

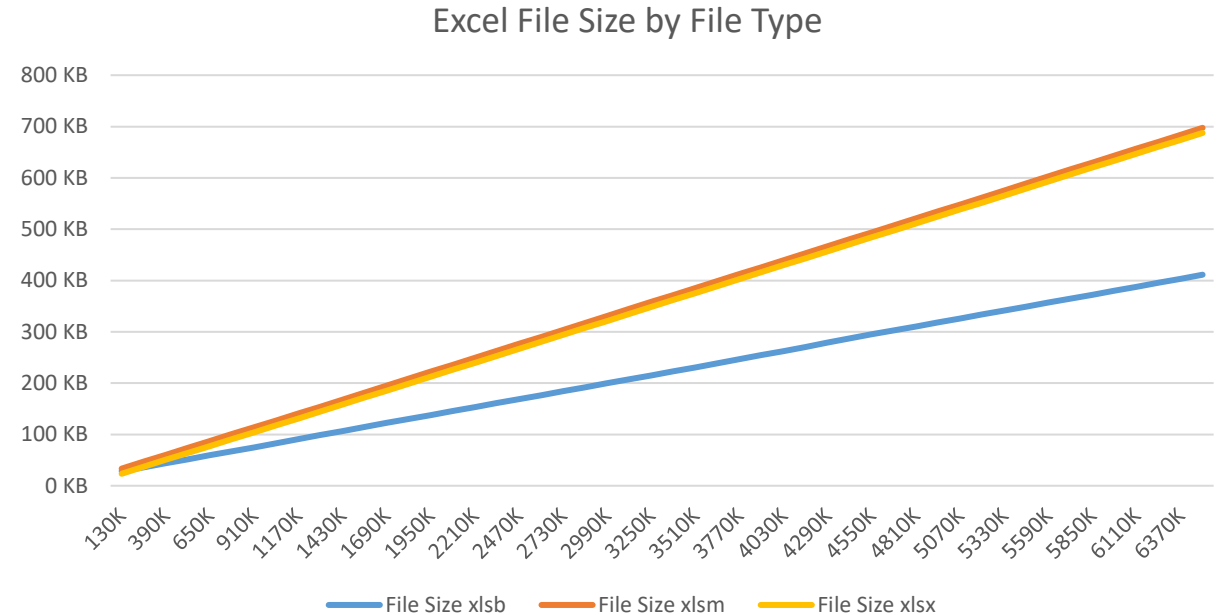
2007 Excel File Type Comparison

Analysis Overview

- Purpose of Analysis: Understand the file size and performance differences between the three Excel file types (XLSB, XLSM, XLSX).
- Testing Methodology
 1. File Size – Each Excel file type was incremented by the same amount of characters and populated cells then saved. The file size was measured for each trial.
 2. Performance –
 1. Save Speeds - A large file size was created for each file type (with an equal number of characters and populated cells). The files were iteratively saved and the save down time was measured and recorded.
 2. Program Use Speeds – A set of the same actions (such as copying cells, adding formatting etc.) were performed and iterated through a macro. The execution times were measured and recorded.

File Size

- Data follows a nearly perfect linear relationship with the number of characters in the file.
- Linear regression was used to get an estimation of the slope. R-squared near 1 and P-Value near 0 due to nearly perfect linear relationship.



