

PROYECTO INTEGRADOR

ANÁLISIS DE DATOS MEDIANTE ESTADÍSTICA DESCRIPTIVA

ESTUDIO SOBRE HÁBITOS DE OCIO

Fecha: 31 / Enero / 2022

Nombre de los estudiantes: BRYAN MANUEL MARTINEZ OLVERA

ALEJANDRO JAHZEEL ESAU TAPIA BUENO

JESUS ALFREDO BECERRIL ORTA

HANNIA NAYELI MC. DONOUGH ONTIVEROS

Nombre del docente: VICTOR MENDOZA MARTINEZ

1. Introducción

El área de bienestar ocupacional de la empresa Automotriz González realiza una encuesta a cierto número de empleados a las cuales se les preguntaron sus hábitos de ocio, principalmente enfocados en el tiempo que dedican en un día a dos actividades lúdicas.

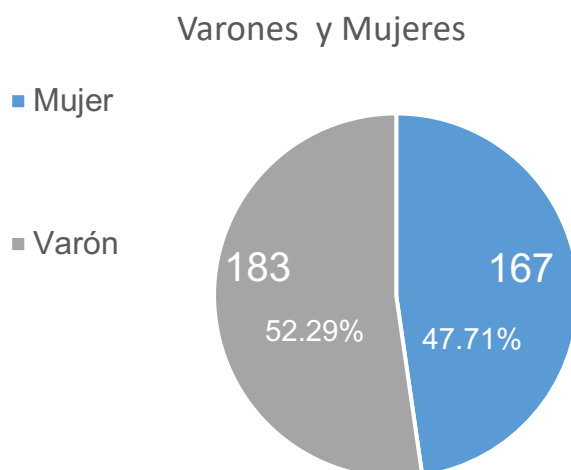
- Número de horas dedicadas a leer en el día anterior a la encuesta.
- Número de horas dedicadas a ver televisión en el día anterior a la encuesta.
- Su situación laboral en el día anterior a la encuesta.
- Número de horas que pasaron realizando alguna otra actividad en el día anterior de la encuesta.
- Obtiene el sexo y edad del personal.

2. Composición de la muestra

Distribución por sexo

La muestra se compone por mas varones que mujeres por una diferencia de 6 hombres mas

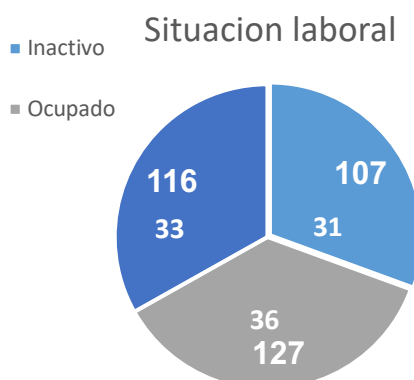
| Sexo | Frecuencia absoluta | Frecuencia relativa |
|--------------|---------------------|---------------------|
| Mujer | 167 | 47.71% |
| Varón | 183 | 52.29% |
| Total | 350 | 100% |



Distribución por situación laboral

La mayor parte de personas se encuentran ocupadas siendo el 36% de la muestra

| Situación laboral | Frecuencia absoluta | Frecuencia Relativa |
|-------------------|---------------------|---------------------|
| Inactivo | 107 | 31% |
| Ocupado | 127 | 36% |
| Desempleado | 116 | 33% |
| Total | 350 | 100% |

