Cálculos Estadísticos en Hidrología

Para extraer conclusiones estadísticas a partir de series de datos de precipitaciones o aforos, es necesario disponer de series históricas de más de 20 ó 30 valores, cuanto mayor sea la serie de datos, mayor será la fiabilidad de las deducciones extraídas.

El tratamiento estadístico que veremos aquí está encaminado a solucionar dos tipos de cuestiones:

- **Valor → Probabilidad:** Evaluar la probabilidad de que se presente en el futuro un caudal (o precipitación) mayor o menor que un determinado valor. Por ejemplo: ¿Qué probabilidad hay de que la aportación anual del Tormes en Salamanca supere los 900 Hm^3?
- **Probabilidad → Valor:** Evaluar qué caudal (o precipitación) se superará un determinado % de los años. Por ejemplo: ¿Qué aportación de un río se superará el 10% de los años?

Utilizaremos dos tipos de datos que requieren distintos tratamientos:

- Valores medios. De una serie de años dispondremos del caudal o precipitación medio de cada año.
- Valores extremos. De una serie de años extraemos el caudal o precipitación del día más caudaloso o lluvioso de cada año.

Conceptos básicos

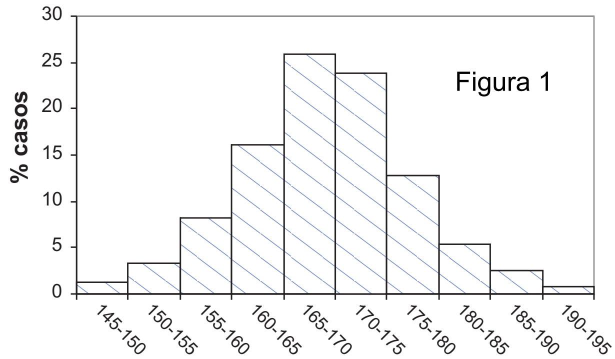
Población y muestra

Población es el conjunto total de datos que queremos estudiar.

A veces disponemos de medidas de toda la **población** estudiada, pero generalmente, esto sería muy difícil (medir la estatura de todos los españoles) o imposible (estudiando el caudal de un río tendríamos que medir los caudales de todos los años pasados y futuros). En estos casos debemos conformarnos con medir una parte de la población (una **muestra**). A partir de la muestra, intentamos extraer estimaciones válidas para toda la población.

Muestra es una pequeña parte de la población que debería ser representativa del total de la población.

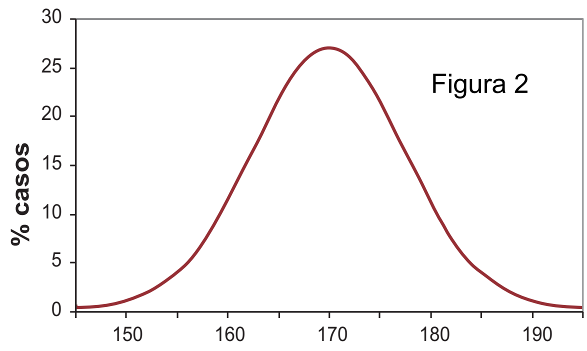
Si yo midiera la estatura de mis alumnos para conocer la estatura media del curso, ellos serían toda la población estudiada. Pero si, a partir de ellos, yo quiero extraer conclusiones sobre la estatura de toda la juventud española, mis alumnos serían solamente una muestra representativa de la población estudiada.



Distribución de los datos

Al observar una serie de datos numéricos (caudales anuales de un río, estatura de personas,...) comprobamos que los valores intermedios son más frecuentes, mientras que los valores elevados o pequeños se presentan con menor frecuencia.

Supongamos que hemos medido la estatura de un gran grupo de personas, y hemos distribuido los valores en grupos de 5 en 5 cm.; después hemos convertido el número de casos de cada intervalo en porcentaje. Su representación gráfica sería similar a la figura 1.

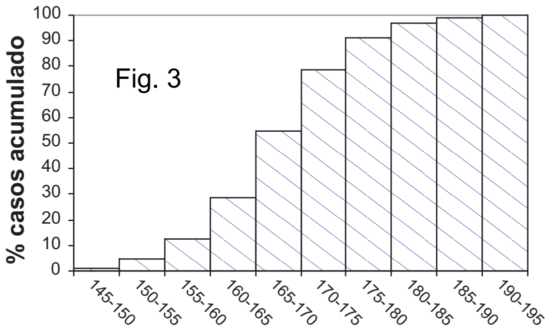


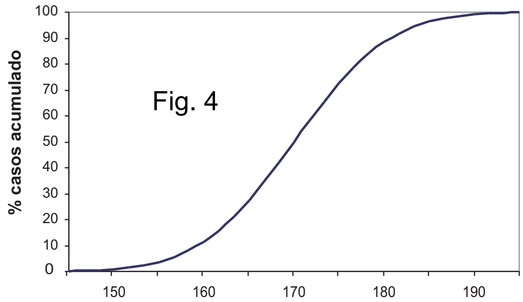
Si en la figura 1 hacemos los intervalos más pequeños y aumentamos el número de valores medidos, el gráfico continuaría con esa forma de campana , pero se suavizaría progresivamente hasta convertirse en una curva continua, como la mostrada en la figura 2.

Una distribución de datos como la figura 2 es muy frecuente en la naturaleza, se denomina *distribución de Gauss* o *normal* . Su ecuación es conocida[1], lo que nos permitirá calcular qué porcentaje se encuentra por encima o por debajo de un valor determinado (por ejemplo, qué porcentaje supera una estatura de 180 cm) .

