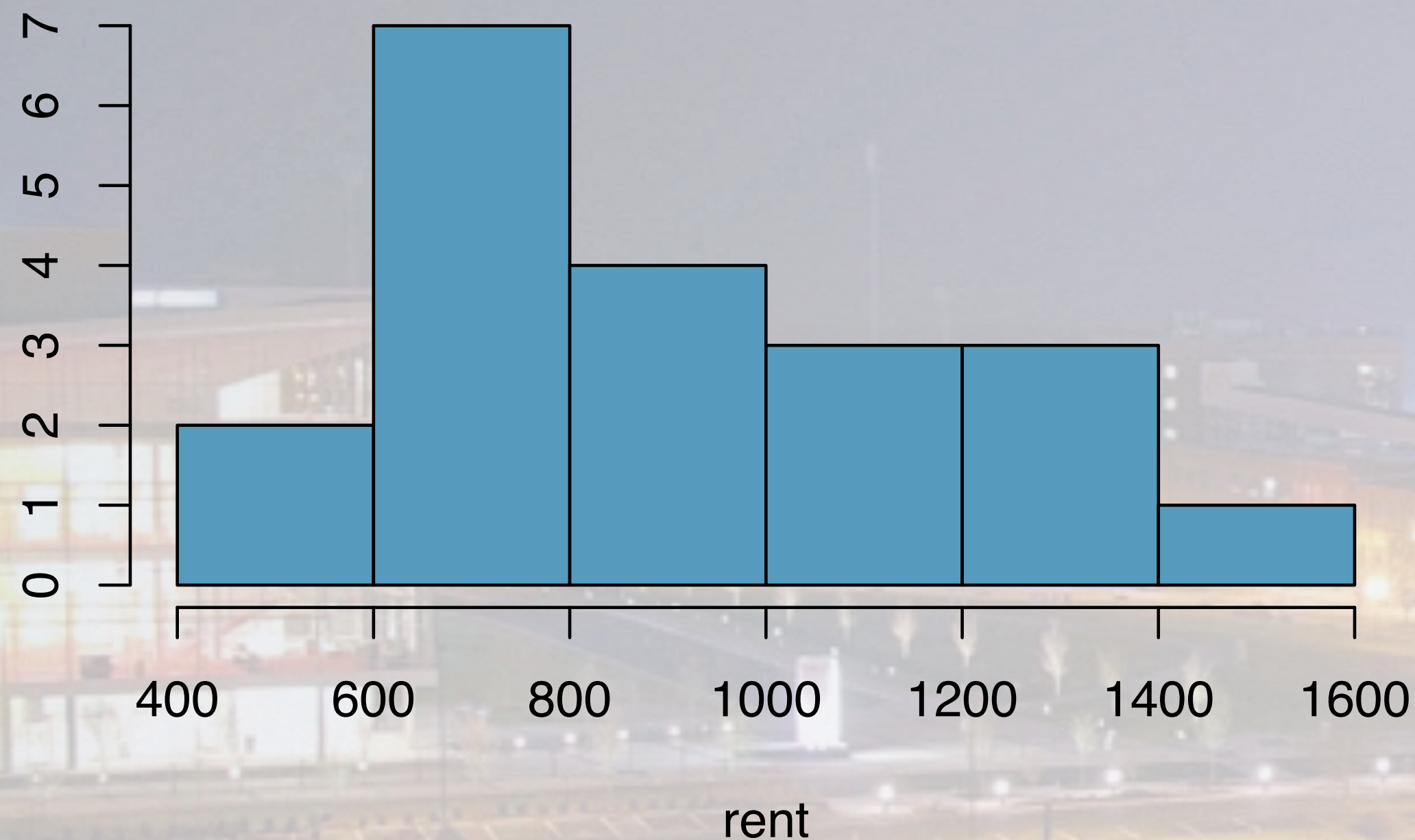


bootstrapping

- ▶ what is bootstrapping?
- ▶ estimation beyond means
- ▶ limitations

rent in durham, nc



Twenty 1+ bedroom apartments were randomly selected on raleigh.craigslist.org. (keyword: **Durham**). Is the mean or the median a better measure of typical rent in Durham?

