

Pràctica T2
Entrega final

LÍNIA FGC Llobregat-Anoia
Documentació

Simulació - Curs 2016/17 Q2
Facultat d'Informàtica de Barcelona
Universitat Politècnica de Catalunya

Membres del grup:
Albert López Alcacer
Iván de Mingo Guerrero
Arnau Blanch Cortès
Grup 12

Índex

Resum executiu	2
Descripció del sistema	3
Objectius	4
Descripció del model	5
El funcionament i interaccions principals entre els dos agents, doncs, es concentra en els dos procediments go-train i go-station.	6
Aspectes metodològics	7
Especificació del model	8
Hipòtesis de modelització	8
Dades d'entrada	9
Variables de sortida	10
Disseny d'experiments	11
Algoritme de Yates	11
Nombre de rèpliques necessàries	12
Resultats de la simulació	13
Experiment 1	13
Experiment 2	13
Experiment 3	14
Experiment 4	14
Resum dels experiments	15
Resultat de l'algoritme de Yates	16
Conclusions i recomanacions	17

Resum executiu

In this project, we wanted to model and analyze the system consisted of the five first stations of the Llobregat-Anoia railway line. The objective of this analysis is the obtain the best combination of train capacity and train frequency in order to minimize the mean waiting time in those five stations.

In the previous deliveries of this project, we specified the model using SDL and later, we implemented it using NetLogo. In the last delivery, we analyzed which was the best option of simulation software to model our system. As the result was NetLogo, we've reused our previous model to generate more than one train. We also added a new queue to the stations in order to calculate the total waiting time in a day and the number of users of the line to be able to calculate the mean waiting time in the line during a day.

After the had improved our previous model implementation, we designed the experiments and used NetLogo to simulate the experiments with a number of replicas calculated in a pilot experiment in order to obtain results with a 5% margin of confidence.

With those results, we processed them using RStudio and then, we used the Yates algorithm to see the effects of each factor.

As this delivery does not require validation of the results, we explained how we would do it in a real case.

Descripció del sistema

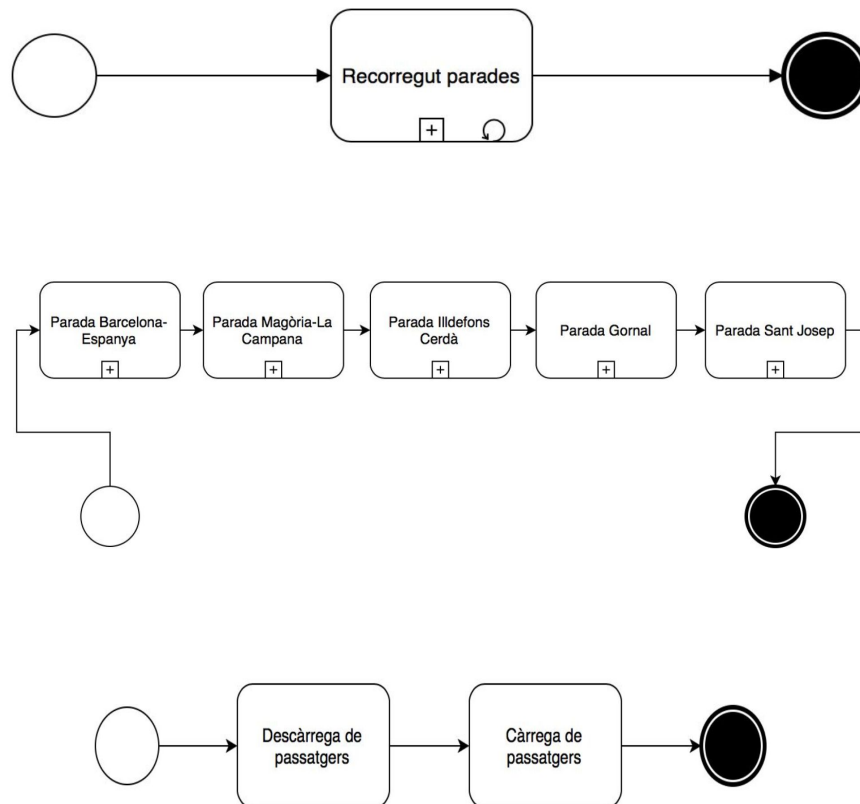
El sistema a modelar es descriu en el fitxer Practica_T1_de_SIM, concretament és la practica 13. En aquest document es parla del principi de la línia de tren de la FGC: Llobregat-Anoia.

