



INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

## Mercado laboral y educación en Nuevo León

Yosshua Eli Cisneros Villasana (179889)

Mario Luis Bravo (183908)

12/05/2021

Inferencia Estadística

# Introducción

Desde los inicios del México independiente Nuevo León fue un estado que se ha favorecido por su cercanía a Estados Unidos. Según datos del INEGI para el 2019, ocupa el 3er lugar por su aportación al PIB con 7.8% y tiene un promedio de escolaridad de 10.7 años ocupando el segundo lugar (el promedio es 9.7), por lo cual es un estado importante en lo que respecta a la economía y educación. En ese sentido, sería de interés estudiar qué es lo que sucede dentro de su mercado laboral, cómo se beneficia de ello su población y como es en sí la misma.

Para el análisis de las diferentes variables que se analizarán se usarán los datos del Censo 2020 que hace el INEGI. Además, es importante tomar en cuenta que estos datos disten de la realidad presente debido a la afectación permanente que sufrió el país por la pandemia.

## Objetivo

El objetivo es hacer un análisis estadístico aplicado integrado por un análisis descriptivo y un análisis inferencial de las variables de ingresos mensuales y niveles de educación de la población del estado de Nuevo León a partir de lo que se aprendió en el curso de Inferencia Estadística en conjunto con la herramienta de software R. El análisis será realizado desde una dimensión económica, tratando de observar la situación de los trabajadores de la entidad a través de la que puede ser las principales variables indicadoras como su ingreso y su tiempo de trabajo. Dado que el fin de un análisis estadístico es poder sacar conclusiones que puedan servir para la toma de decisiones, trataremos de ser lo más explícitos posibles en las interpretaciones de los resultados.

## Lectura de datos

```
ruta<-"Personas19.CSV"

personas<- read.csv(ruta, sep=",", colClasses= c(INGTRMEN = "numeric", NIVACAD = "character",
TAMLOC = "character", TIE_TRASLADO_TRAB="character",
HORTRA="numeric"))[,c('INGTRMEN', 'NIVACAD', 'TAMLOC', 'TIE_TRASLADO_TRAB', 'HORTRA')]
```

```
df <- personas %>% filter(INGTRMEN <= 999998, !is.na(INGTRMEN))
```

```
dim.data.frame(personas)
```

```
## [1] 336666      5
```

El tamaño de la muestra es de 336666 personas. De esas personas se cuentan con 92 variables diferentes. De estas variables usaremos 3 para nuestro análisis: 1) Ingreso mensual 2) Nivel educativo 3) Horas trabajadas a la semana

Para iniciar el análisis descriptivo es necesario filtrar los datos:

```
personasdefinidas<-df%>%
  filter(HORTRA!=999)
dim.data.frame(personasdefinidas)
```

```
## [1] 134448      5
```

El tamaño de nuestra muestra se redujo a 134448 personas, las cuales ya tienen datos definidos.

# Análisis descriptivo univariado

## Ingreso mensual

Comenzaremos nuestro análisis descriptivo con algunos datos superficiales del ingreso mensual de las personas en Nuevo León:

```
summary(personasdefinidas$INGTRMEN)
```

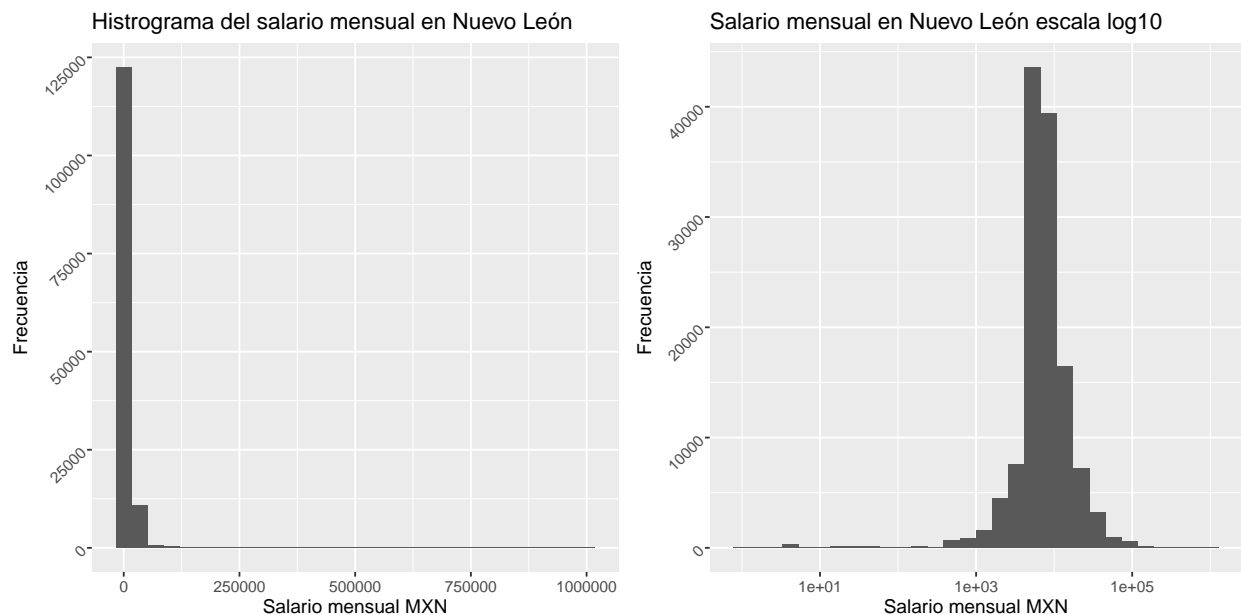
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         0   5000   6880   9615  10320 999998
```

