**UNIVERSIDAD NACIONAL AGRARIA LA MOLINA - ESCUELA DE POSGRADO**

**DOCTORADO EN INGENIERÍA Y CIENCIAS AMBIENTALES**

****

**DISEÑO Y ANÁLISIS DE EXPERIMENTOS EN INGENIERÍA Y CIENCIAS AMBIENTALES**

**Actividad:** Práctica

**Docente:**

Ph.D. Christian René Encina Zelada

**Integrantes:**

-Agatha Prado Gárate

-Gustavo De la Cruz Montalvo

-Jhonsy Omar Silva López

-José Augusto Zevallos Ruiz

**Lima – Perú**

**09 de setiembre del 2024**

**Práctica N.º 02:  
Test t - Student**

**CONTENIDO:**

[I. INTRODUCCIÓN 2](#_Toc176212196)

[II. OBJETIVO 2](#_Toc176212197)

[III. MARCO TEÓRICO 2](#_Toc176212198)

[IV. METODOLOGÍA 4](#_Toc176212199)

[V. RESULTADOS Y DISCUSIONES 5](#_Toc176212200)

[VI. CONCLUSIONES Y RECOMENDACIONES 7](#_Toc176212201)

[VII. REFERENCIAS BIBLIOGRÁFICAS 8](#_Toc176212202)

[VIII. ANEXOS 8](#_Toc176212203)

# INTRODUCCIÓN

El test estadístico t de Student es una herramienta ampliamente utilizada para comparar las medias de dos grupos, con el objetivo de determinar si existe una diferencia significativa entre ellas. En el análisis de variables climáticas, como la precipitación y la temperatura, el uso de este test permite evaluar si los cambios observados entre diferentes periodos son atribuibles al azar o si se deben a variaciones reales en el clima. Su aplicabilidad es especialmente útil en estudios regionales como en el caso de Chachapoyas, donde las fluctuaciones climáticas pueden tener impactos importantes en la agricultura y el medio ambiente.

El test t se basa en la suposición de que los datos siguen una distribución normal, lo que lo hace ideal para el análisis de las variables mencionada, siempre que se cumplan ciertas condiciones. Al aplicar esta prueba en los datos mensuales de Chachapoyas, se puede determinar si las diferencias observadas entre los periodos analizados son estadísticamente significativas o si podrían atribuirse a fluctuaciones naturales del clima.

# OBJETIVO

**Objetivo general**

* Evaluar consistencia de medias en las variables de precipitación y temperatura de la estación meteorológica EMA de Chachapoyas (2022-2023) utilizando el software R, asegurando la precisión y validez de los resultados.

**Objetivos específicos**

* Recopilar y procesar datos de temperatura y precipitación para Chachapoyas en el periodo 2022-2023.
* Evaluación de medias a través del test estadístico t- Student a partir de la población de datos mensuales de temperatura y precipitación.

# MARCO TEÓRICO

**Datos climatológicos**

Para el análisis del clima local y regional, los datos meteorológicos de una estación meteorológica son esenciales, ya que brindan información detallada sobre variables como temperatura, humedad, precipitación, velocidad del viento y radiación solar. Estos datos se recopilan continuamente y se utilizan para una variedad de propósitos, incluida la agricultura y la gestión del agua, la planificación urbana y la modelación climática. Según Hidalgo et al. (2020), "la precisión de los datos de una estación meteorológica es crucial para la validación de modelos climáticos y para la calibración de los sensores remotos utilizados en el monitoreo ambiental". Esto hace que las estaciones meteorológicas sean cruciales tanto para la observación directa del clima como para desarrollo de otras investigaciones a futuro.

**T Student:**

La distribución t es una distribución de probabilidad que estima el valor de la media de una muestra pequeña extraída de una población que sigue una distribución normal en la cual su desviación típica es conocida. La t de Student, inicialmente se diseñó para examinar las diferencias entre dos muestras independientes y pequeñas que tengan distribución normal y homogeneidad en sus varianzas.

Esta prueba se fundamenta en dos premisas; la primera: en la distribución de normalidad, y la segunda: en que las muestras sean independientes. Permite comparar muestras, N ≤ 30 y/o establece la diferencia entre las medias de las muestras. El análisis matemático y estadístico de la prueba con frecuencia se minimiza para N > 30, utilizando pruebas no paramétricas, cuando la prueba tiene suficiente poder estadístico.

**Serie estacionaria:**

Un proceso estocástico, Yt, es estacionario en sentido estricto si y sólo si:

F[Yt1 , Yt2 , . . . , Ytn ] = F[Yt1+k, Yt2+k, . . . , Ytn+k] ∀ (t1, t2, . . . , tn) y k es decir, si la función de distribución de cualquier conjunto finito de n variables aleatorias del proceso no se altera si se desplaza k periodos en el tiempo.

