La distribución normal

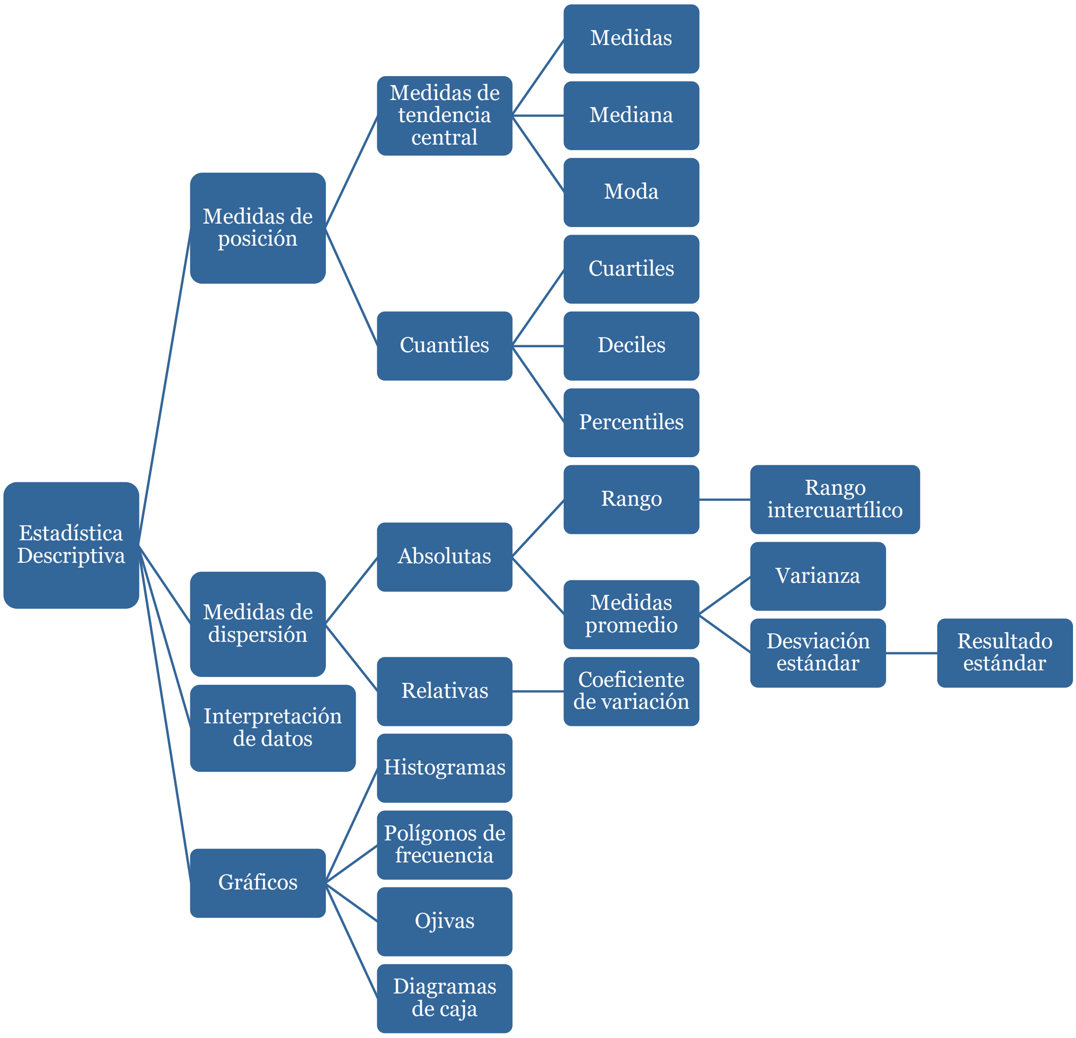
TEMA 7

Ideas clave

7.1. ¿Cómo estudiar este tema?

En este tema hablaremos sobre **la distribución normal y su importancia en la estadística.** Por otro lado se verá **la técnica conocida como normalización que sirve para pasar a una distribución normal.** Por último, se hablará de **la famosa regla 68/95/99,7.**

Para estudiar este tema **deberás comprender las Ideas clave** expuestas en este documento y que han sido elaboradas por el profesor de la asignatura. Estas ideas se van a complementar con lecturas y otros documentos para que puedas ampliar los conocimientos sobre el mismo. También es importante que **veas la lección magistral** grabada para afianzar los conocimientos expuestos en este tema.



7.2. El modelo de distribución normal

Las medidas de forma y concentración

Las medidas de forma y concentración son otro tipo de medidas que sirven para identificar la forma geométrica de una distribución y la uniformidad entre las variables de la muestra.

1. Medidas de forma

Las medidas de posición y dispersión son de gran utilidad ya que proporcionan parámetros que resumen las características de una dispersión, pero tienen el inconveniente de que no aportan información de los datos individuales. Las medidas de forma nos ayudan a saber la estructura de la distribución, es decir si es simétrica o asimétrica, puntiaguda o achatada. A tales medidas se las conoce como medidas de asimetría y medidas de curtosis respectivamente.

A. Medidas de asimetría

Una distribución es simétrica cuando al doblarla por el eje vertical las ramas de la distribución coinciden. Tal distribución se conoce como distribución normal o Gausiana y realmente se trata de una curva de frecuencias.

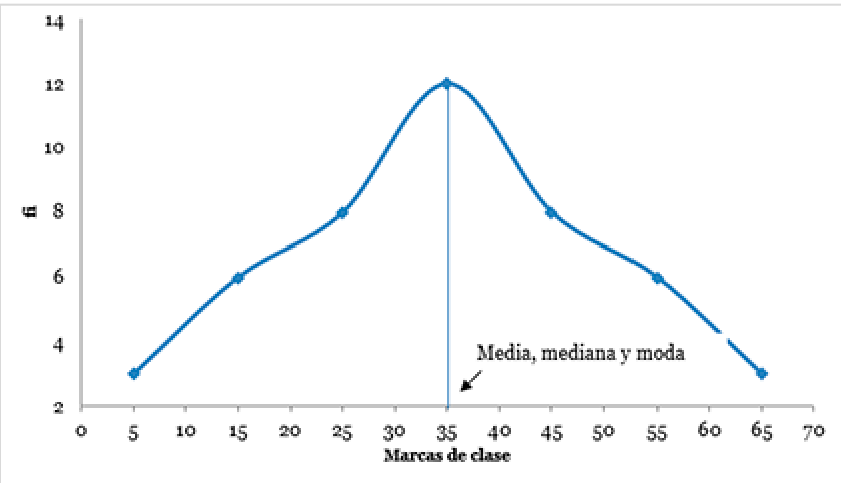


Gráfico 1. Curva de distribución normal o Gausiana.

Cuando las distribuciones son simétricas y de variables continuas, tanto la media como la mediana o la moda caen en el mismo lugar y dividen la distribución en dos partes de igual frecuencia y valores equidistantes dos a dos.

Cuando la distribución no es simétrica se dice que es sesgada. Una distribución sesgada puede ser positivamente o negativamente sesgada, moderadamente o significativamente sesgada y unimodal, bimodal o multimodal. Una distribución positivamente sesgada es aquella en la que el mayor número de frecuencias están a la izquierda de la distribución y la cola está a la derecha; mientras que una distribución negativamente sesgada es aquella en la que el mayor número de frecuencias están a la derecha y su cola está a la izquierda. Por otro lado, una distribución moderadamente sesgada tiene poca asimetría a diferencia de una distribución significativamente sesgada que es muy asimétrica. Por último, una distribución unimodal tiene una moda, una bimodal tiene dos modas y una multimodal tiene varias modas o picos de máxima concentración.

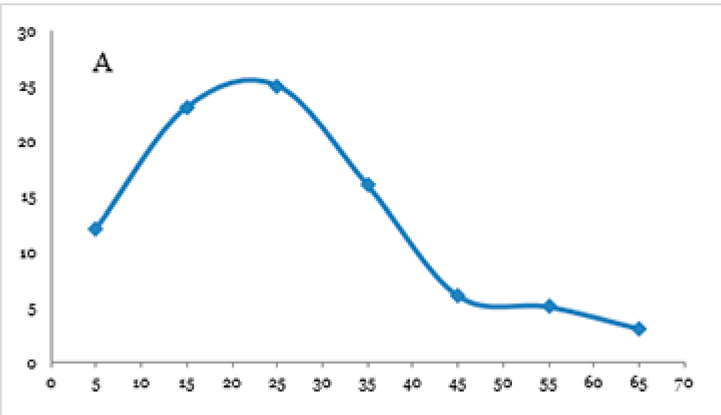


Gráfico 2. Distribución moderada, unimodal y positivamente sesgada.

