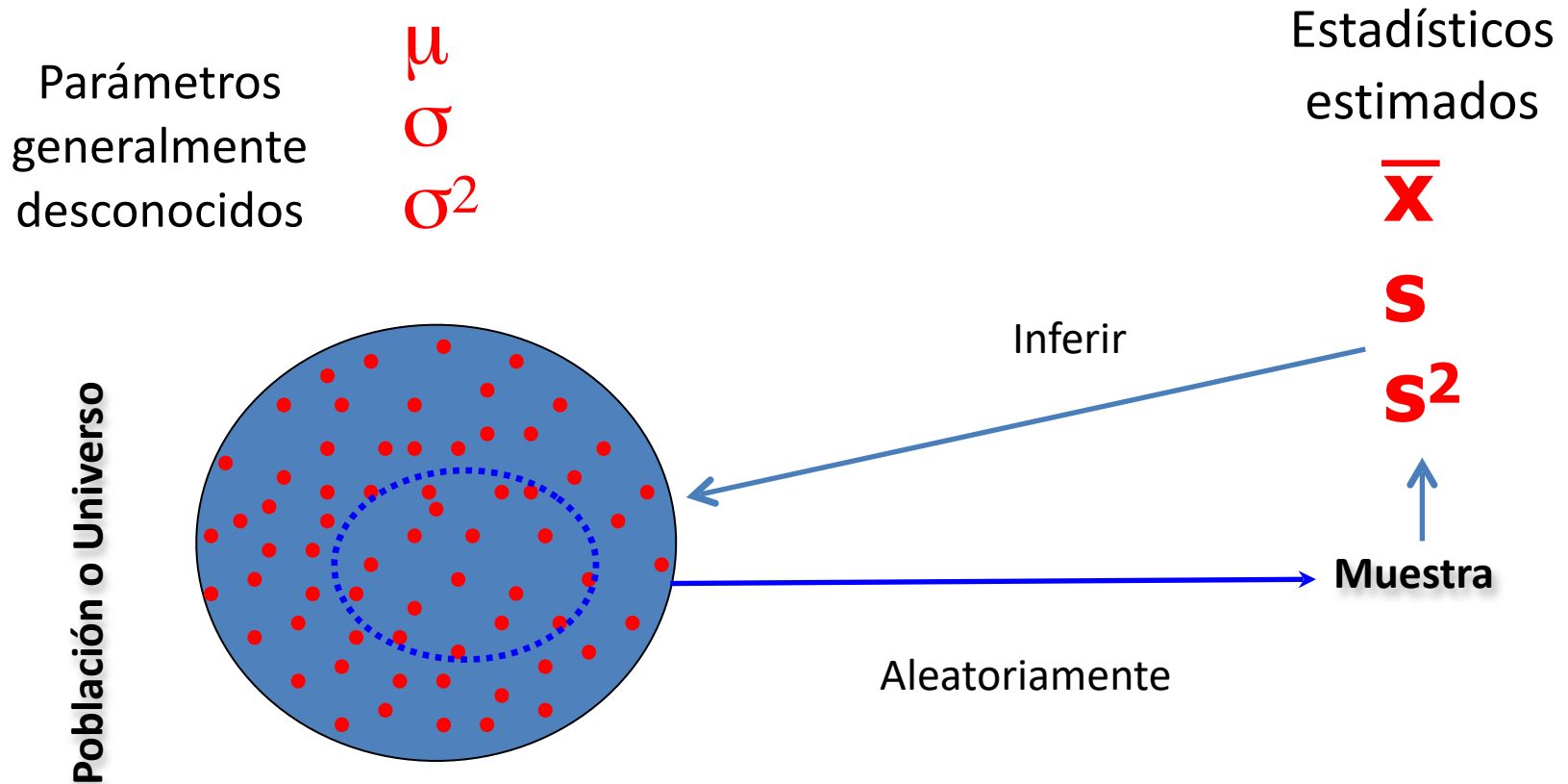


Estadística Descriptiva



Instructor: Dr. Porfirio Gutiérrez González
Correo: pgutierrezglez@gmail.com

Población, Muestra, Parámetros y Estadísticos



Medidas de tendencia central y variabilidad

Para analizar el comportamiento o distribución de un conjunto de datos, comúnmente se inicia calculando los estadístico básicos.

