PROBABILIDAD Y DISTRIBUCIONES

Bioestadística

CLAVE:	PERIODO LECTIVO: primer semestre	HT: 1	HP: 3	TH: 4	CRÉDITOS: 5

LICENCIATURA EN MEDICINA VETERINARIA Y ZOOTECNIA

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

Septiembre de 2018

Dr. en BCA. Jorge Osorio Avalos

Contenido: Unidad 2
Probabilidad y Distribuciones

2.1 Evento
2.2 Definición de probabilidad

Probabilidad y su relación con las Ciencias Biológicas

- Algunos procesos biológicos están afectados por el azar (segregación alélica, determinación sexual).

- Aunque el proceso biológico no está afectado por el azar, el resultado de experimentos si lo están (fluctuaciones del medio ambiente debidas al azar).

- El azar o proceso aleatorio también pueden estar involucrados en un diseño de un experimento. La asignación aleatoria de los sujetos a un tratamiento experimental.

- Las conclusiones del análisis de datos son planteadas en términos de probabilidad.

- Las bases teóricas (modelos) que son utilizados como base para la inferencia estadística, depende de la probabilidad.

Probabilidad: Es un valor numérico que expresa la verosimilidad de un evento {E}.

$$P\{E\} \longrightarrow 0 \leq \{E\} \leq 1$$

Ejemplo: Un evento ocurre o no ocurre;

Lanzar una moneda → Águila (ocurre o no ocurre)

Lanzar una moneda → Sol (ocurre o no ocurre)

Al lanzar una moneda existen dos opciones con respecto al evento sol (ocurre o no ocurre), al igual del evento águila.

$P\{E\} = ½ = 0.5$ → que caiga sol

$P\{sol\} = ½= 0.5$ → que no caiga sol

En una población de Drosophila melanogaster,

- 30% tienen ojos negros (mutación)
- 70% tienen ojos grises (normal)

Si una mosca es extraída al azar, la esperanza de que sea negra:

P {E}: mosca negra P {mosca negra} = 0.3

2.3 Propiedades de la probabilidad

- Para comprender las propiedades de la probabilidad debemos establecer algunos conceptos previamente:

Espacio muestral {E}: es el conjunto de los diferentes resultados que pueden darse en un experimento aleatorio.

Suceso: subconjunto del espacio muestral. Se representa con una letra mayúscula con sus elementos entre llaves y separados por comas.

Unión: la unión de dos sucesos es el suceso que ocurre cuando se da uno de ellos.

Intersección: intersección de dos sucesos, es el suceso que ocurre cuando se dan ambos a la vez.

Interpretación frecuencial de la probabilidad

La probabilidad de un evento {E} tiene significado en la medida que el proceso aleatorio puede repetirse en forma infinita.

La P {E} es interpretada como frecuencia relativa de que ocurra E en una serie de repeticiones del proceso aleatorio.

$$P\{E\} = \frac{\text{\# de veces que ocurre E}}{\text{\# de veces que el proceso aleatorio se repite}}$$

E: águila   proceso aleatorio: **lanzar una moneda.**

$$P\{E\} = 0.5 \qquad \frac{\text{\# águila}}{\text{\# de lanzamientos (volados)}}$$

E: águila dos veces   proceso aleatorio: **lanzar dos veces una moneda.**

$$\frac{\text{\# de veces donde el águila cae 2 veces}}{\text{\# de veces en que la moneda se lanza dos veces}}$$

P {E} = 0.25

E: seleccionar 2 moscas del mismo color (negro)   proceso aleatorio: **Tomar una muestra aleatoria de tamaño n=2**

$$P\{E\} = 0.09 \qquad \frac{\text{\# de veces en que las moscas tienen un mismo color}}{\text{\# de veces que una muestra de tamaño (n=2) es seleccionada}}$$

Variable aleatoria: Una variable cuyo valor depende de un proceso al azar, por ejemplo:

Lanzar un dado, $y = 1, 2, 3, 4, 5, 6$.

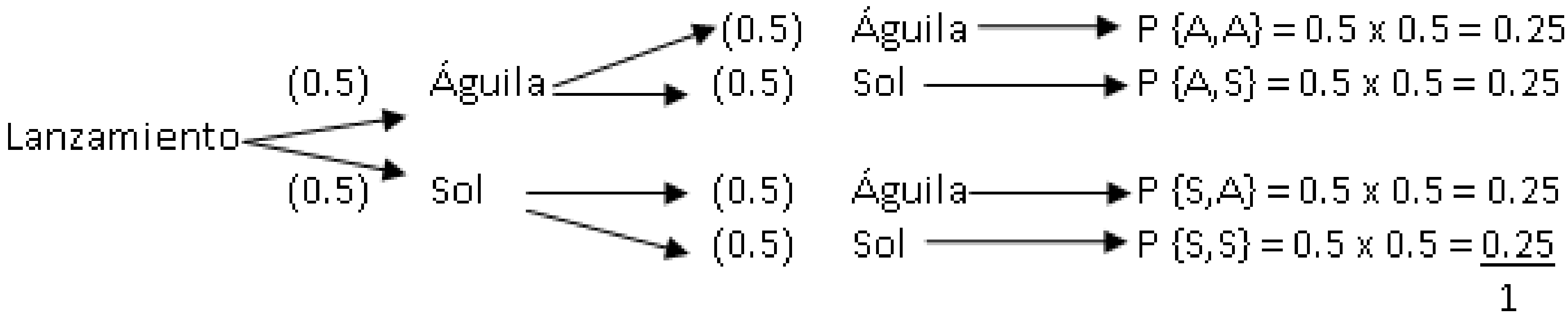
$P\{6\} = 1/6 = 0.167$, $P\{2 \leq y \leq 4\} = 3/6 = 0.5$

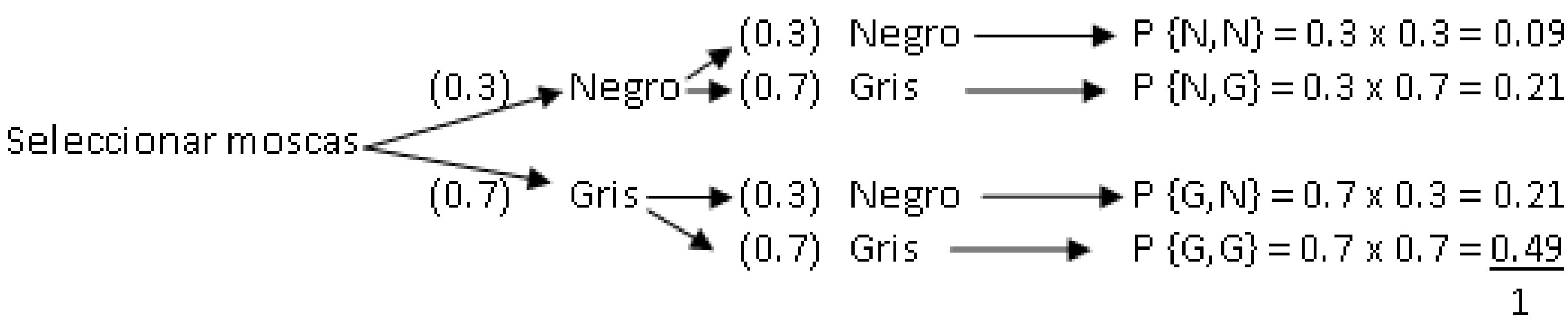
- Otro ejemplo: probabilidad de seleccionar una familia de una población, si el 25% de las familias tiene 2 hijos,

$P\{2\} = 0.25$, y = # de niños en la familia $y = 0, 1, 2, 3, 4$ …….