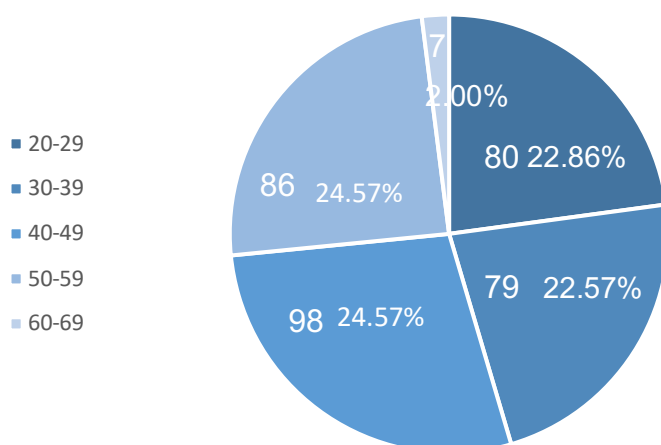


Distribución por edades

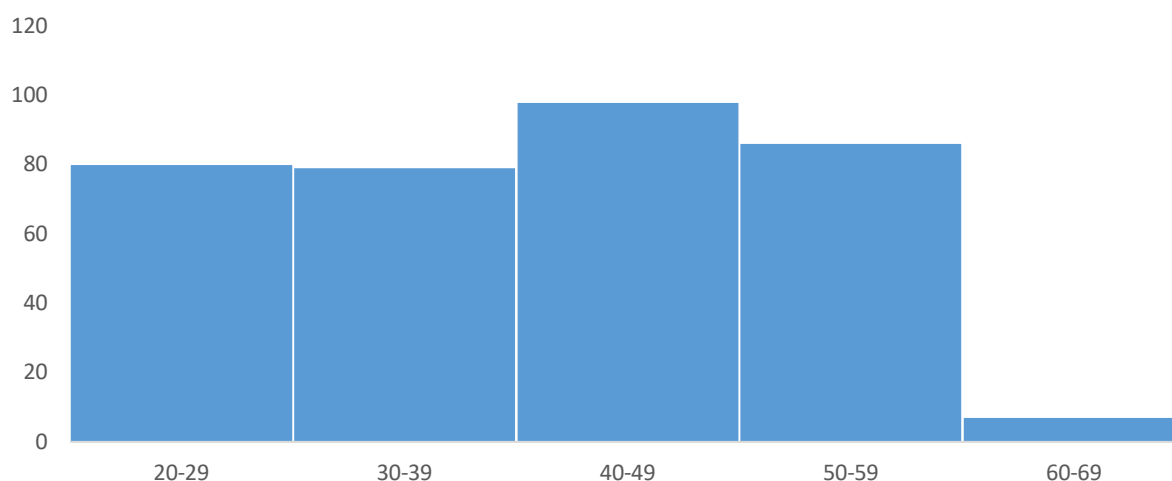
Validamos que entre la mayor parte de los encuestados se encuentra entre los 40y 49 años de edad

| Edad | Frecuencia absoluta | Frecuencia Relativa |
|--------------|---------------------|---------------------|
| 20-29 | 80 | 22.86% |
| 30-39 | 79 | 22.57% |
| 40-49 | 98 | 28.00% |
| 50-59 | 86 | 24.57% |
| 60-69 | 7 | 2.00% |
| Total | 350 | 100% |

Grupo de Edades



Grupos por edad

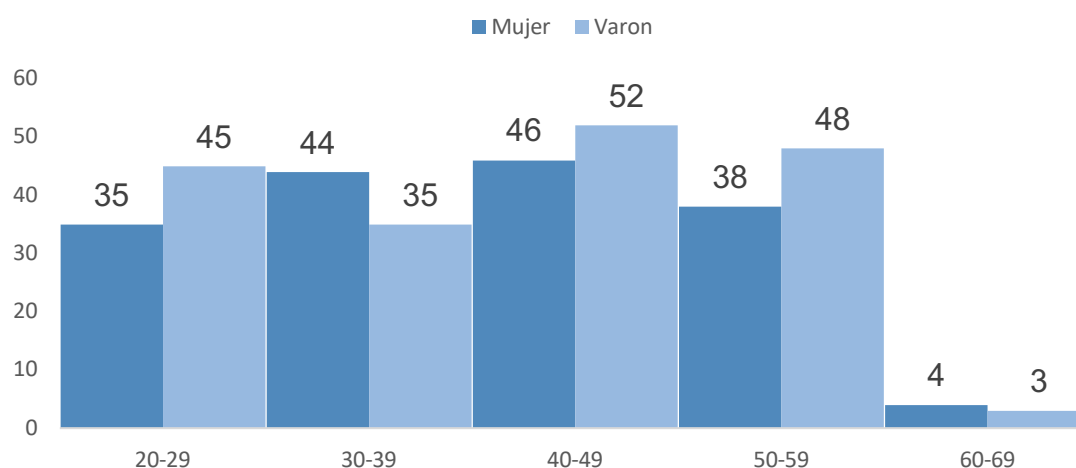


Distribución por edad y sexo

La mayor parte de los encuestados se encuentra entre los 40 y 49 años en ambos sexos

| Sexo | Edades | | | | | Total |
|--------------|-----------|-----------|-----------|-----------|----------|------------|
| | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | |
| Mujer | 35 | 44 | 46 | 38 | 4 | 167 |
| Varón | 45 | 35 | 52 | 48 | 3 | 183 |
| Total | 80 | 79 | 98 | 86 | 7 | 350 |

