Curso de Estadística Aplicada a las Ciencias Sociales

Tema 9. Funciones de probabilidad

(Manual: cap. 15, 18.1 y 6.1 ; Cap. 4 de Agresti)

Tema 9. Funciones de probabilidad

Introducción

- ¿Por qué estudiamos probabilidad?
  - Rama de la matemática distinta de la estadística
  - Muy importante por sí misma
  - También es rama auxiliar para la estadística
- Aquí: sólo nociones muy básicas para ayudarnos a entender estadística inferencial (bloque IV del curso)
- Recordatorio estadística inferencial: de muestras a poblaciones
- De estimadores a parámetros

1. La probabilidad

- ¿Qué es la probabilidad?
  - Hipotético **experimento**, basado en una larga secuencia de observaciones repetidas de un **fenómeno aleatorio** (fenómeno cuyo resultado no se conoce de antemano)
  - Cada observación: puede tener varios valores
  - La **probabilidad** de un particular resultado (o valor) es la **proporción** de las veces que ese resultado sucedería en una larga serie de observaciones repetidas
  - La probabilidad de cada valor: la **frecuencia relativa** que tendría ese valor en una serie muy larga de ocurrencias del fenómeno

1. La probabilidad

- Ejemplos de experimentos aleatorios y probabilidades:
  - Una larga serie de veces en los que se tira una moneda al aire
    - Resultado a largo plazo: aparece cara el 0,5 de las veces
    - Probabilidad de que salga cara en una ocasión concreta: 0,5
  - Una larga serie de tiradas de un dado
    - Resultado a largo plazo: cada lado aparece (1/6=)0,1666 de las veces
    - Probabilidad de que salga, por ejemplo, un dos, en una tirada: 0,1666

1. La probabilidad

- Otro experimento aleatorio muy importante en estadística: sacar muestras de una población

- Ejemplo: una población con la mitad de hombres y de mujeres

  - Extraigo muchas muestras de 1.000 personas

  - Las muestras tendrán diferentes proporciones de hombres y mujeres, que oscilarán alrededor de 0,5

  - Valor más probable para una muestra concreta: 0,5

2. Distribución de probabilidad

- Una **variable aleatoria** es una variable que recoge todos los posibles **valores** de un experimento aleatorio
- Como todas las variables, las variables aleatorias pueden resumirse por una distribución de frecuencias, que recibe el nombre, para variables discretas, de **función de probabilidad** (o función de masa) y para variables continuas **de función de densidad**
- Es decir, si x es una **variable aleatoria discreta**, y sus valores son $x_1, x_2,...x_k$, representamos como $p_1, p_2,...p_k$ las correspondientes probabilidades de que ocurran cada uno de los valores. Dicho de otro modo $p_i=P\{x=x_i\}$
- *Cada probabilidad es un número entre 0 y 1, y la suma de todas las probabilidades es 1*

2. Distribución de probabilidad

- Un experimento aleatorio sencillo: extraer un elemento de una población
- La frecuencia relativa de una característica en una población es también la probabilidad de que si extraemos un elemento de esa población, tenga esa característica.
- Por tanto, la distribución de frecuencias de una variable en una población es también la función de probabilidad de la **variable aleatoria discreta** del experimento consistente en extraer un elemento de la población

2. Distribución de probabilidad

- Ejemplo: NPER en fichero Hogares. Distribución de frecuencias es también la función de probabilidad de la **variable aleatoria discreta** producida por el experimento que consiste en extraer una familia de esa población

NPER	Total
1	0,53
2	0,27
3	0,12
4	0,03
5	0,04
6	0,01
(vacías)	0,00
Total general	1,00

2. Distribución de probabilidad

- Otro experimento aleatorio: tirar al aire una moneda cuatro veces
- Veamos todos los posibles resultados:

  CCCC,CCCX,CCXC,CCXX,CXCC,CXCX,CXXC,CXXX

  XCCC,XCCX,XCXC,XCXX,XXCC,XXCX,XXXC,XXXX

- Todos son igualmente probables
- Cada resultado del experimento es un caso de las variable aleatoria
- Representamos una **variable aleatoria discreta**: número de caras que salen

2. Distribución de probabilidad

Valores	N	P
0	1	0,0625
1	4	0,25
2	6	0,375
3	4	0,25
4	1	0,0625
	16	1

- La función de probabilidad propiamente dicha sólo incluye los valores de *p*, porque los valores de "*N*" que hemos utilizado para calcular *p* son ficticios (nos han ayudado a calcular *p*)
- Como con las distribuciones de frecuencias, la función de probabilidad nos permite calcular la probabilidad acumulada de varios sucesos: ¿cuál es la probabilidad de que salgan no más de dos caras?

2. Distribución de probabilidad

- Las variables aleatorias también pueden ser continuas. La distribución de probabilidad de una variable continua se llama **<u>función de densidad</u>**, y como la distribución de frecuencias, requiere que agrupemos los valores por clases.
- Así, la distribución de frecuencias de la variable GTINE es también la **<u>función de densidad</u>** del gasto total de una familia extraída al azar de la población.

x	P(x)
Hasta 100.000	0,1467
100.001 a 200.000	0,2800
200.001 a 300.000	0,2933
300.001 a 400.000	0,0533
400.001 a 500.000	0,1467
500.001 a 600.000	0,0133
600.001 a 700.000	0,0400
700.001 a 800.000	0,0133
800.001 a 900.000	0,0133
	1,0000

2. Distribución de probabilidad

- También serían **variables aleatorias continuas**:
  - El gasto medio de una selección de 3 familias extraídas al azar de la población del fichero hogares.xls
  - la nota media de un grupo de 10 alumnos extraídos al azar de los cientos de alumnos del fichero notas.xls
- Calcular la función de densidad en esos ejemplos sería muy complicado (no lo hacemos)
- Pero podemos imaginar que tendrá unas bajas probabilidades para valores que se aparten mucho de la media en la población, y altas probabilidades para valores que se parezcan a los de la población

2. Distribución de probabilidad

- Aún más: siempre que hagamos un experimento aleatorio que se componga de **muchos sucesos elementales**, las variables aleatorias producidas serán variables continuas, aunque la característica que estemos estudiando en la población sea una variable discreta
- Ejemplo: una población grande con una distribución de la **variable discreta** NPER (número de personas que aportan ingresos al hogar) como la que aparece en el fichero hogares.xls.

