

## **Unidad 1: Estadística Descriptiva.**

### **1.0 Contenidos, objetivos y palabras clave.**

#### **Contenidos:**

|       |                                                                                   |    |
|-------|-----------------------------------------------------------------------------------|----|
| 1.0   | Contenidos, objetivos y palabras claves                                           | 1  |
| 1.1   | Resumen de inicio. Definiciones y propiedades principales.                        | 2  |
| 1.1.1 | Ejercicios resueltos, paso a paso.                                                | 3  |
| 1.1.2 | Ejercicios propuestos.                                                            | 11 |
| 1.2   | Medidas de posición y dispersión. Definiciones y propiedades principales.         | 12 |
| 1.2.1 | Ejercicios resueltos, paso a paso.                                                | 16 |
| 1.2.2 | Ejercicios propuestos.                                                            | 24 |
| 1.3   | Planteamientos problemáticos para aplicación de propiedades.                      | 25 |
| 1.3.1 | Ejercicios resueltos, paso a paso.                                                | 26 |
| 1.3.2 | Ejercicios propuestos.                                                            | 30 |
| 1.4   | Datos bidimensionales. Tablas de frecuencia conjunta, marginales y condicionales. | 31 |
| 1.4.1 | Ejercicios resueltos, paso a paso.                                                | 32 |
| 1.4.2 | Ejercicios propuestos.                                                            | 39 |

#### **Objetivos:**

- *Aplicar correctamente los modelos de probabilidad en el cálculo de probabilidades en problemas de aplicación.*
- *Calcular percentiles.*
- *Reconocer la conveniencia de la aplicación de distribuciones límite cuando el tamaño  $n$  de la muestra es muy grande.*

#### **Palabras Clave:**

*Población, muestra.  
Variable cuantitativa.  
Variables cualitativas.  
Escala de medición.  
Frecuencias.*

*Medidas de posición.  
Medidas de dispersión.  
Dato bidimensional.  
Distribución Bidimensional.*

## 1.1. Resumen de inicio. Definiciones y propiedades principales.

- ✓ Se llama **población** al conjunto de todos los objetos que constituyen el interés de estudio para un **estadístico** particular.
- ✓ Una **muestra** es cualquier subconjunto de la población.
- ✓ Se llama **muestreo** a la técnica de obtención de la muestra.
- ✓ Tipos de muestreo: **Probabilístico, No probabilístico**.
- ✓ Muestreo Probabilístico: **muestreo aleatorio simple, muestreo sistemático, muestreo estratificado, muestreo por conglomerado**.
- ✓ Muestreo no probabilístico: **muestreo por cuotas**.
- ✓ El ente de trabajo de la estadística es el **dato**.
- ✓ Se llama **Estadística** a la ciencia que se preocupa de estudiar las variables y sus comparaciones o relaciones para explicar su comportamiento actual, posterior o anterior. Para esto último ha desarrollado la **Estadística Inferencial**.
- ✓ El procedimiento básico de la estadística (de interés particular en estadística descriptiva) consiste en **recabar, organizar, analizar e interpretar** datos para extraer de ellos **información lógicamente aceptable**.
- ✓ El dato es un **registro** de la **variable** de interés en el momento en que ha finalizado la experiencia para observar el resultado que se logra. El registro de este resultado se llama dato.
- ✓ Para que sea útil la **información** que se logra a partir de los datos sobre la base de la cual se hará la toma de decisiones, los datos deben organizarse y mostrarse apropiadamente.
- ✓ El **tipo de datos** indicará el **método** que se usará en su análisis.
- ✓ La característica que se va a estudiar en los objetos de la población se llama **variable** y será designada por letras mayúsculas  $X, Y, Z$ , etc.
- ✓ Los **valores** de las variables serán denotados por letras minúsculas y subíndices en la forma  $x_i, y_i, z_i$ , etc., designando con ello el **dato u observación**  $i$ -ésima de la variable  $X, Y, Z$ , respectivamente.
- ✓ Las variables se clasifican en dos categorías generales: **variables cuantitativas y variables cualitativas**.
- ✓ Se llama **variable cuantitativa** a aquella cuyos valores medidos en una **escala numérica** (números reales).

- ✓ Se llama **variable cualitativa** a aquella cuyos valores representan **categorías** o **atributos**, que pueden clasificarse según un criterio o cualidad.
- ✓ Cada una de las distintas categorías en las que se puede presentar en una tabla a las variables categóricas o numéricas discretas, es llamada **clase**.
- ✓ Las variables numéricas continuas se clasifican por **intervalos** de números reales. El punto medio de cada intervalo se llama **marca de clase**.
- ✓ **La tabla de frecuencia** es un ordenamiento en filas y columnas, de clases o intervalos con sus respectivas frecuencias, que sirve para presentar los datos en forma resumida, facilitando así cálculos de indicadores estadísticos y gráficos.
- ✓ Un **gráfico** es una forma visual de representar y resumir datos de la tabla. Puede ser de varios tipos: **gráfico circular**, **gráfico de barras**, **histograma**, **polígono de frecuencias**, **diagrama de caja y bigote**, etc.

### 1.1.1 Ejercicios resueltos, paso a paso:

#### Ejemplo 1.

1. Se obtuvo la siguiente información sobre la atención odontológica realizada a pacientes por alumnos de cuarto a sexto año de la U.A por mes, en la Facultad de dicha universidad, durante los meses de Abril a Diciembre 2010.

- a. **Identifique:** Elemento de la población, variable, tipo de variable, Tipo de escala
- b. **Complete** la tabla de frecuencias con su respectivo título. Anéxele un gráfico adecuado
- c. **Determine el porcentaje de pacientes** atendidos en el mes de Octubre?
- d. **Determine el número de pacientes** atendidos a lo menos en septiembre?
- e. **¿Qué porcentaje de pacientes** fue atendido a lo más en Agosto?

| Mes        | n <sub>i</sub> |
|------------|----------------|
| Abril      | 78             |
| Mayo       | 41             |
| Junio      | 78             |
| Julio      | 97             |
| Agosto     | 82             |
| Septiembre | 28             |
| Octubre    | 53             |
| Noviembre  | 43             |
| Diciembre  | 18             |

### Esquema de desarrollo:

**Resp.: Elemento de la población :** Paciente atendido por alumnos de cuarto a sexto año de la carrera de Odontología,

**Variable:** Mes en que se efectuó la atención.

**Tipo de variable:** Cualitativa.

**Tipo de escala:** Categórica ordinal.

a) **Resp.:**

**Tabla de frecuencias:**

| Distribución de pacientes de alumnos de 4º a 6º año de la carrera de Odontología, según mes de atención, U. de Antofagasta, 2010. |            |          |       |          |
|-----------------------------------------------------------------------------------------------------------------------------------|------------|----------|-------|----------|
| Mes                                                                                                                               | $n_i$      | $f_i$    | $N_i$ | $F_i$    |
| Abril                                                                                                                             | 78         | 0,150579 | 78    | 0,150579 |
| Mayo                                                                                                                              | 41         | 0,079150 | 119   | 0,229729 |
| Junio                                                                                                                             | 78         | 0,150579 | 197   | 0,380308 |
| Julio                                                                                                                             | 97         | 0,187258 | 294   | 0,567567 |
| Agosto                                                                                                                            | 82         | 0,158301 | 376   | 0,725868 |
| Septiembre                                                                                                                        | 28         | 0,054054 | 404   | 0,779922 |
| Octubre                                                                                                                           | 53         | 0,102316 | 457   | 0,882239 |
| Noviembre                                                                                                                         | 43         | 0,083011 | 500   | 0,965250 |
| Diciembre                                                                                                                         | 18         | 0,034749 | 518   | 1        |
| <b>Total de</b>                                                                                                                   | <b>518</b> |          |       |          |

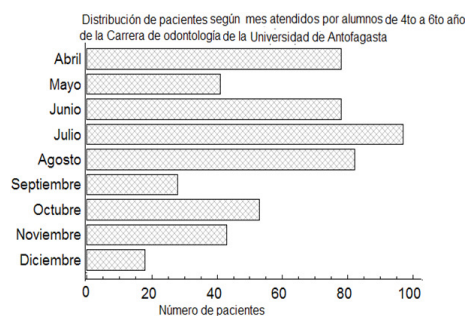
**Recuerda el siguiente plan cuando procedes a solucionar estos problemas:**

**Paso 1.** Identificar los datos que entrega el problema.

**Paso 2.** Identificar las preguntas del problema.

**Paso 3.** Investigar conocimientos adicionales.

**Paso 4.** Proponer y realizar las respuestas a las preguntas.



Presentación de la gráfica de los datos de la tabla.

### Ejemplo 2.

En un hospital de la capital, se informó que de un total de 350 enfermos, su diagnóstico fue: 60 con Cirrosis Hepática, 42 con Úlcera Gástrica, 180 con Úlcera duodenal, 21 con Gastritis erosiva, 16 con Esofagitis erosiva, 11 Yeyunitis hemorrágicas, no precisado 20. Determine:

- Elemento de la población:
- Variable.
- Tipo de variable.
- Tipo de escala.
- Construya una tabla de frecuencias con su título correspondiente.
- Presente estos datos gráficamente.

Esquema de desarrollo.

**Paso 1.** Aunque parezca una insistencia inadecuada, lea el problema detenidamente hasta comprenderlo.

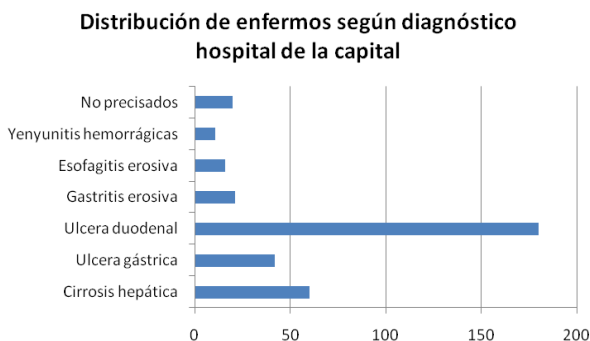
**Paso 2.** Desarrollar las respuestas:

- Resp. **Elemento de la población:** Un enfermo del hospital.
- Resp. **Variable:** Diagnóstico
- Resp. **Tipo de variable:** Cualitativa
- Resp. **Tipo de escala:** Nominal
- Resp. Tabla de frecuencias con un título adecuado:

| Distribución de enfermos según diagnóstico en un hospital de la capital |       |            |
|-------------------------------------------------------------------------|-------|------------|
| Diagnóstico                                                             | $n_i$ | $f_i$      |
| Cirrosis hepática                                                       | 60    | 0,17142857 |
| Úlcera gástrica                                                         | 42    | 0,12       |
| Úlcera duodenal                                                         | 180   | 0,51428571 |
| Gastritis erosiva                                                       | 21    | 0,06       |
| Esofagitis erosiva                                                      | 16    | 0,04571429 |
| Yenunitis hemorrágicas                                                  | 11    | 0,03142857 |
| No precisados                                                           | 20    | 0,05714286 |
| <b>Total de datos:</b>                                                  | 350   |            |

Obsérvese que por la poca información acerca de cuál es la capital a la que hacen referencia estos datos, el mes y/o año en que fueron tomados, la fuente que tomó estos datos y algunos otros indicadores de interés, no se ha podido detallar más esta experiencia estadística. Estos defectos deben corregirse.

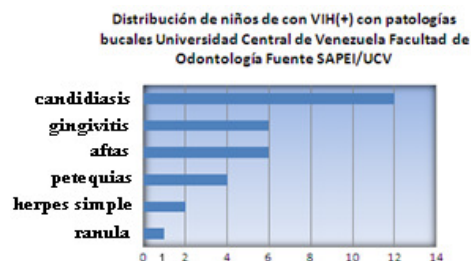
- Resp. Presentación de estos datos gráficamente:



### Ejemplo 3.

Dada la gráfica de la figura,

- a) Determine:
- El elemento de la población.
  - La variable en estudio.
  - El tipo de variable en estudio.
  - El tipo de escala.



- b) Construya una tabla de frecuencias.

Esquema de desarrollo:

**Paso 1.** Una vez más se sugiere que siempre comience por efectuar una lectura detenida del problema para lograr su total comprensión ANTES de comenzar a resolver.

**Paso 2.** Desarrollar las respuestas:

- Resp. **Elemento de la población:** Un niño con VIH(+).
- Resp. **Variable:** Patologías bucales.
- Resp. **Tipo de variable:** Cualitativa.
- Resp. **Tipo de escala:** Nominal.

| Distribución de niños con VIH(+) con patologías bucales, U.Central Venezuela Facultad de Odontología. Fuente: (Sapei/Ucv) |       |            |
|---------------------------------------------------------------------------------------------------------------------------|-------|------------|
| PATOLOGÍAS BUCALES                                                                                                        | $n_i$ | $f_i$      |
| Candiasis                                                                                                                 | 12    | 0,38709677 |
| Gingivitis                                                                                                                | 6     | 0,19354839 |
| Aftas                                                                                                                     | 6     | 0,19354839 |
| Petequias                                                                                                                 | 4     | 0,12903226 |
| herpes simples                                                                                                            | 2     | 0,06451613 |
| Ranulas                                                                                                                   | 1     | 0,03225806 |
| Total                                                                                                                     | 31    |            |

Resp. **Tabla de frecuencias:**

Verifique cada valor de la tabla adyacente, propuesta como respuesta.

### Ejemplo 4.

La tabla adyacente presenta el nivel de colesterol total de 120 trabajadores de la Empresa Minera Cerro Colorado en Agosto de 2005 ubicada a 5000 metros sobre el nivel del mar.

