

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN

ESCUELA DE ESTADÍSTICA

PRIMER TRABAJO

Análisis descriptivo de datos (10%)

SEM. 01 – 2021

# cédula usada	1001463871
----------------	------------

Grupo:	1
--------	---

Nombre Completo	Jean Carlo Montoya Castro	D. I.	1000872579
Nombre Completo	Juan Camilo Molina Roncancio	D. I.	1001463871
Nombre Completo	Lina Juliana Mora Fernandez	D. I.	1005059850

La base de datos con la que le corresponde trabajar se obtiene como una muestra aleatoria de datos de una gran base. Esta base contiene las siguientes variables: **GENERO** (HOMBRE o MUJER), **ESTATURA** (en cm. del estudiante), **MASA** (en kg del estudiante), **EDAD** (en años), **PAPA** (obtenido en el semestre anterior), **VMATRI** (valor de matrícula pagada en el semestre actual, en miles de pesos), **ESTRATO** y **TRAB** (trabaja SI o NO).

Cada pareja debe generar una base de datos, que corresponderá a una muestra de tamaño 100 de la base original. Los datos originales están en el archivo: “**BASE.txt**”, el cual está disponible en Moodle. Para hacerlo debe seguir los siguientes pasos:

Procedimiento para generar archivo de datos para el trabajo

Primero debe descargar la base de datos. Luego En R, ejecutar el comando

```
base <- read.table(file.choose(), header=T)
```

Desplegará un cuadro de exploración y debe buscar el archivo BASE.txt para cargarlo.

Copiar el siguiente código en R sin modificar nada

```
genera <- function(cedula) {  
  set.seed(cedula)  
  data <- base[sample(1:125,100),]  
  data  
}
```

Para crear la base de datos con la cual trabajara, debe ejecutar la siguiente línea:

```
datos <- genera(cedula)
```

“**cedula**” representa el número de cedula de uno de los dos integrantes y **datos** representa el nombre que ustedes darán a su muestra de trabajo. Se recomienda generar los datos una sola vez y en un solo computador, pues el generador de cada computador es diferente y aunque use la misma cédula, puede generar muestras diferentes. Esto es importante, ya que el trabajo es en parejas y debe usar la misma muestra. En el trabajo debe indicar cuál número de cédula usó. Se aclara que ambos estudiantes deben pertenecer al mismo grupo en el cual están matriculados.

Usando la base de datos generada, responda a las siguientes preguntas:

1. (25 pts.) Para las variables **MASA**, **ESTATURA** y **EDAD**, calcule: Media, Mediana y desviación estándar e interprete sus resultados en el contexto de la información.

Para la variable **ESTRATO**, elabore una tabla de porcentajes y el respectivo diagrama de barras. Comente los resultados obtenidos en el gráfico.

2. (25 pts.) Considere la variable **PAPA** y elabore un histograma de porcentajes. Comente sobre la forma obtenida en el histograma. Si se discrimina por **GENERO**, ¿Puede afirmarse que el **PAPA** se comporta para Hombres y Mujeres? Justifique su respuesta usando el gráfico o los gráficos que considere pertinentes.

3. (25 pts.) Considere la variable **VMATRI**. En una tabla muestre la media y desviación estándar discriminando por Estrato. ¿Puede afirmarse que el valor promedio de matrícula pagado por los estudiantes es similar para todos los estratos? Justifique su respuesta usando el gráfico o los gráficos que considere pertinentes.

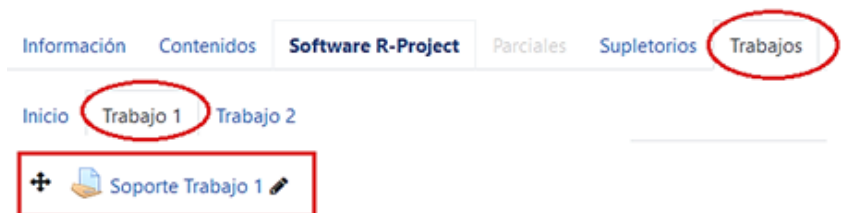
4. (25 pts.) Considere las variables **ESTRATO** y **TRAB**. Elabore una tabla de doble entrada con los respectivos porcentajes. ¿Se puede afirmar que el porcentaje de estudiantes de Estratos 3 y 4 que no trabajan es superior al porcentaje de estudiantes de estratos 3 y 4 que trabajan? Justifique su respuesta.

