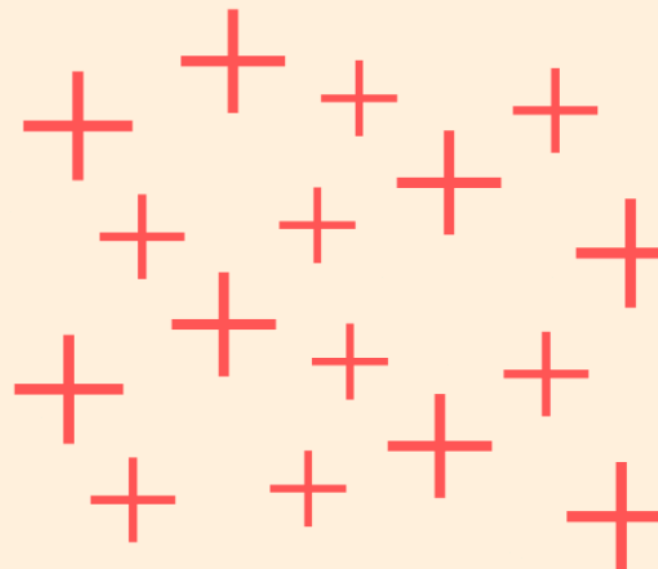


DESEMBRE 2024

ARNAU CULLELL MARTÍNEZ, JOAQUIM FUENTES POU,  
ELOI MAS MARTÍ I ARNAU SÁNCHEZ COLL

# Wifi vs Ethernet



# ÍNDEX

<b>1. Resum</b>	<b>2</b>
<b>2. Introducció</b>	<b>2</b>
2.1. Motivació	2
2.2. Objectius	2
2.3. Hipòtesis	2
<b>3. Mètodes</b>	<b>3</b>
3.1. Variables	3
3.2. Metodologia	3
3.3. Pla de recollida de dades	4
3.4. Anàlisi estadística	5
<b>4. Resultats</b>	<b>6</b>
4.1. Descriptiva de les dades	6
Fig. 1: Boxplot mostrant els temps amb wifi i Ethernet en ms i la diferència entre aquests	7
Fig. 2: Boxplot mostrant els temps amb wifi i Ethernet en ms i la diferència entre aquests un cop feta la transformació.	8
4.2. Resultats BC	8
Fig. 3 : Comprovació de les premisses de linealitat, homoscedasticitat i normalitat	9
Fig. 4: Diagrama de dispersió del Temps en ms de la càrrega de pàgines web amb wifi i Ethernet	10
Fig. 5: Diagrama de dispersió del Temps en ms de la càrrega de pàgines web amb wifi i Ethernet amb transformació logarítmica	10
Fig. 6 : Gràfic Bland-Altman amb transformació logarítmica. La diferència es representa en funció de la mitjana logarítmica de les dues mesures.	11
4.3. Resultats BD	12
<b>5.Discussió:</b>	<b>14</b>
<b>Apèndix</b>	<b>16</b>
Fig. 7 : Eines de desenvolupador del navegador	16
Fig.8 : gràfic de regressió lineal múltiple sobre wifi i Ethernet	17
<b>Taula d'imatges</b>	<b>18</b>

# 1. Resum

Aquest projecte neix amb la voluntat d'analitzar i investigar en profunditat les característiques, avantatges, limitacions i diferències del wifi i l'Ethernet, les dues tecnologies més utilitzades a l'hora de connectar-se a la xarxa.

Per tal d'obtenir conclusions clares, hem realitzat un estudi amb 60 pàgines web diferents com a mostra i hem enregistat el temps que es triga a descarregar una pàgina web amb cada una de les opcions (wifi i Ethernet) escollida a l'atzar.

## 2. Introducció

### 2.1. Motivació

Les principals fonts de motivació han sigut la possible aplicació a la pràctica en un futur, ja que en ser estudiants d'enginyeria informàtica aquest coneixement el podem acabar aplicant al nostre entorn (casa, feina...).

I el repte de comprendre dues tecnologies que utilitzem diàriament, però que no entenem amb profunditat, i saber discernir quina és més adequada en cada situació, perquè ambdues tenen avantatges i limitacions pròpies.

### 2.2. Objectius

Els objectius que ens hem proposat són els següents:

1. Fer una bona recollida de dades que sigui representativa i que ens permeti fer un bon estudi.
2. Decidir quina tecnologia ofereix major velocitat en entrar en una pàgina web de determinades característiques, en el mateix entorn i mateixes condicions, obtenint així dades clares sobre el temps que es triga a descarregar una pàgina web.
3. Aprendre sobre les diverses tecnologies de connexió com per millorar el rendiment del nostre temps a l'hora de treballar.

### 2.3. Hipòtesis

La nostra hipòtesi en primera instància de l'estudi va ser la següent: a l'estar connectat mitjançant un cable a la xarxa, l'Ethernet trigaria menys en descarregar la majoria de les pàgines web, i el wifi més.

## 3. Mètodes

### 3.1. Variables

Aquestes són les tres variables sobre les quals hem basat el nostre treball:

- Variable Y: temps que es triga a descarregar una pàgina web.
- Variable X: wifi o Ethernet.
- Variable Z: mida de la pàgina web.

Es pot representar Y d'aquesta forma:  $Y \sim X + Z$ . Segons com modifiquem X i Z obtindrem diferents valors per Y.

La nostra variable X representa el tipus de mètode emprat, és a dir, wifi o Ethernet. La nostra variable Y es basa en mostres aparellades, ja que estudiem la mateixa pàgina web amb ambdós mètodes: Y1 és el temps que triga l'Ethernet a descarregar-la i Y2 el temps que triga el wifi. La covariable Z és la mida de la pàgina web, volem saber si aquesta influència quin dels dos mètodes va més de pressa.

