8 – Principales distribuciones de probabilidad

Modelos teóricos de distribución de probabilidad

• El conocimiento acumulado en Psicología ha permitido evidenciar como algunas variables de interés en este campo se distribuyen de un modo característico, esto es, tienen una distribución de probabilidad particular que se repite a lo largo del tiempo y para diferentes muestras. De muchos de estos patrones regulares de distribución de probabilidad han sido planteados los modelos teóricos que representan matemáticamente a esas distribuciones y que, por tanto, permiten obtener fácilmente, a partir de una función matemática, cuál será la probabilidad (o probabilidad acumulada) asociada a un valor cualquiera de la variable.

• Dos de los modelos más relevantes en el contexto de la Psicología son el de la distribución binomial, para variables categóricas, y el de la distribución normal, para variables cuantitativas. En los siguientes apartados se describen las características de estos dos modelos teóricos de distribución de probabilidad y se muestra su aplicación en la práctica.

• Otros modelos teóricos de distribución de probabilidad como la distribución *t* de Student, la distribución *ji-cuadrado* y la distribución *F* de Snedecor son también especialmente importantes en el campo de la Estadística, ello debido a que la 'distribución muestral' de algunos estadísticos se ajusta a estos modelos teóricos de distribución de probabilidad. La distribución muestral de un estadístico es un concepto clave en la estadística inferencial que será introducido en un capítulo posterior.

1. La distribución binomial

• Comencemos con un ejemplo: supongamos que se elige al azar una muestra de 6 personas para formar parte de un jurado popular que ha de juzgar a una persona inmigrante y sabemos que, en la población de la que se ha extraído esa muestra, un 30% de las personas son racistas. A partir de estos datos, ¿cuál es la probabilidad de que la mitad o más de los miembros del tribunal sean racistas?

• La respuesta a la pregunta anterior se puede resolver fácilmente si asumimos que la distribución de probabilidad de la variable que nos ocupa se corresponde con la de la distribución binomial. Veamos más formalmente las condiciones para poder asumir que la distribución de probabilidad de una variable se ajusta al modelo teórico de la distribución binomial:

→ 1ª condición: que se tenga una variable categórica dicotómica de la que se conozca su distribución de probabilidad en la población de interés.

Sea la variable dicotómica X [X_1; X_2]: lo primero que se debe decidir es cuál de las dos modalidades de la misma es la que nos interesa, esto es, cuál es la que se corresponde con aquello que nos interesa conocer (sea, por ejemplo, X_1). La probabilidad asociada a esa modalidad se expresa simbólicamente como π (y, complementariamente, a la de X_2 como $1-\pi$):

X_i	$P(X_i)$
X_1	π
X_2	$1-\pi$
	1

Ejemplo: Variable “Ser racista” [Si; No] => modalidad de interés para contestar a la pregunta planteada en el ejemplo: ‘Si’ => $P(X_1) = \pi = 0{,}30$

X_i	$P(X_i)$
Si	0,3
No	0,7
	1

Señalar que en la práctica del análisis de datos es bastante habitual contar con datos de variables dicotómicas (o dicotomizadas), por ejemplo, variables en que se han recogida datos del tipo correcto/incorrecto, a favor/en contra, de acuerdo/en desacuerdo, bien/mal, si/no, curado/no curado, tratamiento/no tratamiento, etc. Destacar también que es bastante frecuente en la literatura de la distribución binomial considerar a esas dos modalidades, de forma genérica, con los términos “Éxito” y “Fracaso”, de modo que $P(Éxito) = \pi$ y $P(Fracaso) = 1-\pi$. Esta práctica puede en ocasiones generar cierta confusión debida a que el significado de las palabras *éxito* y *fracaso* no encaja con el significado de las dos modalidades de muchas variables categóricas –como es el caso de la variable de nuestro ejemplo.

→ 2ª condición: Se realiza *n* veces el experimento aleatorio representado por la variable aleatoria en cuestión, por ejemplo, medir una variable X en una muestra de *n* casos –cada vez que se mide la variable es un experimento aleatorio, pues no sabemos a priori que va a ocurrir con la variable *X*. Se debe satisfacer la condición de que se mantenga π constante a lo largo de la realización de esos experimentos.

Ejemplo: En el caso del jurado popular, se realizaría 6 veces (n = 6) el experimento aleatorio consistente en elegir un miembro del tribunal, pues son seis los miembros del tribunal elegidos al azar de la población. Sabemos que la probabilidad de ser racista en la población es de 0,30 (π = 0,30) y es razonable asumir que esa probabilidad se va a mantener constante a lo largo del proceso de elección de esos seis miembros. Podría no ser razonable tal asunción si, por ejemplo, si dilatara mucho en el tiempo la elección de cada uno de los miembros del tribunal, pues a lo largo de ese tiempo podría cambiar la probabilidad de ser racista en la población.

• Si se cumplen las dos condiciones anteriores, la variable aleatoria "Número de experimentos (casos) en los que se verifica la condición X_1" se distribuye según el modelo teórico de la distribución binomial. Si la distribución de probabilidad de una variable *X* se ajusta al modelo binomial se expresa simbólicamente como: $X \to B(n;\pi)$

Ejemplo: como se cumplen las dos condiciones anteriores, podemos afirmar que la variable "Número de miembros del tribunal que son racistas" (*X*) se distribuye según la distribución binomial con n = 6 y π = 0,30

$$X \to B(6;\ 0{,}30)$$

• La distribución de probabilidad (función de probabilidad) de una variable binomial *X* viene definida por la siguiente <u>función matemática</u>, donde X_i puede oscilar entre 0 y n:

$$P(X_i) = \frac{n!}{X_i!(n-X_i)!} \cdot \pi^{X_i} \cdot (1-\pi)^{n-X_i}$$

Si, por **ejemplo**, queremos conocer la probabilidad de que 2 miembros del tribunal sean racistas:

$$P(2) = \frac{6!}{2!(6-2)!} \cdot 0,30^2 \cdot (1-0,30)^{6-2} = \frac{720}{2 \cdot 24} \cdot 0,30^2 \cdot 0,70^4 = 0,324$$

Si sustituimos en la fórmula del modelo binomial, los distintos valores que puede tomar *X* en nuestro ejemplo, obtendríamos la distribución de probabilidad completa de la variable “Nº de miembros del tribunal que son racistas (*X*):

X_i	$P(X_i)$
0	0,118
1	0,303
2	0,324
3	0,185
4	0,060
5	0,010
6	0,001
	1

• Otro procedimiento alternativo para obtener los valores de la distribución de probabilidad de una variable que se ajuste al modelo binomial es acudir a la tabla de la distribución de probabilidad de este modelo, la cual se puede encontrar en el *Apéndice de Tablas* que suele aparecen en la parte final de muchos libros de estadística. En la misma se pueden encontrar tabuladas las distribuciones de probabilidad de una variable binomial para diferentes valores de π y de *n*. : Observar el fragmento adjunto de la tabla de la distribución binomial.