- Escriba un título a la tabla adjunta y construya un gráfico adecuado para presentar estos datos.
- Calcule e interprete la moda y el primer cuartil del colesterol total.
- Aquellos trabajadores con nivel de colesterol total igual o superior a 200 se consideran con colesterol alterado. ¿Qué porcentaje de trabajadores tienen colesterol alterado?
- Si el 25% de los trabajadores con más bajo colesterol total y el 30% de los trabajadores con más alto colesterol total son considerado de alto riesgo. Indique entre qué valores de colesterol total un trabajador no será considerado de alto riesgo.

| Colesterol Total |       | $n_i$ |
|------------------|-------|-------|
| LI               | LS    |       |
| 107,5            | 134,5 | 1     |
| 134,5            | 161,5 | 8     |
| 161,5            | 188,5 | 22    |
| 188,5            | 215,5 | 29    |
| 215,5            | 242,5 | 22    |
| 242,5            | 269,5 | 22    |
| 269,5            | 296,5 | 12    |
| 296,5            | 323,5 | 3     |
| 323,5            | 350,5 | 1     |
| Total de datos:  |       | 120   |

## Esquema de desarrollo.

**Paso 1.** Leer el problema, **identificar en pocas palabras** todo aquello que se pide hacer. Escribir frases cortas. Tener claridad de que al final **se debe terminar con una respuesta a lo que se pide**.

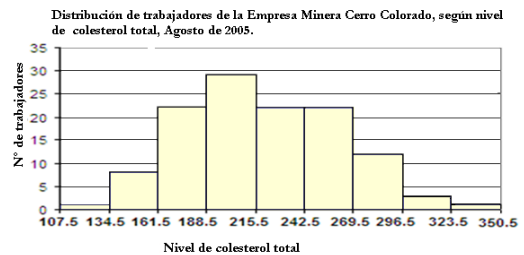
**Paso 2.** Desarrollar las respuestas:

- a) **Se pide:** Escribir un título a la tabla adjunta y construir un gráfico adecuado a estos datos.

Cuando se escriben títulos de tabla es costumbre comenzar con la expresión “**Distribución de**” y a continuación se identifica el tipo de **elementos de la población** que han sido contados como pertenecientes a cada clase o intervalo de la tabla, en la columna de frecuencias absolutas  $n_i$ . Luego se escribe la frase “**según**” y se agrega la **variable en estudio** (colesterol total en este caso).

| Distribución de trabajadores de la Empresa Minera Cerro Colorado, según Nivel de Colesterol (5000 mts s.n.m.), Agosto 2005 |       |     |     |          |     |          |
|----------------------------------------------------------------------------------------------------------------------------|-------|-----|-----|----------|-----|----------|
| L.I.                                                                                                                       | L.S.  | ci  | ni  | fi       | NI  | Fi       |
| 107,5                                                                                                                      | 134,5 | 121 | 1   | 0,008333 | 1   | 0,008333 |
| 134,5                                                                                                                      | 161,5 | 148 | 8   | 0,066667 | 9   | 0,075    |
| 161,5                                                                                                                      | 188,5 | 175 | 22  | 0,183333 | 31  | 0,258333 |
| 188,5                                                                                                                      | 215,5 | 202 | 29  | 0,241667 | 60  | 0,5      |
| 215,5                                                                                                                      | 242,5 | 229 | 22  | 0,183333 | 82  | 0,683333 |
| 242,5                                                                                                                      | 269,5 | 256 | 22  | 0,183333 | 104 | 0,866667 |
| 269,5                                                                                                                      | 296,5 | 283 | 12  | 0,1      | 116 | 0,966667 |
| 296,5                                                                                                                      | 323,5 | 310 | 3   | 0,025    | 119 | 0,991667 |
| 323,5                                                                                                                      | 350,5 | 337 | 1   | 0,008333 | 120 | 1        |
|                                                                                                                            |       |     | 120 |          |     |          |

Seguidamente se deja expresado en **dónde se recopilaron los datos** (Empresa Minera Cerro Colorado), el **año** y el **mes**. También es aconsejable dejar constancia del lugar geográfico (región, país: Antofagasta, Chile, por ejemplo) y cualquier otro dato que permita diferenciar los datos de una fecha y lugar con los de otra investigación distinta.



El gráfico, que no es sino una expresión visual de lo que contiene la tabla, queda como se muestra en la figura adyacente.

- b) **Se pide:** Calcular e interpretar la moda y el primer cuartil del colesterol total.

- a. La moda es el valor de mayor frecuencia absoluta. La mayor frecuencia absoluta en la tabla es  $n_4 = 29$ . Por lo tanto la moda debe ser un valor que pertenece al intervalo  $(188,5; 215,5]$ . Puesto que esta variable está definida por intervalos, es variable continua. Esto quiere decir que la moda será calculada, para  $i = 4$ , con

$$M_o = LI_{(i)} + A_i \cdot \frac{(n_i - n_{i-1})}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

$$\stackrel{i=4}{=} LI_{(4)} + A_4 \cdot \frac{(n_4 - n_{4-1})}{(n_4 - n_{4-1}) + (n_4 - n_{4+1})}$$

$$\stackrel{i=4}{=} LI_{(4)} + A_4 \cdot \frac{(n_4 - n_3)}{(n_4 - n_3) + (n_4 - n_3)}$$

Reemplazando los valores de la tabla se obtiene

$$M_o = 188,5 + 27 \cdot \frac{(29 - 22)}{(29 - 22) + (29 - 22)} = 202$$

**Resp.:** El **colesterol total más frecuente** en esta muestra de 120 trabajadores es de 202 unidades (Esto expresa el cálculo y la interpretación pedidos por el problema).

- b. El primer cuartil es el percentil 25,  $P_{25}$ , calculado como sigue

$$Q_1 = P_{25} = LI_{(k)} + A_k \frac{\left( \frac{n \cdot 25}{100} - N_{k-1} \right)}{n_k}$$

Debemos determinar el orden del intervalo que contiene al cuartil  $Q_1$ . Esto se hace con la expresión  $\alpha = \frac{n \cdot 25}{100}$ , que representa la cuarta parte del total de datos  $n = 120$ . Así,  $\alpha = \frac{120 \cdot 25}{100} = 30$ .

En la columna de las frecuencias absolutas acumuladas  $N_i$ , cuyos valores son ascendentes por representar acumulación de observaciones desde la de más bajo valor, buscamos el primer  $N_i$  que iguale o supere a  $\alpha = 30$ . Éste resulta ser  $N_3 = 31$ . De aquí se deduce que el valor de  $k$  que identifica al intervalo que contiene al primer cuartil es  $k = 4$ . Por lo tanto, la expresión con la que calcularemos el valor del primer cuartil es

$$Q_1 = P_{25} = LI_{(3)} + A_3 \frac{\left( \frac{120 \cdot 25}{100} - N_{3-1} \right)}{n_3} = LI_{(3)} + A_3 \frac{(30 - N_2)}{n_3}$$

Buscando los valores correspondientes en la tabla obtenemos

$$P_{25} = 161.5 + 27 \cdot \frac{(30 - 9)}{22} = 187.27$$

**Interpretación:** Cuando los datos han sido ordenados de menor a mayor, el primer cuartil es la frontera que deja bajo sí a lo más al 25% del total de datos (este es el valor de  $\alpha$ ) y a lo más sobre él al 75% del total de datos.

**Cálculo:** El valor del primer cuartil es de 187,27 unidades de colesterol total.

**Resp.** (Cálculo e interpretación):

El valor del Primer Cuartil o Cuartil Inferior es de 187,27 unidades de colesterol total y deja bajo sí a lo más al 25% de los valores de colesterol total más bajo, y sobre este valor se encuentra a lo más el 75% de los valores de colesterol total más alto.

c) **Se pide:** El porcentaje de trabajadores tienen colesterol **alterado**.

Los trabajadores tienen colesterol alterado cuando su nivel de colesterol total es igual o superior a 200 unidades.

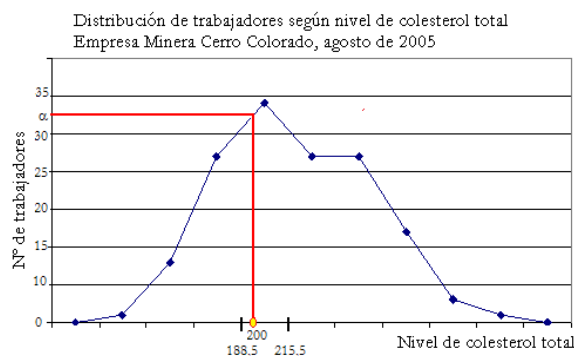
Decir “200 unidades de colesterol total” significa fijar una **frontera** en la muestra de 120 observaciones para calcular qué porcentaje tiene **tanto como o más** de “200 unidades de colesterol total”. Es decir debemos calcular un percentil  $P_i$  de valor 200, desconociendo el valor de  $i$ , y calcular qué porcentaje del total de datos tiene valores menores que él.

Si primero nos abocamos a encontrar cuántos datos, a lo más, son menores que el percentil que buscamos, entonces nos estamos refiriendo al valor  $\alpha$  calculado antes. En esa oportunidad sabíamos que el percentil dejaba a lo más un 25% de los datos bajo sí y escribíamos  $\alpha = \frac{n \cdot 25}{100}$  que puede ser escrito como  $\alpha = \frac{n \cdot 25}{100} = n \cdot \frac{25}{100} = \frac{25}{100} \cdot n$ , que no es sino el 25% del total de datos  $n$  ... ¡¡¡Sorpresa!!!

Si generalizamos el razonamiento anterior, desconociendo el porcentaje, diremos que si  $\alpha = \frac{n \cdot i}{100} = \frac{i}{100} \cdot n$ , lo que se pide encontrar es el valor de  $i$ .

En resumen, sabemos que  $P_i = 200$  y debemos encontrar el valor de  $i$ . Usando un polígono de frecuencias, como el de la figura, podemos ilustrar lo que está sucediendo.

El percentil  $P_i = 200$  pertenece al intervalo  $(188.5; 215.5]$ . Por lo tanto podemos escribir



$$200 = \underset{\text{Lím.Inf.}}{188.5} + \underset{A_k}{27} \cdot \left( \frac{\frac{200}{100} N_{k-1}}{29} \right) \underset{\text{despejando}}{\Rightarrow} \alpha = 43.35 \Leftrightarrow \frac{120 \cdot i}{100} = 43.35 \Leftrightarrow i = 36.13$$

Puesto que este percentil deja bajo él a lo más entre 36 y 37 datos (“36.13” para ser más exactos) que corresponden al 36.13% del total de datos (120), y los trabajadores que tienen **colesterol alterado** tienen un colesterol total de 200 o más unidades, se deduce que el  $(100 - 36.13)\% = 63.87\%$  tiene colesterol alterado.

**Resp. :** El 63.87% de los trabajadores de la Empresa Minera Cerro Colorado tiene colesterol **alterado**.

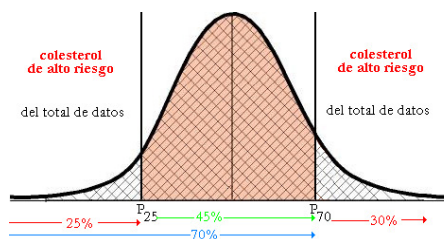
- d) **Se pide:** Entre qué valores de colesterol total un trabajador **no será** considerado de alto riesgo.

El enunciado del problema dice:

“... el 25% de los trabajadores con más bajo colesterol total y el 30% de los trabajadores con más alto colesterol total son considerado de **alto riesgo**...”

Por lo tanto, suavizando el histograma de los datos de este problema, podemos obtener una imagen como la de la figura. En ella podemos dejar fronteras para identificar las zonas de valores de colesterol total de **alto riesgo**.

Por la definición de percentil, es claro que la frontera del superior del 25% más bajo de valores de colesterol total es el percentil 25,  $P_{25}$ . Si por sobre la frontera inferior del 30% de valores más altos de colesterol total están también los valores de colesterol de alto riesgo, entonces el colesterol de alto riesgo es de un 55% del total de valores posibles. En consecuencia el colesterol que **NO ES DE ALTO RIESGO** constituye el 45% del total de valores posibles de colesterol total. La frontera superior de este intervalo es el percentil  $P_{70}$ .



Las fronteras de este intervalo de colesterol total que **NO ES DE ALTO RIESGO** son los percentiles  $P_{25}$  y  $P_{70}$ . Calculamos estos percentiles (esta vez no daremos tanto detalle, debe revisarse la expresión de cálculo dada antes):

$$P_{25} = 161.5 + 27 \cdot \frac{(30 - 9)}{22} = 187.27$$

$$P_{70} = 242.5 + 27 \cdot \frac{(90 - 82)}{22} = 252.32$$

Estamos ahora en condiciones de responder a lo pedido:

**Resp.:** Un trabajador **no será** considerado con colesterol total de alto riesgo cuando sus valores de colesterol total se encuentren entre 187.27 y 252.32 unidades.

El intervalo de valores de colesterol total que no es de alto riesgo es  $[187.27; 252.32]$ .

