

SOLUCIÓN EXAMEN PARCIAL I

Nombres:

Apellidos:

C.I.:

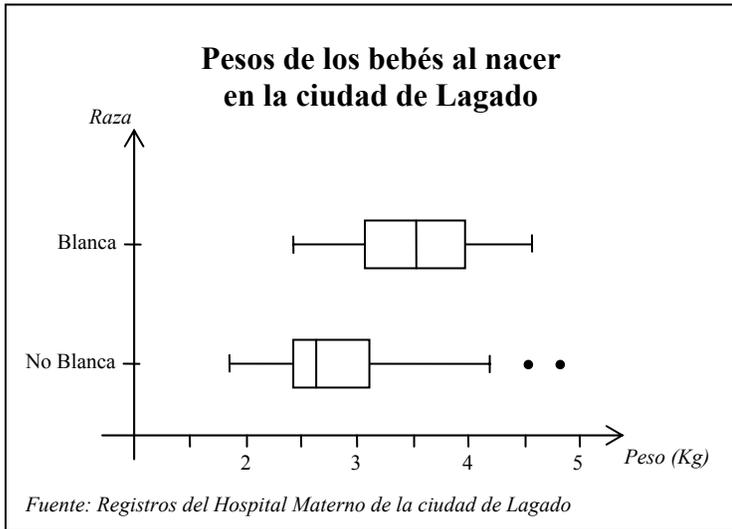
Firma:

Fecha: 07/03/05

MÉTODOS ESTADÍSTICOS I – EXAMEN I

Prof. Gudberto León

PARTE I: En el siguiente gráfico se muestran los diagramas de caja correspondientes a los pesos de los bebés al nacer según raza (blanca y no blanca). Los datos fueron recolectados en el hospital materno de la ciudad de Lagado. Marque con un círculo la respuesta correcta (0,5 puntos c/u).



- I. En cuanto a la dispersión de los pesos en los dos grupos de bebés:
- Los de raza blanca presentan mayor dispersión que los de raza no blanca
 - Los de raza blanca presentan menor dispersión que los de raza no blanca
 - Los de raza blanca presentan aproximadamente la misma dispersión que los de raza no blanca
 - Ambos grupos casi no presentan dispersión
 - La dispersión de los pesos de los bebés de raza blanca y no blanca no se puede visualizar en estos diagramas de caja
- II. La forma de la distribución de los pesos de los bebés de raza no blanca es:
- Simétrica
 - Asimétrica negativa
 - Asimétrica positiva
 - No se puede determinar la forma
 - Estos diagramas de caja no muestran la forma de la distribución de los pesos.
- III. Según los diagramas de caja, el bebé con el menor peso:
- Es de raza blanca
 - Es de raza no blanca
 - Es un valor atípico
 - Es representado por un punto
 - No se puede visualizar en estos diagramas de caja
- IV. En cuanto a la dispersión de la *parte central* de los pesos en los dos grupos de bebés:
- Los de raza blanca presentan aproximadamente la misma dispersión que los de raza no blanca
 - Ambos grupos casi no presentan dispersión
 - Los de raza blanca presentan menor dispersión que los de raza no blanca
 - Los de raza blanca presentan mayor dispersión que los de raza no blanca
 - La dispersión de la *parte central* de los pesos de los bebés de raza blanca y no blanca no se puede visualizar en estos diagramas de caja
- V. El peso que recomiendan los pediatras que debe presentar un bebé al nacer debe encontrarse entre 3 y 4 Kg. para que en el futuro sea sano. Según los diagramas de caja:
- Los bebés de raza no blanca presentan en promedio un peso dentro del rango recomendado por los pediatras.
 - Tanto los bebés de raza blanca como los de raza no blanca presentan en promedio un peso dentro del rango recomendado por los pediatras.
 - Los bebés de raza blanca presentan en promedio un peso dentro del rango recomendado por los pediatras.
 - Ni los bebés de raza blanca ni los de raza no blanca presentan en promedio un peso dentro del rango recomendado por los pediatras.
 - Los pesos promedios de los bebés no se pueden visualizar en estos diagramas de caja

PARTE II: Marque con un círculo la respuesta correcta o llene los espacios en blanco, excepto en pregunta 5. (0,5 puntos. c/u)