Fragmento de la tabla de la distribución binomial

		π													
n	x	0,01	0,05	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	0,95	0,99	x
2	0	980	902	810	640	490	360	250	160	090	040	010	002	0+	0
	1	020	095	180	320	420	480	500	480	420	320	180	095	020	1
	2	0+	002	010	040	090	160	250	360	490	640	810	902	980	2
3	0	970	857	729	512	343	216	125	064	027	008	001	0+	0+	0
	1	029	135	243	384	441	432	375	288	189	096	027	007	0+	1
	2	0+	007	027	096	189	288	375	432	441	384	243	135	029	2
	3	0+	0+	001	008	027	064	125	216	343	512	729	857	970	3
4	0	961	815	656	410	240	130	062	026	008	002	0+	0+	0+	0
	1	039	171	292	410	412	346	250	154	076	026	004	0+	0+	1
	2	001	014	049	154	265	346	375	346	265	154	049	014	001	2
	3	0+	0+	004	026	076	154	250	346	412	410	292	171	039	3
	4	0+	0+	0+	002	008	026	062	130	240	410	656	815	961	4
5	0	951	774	590	328	168	078	031	010	002	0+	0+	0+	0+	0
	1	048	204	328	410	360	259	156	077	028	006	0+	0+	0+	1
	2	001	021	073	205	309	346	312	230	132	051	008	001	0+	2
	3	0+	001	008	051	132	230	312	346	309	205	073	021	001	3
	4	0+	0+	0+	006	028	077	156	259	360	410	328	204	048	4
	5	0+	0+	0+	0+	002	010	031	078	168	328	590	774	951	5
6	0	941	735	531	262	118	047	016	004	001	0+	0+	0+	0+	0
	1	057	232	354	393	303	187	094	037	010	002	0+	0+	0+	1
	2	001	031	098	246	324	311	234	138	060	015	001	0+	0+	2
	3	0+	002	015	082	185	276	312	276	185	082	015	002	0+	3
	4	0+	0+	001	015	060	138	234	311	324	246	098	031	001	4
	5	0+	0+	0+	002	010	037	094	187	303	393	354	232	057	5
	6	0+	0+	0+	0+	001	004	016	047	118	262	531	735	941	6
7	0	932	698	478	210	082	028	008	002	0+	0+	0+	0+	0+	0
	1	066	257	372	367	247	131	055	017	004	0+	0+	0+	0+	1
	2	002	041	124	275	318	261	164	077	025	004	0+	0+	0+	2
	3	0+	004	023	115	227	290	273	194	097	029	003	0+	0+	3
	4	0+	0+	003	029	097	194	273	290	227	115	023	004	0+	4

Ejemplo Buscando en la misma podríamos obtener fácilmente, por ejemplo, la distribución de probabilidad de una variable *X* que se distribuya según el modelo binomial con parámetros $n = 4$ y $\pi = 0{,}50$ [$X \rightarrow B(4;0{,}50)$]

X_i	$P(X_i)$
0	0,062
1	0,250
2	0,375
3	0,250
4	0,062

Un ejemplo de variable que se distribuye según ésta distribución de probabilidad es el "Nº de veces que sale cara al lanzar una moneda al aire 4 veces".

• Una tabla más amplia de la distribución binomial puede encontrarse en las páginas finales de muchos libros de estadística, o bien, podemos recorrer a la obtención informatizada del valor exacto de probabilidad que nos interese. Por ejemplo, en el programa *MS Excel®* basta con introducir en una casilla cualquiera la siguiente expresión con los valores entre paréntesis que nos interese:

```
=DISTR.BINOM(núm_éxito;ensayos;prob_éxito;0)
```

Por ejemplo, si escribimos en una celda `=DISTR.BINOM(2;6;0,30;0)` y pulsamos la tecla *Intro*, automáticamente obtendremos el valor de probabilidad que anteriormente ya calculamos con la fórmula del modelo binomial (P = 0,324). Si el "0" que aparece en último lugar en el paréntesis lo cambiamos por un "1", obtendremos la probabilidad acumulada, esto es, la probabilidad de obtener el valor que nos interesa o uno inferior.

Ejercicio 1: Para la variable "Nº de miembros del tribunal que son racistas" presentada antes, obtener las siguientes probabilidades: (a) de que 4 miembros del tribunal sean racistas; (b) de que, como máximo, 2 sean racistas; (c) de que más de la mitad sean racistas; (d) obtener la esperanza matemática y la varianza de esta variable aleatoria.

Ejercicio 2: Sabiendo que en la población española la proporción de mujeres es de 0,60, (a) ¿cuál es la probabilidad de que al extraer una muestra aleatoria de 7 personas de esa población, ninguna de ellas sea mujer? (b) ¿y cuál la de que todas fuesen mujeres? (c) construir la distribución de probabilidad correspondiente a la variable "Nº de mujeres al extraer al azar una muestra de 7 personas de la población española" (X); (d) obtener la media (o valor esperado), la mediana y la moda de la variable aleatoria X; (e) representar gráficamente la función de probabilidad de X; (f) ídem de la función de distribución de X; (g) obtener la distribución de probabilidad de la variable aleatoria complementaria, esto es, la de la variable "Nº de hombres al extraer al azar una muestra de 7 personas de la población española".

• Es una propiedad de cualquier variable que se distribuye según la distribución binomial que su valor esperado (media) y su varianza se pueden obtener según las siguientes fórmulas:

$$\mu_X = n \cdot \pi$$
$$\sigma^2_X = n \cdot \pi \cdot (1-\pi)$$

Ejercicio 3: Obtener con las anteriores fórmulas la esperanza matemática y la varianza de la variable "Nº de miembros del tribunal que son racistas". Comprobar que coincide con los valores ya obtenidos anteriormente para estos dos índices.

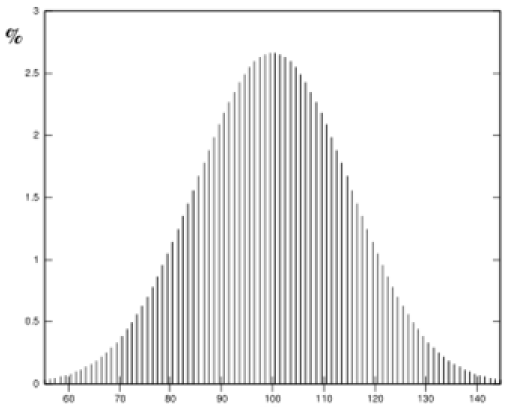
Ejercicio 4: Suponiendo que se contesta completamente al azar a un examen de 10 preguntas de verdadero/falso y de que se corrige puntuando con un 1 los aciertos y con un 0 los errores, obtener la probabilidad de que se saque un 5 en el examen; ¿y cuál es la probabilidad de sacar un 10?; ¿y la de sacar un 5 o más? Obtener el valor esperado de la variable "puntuación en el examen" e interpretarla.

2. La distribución o curva normal

• Se trata de un modelo teórico de distribución de probabilidad para variables aleatorias cuantitativas que se caracteriza, gráficamente, por tener forma similar a la de una campana. Por ello, y por haber sido estudiada inicialmente por el matemático Karl Gauss, se le denomine también como curva o campana de Gauss.

• La importancia de esta distribución reside en el hecho de que diversas variables, como los caracteres fisiológicos y morfológicos de individuos —altura, peso o longevidad—, atributos sociológicos, psicológicos y, en general, variables que vienen determinados por muchos factores, se distribuyen según el modelo de la curva normal.

• Abajo se muestra la representación gráfica de una variable aleatoria que se distribuye según el modelo teórico de la distribución normal (más usualmente dicho, "que se distribuye normalmente").