Si un proceso estocástico es estacionario en covarianza y su distribución es Normal, es estacionario en sentido estricto.

**Serie no estacionaria:**

A diferencia de las series estacionarias, donde la media y la varianza son constantes y finitas y las autocovarianzas no dependen del tiempo sino sólo del número de periodos de separación entre las variables; las series no estacionarias son aquellas que suelen ir cambiando de nivel en el tiempo o porque la varianza no es constante. Cuando una serie no es estacionaria en varianza, es decir, no se puede sostener el supuesto de que ha sido generada por un proceso con varianza constante en el tiempo, la solución es transformar la serie mediante algún método que estabilice la varianza.

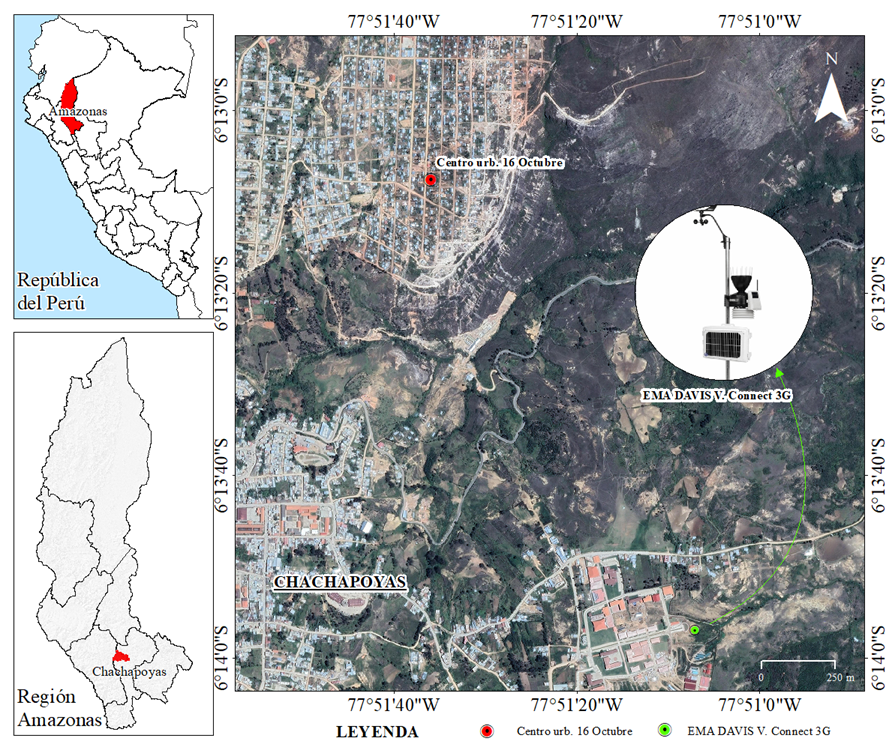
**Diferencia significativa:**

El concepto “significación estadística” se relaciona con la necesidad de “probar hipótesis”. Este proceso se realiza utilizando “pruebas de hipótesis”, las que permiten cuantificar hasta qué punto la variabilidad de la muestra en estudio es responsable de los resultados obtenidos en el estudio. Es así como H0 o hipótesis nula, representa la afirmación de que no hay asociación entre las dos variables; y Ha, o hipótesis alternativa, afirma que existe asociación entre las dos variables. Entonces, la estadística nos permite decidir sobre qué hipótesis debemos elegir, lo que será con el nivel de seguridad que previamente se haya establecido por el equipo de investigación (habitualmente en clínica es 95%). Las pruebas estadísticas funcionan entonces de la siguiente forma: se verifica la magnitud de la diferencia existente entre los grupos a comparar (A y B). Si esta magnitud es mayor que un error estándar definido multiplicado por una seguridad definida, concluimos que la diferencia entre A y B es significativa; por ende, “se rechaza la hipótesis nula” y se “acepta la hipótesis alternativa”.

# METODOLOGÍA

**4.1. Descripción de datos meteorológicos de la EMA Chachapoyas**

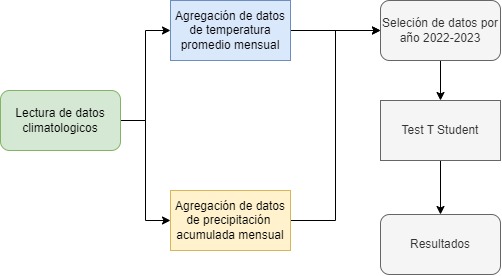
Los datos de la estación meteorológica automática (EMA) de Chachapoyas (**Figura** 1) son recopilados utilizando una estación Davis Connect 3G, que se encuentra instalada en el campus universitario de la Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM). Esta estación registra variables meteorológicas cada 15 minutos, proporcionando información detallada y continua sobre el clima local. Entre las variables medidas se incluyen la temperatura, la humedad relativa, la velocidad y dirección del viento, la precipitación, y la presión atmosférica, entre otras. El uso de una estación automática permite la captura de datos de alta resolución temporal, esencial para estudios climáticos y ambientales en la región, así como para validar modelos meteorológicos y evaluar patrones climáticos locales con mayor precisión.



