Standard Error Distribution

Jimmy Zhang

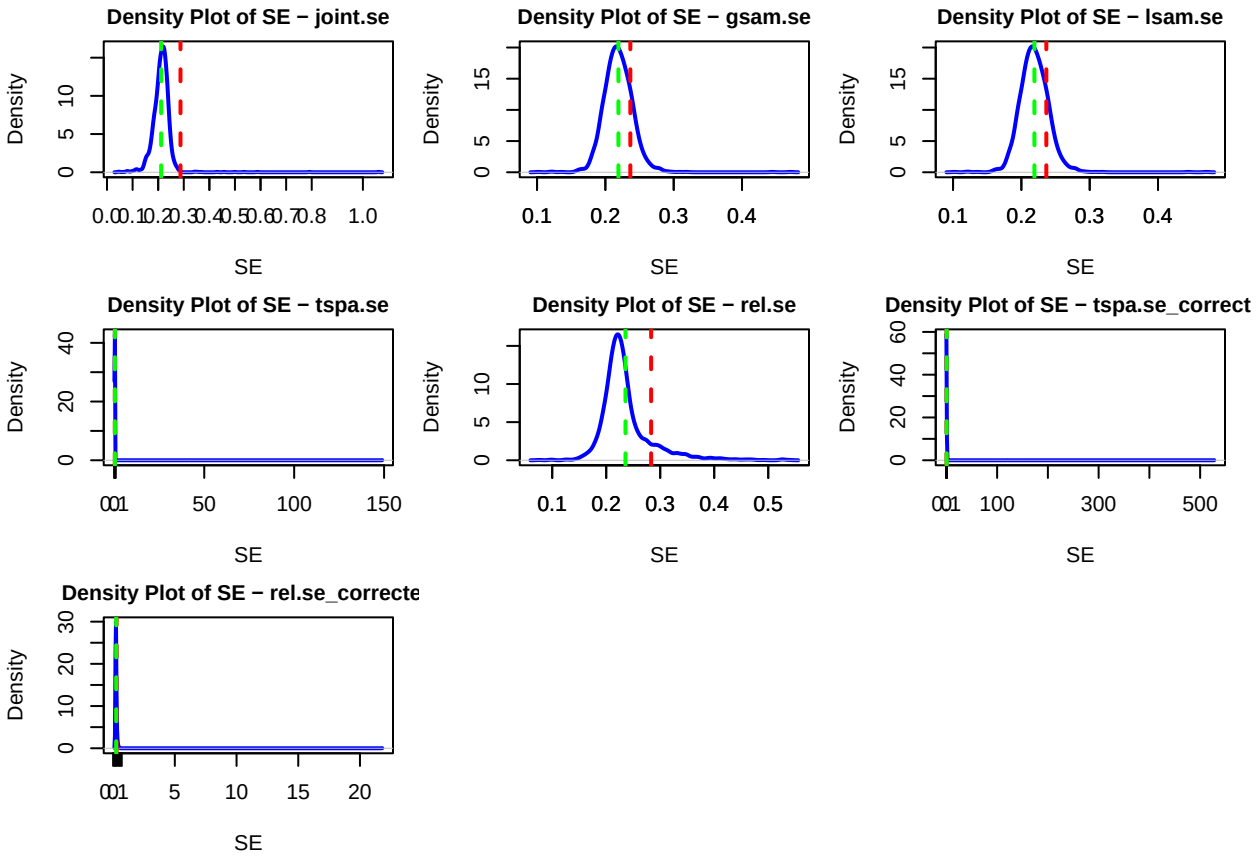
2024-11-29

```
library(here)
library(dplyr)
library(tidyr)
library(forcats)
library(readr)
library(ggplot2)
library(rlang)
```

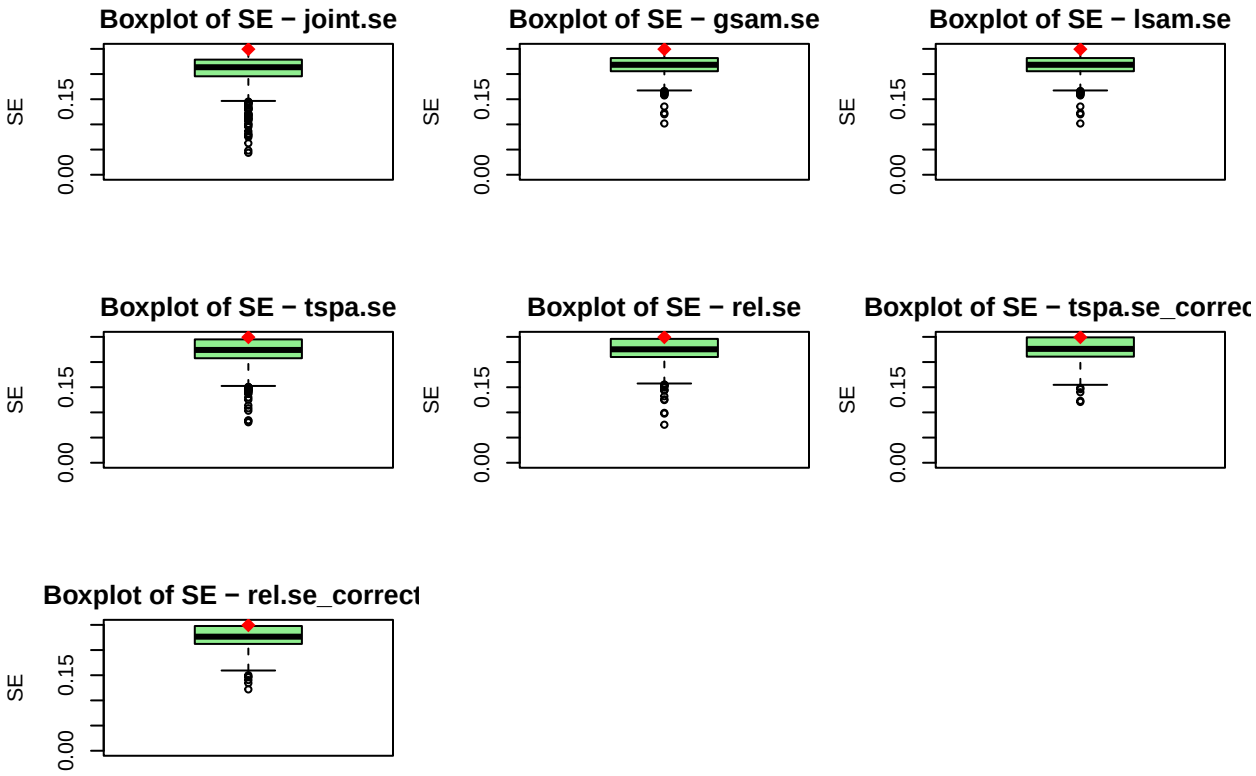
Standard error distributions are examined under conditions with small sample size (i.e., 30, 60, 120) with low reliability ($\rho = 0.7$).

