

9

La Normalidad

En este capítulo se trata el concepto de normalidad en Estadística. Se comienza con un primer cuadro comparativo del concepto, en diversas especialidades o campos, y se propone un método para evitar las confusiones comunes con la normalidad. Luego se analiza con cuidado la idea de “valores normales” en clínica y se da un método para su obtención de acuerdo con las recomendaciones de la bibliografía. Finalmente se trata a la función normal o de Gauss, sus propiedades, usos y manejo de tablas para el cálculo de probabilidades.

9.1 ¿Qué es lo normal?

Cada especialidad parece tener su definición particular del concepto de normalidad. Una de las más difundidas en medicina es considerar al paciente sano como normal. Sin embargo, la Organización Mundial de la Salud define a la *salud* como un estado e bienestar físico, psíquico y social, y no solamente como la ausencia de enfermedad o deterioro. Lo toma como un estado que permite al organismo adaptarse y funcionar en forma adecuada, a las condiciones endógenas y a los factores ambientales a los que está sometido. Por otra parte, el concepto de *enfermedad* puede definirse como una respuesta inadecuada del organismo, o un agotamiento de sus mecanismos de defensa y adaptación, una falta de respuesta a los estímulos a los que está expuesto. El proceso puede ser una perturbación de la estructura y/o función de un órgano, de un sistema o de la totalidad del organismo.

Según Murphy, E.A., del Departamento de Medicina de la J.Hopkins University, con muchos trabajos dedicados a este tipo de clasificaciones, se pueden identificar cinco categorías de manifestaciones que pueden considerarse como una enfermedad, y siete significados diferentes del concepto de normalidad. Además de estas siete definiciones teóricas, todavía se pueden agregar cinco más desde el punto de vista pragmático, como se puede apreciar el Cuadro 9.1 Es necesario detenerse para analizar estos aspectos.

Las manifestaciones que pueden considerarse como enfermedades son cinco:

- 1) Determinadas respuestas homeostáticas (inflamación no supurativa)
- 2) Determinados mecanismos homeostáticos aberrantes (hipertensión u obesidad)
- 3) Respuestas que impiden una adaptación, o bien, respuestas inhabituales frente a los factores que las provocan (golpe de calor, desfallecimiento, hipertensión de origen renal)
- 4) Una respuesta inadecuada de origen endógeno (agammaglobulinemia)
- 5) Una respuesta totalmente anárquica (cáncer o tireotoxicosis).

A toda esta diversidad de aspectos se pueden agregar otros aspectos operativos que permiten identificar individuos, en los cuales es aconsejable una acción médica o sanitaria en forma preventiva. Entonces, a cualquier paciente en alguna de estas situaciones se lo cataloga como “enfermo o no normal”.

Cuadro 9.1 Diferentes aspectos del concepto de normalidad.

Noción	Dónde se usa	Término preferido
Clasificación teórica de Murphy		
1. Probabilística: a) Población total b) Individuos sanos	Estadística	Gaussiano
2. Más representativa en su clase	Ciencias descriptivas	Promedio, modal
3. Comúnmente observable en su clase	Ciencias descriptivas	Habitual
4. La más apta para la supervivencia	Genética, Invest. Operativa	Optimo, más apto
5. Que no causa daños ni implica castigos	Medicina Clínica	Inofensivo, no nocivo
6. A lo que se aspira en general	Política, Sociología	Convencional
7. Más perfecto en su clase	Matemática, estética, moral	Ideal
Clasificación de acuerdo con los aspectos prácticos		
8. Que engloba a los individuos que pueden ser atendidos	Sociología, política, economía administración.	Aceptable
9. Efecto más provechoso que nocivo	Terapéutica, Org. de salud	Máximo efecto
10. Ausencia de signos externos raros	Diagnóstico Profesional	Sano
11. Idem anterior, pero percibidos por uno mismo.	Anamnesis	Saludable
12. Ausencia de signos inaceptables para el prójimo	Vida comunitaria	Admisible

Siguiendo el cuadro de Murphy se pueden tener siete respuestas diferentes a una simple pregunta como: *¿Cuál es el nivel normal de colesterol?*

- 1) Una distribución de probabilidades, en forma de campana, descrita por la función de Gauss con los valores de colesterol en la población de referencia (un estadístico).
- 2) El valor más representativo del colesterol en la población definido por un valor esperado, o promedio general (un demógrafo).
- 3) Los valores de colesterol más comunes definidos por una gamma de límites, o sea los habituales y conocidos “límites de normalidad de los laboratorios” (un bioquímico).

- 4) Los valores de colesterol más adecuados para la reproducción y para la supervivencia de la especie estudiada (un genetista).
- 5) Los valores de colesterol inocuos que causan poco o ningún daño al organismo (un médico).
- 6) La opinión conjunta de una comisión, o sea los valores “aprobados de colesterol” por las autoridades respectivas (un político).
- 7) El valor “ideal” de colesterol (un utópico).

El primer uso de la palabra correspondiente a la Estadística causa el mayor problema porque tiene poco que ver con los demás significados. No hay ninguna razón científica para que un grupo de personas “normales” tenga una distribución gaussiana en alguna magnitud clínica que se les esté midiendo. Lamentablemente, las estadísticas “gaussianas” se emplean frecuentemente para determinar los “límites normales” aplicados a las pruebas del laboratorio de análisis clínicos. Al respecto, Galen hace mención a las conclusiones de Benson, director del laboratorio médico de la Universidad de Minnesota:

“Los límites normales han tenido un papel indefinido, pero tranquilizante en el laboratorio clínico. Asoman en el horizonte de nuestra conciencia, perfectamente simétricos como el Monte Fujiyama, un poco nebulosos en su significado, pero aceptados y reverenciados con gratitud. Sin embargo, lejos de ser puros y simples, como una querida ilusión infantil, examinados de cerca resultan ser insoportablemente complicados, siendo en realidad uno de los problemas más difíciles y que más limitan la utilidad de los datos de un laboratorio clínico”.

Como alternativa al problema de los “límites normales”, varios investigadores han adoptado los *límites de referencia*, de uso cada vez más frecuente en la bibliografía y laboratorios, que se analizarán en el próximo punto.

En uno de los trabajos recomendados por la Clinical Chemistry para el análisis estadístico de los datos del laboratorio, Wu, Towmey y Thiers examinaron publicaciones previas, revelan que en virtud de la *costumbre* las técnicas estadísticas usadas casi invariablemente se basan en el supuesto de normalidad (distribución gaussiana de probabilidad) en la distribución de los datos obtenidos. Sin embargo, estos investigadores encontraron que a menudo los datos clínicos no adhieren a una función de Gauss. Entre los estadísticos hay dos corrientes de pensamiento. Un grupo, con el enfoque clásico, piensa que el supuesto gaussiano es lo suficientemente *robusto* como para aguantar discrepancias usuales, es decir que los modelos *paramétricos* para análisis de datos, basados en un supuesto de normalidad, sirven igual aun cuando tal supuesto no se cumpla estrictamente. Esto, por las condiciones generales del Teorema Central del Límite. Pero hay otra forma de pensar; entre los que no creen en tamaña robustez de los métodos clásicos, piensan que es mejor utilizar los modelos *no paramétricos*, los cuales no requieren el supuesto de normalidad aunque sean menos sensibles que los primeros. Depende de cada caso particular, y del área de estudio específica analizada, para saber cuál de las maneras es mejor. Estos investigadores estudiaron doce magnitudes clínicas (constituyentes bioquímicos) en suero, obtenidos de más de 1400 individuos en una muestra controlada de hombres y mujeres, y ninguno de ellos se ajustaba bien a una distribución normal.