### 1.1.2 Ejercicios Propuestos.

1. Un estudiante de enfermería hace una encuesta en una fábrica, a los 50 empleados que ahí laboran, acerca de del número de cargas familiares de cada uno de ellos; los resultados fueron los siguientes:

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 7 | 2 | 9 | 4 | 5 | 2 | 3 |
| 2 | 3 | 4 | 3 | 7 | 7 | 2 | 9 | 8 | 4 |
| 3 | 7 | 4 | 1 | 2 | 4 | 5 | 5 | 1 | 2 |
| 3 | 3 | 4 | 5 | 7 | 9 | 8 | 7 | 5 | 3 |
| 2 | 2 | 1 | 3 | 5 | 6 | 6 | 2 | 1 | 3 |

- Identificar Población analizada.
  - Elemento de la población.
  - Característica en estudio.
  - Tipo de dato analizado.
  - Construya una tabla de frecuencia completa en clases de longitud 2 anexándole título.
  - Construir un gráfico adecuado.
  - ¿Cuántos empleados tienen 1 ó 2 cargas? ¿Qué porcentaje representa?
  - ¿Cuántos empleados tienen a lo más 6 cargas? ¿Qué porcentaje representa?
  - ¿cuántos empleados tienen más de 4 cargas?
2. Para un estudio de una dieta de escaso contenido proteico se tiene un informe de la cantidad de nitrógeno (en gramos) en 30 muestras de orina de 24 horas, en el consultorio sur de la ciudad de Antofagasta en junio de 2010, con los siguientes resultados

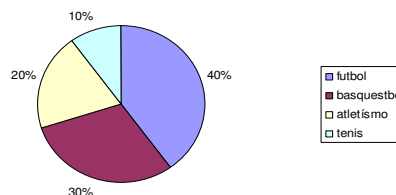
|     |     |     |     |     |     |     |     |     |      |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 2,8 | 3,2 | 3,4 | 3,8 | 4,3 | 5,2 | 7,5 | 8,4 | 8,9 | 9,5  |
| 3,0 | 3,3 | 3,5 | 4,1 | 4,7 | 6,3 | 7,6 | 8,4 | 9,2 | 11,0 |
| 3,0 | 3,4 | 3,7 | 4,2 | 5,0 | 7,0 | 7,8 | 8,5 | 9,4 | 12,4 |

- Identifique: Población, variable en estudio, tipo de variable, tipo de escala
- Construya una tabla de frecuencia completa de 8 categorías, escribiéndole un título para los datos entregados
- ¿Cuántas muestras tienen una cantidad de nitrógeno menor o igual a 8,8 gramos?
- ¿Cuántas muestras tienen una cantidad de nitrógeno mayor a 5.2 gramos?
- ¿Cuántas muestras tienen
- una cantidad de nitrógeno mayor a 5.2 gramos y menor o igual a 8.8 gramos?
- ¿Qué porcentaje de muestra tienen una cantidad de nitrógeno a lo más 10 gramos?
- ¿Qué porcentaje de muestras tiene una cantidad de nitrógeno mayor a 6.4 gramos?
- Calcular e interpretar el percentil 30
- Construya un gráfico adecuado para presentar los datos de la tabla construida en b) e interprételo.

3. Dado el gráfico circular

- ¿Cuál es la variable en estudio?
- ¿Qué tipo de variable es?
- ¿Cuál es la frecuencia absoluta del grupo de fútbol?
- ¿Cuántas personas no prefieren el tenis?

Preferencias de 30 personas en actividades deportivas



## 1.2. Medidas de posición y dispersión. Definiciones y propiedades principales.

- ✓ Llamaremos **medida de posición** a aquel valor que resume alguna cualidad de la variable en términos de su posición en el dominio. Tal valor sirve como representante del total de observaciones respecto de esa cualidad.
- ✓ Las medidas de posición usuales son: **media aritmética, moda y percentil**.
- ✓ Se llama **media aritmética** a la medida de posición que se obtiene sumando todos los valores de la variable en una población o en una muestra, dividida por el número total de datos considerados.
- ✓ En **datos no tabulados** la media aritmética,  $\bar{x}$ , de los valores  $x_1, x_2, x_3, \dots, x_n$  se calcula como

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- ✓ En **datos tabulados**, la media aritmética de los  $n$  observaciones  $x_1, x_2, x_3, \dots, x_n$  agrupados en  $m$  categorías o intervalos con frecuencia absoluta  $n_1, n_2, n_3, \dots, n_m$ , y clases o marcas de clase  $c_1, c_2, c_3, \dots, c_m$  está dada por

$$\bar{x} = \frac{1}{n} \sum_{i=1}^m c_i n_i$$

- ✓ Se llama **media aritmética total** de las  $k$  subpoblaciones de datos  $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \dots, x_{n_1}^{(1)}; x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, \dots, x_{n_2}^{(2)}; \dots; x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_{n_k}^{(k)}$  a

$$\bar{x}_T = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \dots + n_k \bar{x}_k}{n} = \frac{\sum_{j=1}^k n_j \bar{x}_j}{n}$$

$$\text{donde, } \bar{x}_j = \frac{\sum_{i=1}^{n_j} x_i^{(j)}}{n_j}, \quad j=1,2,\dots,k.$$

- ✓ Se llama **moda** al valor de la variable en el que se produce un máximo de la distribución de frecuencias. Es llamado también dato más frecuente, valor más frecuente o valor típico.
- ✓ Existen distribuciones con más de una moda, éstas son llamadas **multimodales** o **plurimodales**.
- ✓ En **datos no tabulados** la moda es el dato más observado.
- ✓ En **datos tabulados de variable categórica** o **numérica discreta** la moda es la clase de mayor frecuencia observada.
- ✓ En **datos tabulados de variable continua** la moda se calcula por la expresión

$$M_o = LI_{(i)} + A_i \cdot \frac{(n_i - n_{i-1})}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

Aquí  $n_i$  es la **mayor frecuencia absoluta**, correspondiente al intervalo de orden  $i$  que contiene a la moda;  $n_{i-1}$  y  $n_{i+1}$  son las frecuencias absolutas **anterior** y **posterior**, respectivamente;  $A_i$  es la **amplitud** del intervalo que contiene a la moda (el de orden  $i$ ) y  $LI_{(i)}$  es el **límite inferior** del intervalo que contiene a la moda.

- ✓ Se llama **percentil** al valor frontera que particiona el ordenamiento de las observaciones en a lo más un  $i$  % bajo él y en un  $(100-i)$  % por sobre él.
- ✓ En **datos no tabulados**  $x_1, x_2, \dots, x_n$ , el percentil  $P_i$  se calcula como una interpolación lineal de la siguiente forma:

$$P_i = \left( (k+1) - \frac{(n+1)i}{100} \right) x_{(k)} + \left( \frac{(n+1)i}{100} - k \right) x_{(k+1)}$$

donde  $\frac{(n+1)i}{100}$  es la posición que ocupa el percentil  $P_i$  entre los datos ordenados  $x_{[1]}, x_{[2]}, \dots, x_{[n]}$  en forma no decreciente, de forma que  $x_{[1]}, x_{[2]}, \dots, x_{[k]}, P_i, x_{[k+1]}, \dots, x_{[n]}$ .

- ✓ En **datos tabulados** el percentil  $P_i$  se calcula con la expresión

$$P_i = LI_{(k)} + A_k \frac{\left( \frac{n \cdot i}{100} - N_{k-1} \right)}{n_k}$$

donde  $n$  es el número total de observaciones;  $\frac{n \cdot i}{100}$  es el valor que permite identificar al intervalo de orden  $k$  en el cual se encuentra el percentil y hasta el cual van acumulados  $\frac{n \cdot i}{100}$  datos;  $LI_{(k)}$  es el límite inferior del intervalo  $k$ ;  $A_k$  es la amplitud del intervalo  $k$ ;  $n_k$  es la frecuencia absoluta del intervalo de orden  $k$  y  $N_{k-1}$  es la frecuencia absoluta acumulada anterior a la del intervalo de orden  $k$ .

✓ **Percentiles importantes:**

- **Mediana:** Es el valor de la variable que particiona la distribución de datos en dos partes dejando en cada una el 50% del total de datos. Se denota por  $Me$ , donde  $Me = P_{50}$ .
- **Cuartiles:** Son valores de la variable que particionan la distribución de datos en cuatro partes dejando en cada una el 25% del total de datos. Se denotan por  $Q_i$ ,  $i=1,2,3$ . Son cuartiles:  $Q_1 = P_{25}$  (lower quartil, cuartil inferior o primer cuartil),  $Q_2 = P_{50}$  (segundo cuartil o mediana),  $Q_3 = P_{75}$  (uper quartil, cuartil superior o tercer cuartil).
- **Quintiles:** Son valores de la variable que particionan la distribución de datos en cinco partes dejando en cada una el 20% del total de datos. Son quintiles:  $P_{20}$  (primer quintil),  $P_{40}$  (segundo quintil),  $P_{60}$  (tercer quintil),  $P_{80}$  (cuarto quintil).
- **Deciles:** Son valores de la variable que particionan la distribución de datos en diez partes dejando en cada una el 10% del total de datos. Se denotan por  $D_i$ ,  $i=1,2,...,9$ . Son deciles  $D_1 = P_{10}$ ,  $D_2 = P_{20}$ , ...,  $D_9 = P_{90}$ .

- ✓ Se llama **medida de dispersión** a aquella medida relaciona la mayor o menor concentración de los datos entorno a un valor central, generalmente la media aritmética
- ✓ Son medidas de dispersión la varianza y la desviación estándar. También el coeficiente de variación (relativa a la media, por “unidades de media”).
- ✓ Se llama **varianza** a la media aritmética de los cuadrados de las diferencias entre los valores de la distribución y su media aritmética. Los valores de la variable pueden estar representados por las observaciones originales (en caso de datos sin tabular) o por sus clases (en caso de variable discreta) o por sus marcas de clase (en caso de variable continua).
- ✓ La **varianza en datos no tabulados** se calcula mediante la expresión

$$V(X) \stackrel{\text{Notación}}{=} S_X^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - (\bar{x})^2$$

- ✓ La **varianza en datos tabulados** se calcula por la expresión

$$S_X^2 = \frac{\sum_{i=1}^m n_i (c_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^m c_i^2 n_i}{n} - (\bar{x})^2$$

donde  $n_i$  es la frecuencia absoluta,  $c_i$  es la clase o marca de clase del

intervalo  $i=1,2,\dots,m$ ,  $m$  es el número de clases o intervalos y  $n$  es el número total de observaciones.

- ✓ Las unidades de la varianza son el cuadrado de las unidades de la variable original. Esto es,  $mts^2$ ,  $cm^2$ ,  $litros^2$ ,  $dólares^2$ ,  $pesos^2$ ,  $Kg^2$ , etc.
- ✓ Se llama **desviación estándar** a la raíz cuadrada positiva de la varianza
- ✓ Las unidades de la desviación estándar son las mismas unidades de la variable original. Esto es,  $mts$ ,  $cms$ ,  $litros$ ,  $dólares$ ,  $pesos$ ,  $Kgs$ , etc.
- ✓ En **datos no tabulados** la desviación estándar se calcula por la expresión

$$S_X = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - (\bar{x})^2}$$

- ✓ En **datos tabulados** la desviación estándar se calcula por la expresión

$$S_X = \sqrt{\frac{\sum_{i=1}^m n_i (c_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^m c_i^2 n_i}{n} - (\bar{x})^2}$$

con las interpretación de símbolos tal como se ha hecho antes.

- ✓ Se llama **coeficiente de variación** al cociente entre la desviación estándar multiplicada por 100, y la media aritmética. Este coeficiente se mide en porcentaje.
- ✓ El **coeficiente de variación** se calcula con la expresión

$$CV_X = \frac{S_X}{\bar{x}} \cdot 100\%$$

- ✓ La **varianza total** de las  $k$  subpoblaciones de datos

$$x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, \dots, x_{n_1}^{(1)}; x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, \dots, x_{n_2}^{(2)}; \dots; x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, \dots, x_{n_k}^{(k)}$$

se calcula por la expresión

$$S_T^2 = \frac{\sum_{i=1}^k n_i [S_i^2 + (\bar{x}_i)^2]}{n} - \left( \frac{\sum_{i=1}^k n_i \bar{x}_i}{n} \right)^2$$

### 1.2.1 Ejercicios resueltos, paso a paso:

#### Ejemplo 1.

La siguiente tabla (prescindiremos por ahora de su título para centrarnos más en los procedimientos) corresponde al porcentaje de cobre obtenido mediante dos procesos de producción, en diferentes muestras de mineral; los resultados se muestran en sus respectivas tablas de frecuencias

- a) Determine e interprete la moda y el cuartil 3 del porcentaje de cobre en ambos procesos.

- b) Si el 20% de las muestras con porcentaje de cobre bajo es considerado como de baja ley, el 10% de las muestras con mayor porcentaje de cobre es considerado como de buena ley y las demás muestras son consideradas como de ley media.

Para ambos procesos determine:

- ¿Cuál es el porcentaje de cobre mínimo para que una muestra sea considerada de buena ley?
- ¿Entre qué valores una muestra será considerada como ley media?
- ¿Cuál es el número estimado de muestras con ley media?

| Porcentaje de cobre obtenido mediante dos procesos de producción |               |                |           |               |                |
|------------------------------------------------------------------|---------------|----------------|-----------|---------------|----------------|
| Proceso 1                                                        |               |                | Proceso 2 |               |                |
| i                                                                | Intervalos    | n <sub>i</sub> | i         | Intervalos    | n <sub>i</sub> |
| 1                                                                | 0.041 – 0.047 | 5              | 1         | 0.059 – 0.063 | 7              |
| 2                                                                | 0.047 – 0.053 | 8              | 2         | 0.063 – 0.067 | 11             |
| 3                                                                | 0.053 – 0.059 | 15             | 3         | 0.067 – 0.071 | 10             |
| 4                                                                | 0.059 – 0.065 | 3              | 4         | 0.071 – 0.075 | 7              |
| 5                                                                | 0.065 – 0.071 | 4              | 5         | 0.075 – 0.079 | 2              |
| 6                                                                | 0.071 – 0.077 | 1              | 6         | 0.079 – 0.083 | 1              |

- Determine la media aritmética y la mediana del porcentaje de cobre en el proceso 1 y compárelas con la moda. ¿Cuál es su opinión respecto a la simetría de la distribución del porcentaje de cobre en el proceso 1? Se deja como inquietud el mismo análisis para el proceso 2
- Si consideramos ambos procesos como uno solo. ¿Cuál será el porcentaje de cobre promedio total?
- Determine un intervalo que contenga el 50% central de la distribución del porcentaje de cobre de las muestras usando el proceso 2.
- Determine un intervalo centrado en la media aritmética y de amplitud 2 desviaciones estándar. ¿Qué porcentaje de observaciones tiene valores de ley en este intervalo?

