

Bootstrap Regression

(A) RESULT TABLES

Residual Bootstrap

	Lower <dbl>	Median <dbl>	Upper <dbl>
Screen Size 75	256.43	328.10	403.76
Screen Size 85	469.56	547.10	641.32
4K Resolution	609.86	688.72	783.62
Sony Brand	206.99	263.65	327.52

Data Bootstrap

	Lower <dbl>	Median <dbl>	Upper <dbl>
Screen Size 75	264.50	329.44	408.72
Screen Size 85	466.58	548.26	641.12
4K Resolution	606.77	690.93	790.22
Sony Brand	205.70	264.63	326.90

Monte Carlo

	Lower <dbl>	Median <dbl>	Upper <dbl>
Screen Size 75	260.01	328.92	406.40
Screen Size 85	465.54	548.12	643.86
4K Resolution	606.65	686.62	786.81
Sony Brand	209.19	263.31	325.05

INTERPRETATION:

The Median values in all the three approaches look quite similar, showing consistency for each attribute.
For Confidence Intervals:

- Residual Bootstrap: The intervals are narrower as compared to others, as the model and its parameters are fixed, while only the residual variation is resampled.
- Data Bootstrap: This approach shows slightly wider intervals, accounting for variability in the data itself.
- Monte Carlo: Confidence intervals for this simulation are similar to the other two methods but seem to slightly widen due to randomness.

(B) EXPLANATION

Bootstrap regression is a non - parametric method used to estimate the confidence interval of a quantity of interest, Z , which may belong to any distribution (may or may not be normal). Essentially, linear regression is fitted on the dataset, the dataset is randomly resampled with replacement many times (say 1000) to create a bootstrap dataset and a linear regression model is fitted on each dataset. Finally, confidence intervals are computed for parameters of the bootstrap distribution by sorting the 10,000 estimates and selecting the 2.5% and 97.5% percentiles.

The following are the two main approaches to bootstrapping commonly used in linear regression:

1. Residual Bootstrap
2. Data Bootstrap

Residual Bootstrapping:

- The general idea of residual bootstrapping is to calculate the residuals from the original estimates of the regression model, resample the residuals, create a new dependent variable, and then use this new dependent variable to run the regression model and get new parameter estimates.
- With these new parameter estimates, the Confidence Intervals can be computed using the quantile function in R. The independent variable (X s) stays the same, only the dependent variable (Y) is new.

Data Bootstrapping:

- This is another method of creating new datasets for bootstrap regression.
- In this approach, both the X and Y variables are resampled with replacement from the (Y, X) matrix, then regression is run for the variables, and the new parameter estimates are utilized to compute Confidence Intervals. The independent and dependent variables are new.

The confidence intervals for residual bootstrapping is mostly seen to be narrower than data bootstrapping as the former has only dependent variables as new, whereas the latter has both variable samples to be new. The Residual bootstrapping method can be more model-specific and data bootstrapping can be more general specifically when there is uncertainty about the error terms.