A pesar de todas estas observaciones, es indudable que por mucho tiempo el supuesto de normalidad estadístico se seguirá usando en los laboratorios. Y por ello debe ser estudiado con cuidado, además del uso intensivo que tiene en la materia Estadística.

9.2 Criterios de normalidad

En principio existen tres criterios de normalidad diferentes, a tener en cuenta:

1. *El criterio matemático o estadístico puro.*
2. *El criterio de referencia a la población de donde se extrajeron los datos.*
3. *El criterio operativo y lógico.*

Los individuos identificados como anormales o enfermos son aquellos que deben recibir asistencia porque son susceptibles de ser tratados. Esta anomalía se asocia a una patología definida por otros estudios. Por lo tanto, un valor “normal”, en el sentido de “sano”, de una función fisiológica o de una morfología determinada, puede tener esos tres significados:

1. Desde el punto de vista *estadístico* un valor puede ser *normal* por la ubicación que él tiene dentro de un intervalo, donde están la mayoría de las observaciones realizadas. Por ejemplo, se ubica en el intervalo que comprende al 95% de las personas consideradas con colesterol normal. Esta definición no es tan aceptable desde el punto de vista epidemiológico o médico. Tomando por caso una población que viva en un medio determinado, como ser el Altiplano boliviano, a miles de metros de altura respecto al nivel del mar, y que posea un patrimonio genético definido. La medida de una función (tensión arterial, número de glóbulos rojos, etc.) variará de un individuo a otro, pero se puede definir un intervalo que englobe al 95% y 99% de la población. Si se adopta en forma mecánica a esos valores como umbrales de normalidad, se corre el riesgo de que al examinar a un habitante costero se lo tome por anémico y se lo trate inútilmente. En las alturas, el efecto de apunamiento no le ocurre a los nativos porque su sangre tiene un número mayor de glóbulos rojos que el promedio de otros habitantes del globo. Al tener mayor capacidad de transportar oxígeno, el efecto de enrarecimiento del aire por la altura se amortigua grandemente, cosa que no le ocurre al habitante de las llanuras. Esta capacidad se transmite genéticamente de generación en generación, y pasa a ser una característica de la población del Altiplano. Comparados con ellos, todos los demás tendrán un hematocrito y un recuento de rojos mucho menor. Entonces, si se toma a los del Altiplano como referencia de normalidad, todos los otros parecerían anémicos. Esto conduce al siguiente punto.

2. La *normalidad* puede variar según las condiciones de vida y desarrollo de los individuos; por ejemplo, es un hecho probado que los descendientes de japoneses que habitan California son mucho más altos en promedio que los que habitan el Japón. La explicación básica parece ser la manera de alimentarse de unos y de otros. En USA, la cantidad ingerida de productos lácteos, cereales y otros alimentos vitaminizados artificialmente, es mucho más grande que la consumida en Japón, por lo que el desarrollo de los huesos no es tan grande. Entonces:

- Cuando una población vive en condiciones adecuadas los valores obtenidos pueden ser considerados óptimos, típicos o ideales, en ausencia de factores nocivos para la magnitud clínica analizada. Se tratará de valores normales en condiciones de vida normales.

- Cuando una población está expuesta a factores ambientales, sanitarios y alimenticios inadecuados, mostrará valores típicos que pueden llegar a diferir bastante de aquellos obtenidos entre los que poseen mejores condiciones de vida. Para estos casos, la normalidad se corresponde con lo

ordinario o habitual, en tal tipo de poblaciones. Entonces, una minoría de la población considerada como ordinaria está sana en realidad.

- Cuando se ajustan los datos obtenidos, según edad y sexo, o cualquier otro criterio adoptado, es necesario considerar el hecho de que se establezcan categorías por la acumulación de situaciones de vida desfavorables. Por ejemplo, si al envejecer un individuo urbano no sigue un régimen alimenticio favorable, ni hace la cantidad de ejercicio físico necesaria, y vive sometido a tensiones psicológicas mucho mayores que las de un individuo en un medio rural, entonces cabe esperar: obesidad y tensión arterial elevada. Ahora, el deterioro que se puede llegar a observar en tales individuos es por el deterioro progresivo de los órganos por envejecimiento, o más bien, es el resultado acumulado de condiciones desfavorables de vida.

3. Un valor es *normal* cuando forma parte de la proporción de observaciones aceptables para toda la comunidad desde un punto de vista individual, social y económico. Habida cuenta de las posibilidades terapéuticas o de otro tipo disponibles.

9.3 Valores de referencia o normales

En una industria farmacéutica dedicada a la fabricación de medicamentos, productos de belleza y derivados, los valores de referencia están referidos a los *límites de tolerancia* de los productos terminados. Cada producto tiene una fórmula con la composición de los productos activos y los complementos. En particular, el certificado de aptitud del medicamento emitido por la autoridad sanitaria es otorgado bajo ciertas condiciones que establecen las cantidades mínimas y máximas de los principales componentes. Dichas cantidades son los límites de tolerancia de fabricación, dentro de los cuales debe encuadrarse el producto para ser considerado “normal” y no ser desechado.

El producto final de un método de análisis clínicos usado en el laboratorio es un resultado relacionado con una muestra de un determinado paciente. El médico interpreta este resultado como indicativo de variación, ausencia de esta, o posible cambio en el estado de salud de su paciente. Para dar sentido a esta interpretación, el médico compara el resultado obtenido con algún *sistema de referencia* indicativo de la “normalidad”. Este ámbito de referencia de un cuadro clínico generalmente se construye con los valores de un intervalo en la magnitud clínica, que incluya al 95% de los individuos considerados sanos de la población. Por desgracia, muchas veces el médico interpreta estos extremos como indicativos de los límites superior e inferior dentro de los que se debe ubicar el paciente para ser considerado sano, según hacen mención Winkel y Statland en sus trabajos. Ellos explican que en la década del setenta el término valores normales se asimilaba al concepto de individuos sanos, lo cual condujo a confusiones en el contexto de la química clínica. Se ha propuesto usar los términos de individuos sanos en lugar de normales, y distribuciones de datos gaussianas en lugar de normales, para evitar este tipo de confusiones. Galen y Gambino (1975) analizaron el tema con detenimiento, y Grasbeck (1969) introdujo el concepto de *valores de referencia* para abolir el término de valores normales. Hay dos tipos de valores de referencia: los basados en una población y los basados en un individuo. Aquí se tratará el primer caso y se definirán estos valores de la manera siguiente:

Los valores de referencia de una magnitud clínica son un grupo de datos de la magnitud medida de un grupo de individuos (o de uno solo de ellos), en un definido estado de salud.

Para utilizar este concepto, el médico evalúa un resultado del laboratorio dentro de un cierto ámbito, para calcular la probabilidad de que el paciente del que se tomó la muestra analizada pertenezca al grupo de individuos sanos. En cambio, cuando el médico compara con los datos del mismo individuo cuando este se encontraba en un estado de buena salud, lo que hace es ver si todavía se mantiene dentro del mismo estado de salud.

No se deben usar los límites de referencia como rígidas fronteras dentro de las cuales se encuentran los pacientes sanos, y los anormales en su exterior.

Por ejemplo, un paciente con un valor sérico de colesterol situado por debajo del límite inferior podría ser considerado como un signo de buena salud, en lugar de lo contrario. Otro caso, no es un buen síntoma para un diabético el tener sus valores de glucosa dentro del recinto de “normalidad”. La idea central, es nunca usar el sistema de referencia en forma aislada sin tener en cuenta todo el cuadro clínico, para no llegar a conclusiones erróneas.

