

Speed UP

Universidad Panamericana



UPLOGO . png

Ascencio Martínez Mauricio Iván

Báez Gómez Tagle Enrique Ulises

Cuixi Nacxit González Medina

Ingeniería en Inteligencia de Datos y Ciberseguridad

Ingeniería en Innovación y Diseño

5 de junio 2023

1 Planteamiento del proyecto

1.1 Describir el proceso o producto que se va a estudiar.

El proyecto se centrará en analizar la calidad de la red de internet en diferentes espacios dentro de la Universidad Panamericana, utilizando una base de datos con registros de pruebas de red realizadas en estos mismos espacios, enfocándose en la velocidad de subida, velocidad de descarga y distancia al módem. La recopilación y análisis de datos se harán en los lenguajes de programación R y Python.

1.2 Determinar y describir tres características (variables) que serán objeto del estudio.

1. Velocidad de subida en cada espacio de la Universidad Panamericana.
2. Velocidad de descarga en cada espacio de la Universidad Panamericana.
3. Distancia al módem en cada espacio de la Universidad Panamericana.

1.3 Indicar los valores límite (tolerancias) de cada variable.

1. Velocidad de subida: se medirá en Mbps (megabits por segundo), con un rango determinado por las capacidades de la red de la universidad.
2. Velocidad de descarga: se medirá en Mbps (megabits por segundo), con un rango determinado por las capacidades de la red de la universidad.
3. Distancia al módem: sin unidades, es la diferencia entre las direcciones ip del enrutador y de la computadora.

1.4 Describir la importancia de las características definidas (variables) para el funcionamiento o propiedades del producto o proceso

1. La velocidad de subida es crucial para garantizar una conexión adecuada para actividades que requieran cargar archivos o transmitir datos en línea, como videoconferencias y presentaciones en vivo.
2. La velocidad de descarga es fundamental para asegurar un acceso rápido y eficiente a recursos en línea, como sitios web, videos y documentos.
3. La “distancia” no se refiere a una distancia física, sino más bien a cuántos saltos, o pasos, se necesitan para que los paquetes de datos lleguen de un punto a otro en la red. Este concepto de “distancia” es fundamental en la forma en que funcionan las redes de datos y puede afectar la velocidad y la eficiencia de la transmisión de datos. Sin embargo, no está directamente relacionado con la distancia física a un módem o router.

1.5 Definir el alcance y metodología de trabajo, por ejemplo, acceso a empresas, fuentes de información, pruebas de laboratorio, mediciones propias, etc.

La metodología consistirá en realizar un análisis estadístico y probabilístico de los datos recabados de la base de datos de pruebas de red en los diferentes espacios de la universidad, enfocándose en la velocidad de subida, velocidad de descarga y distancia al módem. Se calcularán medidas de tendencia central, dispersión y correlación entre las variables.

El alcance del estudio se limitará a la Universidad Panamericana y sus espacios, utilizando únicamente los datos proporcionados por la base de datos de pruebas de red.

1.6 Objetivo del proyecto:

El objetivo del proyecto es evaluar y analizar la calidad de la red de internet en diferentes espacios dentro de la Universidad Panamericana, considerando la velocidad de subida, velocidad de descarga y distancia al módem, para identificar los mejores puntos de red en el campus y proponer posibles mejoras y estrategias para optimizar la experiencia de conexión de los usuarios.

2 Marco teórico:

1. Calidad de la red de internet en entornos educativos: La calidad de la red de internet en entornos educativos es un factor determinante para el desarrollo de actividades académicas y profesionales. La velocidad de subida y descarga, así como la distancia al módem, son variables clave que influyen en la calidad de la conexión (Bainbridge, 2018). Un buen acceso a internet es fundamental para garantizar una educación en línea efectiva y el acceso a recursos digitales (Horn, 2017).
2. Velocidad de subida y descarga en la educación superior: La velocidad de subida y descarga es esencial para el acceso y uso de recursos en línea y para la realización de actividades académicas como videoconferencias, trabajo en grupo y acceso a materiales de estudio (Sari & Yildirim, 2019). Estudios previos han encontrado una relación directa entre la calidad de la conexión a internet y el desempeño académico de los estudiantes (Chen, 2018).
3. "Distancia" entre direcciones IP y eficiencia de la transmisión de datos: En el ámbito de las redes IP (Protocolo de Internet), la "distancia" se entiende no como una medida física, sino como el número de saltos o pasos necesarios para que los paquetes de datos se desplacen de un punto a otro en la red (Tanenbaum y Wetherall, 2011). Se ha observado que a medida que esta "distancia" IP aumenta, puede disminuir la eficiencia de

