FIB-ADEI

(Data: 2/June/2017 18:00-20:00 h Lloc: Aula – B5S101)

Nom de l'alumne:	
DNI:	

Lecturer:	Lídia Montero
Office:	Edifici C5-D217
Rules for the Quiz:	ES PERMÉS DE DUR ELS APUNTS PUBLICATS ES POT DUR CALCULADORA
Duration:	2h 00
Marks:	Before 9/06/17 at 3 pm on ATENEA.
Open-office:	2/June/17 at 3 pm (C5-217).

Problem 1: All qüestions account for 1 point

Data of a traffic simulation experimental design on Connected Car impact in a urban network (Central Bussiness District of Barcelona, Eixample). Eighteen variables were recorded for 30 configurations and 5 replications for each configuration giving a total of 150 executions. Default statistics in traffic microsimulation platforms are usually very rich. Statistics have been collected every 90 sec. Network Performance indicators are global statistics for all drivers over the whole simulation horizon (1h, morning period). Factors involved in the experimental design are:

- Factor TD –Driver Types into Expert-Regular-Tourist into 4 levels. Base level: 40-50-10. Alternative levels to be considered: 20-70-10, 40-40-20 and 60-20-20.
- Factor GP – Guidance Penetration. Percentage of total drivers whose route choice decisions follow those advised by a Navigation Tool modeled through 9 levels. Base level is 0% and alternative levels 10%, 20%, 30%, 50%, 70%, 80%, 90% and 100%.
- Factor DP- Demand Pattern into 4 levels 0, 10, 20 and 30. The base level 0% means historic demand pattern for the simulation horizon, while 10%, 20% and 30% levels refer to the increment over the historic pattern (D) on long OD pairs
- Factor PVD- Probe Vehicle Data Penetration modeled as invariant to driver type into 4 levels: 0%, 10%, 20% and 30%.
- Factor TW-Time-Window length is the rolling horizon interval considered for the estimation of traffic variables from Probe Vehicle Data into 3 levels. At base 3 min and alternative lengths of 1.5 min and 6 min are also considered.
- Factor NS-Navigation Strategy as driving recommendations based on either lane-level or link-level elaborated Probe Vehicle Data. Base level is lane-level.

This exercise focuses on the following **Key Performance Indicators** (KPI list) and defines as selected target **fuelc (liters) - Total fuel consumed**:

- Total Travelled Distance (km) – ttdis.
- Total Travelled Time (h) – ttt.
- **fuelc (l) Total litres of fuel consumed – fuelc**.
- CO2 (kg) – Total co2
- NOx (kg) - Total nox
- Average Travel Time (sec/km) - mtt. Seconds needed to cover one km.
- Mean Delay time (sec/km)-mdelay. Seconds of delay when covering one km.
- Density (veh/km) – density. Number of vehicles per km.
- Average speed (km/h) - mspeed.
- Average Flow (veh/h) Thoroughput measure - thrput.
- Thoroughput rate (%) completed trips divided into total number of trips for 1h – thrputrate.

Each register contains the description of the configuration (according to design factors) and the resulting KPIs.

Source

Montero, L., Linares, M.P., et al. (2018) *An Experimental Design For Simulation Assessment Of Navigation Strategies Based On Probe Car Data*, to be presented at Transportation Research Board Annual Meeting, Washington D.C.

Load TRB2018.RData file in your current R or RStudio session. You have to consider only the variables listed at the beginning of this exercise (KPI list).

1. **Variable fuelc** contains the total fuel consumption for all trips of the selected configuration. Summarize numerically and graphically the response variable. Make an interpretation of the results. Do you think that **fuelc** may be considered normally distributed? Justify your answer with at least two arguments.
2. Which are the variables statistically associated with the response (**fuelc**)? Indicate the suitable measure of association and/or tests that support your answer.
3. The average **fuelc** can be argued to be the same for all **Demand Pattern** levels (**DP)?** Which are the groups that show a significant greater average **fuelc** than the others?
4. Let us perform a one-way analysis of variance on TRB2018 dataset for target **fuelc** on **Demand Pattern** levels (**DP)** factor using the standard general linear model (lm() method). Assess the explicability of the model and interpret the model for prediction purposes writing the equation for each GP level.
5. Use an inferential method to quantify the probability of the null hypothesis of a negligible gross effect for **demand pattern (DP)** factor on total fuel consumption **(fuelc)**.
6. Let us perform a two-way analysis of variance on TRB2018 dataset for target **fuelc on Guidance Penetration (GP) and Demand Pattern (DP) factors** using the standard general linear model (lm() method). Are interactions needed for GP and DP to explain the total fuel consumption in the network? Are net effects for both factors significant?
7. Assess the explicability of the best model selected in question 6 and calculate the point and 95% confidence prediction interval for total fuel consumption in an scenario with a guidance penetration of 80% and demand level of 30%. Try to explain the reason of warning messages appearing for some methods depending on the selected mode.
8. Consider a general linear model for **total fuel consumption (fuelc)** when all numeric variables included in the list of KPI are taken into account. Assess explicability, try to reduce the number of explanatory variables without losing too much explicability and discuss colinearity among explanatory variables.
9. Consider linear models including as explanatory numeric variables **total travel distance (ttdis)** and **speed** and main effect of guidance penetration (GP) factor. Discuss the selected model, while justifying the choice (neither residuals, nor influent data analysis has to be undertaken at this point).
10. Consider the model for **fuel consumption proposed in the previous question**. Does the model fulfill the properties for linear models? Justify your answer using statistical arguments. Indicate the presence of lack of fit configurations and influent data.