4.1. ¿Qué ES UNA DISTRIBUCIÓN DE LA PROBABILIDAD?

INTRODUCCIÓN.

Una distribución de probabilidad indica toda la gama de valores que pueden representarse como resultado de un experimento. Una distribución de probabilidad es similar al distribución de frecuencias relativas .Si embargo, en vez de describir el pasado, describe la probabilidad que un evento se realice en el futuro, constituye una herramienta fundamental para la prospectiva, puesto que se puede diseñar un escenario de acontecimientos futuros considerando las tendencias actuales de diversos fenómenos naturales.

Las decisiones estadísticas basadas en la estadística inferencial son fundamentales en la investigación que son evaluadas en términos de distribución de probabilidades.

En el presente trabajo, se estudia de manera ágil los diverso tipos de distribución probabilística, caracterizaremos cada distribución, la fundamentación matemática de los diversos resultados no se enfocaran en el presente trabajo; sólo me limitaré al estudio descriptivo de la distribución de probabilidades discretas.

¿Qué es una distribución de probabilidad?

Muestra todos los resultados posibles de un experimento y la probabilidad de cada resultado.

¿Cómo generamos una distribución de probabilidad?

Supongamos que se quiere saber el numero de caras que se obtienen al lanzar cuatro veces una moneda al aire?

Es obvio que, el hecho de que la moderna caiga de costado se descarta.

Los posibles resultados son: cero caras, una cara, dos caras, tres caras y cuatro caras.

Si realizamos el experimento obtenemos el siguiente espacio muestral:

$$\Omega = \{cccc, cccs, ccsc, ccss, cscc, cscs, cssc, csss, sccc, sccs, scsc, scss, sscc, sscs, sssc, ssss\}$$

$$n(\Omega) = 16$$

NUMERO DE CARAS	FRECUENCIA	DISTRIBUCIÓN DE PROBABILIDADES
0	1	1/16
1	4	4/16
2	6	6/16
3	4	4/16
4	1	1/16

OBSERVACION

1. La probabilidad de cada resultado especifico va desde cero hasta uno inclusive
2. $$\sum_{k=0}^{4} P(X_k) = 1$$

2 VARIABLE ALEATORIA.-Cantidad que es resultado de un experimento y debido al azar, puede tomar valores diferentes.

Variable aleatoria discreta:- Toma valores claramente separados, generalmente se produce por conteo.

2.1Variable aleatoria continua:-Cantidades que toman infinitos valores, dentro de un rango permitido, generándose una distribución de probabilidades continuas.

2.2Media de una Distribución de Probabilidades.-Valor promedio a largo plazo de la variable aleatoria, también es conocido como valor esperado. Esta media es un promedio ponderado, en el que los valores posibles se ponderan mediante sus probabilidades correspondientes de ocurrencia, se calcula con la formula:

$$\mu = \sum XP(X) \quad \text{..............................(1)}$$

Donde P(X) es la probabilidad que puede tomar la variable aleatoria X.

2.3Varianza.- Mide el grado de dispersión de la distribución de probabilidades, siendo la formula:

$$\sigma^2 = \sum (X-\mu)^2 P(X)$$ ....(2)

También se aplica la fórmula:

$$\sigma^2 = \sum XP(X) - \mu^2$$ .... (3)

Desviación Estándar.-Es la raíz cuadrad del varianza, luego:

$$\sqrt{\sigma^2} = \sigma = \sqrt{\sum XP(X) - \mu^2}$$ .... (4)

4.2. VARIABLES ALEATORIAS

- Una <u>variable aleatoria</u> es una variable que toma valores numéricos determinados por el resultado de un experimento aleatorio. No hay que confundir la variable aleatoria con sus posibles valores. Ejemplos:
  - nº de caras al lanzar 6 veces una moneda (valores: 0, 1, 2…)
  - nº de llamadas que recibe un teléfono en una hora
  - tiempo que esperan los clientes para pagar en un supermercado…

- Las variables aleatorias pueden ser <u>discretas</u> o <u>continuas</u>:
  - <u>Discretas</u>: el conjunto de posibles valores es numerable. Suelen estar asociadas a experimentos en que se mide el número de veces que sucede algo.
  - <u>Continuas</u>: el conjunto de posibles valores es no numerable. Puede tomar todos los valores de un intervalo. Son el resultado de medir.

<u>Ejemplo</u>: Ejercicio 15.2 de Peña y Romo

Clasificar como discretas o continuas las siguientes variables aleatorias:

a) nº de páginas de un libro → discreta

b) tiempo que tarda en fundirse una bombilla → continua

c) nº de preguntas en una clase de una hora → discreta

d) cantidad de agua consumida en un mes → continua

En la práctica se consideran discretas aquellas variables para las que merece la pena asignar probabilidades a todos los posibles sucesos elementales.

Distribución de una variable aleatoria

- Sea x una variable aleatoria discreta. Su distribución viene dada por los valores que puede tomar, x_1, x_2, x_3, …, x_k, y las probabilidades de que aparezcan p_1, p_2, p_3, …, p_k. Estas cantidades $p_i = P\{x = x_i\}$ reciben el nombre de función de probabilidad o función de masa.

Ejemplo:

Variable aleatoria x=nº de caras al lanzar tres veces una

moneda

Posibles valores de x: 0, 1, 2 y 3

Lanzar 3 veces moneda:

E={CCC,CCX,CXC,XCC,XXC,XCX,CXX,XXX}

La variable aleatoria x:

- Toma valor 0 cuando ocurre el suceso {XXX}
- Toma valor 1 cuando ocurre el suceso {XXC,XCX,CXX}
- Toma valor 2 cuando {CCX,CXC,XCC}
- Toma valor 3 cuando {CCC}

La función de probabilidad es:

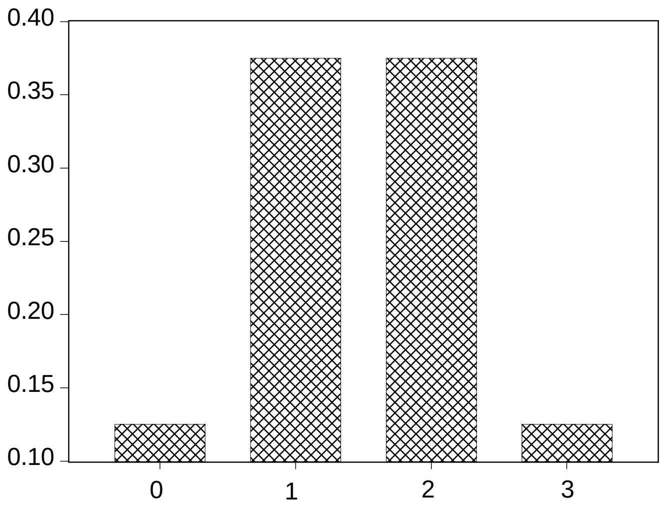
$$p_0 = P\{x = 0\} = 1/8 = 0{,}125$$

$$p_1 = P\{x = 1\} = 3/8 = 0{,}375$$

$$p_2 = P\{x = 2\} = 3/8 = 0{,}375$$

$$p_3 = P\{x = 3\} = 1/8 = 0{,}125$$

Función de probabilidad de *x*:



¿Cuál será la probabilidad de que salgan al menos dos caras?

$$P\{x \leq 2\} = P\{x = 0\} + P\{x = 1\} + P\{x = 2\} = 0{,}125 + 0{,}375 + 0{,}375$$
$$= 0{,}875$$

¿y la probabilidad de que el número de caras esté entre 1 y 2?

$$P\{1 \leq x \leq 2\} = P\{x = 1\} + P\{x = 2\} = 0{,}375 + 0{,}375 = 0{,}75$$

- La probabilidad de que una variable aleatoria *x* tome un valor entre dos cantidades *a* y *b* será:

$$P\{a \le x \le b\} = P\{x = a\} + P\{x = a + 1\} + \ldots + P\{x = b - 1\} + P\{x = b\}$$
$$= \sum_{x_i = a}^{b} P\{x = x_i\}$$

- La función de probabilidad verifica que:

- $p_i = P\{x = x_i\} \ge 0$

- $\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} P\{x = x_i\} = 1$

- La <u>función de distribución</u> o de <u>probabilidad acumulada</u> representa en cada punto x_0 la probabilidad de que la variable tome un valor menor o igual que dicho punto, es decir, $P\{x \le x_0\}$.

