SESIÓN 13

LA TRANSFORMACIÓN Z

I. CONTENIDOS:

1. Estandarización de datos, transformación Z.
2. La distribución normal.
3. La distribución normal estándar.

II. OBJETIVOS:

Al término de la Sesión, el alumno:

- Analizará la relación entre transformación Z y distribución normal estándar.
- Comprenderá bajo un caso práctico de estadística que la transformación Z representa la realidad de un concepto.

III. PROBLEMATIZACIÓN:

Comenta las preguntas con tu Asesor y selecciona las ideas más significativas.

- ¿Qué entiendes por estandarización?
- ¿Cómo representarías gráficamente una distribución datos?

IV. TEXTO INFORMATIVO-FORMATIVO:

1.1. Estandarización de datos, transformación Z

En este capítulo abordaremos el tratamiento de las variables cardinales continuas en distribuciones de probabilidad. La estandarización de datos o transformación Z, es un procedimiento aritmético que se efectúa sobre los datos de una distribución que se aproxima a la distribución normal (en el siguiente apartado se explicará), a fin de que el nuevo conjunto de datos tenga como media el número cero y como desviación estándar el número uno.

Para efectuar una transformación Z, se hace lo siguiente:

1) Se disponen, en una columna, los datos recolectados (X) de la variable cardinal continua.
2) Calcula la sumatoria de ellos y ésta divídela entre el número de datos (N), para así encontrar la media ($\overline{X}$)
3) En otra columna se dispone la diferencia de cada dato (X) con la media ($\overline{X}$): ($X - \overline{X}$)
4) Se eleva al cuadrado cada resultado de la segunda columna, para agregar otra columna que tenga por encabezado $(X - \overline{X})^2$.
5) Calcula la sumatoria de la tercera columna y divide el resultado entre el número de datos (N)
6) Extrae la raíz cuadrada del resultado del paso anterior, teniendo así la desviación estándar (S)
7) Se forma una cuarta columna con los cocientes de los valores de la segunda columna entre la desviación estándar, cada resultado obtenido es un dato Z.
8) Eleva al cuadrado cada dato Z y dispón los resultados en una quinta columna (Z^2)
9) Finalmente, se prueba que la transformación Z, resulta en una nueva distribución con media igual a cero y desviación estándar igual a uno; con las sumatorias de las últimas columnas. Deben cumplirse la siguientes expresiones: $\Sigma Z = 0 \qquad \Sigma Z^2 = N$

2.1. La distribución normal

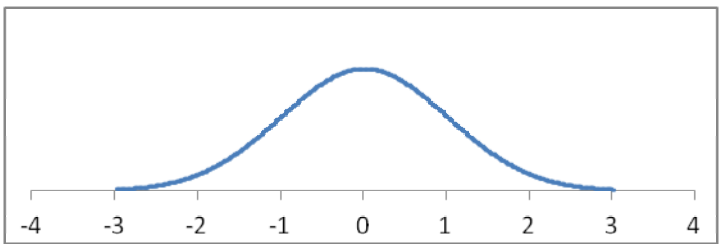
Una distribución normal es aquella formada por una variable aleatoria continua, cuya distribución en una gráfica tiene forma de campana y es simétrica respecto del valor central. Fue propuesta en

1733 por Abraham Moivre, un matemático francés. Se le conoce también como campana de Gauss porque la utilizó para sus estudios de astronomía. La función es un poco compleja:

$$y = \frac{e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}}{(\sigma)\left(\sqrt{2\pi}\right)}$$

Donde el numerador representa datos Z elevados al cuadrado, y el denominador el producto de la desviación estándar por la raíz cuadrada de dos veces el número de Euler por el número pi.

La gráfica, para una media de cero y una desviación estándar de uno, es como se muestra a continuación.



Hay una gran cantidad de fenómenos en la naturaleza que se aproximan a la distribución normal, por ejemplo la estatura de una población humana, el número de hijos, la radiación solar, etc.

3.1. La distribución normal estándar

La distribución normal estándar es una distribución normal con media igual a cero y desviación estándar de uno. De este modo al efectuar la transformación Z a los datos de una distribución aproximadamente normal, se ajusta la distribución a la normal estándar; esto con el fin de hacer un cálculo de probabilidades de ocurrencia de los elementos de la distribución original.