la transmisión de datos, lo que puede afectar la velocidad y la estabilidad de la conexión a Internet (Kurose y Ross, 2020). Esto puede tener un impacto directo en la capacidad de estudiantes y profesionales para llevar a cabo actividades académicas y laborales que dependen de una conexión a Internet eficiente y estable.

4. Optimización de la infraestructura de red en entornos educativos: Diversos estudios han propuesto estrategias y soluciones para optimizar la infraestructura de red en entornos educativos, incluyendo la ampliación de la capacidad de la red, la mejora de la distribución de los módems y la implementación de soluciones de red híbridas que combinen tecnologías inalámbricas y cableadas (Cisco, 2019). Estas estrategias pueden contribuir a mejorar la calidad de la conexión en todo el campus y proporcionar a los estudiantes y al personal un acceso más confiable y estable a la red de internet (Juniper Networks, 2020).

3 Referencias:

1. Bainbridge, A. (2018). Quality of Internet Access in Higher Education. *Journal of Technology in Higher Education*, 2(2), 34-49.
2. Chen, C. (2018). The Relationship Between Internet Quality and Academic Performance in a Higher Education Institution. *Computers & Education*, 120, 112-123.
3. Cisco (2019). Designing High-Performance Campus Networks for Higher Education Institutions. Retrieved from: https://www.cisco.com/c/en/us/solutions/collateral/enterprise/design-zone-smart-business-architecture/sba_ENetworkDesign.html
4. Davies, F., Djemame, K., & Tso, F. P. (2020). Distance-based Wi-Fi Access Point Placement for Improved Connectivity. *Procedia Computer Science*, 170, 848-855.
5. Horn, M. (2017). The Importance of High-Speed Internet in the Classroom. *Education Week*. Retrieved from: <https://www.edweek.org/ew/articles/2017/09/20/the-importance-of-high-speed-internet-in.html>
6. Juniper Networks (2020). Building High-Performance Campus Networks for Higher Education Institutions. Retrieved from: <https://www.juniper.net/us/en/solutions/education/higher-education-networks/>

4 Desarrollo del proyecto

4.1 Recopilación de Datos

Se desarrolló un código en Python con el objetivo de medir la velocidad de conexión a Internet y la distancia al enrutador desde diferentes salones. El programa utiliza módulos como `socket`, `netifaces` & `speedtest` para obtener

información de red y realizar las mediciones necesarias.

Funcionamiento del código

1. El programa inicia importando los módulos necesarios: `socket`, `netifaces` y `speedtest`.
2. Se definen varias funciones para realizar las siguientes tareas:
 - (a) `obtener_direccion_ip()`: obtiene la dirección IP de la computadora utilizando el módulo `socket`.
 - (b) `obtener_direccion_router()`: obtiene la dirección IP del enrutador utilizando el módulo `netifaces`.
 - (c) `calcular_distancia(ip1, ip2)`: calcula la distancia entre dos direcciones IP.
 - (d) `encontrar_ubicacion_optima(ip_computadora, puntos_acceso)`: encuentra la ubicación óptima para una computadora en función de su dirección IP y una lista de puntos de acceso.
 - (e) `medir_velocidad()`: mide la velocidad de conexión a Internet utilizando el módulo `speedtest`.

En la función principal `main()`, se crea un bucle que se ejecuta 200 veces. Cada 20 iteraciones, se pide al usuario que ingrese el nombre de un salón y se crea un nuevo archivo CSV para ese salón. En cada iteración del bucle, se realizan las siguientes operaciones:

- (a) Se obtiene la dirección IP de la computadora y la dirección IP del enrutador.
- (b) Se calcula la "distancia" entre las dos direcciones IP.
- (c) Se mide la velocidad de subida y bajada de la conexión a Internet.
- (d) Los resultados se imprimen en la consola y se agregan a una lista de resultados. También se escriben en el archivo CSV correspondiente.