**Figura 1.** Localización de la estación meteorológica automática – EMA Chachapoyas

**4.3. Análisis de datos**

El flujograma de trabajo describe el proceso de análisis de datos climatológicos comenzando con la lectura de archivos que contienen información de temperatura y precipitación para los años 2022 y 2023. Una vez cargados los archivos, los datos se agregan a un formato mensual, donde para la temperatura se calcula el promedio mensual, mientras que para la precipitación se calcula la suma mensual. Posteriormente, se seleccionan los datos mensuales específicos de los años 2022 y 2023 tanto para temperatura como para precipitación. Con estos datos mensuales, se realiza un análisis estadístico mediante el test de t de Student para evaluar si existen diferencias significativas en las medias entre los dos años. Finalmente, se obtienen los resultados del test para determinar si la diferencia de medias es significativa o no.

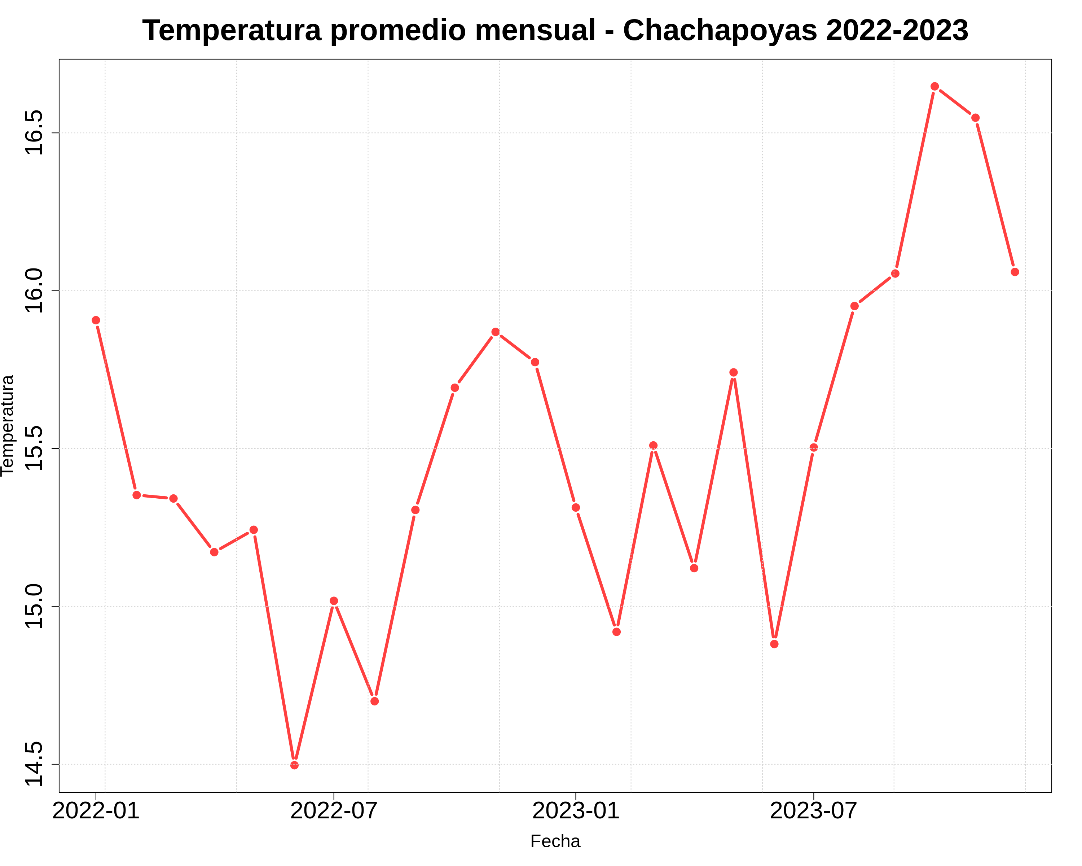


**Figura 2.** Flujograma metodológico empleado en el presente trabajo

# RESULTADOS Y DISCUSIONES

* + - * 1. **TEMPERATURA**

El resultado del test de t de Student para dos muestras indica si existe una diferencia estadísticamente significativa entre las medias de temperatura de los años 2022 y 2023. Aquí está la interpretación detallada de los resultados:

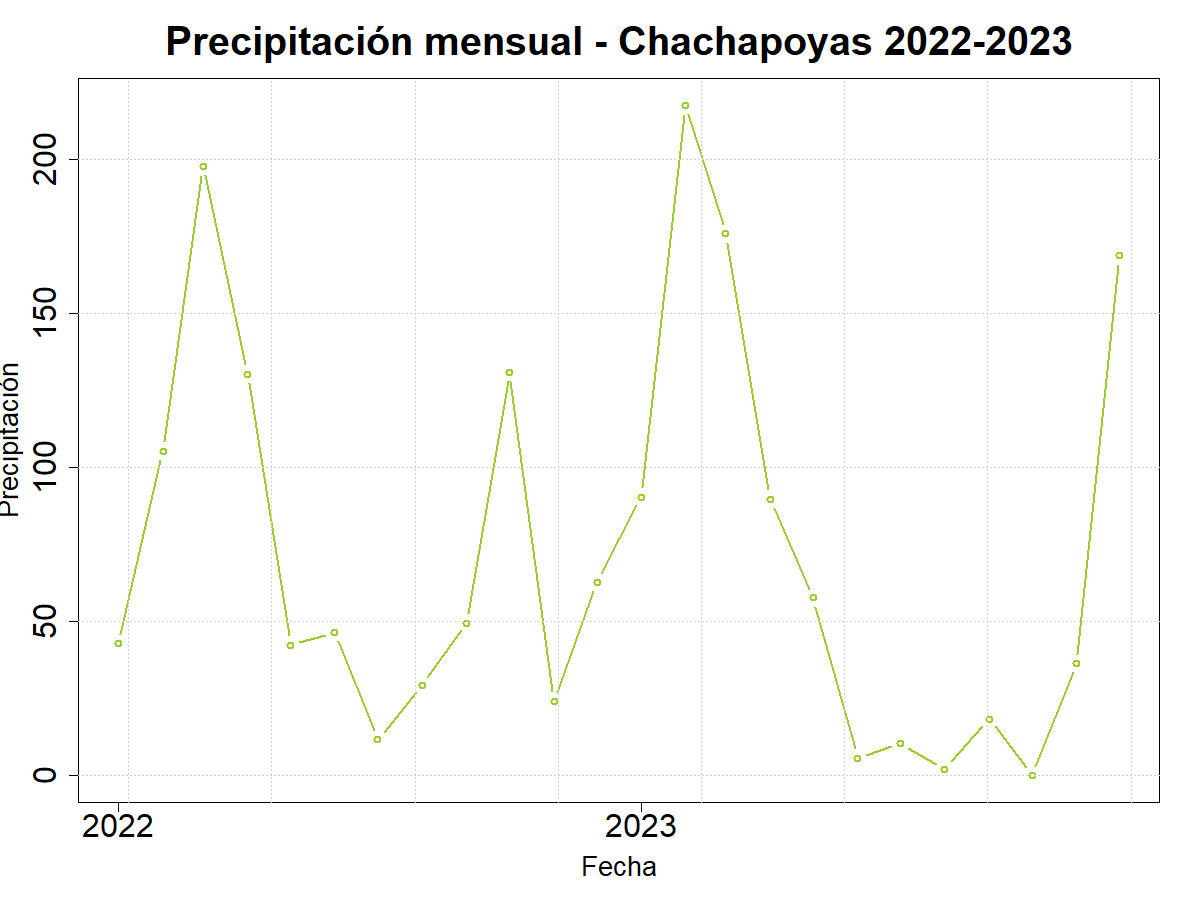


**Figura 3.** Temperatura mensual promedio 2022-2023

**Interpretación de los Resultados:**

1. **Hipótesis nula (H0):** No hay diferencia significativa en las medias de temperatura entre los años 2022 y 2023. En otras palabras, la diferencia en las medias de temperatura es igual a 0.
2. **Hipótesis alternativa (H1):** Existe una diferencia significativa en las medias de temperatura entre los años 2022 y 2023 (la diferencia en las medias no es igual a 0).
3. **Valor del estadístico t:** t = -1.7187
   * Este valor de t indica la magnitud y la dirección de la diferencia entre las medias de las dos muestras. Un valor negativo sugiere que la media de la primera muestra (2022) es menor que la media de la segunda muestra (2023). La magnitud de t no es muy grande, lo que indica que la diferencia entre las medias no es considerablemente grande.
4. **Grados de libertad (df):** df = 22
   * Los grados de libertad están relacionados con el tamaño de la muestra y afectan la distribución de la estadística de prueba.
5. **Valor p (p-value):** p-value = 0.09971
   * El valor p es mayor que 0.05 (nivel de significancia comúnmente utilizado). Esto significa que no hay suficiente evidencia para rechazar la hipótesis nula. En otras palabras, la diferencia observada en las medias no es estadísticamente significativa al nivel del 5%. Sin embargo, el valor p está relativamente cerca de 0.1, lo que podría considerarse marginalmente significativo en algunos contextos.
6. **Intervalo de confianza del 95%:** (-0.8052, 0.0754)
   * Este intervalo de confianza indica que la verdadera diferencia en las medias de temperatura entre 2022 y 2023 podría estar entre -0.8052 y 0.0754. Dado que el intervalo incluye 0, esto respalda aún más la conclusión de que no hay una diferencia significativa en las medias de temperatura entre los dos años.
7. **Medias de las muestras:**
   * **2022 (mean of x):** 15.32258
   * **2023 (mean of y):** 15.68747
   * Las medias de las temperaturas de 2022 y 2023 son 15.32 y 15.69, respectivamente. Esto sugiere que, en promedio, las temperaturas fueron ligeramente más altas en 2023 que en 2022, pero la diferencia no es estadísticamente significativa.
     + - 1. **PRECIPITACIÓN**

