SACRAMENTO STATE

A Comparative Simulation Study on Various Equality of Variance Tests

Feini Pek, Lauren Cappiello
Department of Mathematics and Statistics

Why do we care?

- Variance is a measure of how far data points are from the mean
- Statistics is an essential component of every study
- Numerous statistical testing assume homogeneity of variance
- Experimental data is messy
- Goal : With several choices of equality of variance testing, we want to identify the best test to use given different types of data

Outline

1. Test For Equality of Group Variances
2. Brown Forsythe vs Levene’s Test
3. Bootstrap Testing
4. Methodology
5. Results
6. Conclusion & Discussion
7. Future Directions

Test for Equality of Group Variances

What does the test do?

Identify whether the variances between 2 groups are different

How does the test work?

Null Hypothesis (H_0) : The group variances are equal

Alternative Hypothesis (H_A) : The group variances are not equal

Returns a p-value

Significance level (α) = 0.05

p-value < 0.05 → Reject H_0

p-value > 0.05 → Do not reject H_0

Levene's Test

- Uses deviations from group means
- Gives the best power for symmetric, moderate tailed distributions

Brown-Forsythe

- Uses deviations from group medians
- More robust to skewness and unequal sample sizes

Bootstrap Testing

- **What is Bootstrap Testing?**

A form of hypothesis testing that involves resampling a single data set to create a multitude of simulated samples

- **Advantages of bootstrap Testing**

This approach doesn't assume any underlying distribution of the data since the sampling distribution can be observed.

