

CAPÍTULO V.

La distribución normal

En este capítulo vas a conocer la reina de las distribuciones de probabilidad, a la cual adaptamos casi todos los fenómenos cotidianos susceptibles de estudiar en probabilidad.

Incluso la binomial, cuando “n” es grande, y siempre que $npq > 9$, podemos usar la normal para evitar esos cálculos tan desagradables (hasta para una calculadora común) y que antiguamente eran terriblemente costosos.

Los cálculos de la normal son mucho peores (implican resolver integrales) por lo que te vas a tener que acostumbrar a usar la tabla que tienes detrás de esta introducción (tenla a mano siempre). Sin esta tabla sería una tortura hacer cada ejercicio.

No son ejercicios tan fáciles como los de binomial, pero tienen la ventaja de que casi siempre siguen el mismo patrón: en un apartado tendrás que hallar una probabilidad, y en otro apartado te darán una probabilidad para que halles un parámetro de la normal (en este orden o al revés). Cuando veas unos pocos ejercicios te resultará fácil.

1.4. Distribuciones de probabilidad continuas

La distribución binomial se utiliza para describir fenómenos aleatorios discretos: número de caras, número de curaciones, número de bombillas de buena calidad... No tendría sentido decir que se habían obtenido 0.3 cincos al tirar unos dados. Ya sabes que otras variables aleatorias pueden ser continuas, como la estatura de una persona, la medida de una pieza de fabricación... Vamos a estudiar una distribución de probabilidad continua adecuada para estos casos. Hay más distribuciones de probabilidad discretas y continuas, pero la distribución binomial para variables discretas y la distribución normal para variables continuas son las más importantes, las más utilizadas.

Ya hemos analizado las propiedades de las funciones de cuantía de las variables discretas. Las **funciones de densidad de las variables continuas** $f(x)$ deben verificar también una serie de propiedades que estudiarás con más rigor si estudias estadística en la universidad.

Propiedades de la función de densidad $f(x)$:

- 1) $f(x) \geq 0$. Es natural, pues estamos midiendo probabilidades.
- 2) El área total bajo la curva debe medir 1. Ya que la probabilidad del suceso seguro es 1.

Propiedades de la función de distribución $F(x)$:

- 1) $0 \leq F(x) \leq 1$
- 2) Es una función creciente en todo su dominio de definición
- 3) $\lim_{x \rightarrow \infty} F(x_{\text{máximo}}) = 1$.

Algo que puede sorprenderte es que la probabilidad de que una persona mida exactamente 1.8 metros es 0. ¿Por qué? La razón es que se debe calcular el área de un rectángulo de base 0, que es 0. Es una situación nueva pues hasta ahora parecía que si la probabilidad era nula el suceso era imposible y no es así, lo que se verifica es que si el suceso es imposible entonces la probabilidad es nula.

Tendríamos que calcular esa área en un intervalo, por ejemplo, entre 1.79 y 1.81. Ya sabes que toda medida lleva implícita una cierta imprecisión. Si decimos que Juan mide 1.8 metros como habrá una imprecisión de por ejemplo ± 0.01 , estaremos en un cierto intervalo. No estamos diciendo que no sea posible que Juan mida exactamente 1.8, sino que su probabilidad es nula:

$$P(x = 1.8) = \int_{1.8}^{1.8} f(x) dx = 0.$$

$$P(1.79 < x < 1.81) = \int_{1.79}^{1.81} f(x) dx.$$

Como consecuencia de lo anterior se tiene que:

$$P(c \leq x \leq d) = P(c < x \leq d) = P(c \leq x < d) = P(c < x < d) = \int_c^d f(x) dx.$$

En las distribuciones de variable aleatoria continua, las frecuencias relativas se corresponden cuando se consideran probabilidades con la función de densidad. Para calcular una probabilidad debemos calcular el área bajo la curva función de densidad. Las frecuencias relativas acumuladas se corresponden con lo que denominamos función de distribución de probabilidad.