Ejemplo: nº caras al lanzar tres veces una moneda

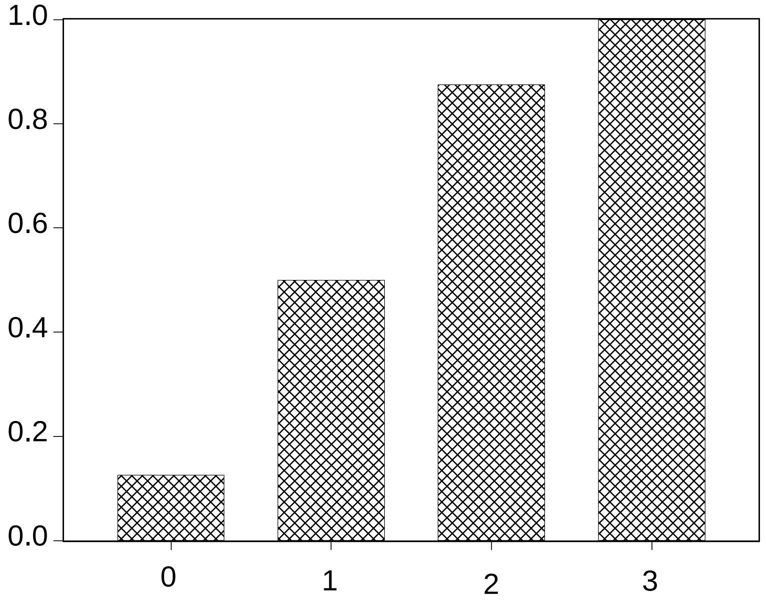
$$P\{x \leq 0\} = P\{x = 0\} = 0{,}125$$

$$P\{x \leq 1\} = P\{x = 0\} + P\{x = 1\} = 0{,}125 + 0{,}375 = 0{,}5$$

$$P\{x \leq 2\} = P\{x = 0\} + P\{x = 1\} + P\{x = 2\} = 0{,}5 + 0{,}375 = 0{,}875$$

$$P\{x \leq 3\} = P\{x = 0\} + P\{x = 1\} + P\{x = 2\} + P\{x = 3\} = 0{,}875 + 0{,}125 = 1$$

Función de distribución de *x*



- Sea *x* una variable aleatoria continua. Si queremos conocer su distribución de probabilidad no nos vale la función de probabilidad empleada con las discretas (cada valor con su probabilidad asociada) porque toma muchos valores. La probabilidad asociada a cada valor es prácticamente nula (la función de distribución es continua).

- Emplearemos la función de densidad. Se interpreta de forma parecida al histograma. Expresa la "densidad" o concentración de probabilidad en cada zona. Expresa las probabilidades por áreas. Sus valores más altos corresponden a zonas en las que es más probable que aparezcan resultados del experimento aleatorio.

Ver Figuras 15.5 y 15.6 de Peña y Romo

Media o esperanza de una variable aleatoria

- La media o esperanza de una variable aleatoria discreta será:

$$E(x) = m_x = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k = \sum_{i=1}^{k} x_i p_i$$

Ejemplo: Ejercicio 15.5 de Peña y Romo

x=resultado de lanzar un dado

La distribución de probabilidad de *x* será:

$$p_1 = P\{x = 1\} = 1/6$$

$$p_2 = P\{x = 2\} = 1/6$$

………………

$$p_6 = P\{x = 6\} = 1/6$$

El valor esperado de *x* será:

$$m_x = \sum_{i=1}^{k} x_i p_i = 1\times\frac{1}{6} + 2\times\frac{1}{6} + 3\times\frac{1}{6} + 4\times\frac{1}{6} + 5\times\frac{1}{6} + 6\times\frac{1}{6} = 3{,}5$$

- La idea de media o esperanza de una variable aleatoria continua es equivalente pero su cálculo es algo más complicado porque requiere emplear el concepto de integral.

- La media de una variable aleatoria puede interpretarse como el valor esperado o medio que toma dicha variable o como el valor central de dicha distribución.

- Propiedades:

  - si *x* e *y* son dos variables aleatorias se cumple que:

    $$m_{x+y} = m_x + m_y$$

  - si *a* y *b* son constantes se cumple que:

    $$m_{ax+b} = am_x + b$$

Ejercicio 15.3 (de Peña y Romo)

Una compañía ha vendido 205 billetes para un avión de 200 plazas.

Sea x la variable aleatoria que expresa el nº de viajeros que va al aeropuerto para viajar en el avión. Su distribución es:

x_i	198	199	200	201	202	203	204	205
p_i	0,05	0,09	0,15	0,20	0,23	0,17	0,09	0,02

a) Hallar la probabilidad de que todos los viajeros que van al aeropuerto tengan plaza.

$$P\{x \le 200\} = P\{x = 198\} + P\{x = 199\} + P\{x = 200\} =$$
$$= 0{,}05 + 0{,}09 + 0{,}15 = 0{,}29$$

b) Obtener la probabilidad de que se quede sin plaza alguno de los viajeros que va al aeropuerto.

$$P\{x > 200\} = P\{x = 201\} + P\{x = 202\} + \ldots + P\{x = 205\} =$$
$$= 0{,}2 + 0{,}23 + 0{,}17 + 0{,}09 + 0{,}02 = 0{,}71$$

$$P\{x > 200\} = 1 - P\{x \le 200\} = 1 - 0{,}29 = 0{,}71$$

c) Calcular el nº esperado de viajeros que acude al aeropuerto.

$$m_x = \sum_{i=1}^{k} x_i p_i = 198 \times 0{,}05 + 199 \times 0{,}09 + 200 \times 0{,}15 + 201 \times 0{,}2 +$$
$$+ 202 \times 0{,}23 + 203 \times 0{,}17 + 204 \times 0{,}09 + 205 \times 0{,}02 =$$
$$= 201{,}44$$

d) ¿Cuál es la probabilidad de que la primera persona de la lista de espera tenga sitio en el vuelo?

$$P\{x \le 199\} = P\{x = 198\} + P\{x = 199\} = 0{,}05 + 0{,}09 = 0{,}14$$

Desviación típica de una variable aleatoria

- La desviación típica de una variable aleatoria es una medida de dispersión de la distribución alrededor de la media. Los valores pequeños indican concentración de la distribución alrededor de la esperanza y los valores grandes corresponden a distribuciones más dispersas.