Para poder construir los intervalos de referencia de una magnitud clínica y poder usarlos convenientemente, siempre se deben explicitar los siguientes cinco puntos:

- 1) *La población de referencia y la forma en que se eligió la muestra estudiada.*
- 2) *Las condiciones ambientales y fisiológicas bajo las cuales se extrajeron las muestras.*
- 3) *La técnica de extracción, el momento de recolección, el transporte, la preparación y la conservación de las muestras tomadas.*
- 4) *El método analítico usado, con datos acerca de su exactitud, precisión y control de calidad*
- 5) *La cantidad de datos observados y los intervalos de referencia derivados.*

Como ejemplo de aplicación se usa un trabajo de Galen (1983) donde el autor evaluó los datos históricos de pacientes externos ambulatorios que buscaban cuidado médico primario de tipo privado. Estos pacientes recibieron un perfil bioquímico de 26 pruebas como parte de una evaluación sanitaria de rutina o una investigación diagnóstica en este marco. Para cada constituyente se identificaron 1000 pacientes para cada grupo de edades y para cada sexo. No hubo subselección por otros factores como ser: origen étnico, ocupación, peso corporal, ubicación geográfica, etc. Los datos se tomaron simplemente por su disponibilidad en cuanto a edad y sexo dentro de los pacientes ambulatorios. No se tuvieron en cuenta adicionales diagnósticas, ambientales y fisiológicas, ni se dispuso de ellas. Las muestras venosas se tomaron después de aplicar torniquete, el suero se separó a los 30 a 45 minutos después de la recolección. La muestras de suero se guardaron a temperatura ambiente y se analizaron entre las horas 12.00 y 18.00. después de su extracción. La documentación estadística de precisión y exactitud de cada procedimiento realizado se acumula diariamente. Se agruparon los datos por sexo, y dentro de cada uno por grupo de edades, comenzando de 10 a 14 años y así sucesivamente con el mismo ancho hasta los de 99 años. Se muestra el cuadro de valores (Cuadro 9.2) de calcio en mg/dl, medido con el método de

auto químico (equipo automático) de la reacción de complejación de sulfonato de alizarina. Teniendo en cuenta la gran cantidad de datos (1000 en cada celda), los autores supusieron una distribución gaussiana porque razonaron "...Como la definición de una población sana o normal, como la población de referencia para estos estudios es larga, costosa y quizás imposible, el análisis de *grandes muestras* evita estos requerimientos...". La idea básica es desarrollar en cada celda un histograma con los 1000 datos disponibles, y obtener los percentiles 0,5; 2,5; 50; 97,5 y 99,5 como los límites de referencia para cada celda.

Cuadro 9.2 Valores de referencia del calcio sérico en mg/dl.

EDAD	HOMBRES				MUJERES			
	Percentiles				Percentiles			
años	0,5	2,5	97,5	99,5	0,5	2,5	97,5	99,5
10 a 14	8,1	9,0	10,7	11,3	7,9	8,9	10,7	11,8
15 a 19	8,2	8,8	10,7	11,4	8,5	8,8	10,7	10,9
20 a 24	8,6	8,9	10,8	11,4	8,2	8,7	10,5	10,8
25 a 29	8,8	8,9	10,6	10,8	8,4	8,8	10,4	10,9
30 a 34	8,7	8,9	10,8	11,3	8,3	8,7	10,4	10,7
35 a 39	8,5	8,9	10,6	11,0	8,5	8,8	10,4	10,8
40 a 44	8,8	8,9	10,6	10,8	8,0	8,5	10,5	10,7
45 a 49	8,7	8,8	10,6	10,9	8,4	8,8	10,6	10,8
50 a 54	8,5	8,8	10,6	11,1	8,1	8,8	10,6	10,9
55 a 59	8,4	8,7	10,5	11,0	8,1	8,8	10,6	10,9
60 a 64	8,2	8,7	10,6	10,9	8,4	8,8	10,7	11,3
65 a 69	8,2	8,6	10,6	10,8	8,5	8,8	10,7	11,3
70 a 74	8,3	8,8	10,7	11,3	8,3	8,8	10,6	11,3
75 a 79	8,0	8,5	10,5	11,0	8,2	8,7	10,6	11,0
80 a 84	7,6	8,5	10,4	10,7	8,3	8,7	10,7	10,9
85 a 89	8,1	8,5	10,6	11,7	7,6	8,6	10,5	11,1

FUENTE: Sonnenwirth AC y Jarret L, *Métodos y diagnósticos del laboratorio clínico*. Pág. 38.

A veces se confunde el término "valor referente", que es un sinónimo de punto de corte, valor crítico o valor de discriminación, con los valores límites del intervalo de referencia. Un punto de corte no puede determinarse en la misma forma que se han determinado los valores de referencia, estudiando individuos sanos o "normales". Se debe determinar estudiando el constituyente en personas sanas y con distintos estados de avance de la enfermedad. De esta forma, se pueden determinar valores referentes o puntos de corte que permitan un máximo de discriminación entre sanos y enfermos; o entre una enfermedad en particular y otras incluidas en el diagnóstico diferencial.

Por ejemplo, si un médico está tratando a una paciente de 52 años, los valores de referencia que debería utilizar están entre 8,8 y 10,6 mg/dl, con una probabilidad asociada del 95% o entre 8,1 y 10,9 mg/dl con una probabilidad del 99%. Para resumir, si se quiere prescindir de la edad o del sexo, se buscan los valores extremos en la tabla para generalizar, al estilo usual de la vieja bibliografía. Se puede ver que entre los valores 7,6 y 11,8 mg/dl se encuentran todos los datos tabulados, luego se podría decir que el Calcio tiene ese rango de valores de referencia en la población humana de donde se extrajo la muestra.

Debido a la forma clásica de definir la “normalidad” en los laboratorios de análisis clínicos, muchos resultados anormales no son identificados como tales, no anuncian estados de enfermedad ni de riesgo para la salud. La respuesta no es tan solo una recolección de datos elaborados con intervalos de referencia. Para un buen diagnóstico de laboratorio se deben considerar varios factores no patológicos, responsables de dar señales de falsos positivos, además de valores referenciales para pruebas individuales de laboratorio que son buenos anunciadores de la enfermedad. Cuando el perfil bioquímico del paciente se evalúe así se habrá avanzado un buen paso en patología clínica.

