

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ**  
**FACULTAD DE CIENCIAS E INGENIERÍA**

**ESTADÍSTICA PARA INGENIERÍA**

1ra Práctica (tipo c)  
 (Ciclo de Verano 2023)

**Indicaciones Generales:**

- Esta práctica debe ser desarrollada por un grupo de máximo 3 alumnos del mismo horario.
- Justifique claramente sus soluciones, presente los códigos y resultados obtenidos en R, interprete estos resultados y escriba respuestas completas teniendo en cuenta el contexto de los ejercicios propuestos; todo esto influirá en la calificación.
- La presentación, la ortografía y la gramática de los trabajos influirán en la calificación.
- La solución debe ser subida en único archivo PDF en la tarea PC1 de su correspondiente horario del curso. La primera hoja debe presentar el código PUCP, nombre completo y correo electrónico de todos los integrantes del grupo.

Puntaje debido al cuestionario: 20 puntos

Puntaje total: 20 puntos

**Cuestionario:**

La base de datos **Encuesta.csv** contiene datos de las siguientes variables medidas en una muestra de 6400 trabajadores:

Variable	Descripción
edad	Edad
estcivil	Estado civil
ingrmens	Ingreso mensual
educacion	Educación
satlab	Satisfacción laboral
genero	Género
familia	Número de integrantes de la familia

Al leer este conjunto de datos mediante la función `read.csv()`, utilice el argumento `stringsAsFactors = TRUE`, de manera que las variables cualitativas sean leídas como factores. En este caso, la sintaxis sería:

```
d <- read.csv("Encuesta.csv", stringsAsFactors = TRUE)
```

Adicionalmente, ejecute el siguiente código para reordenar los niveles de las variables ordinales (educación y satisfacción laboral), de manera que se mantenga el orden correcto de las categorías.

```
d$educacion <- with(d, factor(educacion, levels=c('Bachillerato incompleto',
                                                'Bachillerato completo', 'Universitario incompleto',
                                                'Universitario completo', 'Post-universitario'), ordered=TRUE))
d$satlab <- with(d, factor(satlab, levels=c('Muy insatisfecho', 'Insatisfecho', 'Neutro',
                                                'Satisfecho', 'Muy satisfecho'), ordered=TRUE))
```

### **Pregunta 1 (2.0 puntos)**

Presente una tabla de distribución de frecuencias para el número de integrantes de la familia, además presente una gráfica adecuada.

Presente un análisis de los resultados, incluyendo dos conclusiones importantes.

### **Pregunta 2 (3.0 puntos)**

Presente una tabla de distribución de frecuencias para el ingreso mensual con 10 intervalos de igual amplitud y presente las gráficas correspondientes para las frecuencias simples y frecuencias acumuladas.

Presente un análisis de los resultados, incluyendo dos conclusiones importantes.

### **Pregunta 3 (4.0 puntos)**

Presente gráficas para las frecuencias relativas simples del ingreso mensual por nivel de educación.

Realice un análisis descriptivo comparativo completo para el ingreso mensual por nivel de educación, incluyendo cuartiles, medidas de tendencia central y medidas de dispersión (rango y rango intercuartil).

Con base en los resultados compare la distribución del ingreso mensual entre trabajadores de los diferentes niveles de educación. Presente una conclusión importante.

### **Pregunta 4 (2.0 puntos)**

Genere una nueva variable que indique 1 si la persona tiene un ingreso mensual mayor o igual a 5000 soles, y 0 en caso contrario. Encuentre la proporción de trabajadores que presentan sueldos igual o mayor a 5000 soles, usando directamente los datos y a través de la tabla de frecuencias obtenida en la pregunta 2. Compare estas dos proporciones y, caso sean diferentes, indique a qué se deben esas diferencias.

Para la nueva variable, desarrolle un gráfico de sectores circulares.

### **Pregunta 5 (3.0 puntos)**

Dentro de un conjunto de datos, un dato se considera atípico por el lado superior si el valor del dato supera (estrictamente) el siguiente límite:

$$LS = Q3 + 1.5 \cdot RIC$$

Calcule la media aritmética de los ingresos mensuales de aquellos trabajadores que presentan ingresos mensuales atípicos por el lado superior.

Considerando estos resultados y los resultados de la pregunta 3, indique cuáles son las medidas de tendencia central y las medidas de dispersión que se debería usar para representar estos datos de los ingresos mensuales. Justifique su respuesta.

### **Pregunta 6 (3.0 puntos)**

Para cada nivel de la variable género, elabore un gráfico para la variable edad.

Calcule las estadísticas necesarias de la variable edad por nivel de género.

Realice una comparación en términos de cuartiles, tendencia central y dispersión (rango y RIC).

### **Pregunta 7 (3.0 puntos)**

Evalúe la veracidad o falsedad de cada una de las siguientes afirmaciones. Justifique su respuesta (presente sus códigos y resultados en R como parte de su justificación)

- a) (1.0 punto) Para representar la edad es adecuado usar la media.
- b) (1.0 punto) Se puede afirmar que la edad es la misma para todos los niveles de sexo.
- c) (1.0 punto) La mediana de la variable satisfacción laboral es la categoría "Satisfecho".

Los profesores del curso

San Miguel, 16 de enero del 2023