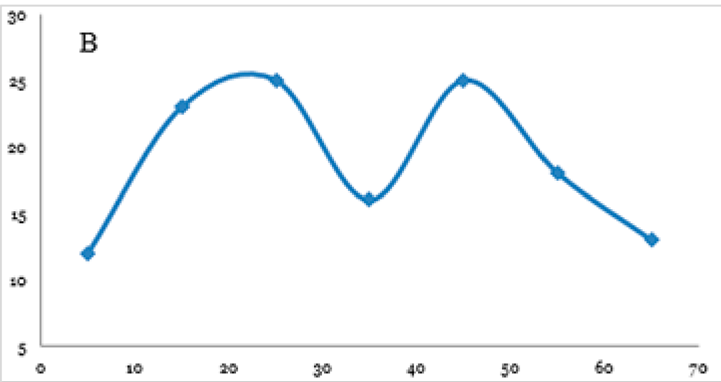


Gráfico 3. Distribución moderada y bimodal.

Las medidas de asimetría son aquellos parámetros que miden la asimétrica de una distribución respecto a la media, la mediana o la moda. A continuación se muestra cómo están dispuestas estas medidas de tendencia central en función del tipo de asimetría.

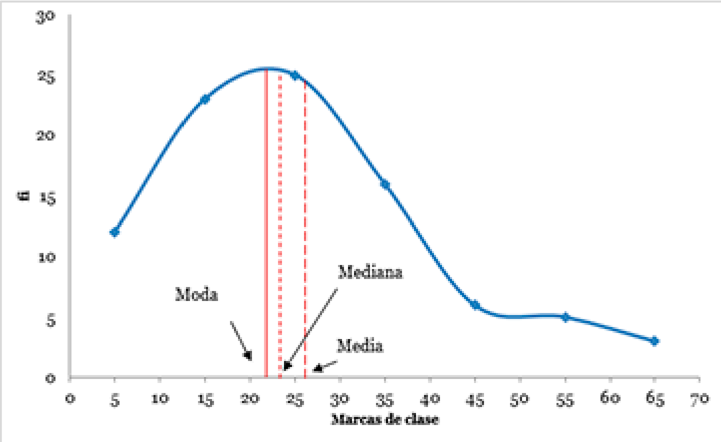


Gráfico 4. Distribución positivamente sesgada.

Tal y como se puede observar, en una distribución positivamente sesgada, la moda se halla en el punto de mayor frecuencia y más hacia la izquierda, la mediana está en el medio y la media a la derecha.

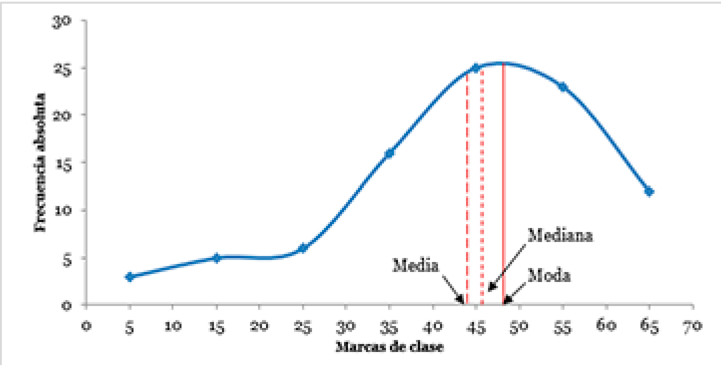


Gráfico 5. Distribución negativamente sesgada.

En una distribución negativamente sesgada, la mediana también se halla en el medio y la moda sigue coincidiendo con el valor de mayor frecuencia pero se encuentra a la derecha y la media a la izquierda.

En estos casos de asimetría, la mediana es la mejor medida de tendencia central por estar en medio. El parámetro más empelado para calcular la asimetría es el coeficiente de Fisher el cual se calcula de la siguiente forma:

$$g_1 = \frac{\frac{\sum_{i=1}^{k}(x_i-\bar{x})^3 \cdot f_i}{n-1}}{\left(\sqrt{\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2 \cdot f_i}{n-1}}\right)^3}$$

g_1 es el coeficiente de Fisher, $(x_i - \bar{x})$ es la distancia entre los valores y la media, f_i es la frecuencia de cada valor y n el número total de observaciones. El término del denominador es la desviación estándar elevada al cubo.

1.

En función del valor obtenido de este coeficiente se puede saber cómo es la asimetría:

- La curva es simétrica si $A_F = 0$.
- La curva es positivamente sesgada si $A_F > 0$.
- La curva es negativamente sesgada si $A_F < 0$.

Ejemplo 1. Estudiar la asimetría de una distribución

A partir de los datos de la siguiente tabla 1 del Tema 7 decidir si hay asimetría y en tal caso el tipo de la misma.

I_i	f_i
4-8	25
9-15	18
16-20	12
21-27	15
28-32	5

Tabla 1. Valores de faltas de asistencia a un curso.

Se calcula la media con las marcas de clase:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot f_i = \frac{6 \cdot 25 + 12 \cdot 18 + 18 \cdot 12 + 24 \cdot 15 + 30 \cdot 5}{75} = 14{,}56$$

Se hace la siguiente tabla:

Intervalo	x	f_i	$x-\bar{x}$	$(x-\bar{x})^2$	$f_i \cdot (x-\bar{x})^2$	$(x-\bar{x})^3$	$f_i \cdot (x-\bar{x})^3$
4-8	6	25	-8,56	73,27	1831,84	-627,22	-15680,55
9-15	12	18	-2,56	6,55	117,96	-16,78	-301,99
16-20	18	12	3,44	11,83	142,00	40,71	488,49
21-27	24	15	9,44	89,11	1336,70	841,23	12618,49
28-32	30	5	15,44	238,39	1191,97	3680,80	18403,99
Total		75			4620,48		15528,42

Aplicando la fórmula 1:

$$g_1 = \frac{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 \cdot f_i}{n-1}}{\left(\sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i}{n-1}}\right)^3} = \frac{\frac{15528,42}{74}}{\left(\sqrt{\frac{4620,48}{74}}\right)^3} = 0,43$$

Puesto que 0,43> 0 la distribución es positivamente sesgada.

B. Medidas de curtosis

Las medidas de curtosis dan información sobre el apuntamiento de una distribución unimodal, campaniforme y moderadamente simétrica. Si alrededor de la media hay muchos valores, el apuntamiento será elevado. De esta manera, las distribuciones se clasifican en: mesocúrticas, que son en las que el apuntamiento es el mismo que el de la distribución normal; leptocúrticas, aquellas que presentan más apuntamiento que la distribución normal; y platicúrticas, aquellas que presentan menos apuntamiento que la distribución normal.

En la siguiente figura se puede ver este tipo de distribuciones:

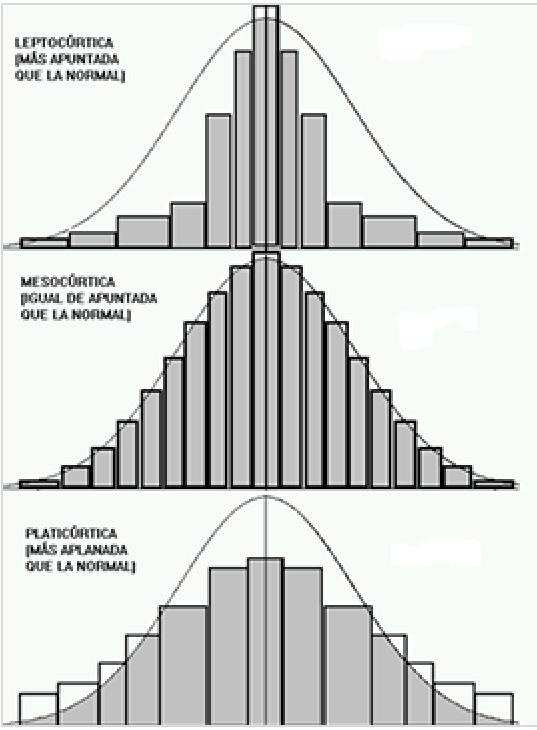


Figura 1. Tipos de apuntamiento. Fuente: uv.es

El grado de apuntamiento de una distribución se mide mediante el coeficiente de curtosis, representado como g_2 y se calcula a partir de la siguiente fórmula:

$$g_2 = \frac{\dfrac{\sum_{i=1}^{k}(x_i - \bar{x})^4 \cdot f_i}{n-1}}{\left(\sqrt{\dfrac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i}{n-1}}\right)^4} - 3$$

2.

- La distribución es normal o mesocúrtica si $g_2=0$.
- La distribución tiene mucho apuntamiento o leptocúrtica si $g_2 > 0$.
- La distribución es achatada o platicúrtica si $g_2 <0$.