Algunos estadísticos que se pueden calcular son:

- **Media, Mediana, Moda (medidas de tendencia central)**
- **Desviación Estándar, Rango, Coeficiente de Variación (medidas de variabilidad)**



Medidas de tendencia central



Media aritmética o promedio

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} = \frac{\sum_{j=1}^n X_j}{n}$$

Mediana

Es un conjunto de números ordenados en orden de magnitud ascendente, es decir de menor a mayor; el dato que ocupa la posición central corresponde a la mediana.

Moda

En un conjunto de números es el valor que ocurre con mayor frecuencia, es decir, es el valor mas frecuente.

La moda puede no existir en la distribución e incluso puede tener 2 o más.

En el caso de una moda la distribución es unimodal; cuando existen dos modas es bimodal; tres modas, trimodal; y así sucesivamente.

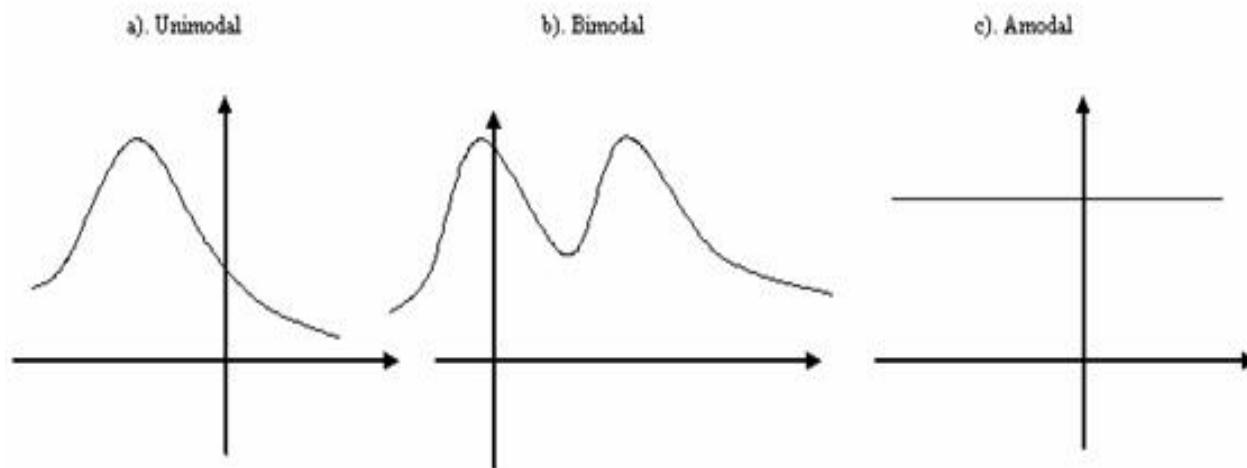
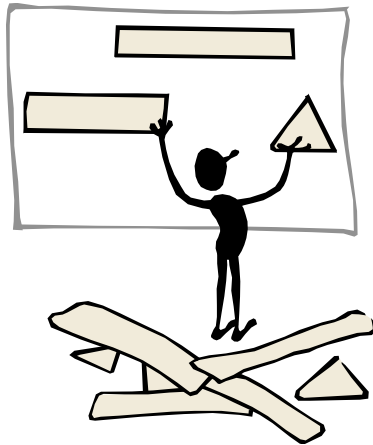


Gráfico de 3 distribuciones diferentes: a). unimodal, b). Bimodal y c). Amodal



Medidas de dispersión o variabilidad

La dispersión o variabilidad de los datos intenta dar una idea de qué tan esparcidos se encuentran los datos en una distribución.



