

**Análisis descriptivo de encuesta en el uso del tiempo  
(ENUT) de la población civil, realizada en el periodo  
2016 – 2017 Colombia**

**Anthony Servitá**

Análisis descriptivo sobre encuesta  
Septiembre 25 - 2020.

DIMPTE, ANDA.

Analisis descriptivo

### **Abstract**

La valoración del uso del tiempo sobre la población en Colombia es una rama de la estadística que se ha venido convirtiendo cada día más necesario para la evaluación sobre como distribuye el tiempo las personas, sobre qué actividades; así como también, las mediciones de excesivas horas de trabajo sobre un sector en específico, o también; las pocas horas de trabajo tomadas por algunos sectores. El cual se considera, que tiene repercusiones debido a que las mismas afectan de manera directa o indirectamente al sector económico en Colombia. Así mismo, la población colombiana presenta resultados suficientes para el años 2016 – 2017 para el cual se toma de evaluación las encuestas sobre el uso del tiempo. Los colombianos mantienen un uso constante, usado en trabajo y en desarrollar otras actividades, no obstante más adelante descubrirán que las horas de trabajo tomadas por los colombianos suelen ser las necesarias para mantener su propia economía, ayudar al crecimiento del país, y a mantener un estado de salud plena.

Analisis descriptivo

## Tabla de Contenidos

Capítulo 1 Introducción e información general .....	4
Bases de datos .....	4
Encuesta nacional de uso del tiempo 2016 – 2017 .....	4
Capitulo I. Uso del tiempo. ....	5
Capitulo H. Uso del tiempo en ingresos y trabajo. ....	5
Capítulo 2 Figuras y tablas .....	6
Análisis de datos .....	6
Análisis descriptivo de la variable horas dedicadas al sueños. ....	6
Análisis descriptivo de la variable uso del tiempo en ingresos y trabajo. ....	12
Capítulo 4 Resultados y discussion.....	25
Lista de Referencias .....	27
Apéndice .....	28

## Capítulo 1

### Introducción e información general

#### **Base de datos en estudio**

El estudio de encuestas realizado por el DIMPE, los datos que se analizaron en este estudio han sido seleccionados desde la base de datos de DANE (archivo nacional de datos) el cual presenta, un abanico extenso con respecto a la base de datos de encuestas realizadas a la población colombiana. En este ámbito, se seleccionó la base de datos referente al uso del tiempo en el periodo 2016 – 2017. El mismo, se encuentra distribuido en 5 capítulos con distintas variables cualitativas y cuantitativas entre ellas.

#### **Base de datos encuesta nacional de uso del tiempo 2016 – 2017 (ENUT)**

El estudio abarca una encuesta nacional en Colombia, sobre el uso del tiempo de las personas en dicha nación. La misma tiene el propósito de entender como las personas usan su tiempo libre, en trabajo o quehaceres del hogar, este último principalmente abarcado con más frecuencia por el género femenino. Un aspecto importante que es necesario resaltar en este artículo, es que para el año 1975 se realiza la primera conferencia mundial sobre la condición jurídica y social de la mujer en México, en lo que se concluye sobre su mismo marco la importancia de la participación de la mujer sobre el desarrollo en la nación; por dicha razón, la encuesta toma un equilibrio orientado a como la actividades realizadas por el hombre o la mujer pueden influenciar sobre el mecanismo de desarrollo de la nación valorando las actividades no remuneradas.

Análisis descriptivo

Para el año 1995 la declaración y plataforma de acción de Beijing se recomendó mejorar la obtención de datos sobre toda la contribución de la mujer y el hombre sobre la economía, incluyendo su participación en el sector no estructurado.

**Capítulo I. Uso del tiempo.** El archivo seleccionado perteneciente a la base de datos, UNET, contiene 396 variables de estudio las mismas contienen datos dirigidos a las actividades de trabajos no remunerados, así como las actividades personales realizadas por los miembros del hogar de 10 años o más .

**Capítulo H. Uso del tiempo e ingresos en trabajo.** sobre este capítulo se han seleccionado 7 variables, las cuales se orientan al lugar donde se realiza el trabajo, horas por semana que consumen los individuos en el trabajo, horas gastadas en la ruta hacia el trabajo, el tipo de trabajo, subsidio de transporte, dinero recibido por el subsidio, pago del mes anterior en el periodo que fue realizada la encuesta.

El objetivo de este análisis descriptivo, es de obtener una base cuantitativa sobre el uso del tiempo en actividades económicas y horas de descanso, realizado por los encuestados en el sector de la poblacional civil de Colombia a partir de la edad de 10 años o más. Para el periodo 2016 – 2017.

## Capítulo 2

### Figuras y tablas

Las figuras seleccionadas para el análisis descriptivo sobre esta base de datos, radica en su utilidad de representar los valores extremos, y las frecuencias en que los valores se concentran. Estos gráficos son, cajas y bigotes (boxplot) e histograma respectivamente.

#### Análisis de datos.

#### Análisis descriptivo de la variable horas dedicada al sueño, del capítulo I.

1.- Conocimiento del problema Los datos generados a continuación, provienen de una base de datos de encuestas de Colombia. los micro-datos y metadatos pueden encontrarse en el siguiente link: <http://microdatos.dane.gov.co/index.php/catalog/552/study-description>

```
datasets <- read.csv2 ("/BD_for analisis/CAPITULO I/CAPITULO I.csv")
head(datasets)
```

##	i..DIRECTORIO	ORDEN	SECUENCIA_P	P1144S1	P1144S1A1	P1144S1A2	P1144S2
## 1	1000000	1	1	1	7	0	1
## 2	1000000	2	1	1	11	0	1
## 3	1000000	3	1	NA	NA	NA	NA
## 4	1000000	4	1	NA	NA	NA	NA
## 5	1000002	1	1	1	8	59	1
## 6	1000004	1	1	1	11	59	1

Descripción de las variables: con respecto a la variable de escogencia, P1144S1 la cual contiene información sobre si los individuos usan tiempo para dormir en siestas y demás. esta variable se corresponde con otras 2 las cuales contienen el total de horas usadas para dormir (P1144S1A1). Esta encuesta se llevó a cabo para la población civil, no institucional con edades iguales o mayores a 10 años.