El sistema contempla que els trens tenen un sentit de marxa, que recorren 5 estacions i que tenen capacitat per a 200 persones.

A cada parada pugen x passatgers i baixen y passatgers. El primer valor depèn del temps que fa que no passa un tren per aquesta, mentre que el segon depèn de la parada. Els valors corresponents estan definits a l'enunciat del sistema.

Finalment, es diu que el temps de sortides de l'estació central segueix una distribució exponencial de 5 minuts.

Podem descriure el sistema amb els següents diagrames en BPMN:



Objectius

L'objectiu general d'aquesta pràctica és modelar el sistema de l'enunciat número 13 *FGC Línia Llobregat-Anoia* per tal de reduir el temps mitjà d'espera. Aquest temps són els minuts (en mitjana) que cal esperar en una estació per poder pujar a un tren al llarg d'un dia en les diverses estacions de la línia.

Per a reduir aquest temps d'espera, volem trobar quina és la combinació de capacitat màxima del trens i la freqüència de sortida que ho permetrà. Tenim la limitació que els trens disponibles amb més capacitat permeten portar 700 passatgers i que el paràmetre més petit que admet la distribució exponencial que segueix la freqüència de sortida dels trens és d'1 minut.

Per a assolir aquests objectius, representarem el model en NetLogo utilitzant simulació basada en agents intel·ligents. En aquest model tenim dos agents diferents: el tren i les parades. Aquests agents interacturan entre sí produint diferents efectes segons les dades d'entrada en cada cas.

Per a poder analitzar el model i complir l'objectiu anterior, hem de poder obtenir dades del model per tal de poder aplicar l'algoritme de Yates per veure els efectes dels dos factors d'entrada.

Descripció del model

El model consisteix en una línia de tren que conté 5 estacions. De l'estació principal surten trens amb una distribució exponencial que podem modificar (dades d'entrada). Així mateix, els trens tenen una certa capacitat, que també podem variar per observar diferents resultats o efectes.

Per a poder obtenir el temps mitjà d'espera dels passatgers a les diferents estacions s'han definit grups de passatgers. Al model tenim (a cada estació) una cua que emmagatzema els valors corresponents a les capacitats dels esmentats grups. Així mateix, es disposa d'un altre cua per emmagatzemar els diferents temps d'espera, associats a cada grup.

