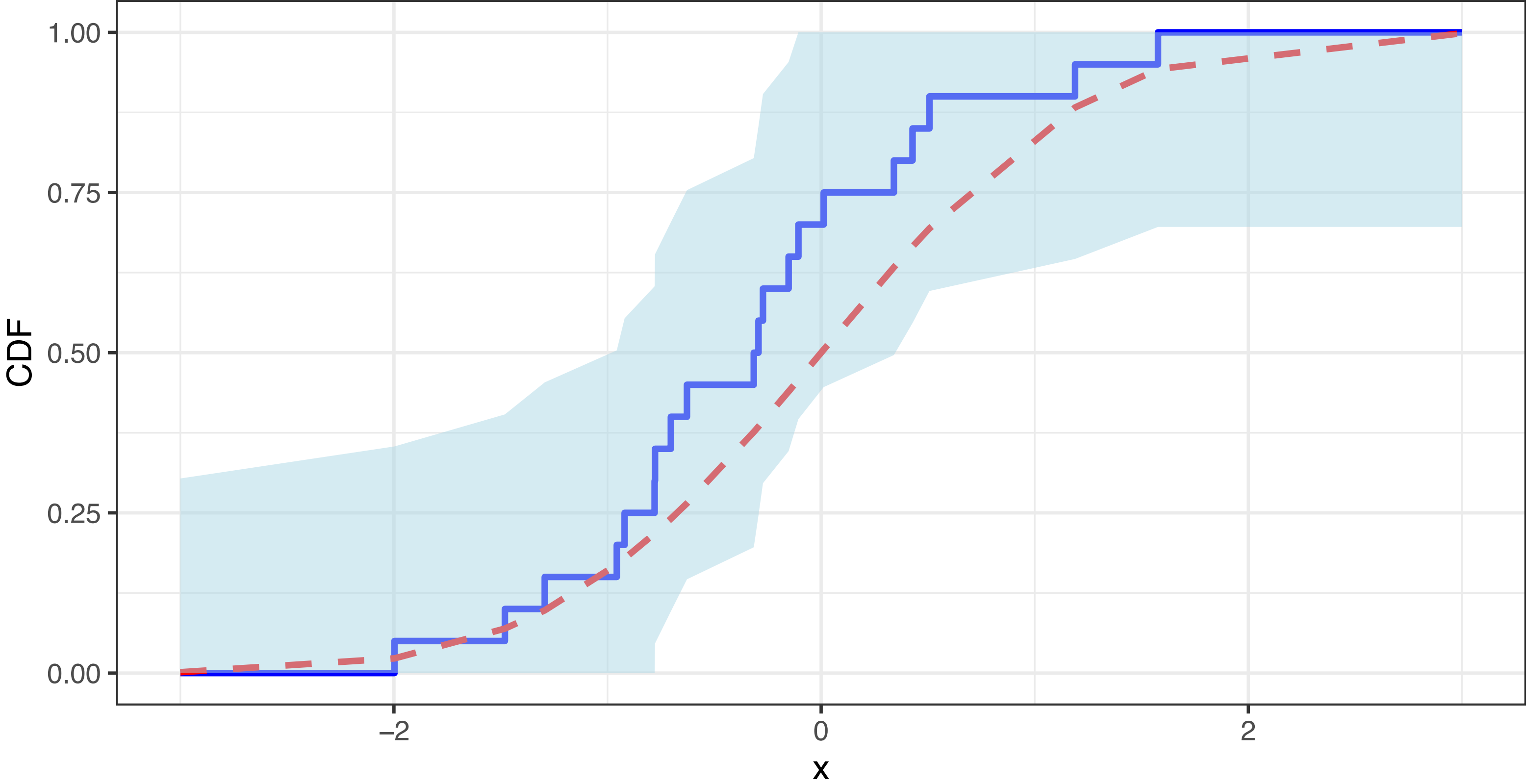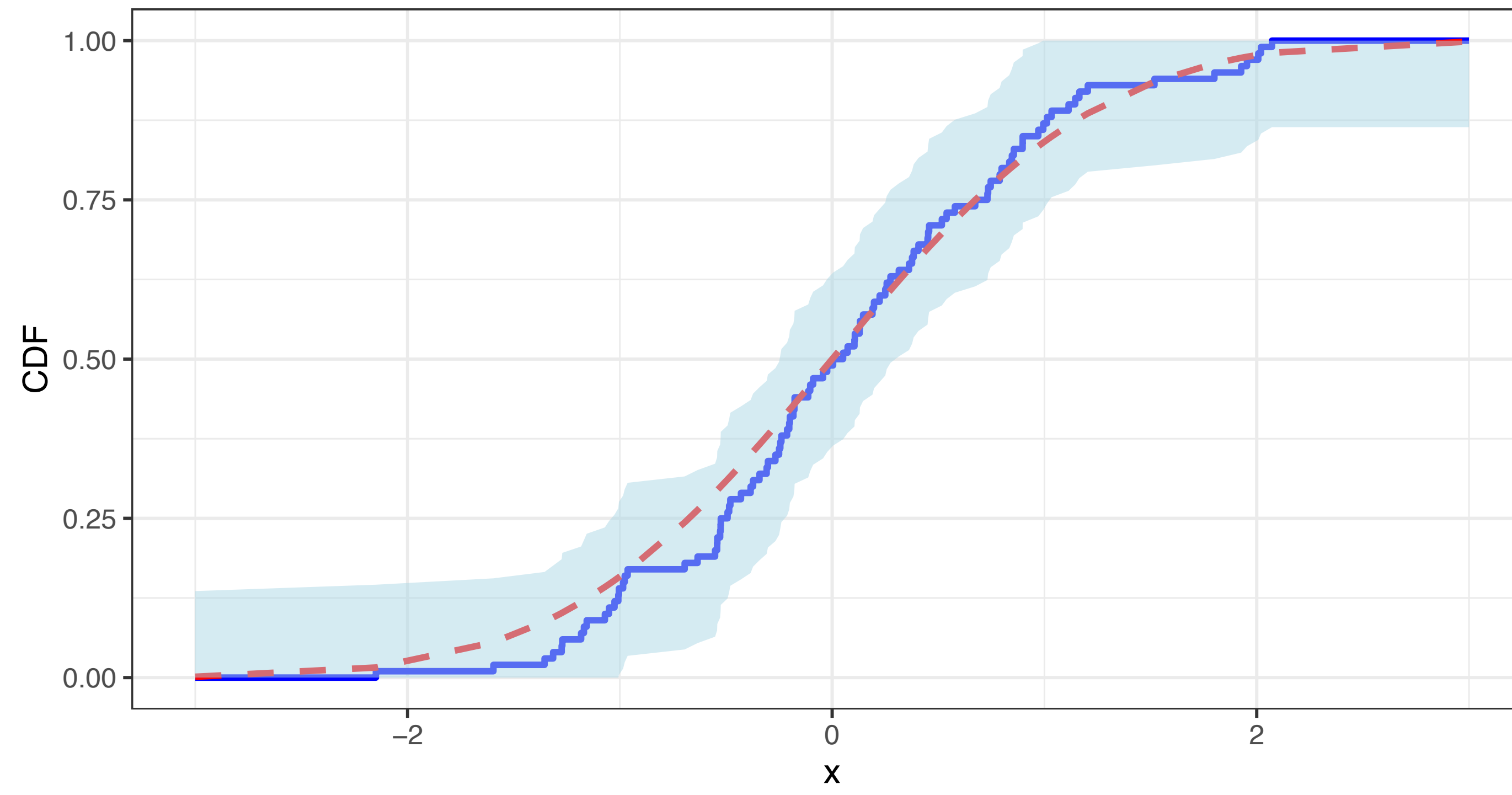# Empirical