#### Análisis descriptivo

## 2.- Conocimiento de los datos

```
df <- datasets %>%
  select(P1144S1, P1140S1A1) %>%
  summarize(glimpse(datasets[, 4:5]))

## Rows: 146,190
## Columns: 2
## $ P1144S1    <int> 1, 1, NA, NA, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, NA, 1...
## $ P1144S1A1 <int> 7, 11, NA, NA, 8, 11, 7, 7, 8, 12, 10, 7, 7, 9, 10, ...
```



Podemos observar que existen un tipo de variables correcto, en este caso las tres variables seleccionadas son del tipo entero también encontramos una cantidad considerable de valores NA lo cual, puede provocar distorsión en nuestros análisis.

```
skim(df)
```

*Data summary*

Name	df
Number of rows	146190
Number of columns	2
<hr/>	
Column type frequency:	
numeric	2
<hr/>	
Group variables	None

**Tabla 1.** Datos descriptivos**Variable type: numeric**

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
P1144S1	22168	0.85	1.00	0.00	1	1	1	1	1	
P1144S1A1	22168	0.85	8.21	1.69	0	7	8	9	20	

Se puede apreciar que existe un tipo de variables correcto, en este caso las tres variables seleccionadas son del tipo entero

Análisis descriptivo

también encontramos una cantidad considerable de valores NA lo cual, puede provocar distorsión en nuestros análisis.

Por otro lado, tenemos un promedio de 8.21 horas consumidas en dormir por parte de los individuos. un 25% de estos individuos consumen 7 horas al día de sueño, un 50% consume 8 horas, hasta un 75% consumen 9 horas en sueño.

## 2.- limpieza y corrección de datos

```
## Eliminar los datos perdidos
df <- na.omit(df)

## revisamos
skim(df)
```

### Data summary

Name	df
Number of rows	124022
Number of columns	2

### Column type frequency:

Numeric	2
---------	---

Group variables	None
-----------------	------

### Tabla 1. Datos descriptivos

#### Variable type: numeric

skim_variable	n_missing	Complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
P1144S1	0	1	1.00	0.00	1	1	1	1	1	— — █
P1144S1A1	0	1	8.21	1.69	0	7	8	9	20	— █ █

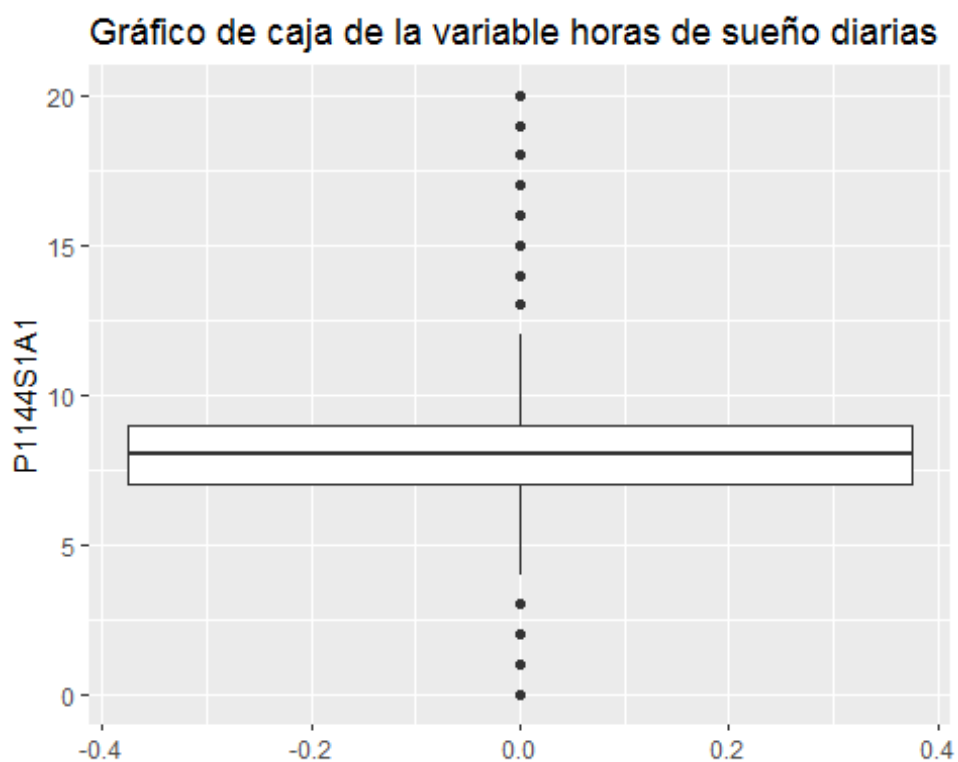
Los datos faltantes fueron eliminados, y como puede apreciarse no hay ninguna modificación en cuanto a sus resultados anteriores, con respecto al medio, desviación y sus percentiles.

## Análisis descriptivo



**Figura 1.** Representación en grafico de cajas

```
ggplot(df, aes(y = P1144S1A1)) +  
  geom_boxplot() +  
  ggtitle("Gráfico de caja de la variable horas de sueño diarias")
```



En el gráfico, el promedio se encuentra cercano a 9 horas de sueños, consumidos por las personas según la muestra del censo. también pueden observarse valores outliers, que indican valores extremos; con esto nos referimos a: mayor cantidad de sueños consumidos por las personas, esto a partir de alrededor de 12 horas en adelante.