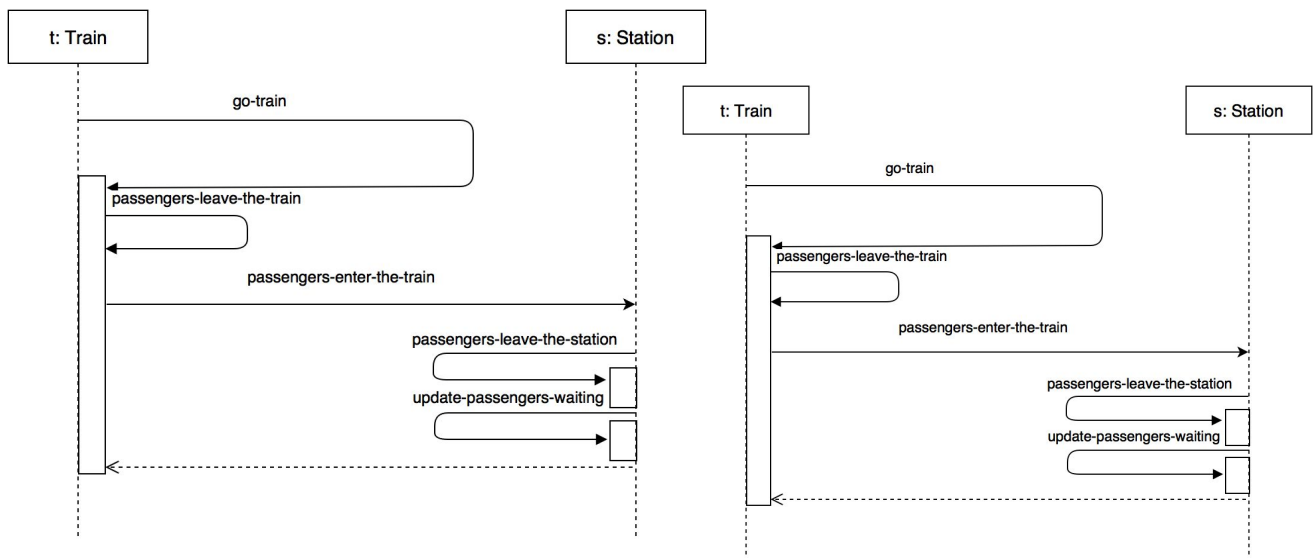
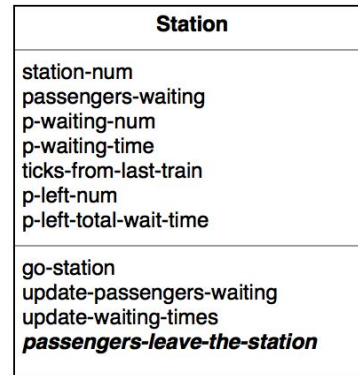
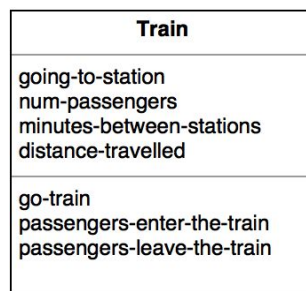
La cua de grups de passatgers s'actualitza a cada estació constantment. De manera que, segons el temps que ha passat des de l'últim tren, es modifica l'últim grup de passatgers afegint-ne de nous (en cas de que hi ha espai en aquest grup, altrament es crea un grup nou). El número de nous passatgers correspon a la distribució segons el temps d'espera, com s'ha dit anteriorment. Per un altre banda també s'actualitza el temps d'espera de cada grup.

Quan un tren arriba a una estació, han de baixar un cert número de passatgers per posteriorment pujar-ne un altre. El primer valor es calcula fent servir les distribucions determinades segons la parada en la que el tren realitza l'acció, com s'especifica a l'enunciat. Per a calcular el segon valor, cal agafar grups de passatgers fins que no hi ha més espai al tren i, en cas de que calgui agafar una part d'un grup, es modifica el tamany d'aquest grup que encara haurà d'esperar al següent tren.

Entre les estacions, es calcula el temps que el tren triga en arribar segons una distribució uniforme entre 1,5 i 3,5 minuts. I quan aquest surt de l'última estació (ja sense passatgers), es destrueix.

Com s'especifica a l'enunciat, els trens no poden xocar entre ells ni adelantar-se. Per aquest motiu el model permet que quan dos trens volen estar a la mateixa posició, aquesta situació s'impedeixi.

Per a poder observar els diferents efectes de les dades d'entrada, es va calculant constantment la mitjana del temps que els passatgers han hagut d'esperar i es reporta en la gràfica de "Mean wait time (total)" i "Mean wait time (per estacions)", a l'hora que es mostra el número de passatgers esperant per estacions a "Passengers waiting" . A més a més, disposem de dues gràfiques per a cada estació mostrant el número total de passatgers i el temps d'espera per als grups de passatgers que hi han.



El funcionament i interaccions principals entre els dos agents, doncs, es concentra en els dos procediments go-train i go-station.

En cada instant de la simulació, es crida als trens per tal d'executar "go-train". Amb això ens assegurem que, una vegada el tren ha arribat a una estació, es calculin el número de passatgers a baixar segons l'estació en la que està.