Condition 1: $\beta = 0$, $N = 30$, $\rho = 0.7$

Density plot of SE distribution (Without removing outliers)



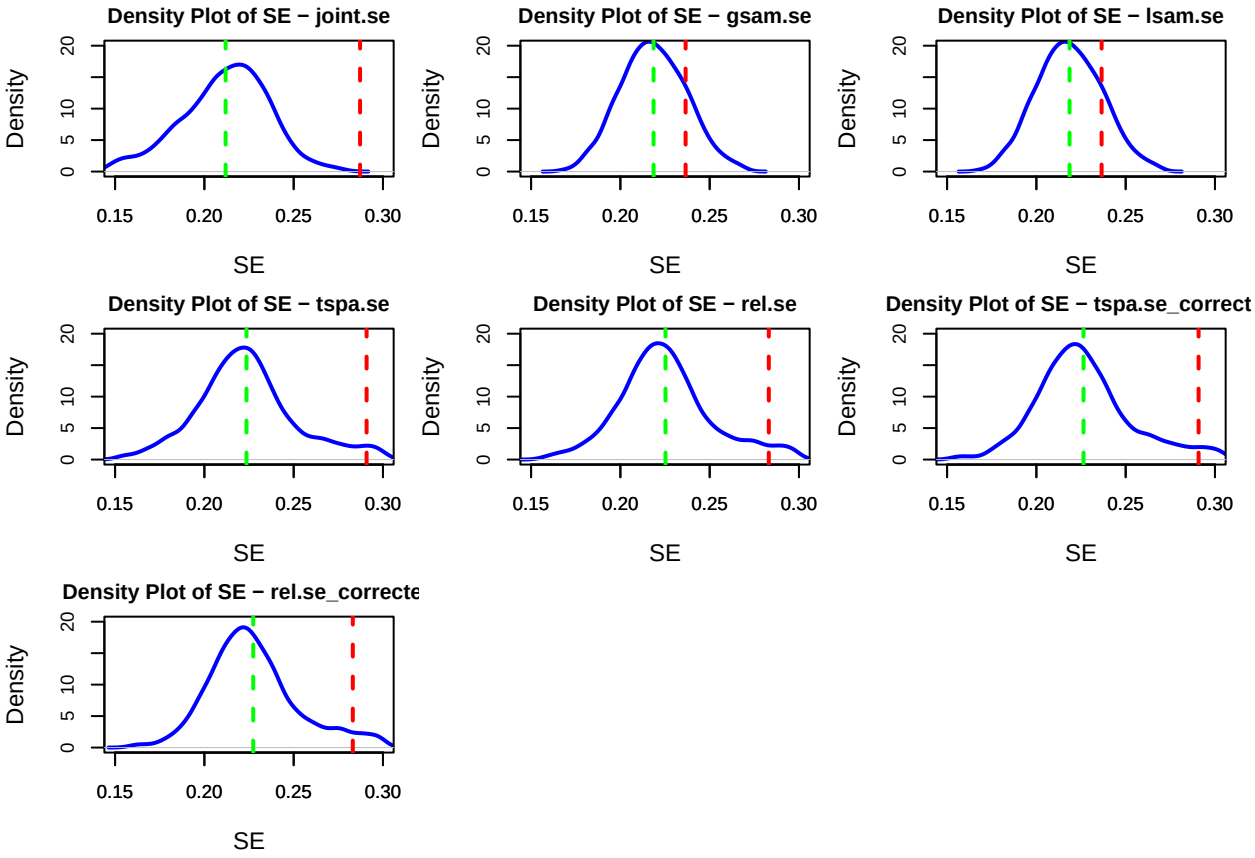
Boxplot of SE values



Outlier proportions

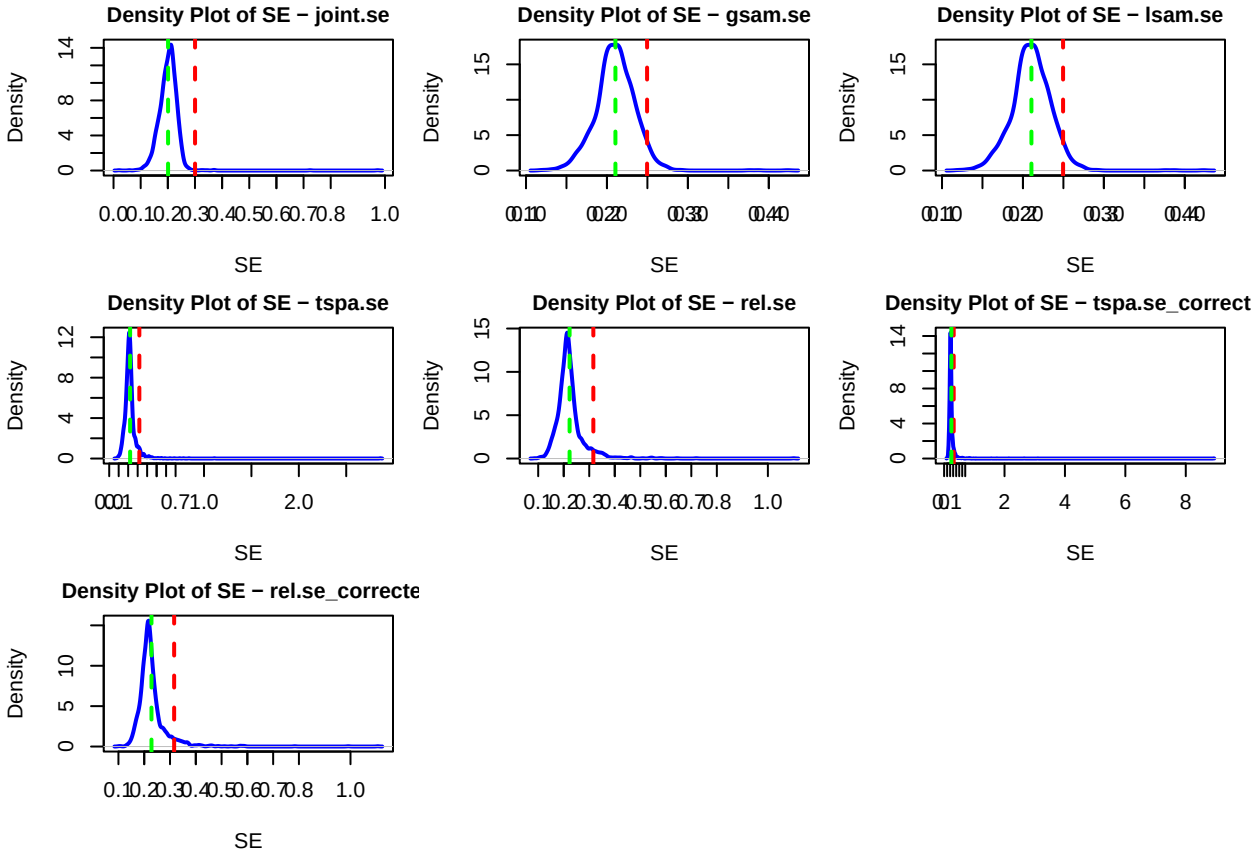
```
##                              Method Outlier_Percentage
## joint.se                   joint.se               2.80
## gsam.se                     gsam.se               1.95
## lsam.se                     lsam.se               1.95
## tspa.se                     tspa.se              10.00
## rel.se                       rel.se               9.60
## tspa.se_corrected tspa.se_corrected              10.10
## rel.se_corrected   rel.se_corrected               9.40
```