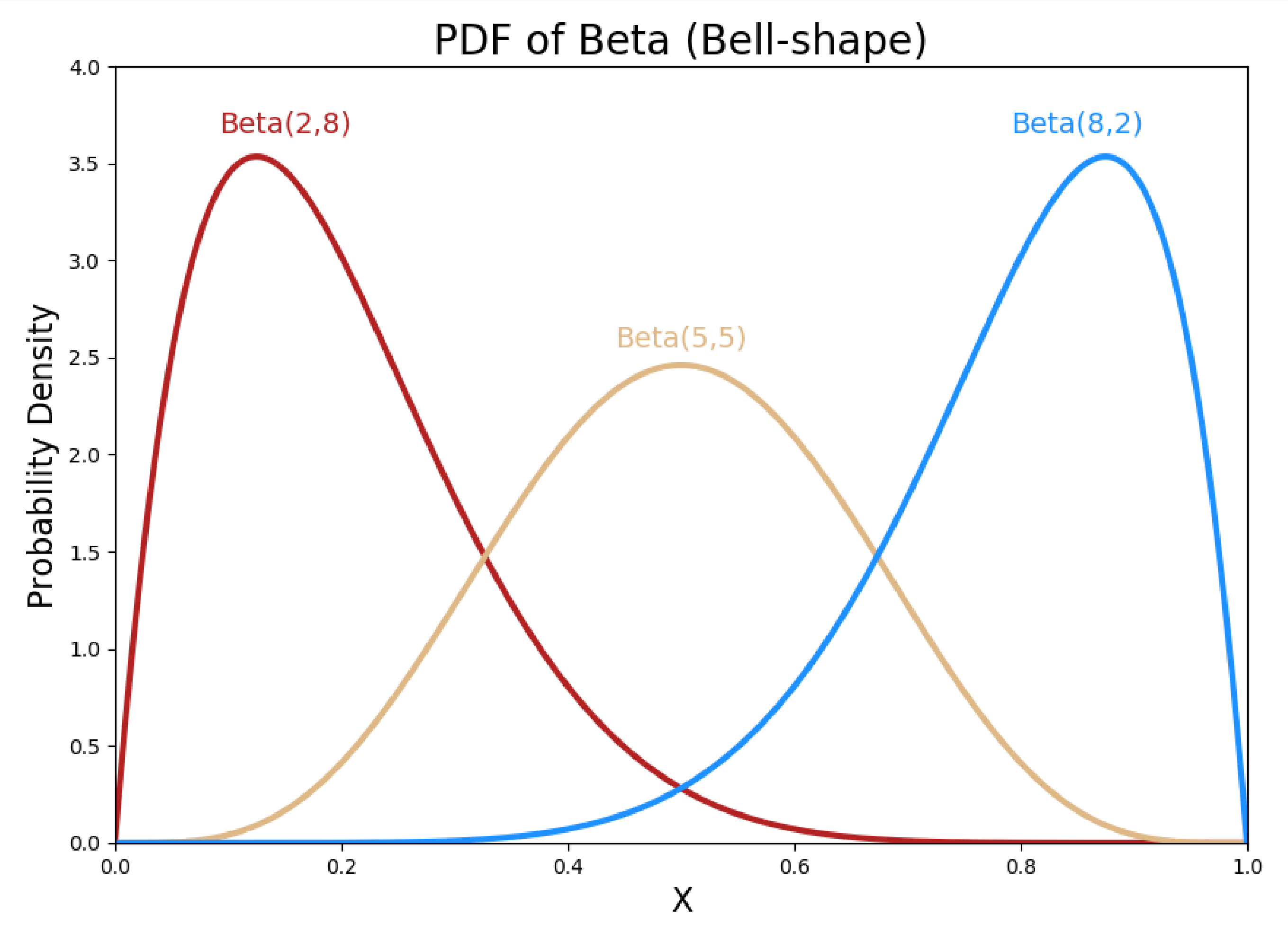
- **In our case:**

We resample our median deviances and construct a sampling distribution from which we will infer a p-value

Methodology

Conduct a simulation study in R

1.Generate data from beta distribution



Source : https://vitalflux.com/beta-distribution-explained-with-python-examples/

Data used in this study:

- Beta distribution of skew 0 – 3
- 5 Variance ratio (1:1 to 1:5)
- Sample size of 50 and 100

We are using:

4 x 5 x 2 = 40 different settings

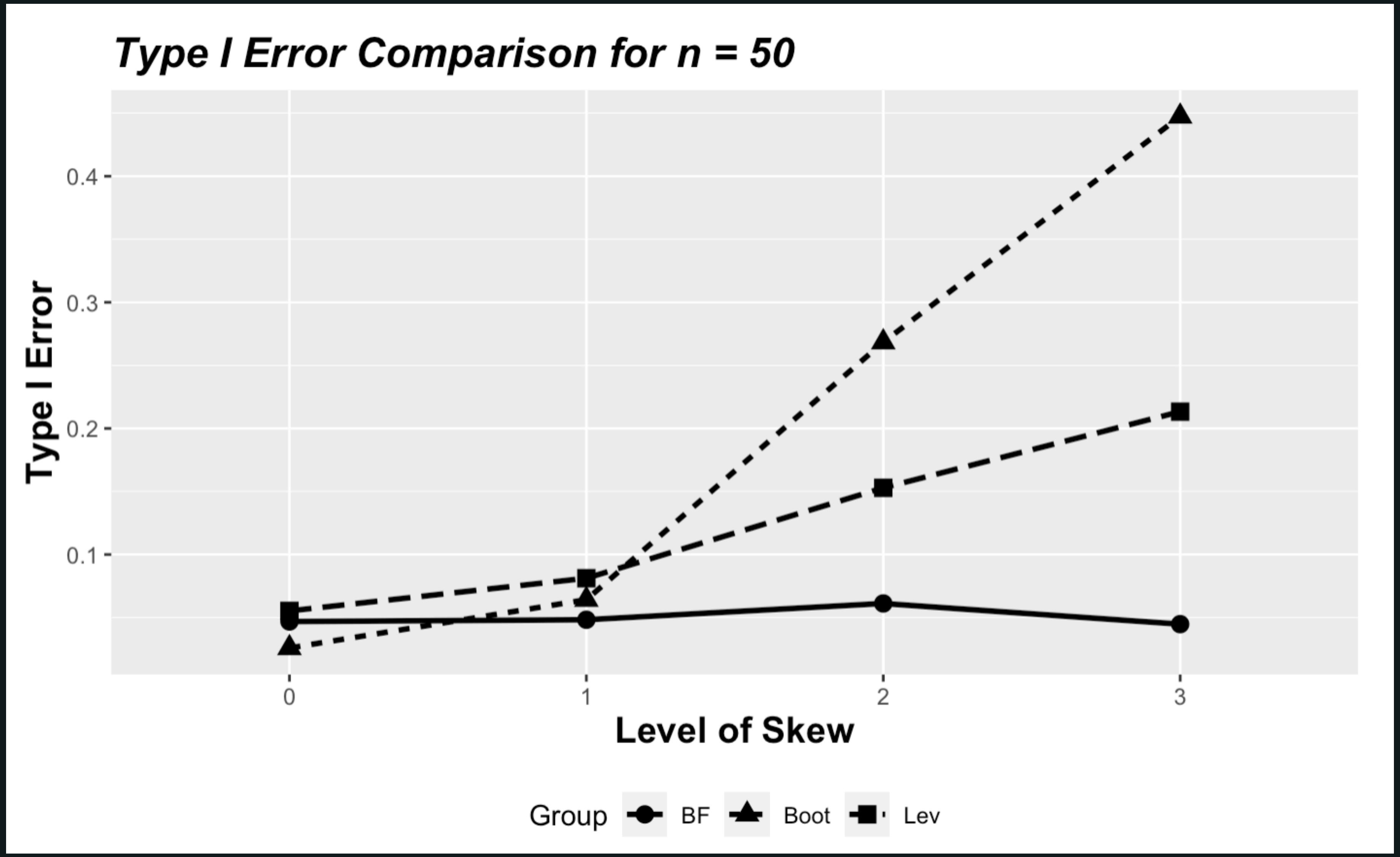
Methodology

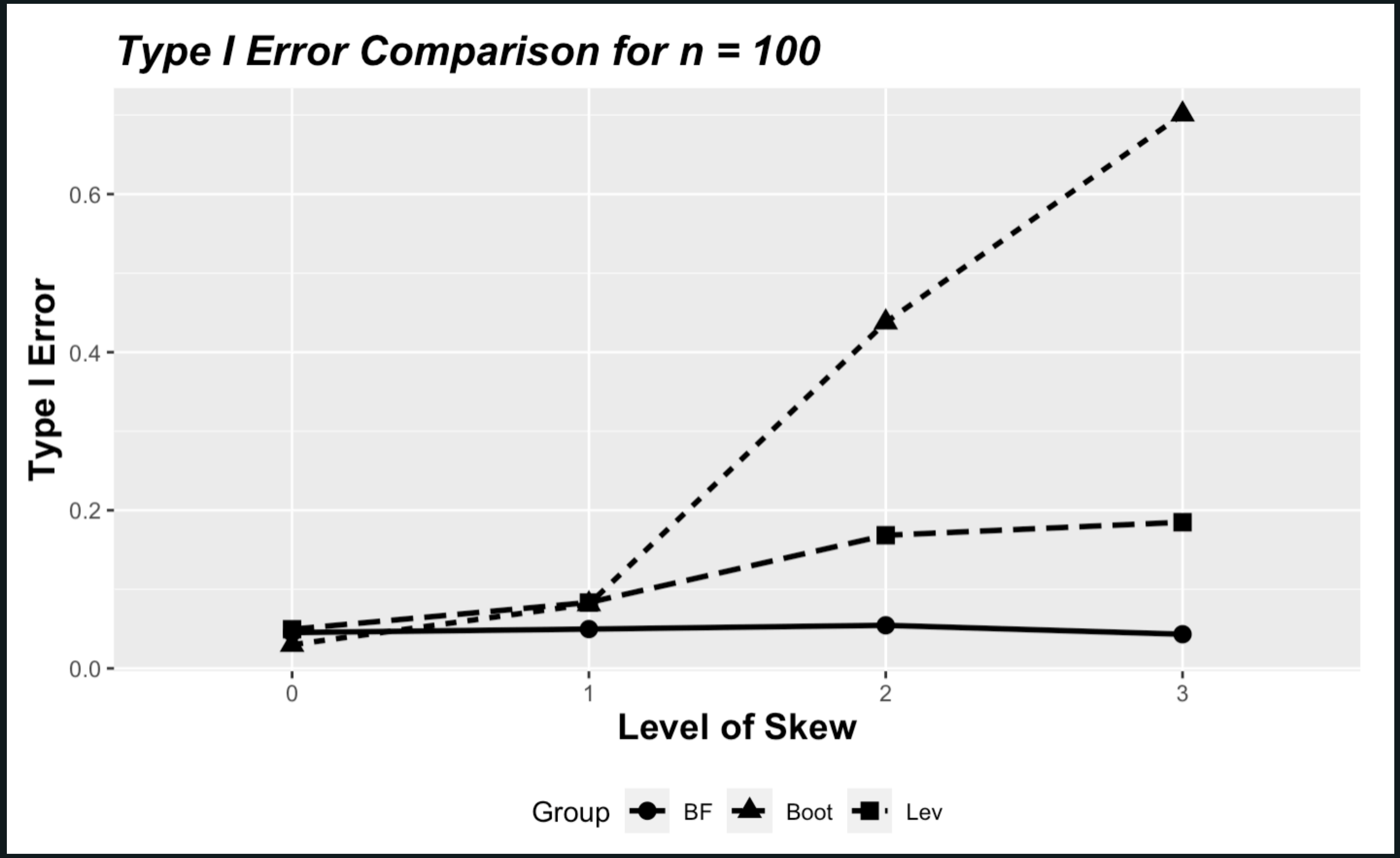
2. Perform all 3 tests on the data
3. Collect p values
4. Repeat 10,000 times
5. Test for type 1 and type 2 errors
   - Type I error = falsely reject null hypothesis
   - Type II error = fail to reject null hypothesis