La función: $F(t) = P(a < x < t) = \int_a^t f(x) dx$ es la función de distribución.

Verifica, como sabes, que $F'(x) = f(x)$, es decir, la función de distribución es la función primitiva de la función de densidad. Conocida una podemos calcular la otra.

Parámetros estadísticos en una distribución continua

Ya sabes que una integral la podemos considerar como una suma. Por eso los parámetros estadísticos en una distribución de probabilidad continua se definen igual que en una distribución discreta cambiando la suma por una integral.

Si el dominio de definición es el intervalo $[a, b]$ entonces la media y la varianza se definen como:

$$\mu = E(x) = \int_a^b x \cdot f(x) \cdot dx = \int_a^b (x - \mu)^2 \cdot f(x) \cdot dx = 0$$

donde $f(x)$ es la función de densidad.

La media y la varianza siguen verificando las mismas propiedades que en el caso discreto.

Actividades resueltas

✚ Definimos la función de densidad $f(x) = (-1/36)(x^2 - 9)$ en el intervalo $[-3, 3]$. Prueba que verifica las condiciones de una función de densidad y calcula la media y la varianza.

1) La función es una parábola. En el intervalo dado $[-3, 3]$ toma valores positivos luego verifica la primera propiedad: $f(x) \geq 0$ en todo el dominio de definición $[-3, 3]$.

2) Para ser función de densidad también debe verificar que $\int_a^b f(x) \cdot dx = 1 = \int_{-3}^3 \left(\frac{-1}{36}(x^2 - 9)\right) \cdot dx$.

Calculamos la integral:

$$\begin{aligned} \int_{-3}^3 \left(\frac{-1}{36}(x^2 - 9)\right) \cdot dx &= \left(\frac{-1}{36}\right) \cdot \int_{-3}^3 (x^2 - 9) \cdot dx = \left(\frac{-1}{36}\right) \cdot \left[\frac{x^3}{3} - 9x\right]_{-3}^3 = \\ &= \left(\frac{-1}{36}\right) \cdot \left[\left(\frac{3^3}{3} - 9 \cdot 3\right) - \left(\frac{(-3)^3}{3} - 9 \cdot (-3)\right)\right] = \left(\frac{-1}{36}\right) \cdot \left[\frac{3^3}{3} + \frac{3^3}{3} + 9(-3 - 3)\right] = \left(\frac{-1}{36}\right) \cdot (18 - 54) = 1 \end{aligned}$$

Cálculo de la media:

$$\begin{aligned} \mu &= \int_a^b x \cdot f(x) \cdot dx = \int_{-3}^3 x \cdot \left(\frac{-1}{36}(x^2 - 9)\right) \cdot dx = \frac{-1}{36} \int_{-3}^3 x \cdot (x^2 - 9) \cdot dx = \frac{-1}{36} \int_{-3}^3 (x^3 - 9x) \cdot dx = \frac{-1}{36} \left[\frac{x^4}{4} - 9 \frac{x^2}{2}\right]_{-3}^3 = \\ &= \frac{-1}{36} \left[\left(\frac{3^4}{4} - 9 \frac{3^2}{2}\right) - \left(\frac{(-3)^4}{4} - 9 \frac{(-3)^2}{2}\right)\right] = 0 \end{aligned}$$

Cálculo de la varianza:

$$\begin{aligned} \sigma^2 &= \int_a^b (x - \mu)^2 \cdot f(x) \cdot dx = \int_{-3}^3 x^2 \cdot \left(\frac{-1}{36}(x^2 - 9)\right) \cdot dx = \frac{-1}{36} \int_{-3}^3 (x^4 - 9x^2) \cdot dx = \frac{-1}{36} \left[\frac{x^5}{5} - 9 \frac{x^3}{3}\right]_{-3}^3 = \\ &= \frac{-1}{36} \left[\left(\frac{3^5}{5} - 9 \frac{3^3}{3}\right) - \left(\frac{(-3)^5}{5} - 9 \frac{(-3)^3}{3}\right)\right] = \frac{-1}{36} \left[2 \frac{3^5}{5} - 18 \frac{3^3}{3}\right] = 1.8 \end{aligned}$$

La media vale 0 y la varianza 1.8. La desviación típica vale aproximadamente 1.34.