Density plot of SE distribution (Removed outliers)

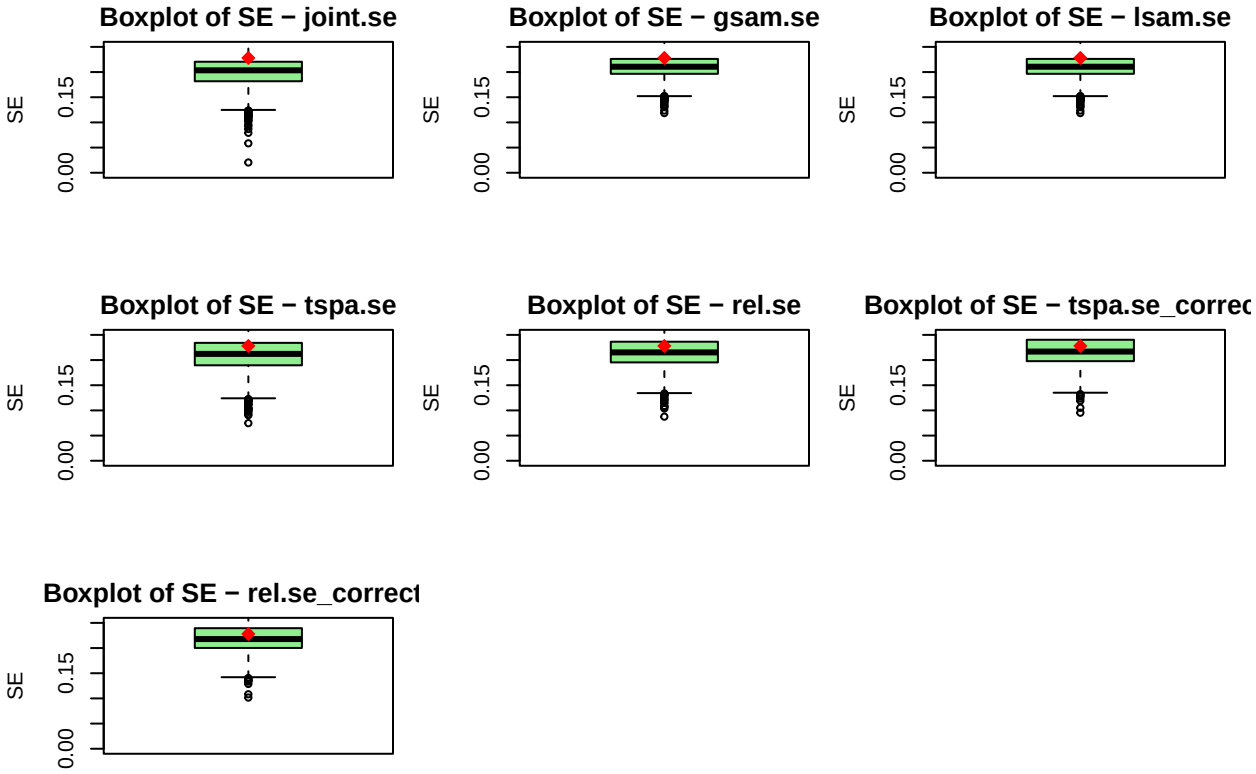


Condition 2: $\beta = 0.3$, $N = 30$, $\rho = 0.7$

Density plot of SE distribution (Without removing outliers)



