Tema 4: Modelos probabilísticos

1. Variables aleatorias:
    a) Concepto.
    b) Variables discretas y continuas.
    c) Función de probabilidad (densidad) y función de distribución.
    d) Media y varianza de una variable aleatoria.

2. Modelos probabilísticos:
    a) Bernoulli.
    b) Binomial.
    c) Normal.
    d) Aproximación de binomial a normal.

Lecturas recomendadas:

- Capítulos 22 y 23 del libro de Peña y Romo (1997)

4.1: Variables aleatorias

- Se llama variable aleatoria a toda función que a cada elemento del espacio muestral asocia un número real.

- Se utilizan letras mayúsculas para designar las variables aleatoria: X, Y, Z; y sus respectivas letras minúsculas para los valores concretos de las mismas: x, y, z.

Variable aleatoria discreta Es la que solo puede tomar una cantidad numerable de valores.

Variable aleatoria continua Es aquella que puede tomar infinitos valores dentro de un intervalo de la recta real.

Función de probabilidad de una v.a. discreta: Es la función que asocia a cada valor x de la v.a. X su probabilidad p.

- Los valores que toma una v.a. discreta X y sus correspondientes probabilidades suelen disponerse en una tabla con dos filas o dos columnas llamada tabla de distribución de probabilidad:

X	x_1	x_2	...	x_n
$P(X=x_i)$	p_1	p_2	...	p_n

Toda función de probabilidad se verifica que $p_1 + p_2 + p_3 + \cdots + p_n = 1$

Función de distribución de una v.a. discreta: Sea X una v.a. cuyos valores suponemos ordenados de menor a mayor. Se llama función de distribución de la variable X a la función que asocia a cada valor de la v.a. la probabilidad acumulada hasta ese valor, es decir, $F(x) = P(X \leq x)$

Media, varianza y desviación típica de una variable aleatoria discreta

Se llama media o esperanza de una v.a. discreta X, que toma los valores x_1, x_2, ...con probabilidades p_1, p_2,... al valor de la siguiente expresión:

$$\mu = \sum_i x_i P(X = x_i) = \sum_i x_i p_i$$

La varianza viene dada por la siguiente fórmula: $\sigma^2 = \sum_i (x_i - \mu)^2 p_i$
que puede calcularse mediante:

$$\sigma^2 = \sum_i x_i^2 p_i - \mu^2$$

Ejemplo: La distribución de probabilidad de una v.a. X viene dada por la siguiente tabla:

x_i	1	2	3	4	5
p_i	0.1	0.3	?	0.2	0.3

¿Cuánto vale P(X=3)?
Calcula la media y la varianza.

4.2: Modelos probabilísticos

Modelos discretos	Modelos continuos
Ensayos de Bernoulli y las distribuciones geométrica y binomial	La distribución normal y distribuciones relacionadas

Ensayos de Bernoulli y las distribuciones geométrica y binomial

Un modelo de Bernoulli es un experimento que tiene las siguientes características:

- En cada prueba del experimento sólo son posibles dos resultados, usualmente denominados éxito y fracaso.
- El resultado obtenido en cada prueba es independiente de los resultados anteriores.
- La probabilidad de éxito es constante, P(éxito) = p, y no varia de una prueba a otra.

La distribución geométrica

Supongamos un modelo de Bernoulli. ¿Cuál es la distribución del número de fracasos, F, antes del primer éxito?

- P(F=0) = P(0 fracasos antes del primer éxito) = p
- P(F=1) = P(fracaso, éxito) = (1-p)p
- P(F=2) = P(fracaso, fracaso, éxito) = $(1-p)^2$ p
- P(F=f) = P(f fracasos antes del primer éxito) = $(1-p)^f$ p para f = 0, 1, 2, …

La distribución de F se llama la distribución geométrica con parámetro p.

$$E[F] = (1-p)/p \qquad V[X] = (1-p)/p^2$$

La distribución binomial

Supongamos un modelo de Bernoulli. ¿Cuál es la distribución del número de éxitos, X, en n intentos?

$$\mathrm{P}(\textit{Obtener} \;\; r \;\textit{éxitos}) = P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

para r = 0,1,2, …, n.

La distribución de X se llama la distribución binomial con parámetros n y p.

