**TRATAMIENTO ESTADISTICO DE DATOS**

**1. OBJETIVOS**

Al final de la práctica el alumno debe ser capaz de:

Conocer y utilizar los parámetros estadísticos básicos para el análisis de información climatológica.

**2. GENERALIDADES**

**PARAMETROS ESTADISTICOS BASICOS**

El análisis estadístico posibilita la obtención de medidas de centralización, dispersión y tendencias

temporales en las series de observaciones de los fenómenos meteorológicos con el fin de analizar la zonación de estos fenómenos. A continuación abordamos muy brevemente algunos de los parámetros estadísticos más usuales en Climatología.

Media: es el cociente entre la suma de todos los valores de la serie y el número de datos de la serie.

****

Mediana: es el valor que divide a una serie ordenada en dos conjuntos de igual probabilidad. La mediana se corresponde con el percentil 50.

Cuantiles: los cuantiles dividen la serie en cuatro (cuartiles), cinco (quintiles), diez (deciles) ó cien

(percentiles) grupos iguales. Los cuantiles representan niveles con una determinada probabilidad de ser sobrepasados, por ello se utilizan con frecuencia para representar gráficamente los regímenes pluviométricos probables. Se pueden obtener según la distribución de frecuencias empíricas de la serie o en el caso de que se requiera acudir a una función de distribución teórica de frecuencias, referirlos a la función correspondiente de ajuste (por ejemplo: normal o gamma incompleta).

Percentiles: valores que dividen el conjunto de datos ordenados en cien partes iguales: P1, P2, ... , P99.

Deciles: valores que dividen el conjunto de datos ordenados en diez partes iguales: D1, D2, ... , D9.

Quintiles: valores que dividen el conjunto de datos ordenados en cinco partes iguales: Q1, Q2, ... , Q4

Cuartiles: valores que dividen el conjunto de datos ordenados en cuatro partes iguales. El rango intercuartílico se obtiene a partir de los cuartiles. Es la diferencia entre el cuartil tres y el uno.

Varianza y la desviación típica: son parámetros de uso común. La varianza es la media aritmética de los cuadrados de las diferencias de cada valor con respecto a la media.



La desviación típica o estándar es la raíz cuadrada de la varianza

Si se emplea el estimador insesgado se divide por n-1. En general, nosotros emplearemos éste, la

raíz cuadrada del estimador insesgado de la varianza muestral.

=

Sesgo: es la mayor o menor simetría o asimetría de una distribución. Si la función de distribución tiene una “cola” más larga hacia la izquierda que hacia la derecha con respecto al máximo central, se dice que la distribución está sesgada a la izquierda o que tiene sesgo negativo. Si la función de distribución tiene una “cola” más larga hacia la derecha que hacia la izquierda con respecto al máximo central, se dice que la distribución está sesgada a la derecha o que tiene sesgo positivo. A las distribuciones de sesgo nulo se las denomina insesgadas.

Coeficiente de variación: es el cociente entre la desviación típica y la media aritmética.

**PERCENTILES**

Los Percentiles (quintiles o deciles) son unos parámetros de dispersión que asocian probabilidades

de ocurrencia a precipitaciones (o variable meteorológica) de un determinado volumen de agua. Así si decimos que el percentil 20 (quintil 1 ó decil 2) para el observatorio de Huayao es de 340 mm, estamos afirmando que existe un 20% de probabilidad de que caiga una precipitación anual del 340 mm.

|  |  |  |
| --- | --- | --- |
| EQUIVALENCIAS | | |
| Percentiles | Deciles | Quintiles |
| 20 | 2 | 1 |
| 40 | 4 | 2 |
| 50 | 5 |  |
| 60 | 6 | 3 |
| 80 | 8 | 4 |

En general, se analizan los percentiles 20, 40, 60 y 80 (que se corresponden, respectivamente, con los quintiles 1, 2, 3 y 4). Los percentiles se calculan para cada mes o para los valores anuales por separado. Estos valores permitirán calificar la sequedad de un determinado año en función del volumen de precipitación caída. Los métodos abordados para el estudio de la distribución temporal de las precipitaciones más generalizados son:

1) distribución de frecuencias empíricas de la serie;

2) distribución teórica de frecuencias gamma; y

3) distribución teórica normal.

**DISTRIBUCION EMPÍRICA**

Los percentiles empíricos se calculan a partir de la función de distribución empírica definida por los valores de la serie con la que se trabaja ordenada desde el valor menor al mayor, y asignando a cada valor ordenado su probabilidad calculada según la expresión:

Prob (X≤xi) = i/(N +1 ).