Actividades propuestas

18. Calcula A para que $f(x) = A(x^2 - 16)$ sea una función de densidad. Determina el dominio. Calcula la media y la varianza.

1.5. Distribución normal

La distribución normal es la distribución más importante tanto en lo que se refiere a la teoría estadística (debido a sus múltiples aplicaciones en inferencia) como en lo que se refiere a sus aplicaciones prácticas. Esta distribución fue propuesta independientemente por *Pierre Simon de Laplace* y *Carl Friedrich Gauss* a finales del siglo XVIII y principios del XIX. Por este motivo, también se la conoce como *distribución de Gauss*. En algunas ocasiones se refiere a ella como *campana de Gauss*, debido a la forma de campana de su función de densidad. Aunque se dice (en broma) que los físicos creen que fue descubierta por un matemático y que los matemáticos opinan que la descubrió un físico.

La expresión de su función de densidad y de su función de distribución es complicada:

$$N(\mu, \sigma): \varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad \Phi(x) = \int_{-\infty}^x \varphi(t) \cdot dt = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Donde μ es la media y σ la desviación típica. Para denotar que la variable aleatoria X sigue una distribución normal de parámetros μ y σ^2 se escribe $N(\mu, \sigma)$.

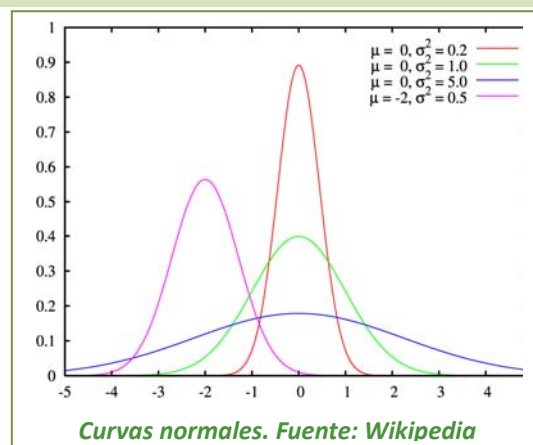
¡No te asustes! ¡No vamos a usar integrales! Son expresiones demasiado complicadas, y además, la integral que aparece no es posible resolverla. Y entonces, ¿qué hacemos? Por ejemplo se podría tabular $N(\mu, \sigma)$, pero serían necesarias infinitas tablas, una para cada uno de los posibles valores de μ y de σ .

Utilizando las propiedades de la esperanza matemática y de desviación típica podemos comprobar que basta con tabular una de ellas, la normal de media 0 y desviación típica 1, $N(0, 1)$, que vamos a denominar **distribución normal estándar**. Por tanto, como la función de distribución no puede calcularse analíticamente, hace que los cálculos de probabilidades en la distribución normal se tengan que realizar utilizando tablas que encontrarás más adelante.

Dada una variable aleatoria x , de media μ y desviación típica σ se llama **variable aleatoria tipificada** a la variable z , obtenida por $z = \frac{x-\mu}{\sigma}$, con lo que se obtiene una variable aleatoria de media 0 y desviación típica 1.

