

Katera pot do fakultete je zares najhitrejša?

Timotej Knez

18. januar 2018

1. Uvod

Pot od doma do fakultete je zame najpogosteje prevožena pot, saj ji sledim prav vsak dan. Zaradi tega si želim to pot čimbolj optimizirati in s tem skrajšati čas, ki ga potrebujem, da pridem do fakultete. Če s tem uspem doseči, da grem lahko od doma le nekaj minut kasneje, lahko s tem dolgoročno prihranim veliko časa.

1.1. Opredelitev problema

Od doma do fakultete vsak dan potujem s kolesom. Tako lahko izberem več poti. Rad bi ugotovil, katera od poti je v resnici najhitrejša, pri čemer pa me ne zanima samo povprečni čas, ki je potreben za pot, ampak predvsem za katero pot je krajši čas, v katerem lahko z 95 odstotno gotovostjo trdim, da bom prišel do fakultete. Zaradi različnih standardnih odklonov časov je mogoče, da to ne bo pot, ki ima najkrajši povprečni čas. Rad bi ugotovil tudi, kako velik vpliv ima hitrost vožnje na čas, ki ga potrebujem, saj sumim, da malo hitrejša vožnja nima posebej velikega vpliva na čas potovanja.

1.2. Cilj projekta

Da bi ugotovili, katera pot je najhitrejša in kako je hitrost povezana s časom poti, bomo naredili program za simulacijo kolesarjenja po poteh, ki nas zanimajo. Z njim bomo lahko dobili zelo velik vzorec podatkov o trajanju voženj po danih poteh. Tako bomo lahko z analizo podatkov odgovorili na zastavljena vprašanja.

2. Predstavitev poti



Slika 1: Zemljevid vseh obravnavanih poti [4]

2.1. Prva pot (zelena):

Prva pot se drži večjih cest. Dolga je 4,34 km, na zgornjem zemljevidu pa je označena z zeleno barvo.

2.2. Druga pot (oranžna):

Druga pot sledi manjšim sprehajalnim potem, kar pomeni, da je manj odvisna od prometa, vendar pa je hitrost vožnje na teh poteh v splošnem počasnejša kot na cestah. Dolga je 4,84 km, na zemljevidu pa je označena z oranžno barvo.

2.3. Tretja pot (rumena):

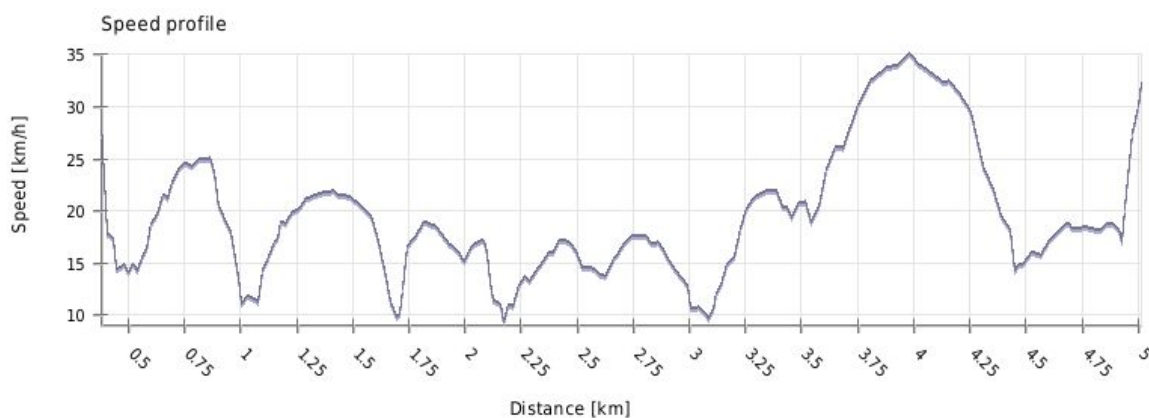
Tretja pot na začetku sledi drugi in se s tem izogne prečkanju železnice ter enemu od semaforjev, nato pa se pridruži prvi. Dolga je 5,15 km, na zemljevidu pa je označena z rumeno barvo.