Como puede observarse, algunas de las características distintivas de este tipo de distribución son:

(1) es unimodal;

(2) es simétrica, situándose el eje de simetría sobre el valor de la media (mediana, moda) de la distribución de la variable;

(3) es asintótica por ambos lados de la distribución;

(4) hace corresponder valores de probabilidad altos para los valores centrales de la variable, mientras que esas probabilidades decaen de forma progresiva a medida que nos alejamos del centro de la distribución, más aceleradamente en la zona central, menos en los extremos.

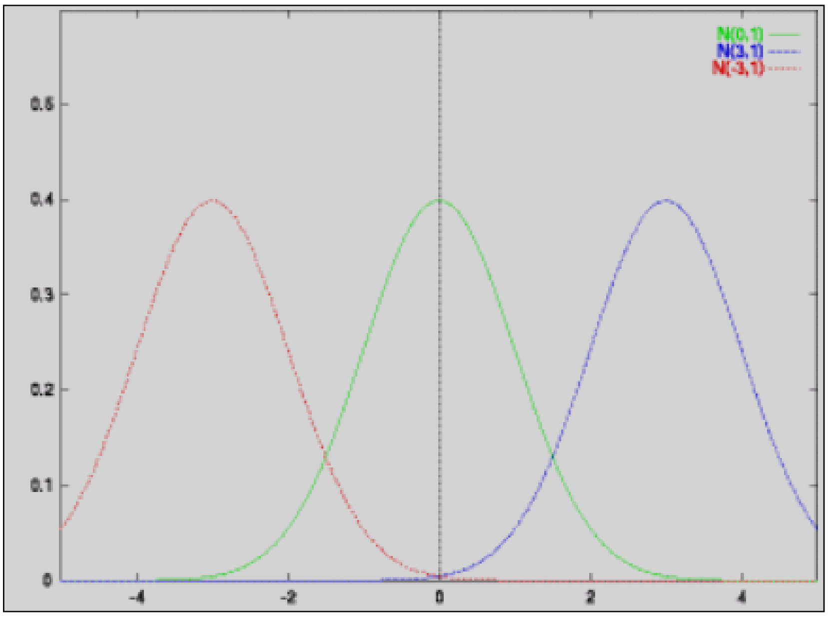
• La distribución de probabilidad (función de densidad de probabilidad) de la curva normal viene definida matemáticamente por la siguiente función matemática, originalmente planteada por Abraham de Moivre en 1733:

$$P(X_i) = \frac{1}{2{,}507 \cdot \sigma_X} \cdot e^{-0{,}5 \cdot \frac{(X_i - \mu_X)^2}{\sigma_X^2}}$$

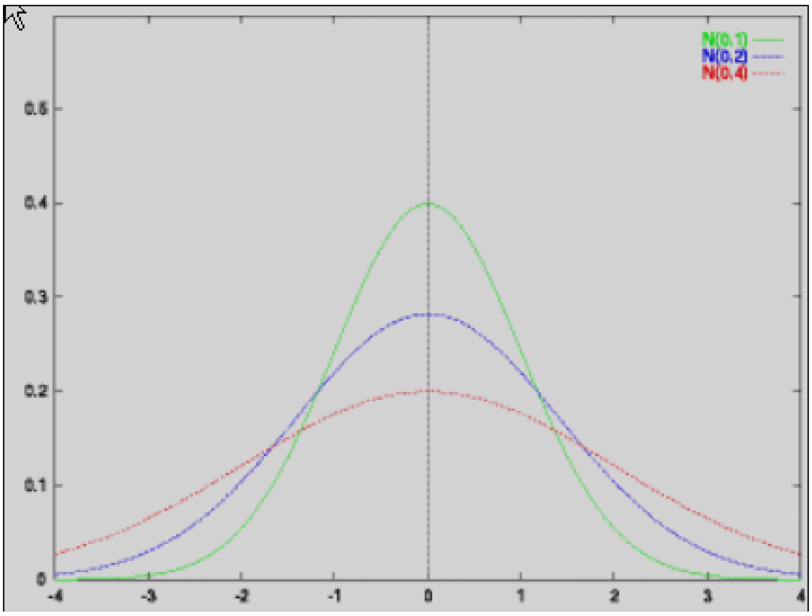
donde X_i es un valor concreto de la variable X, e es una constante matemática (=2,72), y μ y σ son dos parámetros de la función que se corresponden, precisamente, con la media y la desviación típica de X.

• Teniendo en cuenta la fórmula presentada, la distribución normal puede adoptar diferentes formas, tantas como distintos valores de μ y σ se consideren (o sea, infinitas). Todos estos modelos integran la conocida como familia de la distribución normal y para representar, a nivel simbólico, a cada uno de los miembros de esa familia se utiliza la expresión $N(\mu; \sigma)$. A continuación se muestra la representación gráfica de diversos modelos de la familia de la distribución normal en los que se han considerado distintos valores de μ y σ:

- Ejemplos de modelos de distribución normal con distinto valor de μ pero el mismo de σ:



- Ejemplos de modelos de distribución normal con el mismo valor de μ pero distinto de σ:



• Entre los modelos de la familia de la distribución normal, el más relevante en la práctica es el denominado como distribución normal estándar o unitaria, esto es, el modelo de la familia de la distribución normal que tiene $\mu = 0$ y $\sigma = 1$ [$X \rightarrow N(0; 1)$]. Así, es común que en los libros de estadística se incluya en un apéndice final de tablas estadísticas, la correspondiente a la curva normal estándar. Aunque hay variaciones en la forma en que se presenta esta tabla en los libros, es habitual que en la misma podamos consultar, para un rango de valores comprendido habitualmente entre -3 y 3, cuál es el valor de probabilidad y de probabilidad acumulada correspondiente al valor que queramos. En suma, se trata de una representación tabular de la función de probabilidad o de la función de distribución (o de ambas, como en el fragmento que se muestra más abajo) de la curva normal con media 0 y desviación típica igual a 1. Así, si se tiene una variable X que se distribuye según la distribución normal estándar [$X \rightarrow N(0; 1)$], en esta tabla podemos consultar para distintos valores de X, cuál es el valor de probabilidad [$P(X)$] ("*y ordenada*" en la tabla) y de probabilidad acumulada [$P_a(X)$] ("*Area*" en la tabla) correspondiente.

• Una vez haya sido asumido que una determinada variable se distribuye según el modelo de la distribución normal, de forma inmediata se puede dar respuesta fácilmente a diferentes tipos de preguntas como, por ejemplo, la probabilidad asociada a un determinado valor, el porcentaje acumulado o percentil correspondiente a ese valor, el nº de sujetos que es de esperar que tengan un valor inferior o igual a ése, o superior a ése, o entre dos valores determinados, etc. (ver preguntas del ejercicio siguiente).