## Esquema de desarrollo.

**Paso 1.** Lea con detenimiento el problema analizando cada uno de los datos informativos que da. Analice qué se requiere de usted en cada pregunta. Cuando tenga claro todo lo que se pide y lo que se da como información, analice y provéase de otros elementos o conocimientos que necesita para resolver estos planteamientos problemáticos. Diseñe planes de solución para cada caso. Finalmente proceda a ejecutar la solución analizando y fundamentando sus procedimientos, a la par de interpretar la aceptación de los resultados que usted está obteniendo.

**Paso 2.** Preocúpese de identificar claramente poblaciones, elementos de estudio en esas poblaciones, variables, tipos de variable, estadísticos de posición y de dispersión, en fin, todo aquello que le permita entender a cabalidad la situación problemática tanto como lo que se le está pidiendo que resuelva.

**Paso 3.** Desarrollo de las respuestas:

- a) **Se pide:** Determinar e interpretar la **moda** y el **cuartil 3** del porcentaje de cobre en ambos procesos.

Para ello echamos mano de una tabla de cálculos como la que se presenta en la figura.

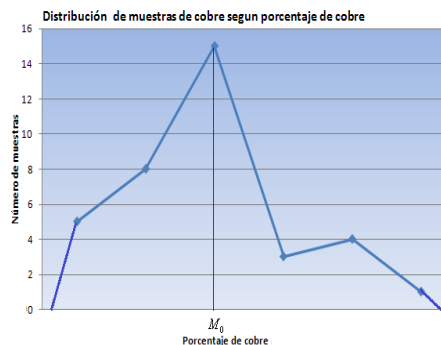
En base a ella podemos calcular la moda del porcentaje de cobre logrado por el proceso 1:

| Proceso 1 |       |       |       |                 |       |        |        |
|-----------|-------|-------|-------|-----------------|-------|--------|--------|
| LI        | LS    | $c_i$ | $n_i$ | $n_i \cdot c_i$ | $N_i$ | $f_i$  | $F_i$  |
| 0,041     | 0,047 | 0,044 | 5     | 0,22            | 5     | 0,1389 | 0,1389 |
| 0,047     | 0,053 | 0,05  | 8     | 0,4             | 13    | 0,2222 | 0,3611 |
| 0,053     | 0,059 | 0,056 | 15    | 0,84            | 28    | 0,4167 | 0,7778 |
| 0,059     | 0,065 | 0,062 | 3     | 0,186           | 31    | 0,0833 | 0,8611 |
| 0,065     | 0,071 | 0,068 | 4     | 0,272           | 35    | 0,1111 | 0,9722 |
| 0,071     | 0,077 | 0,074 | 1     | 0,074           | 36    | 0,0278 | 1      |
| Totales:  |       |       | 36    | 1,992           |       |        |        |

$$M_o = LI_{(i)} + A_i \cdot \frac{(n_i - n_{i-1})}{(n_i - n_{i-1}) + (n_i - n_{i+1})} = 0,053 + 0,006 \cdot \frac{(15 - 8)}{(15 - 8) + (15 - 3)} = 0,055$$

**Resp.:** El porcentaje de cobre más frecuente obtenido por el Proceso 1 es 0.055%.

Una visualización gráfica de lo que hemos calculado se presenta en la siguiente gráfica. En ella se ha utilizado un polígono de frecuencias:



El cuartil 3 del porcentaje de cobre obtenido por el proceso 1 equivale al percentil 75 de esta variable. Por lo tanto desde la misma tabla se obtiene:

$$P_{75} = LI_{(3)} + A_3 \cdot \frac{(\frac{n \cdot 75}{100} - N_2)}{n_3} \stackrel{n=36}{=} 0.053 + 0.006 \cdot \frac{(27-13)}{15} = 0.059$$

**Resp.:** El cuartil 3 del porcentaje de cobre obtenido por el Proceso 1 es 0.059%. Está ligeramente por debajo del punto medio del rango de variación observado. Esto significa que hay una ligera mayor agrupación de porcentajes de cobre obtenidos por el proceso 1 en la mitad más baja de porcentajes del Proceso 1.

A lo más el 75% de las muestras obtenidas por el Proceso 1 tiene porcentaje de cobre menor o igual a 0.059 y por sobre este valor a lo más el 25% de las muestras obtenidas por el Proceso 1 tiene porcentaje de cobre mayor que 0.059

Si se analiza ahora, el Proceso 2 de la misma manera se obtiene:

$$M_o = LI_{(2)} + A_i \cdot \frac{(n_2 - n_1)}{(n_2 - n_1) + (n_2 - n_3)}$$

$$= 0.063 + 0.004 \cdot \frac{(11-7)}{(11-7) + (11-10)}$$

$$= 0.0662$$

| Proceso 2 |       |       |                |                |                                |                |
|-----------|-------|-------|----------------|----------------|--------------------------------|----------------|
| i         | LI    | LS    | c <sub>i</sub> | n <sub>i</sub> | n <sub>i</sub> ·c <sub>i</sub> | N <sub>i</sub> |
| 1         | 0,059 | 0,063 | 0,061          | 7              | 0,427                          | 7              |
| 2         | 0,063 | 0,067 | 0,065          | 11             | 0,715                          | 18             |
| 3         | 0,067 | 0,071 | 0,069          | 10             | 0,69                           | 28             |
| 4         | 0,071 | 0,075 | 0,073          | 7              | 0,511                          | 35             |
| 5         | 0,075 | 0,079 | 0,077          | 2              | 0,154                          | 37             |
| 6         | 0,079 | 0,083 | 0,081          | 1              | 0,081                          | 38             |
| Totales:  |       |       |                | 38             | 2,578                          |                |

**Resp.:** El porcentaje de cobre más frecuentemente obtenido por el Proceso 2 es 0.0662% (numéricamente mayor que el obtenido por el Proceso 1 ¿es esto significativo?)

El cuartil 3 del porcentaje de cobre obtenido por el proceso 2 se logra también desde la tabla:

$$P_{75} = LI_{(4)} + A_4 \cdot \frac{(\frac{n \cdot 75}{100} - N_3)}{n_4} \stackrel{n=38}{=} 0.071 + 0.004 \cdot \frac{(28.5-28)}{7} = 0.074$$

**Resp.:** El cuartil 3 del porcentaje de cobre obtenido por el Proceso 2 es 0.074%. Está ligeramente por sobre el punto medio del rango de variación observado. Esto era de esperarse si la distribución de valores se piensa simétricamente distribuida en el rango de variación observado. Comparado con la obtención de cobre lograda por el Proceso 1 se advierte que este proceso 2 obtiene mejores porcentajes de cobre (¿podría esto afirmarse con autoridad?).

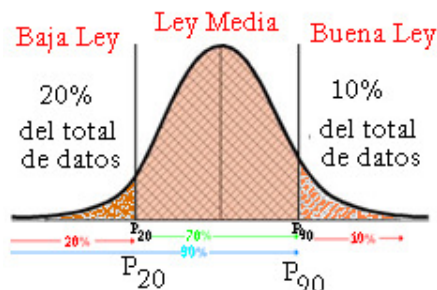
A lo más el 75% de las muestras obtenidas por el Proceso 2 tiene porcentaje de cobre menor o igual a 0.074 y por sobre este valor a lo más el 25% de las muestras obtenidas por el Proceso 2 tiene porcentaje de cobre mayor que 0.074.

b) **Se define:**

- Muestras de **baja ley**: 20% de muestras con más bajos porcentajes de cobre.
- Muestras de **buena ley**: 10% de muestras con más altos porcentajes de cobre.
- Muestras de **ley media**: las restantes muestras.

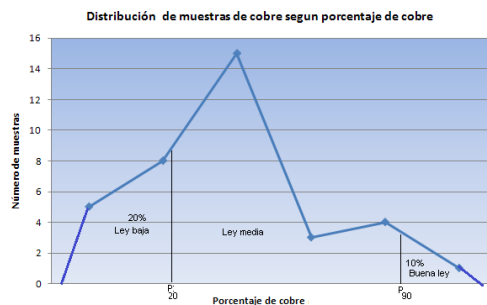
- a. Se pide: **Porcentaje de cobre mínimo** para que una muestra sea considerada de **buena ley**.

El percentil  $P_{20}$  es la frontera superior del intervalo que contiene al 20% de más bajo porcentaje de cobre o cobre de **baja ley**. Se puede deducir que si  $P_{90}$  es la frontera inferior de los porcentajes de ley más altos, entonces el intervalo que define un porcentaje de ley media es  $[P_{20}; P_{90}]$ .



Con una polígono de frecuencias, como antes, se puede visualizar también esta situación (ver figura).

Debemos calcular, entonces  $P_{20}$ , que es el “porcentaje de cobre mínimo para que una muestra sea considerada de buena ley”. Por lo tanto:

**EN EL PROCESO 1:**

Observando la tabla de la derecha se puede deducir que en este proceso

$$P_{20} = 0.047 + 0.006 \cdot \frac{(7.2 - 5)}{8}$$

$$= 0.0487$$

| Porcentaje de cobre obtenido mediante dos procesos de producción |               |       |           |               |       |
|------------------------------------------------------------------|---------------|-------|-----------|---------------|-------|
| Proceso 1                                                        |               |       | Proceso 2 |               |       |
| i                                                                | Intervalos    | $n_i$ | i         | Intervalos    | $n_i$ |
| 1                                                                | 0.041 – 0.047 | 5     | 1         | 0.059 – 0.063 | 7     |
| 2                                                                | 0.047 – 0.053 | 8     | 2         | 0.063 – 0.067 | 11    |
| 3                                                                | 0.053 – 0.059 | 15    | 3         | 0.067 – 0.071 | 10    |
| 4                                                                | 0.059 – 0.065 | 3     | 4         | 0.071 – 0.075 | 7     |
| 5                                                                | 0.065 – 0.071 | 4     | 5         | 0.075 – 0.079 | 2     |
| 6                                                                | 0.071 – 0.077 | 1     | 6         | 0.079 – 0.083 | 1     |

**Resp.:** En el Proceso 1, el porcentaje de cobre mínimo para que una muestra sea considerada de buena ley es 0.0487%.

**EN EL PROCESO 2:**

Se puede deducir que el límite inferior  $P_{20}$  será

$$P_{20} = 0.063 + 0.004 \cdot \frac{(7.6 - 7)}{11} = 0.0632$$

**Resp.:** En el Proceso 2, el porcentaje de cobre mínimo para que una muestra sea considerada de buena ley es 0.0632%.

Con estos porcentajes se puede observar que lo que se llama buena ley en el proceso 2 tiene frontera mínima mayor que en el Proceso 1.

- b. **Se pide:** Entre qué valores una muestra será considerada como de ley media.

Para hacer esta determinación en ambos procesos debemos calcular el percentil  $P_{90}$  ( $P_{20}$  ya está calculado).

#### EN EL PROCESO 1:

Observando la tabla de la derecha se puede deducir que en este proceso

$$P_{90} = 0.065 + 0.006 \cdot \frac{(32.4 - 31)}{4} = 0.0698$$

| Porcentaje de cobre obtenido mediante dos procesos de producción |               |       |           |               |       |
|------------------------------------------------------------------|---------------|-------|-----------|---------------|-------|
| Proceso 1                                                        |               |       | Proceso 2 |               |       |
| i                                                                | Intervalos    | $n_i$ | i         | Intervalos    | $n_i$ |
| 1                                                                | 0.041 – 0.047 | 5     | 1         | 0.059 – 0.063 | 7     |
| 2                                                                | 0.047 – 0.053 | 8     | 2         | 0.063 – 0.067 | 11    |
| 3                                                                | 0.053 – 0.059 | 15    | 3         | 0.067 – 0.071 | 10    |
| 4                                                                | 0.059 – 0.065 | 3     | 4         | 0.071 – 0.075 | 7     |
| 5                                                                | 0.065 – 0.071 | 4     | 5         | 0.075 – 0.079 | 2     |
| 6                                                                | 0.071 – 0.077 | 1     | 6         | 0.079 – 0.083 | 1     |

En el Proceso 1, el porcentaje de cobre máximo para que una muestra sea considerada de buena ley es 0.0698%.

#### EN EL PROCESO 2:

Se puede deducir que el límite inferior  $P_{90}$  será

$$P_{90} = 0.071 + 0.004 \cdot \frac{(34.2 - 28)}{7} = 0.0745$$

En el Proceso 2, el porcentaje de cobre máximo para que una muestra sea considerada de buena ley es 0.0745%, que también es mayor al percentil equivalente del proceso 1.

**Resp.:** Una ley de cobre será declarada como ley media si:

- En el Proceso 1: la ley pertenece al intervalo  $[0.048; 0.0698]$ .
- En el Proceso 2: la ley pertenece al intervalo  $[0.0632; 0.0745]$

c. **Resp.:** El número estimado de muestra en cada proceso está dado por:

- En el proceso 1 es  $\alpha = \frac{36 \cdot 0.7}{100} = 25.2 \approx 25$ .

- En el proceso 2 es  $\alpha = \frac{38 \cdot 0.7}{100} = 26.6 \approx 27$ .

c) **Se pide:** Opinión respecto a la **simetría** de la distribución del porcentaje de cobre

Si se observa detenidamente las columnas de frecuencias absolutas de ambos procesos se advertirá que se trata de distribuciones bajas en los extremos y más altas en la parte central. Puede verse también que son distribuciones **unimodales**. Esto nos lleva a pensar que si estas distribuciones de frecuencias mostraran simetría en sus gráficas debería ocurrir que: el valor de la media aritmética con el valor de la moda y con el valor de la mediana coinciden

#### EN EL PROCESO 1:

**Resp.:**

- La media aritmética del porcentaje de cobre en el proceso 1 es

$$\bar{x} = \frac{\sum_{i=1}^m c_i n_i}{n} = 0.055$$

- La mediana del porcentaje de cobre en el proceso 1 es

$$P_{50} = LI_{(3)} + A_3 \cdot \frac{\left(\frac{36}{2} - N_{k-1}\right)}{n_3} = 0.053 + 0.006 \cdot \frac{(18-13)}{15} = 0.055$$

En la distribución de los porcentajes de cobre en el proceso 1 la moda fue 0.055.