- VI. En estadística, cuando se analiza una muestra aleatoria para generalizar las conclusiones obtenidas a la población, se conoce con el nombre de:
- Muestreo
 - Estadística Descriptiva
 - Inferencia Estadística
 - Censo
 - Parámetro
- VII. La media aritmética de una población es:
- Una medida de dispersión
 - Un estadístico
 - Una medida de forma
 - Un parámetro
 - Un dato

SOLUCIÓN EXAMEN PARCIAL I

VIII. Las amplitudes de clase en una distribución de frecuencias:

- a. Siempre son iguales
- b. Pueden ser diferentes**
- c. Siempre son cinco
- d. Al menos son quince
- e. Son los intervalos de clase

IX. Si en un estudio estadístico se analizan conjuntamente las características *sexo*, *raza*, y *estado civil* de un grupo de personas, se habla de una variable **MULTIVARIANTE (O TRIVARIANTE)**

X. Escriba el formato de una tabla de tres entradas para las variables: *sexo* (M, F), *raza* (Blanca, Negra, Mestiza), y *estado civil* (Soltero, Casado, Viudo, Divorciado).

Un formato posible para una tabla de tres entradas de las variables *sexo*, *raza*, y *estado civil* se presenta en el siguiente cuadro:

	RAZA					
	<i>Blanca</i>		<i>Negra</i>		<i>Mestiza</i>	
	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>	<i>M</i>	<i>F</i>
EDO. CIVIL						
<i>Soltero</i>						
<i>Casado</i>						
<i>Viudo</i>						
<i>Divorciado</i>						

PARTE III:

Nota: Recuerde que este es un examen de desarrollo por lo que **debe** incluir **todos** los pasos necesarios que justifiquen los resultados. Favor encerrar sus respuestas de forma tal que sea fácil de encontrarlas en su desarrollo. **Solamente** se responderán aquellas preguntas tendientes a aclarar enunciados de los problemas.

1. El Departamento de Trabajo de Estados Unidos publica varias clasificaciones de la tasa de desempleo. Recientemente, el departamento registró las siguientes categorías educativas:

Nivel de Educación	% de Desempleados
No terminó el nivel de bachillerato	35
Terminó el nivel de bachillerato	31
Asistió a la universidad pero no recibió el grado	16
Recibió un grado universitario	9
Asistió a postgrado pero no recibió el grado	6
Recibió un título de postgrado	3

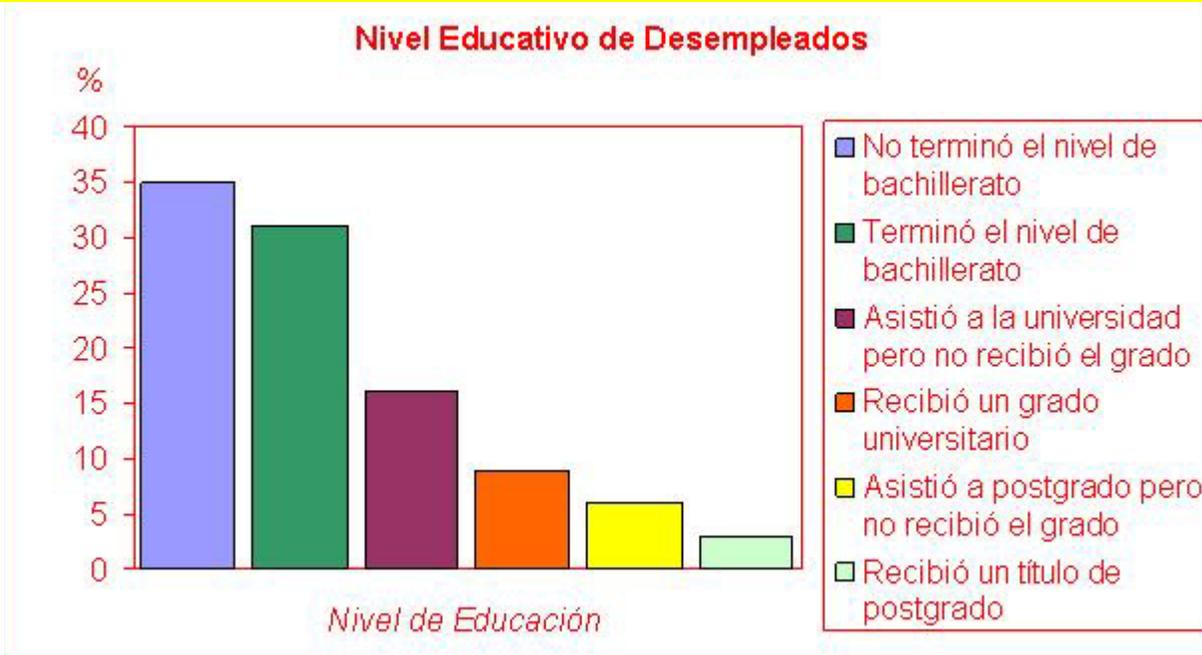
Responda lo siguiente:

- a. Diga de qué tipo son los datos.
- b. Represente gráficamente la información anterior. Interprete en términos del problema.
- c. Calcule e interprete las medidas de posición que usted considere adecuadas.

(2,5 puntos)

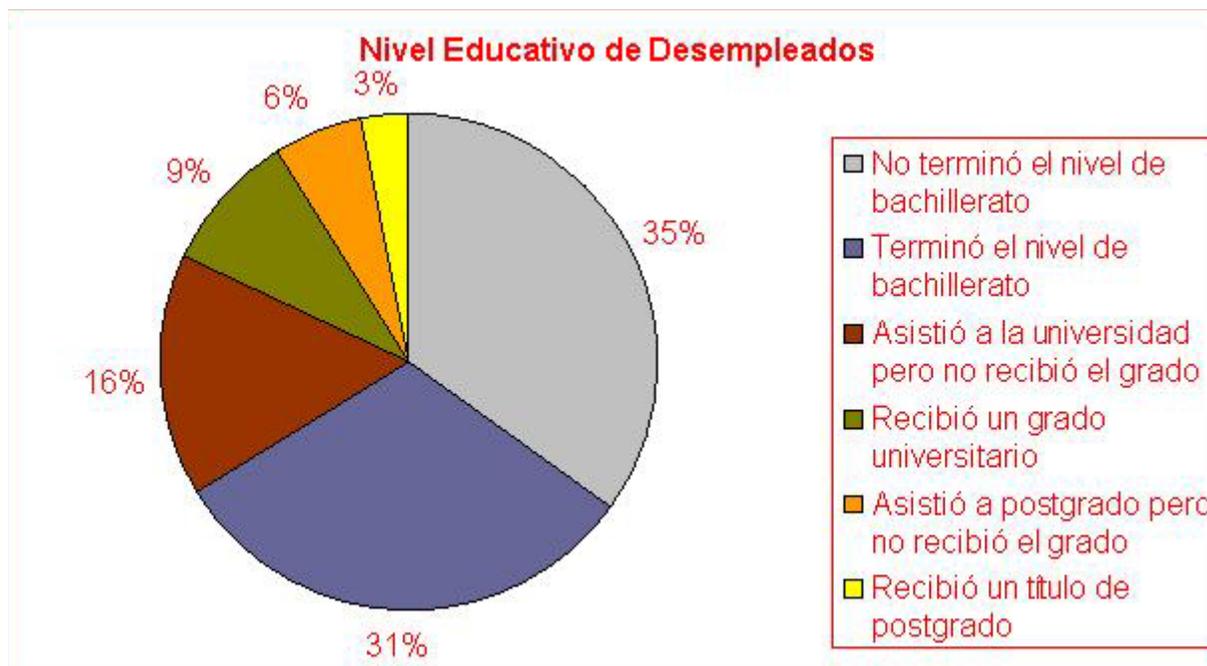
- a. Los datos son **CUALITATIVOS ORDINALES**
- b. A continuación se presenta un gráfico de barras para representar a la variable “Nivel educativo de desempleados”

SOLUCIÓN EXAMEN PARCIAL I



Fuente: Departamento de Trabajo de Estados Unidos

Otro gráfico que se puede usar para representar variable cualitativas ordinales es el de sectores:



Fuente: Departamento de Trabajo de Estados Unidos

En ambos gráficos se observa claramente que a medida que el nivel educativo es más alto la tasa de desempleo disminuye. El desempleo más alto es para las personas que no terminaron el bachillerato y es muy similar a la tasa de desempleo de los bachilleres.