El área bajo la curva normal es la misma, en proporción para todas las distribuciones normales. Si se ha obtenido un dato Z de una distribución, y este resulta Z =-1.3, entonces significa que el dato original se aleja de la media 1.3 desviaciones estándar a la izquierda. Del mismo modo, un dato Z = 2.4 significa que el dato original se aleja 2.4 unidades a la derecha de la media.

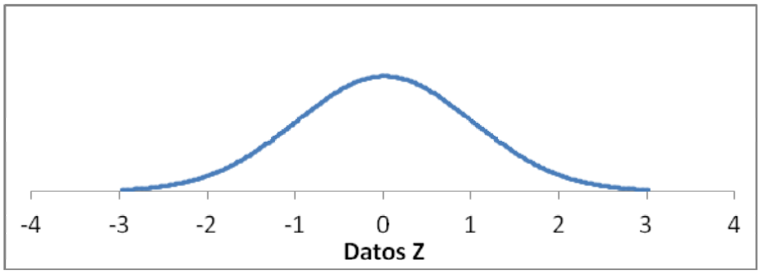
Los porcentajes de área bajo la curva normal estándar se muestran a continuación:

Para Z = 1 o Z = -1, el área bajo la curva desde la media es del 34.13%
Para Z = 2 o Z = -2, el área bajo la curva desde la media es del 47.73%
Para Z = 3 o Z = -3, el área bajo la curva desde la media es del 49.865%
Para Z = 4 o Z = -4, el área bajo la curva desde la media es del 49.99683%

De este modo si queremos saber el área bajo la curva desde Z = -2 hasta Z = 3, sólo debemos sumar los porcentajes 47.73% + 49.865% que resultan 97.595%

Pero si queremos saber el porcentaje de área bajo la curva a la derecha de Z = 1, le restamos al área de la mitad de la curva el área que hay desde la media hasta Z = 1, 50% - 34.13% = 15.87%
Ahora si necesitamos el área bajo la curva a la izquierda de Z = 1, sumamos 50% con 34.13, siendo esto igual a 84.13%

Finalmente, si debemos calcular el porcentaje de área bajo la curva desde Z = 1 hasta Z = 2, entonces debemos restar, 47.73% - 34.13% = 13.6%



De este modo, la recomendación es que se trace la curva normal; luego líneas verticales desde los valores Z que se indiquen en el eje horizontal hasta la curva; finalmente se sombrean las zonas correspondientes a lo que se solicita en el problema, y con esto se decide qué sumar o restar con qué.

Hace falta señalar que para los demás valores que se pueden obtener de una transformación Z, hay una tabla que permite conocer el área bajo la curva normal. Es muy sencilla su lectura, se comienza por la columna de la izquierda y allí se busca el dato Z con su primer cifra decimal, luego se avanza por la fila hasta encontrar la columna cuyo encabezado tiene a la segunda cifra decimal del dato Z. En ese cruce de fila y columna, hay un número al que debemos pensar como porcentaje si le agregamos un punto decimal después de las dos cifras. No se acostumbra poner los puntos decimales con el fin de ahorrar espacio en la tabla.

Según hemos aprendido, las frecuencias relativas de una distribución se pueden asociar con la probabilidad de ocurrencia de cada categoría. Entonces, los porcentajes de área bajo la curva normal también se pueden asociar con la probabilidad de ocurrencia de alguna categoría de la variable cardinal continua.

Ejemplo 1 Las medidas de contenido de nicotina de cada cigarrillo de la marca X se distribuyen de manera aproximadamente normal, tienen una varianza de 16 mg y una media de 24 mg, ¿cuál es la probabilidad de encontrar un cigarrillo con más de 28 mg de nicotina?

Para resolver el problema, primero determinamos la desviación estándar que es la raíz cuadrada de la varianza, luego transformamos a dato Z el valor de 28 mg sustituyendo en la fórmula:

$$Z = \frac{X - \overline{X}}{S} = \frac{28 - 24}{4} = 1$$

Enseguida trazamos la curva normal con una línea vertical en Z = 1 y sombreamos la parte derecha de la línea debajo de la curva porque se quiere encontrar la probabilidad de que haya un cigarrillo con más de 28 mg de nicotina, más por lógica significa a la derecha. Finalmente razonamos que se debe restar al 50%, que es el área bajo la curva a la derecha de la media, el porcentaje para Z = 1 desde la media. Entonces: 50% - 34.13% = 15.87%

Esto significa que hay un 15.87% de probabilidad de encontrar un cigarrillo que tenga más de 28 mg de nicotina en la marca X.

