Supplementary materials for the paper "Testing software for non-discrimination: an updated and extended audit in the Italian car insurance domain"

A Preliminary Exploratory Analysis

A.1 Methodology

Analysis Method We performed a preliminary exploratory analysis by analysing the prices' distribution for each attribute value. As an example, it assesses the spread between the average premium for male and female drivers. We perform a top1 and top5 average analysis, as in the original study. The top1 analysis concentrates exclusively on the most affordable quote for each profile: this is particularly relevant from the viewpoint of a person who is primarily concerned with the lowest insurance cost. In contrast, the top5 analysis compares the averages of the five cheapest quotes for every profile, catching a deeper view of the insurance rates for a specific profile.

A.2 Results

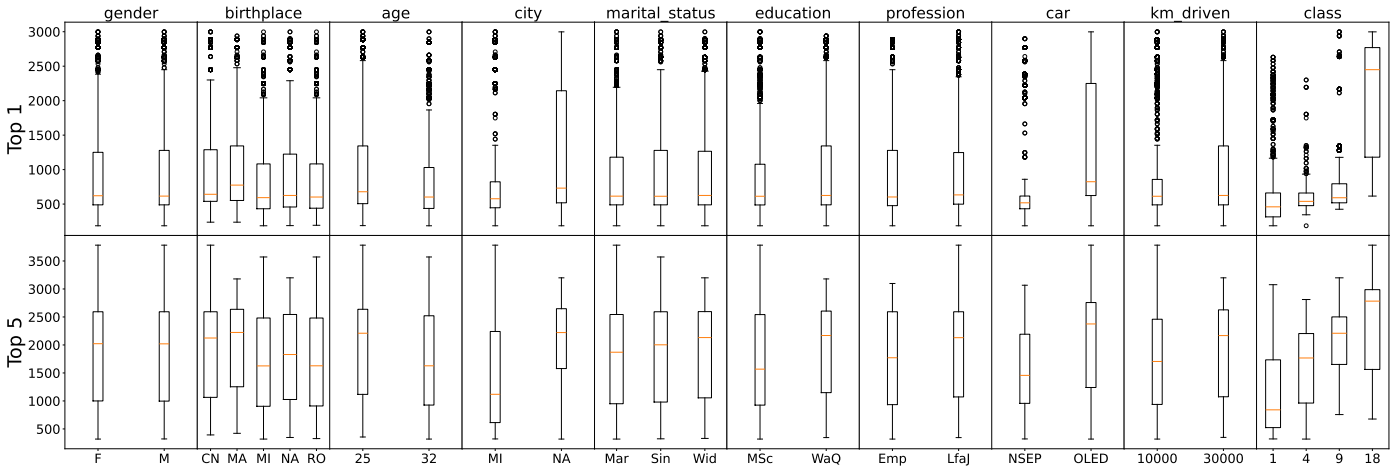

Fig. 1. Distribution of1 the top1 values (the cheapest quote) and of the top5 values (the five cheapest quotes).

This section presents the distributions of quotes for each of the characteristics under analysis, starting with the demographic variables and continuing with the driving variables. Figure 1 shows the box plots for the different features. Overall, the top5 values show a wider distribution, as they include higher quotes. Looking at the individual variables:

- The *gender* analysis of the cheapest insurance quotes shows no differences in median values and distribution.
- *Birthplace*: it emerges that profiles born in Milan are at an advantage compared to others born in Naples, and that profiles born in China and Morocco have higher prices compared to profiles born in Italy.
- Regarding the variable *age*, results show that 25-year-old profiles get significant higher quotes in top5 analysis.
- The *city of residence*, has a significant impact on the prices: on the one hand, the median of Naples is slightly higher than the median of Milan for the top1 offers, but with a much wider variability towards the top of the distribution; on the other hand, if we look at the top5 offers, the difference is very large, ranging from offers below 1000 € (in Milan) to offers above 2000 € (in Naples).
- Observing the cheapest quote of *marital status*, there is no effect on insurance prices. In the top5, there is a slight increase from married to single and widowed profiles.
- *Educational qualification*: profiles without a qualification receive higher quotes in relation to profiles with a master's degree. This result is only visible in the top5 analysis.
- Regarding *employment*, there is no significant difference in the distribution of the cheapest offers. Extending the analysis, a price advantage for employees can be recognised.
- *Car type*: profiles with an OLED (Old, Large Engine, Diesel) car type have higher insurance prices than profiles with an NSEP (New, Small Engine, Petrol) car type.
- *Km travelled in one year*: on average, profiles with 30,000 km/year receive higher prices.
- As expected, the *risk class* turns out to be a variable directly proportional to the final insurance price. The increases from class 4 to class 9, and for 9 to 18, are more pronounced than that from class 1 to class 4.

Overall, the box plots show variations for both demographic variables and driving variables. The first research question will delve into the details of the potential discrimination.