

UNIVERSIDAD DEL NORTE
INFERENCIA ESTADÍSTICA
PARCIAL I.

Fecha de realización: **18-02-2019**



Nombre: _____ Nota: _____

Considere la población de acuerdo a la siguiente información. Se recopilaron unos datos sobre el tratamiento policial de personas detenidas en Barranquilla por simple posesión de pequeñas cantidades de marihuana. Las variables de estudio son:

- Publicado: Si el arrestado fue puesto en libertad o no con una citación; un factor con niveles: No; SI.
- Color: La raza del arrestado; un factor con niveles: Blanco; Negro.
- Año del arresto: De 1997 hasta 2002, un vector numérico.
- Edad: Edad del arrestado.
- Peso: Peso en Kg. del arrestado.
- Sexo: Un factor con niveles: Masculino (M), Femenino (F)
- Empleado: Un factor con niveles: Es empleado (SI), No es empleado (NO).
- Origen: Un factor con niveles: Si, NO
- Registro: Número de bases de datos de la policía (de arrestos anteriores, condenas anteriores, estado de libertad condicional, etc. - 6 en total) en las que apareció el nombre del arrestado; un vector numérico.

La siguiente tabla muestra un resumen de la información sobre las variables:

Publicado	Color	Año	Edad	Peso	Sexo	Empleado	Origen	Registro
No:892	Blanco: 3938	Min: 1997	Min:12	Min: 40	F: 443	NO: 1115	NO: 771	Min: 0
SI: 4334	Negro: 1288	Q1: 1998	Q1: 18	Q1:54.45	M: 4783	SI: 4111	SI: 4455	Q1:0
		Mediana: 2000	Mediana: 21	Mediana: 60				Mediana: 1
		Media: 2000	Media: 23.85	Media: 63				Media: 1.636
		Q3: 2001	Q3: 27	Q3: 68.13				Q3: 3
		Max: 2002	Max: 66	Max: 98.8				Max: 6

Cuadro 1: Variables.

1. Identificar la unidad de análisis correcta para las siguientes preguntas de investigación
 - a) ¿Está en aumento el porte ilegal de marihuana en jóvenes de la ciudad de Barranquilla?
 - b) ¿Existe una fuerte relación entre el porte ilegal de marihuana y la condición laboral?
 - c) ¿El consumo de marihuana afecta el peso corporal de las mujeres?
2. Para los siguientes casos, identifique el tipo de muestreo realizado. Si es probabilístico, identifique si es muestreo (aleatorio simple, estratificado, por conglomerados o sistemático). Justifique su respuesta.
 - a) Caso 1: El objetivo es documentar las razones del consumo o porte ilegal de marihuana en jóvenes con edad entre 12 y 20 años. Para cumplir dicho propósito se selecciono una muestra. Las personas se seleccionaron del reporte de la policía, de acuerdo a sus edades y lugar de residencia, de tal forma que siguieran viviendo en Barranquilla. De esta manera se seleccionaron a 100 jóvenes que seguían viviendo en Barranquilla, los cuales aceptaron dar a conocer sus experiencias.
3. Realice una explicación de cada una de las medidas mostradas en el cuadro 1 para las variable EDAD.
 - b) Caso 2: El objetivo es determinar la relación existente entre el peso de las personas detenidas y las variables asociadas a (Sexo, Empleado y Edad). Se enumera a todos las personas registradas desde el inicio (1997) hasta el final del estudio (2002) y luego se generaría 100 números aleatorios con el programa R con rango entre [1,5226]. Los números aleatorios generados se seleccionan de acuerdo al número asignado a cada persona.

4. Identifique el tipo de variable (cualitativa o cuantitativa, si es cuantitativa: discreta o continua) y la escala de medida, para cada una de las variables del estudio

Variabes	Tipo de variable	Escala de medida
Publicado		
Color		
Año del arresto		
Edad		
Peso		
Sexo		
Empleado		
Origen		
Registro		