Can we apply CLT based methods we have learned so far to construct confidence intervals for both?



bootstrapping

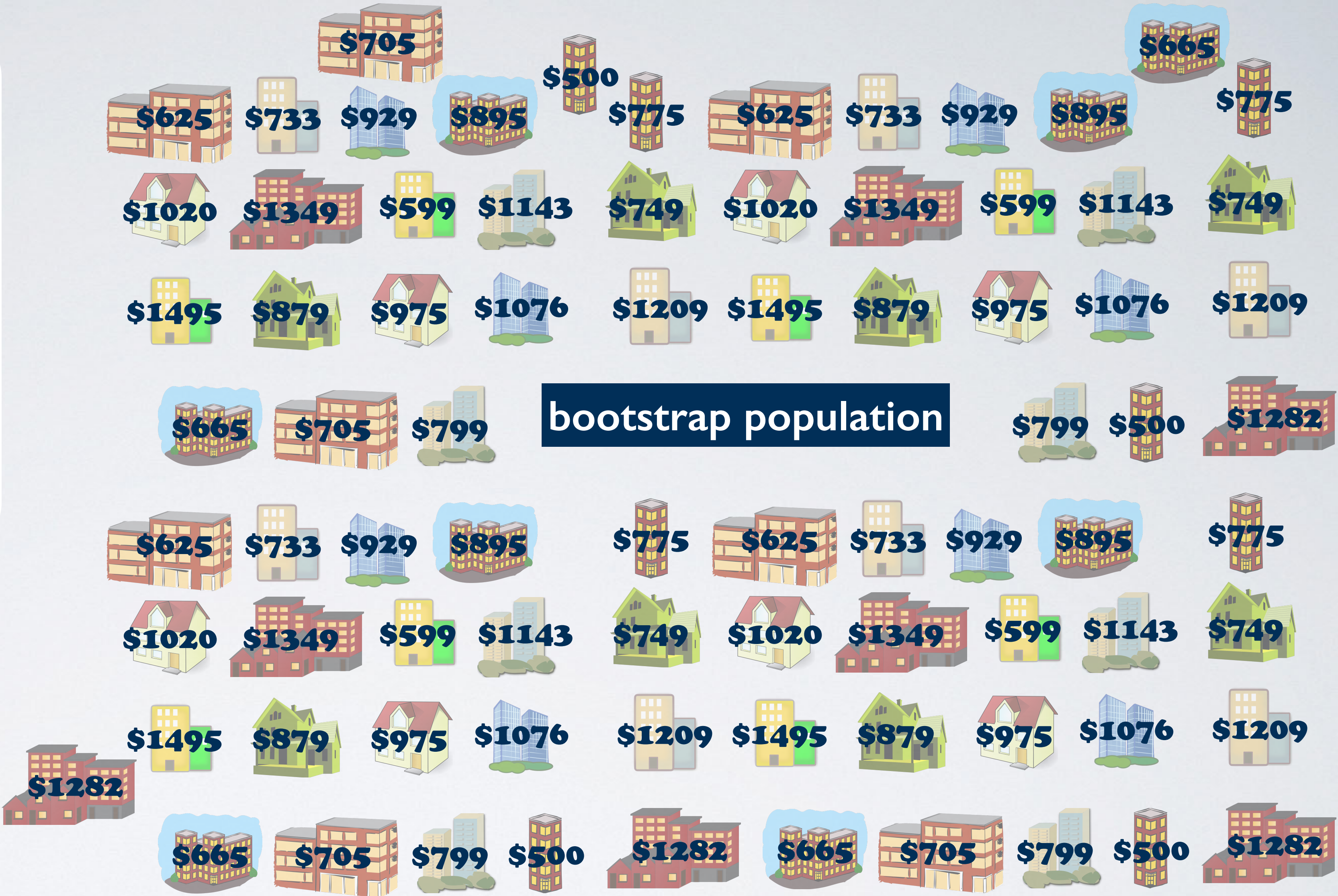
- ▶ An alternative approach to constructing confidence intervals is **bootstrapping**.
- ▶ This term comes from the phrase “*pulling oneself up by one’s bootstraps*”, which is a metaphor for accomplishing an impossible task without any outside help.
- ▶ In this case the impossible task is estimating a population parameter, and we’ll accomplish it using data from only the given sample.