Observaciones:

- 1) La transformación, tipificación, supone una traslación, que cambia el origen de μ a 0, y una contracción o dilatación.
- 2) Se conservan las áreas bajo ambas curvas, una vez que usemos las variables tipificadas.
- 3) La variable aleatoria tipificada es adimensional, pues se obtiene dividiendo magnitudes de la misma dimensión, lo que permite poder comparar variables aleatorias diferentes, como estaturas de una población, y pesos de recién nacidos.
- 4) En la figura del margen puedes observar varias curvas normales, la dibujada en verde es la tipificada. Observa que todas las curvas normales son simétricas, de eje de simetría $x = \mu$ (o $x = 0$ en el caso de $N(0, 1)$). Tienen la media, la moda y la mediana iguales. En los puntos de abscisa $x = \mu - \sigma$ y $x = \mu + \sigma$ tienen un punto de inflexión. Son crecientes hasta $x = \mu$, en ese punto se alcanza un máximo, y decrecientes de $x = \mu$ en adelante.
- 5) La expresión de la función de densidad tipificada es $N(0, 1) = \varphi(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$



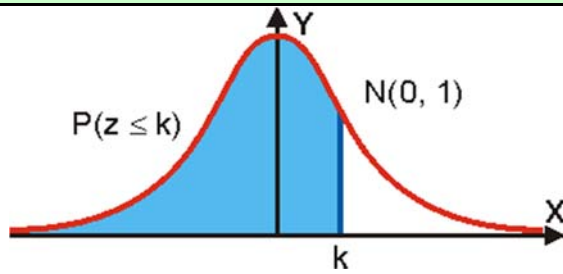
ÁREAS BAJO LA DISTRIBUCIÓN DE PROBABILIDAD NORMAL ESTÁNDAR, $N(0, 1)$ 

Tabla de la UAM: Universidad Autónoma de Madrid

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Vamos ahora a observar con cuidado la tabla para aprender a calcular, con ella, probabilidades.

No están todos los valores. Como el área total bajo la curva es 1, y la curva es simétrica $\phi(-z) = 1 - \phi(z)$.

Actividades resueltas

✚ Utiliza la tabla para calcular las probabilidades: a) $P(z \leq 1)$; b) $P(z \leq 2.46)$; c) $P(z \geq 1)$; d) $P(z \leq -1)$; e) $P(0.5 < z < 1.5)$.

a) $P(z \leq 1)$: Buscamos en la primera columna el 1, y como no tenemos cifras decimales, buscamos en la primera fila el 0. Obtenemos que $P(z \leq 1) = 0.8413$.

b) $P(z \leq 2.46)$: Hacemos lo mismo, buscamos el 2.4 en la primera columna y el 0.06 en la primera fila. Obtenemos $P(z \leq 2.46) = 0.9931$

c) $P(z \geq 1)$: Como el área total es 1 y la curva es simétrica, $P(z \geq 1) = 1 - P(z \leq 1) = 1 - 0.8413 = 0.1587$.

d) $P(z \leq -1)$: Como el área total es 1 y la curva es simétrica, $P(z \leq -1) = 1 - P(z \leq 1) = 1 - 0.8413 = 0.1587$.

e) $P(0.5 < z < 1.5)$: Calculamos $P(0.5 < z < 1.5) = P(z < 1.5) - P(z < 0.5)$. Buscamos en la tabla y obtenemos $P(0.5 < z < 1.5) = P(z < 1.5) - P(z < 0.5) = 0.9332 - 0.6915 = 0.2417$.



Qué es la distribución normal. Pildoras matemáticas

<https://www.youtube.com/watch?v=phY8Z9-TXCY>



Actividades propuestas

19. Utiliza la tabla de la normal tipificada para calcular:

a) $P(z \leq 0.37)$; b) $P(z < 1.51)$; c) $P(z \geq 0.87)$; d) $P(z \leq -0.87)$; e) $P(0.32 < z < 1.24)$.

Para calcular probabilidades en una $N(\mu, \sigma)$ basta tipificar las variables y buscar las probabilidades en la tabla de $N(0, 1)$.

Actividad resuelta

✚ El consumo familiar diario de electricidad (en kW) en cierta ciudad se puede aproximar por una variable aleatoria con distribución normal de media 5.7 kW y desviación típica 1.1 kW. Calcula la probabilidad de que al tomar una persona al azar su consumo esté comprendido entre 5 kW y 6 kW.