NPER	Total
1	0,53
2	0,27
3	0,12
4	0,03
5	0,04
6	0,01
(vacías)	0,00
Total general	1,00

2. Distribución de probabilidad

- Experimento: extraer muestras de 100 familias
- Posibles **variables aleatorias continuas**:
  - "Número de familias con un solo perceptor de rentas". Posibles valores: 0,1,2,3... 100.
  - "Número de familias con dos perceptores de rentas". Posibles valores: 0,1,2,3... 100.
  - "Número medio de perceptores de rentas". Posibles valores: 1; 1,01; 1,02… 5,98; 5,99; 6.
- Siempre que trabajemos con muestras "grandes" tendremos variables aleatorias continuas, aunque la característica en la población sea una variable discreta

2. Distribución de probabilidad

- La función de densidad de la variable aleatoria continua "número de familias con un solo perceptor de renta en una muestra de 100 familias" es un poco complicada de averiguar. No lo vamos a hacer aquí, pero el resultado, agrupando por clases, sería el siguiente (ojo: hay una clase más grande, con 11 elementos):

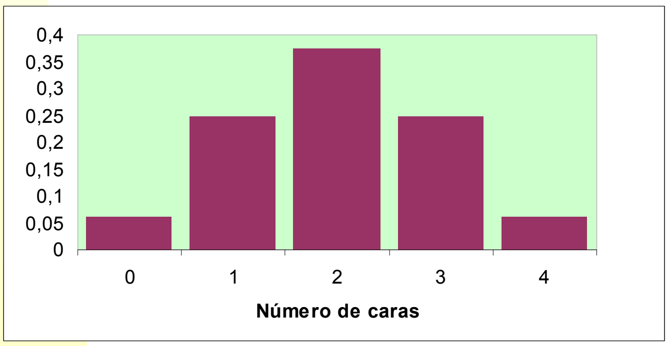
x	P(x)
0 a 9	0,00000000000000000000099449
10 a 19	0,00000000000026362240601399
20 a 29	0,00000100653522911564000000
30 a 39	0,00339883298341082000000000
40 a 49	0,23794411478009400000000000
50 a 59	0,66262726977523900000000000
60 a 69	0,09563339411266160000000000
70 a 79	0,00039536397880665700000000
80 a 89	0,00000001783192258089420000
90 a 100	0,00000000000000001587261712 3

2. Distribución de probabilidad

- Incluso a partir de una variable cualitativa podemos obtener variables aleatorias continuas
- Ejemplo: una población muy grande, tomamos muestras de 1000 personas, y les preguntamos a qué partido van a votar
- Cada muestra de 1.000 personas tendría un número distinto de personas que vota a cada partido
- La variable aleatoria "votantes del PSOE" tendría valores entre 0 y 1000
- La variable aleatoria "votantes del PP" tendría valores entre 0 y 1000
- Y así podríamos hacer una variable aleatoria para cada partido

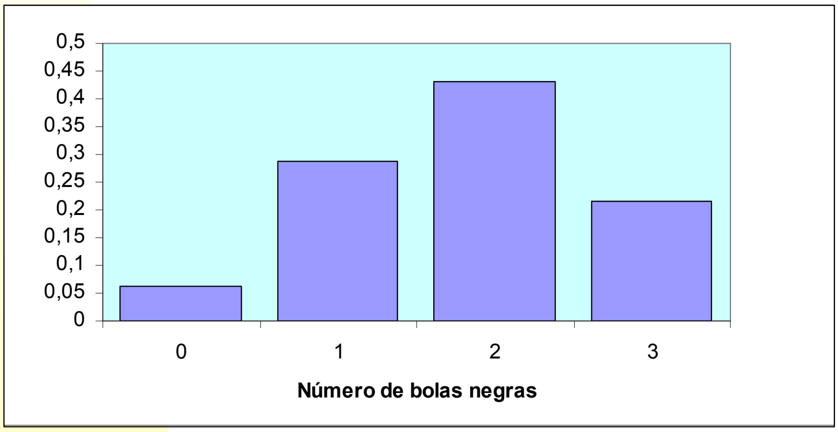
2.1. Representación gráfica

- Al igual que hacíamos con cualquier otra variable, las variables aleatorias pueden representarse también gráficamente
- Por ejemplo, en la transparencia 11 tenemos la función de probabilidad del número de caras que salen al tirar una moneda al aire cuatro veces
- Podemos representar la función de probabilidad con este diagrama de barras



2.1. Representación gráfica

- Podemos hacer lo mismo con todas las **variables discretas**: un diagrama de barras representa la función de probabilidad
- Otro ejemplo: número de bolas negras en muestras de 3 bolas sacadas de una urna con una proporción de 0,6 bolas negras



2.1. Representación gráfica

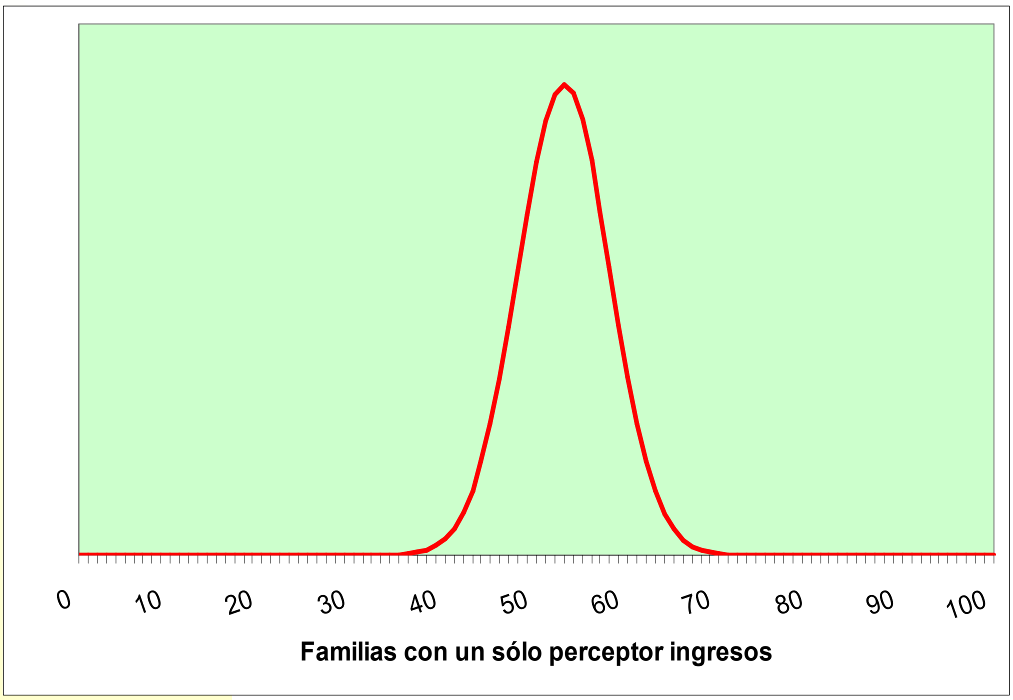
- Pero la mayor parte de las variables aleatorias que nos interesan van a ser variables continuas:
  - todas las que se derivan de variables continuas en la población
  - todas las que se refieren a experimentos aleatorios con fenómenos que contienen muchos sucesos elementales; como por ejemplo muestras de 100 ó 1000 elementos de una población