Función de densidad y función de distribución

Si en lugar de considerar cuantas personas quedan *incluídas **dentro** de cada intervalo* (fig. 1), consideramos cuántas personas quedan *incluídas **hasta** ese intervalo* (es decir, en ese y en todos los anteriores), obtendríamos la figura 3. Y al disminuir el tamaño de los intervalos y aumentar el número de sujetos medidos, obtendríamos la figura 4:





La ecuación de la curva en forma de campana (Figura 2) se denomina *función de densidad*. Si trabajamos con valores acumulados, la ecuación de la curva de la figura 4 es la *función de distribución*.

Si la función de densidad es: *y= f(x)*, la función de distribución será: $F(x) = \int_{-\infty}^{x} f(x) \cdot dx$

Por tanto, las curvas de las figuras 2 y 4 nos proporcionan la misma información, aunque de modo diferente. Vemos un ejemplo: en la figura 5 se comparan las curvas de las figuras 2 y 4, haciendo coincidir sus ejes de abcisas.

En el gráfico acumulado (abajo) leemos que el 90% de esa población tiene una estatura igual o menor que 181 cm, son las coordenadas de un punto de la curva. Para realizar la misma lectura en el gráfico en campana debemos acotar **un área** bajo la curva que sea el 90% del área total bajo la campana (parte rayada en fig. 5, arriba) y leer a qué valor de abcisas corresponde. En ambos casos se obtiene el mismo resultado sobre el eje de abcisas: una estatura de 181 cm.

Estas figuras 2, 4 y 5 representan la ley normal o de Gauss, pero los conceptos explicados (*función de densidad, función de distribución*) son válidos para otros tipos de distribución.

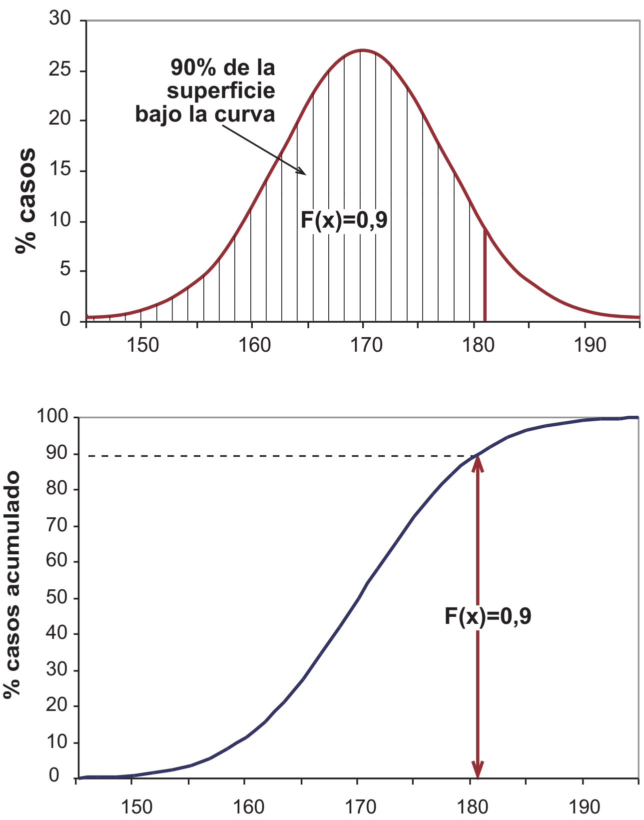


Fig. 5.- Función de densidad y función de distribución

[1] Gauss hizo la justificación matemática en 1809, aunque había sido descrita mucho antes por Moivre en 1733

Distribuciones simétricas y asimétricas

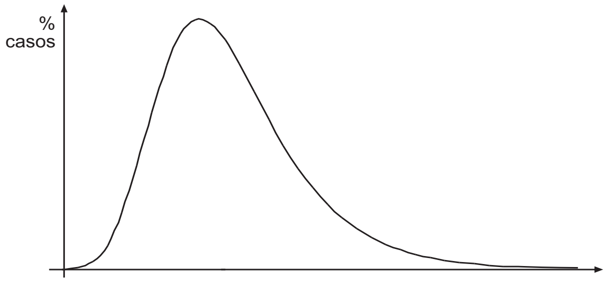


Figura 6.- Distribución asimétrica (esta curva corresponde a la ecuación de Gumbel).

Muchas variables naturales se ajustan a la *distribución simétrica* normal o de Gauss, pero no todas. En ocasiones no hay la misma proporción de valores pequeños que de grandes, eso dará lugar a una *distribución asimétrica*.

Por ejemplo, si representáramos los ingresos de la población de un pais, probablemente la campana no sería simétrica: la riqueza se distribuye con menor equidad que la estatura, y mientras que la proporción de altos y bajos es similar, no así la de ricos y pobres (hay pocos ricos y muchos pobres). Quizá la campana correspondiente tendría una forma similar a la figura 6. Los matemáticos han encontrado las ecuaciones de muchas de estas campanas asimétricas (**Gumbel**, **Pearson III**, etc.). En otras ocasiones, los valores no se ajustan a la distribución de Gauss, pero sus logaritmos sí: se denomina entonces **log-normal**.