Para asegurarse de tener una distribución normal en un experimento cualquiera de un laboratorio, se pueden medir alícuotas de una misma muestra en condiciones estables teniendo en cuenta que: *Mediciones repetidas de la misma magnitud tienen una distribución gaussiana.*

9.4 La función de Gauss o “la normal”

Las distribuciones de probabilidad analizadas hasta ahora eran del tipo de variable discreta. Pero la gran mayoría de las magnitudes clínicas son del tipo continuo, cuya distribución teórica de la variable se llama *función de densidad* como se vió en el capítulo anterior. Recordando el Teorema Central del Límite cuando las muestras son grandes, las funciones de probabilidad discretas tienden a una continua: la función de Gauss. Esta función tiene forma de campana y se expresa con:

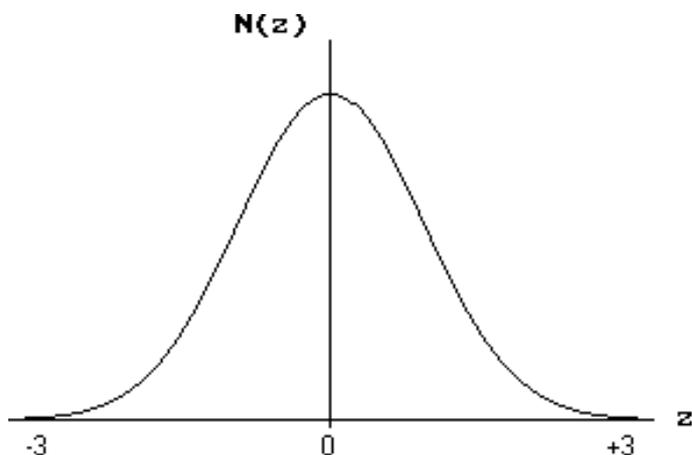
$$N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}$$

donde $N(x)$ indica la altura de la ordenada de la curva acampanada, que representa la función densidad de Gauss. Como se puede apreciar en la fórmula, existen infinitas funciones normales posibles. La manera de identificar en forma unívoca a una de ellas es dando los valores a sus dos parámetros μ y σ . Luego, para destacar este hecho se puede escribir $N_x(\mu, \sigma)$, para denotar a una función normal de la magnitud clínica x , con sus dos parámetros identificatorios.

De toda la familia gaussiana hay una curva en particular que se toma como referencia o patrón de las demás. Es aquella que se obtiene tipificando la variable x , o sea: $N_z(0, 1)$. Esta curva se muestra en el Gráfico 9.1 y sus valores están tabulados en el Tabla 1. Se la denomina curva normal *estandarizada o tipificada*. Se la emplea para resolver cualquier problema de obtención de la probabilidad normal pues no importa la curva normal que se tenga, con el simple recurso de tipificarle la variable se la transforma en la de tablas, y con eso se simplifica la tarea de calcular la integral correspondiente. Se aprecia que la curva es simétrica respecto del eje que pasa por su centro (el origen), y tiene dos puntos de inflexión también simétricos situados a un desvío estándar del eje. El área total bajo la curva es uno, como se puede calcular con:

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2} = 1$$

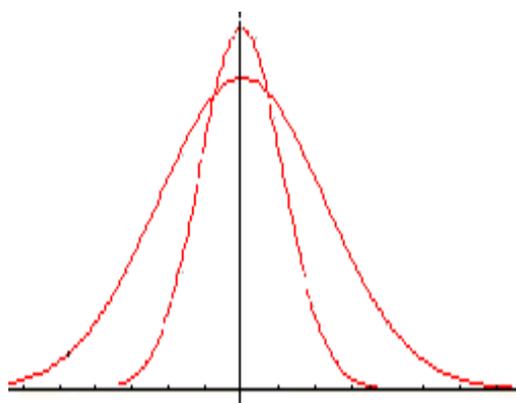
Gráfico 9.1 : Campana de Gauss tipificada.



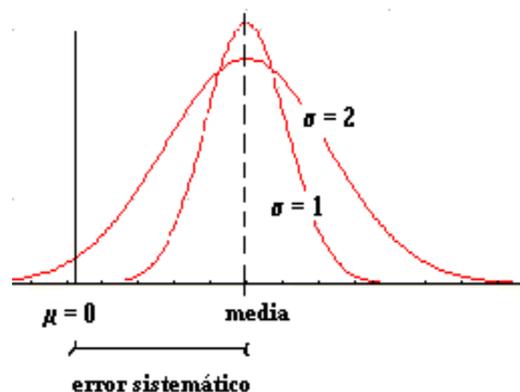
En el Gráfico 9.2 se presentan funciones normales con diferentes valores en sus parámetros. Se aprecia que a medida que se disminuye el valor σ , dejando fijo el valor de μ , la curva se angosta y se hace más alta (leptocúrtica), mientras que si aumenta el valor de su desvío estándar se achata, tiende a pasar de su forma acampanada a otra forma como de plato invertido (platocúrtica). Eso significa que si la dispersión de los valores disminuye, el error casual es menor, la precisión del sistema de medición es mayor, y la curva se afina. Viceversa, cuanto peor es la precisión de la técnica clínica empleada, más aplanada resultará la curva. En la parte superior del gráfico se muestran dos curvas con valor esperado nulo y desvío estándar 1 y 1,5. En la parte derecha del gráfico, se ilustra el caso de los errores sistemáticos. El efecto de la aparición de un error de esta naturaleza en el sistema de medición, hace que la curva gaussiana se corra a derecha o izquierda, *sin cambiar de forma*. En efecto, si aparece un valor positivo μ la curva de referencia $\sigma = 1$ se desplaza a la derecha. En cambio, si la precisión disminuye porque la dispersión se duplicó $\sigma = 2$ la curva se aplanan (cuando la dispersión se achica a la mitad $\sigma = 0,5$ la curva se afina). O sea, si cambia la dispersión de los valores, *cambia la forma de la curva*. Los errores sistemáticos la desplazan sin cambiarla, mientras que los errores casuales la transforman sin desplazarla. El efecto conjunto es el ilustrado más abajo.

Gráfico 9.2 : Efecto de los errores sistemáticos y casuales en la curva de errores de Gauss.

a) Efecto de los casuales



b) Efecto de los casuales y sistemáticos



9.5 Propiedades

Las propiedades siguientes se enuncian sin demostración, las que pueden hallarse en los libros de teoría como el de Cramer.

- 1) La curva normal es simétrica respecto de un eje vertical que pase por su centro.
- 2) El área total bajo la curva es igual a uno, por lo tanto a cada lado del eje de simetría es igual a 0,5 la mitad del área total. El área bajo la curva es igual a la probabilidad.
- 3) Cada curva normal se identifica en forma unívoca con los valores de sus 2 parámetros (μ y σ).
- 4) La función normal tipificada tiene parámetros $\mu = 0$ y $\sigma = 1$, es simétrica respecto al origen y la probabilidad a cada lado del eje de ordenadas vale 0,5 (ver Tabla en el Anexo).
- 5) Las funciones de Gauss tienen dos puntos de inflexión simétricos en $x = \mu \pm \sigma$
- 6) En toda función normal la media coincide con la mediana y con la moda.
- 7) Si x es una variable aleatoria de parámetros $N_x(\mu, \sigma)$ entonces la variable $y = ax + b$, donde a y b son constantes, también tendrá una distribución normal $N_y(a\mu + b, a\sigma)$.
- 8) Si x e y son dos variables aleatorias normales de parámetros $N_x(\mu_x, \sigma_x)$; $N_y(\mu_y, \sigma_y)$, entonces la variable suma de ambas $z = x + y$, también será normal de parámetros dados por la suma de sus valores esperados y de sus varianzas $N_z\{(\mu_x + \mu_y), (\sigma_x^2 + \sigma_y^2)^{1/2}\}$.
- 9) Si x_1 y x_2 son dos variables aleatorias normales independientes, tienen los mismos parámetros (μ, σ), entonces la variable aleatoria $w = (x_1 + x_2)/2$, promedio de ambas, tendrá una distribución normal $N_w(\mu, \sigma/\sqrt{2})$.
- 10) Generalizando, sean n variables aleatorias normales independientes, cada una de ellas proveniente de la misma población, es decir con los mismos parámetros (μ, σ), entonces la variable aleatoria promedio de las mismas también será una función normal con los parámetros:

Si $y = \bar{x} = \sum_{i=1}^{i=n} \frac{x_i}{n}$ entonces y tendrá una distribución normal: $N_y(\mu, \sigma/\sqrt{n})$

- 11) El área bajo la curva normal, correspondiente a los intervalos siguientes, es igual a la probabilidad de que un valor de la variable aleatoria caiga dentro de esos intervalos:

$P\{x \in (\mu \pm \sigma)\} = 0,6828 \Rightarrow$ la probabilidad que el valor caiga en $(\mu \pm \sigma)$ es del 68,3%

$P\{x \in (\mu \pm 2\sigma)\} = 0,9546 \Rightarrow$ la probabilidad que el valor caiga en $(\mu \pm 2\sigma)$ es del 95,5%

$P\{x \in (\mu \pm 3\sigma)\} = 0,9973 \Rightarrow$ la probabilidad que el valor caiga en $(\mu \pm 3\sigma)$ es del 99,7%

9.6 Cálculo de probabilidades con Gauss

A continuación, se muestran algunos ejemplos de aplicación del cálculo de probabilidad usando la función de Gauss con la Tablas 1 y 2 del Anexo de Tablas Estadísticas que corresponden a la Curva de Gauss Estandarizada (con valor esperado nulo y desvío estándar igual a uno).

