HW1: 1.2 Problem 5a, 15b

5. Use three-digit rounding arithmetic to perform the following calculations. Compute the absolute error and relative error with the exact value determined to at least five digits.

a. 133 + 0.921

Exact answer: $p = 133.921$
Three digit rounding: $p^* = 133.921 + 0.5 = 134.421 \to 134$

$$|p - p^*| = |133.921 - 134| = 0.079000$$
$$\frac{|p - p^*|}{|p|} = \frac{0.079}{133.921} = 0.000589900016 = 5.8990 * 10^{-4}$$

15. Use the 64-bit long real format to find the decimal equivalent of the following floating-point machine numbers

b. 1 10000001010 1001001100000000000000000000000000000000000000000000

$$y_{fl} = (-1)^s 2^{c-1023}(1 + f)$$

Let's fill it in

$$s = 1 \to (-1)^1 = -1$$
$$c = 10000001010 \mapsto 1 * 2^{10} + 0 * 2^9 + \cdots + 0 * 2^4 + 1 * 2^3 + 0 * 2^2 + 1 * 2^1 + 0 * 2^0$$
$$= 1034 \to 2^{1034-1023} = 2^{11} = 2048$$
$$f = 10010011000 \cdots \mapsto \frac{1}{2}^1 + \frac{1}{2}^4 + \frac{1}{2}^7 + \frac{1}{2}^8 = 0.57421875 \to (1 + 0.57421875)$$

This yields

$$y_{fl} = -1(2048)(1 + 0.57421875) = -3224$$

Which somehow I didn't realize this would be an integer