El resultado del *Two Sample t-test* (prueba t de dos muestras) proporciona información sobre si existe una diferencia estadísticamente significativa entre las precipitaciones mensuales de los años 2022 y 2023.



**Figura 4.** Precipitación acumulada mensual 2022-2023

**Interpretación de los Resultados**

1. **Valor de t (t-statistic)**: t = 0.0012166  
   El valor de t es cercano a cero, lo que indica que no hay una diferencia significativa entre las medias de las dos muestras (precipitación de 2022 y 2023).
2. **Grados de libertad (df)**: df = 22  
   Este valor indica los grados de libertad para la prueba t, que depende del tamaño de las muestras.
3. **p-value**: p-value = 0.999  
   El p-valor es mucho mayor que el nivel de significancia comúnmente utilizado (α = 0.05). Esto significa que no hay suficiente evidencia para rechazar la hipótesis nula. En otras palabras, la diferencia entre las medias de las precipitaciones mensuales de 2022 y 2023 no es estadísticamente significativa.
4. **Hipótesis alternativa**: true difference in means is not equal to 0  
   La hipótesis alternativa establece que la verdadera diferencia en las medias no es igual a cero. Sin embargo, debido al alto p-valor, no podemos aceptar esta hipótesis.
5. **Intervalo de confianza del 95%**: -56.78665 a 56.85332  
   El intervalo de confianza para la diferencia de medias incluye el cero (-56.79 a 56.85), lo que indica que la diferencia real entre las medias podría ser cero o cualquier otro valor dentro de este rango. Esto refuerza la conclusión de que no hay una diferencia significativa.
6. **Estimaciones de las medias de las muestras**:
   * **Media de 2022 (mean of x)**: 72.85
   * **Media de 2023 (mean of y)**: 72.82
7. Las medias de las precipitaciones mensuales para 2022 y 2023 son casi idénticas (72.85 mm para 2022 y 72.82 mm para 2023), lo que confirma que las precipitaciones promedio fueron muy similares en ambos años.

# CONCLUSIONES Y RECOMENDACIONES

La prueba t es adecuada para situaciones donde se desea evaluar si las diferencias entre los grupos son reales o si pueden atribuirse a variaciones aleatorias. Sin embargo, es importante considerar el tamaño de la muestra y las suposiciones sobre la distribución de los datos para asegurar la validez de los resultados.

Con un valor p de 0.09971, no hay suficiente evidencia para afirmar que la diferencia en las medias de temperatura entre los años 2022 y 2023 es significativa al nivel del 5%. La diferencia observada puede deberse al azar, y no podemos concluir que haya un cambio significativo en las temperaturas entre estos dos años basándonos en estos datos.

No hay evidencia estadística significativa para afirmar que la precipitación mensual promedio de los años 2022 y 2023 sea diferente. Esto sugiere que, en promedio, la cantidad de precipitación fue similar en ambos años. El análisis de la prueba t respalda la hipótesis de que no hubo cambios significativos en las precipitaciones mensuales entre 2022 y 2023.

Dado que no se encontraron diferencias significativas en la precipitación ni en la temperatura, sería recomendable monitorear continuamente estas variables climáticas en futuros periodos. Si bien los datos actuales no muestran un cambio estadísticamente significativo, el clima es un sistema complejo y dinámico, por lo que sería prudente realizar un análisis más prolongado, abarcando varios años, para detectar posibles tendencias a largo plazo. Además, podría ser útil complementar estos estudios con otros factores climáticos o utilizar pruebas más sensibles para captar cambios sutiles.