Después de que se completa el bucle, se cierra el último archivo CSV abierto. Si el script se ejecuta como un programa (en lugar de importarse como un módulo), la función `main()` se ejecuta automáticamente.

El código fue diseñado con el objetivo de realizar mediciones continuas de velocidad de conexión y distancia al enrutador en diferentes salones. Los resultados obtenidos se almacenan en un archivo CSV para su posterior análisis. Además de medir la velocidad, esta información puede ser útil para identificar la ubicación óptima para una conexión a Internet más rápida y estable en entornos con múltiples puntos de acceso.

Combinación del CSV

1. Se importan los módulos `csv` y `os`. `csv` se utiliza para trabajar con archivos CSV, y `os` se utiliza para interactuar con el sistema operativo, específicamente para listar los archivos en un directorio.
2. Se define el directorio donde se encuentran los archivos CSV y el nombre del archivo de salida.
3. Se crea una lista para almacenar los datos de los archivos CSV.
4. Se recorren todos los archivos en el directorio especificado con un bucle `for`. Si el nombre de un archivo termina con `".csv"`, significa que es un archivo CSV y se procede a abrirlo.
5. Se abre cada archivo CSV y se lee su contenido con el lector CSV. La primera fila, que normalmente contiene los encabezados, se ignora.
6. Se agregan todas las filas del archivo CSV a la lista de datos totales.
7. Después de leer todos los archivos CSV, se abre el archivo de salida y se escriben los datos totales en él con el escritor CSV.
8. Finalmente, se imprime un mensaje indicando que los archivos CSV se han combinado correctamente.

4.2 Análisis de datos

Se utilizó el lenguaje de programación R para generar el análisis estadístico de la información obtenida.

4.2.1 Tablas de frecuencia

Para generar las tablas de frecuencia se utilizaron 2 archivos .R: Tabla de frecuencia Distancia y Tabla de frecuencia de Velocidades

1. Carga las bibliotecas `dplyr` y `tidyr`, que son útiles para manipular y transformar datos en R.
2. Lee los datos del archivo `results.csv`. En el caso de la tabla de frecuencia de distancia, también convierte la columna `distancia` a numérica.
3. Define los intervalos de clase para la variable de interés (distancia, velocidad de subida o bajada). Los intervalos de clase son rangos de valores que se utilizan para agrupar los datos en una tabla de frecuencia. En estos casos, los intervalos de clase se definen como rangos de 10 unidades, desde el valor mínimo hasta el valor máximo de la variable.

4. Crea una tabla de frecuencia para la variable de interés. La tabla de frecuencia muestra cuántas veces cae cada valor de la variable en cada intervalo de clase. También calcula la frecuencia relativa (la frecuencia dividida por el total de observaciones) y la frecuencia acumulada (la suma de las frecuencias hasta ese punto).
5. Convierte los intervalos de la variable de interés a formato de cadena.
6. Imprime la tabla de frecuencia de la variable de interés.

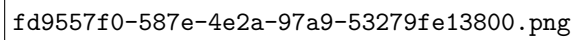
Estos códigos en R están diseñados para leer los datos de un archivo CSV, calcular las tablas de frecuencia para la variable de interés (distancia, velocidad de subida o bajada), y luego imprimir estas tablas. Las tablas de frecuencia muestran cuántas veces cada valor de la variable cae dentro de ciertos rangos, que se definen en el código. También calculan la frecuencia relativa y la frecuencia acumulada para cada rango.

