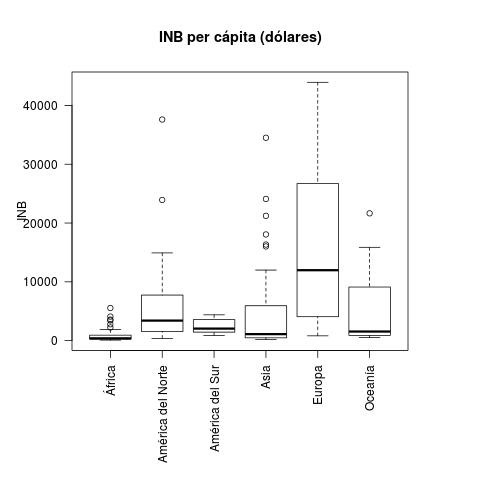
En el **gráfico 3** se encuentran representadas las tasas de mortalidad para menores de 5 años por continente. Primero que nada, se puede observar que Europa, América del sur, América del norte y Oceanía presentan los rangos intercuartiles mas pequeños a pesar de que tienen datos atípicos. Por otro lado, aunque Asia tiene un rango intercuartil grande, la mayoría de sus valores se preservan por debajo de los 150 (A pesar del dato atípico) a diferencia de Africa que, aunque no presenta datos atípicos, su rango intercuartil se encuentra entre 100 y 200 y el rango de sus valores pasa los 250.

**Gráfico 4.** Diagrama de caja para la tasa de mortalidad de menores a 1 año.



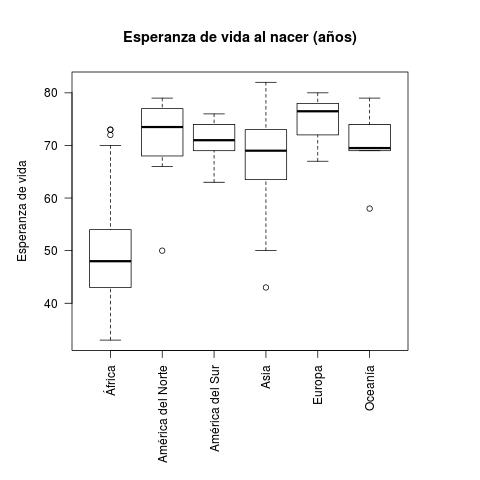
En el **gráfico 4** se visualizan los diagramas de caja correspondientes a la tasa de mortalidad de menores a 1 año. Para este caso, nuevamente Europa, Oceania, América del sur y América del norte tienen rangos intercuartiles pequeños y ubicados en posiciones bajas. Sin embargo, esta vez los datos atípicos están mas cercanos a la gráfica. Asia presenta la misma condición que en la gráfica anterior (Rango intercuartil amplio debido a población total y dispersión de los datos) y se sigue manteniendo en valores bajos. Africa no presenta cambios en cuanto a comportamiento, es decir, carece de datos atípicos pero tiene un rango intercuartil muy amplio y este se ubica en valores altos.

**Gráfico 5.** Diagrama de caja de INB per cápita por continente.



En el **gráfico 5** se aprecian los diagramas de caja del INB per cápita para cada continente. En este caso se puede apreciar que África y América del sur poseen valores muy bajos. En el caso de África, presenta valores atípicos, pero estos no llegan a 10000. Por otro lado, Oceanía, Asia y América del norte parecen tener una proporción parecida entre sí (Exceptuando la posición de la mediana en el rango intercuartil) y tanto Asia como América del norte poseen datos atípicos que se encuentran en valores muy elevados. Europa posee el rango intercuartil mas alto, no posee datos atípicos y alcanza valores altos.