Luego,  $\bar{x} = M_0 = M_e$ , y **existe simetría** de la distribución del porcentaje de cobre.

d. **Resp.:** El porcentaje de cobre promedio total es

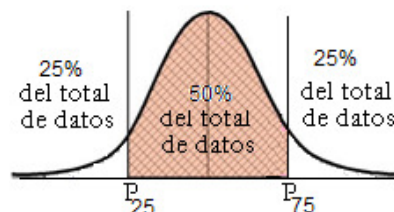
$$\bar{x}_T = \frac{36 \cdot 0.055 + 38 \cdot 0.0678}{74} = 0.0615729$$

- d) **Se pide:** Un intervalo que contenga el 50% central de la distribución del porcentaje de cobre de las muestras usando el proceso 2. Para hacer esta determinación debemos calcular los percentil  $P_{25}$  y  $P_{75}$

**EN EL PROCESO 2:**

$$P_{25} = 0.063 + 0.004 \cdot \frac{(9.5 - 7)}{11} = 0.063909$$

$$P_{75} = 0.071 + 0.004 \cdot \frac{(28.5 - 28)}{7} = 0.074$$

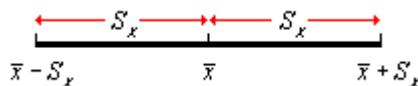


**Resp.:** El intervalo que contiene el 50% central de la distribución del porcentaje de cobre de las muestras en el proceso 2 es  $[0.063909, 0.074]$

- e) **Se pide:** Determinar, en el Proceso 1, un intervalo centrado en la media aritmética y de amplitud 4 desviaciones estándar, además de calcular el porcentaje de observaciones que tiene valores en este intervalo.

La identificación del intervalo pedido (antes que la **determinación**) requiere un pequeño análisis. Que el intervalo esté “centrado en la media”,  $\bar{x}$ , significa que el punto centro del intervalo es el valor de la media aritmética de los valores de porcentaje de cobre obtenidos por el Proceso 1.

Si la amplitud del intervalo es de 2 desviaciones estándar de estas leyes, entonces la **distancia entre los límites** del intervalo pedido es de 2 desviaciones estándar,  $S_x$ . Además, la distancia



desde la media aritmética (que es el punto centro del intervalo) a los extremos es de 1 desviación estándar (la mitad). Por lo tanto los límites de este intervalo serán  $\bar{x} - S_x$  y  $\bar{x} + S_x$  tal como lo indica la figura.

La siguiente tabla permite calcular la media aritmética y la desviación estándar de los valores de ley obtenidos por el Proceso 1. De la tabla se deduce que  $\bar{x} = 0.05533$  es la media aritmética de las leyes obtenidas por el Proceso 1 y que la varianza es

$$S_x^2 = \frac{0.11222}{36} - (0.05533)^2 = 5.556 \cdot 10^{-5}$$

| Porcentaje de cobre obtenido mediante el Proceso 1 |            |       |       |       |           |             |
|----------------------------------------------------|------------|-------|-------|-------|-----------|-------------|
| i                                                  | Intervalos |       | $n_i$ | $c_i$ | $c_i n_i$ | $c_i^2 n_i$ |
|                                                    | Linf       | Lsup  |       |       |           |             |
| 1                                                  | 0,041      | 0,047 | 5     | 0,044 | 0,22      | 0,00968     |
| 2                                                  | 0,047      | 0,053 | 8     | 0,05  | 0,4       | 0,02        |
| 3                                                  | 0,053      | 0,059 | 15    | 0,056 | 0,84      | 0,04704     |
| 4                                                  | 0,059      | 0,065 | 3     | 0,062 | 0,186     | 0,01153     |
| 5                                                  | 0,065      | 0,071 | 4     | 0,068 | 0,272     | 0,0185      |
| 6                                                  | 0,071      | 0,077 | 1     | 0,074 | 0,074     | 0,00548     |
| Suma :                                             |            |       |       |       | 1,992     | 0,11222     |
| Media :                                            |            |       |       |       | 0,05533   |             |
| Varianza :                                         |            |       |       |       | 5,556E-05 |             |
| Desviación Estándar                                |            |       |       |       | 0,00745   |             |

Por lo tanto la desviación estándar de las leyes obtenidas por el Proceso 1 es  $S_x = \sqrt{S_x^2} = \sqrt{5.556 \cdot 10^{-5}} = 0.00745 \%$ .

Esto significa que el intervalo centrado en la media y de longitud 2 desviaciones estándar es

$$[\bar{x} - S_x; \bar{x} + S_x] = [0,04788; 0,06279]$$

Para calcular el porcentaje de observaciones que tiene valores de porcentaje de cobre en este intervalo debe observarse que el límite inferior 0,04788 pertenece al intervalo  $(0,047; 0,053]$  (el orden del intervalo es  $k = 2$ , primer intervalo de la tabla).

Bajo el percentil  $P_i = 0,04788$  hay  $\alpha = \frac{36 \cdot i}{100}$  datos (no conocemos el porcentaje  $i$  que estamos averiguando). Entonces, utilizando los valores en la tabla y la expresión de cálculo de un percentil obtenemos

$$P_i = 0,04788 = 0,047 + 0,006 \cdot \frac{(\alpha - 5)}{8} \Rightarrow \alpha = 6,17333 \%$$

$$\alpha = \frac{36 \cdot i}{100} = 6,17333 \text{ Luego } i = 17,1482$$

Esto significa que bajo el percentil  $P_i = P_{17,1482} = 0,04788$  hay un 17,15% (aprox.) del total de valores de ley obtenidos por el Proceso 1.

De la misma manera trabajamos con el límite superior, 0,06279, que pertenece al intervalo  $(0,059; 0,065]$  (el orden del intervalo es  $k = 4$ , cuarto intervalo de la tabla).

Bajo el percentil  $P_j = 0,06279$  hay  $\alpha = \frac{36 \cdot j}{100}$  datos (no conocemos el porcentaje  $j$  que estamos averiguando). Entonces, utilizando los valores en la tabla y la expresión de cálculo de un percentil se obtiene

$$P_j = 0,06279 = 0,059 + 0,006 \cdot \frac{(\alpha - 28)}{3} \Rightarrow \alpha = 29,895 \text{ datos}$$

$$\alpha = \frac{36 \cdot j}{100} = 29,895 \text{ . Luego } j = 83,0417 \approx 83,05 \%$$

Esto significa que bajo el percentil  $P_j = P_{83,05} = 0,06279$  hay un 83,05% del total de valores de ley obtenidos por el Proceso 1.

Por lo tanto, en el intervalo  $(P_i; P_j] = (0,04788; 0,06279]$  hay  $83,05 - 17,15 = 65,9$  por ciento del total de observaciones de ley obtenida por el Proceso 1.

**Resp.:** El intervalo centrado en la media aritmética y de amplitud 2 desviaciones estándar de ley de cobre obtenida por el Proceso 1 es  $(0,04788; 0,06279]$  cuyos límites

están en porcentaje. Además, el porcentaje de observaciones que tiene valores de ley este intervalo es 65,9%.

## 1.2.2 Ejercicios Propuestos.

1. La siguiente información corresponde a la Albúmina total circulante (en gramos) en 30 pacientes internados en el Hospital Regional de Antofagasta:

| Albúmina |     | $c_i$    | $n_i$ |
|----------|-----|----------|-------|
| 100      | 110 | 105      | 2     |
| 110      | 120 | 115      | 6     |
| 120      | 130 | 125      | 6     |
| 130      | 140 | 135      | 7     |
| 140      | 150 | 145      | 8     |
| 150      | 160 | 155      | 1     |
|          |     | $n = 30$ |       |

- Si el 40% de los pacientes con mayor albúmina son considerados anormales ¿Cuál será su albúmina total circulante mínima para ser considerado anormal?
- Determine la albúmina total circulante más frecuente. .
- Si en el Hospital de Calama los pacientes tienen una albúmina promedio de 115 gramos, con una varianza de 170 gramos<sup>2</sup> ¿Cree usted que los pacientes de Antofagasta son más homogéneos en su albúmina total circulante respecto a los de Calama? Justifique su respuesta.

2. En una muestra de 35 barriles de yodo de una empresa minera de Antofagasta se determina el valor del peso (en kilos) en marzo 2003, obteniéndose los siguientes resultados de la tabla.

|    |    |    |    |    |
|----|----|----|----|----|
| 34 | 39 | 40 | 42 | 44 |
| 35 | 39 | 40 | 42 | 45 |
| 36 | 39 | 41 | 42 | 45 |
| 36 | 39 | 41 | 44 | 47 |
| 38 | 40 | 41 | 44 | 47 |
| 38 | 40 | 41 | 44 | 48 |
| 38 | 40 | 42 | 44 | 50 |

- Identifique los siguientes conceptos:
  - Población analizada
  - Variable en estudio
  - Elementos de la población
  - Tipo de variable.
- Construya una tabla de frecuencia completa de 8 categorías e interprete  $n_3, N_5, f_4, F_6$ .
- Construya un gráfico estadístico adecuado para la tabla construida en la parte a) e interprételo.
- Los barriles cuyo peso pertenece al intervalo que contiene el 50% central de la distribución de los pesos son considerados adecuados. Determine dicho intervalo.
- Si el instrumento que hace la medición está defectuoso y entrega el valor del peso disminuido en 20% de su verdadero valor más 2 kilos. ¿Cuál será el peso medio corregido? (use propiedades)
- Si el 33,5% de los barriles de menor peso y el 15,2% de los barriles con mayor peso son devueltos para nivelar el llenado y el resto es embarcado. Determine entre qué pesos un barril será embarcado.
- Si en una empresa minera de Tocopilla una muestra de 50 barriles proporcionó un peso medio de 45.8 kilos. ¿Cuál será el peso medio total considerando ambas empresas mineras?
- Determine el porcentaje exacto de barriles con un peso entre 39.5 y 42.5 kilos.
- ¿Cuál es el peso más frecuente de los barriles y la mediana? ¿Que puede decir de la simetría de la distribución de los pesos comparándolas con la media aritmética?

### 1.3. Planteamientos problemáticos para aplicación de propiedades.

Los problemas de aplicación que se proponen tienen por objetivo poner al estudiante frente a situaciones en las cuales deberá diseñar apropiadamente las respuestas, además de utilizar una notación adecuada RIGUROSAMENTE para aplicar las propiedades enunciadas al inicio de este capítulo.

#### Resumen de propiedades más importantes.

Si  $x_1, x_2, \dots, x_n$  un conjunto de  $n$  datos de los cuales conocemos su media aritmética  $\bar{x}$  y su varianza  $S_x^2$  podemos resumir las siguientes propiedades:

- a. Propiedad lineal de la media aritmética: Si cada observación  $x_i$  se transforma en  $a + bx_i = y_i$  entonces

$$y_i = a + bx_i \Rightarrow \bar{y} = \overline{a + bx} \Rightarrow \bar{y} = a + b\bar{x}, \text{ donde } a \text{ y } b \text{ son constantes reales.}$$

- b.  $\bar{x}$  es un centro de gravedad:  $\sum_{i=1}^n (x_i - \bar{x}) = 0$ .

- c.  $\sum_{i=1}^n (x_i - k)^2$  obtiene su valor mínimo cuando  $k = \bar{x}$

- d. Si la muestra está particionada en  $k$  submuestras de tamaños  $n_1, n_2, \dots, n_k$  con medias aritméticas conocidas  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ , entonces la media aritmética total de la muestra completa se define como:

$$\bar{x}_T = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \dots + \bar{x}_k n_k}{n} = \frac{\sum_{i=1}^k \bar{x}_i n_i}{n}$$

- e. Sea  $X$  una variable de la que se conoce su varianza  $S_x^2$  y sea  $Y$  una nueva variable con valores  $y = ax + b$  donde  $a$  y  $b$  son constantes conocidas, entonces  $S_Y^2 = S_{a+bX}^2 = a^2 S_X^2$  y  $S_Y = S_{a+bX} = |a| S_X$ .

- f. Si se tiene  $k$  conjuntos de datos con tamaños  $n_1, n_2, \dots, n_k$ , con medias  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$  y varianzas  $S_1^2, S_2^2, \dots, S_k^2$ , respectivamente, entonces la varianza del conjunto de  $n = n_1 + n_2 + \dots + n_k$  datos se llama **varianza total**, y se calcula por la siguiente expresión:

$$S_T^2 = \frac{\sum_{i=1}^k n_i \left[ S_i^2 + (\bar{x}_i)^2 \right]}{n} - \left( \frac{\sum_{i=1}^k \bar{x}_i n_i}{n} \right)^2$$

$$S_T^2 = \frac{\sum_{i=1}^k n_i \left[ S_i^2 + (\bar{x}_i)^2 \right]}{n} - \left( \frac{\sum_{i=1}^k n_i \bar{x}_i}{n} \right)^2$$

### 1.3.1 Ejercicios resueltos, paso a paso:

#### Ejemplo 1.

1. Los sueldos diarios de trabajadores de una pequeña empresa metalmecánica se muestran en la tabla. Debido a las continuas alzas de precios en el rubro alimenticio y transporte, los trabajadores se organizan para plantearle al dueño de la empresa que cada sueldo se reajuste en un 25% más un bono de 3000 pesos. El dueño de la empresa acoge parcialmente la petición de los trabajadores, le ofrece a cada trabajador el sueldo pedido disminuido en un 10%, si este queda en menos de 20.000 pesos, en cambio a los que quedan con el sueldo pedido en más de 20.000 les conserva el sueldo anterior y ofrece un bono de 4.000 pesos.

|        |        |        |        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 8000   | 12.500 | 15.000 | 10.000 | 13.500 | 18.000 | 8000   | 12.500 | 15.000 | 15.000 |
| 14.000 | 21.000 | 16.800 | 14.000 | 15.000 | 10.000 | 14.000 | 21.000 | 16.800 | 16.800 |

- ¿Cuál es el sueldo promedio que aspiran los trabajadores?
- ¿Cuál es sueldo promedio total que le ofrece el empleador?
- ¿Es más heterogéneo el actual sueldo o el que piden los trabajadores?