2.1. Representación gráfica

- La representación gráfica de la función de densidad de las variables aleatorias continuas la hacemos con un histograma “estilizado”: una curva continua y suave que representaría los millones de sucesos elementales posibles
- Ejemplo: función de densidad de la variable aleatoria “número de familias que tienen un solo perceptor de ingresos en una muestra de 100 familias” (si la distribución de NPER en la población es la del fichero hogares.xls) (transparencia 16)

2.1. Representación gráfica

- Esta sería la representación gráfica



2.1. Representación gráfica

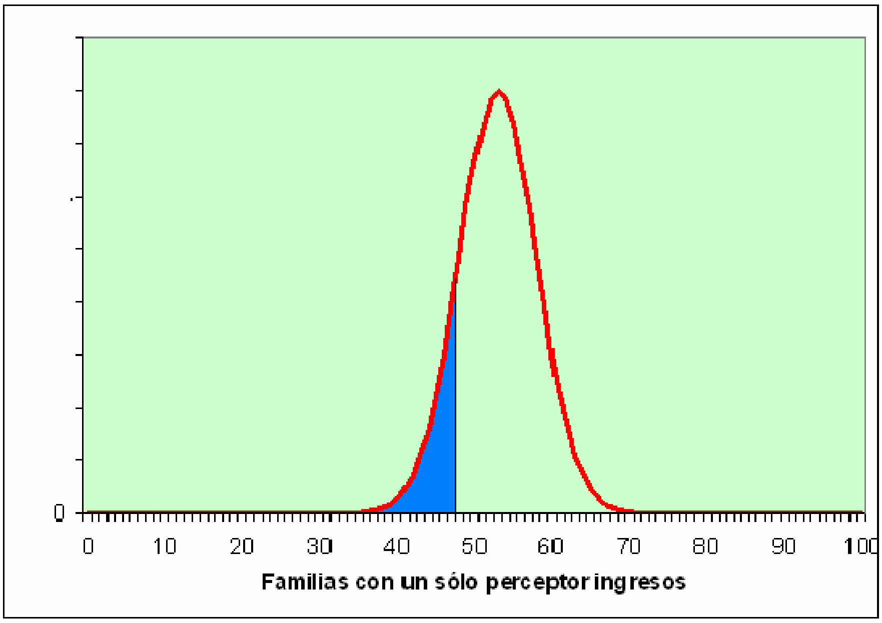
- La función de densidad se interpreta como el histograma o el diagrama de barras: su parte más alta representa los valores más probables
- En el histograma anterior: vemos que habría muy pocas muestras con valores entre 0 y 40 y entre 70 y 100 (comprobar trans. 14); la mayor parte de las muestras tendrían valores entre 50 y 60 aproximadamente.

2.1. Representación gráfica

- No hemos representado los valores en el eje vertical porque lo que importa aquí no es la probabilidad de un valor concreto, sino la probabilidad de un cada rango de valores
- El área bajo la curva representa la suma de las probabilidades de todos los valores (=1); el área bajo la curva entre dos valores representa la probabilidad de que la variable tome un valor en el intervalo definido por esos dos valores

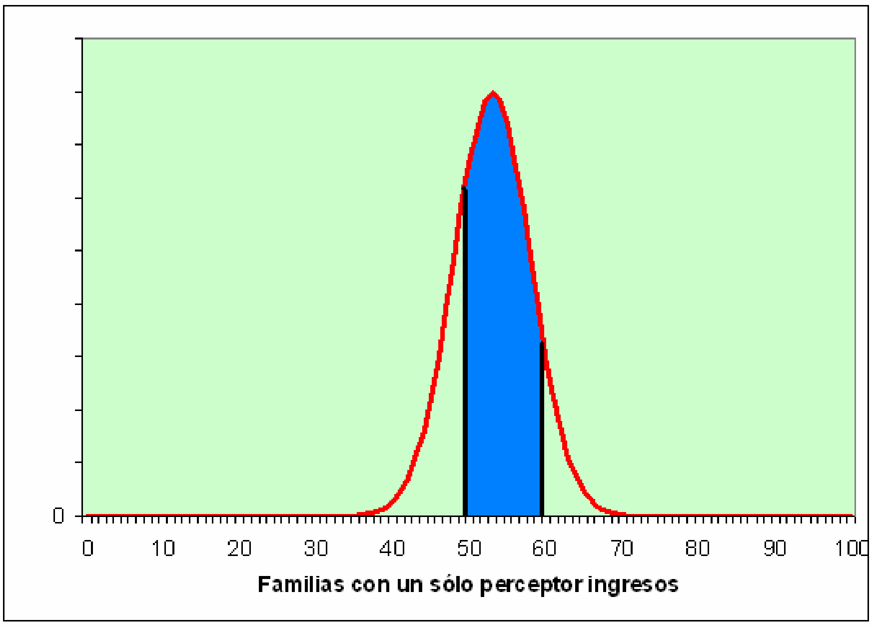
2.1. Representación gráfica

- Ejemplo: el área sombreada, entre los valores 0 y 48 representa el 18,35% del área debajo de la curva; que es la probabilidad de que el valor sea entre 0 y 48 (0,1835)



2.1. Representación gráfica

- Ahora el área sombreada marca la zona entre los valores 50 y 60, y ocupa el 69,27% del área debajo de la curva; esa es la probabilidad (0,6927) de que el valor de la muestra esté entre 50 y 60



2.2. Media o esperanza de una variable aleatoria

- Los mismos conceptos utilizados para resumir información sobre las variables en los temas de estadística descriptiva los podemos usar para variables aleatorias
- La media se calculaba con la fórmula

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{\sum x_i}{N}$$

- Pero si teníamos la distribución de frecuencias de una variable, también podía calcularse con la fórmula

$$\bar{x}_c = \sum c_i f_i$$

2.2. Media o esperanza de una variable aleatoria

- Como la función de probabilidad es una distribución de frecuencias, la media de una variable aleatoria *x* (m_x), se calcula como el sumatorio de los productos de los valores de *x* por sus probabilidades

$$m_x = \sum x_i p_i$$

- Ejemplos: variable aleatoria "número de caras al tirar al aire una moneda 4 veces" (trans. 8). Media es:

$$m_x = (0 \cdot 0{,}0625) + (1 \cdot 0{,}25) + (2 \cdot 0{,}375) + (3 \cdot 0{,}25) + (4 \cdot 0{,}0625) = 2$$

- Interpretación: número esperado medio de caras al tirar al aire una moneda 4 veces