### 3.2. Metodologia

Les variables Y i Z s'han pogut mesurar gràcies a les eines de desenvolupador del navegador (vegeu a l'apèndix la figura 7), i per tant podem confirmar que aquest projecte és reproducible per altres persones, ja que aquestes eines són accessibles per a tothom, i que les pàgines triades són webs públiques. Els resultats obtinguts poden variar, ja que els servidors o el tràfic de la xarxa poden ser factors importants a tenir en compte, i degut això, hem fet la recollida de dades sempre en les mateixes condicions, i només fent servir l'ordinador per l'experiment i no per cap altra tasca.

La nostra eina de mesura (les eines de desenvolupador del navegador) l'hem fet funcionar de la següent forma:

1. Fer click dret i prémer inspeccionar (F12).
2. Anar a l'apartat "Network".
3. Entrar a la pàgina web desitjada introduint directament l'URL.
4. Observar els paràmetres Load (indica temps de càrrega) i Resources (indica la mida dels recursos que s'han carregat).

Hem mesurat la variable Y en mil·lisegons, i la Z en megabytes, ja que són unitats representatives i visuals. Les mesures les hem guardat en una taula d'Excel, on hi ha quatre columnes: Pàgina web, Temps amb Ethernet, Temps amb wifi, Mida. Així

facilitant-nos el trasllat de les dades a R. A més a més, hem inclòs dues columnes addicionals: diferència de temps i l'URL de la pàgina web.

---

### 3.3. Pla de recollida de dades

Per tal d'aconseguir un estudi que pugui garantir la validesa dels seus resultats, hem agafat una mostra representativa de pàgines web que permet comparar de forma precisa el temps de càrrega amb wifi i Ethernet.

La població de referència inclou pàgines web de diferents categories i amb diferents requeriments de recursos de xarxa, així com connexions wifi i Ethernet en un mateix entorn i amb les mateixes condicions.

Per garantir una mostra representativa, hem utilitzat un mostreig estratificat. Per fer-ho, hem organitzat les diferents pàgines web en dues categories: pàgines amb moltes imatges i pàgines amb molt de text. Cada categoria inclou el mateix volum de notícies per assegurar la fiabilitat de la mostra.

Més concretament, hem realitzat l'estudi amb 60 pàgines web diferents, on 30 formaven part de la primera categoria (pàgines amb moltes imatges), i les altres 30 de la segona categoria (pàgines amb molt de text). Llavors, donada una categoria, hem seleccionat de manera aleatòria una pàgina web dins d'aquesta mateixa categoria.

Un mostreig aleatori simple no seria òptim en aquest cas, ja que no garantiria la representació de totes les categories de pàgines web, i la probabilitat de selecció d'una pàgina web no seria la mateixa en tots els casos. Per tant, utilitzar un mostreig estratificat assegura una millor representació dels diferents llocs web.

També ens hem assegurat de l'aleatorietat en usar primer el wifi o l'Ethernet a l'hora de comparar la càrrega de la pàgina web amb els dos mecanismes, evitant així possibles alteracions en els resultats a causa d'un ordre fix de la prova.

A més, hem fet diverses proves per garantir que esdeveniments externs al nostre control no hagin afectat la recollida de dades, com pujades puntuals del trànsit, entre altres factors. El procés de mostreig implica un disseny amb dades aparellades: el temps de càrrega de cada pàgina web es mesura tant amb wifi com amb Ethernet en les mateixes condicions. Això permet comparar el rendiment de les dues connexions per la mateixa pàgina web, ja que les mesures provenen de la mateixa pàgina i en les mateixes condicions.

Cal mencionar que el nombre de mostres ha sigut més gran del que esperàvem inicialment (unes 50 pàgines web). Aquest augment s'ha fet perquè volem obtenir resultats encara més fiables. Calculem que amb aquesta quantitat es podrà veure una diferència, si és que existeix, entre el wifi i l'Ethernet, ja que sovint amb 60 parelles de mostres ja es poden obtenir resultats força robustos.

### 3.4. Anàlisi estadística

Hem utilitzat els següents mètodes per dur a terme l'anàlisi estadística: un interval de confiança del 95% de la mitjana de la diferència de temps entre ambdós mètodes, amb un model ajustat a les nostres dades i el càlcul d'una predicció de les observacions obtingudes. Vam decidir fer l'interval de confiança sobre la mitjana de la diferència de les nostres variables resposta, ja que el nostre estudi és sobre dades aparellades. Per tal de dur-ho a terme hem creat la darrera columna del nostre fitxer Excel restant la columna Ethernet de la columna wifi.

Un cop creada, hem comprovat que compleixi les premisses de validació, que són: l'aleatorietat de les dades i la normalitat. L'aleatorietat de les dades la podem afirmar gràcies a la metodologia seguida en recopilar les dades, explicat més endavant. Quant a la normalitat, l'hem pogut comprovar a partir de les instruccions QQline i QQplot de la diferència entre mètodes (wifi / Ethernet). La normalitat sobre la diferència no es compleix del tot i és per això que, per tal de millorar-ho, hem preferit aplicar una transformació logarítmica sobre la diferència. Un cop comprovades les premisses, hem calculat l'interval de confiança a partir de la variable transformada, i un cop obtinguda la resposta hem desfet la transformació, que en lloc d'indicar la diferència de temps, ens indica el rati entre el temps de wifi i el temps d'Ethernet.

A l'hora de fer la predicció amb un model estadístic, primer hem decidit quin és el model que s'ajusta més a les nostres dades. D'acord amb les nostres variables, ens hem decidit pel model lineal per poder calcular el temps de descàrrega amb wifi a partir del temps de descàrrega amb Ethernet i de la mida de la pàgina web. Aquest model té la següent fórmula:  $\text{Temps(ms)} \sim \beta_0 + \beta_1 \cdot \text{wifi(1)} / \text{Ethernet(0)} + \beta_2 \cdot \text{Mida(MB)}$ .