#### Esquema de desarrollo.

- Paso 1.** Aunque parezca una insistencia inadecuada, lea el problema detenidamente hasta comprenderlo.
- Paso 2.** Identificar los datos que entrega el problema. El sueldo actual de los trabajadores que designaremos por  $x_i$ , el sueldo a que aspiran los trabajadores que lo designaremos

Recuerda el siguiente plan cuando procedes a solucionar estos problemas:

- Paso 1.** Identificar los datos que entrega el problema.
- Paso 2.** Identificar las preguntas del problema.
- Paso 3.** Investigar conocimientos adicionales.
- Paso 4.** Proponer y realizar las respuestas a las preguntas.

por  $y_i$ , el sueldo que ofrece el dueño de la empresa que lo designaremos por  $t_i$ , si el sueldo pedido es menor a \$20.000,  $z_i$ , si el sueldo pedido es mayor a \$20.000

**Paso 3. Identificar las preguntas del problema:**

- Sueldo promedio que aspiran los trabajadores:  $\bar{y}$
- Sueldo promedio total que ofrece el dueño de la empresa:  $\bar{T}$
- Coeficiente de variación:
  - del actual sueldo  $CV_x = \frac{S_x}{\bar{x}} \cdot 100\%$ ,
  - del sueldo que a que aspiran los trabajadores:  $CV_y = \frac{S_y}{\bar{y}} \cdot 100\%$

**Paso 4. Desarrollo de la respuesta:**

- a. Sea  $x_i$  es sueldos de los trabajadores, su sueldo promedio es  $\bar{x} = 14.345$   
 Sea  $y_i$  el sueldo a que aspiran los trabajadores  $y_i = 1,25 \cdot x_i + 3000$   
**Resp:** Sueldo a que aspiran los trabajadores  $\bar{y} = 1,25 \cdot \bar{x} + 3000 = 20.391,25$
- b. El patrón una vez reajustado el sueldo ofrece:

|            |       |       |       |       |       |       |       |       |       |       |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Actual     | 8000  | 8000  | 10000 | 10000 | 12500 | 12500 | 13500 | 14000 | 14000 | 14000 |
| reajustado | 13000 | 13000 | 15500 | 15500 | 18625 | 18625 | 19875 | 20500 | 20500 | 20500 |
| ofrecido   | 11700 | 11700 | 13950 | 13950 | 16763 | 16763 | 17883 | 18000 | 18000 | 18000 |

|            |       |       |       |       |       |       |       |       |       |       |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Actual     | 15000 | 15000 | 15000 | 15000 | 16800 | 16800 | 16800 | 18000 | 21000 | 21000 |
| reajustado | 21750 | 21750 | 21750 | 21750 | 24000 | 24000 | 24000 | 25500 | 29250 | 29250 |
| ofrecido   | 19000 | 19000 | 19000 | 19000 | 20800 | 20800 | 20800 | 22000 | 25000 | 25000 |

Sea  $y_A$  el sueldo diario reajustado menor a \$20.000,

$$t_i = y_A - 0.1y_A = 0.9y_A \Rightarrow \bar{t} = 0,9\bar{y}_A = 0,9 \cdot 16303,5714 = 14673,213$$

Si el sueldo diario reajustado es mayor a \$20.000,

$$z_i = \underset{y > 20000}{x} + 4000 \text{ su promedio es } \Rightarrow \bar{z} = \bar{x} + 4000 = 20338,4615$$

**Resp:** El sueldo promedio total que ofrece el empleador es

$$\bar{T} = \frac{7 \cdot 14673,213 + 13 \cdot 20338,4615}{20} = 18355,625$$

c.  $CV_x = \frac{3520,5879}{14345} \cdot 100 = 0,2454 \cdot 100 = 24,54$

$$y_i = 1,25 \cdot x_i + 3000 \Rightarrow s_y = 1,25 \cdot s_x = 1,25 \cdot 3520,5879 = 4400,7$$

$$CV_Y = \frac{4400,7}{20931,25} \cdot 100 = 0,2102 \cdot 100 = 21,02$$

**Resp:** Efectivamente es más heterogéneo el actual sueldo de los trabajadores pues  $CV_X > CV_Y$ .

### Ejemplo 2.

En una empresa estatal existen tres sindicatos, el de los Supervisores, el de los Ingenieros de Ejecución-Administrativos y el de los Obreros.

Los números de trabajadores de los sindicatos de Supervisores, Ingenieros de Ejecución-Administrativos y el de los Obreros, están en relación 1:3:6.

- Hace dos años, los sindicatos negociaron con la empresa y obtuvieron sueldos promedios de 3 millones, 1 millón doscientos mil y 700 mil pesos, respectivamente. El coeficiente de variación de los sueldos de cada sindicato fue igual a 1. Calcule para el **total de trabajadores** (Supervisores, Ingenieros de Ejecución-Administrativo y Obreros):
  - El sueldo promedio.
  - La desviación estándar de los sueldos.
- Este año la empresa se propone que en la negociación, el sueldo promedio total sea exactamente de 1 millón ciento veinte mil pesos. Negocia por separado con cada sindicato. El primero en negociar es el sindicato de Supervisores, el cual obtiene un reajuste del 10%. El sindicato de Ingeniero de Ejecución-Administrativos, obtiene un 8% de reajuste. El sindicato de Obreros, que aún no ha negociado, conocedor del propósito de la empresa y de los reajustes obtenidos por los otros sindicatos, declara la huelga por cuanto estiman que bajo las actuales condiciones sus asociados quedarán desfavorecidos. ¿Concuerda usted con la estimación del sindicato de Obreros? Justifique su respuesta estadísticamente.

### Esquema de desarrollo.

**Paso 1.** Lea el problema detenidamente hasta comprenderlo.

**Paso 2.** Identificar los datos que entrega el problema.

- ✓  $n_A$ : Número de Supervisores.
- $n_B$ : Número de Ingenieros de Ejecución-Administrativos.
- $n_C$ : Número de Obreros.
- $\bar{x}_A$ : Promedio de sueldos de Supervisores.
- $\bar{x}_B$ : Promedio de sueldos de Ingenieros de Ejecución-Administrativos.

$\bar{x}_C$  : Promedio de sueldos de Obreros.

$CVx_A$  : Coeficiente de variación de sueldos de Supervisores.

$CVx_B$  : Coeficiente de variación de sueldos de Ingenieros de Ejecución-Administrativos.

$CVx_C$  : Coeficiente de variación de sueldos de Obreros.

$\bar{x}_T = \frac{n_A}{n} \cdot \bar{x}_A + \frac{n_B}{n} \cdot \bar{x}_B + \frac{n_C}{n} \cdot \bar{x}_C$  : Sueldo promedio total (del total de trabajadores de la empresa).

**Paso 3. Identificar las preguntas del problema:**

- $\bar{y}$  : Sueldo promedio total en base a la petición de los trabajadores.
- $S_T$  : Desviación estándar total en base a la petición de los trabajadores.
- $\bar{y}_{TE}$  : Sueldo promedio total que ofrece la empresa..
- Coeficiente de variación:
  - Del actual sueldo  $CV_x = \frac{S_x}{\bar{x}} \cdot 100\%$ .
  - Del sueldo que piden  $CV_y = \frac{S_y}{\bar{y}} \cdot 100\%$ .

**Paso 4. Desarrollo y respuesta a lo solicitado:**

$$n_A : n_B : n_C = 1 : 3 : 6 \Leftrightarrow \frac{n_A}{n_A + n_B + n_C} = \frac{1}{10} \Leftrightarrow \frac{n_A}{n} = \frac{1}{10}$$

$$n_A : n_B : n_C = 1 : 3 : 6 \Leftrightarrow \frac{n_B}{n_A + n_B + n_C} = \frac{3}{10} \Leftrightarrow \frac{n_B}{n} = \frac{3}{10}. \text{ Luego, } \frac{n_C}{n} = \frac{6}{10}$$

a.  $\bar{x}_A = 3.000.000, \bar{x}_B = 1.200.000, \bar{x}_C = 700.000$

$$CVx_A = 100 \Leftrightarrow \frac{Sx_A}{\bar{x}_A} = 1 \Leftrightarrow Sx_A = \bar{x}_A, CVx_B = 100 \Leftrightarrow \frac{Sx_B}{\bar{x}_B} = 1 \Leftrightarrow Sx_B = \bar{x}_B$$

$$CVx_C = 100 \Leftrightarrow \frac{Sx_C}{\bar{x}_C} = 1 \Leftrightarrow Sx_C = \bar{x}_C$$

**Resp:** Luego el sueldo promedio total de trabajadores (Supervisores, Ingenieros de Ejecución-Administrativo) es de \$1 080 000. En efecto:

$$\bar{x}_T = \frac{1}{10} \cdot 3.000.000 + \frac{3}{10} \cdot 1.200.000 + \frac{6}{10} \cdot 700.000 = 1.080.000$$

b.

$$S_T^2 = \frac{1}{10} \cdot [(3.000.000)^2 + (3.000.000)^2] + \frac{3}{10} \cdot [(1.200.000)^2 + (1.200.000)^2] +$$

$$+ \frac{6}{10} \cdot [(700.000)^2 + (700.000)^2] - (1.080.000)^2$$

$$S_T^2 = \frac{1}{5} \cdot [(3.000.000)^2] + 3 \cdot [(1.200.000)^2] + 6 \cdot [(700.000)^2] - (1.080.000)^2$$

$$S_T = 1.444.160,6$$

**Resp.** La desviación estándar total (desviación estándar del sueldo de todos los trabajadores de la empresa) es de \$1 444 160,6.

c. Sea  $\bar{y}_{TE}$  el sueldo promedio total ofrecido por la empresa = 1.220.000.

$\bar{y}_A = 1.1 \cdot \bar{x}_A = \$ 3 300 000$  es el sueldo promedio actual obtenido por los Supervisores.

$$\bar{y}_B = 1.08 \cdot \bar{x}_B = 1.296.000$$

$$1.120.000 = \frac{1}{10}(3.300.000) + \frac{3}{10}(1.296.000) + \frac{6}{10}(\bar{y}_C)$$

$$= \frac{3.300.000 + 3.888.000 + 6\bar{y}}{10}$$

$$11.200.000 = 7.188.000 + 6\bar{y}_C \Rightarrow \bar{y}_C = 668.667 \Rightarrow \frac{668.667 - 700.000}{700.000}$$

$$\Rightarrow k = -0.045 = -4.5\%$$

**Resp:** A los Obreros se les rebajaría el sueldo en un 4.5%.

de las muestras en el proceso 2 es  $[0.063909, 0.074]$

$$S_T^2 = \frac{\sum_{i=1}^k n_i [S_i^2 + (\bar{x}_i)^2]}{n} - \left( \frac{\sum_{i=1}^k n_i \bar{x}_i}{n} \right)^2$$

### 1.3.2 Ejercicios Propuestos.

1. Una fábrica vende dos tipos de artículos A y B. La cantidad de artículos del tipo B producidos por la fábrica es la tercera parte que la del tipo A. El costo medio de los artículos del tipo A es de \$4800. El costo medio total (artículos A y B juntos) es de \$4100.

Se requiere aumentar los precios de los artículos para lograr un costo medio total de \$5550 reajustando en un 3% el costo unitario de los artículos tipo A y aumentando K veces el costo unitario de los artículos tipo B.

- a. ¿Cuál es el costo medio inicial de los artículos del tipo B?
  - b. ¿Cuál es el valor de k?
  - c. En qué porcentaje cambió el costo unitario de los artículos tipo B?
2. En el mes de Marzo los trabajadores de una empresa tenían un sueldo promedio de \$350.000 con un coeficiente de variación del 15%. En el mes de Abril se incorporan nuevos trabajadores a la Empresa, en un número igual al 20% de los existentes en Marzo. Los nuevos trabajadores tienen un sueldo promedio de \$270.000 y una desviación estándar de \$30.000. En Septiembre se reajustan los sueldos de los trabajadores antiguos en un 10% y se les entrega un bono de Fiestas Patrias de \$20.000, mientras que para los nuevos trabajadores el aumento fue de un 20% con un bono de \$k. Justificando estadísticamente sus respuestas, responda a las siguientes preguntas:
    - a. ¿Cuál es el porcentaje de aumento de la planilla de sueldos del mes de Abril en relación a la del mes de Marzo?
    - b. ¿Cuál es la desviación estándar de los sueldos, del total de los trabajadores, en el mes de Abril?
    - c. ¿Cuál es el coeficiente de variación de los trabajadores antiguos en el mes de Septiembre?
    - d. Determine el valor de k de modo que ambos grupos de trabajadores, nuevos y antiguos, tengan el mismo promedio de sueldos.
  3. Un empresa realiza su trabajo diario en tres turnos: A, B y C: Los números de trabajadores por turno están en la relación 7:5:4, constituyendo un total de 160 trabajadores. Los salarios por turno, son \$8.000, \$10.000, \$12.000, y sus correspondientes desviaciones estándar, son: \$1.200, \$9.00, \$1.000, respectivamente. Si la empresa piensa hacer un aumento salarial según el siguiente criterio: reajuste del 15% más un bono de \$4.000, para cada trabajador del turno A, reajuste del 18% para cada trabajador del turno B, y solo un bono de \$700 para cada trabajador del turno C.