- c. Como esta variable es cualitativa ordinal, las medidas de posición que se pueden calcular son la moda y la mediana. Para determinar la moda se encuentra en la tabla la modalidad con frecuencia más alta (en este caso porcentaje más alto). Por otra parte, para encontrar la mediana, se calcula la frecuencia relativa acumulada y la mediana es la modalidad que contiene 50% en la frecuencia relativa acumulada

Mo = No terminó el nivel de bachillerato. Es el nivel de educación más común de los desempleados.

SOLUCIÓN EXAMEN PARCIAL I

$M_d =$ Terminó el nivel de bachillerato. El 50% de los desempleados tiene un nivel educativo igual o menor a bachiller y el otro 50% un nivel un nivel educativo igual o mayor a bachiller.

Nivel de Educación	% de Desempleados	Fr _i *100 (%)
No terminó el nivel de bachillerato	35	35
Terminó el nivel de bachillerato	31	66
Asistió a la universidad pero no recibió el grado	16	82
Recibió un grado universitario	9	91
Asistió a postgrado pero no recibió el grado	6	97
Recibió un título de postgrado	3	100

2. De una encuesta realizada a los estudiantes de Métodos Estadísticos I (Sem. A-2005) se ha obtenido la siguiente información sobre sus estaturas (en metros), clasificadas por sexo:

Sexo	Media	Desviación Estándar
F	1,64	0,0525
M	1,75	0,0492

- a. ¿Quiénes presentan menor dispersión en sus estaturas, los varones o las hembras? Justifique su respuesta.
- b. En cual grupo la media es más representativa. Justifique su respuesta.
- c. Después de un examen de Métodos Estadísticos I todos los varones se fueron a celebrar y cenaron con hamburguesa. Estas hamburguesas estaban contaminadas con una bacteria desconocida, por la que todos los muchachos se enfermaron. Sorprendentemente después de seis meses todos los varones habían aumentado exactamente 2 cm en su estatura. ¿Cuál es ahora la media aritmética y la desviación estándar de los masculinos de Métodos Estadísticos I?

(3 puntos)

- a. Para determinar cuál grupo presenta la menor dispersión, se utiliza el coeficiente de variación de cada grupo de estudiantes:

$$CV_F = \frac{S}{\bar{X}} * 100 = \frac{0,0525}{1,64} * 100 = 3,2\%$$

$$CV_M = \frac{S}{\bar{X}} * 100 = \frac{0,0492}{1,75} * 100 = 2,8\%$$

Las estaturas de los estudiantes varones de Métodos Estadísticos I presentan menor variabilidad ($CV_M = 2,8\%$) que las estaturas de las hembras ($CV_F = 3,2\%$).

- b. La media es más representativa en el grupo de los varones ya que presenta menor dispersión con respecto a su media ($CV_M = 2,8\%$ y $CV_F = 3,2\%$) que el grupo de las hembras.
- c. Como todos los varones aumentaron exactamente 2 cm en su estatura, es equivalente a decir que todas las estaturas originales (x_1, x_2, \dots, x_{41}) se modifican sumando a cada una la constante 2: $x_1 + 2, x_2 + 2, \dots, x_{41} + 2$. De esta manera por propiedades de la media, la media aritmética de las estaturas de los varones después que aumentaron 2 cm es : $\bar{x}_{nueva} = 1,75 + 0,02 = 1,77$. Y la desviación estándar, según propiedad, es la misma: $S_{nueva} = 0,0492$

3. Pruebe que
$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

(1,5 puntos)

SOLUCIÓN EXAMEN PARCIAL I

Prueba:

$$\begin{aligned}
 S^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} = \frac{\sum_{i=1}^n x_i^2 - 2\bar{x}\sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2}{n} = \frac{\sum_{i=1}^n x_i^2 - 2\bar{x}\frac{n}{n}\sum_{i=1}^n x_i + n\bar{x}^2}{n} \\
 &= \frac{\sum_{i=1}^n x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2}{n} = \frac{\sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2}{n} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n} = \boxed{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2}
 \end{aligned}$$

4. En la siguiente tabla se presenta la distribución de frecuencias correspondiente al número de hermanos de los estudiantes de Métodos Estadísticos I (Sem A-2005)

Clases	m_i	f_i	fr_i	F_i	Fr_i
[0 - 2)	1	12	0,2927	12	0,2927
[2 - 4)	3	19	0,4634	31	0,7561
[4 - 6)	5	6	0,1463	37	0,9024
[6 - 8)	7	1	0,0244	38	0,9268
[8 - 10)	9	1	0,0244	39	0,9512
[10 - 12)	11	2	0,0488	41	1
	41	1			