CDFs and Bootstrap
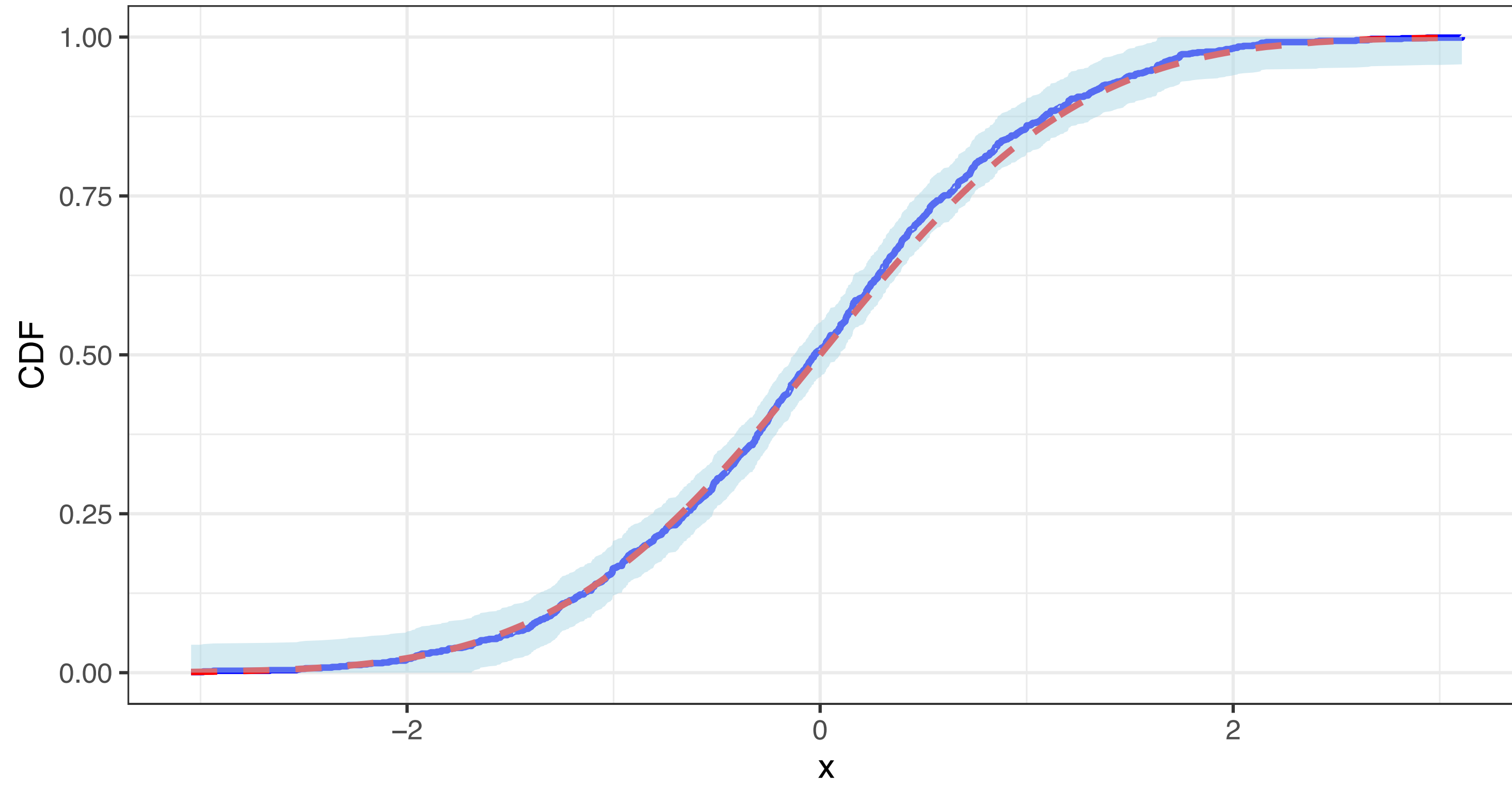
Empirical CDF vs