Debemos calcular $P(5 < x < 6)$ en una distribución $N(5.7, 1.1)$. Tipificamos las variables: $z = \frac{x-\mu}{\sigma} = \frac{x-5.7}{1.1}$ por tanto $z = \frac{x-\mu}{\sigma} = \frac{5-5.7}{1.1} = \frac{-0.7}{1.1} = -0.636 \dots$ y $z = \frac{x-\mu}{\sigma} = \frac{6-5.7}{1.1} = \frac{+0.7}{1.1} = 0.2727 \dots$

Entonces:

$$P(5 < x < 6) = P(-0.636 < z < 0.2727) = P(z < 0.2727) - P(z < -0.636) =$$

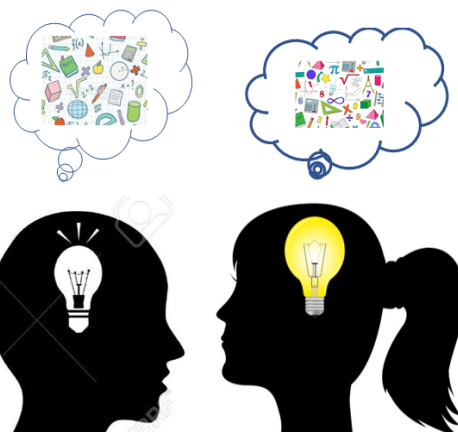
$$P(z < 0.2727) - (1 - P(z < 0.636)) = P(z < 0.2727) - 1 + P(z < 0.636) = 0.6064 - 1 + 0.7389 = 0.3453.$$

Actividades propuestas

20. Se trata a pacientes con trastorno del sueño con un tratamiento que modela el número de días con una distribución normal de media, 290 días y desviación típica, 30. Calcula la probabilidad de que al tomar una persona al azar su tratamiento dure más de 300 días.

21. En una estación meteorológica que las precipitaciones anuales de lluvia tienen una media de 450 mm/m² con una desviación típica de 80 mm/m². Suponemos que la variable aleatoria sigue una distribución normal. Calcula la probabilidad de que: a) Este próximo año la precipitación exceda los 500 mm/m². b) La precipitación esté entre 400 y 510 mm/m². c) La precipitación sea menor de 300 mm/m².

22. En el caso del problema anterior de una $N(450, 80)$ determina la probabilidad de que la variable esté en los intervalos $(\mu - \sigma, \mu + \sigma)$, $(\mu - 2\sigma, \mu + 2\sigma)$, $(\mu - 3\sigma, \mu + 3\sigma)$.



El resultado es el mismo para cualquier normal, verificándose que:

$$P(\mu - \sigma < x < \mu + \sigma) = P(-1 < z < 1) = 0.6826;$$

$$P(\mu - 2\sigma < x < \mu + 2\sigma) = P(-2 < z < 2) = 0.9544;$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = P(-3 < z < 3) = 0.9974$$

como puedes comprobar calculándolo con la tabla pues $P(-a < x < a) = 2 P(x < a) - 1$.

En una distribución normal los valores comprendidos entre $[\mu - \sigma, \mu + \sigma]$ se consideran “normales” (desde el punto de vista estadístico).

- ✓ Un año con precipitaciones entre $[\mu + \sigma, \mu + 2\sigma]$ se considera lluvioso.
- ✓ Un año con precipitaciones entre $[\mu + 2\sigma, \mu + 3\sigma]$ se considera muy lluvioso.
- ✓ Un año con precipitaciones entre $[\mu - 2\sigma, \mu - \sigma]$ se considera seco.
- ✓ Un año con precipitaciones entre $[\mu - 3\sigma, \mu - 2\sigma]$ se considera muy seco.

Y esto mismo se generaliza para cualquier distribución normal.