**Gráfico 6.** Diagrama de caja de esperanza de vida al nacer por continente.



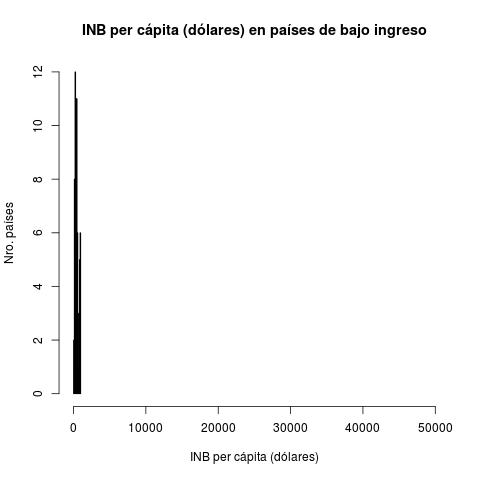
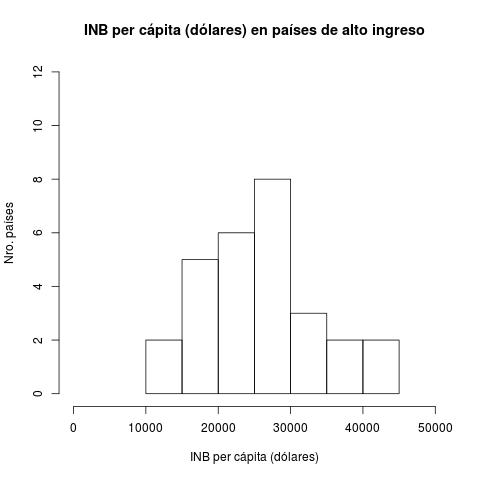
En el **gráfico 6** se muestra la esperanza de vida al nacer en años en cada continente. Esta vez se puede observar que América del norte, Asia y Oceanía poseen datos atípicos hacia valores bajos. La mayoría de los continentes mantienen sus valores entre los 60 y 80 años a excepción de África, cuyo rango intercuartil se encuentra ubicado entre valores muy bajos (40 y 60) en relación a los demás. África posee datos atípicos que están al nivel del resto de la población. Esto concuerda con los resultados previos debido a que África tiene una tasa de mortalidad bastante alta, por lo que se puede esperar una esperanza de vida baja, ya que esta última se calcula tomando como fija la tasa de mortalidad.

**Estudio de distribución considerando alto y bajo ingreso**



**Tabla 2.** INB per cápita para bajo y alto ingreso.

**Gráficos 7 y 8.** INB per cápita para países con bajo y alto ingreso

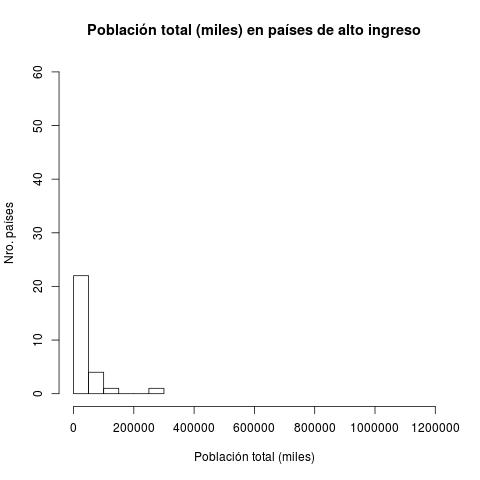
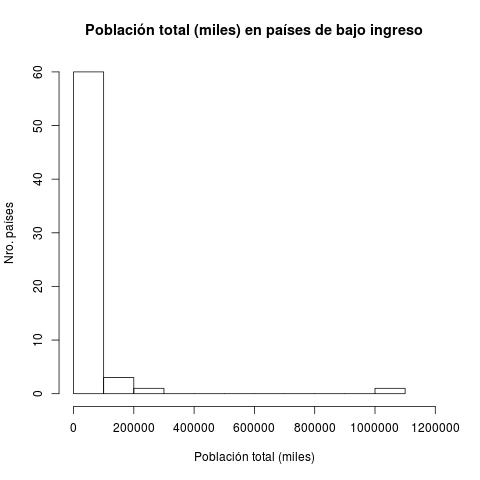


En la **Tabla 2** y en los **Gráficos 7 y 8** se puede apreciar un estudio completo del INB per cápita en países con bajo y alto ingreso. En la **Tabla 2** se puede observar que en los países con bajo ingreso se tienen valores que son cercanos entre sí pero dispersos dentro del rango que se obtuvo de esta variable. Esto se ve respaldado en el **gráfico 7** donde los datos en su totalidad no superan el valor de 10000. Con respecto a los países de alto ingreso, se observa una desviación estándar bastante alta al igual que un rango bastante amplio, por lo que se esperan resultados bastante dispersos y esto se puede comprobar mediante el **gráfico 8** donde gracias a la escala, se puede apreciar mucho mejor la dispersión de los datos procesados de la muestra.