Distribución edad y sexo



Situación laboral y grupos de edad

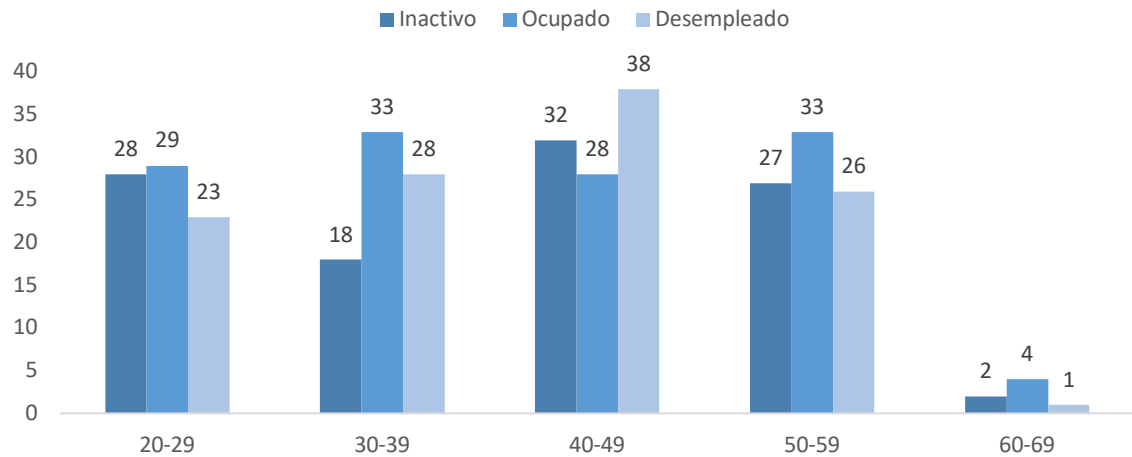
Podemos ver que el mayor número de desempleados se encuentran en el rango de los 40 años de edad

Y mayor parte de ocupados se encuentran entre las edades de 30 a 39 y de 50 a 59 años de edad

| Situación Laboral | Edades | | | | | Total |
|--------------------|-----------|-----------|-----------|-----------|----------|------------|
| | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | |
| Inactivo | 28 | 18 | 32 | 27 | 2 | 107 |
| Ocupado | 29 | 33 | 28 | 33 | 4 | 127 |
| Desempleado | 23 | 28 | 38 | 26 | 1 | 116 |
| Total | 80 | 79 | 98 | 86 | 7 | 350 |



Situación laboral y grupos de edad



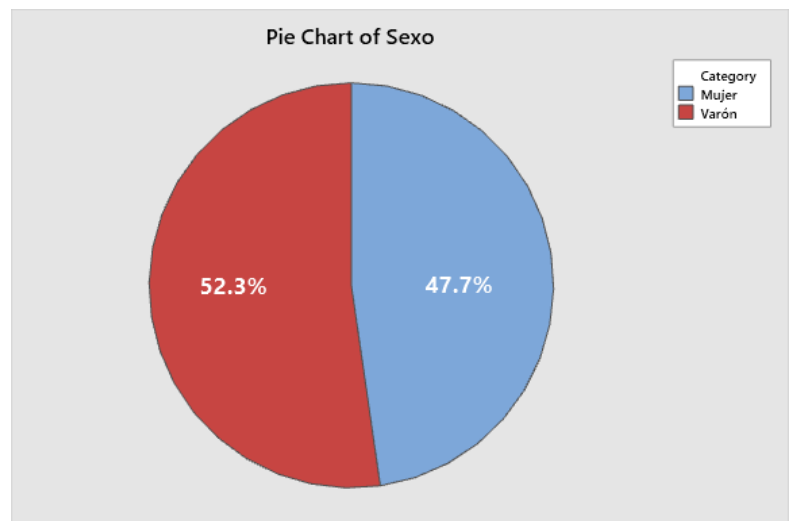
3. Resultados

1. ¿Es fácil determinar cuantos hombres o mujeres se encuestaron?

Si, fácilmente utilizando la base de datos ya obtenida, podemos filtrar la información dentro de las diferentes herramientas y así obtener un resultado.