Per saber si el model és vàlid per fer la predicció, hem de validar les següents premisses: mostra aleatòria, linealitat i normalitat i homocedasticitat dels residus. La normalitat dels residus l'hem comprovada fent un QQplot dels residus, quant a la linealitat i homoscedasticitat dels residus hem fet un gràfic on apareixen els residus enfront de les prediccions del model.

En veure que aquestes premisses no es compleixen, hem dut a terme una transformació logarítmica. I hem fet l'estudi amb la modificació. Finalment, quant a

la independència hem pogut observar que aquesta sí que es compleix, ja que no hem observat cap patró al gràfic on comparem els residus amb l'ordre de recollida de les dades. Per acabar, hem validat les respostes i hem obtingut els resultats que es mostren en el següent apartat.

## 4. Resultats

### 4.1. Descriptiva de les dades

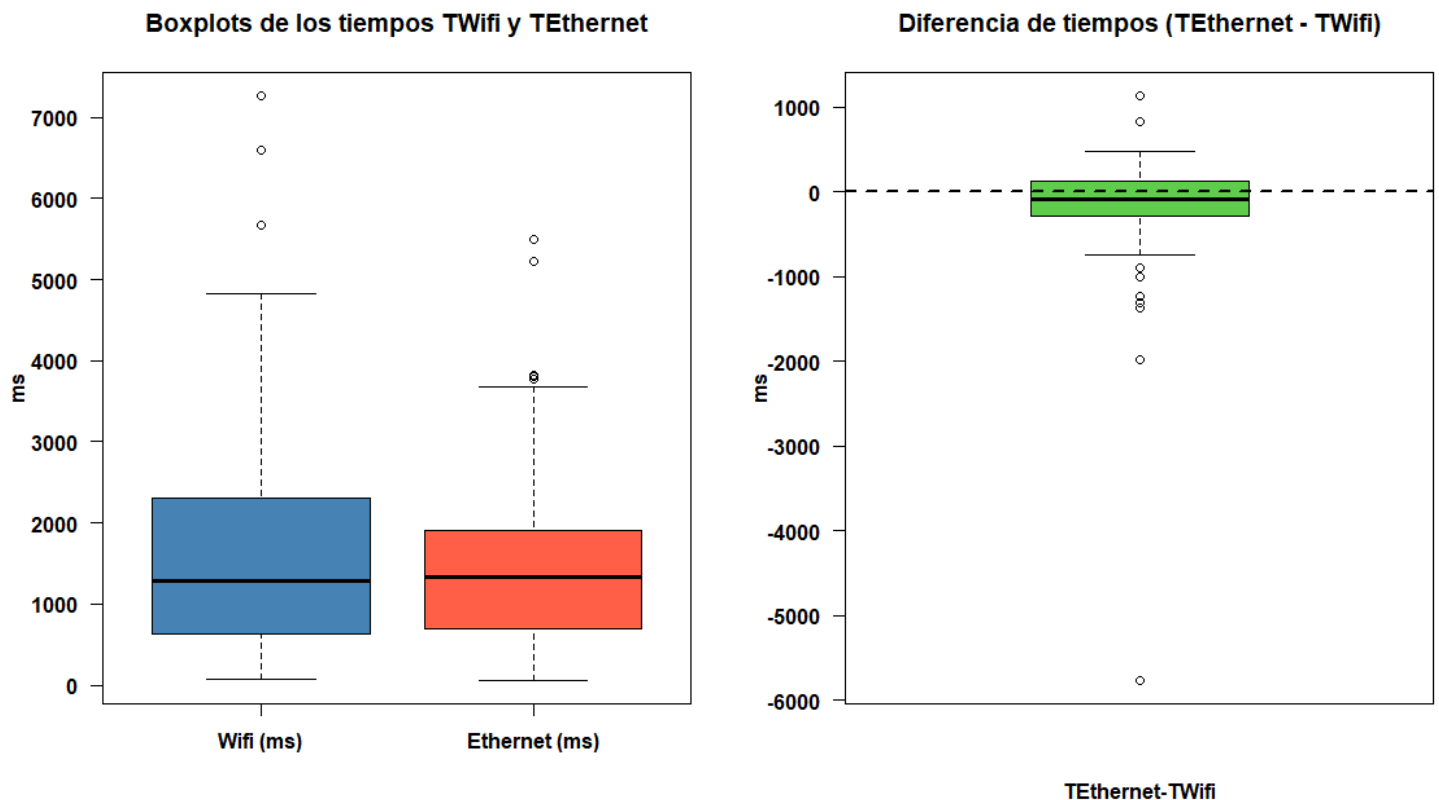
Les nostres variables d'interès són numèriques contínues, ja que es tracten de temps que es poden trobar representats dins d'un rang. Per tant, observarem la tendència central, que inclou la mitjana i la mediana, i també estudiarem la dispersió, mitjançant la desviació estàndard.

	Tendència Central:		Dispersió:
	Mitjana (ms)	Mediana (ms)	Desviació estàndard
<b>wifi</b>	1790.5	1285	1545.33
<b>Ethernet</b>	1551	1330	1226.29
<b>Ethernet - wifi</b>	-239.2	-85.5	888.95

Gràcies al programa R hem aconseguit calcular la mitjana, la mediana i la desviació estàndard de temps (en ms) que es triga a carregar les pàgines amb wifi i Ethernet. Amb aquests resultats, veiem que, com havíem dit a la nostra hipòtesi, la mitjana de temps a carregar les pàgines web que triga el wifi és bastant major que l'Ethernet.

És per això que també observem com la mitjana de la diferència és positiva, ja que el wifi en mitjana és més lent que l'Ethernet. No obstant això, veiem que la mediana és lleugerament inferior en el cas del wifi. I finalment la desviació estàndard de l'Ethernet és menor que la del wifi, la qual cosa significa que els valors del temps que es triga amb wifi seran més llunyans a la mitjana que amb l'Ethernet.