2.2. Media o esperanza de una variable aleatoria

- La variable aleatoria “número de personas que aportan ingresos a una familia extraída al azar” tiene la misma función de probabilidad que la distribución de frecuencia de la variable NPER; por tanto la media será la misma y se puede calcular por el mismo procedimiento (trans. 9)

$$m_x = (1 \cdot 0{,}53) + (2 \cdot 0{,}27) + (3 \cdot 0{,}12) + (4 \cdot 0{,}03) + (5 \cdot 0{,}04) + (6 \cdot 0{,}01) = 1{,}8133$$

2.2. Media o esperanza de una variable aleatoria

- Variables aleatorias continuas:
  - No podemos calcular la media de manera tan fácil (porque tampoco sabemos fácilmente averiguar la función densidad).
  - Pero es posible “intuitivamente” imaginar su valor: el mismo que en la población
- Así, si hago muchas veces el experimento de sacar una persona de una población que tiene la misma distribución de GTINE que conocemos, la media del gasto de todas las personas que saco será la media de GTINE

2.2. Media o esperanza de una variable aleatoria

- Si obtengo muchas muestras de 1000 elementos de la población que tienen una distribución de NPER como en hogares.xls, la media de NPER en la variable aleatoria será 1,8133
- Si obtengo muchas muestras de 1000 elementos de la población que tiene una distribución de NPER como en hogares.xls la media de la variable "número de hogares con un solo perceptor de ingresos" será 530 (0,53*1000)

2.3. Desviación típica, mediana, cuartiles

- Como siempre, la información que la media aporta sobre una variable (valor central), conviene completarla con la información sobre la dispersión, que nos da la desviación típica
- Para **variables discretas**: A partir de la fórmula para la desviación típica cuando tenemos la distribución de frecuencias

$$s_c = \sqrt{\sum (c_i - \overline{x}_c)^2 f_i}$$

- Podemos elaborar la fórmula para la desviación típica de una variable aleatoria discreta

$$\sigma_x = \sqrt{\sum (x_i - m_x)^2 p_i}$$

2.3. Desviación típica, mediana, cuartiles

- Por ejemplo, la desviación típica de la variable aleatoria "número de caras" al tirar 4 veces una moneda (trans. 8) es

$$\sigma_x = \sqrt{\sum (x_i - m_x)^2 p_i} =$$

$$\sqrt{((0-2)^2 \cdot 0{,}0625) + ((1-2)^2 \cdot 0{,}25) + ((2-2)^2 \cdot 0{,}375) + ((3-2)^2 \cdot 0{,}25) + ((4-2)^2 \cdot 0{,}0625)} =$$

$$\sqrt{(4 \cdot 0{,}0625) + 0{,}25 + 0{,}25 + (4 \cdot 0{,}0625)} = \sqrt{0{,}25 + 0{,}25 + 0{,}25 + 0{,}25} = \sqrt{1} = 1$$

- La desviación típica de la variable aleatoria "número de pesonas que aportan ingresos" en una familia (trans. 9) sería:

$$\sqrt{\begin{matrix}((1-1{,}8133)^2 \cdot 0{,}533) + ((2-1{,}8133)^2 \cdot 0{,}2667) + ((3-1{,}8133)^2 \cdot 0{,}1200) + \\ ((4-1{,}8133)^2 \cdot 0{,}0267) + ((5-1{,}8133)^2 \cdot 0{,}0400) + ((6-1{,}8133)^2 \cdot 0{,}0133)\end{matrix}} = 1{,}2984$$

2.3. Desviación típica, mediana, cuartiles

- **Variables aleatorias contínuas**: el valor de la desviación típica no es tan fácil de calcular (no lo vemos)
- Otros conceptos estudiados para las variables normales se pueden aplicar también a las variables aleatorias
- El coeficiente de variación, que es la desviación típica partida por la media, en una variable aleatoria lo expresamos así

$$CV_x = \frac{\sigma_x}{|m_x|}$$

- Mediana, cuartiles, rango intercuartílico… también sería posible (nosotros no los usamos)

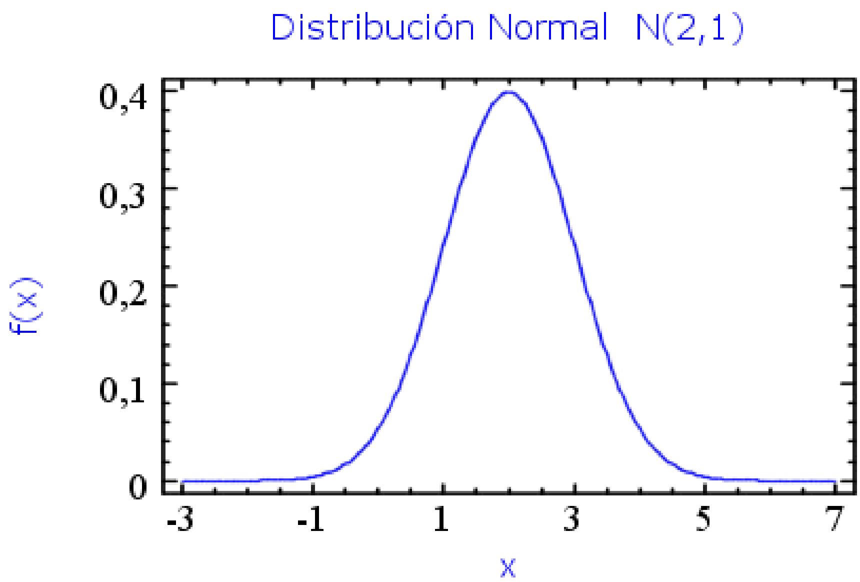
3. La distribución normal

- Existen algunas distribuciones de probabilidad que son especialmente frecuentes: se dan en muchas situaciones
- Algunas son especialmente importantes para la probabilidad (temas 16 y 17 de libro algunos casos: no estudiamos)
- **Distribución normal:** la más importante de todas las distribuciones en estadística
- Modelo para muchas variables que toman valores intermedios con alta probabilidad, y muy poca probabilidad de valores extremos
- Distribución de muchas variables aleatorias es una distribución que se aproxima a la normal
- Punto de partida de muchos métodos de estadística inferencial

3. La distribución normal

- Introducida a principios del siglo XIX por Carl Friedrich Gauss (también llamada "campana de Gauss")
- Originalmente: estudio de distribución de errores de medida
- Modelo para multitud de variables (peso, altura, calificación de examen)
- Características: distribución simétrica unimodal respecto a un valor central, que es a la vez la moda, con frecuencia decreciente a medida que los valores se alejan de la moda, en ambas direcciones, y muy pocos valores extremos
- No cualquier distribución con esas características: una con una "función" (ecuación) particular

