Para conocer mejor la distribución gaussiana, vamos a dejar a un lado las notas obtenidas en el examen y vamos a concentrarnos en las críticas de películas.

Estas son las opiniones (calificadas de 0 a 5) obtenidas por una película, donde 5 es la mejor nota que puede obtener la película: las famosas 5 estrellas que podemos encontrar en todos los sitios de críticas de cine.

| Opinión | Cantidad de votantes |
| --- | --- |
| 5 | 42 |
| 4 | 96 |
| 3 | 132 |
| 2 | 124 |
| 1 | 88 |
| 0 | 58 |

Si hacemos una representación gráfica de estos datos, obtenemos una forma particular: una campana.

Curva de Gauss

Ante este tipo de gráfico, podemos afirmar que la serie de observaciones sigue una ley matemática llamada ley normal o ley de Gauss (en honor a Karl Friederich Gauss (1777-1855)).

En estadística y en probabilidad, la ley normal permite representar muchos fenómenos aleatorios naturales. Cuando una serie de observaciones obedece a la ley normal, se puede afirmar:

* El 50 % de las observaciones están por encima de la media.
* El 50 % de las observaciones están por debajo de la media.
* El 68 % de las observaciones están comprendidas en el intervalo que va desde la media - la desviación típica hasta la media + la desviación típica.
* El 95 % de las observaciones están comprendidas en el intervalo que va desde la media - 2\* la desviación típica hasta la media + 2\* la desviación típica.
* El 99,7 % de las observaciones están comprendidas en el intervalo que va desde la media - 3\* la desviación típica hasta la media + 3\* la desviación típica.

Ahora vamos a hacer algunos cálculos que al mismo tiempo nos permitirán ver cómo utilizar la idea de frecuencia en los cálculos de media y de desviación típica.

| Opinión (Xi) | Cantidad de votantes (Ni) |
| --- | --- |
| 5 | 40 |
| 4 | 99 |
| 3 | 145 |
| 2 | 133 |
| 1 | 96 |
| 0 | 40 |

Las opiniones corresponden a nuestros valores observados denominados Xi, y la cantidad de votantes se equipara a la cantidad de veces en que el valor observado ha sido elegido por los espectadores. Entonces hablamos de frecuencia de elección, que se denomina Ni.

A fin de calcular la media de esta serie de observaciones, para cada observación hay que realizar el producto de las opiniones por la cantidad de votantes:

| Opinión (Xi) | Cantidad de votantes  (Ni) | Productos  (Ni \* Xi) |
| --- | --- | --- |
| 5 | 40 | 200 |
| 4 | 99 | 396 |
| 3 | 145 | 435 |
| 2 | 133 | 266 |
| 1 | 96 | 96 |
| 0 | 40 | 0 |

Luego hay que sumar los productos:

200 + 396 + 435 + 266 + 96 + 0 = 1393

Y a continuación hay que sumar las frecuencias:

40 + 99 + 145 + 133 + 96 + 40 = 553

La media se calcula obteniendo la relación entre estos dos valores, es decir: 1393/553= 2,51.

Ahora vamos a pasar al cálculo de la varianza. Para cada observación obtendremos el producto de la frecuencia por la diferencia elevada al cuadrado entre el valor observado y la media.

Por ejemplo, para la primera observación tenemos:

40\*((5 - 2,51)2) = 246,21

Lo que nos da la siguiente tabla:

| Opinión (Xi) | Cantidad de votantes (Ni) | Productos (Ni \* Xi) | Ni\* ((Xi-media)2) |
| --- | --- | --- | --- |
| 5 | 40 | 200 | 246,21 |
| 4 | 99 | 396 | 217,14 |
| 3 | 145 | 435 | 33,54 |
| 2 | 133 | 266 | 35,82 |
| 1 | 96 | 96 | 221,50 |
| 0 | 40 | 0 | 253,81 |

Esto nos permite calcular la varianza haciendo la suma de la columna que acabamos de crear dividida entre la suma de las frecuencias:

Varianza = (246,21 + 217,14 + 33,54 + 35,82 + 221,50 + 253,81)/553 = 1,82

Por último, podemos terminar con la desviación típica calculando la raíz cuadrada de la varianza:

Lo que da un valor de 1,35 para la desviación típica.

Ahora le toca a usted examinar el reparto de las observaciones en función de las desviaciones entre la media y la desviación típica que permite definir los 68 %, 95 % y 97 % de repartos.

Como ejemplo, podemos comprobar que el 68 % de las observaciones están comprendidas en el intervalo [1,3]. Los límites del intervalo se han determinado mediante la resta de la desviación típica a la media para el límite inferior y la suma de la desviación típica a la media para el límite superior. Lo que nos da los siguientes resultados:

* Cantidad de observaciones total: 553
* Cantidad de observaciones comprendidas entre 1 y 3 = 145 + 133 + 96 = 374
* Un porcentaje de 374/553 = 67,63 %