Com hem justificat en l'apartat anterior, la nostra variable d'interès és contínua. Per tant, usarem el Boxplot següent per fer la part gràfica de la descripció de les nostres dades. En el gràfic veiem que la mediana de l'Ethernet i la del wifi ronden els 1200 microsegons. Pel que fa als valors inferiors i superiors a les mitjanes veiem que es concentren entre 700 i 2300 microsegons en el wifi i entre 750 i 1900 microsegons pel Ethernet. Finalment, hem trobat tres valors atípics aproximadament en cada un dels mètodes.



*Fig. 1: Boxplot mostrant els temps amb wifi i Ethernet en ms i la diferència entre aquests*

Hem aconseguit aquests Boxplots utilitzant l'R (Vegeu Annex) i n'hem obtingut el resultat que veieu a la Fig.1. Amb aquests resultats podem veure com l'Ethernet és més regular, ja que tots els seus resultats estan més concentrats que en cas del wifi, on veiem la caixa més gran i amb pics més elevats, cosa que ens ensenya que és més irregular. En el cas de la diferència de temps, veiem que tendeix a ser negativa, sobretot els pics, ja que el temps en Ethernet en mitjana ja hem vist que era menor. Perquè quedi més clar, utilitzarem el diagrama de dispersió i el Bland Altman, ja que amb el boxplot no ens en podem assegurar.



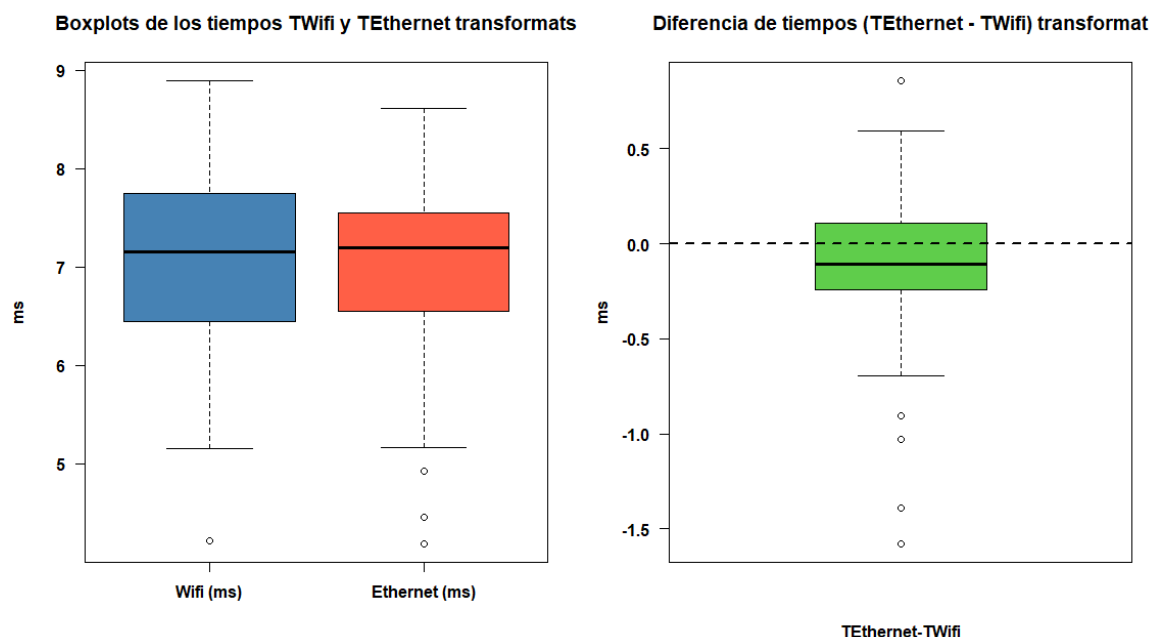
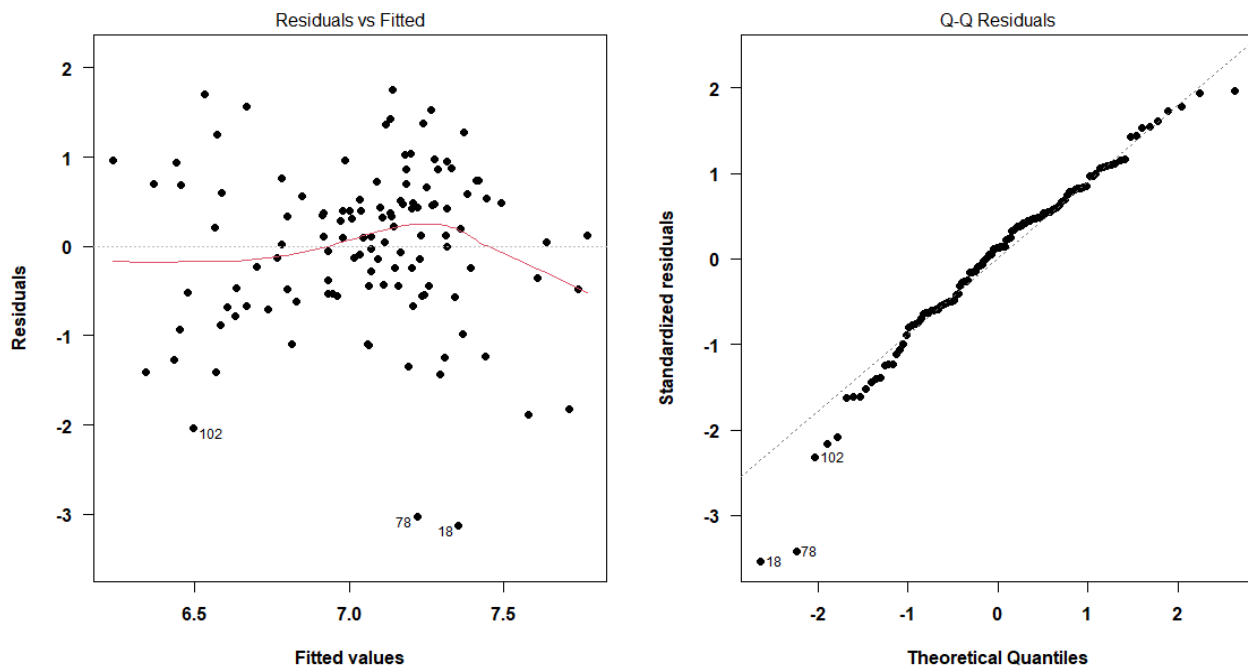


Fig. 2: Boxplot mostrant els temps amb wifi i Ethernet en ms i la diferència entre aquests un cop feta la transformació.