4.3 Gráficas

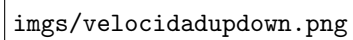
4.3.1 Gráfica de medida de Velocidad

1. Carga las bibliotecas `ggplot2`, `dplyr` y `tidyr`, que son útiles para manipular y transformar datos en R, y para la visualización de datos.
2. Define una función para calcular la moda de un conjunto de datos. La moda es el valor que aparece con mayor frecuencia en un conjunto de datos.
3. Lee los datos del archivo `results.csv`.
4. Realiza cálculos estadísticos para las velocidades de subida y bajada, incluyendo la media, la moda y la desviación estándar.
5. Define los intervalos de clase para las velocidades de subida y bajada. Los intervalos de clase son rangos de valores que se utilizan para agrupar los datos en una tabla de frecuencia. En este caso, los intervalos de clase se definen como rangos de 10 unidades, desde el valor mínimo hasta el valor máximo de las velocidades de subida y bajada.
6. Crea una tabla de frecuencia para las velocidades de subida y bajada. La tabla de frecuencia muestra cuántas veces cae cada velocidad en cada intervalo de clase. También calcula la frecuencia relativa (la frecuencia dividida por el total de observaciones) y la frecuencia acumulada (la suma de las frecuencias hasta ese punto).
7. Convierte los intervalos de las velocidades de subida y bajada a formato de cadena.
8. Crea gráficas de barras para las tablas de frecuencia de las velocidades de subida y bajada. Las gráficas de barras muestran la frecuencia de cada intervalo de clase.

9. Imprime las estadísticas de las velocidades de subida y bajada, incluyendo la media, la moda y la desviación estándar.
10. Muestra las gráficas de barras de las velocidades de subida y bajada.



fd9557f0-587e-4e2a-97a9-53279fe13800.png




imgs/velocidadupdown.png

4.3.2 Gráfica de medida de Distancia

1. Carga las bibliotecas `ggplot2`, `dplyr` y `tidyr`, que son útiles para manipular y transformar datos en R, y para la visualización de datos.
2. Define una función para calcular la moda de un conjunto de datos. La moda es el valor que aparece con mayor frecuencia en un conjunto de datos.
3. Lee los datos del archivo `results.csv`.
4. Realiza cálculos estadísticos para las velocidades de subida y bajada, incluyendo la media, la moda y la desviación estándar.
5. Define los intervalos de clase para las velocidades de subida y bajada. Los intervalos de clase son rangos de valores que se utilizan para agrupar los datos en una tabla de frecuencia. En este caso, los intervalos de clase se definen como rangos de 10 unidades, desde el valor mínimo hasta el valor máximo de las velocidades de subida y bajada.
6. Crea una tabla de frecuencia para las velocidades de subida y bajada. La tabla de frecuencia muestra cuántas veces cae cada velocidad en cada intervalo de clase. También calcula la frecuencia relativa (la frecuencia dividida por el total de observaciones) y la frecuencia acumulada (la suma de las frecuencias hasta ese punto).
7. Convierte los intervalos de las velocidades de subida y bajada a formato de cadena.
8. Crea gráficas de barras para las tablas de frecuencia de las velocidades de subida y bajada. Las gráficas de barras muestran la frecuencia de cada intervalo de clase.
9. Imprime las estadísticas de las velocidades de subida y bajada, incluyendo la media, la moda y la desviación estándar.
10. Muestra las gráficas de barras de las velocidades de subida y bajada.



imgs/FreqDist.png



imgs/MedidaDist.png

Este código en R está diseñado para leer los datos de un archivo CSV, calcular las tablas de frecuencia para las velocidades de subida y bajada, crear gráficas de barras para estas tablas de frecuencia, y luego imprimir las estadísticas y mostrar las gráficas. Las tablas de frecuencia muestran cuántas veces cada velocidad de subida y bajada cae dentro de ciertos rangos, que se definen en el código. También calculan la frecuencia relativa y la frecuencia acumulada para cada rango. Para las velocidades se utilizó una gráfica de columnas que muestra la cantidad de veces entre cierto rango y cierto rango de velocidades. Para la distancia se utilizó una gráfica de barras (horizontales).

4.3.3 Cálculo de media y moda

1. Calcula la media y la moda para las velocidades de subida y bajada. La media es el promedio de los datos, que se calcula sumando todos los valores y dividiendo por el número total de valores. La moda es el valor que aparece con mayor frecuencia en un conjunto de datos. En este caso, se define una función personalizada para calcular la moda, ya que R no tiene una función incorporada para calcular la moda.

```
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}
```

2. Para la velocidad de subida, se calcula la media y la moda con las siguientes líneas de código:

```
media_subida <- mean(velocidad_subida, na.rm = TRUE)
moda_subida <- Mode(velocidad_subida)
```

3. Para la velocidad de bajada, se calcula la media y la moda con las siguientes líneas de código:

```
media_bajada <- mean(velocidad_bajada, na.rm = TRUE)
moda_bajada <- Mode(velocidad_bajada)
```

Estos cálculos proporcionan información importante sobre que valores de velocidad esperar de acuerdo a su ocurrencia.