E[X] = np V[X] = np(1-p)

EJEMPLO

Calcula la probabilidad de que una familia que tiene 4 hijos, 3 de ellos sean varones.

EJEMPLO

La probabilidad de que un alumno de repita curso es de 0,3.

- Elegimos alumnos al azar. ¿Cuál es la probabilidad de que el primero que repita curso sea el tercer alumno que elegimos?
- Elegimos 20 alumnos al azar. ¿Cuál es la probabilidad de que haya exactamente 4 alumnos repetidores?

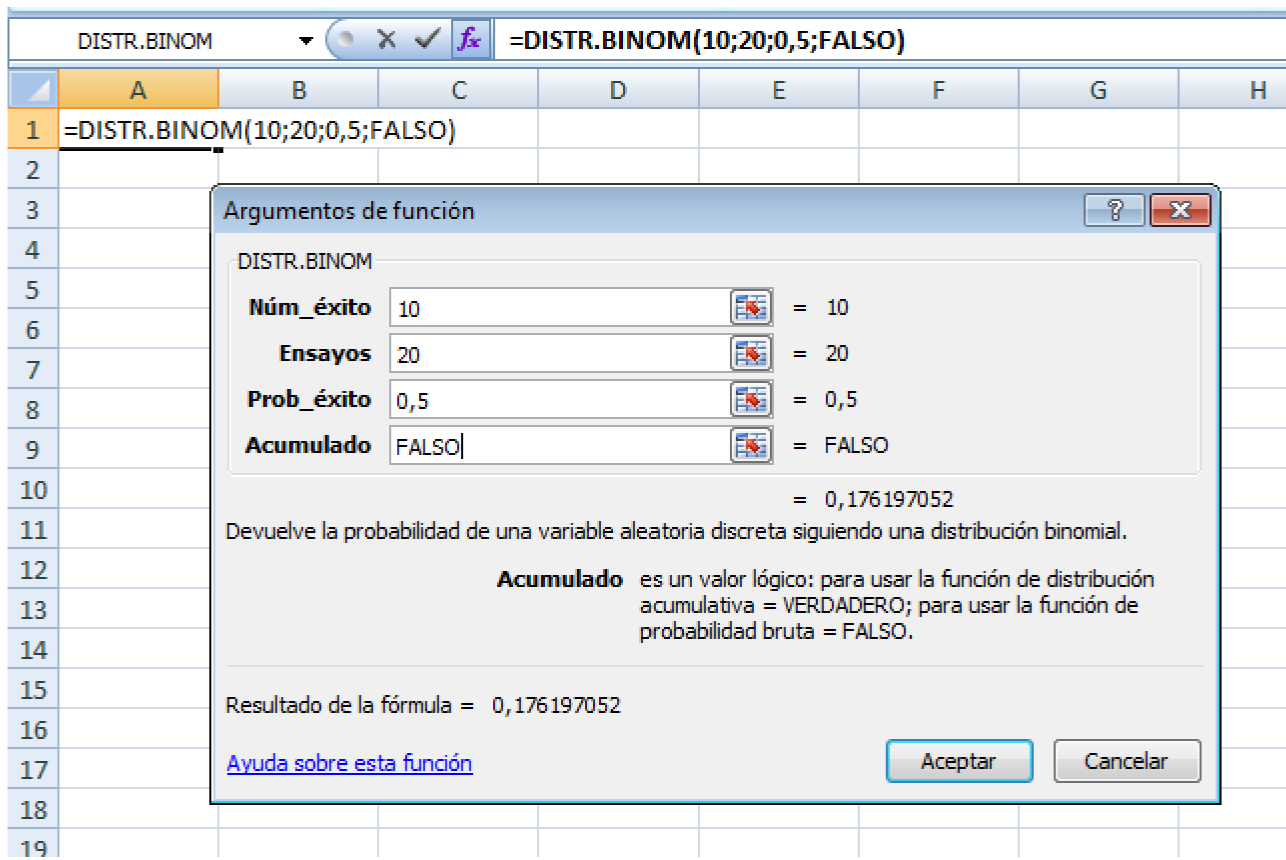
EJEMPLO

La probabilidad de que un reloj salga de fábrica defectuoso es del 4%. Halla el número esperado de relojes defectuosos en un lote de 1000

Cálculo con Excel

Probabilidades binomiales son difíciles de calcular “a mano” salvo en los casos simples de 0 o 1 (n o n-1) éxitos.