**Tabla 2.** Estadística descriptiva: tablas de frecuencias

```

tab_freq <- fdt(df[, 2], breaks = "Sturges")
tab_freq
##   Class limits      f   rf rf(%)      cf  cf(%)
##   [0,1.122)      110 0.00  0.09      110  0.09
##   [1.122,2.244)    95 0.00  0.08      205  0.17
##   [2.244,3.367)   264 0.00  0.21      469  0.38
##   [3.367,4.489)  1079 0.01  0.87     1548  1.25
##   [4.489,5.611)  3578 0.03  2.88     5126  4.13
##   [5.611,6.733) 11502 0.09  9.27    16628 13.41
##   [6.733,7.856) 21415 0.17 17.27    38043 30.67
##   [7.856,8.978) 39202 0.32 31.61    77245 62.28
##   [8.978,10.1)  37551 0.30 30.28   114796 92.56
##   [10.1,11.22)  4312 0.03  3.48   119108 96.04
##   [11.22,12.34) 3612 0.03  2.91   122720 98.95
##   [12.34,13.47)  712 0.01  0.57   123432 99.52
##   [13.47,14.59)  368 0.00  0.30   123800 99.82
##   [14.59,15.71)  143 0.00  0.12   123943 99.94
##   [15.71,16.83)   45 0.00  0.04   123988 99.97
##   [16.83,17.96)   16 0.00  0.01   124004 99.99
##   [17.96,19.08)   13 0.00  0.01   124017 100.00
##   [19.08,20.2)    5 0.00  0.00   124022 100.00

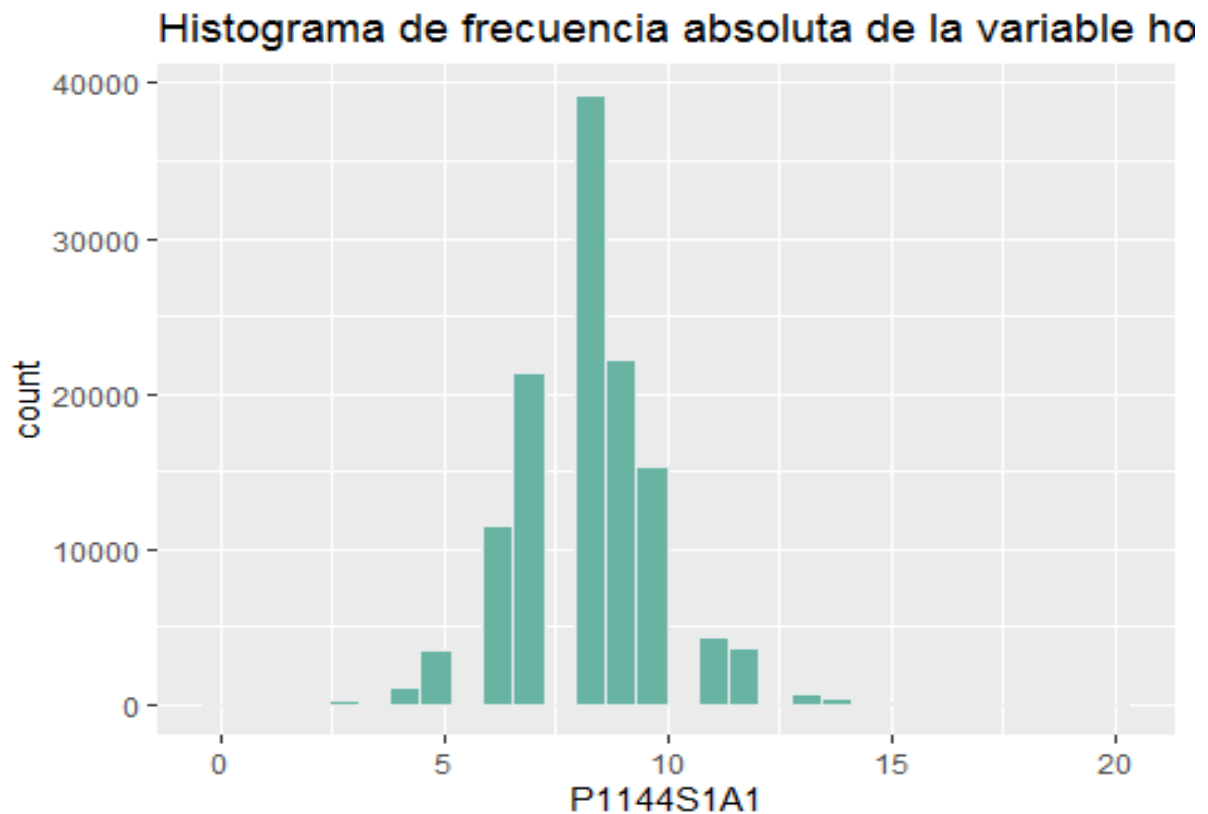
```

la tabla de frecuencia muestra los siguientes resultados: alrededor del 88% de las personas duermen entre 5 a 10 horas, esto se observa gracias a que la mayor concentración de los datos se encuentra entre la clase 6 y 9 de la variable P1144S1A1 la cual mide el tiempo de sueño en horas. en pocas palabras esto se traduce en: 109670 individuos que duermen esta cantidad de horas.

Por otro lado, solo un 4% de las personas duermen entre 3.3 a 5.6 horas diarias, porcentaje que se encuentra entre la clase 4 y 5. así mismo, solo un 7% de la población se duerme alrededor de 10 13.47 horas diarias.

**Figura 2.** Representación gráfica, histograma de frecuencia

```
ggplot(df, aes(x = P1144S1A1)) +  
  geom_histogram(bins = 30, fill = "#69b3a2", color = "#e9ecef") +  
  ggtitle("Histograma de frecuencia absoluta de la variable horas de sueño")
```



La representación gráfica de la variable horas de sueño (P1144S1A1) muestra similitud con la tabla de frecuencias calculada, pues muestra una gran cantidad de datos alojados entre 7 y 9 horas, como ya se había mencionado anteriormente.

## Análisis descriptivo de la variable trabajo uso del tiempo e ingresos.

1.- Entendimiento de la nueva base de datos. para este caso, importaremos el capítulo H. Fuerza de trabajo - uso del tiempo e ingresos

```
datasets_w <- read.csv2("/BD_for análisis/capitulo h uso del tiempo/capitulo h
1.csv")
head(datasets_w)
```

##	i..DIRECTORIO	ORDEN	SECUENCIA_P	P6370S1	P6390S1	P6880	P1151	P1150
## 1	1000000	1	1	NA	NA	NA	NA	NA
## 2	1000000	2	1	4	6211	7	48	2
## 3	1000000	3	1	NA	NA	NA	NA	NA
## 4	1000000	4	1	NA	NA	NA	NA	NA
## 5	1000002	1	1	NA	NA	NA	NA	NA
## 6	1000004	1	1	84	2424	7	48	2

5.- selección de variables de trabajo para relacionar en el uso del tiempo en el trabajo (capítulo H): se han seleccionado por medio de las instrucciones dadas, 7 variables las cuales contienen.