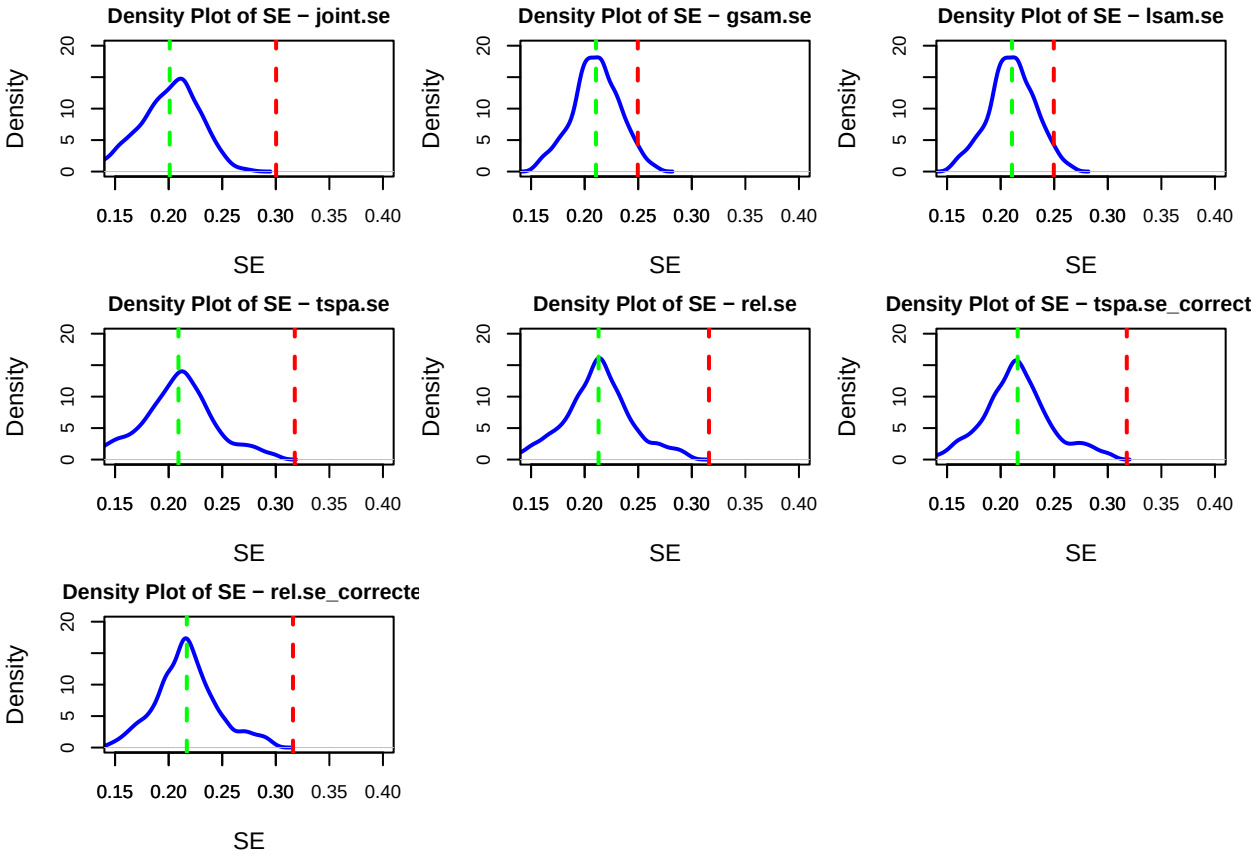
Boxplot of SE values



Outlier proportions

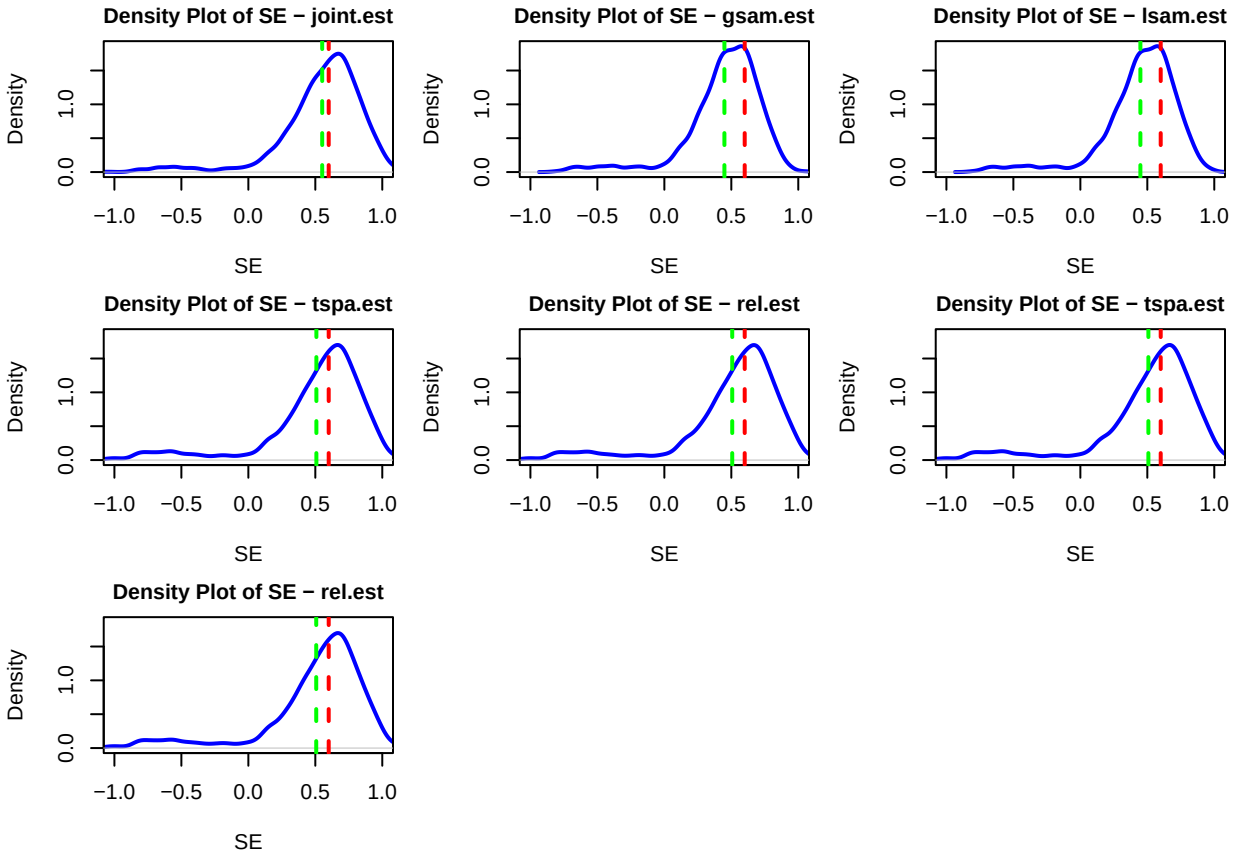
```
##                              Method Outlier_Percentage
## joint.se                   joint.se               1.95
## gsam.se                     gsam.se               1.95
## lsam.se                     lsam.se               1.95
## tspa.se                     tspa.se               8.25
## rel.se                       rel.se               8.10
## tspa.se_corrected tspa.se_corrected               8.75
## rel.se_corrected   rel.se_corrected               8.10
```

Density plot of SE distribution (Removed outliers)

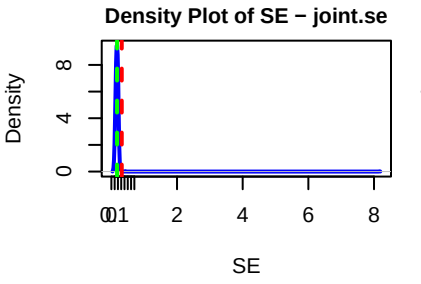


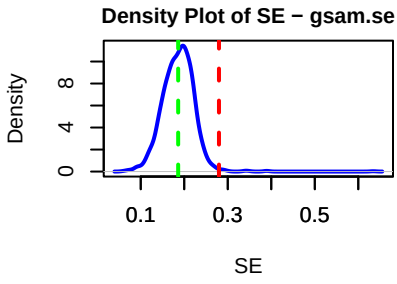
Condition 3: $\beta = 0.6$, $N = 30$, $\rho = 0.7$

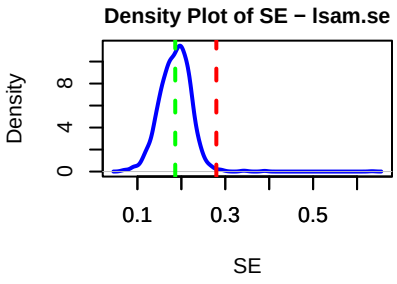
Density plot of estimates distribution

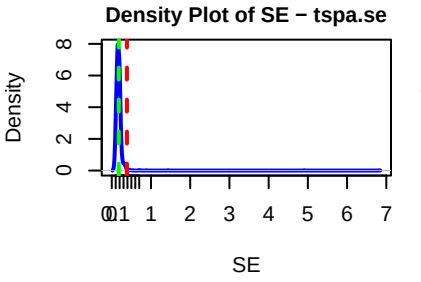


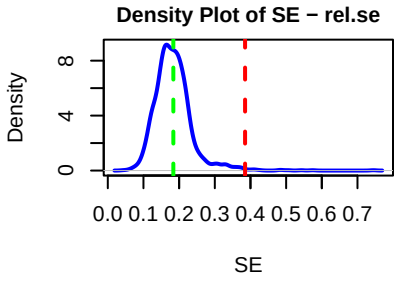
Density plot of SE distribution (Without removing outliers)