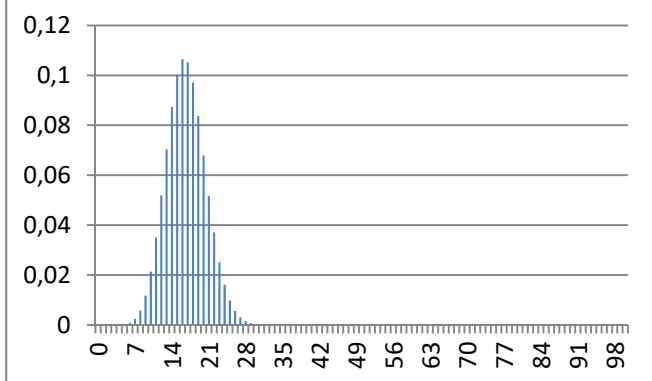
Actividades propuestas

23. En una fábrica de coches se hacen pruebas para conocer el tiempo que tardan sus vehículos en alcanzar la velocidad punta. Se considera que esa variable aleatoria tiempo se distribuye según una distribución normal de media 20 s y desviación típica 2 s. Calcula las probabilidades siguientes: a) Que un vehículo alcance su velocidad punta a los 25 s. b) Alcance su velocidad punta en menos de 25 s. c) La alcance entre 18 s y 22s. d) ¿Qué velocidad punta consideras que tendrán los vehículos rápidos? e) ¿Y los lentos?

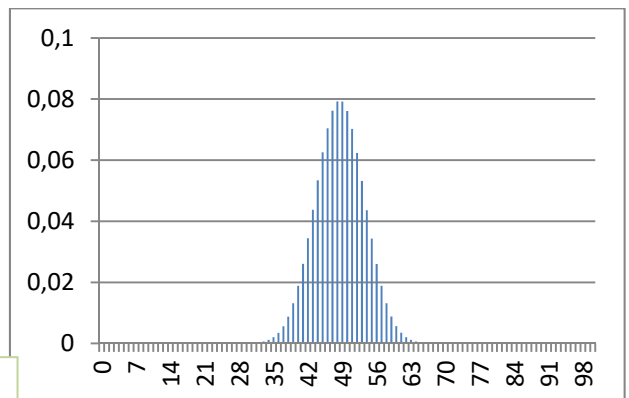
1.6. Aproximación de la binomial a la normal

Hemos visto que la distribución binomial $B(n, p)$ tiene una media $\mu = np$, y una varianza $\sigma = npq$. Queremos analizar en este apartado si la distribución binomial "se ajusta bien" a una normal de igual media y desviación típica. Entenderemos que el ajuste es bueno cuando el área bajo la normal en un cierto intervalo sea *casi igual* al área de los rectángulos de la binomial.

Al estudiar la distribución binomial representamos muchos histogramas de distintas binomiales donde puedes observar que, incluso para valores de n



$B(100, 1/6)$



$B(100, 0.485)$

bajos, el ajuste no es malo. Representamos el histograma de $B(100, 0.485)$ sobre el sexo de los bebés y parece que el ajuste es muy bueno. Al margen puedes observar el histograma del experimento tirar 100 dados y contar el número de cincos: $B(100, 1/6)$ que resultaba muy asimétrico. ¿Qué opinas? ¿Se ajustan a la normal?

Otra forma de hacer la comparación podría ser comparar las áreas en determinados intervalos entre la curva normal y el

histograma de la distribución binomial. Por ejemplo para $B(3, 1/2)$ para $x = 1$ calculamos el área bajo el

histograma para el intervalo $(0.5, 1.5)$ que es 0.38. La media es $\mu = 3/2 = 1.5$ y $\sigma = \sqrt{3 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \sqrt{\frac{3}{4}} = 0.866$

0.866. Tipificamos la normal $N(3/2, 0.866)$ y calculamos:

$$P(0.5 < x < 1.5) = P\left(\frac{0.5-1.5}{0.866} < z < \frac{1.5-1.5}{0.866}\right) = P(-1.1547 < z < 0) = P(z < 0) - P(z < -1.1547) =$$

$$P(z < 0) - (1 - P(z < 1.1547)) = P(z < 0) + P(z < 1.1547) - 1 = 0.5 + 0.8749 - 1 = 0.3749.$$

Hasta en este caso tan desfavorable el ajuste es bueno.