Árboles de probabilidad

Ejemplo: lanzar una moneda dos veces:





Diagnóstico de una enfermedad:

- 8% de la población tiene la enfermedad
- 95% de confiabilidad de la prueba para detectar la enfermedad si la persona está enferma
- 90% de la prueba para detectar que la persona no tiene la enfermedad.

Individuo → (0.08) Enfermo → (0.95) Positivo → Positivo verdadero = 0.08 x 0.95 = 0.076

Individuo → (0.08) Enfermo → (0.05) Negativo → Falso negativo = 0.08 x 0.05 = 0.004

Individuo → (0.92) Sano → (0.1) Positivo → Falso positivo = 0.92 x 0.1 = 0.092

Individuo → (0.92) Sano → (0.9) Negativo → Negativo verdadero = 0.92 x 0.9 = 0.828

1

Otro ejemplo:

Para analizar la eficacia de una prueba de embarazo, 150 mujeres fueron sometidas a dicha prueba, resultando que el 10% de las mujeres se encontraban embarazadas y el 90% no lo estaban. La prueba de embarazo tiene una efectividad del 98%, pero se ha detectado que aquellas mujeres que la prueba indicó que la no se encontraban embarazadas resultó en que si lo estaban.

Mujer → (0.10) Embarazada → (0.98) Positiva → Positiva verdadera = 0.10 x 0.98 =0.098

Mujer → (0.10) Embarazada → (0.02) Negativa → Falsa negativa = 0.10 x 0.02 = 0.002

Mujer → (0.90) No embarazada → (0.01) Positiva → Falsa positiva = 0.90 x 0.01 = 0.009

Mujer → (0.90) No embarazada → (0.99) Negativa → Negativa verdadera = 0.90 x 0.99 = <u>0.891</u>

1.000

¿Cuántas mujeres se encuentran realmente embarazadas?

¿Cuántas mujeres si estaban embarazadas cuando la prueba de embarazo indicó que era negativa?

NOTA: realizar otro ejemplo con el 5% de las mujeres se encontraban embarazadas y el 95% no lo estaban

Percentiles

Son observaciones que dividen un conjunto de datos ordenados, de tal forma que "x" % de los datos caen por debajo de una observación o dato específico, ejemplo: percentil 50 → 50% de los datos caen por debajo de "x" observación (mediana).

Percentil L (P) = $\dfrac{n+1.}{100/P}$

+ Cuartiles.

(Q_1) – Inferior: deja 25% por debajo del valor y 75% por arriba del valor
(Q_2) – Medio: deja 50% por debajo del valor y 50% por arriba del valor
(Q_3) – Medio: deja 75% por debajo del valor y 25% por arriba del valor

Percentil 25 = $\dfrac{n+1}{4}$ Percentil 50 = $\dfrac{n+1}{2}$ Percentil 75 = $\dfrac{n+1}{1.33}$

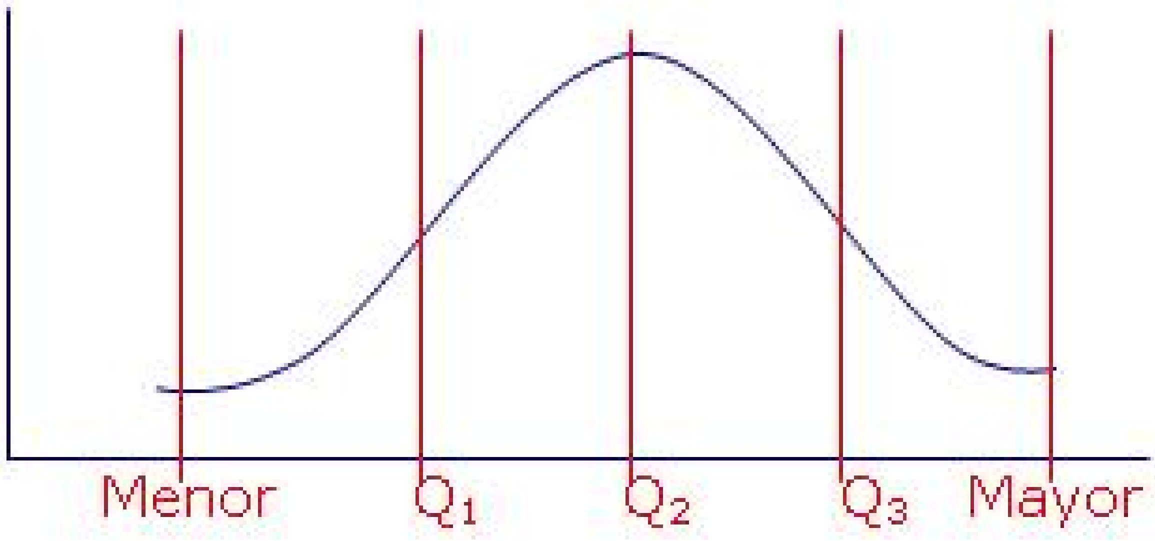
Ejemplo: Concentración de CK (creatinin-fosfokinasa)

$P_{25} = \dfrac{36+1}{4} = 9.25 \downarrow$ **67**

$P_{50} = \dfrac{36+1}{2} = 18.50 \downarrow$ **94**

$P_{75} = \dfrac{36+1}{1.33} = \dfrac{37}{1.33} = 27.8 \downarrow$ **118**

Los 3 cuartiles dividen la distribución de los datos en 4 regiones:



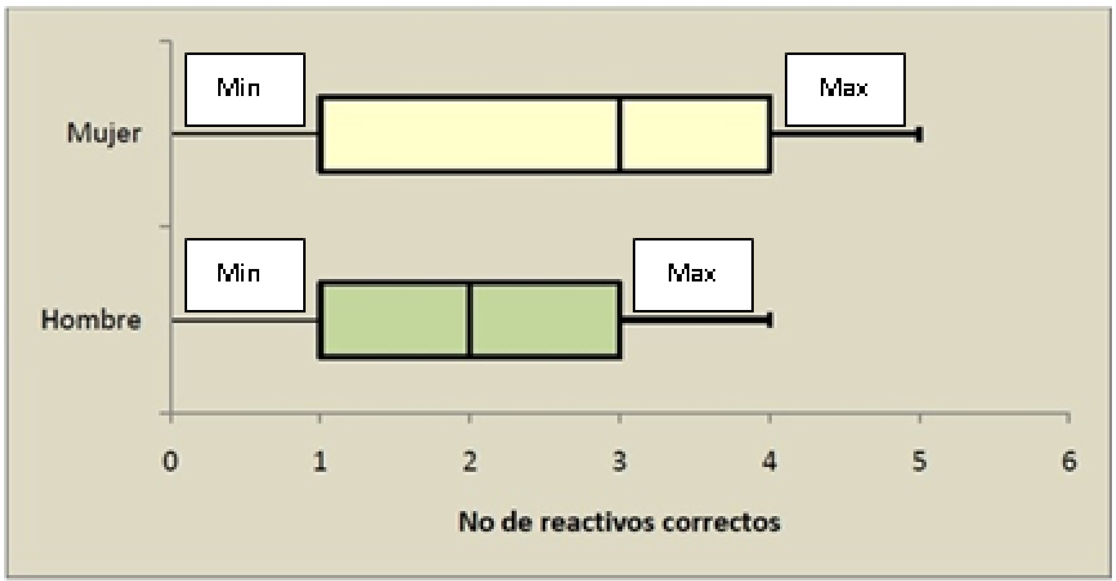
Rango intercuartil = IQR = Q3 − Q1 = 50% de las observaciones.