original sample



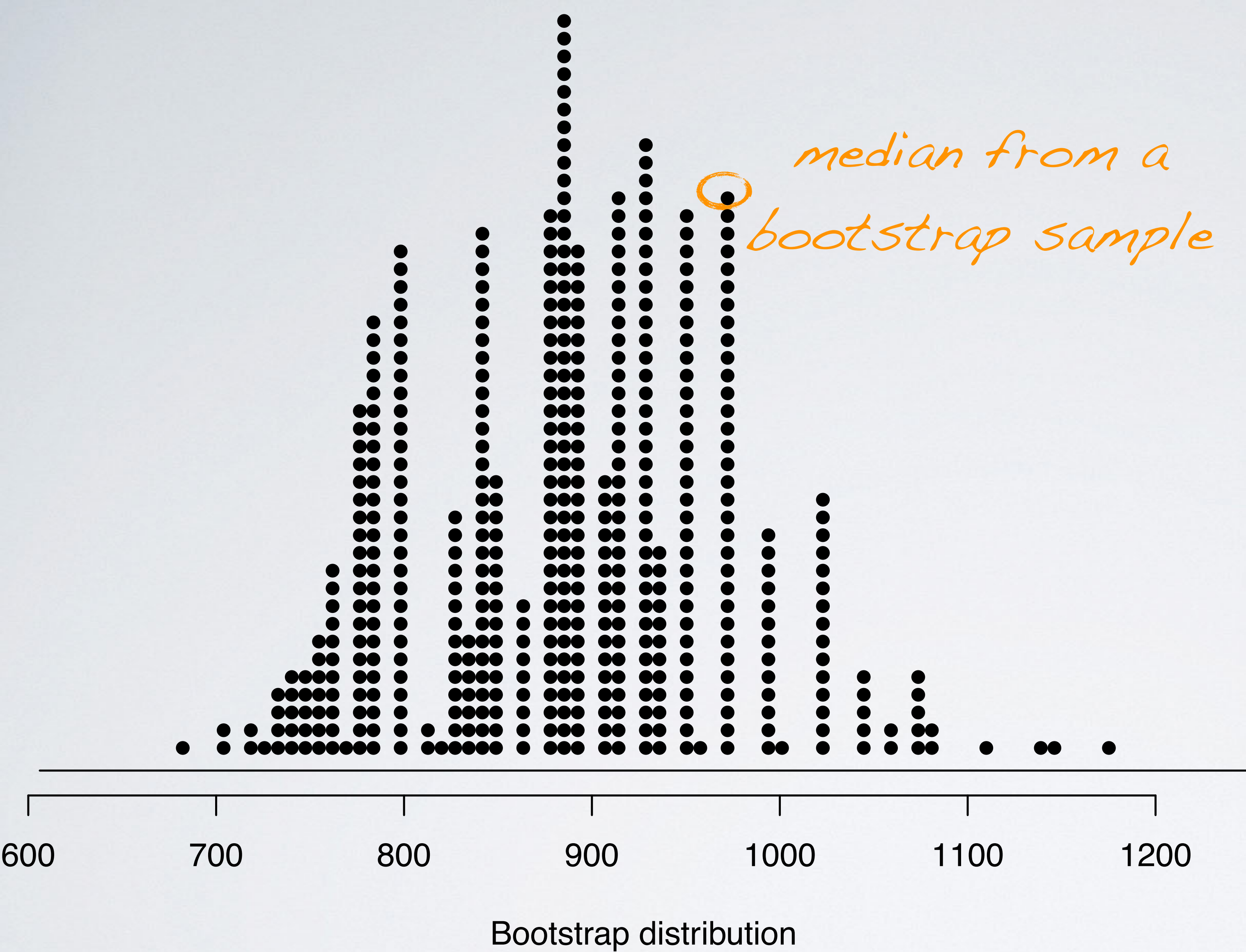
median = \$887

original sample

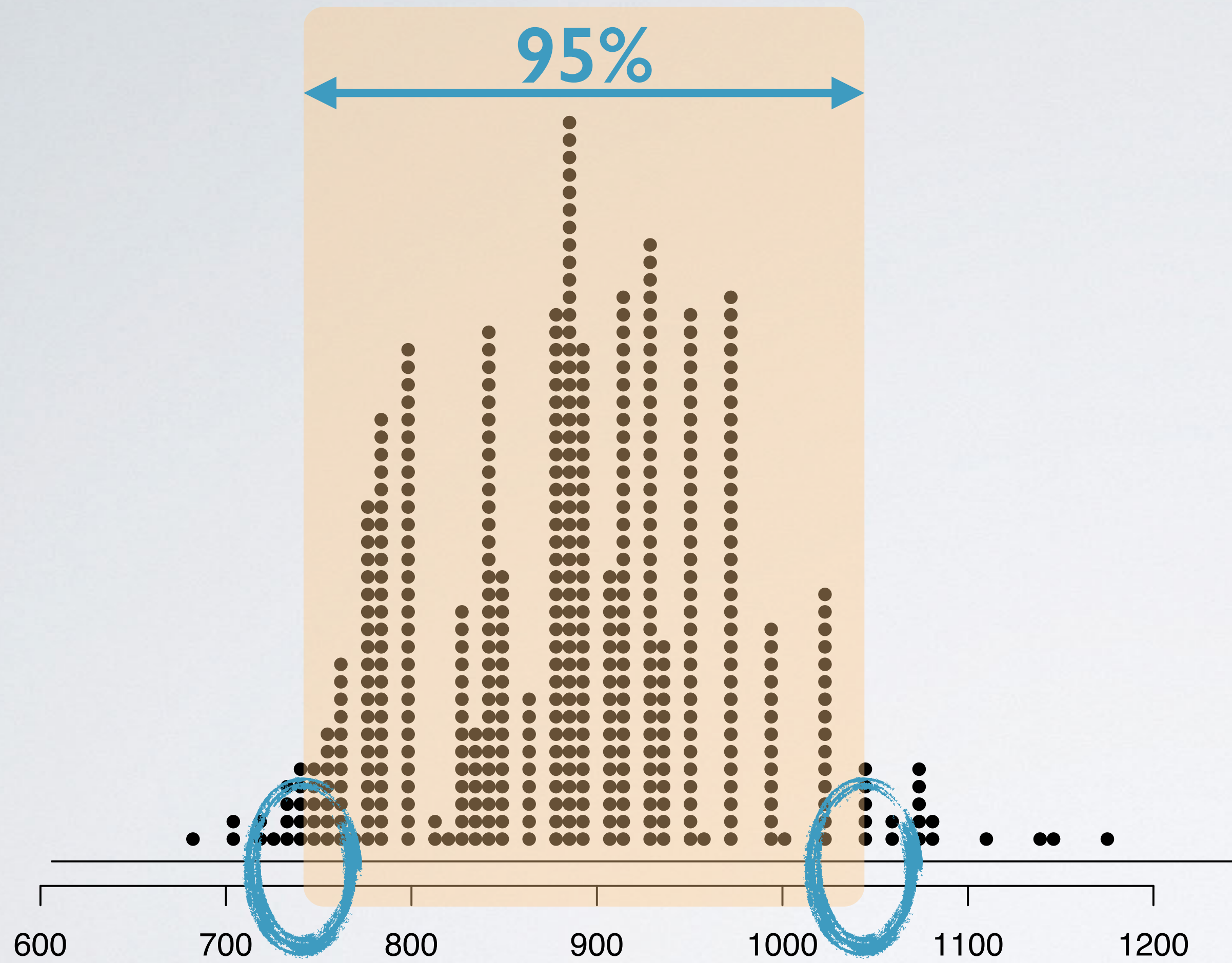


bootstrapping scheme

- (1) take a bootstrap sample - a random sample taken **with replacement** from the original sample, of the same size as the original sample
- (2) calculate the bootstrap statistic - a statistic such as mean, median, proportion, etc. computed on the bootstrap samples
- (3) repeat steps (1) and (2) many times to create a bootstrap distribution - a distribution of bootstrap statistics