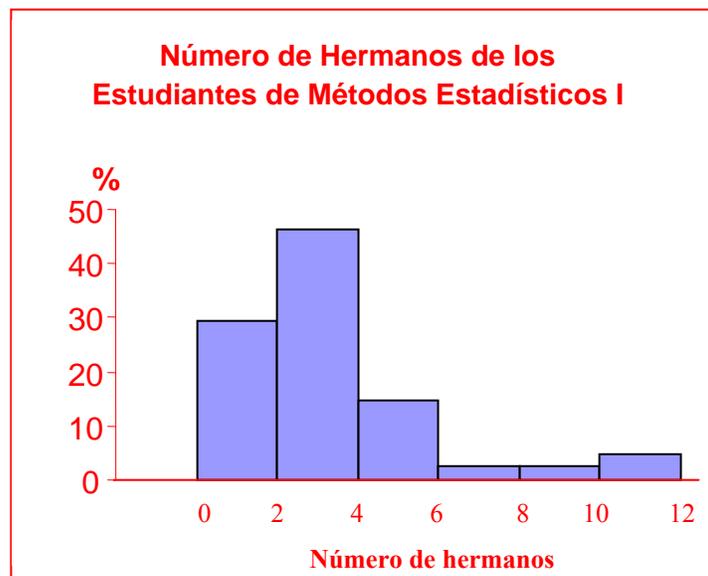
Fuente: Encuesta levantada en Enero 2005

En relación a esta distribución de frecuencias, responda lo siguiente:

- Interprete fr_2 y F_4 en términos del problema.
- Construya el gráfico que considere adecuado para estos datos. Interprete en términos del problema.
- Calcule e interprete:
 - La media aritmética
 - La mediana
 - La desviación estándar
 - El primer cuartil
- ¿Qué porcentaje de estudiantes tiene menos de cinco hermanos?
- ¿El 20% de los estudiantes con mayor número hermanos, tiene más de cuántos hermanos?

(8 puntos)

- $fr_2=0,4634$: El 46,34% de los estudiantes de Métodos Estadísticos I tiene entre 2 y 3 hermanos; ó 0,4634 es la proporción de estudiantes de Métodos Estadísticos I que tiene entre 2 y 3 hermanos.
 $F_4=38$: Treinta y ocho estudiantes de Métodos Estadísticos I tiene menos de 8 hermanos (es decir, 7 o menos hermanos)
- El gráfico adecuado los datos agrupados en una distribución de frecuencias cuyas clases son intervalos es el histograma o su equivalente que es el polígono de frecuencias. A continuación se presenta el histograma respectivo.



SOLUCIÓN EXAMEN PARCIAL I

Fuente: Encuesta levantada en Enero 2005

En este gráfico se observa que la distribución del número de hermanos es *asimétrica por la derecha*. Se puede apreciar que la mayoría de los estudiantes tiene menos de 4 hermanos. También se nota que muy pocos estudiantes poseen más de seis hermanos.

Se puede llegar a estas mismas conclusiones observando el polígono de frecuencias que se presenta a continuación.



Fuente: Encuesta levantada en Enero 2005

c.

Clases	m_i	f_i	fr_i	F_i	Fr_i	$m_i * f_i$	$m_i^2 * f_i$	
[0-2)	1	12	0,2927	12	0,2927	12	12	→Clase del Q_1
[2-4)	3	19	0,4634	31	0,7561	57	171	→Clase mediana
[4-6)	5	6	0,1463	37	0,9024	30	150	→Clase del P_{80}
[6-8)	7	1	0,0244	38	0,9268	7	49	
[8-10)	9	1	0,0244	39	0,9512	9	81	
[10-12)	11	2	0,0488	41	1	22	242	
		41	1			137	705	

i.
$$\bar{x} = \frac{\sum_{i=1}^k m_i f_i}{n} = \frac{137}{41} \Rightarrow \boxed{\bar{x} = 3,34 \text{ hermanos}}$$

El promedio de hermanos en los estudiantes de M.E.I. es de 3,34 hermanos; este valor representa el punto de equilibrio o centro de gravedad de la distribución del número de hermanos.

