DISTRIBUCIONES DE PROBABILIDAD

Una muestra es una colección de mediciones seleccionadas de alguna fuente o población más grande. Mediante el uso de métodos estadísticos, podemos analizar los datos de una muestra y sacar ciertas conclusiones sobre la población total.

Una distribución de probabilidad es un modelo matemático que relaciona el valor de la variable con la probabilidad de ocurrencia de ese valor en la población. En otras palabras, podríamos visualizar el grosor de la capa de pintura de una pieza mecánica como una variable aleatoria, porque toma diferentes valores en la población de acuerdo con algún mecanismo aleatorio, y luego la distribución de probabilidad del grosor de la capa describe la probabilidad de que ocurra cualquier valor de grosor de capa en el población.

Hay dos tipos de distribuciones de probabilidad.

1. Distribuciones continuas. Cuando la variable que se mide se expresa en una escala continua, su distribución de probabilidad se denomina distribución continua. La distribución de probabilidad del espesor de la capa metálica es continua.

2. Distribuciones discretas. Cuando el parámetro que se está midiendo solo puede tomar ciertos valores, como los enteros 0, 1, 2,. . . , la distribución de probabilidad se llama distribución discreta. Por ejemplo, la distribución del número de no conformidades o defectos en las placas de circuito impreso sería una distribución discreta.

Algunos ejemplos:

Distribuciones de probabilidad discretas	**Distribuciones de probabilidad continuas**
Número de caras en tres lanzamientos de una moneda.	El peso de cada estudiante de la clase.
Número de burbujas por envase de vidrio que son generadas en un proceso dado..	La temperatura ambiente en esta aula.
Número de obreros que se ausentaron hoy en el segundo turno.	El diámetro de un engrane en pulgadas
Número de productos defectuosos en un lote de 25 productos.	Concentración en gramos de plata de algunas muestras de mineral

Variables aleatorias discretas

Algunas distribuciones de probabilidad de variables aleatorias discretas, son:

- Distribución Binomial
- Distribución Hipergeométrica
- Distribución de Poisson

En nuestro curso solo veremos la distribución de Poisson.

Variables aleatorias continuas

Algunas distribuciones de probabilidad de variables aleatorias continuas, son:

- Distribución Normal.
- Distribución t de Student

En nuestro curso solo veremos la distribución Normal.

Distribución de Poisson

Definición

La distribución de Poisson se puede entender como un caso particular de la Binomial que utilizamos para determinadas distribuciones en las que el cálculo de la probabilidad es engorroso debido bien a que el número de pruebas es excesivamente elevado o bien a que la probabilidad de éxito es excesivamente baja; en ambos casos la media (n*p) es muy pequeña en relación al número de pruebas (n). En estos casos se puede demostrar que la distribución binomial converge, tiende a comportarse, como una distribución de Poisson.

Como regla práctica entenderemos que es aplicable la distribución de Poisson en aquellas binomiales cuya media tenga un valor inferior a 5 y el número de pruebas sea superior a 30.

Algunos ejemplos de fenómenos que se ajustan a una distribución de Poisson son los siguientes:

- el número de accidentes de tráfico en una ciudad durante una semana.
- el número de emergencias que llegan a un servicio de urgencia hospitalaria.
- el número de robos denunciados en un mes en la ciudad de Madrid.
- el número de llamadas telefónicas que llegan a la centralita de una gran empresa en hora punta.

Todos estos casos pueden ser caracterizados como el número de sucesos de un determinado evento en un período de tiempo. Cada uno de estos eventos toma dos posibles valores, éxito o fracaso, equivalente a recibir una llamada o no recibirla, accidentarse o no accidentarse, etc... Además, la probabilidad del éxito es muy pequeña. Es decir, la probabilidad de tener un accidente es muy baja, la probabilidad de que una persona llama a la centralita de la gran empresa es mínima.

Y, no debemos de olvidar, que n, número de personas que hay en una ciudad, número de personas que circulan en coche, etc... es muy grande. Todo esto nos lleva a que una buena parte de fenómenos de esta naturaleza siguen una distribución del tipo Poisson. Es por ello que esta distribución, también denominada "de los sucesos raros" es particularmente útil para resolver problemas de colas y líneas de espera, temas ambos que se podrán estudiar en otras asignaturas del área de métodos cuantitativos en economía.

Se puede demostrar que la forma que adopta la función de cuantía para una variable aleatoria (X) con una distribución de Poisson de parámetro 8 es:

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

donde λ = n*p,

x! es el factorial de x

e es el número e.