Posteriorment, mitjançant la interacció "passengers-enter-the-train" amb l'estació, fem que es cridi a l'estació a calcular el número de passatgers a pujar al tren mitjançant el procediment "passengers-leave-the-station". Així mateix l'estació aprofita per actualitzar el número de passatgers esperant.

Per un altre banda, durant l'execució del model, els agents de tipus estació s'encarreguen d'anar actualitzant els diferents temps d'espera dels grups de passatgers. Així mateix, també s'actualitzen aquests grups de passatgers segons el temps que fa que un tren no ha interactuat amb l'estació.

Aspectes metodològics

El motiu de fer servir l'eina NetLogo en concret és el resultat de l'estudi que vam fer a la pràctica anterior.

I degut a que fem servir NetLogo com a motor de simulació, estem fent servir un paradigma de modelització basat en agents.

Aquest paradigma ens permet simular accions i interaccions entre trens i parades. A l'hora, podem observar els resultats o efectes d'aquestes interaccions realitzant execucions en les que modifiquem dades d'entrada.

A més a més, fem servir l'eina de disseny d'experiments *Behaviour Space* integrada a NetLogo, per tal d'obtenir els diferents resultats de la simulació.

Especificació del model

Hipòtesis de modelització

Hipòtesis simplificadores

- Quan s'arriba a les 23:00 h, la simulació es para i per tant els trens, encara que tinguin passatgers, es paren al lloc on es trobin en aquell moment.
- Les estacions tenen un aforament il·limitat.

Hipòtesis sistèmiques

- Temps del trajecte entre dues estacions: el temps del viatge entre dues estacions segueix una distribució uniforme entre 1.5 i 3.5 minuts. En la realitat, els viatges entre estacions d'aquesta línia de tren tenen una durada de 2 - 3 minuts.
- Persones esperant un tren: a les estacions s'hi van acumulant persones que porten temps diferents esperant al tren. Cada cop que arriba un tren a una estació, hi pugen primer les persones que porten més temps esperant.
- Horari: els trens comencen a circular a les 8 del matí i finalitzen la jornada a les 23:00 h.
- Els trens no es poden avançar ja que en el nostre model només tenim una via. En el cas que dos trens estiguin a punt de xocar, el tren de darrere s'aturarà fins que pugui avançar.
- Els trens estan aturats a les estacions 12 segons (1 *tick*). Durant aquest temps, hi baixen i hi pugen passatgers.
- La capacitat màxima dels trens disponibles en el mercat (i que, per tant, es podrien utilitzar a la línia de FGC) és de 700 passatgers.
- El paràmetre mínim de la distribució exponencial de les sortides de trens és d'1 minut.

Variables d'estat

- *Going to station*: indica l'estació a la qual es dirigeix un tren.
- *Distance travelled*: indica la distància recorreguda des de l'última estació. Si la distància recorreguda és igual a la distància entre estacions, el tren es troba aturat en una estació.
- *Cua de passatgers d'una estació*: cada estació conté una cua amb els diferents grups de passatgers que hi estan esperant. Està representada per les llistes *p-waiting-num* i *p-waiting-time* que contenen el número de passatgers dels blocs i el temps que porten esperant els passatgers de cada bloc.

Successos

- Els trens surten des de la primera estació (Barcelona-Espanya) amb el número de gent que estava esperant-hi, però no s'hi baixa cap passatger ja que és la primera parada.

- A l'arribar a una estació, primer baixen un nombre de passatgers determinat segons l'estació (taula 1) i, a continuació, hi pugen els que estaven esperant fins que el tren s'omple per complet o ja no hi ha més gent per pujar.
- Quan un tren surt de l'última estació del nostre model *mor*. En la realitat continuaria fins la següent estació, però l'objectiu d'aquesta simulació és simular només el principi de la línia FGC Llobregat-Anoia.