¡En Excel es mucho más sencillo!



Ejemplo (Pregunta de Test)

De todas las ONGs que existen en España, el 30% de ellas se dedican especialmente a la infancia. Se seleccionan 50 ONG’s de manera aleatoria. Calcular el número esperado de ellos que se dedican a la infancia.

Ejemplo (Pregunta de Test)

En promedio, uno de cada diez afiliados de CCOO es un delegado.

a) En unas entrevistas a afiliados CCOO, ¿cuál es la probabilidad de que el primer delegado sea la segunda persona entrevistada?

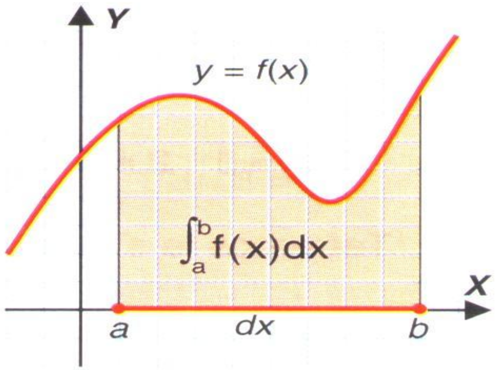
b) Hay 4 afiliados de CCOO en la Chimbomba. ¿Cuál es la probabilidad de que no haya ningún delegado?

c) En una muestra de 100 afiliados, ¿cuál es el número esperado de delegados?

Función de densidad de una variable aleatoria continua

La función de densidad de una v.a. continua cumple las siguientes condiciones:

- Sólo toma valores no negativos, $f(x) \geq 0$.
- El área encerrada bajo la curva es igual a 1.



Función de distribución. Como en el caso de la v.a. discreta, la función de distribución proporciona la probabilidad acumulada hasta un determinado valor de la variable, es decir, $F(x) = P(X \leq x)$

Cumple las siguientes condiciones:

- Su valor es cero para todos los puntos situados a la izquierda del menor valor de la variable.
- Su valor es 1 para todos los puntos situados a la derecha del mayor valor de la variable.

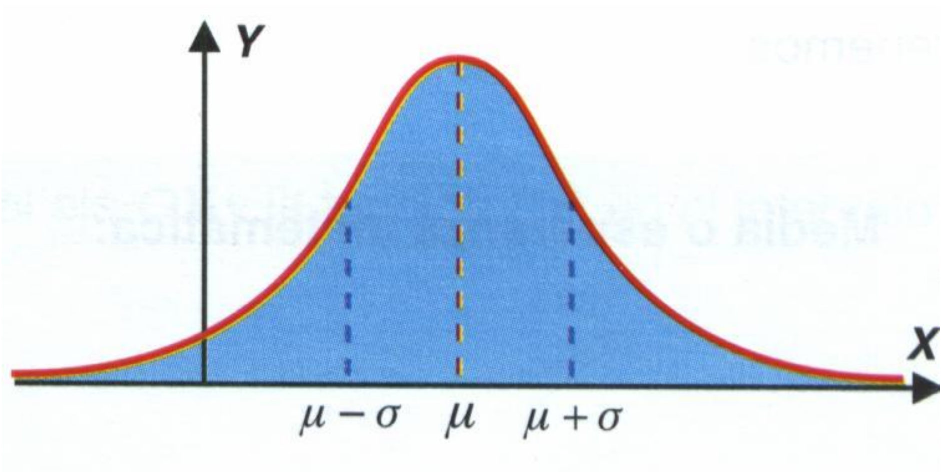
La distribución normal o gaussiana

Hay v.a. continuas cuya función de densidad tiene forma de campana.

Ejemplos:
- La variable peso en una población de personas de la misma edad y sexo.
- La variable altura de la población citada.
- Las notas de una asignatura (*mito urbano*).

Para expresar que una v.a. continua X, tiene una distribución normal de media μ y desviación típica σ , escribimos:

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



La distribución normal estándar

De las infinitas distribuciones normales, tiene especial interés la que tiene media igual a 0 y desviación típica igual a 1, es decir, N(0,1). Esta distribución recibe el nombre de normal estándar o tipificada.