En las distribuciones asimétricas el pico de la curva puede estar desviado hacia la izquierda (como la figura 6) o hacia la derecha. Se denominan respectivamente *positivas* y *negativas*, o se dice que tienen *sesgo positivo* o *negativo.*

En **Hidrología**, los valores **medios** (las precipitaciones o caudales anuales) suelen ajustarse a la distribución simétrica de Gauss, pero los valores **máximos**, no: si consideramos el día más caudaloso o el más lluvioso de cada año de una serie larga de años, no se ajustarán la *distribución normal*, sino probablemente a la campana asimétrica descrita por Gumbel o alguna similar.

Media y desviación estándar

Para caracterizar la distribución de un conjunto cualquiera de medidas es necesario disponer de un valor indicativo de su **tendencia central** y otro valor que refleje la **dispersión**: si los valores están apretados o alejados a ambos lados del valor central.

Para indicar la **tendencia central**, normalmente se utiliza la **media aritmética**, tan intuitiva y que todos conocen: sumar valores y dividir por el número de casos. Pero a veces **la media aritmética nos proporciona una información equívoca**: Supongamos que un multimillonario reside en una aldea de 100 vecinos pobres. Los ingresos medios anuales (por persona) de esa aldea serían muy elevados, pero ese valor nos engañaría respecto a la pobreza de la mayoría de los vecinos. En estos casos es más significativa la **mediana**, que es un valor que deja por encima a la mitad de los casos y por debajo a la otra mitad. En las distribuciones simétricas, la media y la mediana coinciden; en las asimétricas positivas (pico a la izquierda) la mediana es inferior a la media.

La **dispersión de los datos** a ambos lados de la media se evalúa mediante la **desviación estándar (o típica)**. La desviación estándar (***s*** ó σ) se calcula en función de la suma de las desviaciones de cada valor (x) de la media previamente calculada ($\overline{x}$), n es el número total de datos:

$$s_n = \sqrt{\frac{\sum (x-\overline{x})^2}{n}} \qquad (1)$$

Por ejemplo, las dos series de datos siguientes tienen la misma media pero obviamente son muy distintas, en la segunda los datos están más dispersos respecto de la media:

							Media	Desv. estándar
19	20	21	23	24	26	28	23	3,02
5	9	17	23	28	35	44	23	13,94

La desviación estándar no sólo nos indica de un vistazo la dispersión de los datos a ambos lados de su media, sino que es específicamente útil para realizar ciertos cálculos que veremos más adelante.

La fórmula anterior se aplica a la población (es decir, si hemos podido medir todos los datos de la población estudiada, y con ellos aplicamos la fórmula). Pero lo habitual es que dispongamos sólo de los datos de una muestra, y la desviación estándar de esa muestra puede no coincidir con la de toda la población; para moderar este error se utiliza este **estimador de la desviación estándar**:

$$s_{n-1} = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} \qquad (2)$$

Cuando el número de datos (*n*) es grande las fórmulas (1) y (2) proporcionan valores casi idénticos. Estas dos fórmulas se incluyen en las calculadoras científicas como σ_n y σ_{n-1}

El cuadrado de la desviación estándar es la **varianza** (s_n^2 , σ_n^2) , y el cuadrado del estimador que utilizamos para las muestras se denomina **quasivarianza** (s_{n-1}^2 , σ_{n-1}^2).

Normalmente se utiliza la notación $\boldsymbol{\sigma}$ (***s*** griega) cuando se ha calculado con los datos de la **población** y se escribe como ***s*** si se ha calculado con una **muestra**. (Análogamente, suele usarse μ para la media aritmética calculada sobre la población y $\bar{x}$ para la calculada sobre una muestra).

Cálculo de la desviación estándar

Es sencillo calcularla artesanalmente, basta con aplicar la fórmula: primero la media aritmética, luego se va calculando la diferencia entre cada valor y la media, su cuadrado, suma de todos los cuadrados, etc.

Pero lo habitual es realizar el cálculo con una calculadora o con la Hoja de Cálculo en un ordenador.

Con la **calculadora** el proceso se limita a introducir todos los datos, y luego solicitar la media y la desviación estándar con las teclas correspondientes. Aparecen las teclas $\boldsymbol{\sigma_n}$ y $\boldsymbol{\sigma_{n-1}}$, que se refieren respectivamente a las dos fórmulas que hemos visto: con los datos de la población (dividir por n) y con los datos de la muestra (dividir por n-1)

La mejor elección es la **hoja de cálculo** donde se utiliza la fórmula **=DESVESTP( )**, o bien **=DESVEST( )**, refiriéndose, como antes a los datos de la población o de una muestra, respectivamente. En ambos casos, dentro del paréntesis incluiremos las celdillas que deseamos realizar el cálculo, por ejemplo: =DESVEST(A2:A35) , si los valores se encuentran en la columna A, desde A2 hasta A35. La media aritmética se obtiene mediante **=PROMEDIO( )**.