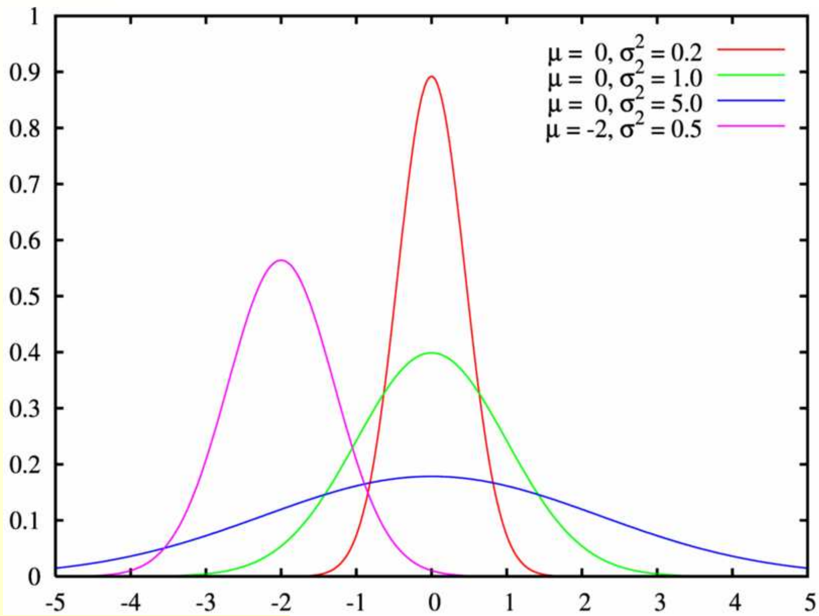
3. La distribución normal



3. La distribución normal

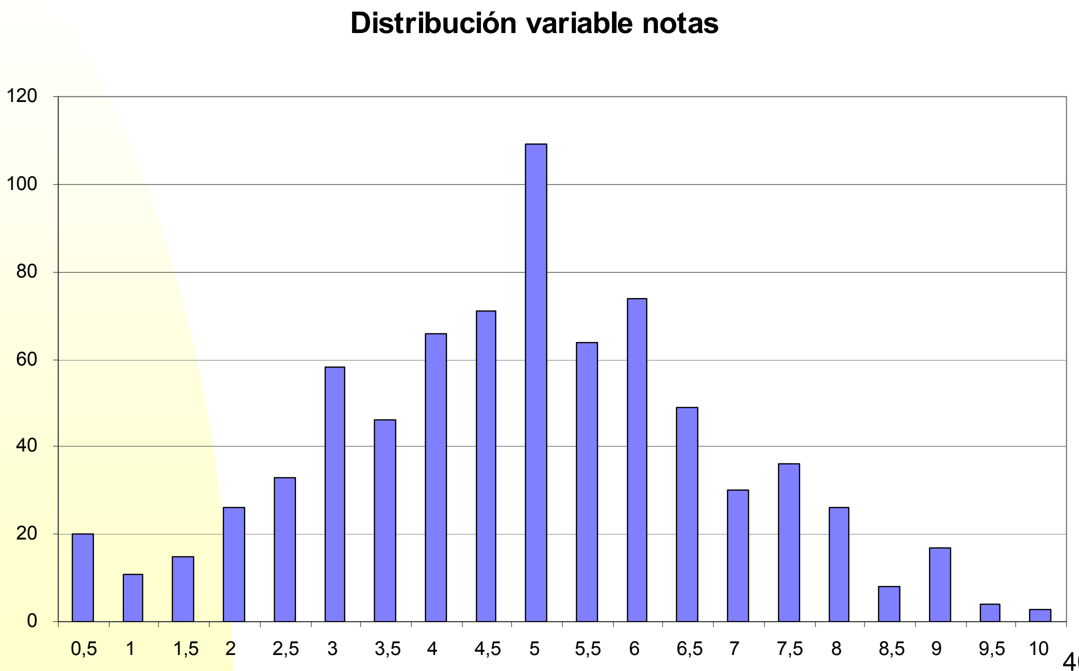
- La distribución normal de una variable *x* se define por dos parámetros,
  - *m* es la media o valor esperado de la variable (puede tomar cualquier valor)
  - σ es la desviación típica (sólo puede tomar valores positivos)
- Una distribución normal puede representarse por *N(m,σ)* (ver transparencia anterior)
- Media = mediana = moda (*m*)

3. La distribución normal



3. La distribución normal

Ejemplo variable "real" con distribución aproximadamente normal



3. La distribución normal

- Por sus propiedades: la mayor parte de los casos tienen valores que están en la "zona central" de la distribución, cercanos al valor de la media
- Más precisamente, si la distribución es justamente la distribución normal, se aplica de manera exacta lo que en la "regla empírica" (tema 5) decíamos que, de manera aproximada, sucedía en las variables con histograma en forma de campana

3. La distribución normal

- Exactamente el 68,3% de los datos caen entre

$$m \pm \sigma$$

- Exactamente el 95,5% de los datos caen entre

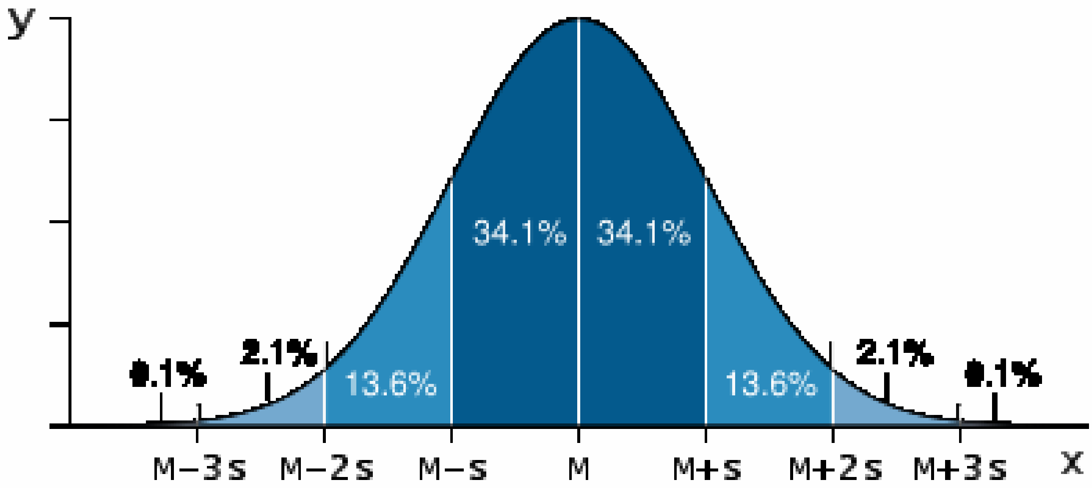
$$m \pm 2\sigma$$

- Exactamente el 99,7% de los datos caen entre

$$m \pm 3\sigma$$

- Por tanto, en consecuencia, sólo 0,3% de los datos tienen un valor que está a más de tres veces la desviación típica de la media