5. Supongamos que en el informe se adhiere la variable centros de detención (CD), en total 7 centros de detención en la ciudad de Barranquilla. En el CD 1 se detuvieron a 350 personas, CD 2 se detuvieron a 866 personas, CD 3 se detuvieron a 742 personas, CD 4 se detuvieron a 1145 personas, CD 5 se detuvieron a 1285 personas, CD 6 se detuvieron a 304 personas y en el CD 7 se detuvieron a 534 personas. Pensamos que puede haber diferencias entre las variables : edades, empleado, sexo y origen, por lo que nos interesa que en la muestra estén representados todos los centro de detención. Si queremos elegir una muestra de 650 personas detenidas, determine:
- ¿Qué tipo de muestreo probabilístico se quiere realizar?
 - Encontrar el número de detenidos que debemos elegir de cada CD para optimizar la muestra, de acuerdo con la distribución de la población en los 7 centros de detención.
6. De acuerdo al siguiente gráfico, determine si la muestra seleccionada con muestreo aleatorio simple, de tamaño $n = 100$ proviene de una población normalmente distribuida. Es decir, tiene evidencias suficientes para aceptar o rechazar la hipótesis de normalidad. De acuerdo Shapiro-Wilk normality test: $W = 0.99228$, $p\text{-value} = 0.8408$

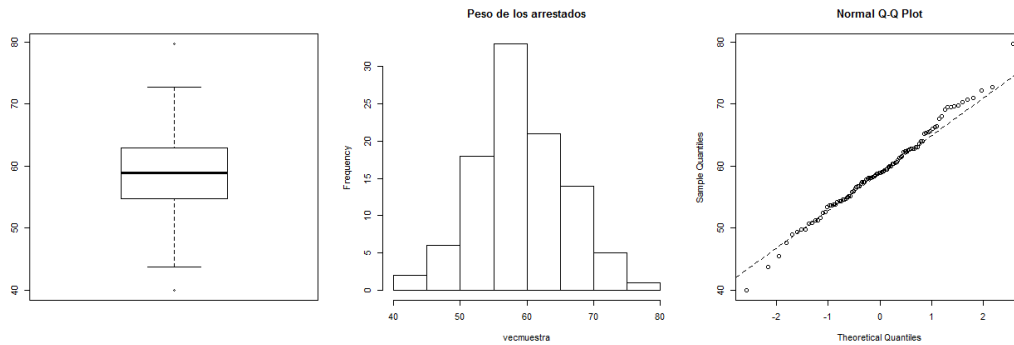


Figura 1: Graficos relacionados con la var Peso.

UNIVERSIDAD DEL NORTE
INFERENCIA ESTADÍSTICA
PARCIAL I.

Fecha de realización: **18-02-2019**



Nombre: _____ Nota: _____

Considere la población de acuerdo a la siguiente información. Se recopilaron unos datos sobre el tratamiento policial de personas detenidas en Barranquilla por simple posesión de pequeñas cantidades de marihuana. Las variables de estudio son:

- Publicado: Si el arrestado fue puesto en libertad o no con una citación; un factor con niveles: No; SI.
- Color: La raza del arrestado; un factor con niveles: Blanco; Negro.
- Año del arresto: De 1997 hasta 2002, un vector numérico.
- Edad: Edad del arrestado.
- Peso: Peso en Kg. del arrestado.
- Sexo: Un factor con niveles: Masculino (M), Femenino (F)
- Empleado: Un factor con niveles: Es empleado (SI), No es empleado (NO).
- Origen: Un factor con niveles: Si, NO
- Registro: Número de bases de datos de la policía (de arrestos anteriores, condenas anteriores, estado de libertad condicional, etc. - 6 en total) en las que apareció el nombre del arrestado; un vector numérico.

La siguiente tabla muestra un resumen de la información sobre las variables:

Publicado	Color	Año	Edad	Peso	Sexo	Empleado	Origen	Registro
No:892	Blanco: 3938	Min: 1997	Min:12	Min: 40	F: 443	NO: 1115	NO: 771	Min: 0
SI: 4334	Negro: 1288	Q1: 1998	Q1: 18	Q1:54.45	M: 4783	SI: 4111	SI: 4455	Q1:0
		Mediana: 2000	Mediana: 21	Mediana: 60				Mediana: 1
		Media: 2000	Media: 23.85	Media: 63				Media: 1.636
		Q3: 2001	Q3: 27	Q3: 68.13				Q3: 3
		Max: 2002	Max: 66	Max: 98.8				Max: 6

Cuadro 2: Variables.