Aquests boxplots, Fig. 2, han estat generats un cop feta la transformació logarítmica. En els del temps es pot observar un comportament similar a l'altre boxplot, Fig. 1, amb la diferència que en ser escala logarítmica veiem proporcions i, per tant, els pics no criden tant l'atenció. Mentre que el de les diferències de temps la interpretació és la mateixa, on s'observa una tendència del wifi generar temps majors.

## 4.2. Resultats BC

**Validació de les premisses:** La premissa de mostra aleatòria es compleix, ja que tant per escollir la pàgina com per escollir el mètode vam utilitzar un “randomizer” que seleccionava un número entre l'1-60 pels webs i entre l'1-2 pels mètodes. La premissa de normalitat l'hem comprovat amb qqnorm2 i qqline, però els resultats no s'adapten a la normal del tot. Aplicant una transformació logarítmica, aconseguim que la distribució segueixi la normal, i per tant farem servir aquesta transformació per l'interval de confiança.



*Fig. 3 : Comprovació de les premisses de linealitat, homoscedasticitat i normalitat*

### IC de la diferència:

Utilitzant aquestes instruccions en R, hem calculat el següent interval de confiança per la diferència entre wifi i Ethernet:  $IC(D, 0.95) = [-0.24, -0.026]$

Aquest interval de confiança ens permet afirmar amb un 95% de seguretat que la  $\mu$  estarà entre aquests dos valors si tinguéssim una infinitat de dades. També ens aporta informació, i com els dos valors són negatius, podem afirmar amb bastant confiança que el Ethernet és més ràpid que el wifi. (en l'R, hem fet temps Ethernet - temps wifi)

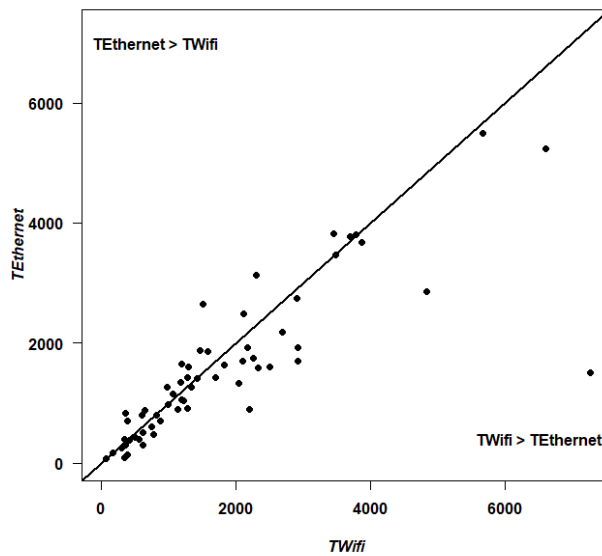


Fig. 4: Diagrama de dispersió del Temps en ms de la càrrega de pàgines web amb wifi i Ethernet

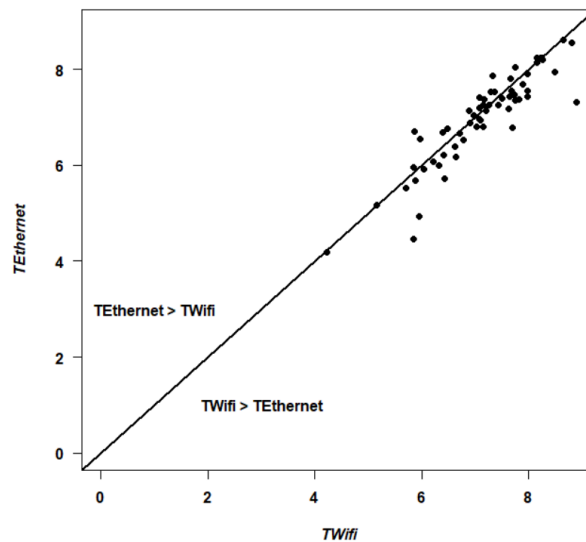
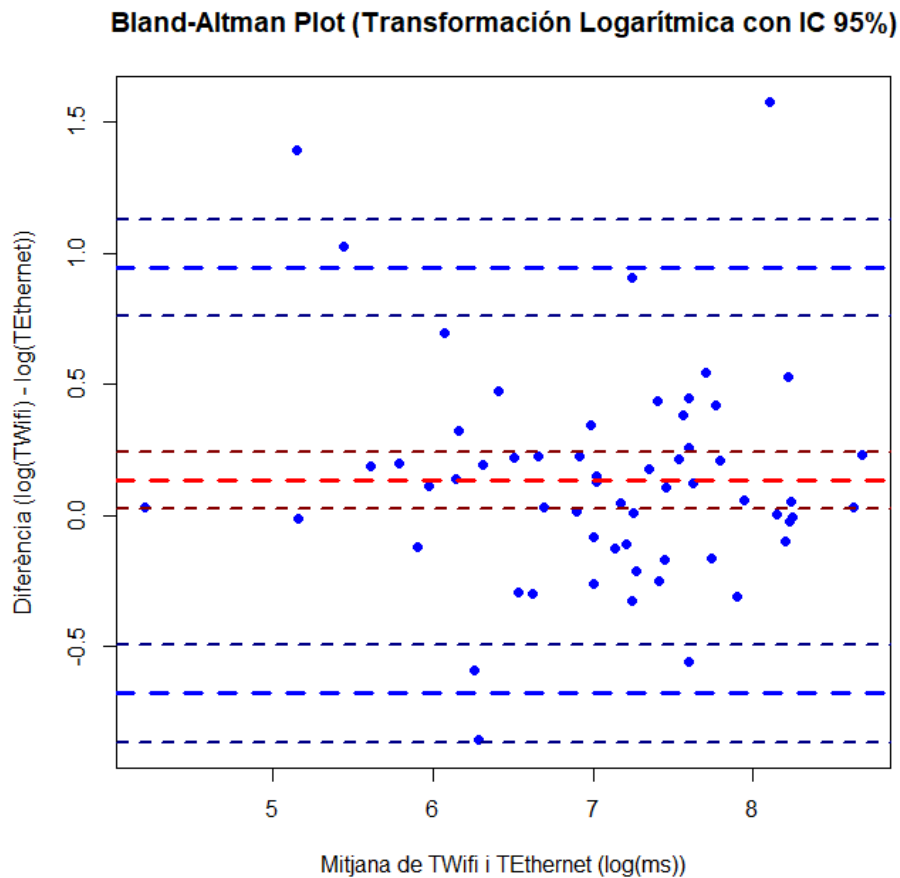