Ejemplo 2. Calcular el apuntamiento de una distribución

A partir de los datos de la siguiente tabla 1 del Tema 7 estudiar el apuntamiento de la distribución:

I_i	f_i
4-8	25
9-15	18
16-20	12
21-27	15
28-32	5

Tabla 2. Valores de faltas de asistencia a un curso.

Se calcula la media con las marcas de clase:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot f_i = \frac{6 \cdot 25 + 12 \cdot 18 + 18 \cdot 12 + 24 \cdot 15 + 30 \cdot 5}{75} = 14{,}56$$

Se hace la siguiente tabla:

Intervalo	x	f_i	$x-\bar{x}$	$(x-\bar{x})^2$	$f_i \cdot (x-\bar{x})^2$	$(x-\bar{x})^4$	$f_i \cdot (x-\bar{x})^4$
4-8	6	25	-8,56	73,27	1831,84	5369,02	134225,51
9-15	12	18	-2,56	6,55	117,96	42,95	773,09
16-20	18	12	3,44	11,83	142,00	140,03	1680,41
21-27	24	15	9,44	89,11	1336,70	7941,23	119118,51
28-32	30	5	15,44	238,39	1191,97	56831,51	284157,54
Total		75			4620,48		539955,06

Aplicando la fórmula 2:

$$g_2 = \frac{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^4 \cdot f_i}{n-1}}{\left(\sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i}{n-1}}\right)^4} - 3 = \frac{\frac{539955{,}06}{74}}{\left(\sqrt{\frac{4620{,}48}{74}}\right)^4} - 3 = -1{,}13$$

Puesto que -1,13 < 0 la distribución es platicúrtica o achatada.

2. Medidas de concentración

Las medidas de concentración dan información sobre el grado de desigualdad que tiene una variable en un conjunto. Estas medidas son empleadas en temas socio-económicos para saber, por ejemplo, el grado de uniformidad entre los salarios de unos trabajadores, entre otras cuestiones. Cuando los datos vienen agrupados en intervalos, es más complejo discernir la uniformidad o no de cada individuo, por lo que estas medidas se emplean más con datos no agrupados.

Se establecen dos tipos de concentración: mínima y máxima concentración. La situación de mínima concentración es similar a una situación de máxima equidad entre los valores y la situación de máxima concentración supone mínima equidad. Esto significa que, si suponemos por ejemplo que en una empresa todos los trabajadores reciben el mismo salario, la equidad es máxima; y al contario, si la mayor parte del salario está repartido entre una sola persona o unos pocos la equidad es mínima.

Para saber el grado de concentración se emplean el **Índice de Gini**, que da **valores numéricos,** y la **curva de Lorenz,** que da **valores gráficos**.

Para el cálculo del Índice de Gini hay que ordenar los datos. Siguiendo con el ejemplo de los salarios, hay que ordenar los preceptores, individuos, de mayor a menor cantidad de salario y calcular los valores acumulados tal y como se muestra en la siguiente tabla:

Valor	Individuos	Individuos acumulados	Valor obtenido	Valor acumulado	$p_i = \frac{precep.acum.}{n} = \frac{F_i}{n}$	$q_i = \frac{valor.acum.}{valor\ total} = \frac{U_i}{\sum_{i=1}^{k} x_i \cdot f_i}$
x_1	f_1	$f_1=F_1$	$x_1 \cdot f_1 = V_1$	$V_1=U_1$	p_1	q_1
x_2	f_2	$f_1+f_2=F_2$	$x_2 \cdot f_2 = V_2$	$V_1+V_2 = U_2$	p_2	q_2
…	…	…	…	….	…	…
x_i	f_i	$f_1+f_2+\ldots=F_i$	$x_i \cdot f_i=V_i$	$V_1+V_2+\ldots.V_i= U_i$	p_i	q_i
…	…	…	…	….	…	…
x_k	f_k	$f_1+f_2+\ldots f_i+\ldots =F_k$	$x_k \cdot f_k=V_k$	$V_1+V_2+\ldots.V_i+ \ldots V_k=U_k$	p_k	q_k
	n		$\sum_{i=1}^{k} x_i \cdot f_i = V$	U		

Tabla 3. Ordenación de los preceptores según el valor de los datos.

Para calcular el Índice de Gini se empela la siguiente fórmula:

$$I_G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

3.

A partir de esta fórmula se deduce que:

- Si todos los trabajadores reciben el mismo salario, la situación es de mínima concentración o máxima igualdad, p_i es igual a q_i y el I_G=0.
- Si el salario se concentra en uno o muy pocos trabajadores, la situación es de máxima concentración o mínima igualdad, q_i sería cero y el I_G=1.
- Para cualquier caso intermedio en el que $p_i > q_i$: $0 < I_G < 1$.

Ejemplo 3. Cálculo del Índice de Gini

A partir de los datos de salarios recibidos por empleados en una empresa calcular el índice de concentración de Gini.

Intervalo	f_i
8.000-10.000	35
10.000-15.000	64
15.000-20.000	23
20.000-30.000	3
30.000-50.000	7

Tabla 4. Salarios anuales recibidos por los empleados de una empresa.

Se construye la siguiente tabla:

x_i	f_i	F_i	$x_i \cdot f_i$	U	$p_i = \frac{F_i}{n}$	$q_i = \frac{Ui}{\sum_{i=1}^{k} x_i \cdot f_i}$	*pi-qi*
9.000	35	35	315000	315000	0,26	0,17	0,10
12.500	64	99	800000	1115000	0,75	0,59	0,15
17.500	23	122	402500	1517500	0,92	0,81	0,11
25.000	3	125	75000	1592500	0,95	0,85	0,10
40.000	7	132	280000	1872500	-	-	-
Total	132		1872500		2,89		0,46

Aplicando la fórmula 3:

$$I_G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} = \frac{0,46}{2,89} = 0,16$$

El índice de Gini muestra que la situación es más próxima a la de mínima concentración, ya que hay bastante uniformidad entre los salarios más bajos.

La curva de Lorenz se obtiene a partir del Índice de Gini y permite visualizar la manera en la que se han llevado a cabo los repartos. Para dibujar la curva, se colocan en el eje de abscisas los valores de p_i y los valores de q_i en el eje de ordenadas. Los valores de p_i y q_i oscilan entre 0 y 1, o entre 0 y 100 si se pone en porcentaje. Al unir los puntos se obtiene una curva que es la curva de Lorenz. Al trazar la recta $y=x$ (recta de equidistribución) la curva de Lorenz quedará por debajo de la recta puesto que los valores de p_i son mayores que los q_i.

Las siguientes imágenes muestran los diferentes tipos de curvas de Lorenz:

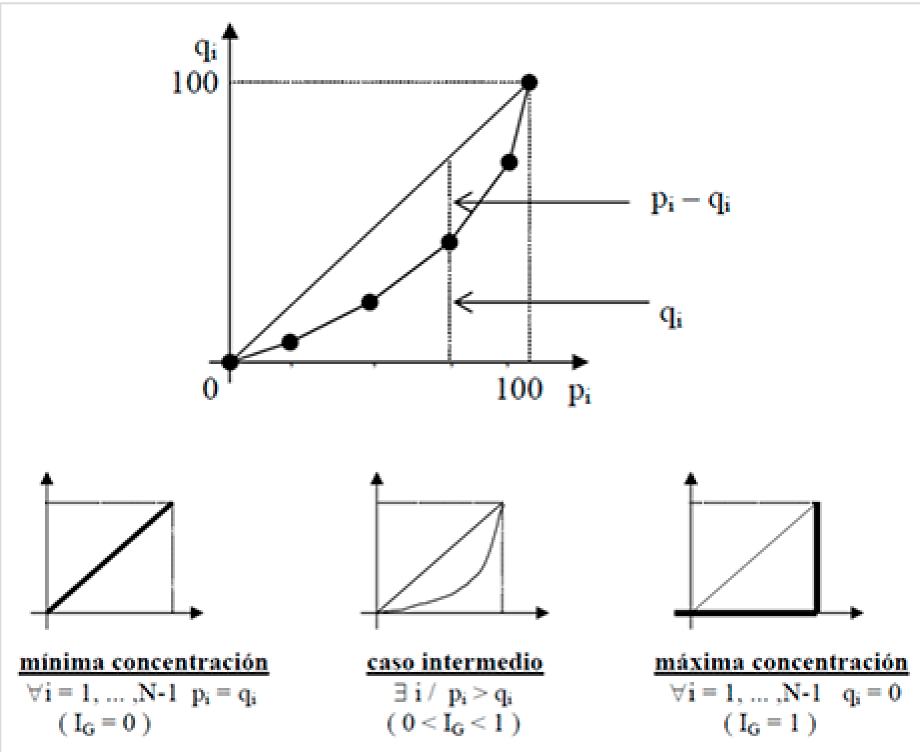


Figura 2. Tipos de curvas de Lorenz. Fuente: ucm.es

Los casos posibles son los siguientes:

- Mínima concentración o máxima igualdad: es la recta de equidistribución ya que los valores de p_i son iguales a los q_i.
- Máxima concentración o mínima igualdad: la curva de Lorenz se transforma en los catetos de un triángulo rectángulo. La diferencia entre los valores de p_i y q_i es la mayor posible y la curva está lo más alejada posible de la diagonal.
- Concentración intermedia: caso intermedio en el que la curva de Lorenz es una función cuadrática. Cuanto más se aleje la curva de la diagonal más situación de máxima concentración habrá.