3. La distribución normal



3. La distribución normal

- Dicho en términos de probabilidad:
  - La probabilidad de que un caso de una variable con distribución normal tenga un valor entre $m-1\sigma$ y $m+1\sigma$ es 0,682
  - La probabilidad de que un caso de una variable con distribución normal tenga un valor entre $m-2\sigma$ y $m+2\sigma$ es 0,955
  - La probabilidad de que un caso de una variable con distribución normal tenga un valor entre $m-3\sigma$ y $m+3\sigma$ es 0,997

3. La distribución normal

- Para cualquier variable con distribución normal, si sabemos la media y la desviación típica podemos calcular la probabilidad de obtener al azar un caso con valores que se aparten 1, 2 ó 3 veces de la desviación típica
- Ejemplo: si altura media de los españoles varones adultos es 1,75m y desviación típica es de 0,05 m (datos imaginarios), y esta variable tiene una distribución normal, un varón español adulto escogido al azar
  - Tiene una probabilidad de 0,683 de medir entre 1,70 y 1,80
  - Tiene una probabilidad de 0,95 de medir entre 1,65 y 1,85
  - Tiene una probabilidad de 0,997 de medir entre 1,60 y 1,90

4. Variables tipificadas

- En el cap. 6 del manual se estudian (nosotros no lo vimos) algunas "transformaciones" de variables
- Pero hay una transformación especial que puede hacerse con una variable, que llamamos "tipificar" la variable (también se puede decir "estandarizar"), que sí nos interesa

4. Variables tipificadas

- Dadas unas observaciones $x_1, x_2, ..., x_n$
- Las observaciones tipificadas se construyen restando a todos los datos la media y dividiendo por la desviación típica

$$z_1 = \frac{x_1 - \bar{x}}{s_x}, z_2 = \frac{x_2 - \bar{x}}{s_x}, \ldots, z_n = \frac{x_n - \bar{x}}{s_x}.$$

- La variable tipificada expresa el número de desviaciones típicas que cada observación dista de la media

4. Variables tipificadas

- Permite comparar posición relativa de datos de diferentes variables
- Ejemplo, comparación de sueldos de un español y un alemán (datos imaginarios)
  - Alemania: sueldo medio es 4.000 € y *s* es 1.000€
  - España: sueldo medio es 2.000€ y *s* es 600€
  - Alemán que gana 4.800€ y español que gana 2.700€
  - ¿Qué sueldo es mejor, en términos comparativos a su país?
  - Datos tipificados son 0,8 y 1,166
  - El sueldo español es comparativamente mejor

4. Variables tipificadas: propiedades

- La media de los datos tipificados es 0
- Lógico, puesto que (tema 5): $\sum (x_i - \bar{x}) = 0$
- Por tanto:

$$\boxed{\bar{z} = \frac{\sum \frac{x_i - \bar{x}}{s_x}}{N} = \frac{\frac{\sum x_i - \bar{x}}{s_x}}{N} = 0}$$

4. Variables tipificadas: propiedades

- La desviación típica de los datos tipificados es 1

$$s_z = \sqrt{\frac{\sum (z_i - \bar{z})^2}{N}} = \sqrt{\frac{\sum (z_i - 0)^2}{N}} =$$

$$= \sqrt{\frac{\sum z_i^{\,2}}{N}} = \sqrt{\frac{\sum \left(\frac{x_i - \bar{x}}{s_x}\right)^2}{N}} = \sqrt{\frac{\frac{\sum (x_i - \bar{x})^2}{s_x^2}}{N}} =$$

$$= \sqrt{\frac{\sum (x_i - \bar{x})^2}{N s_x^2}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N \frac{\sum (x_i - \bar{x})^2}{N}}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = \sqrt{1}$$

5. La distribución normal tipificada

- Como cualquier otra variable, las variables con distribución normal también pueden tipificarse
- El resultado es una variable que tiene la **distribución normal tipificada o estándar**
- Es una distribución con media 0 y desviación típica 1
- Suele representarse con la letra *z*
- Una vez tipificadas, todas las variables con distribución normal tienen exactamente la misma distribución

5. La distribución normal tipificada

- Esto quiere decir que para toda variable con distribución normal, una vez tipificada, para cada valor o conjunto de valores de *z*, sabemos exactamente su frecuencia relativa o su probabilidad
- Podemos reformular lo que sabemos sobre la distribución normal así:
  - 0,683 de los casos tienen un valor tipificado *z* entre -1 y +1
  - 0,955 de los casos tienen un valor tipificado *z* entre -2 y + 2
  - 0,997 de los casos tienen un valor tipificado *z* entre -3 y + 3

5.1. Cálculo probabilidades valores *z*

- Pero el valor tipificado *z* no tiene por qué ser un número entero
- Podemos calcular la probabilidad de cualquier valor de *z,* aunque no sea entero
- Hoy día lo hacemos con ordenadores
- También se puede hacer (antes siempre así) con tablas publicadas en muchos lugares con la frecuencia relativa o probabilidad de diferentes valores de *z* en la distribución normal estándar
- Muchos tablas ligeramente diferentes: todas la misma información

5.1. Cálculo probabilidades valores *z*

- Ejemplo: Tabla en archivo pdf adjunto a este tema
- Significado : frecuencia relativa acumulada de los diferentes valores de *z;* o probabilidad de que la variable *z* tome un valor menor o igual a una cierta cantidad, es decir $P\{z \leq b\}$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	**0,64058**	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490

5.1. Cálculo probabilidades valores z

Cómo interpretar la tabla:

Columna de la izquierda (permite elegir fila): valor de *z* con un decimal

Columnas sucesivas (permite elegir columna): valor del segundo decimal de *z*

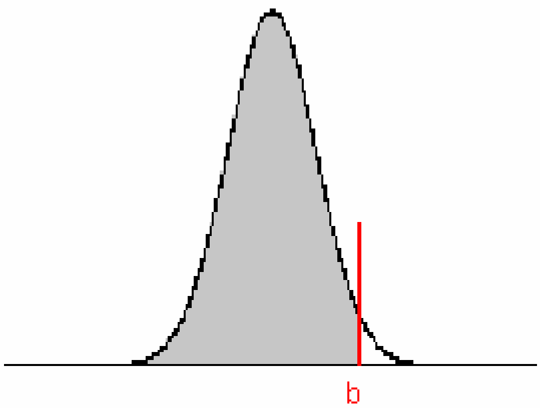
Ejemplo: para *z* 0,36 escogemos la fila 0,3 y la columna 0,06