3. Zbiranje podatkov

Trajanje različnih poti bomo določili z uporabo programa, ki bo simuliral vožnjo s kolesom po teh poteh. Za izdelavo takega programa najprej potrebujemo podatke o več parametrih, ki vplivajo na trajanje vožnje.

3.1. Hitrost vožnje po poti

Prvi podatek, ki ga potrebujemo za simulacijo, je hitrost vožnje po tej poti. Ta je zelo odvisna od poti in je eden od ključnih dejavnikov, ki določajo čas vožnje. Ker hitrost vožnje ni konstantna po celi poti, jo razdelimo na kratke dele in za vsakega posebej določimo hitrost, ki jo normalno dosežem na njem. Podatke o hitrosti vožnje na posameznih delih poti sem dobil tako, da sem z uporabo GPS tehnologije posnel več primerov vožnje po vsaki od poti. Podatke o hitrosti sem kasneje analiziral in določil pričakovano hitrost in njen odklon za vsak del poti.



Slika 2: Primer grafa hitrosti v odvisnosti od razdalje

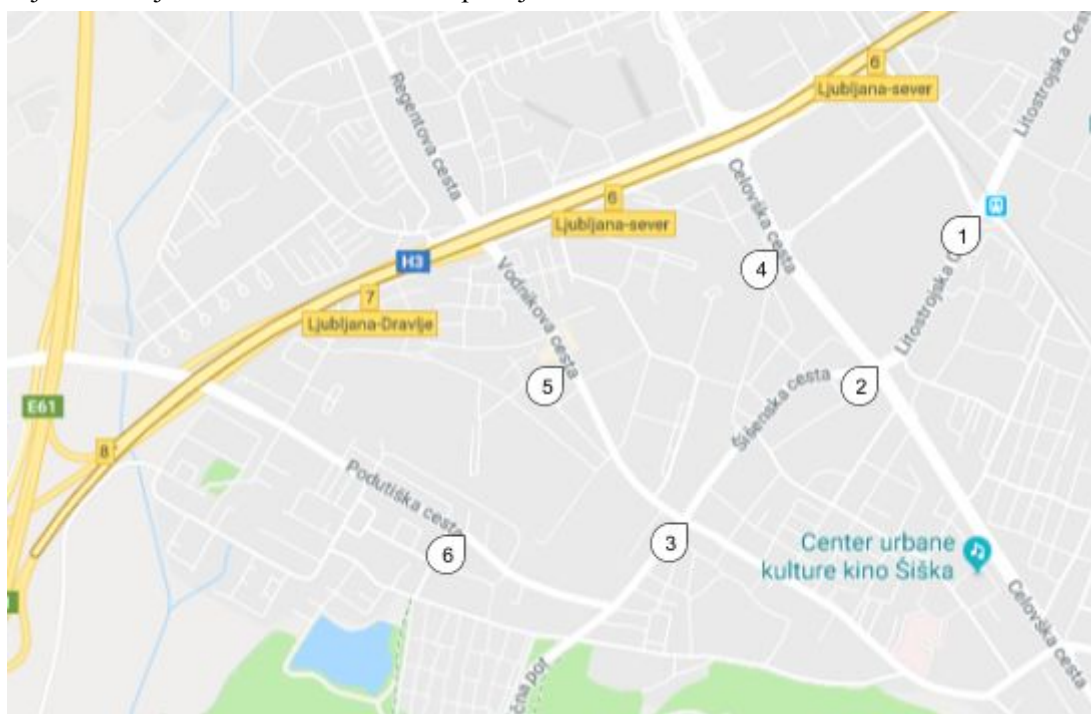
3.2. Trajanje rdečih in zelenih luči na semaforjih

Na skupno povprečno hitrost vožnje brez dvoma močno vplivajo tudi semaforji. Zato potrebujemo podatke tudi o teh. Te podatke sem zbral tako, da sem šel do vsakega semaforja, ki se pojavi na poti in izmeril, koliko časa trajata rdeča in zelena luč. Opazoval sem tudi, ali so morda nekateri semaforji usklajeni med sabo. Izkazalo se je, da sta med sabo povezana oba semaforja na Celovski cesti, ki pa ju nikoli ne prečkam na isti poti, zaradi česar ta povezava ne vpliva na čas moje poti. Časi celotnih ciklov vseh ostalih semaforjev pa so različni med sabo in zato ne morejo biti usklajeni. Čase semaforjev predstavimo v spodnji tabeli.