Rango o amplitud

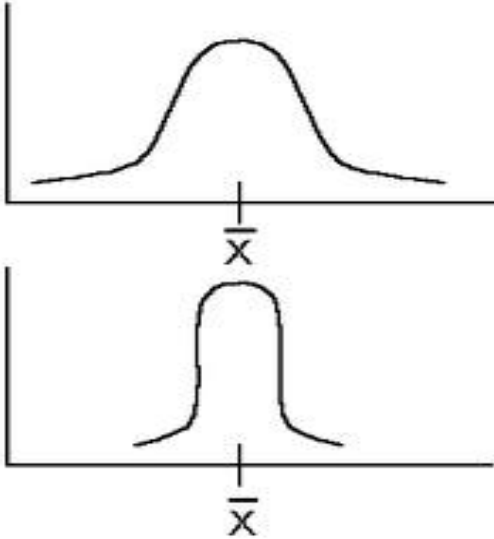
En un conjunto de datos el rango se define como la diferencia existente entre el valor máximo y el valor mínimo del conjunto de datos.

2,4,3,5,4,3,5,7,6,2,4,5,7,4

$$\text{Rango} = R = \text{VALOR}_{\max} - \text{VALOR}_{\min}$$

$$R = 7 - 2$$

$$R = 5$$



Desviación estándar

La desviación estándar representa las desviaciones de cada uno de los números obtenidos con respecto a su media aritmética, dividido entre el total de datos menos 1. Se obtiene de la siguiente manera:

Varianza

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

Desviación Estándar

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

Problema

Un producto debe tener un % vol. de alcohol de 40%, con una tolerancia de $\pm 5\%$. De los muestreos para evaluar la calidad se obtienen los siguientes datos:

41.77	39.28	40.31	34.03	38.89	42.70
39.36	38.83	39.02	35.43	41.81	44.65
39.67	42.12	45.22	42.23	38.80	39.57
40.47	39.52	40.39	38.37	37.26	40.75
42.83	41.66	42.94	38.67	42.69	40.56
37.49	43.59	38.08	39.20	42.07	42.16
39.70	40.38	41.47	41.84	39.48	37.98
39.14	41.03	37.68	41.66	40.68	40.67
41.75	39.81	42.71	39.83	38.17	41.89
41.86	41.77	38.82	40.77	40.10	37.67

Resultados estadísticos

Recuento	60
Promedio	40.3208
Mediana	40.385
Moda	
Varianza	4.29863
Desviación Estándar	2.07331
Coefficiente de Variación	5.14%
Mínimo	34.03
Máximo	45.22
Rango	11.19
Cuartil Inferior	38.955
Cuartil Superior	41.825
Rango Inter cuartílico	2.87



El promedio de % Volumen es 40.321, con esto puedo afirmar que, si se evalúan a otros 60 .

¿Se esperaría que el promedio fuera de 40.321?

¿Se esperaría que la desviación estándar fuera de 2.07?