- El concepto de desviación típica es equivalente en variables aleatorias discretas y continuas, aunque en estas últimas su cálculo es más complicado.

- Si *x* es una variable aleatoria discreta su desviación típica viene dada por:

$$\sigma_x = DT(x) = \sqrt{\sum_{i=1}^{k} \left(x_i - m_x\right)^2 p_i} = \sqrt{\sum_{i=1}^{k} x_i^2 p_i - m_x^2}$$

  y su varianza será:

$$\sigma_x^2 = V(x) = \sum_{i=1}^{k} \left(x_i - m_x\right)^2 p_i = \sum_{i=1}^{k} x_i^2 p_i - m_x^2 = m_{x^2} - m_x^2$$

- Propiedades:
  - si *a* y *b* son constantes se cumple que:

$$\sigma_{ax+b} = |a|\sigma_x$$

$$\sigma^2_{ax+b} = a^2\sigma^2_x$$

- si x e y son dos variables aleatorias independientes se cumple que:

$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y \qquad \text{y} \qquad \sigma_{x+y} = \sqrt{\sigma^2_x + \sigma^2_y}$$

<u>Ejercicio 15.4</u> (de Peña y Romo)

Se lanza tres veces una moneda. Sea x la variable aleatoria que expresa el nº de caras en los tres lanzamientos.

a) Hallar y representar la función de probabilidad de x. (ver Ejemplo pag. 3)

Se lanza 3 veces una moneda:

E={CCC,CCX,CXC,XCC,XXC,XCX,CXX,XXX}

x=0 →{XXX} $\qquad p_0 = P\{x=0\} = 1/8 = 0{,}125$

x=1 →{XXC,XCX,CXX} $\qquad p_1 = P\{x=1\} = 3/8 = 0{,}375$

x=2 →{CCX,CXC,XCC} $\qquad p_2 = P\{x=2\} = 3/8 = 0{,}375$

x=3 →{CCC} $\qquad p_3 = P\{x=3\} = 1/8 = 0{,}125$

b) Calcular el nº esperado de caras al lanzar la moneda. ¿Era previsible el resultado?

$$m_x = \sum_{i=1}^{k} x_i p_i = 0\times 0{,}125 + 1\times 0{,}375 + 2\times 0{,}375 + 3\times 0{,}125 = 1{,}5$$

Sí, ya que en cada lanzamiento P(C)=1/2 y al lanzar tres veces se tiene que $3\times 1/2 = 1{,}5$.

c) Hallar la desviación típica de x

$$\sigma_x = \sqrt{\sum_{i=1}^{k} \left(x_i - m_x\right)^2 p_i} = \sqrt{\begin{array}{l}(0-1{,}5)^2 \times 0{,}125 + (1-1{,}5)^2 \times 0{,}375 \\ + (2-1{,}5)^2 \times 0{,}375 + (3-1{,}5)^2 \times 0{,}125\end{array}} = 0{,}866$$

o bien:

$$\sigma_x = \sqrt{\sum_{i=1}^{k} x_i^2 p_i - m_x^2} = \sqrt{\left(0^2 \times 0{,}125 + \ldots + 3^2 \times 0{,}125\right) - 1{,}5^2} = 0{,}866$$

- La desviación típica es una medida de dispersión que depende de las unidades de medida de la variable. Para evitar este inconveniente podemos emplear el <u>coeficiente de variación</u>. El coeficiente de variación de una variable aleatoria *x* será:

$$CV_x = \frac{\sigma_x}{\left|m_x\right|}$$

Ejercicio 15.7 (de Peña y Romo)

Sea x una variable aleatoria que expresa el nº de personas que habitan en una vivienda elegida al azar. La distribución de probabilidad de x es la siguiente:

x_i	1	2	3	4	5	6	7	8 ó +
p_i	0,230	0,322	0,177	0,155	0,067	0,024	0,015	0,010

a) Comprobar que es una distribución de probabilidad.

Todas las p_i son mayores o iguales que cero y además se cumple que:

$$\sum_{i=1}^{8} p_i = 0{,}23 + 0{,}322 + 0{,}177 + \ldots + 0{,}010 = 1$$

b) Hallar la probabilidad de que el nº de personas que viven en un hogar sea menor o igual que cuatro.

$$P(x \leq 4) = P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4) =$$
$$= 0{,}23 + 0{,}322 + 0{,}177 + 0{,}155 = 0{,}884$$

c) Calcular la probabilidad de que al menos dos personas vivan en una vivienda.

$$P(x \geq 2) = P(x = 2) + P(x = 3) + \ldots + P(x \geq 8) =$$
$$= 1 - P(x < 2) = 1 - 0{,}23 = 0{,}77$$

d) Obtener el nº medio de personas que habitan en una vivienda.

$$m_x = 1 \times 0{,}23 + 2 \times 0{,}322 + 3 \times 0{,}177 + \ldots + 7 \times 0{,}015 + 8 \times 0{,}01 = 2{,}689$$

4.3. MEDIA, VARIANZA Y DESVIACIÓN ESTANDAR

La desviación estándar (**σ**) mide cuánto se separan los datos.

La fórmula es fácil: es la raíz cuadrada de la **varianza**. Así que, "¿qué es la varianza?"

Varianza

La varianza (que es el cuadrado de la desviación estándar: $\boldsymbol{\sigma^2}$) se define así:

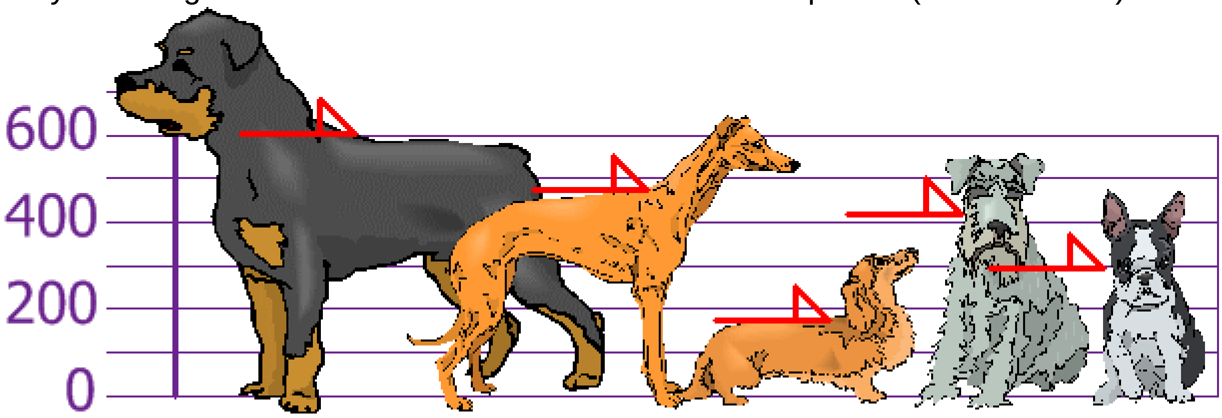
Es la media de las diferencias con la media **elevadas al cuadrado**.

En otras palabras, sigue estos pasos:

1. Calcula la media (el promedio de los números)
2. Ahora, por cada número resta la media y eleva el resultado al cuadrado (la diferencia elevada al cuadrado).
3. Ahora calcula la media de esas diferencias al cuadrado. (¿Por qué al cuadrado?)