Gráfica de caja

- Representación gráfica de valor mínimo, cuartil inferior, mediana y cuartil superior.
- Resumen gráfico de la distribución de datos:

+ rango ⟶ dispersión y forma de distribución
+ IQR
+ mediana ⟶ centro

- Permite comparar la distribución de las dos poblaciones o muestras.



2.3 Propiedades de la probabilidad

Tipos de sucesos

Suceso seguro: se tiene la certeza que se producirá porque contiene todos los resultados posibles de la experiencia (coincide con el espacio muestral).

Suceso imposible: se tiene la certeza de que nunca se puede presentar, ya que no tiene elementos, es el conjunto vacío.

Suceso contrario de A: es el que ocurre cuando no se da A; es su complementario respecto al espacio muestral (A).

Suceso elemental: es el que tiene solo un resultado, es un conjunto unitario.

Sucesos incompatibles: La intersección es conjunto vacío, es decir, no pueden darse dos sucesos al mismo tiempo.

Sucesos compatibles: La intersección de dos sucesos contiene algún elemento.

Las características de la probabilidad son:

- La probabilidad de un suceso es mayor o igual a cero.

- La probabilidad del suceso seguro es uno.

- La probabilidad de la unión de dos sucesos incompatibles es igual a la suma de sus probabilidades.

2.3.1 Notaciones de probabilidad.

1. La suma de las probabilidades de un suceso y su contrario vale 1, por tanto la probabilidad del suceso contrario es:

$$p(\overline{A}) = 1 - p(A)$$

2. Probabilidad del suceso imposible es cero.

$$p(\varnothing) = 0$$

3. La probabilidad de la unión de dos sucesos es la suma de sus probabilidades restándole la probabilidad de su intersección.

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

4. Si un suceso está incluido en otro, su probabilidad es menor o igual a la de éste.

$$Si \;\; A \subset B, \text{entonces} \qquad p(A) \leq p(B)$$

5. Si A_1, A_2, ..., A_k son incompatibles dos a dos entonces:

$$p(A_1 \cup A_2 \cup \cdots \cup A_k) = p(A_1) + p(A_2) + \cdots + p(A_k)$$

6 Si el espacio muestral E es finito y un suceso es $S = \{x_1, x_2, ..., x_n\}$ entonces:

$$p(S) = p(x_1) + p(x_2) + \cdots + p(x_n)$$

2.4 Distribución de probabilidad

2.4.1 Distribución binomial

- Distribución de probabilidad.
- Basada en un modelo de eventos independientes. El modelo de eventos independientes es:
- Secuencia de procesos aleatorios.
- Cada proceso (prueba) tiene dos posibles resultados: éxito (**p**) o fracaso (**q**).
- Los procesos son independientes.

En el caso del albinismo….

- Si dos individuos portadores se "cruzan", cada uno de sus hijos tiene una probabilidad de 0.25 de ser albino……. AA= Normal, aa = albino, Aa =portador (normal).

X	A	a
A	AA	Aa
a	Aa	aa

Albino = 2

Albino = 25%

p = 0.25 (éxito)

q = 0.75 (fracaso)

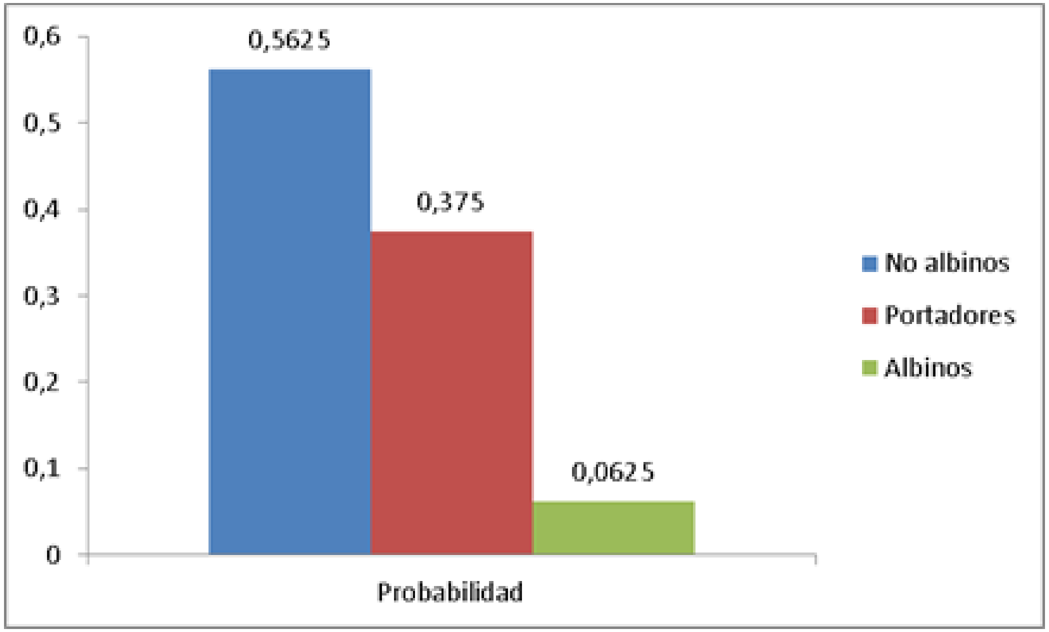
La probabilidad de que se obtengan 2 hijos albinos en dos eventos continuos es:

P (a, a) = 0.25 x 0.25 = 0.0625 (1/16)
P (A, A) = 0.75 x 0.75 = 0.5625 (9/16)
P (A, a) = 0.75 x 0.25 = 0.1875 (3/16)
P (a, A) = 0.25 x 0.75 = <u>0.1875 (3/16)</u>
1.00 = 16/16

Clasificación

Albino (p)	No albino (q)	Probabilidad
0	2	0.5625 (9/16)
1	1	0.375 (6/16)
2	0	0.0625 (1/16)

Distribución de probabilidades para una pareja de portadores en dos partos (2 hijos)



Fórmula para distribución binomial.