- P6880 (tipo de medida discreta (describe una variable cualitativa)): ¿Donde realiza su trabajo?
  1. En esta vivienda
  2. En otras viviendas
  3. En kiosco - caseta
  4. En un vehículo
  5. De puerta en puerta
  6. Sitio al descubierto en la calle (ambulante y estacionario)
  7. Local fijo, oficina, fábrica, etc.
  8. En el campo o área rural, mar o río
  9. En una obra en construcción
  10. En una mina o cantera
  11. Otro, ¿cuál?
- P1151 (tipo cuantitativa aleatoria continua): Cuantas horas a la semana trabaja.

Análisis descriptivo

- P1099S1 (tipo cuantitativa discreta): Horas gastadas para ir al sitio de trabajo.
- P427 (tipo discreto (describe una variable cualitativa)): Clase de trabajo
  1. Obrero o empleado de empresa particular
  2. Obrero o empleado del gobierno
  3. Empleado doméstico
  4. Jornalero o Peón
  5. Trabajador por cuenta propia (Pase a 30)
  6. Patrón o empleador (Pase a 30)
  7. Trabajador de su propia finca o de finca en arriendo o aparcería. (Pase a 30)
  8. Trabajador familiar sin remuneración (Pase a 34)
  9. Trabajador sin remuneración en empresas o negocios de otros hogares. (Pase a 34)
- P6500 (tipo cuantitativa discreta): Antes del descuento, ¿cuánto gano... el mes pasado en este empleo?
- P6545B (tipo de variable discreta numérica (describe una variable cualitativa): subsidio de transporte
- P6545BS1 (tipo cuantitativa aleatoria continua): Cuanto recibe de subsidio por el transporte en su trabajo.

## 2.- Comprensión de los datos

```
df_w <- datasets_w %>%
  select(P6880, P1151, P1099S1, P427, P6500, P6545B, P6545BS1)
glimpse(df_w)

## Rows: 146,190
## Columns: 7
## $ P6880    <int> NA, 7, NA, NA, NA, 7, NA, NA, 7, NA, NA, NA, 1, NA, N...
## $ P1151    <int> NA, 48, NA, NA, NA, 48, NA, NA, 48, NA, NA, NA, 16, N...
## $ P1099S1  <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 0, NA...
## $ P427     <int> NA, 1, NA, NA, NA, 1, NA, NA, 1, NA, NA, NA, 8, NA, N...
## $ P6500    <int> NA, 5500000, NA, NA, NA, 1200000, NA, NA, 1400000, NA...
## $ P6545B   <int> NA, 2, NA, NA, NA, 2, NA, NA, 1, NA, NA, NA, NA, NA, ...
## $ P6545BS1 <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, 70000, NA, NA, NA, NA...
```

## Analisis descriptivo

en observación con estas variables, encontramos que todas corresponden al tipo de variable que se mencionan en los metadatos. por otro lado, hay una cantidad considerable de valores perdidos, lo que dificulta los análisis sobre estas variables.

```
skim(df_w)
```

#### Data summary

Name	df_w
Number of rows	146190
Number of columns	7








#### Column type frequency:

numeric	7
---------	---

Group variables	None
-----------------	------

#### Tabla 3. Datos descriptivos

##### Variable type: numeric

skim_variable	n_missing	mean	sd	p0	p25	p50	p75	p100	hist
P6880	81447	5.61	2.55	1	4	7	7	1.1e+01	
P1151	81447	45.86	17.82	1	40	48	56	1.2e+02	
P1099S1	97296	0.25	0.64	0	0	0	0	1.2e+01	
P427	81447	3.37	2.11	1	1	5	5	9.0e+00	
P6500	114685	1008920.17	1364955.17	99	600000	750000	1100000	7.0e+07	
P6545B	114685	1.59	0.49	1	1	2	2	2.0e+00	
P6545BS1	133154	74706.39	83596.66	99	70500	77700	83000	2.1e+06	

#### Análisis descriptivo

### 3.- Limpieza de datos:

*## Depuración de Los datos*

```
df_w <- na.omit(df_w)
```

*## Transformación de Las variables P6500 Y P6545BS1, La cuales corresponden a los pagos del mes pasado correspondiente a este periodo,*

*## y el dinero recibido por subsidio de transporte respectivamente.*

*## La variable P650 en millones de pesos y P645BS1 en miles de pesos*

```
df_w <- df_w %>%
```

```
  mutate(P6500 = P6500/1000000) %>%
```

```
  mutate(P6545BS1 = P6545BS1/1000)
```

```
head(df_w)
```

```
##   P6880 P1151 P1099S1 P427 P6500 P6545B P6545BS1
## 1     5    44         0    1 1.200         1     72.0
## 2     4    15         0    1 1.200         1     72.0
## 3     7    48         1    2 0.800         1     74.0
## 4     7    48         0    1 0.815         1     77.0
## 5     6    80         0    2 1.900         1     68.0
## 6     7    48         0    1 0.850         1     77.1
```

```
skim(df_w)
```

#### Data summary

Name	df_w
Number of rows	10272
Number of columns	7

---

#### Column type frequency:

numeric	7
---------	---

---

Group variables	None
-----------------	------

### Analisis descriptivo

**Tabla 3.** Datos descriptivos**Variable type: numeric**

skim_variable	n_missing	complete rate	mean	sd	p0	p25	p50	p75	p100	hist
P6880	0	1	6.65	1.36	1.0	7.0	7.0	7	11	
P1151	0	1	51.16	10.98	3.0	48.0	48.0	56	120	
P1099S1	0	1	0.38	0.68	0.0	0.0	0.0	1	12	
P427	0	1	1.10	0.36	1.0	1.0	1.0	1	4	
P6500	0	1	0.88	0.66	0.0	0.7	0.8	1	20	
P6545B	0	1	1.00	0.00	1.0	1.0	1.0	1	1	
P6545BS1	0	1	73.80	80.90	0.1	70.0	77.7	83	2100	

Los valores de la media correspondientes a la variable P6500 vemos un valor de 876608 pesos colombianos que ganó una persona el mes pasado de haberse realizado esta encuesta.

también es notable que existe un promedio de personas que trabajan e oficinas o local comerciales.

en promedio las personas en Colombia gastan 51.2 horas de trabajo a la semana.

estas personas gastan en promedio 0.379 horas en ir al trabajo.

50% de la población en estudio son trabajadores son empleado de empresas particulares

en promedio todos reciben subsidio de transporte y estos mismos reciben un total de 73805 pesos colombianos por mes como subsidio para el transporte.