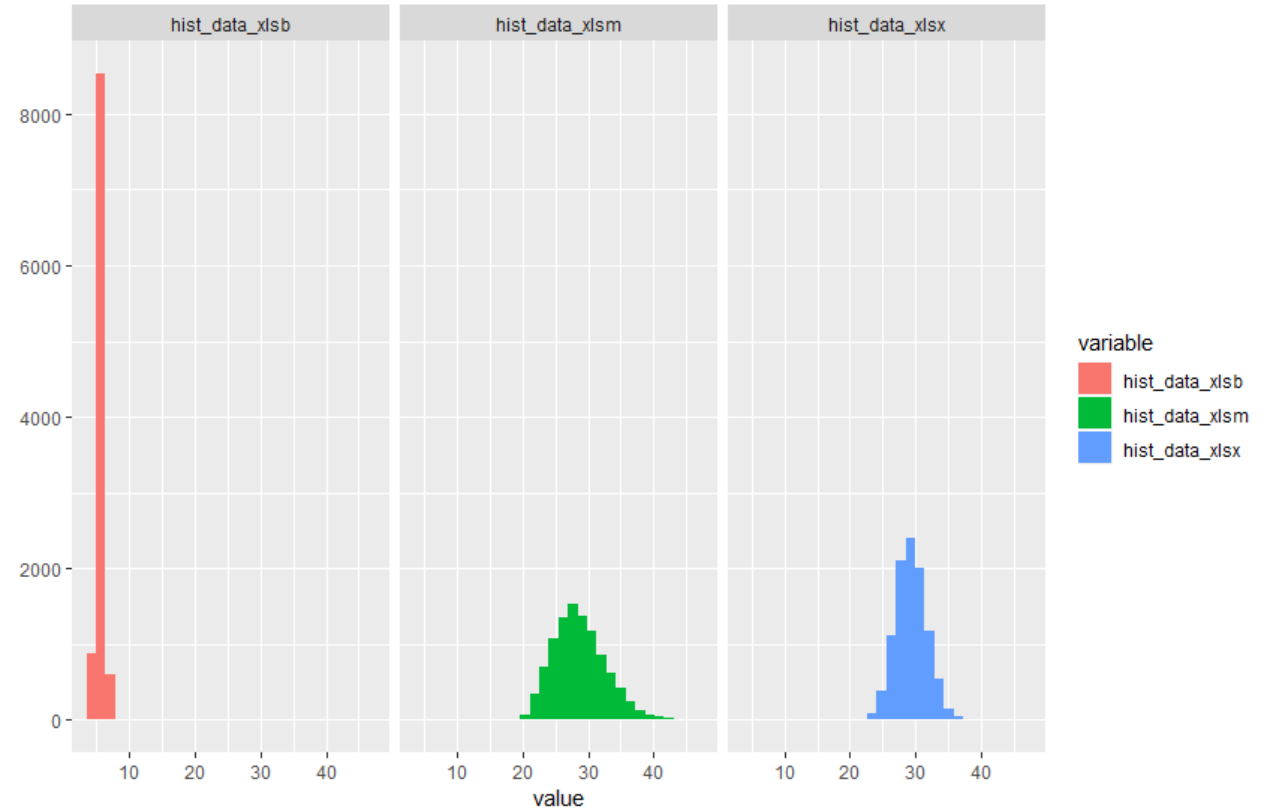
Regression Coefficient Output

File Type	Coefficient
XLSB	.00602
XLSM	.0104
XLSX	.0104

Units = bytes per character

Performance – Save Speed

- A large file was created with the same number of characters and populated cells for each type.
- Only 14 trials were ran per file type due to CPU and time cost.
- Data was bootstrapped with 10,000 iterations to get a distribution for each file type.



Summary Statistics

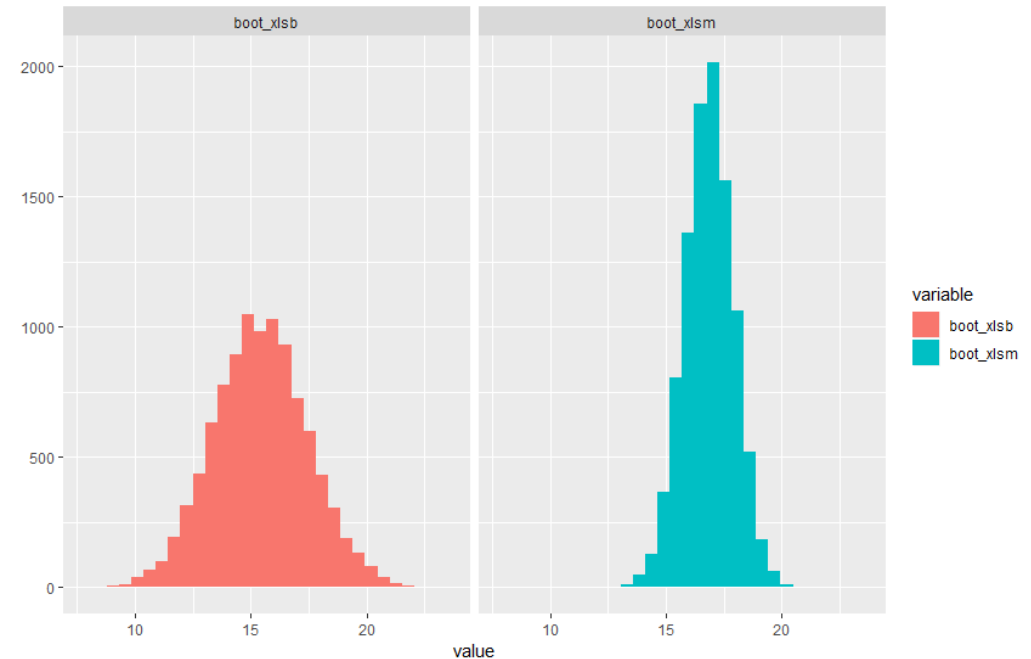
File Type	Median	Mean	Standard Deviation
XLSB	5.73	5.76	0.49
XLSM	28.39	28.78	3.96
XLSX	29.23	29.32	2.39

Units = Seconds

Note: Actual save down times are dependent on multiple variables specific to the machine used. These statistics are used for comparison of the types, not to get a generally applicable estimate of save down times.

Performance – Save Speed Cont.

- Since files with the same number of characters are much smaller for XLSB, the performance of save down time for the same memory sized files was tested.
- Only XLSM and XLSB were tested to save resources.
- XLSM and XLSB have much closer save down times when the files take up the same memory. XLSM has a lower standard deviation.



File Type	Median	Mean	Standard Deviation
XLSB	15.42	15.43	2.03
XLSM	16.90	16.85	1.06

Note: The test was performed on approximately 6.5 MB sized files.

Performance – Use Speed

- For this test, multiple actions were performed in excel iteratively and time to execute was recorded.
- Since this code ran much faster, sufficient trials were performed for a T-Test.
- The T-Test Results to the right show that there is no significant difference in use speed between the three file types. This makes sense because the file type is how the file is saved in long-term memory. What typically determines use speed is RAM, which I believe is utilized the same for the three data types.

T-Test	95% CI--Lower	95% CI--Upper
XLSB vs XLSM	-0.14	0.16
XLSB vs XLSX	-0.15	0.22
XLSX vs XLSM	-0.16	0.11

File Type	Median	Mean	Standard Deviation
XLSB	0.57	0.70	0.49
XLSM	0.65	0.71	0.15
XLSX	0.59	0.73	0.47

Conclusion

- XLSM and XLSX have similar performance and file sizes. XLSX showed less variance in save times but has similar measures of central tendency (mean and median) to XLSM.
- XLSB has much better save time performance and file sizes. XLSB files grow at about 60% the rate of XLSM and XLSX. The average save down size is about 20% of that of XLSM and XLSX. The save down benefit is created by the decreased file size, not a more efficient save down process.
- Due to its superior performance and size, XLSB should be used for all large files where macros are acceptable. For smaller files, XLSM or XLSX (depending on whether or not macros are acceptable in the file) should be used since they are more universally compatible.