Fig. 5: Diagrama de dispersió del Temps en ms de la càrrega de pàgines web amb wifi i Ethernet amb transformació logarítmica

Utilitzant l'R (vegeu Annex), hem aconseguit el Diagrama de dispersió del Temps (en ms) de la càrrega de pàgines web amb wifi i amb Ethernet com es pot veure a la Fig. 4 i 5. Aquí veiem clarament que principalment, tant en wifi com en Ethernet els temps estan concentrats a la línia, és a dir, tenen uns temps semblants i regulars. Tanmateix, tornem a veure que en el cas del wifi, hi ha molts més pics (punts molt separats de la línia) on el temps de càrrega en wifi és molt superior que en Ethernet ( $T_{wifi} >$

TEthernet). Amb això tornem a veure que l'Ethernet és molt més regular que el wifi tot i que majoritàriament tenen uns valors de temps semblants.



*Fig. 6 : Gràfic Bland-Altman amb transformació logarítmica. La diferència es representa en funció de la mitjana logarítmica de les dues mesures.*

Utilitzant l'R (veure Apèndix), hem aconseguit el gràfic Bland-Altman, Fig. 6, que mostra les diferències entre les mesures de wifi i Ethernet després d'aplicar una transformació logarítmica. La línia vermella representa la diferència mitjana, que és propera a zero, indicant que no hi ha una tendència rellevant entre les dues connexions. El fet que l'interval de confiança sigui tan proper al zero en mostra que la diferència proporcional no és molt gran.

Les línies blaves corresponen als límits d'acord del 95%, dins dels quals es troben la majoria dels punts. Això suggereix una bona concordança entre les dues mesures. Tot i que hi ha una lleugera dispersió en valors més alts de la mitjana, això no implica una tendència sistemàtica evident.

En general, les mesures de wifi i Ethernet mostren una bona concordança global. No obstant això, es poden observar més pics a les mesures de wifi, suggerint una major

inestabilitat. Tot i que l'Ethernet en presenti, aquest són proporcionalment menors i menys.

**Interpretació:** Sabem amb un 95% de confiança que la diferència logarítmica val entre -0.24 , -0.026. Com que tenim la diferència logarítmica D com a  $\ln(\text{ethernet}) - \ln(\text{wifi})$ , que és equivalent a  $\ln(\text{ethernet} / \text{wifi})$ , i finalment, com  $e^{-0,24} = 0.7866$  i  $e^{-0,026} = 0.9743$ , llavors Ethernet en mitjana serà entre 0.7866 , 0.9743 vegades més ràpid que wifi . Malgrat que aquest interval no sigui del tot fiable, ja que la distribució no és perfectament normal, l'Ethernet sembla ser bastant més ràpid que el wifi, per tant podem afirmar que aquest és millor.

### 4.3. Resultats BD

**Model ajustat i estimacions del model:** Per aquest estudi, i per intentar veure com es relacionen els temps de descàrrega de wifi i Ethernet i veure si la mida de pàgina web hi té res a veure, hem establert el model lineal múltiple següent:  $\log(\text{Temps}(\text{ms})) \sim \beta_0 + \beta_1 \cdot \text{wifi}(1) / \text{Ethernet}(0) + \beta_2 \cdot \log(\text{Mida}(\text{MB}))$ . En fer el summary 3 del model obtenim les següents estimacions:

	Estimació	Error estàndard
Intercept	6.7225	0.1355
Ethernet	0.1328	0.1637
Mida de pàgina	0.2280	0.0603

Amb  $R^2 = 0.1134$

Correlació = 0.3367

Després de veure tots aquests valors podem arribar a extreure bastant informació.

Primer, ens podem fixar en el fet que  $\beta_1 = e^{0.1328} = 1.1420$ , i aquest valor indica la diferència esperada de temps que hi ha en carregar dues pàgines web iguals, una amb wifi i l'altra amb Ethernet.

Segonament, veiem que  $\beta_2 = e^{0.2280} = 1.25608$ , i aquest valor vol dir el temps esperat que trigaria a carregar una pàgina web per cada MB que tingui, i aquest valor serveix per tant wifi com per a Ethernet.

També podríem comentar el valor de  $\beta_0$ , però no és rellevant, ja que és el valor esperat quan el temps són 0 ms i la mida és 0 MB, cas impossible i absurd.

Finalment, el valor  $R^2$  permet afirmar que el model explica amb un 11.34% la variabilitat del temps. Això vol dir que encara que el wifi i l'Ethernet fan variar el temps de descàrrega de pàgines web, no són el principal factor o element a tenir en compte. El valor de la correlació reafirma aquesta idea, ja que diu que estan relacionades de certa forma. (la correlació és positiva, ja que a mesura que augmenta la mida també augmenta el temps)

**Validació de les premisses:** Per utilitzar aquest model cal validar les premisses de linealitat, normalitat dels residus, homoscedasticitat dels residus i mostra aleatòria. Ja hem justificat abans que la nostra mostra és aleatòria, i podem demostrar-ho amb el gràfic dels residus segons l'ordre d'arribada, que no mostra cap tendència. La linealitat es compleix raonablement ja que els residus en el gràfic Residuals vs Fitted es distribueixen de manera similar per sobre i per sota del 0. L'homoscedasticitat també es compleix força bé ja que la majoria de punts es distancien del 0 de forma semblant, sense zones amb molta dispersió (tret d'uns quants casos extrems). La normalitat dels residus, com veiem al QQplot, no s'acaba de complir.