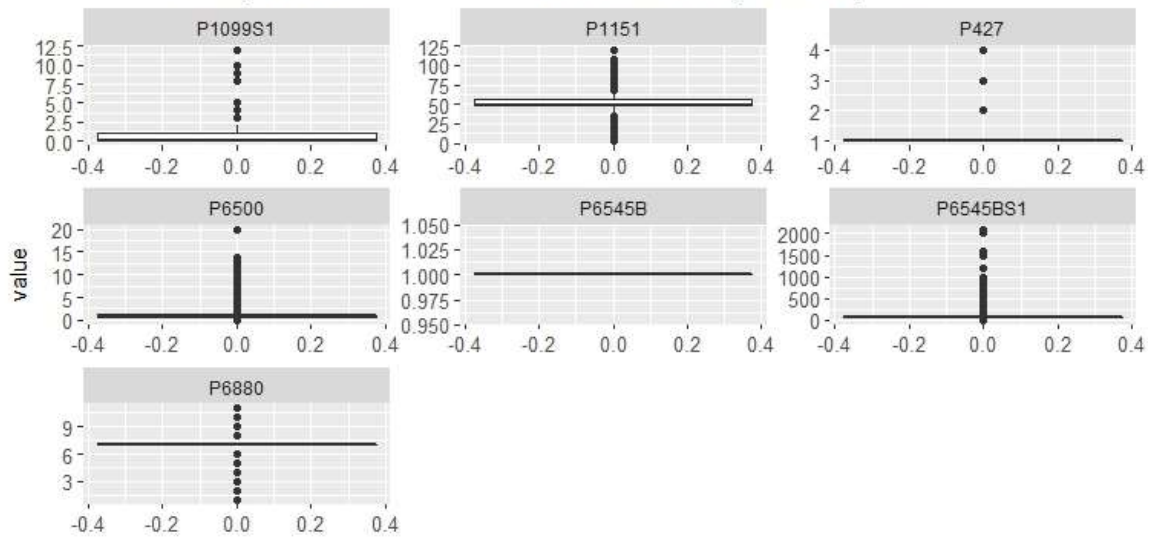
**Figura 3.** Representación grafica por grafico de cajas

```
df_w %>%
  select(P6880, P1151, P1099S1, P427, P6500, P6545B, P6545BS1) %>%
  gather() %>%
  ggplot(aes(y = value)) +
  geom_boxplot() +
  facet_wrap(~key, scales = "free") +
  theme(axis.text = element_text(size = 10)) +
  ggtitle("Gráfico de caja de la variables 'uso de horas de trabajos' del
capítulo H")
```

Análisis descriptivo



Gráfico de caja de la variables 'uso de horas de trabajos' del capítulo H



como se muestra en los gráficos de las 7 variables, poseen una gran cantidad de datos muy dispersos.

**Tabla 4.** Tabla de frecuencia

```

tab_freq_w <- fdt(df_w, breaks = "Sturges")

## analicemos la tabla de frecuencia de la variable P6880 (Donde realiza su tr
abajo)
print(tab_freq_w$P6880)

## $table
##      Class limits      f      rf      rf(%)      cf      cf(%)
## 1  [0.99,1.6647)    52 0.005062305 0.5062305    52 0.5062305
## 2  [1.6647,2.3393) 382 0.037188474 3.7188474   434 4.2250779
## 3  [2.3393,3.014)   14 0.001362928 0.1362928   448 4.3613707
## 4  [3.014,3.6887)    0 0.000000000 0.0000000   448 4.3613707
## 5  [3.6887,4.3633) 550 0.053543614 5.3543614   998 9.7157321
## 6  [4.3633,5.038)  201 0.019567757 1.9567757  1199 11.6725078
## 7  [5.038,5.7127)    0 0.000000000 0.0000000   1199 11.6725078
## 8  [5.7127,6.3873) 258 0.025116822 2.5116822  1457 14.1841900
## 9  [6.3873,7.062) 8297 0.807729751 80.7729751  9754 94.9571651
## 10 [7.062,7.7367)    0 0.000000000 0.0000000   9754 94.9571651
## 11 [7.7367,8.4113)  193 0.018788941 1.8788941   9947 96.8360592
## 12 [8.4113,9.086)   218 0.021222741 2.1222741  10165 98.9583333
## 13 [9.086,9.7607)    0 0.000000000 0.0000000   10165 98.9583333
## 14 [9.7607,10.435)  20 0.001947040 0.1947040  10185 99.1530374
## 15 [10.435,11.11)   87 0.008469626 0.8469626  10272 100.0000000
##
## $breaks
##      start      end      h      right
## 0.9900000 11.110000 0.6746667 0.0000000

```

puede observarse en la clase 9 existe un 80% de los individuos realizan sus trabajos desde zonas descubiertas (ambulantes) y desde la oficina. y solo un 3.7% de estos encuestados trabajan desde su vivienda u otra vivienda. un 5.3% trabajan desde el vehículo (transporte).

**Tabla 5.**

*## analicemos la tabla de frecuencia de la variable P1151 (cuantas horas a la semana trabaja)*

```
print(tab_freq_w$P1151)
```

```
## $table
##      Class limits      f      rf      rf(%)      cf      cf(%)
## 1    [2.97,10.852)    84 0.0081775701 0.81775701    84 0.817757
## 2    [10.852,18.734)  34 0.0033099688 0.33099688   118 1.148754
## 3    [18.734,26.616)  73 0.0071066978 0.71066978   191 1.859424
## 4    [26.616,34.498) 116 0.0112928349 1.12928349   307 2.988707
## 5    [34.498,42.38)  845 0.0822624611 8.22624611  1152 11.214953
## 6    [42.38,50.262) 5824 0.5669781931 56.69781931  6976 67.912773
## 7    [50.262,58.144) 1329 0.1293808411 12.93808411  8305 80.850857
## 8    [58.144,66.026) 1043 0.1015381620 10.15381620  9348 91.004673
## 9    [66.026,73.908)  608 0.0591900312 5.91900312  9956 96.923676
## 10   [73.908,81.79)  108 0.0105140187 1.05140187 10064 97.975078
## 11   [81.79,89.672)  143 0.0139213396 1.39213396 10207 99.367212
## 12   [89.672,97.554)   44 0.0042834891 0.42834891 10251 99.795561
## 13   [97.554,105.44)   15 0.0014602804 0.14602804 10266 99.941589
## 14   [105.44,113.32)    2 0.0001947040 0.01947040 10268 99.961059
## 15   [113.32,121.2)    4 0.0003894081 0.03894081 10272 100.000000
##
## $breaks
##      start      end      h      right
## 2.970 121.200  7.882  0.000
```

mayormente el 79.7% de los encuestados trabajan por semana desde 42.38 - 66 horas un equivalente de 8196 personas. solo el 0.81% trabajan en un intervalo de 3 a 11 horas por semana, esto es un total de 84 encuestados.