Coeficiente de Variación

Si dos series tienen la misma media, su desviación estándar nos indica en cual de las dos los valores está más dispersos a ambos lados de la media. Pero *si las medias son distintas, la comparación de las desviaciones estándar no sirve*. Supongamos que deseamos saber cuál de las dos series siguientes está más dispersa a ambos lados de su media:

							Media	Desv. estándar	C.V
19	20	21	23	24	26	28	23	3,0	0,13
1259	1311	1350	1374	1396	1423	1445	1365,4	64,8	0,05

Aparentemente, la segunda serie prese nta una mayor dispersión (*s* = 64,8 parece muy alta comparada con *s* = 3,0 de la primera). Pero *s*=3,0 en valores que rondan la media de 23 es mayor que *s* = 64,8 en una población de media 1365. Esta idea se cuantifica mediante el **Coeficiente de Variación (C.V.)** :

$$\text{Coeficiente de variación.} = \frac{\text{Desviación Estándar}}{\text{Media aritmética}}$$

Así observamos que la dispersión de la primera muestra es relativamente mayor (CV=0,13) su desviación estándar equivale al 13% de la media, mientras que en la segunda muestra, su desviación estándar es solamente el 5% de su media (CV=0,05)

Puntuaciones estandarizadas

En los cálculos de los siguientes apartados, en lugar de trabajar con *puntuaciones brutas*, debemos utilizar *puntuaciones estandarizadas o tipificadas*. La puntuación estandarizada nos indica cuántas desviaciones estándar se encuentra un valor individual por encima o por debajo de la media.

<u>Ejemplo</u>: Hemos calculado la media y la desviación estándar de los caudales de un río: caudal medio= 97 m^3/seg; desviación estándar 13,4 m^3/seg. En un año húmedo el caudal fue de 112 m^3/seg. Convertir ese valor en puntuación estándar.

<u>Solución</u>: El valor 112 m^3/s supera a la media en: 112-97=15 m^3/s. Calculamos cuántas desviaciones estándar está por encima de la media:

El caudal del río superó a su media en : $\frac{112-97}{13,4}$ =**1,12** desviaciones estándar .

Por tanto, si hablamos de un caudal 0 (valor estandarizado) no significa que el río esté seco, sino que su caudal es igual a la media. La expresión general es:

$$\text{Puntuación estándar} = \frac{\text{Puntuación bruta - Media}}{\text{Desviación estándar}} \approx z = \frac{x - \bar{x}}{s_x}$$

La puntuación estandarizada o tipificada se representa generalmente como ***u*** o ***z***.

Curvas de probabilidad utilizando puntuaciones estandarizadas

En la figura 2, en el eje de abcisas figuraban las estaturas en cm. Un gráfico similar para precipitaciones anuales mostraría en ese eje valores en mm. y para caudales en m^3/s. Ahora que conocemos los valores estandarizados, preferimos utilizar la misma curva de la figura 2 representando en el eje de abcisas valores estandarizados en vez de las estaturas (puntuaciones brutas) los correspondientes: resultaría como se muestra en la figura 7. De este modo, el mismo gráfico será válido para estaturas, precipitaciones, caudales de un arroyo (en el eje de abcisas los valores variarían de 0,2 a 3 m^3/s) o caudales de un gran río (valores de 500 a 8000 m^3/s).

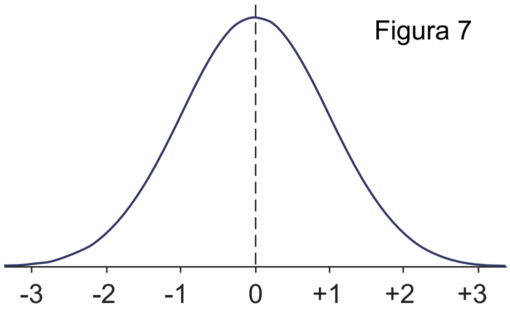


Figura 7

Cálculo de probabilidades con la Ley de Gauss (valores medios)

Asumiendo que la serie de caudales se ajusta a la ley de Gauss, podemos responder a dos tipos de cuestiones:

1. *¿Cuál es la probabilidad de que el caudal supere 40 m^3/seg? O bien: ¿Cada cuántos años se superará el caudal de 40 m^3/seg?*
2. *¿Qué caudal será superado un 2% de los años? O lo que es lo mismo: ¿Cual es el caudal superado cada 50 años?*

Datos necesarios: Media aritmética=29,8 m^3/seg; desv estándar=8,1 m^3/seg

Solución a la Cuestión 1 (del valor a la probabilidad):

1º) Expresamos el caudal de 40 m^3/seg como puntuación estandarizada: $z = \frac{40-29,8}{8,1} = 1,26$.

Esto significa que ese dato individual está 1,26 desviaciones estándar por encima de la media.

2º) Calculamos la probabilidad de que ***z***≥1,26. Como aplicar la ecuación de Gauss no es simple, ésto puede hacerse de dos maneras:

- Con la **Hoja de Cálculo**, escribiendo en EXCEL la siguiente fórmula:
  =1-DISTR.NORM.ESTAND(1,26)
- Aplicando la **Tabla** que se presenta al final (Esta Tabla se construye aplicando la fórmula de Gauss a todos los posibles valores de ***z***).

Para nuestro caso (z =1,26) por cualquiera de los dos procedimientos obtenemos el valor: 0,10383. Por tanto, el 10,38% de los años tendrán un caudal igual o superior a 40 m^3/seg. El caudal citado se superará en promedio cada 10 años.

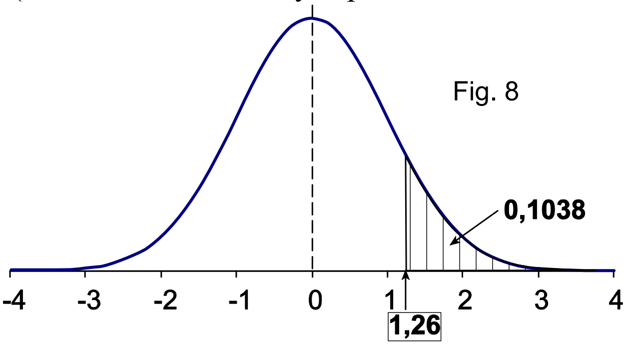


Fig. 8

En la figura 8 vemos que por encima del valor estándar +1,26 se encuentra el 10,38% de la superficie total bajo la curva.