1. Identificar la unidad de análisis correcta para las siguientes preguntas de investigación
 - a) ¿Está en aumento el porte ilegal de marihuana en jóvenes de la ciudad de Barranquilla?
 - b) ¿Existe una fuerte relación entre el porte ilegal de marihuana y la condición laboral?
 - c) ¿El consumo de marihuana afecta el peso corporal de las mujeres?
2. Para los siguientes casos, identifique el tipo de muestreo realizado. Si es probabilístico, identifique si es muestreo (aleatorio simple, estratificado, por conglomerados o sistemático). Justifique su respuesta.
 - a) Caso 1: El objetivo es documentar las razones del consumo o porte ilegal de marihuana en jóvenes con edad entre 12 y 20 años. Para cumplir dicho propósito se selecciono una muestra. Las personas se seleccionaron del reporte de la policía, de acuerdo a sus edades y lugar de residencia, de tal forma que siguieran viviendo en Barranquilla. De esta manera se seleccionaron a 100 jóvenes que seguían viviendo en Barranquilla, los cuales aceptaron dar a conocer sus experiencias.
- b) Caso 2: El objetivo es determinar la relación existente entre el peso de las personas detenidas y las variables asociadas a (Sexo, Empleado y Edad). Se enumera a todos las personas registradas desde el inicio (1997) hasta el final del estudio (2002) y luego se generaría 100 números aleatorios con el programa R con rango entre [1,5226]. Los números aleatorios generados se seleccionan de acuerdo al número asignado a cada persona.
3. Realice una explicación de cada una de las medidas mostradas en el cuadro 2 para las variable PESO.

4. Identifique el tipo de variable (cualitativa o cuantitativa, si es cuantitativa: discreta o continua) y la escala de medida, para cada una de las variables del estudio

VARIABLES	Tipo de variable	Escala de medida
Publicado		
Color		
Año del arresto		
Edad		
Peso		
Sexo		
Empleado		
Origen		
Registro		

Cuadro 3: Variables.

5. Supongamos que en el informe se adhiere la variable centros de detención (CD), en total 7 centros de detención en la ciudad de Barranquilla. En el CD 1 se detuvieron a 350 personas, CD 2 se detuvieron a 866 personas, CD 3 se detuvieron a 742 personas, CD 4 se detuvieron a 1145 personas, CD 5 se detuvieron a 1285 personas, CD 6 se detuvieron a 304 personas y en el CD 7 se detuvieron a 534 personas. Pensamos que puede haber diferencias entre las variables :edades, empleado, sexo y origen, por lo que nos interesa que en la muestra estén representados todos los centro de detención. Si queremos elegir una muestra de 650 personas detenidas, determine:
- ¿Qué tipo de muestreo probabilístico se quiere realizar?
 - Encontrar el número de detenidos que debemos elegir de cada CD para optimizar la muestra, de acuerdo con la distribución de la población en los 7 centros de detención.
6. De acuerdo al siguiente gráfico, determine si la muestra seleccionada con muestreo aleatorio simple, de tamaño $n = 100$ proviene de una población normalmente distribuida. Es decir, tiene evidencias suficientes para aceptar o rechazar la hipótesis de normalidad. De acuerdo Shapiro-Wilk normality test: $W = 0.85597$, $p\text{-value} = 1,983e-08$

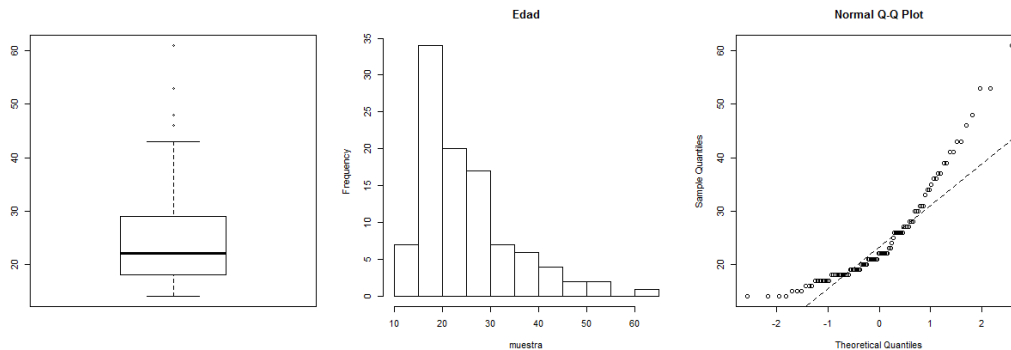


Figura 2: Graficos relacionados con la var Edad.