La forma que adopta la función de distribución es

$$F(x) = \sum_{r \leq x} P(X = r) = \sum_{r \leq x} \frac{e^{-\lambda}\lambda^{r}}{r!}$$

Por tanto, la media y la varianza de esta distribución son idénticas e iguales a λ=np.

Caso Práctico 1

La probabilidad de que un producto salga defectuoso es de 0.012. ¿Cuál es la probabilidad de que entre 800 productos ya fabricados hayan 5 defectuosos?

En este ejemplo vemos nuevamente la probabilidad *p* menor que 0.1, y el producto *n * p* menor que 10, por lo que aplicamos el modelo de distribución de Poisson:

$$\boxed{P\,(x = 5) \;=\; e^{-9{,}6} \;*\; \frac{9{,}6^{5}}{5\,!}}$$

El resultado es P (x = 5) = 0.04602

Por lo tanto, la probabilidad de que haya 5 productos defectuosos entre 800 recién producidos es de 4.6%.

Caso Práctico 2

En estudios anteriores, en una agencia bancaria, en promedio llegan 3 clientes a una ventanilla para ser atendido durante la hora del almuerzo. Si en la actualidad queremos hacer modificaciones en la ventanillas, una de las preguntas que se pueden hacer los del depto. de Mercadeo es:

¿Cuál es la probabilidad de que lleguen dos clientes en un minuto dado?

En el enunciado el promedio es de 3 clientes por minuto; $\lambda=3$; al preguntar por la probabilidad de que 2 clientes lleguen en un minuto dato se quiere calcular x=2

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad \equiv \quad P(X = 2) = \frac{e^{-3}3^{2}}{2!}$$

$$P(X = 2) = \frac{(0.049787)(9)}{2x1} = 0.224$$

La probabilidad que lleguen 2 clientes por minuto es 22.4%

Caso Práctico 3:

Si un banco recibe en promedio 6 cheques sin fondo por día, ¿cuáles son las probabilidades de que reciba:

a) cuatro cheques sin fondo en un día dado

b) 10 cheques sin fondos en cualquiera de dos días consecutivos?

Solución:

a) x = variable que nos define el número de cheques sin fondo que llegan al banco en un día cualquiera = 0, 1, 2, 3, ..., etc, etc.

λ = 6 cheques sin fondo por día e = 2.718

$$p(x=4,\lambda=6)=\frac{(6)^4(2.718)^{-6}}{4!}=\frac{(1296)(0.00248)}{24}=0.13392$$

b) x= variable que nos define el número de cheques sin fondo que llegan al banco en dos días consecutivos = 0, 1, 2, 3, ..., etc., etc

λ = 6 x 2 = 12 cheques sin fondo en promedio que llegan al banco en dos días consecutivos

$$p(x=10,\lambda=12)=\frac{(12)^{10}(2.718)^{-12}}{10!}=\frac{(6.1917364\mathrm{E}10)(0.000006151)}{3628800}=0.104953$$

Nota: λ siempre debe de estar en función de x siempre o dicho de otra forma, debe "hablar" de lo mismo que x.

Variables aleatorias continuas

LA DISTRIBUCION NORMAL

La distribución normal es la más común entre todas las distribuciones de probabilidad utilizadas en Estadística y tiene importantes aplicaciones en la modelización de variables estadísticas asociadas a los elementos de una población. Por ejemplo, las medidas físicas del cuerpo humano en una población, las características psíquicas medidas por test de inteligencia o personalidad, las medidas de calidad en muchos procesos industriales, los errores de las observaciones astronómicas; siguen distribuciones normales.

Una justificación de la frecuente aparición de la distribución normal es el **teorema central del límite** que establece que cuando los resultados de un experimento sean debidos a un conjunto muy grande de causas independientes, que actúan sumando sus efectos, siendo cada efecto individual de poca importancia respecto al conjunto, es esperable que los resultados sigan una distribución normal.

Diremos que una variable aleatoria sigue una distribución normal con parámetros μ y σ^2 y se representa:

$$X \rightarrow N(\mu, \sigma^2)$$

cuando su función de densidad es de la forma:

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

$$\forall\, x \in \Re \textit{ siendo } \sigma \succ 0 \textit{ y } \mu \in \Re$$

en donde R es el conjunto de los números reales.

Como se puede observar, la familia de distribuciones normales depende de los parámetros μ y σ^2, que coinciden con la media y la varianza respectivamente.