- a. ¿Conseguiría la empresa hacer más homogéneos los sueldos de todos los trabajadores en su conjunto?. Justifique estadísticamente su respuesta.
- b. Si la gerencia establece que la planilla diaria no debe aumentar más del 12%, por lo que decide modificar el bono de los trabajadores del turno A. ¿Cuál es el mayor bono, en cantidades enteras, que puede otorgar al turno A, además del reajuste, para lograr el objetivo de la empresa.

#### 1.4. Datos bidimensionales. Tablas de frecuencias conjuntas, marginales y condicionales.

- ✓ Llamaremos **dato bidimensional**,  $(x_i, y_i)$  al registro que presenta una característica conjunta de las variables  $X$  e  $Y$ , contenidas en el vector bidimensional  $(X, Y)$ .
- ✓ Para describir simultáneamente dos o más características estas quedan representadas en **Tablas de frecuencias conjuntas**, cuya construcción es similar al caso unidimensional, dependiendo del origen de cada variable. Estudiaremos solo el caso de dos variables, que llamaremos tabla de doble entrada
- ✓ Llamaremos **frecuencia absoluta conjunta** al número de veces que se presentan conjuntamente los valores  $x_i$  y  $y_j$ . Se representa por  $n_{ij}$
- ✓ Con los valores de las variables como sus respectivas frecuencias, llamadas **Frecuencias marginales** que se encuentran a ubicadas al margen de la tabla de doble entrada podemos construir **distribuciones marginales**
- ✓ Una **distribución condicional** es una distribución unidimensional de una variable condicionada a un valor fijo de la otra variable, en la cual, a su vez se pueden determinar los diferentes estadígrafos de posición y dispersión estudiados.

##### 1.4.2 Ejercicios resueltos, paso a paso:

##### Ejemplo 1.

1. Los siguientes datos corresponden a la edad  $X$  (en años) y el índice de productividad  $Y$  (en porcentaje) de 40 temporeras de la hacienda “Los Aromos”.
  - a. ¿Cuál es el promedio de edad de las temporeras cuyo índice de productividad está entre 40 y 60?

| $x \backslash y$ | 20-40 | 40-60 | 60-80 | 80-100 | Total |
|------------------|-------|-------|-------|--------|-------|
| 16-20            | 0     | 2     | 1     | 2      | 5     |
| 20-24            | 2     | 2     | 2     | 2      | 8     |
| 24-28            | 2     | 3     | 4     | 1      | 10    |
| 28-32            | 3     | 5     | 1     | 1      | 10    |
| 32-36            | 4     | 2     | 1     | 0      | 7     |
| Total            | 11    | 14    | 9     | 6      | 40    |

- ¿Cuál es el índice de productividad más frecuente de las temporeras cuya edad está entre 20 y 28 años.
- ¿Afirmaría usted que la edad de las temporeras presenta mayor dispersión relativa que el índice de productividad?. Justifique su respuesta.
- Determine el índice de productividad mínimo del 30% superior de la distribución de la productividad de las temporeras cuya edad varía de 20 a 28 años.
- Si el 10% de temporeras de menor edad son considerados Principiantes y si el 20% de las temporeras con mayor edad son consideradas Expertas, los demás temporeras son considerados Normales. ¿Entre qué edad las temporeras son consideradas Normales?
- Determine el porcentaje de temporeras cuyo índice de productividad es menor o igual 65.
- Determine un intervalo centrado en la media aritmética de longitud dos errores estándar de estimación para la edad de las temporeras.

### Esquema de desarrollo:

#### Paso 1. Leer el problema:

- ✓ Se da una tabla de doble entrada exhibiendo su distribución conjunta y las marginales.
- ✓ Se trata de una muestra de  $n = 20$  personas en las que se mide su edad ( $X$ ) y su productividad ( $Y$ ).
- ✓ Se pide:
  - Un promedio de edad condicionado a valores de productividad.
  - Una moda de un índice de productividad condicionada a un rango de edad.
  - Cálculo de un coeficiente de variación para comparar la dispersión relativa de la edad con la productividad.
  - Un percentil del índice de productividad.
  - Dos percentiles para determinar un intervalo.
  - Un porcentaje del total de datos asociado a un valor dado de un percentil.
  - Un intervalo de longitud dos errores estándar de estimación centrado en la media observada de la edad.

#### Paso 2. Identificar las variables:

$X$  : Edad de las temporeras de la muestra (en años).

$Y$  : Índice de productividad de la temporera.

#### Paso 3. Desarrollar la respuesta con fundamentos y responder:

- Se pregunta por el promedio de edad de las temporeras condicionada a una productividad está entre 40 y 60

$$\bar{x} = \frac{\sum_{i=1}^5 c_i n_i}{14} = \frac{376}{14} = 26,86 \text{ años}$$

| x/y(40,60) |     |       |       |             |
|------------|-----|-------|-------|-------------|
| L.I        | L.S | $c_i$ | $n_i$ | $c_i * n_i$ |
| 16         | 20  | 18    | 2     | 36          |
| 20         | 24  | 22    | 2     | 44          |
| 24         | 28  | 26    | 3     | 78          |
| 28         | 32  | 30    | 5     | 150         |
| 32         | 36  | 34    | 2     | 68          |
| suma       |     |       | 14    | 376         |

**Resp:** La edad promedio de las temporeras cuyo índice de productividad está entre 40 y 60 es de 26,86 años.

- b. Se pide la moda del índice de productividad condicionado a aquellas temporeras cuya edad está entre 20 y 28 años

$$M_o = LI_{(i)} + A_i \cdot \frac{(n_i - n_{i-1})}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

$$= 60 + 20 \cdot \frac{(6 - 5)}{(6 - 5) + (6 - 3)} = 65$$

| y/x(20-28) |        |    |
|------------|--------|----|
| LI         | LS sup | ni |
| 20         | 40     | 4  |
| 40         | 60     | 5  |
| 60         | 80     | 6  |
| 80         | 100    | 3  |
| Suma       |        | 18 |

**Resp:** El índice de productividad de aquellas temporeras cuya edad está entre 20 y 28 años es de 65.

- c. Se debe verificar si la edad de las temporeras presenta mayor dispersión relativa que el índice de productividad. Para realizar esto se debe calcular el coeficiente de variación de cada variable. Se trabaja, así, con los valores marginales de la tabla dada.

$$\bar{x} = \frac{\sum_{i=1}^5 c_i n_i}{40} = \frac{1064}{40} = 26,6 \text{ años}$$

$$S_x = \sqrt{\frac{\sum_{i=1}^5 n_i (c_i - \bar{x})^2}{40}}$$

$$= \sqrt{\frac{1041,6}{40}} = 5,10 \text{ años}$$

| Edad x |    |                |                |                                |                                                            |
|--------|----|----------------|----------------|--------------------------------|------------------------------------------------------------|
| LI     | LS | c <sub>i</sub> | n <sub>i</sub> | n <sub>i</sub> *c <sub>i</sub> | n <sub>i</sub> *(c <sub>i</sub> - $\bar{x}$ ) <sup>2</sup> |
| 16     | 20 | 18             | 5              | 90                             | 369,8                                                      |
| 20     | 24 | 22             | 8              | 176                            | 169,28                                                     |
| 24     | 28 | 26             | 10             | 260                            | 3,6                                                        |
| 28     | 32 | 30             | 10             | 300                            | 115,6                                                      |
| 32     | 36 | 34             | 7              | 238                            | 383,32                                                     |
| suma   |    |                | 40             | 1064                           | 1041,6                                                     |

Luego el coeficiente de variación de la edad  $X$  es:

$$CV_x = \frac{5,10}{26,6} \cdot 100 = 19,17$$

Análogamente,

$$\bar{y} = \frac{\sum_{i=1}^4 d_i n_i}{40} = \frac{2200}{40} = 55$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^4 n_i (d_i - \bar{y})^2}{40}} = \sqrt{\frac{415}{40}} = 20,37$$

Luego el coeficiente de variación de  $Y$  es:

$$CV_Y = \frac{20,37}{55} \cdot 100 = 37,04$$

| Indice y |     |       |       |                 |                               |
|----------|-----|-------|-------|-----------------|-------------------------------|
| LI       | LS  | $d_i$ | $n_i$ | $d_i \cdot n_i$ | $n_i \cdot (d_i - \bar{y})^2$ |
| 20       | 40  | 30    | 11    | 330             | 6875                          |
| 40       | 60  | 50    | 14    | 700             | 350                           |
| 60       | 80  | 70    | 9     | 630             | 2025                          |
| 80       | 100 | 90    | 6     | 540             | 7350                          |
| suma     |     | 40    | 2200  | 16600           |                               |

**Resp:** El índice de productividad presenta mayor dispersión relativa que la edad de las temporeras, pues  $CV_Y > CV_X$ .

- d. Se pide el índice de productividad mínimo del 30% superior de la distribución de la productividad de las temporeras cuya edad varía de 20 a 28 años. Es decir debemos calcular el percentil 70,  $P_{70}$ :

$$P_{70} = LI_{(3)} + A_3 \cdot \frac{(\frac{n \cdot 70}{100} - N_2)}{n_3}$$

$$= 60 + 20 \cdot \frac{(12,6 - 9)}{6} = 72$$

| y/x(20-28) |     |       |       |       |  |
|------------|-----|-------|-------|-------|--|
| LI         | LS  | $d_i$ | $n_i$ | $N_i$ |  |
| 20         | 40  | 30    | 4     | 4     |  |
| 40         | 60  | 50    | 5     | 9     |  |
| 60         | 80  | 70    | 6     | 15    |  |
| 80         | 100 | 90    | 3     | 18    |  |
| suma       |     |       | 18    |       |  |

**Resp:** El índice de productividad mínimo del 30% más alto, de valores de la distribución de la productividad de las temporeras cuya edad varía de 20 a 28 años es 72.

- e. Se pide entre qué edades se consideran Normales, para esto, debemos calcular el **percentil 10** y el **percentil 80**

$$P_{10} = LI_{(1)} + A_1 \cdot \frac{(\frac{40 \cdot 10}{100} - 0)}{n_1} = 16 + 4 \cdot \frac{(4)}{5} = 16,2$$

$$P_{80} = LI_{(4)} + A_4 \cdot \frac{(\frac{40 \cdot 80}{100} - N_3)}{n_4} = 28 + 4 \cdot \frac{(32 - 23)}{10} = 31,6$$

| Edad x |    |       |       |       |  |
|--------|----|-------|-------|-------|--|
| LI     | LS | $c_i$ | $n_i$ | $N_i$ |  |
| 16     | 20 | 18    | 5     | 5     |  |
| 20     | 24 | 22    | 8     | 13    |  |
| 24     | 28 | 26    | 10    | 23    |  |
| 28     | 32 | 30    | 10    | 33    |  |
| 32     | 36 | 34    | 7     | 40    |  |
| suma   |    |       | 40    |       |  |

**Resp:** Las temporeras son consideradas Normales entre las edades de 16,2 y 31,6

- f. Se pide el **porcentaje** de temporeras cuyo índice de productividad es **menor o igual** que 65, esto quiere decir que el valor del percentil de productividad es 65. Se deduce, entonces,

que existe un valor de porcentaje,  $\alpha$ , tal que  $P_\alpha = 65$ . En consecuencia, se debe hallar el valor de  $\alpha$ .

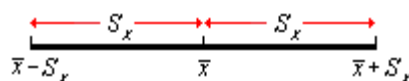
$$P_\alpha = 65 = 60 + 20 \cdot \frac{(\alpha - 25)}{9} \Leftrightarrow \frac{(65 - 60) \cdot 9}{20} = \alpha - 25$$

$$\Rightarrow \alpha = 27,25$$

**Resp:** El porcentaje de temporeras cuyo índice de productividad es menor o igual 65 es de 27,25%.

| Indice y |     |                |                |                |
|----------|-----|----------------|----------------|----------------|
| LI       | LS  | d <sub>i</sub> | n <sub>i</sub> | N <sub>i</sub> |
| 20       | 40  | 30             | 11             | 11             |
| 40       | 60  | 50             | 14             | 25             |
| 60       | 80  | 70             | 9              | 34             |
| 80       | 100 | 90             | 6              | 40             |
| suma     |     |                | 40             |                |

- g. **Se pide:** Determinar un intervalo centrado en la media aritmética y de amplitud 2 desviaciones estándar para la edad de las temporeras



**Resp:** Como en c) obtuvimos los datos la media aritmética y la desviación estándar el intervalo pedido está entre

$$[\bar{x} - s; \bar{x} + s] = [26,6 - 5,10; 26,6 + 5,10] = [21,15; 31,10]$$

### Ejemplo 2.

En el consultorio Norte obtiene la clasificación de 124 pacientes según peso ( $X$ , en kgs.) y su grupo etáreo ( $Y$ ).

| Peso(X) | Grupo etario(Y) |              |        |              | Total |
|---------|-----------------|--------------|--------|--------------|-------|
|         | Joven           | Adulto joven | Adulto | Adulto Mayor |       |
| 40-50   | 0               | 1            | 1      | 0            | 2     |
| 50-60   | 0               | 10           | 2      | 0            | 12    |
| 60-70   | 1               | 25           | 10     | 0            | 36    |
| 70-80   | 0               | 13           | 22     | 1            | 36    |
| 80-90   | 0               | 1            | 22     | 5            | 28    |
| 90-100  | 0               | 0            | 3      | 2            | 5     |
| 100-110 | 0               | 0            | 1      | 4            | 5     |
| Total   | 1               | 50           | 61     | 12           | 124   |

- Calcular e interpretar:  $n_{32}$ ,  $n_{23}$ ,  $n_{2\bullet}$ ,  $n_{\bullet 4}$ .
- ¿Cuál es el peso promedio de los pacientes?
- ¿Cuál es la moda del grupo etario?
- Determine el peso mínimo del 30% superior de la distribución del peso de los pacientes.
- Calcular el peso (kg.) promedio de los pacientes considerados adultos.
- Si el 10% de los pacientes adultos de menor peso son considerados de Bajo peso y si el 20% de los pacientes adultos de mayor peso son considerados Sobre peso y el resto son considerados Normales. ¿Entre que peso los adultos son consideradas Normales?
- Para los pacientes Adulto joven, determine el porcentaje de pacientes con un peso mayor o igual a 83 kilos. **Fundamente su respuesta.**
- Afirmaría usted que Ud. que el peso de los pacientes Adulto Joven es más homogéneo que el de los Adultos?