Solución a la Cuestión 2 (de la probabilidad al valor):

Se trata de repetir el proceso anterior al revés:

1º) Calculamos a qué valor de **z** corresponde la probabilidad 0,02 (o sea: 2%). De nuevo, ésto puede hacerse de dos maneras:

- Con la Hoja de Cálculo, escribiendo en EXCEL la siguiente fórmula: =DISTR.NORM.ESTAND.INV(1-0,02)
- Aplicando la Tabla que se presenta al final, inversamente a como la utilizamos antes: buscar dentro de la tabla la probabilidad requerida (en este ejemplo, 0,02), o la más próxima a ese valor, y desde el interior de la tabla, leer el valor de **z** correspondiente en los bordes de la Tabla.

Para nuestro caso (probabilidad=0,02) por cualquiera de los dos procedimientos obtenemos el valor: 2,05. Finalmente, calculamos a qué puntuación bruta corresponde una puntuación estandarizada de 2,05:

$$2{,}05 = \frac{x - 29{,}8}{8{,}1} \quad ; \quad \text{x} = 46{,}4 \ \text{m}^3/\text{s}$$

Por tanto, el valor que es superado un 2% de los años (cada 50 años) es 46,4 m^3/seg

Las mismas cuestiones con valores son inferiores a la media

En las dos cuestiones anteriores se manejaban caudales superiores a la media. Nos movemos en la mitad derecha de la "campana" de Gauss (ver la figura 5). De hecho, la tabla de valores que utilizamos para resolver las dos cuestiones anteriores, solamente refleja la mitad derecha del gráfico de Gauss; no sería problema construir de una tabla de doble tamaño para manejar también valores **inferiores a la media.**

Si estamos haciendo previsiones de años secos, las preguntas (equivalentes a las cuestiones ***1*** y ***2*** de la página anterior) serán de este tipo:

> ***3.*** *¿Cuál es la probabilidad de que el caudal no alcance los 15 m^3/seg? ?*
>
> ***4.*** *¿Qué caudal no se alcanzará un 10% de los años?*

Se trata de la misma muestra que en los ejemplos ***1*** y ***2*** anteriores (Media aritmética =29,8 m^3/seg; desv estándar=8,1 m^3/seg).

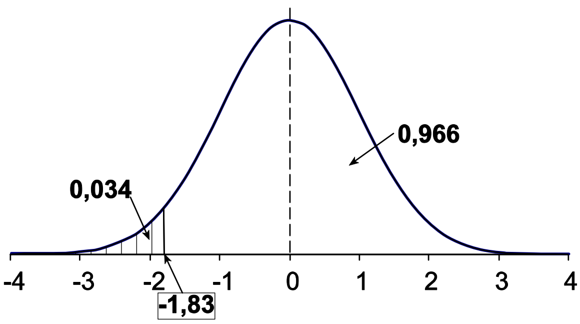
Solución a la Cuestión 3 (del valor a la probabilidad):

1º) Expresamos el caudal de 15 m^3/seg como puntuación estandarizada: $z = \dfrac{15 - 29{,}8}{8{,}1} = -1{,}83$.

Esto significa que ese dato individual está 1,83 desviaciones estándar por debajo de la media.

2º) Calculamos la probabilidad de que **z** $\geq -1{,}83$:

Aplicando la **Tabla** buscamos la probabilidad correspondiente a **z** = –1,83.
Para valores **negativos** de **z** se toma el valor complementario, es decir: si para 1,83 la tabla da 0,034, para –1,83 corresponde 1–0,034=0,966



Por tanto, la probabilidad de superar el caudal de 15 m^3/seg es de 0,966, y la probabilidad de que **no** se supere ese valor será de 1–0,966= 0,034 (Recuperamos el valor 0,034 que nos proporcionó la tabla inicialmente).

Con la **Hoja de Cálculo**, escribiendo en EXCEL la fórmula: =DISTR.NORM.ESTAND(-1,83) nos proporciona directamente la probabilidad de que sea **menor que** –1,83 desviaciones estándar: 0,034

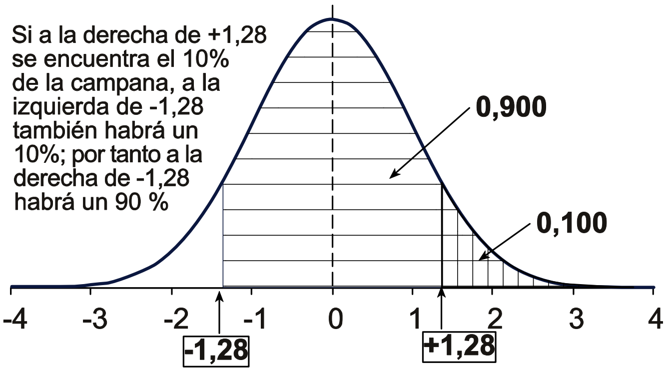
Respuesta final: probabilidad de que **no** alcance 15 m^3/s = 3,4%

Solución a la Cuestión 4 (de la probabilidad al valor):

La cuestión **4** podemos replantearla así: ¿Qué caudal se superará el 90% de los años?