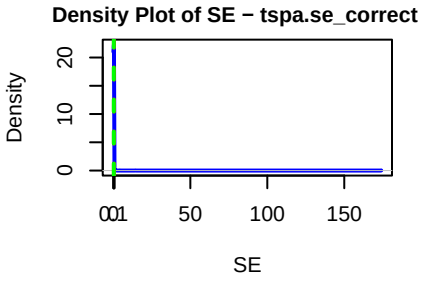


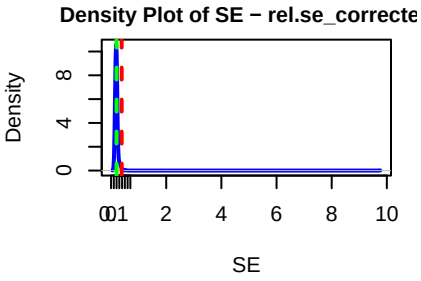




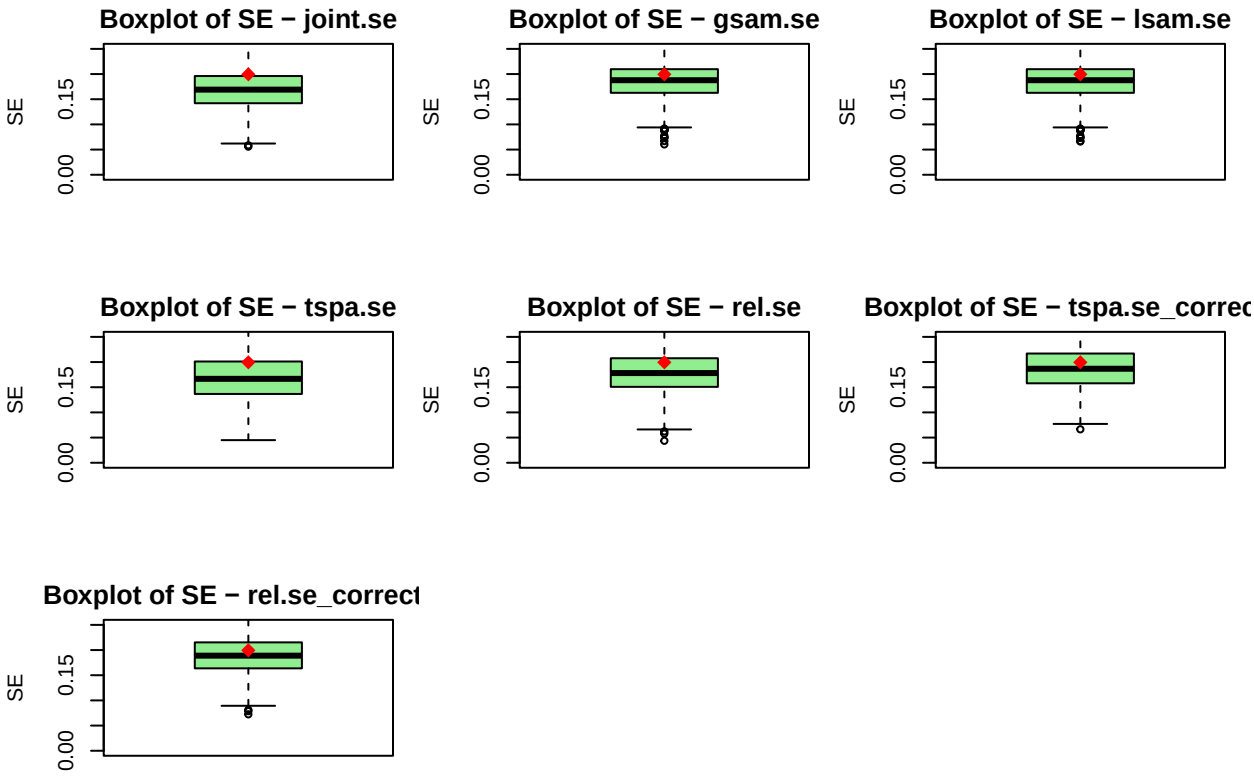








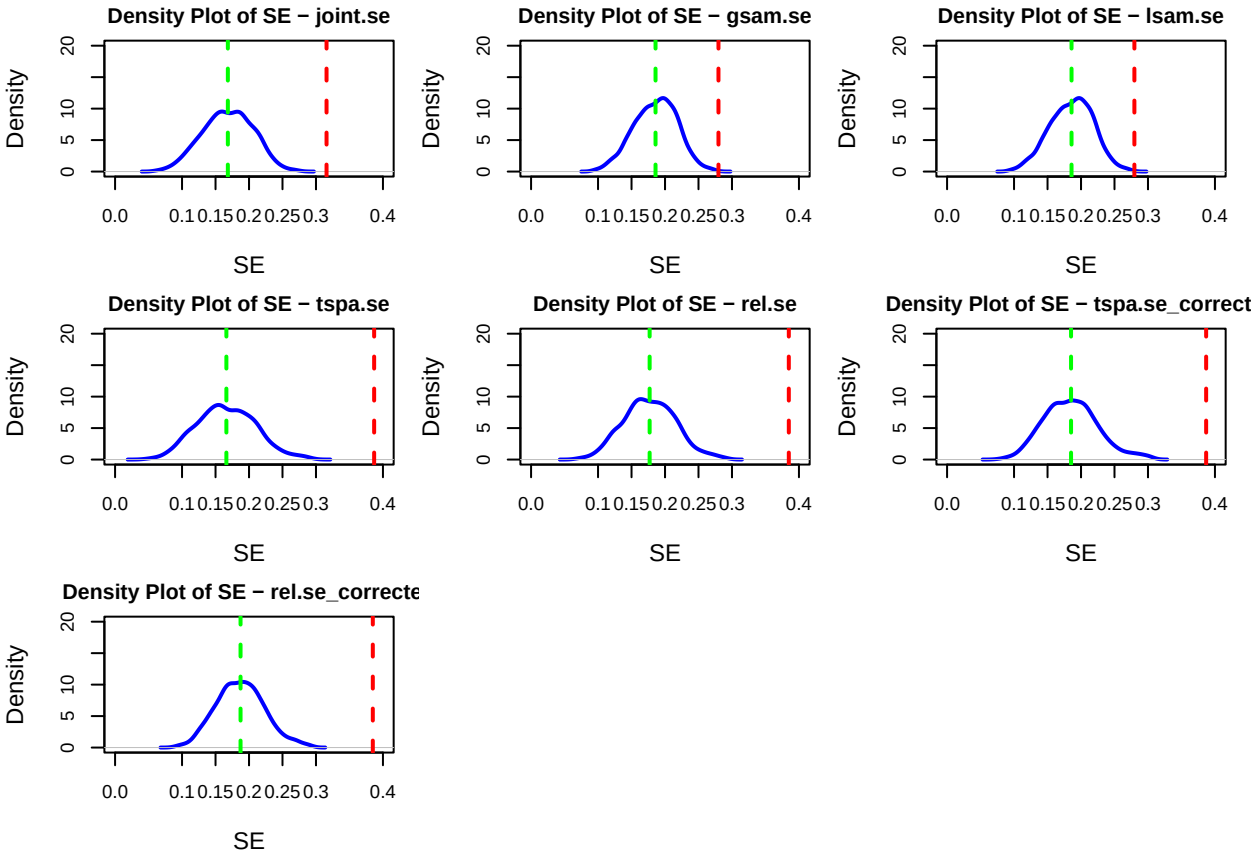
Boxplot of SE values



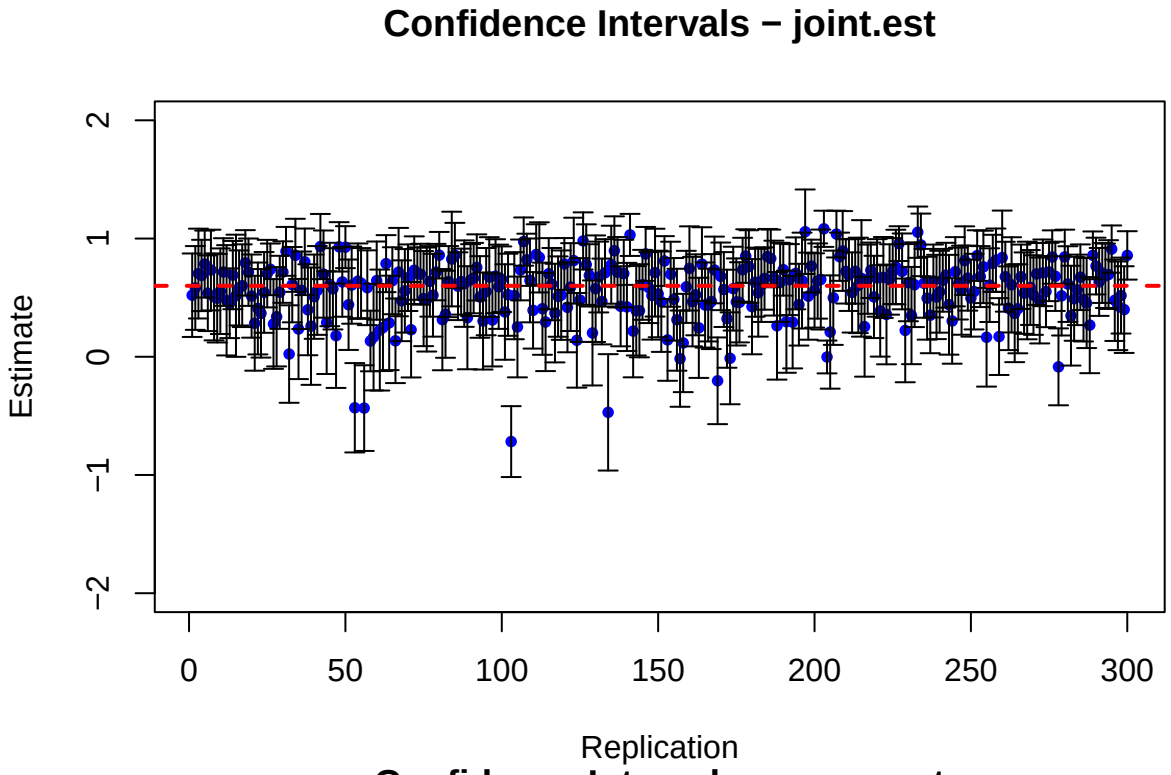
Outlier proportions

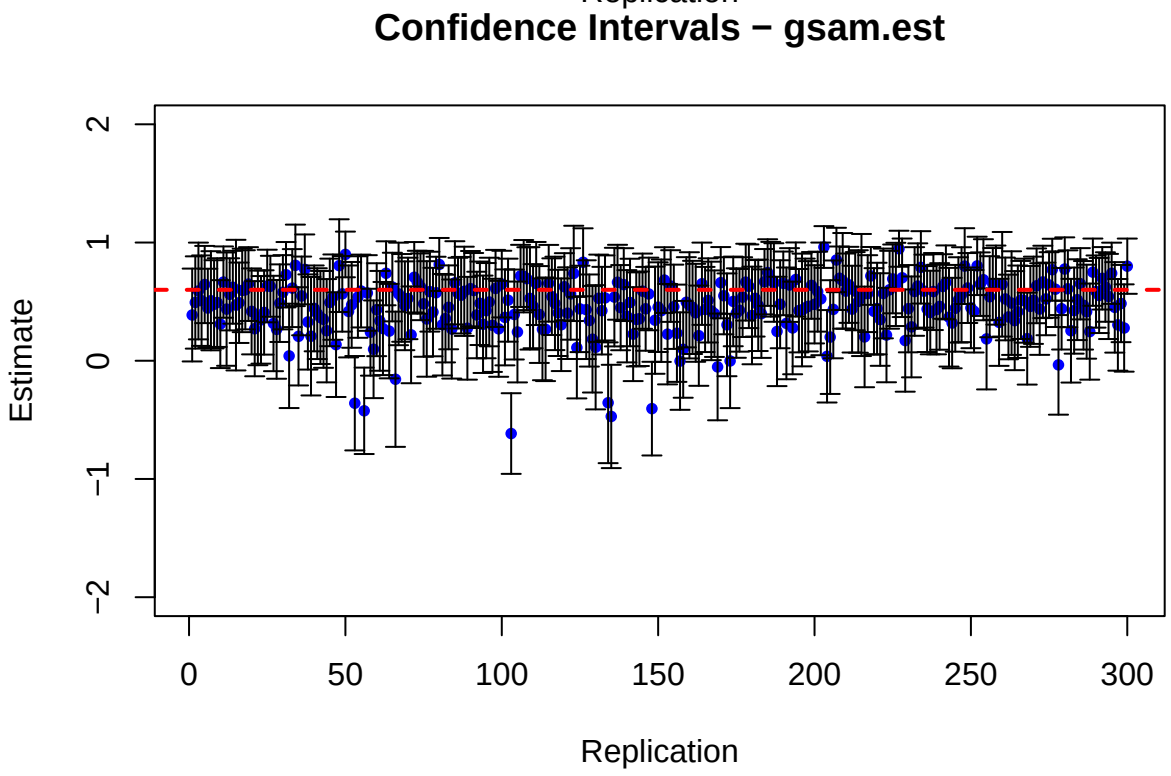
```
##                                Method Outlier_Percentage
## joint.se                    joint.se               0.60