Donde ”i” representa el número de orden que ocupa el valor “x” en la serie de datos ordenada en

orden creciente y “N” el número total de datos. La probabilidad correspondiente al 20, 40, 50, 60 ó 80 por ciento se obtienen por interpolación lineal, considerando las probabilidades asignadas a cada dato ordenado.

Ejercicio 4.2. Se pide calcular los valores de los percentiles 20 y 40 mediante la función de distribución empírica, de la siguiente serie de valores:

**102.2 96.3 377.7 119.9**

**221.1 32 153.8 199**

**261.9 58.7 160 209.8**

**270 60.4 171.9 142**

**138.3 83.5 172.1 148.5**

**13.5 289.4 183.6 269.4**

**18.1 299.9 197.9**

**118 110.5 300.7**

Solución

Se deben ordenar los datos de precipitación en orden creciente, y asignar a cada valor de precipitación su probabilidad empírica en función del orden de situación del valor y del número de datos. Así para los dos primeros valores y para los dos últimos tendremos:

*Nº orden Precipitación Probabilidad*

1 (primero) 13.5 mm Prob (X≤13.5) = i/(N +1) = 1/(30+1) = 0.03226 (3.226 %)

2 (segundo) 18.1 mm Prob (X≤18.1) = i/(N +1) = 2/(30+1) = 0.06452 (6.452 %)

...

29 (vigésimo nono) 300,7 mm Prob (X≤300,7) = i/(N +1) = 29/(30+1) = 0.9354 (93,55 %)

30 (trigésimo) 377,7 mm Prob (X≤377,7) = i/(N +1) = 30/(30+1) = 0.9677 (96,77 %)

Los valores de precipitación ordenados desde el menor al mayor para los treinta años de la serie y

los valores de probabilidad asignados son:

Nº orden Prob % Prec (mm)

1 3.226 13.5

2 6.452 18.1

3 9.677 32

4 12.9 58.7

5 16.13 60.4

6 19.35 83.5

7 22.58 96.3

8 25.81 102.2

9 29.03 110.5

10 32.26 118

11 35.48 119.9

12 38.71 138.3

13 41.94 142

14 45.16 148.5

15 48.39 153.8

16 51.61 160

17 54.84 171.9

18 58.06 172.1

19 61.29 183.6

20 64.52 197.9

21 67.74 199

22 70.97 209.8

23 74.19 221.1

24 77.42 261.9

25 80.65 269.4

26 83.87 270

27 87.1 289.4

28 90.32 299.9

29 93.55 300.7

30 96.77 377.7

El percentil 20 se obtiene por interpolación sabiendo que será un volumen de precipitación entre el valor que está en la posición sexta (19,35 %) y el valor que está en la posición séptima (22,58 %)

P19.35 83.5 P20 86.07

P22,58 96.3

El percentil 40 se obtiene por interpolación sabiendo que será un volumen de precipitación entre el valor que está en la posición duodécima (38,71 %) y el valor que está en la posición décimotercera (41,94 %)

P38.71 138.3 P40 139.78

P41.94 142

**DISTRIBUCION NORMAL**

La ley de distribución normal se define a partir de la media (x) y la desviación típica (σ). La función de distribución de esta ley muestra que:

El 50 % de las observaciones están en el intervalo (x ± 0,68σ)

El 68,3 % de las observaciones están en el intervalo (x ± σ)

El 95 % de las observaciones están en el intervalo (x ± 1,96σ)

El 99 % de las observaciones están en el intervalo (x ± 2,58σ)

El 99,9 % de las observaciones están en el intervalo (x ± 3,29σ)

Para extender los resultados a cualquier serie las variables se tipifican. Llamaremos tipificación de la variable al paso de una variable aleatoria X є N (x, σ) a una nueva variable (la tipificada) Z є N (0, 1). El paso consiste en centrar (hacer la media nula) y reducir (hacer la desviación típica igual a 1).

La nueva variable será: Zi = (xi – x/ σ)

La distribución normal tipificada tiene de media 0 y las tablas de la distribución normal nos dan directamente la probabilidad de obtener un valor inferior o superior a un valor tipificado dado.

Ejercicio

Dados los datos de temperaturas medias (ºC) para el mes de Julio del observatorio de Cáceres. Se pide determinar la probabilidad de que la temperatura media del mes de Julio sea inferior a 26 °C.

