

INFORMÁTICA APLICADA A LA ESTADÍSTICA

SEMANA 3.

Temas	Enlaces Recomendados
Medidas de Dispersión	http://www.slideshare.net/moibemo/tema-3-medidas-de-dispersin http://www.slideshare.net/profe_simmons/estadistica-i-05-presentation
Rango	http://www.youtube.com/watch?v=nInYWSqNsHw http://office.microsoft.com/es-hn/excel-help/calcular-el-numero-mayor-o-menor-de-un-rango-HP010070506.aspx
Desviación Media	
Varianza	http://www.slideshare.net/mmujica/varianza
Desviación Estándar	http://www.fisterra.com/mbe/investiga/10descriptiva/10descriptiva2.pdf

MEDIDAS DE DISPERSIÓN

Se llama dispersión de un conjunto de datos al grado en que los diferentes valores numéricos de los datos tienden a extenderse alrededor del valor medio utilizado. Este grado de dispersión se mide por medio de los indicadores estadísticos llamados medidas de dispersión, entre ellas tenemos el rango, la varianza y la desviación típica.

RANGO

También se conoce como amplitud total o recorrido, se llama rango de una distribución a la diferencia entre el mayor y el menor valor de la variable estadística. Es muy sencillo aplicando la definición, consiste en ordenar los valores de menor a mayor y restar al último el primero.

Observaciones al recorrido:

1. Cuanto menor es el recorrido mayor es el grado de representatividad de los valores centrales.
2. Cuanto mayor es, la distribución está menos concentrada o más dispersa.
3. Tiene la gran ventaja de su sencillez de cálculo.
4. Tiene gran aplicación en procesos de control de calidad,
5. Tiene el inconveniente de que sólo depende de los valores extremos. De esta forma basta que uno de ellos se separe mucho para que el recorrido se vea sensiblemente afectado.

Comparemos, por ejemplo, estas dos series:

Serie 1: 1 5 7 7 8 9 9 10 17

Serie 2: 2 4 6 8 10 12 14 16 18

Ambas series tienen rango 16, pero están desigualmente agrupadas, pues mientras la primera tiene una mayor concentración en el centro, la segunda se distribuye uniformemente a lo largo de todo el recorrido. El uso de esta medida de dispersión, será pues, bastante restringido.

DESVIACIÓN MEDIA

Para conocer con un solo indicador que tan disperso se encuentran un conjunto de datos a un punto de concentración, debemos como primera medida, calcular la distancia de cada dato respecto a una medida de tendencia central. Por ejemplo:

4	5	3
5	3	2
2	2	2
3	5	1
4	1	4

Tenemos que la media aritmética es de aproximadamente 3,0667. El primer dato (4), se aleja de la media en 0,9333 hacia la derecha. Para el segundo dato (5) la distancia es de 1,9333 respecto a la media aritmética:

Note que el tercer dato (3) posee una distancia de 0,0667 hacia la izquierda de la media. Para indicar las distancias de estos puntos, agregaremos el signo negativo, por tanto, la distancia del tercer dato sería $-0,0667$. La representación gráfica de todos los puntos quedaría:

Es por ello que la Desviación media (D_m): Equivale a la división de la sumatoria del valor absoluto de las distancias existentes entre cada dato y su media aritmética y el número total de datos.

Por ejemplo: Tres alumnos son sometidos a una competencia para probar sus conocimientos en 10 materias diferentes, cada una sustentada con 10 preguntas. La idea del concurso es encontrar al alumno más idóneo para representar al colegio en un torneo a nivel nacional.

El número de preguntas buenas por materia se muestra a continuación:

Materia	Carlos	Pedro	Juan
1	2	7	5
2	9	2	6
3	10	2	5
4	2	6	5
5	3	6	5
6	1	3	5
7	9	6	4
8	9	7	5
9	1	6	6
10	4	5	4

Lo primero que analizaremos es la media de los puntajes para cada uno de los alumnos, con el fin de determinar el alumno con mayor promedio de preguntas buenas.

Las medias para los resultados de los alumnos coinciden: los tres alumnos tienen responden en promedio 5 preguntas correctas por prueba. Complementemos el análisis anterior calculando la desviación media:

Carlos muestra una desviación media de 3,9 indicando que los datos se alejan en promedio de la media en 3,9 preguntas buenas. Pedro disminuye su variación (2,9), siendo Juan el que menos variación presenta con 0,9 preguntas tanto por arriba como por debajo de la media aritmética. Se recomienda al colegio elegir como ganador en este caso a Juan, presenta resultados más constantes que los otros dos alumnos, Juan en promedio acierta 5 preguntas buenas con una variación muy baja (rondando entre 4 y 6).

VARIANZA

Una forma natural de medir la dispersión en torno a la media es calcular la media de las diferencias, pero como habrá valores por encima y por debajo de la media que se compensarán, calcularemos mejor el cuadrado de las diferencias. Se define así varianza de una variable estadística, como la media de los cuadrados de las desviaciones de sus valores respecto a su media.

Propiedades de la varianza

1. La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.
2. Si a todos los valores de la variable se les suma un número la varianza no varía.
3. Si todos los valores de la variable se multiplican por un número la varianza queda multiplicada por el cuadrado de dicho número.
4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas varianzas se puede calcular la varianza total.

Ejemplo: La siguiente muestra representa las edades de 25 personas sometidas a un análisis de preferencias para un estudio de mercado.

25	19	21	35	44	20
27	32	38	33	18	30
19	29	33	26	24	28
39	31	31	18	17	30
27					

Lo primero que se debe hacer es calcular la media aritmética, para de esta manera calcular la varianza. La varianza equivale a 51,8567. Por elevar las unidades al cuadrado, carece de un significado contextual dentro del análisis descriptivo del caso.

DESVIACIÓN ESTÁNDAR (TÍPICA)

Esta medida nos permite determinar el promedio aritmético de fluctuación de los datos respecto a su punto central o media. La desviación estándar nos da como resultado un valor numérico que representa el promedio de diferencia que hay entre los datos y la media. Para calcular la desviación estándar basta con hallar la raíz cuadrada de la varianza,

Propiedades de la desviación típica

1. La desviación típica será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.
2. Si a todos los valores de la variable se les suma un número la desviación típica no varía.
3. Si todos los valores de la variable se multiplican por un número la desviación típica queda multiplicada por dicho número.
4. Si tenemos varias distribuciones con la misma media y conocemos sus respectivas desviaciones típicas se puede calcular la desviación típica total.

NETGRAFÍA

- Medidas de dispersión, disponible en:
<http://www.mitecnologico.com/Main/MedidasDeDispersion>
- Rango, disponible en:
<http://thales.cica.es/rd/Recursos/rd99/ed99-0122-04/rango.html>
- Desviación Media, disponible en:
<http://www.eumed.net/libros/2007a/239/5a.htm>
- Desviación Típica, disponible en:
<http://www.eumed.net/libros/2007a/239/5c.htm>
http://www.vitutor.com/estadistica/descriptiva/a_16.html
- Varianza, disponible en:
<http://www.innovanet.com.ar/gis/TELEDETE/TELEDETE/bmatyest.htm>
http://www.vitutor.com/estadistica/descriptiva/a_15.html