Se puede demostrar (utilizando el teorema central del límite) que el ajuste es bueno entre binomial y normal cuando las variables son independientes y además $npq \geq 9$.

Al estudiar la distribución binomial no hicimos los cálculos en muchos de los ejercicios pues eran muy laboriosos. Sin embargo, mirar la tabla de la normal es bastante más rápido y sencillo.

Observa también que no hemos tomado el valor $x = 1$ pues para tomar intervalos le hemos restado a 1 y sumado a 1 la longitud del intervalo: 0.5, y hemos tomado el intervalo $(0.5, 1.5)$.

Actividad resuelta

✚ En una determinada población se divide la población activa en dos grupos, los que trabajan en agricultura y servicios que son un 44 %, y el resto. Se elige al azar una muestra de 200 personas entre la población activa, ¿qué probabilidad hay de que haya entre 80 y 100 personas del primer grupo?

Es un problema de distribución binomial $B(200, 0.44)$ pues una persona o pertenece a dicho grupo, o no pertenece. Sabemos que:

$$p(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x} = \binom{200}{x} \cdot 0.44^x \cdot 0.56^{200-x}$$

Y deberíamos calcular:

$$P(80 \leq x \leq 100) = \sum_{x=80}^{x=100} p(x) = \sum_{x=80}^{x=100} \binom{200}{x} \cdot 0.44^x \cdot 0.56^{200-x}$$

Habíamos advertido que el cálculo era laborioso, pero ahora podemos utilizar el ajuste de la binomial a la normal. Calculamos la media y la desviación típica:

$$\mu = np = 200 \cdot 0.44 = 88,$$

$$\sigma = \sqrt{200 \cdot 0.44 \cdot 0.56} = 7.02,$$

por lo que ajustamos con la normal $N(88, 7.02)$.

Como la longitud de cada intervalo es 1, se añade a cada valor 0.5 para ir desde el extremo del intervalo, y no desde el centro.

$$P(80 - 0.5 \leq x \leq 100 + 0.5)$$

Ahora tipificamos:

$$P\left(\frac{80-88-0.5}{7.02} < z < \frac{100-88+0.5}{7.02}\right) = P(z \leq 1.78) + P(z \leq 1.21) - 1 = 0.9625 + 0.8869 - 1 = 0.8494$$

En el 85 % de los casos habrá entre 80 y 100 personas del primer grupo.

Como $npq = 49.28 \geq 9$, el ajuste es bueno.

Actividades propuestas

24. Se lanza una moneda mil veces, ¿cuál es la probabilidad de que el número de caras obtenidas esté entre 400 y 600? ¿Y de que sea mayor que 800?
25. En una fábrica de bombillas de bajo consumo se sabe que el 70 % de ellas tienen una vida media superior a 1 000 horas. Se toma una muestra de 50 bombillas, ¿cuál es la probabilidad de que haya entre 20 y 30 cuya vida media sea superior a mil horas?, ¿y la probabilidad de que haya más de 45 cuya vida media sea superior a 1 000 horas?
26. Se investigan a pie de urna las preferencias de votos en la Comunidad de Madrid. De 2 000 encuestas 700 votan al partido X. Cuantos tendrían que votar al partido estudiado para que ganara con un 99 % de confianza.

En algunos problemas nos darán ciertas probabilidades para que nosotros obtengamos los parámetros. Este tipo de ejercicios son muy habituales en Selectividad, casi siempre hay un apartado de este tipo. Aquí tenemos dos ejemplos:

60. El peso, en gramos, de los recién nacidos en la maternidad de un hospital sigue una distribución normal de media 3000 g. Si el 95 % de los bebés tiene un peso comprendido entre 2600 y 3400 g ¿Cuál es el valor de σ ?