## gsam.se                      gsam.se               1.40
## lsam.se                      lsam.se               1.40
## tspa.se                      tspa.se               3.70
## rel.se                        rel.se               4.05
## tspa.se_corrected tspa.se_corrected               5.60
## rel.se_corrected   rel.se_corrected               4.80
```

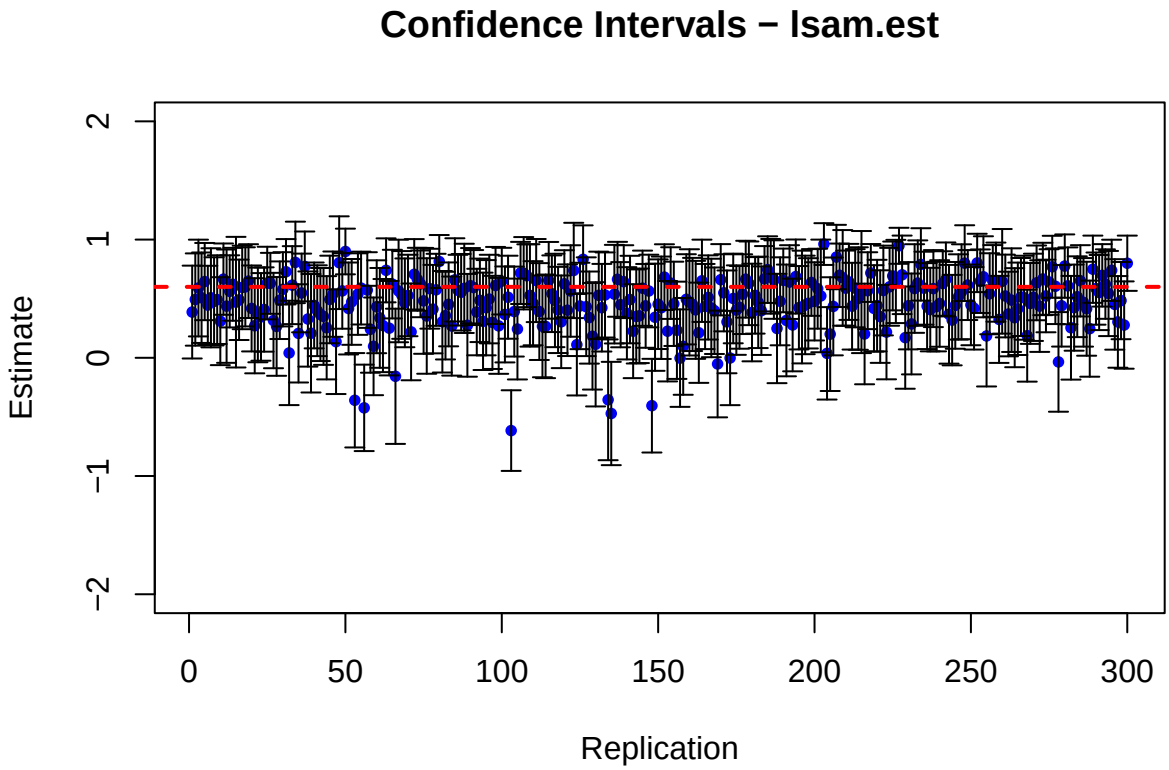
Density plot of SE distribution (Removed outliers)