True CDF (n = 20)
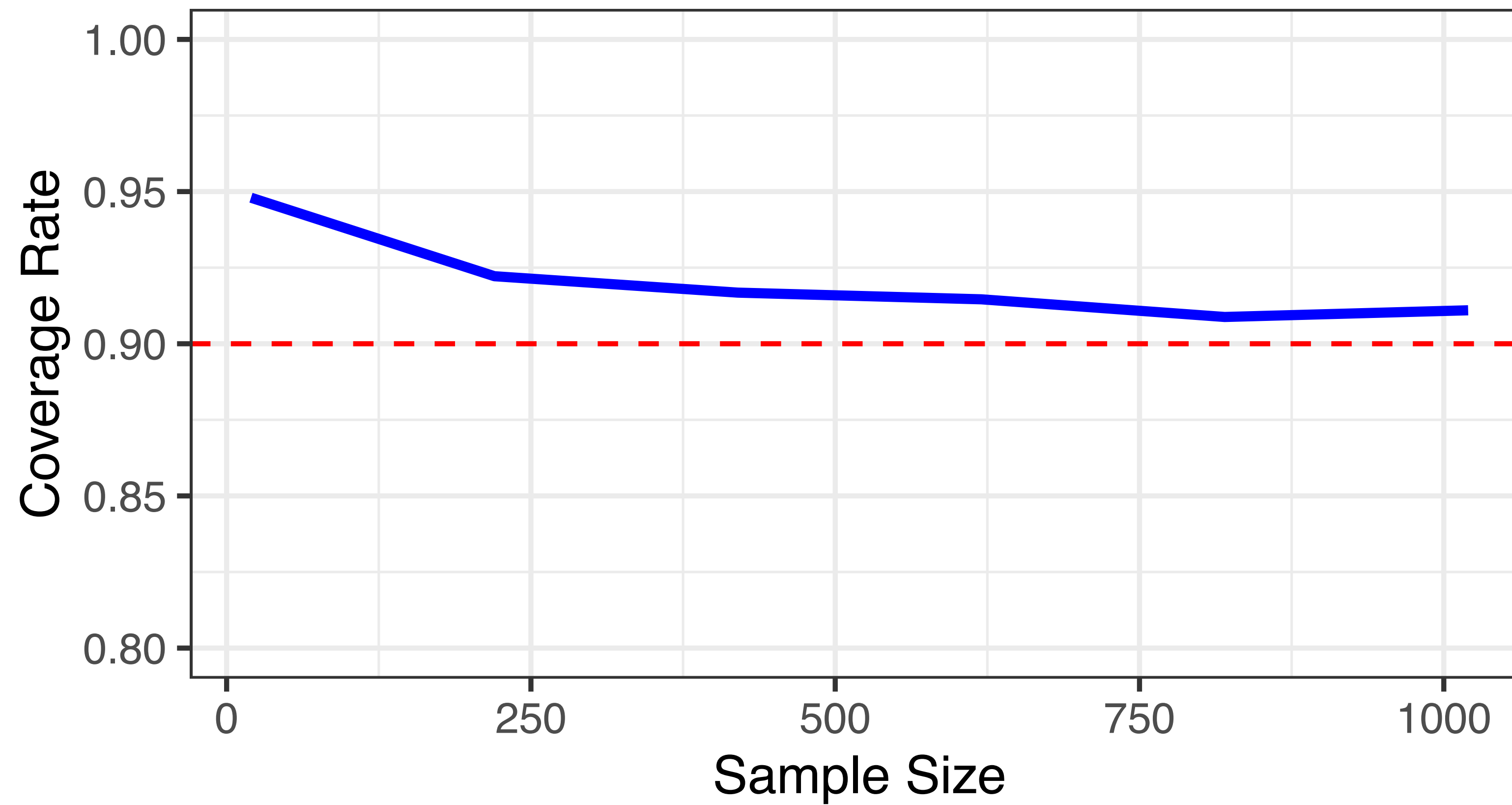
Empirical CDF vs True CDF (n = 100)