Tm tm tm

1971 24,9 1983 23,6 1995 26,4

1972 24,1 1984 25,7 1996 26,2

1973 24,2 1985 25,8 1997 24,8

1974 27,5 1986 27 1998 27,2

1975 26 1987 25,3 1999 27,4

1976 25,9 1988 24,6 2000 24,5

1977 22,4 1989 28,8

1978 25,7 1990 28,2

1979 26,2 1991 27

1980 24,7 1992 27,3

1981 26,2 1993 26,5

1982 24,1 1994 26,6

Media = 25,8 °C

Número de datos: n = 30

Desviación típica = 1,44 °C

Para la temperatura de 26 ºC, la variable tipificada será :

([26-25,8]/1,44) = 0,1388.

En las tablas para un valor de z = 0,138, tenemos que la probabilidad de obtener una valor inferior a Z será 0,556. Luego el 55,6 % de los años la temperatura será inferior a 26 ºC.

Tabla: Valor de la variable tipificada. El valor correspondiente a la fila y la columna nos da la probabilidad de obtener una valor inferior a Z.

**0 0,01 0,02 0,03 0,04 0,05 0,06 0,07 0,08 0,09**

**0** 0,5 0,504 0,508 0,512 0,516 0,5199 0,5199 0,5279 0,5319 0,5359

**0,1** 0,5398 0,5438 0,5478 0,5517 0,5557 0,5596 0,5596 0,5675 0,5714 0,5753

**0,2** 0,5793 0,5832 0,5871 0,591 0,5948 0,5987 0,5987 0,6064 0,6103 0,6141

**0,3** 0,6179 0,6217 0,6255 0,6293 0,6331 0,6368 0,6368 0,6443 0,648 0,6517

**0,4** 0,6554 0,6591 0,6628 0,6664 0,67 0,6736 0,6736 0,6808 0,6844 0,6879

**0,5** 0,6915 0,695 0,6985 0,7019 0,7054 0,7088 0,7088 0,7157 0,719 0,7224

**0,6** 0,7257 0,7291 0,7324 0,7357 0,7389 0,7422 0,7422 0,7486 0,7517 0,7549

**0,7** 0,758 0,7611 0,7642 0,7673 0,7704 0,7734 0,7734 0,7794 0,7823 0,7852

**0,8** 0,7881 0,791 0,7939 0,7967 0,7995 0,8023 0,8023 0,8078 0,8106 0,8133

**0,9** 0,8159 0,8186 0,8212 0,8238 0,8264 0,8289 0,8289 0,834 0,8365 0,8389

**1** 0,8413 0,8438 0,8461 0,8485 0,8508 0,8531 0,8531 0,8577 0,8599 0,8621

**1,1** 0,8643 0,8665 0,8686 0,8708 0,8729 0,8749 0,8749 0,879 0,881 0,883

**1,2** 0,8849 0,8869 0,8888 0,8907 0,8925 0,8944 0,8944 0,898 0,8997 0,9015

**1,3** 0,9032 0,9049 0,9066 0,9082 0,9099 0,9115 0,9115 0,9147 0,9162 0,9177

**1,4** 0,9192 0,9207 0,9222 0,9236 0,9251 0,9265 0,9265 0,9292 0,9306 0,9319

**1,5** 0,9332 0,9345 0,9357 0,937 0,9382 0,9394 0,9394 0,9418 0,9429 0,9441

**1,6** 0,9452 0,9463 0,9474 0,9484 0,9495 0,9505 0,9505 0,9525 0,9535 0,9545

**1,7** 0,9554 0,9564 0,9573 0,9582 0,9591 0,9599 0,9599 0,9616 0,9625 0,9633

**1,8** 0,9641 0,9649 0,9656 0,9664 0,9671 0,9678 0,9678 0,9693 0,9699 0,9706

**1,9** 0,9713 0,9719 0,9726 0,9732 0,9738 0,9744 0,9744 0,9756 0,9761 0,9767

**2** 0,9772 0,9778 0,9783 0,9788 0,9793 0,9798 0,9798 0,9808 0,9812 0,9817

**2,1** 0,9821 0,9826 0,983 0,9834 0,9838 0,9842 0,9842 0,985 0,9854 0,9857

**2,2** 0,9861 0,9864 0,9868 0,9871 0,9875 0,9878 0,9878 0,9884 0,9887 0,989

**2,3** 0,9893 0,9896 0,9898 0,9901 0,9904 0,9906 0,9906 0,9911 0,9913 0,9916

**2,4** 0,9918 0,992 0,9922 0,9925 0,9927 0,9929 0,9929 0,9932 0,9934 0,9936

**2,5** 0,9938 0,994 0,9941 0,9943 0,9945 0,9946 0,9946 0,9949 0,9951 0,9952

**2,6** 0,9953 0,9955 0,9956 0,9957 0,9959 0,996 0,996 0,9962 0,9963 0,9964

**2,7** 0,9965 0,9966 0,9967 0,9968 0,9969 0,997 0,997 0,9972 0,9973 0,9974

**2,8** 0,9974 0,9975 0,9976 0,9977 0,9977 0,9978 0,9978 0,9979 0,998 0,9981

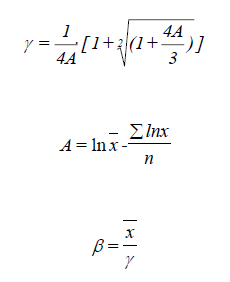