Tally

| Sexo Count | |
|------------|-----|
| Mujer | 167 |
| Varón | 183 |
| N= | 350 |



- Realiza una tabla de frecuencias para el rubro sexo, situación laboral y edad.

Utilizando la herramienta Minitab, abriremos el archivo con todos los datos obtenidos de la encuesta y de esta manera podemos generar una tabla de frecuencia.

Tabla de Frecuencia

| Sexo | Count | Percent | CumCnt | CumPct | Edad | Count | Percent | CumCnt | CumPct |
|-------|-------|---------|--------|--------|------|-------|---------|--------|--------|
| Mujer | 167 | 47.71 | 167 | 47.71 | 20 | 4 | 1.14 | 4 | 1.14 |
| Varón | 183 | 52.29 | 350 | 100.00 | 21 | 12 | 3.43 | 16 | 4.57 |
| N= | 350 | | | | 22 | 3 | 0.86 | 19 | 5.43 |
| | | | | | 23 | 10 | 2.86 | 29 | 8.29 |
| | | | | | 24 | 9 | 2.57 | 38 | 10.86 |
| | | | | | 25 | 8 | 2.29 | 46 | 13.14 |
| | | | | | 26 | 10 | 2.86 | 56 | 16.00 |
| | | | | | 27 | 9 | 2.57 | 65 | 18.57 |
| | | | | | 28 | 8 | 2.29 | 73 | 20.86 |
| | | | | | 29 | 7 | 2.00 | 80 | 22.86 |
| | | | | | 30 | 5 | 1.43 | 85 | 24.29 |
| | | | | | 31 | 10 | 2.86 | 95 | 27.14 |
| | | | | | 32 | 5 | 1.43 | 100 | 28.57 |
| | | | | | 33 | 4 | 1.14 | 104 | 29.71 |
| | | | | | 34 | 13 | 3.71 | 117 | 33.43 |
| | | | | | 35 | 6 | 1.71 | 123 | 35.14 |
| | | | | | 36 | 10 | 2.86 | 133 | 38.00 |
| | | | | | 37 | 9 | 2.57 | 142 | 40.57 |
| | | | | | 38 | 14 | 4.00 | 156 | 44.57 |
| | | | | | 39 | 3 | 0.86 | 159 | 45.43 |
| | | | | | 40 | 11 | 3.14 | 170 | 48.57 |
| | | | | | 41 | 9 | 2.57 | 179 | 51.14 |
| | | | | | 42 | 5 | 1.43 | 184 | 52.57 |
| | | | | | 43 | 10 | 2.86 | 194 | 55.43 |
| | | | | | 44 | 5 | 1.43 | 199 | 56.86 |
| | | | | | 45 | 17 | 4.86 | 216 | 61.71 |
| | | | | | 46 | 10 | 2.86 | 226 | 64.57 |
| | | | | | 47 | 11 | 3.14 | 237 | 67.71 |
| | | | | | 48 | 7 | 2.00 | 244 | 69.71 |
| | | | | | 49 | 13 | 3.71 | 257 | 73.43 |
| | | | | | 50 | 8 | 2.29 | 265 | 75.71 |
| | | | | | 51 | 7 | 2.00 | 272 | 77.71 |
| | | | | | 52 | 13 | 3.71 | 285 | 81.43 |
| | | | | | 53 | 10 | 2.86 | 295 | 84.29 |
| | | | | | 54 | 4 | 1.14 | 299 | 85.43 |
| | | | | | 55 | 11 | 3.14 | 310 | 88.57 |
| | | | | | 56 | 5 | 1.43 | 315 | 90.00 |
| | | | | | 57 | 13 | 3.71 | 328 | 93.71 |
| | | | | | 58 | 6 | 1.71 | 334 | 95.43 |
| | | | | | 59 | 9 | 2.57 | 343 | 98.00 |
| | | | | | 60 | 7 | 2.00 | 350 | 100.00 |
| | | | | | N= | 350 | | | |

| Situación laboral | Count | Percent | CumCnt | CumPct |
|-------------------|-------|---------|--------|--------|
| Desempleado | 116 | 33.14 | 116 | 33.14 |
| Inactivo | 107 | 30.57 | 223 | 63.71 |
| Ocupado | 127 | 36.29 | 350 | 100.00 |
| N= | 350 | | | |



