

<sup>1</sup> Theoretical Bias and variance, as a function of population parameters

<sup>2</sup> Delacre Marie<sup>1</sup>

<sup>3</sup> <sup>1</sup> ULB

Theoretical Bias and variance, as a function of population parameters

### The bias

For all estimators, when the population effect size is null so is the bias. For each estimator, We will subdivide all configurations when there is a non-null population effect into 3 conditions:

- when variances are equal across groups,
- when variances are unequal across groups, with equal sample sizes
- when variances are unequal across groups, with unequal sample sizes

### Cohen's $d_s$

**When variances are equal across populations.** The **bias** of Cohen's  $d_s$  is a function of total sample size (N) and the population effect size ( $\delta_{Cohen}$ ):

- The larger the population effect size, the more Cohen's  $d_s$  will overestimate  $\delta_{Cohen}$ .
- The larger the total sample size, the lower the bias. The bias tends to zero when the total sample size tends to infinity (see Figure 1)

Of course, considering the degrees of freedom, the sample size ratio does not matter... (see Figure 2)

### Glass's $d_s$

Because degrees of freedom depend only on the sample size of the control group, there is no need to distinguish between cases where there is homoscedasticity or heteroscedasticity!

The **bias** of Glass's  $d_s$  is a function of the sample size of the control group ( $n_c$ ) and the population effect size ( $\delta_{glass}$ ):

- The larger the population effect size, the more Glass's  $d_s$  will overestimate  $\delta_{Glass}$ .

- The larger the size of the control group, the lower the bias. The bias tends to zero when the sample size of the control group tends to infinity (see Figure 3)

### Cohen's $d'_s$

**When variances are equal across populations.** When  $s_1 = s_2 = s$ :

$$df_{Cohen's d'_s} = \frac{(n_1 - 1)(n_2 - 1)(2s^2)^2}{(n_2 - 1)s^4 + (n_1 - 1)s^4} = \frac{(n_1 - 1)(n_2 - 1)4s^4}{s^4(n_1 + n_2 - 2)} = \frac{4(n_1 - 1)(n_2 - 1)}{n_1 + n_2 - 2}$$

One can see that degrees of freedom depend only on the total sample size ( $N$ ) and the sample size allocation ratio. As a consequence, the **bias** of Cohen's  $d'_s$  is a function of the population effect size ( $\delta'_{Cohen}$ ), the sample size allocation ratio and the total sample size ( $N$ ).

- The larger the population effect size, the more *Cohen's  $d'_s$*  will overestimate  $\delta'_{Cohen}$
- The further the sample size allocation ratio is from 1, the larger the bias (see Figure 4)
- The larger the total sample size, the lower the bias (see Figure 5)

**When variances are unequal across populations, with equal sample sizes.**

When  $n_1 = n_2 = n$ :

$$df_{Cohen's d'_s} = \frac{(n - 1)^2(s_1^2 + s_2^2)^2}{(n - 1)(s_1^4 + s_2^4)} = \frac{(n - 1)(s_1^4 + s_2^4 + 2s_1^2s_2^2)}{s_1^4 + s_2^4}$$

One can see that degrees of freedom depend only on the total sample size ( $N$ ) and the SD-ratio. As a consequence, the **bias** of Cohen's  $d'_s$  is a function of the population effect size ( $\delta'_{Cohen}$ ), the SD-ratio and the total sample size ( $N$ ):

- The larger the population effect size, the more *Cohen's  $d'_s$*  will overestimate  $\delta'_{Cohen}$
- The further the SD-ratio is from 1, the larger the bias (see Figure 6)
- The larger the total sample size, the lower the bias (see Figure 7)

Note: for a constant SD-ratio, the size of the variance does not matter (see Figure 8)

**When variances are unequal across populations, with unequal sample sizes.** The **bias** of Cohen's  $d'_s$  is a function of the population effect size ( $\delta'_{Cohen}$ ), the total sample size, and the sample sizes and variances pairing :