```
sd(personasdefinidas$INGTRMEN)
```

```
## [1] 17286.2
```

Los datos indican que el promedio de los ingresos mensuales de las personas de Nuevo León es de 9615. De igual manera, la muestra nos indica que el 50% de las personas en Nuevo León ganan menos de 6880 pesos al mes con su trabajo, lo que es una cantidad bastante baja tomando en cuenta que la cantidad máxima de ingresos que registró la encuesta del INEGI es de 999998. Este dato es simbólico, ya que de acuerdo al Diccionario de datos de la encuesta hecha por el INEGI, 999998 representa a personas que ganaron más de 999,997 pesos al mes. Tomando en cuenta el hecho de que el 75% de las personas en Nuevo León ganan menos de 10320 pesos al mes, junto con la media y la mediana antes mencionadas, podemos intuir que es muy probable que en Nuevo León haya una minoría de personas con ingresos mensuales exorbitantes, mientras que la gran mayoría no los tiene. El dato adicional, de 17286.2 pesos, representa la desviación estándar de los ingresos de las personas de Nuevo León. Esta desviación estandar es bastante alta considerando que el promedio de ingresos mensuales no rebasa los \$10,000 al mes. Este último dato confirma la idea de que existen muchas diferencias de ingresos en Nuevo León respecto al promedio.

A continuación observamos un par de gráficos que revelan bastante sobre los salarios de las personas de Nuevo León:



Las dos gráficas muestran el salario mensual de los trabajadores de Nuevo León. La primera de ellas grafica la frecuencia absoluta en escala normal y la segunda en escala logarítmica. Tal como se observa en la primera gráfica, prácticamente la totalidad de las personas en Nuevo León obtienen un ingreso menor a 125,000 pesos mensuales. Este dato en sí mismo no es muy revelador dado que la escala de la gráfica no permite mayor precisión. Contrariamente, la gráfica en escala logarítmica permite visualizar con mayor precisión la distribución de los datos. Si bien existe una marcada concentración a los ingresos no extremos (centrales), se puede apreciar que en Nuevo León si existe gente con ingresos mucho menores al promedio y algunos con ingresos mayores.

## Horas trabajadas a la semana

Antes de pasar a la siguiente variable de análisis, limpiaremos los datos de la base del INEGI.

```
personashordef<-personas%>%  
  filter(HORTRA < 900)  
dim.data.frame(personashordef)
```

```
## [1] 135920      5
```

El número de personas que cumple nuestra restricción, o sea que tiene datos definidos para las horas que trabajan, es de 135920 personas.

```
summary(personashordef$HORTRA)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      0.00  40.00   48.00   43.56   48.00   140.00
```