Tabela 1: Časi trajanja zelene in rdeče luči na semaforjih

SEMAFOR	ČAS RDEČE LUČI [s]	ČAS ZELENE LUČI [s]
Semafor 1	70	20
Semafor 2	108	15
Semafor 3	43	17
Semafor 4	108	15
Semafor 5	77	12
Semafor 6	62	8

Položaje semaforjev v tabeli označimo na spodnji sliki.



Slika 3: Zemljevid lokacij semaforjev [4]

3.3. Pogostost vlakov

Ena od poti prečka tudi železniško progo, zato potrebujemo tudi podatke o pogostosti prihoda vlakov na ta prehod. Podatke o potniških vlakih, ki prečkajo ta prehod, je enostavno dobiti kar iz voznega reda Slovenskih železnic [1], medtem ko podatkov o tovornih vlakih nisem uspel najti. Zato sem si začel zapisovati, koliko potniških in koliko tovornih vlakov sem videl na tej progi, iz česar lahko izračunamo razmerje med njimi. V nekaj tednih zapisovanja sem videl 17 potniških in 16 tovornih vlakov. Izračunamo interval zaupanja za delež potniških vlakov pri 10% tveganju z uporabo enačbe:

$$p \pm z_{\alpha/2} \sqrt{\frac{pq}{2}}$$

Pri tem upoštevamo, da je $p = 17/33$ in $q = 16/33$ in dobimo, da je delež potniških vlakov na intervalu $[0.37, 0.66]$. Glede na dobljeni interval lahko rečemo, da je približno polovica vseh vlakov na tej progi potniških. Iz voznega reda Slovenskih železnic [1] vidimo, da v času, ko grem ponavadi na fakulteto (med 7:30 in 10:30), prehod prečka 8 potniških vlakov. Če upoštevamo, da je $1/2$ vlakov na tej progi potniških, lahko rečemo, da je vseh vlakov v tem časovnem intervalu 16. Torej je povprečni čas med dvema vlakoma približno 11 minut. Potrebujemo tudi podatek, koliko časa so zapornice spuščene, ko pride mimo vlak. Ta podatek sem dobil tako, da sem preprosto izmeril nekaj časov, ko so bile zapornice spuščene, in ugotovil, da so v povprečju spuščene približno 1,5 minut.

3.4. Pospešek pri speljevanju in zaviranju

Potrebujemo tudi podatke o pospeških ob speljevanju in ustavljanju. Pospešek pri speljevanju določimo z analizo podatkov o hitrosti na poti, ki smo jih uporabili že prej. Poiskal sem več točk na grafih, kjer sem opazil speljevanje in izmeril, koliko časa traja, da dosežem končno hitrost in kako visoka je ta hitrost. S temi podatki lahko enostavno izračunamo pospešek ob speljevanju. Izmeril sem tudi razdaljo od semaforja, na kateri začnem običajno zavirati, če je na semaforju rdeča luč. Pospešek pri zaviranju se v programu sproti računa na podlagi hitrosti in razdalje do točke, na kateri se je potrebno ustaviti.