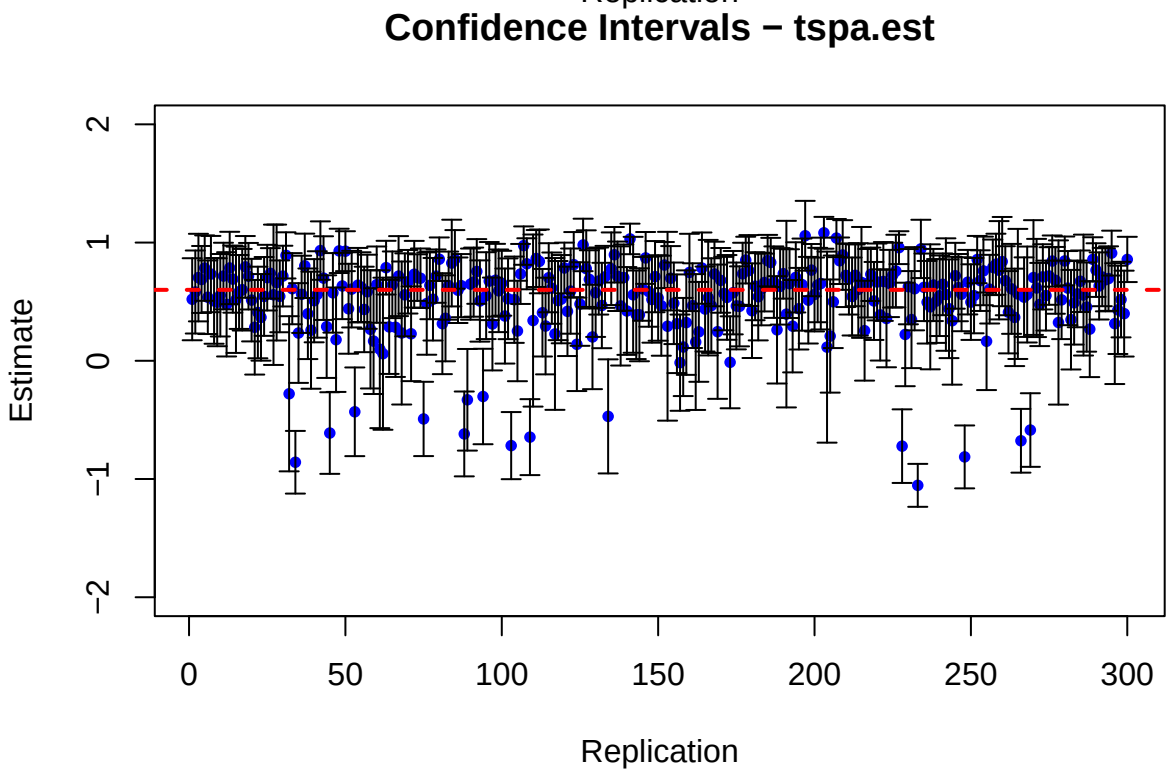


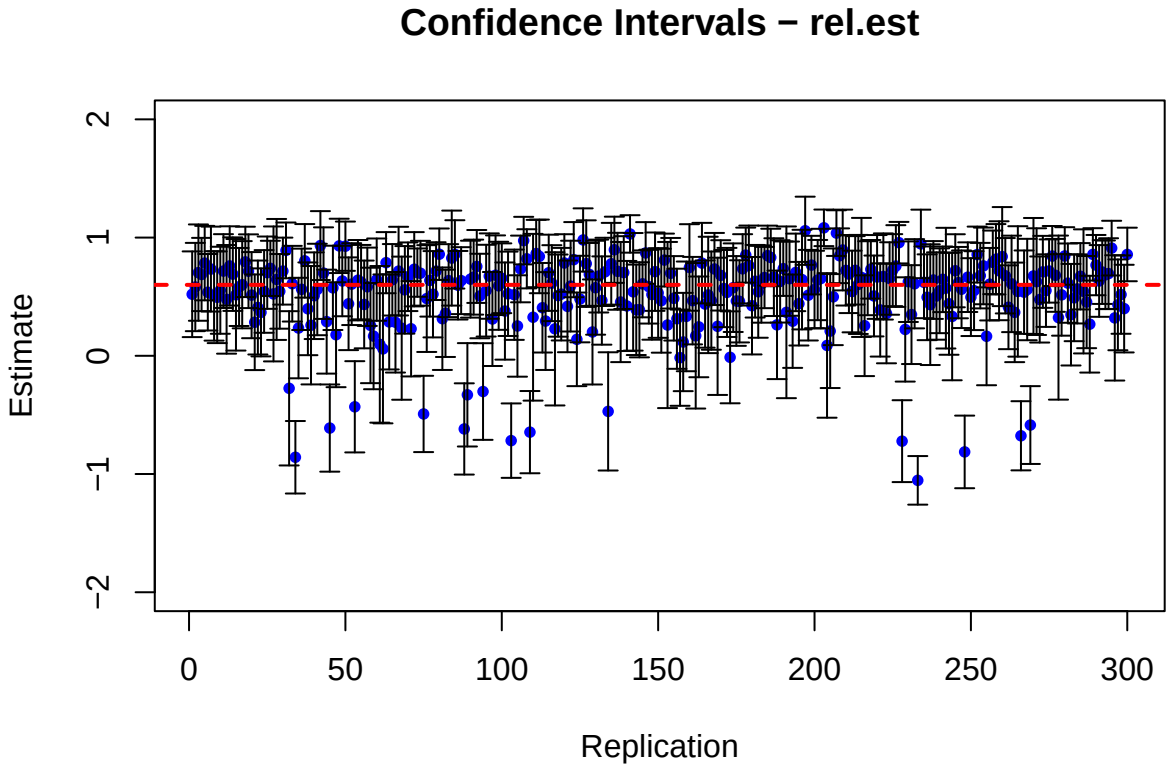
Confidence intervals (Without removing outliers)