**Tabla 6.**

```
## analicemos la tabla de frecuencia de la variable P427 (trabajo)
```

```
print(tab_freq_w$P427)
```

```
## $table
```

##	Class limits	f	rf	rf(%)	cf	cf(%)
## 1	[0.99,1.193)	9489	0.92377336	92.377336	9489	92.37734
## 2	[1.193,1.397)	0	0.00000000	0.000000	9489	92.37734
## 3	[1.397,1.6)	0	0.00000000	0.000000	9489	92.37734
## 4	[1.6,1.803)	0	0.00000000	0.000000	9489	92.37734
## 5	[1.803,2.007)	601	0.05850857	5.850857	10090	98.22819
## 6	[2.007,2.21)	0	0.00000000	0.000000	10090	98.22819
## 7	[2.21,2.413)	0	0.00000000	0.000000	10090	98.22819
## 8	[2.413,2.617)	0	0.00000000	0.000000	10090	98.22819
## 9	[2.617,2.82)	0	0.00000000	0.000000	10090	98.22819
## 10	[2.82,3.023)	162	0.01577103	1.577103	10252	99.80530
## 11	[3.023,3.227)	0	0.00000000	0.000000	10252	99.80530
## 12	[3.227,3.43)	0	0.00000000	0.000000	10252	99.80530
## 13	[3.43,3.633)	0	0.00000000	0.000000	10252	99.80530
## 14	[3.633,3.837)	0	0.00000000	0.000000	10252	99.80530
## 15	[3.837,4.04)	20	0.00194704	0.194704	10272	100.00000

```
##
```

```
## $breaks
```

##	start	end	h	right
##	0.9900000	4.0400000	0.2033333	0.0000000

El 92.3% de los encuestados son empleados de empresas particulares.

**Tabla 7.**

*## analicemos la tabla de frecuencia de la variable P6545BS1 (dinero por subsidio de transporte)*

```
print(tab_freq_w$P6545BS1)
```

```
## $table
##      Class limits      f      rf      rf(%)      cf      cf(%)
## 1  [0.09801,141.5) 10008 9.742991e-01 97.429906542 10008 97.42991
## 2  [141.5,282.9)   132 1.285047e-02  1.285046729 10140 98.71495
## 3  [282.9,424.3)    45 4.380841e-03  0.438084112 10185 99.15304
## 4  [424.3,565.7)     6 5.841121e-04  0.058411215 10191 99.21145
## 5  [565.7,707.1)    15 1.460280e-03  0.146028037 10206 99.35748
## 6  [707.1,848.5)    54 5.257009e-03  0.525700935 10260 99.88318
## 7  [848.5,989.9)     5 4.867601e-04  0.048676012 10265 99.93185
## 8  [989.9,1131)      2 1.947040e-04  0.019470405 10267 99.95132
## 9  [1131,1273)       1 9.735202e-05  0.009735202 10268 99.96106
## 10 [1273,1414)       0 0.000000e+00  0.000000000 10268 99.96106
## 11 [1414,1555)       1 9.735202e-05  0.009735202 10269 99.97079
## 12 [1555,1697)       1 9.735202e-05  0.009735202 10270 99.98053
## 13 [1697,1838)       0 0.000000e+00  0.000000000 10270 99.98053
## 14 [1838,1980)       0 0.000000e+00  0.000000000 10270 99.98053
## 15 [1980,2121)       2 1.947040e-04  0.019470405 10272 100.00000
##
## $breaks
##      start      end      h      right
## 0.09801 2121.00000 141.39347 0.00000
```

un 97.4% de los encuestados afirman que reciben un intervalo de 98.01 - 141491.5 pesos colombianos por subsidio de transporte, el resto de los encuestados aseguran que reciben subsidios monetarios mayores a este.

**Tabla 8.**

*## analicemos la tabla de frecuencia de la variable P6500 (dinero recibido el mes pasado en el periodo de esta encuesta 2016 - 2017)*

```
print(tab_freq_w$P6500)
```