Fragmento de la tabla de la curva normal estándar

z	Área	y ordenada	z	Área	y ordenada
0.31	.6217	.3802	0.76	.7764	.2989
0.32	.6255	.3790	0.77	.7794	.2966
0.33	.6293	.3778	0.78	.7823	.2943
0.34	.6331	.3765	0.79	.7852	.2920
0.35	.6368	.3752	0.80	.7881	.2897
0.36	.6406	.3739	0.81	.7910	.2874
0.37	.6443	.3725	0.82	.7939	.2850
0.38	.6480	.3712	0.83	.7967	.2827
0.39	.6517	.3697	0.84	.7995	.2803
0.40	.6554	.3683	0.85	.8023	.2780
0.41	.6591	.3668	0.86	.8051	.2756
0.42	.6628	.3653	0.87	.8078	.2732
0.43	.6664	.3637	0.88	.8106	.2709
0.44	.6700	.3621	0.89	.8133	.2685
0.45	.6736	.3605	0.90	.8159	.2661
0.46	.6772	.3589	0.91	.8186	.2637
0.47	.6808	.3572	0.92	.8212	.2613
0.48	.6844	.3555	0.93	.8238	.2589
0.49	.6879	.3538	0.94	.8264	.2565
0.50	.6915	.3521	0.95	.8289	.2541
0.51	.6950	.3503	0.96	.8315	.2516
0.52	.6985	.3485	0.97	.8340	.2492
0.53	.7019	.3467	0.98	.8365	.2468
0.54	.7054	.3448	0.99	.8389	.2444
0.55	.7088	.3429	1.00	.8413	.2420
0.56	.7123	.3410	1.01	.8438	.2396
0.57	.7157	.3391	1.02	.8461	.2371
0.58	.7190	.3372	1.03	.8485	.2347
0.59	.7224	.3352	1.04	.8508	.2323
0.60	.7257	.3332	1.05	.8531	.2299
0.61	.7291	.3312	1.06	.8554	.2275
0.62	.7324	.3292	1.07	.8577	.2251
0.63	.7357	.3271	1.08	.8599	.2227
0.64	.7389	.3251	1.09	.8621	.2203
0.65	.7422	.3230	1.10	.8643	.2179
0.66	.7454	.3209	1.11	.8665	.2155
0.67	.7486	.3187	1.12	.8686	.2131
0.68	.7517	.3166	1.13	.8708	.2107
0.69	.7549	.3144	1.14	.8729	.2083
0.70	.7580	.3123	1.15	.8749	.2059
0.71	.7611	.3101	1.16	.8770	.2036
0.72	.7642	.3079	1.17	.8790	.2012
0.73	.7673	.3056	1.18	.8810	.1989
0.74	.7704	.3034	1.19	.8830	.1965
0.75	.7734	.3011	1.20	.8849	.1942

• De modo análogo a lo que se dijo para la distribución binomial, una tabla completa de la distribución normal estándar puede encontrarse en el apéndice de tablas de la mayoría de los libros de estadística, o bien, no resulta difícil encontrar aplicaciones para ordenador que nos permitan la obtención informatizada del valor exacto de probabilidad que nos interese. En el caso del programa *MS Excel®*, basta con introducir en una casilla cualquiera la siguiente expresión con el valor de *z* que nos interese:

=DISTR.NORM.ESTAND.N(*z;0*)

Si el "0" que aparece en último lugar dentro del paréntesis lo cambiamos por un "1", obtendremos la probabilidad acumulada, esto es, la probabilidad de obtener el valor que nos interesa (z) o uno inferior.

Si lo que se desea es seguir el camino inverso, esto es, a partir de un valor de probabilidad acumulada, obtener la puntuación típica que deja por debajo el área correspondiente de la curva normal, podemos utilizar la siguiente función de *MS Excel®*:

=INV.NORM.ESTAND(*probabilidad_acumulada*)

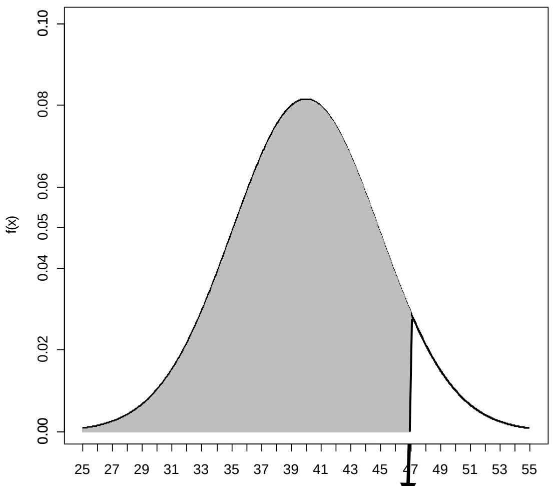
Ejercicio 5: Contesta a las siguientes preguntas relativas a una variable cuantitativa X que para un determinado grupo de sujetos se distribuye según $N(0;1)$ (para cada pregunta se muestra entre corchetes la expresión simbólica de la misma): (a) ¿cuál es la probabilidad acumulada correspondiente a un valor de X igual a 1,18 [$P_a(1{,}18)$]?; (b) ¿qué porcentaje de sujetos tendrán una puntuación inferior o igual a 1,18?; (c) sabiendo que el grupo de sujetos era de 1000 personas (n =1000), ¿cuántos tendrán una puntuación inferior o igual a 1,18?; (d) ¿cuál es la probabilidad de que al extraer un sujeto al azar, éste tenga una puntuación inferior o igual a 1 [$P_a(1)$]?; (e) ¿y de que sea superior a 1 [$1–P_a(1)$]??; (f) ¿y de qué esté entre 1 y 2 [$P(1 \leq X \leq 2)$], sabiendo que $P_a(2) = 0{,}977$? ; (g) ¿y de qué esté entre la media de la distribución y 1 [$P(0 \leq X \leq 1)$]?; (h) ¿a qué valor de la variable X le corresponde una probabilidad acumulada de 0,75 [$P_a(X) = 0{,}75$] (esto es, el 75% de los sujetos obtienen una puntuación inferior o igual a ese valor en la variable => Q_3)?; (i) ¿qué valor de la variable X será sólo superado por el 25% de los sujetos? (j) ¿qué valor corresponde al percentil 25 [$P_a(X) = 0{,}25$]?; (k) ¿cuál es la probabilidad de que al extraer un sujeto al azar, éste tenga una puntuación inferior o igual a -1 [$P_a(-1)$]?; (l) ¿y de que la puntuación sea superior a -1 [$1–P_a(-1)$]?

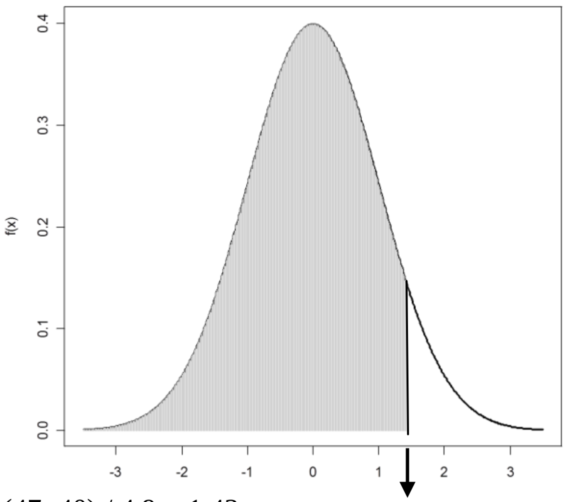
• Una consecuencia práctica derivada de la aplicación del modelo teórico de la distribución normal y, en general, de cualquier modelo teórico de distribución de probabilidad es que, si sabemos o podemos asumir que una variable se distribuye según un modelo teórico, entonces se pueden obtener las probabilidades asociadas a cualquier valor de esa variable y, en consecuencia, la correspondiente distribución de probabilidad. Será suficiente para ello con aplicar la fórmula matemática del modelo de probabilidad correspondiente o, más sencillo, recurrir a una tabla estadística de ese modelo y consultar en la misma los valores que nos interesen.