Existen tablas que permiten calcular probabilidades de la distribución N(0,1).

Por ello cuando tenemos una v.a. X que sigue una distribución normal de media μ y desviación típica σ pasamos a otra variable Z que sigue una distribución N(0,1) mediante la siguiente transformación:

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

Ejemplos

Sea $Z \sim N(0,1)$. Hallar las siguientes probabilidades

- $P(Z < -1)$
- $P(Z > 1)$
- $P(-1{,}5 < Z < 2)$

Hallar los percentiles de 90%, 95%, 97,5% y 99% de la distribución normal estándar.

Estos valores son útiles en los siguientes temas

Sea $X \sim N(2,4)$. Hallar las siguientes probabilidades

- $P(X < 0)$
- $P(-1 < X < 1)$

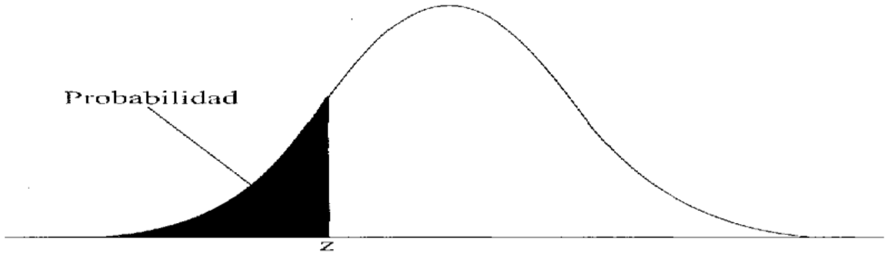


Tabla 3. *Probabilidad de que una variable normal de media cero y desviación típica uno tome un valor menor que z*

z	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
—3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
—3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
—3,2	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
—3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
—3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
—2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
—2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
—2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
—2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
—2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
—2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
—2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
—2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
—2,1	0,0179	0,0174	0,0170	0,0166	0,016	0,0158	0,0154	0,0150	0,0146	0,0143
—2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
—1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
—1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
—1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
—1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
—1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
—1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
—1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0855	0,0869	0,0853	0,0838	0,0823
—1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
—1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
—1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
—0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
—0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
—0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
—0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
—0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
—0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
—0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
—0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
—0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
—0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,721	0,4681	0,4641

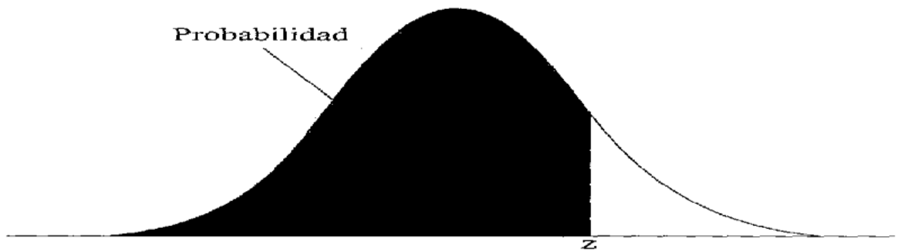


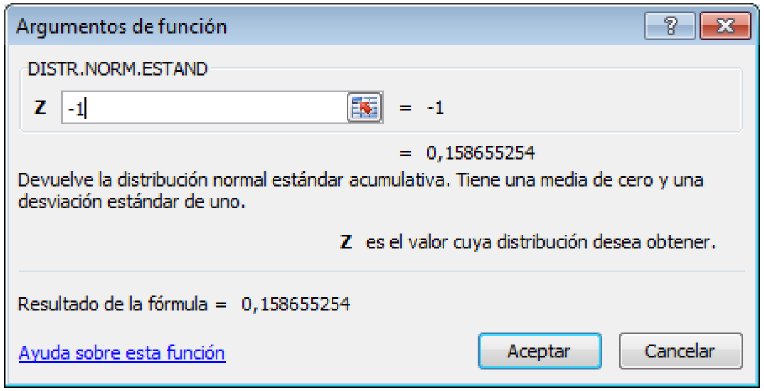
Tabla 3. *(continuación) Probabilidad de que una variable normal de media cero y desviación típica uno tome un valor menor que z*

z	**0,00**	**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	09981,	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	0,9989	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