```
sd(personashordef$HORTRA)
```

```
## [1] 15.53793
```

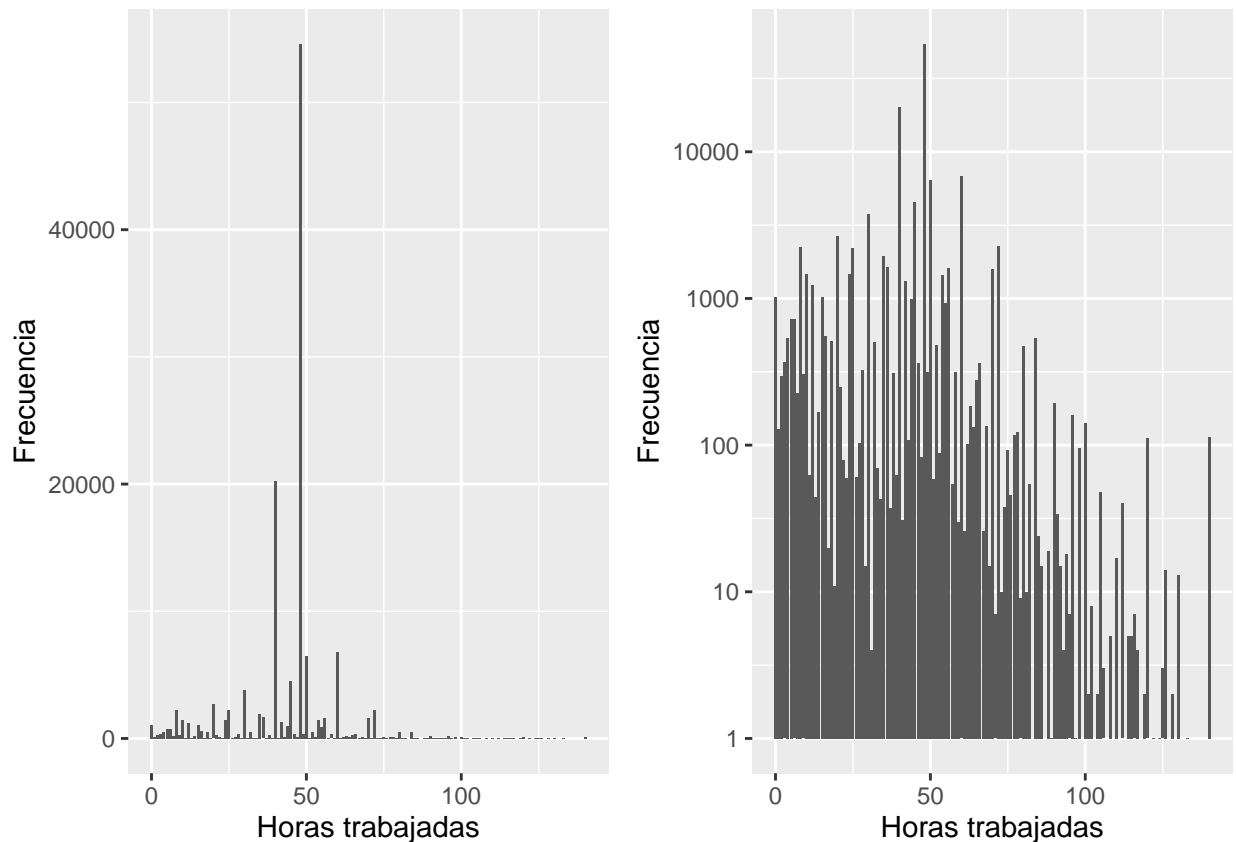
Los datos de las horas trabajadas por semana indican que en Nuevo León el 25% de las personas trabajan como máximo 40hrs por semana, o equivalentemente, 6.6 hrs trabajadas por día (tomando como referencia una semana de 6 días laborales). De igual manera, el 50% de las personas trabajan menos de 48hrs semanales, o equivalentemente, 8hrs diarias con la semana definida de la misma manera que en el caso anterior. Este dato es importante porque nos indica que solo el 50% de las personas trabaja como máximo de hrs al día las 8hrs de una jornada laboral. El otro 50% de los trabajadores en Nuevo León trabaja más de lo establecido por una jornada laboral. Es muy significativo el hecho de que en Nuevo León existe gente que trabaja 140hrs a la semana. Tomando como referencia la misma semana laboral de 6 días, los datos indican que estas personas trabajan por 23.3hrs al día. Esto es muestra de que es posible que exista un ambiente de explotación laboral, lo cual es solo una hipótesis ya que no se puede concluir nada de los tipos de trabajo a los que las personas de Nuevo León se enfrentan, únicamente teniendo esta información.

Es muy llamativo el hecho de que el 3er cuartil, o sea el 75% de las personas que menos trabajan en Nuevo León, cumplen exactamente el mismo número de horas que el 50% que menos trabaja, lo que nos indica que un muy buen porcentaje de personas trabaja menos de 48hrs a la semana.

Es igual de relevante notar el hecho de que la desviación estandar de las horas trabajadas por semana en Nuevo León es de 15.5379hrs. Esta cantidad es elevada, lo que significa que sí hay algo de variabilidad en las horas trabajadas en Nuevo León.

Mostramos de igual manera los gráficos siguientes con la distribución de los datos de las horas trabajadas por semana en Nuevo León:

```
h1 <- ggplot(personashordef, aes(x=HORTRA)) + geom_histogram(stat="count") +
  labs(x="Horas trabajadas", y="Frecuencia")
h2 <- ggplot(personashordef, aes(x=HORTRA)) + geom_histogram(stat="count") +
  scale_y_log10()+labs(x="Horas trabajadas", y="Frecuencia")
multiplot(h1,h2, cols = 2)
```

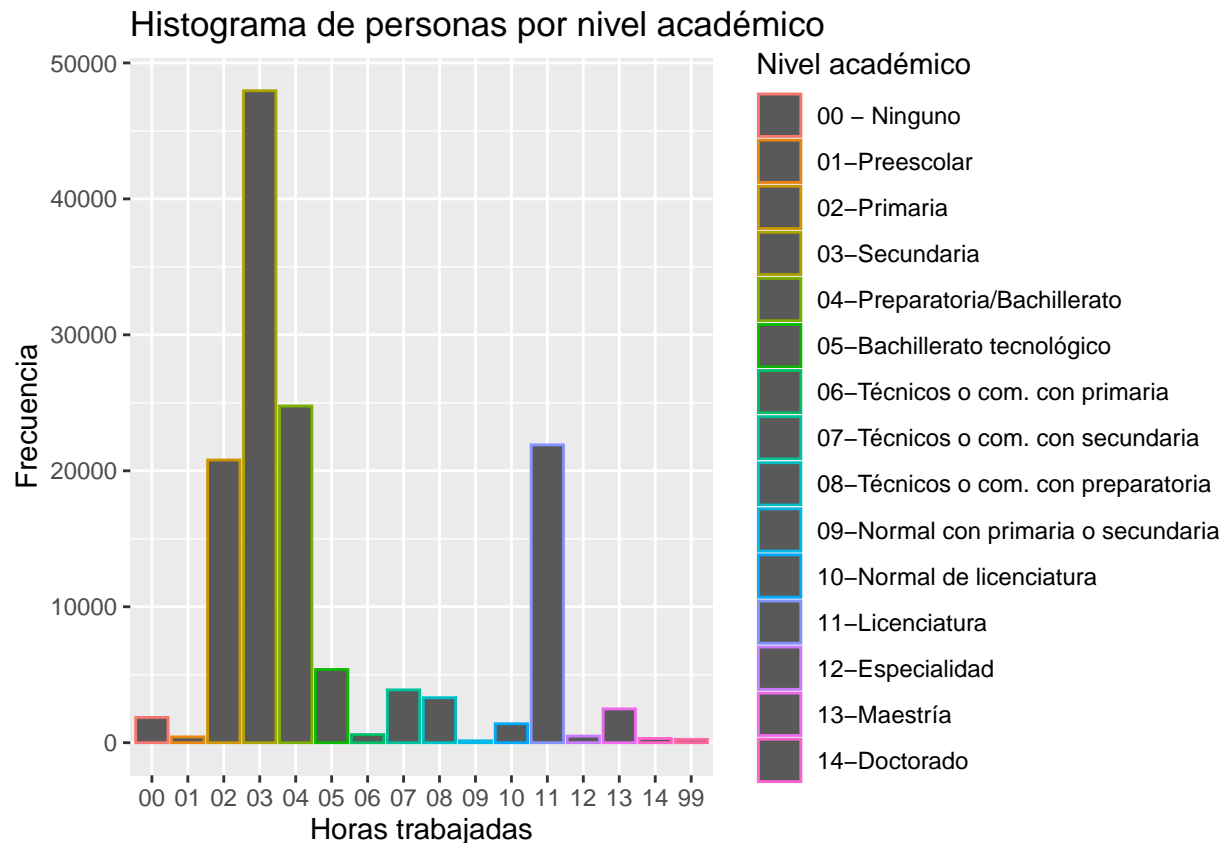


De acuerdo a ambas gráficas la moda de la muestra es de 48hrs a la semana. Esto es congruente con los datos mostrados por la tabla anterior en la que se nos muestra que tanto la mediana como el 3er cuartil de los datos es de 48hrs, lo cual es claramente muestra de una gran concentración de los mismos en este dato. Los datos de las horas trabajadas en la primera gráfica muestran que existe una ligeramente mayor frecuencia en las horas menores a la moda, en comparación a las cantidades de horas mayores. Dado que 48hrs es la mediana y el tercer cuartil, el dato nos indica que el 50% de las personas de la muestra se encuentran debajo de este número, mientras que solo el 25% de las personas trabajan más de 48hrs a la semana. Es llamativa igualmente la concentración de datos en las 40hrs a la semana. Poco más de 20000 personas de la muestra trabajan 40hrs por semana. A pesar de la clara concentración de datos en horas menores a 48 en nuestra muestra, no debe ignorarse el hecho de que el rango de horas trabajadas es bastante amplio. Existe un número de personas, si bien relativamente pequeño, que trabajan 140hrs a la semana. Como también lo existe, aunque un poco mayor, de personas que no trabajan en lo absoluto.

## Nivel académico

Para complementar, mostramos este historgrama con la frecuencia de personas por nivel académico. De acuerdo al diccionario del INEGI, los niveles académicos van desde 0 que representa “Ningún nivel académico” hasta 14 que representa “Doctorado”. Es importante mencionar que el nivel 99 representa a los "no identificados".