El trabajo debe cargarse

en la carpeta

“Soporte Trabajo 1”
antes

de las 4:00 pm:



TRABAJO ESTADISTICA # 1

1) Resúmenes numéricos:

```
> summary(datos)
      GENERO      ESTATURA      MASA      EDAD      PAPA      VMATRI      ESTRATO TRAB
HOMBRE:69  Min.   :143.3  Min.   :44.90  Min.   :16.00  Min.   :2.900  Min.   : 111.0  1: 5  NO:63
MUJER :31  1st Qu.:163.1  1st Qu.:59.55  1st Qu.:20.27  1st Qu.:3.500  1st Qu.: 511.5  2:15  SI:37
          Median :168.5  Median :67.30  Median :21.65  Median :3.700  Median : 946.0  3:40
          Mean   :169.6  Mean   :65.92  Mean   :22.17  Mean   :3.691  Mean   :1162.9  4:28
          3rd Qu.:175.3  3rd Qu.:72.75  3rd Qu.:24.12  3rd Qu.:3.900  3rd Qu.:1748.2  5: 9
          Max.   :192.5  Max.   :82.10  Max.   :30.60  Max.   :4.700  Max.   :4367.4  6: 3

      mean      sd      IQR      cv      0%      25%      50%      75%      100%      n
EDAD      22.173  3.117363   3.85  0.14059274  16.0  20.275   21.65  24.125   30.6  100
ESTATURA 169.551  9.402949  12.25  0.05545794 143.3 163.050 168.50 175.300 192.5 100
MASA      65.916  8.736996  13.20  0.13254742  44.9  59.550   67.30  72.750   82.1  100
```

Tablas resumen:

Masa		
Media	Mediana	Desviación estándar
65.916	67.3	8.736996

Estatura		
Media	Mediana	Desviación estándar
169.551	168.5	9.402949

Edad		
Media	Mediana	Desviación estándar
22.173	21.65	3.117363

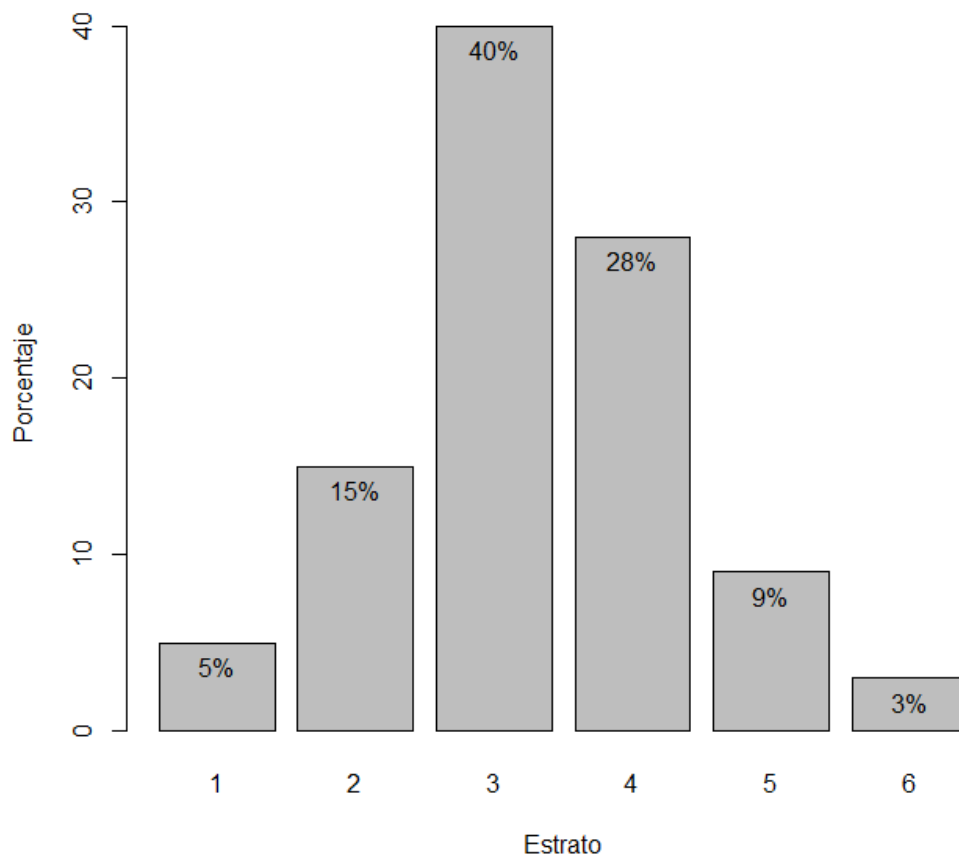
Masa: Para empezar podemos notar que los datos en esta variable son relativamente simétricos la media nos dice que en promedio los estudiantes de la muestra pesan 65.9 kg, siendo esta medida un indicativo de poco interés, sin embargo, podemos notar que la variable tiene una desviación estándar bastante grande de 8.7 kg combinado a la similitud de la media y la mediana podemos concluir que aunque es una distribución relativamente simétrica las masas de los estudiantes por separado pueden variar mucho con respecto a las medidas centrales.