**Explicació de transformacions:** Com hem pogut veure en l'apartat anterior, les premisses del model lineal múltiple no han estat del tot validades. És per això que hem dut a terme la transformació logarítmica següent i hem fet la mateixa anàlisi que en l'apartat anterior.

Codi de la transformació:

```
Recollida_de_dades2$`Temps(ms)` <- log(Recollida_de_dades2$`Temps(ms)`)
Recollida_de_dades2$`Mida(MB)` <- log(Recollida_de_dades2$`Mida(MB)`)
```

Mirant els gràfics resultats observem que la linealitat millora, per tant, hem escollit continuar amb les variables amb la transformació per fer el nostre model.

**Interpretació:** Segons els resultats del model, el temps base per trobar un node amb wifi és de 6.7225 microsegons, i incrementa en 0.1328 microsegons per cada microsegon que triga Ethernet. A més a més, per cada mida de pàgina, el temps decrementa en 0.2280 microsegons. Els resultats són força correctes, ja que veiem que Ethernet tendeix a ser més ràpid que wifi pel model.

**Ús per fer prediccions:** Tal com hem mostrat anteriorment, aquest model ens dona un  $R^2$  amb un valor molt pròxim a 0, cosa que ens indica que el temps que triguen wifi i

Ethernet estan poc relacionats, i que la mida de la pàgina tampoc influeix gaire en el resultat. Per tant, no podem fer una predicció acurada, ja que tan sols una petita part de la variabilitat és explicada pel model.

## 5. Discussió:

La nostra hipòtesi inicial era que el temps de càrrega en pàgines web depenia del tipus de connexió que s'utilitzés, i pensàvem que el temps en el cas de l'Ethernet seria inferior que fent servir el wifi, cosa que hem vist que s'ha complert, ja que la mitjana de temps de càrrega ho corrobora. Tot i això, hem vist que la mediana és lleugerament inferior amb wifi. Veient els diferents gràfics hem pogut entendre aquestes dades, perquè hem vist que el wifi és més irregular, en cas de webs que la seva mida és més petita ho càrrega amb prou facilitat, però en webs amb mida més gran té pics més elevats que fan que la mitjana creixi. D'altra banda, l'Ethernet és molt més regular, tenint molt pocs pics tant en els casos de poca mida com els de molta mida de la pàgina.

Segons els resultats obtinguts al bloc C, podem afirmar amb confiança que Ethernet és més ràpid que wifi per descarregar pàgines web, ja que wifi tendeix a ser entre 0.24 i 0.026 vegades més lent. Malgrat tot, atès el baix  $R^2$  del model lineal, no hem pogut establir cap relació convincent entre el temps que triguen els dos mètodes ni hem pogut trobar cap relació entre els seus temps i la mida de la pàgina, com originalment pensàvem que passaria. Això sembla indicar que les variables estan poc relacionades.

Alguna limitació que hem tingut a l'hora de fer el treball ha estat en el gràfic de regressió lineal, hauríem d'haver separat les variables en dos gràfics diferents i no en el mateix com hem fet, ja que les nostres variables no són independents i això podria causar una alteració en el resultat. Potser també hauríem d'haver agafat un nombre més gran de pàgines web per fer el treball, pel fet que encara que vam augmentar el nombre inicial que teníem previst, al final no hem pogut aconseguir normalitat sense haver de fer una transformació a les dades. Tot i això, amb  $n = 60$  ja hauria de ser una mida prou gran per a treure bones conclusions, però pot ser que amb una  $n$  més gran (com  $n = 100$  per exemple) sí que veuríem canvi. Un factor que no podíem controlar era el fet del recorregut seguit per les consultes per accedir als servidors de la pàgina web, ni el temps de resposta d'aquests.

Finalment, després d'aquest treball, hem après una sèrie de lliçons que abans no sabíem i que ens poden ser molt útils pel futur. Aquestes lliçons van des de saber fer una gran recollida de dades i organitzar-la, fins a tindre un pensament crític i saber que l'Ethernet és més ràpid que el wifi en mitjana.

També cal dir que per futurs treballs, podríem millorar agafant més mostres pel projecte i intentant que es compleixi homoscedasticitat i linealitat des d'un principi. Una altra millora que ja hem tingut en compte, però potser ho hauríem d'haver aprofundit més és la part d'escollir les diferents pàgines web, ja que això pot alterar els resultats finals.

Nosaltres, com hem explicat anteriorment, hem utilitzat un mostreig estratificat, on hem agafat diferents webs en dos blocs, amb mida de pàgina més pesant (on predominen les imatges) i menys pesant (on predomina el text). Tot i que aquesta estratègia és prou bona per poder analitzar els diferents tipus de webs, podríem haver usat una altra tècnica o simplement haver utilitzat aquest mètode però en comptes de separar-ho en dos grups, separar-ho amb més subgrups perquè hi hagués més relació i igualtat.



## Apèndix

Per recollir dades sobre els temps de càrrega per al nostre treball d'estadística, hem fet servir les eines de desenvolupador del navegador. Aquestes eines ens permeten realitzar mesures precises dels temps de càrrega de diverses pàgines web, i així obtenir dades quantitatives. Per exemple:

- **DOMContentLoaded:** Aquest valor representa el moment en què el document HTML inicial està completament carregat i processat, però sense esperar que tots els recursos (com imatges, estils externs o scripts) hagin acabat de carregar-se.
- **Load:** Aquest valor indica el temps que triga a carregar-se tota la pàgina, inclosos tots els recursos addicionals.