Ejemplo 1) En una población gaussiana de parámetros (114 ; 23) se extrae un individuo al azar y se desea calcular las probabilidades de que: a) sea mayor de 100; b) esté comprendido entre 56 y 74.

Para resolver estos casos se deben calcular primero las tipificaciones de los valores dados. Luego hacer un pequeño esquema para entender el problema y por último usar la tabla de valores tipificados para calcular las probabilidades. Tal como se procedió más arriba. Pero ahora, se detallará un poco más:

Los valores tipificados son:

$$Z_1 = (100 - 114)/23 = - 0,6087$$
$$Z_2 = (56 - 114)/23 = - 2,5217$$
$$Z_3 = (74 - 114)/23 = - 1,7391$$

a) Para encontrar las probabilidades respectivas hay dos caminos: el primero es ir a la Tabla de Gauss tipificada e interpolar para calcular el valor pedido; el otro es obtener el valor directamente con una computadora y un programa estadístico como el Excel, el Statistica, y otros.

De tablas para $z = 0,60$ es $p' = 0,2257$ y para $z = 0,61$ es $p'' = 0,2291$, el valor buscado está entre ambos. Para interpolar se razona que, si para una diferencia de $0,01 = 0,61 - 0,60$ corresponde una diferencia de probabilidades $P = 0,2291 - 0,2257 = 0,0034$, para la diferencia buscada $(0,61 - 0,6087) = 0,0013$ corresponderá una probabilidad $0,000442 = (0,0034)(0,0013)/0,01$.

Entonces será:

$$P(Z_1) = P(0,6087) = P(0,61) - 0,000442 = 0,2291 - 0,000442 = 0,228658$$

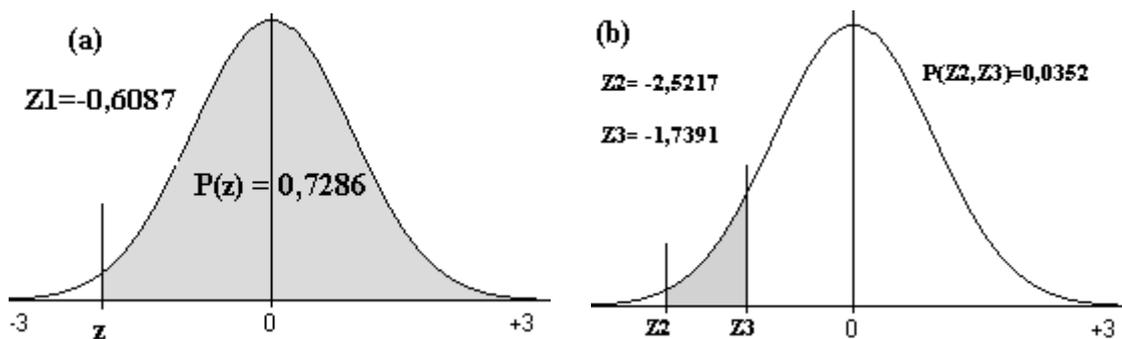
Análogamente, si se hubiese calculado la otra diferencia $0,0087$ el valor correspondiente, de probabilidad es $0,002958$, que debería sumarse al valor p' y el resultado sería el mismo.

Con el programa de Microsoft, Excel 2000, se puede calcular directamente el valor de probabilidad usando las funciones estadísticas que tiene incorporadas. En este caso: Distr. Norm. Estand. Que da un valor $P(Z_1) = P(0,6087) = 0,2286384$ el cual coincide en las cinco primeras cifras con la interpolación de tablas.

Ahora bien, $P(Z_1) = P(-Z_1) = 0,2286$ por la simetría de la función de Gauss. Entonces, el valor pedido en el problema resultará:

$$P(x > 100) = P(Z > -Z_1) = P(-0,6087 < Z < 0) + P(Z > 0) = 0,2286 + 0,5 = 0,7286$$

$$\text{b) } P(56 < x < 74) = P(Z_2 < Z < Z_3) = P(-2,5217 < Z < -1,7391) = 0,49416 - 0,45899$$
$$P(56 < x < 74) = 0,0352$$



Ejemplo 2) Se determinó la Lipemia con un método colorimétrico de un grupo de 200 individuos que padecen hipotiroidismo, obteniéndose un promedio de 13 g/l, con un desvío estándar de 0,1 g/l. Se pide calcular la probabilidad de encontrar un paciente elegido al azar cuyo valor:
 a) esté entre 13 y 13,2 g/l; b) sea mayor de 13,25 g/l; c) esté comprendido entre 13,1 y 13,2 g/l

Método 1) Se supone que los valores poblacionales son $\mu = 13$ g/l y $\sigma = 0,1$ g/l

a) Esté entre 13 y 13,2 g/l

Tipificando la variable resulta $z = \frac{13,2 - 13,0}{0,1} = 2$

Y $z = 0$ para el valor 13, entonces los valores pedidos son:

$$P(13,0 < x < 13,2) = P(0 < z < 2)$$

Buscando en la Tabla 1 se halla: $P(z = 2) = 0,4772$

La probabilidad es casi del 48%.



b) Sea mayor de 13,25 g/l

Tipificando la variable resulta $z = \frac{13,25 - 13,0}{0,1} = 2,5$

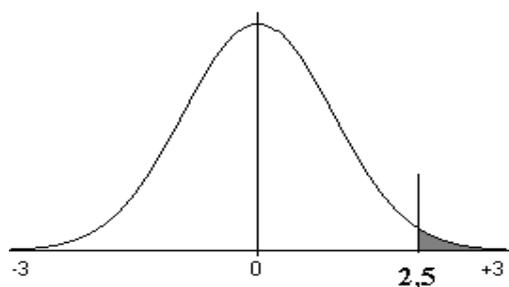
Lo pedido es $P(z > 2,5) = P(x > 13,25)$

Buscando en la Tabla 1 se halla: $P(z = 2,5) = 0,4938$

La mitad del área bajo la curva es 0,5; lo buscado es:

$$P(x > 13,25) = P(z > 2,5) = 0,5 - 0,4938 = 0,0062$$

La probabilidad es del 0,62%.



c) Esté comprendido entre 13,1 y 13,2 g/l

Tipificando resulta $z_1 = \frac{13,1 - 13,0}{0,1} = 1$

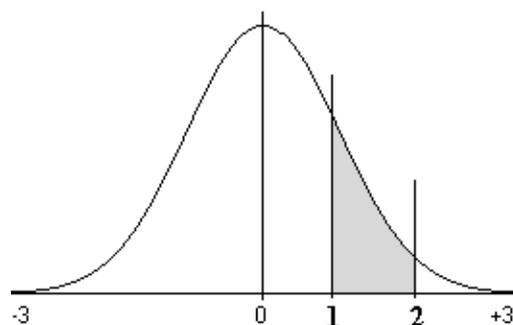
y $z_2 = \frac{13,2 - 13,0}{0,1} = 2$

$$P(13,1 < x < 13,2) = P(1 < z < 2) = P(x < 2) - P(x < 1)$$

O sea,

$$P(1 < z < 2) = 0,4772 - 0,3413 = 0,1359$$

La probabilidad es de 13,6%



Método 2) No se conocen los parámetros poblacionales y se los estima con los valores muestrales. O sea, se toman $\mu \approx \bar{x} = 13,0 \text{ g/l}$ y $\sigma \approx DS = 0,1 \text{ g/l}$

Averiguar: ¿cuántos pacientes tendrán un valor comprendido entre 13,1 y 13,2 g/l?