- The larger the population effect size, the more *Cohen's*  $d'_s$  will overestimate  $\delta'_{Cohen}$
- When there is a positive pairing between sample sizes and variances, one gives more weight to the smallest variance. As a consequence, the denominator in the df computation decreases, the degrees of freedom increase and therefore, the bias decreases (see the two plots on the top in Figure 7). On the other size, where there is a negative pairing between sample sizes and variances, one gives more weight to the largest variance. As a consequence, the denominator in the df computatin increases, the degrees of freedom decrease and therefore, the bias increase (see the two plots in the bottom of Figure 7).
- The larger the total sample size, the lower the bias (illustration in Figure 10)

Note: for a constant SD-ratio, the variance does not matter. (See Figure 11)

**Shieh's  $d_s$**

**When variances are equal across populations.**

**When variances are unequal across populations, with equal sample sizes.**

**When variances are unequal across populations, with unequal sample sizes.**

**The variance**

**Cohen's  $d_s$**

**When variances are equal across populations.**

66 Glass's  $d_s$

67 When variances are equal across populations.

68 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

69 — ...

70 — ...

71 When variances are unequal across populations, with equal sample sizes.

72 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

73 — ...

74 — ...

75 When variances are unequal across populations, with unequal sample sizes.

76 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

77 — ...

78 — ...

79 Cohen's  $d'_s$

80 When variances are equal across populations.

81 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

82 — ...

83 — ...

84 When variances are unequal across populations, with equal sample sizes.

85 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

86 — ...

87 — ...

88 When variances are unequal across populations, with unequal sample sizes.

89 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

90 — ...

91 — ...

92 **Shieh's  $d_s$**

93 **When variances are equal across populations.**

94 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

95 — ...

96 — ...

97 **When variances are unequal across populations, with equal sample sizes.**

98 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

99 — ...

100 — ...

101 **When variances are unequal across populations, with unequal sample sizes.**

102 • The **variance** of Cohen's  $\delta'_s$  is a function of ...:

103 — ...

104 — ...

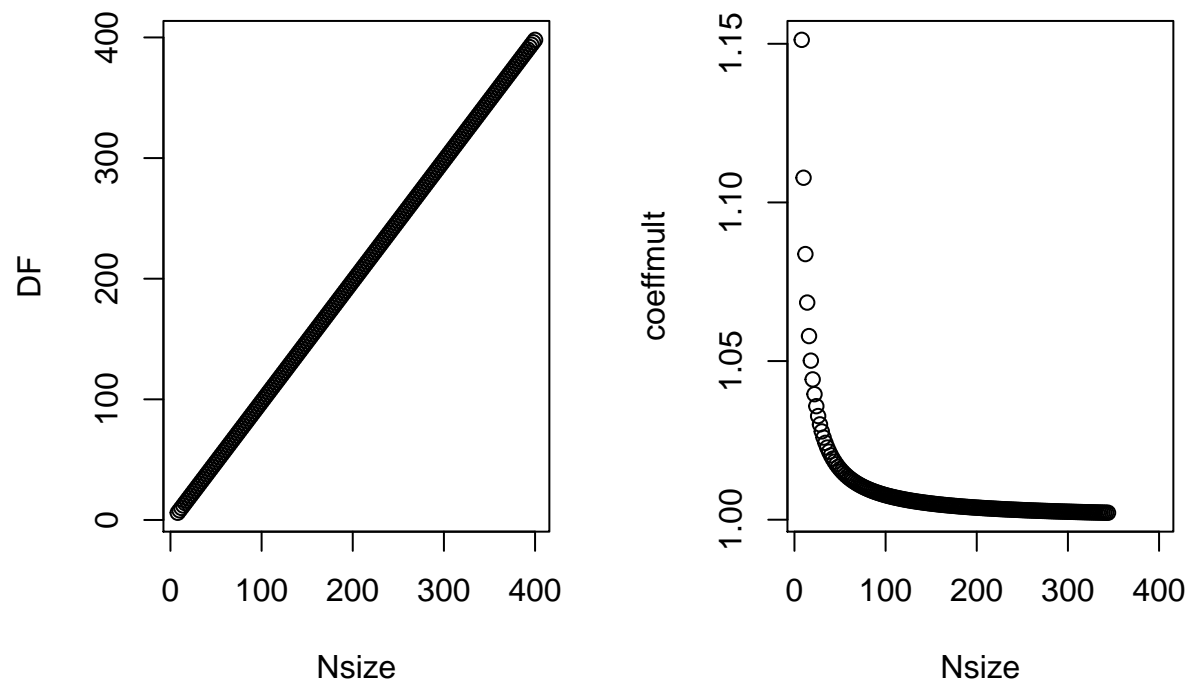


Figure 1. ...