Estació	Barcelona - Espanya	Magòria - La Campana	Ildefons Cerdà	Gornal	Sant Josep
Nº persones que baixen	0	U(50, 75)	U(75, 100)	U(100, 150)	U(100, 200)

Taula 1: Persones que baixen a cada estació.

Activitats

- Els trens avancen per les vies segons el temps mitjà entre estacions (*mean time between stations*) en minuts i el número de *ticks* per minut definit (*ticks per minute*).
- Cada 5 minuts s'actualitza el número de persones esperant segons el temps que ha passat des de l'últim tren (taula 2).

Temps des de l'últim tren	menys de 5 min	entre 5 i 10 min	entre 10 i 15 min	més de 15 minuts
Nº persones esperant	U(100, 200)	U(150, 300)	U(50, 300)	U(200, 400)

Taula 2: Persones que esperen un tren en una estació.

Dades d'entrada

Freqüència de sortida de trens

Fa referència al número de minuts entre les sortides de dos trens i es pot modificar a través d'un *slider*. La dada d'entrada és el paràmetre d'una distribució exponencial, tal i com ho indica l'enunciat d'aquesta pràctica. És a dir, els trens surten cada certs minuts, definits per una distribució exponencial amb mitjana indicada per aquesta dada d'entrada.

Capacitat màxima del tren

La capacitat màxima del tren és una altra dada d'entrada que també es pot modificar a través d'un *slider*. Aquesta dada determina el número màxim de passatgers que pot portar el tren.

Variables de sortida

La nostra variable de sortida és el temps mitjà d'espera dels passatgers a les estacions al llarg d'un dia (de les 8h a les 23h).

Verificació i validació de la simulació

Un model hauria d'estar desenvolupat per a un objectiu o aplicació concrets i la seva validesa determinada respecte aquests. Normalment, determinar si un model és totalment vàlid o no és molt costós i requereix molt de temps. Com més grau de validesa hagi de tenir el model, més valor per l'usuari i més cost tindrà, com es pot veure a la següent figura.

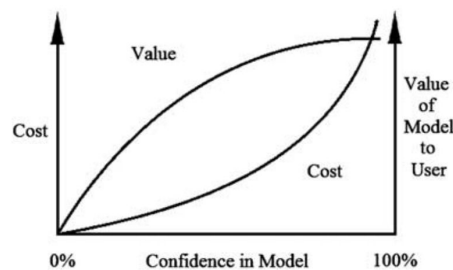


Fig. 1: Valor i cost d'un model segons la seva validesa.

Tenint en compte que el nostre no és un model de gran escala i tenim un equip de desenvolupament petit, faríem aquesta validació a través d'un usuari del model simulat durant el desenvolupament.

Amb aquesta verificació, validació i acreditació del model simulat (VV&A) serem capaços de produir un model que representi el comportament del sistema tant proper com sigui possible i sigui d'utilitat pel client. A més, incrementarem la credibilitat d'aquest per tal que sigui usat per a la gestió i per a la predicció.

Disseny d'experiments

Com que volem reduir el temps d'espera mitjà al mínim, volem obtenir la combinació òptima entre capacitat dels trens i el paràmetre de l'exponencial de sortida.

Per tant, el resultat que analitzarem serà el temps d'espera mitjà al llarg d'un dia (de 8h a 23h). Aquest temps estarà representat en minuts.

Per a esbrinar quins són els efectes en els resultats del model de les dues variables de sortida, hem decidit utilitzar l'algoritme de Yates.

Algoritme de Yates

Per aplicar l'algoritme de Yates, en el nostre cas hem definit dos factors.

El factor A representa la capacitat màxima dels trens. El valor - és la capacitat del cas base, que és de 350 passatgers. El valor + és la capacitat màxima que suposem que pot contenir un tren (700 passatgers).

El factor B representa el paràmetre de la distribució exponencial que segueix la freqüència de sortida de trens (en minuts). El valor - és el paràmetre de l'exponencial del cas base, que és de 4.0 minuts. El valor + és el paràmetre de l'exponencial mínim que suposem que pot tenir (1 minut).