```
ggplot(personashordef, aes(x=NIVACAD, color = NIVACAD)) + geom_histogram(stat="count") +  
  labs(x="Horas trabajadas", y ="Frecuencia") + ggtitle("Histograma de personas por nivel académico")+s
```



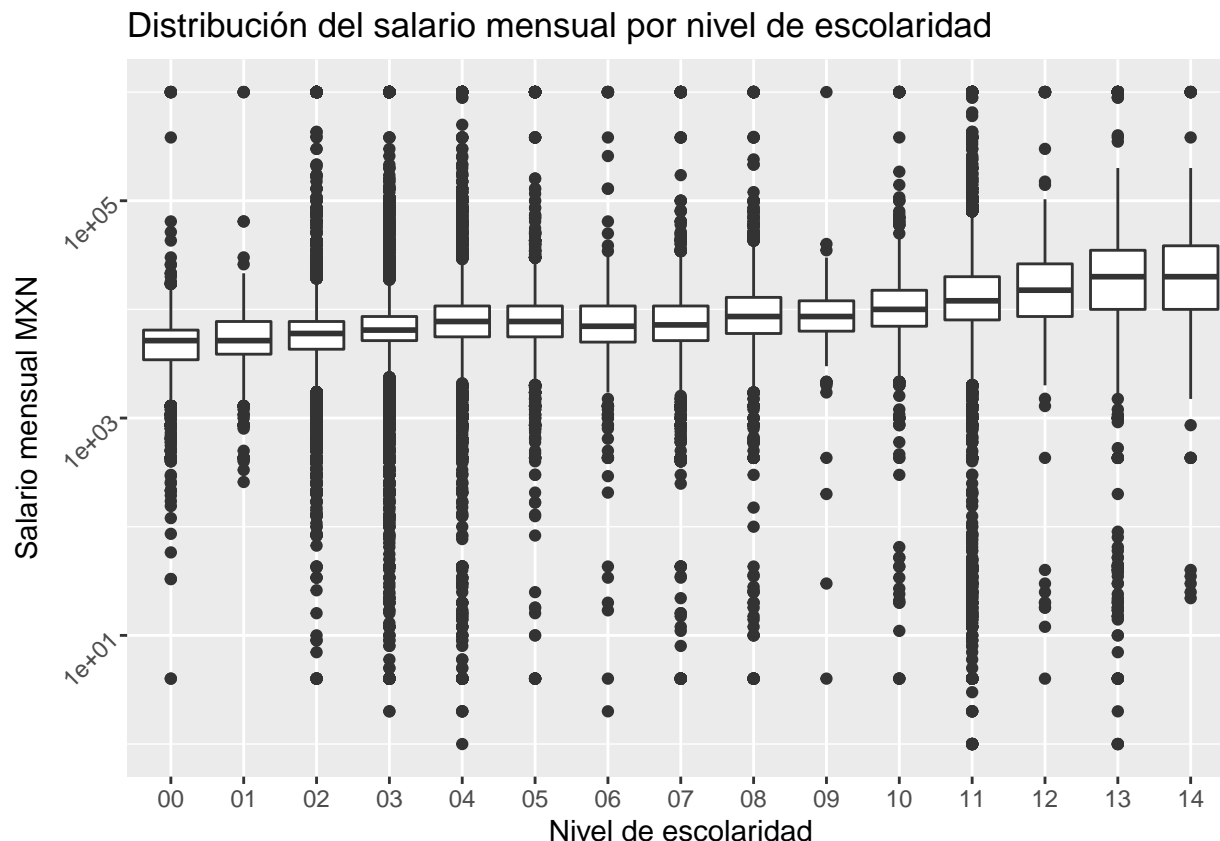
Tal como se puede observar en la gráfica, un gran número de personas se encuentra entre el nivel 3 y el 5, que representa la secundaria y el bachillerato tecnológico, respectivamente. Sin embargo, en Nuevo León existe todavía mucha gente sin ningún grado de escolaridad, y aún más con grados como primaria o secundaria. Hay que reconocer, a pesar de todo lo demás, que existe un gran número de gente con la universidad terminada en Nuevo León, al menos en comparación a otros grados de estudio menores como los estudios técnicos con prepa terminada o las escuelas normales, o en comparación a grados superiores como especialidades, maestrías o doctorados.

## Análisis descriptivo bivariado

Antes de pasar al análisis Inferencial, es necesario establecer una relación importante entre dos variables reveladoras desde una dimensión socioeconomica. En este sentido, analizaremos la relación entre el nivel de escolaridad y el salario mensual.

## Relación entre el salario mensual y nivel de escolaridad

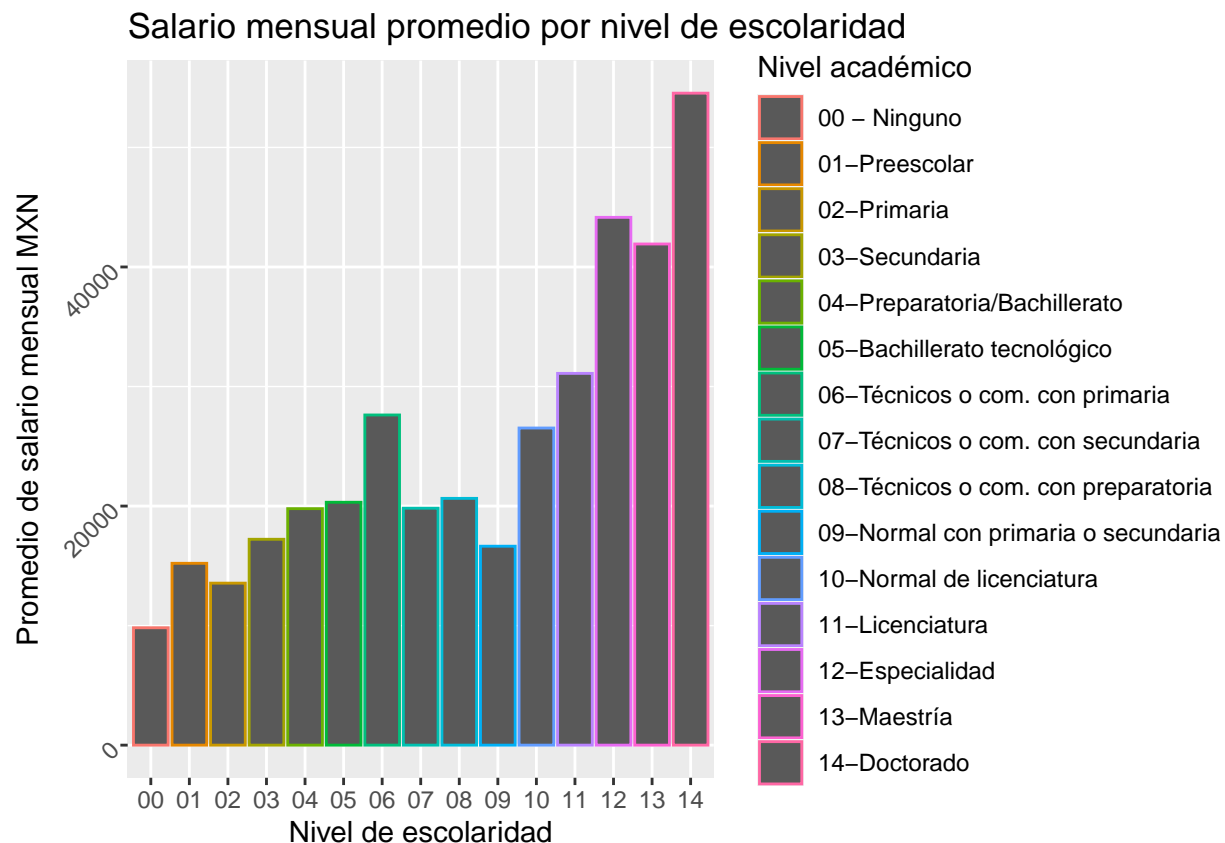
La conversación popular nos lleva a creer que existe una relación directa entre el nivel de escolaridad con los ingresos percibidos. Es decir, que entre más estudie uno, más dinero ganará. ¿Pero es esto cierto? A continuación se muestran un par de gráficas con los datos de ambas variables:



Esta gráfica es extraordinariamente informativa. En primer lugar podemos notar que es claro que las personas que tienen mayor grado de escolaridad tienden a ganar más dinero. Incluso, la gráfica nos muestra que el primer cuartil de los ingresos del nivel más alto de escolaridad es mayor que el tercer cuartil de los ingresos del nivel más bajo. Es interesante notar que a pesar de tener la mediana más alta entre todos los niveles de escolaridad, el nivel 14 o de doctorado es el nivel de escolaridad con la mayor dispersión de los datos. Esto se puede notar porque el rango intercuartílico es el más grande de entre todas los niveles de escolaridad. Esto nos indica de que existen ingresos muy diversos entre las personas que poseen un doctorado, ya que existen incluso una cantidad considerable de putliers con gente que gana mucho menos de la mediana y gente que gana mucho más. Ni siquiera un doctorado trae seguridad económica consigo en Nuevo León, jaja! Contrariamente, el nivel con la menor dispersión de ingresos en el estado de Nuevo León es el nivel de secundaria (nivel 003). Lo cual se nota por la amplitud en los bigotes de la gráfica en su variable y el compactísimo rango intercuartílico. Esto significa que hay muchísima gente con secundaria que gana cantidades no muy diferentes a la mediana de los ingresos en Nuevo León. Naturalmente, un rango intercuartílico tan compacto genera una cantidad grande de outliers, lo cual indica de que a pesar de que mucha gente tiene ingresos cercanos a la mediana, también hay personas que ganan ingresos mucho menores e ingresos mucho mayores teniendo la secundaria. La misma situación es aplicable a la primaria, que sin embargo, posee una mediana ligeramente menor y una muy ligera mayor dispersión de ingresos.

Un fenómeno interesante que puede visualizarse en esta gráfica es el hecho de que en todos los niveles de escolaridad hay personas con el máximo de ingresos, o en su defecto ligeramente por debajo, incluso en el caso de las personas que no tienen ningún nivel de escolaridad (nivel 00). Sin embargo, para el caso de la gente con Doctorados, no existe nadie que obtenga el mínimo de ingresos, ni de cerca. Parece ser que el