# REFERENCIAS BIBLIOGRÁFICAS

Gozáles-Casilmiro, M. P. (n.d.). *Análisis de series temporales: Modelos ARIMA* (Universidad del País Vasco (ed.); 04–09 ed.). Facultad de Ciencias económicas y empresariales. https://gc.scalahed.com/recursos/files/r161r/w25599w/L1EF118\_U3\_W2.pdf

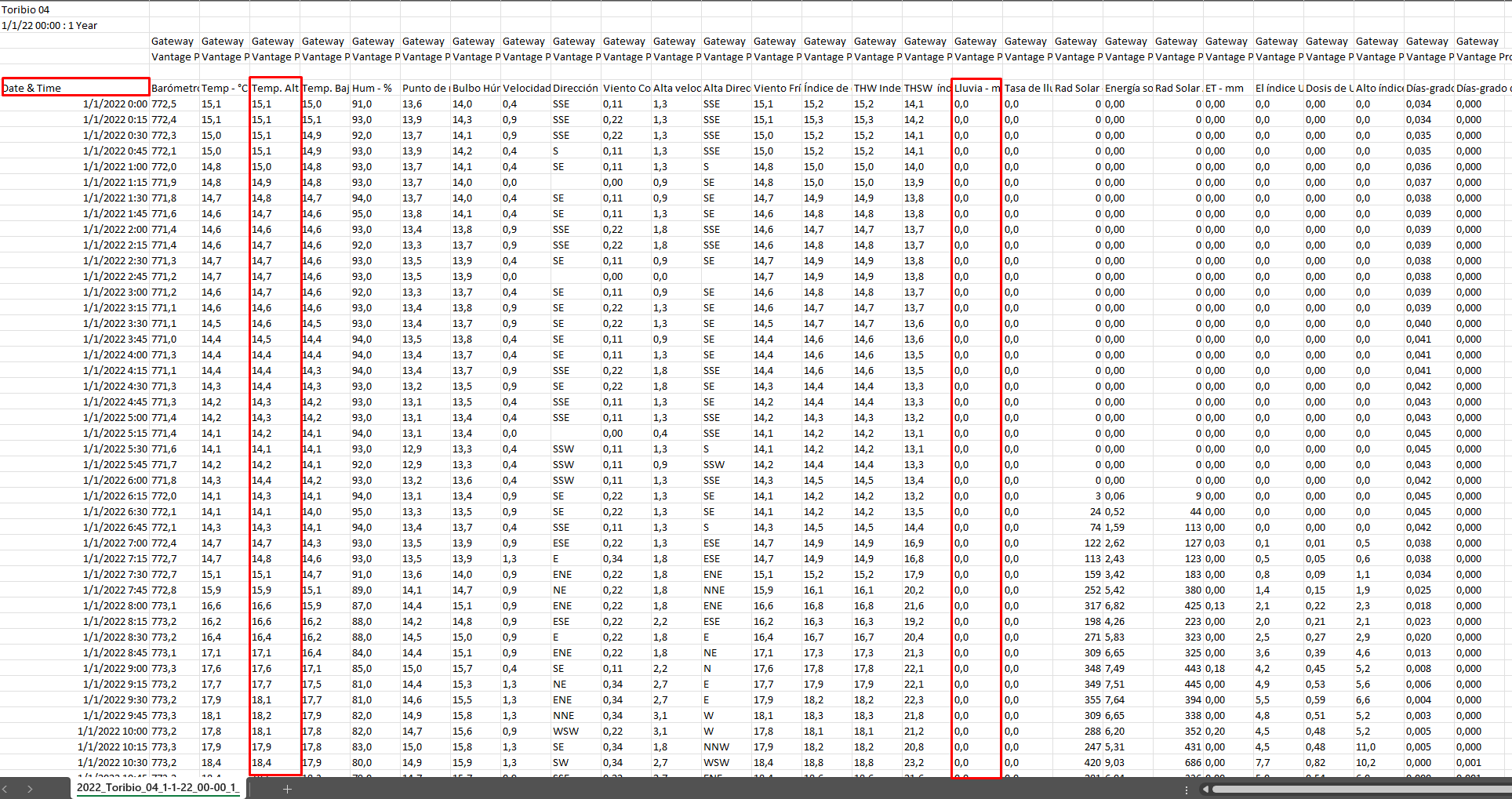
Sánchez Turcios, Reinaldo Alberto. (2015). t-Student: Usos y abusos. Revista mexicana de cardiología, 26(1), 59-61. Recuperado en 09 de septiembre de 2024, de <http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-21982015000100009&lng=es&tlng=es>

(Gozáles-Casilmiro, n.d.)