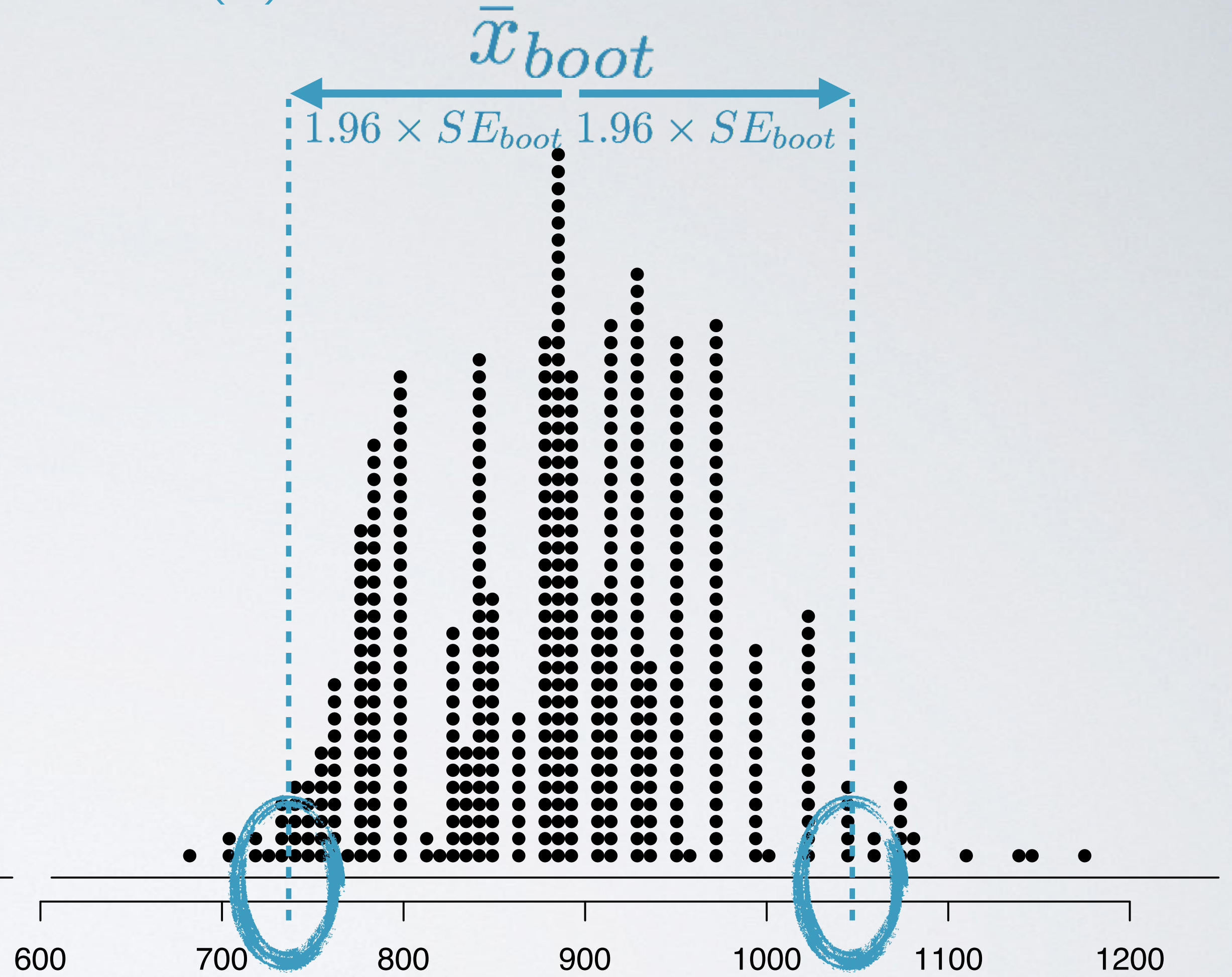


(1) percentile method



Bootstrap distribution

(2) standard error method



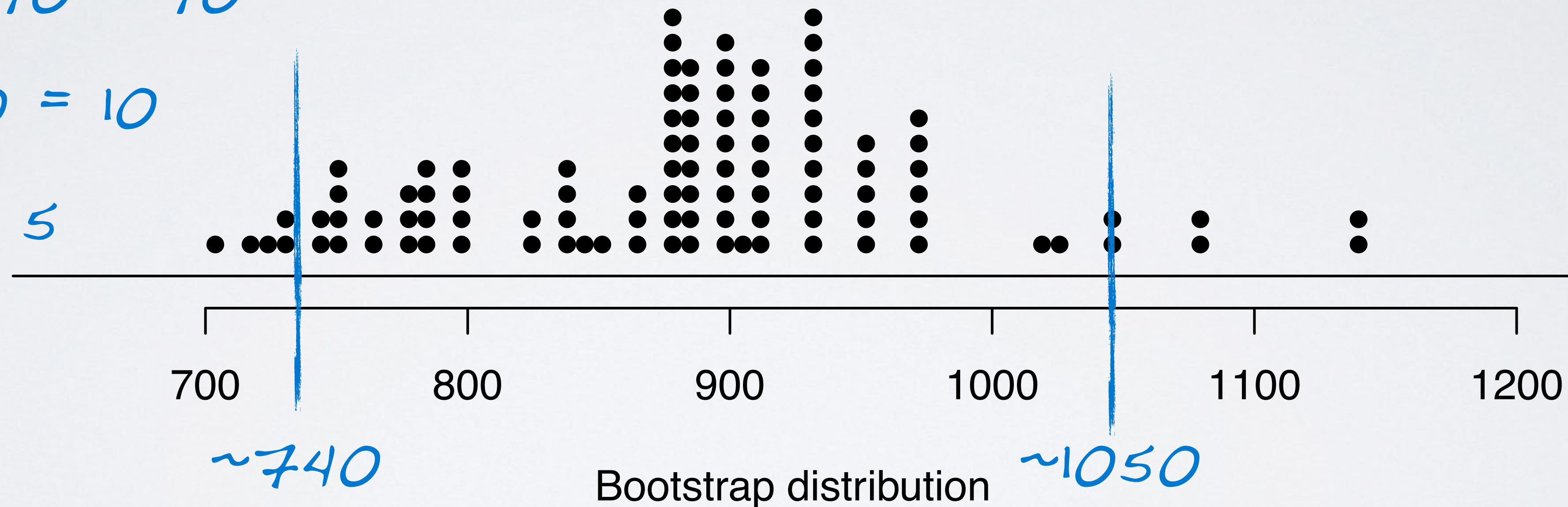
Bootstrap distribution

The dot plot below shows the distribution of medians of 100 bootstrap samples from the original sample. Estimate the 90% bootstrap confidence interval for the median rent based on this bootstrap distribution using the percentile method.