Así mismo, a partir de la curva de Lorenz se puede calcular el Índice de Gini como el área que deja la curva con la diagonal de la siguiente manera:

$$I_G = \frac{\text{área entre curva de Lorenz y la diagonal}}{\text{área de la diagonal con los ejes}}$$

4.

En el caso de saber la ecuación de la curva, el área se resuelve mediante métodos de integración.

Ejemplo 4. Dibujar una curva de Lorenz

A partir de los datos del Ejemplo 3 dibujar la curva de Lorenz.

x_i	f_i	F_i	$x_i \cdot f_i$	U	$p_i = \frac{F_i}{n}$	$q_i = \frac{Ui}{\sum_{i=1}^{k} x_i \cdot f_i}$	pi-qi
9.000	35	35	315000	315000	0,26	0,17	0,10
12.500	64	99	800000	1115000	0,75	0,59	0,15
17.500	23	122	402500	1517500	0,92	0,81	0,11
25.000	3	125	75000	1592500	0,95	0,85	0,10
40.000	7	132	280000	1872500	-	-	-
Total	132		1872500		2,89		0,46

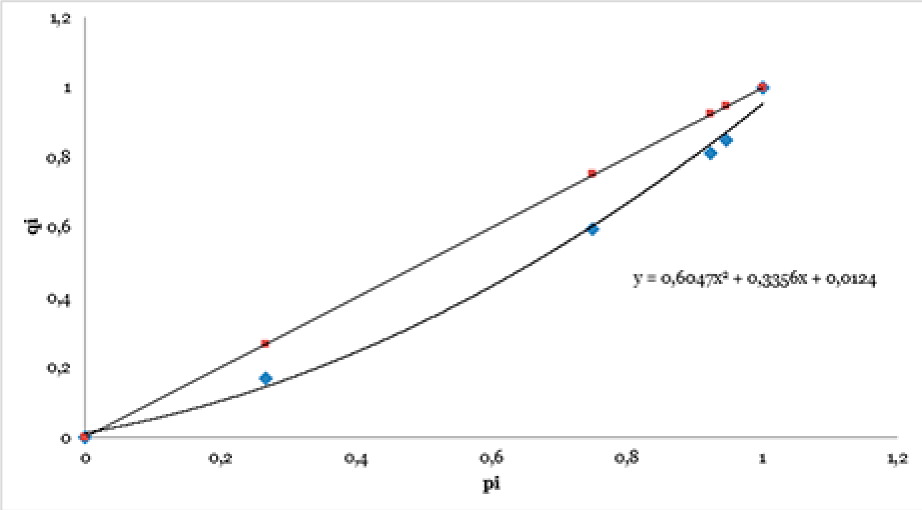


Gráfico 6. Curva de Lorenz para el Ejemplo 5.

La situación es de mínima concentración debido a que el área es pequeña.

La distribución normal

La distribución normal es ampliamente empleada en estadística, ya que muchas situaciones reales se ajustan a este tipo de distribución y, tal y como se verá posteriormente, es de gran utilidad en las distribuciones de muestreo.

Tal y como ya se ha mencionado, una distribución normal es aquella en la que la media cae en el punto medio de la distribución junto con la mediana y la moda y es unimodal. Además, las dos colas simétricas nunca tocan al eje horizontal de la distribución por lo que la probabilidad se extiende indefinidamente. Las curvas de distribución normal vienen definidas por la media y por la desviación estándar por lo que puede haber curvas con igual media pero distintas desviaciones estándar. Cuando la distribución tiene mucho apuntamiento la desviación estándar entre los valores es menor y viceversa.

Tal hecho se puede ver en la figura mostrada a continuación:

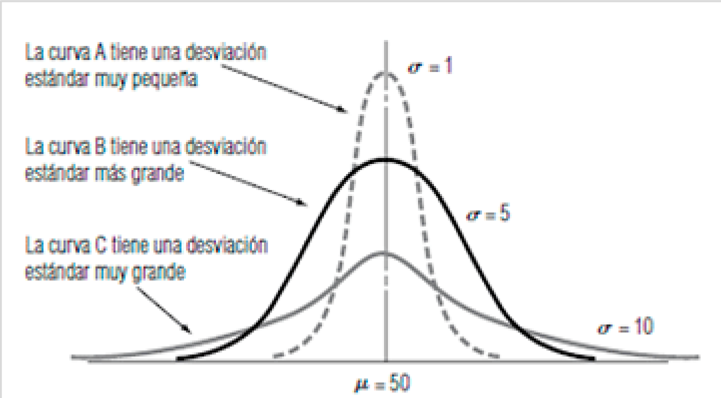


Figura 3. Distribuciones normales con igual media y distintas desviaciones estándar. Fuente: *Estadística para la administración y la economía de Levin y Rubin.*

Independientemente de los valores de las media y la desviacion estándar de la población, el área total bajo la curva siempre es 1 o el 100% y las diferentes áreas que se vaya tomando son probabilidades. Se demuestra que:

- El 68% aproximadamente de los datos caen a ±1 desviación estándar de la media.
- El 95% aproximadamente de los datos caen a ±2 desviaciones estándar de la media.
- El 99,7 % aproximadamente de los datos caen a ±3 desviaciones estándar de la media.

En la siguiente figura se puede ver la significación de lo explicado.

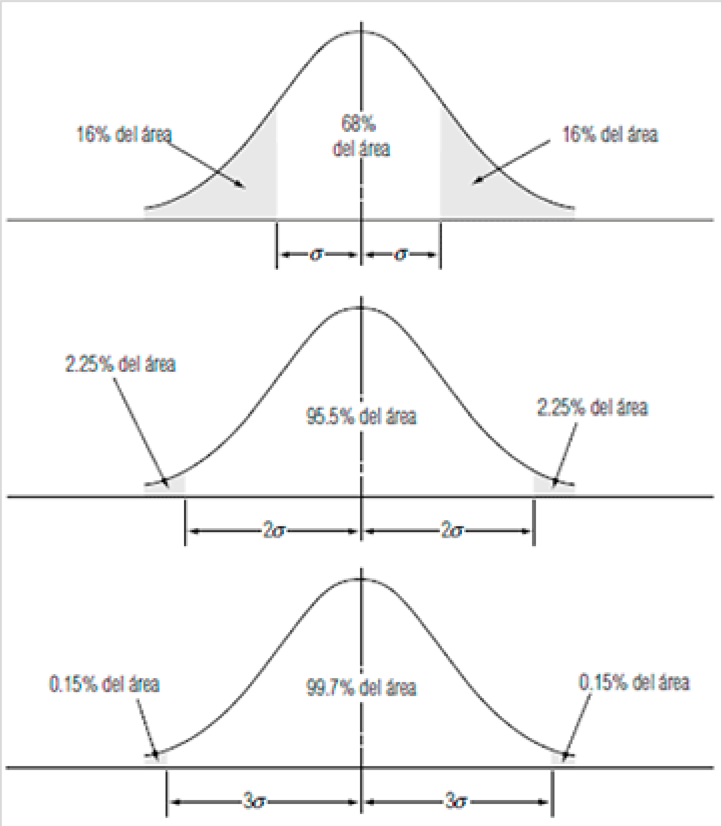


Figura 4. Áreas bajo la curva en la distribución normal. Fuente: *Estadística para la administración y la economía de Levin y Rubin.*

En la práctica, la mayor parte de las situaciones caerán en valores de distancias en desviaciones estándar de la media diferentes. En tales situaciones es preciso hacer uso de tablas que nos permitan saber la probabilidad de los datos tal y como se verá posteriormente.

