Curso de Estadística Básica

SESION 5
TEOREMA DE CHEBYSHEV, REGLA EMPÍRICA Y CASO

MCC. Manuel Uribe Saldaña
MCC. José Gonzalo Lugo Pérez

Objetivo

Comprender y aplicar el teorema de Chebyshev y la regla empírica para una distribución de datos. Aplicar los conocimientos adquiridos en las cuatro sesiones anteriores en la solución de un caso.

Agenda Sesión 4

- Teorema de Chebyshev
- Regla Empírica
- Ejercicios
- Caso

Teorema de Chebyshev

La proporción de cualquier distribución que esté a menos de *k* desviaciones estándar de la media es por lo menos

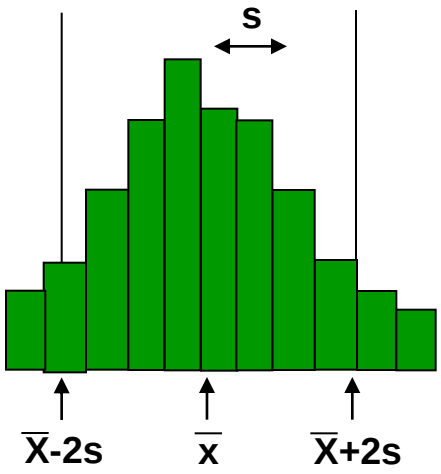
$$1-\frac{1}{k^2}$$

Donde *k* es cualquier número positivo mayor que 1. Este teorema es válido para todas las distribuciones de datos.

Teorema de Chebyshev

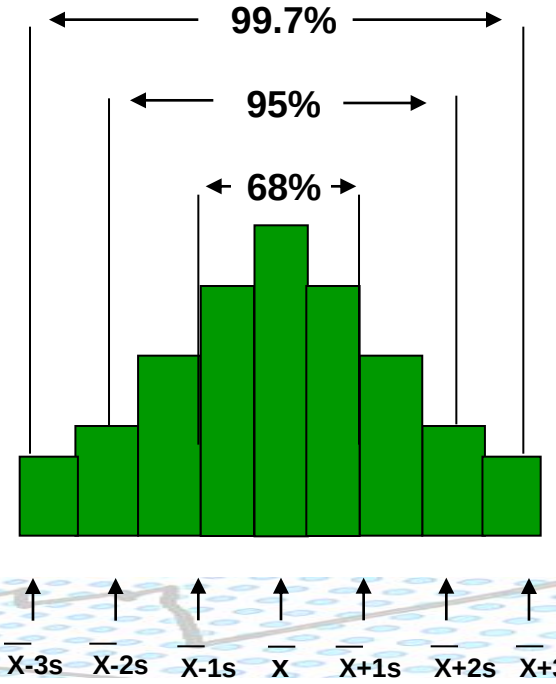
Este teorema establece que a menos de dos desviaciones estándar de la media (k = 2) siempre se encontrará por lo menos el 75% (o más) de los datos.

Por lo menos el 75%



Regla Empírica

Si una variable está distribuida normalmente, entonces: a menos de una desviación estándar de la media hay aproximadamente 68% de los datos; a menos de dos desviaciones estándar de la media hay aproximadamente 95% de los datos; y a menos de tres desviaciones estándar de la media hay aproximadamente 99.7% de los datos. Esta regla es válida específicamente para una distribución normal.



Ejercicio

- Se tiene el siguiente conjunto de datos:

Datos ordenados: puntajes del examen de estadística				
39	64	72	78	89
44	66	72	80	90
47	67	74	82	90
50	68	74	82	91
55	68	75	83	92
58	70	76	85	94
58	70	77	86	95
60	70	77	86	95
63	72	77	88	97
64	72	78	88	98

- Determinar si los datos están o no, aproximadamente, distribuidos normalmente.