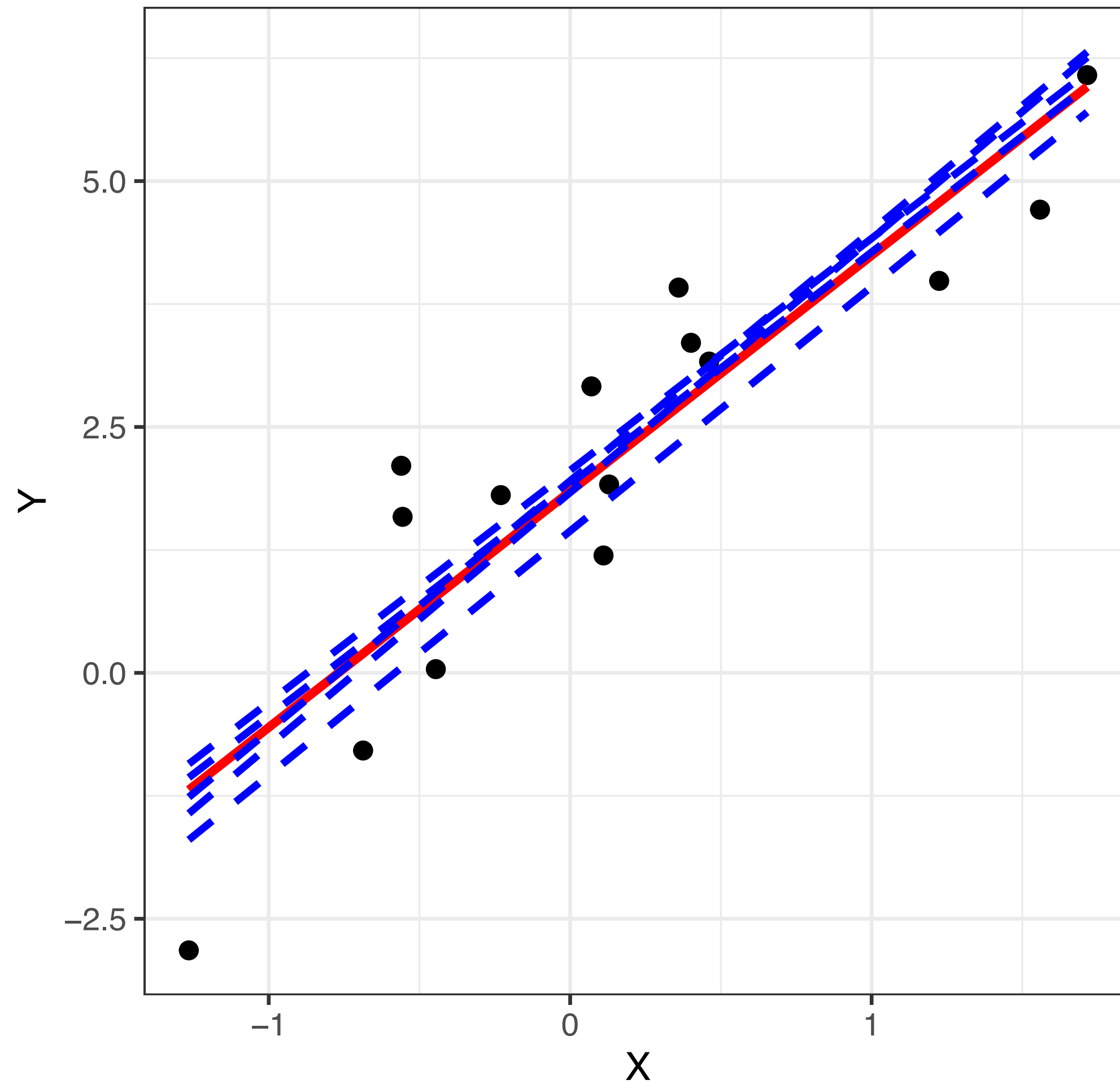
Empirical CDF vs True CDF (n = 1000)