• Ahora bien, ¿cómo aprovechar la tabla de la curva normal unitaria si tengo una variable que, aunque se pueda asumir que se distribuye normalmente, su media y su desviación típica no son precisamente 0 y 1? La respuesta está en <u>transformar los valores de la variable a puntuaciones típicas (*Z*)</u>, con lo que la variable se seguirá distribuyendo según la curva normal, si bien, pasa a tener media 0 y desviación típica 1, haciendo así factible la utilización de la tabla de la curva normal unitaria.

> **Ejemplo**: Sea una variable *X* que se distribuye según el modelo de la curva normal con media igual a 40 y varianza igual a 24 y supongamos que deseamos saber cuál es la probabilidad de obtener un valor inferior o igual a 47 en esa variable [$P(X \leq 47)$].

$X \rightarrow N(40;\ 4{,}9)$ $\qquad\qquad$ $Z \rightarrow N(0;\ 1)$



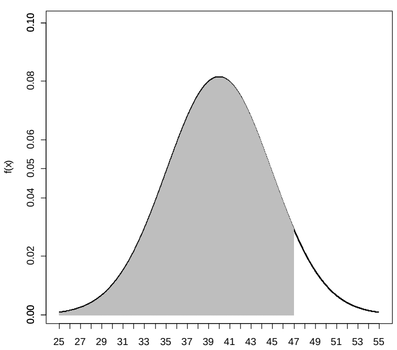


$X_i = 47$ $\quad \xrightarrow{Z_i = (47–40)\ /\ 4{,}9\ =\ 1{,}43} \quad$ $Z_i = 1{,}43$

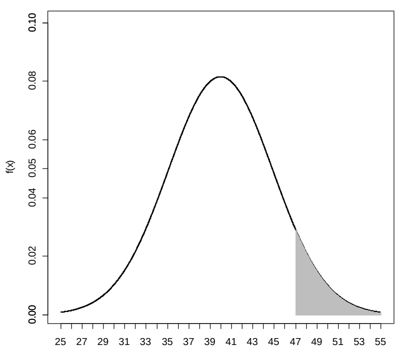
¿$P_a(47)$? $\qquad\qquad$ $P_a(1{,}43) = 0{,}924$

Ejercicio 6: Sea $X \rightarrow N(40;\ 4{,}9)$. Obtener: (a) $P_a(X \leq 43)$; (b) $P_a(X > 47)$; (c) $P(43 \leq X \leq 47)$.

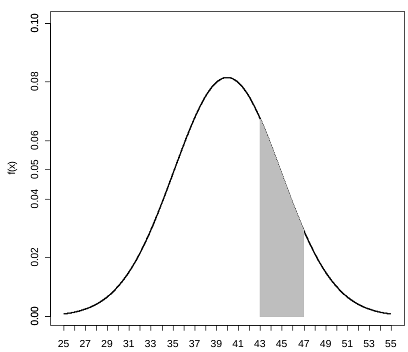
(a)



(b)



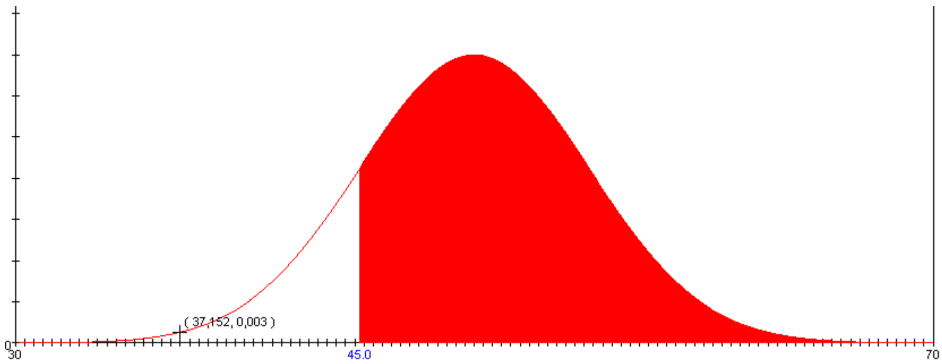
(c)



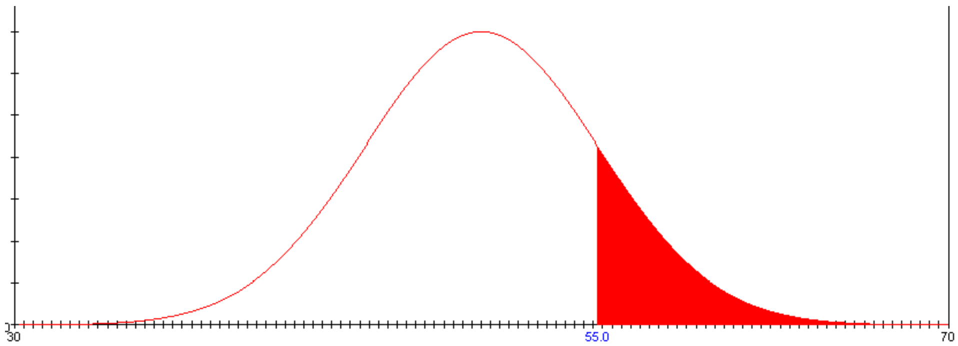
Ejercicio 7. En la población española de neonatos la variable "Talla al nacer" se distribuye normalmente con media= 50 cm y desviación típica igual a 5 cm. Responde a las siguientes preguntas, utilizando las tablas de la distribución normal estandarizada (0,1) o las funciones de Excel de esta distribución:

> **`=DISTR.NORM.ESTAND.N (z;1):`** la función devuelve la probabilidad acumulada para el valor z especificado entre paréntesis.
>
> **`=INV.NORM.ESTAND (prob):`** la función devuelve el valor z al que le corresponde la probabilidad acumulada especificada entre paréntesis.

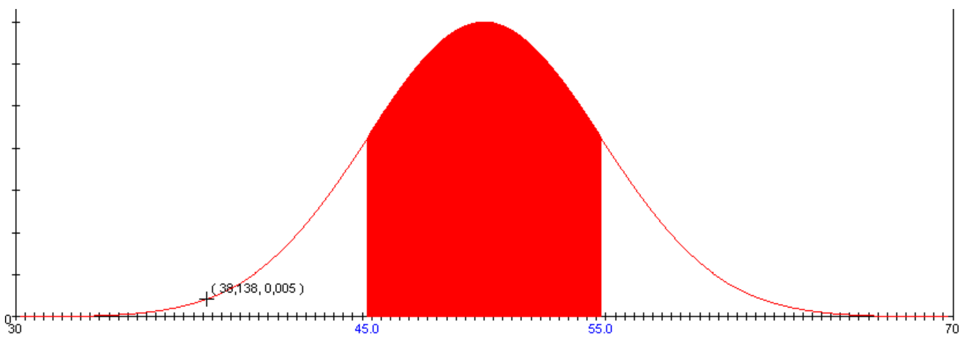
a) ¿Cuál es la probabilidad de que un niño al nacer mida menos de 45 cm? ¿y de que mida más de 45 cm?