ii.
$$Md = LI_m + \left[\left(\frac{n}{2} - Fa_m \right) / fm \right] * C_m$$

SOLUCIÓN EXAMEN PARCIAL I

$$Md = LI_m + \left[\left(\frac{n}{2} - Fa_m \right) / fm \right] * C_m$$

$$Md = 2 + \left[(20,5 - 12) / 19 \right] * 2 \Rightarrow \boxed{Md=2,89 \text{ hermanos}}$$

Alrededor de la mitad de los estudiantes de M.E.I tiene menos de 3 hermanos aproximadamente ($Md=2,89 \approx 3$ hermanos) y la otra mitad de los estudiantes tiene 3 o más hermanos.

iii.
$$S = \sqrt{\sum_{i=1}^n \frac{m_i^2 * f_i}{n} - \bar{x}^2}$$

$$S = \sqrt{\frac{705}{41} - (3,34)^2} = \sqrt{6,04} \Rightarrow \boxed{S = 2,46 \text{ hermanos}}$$

El grado de dispersión de la distribución del número de hermanos de los estudiantes de M.E.I con respecto a su media ($\bar{x} = 3,34$ hermanos) es de 2,46 hermanos.

iv.
$$Q_1 = P_{25} = LI_p + \left\{ \left[n * \left(\frac{25}{100} \right) - F_{ap} \right] / f_p \right\} * C_p$$

$$Q_1 = P_{25} = 0 + \left\{ \left[41 * \left(\frac{25}{100} \right) - 0 \right] / 12 \right\} * 2 \Rightarrow \boxed{Q_1 = 1,71 \text{ hermanos}}$$

d. El porcentaje de estudiantes que tiene menos de 5 hermanos se puede calcular despejando h de la fórmula del percentil P_h :

$$P_h = LI_p + \left\{ \left[n * \left(\frac{h}{100} \right) - F_{ap} \right] / f_p \right\} * C_p$$

$$\Rightarrow h = \left[\left(\frac{P_h - LI_p}{C_p} \right) * f_p + F_{ap} \right] * \frac{100}{n}$$

$$h = \left[\left(\frac{5 - 4}{2} \right) * 6 + 31 \right] * \frac{100}{41} \Rightarrow \boxed{h = 82,93 \%}$$

Entonces 82,93% de los estudiantes de M.E.I tiene menos de cinco hermanos.

e. Se debe calcular P_{80} :

$$P_{80} = 4 + \left\{ \left[41 * \left(\frac{80}{100} \right) - 31 \right] / 6 \right\} * 2 \Rightarrow \boxed{P_{80} = 4,6 \text{ hermanos}}$$

Así, el 20% de los estudiantes que tienen un número mayor de hermanos, tiene 5 o más hermanos ($P_{80} = 4,6 \approx 5$ hermanos).

También se puede decir que el 80% de los estudiantes tiene menos de 5 hermanos.

FORMULARIO:

$$\bar{x} = \frac{\sum_{i=1}^k m_i * f_i}{n}; \quad \bar{x} = \left(\frac{\sum_{i=1}^n x_i * w_i}{\sum_{i=1}^n w_i} \right); \quad S^2 = \sum_{i=1}^n \frac{m_i^2 * f_i}{n} - \bar{x}^2; \quad S^2 = \sum_{i=1}^n \frac{(m_i - \bar{x})^2 * f_i}{n}; \quad CV = \frac{S}{\bar{x}} * 100; \quad Q = \frac{Q_3 - Q_1}{2};$$

$$RQ = Q_3 - Q_1; \quad Md = LI_m + \left[\left(\frac{n}{2} - Fa_m \right) / fm \right] * C_m; \quad P_h = LI_p + \left\{ \left[n * \left(\frac{h}{100} \right) - F_{ap} \right] / f_p \right\} * C_p; \quad \frac{\overline{AR}}{AB} = \frac{\overline{RS}}{BC}; \quad ASP = 3(\bar{x} - Md) / S_*$$

$$\gamma_1 = \left[\frac{\sum_{i=1}^k (m_i - \bar{x})^3 f_i / n}{S_*^3} \right]; \quad \gamma_1 = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^3 f_i / n}{S_*^3} \right]; \quad \beta_2 = \left[\frac{\sum_{i=1}^k (m_i - \bar{x})^4 f_i / n}{S_*^4} \right]; \quad \beta_2 = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^4 f_i / n}{S_*^4} \right];$$

$$C_i = R / K; \quad K = 1 + 3,3 * \text{Log}(n);$$