1º) Calculamos a qué valor de ***z*** corresponde la probabilidad 0,90 (o sea: 90%):

Aplicando la Tabla, buscamos dentro de ella la probabilidad requerida (0,90), pero ese valor no existe en la tabla, así que buscamos el complementario: 1-0,90 =0,10 ; o el más próximo a ese valor, y desde el interior de la tabla, leemos el valor de ***z*** correspondiente en los bordes de la Tabla: 1,28 . Pero ***z*** = 1,28 corresponde a una probabilidad de 0,10; para la probabilidad 0,90 tomamos ***z*** = –1,28



Con la **Hoja de Cálculo**, escribiendo en EXCEL la fórmula: =DISTR.NORM.ESTAND.INV(0,10) se obtiene directamente el valor –1,28

2º) Calculamos a qué puntuación bruta corresponde una puntuación estandarizada de –1,28:

$$-1,28 = \frac{x - 29,8}{8,1} \qquad ; \qquad x = 19,4$$

Por tanto, el valor que no se alcanza el 10% de los años (probabilidad 0,90 de ser superado) es 19,4 m^3/seg.

Probabilidad y periodo de retorno

A lo largo de los apartados anteriores se ha estado utilizando indistintamente probabilidad (por ejemplo: un 2% de los años) y expresiones como "cada 50 años".

Es evidente que si un caudal se iguala o supera (por término medio) cada 10 años, la probabilidad de qu eso suceda es de 0,10 (10%).

Análoga e inversamente, si la probabilidad de que algo suceda es de 0,04 (4%), ello quiere decir que, en promedio, sucederá 4 veces en 100 años, o sea: una vez cada 25 años.

Estos conceptos se relacionan mediante la expresión:

$$\text{Periodo de retorno} = \frac{1}{\text{Probabilidad}}$$

Como : [Probabilidad de que NO se supere] = 1 – [Probabilidad de que se supere], también se cumple:

$$\text{Probabilidad de que se supere} = \frac{1}{\text{Periodo de retorno}} \qquad ; \qquad \text{Probabilidad de que NO se supere} = 1 - \frac{1}{\text{Periodo de retorno}}$$

En Hidrología, para valores con un probabilidad superior al 50% se utiliza más el **periodo de retorno** que la **probabilidad**. Así, se habla de "la crecida de 50 años" en lugar de referirse a "la crecida con probabilidad 0,02"; o se dice "precipitación con retorno de 100 años" en vez de "la precipitación con probabilidad 0,01".

Un ejemplo de serie de valores extremos sería si hubiéramos elegido, de una serie de años, el día mas caudaloso o de mayor precipitación de cada año. Para el estudio de series de valores extremos se utilizan diversas distribuciones; la de utilización más simple es la distribución de Gumbel :

La probabilidad de que se presente un valor **inferior** a ***x*** es:

$$F(x) = e^{-e^{-(x-u)/\alpha}} \qquad (3)$$

siendo: $\alpha = s_x \;/\; \sigma_y$ (4)

$u = \bar{x} - \mu_y \cdot \alpha$ (5)

F(x) = Probabilidad de que se presente un valor igual o menor que *x*.

e = base de los logaritmos neperianos

$\bar{x}$ = media aritmética de la muestra

s_x = desviación estándar de la muestra

σ_y , μ_y = consultar en la tabla adjunta, según el número de datos de la muestra[2]]

nº datos	μ_y	σ_y
10	0,4952	0,9496
15	0,5128	1,0206
20	0,5236	1,0628
25	0,5309	1,0914
30	0,5362	1,1124
35	0,5403	1,1285
40	0,5436	1,1413
45	0,5463	1,1518
50	0,5485	1,1607
55	0,5504	1,1682
60	0,5521	1,1747
65	0,5535	1,1803
70	0,5548	1,1854
75	0,5559	1,1898
80	0,5569	1,1938
85	0,5578	1,1974
90	0,5586	1,2007
95	0,5593	1,2037
100	0,5600	1,2065

Mediante las expresiones anteriores podremos calcular la frecuencia a partir del valor *x*, es decir: calcular con qué frecuencia (o periodo de retorno) se presentará un cierto caudal o precipitación.

Para solucionar el **caso inverso** (qué caudal o precipitación se producirán cada *n* años) debemos despejar *x* en la expresión (3), obteniendo:

$$(x-u) \,/\, \alpha = -\ln(-\ln(F(x)))$$

$$x = -\ln(-\ln(F(x))) \cdot \alpha + u \qquad (6)$$

Ejemplo.- De una serie de 55 caudales máximos (el caudal diario máximo de cada año)[3], hemos calculado:

Media= 21,97 m^3/seg
Desv. estándar=13,22 m3/seg

Calcular: a) Periodo de retorno de un caudal de 60 m^3/s. b) Caudal con retorno de 100 años.

a) Periodo de retorno para que el día más caudaloso del año el caudal supere el valor de 60 m^3/seg

1º) De acuerdo con la tabla adjunta, para 55 datos, tomamos los valores:

$$\mu_y = 0{,}5504 \quad ; \quad \sigma_y = 1{,}1682$$

2º) Calculamos α y u:

[2] Algunos autores utilizan μ_y =0,5772; σ_y = 1,2825 sin considerar el número de datos. Equivale a considerar no la muestra disponible, sino toda la población (nº de datos infinito)

μ_y , σ_y son, respectivamente, la media y la desviación estándar de una serie de valores y_i definida así::

$$y_i = -\ln\left(\ln\left(\frac{N+1}{i}\right)\right) \qquad (i = 1 \text{ a } N \;;\; N = \text{nº de datos de la muestra})$$