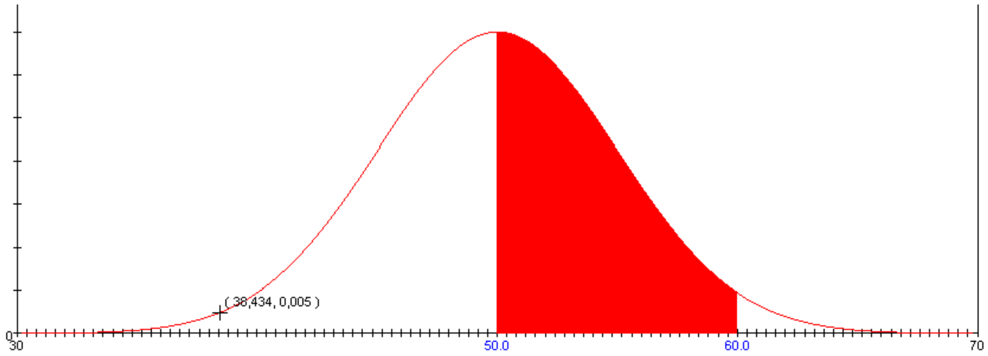
b) ¿Qué porcentaje de niños miden al nacer más de 55 cm?



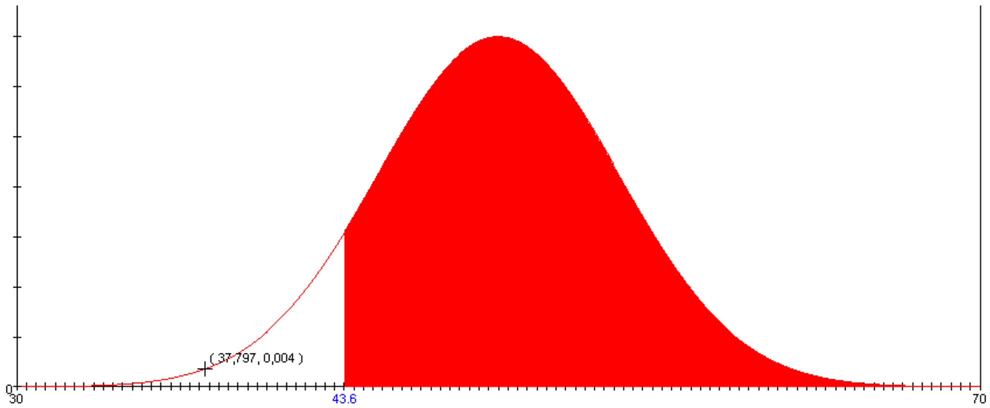
c) ¿Cuál es la probabilidad de que un niño mida al nacer entre 45 y 55 cm? ¿y de que mida menos de 45 cm?



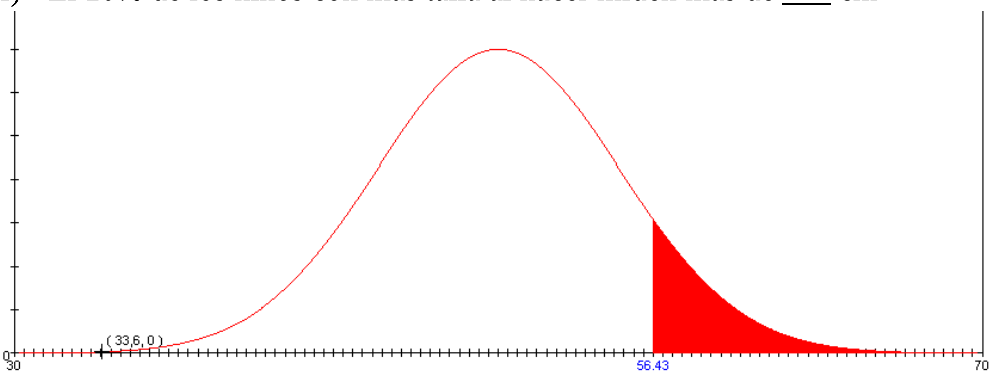
d) ¿Y cuál de que mida entre 50 y 60 cm? ¿y de que mida más de 60 cm?



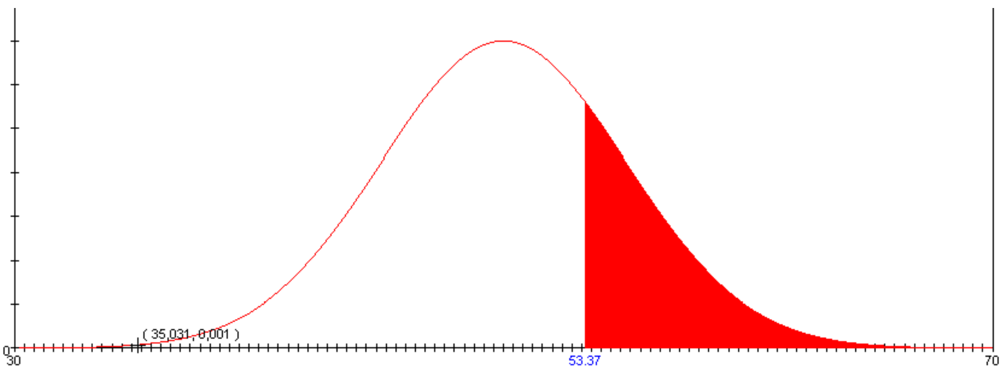
e) El 10% de los niños con menos talla al nacer miden menos de ___ cm



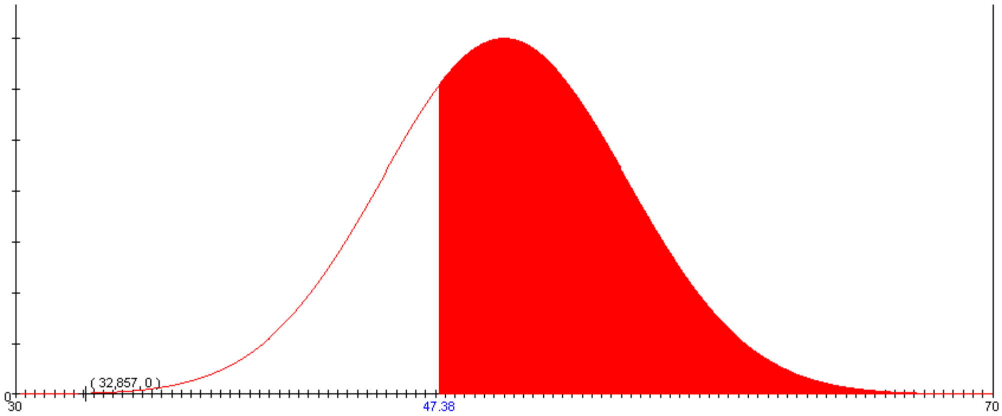
f) El 10% de los niños con más talla al nacer miden más de ___ cm



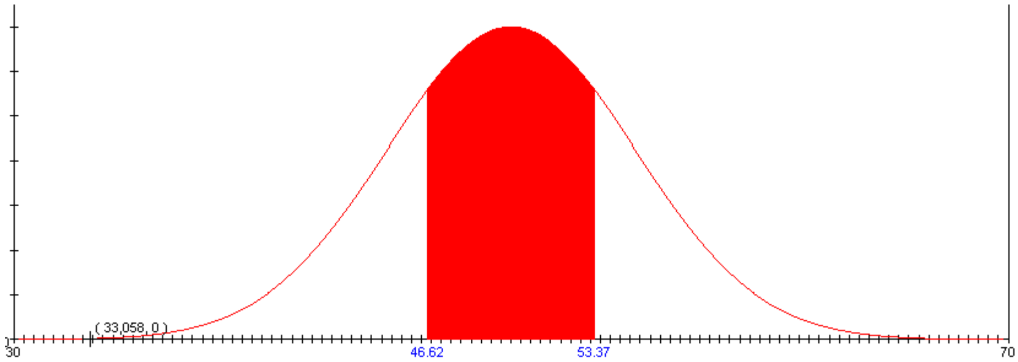
g) El 25% de los niños con más talla al nacer miden más de ___ cm