Estatura: la estatura media en Colombia es de 160 cm, por lo que los estudiantes de la universidad son más altos que la media colombiana con un promedio de 169.5 cm. igualmente podemos observar que la media y la mediana al estar relativamente cerca podemos inferir que los datos están distribuidos simétricamente, además fijándonos en la desviación estándar (unos 8.7 cm aproximadamente) nuevamente podemos concluir que esta variable tiene una variabilidad hacia arriba y hacia abajo con respecto a la mediana bastante alta.

Edad: la edad media es de 22 años, lo cual es relativamente joven, no sabemos en qué porcentaje de la carrera están los estudiantes pero posiblemente ingresaron a la universidad poco después de graduarse del colegio, por lo menos la mitad de los encuestados tienen casi 22 años, por lo que sí ingresaron inmediatamente de obtener el título de bachiller académico están terminando la carrera o llevan un porcentaje bastante avanzado, si por el contrario apenas ingresaron, saldrán con el título antes de los 30 años. Analizando la mediana con respecto a la media y la variación estándar podemos concluir que los datos en cuanto edades están distribuidas de una manera bastante simétrica y con relativa poca variación.

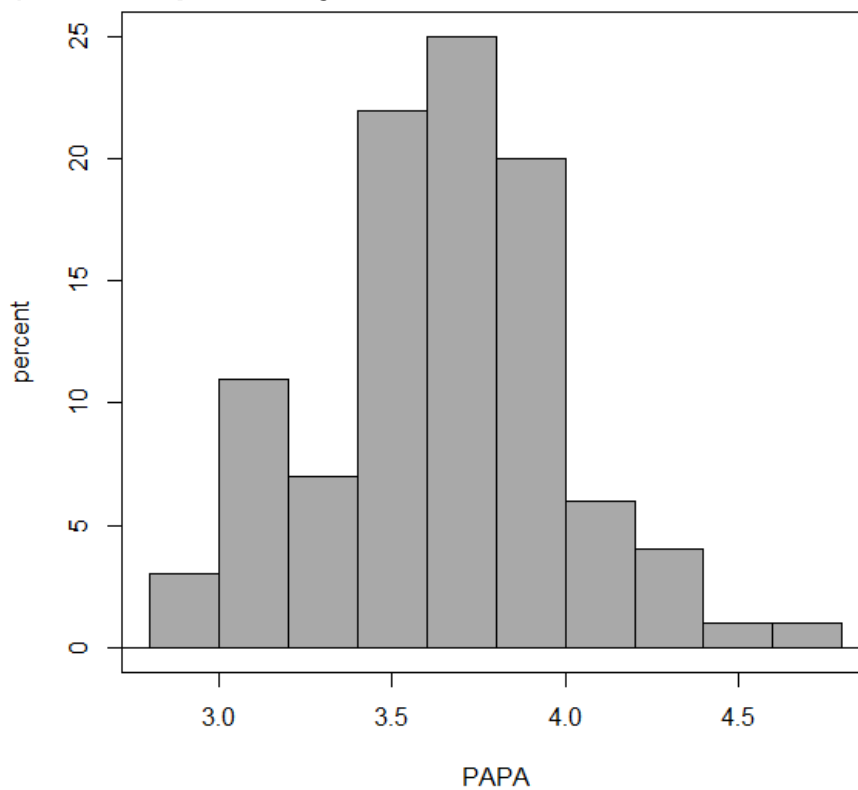
Columna de porcentajes	
Estrato	Porcentaje de estudiantes
1	5%
2	15%
3	40%
4	28%
5	9%
6	3%
Total	100%