### Esquema de desarrollo.

#### Paso 1. Leer el problema.

#### Paso 2. Identificar las variables:

$X$  : Peso (en kg.) del paciente.

$Y$  : Grupo etario del paciente.

#### Paso 3. Desarrollar la respuesta indicando fundamentos y responder.

- a. Identificamos cada frecuencia y luego interpretamos

**Resp:**

$n_{32} = 25$ , es decir hay 25 pacientes

cuyo peso es mayor que 60 kg y menor o igual a 70 kg y está considerado en el grupo etario Adulto joven.

$n_{23} = 2$ , es decir hay 2 pacientes

cuyo peso es mayor que 50 kg. y menor o igual a 60 kg y está considerado en el grupo etario Adulto.

$n_{2\bullet} = 12$ , hay 12 pacientes cuyo peso es mayor que 50 kg. y menor o igual a 60 kg.

$n_{\bullet 4} = 12$ , hay 12 pacientes en el grupo etario Adulto mayor.

- b. Como nos piden el peso promedio de los pacientes, trabajamos con las frecuencias marginales para la variable peso ( $X$ )

**Resp:** El peso promedio de los pacientes es de

$$\bar{x} = \frac{\sum_{i=1}^7 c_i n_i}{124} = \frac{9170}{124} = 73,9516$$

| G. Etario    | $n_i$ |
|--------------|-------|
| Joven        | 1     |
| Adulto Joven | 50    |
| Adulto Joven | 61    |
| Adulto Mayor | 12    |
| Total        | 124   |

- c. Se pide la moda del grupo etario.

**Resp:** Notemos que el mayor grupo de pacientes pertenece a la categoría de adulto, que es la moda en el grupo etario.

El peso mínimo del 30% superior de la distribución del peso de los pacientes, corresponde al percentil 70:

Aplicamos la fórmula del percentil

| Peso | ci  | $n_i$ | $N_i$ |
|------|-----|-------|-------|
| 40   | 50  | 45    | 2     |
| 50   | 60  | 55    | 12    |
| 60   | 70  | 65    | 36    |
| 70   | 80  | 75    | 36    |
| 80   | 90  | 85    | 28    |
| 90   | 100 | 95    | 5     |
| 100  | 110 | 105   | 5     |
|      |     | 124   | 124   |

$$P_{70} = 80 + 10 \cdot \frac{\left(\frac{124 \cdot 70}{100} - 86\right)}{28} = 80 + 10 \cdot \frac{(86,8 - 86)}{28}$$

$$= 80,29$$

**Resp:** El peso mínimo del 30% superior de la distribución del peso de los pacientes es de 80,29 kg. O también podemos decir que el 70% de los pacientes tiene un peso menor o igual a 80,29 kg. y 30% de los paciente tiene un peso mayor que 80,29.

| Peso |     | ci  | n <sub>A</sub> | c <sub>i</sub> · n <sub>A</sub> |
|------|-----|-----|----------------|---------------------------------|
| 40   | 50  | 45  | 1              | 45                              |
| 50   | 60  | 55  | 2              | 110                             |
| 60   | 70  | 65  | 10             | 650                             |
| 70   | 80  | 75  | 22             | 1650                            |
| 80   | 90  | 85  | 22             | 1870                            |
| 90   | 100 | 95  | 3              | 285                             |
| 100  | 110 | 105 | 1              | 105                             |
|      |     |     | 61             | 4715                            |

- d. Para realizar esta operación, construimos una nueva tabla de frecuencias donde la columna de frecuencias absolutas corresponde a la columna de los adultos

**Resp:** Luego el peso promedio de los adultos es:

$$\bar{x}_{y \in \text{Adultos}} = \frac{\sum_{i=1}^8 n_i \cdot c_i}{n} = \frac{4175}{61} = 68.44 \text{ kg}$$

- e. Nos piden calcular el  $P_{10}$  y el  $P_{80}$  en el grupo adulto.

$$P_{10} = 60 + 10 \cdot \frac{\left(\frac{61 \cdot 10}{100} - 3\right)}{10} = 60 + 10 \cdot \frac{(6,1 - 3)}{10} = 63,1$$

$$P_{80} = 80 + 10 \cdot \frac{\left(\frac{61 \cdot 80}{100} - 35\right)}{22} = 80 + 10 \cdot \frac{(48,8 - 35)}{22} = 86,27$$

| Peso |     | c <sub>i</sub> | n <sub>A</sub> | N <sub>i</sub> |
|------|-----|----------------|----------------|----------------|
| 40   | 50  | 45             | 1              | 1              |
| 50   | 60  | 55             | 2              | 3              |
| 60   | 70  | 65             | 10             | 13             |
| 70   | 80  | 75             | 22             | 35             |
| 80   | 90  | 85             | 22             | 57             |
| 90   | 100 | 95             | 3              | 60             |
| 100  | 110 | 105            | 1              | 61             |
|      |     |                | 61             |                |

**Resp:** El peso Normal de un Adulto está entre 63,1 y 86,27 Kg.

- f. Se pide obtener es el porcentaje de pacientes Adultos jóvenes con peso mayor o igual a 83 kilos. Es decir  $P_{\alpha} = 83$ .

Usando fórmula del percentil

$$83 = 80 + 10 \cdot \frac{\left(\frac{50 \cdot \alpha}{100} - 49\right)}{1}, \text{ donde } \alpha = \frac{i \cdot n}{100}$$

| Peso |     | ci  | n <sub>Aj</sub> | N <sub>Aj</sub> |
|------|-----|-----|-----------------|-----------------|
| 40   | 50  | 45  | 1               | 1               |
| 50   | 60  | 55  | 10              | 11              |
| 60   | 70  | 65  | 25              | 36              |
| 70   | 80  | 75  | 13              | 49              |
| 80   | 90  | 85  | 1               | 50              |
| 90   | 100 | 95  | 0               | 50              |
| 100  | 110 | 105 | 0               | 50              |
|      |     |     | 50              |                 |

Se debe despejar  $i$ :

$$\frac{(83-80)}{10} \cdot 1 + 49 = 0,5i \Leftrightarrow i = \frac{49,3}{0,5} \Leftrightarrow i = 98,6,$$

**Resp:** Como se pide determinar el porcentaje de pacientes con un peso mayor o igual a 83 kilos, el complemento es  $(100-98,6)\% = 1,4\%$ .

- g. **Se pide:** Calcular el coeficiente de variación del peso de los pacientes Adulto Joven y el de los Adultos, y compararlos.

Luego,  $CV_{X_{y \in \text{Adultos jóvenes}}} = 11.97$ .

| Peso |     | $c_i$ | $n_{AJ}$ | $n_{AJ} \cdot c_i$ | $n_{AJ} \cdot (c_i - \text{prom})^2$ |
|------|-----|-------|----------|--------------------|--------------------------------------|
| 40   | 50  | 45    | 1        | 45                 | 424,36                               |
| 50   | 60  | 55    | 10       | 550                | 1123,6                               |
| 60   | 70  | 65    | 25       | 1625               | 9                                    |
| 70   | 80  | 75    | 13       | 975                | 1148,68                              |
| 80   | 90  | 85    | 1        | 85                 | 376,36                               |
| 90   | 100 | 95    | 0        | 0                  | 0                                    |
| 100  | 110 | 105   | 0        | 0                  | 0                                    |
|      |     |       | 50       | 3280               | 3082                                 |

Se calcula a continuación el coeficiente de variación de Adultos.

Luego,  $CV_{X_{y \in \text{Adultos}}} = 13.54$ .

| Peso |     | $c_i$ | $n_A$ | $n_A \cdot c_i$ | $n_A \cdot (c_i - \text{prom})^2$ |
|------|-----|-------|-------|-----------------|-----------------------------------|
| 40   | 50  | 45    | 1     | 45              | 1042,972319                       |
| 50   | 60  | 55    | 2     | 110             | 994,1413598                       |
| 60   | 70  | 65    | 10    | 650             | 1511,690406                       |
| 70   | 80  | 75    | 22    | 1650            | 115,8828272                       |
| 80   | 90  | 85    | 22    | 1870            | 1306,046762                       |
| 90   | 100 | 95    | 3     | 285             | 940,3923676                       |
| 100  | 110 | 105   | 1     | 105             | 767,5624832                       |
|      |     |       | 61    | 4715            | 6678,688525                       |

**Resp:** Efectivamente el peso de los pacientes Adulto Joven es más homogéneo el de Adultos, puesto que  $CV_{X_{y \in \text{Adultos}}} > CV_{X_{y \in \text{Adultos jóvenes}}}$

### 1.4.2 Ejercicios Propuestos.

1. Los siguientes datos corresponden a la edad  $X$  (en años) y el ingreso (semanal)  $Y$ , de 58 estudiantes universitarios que laboran en los diferentes supermercados de Antofagasta

- Intérprete:  $n_{3\bullet}$ ,  $n_{\bullet 4}$ ,  $n_{34}$ ,  $n_{43}$ .
- ¿Cuál es el promedio de edad de los estudiantes cuyo ingreso semanal está entre 20.000 y 25.000?
- ¿Cuál es el sueldo más frecuente de los estudiantes cuya edad está entre 24 y 26 años?
- ¿Afirmaría usted que los ingresos de los estudiantes cuya edad está entre 24 y 26 años es más homogéneo que el ingreso de los

| $x \backslash y$ | 10.000-15.000 | 15.000-20.000 | 20.000-25.000 | 25.000-30.000 | Total |
|------------------|---------------|---------------|---------------|---------------|-------|
| 18-20            | 3             | 2             | 1             | 2             | 8     |
| 20-22            | 2             | 2             | 2             | 2             | 8     |
| 22-24            | 4             | 3             | 4             | 1             | 12    |
| 24-26            | 3             | 5             | 1             | 1             | 10    |
| 26-28            | 4             | 6             | 4             | 6             | 20    |
| Total            | 16            | 18            | 12            | 12            | 58    |

estudiantes cuya edad está entre 18 y 20 años?. Justifique su respuesta.

- e. Si el 20% de los estudiantes de menor edad son considerados Nuevos y si el 30% de los estudiantes con mayor edad son considerados Antiguos, los demás estudiantes son considerados Normales. ¿Entre qué edad los estudiantes son consideradas Normales?
- f. ¿Cuántos estudiantes no son considerados antiguos?

2. La siguiente tabla muestra la distribución de ostiones para cosecha registradas en un cultivo en el año 1998, según el peso (en gramos) y tipos de ambiente:

| Peso      | Ambiente       |     |             |          |
|-----------|----------------|-----|-------------|----------|
|           | jaula flotante | mar | laboratorio | estanque |
| 3.0 - 3.8 | 22             | 27  | 21          | 0        |
| 3.8 - 4.6 | 15             | 21  | 13          | 0        |
| 4.6 - 5.4 | 55             | 12  | 15          | 23       |
| 5.4 - 6.2 | 0              | 0   | 25          | 45       |
| 6.2 - 7.0 | 0              | 14  | 14          | 27       |

- a. ¿Cuál es el peso promedio de los ostiones de mar?
- b. ¿Qué porcentaje de ostiones en jaula flotante tiene un peso mayor a 48 gramos?
- c. ¿El peso de los ostiones de mar es más homogéneo que es de laboratorio?. Justifique su respuesta, con el indicador adecuado
- d. ¿De qué ambiente es más frecuente que procedan los ostiones?
- e. ¿Existe simetría en la distribución del peso en los ostiones de laboratorio?

3. La siguiente tabla muestra la distribución de 90 automóviles importados por la Empresa Automotora Autocar considerando el número de cilindros y caballos de fuerza

| Distribución conjunta de automóviles importados según número de cilindros y caballos de fuerza |                    |              |               |               |               |               |       |
|------------------------------------------------------------------------------------------------|--------------------|--------------|---------------|---------------|---------------|---------------|-------|
| Automotora AUTOCAR                                                                             |                    |              |               |               |               |               |       |
| N° Cilindros                                                                                   | Caballos de fuerza |              |               |               |               |               | Total |
|                                                                                                | 54-81.43           | 81.43-108.86 | 108.86-136.29 | 136.29-163.72 | 163.72-191.15 | 191.15-218.58 |       |
| 3                                                                                              | 2                  | 0            | 0             | 0             | 0             | 0             | 2     |
| 4                                                                                              | 7                  | 27           | 15            | 1             | 0             | 0             | 50    |
| 5                                                                                              | 0                  | 0            | 1             | 0             | 0             | 0             | 1     |
| 6                                                                                              | 0                  | 0            | 11            | 15            | 1             | 0             | 27    |
| 8                                                                                              | 0                  | 0            | 3             | 4             | 2             | 1             | 10    |
| Total                                                                                          | 9                  | 27           | 30            | 20            | 3             | 1             | 90    |

- a. ¿Cuál es el número promedio de cilindro de los 90 automóviles?
- b. Calcular la cantidad de automóviles con caballos de fuerza entre 108 y 130?
- c. Se cree que los caballos de fuerza de los automóviles son más heterogéneo en aquellos automóviles de 5 cilindradas que aquellos automóviles de 6 cilindradas.