[3] Datos de Hoyos del Espino, en la cabecera del río Tormes, cuenca receptora 88 km^2

$$\alpha = s_x / \sigma_y = 13{,}22 / 1{,}1682 = 11{,}3168$$

$$u = \bar{x} - \mu_y \cdot \alpha = 21{,}97 - 0{,}5504 \cdot 11{,}3168 = 15{,}741$$

3º) Aplicamos la ecuación de Gumbel (3). La probabilidad de que se presente un caso **menor** que *x* será:

$$F(x) = e^{-e^{-(x-u)/\alpha}} = e^{-e^{-(60-15{,}741)/11{,}3618}} = 0{,}9803 \approx 98{,}03\%$$

Por tanto, la probabilidad de que se presente un caudal mayor que ***x*** será:

$$1- F(x) = 1 -0{,}9803 = 0{,}0197 \ (= 1{,}97\%)$$

Finalmente, el **periodo de retorno** es el inverso de la probabilidad:

Periodo de retorno= 1/0,0197 = 50,8 años

b) Caso inverso: Calcular el caudal con retorno de 100 años

Probabilidad de superar = 1/ periodo de retorno = 1 / 100 = 0,01

Por tanto, buscamos un caudal con una probabilidad de 0,01 de ser superado

Probabilidad de ser inferior = 1 – prob. de ser superior; es decir: $F(x) = 1 - 0{,}01 = 0{,}99$.

Ahora deberíamos calcular α y u mediante las expresiones (4) y (5), pero en este ejemplo ya las hemos calculado en el apartado anterior.

Aplicando la fórmula (6):

$$x = -\ln(-\ln(F(x))) \cdot \alpha + u = -\ln(-\ln(0{,}99)) \cdot 11{,}317 + 15{,}741 = 67{,}8 \text{ m}^3/\text{s}$$

Para el cálculo de probabilidades de valores extremos se utilizan diversas distribuciones, entre las que destacan, como más utilizadas, la **log-normal** (los logaritmos de los valores son los que se ajustan a la ley de Gauss) o la ley **Pearson III**, adoptada por las agencias federales en USA. Ver, por ejemplo en Viessman, 2003, capítulo 3.

En España los organismos oficiales para precipitaciones máximas aplican la distribución **SQRT-max**[4]

Riesgo de fallo

Se denomina **riesgo de fallo** a la probabilidad de que se produzca un suceso con un periodo de retorno *T* en alguno de los próximos *n* años.

Supongamos que hemos calculado un cierto caudal que corresponde al retorno de 50 años. La probabilidad de que se produzca el año próximo será de 0,02 (=1/50); y la probabilidad de que se produzca el siguiente año será de 0,02 y así cada año. Queremos conocer la probabilidad de que se alcance ese caudal en los próximos *n* años:

Probabilidad de que un suceso de retorno *T* se produzca el próximo año .... $1/T$

“ “ “ NO se produzca el próximo año [5] .... $1-(1/T)$

“ “ “ NO se produzca los próximos dos años[6] .... $[1-(1/T)].\ [1-(1/T)]$

“ “ “ NO se produzca los próximos n años .... $[1-(1/T)]^n$

“ “ “ SI se produzca los próximos n años [5] .... $1-[1-(1/T)]^n$

Vamos a denominar a la última expresión obtenida arriba es el **riesgo de fallo (R)**, es decir: la probabilidad de que **sí** se produzca alguna vez un suceso de **periodo de retorno T** a lo largo de un periodo de **n** años:

---

[4] Imposible el cálculo manual, ver una aproximación con Excel en http://hidrologia.usal.es (Seción “Complementos”)

[5] Las probabilidades de dos **sucesos complementarios** (debe suceder uno u otro) suman 1. Por ejemplo: probabilidad de obtener un 3 en un dado= 1/6. Probabilidad de obtener un valor distinto de 3= 1-1/6 = 5/6

[6] La probabilidad de que se produzcan dos **sucesos independientes** es el producto de sus probabilidades; por ejemplo: probabilidad de obtener un 3 en un dado= 1/6. Probabilidad de obtener dos 3 seguidos = 1/6.1/6 =1/36

$$R = 1 - \left(1 - \frac{1}{T}\right)^n$$

Ejemplo: Se va a construir un canal cuya *vida útil* es de 75 años. Si el caudal supera el valor correspondiente al periodo de retorno de 100 años, se desbordará. Calcular la probabilidad de que se produzca un desbordamiento en alguno de los próximos 75 años

$$R = 1 - \left(1 - \frac{1}{100}\right)^{75} = 0{,}529 = 52{,}9\%$$

Por tanto, existe un 52,9% de probabilidad de que el caudal de retorno 100 años se alcance en alguno de los próximos 75 años.

Se produce la siguiente paradoja: si consideramos un caudal con retorno de 100 años, parece seguro que se presente en alguno de los próximos 100 años. Pero si aplicamos la fórmula anterior, haciendo T= 100 y n=100, y obtenemos 0,633 , es decir solamente un 63,3 %

Cálculo inverso: evaluación del periodo de retorno a partir del riesgo

Muchas veces, esta sencilla fórmula debe aplicarse a la inversa: despejar T a partir del riesgo de fallo (R) y del número de años (n). Depejando T en la última fórmula obtenemos:

$$T = \frac{1}{1 - \exp\left(\frac{\ln(1-R)}{n}\right)}$$

donde: R =riesgo de que se produzca el suceso de probabilidad $1/T$ durante los próximos n años

T = periodo de retorno en años

$\exp(x) = e^x$

Ejemplo: Se está diseñando una obra cuya vida útil se calcula en 50 años y se admite que en ese periodo el riesgo sea de un 10% (probabilidad de que en esos 50 años se produzca un caudal superior a un valor determinado). Calcular dicho caudal.