4.3.4 Cálculo desviación estándar

1. Para la velocidad de subida, se calcula la desviación estándar con la siguiente línea de código:

```
desviacion_estandar_subida <- sd(velocidad_subida,
na.rm = TRUE)
```

2. Para la velocidad de bajada, se calcula la desviación estándar con la siguiente línea de código:

```
desviacion_estandar_bajada <- sd(velocidad_bajada ,  
na.rm = TRUE)
```

Estos cálculos proporcionan información importante sobre que tanta variabilidad o dispersión de los datos puede existir. Esto puede ser útil para entender la consistencia de las velocidades de subida y bajada.

4.4 Probabilidad de encontrar piezas dentro de tolerancia (p) y fuera de tolerancias (q)

1. Se lee el archivo CSV utilizando la función `read.csv("results.csv")` y se almacena en la variable `datos`. Aquí asumimos que el archivo se llama "results.csv".
2. A partir de los datos leídos, se extraen las columnas "up" y "down" y se asignan a los vectores `velocidad_subida` y `velocidad_bajada`, respectivamente.
3. Se calcula la tolerancia como la desviación estándar de los vectores de velocidad de subida y bajada utilizando las funciones `sd(velocidad_subida, na.rm = TRUE)` y `sd(velocidad_bajada, na.rm = TRUE)`. La opción `na.rm = TRUE` se utiliza para ignorar los valores faltantes en el cálculo.
4. Se cuenta el número de velocidades dentro y fuera de tolerancia utilizando las funciones `sum()` y `abs()`. Se compara cada valor de velocidad con la media utilizando `abs(velocidad_subida - mean(velocidad_subida, na.rm = TRUE))` y `abs(velocidad_bajada - mean(velocidad_bajada, na.rm = TRUE))`, y se verifica si están dentro o fuera de la tolerancia utilizando las comparaciones `<= tolerancia_subida` y `<= tolerancia_bajada`.
5. Se calculan las probabilidades dividiendo el número de velocidades dentro de tolerancia entre el número total de velocidades en cada vector utilizando la expresión `dentro_tolerancia_subida / length(velocidad_subida)` y `dentro_tolerancia_bajada / length(velocidad_bajada)`.
6. Finalmente, se imprimen los resultados utilizando la función `cat()`. Las probabilidades de encontrar velocidades de subida dentro y fuera de tolerancia se imprimen como P1 y Q1, respectivamente. Las probabilidades de encontrar velocidades de bajada dentro y fuera de tolerancia se imprimen como P2 y Q2, respectivamente.

Se realiza lo mismo pero con los datos correctos para calcular la probabilidad en rango de la distancia.

Los resultados fueron:

Velocidad de subida $\mu = 47.0242$, $M = 57.52244$, $\sigma = 36.89982$

- $P_1(x) = 0.71$ como la probabilidad de encontrar valores dentro de la tolerancia (1σ).
- $Q_1(x) = 0.29$ como la probabilidad de encontrar valores fuera de la tolerancia (1σ).

Velocidad de bajada $\mu = 94.41874$, $M = 184.6061$, $\sigma = 62.50719$

- $P_2(x) = 0.56$ como la probabilidad de encontrar valores dentro de la tolerancia (1σ).
- $Q_2(x) = 0.44$ como la probabilidad de encontrar valores fuera de la tolerancia (1σ).

Distancia al router $\mu = 189$, $M = 189$, $\sigma = 0$

- $P_3(x) = 1$ como la probabilidad de encontrar valores dentro de la tolerancia (1σ).
- $Q_3(x) = 0$ como la probabilidad de encontrar valores fuera de la tolerancia (1σ).