Cabe destacar que dos distribuciones normales que tengan diferente media y diferente desviación estándar pueden tener el mismo área bajo la curva si las áreas sombreadas están a los mismos valores de desviaciones estándar de la media tal y como se muestra en la siguiente figura:

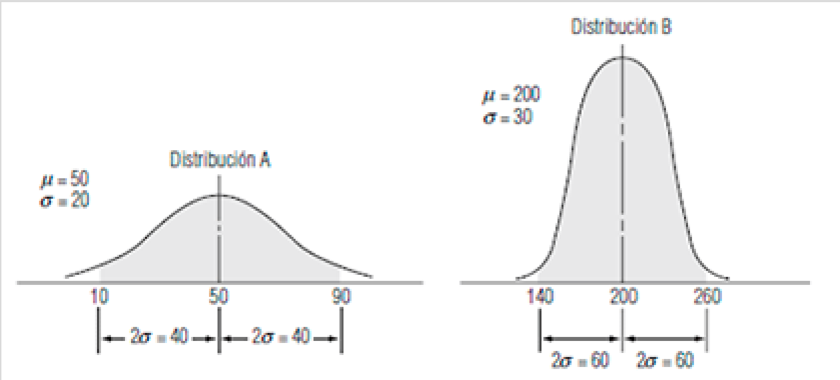


Figura 5. Distribuciones con distinta media y desviación estándar pero misma área bajo la curva. Fuente: *Estadística para la administración y la economía de Levin y Rubin.*

Por este motivo, se habla de distribución de probabilidad normal estándar para encontrar las áreas bajo la curva de cualquier curva normal.

7.3. La estandarización mediante puntuaciones z

En este apartado se aprenderá a hacer uso de la distribución de probabilidad normal estándar.

Lo primero es introducir la estandarización mediante un parámentro que se denota por Z. Este parámetro se emplea como sustituyente al número de desviaciones estándar respecto a la media, con el objetivo de hacer una estandarización de tales valores como números enteros de 0 a 3 en vez de tener que calcular para cada caso en particual el valor de las desviaciones estándar.

Para calcular Z se emplea la siguiente fórmula:

$$Z = \frac{x - \mu}{\sigma}$$

x es la variable aleatoria en cuestión, μ y σ son la media y la desviacion estándar de la distribución y Z es el número de desviaciones estándar a las que la variable de estudio se halla de la media.

5.

En la siguiente figura se visualiza lo que se pretende obtener con la estandarización mediante puntuaciones Z.

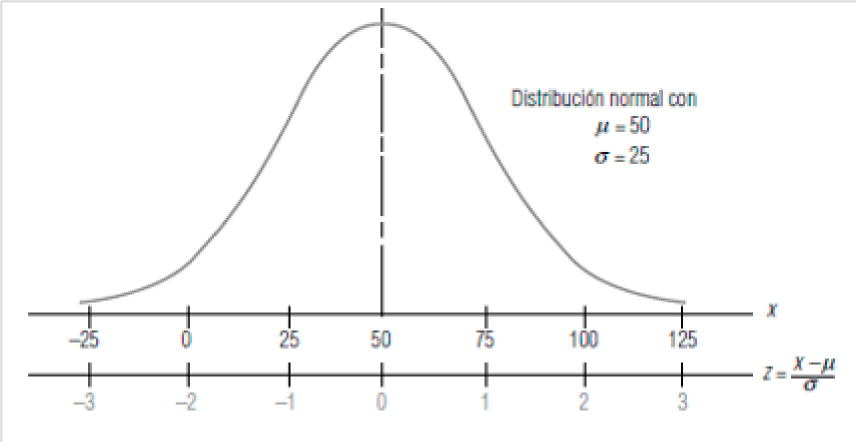


Figura 6. Estandarización con Z. Fuente: *Estadística para la administración y la economía de Levin y Rubin.*

7.4. La tabla de probabilidad normal

A continuación se presenta la tabla de distribución de probabilidad normal estándar.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	6.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

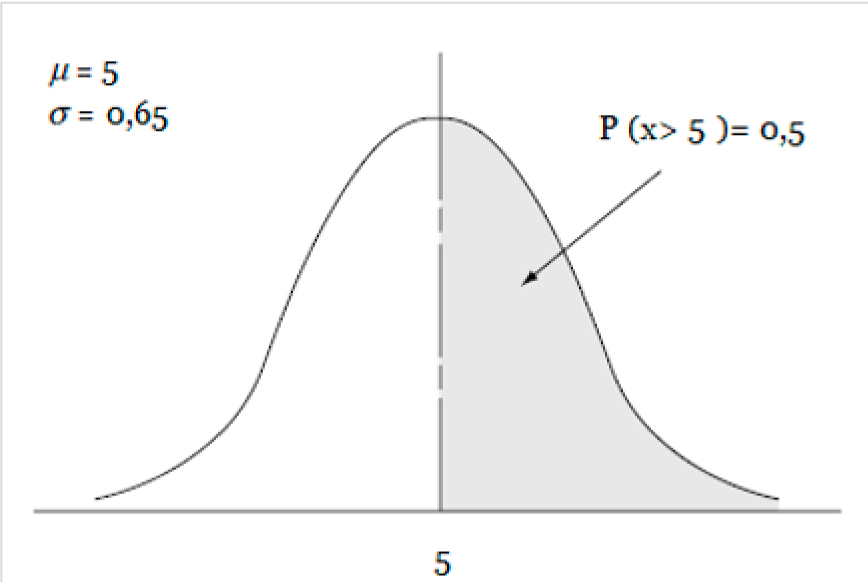
Figura 7. Tabla de distribución de probabilidad normal estándar.

Tal y como se puede ver, la tabla está organizada en valores de *Z* y da los valores de áreas bajo la curva siendo el 0.0 el que coincide con la media. Tales valores son solo las áreas de la mitad de la curva, pero debido a que la distribución es simétrica se puede usar para calcualr probabilidades que involucren el otro lado de la distribución.

Ejemplo 5. Aplicaciones de la distribución normal

Las horas precisas para la maquetación de unos artículos periodísticos en varios días siguen una distribución normal en la que la media es 5 horas y la desviación 0,65.

A. ¿Cuál es la probabilidad de que un día se hayan necesitados más de 5 horas?



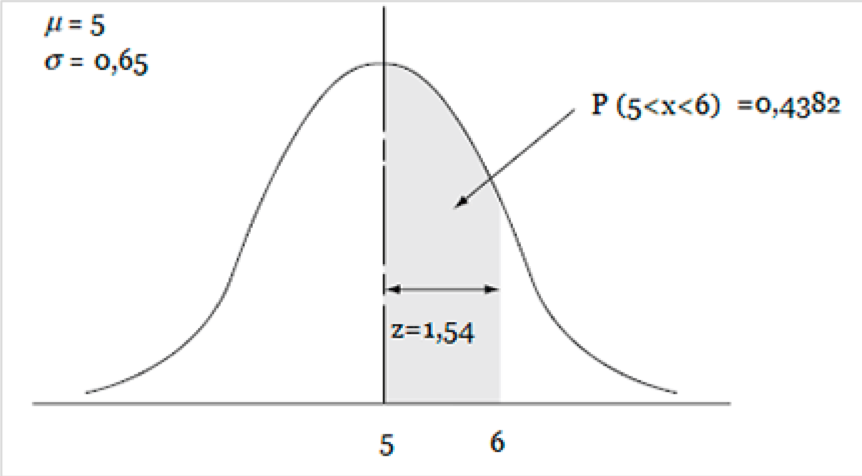
La probabilidad es 0,5 ya que la mistad del área bajo la curva se halla a ambos lados de la media.

B. ¿ Cuál es la probabilidad de que de un día se hayan necesitado entre 5 y 6 horas?

Hay que calcular el valor de Z:

$$z = \frac{x - \mu}{\sigma} = \frac{6 - 5}{0,65} = 1,54\ desviaciones\ estándar$$

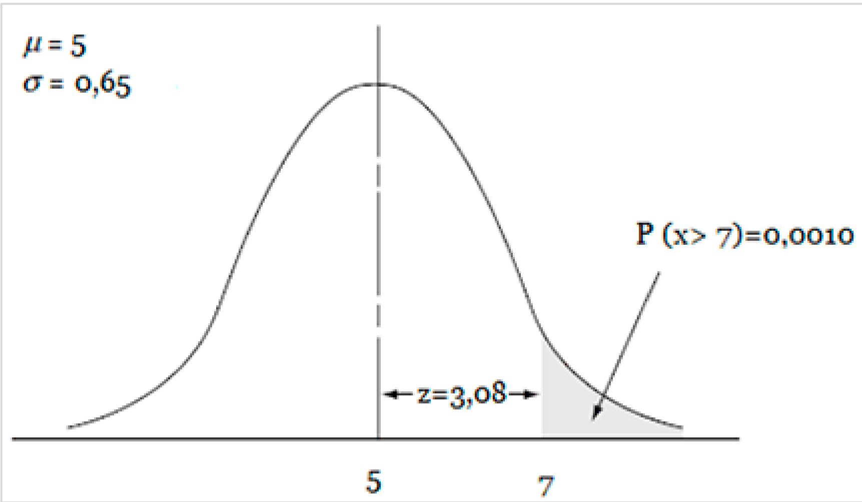
Hay que buscar en la tabla el valor de z=1,54. Para ello miramos en la columna de z hasta la fila con valor 1.5 y hacemos coincidir con la columna 0,04. La probabilidad es 0,4382.