Para la variable aleatoria la probabilidad de que “n” procesos aleatorios resulten en “j” éxitos (y – n – “j” fracasos), está dada por la siguiente ecuación:

P (j éxitos) $= nCj\ [P^{J}\ (1-p)^{n-J}]$ nCj = Coeficiente binomial

$$nCj = \frac{n!}{j!\,(n-j)!}$$

! = factorial

Factorial 0! = 1, 1! = 1 (1 x 1), 2! = 2 (2 x 1), 3! = 6 (3 x 2 x 1), 4! = 24 (4 x 3 x 2 x 1), 5! = 120 (5 x 4 x 3 x 2 x 1), 6! = 720 (6 x 5 x 4 x 3 x 2 x 1)

Ejemplo………

Si "n" = 5 (procesos aleatorios), j: 0, 1, 2, 3, 4, 5

Si j = 0……..n-j = 5 – 0 = 5 $\frac{5!}{0!(5!)} = \frac{120}{120} = 1$

Si j = 1……..n-j = 5 – 1 = 4 $\frac{5!}{1!(4!)} = \frac{120}{24} = 5$

Si j = 2……..n-j = 5 – 2 = 3 $\frac{5!}{2!(3!)} = \frac{120}{12} = 10$

Si j = 3……..n-j = 5 – 3 = 2 $\frac{5!}{3!(2!)} = \frac{120}{12} = 10$

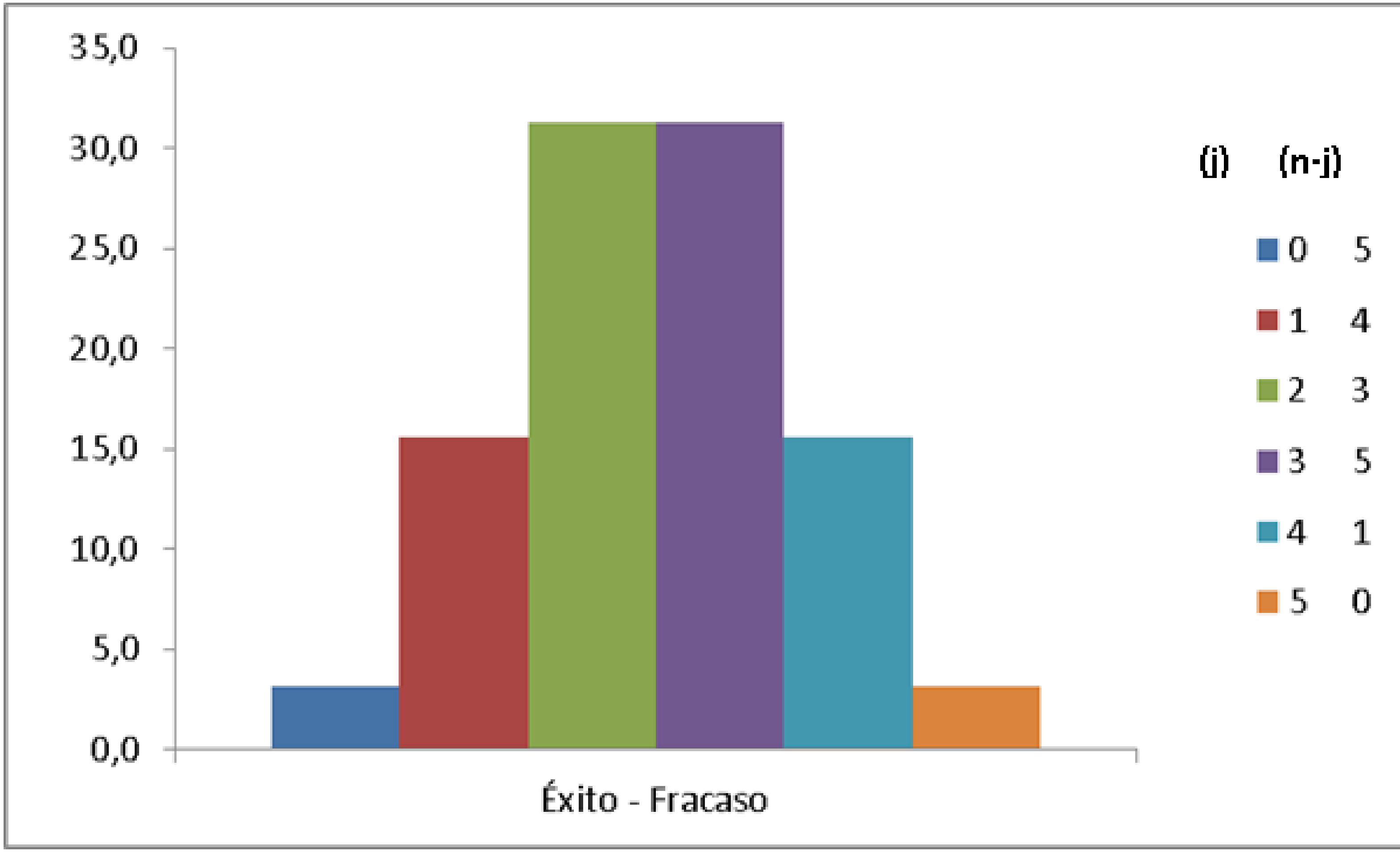
Si j = 4……..n-j = 5 – 4 = 1 $\frac{5!}{4!(1!)} = \frac{120}{24} = 5$

Si j = 5……..n-j = 5 – 5 = 0 $\frac{5!}{5!(0!)} = \frac{120}{120} = 1$

Solución…….

Éxito (j)	Fracaso (n-j)	$nCj[P^j\ (1\text{-}p)^{n\text{-}j}]$	$nCj[P^j\ (1\text{-}p)^{n\text{-}j}]$	$nCj[P^j\ (1\text{-}p)^{n\text{-}j}]$	Resultado = %
0	5	$1\ [(0.5)^0\ (1 – 0.5)^{5-0}] =$	$1\ [(0.5)^5] =$	1 [(1) (0.03125)] =	0.03125 = 3.125%
1	4	$5\ [(0.5)^1\ (1 – 0.5)^{5-1}] =$	$5\ [(0.5)\ (0.5)^4] =$	5 [(0.5) (0.0625)] =	0.15625 = 15.625%
2	3	$10\ [(0.5)^2\ (1 – 0.5)^{5-2}] =$	$10\ [(0.25)\ (0.5)^3] =$	10 [(0.25) (0.125)] =	0.3125 = 31.25%
3	2	$10\ [(0.5)^3\ (1 – 0.5)^{5-3}] =$	$10\ [(0.125)\ (0.5)^2] =$	10 [(0.125) (0.25)] =	0.3125 = 31.25%
4	1	$5\ [(0.5)^4\ (1 – 0.5)^{5-4}] =$	$5\ [(0.0625)\ (0.5)^1] =$	5 [(0.0625) (0.5)] =	0.15625 = 15.625%
5	0	$1\ [(0.5)^5\ (1 – 0.5)^{5-5}] =$	$1\ [(0.03125)\ (0.5)^0] =$	1 [(0.03125) (1)] =	0.03125 = 3.125%

Graficando……



Otro ejemplo:

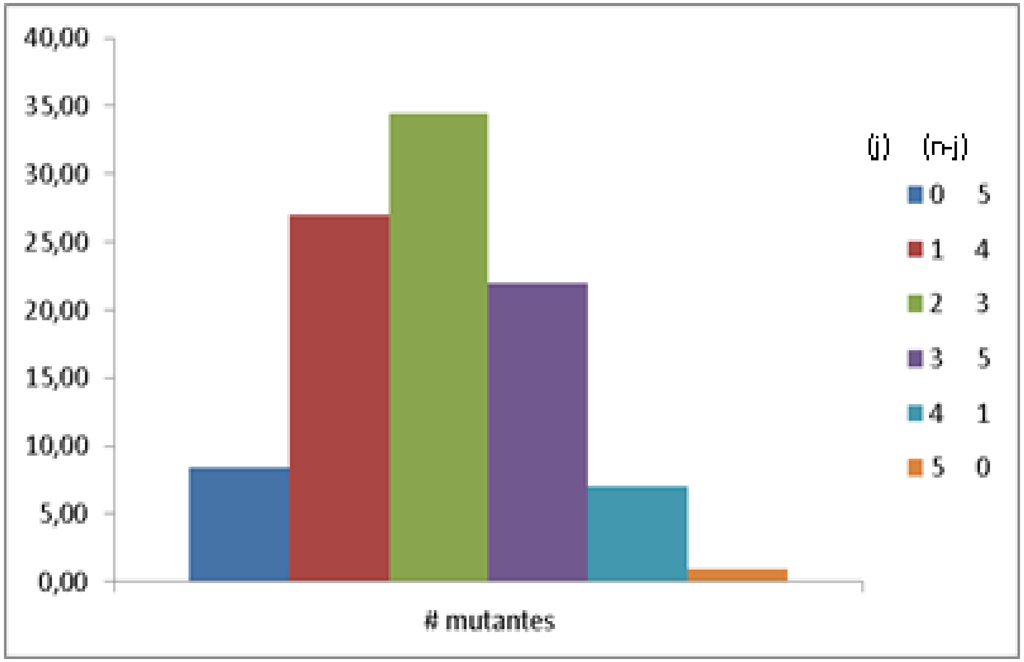
Muestra aleatoria de n = 5, de una población donde el 39% son animales mutantes.

$n = 5 \quad p = 0.39$

Solución.......

Éxito (j)	Fracaso (n-j)	$nCj[P^j (1-p)^{n-j}]$	$nCj[P^j (1-p)^{n-j}]$	$nCj[P^j (1-p)^{n-j}]$	Resultado = %
0	5	$1 [(0.39)^0 (1-0.39)^{5-0}] =$	$1 [(1) (0.61)^5] =$	1 [(1) (0.0844)] =	1 (0.844) = 0.0844 = 8.44%
1	4	$5 [(0.39)^1 (1-0.39)^{5-1}] =$	$5 [(0.39) (0.61)^4] =$	5 [(0.39) (0.1384)] =	5 (0.054) = 0.270 = 27.00%
2	3	$10 [(0.39)^2 (1-0.39)^{5-2}] =$	$10 [(0.1521) (0.61)^3] =$	10 [(0.1521) (0.227)] =	10 (0.0345) = 0.345 = 34.50%
3	2	$10 [(0.39)^3 (1-0.39)^{5-3}] =$	$10 [(0.0593) (0.61)^2] =$	10 [(0.059) (0.3721)] =	10 (0.22) = 0.220 = 22.00%
4	1	$5 [(0.39)^4 (1-0.39)^{5-4}] =$	$5 [(0.023) (0.61)^1] =$	5 [(0.023) (0.61)] =	5 (0.01403) = 0.070 = 7.00%
5	0	$1 [(0.39)^5 (1-0.39)^{5-5}] =$	$1 [(0.009) (0.61)^0] =$	1 [(0.009) (1)] =	1 (0.009) = 0.009 = 0.9%

Graficando.......



2.4.2 Distribución normal

La distribución de una variable cuantitativa (y) está definida por:

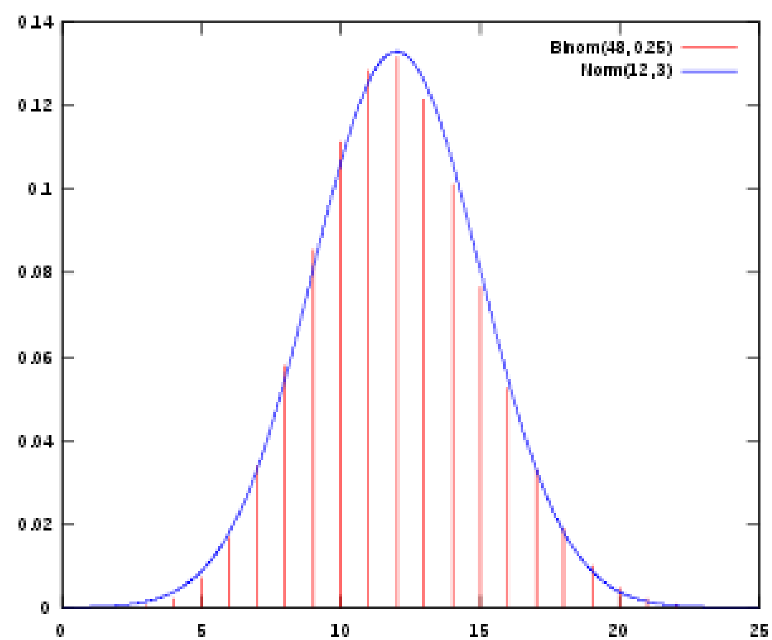
a) Media (μ)

b) Desviación estándar (σ)

c) Forma de la distribución: Función de densidad, que representa las frecuencias relativas como áreas bajo la curva.

La curva de distribución normal es:

- Simétrica y
- de forma de campana
- la cual va a estar definida por la media (μ) y la desviación estándar (σ).



Si la distribución de la variable (y) sigue una distribución normal con media (μ) y desviación estándar (σ), esto se escribe:

$$y \sim N(\mu,\sigma)$$

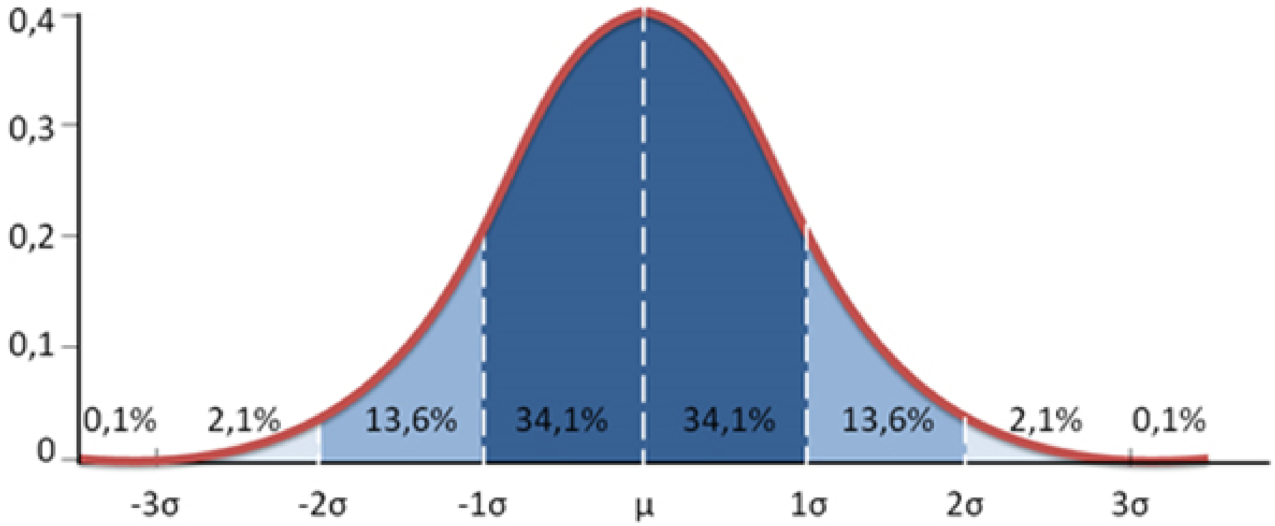
Si "y" tiene una distribución normal (μ,σ), entonces la densidad que describe a la curva de distribución de "y" está dada por la siguiente fórmula:

$$f(y) = \left[\frac{1}{\sigma\sqrt{2\pi}}\right] = (e)^{-1/2\,\frac{(y-\mu)2}{\sigma}}$$

$\pi = 3.1416$

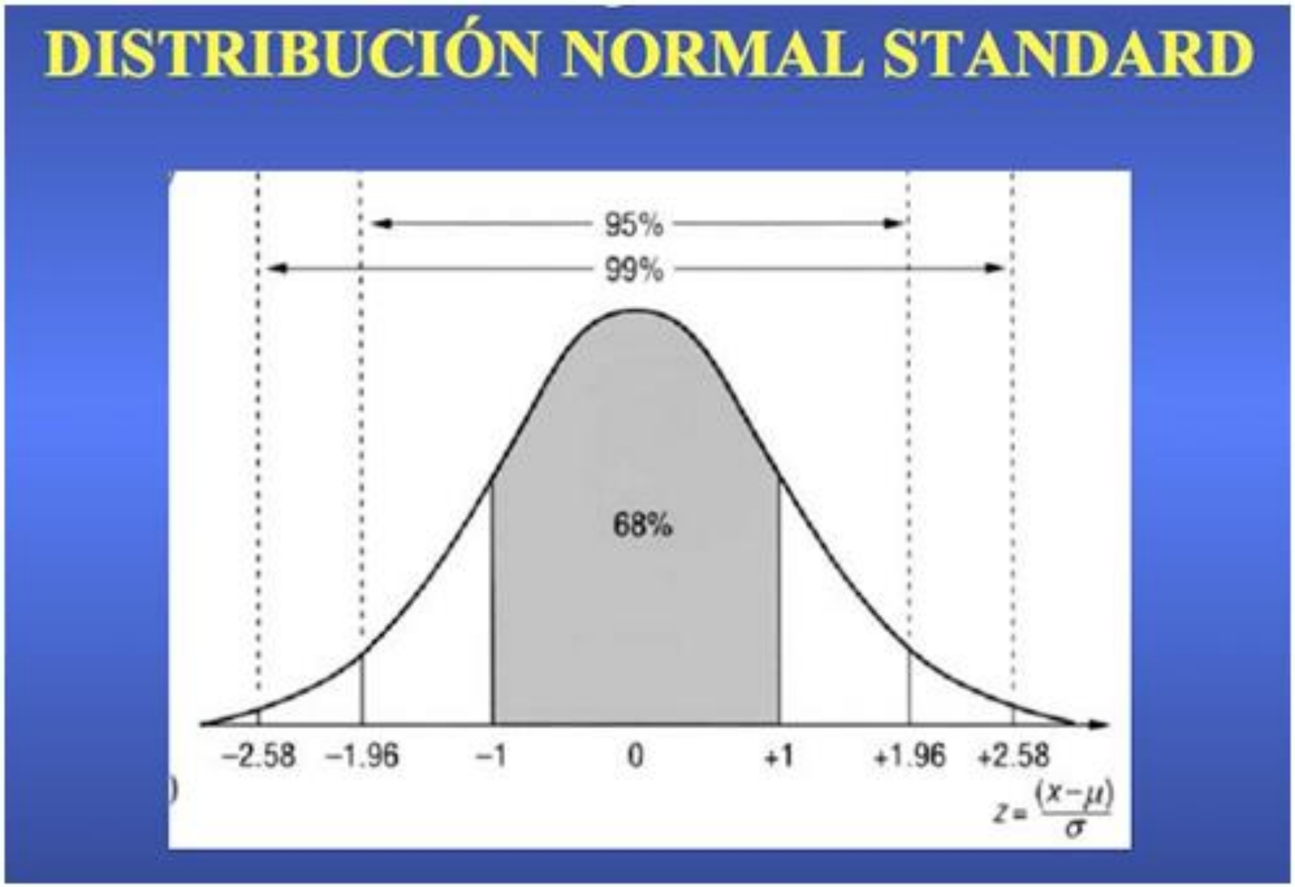
$e = 2.71$

f (y): Función de densidad de la distribución y expresa la altura de la altura como función de la posición de y sobre el eje de los valores "y".



- Simétrica ⟶ centrada en la μ (media).
- Forma de campana
- Se extiende de -0.0 a + 0.0
- El área bajo la curva es igual a 1.
- Asintótica (es una curva a la cual se acerca indefinidamente otra, sin llegar jamás a coincidir)

$\mu \pm 1\ \sigma = 68\%$ $\mu \pm 2\ \sigma = 95\%$ $\mu \pm 3\ \sigma = 99\%$



Área bajo la curva normal.

- Las áreas bajo la curva han sido calculadas y tabuladas en tablas.
- Utiliza una escala basada en desviaciones estándar con respecto a la media.
- Es necesario cambiar la escala de variable o "**estandarizada**".
- Las variables estandarizadas se denota como "z" y se refiere como normal estándar.

$$Z = \frac{y - \mu}{\sigma} \quad o \quad z = \frac{y - \bar{y}}{s}.$$

$$Z \sim N(0,1) \longrightarrow \text{Curva normal estándar}$$

- Al estandarizar una observación, sed mide la distancia que han entre la observación y la media (desviación) en desviaciones estándar.

