Fast estimates of proportions and occupancy numbers of hotels

1 Introduction

The number of overnight stays at hotels is usually reported based on data collected from all hotels in Iceland. This however results in a longer time to publication of results.

Statistics Iceland aims to provide information much faster. This is done by using the reports of a sample of the total number of hotels, stratified and weighted appropriately for estimating the proportions of occupied beds.

The occupancy numbers, i.e. total number of overnight stays, are then estimated based on the sample proportions and the now-casting of the total number of beds in all hotels.

2 Testing the method on past data

The true value of the occupancy rates (p) is, with α confidence, contained by the confidence intervals of the sample estimates $\hat{p}$ (when the total population N is finite), with the lower and upper bounds given by:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot \sqrt{\frac{1-n/N}{1-1/N}} \tag{1}$$

In order to evaluate how good the estimate

$$\hat{p} = n^{occupied}/n \tag{2}$$

is, we compared the observed values of the rates (proportions) for past data with the confidence intervals (CI) of the $\hat{p}$. Indeed, the observations are contained by these intervals and one may note that when testing these hypotheses a Bonferroni multiplicity adjustments dependent on the number of tests was applied.

In the equations above, n is the sample total of beds (for the whole month), $n^{occupied}$ is the number of occupied beds (during the same month) in the sample.

3 Estimating the current proportions and total number $\hat{N}^{occup}$

The proportions are calculated according to the eq.(1) and (2). The total number of overnight stays is estimated by:

$$\hat{N}^{occup} = \hat{p} \cdot \tilde{N} \tag{3}$$

where $\hat{p} = n^{occupied}/n$ (all small n s are measured on sample) and $\tilde{N}$ is the predicted (forecast) of the total number of available beds in the population of beds.

The error of $\hat{N}^{occup}$ is:

$$e(\hat{N}^{occup}) = e(\hat{p})\tilde{N} + e(\tilde{N})\hat{p} \tag{4}$$

where $e(Z)$ denotes the error of variable Z.

By using the data of monthly numbers of beds for the past five years $N(t)$ (a time series exhibiting both trend and seasonality) and an exponential state space model forecast with horizon h=1, one may predict the current value of $\tilde{N}$ and the lower and upper limits of its 95% prediction interval. This is shown in Figure 1, for the month of March 2020.

The results are then used in the two equation above for calculating the current estimates and the corresponding errors.

4 Important remarks

1. The samples used monthly are not random but an appropriate stratification and weighting scheme compensates for that.

2. When testing the fitting of estimayes for the past data, we adjust for multiplicity by using Bonferroni adjustment (by using $\alpha/2m$ where m is the number of tests)

3. When calculating the current CI, $z_{\alpha/2}$ is the the critical value corresponding to the desired confidence for one point estimate.

4. The problem of extreme values and outliers will modify our approach for the future months.

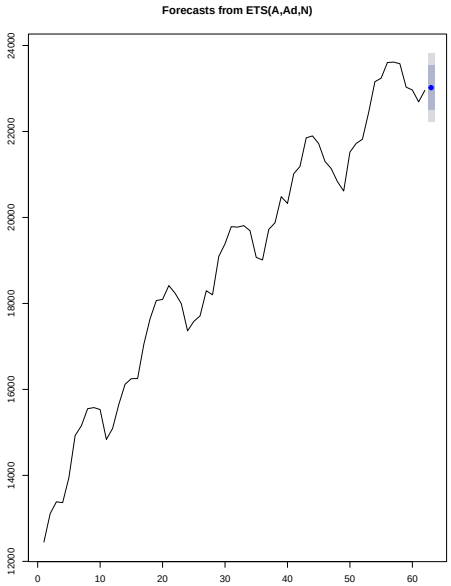

Figure 1: Now-casting the total number of hotel beds