Ejemplo

Tú y tus amigos habéis medido las alturas de vuestros perros (en milímetros):



Las alturas (de los hombros) son: 600mm, 470mm, 170mm, 430mm y 300mm.

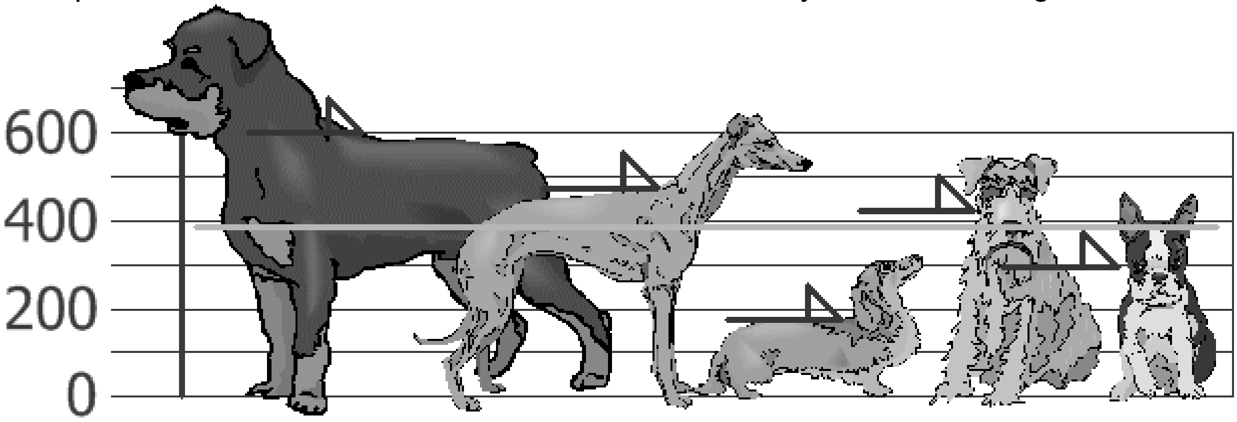
Calcula la media, la varianza y la desviación estándar.

Respuesta:

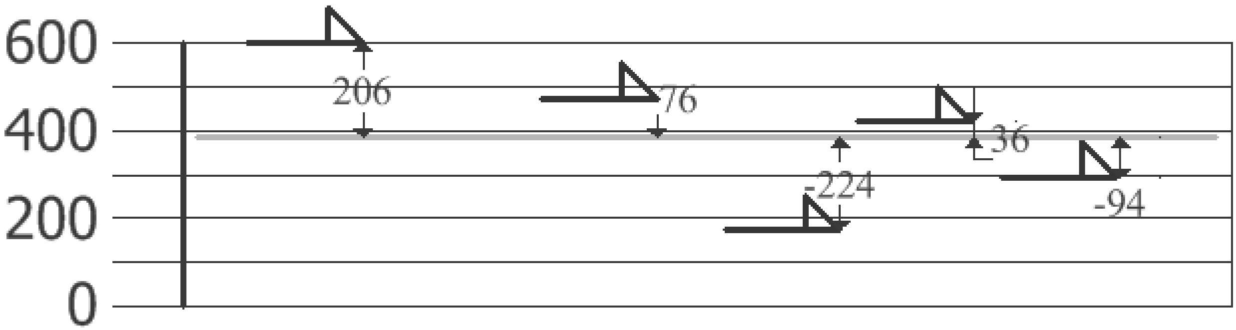
Media = 600 + 470 + 170 + 430 + 300 = 1970 = 394

5 5

así que la altura media es 394 mm. Vamos a dibujar esto en el gráfico:



Ahora calculamos la diferencia de cada altura con la media:



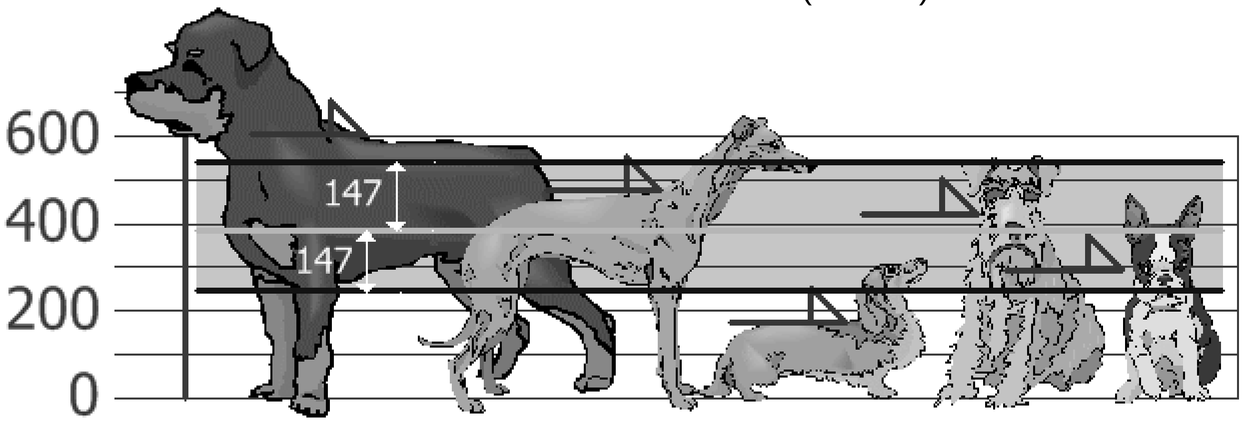
Para calcular la varianza, toma cada diferencia, elévala al cuadrado, y haz la media:

$$\text{Varianza: } \sigma^2 = \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5} = \frac{108{,}520}{5} = 21{,}704$$

Así que la varianza es 21,704.

Y la desviación estándar es la raíz de la varianza, así que:

Desviación estándar: **σ = √21,704 = 147** y lo bueno de la desviación estándar es que es útil: ahora veremos qué alturas están a distancia menos de la desviación estándar (147mm) de la media:



Así que usando la desviación estándar tenemos una manera "estándar" de saber qué es normal, o extra grande o extra pequeño.

Los Rottweilers **son** perros grandes. Y los Dachsunds **son** un poco menudos... ¡pero que no se enteren!

***Nota: ¿por qué *al cuadrado*?**

Elevar cada diferencia al cuadrado hace que todos los números sean positivos (para evitar que los números negativos reduzcan la varianza)

Y también hacen que las diferencias grandes se destaquen. Por ejemplo 100^2=10,000 es mucho más grande que 50^2=2,500.

Pero elevarlas al cuadrado hace que la respuesta sea muy grande, así que lo deshacemos (con la raíz cuadrada) y así la desviación estándar es mucho más útil

4.4. DISTRIBUCIÓN DE PROBABILIDAD

4.4.1. BINOMINAL

DISTRIBUCIÓN DE LA PROBABILIDAD BINOMIAL

Esta distribución es la que mejor se ajusta a la distribución de probabilidades de variable discreta.

Si lanzamos dos monedas al aire, se tiene el siguiente espacio maestral:

$$\Omega = \{cc, cs, sc, ss\}$$

$$n(\Omega) = 4$$

Si p es la probabilidad de obtener una cara(c) al considerar una sola moneda y q la probabilidad de que salga sello(s); entonces p=q= ½; luego:

$$pp = p^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$$2pq = 2(\frac{1}{2})(\frac{1}{2}) = 2\left(\frac{1}{2}\right)^2 = \frac{2}{4}$$

2

$$qq = q^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

Con el binomio de Newton deducimos lo siguiente:

$$(p+q)^n = \sum_{x=0}^{n} C_x^n p^x q^{n-x} \qquad \text{……………………………………………………………(5)}$$

Luego, la distribución de probabilidad binomial esta dada por:

$$P(X=x)=C_x^n p^x q^{n-x} = C_x^n p^x (1-p)^{n-x}$$ ……………………………………. (6)

Donde:

p: Probabilidad de éxito de cada ensayo.

n: Número de ensayos.

x: Número de exitos.