LA DISTRIBUCION NORMAL TIPIFICADA O ESTANDARD

Un caso particular de distribución normal es aquella en la cual la media vale cero y la varianza 1. En este caso decimos que la variable z se distribuye como una variable normal tipificada o standard y la denotamos por

$$z \rightarrow N\,(0,1)$$

En este caso, su función de densidad toma la forma:

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{x^2}{2}} \quad \forall\, x \in \Re$$

En general, se convierte a unidades estándar al restar la media y dividir entre la desviación estándar. Por consiguiente, si x es una unidad seleccionada de una población normal con media μ y varianza σ^2, la unidad estándar equivalente a x es el número z, donde

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

PROPIEDADES

(1) Es simétrica respecto de x = 0, ya que:

$$f(-x) = f(x)$$

(2) Alcanza un máximo en x = 0 y vale

$$f(0) = \frac{1}{\sqrt{2\varpi}}$$

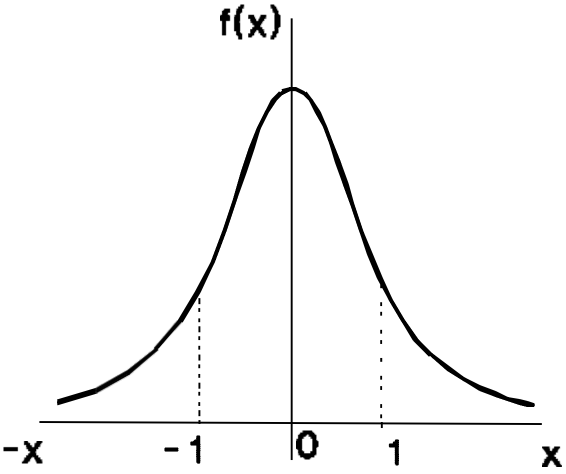
(3) Es creciente para x < 0 y decreciente para x > 0.

(4) Los puntos de abcisas 1 y -1 son de inflexión de la función.

(5) La recta y = 0 es asíntota de la función, pues

$$\lim_{x\to\infty} f(x)=\lim_{x\to-\infty} f(x)=0$$

Su representación gráfica es:



Las áreas debajo de la curva normal estándar (media 0, varianza 1) se han tabulado extensivamente. Una tabla común, denominada tabla normal estándar, o tabla z.

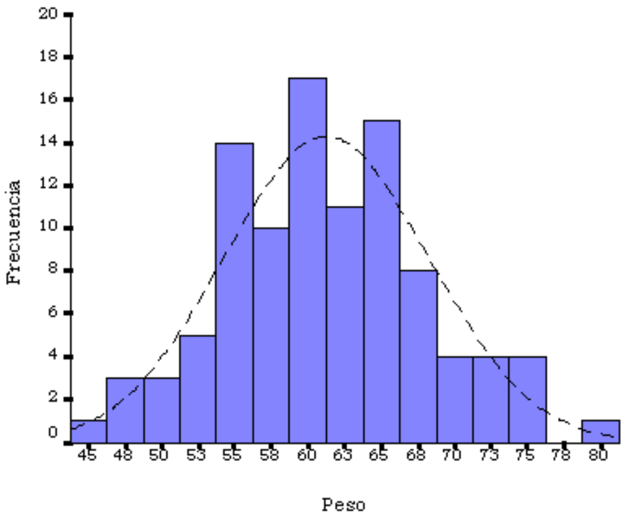
Para determinar las áreas debajo de una curva normal con diferentes media y varianza, se convierten a unidades estándares y se utiliza la tabla z. La tabla proporciona las áreas en la cola izquierda de la curva para valores de z. Es posible calcular otras áreas al sustraer o usando el hecho de que el área total debajo de la curva es igual a 1.

Teorema del Límite Central

Nos indica que, bajo condiciones muy generales, según aumenta la cantidad de datos, la distribución de la suma de variables aleatorias tendera a seguir hacia una distribución normal.

En otras palabras el Teorema del Límite Central garantiza una distribución normal cuando el tamaño de la muestra es suficientemente grande.

En el siguiente histograma podemos observar la distribución de frecuencias por peso de acuerdo a la edad. De acuerdo a este teorema según aumenten la cantidad de dato se podrá trazar una curva que tome cada vez más formación en forma campana.