Valor de la probabilidad: 0,64058

Interpretación: la frecuencia relativa acumulada del valor de *z* 0,36; o la probabilidad de que el valor de *z* sea menor o igual a 0,36 es 0,64058

5.1. Cálculo probabilidades valores z

- Cuando calculamos la probabilidad de que z sea igual o menor a b calculamos el área de la curva que queda “a la izquierda del valor b”



5.1. Cálculo probabilidades valores z

- La misma tabla presentada de forma ligeramente distinta
- Fuente: http://www.digitalreview.com.ar/distribucionnormal/

X	Área	X	Área	X	Área	X	Área	X	Área	X	Área	X	Área	X	Área
-4,00	0,0000	-3,00	0,0013	-2,00	0,0228	-1,00	0,1587	0,00	0,5000	1,00	0,8413	2,00	0,9772	3,00	0,9987
-3,99	0,0000	-2,99	0,0014	-1,99	0,0233	-0,99	0,1611	0,01	0,5040	1,01	0,8438	2,01	0,9778	3,01	0,9987
-3,98	0,0000	-2,98	0,0014	-1,98	0,0239	-0,98	0,1635	0,02	0,5080	1,02	0,8461	2,02	0,9783	3,02	0,9987
-3,97	0,0000	-2,97	0,0015	-1,97	0,0244	-0,97	0,1660	0,03	0,5120	1,03	0,8485	2,03	0,9788	3,03	0,9988
-3,96	0,0000	-2,96	0,0015	-1,96	0,0250	-0,96	0,1685	0,04	0,5160	1,04	0,8508	2,04	0,9793	3,04	0,9988
-3,95	0,0000	-2,95	0,0016	-1,95	0,0256	-0,95	0,1711	0,05	0,5199	1,05	0,8531	2,05	0,9798	3,05	0,9989
-3,94	0,0000	-2,94	0,0016	-1,94	0,0262	-0,94	0,1736	0,06	0,5239	1,06	0,8554	2,06	0,9803	3,06	0,9989
-3,93	0,0000	-2,93	0,0017	-1,93	0,0268	-0,93	0,1762	0,07	0,5279	1,07	0,8577	2,07	0,9808	3,07	0,9989
-3,92	0,0000	-2,92	0,0018	-1,92	0,0274	-0,92	0,1788	0,08	0,5319	1,08	0,8599	2,08	0,9812	3,08	0,9990
-3,91	0,0000	-2,91	0,0018	-1,91	0,0281	-0,91	0,1814	0,09	0,5359	1,09	0,8621	2,09	0,9817	3,09	0,9990
-3,90	0,0000	-2,90	0,0019	-1,90	0,0287	-0,90	0,1841	0,10	0,5398	1,10	0,8643	2,10	0,9821	3,10	0,9990
-3,89	0,0001	-2,89	0,0019	-1,89	0,0294	-0,89	0,1867	0,11	0,5438	1,11	0,8665	2,11	0,9826	3,11	0,9991
-3,88	0,0001	-2,88	0,0020	-1,88	0,0301	-0,88	0,1894	0,12	0,5478	1,12	0,8686	2,12	0,9830	3,12	0,9991
-3,87	0,0001	-2,87	0,0021	-1,87	0,0307	-0,87	0,1922	0,13	0,5517	1,13	0,8708	2,13	0,9834	3,13	0,9991
-3,86	0,0001	-2,86	0,0021	-1,86	0,0314	-0,86	0,1949	0,14	0,5557	1,14	0,8729	2,14	0,9838	3,14	0,9992

5.1. Cálculo probabilidades valores z

Otras tablas dan, al revés, la probabilidad de que el valor sea mayor que z:

Elegimos fila: valor de *z* con un decimal

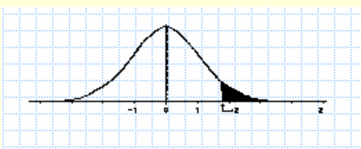
Elegimos columna: valor del segundo decimal de *z*

Ejemplo: para *z* 0,36 escogemos la fila 0,30 y la columna 0,06

Valor de la probabilidad: 0,3594

Interpretación: la probabilidad de un valor z mayor que 0,36 es 0,3594

(Fuente:http://www.unlu.edu.ar/~mapco/apuntes/630/mapco630.htm)



+	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.50	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611

5.1. Cálculo probabilidades valores z

- Algunas tablas dan todos los valores z, negativos y positivos
- Tabla del archivo pdf: desde -4,09 a +4,09
- Otras: sólo la mitad de la tabla, porque las dos mitades son simétricas
- Ejemplo:
  - P (z<-1,73)= 0,04182
  - P (z>1,73)= 0,95818
  - Las dos suman uno
  - Por eso, una tabla que dé sólo la probabilidad de los valores z positivos, permite calcular también los valores negativos

5.1. Cálculo probabilidades valores *z*

- También se puede hacer con Excel
- =DISTR.NORM.ESTAND(z)
- Nos da la frecuencia relativa acumulada del valor z que le damos
  - =DISTR.NORM.ESTAND(0)=0,5
  - =DISTR.NORM.ESTAND(0,3)=0,6179
  - =DISTR.NORM.ESTAND(-0,89)=0,1867
  - =DISTR.NORM.ESTAND(-2,98)=0,0014
  - =DISTR.NORM.ESTAND(0,4678)=0,68003618
- Se pueden comprobar resultados con tablas anteriores

5.1. Cálculo probabilidades valores z

- Por las propiedades de la probabilidad, a partir de las tablas o de los resultados de Excel podemos calcular otros tipos de probabilidades
- Si la tabla (o Excel) nos da la probabilidad de que *z* sea **menor** que un determinado valor, la probabilidad de que z sea **mayor** que ese valor la calculamos restando la primera de uno

$$P\{z>b\}=1-P\{z\leq b\}$$

- Ejemplo: probabilidad de que *z* mayor que 0,36 es 1-0,64058= 0,35942
- Es decir, para toda distribución normal: el 35,94% de los casos están por encima de la media, a una distancia superior a 0,36 veces la desviación típica

5.1. Cálculo probabilidades valores z

- Probabilidad de que *z* esté **entre dos valores cualesquiera**: hallamos la probabilidad de que sea menor o igual al valor mayor y le restamos la de que sea menor o igual al valor menor

$$P\{a<z\leq b\}=P\{z\leq b\}-P\{z\leq a\}$$

- Ejemplo: probabilidad de que *z* esté entre -0,5 y +0,5 es: 0,69146-0,30854= 0,38292
- Es decir: para toda distribución normal, el 38,3% de los casos están hasta 0,5 veces la desviación típica de la media, hacia arriba o hacia abajo