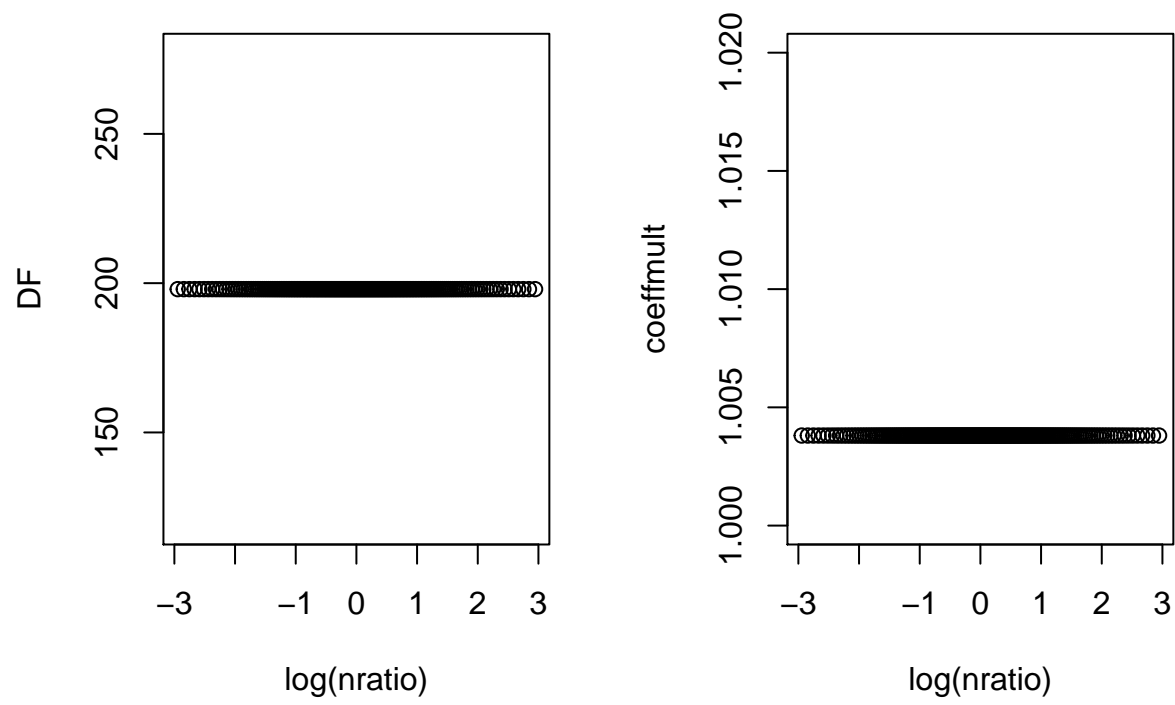


Figure 2. ...