A continuación se muestra la tabla que se ha descrito para la distribución normal estándar.

z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0.0	0000	0040	0080	0120	0159	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0739	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2133	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2518	2549
0.7	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3718	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4083	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1.6	4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4758	4762	4767
2.0	4773	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4898	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4980	4980	4981
2.9	4981	4982	4983	4984	4984	4984	4985	4985	4986	4986
3.0	49865	4987	4987	4988	4988	4988	4989	4989	4989	4990
3.1	49900	4991	4991	4991	4992	4992	4992	4992	4993	4993
3.2	4993129									
3.3	4995166									
3.4	4996631									
3.5	4997674									
3.6	4998409									
3.7	4998922									
3.8	4999277									
3.9	4999519									
4.0	4999683									
4.5	4999966									
5.0	4999997133									

Ejemplo 2 Se recopilaron las estaturas de 24 niños al nacer, como se muestran a continuación. Transforma a datos Z, y calcula la probabilidad de que un niño tenga al nacer más de 51 cm de longitud para esa población.

47.2	48.5	49.0	46.3	47.1	48.0	48.9	47.1	48.5	46.7	45.4	47.0
48.0	45.3	47.3	50.8	46.7	48.9	49.9	46.9	48.0	46.3	46.7	50.2

Se procede tal como se señaló al principio de la clase:

X	$X-\overline{X}$	$(X-\overline{X})^2$	Z	Z^2
47,2	-0,50	0,25	-0,35	0,13
48,5	0,80	0,65	0,57	0,33
49	1,30	1,70	0,93	0,87
46,3	-1,40	1,95	-1,00	0,99
47,1	-0,60	0,36	-0,43	0,18
48	0,30	0,09	0,22	0,05
48,9	1,20	1,45	0,86	0,74
47,1	-0,60	0,36	-0,43	0,18
48,5	0,80	0,65	0,57	0,33
46,7	-1,00	0,99	-0,71	0,51
45,4	-2,30	5,27	-1,64	2,69
47	-0,70	0,48	-0,50	0,25
48	0,30	0,09	0,22	0,05
45,3	-2,40	5,74	-1,71	2,93
47,3	-0,40	0,16	-0,28	0,08
50,8	3,10	9,64	2,22	4,92
46,7	-1,00	0,99	-0,71	0,51
48,9	1,20	1,45	0,86	0,74
49,9	2,20	4,86	1,57	2,48
46,9	-0,80	0,63	-0,57	0,32
48	0,30	0,09	0,22	0,05
46,3	-1,40	1,95	-1,00	0,99
46,7	-1,00	0,99	-0,71	0,51
50,2	2,50	6,27	1,79	3,20
1144,7		47,05	0,00	24,00

Media	47,70
Desviación Estándar	1,40

Para la segunda parte del problema, se transforma a dato Z, el valor de 51 cm. tal como se explicó en el primer ejemplo:

$$Z = \frac{X - \overline{X}}{S} = \frac{51 - 47.70}{1.4} = 2.36$$

Luego se consulta la tabla de la distribución normal estándar. Resulta un porcentaje de 49.09%.

Significa que desde la media hasta Z = 2.36 hay un 49.09% de área bajo la curva; entonces el resto es 0.91% que es la probabilidad de que en esa población tenga al nacer más de 51 cm de estatura un niño.

V. ESTRATEGIAS CENTRADAS EN EL APRENDIZAJE:

A. Realiza las siguientes actividades.

1. Transforma a datos Z cada una de las siguientes distribuciones y comprueba en cada caso que:
$\Sigma z = 0$ y $\Sigma z^2 = N$
   a) 1, 3, 2, 5, 6, 7, 4, 3
   b) 21, 25, 20, 19, 21, 25, 23

2. Determina el área bajo la curva normal limitada como se indica:
   a) Entre la media y Z = 2
   b) A la izquierda de Z = 1.52
   c) Por arriba de Z = 1
   d) Entre Z = -0.5 y Z = 0.8

3. Una máquina despachadora de refrescos está ajustada para servir en promedio 200ml por vaso. Si la cantidad de refrescos es normalmente distribuida con una desviación estándar de 15 ml:
   a) ¿Qué porcentaje de los vasos contendrá más de 224 ml?
   b) ¿Cuál es la probabilidad de que un vaso contenga entre 191 y 209 ml?
   c) ¿Cuántos vasos probablemente se derramen si se utiliza vasos de 230 ml en los siguientes 1000 refrescos?
   d) ¿Bajo que valor se obtiene el 25% más pequeño de los refrescos?