MATH 111: HW 10 SOLUTIONS AND COMMENTS

7.6 #1. (I won't draw the histogram here.) The distribution is concentrated in the range 10-24, with several outlying large prices. The maximum of the distribution is 9 people in the 20-24 range.

7.6 #2. The median is 20, and the mean is approximately 20.87. The mean is higher than the median because of the outlying large prices. To answer the question "How much did a typical person with Internet access pay for that service in 2001?" it is better to give the answer that is not skewed by outliers: hence, the median.

7.6 #3. The minimum, first quartile, median, third quartile, and maximum are: 5, 12, 20, 21, and 58, as indicated here:

$$\mathbf{5}\ 9\ 10\ 10\ 10\ \mathbf{12}\ 12\ 19\ 19\ 19\ 20\ \mathbf{20}\ 20\ 20\ 20\ 21\ 21\ \mathbf{21}\ 22\ 33\ 37\ 42\ \mathbf{58}\,.$$

The summary does give a similar impression to the histogram: it is clear from the summary that there is a concentration of Internet subscribers paying in the $20-21 range, but that it is roughly as common to pay somewhat lower fees (in the $12-20 range). The spread in the lower half of the data is fairly even, but the spread in the highest quarter of the data is very large.

7.6 #5. The scatterplot indicates that roughly (but not always) the miles per gallon decreases as the weight increases.

7.6 #11. This question is largely a test of your common sense in using data. We saw roughly that miles per gallon decreases as weight increases, so it makes sense to predict that the miles per gallon for a 6000 pound car probably should be no more than 14 mpg. However, the data we're given is not useful for saying anything more precise than this.

Though the data is roughly (but only roughly) in a line, if you try extending that line you'll see that this would predict a mpg of around zero, which probably can't be right. Even if it were to give a reasonable answer, though, we shouldn't take this as a reliable predictor. A car that heavy would probably be a significantly different sort of machine than the most of the cars in the data we've been given — there's no reason to think that information about 2000 pound cars should tell us about the mileage of a 6000 pound car. (In fact, the 4000 pound car weighs significantly more than the other cars in the data. How do we know that the mileage of the 6000 pound car wouldn't just be very similar to the mileage of the 4000 pound car? We don't.) There is no strong evidence in favour of any prediction you might try to make here, other than that the miles per gallon will probably be no more than 14.

On the other hand, a 2400 pound car would be right in the middle of the data we already have, so this data would be much more useful for predicting the mileage of a 2400 pound car than a 6000 pound car.

7.7 #1. When you breed a red-flowered snapdragon with a white-flowered snapdragon, the red flower contributes a red gene and the white flower contributes a white gene. Therefore the offspring has a red gene and a white gene, and is a pink flower.

7.7 #7. The question asks about the offspring when a second-generation flower from #1 — i.e. a pink flower — breeds with a red flower. The red flower will always contribute a red gene. The pink flower contributes its red gene 50% of the time, and its white gene 50% of the time.

Therefore, 50% of the time the offspring will have two red genes, and will be red; 50% of the time the offspring will have one red gene and one white gene, and will be pink; and the offspring will never be white.

7.7 #14. Here's a more blatant version of exactly the same problem. Suppose one person takes a survey of four people, and three of them like green catsup (so 75% of the people surveyed like green catsup). Suppose another person takes a survey of 1000 people, and 250 of them like green catsup (so 25% of the people surveyed like green catsup). Is the true percentage of people who like green catsup more likely to be closer to 25% or to 75%? This is clear — it is much more likely to be closer to 25%.

Though the numbers are closer together, the situation in the textbook's problem is basically the same. Since the second survey is larger and has a much smaller margin of error, the true percentage in the whole population is likelier to be closer to the result of the second survey than to the first survey.

7.7 #18. This is an example of *post hoc ergo propter hoc*. Remember, *post hoc ergo propter hoc* refers to a specific kind of **bad reasoning**: reasoning in which one leaps to the conclusion that one event caused another, on the basis that the one event came before the other.

Although the author found the penny before winning the lottery, there is no evidence to suggest a **causal relationship** between the two: that is, there is no evidence that finding the penny actually caused the lottery win, rather than merely preceding it. This is a clear example of *post hoc ergo propter hoc*.

7.7 #19. This is also an example of *post hoc ergo propter hoc*.

In this case, the author has concluded that the dog made him sick, but his only evidence is that first the dog licked the author's hand and later the author became sick. Even though it is **possible** that this really is why the author became sick, there are also many other reasons that the author could have become sick, and it is faulty reasoning to conclude that he "must have caught some bug from that darn dog lick".