Estrato socioeconómico de los estudiantes



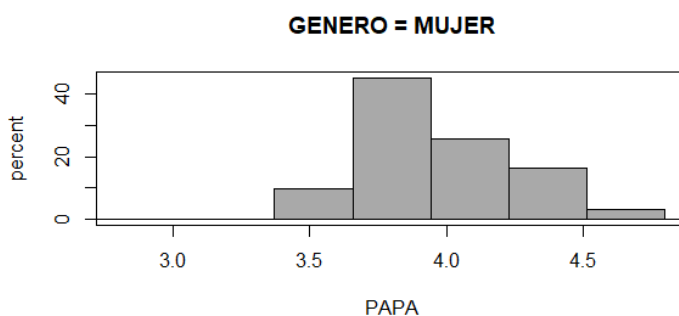
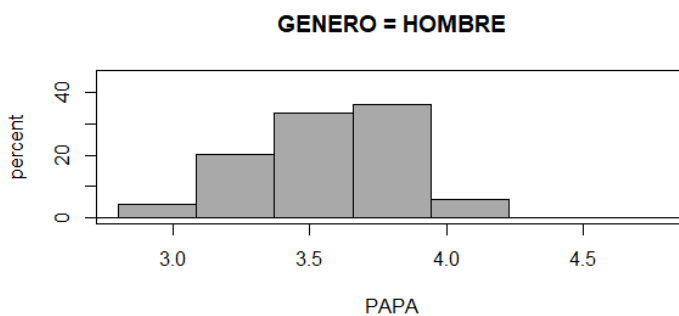
Según lo que observamos en el diagrama de barras podemos concluir que la mayoría de estudiantes son de estratos 3 y 4 (casi un 70%), además hay más personas de estrato 1 y 2 que de 5 y 6. Podemos inferir que el estrato socioeconómico puede afectar al ingreso o no a la universidad.

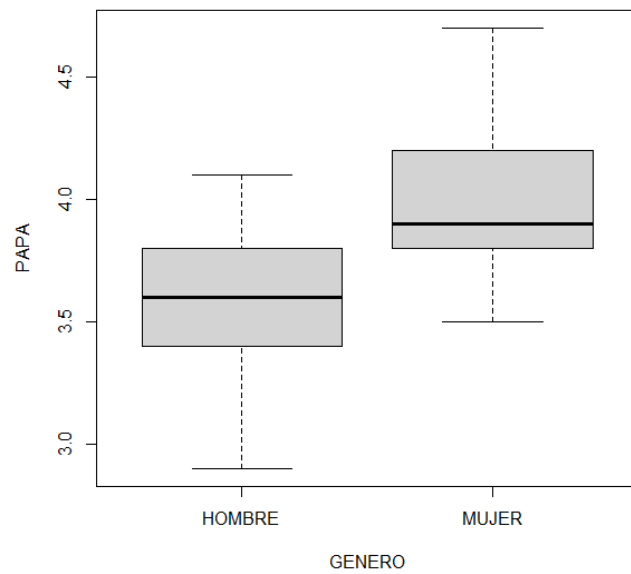
2)Tabla de porcentaje del PAPA



En el histograma se ve que el PAPA entre 3.0 y 4.0 es el más frecuente entre los estudiantes, probablemente porque es el necesario para permanecer como estudiante de la universidad, PAPAs más altos que 4 son poco comunes entre los encuestados, inclusive no hay nadie con un PAPA de 5 mostrando la dificultad para subirlo en la universidad.