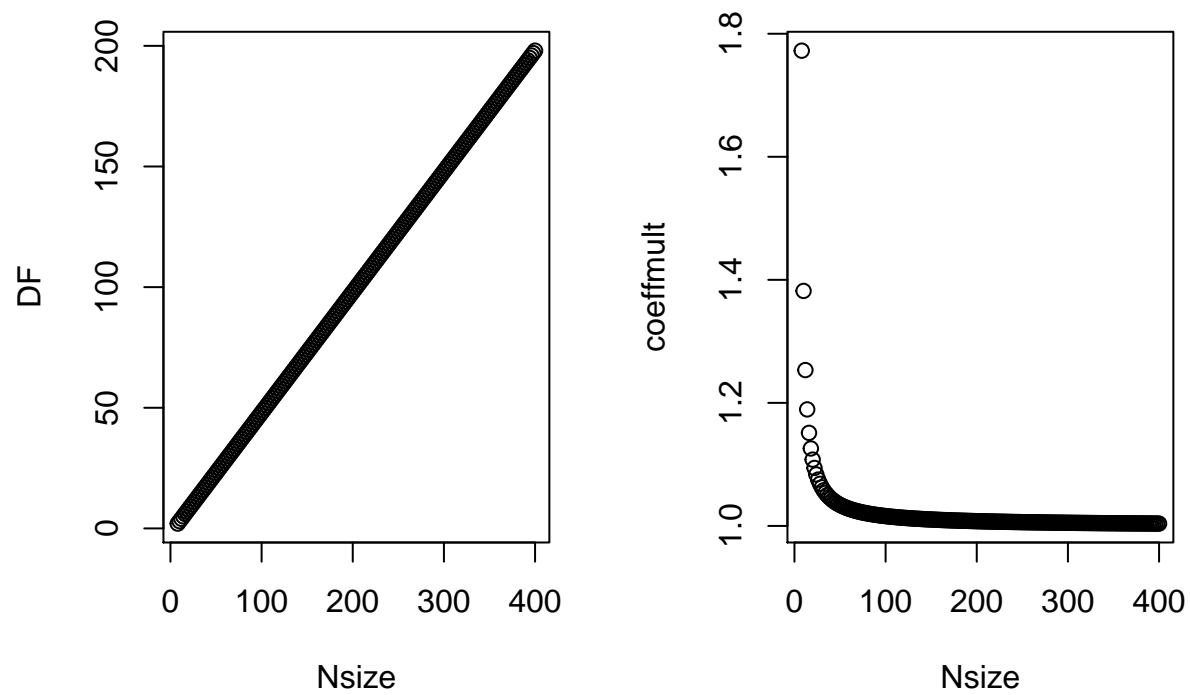


Figure 3. ...

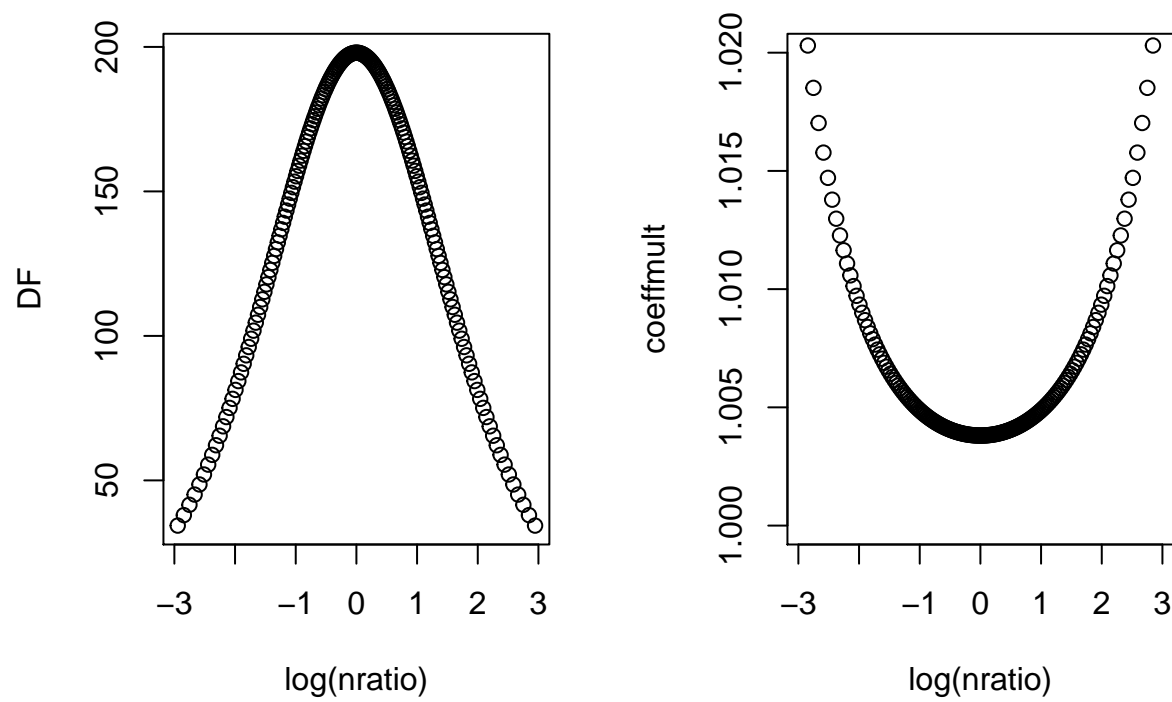


Figure 4. ...