5.1. Cálculo probabilidades valores z

- Podemos calcular en la tabla la probabilidad de que *z* esté entre -1 y +1 y nos dará la misma probabilidad que ya sabemos que en todas las distribuciones normales cae entre la media y una desviación típica arriba y abajo: 0,683
- Efectivamente: probabilidad de *z* menor o igual a 1 es 0,84134 y probabilidad de *z* menor o igual a -1 es 0,15866. La diferencia es: 0,68268
- La probabilidad de que *z* esté entre -2 y +2 es: 0,97725-0,02275= 0,9545
- La probabilidad de que *z* esté entre -3 y +3 es: 0,99865-0,00135= 0,9973

5.2. Probabilidades distribución normal

- Dada **cualquier distribución normal**, entonces, podemos calcular, con tablas u ordenadores, la probabilidad de cualquier rango de valores, convirtiendo los valores que definen el rango en **valores tipificados**
- Si x es una variable normal con media m_x y desviación típica σ_x, la probabilidad de que x tenga un valor igual o menor a b, $P\{x \leq b\}$ será igual a

$$P\{x \leq b\} = P\left\{z \leq \frac{b - m_x}{\sigma_x}\right\}$$

5.2. Probabilidades distribución normal

- Con ejemplo transparencia 45: si la estatura de los varones españoles adultos tiene una distribución normal, con media 175cm y desviación típica 5 cm. ¿Cuál es la probabilidad de que un varón adulto escogido al azar mida 168 cm o menos?

$$P\left\{z \leq \frac{b - m_x}{\sigma_x}\right\} = P\left\{z \leq \frac{168 - 175}{5}\right\} = P\{z \leq -1{,}4\}$$

- En la tabla (o con Excel) vemos que la probabilidad de que z sea menor o igual a -1,4 es 0,08076

- 0,08076 sería la probabilidad de que un varón adulto escogido al azar midiera 168 cm o menos

5.2. Probabilidades distribución normal

- Con la misma lógica: Si *x* es una variable normal con media m_x y desviación típica σ_x, la probabilidad de que *x* tenga un valor **<u>mayor que</u>** *b*, *P{x>b}* será igual a

$$P\{x > b\} = P\left\{z > \frac{b - m_x}{\sigma_x}\right\} = 1 - P\left\{z \leq \frac{b - m_x}{\sigma_x}\right\}$$

- Para el mismo ejemplo de la altura de españoles adultos, probabilidad de que un español mida más de 184 cm

$$1 - P\left\{z \leq \frac{b - m_x}{\sigma_x}\right\} = 1 - P\left\{z \leq \frac{184 - 175}{5}\right\} = 1 - P\{z \leq 1{,}8\} = 1 - 0{,}9641 = 0{,}0359$$

5.2. Probabilidades distribución normal

- Análogamente, podemos calcular la probabilidad de que *x* tenga **un valor entre a y b**, *P{a≤x≤b}*

$$P\{a \le x \le b\} = P\left\{\frac{a-m_x}{\sigma_x} \le z \le \frac{b-m_x}{\sigma_x}\right\} =$$

$$P\left\{z \le \frac{b-m_x}{\sigma_x}\right\} - P\left\{z \le \frac{a-m_x}{\sigma_x}\right\}$$

- Probabilidad de que un varón adulto mida entre 162 y 182 cm

$$P\left\{z \le \frac{b-m_x}{\sigma_x}\right\} - P\left\{z \le \frac{a-m_x}{\sigma_x}\right\} = P\left\{z \le \frac{182-175}{5}\right\} - P\left\{z \le \frac{162-175}{5}\right\} =$$

$$P\{z \le 1{,}4\} - P\{z \le -2{,}6\} = 0{,}9192 - 0{,}0047 = 0{,}9145$$

5.3. Valor *z* para una probabilidad

- La tabla de valores de probabilidades de la distribución normal estándar también se puede utilizar "al revés": partiendo de una probabilidad, buscar el valor de *z*

- Varios métodos estadísticos funcionan de esta manera: partimos de una probabilidad, y queremos hallar el valor de *z*

- Por ejemplo, podemos intentar averiguar cuál es el valor de *z* tal que sólo un 1% de los casos tienen un valor mayor

5.3. Valor *z* para una probabilidad

- Si estamos interesados en buscar sólo el 1% de los valores "más altos" (sólo valores con valor de *z* positivo), buscamos en la tabla la probabilidad 0,99 y encontramos que el valor correspondiente de *z* es 2,33 (exactamente la probabilidad de ese valor de *z* es 0,99010)

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520

5.3. Valor *z* para una probabilidad

- Es decir: en una distribución normal el 1 por ciento de los sujetos con valores más altos se aparta de la media 2,33 veces la desviación típica (o más)
- Dicho de otra manera: en una distribución normal tipificada, el percentil 99 tiene el valor 2,33
- Con Excel:

  =DISTR.NORM.ESTAND.INV(0,99)= 2,326347874

5.3. Valor *z* para una probabilidad

- Si nos interesa calcular el valor *z* del 1% de valores extremos tanto por arriba como por abajo, como la distribución es simétrica, tendremos que buscar el valor de *z* tal que 0,5% de los valores queden por debajo de ese valor con el signo negativo, y 0,5% de los valores queden por encima de ese valor con el signo positivo
- Lo que es lo mismo: buscar el valor de *z* que tiene en la tabla la probabilidad 0,005 y el que tiene la probabilidad 0,995
- Vemos que el valor está en –2,58 (0,00494) en valor negativo; y su simétrico 2,58 (0,99506)
- Excel: =DISTR.NORM.ESTAND.INV(0,005)=-2,5758

Resumen

- Probabilidad
- Variable aleatoria
- Distribución de probabilidad
  - Variables aleatorias discretas y continuas
  - Función de probabilidad y función de densidad
  - Representación gráfica
  - Área del histograma y probabilidad
  - Media de una variable aleatoria
  - Relación de la media con valores en la población
  - Desviación típica de variable aleatoria
  - Otros parámetros de variable aleatoria

Resumen

- La distribución normal
  - Importancia
  - Propiedades
- Variables tipificadas
  - Procedimiento
  - Propiedades
- Distribución normal estándar (valores *z*)
  - Probabilidades de valores *z*
  - Valor *z* para una probabilidad

Ejercicios recomendados

- Del manual:
  - 15.2
  - 15.3
  - 15.4
  - 15.7
  - 18.1 a 18.8

Ejercicios recomendados

- De los exámenes:
  - Feb02: 6
  - Jun02: 7
  - Feb03, Sep03: 6
  - Feb04: 10
  - Jul04: 9
  - Feb05, Jul05: 10,11
  - Feb06: 7
  - Jul06: 8
  - Ene07, Jul07, Ene08, Jun08: 7