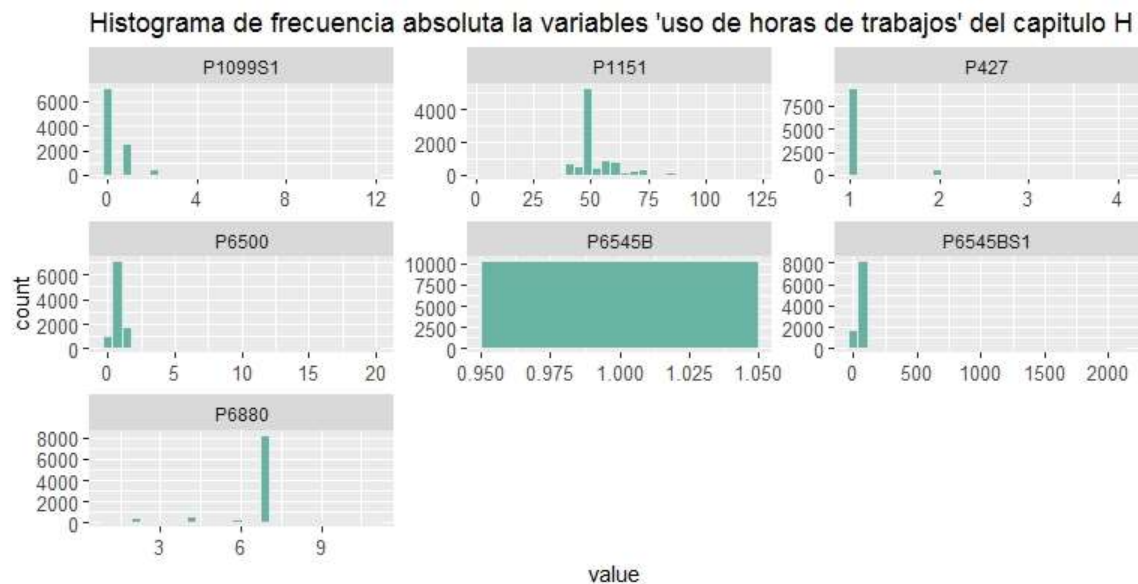
```
## $table
##      Class limits      f      rf      rf(%)      cf      cf(%)
## 1  [9.801e-05,1.347) 9323 9.076129e-01 90.761292835 9323 90.76129
## 2    [1.347,2.693)  821 7.992601e-02  7.992601246 10144 98.75389
## 3    [2.693,4.04)   90 8.761682e-03  0.876168224 10234 99.63006
## 4    [4.04,5.387)    6 5.841121e-04  0.058411215 10240 99.68847
## 5    [5.387,6.733)    7 6.814642e-04  0.068146417 10247 99.75662
## 6    [6.733,8.08)   10 9.735202e-04  0.097352025 10257 99.85397
## 7    [8.08,9.427)    6 5.841121e-04  0.058411215 10263 99.91238
## 8    [9.427,10.77)   3 2.920561e-04  0.029205607 10266 99.94159
## 9    [10.77,12.12)   3 2.920561e-04  0.029205607 10269 99.97079
## 10   [12.12,13.47)   1 9.735202e-05  0.009735202 10270 99.98053
## 11   [13.47,14.81)   1 9.735202e-05  0.009735202 10271 99.99026
## 12   [14.81,16.16)   0 0.000000e+00  0.000000000 10271 99.99026
## 13   [16.16,17.51)   0 0.000000e+00  0.000000000 10271 99.99026
## 14   [17.51,18.85)   0 0.000000e+00  0.000000000 10271 99.99026
## 15   [18.85,20.2)    1 9.735202e-05  0.009735202 10272 100.00000
##
## $breaks
##      start      end      h      right
## 0.00009801 20.2000707 1.34666060 0.00000000
```

De los 9323 encuestados ganaron el mes pasado un rango de 1346758 - 2693419 pesos colombianos, lo que equivale al 98% de la muestra total.

**Figura 4.** Representación gráfica.

```
df_w %>%
  select(P6880, P1151, P1099S1, P427, P6500, P6545B, P6545BS1) %>%
  gather() %>%
  ggplot(aes(value)) +
  geom_histogram(bins = 30, fill = "#69b3a2", color = "#e9ecef") +
  facet_wrap(~key, scales = "free") +
  theme(axis.text = element_text(size = 10)) +
  ggtitle("Histograma de frecuencia absoluta la variables 'uso de horas de tra
bajos' del capítulo H")
```

Análisis descriptivo



Exportación de los datos de los datos df\_w

```
names(df_w) <- c("Lugar_t", "H_semana_t", "H_gastadas_t", "C_trabajo", "Ganancia_pre_desc", "Sub_transporte", "Dinero_sub")
names(df_w)

## [1] "Lugar_t"          "H_semana_t"       "H_gastadas_t"
## [4] "C_trabajo"        "Ganancia_pre_desc" "Sub_transporte"
## [7] "Dinero_sub"

write.csv(df_w, file = "/BD_for análisis/capitulo h uso del tiempo/encuesta del uso del trabajo en horas.csv")
```

las variables han sido nombradas de manera simultaneas a como se importaron en la base de datos.

- Lugar\_t: Donde se realiza su trabajo
- H\_semana\_t: cuantas horas a la semana trabaja
- H\_gastadas\_t: horas gastadas para ir al sitio de trabajo
- C\_trabajo: tipo de trabajo
- Ganancia\_pre\_desc: cuando gano el mes pasado.
- Sub\_transporte: Subsidio de transporte
- Dinero\_sub: dinero recibido por el subsidio para el transporte.

Análisis descriptivo

## Relaciones entre las variables:

¿cuantas horas de trabajo cumplen a la semana las personas que trabajan desde la oficina?

```
df_w %>%
  filter(P6880 == 7) %>%
  select(P1151) %>%
  summarize(mean(P1151))

##      mean(P1151)
## 1      50.69953
```

¿qué tipo de trabajo tienen estas personas que trabajan en oficina?

```
df_w %>%
  filter(P6880 == 7) %>%
  select(P427) %>%
  summarize(mean(P427)) %>%
  round(digits = 0)

##      mean(P427)
## 1              1
```

Las personas que trabajan desde la oficina son empleados de empresas regular.

¿Cuánto ganan en promedio (pesos colombianos) las personas que trabajan desde la oficina?

```
df_w %>%
  filter(P6880 == 7) %>%
  select(P6500, P427) %>%
  summarize(mean(P6500) * 1000000)

##      mean(P6500) * 1e+06
## 1              875393.2
```

¿Cuánto reciben estas personas de subsidio (pesos colombianos) para transporte?

Análisis descriptivo



```
df_w %>%
  filter(P6880 == 7) %>%
  select(P6545BS1, P427) %>%
  summarize(mean(P6545BS1) * 1000)

##    mean(P6545BS1) * 1000
## 1                72950.62
```

## Capítulo 4

### Resultados y discusión.

Este estudio representa un breve análisis descriptivo sobre la base de datos del uso del tiempo de los individuos de la población civil no institucional residente en todo el territorio nacional, como puede apreciarse, los resultados de los análisis realizados para la variable tiempo en horas de sueño que utilizan las personas según las encuestas de este estudio dieron un promedio de 8 horas. Lo que se considera como tiempo suficiente de descanso; así mismo, hay que resaltar, la evidencia propuesta en la *tabla 2* de frecuencias la cual arroja un 88% de los individuos de las encuestas descansan entre 5 y 10 horas por día de sueño. En la *figura 2* el histograma de frecuencia también muestra esta tendencia a que los individuos suelen dormir entre 5 y 10 horas al día.

En los análisis realizados del capítulo H tiempo en trabajo he ingreso se obtuvieron en los resultados, un 80% de los encuestados trabajan desde las oficinas, lo que puede dar a

Analisis descriptivo

entender que existe una mayor población de personas trabajadoras para empresas, que cuentan con un ingreso mensual de casi 900 pesos colombianos, según los resultados obtenidos en el apartado *relaciones entre variables*. Hay que destacar que en la *tabla 8* podemos observar que el 97% de la población ganan alrededor de 2 millones de pesos. Finalmente, la clase obrera (trabajadores en empresas regulares) son las más abundantes.