1. Identificar la unidad de análisis correcta para las siguientes preguntas de investigación
 - a) ¿Está en aumento el porte ilegal de marihuana en jóvenes de la ciudad de Barranquilla? **R/** Muestra de personas (hombres y mujeres) de distintas edades, para comparar las edades de aquellos que fueron hallados con marihuana ilegal.
 - b) ¿Existe una fuerte relación entre el porte ilegal de marihuana y la condición laboral? **R/**. Grupo de personas con o sin empleo, a los cuales se les encontró marihuana.
 - c) ¿El consumo de marihuana afecta el peso corporal de las mujeres? **R/** Muestra de personas (hombres y mujeres), para comparar los distintos pesos y ver si existe afectaciones en el peso de las mujeres.
2. Para los siguientes casos, identifique el tipo de muestreo realizado. Si es probabilístico, identifique si es muestreo (aleatorio simple, estratificado, por conglomerados o sistemático). Justifique su respuesta.
 - a) Caso 1: No probabilístico
 - b) Caso 2: Probabilístico. Muestreo aleatorio simple.
3. Realice una explicación de cada una de las medidas mostradas en el cuadro 2 para las variable PESO-Edad.
 - a) Peso: De las personas encontradas con marihuana ilegal, se encontró una persona con un peso mínimo de 40 kg y uno máximo de 98.8 kg. De acuerdo al cuartil 1, aproximadamente 1307 personas tienen un peso menor a 54.45 kg. Un 25 % adicional tiene un peso menor a 60 y el 75 % de todos los capturados tiene un peso menor a 68.13 kg. El peso promedio se encuentra entre 63 kg.
 - b) Edad: De las personas encontradas con marihuana ilegal, se encontró una persona con una edad mínimo de 12 años y uno máximo de 66 años. De acuerdo al cuartil 1, aproximadamente 1307 personas tienen una edad menor a 18 años. Un 25 % adicional tiene una edad menor a 21 años y el 75 % de todos los capturados tiene una edad menor a 27 años. La edad promedio se encuentra es de 24 años.
4. Identifique el tipo de variable (cualitativa o cuantitativa, si es cuantitativa: discreta o continua) y la escala de medida, para cada una de las variables del estudio

Variables	Tipo de variable	Escala de medida
Publicado	Cualitativa	Nominal
Color	Cualitativa	Nominal
Año de arresto	Cuantitativa-Discreto	Intervalo
Edad	Cuantitativa-Discreto	Razón
Peso	Cuantitativa-Continua	Razón
Sexo	Cualitativa	Nominal
Empleado	Cualitativa	Nominal
Origen	Cualitativa	Nominal
Registro	Cuantitativa-Discreto	Razón

Cuadro 4: Variables.

5. Supongamos que en el informe se adhiere la variable centros de detención (CD), en total 7 centros de detención en la ciudad de Barranquilla. En el CD 1 se detuvieron a 350 personas, CD 2 se detuvieron a 866 personas, CD 3 se detuvieron a 742 personas, CD 4 se detuvieron a 1145 personas, CD 5 se detuvieron a 1285 personas, CD 6 se detuvieron a 304 personas y en el CD 7 se detuvieron a 534 personas. Pensamos que puede haber diferencias entre las variables :edades, empleado, sexo y origen, por lo que nos interesa que en la muestra estén representados todos los centro de detención. Si queremos elegir una muestra de 650 personas detenidas, determine:
 - a) ¿Qué tipo de muestreo probabilístico se quiere realizar?
 - b) Encontrar el número de detenidos que debemos elegir de cada CD para optimizar la muestra, de acuerdo con la distribución de la población en los 7 centros de detención.

CD	N	fh	nh
CD1	350	0.124378109	43.53233831
CD2	866	0.124378109	107.7114428
CD3	742	0.124378109	92.28855721
CD4	1145	0.124378109	142.4129353
CD5	1285	0.124378109	159.8258706
CD6	304	0.124378109	37.81094527
CD7	534	0.124378109	66.41791045

Cuadro 5: Variables.

6. De acuerdo al siguiente gráfico, determine si la muestra seleccionada con muestreo aleatorio simple, de tamaño $n = 100$ proviene de una población normalmente distribuida. Es decir, tiene evidencias suficientes para aceptar o rechazar la hipótesis de normalidad.
- a) Peso: Para un $\alpha = 0,05$, no existe suficiente evidencia para rechazar la hipótesis de normalidad de la población. Además se puede observar en los gráficos: diagrama de cajas e histograma que tienen un comportamiento muy parecido a la distribución normal.
 - b) Edad: Para un $\alpha = 0,05$, existe suficiente evidencia para no aceptar la hipótesis de normalidad. Además se puede observar en los gráficos: diagrama de cajas e histograma que su comportamiento no se parece a la distribución normal.