**2,9** 0,9981 0,9982 0,9982 0,9983 0,9984 0,9984 0,9984 0,9985 0,9986 0,9986

**3** 0,9987 0,9987 0,9987 0,9988 0,9988 0,9989 0,9989 0,9989 0,999 0,999

**DISTRIBUCION GAMMA**

La distribución gamma se ajusta, en general, mejor que otras distribuciones teóricas para el caso de las precipitaciones. La distribución gamma incompleta es asimétrica y es adecuada para las variables cuyo límite inferior es cero en numerosos casos. Es aplicada por el INM en las series con menos de 30 años.

La distribución gamma se define a partir de los parámetros de forma (alfa) y de escala (beta). Estos parámetros se pueden estimar mediante la aproximación de Thom (1958):



Donde β es el parámetro de distribución beta y γ es el parámetro alfa. Se destaca que con valores iguales a cero no es posible el cálculo del valor A pues el logaritmo de cero es infinito. En el caso de que aparezcan valores nulos hay que crear una función mixta compuesta de la probabilidad del valor nulo y la probabilidad del valor no nulo:

“q” y “p” = 1-q.

Lo veremos más claro con un ejemplo:

Ejemplo:

Con los siguientes datos de precipitación del mes de Julio se pide calcular los percentiles 20, 40, 60 y 80, mediante el empleo de la ley de distribución Gamma.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0 | 44.8 | 3.2 | 2.8 | 0 |
| 0 | 8.7 | 2.5 | 68.6 | 5.6 |
| 9.4 | 10 | 8.2 | 71.2 | 4 |
| 6 | 2.8 | 37.1 | 9.7 | 37.9 |
| 0 | 12.3 | 16.7 | 72.9 | 2.6 |
| 10.5 | 3.9 | 4.8 | 13.8 |  |

Solución.

El número de datos de la serie es de 29. Podemos observar que en algunos años durante el mes de Julio no hubo precipitación. Como con los valores iguales a cero no es posible el cálculo del valor A pues el logaritmo de cero es infinito. Hay que crear una función mixta compuesta de la probabilidad del valor nulo “q” y la del valor no nulo “p = 1-q”.

H(X) = q + p · G(X) Función mixta

q: probabilidad de que se presente un valor cero (sin precipitación). Fácil de calcular considerando los ceros existentes con respecto al total de datos.

p = 1-q

Como del total de 29 datos tenemos 4 con cero, tenemos:

q = 4/29 = 0.1379 (13.79)

p = 1- q = 25/29 = 0.8620 (86.21)

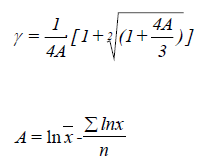
Así eliminamos los ceros y hacemos los cálculos sólo para los 25 valores restantes (función G(X) que afecta a “p”), posteriormente al final consideraremos la función mixta (H).

H(X) = q + p · G(X) Función mixta

Suma de los 25 datos = 470

Media = 470/25 = 18.8

Las formulaciones a emplear son:



Luego para calcular A es necesario calcular el logaritmo neperiano de todos los valores (los 25 no cero). Así:

ln (media) = 2.9338

Suma (lnx) = 57,11256

Luego A es igual a: A = 2,9338 – (57,11256/25) = 0.649

Tomando el valor de A obtenemos fácilmente el valor del parámetro alfa “γ” y el valor del parámetro de distribución beta “β” :

Alfa = 0.9109

Beta = 20.6393

Para calcular los percentiles se puede acudir al empleo de tablas o ábacos o emplear un programa de hojas de cálculo como el Excel. Si usamos el Excel hay que usar la función:

[=DISTR.GAMMA.INV(probabilidad;alfa;beta)]. Los parámetros de la distribución gamma incompleta alfa y beta ya están calculados, sólo se necesita considerar las probabilidades. Así:

Percentil 20 Es la probabilidad igual a 0,20 Como trabajamos con una función mixta :

H(X) = q + p · G(X)

Siendo q la probabilidad de que se presente un valor cero (sin precipitación) y p = 1-q.