En la fórmula anterior basta con hacer: R = 0,10; n = 50 años; y despejar T. Con estos datos obtenemos un periodo de retorno T = 475 años.

En este ejemplo, el paso siguiente sería estudiar estadísticamente las series históricas de caudales de ese cauce para evaluar el caudal correspondiente a un retorno de 475 años.

Ley de Gauss: Probabilidad de que *z* sea mayor o igual a ...

(Las columnas indican la segunda decimal. Ejemplo: Probabilidad de que z sea > 1,41 es 0,07927)

Para valores de *z* negativos, tomar **1-tabla**. Ejemplo: Probabilidad de que z sea ≥ – 1,41 es 1 – 0,07927 = 0,92073

Para probabilidades > 0,50, el valor de **z** será el indicado por la tabla para la probabilidad complementaria, pero con signo –

Ejemplo : Valor de z con probabilidad de ser superado de 0,80. Para la probabilidad complementaria (0,20) la tabla indica *z*=0,84. Por tanto para probabilidad 0,80 adoptaremos –0,84

	0	1	2	3	4	5	6	7	8	9
0,0	0,50000	0,49601	0,49202	0,48803	0,48405	0,48006	0,47608	0,47210	0,46812	0,46414
0,1	0,46017	0,45620	0,45224	0,44828	0,44433	0,44038	0,43644	0,43251	0,42858	0,42465
0,2	0,42074	0,41683	0,41294	0,40905	0,40517	0,40129	0,39743	0,39358	0,38974	0,38591
0,3	0,38209	0,37828	0,37448	0,37070	0,36693	0,36317	0,35942	0,35569	0,35197	0,34827
0,4	0,34458	0,34090	0,33724	0,33360	0,32997	0,32636	0,32276	0,31918	0,31561	0,31207
0,5	0,30854	0,30503	0,30153	0,29806	0,29460	0,29116	0,28774	0,28434	0,28096	0,27760
0,6	0,27425	0,27093	0,26763	0,26435	0,26109	0,25785	0,25463	0,25143	0,24825	0,24510
0,7	0,24196	0,23885	0,23576	0,23270	0,22965	0,22663	0,22363	0,22065	0,21770	0,21476
0,8	0,21186	0,20897	0,20611	0,20327	0,20045	0,19766	0,19489	0,19215	0,18943	0,18673
0,9	0,18406	0,18141	0,17879	0,17619	0,17361	0,17106	0,16853	0,16602	0,16354	0,16109
1,0	0,15866	0,15625	0,15386	0,15151	0,14917	0,14686	0,14457	0,14231	0,14007	0,13786
1,1	0,13567	0,13350	0,13136	0,12924	0,12714	0,12507	0,12302	0,12100	0,11900	0,11702
1,2	0,11507	0,11314	0,11123	0,10935	0,10749	0,10565	0,10383	0,10204	0,10027	0,09853
1,3	0,09680	0,09510	0,09342	0,09176	0,09012	0,08851	0,08692	0,08534	0,08379	0,08226
1,4	0,08076	0,07927	0,07780	0,07636	0,07493	0,07353	0,07215	0,07078	0,06944	0,06811
1,5	0,06681	0,06552	0,06426	0,06301	0,06178	0,06057	0,05938	0,05821	0,05705	0,05592
1,6	0,05480	0,05370	0,05262	0,05155	0,05050	0,04947	0,04846	0,04746	0,04648	0,04551
1,7	0,04457	0,04363	0,04272	0,04182	0,04093	0,04006	0,03920	0,03836	0,03754	0,03673
1,8	0,03593	0,03515	0,03438	0,03362	0,03288	0,03216	0,03144	0,03074	0,03005	0,02938
1,9	0,02872	0,02807	0,02743	0,02680	0,02619	0,02559	0,02500	0,02442	0,02385	0,02330
2,0	0,02275	0,02222	0,02169	0,02118	0,02068	0,02018	0,01970	0,01923	0,01876	0,01831
2,1	0,01786	0,01743	0,01700	0,01659	0,01618	0,01578	0,01539	0,01500	0,01463	0,01426
2,2	0,01390	0,01355	0,01321	0,01287	0,01255	0,01222	0,01191	0,01160	0,01130	0,01101
2,3	0,01072	0,01044	0,01017	0,00990	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
2,7	0,00347	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,00280	0,00272	0,00264
2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139
3,0	0,00135	0,00131	0,00126	0,00122	0,00118	0,00114	0,00111	0,00107	0,00104	0,00100

$$\boxed{z = \frac{x - \overline{x}}{s_x}}$$

donde:
x =puntuación bruta
z = puntuación estandarizada
$\overline{x}$ = media aritmética
s_x =desviación estándar

Representación gráfica de la probabilidad proporcionada por esta tabla:

Si toda el área bajo la curva de Gauss vale 1 (ya que bajo la curva se encuentran el 100% de los casos), la tabla nos da la parte de dicha área superior a la puntuación dada.

A la derecha vemos un ejemplo: para **z = 1,5** la tabla nos proporciona el valor **0,0668** que es la parte rayada del dibujo (el **6,68%** de la superficie total bajo la curva) y representa **los casos que superan a la media en 1,5 desviaciones estándar**