ESTANDARIZAR LOS DATOS

1. PUNTAJES Z

Si lo que deseamos es comparar datos de más de una distribución es necesario estandarizar los datos de tal suerte que efectivamente los datos sean comparables. Para poder hacerlo necesitamos calcular los **puntajes Z (***Z-scores***)**. Una variable donde las observaciones han sido convertidos en puntajes Z se conoce con el nombre de variable estandarizada (o normalizada).

Ejemplo

Supongamos que dos estudiantes están discutiendo a quién le fue mejor en sus exámenes. María obtuvo 65 en su examen de Políticas Públicas y Alejandra obtuvo 60 en su examen de Economía. ¿A cuál de las dos alumnas le fue mejor?

Podríamos afirmar, a simple vista que a María le fue mejor que a Alejandra porque su puntaje es más alto. Sin embargo, ¿qué pasa si a la mayoría de los estudiantes les fue muy bien en el examen de Políticas Públicas, mientras que muchos estudiantes les fue mal en el examen de Economía?

Para poder contestar la pregunta de forma adecuada necesitamos saber como les fue a los otros estudiantes de ambos cursos. Al respecto, supongamos que logramos conseguir dicha información y encontramos que para el curso de Políticas Públicas el puntaje medio fue de 60 con una desviación estándar de 5 y para el curso de Economía el puntaje medio fue de 50 con una desviación estándar de 6.

La información anterior es suficiente para poder calcular un puntaje Z ya que la formula correspondiente es:

$$Z_i = \frac{X_i - \bar{X}}{DE}$$

X_i = es un puntaje específico
$\bar{X}$ = es la media para todo el grupo
DE = es la desviación estándar para todo el grupo

Para el ejemplo de María el puntaje Z es:

$$Z = \frac{65 - 60}{5} = 1.00$$

y para Alejandra el puntaje Z es:

$$Z = \frac{60 - 50}{6} = 1.67$$

Alejandra tuvo un puntaje Z más alto y entonces, efectivamente, a ella le fue mejor que a María tomando en consideración a los demás estudiantes del curso. Los resultados anteriores nos dicen que el puntaje de Alejandra es 1.67 desviaciones estándar por arriba de la media en el curso de Economía y el resultado de María fue de 1 desviación estándar por arriba de la media en el curso de Políticas Públicas.

Adicionalmente, podemos afirmar que a ambas estudiantes les fue bien porque ambos puntajes se ubican por arriba de la media y por lo tanto obtuvieron un puntaje Z positivo.

Supongamos que queremos comparar el resultado de una tercera estudiante, Liliana. Ella obtuvo un puntaje de 47 en el examen de Economía. El puntaje Z de Liliana sería:

$$Z = \frac{47 - 50}{6} = -0.5$$

Liliana obtuvo un puntaje Z negativo que significa que su puntaje se ubica por debajo de la media para el grupo. De hecho, su puntaje fue exactamente 0.5 desviaciones estándar por debajo de la media. La desviación estándar para el grupo fue de 6, así que 0.5 de una desviación estándar es 3. El puntaje de 47 está tres puntos por debajo de la media (que en este caso es de 50).

Del ejemplo anterior, podemos apreciar cómo interpretar los puntajes Z:

-Un puntaje Z mide el número de desviaciones estándar que una observación se aleja de la media.

-Un puntaje Z positivo muestra que la observación es más grande que la media (se ubica por arriba del promedio).

-Un puntaje Z negativo muestra que una observación es más pequeña que la media (se ubica por abajo del promedio).

-El puntaje Z será de 0 si la observación es igual a la media.

La mayoría de los puntajes Z se ubicarán dentro del rango de [Z = -2] a [Z = 2]. Los valores alejados más de 2 desviaciones estándar de la media tienden a ser valores extremos (*outliers*).

Una observación importante con respecto de los puntajes Z es que su utilización asume que los resultados para ambos exámenes presentaron una distribución normal. En otras palabras, un histograma de los resultados mostraría una curva simétrica y también explica porque se usa la media y la desviación estándar como medidas de promedio y dispersión.

2. La media y la desviación estándar de una variable estandarizada

Una variable estandarizada tiene ciertas propiedades, como se podrá apreciar a continuación.

Si retomamos los resultados del peso de las corredoras (antes de que perdieran peso) para mostrar algunas propiedades de los puntajes Z. El peso promedio de las corredoras fue de 62 kg con una desviación estándar de 5 kg. Los puntajes Z serían los siguientes:

X	$Z = \frac{X_i - \bar{X}}{DE}$
55	-1.4
59	-0.6
63	0.2
66	0.8
67	1.0

Si sumamos todos los puntajes Z obtenemos un valor de 0. Así si quisiéramos calcular la media de los puntajes Z obtendríamos como respuesta 0 (0 ÷ 5 = 0). La media de un conjunto de puntajes Z es siempre de cero (si los calculamos correctamente).

Ahora bien, para calcular la desviación estándar de los puntajes Z, tendríamos que restar la media (media de puntaje Z - 0) de cada puntaje Z y posteriormente sacar el cuadrado.