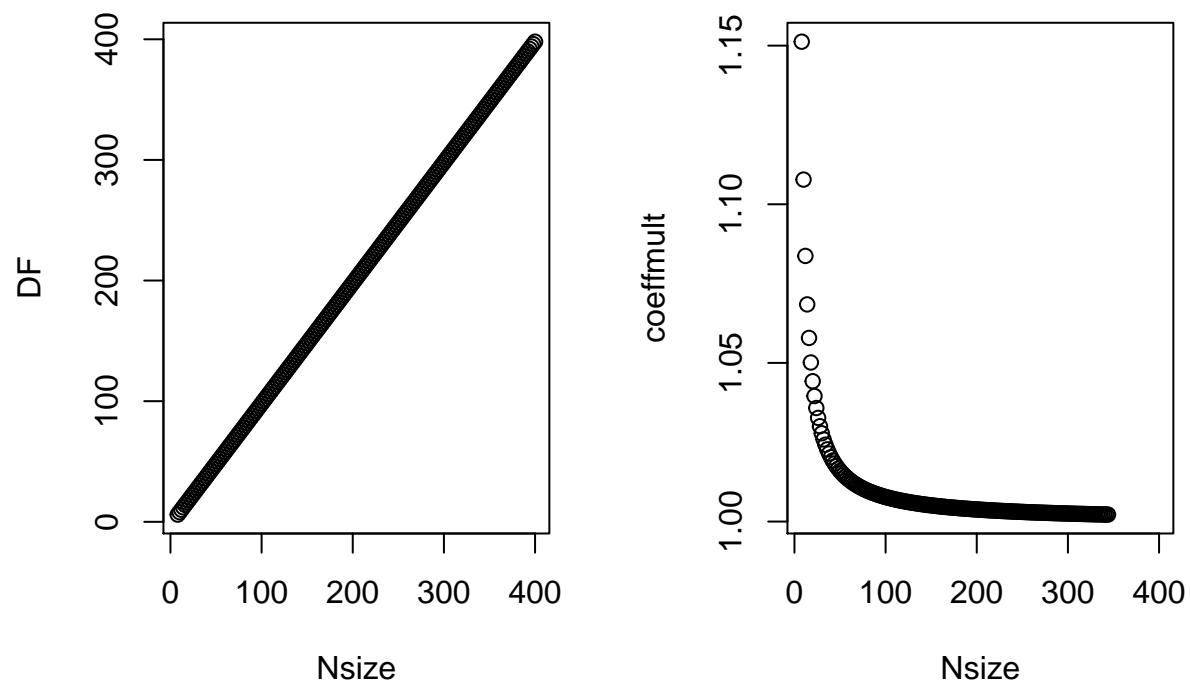


Figure 5. ...