Doctorado sí sirve a final de cuentas. Para el caso de los licenciados la situación no es tan alentadora.(nivel 11) Si bien su mediana es más alta que la de niveles de educación inferior, existe una mayor variabilidad de los datos respecto a la mediana en comparación a niveles inferiores, y existe una gran cantidad de outliers por encima y por debajo de los bigotes de la variable. Esto nos indica que hay bastantes personas que ganan mucho menos dinero al mes por trabajo y gente que gana mucho más.



Tal como muestra la gráfica de barras, existe una relación directa clara entre los ingresos y el nivel académico o escolaridad. Los datos indican que a mayor nivel académico (más preparación y años de estudio) mayor es el promedio del sueldo mensual, al menos en el estado de Nuevo León ocurre este fenómeno con certeza. De hecho, es interesante el hecho de que existe una diferencia significativa entre el promedio de ingresos de licenciatura y el promedio de ingresos de especialidad, siendo mucho mayores los segundos, y aún otra mayor entre maestría y doctorado. Es curioso el hecho de que en el estado de Nuevo León, hacer una maestría, habiendo completado una especialidad, representa una disminución en el salario promedio. El hecho de que exista una diferencia considerable entre los salarios promedio de licenciados y licenciados especializados puede deberse a la escasez relativa de estos últimos respecto a los primeros. Esto visto desde una perspectiva meramente económica. De igual manera, puede ser este mismo fenómeno económico uno de los determinantes del salario promedio tan elevado de los Doctores en Nuevo León, en comparación a unos abundantes maestros en la misma entidad.



## Análisis inferencial

La siguiente parte de nuestro análisis corresponde a un análisis inferencial dividido en dos partes: estimaciones por intervalo y pruebas de hipótesis de diferentes parámetros de la población.

### Estimación por intervalos

Antes de iniciar la estimación por intervalos es necesario mencionar que estas se harán con un 95% de confianza. La primera de ellas corresponderá a una estimación de dos colas de la media de ingresos laborales en Nuevo León. La segunda será correspondiente a la media de horas trabajadas en el estado de Nuevo León por los distintos grupos de trabajadores.

#### Media de ingreso mensual

Con 95% de confianza el promedio de ingreso por trabajo de las personas en Nuevo León

$$\mu_{ing} \in (9513.9749166, 9698.5471118)$$

En este sentido, puede afirmarse, con alto nivel de confianza, de que el salario promedio de las personas en Nuevo León no es tan alto en realidad pues se encuentra dentro de dicho intervalo de valores. Es interesante notar que el rango otorgado por el crítico inferior y el superior no es de más de 200 pesos. Tomando en cuenta que los valores inferior y superior son de casi 10000 pesos, se puede decir con un 95% de confianza que nos acercamos bastante a un valor preciso.

#### Media de horas trabajadas

De igual manera que con el ingreso laboral de las personas de Nuevo León, quisimos estimar el promedio de horas trabajadas por las personas de la entidad. Para este caso, utilizamos de igual manera un 95% de confianza. Esta estimación dio como resultado que con el 95% de confianza el promedio de horas trabajadas en Nuevo León

$$\mu_{horas} \in (44.7568013, 45.1797424)$$

Es muy interesante el hecho de que el intervalo sea muy chico, lo cual nos indica que con el 95% de confianza tenemos bastante precisión en nuestra estimación.