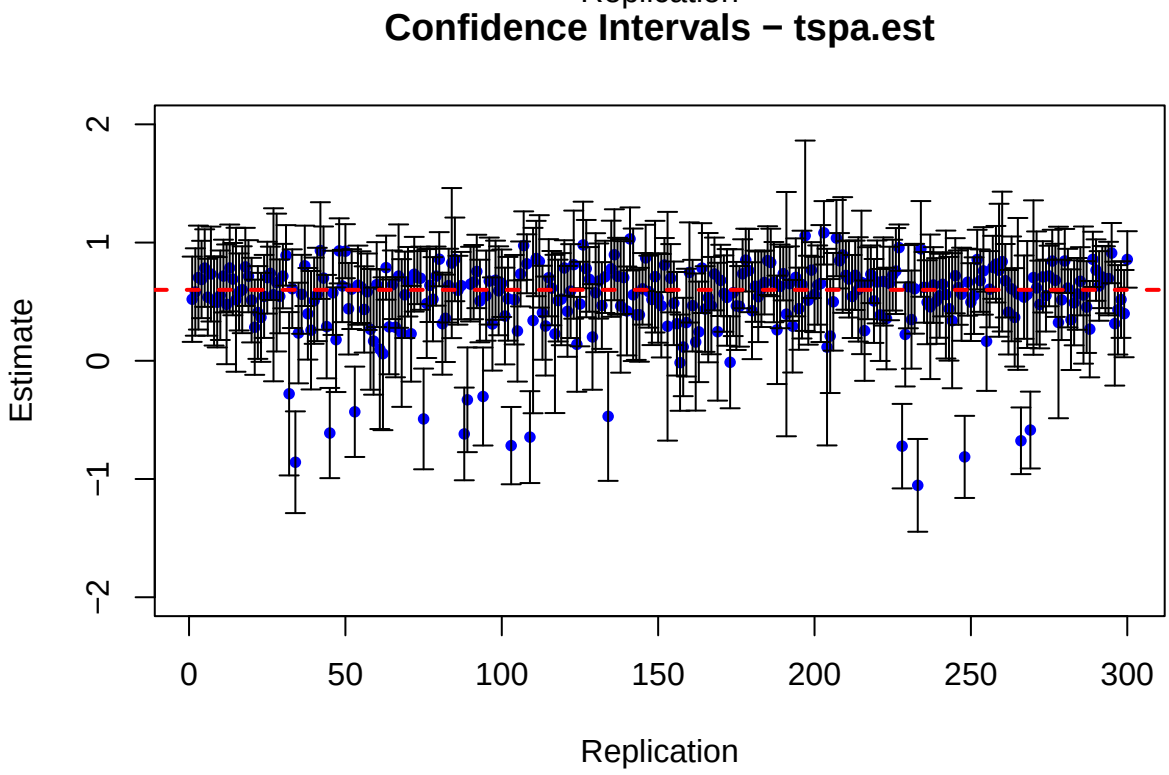


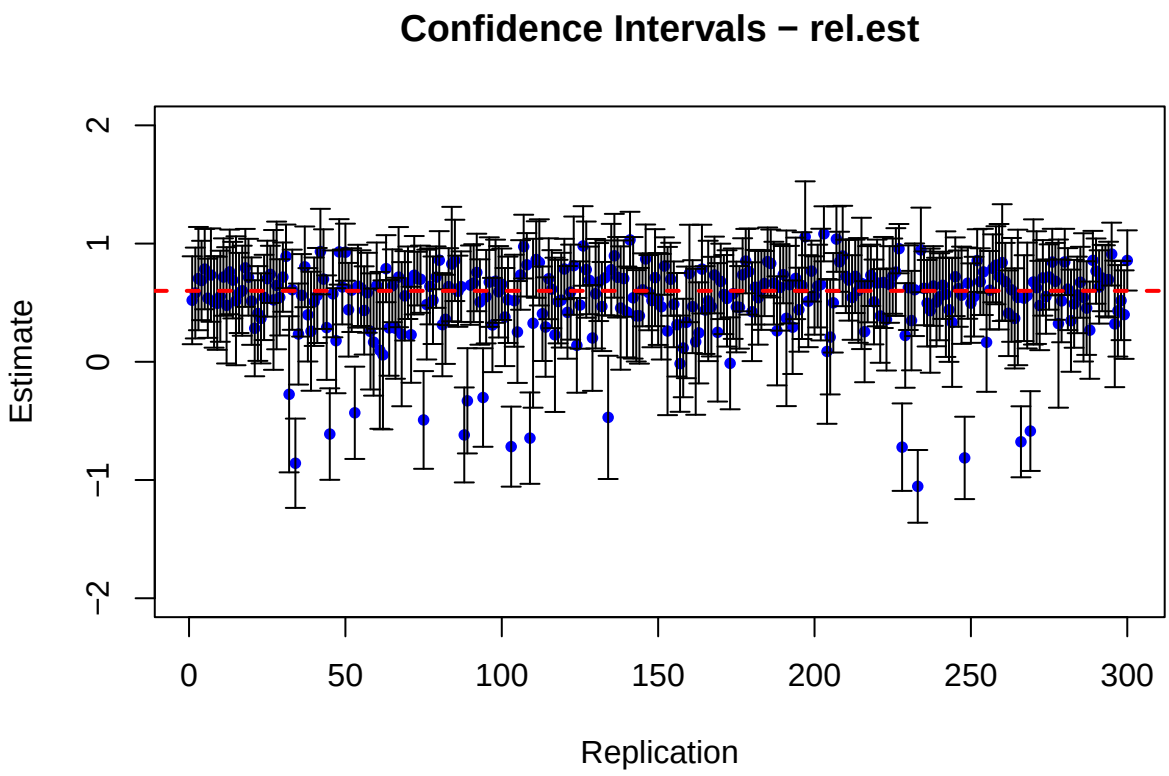



Confidence Intervals − lsam.est
Estimate
2
1
0
−1
−2
0
50
100
150
200
250
300
Replication


Confidence Intervals − tspa.est
Estimate
2
1
0
−1
−2
0
50
100
150
200
250
300
Replication


Confidence Intervals − rel.est
Estimate
2
1
0
−1
−2
0
50
100
150
200
250
300
Replication


Confidence Intervals − tspa.est
Estimate
2
1
0
−1
−2
0
50
100
150
200
250
300
Replication


Confidence Intervals − rel.est
Estimate
2
1
0
−1
−2
0
50
100
150
200
250
300
Replication

The green line represents the mean of estimated standard errors, while the red line is the empirical standard deviation of path coefficients. The green line (mean standard error) is consistently positioned to the left of the red line (empirical standard deviation), indicating that the model underestimates the true standard error.

However, the magnitude of this underestimation varies between methods:

- gsam.se and lsam.se: The difference between the red and green lines is relatively small compared to other methods, indicating that these methods might provide a closer estimate of the true variability.
- rel.se and tspa.se_corrected: There is a more substantial difference between the red and green lines, suggesting that these methods tend to underestimate the standard error more compared to others.

For shape and spread of density curves:

- The density curves for joint.se, gsam.se, lsam.se, and tspa.se are relatively similar, with a peak around 0.2 to 0.3 and a sharp decline, indicating that these methods have similar distributions of standard errors.
- The tspa.se_corrected and rel.se methods show a slightly broader spread, suggesting greater variability in the calculated standard errors.