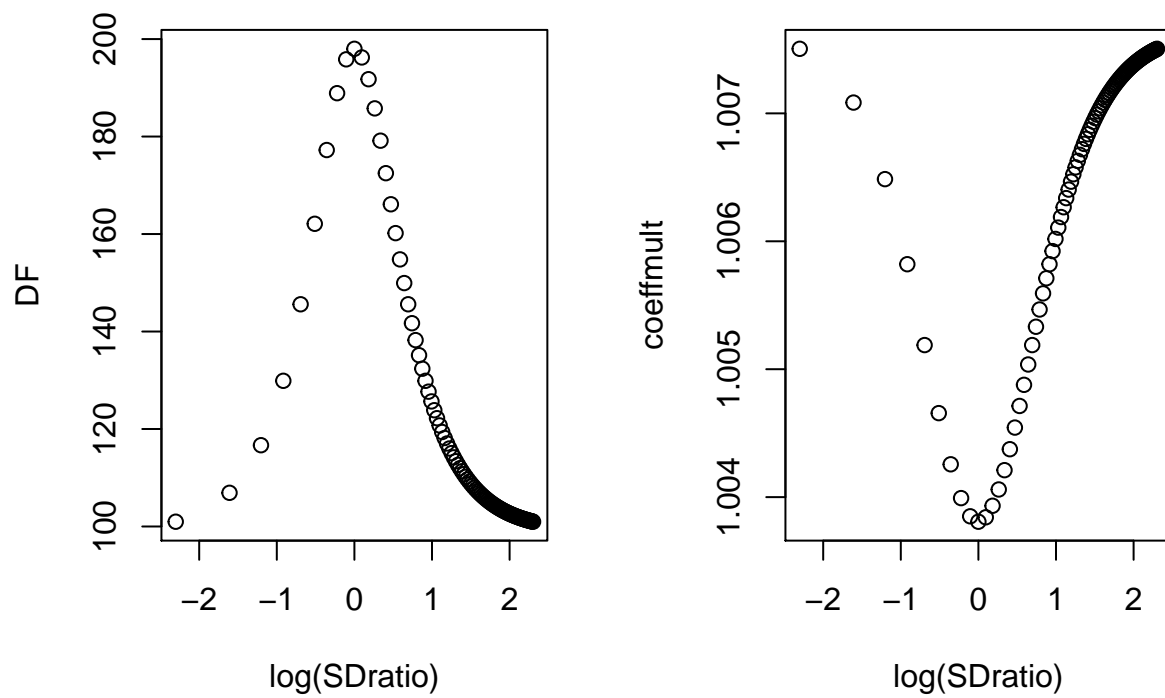


Figure 6. bias of Cohen's  $d'_s$  as a function of the total sample size, when variances are equal across groups

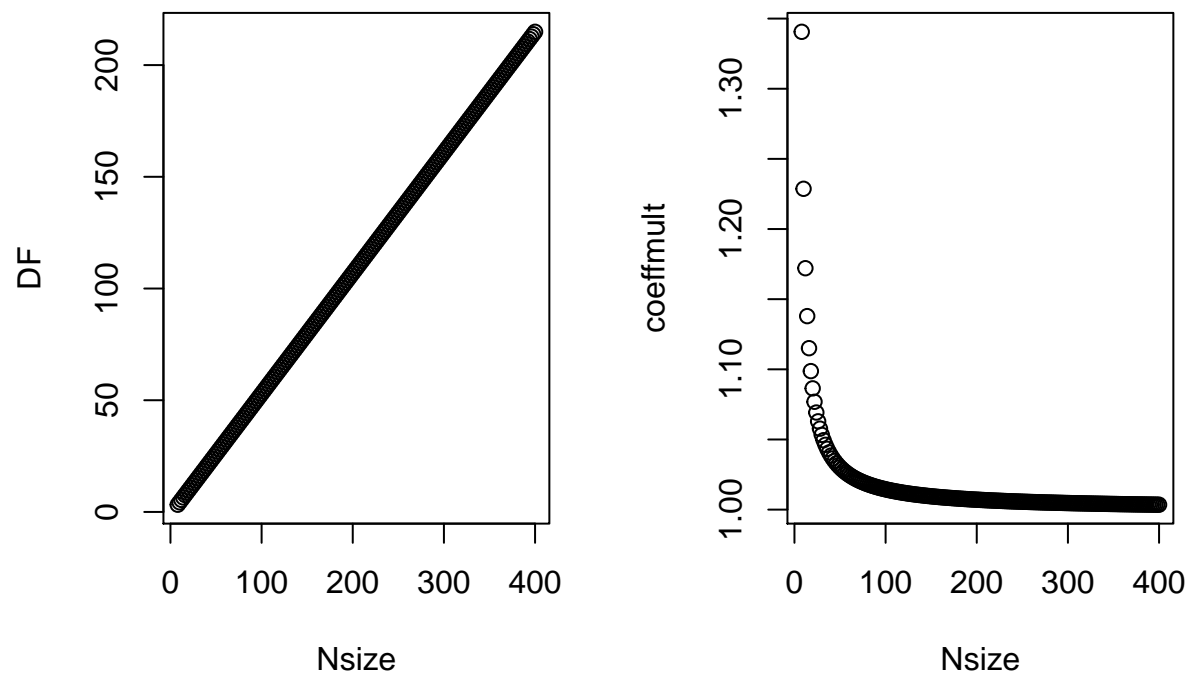


Figure 7. bias of Cohen's  $d'_s$  as a function of the total sample size, when variances are equal across groups