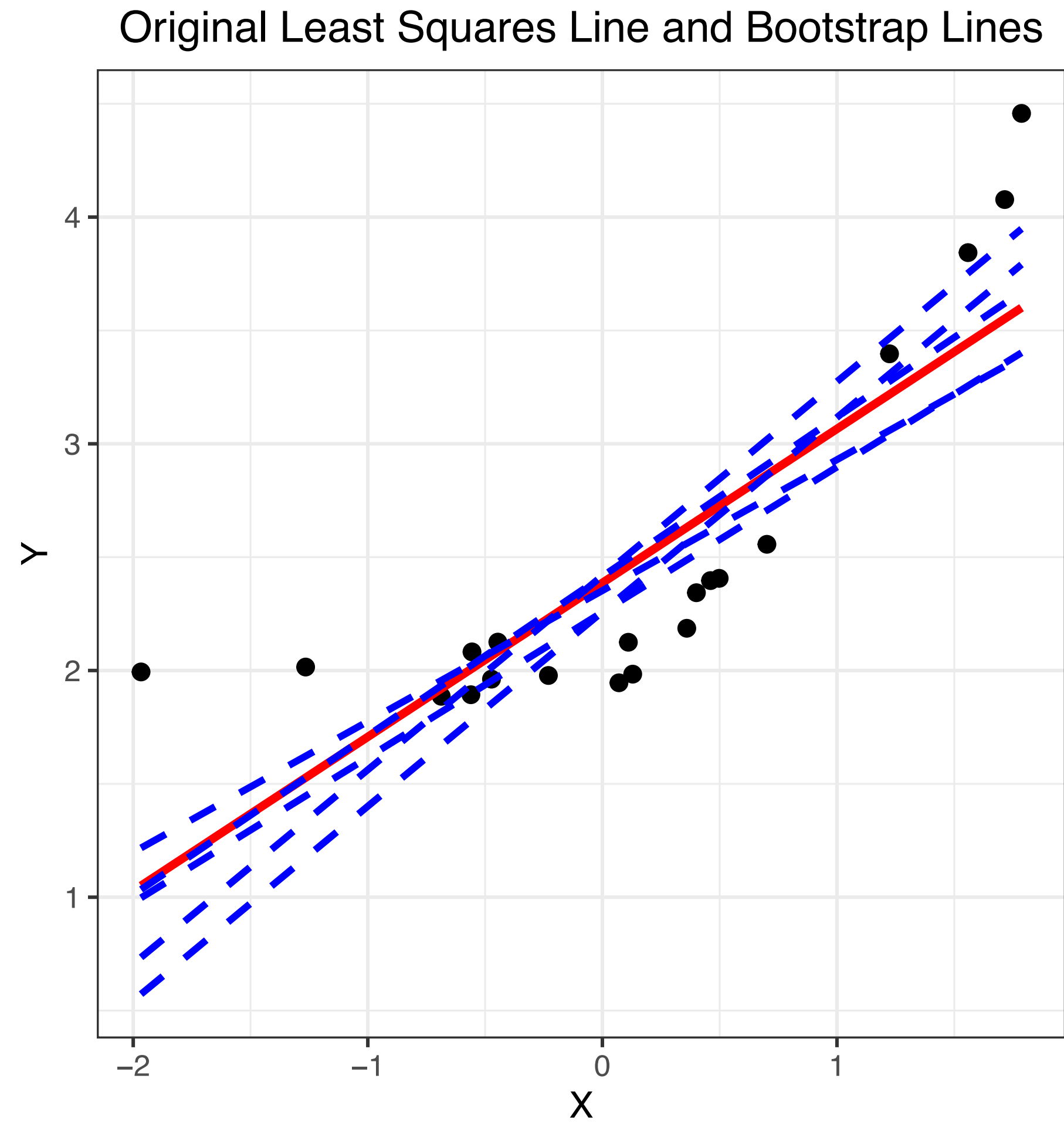Coverage Rate of DKW Confidence Bands

Linear model

Original Least Squares Line and Bootstrap Lines

$$Y_i = 2 + 3 * X_i + \epsilon_i \text{ where } \epsilon_i \sim \mathcal{N}(0,1)$$