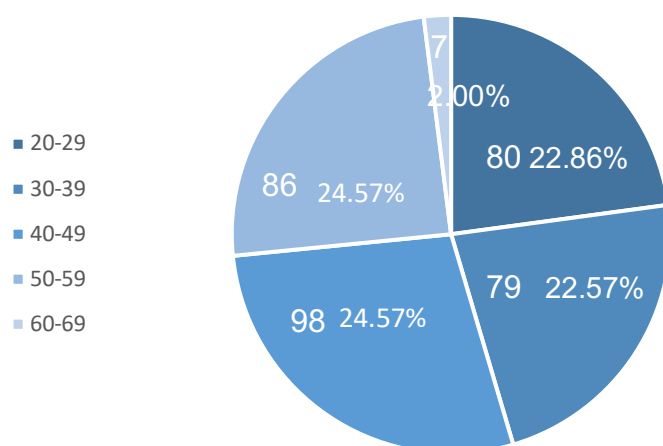
3. Genera una tabla agrupando valores (realiza 5 grupos de edad) para que puedas generar a partir de ella un histograma.

Distribución por edades

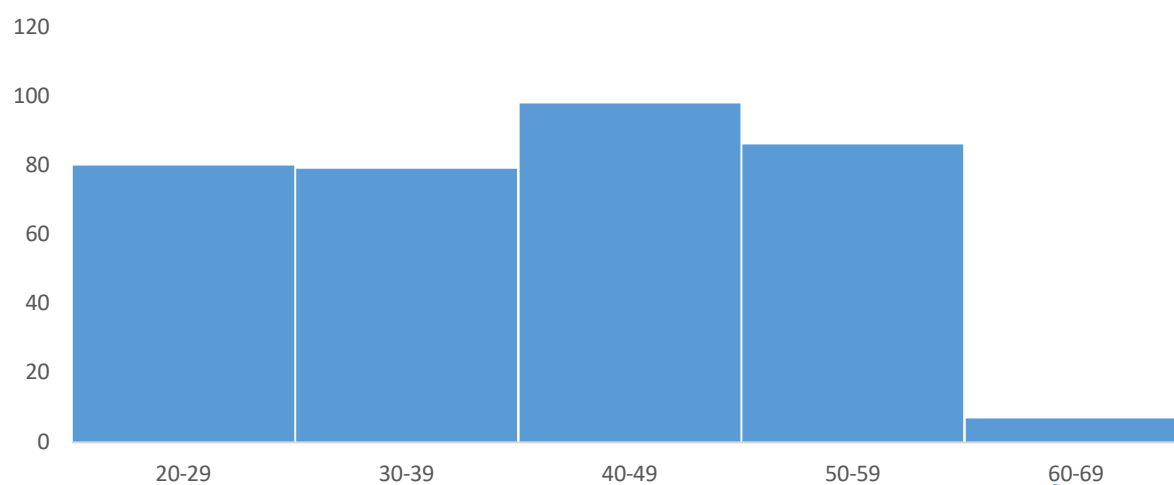
Validamos que entre la mayor parte de los encuestados se encuentra entre los 40 y 49 años de edad.

| Edad | Frecuencia absoluta | Frecuencia Relativa |
|--------------|---------------------|---------------------|
| 20-29 | 80 | 22.86% |
| 30-39 | 79 | 22.57% |
| 40-49 | 98 | 28.00% |
| 50-59 | 86 | 24.57% |
| 60-69 | 7 | 2.00% |
| Total | 350 | 100% |

Grupo de Edades



Grupos por edad



4. Con base en lo realizado anteriormente contesta las siguientes preguntas:

a. ¿Cuál es el tamaño de la muestra?

El tamaño de la muestra es de 350 personas encuestadas.

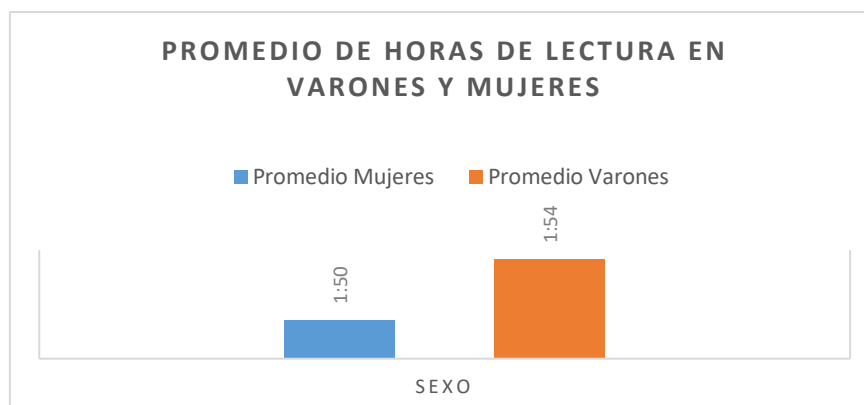
b. ¿Se aplicó la encuesta a más hombres o mujeres?