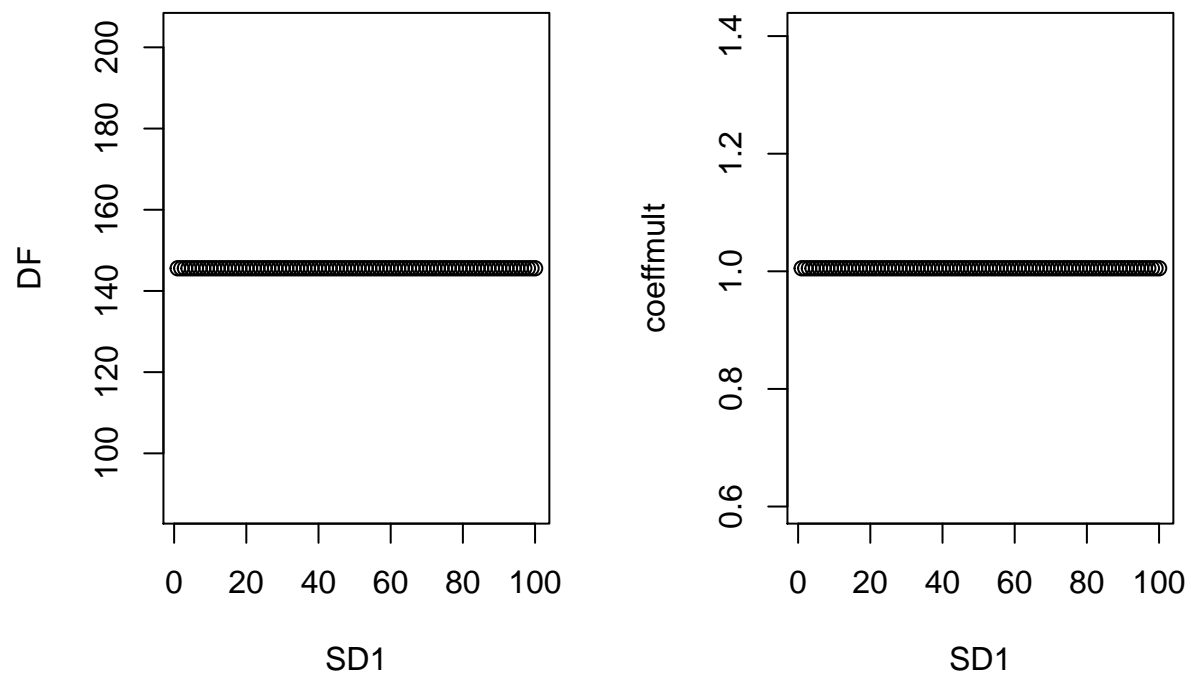


Figure 8. bias of Cohen's  $d'_s$  as a function of the total sample size, when variances are equal across groups