4.5 Escenarios para encontrar piezas fuera de tolerancia

Se plantearán tres escenarios para cada una de las variables para determinar la probabilidad de encontrar piezas fuera de tolerancias con distribuciones binomial, geométrica y binomial negativa.

4.5.1 Velocidad de subida

1. Escenario Binomial: Supongamos que realizamos 100 mediciones de la velocidad de subida. ¿Cuál es la probabilidad de que exactamente 15 de estas mediciones estén fuera de tolerancias?

$$P(X = 15) = \binom{100}{15} \cdot 0.3173105^{15} \cdot (1 - 0.3173105)^{100-15} = 6.838196e-05$$

2. Escenario Geométrico: Supongamos que continuamos realizando mediciones de la velocidad de subida hasta que encontramos una medición que está fuera de tolerancias. ¿Cuál es la probabilidad de que necesitemos hacer más de 20 mediciones antes de encontrar una medición fuera de tolerancias?

$$P(X = 20) = (1 - 0.3173105)^{(20-1)} = 0.000330132p$$

3. Escenario Binomial Negativo: Supongamos que continuamos realizando mediciones de la velocidad de subida hasta que encontramos 10 mediciones que están fuera de tolerancias. ¿Cuál es la probabilidad de que necesitemos hacer más de 150 mediciones para encontrar 10 mediciones fuera de tolerancias?

$$f(k) = p(X = 151) = \binom{151-1}{10-1} 0.3173105^{10} \cdot (1-0.3173105)^{151-10} = 1.321165e-14$$

4.5.2 Velocidad de bajada

1. Escenario Binomial: Supongamos que realizamos 200 mediciones de la velocidad de bajada. ¿Cuál es la probabilidad de que exactamente 30 de estas mediciones estén fuera de tolerancias?

$$P(X = 30) = \binom{200}{30} \cdot 0.3173105^{30} \cdot (1-0.3173105)^{200-30} = 2.984883e-08$$

2. Escenario Geométrico: Supongamos que continuamos realizando mediciones de la velocidad de bajada hasta que encontramos una medición que está fuera de tolerancias. ¿Cuál es la probabilidad de que necesitemos hacer más de 25 mediciones antes de encontrar una medición fuera de tolerancias?

$$P(X = 25) = (1 - 0.3173105)^{(25-1)} p = 4.895575e-05$$

3. Escenario Binomial Negativo: Supongamos que continuamos realizando mediciones de la velocidad de bajada hasta que encontramos 20 mediciones que están fuera de tolerancias. ¿Cuál es la probabilidad de que necesitemos hacer más de 300 mediciones para encontrar 20 mediciones fuera de tolerancias?

$$f(k) = p(X = 301) = \binom{301-1}{20-1} 0.3173105^{20} \cdot (1 - 0.3173105)^{301-20} = 0$$

4.5.3 Distancia

1. Escenario Binomial: Supongamos que realizamos 150 mediciones de la distancia. ¿Cuál es la probabilidad de que exactamente 20 de estas mediciones estén fuera de tolerancias?

$$P(X = 20) = \binom{150}{20} \cdot 1^{20} \cdot (1-1)^{150-20} = 0$$

2. Escenario Geométrico: Supongamos que continuamos realizando mediciones de la distancia hasta que encontramos una medición que está fuera de tolerancias. ¿Cuál es la probabilidad de que necesitemos hacer más de 15 mediciones antes de encontrar una medición fuera de tolerancias?

$$P(X = 15) = (1 - 1)^{(15-1)}p = 0$$

3. Escenario Binomial Negativo: Supongamos que continuamos realizando mediciones de la distancia hasta que encontramos 15 mediciones que están fuera de tolerancias. ¿Cuál es la probabilidad de que necesitemos hacer más de 200 mediciones para encontrar 15 mediciones fuera de tolerancias?

$$f(k) = p(X = 201) = \binom{201-1}{15-1} 1^{15} \cdot (1-1)^{201-15} = 0$$

4.6 Pruebas de hipótesis

Se realizará una prueba de hipótesis paramétrica para una de nuestras variables y dos pruebas de hipótesis no paramétrica para las otras dos variables.