Tomando una aproximación de 0,1 g/l en el sistema de medición, los valores medidos entre 13,1 y 13,2 pueden realmente tener cualquier valor comprendido entre los límites reales, o sea, entre 13,05 y 13,25. Por lo tanto será

$$z_1 = \frac{13,05 - 13,0}{0,1} = 0,5$$

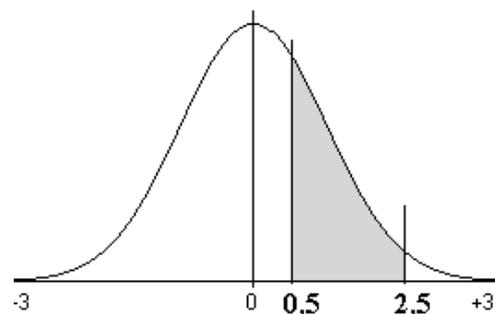
y por el cálculo anterior $z_2 = 2,5$

Luego la probabilidad es:

$$P(13,05 < x < 13,25) = P(0,5 < z < 2,5) = 0,4938 - 0,1915 = 0,3023$$

Y el valor esperado será

$$E = N \cdot P(13,05 < x < 13,25) = 200 \cdot 0,3023 \approx 60 \text{ pacientes}$$



Nota: Cuando se usan los valores de la muestra para estimar los valores desconocidos de la población, entonces se debe incrementar (o disminuir) en la mitad de la menor unidad de la escala de medición, a los valores buscados.

Ejemplo 3) Los límites de tolerancia del componente Sulfato de Neomicina en un antibiótico fueron fijados en (40 ; 60) mg. Se sabe que el valor promedio histórico de proceso de fabricación es 50 mg con un desvío de 4 mg. Calcular el número esperado de defectuosos en un lote de 9000.

Los valores de tolerancia tipificados son:

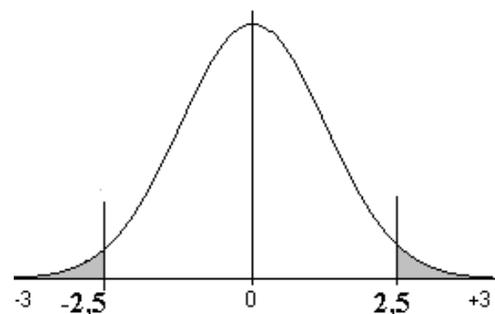
$$Z_1 = (40 - 50)/4 = -2,5 \quad \text{y} \quad Z_2 = (60 - 50)/4 = 2,5$$

La zona de rechazo es por fuera de estos límites, entonces la probabilidad de rechazo es:

$$P(\text{rechazo}) = 1 - P(-2,5 < z < 2,5) = 1 - 2 P(0 < z < 2,5) = 1 - 2(0,4938) = 1 - 0,9876 = 0,0124$$

El número esperado de defectuosos será:

$$D = N \cdot P(\text{rechazo}) = 9000 (0,0124) = 111,6$$



O sea, se esperan rechazar 112 comprimidos de los 9000.

9.7 Aproximaciones con la función de Gauss

Se puede usar la función de Gauss para ajustarla a una serie de datos agrupados en una tabla de frecuencia de una magnitud clínica de tipo asumida como gaussiana. El supuesto básico es que la muestra fue extraída de una población normal. Para ilustrar el proceso se tomarán los datos de peso de 195 estudiantes varones, que se muestran en el Cuadro 9.4. En la primera columna figuran los límites reales de clase, y en la segunda las frecuencias respectivas. Con esos datos se calcula el promedio de 69,47 kg con un desvío típico de 2,82 kg. Para determinar la frecuen-

cia esperada teóricamente, asumiendo una distribución normal de la variable, se debe determinar primero la probabilidad de cada clase y luego multiplicarla por el número total de casos (N=195). Debe tenerse cuidado con el primero y el último de los intervalos, pues si no se tiene en cuenta las colas de la distribución que tienden asintóticamente a infinito, la suma de la probabilidad total no será la unidad. Para esquivar ese problema se toma toda el área a la izquierda del límite real superior de la primera clase ($-\infty, 63,5$), para el primer caso. Luego, para el segundo intervalo se calcula la probabilidad entre ($-\infty, 65,5$) y se le resta el valor anterior, con lo que se determina la probabilidad del segundo intervalo, y así sucesivamente hasta el último, donde se debe incluir el área a la derecha, obteniéndola por diferencia con la unidad.

Cuadro 9.4: Ajuste a una distribución de pesos de 195 varones.

Intervalos	Frecuencia Observada	Valor de z para LRS	Área acumulada en ($-\infty, LRS$)	Probabilidad en cada clase	Frecuencia Esperada
hasta 63,5	3	-2,12	0,0170	0,0170	3,3
63,5 - 65,5	14	-1,41	0,0793	0,0623	12,2
65,5 - 67,5	30	-0,70	0,2420	0,1627	31,7
67,5 - 69,5	48	+0,01	0,5040	0,2620	51,1
69,5 - 71,5	48	+0,072	0,7642	0,2602	50,7
71,5 - 73,5	39	+1,43	0,9236	0,1594	31,1
73,5 - 75,5	11	+2,14	0,9838	0,0602	11,7
75,5 y más	2		1,0000	0,0162	3,2

La aproximación de una distribución de frecuencias reales a una teórica puede hacerse, siempre y cuando se asuma que las muestras provienen de una población normal.

Los procesos del tipo Bernoulli, representados mayormente por la probabilidad binomial, también pueden aproximarse con una función normal de parámetros: $\mu = np$ y $\sigma = \sqrt{npq}$. Siempre que las muestras sean lo suficientemente grandes, basta que $np > 5$ y $nq > 5$ para considerar buena a la aproximación. Lo curioso es que una variable de tipo discreta se puede aproximar con una de tipo continua. Por ello, se debe efectuar una corrección por *continuidad* que consiste en sumarle o restarle 0,5 a los valores enteros en los intervalos buscados. Esto es, los valores 0 y n son los máximos y mínimos de una binomial al aproximar las correspondientes probabilidades se tomará: toda el área a la izquierda de 0,5 para el valor $x = 0$, y toda el área a la derecha de $(n-0,5)$ para aproximar el valor de $x = n$. Muy similar a lo visto en el punto anterior como *Método 2*. En el Cuadro 9.5 se toma un caso extremo para ver cómo resulta la peor aproximación en el límite: 10 lanzamientos de una moneda ($np = nq = 5$).

Cuadro 9.5: Aproximación entre la normal y la binomial.