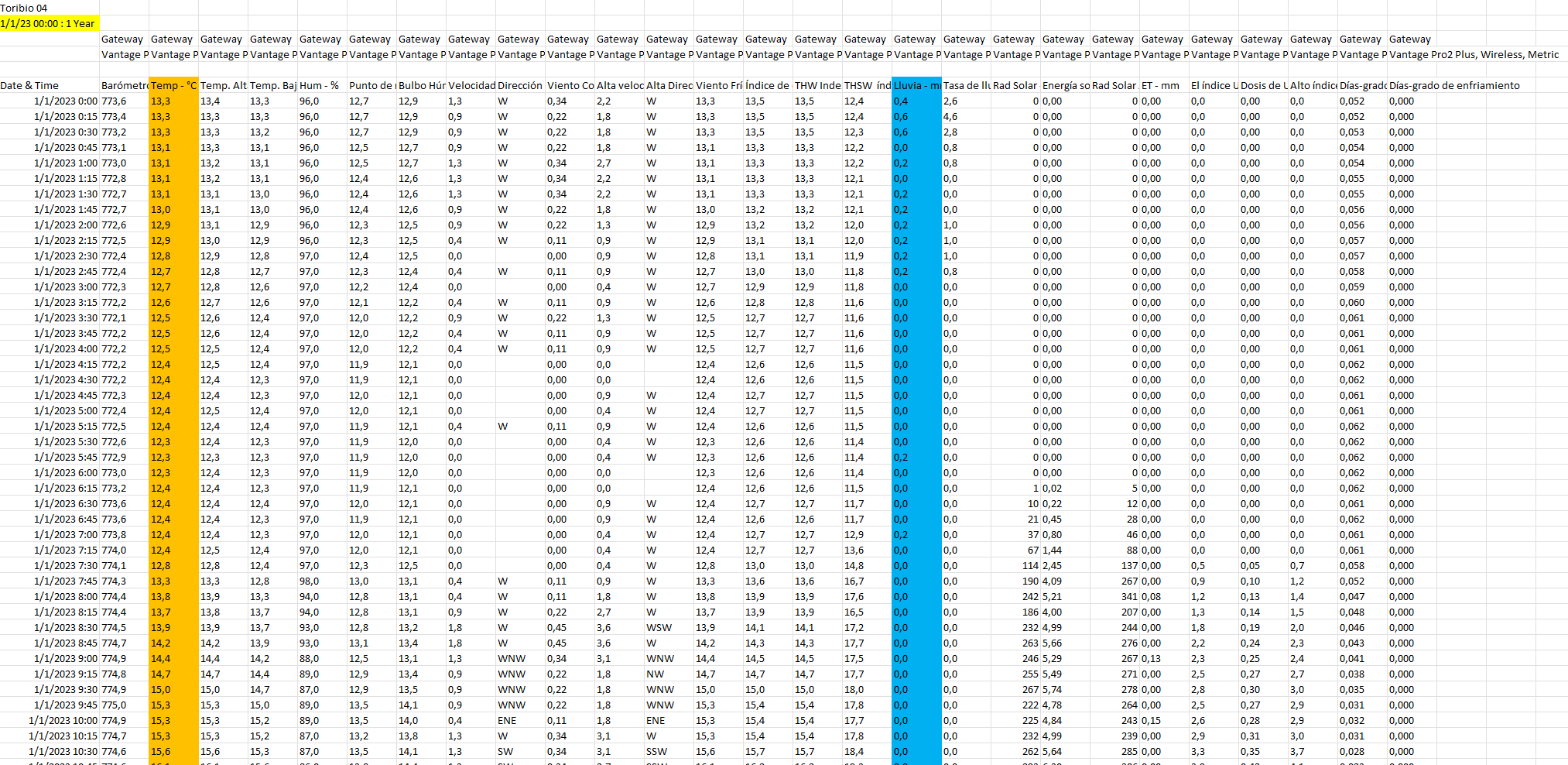
Gozáles-Casilmiro, M. P. (n.d.). Análisis de series temporales: Modelos ARIMA (Universidad del País Vasco (ed.); 04–09 ed.). Facultad de Ciencias económicas y empresariales. https://gc.scalahed.com/recursos/files/r161r/w25599w/L1EF118\_U3\_W2.pdf

# ANEXOS

**Anexo 1.** Datos meteorológicos de 2022 y 2023



Datos meteorológicos de 2022



Datos meteorológicos de 2023

**Anexo 2.** Script desarrollado en el software R

setwd**(**"D:/Proyectos\_GitHub/Diseno\_experimental/tarea2"**)**

#librerias para series de tiempo

library**(**lubridate**)**

library**(**zoo**)**

#librerias para graficos

library**(**plotly**)**

library**(**ggplot2**)**

library**(**readr**)** # Asegúrate de tener esta librería para usar `parse\_number`

