

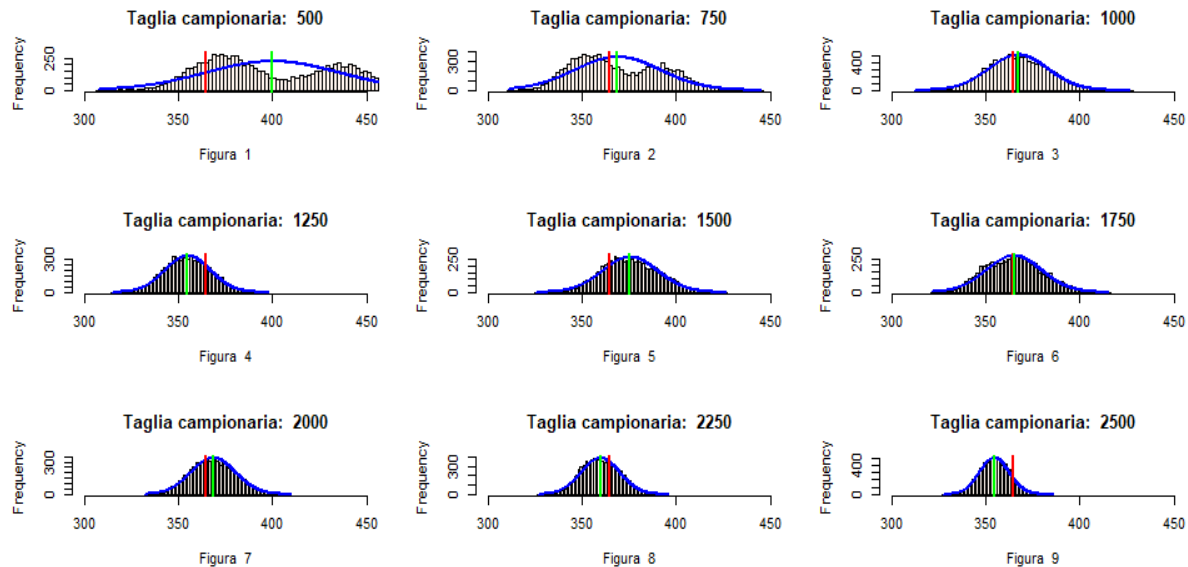
# Analisi Bootstrap

D'alto Jacopo, Mattazzi Anna Chiara, Quercini Luca, Spinelli Sonia

Si effettua un'analisi bootstrap per il dataset sugli alloggi di Airbnb londinesi, già analizzato in precedenza. Lo scopo è quello di studiare l'effetto del Teorema del Limite centrale sullo stimatore del parametro media di `realSum` al variare della taglia campionaria e della bontà del campione iniziale.

## AL VARIARE DELLA TAGLIA CAMPIONARIA

Per studiare il bootstrap all'aumentare della numerosità del campione, sono selezionate 9 taglie, da 500 a 2500 con passo 250. Si effettua, quindi, un ricampionamento di due quinti delle unità campionate, in modo da avere un sottocampione significativamente più piccolo della popolazione accessibile. Il risultato ottenuto è riportato di seguito. Si nota che il teorema del limite centrale funziona meglio per taglie campionarie alte.



Per esempio le *figure 1 e 2* hanno taglie campionarie basse e, come si può osservare, sono più facilmente soggette a bias. La loro distribuzione è influenzata dagli outliers e hanno alta varianza. Dalla taglia 1000 in poi la distribuzione è approssimabile con una gaussiana di uguale media e varianza. Oltre a ciò, si nota che all'aumentare della numerosità del campione la varianza osservata diminuisce.

Più unità sono campionate, maggiore è la precisione della stima della media. Tuttavia, è opportuno tener conto della fluttuazione aleatoria del campionamento.

## AL VARIARE DELLA BONTÀ DEL CAMPIONE

Al fine di mettere in evidenza l'importanza della rappresentatività del campione per il metodo bootstrap, si decide di costruire il campione peggiore possibile. Vengono selezionate le unità sperimentali del primo e quarto quartile, ovvero i valori estremi, per un totale di 2676 unità.

Dal grafico in *figura 10*, è evidente che il bias è alto, perché la media stimata con il bootstrap si discosta molto dal valore reale: con un campione non rappresentativo si perde accuratezza.

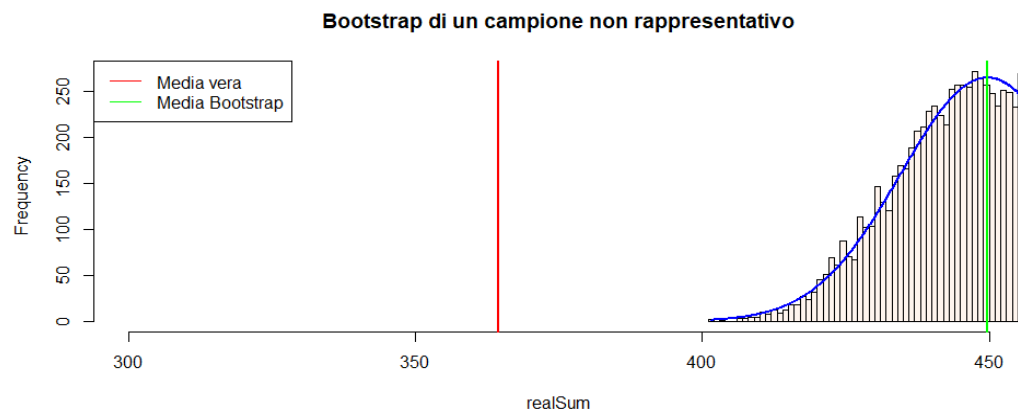


Figura 10

In conclusione, come evidenziato in precedenza, la bontà dell'analisi bootstrap dipende dalla taglia campionaria scelta e dalla rappresentatività del campione stesso.

Per taglie campionarie basse, infatti, si ottiene una stima meno precisa e una varianza molto alta. Invece, se il campione non è rappresentativo la distorsione è alta e la stima non è accurata.