**Tabla 3.** Población total en países con alto y bajo ingreso.

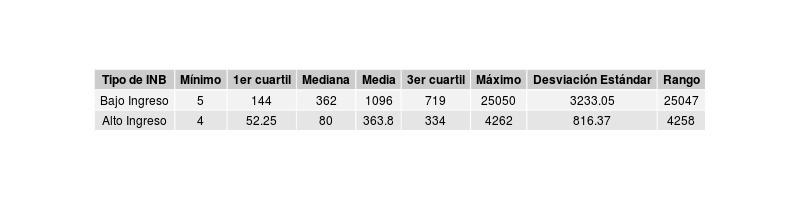


**Gráficos 9 y 10.** Población total para bajo y alto ingreso.

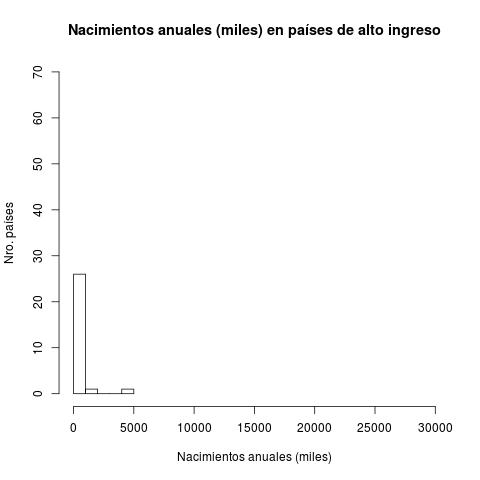
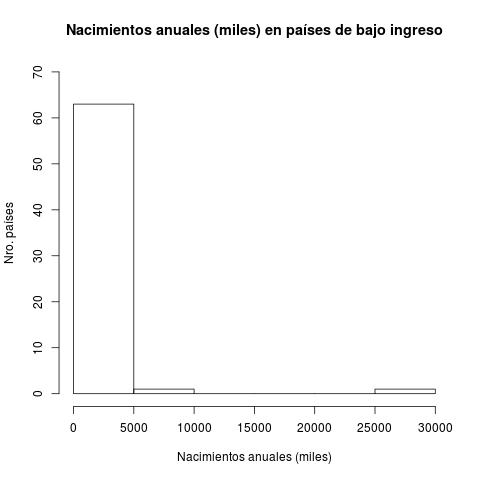


En la **tabla 3** se puede apreciar el estudio descriptivo de la población total de países con bajos y altos ingresos. Con respecto a los países de bajo ingreso, se puede observar que la población total es mas pequeña en la mayor cantidad de países de esta categoría, sin embargo, en el **gráfico 9** se puede apreciar que existe una cantidad de países (entre 1 y 10) cuya población total se ubica entre 1000 millones y 1200 millones de personas. En el **gráfico 10** se observan a los países con alto ingreso y se puede apreciar claramente que ningún país de esta categoría posee una población total mayor a 400 millones y presentan el mismo comportamiento que las de bajo ingreso (Gran parte de sus países tienen una cantidad baja de población total)

**Tabla 4.** Tabla de nacimiento anual para países con alto y bajo ingreso



**Gráficos 11 y 12.** Nacimiento anual para países con alto y bajo ingreso.



En la **tabla 4** se plantean los nacimientos anuales para los países de alto y bajo ingreso. Esta vez se observa que los mínimos en ambas categorías presentan valores muy cercanos y en los **gráficos 11 y 12** se puede ver que ambos grupos presentan un comportamiento similar pero con diferentes escalas. En cuanto a países de bajo ingreso, hay alrededor de 60 países en los que no se registran más de 5 millones de nacimientos al año pero hay entre 1 y 10 países que registran mas de 25 millones. Los países con altos ingresos en el peor de los casos a penas registran un poco mas de 5 millones de nacimientos al año, pero hay entre 20 y 30 países que no registran ni la mitad de dicha cantidad.