#### Diferencia de medias de ingreso entre personas que estudiaron licenciatura o más y las que no

El resultado es que con 95% de confianza podemos afirmar que la diferencia del promedio de ingresos entre personas que estudiaron la licenciatura o un grado superior y las que estudiaron hasta grados menores de licenciatura es el siguiente:

$$\mu_{Lic} - \mu_{sinLic} \in (9153.7426071, 9915.4554438)$$

Este intervalo quiere decir que el promedio de los ingresos por trabajo de nuestro grupo de educación superior difiere entre 9153.7426071 y 9915.4554438 pesos del promedio de ingresos de gente que no tiene educación superior en el estado de Nuevo León, con un 95% de confianza. Dado que la diferencia es positiva y alcanza como mínimo 9153 pesos, se puede concluir que existe entrar en la educación superior hizo una diferencia clara en el salario de los nuevoleonenses, todo esto tomando en cuenta que manejamos únicamente un 95% de confianza. Sin embargo, a pesar de que no hay seguridad absoluta, los datos sí muestran una diferencia favorable a las personas con educación superior en su ingreso promedio.

## Prueba de hipótesis

### 1) Hipótesis de horas de trabajo por nivel de escolaridad

Tal como en nuestra última estimación por intervalos, dividimos el grupo de trabajadores nuevoleonese en dos partes: aquellos que tienen un grado igual o superior a la licenciatura y aquellos que tienen grados menores a la misma. Nos interesa probar que el número de horas trabajadas entre los dos grupos es diferente entre sí. Para esto, tomamos como hipótesis nula la igualdad del promedio de horas trabajadas por ambos grupos y como alternativa la no igualdad de manera que nuestras hipótesis son las siguientes:

$$H_0 : \mu_L - \mu_{NL} = 0$$

$$H_1 : \mu_L - \mu_{NL} \neq 0$$

donde  $\mu_L$  representa el promedio de horas trabajadas por personas que tiene al menos Licenciatura y  $\mu_{NL}$  representa el promedio de horas trabajadas por personas que tienen un grado menor que una Licenciatura.

### 2) Nivel de significancia

Tomaremos un nivel de significancia del .05, osea un nivel de confianza de la prueba de .95. De forma que tenemos lo siguiente:

$$\alpha = 0.05$$

### 3) Estadístico

Como  $n$  es muy grande podemos decir, sin pérdida de generalidad, que  $S^2 \approx \sigma^2$ . Entonces,

$$Z = \frac{(\bar{\mu}_L - \bar{\mu}_{NL}) - (\mu_L - \mu_{NL})}{\sqrt{\frac{\sigma_{\mu_L}^2}{n_L} + \frac{\sigma_{\mu_{NL}}^2}{n_{NL}}}} = \frac{(43.2118518 - 45.3927626) - 0}{\sqrt{\frac{1438.2465771}{26207} + \frac{1597.7499192}{108437}}} = -8.26$$

### 4) Región de rechazo

Como  $Z_{0.025} = -1.96 \Rightarrow RR : (-\infty, -1.96) \cup (1.96, \infty)$

### 5) Conclusión

Dado que  $Z \in RR \Rightarrow$  Rechazamos  $H_0$  y con 95% de confianza  $\mu_L \neq \mu_{NL}$ . Es decir, no trabajan el mismo número de horas con solo un .05 de significancia.

Dicho de otra manera, como nuestro estadístico pertenece a la región de rechazo otorgada por el nivel de significancia deseado, es claro que se rechaza la hipótesis nula (que nos dice que el promedio de horas trabajadas es el mismo) y por consecuencia no se rechaza la hipótesis alternativa que indica que son diferentes las horas trabajadas. De esta hipótesis no se puede sacar la conclusión de que el promedio de horas trabajadas de uno de los grupos seleccionados sea mayor o menor que el del otro, sino solamente que son diferentes unos y otros.

### 1) Hipótesis de ingresos mensuales por nivel de escolaridad

Como segunda prueba de hipótesis queremos confirmar el hecho de que sí existe diferencia en los ingresos promedio de los dos grupos anteriores: aquellos que cuentan con educación superior (licenciatura o más grados) y aquellos que no cuentan con educación superior. Por este motivo, tomamos como hipótesis nula la igualdad entre los promedios de ingreso entre ambos grupos e hipótesis alternativa la no igualdad:

$$H_0 : \mu_L - \mu_{NL} = 0$$

$$H_1 : \mu_L - \mu_{NL} \neq 0$$

donde  $\mu_L$  representa el promedio de horas trabajadas por personas que tiene al menos Licenciatura y  $\mu_{NL}$  representa el promedio de horas trabajadas por personas que tienen un grado menor que una Licenciatura.

### 2) Nivel de significancia

Al igual que en nuestra prueba anterior, tomamos un nivel de significancia del .05, lo que nos indica que

$$\alpha = 0.05$$

### 3) Estadístico

Como  $n$  es muy grande podemos decir, sin pérdida de generalidad, que  $S^2 \approx \sigma^2$ . Entonces,

$$Z = \frac{(\bar{\mu}_L - \bar{\mu}_{NL}) - (\mu_L - \mu_{NL})}{\sqrt{\frac{\sigma_{\mu_L}^2}{n_L} + \frac{\sigma_{\mu_L}^2}{n_{NL}}}} = \frac{(1.7285053 \times 10^4 - 7750.4543191) - 0}{\sqrt{\frac{9.6035094 \times 10^8}{26207} + \frac{1.208733 \times 10^8}{108437}}} = 49.08$$

### 4) Región de rechazo

Con el alpha requerido obtenemos la siguiente región de rechazo:

$$\text{Como } Z_{0.025} = -1.96 \implies RR : (-\infty, -1.96) \cup (1.96, \infty)$$

### 5) Conclusión

Dado que  $Z \in RR \implies$  Rechazamos  $H_0$  y con 95% de confianza  $\mu_L \neq \mu_{NL}$ . Es decir, en promedio no tienen el mismo ingreso.

Dicho de otra manera, dado que nuestro estadístico cae en la región de rechazo, se rechaza la hipótesis nula y por lo tanto no se rechaza la alternativa que es correspondiente a la no igualdad en los promedios. Esto nos dice que efectivamente si existe evidencia para afirmar que el promedio de ingresos obtenidos por gente con educación superior no es igual al promedio de ingresos obtenidos por gente sin esa educación superior, al menos en Nuevo León y con 95% de confianza. De nuevo, de esta prueba no se puede concluir cuál de los promedios es mayor al otro, sino solamente que son diferentes. Intuitivamente, lo más lógico es pensar que el de las personas con educación superior es mayor, sin embargo, aunque puede que sea cierto, ese resultado no se deriva de esta prueba de hipótesis.

## Conclusión

Para concluir nuestro trabajo consideramos fundamental recordar la relevancia de los análisis estadísticos. Un análisis de esta naturaleza obtiene su importancia del hecho de la necesidad de la toma de decisiones fundamentadas. Por estos motivos, la estadística, como rama de las matemáticas, se ha dedicado a desarrollar métodos para analizar los hechos a través de datos como los utilizados en nuestro trabajo. Es claro que en la realidad mexicana la situación económica, social y política es complicada. Parte de esta situación es consecuencia de una serie de decisiones tomadas sin fundamento alguno, ignorando la realidad mostrada por los datos en infinidad de dimensiones. El caso de Nuevo León es relevante puesto que representa uno de los estados más grandes del país en términos de población, aportación al PIB, entre otros aspectos. Es por esto que es especialmente necesario tomar las decisiones adecuadas para promover su desarrollo como entidad, fundamentándolas en análisis serios.

El hecho de que el promedio de los ingresos en ese estado se encuentre en un rango que no sobrepasa los 10,000 pesos muestra la dificultad de la situación en el que es uno de los estados más ricos del país, y es señal de la necesidad de llevar a cabo acciones que corrijan estos aspectos de la realidad de esta entidad para promover su desarrollo y con ello mismo el del país.

Análisis hechos a través de programas como R brindan una base sólida para la realización de acciones encaminadas no solo a mejorar la realidad mexicana en las distintas dimensiones que sean necesarias, sino para cualquier decisión fundamentada en cualquier ámbito de empresas, familias, personas, entre otros.