We test type I error for data with Variance ratio 1:1

We test type II error for data with Variance ratio 1:2 – 1:5

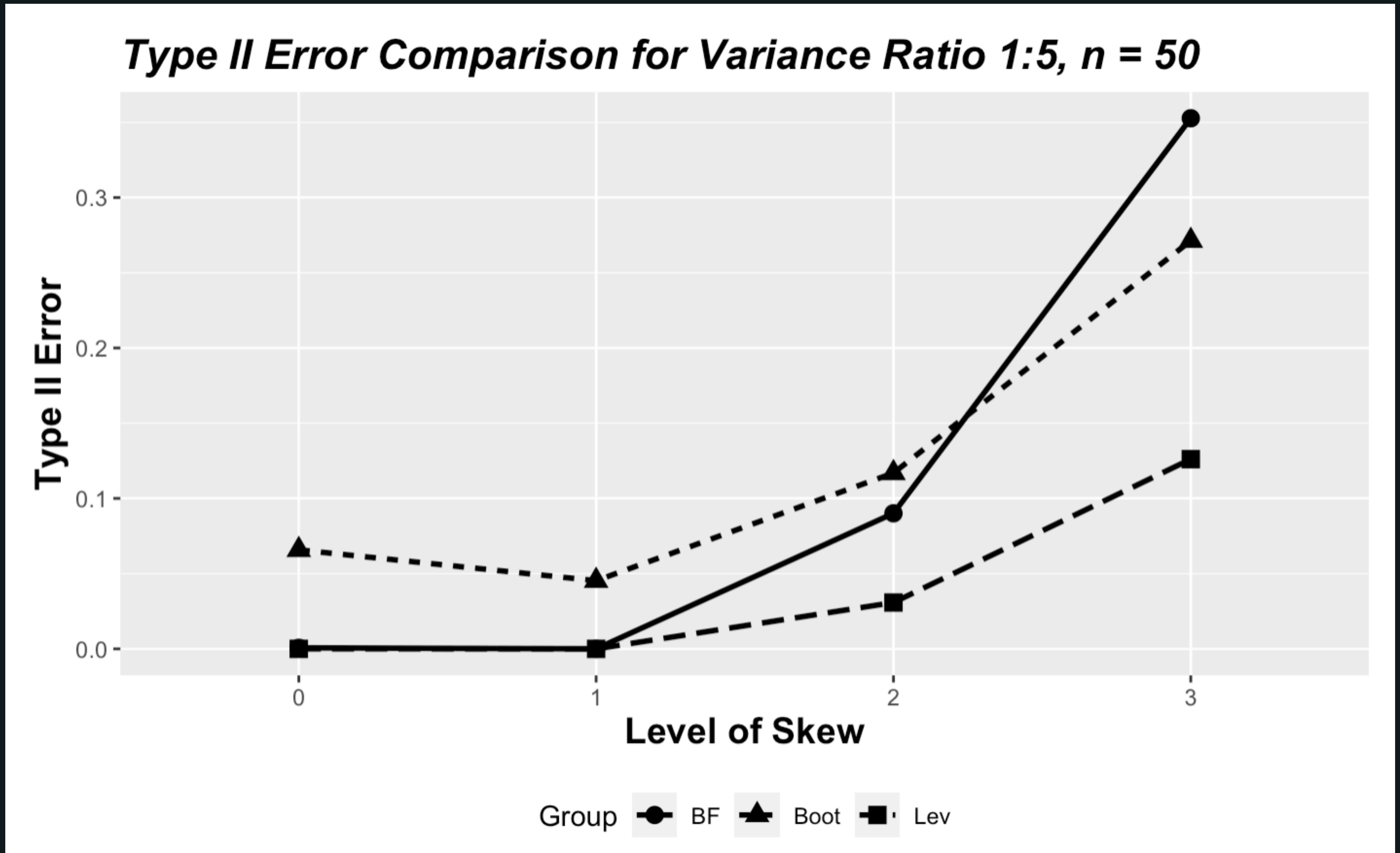
Results

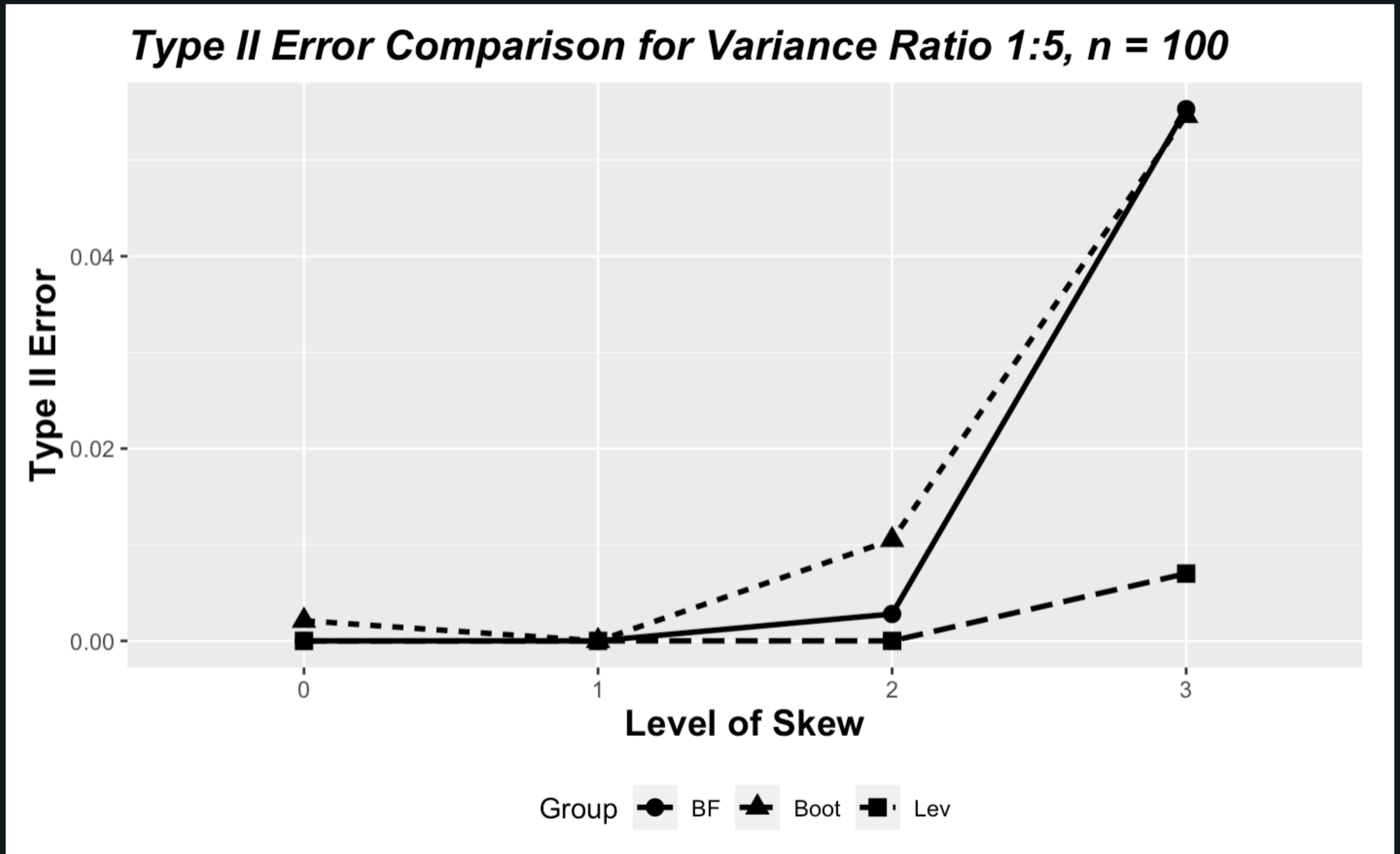




- Brown Forsythe Test returns the lowest Type I Error level, of average 0.049 for all level of skewness and both n = 50 and n = 100
- All the test performed similarly for skew level 0,1, but error level shoots up for both Levene’s and bootstrap when skew ≥ 2.

Results





- All the test return a low level of Type II error even when the data is skewed
- Since the error level < 0.35 when n = 50, and error level < 0.06 when n = 100, the test are sensitive to sample size
- As the variance ratio increases, type II error decreases

Conclusion & Discussion

- BF outperforms the other tests across all level of skews, for groups with equal sample sizes
- Unexpected type I error for bootstrap testing
- If type I error is larger, then it is expected for type 2 to be smaller

Future Directions

- Improve bootstrap testing performance
- Evaluate the tests' performances on datasets with different sample sizes

THANK YOU!