Relative frequency density (RFD), revisited.

Suppose we have a data set

$$X = \{x_1, x_2, x_3, \dots, x_n\}$$

of n real numbers, a range $[a,b)$ of possible values of X, and f of the points of X are actually in $[a,b)$. Write $B = b-a$, so B is the bin width (= length of the range).

We define the RFD of the range $[a,b)$ by

$$RFD = \frac{f}{B \cdot n}. \qquad (*)$$

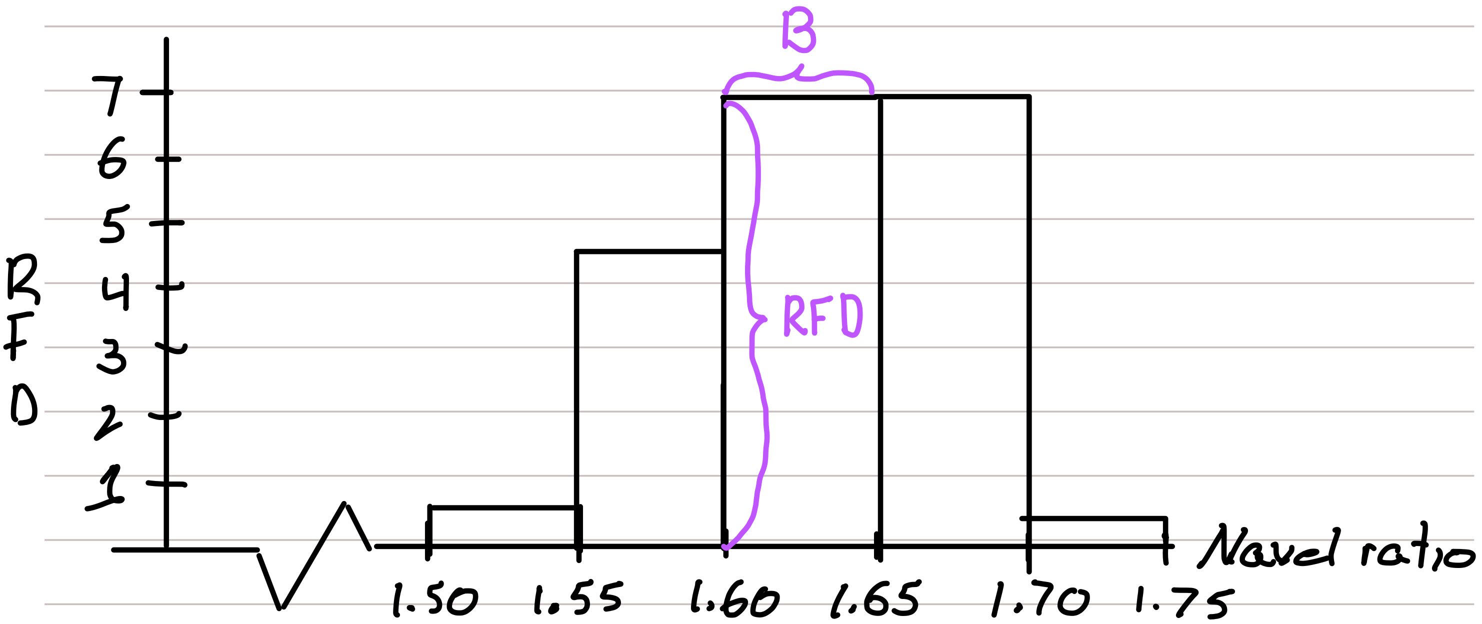
Example.

We recorded navel ratios (= height divided by height of navel) of 48 CU students. Here's an RFD table:

Range $[a,b)$	Frequency f	RFD
[1.50, 1.55)	2	2/(.05·48) = 0.83
[1.55, 1.60)	11	11/(.05·48) = 4.58
[1.60, 1.65)	17	17/(.05·48) = 7.08
[1.65, 1.70)	17	17/(.05·48) = 7.08
[1.70, 1.75)	1	1/(.05·48) = 0.42

Here's an RFD histogram:

48 Navel Ratios



Again: why do we care about RFD?

Answer: in an RFD histogram, probability = area!

To see this, recall (*):

$$RFD = \frac{f}{B \cdot n}.$$

Multiply by B:

$$B \cdot RFD = \frac{f}{n}.$$

$B \cdot RFD$ = area of the bar over the range in question (see histogram above)

$\frac{f}{n}$ = proportion of the data in X that lies in the given range = probability that a point in X, chosen at random, will be in that range

Area = probability!

Example.

For the above data set X, we compute:

$P(1.55 \le X < 1.65)$ = area of bars above $[1.55, 1.65)$

$= 0.05 \cdot 4.58 + 0.05 \cdot 7.08$

$= 0.583 = 58.3\%$

(the probability that a randomly chosen point from X will lie in $[1.55, 1.65)$)

For ranges that don't line up perfectly with bins, we can approximate, e.g.

$P(1.57 \le X < 1.61)$

$\approx 0.03 \cdot 4.58 + 0.01 \cdot 7.08 = 0.208 = 20.8\%$

(0.03: length of range $[1.57, 1.60)$; 0.01: length of range $[1.60, 1.61)$)