En base a los datos tomados de la encuesta y con ayuda de la aplicación Excel es posible determinar de manera simple por medio de las funciones Graficar y Contar la Cantidad y porcentaje de Mujeres y Varones participantes. De los 350 encuestados el 52% fueron hombres con un total de 183 varones participantes mientras que el 48% restante fueron mujeres con un total de 167 participantes femeninas.

c. ¿Se aplicó la encuesta a más hombres o mujeres?

Con la función promedio de igual manera es posible generar el promedio de las personas encuestadas. Con la finalidad de obtener el promedio de horas dedicadas a la lectura por parte de los participantes en sus momentos de ocio siendo 1:54 en Varones y 1:50 en Mujeres.



d. ¿La edad es relevante para dedicar más tiempo a la lectura?

Para poder responder la pregunta primero clasifiqué a los encuestados en 4 grupos de edad:

Tabla de frecuencia para cada grupo de edad

| Gpo Edad | Total | Porcentaje | Acumulado | % Acumulado |
|----------|-------|------------|-----------|-------------|
| 20 - 30 | 85 | 24.29 | 85 | 24.29 |
| 30 - 40 | 85 | 24.29 | 170 | 48.57 |
| 40 - 50 | 95 | 27.14 | 265 | 75.71 |
| 50 - 60 | 85 | 24.29 | 350 | 100.00 |
| N= | 350 | | | |

Decidí separar el total de registros en grupos de edad de diez años para poder interpretar más fácilmente los resultados (casi todos los grupos de edad tienen la misma cantidad de registros)

A continuación, procedí a obtener algunos indicadores estadísticos básicos para cada grupo de edad (media, mediana, desviación estándar, máximo, mínimo, rango y rango intercuartil)

Resumen de tiempo de lectura por grupo de edad en horas

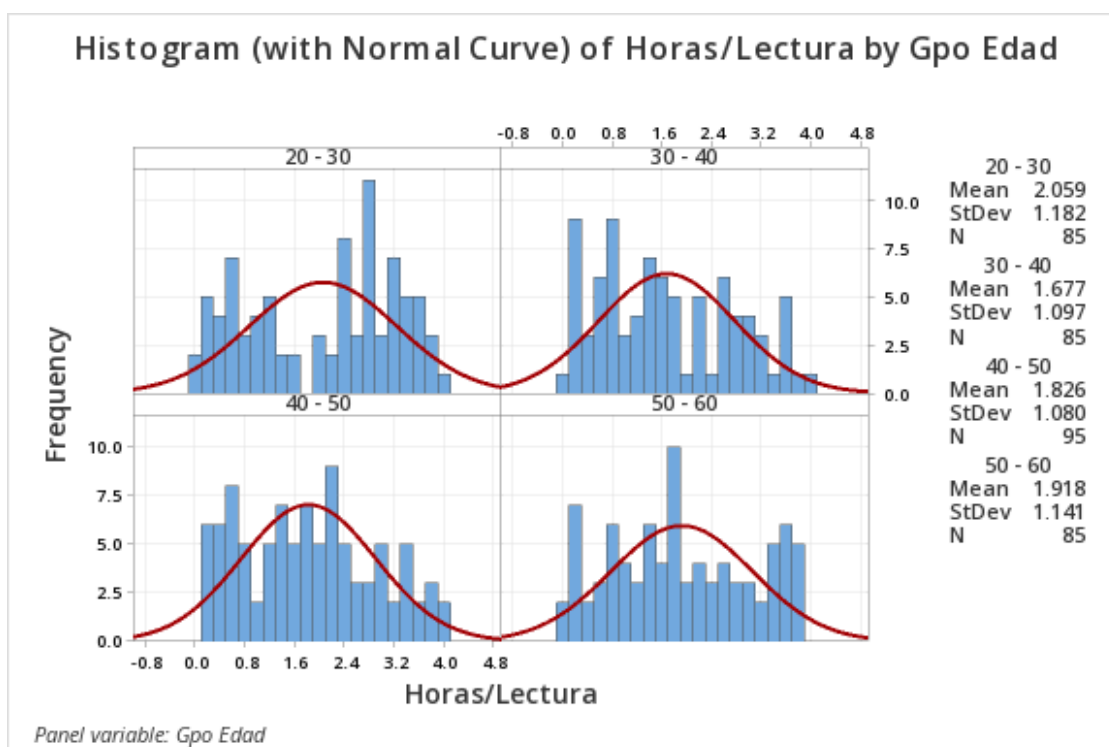
| Variable | Gpo Edad | Media | Desv. Std. | Varianza | Suma | Mínimo | Mediana | Máximo | Rango | RIQ |
|---------------|----------|-------|------------|----------|---------|--------|---------|--------|-------|-------|
| Horas/Lectura | 20 - 30 | 2.059 | 1.182 | 1.397 | 175.033 | 0.083 | 2.383 | 3.900 | 3.817 | 2.192 |
| | 30 - 40 | 1.677 | 1.097 | 1.204 | 142.583 | 0.050 | 1.500 | 3.967 | 3.917 | 1.883 |
| | 40 - 50 | 1.826 | 1.080 | 1.166 | 173.500 | 0.100 | 1.817 | 3.967 | 3.867 | 1.800 |
| | 50 - 60 | 1.918 | 1.141 | 1.302 | 163.067 | 0.050 | 1.833 | 3.833 | 3.783 | 1.958 |