OBSERVACIÓN

(1) $(\frac{1}{2}+\frac{1}{2})^3 = \frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}$

(2)Si p=q=1/2, el histograma de las distribuciones binomiales son simétricas.

Si el experimento se repite r veces con n ensayos ; entonces se tiene:

$$r(p+q)^n = r\sum_{x=0}^{n} C_x^n p^x q^{n-x} = \sum_{x=0}^{n} rC_x^n p^x q^{n-x}$$ ……………………………. (7)

Luego se deduce que:

$$P(X=x) = rC_x^n p^x (1-p)^{n-x}$$ ………………………………. (8)

MEDIA DE LA DISTRIBUCIÓN BINOMIAL

Esta dada por:

$\mu = np$ ………………………………………. (9)

VARIANZA DE LA DISTRIBUCIÓN BINOMIAL

$$\sigma^2 = np(1-p)$$ ………………………………………………. (10)

4.4.2. HIPERGEOMÉTRICA

Esta distribución se aplica cuando el muestreo se realiza sin repetición y la probabilidad de éxito no permanece constante de un ensayo a otros calcula mediante la fórmula:

$$P(x) = \frac{C_x^s C_{n-x}^{N-s}}{C_n^N}$$ ………………………… (12)

Donde:

N: Tamaño de la población

S: Cantidad de éxitos en la población

X: Número de éxitos en la muestra.

n : Tamaño de la muestra.

n>=0.05N

4.4.3. POISOON

Describe la cantidad de veces que ocurre un evento en un intervalo determinado (tiempo, volumen, temperatura, etc...).La distribución se basa en dos supuestos:

1°) La probabilidad es proporcional a la extensión del intervalo.

2°) Los intervalos son independientes.

Esta distribución es una forma límite de la distribución binomial, cuando la probabilidad de éxito es bien pequeña y n es grande ,a esta distribución se llama "Ley de eventos improbables", lo cual significa que la probabilidad de p es bien pequeña .La probabilidad de Poisson es una probabilidad discreta; puesto que se forma por conteo

$$P(x) = \lim_{n\to\infty} C_x^n p^x (1-p)^{n-x}$$ ……………………… (13)

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$ ……………………………(14)

Donde:

μ: Media del número de ocurrencias.

e: Constante de Euler.

x : Número de ocurrencias

6.1Media:-Esta dado por:

$\mu = np$

$\sigma^2 = np$.

4.4.4. CONTINUA

Características:

1. Es generada por una variable continua (x).

   x® Es una variable que puede tomar tanto valores enteros como fraccionarios.

   x® 1.0, 3.7, 4.0, 4.6, 7.9, 8.0, 8.3, 11.5, ...,¥

2. f(x)³ 0 Las probabilidades asociadas a cada uno de los valores que toma x deben ser mayores o iguales a cero. Dicho de otra forma, la función de densidad de probabilidad deberá tomar solo valores mayores o iguales a cero. La función de densidad de probabilidad sólo puede estar definida en los cuadrantes I y II.

3. $\int_{-\infty}^{\infty} f(x)dx = 1$ La sumatoria de las probabilidades asociadas a cada uno de los valores que toma x debe ser igual a 1. El área definida bajo la función de densidad de probabilidad deberá ser de 1.

Hasta el momento se han considerado las distribuciones de probabilidad para variables discretas, donde se podía asignar el valor que toma la función de probabilidad cuando la variable aleatoria tomaba un valor en concreto. Sin embargo, al considerar las variables continuas se encuentra uno el problema de que, lo más probable, los datos que se puedan recabar no sean completamente exactos, o dos o más de ellos no coincidan, por lo que se tienen que trabajar en intervalos y, en ese momento, modelar una función se convierte en un problema serio.

Sin embargo, se pueden realizar aproximaciones y describir la probabilidad a través de modelos teóricos de probabilidad cuya gráfica es una línea continua, a diferencia de las variables discretas que le corresponde un histograma.

Para clarificar cómo se realiza esta aproximación al modelo teórico consideremos el siguiente caso:

Se han registrado los tiempos que le tomó a una empresa de mensajería entregar 190 paquetes con destinatarios diferentes dentro de una misma ciudad. Los datos se han agrupado en una distribución de frecuencias considerando intervalos de cinco días como sigue:

Tiempo de entrega (días)	**No. de paquetes**

[0,5)	115
[5,10)	31
[10,15)	17
[15,20)	12
[20,25)	10
[25,30)	5

Supongamos que un posible cliente, conociendo esta información, quisiera saber qué probabilidad tiene de que su paquete sea entregado en dos días. El problema es que al manejar intervalos de cinco días estamos suponiendo que dentro de cada intervalo los datos se distribuyen uniformemente, cosa que no es real.

Podríamos aumentar la muestra y seguir recogiendo información para hacer una distribución de frecuencias similar a la anterior, pero se tendría el mismo problema: dentro de cada intervalo se está presuponiendo que los datos se distribuyen uniformemente.

Otra posible solución es reducir la amplitud de los intervalos, de tal suerte que podríamos tomar una amplitud de tres días por intervalo y hacer la siguiente distribución de frecuencias:

Tiempo de entrega (días)	**No. de paquetes (frec.)**
[0,3)	93
[3,6)	30
[6,9)	18
[9,12)	13
[12,15)	9
[15,18)	8
[18,21)	6

[21,24)	6
[24,27)	4
[27,30)	3

Al seguir reduciendo la amplitud a dos días se obtiene la distribución:

Tiempo de entrega (días)	**No. de paquetes (frec.)**
[0,2)	76
[2,4)	29
[4,6)	18
[6,8)	13
[8,10)	10
[10,12)	8
[12,14)	6
[14,16)	6
[16,18)	5
[18,20)	4
[20,22)	4
[22,24)	4
[24,26)	3
[26,28)	2
[28,30)	2

Y al reducirla a intervalos de un día se tiene la distribución:

Tiempo de entrega (días)	**No. de paquetes (frec.)**
[0,1)	51
[1,2)	25
[2,3)	17
[3,4)	12
[4,5)	10
[5,6)	8
[6,7)	7
[7,8)	6
[8,9)	5
[9,10)	5
[10,11)	4
[11,12)	4
[12,13)	3
[13,14)	3
[14,15)	3
[15,16)	3
[16,17)	3
[17,18)	2
[18,19)	2
[19,20)	2

[20,21)	2
[21,22)	2
[22,23)	2
[23,24)	2
[24,25)	2
[25,26)	1
[26,27)	1
[27,28)	1
[28,29)	1
[29,30)	1

Ahora, veamos. Lo que le interesa al futuro cliente es la probabilidad de que se haga una entrega en un cierto tiempo, por lo que habría que considerar las frecuencias relativas y, como antes, reducir la amplitud de los intervalos. Con esto se obtendrían las siguientes distribuciones de frecuencias:

Intervalos de cinco días

Intervalo	frec.	frec. rel.
[0,5)	115	0.605
[5,10)	31	0.163
[10,15)	17	0.089
[15,20)	12	0.063
[20,25)	10	0.053
[25,30)	5	0.026

Intervalos de tres días

Intervalo	frec.	frec. rel.
[0,3)	93	0.489
[3,6)	30	0.158
[6,9)	18	0.095
[9,12)	13	0.068
[12,15)	9	0.047
[15,18)	8	0.042
[18,21)	6	0.032
[21,24)	6	0.032
[24,27)	4	0.021
[27,30)	3	0.016

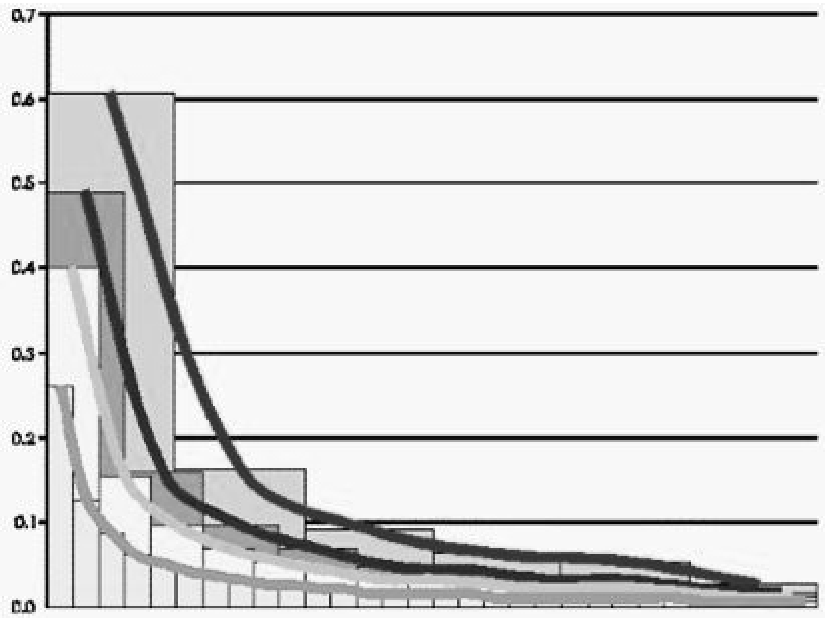
Intervalos de dos días

Intervalo	frec.	frec. rel.
[0,2)	76	0.400
[2,4)	29	0.153
[4,6)	18	0.095
[6,8)	13	0.068
[8,10)	10	0.053
[10,12)	8	0.042
[12,14)	6	0.032
[14,16)	6	0.032
[16,18)	5	0.026
[18,20)	4	0.021
[20,22)	4	0.021
[22,24)	4	0.021
[24,26)	3	0.016
[26,28)	2	0.011
[28,30)	2	0.011

Intervalos de un día

Intervalo	frec.	frec. rel.
[0,1)	51	0.268
[1,2)	25	0.132
[2,3)	17	0.089
[3,4)	12	0.063
[4,5)	10	0.053
[5,6)	8	0.042
[6,7)	7	0.037
[7,8)	6	0.032
[8,9)	5	0.026
[9,10)	5	0.026
[10,11)	4	0.021
[11,12)	4	0.021
[12,13)	3	0.016
[13,14)	3	0.016
[14,15)	3	0.016
[15,16)	3	0.016
[16,17)	3	0.016
[17,18)	2	0.011
[18,19)	2	0.011
[19,20)	2	0.011
[20,21)	2	0.011
[21,22)	2	0.011
[22,23)	2	0.011
[23,24)	2	0.011
[24,25)	2	0.011
[25,26)	1	0.005
[26,27)	1	0.005
[27,28)	1	0.005
[28,29)	1	0.005
[29,30)	1	0.005

Y podríamos graficar tal información en histogramas para poder ver cómo se aproximan, si es que ocurre, los valores a una curva continua:



donde las barras rosas (y la línea roja) corresponden a los intervalos de cinco días; las barras y línea azules, a los intervalos de tres días; las barras y línea amarillas, a los intervalos de dos días; y las barras y líneas verdes, a los intervalos de un día.

Se han incluido de una vez las líneas que unen los puntos medios de las barras del histograma porque se puede ver que las barras de las frecuencias relativas se "**achaparran**" y las líneas graficadas están tan separadas del lado izquierdo (en este caso) que no se puede hablar de una aproximación continua a una sóla línea.

Una posible solución es utilizando la **densidad del intervalo**, que se va a definir como el cociente de la frecuencia relativa entre la amplitud del intervalo:

$$\text{densidad del intervalo} = \frac{\text{frecuencia relativa}}{\text{amplitud del intervalo}}$$

(De hecho, existe la función de densidad de una distribución de probabilidad, de donde se deriva esta definición de densidad del intervalo.)

De esta manera, a las distribuciones de frecuencias anteriores se les puede añadir la columna correspondiente a la densidad:

Intervalos de cinco días

Intervalo	**frec.**	**frec. rel.**	**densidad**
[0,5)	115	0.605	0.121
[5,10)	31	0.163	0.033
[10,15)	17	0.089	0.018
[15,20)	12	0.063	0.013
[20,25)	10	0.053	0.011
[25,30)	5	0.026	0.005

Intervalos de tres días

Intervalo	**frec.**	**frec. rel.**	**densidad**
[0,3)	93	0.489	0.163
[3,6)	30	0.158	0.053
[6,9)	18	0.095	0.032

[9,12)	13	0.068	0.023
[12,15)	9	0.047	0.016
[15,18)	8	0.042	0.014
[18,21)	6	0.032	0.011
[21,24)	6	0.032	0.011
[24,27)	4	0.021	0.007
[27,30)	3	0.016	0.005

Intervalos de dos días

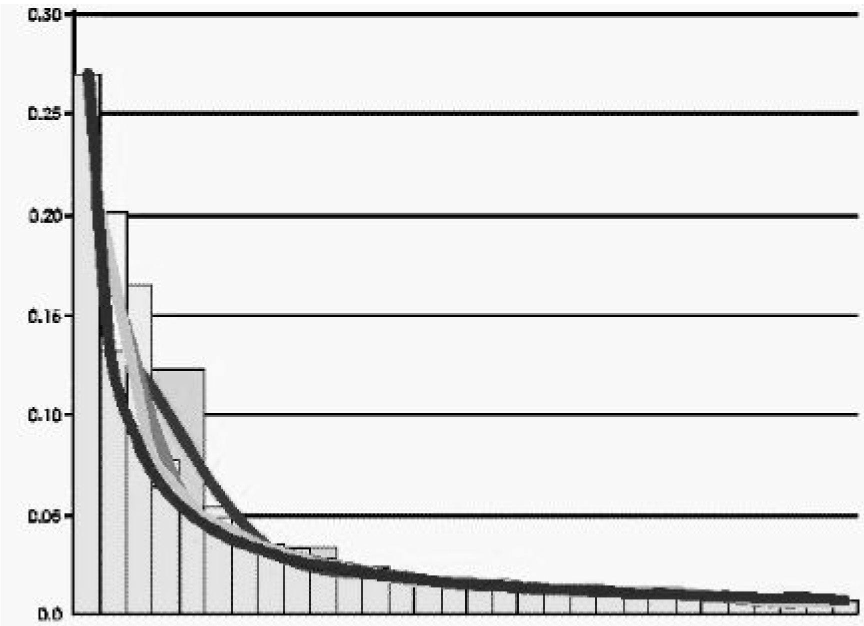
Intervalo	frec.	frec. rel.	densidad
[0,2)	76	0.400	0.200
[2,4)	29	0.153	0.076
[4,6)	18	0.095	0.047
[6,8)	13	0.068	0.034
[8,10)	10	0.053	0.026
[10,12)	8	0.042	0.021
[12,14)	6	0.032	0.016
[14,16)	6	0.032	0.016
[16,18)	5	0.026	0.013
[18,20)	4	0.021	0.011
[20,22)	4	0.021	0.011
[22,24)	4	0.021	0.011
[24,26)	3	0.016	0.008
[26,28)	2	0.011	0.005
[28,30)	2	0.011	0.005