C. ¿Cuál es la probabildad de que un día se hayan requerido 7 horas?

$$z = \frac{x-\mu}{\sigma} = \frac{7-5}{0,65} = 3,08 \; desviaciones \; estándar$$

Comparando el valor de z en la tabla se tiene que la probabilidad es 0,4990. Sin embargo, este dato muestra la probabilidad de que los datos se hallen entre 5 y 7 horas tal y como se ha visto en el apartado B con 6 horas. Tal y como muestra la siguiente figura, la probabilidad de que en un día se hayan requerido más de 7 horas es 0,5 -0,4990, es decir, 0,0010. Lo cual tiene más lógica que pensar que tal probabilidad fuese 0,4990, ya que 7 se aleja bastante de la media.

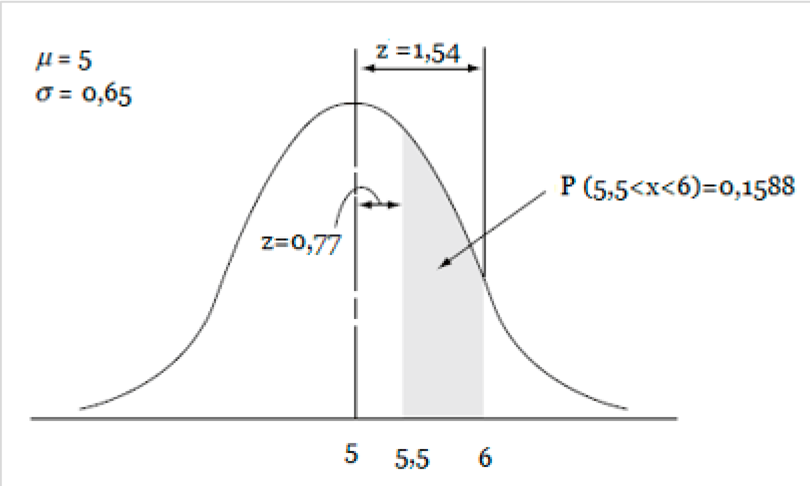


D. ¿Cuál es la probabilidad de que un día haya costado entre 5,5 y 6 horas?

Puesto que se trata de dos datos que no coinciden con la media hay que calcular dos valores de z:

$$z = \frac{x-\mu}{\sigma} = \frac{6-5}{0,65} = 1,54 \qquad\qquad z = \frac{x-\mu}{\sigma} = \frac{5,5-5}{0,65} = 0,77$$

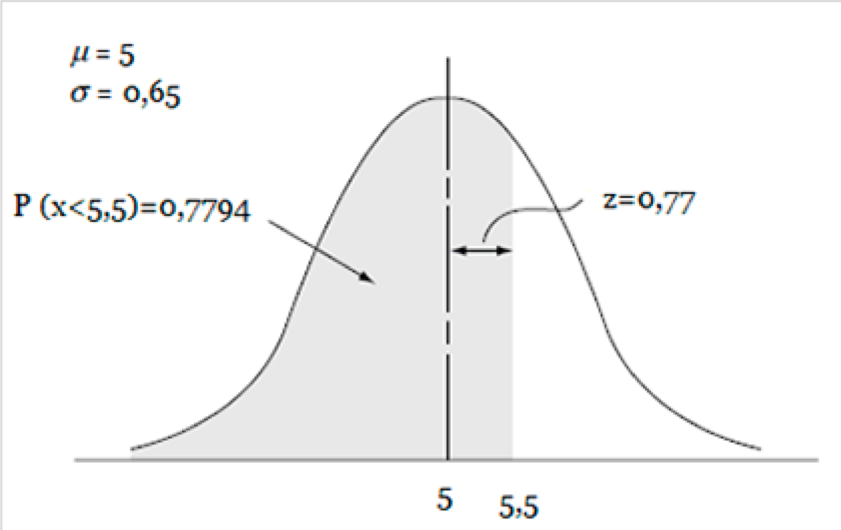
z=1,54 corresponde a una probabilidad de 0,4382 y z=0,77 a una probabilidad de 0,2794. La diferencia entre el área desde la media hasta 6 y el área desde la media hasta 5,5 da el área entre 5,5 y 6 → A=0,4382-0,2794=0,1588, tal y como se muestra a continuación:



E. ¿Cuál es la probabilidad de que un día se haya tardado menos de 5,5 horas?

$$z = \frac{x-\mu}{\sigma} = \frac{5,5-5}{0,65} = 0,77$$

z=0,77 corresponde a una probabilidad de 0,2794. La probabilidad de que se haya tardado menos de 5,5 horas es 0,5+0,2794=0,7794.

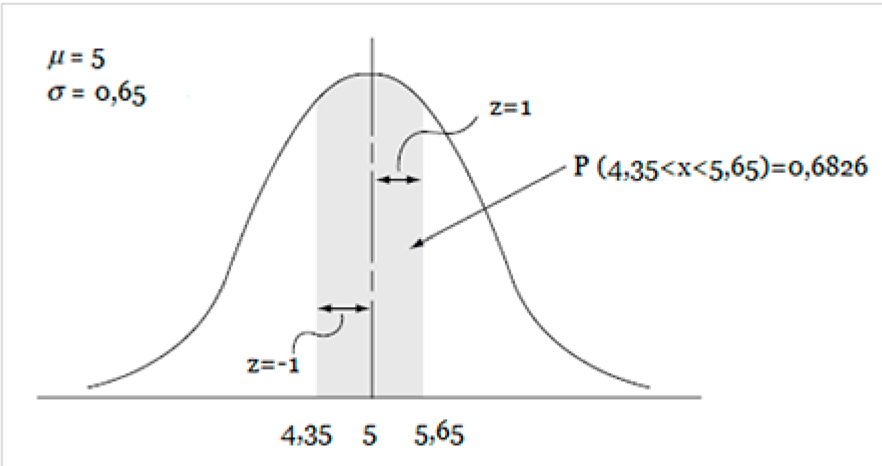


¿Cuál es la probabilidad de que se hayan requerido entre ± 1 σ horas?

Tales valores son 4,35 y 5,65.

$$z = \frac{x-\mu}{\sigma} = \frac{4,35-5}{0,65} = -1 \qquad\qquad z = \frac{x-\mu}{\sigma} = \frac{5,65-5}{0,65} = 1$$

El área que corresponde a z=-1 (se toma 1) es 0,3413 y para z=1 es 0,3413, el área total es 0,3413+0,3413=0,6826.



Las distribuciones de muestreo normales

Una distribución de muestreo es una distribución de probabilidad de las medidas de tendencia central de un conjunto de muestras, generalmente las medias. Cuando llevamos a cabo un muestreo para tomar diferentes muestras de una población, cada una tiene un valor medio y una desviación estándar. La distribución de todas esas medias es la distribución de muestreo de la media. Del mismo modo que hay distribuciones de muestreo de medias, también hay distribuciones de muestreo de proporciones en las que en vez de tomar los valores medios, se toman los valores de las proporciones.

Nos centraremos en las distribuciones de muestreo de la media las cuales vienen definidas por la media y la desviación estándar, también llamado error estándar. El error estándar da información de la magnitud del error y de la precisión que supone el hecho de tomar ese estadístico a la hora de hacer inferencia en la población.

Cabe destacar que una distribución de medias que sea poco extendida, es decir, que tenga valores menores de errores estándar, supone una mejor estimación para hallar la media poblacional.

Ejemplo 6. Distinción entre una muestra de una distribución muestral y de una población

Se desea saber las personas que quieren asistir a ver una película el día del estreno. Para ello se eligieron 100 localidades y se entrevistó a 200 personas por localidad, ¿esa muestra es de la población o de una distribución?

Se trata de una muestra de una distribución muestral de las medias de muestras de tamaño 200, obtenida de la población de personas en las localidades.

Si en una población con media μ y desviación estándar σ se toman diferentes muestras aleatorias, cada una tendrá valores diferentes de media y desviación estándar. En la siguiente figura del libro se observan las diferentes distribuciones de unas muestras aleatorias y la distribución de muestreo de la media obtenida a partir de ellas.