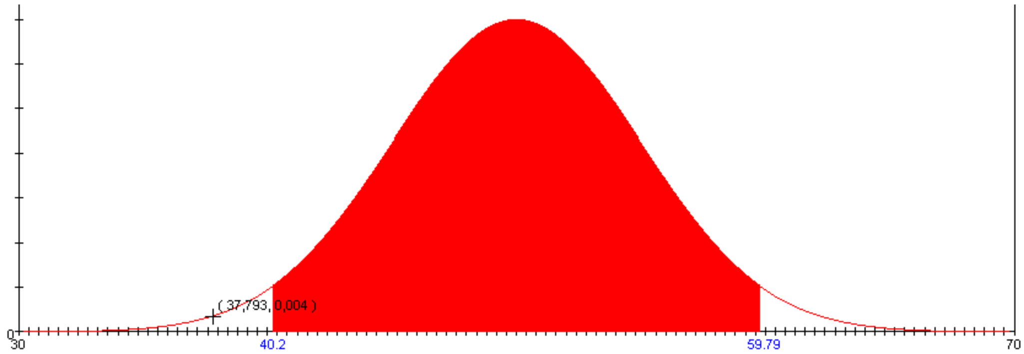
h) ¿A qué valor corresponde el percentil 30 de esta distribución?



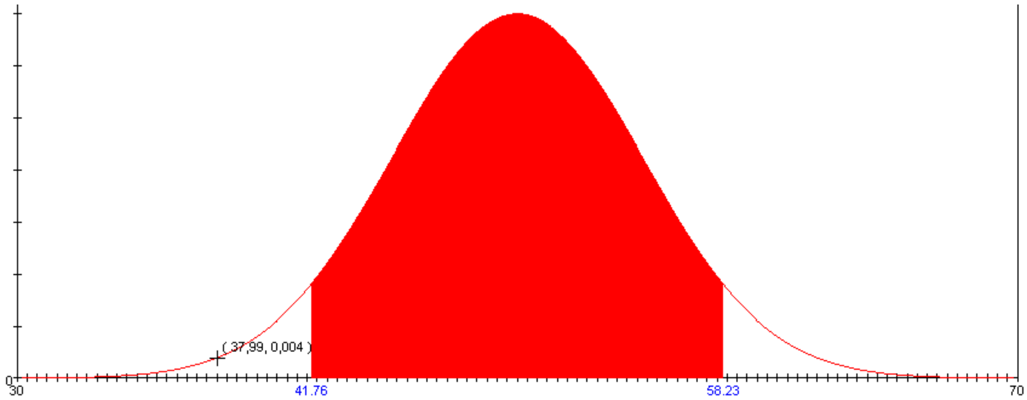
i) El 50% central de los niños miden entre ____ y ____ cm



j) ¿Entre qué puntuaciones típicas (z) se encuentra el 95% de los niños? ¿A qué talla corresponden esos valores?



k) ¿Entre qué puntuaciones típicas (z) se encuentra el 90% de los niños? ¿A qué talla corresponden esos valores?



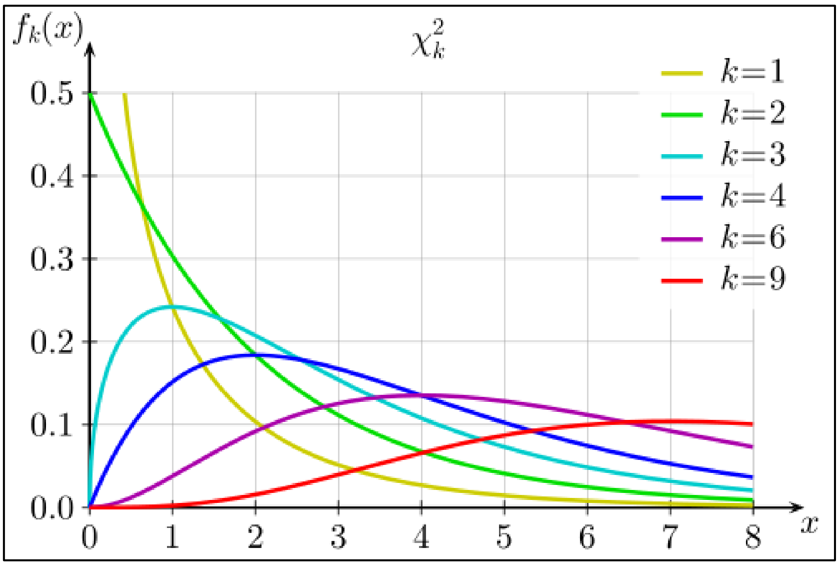
Ejercicio 8. La variable “Peso” en la población de mujeres europeas es N(65;10).
(a) ¿Qué porcentaje de mujeres pesa menos de 50 kilos?, (b) ¿Cuál es probabilidad de que una mujer de esta población pese más de 70 kilos?, (c) ¿Cuál es la probabilidad de que pese entre 60 y 70 kilos?, (d) El 5% de las mujeres con mayor peso, pesan más de ___ kilos, (e) El 5% de las mujeres con menor peso, pesan menos de ___ kilos, (f) ¿Cuál es el Q1 de esta distribución? (g) El 95% central de las mujeres pesan entre ___ y ___ kilos, (h) El 90% central de las mujeres pesan entre ___ y ___ kilos.

Ejercicio 9: Supongamos que un conocido nos dice que ha obtenido en un test de inteligencia una puntuación *CI* igual a 95. Asumiendo que las puntuaciones en un test de inteligencia se distribuyen normalmente y sabiendo que las puntuaciones *CI* tienen media 100 y desviación típica 15, ¿qué le podemos decir acerca de su puntuación?, más concretamente, (a) ¿qué porcentaje de sujetos es de esperar que obtengan un valor inferior o igual a 95?, o (b) ¿qué porcentaje de sujetos es de esperar que obtengan un valor superior a 95? Supongamos también que nos pregunta (c) ¿qué puntuación CI habría que sacar en el test de inteligencia para estar en el 30% inferior? (puntuación de *CI* que deja el 30% de sujetos por debajo); (d) ¿y para estar en el 10% superior? (puntuación de *CI* que es superada solo por el 10% de los sujetos) (e) ¿entre qué valores de *CI* se encuentra el 50% central de los sujetos?

3. Las distribuciones χ^2, *t* y *F*

3.1. La distribución χ^2 (ji-cuadrado) constituye, en realidad, una familia de distribuciones de probabilidad. Cada uno de los miembros de esta familia viene determinado por un parámetro *k* que hace referencia al número de grados de libertad de la distribución. Se suele hacer referencia a cualquier miembro de esta familia de distribuciones con la expresión χ_k^2, donde *k* expresa el número de grados de libertad. Por ejemplo, χ_{18}^2 representa a la distribución χ^2con 18 grados de libertad.

• En la figura inferior aparece reprsentada gráficamente la función de densidad de probabilidad de las distribuciones χ^2 con 1, 2, 3, 4, 6 y 9 grados de libertad. Como puede observarse, la familia de distribuciones χ^2 es asimétrica positiva, si bien, a medida que *k* es mayor, la distribución tiende a ser más simétrica. Para valores de *k* superiores a 50, la distribución χ^2 se puede considerar igual a la distribución normal.