4. Izdelava računalniškega modela vožnje

Ko imamo vse potrebne podatke, lahko začnemo izdelovati računalniški model vožnje. Želimo, da program upošteva čim več dejavnikov, ki vplivajo na hitrost vožnje in posledično na čas, ki ga potrebujem za pot. Prvi element, ki ga mora program upoštevati, je hitrost vožnje na posameznih delih poti. To program doseže tako, da s pospeškom, ki smo ga izmerili prej, pospešuje oziroma zavira do hitrosti, ki je določena za ta del poti. S tem ko upoštevamo pospeške, dobimo bolj realistično simulacijo vožnje, kot bi jo, če bi se hitrost v hipu spremenila na končno. Naslednji pomembni elementi so semaforji in vlak. Ko pri teh kolesar pride na določeno razdaljo, začne zavirati, če je semafor rdeč ali če je na prehodu čez železnico vlak. Zavira s pospeškom, ki je potreben, da se ustavi ravno do semaforja oziroma prehoda čez železnico. Če je ta pospešek prevelik, pomeni, da je bil kolesar v trenutku, ko se je prižgala rdeča luč, preblizu semaforja, da bi varno ustavil, in zato nadaljuje brez zaviranja. Upoštevamo tudi to, da začne ponovno pospeševati, če semafor med zaviranjem postane zelen ali če se odpre prehod čez železnico. Za razdaljo, na kateri začne kolesar zavirati, sem na podlagi zbranih podatkov določil 130 m. Intervale rdečih in zelenih luči na semaforjih smo določili že z merjenjem trajanja posameznih luči na teh semaforjih, medtem ko so pri vlakih intervali zapiranja zapornic nekoliko bolj nepredvidljivi. Odločil sem se, da bom čas med dvema vlakoma določil po eksponentni porazdelitvi s pričakovano vrednostjo 11 minut, ki smo jo določili prej. Tako naredimo program, ki lahko simulira kolesarjenje po dani poti in nam s tem da velik vzorec podatkov o trajanju posamezne poti.

5. Izbira najhitrejše poti

Primerjali bomo tri že prej opisane poti, ki se zdijo najbolj perspektivne. Odločil sem, da bom za vsako pot izvedel 100.000 simulacij, saj s tako velikim vzorcem dobim precej zanesljive podatke.

5.1. Predstavitev podatkov

Za vsako pot izračunamo povprečni čas trajanja in njegov standardni odklon. Za to uporabimo naslednji formuli.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Izračunamo tudi intervale zaupanja za pričakovano vrednost pri 5% tveganju. Pri tem uporabimo formulo:

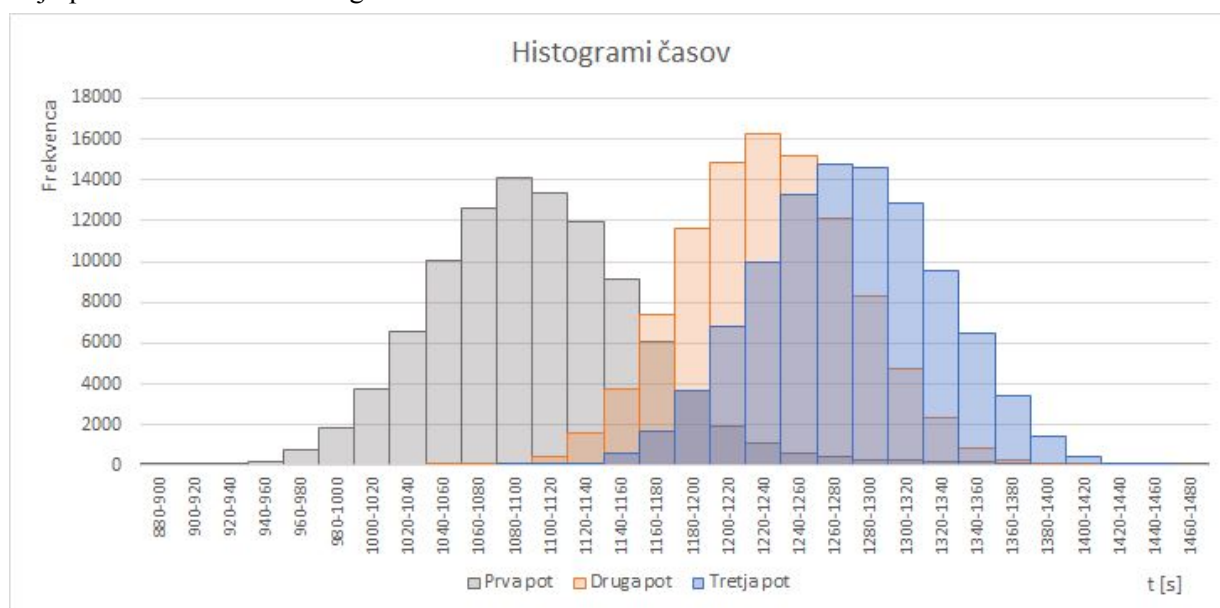
$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Podatke o poteh predstavimo v spodnji tabeli:

Tabela 2: Pričakovana vrednost časov

	povprečni čas	standardni odklon časov	Interval zaupanja za pričakovani čas
Prva pot	1103,3 s (18 min 23 s)	64,7 s	[1102`8, 1103`7]
Druga pot	1231,6 s (20 min 32 s)	47,3 s	[1231`3, 1231`9]
Tretja pot	1277,1 s (21 min 17 s)	50,8 s	[1276`8, 1277`4]

Podatke predstavimo s pomočjo histograma. Čase sem razdelil v razrede široke 20 sekund in jih za lažjo predstavo združil na en graf.



Slika 4: Histogrami časov za vse tri poti

5.2. Preverjanje domneve o normalni porazdelitvi časov

Ko pogledamo zgornje histograme, opazimo, da močno spominjajo na normalno porazdelitev. Tako postavimo domnevo, da so časi porazdeljeni s porazdelitvijo $N(\mu, \sigma)$. To domnevo preverimo s pomočjo hi kvadrat testa. Za test uporabimo statistiko

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - f'_i)^2}{f'_i},$$

kjer je f_i izmerjeno število vzorcev v danem razredu, f'_i pa teoretično izračunano število vzorcev v istem razredu, če bi se vzorci porazdelili normalno. Slednjo lahko izračunamo s pomočjo formule

$$f'_i = n * \left(\Phi\left(\frac{b - \bar{x}}{\sigma}\right) - \Phi\left(\frac{a - \bar{x}}{\sigma}\right) \right),$$

kjer je n = velikost celotnega vzorca, a = spodnja meja razreda in b = zgornja meja razreda.

Testna statistika se porazdeli po χ^2 s $k - 1$ prostostnimi stopnjami, kjer je k enak številu razredov, v katere smo razdelili podatke. Domnevo preverimo pri 5% stopnji značilnosti. Pri tem sta ničelna in alternativna domneva

$$H_0 : \chi^2 = 0 \quad \text{in} \quad H_a : \chi^2 \neq 0,$$

kritična vrednost pa je enaka

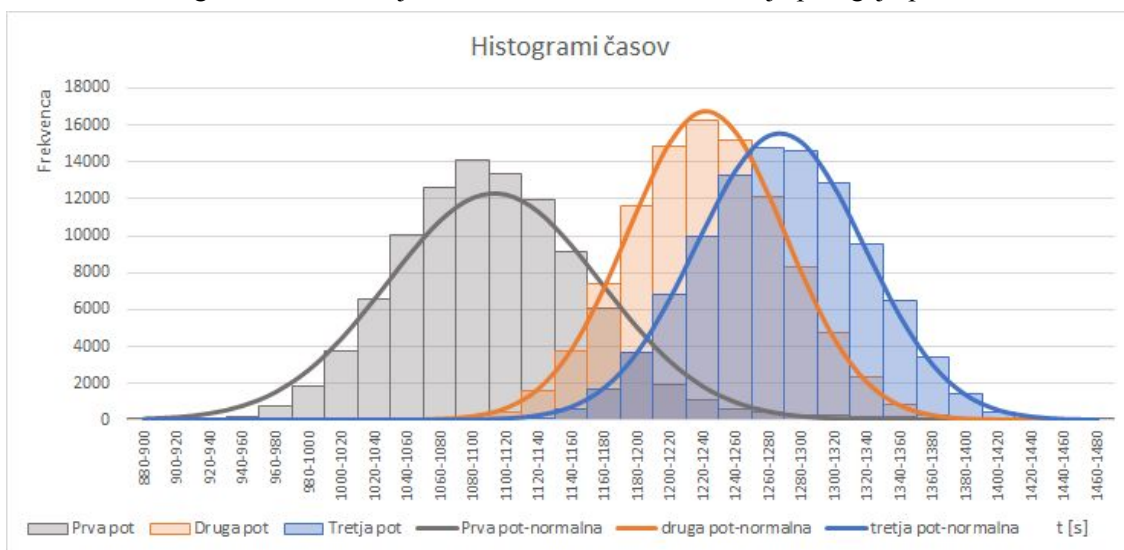
$$\chi^2_{1-\alpha}(k-1) = \chi^2_{0,95}(29) = 42,6.$$

Sedaj lahko izračunamo testno statistiko za vse poti in rezultate predstavimo v tabeli.