Revisando la tabla anterior podemos asegurar que el grupo de edad de 20 a 30 años lee más horas en promedio (2.059 horas o 02:03). El valor promedio por sí solo nos ofrece solo una perspectiva, es necesario analizar como están distribuidos los datos.

Para el grupo de edad de 20 a 30 años la desviación estándar y la varianza son de 1.182 y 1.397 respectivamente, con lo anterior se observa que los datos **están más alejados de la media** que el resto de los grupos de edad.

Sin embargo, al analizar el rango (3.817) y el rango intercuartil (2.192) podemos concluir que realmente que no existe una distorsión de la media debido a valores extremos (los llamados *outliers* en inglés) y que de hecho si existen muchos valores altos, como se puede observar en las siguientes graficas:

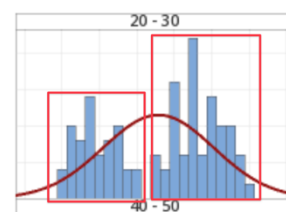




Gráfica 4.4.1

En la gráfica 4.4.1 se observa que en el grupo de edad de 20 a 30 años hay muchos registros con duraciones de más de 2 horas, incluso se observa que la curva de distribución normal se inclina hacia la derecha (la mediana es de 2.383 horas o 2:22).

La razón de que la desviación estándar y la varianza sean más altos en ese grupo de edad es la existencia de 2 subgrupos más o menos definidos, en un grupo que lee mucho y otro grupo que no lee mucho y casi no hay registros al centro de la distribución.



También observamos que la curva de distribución para el grupo de edad de 30 a 40 años se inclina a hacia la izquierda (menos horas de lectura) lo anterior es consistente con una mediana menor (1.5 horas o 1:30) y además que es el grupo con la media más baja (1.677 o 1:40) de horas de lectura. Para finalizar el grupo de 20 a 30 es el grupo que más horas de lectura en total tiene (175) y el grupo de 30 a 40 el que menos (142).

Con lo anterior en mi opinión **la edad es relevante para dedicar más tiempo de lectura.**



e. ¿Quiénes ven más horas de televisión, los hombres o las mujeres sin empleo?

El primer paso para responder esa pregunta es obtener el total de desempleados y la cantidad de hombres y mujeres y su proporción. Creé un subconjunto de datos que solo incluía los desempleados y la proporción de hombres y mujeres es la siguiente:

Desempleados

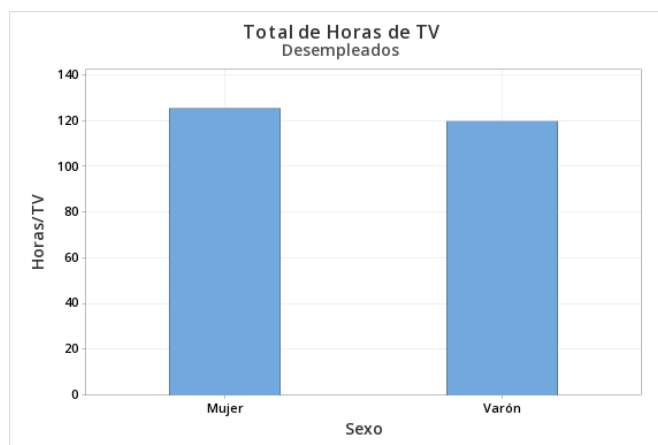
| Sexo | Total | Porcentaje | Acumulado | % Acumulado |
|-------|-------|------------|-----------|-------------|
| Mujer | 55 | 47.41 | 55 | 47.41 |
| Varón | 61 | 52.59 | 116 | 100.00 |
| N= | 116 | | | |

Para ése mismo subconjunto de desempleados obtuve métricas de tendencia central y de variabilidad, así como la suma de horas para cada género:

Resumen de tiempo de horas de TV por Sexo

| Variable | Sexo | Media | Desv. Estándar | Varianza | Total | Mínimo | Máximo | Rango | RIC |
|----------|-------|-------|----------------|----------|---------|--------|--------|-------|-------|
| Horas/TV | Mujer | 2.284 | 1.263 | 1.596 | 125.633 | 0.150 | 3.983 | 3.833 | 2.600 |
| | Varón | 1.967 | 1.072 | 1.150 | 119.967 | 0.017 | 3.967 | 3.950 | 1.567 |

Como podemos observar las mujeres desempleadas en promedio pasan más tiempo mirando la televisión que los hombres desempleados.



Gráfica 4.5.1

En la grafica 4.5.1 podemos observar que el total de horas que vieron las mujeres desempleadas fue mayor (125.63) al de los hombres (119.96) a pesar de haber más varones que mujeres (55 vs 61).

Teniendo en cuenta lo anterior podemos asegurar que **las mujeres desempleadas vieron más horas de TV que los hombres desempleados.**



4. Conclusiones

En esta primera etapa del proyecto integrador comprendimos los diversos datos que se pueden obtener mediante el uso de la estadística descriptiva con el fin de así poder comprenderlos y desarrollar un mejor entendimiento de sus diferentes características ofreciéndonos un mejor manejo de la información por medio de ayudas visuales como las gráficas y concentrando gran cantidad de información de manera sintetizada o resumida la cual puede llevar a un uso más eficiente del conocimiento adquirido independientemente del ámbito en que se desempeñe.

También aplicamos distintas técnicas y programas como Excel comprendiendo la correcta representación de datos cuantificables en diferentes puntos como recursos humanos y conceptos más abstractos como el tiempo, la edad y el género. Dándonos como resultado que la estadística descriptiva es aplicable en la mayoría de las áreas en las que se use la recopilación de datos. Incluso las grandes ciudades como Seúl lo utilizan para detectar áreas de oportunidad y usar mejor sus recursos como el transporte público y la logística, amplificar su alcance reducir costos y la contaminación al mismo tiempo.

En la actualidad es un hecho lo importante que puede llegar a ser la información. Incluso en áreas como la política cada sexenio en México se ven las encuestas sobre los candidatos a la presidencia en toda la República Mexicana. De igual forma en la mayor parte del mundo. En los deportes, en el sector financiero, en el sector industrial, en el sector salud, etc. Actualmente con el uso de las nuevas tecnologías y las redes sociales, estamos en contacto con encuestas, datos, estadísticas que recopilan nuestra información con la finalidad de brindar un mejor servicio, o simplemente dar un uso más eficiente a sus recursos. E incluso en nuestra vida cotidiana es posible implementar la estadística descriptiva para llevar un mejor control de cualquier cosa.



Bibliografía.

Griffiths, D. (2009). *Head first statistics*. O'Reilly Media. (45-118)

Kokoska, S. (2015). *Introductory statistics: A problem-solving approach*. W.H. Freeman & Company. (27-45)

Tabla de Frecuencias en Excel Para Datos Agrupados - YouTube. (s. f.). Recuperado el 29 de Enero de 2022 de <https://www.youtube.com/watch?v=YIAWyP5rsSw>

¿Cómo se organiza la información en una tabla de frecuencia? (s. f.). A.L.EPH. Recuperado 28 de enero de 2022, de <https://aleph.org.mx/como-se-organiza-la-informacion-en-una-tabla-de-frecuencia>

Evidencia de reunión de equipo