• Los libros de estadística suelen incluir tablas con información acerca de la distribución de probabilidad de diversos de los miembros de la familia de distribuciones χ^2. Podemos obtener los valores correspondientes a la función de densidad de probabilidad para cualquiera de ellas con el programa *MS Excel*® introduciendo en una casilla cualquiera la siguiente expresión con los valores de χ^2 y k que nos interese:

=DISTR.CHICUAD(χ^2;*k*;0)

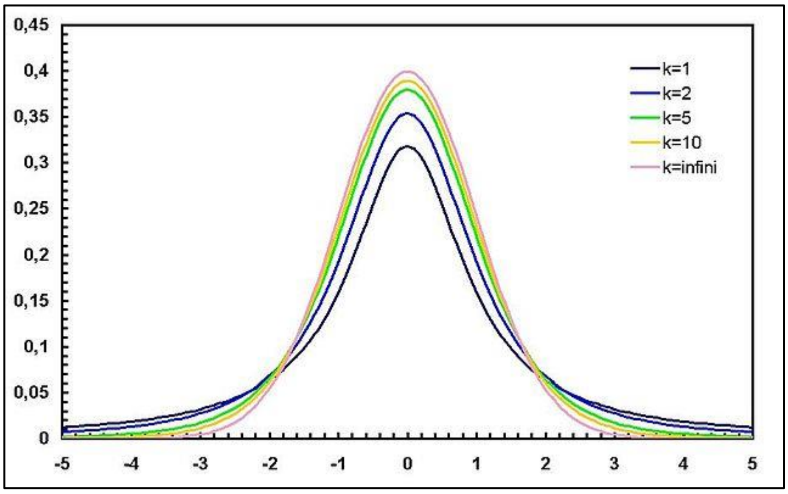
Si el “0” que aparece en último lugar dentro del paréntesis lo cambiamos por un “1”, obtendremos la probabilidad acumulada, esto es, la probabilidad de obtener un valor como χ^2 o inferior.

Inversamente, si lo que se desea es obtener, a partir de un valor de probabilidad acumulada, el valor de la distribución χ^2 al que corresponde esa probabilidad acumulada, podemos utilizar la siguiente función de *MS Excel*®:

=INV.CHIQUAD(*probabilidad_acumulada;k*)

3.2. La <u>distribución *t* de Student</u> representa también a una familia de distribuciones de probabilidad y, como en la distribución χ^2, cada uno de los miembros de esta familia viene determinado por un parámetro k que hace referencia al número de grados de libertad de la distribución. Se suele hacer referencia a cualquier miembro de esta familia de distribuciones con la expresión t_k , donde k expresa el número de grados de libertad. Así, t_{10} representa a la distribución t con 10 grados de libertad.

• En la figura inferior aparece reprsentada la función de densidad de probabilidad de las distribuciones t con 1, 2, 5, 10 y un valor muy grande de grados de libertad. Como puede observarse, la familia de distribuciones t es muy similar a la distribución normal, aunque más platicúrtica; no obstante, a medida que *k* es mayor, la distribución *t* tiende a ser más mesocúrtica y para valores de *k* superiores a 30, las diferencias con la distribución normal se hacen inapreciables.



• Aparte de en el apéndice de algún libro de estadística, podemos obtener fácilmente con el programa *MS Excel*® los valores correspondientes a la función de densidad de probabilidad de cualquier distribución t_k. Para ello, debe introducirse en una casilla cualquiera de la hoja de cálculo la siguiente expresión con los valores de *t* y *k* que nos interese:

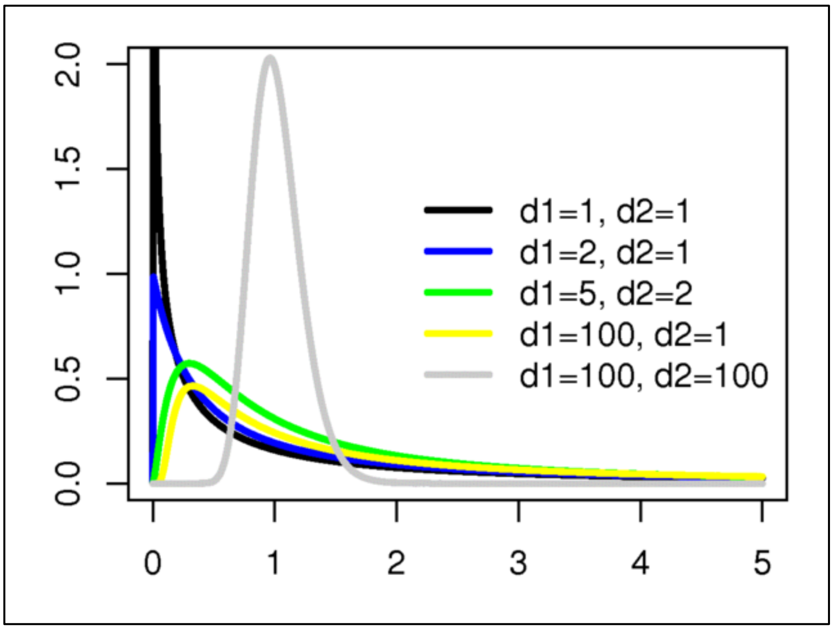
=DISTR.T.N(*t;k;*0)

Si el "0" que aparece dentro del paréntesis lo cambiamos por un "1", obtendremos la probabilidad acumulada, esto es, la probabilidad de obtener el valor de *t* que nos interese o uno inferior.

Inversa a la anterior, si lo que se desea es obtener el valor de la distribución *t* que tiene un determinado valor de probabilidad acumulada, la función de *MS Excel*® a utilizar es:

=INV.T(*probabilidad_acumulada;k*)

3.3. La <u>distribución *F* de Snedecor</u> representa a una familia de distribuciones de probabilidad cuyos miembros vienen caracterizados por dos parámetros, los conocidos como el número de grados de libertad del numerador (*d1*) y el número de grados de libertad del denominador (*d2*). Se suele hacer referencia a cualquier miembro de esta familia con la expresión $F_{d1,d2}$ (por ejemplo, $F_{5,10}$ representa la distribución *F* con 5 y 10 grados de libertad. En la figura inferior aparece reprsentada la función de densidad de probabilidad de las distribuciones $F_{1,1}, F_{2,1}$, $F_{5,2}$, $F_{100,1}$ y $F_{100,100}$.



• Podemos obtener con el programa *MS Excel®* los valores correspondientes a la función de densidad de probabilidad de cualquier distribución $F_{d1,d2}$ a través de la siguiente expresión:

`=DISTR.F.N(F;d1;d2;0)`

Si el “0” lo cambiamos por un “1”, obtendremos la probabilidad acumulada correspondiente al valor de *F* que nos interese. Si lo que se desea es obtener el valor de la distribución *F* que tiene un determinado valor de probabilidad acumulada, la función de *MS Excel®* a utilizar es:

`=INV.F(probabilidad_acumulada;d1;d2)`

Referencias

Barón-López, J. (2005). Bioestadística: métodos y aplicaciones. Apuntes y material disponible en http://www.bioestadistica.uma.es/baron/apuntes/

Botella, J., León, O. G., San Martín, R. y Barriopedro, M. I. (2001). *Análisis de datos en psicología I: teoría y ejercicios.* Madrid: Pirámide.