Tabela 3: Vrednosti testnih statistik za preverjanje normalnosti porazdelitve

	Prva pot	Druga pot	Tretja pot
Vrednost testne statistike	3959	277	431

Pri vseh treh poteh je kritična vrednost enaka, kar pomeni, da lahko domnevo o normalni porazdelitvi časov pri vseh treh poteh zavrnemo, saj so testne statistike v kritičnem območju. S tem ugotovimo, da se pri nobeni poti časi ne porazdeljujejo povsem normalno. Izračunane normalne porazdelitve lahko dodamo še na histograme in tako bolj vidimo, kako dobro se krivulje prilegajo podatkom.



Slika 5: Histogrami časov s prikazanimi pričakovanimi normalnimi porazdelitvami

Opazimo, da se podatki prve poti močno razlikujejo od pričakovane krivulje, kar je verjetno posledica tega, da so nekateri časi pri tej poti precej večji od povprečja, zaradi česar porazdelitev ni povsem simetrična. Izračunamo lahko tudi koeficient asimetrije po formuli

$$A(X) = \frac{E((X - E(X))^3)}{D(X)^{3/2}}$$

Tabela 4: Asimetričnost razporeditve časov

	Prva pot	Druga pot	Tretja pot
Asimetričnost	1,48	0,07	-0,01

Če upoštevamo, da ima simetrična porazdelitev asimetričnost enako 0, vidimo, da je prva pot res precej bolj asimetrična od ostalih dveh, kar potrjuje, da je bolj oddaljena od normalne porazdelitve.

5.3. Določanje najhitrejše poti

To nas pripelje do prvotnega vprašanja, to je, katera pot je pravzaprav najhitrejša. Na to vprašanje lahko enostavno odgovorimo tako, da pogledamo, katera od poti v povprečju zahteva najmanj časa. Povprečne čase smo že izračunali, zato hitro ugotovimo, da je vrstni red naslednji:

1. Prva pot $\mu = 1103$ s (18 minut in 23 sekund)
2. Druga pot $\mu = 1232$ s (20 minut in 32 sekund)
3. Tretja pot $\mu = 1277$ s (21 minut in 17 sekund)

Mnogo bolj zanimivo primerjavo pa naredimo, če namesto povprečnih časov vzamemo čas, za katerega lahko z 95% gotovostjo trdimo, da bo pot trajala manj kot toliko časa. Ta primerjava je smiselna, ker s tem podatkom lahko določimo, kdaj se je potrebno odpraviti na pot, če želimo biti 95% prepričani, da ne bomo zamudili. Ker ne poznamo porazdelitve, po kateri se porazdeljujejo časi, je ta čas najbolj enostavno poiskati tako, da določimo kvantil reda 0,95. Tega najlažje poiščemo, če vse čase, ki smo jih dobili, uredimo po vrsti in gledamo podatek na mestu $n * 0,95$, kjer je n število podatkov. V našem primeru gledamo 95.000. podatek. Tako dobimo naslednji vrstni red:

1. Prva pot $x_{0,95} = 1204$ s (20 minut in 4 sekunde)
2. Druga pot $x_{0,95} = 1311$ s (21 minut in 51 sekund)
3. Tretja pot $x_{0,95} = 1361$ s (22 minut in 41 sekund)