Un cop definides les variables, hem de fer totes les combinacions possibles entre aquests. En el nostre cas, com que tenim només 2 variables d'entrada, tindrem 2^2 experiments possibles.

Per tant, tindrem 4 experiments:

- **Experiment 1:**
 - Capacitat màxima dels trens (A) = 350 persones (-)
 - Paràmetre de la distribució de les sortides (B) = 4 minuts (-)
- **Experiment 2:**
 - Capacitat màxima dels trens (A) = 350 persones (-)
 - Paràmetre de la distribució de les sortides (B) = 1 minut (+)
- **Experiment 3:**
 - Capacitat màxima dels trens (A) = 700 persones (+)
 - Paràmetre de la distribució de les sortides (B) = 4 minuts (-)
- **Experiment 4:**
 - Capacitat màxima dels trens (A) = 700 persones (+)
 - Paràmetre de la distribució de les sortides (B) = 1 minut (+)

Nombre de rèpliques necessàries

Per calcular el número de rèpliques necessàries per la nostra simulació hem fet una primera simulació (prova pilot) per a cadascun dels quatre experiments. A cada prova pilot s'han fet 50 rèpliques. A partir d'aquí, hem calculat la mitjana de cada prova pilot:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

A continuació, calculem la variància mostral:

$$S^2 = \frac{\sum_{i=1}^n x_i - \bar{X}}{n-1}$$

A més, per saber l'interval de confiança, apliquem una distribució *t* de *Student* de *n-1* graus de llibertat:

$$\bar{X} \pm t_{1-\alpha/2, n-1} \sqrt{\frac{S^2}{n}}$$

En primer lloc, calcularem el semi-interval de confiança de la prova pilot:

$$h = t_{1-\alpha/2, n-1} \frac{S}{\sqrt{n}}$$

Finalment, el nombre de rèpliques necessàries es pot calcular de la següent manera:

$$n^* = n \left(\frac{h}{h^*} \right)^2$$

Per a calcular l'interval de confiança desitjat h^* , hem aplicat un interval de confiança del 5% sobre la mitjana de la prova.

Tots aquests càlculs s'han realitzat a través d'un *script* d'*R* adjuntat en aquesta pràctica, juntament amb els resultats.

Resultats de la simulació

Després d'haver de dissenyat els experiments, per a cada experiment hem calculat el nombre de rèpliques necessàries per a obtenir un interval de valors amb una confiança del 95%. Aleshores, hem executat el número resultant de rèpliques i hem processat els resultats i els hem analitzat amb l'algoritme de Yates.

Per a obtenir els resultats, hem afegit diversos experiments amb el *Behaviour Space* de NetLogo per a obtenir els resultats en un CSV. Aquests resultats els hem passat a un fitxer Excel per tal de poder-los processar bé.

Amb RStudio, hem creat un script per calcular el nombre de rèpliques necessàries per a cada experiment i un script per a obtenir la mitjana i l'interval de confiança del resultat dels experiments. Aquests scripts generen un fitxer amb resultats detallats per a cada fitxer d'entrada i un fitxer amb un resum dels resultats.

Tots aquests fitxers i scripts estan adjuntats a aquesta pràctica.

Experiment 1

L'experiment 1 representa el cas on la capacitat d'un tren és de 350 passatgers i el paràmetre de la distribució de les sortides és de 4 minuts. Aquest és el cas base.

En la prova pilot d'aquest cas hem realitzat 50 rèpliques i hem obtingut els següents resultats:

```
Mitjana prova pilot = 3.917167
Desviació prova pilot = 1.364336
Semi-amplada de l'interval de confiança de la prova pilot = 0.38774
Interval de confiança ( $\alpha = 0.05$ ) = [ 3.529427 , 4.304907 ]
Semi-amplada de l'interval de confiança desitjat (95% de confiança) = 0.1958583
Nombre de rèpliques necessàries = 195.9599 -> 196
```

A partir d'aquí, hem realitzat un altre cop d'experiment amb 196 rèpliques per tal d'obtenir l'interval de confiança desitjat (5%). Hem obtingut els següents resultats:

```
Mitjana = 3.842326
Desviació = 1.11205
Interval de confiança ( $\alpha = 0.05$ ) = [ 3.842326 , 3.842326 ]
```

Experiment 2

L'experiment 2 representa el cas on la capacitat d'un tren és de 350 passatgers i el paràmetre de la distribució de les sortides és el mínim (1 minut).