Caso Práctico 1

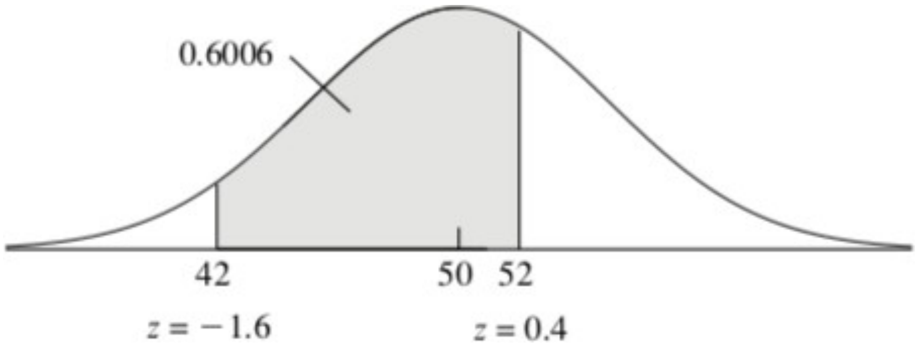
Los tiempos de vida de las baterías en cierta aplicación se distribuyen normalmente con media de 50 horas y desviación estándar de cinco horas. Determine la probabilidad de que se elija aleatoriamente una batería que dure entre 42 y 52 horas.

Solución

Sea X el tiempo de vida de una batería elegida aleatoriamente. Entonces $X = N(50, 5^2)$. La figura muestra la función de densidad de probabilidad de la población $N(50, 5^2)$. El área sombreada representa $P(42 < X < 52)$, la probabilidad de que una batería seleccionada de forma aleatoria tenga una duración entre 42 y 52 horas. Para calcular esta área, se hará uso de la tabla z. Primero se necesita convertir las cantidades 42 y 52 a unidades estándar. Se tiene

$z= (42-50) / 5 =-1.60$ y $z= (52-50) / 5 =0.40$

De la tabla *z*, el área a la izquierda de *z* =1.60 es 0.0548, y el área a la izquierda de *z* 0.40 es 0.6554. La probabilidad de que una batería tenga tiempo de vida entre 42 y 52 horas es 0.6554 - 0.0548= 0.6006.



Caso Práctico 2:

Un proceso fabrica cojinetes de bolas cuyos diámetros se distribuye normalmente con media de 2.505 cm y desviación estándar de 0.008 cm. Las especificaciones requieren que el diámetro esté dentro del intervalo 2.5 $\pm$0.01 cm. ¿Qué proporción de cojinetes de bolas cumple con la especificación?

Solución

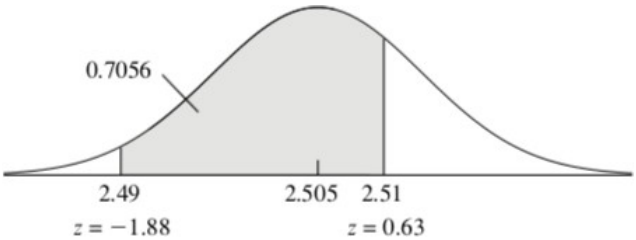
Sea *X* el diámetro de un cojinete de bolas seleccionado aleatoriamente. Entonces *X*= *N*(2.505, 0.008^2). La figura presenta la función de densidad de probabilidad de la población *N*(2.505, 0.008^2). El área sombreada representa *P*(2.49< *X* <2.51), que es la proporción de cojinetes de bolas que cumplen con la especificación.

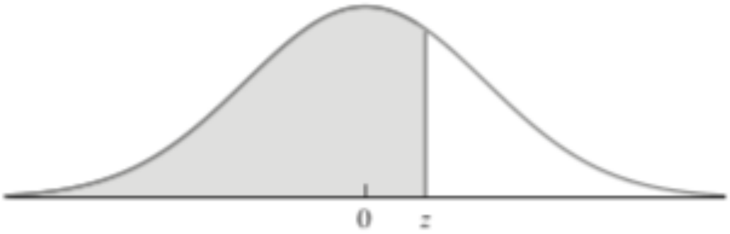
Se calcula los *z* de 2.49 y 2.51:

z=(2.49 – 2.505) / 0.008 = - 1.88 y z= (2.51 – 2.505) / 0.008 = 0.63

El área a la izquierda de *z*= - 1.88 es 0.0301. El área a la izquierda de *z*= 0.63 es 0.7357.

El área entre *z*= 0.63 y *z*=- *1*.88 es 0.7357- 0.0301=0.7056. Aproximadamente 70.56% de los diámetros satisface la especificación





z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

Veamos como calculamos $z(|x|<1)$ con el uso de esta tabla. Esta probabilidad es la probabilidad de que la variable tome valores entre –1 y 1. Dado que sabemos que la variable normal es simétrica, la probabilidad entre 0 y 1 es la misma que la que hay entre –1 y 0. La probabilidad entre 0 y 1 la calculamos directamente de la tabla buscando el cruce entre la fila 1.0 y la columna 0.00. Por tanto vale 0.8413. En consecuencia, la probabilidad pedida será igual a:

$$z(|x|<1)=z(-1,0)+z(0,1)=2*z(0,1)=2*0.8413$$

No existe una única tabla de la distribución N(0,1). Todas ellas nos dan la misma información pero de distinta forma. **Para los ejemplos planteados y lo detallado en el curso usamos aquella tabla que indica el área de z a menos infinito.**

