Recitation 5 - Computer Organization, Spring 2013

Problem 1. An IEEE floating point representation uses 4 *exp* bits and 5 *frac* bits.

1) How many bits are needed to store these numbers?
2) What is the bias?
3) How many denormalized values are there?
4) What is the binary representation of the smallest denormalized value that is greater than 0?
5) What is the smallest denormalized value that is greater than 0?
6) What is the binary representation of the smallest normalized value that is greater than 0?
7) What is the smallest normalized value that is greater than 0?
8) What is the binary representation of the largest normalized value?
9) What is the largest normalized value?
10) How would the number 69 be represented? (Give the answer in binary and hex.)
11) How would the number 68 be represented? (Give the answer in binary and hex.)
12) How would the number -6.25 be represented? (Give the answer in binary and hex.)
13) The bits corresponding to 0x10 are stored in a variable that represents one of the numbers. What is its value?
14) The bits corresponding to 0x34a are stored in a variable that represents one of the numbers. What is its value?

Problem 2. Assume variables *x*, *f*, and *d* are of type int, float, and double, respectively. Their values are arbitrary, except that neither *f* nor *d* equals $+\infty$, $-\infty$, or NaN . For each of the following C expressions, either argue that it will always be true (i.e., evaluate to 1) or give a value for the variables such that it is not true (i.e., evaluates to 0).

1) *x* == (int)(double) *x*
2) *x* == (int)(float) *x*
3) *d* == (double)(float) d
4) *f* == (float)(double) *f*
5) *f* == -(-*f*)
6) 1.0/2 == 1/2.0
7) *d* × d >= 0.0
8) (*f*+*d*) - *f* == *d*