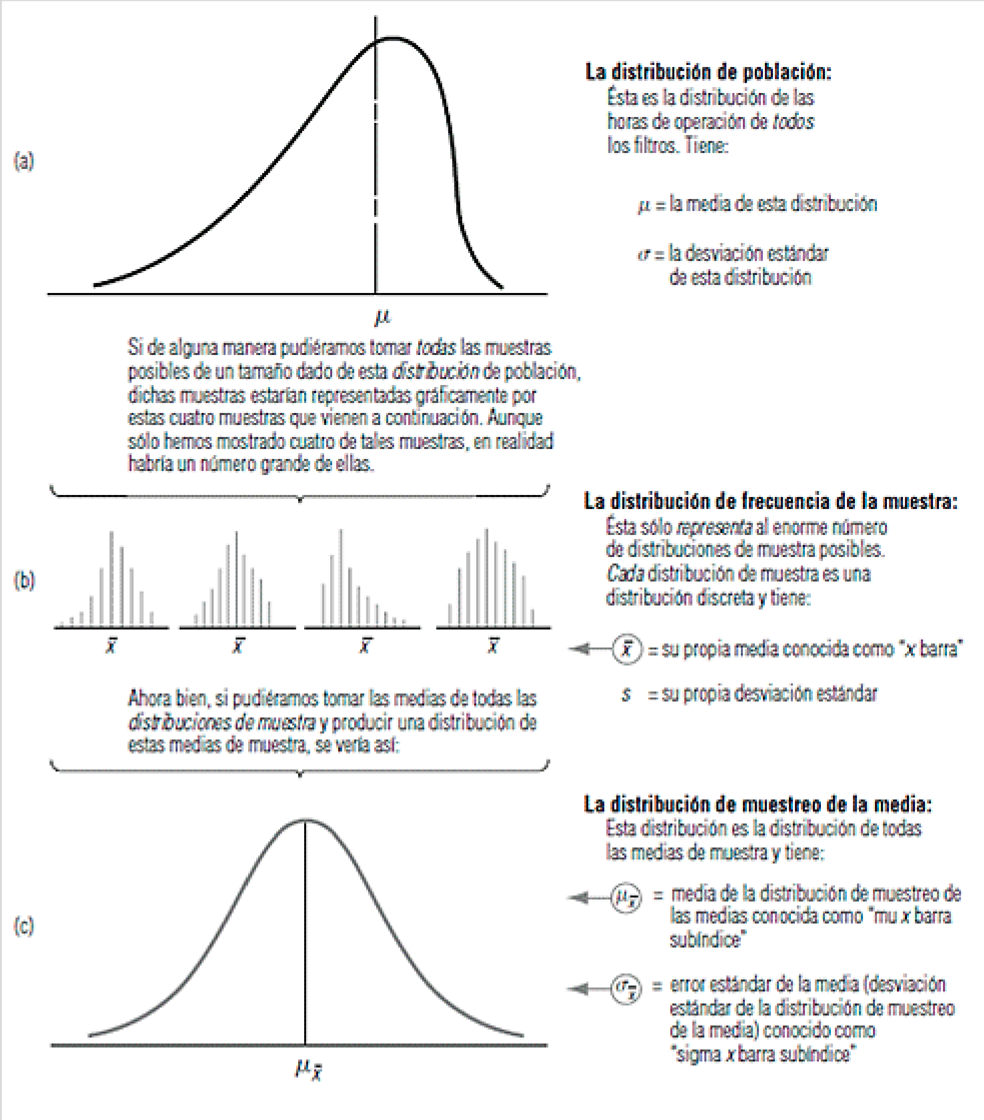


Figura 8. Distribución de muestreo obtenida a partir de diferentes distribuciones de frecuencias. Fuente: *Estadística para la administración y la economía de Levin y Rubin.*

La distribución de muestreo viene definida por el valor esperado de la media $\mu_{\bar{x}}$, promedio de todas las medias muestrales y por la desviación estándar de la media de todas las desviaciones estándar $\sigma_{\bar{x}}$.

Supongamos una población distribuida normalmente en la que la media es 300 y la desviación estándar es 60. Se toman muestras de 15 en 15 y se calculan las medias muestrales las cuales podrían por debajo o por encima de 300 con igual proporción. Se demuestra que el valor esperado de la media de la distribución muestral es igual al valor de la media poblacional:

$$\mu_{\bar{x}} = \mu$$

6.

Por el contrario, el error estándar de la media o desviación estándar de la distribución de muestreo de la media es menor que la desviación estándar de los elementos individuales de la población puesto que la diferencia es menor si se trata de las medias de las muestras en vez de los datos individuales de la población. En la siguiente figura se puede apreciar lo explicado.

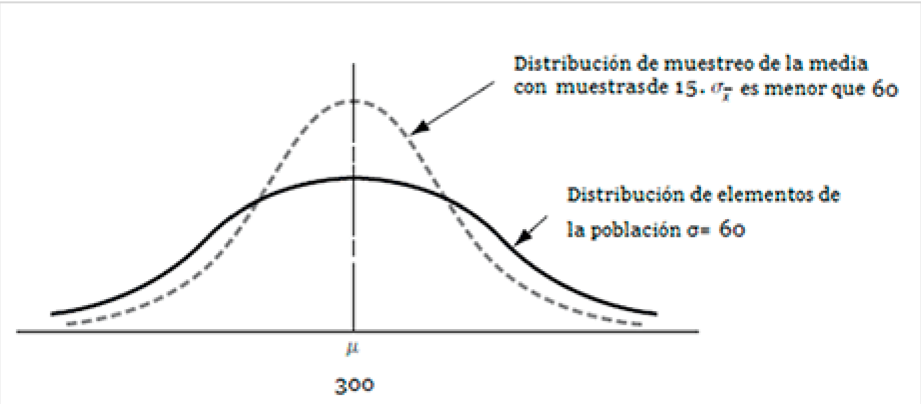


Figura 9. Demostración de que la desviación estándar de la distribución es menor que la de la población.

Si las muestras se tomasen de 30 en 30, la desviación estándar de la población no variaría, pero sí lo haría la dispersión entre las medias de las muestras, ya que, al haber menos muestras a tomar, el valor sería menor. Este hecho se puede ver en la siguiente figura.

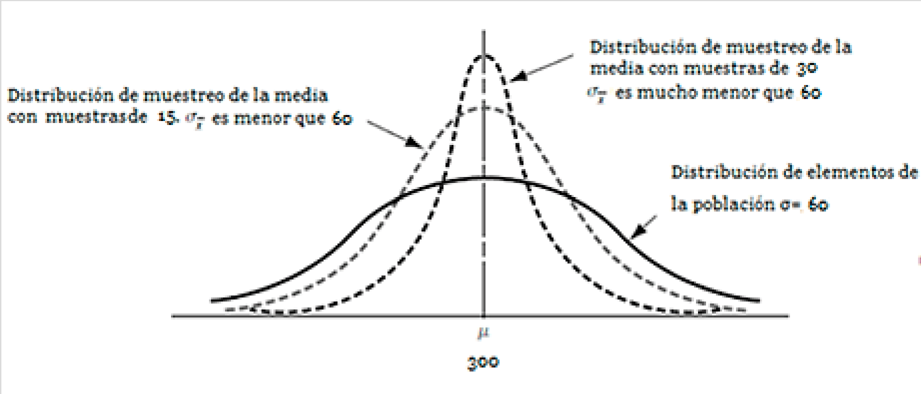


Figura 10.Demostración de cómo disminuye la desviación estándar de la distribución muestral cuando se reduce el número de muestras.

La desviación estándar de una población viene dada por la siguiente expresión:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

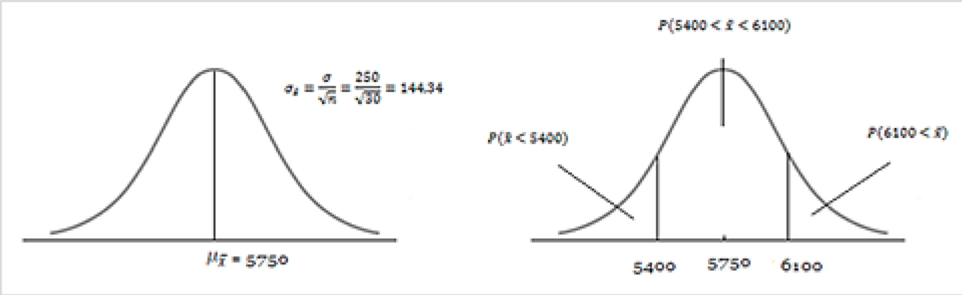
7.