Tenemos que: q = 4/29 = 0.1379 (13.79) ; y, p = 1- q = 25/29 = 0.8620 (86.21)

La precipitación que corresponde a una probabilidad del 0,2 será:

H(X) = q + p · G(X) = 0,1379 + 0.8620 · G(X) = 0.2 (20 %)

Al valor de la probabilidad del 20 % para la función mixta le corresponde una probabilidad referida sólo a los valores no nulo de:

G(X) = (0.2 – 0.1379)/0.8620 = 0,072.

No olvidemos que trabajamos sólo con los valores no nulos.

La función Excel a aplicar será: =DISTR.GAMMA.INV(0.072; 0.9109; 20.6393). Así:

Percentil 20 = 1,1 mm

Para el resto será:

G(X) = (0.4 – 0.1379)/0.8620 = 0.3040 → =DISTR.GAMMA.INV(0.3040; 0.9109; 20.6393) → Percentil 40 = 6.3

G(X) = (0.6 – 0.1379)/0.8620 = 0.5360 → =DISTR.GAMMA.INV(0.536; 0.9109; 20.6393) → Percentil 60 = 14

G(X) = (0.8 – 0.1379)/0.8620 = 0.768 → =DISTR.GAMMA.INV(0.768; 0.9109; 20.6393) → Percentil 80 = 27.5

Los valores son:

Percentil 20 (Quintil 1) = Q1 = 1 mm

Percentil 40 (Quintil 2) = Q2 = 6 mm

Percentil 60 (Quintil 3) = Q3 = 14 mm

Percentil 80 (Quintil 4) = Q4 = 28 mm

**3. MATERIALES Y PROCEDIMIENTOS**

**3.1 Materiales**

* Datos de precipitación.
* Calculadora o computadora.
* Software EXCEL.
* Materiales de escritorio.

**3.2 Procedimientos**

Desarrolle los siguientes ejercicios:

3.2.1.- Ingrese a la siguiente dirección del Servicio Nacional de Meteorología e Hidrología (SENAMHI) “Condiciones del Tiempo Actual”: <http://www.senamhi.gob.pe/main_mapa.php?t=dHi>

Ubique la estación cuyas características son:

Departamento: Loreto

Provincia: Maynas

Distrito: Fernando Lores

Nombre: Tamshiyacu

Tipo: Convencional Meteorológica

Código: 000172

Seleccione la serie de valores de precipitación (mm), registrados mensualmente durante el año 2010 para las 7 y 19 horas y calcule:

1. Percentiles 20, 40, 50 y 60 mediante la función de distribución empírica de la serie de valores.
2. Media, mediana, varianza y desviación típica

Notas:

* Si en el casillero no aparece dato, considere precipitación CERO. El cero es dato.
* Si un valor de precipitación se repite, tome un solo valor para construir la serie. Ej: 0, 0, 0, 0. Para la serie tome solo un cero y lo ubica en la posición de la serie creciente.
* Si aparece como dato -888 o similar, no considere el dato, es nulo (no CERO).
* Si aparece TZ (traza), considere precipitación igual a 0.5 mm
* Para interpolar, considere el siguiente ejemplo:

Nº orden Prob % Prec (mm)

1 3.226 13.5

2 6.452 18.1

3 9.677 32

4 12.9 58.7

5 16.13 60.4

6 19.35 83.5

7 22.58 96.3

Nos piden el percentil 20 (Prob % 20.0), este se ubica entre el número de orden 6 y 7, para lo cual hacemos lo siguiente:

Nº orden Prob % Prec (mm)

6 19.35 83.5

20.00 X

7 22.58 96.3

Despejamos el valor de X y obtenemos el percentil 20 (P20 = 86.07).

3.2.2.- Calcular los valores de los percentiles 20, 40, 60 mediante la función de distribución empírica de la siguiente serie de valores:

|  |  |  |  |
| --- | --- | --- | --- |
| 102.2 | 96.3 | 377.7 | 119.9 |
| 221.1 | 32 | 153.8 | 199 |
| 261.9 | 58.7 | 160 | 210 |
| 270 | 60.4 | 171.9 | 142 |
| 138.3 | 83.5 | 172.1 | 148 |
| 13.5 | 289.4 | 183.6 | 256 |
| 18.1 | 299.9 | 197.9 | 123 |
| 118 | 110.5 | 300.7 | 233 |

**4. CONCLUSIONES**

Como conclusión de la práctica, el alumno debe saber aplicar correctamente los parámetros estadísticos básicos en climatología.