$$100 \times 0.90 = 90$$

$$100 - 90 = 10$$

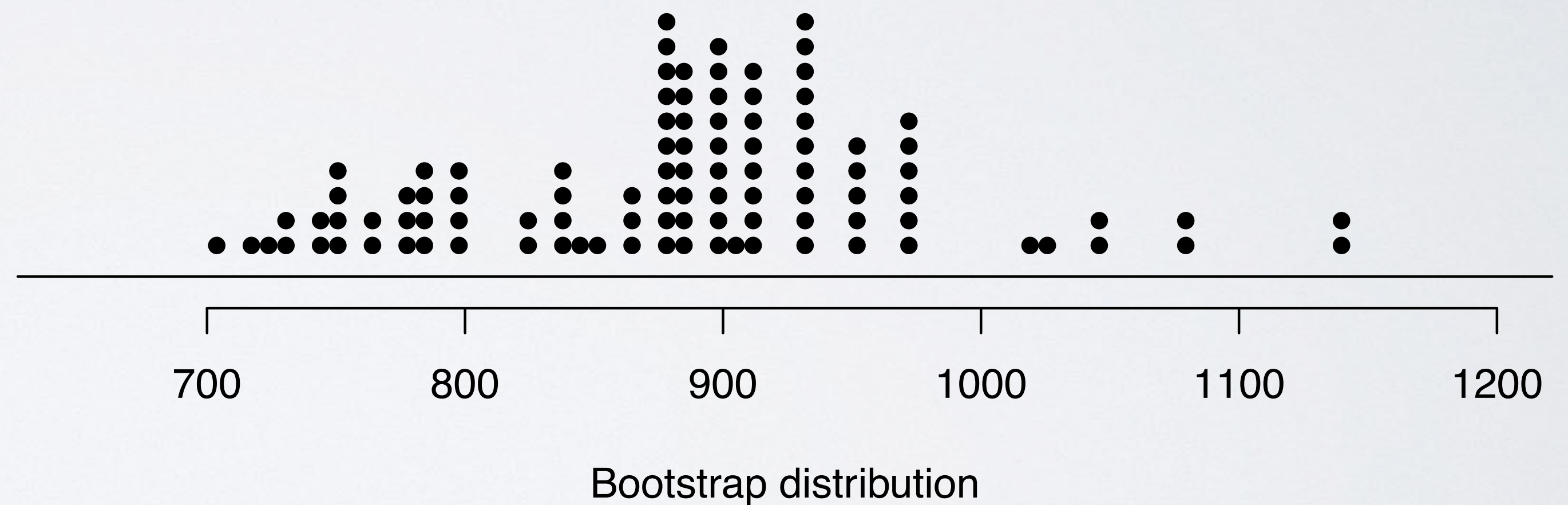
$$10 / 2 = 5$$



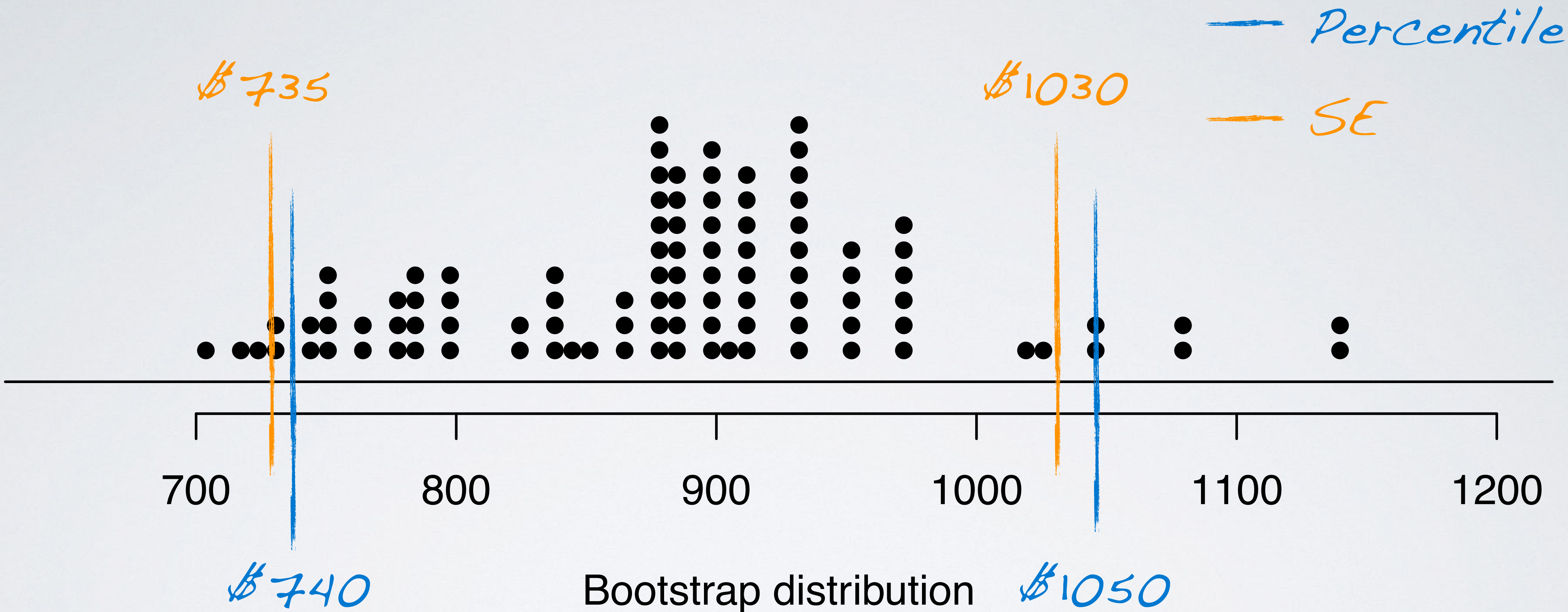
The dot plot below shows the distribution of medians of 100 bootstrap samples from the original sample. Estimate the 90% bootstrap confidence interval for the median rent based on this bootstrap distribution using the standard error method.

$$\begin{aligned} \bar{x}_{boot} \pm z^* SE_{boot} &= \\ &= 882.515 \pm 1.65 \times 89.5758 \\ &\approx (734.7, 1030.3) \end{aligned}$$

Boot. mean = 882.515
Boot. SE = 89.5758



comparison: percentile vs. SE methods



bootstrapping limitations

- ▶ Not as rigid conditions as CLT based methods.
- ▶ However if the bootstrap distribution is extremely skewed or sparse, the bootstrap interval might be unreliable.
- ▶ A representative sample is required for generalizability. If the sample is biased, the estimates resulting from this sample will also be biased.

bootstrap vs. sampling distribution

- ▶ Sampling distribution created using sampling (with replacement) from the population.
- ▶ Bootstrap distribution created using sampling (with replacement) from the sample.
- ▶ Both are distributions of sample statistics.