## Lista de referencias

Página oficial de ANDA. Descripción de variables.

<http://microdatos.dane.gov.co/index.php/catalog/552/datafile/F31/?offset=300&limit=100#page=F30&tab=data-dictionary> American Psychological Association. (2010). *Manual de Publicaciones de la American Psychological Association* (6 ed.). (M. G. Frías, Trad.) México, México: El Manual Moderno.

Página oficial de ANDA. Encuesta nacional de uso del tiempo 2016 - 2017. base de datos.[http://microdatos.dane.gov.co/index.php/catalog/552/get\\_microdata](http://microdatos.dane.gov.co/index.php/catalog/552/get_microdata)

Página oficial de ANDA. Encuesta nacional de uso del tiempo 2016 – 2017. Descripción de la operación estadística.

<http://microdatos.dane.gov.co/index.php/catalog/552/study-description>

## Apéndice

**Figura 1.** Grafico de caja, visualización de los valores extremos.

Figura 2.

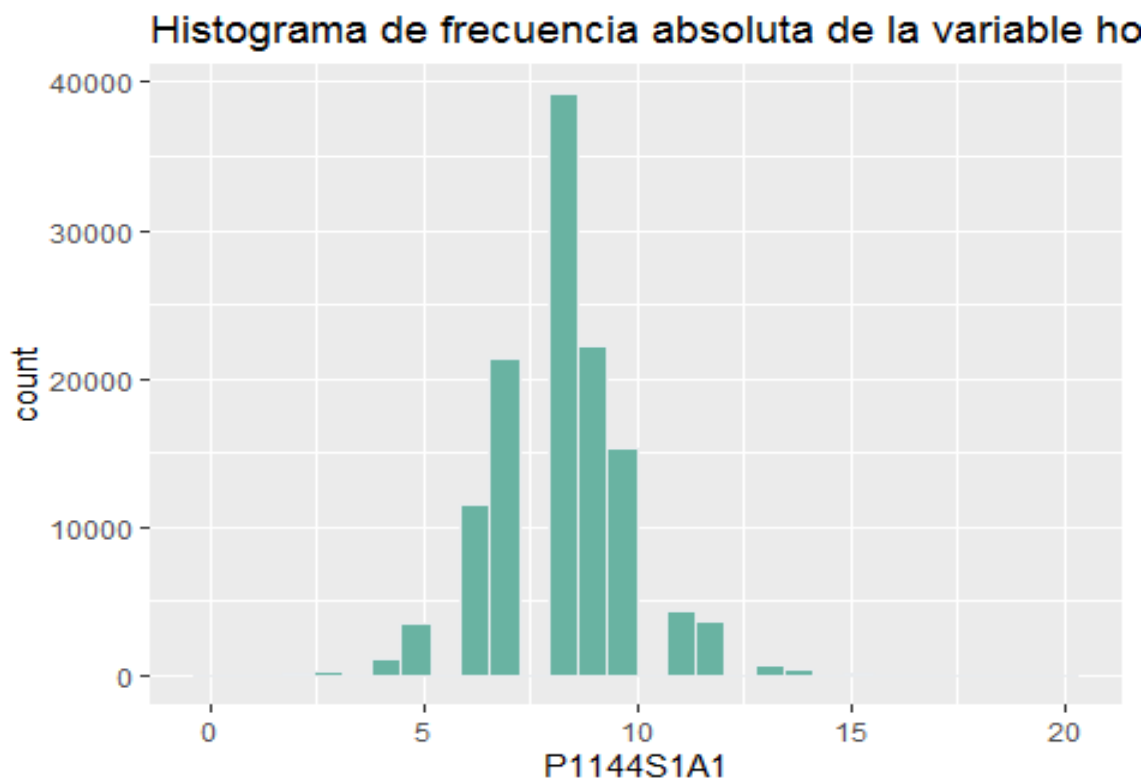
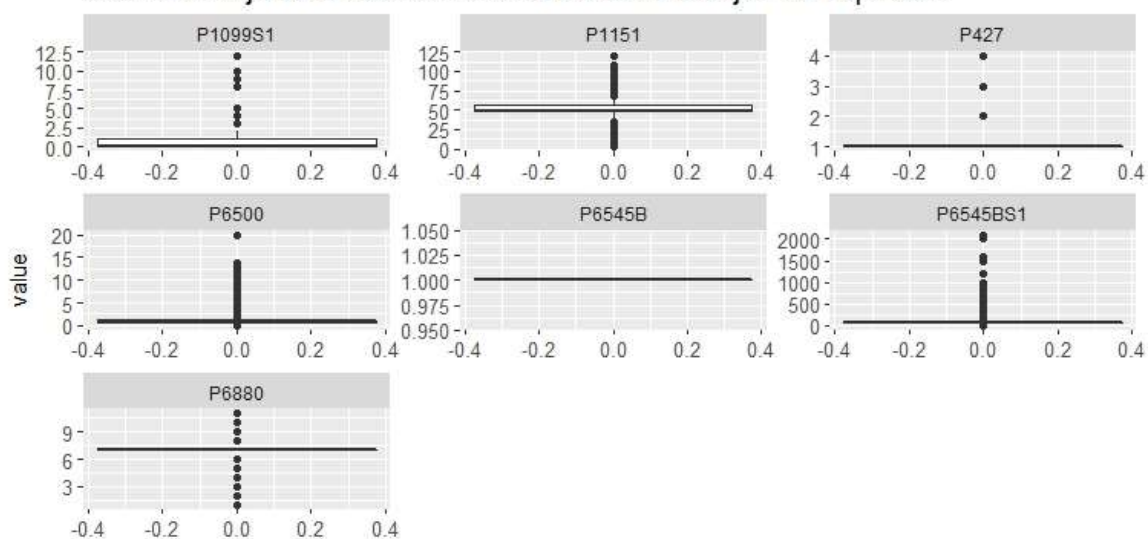


Figura 3. Gráfico de caja y bigotes, observación de los valores extremos

Gráfico de caja de la variables 'uso de horas de trabajos' del capítulo H



Análisis descriptivo

**Figura 4.** Histograma de frecuencia en las 7 variables, obsérvese la frecuencia en la variable P6880; la cual tiene lugar en “empleados de trabajo de oficina”.