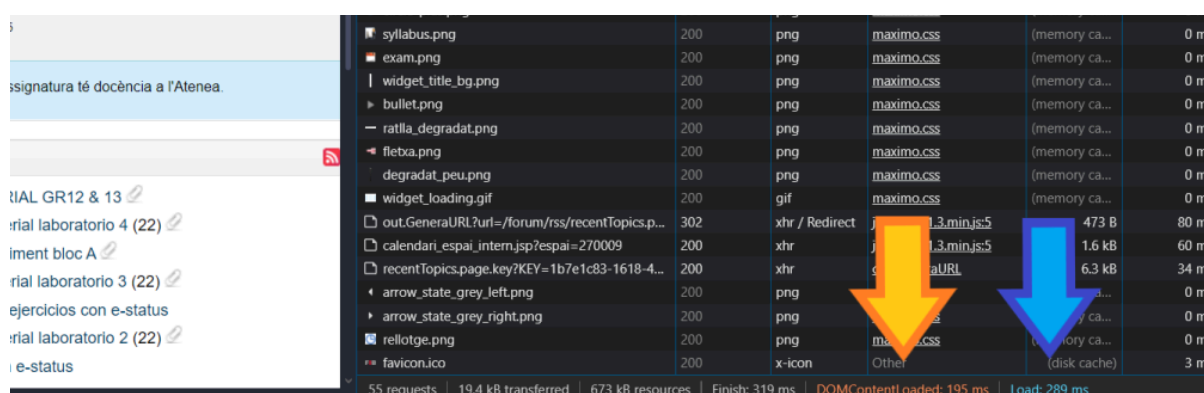


Fig. 7 : Eines de desenvolupador del navegador

El temps que es tarda en descarregar una pàgina web amb Wi-fi(1) i Ethernet(0)

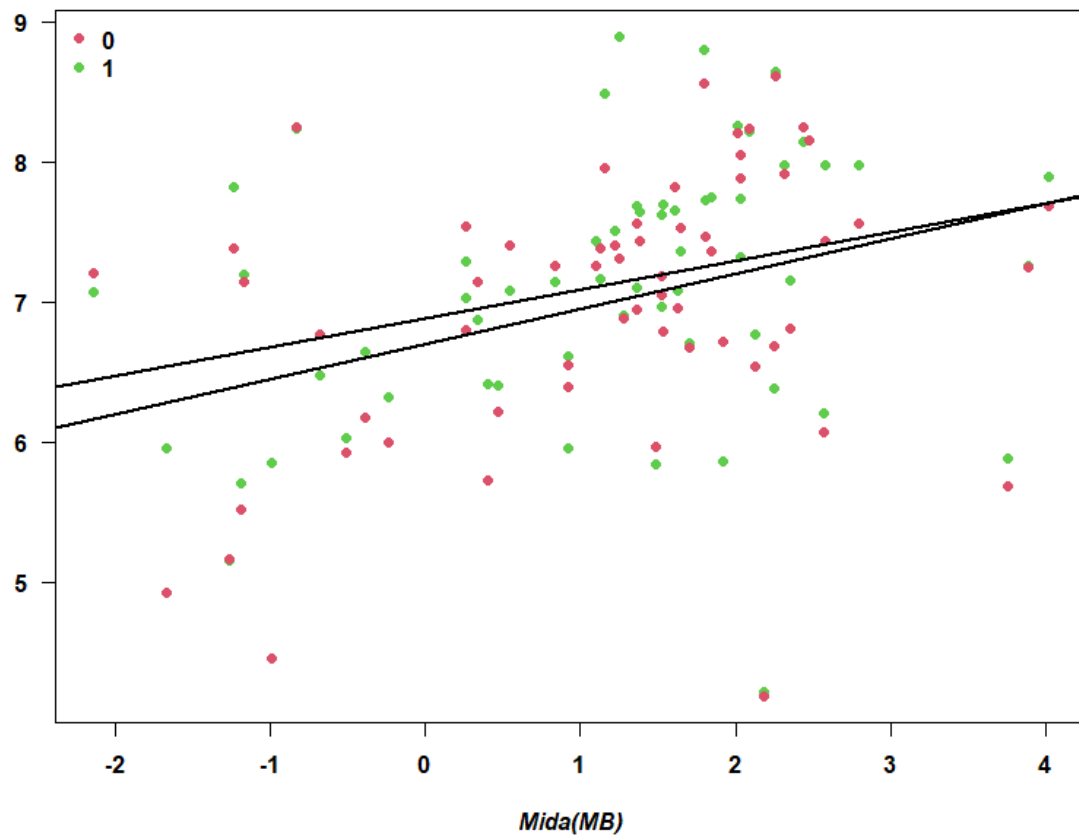


Fig.8 : Gràfic de regressió lineal múltiple sobre wifi i Ethernet

## Taula d'imatges

Fig. 1: Boxplot mostrant els temps amb wifi i Ethernet en ms i la diferència entre aquests	7
Fig. 2: Boxplot mostrant els temps amb wifi i Ethernet en ms i la diferència entre aquests un cop feta la transformació.	8
Fig. 3 : Comprovació de les premisses de linealitat, homoscedasticitat i normalitat	9
Fig. 4: Diagrama de dispersió del Temps en ms de la càrrega de pàgines web amb wifi i Ethernet	10
Fig. 5: Diagrama de dispersió del Temps en ms de la càrrega de pàgines web amb wifi i Ethernet amb transformació logarítmica	10
Fig. 6 : Gràfic Bland-Altman amb transformació logarítmica. La diferència es representa en funció de la mitjana logarítmica de les dues mesures.	11
Fig. 7 : Eines de desenvolupador del navegador	16
Fig.8 : gràfic de regressió lineal múltiple sobre wifi i Ethernet	17

### **Codi R en l'Annex (document apart)**