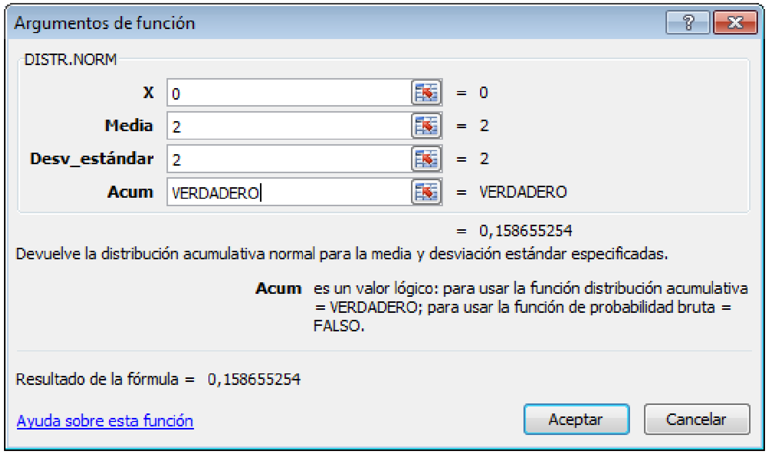
Cálculo con Excel

Es más fácil hacer los cálculos en Excel …

con la distribución normal estándar …



… o directamente con la distribución original.



Aproximación de la distribución binomial mediante la normal

Cuando n es suficientemente grande el comportamiento de una distribución binomial, $X \sim B(n, p)$, es aproximadamente igual a una distribución normal, con media np y varianza np(1-p).

Se considera que esta aproximación es "buena" cuando $np > 5$ y $n(1-p) > 5$.

EJEMPLO

Se lanza una moneda correcta al aire 400 veces. Calcula la probabilidad de obtener un número de caras comprendido entre 180 y 210, ambos inclusive.

La solución exacta haciendo el cálculo en Excel con la distribución binomial es 0,833.

Utilizando la aproximación normal, la solución aproximada es de 0,819.

Se puede mejorar la aproximación con una *corrección de continuidad* y en este caso las soluciones exactas y aproximadas coinciden hasta tres decimales … pero ¿porqué usarla?

El uso real de esta aproximación se verá en los siguientes temas.

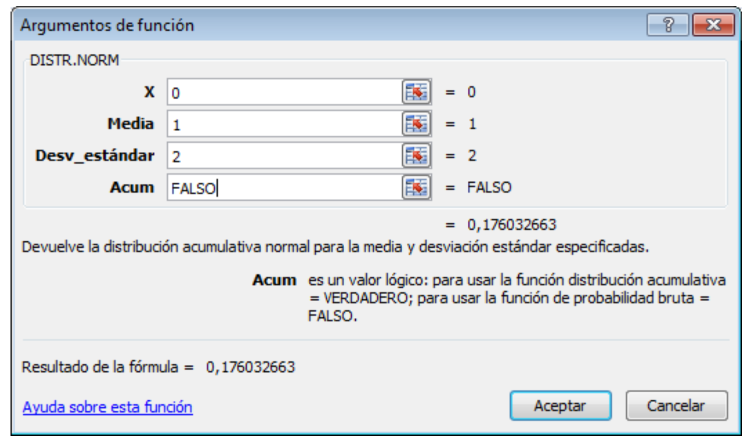
Ejemplo (Pregunta de examen)

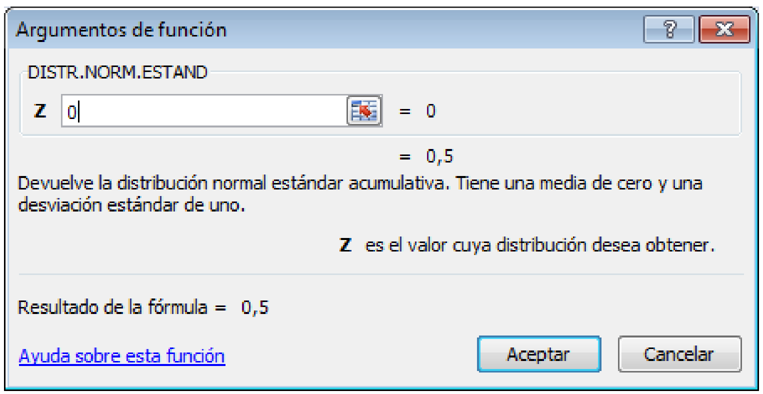
Según un sondeo del CIS, la evaluación media de satisfacción con Mariano Rajoy es de 3,09 con una desviación típica de 2,5. Suponiendo que las evaluaciones siguen una distribución normal, si se elige una persona al azar, la probabilidad de que dé a Rajoy una evaluación de menos de 3,09 es:

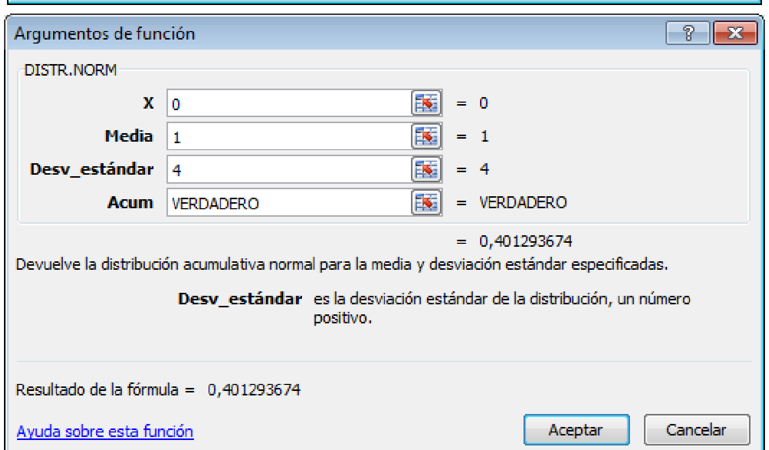
a) 0,5.
b) 0.
c) 1.236
d) 1.

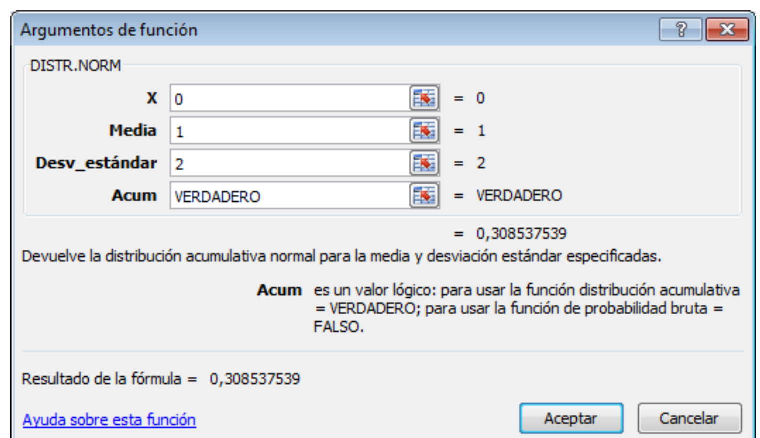
Ejemplo (Pregunta de examen)

La tasa de inflación en un país sigue una distribución normal con media 1 y varianza 4. Señala cuál de las siguientes fórmulas en Excel proporciona la probabilidad de que la inflación sea negativa.









Ejemplo: (Pregunta de examen)

Los siguiente tabla proporciona las valoraciones de los ministros del gobierno en el último barómetro del CIS:

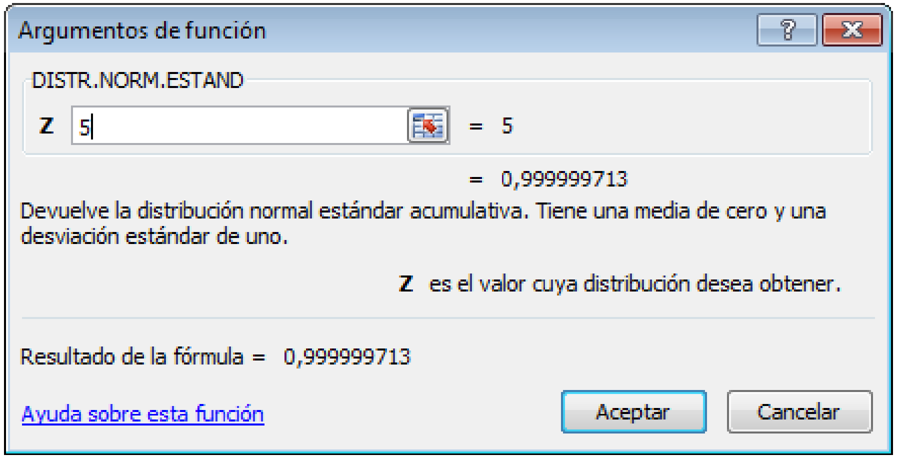
a) Suponiendo que las valoraciones de Carme Chacón siguen una distribución normal con media 4,55 y desviación típica 2,6, calcular la probabilidad de que un español la valore con más de un 4,55.