$$n = 15$$
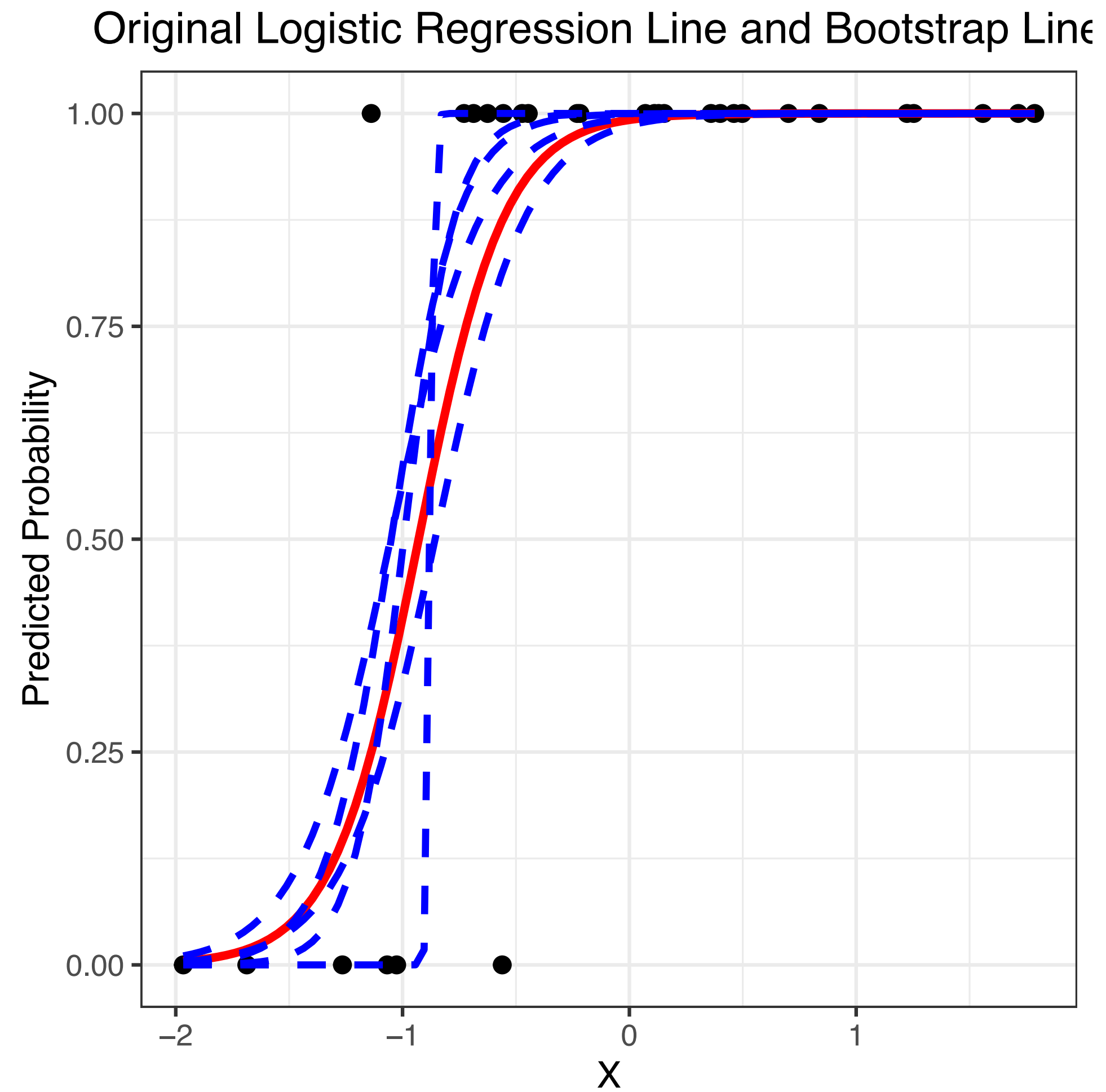
Misspecified linear model

Original Least Squares Line and Bootstrap Lines

$$Y_i = 2 + \max(0, X_i)^{1.5} + .1 \cdot \epsilon_i \text{ where } \epsilon_i \sim \mathcal{N}(0,1)$$

