Part 1: Multiple Choice

1. B	5. C	9. D	13. B	17. D
2. B	6. C	10. D	14. D	18. C
3. A	7. D	11. C	15. C	19. B
4. C	8. C	12. B	16. A	20. D

Part 2: Short Answer

1. (a) To perform the randomization, assign each child a number, say 001, 002, 003, ..., 300. Use R (or the Table of Random Digits) to select 100 children at random. These children are assigned to Group A (Ibuprofin). Repeat for the next 100 children and assign to Group B (Acetaminophen). The remaining 100 children are assigned to Group C (Codeine). This is a completely randomized design−individuals (children) are assigned to treatments (drugs) at random (and there is no blocking).
(b) The response variable should measure how effective the drug is at reducing pain.

- One possible choice could be "Pain reduction status" (i.e., "Was your pain reduced?"); 1 = Yes, 0 = No. This would be measured on each child.
- Another possible choice could be a numerical rating of pain, taken before the drug (0-10) and after the drug (0-10). For example, if a child's pain was 8 before the drug and 2 after the drug. The numerical rating would be $8 - 2 = 6$. Positive values of this variable mean the drug is "working." Zero values or negative values mean the drug is not working or making things worse.

(c) The 1/0 variable above (pain reduction status) is **categorical**−individuals are simply put into categories (1 = pain was reduced; 0 = pain was not reduced). The numerical rating variable in part (b) is **quantitative**. The variable takes on different numerical values which imply different amounts of physical pain.

2. Here are the answers to the questions in each bullet:

- Dr. Potti proposed that cancer treatments could be tailored specifically to an individual's tumor and its genetic makeup. Therefore, instead of giving a general treatment to a population of cancer patients, a treatment would be based on each individual's specific cancer characteristics. Every patient's DNA is different, so different treatments should be used for different patients.
- To make his research hypothesis seem like it was "working," Dr. Potti manipulated the patients' data so that they would favor his approach. When a patient's data didn't support his theory, he changed the data.
- Patients were getting ineffective or dangerous treatments; some patients died; Dr. Potti "resigned" from Duke; i.e., he was fired; all of his publications were retracted by Duke, Dr. Potti/Duke was sued, etc.

3. (a) Innumeracy is a lack of understanding of basic mathematics (e.g., how large numbers are, etc.).
(b) One famous example we talked about in class was how "the number of children gunned down in the United States" doubles every year. To innumerate people, this may sound plausible. However, this claim is ridiculous.
(c) Dr. Best's book is about the media; e.g., how statistics are reported in the media, how statistics can be used to trick innumerate readers into accepting a certain social outcome (e.g., 41% of biology teachers in Louisiana...., 31 million Americans hungry, etc.), how members of the media themselves are innumerate, etc.

4. (a) For this histogram,

- The **center** of the distribution is around 3.25 lb.
- The **spread** in the distribution goes from about 2.0 lb to 4.5 lb. No cattle have gains outside this range.
- The **shape** of the distribution is **symmetric**.
- There are no outliers. In this example, a clear outlier would be a cattle whose weight gain was 10 lb or −5 lb (i.e., the cattle actually lost weight).

(b) The population density curve would be a smooth curve that "approximates" the histogram (see examples in Chapter 11 in the notes). A population density curve (in this example) would describe the average daily weight gain for all cattle in the population. The histogram is only for the SRS of 200 cattle.

5. (a) First of all, the vertical axis is not labeled (although the title says what the numbers "6,000,000" and "7,066,000" mean). The height of the second bar looks like it is 2-3 times larger than that of the first bar. However, if you calculate the **percentage increase** in enrollment, for part (b), we get

$$\begin{aligned}\text{percentage change} &= \frac{\text{amount of change}}{\text{starting value}} \times 100\% \\ &= \frac{1{,}066{,}000}{6{,}000{,}000} \times 100\% \\ &\approx 0.178 \times 100\%\end{aligned}$$

or about 17.8%. So, the March 31 enrollment is only a 17.8% increase when compared to March 27. It is not over a 100% increase, which is what the bar heights suggest. Misleading!