**Intervalos de confianza**

**Tabla 5.** Intervalo de confianza para media de INB per cápita.



Se realizó un intervalo de confianza del 95% para la media del Ingreso Nacional Bruto en países de bajo ingreso y otro intervalo de confianza, igualmente del 95%, para la media del INB en países de alto ingreso. Los resultados obtenidos fueron los siguientes:

**Países de bajo ingreso**

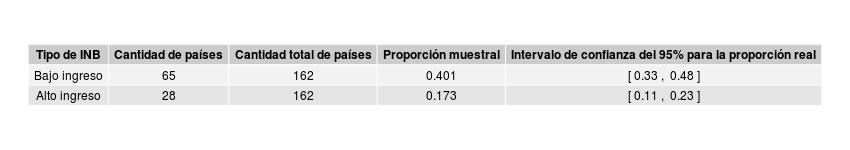
En esta categoría clasifican aquellos países cuyo Ingreso Nacional Bruto está por debajo de 1035 dólares anuales. De los datos suministrados por UNICEF 65 de los 162 países estudiados clasifican dentro de esta categoría. El estudio estadístico de dichos países determinó que la media muestral para el INB era de 448.62$ anuales con una desviación estándar muestral de 251.74$ en países de bajo ingreso. Con estos datos, y dado que el tamaño de la muestra era 65, se determinó el intervalo de confianza del 95% con una distribución normal de la media. Se obtuvo el intervalo [387.42, 509.81].

**Países de alto ingreso**

En esta categoría clasifican aquellos países cuyo INB está por encima de 12616$ anuales. De los datos suministrados por UNICEF 28 de los 162 países estudiados se encuentran dentro de esta categoría y se determinó que la media muestral para el INB era de 26010$ anuales con desviación estándar muestral de 8403.72$. Debido a que el tamaño de la muestra era menor a 30 se utilizó la distribución t de Student para determinar el intervalo de confianza. Se obtuvo [22751.38, 29268.62].

Se observó que casi la mitad de los países clasifican como bajo ingreso y sólo 28, menos de 1/5 del total, poseen alto ingreso. Más Aún, el pequeño grupo de países de alto ingreso posee un INB mucho mayor al de los países de bajo ingreso, aproximadamente 50 veces mayor, siendo esto muestra de la desequilibrada repartición de la riqueza a nivel mundial. Se puede concluir que la mayor parte de la población posee bajos ingresos y sólo es en un pequeño grupo de países en donde se encuentra la mayor concentración de riqueza.

**Tabla 6.** Intervalo de confianza para proporción entre países con bajo y alto ingreso.



Se determinó un intervalo de confianza del 95% para la proporción de países con bajo ingreso en relación a la cantidad total de países y otro intervalo de confianza, igualmente del 95%, para la proporción de países con alto ingreso. Se obtuvieron los siguientes resultados:

**Países de bajo ingreso**

De los datos suministrados por UNICEF 65 de los 162 países estudiados poseen ingreso anual menor a 1035$. El estudio estadístico de los datos determinó que la proporción muestral de países con bajo ingreso es 0.401. En base a la información obtenida de la muestra se obtuvo el intervalo de confianza del 95% [0.33, 0.48] para la proporción real.

**Países de alto ingreso**

A partir de los datos suministrados por UNICEF 28 de los 162 países estudiados poseen ingreso anual mayor a 12616$ y se determinó que la proporción muestral de países con alto ingreso es 0.173 del total de países. En base a la información obtenida de la muestra se obtuvo el intervalo de confianza del 95% [0.11, 0.23] para la proporción real de países con alto ingreso.

Se observó que, con 95% de confianza, entre 10% y 20% de los países clasifican como alto ingreso y entre 30% y casi 50% de los países clasifican como bajo ingreso. Por lo tanto, gran parte de los países poseen bajo ingreso y sólo una pequeña parte posee alto ingreso. Más aún, la cantidad de países con bajo ingreso es aproximadamente el doble de la cantidad de países con alto ingreso. Se puede concluir que existe una marcada diferencia de ingresos a nivel mundial definida por un pequeño grupo de países recibiendo grandes ingresos y una gran parte de países recibiendo bajos ingresos.

**BIBLIOGRAFIA**

* Dennis D. Wackerly, William Mendenhall, Richard L. Scheaffer - Mathematical statistics with applications - Thomson Brooks\_Cole (2008)
* RStudio – Open source and enterprise-ready professional software for R. <https://www.rstudio.com/>
* R: The R Project for Statistical Computing. <https://www.r-project.org/>
* <http://ccp.ucr.ac.cr/cursoweb/3411tmi.htm>
* http://definicion.de/tasa-de-natalidad/