$$n = 20$$

Logistic regression

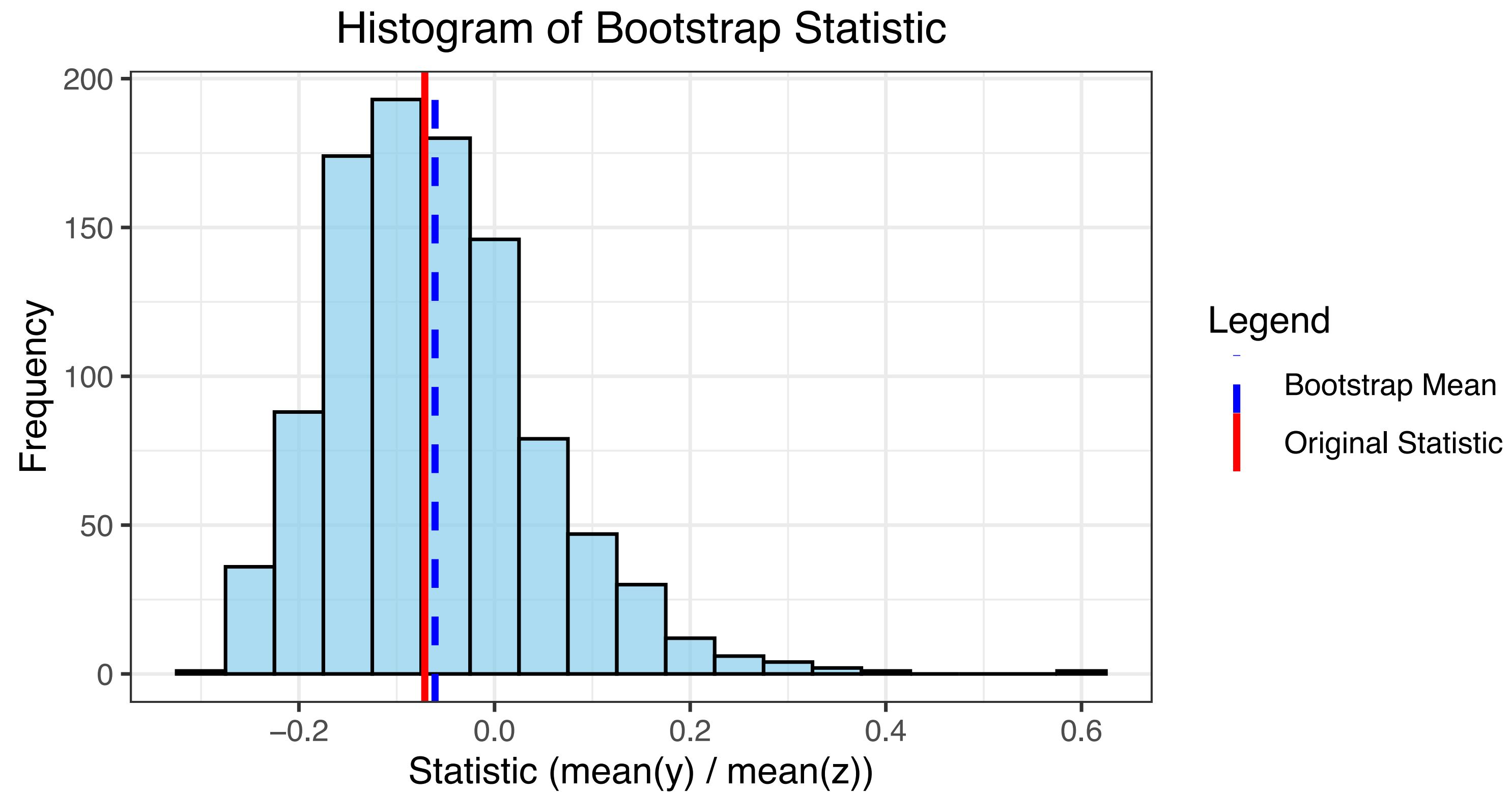Original Logistic Regression Line and Bootstrap Line