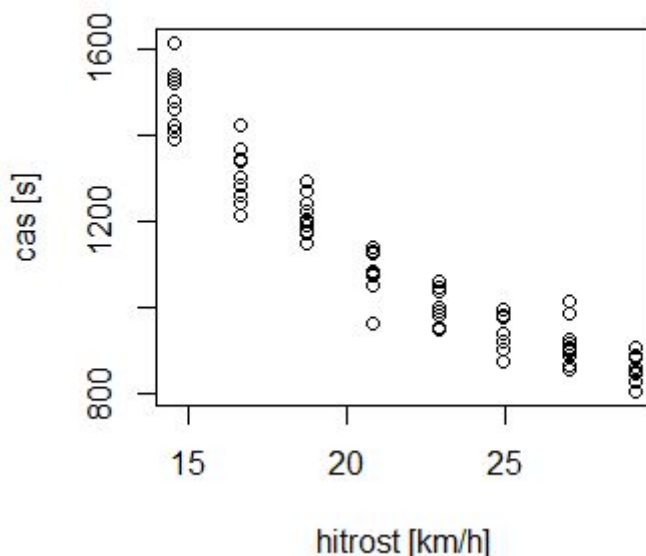
Vrstni red poti je enak, kot ko smo opazovali povprečje časov, vendar pa opazimo, da so tukaj razlike med časi manjše. Tako se pojavi ideja, da morda obstaja gotovost, pri kateri se vrstni red zamenja. Ko to preverimo, ugotovimo, da druga pot prehiti prvo, če je verjetnost, s katero želimo biti točni, večja od 99,3%, tretja pa prehiti prvo, če je verjetnost večja od 99,6%. Če gledamo kvantil reda 0,999, je tako vrstni red:

1. Druga pot $x_{0,999} = 1375$ s (22 minut in 55 sekund)
2. Tretja pot $x_{0,999} = 1423$ s (23 minut in 43 sekund)
3. Prva pot $x_{0,999} = 1527$ s (25 minut in 27 sekund)

Kljub temu lahko zaključimo, da je prva pot najhitrejša pod pogoji, ki smo jih določili na začetku.

6. Povezava med hitrostjo vožnje in časom prihoda

Ko smo ugotovili, katera pot je najhitrejša, nas zanima še, koliko na čas vožnje v resnici vpliva hitrost vožnje in pospeševanja. Da lahko to ugotovimo, sem v program za simulacijo vožnje dodal še možnost za nastavljanje vrednosti, s katero se pomnožijo vse hitrosti in pospeški na poti. Tako lahko simuliramo vožnjo, ki bi jo dosegli z močnejšim ali šibkejšim poganjanjem. Čase bomo merili na prvi poti, saj je ta izbrana kot najhitrejša. Rezultate predstavimo v obliki razsevnega diagrama.



Slika 6: Razsewni diagram časov v odvisnosti od hitrosti

Najprej preverimo ali obstaja linearna odvisnost med hitrostjo in časom, potrebnim za pot. To naredimo tako, da postavimo domnevo, da čas in hitrost nista linearno povezana in poskusimo to domnevo ovreči pri 5% stopnji značilnosti.

Najprej izračunamo Pearsonov koeficient korelacije. V našem primeru x predstavlja povprečno končno hitrost vožnje, y pa čas vožnje.

$$\bar{x} = 21,8 \quad \bar{y} = 1100$$

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 * \sum_{i=1}^n (y_i - \bar{y})^2}} = -0,95$$

Vidimo, da je Pearsonov koeficient korelacije blizu -1, kar nakazuje na to, da sta vrednosti med sabo povezani. Sedaj lahko poskusimo zavreči domnevo, da vrednosti nista linearno povezani, in s tem potrdimo povezavo med njima.

$$H_0 : \rho = 0 \quad H_a : \rho \neq 0$$

$$TS = \frac{r_{XY} * \sqrt{n-2}}{\sqrt{1-r_{XY}^2}} = -26,9$$

Testna statistika se porazdeljuje po Studentovi porazdelitvi, tako dobimo kritično območje

$$\pm t_{\alpha/2}(n-2) = \pm t_{0,975}(78) = \pm 1,99$$

Testna statistika je v kritičnem območju, zato lahko zavrnemo hipotezo, da med hitrostjo in časom ni linearne korelacije.