Caso 1: Se lanza una moneda 10 veces, obtener las probabilidades exactas del número de caras con la binomial: $P_{bi} = P(x < 11 / n = 10; p = 1/2)$, para comparar con las aproximadas N ($\mu = 5 ; \sigma = 1,58$)

x	0	1	2	3	4	5	6	7	8	9	10
Binomial:	0,001	0,010	0,044	0,117	0,205	0,246	0,205	0,117	0,044	0,010	0,001
Normal :	0,002	0,011	0,044	0,114	0,203	0,251	0,203	0,114	0,044	0,011	0,002

Caso 2: Se deben adivinar los símbolos que se sacan al hacer una extracción al azar en las cartas Zener: como hay 5 símbolos, $p = 1/5 = 0,2$. Se efectúan 25 extracciones, calcular las probabilidades de tener hasta 9 éxitos, en forma exacta y aproximando con Gauss. $N(\mu = 5; \sigma = 2)$

x	0	1	2	3	4	5	6	7	8	9
Binomial:	0,004	0,024	0,071	0,136	0,187	0,196	0,163	0,111	0,062	0,029
Normal :	0,009	0,027	0,065	0,121	0,176	0,199	0,176	0,121	0,065	0,027

Cuando sea $np > 5$ y $nq > 5$, se puede aproximar la binomial con una normal.

En un proceso del tipo Poisson, como se puede aproximarla con la binomial es de esperar que se pueda usar la normal para aproximar una Poisson. En efecto, si el valor esperado es mucho mayor que 1, la aproximación es aceptable. En la binomial sería $\mu = np$, entonces si es $n \gg 1$ y $p \ll 1$, su producto resultará $np > 1$, así se tendrá una Poisson con $\sigma = \mu = np$.

Cuadro 9.6: Aproximaciones entre la Binomial, Poisson y la de Gauss.

Caso 1: La probabilidad de éxito en un proceso del tipo Bernoulli es muy pequeña ($p = 0,01$). Si se repite 100 veces el experimento, calcular la probabilidad de obtener 0,1,2,3, 4 y 5 éxitos. Donde las probabilidades se sacan con : $P_{PO}(x / \sigma^2 = \mu = 1)$, $P_{bi} = P(x / n = 100; p = 0,01)$, $N(\mu = 1; \sigma = 1)$

x	0	1	2	3	4	5
Binomial:	0,366	0,370	0,185	0,061	0,015	0,003
Poisson:	0,368	0,368	0,184	0,061	0,015	0,003
Normal :	0,309	0,383	0,242	0,061	0,005	0,000

La aproximación entre la Poisson y la normal podría ser mejor si np hubiese sido mucho mayor que 1.

Caso 2: La probabilidad de que un individuo sufra una reacción alérgica por una inyección es de 0,01. Calcular la probabilidad de que entre 8 y 14 pacientes tengan una reacción, sobre un total de 1000. Para resolver, se usa una Poisson de parámetro $\mu = np = 10$

x	8	9	10	11	12	13	14	Total
Poisson:	0,1126	0,1251	0,1251	0,1137	0,0948	0,0729	0,0521	0,6963

Luego la probabilidad de acuerdo con Poisson es de 69,63%

Usando la función normal será $z_1 = (7,5 - 10) / \sqrt{10} = -0,79$ y $z_2 = (14,5 - 10) / \sqrt{10} = 1,42$

$P(8 < x < 14) = 0,2852 + 0,4222 = 0,7072$. O sea un 70,7% (la diferencia es del 1,5%).

La función de Gauss descubierta alrededor de 1810, no solo es la más tradicional en Estadística sino la de uso más difundido por las posibilidades de simplificar los cálculos a través de las aproximaciones. Como se verá más adelante, casi todos los modelos estadísticos tienen una probabilidad teórica propia que se aproxima a la función de Gauss en condiciones bastante generales y cuando las muestras son lo *suficientemente grandes*. Esto significa para el estudiante, que esta función la usará en forma reiterada y por lo tanto debe asegurarse de entender muy bien su empleo, propiedades y aplicaciones.

9.8. Correcciones por continuidad

Como se verá más adelante la función normal se usa como aproximación asintótica en muchos modelos de estadística, tanto la paramétrica como la no-paramétrica. Y como se vio más arriba se usa como aproximación a las distribuciones binomiales y de Poisson. En resumen, cuando el valor del tamaño muestral n crece indefinidamente, la forma de la función binomial se parece cada vez más a la forma de la normal. Y esto ocurre para cualquier valor de la probabilidad p . Cuando los valores de la probabilidad son más cercanos a $p = 0,5$ tiende más rápido, que cuando los valores de p están cercanos a 0 o 1, pues todas las distribuciones binomiales con $p = 0,5$ tienen la gran ventaja de ser simétricas. Así, si n es lo suficientemente grande se puede imaginar la binomial como muy parecida a la normal.

Cuando la variable que se está usando es una frecuencia r (0, 1, 2, ..., n) la magnitud se considera discreta, pero las tablas de la función Normal son para magnitudes de tipo continuo, entonces existe una cierta corrección que se le puede hacer, para lograr un mejor ajuste entre ambas distribuciones. Esta se llama: *corrección por continuidad*. Se basa en la idea que el valor de la probabilidad binomial para un valor r en realidad es aproximadamente la probabilidad del intervalo entre $(r - 1/2)$ y $(r + 1/2)$ porque la integral de la función normal para un punto es nula, y siempre se debe tomar la integral de algún intervalo para tener un resultado diferente de cero. Por consiguiente, de acuerdo a la hipótesis nula efectuada la corrección por continuidad se realiza así:

1) *Corrección de Yates*: Al valor tipificado $z = (r - \mu) / \sigma$ se lo corrige con

$$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma \quad \text{donde } \sigma = \sqrt{npq}$$

Caso 1: Si se estima que la probabilidad de nacer varón es $p = 0,5$. Se pide calcular la probabilidad de que en 40 nacimientos en una maternidad nazcan por lo menos 14 varones:

Ho : $r \leq n \cdot p$ donde $\mu = n p = 20$ y $\sigma = \sqrt{npq} = 3,162$ entonces

$$z = (r - \mu) / \sigma = (14 - 20) / 3,162 = - 1,8975 \quad \text{y entonces la probabilidad es } 0,02888$$

Pero si se hace la corrección por continuidad el resultado sería:

$$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [6 - 0,5] / 3,162 = 1,739 \quad \text{y entonces la probabilidad es } 0,0410$$

Comparando con la probabilidad exacta calculada por la fórmula binomial $P_{Bi} = 0,40345234$. De donde se puede ver que sin la corrección de Yates la aproximación no es muy aceptable. Si en cambio se pide calcular la probabilidad de que nazcan 26 o más varones es:

Ho : $r \leq n \cdot p$ entonces y resulta igual a la anterior pues $P_{Bi}(r \leq 14) = P_{Bi}(r \geq 26)$

$$z = (r - \mu) / \sigma = (26 - 20) / 3,162 = 1,8975 \quad \text{y así la probabilidad es la misma } 0,0289.$$

Aplicando nuevamente la corrección resulta $Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = 1,739$ y entonces la probabilidad es 0,0410 muy parecida a la exacta $P_{Bi} = 0,0403$.