Intervalos de un día

Intervalo	frec.	frec. rel.	densidad
[0,1)	51	0.268	0.268
[1,2)	25	0.132	0.132
[2,3)	17	0.089	0.089
[3,4)	12	0.063	0.063
[4,5)	10	0.053	0.053
[5,6)	8	0.042	0.042
[6,7)	7	0.037	0.037
[7,8)	6	0.032	0.032
[8,9)	5	0.026	0.026
[9,10)	5	0.026	0.026
[10,11)	4	0.021	0.021
[11,12)	4	0.021	0.021
[12,13)	3	0.016	0.016
[13,14)	3	0.016	0.016
[14,15)	3	0.016	0.016
[15,16)	3	0.016	0.016
[16,17)	3	0.016	0.016
[17,18)	2	0.011	0.011
[18,19)	2	0.011	0.011
[19,20)	2	0.011	0.011

Intervalo	frec.	frec. rel.	densidad
[20,21)	2	0.011	0.011
[21,22)	2	0.011	0.011
[22,23)	2	0.011	0.011
[23,24)	2	0.011	0.011
[24,25)	2	0.011	0.011
[25,26)	1	0.005	0.005
[26,27)	1	0.005	0.005
[27,28)	1	0.005	0.005
[28,29)	1	0.005	0.005
[29,30)	1	0.005	0.005

y realizar los histogramas correspondientes, que quedan como sigue:



donde las barras rosas, y la línea roja, corresponden a los intervalos de cinco días; las barras y línea verdes, a los intervalos de tres días; las barra y línea amarillas, a los intervalos de dos días; y las barras y línea azules, a los intervalos de un día.

Igual que en el caso anterior, se han graficado simultáneamente las barras y las líneas que unen los puntos medios de éstas para observar que con la densidad sí se aproximan los histogramas a una línea continua (que la mejor aproximación presentada es la línea azul) cuando los intervalos se reducen continuamente.

El resultado es una línea continua que es la gráfica de una cierta función denominada función de densidad de la distribución probabilística.

Ahora, considerando la manera en que se definió la densidad de un intervalo como:

$$\text{densidad del intervalo} = \frac{\text{frecuencia relativa}}{\text{amplitud del intervalo}}$$

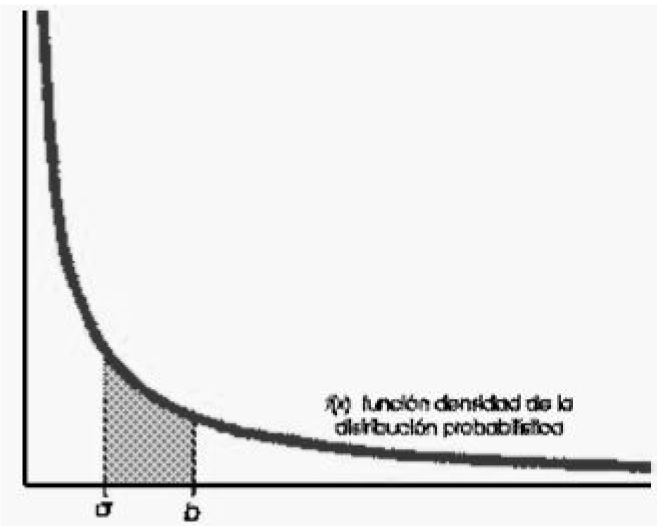
y recordando que la frecuencia relativa es la probabilidad de un evento (en el ejemplo de la mensajería sería la probabilidad de entregar un paquete dentro de un intervalo dado de tiempo):

$$\text{frecuencia relativa} = \frac{\text{frecuencia de un evento}}{\text{numero total de eventos}} = \text{probabilidad del evento}$$

Entonces, despejando en el primer cociente la frecuencia relativa e igualando con esta segunda expresión obtenemos que

probabilidad del evento = (*densidad del intervalo*)· (*amplitud del intervalo*)

Es decir, que la probabilidad de que ocurra un evento corresponde al área de las barras del histograma hecho tomando en cuenta la densidad de los intervalos; y que cuando tales intervalos tienen una amplitud que tiende a cero, y la gráfica se convierte en la curva continua de la función de densidad, entonces la probabillidad de que un evento ocurra en un intervalo (*a*,*b*) es el área bajo la curva de la función en ese intervalo:



y, por tanto, el cálculo de tal probabilidad se realiza utilizando cálculo integral:

$$P(a < x < b) = P(x \in (a,b)) = \int_a^b f(x)\,dx$$

donde *f*(*x*) es la función de densidad de la distribución probabilística correspondiente.

Hay que estar conscientes de que en el caso de las variables continuas sólo se puede calcular la probabilidad de que un evento caiga dentro de un intervalo, debido a que la exactitud de los instrumentos de medición siempre es relativa y muy lejana a la "exactitud" de los cálculos matemáticos.

Por esto, la probabilidad de que la variable aleatoria tome un valor exacto es nula:

$$P(x = a) = \int_a^a f(x)\,dx = 0$$

Esto se puede explicar de la siguiente manera: si, como ya dijimos, la probabilidad (frecuencia relativa) es igual a la densidad del intervalo por la amplitud del intervalo, entonces no importa qué tan grande sea la densidad de tal intervalo porque, como ya también se dijo, por ser variable continua la amplitud del intervalo tiende a cero y, por tanto, la probabilidad es igual a cero.

Modelos de distribución de probabilidad de variables continuas

Al igual que en el caso de las distribuciones de probabilidad de variables discreta, en el caso de las distribuciones de probabilidad de variables continuas se tienen varios modelos teóricos que en seguida presentamos.

A la derecha de cada modelo aparece la función de densidad correspondiente a cada modelo.