Naslednji korak je, da izračunamo linearno regresijsko funkcijo. Regresijska premica bo imela obliko

$$\hat{Y} = a + bX,$$

kjer sta a in b:

$$b = \frac{c_{xy}}{c_x^2} \text{ in } a = \bar{Y} - b\bar{X}.$$

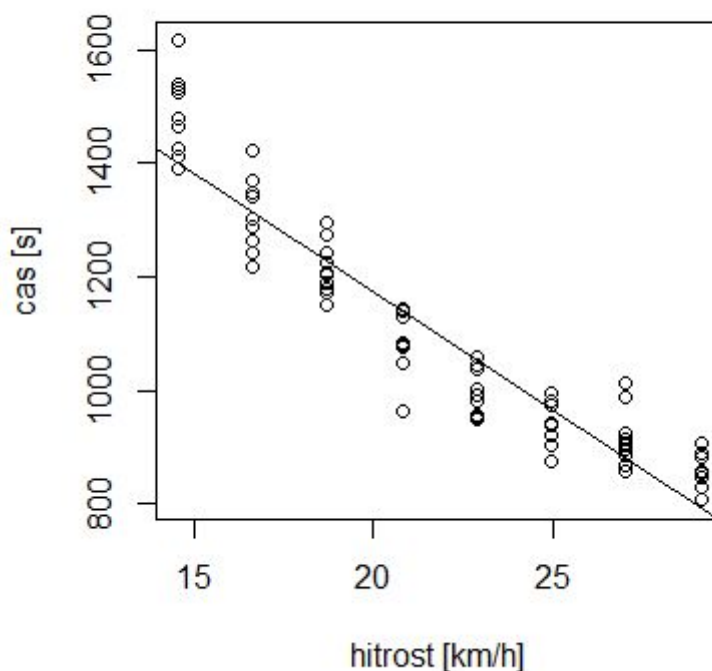
$$c_x^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = 1817 \text{ in } c_{xy} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = -75688$$

$$b = -42 \quad a = 2010$$

Premica, ki jo tako dobimo, je

$$Y = 2010 - 42 * X.$$

Regresijsko premico lahko dodamo še na razsevni diagram in vidimo, ali se res prilega podatkom.



Slika 7: Razsevni diagram z narisano regresijsko premico

Sedaj vidimo, da se, če končno hitrost vožnje iz 20 km/h povečam na 25 km/h, čas zmanjša iz 19,5 minut na 16 minut, kar pomeni, da je z malenkost hitrejšo vožnjo v resnici mogoče znatno zmanjšati čas potovanja.

7. Zaključek

Preučili smo tri kandidate za najhitrejšo pot do fakultete in ugotovili, da je prva pot pod pogoji, ki smo jih določili, najhitrejša, kar se sklada tudi z mojimi pričakovanji. Ugotovili pa smo tudi več presenetljivih stvari, ki jih nisem pričakoval. Najbolj me je presenetilo, da smo ugotovili, da se časi poti ne razporejajo povsem normalno, saj sem bil ob pogledu na histograme prepričan, da je prikazana porazdelitev normalna. Verjetno bi se normalni porazdelitvi približali, če bi bile simulacije poti sestavljene iz še več med seboj neodvisnih naključnih dejavnikov. Druga presenetljiva ugotovitev je bila, da se ob povečanju gotovosti, s katero želimo biti točni, vrstni red poti spremeni, kar pomeni, da nobena pot ni absolutno najhitrejša. Ugotovili smo tudi, da sta hitrost vožnje in trajanje poti linearno povezani in da je mogoče z malo hitrejšo vožnjo opazno skrajšati čas potovanja. Če grem torej od doma kakšno minuto prepozno, je še vedno mogoče prispeti pravočasno, če vozim nekoliko hitreje.

8. Viri

- [1] (2017) Vozni red slovenskih železnic. Dostopno na:
http://www.slo-zeleznice.si/images/Potniski_fotografije/potniski_pdf/vozni-red-2017-2018/03_ljubljana_jesenice.pdf
- [2] A. Jurišić. Verjetnostni račun in statistika, zapiski, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, 2017.
- [3] (2017) uTrack - online GPX track report generator. Dostopno na: <http://utrack.crempa.net/>
- [4] (2018) Google Zemljevidi. Dostopno na: <http://maps.google.com/>
- [5] (2010) Momenti porazdelitev. Dostopno na:
<http://km.fgg.uni-lj.si/PREDMETI/sei/Momenti%20porazdelitev.PDF>