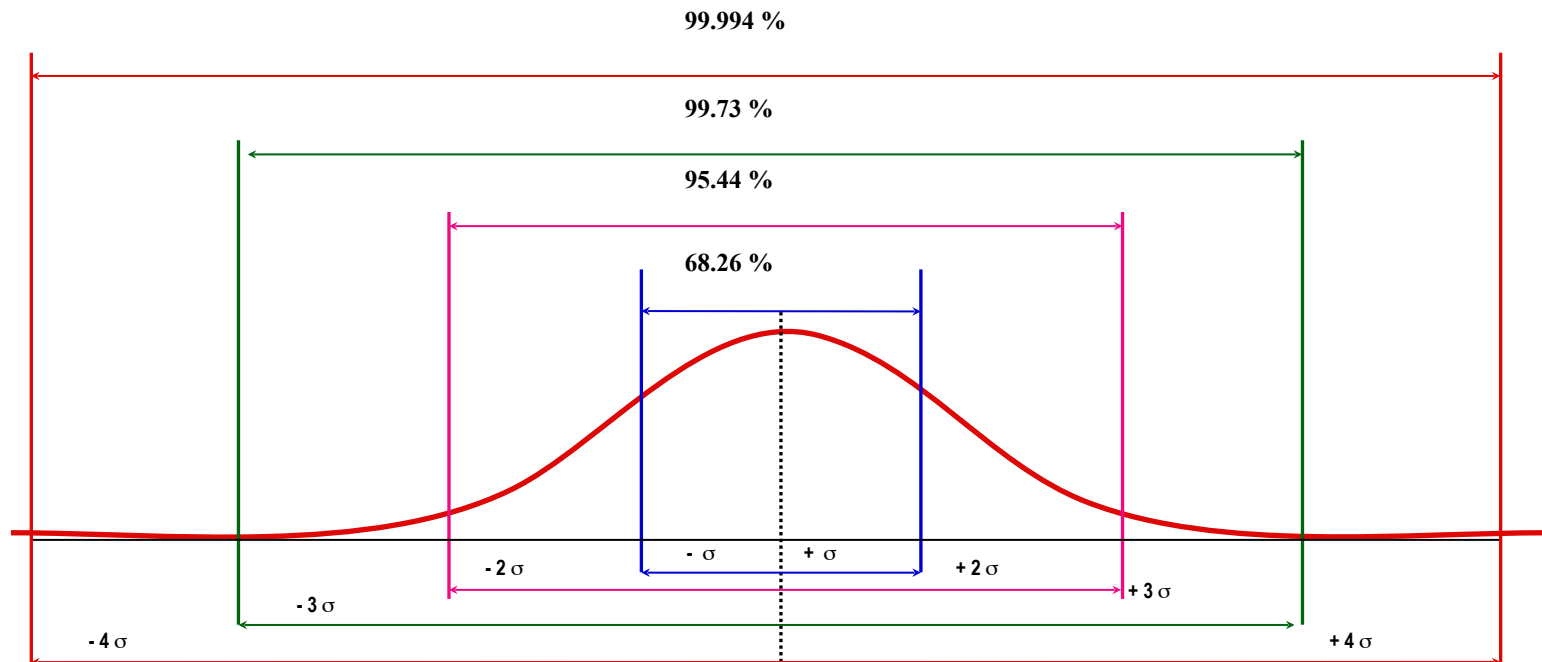




Regla empírica.

Muchos de los datos que surgen en la práctica se ha observado.

- $\bar{X} - S$ y $\bar{X} + S$ está el 68%
- $\bar{X} \pm 2S$ está el 95%
- $\bar{X} \pm 3S$ el 99.7%



- Regla empírica



$$\bar{X} \pm 3S \text{ el } 99.7\%$$

$$\bar{X}-3(S)=40.32-3*2.07=34.11$$

$$\bar{X}+3(S)=40.32+3*2.07=46.53$$

Intervalo [34.11, 46.53]