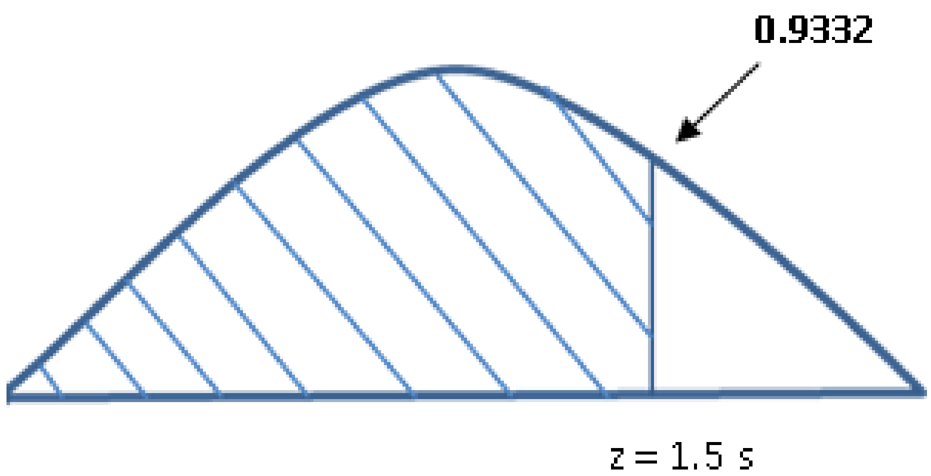
Presentación del caso concentración de CK (U/L, suero de sementales ovinos):

		x_1	x_2	x_3
$\bar{X} = 98.25$	$s = 40.38$	$x_i = 25$	$x_i = 94$	$x_i = 119$

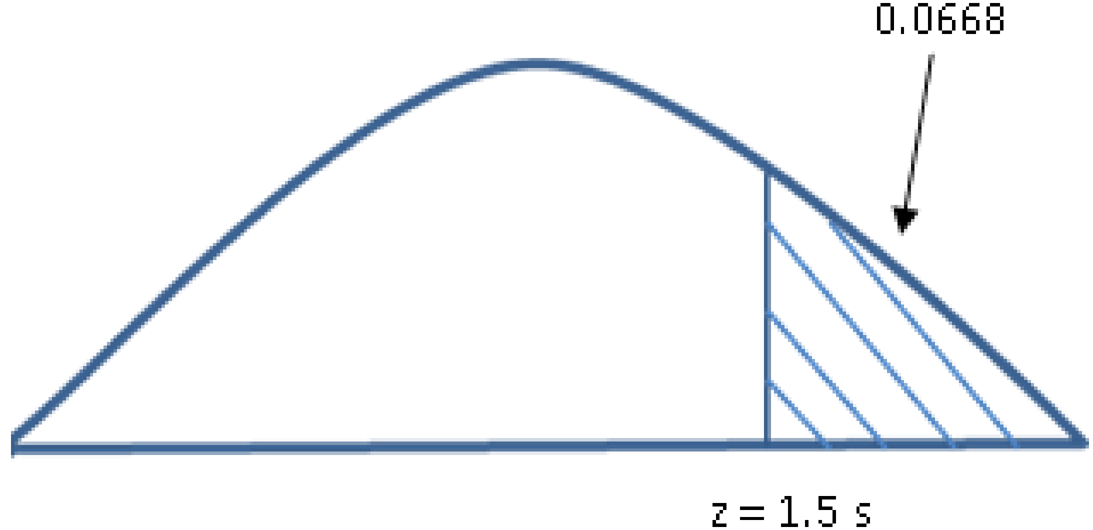
$$z = \frac{25 - 98.27}{40.38} = \mathbf{-1.81} \qquad z = \frac{94 - 98.27}{40.38} = \mathbf{-0.118} \qquad z = \frac{119 - 98.27}{40.38} = \mathbf{0.51}$$

- Si queremos conocer un área por debajo de un valor de “z”

z= 1.5 s

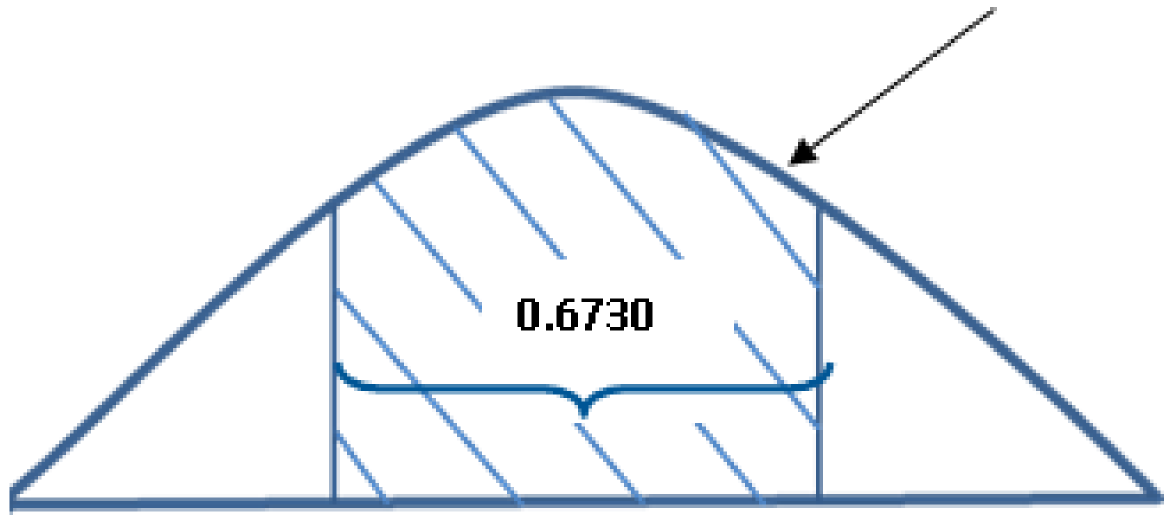


- Si queremos conocer el área por arriba de un valor de "z", $1 - 0.9332 = \mathbf{0.0668}$



- Si queremos conocer el área entre 2 valores "z", $z = -1.2$ y $z = 0.8$,

$0.8 = 0.7881$, $-1.2 = 0.1151$, $0.7881 - 0.1151 =$ **0.6730**



En una población de peces se sabe que la longitud se distribuye de forma normal.

$\overline{X} = 54.0$ mm $\qquad s = 4.5$ mm

a) Qué porcentaje de los peces mide 60 mm o menos?

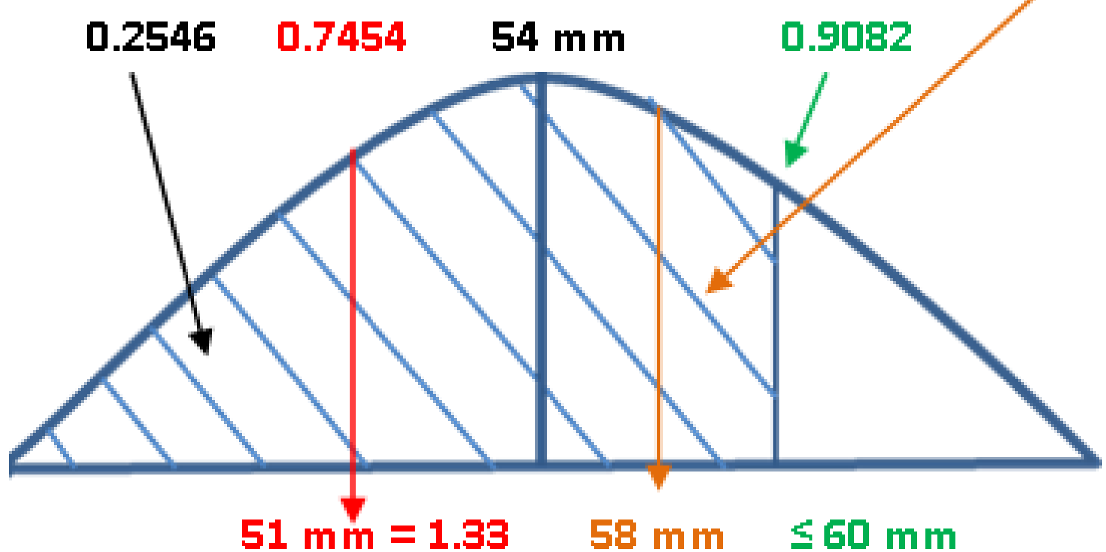
b) Qué porcentaje de los peces son más largos de 51 mm?