$$P(Y_i = 1) = \frac{1}{1 + \exp(-(2 + 3X_i))}$$

$$n = 30$$

Pivotal bootstrap

Histogram of Bootstrap Statistic

Legend

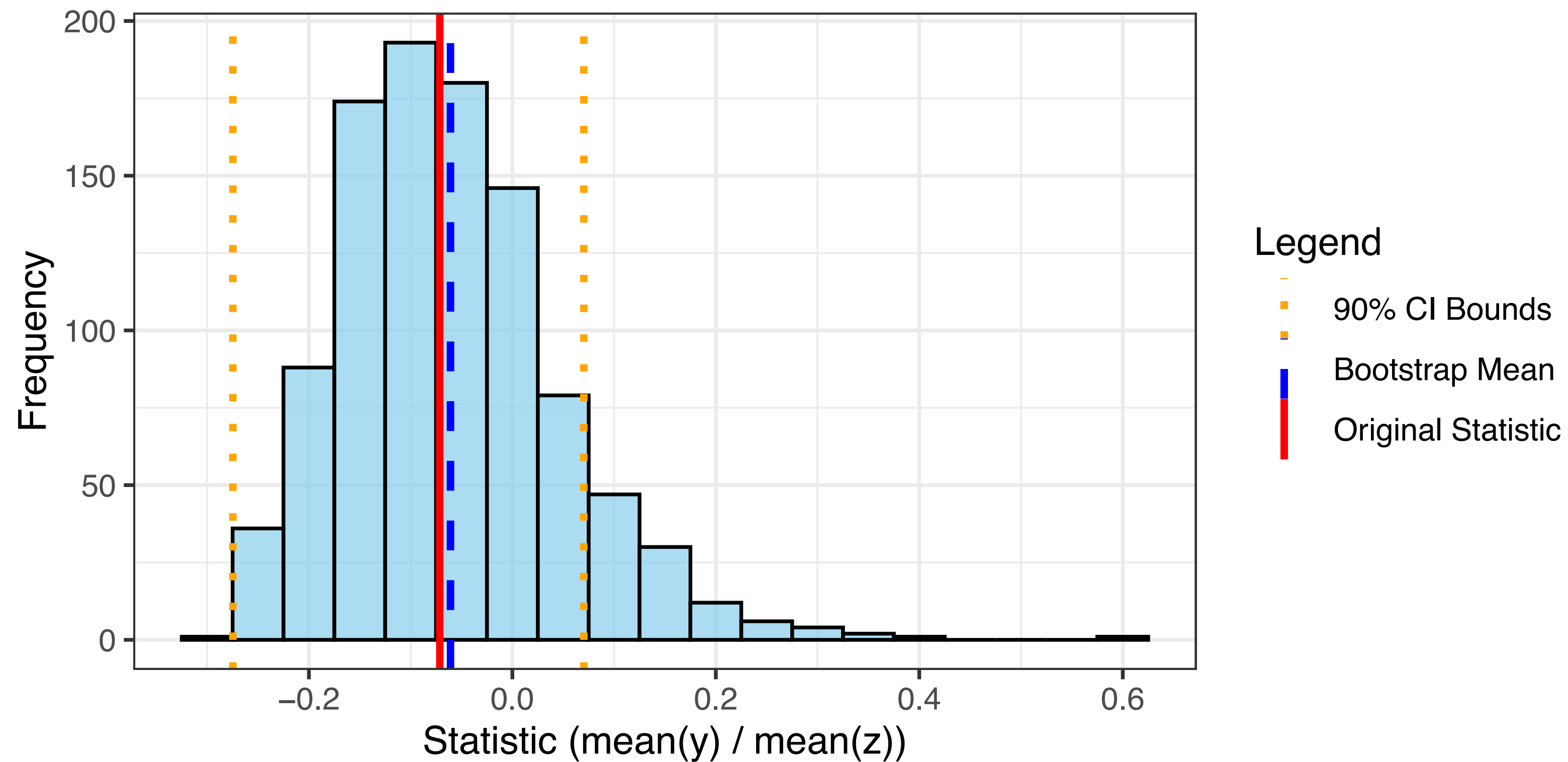- Bootstrap Mean
- Original Statistic

Data: $(Y_1, Z_1), \ldots, (Y_n, Z_n)$

Statistic: $T(\text{data}) = \dfrac{\bar{Y}}{\bar{Z}}$

1. slight upward bias of the statistic

2. sampling distribution skew right

Pivotal bootstrap

Histogram of Bootstrap Statistic with Pivotal Bootstrap 90% CI

Data: $(Y_1, Z_1), \ldots, (Y_n, Z_n)$

Statistic: $T(\text{data}) = \dfrac{\bar{Y}}{\bar{Z}}$

1. slight upward bias of the statistic -> we should move CI slightly to left

2. sampling distribution skew right -> should have wider CI on <u>left</u>