- Uniforme. Es la distribución en donde todos los eventos tienen la misma probabilidad.

$$f(x) = \frac{1}{\theta_2 - \theta_1};\quad \theta_1 \leq x \leq \theta_2$$

- Exponencial. Se utiliza para estudiar el tiempo entre dos sucesos. La función de *Excel* que le corresponde es DISTR.EXP.

$$f(x) = \frac{1}{\beta} e^{\frac{-x}{\beta}};\quad 0 \leq x < \infty$$

- Beta. Sirve para el estudio de variaciones, a través de varias muestras, de un porcentaje que representa algún fenómeno. La función DISTR.BETA del *Excel* sirve para obtener sus valores; y la función DISTR.BETA.INV proporciona los valores inversos de la función, es decir, se utiliza como parámetro la imagen de la función y regresa la variabla independiente.

$$f(x) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] x^{\alpha-1}(1-x)^{\beta-1};$$
$$0 < x < 1$$

- Gamma. Se utiliza para estudiar variables cuya distribución puede ser asimétrica. La función de *Excel* que le corresponde es DISTR.GAMMA; y la función DISTR.GAMMA.INV es la inversa de la anterior.

$$f(x) = \left[\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right] x^{\alpha-1} e^{\frac{-x}{\beta}};$$
$$0 < x < \infty$$

- ji cuadrada (c^2). Es una distribución asociada a la prueba c^2, y se usa para comparar los valores observados con los esperados. La función DISTR.CHI de *Excel* sirve para este

$$f(\chi^2) = \frac{(\chi^2)^{\left(\frac{\nu}{2}\right)-1} e^{\frac{-\chi^2}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)};$$
$$\chi^2 > 0$$

- Normal. Es la distribución más utilizada porque la mayoría de las variables utilizadas en fenómenos sociales se distribuyen aproximadamente siguiendo este modelo. Es la que tocaremos a continuación y se le llama comúnmente distribución normal.

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

Cálculo de media y desviación estándar para una distribución continua

1. Media o valor esperado de x.- Para calcular la media de una distribución de probabilidad continua se utiliza la siguiente fórmula:

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx$$

Donde:

m = E(x) = media o valor esperado de la distribución

x = variable aleatoria continua

f(x) = función de densidad de la distribución de probabilidad

2. Desviación estándar.- La fórmula para determinar la desviación estándar de una distribución continua es;

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 * f(x)dx$$

luego:

$$\sigma = \sqrt{\sigma^2}$$

Ejemplo:

1. Para la siguiente función,

$f(x) = \frac{1}{9}x^2$ cuando 0£ x £ 3, f(x) = 0 para cualquier otro valor

a. Diga si esta función nos define una distribución de probabilidad.
b. Si la función define una distribución de probabilidad, entonces, determine su media y desviación estándar.
c. Determine la probabilidad de que 1£ x < 2.

Solución:

a. Para verificar que la función nos define una distribución de probabilidad, es necesario que cumpla con las características que se habían mencionado.
   1. x ® sí es una variable continua porque puede tomar cualquier valor entre 0 y 3
   2. f(x)³ 0, lo que se comprueba si damos diferentes valores a x para ver que valores toma f(x), dándonos cuenta de que efectivamente f(x) solo toma valores mayores o iguales a cero.
   3.

x	f(x)
0	0.0
0.5	0.02778
1.0	0.11111

1.4	0.21778
2.1	0.49
2.7	0.81
3.0	1.0

5. Para comprobar que la sumatoria de las probabilidades que toma cada valor de x es de 1, se integra la función de 0 a 3 como se muestra a continuación:

$$A = \int_{-\infty}^{\infty} f(x)dx = \int_{0}^{3} \frac{1}{9}x^{2}\,dx = \frac{1}{9}\left(\frac{x^{2+1}}{2+1}\right) \Big| = \frac{1}{27}(3^{3} - 0^{3}) = \frac{1}{27}(27 - 0) = 1$$

A= área bajo la función

Con las operaciones anteriores comprobamos que la función $\frac{1}{9}x^{2}$ sí nos define una distribución de probabilidad continua.

b. Cálculo de media y desviación estándar.

$$\mu = \int_{-\infty}^{\infty} x * f(x)dx = \int_{0}^{3} x\left(\tfrac{1}{9}x^{2}\right)dx = \int_{0}^{3} \frac{1}{9}x^{3}\,dx = \frac{1}{9}\left(\frac{x^{4}}{4}\right) \Big| =$$

$$= \frac{1}{36}(3^{4} - 0^{4}) = \frac{1}{36}(81 - 0) = \frac{81}{36} = 2.25$$

$$\sigma^{2} = \int_{-\infty}^{\infty} (x - \mu)^{2} * f(x)dx = \int_{0}^{3} (x - 2.25)^{2} * \left(\frac{1}{9}x^{2}\right)dx =$$

$$= \int_{0}^{3} (x^{2} - 4.5x + 5.0625)\left(\frac{1}{9}x^{2}\right)dx = \int_{0}^{3} \left(\frac{x^{4}}{9} - \frac{x^{3}}{2} + \frac{5.0625x^{2}}{9}\right)dx =$$

$$= \frac{x^{5}}{45} \Big| - \frac{x^{4}}{8} \Big| + \frac{5.0625x^{3}}{27} \Big| = \frac{(3)^{5}}{45} - \frac{(3)^{4}}{8} + \frac{5.0625(3)^{3}}{27} =$$

$$= \frac{243}{45} - \frac{81}{8} + \frac{136.6875}{27} = 5.4 - 10.125 + 5.0625 = 0.3375$$

$$\sigma = \sqrt{\sigma^{2}} = \sqrt{0.3375} = 0.5809$$

Las barras nos indican la evaluación de la integral entre 0 y 3.

c)

$$p(1 \le x < 2) = \int_1^2 f(x)dx = \int_1^2 \frac{1}{9}x^2 dx = \frac{1}{9}(\frac{x^3}{3}) \Big| = \frac{1}{9}(\frac{2^3}{3} - \frac{0^3}{3}) = \frac{1}{9}(\frac{8}{3}) = \frac{8}{27} = 0.2963$$

La barra nos indica la evaluación de la integral de 1 a 2.

Con las operaciones anteriores nos damos cuenta que para evaluar probabilidades para variables de tipo continuo, es necesario evaluar la función de densidad de probabilidad en el rango de valores que se desea; que vendría siendo el área que se encuentra entre f(x) y el eje de las x y entre el rango de valores definidos por la variable x.

1. Suponga que el error en la temperatura de reacción, en ^{o}C, para un experimento controlado de laboratorio es una variable aleatoria continua x, que tiene la función de densidad de probabilidad:

   $f(x) = \frac{x^2}{3}$, para -1< x < 2 y f(x)= 0 en cualquier otro caso

   a. Verifique la tercera condición de la definición de una distribución de probabilidad continua.
   b. Determine la media o valor esperado de la distribución de probabilidad.
   c. Encuentre la probabilidad de que 0< x £ 1.

Solución:

a. Como la tercera condición es que la sumatoria de las probabilidades asociadas a cada uno de los valores que toma x debe de ser 1, esto se comprueba de la siguiente manera:

$$A = \int_{-1}^2 \frac{x^2}{3}dx = \frac{1}{3}(\frac{x^3}{3}) \Big| = \frac{1}{3}(\frac{2^3}{3} - \frac{-1^3}{3}) = \frac{8}{9} + \frac{1}{9} = \frac{9}{9} = 1$$

b. $$\mu = E(x) = \int_{-\infty}^{\infty} x * f(x)dx = \int_{-1}^2 x(\frac{x^2}{3})dx = \int_{-1}^2 \frac{x^3}{3}dx = \frac{1}{3}(\frac{x^4}{4}) \Big| =$$

$$= \frac{1}{3}(\frac{2^4}{4} - \frac{-1^4}{4}) = \frac{16}{12} - \frac{1}{12} = \frac{15}{12} = 1.25$$

c. $$p(0 < x \le 1) = \int_0^1 \frac{x^2}{3}dx = \frac{1}{3}(\frac{x^3}{3}) \Big| = \frac{1}{3}(\frac{1^3}{3} - \frac{0^3}{3}) = \frac{1}{3} * \frac{1}{3} = \frac{1}{9} = 0.11111$$