Z	$(Z_i - \bar{Z})^2$
-1.4	1.96
-0.6	0.36
0.2	0.04
0.8	0.64
1.0	1.00
	$\sum$ 4.00

$$DE = \sqrt{\left[\frac{\sum (Z_i - \bar{Z})^2}{n-1}\right]} = \sqrt{\left[\frac{4.00}{4}\right]} = \sqrt{1} = 1$$

La desviación estándar de los puntajes Z es de 1. En realidad, no es necesario hacer el cálculo, pero la idea es mostrar que este resultado, aplicado a cualquier conjunto de datos (independientemente de su forma) nos muestra la regla de datos estandarizados:

-La media de una variable estandarizada es *cero*
-La desviación estándar de una variable estandarizada es de *uno*

3. Cálculo de un índice de puntaje Z

En el cuadro siguiente se muestra la distribución de puntajes medios de varios cursos tomados por estudiantes de primer año. Todos tomaron cuatro materias y lo que deseamos es comparar dos estudiantes. ¿Por qué no sería adecuado averiguar el puntaje promedio de cada estudiante para hacer la comparación?

1. Distribución de puntajes de seis cursos de primer año

Curso	Puntaje medio (%)	Desviación estándar
Métodos cuantitativos	65	2
Políticas Públicas	55	5
Sociología	54	4
Psicología	49	3
Economía	51	6
Demografía	53	4

Los resultados favorecerían a aquellos estudiantes en que el grupo tuvo puntajes medios altos. Necesitamos un puntaje para cada curso que indique la posición relativa del estudiante en dicho curso. Entonces necesitamos estandarizar los datos y calcular los puntajes Z.

El cuadro 2 muestra los puntajes obtenidos por los dos estudiantes

2. Puntajes obtenidos por dos estudiantes de primer año

Curso	Puntaje estudiante A (%)	Puntajes estudiante B (%)
Métodos cuantitativos	67	--
Políticas Públicas	53	52
Sociología	56	54
Psicología	43	--
Economía	--	57
Demografía	--	57

¿Quién tuvo mejores puntajes en general? Para saberlo es necesario calcular el puntaje Z para cada puntaje y posteriormente combinarlos.
Por ejemplo, el estudiante A obtuvo 67 en el curso de Métodos cuantitativos. El puntaje medio para ese curso fue de 65 con una desviación estándar de 2. El puntaje Z en este caso es:

$$Z = \frac{67 - 65}{2} = 1.00$$

Este puntaje nos dice que el puntaje del estudiante A fue de una desviación estándar por arriba del puntaje medio para el curso en cuestión. En otras palabras, el estudiante tuvo un buen puntaje.

Ahora necesitamos saber el puntaje Z medio para cada estudiante. Para esto sumamos los puntajes Z y los dividimos por cuatro (el número de exámenes presentados).

$$\text{Estudiante A: Puntaje Z medio} = \frac{1.0 + (-0.4) + 0.5 + (-2.0)}{4} = -0.225$$

$$\text{Estudiante B: Puntaje Z medio} = \frac{(-0.6) + 0.0 + 1.0 + 1.0}{4} = 0.35$$

Ahora podemos concluir que el estudiante B le fue mejor que el promedio en los cursos que tomó porque el puntaje Z medio es positivo (0.35) y el estudiante A estuvo por debajo de la media cuando consideramos todos los cursos que tomó porque el puntaje Z medio es negativo (-0.225). El estudiante B tuvo mejor desempeño en todos sus exámenes que el estudiante A.

Ejemplo

A un grupo de derechos humanos se le ha solicitado que dé una idea de cómo se comparan países asiáticos en términos de empoderamiento femenino. Usted obtiene datos comparables en tres temas: el porcentaje de ingreso ganado por mujeres, el porcentaje de lugares en el parlamento ocupados por mujeres y, por último, la taza de mortalidad materna durante el alumbramiento. Los datos de 6 países asiáticos se muestran en el cuadro siguiente.

Medidas de empoderamiento femenino en 8 países asiáticos

País	% ingreso	% de lugares en el parlamento	% tasa de mortalidad materna[1]
Filipinas	31	8.5	280
India	26	7.3	570
Tailandia	37	6.6	200
Malasia	30	10.30	80
Indonesia	33	12.60	650
China	38	21.0	95
Bangladesh	23	9.1	850
Pakistan	21	3.4	340

[1] Por 100 000 nacimientos vivos

Calcule un índice de empoderamiento para las mujeres para los 6 países utilizando los puntajes Z. ¿En qué países las condiciones son más favorables para las mujeres? ¿En qué países las condiciones son menos favorables?

a. Calcule el puntaje medio para cada variable
b. Calcule la desviación estándar para cada variable
c. Calcule los puntajes Z (si es necesario revise que los haya calculado correctamente)
d. Reflexione acerca de lo que significan las variables. Si el porcentaje de ingreso y el porcentaje de lugares en el parlamento son altos, esto sería bueno para las mujeres. Sin embargo, si la tasa de mortalidad materna es alta, esto es malo para las mujeres. Por lo tanto, revierta los signos de los puntajes Z para la mortalidad materna de tal

suerte que los positivos se conviertan negativos y los negativos se conviertan en positivos. Ahora un puntaje Z positivo para la mortalidad materna significa que la mortalidad es baja. Un puntaje Z positivo para las tres variables ahora indica un buen escenario para las mujeres.

e. Encuentre el puntaje Z medio para cada país.
f. Interprete sus respuestas. ¿Qué significa o qué indica un valor de índice alto o un valor de índice bajo?