Ejemplo 7. La población de personas que acude a un restaurante de comida rápida diariamente se distribuye de forma normal con una media de 750 y una desviación estándar de 120. Si para hacer un determinado análisis se toman 20 muestras de 15 personas cada una, ¿cuál es la media y la desviación estándar de la distribución muestral?

$$\mu_{\bar{x}} = \mu = 750$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{15}} = 31$$

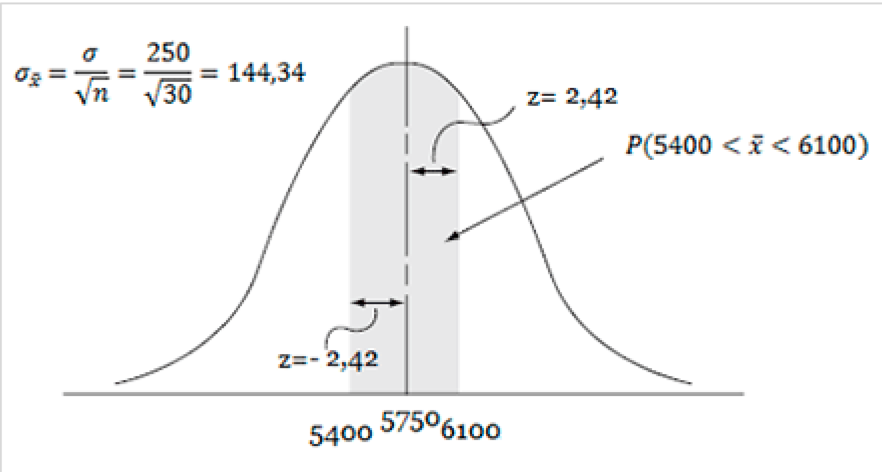
Ejemplo 8. Un periodista quiere escribir un artículo sobre la cantidad de residuos reciclados en una comunidad en la que la media es 5750 kg y la desviación estándar 250. Para estudiar cómo se distribuye el reciclaje en las diferentes regiones toma muestras de 30 empresas de reciclaje y considera que las medias muestrales serán buenas estimaciones de la media poblacional si difieren de ella en ± 350 toneladas. ¿Cuál es la probabilidad de que la media muestral se halle a ese valor de la media poblacional?



Lo que se pretende saber es la probabilidad de que los valores de la media estén entre 5400 y 6100.

$$z = \frac{x-\mu}{\sigma} = \frac{5400-5750}{144{,}34} = -2{,}42 \qquad z = \frac{x-\mu}{\sigma} = \frac{6100-5750}{144{,}34} = 2{,}42$$

Comparando en la tabla de distribucion de probabilidad normal estándar resulta que la probabilidad total es 0,4922+0,4922=0,9844. Tal porcentaje es bastate elevado por lo que la estimación es adecuada.



7.5. La regla 68/95/99,7

Tal y como se ha mencionado anteriormente, en cualquier tipo de distribución normal e independendientemente de los valores de las media y la desviación estándar de la población, se cumple que:

- El 68% aproximadamente de los datos caen a ±1 desviacion esátandar de la media.
- El 95% aproximadamente de los datos caen a ±2 desviaciones estándar de la media.
- El 99,7 % aproximadamente de los datos caen a ±3 desviaciones estándar de la media.

Esto se conoce como la regla 68/95/99,7. Si por ejemplo, las alturas de las chicas de una clase siguen una distribución normal en la que la media es 1,65 y la desviación estándar 0,05 m, se puede decir que el 95 % de las chicas miden entre 1,65-0,10 y 1,65+0,1, es decir, el 95% de las chicas miden entre 1,55 y 1,75.

Ejemplo 9. Las alturas de unos jugadores de baloncesto siguen una distribución normal con media 1,77 y desviación estándar 0,07.

A. ¿Qué porcentaje de jugadores mide más 1,91 m?

» El intervalo para una desviacion estándar es:
  1,77 -0,07 hasta 1,77+0,07, es decir, 1,70-1,84

» El intervalo para dos desviaciones estándar es:
  1,77-0,14 hasta 1,77+0,141, es decir, 1,63-1,91

» El intervalo para tres desviaciones estándar es:
  1,77-0,21 hasta 1,77+0,21, es decir, 1,56-1,98

El valor 1,91 está en el límite superior de dos desviaciones estándar. Este intervalo recoge al 95% de los datos, por lo que los datos que están fuera supononen el restante 5%. La mitad de este 5% es para los datos menores a los del intervalo y la mitad para los mayores, por lo que el porcentaje de que la altura sea superior a 1,91 m es 2,5%.

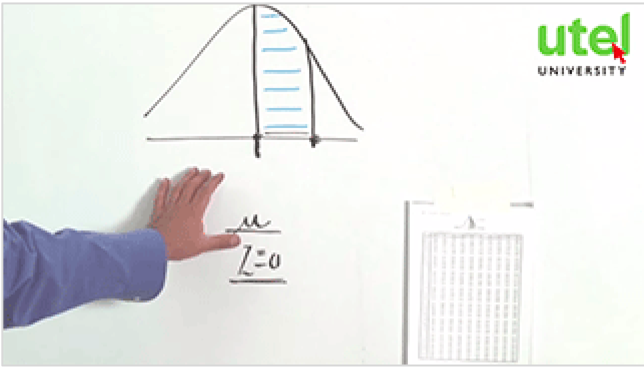
B. ¿Qué porcentaje de jugadores mide menos de 1,70?

1,70 se encuentra en el límite del primer intervalo. El 68% de los datos se hallan entre 1,70 y 1,84 m por lo que el restante 32% estará fuera de estos límites. El 16% de los jugadores mide menos de 1,70m.

Lo + recomendado

No dejes de ver…

Estadística y probabilidad. Tabla de distribución normal



Este vídeo muestra un tutorial para hallar el área bajo la curva en una distribución normal.

Accede al vídeo a través del aula virtual o desde la siguiente dirección web:
http://www.utel.edu.mx/blog/infografias-utel/estadistica-y-probabilidad-tabla-de-distribucion-normal/

+ Información

A fondo

Desigualdad global. La distribución del ingreso en 141 países

Este artículo muestra el uso del índice de Gini en el mundo real.

Accede al artículo a través del aula virtual o desde la siguiente dirección web: http://www.unicef.org/socialpolicy/files/Desigualdad_Global.pdf

Bibliografía

Arias, L., Portilla, L. M. y Bernal, M. E. (2008). Los costos y su manejo con el control estadístico de procesos, con ayuda de la distribución normal. *Scientia et Technica*, 38, 259-264.

Test

1. El coeficiente de asimetría de los siguientes datos es:

3	4	5	7	8	9	12	3	13	15	7	6	3	5	8	9

A. 0.
B. 0,635.
C. 0,478.
D. 1,59.

2. Los siguientes datos muestran una distribución:

3	4	5	7	8	9	12	3	13	15	7	6	3	5	8	9

A. Platicúrtica.
B. Normal.
C. Mesocúrticia.
D. Leptocúrtica.

3. El Índice de Gini de estos datos es:

3	4	5	7	8	9	12	3	13	15	7	6	3	5	8	9

A. 0,53.
B. 0.
C. 0,27.
D. 1.

4. En una distribución positivamente sesgada:

A. La mediana está en el medio
B. La media está a la derecha
C. G1>0
D. Todas son ciertas

5. En una distribución negativamente sesgada:

A. La moda está en el medio.

B. G2<0.

C. Está algo concentrada.

D. Todas son falsas.

6. La forma de una distribución da información de si:

A. Es simétrica o no.

B. Tiene apuntamiento o no.

C. Hay igualdad entre los datos o no.

D. A y b son ciertas.

7. La distribución normal:

A. En ella coinciden la media, la mediana y la moda.

B. Es unimodal.

C. Hay múltiples distribuciones normales.

D. A y b son ciertas.

8. La regla 68/95/99,7:

A. Permite calcular cualquier probabilidad bajo la curva

B. Solo permite calcular determinadas probabilidades bajo la curva.

C. Se basa en la desviación estándar.

D. A y c son ciertos.

9. En una distribución normal, ¿cuál es el área que cae a la izquierda de Z= 1,43?

A. 0,0764.

B. 0,8765.

C. 0,9236.

D. No se puede saber.

10. Una población tiene una media de 200 y una desviación estándar de 15. Si se toman muestras de 10 en 10, señale la opción correcta:

A. La media de la distribución muestral es 200 y la desviación estándar 15.
B. La probabilidad de que la media muestral se halle a ±5 de la media poblacional es 0,2586.
C. La probabilidad de que la media muestral se halle a ±5 de la media poblacional es 0,7062.
D. A y c son ciertas.