Considera la variable aleatoria X : "el peso, en gramos, de los recién nacidos". X tiene una distribución $N(\mu = 3000; \sigma)$. Si el peso del 95% de los bebés está entre 2600 y 3400 g, ello significa que:

$$P(2600 < X < 3400) = 0,95 \Rightarrow P\left(\frac{2400 - 3000}{\sigma} < Z < \frac{3400 - 3000}{\sigma}\right) = 0,95 \Rightarrow P\left(-\frac{600}{\sigma} < Z < \frac{600}{\sigma}\right) = 0,95$$

Ahora bien, como $P(-a < Z < a) = 2 \cdot \Phi(a) - 1$ cualquiera que sea $a > 0$, se tiene que

$$2\Phi\left(\frac{600}{\sigma}\right) - 1 = 0,95 \Rightarrow \Phi\left(\frac{600}{\sigma}\right) = 0,975 \Rightarrow \frac{600}{\sigma} = 1,96 \Rightarrow \sigma = 306,12$$

61. La producción de trigo por hectárea de terreno en una comarca sigue una distribución $N(\mu, \sigma)$ Los datos históricos indican que solo en el 10 % de los años la producción supera los 4000 kg/ha, mientras que en el 60 % de los años queda por debajo de los 3200 kg/ha.

- a) Calcula la media y la desviación típica de la distribución.
b) Calcula la probabilidad de que la producción alcance los 3500 kilos por hectárea en una un año elegido al azar.

- a) Considerando la variable aleatoria X : "producción de trigo por hectárea en la comarca", cuya distribución es $N(\mu, \sigma)$. Que el 10% de los años la producción supere los 4000 kg/ha significa que: $P(X > 4000) = 0,1$, Mientras que el 60% de los años la producción queda por debajo de los 3200 kg/ha significa que: $P(X < 3200) = 0,6$.

Con la información disponible se tiene que: $P(X > 4000) = 0,1 \Rightarrow P\left(Z > \frac{4000 - \mu}{\sigma}\right) = 0,1 \Rightarrow \frac{4000 - \mu}{\sigma} = 1,281$

$$P(X < 3200) = 0,6 \Rightarrow P\left(Z < \frac{3200 - \mu}{\sigma}\right) = 0,6 \Rightarrow \frac{3200 - \mu}{\sigma} = 0,253$$

El sistema de dos ecuaciones con dos incógnitas, queda:
$$\begin{cases} \mu + 1,281\sigma = 4000 \\ \mu + 0,253\sigma = 3200 \end{cases}$$

cuya solución es $\mu = 3003,11$ y $\sigma = 778,21$ kg/ha.

- b) Con los parámetros calculados en el apartado a), se tiene que:

$$P(X \geq 3500) = P\left(Z > \frac{3500 - 3003,11}{778,21}\right) = P(Z > 0,64) = 1 - \Phi(0,64) = 1 - 0,7389 = 0,2611$$

A continuación te proponemos dos ejercicios parecidos:

65. El encargado de una plantación de chopos asegura que, en este momento, el diámetro de los árboles sigue una distribución normal de media 20 cm y que el 90 % de ellos tiene un diámetro inferior a 25 cm.
- a) Calcula la desviación típica de la distribución.
b) Calcula la probabilidad de que un árbol tenga más de 22 cm de diámetro.
69. La anchura de las hojas de una especie de árbol siguen una distribución normal de media $\mu = 4$ cm. Además, se ha observado experimentalmente que la anchura del 90 % de las hojas es inferior a 5 cm.
- a) Halla la varianza de la distribución.
b) Calcula la probabilidad de que la anchura de una hoja elegida al azar sea mayor de 6 cm.
c) ¿Cuál es la anchura máxima de las hojas que se encuentran entre el 30% con menor anchura?