2) *Correcciones de Molenaar*: Al valor tipificado $z = (r - \mu) / \sigma$ se lo corrige con

$$Z_{\text{corr.}} = [(4r + 3) \cdot (1-p)]^{1/2} - [(4n - 4r - 1)p]^{1/2}$$

Para el caso anterior donde $r = 14$ es $Z_{\text{corr.}} = [(56 + 3) \cdot (0,5)]^{1/2} - [(160 - 56 - 1)p]^{1/2} = -1,74496$

Este valor da una probabilidad de 0,405 que es mucho más cercana al valor exacto que la efectuada con la corrección de Yates. Mientras que si el valor de p es muy cercano a cero o uno, se debería usar una fórmula corregida diferente dada por:

$$Z_c = [0,5 \cdot (8r + 5) \cdot (1-p)]^{1/2} - [0,5 \cdot (8n - 8r - 3)p]^{1/2}$$

Por ejemplo, sea una probabilidad de éxito de 0,05 en un proceso binomial de 50 pruebas y hay que calcular la probabilidad de tener por lo menos 2 éxitos. Entonces, la probabilidad exacta da un resultado de $P_{Bi} = 0,54053$, mientras que la aproximación normal da:

Yates: $Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [|2 - 2,5| - 1/2] / 2,121 = 0$; lo que da $P \approx 0,5$

Molenaar: $Z_c = [0,5 \cdot (8r + 5) \cdot (1-p)]^{1/2} - [0,5 \cdot (8n - 8r - 3)p]^{1/2}$

$Z_c = [0,5 \cdot (16 + 5) \cdot (1-0,05)]^{1/2} - [0,5 \cdot (400 - 16 - 3) \cdot 0,05]^{1/2} = 0,06397$; lo que da $P \approx 0,5255$

Y con la otra es $Z_{\text{corr.}} = [(4r + 3) \cdot (1-p)]^{1/2} - [(4n - 4r - 1)p]^{1/2} = 0,1423$ O sea $P \approx 0,5566$

Este ejemplo muestra que la corrección de Molenaar es mejor que la de Yates en ambos casos. Y que la segunda ajusta mejor, cuando las probabilidades son muy pequeñas o muy grandes.

Lo mismo puede hacerse con la aproximación normal a la probabilidad de Poisson. Suponiendo que la variable X sea de tipo Poisson, el valor tipificado para usar la aproximación normal es $z = (X - \mu) / \sigma = (X - \mu) / \mu^{1/2}$. Y su valor corregido será:

1) *Corrección de Yates*: $z = [|X - \mu| - 1/2] / \mu^{1/2}$

Por ejemplo, si el valor medio de un proceso de Poisson es $\mu = 5$ y se desea calcular la probabilidad de a) no obtener ningún éxito, b) obtener a lo sumo 2 y c) por lo menos 8 éxitos. Entonces, las probabilidades exactas son: $P_{PO}(x = 0) = 0,0067$; $P_{PO}(x \leq 2) = 0,1246$; $P_{PO}(x \geq 8) = 0,1334$ y efectuando las aproximaciones con la función normal, los valores resultan ser:

a) $z = [|0 - 5| - 1/2] / 5^{1/2} = 2,013$ lo que equivale a una probabilidad de 0,0221

b) $z = [|2 - 5| - 1/2] / 5^{1/2} = 1,118$ y corresponde a una probabilidad de 0,1318

c) $z = [|8 - 5| - 1/2] / 5^{1/2} = 1,118$ igual que la anterior

2) *Correcciones de Molenaar*: $z = 2 [X + 0,75]^{1/2} - 2 [\mu]^{1/2}$

a) $z = 2 [0 + 0,75]^{1/2} - 2 [5]^{1/2} = - 2,740$ o sea $P \approx 0,0031$

b) $z = 2 [2 + 0,75]^{1/2} - 2 [5]^{1/2} = - 1,156$ o sea $P \approx 0,1239$

c) La $P_{PO}(x \geq 8) = 1 - P_{PO}(x \leq 7)$ Con la aproximación es $z = 1,0956$ o sea $P(7) \approx 0,8633$

Luego la $P(x \geq 8) \approx (1 - 0,8633) = 0,1367$. Nuevamente esta corrección es la mejor de ambas.

9.9 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | | |
|---|-------|---|--|
| 1) Explicar lo que se entiende por "normal" | | | |
| 2) Las estadísticas gaussianas sirven para determinar los límites normales. | V | F | |
| 3) Se estila adoptar los límites de referencia como "límites normales". | V | F | |
| 4) En análisis clínicos conviene más usar los modelos no paramétricos. | V | F | |
| 5) Los criterios de normalidad son tres: | | | |
| 6) Los valores de referencia de una magnitud clínica dependen de la población referencial. | V | F | |
| 7) Para construir los intervalos de referencia se deben explicar los 5 puntos siguientes: | | | |
| 8) La función de Gauss es simétrica y el área total bajo la curva es uno. | V | F | |
| 9) La curva de Gauss tiene 2 puntos de inflexión. | V | F | |
| 10) El efecto de los errores casuales es desplazar la curva de Gauss sin cambiar su forma. | V | F | |
| 11) El efecto de los errores sistemáticos es modificar el ancho de la curva de Gauss. | V | F | |
| 12) La curva de Gauss queda unívocamente identificada con 2 parámetros que son: | | | |
| 13) En la curva Normal la media coincide con la mediana y con la moda. | V | F | |
| 14) La suma de 2 funciones normales es otra función normal. | V | F | |
| 15) La probabilidad de que un valor caiga en un intervalo dado por ± 2 desvíos es: | | | |
| 16) Si las muestras provienen de una población normal se puede aproximar con la empírica. | V | F | |
| 17) Si n es lo suficientemente grande se puede aproximar la binomial con la normal. | V | F | |
| 18) Si $np > 5$ y $nq > 5$ se puede hacer la aproximación normal a la binomial. | V | F | |
| 19) Si $n \gg 1$ y $p \ll 1$ tal que $np = cte$. Se puede aproximar la normal a la Poisson. | V | F | |
| 20) La suma de n variables aleatorias independientes y normales es otra función normal. | V | F | |

2) Calcular las probabilidades o área bajo la curva de Gauss de cada uno de los siguientes casos:

- a) Z entre 0 y 1,6; b) Z entre $-2,1$ y 0; c) Z entre $-1,96$ y $1,96$; d) Z entre $-2,04$ y $1,56$;
e) a la izquierda de $Z = -1,73$; f) a la derecha de $Z = -0,56$; g) a la derecha de $Z = 1,89$.

3) Determinar los valores de Z conociendo los valores de probabilidad gaussiana de estos casos:

- a) entre 0 y Z es 0,2734; b) a la izquierda de Z es 0,9463; c) A la derecha de Z es 0,7257;
d) entre $-1,5$ y Z es 0,4332; e) entre $1,27$ y Z es 0,2146.

4) Suponiendo que el peso de los estudiantes de la Facultad se distribuya normalmente con media de 63 Kg y desvío de 7 Kg. Hallar la probabilidad de que uno de ellos escogido al azar pese:

- a) más de 75 Kg; b) menos de 55 Kg; c) entre 55 y 74 Kg; d) exactamente 70 Kg

5) Suponiendo que la altura de los mismos estudiantes se distribuya normalmente con una media de 1,68 m y un desvío de 0,17 m y que sean 4.500 en total. Hallar cuántos de ellos tienen una estatura: a) de más de 1,8 m; b) de menos de 1,6 m; c) entre 1,65 y 1,75 m; d) de 1,72 m

6) Haciendo los supuestos de normalidad correspondientes, hallar las frecuencias esperadas en todos los problemas de los capítulos anteriores donde se hallaron con Binomial, Poisson, etc.

7) Aplicar a los casos anteriores las correcciones de Yates y de Molenaar para decidir cual es la mejor de ambas y en que casos conviene usar una u otra.

8) ¿Qué significan los dos puntos de inflexión?