Pasos

1. Obtener un diagrama de frecuencias agrupadas con Marca de Clase
2. Obtener la Media y la Desviación Estándar
3. Calcular los intervalos que van de una desviación estándar por debajo y por arriba de la media
4. Observar en la tabla los elementos que lo conforman
5. Obtener el porcentaje de los datos contenidos en el intervalo.
6. Repetir los pasos 3, 4 y 5 para dos desviaciones estándar y tres desviaciones estándar
7. Construir una tabla que incluya el porcentaje de la regla empírica y el porcentaje encontrado
8. Concluir

Ejercicios

El tiempo promedio de limpieza para un equipo de una empresa de tamaño mediano es de 84 horas y la desviación estándar es de 6.8 horas. En el supuesto de que la regla empírica es idónea, conteste:

a) ¿En qué proporción del tiempo el equipo limpia la planta en 97.6 o más horas?

b) ¿En qué intervalo caerá el tiempo total de limpieza 95% de las veces?

Ejercicios

A 60 estudiantes universitarios de nuevo ingreso se les solicitó indicar el número de niños que hay en su familia (número de hermanos y hermanas más 1). Los datos recolectados fueron los siguientes:

1	6	3	5	5	3	4	1	2	7	3	2
3	4	5	3	1	3	2	1	4	4	2	2
3	9	4	3	3	5	3	5	7	3	1	1
3	5	2	6	4	3	3	3	3	3	2	3
4	3	5	7	3	2	1	2	3	2	4	3

CONTINUA →

Ejercicios

a) Elabore una distribución de frecuencias agrupadas de estos datos.

b) Use la distribución de frecuencias agrupadas del inciso a) para encontrar la media y la desviación estándar de estos datos.

c) Encuentre los valores de –s y +s con respecto a la media

d) ¿Cuántos de los 60 datos tienen valores en este intervalo?¿Qué porcentaje de la muestra es esto?

e) Encuentre los valores de -2s y +2s con respecto a la media

f) ¿Cuántos de los 60 datos tienen valores en este intervalo?¿Qué porcentaje de la muestra es esto?

g) Encuentre los valores de -3s y +3s con respecto a la media

h) ¿Qué porcentaje de la muestra tiene valores en este intervalo?

i) Compare las respuestas encontradas en los incisos f) y h) con los resultados predichos por el teorema de Chebyshev

j) Compare las respuestas encontradas en los incisos d), f) y h) con los resultados predichos por la regla empírica. ¿Los resultados sugieren una distribución aproximadamente normal?

Caso

En una empresa se han registrado los retardos de los trabajadores en sus diferentes jornadas a través del checador de Recursos Humanos. Se emitió un reporte con estos tiempos, siendo la tolerancia de 1 a 15 minutos y se obtuvieron los siguientes datos:

6	6	7	8	6	7	12	6	6	10
10	6	7	5	5	4	6	5	12	9
9	10	6	9	10	12	9	4	2	5
10	7	7	3	10	7	7	9	6	4
13	7	7	7	11	8	8	11	4	4
5	8	6	6	8	4	7	11	2	6
5	8	3	5	8	8	8	8	12	9
8	7	6	10	8	2	6	8	10	11
8	12	4	8	5	7	7	9	11	7
9	7	8	10	5	10	8	7	10	7

Caso

Calcular:

- Una tabla de frecuencias no agrupadas
- Un histograma
- Obtener la media, varianza y desviación estándar de la tabla de frecuencias
- Obtener la mediana, Q1 y Q3
- Obtener la moda
- Encuentre los valores –s y +s con respecto a la media
- ¿Qué porcentaje de los 100 datos se encuentra en este intervalo?
- Encuentre los valores -2s y +2s con respecto a la media
- ¿Qué porcentaje de los 100 datos se encuentra en este intervalo?
- Encuentre los valores -3s y +3s con respecto a la media
- ¿Qué porcentaje de los 100 datos se encuentra en este intervalo?
- Compare los resultados de los incisos g), i) y k) con el teorema de Chebyshev en una tabla
- Compare los resultados de los incisos g), i) y k) con la regla empírica en una tabla
- ¿Los resultados sugieren una distribución aproximadamente normal?
- ¿En qué proporción del tiempo los trabajadores llegan con 10 minutos de retardo o más?