**El 99.7% de la producción del producto tienen %
volumen de 34.11 a 46.53**

Histograma para el % de volumen de alcohol

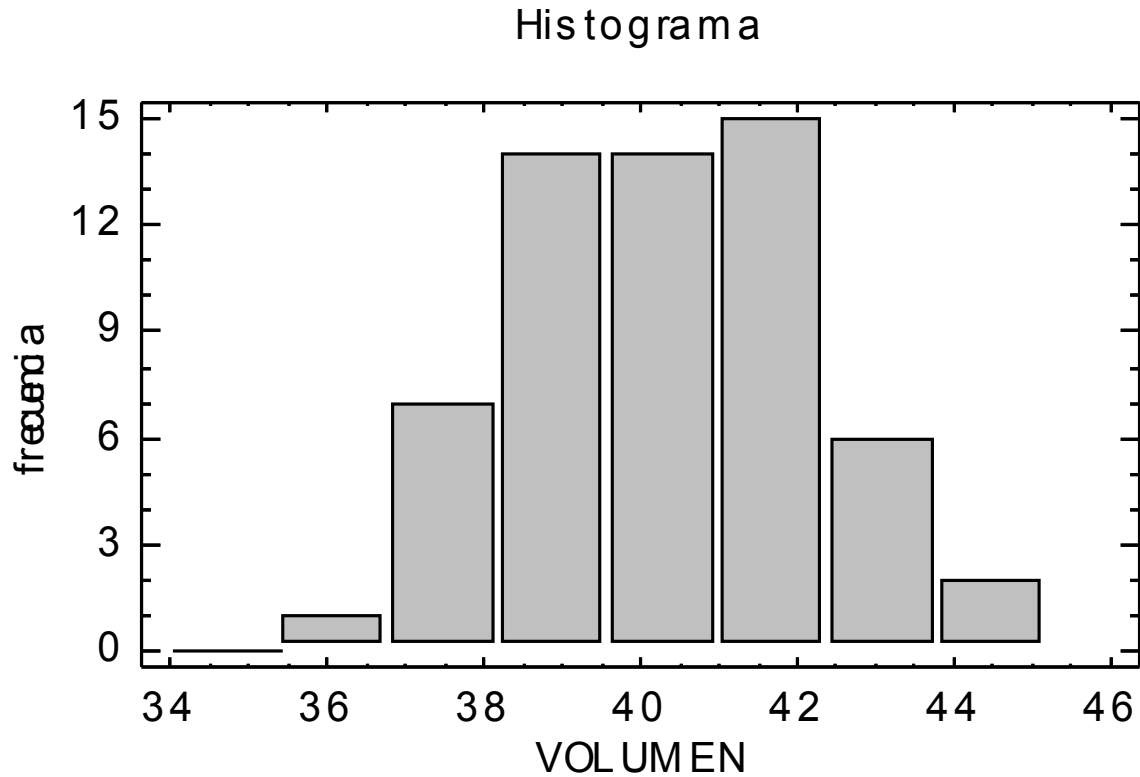
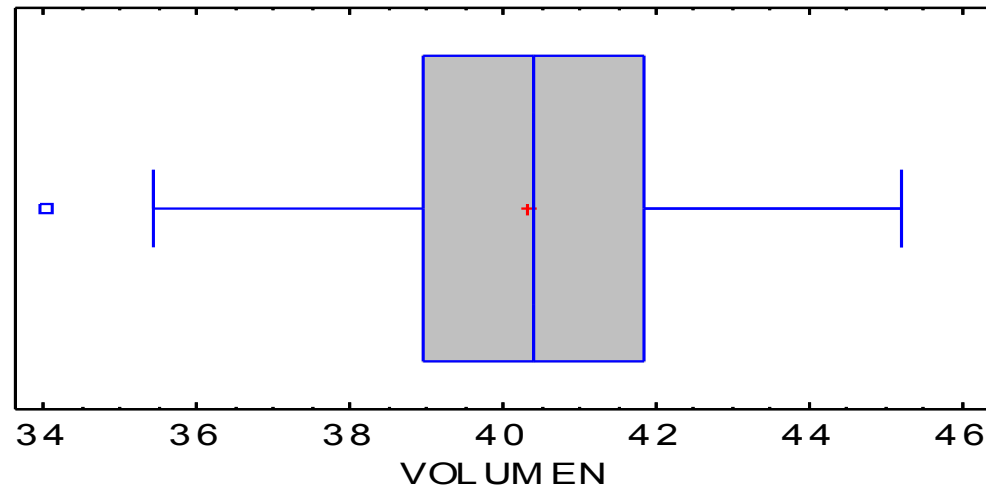


Diagrama de Caja para el % Volumen de alcohol

Gráfico de Caja y Bigotes



Mínimo	34.03
Máximo	45.22
Cuartil Inferior	38.955
Cuartil Superior	41.825
Rango Intercuartílico	2.87