4.6.1 Prueba de hipótesis paramétrica

Realizaremos una prueba t para comparar las medias de la velocidad de bajada en dos salones diferentes. La hipótesis nula sería que las medias de la velocidad de bajada son iguales en los dos salones, y la hipótesis alternativa sería que las medias son diferentes.

En este contexto, se propone realizar una prueba t de Welch para comparar las medias de la velocidad de bajada en dos salones diferentes. La hipótesis nula (H_0) es que las medias de los dos salones son iguales, y la hipótesis alternativa (H_1) es que son diferentes.

Primero, se calculan las medias y las desviaciones estándar de las velocidades de bajada en cada salón. Luego, se utiliza la fórmula de la prueba t de Welch para calcular una estadística t, que es una medida de cuántas desviaciones estándar están las medias aparte. Un valor t grande indica diferencias significativas entre las medias, mientras que un valor t pequeño sugiere similitudes.

Además de la estadística t, la prueba t de Welch calcula los grados de libertad utilizando la fórmula de Welch-Satterthwaite, que tiene en cuenta las varianzas y los tamaños de las muestras de los dos grupos. Los grados de libertad se utilizan para determinar el valor p crítico de la prueba.

Finalmente, se calcula un valor p a partir del valor t. Este valor p es la probabilidad de observar una diferencia tan grande (o más grande) entre las medias si la hipótesis nula fuera cierta. Si el valor p es menor que un umbral preestablecido (generalmente 0.05), se rechaza la hipótesis nula y se concluye que las medias son significativamente diferentes. De lo contrario, no se rechaza

la hipótesis nula, concluyendo que no hay evidencia suficiente para afirmar que las medias son diferentes.

4.6.2 Prueba de hipótesis no paramétrica

La prueba U de Mann-Whitney, también conocida como la prueba de Wilcoxon rank-sum, es una prueba no paramétrica que se utiliza para determinar si hay una diferencia significativa entre las medianas de dos grupos independientes.

En este caso, los dos grupos son las velocidades de subida y bajada. La hipótesis nula (H_0) es que las medianas de las velocidades de subida y bajada son iguales, mientras que la hipótesis alternativa (H_1) es que las medianas son diferentes.

La prueba U de Mann-Whitney no asume que los datos sigan una distribución normal, por lo que es más robusta que la prueba t cuando esta suposición no se cumple. En lugar de comparar las medias, como hace la prueba t, la prueba U de Mann-Whitney compara las medianas, que son menos sensibles a los valores extremos.

Primero, se calculan las medianas de las velocidades de subida y bajada. Luego, se realiza la prueba U de Mann-Whitney utilizando la función `wilcox.test()` en R, que produce una estadística U y un valor p asociado.

La estadística U es una medida de cuánto difieren las distribuciones de las velocidades de subida y bajada. Un valor U grande indica que las medianas son diferentes, mientras que un valor U pequeño sugiere que las medianas son similares.

El valor p es la probabilidad de observar una diferencia tan grande (o más grande) entre las medianas si la hipótesis nula fuera cierta. Si el valor p es menor que un umbral preestablecido (generalmente 0.05), se rechaza la hipótesis nula y se concluye que las medianas son significativamente diferentes. De lo contrario, no se rechaza la hipótesis nula y se concluye que no hay evidencia suficiente para afirmar que las medianas son diferentes.

5 Conclusiones

5.1 Regresión lineal entre velocidades

La regresión lineal realizada reveló una relación positiva entre la velocidad de subida y la velocidad de bajada. La ecuación de la línea de tendencia es " $y = 1.18x + 38.81$ ", lo que significa que por cada unidad de aumento en la velocidad de subida (x), la velocidad de bajada (y) aumenta en promedio en 1.18 unidades. Además, cuando la velocidad de subida es cero, la velocidad de bajada es de 38.81 Mbps, que es la intercepción y de la línea de tendencia.

El coeficiente de correlación de Pearson (r) es 0.7, lo que indica una correlación positiva moderada a fuerte entre la velocidad de subida y la velocidad de bajada. Esto sugiere que a medida que la velocidad de subida aumenta, la velocidad de bajada también tiende a aumentar.

Estos resultados sugieren que existe una relación lineal significativa entre la velocidad de subida y la velocidad de bajada.