Porcentaje del PAPA en cuanto al género:



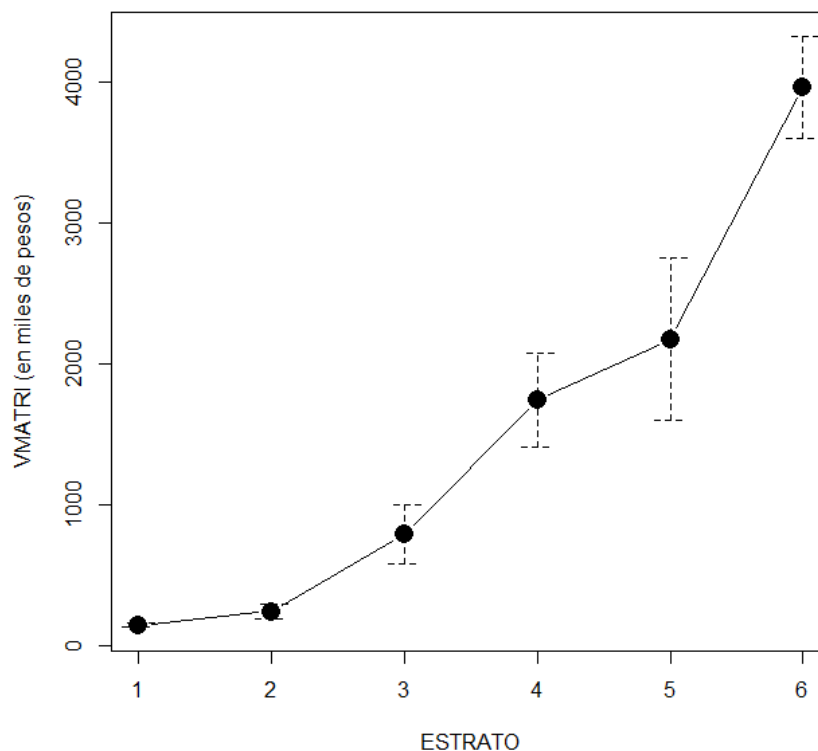


Conclusión: en las gráficas se puede observar que en términos de PAPA, a las mujeres les va mejor que a los hombres, teniendo un mínimo de 3.5 hasta uno mayor que 4.5 y los hombres van desde 3.0 hasta 4.0 como máximo; gracias a la gráfica de cajas, podemos ver que las medias tienen una separación no muy grande pero tienden hacia el promedio 3.5 que no es uno demasiado bueno pero es suficiente para mantener el estado de “estudiante”.

3)

	Estrato					
	1	2	3	4	5	6
Media de VMATRI	144.3000	241.1000	792.4425	1741.5429	2176.8000	3967.4000
Desviación estándar de VMATRI	12.64456	56.45305	208.90146	331.53919	578.13444	363.12423

Grafica de medias



En la tabla se evidencia que el valor promedio sube exponencialmente a medida que sube de estrato, por lo que el valor de la matrícula no es igual para personas de estratos diferentes, a mayores ingresos, mayor valor de la matrícula; vemos además que las personas de estrato 1 y 2 tienen una desviación estándar relativamente baja comparado con los otros estratos, por lo que el valor de la matrícula de la media es bastante similar y debido a la forma de asignar el valor de la matrícula en la universidad, su situación financiera también es similar; con los estratos 3,4,5,6 es diferente, es posible que los trabajos sean tan variados que por eso la desviación es tan grande, de ahí a que el valor de la matrícula de la media pueda tener una diferencia de hasta 500.000 pesos colombianos.

4)

Tabla de doble entrada: ESTRATO vs. TRAB.			
Estrato	No	Si	Total
Estrato 1	3%	3%	5%
Estrato 2	9%	6%	15%
Estrato 3	26%	14%	40%
Estrato 4	17%	11%	28%
Estrato 5	6%	3%	9%
Estrato 6	2%	1%	3%
Total	63%	37%	100%

Según la tabla podemos afirmar que hay más estudiantes que no trabajan de estratos 3 y 4 de los que trabajan de estratos 3 y 4 con un 43% contra un 25% respectivamente. Esto puede relacionarse al hecho de que probablemente al tener una posición económica ligeramente superior muchos prefieren (muchas veces apoyados por sus padres) enfocarse principalmente en la universidad, por lo tanto no se ven en la necesidad de realizar las dos actividades al mismo tiempo. Esto se puede ver reflejado en la cantidad de tiempo disponible que poseen las personas que trabajan y estudian simultáneamente, beneficiando su promedio académico.