b) En un grupo de 3 españoles, ¿cuál es la probabilidad de que todos valoren a Carme con más de un 5?*

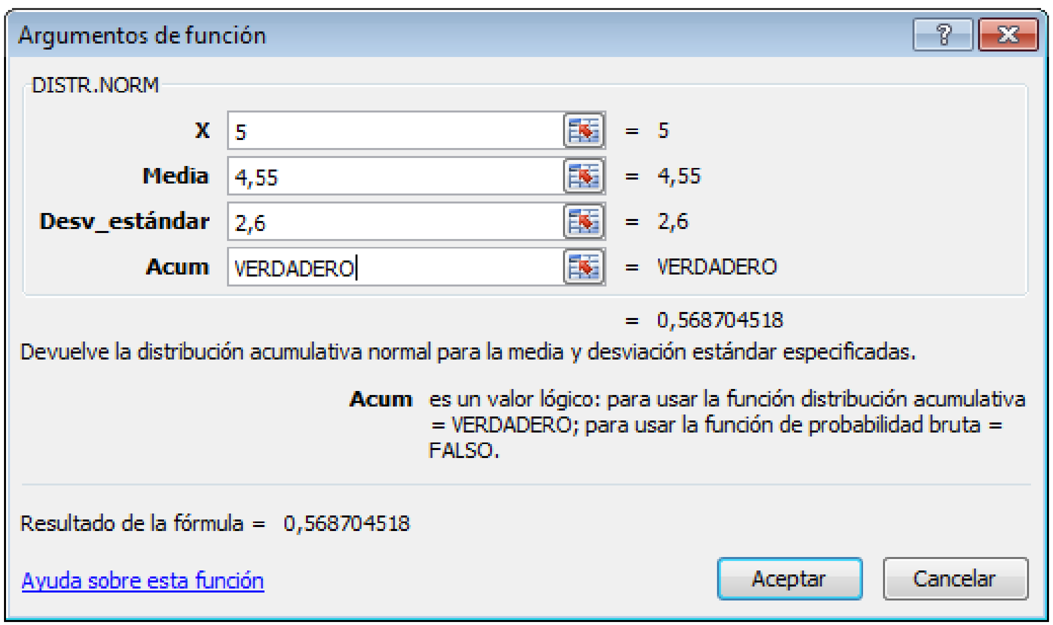
c) La ministra peor valorada es Ángeles González Sinde. Suponiendo que las valoraciones siguen una distribución normal, calcular la probabilidad de que un español la proporcione una valoración de exactamente 5.

	Media	Desviación típica	(N)
Rosa Aguilar	4.12	2.31	(1092)
José Blanco	3.69	2.49	(1672)
Francisco Caamaño	3.49	2.29	(830)
Carme Chacón	4.55	2.60	(1943)
Manuel Chaves	3.40	2.54	(1816)
Ángel Gabilondo	3.89	2.52	(926)
Cristina Garmendia	3.62	2.31	(685)
Valeriano Gómez	3.19	2.35	(616)
Ángeles González Sinde	2.55	2.17	(1048)
Ramón Jáuregui	3.78	2.40	(976)
Trinidad Jiménez	3.94	2.38	(1548)
Leire Pajín	2.76	2.54	(1691)
Alfredo Pérez Rubalcaba	4.72	2.87	(2012)
Elena Salgado	3.74	2.50	(1443)
Miguel Sebastiánn	3.34	2.38	(962)

*Ver la información en la siguiente Transparencia.



0,999999713



0,568704518