En la prova pilot d'aquest cas hem realitzat 50 rèpliques i hem obtingut els següents resultats:

Mitjana prova pilot = 1.696922
Desviació prova pilot = 0.05700392

Semi-amplada de l'interval de confiança de la prova pilot = 0.01620034
Interval de confiança ($\alpha = 0.05$) = [1.680722 , 1.713123]
Semi-amplada de l'interval de confiança desitjat (95% de confiança) = 0.08484611

Nombre de rèpliques necessàries = 1.822863 -> 2

A partir d'aquí, hem realitzat un altre cop d'experiment amb 2 rèpliques per tal d'obtenir l'interval de confiança desitjat (5%). Hem obtingut els següents resultats:

Mitjana = 1.702296
Desviació = 0.02038861
Interval de confiança ($\alpha = 0.05$) = [1.702296 , 1.702296]

Experiment 3

L'experiment 3 representa el cas on la capacitat d'un tren és la màxima (700 passatgers) i el paràmetre de la distribució de les sortides és de 4 minuts.

En la prova pilot d'aquest cas hem realitzat 50 rèpliques i hem obtingut els següents resultats:

Mitjana prova pilot = 2.299412
Desviació prova pilot = 0.06902241

Semi-amplada de l'interval de confiança de la prova pilot = 0.01961595
Interval de confiança ($\alpha = 0.05$) = [2.279796 , 2.319028]
Semi-amplada de l'interval de confiança desitjat (95% de confiança) = 0.1149706

Nombre de rèpliques necessàries = 1.45551 -> 2

A partir d'aquí, hem realitzat un altre cop d'experiment amb 2 rèpliques per tal d'obtenir l'interval de confiança desitjat (5%). Hem obtingut els següents resultats:

Mitjana = 2.397441
Desviació = 0.09390713
Interval de confiança ($\alpha = 0.05$) = [2.397441 , 2.397441]

Experiment 4

L'experiment 3 representa el cas on la capacitat d'un tren és la màxima (700 passatgers) i el paràmetre de la distribució de les sortides és el mínim (1 minut).

En la prova pilot d'aquest cas hem realitzat 50 rèpliques i hem obtingut els següents resultats:

Mitjana prova pilot = 1.486156

Desviació prova pilot = 0.01254808

Semi-amplada de l'interval de confiança de la prova pilot = 0.003566124

Interval de confiança ($\alpha = 0.05$) = [1.48259 , 1.489723]

Semi-amplada de l'interval de confiança desitjat (95% de confiança) = 0.07430782

Nombre de rèpliques necessàries = 0.1151579 -> 1

A partir d'aquí, hem realitzat un altre cop d'experiment amb 1 rèplica per tal d'obtenir l'interval de confiança desitjat (5%). Hem obtingut els següents resultats:

Mitjana = 1.480008

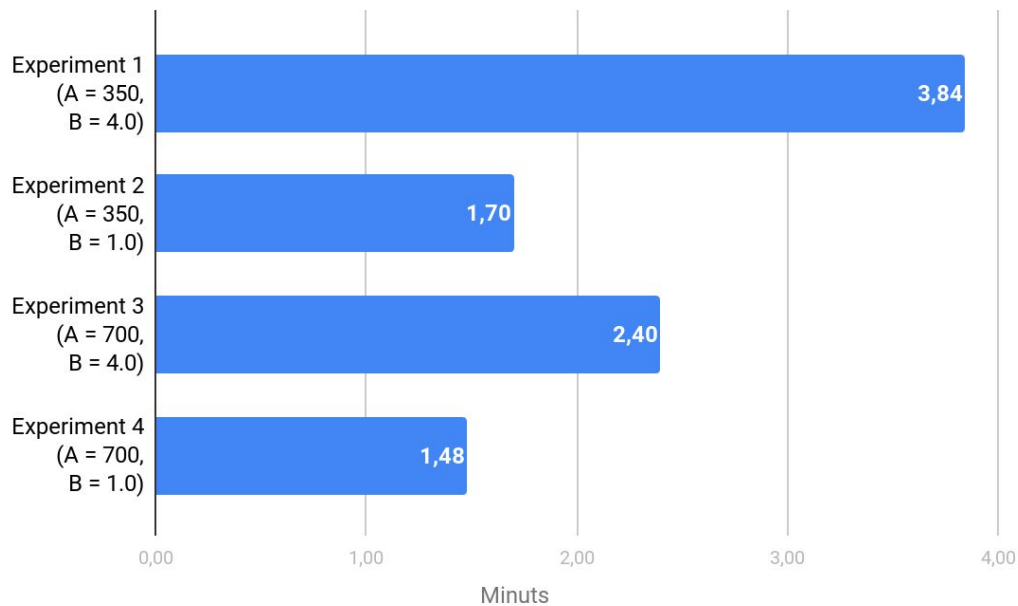
Desviació = NA

Interval de confiança ($\alpha = 0.05$) = [1.480008 , 1.480008]

Resum dels experiments

En la següent gràfica tenim una representació dels resultats obtinguts en els 4 experiments:

Temps d'espera mitjà (min)



Resultat de l'algoritme de Yates

Amb els resultats obtinguts als quatre experiments, hem aplicat l'algoritme de *Yates*. La taula resultant és la següent:

Exp.	A	B	Resultat	(1)	(2)	div	efecte	id
1	-	-	3,8423	5,5446	9,4221	4	2,35551775	Mitjana
2	-	+	1,7023	3,8774	-3,0575	2	-1,5287315	A
3	+	-	2,3974	-2,1400	-1,6672	2	-0,8335865	B
4	+	+	1,4800	-0,9174	1,2226	2	0,6112985	AB

Taula 3: Resultat de l'algoritme de *Yates*.

Com que volem obtenir el temps d'espera mitjà mínim, haurem de buscar les combinacions amb efecte menor.

Podem destacar l'experiment 2 (A) amb un resultat de 1,7 minuts i un efecte de -1,53 i l'experiment 3 (B) amb resultat de 2,4 minuts i un efecte de -0,83.

Amb aquests efectes caldria continuar en l'anàlisi dels resultats fent una comparativa de les variàncies, però això no entra en l'àmbit d'aquesta pràctica.

Conclusions i recomanacions

Gràcies a la modelització realitzada amb NetLogo, l'eina que vam escollir a l'entrega anterior, i l'aplicació de l'algoritme de Yates, s'han pogut obtenir uns resultats que permeten respondre al problema especificat als requeriments de la problemàtica.

L'elecció de NetLogo ha estat molt beneficiosa, ja que al poder reutilitzar la implementació que ja teníem, hem reduït el temps que haguéssim hagut d'invertir. A més, amb els gràfics que hi hem afegit, ens han permès veure a simple vista l'evolució del temps mitjà d'espera i del nombre de passatgers en una estació per fer-nos una idea de com es comportaria el model en la realitat. D'aquesta forma, hem pogut veure quan s'acumulava una quantitat desorbitada de persones en una estació concreta.

Dels quatre experiments del disseny realitzat, hem obtingut una sèrie de resultats que gràcies a RStudio hem pogut analitzar de forma senzilla. Segons els resultats, per tal de que els passatgers hagin d'esperar el mínim de temps possible, les combinacions de factors a tenir en compte degut als seus efectes en el temps mitjà d'espera són les dels experiments 2 (A) o 3 (B).

Per tant, en una hipotètica continuació de l'estudi caldria seguir l'anàlisi dels resultats a partir d'aquests dos resultats de la taula de Yates.