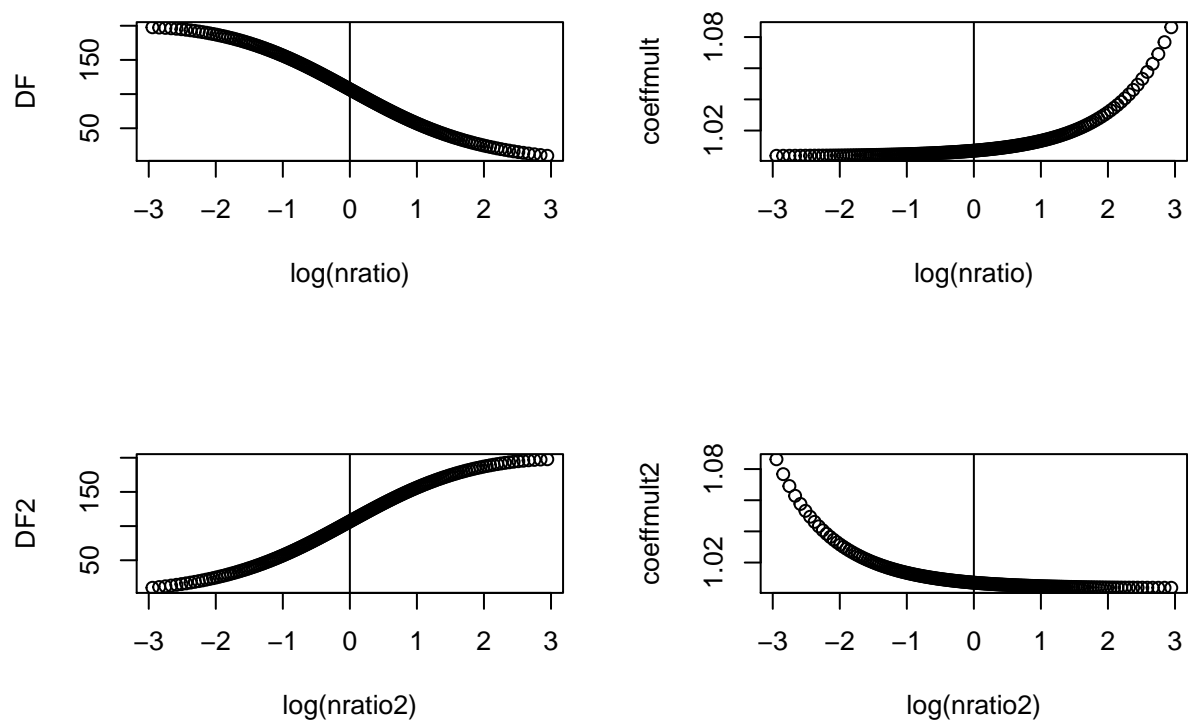


Figure 9. bias of Cohen's  $d'_s$  as a function of the total sample size, when variances are equal across groups

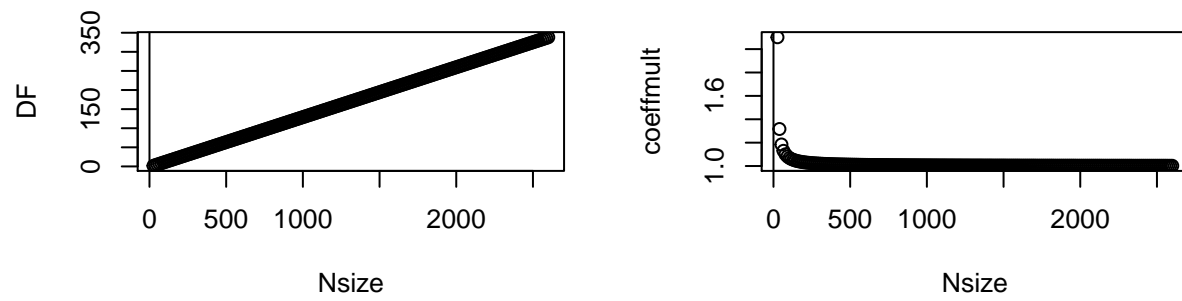
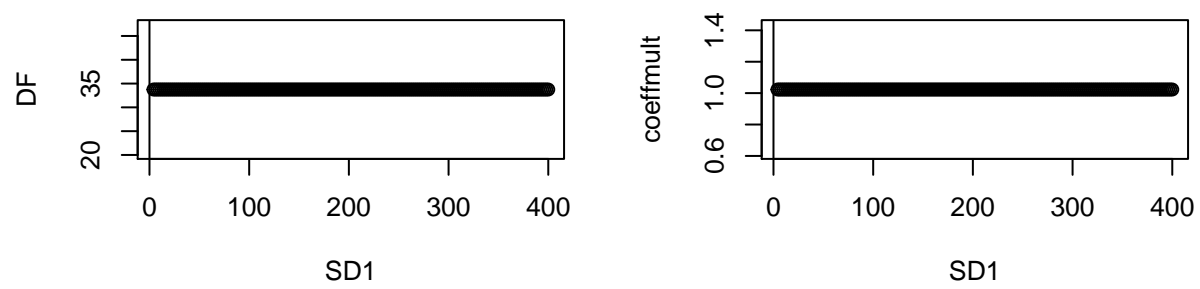


Figure 10. bias of Cohen's  $d'_s$  as a function of the total sample size, when variances are equal across groups





*Figure 11.* bias of Cohen's  $d'_s$  as a function of the total sample size, when variances are equal across groups