Hasta ahora hemos calculado la probabilidad que hay entre dos valores posibles de la variable, sin embargo otras veces tendremos que calcular el valor x_1 conocida su probabilidad. Es decir, dado $z(x_1)$,por ejemplo $z(x_1) = 0.90$, que lo representamos como $z_{0.90}$ tendremos que buscar en las tablas el valor de x_1 (la abcisa) que nos da una probabilidad (área) de 0.90.

En este caso vemos que no existe en las tablas una probabilidad exactamente igual a 0.90, lo que tenemos es:

A z=1.28 le corresponde una función de distribución de 0.8997

A z=1.29 le corresponde una función de distribución de 0.90147

Para calcular el valor exacto de x_1 tenemos que interpolar y lo hacemos empleando el método:

Paso1.-Calculamos las diferencias siguientes:

Diferencia de abcisas	Diferencia de probabilidad	
1.29	0.90147	0.90147
<u>-1.28</u>	<u>-0.8997</u>	<u>-0.90</u>
0.01	0.00177	0.00147

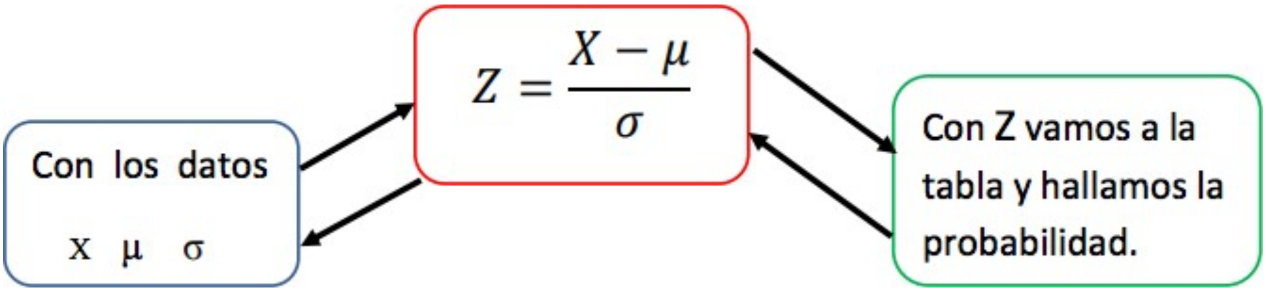
Paso 2.-Hacemos la siguiente proporción:

Diferencia de abcisas	Diferencia de probabilidad	
0.01	0.00177	
		x = 0.0083051
x	0.00147	

Paso 3.- $x_1 = 1.29 - 0.0083051 = 1.28169$;

0.90 = z(1.22)

En resumen:



Caso Práctico 3:

En una ciudad se estima que la temperatura máxima en el mes de junio sigue una distribución normal, con media 23° y desviación típica 5°. Calcular el número de días del mes en los que se espera alcanzar máximas entre 21° y 27°.

Solución

En una ciudad se estima que la temperatura máxima en el mes de junio sigue una distribución normal, con media 23° y desviación típica 5°.

Calcular el número de días del mes en los que se espera alcanzar máximas entre 21° y 27°.

Utilizando la formula $Z = \frac{X - \mu}{\sigma}$, vamos a sustituir el valor de la media (23), y la desviación típica (5).

$$P(21 \leq X \leq 27) = P\left(\frac{(21-23)}{5} \leq Z \leq \frac{(27-23)}{5}\right)$$

$$= P(-0.4 \leq Z \leq 0.8)$$

$$= P(Z \leq 0.8) - P(Z \geq -0.4)$$

$$= P(Z \leq 0.8) - (1 - P(Z \leq 0.4))$$

Buscamos los valores correspondientes en la tabla de distribución normal:

$$P(Z \leq 0.8) = 0.7881 \qquad \text{y} \qquad P(Z \leq 0.4) = 0.6554$$

Por lo tanto

$$30 \cdot P(21 \leq X \leq 27) = 30 \cdot P\left(\frac{(21-23)}{5} \leq Z \leq \frac{(27-23)}{5}\right)$$

$$\begin{aligned} &= (30)(0.7881 - (1 - 0.6554)) \\ &= (30)(0.4425) \\ &= 13 \end{aligned}$$

Esto quiere decir, que en todo el mes, solo 13 días alcanzaran temperaturas entre 21 y 27 grados.