c) Qué porcentaje de los peces miden entre n58 y 60 mm de largo?

a) $\overline{X} = 54.0$ mm, $\quad z = \dfrac{60 - 54}{4.5} = \dfrac{6}{4.5} = 1.33 = z =$ **0.9082**

b) $\dfrac{51 - 54}{4.5} = \dfrac{-3}{4.5} = -0.66 = z =$ **0.2546**, $= 1.0 - 0.2546 =$ **0.7454**

c) $z = \dfrac{58 - 54}{4.5} = \dfrac{4}{4.5} = 0.888 = z = 0.8106 \qquad 0.9082 - 0.8106 =$ **0.0976**



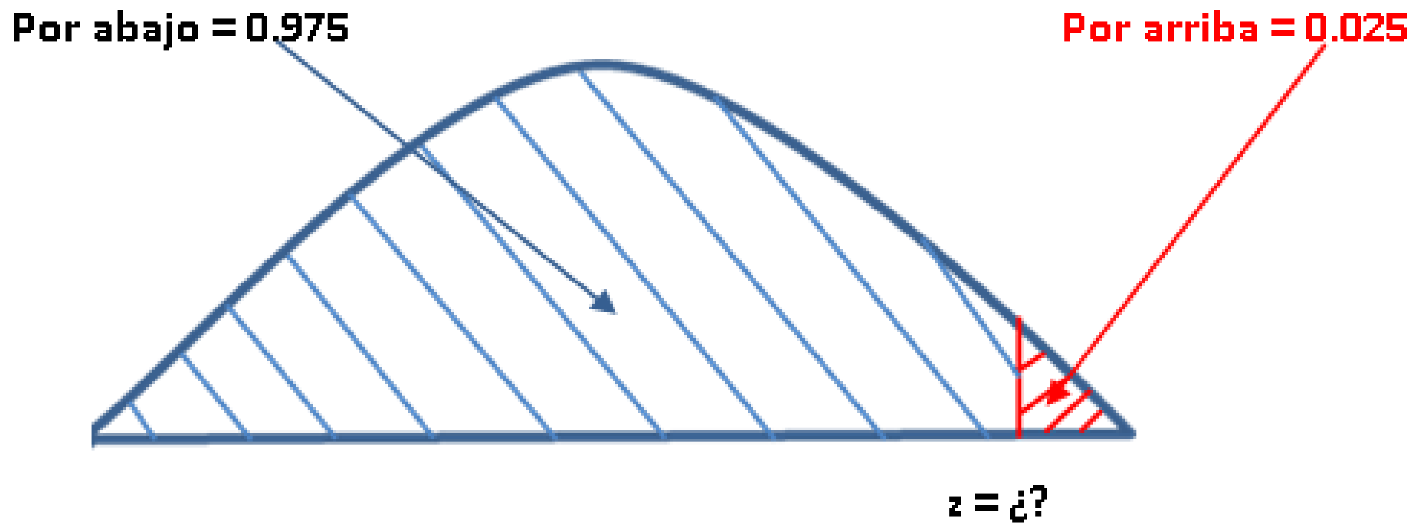
En términos de probabilidad

Y = el largo de un pez seleccionado aleatoriamente de la población.

$p = \{y \leq 60\} = 0.9082$ $p = \{y \geq 51\} = 0.7454$ $p = \{58 \leq y \leq 60\} = 0.0976$

El valor "z" corresponde a un área bajo la curva. El valor "z" que corresponde al punto de corte (valor crítico) que separa al 2.5% superior de la distribución.

1.0 – 0.025 = 0.975, buscando el valor de **z** en tablas = **1.96**



¿Cuál es el estadístico descriptivo que divide a la distribución de los datos?

Cuál es el percentil 70 para la longitud de la población de peces?

1.0 – 100% $\overline{X} = 54$ mm $s = 4.5$ mm

0.7 – 70%

0.700 0.6985 $z = 0.52$

* Convertir los valores z a la escala en que se mide la variable

$y = (s)(z) + \mu$ $y = (4.5)(0.52) + 54 = 56.3$ mm $P_{(70)}$

- Muchos procedimientos estadísticos están basados en tener datos de una población con ***distribución normal***.

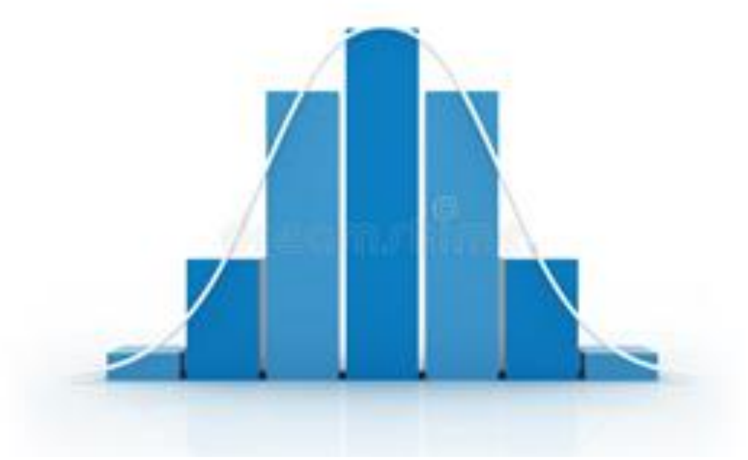
Existen formas de probar si los datos tienen una distribución lo suficientemente parecida a un modelo de distribución normal.

1.- Si y sigue una distribución normal.

- 68% de y se encuentran en $\bar{y} \pm 1$ s
- 95% de y se encuentran en $\bar{y} \pm 2$ s
- 99.7% de y se encuentran en $\bar{y} \pm 3$ s

2.- Gráfica de probabilidad normal.

En ocasiones es difícil saber si un conjunto de datos sigue una distribución normal únicamente utilizando un Histograma.



Referencias bibliográficas

- Aviva P, Watson P (2006). Statistics for veterinary and animal science. Oxford, Ames, Iowa. Blackwell Publishing. (ISBN: 9781405127813)

- Daniel, Wayne W (1997). Bioestadística: Base para el Análisis de las Ciencias de la Salud. 3a ed. Ed UTEHA. México. (ISBN: 968-18-596-X).

- Navarro Fierro, Ricardo (1988). Introducción a la Bioestadística. Ed. Mcgraw-Hill. México, DF. (ISBN: 9789684223875)

- Steel, Robert George Douglas Torrie, James Hiram, Martínez B, Ricardo, TR (1987). Bioestadística: Principios y Procedimiento. Ed McGraw-Hill México, DF. (ISBN: 9789684514959)