#aa1 es temperatura del ano 2022

#aa2 es temperatura del ano 2023

aa1**<-**read.csv**(**'2022\_Toribio\_04\_1-1-22\_00-00\_1\_Year\_1719460093\_v2.csv', header **=** F,skip**=**6**)**

aa1**<-**aa1**[**1**:(**dim**(**aa1**)[**1**]-**1**)**,**]**

aa2**<-**read.csv**(**'2023\_Toribio\_04\_1-1-23\_00-00\_1\_Year\_1719460101\_v2.csv', header **=** F,skip**=**6**)**

aa2**<-**aa2**[**1**:(**dim**(**aa2**)[**1**]-**1**)**,**]**

dates1 **<-** as.POSIXct**(**aa1**$**V1, format **=** "%d/%m/%Y %H:%M"**)**

dates2 **<-** as.POSIXct**(**aa2**$**V1, format **=** "%d/%m/%Y %H:%M"**)**

# Combinar datos

aa **<-** rbind**(**aa1, aa2**)**

dates **<-** c**(**dates1, dates2**)**

# Crear objeto zoo para la temperatura (V3 se supone que es la columna de temperatura)

datazoo **<-** zoo**(**readr**::**parse\_number**(**aa**$**V3**)** **/** 10, order.by **=** dates**)**

# Calcular la media mensual

mes\_temp **<-** aggregate**(**datazoo, format**(**index**(**datazoo**)**, "%Y-%m"**)**, mean**)**

# Convertir a yearmon

mes\_temp **<-** zoo**(**coredata**(**mes\_temp**)**, order.by **=** as.yearmon**(**rownames**(**mes\_temp**)))**

# Separar las temperaturas de 2022 y 2023

temp\_2022 **<-** mes\_temp**[**1**:**12**]**

temp\_2023 **<-** mes\_temp**[**13**:**24**]**

# Realizar el test de t de Student

t\_test\_result **<-** t.test**(**coredata**(**temp\_2022**)**, coredata**(**temp\_2023**)**, var.equal **=** **TRUE)**

print**(**t\_test\_result**)**

#Precipitacion

datazoo **<-** zoo**(**readr**::**parse\_number**(**aa**$**V18**)/**10, order.by **=** dates**)**

yyddmm**=**paste0**(**sprintf**(**"%02d", year**(**dates**))**,sprintf**(**"%02d", month**(**dates**))**,sprintf**(**"%02d", day**(**dates**)))**

yyddmm2**=**c**(**rep**(**"211231",28**)**,yyddmm**[**1**:(**length**(**yyddmm**)-**28**)])**

df**<-**data.frame**(**data**=**datazoo,dates**=**dates,yyddmm**=**yyddmm2**)**

day\_pp**<-**aggregate**(**datazoo **~** yyddmm, data**=**df, FUN**=**sum**)**

day\_pp**<-**day\_pp**[**2**:**dim**(**day\_pp**)[**1**]**,**]**

str**(**day\_pp**$**datazoo**)**

#---------------- Estimación de acumulados diarios ----------------------------

datesdays**<-**seq**(**as.Date**(**'2022-01-01'**)**,as.Date**(**'2023-12-31'**)**,by**=**'days'**)**

datazoo\_day **<-** zoo**(**day\_pp**$**datazoo, order.by **=** datesdays**)**

plot**(**datazoo\_day,xlab**=**"Fecha",ylab**=**"Precipitación",main**=**"Precipitación diaria - Chachapoyas 2022-2023",

cex.lab**=**1.7, cex.axis**=**2, cex.main**=**2.5,col**=**"olivedrab3",lwd**=**2.5**)**

grid**()**

# Corregir la función aggregate para sumar la precipitación mensual

monthly\_sum **<-** aggregate**(**day\_pp**$**datazoo, by **=** list**(**format**(**index**(**datazoo\_day**)**, "%Y-%m"**))**, FUN **=** sum**)**

names**(**monthly\_sum**)** **=** c**(**'Fecha','pp'**)**

# Separar precipitacion de 2022 y 2023

pcp\_2022 **<-** monthly\_sum**[**1**:**12, **]**

pcp\_2023 **<-** monthly\_sum**[**13**:**24, **]**

# Corregir la conversión de las fechas a Date

monthly\_sum**$**Fecha **<-** as.Date**(**paste0**(**monthly\_sum**$**Fecha, "-01"**))**

# Graficar la precipitación mensual

plot**(**x **=** monthly\_sum**$**Fecha, y **=** monthly\_sum**$**pp,

xlab **=** "Fecha", ylab **=** "Precipitación",

main **=** "Precipitación mensual - Chachapoyas 2022-2023",

cex.lab **=** 1.7, cex.axis **=** 2, cex.main **=** 2.5,

col **=** "olivedrab3", lwd **=** 2.5, type **=** "b"**)** # type = "b" para líneas y puntos

grid**()**

# Realizar el test de t de Student

pp\_test\_result **<-** t.test**(**coredata**(**pcp\_2022**$**pp**)**, coredata**(**pcp\_2023**$**pp**)**, var.equal **=** **TRUE)**

print**(**pp\_test\_result**)**